

# Om lærepladsspørgsmålet

Karsten Albæk

Økonomisk Institut, Københavns Universitet, E-mail: Karsten.Albaek@econ.ku.dk

*SUMMARY: The paper gives an overview of the development of the education of apprentices in Denmark since 1931. The development during the late 1990ies was very favourable, as about 45 per cent of a youth cohort began an education as apprentices. It is the highest registered share over a longer time span. As a consequence of the slump, the share dropped from 46 to 41 per cent from 2001 to 2002. The drop on 5 percentage points corresponds to about 3000 young persons, who cannot obtain the education at the employer that they wish. Quantitative analyses are carried out in order to trace the factors that influence the inflow of apprentices. An evaluation is conducted of the different proposals to alleviate the lack of training slots for apprentices.*

---

## 1. Indledning

For øjeblikket er virksomhederne i Danmark meget påholdende med at ansætte ny arbejdskraft, og det er bl.a. til ulempe for grupper, der skal ud på arbejdsmarkedet for første gang. Et særligt alvorligt problem i denne forbindelse er de unge, der søger en læreplads, idet en manglende læreplads kan betyde, at de ikke færdiggør en erhvervsuddannelse.

Mangel på lærepladser kan forsøges afhjulpet på forskellig vis, og dette indlæg indeholder en gennemgang af fordele og ulemper ved de forskellige metoder. En sådan vurdering kræver imidlertid, at den aktuelle situation vurderes i forhold til de tidligere erfaringer på området. Artiklen giver således også et overblik over den overordnede udvikling på området og et forsøg på at forklare udviklingen, bl.a. set i relation til de forskellige offentlige indgreb. Tankegangen er at prøve at se på de generelle træk i udviklingsforløbet, som let drukner i detailobservationer på dette komplekse uddannelsesområde.

Der gennemføres kvantitative analyser til at spore effekten på lærlingetilgangen af forskellige forklarende faktorer som konjunkturudvikling, ungdomsårgange og omkostninger ved at beskæftige lærlinge. Et hovedresultat er, at omkostningerne har en væsentlig betydning i forbindelse med arbejdsgivernes efterspørgsel efter lærlinge.

---

Tak til en referee og til tidsskriftets redaktør for kommentarer. Endvidere tak til Heino B. Nielsen for nyttige kommentarer til de økonometriske aspekter af manuskriptet og til Henrik Hansen for diskussioner om emnet i forbindelse med et tidligere samarbejde på området. Artiklen er en revideret og forkortet udgave af Albæk (2004). Der henvises til dette manuskript for gennemgang af datagrundlag og -oparbejdning.

Det gør det muligt at påvirke lærlingetilgangen via indgreb, som ændrer omkostningerne ved at beskæftige lærlinge.<sup>1</sup>

Til afhjælpning af manglen på lære- og praktikpladser blev det i 1990 vedtaget at iværksætte praktikpladskompenserende undervisning, og ordningen er senere gjort permanent. Der er imidlertid flere problemer forbundet med denne undervisning, og regeringen har ved årsskiftet 2003-04 besluttet at reducere adgangen til skolepraktik, hvis der ikke findes en læreplads hos en arbejdsgiver. Det er derfor ønskeligt at se på alternativerne, idet en reduktion af omfanget af skolepraktik isoleret set ikke bidrager til at flere unge gennemfører en erhvervsuddannelse.

Indlægget er disponeret som følger. I afsnit 2 gennemgås udviklingen i centrale størrelser på området, så som tilgang af lærlinge, ungdomsårgange og omkostninger. Diskussionen om lærepladser bygger i et vist omfang på formodninger, som forekommer mindre plausible, når man betragter udviklingen over en længere periode. Årsagerne til udviklingen på området vil blive belyst, så langt det nu er muligt, ved alene at sammenholde udviklingen i de enkelte tidsserier. I afsnit 3 gennemgås resultaterne af forsøgene på at kvantificere konsekvenserne af de forskellige forhold, der bestemmer tilgangen af lærlinge. Lærlingenes alder og produktivitet behandles i afsnit 4. I afsnit 5 foretages en diskussion af de forskellige løsningsmetoder. Der afrundes i afsnit 6 med en konklusion, der omtaler lærepladsproblemet i relation til den aktuelle arbejdsmarkedspolitiske debat.<sup>2</sup>

## 2. Udviklingen på lærlingeområdet

I dette afsnit gennemgås udviklingen i centrale størrelser på området. Sigtet er at give et overblik over udviklingen på området, samt diskutere årsagerne til udviklingsforløbet. De størrelser, der behandles, er lærlingetilgang og beskæftigelse, lærlingelønninger før og efter tilskud, ungdomsårgangenes størrelse og den andel af en ungdomsårgang, som fuldfører en lærlingeuddannelse.

Udviklingen illustreres ved hjælp af grafer, mens de forholdsvis omfattende data-

---

1. Tidligere arbejder om de danske lærlingeuddannelser inkluderer den klassiske undersøgelse Arbejds- og Socialministeriet (1945) samt det monumentale værk af Houman Sørensen m.fl. (1983), som har en bredere tilgang til emnet. På mikrodata findes Bjørn (1996) samt Westergaard-Nielsen og Rasmussen (1999), som behandler fordelingen af lærlinge på forskellige typer af arbejdsgivere.

2. Fra udgivelsen af Albæk (2004) til revisionen af manuskriptet i forbindelse med publiceringen i Nationaløkonomisk Tidsskrift er der sket en væsentlig nyorientering i politikken på området. Det hedder således i regeringsgrundlaget, VK Regeringen II (2005), s. 26. »Regeringen vil [...] inden for rammerne af Arbejdsgivernes Elevrefusion (AER) foreslå at etablere en ordning, hvor virksomheder, som opretter ekstra praktikpladser, modtager et kontant beløb som anerkendelse herfor. Den træder i kraft umiddelbart.« På tidspunktet for revisionen af manuskriptet er det ikke forfatteren bekendt, at der er gennemført en sådan ordning. Det er valgt at bibeholde manuskriptets karakter af problem- og løsningsidentifikation. Så meget mere, som det vel ikke kan forventes, at alle på en gang bifalder nyorienteringen af politikken på området.

serier er medtaget i Albæk (2004). Her er også gjort rede for kilder og oparbejdning af data, som derfor alene omtales i teksten, hvis det skønnes af væsentlig betydning for forståelsen.

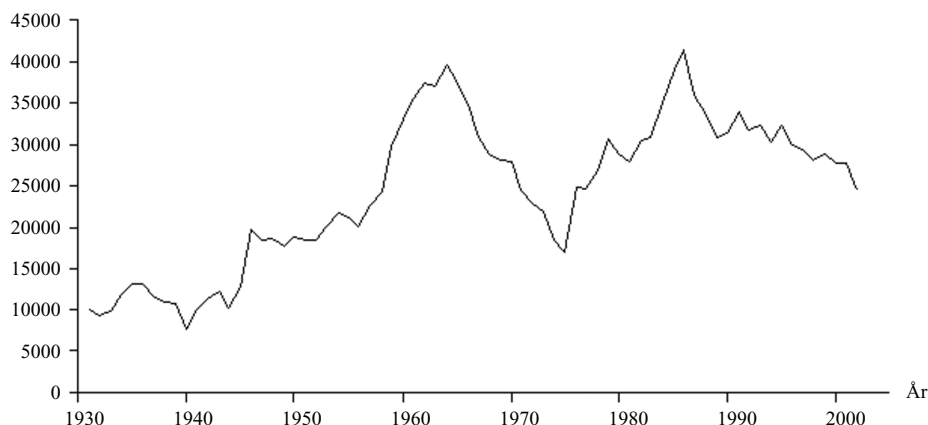
Først ses der på *tilgangen af lærlinge*. Der ses alene på tilgang af lærlinge til arbejdsgivere, idet det er forsøgt at holde skolepraktik ude af opgørelserne. Anledningen er, at det er arbejdsgivernes beslutninger om at ansætte lærlinge og besætte ledige praktikpladser, som er det centrale ved overvejelser om indgreb til at fremskaffe flere lærepladser.

Af figur 1 fremgår det, at der har været ganske betydelige svingninger i tilgangen af lærlinge. Fra begyndelsen af 1990'erne har der været en faldende tendens i tilgangen fra lidt over 30.000 til ca. 25.000 i år 2002. I midten af 1980'erne var tilgangen helt oppe på lidt over 40.000, men et sådant kraftigt udsving er imidlertid ikke usædvanligt, når udviklingen over en længere periode betragtes. I begyndelsen og midten af 1960'erne var man også oppe omkring 40.000 i tilgang, hvorefter tallet over en årrække blev mere end halveret til et niveau under 20.000 i midten af 1970'erne. Længere tilbage i tiden var niveauet lavere, idet erhvervsstrukturen var mere landbrugsdomineret. I 1930'erne og begyndelsen af 1940'erne svingede tilgangen omkring 10.000, for at vokse til næsten 20.000 i slutningen af 1940'erne og 1950'erne.

Af figuren fremgår det, at konjunkturerne har en mærkbar indflydelse på tilgangen af lærlinge. Ved de store konjunkturtilbageslag i 1955-56 og 1963 faldt lærlingetilgangen, og det samme gjorde sig gældende både efter den første og den anden »oliekrise« i henholdsvis 1974-75 og 1979-81.

Beskæftigelsen af lærlinge svarer cirka til tilgangen ganget med den gennemsnitlige læretid, og en graf over lærlingebeskæftigelsen har nogenlunde det samme forløb over tiden som tilgangen af lærlinge blot på et højere niveau. Tilgangen og dermed også beskæftigelsen af lærlinge fluktuerer betydeligt mere end beskæftigelsen af andre lønmodtagere, og lærlingenes andel af den samlede beskæftigelse har således bevæget sig betydeligt over tiden. Således udgjorde lærlingebeskæftigelsen ca. 5 pct. af arbejdsstyrken ved toppunktet i midten af 1960'erne, mens den tilsvarende andel i slutningen af 1970'erne var 2,5-3 pct.

Ud fra denne udvikling kan det udledes, at der i perioden har været en ganske betydelig fleksibilitet hvad angår tilgang og beskæftigelse af lærlinge. I nogle perioder er der en meget stor tilgang og beskæftigelse af lærlinge, i andre perioder en forholdsvis begrænset tilgang og beskæftigelse. To forhold er bemærkelsesværdige. For det første størrelsen af udsvingene, idet der næsten er tale om fordoblinger og halveringer af tilgang og beskæftigelse af lærlinge i forhold til andre grupper på arbejdsmarkedet. For det andet hastigheden, ændringerne fra top til bund foregår inden for en periode på 5-10 år.



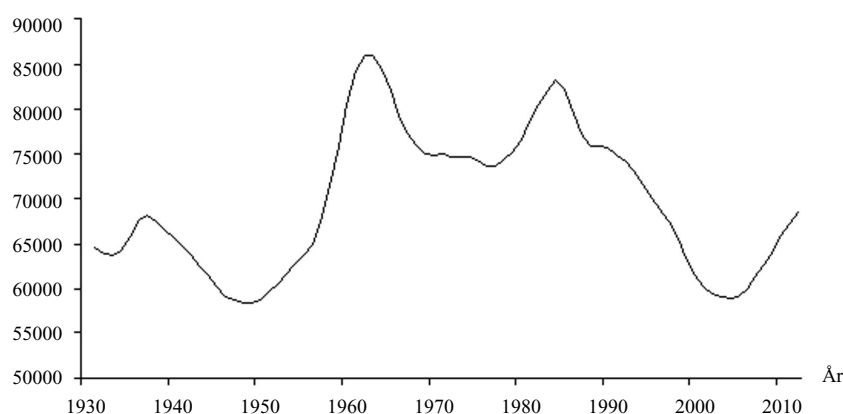
Figur 1. Tilgangen til lærlingeuddannelserne, 1931-2002.

I debatten om lærepladsproblemet hævdes det undertiden, at arbejdsgiverne ikke har mulighed for at beskæftige flere elever, at der ikke er plads til de unge uddannelsessøgende ude på arbejdspladserne. Det fremgår af udviklingen på området, at denne påstand næppe er holdbar. I midten af 1960'erne var én ud af 20 lønmodtagere under uddannelse, så faldt andelen til én ud af 40, for herefter at vokse og så falde igen.

Hidtil har det danske arbejdsmarked altså vist en betydelig absorberingsevne. Der har tidligere været beskæftiget langt flere lærlinge, end tilfældet er nu. På denne baggrund må det antages, at der også fremover vil være en betragtelig absorberingsevne til stede, at der er plads til betydeligt flere lærlinge end der beskæftiges for øjeblikket.

Det er naturligt at sætte antallet af unge, som får en læreplads, i relation til *ungdomsårgangenes størrelse*. Både fordi dette afspejler den chance, de unge har for at opnå en læreplads, men også fordi dette siger noget om tilgangen af faglært arbejdskraft. I figur 2 er indtegnet et mål for ungdomsårgangenes størrelse på de alderstrin, hvor en lærlingeuddannelse påbegyndes. Betegnelsen ungdomsårgang dækker således over et vejet gennemsnit af de alderstrin, hvor lærlingene opnår læreplads.

Fra et niveau på godt lidt over 65.000 i slutningen af 1930'erne falder ungdomsårgangene til knap 60.000 i slutningen af 1940'erne, og vokser herefter kraftigt til godt 85.000 i begyndelsen af 1960'erne. Der indtræder et fald til et niveau på ca. 75.000 i perioden fra slutningen af 1970'erne. Nye store årgange på knapt 85.000 træder ind på arbejdsmarkedet i begyndelsen af 1980'erne, mens niveauet i slutningen af 1980'erne falder til knap 75.000. I 1990'erne kommer meget små årgange ud på arbejdsmarkedet, og niveauet ved årtusindeskiftet ligger på knap 60.000. Det er således et ganske betydeligt fald i ungdomsårgangene gennem 1990'erne, med dette vender igen hen mod slutningen af det første årti i dette årtusinde.

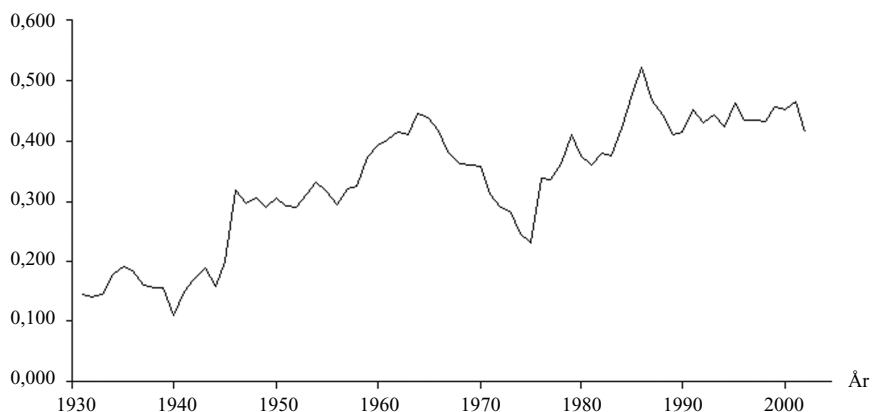


Figur 2. Ungdomsårgangene, 1931-2012.

Ved sammenligning af figur 1 og figur 2 fremgår det, at der er et sammenfald mellem det tidspunkt, hvor ungdomsårgangene topper i begyndelsen af 1960'erne, og det tidspunkt, hvor tilgangen af lærlinge når sit maksimum. Og da de næste store ungdomsårgange kom ud på arbejdsmarkedet i begyndelsen af 1980'erne, ligger lærlingetilgangen også på et relativt gunstigt niveau sammenlignet med de foregående år. Det samme gør sig gældende ved udviklingen fra begyndelsen af 1990'erne til begyndelsen af det nye årtusinde. Der er både tale om et fald i ungdomsårgangene og lærlingetilgangen. I store træk er bevægelserne i de to kurver identiske, dog med visse niveauforskydninger – små ungdomsårgange falder sammen med lav lærlingetilgang, mens store ungdomsårgange falder sammen med stor lærlingetilgang.

I midten af 1980'erne var der visse forhåbninger om, at lærepladsproblemet var løst. Hvis det store tilgangstal kunne holdes, når ungdomsårgangene faldt i slutningen af 1980'erne, ville en langt større andel af de unge kunne få en faglig uddannelse. Imidlertid faldt tilgangstallet, som vi netop har set, med ca. 25 pct., samtidig med at ungdomsårgangene blev mindre. Som det fremgår af figur 2, står vi over for en ny vækst i størrelsen af ungdomsårgangene i det næste årti. På baggrund af den hidtidige udvikling er der imidlertid ingen grund til at forvente, at dette skulle gøre det vanskeligere for de unge at få en læreplads.

Konsekvenserne af bevægelserne i ungdomsårgange og lærlingetilgang afspejler sig i *andelen af en ungdomsårgang med lærlingeuddannelse*. I figur 3 er indtegnet forholdet mellem lærlingetilgangen og ungdomsårgangen, dvs. tallene i figur 1 divideret med tallene i figur 2. Et iøjnefaldende træk er niveauforskellen fra før 1946, hvor andelen af unge med en lærlingeuddannelse udgjorde ca. 15 pct., til niveauet fra 1946 og fremefter, hvor andelen ligger omkring 30 pct. Et højdepunkt kom i midten af



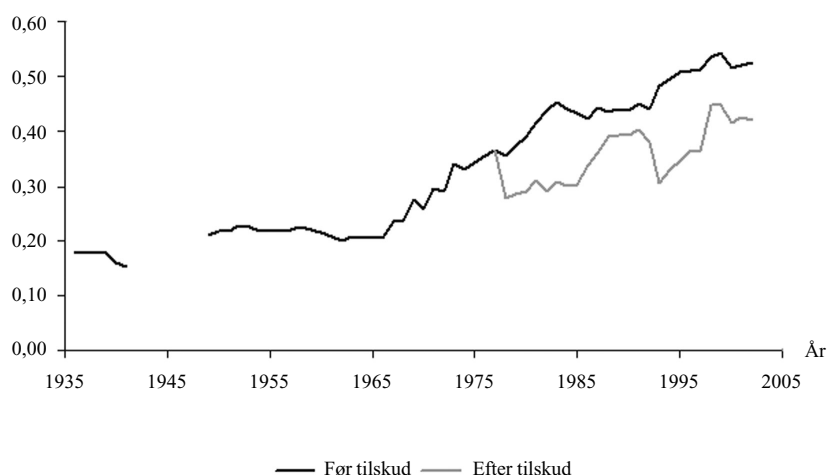
Figur 3. Tilgangen til lærlingeuddannelserne som andel af ungdomsårgang, 1931-2002.

1960erne med en andel på over 40 pct. efterfulgt af et kraftigt fald til under 25 pct. i midten af 1970erne. I begyndelsen af 1980erne lå andelen af unge, der begyndte en lærlingeuddannelse, på ca. 35 pct. for så at vokse til omkring 50 pct. i midten af 1980erne, hvorefter et fald satte ind.

Udviklingen i 1990erne er imidlertid bemærkelsesværdig, idet der er tale om en svag vækst i andelen op til 46 pct. i 2001: Der har ganske vist været et fald i tilgangen, men faldet i ungdomsårgangen har været endnu større, så andelen af en ungdomsårgang, der begynder en lærlingeuddannelse, er vokset. Et par gange har andelen af en ungdomsårgang med lærlingeuddannelse været højere end i slutningen af 1990erne, men det er første gang, at andelen er fastholdt på et så højt niveau over en længere årrække. Dette positive mønster brydes imidlertid i forbindelse med den aktuelle stagnation i økonomien, og det sidste år i observationsperioden viser således et fald i andelen fra 2001 til 2002 på hele 5 procentpoint. Som det fremgår af grafen, er de 41 pct. i tilgang for 2002 imidlertid ganske højt set i historisk perspektiv.

Det skal endvidere fremhæves, at andelen af unge, der har begyndt et læreforhold i de senere år, sandsynligvis er undervurderet i figur 3. Undervisningsministeriet har i de senere år udarbejdet en alternativ opgørelse over omfanget af læreforhold, som i modsætning til opgørelserne i figur 1 og 3 også omfatter de skolepraktikanter, der i løbet af skolepraktikken opnår en almindelig læreplads. Denne opgørelse ligger cirka 10 pct. over tallene fra Danmarks Statistik, svarende til en forøgelse på næsten 5 procentpoint i andelen af unge, der får en læreplads. For at give et retvisende billede af den tidsmæssige udvikling på området, er der imidlertid undladt at korrigere for dette forhold, jævnfør forklaringen i Albæk (2004), appendiks C.

I 1950erne så man med en vis ængstelse på den forestående store tilgang af unge til



Figur 4. Lærlingelønnens andel af voksenlønnen før og efter tilskud, 1936-2002.

arbejdsmarkedet, hvilket gav sig udtryk i debatten om de store årgange, hvor et vægtigt bidrag var flere kommissionsbetænkninger. Da de store årgange så dukkede op, viste der sig imidlertid ingen problemer.

For lærlingeuddannelsens vedkommende skulle problemer med store årgange give sig udtryk i en relativ lav andel af ungdomsårgangene med en lærlingeuddannelse, mens ingen problemer med store årgange skulle resultere i en konstant andel. Af figur 3 sammenholdt med figur 2 fremgår imidlertid det paradoksale fænomen, at bortset fra den seneste udvikling, så har andelen af unge, som gennemfører en lærlingeuddannelse, været størst, når ungdomsårgangene er store. De store årgange i begyndelsen af 1960erne opnåede som nævnt en lærlingeandel på lidt over 40 pct., og tilsvarende har de store årgange i 1980erne en lærlingeandel på næsten 50 pct., mens andelen i de omkringliggende perioder var mindre.

Som forklaring på dette fænomen kan nævnes, at de store årgange har været så usædvanligt heldige at komme ud på arbejdsmarkedet i højkonjunkturperioder. Både midten af 1960erne og midten af 1980erne var karakteriseret ved en relativ kraftig økonomisk vækst. Undtagelsen er som nævnt den seneste udvikling, hvor den aktuelle ungdomsårgang af begrænset størrelse har haft en høj andel, der begynder en lærlingeuddannelse. Andelen af en ungdomsårgang, som fuldfører en lærlingeuddannelse, er lidt lavere end tallene i figur 3 som følge af frafald under uddannelsen.

Et væsentligt element i forbindelse med udviklingen på lærepladsområdet er *omkostningsudviklingen*. Det er ikke kun lønningerne, som har spillet en rolle, idet forskellige tilskudsordninger har været gældende.

Ligesom de øvrige tidsserier fremviser også de relative lærlingelønninger store varia-

tioner, hvilket ses af figur 4. I forhold til gennemsnits aflønningen for en voksen arbejder udgjorde lærlingelønningerne lidt over 20 pct. i perioden fra 1949-66, men med en svagt faldende tendens fra slutningen af 1950erne. Før 1949 er der statistik på Dansk Arbejdsgiverforenings område for perioden 1936-41, hvor niveauet var lidt lavere, 15-20 pct. af gennemsnits aflønningen inden for industri og håndværk.

Fra 1967 sætter en betydelig stigning ind med et foreløbigt højdepunkt på 45 pct. i 1983, efterfulgt af et moderat fald. En vis del af denne udvikling kan tilskrives forskydninger i lærlingebestanden. Lønsatserne er gradueret efter læreår, men kan også være gradueret efter alder, typisk med en lavere aflønning til lærlinge under 18 år, og da der inden for flere fag blev indført aldersgraderede skalaer i 1970erne, vil alene det forhold, at lærlingene er blevet ældre, indebære en forøgelse af gennemsnitslønnen. Det samme gælder indførelsen af efg, idet efg-eleverne, der for alvor begyndte at betyde noget i slutningen af 1970erne, typisk sprang over 1. løntrin på skalaen, som lå betydeligt under lønnen i de senere læreår.

Langt den største del af lønstigningerne kan dog tilskrives satsforhøjelser. Ved at betragte figur 4 ses det, at forløbet fra midten af 1960erne er karakteriseret ved store relative lønforbedringer hvert andet år efterfulgt af mindre relative formindskelser. Toppunkterne i figuren ligger i overenskomstårene, og satsforhøjelser i overenskomsterne registreres fuldt ud, idet lærlingelønningerne opgøres i årenes tredje kvartaler. Perioden må således have været karakteriseret ved kraftige relative forhøjelser i overenskomstsatserne efterfulgt af en relativ formindskelse i løbet af overenskomstperioden som følge af mindre relativ vækst i de to andre lønkomponenter, dyrtidsregulering og lønglidning. En mere direkte vurdering er foretaget ved at gå ind i overenskomsterne for et par udvalgte fag og beregne forhøjelserne i de enkelte satser, og denne fremgangsmåde bekræfter, at satsforhøjelserne har spillet en dominerende rolle.<sup>3</sup>

Dette synes bekræftet af udviklingen i 1980erne, hvor reduktionen i de relative lærlingelønninger falder sammen med den borgerlige regerings indkomspolitik i begyndelsen og midten af 1980erne. På lærlingeområdet synes der kun at være begrænset lønglidning, og satsændringer ved overenskomsterne spiller en altdominerende rolle. Efter slækkelsen af lønrammerne genoptages det gamle mønster, der ses mindre toppe i overenskomstårene 1987, 1989 og 1991.

Året 1993 udviser en brat stigning på 4,2 procentpoint i den relative lærlingeløn og dette efterfølges af yderligere vækst, således at lønnen til lærlinge ifølge figur 4 skulle udgøre mere en 50 pct. af lønnen til voksne arbejdere i de seneste observationsperioder. Desværre er der problemer med opgørelserne fra og med 1993 med den konsekvens, at der kan sættes spørgsmålstegn ved, om udviklingen i denne periode som an-

---

3. Der er tale om opgørelser af satsstigninger for tømrerlærlinge, smedelærlinge og HK-lærlinge fra 1965 til 1983, se Albæk (1988), s. 16. En sådan opgørelse er ikke gennemført på mere aktuelle data.



givet i figur 4 giver et dækkende billede af udviklingen. For det første indførte Dansk Arbejdsgiverforening et nyt statistiksystem fra og med 1993, og dette vanskeliggør sammenligninger af lønniveauerne før og efter 1993. For det andet er 1990'erne karakteriseret ved en vækst i andelen af voksenlærlinge med en alder på 25 år og derover, og denne kategori har typisk en højere løn end lærlinge under 25 år. I Albæk (2004), appendiks C, er der forsøgt at danne et skøn over størrelsesordenen ud fra de forhåndenværende oplysninger, og ifølge disse beregninger kan effekten af det forøgede antal voksenlærlinge nå op i nærheden af en vækst i den relative lærlingeløn på 4 procentpoint. Hvis det lægges til hoppet på 4 procentpoint i 1993 får man det resultat, at næsten hele væksten i den relative lærlingeløn efter 1992 kan tilskrives aspekter af den måde, lærlingelønningerne opgøres på, og det er således uklart i hvilket omfang, der har været tale om en vækst efter 1992.

Fra 1978 blev tilskuddene indført: for det første et tilskud til refusion af arbejdsgivernes lønudgifter under lærlingenes og elevernes skoleophold (AER – arbejdsgivernes elevrefusion), og for det andet et tilskud til oprettelse af ekstraordinære lærepladser. Det sidste var først på 15.000 kr. pr. læreforhold og 16.000 kr. fra 1979, men blev i 1982 forhøjet til 30.000 kr. samtidig med, at tilskudskriteriet blev ændret.

Tilskuddene har betydet, at arbejdsgivernes omkostninger har været mindre end lønnen, som illustreret ved den nederste kurve i figur 4, hvor forskellen op til den øverste kurve er tilskuddet for de arbejdsgivere, der har fået fuldt tilskud.<sup>4</sup> For perioden 1982-85 var der tale om, at knap en tredjedel af lønomkostningerne blev dækket af tilskud. Fra 1986 blev tilskuddet til ekstraordinære lærepladser halveret til 15.000 kr., for at blive fjernet pr. 31. december 1987, og omkostningskurven bevæger sig således brat opad. De ekstraordinære tilskud blev indført igen under lavkonjunkturen i begyndelsen af 1990'erne med et beløb på 20.000 kr. per elev i 1993, 16.000 kr. per elev per år, der blev ansat i 1994 eller 1995, samt 12.000 kr. per elev ansat i 1996, hvorefter tilskuddet blev afskaffet.<sup>5</sup>

Et første indblik i årsagerne til variationerne i lærlingetilgangen kan fås ved at sammenligne udviklingen i tilgangen i figur 1 med lønudviklingen i figur 4. Det fremgår, at der er tale om en vis samvariation. Den kraftige beskæftigelsesvækst frem til midten af 1960'erne fandt sted samtidig med et – omend ganske svagt – fald i de relative lærlingelønninger. Halveringen af beskæftigelsen i det følgende årti falder sammen

4. Ifølge Arbejdsdirektoratets lære- og praktikpladsopgørelser er det for perioden 1983-86 af størrelsesordenen 2/3 af samtlige lære- og praktikpladser, der har fået det ekstraordinære tilskud.

5. Tilbage er alene AER-tilskuddet, hvis størrelse illustreres af forskellen mellem de to kurver. Man kan diskutere, om det er den rigtige måde at illustrere AER-tilskuddet på, idet timeomkostningerne i produktionen for arbejdsgiverne snarere udgør den udbetalte timeløn, når omkostningerne ved skoleophold dækkes. Alternativet ville være at forrykke hele den øverste kurve opad med en faktor svarende til udgifterne ved skoleophold.

med en fordobling af den relative lærlingeløn, og beskæftigelsesforbedringen i slutningen af 1970erne indtræder parallelt med den tilskudsfinansierede omkostningsformindskelse.

Tilsvarende er der et sammenfald mellem en særdeles gunstig udvikling i lærlingetilgangen i den første halvdel af 1980erne og et moderat niveau for lærlingeomkostningerne sammenholdt perioderne før og efter. Det drastiske fald i lærlingetilgangen i slutningen af 1980erne falder sammen med en betragtelig vækst i omkostningerne. Fra 1993 og fremefter holdt lærlingetilgangen sig oppe på et pænt højt niveau på trods af lavkonjunktoren præget af endog særdeles høje arbejdsløshedsprocenter set i historisk perspektiv.

Udgangspunktet for store dele af overvejelserne omkring lærepladsspørgsmålet er de mange unge, som ikke får en plads. Det er derfor relevant at se på størrelsesordenen af *søgningen til lærlingeuddannelserne*. Der henvises til behandlingen i Albæk (2004) for behandling af dette emne, inklusive registreringsprocedurerne. Hovedkonklusionen er, at der synes at have været et overskud af lærepladssøgende for de tidsrum, hvor man har opgørelser.<sup>6</sup>

Gennemgangen i dette afsnit har tilsigtet at give et overblik over udviklingen i forskellige forhold, der er afgørende for omfanget af lærlingeuddannelserne. Endvidere er det forsøgt at give et indblik i samspillet mellem disse forhold. Det synes klart, at man for at forklare *hovedtendenserne på området* må inddrage tre forhold: konjunkturerne, ungdomsårgangenes størrelse samt omkostningerne.

Når virksomhederne får flere ordrer ind, opretter de også flere lærepladser, og når den økonomiske aktivitet dæmpes, falder antallet af uddannelsespladser. Sammenhængen fremgår klart ved at gå tilbage i tiden og se på forløbet ved større tilbageslag i økonomien.

Når ungdomsårgangene er store, ansætter virksomhederne mange lærlinge, og når ungdomsårgangene er små, tager virksomhederne få lærlinge ind. Det har i det mindste været sammenhængen indtil nu, hvor vi har to en halv bølge af ungdomsårgange at bygge på: de store ungdomsårgange i begyndelsen af henholdsvis 1960erne og 1980erne, og den lille ungdomsårgang omkring årtusindeskiftet.

Færre lærlinge ansættes, når lønomkostningerne er høje – lave omkostninger giver mange lærlinge. Tilgangsformindskelsen gennem 1970erne faldt sammen med stigende omkostninger. I de gunstige år i 1980erne blev omkostningerne holdt nede ved tilskud, mens tilgangen faldt drastisk efter afskaffelsen af tilskuddene i slutningen af 1980erne. Lavkonjunktoren i begyndelsen og midten af 1990erne med arbejdsløshed

---

6. I Albæk (2004) gøres der også opmærksom på, at de nuværende registre synes at være underudnyttede, hvad angår analyser af ventetider i køen af lærepladssøgende.

over 10 pct. er ikke til at spore i lærlingetilgangen, men her blev omkostningerne også holdt nede med ekstraordinære tilskud.

### 3. Estimationsresultater for lærlingetilgangen

I det indledende afsnit blev der givet et overblik over udviklingen i forskellige forhold, der er afgørende for omfanget af lærlingeuddannelser, samt et indblik i samspillet mellem disse forhold. En nøjere vurdering af dette samspil forudsætter imidlertid en præcisering af de sammenhænge, der antages at gælde på området.

I det følgende formuleres en empirisk model til beskrivelse af efterspørgselen efter lærlinge, hvorefter estimation af modellen gennemføres. Tidligere er der i Albæk (1992) gennemført analoge estimationsresultater for udviklingen i lærlinge-beskæftigelsen.

Estimationerne tager teoretisk udgangspunkt i en variant af en to-faktor CES-produktions-funktion, hvor beskæftigede lærlinge og andre arbejdere producerer output

$$Q = [\delta(e^{\mu t} N)^{-\rho} + (1 - \delta)(e^{\lambda t} A^b)^{-\rho}]^{-1/\rho}, \quad (1.1)$$

hvor  $t$  er tiden,  $\mu$  og  $\lambda$  er faktorspecifikke tekniske fremskridt,  $\delta$  angiver fordelingsparameteren, og substitutionselasticiteten er givet ved  $\sigma = 1/(1 + \rho)$ .

I ligevægt antages relationen mellem tilgang af lærlinge  $A$  og bestand af lærlinge  $A^b$  at kunne beskrives ved

$$A^b = A \exp(\alpha + \beta t), \quad (1.2)$$

hvor faktoren til  $A$  på højresiden repræsenterer den gennemsnitlige læretid.

Første ordens betingelsen for tilgangen af lærlinge resulterer i følgende ligning, der beskriver efterspørgselen efter lærlinge

$$\ln A = \sigma \ln(1 - \delta) - \alpha + \ln Q - \sigma \ln w_A^n + ((\sigma - 1)\lambda - \beta)t, \quad (1.3)$$

hvor  $w_A^n$  er lærlingenes realløn efter fradrag af tilskud. Anvendelse af en tilsvarende betingelse for voksne arbejdere resulterer i en ligning, der beskriver den relative efterspørgsel efter lærlinge

$$\ln \frac{A}{N} = \sigma \ln \frac{1 - \delta}{\delta} - \alpha - \sigma \ln \frac{w_A^n}{w_N} + ((\sigma - 1)(\lambda - \mu) - \beta)t, \quad (1.4)$$

hvor  $w_N$  er reallønnen for andre arbejdere.

Ved estimation af førsteordensbetingelsen (1.3) indgår bruttonationalproduktet i faste priser som forklarende variabel, og lønnen til lærlinge er deflateret med den implicitte BNP-deflator. Som forklarende variabel er endvidere medtaget af ungdomsårgangen,  $U$ . Ved estimation af aldersspecifik beskæftigelse er det en typisk procedure at inkludere størrelsen af befolkningen i de relevante aldersklasser. Ungdomsårgangen leaved én periode har den bedste forklaringskraft i regressionerne og er derfor anvendt. Denne variabel er konstrueret af forfatteren, og årsagen til periodeforskydningen kan enten tilskrives målefejl ved konstruktionen (aldersbetingede overgangsfrekvenser er ikke observerede i perioden 1957-74), eller en tidsforskydning i effekten fra ungdomsårgangenes størrelse. Logaritmen til alle de nævnte variable er  $I(1)$  variable ifølge Dickey-Fuller tests. En regression af  $A$  på  $Q$ ,  $w_A^n$ ,  $U$  og en trend giver residualer, som kan afvises at være  $I(1)$ . En regression af ændringerne i residualerne fra ligningen på ændringerne lagget én og tre perioder samt det laggede niveau af residualerne giver koefficienter, som accepteres at være forskellige fra nul på et 5 procents signifikansniveau med anvendelse af normalfordelingens kritiske værdier. Koefficienten til den laggede værdi af residualerne har en  $t$ -værdi på  $-4,54$ , og McKinnons 5 procents kritiske værdi for afvisning af nulhypotesen om tilstedeværelsen af en enhedsrod er  $-4,12$  i dette tilfælde.

Den foretrukne fejlkorrigeringsmodel for efterspørgselen på baggrund af førsteordensbetingelsen for efterspørgsel efter lærlinge er følgende:

$$\Delta \ln A = -2,702 - 0,291 \ln(w_A^n)_{-1} - 0,278(\ln A - \ln Q)_{-1} + 0,348 \ln U_{-1} + 0,005t \quad (1.5)$$

(1,284) (0,115) (0,103) (1,140) (0,003)

$$+ 0,487 \Delta \ln w_A^n + 0,242 \Delta \ln A_{-1} + 1,400 \Delta \ln Q + 0,323 D76$$

(0,111) (0,095) (0,451) (0,072)

$$R^2 = 0,661 \quad \hat{\sigma} = 0,060 \quad T = 1950 - 2002$$

$$F_{ar}(2,42) = 2,674 \quad F_{arch}(1,42) = 0,806 \quad \chi_{nd}^2(2) = 0,818$$

$$F_{hc}(15,28) = 0,764 \quad F_{RESET}(1,43) = 5,375$$

Koefficienten til  $Q$  er restrikeret til én som en implikation af antagelsen om homogenitet af produktionsfunktionen (1.1), og denne restriktion accepteres på et 5 procents signifikans niveau i niveauregressionen. Alle koefficienterne er signifikant forskellige fra nul. Der er medtaget en række tests for misspecifikation, som med en enkelt undtagelse angiver, at fravær af problemer ikke kan forkastes.  $F_{ar}$  er et F-test for anden ordens autokorrelation,  $F_{arch}$  er et F-test for ingen autokorreleret betinget heteroskedasticitet,  $\chi_{nd}^2$  et chi i anden test for normalitet af residualerne,  $F_{hc}$  et test for ingen heteroskedasticitet,  $F_{RESET}$  og er Ramseys test for misspecifikation. Sidstnævnte test

angiver, at der er problemer med modelspecifikationen, men det er ikke tilfældet, hvis nationalproduktet indgår i ikke-restrikeret form.

Division af koefficienten til lønnen med koefficienten til fejlkorrektionsleddet giver en elasticitet på langt sigt på -1,050. Langsigtskoefficienten til ungdomsårgangen er estimeret til 1,252. Dette er lidt over det forventede. Man skulle forvente en værdi på maksimalt én for betydningen for ungdomsårgangen, idet dette svarer til en procentvis forøgelse i lærlingebeskæftigelsen på samme omfang som ændringen i ungdomsårgangen. Som nævnt er antallet af små og store ungdomsårgange i data imidlertid begrænset. Dette indebærer, at begivenheder, som er specifikke for tidsperioderne med små og store ungdomsårgange kan influere på resultaterne i det omfang disse begivenheder ikke fanges fuldt ud af andre variable i ligningerne. Ifølge standardfejlen til  $U$  kan hypotesen om en langsigtskoefficient på én ikke forkastes.

Endvidere er der medtaget ændringer i variablene for at forbedre tilpasningen på kort sigt. Endelig er der medtaget en dummy, som beskriver den ekstraordinært høje vækst i lærlingetilgangen fra 1975 til 1976, og som det ikke er muligt at indfange med nogle af variablene i data. Inkluderingen har ingen væsentlig indflydelse på størrelsen af de estimerede koefficienter, men forbedrer ligningens statistiske egenskaber betydeligt.

En anden fremgangsmåde er at estimere efterspørgsel efter lærlinge relativt til andre arbejdere med udgangspunkt i ligning (1.4). Også inkluderet i denne regression er  $U/P$ , ungdomsårgangen relativt til befolkningen i alderen 20-64. Variablene kan ikke afvises at være  $I(1)$  variable ifølge Dickey-Fuller tests. En regression af  $A/N$  på  $w_A^n/w_N$ ,  $U/P$  og en trend giver residualer, som ikke kan afvises at være  $I(1)$ . En regression af ændringerne i residualerne på det laggede niveau af residualerne giver en  $t$ -værdi på -3,55 og da McKinnons 5 procents kritiske værdi for afvisning af nulhypotesen om tilstedeværelsen af en enhedsrod er -3,78 i dette tilfælde, er variablene således ikke langt fra at kointegrere.

Det foretrukne resultat af en autoregressiv fordelt lag model er:

$$\ln \frac{A}{N} = -1,165 - 0,319 \ln \frac{w_A^n}{w_N} + 0,743 \ln \left( \frac{A}{N} \right)_{-1} + 0,005t + 0,281 \ln \left( \frac{U}{P} \right)_{-1} \quad (1.6)$$

(0,661) (0,118) (0,088) (0,002) (0,002)

$$+ 0,275 \Delta \ln \left( \frac{A}{N} \right)_{-1} + 0,327 D76$$

(0,107) (0,075)

$$R^2 = 0,908 \quad \hat{\sigma} = 0,0649 \quad T = 1950 - 2002$$

$$F_{\text{ar}}(2,44) = 2,579 \quad F_{\text{arch}}(1,44) = 0,694 \quad \chi_{\text{nd}}^2(2) = 0,561$$

$$F_{\text{he}}(11,34) = 0,530 \quad F_{\text{RESET}}(1,45) = 2,330$$

Alle koefficienterne til de forklarende variable er signifikant forskellige fra nul på et 5 procents niveau med anvendelse af de kritiske værdier fra normalfordelingen. Testene for misspecifikation angiver, at fravær af problemer ikke kan forkastes.

Regressionsligningen kan fortolkes som at lærlingetilgangen tilpasser sig en langtidsligevægt efter en »partial adjustment« mekanisme, hvor estimatet på tilpasningsparametren er én minus koefficienten til den laggede værdi af den afhængige variabel. Estimatet af lønelasticiteten er på langt sigt er således -1,241, mens effekten af ungdomsårgangen relativt til befolkningens størrelse er 1,093.

En hovedkonklusion fra de forskellige forsøg på estimationer er, at alle tre faktorer har en mærkbar betydning for lærlingebeskæftigelsen. Næsten uanset hvordan man vrider tallene, bliver resultatet, at alle tre faktorer har betydning. Men størrelsen varierer med fremgangsmåden. I Albæk (1992) blev der foretaget estimationer på opdaterede data, og der blev gennemført visse justeringer i specifikationerne. Der er visse forskelle mellem de resultater, der præsenteres i det følgende, og resultaterne i Albæk (1988) og (1992), men størrelsesordenen af de forskellige elasticiteter er nogenlunde den samme.

Konjunkturerne er målt ved udviklingen i nationalproduktet, og resultatet af estimationerne er en elasticitet på 1,4 på kort sigt. En forøgelse i nationalproduktet på én pct. forøger således lærlingetilgangen med lidt mere end én pct. inden for et år. På langt sigt er elasticiteten bundet til én.

For omkostningernes vedkommende giver den ene relation en langsigtselasticitet på ca. 1,1, mens den anden relation giver en langsigtselasticitet på 1,2. Der er altså ingen nævneværdig forskel mellem de to fremgangsmåder. Til sammenligning var resultaterne i de foretrukne relationer i Albæk (1992) henholdsvis 1 og 0,5. Man kan forsøge at sammenligne resultaterne med undersøgelser på tidsserier for andre lande med henblik på at vurdere, om resultaterne er plausible, men det har kun i begrænset omfang været muligt at finde tilsvarende undersøgelser for andre lande. I sin analyse af tilgangen af lærlinge i de britiske »engineering industries« opnår Stevens (1994) elasticiteter i omegnen af én, men har visse problemer med signifikans.

En interessant, men samtidigt også problematisk variabel, er ungdomsårgangene. Vi har for øjeblikket lave ungdomsårgange, og står i det kommende årti over for et stigende antal unge mennesker, som kommer ud på arbejdsmarkedet. Spørgsmålet er, om man kan forvente, at dette skaber problemer på lærepladsområdet. Langsigtselasticiteten i begge relation giver næsten det samme, den ene ligger på 1,3 og den anden på 1,1, hvilket vil sige, at en forøgelse af ungdomsårgangene med én pct. trækker lærlingetilgangen op med lidt mere end én pct.<sup>7</sup>

---

7. Hovedårsagen til estimater lidt større end én er sandsynligvis, at konjunkturudviklingen tilfældigvis har været gunstig i tidsrummene med store ungdomsårgange, jævnfør omtalen i afsnit 2. Estimerne er ikke signifikant forskellige fra én.

Dette resultat svarer på sin vis til det indtryk man får ved at sammenligne graferne over tilgang og ungdomsårgangenes størrelse i henholdsvis figur 1 og figur 2 i forrige afsnit. Der er den ubehagelighed, at det ikke er oplagt, hvad det er for en økonomisk mekanisme, som genererer sammenhængen mellem de to variable. Det kan nævnes, at medtagelse af udbudsvariable i aldersbetingede estimationer af beskæftigelsen er almindelige i amerikanske undersøgelser, men at udbudsvariablen (størrelse af aldersklasserne) sjældent kommenteres. Det er en nærliggende tanke at sammenligne ungdomsårgange med effekterne af ændringer i udbuddet af arbejdskraft. De fleste økonomer har den opfattelse, at øget arbejdsudbud primært giver sig udslag i øget beskæftigelse og kun i mindre udstrækning i øget arbejdsløshed, men denne sammenhæng bygger på en velbeskrevet økonomisk mekanisme, nemlig at øget arbejdsudbud og arbejdsløshed giver et lønpres i nedadgående retning, som resulterer i større beskæftigelse. Denne mekanisme må skønnes kun i begrænset omfang at gøre sig gældende på lærlingeområdet, idet minimumslønnen er fastlagt ved overenskomsterne og udstrakt til det ikke overenskomstdækkede arbejdsmarked ved lovgivning.

Endelig viser elasticiteten til trenden, at lærlingetilgangen er vokset med cirka  $\frac{1}{2}$  pct. om året, når der er taget højde for de øvrige forklarende faktorer påvirkning af tilgangen. Dette skal forstås som en beskrivelse af den meget gunstige udvikling på lærepladsområdet, som er fremgået i det forrige. Man kan ikke forvente, at udviklingen af sig selv fortsætter i samme opadgående takt fremover (alene af den grund, at der visse grænser for, hvor meget tilgangen kan vokse).

Vedrørende anvendelsen af regressionsresultaterne er det især ungdomsårgangene og omkostningerne, der er relevant. Hvor væsentlig lærepladsområdet end måtte synes, bestemmes den økonomiske udvikling og politik af mere overordnede forhold. For ungdomsårgangenes vedkommende kan det konkluderes, at der ikke er nogen grund til at forvente, at de stigende ungdomsårgange i det næste årti i sig selv vil skabe problemer på lærepladsområdet. Ud fra estimationsresultaterne er det mest plausible, at tilgangen til lærlingeuddannelserne vil vokse med ungdomsårgangene i en sådan takt, at der er en uændret andel af ungdomsårgang, som gennemfører en lærlingeuddannelse. Spørgsmålet er, om man er tilfreds med den nuværende andel af en ungdomsårgang, som tager en lærlingeuddannelse. En formindskelse af omkostningerne vil kunne afhjælpe mangelen på lærepladser. Der er usikkerhed om den nøjagtige størrelse af effekten, men ingen tvivl om retningen.

#### **4. Lærlingenes alder og produktivitet**

Et element i den samlede vurdering af lærlingeuddannelserne i det danske uddannelsessystem er lærlingenes alder og produktivitet, dvs. det bidrag, de yder til produk-

tionen under læreforholdet. Det følgende giver et overblik over udviklingen i lærlingenes alder, der på det seneste er vokset betragteligt.

Som det er fremgået af det forrige, har der over tiden været en meget gunstig udvikling på lærepladsområdet, idet antallet af lærepladser i forhold til ungdomsårgangene er gået op. Det har imidlertid fundet sted samtidigt med, at der har været en betydelig forøgelse i de relative omkostninger ved at beskæftige lærlinge. Som det fremgår af ligning (1.4), vil den relative forøgelse af lærlingelønningerne på højresiden trække i retning af en formindskelse af den relative lærlingetilgang på vestresiden. Dette kan modvirkes, hvis det sidste led i ligningen er positivt, og da estimeret af substitutionselasticiteten ( $\sigma$ ) er større end en, er den nødvendige betingelse, at forøgelsen i lærlingenes produktivitet ( $\lambda$ ) er større end forøgelsen af andre arbejders produktivitet ( $\mu$ ).

Det er der imidlertid også god grund til at formode at lærlingenes produktivitet er gået op. I begyndelsen af 1950erne var lærlingene ganske unge, og de fleste må skønnes alene at have 7 års obligatorisk skolegang bag sig. Lærlinge har nu alle gået i 9 år i grundskole, og en god del af dem har også deltaget i yderligere almen skolegang. Hertil kommer den forbedrede kvalitet i skoleundervisningen i forbindelse med læreforholdet. Op til midten og slutningen af 1950erne gik lærlingene på skole om aftenen efter at have arbejdet hos mester om dagen, og det satte visse grænser for intensiteten af indlæringen. Hertil kommer, at hovedparten af undervisningen var ren kateterundervisning, idet udstyret og apparaturet på de tekniske skoler var begrænset, i det omfang der overhovedet var noget. Dette står i kontrast til nutidens tekniske skoler, der er velforsynede med udstyr, apparatur og materialer til dagundervisning af lærlinge.

Et overblik over udviklingen i lærlingenes alder fås af tabel 1. Det ses således, at den typiske alder for indgåelse af en lærekontrakt i 1956 var 16 år, som én tredjedel af alle lærlingene havde, mens lidt mere end én tredjedel var yngre, og lidt mindre end én tredjedel var ældre. Sammenlignet med 1939 er der tale om en mindre vækst i lærlingenes alder. I 1974 var den typiske alder ved indgåelse af lærekontrakten vokset til 18 år, som 40 pct. af lærlingene havde, mens 38 pct. var yngre, og 22 pct. ældre.<sup>8</sup> I 1984 var den typiske alder stadig 18 år, hvor 35 pct. af lærlingene indgik læreforhold, men da var kun 21 pct. yngre, mens 44 pct. var ældre.

Udviklingen i aldersfordelingen op til 1984 er stadig forenelig med, at lærlingene har gået noget længere i almen skole, og at det kan tage noget tid at finde frem til den rigtige læreplads inden for det rigtige fag. Det kan derimod ikke siges om aldersfordelingerne for 1997 og 2002. Aldersfordelingerne er så flade, at man næppe kan tale om en typisk alder for indgåelse af lærekontrakt. Det er især bemærkelsesværdigt, at der

---

8. Der var ingen løbende aldersstatistik før indførelsen af den CPR-nummer baserede statistik i begyndelsen af 1970erne, udover 1939 og 1956 eksisterer der alene en opgørelse for 1957, så det er ikke muligt at spore det tidsmæssige forløb nøjere.



Tabel 1. Lærlingene fordelt på alder ved påbegyndelse af læreforholdet i udvalgte år: Procent.

Alder, år:	1939	1956	1974	1984	1997	2002
14	11	8	0	0	0	0
15	31	27	0	0	0	0
16	32	33	7	3	4	5
17	21	25	31	18	18	16
18	5	6	40	35	17	14
19	0	1	13	20	11	11
20	0	0	4	10	11	9
21-22	0	0	3	9	15	14
23-24	0	0	1	3	7	7
25-30	0	0	0	2	8	12
31-	0	0	0	1	8	13
I alt	100	100	100	100	100	100

Kilde: Danmarks Statistikbank og Statistiske Efterretninger, diverse årgange.

tilsyneladende har været en ganske kraftig vækst i andelen, som har en meget høj alder, således skulle andelen, som ved indgåelse af lærekontrakt havde en alder på 25 år og derover, være helt oppe på 25 pct. i 2002 ifølge opgørelsen i tabel 1, en vækst fra 16 pct. i 1997, mens andelen i 1984 var nede på 3 pct.

Væksten i alderen kan delvist tilskrives forskydninger i sammensætningen i tilgangen, men synes især at stamme fra en vækst i alderen inden for de enkelte grupper. Det er muligt at foretage en opdeling af lærlinge i forskellige grupper, se Albæk (2004) appendiks B.

En beregning på grundlag af aldersopgørelsen i tabel 1 giver det resultat, at den gennemsnitlige alder ved tilgang til anden del er vokset med ét år fra 1997 til 2002, fra 20,9 år til 21,9 år. Der er imidlertid tale om et skøn, idet aldersopgørelserne i kilden (Danmarks Statistikbank) er opdelt på intervaller, og beregningen er således ikke nøjagtig. Et andet mål for alderen ved tilgangen er medianen, som er mere præcis i denne sammenhæng, og her er der en vækst på 0,6 år, fra 19,0 i 1997 til 19,6 i 2002. Væksten i lærlingenes alder fremgår også af andre opgørelser, således hedder det i AER's årsberetning for 2001, side 25: »...., at gennemsnitsalderen for elever, der indgår en uddannelsesaftale med en arbejdsgiver, på blot 5 år er steget med ca. 2 år, fra 20,5 til 22,4 år.«

AER har mulighed for at foretage en mere præcis opgørelsen end brugerne af Statistikbanken, så hvis man tager udgangspunkt i de 22,4 år, og tillægger et skøn over den gennemsnitlige restlæretid på f.eks. 2 år (hvilket ikke burde være for højt sat), kommer man op på en gennemsnitlig alder ved færdiggørelse af et læreforhold på 24,4 år. Det-

te er et betænkeligt højt tal, der er tæt på alderen for en student, som gennemfører en lang videregående uddannelse på normeret tid, og som ikke både har taget tiende klasse og et sabbatår efter studentereksamen.

I den forbindelse er det værd at fremhæve, at det er den gennemsnitlige alder ved færdiggørelse af læreforholdet, som er afgørende for de antal år, lærlingene kommer til at arbejde med den høje produktivitet som faglærte arbejdere, der er sigtet med investeringen i uddannelsen. Når op mod halvdelen af en ungdomsårgang gennemløber en lærlinguddannelse, udløser en vækst i gennemsnitsalderen et ikke ubetydeligt produktionsstab for samfundshusholdningen.

Det må derfor være hensigtsmæssigt at overveje foranstaltninger med det sigte at bremse væksten i den gennemsnitlige lærlingealder, og måske prøve at føre alderen tilbage til et tidligere kendt niveau. For eksempel kan det noteres, at den særligt gunstige tilskudsordning til voksnelever giver unge under 25 år en tilskyndelse til at vente med at indgå et læreforhold, indtil der er mulighed for at opnå den høje lønsats. I hvilket omfang støtteordningen til voksnelever rent faktisk har bidraget til at udskyde de unges start på en lærlinguddannelse, kan der ikke siges noget om på det foreliggende datagrundlag.

### **5. Løsningsmuligheder på lærepladsproblemet**

Før der ses på løsningsmulighederne, er det måske hensigtsmæssigt at se lidt nærmere på behovet for at iværksætte løsninger på et utilstrækkeligt antal lærepladser. Der er flere forskellige indfaldsvinkler til denne problemstilling.

Gennem adskillige år er presset for at finde løsninger kommet fra de mange unge, som har gennemført første del af en lærlinguddannelse på skolerne, men som ikke efterfølgende har kunnet opnå plads hos en arbejdsgiver. På det seneste er det omkostningsudviklingen i forbindelse med den praktikpladskompenserende undervisning, som har gjort det påkrævet at overveje alternative løsningsmuligheder.

Imidlertid er det værd at fremhæve forbindelsen til en bredere debat om udviklingen på det danske arbejdsmarked i de senere år. Et tema i Det økonomiske Råds rapporter har været, at manglende uddannelse indebar en ganske stor risiko for arbejdsløshed, og at en ekstra uddannelsesindsats også burde ses i dette perspektiv. I Albæk m.fl. (1992) blev det fremhævet, at netop lærlinguddannelserne er et vigtigt element i en sådan indsats, idet disse uddannelser må anses for at være attraktive for de grupper, som for indeværende ikke får en erhvervsuddannelse. Derimod spiller kapaciteten på visse attraktive videregående uddannelser næppe den store rolle i denne forbindelse. En forøget optagelseskapacitet på lærlinguddannelserne må således ses som et afgørende element i at formindske den andel af en ungdomsårgang, som ikke får en er-

hvervsuddannelse. Set i et indkomstfordelingsperspektiv giver en øget andel af unge med erhvervsuddannelse mulighed for en mere jævn indkomstfordeling i fremtiden.

En række muligheder har indgået i debatten om lærepladser, og disse gennemgås i det følgende. Der er ikke tale om nogen komplet diskussion, idet nogle af mulighederne ikke har været tillagt større vægt på det seneste. For en mere indgående diskussion af de forskellige alternativer henvises til Albæk (1988).

En mulighed, som har været anvendt flere gange, er *henstillinger og kampagner*. I foråret 1990 blev det vedtaget, at der skulle oprettes skolepraktik finansieret af arbejdsgiverne, hvis de ikke inden efteråret havde oprettet det tilstrækkelige antal praktikpladser. Vedtagelsen blev efterfulgt af omfattende kampagner og henstillinger fra arbejdsgiverorganisationerne til deres medlemmer om at tage flere lærlinge og praktikanter. Aviserne var fyldt med annoncer med billeder af håbefulde unge mennesker i forgrunden og myndige mestre med bagerhuer og slagterhatte i baggrunden. Der var tale om en kampagne af størst tænkelig skala med opbakning fra alle autoriteter på området.

Når man efterfølgende gør udviklingen op, kan man konstatere, at kampagnen synes at have haft en begrænset effekt. Der var en vækst i tilgangstallet fra 1989 til 1990 af størrelsesordenen 400 lærlinge, og denne forøgelse havde som bekendt ikke en tilstrækkelig størrelsesorden til at undgå indførelsen af skolepraktikken. Dette tyder på, at kampagner og henstillinger nok kan have en midlertidig effekt, men at man ikke kan bero på dette middel i en ordning, som skal være af mere varig karakter.

En mere formel måde at vurdere effekten af kampagner er at prøve at inkludere dummier i regressionerne for perioder med kampagner. Sådanne eksperimenter giver det resultat, at man ikke kan afvise hypotesen, at kampagner ikke har nogen effekt på tilgangen af lærlinge. I sampleperioden er supplerende dummier på grænsen til signifikans i to tilfælde med koefficienter omkring to gange regressionsligningernes standardfejl (og hvor dummierne næppe kan siges at indfange effekterne af kampagner). I det omfang kampagner har en effekt på tilgangen af lærlinge, synes størrelsesordenen at være moderat.

Med »Lov om tilvejebringelse af lære- og praktikpladser« blev der åbnet mulighed for, at skoleuddannelse kan erstatte praktikophold, også kaldet *skolepraktik*. Det er ikke noget, man var særlig glad for, og i lovbemærkningerne blev det nævnt som en nødløsning. Det er også bestemt, at eleverne har pligt til at forlade skolepraktikken, hvis de får anvist en praktikplads.

Der er både problemer af pædagogisk og økonomisk karakter ved skolepraktik. De økonomiske består især i omkostningsproblemet. Omkostningerne ved skolepraktik består dels af udgifter til lokaler, maskiner, materialer, lærerlønninger m.m., og det er forholdsvis dyrt sammenlignet med andre uddannelser. At lære en murerlærling at

lægge mursten er ikke auditorieundervisning, men kræver en mere intensiv undervisning. Hertil kommer udgifterne til elevernes indkomst, enten i form af SU eller en anden aflønning. Opgørelser over omkostningerne til skolepraktik viser da også, at de overstiger omkostningerne ved mange universitetsstudier.

Under et læreforhold afholder virksomhederne udgifterne, også til lønnen, og kun under skoleophold i læreperioden påføres det offentlige udgifter til drift og lønrefusion til arbejdsgiverne. Udgifterne til skolepraktik er derfor langt større end udgifterne under et læreforhold hos en arbejdsgiver, og det synes derfor en god forretning at lade arbejdsgiverne stå for uddannelsen mod at betale en del af udgifterne. I forvejen dækker det offentlige størstedelen af udgifterne til forskellige former for uddannelse i det danske samfund, så hvorfor ikke dække en del af udgifterne til den uddannelsesaktivitet, der foregår i private virksomheder.

*Kvoteordning* går ud på, at virksomhederne pålægges at ansætte et antal lærlinge, afhængigt af den enkelte virksomheds størrelse. For øjeblikket er kontrakten frivillig: arbejdsgiveren vil gerne beskæftige eleven, og eleven vil gerne ansættes. Ved kvoteordningen tvinges arbejdsgiveren til at beskæftige et ungt menneske, også i det tilfælde, hvor han eller hun helst er fri for det. Det er oplagt, at dette kan betyde, at ordningen ikke kommer til at fungere på en rimelig måde.

Hertil kommer, at kvoteordningen ikke er gratis, selvom udgifterne ikke står opført i de offentlige budgetter. Arbejdsgiverne påføres udgifter, og det synes mere oplagt at inddrage disse udgifter via en beskatning af de mere vrangvillige og udbetale beløbet til de arbejdsgivere, som gerne vil beskæftige lærlinge.

Endvidere risikerer kvoteordninger let at få karakter af et mekanisk instrument. Arbejdsgiverne har værsågod at ansætte f.eks. én lærling for hver tiende ansat, uanset om virksomheden er egnet til at uddanne lærlinge og uanset om der også efterfølgende er behov for den type af lærlinge, som virksomheden kan uddanne.

Fra tid til anden er kvoteordninger blevet nævnt i debatten som en mulighed for at få langtidsledige beskæftiget i private virksomheder. Det kan fremføres, at argumenterne for kvotering på lærlingeområdet synes svagere, idet der her er en ganske betydelig fleksibilitet i beskæftigelsen, som gennemgået i det forrige. Tilgangen og beskæftigelsen har varieret betydeligt over tid, og har været betydeligt højere, end den er for øjeblikket – spørgsmålet er, hvordan man påvirker virksomhederne til at komme op på et tidligere kendt niveau.

Den sidste mulighed er *løntilskud og løntilbageholdenhed*. Set fra den enkelte arbejdsgivers synsvinkel virker løntilskud og løntilbageholdenhed ens: omkostningerne ved beskæftigelse af en lærling eller elev falder. Spørgsmålet er så, om lærlingetilgangen stiger, når lønomkostningerne falder. I denne artikel er der præsenteret argumenter for, at der foreligger en sådan sammenhæng, og synspunktet burde egentlig være ukontroversielt.

Ikke mindst fra borgerlig side har man peget på de høje ungdomslønninger som en væsentlig forklaring på ungdomsarbejdsløsheden og foreslået introduktionsløn indført som middel til at hjælpe de unge over lønhoppet ved 18-årsgrænsen. Men hvis lavere omkostninger betyder, at arbejdsgiverne tager flere unge ufaglærte ind, så må de jo også tage flere lærlinge ind, hvis lærlingene bliver billigere at beskæftige – problemstillingen er helt ens.

Løntilskuddene blev indført i slutningen af 1970erne, hvor socialdemokraterne var ved regeringsmagten. Siden er der alene i den permanente tilskudsordning udbetalt næsten 20 mia. kr. i løbende priser i tilskud til beskæftigelse af lærlinge, og hertil kommer beløbene til de ekstraordinære tilskud, der har været iværksat i perioder med behov for en særlig indsats. Men hvis løntilskud virker, så virker løntilbageholdenhed også, og overenskomstafslutninger med høje lønstigninger til lærlinge må derfor have formindsket uddannelsesmulighederne.

Der burde således være bred opbakning bag det synspunkt, at omkostningsformindskelse kan afhjælpe lærepladmangelen, men de ubehageligheder, der er forbundet med midlerne til gennemførelse af en omkostningsreduktion, gør en tøven forståelig. Løntilskud til en så stor gruppe på arbejdsmarkedet er en dyr foretælse, og reduktion af lønningerne for personer, der sammenlignet med andre på arbejdspladsen har forholdsvis lave indkomster, kan næppe forventes at være populær.

Hovedproblemet ved at formindske omkostningerne alene ved at yde løntilskud er, at tilskuddet løber op i ganske store summer. Således svarede udgifterne til løntilskud til lærlinge i midten af 1980erne rundt regnet til halvdelen af udgifterne ved at drive alle landets universiteter.

Hertil kommer, at det ikke er sikkert, at lærlingelønnen holder sig i ro efter indførelsen af et sådant tilskud. Som det fremgik af afsnit 2, er lønningerne til lærlinge vokset betragteligt siden man begyndte at udbetale tilskud i slutningen af 1970erne. Man kan på ingen måde udelukke, at indførelse af tilskud til lærlinge påvirker løndannelsen for denne gruppe på arbejdsmarkedet. Faktisk er det netop en hovedtankegang i nyere økonomisk litteratur, at man i en økonomi med kollektiv løndannelse bør forsøge at tage højde for, hvordan løndannelsen påvirkes af forskellige indgreb.

Et resultat er, at hvis fagforeningerne har en målsætning om en lærlingebeskæftigelse af en vis størrelse, og fagforeningerne samtidig har kontrol over lønningerne, så vil tilskud ikke have nogen betydning for beskæftigelsen. Tankegangen er følgende: Fagforeningen vælger en lønsats, så der set fra dennes synspunkt er en optimal kombination mellem beskæftigelse af voksne og lærlinge. Statsmagten har imidlertid den opfattelse, at lærlingebeskæftigelsen er for lille, og giver derfor et subsidie til beskæftigelse af lærlinge. Resultatet bliver, at virksomhederne ansætter relativt flere lærlinge og relativt færre af fagforeningens medlemmer. Ud fra fagforeningens synspunkt er

dette inoptimalt, og den genopretter ligevægten ved at lade lærlingenes løn stige svarende til beskæftigelsesubsidiet. Det er værd at fremhæve, at hele tankegangen forudsætter, at fagforeningen er glad for lærlinge, ellers kunne den jo blot lade lønnen stige så meget, at de slet ikke blev ansat. I Albæk (1988) er der opstillet en formel model, som giver de beskrevne resultater.

Hvor meget der er om en sådan tankegang, er svært at sige, og der kan diskuteres for og imod. Imidlertid må det anses for betænkeligt at indføre en ordning, hvor lønhøjelser aftalt ved overenskomstforhandlinger finansieres af statskassen. Dette gælder, uanset hvordan finansieringen af tilskuddene udformes.

Hvis omkostningerne ved beskæftigelse af lærlinge skal sænkes, er formindskelse i lønningerne det eneste alternativ til tilskud. Et argument imod dette er, at lærlingelønnen altså stadig ligger en del under lønnen til andre arbejdere. Spørgsmålet er imidlertid, hvem man skal sammenligne med. Hvis man sammenligner med andre uddannelsessøgende, er lærlinge og elever forholdsvis godt stillede. Aflønningen er noget højere end SU. Når lærepladsen først er fundet, og prøvetiden overstået, er der rimelig stor sikkerhed for gennemførelse. Uddannelsestiden med lav indkomst er forholdsvis kort: gennemsnittet for lærekontrakten ligger på  $2\frac{1}{2}$ -3 år.

Et andet argument mod formindskelse af lærlingelønningerne er, at dette vil formindske incitamentet til at tage en uddannelse. Dette er et tema, som har haft en vis placering i den arbejdsmarkedspolitiske debat. Der er imidlertid flere argumenter imod den tanke, at omfanget af lærlinguddannelserne skulle være begrænset som følge af manglende incitamentet til at tage en sådan uddannelse. For det første er der en ganske stor kø af unge, som ikke kan få en plads. For det andet ville forhøjelser af lærlingelønningerne udløse en større tilgang, hvis det er udbuddet af unge, der er begrænsningen, og ikke virksomhedernes efterspørgsel, og dette synes ikke at være tilfældet. For det tredje er det tilsyneladende ikke de unges egen opfattelse. I en undersøgelse foretaget af Amtskommunernes og Kommunernes Forskningsinstitut (1990, s. 45), blev de unge spurgt, om de ville tage en plads, hvis lønnen var 15 pct. lavere. To tredjedele svarede ja, både blandt dem, som havde fået, og blandt dem som stadig søgte.

## 6. Konklusion

Det overordnede billede af udviklingen på lærlingeområdet op mod årtusindeskiftet er meget positivt. Over en årrække har andelen af en ungdomsårgang, som begynder på en lærlinguddannelse, holdt sig på et niveau på omkring 45 pct. To gange tidligere har der været en tilgang til lærlinguddannelserne på det niveau, men det var kun for et par enkelte år, og ikke for en længere periode.

Et problem, som har udviklet sig betydeligt på det seneste, er den høje alder for lærlingene: gennemsnitsalderen ved færdiggørelse af en uddannelse som faglært arbejder er tilsyneladende tæt på de 25 år. Det svarer til den alder, som en god del af kan-

didaterne fra universitet burde have, hvis der var et rimeligt effektivt gennemløb på dette område. En akademiker har imidlertid et normeret uddannelsesforløb på 8 år efter folkeskolen, og det er jo ikke tilfældet for en lærling.

Når op mod halvdelen af en ungdomsårgang gennemløber en lærlingeuddannelse, indebærer den sene alder som færdiguddannet et betydeligt produktionstab for samfundet – det antal år, hvor investeringen i lærlingeuddannelsen giver afkast, bliver mindre. Det må derfor være væsentligt at overveje initiativer, som kan bremse væksten i lærlingenes alder, og helst føre den tilbage til et tidligere kendt niveau. Dette gælder også af hensyn til at forbedre mulighederne for at skaffe flere lærepladser: det er lettere at holde indkomstniveauet lavt for en ung under 20 år i forhold til en 25-årig, og dermed spare på både lønninger og offentlige tilskud.

Fra 2001 til 2002 indtraf en betragtelig tilbagegang i tilgangen af lærlinge på 5 procentpoint, som reducerede andelen af en ungdomsårgang, der begynder en lærlingeuddannelse, fra 46 til 41 pct. De foreliggende opgørelser angiver, at der er indtruffet en yderligere reduktion siden 2002.

Det noteres, at de indgreb, som regeringen har gennemført i samarbejde med forligspartierne på området, Det Radikale Venstre og Dansk Folkeparti, næppe kan anses for tilstrækkelige. Hovedelementerne er en forbedring af incitamentsstrukturen, således at elever i skolepraktik ikke mere får den samme løn som elever, der har opnået en ordinær læreplads. Endvidere er det aftalt at begrænse omfanget af pladser på skolepraktik, men dette indgreb er jo ikke noget, som i sig selv skaber flere uddannelsespladser til de unge.

Det rejser spørgsmålet, om den aktuelle situation berettiger til et indgreb på området. Besvarelsen af spørgsmålet afhænger bl.a. af, hvor alvorligt man opfatter tilbagegangen i tilgangen af lærlinge. På den ene side kan man pege på, at de 41 pct. faktisk stadig er pænt højt sammenlignet med tidligere. På den anden side ville det jo ikke gøre noget, at man forsøgte at holde sig oppe på de cirka 45 pct., når det nu er lykkedes at komme helt op på dette niveau. Det drejer sig faktisk om ganske mange unge mennesker, som ikke kan få den erhvervsuddannelse hos en arbejdsgiver, som de ønsker sig. De relevante ungdomsårgange ligger for øjeblikket på næsten 60.000, og de 5 procentpoint svarer derfor til ca. 3.000 unge, som ikke kan begynde på deres ønskede uddannelse. Og det er vel at mærke 3.000 per år i perioden, dykket på de 5 procentpoint varer.

Det forhold, at unge har svært ved at få en uddannelse som faglært arbejder, når de er så uheldige at skulle søge læreplads i en lavkonjunktur, er en indbygget uhelds- og uheldsmæssighed i den type af samfundssystem, som vi befinder os i.<sup>9</sup> I det omfang man me-

---

9. Dette forhold er da også blevet noteret med omhu af tilhængerne af et enhedsskolesystem, hvor erhvervsuddannelserne og gymnasieuddannelserne slås sammen i et undervisningssystem baseret på ren skoleundervisning, se især diskussionen hos Erik Jørgen Hansen (2003).

ner, at dette er et problem, er det værd at overveje, hvad man kan gøre ved det. I en lavkonjunktur skulle man synes, at det netop er en god idé at prøve at kompensere for de uddannelsespladser, som automatisk falder bort hos private arbejdsgivere, med mindre der foretages indgreb.

Overvejelserne om et eventuelt indgreb og doseringen af dette må selvsagt også afhænge af forventningerne til konjunkturudviklingen. Hvis den aktuelle afmatning i økonomien afhjælpes, vil dette have en gunstig effekt på lærlingetilgangen. I de senere år har økonomer udviklet en betydelig skepsis med hensyn til det hensigtsmæssige i at forsøge at regulere konjunkturerne ved hjælp af generelle finanspolitiske indgreb, men argumenterne for at være tilbageholdende med hensyn til indgreb kan næppe overføres til lærepladsområdet. Bortfaldet af lærepladser i lavkonjunkturer er primært efterspørgselsbestemt, og det er f.eks. vanskeligt at se, hvordan indgreb til afhjælpning af dette skulle blokere for tilpasninger på udbudssiden, som kunne bidrage til forbedring af lærepladssituationen.

Ved gennemgangen af de forskellige løsningsmuligheder synes konklusionen at måtte være, at den mest oplagte metode til forbedring af lærepladssituationen er en formindskelse af arbejdsgivernes omkostninger ved beskæftigelse af lærlinge. At omkostningerne spiller en rolle for størrelsen af tilgangen af lærlinge synes at fremgå alene ved en sammenligning mellem de forskellige grafer i afsnit 2, men dette bekræftes også af forsøgene på at spore de nøjere sammenhænge mellem de størrelser, som påvirker lærlingetilgangen og lærlingebeskæftigelsen.

Omkostningsreduktion er en metode til at fremkalde flere lærlingepladser, som nu har en permanent status, og det må indebære, at beslutningstagerne på området har den opfattelse, at metoden virker. Det er endvidere en metode, der tidligere har været anvendt i forbindelse med særlige problemer for lærlingetilgangen. I den forbindelse noteres også, at det midlertidige tilskud blev gradvist reduceret ned til nul i takt med forbedringen på arbejdsmarkedet i slutningen af 1990'erne.

Ved de sidste overenskomstforhandlinger blev der desværre aftalt lønstigninger for lærlinge på 4,5 pct. om året over en 3-årig periode. Det må skønnes at ligge noget over den forventede udvikling for andre arbejdere. Det er ikke enkelt at skønne over omkostningsudviklingen for andre arbejdere, men hvis den f.eks. holder sig på 3 pct. om året, så vil lærlingelønningerne jo efter 3 år være steget med hele 4,5 pct. i forhold til andre arbejdere. Konsekvensen er, at offentlige indgreb i form af løntilskud bliver endnu dyrere, end de ellers ville have været. Det er ikke klart for denne forfatter, hvad der er bevæggrunden hos arbejdsmarkedets parter for at forøge omkostningerne ved at beskæftige lærlinge i den nuværende situation med udbredt mangel på lærepladser. Som det er fremgået af gennemgangen af lønudviklingen, er det satsforhøjelser, som spiller den væsentligste rolle, og der ligger således et betydeligt ansvar for udviklingen på området hos arbejdsmarkedets parter.



## Litteratur

- Albæk, K. 1988. *En økonomisk analyse af lære- og praktikpladsproblemet*. Licentiat-afhandling. Økonomisk Institut, Københavns Universitet.
- Albæk, K. 1992. Lære- og praktikpladsspørgsmålet. Redegørelse til Undervisningsministeriets udvalg vedrørende alternativer til praktikpladskompenserende undervisning. I *Rapport fra udvalget vedrørende alternativer til praktikpladskompenserende undervisning*, Bind II: Bilag, s. 1-52.
- Albæk, K. 2004. Om lærepladsspørgsmålet. Blåt Memo nr. 212. Økonomisk Institut, Københavns Universitet, september 2004. Kan hentes fra <http://www.econ.ku.dk/wpa/blue/bluedefault.htm>.
- Albæk, K., E. S. Madsen, K. Pedersen. 1992. *Veje til fuld beskæftigelse*. I K. Albæk, m.fl. Kampen mod ledigheden. Spektrum. Rockwool Fondens priskonkurrence om en analyse af ledighedsproblemet i Danmark.
- Amtskommunernes og Kommunernes Forskningsinstitut. 1990. *Manglen på lære- og praktikpladser*. København.
- Arbejds- og Socialministeriet. 1945. *Lærlingetilgangen 1935-1944 og Aarsagerne til dens bevægelse*. Arbejds- og Socialministeriets økonomisk-statistiske undersøgelser, nr. 9. Særtryk af Socialt tidsskrift.
- Arbejdsgivernes elevrefusion (AER). *Årsberetning*. Diverse årgange.
- Bjørn, N. H. 1996. The Apprenticeship System – Supply and Demand. Chapter 7 in Eskil Wadensjö, red.: *The Nordic Labour Markets in the 1990's*, Part II, Amsterdam: Elsevier.
- Danmarks statistik. *Statistikbanken*.
- Danmarks statistik. *Statistiske efterretninger*. Diverse årgange.
- Hansen, E. J. 2003. *Uddannelsessystemerne i sociologisk perspektiv*. København.
- Houman Sørensen, J. m.fl. 1983. Lærlinge, uddannelse og udbytning: om lærlingeuddannelsenes økonomiske, politiske og ideologiske funktioner. Aalborg Universitetscenter.
- Stevens, M. 1994. An investment model for the supply of training by employers. *Economic Journal*, 104, s. 556-70.
- VK Regeringen II. 2005. *Nye mål. Regeringsgrundlag*. København, Februar 2005.
- Westergaard-Nielsen, N. og A. Rue Rasmussen. 1999. The Impact of Subsidies on Apprenticeship Training. *Research in Labour Economics*, Vol. 1B, s. 359-75.

# Miljømodeller til ADAM

Frits Møller Andersen, Kenneth B. Karlsson, Dorte Grinderslev  
Forskningscenter Risø, Systemanalyseafdelingen

Morten Werner

Danmarks Statistik, Økonomiske modeller

Trine S. Jensen

Danmarks Miljøundersøgelser, Systemanalyseafdelingen

*SUMMARY: To support environmental analyses and policy at a national level, forecasts of environmental indicators consistent with economic indicators are required. However, environmental indicators are physical measures linked to physical activities not specified in economic models. One way to handle this is to develop environmental satellite models linked to economic models. This paper presents a set of satellite models determining emissions of greenhouse gasses and acid gasses linked to the Danish macro-econometric model ADAM.*

---

## 1. Indledning

For at udviklingen i miljøforhold kan indgå i den politiske prioritering på linie med udviklingen i økonomiske indikatorer som produktion, beskæftigelse og indkomst, er det nødvendigt at fremskrive udviklingen i en række miljøindikatorer. Men miljøindikatorer er fysiske størrelser ofte beregnet for ganske specifikke fysiske aktiviteter, der ikke fremskrives i økonomiske modeller.

I denne artikel gennemgås et system af miljømodeller til ADAM, der søger at kombinere overordnede økonomiske indikatorer og detaljerede miljørelaterede fysiske aktiviteter for at fremskrive miljøindikatorer konsistent med økonomiske fremskrivninger. Det er valgt at udvikle modelsystemet til ADAM, da ADAM er en offentligt tilgængelig model, der anvendes til officielle fremskrivninger af den økonomiske udvik-

---

De beskrevne miljømodeller til ADAM er udviklet i et samarbejde i AMOR-centret under det Strategiske Miljøforskningsprogram mellem Forskningscenter Risø, Danmarks Miljøundersøgelser, Fødevarøkonomisk Institut og Danmarks Statistik. Miljømodellerne er dokumenteret i Andersen m.fl. (2001 og 2003) og Andersen (2001). Modellernes komplette sæt af formler m.v. er offentliggjort sammen med dokumentationen af ADAM på Danmarks Statistiks hjemmeside [www.dst.dk](http://www.dst.dk), og modellerne stilles til rådighed for ADAM's abonnenter. Vi skylder en tak til to anonyme referees for mange gode kommentarer til en tidligere version af denne artikel. Bemærk, at synspunkterne i artiklen er forfatterne og ikke nødvendigvis deles af de institutioner, forfatterne er tilknyttet.

ling i Danmark. Miljømodellerne fremskriver aggregerede nationale indikatorer for drivhuseffekt, forsurening og næringsstofbelastningen af vandmiljøet, således at disse miljøforhold kan indgå i en overordnet prioritering f.eks. i forbindelse med en miljøvurdering af finansloven. Specielt for næringsstofbelastningen af vandmiljøet er belastningens geografiske fordeling miljømæssigt meget væsentlig. Nationale udledninger af næringsstoffer er således kun en meget overordnet indikator på problemer relateret til vandmiljøet.

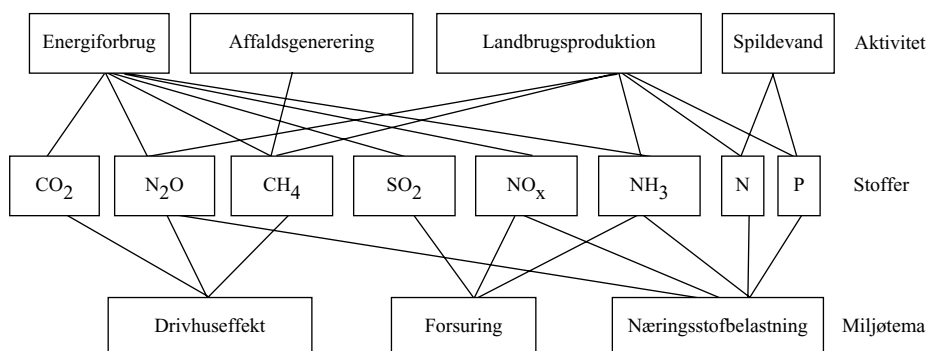
Ved at se på flere miljøindikatorer samtidigt tages højde for, at forskellige miljøindikatorer er koblet forskelligt til den økonomiske udvikling, og at det for en overordnet vurdering derfor sjældent er tilstrækkeligt blot at se på én indikator, f.eks. emissionen af CO<sub>2</sub>. Ses eksempelvis på udviklingen i emissionen af drivhusgasser, vil emissionen af de forskellige gasser udvikle sig forskelligt, og i nogle situationer vil emissionen af enkelte stoffer stige, mens emissionen af andre stoffer vil falde.

Der gives en oversigt over modelsystemet i afsnit 2, og de betragtede miljøtemaer opridses i afsnit 3. Skønt de udviklede miljømodeller til ADAM inkluderer næringsstofbelastningen af vandmiljøet, fokuseres i denne artikel på drivhusgasser og forsurende gasser. I afsnit 4 foretages en række simuleringseksperimenter med modelsystemet for at illustrere anvendelsesmulighederne, og endelig konkluderes i afsnit 5.

## 2. Emissioner og modeller

Inden for fremskrivning af miljøindikatorer eksisterer to meget forskellige traditioner. I den ene tradition kobles emissionen af enkelte stoffer til den økonomiske udvikling på et aggregeret niveau, f.eks. kobles emissionen af energirelaterede emissioner af CO<sub>2</sub>, SO<sub>2</sub> og NO<sub>x</sub> til integrerede energi-økonomi modeller. Eksempler inden for denne tradition er E3ME – An Energy Environmental Economy Model for Europe, Barker m.fl. (1998), den norske MSG-EE model, Alfsen m.fl. (1996) og GEM-E3 modellen, Capros m.fl. (1997). Den anden tradition tager udgangspunkt i, at miljøindikatorer er fysiske størrelser, der ofte beregnes for ganske specifikke fysiske aktiviteter f.eks. kg NH<sub>4</sub> (metan) pr. malkeko pr. år. Inden for denne tradition kobles udviklingen i emissionerne til tekniske forudsætninger som eksempelvis udviklingen i malkekøernes årlige produktion og aggregerede økonomiske indikatorer. I forbindelse med beregningen af emissioner fra forskellige kilder sikres ikke nødvendigvis makroøkonomisk konsistens. Eksempler inden for denne tradition er AEA Technology (1998), ECOFYS (1998), IPCC (1996a, b) og de danske indberetninger til FN, (Miljøstyrelsen 2003).

De udviklede miljømodeller, beskrevet i denne artikel, søger at kombinere de to traditioner. Der tages udgangspunkt i traditionen for integreret modellering, og denne udvides med flere typer af emissioner samt med etablering af links til fysiske størrelser og detaljerede emissionstekniske beskrivelser.



Figur 2.1 Miljøtemaer, emissioner og aktiviteter.

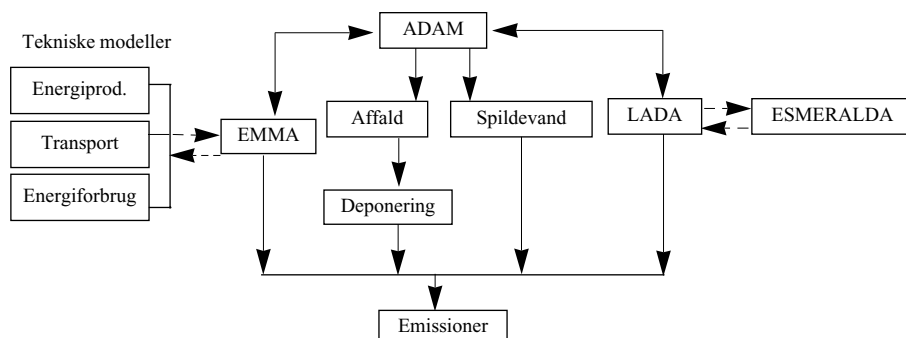
Anm.: CO<sub>2</sub>: kuldioxid, N<sub>2</sub>O: lattergas, CH<sub>4</sub>:metan, SO<sub>2</sub>: svovldioxid, NO<sub>x</sub>: kvælstofoxider, NH<sub>3</sub>: ammoniak, N: kvælstof, P: fosfor.

De overordnede sammenhænge mellem miljøtemaerne drivhuseffekt, forsuring og næringsstofbelastning, emission af stoffer og aktiviteter er vist i figur 2.1. Som det fremgår af figuren, er hvert miljøtema relateret til emissionen af flere forskellige stoffer, og de enkelte aktiviteter bidrager til emissionen af mange forskellige stoffer. Opstillingen af miljømodeller til ADAM starter med en detaljering og transformering af de økonomiske aktiviteter til specifikke fysiske aktiviteter, hvorefter emissionen af de enkelte stoffer kobles til de detaljerede fysiske aktiviteter. Som det ses af figur 2.1, er det primært energiforbrug, affaldshåndtering, spildevand og landbruget, det er nødvendigt at modellere mere detaljeret, end det er gjort i ADAM.

Modelmæssigt er det valgt at udvikle miljømodellerne som en række satellitmodeller til ADAM, dvs. som for- og eftermodeller. Herved opnås, at ADAM kan simuleres uafhængigt af miljømodellerne, således at disse kun simuleres i forbindelse med specifikke miljøvurderinger. EMMA køres dog iterativt med ADAM således, at energiforbrug og produktion i forsyningssektoren bestemt i EMMA føres tilbage til ADAM. Miljømodellerne til ADAM er vist skematisk i figur 2.2.

ADAM (Annual Danish Aggregated Model) er den mest omfattende danske makroøkonometriske model. Modellen beskriver dansk økonomi opdelt på 19 brancher og 12 forbrugskategorier, Dam (1996).<sup>1</sup> Modellens relationer (bl.a. privat forbrug, import- og eksportkomponenter, erhvervenes efterspørgsel efter kapital og arbejdskraft, prisrelationer og lønrelation) er hovedsageligt estimeret på baggrund af tidsserier fra

1. Til eksperimenterne i denne artikel er anvendt ADAM version Februar 2002 dokumenteret i Kristensen (2002).



Figur 2.2 Modelstruktur for miljørelaterede satellitmodeller til ADAM.

Nationalregnskabet. På kort sigt er modellen karakteriseret ved en keynesiansk indkomstmultiplikator. På mellemfristet sigt er der crowding out via løn og udenrigshandel, således at modellen på lang sigt har klassiske egenskaber.

EMMA (Energi- og eMissionsModeller til ADAM) er ligeledes en makroøkonometrisk model, der beskriver energiforbruget i de enkelte brancher og i husholdningerne fordelt på syv energityper. Modellens efterspørgselsrelationer er estimeret på tidsserier fra Nationalregnskabet og Danmarks Statistiks Energibalancer. I forhold til tidligere versioner af modellen, beskrevet i bl.a. Andersen m.fl. (1997, 1998) er der videreudviklet på bestemmelsen af erhvervenes og husholdningernes energiefterspørgsel.<sup>2</sup> For erhvervene er de langsigtede priselasticiteter mindsket, således at i gennemsnit over erhvervene er den  $-0,07$  for elforbruget og  $-0,10$  for et aggregat af øvrig energi; tilpasning til langsigtet niveau er på knapt 10 år for el og hurtigere for øvrig energi. Varmeefterspørgslen i husholdningerne tager nu højde for varmebidraget fra elforbrug, og elefterspørgslen er estimeret simultant med bestanden af elapparater. Den langsigtede elpriselasticitet er  $-0,49$ , og effekten det første år er ca. det halve, Grinderslev m.fl., (2004). EMMA's forsyningsmodel modellerer det danske energisystem med repræsentative værker, der producerer den efterspurgte el og fjernvarme. Energiforbruget opgøres i den fysiske enhed Tera Joule (TJ). EMMA kan desuden kobles til tekniske modeller for energisystemet, transport og energiforbrug, hvorved detaljeret information fra bottom-up modeller kan udnyttes til fremskrivning af aggregerede størrelser, Karlsson (2003a). Eksempelvis kan markedsprisen for el hentes fra Energistyrelsens energiforsyningsmodel RAMSES.<sup>3</sup>

2. Til eksperimenterne i denne artikel er anvendt EMMA version 2003 dokumenteret i Karlsson (2003b).

3. [www.ens.dk](http://www.ens.dk)

De tekniske modeller anvendes til, på basis af den overordnede økonomiske udvikling givet fra en fremskrivning med ADAM, at danne aggregerede effektivitetsindeks til EMMA's energiftherspørgselsligninger. Dvs. fremtidig teknologisammensætning og -udvikling kan håndteres detaljeret i de tekniske modeller og overføres til EMMA. Dette er ikke mindst vigtigt for energiforsyningssektoren, hvor en detaljeret modellering er nødvendig for at kunne bestemme brændselsforbrug og emissioner. Aggregerede virkningsgrader og emissionskoefficienter kan derefter overføres til EMMA's simplificerede forsyningsmodel og dermed sikre, at brændselsforbrug og emissioner beregnes på et veldokumenteret grundlag.

Affaldsgenereringen og -deponeringen modelleres som et simpelt accounting system og anvendes i det samlede modelsystem udelukkende til bestemmelse af CH<sub>4</sub> emissionerne fra lossepladser. Modellen er baseret på affaldsdata fra Miljøstyrelsens affaldsstatistik »ISAG« og er en forsimplet version af Miljøstyrelsens affaldsmodel beskrevet i Andersen m.fl. (1999). Modellen skelner mellem syv affaldskilder, der kobles til forskellige økonomiske aktiviteter i ADAM, og grundlæggende forudsættes affaldskoefficienterne at være konstante. Deponeringsandelene er politiske målsætninger og derfor eksogene.

Mængden af spildevand er baseret på Miljøstyrelsens opgørelser over spildevand fra punktkilder og modelleres som et simpelt accounting system relateret til befolkningen, typer af husstande samt for industrielle punktkilder til produktionen i ADAM's erhverv.

Landbruget beskrives i modelsystemet af en delmodel for landbrugsproduktionen i ADAM og i modellerne LADA og ESMERALDA, der kan anvendes både som for- og eftermodeller til ADAM. I ADAM vælger landbruget produktion og hjemmemarkedspris givet verdensmarkedsprisen på landbrugsvarer og erhvervets produktionsomkostninger. Således er der en negativ sammenhæng mellem omkostninger og landbrugsproduktionen og en positiv sammenhæng mellem verdensmarkedspris og den indenlandske pris på landbrugets produkter. Modelleringen betyder, at der ikke vil være temporære positive effekter på landbrugsproduktionen af et positivt indenlandsk efterspørgselsstød.

I LADA opdeles ADAM's landbrugsproduktion i fem underbrancher: vegetabilsk produktion, mælk og kvægproduktion, svin, fjerkræ og øvrig landbrugsproduktion. Landbrugsarealet er en begrænsende faktor i den vegetabilske sektor og kvægsektoren, således at produktionen i den vegetabilske sektor for givet landbrugsareal ikke kan udvides uden en indskrænkning af produktionen i kvægsektoren. LADA og koblingen mellem ESMERALDA, LADA og ADAM er beskrevet nærmere i Andersen m.fl. (2001).

LADA kan kobles med den mere detaljerede landbrugsmodel ESMERALDA, Jensen (1996), hvor landbrugsproduktionen og faktorefterspørgslen er modelleret for 16 driftsgrene på baggrund af estimerede relationer. ESMERALDA transformerer endvidere produktionen i de 16 driftsgrene til fysiske enheder, dvs. antal dyr og tons høst af forskellige afgrøder.

Det samlede kompleks af landbrugsmodeller kan anvendes på flere måder. Til fremskrivninger med særlig fokus på landbruget og til eksperimenter, der forventes at forvride landbrugserhvervets sammensætning f.eks. en afgift på kunstgødning, kan der tages udgangspunkt i ESMERALDA. I dette tilfælde fungerer LADA som aggregeringsmodul, mens ADAM bestemmer ændringer i niveauet for landbrugsproduktionen for en givet sammensætning. Betragtes eksperimenter, der påvirker produktionen i LADAs undersektorer direkte via regler, f.eks. en lavere mælkekvote, kan der tages udgangspunkt i LADA. Endelig kan man, som det er tilfældet i eksperimenterne i afsnit 4 anvende modellerne som eftermodeller til omregning af ADAM's landbrugsproduktion til fysiske enheder og deraf afledte emissioner. I det sidste tilfælde antages det altså, at sammensætningen af landbrugsproduktionen er konstant.

Emissionsmodellerne indeholder grundlæggende emissionskoefficienter koblet til de enkelte fysiske aktiviteter. Typisk defineres emissionskoefficienter som kg stof pr. fysisk aktivitet, f.eks. kg CO<sub>2</sub> pr. TJ kul forbrugt eller kg CH<sub>4</sub> pr. malkeko pr. år. Emissionskoefficienterne beregnes ud fra tekniske forhold som kulstofindhold i brændsler, forbrændingsbetingelser på kraftværker eller mælkeproduktion, fedtindhold, staldindretning og dyrkningsforhold i landbruget. I eksperimenterne i afsnit 4 forudsættes emissionskoefficienterne uændrede, men modellen giver mulighed for at ændre disse eksogent og at beregne de emissionsmæssige konsekvenser heraf.

### 3. Miljøtemaer

#### 3.1. Drivhusgasser

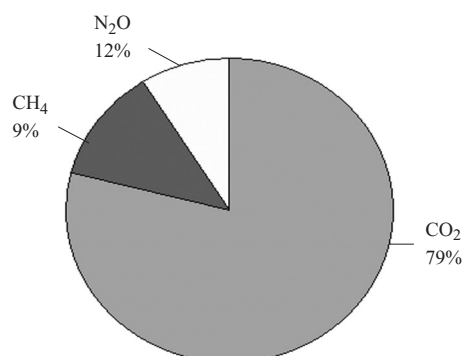
De væsentligste drivhusgasser er CO<sub>2</sub>, CH<sub>4</sub> og N<sub>2</sub>O.<sup>4</sup> Emissionen af 1 kg af de forskellige stoffer har meget forskellig drivhuseffekt, og i internationale klimaforhandlinger vægtes emissionen af de tre gasser til et Global Warming Potential mål (GWP).<sup>5</sup>

De enkelte stoffers bidrag til de danske drivhusgasemissioner i 2002 er vist i figur 3.1, og de forskellige aktiviteters bidrag til emissionen af de enkelte stoffer er vist i tabel 3.1.

---

4. Udover CO<sub>2</sub>, CH<sub>4</sub> og N<sub>2</sub>O inkluderer Kyoto protokollen en lang liste af fluorgasser (HFC'er, PCF'er og SF<sub>6</sub>). For Danmark bidrager disse gasser med omkring 1% af de samlede drivhusgasemissioner, og de er ikke inkluderet i modellen.

5. GWP beregnes som  $GWP = CO_2 + 21 CH_4 + 310 N_2O$ , hvor emissionen af de enkelte stoffer måles i kg af stoffet.



Figur 3.1 Fordeling af danske drivhusgasemissioner i 2002 målt i CO<sub>2</sub> ækvivalenter.

Tabel 3.1 Aktiviteternes bidrag til emissionen drivhusgasser i Danmark, 2002.

	GWP	CO <sub>2</sub>	CH <sub>4</sub>	N <sub>2</sub> O
Energi	0,80	0,97	0,13	0,20
Landbrug	0,15		0,67	0,80
Lossepladser	0,02		0,20	
Andre kilder	0,03	0,03		
Total	1,00	1,00	1,00	1,00

Som det fremgår af figur 3.1 og tabel 3.1, er CO<sub>2</sub> den væsentligste drivhusgas og primært relateret til forbruget af energi. Emissionerne af CH<sub>4</sub> og N<sub>2</sub>O bidrager med hhv. 9% og 12% af de samlede drivhusgasemissioner i 2002 og er primært relateret til landbrugsproduktionen. De landbrugsmæssige emissioner af CH<sub>4</sub> er udelukkende knyttet til husdyrproduktionen, mens emissionerne af N<sub>2</sub>O overvejende er knyttet til den vegetabiliske produktion. CH<sub>4</sub> emissionerne relateret til lossepladser kommer fra organisk materiale, hvor en del af kulstofindholdet under forrådnelse omdannes til CH<sub>4</sub> og emitteres. Da der lovgivningsmæssigt er gennemført begrænsninger i mængden af organisk materiale, der må deponeres, er CH<sub>4</sub> bidraget fra lossepladser aftagende.

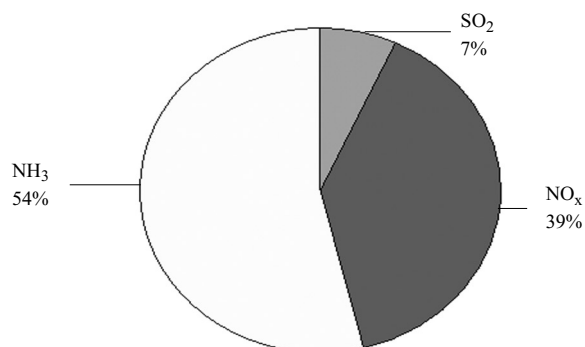
### 3.2. Forsuring

De væsentligste forsurende gasser er SO<sub>2</sub>, NO<sub>x</sub> og NH<sub>3</sub> (svovldioxid, kvælstofoxider og ammoniak). Et mål for den samlede forsurening fås ved at vægte emissionen af de enkelte til forsuringsækvivalenter (PAE: Potential Acid Equivalents).<sup>6</sup>

Målt i forsuringsækvivalenter viser figur 3.2 og tabel 3.2 de enkelte stoffers og aktiviteterets bidrag til de samlede forsurende emissioner i 2002.

6. PAE beregnes som  $PAE = \frac{SO_2}{64} \cdot 2 + \frac{NO_x}{46} + \frac{NH_3}{17}$ , hvor emissionerne af de enkelte stoffer måles i kg af stoffet.





Figur 3.2 De enkelte stoffers bidrag til de samlede danske emissioner af forsurende gasser i 2002.

Tabel 3.2 Aktiviteternes bidrag til emissionen forsurende gasser i Danmark, 2002.

	PAE	SO <sub>2</sub>	NO <sub>x</sub>	NH <sub>3</sub>
Energi	0,48	1,00	1,00	0,03
Landbrug	0,52			0,97
Lossepladser				
Andre kilder				
Total	1,00	1,00	1,00	1,00

Som det fremgår af figur 3.2 og tabel 3.2 er emissionerne af SO<sub>2</sub> og NO<sub>x</sub> relateret til energiforbruget og udgør tilsammen 46% af de samlede danske emissioner af forsurende gasser. For både SO<sub>2</sub> og NO<sub>x</sub> kommer en væsentlig del (hhv. 43% og 26%) fra kraftværkerne, der reguleres ved kvoter. Andelen af SO<sub>2</sub> og NO<sub>x</sub> ud af de samlede forsurende emissioner har været aftagende, ligesom bidraget fra kraftværkerne har været faldende de seneste år. Kraftværkerne kan regulere forbrændingsbetingelserne og rense røggasserne og dermed regulere deres emissioner af SO<sub>2</sub> og NO<sub>x</sub>. Derudover reguleres SO<sub>2</sub> emissionerne ved at reducere det tilladte svovlindhold i de enkelte brændsler. En væsentlig del af NO<sub>x</sub> emissionerne kommer fra vejtransporten (ca. 35%), hvor emissionerne reguleres ved krav om katalysatorer på nye benzindrevne biler og ved normer for dieseldrevne biler.

Den altovervejende del af NH<sub>3</sub> emissionerne kommer fra landbruget, hvor ca. <sup>3</sup>/<sub>4</sub> er fordampning fra husdyrgødning. Resten er relateret til kvælstofomsætningen i den vegetabiliske produktion. NH<sub>3</sub> emissionerne kan reduceres ved at ændre staldindretning, håndtering og anvendelse af husdyrgødning.

#### 4. Multiplikatorer i det samlede modelsystem

For at illustrere modelsystemets samlede egenskaber samt at vise, hvilken type analyser modelsystemet kan anvendes til at foretage, gennemgås i dette afsnit dels et standard offentligt varekøbseksperiment, dels et eksperiment, hvor der introduceres CO<sub>2</sub>-kvoter og handel med disse, og endeligt et eksperiment, hvor der udbygges med ekstra vindmøllekapacitet.

Eksperimenterne er kørt på ADAM-version februar 2002. Grundforløbet afspejler finansredegørelsen 2002, dog modificeret således at forløbet er konsistent med det samlede modelsystem. Det er i forbindelse med kørslerne valgt at antage, at sammensætningen af landbrugsproduktionen er uændret, mens niveauet bestemmes af ADAM. Som nævnt giver modelsystemet mulighed for, at udviklingen i og sammensætningen af ADAM's samlede landbrugsproduktion bestemmes af en detaljeret landbrugsfremskrivning baseret på ESMERALDA.

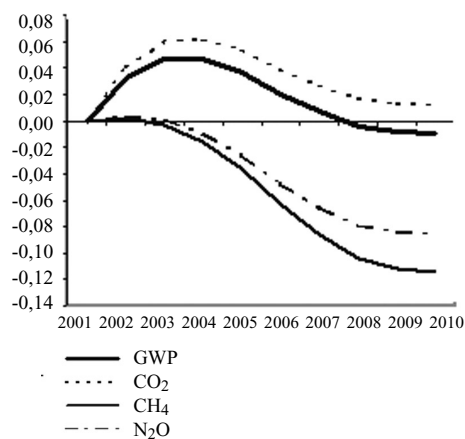
##### 4.1. Offentligt varekøbseksperiment

I dette eksperiment hæves det offentlige forbrug permanent med 1 mia. 1995-kr. Den finanspolitiske ekspansion finansieres via nedbringelse af det offentlige overskud. Som det fremgår af tabel 4.1 er den umiddelbare effekt af stødet (første år 2002), at produktionen i den offentlige og private servicesektor øges med hhv. 0,34% og 0,18%. Aktiviteten i fremstillingserhvervene øges med 0,04%, mens landbrugsproduktionen er uændret. Totalt set øges BNP med 0,08%, og energiforbruget øges med 0,04%. Figur 4.1 viser, at CO<sub>2</sub> emissionerne øges svarende til det øgede energiforbrug, samt at emissionen af CH<sub>4</sub> og N<sub>2</sub>O, der primært er relateret til landbrugsproduktionen, stort set er uændrede (som det ses af tabel 3.1, kommer hhv. 67% og 80% af CH<sub>4</sub> og N<sub>2</sub>O emissionerne fra landbruget). Med hensyn til emissionen af forsurende gasser viser figur 4.2, at de energirelaterede emissioner af SO<sub>2</sub> og NO<sub>x</sub> øges, mens de overvejende landbrugsrelaterede emissioner af NH<sub>3</sub> stort set er uændrede.

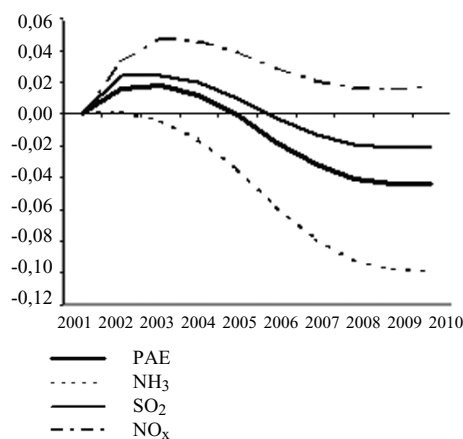
På sigt giver det øgede offentlige forbrug anledning til crowding-out som følge af øget indenlandsk lønpres og deraf følgende tab af konkurrenceevne. Konkurrenceevnetabet fører til, at produktionen i landbruget og fremstillingserhvervene fra år 2005 reduceres i forhold til grundforløbet. Produktionen i serviceerhvervene er større end i grundforløbet gennem hele perioden blandt andet som følge af øget indenlandsk efterspørgsel. Efter otte år (i år 2010) er stigningen i BNP reduceret til 0,02%, og stigningen i energiforbruget er reduceret til 0,01%. Emissionsmæssigt betyder dette, at stigningen i CO<sub>2</sub> emissionerne efter otte år (2010) kun er på 0,01%, og at de landbrugsrelaterede emissioner af CH<sub>4</sub> og N<sub>2</sub>O reduceres ganske betydeligt (hhv. med 0,12% og 0,08%). Totalt set er effekten, at den umiddelbare stigning i GWP på sigt ændres til et fald. Tilsvarende reduceres emissionen af forsurende gasser på sigt. Emissionen af SO<sub>2</sub> og NH<sub>3</sub> reduceres pga. det mindre forbrug af brændsler og den reducerede land-

Tabel 4.1 Offentligt varekøbseksperiment.

Ændringer i %	År		
	2002	2005	2010
BNP i faste priser	0,08	0,03	0,02
Privat forbrug i faste priser	0,03	0,03	0,05
Beskæftigelse	0,04	0,04	-0,01
<i>Produktion i faste priser i:</i>			
Landbrug	0,00	-0,08	-0,14
Serviceerhverv	0,18	0,13	0,11
Offentlig service	0,34	0,33	0,33
Fremstillingserhverv	0,04	-0,09	-0,14
Gennemsnitlig energipris	0,00	0,02	0,03
Danmarks endelige energiforbrug	0,04	0,03	0,01



Figur 4.1 Ændring i emission af drivhusgasser målt i % af grundforløb.



Figur 4.2 Ændring i emission af forurenende gasser målt i % af grundforløb.

brugsproduktion. Når der ikke er en tilsvarende reduktion i  $\text{NO}_x$  emissionerne, skyldes det, at en væsentlig del af  $\text{NO}_x$  emissionerne er relateret til transportenergiforbruget, der i eksperimentet er øget med 0,07% i år 2010.

Ud fra dette eksperiment er en væsentlig konklusion, at hvis man vil se på miljømæssige effekter af den økonomiske politik, er det ikke tilstrækkeligt at se på emissionen af et enkelt stof eller blot på de energirelaterede emissioner. Ser man udelukkende på  $\text{CO}_2$  emissionerne, vil et øget offentligt forbrug øge miljøbelastningen, også på sigt. Tager man derimod hensyn til de landbrugsrelaterede emissioner, så reducerer et øget offentligt forbrug de samlede emissioner af drivhusgasser på sigt. Af figur 4.2

fremgår, at det samme gør sig gældende, hvis man ser på emissionen af forsurende gasser.

En grundlæggende forklaring på de samlede effekter er, at det offentlige forbrug er mindre miljøbelastende, end den produktion et øget offentligt forbrug fortrænger.

#### *4.2. Indførelse af international CO<sub>2</sub>-kvotehandel*

I dette eksperiment indføres et CO<sub>2</sub>-kvotehandelsystem, hvor kvotekøbet dækker mankoen mellem den faktiske udledning af drivhusgasser i Danmark og den tilladte ifølge de i EU aftalte reduktionsforpligtigelser. Alle drivhusgasser er omfattet af reduktionsforpligtelsen, men i de her gennemførte beregninger er omkostningerne til kvotekøbene udelukkende pålagt energiforbruget i forhold til kulstofindholdet i de anvendte brændsler. Modelsystemet kan håndtere, at det f.eks. pålægges landbruget at købe et bestemt antal kvoter pr. ko, men i dette eksperiment er det valgt at håndtere CO<sub>2</sub>-kvotekøbene som en CO<sub>2</sub>-afgift på brændsler, hvor provenuet sendes ud af landet (køb af udenlandske kvoter). Kvotekøb svarende til den samlede manko for drivhusgasser finansieres dermed via brændselsafgifter.

I Karlsson og Kristensen (2004) præsenteres to scenarier for kvoteprisen; en lav kvotepris på 40 kr. pr. ton CO<sub>2</sub>-ækvivalent og en høj kvotepris på 100 kr. pr. ton CO<sub>2</sub>-ækvivalent målt i faste 1995-kr. I det følgende scenario anvendes den lave kvotepris. Til bestemmelse af mankoen for den danske udledning er der regnet med et krav om 21% reduktion af de danske drivhusgasudledninger i år 2008-12 i forhold til niveauet i 1990. I EU er der indført en prøveperiode i 2005 til 2007 for anvendelse af Kyoto-mekanismerne (JI, CDM og CO<sub>2</sub>-kvotekøb) til opfyldelse af landenes reduktionsmål. I denne periode vurderes landene ligeledes på, om de er på vej mod deres aftalte reduktionsmål, dvs. hvor stort gabet er mellem den faktiske udledning og en lige linie mellem deres udledning i 1990 og målsætningen i 2010. I scenariet regnes det med, at Danmark også skal opfylde reduktionsmålet i prøveperioden og derfor allerede anvender kvotekøb fra 2005 til at lukke gabet.

I scenariet er mankoen 13,5 mio. ton CO<sub>2</sub>-ækvivalenter første år (2005), hvilket med en kvotepris på 40 kr. pr. ton medfører en gennemsnitlig stigning i energiprisen på 1,8%.

Over tid øges mankoen og dermed også udgiften til køb af CO<sub>2</sub>-kvoter, dvs. den gennemsnitlige energipris inklusiv udgifter til CO<sub>2</sub>-kvoter stiger. Da kravet om reducerede drivhusgasudledninger dækkes ved køb af udenlandske CO<sub>2</sub>-kvoter, er effekterne på det hjemlige energiforbrug og emissioner derfor forholdsvis beskedne (se figur 4.3 og 4.4). De priser, danske erhverv konkurrerer med, antages at stige med 45% af stigningen i de danske priser, idet det danske behov for at købe CO<sub>2</sub>-kvoter er større end i andre lande som følge af en større reduktionsforpligtelse (EU's samlede reduktionsforpligtelse er på

8%). Endvidere forudsættes, at kvotesystemet fører til et væksttab på de danske eksportmarkeder på gennemsnitligt 0,03%, jf. EU-Kommissionen (2002).<sup>7</sup>

Krav om kvotekøb øger som nævnt omkostningen til energi. Den direkte effekt af kvoterne er derfor højere omkostninger i erhvervene som følge af højere energipriser. Idet det i udlandsforudsætningerne er antaget, dels at Danmark rammes hårdere af systemet af omsættelige kvoter end vores gennemsnitlige samhandelspartner, og dels at systemet fører til lavere vækst hos aftagerlandene, fås, at den indenlandske aktivitet falder som følge af forringet konkurrenceevne og lavere vækst på eksportmarkederne. Således falder beskæftigelsen indtil omkring år 2008 med 0,06%, se tabel 4.2. Den faldende beskæftigelse medfører aftagende lønpres og dermed en reetablering af konkurrenceevnen over for udlandet, hvorfor den negative effekt på produktionen i eksporterhvervene (fremstillingserhvervene) er mindre i den sidste del af perioden. Eksporten kommer dog ikke op på samme niveau som i grundforløbet blandt andet som følge af væksttabet på de danske eksportmarkeder.

Privatforbruget rammes relativt hårdt af stigende energipriser, dels som følge af en stigning i det generelle prisniveau (erhvervene vælter deres ekstra energiomkostninger over på forbrugerprisen) og dels direkte på deres brændselsforbrug. Samlet giver de indførte kvoter anledning til en reduktion i BNP på 0,08% i år 2007 og 0,06% i år 2010.

Som det fremgår af tabel 4.2 rammes fremstillingserhvervene hårdt i starten, i år 2005 med en nedgang på 0,05% og i år 2007 på 0,16%. Efterhånden som kvotebetalingen stiger, rammes også de øvrige erhverv. Serviceerhvervenes produktion ender i år 2010 med en nedgang på 0,05%, fremstillingserhvervenes produktion falder 0,12%, og landbrugets produktion falder i år 2010 med 0,09%.

Det faldende aktivitetsniveau – og dermed faldende energiforbrug – resulterer i en nedgang i udledningen af drivhusgasser og forsurende gasser. Som tidligere nævnt er hele omkostningen til kvotekøbene i dette eksempel lagt på brændsler. I figur 4.3 ses derfor, at CO<sub>2</sub>-udledning falder væsentligt mere end udledningen af metan (CH<sub>4</sub>) og lattergas (N<sub>2</sub>O). Dette skyldes, at de højere energipriser giver anledning til en faldende aktivitet og substitution væk fra energi, mens de ikke-energi-relaterede emissioner (metan og lattergas) kun falder i takt med den faldende landbrugsproduktion.

For de forsurende gasser er det ligeledes de energi-relaterede emissioner (SO<sub>2</sub> og NO<sub>x</sub>), der falder markant, mens ammoniakudledningen (NH<sub>3</sub>) falder i takt med landbrugsproduktionen – se figur 4.4. Når SO<sub>2</sub>-udledningen falder noget mere end NO<sub>x</sub>-udledningen, skyldes det, at der er et sammenfald mellem CO<sub>2</sub>-tunge brændsler og svovlholdige brændsler. Disse brændsler rammes hårdest prismæssigt som følge af

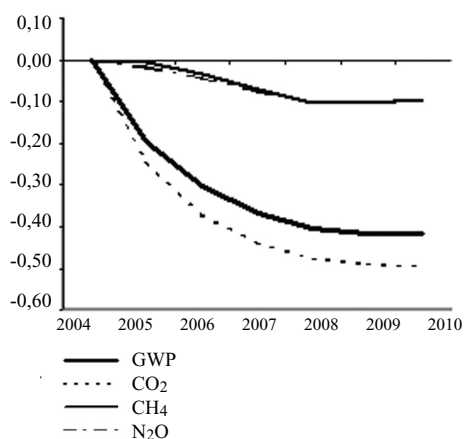
---

7. Udlandsforudsætningerne kan forfines væsentligt, men som følge af eksperimentets illustrative karakter følges dette ikke her.

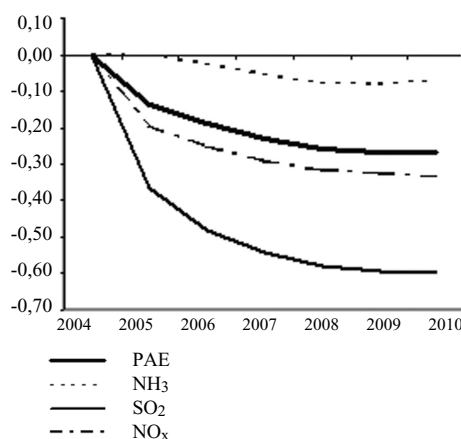
Tabel 4.2 Effekt af indførelse af CO<sub>2</sub>-kvoter.

Ændringer i %	År		
	2005	2007	2010
Manko i % af samlet GWP <sup>(1)</sup>	19	22	25
Bruttonationalproduktet	-0,04	-0,08	-0,06
Privat forbrug i faste priser	-0,09	-0,07	-0,10
Beskæftigelse	-0,02	-0,06	-0,06
<i>Produktion i faste priser i:</i>			
Landbrug	0,00	-0,07	-0,09
Serviceerhverv	-0,03	-0,06	-0,05
Offentlig service	0,00	0,00	0,00
Fremstillingserhverv	-0,05	-0,16	-0,12
Gennemsnitlig energipris	1,84	1,90	1,99
DK's endelige energiforbrug	-0,23	-0,40	-0,46

Note: (1) Mankoen's andel af GWP illustrerer, hvor meget Danmark overskrider udledningsmålsætninger.



Figur 4.3 Ændring i emission af drivhusgasser målt i % af grundforløb.



Figur 4.4 Ændring i emission af forurenende gasser målt i % af grundforløb.

den pålagte CO<sub>2</sub>-afgift, og forbruget af disse brændsler reduceres derfor relativt mere end forbruget af andre brændsler.

Modellsimuleringen peger på, at kvotesystemet fører til et BNP-tab i år 2010 på 0,06% i forhold til forløbet, hvor reduktionsforpligtigelserne ikke overholdes. Imidlertid er det vigtigt at slå fast, at resultaterne gælder under den forudsatte udvikling i brændselspriser (ca. 2% stigning p.a.), og forudsat der er kvoter at købe til den forudsatte kvotepris. I scenarierne bibeholdes det eksisterende landbrugs-, energi- og trans-

portsystem, og dermed er der ikke påbegyndt en omstilling af disse systemer, som kan sikre en fremtidig indenlandsk reduktion af drivhusgasudledningen. Scenarierne siger derfor ikke noget om, hvordan Danmark står over for en ny periode med formodet skrappe reduktionskrav og mangel på udbudte kvoter. Dette vil kræve analyser, der kører længere frem i tid, hvilket det præsenterede modelsystem også vil kunne anvendes til ved at inddrage information fra bl.a. detaljerede energisystemmodeller, jf. Karlsson (2003a).

Ved anvendelse af en højere kvotepris vil resultaterne stort set kunne skaleres. Ved en kvotepris på 100 kr. pr. ton CO<sub>2</sub>-ækvivalenter reduceres BNP og beskæftigelsen med 0,16% i år 2010.

#### 4.3. Udbygning med ekstra vindkraft

I dette eksperiment foretages en udbygning i energiforsyningssystemet med ekstra vindkraft. Udbygningen foretages løbende, og kapaciteten er i år 2010 på 1000 MW eller 25% højere end i grundforløbet. De ekstra vindmøller antages havplacert, og ekstra omkostninger til investeringer, drift og vedligehold af energiforsyningssystemet antages at være 11 mio. kr. pr. installeret ekstra MW vind. Denne udbygning resulterer i en årlig ekstra investering på ca. 1,5 mia. 1995-kr.

De ekstra investeringer i energisektoren giver i første omgang (år 2004 til 2007) anledning til en øget beskæftigelse (0,14%) og som følge heraf et øget privatforbrug (0,09%). Den øgede aktivitet giver gennem hele forløbet anledning til en forringelse af konkurrenceevnen over for udlandet, som følge af indenlandsk lønpres. Dette kommer blandt andet til udtryk som faldende produktion i eksporterhvervene (landbrug og fremstillingserhverv), der permanent mister eksportandele. Eksporten falder gennem perioden med ca. 0,2%.

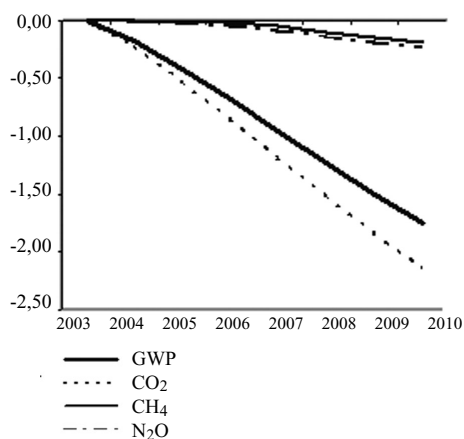
Den gennemsnitlige energipris stiger kun med 0,02% i år 2007 og 0,01% i år 2010. Dette skyldes bl.a. en antagelse om, at en dansk udbygning med vindkraft i nævnte omfang ikke vil have indflydelse på elprisen. De udførte beregninger er derfor kørt med en eksogen elpris, som følger en antaget spotmarkedspris. Investeringen i vindmøllerne finansieres dermed i scenariet via en reduktion af profitten i elforsyningssektoren. Stigningen i den gennemsnitlige energipris er dermed en sammensætnings-effekt, der først og fremmest kommer fra et øget transportenergiforbrug (stiger med 0,08% i år 2007 og 0,06% i 2010).

Figur 4.5 viser, at CO<sub>2</sub>-udledningen falder støt i takt med udbygningen med ekstra vindkapacitet og ender med at være 2,14% lavere end i grundforløbet. De øvrige drivhusgasser (CH<sub>4</sub> og N<sub>2</sub>O) følger derimod udviklingen i landbruget. SO<sub>2</sub> og NO<sub>x</sub> udledningen falder ikke svarende til den reducerede elproduktion på de centrale kraftværker, da der er antaget en fast udledningstilladelse for SO<sub>2</sub> og NO<sub>x</sub> på disse værker.

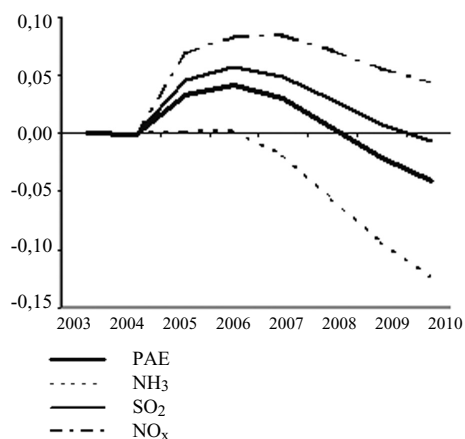
Tabel 4.3 Ekstra udbygning med vindkraft.

Ændringer i %	År		
	2004(1)	2007	2010
Ekstra vindkapacitet ifht. grundforløb i MW samt i %	143MW ~4,6%	571MW ~16,0%	1000MW ~25,0%
Bruttonationalproduktet	0,00	0,10	0,03
Privat forbrug i faste priser	0,00	0,09	0,06
Beskæftigelse	0,00	0,14	0,06
<i>Produktion i faste priser i:</i>			
Landbrug	0,00	-0,03	-0,17
Serviceerhverv	0,00	0,13	0,08
Offentlig service	0,00	0,00	0,00
Fremstillingserhverv	0,00	0,14	-0,20
Gennemsnitlig energipris	0,00	0,02	0,01
DK's endelige energiforbrug	0,00	0,07	0,02

Note: (1) I det kørte eksperiment er der effekter første år (2004), men først ude på tredje decimal og derfor ikke synlig i tabellen.



Figur 4.5 Ændring i emission af drivhusgasser målt i % af grundforløb.



Figur 4.6 Ændring i emission af forurenende gasser målt i % af grundforløb.

Udledning af NH<sub>3</sub> følger landbrugsproduktionen og ender 0,12% lavere end grundforløbet i år 2010. Samlet giver dette et stigende PAE indeks i første halvdel af perioden, men ender med en nettonedgang på 0,05%.

Udbygningen med ekstra vindmøllekapacitet har på kort sigt en betydelig positiv beskæftigelseseffekt, og gennem hele forløbet mindskes udledningen af drivhusgasser.



Konkurrenceevnen rammes negativt, da de øgede indenlandske investeringer flytter arbejdskraften væk fra eksportrelaterede produkter. Den relative nedgang i fremstillingserhvervenes produktion dækker derfor over, at nogle erhverv (herunder bl.a. leverandører til byggeri) har en stor relativ fremgang som følge af investeringerne i vindmøller.

## 5. Konklusion

Det præsenterede modelsystem er udviklet til overordnede nationale analyser af, hvordan den økonomiske udvikling påvirker udviklingen i centrale miljøindikatorer, samt til at analysere de samfundsøkonomiske effekter af overordnede miljøtiltag. For at få konsistens i fremskrivningen af flere forskellige miljøindikatorer er det nødvendigt at anvende samme økonomiske fremskrivning, og for at få en vurdering af de samlede miljømæssige effekter af økonomiske ændringer er det væsentligt at se på udviklingen i emissionerne af flere forskellige stoffer, som det fremgår af de gennemgåede multiplikatoreksperimenter.

De gennemgåede eksperimenter skal ses som eksempler på mulighederne med ADAM og de tilhørende miljømodeller. Eksperimenterne illustrerer bl.a. vigtigheden af at kunne håndtere sammensætningseffekter erhvervene imellem. I varekøbs- og vindmølleeksperimentet sker der en forskydning væk fra landbruget og andre eksport-erhverv, og dermed er der bl.a. en ændret sammensætning af de danske udledninger af drivhusgasser.

Analysemulighederne med modelsystemet er mangfoldige. Lige fra implementering af en enkelt miljøafgift til hele skattereformer, samfundsøkonomiske konsekvensberegninger for opfyldelsen af diverse miljømålsætninger, fremskrivning af miljøindikatorer (drivhusgasser, forsurende gasser og næringsstofbelastning) baseret på den førte finanspolitik og desuden muligheden for at omsætte scenarier fra tekniske modeller til makroøkonomiske konsekvensberegninger.

### Litteratur

- AEA Technology. 1998. *Options to Reduce Methane Emissions (Final Report)*. A Report Produced for DGXI. Judith Bates. November 1998.
- Alfsen, K. H., T. Bye, E. Holmøj, red. 1996. *MSG-EE: An Applied General Equilibrium Model for Energy and Environmental Analyses*, Social Economic Studies 96. Statistics Norway.
- Andersen, F. M., D. Grinderslev, M. Werner. 2003. Environmental Satellite Models for a Macroeconomic Model. *Environmental and Resource Economics* 24, s. 197-212.
- Andersen, F. M. 2001. *Miljømodeller og makroøkonomiske analyser i Bæredygtighed, økonomi og velfærd*, red., P. Andersen, Mortensen J. B. Nielsen, H. Ø.) Det Strategiske Miljøforskningsprogram.
- Andersen, F. M., red. Werner, M. Jensen, J. D. Jensen, T. S. Henriksen, G. T. Olsen, A. Illerup, J. B. Nielsen, C. Winther, M. 2001. *Environmental satellite models for ADAM*. Statistics Denmark.
- Andersen, F. M. Fenhann, J. Larsen, H. V. Schleisner, L. 1999. *A scenario model for the generation of waste*. Danish Environ-

- mental Protection Agency, Copenhagen, 1999, Environmental project, nr. 434.
- Andersen, F. M., H. Klinge Jacobsen, P. E. Morthorst, A. Olsen, M. Rasmussen, T. Thomsen, P. Trier. 1998. EMMA: *En energi- og miljørelateret satellitmodel til ADAM*. Nationaløkonomisk Tidsskrift 136, s. 333-349.
- Andersen, F. M., H. K. Jacobsen, P. E. Morthorst, A. Olsen, M. Rasmussen, T. Thomsen, P. Trier. 1997. *Energi- og emissionsmodeller til ADAM*. Danmarks Statistik.
- Barker, T., Gardiner B., Milas C. 1998. *An Energy-Environmental-Economy Model for Europe E3ME Version 2.1*. Cambridge Econometrics.
- Capros, P., T. Georgakopolus, D. Van Rege-morter, S. Proost, T. Schmidt, og K. Conrad. 1997. *The GEM-E3 General Equilibrium Model for the European Union*. Economic & Financial Modelling, Summer/Autumn 1997.
- Dam, P.U., red. 1996. *ADAM en model af dansk økonomi, marts 1995*. Danmarks Statistik.
- Danmarks Miljøundersøgelser. 2004. [http://www2.dmu.dk/1\\_Viden/2\\_miljoe-tilstand/3\\_luft/4\\_adaei/greenhouse\\_gases\\_en.asp#emissions](http://www2.dmu.dk/1_Viden/2_miljoe-tilstand/3_luft/4_adaei/greenhouse_gases_en.asp#emissions) (juli 2004).
- ECOFYS. 1998. *Emission Reduction Potential and Costs for Methane and Nitrous Oxide in the EU-15*. Hendriks, C. A., D. Jager, and K. Blok. ECOFYS, Utrecht, The Netherlands.
- EU-Kommissionen. 2002. [http://www.europa-eu-n.org/articles/lt/article\\_1421\\_lt.htm](http://www.europa-eu-n.org/articles/lt/article_1421_lt.htm). (September 2004).
- Grinderslev, D. Karlsson, K. Bjørsted B. 2004. *Household Demand for Electricity and Electrical Appliances*. Conference Proceedings, 6. IAEE European conference, Zürich 2004.
- IPCC. 1996, a. *IPCC Guidelines for National Greenhouse Gas Inventories*. OECD.
- IPCC. 1996, b. *Climate Change 1995. The Science of Climate Change*. Contribution of Working Group I to the Second Assessment Report of the Intergovernmental Panel on Climate Change. Ed. Houghton, J. T. m.fl. Cambridge University Press, 572 s.
- Jensen, J. D. 1996. *An Applied Econometric Sector Model for Danish Agriculture (ESMERALDA)*. Danish Institute of Agricultural and Fisheries Economics, Report nr. 90.
- Karlsson, K., T. M. Kristensen. 2004. Samfundsøkonomiske omkostninger og CO<sub>2</sub>-reduktion ved varmebesparelser i bolig-massen. *Arbejdsrapport* KKA-09.02.04, [www.dst.dk/adam](http://www.dst.dk/adam), Økonomiske modeller, Danmarks Statistik.
- Karlsson, K. 2003, a. *Modeller i energiplanlægning med henblik på bæredygtig udvikling*. Ph.d.-afhandling, BYG.DTU, Danmarks Tekniske Universitet.
- Karlsson, K. 2003, b. EMMA – Oversigtsrapport. *Arbejdsrapport* KKA-21.10.03, [www.dst.dk/adam](http://www.dst.dk/adam), Økonomiske modeller, Danmarks Statistik.
- Kristensen, T. M. 2002. Vedr. ADAM, februar 2002. *Arbejdsrapport* TMK-23.09.02, [www.dst.dk/adam](http://www.dst.dk/adam), Økonomiske modeller, Danmarks Statistik.
- Miljøstyrelsen. 2003. *Denmark's Third National Communication on Climate Change*.

# Bør virksomheder oplyses om sandsynligheden for kontrol?

Lars Gårn Hansen

akf, amternes og kommunernes forskningsinstitut, E-mail: lgh@akf.dk

Signe Krarup

akf, amternes og kommunernes forskningsinstitut, E-mail: sk@akf.dk

*SUMMARY: Regulatory authorities often both have limited monitoring budgets and must apply preset punishments resulting in sub-optimal compliance rates. An important task of regulatory authorities is therefore to identify monitoring strategies that maximise compliance under these constraints. The literature often assumes that the monitoring probabilities chosen by authorities are known to firms and that uncertainty reduces welfare. However, looking at areas of monitoring in Denmark, authorities expect the effect of the monitoring to be larger when monitoring probabilities are not disclosed to firms. In this article we analyse the effect on compliance rates of disclosing monitoring probabilities to the regulated firms.*

---

## 1. Introduktion

I et moderne samfund som det danske er virksomheder og privatpersoner underlagt offentlig regulering på snart sagt alle områder af deres virke. For eksempel er virksomheder pålagt skatter (moms, punktafgifter, indkomstskat mv.) og underlagt regler på områder som miljø, arbejdsmiljø, levnedsmidler mv. Om end visdommen i den enkelte afgift eller regel kan diskuteres, er der næppe tvivl om, at et moderne samfund forudsætter et væsentligt omfang af offentlig regulering og offentligt finansierede udgifter (og dermed skatter). I Danmark findes et antal offentlige myndigheder, der hver på deres område varetager kontrol- og håndhævelsesopgaver, og der medgår et ikke ubetydeligt ressourceforbrug hertil. Vi har undersøgt tre danske kontrolmyndigheder med et samlet ressourceforbrug på skønsmæssigt 1 mia. kr. pr. år. I praksis medfører kontrollen ofte langt fra 100% regeloverholdelse, og den mangelfulde regeloverholdelse giver (formodentlig) anledning til et samfundsøkonomisk velfærdstab. Dermed synes det relevant at overveje, om der fra økonomisk teori kan udledes brugbare principper for, hvorledes en effektiv offentlig overholdelseskontrol kan tilrettelægges.

---

Denne artikel er blevet tilvejebragt på baggrund af finansiering fra Statens Samfundsvidenskabelige Forskningsråd (SSF) og akf, amternes og kommunernes forskningsinstitut.

I den tidlige økonomiske kontrol og håndhævelseslitteratur (Becker 1968) antages, at den forventede gevinst er afgørende for en rationel økonomisk agents beslutning om, hvorvidt en regel skal overholdes eller ej. Dermed kan en høj overholdelsesgrad sikres med selv en beskeden kontrolindsats (og dermed ringe opdagelsessandsynlighed), blot straffen for overtrædelse er tilstrækkelig stor. I praksis sætter almindelige retsprincipper og rimelighedsbetragtninger dog grænser for, hvor høj en straf kan være, idet den typisk skal stå i et passende forhold til overtrædelsen.<sup>1</sup> Af denne grund må der på de fleste områder foretages en reel afvejning mellem de ressourcer, der bruges på kontrol, og graden af regeloverholdelse. Den enkelte kontrollerende myndighed har tillige et begrænset budget til rådighed. Under denne dobbelte restriktion bliver den centrale opgave for kontrolmyndigheden at tilrettelægge kontrolindsatsen, så der opnås størst mulig regeloverholdelse med de givne kontrolressourcer og strafferammer.<sup>2</sup> Der er fremkommet en række artikler, der viser, hvordan man under disse restriktioner kan designe kontrolstrategier, der sikrer, at velfærdsgevinsten ved kontrollen maksimeres (eller velfærdsmkostninger forbundet med, at reguleringen ikke overholdes, minimeres) inden for det givne budget. I Greenberg (1984), Russell (1990) og Scotchmer (1987) vises, at en kontrolstrategi, der indebærer, at nogle kontrolleres oftere end andre, vil betyde, at man samlet set kan opnå en højere grad af regeloverholdelse, end hvis alle blev kontrolleret med den samme sandsynlighed.

I størstedelen af litteraturen går man ud fra, at de regulerede har kendskab til kontrolstrategien og handler optimalt ud fra dette.<sup>3</sup> Dette synes imidlertid ikke at afspejle virkeligheden i Danmark, og flere artikler om kontrolpolitik, Polinsky & Shavell (2000); Reinganum & Wilde (1988), peger på, at usikkerhed om kontrolsandsynligheden hos de kontrollerede også synes at være almindelig i mange andre lande.

En gren af litteraturen, se f.eks. Russell (1990) og Scotchmer & Slemrod (1989), analyserer konsekvenserne af usikkerhed om kontrolresultatet (altså det forhold, at kontrolmyndigheden kan tage fejl og nogen gange godkende regelovertrædelser og andre gange pålægge sanktioner, selv om reglerne er overholdt). Dette er en interessant

---

1. Dog er offentliggørelsen af fødevarekontrolresultater et eksempel på, at det nogen gange kan lade sig gøre at finde »acceptable« sanktioner med stor straffevirkning i tråd med Beckers oprindelige ide.

2. Da den forventede straf er sandsynligheden for opdagelse gange straffen givet opdagelse, vil en lav opdagelsessandsynlighed kunne opvejes af en høj straf. Hvis man f.eks. indførte dødsstraf for at køre med S-tog uden billet, kunne man givetvis opnå en høj grad af overholdelse selv med en væsentligt lavere kontrolindsats end i dag. Det nuværende bødeniveau (på 500 kr.) må imidlertid opfattes som givet ud fra principielle rets- og rimelighedsovervejelser, hvorfor kontrolopgaven består i at tilrettelægge den givne kontrolindsats, så den får størst mulig præventiv virkning (f.eks. kan indsatsen differentieres efter strækning, årstid, tidspunkt på døgnet mv.).

3. Tidlige normgivende arbejder i denne tradition er bl.a. Becker (1968) og Greenberg (1984). Desuden viser Lando & Shavell (2004), at gevinsten ved at målrette kontrol bl.a. forudsætter, at de kontrollerede har kendskab til deres kontrolsandsynlighed.

problemstilling, men en anden end usikkerhed om kontrol*sandsynligheden*, der er i fokus her.

Der er kun få artikler, der fokuserer på usikkerhed om kontrol*sandsynligheden*. I disse analyseres *konsekvenserne* af usikkerhed om kontrol*sandsynligheden* for tilrettelæggelsen af den optimale kontrolpolitik, idet det *antages*, at usikkerhed om kontrol*sandsynligheden* er velfærdsreducerende. Således undersøger Bebchuk & Kaplow (1992), hvilken håndhævelsespolitik der er optimal, når individer ikke kender kontrol*sandsynligheden* med sikkerhed. De viser, at hvis kontrol*sandsynligheden* øges, og straffen tilsvarende sættes ned, vil de negative konsekvenser af usikkerhed kunne reduceres. I Garoupa (1999) generaliseres resultaterne i Bebchuk & Kaplow, idet der her ses på konsekvenserne af usikkerhed i forhold til både kontrol*sandsynligheder* og sanktioner. Han viser, at dette betyder, at nogle individer vil have incitament til at indhente information om myndighedens kontrolstrategi. Når dette sker, vil det ikke længere være optimalt at fastsætte de højest mulige bøder, som Becker anbefalede. I en relateret artikel viser Garoupa (1998), at usikkerhed også betyder, at det er optimalt, hvis straffen for overtrædelse afhænger af og stiger med individers velstand. Også Ben-Shahar (1997) generaliserer resultaterne i Bebchuk & Kaplow, men på en anden måde. I hans model antages det, at usikkerheden bestemmes endogent, hvilket betyder, at det er optimalt at fastsætte højere kontrol*sandsynligheder* end ellers. I alle tre artikler antages, at usikkerhed er en ulempe, og det undersøges, hvordan bøder og kontrol*sandsynligheder* skal tilpasses det givne usikkerhedsniveau. I denne artikel fokuserer vi på effekten af usikkerhed om kontrol*sandsynligheden* ved givne bøder og kontrol*sandsynligheder*, og denne tilgang er ny i litteraturen.

Det er myndighedens tilrettelæggelse af kontrolindsatsen, der er vores fokus. Vi forudsætter, at kontrolmyndigheden har opdelt virksomhederne i en række undergrupper, og vi betragter kontrolstrategien for en af disse undergrupper, hvor myndigheden ud fra overordnede kriterier har fastlagt en bestemt kontrol*sandsynlighed*. Vi vil forsøge at vurdere, om den kontrollerende myndighed bør søge at skjule den *sandsynlighed*, hvormed der kontrolleres, eller om myndigheden tværtimod bør sørge for, at de kontrollerede virksomheder i den pågældende undergruppe får så præcise oplysninger herom som muligt.<sup>4</sup>

Ser man på dansk kontrolpraksis, fastlægges der vejledende tilsynsfrekvenser for forskellige hovedgrupper af virksomheder. Dette betyder, at virksomhederne i et vist

---

4. Når myndigheden giver virksomheden præcise oplysninger om kontrol*sandsynligheden*, betyder det ikke, at det enkelte kontrolbesøg forhånds anmeldes eller lignende. Det, virksomheden får oplysninger om, er alene den *sandsynlighed*, hvormed den kan forvente at blive kontrolleret (f.eks. at den kan forvente kontrolbesøg i gennemsnit hvert andet år). Vi ser endvidere bort fra den mulighed, at myndigheden bevidst vildleder virksomhederne (f.eks. ved at udmelde en for høj kontrol*sandsynlighed*). Dels vil en sådan adfærd næppe være moralsk acceptabel (hvis overhovedet lovlig), dels er der risiko for, at vildledningen bliver afsløret, hvorfor en sådan strategi næppe er holdbar i det lange løb.

omfang kan udregne den gennemsnitlige kontrolsandsynlighed for den hovedgruppe, de tilhører. Det er imidlertid klart, at kontrolmyndighederne kan differentiere kontrolindsatsen yderligere svarende til, at virksomhederne i en hovedgruppe opdeles i undergrupper med forskellige kontrolsandsynligheder.<sup>5</sup> Kontrolmyndighederne informerer ikke aktivt de kontrollerede virksomheder om den sandsynlighed, hvormed de kontrolleres, og er heller ikke ved henvendelse villig til at oplyse herom. Bag denne praksis ligger en formodning om, at størst mulig usikkerhed om kontrolsandsynligheden hos de regulerede virksomheder øger regeloverholdelsen.<sup>6</sup>

Manglende information betyder, at den enkelte virksomheds skøn over kontrolsandsynligheden typisk afviger fra den faktiske. Imidlertid virker det ikke plausibelt, at virksomhederne i gennemsnit systematisk og vedvarende skulle overvurdere eller undervurdere den faktiske kontrolsandsynlighed. Hvis manglende information skal give en stabil og vedvarende forøgelse i regeloverholdelsen, må baggrunden derfor være, at den øgede variation i virksomhedernes skøn omkring den sande værdi, som usikkerhed medfører, i sig selv øger regeloverholdelsen (selv om virksomhederne i gennemsnit skønner rigtigt). Om det faktisk forholder sig sådan, har imidlertid (så vidt vi ved) aldrig været undersøgt i den økonomiske kontrol- og håndhævelseslitteratur.

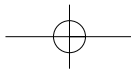
Med udgangspunkt i en simpel model viser vi, at forskellige oplysningsstrategier påvirker andelen af virksomheder, der vælger at overholde reglerne, forskelligt. Mere præcist viser vi, at effekten afhænger af, hvilken sammenhæng der er mellem kontrolsandsynligheden og overholdelsesandelen. Vores resultater peger på, at hvis denne sammenhæng er konveks (f.eks. hvis overholdelsesandelen stiger eksponentielt med kontrolsandsynligheden), vil usikkerhed alt andet lige medføre, at flere virksomheder vælger at overholde reglerne. Hvis denne sammenhæng derimod er konkav, vil det omvendt gælde, at den største grad af regeloverholdelse alt andet lige opnås, når virksomhederne oplyses om deres faktiske kontrolsandsynlighed. Dette teoretiske resultat gælder under ret generelle forudsætninger.

For de danske kontrolområder, vi har undersøgt (miljøtilsyn, fødevarerkontrol og kontrol med betaling af punktafgifter), har vi ikke haft adgang til data, der muliggør en egentlig undersøgelse af, om sammenhæng mellem kontrolsandsynligheden og overholdelsesandelen er konveks eller konkav. Blandt andet har det kun været muligt at indhente aggregerede oplysninger om kontrolsandsynligheder og overholdelsesande-

---

5. Myndigheden kan for eksempel inddele en hovedgruppe af forurenende virksomheder i undergrupper efter, hvor sårbar recipienten vurderes at være, og allokere flest ressourcer til kontrol af udledninger til de mest sårbare recipienter.

6. Flere steder i litteraturen (se bl.a. Polinsky & Shavell 2000; Reinganum & Wilde 1988) nævnes det, at kontrolmyndigheder forventer en større grad af regeloverholdelse, hvis de hemmeligholder kontrolsandsynligheder for virksomhederne. Denne forventning synes også at være gældende inden for miljø-, skatte- og fødevarerkontrollen i Danmark, hvilket dels fremgår af rapporter fra de relevante myndigheder, dels oplyses ved henvendelse til disse.

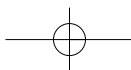
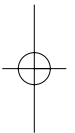
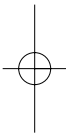


le. Hvis sammenhængen for en eller flere undergrupper af regulerede virksomheder inden for disse områder faktisk er konkav, vil regeloverholdelsen kunne øges, hvis man i stedet for den nuværende »hemmeligholdelsesstrategi« oplyste virksomhederne om deres faktiske kontrolsandsynlighed. Der kan derfor være grund til at foretage egentlige undersøgelser heraf på alle tre områder. Strukturen i artiklen er som følger. I afsnit 2 gives der nogle eksempler på dansk kontrolpraksis inden for udvalgte områder. Dernæst opstilles i afsnit 3 en simpel model, der belyser konsekvenserne af forskellige oplysningsstrategier. De praktiske implikationer af vores resultater diskuteres i afsnit 4. Der konkluderes og perspektiveres i afsnit 5.

## 2. Nogle eksempler fra dansk kontrol- og håndhævelsespraksis

Vi har undersøgt, hvilken kontrol- og håndhævelsespraksis der findes inden for tre af de tunge reguleringsområder i forhold til danske virksomheder (told og skat, miljøtilsynet og fødevarerkontrollen). Oversigten og gennemgangen er ikke komplet eller for så vidt repræsentativ for den danske regulering i forhold til virksomheder, men den giver dog et vist indtryk af ressourceforbrug og spredningen i regeloverholdelse og håndhævelsesstrategier.

På *miljøområdet* skal forurenende virksomheder bl.a. overholde en række miljøkrav, der er specificeret i miljølovgivningen eller i deres miljøgodkendelse. Det er miljøministeren, der udfærdiger en liste over de særligt forurenende virksomheder, anlæg og indretninger, der er omfattet af godkendelsespligten. Disse virksomheder omtales som listevirksomheder. Listen specificerer også, om det er amtet eller kommunen, der er godkendelses- og tilsynsmyndighed. Derudover fører kommunerne tilsyn med de virksomheder, der er omfattet af branchebekendtgørelserne, f.eks. pelsdyrfarme og autoværksteder. Hvorledes kontrollen skal tilrettelægges, fremgår ikke klart af miljølovgivningen, men af forskellige vejledninger fremgår det dog, at tilsynsmyndigheden har pligt til at føre et aktivt og opsøgende tilsyn, og at tilsynet skal ske efter behov, dvs. at de forhåndenværende ressourcer skal anvendes til de opgaver, der har størst miljømæssig betydning. I en aftale mellem KL og Miljø- og Energiministeriet i 1996 anbefales det, at kommunerne mindst fører tilsyn med bestemte typer af virksomheder hhv. hvert andet eller tredje år, jf. Miljøstyrelsen (2003). Der er således fastsat retningslinjer for den gennemsnitlige kontrol af virksomheder inden for forskellige overordnede grupper. Inden for disse overordnede grupper kan (og skal) tilsynsmyndigheden differentiere kontrolindsatsen bl.a. på baggrund af virksomhedernes miljøbelastning og potentielle skader ved regelovertrædelser, hvorved den enkelte virksomheds kontrolsandsynlighed kan afvige markant fra gennemsnittet for den overordnede gruppe, den tilhører. Miljømyndigheden oplyser ikke de enkelte virksomheder om deres kontrolsandsynlighed og har som politik ikke at oplyse herom, hvis en virksomhed skulle forespørge.



Tabel 1. Fakta om udvalgte kontrolområder i 2001.

	Amters og kommuners miljøtilsyn <sup>1</sup>	Fødevarerkontrollen	Told & Skats restanceinddrivelse <sup>2</sup>
Udgifter til kontrol	372 mio. kr. <sup>3</sup>	316 mio. kr.	357 mio. kr. <sup>4</sup>
Regler/anbefalinger mht. kontrolsandsynlighed i love, bekendtgørelser mv.	Anbefales mindst ét tilsyn hvert andet eller tredje år for bestemte typer af virksomheder	Anbefales mindst ét tilsyn pr. år for bestemte typer af virksomheder	Ikke tilgængelig
Kan myndigheden differentiere kontrolindsatsen i forhold til disse regler/anbefalinger?	Ja, differentiering bl.a. i forhold til potentiel miljøskade	Ja, differentiering bl.a. i forhold til skadesomfanget ved regelovertrædelser	Ja, differentiering bl.a. i forhold til risiko for provenutab
Oplyses den enkelte virksomhed om dens kontrolsandsynlighed? <sup>5</sup>	Nej	Nej	Nej
Indikator for kontrolsandsynlighed	48%	84% <sup>6</sup>	8% <sup>7</sup>
Indikator for overholdelsesandel	55%	95%	51% <sup>8</sup>

Kilde: Miljøstyrelsen (2003), Fødevaredirektoratet (2002), Rigsrevisionen (2002) & (2003).

1. Miljøtilsynet omfatter her kommuners og amters samlede tilsyn med liste-, anmelde- og branchevirksomheder.
2. Denne omfatter Told & Skats kontrol med virksomhedernes betaling af moms, punktafgifter, lønsumsafgifter og arbejdsgivernes indberetning af lønmodtagernes indkomst til skattevæsenet, jf. Told & Skat (2004).
3. Ressourceforbruget opgøres i 2001 til i alt 806 årsværk. Da et årsværk anvendt til løsningen af opgaver i de kommunale miljøtilsyn i 2001 blev opgjort til at koste 461.000 kr. (jf. Miljøstyrelsen 2003), opgøres det samlede ressourceforbrug her til 372 mio.kr. Her skal der tages det forbehold, at der kan være forskel på udgifterne til et årsværk ansat i hhv. kommunale og amtskommunale miljøforvaltninger.
4. Told & Skat opgør kun ressourceanvendelsen i årsværk. Ressourceforbruget er omregnet til kroner ved brug af de gennemsnitlige omkostninger ved et årsværk anvendt til miljøtilsynet i kommunerne.
5. Myndighedernes oplysningsstrategi, samt baggrunden for denne, fremgår i nogle tilfælde af deres rapporter. Hvor dette ikke har været tilfældet, er oplysningerne indhentet ved at spørge myndighederne direkte.
6. Denne indikator udtrykker den gennemsnitlige andel af kontrollerede virksomheder for samtlige fødevareraktiviteter. Kontrolsandsynligheden kan variere meget mellem forskellige typer af fødevareraktiviteter.
7. De tal fra Told & Skat, der vedrører kontrolsandsynlighed og overholdelsesandel, vedrører kun kontrollen med virksomhedernes betaling af punktafgifter, idet oplysninger for hele Told & Skats område ikke har været tilgængelige. Det skal desuden bemærkes, at tallene her er fra 2000.
8. Told & Skat opgør ikke antallet af sanktioner, der blev anvendt i forbindelse med kontrollen med virksomheder, men kun antallet af efterreguleringer. Dette er derfor nok et underkantsskøn for den faktiske overholdelsesandel.



Danske virksomheder er desuden pålagt en lang række *skatter og afgifter*. Derfor foretager skattemyndighederne (dvs. statens told- og skatteregioner) kontrol med, om skatter og afgifter afregnes korrekt til statskassen, og med, at grundlaget for eventuelle skatte- eller afgiftsfritagelser er opgjort korrekt. Lovgivning og vejledninger fastsætter ikke konkrete kontrolsandsynligheder, men præciserer, at kontrollen skal tilrettelægges, så den tager hensyn til risikoen for provenutab og retssikkerhed, jf. Rigsrevisionen (2003). I praksis betyder dette, at virksomhederne inddeles i grupper efter, om der er en stor eller lille risiko for skatteunddragelse og efter unddragelsens forventede størrelse. Told & Skat oplyser ikke virksomhederne om, hvilken gruppe de befinder sig i, eller med hvilken sandsynlighed de udsættes for kontrol, og oplyser heller ikke herom, hvis en virksomhed skulle forespørge, jf. Told & Skat (2004).

På *fødevarerområdet* kontrollerer 11 statslige fødevareregioner, om fødevarer virksomheder overholder reglerne på området. Disse regler skal sikre, at behandling og salg af fødevarer foregår på en hygiejnisk forsvarlig måde og gennemføres således, at fødevarerne ikke udsættes for forurening eller anden skadelig påvirkning eller påvirker andre fødevarer i skadelig retning. Fødevaredirektoratet fastsætter vejledende tilsynsfrekvenser for forskellige overordnede grupper af fødevarer virksomheder. Inden for disse overordnede grupper skal tilsynsmyndigheden differentiere kontrolindsatsen efter en behovsvurdering (en vurdering af skadesomfanget ved regelovertrædelser), bl.a. anbefales en relativt højere kontrolindsats på fødevarer producerende virksomheder. Fødevarer kontrollen oplyser ikke de enkelte virksomheder om deres kontrolsandsynlighed og har som politik ikke at oplyse herom, hvis en virksomhed skulle forespørge.

Ved fortolkning af tabellens første række skal det nævnes, at ressourceforbruget for nogle områder er omregnet fra årsværk til kroner for at gøre en sammenligning mulig, hvorfor opgørelsen er behæftet med en vis usikkerhed. Imidlertid er det klart, at ressourceforbruget i forbindelse med tilsyn med virksomhederne inden for miljø-, fødevarer- og skatteområdet er betydeligt. Et groft skøn over de anvendte kontrolressourcer på de tre områder i 2001 er godt 1 mia. kr.<sup>7</sup>

Det er normen, at virksomhederne opdeles i forskellige overordnede grupper, hvor sandsynligheden for kontrol afhænger af, hvilken gruppe virksomheden er placeret i (tabellens 2. række). Den enkelte virksomhed kan i nogle tilfælde udlede den gennemsnitlige tilsynsfrekvens for den overordnede gruppe, den tilhører, ud fra love, bekendtgørelser og vejledninger. Imidlertid foretager tilsynsmyndigheden i alle disse tilfælde en differentiering af kontrolindsatsen inden for de overordnede grupper (tabellens 3.

---

7. De opgjorte omkostninger afspejler ressourceforbruget til kontrol. Finansieringen varierer fra område til område. F.eks. betaler virksomheder en del af de udgifter, der er knyttet til godkendelse og tilsyn på miljøområdet. Også fødevarer virksomhederne betaler et gebyr til fødevareregionerne, der skal dække disses udgifter til tilsyn og kontrol.

række). Da det samtidig er fast praksis ikke at oplyse virksomhederne om deres kontrolsandsynlighed (4. række), må vi konkludere, at virksomhederne på alle de udvalgte områder oplever en vis usikkerhed med hensyn til den kontrolsandsynlighed, de udsættes for. En usikkerhed, der formodentlig vil kunne mindskes betydeligt, hvis kontrolmyndighederne oplyste herom.

I tabellens 5. række opgøres en indikator for den gennemsnitlige kontrolsandsynlighed for hvert område. Kontrolgruppen omfatter alle de virksomheder, der er forpligtet til at overholde det betragtede regelsæt og derfor kan risikere at blive kontrolleret i det pågældende år. Som indikator benyttes antallet af kontrollerede virksomheder i forhold til det samlede antal virksomheder i kontrolgruppen. Indikatoren skal tages med forbehold, dels på grund af datausikkerhed, men også mere grundlæggende på grund af usikkerhed om den relevante periodelængde. Hvis Told & Skat ved et årligt besøg kan afsløre alle overtrædelser, som en virksomhed har lavet i løbet af det aktuelle regnskabsår, vil et år være en rimelig periode at opgøre antallet af kontrolbesøg over. Hvis hver virksomhed besøges en gang om året, vil dette svare til en fuldstændig eller 100% kontrol. I modsætning hertil kan man forestille sig, at et årligt besøg af fødevarekontrollen ikke opleves som en intensiv 100% kontrol, hvis de regler, der kontrolleres, primært handler om den daglige hygiejne. Her kunne det være, at man skal op på betydeligt mere end et årligt besøg, før man kan sige, at kontrollen er 100%. Vi har imidlertid valgt at benytte året som opgørelsesperiode vel vidende, at dette kan betyde en vis skævvridning. Man kan formode, at vi dermed overvurderer fødevarekontrolsandsynligheden. Fejlen ved opgørelsen for de øvrige områder er nok mindre, men vi kan ikke udelukke, at kontrolsandsynligheden også overvurderes her.

I tabellens sidste række anvendes antallet af myndighedsreaktioner i forhold til antallet af tilsyn som indikator for virksomhedernes overtrædelsesomfang inden for det pågældende område. Overholdelsesandelen defineres som antallet af tilsyn, som ikke har medført en myndighedsreaktion. Denne indikator skal også tages med forbehold, idet det ikke kan udelukkes, at praksis for myndighedsreaktioner inden for områderne kan være forskellig.<sup>8</sup> Imidlertid synes der (måske bortset fra fødevareområdet) at være et stort potentiale for at forbedre regeloverholdelsen.

For de tre områder, hvor indikatorer for såvel kontrolsandsynlighed som overholdelsesandel haves, synes disse at følges ad, som man umiddelbart skulle forvente. Overholdelsesandelen på miljø- og skatteområdet er godt 50%, mens næsten alle virksomheder overholder regelsættet på fødevareområdet samtidig med, at en stor del af virksomhederne kontrolleres på fødevareområdet (84%), mens hhv. 48% og 8% af virksomhederne kontrolleres på miljø- og skatteområdet. Specielt synes den høje kon-

---

8. Det kan f.eks. være praksis at give uofficielle advarsler frem for myndighedsreaktioner ved mindre forseelser mv., ligesom denne praksis kan variere fra område til område.

trolsandsynlighed at hænge godt sammen med den høje overholdelsesandel på fødevarerområdet. Dette kunne tyde på, at den formodede fejl ved opgørelsen af kontrolsandsynligheden for fødevarerområdet ikke er så alvorlig. Omvendt kan der være andre forklaringer på den høje overholdelsesandel end en høj kontrolsandsynlighed, f.eks. at den oplevede straf ved overtrædelser i nogle tilfælde er ganske stor.<sup>9</sup>

### 3. Konsekvenser af valg af forskellige oplysningsstrategier

I dette afsnit opstilles en formel model, som skal belyse konsekvenserne af myndighedens valg af forskellige oplysningsstrategier.

#### *Reguleringssituationen*

Vi betragter en myndighed, der skal kontrollere en given undergruppe af virksomheder med hensyn til overholdelse af et bestemt regelsæt. Vi antager, at denne undergruppe består af  $N$  virksomheder. Endvidere antager vi, at myndigheden, ud fra overordnede kriterier, har allokeret et givet budget til at kontrollere de  $N$  virksomheder. Dette svarer til, at myndigheden har allokeret et bestemt antal kontrolbesøg ( $V$ ) til den pågældende undergruppe, hvilket svarer til en kontrolsandsynlighed på  $\bar{S} = V/N$  for den pågældende undergruppe af virksomheder. Den forventede bøde/straf ( $C$ ) er eksogen bestemt.

Kontrolmyndighedens problem er at vælge den oplysningsstrategi, der maksimerer andelen af virksomheder, der overholder reglerne.<sup>10</sup> Vi antager, at myndigheden kan vælge mellem to oplysningsstrategier:

- (U) ingen informeres om kontrolsandsynligheden, eller
- (M) alle virksomheder informeres om kontrolsandsynligheden  $\bar{S}$ .

I almindelighed danner virksomhederne forventninger til kontrolsandsynligheden, og vi lader  $S_i^{for}$  angive den kontrolsandsynlighed, som virksomhed  $i$  forventer, idet vi her antager, at virksomhederne er risikoneutrale. Den forventede kontrolsandsynlighed ( $S_i^{for}$ ) antages at afhænge af såvel den faktiske kontrolsandsynlighed ( $\bar{S}$ ) som myndighedens oplysninger herom.

9. Det har således været praksis, at de overtrædelser, fødevarerkontrollen finder, bliver offentliggjort. Denne »sanktion« indgår end ikke i opgørelsen af myndighedsreaktioner, men har sandsynligvis en stor indflydelse på virksomhedernes overholdelse af reguleringen på dette område. Selv om denne »straf« ud fra almindelige retsprincipper opfattes som rimelig i forhold til forseelsen, kan konsekvenserne af at offentliggøre et dårligt kontrolresultat være ganske voldsomme for f.eks. en restaurant eller slagterforretning.

10. Andre dele af litteraturen antager, at myndigheden i stedet maksimerer velfærden. I dette tilfælde vil det ikke nødvendigvis være optimalt at maksimere graden af overholdelse, jf. f.eks. Garoupa (1997). En antagelse om, at myndigheden i stedet maksimerer overholdelsesgraden, er dog ikke urealistisk, når man kigger på en situation, hvor ansvaret for kontrollen er uddelegeret til tredjepart, f.eks. en regional myndighed, som inden for fastslagne rammer skal tilrettelægge kontrollen.

Ud over, at virksomhederne kan have forskellige forventninger til kontrolsandsynligheden, er de forskellige med hensyn til den gevinst, de opnår ved at overtræde det regelsæt, som kontrollen vedrører. Vi antager, at virksomhed  $i$ 's gevinst ved at overtræde et regelsæt er  $G_i$ . Herved kan betingelsen for, at virksomhed  $i$  overholder regelsættet skrives som:

$$G_i \leq S_i^{for} C \quad (1)$$

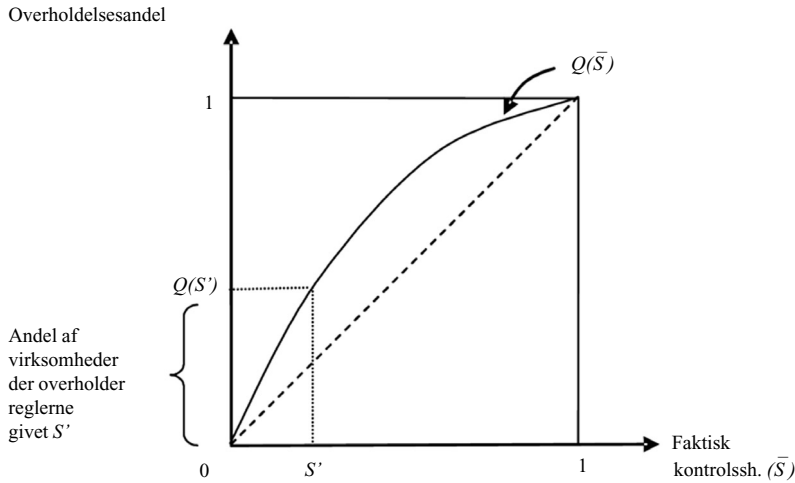
Dette betyder, at når sanktionen er eksogent bestemt, og den forventede kontrolsandsynlighed kan variere mellem virksomhederne, vil den forventede straf for ikke at overholde reglerne variere mellem virksomhederne inden for den pågældende gruppe. Dette på trods af, at den faktiske kontrolsandsynlighed er givet og ens for alle virksomheder i undergruppen.

Når lighedstegnet i ligning (1) gælder, defineres den mindste forventede kontrolsandsynlighed, der medfører, at virksomhed  $i$  overholder regelsættet. Lad  $S_i^{min}$  betegne den således definerede *mindste kontrolsandsynlighed*.  $S_i^{min}$  er en teknisk egenskab givet ved virksomhedens produktions- og afsætningsvilkår (og holdninger til lovovertrædelser mv.) og må derfor formodes at være *uafhængig* af såvel den faktiske kontrolsandsynlighed ( $\bar{S}$ ), myndighedens oplysningsstrategi, som virksomhedens forventning til kontrolsandsynligheden ( $S_i^{for}$ ).<sup>11</sup>

Hver virksomhed er således karakteriseret ved værdierne  $S_i^{min}$  og  $S_i^{for}$ , hvorfor undergruppen af virksomheder kan beskrives ved den tilhørende todimensionale sandsynlighedsfordeling over  $S^{min}$  og  $S^{for}$ , som vi betegner  $H(S^{for}, S^{min})$ . Vi betegner de tilhørende marginale fordelinger  $F(S^{for})$  og  $Q(S^{min})$  samt de modsvarende tæthedsfunktioner  $h(S^{for}, S^{min})$ ,  $f(S^{for})$  og  $q(S^{min})$ . Ligeledes lader vi  $F(S^{for} | S^{min})$  betegne den betingede sandsynlighedsfunktion svarende til  $F(S^{for})$ .

Den marginale  $Q(S^{min})$ -fordeling angiver den andel af virksomhederne, hvis mindste kontrolsandsynlighed er mindre end  $S^{min}$ . Hvad fordelingen mere intuitivt udtrykker ses nemmest, hvis vi for et øjeblik antager, at den faktiske kontrolsandsynlighed  $\bar{S}$  er kendt af alle virksomheder (således at  $S_i^{for} = \bar{S}$  for alle  $i$ ). I denne situation angiver  $Q(S^{min})$ -fordelingen andelen af virksomheder, hvor  $S_i^{min} \leq S_i^{for} = \bar{S}$ , hvilket netop er den andel af virksomhederne, der overholder reglerne ved den gældende kontrolsandsynlighed. Figur 1 illustrerer et muligt forløb for denne fordeling.

11. Bemærk, at i forhold hertil gøres antagelserne om, at virksomhederne er risikoneutrale og har punktformede forventninger uden tab af generalitet. Hvis virksomhederne var risikoaverse og dannede en sandsynlighedsfordeling over mulige kontrolsandsynligheder skulle (1) blot skrives som:  $E[S_i^{for} U(G_i - C) + (1 - S_i^{for}) U(G_i)] \leq 0 \Leftrightarrow E[S_i^{for}] (U(G_i - C) - U(G_i)) + U(G_i) \leq 0$  hvorved  $S_i^{min} = U(G_i) / (U(G_i) - U(G_i - C))$ .



Figur 1. Eksempel på sammenhængen mellem kontrolsandsynlighed og overholdelsesandel.

Langs  $x$ -aksen angives mulige værdier for den faktiske kontrolsandsynlighed (der kan spænde fra 0 til 1), mens  $y$ -aksen angiver den andel af virksomheder i undergruppen, som vil overholde regelsættet ved den angivne kontrolsandsynlighed. Hvis den faktiske kontrolsandsynlighed f.eks. er  $S'$ , vil der være en andel  $Q(S')$  af virksomhederne, der overholder reglerne.<sup>12</sup>

Den anden fordelingsdimension ( $S^{for}$ ) udtrykker virksomhedernes forventede kontrolsandsynlighed, som kan være forskellig fra den faktiske kontrolsandsynlighed,  $\bar{S}$ , når der er usikkerhed med hensyn til kontrolsandsynligheden. Opgørelsen af den andel af virksomhederne, der overholder reglerne ved den gældende kontrolsandsynlighed (som vi betegner  $\bar{A}$ ), bliver mere kompliceret, når der er usikkerhed om kontrolsandsynligheden:

$$\bar{A} = \int_0^1 q(S)(1 - F(S | S))dS \quad (2)$$

12. Ved tegning af figuren har vi antaget, at  $Q(0) = 0$  og  $Q(1) = 1$ . Dermed antager vi, at ingen virksomheder vil overholde reglerne, hvis der ingen kontrol er, og at strafferammen er udformet således, at alle virksomheder overholder reglerne, når kontrolsandsynligheden er 1. Dette er ikke nødvendigvis tilfældet. Man kan godt forestille sig, at nogle virksomheder ikke har gevinst ved at overtræde og derfor overholder reglerne, selv om der ikke er kontrol, ligesom man kan forestille sig, at nogle virksomheder har så store gevinster ved at overtræde, at de vil bryde reglerne, selv om sandsynligheden for opdagelse er 1. Det eneste, vi kan være sikre på, er, at  $Q$ -funktionen er monotont stigende.

Ideen er, at man grupperer virksomhederne efter deres  $S^{min}$ -værdi, opgør den andel af hver gruppe, der forventer en kontrolsandsynlighed større end  $S^{min}$ , og summerer over grupperne. Der integreres over alle mulige værdier af  $S^{min}$ , og  $q(S)$  angiver andelen af virksomheder, der har den aktuelle  $S^{min}$ -værdi. Den betingede fordeling  $F(S^{for} | S)$  angiver fordelingen af forventede kontrolsandsynligheder for gruppen af virksomheder med den aktuelle  $S^{min}$ -værdi. Vi er interesseret i den andel heraf, der har en forventet kontrolsandsynlighed ( $S^{for}$ ), der er større end  $S$  (altså  $(1 - F(S | S))$ ), fordi netop disse vil overholde reglerne.

#### *Det generelle tilfælde*

Da virksomheder kun kan forvente kontrolsandsynligheder mellem 0 og 1, er  $F(1) = 1$  pr. definition. Vi vil endvidere antage, at der foretages så megen kontrol, at ingen virksomheder forventer en kontrolsandsynlighed på 0, og dermed antager vi, at  $F(0) = 0$  uanset myndighedens oplysningsstrategi. Antagelsen betyder, at vores resultater ikke gælder i situationer, hvor kontrollen er så lille og usynlig, at en ikke ubetydelig del af virksomhederne forventer en kontrolsandsynlighed på 0.

Endvidere vil vi antage, at middelværdien af virksomhedernes forventede kontrolsandsynligheder er lig den faktiske kontrolsandsynlighed uanset myndighedens oplysningsstrategi (altså at  $E[S^{for}] = \bar{S}$ ). Dette er en standardantagelse i litteraturen (se bl.a. Bebchuk & Kaplow 1992; Garoupa 1998 og Garoupa 1999), der bunder i, at virksomhederne oplever og dermed observerer den faktiske kontrolindsats. Selv om virksomhederne i afgrænsede perioder oplever forskellige kontrolsandsynligheder, vil virksomhederne i gennemsnit opleve den faktiske kontrolsandsynlighed  $\bar{S}$ . Det virker derfor ikke plausibelt, at virksomhedernes forventninger på langt sigt skulle afvige systematisk fra den faktiske kontrolsandsynlighed.<sup>13</sup>

Endelig vil vi i det følgende benytte den afgørende forudsætning, at fordelingerne af  $S^{min}$  og  $S^{for}$  er uafhængige, således at:

$$H(S^{for}, S^{min}) = F(S^{for}) Q(S^{min}) \quad (3)$$

og der for de tilhørende tæthedsfunktioner gælder:

$$h(S^{for}, S^{min}) = f(S^{for}) q(S^{min}) \quad (4)$$

---

13. Betragtes flere perioder, hvor kontrolmyndigheden løbende ændrer på den faktiske kontrolsandsynlighed, kan man forestille sig, at virksomhederne kun langsomt lærer de nye kontrolsandsynligheder og dermed, at de på kort sigt vil have systematisk forkerte forventninger til kontrolsandsynligheden. På langt sigt vil sådanne variationer dog ikke give anledning til systematisk bias.

Uafhængighedsantagelsen siger, at der er uafhængighed mellem virksomhedernes forventede kontrolsandsynlighed  $S_i^{for}$  og den (mindste) kontrolsandsynlighed, der sikrer, at virksomhederne vil overholde reguleringen,  $S_i^{min}$ . Sidstnævnte bestemmes af den eksogent givne straf, virksomheden får, hvis kontrollen afslører, at regelsættene er overtrådt, samt af virksomhedernes forventede gevinst ved at overtræde reguleringen. Denne gevinst er bestemt ud fra produktions- og afsætningsvilkår i virksomhederne. Uafhængighedsantagelsen siger dermed, at der er uafhængighed mellem virksomhedsspecifikke tekniske forhold og virksomhedernes forventede kontrolsandsynlighed. Selv om antagelsen er standard i litteraturen (se f.eks. Ben-Shahar 1997 og Bebchuk & Kaplow 1992), kan man dog ikke helt afvise, at virksomhedens incitament til at opsøge/indhente information om myndighedens kontrolstrategi kan afhænge af  $S^{min}$ . Hvis virksomheden tror, at  $\bar{S}$  ligger tæt på  $S^{min}$ , kan den mene, at det er vigtigere at få en præcis ide om størrelsen af  $\bar{S}$  for dermed at kunne afgøre, om reglerne skal overholdes. Hvis virksomheden omvendt tror, at der under alle omstændigheder er en stor forskel, og den derfor ikke er i tvivl om, hvorvidt reglerne skal overholdes, er incitamentet til at søge oplysninger om  $\bar{S}$  måske mindre. Dette forhold kunne give anledning til nogen afhængighed mellem  $S^{min}$  og  $\bar{S}$ . Omvendt vil et forkert valg ikke påvirke virksomhedens nytte særlig meget, hvis  $\bar{S}$  ligger tæt på  $S^{min}$ , hvorfor incitamentet til at opsøge information ikke kan være stort i dette tilfælde. Derfor virker det sandsynligt, at en eventuel afhængighed er lille. Alt i alt virker denne centrale antagelse ikke oplagt urimelig. Antagelsen betyder, at  $F(S^{for} | S^{min}) = F(S^{for})$ , hvorved:

$$\bar{A} = \int_0^1 q(S)(1 - F(S | S))dS$$

Ved delvis integration fås, at:

$$\bar{A} = \int_0^1 Q(S)(1 - F(S)) + \int_0^1 Q(S)f(S)dS$$

Idet  $F(1) = 1$  og  $F(0) = 0$  har vi, at:

$$\bar{A} = Q(0) + \int_0^1 Q(S)f(S)dS \quad (5)$$

Myndighederne ønsker at benytte den oplysningsstrategi, der giver den højeste grad af regeloverholdelse (altså sikrer den største  $\bar{A}$ -værdi).

Under Strategi  $U$  får virksomhederne ingen information fra myndigheden om værdien af  $\bar{S}$ . Selv om den tilhørende tæthedsfunktion  $f^u(S^{for})$  (jf. antagelsen ovenfor) har

middelværdien  $\bar{S}$ , kan variansen omkring  $\bar{S}$  være stor. Tøp-tegnet i funktionsudtrykket angiver, at det er den til Strategi  $U$  hørende tæthedsfunktion. Vi benævner den tilhørende overholdelsesgrad

$$\bar{A}^u = Q(0) + \int_0^1 Q(S) f^u(S) dS \quad (6)$$

Under Strategi  $M$  oplyses virksomhederne om den faktiske  $\bar{S}$ -værdi. Myndighedens information til virksomhederne må forventes at påvirke de enkelte virksomheders forventede kontrolsandsynlighed. I det ekstreme tilfælde, hvor alle virksomheder har fuld tillid til myndighedens information, vil alle revidere deres forventning til værdien  $\bar{S}$ , hvorved hele sandsynlighedsmassen i den tilhørende  $f^m(S^{for})$  tæthedsfunktion vil samle sig i punktet  $\bar{S}$ . Hvis virksomhederne har mindre tillid til myndighedens information, vil forventningstilnærmelsen mod  $\bar{S}$  være mindre, men vi antager, at hvis myndighedsinformationen medfører, at en virksomhed reviderer sin forventning, vil denne altid bevæge sig mod  $\bar{S}$ . Dermed vil tæthedsfunktionen  $f^m(S^{for})$  kunne dannes ud fra tæthedsfunktion  $f^u(S^{for})$  alene ved transformationer, der flytter sandsynlighedsmasse mod  $\bar{S}$  (for hver virksomhed, der justerer sin forventning tættere til  $\bar{S}$ , svarer en transformation, der flytter den til virksomheden hørende sandsynlighedsmasse mod  $\bar{S}$ ). Variansen omkring  $\bar{S}$  kan fortsat være stor, men vil være mindre end i situationen uden information. Ud over dette og den fælles middelværdi  $\bar{S}$  pålægger vi ikke funktionerne restriktioner. Tæthedsfunktionerne kan for eksempel være asymmetriske, have flere toppe, o.l.

Vi benævner overholdelsesgraden under Strategi  $M$ :

$$\bar{A}^m = Q(0) + \int_0^1 Q(S) f^m(S) dS \quad (7)$$

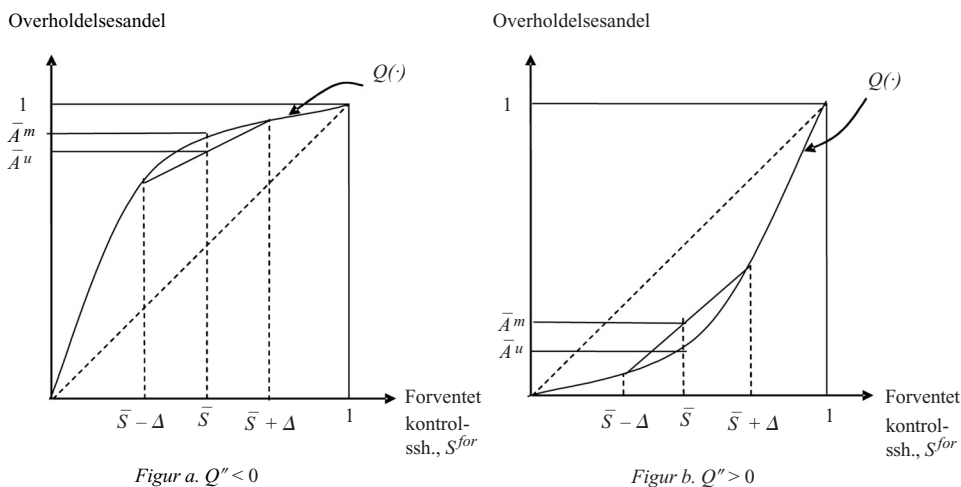
Myndighedens problem er altså at afgøre, om  $\bar{A}^u$  eller  $\bar{A}^m$  er størst i den konkrete situation. I forhold hertil er følgende sætninger nyttige (de er bevist i appendikset):

*Sætning 1:*  $Q'' > 0$  så er  $\bar{A}^u > \bar{A}^m$

*Sætning 2:*  $Q'' < 0$  så er  $\bar{A}^u < \bar{A}^m$

Det fremgår således, at hvis  $Q(\cdot)$  er konveks i det relevante område omkring  $\bar{S}$ , giver hemmeligholdelse af kontrolsandsynligheden den største grad af regeloverholdelse. Hvis  $Q(\cdot)$  omvendt er konkav i det relevante område omkring  $\bar{S}$ , giver information





Figur 2. Implikationer af usikkerhed.

om kontrolsandsynligheden den største grad af regeloverholdelse. Resultatet gælder også, hvis virksomhederne i stedet for punktforventninger til kontrolsandsynligheden er karakteriseret ved en sandsynlighedsfordeling over mulige værdier og risikoaversion.

#### Et illustrativt eksempel

Man kan illustrere konsekvenserne af at vælge forskellige oplysningsstrategier ved hjælp af figur 2. Dette er gjort for såvel en konkav som en konveks sammenhæng mellem den forventede kontrolsandsynlighed og den dertil hørende overholdelsesandel.

Figurerne viser betydningen af en symmetrisk reduktion af virksomhedernes usikkerhed med hensyn til kontrolsandsynligheden i et specialtilfælde. I udgangssituationen (Strategi  $U$ , uden information) forventer halvdelen af virksomhederne en kontrolsandsynlighed på  $\bar{S} - \Delta$ , mens den anden halvdel forventer  $\bar{S} + \Delta$ . Under Strategi  $U$  kan overholdelsesandelen for hver af de to grupper aflæses på  $y$ -aksen, og den gennemsnitlige overholdelse beregnes som et simpelt gennemsnit af de to aflæste størrelser, jf. figurerne. Under Strategi  $M$  (med information) forventer alle virksomheder den faktiske kontrolsandsynlighed  $\bar{S}$ . I dette simple tilfælde er det oplagt, at konkavitet medfører, at  $\bar{A}^u < \bar{A}^m$ , mens konveksitet medfører  $\bar{A}^u > \bar{A}^m$ . Det vil sige, at myndigheden kun bør vælge at hemmeligholde kontrolsandsynligheden, hvis der er en konveks sammenhæng mellem denne og overholdelsesandelen. Selv om antallet af virksomheder, der vil overestimere kontrolsandsynligheden, når denne hemmeligholdes, svarer til antallet, der vil underestimere, er virkningerne på regeloverholdelsen forskellige.

Konveksitet betyder, at andelen af virksomheder, der begynder at overholde reglerne, fordi de overestimerer kontrolsandsynligheden, er større end andelen af virksomheder, der begynder at bryde reglerne, fordi de underestimerer kontrolsandsynligheden. Når  $Q$ -funktionen er konkav, er forholdet omvendt.

#### 4. Praktiske implikationer: nogle indledende overvejelser

Vi har vist, hvordan oplysning om kontrolsandsynligheden påvirker overholdelsen. Dette resultat kan være en del af grundlaget for kontrolmyndighedernes valg af oplysningsstrategi, hvis en myndighed har adgang til de nødvendige data og analysekapacitet. I denne situation kan den til kontrolgruppen hørende  $Q$ -funktion muligvis estimeres så præcist, at dens konveksitet/konkavitet kan fastlægges med rimelig sikkerhed. Dette kræver en disaggregeret dataserie med samhörørende punkter af kontrolsandsynlighed og overholdelsesandele af en kvalitet, der tillader benyttelse af de relevante estimeringsteknikker<sup>14</sup> til estimation af ligning (6):

$$\bar{A}^u = Q(0) + \int_0^1 Q(S) f^u(S) dS$$

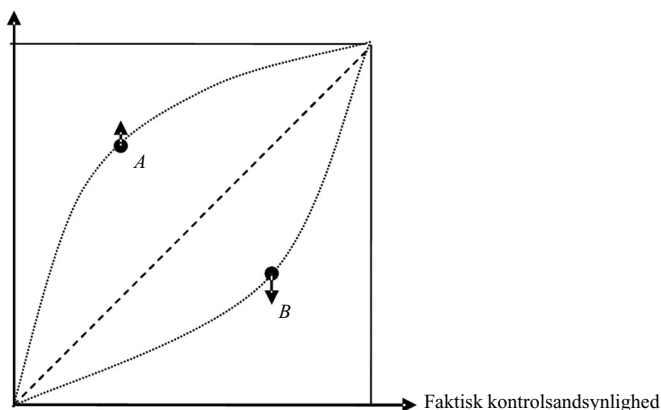
##### *En simpel procedure*

I praksis kan det være vanskeligt for en myndighed at foretage en egentlig estimation af  $Q$ -funktionen på grund af utilstrækkelige data eller analysekapacitet. Det bliver dermed relevant at spørge, om simple analyser med mere mangelfulde data kan give en indikation af, hvilken oplysningsstrategi der bør foretrækkes. Lad os derfor se på en situation, hvor myndigheden kun observerer et datapunkt (samhörørende værdier for kontrolsandsynlighed og estimeret overholdelsesandel) for en bestemt gruppe af kontrollerede virksomheder. Myndigheden antages også at have en rimelig ide om, hvor stor en andel af virksomheder der vil overholde det kontrollerede regelsæt, hvis der ikke gennemføres nogen kontrol overhovedet, samt hvis alle virksomheder kontrolleres. En formodning om, at ingen overholder regelsættet, når der ikke kontrolleres, og alle overholder regelsættet, når alle kontrolleres, kunne for eksempel være rimelig i nogle situationer. Hermed har myndigheden tre datapunkter som udgangspunkt for sit bud på  $Q$ -funktions form omkring det observerede punkt. Gennem disse punkter kan der fastlægges kurver, jf. figur 3.

Man kan naturligvis ikke vide, hvilken form den sande kurve har omkring det observerede punkt, men hvis det observerede punkt (f.eks. punkt  $A$ ) ligger over linjen gennem  $Q(0)$  og  $Q(1)$ , virker det sandsynligt, at  $Q(\cdot)$  er konkav omkring det observerede

14. Bemærk, at estimationen dels skal ske med en så fleksibel funktionel form, at konveksitet/konkavitet kan estimeres, dels skal denne kunne håndtere det forhold, at virksomhedernes skøn over kontrolsandsynligheden ikke observeres direkte.

Overholdelsesandel



Figur 3. Estimering af  $Q(\cdot)$  funktionen.

de datapunkt. Hvis det observerede punkt omvendt ligger under linjen (som punkt  $B$ ), virker det mere sandsynligt, at  $Q(\cdot)$  er konvex omkring det observerede datapunkt.

Som tidligere nævnt observerer vi ikke  $Q(\bar{S})$ , men i stedet

$$\bar{A}^u = Q(0) + \int_0^1 Q(S) f^u(S) dS.$$

Da vi ved at  $Q(\bar{S}) > \bar{A}^u$ , når  $Q(\bar{S})$  er konkav, vil den sande  $Q(\bar{S})$  værdi i dette tilfælde ligge over det observerede punkt og derfor længere inde i det konkave område (indikeret med den opadpegende pil fra punkt  $A$ ). Modsat ved vi, at når  $Q(\bar{S})$  er konvex, vil  $Q(\bar{S}) < \bar{A}^u$ , hvilket betyder, at den sande  $Q(\bar{S})$  værdi her vil ligge længere inde i det konvekse område (hvilket den nedadgående pil ud fra punkt  $B$  illustrerer).

Hvis en myndighed, der skal vælge oplysningsstrategi, ikke har anden information end den ovenfor nævnte, kan denne simple procedure give en indikation af, om konkavitet eller konvexitet er det mest sandsynlige. Man kan givet forestille sig andre procedurer, og man kan givetvis også raffinere ovennævnte procedure (f.eks. hvis myndigheden har adgang til mere information). Den pointe, vi gerne vil understrege, er, at det teoretiske resultat, der er vist ovenfor, kan udnyttes til praktisk politikvejledning også i mere realistiske situationer, hvor den relevante myndighed har mangelfulde data og/eller analysekapacitet.

#### En illustration

Det har ikke været muligt at indhente detaljerede, men kun aggregerede oplysninger om kontrolindsatsen mv. for de tidligere beskrevne kontrolområder. I det følgende

illustreres en praktisk anvendelse af vores resultater på disse aggregerede data. Dette illustrerer, hvordan man *kan* anvende den gennemgåede simple procedure, men analysen kan ikke bruges som grundlag for en konkret politikanbefaling – blandt andet fordi data for hvert område er aggregeret over flere kontrolgrupper.

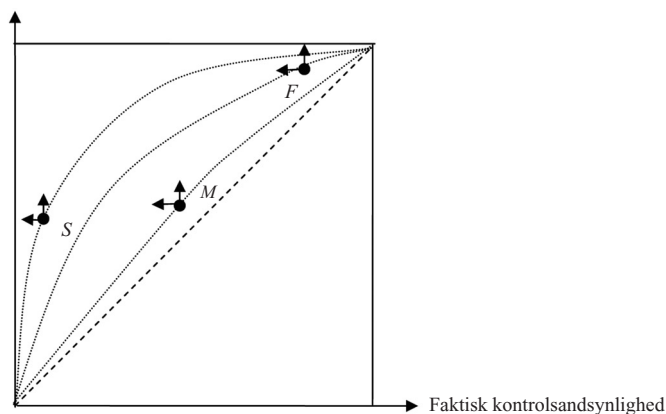
I figur 4 er de tre områder – beskrevet i afsnit 2 – indplaceret med kontrolsandsynlighed ud ad  $x$ -aksen og overholdelsesandel ud ad  $y$ -aksen (svarende til figurerne 1 og 2). I figur 4 betegner punktet "M" miljøtilsynsområdet, "F" fødevarekontrollen, og "S" skatteområdet, idet vi har benyttet de i tabel 1 opgjorte indikatorer for kontrolsandsynlighed og overholdelsesandel samt det forhold, at  $\bar{S} = E[S_i^{for}]$ . Ud over at medtage pile, der illustrerer underestimering af  $Q(\bar{S})$  (jf. punkt A i figur 3), har vi tilføjet vandrette pile ud fra punktet "F", "M" og "S", der indikerer, at  $\bar{S}$  er overvurderet for alle tre områder (jf. ovenfor). Det skal dog bemærkes, at dette i højere grad er gældende for fødevareområdet end for skatteområdet. De stiplede kurver, der er tegnet gennem hvert af punkterne, illustrerer en mulig  $Q(\cdot)$ -funktion for hvert af de tre områder. For et givet område er den tilhørende  $Q(\cdot)$ -kurve tegnet, så den går gennem det observerede punkt ( $M$ ,  $F$  eller  $S$ ) og punkterne  $Q(0) = 0$  (svarende til, at ingen virksomheder vil overholde reglerne, hvis der overhovedet ikke er kontrol) og  $Q(1) = 1$  (svarende til, at alle virksomheder vil overholde reglerne, hvis der er 100% kontrol). Fastlæggelsen af randpunkterne må bero på myndighedens konkrete vurdering af forholdene for den undergruppe af regulerede virksomheder, der undersøges. I nogle situationer kan det f.eks. være mere realistisk at forvente, at en eller anden positiv andel af virksomhederne vil overholde reglerne selv uden kontrol.

Hvis man for hvert af de illustrerede områder antager, at disse tre punkter (de to yderpunkter og det observerede punkt) ligger på den tilhørende  $Q(\cdot)$ -kurve, illustrerer figuren, at man for alle observerede punkter i den øvre trekant umiddelbart vil forvente, at den tilhørende  $Q(\cdot)$ -kurve er konkav omkring det observerede punkt. Det er naturligvis ikke givet, at det er sådan.  $Q(\cdot)$ -kurverne kan have vendetangenter, hvorfor lokal konvexitet omkring et punkt i øvre trekant ikke kan udelukkes. Hvis man imidlertid kun kender de tre nævnte kurvepunkter, virker konkaviteten umiddelbart som det bedste bud, man kan give på kurvens form omkring det observerede punkt.

Hvis de observerede punkter i figur 4 havde afspejlet forholdene i konkrete undergrupper af virksomheder med ens kontrolsandsynlighed, ville dette indikere, at det var mest sandsynligt, at alle tre kurver var konkave omkring det observerede punkt. I givet fald ville regeloverholdelsen kunne øges, hvis man i stedet for den nuværende »hemmeligholdelsesstrategi« oplyste virksomhederne om deres faktiske kontrolsandsynlighed.

Da datapunkterne i alle tre tilfælde imidlertid afspejler aggregerede tal, kan analysen ikke bruges som grundlag for en konkret politikanbefaling. Der synes imidlertid at,

Overholdelsesandel



Figur 4. Praktiske implikationer.

være grund til at foretage egentlige undersøgelser på alle tre områder baseret på de mere detaljerede oplysninger om kontrolsandsynligheder og overholdelsesandele, som de pågældende myndigheder må formodes at have adgang til.

### 5. Konklusion og perspektivering

Kontrolmyndigheder har som oftest begrænsede ressourcer og udefra givne sanktionsmuligheder, hvilket betyder, at de i praksis langt fra altid kan sikre 100% overholdelse af det regelsæt, de kontrollerer. Det er derfor hensigtsmæssigt at differentiere kontrollen efter de incitament, virksomheder har, og de skader, eventuelle overtrædelser kan forvolde, for derved at øge overholdelsesandelen og minimere konsekvenserne af overtrædelser. Dette anbefales i størstedelen af litteraturen på området og er i praksis det, kontrolmyndighederne faktisk gør inden for en række områder i Danmark (om det så gøres helt optimalt, har vi ikke undersøgt).

Formålet med denne artikel har alene været at foretage en sammenligning af konsekvenserne af at hemmeligholde omfanget af kontrolindsatsen med en strategi, hvor kontrolomfanget oplyses over for de virksomheder, der kontrolleres. Det viser sig, at man generelt kan afgrænse, hvornår det er en fordel at oplyse om kontrolsandsynligheden, og hvornår det er en fordel at holde denne hemmelig. Mere præcist viser vi, at effekten afhænger af, hvilken sammenhæng der er mellem kontrolsandsynligheden og overholdelsesandelen. Vores resultater peger på, at hvis denne sammenhæng er konveks, vil usikkerhed alt andet lige medføre, at flere virksomheder vælger at overholde reglerne. Hvis denne sammenhæng derimod er konkav, vil det omvendt gælde, at den

største grad af regeloverholdelse alt andet lige opnås, når virksomhederne oplyses om deres faktiske kontrolsandsynlighed. Endvidere synes det muligt empirisk at afgøre, hvilken oplysningsstrategi der er den foretrukne for en given kontrolsituation ved hjælp af data, kontrolmyndighederne må formodes at have adgang til.

Vores teoretiske hovedresultat gælder ret generelt – bl.a. også, hvis nogle virksomheder er risikoaverse. Det væsentligste forbehold synes at være knyttet til vores forudsætning om fordelingsuafhængighed mellem forventet og mindste kontrolsandsynlighed. Resultatet forudsætter også, at gruppen af virksomheder ikke i gennemsnit og på langt sigt systematisk over- eller undervurderer kontrolsandsynligheden, ligesom det antages, at bevidst vildledning ikke er en farbar strategi for den regulerende myndighed. Om end man nok kan finde reguleringssituationer, hvor disse antagelser ikke holder, forekommer de dog i almindelighed at være rimelige. Alt i alt virker det centrale, teoretiske resultat rimeligt robust.

For de danske kontrolområder, vi har undersøgt (miljøtilsyn, fødevarerkontrol og kontrol med betaling af punktafgifter), har vi ikke haft adgang til data, der muliggør en egentlig undersøgelse af, om sammenhæng mellem kontrolsandsynligheden og overholdelsesandelen er konveks eller konkav. Blandt andet har det kun været muligt at indhente aggregerede oplysninger om kontrolsandsynligheder og overholdelsesandele. Imidlertid illustreres en simpel metode, hvormed en myndighed, der har adgang til disaggregerede data, vil kunne afgøre, om det er mere sandsynligt, at den analyserede gruppe af virksomheder er karakteriseret ved konveksitet eller konkavitet.

Det bør dog understreges, at selv, hvis det viser sig, at der er reguleringsmæssige fordele knyttet til at oplyse en gruppe af virksomheder om deres kontrolsandsynlighed, kan der være andre forhold, der gør dette vanskeligt eller uhensigtsmæssigt at gennemføre i praksis. For eksempel kan der være juridiske problemer forbundet med, at en myndighed oplyser nogle virksomheder, men ikke andre, om deres kontrolsandsynlighed, eller ved at en myndighed via offentliggjorte kontrolsandsynligheder »erkender«, at virksomheder i forskellige undergrupper forskelsbehandles. De juridiske og administrative konsekvenser af at offentliggøre kontrolsandsynligheder bør derfor også kortlægges, inden en omlægning gennemføres.

### Appendiks

#### Bevis af sætning 1

Betragt mængden af transformationer af sandsynlighedsmasse mod  $\bar{S}$ , der fører  $f^u(S^{for})$  over i  $f^m(S^{for})$ . Denne mængde kan indeholde to typer af transformationer:

- (L) transformationer af sandsynlighedsmasse fra  $S^{for}$ -værdier mindre end  $\bar{S}$  til  $S^{for}$ -værdier tættere på, men stadig mindre end eller lig med  $\bar{S}$ .
- (R) transformationer af sandsynlighedsmasse fra  $S^{for}$ -værdier større end  $\bar{S}$  til  $S^{for}$ -værdier tættere på, men stadig større end eller lig med  $\bar{S}$ .

En vilkårlig transformation af den første type (L) flytter sandsynlighedsmassen  $m_j$  fra  $S^{for}$ -værdien  $\bar{S}_j^- < \bar{S}$  til  $S^{for}$ -værdien  $\bar{S}_j^- + d_j \leq \bar{S}$ , hvor  $d_j > 0$ . En vilkårlig transformation af den anden type (R) flytter sandsynlighedsmassen  $m_i$  fra  $S^{for}$ -værdien  $\bar{S}_i^+ > \bar{S}$  til  $S^{for}$ -værdien  $\bar{S}_i^+ - d_i \geq \bar{S}$ , hvor  $d_i > 0$ . Lad der være  $M$  type L transformationer ( $j = 1 \dots M$ ) og  $N$  type R transformationer ( $i = 1 \dots N$ ).

Den ændring i overholdelsesandelen ( $\Delta \bar{A}$ ), som denne mængde af transformationer medfører, kan da udtrykkes således:

$$\Delta \bar{A} = \sum_{j=1}^M (Q(\bar{S}_j^- + d_j) - Q(\bar{S}_j^-)) m_j + \sum_{i=1}^N (Q(\bar{S}_i^+ - d_i) - Q(\bar{S}_i^+)) m_i$$

Da  $Q' > 0$  vil  $Q'(\bar{S}) d_j > (Q(\bar{S}_j^- + d_j) - Q(\bar{S}_j^-)) > 0$  for  $j = 1 \dots M$  og  $0 > -Q'(\bar{S}) d_i > (Q(\bar{S}_i^+ - d_i) - Q(\bar{S}_i^+))$  for  $i = 1 \dots N$ , hvorved vi har, at:

$$\Delta \bar{A} < \sum_{j=1}^M (Q'(\bar{S}) d_j) m_j - \sum_{i=1}^N (Q'(\bar{S}) d_i) m_i$$

som kan omskrives til:

$$\Delta \bar{A} < Q'(\bar{S}) \left( \sum_{j=1}^M d_j m_j - \sum_{i=1}^N d_i m_i \right)$$

Da tæthedsfunktionerne  $f^u(S^{for})$  og  $f^m(S^{for})$  begge har middelværdien  $\bar{S}$ , må det gælde, at:

$$\sum_{j=1}^M d_j m_j = \sum_{i=1}^N d_i m_i$$

hvorved  $\Delta \bar{A} < 0 \Leftrightarrow \bar{A}^u > \bar{A}^m$ .

Bemærk at beviset gælder for en fordeling der er diskret, hvilket enhver fordeling af en bestemt gruppe af virksomheders forventede kontrolsandsynligheder nødvendigvis må være.

#### *Bevis af sætning 2*

Følger analogt beviset for sætning 1.

#### *Litteratur*

- Bebchuk, L.A. og Kaplow L. 1992. Optimal Sanctions when Individuals are Imperfectly Informed about the Probability of Apprehension, *Journal of Legal Studies*, vol. 21, s. 365-70.
- Becker, G. 1968. Crime and Punishment: An Economic Approach, *Journal of Political Economic Review*, vol. 6, s. 169-217.
- Ben-Shahar, O. 1997. Playing Without a Rulebook: Optimal Enforcement When Individuals Learn the Penalty Only by Committing the Crime, *International Review of Law and Economics*, vol. 17, s. 409-21.
- Fødevarerdirektoratet. 2002. *Virksomhedsregnskab for Fødevarerdirektoratet 2001*, Ministeriet for Fødevarer, Landbrug og Fiskeri.
- Garoupa, N. 1997. The Theory of Optimal Law Enforcement, *Journal of Economic Surveys*, vol. 11, nr. 3, s. 267-95.
- Garoupa, N. 1998. Optimal Law Enforcement and Imperfect Information when Wealth Varies among Individuals, *Economica*, vol. 65, s. 479-90.
- Garoupa, N. 1999. Optimal Law Enforcement with Dissemination of Information, *European Journal of Law and Economics*, vol. 7, s. 183-96.
- Greenberg, J. 1984. Avoiding Tax Avoidance: A (Repeated) Game-Theoretic Approach, *Journal of Economic Theory*, vol. 32, s. 1-13.
- Lando, H. og Shavell, S. 2004. The Advantage of Focusing Law Enforcement Effort, *International Review of Law and Economics*, vol. 24, s. 209-218.
- Miljøstyrelsen. 2003. *Miljøtilsyn 2001. Orientering nr. 8 fra Miljøstyrelsen*.
- Polinsky, A.M. og Shavell, S. 2000. The Economic Theory of Public Enforcement of Law, *Journal of Economic Literature*, vol. XXXVIII, s. 45-76.
- Reinganum, J.F. og Wilde, L.L. 1988. A Note on Enforcement Uncertainty and Taxpayer Compliance, *The Quarterly Journal of Economics*, vol. 103, nr. 4, s. 793-98.
- Rigsrevisionen. 2002. *Beretning til statsrevisorerne om kontrol af punktafgifter, RB C601/02 – Januar 2002*.
- Rigsrevisionen. 2003. *Beretning om statens restanceinddrivelse. 13/02*.
- Russell, C.S. 1990. Game Models for Structuring Monitoring and Enforcement Systems, *Natural Resource Modelling*, vol. 4, s. 143-73.
- Scotchmer, S. 1987. Audit Classes and Tax Enforcement Policy, *American Economic Review*, vol. 77, nr. 2, s. 229-33.
- Scotchmer, S. og Slemrod, J. 1989. Randomness in Tax Enforcement, *Journal of Public Economics*, vol. 38, s. 17-32.
- Told & Skat. 2004. *Oplysninger indhentet fra Told & Skat*.



# Short-run and long-run relationships in the consumption of alcohol in the Scandinavian countries

Jan Bentzen

Department of Economics, The Aarhus School of Business, Aarhus C, E-mail: [jb@asb.dk](mailto:jb@asb.dk)

Valdemar Smith

Department of Economics, The Aarhus School of Business, Aarhus C, E-mail: [vs@asb.dk](mailto:vs@asb.dk)

*SUMMARY: Traditionally, the Scandinavian countries have been characterized as spirits and beer consuming countries and a century ago the historical background was decades of relatively heavy drinking behaviour with spirit as the preferred beverage. It might be expected that alcohol consumption – especially in the last part of the 19<sup>th</sup> century and the beginning of the 20<sup>th</sup> century – would behave in a counter-cyclical manner, i.e. heavy drinking during severe recessions characterized by harsh economic conditions. Using long-run time series data for alcohol consumption levels in the Scandinavian countries the question of a counter-cyclical or pro-cyclical behaviour is addressed – with the business cycle measured as the GDP – and the empirical findings are that generally, alcohol consumption behaves in a pro-cyclical manner in the short run, and with no long-run relationship concerning real income.*

---

## 1. Introduction

The Scandinavian countries are usually characterized as spirits and beer consuming countries because these two beverages have been the far most popular alcoholic drinks through centuries. Until a few decades ago wine was consumed only in very modest quantities in these northern, climatically rather cold countries. For obvious reasons wine production is not possible or efficient in the Scandinavian countries and of course this has influenced the drinking behaviour. But from the beginning of the 1960s wine consumption suddenly increased in all of the Scandinavian countries and during the next three decades wine became a widespread, popular beverage – even with per capita consumption levels (measured in pure alcohol) surpassing spirits. At present, beer is still the most common alcoholic beverage in all of the three countries, but in relatively

---

We thank participants at the VDQS Oenometri XI Conference in Dijon, May 2004, for helpful comments on an earlier version of the paper.

few years from now wine may well be the most popular, alcoholic beverage in one or more of the Scandinavian countries.

The drinking patterns in the Scandinavian countries have evolved – or converged – towards a continental behaviour as far as alcohol consumption is concerned. Additionally, in the wine consuming countries of Southern Europe beer has recently gained popularity and therefore a much more uniform pattern of alcohol consumption is seen today among the European countries. However, this historical shift in the drinking behaviour in Scandinavian countries – i.e. wine consumption becoming widespread and probably substituted for spirits and/or beer – is only the latest episode in a long development and history of each country's alcohol consumption.

Significant and even dramatic changes in alcohol consumption did in fact take place regularly during the latest couple of decades in these countries and the aim of this paper is to shed light on the long-run movements in wine, spirits and beer consumption in the Scandinavian countries. Using long-run time series data for alcohol consumption levels, i.e. from the middle of the 19<sup>th</sup> century, the question of a counter-cyclical or pro-cyclical behaviour of alcohol consumption is addressed – with the business cycle measured as the real GDP.

## **2. Long-run trends in alcohol consumption in the Scandinavian countries**

In the 19<sup>th</sup> century considerable amounts of both spirits and beer were consumed – even compared to the present levels of alcohol consumption – but there is no long-run tradition for wine drinking in any of the Scandinavian countries. Until the late 1960s wine was considered partly as a luxury good, which was regularly consumed by only a small part of the population. Like other alcoholic beverages wine has usually been heavily taxed and therefore relatively good statistical sources are available when it comes to alcohol consumption levels. Thus, for Norway and Sweden per capita consumption levels of the specific beverages are published back to the 1850s and for Denmark data are available since the beginning of the 1890s with some exceptions.<sup>1</sup>

Figure 1 shows the long-run development in the per capita alcohol consumption levels in each of the Scandinavian countries (litres per capita aged fifteen years and above) – the statistical sources concerning alcohol consumption are presented in more detail in the data appendix. It is clear that significant changes have taken place in the overall alcohol consumption over the past 100-150 years – most dramatic for Denmark. In fact, the Danish per capita alcohol consumption level was as high as 15 litres in the late 19<sup>th</sup> century, which is even higher than today's consumption level. However, from the beginning of the 20<sup>th</sup> century the alcohol consumption dropped drastically to a level around 5-6 litres per capita where it stayed until the 1960s. Since then the

---

1. A copy of the time-series data for alcohol consumption is available upon request to the authors.

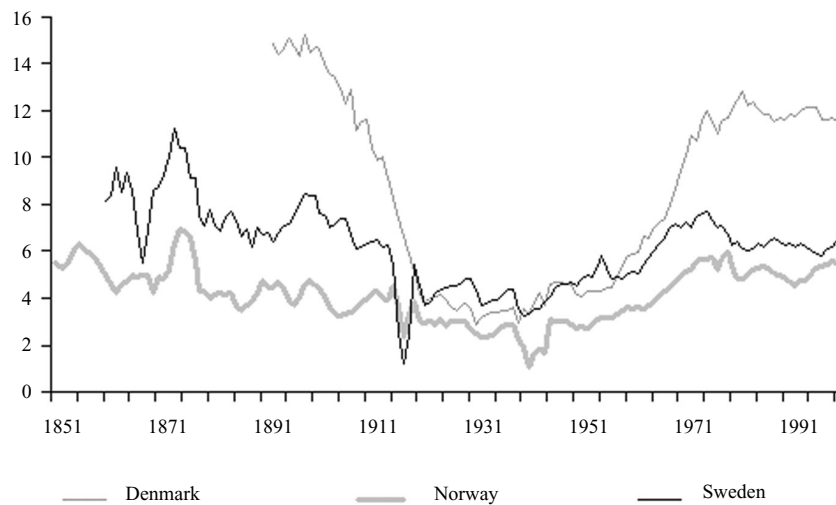


Figure 1. Alcohol consumption in the Scandinavian countries 1851-2002 (litres per capita, 100 pct alcohol, 15 years+).

Notes: The data are calculated as litres per capita for the part of the population aged 15 years or above.

Sources: Statistical sources as listed in the data appendix.

Danish alcohol consumption has increased mainly driven by higher consumption of beer and wine.

Of course the official data for alcohol consumption as presented here suffer from loss of information about illegal home production (especially spirits in Norway and Sweden) – and for the most recent part of the time period the question of border trade, i.e. private import of alcohol. For Sweden a restrictive legislation on home production of spirits and state control regarding the sales of alcohol were introduced around 1860-65 and similarly for Norway, where high taxes on home production of spirits were introduced in 1848 – and thereafter the official home production of alcohol strongly diminished. Most likely, the level of illegal production of spirits continued during the following century, but only little reliable information is available to make adjustments to the official alcohol consumption statistics. One source, Nordlund (1996), concludes from earlier surveys dealing with non-reported consumption of alcohol that this amounts to approximately 25-30 per cent of the total quantity of alcohol consumed in Norway. This problem is also addressed in Greenfield et al. (2003) and has to be taken into account especially when the data for alcohol consumption span e.g. periods of prohibition or other 'radical' alcohol policies are introduced.

Contrary to the situation in many other countries there have been no periods of prohibition in Denmark and even though the temperance movement existed from the late

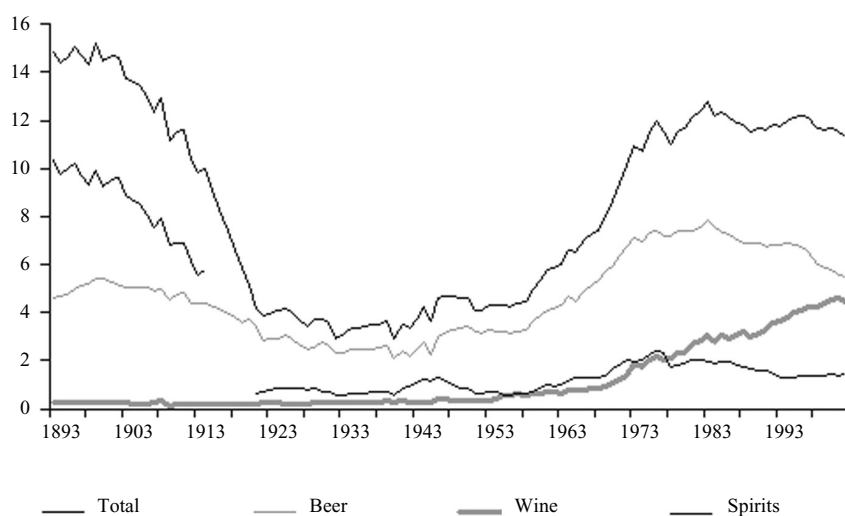


Figure 2. Consumption of alcoholic beverages, Denmark 1893-2002 (litres per capita, 100 pct alcohol, 15 years+).

Sources: Statistical sources as listed in the data appendix.

part of the 19<sup>th</sup> century it never had any decisive influence on Danish alcohol policies. Even though the temperance movements asked for prohibition in connection with World War I no policy actions in terms of ban and similar actions were taken. However, significant tax increases on alcohol were implemented in 1917, especially for spirits where prices increased ten-fold during a few months, thereby ending an era of very high levels of spirits consumption. The Danish alcohol taxes stayed at a relatively high level until Denmark joined the European Union in 1972 and the main reason for the high alcohol taxes and duties was predominantly fiscal.

The Swedish experience is in accordance with the development in Denmark. Despite rather strong temperance movements, direct prohibition was not implemented. Instead taxes on alcohol were increased and in connection with World War I production and sale of alcohol was monopolised by the state from 1919, and the retail sale of alcohol was rationed by the introduction of the so-called 'Bratt System' – a ration book for each individual regulating the amounts of alcohol bought. The system was later abandoned in the 1950s.

Two centuries ago Norway experienced significant alcohol problems and annual per capita alcohol consumption was as high as 13 litres around 1830-40. Naturally, the temperance movements gained momentum and in 1919 direct prohibition was decided. However, due to protests from wine producing countries, the ban on wine ceased in

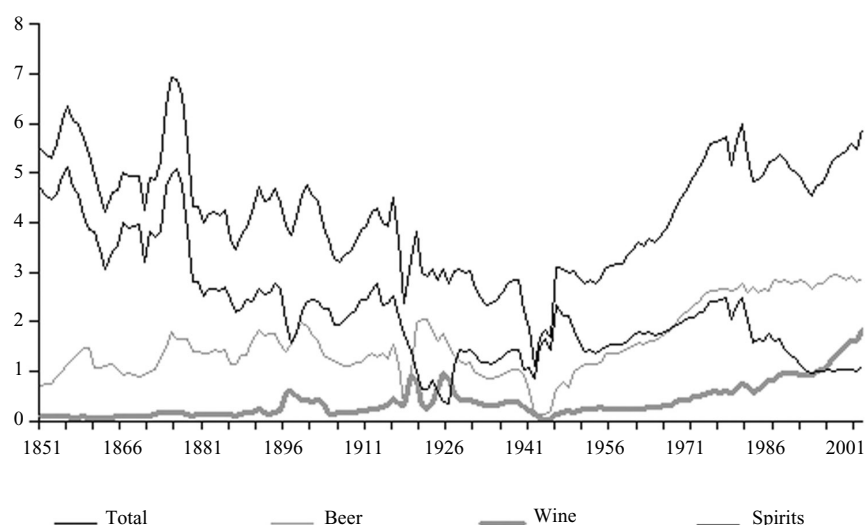


Figure 3. Consumption of alcoholic beverages, Norway 1851-2002 (litres per capita, 100 pct alcohol, 15 years+).

Sources: Statistical sources as listed in the data appendix.

1922 and so did the ban on strong wine in 1923. Finally, in 1926 the sale of strong alcohol (spirits) was allowed again. At the same time the retail and wholesale alcohol sales system changed to a system of state monopoly which still exists. Thus, for all the Scandinavian countries increased taxation of alcohol and other regulatory policies towards production and sale of alcohol was introduced during WWI and the following decade which had profound influence on the levels of alcohol consumption as exhibited in figure 1. For Norway further increases in taxes, rationing and closing of sales outlets were introduced during WWII and the following years with obvious effects on the level of consumption (figure 1).

Despite of the many restrictions concerning alcoholic beverages, spirits and beer have always been produced in all of the Scandinavian countries and consequently, they have been integrated parts of the food and drinking culture for centuries. Figures 2-4 also show the significant structural shifts with strongly increasing shares for wine and the marked declines for spirits in Norway and Sweden.

Today, beer is in fact the most popular drink but this might look different in the future if the present trends in beverage shares continue. Denmark is the first of the Scandinavian countries – in the post WWII period – to become a place where wine is regularly consumed and in levels comparable to other European countries. Being a part of continental Europe Denmark has adhered to relatively more liberal alcohol policies com-

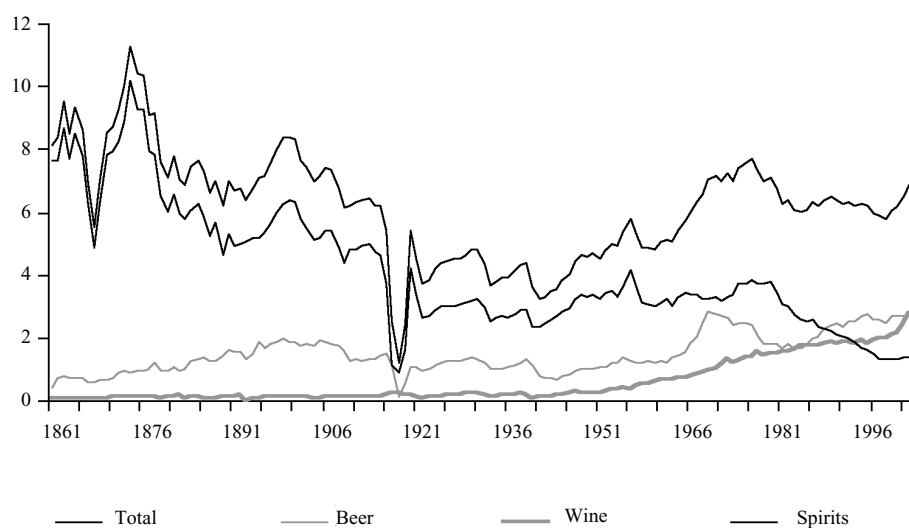


Figure 4. Consumption of alcoholic beverages, Sweden 1861-2002 (litres per capita, 100 pct alcohol, 15 years+).

Sources: Statistical sources as listed in the data appendix.

pared to Norway and Sweden, which may explain the higher Danish consumption levels. Of course, Denmark lags behind countries in Southern Europe concerning wine consumption, e.g. Italy with approximately 50 litres of wine per capita and for France the annual level is 60 litres per capita (direct measures in volumes), cf. also Anderson (2004).

In spite of the restrictive alcohol policies – i.e. regulations and taxation – all Scandinavian countries seem to have entered tracks of strongly increasing wine consumption levels, where Norway with the most restrictive policy seems to lag most behind. The main reasons for increases in wine consumption are probably the rising living standards stemming from the strong growth rates of real incomes and more open economies, i.e. especially the tendencies developed from the early 1960s – the Scandinavian populations travelling on holidays to Southern Europe – where wine was also abundantly and cheaply available compared to the conditions in the home countries! As no grape wine production takes place in the Scandinavian countries consumer habits or preferences for wine had to be adopted from outside.

#### *The national retail systems*

As mentioned above state monopolies in production, trade and sales were established in Norway and Sweden at the beginning of the 20<sup>th</sup> century for alcoholic beverages,

although there were some exemptions for beer. Still today, Norway and Sweden have state monopolies, although they are only in effect regarding retail sales of alcohol (again with some exemptions for beer). For the production and wholesale of alcohol – beer, wine or spirits – a licence is needed and furthermore, spirits production is a state monopoly in Norway. Denmark is the exception with none of these regulatory systems. However, all the Scandinavian countries have used heavy taxation in order to create considerable fiscal revenues and – usually presented as an important motive – also to curb all kinds of alcohol consumption, cf. Bentzen and Smith (2004).

Sweden has organized the retail system of alcoholic beverages as a state-owned monopoly 'Systembolaget' operating through shops (less than 500) or local agents in 575 communities. The monopoly covers spirits, wine and strong beer. But light beer is available in other shops too and the Swedish breweries can sell strong beer (above 3.5% alcohol) directly to restaurants. At the retail level spirits, wine and strong beer may only be sold to persons aged 20 years or more. Today, there are no quantitative restrictions on the amounts of wine individuals may buy, but as the 'Systembolaget' is a state monopoly with a policy of supporting 'a healthy drinking culture' they are not striving to attract customers to the shops.

Since the 1950s, some effort has been devoted to the objective of making customers substitute wine for spirits because wine is considered healthier relative to spirits and therefore a relatively faster introduction of wine in the consumption of alcoholic beverages has taken place despite the otherwise restrictive retail sales system. Until 1994 'Systembolaget' had an effective monopoly also in the case of quantities sold to e.g. restaurants, but this part of the monopoly ceased in 1995 and there are currently 200 licensed, private import companies supplying wines to 'Systembolaget', restaurants, hotels, etc.

After Sweden joined the European Union in 1995 problems were expected for the sales monopoly in the market-oriented community. However, in 1997 the EU Court of Justice ruled that 'Systembolaget' was not in contradiction with an EU membership because it was created due to public health considerations. Moreover the Swedish system was not found to discriminate between foreign and Swedish products. Still a gradual liberalization of the border trade regulations, i.e. personal imports of (cheaper) alcoholic beverages, has taken place and more important the monopolies in import/export, wholesales and production (spirits) have disappeared.

Norway has a system rather similar to the Swedish system and 'Vinmonopolet' was established in order to control a widespread misuse of spirits – 'aqua vitae'. Norway is not a member of the European Union, but partly due to trade-agreements – and the general liberalization of international trade – the state-monopoly was split up in the mid 1990s and, similar to the changes in Sweden, 'Vinmonopolet' is today only a retail

sales monopoly. In total, there is less than 200 sales outlets, and the low density of alcohol stores combined with the geography of Norway, i.e. mountains, forests, a lot of snow in the wintertime, etc. makes it difficult for people living outside urban areas to buy alcohol.

Traditionally, Denmark has adhered to liberal, market-oriented systems. Taxes have been used primarily for fiscal purposes and except for a minimum age of 18 years for sale at restaurants, there have generally been no impediments to alcohol consumption. Recently, a limit of 15 years was introduced for the sale of alcohol from retail shops. Furthermore, after the Danish membership of the European Union in 1972, free trade and harmonization of tax systems have contributed to the continuous increase in the absolute alcohol consumption levels, especially for wine – cf. figure 2 – which shows a significant increase in the consumption of wine from the early seventies. In fact, a tax cut of 75 per cent on wine was effectuated after 1972, and the tax rates on beer and spirits were fixed in absolute terms, and again in the 1990s a further lowering of Danish alcohol taxes was effectuated in order to complete the harmonization of the Danish tax levels towards other countries in the European Union. Accordingly, it is evident that much more attention has been paid to fiscal consideration and obligations to the EU-membership than to health and social arguments in the Danish alcohol policies.

Norway and Sweden both decided not to enter the EU in 1972 and hence, both countries were able to formulate their own fiscal policy, especially with regard to the taxation of alcoholic beverages. In Sweden real alcohol prices were relatively constant during the 1960s, but they have been rising since the beginning of the 1970s. Thus fiscal considerations have been of significant importance to the Swedish taxation of alcoholic beverages. Norway also decided to stay outside the EU in 1994, by a referendum, and therefore, Norway has no legal problems withholding a high level of taxation on alcohol – but there are increasing problems concerning border trade in relation to both Denmark and Sweden.

### **3. Alcohol consumption and the business cycle – a short-run relationship**

From an economic point of view there might be a positive relationship between alcohol consumption and real income assuming that alcohol is a normal good. Alternatively, a counter-cyclical relationship might also be present if for example recessionary periods are accompanied by more heavy drinking behaviour and thus, the relationship between level of income and alcohol consumption might not be present. In order to address the question of the relationship between the general business cycle – measured as real GDP per capita – and the consumption of alcohol, scatter plots of log differences of these variables are exhibited in the figures 5, 6 and 7 for the respective countries.



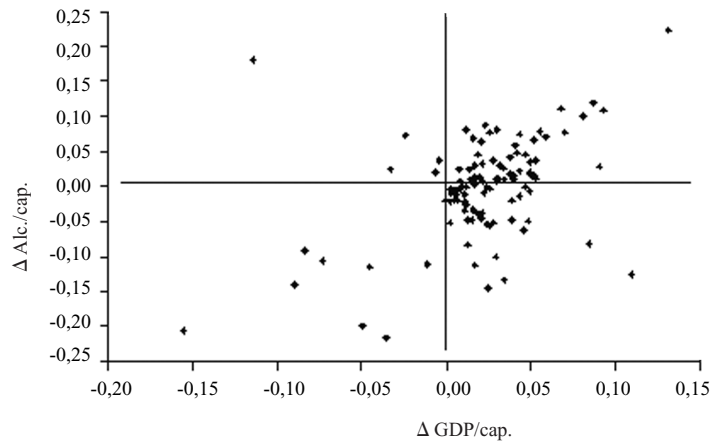


Figure 5. Scatter plot of real GDP per capita and alcohol consumption per capita (15 years+) for Denmark, 1893-2002 (values in log differences).

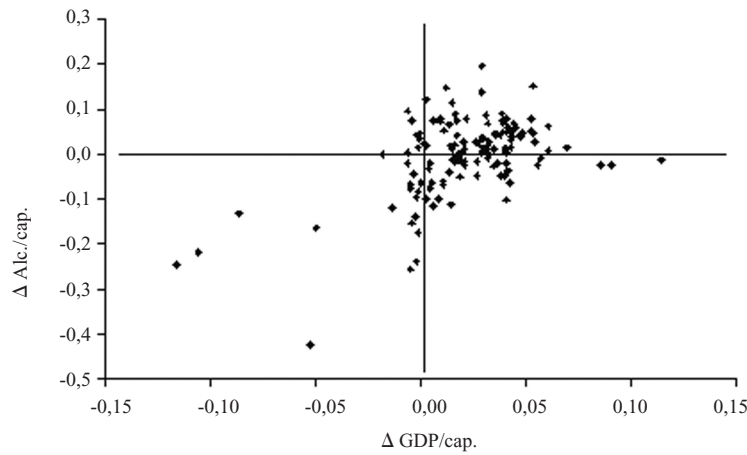


Figure 6. Scatter plot of real GDP per capita and alcohol consumption per capita (15 years+) for Norway, 1865-2002 (values in log differences).

Note: A few outliers with extreme values are not included in the graph.

When using values in log differences, only short-run relationships are revealed as information about level values are filtered away via the difference operator.

The figures 5-7 reveal that there is no doubt about a pro-cyclical relationship between the growth rates of GDP and alcohol consumption. This result is also found for sub-samples of the data – even for data related to the 19<sup>th</sup> century – and therefore not much evidence is found in favour of a hypothesis of 'heavy drinking in harsh times' as

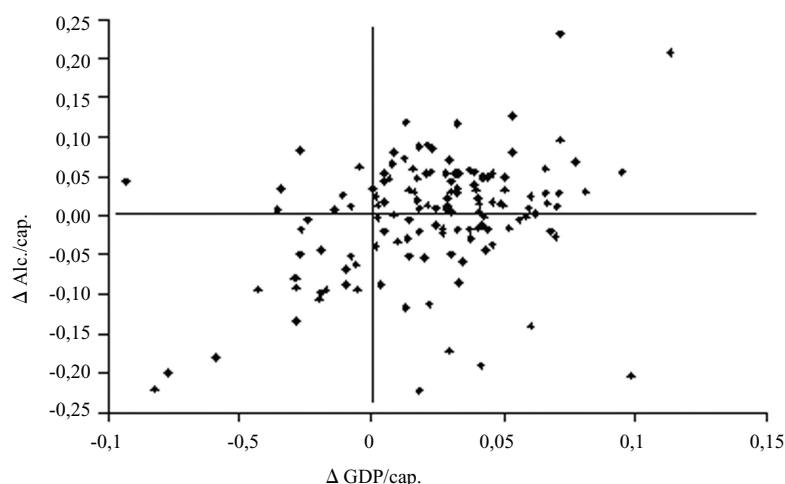


Figure 7. Scatter plot of real GDP per capita and alcohol consumption per capita (15 years+) for Sweden, 1861-2002 (values in log differences).

Note: A few outliers with extreme values are not included in the graph.

an escape from poverty and severe social problems. For the specific beverages, i.e. beer, wine and spirits, the cross correlation coefficients are reported in table 1.

The conclusion from table 1 is in accordance with the plots in the figures above as all correlations are found to be positive. Hence, in the short run there is a pro-cyclical pattern for the overall consumption of alcohol as well as for the specific alcoholic beverages. In the long run there is less doubt about substitution processes taking place between beer, wine and spirits as exhibited in the graphs in figures 2-4. Looking at these data covering more than a century, the general pattern for all the Scandinavian countries is a steady decline in the consumption of spirits in favour of increasing consumption of beer and wine – where the consumption levels of the latter has increased tremendously since the 1960s and probably will be the most important alcoholic beverage in the future evaluated from the present trends in consumer behaviour.

From an economic point of view alcohol consumption is expected to be influenced by real income and real prices of the specific alcoholic beverages – as well as other variables, of course. From the presented graphs and correlation analysis there seems to be evidence of a positive income effect, but data for prices of the alcoholic beverages are not available for the long time span covered by the consumption and income data already exhibited.<sup>2</sup> Therefore, in order to obtain an estimate of the (short-run) income effect in alcohol consumption an autoregressive distributed lag model (ARDL) is applied to the data:

2. See Bentzen et al. (1997, 2004) and Milhøj (1993, 1995) for prices concerning alcoholic beverages and price elasticities.

Table 1. Cross correlations between GDP per capita and per capita consumption of beer, wine and spirits (Log differences).

	Denmark	Norway	Sweden
GDP : Beer	0.461	0.345	0.244
GDP : Wine	0.178	0.406	0.118
GDP : Spirits	0.201	0.154	0.359

Note: The data used when calculating the correlation coefficients are 1893-2002 for Denmark, 1865-2002 for Norway (the data for GDP only covers this period, but the alcohol consumption data goes back to 1851) and 1861-2002 for Sweden.

$$\Delta \ln A_t = \alpha + \sum_{i=0}^p \beta_i \Delta \ln Y_{t-i} + \sum_{i=1}^q \delta_i \Delta \ln A_{t-i} + \varepsilon_t \quad (1)$$

First differences of income ( $Y$ ) and alcohol consumption ( $A$ ) measured in per capita terms are included in the RHS of (1) with a suitable number of lags following the general-to-specific estimation technique. The income variable is assumed exogenous and the parameter estimates are reported in table 2.

The resulting parameter estimates are found to be very significant, although the models for Norway and Sweden still suffer from ARCH errors, which will remain even when extending the number of lags. Assuming the ARDL models are reasonably well representing short-run relationships in alcohol consumption – in spite of the missing price variables, which may cause bias in the parameter estimates – the best available estimates of the income elasticities are the  $\beta_0$ -parameter estimates. Thus, the short-run income elasticities are relatively high, especially for Norway and Sweden, and even if the magnitudes are somewhat exaggerated due to missing variables, the conclusion is still a strongly pro-cyclical alcohol consumption pattern.

In the long run no evidence for a positive relationship between income and alcohol consumption levels is expected – for obvious reasons as there will be a natural limit to the human intake of alcohol. Evaluated from table 2 the long-run elasticity between the *growth rates* of income and alcohol consumption will be zero for all three countries. As first differences of the variables are estimated in (1), only relationships between growth rates can be evaluated – not level values – and the parameter estimates in table 2 all imply a zero long-run relationship between alcohol and real income. Consequently, the next question to address in the following part 4 will be whether there is some kind of 'mean reversion' in the long-run level of alcohol consumption for the Scandinavian countries.

Table 2. Parameter estimates of the ARDL models.

	Denmark		Norway		Sweden	
$\alpha$	-0.019 *	(-2.78)	-0.030 *	(-2.68)	-0.034 *	(-3.07)
$\beta_0$	0.806 *	(5.37)	1.480 *	(5.06)	1.444 *	(5.46)
$\delta_1$	–	–	–	–	0.321 *	(4.87)
$\delta_2$	0.252 *	(3.01)	–	–	-0.512 *	(-7.80)
$\bar{R}^2$	0.25		0.15		0.44	
DW	1.82		2.20		1.93	
LM(2)	1.29	[0.52]	2.08	[0.35]	3.57	[0.17]
LMARCH(2)	0.35	[0.84]	5.41	[0.07]	39.72	[0.00]

Note: *t*-values of the parameter estimates in parenthesis and with a \* indicating significance at least at the 5 per cent level. LM(2) and LMARCH(2) are Lagrange Multiplier tests for second order residual autocorrelation and second order autoregressive conditional heteroscedasticity, respectively, with *p*-values in parenthesis [ ].

#### 4. Alcohol consumption in the long run: a mean reversion process or a random walk?

When evaluating the data for alcohol consumption from figure 1 these variables seem very much to exhibit random walk behaviour, i.e. be non-stationary I(1)-variables. Testing for unit root behaviour by the Dickey-Fuller test statistic is reported in table 3.

The conclusion from the DF/ADF-tests is in all three cases non-stationarity as the critical value is -2.88 at the 5% level of significance according to MacKinnon (1991). When taking a closer look at figure 1 an alternative interpretation of the unit root hypothesis – especially for Norway and Sweden – might be an asymmetric adjustment process towards a long-run level of alcohol consumption that is non-zero. Such an alternative hypothesis to the random walk behaviour as tested for in table 3 might be the so-called threshold autoregressive model (TAR-model) as presented in Enders and Granger (1998), see also Enders and Siklos (2001). In the present case of alcohol consumption levels the most obvious version of these models will be the TAR model with a linear attractor – the latter being the average long-run level of alcohol consumption. The threshold model (TAR) with a linear attractor is defined by (2) and (3), see Enders and Granger for further details:

$$\Delta \ln A_t = I_t \gamma_1 (\ln A_{t-1} - \bar{A}) + (1 - I_t) \gamma_2 (\ln A_{t-1} - \bar{A}) + \sum_{i=1}^p \omega_i \Delta \ln A_{t-i} + \varepsilon_t \quad (2)$$

where the linear attractor ( $\bar{A}$ ) is the average value of  $A_t$  in logs and  $\varepsilon_t$  is a white noise disturbance term. The indicator function,  $I_t$ , is defined as:

*Table 3. Unit root test of alcohol consumption levels (log values).*

	Denmark	Norway	Sweden
<i>DF/ADF</i>	-1.068 {0}	-2.212{0}	-2.342{2}
<i>N</i>	109	151	139

*Note:* *N* is the number of observations, and the number of lags in the augmented DF-test indicated by the {}-parenthesis.

$$I_t = \begin{cases} 1 & (\ln A_{t-1} \geq 0) \\ 0 & (\ln A_{t-1} < 0) \end{cases} \quad (3)$$

An extension of this model is the M-TAR-model (Momentum TAR) where the indicator function is changed to:

$$I_t = \begin{cases} 1 & (\Delta \ln A_{t-1} \geq \bar{A}) \\ 0 & (\Delta \ln A_{t-1} < \bar{A}) \end{cases} \quad (4)$$

Hence, the model in (2) along the two versions of the indicator function, i.e. (3) and (4) respectively, is used to test the unit root hypothesis versus the TAR and M-TAR models. The test statistics are reported in table 4, where the Enders-Granger  $\Phi\mu$ -statistic ( $\Phi\mu^*$  for M-TAR) concerns the rejection of the unit root null hypothesis in favour of the TAR and M-TAR models from (2).

For Denmark and Norway the unit root behaviour of alcohol consumption levels is not rejected in favour of the alternative models, but for Sweden the  $\Phi\mu^*$ -statistic is above the 5 per cent critical value (approx. 4.72) and therefore the M-TAR model cannot be ruled out in this case. Hence, for Sweden the long-run development of alcohol consumption may be described as an asymmetric adjustment process with a linear attractor given by the average level of alcohol consumption for the whole sample period 1861-2002.

## 5. Conclusion

The aim of this analysis has been to analyse both the cyclical aspects and the long-run movements in wine, spirits and beer consumption in the Scandinavian countries. Using long-run time series data for alcohol consumption levels, i.e. from the middle of the 19<sup>th</sup> century the question of a counter-cyclical or pro-cyclical behaviour is addressed in particular. Although the Scandinavian countries are quite similar in many respects as

Table 4. Test statistics of the TAR and M-TAR models.

	Denmark		Norway		Sweden	
	TAR	M-TAR	TAR	M-TAR	TAR	M-TAR
$\gamma_1$	-0.015	-0.013	-0.027	0.051	-0.050	0.001
$\gamma_2$	-0.011	-0.012	-0.088	-0.081	-0.101	-0.182
Lags( $p$ )	0	0	0	0	2	3
$\Phi\mu$	0.517	–	2.969	–	3.015	–
$\Phi\mu^*$	–	0.509	–	2.581	–	6.853 *

Note: Critical values for  $\Phi\mu$  and  $\Phi\mu^*$  reported in Enders and Granger (1998), with corrections in Enders (2001). An \* indicates a significant test-statistic at least at the 5 per cent level of significance.

concern history, culture, languages and lifestyles they do in fact differ with respect to drinking habits and alcohol and wine consumption levels.

The conclusion from the correlation analysis between growth rates of income and alcohol consumption points in the direction of a strong positive relationship. Moreover, in the short run there is a pro-cyclical pattern for the overall consumption of alcohol as well as for the specific alcoholic beverages of beer, wine and spirits. Using an autoregressive distributed lag model (ARDL) the short-run income elasticities are also found to be relatively high, especially for Norway and Sweden, again suggesting a strongly pro-cyclical alcohol consumption pattern.

The outcomes of the analysis of estimating the ARDL models indicate that the long-run elasticity between the *growth rates* of income and alcohol consumption will be zero for all three countries. The analysis also deals with the question of whether alcohol consumption in the long run can be considered as a mean reversion process or as a pure random walk. Using the TAR-model (threshold autoregressive model) and M-TAR-model (Momentum TAR) the results suggest that alcohol consumption in Sweden can be described as an asymmetric adjustment process with a linear attractor given by the average level of alcohol consumption for the whole-sample period 1861-2002. However, alcohol consumption in Denmark and Norway exhibit more evidence of pure random walk behaviour.

#### References

- Anderson, K., ed. 2004. *The World's Wine Markets. Globalization at Work*. Edward Elgar, London.
- Bentzen, J., N. Nannerup and V. Smith. 1997. Alcohol consumption and drunken driving in the Scandinavian countries. *Working Paper 97-1*, Department of Economics, The Aarhus School of Business.
- Bentzen, J. and V. Smith. 2004. Wine in the Nordic countries: Historical and recent trends in the development of wine consumption, prices and taxation. In: Anderson, K., ed. (2004): *The World's Wine Markets. Globalization at Work*. Edward Elgar, London.
- Enders, W. and C. W. J. Granger. 1998. Unit-

- Root Tests and Asymmetric Adjustment With an Example Using the Term Structure of Interest rates. *Journal of Business & Economics Statistics*, 16 (3), s. 304-11.
- Enders, W. 2001. Improved critical values for the Enders-Granger unit-root test. *Applied Economics Letters*, 8, s. 257-61.
- Enders, W. and P. L. Siklos. 2001. Cointegration and Threshold Adjustment. *Journal of Business & Economics Statistics*, 19(2), s. 166-76.
- Greenfield, T. K. and W. C. Kerr. 2003. Tracking Alcohol Consumption Over Time. *Epidemiology in Alcohol Research*, 27, s. 30-38.
- Nordlund, S. 1996. Rusmiddelbruk i Norge. *Norsk Epidemiologi*, 6, s. 3-12.
- MacKinnon, J. G. 1991. *Critical values for cointegration tests*. In Engle, R.F. & Granger, C. W. J. Long-run Economic Relationships. Oxford Economic Press.
- Milhøj, A. 1993. Virkningen af afgiftsnedsettelser på øl- og vinsalget i Danmark. *Nordisk Alkohol Tidsskrift*, 10, s. 319-29.
- Milhøj, A. 1995. Udviklingen i danske familiers øl- og vinkøb 1990-1993. *Nordisk Alkohol Tidsskrift*, 12, s. 181-93.

**Appendix: Data sources**

Statistisk Aarbog (The Danish Statistical Yearbook), var. issues 1893-1965, Statistics Denmark.

*www.statistikbanken.dk* (The Statistical Databank), Statistics Denmark.

*www.sbb.no* (The Statistical Databank), Statistics Norway.

*www.sbc.se* (The Statistical Databank), Statistics Sweden.

Historisk Statistikk (Historical statistics), Statistics Norway.

Statistisk Aarsbok (The Swedish Statistical Yearbook), var. issues 2000-2004, Statistics Sweden.

Sales statistics for spirits, wine and beer in Sweden (Alkoholinspektionen, *www.fhi.se*), Sweden.

Nordic Alcohol Statistics 1994-1998 (Stakes 2000), Finland.

Hansen, S. Aa. 1974. Economic growth in Denmark (in Danish), Copenhagen.

Kaergaard, N. 1991. Economic Growth (in Danish), Copenhagen.

Krantz, O. 2001. Swedish Historical National Accounts 1800-1998 – Aggregated Output Series. Working Paper, Department of Economic History, University of Umeaa, Sweden.

The above mentioned statistical sources have been used for the construction of annual data for alcohol consumption levels measured in litres of pure alcohol – including the specific beverages of beer, wine and spirits – and data for real GDP and population (aged 15 years and above).

Data for Denmark before 1960 were collected from the annual statistical yearbooks and in some cases some data had to be estimated, especially when converting the volumes in litres of beer, wine and spirits into litres of 100 per cent alcohol. Data for alcohol consumption covers the time period 1893-2002 for Denmark (with the exception that data for spirits is missing for 1915, 1917-18 and 1921), 1851-2002 for Norway (data missing for 1998, but estimated) and 1861-2002 for Sweden.

Data for real GDP and population (+15 years and above) for Denmark and Sweden correspond to the time periods covered by the alcohol consumption statistics. For Norway data cover the time period 1865-2002 (i.e. shorter time span than the alcohol data series) and data for GDP are missing for 1940-1945 (estimated for the purpose here as the level of GDP in the late 1930s correspond well to the level immediately after WWII).



# The Effect of Labour Market Conditions on the Time-to-Completion of Higher Education in Denmark

Nabanita Datta Gupta

The Danish National Institute of Social Research, E-mail: [ndg@sfi.dk](mailto:ndg@sfi.dk)

Mark Yuying An

Fannie Mae, U.S.A., E-mail: [mark\\_y\\_an@fanniemae.com](mailto:mark_y_an@fanniemae.com)

*SUMMARY: The effect of labour market conditions on the time-to-completion of higher education is analysed on a sample of high school graduates and equivalents drawn from the 0.5% Danish Longitudinal Sample database. A semi-parametric hazard model of education duration with correction for unobserved heterogeneity is proposed and estimated. Results indicate that completion hazard is significantly affected by individual background characteristics, program-specific factors and labour market variables. The expected wage upon completion significantly increases degree completion hazard while the higher the individual's own earnings while studying, the lower the completion hazard. However, the unemployment degree that students may expect to face upon graduation does not, as expected, have a delaying impact on education completion.*

---

## I. Introduction

This paper analyses whether labour market conditions affect the completion time of higher education of Danish high school graduates. Danish students take a relatively longer time in completing higher education compared with students in other OECD countries. For example, the average age at long-term (university) higher education completion in Denmark is 29.4 years, compared to an average of 25.2 years in Holland.<sup>1</sup> A recent study by the Danish Employers Federation, Dansk Arbejdsgiverfor-

---

Thanks to Ronald G. Ehrenberg, Steve Bronars, Niels Westergaard-Nielsen and participants of the CLS Workshop on Higher Education, Aarhus, June 1998, and to Anne Møller Danø, Helena Skyt Nielsen, Michael Rosholm and other participants of the Rørvig workshop, April 2002 for useful comments. Use of the data is made possible by financial support from the Danish National Research Foundation.

Mark Y. An was visiting the Economics Department, University of Aarhus, and the Center for Labour Market and Social Research (CLS), Aarhus, at the start of this project.

1. OECD Education at a Glance, Education Indicators, 2001.

ening (2004) reports that graduates in Denmark take on average 5<sup>1/2</sup> years more than necessary in completing their higher education.<sup>2</sup> Frequently cited reasons for this phenomenon are first of all a late school start<sup>3</sup> plus a delayed start of higher education with the widely used practice of a 'sabbatical year' of work or study abroad immediately following high school graduation. Also, the fact that a growing fraction of high school students spend an extra (voluntary) year in secondary school before going on to high school, and because Danish students of higher education frequently take longer than the norm over their degree completion because they either change educational programs or take non-credit courses. To some extent, this type of behaviour may reflect rational responses to changing conditions in the labour market so that students take longer over their education if they expect to be unemployed or face low wages upon completion. Another effect arising from the labour market could be that Danish students frequently take up paid work in the labour market alongside their education and this has a delaying impact on education completion. Thus, an important question would be to try to assess and quantify the relative importance of these labour market factors on degree completion time.

In a scenario of tightening labour market conditions in Denmark because of falling unemployment rates and smaller entering youth cohorts, policy makers warn that lengthening degree completion times only exacerbates the excess demand for educated labour. In fact, according to recent projections released by the Danish Employers Federation (2004), moving students quicker through the educational system would add 22,000 jobs over the next 20 years. Viewed from the micro perspective too, the costs to lengthening degree completion may be considerable. For example, the returns to a bachelor degree in business falls from 14% to 4% if the education is started at age 33 instead of age 18. The same delay in obtaining an electrician's education results in a fall from 7.5% to 4% (Danish Council of Economic Advisors Spring 1995 Report).

A few previous studies have analysed the question of the factors that affect the time to degree completion, particularly in the case of doctoral study. Tuckerman et al. (1990) estimates a regression model based on aggregate (field-level) time-series data and Abedi and Benkin (1987) analyse data from a single Ph.D. granting institution for a 10-year period. A few studies have also analysed completion rates, instead of focusing only on completers or degree recipients, as this may lead to selection bias. Bowen and Rudenstine (1992) utilize data on entrants to graduate programs in six fields in 10 major research universities in the U.S. and find that the type of financial support signifi-

---

2. This study projects that as a consequence of delaying education completion, between 2002 and 2020, Denmark will have fallen in rank from 11<sup>th</sup> to 14<sup>th</sup> globally, in terms of the share of workers with higher education.

3. Formal schooling starts at age 7 in Denmark, although most children are enrolled in kindergarten/pre-school between ages 3-6.

cantly affects mean differences in completion rates and time-to-degree across these groups. Booth and Satchell (1996) analyse some of the same issues on British data on one entering Ph.D. cohort in 1980. The most complete study in this area to date is by Ehrenberg and Mavros (1995), who model the time to completion of higher education using a discrete-time proportional hazard framework in which they analyse the effect of changes in financial support, changes in student ability and changes in academic labour market conditions on the time-to-degree completion and completion rates. Their data is restricted to all entrants to PhD programs in 4 fields (Economics, English, Mathematics and Physics) during the 1962 to 1986 period, in one single major U.S. University, Cornell University. While their study incorporates drop-out behaviour by modelling the competing risks of either degree completion or dropping out, they are unable to control for unobserved ability that may arise due to unobserved heterogeneity.

In this paper, we consider the role of labour market conditions as well as student background characteristics, student ability and program-specific factors on the time to degree completion within the framework of a hazard model. Our focus is not just doctoral students from a single institution or field, but all students that enter higher-education programs in Denmark. In addition, unlike most of the previous studies mentioned above, we correct for differences in individual unobserved ability that may affect the time to degree completion. Unobserved ability differences among students can confound the effect of education on labour market outcomes as high-ability individuals tend to perform better in school and in the labour market. Wage studies of the effect of schooling on earnings typically have addressed this problem by including separate measures of ability in wage regressions, or by analysing how wages are affected when a random event unrelated to ability affects their schooling, or even by analysing differences in schooling among family members who have the same inherited characteristics, such as twins, see Card (2001). The same problem can arise when estimating the effect of observed characteristics on degree completion as omitted ability can bias the returns to observed characteristics. That is, higher-ability students may be quicker to complete their education because they experience lower psychic costs (i.e. have an easier time learning etc.) and because their forgone earnings are likely higher. In a recent paper, van Ours and Ridder (2003) analyze the supply of Ph.D.s in economics in the Netherlands within a competing risks hazard model and find that students who are more likely to take a long time to graduate are the ones most likely to drop out. Selection is important, as good supervisors attract better students, and these students typically also have a higher probability of staying on the program. Similarly, we propose a correction for unobserved heterogeneity in higher education completion. In doing so, we apply a proportional hazard modelling framework and an estimation technique of fitting the model with grouped duration data developed by An (2000). This model

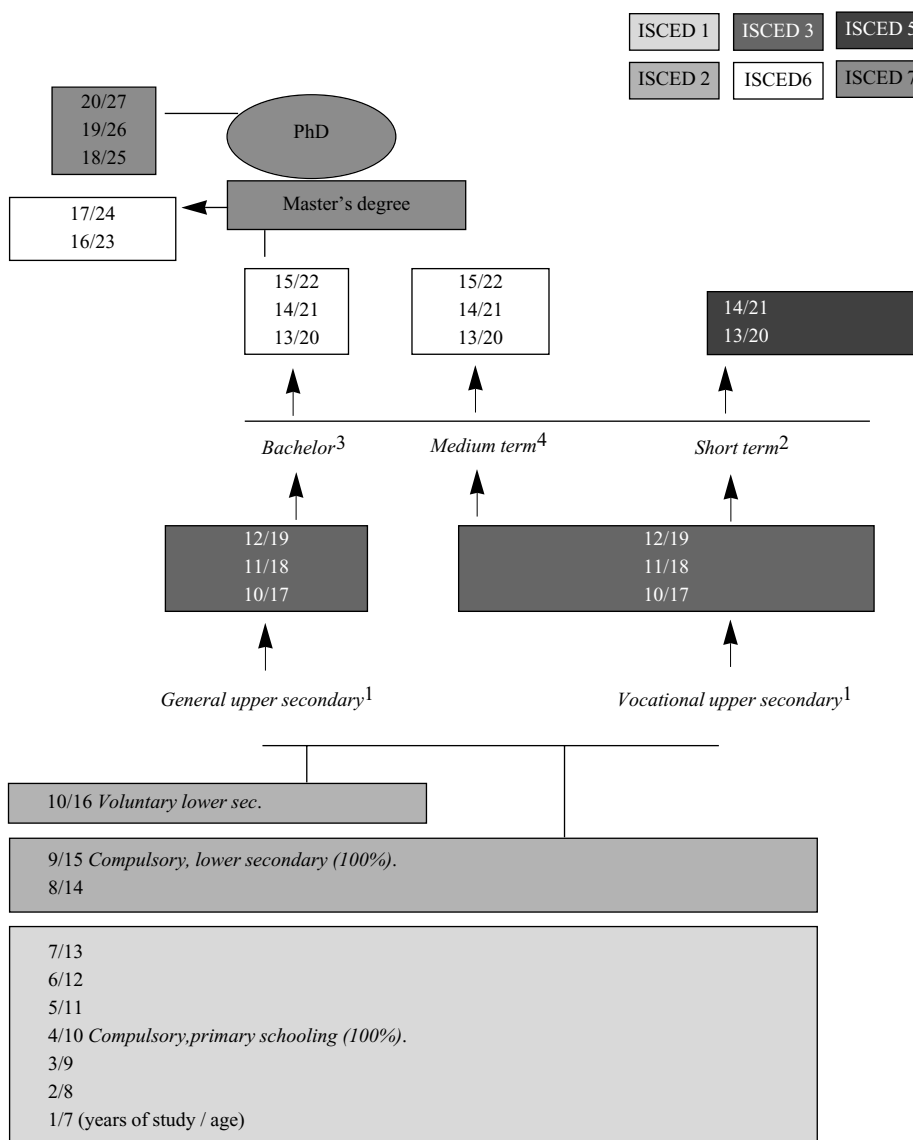


Figure 1. An overview of the Danish educational system.

1. Passing the school leaving examination at the end of basic school qualifies for Gymnasium.
2. HF (2 years) requires 10th grade. HHX and HTX (general vocational upper secondary) require 1 year of vocational training after basic school.
3. General or general vocational upper secondary education is required.
4. Apprenticeship / Skilled workers. Combined theoretical school education and practical training.
5. Bachelor's degree required to get admission to Master's degree courses.

Source: Christensen and Westergaard-Nielsen, (1999).

allows flexible functional forms for both the baseline hazard and the distribution of the unobserved heterogeneity.

The rest of the paper is organized as follows: Section II gives an overview of the educational system in Denmark, Section III lays out a simple theoretical framework for the subsequent reduced form empirical model, Section IV specifies the duration model which is the basis for the analysis, Section V discusses the data, Section VI presents sample descriptive statistics and Section VII the results of estimation. Section VIII concludes.

## II. Higher Education in Denmark

Figure 1 gives an overview of the Danish educational system. In the Danish system, children enter school at age 7 and must take 7 years of compulsory primary schooling plus 2 years of compulsory lower secondary schooling. Following the basic 9 years of compulsory schooling, the routes to higher education are either to directly proceed to a 3-year high school (*gymnasium*) program or to take an extra (voluntary) 10<sup>th</sup> year of lower secondary school before going on to the 3-year high school. A 2-year high school equivalent degree (*HF*) requires the 10<sup>th</sup> grade of secondary school. Vocation-based high school degrees such as in business management or technical fields require one year of basic training following compulsory school. Following high school, bachelor programs are typically 3 or in some cases, 2 years in length. Masters programs are 2 years and the Ph.D. is a 3-year program. Figure 1 also gives corresponding ISCED levels.<sup>4</sup>

## III. Theoretical Considerations

Individuals (students) take into account future earnings and unemployment risk when deciding how many years they take to complete a given higher educational program, that is, whether or not to extend the time taken to complete the educational program. We ignore the issue of dropouts, i.e. persons who begin an educational program but do not complete it.

Let  $V$  be the value to an individual say high school graduate, of completing a higher educational program in  $t$  years.

$$V(t) = V(Y, \pi, UIB, C, wH, G, Z, p, U) \quad (3.1)$$

The (net present) value depends on the future earnings  $Y$  after completion,  $\pi$ , the probability of employment and  $UIB$ , the unemployment benefit. It also depends on costs of completion ( $C$ ), both out-of-pocket costs plus forgone labour market earnings

---

4. ISCED = International Standard Classification of Education.

incurred each year spent studying. As students are liquidity constrained, they need to supplement their income with labour market earnings equal to  $wH$ , where  $w$  is the current wage rate and  $H$  the hours supplied to the labour market while studying, and the financial aid grant  $G$ , which may be a declining function of the number of years taken to complete the education. Further, there could be  $Z$ , individual background factors,  $P$  certain program-specific factors and  $U$ , which is individual unobserved ability affecting the value.

The problem would be to choose both  $t$ , the years taken to complete the education, and  $H$ , the hours of concurrent work, given the predetermined variables in the value function. In the empirical formulation, we abstract away from the hours choice, treating current earnings as a given, and focus only on the choice of  $t$ . A more complete study could potentially consider the joint determination of both variables.

The maximized value of completion  $V(t^*)$  is attained after  $t^*$  periods in the program, where  $t^*$ , based on the above model depends on earnings before and after completion, probability of unemployment after completion, unemployment benefits, costs, financial aid, individual background factors, spell-specific factors and individual unobserved heterogeneity.

While the theoretical considerations above are supposed to show the tradeoffs between for example, working and studying and other factors affecting students' choices, in the empirical specification, we estimate the distribution of completion times conditional on characteristics, by constructing suitable regressors in an empirical model of observed durations until completion. As direct costs of education and unemployment benefits do not vary much across individuals in Denmark, these are suppressed in the empirical work. Thus, the main arguments should influence completion time as follows:

$$t^*(X) = t^*(\overset{-}{Y}, \overset{+}{\pi}, \overset{+}{wH}, \overset{+}{G}, \overset{-}{U}, \dots) \quad (3.2)$$

#### IV. Empirical Model

For the empirical analysis, we adopt a hazard-function approach to model duration of education spells. Let  $T_n$  be the duration in years for individual  $n$ .  $X_n(t)$  is a vector of observed covariates, some of which may be time-varying. Let  $U_n$  denote unobserved heterogeneity with distribution function  $G(u)$ .

##### (A) Model Specification

We start with the specification of the hazard function  $h(t)$  for  $T_n$ , that is,  $h(t)$  measures the conditional probability density function, given the spell is at least  $t$ :  $h(t) = f(t)/P(T > t)$ . The following assumptions are made.

*Assumption 1 [Proportional Hazard].* The hazard function conditional on both observed and unobserved heterogeneities  $(X_n, U_n)$  is of the following proportional hazard (PH) type,

$$h(t|X_n(t), U_n) = \lambda(t)e^{X_n(t)\beta} e^{U_n} = \lambda(t)e^{X_n(t)\beta + U_n} \quad (4.1)$$

Here  $\lambda(t)$  is the *baseline hazard function* that characterizes the basic shape of the hazard function and is common across all individuals. Let  $\Lambda(t) = \int_0^t \lambda(s) ds$  denote the *integrated baseline hazard*.

The expression  $\exp\{X_n(t)\beta\}$  is a customary function used to characterize the effect of observed covariates  $X_n(t)$  on the hazard up to a finite set of unknown regression coefficients  $\beta$ . The effect of the unobserved heterogeneity  $U$  enters multiplicatively via the exponential functional form. In this form  $U$  has a natural interpretation as an omitted variable.

*Assumption 2 [Heterogeneity].* The random variable  $U_n$  is independent of  $X_n$ , satisfying a normalization condition  $E[U_n] = \int_{\Omega} u dG(u) = 0$ , where  $\Omega$  is the distribution support of  $U_n$ .<sup>5</sup> This set-up allows unobserved heterogeneity to be absent. In that case, the distribution  $G(u)$  will be degenerate with a unit point mass at 0, that is,  $P(U=0) = 1$ .

*Assumption 3 [Grouping].* The data on the duration variable  $T$  is grouped into intervals bounded by integers. For each individual in the sample, instead of directly observing  $T_n$  in continuous time, we observe a positive integer  $K_n$  such that it is known that either  $T_n$  falls in the interval  $(K_n, K_n+1)$  (for a completed spell), or  $T_n$  is at least  $K_n$  (for a right-hand censored spell).

*Assumption 4 [Covariates].* The covariate vector  $X(t)$

- (a) is weakly exogenous to the model parameters, and
- (b) is either time-invariant or piecewise constant within each time interval.

In the most general case, neither the baseline hazard function  $\lambda(t)$  nor the heterogeneity distribution  $G(u)$  is specified. The only parametric part of the model is the observed covariates effect  $\beta$ .

---

5. Thus, we adopt the customary assumption in the duration modelling literature that the unobserved heterogeneity is independent from the observed heterogeneity. This assumption on the one hand is more general than assuming that heterogeneity is absent and on the other hand is a restriction in itself. Because we work with several spell-specific regressors and several person-spell-year-specific time-varying variables, modelling dependency between observed regressors and unobserved heterogeneity, as suggested by one referee, would make the model overly parameterized.

*(B) Likelihood Function for A Simple Case*

We now discuss the likelihood-based inference of the above model in the case of a sample of grouped duration data. To ease exposition we first take the simple case where

- (i) There is no unobserved heterogeneity, that is,  $P(U_n = 0) = 1$ .
- (ii) All the covariates are time-invariant that is  $X_n(t) = X_n$  for all  $t$

In this simple set-up the relevant hazard function is

$$h(t|X_n(t), U) = h(t|X_n) = \lambda(t)e^{X_n\beta}. \quad (4.2)$$

The corresponding survivor function,  $S(t|X_n) = P(T_n > t | X_n)$  can be easily derived, using the proportional hazard assumption,

$$S(t|X_n) = \exp\{-\Lambda(t)e^{X_n\beta}\}. \quad (4.3)$$

The contribution to the likelihood of an observation with a completed spell is the probability that  $T_n \in [K_n, K_n + 1)$  conditional on  $X_n$ :

$$\begin{aligned} P(T_n \in [K_n, K_n + 1) | X_n) &= S(K_n | X_n) - S(K_n + 1 | X_n) \\ &= \exp\{-\Lambda(K_n)e^{X_n\beta}\} - \exp\{-\Lambda(K_n + 1)e^{X_n\beta}\} \end{aligned} \quad (4.4)$$

Similarly, the contribution to the likelihood function of the right censored spell is the probability that  $T_n \in [K_n, \infty)$  conditional on  $X_n$ :

$$P(T_n \in [K_n, \infty) | X_n) = S(K_n | X_n) = \exp\{-\Lambda(K_n)e^{X_n\beta}\}. \quad (4.5)$$

Clearly under data grouping A4, the sample likelihood function depends on the unknown baseline hazard only through the discrete values of the integrated baseline hazard function evaluated at integers - distinct values of  $\{K_n\}$ 's.

The likelihood function for the entire sample will be the product of terms either (4.4) or (4.5), depending upon the censoring condition for each individual in the sample.

Suppose there are  $m$  distinct integer values represented in the entire sample ( $n = 1, 2, \dots, N$ ) and, without loss of generality, suppose the set of integers are  $\{1, 2, 3 \dots m\}$ .

$$\text{Let } \kappa_0 = \Lambda(0) = 0, \kappa_1 = \Lambda(1), \dots, \kappa_m = \Lambda(m).$$

The sample log likelihood function can then be expressed as a function of the regression coefficients  $\beta$  and the baseline-hazard parameters  $\kappa = (\kappa_1, \kappa_2, \dots, \kappa_m)$ . The



maximum likelihood estimator for the extended parameter vector  $(\beta', \kappa_1, \kappa_2, \dots, \kappa_m)'$  maximizes the sample log likelihood subject to  $0 \leq \kappa_1 \leq \kappa_2 \leq \dots \leq \kappa_m \leq \infty$ .

To get around these inequality constraints, it is customary to adopt an alternative parametrization. Let  $\gamma_1 = \log(\kappa_1)$  and  $\gamma_j = \log(\kappa_j - \kappa_{j-1})$  for  $j = 2, 3, \dots, m$ . With the new parametrization the log likelihood function can be maximized in terms of  $\delta = (\beta', \gamma_1, \dots, \gamma_m)'$ .

(C) *Incorporating Heterogeneity Terms*

In the presence of unobserved heterogeneity, since  $U$  is not observed, the contribution to the likelihood of an observation with a completed spell is of the following integral form:

$$\begin{aligned} P(T_n \in [K_n, K_n + 1) | X_n) &= E_G[S(K_n | U)] - E_G[S(K_n + 1 | X_n, U)] \\ &= \int_{\Omega} \exp(-e^{X_n \beta} \Lambda(K_n) e^u) dG(u) - \int_{\Omega} \exp(-e^{X_n \beta} \Lambda(K_n + 1) e^u) dG(u) \end{aligned} \quad (4.6)$$

There are three alternatives proposed in the literature in dealing with this integration problem:

- (i) Make a transformation of the random variable from  $U$  to  $W = e^U$ . Assume  $L(w)$  to be the distribution function of  $W$ . In terms of  $W$ , (4.6) becomes

$$\begin{aligned} P(T_n \in [K_n, K_n + 1) | X_n) &= E_L[S(K_n | X_n, W)] - E_L[S(K_n + 1 | X_n, W)] \\ &= \int \exp(-e^{X_n \beta} \Lambda(K_n) w) dL(w) - \int \exp(-e^{X_n \beta} \Lambda(K_n + 1) w) dL(w) \end{aligned} \quad (4.7)$$

Each of the two integrals is exactly the moment generating function (m.g.f.) of the distribution function  $L$ . Due to this m.g.f. representation, parametric distributions that have tractable analytical m.g.f.'s are convenient choices for the heterogeneity distribution. A particular example is the (unit-mean) gamma density for  $W$ :

$$dL(w) / dw = \psi(w; \theta) = \frac{\theta^\theta}{\Gamma(\theta)} w^{\theta w - 1} e^{-\theta}, \quad w \geq 0 \quad (4.8)$$

whose m.g.f. is  $m_L(t) = (1 - t/\theta)^{-\theta}$ . Adopting this gamma distribution, the sample log likelihood function has analytical form in terms of  $(\delta, \theta)$ . Maximization of this likeli-

hood function results in consistent and asymptotically normal estimates for the extended parameter vector  $(\delta, \theta)$ .

- (ii) Assume  $U$  has a discrete distribution with finite distribution support. That is, there is a set of  $J$  values  $(u_1, u_2, \dots, u_J)$  with corresponding probability mass  $(p_1, p_2, \dots, p_J)$ . Adopting this discrete distribution, the integrations in (4.6) become simple weighted averages. The sample log likelihood function also has an analytical form in terms of  $(\delta, u_1, u_2, \dots, u_J, p_1, p_2, \dots, p_J)$ . Maximization of this likelihood function, under restrictions that all the  $p$ 's are non-negative and sum to 1, and that  $U$  is mean zero, results in consistent estimates for the extended parameter vector with  $2 \cdot J - 2$  additional free parameters.<sup>6</sup>
- (iii) In the case that the distribution function for  $U$  makes the integrals in (4.6) non-analytical, one can follow the maximum simulated likelihood method via simulation of the heterogeneity distribution function.

In the empirical analysis of the present paper we take alternative (ii), of a non-parametric discrete distribution function in terms of  $U$  with  $J = 3$ .

#### (D) Treating Time-Varying Covariates

Because all the covariates are assumed to have piece-wise constant values conforming to intervals bounded by integers, the survivor function (4.3) becomes, for  $t = k$ , an integer

$$S(k|X_n(t)) = \exp\left\{-\sum_{j=1}^k \phi_{jn}(\kappa_j - \kappa_{j-1})\right\}. \quad (4.9)$$

Where  $\phi_{jn}$  is the constant value  $\exp\{X_n(t)\beta\}$  for  $t$  between  $j-1$  and  $j$ ; and  $\kappa_j$  is, as before, the integrated based hazard  $\Lambda(j)$  evaluated at integer  $j$ . With this notation, extension of (4.4) and (4.5) is easily done.

## V. The Data

The 0.5% Danish Longitudinal Sample, collected by Statistics Denmark, is the data source for this study. This is a 0.5% representative longitudinal sample of the 18-65 population in the years 1980-1995. This is not a survey-based data, but is drawn instead from administrative registers. New entrants are added each year after 1980 in order to maintain the representative nature of the data.

---

6. Note that modelling unobserved heterogeneity in this manner, i.e. assuming that it follows a distribution with discrete support, is a common practice after the seminal work of Heckman and Singer (1984). As one of the referees correctly pointed out, asymptotic normality of the MLE under this treatment has not been formally proved. So, caution should be used when interpreting the T-ratios for these parameters. We report only standard errors.

For the purposes of this paper, we extract and pool data for all 18-20 year-olds, who graduated from high school in each year starting from 1980 and up to and including 1988.<sup>7</sup> These persons are followed from the year they graduate up to the last year of the sample period, 1995. An individual is considered a high school graduate if the highest degree attained in each year 1980-1988 is either a high school (*gymnasium*) degree, a high school equivalent degree (*HF*), or a degree granted from a technical high school, (*HTX*). Our sample size consists of 693 high school graduates.

The extracted sample consists of 901 spells of higher education taken by 693 high school graduates who graduated sometime in the period 1980-1988.<sup>8</sup> This works out to a total of 3,064 person-year observations. Of these 901 spells which on average have a duration of 3.401 periods, 481 spells are completed by the survey end (1995) with an average duration of 3.769 periods, 158 are censored with average duration of 4.475 periods and 262 are incomplete or drop-out spells with average duration of 2.076 periods. That is, on average most drop-outs occur just after around 2 years of higher study.

Sample size limitations preclude disaggregation of the sample by type of program (short, medium, long) but as different programs can be expected to have different average completion times, as the next best solution, program-specific baseline hazards are allowed, that is short, medium and long programs are assumed to have different hazard shapes, although the effect of the covariates are still assumed to be the same.

Covariates that are hypothesized to significantly affect durations are background variables, previous high school type and labour market variables. Among the background variables are gender, living arrangements, that is, whether the individual is cohabiting or not, and the presence of small children aged 0-2 years in the household. Information on enrolment status in the year after graduating from high school is also available, allowing a test of whether taking a sabbatical year affects time to subsequent education completion (BREAK). We also include some program-specific information, for example for whether or not the spell is the first educational program (FP) and the time in years between programs (TBP) and its square in order to capture the effect of breaks taken in between educational programs and the effect of switching programs including non-linearities.<sup>9</sup> Financial aid and specific institutional details of the Danish

---

7. Note: The normal age of graduation from high school in Denmark is 19 years.

8. By higher education is meant either short-cycle, medium cycle or master's degree/Ph.D. courses, that is, ISCED levels 5-7. Short cycle courses, which are typically non-university educational courses, include courses at the middle technician's level in areas such as food, construction, clothing, electronics, medical technology etc. Medium cycle higher-education courses (university equivalent) are qualifying courses in areas such as nursing, teaching, journalism and include bachelor level educational programs as well. Long-cycle tertiary courses include traditional masters' and Ph.D. degree courses. Thus, by this definition, after-school vocational/apprenticeship courses are not considered higher education.

9. Other studies e.g. Albæk (2001) have shown a non-linear effect of time between high school and university on the probability of completing the first of studies in the economic program at the University of Copenhagen.

educational system could potentially also affect time to degree completion although there is unlikely to exist much variation in the universal financial aid packages (*SU*) available to Danish students. Two variables capture financial aid generosity: *GRANT*, which is the individual's (annual) grant entitlement, obtained from *SU* publications, which varies by age (threshold 18), residential arrangements (living with parents or not) and year, and *WINDOW*, an indicator of being in the eligible window to receive financial aid, which for most years and for most individuals in our sample, is a period of 70 months.

With regard to student ability variables, as student test score information was unavailable in the 0.5% sample, we proxy student ability by type of high school degree obtained, i.e. a regular high school degree, a high school equivalent degree or a degree from a technical high school. Ideally, we would like to have for example verbal and math ability variables which previous studies have shown to be highly correlated with performance in higher-educational programs, see for example Albæk (2002), for Denmark, but these data are unavailable in the 0.5% sample. Of course, a more fundamental control for ability is made by allowing for the existence of unobserved heterogeneity using a three-point discrete distribution.

Finally, labour market variables that are assumed to significantly affect education completion times are the individual's real annual labour market earnings in kroner, real average (log) wages of workers who have completed the same degree and average weekly unemployment degrees of similar degree holders.<sup>10</sup> While information on the individual's own labour market earnings is available in the 0.5% sample data, the last two are aggregates created from a 2% sample of the population (Danish IDA data), for each type of degree (short, medium, long), for each year in the sample period which are merged to our estimation sample by educational type and year. In addition, cohort size effects are captured by year of high school graduation variable, which varies by definition from 1980-1988.

Background variables such as gender, type of high school degree, year of high school graduation and taking a sabbatical year after high school are time invariant. The time-varying covariates are cohabitation status, presence of small children, the student's real annual labour market earnings, real average wage of same degree holders, average unemployment degree of same degree holders, type of education (short, medium or long), time between programs, first program or not and the financial aid variables.

---

10. In Denmark, unemployment is typically recorded as a degree (0-1000) indicating the fraction of the year that the individual has been unemployed.

Table 1. Sample Descriptive Statistics.

Variable	Mean	Std. Deviation	Min	Max
<i>1. Individual background variables</i>				
Female <sup>a</sup>	0.5649	0.4958	0	1
Age <sup>c</sup>	22.7580	2.9274	18	33
Cohabitor Status <sup>c</sup>	0.3398	0.4736	0	1
Kids less than 2 yrs <sup>c</sup>	0.0565	0.2308	0	1
Graduation year-HS <sup>a</sup>	82.8106	2.5466	80	88
Break <sup>a</sup>	0.507	0.500	0	1
<i>2. Student ability variables</i>				
High school <sup>a</sup>	0.8091	0.3930	0	1
Tech. High school <sup>a</sup>	0.0360	0.1863	0	1
<i>3. Educational program-specific variables</i>				
Medium program <sup>b</sup>	0.4528	0.4978	0	1
Long program <sup>b</sup>	0.3552	0.4786	0	1
Time between programs <sup>b</sup>	0.5649	1.2019	0	11
First program <sup>b</sup>	0.6104	0.4877	0	1
SU-Grant/10,000 <sup>b</sup>	2.348	0.605	1.00	3.33
SU-Window <sup>b</sup>	0.794	0.404	0	1
<i>4. Labour market variables</i>				
Annual earnings/10,000 <sup>c</sup>	3.2397	3.5747	0	25.0511
Expected wage <sup>c</sup>	4.3858	0.1676	3.9660	4.6068
Expected unempl./100 <sup>c</sup>	0.3153	0.0711	0.1735	0.4724

<sup>a</sup>Person-specific, N=693; <sup>b</sup>Spell-specific, N=901; <sup>c</sup>Year-specific, N=3064.

## VI. Descriptive Statistics

Table 1 provides descriptive statistics at the person-specific, spell-specific and year-specific levels.

A little more than half the sample is female. Per sample construction, age is between 18 and 33 and the average age of individuals is about 23. About a third of the sample is cohabiting and 6% have small children in the 0-2 age-group. Approximately 51% take a sabbatical year immediately following high school graduation. These means are based on all person-year observations. The vast majority (81%) are graduates from the regular high school (*gymnasium*) and 4% from technical high schools, implying that 15% hold a high school equivalent degree. In terms of type of higher-educational program, 45% of observed spells are medium length programs, 35% are long educational programs and the rest are short higher educational courses. The average time between spells is about 6-7 months, and 61% of the observed spells are the first program for the individual. The average grant per educational spell is 23,480 DKK and about 80% of spells fall within the 70 months window. Real annual labour

*Table 2. Distribution of Observed Duration (901 Spells in Total).*

Program Type					
	Hazard	No. Completed	No. Censored	No. Total	No. at Risk
<b>Short</b>					
1	0.0289	5	37	42	173
2	0.1603	21	23	44	131
3	0.7586	66	6	72	87
4	0.9333	14	1	15	15
5	–	0	0	0	0
6	–	0	0	0	0
7	–	0	0	0	0
8	–	0	0	0	0
<b>Medium</b>					
	Hazard	No. Completed	No. Censored	No. Total	No. at Risk
1	0.0294	12	60	72	408
2	0.0536	18	43	61	336
3	0.3309	91	27	118	275
4	0.6369	100	19	119	157
5	0.5263	20	6	26	38
6	0.1667	2	4	6	12
7	0.1667	1	3	4	6
8	0.5000	1	1	2	2
<b>Long</b>					
	Hazard	No. Completed	No. Censored	No. Total	No. at Risk
1	0.0125	4	48	52	320
2	0.0336	9	34	43	268
3	0.0933	21	22	43	225
4	0.0714	13	12	25	182
5	0.1146	18	6	24	157
6	0.2406	32	14	46	133
7	0.2184	19	16	35	87
8	0.2692	14	38	52	52

market earnings (1990 prices) are roughly 32,000 DKK for the average student. Expected wage (average wage of same degree holders) is about 80 DKK per hour and expected unemployment degree (average unemployment degree of same degree holders) is 31.5.

Table 2 shows the distribution of observed durations by program type. The information presented in this table is used to calculate the Kaplan-Meier estimates of the hazard rates of completion, which are shown graphically in Figure 1. Note that the maximum observed duration of short programs is 4 periods, while the maximum observed duration for medium and long programs is a spell length of 8.

According to Figure 2, non-parametric Kaplan-Meier estimates of hazard rates are shown separately for short, medium and long programs. The hazard rate of completion rises sharply between 2 and 3 years in the program in the case of short programs i.e. as expected, short programs finish fast. For medium programs, the hazard rates shows a

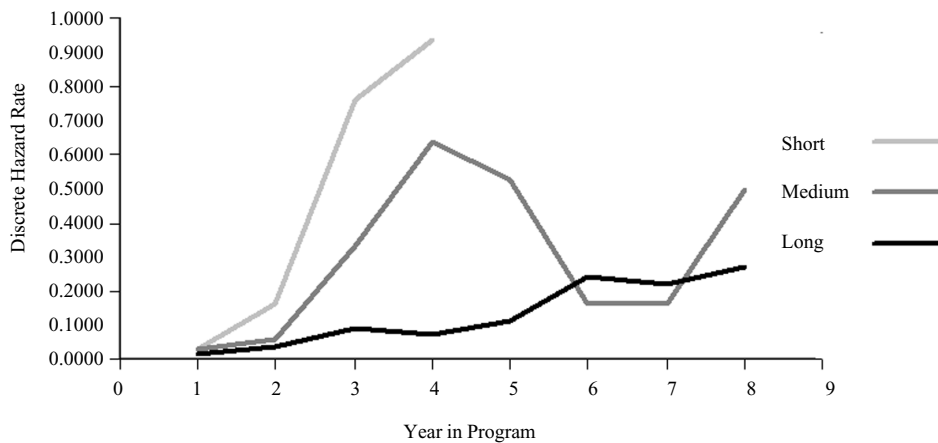


Figure 2. Kaplan-Meier Estimates of Hazard Rate By Program Type.

clear peak at 4 years, and a smaller peak at 8 (although the number of observations are low in the range 7-8 years) while for long programs, the exit rate increases at 6 years and stays constant at this higher-level thereafter.

## VII. Estimation Results

Given the large differences in Kaplan-Meier hazard rates among the three types of programs, in our actual model specification, we assume short, medium, and long programs have different program-specific baseline hazard functions. Table 3 presents the determinants of the completion hazard of four different model specifications. In Model 1, we consider the effect of student background characteristics, previous high school type, program-specific factors and labour market variables.<sup>11</sup> In Model 2, we omit previous high school type as this turns out to be a poor proxy for ability in (1) and add instead the squared term to time-between programs. In Model 3, we further omit cohabitation status and effect of small children, as these are potentially endogenous, because it may be a selected group of students who choose to cohabit or have children. Model 3 also relaxes the assumption that gender differences are captured only through a shift in the intercept term and instead allows for gender-specific interactions in all the covariates. Finally, in Model 4, we retain the gender-specific interactions and

11. In a previous version, we tried two different ways of capturing expected labour market conditions: first, by averaging over all employed workers in the IDA 2% extract who have the same educational type as the individual in any year, and second, by taking average of wages and unemployment degree of workers with the same degree type as the individual in a given year but with less than 5 years accumulated tenure. The second measure fits the data better for all models estimated and is the measure that is applied here, suggesting that mainly early career prospects affect degree completion times.

Table 3. Estimation Results.

	Model 1				Model 2				Model 3				Model 4				
	W/out H		With H		W/out H		With H		W/out H		With H		W/out H		With H		
	Para	S.D	Para	S.D	Para	S.D	Para	S.D	Para	S.D	Para	S.D	Para	S.D	Para	S.D	
<i>I. Covariates</i>																	
Woman	-0,119	0,089	-0,134	0,111	-0,101	0,089	-0,233	0,139	-	-	-	-	-	-	-	-	-
hsyr	-0,137	0,007	-0,160	0,070	-0,104	0,062	-0,099	0,068	-0,009	0,067	-0,009	0,092	-0,015	0,069	-0,015	0,058	-
Break	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
hsdeg	-0,452	0,118	-0,529	0,153	-	-	-	-	-	-	-	-	-	-	-	-	-
hstek	-0,328	0,260	-0,348	0,317	-	-	-	-	-	-	-	-	-	-	-	-	-
age	-0,007	0,026	-0,010	0,067	0,031	0,062	-0,009	0,078	0,083	0,075	0,083	0,095	0,070	0,068	0,070	0,064	-
Cyr	0,027	0,007	0,044	0,068	-0,004	0,065	0,027	0,077	-0,093	0,069	-0,093	0,098	-0,101	0,066	-0,101	0,057	-
tbp	-0,439	0,102	-0,506	0,132	-0,050	0,298	-0,938	0,382	0,544	0,579	0,544	0,624	0,502	0,583	0,502	0,619	-
tbp_sq	-	-	-	-	-0,088	0,069	0,041	0,076	-0,221	0,142	-0,221	0,154	-0,216	0,145	-0,216	0,154	-
fp	-0,609	0,161	-0,724	0,254	-0,293	0,262	-1,621	0,428	0,082	0,486	0,082	0,521	0,145	0,496	0,145	0,528	-
cohab	0,142	0,089	0,142	0,107	0,140	0,091	0,176	0,117	-	-	-	-	-	-	-	-	-
kidst2	-0,396	0,173	-0,457	0,210	-0,360	0,183	-0,536	0,235	-	-	-	-	-	-	-	-	-
estwage2	3,981	0,450	4,116	0,623	4,037	1,067	4,428	1,893	4,062	0,921	4,062	1,607	3,349	1,342	3,349	0,671	-
uem2/100	2,010	0,828	2,245	1,211	1,870	1,046	2,001	1,381	1,598	1,411	1,598	1,488	1,686	1,453	1,686	1,424	-
erm/10k	-0,073	0,014	-0,086	0,021	-0,075	0,014	-0,109	0,020	-0,053	0,020	-0,053	0,020	-0,052	0,020	-0,052	0,020	-
grant	-	-	-	-	-	-	-	-	-	-	-	-	0,196	0,173	0,196	0,171	-
window	-	-	-	-	-	-	-	-	-	-	-	-	-0,285	0,221	-0,285	0,226	-
hsyr	-	-	-	-	-	-	-	-	-0,175	0,036	-0,175	0,107	-0,140	0,080	-0,140	0,030	-
Break	-	-	-	-	-	-	-	-	-	-	-	-	0,039	0,124	0,039	0,144	-
age	-	-	-	-	-	-	-	-	0,027	0,041	0,027	0,106	0,037	0,082	0,037	0,050	-
Cyr	-	-	-	-	-	-	-	-	0,054	0,041	0,054	0,111	0,025	0,076	0,025	0,047	-
tbp	-	-	-	-	-	-	-	-	-0,259	0,362	-0,259	0,372	-0,488	0,359	-0,488	0,363	-
tbp_sq	-	-	-	-	-	-	-	-	-0,050	0,086	-0,050	0,084	0,000	0,081	0,000	0,078	-
fp	-	-	-	-	-	-	-	-	-0,322	0,313	-0,322	0,328	-0,154	0,319	-0,154	0,328	-

Fortsættes næste side ...



fortisat ...

	Model 1				Model 2				Model 3				Model 4			
	W/out H		With H		W/out H		With H		W/out H		With H		W/out H		With H	
	Para	S.D	Para	S.D	Para	S.D	Para	S.D	Para	S.D	Para	S.D	Para	S.D	Para	S.D
estwage2	-	-	-	-	-	-	-	-	4,599	0,981	4,599	1,689	3,425	1,465	3,425	0,712
uem2/100	-	-	-	-	-	-	-	3,177	1,341	3,177	1,477	3,509	1,443	3,509	1,427	
ern/10k	-	-	-	-	-	-	-	-0,092	0,019	-0,092	0,020	-0,100	0,018	-0,100	0,020	
grant	-	-	-	-	-	-	-	-	-	-	-	0,296	0,225	0,296	0,200	
Window	-	-	-	-	-	-	-	-	-	-	-	-0,746	0,214	-0,746	0,223	
<b>2. Log of Baseline Hazard</b>																
<b>a) For Short Program</b>																
bs1	-10,46	2,03	-10,54	3,91	-12,12	4,54	-16,56	5,07	-13,93	3,12	-13,93	5,93	-9,84	5,78	-9,84	3,15
bs2	-8,66	2,01	-8,68	3,90	-10,35	4,54	-14,57	5,08	-12,11	3,09	-12,11	5,91	-8,00	5,72	-8,00	3,12
bs3	-6,53	1,99	-6,26	3,94	-8,21	4,54	-12,02	5,03	-10,03	3,08	-10,03	5,93	-5,91	5,71	-5,91	3,12
bs4	-5,60	2,03	-5,07	4,02	-7,32	4,53	10,84	90,69	-9,24	3,07	-9,24	5,90	-5,41	5,74	-5,41	3,15
<b>b) For Medium Program</b>																
bm1	-10,72	2,10	-10,80	3,94	-12,42	4,63	-17,03	5,02	-14,18	3,15	-14,18	6,05	-9,98	5,82	-9,98	3,18
bm2	-10,06	2,10	-10,12	3,95	-11,81	4,64	-16,21	5,03	-13,57	3,16	-13,57	6,06	-9,34	5,82	-9,34	3,17
bm3	-8,09	2,09	-8,03	3,96	-9,83	4,64	-14,13	5,04	-11,61	3,16	-11,61	6,07	-7,41	5,81	-7,41	3,16
bm4	-7,10	2,09	-6,78	4,01	-8,87	4,65	-12,61	5,04	-10,65	3,17	-10,65	6,08	-6,46	5,82	-6,46	3,15
bm5	-7,36	2,11	-6,89	4,06	-9,14	4,67	-11,20	5,06	-10,89	3,16	-10,89	6,11	-6,88	5,81	-6,88	3,14
bm6	-8,55	2,21	-8,01	4,15	-10,35	4,73	-9,76	5,12	-12,09	3,24	-12,09	6,16	-8,02	5,83	-8,02	3,21
bm7	-8,71	2,32	-8,19	4,19	-10,50	4,79	7,75	90,94	-12,22	3,35	-12,22	6,21	-8,48	5,85	-8,48	3,29
bm8	-7,40	2,32	-6,84	4,20	-9,18	4,82	9,18	90,96	-11,05	3,28	-11,05	6,22	-7,23	5,90	-7,23	3,28
<b>c) For Long Program</b>																
bl1	-12,80	2,21	-12,95	4,11	-14,54	4,92	-19,32	5,01	-16,49	3,37	-16,49	6,47	-12,03	6,18	-12,03	3,34
bl2	-11,87	2,19	-12,02	4,10	-13,64	4,91	-18,38	5,06	-15,58	3,38	-15,58	6,48	-11,16	6,17	-11,16	3,30
bl3	-10,81	2,18	-10,90	4,11	-12,59	4,91	-17,26	5,02	-14,52	3,38	-14,52	6,49	-10,11	6,18	-10,11	3,29

fortisettes neste side ...

fortsat ...

	Model 1		Model 2		Model 3		Model 4										
	W/out H Para	With H S.D	W/out H Para	With H S.D	W/out H Para	With H S.D	W/out H Para	With H S.D									
b14	-11,10	2,19	-11,13	4,12	-12,88	4,93	-17,49	5,03	-14,81	3,37	-14,81	6,51	-10,43	6,18	-10,43	3,29	
b15	-10,63	2,19	-10,61	4,13	-12,40	4,94	-16,94	5,03	-14,32	3,40	-14,32	6,53	-9,94	6,16	-9,94	3,27	
b16	-9,83	2,20	-9,72	4,14	-11,61	4,94	-16,01	5,02	-13,53	3,41	-13,53	6,53	-9,19	6,16	-9,19	3,26	
b17	-9,97	2,20	-9,76	4,16	-11,75	4,95	-16,00	5,00	-13,66	3,41	-13,66	6,55	-9,79	6,14	-9,79	3,27	
b18	-9,78	2,21	-9,49	4,19	-11,58	4,97	-15,56	4,97	-13,50	3,40	-13,50	6,58	-9,66	6,16	-9,66	3,26	
<i>3. Heterogeneity Distribution</i>																	
p1			0,70				0,08					0,00				0,00	
p2			0,00				0,15					0,00				0,00	
p3*			0,30				0,77					1,00				1,00	
v1			-0,36	0,40			-21,05	86,01				-2,93	182,68			-3,21	15,65
v2			-2,93	34,60			0,16	7,12				-2,90	174,25			-3,26	16,44
v3*			0,83				2,05					0,00				0,00	
<i>4. Model Fit</i>																	
# of Para			37				36					42				48	
Log Lik			-916,41				-920,50					-917,875				-911,35	
				41				40					46				52
				-915,42				-918,36					-917,875				-911,35

1. p3 and v3 are not freely estimated. They are subject two constraints defined in the paper.

introduce the effect of financial aid variables and sabbatical year. All models are estimated with and without the correction for unobserved individual heterogeneity. In each case, a positive sign on a coefficient indicates that (education) duration decreases, i.e. completion is quicker, while a negative sign indicates that the duration becomes longer.

Results of the first model show that the female indicator is not significant. This is not a surprising result since studies have shown that the gender gap in education no longer exists in countries such as US, Belgium, Spain, Canada and the Scandinavian countries. According to the OECD study, women in the 25-34 year old age group in Denmark even have a slight educational advantage over men (OECD Indicators, 1997). However, this model only allows for a gender difference through a shift of intercept and there could potentially be gender differences in covariate effects, which are not being allowed for here. The later the high school year of graduation, the longer is the completion time, indicating the importance of cohort effects. Graduating from a regular high school (*gymnasium*) is significant and negative, i.e. making duration longer, even after allowing for program-specific baseline hazards. However, graduating from technical high school is not significant. These findings indicate that the individual's type of high school is likely not a good proxy for underlying ability and therefore we drop them in the next model. Age does not seem to be significantly related to completion risk. The later is the calendar year, the lower the completion hazard, which may reflect the effect of changing labour market conditions or increasing generosity of the government-financed universal financial aid package (*SU*) over time. The longer is the time between programs, the lower the risk that a subsequent spell will be completed, suggesting depreciation of studying skills when not put to use. However, if the spell is the first program, it has a lower hazard of completion. Both of these spell-specific variables are significant. Cohabiter status speeds up degree completion but the effect is not significant, while the presence of small children significantly reduces the completion hazard as expected. The labour market variables are all strongly significant: The higher the average wage of workers of similar educational background, the greater is the completion hazard. The higher an individual's labour market earnings while studying, the lower is the hazard of completion. Unexpectedly, unemployment upon completion based on the unemployment experiences of workers holding a similar degree significantly increases completion hazard. How do these results change when we allow for unobserved heterogeneity with 3 discrete points of support? In column (2), model 1, we see that the fit improves slightly (but not significantly so when doing the likelihood ratio test) and although heterogeneity parameters themselves are not statistically significant, probabilities  $p_1$  and  $p_3$  are non-zero indicating some type of unobserved heterogeneity is present. Most of the estimated coefficients have the

same signs and significance, except that calendar year and the expected unemployment degree are no longer significant.

In Model 2, we drop previous high school type and add the squared term to time between programs. The first column without heterogeneity does not fare especially well, and only future expected wages, current earnings and the effect of children show up as being statistically significant. In the second column when unobserved heterogeneity is allowed, time between programs and the effect of first program are significant. In general, Model 2 is worse than Model 1 in terms of fit. Note that for Models 1 and 2, unobserved heterogeneity can be said to be present, because although the heterogeneity distribution parameters ( $v$ 's) are not all statistically significant, the distribution at least is not degenerate, with some variation and non-zero probabilities ( $p$ 's).

In Model 3, we relax the assumption of no gender difference in covariates and allow a full set of gender interactions. The fact that even some of the gender-specific interactions are significant indicates that men and women do have different hazard functions and covariate effects which a single intercept term could not adequately capture. In Model 3, the labour market variables continue to exert a significant influence on time to completion, with higher expected wages speeding up the time to completion and higher current earnings while studying slowing down completion times for both men and women. However, only men experience a speeding up due to higher expected unemployment rates, a result that appears contrary, but may be a result of the confounding effects of different types of higher educational courses that have been aggregated within each program type (short, medium, long) in the present analysis. Calendar year and time between programs are significant for women, but weakly so, but both with a negative sign indicating that completion times have been lengthening over time and are longer, the longer the time that elapses between programs.

When controlling for heterogeneity in column 2, results do not move and the heterogeneity distribution collapses at unit mass at  $P(V=0) = 1$ , i.e. a degenerate distribution. All the labour market factors have the same signs and significance as in column 1. For women, time between programs remains weakly significant.

In the final model, Model 4, we try adding the effect of policy-related variables to the previous model, i.e. financial aid variables (GRANT and WINDOW) and the effect of taking a sabbatical year (BREAK). In general, these factors are not statistically significant, except for men, being within the 70 months window means significantly longer completion times, giving some evidence that financial aid policy does seem to work as intended, but for this group only. The labour market variables continue to exert the same signs and significance as in Model 3, and just as in Model 3, calendar year and time between programs tends to slow down completion for women, and high-school year and time between programs the same for men, but these effects are moderate. In general,

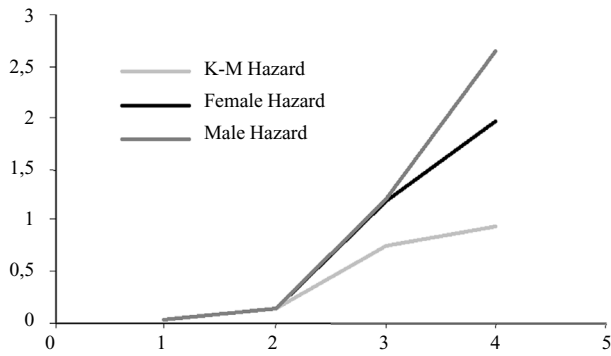


Figure 3. Estimated Hazard Rate Evaluated at Mean Values of Regressors (Short Program Model 4).

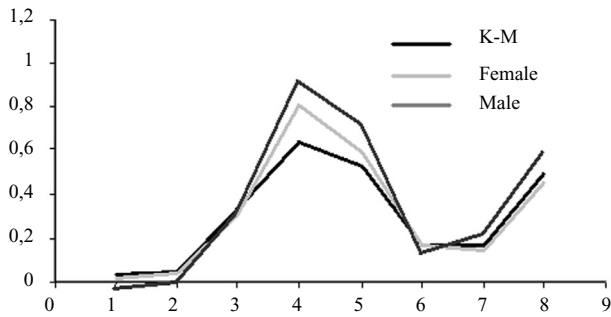


Figure 4. Estimated Hazard Rate Evaluated at Mean Values of Regressors (Medium Program Model 4).

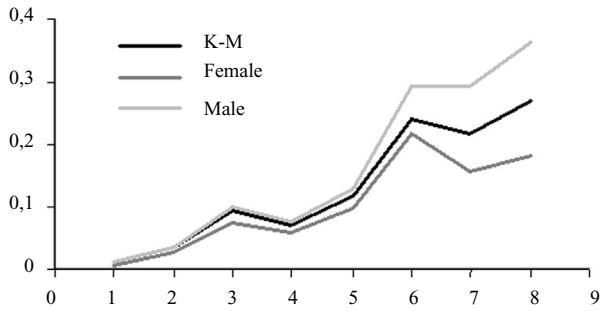


Figure 5. Estimated Hazard Rate Evaluated at Mean Values of Regressors (Long Program Model 4).

Model 4 fits better than Model 3, but here again, there is no unobserved heterogeneity left in the model once gender-specific covariate effects are allowed.

On the basis of the estimates from Model 4 without heterogeneity, we plot baseline hazards at the mean values of regressors separately for each program-specific hazard, separately for males and females. These appear as Figures 3-5.

As is evident from these figures, the shape of the estimated base-line hazards after controlling for covariates closely mimic the non-parametric K-M hazard for each of the three program types, particularly in the ranges where the bulk of the observations lie, the first 2 years for short programs, the first 3 years for medium programs and the first 5 years for long programs. Beyond these ranges, there could be some degree of extrapolation based on few observations. There is also a tendency for male hazards to lie above female hazards, but again, this is most pronounced in the final years of each program type, in which the number of observations are low.

### **VIII. Conclusions**

In this paper we investigate the determinants of the time-to-completion of higher education of a sample of high school graduates and equivalents from the 0.5% Danish Longitudinal Sample. We estimate a semi-parametric hazard model with correction for unobserved heterogeneity. Findings that are robust across a number of different specifications show that completion hazard is significantly affected by student background variables, higher educational spell-specific factors and (especially) effects originating from the labour market. In particular, the expected wage upon completion is strongly significant and positive, i.e. speeding-up degree completion while the higher the individual's own earnings while studying, the lower the completion hazard. Unexpectedly, the higher the expected unemployment degree in a student's given education type, the quicker the time to completion, but this effect is only present for men, and may be a result of the confounding effects of aggregating different higher educational courses within each program type (short, medium, long). Other findings are that completion times have been increasing over time perhaps due to changing societal norms or due to increasing generosity of financial aid not being captured by the measures used in the analysis, and, that students in Denmark take longer to complete the longer the time that elapses between programs. Interestingly, the level of SU entitlement does not have much of an effect and nor does taking a sabbatical year immediately following high school. However, (male) students are faster to complete once SU support is exhausted, which on average is a period of 70 months for most spells in our sample period. Policy implications would point to continuation of incentive-based financial support to students of higher education to encourage degree completion and to reduce the propensity to seek concurrent work, which our results show have significant delaying

impacts on degree completion. In addition, universities should offer intensified counselling to students who experience a lapse of time between programs, as, according to our results, this group is likely to experience long completion times.

#### References

- Abedi, J. and E. Benkin. 1987. The Effects of Students, Academic and Financial and Demographic Variables on Time to Doctorate. *Research in Higher Education*, 17(1), s. 3-14.
- Albæk, K. 2001. Hvem består på politstudiet, *Nationaløkonomisk Tidsskrift*, nr. 139, s. 208-22.
- Albæk, K. 2002. Incitamentsstrukturen i gymnasiet i relation til de videregående uddannelser, *Samfundsøkonomen*, nr. 5, s. 20-27.
- An, M. Y. 2000. A Semiparametric Distribution for Willingness to Pay and Statistical Inference with Dichotomous Choice Contingent Valuation Data, *American Journal of Agricultural Economics*, 82 (August 2000), s. 487-500.
- Ehrenberg, R. G., and P. G. Mavros. 1995. Do Doctoral Students' Financial Support Patterns Affect Their Times-to-Degree and Completion Probabilities?, *Journal of Human Resources*, 30(3), s. 581-609.
- Booth, A. L. and S. E. Satchell. 1996. British PhD Completion Rates: Some Evidence from the 1980s. *Higher Education Review*, 28(2).
- Bowen, W. G., and N. L. Rudenstine. 1992. *In Pursuit of the Ph.D.* Princeton, N.J.: Princeton University Press.
- Card, D. 2001. The Causal Effect of Education on Earnings, in *Handbook of Labor Economics*, ed. Orley Ashenfelter and David Card (New York: North-Holland).
- Christensen, J. J. and N. Westergård-Nielsen. 1999. The Educational Structure and the System of Financial Support in Denmark. Aarhus School of Business, manuscript.
- Danish Council of Economic Advisors Report, Spring 1995, Det Økonomiske Råd Formandskabet.
- Dansk Arbejdsgiverforening. 2004. Arbejdsmarkedsrapport. Uddannelse. DA Forlag.
- Heckman, J. J., and B. Singer. 1984. A Method for Reducing the Impact of Distributional Assumptions in Econometric Models of Duration Data, *Econometrica*, 52, s. 271-320.
- SU Støtte og SU gæld, 1997, SUS nr. 5, august 1998.
- Tuckman, H., S. Coyle and Y. Bae. 1990. On The Time to the Doctorate. Washington D.C., National Academy Press.
- Van Ours, J. C. and G. Ridder. 2003. Fast track or failure: A study of the graduation and dropout rates of Ph.D. students in economics, *Economics of Education Review*, Vol. 22, 157-66.

# Debat og kommentarer

## Keynes-inspireret makroøkonomisk teori

Jesper Jespersen

Institut for Samfundsvidenskab, Roskilde Universitetscenter, E-mail: [jesperj@ruc.dk](mailto:jesperj@ruc.dk)

### En kort introduktion

#### a. Metodologiske forskelle

En stabil makroøkonomisk udvikling er bredt accepteret som en nødvendig forudsætning for at sikre fortsat velstand og velfærd i et demokratisk samfund baseret på markedsøkonomi, privat ejendomsret og et stort antal lønmodtagere. Alene af den grund er det vigtigt at forstå de samfundsmæssige processer, der er bestemmende for den makroøkonomiske udvikling i beskæftigelse, arbejdsløshed, betalingsbalance, inflation og offentlige budgetter.

Formålet med denne artikel er at give en del af den teoriehistoriske baggrund for den stadig aktuelle metodologiske diskussion, der udspiller sig selv blandt makroøkonome med hensyn til at analysere makroøkonomisk stabilitet. En væsentlig kilde til forståelse af denne på mange måder fundamentale makroøkonomiske uenighed ligger i de ontologiske antagelser, der som udgangspunkt benyttes til at karakterisere den makroøkonomiske virkelighed. Inden for den nyere litteratur anvendes ligefrem metaforen »Einsteinian versus Newtonian macroeconomics« jf. Togati (1998) til at karakterisere en væsentlig forskel. Heri gøres det metodologiske udgangspunkt for den makroøkonomiske analyse til *den* afgørende forskel i forståelsen af det makroøkonomiske systems funktion, hvilket har afgørende betydning for analysen af den makroøkonomiske virkelighed.

Inspiration til denne erkendelse af metodologiens betydning har jeg bl.a. fået fra læsningen af Keynes egen teoretiske udvikling, navnlig forordet til *The General Theory of Employment, Interest and Money* (GT) udgivet i 1936 er instrumentelt, jfr. Jespersen (2002). Frem til begyndelsen af 1930erne var Keynes metodologisk stadig forankret i den (i Cambridge) dominerende Marshallianske metodemæssige tradition. Det er først med skiftet *fra* at opfatte det makroøkonomiske system som styret af markedskræfterne frem mod en generel ligevægt (billedligt talt, som tyngdekraften styrer

---

Denne artikel er udarbejdet under mit ophold ved University of Cambridge, Churchill College med støtte fra Carlsbergfondet.



det astronomiske system), *til* en opfattelse af det makroøkonomiske system som værende åbent dvs. underbestemt, at Keynes metodemæssigt bryder nye veje. Forståelsen af et åbnet makroøkonomiske systems funktion stiller krav om en helt ny epistemologi (teori og metode), hvilket har været en del af den makroteoretiske diskussion lige siden.<sup>1</sup>

*b. Makroteori som generaliseret mikroteori*

Makroøkonomisk teori blev først etableret som en egentlig selvstændigt forskningsfelt i mellemkrigstiden. Makroøkonomiske modeller, der havde det primære formål at analysere udviklingen, navnlig i arbejdsløshed (og inflation), blev i stigende grad en del af den økonomiske forskning i denne periode. Udgangspunktet for denne forskning var fortsat det eksisterende neoklassiske fundament, der satte fokus på markedsmæssige processer. Her var udbud, efterspørgsel og markedspris på de enkelte, ofte adskilte, markeder i centrum for den makroøkonomiske teori. I den angelsaksiske tradition,<sup>2</sup> der var blevet grundlagt mere end 100 år tidligere af Smith og Ricardo, og videreført af bl.a. Alfred Marshall og Arthur Pigou (i Cambridge) forblev denne markedsmæssige (mikroøkonomiske) analyseform grundlaget for den makroøkonomiske teori, der blev anvendt som basis for den økonomisk politiske rådgivning vedrørende afhjælpning af bl.a. arbejdsløshed, der om noget var datidens svøbe.

Mellemskrigstidens makroøkonomiske teori bestod af en række ikke fuldt integrerede – og derfor partielle – analyser, af de faktorer der påvirker udviklingen i de enkelte makroøkonomiske variable. Hvert »makromarked«, dvs. arbejds-, kapital-, vare- og valutamarked blev analyseret uden eksplicit hensyntagen til den gensidige interaktion, dvs. som om det var et »mikro-marked«. Denne fremgangsmåde blev metodemæssigt legitimeret gennem antagelsen om, at »alt andet forblev lige« – den ofte benyttede *ceteris paribus* forudsætning. Den makroøkonomiske teori kom herved til at bestå af partielle (dvs. uafhængige) analyser af de enkelte markeder, herunder arbejds-, vare- og kreditmarkederne. Tilsvarende blev det absolutte prisniveau bestemt uafhængigt af den realøkonomiske udvikling, idet den monetære kvantitetsligning (*the quantitative*

1. En diskussion, der – som det fremgår af den nys udkomne lærebog (Sørensen og Whitta-Jacobsen (2005)) – bestemt ikke er bragt til en endelig afklaring.

2. Jeg skriver den angelsaksiske tradition, fordi der på Kontinentet allerede i slutningen af det 19. århundrede var blevet udarbejdet en teoretisk ramme for, hvorledes analysen af de enkelte markeder kunne sammenknyttes til én samlet model. Dette arbejde blev indledt af franskmænden Léon Walras og videreført af italieneren Vilfredo Pareto, der begge var tilknyttet universitetet i Lausanne. Dette arbejde påkaldte sig dog ikke større opmærksomhed før efter 2. verdenskrig, hvor arbejdet med opstilling af en generel ligevægtsteori (generaliseret mikroteori), der senere skulle blive prototypen for neoklassisk makroteori, var blevet påbegyndt i den angelsaksiske verden. Det er således et karakteristisk udtryk for denne sene sammensmeltning af de to forskningsprogrammer, at Walras' bog fra 1874 først blev oversat til engelsk i 1954.

*theory of money*) antoges udtømmende at beskrive den snævre årsagsmæssige sammenhæng mellem penge og priser.

Dette er meget kort fortalt den spinkle makroøkonomiske forståelsesramme, som var til rådighed, da uvejret for alvor brød løs efter krakket på børsen i Wall Street i 1929. Det findes bl.a. beskrevet i A.C. Pigou's bog fra 1933 *The Theory of Unemployment*. Som en konsekvens af den partielle, mikroøkonomiske markedsanalyse måtte Pigou nødvendigvis nå til den konklusion, at den vedvarende høje arbejdsløshed skyldtes manglende fleksibilitet på selve arbejdsmarkedet, dvs. træghed i tilpasningen af reallønnen. Når ubalancen på arbejdsmarkedet forblev uforandret stor i gennem en årrække, måtte det skyldes, at reallønnen ikke faldt (tilstrækkeligt), hvilket blev demonstreret i en arbejdsmarkedsmodel. Her var det ikke den svigtende britiske eksport, men derimod erhvervslivets svigtende indtjening, som var årsagen til den for lille efterspørgsel efter arbejdskraft. En konklusion der som nævnt var udledt fra en partiel, mikroøkonomisk funderet arbejdsmarkedsanalyse med beskednen empirisk afprøvning.

Inden for denne forskningstradition kom, som det vil blive påvist nedenfor, udgivelsen af *The General Theory of Employment, Interest and Money* i 1936 ikke til at spille nogen afgørende rolle, som ikke mindst Keynes selv havde håbet på. De specielle keynesianske begreber, navnlig forbrugsfunktion, likviditetspræference og effektiv efterspørgsel, blev dog i løbet af 1950erne indarbejdet som en del af den neoklassiske makroteori. I første omgang i form af den såkaldte »neoklassiske syntese«, etableret omkring IS-LM modellen, hvor pris- og lønstivhed kunne forklare arbejdsløshed på kort sigt, her var navnlig Modigliani (1944) instrumental, se også Modigliani (1999). Efterfølgende blev den neoklassiske syntese indoptaget i de generelle ligevægtsmodeller, der benyttes til at forklare den strukturelt betingede makroøkonomiske udvikling, som det f.eks. også kendes fra den danske DREAM-model, Pedersen og Rasmussen (2000).

### *c. Keynes-inspireret teori og metode*

Udgivelsen af John Maynard Keynes' hovedværk *The General Theory of Employment, Interest and Money* (GT) gjorde dog en markant forskel uden for den neoklassiske forskningstradition. Her præsenteredes for første gang en egentlig *makroøkonomisk metode*, der udgjorde fundamentet for en grundlæggende ny forståelse af det makroøkonomiske kredsløb. Metodemæssigt stillede Keynes en række krav, der skulle være opfyldt for at analysen overhovedet kunne karakteriseres som værende distinkt makroøkonomisk:

- For det første skal analysen omfatte *the economy as a whole* – de enkelte markeder kan ikke analyseres uafhængigt af hinanden; for »alt andet lige« forudsætninger

gælder ikke på makroplan. Som noget nyt lykkedes det Keynes at integrere de reale og de finansielle sektorer, hvilket han understregede i titlen, *a monetary theory of production*, på en af de artikler, der pegede frem mod GT.

- For det andet skal den makroøkonomiske teori afspejle det forhold, at forventningsdannelsen ifølge sagens (eller rettere samfundsøkonomiens) natur må være karakteriseret ved (delvis) *usikkerhed* med hensyn til, hvad fremtiden vil bringe. Keynes har den ofte, med rette, fremførte sentens, at vedrørende fremtiden gælder det, *that we simply don't know*. Dette udelukker, at den (mikro)økonomiske teori, der forudsætter fuld information om fremtiden, kan udgøre det teoretiske grundlag for Keynes' makroteori. De makroøkonomiske adfærdsrelationer må udledes under hensyntagen til denne fundamentale usikkerhed og til de institutionelle forhold der i øvrigt eksisterer i den virkelige økonomi.
- Som en konsekvens af navnlig punkt 2 kan det *ikke* a priori antages, at det økonomiske system *as a whole er selvregulerende*, endsige at det makroøkonomiske system blot ved hjælp af de iboende markeds kræfter skulle kunne nå en generel ligevægt bl.a. karakteriseret ved fuld beskæftigelse. Der er på det hidtidige empiriske grundlag intet, som tilsiger, at der overhovedet kan defineres en relevant makroligevægt for det økonomiske system. Hertil kommer, at den makroøkonomiske egendynamik vil kunne stille sig hindrende for, at en sådan makroligevægt overhovedet ville kunne nås – uanset om alle pris- og lønvariable er fuldt fleksible.

Forståelsen af Keynes' makroøkonomisk ontologi fordrer et nyt metodologisk grundlag. For at kunne inddrage fænomenet *fundamental usikkerhed* udvikler Keynes en ny makroøkonomisk analysemetode, der skal gøre det nye analytiske begreb *effective demand* operationel.

*Fundamental usikkerhed* udspringer af to afgørende egenskaber ved den makroøkonomiske virkelighed, der gensidigt forstærker hinanden: for det første det forhold at selv en mikroøkonomisk perfekt fungerende markedsøkonomi ikke kan antages at konvergere mod en generel markedsligevægt – end ikke »in the long run« – af den simple grund, at selv nok så »rationelle« aktører ikke kan have sikker information om fremtiden. Eller sagt med andre ord den økonomiske teori på makroniveau må afspejle denne systemisk betingede uvidenhed, der er et ontologisk grundvilkår, hvis teorien skal kunne forklare virkeligheden. Generelle ligevægtsmodeller har ikke hidtil analytisk inddraget fundamental usikkerhed, hvilket bl.a. forklarer, hvorfor den *effektive efterspørgsel* ikke er indeholdt i f.eks. DREAM-modellen.

Disse metodologiske forskelle mellem generaliseret mikroteori og Keynes' makroteori var en væsentlig del af den makroøkonomiske diskussion, der opstod i kølvandet

på offentliggørelsen af GT – og det har det været lige siden. Det er næppe nogen overdrivelse at tale om »diskussionen, der ikke ville dø«.

I den netop udkomne makroøkonomiske lærebog (Sørensen og Whitta-Jacobsen (2005) skriver forfatterne: *let us state two trivial, but nevertheless important, and highly related points:*

- In a labour market characterized by full competitive wage flexibility, there can be no unemployment, since the nominal wage rate adjusts immediately to generate a real wage that equates the competitive (wage-taking) supply of and demand for labour.
- For unemployment to be possible, there must be factors preventing the (real) wage from adjusting to a level that would equate the competitive supply of labour and the competitive demand for labour (s. 328)

Vedrørende betydningen af penge (og dermed implicit den finansielle sektor) gør forfatterne følgende antagelse:

- Most economists think that in the long run money does not matter for real economic variables (s. 68).

#### **Det er vigtigt at sondre mellem ny-keynesianere og post-keynesianere**

Det er dog nødvendigt indledningsvist at foretage en præcisering – for at undgå den sproglige forvirring, der er opstået i kølvandet på den tvetydige brug af begrebet »keynesiansk økonomi«. Der har i 1980erne og 1990erne udviklet sig den sprogbrug, navnlig i amerikanske lærebøger, at benytte betegnelsen keynesiansk makroteori, om de specielle modeller, hvor de »nominelle priser og lønninger (på kort sigt) er uflexible/stive, og hvor virksomhederne derfor ændrer deres produktion som følge af ændringer i efterspørgslen«, jfr. f.eks. Romer (1996, s.197).<sup>3</sup> Afvigelser i produktion og beskæftigelse fra langsigts-ligevægten opnås her modelmæssigt ved en antagelse om, at virksomhedernes aggregerede udbudskurve på kort sigt kan have en vis realpris-elasticitet, mens udbudet på langt sigt udelukkende vil være bestemt af tekniske/strukturelle forhold. Disse kortsigtede fluktuationer i beskæftigelsen i denne såkaldte Keynes-model skyldes, i følge Romer, at Keynes baserede sin »generelle teori« på den »specielle« forudsætning, at det nominelle lønniveau på kort sigt er fast (rigid), s. 215. Disse lærebogsfremstillinger benytter således begrebet »keynesiansk« om den specielle variant af neoklassisk makroteori, hvor ufrivillig arbejdsløshed er forårsaget af løn- og

---

3. Se f.eks. Romer, 1996/2001, kap. 5, der netop bærer overskriften *Traditional Keynesian Theories of Fluctuations*.

prisstivheder.<sup>4</sup> I denne litteratur benyttes betegnelsen »ny-keynesiansk« som oftest som modstillingen til monetaristisk (eller ny-klassisk) makroteori, hvor der forudsættes fuld pris- og lønflexibilitet og dermed fuld tilpasning til generel ligevægt på kort (eller kortere) sigt. I disse neoklassiske lærebøger kan forskellen mellem ny-keynesianere og monetarister/ny-klassikere sammenfattes således:

- Monetaristerne fraråder makroøkonomiske indgreb. I deres modelverden er det markedskræfterne, der sikrer den makroøkonomiske stabilitet – bortset dog fra centralbankens forpligtigelse til at holde pengemængden og dermed inflationen under kontrol.
- Heroverfor står i lærebogsfremstillingen de såkaldte (ny)keynesianerne, der som en konsekvens af pris- og lønstivhed med varierende styrke konkluderer, at der på kort sigt med fordel kan føres en aktiv makroøkonomisk politik, når økonomien udsættes for eksterne stød til efterspørgslen, f.eks. i form af international afmatning, børskraks og lignende. »Keynesianerne« anbefaler også brug af (midlertidig) finanspolitik, dog under forudsætning af, at den langsigtede finanspolitiske holdbarhed ikke sættes over styr. Risikoen herfor har ført til en anbefaling af et loft over underskuddets størrelse, for herved at forebygge en evt. makroøkonomisk ustabilitet fra tilvækst i den offentlige sektors gæld.

Problemet er, at betegnelsen »keynesiansk« mildt sagt ikke benyttes entydigt i den makroøkonomiske litteratur. Som det vil fremgå af de efterfølgende afsnit, skal der metodemæssigt sondres mellem »gammel-keynesianere«, »ny-keynesianere« og »post-keynesianere«.<sup>5</sup> Lærebogsfremstillingerne er sjældent udtømmende med hensyn til at give en dækkende fremstilling af de forskellige Keynes-inspirerede makroøkonomiske teorier. Navnlig de metodologisk set helt afgørende forskelle mellem disse tre retninger er ofte blevet negligeret. Prædikatet 'keynesiansk' er blot blevet benyttet som en samlebetegnelse for makroøkonomisk teori, der i varierende grad anbefaler en aktiv stabiliseringspolitik i modsætning til nyklassisk (og delvis monetarisisk) teori der har policy ineffectiveness som et karakteristisk fælles træk.<sup>6</sup>

---

4. En terminologi, der også anvendes af Sørensen og Whitta-Jacobsen (2003); derimod er der i lærebogen (2005) på dette terminologiske punkt sket en revision. De tidligere såkaldte 'keynesianske' modeller, hvori konsekvenserne af antagelsen af nominelle pris- og lønstivheder i en ligevægtsmodel analyseres, nu benævnes som »short run« eller »business cycle« modeller. Etiketten er ændret, hvilket – uanset at det analytiske indhold er uændret – ud fra en teoriehistorisk synsvinkel må anses for en klarifikation.

5. Se f.eks. Begg, Dornbusch & Fischers lærebog i *macroeconomics*, hvor den post keynesianske forskningstradition anerkendes dog under prædikatet 'extreme Keynesian'.

6. I Estrup m.fl. (2004) er udviklingen i makroøkonomisk teori sat ind i en bredere, om end kortfattet, teoriehistorisk ramme.

*a. Den neoklassiske syntese – »gammel« keynesianere*

Som nævnt satte Keynes' bog en proces i gang inden for makroøkonomisk teoridannelse. Det var navnlig 'den monetære produktionsteori' og fraværet af selvregulering og dermed etablering af ligevægt med fuld beskæftigelse, der umiddelbart fængede blandt de 'yngre' økonomer. John Hicks' IS-LM diagram kom til at stå som eksponent for den 'fællesmængde', som 'keynesianerne' på daværende tidspunkt i opposition til datidens main stream økonomer kunne enes om måtte være hovedbudskabet i Keynes' nye makroteori. Retfærdighedsvist skal det her tilføjes, at Keynes i hvert fald ikke tog afstand fra Hicks' fremstilling.<sup>7</sup> Dette er for så vidt også ganske forståeligt, da ISLM-modellen tildeler Keynes' nye analytiske begreb *effective demand* en prominent placering. Det er denne 'effektive efterspørgsel efter varer (og tjenesteydelser), der bestemmer størrelsen af produktionen og dermed også (indirekte) beskæftigelsen. Men Hicks' fremstilling lod Keynes' metodologisk i stikken ved at hævde, at både den hidtidige neoklassiske makroteori og Keynes' nye makroteori kunne rummes inden for rammerne af én og samme model. Forskellen kunne, ifølge Hicks, begrænses til spørgsmålet om, hvilke variable der skulle (og kunne) indgå i de makroøkonomiske adfærdsrelationer. Herved blev der i realiteten taget det første skridt i retning af den *neoklassiske syntese* (eller rettere den neoklassiske metodologiske dominans) som kom til at sætte den makroøkonomiske dagsorden de første 30 år af efterkrigstiden.

Franco Modigliani (1944) tog det næste skridt ved at reintrodere det neoklassiske arbejdsmarked og Pigous argument, at kun stivhed i realløn kan begrunde<sup>8</sup> vedvarende arbejdsløshed. Men Modigliani havde (ligesom Pigou) det problem, at den effektive efterspørgsel var blevet en integreret del af den makroøkonomiske teori. Der skulle derfor etableres en sammenhæng mellem (ændringer i lønnen) og den effektive efterspørgsel efter varer (og tjenester). En fleksibel realløn kan ikke i sig selv forklare, hvorledes den effektive efterspørgsel på varemarkedet ændres. Denne brik faldt først logisk på plads ved tilføjelsen af realkasse-effekten (eller Pigou-effekten som den ikke ukarakteristisk også kaldes) i Patinkin (1956).

*Den neoklassiske syntese* havde herved bragt Keynes' makroteori tilbage i 'folden', hvor det makroøkonomiske system er selvregulerende, hvor usikkerhed om fremtiden ikke hindrer tilpasningen, hvor analyse af 'the economy as a whole' kun er relevant på kort sigt, indtil 'hele' økonomien har nået at tilpasse sig. På et punkt forbliver 'gammel-

7. Skidelsky (1992) har en grundig gennemgang af Keynes' korrespondance med Hicks. Warren Young (1987) har skrevet en hel bog om ISLM-modellens tilblivelseshistorie med afsæt i »Oxford-mødet« i september 1936, hvor foruden John Hicks også Roy Harrod og James Meade præsenterede deres »Keynes-model«.

8. At der ikke er tale om en tilfældig lapsus, men en bevidst teoriudvikling fremgår af Modigliani (1999), hvor han i et interview konkluderer sin keynesianske position.

keynesianerne<sup>9</sup> dog tro mod deres intellektuelle ophav: At tilpasningen til den langsigtede ligevægt kan påvirkes ved brug af økonomisk politik – navnlig finanspolitik.

*b. De keynesianske makroøkonometriske modeller og Lucas-kritikken*

Det var også den neoklassiske syntese, som udgjorde det teoretiske grundlag for de såkaldte *makroøkonometriske* modeller, der blev udarbejdet i 1960erne og navnlig i 1970erne i takt med at regnekapaciteten på computerne blev forøget dramatisk. Her var det især navne som Franco Modigliani og Lawrence Klein, som stod for denne empirisk forankrede matematiske modeludvikling.<sup>10</sup> De *'keynesianske'* makroøkonometriske modeller blev i slutningen af 1970erne udsat for en betydelig kritik fra en kreds af stærkt markedsorienterede økonomer, der efterlyste et eksplicit mikroøkonomisk fundament for den makroøkonomiske teori. Disse såkaldte *nyklassiske* økonomer, bl.a. Robert Lucas, Thomas Sargent, Kydland og Precott var navnlig kritiske over for den usystematiske forventningsdannelse hos keynesianerne (og for den sags skyld også hos monetaristerne). De mente, at den ad hoc-prægede forventningsdannelse burde erstattes med en *'rationel forventningsdannelse'*, hvor de økonomiske aktører antages at kende den langsigtede ligevægtsløsning, som så kan lægges til grund for den rationelle forventningsdannelse. Denne forudsætning indebærer, at der (gen)indføres en forudsætning om fuld forudseenhed, hvilket gør det metodisk konsistent at arbejde med *generel ligevægt*, som referencemodellen. Makroøkonomisk teori var hermed blevet ført tilbage til sit prækeynesianske udgangspunkt som en konsekvens af, at den rationelle forventningshypotese gik sin sejrsgang. I slutningen af 1980erne formulerede Lucas det således:

The most interesting recent developments in macroeconomic theory seem to me describable as the reincorporation of aggregative problems such as inflation and the business cycle within the general framework of 'microeconomic' theory. If these developments succeed, the term 'macroeconomics' will simply disappear from use and the modifier 'micro' will become superfluous. We will simply speak, as did Smith, Ricardo, Marshall and Walras, of economic theory. [Lucas, 1987, s.108].

*c. Ny 'keynesianere' → 'Ny pigouvianere'?*

Nyklassikerne havde sat en ny dagsorden med kravet om et mikroøkonomisk fundament for de generelle ligevægtsmodeller, der blev understøttet af den rationelle for-

9. Listen over 'gammel-keynesianere' er lang. Her skal nævnes Paul Samuelson, James Tobin, James Meade, Robert Solow, Lawrence Klein og naturligvis John Hicks (alle nobelpristagere).

10. Vi vender tilbage til modeldiskussionen, men nævnes bør det, at Danmark også fik sin makroøkonometriske model, ADAM, som professor Ellen Andersen tog initiativet til i slutningen af 1960erne, og som senere blev koblet til Kleins verdensmodel 'project Link'.

ventningshypotese. Som bl.a. Andersen (2000) bemærker i sin oversigtsartikel, er hypotesen ikke specielt realistisk, men giver den analytiske fordel, at det kan udelukkes, at en systematisk fejlspecifikation i forventningsdannelsen kan begrunde afvigelser fra den generelle ligevægt.

Ny-keynesianernes forskningsstrategi blev dernæst at udvikle modeller, der kan analysere trægheder i tilpasningen til den generelle ligevægt, der ikke bryder med antagelsen om rationel mikroøkonomisk adfærd. Når sådanne trægheder kan identificeres, vil der som hovedregel være to politik-optioner, der kan forfølges. Enten kan den ofte institutionelt begrundede årsag til trægheden søges fjernet gennem strukturreformer. F.eks. vil fagbevægelsen være et eksempel på en institution, der kan blokere for 'the first best solution' ved at bremse den 'nødvendige' reallønstilpasning til opnåelse af fuld beskæftigelse. Hvis det ikke er muligt politisk at komme igennem med en sådan strukturændring kan tilpasningen til en ny (eller rettere til den gamle) ligevægt fremskyndes gennem en midlertidig efterspørgselsorienteret politik – heraf tilnavnet (ny) keynesianere.

I betragtning af at ny-keynesianere har givet køb på mindst to af Keynes' metode-mæssige innovationer: (1) at langsigtligevægten er ukendt p.gr.a. usikkerhed, og muligvis slet ikke eksisterer, og (2) at systemet ikke er selvregulerende (med eller uden trægheder), kan det uden overdrivelse konkluderes, at ny-keynesianere i deres metode-mæssige tilgang ligger tættere på traditionel neoklassisk teori end på Keynes' originale makroøkonomiske teori. Set i dette perspektiv er der en slående lighed med de teoretiske konklusioner i Pigous bog fra 1933, hvor han også havde betydelige vanskeligheder med at få teori og virkelighed til at hænge sammen. Pigou var, som nævnt, metodologisk neoklassiker (en loyal elev af Marshall), men tilstrækkelig meget realist til at konstatere, at de teoretiske konklusioner ikke svarede til mellemkrigstidens virkelighed. På tilsvarende måde er ny-keynesianerne i stigende grad blevet konfronteret med det empiriske 'problem', at arbejdsløsheden i de fleste europæiske lande vedvarende har ligget på et niveau, der i hvert fald antalmæssigt har klare mindelser om mellemkrigstiden. Set i et teorihistorisk lys ville jeg egentlig mene, at ny-keynesianere, snarere burde benævnes præ-keynesianere eller måske endnu mere betegnende for 'ny-pigouvianere' for at undgå forveksling med post-keynesianere.

*d. Post-keynesianere*<sup>11</sup>

*On the one side are those who believe that the existing economic system is, in the long run, a self-adjusting system, though with creaks and groans and jerks and interrupted by time lags,*

---

11. Skulle læseren have mod på en ganske detaljeret gennemgang af forskellige undskoler og deres indbyrdes forhold inden for den post keynesianske retning, så kan J.E.King (2002) *A History of Post Keynesian Economics since 1936* anbefales på det varmeste.



*outside interference and mistakes.... On the other side of the gulf are those that reject the idea that the existing economic system is, in any significant sense, self-adjusting. The strength of the self-adjusting school depends on it having behind it almost the whole body of organised economic thinking of the last hundred years, (Keynes, CWK XIII, s. 485).*

Jeg vil vove den påstand, at det vil være vanskeligt at finde en bedre, kortfattet præcisering af den grundlæggende forskel mellem ny-keynesianere og post-keynesianere, som samtidigt forklarer, hvorfor ny-keynesianerne/ny-pigouvianere fortsat betragtes som the main stream economics. De har ét ben i hver lejr: det ene står i den generelle ligevægtsteori, det andet står i anbefalingen af en aktiv økonomisk politik. Post-keynesianerne afviser at benytte en generel ligevægtsmodel og dermed også 'laboratorium-forsøg', hvilket får deres analyser til at fremstå mindre formaliseret i deres forsøg på at komme tættere på virkeligheden. Det har dog den konsekvens, at politik-anbefalinger bliver mindre præcise og mere afdæmpede. Paradoksalt nok har denne metodologisk set mere ydmyge fortolkning af de analytiske modellers udsagnskraft om virkeligheden – gennem en betoning af usikkerhedens betydning – indebåret, at den post-keynesianske teoriretning i traditionelle lærebøger betegnes som 'fundamentalistisk' og 'ekstrem' i den omfang den overhovedet omtales.

Post-keynesianere kan bedst afgrænses ved metodemæssigt at tage Keynes på ordet. I mellemkrigstiden var det empirisk åbenbart, at det makroøkonomiske system i de industrialiserede lande ikke var selvregulerende – arbejdsløsheden var permanent høj og fattigdommen udtalt. De første 20 år af efterkrigstiden så det derimod ud, som om arbejdsløsheden var overvundet ikke på grund af markedøkonomiens selvregulerende kræfter, men på grund af en aktiv økonomisk politik suppleret med udbygning af en omfattende velfærdsstat. Denne 20-årige periode skulle dog vise sig blot at være et isoleret intermezzo i disse landes økonomiske historie. Den makroøkonomiske ustabilitet tog til i løbet af 1960erne og manifesterede sig stærkt i de efterfølgende årtier. For en række europæiske lande voksede arbejdsløsheden til et niveau, der som nævnt havde mindelser om tilstanden i mellemkrigstiden.

Indtil slutningen af 1950erne levede gammel-keynesianere og post-keynesianere dog så nogenlunde fredeligt side om side, idet begge grupper var optaget af at analysere det nye begreb 'effektiv efterspørgsel' inden for rammerne af IS-LM modellen. Meget tid gik med at diskutere hældningerne på disse to kurver, og om penge- eller finanspolitik var det mest effektive instrument til at holde arbejdsløsheden stangen.

Det var først med Solows (1956) neoklassisk inspirerede vækstmodel, at det metodemæssige skisma blev definitivt blotlagt. Spørgsmålet om, den generelle ligevægtsmodel er en relevant (dvs. realistisk) ramme for forståelsen af makroøkonomisk vækst, stod her centralt.

Joan Robinson og Nicholas Kaldor i Cambridge mente 'nej', hvilket førte til den såkaldte *capital controversies*, der afdækkede den afgørende rolle, som den anvendte metode spiller for de analytiske konklusioner vedr. vækstens og den funktionelle indkomstfordelings determinanter. Den metodiske strid centrerede omkring det spørgsmål, hvorvidt den neoklassiske analyse har gyldighed, når makroøkonomien befinder sig uden for den generelle ligevægt – mens den, om så at sige, 'er på vej'. Hvordan kan vi afdække og beskrive et økonomisk system, hvor 'marginalitetsbetingelserne' ikke pr. automatik er opfyldt.<sup>12</sup> Robinson forsøgte at holde Solow fast på, at det (i bedste fald) kun var i ligevægten, at den neoklassiske teori havde gyldighed. Uden for ligevægten var det derimod ikke umiddelbart metodisk korrekt at antage, at her havde de neoklassiske konklusioner baseret på generel ligevægtsteori også gyldighed. Denne diskussion mellem Cambridge (US) og Cambridge (UK) førte ikke til nogen definitiv metodologisk afklaring. Neoklassikerne fortsatte, måske i mangel af bedre, måske fordi de fandt indvendingerne af mindre betydning, med at benytte den neoklassiske ligevægtsteori også uden for ligevægt.

Post-keynesianernes analytiske gennemslagskraft var stækket af, at de ikke havde et konstruktivt alternativ at sætte i stedet for Solows vækstmodel. Og her var der ikke megen hjælp umiddelbart at hente hos Keynes, der primært havde sat fokus på arbejdsløsheden, der var 1930erne alt overskyggende problem; men som i 1950ernes og 1960ernes vækstøkonomi syntes at høre en fjern fortid til.<sup>13</sup>

Den erkendelsesmæssige kerne, der adskilte Keynes fra sine forgængere (eller rettere samtidige) var som nævnt fænomenet '*usikkerhed*'. Det forhold at vi ikke har – og heller ikke kan opnå – sikker viden om fremtiden. Anerkendelse af 'usikkerhedens' analytiske relevans har betydning for forståelsen og beskrivelsen af, hvorledes en række samfundsøkonomiske institutioner 'fungerer', og hvorledes de med fordel kan organiseres. 1936-bogen er spækket med eksempler på, hvorledes usikkerhed gennemtrænger al økonomisk aktivitet. Bogen er derimod langt mindre formidlende med hensyn til en præcisering af, hvilke metodemæssige konsekvenser, der bør drages af 'usikkerhedens' ontologi for den makroøkonomiske teori.

12. Inden for velfærdsteori betegnes den tilsvarende problematik 'the theory of the second best solution', hvilket mere korrekt burde betegnes som 'the lack of a theory of the second best solution'.

13. Set i retrospektiv er det åbenbart, at de tre 'gamle' post keynesianere (Robinson, Kaldor og Sraffa) var uforberede på diskussionen af the long run, de havde alle tre en forestilling om, at der måtte 'eksistere' en form for langsigts ligevægt defineret i logisk tid (det er mest udtalt for Sraffa). Deres indsigelser gik overvejende på, om denne langsigts ligevægt er relevant i historisk tid. Joan Robinson sammenfatter denne modstilling i titlen på et af hendes sene essay 'History vs. Equilibrium'. Kaldor er bl.a. optaget af spørgsmålet om, hvad nu hvis der er increasing returns to scale, så er den neoklassiske ligevægtsteori ikke dækkende. Sraffa står som eksponent for den såkaldte *neoricardianske* skole, der er fortsat med at benytte en ligevægtsmetode og dermed har skilt sig ud fra (i hvert fald) nyere post-keynesiansk makroteori, jfr. King (2002).

Begrebet 'usikkerhed' får en fundamentalt forskellig epistemologisk betydning når den generelle ligevægtsmodel opgives som en relevant analytisk ramme. I så fald kan usikkerhed ikke reduceres til risiko. De 'store tals lov' har kun gyldighed, hvis den bagvedliggende stokastiske proces er konstant og udfaldene er uafhængige. Er dette ikke tilfældet kan end ikke statistisk usikkerhed reduceres til sikker viden på makroniveau. Denne metodologiske forskel har betydning for den information, som rationelle aktører kan forudsættes at have ved beskrivelse af virkeligheden. Antages fremtiden at være ukendt, fordi *we simply do not know* om f.eks. bunden vil gå ud af aktie- og obligationsmarkedet, så er det rationelt at efterspørge penge, der giver et lavt, men dog (relativt) sikkert afkast. Den finansielle sektor bør ansues som delvist begrundet i den usikkerhed der knytter sig til den fremtidige inflation. Der er en analytisk verden til forskel på at antage at inflationen er bestemt af den eksogent bestemte pengemængde, eller den er et resultat af et kompleks og uforudsigeligt sammenspil af makroøkonomiske faktorer.

Ikke-kvantificerbar usikkerhed – det Keynes kaldte *fundamental uncertainty* – øver indflydelse på både penge og rente og dermed på den effektive efterspørgsel efter varer og tjenester, hvilket forklarer titlen *Employment, Interest and Money*. Til forståelse af 'usikkerhedens' årsager og deraf følgende epistemologiske konsekvenser har bl.a. Paul Davidson (1973) og Victoria Chick (1982) i deres understregning af den afgørende rolle som penge og kredit spiller i en moderne makroøkonomi givet væsentlige bidrag.

Den faktiske økonomiske udvikling op gennem 1970erne havde gjort det åbenbart, at den 'neoklassiske syntese' nu repræsenteret ved AS-AD modellen med en påhægtet long run lodret 'Phillipskurve' som kernen i de store økonometriske modeller var kørt empirisk fast. Kombinationen af inflation og arbejdsløshed (den såkaldte 'stagflation')<sup>14</sup> og den tilsyneladende finanspolitiske impotens<sup>15</sup> havde, som beskrevet ovenfor, efterladt et 'keynesiansk' tomrum. Der var her 50 år efter publiceringen af GT et stærkt behov for at få klarlagt, på hvilke punkter den 'neoklassiske syntese' havde kørt Keynes' makroøkonomiske teori, og navnlig metode, af sporet. For at gennemføre denne 'relancering' var det nødvendigt at gå ned til de videnskabsteoretiske rødder. Den neoklassiske syntese ansuede det makroøkonomiske system som et 'urværk', der kun-

14. 'Stagflation' er en sammentrækning af 'stagnation' og 'inflation'. Empirisk er det dog forkert at karakterisere 1970erne og 1980erne som en egentlig stagnationsperiode, idet væksten i BNP fortsat var ganske betydelig, bort mere svingende end tidligere og ikke kraftig nok til at opsuge den arbejdsløshed, der blev genereret i år med en vækst, der var mindre end produktivitet og tilvækst til arbejdsstyrken.

15. Det er et empirisk uafklaret spørgsmål i hvilken grad det var styrken i den økonomiske politik der ændrede sig op gennem denne periode eller om det snarere var de analytiske modeller, der ændrede sig? I en generel ligevægtsmodel vil efterspørgselspolitik pr. definition være uden effekt, jfr. Lucas policy ineffectiveness konklusion.

ne forklares på et (logisk) positivistisk grundlag. Keynes havde derimod på et tidligt tidspunkt advaret mod denne naturvidenskabelige analogi:

*...economics is essentially a moral science and not a natural science...it deals with introspection and values .. and with motives, expectations, psychological uncertainties. It seems to me that economics is a branch of logic, a way of thinking; and that you [Roy Harrod, JJ] do not repel sufficiently firmly attempts ... to turn it into a pseudo-Natural-science..... Progress in economics consists almost entirely in a progressive improvement in the choice of model. The grave fault of the later classical school, exemplified by Pigou, has been to overwork a too simple or out-of-date model. Economics is a science of thinking in terms of models joined to the art of choosing models which are relevant to the contemporary world.*

(From two letters to Roy Harrod commenting upon his presidential address 'Scope and Method of Economics' in July 1938, CWK:XIV, s. 296-7 og 300).

*The object of our analysis is, not to provide a machine, or method of blind manipulation, which will furnish an infallible answer, but to provide ourselves with an organised and orderly method of thinking out particular problems; and, after we have reached a provisional conclusion by isolating the complicating factors one by one, we then go back on ourselves and allow, as well as we can, for the probable interactions of the factors among themselves. This is the nature of economic thinking, Keynes (1936, p. 297).*

Det makroøkonomiske system er med andre ord, ifølge Keynes, åbent. Der kan ikke én gang for alle opstilles én generel (ligevægts)model, der med et urværks præcision kan beskrive og forudsige det makroøkonomiske forløb. Og hvorfor ikke? Hertil svarer den nye generation af post-keynesianere, at den makroøkonomiske epistemologi må afspejle det forhold, at den menneskelige adfærd ikke er mekanisk. At økonomi grundlæggende er en humanvidenskab (moral science) med deraf følgende tilfældige elementer – tilfældige forstået som uforklarede snarere end usystematiske handlinger. Systematikken kan bl.a. bestå i gensidig påvirkning i form af altruisme, flokadfærd, spontanitet. Det er fænomener, der er fuldt forenelige med rationel adfærd, men som blot ikke udtømmende kan beskrives ved *individuel* nyttemaksimering under fuld viden. Tværtimod er fremtiden usikker både på det individuelle og på makroniveau – så alene af den grund kan altruisme og gruppesammenhold bidrage til at mildne konsekvenserne, når det uforudseelige 'uheld er ude'. Et forhold som bl.a. opbygningen af velfærdsstaten er et af de rationelle svar på, og som har ændret både mikro- og makro- adfærden.

Som en konsekvens af den ændrede opfattelse af, hvorledes den makroøkonomiske ontologi skal beskrives, drejes det videnskabsteoretiske grundlag i retning af realisme;

men ikke kun den *realisme*, der direkte kan observeres. Ontologisk ved vi, at der er andre faktorer, der er medbestemmende for det makroøkonomiske forløb end blot mikroøkonomisk betinget adfærd. Her er de økonomiske strukturer, informationsformidling, kredit- og arbejdsmarkedernes organisering, konkurrence fra udlandet, den økonomiske politik og erhvervslivets organisering af væsentlig betydning, men samtidig ikke altid direkte observerbar. Betydningen af disse institutionelle faktorer befinder sig i en form for empirisk gråzone, der i den terminologi som benyttes af *kritisk realisme* karakteriseres som det transcendent domæne, jfr Lawson (1997). Her må formuleres hypoteser om de økonomiske strukturers indflydelse på det makroøkonomiske forløb, som så på bedste popperianske vis må søges falsificeret. Et er dog givet, at disse økonomiske strukturer (lige som den mikroøkonomiske adfærd) er under konstant forandring. Derfor må den analytiske ramme udformes, så ændringer i disse faktorer kan indoptages, nye institutioner inkorporeres, ligesom forventningerne til den økonomiske politik (og fremtiden i det hele taget) umuligt kan være stationær. Alt sammen forhold der peger væk fra antagelsen om generel ligevægt og i retning af en åben ikke for rigid modelstruktur, der løbende kan revideres.

Dette er nogle af kernepunkterne i den nye videnskabsteoretiske tilgang, der er blevet udviklet inden for den post keynesianske makroteori igennem de seneste 10-15 år med vigtige bidrag fra Rod O'Donnell (1989), Tony Lawson (1997), Sheila Dow (1996) og (2002).

### **Makroøkonomi metode i moderne kontekst**

Intentionen med denne artikel er således ikke at gennemføre en egentlig analyse af Keynes' makroøkonomiske teori og metode. Det er ikke mit ønske her at give endnu en mulig tolkning af, 'what Keynes really meant'. Det er primært den inspiration, som læsning af Keynes' hovedværker stadig kan give nutiden, der begrundes mine referencer til et nu 70 år gammelt værk. For det er uomtvisteligt, at Keynes satte en ny metodologisk dagsorden til forståelse af makroøkonomiske ustabilitet – navnlig af arbejdsløshedens årsager.

Jeg gør mig dog ingen illusion om, at Keynes' makroøkonomiske metode ukritisk kan overføres til brug for forståelsen af de åbenbare balanceproblemer, der eksisterer her ved indgangen til det 21. århundrede. På den anden side er det, som argumenteret ovenfor, et endnu ikke gennemført forskningsprojekt at vurdere i hvilken grad den post-keynesianske metodologi konstruktivt kan bidrage til en bedre forståelse af årsagerne til og konsekvenserne af de aktuelle ubalancer både på nationalt, regionalt og globalt niveau. Når Keynes understreger betydningen af at se på 'the economy as a whole', så opfyldes dette metodiske krav ikke ved blot at afløse den hidtidige partielle ligevægtsteori med en generel ligevægtsmodel. For Keynes kunne kort sigt og langt-

sigts modeller ikke holdes adskilt – arbejdsløshed kunne også på lang sigt være en 'normal' tilstand. Den delvist ukendte fremtid ville bestemt ikke være upåvirket af det aktuelle investeringsniveau. Selv om det ikke kommer eksplicit frem, så ligger der i de ovenfor citerede metodemæssige afsnit en afstandtagen fra økonomisk analyse baseret på urværksmodeller, idet ethvert udviklingsforløb har unikke træk og vil være path-dependent.

I dette metodiske perspektiv kan en fornyet læsning af Keynes' egne værker give inspiration til udvikling af en metodologi, der er i bedre overensstemmelse med den åbne makroøkonomis ontologi. Et metodologisk arbejde der endnu er i sin vorden, men som klart peger udover den konventionelle analyseform, hvor væksttend og konjunkturbevægelser analyseres adskilt, og hvor der knap nok er levnet plads til, at den finansielle sektor kan have en realøkonomisk betydning.

### **Sammenfatning**

Keynesiansk-inspirerede makroøkonomer har, kort fortalt, genstandsfelt og økonomisk politik til fælles, men afviger markant med hensyn til den anvendte analysemetode.

*Genstandsfeltet* for den makroøkonomiske analyse omfatter stort set uden undtagelse hensynet til beskæftigelse og arbejdsløshed. Påvirkningen af beskæftigelsesniveauet, opgjort i personer, står centralt i Keynes-inspireret makroøkonomisk analyse. Hertil kommer, ligeledes som et bredt accepteret synspunkt at opnåelse af fuld beskæftigelse er et væsentligt mål, som den økonomiske politik, i varierende grad, kan bidrage til at fremme.

På disse to punkter adskiller Keynes-inspireret makroøkonomisk teori sig klart fra den nyklassiske teori. Som nævnt antages det i nyklassiske makromodeller, at arbejdsmarkedet altid er i ligevægt – ligesom den øvrige makroøkonomi. Denne modelmæssige tilgang indebærer, at et fald i beskæftigelsen er et udtryk for ændrede præferencer hos arbejdstagerne og repræsenterer derfor ikke en tilstand med ufrivillig arbejdsløshed. Dette begrundes også den nyklassiske konklusion, at økonomisk politik vil (i bedste fald) være 'ineffective', idet den vil gribe forstyrrende ind i en på forhånd tilpasset ligevægt. I en generel ligevægtsmodel, hvor aktørerne ydermere antages at have fuld viden med hensyn til fremtiden (antagelsen om rationelle forventninger), og hvor der ikke er markedsmæssige eksternaliteter, vil økonomisk politik være formålsløs med hensyn til at etablere makroøkonomisk stabilitet. Her vil markedsmekanismen være politikkerne overtruffen.

Derimod skiller vandene sig blandt de Keynes-inspirerede makroøkonomer med hensyn til metodologien. Ny-keynesianerne benytter (ligesom nyklassikerne) sig af en generel ligevægtsmodel og accepterer ofte hypotesen om rationelle forventninger

som et analytisk relevant udgangspunkt. Når ny-keynesianerne alligevel når til det resultat, at der er rum for en aktiv økonomisk politik, så skyldes det, at de sætter fokus på en række af markedsmæssige imperfektioner og eksternaliteter, i form af f.eks. transaktionsomkostninger, asymmetrisk information, moralsk svigt/svig (bl.a. hos forsikringstagere), interesseorganisationer m.fl., der kan fordrer en kollektiv politisk indsats for at genskabe generel ligevægt.

Post-keynesianerne er ikke uenige i, at en aktiv økonomisk politik kan være nødvendig for at genskabe makroøkonomisk stabilitet. Uenigheden er derimod begrundet i den anvendte analytiske metode. Denne metodiske forskel kan være så betydningsfuld, at de politik-anbefalinger, der fremføres er forskellige ikke mindst ved tilrettelæggelsen af den mere langsigtede økonomiske politik. Her kan forskellen mellem på den ene side en lukket modelstruktur, hvor ligevægten dominerer på længere sigt og på den anden side en åben ontologi, hvor fremtiden er usikker, komme til at dominere. I generel ligevægtsteori antages det, at den langsigtede udvikling kendes med betydelig sikkerhed, her er det de mere kortsigtede trægheder, der er svære at fastlægge mere præcist. I skarp kontrast til denne konklusion hæfter post-keynesianerne sig navnlig ved, at usikkerheden vil udgøre et stigende analytisk problem, jo længere planlægningshorisonten strækkes ud i den åbne og notorisk ukendte fremtid.

Dette er kernen i den metodologiske uenighed, der forsat hersker blandt Keynes-inspirerede makroøkonomer.

#### Litteratur

- Andersen, T. M. 2000. Makroteori i Chr. Hjorth-Andersen, red., *Udviklingslinjer i økonomisk teori*, København: DJØF's Forlag.
- Carabelli, A. 1988. *On Keynes's Method*, London: MacMillan.
- Chick, V. 1983. *Macroeconomics after Keynes – a Reconsideration of the General Theory*, Oxford: Philip Allan Publishers.
- Clower, R.W. 1998. Keynes in retrospect i Ph. Arestis, red., *Method, Theory and Policy in Keynes – Essays in Honour of Paul Davidson*, vol. 3, Cheltenham: Edward Elgar.
- Davidson, P. 1972. *Money and the real World*, London: Macmillan.
- Davidson, P. 1994. *Post-Keynesian Macroeconomic Theory: A Foundation for successful Economic Policies for the Twenty-First Century*, Aldershot: Edward Elgar.
- DeLong, J. B. 1999. The Triumph of Monetarism, *Journal of Economic Perspectives*, 14, winter, s. 83-94.
- Dow, S. C. 1996. *The Methodology of Macroeconomic Thought: a Conceptual Analysis of Schools of Thoughts in Economics*, Cheltenham: Edward Elgar.
- Dow, S. C. 2002. *Economic methodology: an inquiry*, Oxford: Oxford University Press.
- Dow, S. C. og J. Hillard, red. 1995. *Keynes, Knowledge and Uncertainty*, Aldershot: Edward Elgar.
- Estrup, H. m.fl. 2004. *Økonomisk Teorihistorie – en kort introduktion*, København: DJØF's Forlag.
- Fitzgibbons, A. 1988. *Keynes's Vision: a New Political Economy*, Oxford: Clarendon Press.
- Fuglsang, L. og P. Bitsch Olsen, red. 2004. *Videnskabsteori for Samfundsvidenskaberne*, 2. udg. Frederiksberg: Samfundsfagslitteratur.
- Hahn, F. og Solow, R. 1996. *A Critical Essay on Modern Macroeconomic Theory*, Cambridge, US: The MIT Press.
- Harcourt, G. C. og Riach, P. A. 1997. *A Se-*

- cond Edition of the General Theory*, to bind, London: Routledge.
- Jespersen, J. 2004. *Introduktion til makroøkonomisk teori*, (2. udg.), København: DJØF's Forlag.
- Jespersen, J. 2002. *Keynes – den makroøkonomiske teoris oprindelse og udvikling*, DJØF's Forlag.
- Jespersen, J. 2003. Makroøkonomisk metode: Keynes og lærebøgerne – to adskilte verdener, *Økonomi og Politik*.
- Journal of Post Keynesian Economics. 1999. Post Keynesianism and critical realism, special issue, Fall/ vol. 22, nr.1, s. 3-131 med bidrag fra bl.a. Tony Lawson, Sheila Dow, Roy Rotheim og Paul Davidson.
- Keynes, J. M. 1936. *The General Theory of Employment, Interest and Money*, London: Macmillan.
- Keynes, J. M. 1973. The Collected Writings of John Maynard Keynes, redigeret af D. Moggridge, London: Macmillan, St. Martin's Press. Af særlig relevans for udviklingen af Keynes' nye metodologi er The General Theory and After, part one and two, hhv. bind XIII: Preparation, og bind XIV: Defence and Development.
- King, J. E. 2002. *A History of Post Keynesian Economics since 1936*, Cheltenham: Edward Elgar.
- Lawson, T. 1997. *Economics and Reality*, London: Routledge.
- Modigliani, F. 1944. Liquidity Preference and the theory of interest and money, *Econometrica*, January.
- Lucas, R.E., Jr. 1987. *Models of Business Cycles*. Oxford: Basil Blackwell.
- Modigliani, F. 1999. interview i *Conversations with Leading Economists*, redigeret af B. Snowdon og H.R.Vane, Cheltenham: Edward Elgar.
- O'Donnell, R. M. 1989. *Keynes: Philosophy, Politics and Economics. The Philosophical Foundations of Keynes's Thought and their Influence on his Economics and Politics*, London: Macmillan.
- Patinkin, D. 1956. *Money, Interest and Prices: an Integration of Monetary and Value Theory*. New York: Harper & Row.
- Pedersen, L. H. og M. Rasmussen. 2000. DREAM-modellen, *working paper*, Danmarks Statistik.
- Pigou, A. C. 1933. *Theory of Unemployment*, London: Macmillan.
- Romer, D. 1996/2001. *Advanced Macroeconomics*, New York: The McGraw-Hill Co, (stort set uændret 2. udgave).
- Rotheim, R., red. 1998. *New Keynesian Economics/Post Keynesian Alternatives*, London: Routledge.
- Skidelsky, R. 1983/1992. John Maynard Keynes: Hopes betrayed, 1883-1920, London: Macmillan Press.
- Skidelsky, R. 1992. *The Economist as Saviour, 1920-37*, London: Macmillan Press.
- Snowdon, B., H. Vane and P. Wynarczyk. 1994. *A Modern Guide to Macroeconomics – an Introduction to Competing Schools of Thought*, Cheltenham: Edward Elgar.
- Solow, R. M. 1956. A Contribution to the Theory of Economic Growth, *Quarterly Journal of Economics*, 70, s. 65-94.
- Sørensen, P. B. og H. J. Whitta-Jacobsen. 2003. Macroeconomics for the Long Run and for the Short Run, *lecture notes*, Københavns Universitet, Økonomisk Institut.
- Sørensen, P. B. og H. J. Whitta-Jacobsen. 2005. *Introducing Advanced Macroeconomics: Growth and Business Cycles*, London/New York: The McGraw-Hill Companies.
- Togati, D.T. 1998. *Keynes and the Neoclassical Synthesis – Einsteinian versus Newtonian macroeconomics*, London: Routledge.



## Appendiks:

*Tabel 1.1. Skematisk oversigt over makroøkonomiske positioner.<sup>1</sup>*

	Nyklassikere	Ny-keynesianere <sup>3</sup>	Post-keynesianere
Videnskabsteori:	Idealisme	(logisk) positivisme	(kritisk) realisme
Modelbillede:	urværk	'grus i urværket'	organisme
Modelramme:	lukket	lukket	åben
Modelindhold:	generel ligevægt	generel ligevægt med trægheder	kausalanalyse path-dependent
Adfærdsrelationer:	market clearing	mikroøkonomisk optimering	makroøkonomisk repræsentation
Forventningsmodel:	rationel forventningsdannelse	rationel forventningsdannelse	al tilrådighed information
Vidensniveau:	fuld information	asymmetrisk information	begrænset information
Genstandsfelter: <sup>2</sup>	markedsstruktur vækst	beskæftigelse inflation	arbejdsløshed fordeling
Efterspørgselspolitik:	ineffektiv	midlertidig	permanent

1. Denne skematisk oversigt må i sin natur være lapidarisk og prætenderer således ikke at være udtømmende. Den skal udelukkende tjene det formål at give nogle centrale og genkendelige positioner, der markerer væsentlige lighedspunkter og navnlig forskelle.

2. Her er det vanskeligt at angive dækkende genstandsfelter for de tre makroøkonomiske teoriretninger, men de forskellige genstandsfelter giver også en (delvis) begrundelse for forskelle i valget af model.

3. Med mindre modifikationer falder 'gammel-keynesianerne' (den neoklassiske syntese) i denne kategori. Det er i deres skepsis over for hypotesen om 'rationel forventningsdannelse' og kravet om et eksplicit 'mikroøkonomisk fundament' at de adskiller sig fra 'ny-keynesianere'.

## Bog anmeldelser

Hector Estrup, Jesper Jespersen og Peter Nielsen, *Den økonomiske teoris historie – en introduktion*. Jurist- og Økonomforbundets Forlag 2004, 174 s. Anmeldt af Kurt Pedersen.

Det vil være læserne af nærværende tidsskrift bekendt, at Jurist- og Økonomforbundets Forlag løbende har udgivet otte bøger om *økonomiens konger*. Serien er nu blevet afrundet med et oversigtsværk, der har til hensigt at få etableret en kongerække. En anerkendelsesværdig intention, som bringer antallet af nye, dansksprogede teorihistorier op på fire.

De tre forfattere har delt arbejdet mellem sig efter hovedinteresser. Hector Estrup tager det lange træk fra den græske oldtid over klassikerne til den neoklassiske teori. Peter Nielsen tager sig af Marx og marxisterne samt institutionel og evolutionær økonomi. Endelig behandler serieredaktør Jesper Jespersen den makroøkonomiske teori (inklusive neoklassiske varianter). Endvidere har han bidraget med en indledning om økonomiens slægtstavle og en afslutning om forskningsprogrammer og økonomisk metode.

Således er Estrups område den succesombruste *main stream* tænkning, som løbende udfordres af forskellige heterodokse grupperinger. For en overordnet redaktionel-strategisk betragtning er det en interessant konstatering, at »kongeriget« – neoklassikerne – ikke er repræsenteret med bare en enkelt konge, men at seriens otte »konger« fordeles sig på tre klassikere (fire, hvis Marx tælles med), to heterodokse (tre, hvis Marx tælles med her), samt Keynes og Friedman. Altså tre døde konger, tre uden kongerige og to aktionærer i den hellige og almindelige makrofusion. Således ophøjes mainstream økonomien til at være en del af den akademiske republik!

Fremstillingen af de klassiske økonomer, deres verden, problemstillinger og tænkning

er vel Estrups paradedisciplin. Økonomi og moral gøres til et grundlæggende tema, og fremstillingen fra Smith til Stuart Mill er knap, præcis og veldisponeret. Det er ingen ringe litterær præstation at få den klassiske teori afviklet med mening på blot 20 sider.

Stilen holdes i det følgende kapitel om neoklassikerne, som håndteres overvejende kronologisk med let, men sikker hånd. Gossen og grænsenyttens, Cournot og monopolpriserne, Walras og den økonomiske arkitektur, Menger og subjektiviteten, Clark og indkomstfordelingen, og Marshall og markedet. Det samlede punkt – økonomisk analyse som en optimeringsøvelse – præsenteres, og på s. 83-84 diskuteres ånden i den neoklassiske maskine – modellernes rationalitet. Ikke urimeligt placeres P. A. Samuelson her som det tyvende århundredes største neoklassiker.

Overgangen til kapitlet om Marx betegner noget af en kovending. Hvor Estrup omfatter sine klassikere og neoklassikere med suverænt overblik og velvillig distance, bliver Peter Nielsen personlig og propagandistisk. Det kommunistiske Manifest er »flot og fejende«, Marx »beskyder« den klassiske politiske økonomi og er i den forbindelse »fuld af skelsættende teorielementer«. Marx stiller sig ikke tilfreds med glansbilleder, men begriber det *dybe* niveau. Der er for meget propaganda, ikke mindst fordi indholdet i alt væsentligt ikke kommer ud over vareanalysen i Kapitalens første bind. Der er mindelser om kapitallogikkerne anno 1970.

Selvom eftertiden efter Nielsens mening misbrugte Karl Marx, er han stadig en »uomgængelig hovedskikkelse i tænkningens historie« – det er »ironisk nok« kun de (fandens?) økonomer, der ikke vil begribe det. Anprisningen af Marx fortsætter i det følgende kapitel om institutionel og evolutionær økonomi, hvor en blandet flok af de heterodokse spredes som jernspåner omkring den marxske magnet. Stort set er alle de heterodokse tænkssomme og gode, og bortset fra

J. A. Schumpeter (der havde en stor svaghed for Walras såvel som for kapitalismen) har de stort set alle en høne (eller en høsegård) at plukke med de stride neoklassikere.

Galbraith ålede det kapitalistiske samfund, fordi selv ganske ligegyldige varer annonceres med en energi som gjaldt det genopstandelsen af Kristus og samtlige 12 disciple. Denne karakteristik kunne for så vidt godt passe på Nielsens to kapitler: Marx og hans tolv disciple (Lenin, Hilferding, Luxembourg, Schmoller, Veblen, Adorno, Fromm, Mitchell, Commons, Galbraith, Heilbroner og Schumpeter – sidstnævnte vel i rollen som *JudAS*) markedsføres med langt flere plusord end egentlig (for)brugeroplysning.

Jespersens kapitel om den moderne makroøkonomi giver en forholdsvis kortfattet fremstilling af de mere eller mindre konflikterende skoler, der har udviklet sig i med- og modspil. Ingen vil vel på forhånd nære tvivl om forfatterens keynesianske twist, som da også bestemmer kapitlets optik. Det er ros-værdigt, at fremstillingen strækker sig helt op til nutidens økonomer og deres positioner. Givet den begrænsede plads skiller forfatteren sig pænt fra opgaven.

Både i starten og slutningen af bogen behandler Jespersen metodespørgsmål og bibringer herigennem fremstillingen en nyttig ramme. Naturligt nok kredses der om konflikten mellem de deduktive neoklassikere – idealisterne – og de induktive historisk-institutionalistiske økonomer. På side 166 får han

anbragt Marx som idealist i selskab med Ricardo, Walras og Lucas, hvor medforfatter Nielsen på side 90 rubricerede selvsamme som videnskabelig anti-idealist. Tese og modtese?

Af særlig interesse er Jespersens overvejelser over Keynes metode. »*Han gav sig ikke af med egentlige forudsigelser, men skitserede nogle udviklingstendenser*« – og var de skræmmende, så havde han en plan (s. 89). Det kunne næsten lyde som om Keynes opfandt en makroøkonomisk scenarie-teknik. Det er positivt, at metodeovervejelser og videnskabsteori i stigende omfang finder vej til de teoriehistoriske fremstillinger – og det ville være glimrende, om Jespersen engang ved lejlighed gav emnet en mere systematisk behandling.

I forordets første sætning betegnes værket som en appetitvækker, og det er da også forholdsvis letlæst. Man kan tvivle på, hvor meget læsere uden synderlig forhåndsviden får ud af øvelsen, men det ligger i en så kort fremstillings natur. Alvorligere er de markante stilbrud, som forfatterne med fordel kunne have nedtonet. Dertil kommer, at økonomernes – eller deres skolars – stridigheder gøres til genstand for en lidt overdreven dramatisering. Forholder det sig ikke i virkeligheden sådan, at de mangeartede tilgange byder på oplagte muligheder for befrugtning? Og er det ikke netop derfor det kan være vanskeligt at få øje på bare et enkelt paradigmeskift af kuhnske dimensioner i den økonomiske teoris historie?



# Makroøkonomiske forudsigelser baseret på diffusionsindeks

Christian Dahl

Purdue University, E-mail: [dahlc@mgmt.purdue.edu](mailto:dahlc@mgmt.purdue.edu)

Henrik Hansen

Økonomisk Institut, Københavns Universitet, E-mail: [henrik.hansen@econ.ku.dk](mailto:henrik.hansen@econ.ku.dk)

John Smidt

Det Økonomiske Råds Sekretariat, E-mail: [jsm@dors.dk](mailto:jsm@dors.dk)

*SUMMARY: This paper presents simulated out-of-sample forecasts based on diffusion indexes for monthly Danish macroeconomic data. The diffusion indexes are derived from 246 series (172 monthly and 74 quarterly series). The primary focus is on forecasts of unemployment, industrial turnover and inflation at horizons of 6, and 12 months over the period 1995-2003. It is shown that it is possible to obtain mean squared forecast errors (MSFE) of diffusion index forecasts that are smaller than the MSFE of a standard autoregressive model. However, the gain in forecasting accuracy is not robust as the results are very sensitive to the specification of the forecasting equation. As it is difficult to establish general rules with respect to the parameterization of the forecasting equation this poses a serious problem in real time forecasts. Thus, the basic results based on Danish data are not as good as has been seen in the literature. The disappointing results are probably related to very volatile estimates of the diffusion indexes. Based on this observation, a possible improvement is suggested, namely calculation of the diffusion indexes on the basis of filtered data. Results based on this idea seem much more promising than the basic setting. Using filtered data the diffusion index model leads to a reduction of MSFE of around 10 to 15 per cent relative to a standard autoregressive model.*

---

## 1. Indledning

Forudsigelser af økonomiske størrelser er af interesse af mange grunde. Aktører på de finansielle markeder handler på baggrund af forventninger til fremtidige aktiekurser, renter og valutakurser. Erhvervsfolk fastlægger produktionsomfanget og beslutter

---

Denne artikel er et resultat af et arbejde, der har været støttet af Statens Samfundsvidenskabelige Forskningsråd. Et arbejdspapir fra Det Økonomiske Råds Sekretariat indeholder yderligere analyser og dokumentation, jf. Dahl, Hansen og Smidt (2005a). Bl.a. indeholder arbejdspapiret analyser af flere serier, resultater

*Fortsættes næste side...*

sig for investeringer på baggrund af den forventede udvikling i efterspørgsel og omkostningsforhold. Politikerne tager beslutninger om økonomisk politiske indgreb på baggrund af kort- og langsigtede prognoser for dansk økonomi. Selv den enkelte forbruger baserer sine indkøb på mere eller mindre sofistikerede forventninger til priser og indkomst i fremtiden – i hvert fald i teorien.

Der findes mange metoder til at lave forudsigelser. Typisk tages udgangspunkt i fortiden og de aktuelle tendenser. På baggrund af historiske data forsøger man at finde mønstre, der – hvis de også gælder fremover – kan bruges til at prognosticere den fremtidige udvikling. Metoderne kan være mere eller mindre avancerede – lige fra tommelfingerregler og simple ekstrapolationer til meget avancerede matematiske og statistiske metoder.

Sammenhængende forudsigelser af centrale makroøkonomiske størrelser som BNP, arbejdsløshed, inflation, betalingsbalance osv. baseres ofte på makroøkonometriske modeller. I Danmark er de mest kendte modeller ADAM, SMEC og MONA. Disse modeller baserer sig på en blanding af teori og data. Modellerne karakteriseres som »kortsigtsmodeller«, men ingen af modellerne har som primært sigte at være gode til forudsigelser. Selvom modellerne anvendes som støtte ved udarbejdelsen af forudsigelser, er det primære formål med modellerne snarere at danne grundlag for konsekvensberegninger af forskellig karakter.

Et særligt problem i forhold til at anvende disse modeller til forudsigelser er, at de er baseret på årsdata (ADAM og SMEC) eller kvartalsdata (MONA). Ingen af modellerne tager således direkte højde for informationen i traditionelle nøgletal, der typisk ligger på månedsbasis. I praksis må modelbrugeren derfor selv forsøge at indarbejde den nyeste konjunkturinformation fra konjunkturbarometre, forbrugerforventninger, månedlige ledighedstal, forbrugerpriser osv. Ofte foregår indarbejdelsen af denne type information på mere eller mindre ad hoc vis, og der kan derfor være behov for på lidt mere systematisk vis at forsøge at uddrage informationsindholdet, der bl.a. ligger i de månedlige nøgletal.

I denne artikel undersøges en forholdsvis ny metode til at forudsige økonomiske tidsserier på danske data, der bl.a. kan anvendes i kombination med de store makroøkonomiske modeller. Den grundlæggende ide bag metoden er, at udviklingen i et meget stort antal økonomiske tidsserier kan beskrives ved et lille antal underliggende fælles faktorer (»diffusionsindeks«). Disse faktorer kan betragtes som udtryk for fælles konjunkturbevægelser. De fælles faktorer, der i udgangspunktet ikke er observerbare,

---

*fortsat...*

for 1 måneds forudsigelseshorisont, supplerende følsomhedsanalyser samt en omfattende litteraturgennemgang. Programkoden til dannelse af diffusionsindeksene mv. samt de tilhørende data udleveres ved henvendelse til Det Økonomiske Råds Sekretariat. Beregningerne er foretaget i Ox, jf. Doornik (2001). To anonyme referees takkes for en række forslag til forbedringer og ideer til videre afestning.

identificeres ved hjælp af statistiske metoder, hvorefter faktorerne anvendes til konstruktion af forudsigelser af de enkelte tidsserier af interesse. Metoden adskiller sig fra traditionelle, regressionsbaserede forudsigelser på ledende indikatorer ved, at man ikke på forhånd skal udvælge et begrænset antal serier. Tankegangen bag diffusionsindeksene er, at man kan anvende *mange* tidsserier ved dannelsen af forudsigelserne, og at man derved undgår en tidskrævende proces med at finde de ledende indikatorer og estimere forudsigelsesligninger.

Forudsigelser baseret på diffusionsindeks er beskrevet af Stock og Watson (1998, 2002a og 2002b). Metoden har med stor succes været anvendt på amerikanske data, og den har også vist sig brugbar for andre lande som UK, Canada, Japan og Euroland.<sup>1</sup> Succesen, i form af mindsket forudsigelsesusikkerhed, er dog sjældent så stor som for de amerikanske data, og flere af studierne viser, at diffusionsindeksmodellen kan være meget følsom over for forskelle i modelspecifikationen.

I denne artikel vises, at det for Danmark er vanskeligt at konstruere forudsigelser baseret på diffusionsindekset, der er bedre end forudsigelser fra en almindelig autoregressiv model. Det er ganske vist muligt at opstille forudsigelser af udviklingen i ledigheden, industriens omsætning og nettopriserne for perioden 1995-2003 og med forudsigelsehorisonter på 6 og 12 måneder, hvor den gennemsnitlige kvadrerede forudsigelsesfejl (MSFE) fra diffusionsindeksmodellen er lidt mindre end MSFE fra en almindelig autoregressiv model. Disse resultater er i det basale set-up imidlertid ikke robuste over for specifikationen af forudsigelsesligningerne. Følsomheden indebærer, at et »forkert« valg af parametre i diffusionsmodellen kan lede til dårligere forudsigelser i forhold til en autoregressiv model. Den manglede robusthed over for parametriseringen gør, at det næppe er sandsynligt, at diffusionsindeksmodellens forudsigelsesegenskaber i praksis vil være bedre end andre forudsigelsesmetoder. Samtidig er gevinsten ved at bruge diffusionsindeksmodellen på danske data mindre end fundet for andre lande.

En forklaring på de mindre gunstige resultater for Danmark kan være, at de anvendte danske data tilsyneladende er ganske støjfyldte, hvilket indebærer, at signalet fra de underliggende, fælles konjunkturbevægelser, som udnyttes i diffusionsindeksmodellen, ikke fremstår særligt klart. Dette illustreres bl.a. af, at de beregnede diffusionsindeks er ganske volatile. Formentlig er det dårlige »signal/støj-forhold« en af årsagerne til, at de her præsenterede resultater ikke er lige så positive som en række internationale studier. Analyser præsenteret i artiklen viser således, at forbedringen af forudsigelsesegenskaberne er mere robust, når diffusionsindeksene beregnes på baggrund af filtrerede (støjrensede) tal.

---

1. Artis, Banerjee og Marcellino (2004) analyserer eksempelvis data fra UK, Brisson, Campbell og Galbraith (2003) ser på canadiske data, Shintani (2003) på japanske data, og Angelini, Henry og Mestre (2001) ser på inflationen i Euro-området.

Artiklen er struktureret som følger. Diffusionsindeksmodellen introduceres i afsnit 2, og de anvendte data præsenteres i afsnit 3. Afsnit 4 giver en præcis beskrivelse af forudsigelsesproceduren og evalueringen af de to forudsigelsesmodeller, der betragtes i analysen, og resultaternes følsomhed over for en række valg i forbindelse med specificationen af forudsigelsesmodellen vises. Afsnit 5 viser, hvorledes en udvidelse af diffusionsmodellen, med en forudgående udglatning af de mange makroøkonomiske serier, kan forbedre modellens egenskaber. Endelig gives en kort opsummering og perspektivering i afsnit 6.

## 2. Diffusionsindeksmodellen

Som nævnt i indledningen kan forudsigelser laves på mange måder. Inden for makroøkonomi, hvor man oftest arbejder med tidsserier, er statistiske eller økonometriske metoder fremherskende, jf. f.eks. Clements og Hendry (1998). Men selv når man begrænser sig til disse forudsigelsesmetoder – oftest forskellige udgaver af lineære og ikke-lineære regressionsmodeller – skal der foretages en række (arbitrære) valg bl.a. med hensyn til udvælgelse af indikatorer, transformationer heraf og specification af dynamikken, herunder eksempelvis lagstrukturen. Der har derfor i mange år været stor interesse i at finde rent databaserede forudsigelsesmodeller enten egentlige tidsrækkemodeller eller databaserede indikatormodeller. Diffusionsindeksmodellen kan ses som det seneste skud på stammen i klassen af databaserede indikatormodeller.

For at eksemplificere forudsigelsesproblemet kan man se på en simpel model for en variabel (f.eks. vækstraten i industriens omsætning). I forudsigelser over korte horisonter kan man ofte tænke på en given serie som en sum af to elementer: Et konjunkturalelement og et såkaldt idiosynkratisk-element (støj). Denne opdeling i konjunktur og støj kan i en forudsigelsesmodel formaliseres ved:

$$y_{t+1} = \alpha c_t + \varepsilon_{t+1}, \quad (1)$$

hvor  $y_t$  er den variabel, man ønsker at forudsige. Den forklarende variabel,  $c_t$ , er konjunkturalelementet, mens støjen angives af  $\varepsilon_{t+1}$ , som antages at være ukorreleret over tid. Parameteren  $\alpha$  angiver konjunkturrens indflydelse på den afhængige variabel. Problemet i denne simple forudsigelsesmodel er, at konjunkturalelementet ikke er direkte observerbart.

Udgangspunktet for de fleste indikatormodeller er, at man søger at approksimere konjunkturalelementet ved brug af indikatorer. Man observerer således over tid en samvariation mellem konjunkturfølsomme serier, som f.eks. industriomsætningen, på den ene side og en række konjunkturindikatorer på den anden. En lineær model for det uobserverbare konjunkturalelement kan f.eks. formuleres som:



$$c_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \dots + \beta_k x_{kt} + w_t, \quad (2)$$

hvor  $x$ 'erne er konjunkturindikatorerne. Hvis indikatorerne er gode, er de tæt korreleret med konjunkturalelementet, hvilket vil give en lille residual,  $w_t$ , og dermed en lille residualvarians.

Indikatormodellen for konjunkturalelementet kan indsættes i forudsigelsesligningen, der således bliver en almindelig regressionsmodel. Forudsigelsesfejlen vil ud over den grundliggende fejl,  $\varepsilon_{t+1}$ , også afhænge af konjunkturrens indflydelse,  $\alpha$  samt residualen i indikatormodellen,  $w_t$ :

$$y_{t+1} = \alpha\beta_1 x_{1t} + \alpha\beta_2 x_{2t} + \dots + \alpha\beta_k x_{kt} + (\alpha w_t + \varepsilon_{t+1}). \quad (3)$$

Ofte forbedres en indikatormodel, når man medtager mange indikatorer frem for få, idet indikatormodellens residualer dermed bliver mindre. Et problem kan imidlertid være, at sammenhængen mellem variablen, der skal forudsiges, og indikatorerne ikke er stabil over tid. Et andet problem er, at mængden af indikatorer nødvendigvis må begrænses til et antal, som er (væsentligt) mindre end antallet af observationer for variablen, man ønsker at forudsige. Hermed opstår et problem med at udvælge de bedste konjunkturindikatorer, og da indikatormodellerne sjældent er baseret på økonomiske teorier, bliver dette valg af specifikke indikatorer mere eller mindre arbitrært.

For at undgå det arbitrære udvalgsproblem, kan man i princippet forsøge at udvælge de relevante indikatorer ved at estimere en række forskellige indikatormodeller og derefter sammenligne disse modellers forudsigelsesegenskaber. Problemet med denne strategi er, at man meget hurtigt står over for at skulle afteste så mange modeller, at det bliver umuligt at gennemføre indenfor en kort tidsramme – selv med den nuværende computerteknologi. Hvis man f.eks. ønsker at finde den bedste forudsigelsesmodel for industriomsætningen på baggrund af de 245 andre serier, der indgår i den samlede indikatordatabase, vi benytter i dette arbejde, og man for at afgrænse søgeprocessen begrænser sig til altid at inddrage netop 5 indikatorer, vil der være over 7 milliarder forskellige regressionsmodeller, som skal estimeres og sammenlignes. Dette kan nedbringes til lige under 2,5 millioner regressioner, hvis man begrænser sig til 3 indikatorer, mens introduktion af laggede værdier naturligvis vil få problemstillingen til at nå endnu mere astronomiske størrelser. Udfordringen er altså at begrænse antallet af mulige indikatorer med mindst muligt tab af information om konjunkturalelementet.

Diffusionsindeksmodellen, som den er udviklet i Stock og Watson (1998 og 2002a,b), kan opfattes som en måde at reducere dimensionen af indikatormodellen ved at begrænse antallet af konjunkturindikatorer i regressionsmodellen samtidig med, at informationen fra alle relevante serier bibeholdes. Den grundliggende ide i diffu-

sionsindeksmodellen er, at konjunktorelementet antages at indgå i (næsten) alle økonomiske serier. Det er netop dette kendetegn, der adskiller konjunktoren fra andre, mere specifikke, elementer. Dette betyder, at man i forhold til ligning (1) forestiller sig, at konjunktorelementet er fælles for alle serier, mens støjleddet er seriespecifikt. En samlet model for  $N$  økonomiske tidsserier antages dermed at kunne beskrives ved:

$$x_{it} = \sum \lambda_{ij} F_{jt} + e_{it}, \quad i=1, \dots, N, \quad (4)$$

hvor de  $r$  ( $\ll N$ ) serier (eller konjunkturfaktorer),  $F_{1t}, \dots, F_{rt}$ , tilsammen repræsenterer det ikke-observerbare konjunktorelement,  $\lambda_{ij}$  angiver konjunkturfaktorerens indflydelse på hver af serierne, og  $e_{it}$  er de seriespecifikke støjled.

Modellen i (4) er et eksempel på en faktormodel, der er en velkendt statistisk model med bred anvendelse i bl.a. psykologien og sociologien. Hvis modellen faktisk er som beskrevet i (4), kan man estimere de  $r$  uobserverbare konjunkturfaktorer  $F_{jt}$  på baggrund af  $T$  observationer af de  $N$  serier. Dermed er konjunkturinformationen i de  $N$  serier reduceret til  $r$  serier.

De uobserverbare faktorer kan estimeres på flere måder afhængig af antagelserne om faktorerne og de idiosynkratiske støjled, jf. Anderson (1984b). De fleste estimationsmetoder bygger på iterative algoritmer, men hvis der er mange observationer, er en hurtig og konsistent estimator af de uobserverbare faktorer de første principale komponenter i korrelations-matricen for alle  $x$ 'erne (jf. nedenfor).

Når de  $N$  serier ( $x_{1t}, \dots, x_{Nt}$ ) observeres over  $T$  perioder og samles i et panel, kan faktormodellen skrives i kompakt form som:

$$X = FA' + e \quad (5)$$

hvor de økonomiske serier er samlet i  $T \times N$ -matricen  $X$ , således at hver række i  $X$  repræsenterer en serie, og hver søjle er en tidsobservation. Konjunkturfaktorerne er tilsvarende samlet i  $T \times r$ -matricen  $F$ , og de enkelte parametre (de såkaldte loadings eller faktorfølsomheder  $\lambda_{ij}$ ) er samlet i  $N \times r$ -matricen  $A$ .

I den klassiske faktormodel, jf. Anderson (1984a,b) eller Lawley og Maxwell (1971), antages de idiosynkratiske elementer ( $e_{it}$ ) at være uafhængige på tværs af serierne og mellem de enkelte observationer for den enkelte serie, svarende til en diagonal kovariansmatrix for  $e$ . For makroøkonomiske tidsserier er det vanskeligt at opretholde en sådan uafhængighedsantagelse. Men et af bidragene i Stock og Watson (2002b) er netop at vise, at man kan tillade en vis korrelation i de idiosynkratiske led *både* over tid og på tværs af serierne og alligevel opnå konsistent estimation af konjunkturfakto-

rne, når antallet af serier,  $N$ , og antallet af observationer,  $T$ , begge er store.<sup>2</sup> En væsentlig betingelse for konsistent estimation er, at konjunkturfaktorerne og de idiosynkratiske komponenter er stationære og ergodiske. Dette betyder specielt, at alle serier i faktormodellen skal være stationære. Derfor finder man i alle anvendelser af diffusionsindeksmodellen, at første trin er en transformation af serierne, således at de kan forventes at være stationære. De hyppigste transformationer er første differenser af serierne enten direkte eller efter en log-transformation.

Estimationen af konjunkturfaktorerne i Stock og Watson (1998, 2002a,b) bygger grundlæggende på, at faktormodellen kan opfattes som en regressionsmodel. Hvis faktorerne antages at være kendte, er modellen i (5) blot en multivariat lineær regressionsmodel. Parametrene i en sådan model ( $\Lambda$ ) kan estimeres med mindste kvadraters metode.

Standardiseres alle serierne i  $X$ , bliver residualvariansen (approksimativt) ens på tværs af ligningerne, hvorved mindste kvadraters metode indebærer, at man skal løse følgende minimeringsproblem:

$$\min_{\Lambda} \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T e_{it}^2 = \frac{1}{NT} \text{tr}(e'e) = \frac{1}{NT} \text{tr}\{(X - F\Lambda)'(X - F\Lambda)\} \quad (6)$$

hvor  $\text{tr}()$  er sporet af matricen i parentes. Løsningen for  $\Lambda$  kan findes med almindelig regression:

$$\hat{\Lambda}'(F) = (F'F)^{-1} F'X \quad (7)$$

Denne estimator kan sættes ind i (6), hvilket giver følgende udtryk, der skal minimeres med hensyn til  $F$ :

$$\begin{aligned} \frac{1}{NT} \text{tr}(e'e) &= \frac{1}{NT} \text{tr}\{X - F(F'F)^{-1} F'X\}'(X - F(F'F)^{-1} F'X)\} \\ &= \frac{1}{NT} \text{tr}(XX') - \frac{1}{NT} \text{tr}(F'XX'F (F'F)^{-1}) \end{aligned} \quad (8)$$

---

2. Korrelation mellem de idiosynkratiske led blev behandlet af Chamberlain og Rothschild (1983) og af Connor og Korajczyk (1986, 1993). Disse forfattere viste, at der kan tillades en vis korrelation i de idiosynkratiske led, når antallet af serier,  $N$ , går mod uendelig. Deres faktormodeller, som anvendes i finansieringslitteraturen, benævnes ofte approksimative faktormodeller.

Det fremgår, at minimering af den normerede residualkvadratsum er ækvivalent med maksimering af andet led i anden linie.

Det er nødvendigt at indføre en normalisering af konjunkturfaktorerne, og (8) viser, at  $F'F = T \times I_r$  er en god normalisering, da maksimering af andet led dermed reduceres til:

$$\begin{aligned} \max_F \quad & \text{tr}[F'(\frac{1}{N} XX') F] & (9) \\ \text{givet} \quad & F'F / T = I_r \end{aligned}$$

Løsningen til dette maksimeringsproblem er givet i f.eks. Rao (1973, afsnit 1f.2), og indebærer, at  $F$  estimeres ved de  $r$  første egenvektorer i  $T \times T$ -matricen  $\frac{1}{N} XX'$ , når egenvektorer og egenverdier er ordnet efter egenverdiernes størrelse.

Det er værd at bemærke, at man med denne procedure skal finde egenvektorer i en symmetrisk  $T \times T$ -matrix, uanset hvor mange serier der indgår i faktormodellen (4). Det betyder, at man uden beregningsmæssige problemer kan introducere nærmest uendeligt mange serier i datagrundlaget for beregningen af diffusionsindeksene.

Da faktormodellen har en bi-lineær struktur, kan man egentlig ikke identificere de enkelte konjunkturfaktorer, og disse vil typisk være linearkombinationer af de estimerede faktorer. Identifikationsproblemet løses ved betingelsen  $F'F / T = I_r$ , men denne normalisering er en ren matematisk forenkling, som ikke – i økonomisk forstand – identificerer de enkelte konjunkturkomponenter. Det er derfor ikke nødvendigvis meningsfyldt at tale om en bestemt faktor som f.eks. en real konjunkturfaktor og en anden faktor som en nominel konjunkturfaktor. Normaliseringen giver dog ortogonale konjunkturfaktorer, som er ordnet efter deres betydning for den samlede variation i  $X$ -matricen. Det er i denne sammenhæng vigtigt at være opmærksom på, at faktorerne kan »bytte plads«, når de estimeres på forskellige perioder.

I forhold til almindelige indikatormodeller med observerbare indikatorer er gevinsten i den diffusionsindeksbaserede forudsigelse, at man ikke behøver indskrænke antallet af mulige indikatorer på forhånd. Alle serier med mulig information om konjunkturudviklingen kan inkluderes i faktormodellen (4) og kan dermed bidrage til en bedre estimation af konjuncturelementet i forudsigelsesligningen (1). Hvis f.eks.  $N = 246$  og  $r = 20$ , som i dette studie, har man reduceret den førømtalte indikator-søgeproces til en afsøgning blandt 20 serier frem for 245 serier. Endelig har diffusionsmodellen den fordel, at man relativt enkelt kan inddrage serier, hvor visse observationer mangler, som f.eks. kvartalsserier i forudsigelser af månedsserier. Behandlingen af manglende observationer er nærmere beskrevet i Dahl m.fl. (2005a).

Tabel 1. Datagrundlag for beregning af diffusionsindeks.

Variabelgruppe	Antal serier		
	Måned	Kvartal	I alt
1. Produktion og konjunkturbarometer for industri mv.	9	14	23
2. Beskæftigelse og ledighed	5	20	25
3. Detailomsætning og industriomsætning	16	0	16
4. Forbrug og bilkøb	2	10	12
5. Byggeri, boligmarked og konjunkturbarometer for byggeri	14	3	17
6. Lagerinvesteringer	0	5	5
7. Ordreindgang og -beholdning	30	0	30
8. Renter, valutakurser, aktiekurser mv.	29	0	29
9. Pengemængde mv.	4	0	4
10. Priser	25	9	34
11. Løn	0	7	7
12. Import, eksport, udenlandske data, tillidsindikatorer mv.	38	6	44
I alt	172	74	246

Anm.: En præcis databeskrivelse findes i Dahl m.fl. (2005a).

### 3. Datagrundlag og beregnede fællesfaktorer

Datagrundlaget for beregningen af diffusionsindeks for danske data består af 246 tidsserier – 172 månedsserier og 74 kvartalsserier. Tidsserierne dækker centrale områder af dansk økonomi, jf. tabel 1, og går for de fleste seriers vedkommende tilbage til 1980'erne.

Datagrundlaget inkluderer både serier, der kan siges at være af interesse i sig selv (f.eks. arbejdsløshed, forbrugerpriser og BNP), og serier, der vurderes at være relevante for at kunne beskrive og forudsige de i sig selv interessante serier (f.eks. konjunkturbarometre). Udvalget af data er i praksis foretaget med udgangspunkt i en databank bestående af ca. 1.300 månedsserier og ca. 1.300 kvartalsserier, der dækker centrale danske og internationale økonomiske nøgletal. Kilderne til denne databank er Danmarks Statistiks Tidsserie Databank (DSTB) samt EcoWin.

Grupperingen af variablerne i tabel 1 er inspireret af Stock og Watson (1998 og 2002a), og der er ved udvælgelsen af data tilstræbt en vis »balance« mellem variabelgrupperne. Andre forhold, der har spillet en rolle ved udvælgelsen af data er tidsserierne længde, frekvens og det typiske publikationslag. Som et eksempel på serier, der ikke indgår i datasættet pga. for kort længde, kan nævnes Serviceerhvervenes konjunkturindikatorer. Disse serier må oplagt formodes at indeholde information om den aktuelle konjunktursituation, men de er først offentliggjort fra april 2000. Kun undtagelsesvist er der medtaget tidsserier, der ikke er tilgængelige tilbage til i hvert fald 1988, hvor det officielle kvartalsvise nationalregnskab begynder.

Valget af hvilke serier, der skal indgå i datagrundlaget, er baseret på et mere eller mindre subjektivt valg. Principielt kan metoden, som nævnt, inddrage nærmest uendeligt mange data, da egenværdiproblemet jo relaterer sig til en  $T \times T$ -matrix og ikke en  $N \times N$ -matrix, jf. afsnit 2. Den grundliggende tankegang tilsiger, at man bør inddrage mange serier, idet logikken er: Jo flere serier, jo mere information om de underliggende, fælles bevægelser, der skal fanges i fællesfaktorerne. Imidlertid indebærer flere serier også mere støj, og det er ikke oplagt, at signal/støj-forholdet forbedres, jf. bl.a. Boivin og Ng (2003). Introduktion af serier, der er præget af ekstraordinære hændelser som f.eks. strejker eller ekstreme naturgivne forhold, vil bidrage til en ny type variation, der ikke nødvendigvis er relevant i relation til at beskrive de underliggende konjunkturelle bevægelser i de centrale dataserier. Tilsvarende vil serier, der er helt perifere i forhold til de overordnede konjunkturbevægelser, udelukkende bidrage med støj. En sådan støj, der er totalt irrelevant – og dermed ukorreleret med de fælles konjunkturbevægelser – vil principielt ikke indgå de første centrale fællesfaktorer – netop fordi støjen er ukorreleret med konjunkturserierne, men med et begrænset antal relevante serier (og en endelig tidsperiode) vil selv irrelevant støj i praksis kunne påvirke resultatet.<sup>3</sup>

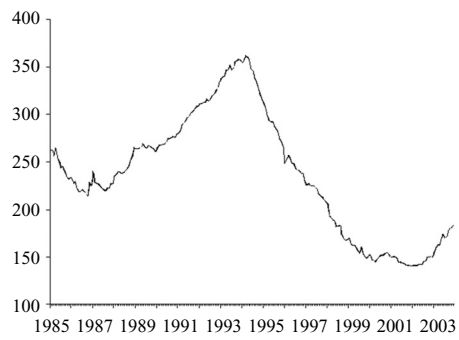
Som udgangspunkt er alle serier sæsonkorrigerede enten fra primærkilden eller ved egen brug af X11. Undtaget herfra er renter samt valuta- og aktiekurser, som det er valgt ikke at sæsonkorrigere. Serierne er gennemgået for oplagte databrud, og i nogle tilfælde er serierne blevet forlænget ved kædning. Serierne er (efter sæsonkorrektion) om nødvendigt transformeret for at opnå stationaritet. De fleste serier indgår i logaritmiske ændringer. Nogle serier indgår dog i absolutte ændringer – bl.a. ledighed og renter, mens andre serier indgår utransformerede – bl.a. forbrugerforventninger, konjunkturindikatorer, lagerinvesteringer og rentespænd. Formelle test af stationaritet indikerer, at de således transformerede serier i det store og hele kan betragtes som stationære, jf. Dahl m.fl. (2005a).

I figur 1 vises tre centrale variabler, der i de følgende afsnit vil blive forsøgt forudsagt med udgangspunkt i diffusionsindekset. Det fremgår af figurerne til højre, at de ændringer, som indgår i beregningen af diffusionsindekset – og som er de størrelser, der skal forudsiges – er ganske volatile størrelser.

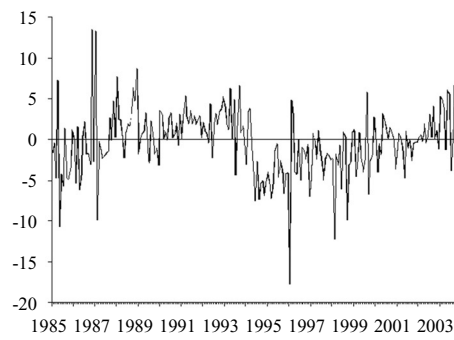
På baggrund af det beskrevne datagrundlag er der beregnet et antal fællesfaktorer, som anvendes i det videre arbejde om forudsigelser. Den betydelige volatilitet i serier-

---

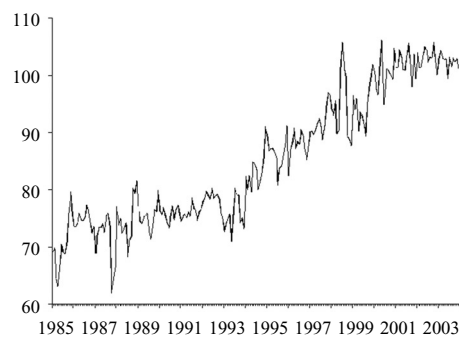
3. I Dahl m.fl. (2005a) vises bl.a. betydningen af at anvende et markant mindre datagrundlag – konkret 100 udvalgte serier mod 246 i de her viste resultater. Overordnet påvirkes forudsigelsesegenskaber ikke meget. Dette kan opfattes som en indikation af, at metoden er relativt robust over for at medtage mange (og måske potentielt irrelevante) serier. Omvendt kan de små forskelle naturligvis også tages som udtryk for, at der ikke er den store gevinst ved at medtage et stor antal serier i datagrundlaget for beregningen af fællesfaktorerne.



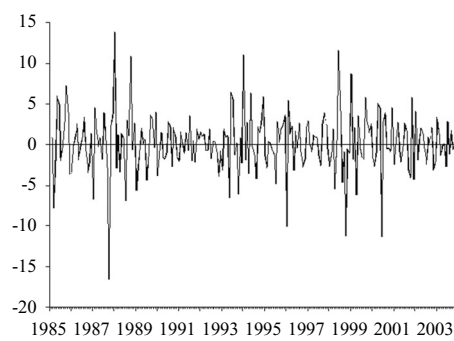
Ledighed, 1.000 personer



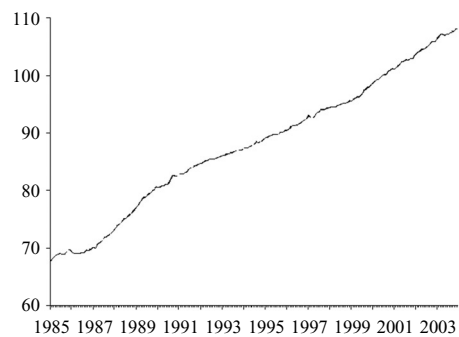
Ændring i ledighed



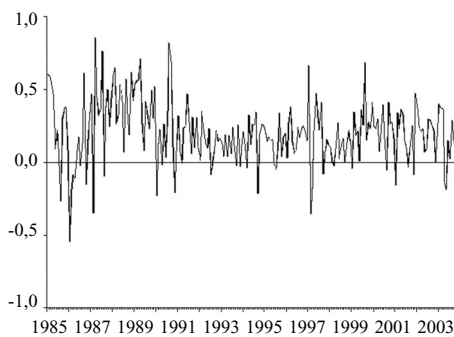
Industriomsætning, faste priser, 2000 = 100



Procentvis ændring i industriomsætning

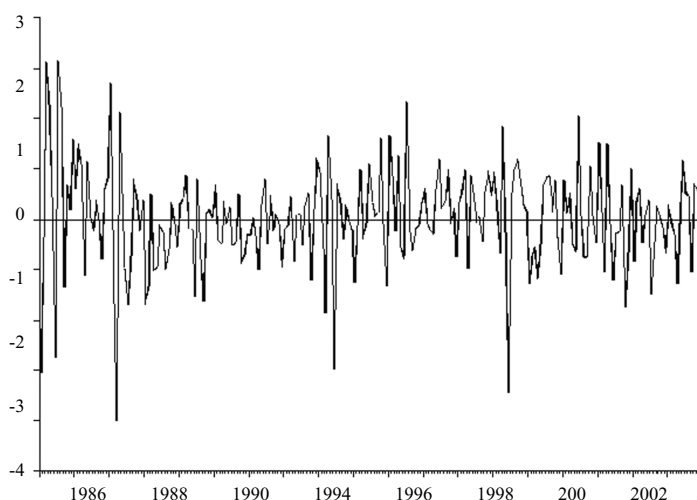


Nettopriser, 2000 = 100



Procentvis ændring i nettopriser

Figur 1. Centrale variabler, der skal forudsiges.



Figur 2. Den første fællesfaktor.

ne, der fremgår af figur 1, bidrager naturligt til, at de resulterende fællesfaktorer ligeledes bliver relativt volatile, jf. figur 2.

Som nævnt er fællesfaktorerne per konstruktion ordnet sådan, at den første fællesfaktor forklarer den største del af variationen i de underliggende data, den anden forklarer næstmest osv. For hele perioden forklarer den første fællesfaktor ca. 10 pct. af variationen i data. De seks første forklarer ca. 42 pct., de ti første ca. 55 pct. og de 20 første ca. 72 pct. af variationen i de underliggende data.<sup>4</sup>

Når fællesfaktorerne er beregnet på hele perioden (1985m1-2003m12), er faktorerne en funktion af data i hele perioden. For at få et indtryk af de estimerede faktorerers sampleafhængighed kan man f.eks. sammenligne den første fællesfaktor, beregnet på hele perioden, med den første fællesfaktor, beregnet på et del-sample.  $R^2$  mellem den første fællesfaktor beregnet på den fulde periode og beregnet på perioden frem til 2000 er 0,80. Den høje korrelation må dog nok mest opfattes som et tilfælde, idet fællesfaktorerne, som beskrevet i afsnit 2, kan »bytte plads«. Dette eksemplificeres af, at den kvadrerede korrelation mellem fællesfaktor nummer 2 beregnet på fuld periode og faktoren beregnet frem til 2000 kun er 0,02.

I tabel 2 vises, hvorledes hver enkelt fællesfaktor beregnet til og med 1994m1 hhv. 2000m1 er korreleret med alle fællesfaktorerne beregnet på fuld periode, dvs. frem til og med 2003m12. Denne beregning er ikke påvirket af fællesfaktorerne præcise ræk-

4. Stock og Watson (1998) angiver til sammenligning, at de første seks fællesfaktorer forklarer 47 pct. af variationen i deres underliggende datasæt (171 amerikanske tidsserier), og Artis m.fl. (2004) angiver, at de seks første faktorer forklarer ca. 50 pct. af variationen (81 serier fra UK).



Tabel 2. Kvadreret korrelation mellem fællesfaktorer beregnet på fuld periode og delperioder.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	Alle
1994m1	92	82	70	87	86	90	75	93	86	79	84	80	60	80	70	84	71	77	50	45	77
2000m1	94	80	94	82	91	78	82	87	93	94	93	90	96	77	71	91	88	61	73	78	85

Anm.: Tabellen viser  $100 \cdot R^2$  i regressionen af den enkelte fællesfaktor beregnet op til og med tidspunktet, der er angivet i forspalten, på en konstant og alle 20 fællesfaktorer beregnet på hele perioden (dvs. til og med 2003m12). Den sidste søjle, benævnt »Alle« indeholder den kvadrerede trace-korrelation, som kan opfattes som en multipel  $R^2$ .

kefølge. Det fremgår af tabellen, at typisk 80-95 pct. af den variation, der ligger i fællesfaktorerne beregnet på en kortere periode bibeholdes i det rum, der udspændes af fællesfaktorerne beregnet på fuld periode. I tabellens sidste søjle er vist korrelationen mellem det samlede rum udspændt af de 20 fællesfaktorer beregnet for hhv. delperioden og hele perioden. Samlet må tabel 2 tages som udtryk for, at der synes at være en rimelig stabilitet i det variationsområde, der beskrives af fællesfaktorerne, når beregningsperioden varieres.

#### 4. Forudsigelser på baggrund af diffusionsindekset

I dette afsnit præsenteres forudsigelser af tre månedsserier på baggrund af diffusionsindekset – og til sammenligning på baggrund af en almindelig autoregressiv model. Indledningsvis beskrives fremgangsmåden bag forudsigelsessammenligningen. Herefter vises resultaterne af forudsigelserne. Perioden er 1995m1-2003m12, og der betragtes to forudsigelseshorisoner: 6 og 12 måneder. Endelig diskuteres betydningen af en række parametervalg.

##### 4.1 Nærmere om fremgangsmåden

Vurderingen af diffusionsindekset baseres i det følgende primært på modellens egenskab til at forudsige *ex ante*, dvs. »out-of-sample«. Der lægges derved mindre vægt på modellernes evne til at beskrive data i estimationsperioden. Den anvendte metode forsøger at tilnærme vurderingen til de omstændigheder, man som prognose-mager står overfor i realtid, herunder ikke mindst, at man ikke kender den sande model eller dens parametre.

Begrebet »out-of-sample« indebærer, at man rullende estimerer en forudsigelsesmodel på baggrund af data, der antages kendt på tidspunkt  $t$ , og anvender denne model til at forudsige en given variabel på et tidspunkt i fremtiden  $t+h$ . Den forudsigelsesmodel, der benyttes i det følgende, kan skrives som:

$$y_{t+h} - y_t = \alpha^h + \sum_{i=0}^{Ymax} \beta_i^h (y_{t-i} - y_{t-i-1}) + \sum_{k=0}^{Fmax} \sum_{j=0}^{Flag} \delta_{kj}^h F_{k, t-j} + \varepsilon_{t+h} \quad (10)$$

Dette er en autoregressiv, fordelt lag model, hvor der inkluderes op til  $Ymax$  lags af den variabel, som skal forudsiges, og op til  $Fmax$  konjunkturfaktorer, idet hver af konjunkturfaktorerne tillades at indgå med op til  $Flag$  lags.

Det bemærkes, at forudsigelserne genereres ved almindelig regression uanset tids-horisonten (en såkaldt multi-step forudsigelse, når  $h$  er større end 1), hvilket indikeres af toptegnet  $h$  på alle parametrene. Multi-step forudsigelsen er et alternativ til »almindelige« forudsigelser med dynamiske modeller, hvor der dannes én-periode-forudsigelser, som herefter itereres frem til den ønskede horisont. Der kan være fordele ved multi-step forudsigelser, hvis forudsigelsesmodellen ikke er identisk med den data-genererende proces (jf. Clements og Hendry (1998) kap. 11), men i diffusionsindeks-modellen er den primære årsag, at man undgår at skulle opstille forudsigelsesmodeller for konjunkturfaktorerne.

I »out-of-sample«-forudsigelsesanalyser anvendes den estimerede forudsigelses-model og data op til og med tidspunkt  $t$  til forudsigelse af  $y$  på tidspunkt  $t+h$ . Herefter rykkes tidsangivelserne én periode frem, estimationerne gentages, og  $y_{t+1+h}$  forudsiges. Denne beregningsgang gentages, idet tidsangivelserne stadigt rykkes frem. Til sidst har man en hel tidsserie af »out-of-sample«-forudsigelser.

Ved forudsigelser baseret på diffusionsindeks indskydes et ekstra trin, der estimerer disse indeks. Beskrevet lidt mere præcis består hvert enkelt trin i »out-of-sample«-forudsigelserne af følgende elementer:

1. Beregn fællesfaktorer (diffusionsindeks) ved hjælp af data op til og med tidspunkt  $t$ . Konkret beregnes 20 fællesfaktorer. Alle data op til tidspunkt  $t$  antages kendt, og de beregnede indeks betegnes  $F_{kt}$ ,  $k = 1, \dots, 20$ .
2. Estimér forudsigelsesligningen (10) for perioden  $T_0$  til  $t-h$ , hvor  $T_0$  er estimations-periodens start ved hjælp af OLS. For hver enkelt estimationsperiode foretages en generel-to-specific nedtestning baseret på BIC. Brugen af BIC-kriteriet indebærer i praksis, at de variabler, der indgår i den foretrukne specifikation af forudsigelsesligningen, har numeriske  $t$ -værdier på omkring  $2^{1/2}$  eller større.
3. Anvend de estimerede parametre fra den nedtestede model samt data op til og med periode  $t$  til at danne en forudsigelse af  $y$  på første forudsigelsestidspunkt,  $t+h$ .

Herefter øges  $t$  med én og beregningerne i trin 1-3 gentages indtil sluttidspunktet,  $t=T-h$ , hvorved man sluttelig har en serie af forudsigelser,  $y_{t+h}^e$ , med  $t = t_1, \dots, T-h$ .

Afslutningsvis anvendes de faktiske og de forudsagte værdier af  $y$  til at beregne den gennemsnitlige kvadrerede forudsigelsesfejl (MSFE):

$$MSFE = \frac{1}{T-h-t_1+1} \sum_{t=t_1}^{T-h} (y_{t+h} - y_{t+h}^e)^2 \quad (11)$$

Til sammenligning med forudsigelser baseret på diffusionsmodellen beregnes i det følgende også forudsigelser baseret på almindelige autoregressive modeller, dvs. forudsigelsesmodellen (10) med alle  $\delta_{kj}$  lig med 0. Disse forudsigelser er ligeledes baseret på »out-of-sample«-metoden.

I appendikset vises som eksempel nogle få af de estimerede forudsigelsesligninger for industriens omsætning, hvor forudsigelsehorisonten er 12 måneder.

Metoden er beskrevet mere indgående i Dahl m.fl. (2005a).

#### 4.2 Forudsigelse af ledighed, industriomsætning og nettopris 6 og 12 måneder frem<sup>5</sup>

Tabel 3 viser resultaterne af forudsigelser baseret på diffusionsindekset (»DI-model«). Til sammenligning vises resultaterne baseret på en autoregressiv model (»AR-model«). For hver variabel/horisont er angivet MSFE, for de to modeller, og den relative MSFE er beregnet. For hver model er angivet resultatet dels for en »optimal« parametrisering (hvor parametriseringen er valgt, sådan at MSFE minimeres), dels for en på forhånd fastlagt parametrisering. Med parametrisering menes her den maksimale laglængde for den endogene variabel ( $Y_{max}$  i ligning (10)), det maksimale antal fællesfaktorer ( $F_{max}$ ) samt den maksimale laglængde af fællesfaktorerne ( $Flag$ ).

En relativ MSFE mindre end 1 indikerer, at diffusionsmodellens forudsigelser i gennemsnit er mere præcise end forudsigelserne fra den autoregressive model. Det fremgår af tabel 3, at med den optimale parametrisering (dvs. hvor parametriseringen er valgt, så MSFE minimeres) er den relative MSFE i alle tilfælde mindre end 1. Dette er udtryk for, at det er muligt at specificere en forudsigelsesligning (10), hvor diffusionsindeksmodellen giver bedre out-of-sample-forudsigelser end den autoregressive model. Samtidig viser tabellen imidlertid også, at diffusionsindeksmodellens forudsigelser ikke kan forventes at være bedre, hvis man på forhånd lægger sig fast på givne værdier af  $Y_{max}$ ,  $F_{max}$  og  $Flag$ . I dette tilfælde må det konstateres, at diffusionsindeksmodellen rent faktisk giver dårligere resultater end den autoregressive model. I det følgende afsnit diskuteres implikationen af disse resultater nærmere.

5. I Dahl m.fl. (2005a) vises tilsvarende resultater for forudsigelser for en række andre månedsserier samt forudsigelser med en horisont på 1 måned.

Tabel 3. MSFE fra forudsigelser for perioden 1995m1-2003m12.

Parametrisering:	DI-model		AR-model		Relativ (DI/AR)	
	Optimal	Fast	Optimal	Fast	Optimal	Fast
<i>Ændring i ledighed</i>						
6 måneder	144,3	168,8	152,8	152,9	0,94	1,10
12 måneder	772,5	873,3	871,4	871,4	0,89	1,00
<i>Vækst i industriens omsætning i faste priser</i>						
6 måneder	14,9	16,7	15,5	16,2	0,96	1,03
12 måneder	19,9	22,6	22,0	22,6	0,90	1,00
<i>Vækst i nettopriserne</i>						
6 måneder	0,160	0,201	0,183	0,201	0,87	1,00
12 måneder	0,381	0,557	0,505	0,526	0,75	1,06

Anm.: Med »fast« parametrisering menes her  $Fmax = 4$ ,  $Ymax = 4$  og  $Flag = 0$ . Med »optimal« parametrisering menes, at der er søgt over en række værdier af  $Fmax$ ,  $Ymax$  og  $Flag$ ; værdierne af disse parametre er valgt, så MSFE for den betragtede »out-of-sample«-periode minimeres, jf. beskrivelsen i afsnit 4.3.

#### 4.3 Om valg af parametre mv.

I forbindelse med praktisk anvendelse af diffusionsindeksmodellen er der en lang række forhold, en prognosemager skal tage stilling til. Blandt disse forhold er

- (1) Valg af grundliggende data,
- (2) Valg af startperiode for dannelse af fællesfaktorer og estimation af forudsigelsesligning,
- (3) Valg antal faktorer i faktormodellen,
- (4) Valg af maksimalt antal fællesfaktorer og lags i forudsigelsesligningen,
- (5) Valg af nedtestningsprocedure.

Det har vist sig, at ikke mindst valget af parametre i forudsigelsesligningen, dvs. valget af det maksimale antal fællesfaktorer ( $Fmax$ ), det maksimale antal lags heraf ( $Flag$ ) samt det maksimale antal laggede endogene ( $Ymax$ ), er vigtigt.

I tabel 3 er under betegnelsen »Fast« vist resultaterne af ét konkret valg, hvor  $Fmax$  sættes til 4,  $Ymax$  til 4 og  $Flag$  til 0. Det er oplagt, at andre valg af parametre vil lede til andre resultater.

Herudover vises resultaterne af en systematisk afsøgning efter den »optimale« parametrisering. Konkret er der for diffusionsindeksmodellen for hver variabel/horisont afsøgt over alle kombinationer af  $Ymax = 2, 3, 4, 5, 6$ ,  $Fmax = 4, 6, 8, 10, 12, 15, 20$  og  $Flag = 0, 1$ . Dette indebærer, at der for hver variabel og for hver forudsigelseshorisont er afprøvet 70 parameterkombinationer. For hver kombination er »out-of-sample«-forudsigelserne dannet, og i tabellen er angivet den laveste MSFE ud af disse 70 beregninger.

Tabel 4. Parametrisering af »optimale« modeller.

	Horisont	DI-model			AR-model
		$F_{max}$	$Flag$	$Y_{max}$	$Y_{max}$
Ledighed	6	10	0	2	3
Ledighed	12	15	0	2	4
Industriomsætning	6	8	1	6	6
Industriomsætning	12	10	0	3	3
Nettopriser	6	6	0	4	8
Nettopriser	12	6	0	2	2

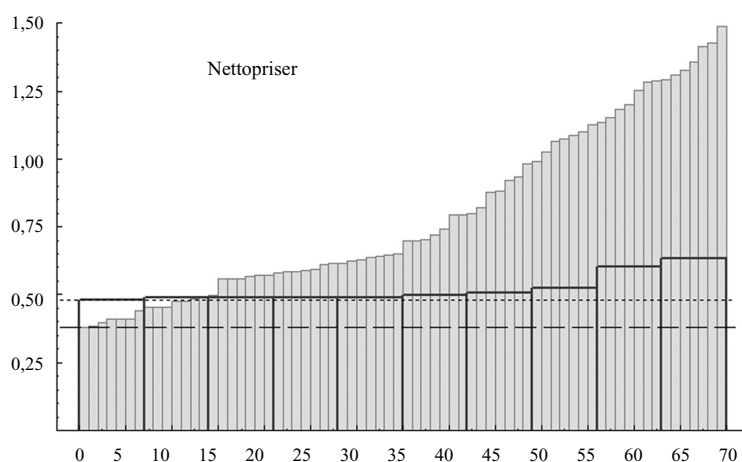
Ann.:  $F_{max}$  angivne det maksimale antal fællesfaktorer i ligning (10).  $Flag$  angiver det maksimale lag i fællesfaktorerne i ligning (10).  $Y_{max}$  angiver det maksimale antal laggede  $y$ 'er i ligning (10).

For den autoregressive model er alle kombinationer af  $Y_{max} = 1, 2, 3, 4, 5, 6, 8, 10, 12, 15$  afprøvet for hver variabel og for hver forudsigelseshorisont. Det skal understreges, at resultaterne, der opnås på denne måde, ikke må opfattes som realistiske (opnåelige i praksis). Det skyldes, at parametriseringen er valgt ved en systematisk afsøgning af »out-of-sample«-egenskaberne, som udvælger den bedste parameterkombination. Dette er gjort for at systematisere undersøgelsen af diffusionsmodellens mulige fordele, men proceduren er selvsagt ikke mulig i reeltidsforudsigelser. De »optimale« parametervalg er angivet i tabel 4.

Det fremgår, at den beskrevne afsøgning typisk giver anledning til, at der tillades relativt mange fællesfaktorer i den generelle specifikation af forudsigelsesligningen. Dette kan tages som udtryk for, at der i visse tilfælde kan være et informationsindhold i at medtage yderligere fællesfaktorer, men det viser sig, at det er forbundet med betydelig risiko for at få særligt dårlige forudsigelser, når antallet af parametre i forudsigelsesligningen øges. Dette gælder til trods for, at der foretages en traditionel nedtestning af estimationsligningen fra generel til specifik.

I figur 3 er for nettopriserne vist alle de beregnede MSFE'er, der er fremkommet som led i den beskrevne afsøgning efter den optimale parametrisering. For DI-modellen er der, som nævnt, afprøvet 70 parameterkombinationer. De resulterende MSFE'er er sorteret, og MSFE fra hver parameterkombination repræsenteres i figuren af én af de smalle, grå søjler. Den første af disse søjler repræsenterer den parameterkombination, der har den laveste MSFE (0,381, jf. tabel 3); denne værdi er i figur 3 markeret med en vandret stiplede linje. Det konkrete valg af parametre svarende til denne søjle fremgår af nederste række i tabel 4.

De bredere, fedt optegnede bokse repræsenterer tilsvarende MSFE'erne fra AR-modellen. Der er, som beskrevet, afprøvet 10 parametriseringer af AR-modellen. Den første af disse søjler repræsenterer den maksimale laglængde,  $Y_{max}$ , der har den lave-



Figur 3. Fordeling af MSFE for afprøvede parameterkombinationer.

Anm.: Forudsigelseshorizonten er 12 måneder. Se i øvrigt teksten for forklaring af figuren.

ste MSFE (0,505; jf. tabel 3). Denne værdi er ligeledes markeret med en vandret stiplet linje; som det fremgår af tabel 4 opnås denne MSFE med  $Y_{max} = 2$ .

Figur 3 afslører, at der er stor forskel på den MSFE, der fremkommer ved den »optimale« parametrisering (den første søjle) og den dårligste parametrisering (den sidste søjle). Det fremgår således, at den bedste parameterkombination, som nævnt, giver en MSFE på 0,38 mens den dårligste parameterkombination giver en MSFE, der er næsten 4 gange så stor: 1,49. Konkret fås dette dårlige resultat, når man tillader op til 20 fællesfaktorer, mod det optimale valg på 6, og generelt karakteriseres søjlerne længst til højre af, at der indgår mange fællesfaktorer i den generelle specifikation af forudsigelsesligningen.<sup>6</sup> Det fremgår endvidere af figuren, at AR-modellen er væsentligt mindre følsom over for variation i parametervalget.

Disse resultater understreger følsomheden over for parametriseringen, dvs. specifikationen af den generelle forudsigelsesmodel. Det er for alle de i tabel 3 betragtede serier muligt – ved en systematisk afsøgning – at finde en specifikation (laglængde eller et antal fællesfaktorer), der gør DI-modellen bedre end AR-modellen. Men det modsatte kan lige så let lade sig gøre. Risikoen for dårligere resultater øges mærkbart, når der tillades mange forklarende variable, og især indebærer kombinationen af mange fællesfaktorer og lags heraf et stort antal parametre i den generelle modelspecifika-

6. Variationen i MSFE for ledigheden er langt mere begrænset end variationen i MSFE for nettoprisindekset (den værste modelspecifikation er 35 procent større end den bedste). Variation i MSFE for industriomsætningen er midt i mellem, idet den dårligste MSFE er dobbelt så stor som den bedste. Se Dahl m.fl. (2005a).

tion, hvilket selvsagt øger risikoen for »over-fit« og dermed dårlige »out-of-sample«-egenskaber.<sup>7</sup>

Som diskuteret i Dahl m.fl. (2005a), er det meget vanskeligt – for ikke at sige umuligt – at opstille generelle retningslinjer for valg af parametrene i forudsigelsesligningen. I tabel 3 var som alternativ til »det optimale« (men uopnåelige) valg angivet resultaterne af et mere eller mindre tilfældigt for alle variabler og horisonter fast valg på 4 fællesfaktorer og 4 laggede endogene. Et mere informeret valg kunne bestå i for hver variabel/horizont at finde den optimale parametrisering for perioden 1995-1999, og derefter fastholde denne for perioden 2000-2003. Denne fremgangsmetode indebærer, at der for perioden 1995-1999 gennemføres et stort antal beregninger for en given kombination af variabel og horisont, hvorefter der foretages én »out-of-sample«-beregning for perioden 2000-2003. Tabel 5 viser resultatet af denne beregning.

Tabel 5 viser flere ting. For det første fremgår det ved sammenligning med tabel 3, at forudsigelsesfejlene generelt er mindre for perioden 2000-2003 end for den fulde periode 1995-2003. Dette gælder både for DI- og AR-modellen, men det fremgår, at det relative forhold generelt skifter til ugunst for diffusionsindeksmodellen. Dette indikerer, at den relativt gode performance for diffusionsindekset, når den »optimale« parametrisering betragtes i tabel 3, primært er udtryk for gode forudsigelsesegenskaber i perioden 1995-1999. For det andet fremgår det, at vælges parametriseringen som beskrevet på baggrund af »out-of-sample«-egenskaberne i perioden 1995-1999 (jf. søjlen »fast«, så bliver DI-modellens forudsigelsesegenskaber markant dårligere, end når der vælges parametrisering på baggrund af selve »out-of-sample«-perioden (jf. søjle »optimal«). Derimod er parametriseringen af AR-modellen, som det kunne ventes, ikke af så stor betydning. MSFE øges således kun lidt for AR-modellen, når parametriseringen fastholdes, hvilket indebærer, den relative MSFE klart forskydes til fordel for AR-modellen.

Dette bekymrende resultat er en illustration af, hvor vigtigt designet af en »out-of-sample«-evaluering er. Det bør bemærkes, at en væsentlig årsag til det meget dårlige resultat for DI-modellen er, at det valg af maksimalt antal fællesfaktorer og lags, der foretages på baggrund af perioden 1995-1999, typisk indebærer et meget stort antal variabler i forudsigelsesligningen. I flere tilfælde medtages 10, 15 eller 20 fællesfaktorer og deres lags i den generelle specifikation, hvilket oplagt indebærer en stor risiko for dårlige resultater, jf. ovenfor. Hvis man på forhånd havde begrænset det maksimale antal fællesfaktorer og/eller undladt at tillade lags heraf, ville resultatet ikke have

---

7. Det bemærkes her, at der er blevet beregnet 20 fællesfaktorer til trods for, at det optimale antal er i størrelsesordenen 8 i henhold til et kriterium udviklet af Bai og Ng (2002). Beregninger af »out-of-sample«-forudsigelser baseret på 8 fællesfaktorer giver dog anledning til MSFE'er, der er stort set svarende til de her afrapporterede. En væsentlig forskel er imidlertid, at de dårligste resultater typisk ikke genfindes.

Tabel 5. MSFE fra forudsigelser for perioden 2000m1-2003m12.

Parametrisering:	DI-model		AR-model		Relativ (DI/AR)	
	Optimal	Fast	Optimal	Fast	Optimal	Fast
<i>Ændring i ledighed</i>						
6 måneder	64,6	153,4	56,5	67,4	1,14	2,28
12 måneder	342,2	612,2	237,4	294,4	1,44	2,08
<i>Vækst i industriens omsætning i faste priser</i>						
6 måneder	9,8	15,4	10,2	11,4	0,96	1,35
12 måneder	10,5	19,9	13,1	13,1	0,80	1,52
<i>Vækst i nettopriserne</i>						
6 måneder	0,155	0,168	0,140	0,155	1,11	1,08
12 måneder	0,247	0,453	0,287	0,336	0,86	1,35

Anm.: Med »fast« parametrisering menes her, at  $F_{max}$ ,  $Y_{max}$  og  $Flag$  er fastlagt som de værdier, der minimerer MSFE for perioden 1995-1999, mens »optimal« parametrisering indebærer, at disse parametre er fastlagt, således at MSFE for den betragtede »out-of-sample«-periode minimeres, jf. teksten.

været nær så dårligt. Det skal således understreges, at de specifikationer, der her er valgt på baggrund af forudsigelsesegenskaberne for 1995-1999 ikke nødvendigvis er dem, som man ville have (eller burde have!) valgt, hvis man i 1999 skulle have fastlagt specifikationen for forudsigelsesligningen for de følgende fire år.

Samlet må det på baggrund af ovenstående konkluderes, at selv om det ved en systematisk afsøgning over parametriseringer af den generelle forudsigelsesmodel er muligt at finde en specifikation, hvor MSFE fra DI-modellen er mindre end fra den autoregressive model, så er dette resultat næppe relevant i praktisk anvendelse. Det skyldes, at udledningen af den »optimale« parametrisering er baseret på en afsøgning af »out-of-sample«-egenskaberne, hvilket oplagt ikke er muligt i praksis. Den afslørede store følsomhed over for parametriseringen af forudsigelsesligningen må opfattes som et stort problem, og den nedslående konklusion synes at være, at diffusionsindeksmodellen i den hidtil betragtede form tilsyneladende ikke har meget at byde på, på danske data.

### 5. Forudsigelser baseret på filtrerede data

De data, der hidtil har været anvendt, er, som beskrevet i afsnit 3, typisk sæsonkorrigerede og transformeret til stationaritet ved at tage ændringer (evt. i logaritmer). Hvis man opfatter en given serie som bestående af fire (ikke-observerbare) komponenter – sæson, trend, konjunkturcykel og en irregulær komponent – er der derved fjernet to komponenter, nemlig sæson og trend. De data, der indgår i dannelsen af fællesfaktorerne, kan dermed omfattes som summen af en konjunkturkomponent og en irregulær komponent.



Tankegangen bag diffusionsindekset er, som beskrevet i afsnit 2, at der er et begrænset antal fælles bevægelser i data, og det er naturligt at opfatte disse som konjunkturbevægelserne. I det omfang den irregulære komponent (støjen) i serierne er fremtrædende, vil de beregnede fællesfaktorer være påvirket af denne støj, mens konjunkturbevægelsen (det egentlige signal) trænges i baggrunden.

Som beskrevet er de anvendte data ganske støjfyldte, og et dårligt signal/støj-forhold kan være en del af forklaringen på de knap så gunstige resultater for diffusionsmodellen på danske data relativt til de internationale erfaringer. Eksempelvis er variationen i BNP-vækstraten (målt ved spredningen) ca. dobbelt så høj i Danmark som i USA, og variationen i den danske industriomsætning er ca. syv gange så stor som variationen i den amerikanske industriproduktion. En nærliggende tanke er derfor at forsøge at filtrere støjen ud af serierne – analogt til at trend og sæson er fjernet. Herved vil de underliggende konjunkturbevægelser muligvis stå klarere, og signal/støj-forholdet forøges.

En metode til at dekomponere de trend- og sæsonrensede data i en konjunkturkomponent og en irregulær komponent er at opstille en model for data i såkaldt State Space-form.<sup>8</sup> En dekomponering (eller filtrering) kan herefter foretages af hver enkelt original (sæson- og trendkorrigeret) serie, hvorved der opnås et estimat af den cykliske komponent. Filtringen foretages dermed univariat (serie for serie), hvilket indebærer, at den information om faktorstrukturen, der i øvrigt postuleres i forbindelse med beregningen og anvendelsen af fællesfaktorerne, ikke anvendes.<sup>9</sup>

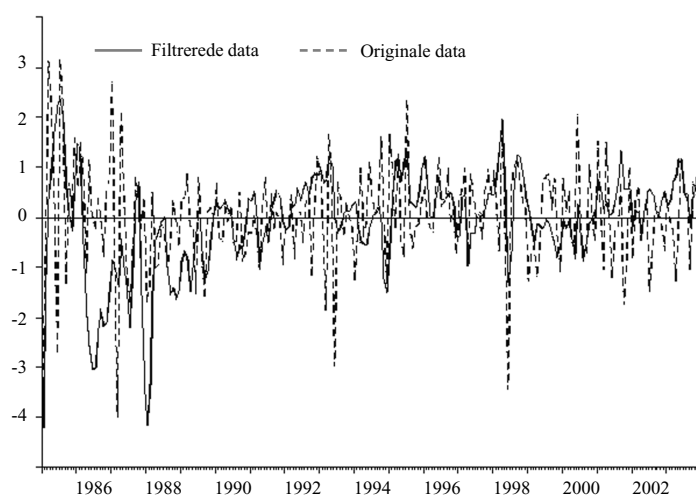
Filtringen af de grundliggende data er foretaget på basis af hele perioden – fra 1980 frem til og med 2003. Herved anvendes i begyndelsen af perioden information, der reelt set ikke var tilgængelig på det relevante tidspunkt. På denne måde vil den estimerede konjunkturkomponent i begyndelsen af perioden afvige fra det estimat, man kunne have lavet, hvis man kun havde anvendt den information, der var tilgængelig i begyndelsen af perioden. Det er oplagt, at man ved at anvende information i hele perioden ved en initial datatransformation giver et for positivt billede af forudsigelsesegenskaberne i forhold til virkeligheden. Bl.a. derfor skal denne analyse primært opfattes som en illustration af den potentielle værdi i at fokusere på konjunkturkomponenten ved forudsigelserne.

Efter filtringen er fællesfaktorerne beregnet på samme måde som tidligere. Et udkomme af proceduren er, at de resulterende fællesfaktorer bliver mere bløde, jf. figur

---

8. Estimationen af de ikke observerbare komponenter foretages ved brug Kalman-filtret. Metoden er udførligt beskrevet i Harvey (1989) og senest af Durbin og Koopman (2001). Beregningerne er foretaget i *Ox* ved hjælp af *SsfPack*, som er beskrevet i Koopman, Shephard og Doornik (1999); *SsfPack* samt dokumentation er frit tilgængelig og kan findes på <http://www.ssfpack.com>.

9. Camacho og Sancho (2001) anvender også filtrerede data i en faktoranalyse af spanske data. Forfatterne viser bl.a., at de spanske data er langt mere volatile end eksempelvis amerikanske data, hvorfor de finder det velbegrundet at fokusere på konjunkturkomponenten.



Figur 4. Første fællesfaktor med originale data hhv. med filtrerede data.

Anm.: Fællesfaktoren med originale data – den stiplede kurve – er identisk med den, der blev vist i afsnit 2.

4, som viser den første estimerede fælles konjunkturfaktor hhv. med originale data og filtrerede data.

Resultatet af »out-of-sample«-forudsigelser baseret på de filtrerede data vises i tabel 6. Det fremgår af tabellen, at MSFE gennemgående er mindre, når fællesfaktorerne beregnes på filtrerede data. Eksempelvis reduceres MSFE for forudsigelsen af ledigheden på 6 måneders horisont til 142,5 med filtrerede data fra 168,8, når de originale data anvendes (jf. tabel 3), hvilket svarer til en relativ MSFE på 0,84.

Sammenlignes med AR-modellens MSFE, jf. tabellens to sidste søjler, ses, at diffusionsindeksmodellen systematisk har de laveste forudsigelsesfejl. Dette gælder både, når det »optimale« (men i praksis uopnåelige) parametervalg anvendes, og når resultaterne af den faste parametrisering betragtes. Især resultatet, der fremgår af den sidste søjle – med fast parametrisering – må opfattes som lovende for diffusionsindeksmodellen. De bedre forudsigelsesegenskaber, der opnås, når støjen (den irregulære komponent) fjernes fra serierne, kan formentlig opfattes som udtryk for, at signalet (konjunkturbevægelsen) går klarere igennem til fællesfaktorerne, og forudsigelsesfejlene derved bliver mærkbart mindre. Gennemsnitligt er MSFE 15 pct. lavere med diffusionsindeksmodellen baseret på filtrerede data end for den autoregressive model, jf. sidste søjle.

Det skal gentages, at den anvendte fremgangsmåde, hvor de grundliggende data filtreres på den fulde periode, kan være problematisk. Det er bl.a. kendt fra trendestimation, at det især i slutningen af den periode, hvor der estimeres, kan være vanskeligt at sikre stabile estimater. Formentlig er problemet ikke helt så stort i forbindelse med

Tabel 6. Filtrerede data. MSFE fra forudsigelser for perioden 1995m1-2003m12.

Parametrisering:	DI-model		Relativ: DI-filter/ DI-original		Relativ: DI-filter/ AR-model	
	Optimal	Fast	Optimal	Fast	Optimal	Fast
<i>Ændring i ledighed</i>						
6 måneder	115,4	142,5	0,80	0,84	0,76	0,93
12 måneder	704,0	730,3	0,91	0,84	0,81	0,84
<i>Vækst i industriens omsætning i faste priser</i>						
6 måneder	13,5	14,1	0,91	0,84	0,87	0,87
12 måneder	15,3	16,3	0,77	0,72	0,70	0,72
<i>Vækst i nettopriserne</i>						
6 måneder	0,171	0,176	1,07	0,88	0,93	0,88
12 måneder	0,470	0,480	1,23	0,86	0,93	0,91

Anm.: Med »fast« parametrisering menes her  $F_{max} = 4$ ,  $Y_{max} = 4$  og  $Flag = 0$ . Der henvises til afsnit 4 for en nærmere beskrivelse af parametrisering og testprocedure.

filtrering på stationære serier, men det udestår at afklare, hvor problematisk det i praksis er, at filtreringen er foretaget på hele perioden frem for at blive foretaget rekursivt.

## 6. Opsamling og perspektiver

Forudsigelser af makroøkonomiske tidsserier som f.eks. ledighed, industriomsætning og inflation indgår i konjunkturvurderinger, der foretages af banker, interesseorganisationer og offentlige institutioner. Det er naturligvis vigtigt at anvende den til rådighed værende information bedst muligt, herunder at anvende de statistiske metoder der giver de bedste resultater.

I nærværende arbejde er det vist, at man med en systematisk søgning over specifikationer af forudsigelsesligningen kan opnå bedre forudsigelser med diffusionsindeksmodellen end med en almindelig autoregressiv model. Den systematiske afsøgning over specifikationer indebærer imidlertid, at information fra »out-of-sample«-perioden anvendes til fastlæggelse af modelspecifikationen, hvilket oplagt er i modstrid med »out-of-sample«-tankegangen. Fastlægges parametriseringen alternativt på forhånd, forværres diffusionsindeksmodellens forudsigelsesegenskaber. Tages udgangspunkt i en given parametrisering, hvor det maksimale antal fællesfaktorer og lags fastlægges på baggrund af en analyse af forudsigelsesegenskaberne i en tidligere periode, klarer DI-modellen sig faktisk dårligere end den autoregressive model. En forklaring på de markant dårligere resultater i dette set-up kan være, at de således fastlagte forudsigelsesligninger indeholder mange parametre (op til 20 fællesfaktorer plus lags heraf).

I betragtning af den store betydning parametriseringen kan have for forudsigelsesegenskaberne er det bekymrende, at det er vanskeligt at angive generelle retningslinjer for det optimale valg af antal fællesfaktorer, antal lags osv. Analyser i Dahl m.fl. (2005a) tyder dog på, at det sandsynligvis er hensigtsmæssigt at begrænse antallet af fællesfaktorer (og lags) betragteligt i forhold til det interval, der er blevet afsøgt i nærværende arbejde (hvor der er anvendt op til 20 fællesfaktorer). Resultaterne i Dahl m.fl. (2005a) peger endvidere på, at introduktion af laggede fællesfaktorer i forudsigelsesligningen næppe er vejen frem – snarere tværtimod.

I lyset af disse resultater må det konkluderes, at selv om det ved en systematisk afsøgning over parametriseringer af den generelle forudsigelsesmodel er muligt at finde en specifikation, hvor MSFE fra DI-modellen er mindre end fra den autoregressive model, så er dette resultat næppe relevant i praktisk anvendelse. Det skyldes, at den »optimale« parametrisering ikke kendes på forhånd. Den afslørede følsomhed over for parametriseringen af forudsigelsesligningen må derfor opfattes som et stort problem.

En bemærkelsesværdig egenskab ved de beregnede fællesfaktorer er, at de er stærkt volatile (uden nævneværdig persistens). Den betydelige støj, der tilsyneladende er til stede i de estimerede fællesfaktorer, indebærer, at den underliggende konjunkturbevægelse, ikke kommer tydeligt frem. Signal/støj-forholdet er for ringe. Dette er et alvorligt problem, fordi det netop er de fælles, underliggende konjunkturbevægelser, der skal drive forudsigelserne i diffusionsindeksmodellen. Hvis der ikke er persistens i fællesfaktorerne, kan det ikke forventes, at de vil have et nævneværdigt bidrag til forudsigelsen af centrale økonomiske variabler. På denne baggrund blev det forsøgt at anvende filtrerede (støjrensede) dataserier ved beregningen af fællesfaktorerne. Dette gav anledning til mærkbart mindre volatile faktorer og bedre forudsigelsesegenskaber for DI-modellen. Forudsigelsesfejlen fra DI-modellen baseret på dette datagrundlag er således ca. 15 pct. bedre end forudsigelser fra den autoregressive model, selv når parametriseringen er fastlagt på forhånd. Den konkrete filtrering kan dog ikke opfattes som uproblematisk, da den baseres på hele perioden, og dermed principielt er i modstrid med »out-of-sample«-tankegangen, og der kan derfor være behov for yderligere analyser baseret på rekursivt filtrerede serier.

Samlet må det konkluderes, at gevinsten ved at anvende DI-modellen på danske tal ikke er så stor, som det er fundet i studier på tal fra bl.a. USA og UK. Samtidig er gevinsten meget afhængig af parametriseringen, og der er ikke fundet gode retningslinjer for fastlæggelsen heraf. Derved afhænger resultaterne af, hvorledes »out-of-sample«-beregningerne designes, og der kan derfor være god grund til at følge DI-modellens forudsigelser i ægte realtidsforudsigelser, hvor konkrete valg af parametre, nedtestningsstrategi, datarevisioner og publikationslag vil spille en rolle.

Det er ikke muligt på baggrund af nærværende arbejde at konkludere, hvorfor DI-

modellen ikke klarer sig så overbevisende på danske data, som det er blevet demonstreret i litteraturen på data fra andre lande. Den mest oplagte forklaring er nok, at de danske tal tilsyneladende er mere støjfyldte. Dette gør dem naturligvis sværere at forudsige, men nok så væsentligt betyder støjen, at »signalet« i de beregnede fællesfaktorer bliver for ringe. Det er formentlig på denne baggrund, at de relativt opløftende resultater med filtrerede data skal ses. Det er ikke muligt på baggrund af nærværende arbejde at konkludere, hvorfor DI-modellen ikke klarer sig så overbevisende på dansk data, som det er blevet demonstreret i litteraturen på data fra andre lande. Den mest oplagte forklaring er nok, at de danske tal tilsyneladende er mere støjfyldte. Dette gør dem naturligvis sværere at forudsige, men nok så væsentlig betyder støjen, at »signalet« i de beregnede fællesfaktorer bliver for ringe. Det er formodentlig på denne baggrund, at de relativt opløftende resultater med filtrerede data skal ses. Indledende arbejde med forudsigelser af kvartalsserier bekræfter de positive resultater med filtrerede data. Analyser af »out-of-sample«-forudsigelser af serier fra det kvartalsvise nationalregnskab, herunder BNP og den samlede beskæftigelse, viser således, at hvor diffusionsindeksmodellen på ikke-filtrerede data kun dårligt kan slå forudsigelserne, så giver anvendelsen af filtrerede data anledning til, at forudsigelsesfejlen reduceres med op til 30-35 pct. i forhold til forudsigelser fra diffusionsindeksmodellen på ikke-filtrerede data og forudsigelser fra en simpel, autoregressiv model, jf. Dahl m.fl. (2005b).

#### Litteratur

- Anderson, T. W. 1984a. *An Introduction to Multivariate Statistical Analysis* (Second Edition). New York: John Wiley & Sons.
- Anderson, T. W. 1984b. Estimating linear statistical relationships. *Annals of Statistics*, 12, s. 1-45.
- Angelini, E., J. Henry og R. Mestre. 2001. Diffusion index-based inflation forecasts for the Euro area. Working Paper No. 61. *ECB Working Paper Series*.
- Artis, M., A. Banerjee og M. Marcellino. 2004. Factor forecasts for the UK. Working Paper, Department of Economics, European University Institute. Kommer i *Journal of Forecasting*.
- Bai, J. og S. Ng. 2002. Determining the number of factors in approximate factor models. *Econometrica*, 70, s. 191-221.
- Boivin, J. and S. Ng. 2003. Are More Data Always Better For Factor Analysis? *Working Paper* 9829. National Bureau of Economic Research. Cambridge, MA.
- Brisson, M., B. Campbell og J.W. Galbraith. 2003. Forecasting Some Low-predictability Time Series Using Diffusion Indices. *Journal of Forecasting*, 22, s. 515-31.
- Camacho, M. og I. Sancho. 2003. Spanish diffusion indexes. *Spanish Economic Review*, 5, s. 173-203
- Chamberlain, G. og M. Rothschild. 1983. Arbitrage factor structure, and mean-variance analysis of large asset markets. *Econometrica*, 51, s. 1281-1304.
- Clements, M. P. og D. F. Hendry. 1998. *Forecasting Economic Time Series*. Cambridge: Cambridge University Press.
- Connor, G. og R. A. Korajczyk. 1986. Performance measurement with the arbitrage pricing theory. *Journal of Financial Econometrics*, 15, s. 373-94.
- Connor, G. og R. A. Korajczyk. 1993. A test for the number of factors in an approximate factor model. *Journal of Finance*, 48, s. 1263-91.

- Dahl, C., H. Hansen og J. Smidt. 2005a. Makroøkonomiske forudsigelser baseret på diffusionsindeks. *Arbejdsrapport* 2005:1. Det Økonomiske Råds Sekretariat.
- Dahl, C., H. Hansen og J. Smidt. 2005b. The cyclical component factor model. Upubliceret manuskript.
- Doornik, J. 2001. *Ox An Object-oriented Matrix Programming Language*. London: Timberlake Consultants Ltd.
- Durbin, J., og S. J. Koopman. 2001. *Time Series Analysis by State Space Methods*, Oxford University Press.
- Harvey, A. C. 1989. *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press.
- Koopman, S. J., N. Shephard og J. A. Doornik. 1999. Statistical algorithms for models in state space form using SsfPack 2.2, *Econometrics Journal*, 2, s. 113-66.
- Lawley, D. N. og A. E. Maxwell. 1971. *Factor Analysis as a Statistical Method*. New York: Elsevier Publishing.
- Rao, C. R. 1973. *Linear Statistical Inference and Its Applications* (Second Edition). New York: Wiley & Sons.
- Stock, J. H. og M. W. Watson. 1998. Diffusion indexes. *Working Paper* 6702. National Bureau of Economic Research. Cambridge, MA.
- Stock, J. H. og M. W. Watson. 2002a. Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economics Statistics*, 20, s. 147-162.
- Stock, J. H. og M. W. Watson. 2002b. Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97, s. 1167-79.
- M. Shintani. 2003. Nonlinear forecasting analysis using diffusion indexes: an application to Japan. *Working Paper* No. 03-W22R. Department of Economics, Vanderbilt University.

### Appendiks

Nedenstående regressioner viser nogle eksempler på de estimerede forudsigelsesligninger for industriens omsætning ( $y$ ), hvor forudsigelseshorisonten er 12 måneder. Konkret vises de nedtestede forudsigelsesligninger for det tilfælde, hvor der maksimalt tillades 4 fællesfaktorer og 4 laggede værdier af industriproduktionen i den generelle specifikation.

Regressionerne viser resultatet for estimationsperioden frem til 1994(1), der anvendes til den første »out-of-sample«-forudsigelse af 1995(1), og den sidste estimationsperiode frem til 2002(12), der anvendes til forudsigelse af 2003(12). Resultaterne er angivet ved de estimerede parametre med heteroskedasticitets- og autokorrelationskonsistente  $t$ -værdier i parentes.

Det fremgår af Durbin-Watson teststørrelserne, at der er betydelig autokorrelation i de estimerede ligninger. Dette er ikke overraskende, idet den afhængige variabel er en 12 måneders differens, mens de forklarende variabler er 1 måneders differenser. Autokorrelationen påvirker som bekendt ikke konsistensen af de estimerede parametre, og der er taget højde for autokorrelationen i de beregnede  $t$ -værdier.

Yderligere er det værd at bemærke, at de estimerede diffusionsindeks, der indgår i de to forudsigelsesligninger, har forskellige numre (hhv. F2 samt F3 i den korte regression og F2 samt F4 i den lange regression). Dette betyder ikke nødvendigvis, at der er sket et »strukturelt brud« i ligningerne, idet diffusionsindeksene kan »skifte plads« som beskrevet i teksten.

*Diffusionsindeksmodel med fællesfaktorer beregnet på filtrerede data (svarende til resultaterne beskrevet i afsnit 5). ( $F_{max} = 4$ ,  $Y_{max} = 4$  og  $Flag = 0$ ):*

$$y_{t+12} - y_t = 1,24 - 0,94 \cdot y_t - y_{t-1} - 0,70 \cdot y_{t-1} - y_{t-2} - 0,65 \cdot y_{t-2} - y_{t-3} - 0,41 \cdot y_{t-3} - y_{t-4}$$

$$(1,5) \quad (-3,9) \quad (-2,8) \quad (-3,7) \quad (-2,2)$$

$$+ 1,84 \cdot F2_t - 2,92 \cdot F3_t$$

$$(1,9) \quad (3,4)$$

$$n: 1986(5)-1994(1) \quad s = 4,93 \quad R^2 = 0,45 \quad F(6,86) = 11,9 \quad DW = 0,70$$

$$y_{t+12} - y_t = 2,76 - 0,92 \cdot y_t - y_{t-1} - 0,76 \cdot y_{t-1} - y_{t-2} - 0,64 \cdot y_{t-2} - y_{t-3} - 0,37 \cdot y_{t-3} - y_{t-4}$$

$$(5,0) \quad (-7,5) \quad (-6,5) \quad (-6,1) \quad (-3,6)$$

$$- 1,69 \cdot F2_t + 2,42 \cdot F4_t$$

$$(-3,2) \quad (4,2)$$

$$n: 1986(5)-2002(12) \quad s = 4,75 \quad R^2 = 0,46 \quad F(6,193) = 27,5 \quad DW = 0,71$$

*Autoregressiv model. (Ymax = 4):*<sup>10</sup>

$$y_{t+12} - y_t = 0,70 - 0,42 \cdot y_t \cdot y_{t-1} - 0,37 \cdot y_{t-1} \cdot y_{t-2} - 0,61 \cdot y_{t-2} \cdot y_{t-3}$$

$$(0,8) \quad (-1,7) \quad (-1,6) \quad (-2,9)$$

$$n: 1986(5)-1994(1) \quad s = 5,83 \quad R^2 = 0,21 \quad F(3,89) = 7,8 \quad DW = 0,70$$

$$y_{t+12} - y_t = 2,52 - 0,63 \cdot y_t \cdot y_{t-1} - 0,61 \cdot y_{t-1} \cdot y_{t-2} - 0,60 \cdot y_{t-2} \cdot y_{t-3} - 0,28 \cdot y_{t-3} \cdot y_{t-4}$$

$$(3,7) \quad (-5,1) \quad (-4,9) \quad (-5,3) \quad (-2,9)$$

$$n: 1986(5)-2002(12) \quad s = 5,49 \quad R^2 = 0,27 \quad F(4,195) = 18,2 \quad DW = 0,57$$

---

10. I dette tilfælde – forudsigelse af industriens omsætning med en horisont på 12 måneder – bliver forudsigelsesligningen for diffusionsmodellen på originale data sammenfaldende med AR-modellen, fordi alle de originalt beregnede fællesfaktorer testes ud.



# Interest-Rate Development in Denmark 1875-2003 – A Survey

Kim Abildgren

Danmarks Nationalbank, E-mail: [kpa@nationalbanken.dk](mailto:kpa@nationalbanken.dk)

*SUMMARY: This article paints a broad picture of the interest-rate development in Denmark since 1875. Three different short-term interest-rate series (the official discount rate, the private banks' average deposit rate, and the market rate of discount/money market rate) and two different long-term interest-rate series (the government bond yield and the yield on mortgage-credit bonds) are constructed and some »stylised facts« about the development in real interest rates are presented and discussed.*

---

## 1. Introduction

The existence of financial markets and financial instruments facilitates an efficient allocation of savings from economic agents with a savings surplus to economic agents with savings deficits. Although legal regulations and other institutional factors, to a certain and time-varying degree, have influenced the allocation process, nominal and real interest rates have always played a crucial role for both real investments and financial portfolio decisions.

This article paints a broad picture of the interest-rate development in Denmark since the introduction of the krone as the Danish currency unit in 1875, where Denmark also changed her monetary standard from silver to gold. Furthermore, some »stylised facts« about the development in real interest rates in Denmark are presented and discussed.<sup>1</sup>

The structure of the article is as follows: Section 2 presents the development in Danish interest rates and inflation since 1875. Section 3 explores the development in different measures of real interest rates and looks at inflation expectations, while section

---

This article is based on the work presented in Abildgren (2005). The author is grateful to the editor of this journal and two anonymous referees, whose comments and suggestions have improved the article in the review process. Furthermore, the author wishes to thank colleagues from Danmarks Nationalbank for useful comments. Views and conclusions expressed in this article are those of the author and do not necessarily represent those of Danmarks Nationalbank. The author alone is responsible for any remaining errors.

1. For other long-span studies of interest-rate development in Denmark, one may refer to Andersen (ed.) (1947), Christiansen and Lystbæk (1994), Møller and Topp (2003), Nielsen and Risager (2001), Oldam (1963), Parum (1999a), Parum (1999b), Pedersen (1930), Statistics Denmark (1969) and Sørensen (1995). For broad studies on the international development in interest rates covering the period since 1875, one may refer to e.g. Bordo and Jonung (1997) and Homer and Sylla (1996).

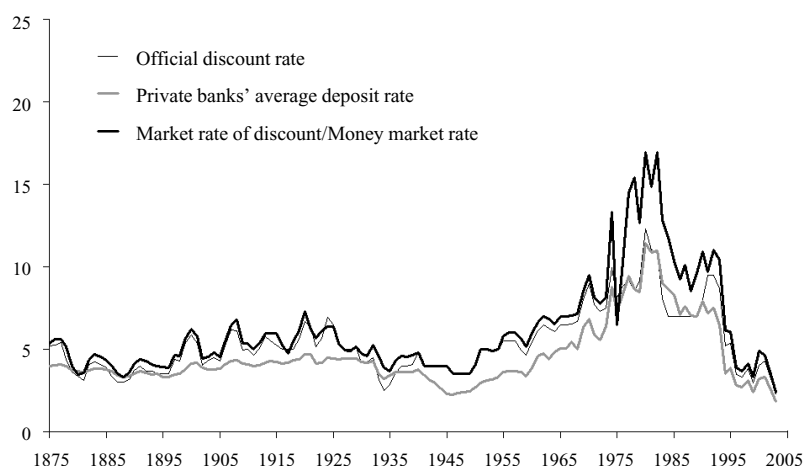


Figure 1. Short-term interest rates 1875-2003, per cent per annum.

Source: Chart 1a in Abildgren (2005).

4 summarises some of the main findings. Finally, annex A describes some issues related to comparability of interest rates across time and outlines the main characteristics of the data set on interest rates constructed for the analysis in the article. The interest rates are listed in annex B.

## 2. Development of nominal interest rates and inflation since 1875 – An overview

Figures 1 and 2 show the development in a range of short-term and long-term nominal interest rates in Denmark since the introduction of the krone as the Danish currency unit in 1875 when Denmark also changed her monetary standard from silver to gold. In the period 1875-1945 the average short-term and long-term nominal interest-rate level was around 4 to 5 per cent per annum. An upward trend in nominal interest rates during the 1960s and 1970s was followed by a downward trend during the 1980s and 1990s.

Table 1 presents a range of summary descriptive statistics on nominal interest rates and inflation broken down by sub-periods determined by in the Danish exchange-rate policy and the degree of restrictions on cross-border capital mobility.<sup>2</sup>

During the Classical Gold Standard period 1875-1913 Denmark participated in the Scandinavian Currency Union based on gold together with Sweden and (from 1877) Norway. During this period all of Denmark's other main trading partners participated

2. The subdivision by degree of cross-border capital mobility is very rough («free» or «restrictions»). However, in Denmark – like many other countries – the reintroduction of free cross-border capital movement in the post-World War II period has been a gradual process. For a short and fact-oriented chronology of the Danish exchange-rate policy and the development in restrictions on cross-border capital movements since 1875, see Abildgren (2004a).

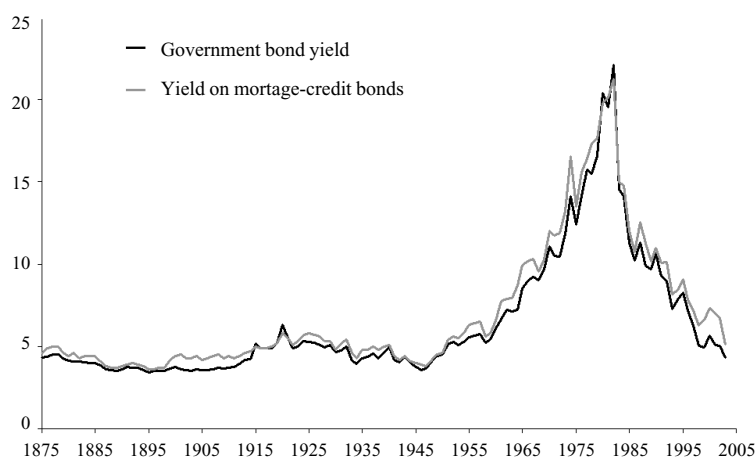


Figure 2. Long-term interest rates 1875-2003, per cent per annum.

Source: Chart 1b in Abildgren (2005).

in the international fixed-exchange-rate Gold Standard system as well. The system was characterised by free movements of capital (including free private import and export of gold in coins and bars). The price level in Denmark was roughly unchanged in the period 1875-1913, and the long-term interest-rate spreads between Denmark and other countries were fairly stable.<sup>3</sup> The Danish government bond yield reached a post-1875 all time low of 3.43 per cent per annum in 1895.

The period 1914-1945 saw rather frequent changes in the monetary regime. World War I *de facto* terminated the Scandinavian Currency Union and the international Classical Gold Standard, and ended free cross-border capital movements. After the War Denmark and its main trading partners gradually returned to the Gold Standard and restored the free cross-border movements of capital,<sup>4</sup> but the system collapsed again after a few years when the UK went off gold in September 1931. Denmark left the Gold Standard later within the same month, and in 1932 a comprehensive exchange-control system was introduced. Apart from a major Danish devaluation in 1933, the Danish krone was pegged rather closely to the British pound for most of the remaining period until the outbreak of World War II. The average Danish inflation rate in the period 1914-1945 was 3.8 per cent, and inflation rates were highly volatile with 10 years of deflation during the period 1921-1933 and a post-1875 all time high rate of inflation

3. Abildgren (2005) contains charts with the short-term and long-term Danish interest-rate spread vis-à-vis Germany, UK, USA, Norway and Sweden since 1875.

4. In January 1927 the Danish krone returned to the Gold Standard at the pre-war parity and the Danish ban on exports of gold in coins and bars was removed vis-à-vis countries with gold-encashment of their currencies.

*Table 1. Interest rates and inflation 1875-2003 – Summary statistics.*

		Market rate of discount/ money market rate			Government bond yield			CPI inflation		
		Mean	Max	Min	Mean	Max	Min	Mean	Max	Min
<i>Sub-periods determined by exchange-rate regime:</i>										
1875-1913	The Classical Gold Standard	4.73	6.78	3.34	3.81	4.50	3.43	0.0	8.5	-10.6
1914-1945	World Wars and inter-war period	5.05	7.31	3.68	4.75	6.30	3.75	3.8	24.4	-15.0
1946-1971	Bretton Woods	5.92	9.50	3.50	6.54	11.07	3.55	4.4	11.7	-0.7
1972-2003	European exchange- rate co-operation	9.19	16.93	2.38	10.93	22.11	4.31	5.6	15.2	1.3
	1972-1978 The Currency Snake	10.84	15.42	6.47	13.45	15.71	10.44	10.1	15.2	6.6
	1979-1998 ERM I	9.98	16.93	3.66	11.52	22.11	5.03	4.9	12.3	1.3
	1999-2003 ERM II	3.74	4.91	2.38	5.01	5.66	4.31	2.5	2.9	2.1
	1972-1986 The devaluation/ »soft peg« period	12.09	16.93	6.47	14.87	22.11	10.20	9.1	15.2	3.6
	1987-2003 The unchanged parity/ »hard peg« period	6.64	11.04	2.38	7.45	11.29	4.31	2.6	4.8	1.3
<i>Sub-periods determined by degree of restrictions on cross-border capital mobility:</i>										
1875-1913	Free cross-border capital movements	4.73	6.78	3.34	3.81	4.50	3.43	0.0	8.5	-10.6
1914-1926	Restrictions on cross-border capital movements	5.92	7.31	4.77	5.14	6.30	4.25	5.5	19.3	-15.0
1927-1931	Free cross- border capital movements	4.87	5.15	4.60	4.91	5.10	4.65	-3.0	-0.6	-5.7
1932-1988	Restrictions on cross-border capital movements	7.26	16.93	3.50	8.33	22.11	3.55	5.7	24.4	-0.7
1989-2003	Free cross-border capital movements	6.28	11.04	2.38	7.04	10.63	4.31	2.4	4.8	1.3
1875-2003	Total	6.16	16.93	2.38	6.36	22.11	3.43	3.2	24.4	-15.0

Source: Abildgren, K. (2005).

at 24.4 per cent per annum in 1940. However, compared with the Classical Gold Standard period, the nominal interest-rate level was only slightly higher and fairly stable.

In the period 1946-1971 Denmark participated in the Bretton Woods fixed exchange-rate system established under the auspices of the International Monetary Fund. The US dollar was the anchor currency of the system. In the late 1940s the UK was still Denmark's largest trading partner and the devaluation of the British pound by 30.5 per cent in September 1949 was followed fully by Denmark. During the 1950s and 1960s Denmark's trade pattern gradually changed towards higher export shares to continental Europe, and the devaluation of the British pound in November 1967 by 14.3 per cent vis-à-vis the US dollar was only followed partly by Denmark (7.9 per cent). During the Bretton Woods period some capital-account transactions (mainly in relation to short-term commercial credits, financial loans and non-financial direct investments) were liberalised but most portfolio investments to and from Denmark still required permission from the Danish monetary authorities. In the Bretton Woods period the average Danish inflation level was only slightly higher than in the period 1914-1945, but during the 1960s there was a sustained upward trend in inflation rates as well as in nominal interest rates.

After the breakdown of the Bretton Woods system in the beginning of the 1970s, the Danish exchange-rate policy became part of the European exchange-rate co-operation, first within the »Currency Snake« founded in 1972 and subsequently from 1979 within the European Exchange Rate Mechanism (ERM).

The post-1971 period also saw a gradual process of deregulation of the remaining Danish restrictions on capital-account transactions. From December 1974 non-residents were given free access to buy Danish krone-denominated exchange-listed bonds (with an original maturity of more than 2 years). However, in February 1979 the free access was abolished again, but it was reintroduced in May 1983. The last restrictions on capital account transactions in Denmark<sup>5</sup> were removed in October 1988.

The oil price shocks of the 1970s and frequent devaluations of the krone during the late 1970s and the beginning of the 1980s caused a continuation of the upward trend in inflation and a widening of the long-term interest spread between Denmark and its main trading partners. Danish government bond yields reached a post-1875 all time high of 22.11 per cent in 1982. The government debt increased rapidly, and a fear that Denmark was on the verge of »state bankruptcy« began to rise. In the beginning of the 1980s the yield on long-term Danish government bonds exceeded the yield on long-

---

5. Mainly concerning money market papers, Danish banks' foreign-exchange loans to residents, loans in kroner to residents from Danish banks' foreign units, private individuals' loans abroad and private individuals' access to open accounts in foreign banks. For a review of the liberalisation cross-border capital movements in Denmark in the period 1950-1985, cf. Danmarks Nationalbank (1986) and Chapter II in Det Økonomiske Råd. Formandskabet (1985).

term Danish mortgage-credit bonds for the first time since the period around World War I, cf. figure 2. Even though a careful interpretation has to be applied<sup>6</sup> this highlights the extent of the crisis in the Danish economy at the beginning of the 1980s.

In September 1982 the incoming Danish government announced the abolition of devaluation as an economic-policy instrument. The Deutsche Mark was revalued several times within the ERM in the period 1982-1987, including vis-à-vis the krone, but not on the initiative of Denmark. The last realignments of the central parity for Danish kroner vis-à-vis Deutsche Mark within ERM occurred in the beginning of 1987. Since then Denmark pursued a »hard« peg against the D-mark and later the euro, despite the widening of the fluctuation bands in the ERM in 1993 and major devaluations by some of Denmark's main trading partners. The increased credibility of the Danish fixed-exchange-rate policy and the international decline of inflation rates during the 1980s and the beginning of the 1990s caused a marked downward trend in both inflation and nominal interest rates in Denmark. The long-term interest spread between Denmark and Germany decreased rapidly from more than 13 per cent in 1982 to less than 1 per cent in 1991 and 0.21 per cent in 2003. The period since 1987 has seen an average inflation level in Denmark of 2.6 per cent, and inflation volatility has been low. In 2003 the Danish short-term interest rate reached a post-1875 all time low, while long-term government bond yields were still above the average level during the Classical Gold Standard.

### 3. Real interest rates and inflation expectations

A »classical« proposition in the theory of finance – the Fisher equation – states that the nominal interest rate approximately equals the sum of the expected inflation and the *ex ante* real interest rate.<sup>7</sup> Other factors such as premiums for interest-rate risk, credit risk, illiquidity, tax treatment<sup>8</sup> *etc* might also influence the nominal interest rate. However, it may still be useful to review the information about the expected inflation and the *ex ante* real interest rate that can be derived from the nominal interest rates using the simple Fisher equation.

Neither expected inflation nor the *ex ante* real interest rate is directly observable – at least not for a time span covering the whole period since 1875. For shorter historical

---

6. Due to the different characteristics of the government bonds and the mortgage-credit bonds from which the yields in figure 2 have been derived.

7. The idea behind this relationship – a distinction between a nominal and a real interest rate – can be traced back at least as far as the works of William Douglas around 1740, cf. Humphrey (1983). In its »modern« form the proposition is mainly associated with Fisher (1896).

8. This article does not cover the importance of tax deductibility in relation to measurement of the real interest rate development. An analysis of the development in Danish real interest rates before and after tax during the period 1953-1984 are found in Chapter V in Det Økonomiske Råd. Formandskabet (1985). The period 1960-2000 is covered by Pedersen (2001).

time-periods one may try to measure inflation expectations more directly from the yield on inflation-index-linked bonds.<sup>9</sup> However, in the case of Denmark such an approach is complicated by a rather illiquid market for index-linked bonds. Furthermore, special tax provisions may distort the results, cf. e.g. Topp (1996) and Hansen (2004).

Another approach could be to try to utilise information about price expectations from consumer surveys, cf. e.g. Christensen (1996) and Knudsen (2002). However, in Denmark such surveys only cover the period since the middle of the 1970s.

A third approach could be to use »independent« (i.e. non-governmental) inflation forecasts from macroeconomic projections as a measure of expected inflation. In Denmark such forecasts are available from the Danish Economic Council since the beginning of the 1960s, cf. Det Økonomiske Råd. Formandskabet (1987).

In order to derive *ex ante* real interest rates for the entire period since 1875 from the nominal interest rates, one needs to make some more crude assumptions regarding inflation expectations. Figure 3 shows three different indicators of the real-interest-rate development in Denmark in the period since 1875:

- The »real short-term interest rate« is measured as the difference between the contemporaneous short-term nominal interest rate and the contemporaneous rate of consumer price inflation.
- The »real long-term interest rate (SE)« is measured as the difference between the contemporaneous nominal long-term interest rate and the contemporaneous rate of consumer price inflation. This corresponds to an *ex ante* long-term real interest rate under the assumption of »static expectations« about future long-term inflation development.
- The »real long-term interest rate (PF)« is measured as the difference between the contemporaneous long-term interest rate and annual average consumer price inflation 7 years ahead.<sup>10</sup> This corresponds to an *ex ante* long-term real interest rate under the assumption of »perfect foresight« (PF) (or »rational expectations«) about future long-term inflation development.

All the three measures for the real interest rate in Denmark show an average around 3 per cent per annum<sup>11</sup> for the period since 1875. Taken at face value the »real long-

9. A commonly used measure for inflation expectations derived from inflation-index-linked bonds is the so-called »break-even inflation«, i.e. the rate of expected inflation at which the return on an inflation-index-linked bond is equal to the return on an equivalent nominal bond.

10. The horizon of 7 years has been chosen because it roughly corresponds to the Macaulay Duration of a 10-year par bullet bond at an interest rate level (6.36 per cent per annum) equal to the average annual yield on Danish government bonds for the period 1875-2003, cf. table 1.

11. The real short-term interest rate is 2.9 per cent per annum, the real long-term interest rate (SE) is 3.1 per cent per annum, and the real long-term interest rate (PF) is 3.1 per cent per annum.

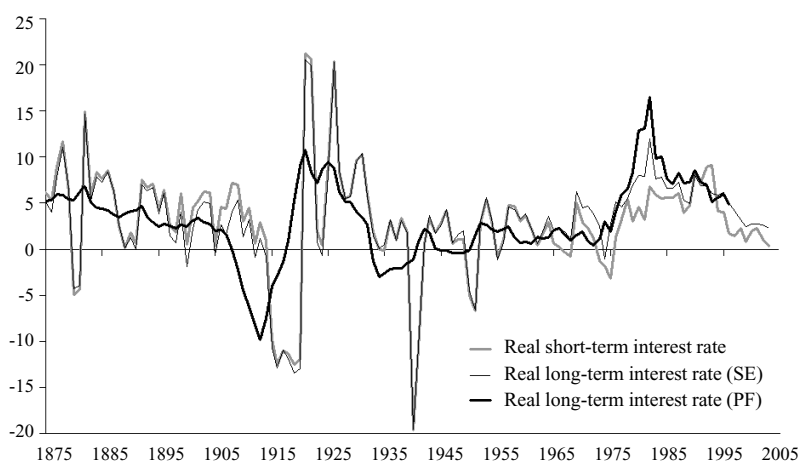


Figure 3. Real interest rates 1875-2003, per cent per annum.

Notes: The »real short-term interest rate (SE)« is measured as the difference between the contemporaneous nominal market rate of discount/money market rate and the contemporaneous rate of consumer price inflation. The »real long-term interest rate (SE)« is measured as the difference between the contemporaneous nominal government bond yield and the contemporaneous rate of consumer price inflation. The »real long-term interest rate (PF)« is measured as the difference between the contemporaneous nominal Government bond yield and the annual average consumer price inflation 7 years ahead. Therefore, the last observation is 1996.

Source: Chart 7a and 8a in Abildgren (2005).

term interest rate (PF)« indicates a rather high level of long-term real interest rates during the late 1980s and the first half of the 1990s.

As an alternative to deriving indicators for the *ex ante* real interest rate from the nominal interest rate, one can try to derive proxies for expected inflation from the nominal long-term interest rate by deducting a measure for the expected real long-term interest rate. Dewald (2003) studies financial market inflation expectations in 13 countries (including Denmark) over the period 1880-2001. One of the measures of the expected real long-term interest rate in each country presented in this study is the country's own 10-year-ahead real GDP growth trend (the »country growth approach«). The underlying argument is the »Golden Rule« within Neoclassical Growth Theory – according to which the steady state real interest rate equals the annual growth rate of real output<sup>12</sup> – combined with an assumption of »perfect foresight« (or »rational expectations«) about future economic growth. Such an approach may be reasonable in periods with restrictions on cross-border capital movements, but less obvious in periods with free cross-border capital movements. In the latter case one would expect real interest rates to be equalised across countries. Dewald (2003) therefore also presents alternative calculations where the expected real long-term interest rate in each country is

12. Cf. e.g. Mankiw (1992) or Barro and Sala-i-Martin (2004) for a presentation of the Golden Rule.



equal to a cross-country average 10-year-ahead real GDP growth trend (the »world growth approach«). However, in practice the two sets of calculations show similar results,<sup>13</sup> and they indicate a rather high degree of persistence in the expected inflation rate compared to the development in the actual inflation rate. This could indicate the presence of a very long learning process in the formation of inflation expectations. Similar results for more recent periods are found in the case of Denmark, cf. e.g. Christensen (1996) and Knudsen (2002) and the references therein. The relatively high long-term real-interest-rate level during the late 1980s and the first half of the 1990s – implied by the »traditional« measures in figure 3 – might, therefore, reflect a high degree of persistence in inflation expectations.

#### 4. Summary of the main findings

The aim of this article has been to paint a broad picture of the interest-rate development in Denmark in the period 1875-2003.

In the period 1875-1945 the average short-term and long-term nominal interest-rate level was around 4 to 5 per cent per annum. An upward trend in nominal interest rates during the 1960s and 1970s was followed by a downward trend during the 1980s and 1990s. During the last five years (1999-2003) the average short-term and long-term nominal interest-rate levels have again been around 4-5 per cent per annum.

Traditional measures of the *ex ante* real interest rate (nominal interest rate less the contemporaneous rate of inflation) show the average short-term and long-term real interest rates in Denmark to be around 3 per cent per annum for the period since 1875. Furthermore, such calculations indicate a rather high long-term real-interest-rate level during the late 1980s and the first half of the 1990s. However, the latter result may reflect a high degree of persistence in inflation expectations.

---

13. Cf. also Bordo and Dewald (2001).

#### References

- Abildgren, K. 2004a. A chronology of Denmark's exchange-rate policy 1875-2003, *Danmarks Nationalbank Working Paper*, No. 12, April.
- Abildgren, K. 2004b. Nominal and real effective krone rate indices for Denmark 1875-2002, *Danmarks Nationalbank Working Paper*, No. 13, April.
- Abildgren, K. 2005. A historical perspective on interest rates in Denmark 1875-2003, *Danmarks Nationalbank Working Paper*, No. 24, February.
- Andersen, P. N. (ed.). 1947. *Laanerenten i Danmark. En teoretisk og historisk undersøgelse med særligt henblik paa renteutviklingen i Danmark*, Copenhagen: Nordisk Livsforsikrings-Aktieselskab og Nordisk Ulykkesforsikrings-Aktieselskab.
- Barro, R. J. and X. Sala-i-Martin. 2004. *Economic Growth, Second Edition*, London: MIT Press.
- Bordo, M. and W. G. Dewald. 2001. Bond Market Inflation Expectations in Industrial Countries: Historical Perspectives, *NBER Working Paper*, No. 8582, November.
- Bordo, M. D. and L. Jonung. 1997. The histo-

- ry of monetary regimes – some lessons for Sweden and the EMU, *Swedish Economic Policy Review*, Vol. 4, p. 285-358.
- Christensen, A. M. 1996. Households' Inflation Expectations, *Danmarks Nationalbank Monetary Review*, November, p. 57-64.
- Christiansen, J. and B. Lystbæk. 1994. Afkast og risiko på aktier og obligationer 1915-1993, *Finans/Invest*, No. 3, p. 10-13.
- Danmarks Nationalbank. 1986. Foreign-Exchange Liberalization and Capital Movements, *Danmarks Nationalbank Monetary Review*, February, p. 9-18.
- Det Økonomiske Råd. Formandskabet. 1985. *Dansk pengepolitik under forvandling – En strukturrapport udarbejdet efter anmodning fra folketingets politisk-økonomiske udvalg*, Copenhagen: Akademisk Forlag.
- Det Økonomiske Råd. Formandskabet. 1987. *Råd og realiteter 1962-1987*, Copenhagen: Det Økonomiske Råd.
- Dewald, W. G. 2003. Bond Market Inflation Expectations and Longer-Term Trends in Broad Monetary Growth and Inflation in Industrial Countries 1880-2001, *ECB Working Paper*, No. 253, September.
- Fisher, I. 1896. *Appreciation and Interest*, New York: MacMillan.
- Hansen, B. W. 2004. Index-Linked Bonds in Portfolio Decisions, *Danmarks Nationalbank Monetary Review*, 2nd Quarter, p. 127-39.
- Hansen, P. H. and H. C. Johansen. 1994. *Det danske finansielle system ca. 1850-1992*, Det nye pengesamfundet. Research on Banking, Capital and Society, Report No. 53, Oslo: Norges Forskningsråd.
- Hansen, S. Aa. and K. E. Svendsen. 1968. *Dansk pengehistorie 1700-1914*, Copenhagen: Danmarks Nationalbank.
- Hoffmeyer, E. 1960. *Strukturaendringer på penge- og kapitalmarkedet. Et studie i anledning af sparekassernes 150 års jubilæum*, Copenhagen: Sparevirkes Forlag.
- Hoffmeyer, E. and E. Olsen. 1968. *Dansk pengehistorie 1914-1960*, Copenhagen: Danmarks Nationalbank.
- Homer, S. and R. Sylla. 1996. *A History of Interest Rates*, Third Revised Edition, New York: Rutgers University Press.
- Humphrey, T. M. 1983. The Early History of the Real/Nominal Interest Rate Relationship, *Federal Reserve Bank of Richmond Economic Review*, May/June, p. 2-19.
- Knudsen, D. 2002. Vurdering af forventet inflation og realrente, *Nationaløkonomisk Tidsskrift*, Vol. 140(1), p. 18-34.
- Mankiw, N. G. 1992. *Macroeconomics*, New York: Worth Publishers.
- Mikkelsen, R. 1993. *Dansk pengehistorie 1960-1990*, Copenhagen: Danmarks Nationalbank.
- Møller, M. and N.-H. Topp. 2003. Er renten historisk lav? – eller har vi for mange samtidshistorikere?, *Tidsskrift for Landøkonomi*, p. 326-35.
- Nielsen, S. and O. Risager. 2001. Stock Return and Bond Yields in Denmark 1922-1999, *Scandinavian Economic History Review*, Vol. 49(1), p. 63-82.
- Oldam, J. W. 1963. Danmarks høje renteniveau: En særlig tradition?, *Erhvervs-historisk årbog*, p. 119-59.
- Parum, C. 1999a. Historisk afkast af aktier og obligationer i Danmark, *Finans/Invest*, No. 3, p. 4-13.
- Parum, C. 1999b. Estimation of realkreditobligationsafkast i Danmark i perioden 1925-1998, *Finans/Invest*, No. 7, p. 12-15.
- Pedersen, E. H. 2001. Development in and Measurement of the Real Interest Rate, *Danmarks Nationalbank Monetary Review*, 3rd Quarter, p. 71-90.
- Pedersen, J. 1930. Forholdet mellem renten af laan med kort og lang løbetid i perioden 1855-1930, *Økonomi og Politik*, Vol. 4, p. 268-82.
- Statistics Denmark. 1969. *Kreditmarkedsstatistik*, Statistiske Undersøgelser No. 24, Copenhagen: Statistics Denmark.
- Sørensen, B. G. 1995. Konverteringsbølgen, *Samfundsøkonomen*, No. 5, p. 32-40.
- Topp, J. 1996. Indicators of the Market's Interest-Rate and Inflation Expectations in Denmark, *Danmarks Nationalbank Monetary Review*, May, p. 46-58.

**Annex A:***Historical interest-rate data for Denmark – some methodological notes and issues related to interpretation*

In general interest rates depend on the characteristics of the underlying financial assets from which the interest rates are derived. The level of interest rates thus depends on the maturity of the underlying asset and its cash-flow profile, the level of credit risks associated with the debtor, the degree of tradability and liquidity of the asset, and the tax treatment of the cash flows from the asset. Furthermore, more specific details in the contract related to the underlying asset (e.g. call provisions or provisions on collateral) might influence the interest rate. Finally, recorded interest rates depend on which side of the market they are quoted (bid, offer or mid prices) and the day-count convention used for the quotation.

Even for a shorter time-span – such as three or four decades – it is not an easy task to find data on representative interest rates that are fully comparable across time. The choice of data for an exercise covering a time span of more than 125 years is, to an even greater extent, determined by data availability, leaving consistency to be an important but secondary concern. This introduces a certain degree of measurement error, and the analysis presented in this article can only be considered as a crude review of the broad trends in short-term and long-term interest rates in Denmark since 1875.

The first year in the period covered, 1875, was the year when the krone was introduced as the Danish currency unit and Denmark changed her monetary standard from silver to gold. Furthermore, the last part of the 19th century was the period in which national financial markets in Denmark were being developed.<sup>14</sup> Before this period segmentation of the regional financial markets prevented differences in interest rates from being arbitrated away.

Another aspect in relation to the interpretation of historical interest rate data concerns the extent to which the interest rates are market based. Regarding interest rate conditions in Denmark since 1875 two things should be noted:

- An Act of 1 May 1933 (Act No. 143 of 1 May 1933 on Formation of an Emergency Fund and on Reduction of the Interest Rate on Deposits with Banks and Savings Banks) introduced maximum interest rates on deposits with banks and savings banks. These provisions were removed again in July 1935. During the period 1935-1973 savings banks and the commercial banks had an internal agreement regarding maximum interest rates on deposits.<sup>15</sup>
- Until 1942<sup>16</sup> there were some legal provisions regarding maximum interest rates on

---

14. Cf. e.g. Hansen and Johansen (1994).

15. Cf. Hoffmeyer (1960) and Mikkelsen (1993).

16. Act No. 285 of 3 July 1942 on the Interest Rate on Loans Secured by Real Property.

*Table A.1. Danish interest rates – an overview.*


---

Data series	Concept
Official discount rate	The discount rate of the Nationalbank (the central bank of Denmark).
Private banks' average deposit rate	Weighted average of deposit interest rates in savings banks and commercial banks.
Market rate of discount/ money market rate	1875-1940 and 1950-1972: Commercial banks' rate of discount for commercial bills of exchange. 1941-1949: Danmarks Nationalbank's lending rate. Since 1973: 3-month uncollateralised inter-bank interest rate.
Government bond yield	1875-1985: Yield to maturity on long central government bonds. Since 1986: Yield to maturity on 10-year central government bonds.
Yield on mortgage- credit bonds	1875-1959: Average yield to maturity on long callable mortgage-credit bonds. 1960-1972: Yield to maturity on 30-year callable mortgage-credit bonds. Since 1973: Yield to maturity on 20-year callable mortgage-credit bonds.

---

loans secured by real property. However, loans raised through mortgage-credit institutes were exempted from these provisions.<sup>17</sup>

In addition, one also has to take into consideration the presence of restrictions on capital account transactions, cf. the main text of the article. Finally, it should be mentioned that the Danish monetary authorities in some historical periods, e.g. in the 1960s, made use of interventions in the bond market in order to influence the bond yields.

Three different short-term interest-rate series (the official discount rate, the private banks' average deposit rate, and the market rate of discount/money market rate) and two different long-term interest-rate series (the government bond yield and the yield on mortgage-credit bonds) have been constructed for the analysis in this article. Table A.1 gives an overview of the main characteristics of the series.

All the interest rate data presented are annual averages. In some cases, interpolations have been necessary in order to splice old and new data series into comparable time series. The sources and compilation methods are outlined in detail in Abildgren (2005) who also lists and documents time series on short-term and long-term interest rates in Germany, United Kingdom, United States, Norway and Sweden. The time series data on consumer price inflation in Denmark used in figure 3 and in table 1 are listed and documented for the period 1876-2002 in Abildgren (2004b). The data on consumer price inflation in 1875 and 2003 is based on the sources listed in Abildgren, *op. cit.*

---

17. Cf. Hansen and Svendsen (1968) and Hoffmeyer and Olsen (1968).

**Annex B:***Danish interest rates 1875-2003, annual averages, per cent per annum.*

Year	Official discount rate	Private banks' average deposit rate	Market rate of discount/Money market rate	Government bond yield	Yield on mortgage-credit bonds
1875	5.22	4.00	5.38	4.33	4.60
1876	5.26	4.02	5.60	4.38	4.90
1877	5.38	4.09	5.62	4.50	5.00
1878	4.41	3.98	5.13	4.50	5.00
1879	3.62	3.71	4.00	4.30	4.60
1880	3.39	3.67	3.48	4.13	4.40
1881	3.12	3.59	3.57	4.08	4.60
1882	4.08	3.72	4.35	4.08	4.30
1883	4.25	3.83	4.67	4.05	4.40
1884	4.07	3.82	4.53	4.00	4.40
1885	3.85	3.80	4.34	4.00	4.40
1886	3.39	3.67	4.05	3.85	4.10
1887	3.00	3.41	3.50	3.60	3.80
1888	3.00	3.29	3.34	3.55	3.70
1889	3.16	3.36	3.57	3.53	3.70
1890	3.73	3.53	4.11	3.60	3.80
1891	4.00	3.65	4.39	3.75	3.90
1892	3.68	3.55	4.28	3.73	4.00
1893	3.68	3.47	4.08	3.70	3.90
1894	3.53	3.50	4.00	3.58	3.80
1895	3.50	3.31	3.92	3.43	3.60
1896	3.52	3.33	3.86	3.50	3.60
1897	4.41	3.49	4.62	3.50	3.70
1898	4.28	3.52	4.58	3.50	3.70
1899	5.40	3.77	5.67	3.65	4.20
1900	5.87	4.14	6.22	3.78	4.40
1901	5.33	4.21	5.75	3.60	4.50
1902	4.05	3.90	4.46	3.55	4.30
1903	4.31	3.80	4.54	3.50	4.30
1904	4.50	3.79	4.79	3.60	4.40
1905	4.29	3.81	4.55	3.55	4.20
1906	5.22	4.08	5.45	3.58	4.30
1907	6.18	4.30	6.42	3.63	4.40
1908	6.14	4.36	6.78	3.70	4.50
1909	4.94	4.11	5.34	3.68	4.30
1910	5.00	4.07	5.35	3.70	4.40
1911	4.62	3.97	4.99	3.75	4.30
1912	5.06	4.03	5.38	3.93	4.40
1913	5.75	4.18	5.98	4.20	4.60
1914	5.52	4.28	5.97	4.25	4.70
1915	5.27	4.25	5.98	5.20	5.00
1916	5.00	4.14	5.28	4.90	4.90

*Continued ...*

*continued ...*

Year	Official discount rate	Private banks' average deposit rate	Market rate of discount/Money market rate	Government bond yield	Yield on mortgage-credit bonds
1917	5.00	4.20	4.77	4.88	4.90
1918	5.00	4.32	5.47	4.90	5.00
1919	5.61	4.38	6.11	5.20	5.20
1920	6.71	4.70	7.31	6.30	5.80
1921	6.35	4.68	6.20	5.53	5.50
1922	5.16	4.14	5.64	4.90	5.10
1923	5.67	4.19	6.11	5.00	5.30
1924	6.96	4.47	6.39	5.30	5.70
1925	6.49	4.46	6.39	5.28	5.80
1926	5.24	4.38	5.33	5.25	5.70
1927	5.00	4.45	4.94	5.10	5.60
1928	5.00	4.42	4.92	4.93	5.30
1929	5.13	4.46	5.15	5.10	5.30
1930	4.19	4.22	4.73	4.65	4.80
1931	4.22	4.16	4.60	4.75	5.20
1932	4.50	4.26	5.25	5.00	5.40
1933	3.17	3.49	4.53	4.13	4.60
1934	2.50	3.22	3.87	3.95	4.30
1935	2.86	3.43	3.68	4.28	4.80
1936	3.56	3.62	4.31	4.38	4.80
1937	4.00	3.62	4.58	4.55	5.00
1938	4.00	3.62	4.55	4.28	4.80
1939	4.08	3.65	4.64	4.63	5.00
1940	4.79	3.76	4.81	4.93	5.10
1941	4.00	3.44	4.00	4.20	4.40
1942	4.00	3.13	4.00	4.05	4.20
1943	4.00	2.96	4.00	4.38	4.40
1944	4.00	2.61	4.00	3.98	4.10
1945	4.00	2.27	4.00	3.75	4.00
1946	3.52	2.26	3.52	3.55	3.90
1947	3.50	2.32	3.50	3.65	3.80
1948	3.50	2.39	3.50	4.10	4.20
1949	3.50	2.46	3.50	4.43	4.50
1950	4.07	2.68	4.08	4.53	4.60
1951	5.00	2.94	5.00	5.13	5.40
1952	5.00	3.14	5.00	5.28	5.60
1953	4.86	3.17	4.88	5.10	5.50
1954	5.02	3.29	5.00	5.28	5.80
1955	5.50	3.63	5.79	5.55	6.30
1956	5.50	3.66	6.00	5.68	6.40
1957	5.50	3.69	6.00	5.75	6.50
1958	4.96	3.61	5.65	5.23	5.60
1959	4.64	3.36	5.15	5.40	5.80

*Continued ...*

*continued ...*

Year	Official discount rate	Private banks' average deposit rate	Market rate of discount/Money market rate	Government bond yield	Yield on mortgage-credit bonds
1960	5.47	3.89	5.96	6.10	6.56
1961	6.11	4.58	6.63	6.68	7.74
1962	6.50	4.74	7.00	7.24	7.91
1963	6.25	4.38	6.81	7.11	7.95
1964	6.06	4.83	6.54	7.23	8.70
1965	6.50	5.04	7.00	8.49	9.94
1966	6.50	5.06	7.00	8.98	10.17
1967	6.53	5.45	7.04	9.21	10.32
1968	6.66	4.98	7.17	9.03	9.57
1969	8.03	6.31	8.50	9.69	10.29
1970	9.00	6.83	9.50	11.07	12.02
1971	7.70	5.91	8.19	10.50	11.75
1972	7.28	5.57	7.77	10.44	11.88
1973	7.50	6.39	8.10	11.83	13.21
1974	9.94	8.79	13.34	14.13	16.54
1975	8.12	7.32	6.47	12.39	13.49
1976	8.82	8.32	10.28	14.19	15.60
1977	9.17	9.42	14.48	15.71	16.38
1978	8.57	8.62	15.42	15.48	17.33
1979	9.12	8.47	12.63	16.57	17.61
1980	12.28	11.45	16.93	20.38	19.78
1981	11.00	10.85	14.84	19.55	20.11
1982	10.91	10.97	16.92	22.11	21.24
1983	8.06	9.02	12.81	14.55	14.97
1984	7.00	8.62	11.77	14.12	14.78
1985	7.00	8.34	10.33	11.33	12.03
1986	7.00	7.08	9.23	10.20	10.77
1987	7.00	7.62	10.11	11.29	12.55
1988	7.00	7.02	8.53	9.87	11.26
1989	7.00	7.00	9.59	9.70	10.16
1990	8.03	7.90	10.89	10.63	10.98
1991	9.50	7.20	9.70	9.27	10.09
1992	9.50	7.50	11.04	8.99	10.14
1993	8.69	6.50	10.41	7.28	8.16
1994	5.21	3.50	6.13	7.85	8.39
1995	5.36	3.90	6.07	8.27	9.09
1996	3.46	2.80	3.87	7.19	7.84
1997	3.31	2.70	3.66	6.26	7.20
1998	3.80	3.10	4.15	5.03	6.27
1999	2.95	2.40	3.31	4.94	6.60
2000	4.02	3.20	4.91	5.66	7.33
2001	4.29	3.30	4.62	5.09	7.05
2002	3.22	2.55	3.48	5.05	6.69
2003	2.26	1.83	2.38	4.31	5.11

# The Effect of the '92-reform of the Voluntary Early Retirement Pension Program on Retirement Age – A Natural Experiment

Mona Larsen

The Danish National Institute of Social Research, E-mail: ml@sfi.dk

*SUMMARY: This paper examines the effect on the retirement age of a change in the popular Danish early retirement program, known as the Voluntary Early Retirement Pension (VERP) or »efterløn«. In order to reduce early retirement through VERP, the government changed the program in 1992 by increasing the incentives to delay retirement until age 63. To study the effect of this change on retirement outcomes, a natural experiment based on longitudinal register data covering two percent of the population for the period 1980-1998 is conducted. The results suggest that the retirement response to the 1992 changes was relatively small.*

---

## 1. Introduction

Early retirement schemes are regarded as crucial pull-factors that make senior workers retire early, see e.g. Börsch-Supan (2000); Gruber and Wise (1999) and Blöndal and Scarpetta (1998). What's more, these schemes are often considered to be generous, that is, tantamount to large disincentives to continue working after the early retirement age, see e.g. Bratberg et al. (2004). Conversely, previous international studies indicate that the retirement response to *changes* in social security benefits is either limited or non-existent. According to Krueger and Pischke (1992), there is no evidence that reducing Social Security benefits slows the trend to earlier retirement. Although, Mitchell and Fields (1984) show that government practices that alter the rewards for retirement will influence older workers' retirement behaviour in predictable ways, their results

---

This work is part of the research of the Graduate School for Integration, Production and Welfare. Financial support from the Danish Social Science Research Council is gratefully acknowledged. This work benefited from comments by Jeffrey Smith, Mette Ejrnæs, Paul Bingley, Peder Pedersen, Nabanita Datta Gupta and Paul Schou on earlier drafts of this paper. Thanks also to the participants at the IZA Summer School 2003, the CEBR Conference 2003 in Copenhagen, and the First Meeting with the Strategic Advisory Board of the Danish National Institute of Social Research 2003 in Korsør and to two anonymous referees for helpful comments and suggestions. All remaining errors are my own.



indicate a relatively small response. In general, Mitchell and Fields find that rather large changes in policy variables such as taxes or benefits would be required to elicit substantial changes in retirement ages. Similarly, Börsch-Supan (2000) finds that introducing an actuarially fair benefit formula induces a shift of the cumulative distribution of retirement age to the right. The largest increase is found for very early retirement. However, calculations on the basis of Börsch-Supan (2000, Figure 8) show that the marginal effects of introducing such a benefit formula are limited. The average retirement age only rises by 0.45 years.<sup>1</sup>

The response to changes in social security benefits has also been examined in the Danish context. Danø et al. (2000) have modelled the main elements of the Voluntary Early Retirement Pension (VERP, *efterløn*) reform in 1999 in a quantitative study to simulate the effects of the policy changes. They show that the reform increases the average retirement age. The effects differ considerably according to different levels of education and income and the effect is larger for women than for men. Christensen and Datta Gupta (2000) use micro data on couples to simulate the effect of two different types of pension reforms. They find that delaying VERP eligibility from 60 to 62 leads to a greater increase in the average retirement age (0.6 years for husbands and 0.75 years for wives) than a reduction in benefits (0.22 years for husbands and 0.65 years for wives). Finally, Bingley et al. (2003) show that the provisions of the social security system play an important role in determining retirement behaviour in Denmark. Among other things, they simulate the effect of raising the eligibility age of all early retirement programs as well raising the normal retirement age by three years. They find quite a large delaying impact of this three-year reform on the retirement of both men and women.

The focus in this paper is on the effect of changes in the VERP program, the most popular early retirement scheme in Denmark,<sup>2</sup> on the average retirement age. When the program was introduced in 1979, the intention was to offer some form of early retirement for workers who were worn-out as a result of a physically or psychologically demanding working-life. However, today the program is widely used even among relatively healthy older workers. To reduce early retirement through VERP, this program was changed in 1992 and again in 1999. Although the effect of the 1999 changes is of current interest in Denmark, this paper focuses on the effect of the 1992 changes because register data from more years are available. In 1992, the incentives to delay retirement until at least the age of 63 were increased. Before 1 March 1992, VERP benefits were equivalent to unemployment benefits for the first two and a half years; after

---

1. I owe this point to Nelissen (2002).

2. In 2001, almost 160,000 individuals received VERP benefits, see Department of Unemployment Insurance (2001). That is, more than 40 percent of the 60-66-year-olds.

two and a half years, VERP benefits were reduced to a maximum of 80 percent of the unemployment benefit. After March 1, 1992, the reduction in VERP benefits after two and a half years was removed for individuals who delayed retirement until at least age 63. They instead received a benefit equivalent to the unemployment benefit over the whole VERP period (for further details, see Section 2).

While previous Danish studies of the response to changes in social security are based on simulations, this paper reports results of a natural experiment. Such an experiment may be useful in this regard because it allows inference of causal responses in retirement behaviour to policy changes. By using the variation in explanatory variables generated by policy changes, this approach allows the variation that is plausibly exogenous to be obtained. Further, this variation is readily examined without the need for strict structural modelling assumptions. A difference-in-differences estimator that is adjusted for differential trends is applied. Researchers' use of the difference-in-differences framework in examining the response to changes in social security benefits is limited<sup>3</sup> because social security programs most often are national programs in which system parameters apply uniformly to all individuals and reforms are universal. It is, however, possible to create a comparison group and apply the approach in this context because not all individuals are eligible for the VERP benefits.<sup>4</sup> The comparison group is people eligible for the Public Employee Pension (PEP, *tjenestemandspension*) (for further information about this group, see Section 3).

The analysis is based on longitudinal register data. The dataset covers a representative two percent sample of individuals and their spouses from 1980-1998. The results control for the announcement of the policy change. Examining two different VERP-sample groups, changing the time period and conducting separate analyses for each of the ages in focus constitute further controls.

The remainder of the paper is organized as follows: The VERP program and the 1992 changes are described in Section 2 and the PEP program in Section 3. Data are described in Section 4 and empirical models are presented in Section 5. Section 6 presents the results and Section 7 summarizes and discusses them.

## **2. Voluntary Early Retirement Pension (VERP, *efterløn*) and the 1992 changes**

Not all individuals are eligible for the VERP benefits. To be entitled to these benefits, a worker must meet a number of conditions. The conditions have been changed several times since the introduction of the program in 1979. To be entitled to VERP on 1 March 1992, at which time the changes in the VERP program came into force, the worker must:

---

3. An example is Baker and Benjamin (1999), who apply the difference-in-differences approach to examine the introduction of early retirement provisions in Canada's two public pension plans.

4. The rules for entitlement to VERP benefits are described in Section 2 below.

- (1) be between 60 and 66 years of age.
- (2) have been a member of an unemployment insurance fund for 20 out of the last 25 years (however, for people born before March 1952, the condition was 10 of the last 15 years, provided that the membership in the unemployment insurance fund was continuous from 31 March 1992 until transition to VERP).<sup>5</sup>
- (3) have been working full-time for at least 26 weeks in the last three years,<sup>6</sup> be entitled to unemployment insurance benefit, or have received transitional benefits (the Transitional Benefit Program (TBP, *overgangsydelse*) was introduced 1 March 1992 and was in force from 1992 to 1996).<sup>7</sup> After 1 January 1995, people who have received part-time VERP were also entitled to VERP benefits.
- (4) live in Denmark.

In principle, it is possible for people aged 60-66 years to leave the VERP program once and re-enter later, provided that the third condition mentioned above is fulfilled once again. In practice, however, very few make use of this possibility, and therefore, such transitions are ignored in the analyses.

In 1992, the normal retirement age was 67 (65 from 2004). At this age, VERP benefit recipients are transferred to old age pensions. The VERP program is generous in the sense that the amount of benefits does not affect the old age pension, which is independent of previous earnings. The replacement rate, when VERP benefits are received, is on average 70 percent of labour market income for low-income workers and 40 percent for high-income workers, cf. Bingley et al. (2003).

In the period studied, the initial VERP benefits did not depend on the age at entry. Instead, it was reduced as a function of time spent in the program. Before 1 March 1992, VERP benefits for the first two and a half years were equivalent to unemployment benefits, which amount to 90 percent of previous earnings, subject to a ceiling.<sup>8</sup>

After two and a half years, VERP benefits were reduced to a maximum of 80 percent of unemployment benefits<sup>9</sup> (and, after 1 January 1994, raised to a maximum of 82 percent). After 1 March 1992, the reduction in VERP benefits after two and a half years was removed for workers who delayed retirement until at least age 63. These workers received a benefit equivalent to the unemployment benefit during the whole VERP period.<sup>10</sup> In other words, annual VERP benefits for people retiring at age 63-64

---

5. Before 1 March 1992, it was 10 out of 15 years for everyone.

6. After 1 January 1997, the condition was changed to at least 52 weeks in the last three years.

7. For more information about this program, see Section 4.

8. This ceiling is reached quite fast. In fact, the gross compensation percentage for an average production worker (as defined by the OECD) was 53% in 1997, see Hansen (1999).

9. If benefits were below the maximum of 80 percent of unemployment benefits the first two and a half years, benefits were not reduced.

10. The rules applied until the reform of the VERP in 1999.

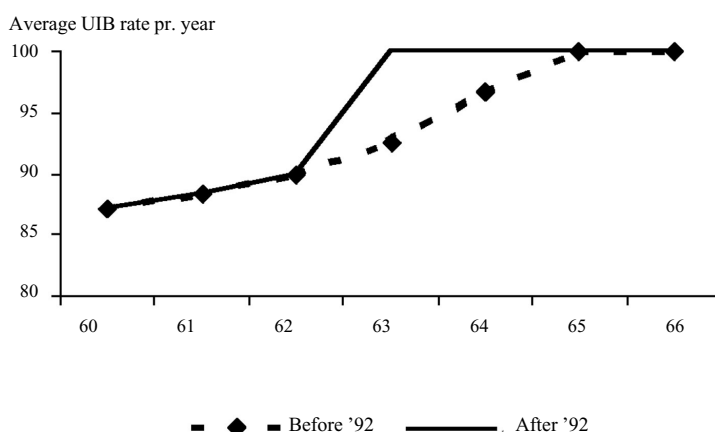


Figure 1. The average annual rate of VERP benefits before and after the changes of the VERP program in 1992. Percent of unemployment benefit. By retirement age.

Note: Reduced VERP benefits correspond to maximum 80 percent of unemployment benefit.

were increased on average, while the situation was unchanged for people retiring at age 60-62 and age 65-66 (see Figure 1 below).

The 1992 changes meant that the financial incentives to delay early retirement from 60-62 to 63 were increased. The effect of the incentives to retire at age 64 is uncertain. Certainly, VERP benefits for age 64 retirement were increased compared to the situation before the change, and the benefit amount remained the same for age 65-66 retirees, thereby increasing the incentives for age 64 retirement. However, the benefit amount was also the same for retirement at age 63, which decreases the incentives for age 64 retirement. Typically, pension benefits are actuarially adjusted so that they are typically higher per year when retirement is delayed; see e.g. Mitchell and Fields (1984). However, people who delayed retirement from 63 to 64 were not compensated for receiving benefits for a shorter period. Similarly, people who delayed retirement until age 65 or 66 did not get any compensation either, and therefore, the incentives to delay retirement until these ages were decreased. All in all, the aggregate effect of the changes in the VERP policy on the average retirement age is uncertain. Provided that the changes in the financial incentives had the intended effect, the effect on the average retirement age depends on the distribution of the retirement age before 1992.

Actual benefit levels were the same at age 63 and 64 before and after the reform for age 63 retirement. At age 65, benefits were around 11,000-13,000 Dkr. (1993-prices) higher per year depending on previous earnings<sup>11</sup> after the reform, while this diffe-

11. In 1993, the ceiling for unemployment benefits was equal to 133,025 DKK. This amount was received if previous earnings were equal to 147,806 DKK or above.

rence was twice as big at age 66. For age 64 retirement, post benefits were the same at age 64 and 65, while at age 66, the gap was around 11,000-13,000 DKK per year. However, irrespective of retirement age, if previous earnings were equal to or below 118,244 DKK (corresponding to 80 percent of the ceiling for VERP benefits), VERP benefits were the same before and after the reform.

### 3. Public Employee Pension (PEP, tjenestemandspension)

PEP is available for public servants who are state-employed or employed in the public school system, the national church, or the Danish Parliament. A public servant is entitled to PEP from the age of 60, and retirement is mandatory at the age of 70 unless lower age limits for the job in question are specified. Furthermore, to receive PEP, at least 10 years of service is required.<sup>12</sup> PEP is the only Danish public sector program that ties a worker's benefits to their past record, see e.g. Bingley et al. (2003), that is, the programs are actuarially adjusted, as there is a reduction in the benefit amount for retirement before age 67.<sup>13</sup> The reduction depends on the retirement age and the number of »pension years« accumulated at the point of retirement.<sup>14</sup> »Pension years« refer to the number of years that a public servant has been employed full-time in public service from the age of twenty-five. The number of pension years cannot exceed 37. Instead of receiving old age pensions at age 67, PEP recipients continue to receive PEP for the rest of their lives.

People entitled to PEP are used as the comparison group to people entitled to VERP. First, these individuals are not affected by the change in the VERP policy in 1992. Second, it seems reasonable to assume that the composition of the group of individuals entitled to PEP is exogenous. Although people could have been jumping between VERP and PEP, it is unlikely that the composition of the two groups changed due to the policy changes in 1992 because rules dictate that eligibility to VERP and PEP respectively is a function of a long labour market history going back in time. Another objection could be that the comparison group is not pure in the sense that, in addition to people eligible for PEP, other public employees are also included in this group. However, the composition of the comparison group is fairly stable over time (see Table A.1 in the Appendix). Consequently, this group in its entirety seems to be a suitable comparison group.

### 4. Data

The data used are obtained from a Danish longitudinal register dataset created for

---

12. However, particular rules apply to public servants with at least three years of service who leave their job before the early retirement age.

13. After 1 July 1999, this was changed to age 65 for people born after 1 July 1939. If the mandatory retirement age is lower than 67 (65), there is a reduction in benefit amount for retirement before this age.

14. After 1 January 1994, the maximum benefit was 57 percent of the wage when retiring.

administrative purposes. The dataset contains a representative two percent sample of individuals and their spouses for 1980-1998. In this paper, the focus is on 60-66-year-olds who were in the labour market when aged 59. Therefore, information about people born in the period from 1921 to 1938 is applied, resulting in a sample in which individuals in the oldest cohort were 59 years old in 1980, while individuals in the youngest cohort were 60 years old in 1998. This selection implies that a representative sample of 60-66-year-olds who were in the labour market when aged 59 is available for the period 1987-1998. This sample includes 10,922 individuals.

To analyse the effect of the change in the VERP policy in 1992, a natural experiment is possible. The people in the VERP sample group are entitled to VERP, while people eligible for PEP constitute the control group. Entitlement to VERP is determined on the basis of information about receipt of Voluntary Early Retirement Pension benefits and retrospective information about membership of an unemployment insurance fund. More precisely, individuals are considered as entitled to VERP if they received Voluntary Early Retirement Pension benefits during the sample period or had been a member of an unemployment insurance fund for at least 10 years until age 60 or, if left-truncated, every year from 1980 until age 60. In counting the number of membership years, the condition of continuous membership on and after 1992 is taken into account. Two VERP sample groups are constructed. A broader group (VERP1) consists of individuals who fulfil the conditions just mentioned, while a smaller group (VERP2) consists of a subsample of individuals from VERP1 who have been publicly employed for at least 10 years until age 60 or, if left-truncated, every year from 1980 until age 60. VERP1 makes up 78 percent of the selected sample, while VERP2 accounts for 26 percent. In the residual group, eligibility for PEP is determined on the basis of information about receipt of a public employment pension and retrospective information about public employment.<sup>15</sup> If individuals have received a public employment pension during the sample period<sup>16</sup> or had been publicly employed according to the same criteria as individuals in the VERP2 group, they are included in the PEP group (PEP). PEP accounts for 12 percent of the selected sample. For number of observations in each of the three groups for each of the years in the sample period, see Table A.1 in the Appendix.

The idea behind constructing two VERP sample groups is to be able to control the robustness of the results to selections made to increase comparability between the VERP sample and the control group. However, two control groups are not constructed

---

15. In this way, data on public servants is captured. However, some public employees not entitled to VERP are also included, because the available data do not enable a more precise selection of individuals to this group.

16. Unfortunately, information about receipt of public employment pensions is only available for the years 1997 and 1998.

because while the number of individuals in the selected sample who are entitled to VERP is relatively large, the number of individuals included in PEP is too small to impose further restrictions on this group. Another potential control group might have consisted of people entitled to neither VERP nor PEP. However, this group is too heterogeneous to be suitable for this purpose. The point of restricting VERP2 to public employees with long tenures is to make the sample group more comparable with respect to education, income, previous unemployment and sector – all factors that affect retirement behaviour. The larger comparability with regards previous unemployment and sector is obtained through the selection criteria. These criteria have also resulted in larger comparability as intended, although a relatively large difference still remains with respect to education and income in particular (see Table A.1 in the Appendix). Conversely, however, VERP2 differs more from PEP with respect to gender than VERP1. Thus, VERP2 is dominated by women, while a minor predominance of men is found in PEP and VERP1. Gender has also a pronounced effect on retirement behaviour and larger comparability would be obtained if separate analyses for men and women were carried out and if the sample group was restricted further with respect to education and/or income. However, small sample size makes separate analyses impossible. Instead, differences with respect to gender, education and income are at least partly controlled for since they are included as explanatory variables in the analysis.

The sample is restricted to individuals who were wage earners or temporarily out of employment due to unemployment or leave at age 59 (individuals, who were outside the labour force at age 59 are excluded because they were not entitled to either VERP or PEP at age 60). To make the individuals in the two VERP groups as comparable as possible to the group entitled to PEP, which consists exclusively of wage earners, self-employed and assisting spouses are also excluded. Labour market status at age 59 and the retirement year are determined from annual information about the primary labour market status at the end of November.

### **5. Empirical models**

The focus is on the average effect of the change in the program on people with access to VERP rather than only on people, who chose to participate in the program. To examine the effect of the change, two cross-section samples of 60-66-year-olds entitled to VERP before and after 1992 are compared to the control group on the basis of two cross-section samples of 60-66-year-olds eligible for PEP.

A difference-in-differences framework is applied. The legitimacy of the difference-in-differences approach is grounded on three assumptions: The change in the VERP policy provided exogenous variation in the outcome of retirees; there was no program

specific shock to behaviour co-incident with the 1992 change in the VERP, and secular trends in labour market behaviour are common to individuals in both groups.

As for the first assumption, the Minister of Labour brought in a bill about changing the VERP program in October 1991 and the law was passed in December 1991. According to the Minister, the idea of changing the program arose from an article in a newspaper half a year earlier, cf. »Ministry of Labour« (1991). The timing of these events might imply that the results are biased by the actions of individuals who in 1991 anticipated the 1992 changes to the VERP. However, two opposite tendencies are present and actually, these might have cancelled each other out.<sup>17</sup>

The second assumption that there was no program specific shock to behaviour co-incident with the 1992 changes in the VERP that affected the two groups differently might be problematic. In fact, at the same time as the VERP program was changed in 1992, the Transitional Benefit Program (TBP, *overgangsydelse*) was introduced. This program, a sort of extension of the VERP program, was available from 1992 to 1996 for people 55-59 years old (and from 1994 for the 50-54-year-olds), who were members of an unemployment insurance fund and had been unemployed for at least 12 out of 15 months. That is, people entitled to TBP constitute a subset of people who are entitled to VERP at age 60. Therefore, the existence of the TBP program might cause a potential sample selection problem since people who choose this program are not selected for the VERP groups in the analysis because they are outside the labour force at age 59. However, the entry of people into the TBP program was relatively limited until 1995. In fact, only up to 2 percent of the 55-59-year-olds received TBP benefits from 1992-1994, while the percentage rose to 5 and 9 percent in 1995 and 1996, respectively. Therefore, to minimize the potential sample selection problem and due to small sample size for the control group, information after 1994 is not utilized in the analyses.

The third assumption is that secular trends in labour market behaviour are common to individuals in both the VERP sample and the control group. Comparing the development of the average retirement age for the control group and the two VERP sample groups, respectively, can test this assumption (see Figure 2 below). The largest difference is found when PEP is compared to VERP1. Part of this difference might be that, contrary to people entitled to VERP, public servants in general suffer few layoffs

---

17. Because the 1992 reform meant that nobody entitled to VERP in 1991 was financially worse off if they delayed retirement until after the 1992 changes, while in particular the 60-62-year-olds were better off, the potential effect of the announcement on the retirement age in 1991 would be positive. However, in 1991 it was also suggested that the age at which VERP benefits were first available should be raised from 60 to 62 years. For a while, the Minister of Labour received many letters from angry citizens who believed that this change would come into force in January 1992, cf. »Ministry of Labour« (1991). These expectations in 1991 might have induced some people to retire earlier than planned, thereby having a negative effect on the retirement age in 1991.



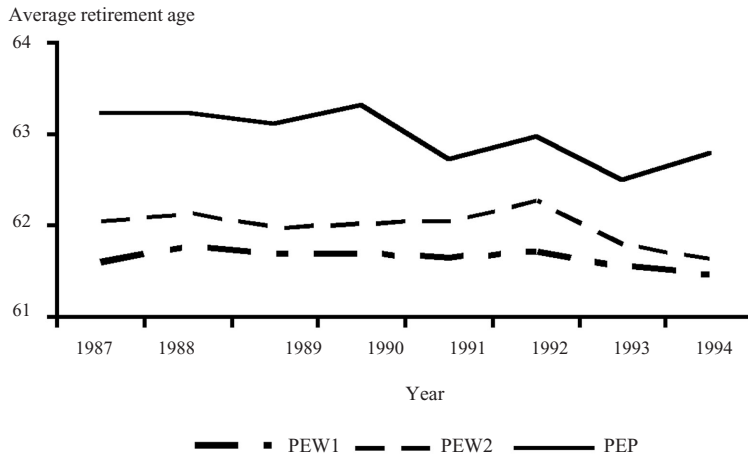


Figure 2. Average retirement age for 60-66-year-olds, VERP1, VERP2 and PEP, 1987-1994.

and have a high degree of job protection. The curves for VERP2 and PEP are more similar although not completely parallel either. Part of this difference might be explained by different gender composition of the two groups. Therefore, the overall picture is that individuals eligible for PEP do not seem to experience all of the other influences affecting people entitled to VERP.

One way to eliminate the different trend is to control for this difference by including a third dummy capturing the trend adjustment in the analysis. In this way the difference-in-differences estimator is adjusted for the differential trends. As stated in Blundell and Costa Dias (2000), formally, the problem can be written as:

$$E(\tilde{\alpha}_{DD}) = \alpha + (k_T - k_C)(\theta_{t_1} - \theta_{t_0}) \quad (1)$$

where  $\tilde{\alpha}$  is the impact of the change, which in this context measures the excess outcome growth for the VERP sample compared to the control group,  $\theta_t$  is the macroeconomic effect and  $k_T$  and  $k_C$  indicate the differential macro effect across the VERP sample ( $T$ ) and the control ( $C$ ) groups. The true effect of the change is only recovered when  $k_T = k_C$ . According to Blundell and Costa Dias (2000) a possible solution to this problem is to take another time interval,  $t^*$  to  $t^{**}$ , over which a similar macro trend has occurred. More precisely, a period for which the macro trend matches the term  $(k_T - k_C)(\theta_{t_1} - \theta_{t_0})$  in (1) is chosen. The differential adjusted difference-in-differences estimator (TADD), which will consistently estimates  $\alpha$ , takes the following form:

$$\tilde{\alpha}_{TADD} = [\bar{Y}_{t_1}^T - \bar{Y}_{t_0}^T] - (\bar{Y}_{t_1}^C - \bar{Y}_{t_0}^C) - [\bar{Y}_{t_{**}}^T - \bar{Y}_{t_*}^T] - (\bar{Y}_{t_{**}}^C - \bar{Y}_{t_*}^C) \quad (2)$$

where  $\bar{Y}^T$  and  $\bar{Y}^C$  are the mean outcomes for the VERP sample and the control group respectively.

In order to identify a period in which a macro trend similar to the trend around 1992 has occurred, a figure corresponding to Figure 2 but excluding observations for 1992 is constructed (see Figure A.1 in the Appendix). If observations for 1990 and/or 1994 are included, it is not straightforward to identify a period in which the macro trend is similar to the trend around 1992. Therefore, the period 1991-1993 is used. Comparing the development in this period to the development in the preceding years, the period 1988-1989 seems to be the most appropriate choice. In other words,  $t_*$  is set equal to 1988, while  $t_{**}$  is set equal to 1989. By using observations for 1991, it is assumed that the announcement of the policy change in 1992 did not affect the estimated effect of this change, while the use of observations for 1993 implies that only short run effects are examined.<sup>18,19</sup>

The trend adjusted difference-in-differences estimate of the effect of the change in the VERP policy in 1992 is computed in an OLS regression on repeated cross sections consisting of pooled micro data for groups and years. The regressors consist of dummies for years and groups, a dummy to adjust for the different macroeconomic effect and interaction terms for these dummies. A vector of individual characteristics is also added:

$$Y_{ijt} = \beta_1 X'_{ijt} + \beta_2 AFTER_t + \beta_3 VERP_i + \beta_4 ADJUST_j + \beta_5 (AFTER_t * VERP_i) + \beta_6 (AFTER_t * ADJUST_j) + \beta_7 (VERP_i * ADJUST_j) + \beta_8 (AFTER_t * VERP_i * ADJUST_j) + \varepsilon_{ijt} \quad (5)$$

The dependent variable ( $Y_{it}$ ) is the retirement age for individuals retiring the year in question in the analysis of the whole sample of 60-66-year-olds. The included individual characteristics ( $X_{it}$ ) are gender, education, cohabitation status, experience, lagged

18. The choice of the previous period (1988-1989) rests on Figure A.1. However, one might argue that this choice is made unconditional of all other kinds of characteristics such as macro shocks and trends in the composition of the groups with respect to e.g. education. Nevertheless, controlling for the unemployment level by gender and for changes in the vocational/higher education ratio within each group over time, the period 1988-1989 still seems to be an appropriate choice in the case of VERP2. Conversely, VERP1 and PEP do not move in parallel in this period. However, the small sample size for PEP implies that caution must be taken in interpreting these results.

19. A similar approach has been applied in Blundell and Costa Dias (2000) in an evaluation of the New Deal for Young People, an initiative of the UK government to help young unemployed people make their way into or back to work. In this study, the outcome of interest is the outflow from the jobseeker's allowance claimant count.

unemployment and lagged income. The estimator is calculated on the basis of information for 1988, 1989, 1991 and 1993. The dummy for year ( $AFTER_t$ ) is equal to one for observations for 1989 and 1993, the dummy for group ( $VERP_i$ ) is equal to one for individuals in VERP1 and VERP2 respectively and the dummy for adjusting for different trend ( $ADJUST_j$ ) equals one for observations for 1991 and 1993. The estimate of the coefficient on the interaction term equal to the product of the dummies for years, groups and for adjustment for different trends ( $\beta_8$ ) indicates whether the policy change had the intended effect.

As the effect of the changes in the VERP program might be the delay of retirement from age 60-62 to age 63 with an average of zero because retirement at age 65-66 at the same time was hastened, separate pooled probit analyses are also carried out for each of the retirement ages 60-66 years. In these analyses, the dependent variable is a dummy that equals one if retirement takes place at the age in question.

## 6. Results

To examine the effect of the increase in delaying incentives in the VERP program on the retirement age, a difference-in-differences estimator that is adjusted for differential trends is applied. However, before turning to the results of this analysis, the distributions of retirement age before and after 1992 are presented.

### 6.1. Distribution of retirement age

The changes of the VERP program in 1992 caused the financial incentives for retiring at age 60-62 and age 65-66 to decrease, and the incentive for retiring at age 63 to increase, while the effect on retirement at age 64 was uncertain. Therefore, provided that the changes in the financial incentives have the intended effect, the effect on the average retirement age depends on the distribution of retirement age before the changes to the VERP policy.

Before 1992, the majority of individuals entitled to VERP retired before the age of 63 (see Figure 3<sup>20</sup> above). Therefore, if the changes in the financial incentives in 1992 had the intended effect, the average retirement age increased. For VERP1, however, the distribution of the retirement age after 1992 was very similar to the situation before 1992. For VERP2, the share of people retiring at age 62 certainly decreased, but at the same time age 65 retirement also decreased, age 60 retirement increased while the tendency to retire at age 63 remained almost unchanged. These results indicate that the changes in the financial incentives did not work as intended.

---

20. The columns for the average retirement age before 1992 are based on pooled information for the period 1987-1990. Information about 1991 is excluded due to the potential announcement effect. The »after 1992« columns are based on pooled information for the period 1993-1998.

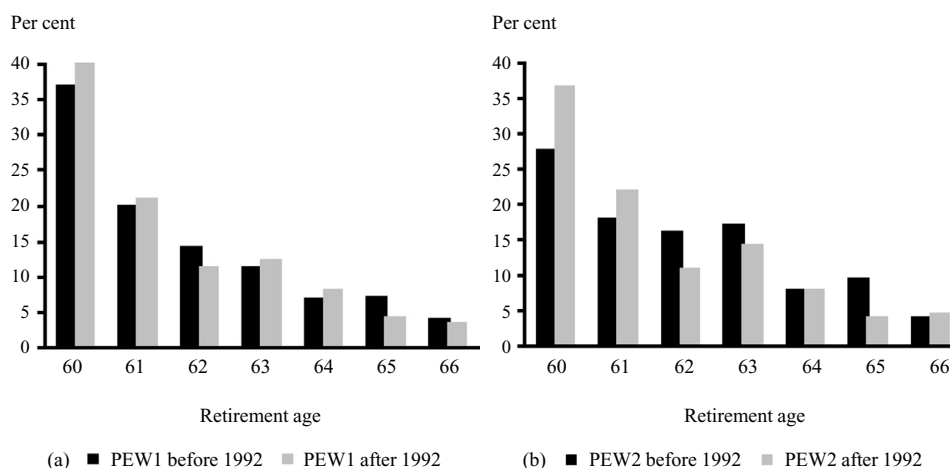


Figure 3. Distribution of the retirement age among 60-66-year-olds before and after 1992, (a) VERP1, (b) VERP2.

### 6.2. Trend adjusted difference-in-differences estimates

That Figures 3a and 3b indicate that the changes in 1992 did not work as intended might result from other influences affecting not only the VERP sample group but also other individuals. Therefore, trend adjusted difference-in-differences analyses are conducted in which the comparison group is used to control for these other influences.<sup>21</sup>

Information about gender, education, cohabitation status, experience, lagged unemployment and lagged income is included in these and the subsequent analyses. The estimated coefficients on these variables are not reported in the tables but are available on request.

The difference-in-differences estimates that are adjusted for differential trends on the sample of 60-66-year-olds suggest that the average retirement age did not increase as intended (see Table 1 below). Certainly, a positive coefficient for the interaction term equal to the product of the dummies for year, group and adjustment for different trend is found for VERP1 as well as VERP2. However, despite the inclusion of the adjustment variable, the effect on the average retirement age is insignificant for both groups. The retirement age is lower for individuals entitled to VERP as expected.

The results suggest that the change of the VERP policy in 1992 did not increase the average retirement age significantly as intended. However, the effect of the change

21. This analysis might suffer from the Ashenfelters dip problem pointed out by one of the anonymous referees. If the behaviour of the two groups differs substantially before the reform, the identified effects might be catch-up effects rather than effects of the reform. Because of the number of control variables included in the analysis (see note to Table 1), group differences in behaviour arising due to other reasons have been adequately controlled for.

*Table 1. Trend adjusted difference-in-differences estimates of the effect on retirement age of the 1992 changes to the VERP policy, VERP1 vs. PEP, VERP2 vs. PEP. Pooled OLS results for 60-66-year-old retirees based on observations for 1988, 1989, 1991 and 1993.*

	<i>VERP1</i>	<i>VERP2</i>
VERP	-1.291 *** (0.279)	-0.968 *** (0.326)
AFTER	-0.151 (0.352)	-0.186 (0.369)
VERP*AFTER	0.107 (0.376)	-0.001 (0.447)
ADJUST	-0.488 (0.356)	-0.396 (0.372)
VERP*ADJUST	0.478 (0.376)	0.396 (0.438)
AFTER*ADJUST	-0.018 (0.511)	0.043 (0.533)
VERP*AFTER*ADJUST	0.022 (0.540)	0.014 (0.628)
Number of observations	1956	716

\*: Significant at the 10 percent level; \*\*: significant at the 5 percent level; \*\*\*: significant at the 1 percent level.

*Note:* Additional controls include gender, education, cohabitation status, experience, lagged income and lagged unemployment.

might have been that retirement was in fact delayed from age 60-62 until the age of 63, while retirement at age 65-66 at the same time was hastened. Such shifts might explain why the effect was zero on average. To control for this, separate analyses for each of the ages 60 to 66 years are conducted. The focus is on individuals who were in the labour market at age  $x-1$ .

First of all, these results suggest that the change in VERP policy increased the incentives for retiring at age 63 and 64 as intended (see Table 2 above). The increase in age 64 retirement suggests that the fact that the VERP benefits increased on average when retirement occurred at age 64 was more important than the fact that people who delayed retirement from age 63 to age 64 were not compensated for receiving benefits for a shorter period. The results indicate that the changes in financial incentives for retiring at age 63-64 have been sufficient to change the retirement behaviour in the intended direction for individuals in the VERP group around these ages.

Table 2. Trend adjusted difference-in-differences estimates of the effect of the changes to the VERP policy in 1992 on retirement at age  $t$  conditioned on being in the labour market at age  $t-1$ , VERP1 vs. PEP, VERP2 vs. PEP. Pooled probit results based on observations for 1988, 1989, 1991 and 1993.

	60	61	62	63	64	65	66							
	Coef.	Marg. eff.	Coef.	Marg. eff.	Coef.	Marg. eff.	Coef.							
	Marg. eff.		Marg. eff.		Marg. eff.		Marg. eff.							
VERP1														
VERP	0.591*** (0.201)	0.183*** (0.052)	0.664*** (0.239)	0.173*** (0.050)	0.778*** (0.254)	0.185*** (0.047)	0.516*** (0.249)	0.140** (0.060)	0.366 (0.263)	0.099 (0.066)	0.399 (0.266)	0.118 (0.074)	-0.222 (0.285)	-0.069 (0.090)
AFTER	-0.000 (0.287)	0.000 (0.102)	0.075 (0.310)	0.023 (0.096)	0.236 (0.302)	0.069 (0.087)	0.287 (0.305)	0.087 (0.093)	0.396 (0.307)	0.115 (0.090)	0.459 (0.298)	0.143 (0.092)	-0.703* (0.370)	-0.210* (0.108)
VERP*AFTER	0.047 (0.302)	0.017 (0.107)	-0.131 (0.330)	-0.040 (0.100)	-0.697* (0.329)	-0.192** (0.085)	-0.600* (0.345)	-0.167* (0.088)	-0.472 (0.355)	-0.128 (0.089)	-0.544 (0.352)	-0.159 (0.095)	0.781* (0.444)	0.254* (0.150)
ADJUST	0.471* (0.268)	0.165* (0.092)	0.327 (0.301)	0.101 (0.093)	0.198 (0.323)	0.058 (0.095)	0.104 (0.304)	0.031 (0.092)	0.273 (0.309)	0.079 (0.090)	-0.014 (0.350)	-0.004 (0.110)	0.097 (0.329)	0.029 (0.099)
VERP*ADJUST	-0.367 (0.282)	-0.129 (0.098)	-0.067 (0.319)	-0.021 (0.098)	-0.183 (0.348)	-0.052 (0.098)	-0.023 (0.335)	-0.007 (0.101)	-0.254 (0.353)	-0.071 (0.095)	-0.019 (0.402)	-0.006 (0.125)	0.276 (0.407)	0.086 (0.130)
AFTER*ADJUST	-0.343 (0.396)	-0.116 (0.127)	0.019 (0.422)	0.006 (0.132)	-0.387 (0.460)	-0.103 (0.111)	-0.453 (0.440)	-0.124 (0.108)	-0.606 (0.454)	-0.150 (0.093)	-0.655 (0.474)	-0.179 (0.109)	0.104 (0.528)	0.032 (0.166)
VERP*AFTER*ADJUST	0.422 (0.415)	0.156 (0.158)	0.169 (0.450)	0.054 (0.149)	0.496 (0.497)	0.160 (0.174)	1.000** (0.492)	0.352** (0.183)	1.136** (0.518)	0.400** (0.192)	0.690 (0.550)	0.244 (0.209)	-0.052 (0.620)	-0.015 (0.182)
Number of obs.	2121	1512	1097	880	708	545	428							

Continued ...

continued ...

	60	61	62	63	64	65	66
	Coef.	Coef.	Coef.	Coef.	Coef.	Coef.	Coef.
	Marg. eff.	Marg. eff.	Marg. eff.	Marg. eff.	Marg. eff.	Marg. eff.	Marg. eff.
VERP2							
VERP	0.365 (0.232)	0.409 (0.269)	0.666** (0.280)	0.764*** (0.279)	0.466 (0.323)	0.132 (0.318)	-0.244 (0.372)
AFTER	-0.008 (0.292)	0.097 (0.317)	0.206 (0.304)	0.299 (0.309)	0.395 (0.312)	0.443 (0.302)	-0.719* (0.377)
VERP*AFTER	0.182 (0.341)	0.145 (0.374)	-0.521 (0.367)	-1.020** (0.410)	-0.149 (0.438)	0.006 (0.447)	0.778 (0.538)
ADJUST	0.437 (0.275)	0.403 (0.300)	0.160 (0.326)	0.099 (0.308)	0.266 (0.314)	-0.035 (0.353)	0.095 (0.332)
VERP*ADJUST	-0.229 (0.321)	0.188 (0.355)	-0.245 (0.396)	-0.183 (0.377)	-0.327 (0.422)	0.257 (0.470)	0.120 (0.517)
AFTER*ADJUST	-0.432 (0.405)	-0.191 (0.430)	-0.331 (0.464)	-0.441 (0.445)	-0.598 (0.464)	-0.697 (0.479)	0.073 (0.536)
VERP*AFTER*	0.322 (0.467)	-0.051 (0.507)	0.587 (0.559)	1.170** (0.576)	1.148* (0.623)	-0.169 (0.671)	0.412 (0.741)
ADJUST							
Number of obs.	839	745	562	491	394	323	267

\*: Significant at the 10 percent level; \*\*: significant at the 5 percent level; \*\*\*: significant at the 1 percent level.

Note: Additional controls include gender, education, cohabitation status, experience, lagged income and lagged unemployment.

Second, by contrast, transition to retirement at age 60-62 and age 65-66 seems not to have been affected significantly. The results mitigate the conclusion that the change in the VERP policy did not have any effect. However, the results suggest that the aggregate effect on the retirement age was rather limited.

### **7. Concluding remarks**

The focus of this paper is on the VERP program, the most popular early retirement program in Denmark, and on the effect of a 1992 change on retirement age. To reduce early retirement through VERP, the government increased the incentives for delaying retirement until age 63.

All in all, results from a natural experiment suggest that the retirement response to the change in VERP policy was relatively small at least in the short run. Distributions of retirement age indicate that the changes did not have the intended effect. However, difference-in-differences estimates that are adjusted for differential trends suggest that the change, as intended, increased the incentives for retiring at age 63-64. But these analyses also suggest that the aggregate effect was small since retirement at age 60-62 and at age 65-66 was not affected significantly. Consequently, the retirement age for people entitled to VERP did not on average increase. However, it cannot be ruled out that in the longer run, a larger effect might have arisen.

The results are similar to those found in previous international studies, thus confirming that only rather large policy changes can elicit substantial changes in retirement ages. The 1992 changes in financial incentives seem to have been too small to significantly delay retirement, at least in the short run. Another reason for the small effect might be that there was no stick, only carrots. Although people entitled to VERP got a »reward« if they delayed retirement until 63, if they retired at age 60-62, the VERP benefits remained exactly the same as before 1992. Moreover, the carrots were rather small. If retirement was delayed until 63, the change first set in 2<sup>1/2</sup> years after the transition to retirement. In addition, VERP benefits did not increase in absolute terms after 2<sup>1/2</sup> years. Instead, the reduction in benefits after 2<sup>1/2</sup> years, which was in force for everyone receiving VERP benefits before 1992, was removed for this group. Therefore, the changes were hardly noticeable in the individual case.

Contrary to the policy aim, the changes to the VERP policy in 1992 might actually have caused public expenditures to increase as a result of the »no stick, only carrots« construction of the policy change, if the effect on retirement behaviour was limited, as the results suggest. If so, the 1992 changes increased the burden on the providers instead of diminishing it.

The results that this paper presents deviate from those of previous Danish studies. Presumably, the most important explanation for this difference is that the 1992 changes



in the VERP program were too small to affect the average retirement age. These changes were much less far-reaching than the reforms considered in the previous studies. In part, the deviation could be due to the fact that, unlike previous Danish studies, this study employs a natural experiment to study retirement behaviour and thereby the resulting estimates can be thought of as causal responses of retirement to policy changes.

The 1999 changes in VERP implied, among other things, that VERP is moving from a pay-as-you-go system toward a funded system. In addition, benefits now depend on the age at entry instead of being reduced as a function of time spent in the program, benefits are means-tested against income from all other pension schemes for people aged 60-61 and a tax premium has been introduced for people who are entitled to early retirement benefits at age 60 but who continue to work at least to the age of 62. Although these changes were more extensive than the changes conducted in 1992, so far the effect of this reform on the average retirement age seems to be modest, see Jørgensen et al. (2005). However, no causal analysis of the effect of this reform has been conducted and, therefore, the effect is uncertain. What's more, it seems reasonable to believe that the effect of this reform might increase in the future. The amount of income from all other pension schemes is larger among younger generations and, therefore, the effect of means-testing for people aged 60-61 will presumably increase.

#### References

- Baker, M. and D. Benjamin. 1999. Early Retirement Provisions and the Labor Force Behavior of Older Men: Evidence from Canada, *Journal of Labor Economics*, vol. 17, no 4, pt. 1, p. 724-56.
- Bingley, P., N. Datta Gupta and P. J. Pedersen. 2003. The Impact of Incentives on Retirement in Denmark, in: Wise, J. and Gruber, D. (eds.), *Social Security and Retirement Around the World: Microestimation*, NBER.
- Blöndal, S. and S. Scarpetta. 1998. The Retirement Decision in OECD Countries, *Aging Working Papers* 1.4, Paris.
- Blundell, R. and M. Costa Dias. 2000. Evaluation Methods for Non-Experimental Data, *Fiscal Studies*, vol. 21, no. 4, p. 427-68.
- Börsch-Supan, A. 2000. Incentive Effects of Social Security on Labour Force Participation: Evidence in Germany and Across Europe, *Journal of Public Economics* 78, p. 25-49.
- Bratberg, E., T. H. Holmås and Ø. Thøgersen. 2004. Assessing the effects of an early retirement program, *Journal of Population Economics* 17, p. 387-408.
- Christensen, B. J. and N. Datta Gupta. 2000. Effekten af pensionsreform på danske ægteparers udtræden af arbejdsmarkedet, *Nationaløkonomisk Tidsskrift* 138, p. 222-42.
- Danø, A. M., M. Ejrnæs, and L. Husted. 2000. Hvordan påvirker efterlønsreformen de ældres tilbagetrækningsalder? *Nationaløkonomisk Tidsskrift* 138, p. 205-21.
- Department of Unemployment Insurance. 2001. *Efterløns- og overgangsydelsesstatistik for 1. halvdel af 2001*, digital document: [www.adir.dk](http://www.adir.dk).
- Gruber, J. and D. Wise, eds. 1999. *Social Security and Retirement around the World*, Chicago. National Bureau of Economic Research, p. 1-35.
- Hansen, H. 1999. *Elements of Social Security, A comparison covering: Denmark, Swe-*

- den, Finland, Austria, Germany, The Netherlands, Great Britain and Canada*, The Danish National Institute of Social Research 99:14, Copenhagen.
- Jørgensen, M., M. Larsen and M. Rosenstock. 2005. *Et længere arbejdsliv. Tilbagetrækningsordninger og arbejdspladsens muligheder*. The Danish National Institute of Social Research, Copenhagen, forthcoming.
- Krueger, A. B. and J.-S. Pischke. 1992. The Effect of Social Security on Labor Supply: A Cohort Analysis of the Notch Generation, *Journal of Labor Economics*, vol. 10, no 4, p. 412-37.
- Mitchell, O. S. and G. S. Fields. 1984. The Economics of Retirement Behavior, *Journal of Labor Economics*, vol. 2, no 1, p. 84-105.
- Nelissen, J. H. M. 2002. Early Retirement: The Impact of Changes in the Benefit Level, *Tinbergen Institute Discussion Paper TI 2002-031/ 3*, Rotterdam.
- Ministry of Labour. 1991. *Sortbog. Lov om ændring af lov om arbejdsformidling og arbejdsløshedsforsikring m.v. Lovforslag nr. L 29*. Copenhagen.

**Appendix***Table A.1. Descriptive statistics for employed and retired 60-66-year-olds who were in the labour market at the age of 59, VERP1, VERP2 and PEP, 1987-1998. Means and standard deviations (in parentheses).*

	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998
<b>VERP1</b>												
Women	0.43 (0.49)	0.43 (0.50)	0.44 (0.50)	0.45 (0.50)	0.45 (0.50)	0.46 (0.50)	0.46 (0.50)	0.46 (0.50)	0.47 (0.50)	0.48 (0.50)	0.48 (0.50)	0.48 (0.50)
Vocational training	0.31 (0.46)	0.32 (0.47)	0.33 (0.47)	0.34 (0.47)	0.36 (0.48)	0.36 (0.48)	0.38 (0.48)	0.38 (0.49)	0.39 (0.49)	0.39 (0.49)	0.39 (0.49)	0.40 (0.49)
Higher education	0.08 (0.28)	0.08 (0.28)	0.08 (0.27)	0.09 (0.28)	0.09 (0.29)	0.10 (0.30)	0.10 (0.30)	0.11 (0.31)	0.12 (0.32)	0.13 (0.34)	0.14 (0.34)	0.15 (0.35)
Single	0.26 (0.44)	0.25 (0.44)	0.26 (0.44)	0.26 (0.44)	0.26 (0.44)	0.25 (0.43)	0.25 (0.43)	0.25 (0.44)	0.26 (0.44)	0.26 (0.44)	0.26 (0.44)	0.25 (0.44)
Experience	15.76 (6.44)	16.30 (6.65)	16.87 (6.92)	17.46 (7.25)	18.07 (7.57)	18.68 (7.86)	19.31 (8.14)	19.91 (8.33)	20.54 (8.55)	21.14 (8.78)	21.85 (8.99)	22.65 (9.17)
Lagged unemployment	0.06 (0.19)	0.06 (0.19)	0.06 (0.20)	0.07 (0.20)	0.08 (0.22)	0.08 (0.23)	0.08 (0.22)	0.08 (0.23)	0.10 (0.27)	0.10 (0.27)	0.09 (0.25)	0.07 (0.22)
Lagged income (log)	7.05 (5.49)	6.89 (5.57)	6.73 (5.66)	6.84 (5.66)	7.01 (5.64)	6.78 (5.69)	6.72 (5.73)	6.59 (5.75)	6.37 (5.77)	6.29 (5.78)	6.42 (5.79)	6.55 (5.85)
Number of observations	3068	3054	3063	3062	3077	3059	3132	3179	3270	3321	3353	3385
<b>VERP2</b>												
Women	0.62 (0.48)	0.63 (0.48)	0.64 (0.48)	0.65 (0.48)	0.64 (0.48)	0.64 (0.48)	0.65 (0.48)	0.66 (0.48)	0.67 (0.47)	0.68 (0.47)	0.68 (0.47)	0.70 (0.46)
Vocational training	0.26 (0.44)	0.27 (0.45)	0.28 (0.45)	0.28 (0.45)	0.29 (0.45)	0.30 (0.46)	0.32 (0.47)	0.32 (0.47)	0.32 (0.47)	0.33 (0.47)	0.33 (0.47)	0.34 (0.47)
Higher education	0.13 (0.34)	0.14 (0.35)	0.14 (0.35)	0.16 (0.37)	0.17 (0.37)	0.18 (0.38)	0.19 (0.39)	0.21 (0.41)	0.22 (0.47)	0.22 (0.42)	0.23 (0.42)	0.24 (0.43)
Single	0.31 (0.46)	0.30 (0.46)	0.31 (0.46)	0.31 (0.46)	0.30 (0.46)	0.29 (0.45)	0.29 (0.45)	0.29 (0.45)	0.32 (0.47)	0.31 (0.46)	0.30 (0.46)	0.30 (0.46)
Experience	15.62 (6.27)	16.42 (6.36)	17.04 (6.59)	17.77 (6.84)	18.74 (7.00)	19.44 (7.25)	20.08 (7.33)	20.73 (7.47)	21.28 (7.61)	21.76 (7.77)	22.30 (7.92)	23.00 (8.08)
Lagged unemployment	0.02 (0.11)	0.02 (0.11)	0.02 (0.11)	0.02 (0.10)	0.02 (0.11)	0.03 (0.13)	0.02 (0.12)	0.04 (0.15)	0.04 (0.18)	0.05 (0.20)	0.05 (0.18)	0.03 (0.16)
Lagged income (log)	7.64 (5.34)	7.60 (5.42)	7.42 (5.57)	7.54 (5.58)	7.97 (5.46)	7.92 (5.46)	7.89 (5.49)	7.77 (5.55)	7.87 (5.51)	7.74 (5.55)	7.63 (5.60)	7.69 (5.64)
Number of observations	1071	1047	1034	1007	994	979	973	961	1015	1067	1118	1165

*Continued ...*

continued ...

	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998
PEP												
Women	0.46 (0.50)	0.43 (0.49)	0.42 (0.49)	0.40 (0.49)	0.40 (0.49)	0.41 (0.49)	0.41 (0.49)	0.41 (0.49)	0.43 (0.50)	0.44 (0.50)	0.46 (0.50)	0.46 (0.50)
Vocational training	0.20 (0.40)	0.22 (0.41)	0.22 (0.41)	0.23 (0.42)	0.23 (0.42)	0.24 (0.43)	0.22 (0.42)	0.20 (0.40)	0.19 (0.39)	0.20 (0.40)	0.20 (0.40)	0.20 (0.40)
Higher education	0.33 (0.47)	0.33 (0.47)	0.35 (0.48)	0.35 (0.48)	0.38 (0.49)	0.39 (0.49)	0.41 (0.49)	0.45 (0.50)	0.47 (0.50)	0.49 (0.50)	0.51 (0.50)	0.52 (0.50)
Single	0.23 (0.42)	0.22 (0.41)	0.22 (0.41)	0.19 (0.39)	0.20 (0.40)	0.22 (0.41)	0.21 (0.41)	0.22 (0.42)	0.21 (0.41)	0.22 (0.41)	0.24 (0.42)	0.25 (0.43)
Experience	17.33 (8.05)	18.44 (8.21)	19.14 (8.49)	19.99 (8.75)	20.97 (8.66)	21.71 (8.81)	22.18 (9.23)	22.99 (9.35)	23.74 (9.65)	24.62 (9.89)	25.32 (10.18)	25.77 (10.55)
Lagged unemployment	0.00 (0.02)	0.01 (0.07)	0.01 (0.07)	0.01 (0.07)	0.01 (0.08)	0.01 (0.10)	0.01 (0.08)	0.01 (0.05)	0.01 (0.07)	0.01 (0.09)	0.01 (0.09)	0.02 (0.12)
Lagged income (log)	9.50 (4.51)	9.70 (4.49)	9.70 (4.54)	9.46 (4.74)	9.55 (4.72)	9.36 (4.89)	9.19 (5.00)	9.14 (4.99)	9.16 (5.01)	9.28 (4.88)	9.46 (4.78)	9.29 (4.94)
Number of observations	567	555	530	524	500	493	499	471	439	435	417	418

Average retirement age

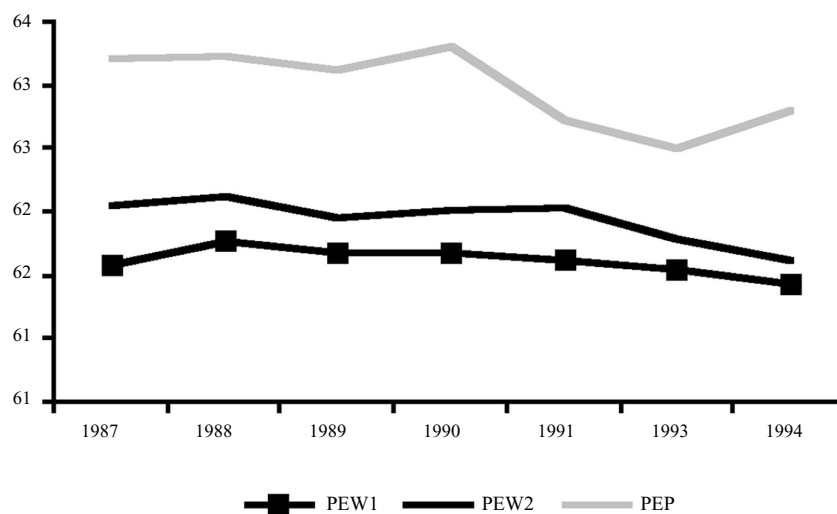


Figure A.1. Average retirement age for 60-66-year-olds, VERP1, VERP2 and PEP, 1987-1991, 1993-1994.

# Debatten om fremtidens velfærd

## Demografi, velstandsdilemma og makroøkonomiske strategier

Torben M. Andersen

Nationaløkonomisk Institut, Aarhus Universitet, E-mail: [tma@econ.au.dk](mailto:tma@econ.au.dk)

Lars Haagen Pedersen

Velfærdskommissionens sekretariat, E-mail: [lh@velfaerd.dk](mailto:lh@velfaerd.dk)

*No one likes to deliver bad news, particularly not individuals hoping to be re-elected. If the bad news is in the form of a projection, then there is a risk that the political competition will deny the bad news and the need to do something about it. Diamond (2003, s. 121)*

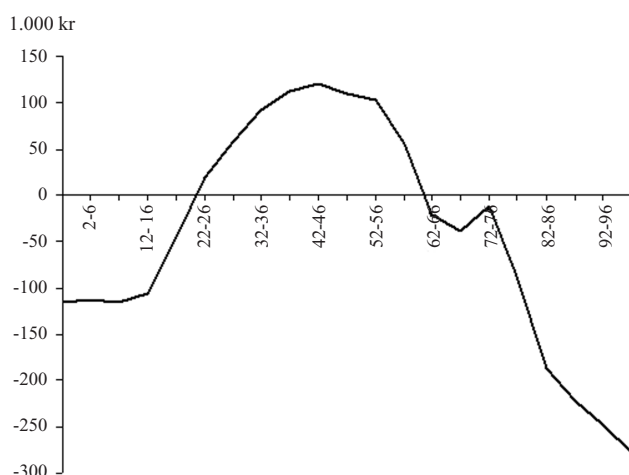
### 1. Indledning

Et væsentligt træk ved den danske velfærdsmodel er den sociale kontrakt mellem generationer. De fleste velfærdsordninger har universelle træk med individuelle rettigheder og kollektiv finansiering via skatter. Et veludbygget velfærdssamfund baseret på dette princip vil helt naturligt udvise en stærk alders- eller livscyklusprofil i forhold til både træk på velfærdsordninger og finansielle bidrag via forskellige former for skattebetalinger. Det gennemsnitlige nettomellemværende med den offentlige sektor er derfor stærkt aldersafhængigt, jf. figur 1. Børn/unge og ældre er nettomodtagere, mens personer i »midtergruppen« er nettobidragydere. Ud over denne alders- eller generationsprofil er der selvfølgelig også på et givet tidspunkt omfordeling langs andre dimensioner som indkomst, arbejdsevne, helbred osv.

På et givet tidspunkt vil de positive nettobidrag fra »midtergruppen« således være det finansielle grundlag for de nettobidrag, som børn/unge og ældre modtager. Dette udstrækker således det traditionelle overlappende generationsprincip inden for den enkelte familie, hvor »forældrene tager sig af deres børn/unge samt deres egne forældre«, til at gælde på samfundsplan. I sit udgangspunkt er den sociale kontrakt i velfærdssamfundet af »pay as you go« (PAYG) typen, dvs. de løbende skatteindtægter, som overvejende kommer fra de erhvervsaktive personer, er det finansielle grundlag for at tage sig af børn/unge og ældre.

---

Torben M. Andersen er formand for Velfærdskommissionen, og Lars Haagen Pedersen er chef for kommissionens sekretariat. Synspunkterne i dette papir er forfatternes og ikke nødvendigvis sammenfaldende med Velfærdskommissionens.



Figur 1. Nettobidrag til den offentlige sektor afhængig af alder.

Anm.: Det gennemsnitlige nettobidrag for aldersgrupperne 67-76 år er højere end for omkringliggende aldersgrupper pga. tekniske forudsætninger i DREAM vedrørende placeringen af skat og arveafgift på kapitalpensioner.

Kilde: Egne beregninger på baggrund af DREAM-modellen.

En stringent indretning af velfærdsamfundet under en PAYG-betingelse, i form af et krav om et balanceret budget, vil være u hensigtsmæssig. Det er velkendt fra litteraturen, at variationer i f.eks. de enkelte generationers størrelse under et PAYG-system vil betyde enten variationer i indbetalinger (skatter) eller i de tilbudte velfærdsordninger og ydelser, se f.eks. Bohn (2001). Herved vil usikkerhed på samfunds- eller makroøkonomisk niveau projiceres over i usikkerhed for den enkelte, med muligheder for variation i velfærdsordninger og beskatning fra det ene år til det andet. Der er således en række argumenter for, at den offentlige sektor skal udnytte kapitalmarkederne via overskud/underskud på de offentlige finanser til at sikre et mere »glat« forløb for skattebetalinger og velfærdsordninger.

Det er imidlertid en grundlæggende pointe i forhold til mulighederne for at udglatte eller diversificere stød, at dette kun er muligt for midlertidige forandringer, og dermed ikke i forhold til permanente forandringer.<sup>1</sup> I forhold til mere permanente ændringer er det derfor nødvendigt at foretage tilpasninger af enten velfærdsordningerne og/eller finansieringsgrundlaget. Der er særligt to permanente ændringer, der er relevante for indretningen af velfærdsordningerne. Den ene er den forventede ændring i befolkningens sammensætning som følge af længere forventet levetid, og den anden er den forventede fortsatte materielle velstandsstigning i samfundet.

1. Dette er velkendt fra forsikringslitteraturen. For mulighederne for diversifikation af stød via det offentlige budget se f.eks. Gordon og Varian (1988), Hansson og Stuart (1989), og Andersen og Dogonowski (2002).

Det er velkendt og veldokumenteret, at vi står over for en markant ændring i de demografiske forhold.<sup>2</sup> En vigtig årsag til dette er, at den forventede levetid blandt ældre er voksende, og at den samlede fertilitet ikke forventes at stige tilstrækkeligt til at sikre reproduktionen.<sup>3</sup> I forhold til diskussionen om tilrettelæggelsen af den økonomiske politik er det centralt, at ændringen i befolkningens sammensætning forventes at være permanent. Det betyder, at grundlaget for den sociale kontrakt i velfærdssamfundet forandres væsentligt, fordi antallet af erhvervsaktive vil falde, mens antallet af ældre vil stige. Derfor er det nødvendigt at vurdere behovet og mulighederne for tilpasning af den sociale kontrakt.

Den vedvarende tendens til stigende produktivitet over tid er ligeledes veldokumenteret. Denne leder til stigende velstand over tid. Den sociale kontrakt i det danske velfærdssamfund betyder, at velstanden fordeles på alle befolkningsgrupper. Samtidig har der historisk været en tendens til, at udgifterne til den offentlige service til den enkelte person er vokset hurtigere end indkomsterne i økonomien. Den forventede fortsatte stigning i velstanden kan også i fremtiden forventes at lede til en relativ stigning i efterspørgslen efter velfærdsservice og sundhedsydelser. Samtidig fører velstandsstigningen til et ønske om mere fritid, hvilket mindsker skatteprovenuet. Det er således et spørgsmål, om den eksisterende sociale kontrakt er robust over for den fortsatte stigning i velstanden.

Formålet med dette papir er for det første at se nærmere på de demografiske forandringer og implikationerne for de offentlige finanser. Et centralt spørgsmål er, om de eksisterende velfærds- og finansieringsordninger kan fastholdes til trods for store variationer i demografien via variationer over tid i offentlig saldo og gæld. Dette rejser spørgsmålet om finanspolitisk holdbarhed: Vil de eksisterende ordninger kunne finansieres, eller vil der opstå finansielle problemer? Hvis det sidste er tilfældet, hvad er da størrelsen af problemet? Konklusionen er, at den demografiske udvikling fører til et væsentligt finansieringsproblem, hvilket rejser spørgsmålet om, hvordan velfærdssamfundets sociale kontrakt skal tilpasses. Fastlæggelsen af en økonomisk politik i forhold hertil forudsætter en eksplicit stillingtagen til den underliggende usikkerhed. Selvom trenden er klar, er der usikkerhed, og den økonomisk politiske reaktion bør forholde sig til begge dele.

Papirets andet formål er at vurdere betydningen af den fortsatte stigning i den materielle velstand for de offentlige finanser. Velstandsstigningen betyder på den ene side, at den demografiske udfordring mere er et generationsfordelings- end et forsørgelses-spørgsmål, og på den anden side, at velfærdssamfundet også kan blive presset af øn-

---

2. Se også f.eks. DØR (2004), OECD (2005).

3. I forhold til de økonomiske konsekvenser er middellevetidsstigningen den afgørende faktor, jf. afsnit 2 og følsomhedsberegningerne i afsnit 8.

sker om øget service og mere fritid. Konklusionen er, at en fortsat relativ stigning i den offentligt finansierede velfærdsservice og en gradvis reduktion i arbejdstiden tilsammen vil føre til et meget betragteligt finansieringsproblem, som kommer i tillæg til det demografisk betingede pres.

Papirets tredje formål er at diskutere tilrettelæggelsen af den økonomiske politik i lyset af den nuværende indretning af velfærdssamfundet og den forventede udvikling i demografi og materiel velstand.

Udgangspunktet for diskussionen er en vurdering af konsekvenserne af eksisterende velfærdsordninger. Dette gøres dels i et forløb, hvor der er set bort fra den historiske trend i velfærdsservice og årlig arbejdstid, dels i et forløb, hvor disse udviklinger er inkluderet. For et givet fremtidigt forløb vurderes de finansielle problemers omfang, og der skabes en referenceramme for en diskussion af mulige reformer.

Som illustration vælger vi i dette papir at se på situationer, hvor stigningen i de offentlige udgifter finansieres ved stigninger i den statslige bundskattesats. Det understreges, at dette er illustrationer. Det er naturligvis muligt at ændre den økonomiske politik, så f.eks. besparelser på offentlig service eller omlægninger af overførselsindkomsterne kan bidrage til at løse finansieringsproblemet.

Skattestigningen betragtes under to forskellige tilgangsvinkler. Den ene er en finansieringsløsning af PAYG-typen med gradvise skattestigninger i takt med de demografiske forandringer (og den forøgede velfærdsservice) for at sikre finansieringen af eksisterende velfærdsordninger. Den anden er en umiddelbar og permanent ændring i skatterne for at sikre finanspolitisk holdbarhed, dvs. at nutidsværdien af forventede indtægter svarer til nutidsværdien af de forventede udgifter. De to tilgangsvinkler har vidt forskellige implikationer for risikospredning og fordelingen af finansieringskravene på nulevende og fremtidige erhvervsaktive. Dette danner udgangspunktet for en diskussion af en økonomisk politisk strategi i forhold til det finansieringsproblem, velfærdssamfundet står over for. Velfærdssamfundet står herudover over for en række andre udfordringer, f.eks. globalisering og øget individualisering. Disse emner tages ikke op her, se i stedet Velfærdskommissionen (2004, 2005a).

Dette papir er organiseret som følger: Afsnit 2 giver en kort introduktion til de demografiske fremskrivninger, der ligger til grund for de efterfølgende analyser. I afsnit 3 fastlægges fortolkningen af uændrede regler og de umiddelbare effekter af de demografiske forskydninger på de offentlige finanser. I afsnit 4 opstilles to muligheder for vurdering af finansieringsproblemet knyttet til de demografiske forskydninger og velstandsdilemmaet, nemlig, PAYG-finansiering og den holdbare skat. Resultaterne for en PAYG-finansiering præsenteres i afsnit 5, og resultaterne for en holdbar skat i afsnit 6, mens de to tilgangsvinkler sammenlignes i afsnit 7. Følsomhedsberegning og usik-



kerhed knyttet til en vurdering af det underliggende finansieringsproblem præsenteres i afsnit 8, mens afsnit 9 indeholder en diskussion af mulige strategier, og der formuleres en pragmatisk »tilpasningsstrategi«. Afsnit 10 giver nogle få afsluttende bemærkninger.

## 2. Demografi

I forhold til den sociale kontrakt, som er indbygget i velfærdssamfundet, har Danmark i udgangssituationen en meget gunstig alderssammensætning af befolkningen. Årgangene af nettobetelere er store, mens årgangene af nettomodtagere af ydelser og service er relativt små. Denne situation er imidlertid gradvist ved at ændre sig, og ifølge Velfærdskommissionens befolkningsfremskrivning er den underliggende trend en stigning i antallet af nettomodtagere, mens antallet af nettobetelere vil falde.

En del af denne udvikling hænger sammen med ekkovirkninger af de store efterkrigsårgange og de små årgange som følge af den lave fertilitet i 1980'erne. Man kunne derfor argumentere for, at der er tale om temporære forskydninger i befolkningens alderssammensætning. Det er imidlertid ikke tilfældet. Ekkovirkningerne er blot med til at fremrykke en generel tendens til, at andelen af ældre i befolkningen vokser. Det skyldes, at befolkningen ikke længere er voksende, samt at middellevetiden formentligt vil fortsætte med at vokse primært som følge af stigende restlevetid for ældre aldersgrupper.<sup>4</sup> Denne udvikling betyder en permanent tendens til stigning i ældreandelen i befolkningen.

Fremskrivningen af befolkningens størrelse og alderssammensætning baseres på antagelser om den fremtidige fertilitet, udviklingen i den alders- og kønsbetingede dødelighed og på nettoindvandringen. I det følgende gennemgås kort forudsætningerne bag Velfærdskommissionens befolkningsfremskrivning.<sup>5</sup>

Den samlede fertilitet fastholdes på det nuværende niveau omkring 1,7 barn pr. kvinde. Dette er en konsekvens af en sammensætningseffekt. På den ene side indebærer den estimerede udvikling en svag tendens til fald i den samlede fertilitet for alle befolkningsgrupper fordelt efter herkomst. På den anden side vil indvandrere og efterkommere fra mindre udviklede lande komme til at udgøre en stigende andel af befolkningen, og da disse grupper har den højeste fertilitet, tenderer dette til at øge den samlede fertilitet. Som en konsekvens af disse to modsatrettede effekter fastholdes det gennemsnitlige niveau, jf. Markeprand, Pedersen og Stephensen (2003).

---

4. Den fremtidige udvikling i middellevetiden står i nogen grad i modsætning til den historiske stigning. Sidstnævnte er karakteriseret ved, at middellevetidsstigningerne i høj grad hang sammen med fald i spædbørnsdødeligheden. Denne udvikling har ikke samme betydning for befolkningens alderssammensætning som den fremtidige udvikling.

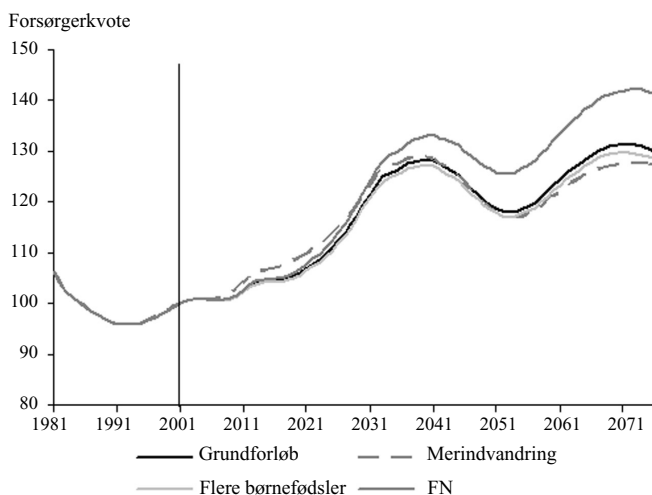
5. Kapitel 10 i Velfærdskommissionen (2004) indeholder en grundig beskrivelse af usikkerhederne ved befolkningsudviklingen.

Udviklingen i den alders- og kønsbetingede dødelighed er estimeret ved anvendelse af metoden fra Lee og Carter (1992) og dokumenteret i Haldrup (2004). Fremskrivningen indebærer, at de seneste års relativt begrænsede stigning i middellevetiden antages at fortsætte, således at den fremtidige årlige vækst ligger på mellem 0,08 og 0,09 år, eller ca. 1 måned om året. Den danske vækst i middellevetiden har i perioden fra 1960-2000 været lav i internationalt perspektiv. Blandt OECD-landene har kun Slovakiet og Ungarn oplevet en lavere vækst end Danmark (sidstnævnte endda kun for mænd). Fremskrivningen indebærer, at middellevetiden i Danmark i år 2040 svarer til middellevetiden i Sverige i dag. Ifølge FN's befolkningsfremskrivning forventes væksten i middellevetiden i det øvrige Vesteuropa at ligge på omkring 0,2 år pr. år. Dette niveau svarer til den historiske vækst i middellevetiden i disse lande. Med Velfærdskommissionens fremskrivning antages således, at Danmark fortsat vil have lavere vækst i middellevetiden, end der forventes for de øvrige Vesteuropæiske lande. Da middellevetiden i disse lande i forvejen er højere end i Danmark, indebærer fremskrivningen, at denne forskel udbygges. De seneste tre års vækst i middellevetiden i Danmark har været højere end den seneste historiske udvikling og har med en vækst på 0,2-0,3 år pr. år ligget på det vesteuropæiske gennemsnit. Fastholdes de seneste års udvikling, indebærer Velfærdskommissionens fremskrivning derfor en undervurdering af væksten i middellevetiden.<sup>6</sup>

Fremskrivningen af ind- og udvandring er usikker, bl.a. fordi den både er påvirket af udlændingepolitikken i såvel Danmark som udlandet og forekomsten af krigs- eller katastrofesituationer. Det er vanskeligt at basere fremskrivningen på en historisk fremskrivning pga. ændringen i den danske udlændingelovgivning medio 2002, som har medvirket til et fald i indvandringsniveauet.<sup>7</sup> I Velfærdskommissionens fremskrivning lægges til grund, at det nuværende niveau for bruttoindvandringen stort set fastholdes, og at fordelingen på typer af opholdsgrundlag og sammensætningen efter oprindelsesland forbliver uændret. Dog forudsættes en fortsat vækst i antallet af familiesammenførte som følge af, at antallet af potentielle modtagere af familiesammenførte vokser. For udvandringen antages, at de alders-, køns- og herkomstbetingede udvandrings sandsynligheder fastholdes på et niveau, der svarer til gennemsnittet i de seneste 5 år. Forudsætningerne fører til en forventet årlig nettoindvandring på omkring 7.000 personer. Der er set bort fra en mulig tendens til øget vandring som følge af globaliseringen. Der kan derfor være en tendens til at undervurdere vandringerne – måske særligt vedrørende udvandringen.

6. En del af den forøgede stigning skyldes »engangs-fald« i spædbørnsdødeligheden og påvirker derfor ikke aldersfordelingen på samme måde som en stigning i restlevetiden for ældre, jf. Danmarks Statistik (2005).

7. For en gennemgang af udviklingen i indvandringen og dennes sammensætning efter opholdsgrundlag, herkomst og alder henvises til Velfærdskommissionen (2005a).



Figur 2. Den demografiske forsørgerbrøk i hovedforløbet og under alternative antagelser.

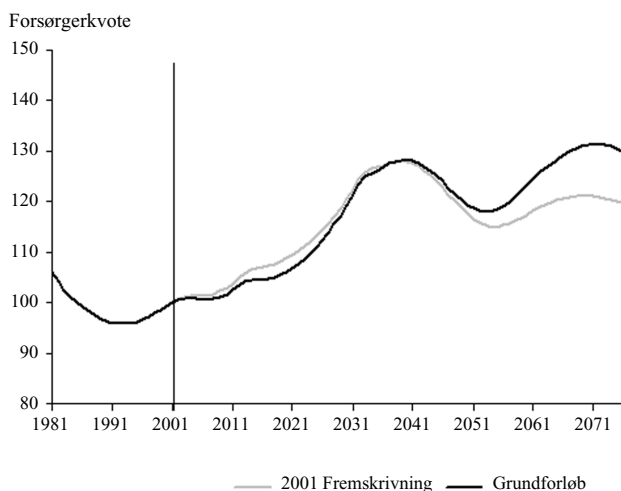
Kilde: Velfærdskommissionens befolkningsfremskrivning og UN (2004).

Ovenstående forudsætninger betyder, at den samlede befolkning reduceres gradvist fra 2010 og frem. I 2026 forventes den samlede befolkning at have samme størrelse som i 2003, dvs. omkring 5,4 mio. personer. I år 2040 er befolkningen reduceret til omkring 5<sup>1</sup>/<sub>4</sub> mio. Faldet forventes at fortsætte, således at befolkningen er under 5 mio. i 2070.

De offentlige finanser afhænger ikke primært af befolkningens størrelse men af dens sammensætning efter oprindelse og alder, jf. livscyklussammenhængen i figur 1. Fremskrivningen betyder, at antallet af personer i alderen 15 til 64 år falder med 10 procent fra 2002 til 2040 og med 16 procent frem til 2080.<sup>8</sup> I de samme to perioder stiger antallet af ældre (65+) med henholdsvis 52 og 47 procent. Disse to forhold får den demografiske forsørgerbrøk til at stige med 27 procent fra 2002 til 2040 og med 28 procent til 2080. Den demografiske forandring er således ikke midlertidig men permanent.<sup>9</sup> Samtidig stiger indvandrere og efterkommere fra mindre udviklede landes andel af befolkningen meget betydeligt. I 2003 udgjorde disse befolkningsgrupper 4,7 procent af befolkningen, og denne andel vil vokse til 11,8 procent i 2040. Dette har særlig betydning for de offentlige finanser, da gennemsnittet af personer i disse to be-

8. Den demografiske forsørgerbrøk defineres som antallet af børn, unge (0-14-årige) og ældre (65+ årige) i forhold til antallet af personer i alderen 15-64 år.

9. Befolkningsfremskrivninger har hidtil haft den implikation, at ændringerne i forsørgerkvoten var mere midlertidige.



Figur 3. Den demografiske forsørgerbrøk i grundforløbet og i fremskrivningen fra 2001.

Kilde: Velfærdskommissionens befolkningsfremskrivning og 2001 befolkningsfremskrivningen (DREAM).

folkningsgrupper har lav arbejdsmarkedstilknytning, og de er derfor i betydeligt omfang nettomtagere af offentlige ydelser over livet.

På grund af den betydelige usikkerhed på hver af de tre typer befolkningsændringer – fertilitet, levetid og indvandring – er udviklingen i den samlede befolkning på længere sigt også usikker. Usikkerhed omkring de tre centrale demografiske faktorer har imidlertid meget forskellig betydning for udviklingen i befolkningens alderssammensætning. Følsomheden af den demografiske forsørgerbrøk i forhold til disse tre demografiske faktorer fremgår af figur 2. Ændringerne er kvantificeret, således at de har sammenlignelig effekt på den samlede befolkning på langt sigt. I alle alternativer er befolkningen mellem 4 og 5 procent større efter 75 år. Den demografiske forsørgerbrøk er relativt robust over for ændringer i fertilitet og indvandring, men yderst påvirkelig af ændringer i middellevetiden.

Årsagen til at middellevetiden har en kvalitativt anderledes effekt på den demografiske forsørgerbrøk end fertilitet og indvandring er, at stigninger i de to sidstnævnte fører til, at der bliver flere personer i alle faser af livscyklusen, mens øget middellevetid i fremtiden hovedsageligt fører til, at antallet af personer på 60 år og derover stiger. Øget middellevetid påvirker således stort set alene tælleren i den demografiske forsørgerbrøk, mens fertilitet og indvandring påvirker såvel tæller som nævner. Konklusionen er derfor, at stigninger i middellevetiden direkte påvirker forholdet mellem alders-

grupperne i befolkningen og dermed alt andet lige får stor effekt på velfærdssamfundets implicite sociale kontrakt.

Der knytter sig betydelig usikkerhed til udviklingen i middelevetiden, hvilket også afspejles i, at den forventede udvikling i befolkningens alderssammensætning har ændret sig i de senere år, jf. figur 3, der sammenligner Velfærdskommissionens fremskrivning med DREAMs<sup>10</sup> fremskrivning fra 2001.

Velfærdskommissionens fremskrivning er baseret på en større vækst i middelevetiden end DREAMs fremskrivning fra 2001. Det betyder, at vurderingen af udviklingen i den demografiske forsørgerbrøk er ændret. Med den tidligere befolkningsfremskrivning var det demografiske problem overvejende af temporær karakter, mens den permanente effekt på forsørgerbrøken var relativt beskeden. Med en øget stigning i middelevetiden bliver der en betydelig større permanent effekt på forsørgerbrøken. Med andre ord er der en mere markant permanent stigning i forsørgerbrøken i Velfærdskommissionens fremskrivning. Forskellen mellem de to befolkningsfremskrivninger understreger både betydningen af, om det demografiske problem er af midlertidig eller mere permanent karakter, men også usikkerheden knyttet til vurderingen. Denne usikkerhed med hensyn til trenden i forsørgerbrøken er afgørende for diskussionen af robustheden af den økonomiske politik, jf. nedenfor.

### 3. Uændrede velfærdsordninger

Udgangspunktet for at vurdere konsekvenserne af de demografiske forskydninger er at betragte et forløb med uændrede velfærds- og skatteordninger. Herved skabes en referenceramme for at vurdere konsekvenserne af de demografiske ændringer og et grundlag for diskussionen af mulige økonomisk politiske reaktioner.

Uden nærmere analyse følger det umiddelbart ved kombination af den demografiske udvikling vist i figur 2 og de aldersafhængige nettobidrag til den offentlige sektor vist i figur 1, at de offentlige finanser vil blive belastet. Umiddelbare overslag over de økonomiske konsekvenser kan nemt gøres ved at kombinere ændringer i antallet af erhvervsaktive og ældre med de gennemsnitlige nettobidrag.<sup>11</sup>

En mere dybtgående vurdering må imidlertid nødvendigvis være baseret på en grundig modellering af ordninger og regler i det danske velfærdssystem. Ikke mindst

10. DREAM er en forkortelse for Danish Rational Economic Agents Model. For en beskrivelse af modellen, se Knudsen m.fl. (1998, 1999) og Pedersen, Stephensen og Trier (1999). Generel information om modellen findes på <http://www.dreammodel.dk>

11. I 2041 er der ca. 350.000 færre erhvervsaktive med et gennemsnitligt årligt nettobidrag på ca. 100.000 kr. og ca. 400.000 flere over 65 år med et gennemsnitligt nettotræk på ca. 100.000. De offentlige finanser i 2041 vil derfor umiddelbart forværres med 75 mia. kr. i forhold til i dag. I grundforløbet med uændrede regler, jf. nedenfor, er budgetforværringen til 2041 ca. 90 mia. kr.. En væsentlig årsag til denne forskel er, at »kuvertberegningen« ikke tager højde for at indvandrere og efterkommere kommer til at udgøre en større andel af befolkningen.

er det vigtigt at tage hensyn til samspillet med andre faktorer, f.eks. opbygning af pensionsformue og implikationer heraf for det fremtidige skatteprovenu og pensioner. Velfærdskommissionens analyser er baseret på den økonomiske model DREAM, der blandt andet udmærker sig ved en grundig modellering af eksisterende ordninger og regler, samt en overlappende generationsstruktur. Sidstnævnte er helt centralt for at gennemføre fremadrettede analyser af demografiske forskydninger. Særligt ved ændringer i den økonomiske politik er det vigtigt med en modelbaseret analyse, dels for at konkretisere virkningerne af ændret politik, dels for at analysere mulige adfærdsvirkninger heraf. En vurdering af f.eks. det »holdbare skattniveau«, jf. senere, er nødvendigvis modelbaseret.

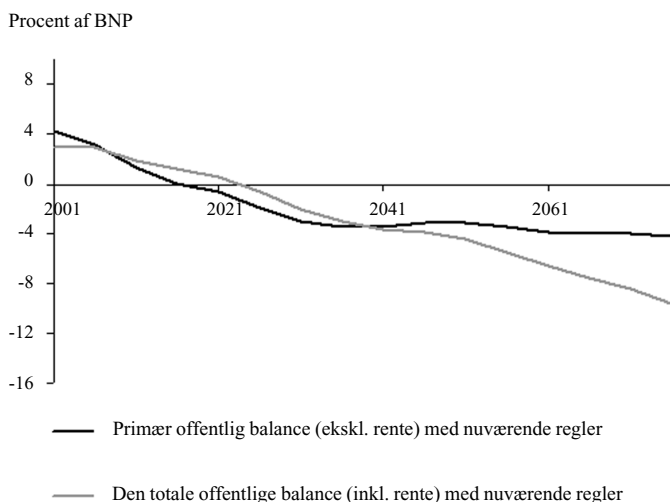
Det beregningsmæssige udgangspunkt er derfor en fastlæggelse af, hvad der skal forstås ved »eksisterende ordninger« og dermed fremskrivningens fortolkning af »uændrede regler«. <sup>12</sup> De væsentlige antagelser i Velfærdskommissionens beregninger er:

- Overførselsindkomster pr. individ reguleres ifølge satsreguleringsloven. Det antages, at satspuljen fremover anvendes til reguleringen. Det svarer til, at overførselsindkomsterne reguleres i takt med lønudviklingen efter fradrag for pensionsindbetalinger. Der er herudover taget højde for, at den øgede udbredelse af udbetalinger fra arbejdsmarkedspensioner fører til en reduktion i de indkomstafhængige overførsler til pensionister. <sup>13</sup>
- Udbredelsen af de enkelte typer af overførselsindkomst er til personer uden for arbejdsstyrken antages at være konstant i fremtiden. Det svarer til, at andelen af en given befolkningsgruppe (fordelt på køn, alder og oprindelse), som modtager en given ydelse, er uændret over tid. Der er herudover taget højde for effekten af allerede gennemførte ændringer i reglerne (som endnu ikke har fået fuld effekt) for modtagelse af overførselsindkomst og visse andre justeringer, jf. Velfærdskommissionen (2004, kap. 11).
- Forbruget af offentlig service (individuel offentlig forbrug) har en bestemt fordeling i forhold til henholdsvis alder, køn og oprindelse. Den gennemsnitlige udgift pr. person i en given gruppe reguleres med produktivitetsstigningen og inflationen, svarende til en nominel vækst på omkring 4 procent pr. år. Reguleringen svarer stort set til, at de gennemsnitlige udgifter pr. individ i en given befolkningsgruppe reguleres med lønudviklingen.

---

12. Der anvendes bevidst ikke betegnelsen »uændret politik«. Det skyldes, at regeringens 2010-plan har forudsætninger, der frem til 2010 afviger fra de her anvendte. 2010-planen indebærer økonomiske indgreb (der ikke er udmøntet), som skal ændre forholdet mellem antallet af personer i privat beskæftigelse og overførselsindkomstmodtagere, således at der flyttes ca. 60.000 personer fra den sidste gruppe til den første. Endvidere skal der frem til 2010 føres en stram udgiftspolitik, som betyder, at væksten i den offentlige service er lavere end svarende til den her anvendte forudsætning.

13. Se Velfærdskommissionen (2005d) for en diskussion af satsreguleringsordningen.



Figur 4. Udviklingen i den primære og den samlede budgetsaldo.

Kilde: Egne beregninger på DREAM-modellen.

- Kollektivt offentligt forbrug reguleres med væksten i BNP.
- Offentlige investeringer fastlægges, så kapital/output-forholdet i den offentlige sektor er konstant.
- Skatte- og afgiftssatser antages at blive fastholdt uændrede, bortset fra mængdeafgifter, der reguleres som følge af prisstigninger.<sup>14</sup>
- Uændret årlig arbejdstid og deltidsandel.

Sammenfattende tager disse forudsætninger udgangspunkt i, at der i fremtiden skal være adgang til de samme velfærdsordninger og tilbud som i dag, at bidragene til finansiering via skatter er uændrede, samt at indkomstfordelingen i bred forstand (erhvervsaktive vs. ikke erhvervsaktive) er uændret.

En fremskrivning baseret på disse forudsætninger giver et forløb for de offentlige indtægter og udgifter som vist i figur 4. Figuren afspejler, at den trendmæssige stigning i den demografiske forsørgerbrøk bliver overført til et trendmæssigt fald i den offentlige sektors primære saldo. Den primære saldo udviser i udgangssituationen et overskud. I takt med at de demografiske forskydninger slår igennem, vil der imidlertid gradvist ske en bevægelse hen imod en situation med underskud på knap 4 procent af BNP omkring år 2060. Denne sammenhæng mellem den demografiske udvikling og

14. Regeringens skattestop indebærer, at realværdien af boligbeskatningen og visse afgifter gradvist reduceres. Skattestoppet er i beregningerne fastholdt til 2010.

den primære offentlige saldo skyldes som nævnt, at den sociale kontrakt indebærer en omfordeling fra midtergenerationen (som reduceres) til de yngre og de ældre generationer (som bliver større over tid).

Sammenhængen betyder, at usikkerhed om middellevetidsudviklingen – der medfører usikkerhed om trenden i forsørgerbrøken – fører til usikkerhed om trenden i det primære offentlige budget.

Beregningen sammenfattet i figuren viser Velfærdskommissionens fremskrivning baseret på antagelserne om uændrede regler. Dette forløb er grundlaget for at diskutere, hvordan forskellige alternative finansieringsformer påvirker makroøkonomien og fordelingen over generationer. På baggrund heraf diskuteres forhold af betydning for at fastlægge den optimale politikrespons. Diskussionen inddrager såvel den fundne trend i det offentlige budget som den usikkerhed, der knytter sig til fremskrivningen.

### *3.1 Velstandsdilemmaer*

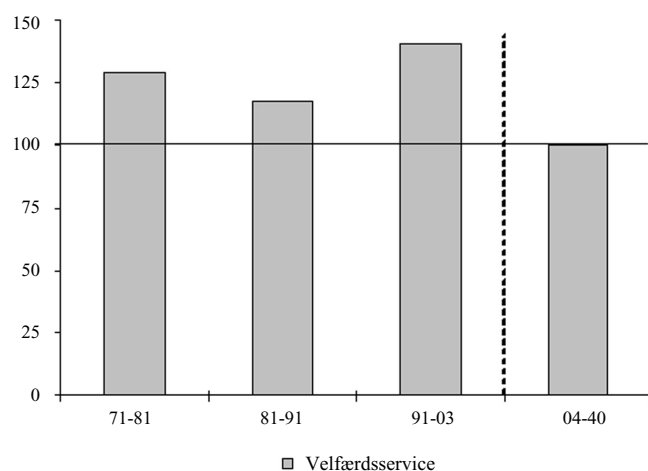
Det ovenfor præsenterede grundforløb tager udgangspunkt i aktuelle velfærdsordninger m.m. Historisk har udviklingen af velfærdssamfundet været påvirket af udviklingen i den materielle velstand, og fremadrettet forudsættes en fortsat velstandsfremgang drevet af årlige produktivitetstigninger på 2 procent om året. Det rejser spørgsmålet, om grundforløbet har en »status quo«-bias ved ikke at tage højde for, hvorledes fortsatte velstandstigninger vil påvirke krav til og behov for såvel offentlig velfærdsservice som fritid. To forhold der er af afgørende betydning for fremtidig udvikling i offentlige udgifter og indtægter.

Grundforløbets tolkning af uændrede regler indebærer, at offentlig service (pr. person i en given alder, køn og oprindelse) vokser med den generelle vækst i økonomien. Hertil kommer, at antagelsen om konstant arbejdstid pr. år betyder, at arbejdstiden forudsættes at være uafhængig af indkomst. Begge disse antagelser er udtryk for et væsentligt anderledes forløb end det historisk observerede.

Udgifterne til velfærdsservice er vokset fra 13,3 procent af BNP i 1971 til 17,8 procent i 2001. Det svarer til en mervækst i forhold til BNP på 0,15 procent pr. år. I grundforløbet antages det, at udgifterne til velfærdsservice pr. person vokser med produktivtetsvæksten plus inflationen (svarende til timelønnen). For at vurdere denne antagelse er udgiftsvæksten i perioden fra 1971 til 2001 korrigeret for den demografiske udvikling i perioden for at få et billede på den underliggende vækst. Resultatet er, at udgiftsvæksten til velfærdsservice pr. person har ligget betydeligt over væksten i timelønnen for en offentlig ansat, jf. figur 5.

For hver af de tre 10-års perioder er udgiftsvæksten mellem 17 og 40 procent højere end væksten impliceret af antagelsen om uændrede regler. I Velfærdskommissionen (2004) kapitel 6 gives en oversigt over mulige forklaringer på denne mervækst. Her





Figur 5. Udgiftsvækst korrigeret for demografi i procent af vækst i timeløn for offentligt ansatte.

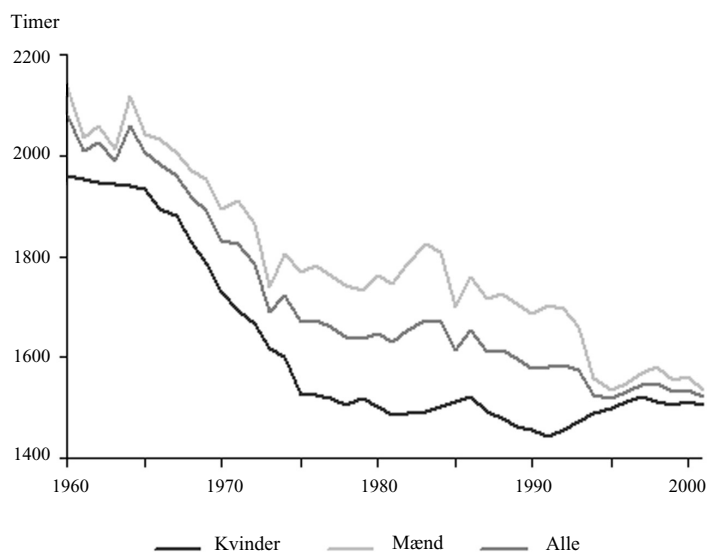
Anm.: Korrektionen af udgifterne er foretaget ved at sammenholde de alders- og kønsfordelte udgifter til velfærdsservice 2001 med befolkningssammensætningen i øvrige år. Korrektionen af BNP er foretaget ved at sammenholde beskæftigelsesfrekvenserne i 2001 med befolkningssammensætningen i øvrige år.

Kilde: Danmarks Statistik, ADAMs databank og egne beregninger bl.a. på baggrund af DREAM-modellen.

skal blot henvises til to grundlæggende årsager, kendt som hhv. Wagners og Baumols lov. Wagners lov refererer til det empiriske fænomen, at væksten i den reale indkomst fører til en mere end proportional stigning i efterspørgslen (dvs. en indkomstelasticitet større end 1) efter service eller velfærdsydelser. Da disse i stort omfang stilles til rådighed af den offentlige sektor, får dette særlig betydning for de offentlige finanser. Baumols lov henviser til det fænomen, at der er en langsommere produktivitetsudvikling i serviceerhverv, hvilket fører til en relativ prisstigning på serviceydelser.

Samtidig må den stigende velstand forventes at føre til et ønske om reduktioner i arbejdstiden f.eks. den årlige arbejdstid eller tidligere tilbagetrækning fra arbejdsmarkedet. Velstandsstigningen må forventes delt mellem øget forbrug og fritid. Betragtes den historiske udvikling har der været en generel faldende trend i det årlige antal arbejdstimer. Fra 1971 til 2001 er arbejdstiden i gennemsnit faldet med 0,5 procent årligt. Der er dog siden midten af 1970'erne en tendens til en lidt fladere trend. I denne periode ligger den gennemsnitlige årlige reduktion på mellem 0,3 og 0,4 procent pr. år, jf. figur 6.

Øget velstand skaber således det finansielle dilemma for velfærdssamfundet, at det både øger udgiftspresset og mindsker skattebasen (fritid er ubeskattet). Dette er knyttet til indretningen af velfærdssamfundet, hvor den enkeltes rettigheder til ydelserne er



Figur 6. Udviklingen i den årlige arbejdstid 1960 – 2001.

afkoblet fra betalingen via skatterne. Det skaber en eksternalitet (common pool-problem), idet den enkelte ikke tager højde for arbejdstidens betydning for finansieringen af det offentlige forbrug. Den enkelte afvejer fritid i forhold til private forbrugsmuligheder, uden at tage højde for konsekvenserne for de offentlige indtægter og dermed kollektive forbrugsmuligheder. Samfundets muligheder for at anvende velstandsstigninger på øget service og øget fritid er ikke sammenfaldende med de privatøkonomiske. Den lavere årlige arbejdstid muliggjort af øget vækst betyder derfor et dobbelt pres på finansieringen af velfærdsydelserne, fordi udvikling ikke automatisk fører til en lavere efterspørgsel efter offentlig service. Samtidig bliver denne service relativt dyrere at producere (Baumols lov).

Det er en konsekvens, at forudsætningen om uændrede regler kan dække over et stigende spænd mellem ønsker og forventninger på den ene side og den faktisk leverede offentlige velfærdsservice på den anden. Fortsat økonomisk vækst vil derfor skabe et underliggende pres for en stigning i den offentlige services andel af den samlede økonomi. At modvirke dette stiller i sig selv store krav til den økonomisk politiske styring. Selvom en sådan styring eventuelt kan implementeres, vil et stigende spænd mellem efterspørgslen og det offentlige udbud af velfærdsservice have afgørende betydning for opbakningen til velfærdssamfundet. Det er således i sig selv en central udfordring at adressere dette problem.

Historisk har dette velstandsdilemma ikke været så tydeligt, dels fordi udviklingen af velfærdssamfundet er en relativ ny ting, dels fordi øget erhvervsdeltagelse for kvinder samt øget beskatning har medvirket til at sikre finansieringen. Fremadrettet rejser dette en række spørgsmål dels i forhold til forventningerne til velfærdssamfundet, dels fordi mulighederne for at øge erhvervsdeltagelse og beskatning ikke er de samme som tidligere.

Velstandsdilemmaet – øget behov og krav til service samt mere fritid – vil forstærke det finansielle pres forårsaget af de demografiske forskydninger. Til illustration heraf er det som alternativ antaget, at udviklingen i den offentlige velfærdsservice følger den historiske trendmæssige udvikling, således at den vokser med 0,15 procent mere pr. person pr. år. Tilsvarende er det antaget, at der er en trendmæssig reduktion i den årlige arbejdstid. Denne er antaget at være på ca. halvdelen af det historisk observerede, således at denne falder med 0,15 procent eller ca. 2 timer om året de første år.<sup>15</sup> Den forøgede velfærdsservice og reducerede arbejdstid fører til en betydelig mere skarp nedadgående trend i den samlede budgetbalance, jf. figur 7.

Den samlede konklusion er derfor, at såvel den demografiske udvikling som velstandsdilemmaet sætter de offentlige finanser under pres. Den permanente vækst i middellevetiden fører til en faldende trend i den offentlige budgetsaldo for uændrede regler. Hvis velfærdsservice og arbejdstid udvikler sig svarende til den historiske trend, forstærkes den trendmæssige forværring af de offentlige finanser betydeligt.

#### **4. Finansieringskrav og -muligheder**

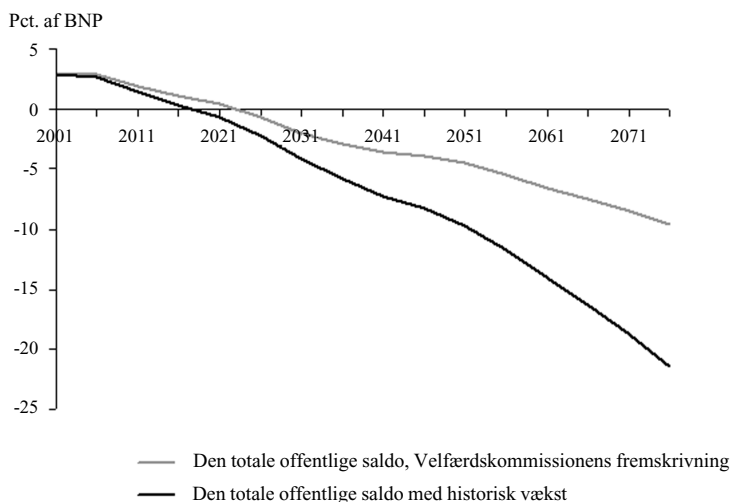
Det faktum, at velfærdssamfundet står over for et finansielt problem rejser to hovedspørgsmål. For det første, hvad er størrelsesordenen af problemet og dermed af de nødvendige tilpasninger? For det andet, hvorledes skal tilpasningen placeres tidsmæssigt?

I udgangssituationen er de offentlige finanser i en gunstig position. Fremskrivningerne fremlagt her skal således ikke tolkes som »fallitscenarier«. Udgangspunktet for diskussionen er netop de muligheder, der skabes som følge af øget middellevetid og øget materiel velstand. Begge forhold er af central betydning for velfærden. Disse forbedringer efterlader imidlertid nogle udfordringer for velfærdssamfundet. Sigtet med det følgende er derfor at få en indikation på omfanget af de nødvendige tilpasninger. Hvordan og hvornår disse skal gennemføres er de interessante økonomisk politiske spørgsmål.

Tilpasning kan sikres ved en uendelighed af politikker i form af ændringer i instrumenter af betydning for indtægter eller udgifter nu eller i fremtiden. For overskuelighedens skyld er det derfor nødvendigt at pålægge analysen disciplin og samtidig sikre,

---

15. Arbejdstidsreduktionen er antaget ikke at føre til ændringer i produktiviteten, hvilket kan føre til en overvurdering af de negative finansielle konsekvenser af denne.



Figur 7. Samlet budgetsaldo med og uden trend i velfærdsservice og arbejdstid.

Kilde: Egne beregninger på DREAM-modellen.

at tilpasningskravene kan udtrykkes på en fortløkkelig måde. I det følgende vil tilpasningskravene blive vurderet ved tilpasningen i bundskattesatsen. Da tolkningen af »uændrede regler« indebærer uændrede skattesatser, fremkommer der ved sammenligning af den initiale skattesats og den beregnede skattesats en indikator for tilpasningsbehovets størrelse. Denne indikator er et værktøj til vurdering af det underliggende tilpasningsbehov. Da indikatoren pr. konstruktion er baseret på én variabel, f.eks. en skattesats, er det umiddelbart oplagt, at man ikke herfra automatisk kan slutte til den optimale politikreaktion.

I princippet kan indikatoren beregnes for et vilkårligt økonomisk politisk instrument. Indikatoren bliver imidlertid kun interessant, såfremt den er udmålt på et instrument af tilstrækkelig tyngde. Det kan være en central skattesats udmålt på en relativ bred skattebase, som f.eks. bundskattesatsen. Dette har også den fordel, at det er et instrument, der umiddelbart kan kontrolleres økonomisk politisk.

Indikatoren beregnet ved bundskattesatsen har den fordel, at den er meget konkret. I forhold til den økonomisk politiske debat er det imidlertid vigtigt at understrege, at forskellige tilpasningsmuligheder kan bringes i anvendelse. For at illustrere dette kan der også beregnes en indikator udmålt på f.eks. offentlige udgifter eller privat beskæftigelse. Velfærdskommissionen (2004) ser på alle tre typer indikatorer. For de to sidstnævnte gælder imidlertid, at de er mindre tæt knyttet til et konkret økonomisk politisk instrument og derfor sværere at tolke.

I relation til den tidsmæssige placering af tilpasningen er der selvsagt mange muligheder. For at vurdere og diskutere betydningen heraf betragtes i det følgende to yderligere muligheder.

*PAYG-finansiering:* Her antages der en løbende tilpasning af bundskattesatsen, således at velfærdsordninger er sikret finansiering, dvs. de offentlige udgifter i et givet år finansieres ved offentlige indtægter i samme år. Konkret antages det, at politikken tilrettelægges, således at den offentlige finansielle formue målt som andel af BNP er uændret igennem det beregnede forløb. Med denne finansiering udnyttes kapitalmarkederne ikke til at fordele tilpasningsbehovet over tid.

*Holdbar skat:* Her tilvejebringes finansieringen ved en permanent ændring i skattesatsen, således at de fremtidige udgifter er sikret finansiering, dvs. den intertemporale budgetrestriktion er overholdt. Da den underliggende profil i den finansielle udvikling er en bevægelse fra overskud til underskud, implicerer denne tilgangsvinkel, at kapitalmarkederne udnyttes til at fordele tilpasningsbyrden over tid.

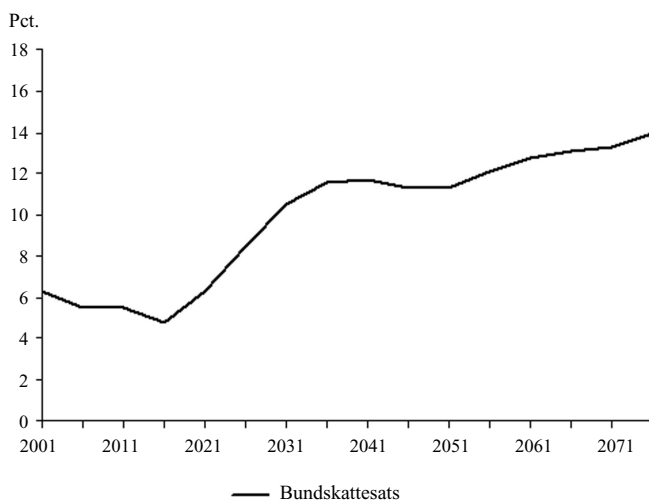
I det følgende præsenteres først beregninger under disse to forskellige tilgangsvinkler, som udspænder yderpunkterne i den mulige tidsmæssige placering af tilpasningskravet udmålt ved ændringer i bundskattesatsen. Dernæst præsenteres følsomhedsberegninger. Efterfølgende diskuteres fordele og ulemper ved disse to fremgangsmåder.

### **5. Pay as you go-finansiering**

Den gradvise ændring i befolkningens sammensætning i retning af flere ældre og færre i den erhvervsaktive alder, og den deraf følgende forværring af de offentlige finanser stiller under PAYG-finansiering krav om løbende tilpasninger af bundskatten. Kravene til bundskatten fremgår af figur 8.

PAYG-finansieringen antages at blive introduceret fra 2011, hvor den offentlige sektor i fremskrivningen fortsat har et overskud, der er større end det, der skal til for at fastholde en konstant formue i procent af BNP. Det betyder, at bundskatten initialt kan reduceres og vil ligge under det nuværende niveau i en ca. 10-årig periode. Niveaulet vil derefter være hastigt stigende til knap 12 procent – eller 6,5 procentpoint over det nuværende niveau – i 2041. I de følgende 20 år stabiliseres niveaulet, hvorefter stigningen forsætter, og bundskattesatsen når over 16 procent i slutningen af århundredet.

Profilen for PAYG-bundskatten i figur 8 afspejler den underliggende profil i den demografiske forsørgerbrøk, jf. figur 2. Da den demografiske forskydning frem til 2040 betyder en stigning i antallet af ældre pga. de store efterkrigs-generationer og et fald i arbejdsstyrken som følge af de små årgange fra 1980'erne, implicerer PAYG-finansiering, at de fremtidige små generationer – alt andet lige – kommer til at opleve en højere beskatning end nuværende store generationer. I perioden fra 2040 til 2060 er der en svag modsat bevægelse i generationernes størrelse, hvilket betyder, at PAYG-



Figur 8. Udviklingen i bundskattesatsen med PAYG-finansiering med skattetilpasning uden trend som følge af velstandsstigning.

Kilde: Egne beregninger på DREAM-modellen.

skattesatsen stabiliseres på trods af den fortsatte vækst i middelevetiden. Fra omkring 2060 fortsætter stigningen i PAYG-skatten som følge af middelevetidsstigningen.

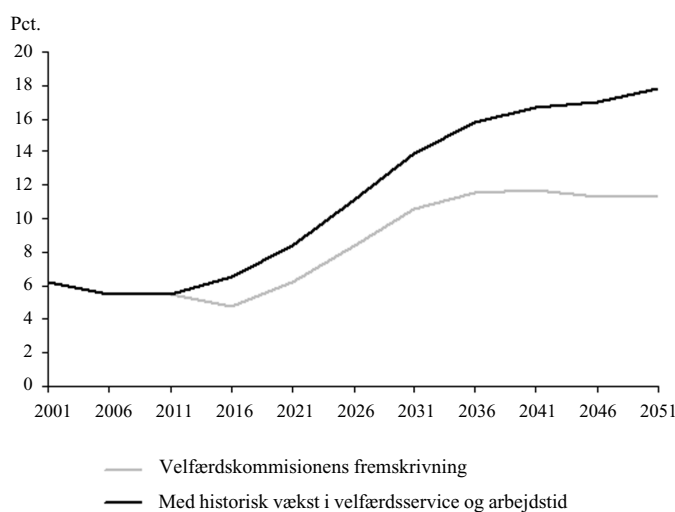
Årsagen til, at middelevetidsstigningerne påvirker PAYG-skatten, er, at den andel af livet, hvor den enkelte er i arbejdsstyrken, gradvist reduceres under antagelse af givne regler for tilbagetrækning. Det fører til, at forholdet mellem potentielt forsørgede og potentielle forsørgere er stigende over tid. Da PAYG-finansieringen og den sociale kontrakt i det danske velfærdssystem indebærer, at de, der aktuelt er beskæftigede, skal finansiere de offentlige udgifter, må den enkelte beskæftigedes skattebetaling stige over tid.

### 5.1 PAYG-finansiering og velstandsdilemmaet

Såfremt der ud over de demografiske forskydninger også tages hensyn til velstandsdilemmaet, jf. afsnit 3.1, bliver den nødvendige skattestigning under PAYG-finansiering selvsagt større.

Konsekvenserne for bundskatten af at sikre finansieringen til at håndtere velstandsdilemmaet (mere service og fritid) fremgår af figur 9. I 2021 er bundskattesatsen steget til over 8 procentpoint, hvilket er 2 procentpoint mere end i grundforløbet. I 2051 er forskellen i bundskattesatsen mellem de to forløb vokset til mere end 6 procentpoint.

Konklusionen er således, at udviklingen i den offentlige sektors finansieringsbehov er meget følsom over for Velfærdscommissionens forudsætning om, at den historiske



Figur 9. Udviklingen i bundskattesatsen ved PAYG-finansiering i forløbet med og uden trendmæssig udvikling i velfærdsservice og arbejdstid.

Kilde: Egne beregninger på DREAM-modellen.

trend i udgifterne til velfærdsservice og i arbejdstiden kan brydes. Hvis dette ikke er tilfældet, opstår behovet for at øge bundskattesatsen tidligere, og den nødvendige stigningstakt i skattesatsen bliver højere.

## 6. Holdbar skattesats og finanspolitisk holdbarhed

PAYG-finansiering sikrer på periode for periode-basis, at der er indtægtsmæssigt grundlag for de udgifter, velfærdsordninger afstedkommer. I det følgende afsnit diskuteres det modsatte yderpunkt, hvor kravet er, at den offentlige sektor overholder sin intertemporale budgetbetingelse. Herved kan man udnytte kapitalmarkederne til at overføre underskud eller overskud mellem perioderne. Kravet er, at nutidsværdien af forventede fremtidige udgifter er dækket af nutidsværdien af forventede fremtidige indtægter plus eventuel nutidig formue. Den intertemporale budgetrestriktion for den offentlige sektor kan i summarisk form skrives som

$$PV(B(t)) + D(0) \geq 0$$

hvor  $PV(B(t))$  angiver nutidsværdien ved givne rentesatser af primære budgetsaldi fra nu af og frem i tid, mens  $D(0)$  er den initiale finansielle gæld.

Tilgangsvinklen er således fremadskuende med udgangspunkt i den offentlige sek-

tors intertemporale budgetrestriktion.<sup>16</sup> Finanspolitisk holdbarhed er en markedstest, hvor det vurderes, om den udvikling i offentlige udgifter og indtægter, der følger af den fastlagte politik og den økonomiske udvikling, er forenelig med den intertemporale budgetrestriktion og dermed solvens. At vurdere den finanspolitiske holdbarhed er således en naturlig del af den mellemfristede økonomiske politik og kan medvirke til at undgå pludselige/store ændringer i den økonomiske politik.

Som et mål for den nødvendige tilpasning defineres den holdbare (bund)skattesats som den konstante værdi af (bund)skattesatsen, der netop sikrer, at den intertemporale budgetbetingelse er opfyldt. Ved beregning af den holdbare skattesats vurderes, på hvilket niveau bundskattesatsen permanent skal ligge for at sikre, at der er indtægtsmæssig dækning af en given udgiftsprofil skabt af i øvrigt uændrede velfærdsordninger og skatte- og afgiftssatser.

Hvis den således beregnede holdbare bundskattesats er højere end satsen i udgangspunktet, er finanspolitikken ikke holdbar. Ved fravær af finanspolitisk holdbarhed kan man af beregningerne udlede, at en tilpasning nødvendigvis må forekomme på et tidspunkt, dvs. at forløbet for offentlige udgifter og indtægter ikke kan realiseres. Beregningerne siger imidlertid intet om, hvornår tilpasningen senest må komme, eller at den nødvendigvis skal ske ved f.eks. en skattestigning. Holdbarhedsberegningen er således et diagnoseværktøj, der vedrører mulighedsområdet for den økonomiske politik ved konkret at vurdere, om et planlagt politikforløb kan realiseres i den forstand, at det er konsistent med solvens, dvs. om det implicerede udgifts- og indtægtsforløb kan realiseres. Når der tages udgangspunkt i »uændrede regler«, er formålet at vurdere, om der på et tidspunkt vil være finansielle grunde til at ændre disse. Beregningens styrke er at sammenfatte tilpasningsbehovet i ét tal, dvs. dette skal tolkes som en indikator for problemets størrelsesorden.

Hvorvidt et givet forløb med en holdbar økonomisk politik er retfærdigt eller i overensstemmelse med økonomisk politiske præferencer, er et andet spørgsmål, som bør vurderes uafhængigt af holdbarhedsindikatoren. Det er således vigtigt at sondre mellem på den ene side, om et givet forløb er realiserbart, dvs. opfylder finanspolitisk holdbarhed, og på den anden side den optimale politik (der pr. konstruktion opfylder kravet om finanspolitisk holdbarhed), jf. afsnit 8.

### *6.1 Holdbarhedsberegning*

I det følgende præsenteres hovedresultaterne af holdbarhedsberegninger ved kravene

---

16. Da spørgsmålet om finanspolitisk holdbarhed er knyttet til fremtidige ændringer i befolkningens sammensætning, er det nødvendigt at benytte et fremadskuende perspektiv. Den traditionelle tilgangsvinkel til spørgsmålet om finanspolitisk holdbarhed er enten baseret på historiske data eller den underliggende dynamik indbygget i det offentlige budget under stiliserede antagelser om offentlig budgetsaldo eller ønsket gælds niveau. Se Burnside (2004) for en introduktion.



til en bundskatteændring. En mere specifik redegørelse for forløbene ud over de umiddelbare konsekvenser for de offentlige finanser findes i Velfærdskommissionen (2004).

Et forløb med »uændrede regler« er ikke foreneligt med finanspolitisk holdbarhed. Den nødvendige stigning i bundskattesatsen er på 8,7 procentpoint, dvs., for at sikre finanspolitisk holdbarhed er en betydelig stigning i bundskattesatsen nødvendig.

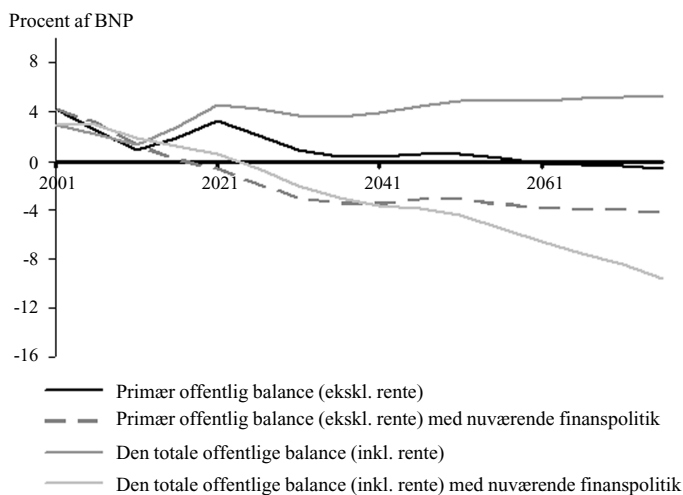
Finanspolitisk holdbarhed opnås således ved permanent at øge de offentlige indtægters andel af BNP. Stigningen i skatteindtægterne, der rent teknisk er gennemført som en gradvis stigning i perioden fra 2011 til 2021, fører til en stigning i den primære offentlige saldo i denne periode. Fra 2021 svarer udviklingen til forløbet uden finansiering, men på et permanent højere niveau. Der er således tale om en »parallelforskydning« af den primære saldo, mens den aftagende trend fastholdes. Parallelforskydningen er netop så stor, at den formue, der akkumuleres, kan finansiere de fremtidige primære underskud, der følger af den aftagende trend. I beregningerne svarer dette til, at den primære saldo forbliver positiv til omkring 2061. Som en konsekvens heraf udviser det samlede offentlige budget overskud i hele perioden, jf. figur 10. Det er en konsekvens af denne beregning, at forholdet mellem privat og offentligt forbrug forskydes i retning af en forøget andel af offentligt forbrug.

Da det underliggende problem er en gradvis ændring i befolkningens alderssammensætning, som fører til et permanent højere forsørgerforhold, jf. figur 2, indebærer forløbet med en »holdbar skat« – som følge af »parallelforskydningen« af den primære saldo – en betydelig konsolidering af de offentlige finanser, inden de demografiske forandringer for alvor slår igennem. Skattestigningen sikrer en gradvis opbygning af en betragtelig offentlig formue (inkl. ATP),<sup>17</sup> jf. figur 10.<sup>18</sup> Det fremgår af figuren, at formuen er på mere end 100 procent af BNP i 2061. I tilfældet uden finansiering vil gælden på dette tidspunkt være på knap 100 procent af BNP med hastigt stigende gælds niveau. Pr. definition har PAYG-forløbet en uændret offentlig formue fra 2011, svarende til 17 procent af BNP, og ligger således midt mellem de to øvrige forløb.

---

17. Det anvendte formuebegreb er nationalregnskabet akkumulerede nettofordringserhvervelse (inklusive ATP). På grund af den veldefinerede sammenhæng med nationalregnskabet er dette det relevante mål til analytiske formål. Det er ikke sammenfaldende med ØMU-gældsbebegret, (der er et bruttogældsbebegreb) som anvendes i Vækst og Stabilitetspagten, og som også ligger til grund for gældsmålsætningen i regeringens 2010 plan.

18. En konsekvens heraf er en ganske betydelig nettofordringserhvervelse på udlandet, idet »udenlandske tilgodehavender« vil overstige 170 procent af BNP i 2061. Dette implicerer samtidig en ganske betydelig omstilling for at realisere de tilsvarende store overskud på handelsbalancen. Størrelsen af disse forandringer er så stor, at man kan sætte spørgsmålstegn ved, om stødet har en størrelse der gør, at den sædvanlige antagelse om approksimativt uændrede strukturer kan opretholdes. Dette problem er især vigtigt, såfremt man økonomisk politisk ønsker at følge en strategi, der indebærer en så kraftig konsolidering af den offentlige sektor.



Figur 10. Primær og samlet budgetsaldo med hhv. uændrede regler og holdbar finanspolitik finansieret ved en permanent skattestigning.

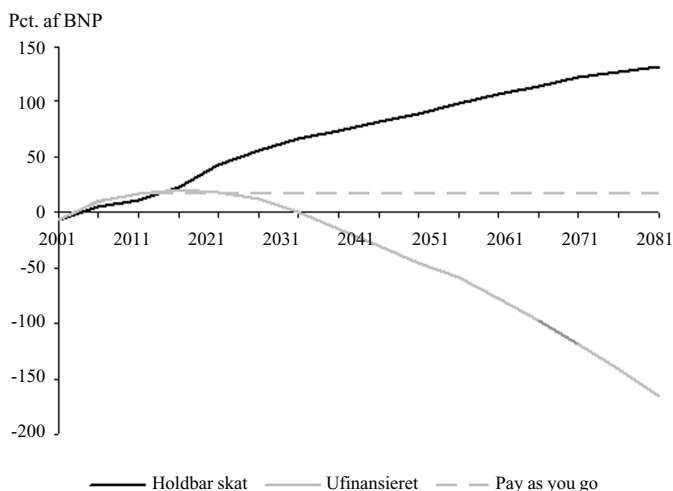
Kilde: Egne beregninger på DREAM-modellen.

## 6.2. Velstandsdilemmaer

Betydningen af stigende velstand for finansieringskravene til velfærdssamfundet kan også vurderes ved ændringer i den holdbare skattesats. Ved en underliggende trendvækst i offentlig individuel service på 0,15 procentpoint om året øges den nødvendige skattestigning fra 8,7 til 15,1 procentpoint. Tilsvarende vil en trend med 0,15 procentpoint årlig reduktion i arbejdstiden kræve en skattestigning til 15,4. Inddrages begge dimensioner af velstandsdilemmaet samtidigt, bliver kravet en permanent stigning i bundskattesatsen på 21,9 procentpoint. Der er tale om ganske betragtelige stigninger for begge dimensioner af velstandsdilemmaet – mere service og mere fritid. Bemærk, at vurdering af størrelsesordenen er baseret på, at velstandsdilemmaet her er indregnet som en vedvarende trend frem til år 2100.

## 7. Sammenligning af PAYG og holdbar skat

I det forudgående har beregningerne af den holdbare skat alene været betragtet som en indikator for den samlede størrelsesorden af tilpasningsproblemet udtrykt ved et enkelt tal. Holdbarhed opnået ved permanente ændringer i beskatning (eller tilsvarende ændringer i velfærdsordninger) kan imidlertid også betragtes som en økonomisk politisk strategi, der har til formål at sikre, at velfærds- og skatteordninger efter engangsændringen kan fastholdes uforandrede over tid. Grundet den underliggende profil i de demografiske forandringer bliver dette en opsparingsstrategi, jf. nedenfor.



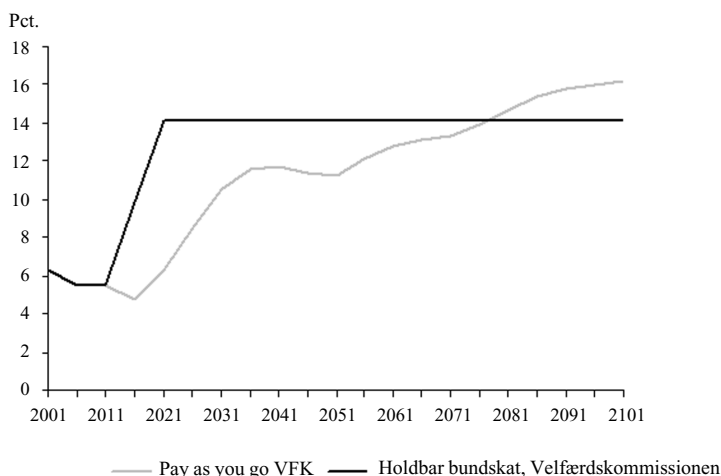
Figur 11. Offentlig formue (inkl. ATP) i procent af BNP med alternative finansieringsantagelser.

Kilde: Egne beregninger på DREAM-modellen.

En økonomisk politisk strategi, som er baseret på, at holdbarhed opnås ved en permanent ændring, kan opfattes som det modsatte yderpunkt i forhold til en PAYG-strategi, hvor holdbarhed sikres ved løbende at tilpasse udgifter og indtægter for at holde en fast gæld i procent af BNP.

Ved henholdsvis PAYG-finansiering og holdbar skat er der i afsnit 5 og 6 analyseret to måder til sikring af de finansielle grundlag for eksisterende velfærdsordninger. Figur 12 viser udviklingen i bundskattesatsen under de to tilgangsvinkler til finansiering. Det fremgår, at den holdbare skat i den første del af fremskrivningen overstiger skatten under PAYG-finansiering, og omvendt i den senere del. Man kan løst tolke den holdbare skat som et vejet gennemsnit (via nutidsværdiberegning) af skattesatsen under PAYG-finansiering. Figuren viser tydeligt, at den holdbare skat implicerer skatteudglatning, mens PAYG-finansiering implicerer en varierende skat over tid med en opadgående trend. Finansieringen med holdbar skat indebærer, at skattebetalingerne frem til 2080 er højere end med PAYG-finansieringen.

Betragtes fordelingen af skattebetalingen over generationer, betyder finansieringen med holdbar skat, at alle generationer født før 2030 vil få højere nettobetaling til den offentlige sektor over livet sammenlignet med PAYG-finansiering. Omvendt vil generationer født efter 2030 have lavere nettobetaling til den offentlige sektor i situationen med holdbar skat end i situationen med PAYG-finansieringen, jf. figur 13.



Figur 12. Bundskattesatsen ved PAYG-finansiering og ved holdbar finanspolitik finansieret ved permanent skattestigning.

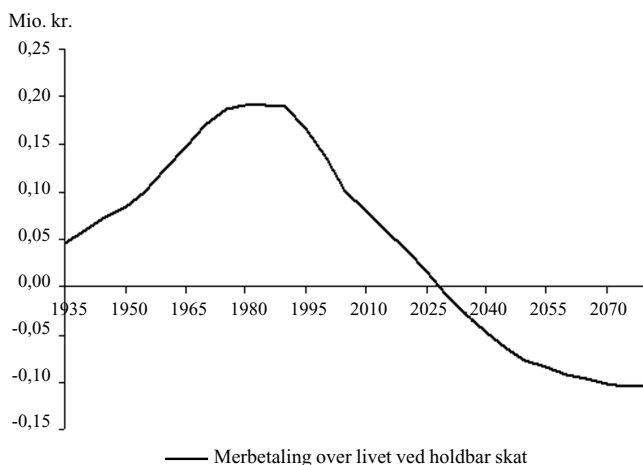
Kilde: Egne beregninger på DREAM-modellen.

Generationerne født mellem 1970 og 2000 oplever den højeste forskel i nettobetalingerne til det offentlige over livet. For disse generationer svarer forskellen til en diskonteret værdi på knap 0,2 mio. kr. (i »lønfaste« priser) pr. person over livet. Når forskellen mellem de to forløb er mindre for generationer, der i dag er ældre, skyldes det, at disse generationer udsættes for forskellen i beskatning i en kortere del af deres liv. Gevinsten ved finansieringen med holdbar skat i forhold til PAYG-finansiering for personer født efter 2060 svarer til en diskonteret værdi på omkring 0,1 mio. kr. pr. person over livet.

Fordelingen af gevinster og tab i forhold til den offentlige sektor kan tilnærmelsesvist anvendes som mål for den relative nytteændring for den enkelte generation. Det skyldes, at de afledte virkninger på priser og lønninger af skatteforskellene ved de to finansieringsformer er begrænsede målt i forhold til den direkte virkning af skatteforskellene på agenternes budgetrestriktion.<sup>19</sup> Grundlæggende betyder finansieringen med holdbar skat derfor, at de nulevende og de umiddelbart kommende generationer stilles værre end i en situation med en PAYG-finansiering. Og som konsekvens heraf stilles fremtidige generationer bedre.

Denne entydige fordeling af gevinster og tab over tid skyldes, at PAYG-finansieringen på grund af den stigende levealder fører til en permanent stigende bundskattesats.

19. Se Jensen, Nødgaard og Pedersen (2001) for en sammenligning af mål baseret på generationsregnskab og ækvivalent variation.



Figur 13. Merbetaling til den offentlige sektor over livet fordelt på fødselsårgange. Finansiering med holdbar skat målt i forhold til PAYG-finansiering (mio. kr. pr. person).

Kilde: Egne beregninger på DREAM-modellen.

Dette understreger, at valget af finansieringsstrategi kan få afgørende betydning for den intergenerationelle fordeling.

### 8. Følsomhedsberegninger og usikkerhed

Beregningerne ovenfor er baseret på et grundforløb, som fremkommer fra den underliggende demografiske fremskrivning og tolkningen af uændrede regler, jf afsnit 3. På en række områder er der usikkerhed omkring den fremtidige udvikling, og det er derfor af stor betydning at afklare beregningernes følsomhed over for alternative antagelser. Velfærdscommissionen (2004) præsenterede en række følsomhedsberegninger, og den efterfølgende diskussion om fremskrivningen har i høj grad været centreret om disse følsomhedsberegninger. Beregningerne nedenfor tjener således det dobbelte formål at vise betydningen af forskellige forudsætninger samt at pege på den usikkerhed, der er knyttet til vurderingen af finansieringsproblemetets størrelse.

Tabel 1 giver en oversigt over nogle af de væsentlige følsomhedsberegninger. De er alle alene repræsenteret ved den nødvendige ændring i den holdbare skat (bundskattesatsen), men på lige fod med de viste beregninger ovenfor kan de alle også repræsenteres ved den nødvendige permanente besparelse på de offentlige udgifter eller den permanente stigning i privat beskæftigelse samt ændringer under PAYG-finansiering. For en nærmere diskussion af de enkelte forudsætninger se Velfærdscommissionen (2004, 2005d).

*Tabel 1. Følsomhedsberegninger – ændringer i nødvendig stigning i bundskattesats under alternative forudsætninger.*

	Skattestigning i pct. point
<i>Grundforløb</i>	8,7
<i>Demografi</i>	
Konstant middellevetid	1,9
DREAMs middelevetidsforudsætning <sup>a</sup>	5,8
FNs middelevetidsvækst	16,0
Ekstra fødsler, fertilitet stiger fra 1,7 til 1,8	8,9
5000 ekstra indvandrere fra mere udviklede lande p.a.	8,5
5000 ekstra indvandrere fra mindre udviklede lande p.a.	10,7
5000 ekstra udvandrere, der vender tilbage svarende til historisk adfærd p.a.	9,0
5000 ekstra udvandrere, der ikke vender tilbage p.a.	10,1
<i>Velfærdsservice og fritid</i>	
En ekstra ferieuge	9,9
0,15 pct. arbejdstidsreduktion pr. år	15,1
0,15 pct. mervækst i individuel service	15,4
<i>Arbejdsmarked</i>	
1 pct. point højere ledighed	9,9
Højere erhvervsdeltagelse for efterkommere fra mindre udviklede lande	7,1
<i>Rente og vækst</i>	
1 pct. point højere rente	5,0
0,5 pct. point ekstra BFI vækst	12,8
<i>Sundhed</i>	
Restlevetid- og aldersafhængighed af sundheds-, sociale og plejeydelser	7,6
Både restlevetids- og aldersafhængighed af sundheds-, sociale og plejeydelser og udgiftsvækst på sundhedsydelser på 0,5 pct. point mere end produktivitet	22,1

*Note:* (a) Forudsætning anvendes af Finansministeriet (2004) og DØR (2004).

*Anm.:* Alle beregninger på nær sidste række i tabellen har tidligere været offentliggjort i enten Velfærdskommissionen (2004) eller Velfærdskommissionen (2005a). Beregningerne af effekten af uændret middellevetid er ændret i forhold til det tidligere offentliggjorte niveau på 1,6. Middellevetiden var teknisk blevet fastholdt på et forkert niveau.

*Kilde:* Velfærdskommissionen (2004, 2005).

Ved læsning af tabellen kan der være grund til at påpege, at tabellens følsomhedsberegninger ikke kan lægges sammen vilkårligt. Der vil generelt være meget store samspilseffekter. I Velfærdskommissionen (2004) side 365 er der således set på to forskelle til Det Økonomiske Råd (DØRs) antagelser. Der tages udgangspunkt i DØRs antagelser, og analysen vedrører den samlede effekt af Velfærdskommissionens forudsætning om større vækst i middellevetiden og højere økonomisk vækst (som er 1/2 procent højere). Effekten af at indføre begge forudsætninger i DØRs analyse er en forværring af holdbarhedsproblemet med 1,92 procent af BNP målt i forhold til DØRs

forudsætninger. Hvis kun middellevetiden ændres, er forskellen 0,65 procent af BNP, og hvis kun væksten ændres, er forskellen 0,56 procent af BNP. Knap halvdelen af den samlede effekt er således en samspilseffekt. Dette samspil opstår via den lavere diskonteringsrente pga. den lavere vækstkorrigerede realrente og det faktum, at der er en gradvis men permanent stigning i middellevetiden.

Tabellen understreger den store betydning både ændringer i levetid og velstandsdilemmaet (mere service, mindre fritid) har for de offentlige finanser, jf. diskussionen ovenfor. Det bemærkes, at betydningen af forudsætninger omkring rente og vækst primært skal ses i relation til, at tabellen viser finansieringsproblemet målt ved den holdbare skat. For en vurdering af den underliggende profil i den offentlige sektors primære saldo har disse forudsætninger kun en meget beskeden betydning (se Velfærdskommissionen (2005d)). Dette understreger betydningen af rente- og vækstforudsætninger såfremt man vælger en opsparingsstrategi, og derfor også det hensigtsmæssige i at vælge en sådan strategi, jf. nedenfor.

### **9. Diskussion af den optimale økonomisk politiske strategi**

Den forudgående analyse af to yderpunkter i valget af strategier danner baggrund for at diskutere den mulige økonomisk politiske reaktion. På den ene side vil en PAYG-finansiering sikre finanspolitisk holdbarhed ved på år til år-basis at justere beskatningen i forhold til udgiftsudviklingen (stabilisering af den offentlige gæld, men variationer i velfærdsordninger/skatter). På den anden side vil en politik i overensstemmelse med kravene til holdbar finanspolitik være baseret på et langsigtet perspektiv og dermed permanente ændringer i beskatningen (variationer i offentlig gæld/formue, men stabile ordninger/beskatning). Da den underliggende demografiske udvikling betyder en gradvis stigning til en permanent højere forsørgerbrøk, implicerer denne strategi nødvendigvis en initial stor konsolidering, jf. ovenfor om opsparingsstrategi.

Beregningerne er ovenfor vist for ændringer i bundskattesatsen, men tilpasningerne kan selvfølgelig for begge strategier foretages ved justeringer ikke alene i beskatningen men også i velfærdsordninger i bred forstand. Kravet er, at disse reformer skaber et provenu ækvivalerende de præsenterede ændringer i bundskattesatsen.

Disse to yderpunkter understreger, at den afgørende økonomiske beslutning er en afvejning mellem successive tilpasninger af velfærdsordningerne og/eller disses finansiering i forhold til en strategi tilstræbende ensartede velfærdsordninger og finansiering på tværs af fremtidige generationer baseret på permanente ændringer.

Udgangspunktet for en diskussion af valg af strategi er at adressere udfordringer skabt af de demografiske forandringer (primært drevet af øget middellevetid) og velstandsdilemmaet skabt af fortsat øget materiel velstand. Som påpeget ovenfor giver

begge disse forhold anledning til en »negativ« trend i de offentlige finanser. Da disse forhold udvikler sig over tid, vedrører en væsentlig del af diskussionen spørgsmål knyttet til intergenerationel fordeling. Samtidig er der usikkerhed knyttet til den fremtidige udvikling og dermed også et spørgsmål om risikodeling og tilpasning til denne risiko. Som udgangspunkt for en diskussion af fastlæggelse af en økonomisk politisk strategi diskuteres først principielle aspekter knyttet til intergenerationel fordeling og risikospredning.

### *9.1 Intergenerationel fordeling*

Spørgsmålet om den intergenerationelle fordeling vedrører generelt de muligheder, nulevende generationer efterlader til fremtidige generationer. Det gælder ikke alene i relation til offentlige finanser, men videre til spørgsmål om infrastruktur, human kapital og natur- og miljøressourcer, jf. f.eks. Velfærdskommissionen (2004) kap. 4.

Det falder uden for denne artikels rammer at give en udtømmende diskussion af disse forhold, og det følgende er afgrænset til en vurdering af de intergenerationelle implikationer af velfærdsordninger og deres finansiering.

For at vurdere den intergenerationelle fordeling er det nødvendigt at afklare såvel mulighedsområdet som de politiske præferencer/ønsker knyttet til den intergenerationelle fordeling. Førstnævnte er reelt sammenfattet i den intertemporale budgetrestriktion for den offentlige sektor og er derfor en del af den forudgående analyse. Den intertemporale budgetrestriktion angiver mulighederne for at overflytte ressourcer over tid og dermed mellem generationer givet markedsforhold (rente) osv.

Det er forbundet med betydeligt større vanskeligheder at fastlægge de politiske præferencer knyttet til den intergenerationelle fordeling, også selvom mange økonomisk politiske beslutninger implicit afspejler intergenerationelle afvejninger.

En traditionel approach til at indkredse forhold af betydning for tilrettelæggelsen af den økonomiske politik er at formulere en samfundsvelfærdsfunktion, som respekterer de enkelte generationers nyttefunktioner, og som sammenvejer disse med en subjektiv tidspræferencerate afspejlende den politiske vægtning af nytten for forskellige generationer.<sup>20</sup> Selvom denne tilgangsvinkel ikke er uproblematisk,<sup>21</sup> er den nyttig til at identificere nogle af de centrale forhold af betydning for den politiske beslutningsproces.

---

20. Dvs. den subjektive tidspræferencerate for en given generation kan afvige fra tidspræferenceraten mellem generationer, jf. f.eks. Blanchard og Fischer (1989).

21. Det traditionelle approach har givet anledning til en diskussion om intergenerationel rimelighed og retfærdighed (se f.eks. Møller (2004) for en introduktion og referencer). Den traditionelle tilgangsvinkel til en formulering af en samfundsvelfærdsfunktion, jf. ovenfor, kan kritiseres for at forudsætte samme præferencerelationer på tværs af generationer, ligesom sammenvægtningen af forskellige generationers nytte ud fra et diskonteringsprincip er problematisk.



Den optimale politik maksimerer samfundsvelfærdsfunktionen under bibetingelse af den intertemporale budgetrestriktion.<sup>22</sup> Herved fastlægges en profil for forbrugsmuligheder osv. for nulevende og fremtidige generationer. Denne tilgangsvinkel giver en række interessante resultater.

Et centralt resultat er, at det ikke vil være optimalt at følge en PAYG-politik, da den optimale løsning generelt vil implicere, at de internationale kapitalmarkeder udnyttes til at omallokere over tid/generationer. Endvidere kan der være et efficienstab ved en PAYG-strategi, som følge af udsving over tid i forvridende skatter, jf. Barro (1979).

Et andet vigtigt resultat er, at den optimale udvikling i forbrugsmuligheder m.m. over tid og dermed generationer afhænger af forholdet mellem den objektive og subjektive diskonteringsrate. Ved en objektiv diskonteringsrate forstås de markeds-mæssige muligheder for at overflytte ressourcer over tid, dvs. markedsrenten. Den subjektive diskonteringsrate afspejler præferencerne for nutidigt i forhold til fremtidigt forbrug. Er den subjektive diskonteringsrate høj, tillægges fremtidige forhold (og dermed forholdene for fremtidige generationer) ikke så stor vægt og omvendt. Et argument om udelukkende at betragte en given horisont svarer derfor til asymmetrisk vægtning af forholdene i perioden før (vægt=1) og efter (vægt=0) det givne tidspunkt. Sammenfattende kan man sige, at den subjektive tidspræferencerate påvirker objektfunktionen, mens den objektive/markeds-mæssige fastlægger budgetrestriktionen.

Den optimale forbrugsprofil over tid og dermed på tværs af generationer bestemmes af forholdet mellem den objektive og subjektive diskonteringsrate<sup>23</sup> (Ramsey-reglen). Er den objektive diskonteringsrate lig den subjektive tilsiger den optimale politik en konstant forbrugsprofil, dvs. nulevende og fremtidige generationer skal have samme forbrug. Såfremt den objektive diskonteringsrate er mindre (større) end den subjektive, er den optimale forbrugsprofil aftagende (stigende) over tid, dvs. denne situation vil tendere til, at en optimal forbrugsprofil medfører højere (lavere) forbrug til nulevende generationer i forhold til fremtidige generationer.

Det følger af ovenstående, at et rentefald alt andet lige vil trække i retning af en ændret forbrugsprofil til fordel for nulevende generationer, dvs. mindre opsparing. Tilsvarende vil en høj subjektiv/politisk diskonteringsrate betyde mindre vægtlægning på vilkårene for fremtidige generationer og dermed trække i retning af mindre behov for konsolidering.

---

22. Et eksempel er Flodén (2003), hvor den subjektive tidspræference er lig husholdningernes tidspræference (husholdninger er modelleret som dynastier). Det vises, at den optimale politik implicerer en permanent skattestigning og en initial konsolidering. Den optimale skat er approksimativt konstant over tid. Modellen løses på globalt plan, dvs. der tages højde for påvirkningen af det internationale renteniveau i tilpasningsforløbet. For Danmark vurderes det, at en permanent skattestigning på 6 % i beskatningen af arbejdsindkomst vil være optimal.

23. Denne vil afhænge af såvel den subjektive tidspræferencerate for en given generation og den politiske tidspræferencerate mellem generationer, jf. Andersen (2005).

Såfremt forskellen mellem den reale markedsrente og den subjektive tidspræferencerate svarer til vækstraten vil den optimale forbrugsprofil indebære en vækst svarende til produktivitetsvæksten. I denne situation får den enkelte generation et forbrug (målt relativt til andre generationer), der svarer til generationens relative produktivitet. Det er også en implikation, at en forventning om øget fremtidig vækst for konstant markedsrente vil føre til et ønske om at hæve forbruget for de nuværende generationer, idet den optimale profil i forbruget over generationer er uændret. Forventning om højere fremtidig vækst reducerer derfor den optimale opsparing som følge af et underliggende ønske om at flytte forbrugsmuligheder til nulevende generationer.<sup>24</sup>

Et særligt interessant spørgsmål er, hvorledes stigende levetid påvirker den intergenerationelle fordeling. Forudsat at længere levetid værdsættes, vil fremtidige generationer umiddelbart få en velfærdsgevinst, som nulevende generationer af gode grunde ikke kan få. Samtidig øges værdien af forbrugsmuligheder (man skal have noget af leve for), men også mulighederne for at være aktiv på arbejdsmarkedet (forudsat intakt arbejdsevne). Hvorledes skal den intergenerationelle fordeling tilpasses dette? Det er interessant at notere, at en samfundsplanlægger med neutral vægtning af generationer (objektiv og subjektiv diskontering er sammenfaldende) ved en underliggende opadgående trend i levealder vil tilstræbe, se Andersen (2005) (i) at udjævne forbrugsmulighederne, hvilket implicerer, at nulevende generationer i et vist omfang skal medvirke til at spare op til fremtidige generationer, og (ii) at tilbagetrækningsalderen (og dermed »pensionistperioden«) udvikler sig proportionalt til levealderen, og dermed kommer fremtidige generationer også i stort omfang til selv at sikre sig det øgede forsørgelsesgrundlag, som den længere levetid forudsætter.

Konsekvenserne af øget levetid kan også tolkes som en ændring i den subjektive tidspræferencerate, hvilket for den enkelte involverer spørgsmålet om afvejning mellem forbrugsmuligheder i dag og i fremtiden også overlevelsessandsynligheden (se f.eks. Blanchard and Fischer (1989) for en lærebogsfremstilling), dvs. sandsynligheden for at få glæde af den fremtidige forbrugsmulighed.<sup>25</sup> Den effektive subjektive diskontering for generationen svarer derfor til summen af den subjektive tidspræferencerate og den forventede dødsrisiko. Stigende middellevetid fører til, at den forventede dødsrisiko falder og således til en lavere effektiv subjektiv diskontering. Det betyder, at generationen

---

24. Se f.eks. Hansson og Stuart (1989) for en analyse af, hvorledes sociale sikringssystemer af PAYG-typen på tværs af generationer kan etableres og implikationerne heraf for fordeling af materiel velstand på tværs af generationer. Etableringen af et socialt sikringssystem med PAYG-elementer kan tolkes som en måde, hvorpå den første generation får andel i fremtidige generationers materielle velstand. Dette er særligt tydeligt i systemer med defined contributions, hvorved benefits bliver en funktion af de unge generationers indkomst.

25. Undtagelsen til dette forekommer, hvis generationen har altruistisk arvemotiv, hvor fremtidige generationer i »familiedynastiet« vægtes på samme måde som generationen selv. Dette er først vist i Barro (1974).

vil ønske en større stigning i forbruget over tid, hvilket alt andet lige trækker i retning af en forøget opsparing. Det er ikke oplagt, at øget middellevetid på samme måde skal påvirke den samfundsmæssige diskontering mellem generationer. Da generationernes effektive diskontering falder, følger det, at den samlede subjektive diskontering er faldende med stigende levetid, således at forventede levetidsstigninger – alt andet lige – fører til et ønske om at flytte forbrugsmuligheder til fremtidige generationer.

Sammenfattende kan det konkluderes, at den traditionelle approach til fastlæggelse af den optimale intergenerationelle fordeling ikke giver støtte til hverken PAYG-strategien eller den rene opsparingsstrategi. Førstnævnte udnytter ikke markedets muligheder for overflytning af ressourcer over tid og generationer. Den optimale politik kan give støtte til en vis konsolidering, dvs. opsparing af nulevende generationer, som følge af stigende levetid. Den understøtter imidlertid ikke en ren opsparingsstrategi, da det også er en implikation, at de generationer, der får glæden af øget levetid, må bidrage til finansiering heraf f.eks. ved senere tilbagetrækning.

Den afgørende implikation af ovenstående er, at den intergenerationelle fordelingsprofil fastlægges af den politiske vægtlægning på konsekvenserne for forskellige generationer. Dette er et fundamentalt politisk spørgsmål.

I forhold til disse principielle overvejelser kan man pragmatisk konstatere, at der ikke historisk har været meget store intergenerationelle udligninger via de offentlige finanser ud over den, der har været implikationen af en overvejende PAYG-baseret opbygning af velfærdsordninger.<sup>26</sup> Herunder er der ikke nogen tendens til, at hidtidige generationer systematisk har fremrykket forventede fremtidige produktivetsgevinster. Væksten i velfærdsordningerne skabt af produktivetsgevinsterne, er hidtil blevet tildelt generationer, der har modtaget ydelser i de perioder, produktivetsgevinsterne fremkom. Såfremt det er en præmis, at nulevende generationer ikke skal agere anderledes end tidligere generationer, giver dette et argument for ikke at foretage store omfordelinger mellem generationer. Dette kan tolkes på den måde, at den samfundsmæssige tidspræferencerate er relativ lav.

I forhold til denne normative approach kan indvendes, at det politiske system typisk agerer under en væsentlig kortere tidshorisont. Man kan dog ikke ubetinget sige, at der er en meget kortsigtet bias i den økonomiske politik, jf. f.eks. miljø- og ressourcedebatten. Forskellige antagelser kan gøres om horisonten for politiske beslutninger og vægtlægningen på forskellige generationers muligheder. Det afgørende for problematikken omkring den ændrede demografiske struktur er imidlertid at pege på, at konsekvenserne er langsigtede, og udskydelse af problemerne er ensbetydende med, at en

---

26. Man kan argumentere for, at den betydelige offentlige gæld, der blev opbygget fra midten af 1970erne til midten af 1980erne, udgør en markant undtagelse fra dette synspunkt. Politikken var dog formentlig et ikke vellykket forsøg på en udglatning af et fænomen, der blev vurderet til at være temporært.

større byrde og usikkerhed skubbes fremad i tid. Centrale forhold knyttet til velfærdssamfundets sociale kontrakt har i kraft af den overlappende generationsstruktur i sagens natur en lang horisont. I modsætning til andre politikområder, hvor f.eks. 2040 kan opfattes som en fjern fremtid, drejer det sig på velfærdsområdet om f.eks. de pensionsvilkår, 30-35-årige kan regne med vil gælde for dem.

### *9.2 Risikodeling*

Usikkerhed er et fundamentalt aspekt ved problemstillingen. Der er på en række punkter usikkerhed knyttet til den fremtidige udvikling. Usikkerheden kan ikke fjernes, men den kan i et vist omfang spredes over personer og tid/generationer. Alle strategier har implikationer for fordeling og spredning af risici. Det er derfor vigtigt at klarlægge, hvorledes forskellige strategier påvirker spredningen af risiko over personer og tid.

Usikkerheden kommer til udtryk ved følsomhedsberegningerne, jf. afsnit 7, der viser betydningen af ændringer i forskellige forhold i forhold til grundforløbet. Som et eksempel er der usikkerhed knyttet til vurderingen af fremtidige sundhedsudgifter. Der er således argumenter for, at de både kan blive større og mindre end antaget i grundforløbet. Såfremt de bliver mindre, vil finansieringsproblemet have været overvurderet, og omvendt. Konsekvenserne af de to typer af fejl er imidlertid ikke nødvendigvis symmetriske. Undervurderer vi i dag behovet og dermed udgifterne på sundhedsområdet, vil fremtidige generationer komme til at bære konsekvenserne, dvs. de vil være mere syge, og vi har efterladt dem med et større finansieringsproblem. Usikkerhed i fremskrivninger skal derfor ikke alene vurderes i forhold til implikationer for det finansielle holdbarhedsproblem, men også i forhold til hvilke konsekvenser disse risici vil have fremadrettet.

Indledningsvis er det vigtigt at notere, at det via det offentlige budget er muligt at etablere en vis risikospredning over tid og dermed generationer. Intergenerationel risikofordeling er sværere at foretage på markedsvilkår, da nulevende generationer ikke kontraktligt kan binde fremtidige generationer.<sup>27</sup> Via velfærdsordninger og det offentliges budget er der imidlertid intergenerationelle sammenhænge og dermed også muligheder for at etablere risikospredning på tværs af generationer (også med ikke overlappende livsperioder). Som velkendt er det muligt at diversificere midlertidige, men ikke permanente stød.

Den betydning, forskellige strategier har for risikoprofilen, kommer tydeligt frem ved at sammenligne PAYG- og opsparingsstrategien.

---

27. Intergenerationel risikofordeling kan via private markeder foretages mellem overlappende generationer levende på samme tidspunkt. Ligeledes sker der intergenerationel risikofordeling via arv m.m.

En ren PAYG-strategi fordrer løbende balance i de offentlige budgetter og dermed en løbende tilpasning af beskatning og/eller ydelser. Om tilpasningen sker via ændringer i beskatning eller ydelser har stor betydning for, om risici bæres af de yngre eller ældre generationer. Optimal risikodeling på et givet tidspunkt kræver løst formuleret, at marginalnyttens af indkomst er den samme for alle generationer, der lever på samme tid, dvs. det vil generelt være optimalt at lade både »beskatning og ydelser variere«, jf. Bohn (2001). Med denne strategi kan man sige, at makroøkonomiske risici umiddelbart transformeres til individuel risiko i form af løbende ændringer i velfærdsordningerne og beskatning. Dette gælder, uanset om ændringerne har midlertidig eller permanent karakter, da der pr. definition ikke er noget fremadskuende element i tilpasningen. Risikodeling er alene afgrænset til generationer i live på et givet tidspunkt. Den rene PAYG-løsning har derfor den ulempe, at den ikke udnytter mulighederne for risikospredning, og samtidig kan den betyde (tilfældig) omfordeling mellem generationer.

Opsparingsstrategien baseret på permanente ændringer i ydelser eller beskatning tilstræber uændrede ordninger på tværs af tid og dermed generationer, men bliver samtidig en opsparingsstrategi, da stigningen i middellevetiden betyder en gradvis stigning til en permanent højere forsørgerkvote. En økonomisk politik baseret på ensartede regler og ordninger over tid mindsker den individuelle risiko, idet tilpasningen tilstræbes fordelt over tid og dermed alle generationer. Denne strategi er fremadskuende og betyder en udligning over tid af både midlertidige og permanente ændringer. I forhold til de midlertidige ændringer betyder dette, at mulighederne for risikospredning udnyttes, mens det i forhold til permanente ændringer kan betyde en væsentlig omfordeling, jf. afsnit 7. Strategiens ulempe er, at forsikring og fordeling ikke kan adskilles, og derfor implicerer denne løsning en potentiel stor omfordeling til fremtidige generationer med længere levetid.

### *9.3 PAYG eller opsparing?*

I det forudgående er to yderpunkter sammenlignet, nemlig den løbende finansiering via en såkaldt PAYG-finansiering og en opsparingsstrategi med en permanent tilpasning for at sikre velfærdsordningernes finansielle grundlag. Begge disse to yderpunkter lider af alvorlige problemer både principielt og i forhold til implementering.

PAYG-modellen udnytter ikke alle muligheder for overflytning af forbrugsmuligheder via kapitalmarkederne. Dermed udnyttes mulighederne for risikodeling heller ikke, hvilket kan betyde, at risiko transformeres til stor individuel risiko omkring fremtidige vilkår. PAYG-strategien er imidlertid mindre informationsmæssigt krævende, da den er baseret på en »periode pr. periode«-betragtning.

Vælges opsparingsstrategien, implicerer det en permanent tilpasning til de underliggende forandringer. Tilgangsvinklen er fremadskuende, og dermed er udgangspunk-

tet at tage højde for fremtidige ændringer. Grundet trenden i demografien implicerer denne et stort konsolideringsbehov – en opsparingsstrategi – som vist ovenfor. Det skyldes, at man med den økonomiske politik søger at »udglatte« et problem med en væsentlig permanent komponent.<sup>28</sup>

Ud over de fordelingsmæssige implikationer diskuteret ovenfor, er der en række øvrige spørgsmålstejn ved denne opsparingsstrategi.

For det første afhænger størrelsen af den nødvendige opsparing af den underliggende fremskrivning. For at fastlægge den nødvendige opsparing skal der tages stilling til, hvilken fremskrivning eller prognose, der skal lægges til grund for den økonomiske politik. Over tid vil ny information fremkomme, og fremskrivningerne ændres, og kravet til den nødvendige konsolidering vil derfor løbende blive justeret. Ændringer i befolkningsprognoserne for Danmark inden for en kort årrække er et godt eksempel på, at vurderingen af den fremtidige udvikling kan ændres markant på kort tid, jf. afsnit 2. Såfremt niveauet for velfærdsordninger og finansieringsbidrag løbende fastlægges ud fra de seneste fremskrivninger, vil der være risiko for hyppige ændringer i regler og vilkår,<sup>29</sup> hvilket går imod udgangspunktet om at sikre stabile vilkår. Sådanne ændringer kan undgås via »precautionary savings«, dvs. en øgning af konsolideringskravet ud over niveauet krævet i forhold til grundforløbet. Med en sådan stødpude vil det være muligt at absorbere udsving af en vis størrelse uden at ændre i velfærdsordningerne. Desto mere usikkerhed eller modvilje mod at ændre velfærdsordningerne, desto større er kravet til denne »ekstra« konsolidering.<sup>30</sup>

For det andet – og modsatrettet – er der en »option value« ved at udsætte den nødvendige stramning i finanspolitikken. Argumentet er en analogi til teorien om irreversible investeringer under usikkerhed, jf. Dixit & Pindyck (1994). Kombinationen af irreversibilitet og usikkerhed betyder, at det forventede afkast skal overstige de forventede omkostninger, for at det er optimalt at gennemføre investeringen. I analogien knytter usikkerheden sig til størrelsen af den fremtidige middellevetidsvækst, og irreversibiliteten fremkommer ved, at nuværende generationer ikke kan kompenseres, når usikkerheden er afsløret, fordi de vil være afdøde. Den forventede gevinst (for fremtidige generationer) ved at udjævne finansieringen og fastholde »uændrede regler« skal derfor ses i forhold til omkostningerne (for nuværende generationer). Det taler for

---

28. I forhold til temporære variationer vil mulighederne for en sådan udglatning være meget bedre.

29. Ved midlertidige stød vil ændringerne være minimale, da konsekvenserne spredes over tid. Ved permanente ændringer (også forventede fremtidige) vil der være store ændringer, da konsekvenserne bæres af alle. Dette efterlader imidlertid et fundamentalt informationsproblem: Med hvilken sikkerhed ved vi, om en given ændring er midlertidig eller permanent?

30. Bemærk dog, at den teoretiske litteratur har påvist, at der ikke entydigt kan argumenteres for, at usikkerhed stiller krav om »precautionary savings«.

at vælge en strategi, hvor den økonomiske politik er fastlagt ud fra langsigtede trends med justering fra tid til anden i takt med at yderligere information akkumuleres.<sup>31</sup>

Endelig er der et fundamentalt tids-inkonsistens problem knyttet til begge strategier. PAYG-strategien er baseret på en løbende justering, i takt med at finansieringsproblemet accentueres. Denne strategi kan derfor virke tillokkende, da den først kræver initiativer i fremtiden. Dette efterlader et åbent spørgsmål, om disse initiativer vil blive taget i overensstemmelse med strategien, eller om tilpasningen vil blive udskudt.

Tilsvarende kan man stille spørgsmålstegn ved mulighederne for at implementere opsparingsstrategien. Ganske vist fordrer denne strategi en øjeblikkelig tilpasning, og dermed kan reformer som sådan ikke udskydes. Strategien implicerer imidlertid også, at man over en længere årrække foretager en betydelig konsolidering af de offentlige finanser samtidig med, at eksisterende velfærdsordninger fastholdes uændrede. Der vil kun være råderum til forbedringer, såfremt der samtidig sikres en finansiering. Det er et åbent spørgsmål, om det vil være politisk muligt at fastholde et sådant forløb.<sup>32</sup> Problemer knyttet til tids-inkonsistens i forhold til konsolidering kan løses ved at fondsbasere konsolideringen for dermed at sikre, at midlerne kun kan frigøres til de tiltænkte formål. Dette forudsætter, at en del af de løbende skatteindtægter overføres til en selvstændig fond med klare og afgrænsede mål for midlernes anvendelse. De politiske troværdighedsproblemer i forbindelse med denne type konstruktioner kan være betydelige, jf. debatten om den sociale pensionsfond. Selvom det skulle vise sig politisk muligt at etablere en sådan konstruktion, er det ikke givet, jf. argumentationen ovenfor, at det vil være ønskværdigt.

Som alternativ til yderpunkterne i form af PAYG- eller opsparingsstrategien diskuteres i det følgende en mere pragmatisk tilgangsvinkel.

#### *9.4 En pragmatisk tilgang – tilpasningsstrategien*

Velfærdsordninger samt skatteregler m.m. fastlægger regler for den enkeltes rettigheder og muligheder samt bidrag til finansieringen. Da der naturligt er en betydelig aldersafhængighed i brug af velfærdsordninger og bidrag i form af skattebetalinger, etableres der et væsentligt intergenerationelt element i velfærdssamfundet, jf. indled-

---

31. Auerbach og Hassett (2002a,b) betragter en situation, hvor der er politisk modvilje eller omkostninger ved hyppige politikændringer. De viser, at dette kan være en grund til træghed i ændringer i den økonomiske politik. Informationsargumentet for at udsætte ændringer i den økonomiske politik skal dog ses i sammenhæng med risiko knyttet til, at man politisk skubber problemerne foran sig.

32. Der kan drages paralleller til både USA og Norge. Konsolideringen af offentlige finanser i USA i slutningen af 1990'erne var delvist motiveret af kravet til konsolidering for at imødegå konsekvenserne af de demografiske ændringer, men er siden blevet omsat i bl.a. skattelettelser, Nordhaus (2003). I Norge er Petroleumsfonden tiltænkt at kunne finansiere en del af udgiftsstigningerne som følge af ændret demografi, men der er løbende forslag om alternative anvendelser af midlerne.

ningen om den sociale kontrakt. Et forløb med uændret politik repræsenterer således en situation, hvor man over tid og på tværs af generationer fastholder de samme regler. Det kan være en politisk målsætning at sikre en sådan ensartethed i velfærdsordningerne på tværs af generationer. Dette taler for, at finansieringen i videst muligt omfang sker ved permanente ændringer i enten ydelser eller beskatning.

Samtidig har der som nævnt ikke hidtil været nogen systematisk tendens til, at den offentlige sektor anvendes til at udjævne forbruget over mange generationer ved enten at opbygge formue eller gæld. Det er således tilsyneladende en underliggende politisk målsætning, at en given generation (og via den implicite kontrakt også den umiddelbart foregående og efterfølgende generation) får glæde af den produktivitet fremgang, som skabes i generationens levetid.

Fremadrettet kan begge disse principper ikke opretholdes samtidig pga. middelelevtidsstigningen. Det skyldes, at det medfører en permanent stigning i forsørgerforholdet. Den permanente tilpasning i den økonomiske politik, der følger af princippet om uændrede ordninger på tværs af generationer, indebærer en stor initial konsolidering for at kunne dække de fremtidige udgiftsstigninger. Dette er imidlertid i strid med sidstnævnte princip om at undgå betydelig omfordeling i form af overførsler fra nuværende til fremtidige generationer.

Den centrale udfordring er at etablere en vis risikodeling samt stabilitet i velfærdsordningerne over tid under overholdelse af de finansieringsmæssige krav. Herved sikres, at makroøkonomisk usikkerhed ikke transformeres til usikkerhed for den enkelte i form af usikkerhed omkring, hvilke ordninger der vil gælde selv i en nær fremtid. Endvidere skal ordningerne indrettes således, at de ikke forårsager betydelig omfordeling i forhold til trend-ændringer. Det betyder konkret, at politikken skal indrettes, så nulevende generationer ikke automatisk kommer til at tage del i finansieringen af forhold, der ligger ude i en fjern fremtid. Omvendt er det nødvendigt, at politikken indrettes i forhold til de underliggende trends i udviklingen.

For at sikre en vis risikodeling samt stabilitet i ordningerne uden samtidig at foretage væsentlig omfordeling mellem generationer forudsættes en skarpere sondring mellem trends og afvigelser fra trend. Førstnævnte har overvejende implikationer for den intergenerationelle fordeling, mens sidstnævnte er knyttet til risiko og risikospredning.

En strategi, der sikrer en tilpasning til underliggende trends, men samtidig muliggør udjævning over tid og generationer af afvigelser fra trends, kan betegnes en »tilpasningsstrategi«. En sådan strategi sikrer, at der i velfærdsordningerne indbygges en automatik, således at velfærdsordningerne automatisk justeres i forhold til væsentlige trend-ændringer. Den konkrete problemstilling med en opadgående trend i levetiden kan modvirkes ved at indbygge en mekanisme i ordningerne, som sikrer, at den enkel-



te generations nettobidrag er relativt uafhængigt af udviklingen i levetiden.<sup>33</sup> Dette kan f.eks. ske ved at indeksere aldersgrænser for tilbagetrækningsordningerne til middellevetiden. Det skal dog understreges, at en sådan indeksering ikke i sig selv er tilstrækkelig til at sikre finanspolitisk holdbarhed. Årsagen er, at finansieringsproblemet ikke alene skabes af tilbagetrækningsalder (set i forhold til levetid), men også af trækket på forskellige former for velfærdsservice. Samtidig er alle ikke erhvervsaktive, og senere tilbagetrækning må også forventes at øge antallet af f.eks. førtidspensionister. Levetidsindeksering er derfor en vigtig del af en tilpasningsstrategi, men den kan ikke stå alene som løsningen på den trendmæssige udvikling i nettoudgifterne.

En tilpasningsstrategi har den fordel, at usikkerheden knyttet til at forudsige den fremtidige udvikling i middellevetiden får mindre betydning for de økonomisk politiske krav her og nu. Skulle levetiden vokse mere eller mindre, end det vurderes i dag, vil der automatisk ske en justering. En sådan ordning vil ikke fjerne behovet for at overveje risikodeling. Befolkningsprognoser revideres løbende og med væsentligt kortere intervaller end middellevetiden for en nyfødt. Der kan derfor være et argument for at sprede noget af denne risiko. I forhold til usikkerheden omkring den præcise ændring i levetiden er der et forsikringsargument ligesom i almindelige pensionsforsikringer, nemlig »risikoen« for en lang levetid og dermed behovet for et forsørgelsesgrundlag. Risikodeling uden omfordeling<sup>34</sup> fordrer risikospredning i forhold til hændelser, den enkelte generation ikke har haft mulighed for at forudsige og tilpasse sig, se f.eks. Andersen (2005). I forhold til levetid kan der argumenteres for risikospredning i forhold til ændringer, der indtræffer, efter at pensionsforhold (inkl. tilbagetrækning) er tilrettelagt og overvejende irreversible. I praksis betyder dette, at den enkelte med en horisont på måske 5 til 10 år skal kende sine pensionsvilkår (set i forhold til de ændringer i levetid vi står overfor, er den tilbageværende risiko minimal).

Tilpasningsstrategien kræver ikke løbende balance på de offentlige budgetter, men muliggør variationer over tid. Derimod forudsætter tilpasningsstrategien for at være forenelig med finanspolitisk holdbarhed, at den er tilrettelagt på en sådan måde, at der ikke er nogen systematisk tendens til underskud eller gældsakkumulering (trends elimineres).

### *9.5 Tilpasningsstrategi og velstandsdilemmaer*

Ovenstående har primært omhandlet en økonomisk-politisk strategi i forhold til de demografiske forandringer. En anden underliggende trend skabes af stigende velstand. På den ene side vil vækst ikke automatisk skabe et øget finansielt råderum, tværtimod er der en tendens til, at det forværres, jf. afsnit 8. På den anden side skaber øget vækst

---

33. Forudsætter, at levetidsstigningen af fremtidige generationer opfattes som en velfærdsfremgang.

34. I den forstand, at der ikke er et ex-ante forventet tilskud til en bestemt generation.

et velstandsdilemma i form af øget behov og krav til service og mere fritid, jf. afsnit 3.1. Styrken af denne trend er ganske betragtelig, jf. beregningerne præsenteret i afsnit 5 og 6, der viser kravene til finansiering både under PAYG-finansiering og den holdbare skat i forhold til velstandsdilemmaet.

Hvis man baserer velfærdspolitikken på uændrede regler og fortsat lader gevinsterne ved økonomisk vækst være knyttet til de enkelte generationer, indebærer velstandsdilemmaet en risiko for en »systemfejl«. Årsagen er, at det vil efterlade et stadigt stigende gab mellem det serviceudbud, den offentlige sektor er i stand til at finansiere, og det serviceudbud, der efterspørges.<sup>35</sup>

I det omfang efterspørgslen ikke imødekommes ved et offentlig udbud, vil der opstå et pres for at få denne dækket på anden vis. Det kan føre til parallelle private markeder eller inden for områder som sundhed til privatfinansierede løsninger på udenlandske hospitaler. En sådan udvikling kan komme i konflikt både med en effektiv udnyttelse af ressourcerne og med målsætningerne om lige adgang til grundlæggende serviceydelser som sundhed og pleje.

Det har været fremført, at det øgede udgiftstryk, øget serviceefterspørgsel og mere fritid kan finansieres via skattestigninger. Dette er selvfølgelig et politisk valg. Med en underliggende velstandsstigning vil der trods skattestigninger fortsat være muligheder for stigninger i den disponible indkomst, jf. Velfærdskommissionen (2004). Det er imidlertid et åbent spørgsmål, om øget materiel velstand i sig selv øger villigheden til skattebetaling. De senere års udvikling giver ikke umiddelbart støtte til, at dette vil være tilfældet. Endvidere vil eksternaliteten knyttet til fritid og andre forvriddinger være stigende i skatternes niveau, samtidig med at skattesystemet presses direkte og indirekte i kraft af globaliseringen, jf. Velfærdskommissionen (2005a).

Historisk har velstandsdilemmaet ikke været så tydeligt, da en løbende udvidelse af velfærdssamfundet har været mulig bl.a. i kraft af skattestigninger og udvidelser af arbejdsstyrken (primært via øget erhvervsfrekvens for kvinder). Fremadrettet står vi derfor over for et vanskeligt problem, hvor de historiske tilpasningsmuligheder ikke i samme udstrækning kan anvendes.

En langsigtet strategi, der gør velfærdsordningerne robuste over for trends, indebærer, at der indbygges mekanismer, som sikrer, at den vækst i serviceefterspørgslen, der ligger ud over den generelle vækst i indkomsten, kan imødekommes uden at skabe et permanent finansieringspres. At sikre dette samtidig med en lige adgang til centrale serviceydelser er en af hovedudfordringerne for velfærdssamfundet i fremtiden. At

---

35. Den aktuelle økonomiske politik, som formuleret i 2010-planen, indebærer et endnu større velfærdsdilemma, idet planen indebærer, at væksten i udgifterne til offentlig service skal være lavere end væksten i indkomsten.

løse dette grundlæggende dilemma er langt vanskeligere end at tilpasse de eksisterende velfærdsordninger til den stigende levetid.

### 10. Konklusion

De forestående demografiske forandringer er markante. Der vil indtræde en stor forskydning i forholdet mellem aldersgrupper og dermed en permanent stigning i forsørgerforholdet. Med et veludviklet velfærdssamfund som det danske og en væsentlig social kontrakt i velfærdsordningerne vil dette uvilkårligt ændre det finansielle grundlag for velfærdssamfundet. Uændret politik og dermed fremadrettet de samme velfærdsordninger og bidrag i form af skattebetalinger m.m. vil ikke finansielt være mulig. Systematiske underskud og stigende gæld vil på et tidspunkt kræve tilpasninger af velfærdssamfundet enten i form af ændringer i velfærdsordninger og/eller ændret finansieringsgrundlag. Beregninger af størrelsen af de nødvendige tilpasninger for at sikre finanspolitisk holdbarhed viser et problem af væsentligt omfang.

Yderligere står vi over for det velstandsdilemma, at øget økonomisk vækst vil forstærke krav og behov til offentlig service og til fritid. Dette vil forstærke det finansieringsproblem, vi står over for. Velstandsdilemmaet vedrører centrale spørgsmål for vores velfærdssamfund: Løser velfærdssamfundet de rigtige opgaver og på en tilfredsstillende måde? Det er vigtigt at slå fast, at disse problemer ikke løser sig selv. Hvis man ikke forholder sig til spørgsmålene, vil der være risiko for stigende frustration i befolkningen, i takt med at de oplever et spænd mellem »ønsker« og realiteter. Samtidig må man forvente, at der på mange områder vil være nogle, som finder »private« løsninger på problemet. Begge dele er problematiske for velfærdssamfundet, da det kan føre til, at flere og flere begynder at sætte spørgsmålstegn ved, hvad de egentlig får for skatterne, og dette kan påvirke opbakningen til velfærdssamfundet.

Den politiske reaktion på udfordringerne og ikke mindst timingen af ændringer i den økonomiske politik er selvfølgelig ultimativt et politisk spørgsmål. Da problemet vedrører centrale aspekter ved velfærdssamfundet, og da den sociale kontrakt spiller en stor rolle her, kan problemstillingen ikke uden videre elimineres med henvisning til at »det vedrører en fjern fremtid«. Den sociale kontrakt medfører et kompliceret samspil på et givet tidspunkt mellem nulevende tilhørende forskellige aldersgrupper og over tid mellem forskellige generationer. Den sociale kontrakt implicerer blandt andet, at uddannelse finansieres for børn og unge ud fra en forventning om, at de i fremtiden vil være skattebetalere, ligesom erhvervsaktive i dag betaler skatter i forventning om som ældre at kunne trække på velfærdssamfundet. Velfærdssamfundets indretning om 10, 20 og 30 år vedrører i mindst lige så stor grad allerede fødte som de endnu ufødte børn. Fravær af initiativer og udskydelse af problemerne skaber således usikkerhed for borgerne, både i forhold til hvilke velfærdsordninger de kan regne med og til skatte-

betalinger. En konsekvens af den sociale kontrakt er således, at man nødvendigvis må være fremadskuende i overvejelserne om velfærdssamfundets indretning. Ellers efterlader man usikkerhed og utryghed, hvilket netop er i modstrid med et af velfærdssamfundets hovedformål.

Diskussion om behovet for reformer er ikke et spørgsmål om i detaljer at planlægge den økonomiske politik i en lang årrække frem. Det har ingen mening. Dertil er usikkerheden og omfanget af strukturelle ændringer m.m. for stor. Derimod er det nødvendigt at indrette den økonomiske politik ud fra nogle pejlemærker og sikre troværdighed om de ordninger, der er etableret. Det implicerer bl.a., at man må forholde sig til klare trends i forsørgerforholdet, herunder længere levetid.

I forhold til alle fremskrivninger og overvejelser om økonomisk politiske strategier er der et tids-konsistensproblem: Vil fremtidens politikere også følge det udlagte spor? Dette troværdighedsproblem kan ikke elimineres, men nogle forløb vil være mere troværdige end andre. Det er ikke i sig selv noget mål at udelukke fremtidige ændringer i den økonomiske politik, men derimod at sikre at det sker, fordi man prioriterer anderledes, og ikke fordi man skal løse problemer, der er overladt til fremtiden, fordi tidligere generationer ikke har taget dem alvorligt eller har udskudt dem.

Afslutningsvis kan der være grund til at understrege, at de centrale problemstillinger diskuteres ovenfor udspringer af potentielle velfærdsforbedringer i kraft af øget levetid og fortsat stigende materiel levestandard. Det er således en positiv problemstilling, der skal løses.

#### Litteratur

- Andersen, T. M. 2005. Social Security and Longevity, *Working paper*.
- Andersen, T. M., and R. Dogonowski. 2002. Social Insurance and the Public Budget, *Economica*, 69, s. 415-32.
- Andersen, T. M., S. E. H. Jensen og L. H. Pedersen. 2004. The Welfare State and Strategies Towards Fiscal Sustainability in Denmark, *Arbejdsrapport 2004:4 Velfærdskommissionen*.
- Auerbach, A. J., and K. Hassett. 2001. Uncertainty and the Design of Long-Run Fiscal Policy, in A. J. Auerbach, and R. D. Lee, red. *Demographic Change and Fiscal Policy*, Cambridge University Press, Cambridge, UK.
- Auerbach, A., and K. Hassett. 2002a. Fiscal Policy and Uncertainty, *International Finance*, 5(2), s. 229-42.
- Auerbach, A., and K. Hassett. 2002b. Optimal Long-Run Fiscal Policy: Constraints, Preferences and the Resolution of Uncertainty, *NBER Working Paper 7036*, National Bureau of Economic Research, Boston, Mass.
- Barro, R. 1974. Are Government Bonds Net Wealth?, *Journal of Political Economy*, 82, s. 1095-1117.
- Barro, R. 1979. On the Determination of Public Debt, *Journal of Political Economy*, 87, s. 940-71.
- Blanchard, O., and S. Fischer. 1989. *Lectures on Macroeconomics*, MIT Press.
- Bohn, H. 2001. Social Security and Demographic Uncertainty: The Risk-Sharing Properties of Alternative Policies, in Campbell, J. and M. Feldstein, red. *Risk Aspects of Investment Based Social Security Reform*, University of Chicago Press, Chicago.

- Burnside, C. 2004. Fiscal Sustainability in Theory and Practice, bogmanuskript, Duke University.
- Danmarks Statistik. 2005. Middellevetiden 2003-2004, *Nyt fra Danmarks Statistik*, nr. 195, maj 2005.
- Diamond, P. 2003. Taxation, *Incomplete Markets and Social Security*, Munich Lectures in Economics, MIT Press.
- Dixit, A. K., and R. S. Pindyck. 1994. *Investment under Uncertainty*, Princeton University Press.
- Flodén, M. 2003. Public Saving and Policy Coordination in Aging Economies, *Scandinavian Journal of Economics*, 105, s. 379-400.
- Finansministeriet. 2004. *Finansredegørelse 2004*, København.
- Gordon, R., and H. Varian. 1988. Intragenerational Risk Sharing, *Journal of Public Economics*, 37, s. 185-202.
- Haldrup, N. 2004. Estimation af middellevetider for mænd og kvinder i Danmark, 2002-2010, baseret på Lee-Carter metoden, Arbejdsrapport 2004:3, Velfærdskommissionen.
- Hansen, N. P., S. E. H. Jensen, and M. Junge. 1999. Government Solvency, Social Security and Debt Reduction in Denmark, in T. M. Andersen, S. E. H. Jensen, and O. Risager: *Macroeconomic Perspectives on the Danish Economy*, kap. 7. Macmillan Press, London, UK.
- Hansson, I. and C. Stuart. 1989. Social Security as Trade Among Living Generations, *American Economic Review*, 79, s. 1182-95.
- Knudsen, M., L. Pedersen, T. Petersen, P. Stephensen, and P. Trier. 1998. Danish Rational Economic Agents Model – DREAM Ver. 1.2, Working paper, [www.dreammodel.dk](http://www.dreammodel.dk), DREAM, København.
- Knudsen, M. B., L. H. Pedersen, T. W. Petersen, P. Stephensen, and P. Trier. 1999. Dynamic Calibration of a CGE-Model with a Demographic Application, Working paper, [www.dreammodel.dk](http://www.dreammodel.dk), DREAM, København.
- Lee, R., and L. Carter. 1992. Modelling and Forecasting the Time Series of U.S. Mortality, *Journal of the American Statistical Association*, 87, s. 659-71.
- Markeprand, T., L. H. Pedersen, og P. Stephensen. 2003. DREAMs disaggregerede befolkningsfremskrivning til år 2100. Resultater og Metode.
- Møller, F. 2004. Bør der diskonteres i forbindelse med velfærdsøkonomisk projektvurdering, *Nationaløkonomisk Tidsskrift*, 142, s. 95-123.
- Nordhaus, W. D. 2003. Comment on A.J. Auerbach, Fiscal Policy, Past and Present, *Brookings Papers on Economic Activity*, 1, s. 130-34.
- OECD. 2005. *Economic Surveys – Denmark*, Paris.
- Pedersen, L., P. Stephensen, and P. Trier. 1999. A CGE Analysis of the Danish Ageing Problem, Working paper, [www.dreammodel.dk](http://www.dreammodel.dk), DREAM, København.
- United Nations. 2004. World Population in 2300. UN, New York.
- Velfærdskommissionen. 2004. *Fremtidens velfærd kommer ikke af sig selv*. Analyse-rapport. Velfærdskommissionen, København.
- Velfærdskommissionen. 2005a. *Fremtidens velfærd og globaliseringen*. Analyserapport. Velfærdskommissionen, København.
- Velfærdskommissionen. 2005b. *Fremtidens velfærd – sådan gør andre lande*. Analyse-rapport. Velfærdskommissionen, København.
- Velfærdskommissionen. 2005c. *Fremtidens velfærd og globaliseringen*. Debatoplæg. Velfærdskommissionen, København.
- Velfærdskommissionen. 2005d. *Demografi, velstands dilemma og makroøkonomiske strategier – teknisk analyserapport*, København.
- Økonomiske Råd, Det. 2004. *Dansk Økonomi*, forår 2004, København.

## Velfærdskommissionens analyse af finanspolitisk holdbarhed – før og nu

Christen Sørensen

Institut for Regnskab og Finansiering, E-mail: chr@sam.sdu.dk

### 1. Indledning

Velfærdskommissionen skal i henhold til dets kommissorium bidrage til en »bred og åben dialog med borgerne og Folketingets partier om fremtidige velfærdsreformer. Det er derfor regeringens ønske, at kommissionen arbejder åbent med henblik på en løbende debat om kommissionens arbejde og temaer«.

At deltage i denne diskussion er såvel spændende som udfordrende. Dette blev allerede klart for mig, da jeg indleverede mit manuskript: »Velfærdskommissionens analyse af finanspolitisk holdbarhed – er den holdbar?« i december 2004 til redaktionen af Nationaløkonomisk Tidsskrift. Redaktionen ville nemlig først bringe mit indlæg i forlængelse af et indlæg fra Velfærdskommissionen. Dette bidrag skulle foreligge primo april 2005, men først i slutningen af juni forelå et endda ufærdigt manuskript – og ikke fra Velfærdskommissionen, men fra dets formand og dets sekretariatsleder. Udsættelsen betød endvidere, at jeg bl.a. besluttede at udgive det fremsendte manuskript i mit instituts interne skriftserie, jf. Sørensen (2005). Og da det endelige manuskript kom i slutningen af august, var det på afgørende punkter ændret. Det er nemlig nu i al væsentlighed sammenfaldende med det sammendrag af en teknisk analyserapport »Befolkningsudvikling, velstandsdilemma og makroøkonomiske strategier«, som Velfærdskommissionen efterfølgende har udgivet – og holdt møde med repræsentanter fra pressen om – selv om resten af rapporten endnu ikke er frigivet. Så læserne af Nationaløkonomisk Tidsskrift får lejlighed til at få denne sammenfatning i stort set samme udgave. Ikke, at der ikke kan peges på forskelle, jf. konklusionen, men om de er bevidste eller ubevidste, står mig ikke klart. Men det er nok min fejl.

Forfatterne til bidraget i Nationaløkonomisk Tidsskrift, Torben M. Andersen og Lars Haagen Pedersen understreger dog, at »synspunkterne i dette papir er forfatternes og ikke nødvendigvis sammenfaldende med Velfærdskommissionens«. Denne fremhævelse kan skyldes, at dette indlæg – efter min opfattelse – er ret så forskelligt fra Velfærdskommissionens første rapport fra maj 2004: »*Fremtidens velfærd kommer ikke af sig selv*«. Og det var denne rapport, jeg anmodede om at kommentere i Nationaløkonomisk Tidsskrift. Derimod er bidraget som allerede anført stort set sammenfaldende med sammenfatningen af den tekniske analyserapport, og overskriften i de to

indlæg er da også praktisk taget ens. Dette er ikke den eneste grund til, at det er svært – i hvert fald for mig – at kommentere på »*Demografi, velstandsdilemma og makro-økonomiske strategier*«. For Velfærdskommissionens formand udtaler jo gang på gang, at der ikke er en forskel mellem den første rapport, og hvad der nu fremføres. Hvordan dette så hænger sammen med, at der i det fremsendte indlæg direkte anføres »at de centrale problemstillinger ... udspringer af potentielle velfærdsforbedringer i kraft af øget levetid og fortsat stigende materiel velfærd. Det er således en positiv problemstilling, der skal løses« og ikke mindst med »I det forudgående er to yderpunkter sammenlignet, nemlig den løbende finansiering via en såkaldt PAYG-finansiering eller en opsparingsstrategi med en permanent tilpasning for at sikre velfærdsordningernes finansielle grundlag. Begge disse to yderpunkter lider af alvorlige problemer både principielt og i forhold til implementering«, er for mig svært at forstå, især når de til yderpunkterne knyttede problemer får forfatterne til at anbefale en tilpasningsstrategi, jf. også nedenfor. Jeg erindrer ikke disse formuleringer om alvorlige problemer munden ud i en tilpasningsstrategi i den første rapport med en 1000-årig horisont. Men det er da givetvis mig, der ikke har læst grundigt nok på lektien. Men det bliver nu heller ikke lettere at forstå overensstemmelsen, når henses til det indtryk, som Velfærdskommissionens præsentation fik pressen til at formidle efter publikationen af den første rapport – og som fik lov til meget, meget længe at være uimodsagt. Så derfor er det ikke let at vide, hvad der egentlig skal kommenteres på – også fordi det som anført alene er sammenfatningen af den tekniske analyserapport, der er offentliggjort, når dette skrives.

Derfor har jeg besluttet, at jeg vil tage udgangspunkt i et kort resume af »*Velfærdskommissionens analyse af finanspolitisk holdbarhed – er den holdbar?*« inden for den sidemæssige begrænsning, som jeg er bevilliget af redaktionen. Dette gør det også muligt, at jeg i afsnit 4 direkte kan give de indrømmelser til Velfærdskommissionen, som følger af den ovenfor anførte forhistorie. Det kunne måske inspirere Velfærdskommissionen – hvem ved.

Da det finanspolitiske holdbarhedsproblem i al væsentlighed udspringer af den forventede stigende levetid, når der tages udgangspunkt i Velfærdskommissionens grundlæggende forudsætninger i »*Fremtidens velfærd kommer ikke af sig selv*« om et fastholdt offentligt serviceniveau og et fastholdt skattetryk, opridses antagelserne om stigende levetid i afsnit 2. Herefter følger en gennemgang af terminalproblemet i relation til finanspolitisk holdbarhed, der var og er mit væsentligste kritikpunkt mod »*Fremtidens velfærd kommer ikke af sig selv*«. I afsnit 4 resumeres min kritik af Velfærdskommissionens beregningsforudsætninger med tilhørende indrømmelser. Herefter følger en kort konklusion.

Tabel 1. Krav til finanspolitisk holdbarhed ved alternative middellevetider.

	Nødvendig stigning i bundskatteprocenten, procentpoint	Nødvendig beskæftigelsesstigning, antal personer
<i>Middellevetid:</i>		
Som i grundscenariet	8,7	280.000
Uændret middellevetid	1,6 <sup>(b)</sup>	67.000
DREAMs middellevetidsforudsætninger <sup>(a)</sup>	5,8	195.000
FN's middellevetidsforudsætninger	16,0	492.000

*Noter:* (a) Disse forudsætninger ligger til grund for Det økonomiske Råds og Finansministeriets beregninger i relation til finanspolitisk holdbarhed, hvorefter middellevetiden er 80,4 år for mænd og 82,0 år for kvinder i år 2100. Det er altså primært for kvinder, at Velfærdskommissionen antager længere middellevetid. (b) Er senere korrigeret til 1,9. Det er dette tal, der er anvendt i teksten.

*Kilde:* Velfærdskommissionen (2004, s. 26 og s. 249-56).

## 2. Udviklingen i middellevetiden

I grundscenariet antager Velfærdskommissionen, at middellevetiden gradvist stiger for såvel mænd som kvinder. For mænd forventes der således en middellevetid på godt 80 år i år 2100, dvs. en stigning på knap 6 år i forhold til år 2000, mens middellevetiden for kvinder antages at blive godt 87 år i år 2100, eller 7,7 år mere end i år 2000. Velfærdskommissionen (2004, s. 21) anfører og begrundet, at dette er en relativ beskedent stigning, hvorfor der også gennemregnes et alternativ, der svarer til FN's prognose for stigningen i middellevetiden i Vesteuropa som helhed, hvorefter middellevetiden for mænd og kvinder skulle stige med henholdsvis godt 10 år og 11 år. Betydningen af antagelserne om middellevetiden for det finanspolitiske holdbarhedsproblem belyses endvidere af, at det beregnes, hvor stort det finanspolitiske holdbarhedsproblem er i to andre alternativer: med uændret middellevetid og med den stigning i middellevetiden, der ellers er regnet med i anvendelsen af DREAM m.v., jf. tabel 1.

Det er meget muligt og måske endda det mest sandsynlige, at Velfærdskommissionens og måske endog FN's antagelser om udviklingen i middellevetiden vil blive de bedste skøn. Men dette forstærker blot nødvendigheden af eksplicit at inddrage et yderligere instrument til sikring af finanspolitisk holdbarhed: ikke mindst aldersgrænsen for overgang til alderspension, især når stort set hele det beregnede finanspolitiske holdbarhedsproblem hidrører fra stigende levetid, nemlig 6,8 procent ud af de 8,7 procent.

Derfor var det heller ikke let at forstå, hvorfor der ikke var foretaget beregninger heraf i Velfærdskommissionen (2004). Dette er der dog rettet noget op på i den netop »udkomne« analyse rapport, jf. Velfærdskommissionen (2005).



### 3. Terminalproblemet og finanspolitisk holdbarhed

I min gennemgang af Velfærdskommissionens første rapport var der to hovedelementer i min kritik. For det første kritiserede jeg de valgte beregningsforudsætninger for prognoseperioden på fire punkter: (1) fremskrivningen af udgifter til individuel service, (2) fremskrivningen af befolkningens uddannelsesniveau, (3) fremskrivningen af erhvervsfrekvensen for efterkommere af indvandrere fra ikke vestlige lande og (4) den anvendte vækstkorrigerede realrente. For det andet kritiserede jeg, at målingen af kravene til finanspolitisk holdbarhed var sket over en 1000-årig periode – dette blev benævnt terminalproblemet. I min sammenfatning heraf blev der anført: »Det væsentligste kritikpunkt er efter min vurdering, at terminalproblemet helt er ignoreret i det empiriske arbejde«.

Jeg vil derfor her begynde med det væsentligste, nemlig terminalproblemet. Dette gør jeg også, fordi der ikke direkte kommenteres på dette i bidraget fra Torben M. Andersen og Lars Haagen Pedersen, hvorfor den væsentligste del af min kritik ikke er kommenteret. Og dog. Hvordan skal jeg f.eks. ellers forstå anbefalingen af tilpasningsstrategien i deres papir? Og i sammenfatningen af den tekniske analyserapport står der: »Renteantagelsen har *ingen* reel betydning for det egentlige problem, der består i at de offentlige finanser systematisk forværres, og at der derfor vil være tendens til fremtidige systematiske budgetunderskud. ... Det fremgår, at forløbene af den offentlige sektors primære saldo under de to sæt forudsætninger for alle praktiske formål er de samme.« Er dette ikke en anden måde at formulere dette, jeg har anført, især når den ledsagende figur kun går frem til 2076? Men det kan det jo ikke være, hvis det, der nu gives udtryk for, var det samme som i den oprindelige rapport – eller hvad?

Terminalproblemet i Velfærdskommissionens holdbarhedsberegninger hænger sammen med, at der i disse beregninger anlægges en i princippet uendelig tidshorisont – en 1000-årig periode – samtidig med den gradvise stigning i levealderen. Dette kombineret med, at Velfærdskommissionen foretager et permanent engangsløft i de undersøgte indgreb, når det beregnede finanspolitiske underskud i grundforløbet skal dækkes, resulterer i resultater, der godt nok er matematisk korrekte, men som ikke desto mindre mangler økonomisk relevans, som det vel også nu erkendes i forbindelse med fremhævnningen af tilpasningsstrategien.<sup>1</sup>

Det mest iøjnefaldende eksempel herpå fremkommer i alternativet, hvor konsekvensen af en stigning i vækstraten med 1/2 procentpoint simuleres. Dette vil kræve, at f.eks. bundskatteprocenten yderligere skal forøges med 4,1 procentpoint fra 8,7 pct. til 12,8 pct. Årsagen hertil er, at der i henhold til Velfærdskommissionens antagelser med bl.a. fastholdte tilbagetrækningsgrænser er offentlige nettoudgifter knyttet til alle ny-

---

1. Engangsløft i de undersøgte indgreb foretages med to modifikationer: indgrebene iværksættes først i 2011 og indføres endvidere lineært i perioden 2011 til 2021.

fødte og dermed til alle fremtidige generationer. Dette betyder, at de nuværende generationer skal foretage en så stor offentlig opsparing, at fremtidige underskud kan finansieres. Dette ses f.eks., hvis situationen i 2041 (i tilfældet, hvor bundskatteprocenten er øget med 8,7 procentpoint) betragtes. Generationerne i 2041 overtager så ikke blot det offentlige produktionsapparat helt gældfrit, men overtager også offentlige finansielle nettofordringer svarende til 82 pct. af BNP. Dette skal endvidere ses i lyset af, at Danmark i 2041 i dette forløb har udenlandske tilgodehavender opgjort netto svarende til 120 pct. af BNP, jf. tabel 2.

Hvor paradoksale Velfærdskommissionens forudsætninger – med fastholdte tilbagetrækningsaldrer, selv om levetiden stiger markant – er, kan ses, hvis der betragtes et tilfælde med faldende vækst. En øget vækstrate med  $1/2$  procentpoint forøger som nævnt holdbarhedsproblemet som målt ved bundskatteprocenten med 4,1 procentpoint. Heraf følger med forbehold for ikke-linearitet at holdbarhedsproblemet på 8,7 procentpoint ifølge Velfærdskommissionen kunne løses ved, at vi slappede mere af på jobbet og dermed reducerede den årlige vækstrate med godt 1 procentpoint!

Der skal ikke blot kritiseres, men også fremføres forslag til, hvordan mere retvisende beregninger kunne være gennemført.

For det første savnes der i Velfærdskommissionen helt en diskussion af, hvordan den offentlige sektors status skal opgøres. Det antages uden videre, at det alene er de finansielle poster, der skal indgå. En sådan trunckeret formueopgørelse er på ingen måde oplagt. Manglen på en problematisering heraf understreges yderligere af, at den offentlige sektors kapitalapparat primo 2003 er opgjort til 36 pct. af BNP alt sammen i løbende priser.<sup>2</sup>

For det andet burde der i de empiriske beregninger være taget stilling til, hvad de kommende og mere velhavende og længere levende generationer kan forvente at få overdraget af offentlig formue opgjort på den ene eller anden måde.

Bedømt ud fra fødende kvinders alder er der ca. 30 år mellem generationerne. I det følgende vil jeg derfor gå ud fra, at det er relevant at se på »overdragelsessituationen« ca. 30 år fra nu. Det år, som der er publiceret tal for i Velfærdskommissionens rapport, og som er nærmest det således definerede overdragelsesår, er 2041.<sup>3</sup>

Det er på ingen måde en selvfølge, at beregninger af finanspolitisk holdbarhed skal ske med en i princippet uendelig tidshorisont, jf. bl.a. Broda og Weinstein (2004). Broda og Weinstein analyserer f.eks. holdbarheden af den japanske finanspolitik med en tidshorisont på hhv. 36 og 96 år. I relation til Velfærdskommissionens opgave kunne

2. Se Statistiske Efterretninger. *Nationalregnskab og betalingsbalance* 2003:15 og 2004:14.

3. Valget af dette år kan også ses i lyset af, at såvel Velfærdskommissionen som Finansministeriet i deres analyser af det finanspolitiske holdbarhedsproblem ikke simulerer indgreb med henblik på at sikre finanspolitisk holdbarhed før i 2011.

Tabel 2. Den offentlige sektors nettofordringer og udenlandske tilgodehavender i pct. af BNP i 2041.

	Den offentlige sektors nettofordringer <sup>(a)</sup>	Udenlandske tilgodehavender
1. Grundforløbet	-30,0	-12,9
2. Grundforløbet + 8,7 procentpoint højere bundskatteprocent	81,7	120,8
3. Grundforløbet + off. besparelser	84,9	93,6
4. Grundforløbet + stigende beskæftigelse	69,6	73,2
5. (2) + 1/2 procentpoint højere vækst <sup>(b)</sup>	148,9	226,0

Noter: (a) Det skal fremhæves, at ATP er inkluderet i den offentlige sektor. (b) I dette tilfælde er bundskatteprocenten yderligere forøget med 4,1 procentpoint fra 8,7 procent til 12,8 procent.

Kilde: Velfærdscommissionen (2004, kapitel 12).

dette være sket med eksempelvis år 2041 og år 2100 som terminalår. I det følgende koncentrerer opmærksomheden om år 2041 som terminalår.

Af tabel 2 fremgår, at den offentlige sektor i 2041 i grundforløbet har en nettogæld på 30 pct. af BNP, hvilket kun marginalt overstiger værdien af det offentlige kapitalapparat.<sup>4</sup>

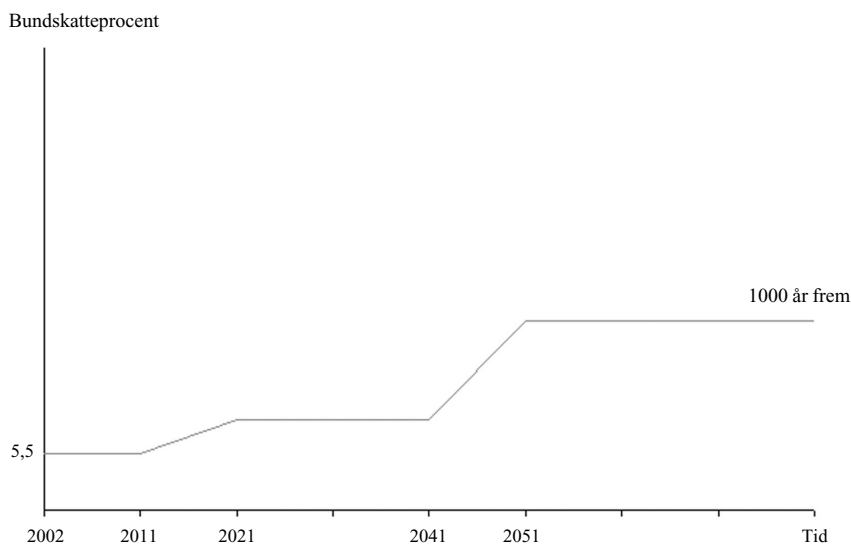
Ud fra dette synspunkt overfører den nuværende generation, såfremt den økonomiske politik fastholdes, derfor ikke et finanspolitisk holdbarhedsproblem til de næste generationer.<sup>5</sup> Det kan naturligvis diskuteres, om der ikke hellere skulle tages udgangspunkt i en trunkeret offentlig formueopgørelse som opgjort ved den finansielle nettostilling. Her skylder Velfærdscommissionen, men også Det økonomiske Råds

4. Primo 2001 udgjorde det offentlige kapitalapparat 36 pct. af BNP. I grundforløbet stiger real BFI med 83,2 pct. fra 2001 til 2041, mens det reale offentlige kapitalapparat stiger med 40,5 pct. i samme periode, jf. Velfærdscommissionen (2004, tabel 12.2 s. 316). Skønsmæssigt udgør det offentlige kapitalapparat derfor

$$36 \text{ procent} \cdot \frac{1,405}{1,832} = 28 \text{ pct. af BNP i 2041.}$$

Indtil 2032 har den offentlige sektor inkl. ATP endvidere flere finansielle aktiver end passiver i grundforløbet.

5. Det finanspolitiske holdbarhedsproblem kan også sættes i perspektiv af, at Det økonomiske Råds formandskab (2004, s. 135) har beregnet, at: »Den nuværende regering har, siden den tiltrådte i 2001, ført en lempelig skattepolitik, der samlet har forværret holdbarheden svarende til en stigning i bundskatten på lidt over 2 1/2 procentpoint. Den lempelige skattepolitik består hovedsagelig af to elementer: skattestoppet og en ufinansieret lempelse af personskatten som følge af skattepakken.«



Figur 1. Grundforløb med ingen nettogæld i 2041.

formandskab og Finansministeriet, en forklaring. Men tages der udgangspunkt i den således definerede trunkeerede offentlige formuestilling, er der i værste fald et mindre finanspolitisk holdbarhedsproblem svarende til en nettogæld på 30 pct. af BNP i 2041, jf. tabel 2. Men dette berettiger ikke et vedvarende løft i bundskatteprocenten på 8,7 procentpoint for den nuværende generation.

I konsekvens af ovenstående – og den nu anbefalede overgang til tilpasningsstrategien – bør Velfærdskommissionen oplyse resultaterne af, hvad stigningen i bundskatteprocenten skal være med et mellemliggende slutpunkt i 2041 og eksempelvis også i 2061, 2076 samt 2100, når den offentlige sektors finansielle nettogæld sættes til nul, og der regnes med et fastholdt offentligt serviceniveau samt fastholdt skattetryk i øvrigt. Med 2041 som eksempel er dette illustreret i figur 1, hvor der også efter 2041 er anvendt en ti-årig indfasningsperiode.

Det fremgår af ovenstående, at der også bør foretages beregninger, hvor i stedet den offentlige sektors nettoformue sættes til nul. Dette betyder, at værdien af reale aktiver også indregnes. Resultatet vil efter min vurdering blive som vist i tabel 3, altså kun en nødvendig stigning i bundskatteprocenten efter 2041, hvis disse forudsætninger lægges til grund.

I såvel Velfærdskommissionens grundforløb, som i de her skitserede alternative forløb, bør der endvidere indlægges en stigende tilbagetrækningsalder, herunder pensionsalder – for yngre og kommende generationer. Det skyldes ikke mindst, at 6,8 pro-

Tabel 3. Nødvendig stigning i bundskatteprocenten til sikring af finanspolitisk holdbarhed. Procentpoint.

	Med uændret pensionsaldersgrænse	Med gradvist stigende pensionsalder fra 2021
Velfærdskommissionens oprindelige grundforløb	8,7	?
Med terminalpunkt i 2041, hvor den off. finansielle nettogæld er nul		
– indtil 2041	?	?
– efter 2041	?	?
Med terminalpunkt i 2041, hvor den off. formue er nul		
– indtil 2041	0 (mit skøn)	?
– efter 2041	?	?

Ann.: Tilsvarende beregninger bør gennemføres, hvor 2041 er erstattet af bl.a. 2061, 2076 og 2100.

cent af de 8,7 procent alene skyldes længere levetid. Hvis det eksempelvis lægges til grund, at dette kommer til at gælde for personer, der nu er 50 år eller derunder, vil der gå 15 år, inden pensionsalderen gradvist begynder at stige. Med en vedtagelse i 2006 er det altså fra og med 2021. Og det er meget vigtigt, at ikke mindst ændringer i pensionsalder meddeles i god tid, så forventninger og adfærd kan indrettes herefter.

Fra og med 2021 kan pensionsalderen eksempelvis beregnes således, hvis pensionsfasen skal have samme relative længde som nu:

$$\text{Pensionsalder år } t = \frac{\text{Middellevetid i år } t}{\text{Middellevetiden i år } t \text{ ved pensionsgrænsen}} = \frac{\text{restlevetiden for 65-årige i 2001}}{65 + \text{restlevetiden for 65-årige i 2001}} \cdot \frac{\text{Middellevetiden i år } t}{\text{Middellevetiden i år } t \text{ ved pensionsgrænsen}}$$

$t = 2021, 2022$  osv.

Antages det, at middellevetiden i år 2021 er 85 år ved pensionsgrænsen, kan pensionsalderen i 2021 beregnes til 67,38 år, da restlevetiden for 65-årige var ca. 17 år i 2001. Nedrundet bliver det til 67 år og 4 mdr. og så fremdeles.<sup>6</sup>

I tabel 3 har jeg sammenfattet, hvad jeg finder det meget væsentligt også at få oplyst, hvis vi skal have en seriøs velfærdsdebat.

6. I Velfærdskommissionen (2005) beregnes stigningen i pensionsalderen ud fra stigningen i middellevetiden. Det relevante er dog, som gjort her, at tage udgangspunkt i restlevetiden ved pensionsgrænsen.

*Tabel 4. Krav til finanspolitisk holdbarhed belyst ved tre instrumenter af tre institutioner.*

	Velfærds- kommissionen	Det økonomiske Råds formandskab <sup>(b)</sup>	Finansministeriet
Stigning i bundskatteprocenten, procentpoint	8,7	4,2	3,5
Udgiftsbesparelser, pct. af BNP <sup>(a)</sup>	3,7	1,9	1,5
Beskæftigelsesstigning, 1000 personer	280	140	90

*Noter:* (a) Via kollektivt offentligt forbrug. (b) Indgrebene foretages som *permanente* engangsændringer fra 2007, idet beskæftigelsesstigningen dog indføres i årene 2007-2011.

*Kilder:* AE-rådet (2004a, s. 40).

Kort sagt. En ordentlig hensyntagen til terminalproblemet vil efter min opfattelse fordrer:

- at der tages stilling til, hvilket offentligt formuebegreb der er relevant i overdragelsessituationen mellem generationer,
- at der simuleres på en sådan måde, at der i lyset af ikke mindst stigende levealder sker en fair fordeling af finansieringen af den offentlige sektor mellem generationerne. Dette kunne foretages ved at definere, hvordan overdragelsessituationen skulle være i f.eks. 2041, 2061, 2076 og 2100, jf. anmærkningen til tabel 3.

Og der bør givetvis også inddrages andre momenter i en diskussion af terminalproblemet: f.eks. usikkerheden om stigningen i levetiden.

#### 4. Velfærdskommissionens resultater

Velfærdskommissionens centrale resultater i relation til finanspolitisk holdbarhed er resumeret i tabel 4, hvor også de tilsvarende beregningsresultater fra Det økonomiske Råds formandskab og Finansministeriet er anført.

Kravene til finanspolitisk holdbarhed er illustreret ved tre former for indgreb i tabel 4. Velfærdskommissionen forudsætter i alle tre tilfælde, at indgrebene indføres lineært fra 2011 til 2021. Det fremgår eksempelvis af tabel 4, at bundskatteprocenten fra og med 2021 *permanent* skal være 8,7 procentpoint højere end i udgangssituationen, hvor den er 5,5 procent, for at sikre finanspolitisk holdbarhed.

Det fremgår endvidere af tabel 4, at Velfærdskommissionen anslår det finanspoliti-

Tabel 5. Korrektioner af Velfærdskommissionens holdbarhedsberegninger.

	Nødvendig stigning i bundskatteprocenten, pct.point	Nødvendige off. besparelser i pct. af BNP <sup>(a)</sup>
Velfærdskommissionens beregninger	8,7	3,7
<i>Korrektioner:</i>		
Højere vækstkorrigeret realrente (0,65 → 1,4 pct. p.a.)	-2,8	-1,2
Delvist restlevetidsbestemte udgifter til individuel off. service	-2,1	-0,9
Mere retvisende fremskrivning af befolkningens uddannelsesniveau	-0,9	-0,4
Højere erhvervsfrekvens for efterkommere af indvandrere fra mindre udviklede lande	-0,8	-0,4
Korrigeret holdbarhedsproblem	2,1	0,8

Note: (a) Besparelser via kollektivt offentligt forbrug.

Kilde: AE-rådet (2004a, s. 59).

ske holdbarhedsproblem til at være ca. dobbelt så stort som de to andre aktører. Årsagerne til disse forskelle er nærmere forklaret i AE-rådet (2004a, s. 40-59).<sup>7</sup>

Tabel 5 giver en sammenfattende oversigt over de korrektioner af Velfærdskommissionens centrale beregningsresultater, som min gennemgang i Sørensen (2005) gav anledning til.

Efter korrektion for overgang til: (1) højere vækstkorrigeret realrente, (2) delvist restlevetidsbestemte udgifter til individuelle offentlige udgifter, (3) mere retvisende fremskrivning af befolkningens uddannelsesniveau og (4) højere erhvervsfrekvens for efterkommere af indvandrere fra mindre udviklede lande fremkom der et finanspolitisk holdbarhedsproblem svarende til enten en permanent stigning i bundskatteprocenten på 2,1 procentpoint eller til permanent lavere offentlige udgifter svarende til 0,8 pct. af BNP.

Her skal jeg komme med to uforbeholdne indrømmelser til Velfærdskommissionen. Grundet simultaniteten mellem de enkelte korrektioner, burde jeg ikke have anført, at det korrigerede holdbarhedsproblem kunne beregnes ved at trække disse fire særskilt beregnede korrektionsposter fra. Grundet en betydelig simultanitet halveres holdbarhedsproblemet alene som følge af disse fire korrektioner, jf. oplysninger fra Velfærds-

7. Se også Finansministeriet (2004) s. 180-85.

kommissionen. For det andet frafalder jeg kritikken af den anvendte størrelsesordenen af den vækstkorrigerede realrente, idet overgangen til et lavinflationsregime, som det også er grund til fortsat at forvente, gør, at historiske – og højere tal – for den vækstkorrigerede realrente ikke kan være retningsgivende. Til gengæld betyder dette blot, at terminalproblemet må tillægges større vægt. For tænkt, hvis der måtte regnes med en negativ vækstkorrigeret realrente i grundforløbet!

Der er givetvis også områder, hvor Velfærdskommissionen har undervurderet kravene til de offentlige finanser. Et sådant område kan være de kollektive serviceudgifter, hvor Velfærdskommissionen har antaget, at disse følger udviklingen i nominel BNP, se nærmere herom i Sørensen (2005).

### 5. Konklusion

Min væsentligste kritik af Velfærdskommissionens hidtidige arbejde er som anført knyttet til behandlingen af terminalproblemet. I tilknytning hertil burde der som anført i afsnit 3 være gennemført beregninger, hvor bl.a. pensionsalderen var forhøjet i takt med stigningen i restlevetiden ved pensionsgrænsen.

I Torben M. Andersens og Lars Haagen Pedersens indlæg er dette afvist med følgende formulering:

»Indikatoren beregnet ved bundskattesatsen har den fordel, at den er meget konkret. I forhold til den økonomisk politiske debat er det imidlertid vigtigt at understrege, at forskellige tilpasningsmuligheder kan bringes i anvendelse. For at illustrere dette kan der også beregnes en indikator udmålt på f.eks. offentlige udgifter eller privat beskæftigelse. Velfærdskommissionen (2004) ser på alle tre typer indikatorer. For de to sidstnævnte gælder imidlertid, at de er mindre tæt knyttet til et konkret økonomisk politisk instrument og derfor sværere at fortolke.«

Heri kan jeg ikke være enig fsva. privat beskæftigelse. Tilbagetrækningsaldre er endog meget konkrete, især med et udbudsorienteret udgangspunkt, som er det relevante i meget lange fremskrivninger.

Og her er de to forfattere nok også i modstrid med Velfærdskommissionen (2005), hvor det anføres:

»I det følgende betragtes to måder til sikring af senere tilbagetrækning. De illustrerer begge en størrelsesorden på lige fod med de beregnede skattestigninger præsenteret ovenfor. En sammenligning af ændret tilbagetrækning og skatteberegningerne ovenfor understreger samtidigt, at der skal tages stilling til et væsentligt politisk prioriteringsspørgsmål: Skal tilpasningen ske ved at de erhvervsaktive betaler mere i skat



mens de er på arbejdsmarkedet, eller skal de erhvervsaktive blive længere tid på arbejdsmarkedet i takt med den stigende levealder?»

Dette kunne jeg ikke have udtrykt bedre.

Velfærdskommissionen (2005) afviser, at de har overvurderet holdbarhedsproblemet. Her kan jeg kun komme ind på den sprængning af den grundlæggende forudsætning om uændret offentligt serviceniveau, som sker ved at inddrage det såkaldte velstandsdilemma i relation til individuelle sundheds- og plejeudgifter. Det indrømmes, at det havde været bedre at anvende restlevetiden frem for alderen til at prognosticere disse udgifter, når netop levetiden stiger væsentligt, hvorfor vel også helbredstilstanden ved en given alder forbedres. Men ved at inddrage de serviceforbedringer, der kan blive en følge af nye og bedre behandlingsformer, kan Velfærdskommissionen selvfølgelig komme frem til – som tidligere formanden for Velfærdskommissionen – at de givetvis undervurderer de fremtidige udgifter til individuelle sundheds- og plejeudgifter. At de så sprænger forudsætningerne for deres egne analyser nævnes ikke. Der er vel ikke noget mærkeligt i, at der skal betales for ekstra goder, som Velfærdskommissionen (2005) da også selv anfører via bl.a. henvisning til en »grundpakke« kombineret med »tilkøb«. Hertil kommer, at Velfærdskommissionen (2005) anvender kritiske forudsætninger vedr. de fremtidige udgifter til individuelle sundheds- og plejeudgifter. Dette fremgår bl.a. af et omfattende EU-forskningsprojekt, hvor forholdene i Danmark også er inddraget. Et hovedresultat er her, at bedre helbred stort set opvejer virkninger af stigende alder i forhold til disse udgifter, jf. Westerhout og Pellikaan (2005).

Så det er ikke altid let at holde fast i, at oprindelige tilgangsvinkler og analyseresultater fastholdes. Også her må der prioriteres!

#### Litteratur

- AE-rådet. 2004a. *Økonomiske Tendenser* 2004. København.
- AE-rådet. 2004b. *Tre spørgsmål til Velfærdskommissionen*. København, 21. juni 2004.
- Arendt, J. N., E. B. Hansen og T. K. Thrane. 2002. Kan forbedringer i ældres levevilkår lette forsørgerbyrden? *Nationaløkonomiske Tidsskrift* 2002:140:3, s. 289-311.
- Broda, C. og D. E. Weinstein. 2004. Happy news from the Dismal Science: Reassessing the Japanese Fiscal Policy and Sustainability. *NBER Working Paper Series nr. 10988*. Cambridge, MA. December 2004.
- Economic Policy Committee. 2001. *Budgetary Challenges Posed by Ageing Populations: The Impact on Public Spending on Pensions, Health and Long-Term Care for the Elderly and Possible Indicators of the*
- Long-Term Sustainability of Public Finances*. EPC/ECFIN/655/01-EN final. October.
- Finansministeriet. 2004. *Finansredegørelse 2004*. København.
- Jacobzone, S., E. Cambois og J. M. Robine. 2000. Is the health of older persons in OECD countries improving fast enough to compensate for population ageing? *OECD Economic Studies* 2000:30, s. 149-90.
- Madsen, J. og N. Serup-Hansen. 2000. Alder, død og sundhedsomkostninger. Det Samfundsvidenskabelige Fakultet, Syddansk Universitet. *Health Economics Papers* 2000:4.
- Nielsen, S. and O. Risager. 2001. Stock Returns and Bond Yields in Denmark 1922-

1999. *Scandinavian Economic History Review* 2001:49:1, s. 63-82.
- Seshamani, M. og A. Gray. 2004a. Ageing and health-care expenditure: the red herring argument revisited. *Health Economics* 2004:13, s. 303-14.
- Seshamani, M. og A. Gray. 2004b. A longitudinal study of the effects of age and time to death on hospital costs. *Health Economics* 2004:23, s. 217-35.
- Sørensen, C. 2005. Velfærdskommissionens analyse af finanspolitisk holdbarhed – er den holdbar? Institut for Regnskab og Finansiering. *Working Paper No. 01/2005*. Odense.
- Undervisningsministeriet. 2003. *Tal der taler*. København.
- Velfærdskommissionen. 2004. *Fremtidens velfærd kommer ikke af sig selv*. Endelig udgave. København, maj 2004.
- Velfærdskommissionen. 2005. *Befolkningsudvikling, velstanddilemma og makroøkonomiske strategier. Sammenfatning af teknisk analyserapport*. København, september 2005.
- Westerhout, E. og F. Pellikaan. 2004. *Alternative scenarios for health, life expectancy and social expenditure*. Centraal Planbureau, AGIR Workshop London (Work in progress).
- Westerhout, E. og F. Pellikaan. 2005. *Can we afford to live longer in better health?* CPB Document No. 85, june 2005.
- Zweifel, P., S. Felder og M. Meier. 1999. Ageing of Population and Health Care Expenditure: a red herring? *Health Economics* 1999:8, s. 485-96.
- Økonomiske Råd, Det. Formandskabet. 2004. *Dansk Økonomi, forår 2004*. København.
- Økonomiske Råd, Det. Sekretariatet. 2004. *Finanspolitikens holdbarhed*. København, 1. juli 2004.

## Makroøkonomisk analyse, finanspolitisk holdbarhed og DREAM-modellen: en umage cocktail

Jesper Jespersen

Institut for Samfundsvidenskab, Roskilde Universitetscenter, E-mail: [jesperj@ruc.dk](mailto:jesperj@ruc.dk)

Velfærdskommissionen (VK) har i sit hidtidige arbejde koncentreret sig om at gennemføre fremskrivninger af udviklingen i nogle få centrale makroøkonomiske variable under antagelse af, at både privat og offentligt forbrug øges med den underliggende trend i produktiviteten. Hovedresultaterne fra denne fremskrivning frem til år 2061 er gengivet i kapitel 12 i VK (2004a). Ved brug af den generelle ligevægtsmodel DREAM er spørgsmålet om *finanspolitisk holdbarhed* efterfølgende blevet vurderet. Den fremlagte beregning viser, at konklusionerne ikke mindst vedr. finanspolitisk holdbarhed er meget følsom over for antagelser om middellevetid, erhvervsfrekvens og den trendmæssige opskrivning af væksten i offentlige udgifter og indtægter.

I VK (2004a) nås det for så vidt trivielle resultat, at de offentlige udgifter på et tidspunkt ude i fremtiden vil overhale indtægterne, når de personrelaterede udgifter og indtægter opskrives med den underliggende væksttrend, og der samtidig sker en markant demografisk forskydning fra erhvervsaktive til pensionistgruppen. Denne beregning minder om Thomas Robert Malthus' påvisning af, at fødevareproduktionen set i et 100 årigt perspektiv ikke var holdbar – alt andet lige.

VK når frem til, at finanspolitisk holdbarhed vil forudsætte en tilpasning af de offentlige nettoindtægter svarende til en stigning i bundskatten i perioden 2011-21 på i alt 8,5 procentpoint under de angivne forudsætninger.

Med udgangspunkt i disse beregninger gennemført på den generelle ligevægtsmodel DREAM formulerede VK følgende utvetydige konklusion: *»vore hovedkonklusioner er klare: Danmark er ikke på holdbar kurs. På bare 10-20 års sigt vil vi opleve et stort misforhold mellem indtægter og udgifter i den offentlige sektor. Vi er nødt til at justere på velfærdssamfundet«*. (2004b) s.1-3. I Andersen, Jensen og Pedersen (2004) s. 3, der bygger på den samme analyse, er konklusionen formuleret således: *»The main findings of the paper are that maintaining current welfare arrangements leaves a sub-*

---

Jeg har valgt at basere denne kommentar ikke alene på Andersen og Pedersen (2005), men også at inddrage Velfærdskommissionens øvrige analyser: VK (2004a), VK (2004b), Andersen, Jensen og Pedersen (2004) samt af DREAM-modellens struktur beskrevet i bl.a. Pedersen, Stephensen og Trier, DREAM (1999). Mine synspunkter er uddybet i bl.a. Jespersen (2004) og Jespersen (2005).

*stantial problem of fiscal sustainability*«. Der er således ingen tvivl om, at VK tillægger begrebet finanspolitisk holdbarhed en særdeles stor analytisk betydning.<sup>1</sup>

Andersen og Pedersen (2005) benytter således Velfærdskommissionens beregninger, gengivet (med mindre justeringer) i tabel 1 til besvarelsen af følgende tre spørgsmål: 1. »spørgsmålet om finansiel holdbarhed« ... 2. »størrelsen af problemet [angivet ved holdbarhedsindikator]«... og 3. »tilrettelæggelsen af den økonomiske politik [i form af et engangsløft i skattesatsen, en Pay-As-You-Go strategi (PAYG) eller en mellemform]«, s. 4.

Jeg vil på grund af den begrænsede plads kun diskutere følgende tre forhold:<sup>2</sup>

- Er begrebet *finanspolitisk holdbarhed* et relevant kriterium for en dynamisk analyse af velfærdssamfundets langsigtede finansierings- og tilpasningsbehov?
- Hvorledes kan de såkaldte *holdbarhedsindikatorer* beregnet inden for rammerne af en generel ligevægtsmodel (in casu DREAM-modellen) fortolkes?
- Giver analysen af de økonomisk politiske strategier analyseret ved hjælp af DREAM-modellen en dækkende beskrivelse af den dynamiske tilpasning?

#### **Finanspolitisk holdbarhed – et relevant begreb?**

Finanspolitisk holdbarhed er i sin oprindelse et simpelt driftsøkonomisk begreb, som enhver virksomhedsleder og bankbestyrer vil nikke genkendende til. Det er spørgsmålet, om der er balance mellem virksomhedens fremtidige indtægter og udgifter. Spørgsmålet besvares ved at beregne nutidsværdien af alle kendte og forventede indtægter og udgifter tilbagediskonteret med markedsrenten. Begrebet er enkelt og operationelt på virksomhedsniveau, hvor de samfundsøkonomiske parameter kan tages som givne – i hvert fald som uafhængige af, hvilke beslutninger virksomheden træffer.

Benyttes begrebet for hele den offentlige sektor, så gør det sig derimod ikke gældende, at den samfundsøkonomiske udvikling er uafhængig af den førte politik. Ved ukritisk at overføre begrebet finansiel holdbarhed fra mikroøkonomi til makroøkonomi begås den såkaldte *atomistiske fejlslutning*, jfr. f.eks. Jespersen (2001).

Der er generelt en betydelig risiko for, at den atomistiske fejlslutning opstår, når der

1. Det er et begreb, der kun optræder sporadisk i den teoretiske litteratur. Man vil således lede forgæves efter det i f.eks. den netop udkomne lærebog Sørensen og Jacobsen (2005), derimod er det (indirekte) nævnt i Snowdon og Vane (2002), s. 246 som en del af »generations-regnskabs approach'et« udviklet under antagelse af rationelle forventninger og generel ligevægt.

2. Jeg antager, at Christen Sørensen i sin kommentar vil diskutere de beregningsmæssige forudsætningsrealisme. Sammenlignet med bl.a. DØR (2002), hvor Torben M. Andersen i øvrigt også var formand, er forudsætningerne blevet gjort væsentligt mere pessimistiske. I 2002-rapporten fra DØR var holdbarhedsindikatoren kun 1/3 af størrelsen i VK (2004a), og i Finansredøgørelsen (2002) er holdbarhedsindikatoren 0.

drages makroøkonomiske konklusioner på basis af en generel ligevægtsmodel baseret på repræsentativ mikroøkonomisk agent-teori, market clearing/ligevægt og rationelle forventninger.<sup>3</sup> Velfærdskommissionen benytter sig af finansministeriets DREAM-model til beregning af finanspolitisk holdbarhed.

I en generel ligevægtsmodel med permanent market clearing/ligevægt – herunder velfungerende kapitalmarkeder, har den offentlige sektor per definition ikke noget finansieringsproblem. Et underskud på de offentlige budgetter må nødvendigvis modsvares af et overskud i den private sektor (i en lukket økonomi). I DREAM-modellen (en åben økonomi) fordeles finansieringen af det offentlige budgetunderskud sig mellem den private sektor og udlandet afhængigt af den private sektors opsparingsadfærd. I de beregninger VK har fremlagt (2004a) er der tilnærmelsesvist en en-til-en sammenhæng mellem ændringer i den offentlige sektors saldo og saldoen på betalingsbalancens løbende poster, jfr. tabel 12.2-12.6 (hvor »uændret politik« sammenlignes med »skattestigning«). Det økonomisk-politiske problem i grundforløbet er således *ikke* finansiering af det offentlige underskud, men derimod et voksende betalingsbalanceunderskud og øget *udenlandsk* gældsætning, der bringer mindelser om 1980erne tilbage – hvis altså DREAM-modellen var et retvisende spejlbillede af virkeligheden.

En generel ligevægtsmodel er en analytisk konstruktion ofte uden eksplicit modellering af den finansielle sektor. En ikke-skattefinansieret finanspolitisk ekspansion overfører reale ressourcer fra udlandet til indlandet, idet beskæftigelsen er bundet af arbejdsmarkedsstrukturer (og evt. en eksogent fastsat løn). Den eneste arbejdsmarkedseffekt, som DREAM-modellen registrerer af en skattestigning, er den antagne reduktion af arbejdsudbudet. Bortset herfra betyder øgede skatter i en generel ligevægtsmodel blot en reduktion af betalingsbalanceunderskuddet. Inden for modellens rammer er det ikke muligt at analysere den dynamiske flow-tilpasning i den private sektor af en ændret finanspolitik.

Hvis analysemodellen derimod blev frigjort fra forudsætningen om generel flow-ligevægt ville ændringer i den effektive efterspørgsel kunne analyseres i form af bl.a. ændring i den registrerede arbejdsløshed. Det ville være naturligt at antage, at erhvervsfrekvensen (også) for de svagere grupper vil øges, hvis presset på arbejdsmarkedet stiger, hvorved forringelsen af betalingsbalancen dæmpes. På den anden side er der en risiko for, at en finanspolitisk ekspansion vil føre til en stigning i inflationen og en efterfølgende forringet konkurrenceevne. Spændinger og ubalancer som en generel ligevægtsmodel ofte fejlagnosticerer som et spørgsmål om finanspolitisk holdbarhed og brud på den intertemporale budgetrestriktion for den offentlige sektor. Den egentlige makroøkonomiske ubalance skal derimod snarere søges på betalingsbalancen eller i form af et øget inflationært pres. En finanspolitisk stramning i form af en

---

3. En problemstilling, som jeg diskuterede med Toke Ward Pedersen tilbage i 1999, jfr. Jespersen (1999).

balancering af budgettet (en ikke-negativ nutidsværdi) er i dette perspektiv ikke nødvendigvis den eneste endelige optimale policy-option. Alt andet forbliver nemlig ikke uændret i den virkelige økonomi, når der føres diskretionær økonomisk politik. Her minder VKs vækst-scenarie på mange måder om situationen i 1960'erne, hvor arbejdsløsheden faldt samtidigt med, at betalingsbalanceunderskuddet blev cementeret, (hvilket kunne ske igen, når bl.a. oliefeltene i Nordsøen tømmes om 20-30 år) og inflationen begyndte at kravle op. Det er den egentlige makroøkonomiske udfordring, som VK burde anvise veje til at forebygge.

Det er på baggrund af ovenstående argumentation et åbent spørgsmål, om Andersen og Pedersen (2005) s. 21 overhovedet har gennemført en egentlig analyse af spørgsmålet om *finanspolitikens holdbarhed*, når de skriver, at [det fordrer] *en markedstest, hvor det vurderes, om den udvikling i offentlige udgifter og indtægter, der følger af den fastlagte politik og den økonomiske udvikling, er forenelig med solvens, s.* Ved at benytte DREAM-modellen, hvor der er forudsat generel ligevægt både før og efter den finanspolitiske ændring, sker der netop *ikke* en test af reaktionen på hverken arbejds-, finans- eller valutamarkedene af den markedsmæssige dynamik. Det er på disse markeder, at den makroøkonomiske ubalance kan registreres, som en konsekvens af ændringer i den indenlandske investerings-/opsparingsadfærd (forårsaget af den økonomiske politik) og i den internationale konkurrenceevne.

### **Følsomhedsberegninger**

Det er velkendt at beregningsresultater er bestemt af (1) *beregningsmodellen* og (2) *beregningsforudsætningerne*. Herved kommer DREAM-modellen til at spille en afgørende rolle for de følsomhedsberegninger som VK har gennemført og som Andersen og Pedersen præsenterer i tabel 1. Anvendelsen af en generel ligevægtsmodel er en medvirkende forklaring på flere – i forhold til virkeligheden – mildt sagt overraskende analyseresultater, der dog kan forklares som en konsekvens af en *atomistisk fejlslutning*. Det fejlræsonnement, der som nævnt begås, når generaliseret mikroteori forveksles med egentlig makroteori, jfr. VK (2004a) og Andersen og Pedersen. (2005), tabel 1, s. 28. hvoraf jeg blot her skal diskutere de to, der umiddelbart springer mest i øjnene: (1) *Højere rente betyder en forbedring af den finanspolitiske holdbarhed* og (2) *forøget vækst forværrer den finanspolitiske holdbarhed*.

Inden for rammerne af en generel ligevægtsmodel, antages det, at den samlede produktion og beskæftigelse er bestemt fra udbudssiden – ikke mindst hvis de mikroøkonomiske aktører har rationelle forventninger. Den offentlige sektor analyseres som en erhvervssektor, der dog overvejende finansieres ved skatter. Ændret aktivitet i den offentlige sektor påvirker ikke den samlede produktion og beskæftigelse. Da den offentlige sektor i dag har en positiv nettoformue, vil en højere realrente have den økonomiske

konsekvens i DREAM-modellen, at den forbedrer statsfinanserne. En højere realrente betyder derfor større finanspolitisk holdbarhed. En stigning i renten med 2-3 procentpoint vil gøre finanspolitikken holdbar, jfr. tabel 1. Men heller ikke i denne analyse er der gennemført en markedstest: (1) Hvad betyder højere rente for udviklingen i de private investeringer og det private forbrug? (2) Hvorledes påvirker et højere renteniveau husholdningernes disponible indkomster? Er det realistisk at antage, at husholdningerne kender den nye langsigts-ligevægt og tør handle herudfra? Historiske erfaringer tilsiger, og analyser med andre modeltyper end generelle ligevægtsmodeller når til andre resultater: at en højere rente lige så vel kan føre til en svækket konjunkturudvikling, højere arbejdsløshed og forringede offentlige finanser – naturligvis afhængig af, hvor stor den offentlige nettoformue er. Den analytiske fejl, som VK synes at begå, er at forveksle statskassen med en privat pensionskasse. For en sådan vil det højere realrente gøre det lettere at finansiere de allerede afgivne pensionstilsagn, idet pensionskassens aktiviteter ikke påvirker det makroøkonomiske forløb.

Det forekommer tilsvarende overraskende, at øget vækst ligefrem skulle forøge den offentlige sektors finansieringsproblem. Atter findes (en del af) forklaringen i DREAM-modellen, hvor arbejdsløsheden er upåvirket af den højere vækst. Og den aflastning af de offentlige budgetter, som en reduceret arbejdsløshed ville indebære, indgår ikke i beregningen. Derimod antager VK, at de offentlige udgifter, lige fra dagpengesatser til indkøb af militært udstyr, forøges i samme takt som væksten i BNP. Øges væksten i BNP, så øges væksten i de offentlige udgifter per person tilsvarende. Under disse (ekstreme) forudsætninger vil øget vækst naturligvis ikke kunne aflaste de offentlige finanser. Tværtimod er de forhåndsprogrammerede til at blive forringet i og med, at enhver borger er forudsat at være en nettobelastning for de offentlige kasser. Forringelsen forstærkes yderligere af, at det vækstkorrigerede afkast i den offentlige »pensionskasse« falder, når væksten stiger.

Disse mildt sagt kontra-intuitive resultater springer umiddelbart i øjnene. Vendes problemstillingen på hovedet, så vil et fald i væksten fra 2 pct. til 1 pct. f.eks. forårsaget af en stigning i renten på 3 procentpoint ifølge DREAM-modellen helt kunne fjerne det finanspolitiske holdbarhedsproblem, jfr. tabel 1.<sup>4</sup>

### **Diskussion af langsigtede økonomisk-politiske strategier**

En stor del af Andersen og Pedersen (2005) er reserveret en afvejning af fordelene ved hhv. en strategi baseret på en »én-gangsskattestigning« og en »Pay-As-You-Go (PAYG) finansiering«.

---

4. Hvis det derimod blev forudsat, at en øget vækst i produktiviteten på 1/2 pct. p.a. ikke automatisk blev omsat i en tilsvarende øget vækst i de offentlige udgifter, så ville holdbarhedsproblemet ifølge modellen opløse sig af sig selv, jfr. VK (2004a, tabel 12.1) – selvom arbejdsløsheden forbliver uændret.

Tager vi skattestigningen først, så analyseres den i DREAM-modellen, hvor der *ingen* afledte virkninger er på det indenlandske produktionsniveau og beskæftigelse (bortset fra den postulerede forvriddende effekt på arbejdsudbudet af højere skat). Det af skatten mindskede privatforbrug bliver substitueret af en tilsvarende stigning i nettoeksporten, jfr. (2004a, tabel 12.5). I DREAM-modellen er afsætningen på verdensmarkedet ubegrænset.<sup>5</sup> Arbejdskraften vandrer ubesværet fra indenlandsk produktion til eksportsektoren! På tilsvarende måde gælder det i VK (2004 a), at de offentlige besparelser, der kan sikre finanspolitisk holdbarhed, ikke fører til en stigning i ledigheden, men derimod til en øget nettoeksport. Disse resultater antager en hidtil uset fleksibilitet på arbejdsmarkedet, samtidig med at erhvevsfrekvenserne forbliver uændrede. Disse beregninger synes at fortælle mere om DREAM-modellens specielle egenskaber, end om realistiske reaktionsmønstre i dansk økonomi.

PAYG-strategien bygger på den forudsætning, at hvis der er balance på den offentlige sektors budget år for år, så vil der automatisk være finanspolitisk holdbarhed. Det er jo nærmest definitionen på dette begreb. Formålet med en sådan løbende balance skulle angiveligt være at sikre den makroøkonomiske stabilitet, idet DREAM-modellens generelle ligevægt automatisk sikrer ligevægt i den private sektor (opsparing = investering) og på arbejdsmarkedet (her defineret som 5 pct.'s arbejdsløshed). Sammenlignet med én-gangsskatten har PAYG den egenskab, der er illustreret i papirets figur 12, at skatteforhøjelsen kommer mere glidende i takt med, at de offentlige udgifter øges, hvilket vil være mindre belastende for den nuværende generation, (og samtidig skabe et mindre kortsigtet tilpasningspres).

Men også her gælder det, at analysens relevans er begrænset af antagelsen om generel ligevægt og fuldstændig arbejdsmarkedsfleksibilitet.

### Konklusion

Velfærdskommissionen (2004a, 2004b) såvel som Andersen og Pedersen (2005) analyserer fremtidens udfordringer til de offentlig finanser inden for rammerne af den generelle ligevægtsmodel DREAM. Inden for modellens rammer defineres begrebet *finanspolitisk holdbarhed*. Begrebets relevans for dansk økonomi diskuteres ikke.

Tillad mig bl.a. som en konsekvens heraf afslutningsvist at citere fra papirets konklusion (s. 42): »*Diskussion om behovet for reformer er ikke et spørgsmål om i detaljer at planlægge den økonomiske politik i en lang årrække frem. Det har ingen mening, dertil er usikkerheden og omfanget af strukturelle ændringer m.m. for stor. Derimod er det nødvendigt at indrette den økonomiske politik ud fra nogle pejlemærker og sikre troværdighed om de ordninger, der er etableret*«. Heri er jeg fuldstændig enig, hvorfor

---

5. I privat korrespondance med Lars Haagen Pedersen er der åbnet mulighed for, at priselasticiteten i nettoeksporten ikke er uendelig, men -5. I så fald vil overflytningen kræve en (mindre) reallønsnedgang.



hele diskussionen om en permanent indkomstskattestigning på 8,7 procentpoint sat over for et krav om løbende balance på de offentlige budgetter analyseret i DREAM-modellens optik virker ligeså verdensfjern, som den er grundigt diskuteret.

Målet med Velfærdskommissionens arbejde burde således snarere have været en realistisk analyse af de balanceproblemer, som dansk økonomi står over for i de kommende 10-15 år, set i lyset af de demografiske og globale udfordringer. Intet politisk parti vil eller kan i dag forpligtige sig til en mere langsigtet politik – bortset fra nogle få investeringsbeslutninger (miljø, infrastruktur og byplanlægning), der rækker ud over en tidshorisont med den længde. Det er i realiteten meget få politiske beslutninger, der har en karakter, der i et længere perspektiv gør dem irreversible, hvilket ydermere skal sammenholdes med den (formentlig) fortsat stigende materielle velstand.

#### Litteratur

- Andersen, T. M. og L. H. Pedersen. 2005. *Demografi, velstandsdilemma og makro-økonomiske strategier*, Nationaløkonomisk Tidsskrift 143, s. 189-229.
- Andersen, T. M., S. E. Hougaard Jensen og L. H. Pedersen. 2004. *The Welfare State and Strategies towards Fiscal Sustainability in Denmark*, paper presented at CESifo-LBI conference, Munich, 22-23 October.
- Finansministeriets hjemmeside ([www.fm.dk](http://www.fm.dk)) med henvisninger til en nærmere beskrivelse af DREAM-modellen, der dog også har sin egen hjemmeside ([www.Dream-model.dk](http://www.Dream-model.dk)).
- Jespersen, J. 1999. Hvor solidt er CGE-modellernes teoretiske og empiriske grundlag? – en kommentar til en ny modeltradition. *Nationaløkonomisk Tidsskrift* 137, s. 213-17 og s. 220-21.
- Jespersen, J. 2001. Den atomistiske fejlslutning, kapitel 6 i Kristensen, J. E. og C. Fenger-Grøn, *Kritik af den økonomiske fornuft*, København: Hans Reitzels Forlag.
- Pedersen, L. H., P. Stephensen og P. Trier. 1999. A CGE Analysis of the Danish Ageing Problem, Working Paper, DREAM, København ([www.dreammodel.dk](http://www.dreammodel.dk)).
- Petersen, J. H. og K. Petersen, red. 13 udfordringer til den danske velfærdsstat, Odense: Syddansk Universitetsforlag, 2004.
- Snowdon, B. og H. Vane, red. 2002. *An Encyclopedia of Macroeconomics*, Cheltenham: Edward Elgar.
- Sørensen, P. B. og H. J. Whitta-Jacobsen. 2005. *Introducing Advanced Macroeconomics: Growth and Business Cycles*, London. McGraw-Hill Education.
- Velfærdskommissionen. 2004a. *Fremtidens velfærd kommer ikke af sig selv* – analyse-rapport, København.
- Velfærdskommissionen. 2004b. *Fremtidens velfærd kommer ikke af sig selv* – debat-pjece, København.
- Økonomiske Råd, Det. Formandskabet. 2002. *Dansk økonomi – efteråret 2002*, København.

## Velfærdskommissionens makroøkonomiske analyser og anbefalinger

Claus Vastrup

Institut for Økonomi, Aarhus Universitet, E-mail: [cvastrup@econ.au.dk](mailto:cvastrup@econ.au.dk)

### Baggrund og kommissorium

Begrundelsen for nedsættelsen af Velfærdskommissionen var de ændringer, som især den forventede stigning i de ældres andel af den samlede befolkning vil have for de offentlige finanser på længere sigt. Det følgende er en diskussion af nogle elementer fra kommissionens arbejde med disse makroøkonomiske balanceproblemer, fortrinsvis som de er fremstillet af Andersen og Pedersen (2005).

Både kommissionen og de to forfattere tager udgangspunkt i det forhold, at en stigning i den andel af befolkningen, der som ældre modtager offentlige pensioner og andre ydelser finansieret ved samtidig og uændret skatteopkrævning, vil føre til en forringelse af de offentlige finanser. En fortsat stigning i de ældres andel af befolkningen er derfor ensbetydende med, at der før eller siden vil opstå et offentligt budgetunderskud. Dette vil kræve finansiering enten i form af en stigning i beskattningen af den øvrige befolkning, en sænkning af de offentlige pensioner eller en forskydning i fordelingen af befolkningen mellem modtagere af offentlig pension og den øvrige befolkning. Det sidste tænkes almindeligvis at finde sted gennem de erhvervsaktives senere overgang til pension. En fjerde mulig finansieringskilde er en reduktion af det offentlige forbrug og andre overførselsindkomster. Den sidste kilde er en uholdbar forøgelse af den offentlige gæld eller anvendelse af seddelpressen og dermed inflation, hvilke jeg vil lade ude af betragtning.

Kommissionen har ligesom andre fremlagt beregninger over den fremtidige udvikling i de offentlige budgetter. Som følge af et gunstigt udgangspunkt og en beregningsmæssig forudsætning om fuld beskæftigelse forventes de offentlige budgetunderskud først at opstå om 15-20 år, men vil herefter være permanente og fortsat stigende. Til finansiering af disse underskud opstilles som yderpunkter to strategier. Den ene er en såkaldt opsparingsstrategi med hurtige forhøjelser af skattesatserne til et holdbart og derfor et konstant, men også et højt niveau, der kan finansiere en efterfølgende, stigende offentlig pensionsbyrde med uændrede tilbagetræknings- og pensionsregler. Den anden er en såkaldt pay-as-you-go strategi, hvor senere og gradvise forøgelser af skattesatserne alene sikrer løbende balance på de offentlige finanser. Denne strategi sikrer holdbar finansiering, men fører ikke til offentlig opsparing. Med uændrede tilbage-

træknings- og pensionsregler fører pay-as-you-go strategien i det lange løb derimod til væsentlige skattestigninger.

De to strategier (og det som ligger herimellem) er således begge ensbetydende med skattestigninger, således at den afgørende forskel vedrører tidspunktet for den forudsatte skatteopkrævning og dermed spørgsmålet om (størrelsen af den) offentlig opsparing. På trods af kommissoriet og selvom andre former for finansiering af de fremtidige budgetunderskud vil føre til andre værdier af arbejdsudbud, produktion mv. har skatteforhøjelser og offentlig opsparing (indtil videre) været hovedemnet for Velfærdskommissionens overvejelser om den makroøkonomiske strategi. Det er en indskrænkning i behandlingen af emnet, som går igen hos Andersen og Pedersen (2005).

Jeg vil i det følgende først og kort diskutere nogle spørgsmål i forhold til opgørelsen af de fremtidige budgetunderskud og dernæst og lidt længere diskutere nogle forhold vedrørende Velfærdskommissionens analyser og forslag til strategi for offentlig opsparing og eliminering af fremtidige budgetunderskud. Dernæst konkluderes om forslag og udeladte emner.

### **De offentlige budgetunderskud**

Ændringer i skatter og offentlige udgifter kan begrundes såvel med (ændringer i) politiske præferencer som med forskellige former for økonomisk nødvendighed. Velfærdskommissionen har ved at koncentrere sine beregninger og især sin argumentation om manglende holdbarhed i udviklingen i de offentlige finanser lagt vægten på en budgetmæssig begrundelse. Derfor vil en tilbagevisning eller sandsynliggørelse af et mindre budgetmæssigt behov kunne fjerne eller reducere motivationen til at ændre tilbagetræknings-, pensions- og skatteregler eller andre forhold i den offentlige sektor. Det er ikke en rimelig indsnævring af problemstillingen, idet der findes andre former for økonomisk begrundede behov for ændringer i den offentlige sektor end de budgetmæssige (således som det var tilfældet i f.eks. 1998), ligesom der kan være rent politisk betingede ønsker om ændringer.

Velfærdskommissionen beregner størrelsen af den offentlige budgetsaldo for en i princippet uendelig årrække ud i fremtiden med en forudsætning om uændret økonomisk politik. Det sker dels for at opgøre tilpasningsbehovets størrelse og dels for at have et budgetmæssigt udgangspunkt for forslag om ændringer. I den forbindelse er det selvfølgelig væsentligt, men også vanskeligt at fastslå, hvad der skal forstås ved uændret økonomisk politik. Er det fuldstændig uændrede regler og satser, eller skal man tage hensyn til forventede politiske reaktioner og dermed anvende såkaldte politiske reaktionsfunktioner sammen med de rent økonomiske sammenhænge?

Grundlaget for beskrivelsen af (fremtidige) politiske reaktioner er som regel empirisk og mindre velfunderet end beskrivelsen af den private sektors økonomiske reak-

tioner. Derfor bliver beregninger af tilpasningsbehovet også mere usikkert ved at inddrage nogle forventede politiske reaktioner i beregningerne. Omvendt vil manglende hensyntagen til forventede politiske reaktioner og dermed en strikt anvendelse af uændrede satser og regler medføre, at resultaterne bliver mere urealistiske, især når man som her opererer med et langt tidsperspektiv. Ved afvejningen af disse hensyn må man tage formålet med beregningerne i betragtning. Når dette er udformningen af et grundlag for politisk stillingtagen og ikke en bred samfundsmæssig fremstilling f.eks. af et historisk forløb, bør man være forsigtig med at inddrage politiske reaktionsfunktioner.

Det har Velfærdskommissionen været ved at lave to sæt beregninger henholdsvis med uændrede regler og med en politisk begrundet udvikling i velfærdsordningerne, men man kunne have ønsket sig en nærmere diskussion af problemstillingen, herunder hvorfor kommissionen ikke har inkluderet andre forventede ændringer i politiske beslutninger om det offentlige budget. For så vidt angår dens anvendelse af politiske beslutninger om velfærdsordningerne gælder det især forudsætningen om, at fortsat velstandsstigning vil forøge »krav og behov« til den offentlige velfærdsservice, hvilket forfatterne betegner som et velstandsdilemma og fortolker som en simpel realindkomstafhængig og derfor trendmæssig stigning også i udbuddet af denne service. Andersen og Pedersen (2005) anfører som begrundelse for den indlagte merstigning, at »et stigende spænd mellem efterspørgslen og det offentlige udbud af velfærdsservice (vil) have afgørende betydning for opbakningen til velfærdssamfundet« (afsnit 3.1).

Ideen med offentligt udbudte og ikke markedsmæssigt bestemte ydelser er at sætte den indkomstafhængige efterspørgsel ud af kraft og overlade fordeling, mængde og kvalitet i udbuddet af sådanne ydelser til politisk beslutning og prioritering. I modsat fald kunne offentlige opgaver i stort omfang privatiseres. Der er derfor behov for en skarpere påpegning af, hvor det økonomiske slutter og beskrivelsen af de politiske prioriteringer begynder, da det er de sidste, som er til diskussion. Det gælder begge sæt af beregninger.

Da offentlig opsparing er hovedemnet i den makroøkonomiske del af Velfærdskommissionens arbejde, er dens forudsætning om et perfekt kapitalmarked central for alle kommissionens opgørelser og analyser. Som følge af de meget store udenlandske formuedannelser (mere end 170 pct. af BNP i 2061) og den tilsvarende indenlandske, offentlige formuedannelse (mere end 100 pct. af BNP i 2061) i forbindelse med opsparingsstrategien er det også en krævende forudsætning, der ikke nødvendigvis vil være opfyldt (som antydnet af Andersen og Pedersen (2005) note 18, jfr. også nedenfor). Strategier med et væsentligt opsparingselement kræver derfor en selvstændig vurdering af virkningerne på kapitalmarkederne og de heraf følgende institutionelle og velfærdsmæssige konsekvenser.

### **Forslag og analytisk grundlag**

Andersen og Pedersen (2005) formulerer i brede termer Velfærdskommissionens forslag til reformer som en pragmatisk tilpasningsstrategi, der angives at være en midtervej mellem de to yderpunkter. Disse yderpunkter er på den ene side opsparingsstrategien, der gennem en betydelig initial stigning i skattesatserne har til formål at udjævne byrdefordelingen mellem kommende generationer, og på den anden side pay-as-you-go strategien, der ikke kræver opsparing, men alene en gradvis tilpasning af skattesatserne til ændringer i de demografiske og andre forhold. Den omfordeler derfor ikke fra nuværende til fremtidige generationer og er ikke så krævende som opsparingsstrategien med hensyn til information om udviklingen i pensionsbyrden, men den udjævner på den anden side heller ikke byrderne mellem fremtidige skatteydere.

For at holde sig på den foreslåede midtervej skal politikerne skelne mellem på den ene side udviklingen i relevante trends for pensionsbyrden (demografi mv.), der ved indeksering skal medføre ændringer i velfærdsordningerne (pensionsregler og/eller offentlige ydelser) samt en tilpasning af skattesatserne på en sådan måde, at der på det grundlag ikke opstår offentlige over- eller underskud, og på den anden side (stokastiske) afvigelser fra denne trendbestemte udvikling, som ikke skal medføre ændringer i velfærdsordninger og skattesatser. Som følge af disse afvigelser kan der opstå over- eller underskud på de offentlige finanser, men ikke systematisk offentlig op- eller nedsparring. Velfærdskommissionen foreslår i tillæg til denne konstruktion en offentlig formuedannelse baseret på et forsigtighedsmotiv, som følge af de nævnte afvigelser.

Foruden den politiske vilje (og troværdighed) til at skelne rigtigt mellem udvikling i de relevante trends og afvigelser herfra er valget af trendkriterier og det informationsmæssige grundlag for den konkrete opgørelse af deres værdi afgørende for en sådan strategi. Andersen og Pedersen (2005) har kun i skitseform redegjort for Velfærdskommissionens forslag og bortset fra levetid ikke angivet, hvilke trendkriterier der anses for relevante. De har derfor ikke gennemført beregninger, som kan angive noget om størrelsesorden og varighed af de midlertidige over- og underskud eller om størrelsesordenen af offentlig opsparing som følge af forsigtighedsmotivet. Det samme gælder spørgsmålet om, hvorledes og hvor ofte indeksering efter trendkriterierne skal finde sted. Man savner også en redegørelse for relevante erfaringer med indeksering på andre områder som f.eks. løn og obligationer. Det er derfor vanskeligt at vurdere, hvad der økonomisk er vundet ved kommissionens forslag om en sådan indeksering i forhold til den i dag anvendte politiske regulering af pensionsforholdene. Der er behov for en sådan sammenligning af politologiske og økonomiske mekanismer (og deres troværdighed) på pensionsområdet.

Velfærdskommissionens forslag om nye økonomiske mekanismer forudsætter, at betingelserne for opstilling af OLG-modellerne er opfyldt. Disse forudsætninger er

foruden et perfekt kapitalmarked bl.a., at de private agenter ikke har et arvemotiv og således ikke har intergenerationelle præferencer. I modsætning hertil forudsætter Andersen og Pedersen (2005), at den politiske præferencefunktion, der formelt anvendes til bestemmelse af den optimale udformning af de samfundsmæssige velfærdsordninger, har sådanne intergenerationelle træk og præferencer og derfor, hvad man kunne kalde et »politisk arvemotiv«. Det sker i form af mange henvisninger til og diskussion af den såkaldte generationskontrakt og f.eks. opretholdelse af en forudsætning om satsregulering af overførselsindkomsterne. Derimod kommenterer eller henviser forfatterne ikke til forudsætningen om fravær af et arvemotiv i deres OLG-beskrivelse af den private sektor. Denne forudsætning er afgørende for den offentlige gælds realøkonomiske betydning i modellerne.

Det samtidige fravær af et arvemotiv hos private agenter og tilstedeværelsen af et helt analogt motiv hos politikerne kræver en nærmere diskussion. Det er nemlig ikke muligt umiddelbart at udlede en samfundsmæssig og dermed politisk præferencefunktion med intergenerationelle præferencer fra private agenter præferencefunktion med endelig levetid, hvis ikke sidstnævnte indeholder et arvemotiv. Omvendt bortfalder grundlaget for OLG-beskrivelserne og dermed grundlaget for generationsregnskabet og formelle analytiske udsagn om virkningen af rente, længere levetid og behovet for risikodeling, hvis private agenter forudsættes at have et arvemotiv.

En ofte anvendt fremgangsmåde til konstruktion af aggregeret adfærd og dermed en samfundsmæssig præferencefunktion er at antage, at private agenter, når de er født, som følge af konstant periodevis døds sandsynlighed har mulighed for at leve evigt. Herved bliver det modelmæssig muligt at aggregere alle variable. Det er imidlertid en diskussion værd, om en sådan approksimation, der medfører, at modeller specificeres i variable aggregeret på tværs af generationer, giver et anvendeligt analytisk grundlag, når emnet er den økonomiske fordeling mellem disse generationer. Aggregerede variable giver normalt ikke information om delmængder i aggregatet. Det er ikke oplyst, hvordan Velfærdskommissionen med henvisning til Andersen (2005) har løst det skitse-rede problem.

Om det empiriske kan bemærkes, at ved en årlig levetidsforlængelse på f.eks. 2 måneder i forhold til en gennemsnitlig levetid på 80 år, er ændringen i modellens døds sandsynlighed af meget begrænset størrelse og dermed betydning sammenlignet med f.eks. ændringer i den årlige rente og produktivtetsstigning. De nævnte variable indgår på analog vis i den samlede diskonteringsfaktor i modeller formuleret med nævnte udgangspunkt. Når emnet er omfordeling og levetid, er det derfor både af analytiske og empiriske grunde formentlig rimeligere at søge andre steder end i den nævnte aggregering, når det gælder argumenter for at anvende en aggregeret præferencefunktion med et intergenerationelt indhold til bestemmelse af den optimale politik. Det tilsvarende gæl-

der derfor begrundelsen for eksistensen af en generationskontrakt og et »politisk arvemotiv«.

En gammeldags mulighed kunne være at indføre det »politiske arvemotiv« som et direkte postulat om politikernes adfærd. Empirisk forekommer det imidlertid heller ikke særlig plausibelt, idet der historisk ikke har været større omfordelinger eller udligninger mellem generationer via de offentlige finanser, hvilket Andersen og Pedersen (2005) også gør opmærksom på. Det gør heller ikke en forudsætning om et »politisk arvemotiv« plausibelt, at de politisk fastlagte vilkår for offentlig pension, herunder tilbagetrækningsalderen, er blevet ændret i både den ene og den anden retning, hvilket ikke umiddelbart synes at kunne forklares ved permanente og stabile politiske præferencer med et »arvemotiv«. Forudsætningen om en stabil generationskontrakt fortolket som et arvemotiv i den politiske præferencefunktion til udledning af de mest hensigtsmæssige former for politisk indgreb er derfor tvivlsom. De omfordelinger over de offentlige finanser, som henvisninger til generationskontrakten skal forklare, er snarere udtryk for almindelige resultater af en politisk proces, som kan variere og ændres over tid.

Det diskuteres stadig, om OLG-modellernes forudsætning og dermed en af forudsætningerne for generationsregnskabet om (fuldstændig) fravær af ricardiansk ækivalens er en holdbar antagelse i empirisk henseende. Det er som nævnt også en forudsætning for OLG-modellerne, og hvad deraf følger, at alle private agenter har adgang til at låne ind og ud på et perfekt kapitalmarked. Der er forhold og empiriske resultater, som tyder på, at denne forudsætning heller ikke er fuldt ud opfyldt. Andre forhold kan også begrunde forsigtighed ved anvendelse af OLG-modeller (for en kritisk oversigt se Buiters (1997)), men hvis det vurderes at være relevant at anvende forudsætningerne og dermed en OLG-model, skal man selvfølgelig anvende alle konklusioner, der følger af de nødvendige forudsætninger.

Andersen og Pedersen (2005) synes at overse, at de eksisterende offentlige pensionsordninger er med pay-as-you-go finansiering, der i henhold til den simpleste OLG-model fortrænger privat opsparing. Det er derfor nødvendigt for kommissionen at tage stilling til, om den nuværende, samlede opsparing og formentlig også kapitalapparatet samfundsmæssigt er af hensigtsmæssig størrelse eller for lille. Det sidste kunne begrunde ændringer i offentlige pensionsordninger og deres finansiering, således som det var tilfældet ved introduktionen af arbejdsmarkedspensionerne. Tilsvarende er det (som forfatterne gør) i diskussionen af fordelingen mellem generationer nødvendigt at være opmærksom på, at de nuværende generationer som følge af den eksisterende pay-as-you-go finansiering allerede har bidraget til (og modtaget fra) omfordelingen via de offentlige finanser.

Det fremgår ikke af diskussionen om de OLG-modeller, som Andersen og Pedersen

(2005) og dermed Velfærdskommissionen anvender i deres argumentation, alle er formuleret for et lille, åbent samfund. Det er et relevant spørgsmål, idet det f.eks. fra litteraturen om økonomisk vækst er velkendt, at åbning af sådanne modeller mod omverdenen ved introduktion af varehandel og adgang til et perfekt internationalt kapitalmarked giver nogle uplausible resultater bl.a. i form af øjeblikkelig konvergens for reale størrelser (Barro og Sala-i-Martin (2004)). Disse resultater har bl.a. ført til introduktion af en antagelse om kreditrestriktioner på kapitalmarkederne, men herved ændres som tidligere nævnt en væsentlig antagelse for både opgørelsen af den intertemporale budgetrestriktion og den velfærdsmæssige vurdering af de opsparingsbaserede strategier.

Andersen og Pedersen (2005) begrundet ikke Velfærdskommissionens forslag om en tilpasningsstrategi med henvisning til empiriske resultater om de centrale antagelser i deres analytiske modelapparat. Mange modeller kan opstilles om det samme forhold, og analytiske modeller giver ikke nødvendigvis entydige resultater, hvorfor konkrete anbefalinger som regel er afhængige af empiriske vurderinger af forhold i modellerne. For at antage, at to eller flere effekter ophæver hinanden, er det heller ikke tilstrækkeligt, at der analytisk findes modsat rettede effekter. Andersen og Pedersen (2005) mangler derfor at sandsynliggøre ved (henvisning til) empiriske analyser, at den af Velfærdskommissionen foreslåede tilpasningsstrategi og ikke andre strategier er at foretrække.

### **Konklusion**

På trods af omfattende beregninger vedrørende den offentlige budgetrestriktion har Velfærdskommissionen (indtil videre) ikke underbygget den foreslåede makroøkonomiske tilpasningsstrategi med henvisninger til beregninger eller empiriske forhold, der kan sandsynliggøre, at deres forudsætninger er væsentlige, og at netop deres forslag er den samfundsmæssigt optimale strategi. Det er formentlig årsagen til, at kommissionen betegner sit forslag om beskatning og offentlig opsparing som »en pragmatisk tilgang« til løsning af balanceproblemerne. Behov for en pragmatisk reformproces følger også af uafklarede spørgsmål i den analytiske begrundelse for og den praktiske udformning af den foreslåede strategi.

Det anførte er ikke ensbetydende med, at de økonomiske problemer, som er belyst ved kommissionens og andres beregninger vedrørende holdbarheden af det offentlige budget, ikke skal konfronteres i en reformproces. Virkningen er alene, at denne proces for så vidt angår de makroøkonomiske aspekter bør være ubundet af de begrænsninger, som kunne følge af henvisninger til Velfærdskommissionens beregninger, generationskontrakten, forsigtighedsmotivet, forslag om flerstrengt indeksering mv. Der er således behov for yderligere makroøkonomiske overvejelser vedrørende både størrel-



sen af den offentlige opsparing, beskatningen, arbejdsudbuddet og andre realøkonomiske variable samt til overvejelser om den politiske holdbarhed af kommissionens og andres reformforslag, den private sektors forventninger og reaktioner i den forbindelse, ejendomsretten til de opsparede midler og andre forhold på kapitalmarkedet, hvor opsparingen jo skal anbringes.

*Litteratur*

- Andersen, T. M. 2005. Social Security and Longevity. Anført som Working Paper i Andersen og Pedersen (2005), men ikke offentliggjort.
- Andersen, T. M. og L. H. Pedersen. 2005. Demografi, Velstandsdilemma og Makroøkonomiske Strategier. *Nationaløkonomisk Tidsskrift* 143, s. 189-229.
- Barro, R. J. og X. Sala-i-Martin. 2004. *Economic Growth* (2. udgave). MIT Press, Cambridge, Mass.
- Buiter, W. H. 1997. Generational Accounts, Aggregate Saving and Intergenerational Distribution. *Economica* 64, s. 605-26.

## Splinten og bjælken

Jørgen Goul Andersen

Institut for Økonomi, Politik og Forvaltning, Aalborg Universitet, E-mail: [goul@socsci.aau.dk](mailto:goul@socsci.aau.dk)

Velfærdskommissionens beregninger vedrørende den fremtidige forsørgerbyrde bygger på en serie tvivlsomme forudsætninger, der tenderer mod at overdrive den i dansk sammenhæng beskedne udgiftsbelastning som følge af højere levealder. Samtidig kan den overdrevne fokus på demografien medvirke til, at opmærksomheden forskydes fra andre og større farer. Sat på spidsen advarer man mod splinten, men overser bjælken. Endelig kunne man have ønsket et bredere udsyn, der satte fokus på velfærdsudfordringer og ikke kun på spørgsmålet om, hvordan velfærdsstatens skadevirkninger kan begrænses.

Det skal dog indledningsvist fremhæves, at Velfærdskommissionen på dette punkt leverer et særdeles grundigt analysearbejde. Man har fået udarbejdet en langt mere realistisk befolkningsfremskrivning, der viser en større stigning i levealderen og dermed i antallet af ældre end tidligere prognoser. Kommissionen punkterer også myter om, at øget indvandring eller øget fertilitet kan afhjælpe forsørgelsesproblemet. Endelig skal Velfærdskommissionen roses for at have fremlagt en række følsomhedsberegninger, der viser betydningen af en række antagelser.

Til gengæld kunne man godt have undværet den indledende bemærkning »Et væsentligt træk ved den danske velfærdsmodel er den sociale kontrakt mellem generationer«. Det danske særtræk er jo, at man i ualmindelig høj grad har ophævet denne kontrakt, fordi fremtidens pensionister overvejende forsørges af deres egen pensionsopsparing. I modsætning til de fleste lande *har* Danmark gennemført de store velfærdsreformer, der skal forhindre, at det stigende antal ældre overbelaster pensionsudgifterne. Det er i øvrigt sket i bedste overensstemmelse med Verdensbankens anbefalinger.

I Danmark er det ukorrekt at betegne fremtidens pensionister under ét som »forsørgede«. Flere og flere pensionister bliver tværtimod »forsørgere« gennem det meste af pensionisttilværelsen. En meget stor del skal kun have folkepensionens grundbeløb på p.t. 4.741 skattepligtige kroner om måneden, samtidig med at de betaler skat af den udbetalte pension fra ATP, SP, arbejdsmarkedspension og individuelle pensioner. Det er hovedårsagen til, at man i udlandet betragter Danmark som et af de få lande, der i det store og hele har løst forsørgelsesproblemet.

Danmark er i øvrigt ikke alene begunstiget af ekstremt billige offentlige pensioner – der målt som andel af BNP koster væsentligt mindre i 2003 (4,5 pct.), end dengang

den fulde folkepension blev indført i 1970 (5,6 pct.), jf. DØR (2005, s.105). Danmark er også begunstiget af den fremtidige befolkningsudvikling. I 2050 ventes ældrebrøken (forholdet mellem 65+ årige og 15-64 årige) i Danmark at stige til 40-45 pct. Men f.eks. i Spanien og Italien er tallet over 65 pct. og i Japan over 70 pct. (OECD, 2005a).

Hvordan man end vender og drejer det, er Velfærdskommissionens beregnede manko på ca. 4 pct. af BNP for at opnå bæredygtighed da heller ikke nogen gyser. Det er ikke mere, end Finansministeriets »budgetbisser« godt kunne finde over en 10-årig periode. Det er ikke større end forskellen mellem en socialdemokratisk og en borgerlig regering (renset for konjunktur). Det er ikke større end faldet i de offentlige udgifters andel af BNP fra 1996/97 til i dag. Og det er ikke ret meget større end forskellen mellem sundhedsudgifterne i Danmark og Schweiz. Det er måske derfor, man i den folkelige agitation har fundet det nødvendigt at forstørre tallene lidt ved at omregne fra BNP-procenter til stigning i bundskatten – en pædagogisk manøvre, læserne af Nationaløkonomisk Tidsskrift dog godt kunne have været sparet for.

Overordnet er det dog klogt at mane til forsigtighed. Vi står i disse år i en unik demografisk situation med en sammenklumpning af befolkningen i de erhvervsaktive aldersklasser, vi aldrig har oplevet før og aldrig kommer til at opleve igen. Det giver en unik chance for at afvikle statsgæld og polstre sig til fremtiden. Samtidig begunstiges landet i de nærmeste år af uventet store engangsindtægter fra nordsøolien – som yderligere kan bidrage til en forestilling om at kunne gå på vandet. F.eks. ser 2010-planens statsgælds målsætning ligefrem ud til at kunne opfyldes før tiden – bl.a. pga. olien, der også giver uventet hjælp til bæredygtigheden i Velfærdskommissionens regnestykke. Provenuet burde ligesom i Norge via en oliefond komme senere generationer til gavn, fordi det er én generation, der brænder gevinsten af.

Forsigtighed er en dyd. Men billedet er formentlig endnu lysere, end Velfærdskommissionens beregninger tyder på. Der indgår nemlig – ofte implicit – en serie problematiske antagelser:

- fuld lønregulering af overførselsindkomster
- flere år til livet, men ikke mere liv til årene
- uændret uddannelsesniveau i forhold til nuværende 30-årige
- ingen effekt af ændrede incitamentter til tilbagetrækning
- ingen forbedring af erhvervsdeltagelsen blandt indvandrere
- ingen produktivitetstigninger i offentlig service

*Fuld lønregulering af overførselsindkomster*

Antagelsen om, at overførselsindkomster reguleres i forhold til arbejdslønnen (mi-

nus pensionsbidrag) er problematisk af to grunde. For det første fragår normalt 0,3 pct. af stigningen til den såkaldte satsreguleringspulje. Det er indskærpet, at disse penge bør gå tilbage til særlige indsatser for modtagere af overførselsindkomst. Puljen, der nu er på ca. 6 mia. kr. og vokser med ca. 600 mio. kr. årligt, kan dog tænkes i nogen grad at fortrænge udgifter, der ville være afholdt alligevel. For det andet bygger satsreguleringen på en sammenvejning af lønstigningen for arbejdere og funktionærer på basis af faste vægte. Men derved udgår en del af den lønstigning, der skyldes kvalifikationsforbedringer, og som afspejler sig i en stigende andel funktionærer.

Det betyder tilsammen, at overførselsindkomsterne sakker bagud med skønsmæssigt 0,5 pct. om året – omkring 20 pct. over 40 år. Det er ca. 200 mia. af overførselsindkomsterne (2004), der reguleres med satsreguleringen – af resten reguleres nogle bedre, andre dårligere. Alene tricket med de faste vægte reducerer Velfærdskommissionens manko med et betragteligt beløb, resten afhænger af, hvordan man vil vurdere fortrængningseffekten.

At overførselsindkomsterne sakker betydeligt bagud, er dog uomtvisteligt, og det burde Velfærdskommissionen til gengæld diskutere som et problem. Det kan føre til fattigdom, og det kan også betyde et kvalitativt skift i f.eks. dagpengesystemets tilstrækkelighed. Endelig burde man også diskutere hensigtsmæssigheden af, at satsreguleringspuljen vokser og vokser.

#### *År til livet uden liv til årene*

Antagelsen om, at pleje- og behandlingsbehovet følger alderen, er også problematisk. Man må forvente, at der bliver føjet »liv til årene« i takt med, at der føjes »år til livet«. Specielt det sidste leveår er omkostningskrævende. Ét yderpunkt er at beregne udgiften ud fra alderen, et andet yderpunkt er at beregne udgiften ud fra restlevetiden. Virkeligheden ligger et sted derimellem, men Velfærdskommissionens hovedskøn bygger på det mest pessimistiske yderpunkt.

Igen burde man dog diskutere den velfærdspolitiske udfordring, nemlig at sikre bedre folkesundhed og sociale netværk mv. blandt de ældre. Det giver både økonomiske besparelser og øget velfærd, men kommissionens perspektiv skygger for de relevante spørgsmål.

#### *Uændret uddannelsesniveau*

Velfærdskommissionen antager uændret tilbagetrækningsmønster for de enkelte uddannelsesgrupper – hvilket måske er en kende optimistisk. Men til gengæld »fastfryser« man uddannelsessammensætningen ud fra niveauet for de nuværende 30-årige. Det ignorerer for det første, at mange først afslutter deres uddannelse senere – også fordi vi allerede har og i stigende grad får »livslang uddannelse«. Og for det andet må man for-

vente en væsentlig forbedring af uddannelsesniveaut fremover. Ellers får Danmark større økonomiske bekymringer end ældrebomben. Igen burde man her pointere den centrale udfordring: Øget – og livslang – uddannelse.

#### *Incitament til tilbagetrækning*

Vi ved næsten ingen ting om tidlig tilbagetrækning i Danmark – hvilket nok er årsagen til, at der florerer der en masse myter og skråsikre holdninger om et fænomen, der er utrolig komplekst. Derfor ved vi heller ikke, hvordan arbejdsmarkedspensionernes stigende vægt vil påvirke tilbagetrækningsmønstret. Velfærdskommissionen har ikke indregnet nogen effekt. Men tror man på incitament, vil det danske pensionsystem fremover give stærke incitament til senere tilbagetrækning, fordi den årlige arbejdsmarkedspension er direkte relateret til tilbagetrækningsalderen: Jo længere, man holder ud, jo højere bliver den årlige pension. Nu er det med pension ligesom med indkomstskat: Vil lavere skat få folk til at arbejde mere (pga. incitament), eller vil det få dem til at arbejde mindre (fordi de kan tjene den nødvendige indkomst på færre timer)? Vi ved det ikke. Men det nye svenske pensionssystem hyldes ofte for dets incitament til at udsætte tilbagetrækningen, og Velfærdskommissionen forbigår, at det danske system i så henseende bliver en modificeret udgave af det svenske.

Den velfærdspolitiske udfordring er at øge tilbagetrækningsalderen, specielt for de mange, der kan og vil arbejde. Her savner man langt større fokus på udstødning og barrierer. Når det gælder dem, der kan, men ikke ønsker at arbejde, savner man empirisk viden både om, hvordan efterlønnen fungerer og om, hvordan de forskellige former for private pensionsordninger (ikke kun arbejdsmarkedspensioner, men bl.a. også de hastigt voksende ratepensioner) påvirker tilbagetrækningen.

#### *Erhvervsdeltagelsen blandt indvandrere*

Så vidt det kan læses ud af Velfærdskommissionens rapporter, har man mekanisk fremskrevet de nuværende alders-, køns-, uddannelses- og nationalitetsbetingede erhvervsfrekvenser for indvandrere. I givet fald strider det faktisk mod en veldokumenteret viden om, at indvandreernes erhvervsdeltagelse stiger i takt med deres opholdstid i landet. Da opstrammingerne betyder, at den gennemsnitlige opholdstid i landet stiger ganske kraftigt, burde det, ud fra hvad vi ved, automatisk føre til større erhvervsdeltagelse for indvandrere. Hertil kommer så, hvad en målrettet integrationsindsats kan kaste af sig; det er naturligvis – som også Velfærdskommissionen påpeger – en af de store velfærdspolitiske udfordringer.

#### *Ingen produktivetsforbedringer i offentlig service*

Endelig – og måske mest alvorligt, set fra et politologisk perspektiv – har Velfærds-

kommissionen indirekte indlagt en antagelse om, at produktiviteten i den offentlige sektor ikke kan forøges. Man antager nemlig, at de offentlige serviceudgifter passivt følger den demografiske udvikling. Individuelt offentligt forbrug pr. person i en given gruppe fremskrives derefter stort set svarende til lønudviklingen, og kollektivt offentligt forbrug fremskrives svarende til væksten i BNP.

Nu er det velkendt, at det er svært at opnå produktivetsforbedringer inden for offentlig service. Men det er ikke ensbetydende med, at det ikke er muligt, og det opnås ikke kun ved, at folk skal »løbe hurtigere«. I dele af omsorgssektoren er der måske grænser for, hvor meget produktiviteten kan øges, uden at det går ud over service-niveauet. Men de fleste steder er rammerne ganske vide. Hertil kommer, at offentligt konsum ikke kun går til aflønning af ansatte. Der er også et stort forbrug af varer – som ofte kan have en produktivetsudvikling og dermed billiggørelse, der svarer til, hvad vi oplever i det private forbrug. Delvist overlappende hermed er der store dele af det kollektive konsum, der ikke har udviklet sig og ikke vil udvikle sig proportionalt med BNP. Forsvaret er et klassisk eksempel, men også f.eks. vejvæsen lægger beslag på en klart faldende andel af BNP.

Produktivetsgevinster i den offentlige service kommer dog sjældent af sig selv. De kommer, populært sagt, når de tages hjem. Det ved politikerne – og om nødvendigt skærer de også med serviceforringelser til følge. Men meget afhænger af, hvor styrbare systemerne er. Bl.a. er det påfaldende, at sundhedsudgifternes andel af BFI i Danmark stadig ligger under 1981-niveau og indtil 2001 endda på 1972-niveau, trods eksplosionen i behandlingsformer og trods 1/3 flere ældre i perioden. I alle de nordiske lande oplevede man i 1990'erne svagt faldende sundhedsudgifter, målt som andel af BNP. Det står i skærende kontrast til f.eks. Schweiz og Tyskland, hvor politikerne helt har mistet kontrollen. Det er imidlertid almindelig kendt, at der er ringe sammenhæng mellem sundhedsudgifternes størrelse og sundhedsvæsenets kvalitet fra land til land – styringssystemet er afgørende. Siden 2001 har der i øvrigt også været eksplosiv vækst i de danske sundhedsudgifter.

Kort sagt følger de offentlige udgifter ikke automatisk udviklingen i behovene. Det er i øvrigt lykkedes Velfærdskommissionen at gennemgå store dele af den righoldige litteratur om årsagerne til den offentlige sektors vækst uden at opdage den politologiske hovedkonklusion, nemlig at væksten ikke kan forklares ud fra udviklingen i behovene. Det er bl.a. på den baggrund, mange politologer vil sige, at behovene er den mindste faktor i ligningen, og at fremskrivninger som Velfærdskommissionens bortleder opmærksomheden fra de reelle udfordringer, specielt i den servicetunge skandinaviske velfærdsmodel (hvor det i den overførselstunge kontinentaleuropæiske model i højere grad er overførslerne, der er udfordringen).

Til gengæld er Velfærdskommissionen overraskende blind over for behovssiden på

det ene punkt, hvor den kunne fortjene berettigelse. Man konstaterer, at udgifterne til velfærdsservice er vokset fra 13,3 pct. af BNP i 1971 til 17,8 pct. i 2001. Det svarer til en mervækst i forhold til BNP på 0,15 pct. om året, som primært forklares med Wagners lov og Baumol-effekten. Baumol-effekten er tildels reel her, men at bringe Wagners lov ind i billedet, når det gælder væksten i den offentlige service fra 1971 til 2001, må forundre. Det er den periode, hvor kvinderne kom ind på arbejdsmarkedet, og hvor man i Danmark og de øvrige nordiske lande *valgte* at tilpasse den offentlige service. Det er ikke nogen naturlov, at der sker en sådan tilpasning, men har man først valgt det, er der ofte ingen vej tilbage. Børnepasningsområdet er det klareste eksempel.

Der vil altid være pres for mere service. Men hvor der på børnepasningsområdet hidtil har været pres både for en højere dækningsgrad og et højere serviceniveau, begrænser presset sig i dag til et højere serviceniveau. Dækningen er nemlig praktisk taget 100 pct. Det er kort sagt en éngangsforeteelse, at den offentlige service tilpasses til kvinders erhvervsarbejde. Her er der ikke grund til at forvente nogen gentagelse, som Velfærdscommissionen gør ved at lave alternative beregninger ud fra historiske vækstrater for at vise, hvor konservativt, man skønner. En dekomponering efter område tegner, som illustreret ved sundhedssektoren ovenfor, et ganske andet billede. Sundhedssektoren er den sektor, hvor Wagners lov passer bedst, og hvor det langsigtede efterspørgselpres er størst, men her er udgifterne netop *ikke* steget som andel af BNP fra 1971 til 2001.

Også presset for kortere arbejdstid må forventes at aftage i takt med den nedsættelse, der allerede er sket. Igen er den velfærdspolitiske udfordring at bidrage hertil ved at skabe større fleksibilitet for den enkelte, f.eks. via timebanker o.l., så der kan skabes rum til bl.a. øgede fritidsaktiviteter uden at sænke arbejdstiden.

Velfærdscommissionens perspektiv sætter groft sagt fokus på, hvordan man kan begrænse velfærdsstatens økonomiske skadevirkninger. Det er der tre problemer med. For det første overvurderer man efter alt at dømme en række problemer – der selv med Velfærdscommissionens beregninger tager sig ret beskedne ud. For det andet skygger perspektivet for at formulere en række velfærdspolitiske udfordringer, der tilmed kan have win-win karakter: Bedre velfærd som sigte – og bedre økonomi som sidegevinst.

Det alvorligste er dog de advarsler, man ikke formulerer. Er styrbare systemer (f.eks. på sundhedsområdet, hvor advarselslampen allerede blinker) den ene velfærdspolitiske hovedudfordring, så er styrbare politikere en anden. Komparativt set er danske politikere engle, når det gælder økonomisk ansvarlighed. Men fristelserne er der konstant. I *rational choice* forstand er det ikke rationelt for en politiker at gennemføre upopulære nedskæringer nu og spare alle pengene op, for at senere generationer af politikere kan nyde frugterne af den finanspolitiske bæredygtighed. Så hellere tilbage-

diskontere en del af de forventede fremtidige glæder som et større »økonomisk råderum« her og nu.

Der verserer p.t. optimistiske skøn over gevinsten ved at ændre efterlønnen, som ville få forskere i tilbagetrækning verden over til at strømme til Danmark. Næst efter Sverige er Danmark nemlig i forvejen det EU-land, der har flest 55-64 årige mænd i beskæftigelse. De eneste rige OECD-lande, der ligger marginalt højere, er Norge, Schweiz, Japan, New Zealand og Island – og i alle disse lande er beskæftigelsen i øvrigt faldende for de 55-64 årige. Den tanke er ikke så fjern, at man beskærer efterlønnen, bruger en del af det forventede provenu til skattelettelser/serviceforbedringer her og nu, og efterfølgende må konstatere, at virkningen ikke blev den forventede. Der skal ikke gå så meget galt, før man ligefrem kan *forringe* den finanspolitiske bæredygtighed frem for at forbedre den.

Den største fare ved den overdrevne fokusering på demografien er, at det let kan svække opmærksomheden på de store trusler. Dem hører demografien ikke til i det danske tilfælde – selv om det skal medgives, at senere tilbagetrækning fra og (måske mere realistisk) tidligere start på arbejdsmarkedet vil være en økonomisk og i mange tilfælde velfærdspolitisk gevinst.

#### Litteratur

- OECD. 2005a. *Society at a Glance*. Paris: OECD. Økonomiske Råd, Det. 2005. Dansk Økonomi. Forår 2005. København. Det Økonomiske Råd.
- OECD. 2005b. *Employment Outlook*. Paris: OECD.



## Velferdsstatens finansieringsproblem: Utsikter i »annerledeslandet« Norge

Erling Holmøy

Forskningsleder, Statistics Norway, E-mail: [erl@ssb.no](mailto:erl@ssb.no)

### 1. Innledning

Torben Andersen og Lars Haagen Pedersen (heretter AP) gir i sin artikkel en grundig analyse av problemene man på lang sikt vil få med å finansiere den danske velferdsstaten. Jeg har tro på at mulighetene for gjennomslag i den politiske debatten øker når man kan supplere kvalitative momenter med den typen godt underbygd tallfesting av de relevante effektene som Velferdskommisjonen (VFK) har gjennomført. Tall er ofte de beste adjektiver. Selv om de alltid vil være usikre, finnes det ingen egentlige alternativer til langsiktige fremskrivninger. Den rasjonelle holdningen til usikkerhet er ikke å tillegge prediksjonene liten eller ingen vekt, og usikkerhet er heller ikke det samme som bias. Fastlegging av realistiske størrelsesordener på problemene og relevante effekter disiplinere debatten, det muliggjør en bedre prioritering av hvilke usikre forhold bør avklares ytterligere, og av hvor det er viktigst med politiske tiltak, gitt dagens informasjon. Det er derfor meget interessant at man i Danmark har utviklet et så avansert og relevant modellverktøy som DREAM for analyser av så aktuelle og tempererte problemstillinger som finansiering av velferdsstaten reiser.

Min vurdering er at de VFK-analyser som AP baserer seg på virker meget kompetent utført. Man kan selvsagt problematisere de fleste forutsetninger, men det er ikke enkelt å argumentere overbevisende for konkrete alternative tallanslag. Noen viktige forutsetninger er dessuten valgt for å ha et gjenkjennelig og lett tolkbart utgangspunkt for diskusjon, heller enn å være forsøk på mest mulig realistisk prognose. Jeg tenker da særlig på forutsetninger knyttet til videreføring av »dagens politikk«. Gjennom sensitivitetsanalysene kartlegger VFK/AP viktige trade-offs og dermed et felles sett av premisser for videre og mer konkret politikktutforming. Alt i alt er min personlige oppfatning at VFK's grunnforløp heller undervurderer de langsiktige statsfinansielle problemer enn det motsatte. Det skyldes først og fremst at jeg tror at det er urealistisk at det ikke vil skje en vekst i individuell velferdsservice. Forventet økning i de eldres levealder har sannsynligvis også en negativ bias i VFKs grunnforløp. Det aner meg at denne oppfatningen om eventuell bias deles av AP.

De relevante momentene i en diskusjon av VFK's fremskrivninger er meget godt ivarett av AP og i VFK's publikasjoner. I stedet for å gjenta disse vil jeg i det følgende presentere resultater og konklusjoner fra relaterte beregninger for Norge. I denne

forbindelse vil jeg kort kommentere noen forskjeller mellom de danske og norske analysene, både når det gjelder bruken av modeller, forutsetninger og resultater. Til slutt vil jeg diskutere noen av de felles metodemessige problemer alle møter som ønsker å uttale seg kvantitativt om en relativt fjern fremtid, slik man tvinges til for å gi relevant innsikt om sammenhenger mellom demografi, makroøkonomi og offentlige finanser.

## 2. Offentlige finanser i Norge

Norge omtales ofte som et »annerledes-land«. Det er iallfall riktig når det gjelder offentlige finanser, fordi den norske stat mottar det aller meste av kontantstrømmen fra olje- og gassutvinningen. Betydningen av petroleumsinntektene er et særnorsk innslag i diskusjonen om offentlige finanser. Spesielt blir spørsmål knyttet til fordeling mellom generasjonene i norsk sammenheng nødvendigvis også et spørsmål om hvilke generasjoner som skal få mest av denne valutagaven som de ikke-fornybare petroleumsressursene representerer.

Statens kontantstrøm fra petroleumsvirksomheten investeres i et eget Statens Petroleumsfond. Kapitalen i dette fondet tilsvarer omtrent den offentlig finansformue og forventes ved utløpet av 2005 å være ca. 1300 mrd. 2005-NOK, tilsvarende ca. 71 prosent av forventet BNP og ca. 0,28 millioner 2005-NOK per innbygger. Statens kontantstrøm fra petroleumsvirksomheten utgjør ca. 41,5 prosent av statsbudsjettets utgifter.<sup>1</sup> Norske økonomer har en nærmest permanent og vanskelig oppgave i å overbevise opinionen om at disse inntektene representerer en omveksling av formue – ikke ordinær inntekt. Spesielt kan dagens høye produksjon gi et misvisende bilde av hvor stor den norske petroleumsformuen er. Regjeringens »Perspektivmelding« fra 2004<sup>2</sup> beregnet denne formuen (definert som neddiskontert ressursrente fra gjenværende petroleumsressurser) til ca. 0,5 mill. 2003-NOK per innbygger. Til sammenligning anslås verdien av humankapitalen til vel 10 ca. mill. 2003-NOK per innbygger. Tallene er usikre, og petroleumsformuen øker selvsagt dersom dagens høye oljepriser legges til grunn. Likevel rokkes ikke konklusjonen: Også i Norge må det først og fremst arbeides dersom man også på lang sikt vil ha et høyt privat og offentlig forbruk.

Spørsmålet om hvordan petroleumsinntektene skal brukes over tid, dvs. fordeles på generasjoner, har vært en gjenganger i den økonomiske debatt i Norge de siste 30 år. Fra og med 2002 har en såkalt »Handlingsregel for finanspolitikken« fastlagt en norm for hvor mye petroleumsinntekter som kan brukes løpende. Utgangspunktet er som nevnt at statens løpende petroleumsinntekter investeres i petroleumsfondet. Regelen sier: (a) kun realavkastningen, for tiden 4 prosent, av forventet fondskapital ved utgangen av inneværende år kan plusses på det offentliges øvrige inntekter for kommende

---

1. Ifølge Reviderte nasjonalbudsjett for 2005, Stortingsmelding nr. 2 (2004-2005), Finansdepartementet.

2. Stortingsmelding nr. 8 (2004-2005), Finansdepartementet.

budsjettår; (b) i et konjunkturtelt »normalår« skal denne summen dekke de løpende offentlige utgifter; (c) det kan drives motkonjunkturpolitikk, slik at regelen kan fravikes utenfor normalår, men over tid på en symmetrisk måte. Nasjonalbudsjettet for 2005 anslår at den »tillatte« bruken av petroleumsinntekter finansierer et »strukturelt oljekorrigert« budsjettunderskudd tilsvarende ca. 4,8 prosent av BNP for fastlands-Norge frem til 2010. Regelen innebærer at olje og gass gradvis konverteres til finansielle fordringer på utlandet. På lang sikt konvergerer regelen mot at Norge konsumerer permanentinntekten av en petroleumsformuen, slik at alle generasjoner får glede av denne naturressursen. Petroleumsfondet vil opprettholde sin verdi (målt i internasjonal kjøpekraft), men vil utgjøre en stadig mindre andel av en voksende nasjonalinntekt. Så lenge regelen har eksistert har budsjettunderskuddet klart oversteget det handlingsregelen tillater. For 2005 forventes den »ulovlige« bruken av petroleumsinntekter å bli 24 milliarder NOK, nær 1/3 av det anslåtte oljekorrigerte underskuddet. Flere observatører og politikere har dermed ment at handlingsregelen er død, mens andre mener at regelen fortsatt bidrar til budsjett disiplin.

For mange kan dagens høye petroleumsinntekter gjøre det vanskelig å tro at Norge vil ha problemer med å finansiere sin velferdsstat etter hvert som aldringen gjør seg gjeldende etter 2010. De scenarier som presenteres i det følgende, indikerer imidlertid at problemene med å finansiere videreføring av dagens velferdsordninger vil være vel så store i Norge som det VFK finner for Danmark, selv under forsiktige antakelser om veksten i offentlig forbruk. Som i andre land er altså fremtidsbildet av norske offentlige finanser radikalt forskjellig fra det som de løpende budsjettall for tiden tegner. Petroleumsinntektene gjør at kontrasten er særlig sterk i Norge. Dermed er det også en enda sterkere behov for langsiktige fremskrivninger i Norge enn i andre land; de er tross alt det eneste verktøy vi har for å innarbeide hensyn til fremtidens generasjoner i dagens vurderinger. Før jeg går nærmere inn på drivkreftene bak utviklingen for offentlige finanser i Norge, vil jeg si litt om det modellapparat som benyttes i Norge.

### 3. Modellverktøy i danske og norske beregninger

Norge har en lang tradisjon i å bygge og anvende empiriske modeller for fremskrivninger på kort og lang sikt. Det er særlig Finansdepartementet og forskere i SSB som mer eller mindre regelmessig utarbeider slike fremskrivninger. De baseres på modeller utviklet i SSB. For langsiktige fremskrivninger (20-60 år frem i tid) benyttes den generelle likevektsmodellen MSG6.<sup>3</sup> Modellen gir en relativt disaggregert beskrivelse av produktmarkedene og offentlige utgifter og inntekter. Fleksible relative priser realise-

---

3. MSG er en forkortelse for Multi Sectoral Growth. MSG6 er sjette generasjon av MSG-modellen. Den er meget forskjellig fra MSG1, utviklet av professor Leif Johansen (Johansen, 1960. *A Multi-Sectoral Study of Economic Growth*, Contributions to Economic Analysis 21, North-Holland, Amsterdam). Ulike generasjoner av likevektsmodeller utviklet fra MSG1 har regelmessig vært brukt av Finansdepartementet siden 1968.

rer full ressursutnyttelse og langsiktig balanse i utenrikshandelen. Aktørenes adferd følger konvensjonell mikrøkonomisk teori.

Selv om både MSG6 og DREAM er generelle likevektsmodeller, er de likevel forskjellige på viktige områder. DREAM deler husholdningssektoren opp i overlappende generasjoner (OLG), mens denne er helt aggregert beskrevet i MSG6. MSG6 gir derimot en mer detaljert beskrivelse av produksjonssektoren enn DREAM. Videre inneholder DREAM en detaljert beskrivelse av pensjonssystemet og andre offentlige overføringer til husholdningene, noe som kreves for at modellen skal være relevant for analyser av offentlige finanser. I seg selv inneholder MSG6 verken relasjoner som fremskriver befolkningen og dens sammensetning, eller pensjonssystemets budsjett- og incentivvirkninger. I stedet simuleres MSG6 i et samspill med en rekke andre spesialmodeller i en »bottom-up approach«. MSG6-fremskrivningene baseres på detaljerte »offisielle« befolkningsfremskrivninger fra SSB. MSG6-anslagene på offentlige utgifter til ulike pensjoner og overføringer, før indeksering, hentes fra en dynamisk mikrosimuleringsmodell som gir et meget detaljert bilde av pensjonsreglene og økonomiske livsløp for et relativt stort utvalg av befolkningen. Detaljerte skattemodeller benyttes til å estimere gjennomsnittlige og marginale skattesatser. Fremskrivningene av offentlige sysselsetting innenfor ulike servicesektorer baseres på en spesialmodell som gjør det mulig å dekomponere veksten i ressursbruk i bidrag fra endringer i bl.a. antall individer i ulike aldersgrupper og standard (timeverk per bruker).

Når det ikke er tilbakevirkninger fra variablene som bestemmes i MSG6 til mekanismene i spesialmodellene, oppstår ingen inkonsistens ved en slik rekursiv bruk av flere modeller. Der inkonsistensen kan være alvorlig, er virkningen fra opptjening av individuelle inntekter til pensjonsrettigheter, arbeidstilbud og privat sparing. Dette problemet er forsøkt redusert gjennom iterativ (i stedet for ren rekursiv) modellbruk. Tapet av konsistens må veies opp mot at spesialmodellene fanger opp detaljer som ville forsvinne i en mer aggregert modell. Disse detaljene påvirker ikke bare nøyaktigheten i resultatene, men også hvorvidt modellapparatet samlet sett er operativt i forhold til å spesifisere f.eks. konkrete endringer i pensjonssystemet. Samlet sett har vi i SSB konkludert med at vi vil bygge opp en OLG-modell a la DREAM for analyser relatert til fiscal sustainability.

#### **4. Er de norske skattesatsene og velferdsordningene opprettholdbare?**

Som sammenligningsgrunnlag for analyser av ulike reformer ble det utarbeidet en referansebane basert på bl.a. videreføring av dagens pensjons- og skattesystem. Viktige trekk ved denne banen er:

– Beregningene er basert på middelalternativet i befolkningsfremskrivningene fra

SSB.<sup>4</sup> Her øker forventet levealderen for menn fra 77,0 år i 2003 til 84,2 år i 2050, dvs. 7,2 år. Den tilsvarende økningen for kvinner er 6,2 år, fra 81,9 år til 88,1 år. I gjennomsnitt øker dermed middellevetiden med ca. 0,14 år per år, mens den tilsvarende økningen i VFKs grunnforløp er 0,08-0,09 år per år. Tallet på alderspensjonister mer enn doubles over samme periode. Forholdstallet mellom antallet i aldersgruppen 20-66 år og antallet eldre enn 66 år faller fra 5 i dag til 3 i 2050. Sysselsettingen endres lite over beregningsperioden. Målt i timeverk er sysselsettingen 11,1 prosent høyere i 2050 enn i 2002.

- Som i VFKs grunnforløp forutsettes konstant standard på individuelle offentlige tjenester. Aldringen fører til at offentlig konsum i faste priser vokser med 0,6 prosent per år frem til 2020. Deretter øker vekstraten til mellom 1,4 og 1,1 prosent frem til 2050.
- Til tross for aldringen, kan det private forbruket per innbygger vokse med en gjennomsnittlig vekstrate på 2,8 prosent i perioden frem til 2050, hvilket impliserer en doubling i løpet av 25 år. Den tilsvarende realveksten i BNP er 1,7 prosent. Det er produktivitsveksten sammen med forutsetningen om svak vekst i offentlig konsum som er avgjørende for veksten i rommet for privat forbruk.<sup>5</sup> I MSG6 er arbeidsproduktiviteten endogen. Den vokser med ca. 2,5 prosent som årlig gjennomsnitt. Det er ca. 0,5 prosentpoeng mer enn den tilsvarende veksten i VFKs grunnforløp. En naturlig tolkning er at en økende og del av veksten i privat forbruk omfatter markedsbaserte helse- og omsorgstjenester.
- Den nominelle renten er satt til 5,5 prosent, mens VFK forutsetter 5,0 prosent. Vi har forutsatt at verdensmarkedsprisene på *traded goods* vokser årlig med 1,5 prosent. I MSG6 vil lønnsveksten ligge nær summen av denne prisveksten og veksten i arbeidsproduktiviteten, dvs. på ca. 4,0 prosent. Den tilsvarende hovedmekanisme gjelder også i DREAM. Også i VFKs grunnforløp burde lønnsveksten ligge nær 4,0 prosent som følge av 2,0 prosent vekst i arbeidsproduktiviteten og 2,0 prosent internasjonal inflasjon. Som jeg kommer tilbake til avslutningvis, har forskjellen mellom rente og lønnsvekst avgjørende betydning for beregningene av holdbar skatt.

Mens AP benytter endringen i den danske bundskattesatsen som mål på finansieringsbehovet, tilpasses arbeidsgiveravgiften endogent i våre beregninger. For å oppfylle handlingsregelen for finanspolitikken må arbeidsgiveravgiften økes hvert år etter 2020 fra dagens gjennomsnitt på 13,1 prosent. Satsen vil passere 25 prosent rundt

4. Statistisk sentralbyrå (2002): Befolkningsframskrivninger. Nasjonale og regionale tall, 2002 – 2050. [www.ssb.no/emner/02/03/folkfram/](http://www.ssb.no/emner/02/03/folkfram/).

5. Produktiviteten av *alle* innsatsfaktorer (Total Factor Productivity) som brukes i private bedrifter er antatt å øke med 1,3 prosent per år.

2050, og det i relativt stor fart.<sup>6</sup> Dagens norske skattesatser og velferdsordninger kan med andre ord ikke opprettholdes på lang sikt. En viktig årsak til dette er at levealderen fortsetter å øke i hele dette århundret, samtidig som dagens pensjonssystem er ikke-aktuarisk i den forstand at et økende antall år som pensjonist ikke påvirker den årlige pensjonsytelsen. Våre beregninger bekrefter et viktig poeng hos AP: Fortsatt økonomisk vekst vil i seg selv ikke finansiere veksten i offentlige utgifter, gitt at produktivitsveksten spres til pensjonistene via lønnsindekserte pensjoner. I våre beregninger vil økt produktivitet endog øke de lønnsavhengige utgiftene mer enn skatteinntektene pga. omfanget av offentlige rente- og petroleumsinntekter i budsjettbalansen. Budsjettvirkningen av økt produktivitet, og dermed lønn, avhenger imidlertid av hvordan overføringene indekseres. I USA er prisindeksering det vanligste, slik at produktivitsvekst der vil ha en mer positiv budsjetteffekt enn i Norge. I forslaget til pensjonsreform i Norge vil man indeksere pensjonsytelsene i forhold til et uveid gjennomsnitt av lønns- og prisvekst, i stedet for lønnsvekst.

Perspektivmeldingen viser hvordan det beregnede udekkede finansieringsbehovet fremover avhenger av usikre forhold som oljepris, rente, demografi, produktivitsvekst, arbeidstid og helsetilstand. Beregningene illustrerer at det vil oppstå et betydelig udekket finansieringsbehov innenfor vide intervaller av anslag for disse størrelsene. Finansieringsbehovet kan unngås ved et heldig sammenfall av utfall, mens uheldige sammenfall kan gjøre det meget stort. Det konkluderes (s. 93): »Samlet etterlater beregningene liten tvil om at det er påkrevd med nye tiltak for å sikre den langsiktige balansen i offentlige finanser.«

## 5. Reformen

Det er i Norge som i Danmark tunge trender som bidrar til å skape et voksende offentlig budsjettunderskudd dersom politikken ikke legges om. Problemenes art og størrelsesorden krever endringer i de »regelstyrte« utgiftene og inntektene. Da snakker man i praksis om pensjonsreformer, lavere ambisjonsnivå for offentlig finansiering av service innenfor helse og omsorg, og/eller skattereformer. I det følgende gir jeg en kort oversikt over noen relevante norske analyser av pensjons- og skattereformer.

### *Betydningen av pensjonssystemet*

Også Norge har sin pensjonsdebatt. En bredt sammensatt Pensjonskommisjon leverte sin innstilling i 2004.<sup>7</sup> Den avtroppende Bondevik-II regjeringens reformforslag lå tett opp til kommisjonens hovedforslag »Modernisert folketrygd« (MF). Sentrale

6. Perspektivmeldingen opererer med et lignende basisforløp og anslår et udekket finansieringsbehov i 2050 tilsvarende 5,5 prosent av BNP for fastlands-Norge.

7. NOU 2004:1 »Modernisert folketrygd«.

elementer i dette forslaget omfatter: (a) Folketrygden gjøres mer aktuarisk ved at opp-tjent pensjonsformue på avgangstidspunktet må fordeles på forventet antall år som pensjonist. Det innebærer at fritiden som følge av økende levealder og førtidspensjonering i større grad må betales av den enkelte. Videre skal dagens »besteårsregel« erstatte av en »alleårsregel« slik at alle inntektskroner (innenfor visse grenser) skal gi pensjonsrettigheter. (b) Dagens lønnsindeksering erstattes av at årlig pensjonsytelse oppjusteres med et gjennomsnitt av lønns- og prisvekst. (c) Folketrygden skal fortsatt finansieres PAYGO, dvs. at den samlede fonderingen av samlede fremtidige offentlige utgifter fortsatt skjer i henhold til handlingsregelen. Stortingets pensjonsvedtak i mai 2005 aksepterte viktige elementer i forslaget, men regler for bl.a. førtidspensjonsordninger og opptjening av pensjonsrettigheter er for tiden gjenstand for nye utredninger.

I Fredriksen, Heide, Holmøy og Solli (2005)<sup>8</sup> beregnet vi virkningene av Pensjonskommissjonens forslag. Når det gjelder virkningene MF er hovedkonklusjonen at reformen vil gi et meget stort bidrag til å løse det offentliges finansieringsproblem; rommet for å senke arbeidsgiveravgiften øker over tid i takt med økningen i levetid. I 2050 kan denne skattesatsen senkes fra 25 prosent i 2005 ved videreføring av dagens system til 11 prosent med MF, mot ca. 13 prosent i dag. Reduksjonen er mulig fordi arbeidstilbudet øker med nær 10 prosent, og fordi de offentlige utgiftene til alderspensjon målt i faste kroner reduseres med ca. 16 prosent i 2050. Sterkere incentiver til å utsette avgangen bidrar direkte til å øke arbeidstilbudet med 4,1 prosent i 2050. Inkluderes likevektseffektene via redusert skatt og en viss nedgang i reallønn, kan 5,6 prosentpoeng av den totale sysselsettingsøkningen i 2050 føres tilbake til økt avgangsalder. I tillegg kommer bidrar sterkere sammenheng mellom arbeidsinnsats og pensjon til å øke arbeidstilbudet med 4,2 prosentpoeng via økt effektiv reallønn. Forslaget om å bygge inn en mekanisme i regelverket som i realiteten overfører det meste av kostnaden ved å ha fritid som pensjonist fra skattebetalerne til den enkelte, kan ses som et ledd i den tilpasningsstrategi som AP anbefaler. Spesielt vil trendmessig økning i midlellelivetiden automatisk finansieres dersom regelverket først er innført. Imidlertid er nettopp dette elementet i pensjonsreformen meget kontroversielt, og den regelendringen som ble forutsatt i våre beregninger, blir sannsynligvis ikke vedtatt.

#### *Betydningen av arbeidstilbud og skattesystem*

Som et element i det såkalte »velferdsdilemma« hevder AP at videre økonomisk vekst vil bidra til å øke etterspørselen etter fritid. En slik negativ partiell inntektseffekt på arbeidstilbudet står sterkt både teoretisk og empirisk. I Norge pusses det nå (i forbindelse med Stortingsvalget og rekordhøye oljepriser) støv av tidligere forslag om

---

8. Fredriksen, D., K. M. Heide, E. Holmøy og I. F. Solli (2005): Makroøkonomiske virkninger av pensjonsreformer. Beregninger basert på forslag fra Pensjonskommissjonen. Rapport 2005/2, Statistisk sentralbyrå.

innføring av seks-timers dag med full lønnskompensasjon. Den langsiktige trend i favør av mer fritid kan imidlertid også skyldes forhold som i sum har økt den nytte man kan ha av fritid. Den økonomiske veksten i fremskrivningene går også hånd i hånd med en sterk økning i konsumentenes reallønn etter skatt som isolert sett vil det stimulere arbeidstilbudet. Hvilken effekt som vil dominere, vil være et empirisk spørsmål. I Aaberge, Colombino, Holmøy, Strøm og Wennemo (2004)<sup>9</sup> belyser vi dette spørsmålet ved å kombinere MSG6 med en meget detaljert mikroøkonometrisk modell for individuelt arbeidstilbud. Arbeidstilbudsmodellen representerer på mange måter state-of-the-art på feltet, men inkluderer ikke valg av pensjoneringstidspunkt. Vi finner at nettovirkningen er en økning på 4,2 prosent i samlet arbeidstilbud i 2050 når vi tar hensyn til at individene tilpasser seg til at deres reallønnsatts og reelle arbeidsuavhengige inntekt da er henholdsvis 209 og 190 prosent høyere enn i 1995. Dette gjør det mulig å senke anslaget på arbeidsgiveravgiften i 2050 fra 25 til 21 prosent. Beregningene i Aaberge *et al.* (2004) viser imidlertid at skattereformer som går i retning av en utflating av skattesystemet kan gi betydelige lettelser i den fremtidige skattebyrden. Dersom man som et stilisert eksempel beskattet alle inntektskroner, uansett inntektskilde, med en felles flat skattesatt, kunne denne satsen settes til 18,3 prosent i 1995, mens den tilsvarende gjennomsnittssatt da var 24,0 prosent. Med et flatt skattesystem ville skattesatsen måtte økes til 22,9 prosent i 2050.

## 6. Noen metodiske utfordringer

Ved beregning av en konstant »holdbar skatt« som sørger for at nåverdien av det primære budsjettunderskuddet er lik dagens offentlige netto finansformue, er tidshorisonten uendelig. Ved lav effektiv diskonteringsrente konfronteres man da *eksplicit* med noen grunnleggende problemer som kan gjøre resultatene vanskelige å tolke. Jeg velger å avslutte med en kort forklaring av disse problemene fordi jeg mener at de har generell relevans.

Anta for enkelthets skyld at enhver utgifts- og inntektskomponent i den primære budsjettbalanse vokser med en konstant vekstrate i nominell verdi. Nåverdien beregnet over en uendelig horisont fra periode 0 av komponent  $X$  er da lik  $X_0/(r-g_X)$  der  $r$  er nominell renten og  $g_X$  er den nominelle vekstraten for  $X$ . Vi har benyttet 5,5 prosent som et realistisk estimat på renten. I våre fremskrivninger vil vekstratene for de viktigste offentlige inntekter og utgifter ligge nær 4,5 prosent. Det viktigste bidraget til dette er en nominell lønnsvekst på 4,0 prosent.<sup>10</sup> I tillegg vokser utgiftene som følge av al-

9. Aaberge, R., U. Colombino, E. Holmøy, B. Strøm og T. Wennemo (2004): Population Ageing and Fiscal Sustainability: An Integrated Micro-Macro Analysis of Required Tax Changes. Discussion Papers 367, Statistisk sentralbyrå.

10. Lønnsveksten på 4,0 prosent per år er i hovedsak et resultat av prisene på tradet goods vokser med 1,5 prosent og at den endogene veksten i arbeidsproduktivitet er ca. 2,5 prosent.



dring og modning av Folketrygden. Den resulterende vekstkorrigerede rente blir da så lav som 1 prosent.<sup>11</sup> Lav effektiv diskontering betyr at en meget fjern og ukjent fremtid teller tungt i nåverdiberegningene. Nåverdiene blir meget sensitive for meget små endringer i  $X_0$ , og særlig i  $r$  og  $g_X$ . Begge deler gjør resultatene vanskelige å tolke og bruke. Det er heller ikke opplagt at vi generelt har  $r-g>0$ . Det er nok at  $r-g<0$  for en komponent for at en holdbar skatt ikke lar seg beregne. Spesielt kan dette gjelde enkelte utgiftskomponenter, uten at dette har noe å gjøre med at økonomien som helhet er dynamisk ineffisient.

VFK forutsetter en nominell rente på 5,0 prosent, 0,5 prosentpoeng lavere enn vårt estimat. Begge anslag er i godt samsvar med tilsvarende forutsetninger i andre studier. Avkastningen på offentlig finansformue får et tillegg pga. meravkastning på aksjeandelen av formuen. Jeg er kritisk til at denne meravkastningen inkluderes i  $r$ . I den grad den reflekterer en risikopremie, bør risikopremien tas med på utgiftsdsiden – ikke bare som ekstra avkastning. Selv den meravkastningen som ikke kan forklares med risiko («Equity Premium Puzzle»), er problematisk. Det er jo et faktum at aktiva med 5,0 prosent avkastning overlever som et konkurransedyktig plasseringsalternativ. I beregningene synes det som om man kan skaffe den danske stat en gratis lunsj ved å plassere alle fordringer i det aktiva som har høyest rente. Inflasjonen påvirker  $r$  og alle  $g$ 'ene like mye og er irrelevant for  $r-g$ , så lenge man – korrekt – benytter samme deflator for alle inntekter og utgifter. Den nominelle lønnsveksten i VFKs grunnforløp synes å være meget nær de 4,0 prosent vi får i våre norske beregninger.<sup>12</sup> Dette skulle bety at også de danske beregninger av holdbar skatt møter de samme problemene med ikke-robuste resultater, iallfall hvis renten settes lik 5,0 prosent på all offentlig finansformue.

De problemene som følger av lav effektiv diskontering er grunnleggende; de er *ikke* knyttet spesifikt til beregning av holdbar skatt som metode. Ved beregninger av holdbar skatt konfronteres man eksplisitt med disse problemene siden metoden per definisjon har en uendelig tidshorisont. Dette betyr selvsagt ikke at det er galt å anlegge et uendelig langt tidsperspektiv. Tvert imot: Lav effektiv diskonteringsrate gjør at forhold langt frem i tid får – og skal ha – stor vekt. Også analyser basert på PAYGO finansiering av et løpende krav til offentlig budsjettbalanse møter de samme grunnleggende problemer dersom fremskrivningene føres langt nok frem i tid. At man i praksis avbryter fremskrivningene på et arbitrært tidspunkt før disse problemene er blitt synlige, er i så måte en svakhet ved PAYGO-analysen. Hvis økonomien på dette tidspunktet er inne i en trendmessig utvikling der skattesatsen må økes hvert eneste år, kan man si mer enn at dagens skattesatser/velferdsordninger (eller Norges finanspolitiske handlingsregel) ikke kan opprettholdes; man kan i tillegg si at utviklingen tilhører en bane som på sikt

---

11. Skatteeffekter tilsier ytterligere reduksjon av den vekstkorrigerede renten.

12. VFK antar 2,0 prosent vekst i arbeidsproduktivitet og 2,0 prosent utenlandsk inflasjon.

ikke kan realiseres som verken en matematisk eller realistisk likevekt. Her er det flere forhold enn suksessive økninger i skattesatsen(e) som kan føre til ikke-eksistens. I de norske beregningene som er nevnt over, legger offentlig sektor beslag på en stadig større del av sysselsettingen. Det vil kun være et spørsmål om tid når en slik trend ikke lenger lar seg etablere som en markedslikevekt.

Mer generelt: fremskrivninger med en uendelig tidshorisont vil før eller senere kollapse dersom ikke modellen selv genererer en steady-state utvikling der alle størrelser vokser med en felles rate (som er mindre enn renten). Steady state vil ikke nås uten restriktive forutsetninger om modellens virkemåte og eksogene anslag. Men når perspektivet blir svært langsiktig, er alternativet til steady state en form for kollaps som er enda mer urealistisk. Når det gjelder offentlige utgifter, er det imidlertid ingen markedsmekanismer som sikrer at veksten i disse utgiftene bringes ned på den uniforme vekstraten i steady state. I stedet må man (implisitt) forutsette at det »ikke for sent« skjer en korleksjon av politikken. En politikkreksjon som innebærer nedjustering av veksten i offentlige utgifter fra et arbitrært fremtidig tidspunkt, bidrar i seg selv til å løse eller redusere holdbarhetsproblemet som beregningene skal belyse når den effektive neddiskontering er svak. Valget av tidspunkt for en slik justering vil ha noe av det samme arbitrære preget over seg som valget av endelig tidshorisont i en PAYGO-fremskrivning. Løsningen for den holdbare skatt er med andre ord betinget på et eksogent valg av tidshorisont for hvor lenge trender kan få lov til å avvike fra steady state veksten.

### 7. Avslutning

De problemene som her er mer skissert enn analysert, rokker likevel ikke ved at beregningene fra VFK bør tillegges stor vekt i den danske debatten. En imponerende mengde detaljert og relevant informasjon er systematisert og transformert via DREAM og god modellbruk til meget opplysende nøkkeltall. De valgene som *må* foretas ved konkrete beregninger vil alltid kunne diskuteres. Som sagt, tror jeg at fremskrivningene fra VFK heller undervurderer problemene enn det motsatte. Den ovennevnte påpekning av at ikke-holdbare veksttrender er avbrudt »før det er for sent« trekker i samme retning. Det bør også være interessant i den danske debatt at også Norge – med sin store statlige petroleumsformue – står overfor problemer med å finansiere velferdsstaten som er av samme art og størrelsesorden som i Danmark. Det er begrenset i hvor stor grad et *engangsskift* i form av en valutagave kan demme opp for den langsiktige virkningen av tunge veksttrender.

## Debat om fremtidens velfærd – opsamling og replik

Torben M. Andersen

Institut for Nationaløkonomi, Aarhus Universitet, *E-mail: tandersen@econ.au.dk*

Lars Haagen Pedersen

Velfærdskommissionens sekretariat, *E-mail: lhp@velfaerd.dk*

Formålet med de fremskrivninger, som er fremlagt i vores artikel, »Demografi, velstands dilemma og makroøkonomiske strategier« i dette tidsskrift er at skabe et bedre grundlag for en diskussion af tilrettelæggelsen af den økonomiske politik. Dette indebærer både en vurdering af størrelsesordener og af usikkerhed/følsomheder. Samtidig må man nødvendigvis tage udgangspunkt i en fortolkning af uændret økonomisk politik. I vores beregninger har vi valgt at illustrere disse problemstillinger ved ændringer i bundskattesatsen. Som det gentagne gange er fremhævet, er dette ikke for at anbefale en endimensionel politikreaktion, men for at formulere størrelsesordener i en variabel, der direkte kan kontrolleres af politikerne.

Det er ikke formålet med fremskrivningerne at lave egentlige prognoser eller at forudsige fremtidige politikreaktioner, men netop at skabe et grundlag for at diskutere økonomisk politiske initiativer. Det er derfor ikke relevant at indarbejde fremtidige politikændringer, som ønsket af Christen Sørensen (2005) (herefter CS) og Claus Vastrup (2005) (herefter CV). Sådanne ændringer i den økonomiske politik kan derimod eventuelt blive resultatet af de analyser, vi har fremlagt.

Det følgende vil derfor alene kommentere den diskussion, der direkte vedrører beregningerne, forudsætningerne herfor, fortolkning af resultaterne samt analysemetoderne.

Vores analyse viser, at indretningen af det danske velfærdssamfund udfordres dels af den stigende levealder og den heraf følgende permanente ændring i befolkningens alderssammensætning og dels af den stigende velstand. Begge ændringer er således positive udfordringer for den fremtidige velfærd i Danmark. Diskussionen om den fremtidige indretning af velfærdssamfundet er derfor en fordelingsdiskussion, og ikke et nedskærings- eller fallitscenario, som det ofte fremstilles.

Der er flere dimensioner i fordelingsdiskussionen. For det første er det en diskussion om den intergenerationelle fordeling. Hvor meget skal de enkelte generationer modtage af offentlige ydelser, og hvor meget skal generationerne bidrage? CS og Erling Holmøy (2005) (herefter EH) tager denne del af problemstillingen op. En anden vigtig del af fordelingsdiskussionen handler om størrelsen og sammensætningen af de offentlige udgifter, dvs. om størrelsen af skatte trykket og om forholdet mellem overførselsindkomster og offentlig service. Velfærdskommissionen (herefter VFK)

har fremhævet, at tilpasningen i den offentlige sektor især bør ske ved at reducere antallet af overførselsindkomstmottagere og øge antallet af beskæftigede i den private sektor. Årsagen til, at der er behov for denne type tilpasning, er, at kombinationen af gældende regler og et stærkt stigende antal ældre fører til, at en faldende andel af befolkningen er i privat beskæftigelse, og at en stigende andel af befolkningen vil modtage overførselsindkomst. Herunder vil især antallet af folkepensionsmottagere stige.

Jørgen Goul Andersen (2005) (herefter JGA), Jesper Jespersen (2005) (herefter JJ) og CS diskuterer alle, hvor stor en effekt denne ændring i befolkningens størrelse og sammensætning vil have for fordelingen af offentlige udgifter og indtægter i fremtiden. Derimod synes der ikke at være nogen grundlæggende uenighed om, at tilpasningen især skal ske ved at søge at reducere antallet af overførselsindkomstmottagere og forøge antallet af beskæftigede i den private sektor.

I tillæg til disse diskussioner om tilrettelæggelsen af den fremtidige økonomiske politik rejser især CV og JJ spørgsmål om den makroøkonomiske analyse, der ligger til grund for politikanbefalingerne.

Denne replik er derfor opdelt i tre dele. Første del vedrører den intergenerationelle fordeling, mens anden del vedrører forudsætninger bag VFKs fremskrivning og dermed den fremtidige fordeling af offentlige udgifter og indtægter. Herunder behandles betydningen af forskellige antagelser vedrørende bl.a. demografi og arbejdsstyrke i VFKs analyser. Tredje del behandler de metodemæssige diskussioner.

### **Intergenerationel fordeling**

I Norge er der truffet en politisk beslutning om, at indtægterne fra olie- og gasudvindingen opsamles i en petroleumsfond. For at sikre, at alle fremtidige generationer får del i gevinsterne fra den udtømmelige ressource i Nordsøen, udbetales principielt »kun« realafkastet af formuen til dækning af offentlige udgifter i et givet år. EH påpeger i sit indlæg, at det er et betydeligt politisk problem at fastholde målsætningen om kun at udbetale beløb svarende til realafkastet. Dette er en vigtig observation i forbindelse med diskussionen af den fremtidige økonomiske strategi i Danmark: Det er vanskeligt at fastholde en positiv offentlig opsparing i en situation, hvor den offentlige sektor i forvejen har en positiv formue, og hvor der føres en relativ stram finanspolitik.

Dette problem må forventes at blive mindst lige så stort, hvis Danmark vælger en opsparingsstrategi, og der akkumuleres en offentlig formue til finansiering af fremtidige udgifter. Det hænger sammen med, at der i Norge er et »retfærdighedsargument« for at sikre, at de »ekstraordinære« indtægter fra de udtømmelige ressourcer i Nordsøen i disse år kommer alle fremtidige generationer til gode. I Danmark findes ikke »ekstraordinære« indtægter i samme omfang som i Norge. Hvis Danmark gennemfører en opsparingsstrategi, vil der derfor ikke være så stærkt et retfærdighedsargument som i

Norge.<sup>1</sup> Tværtimod kan der argumenteres for, at nuværende generationer kommer til at medfinansiere, at fremtidige generationer kan opnå en større andel af livet som tilbagetrukket fra arbejdsmarkedet.

Den norske beslutning om fordeling af afkastet betyder, at alle fremtidige generationer får andel i værdien af den udtømmelige ressource fra Nordsøen. Men reglen betyder også, at fremtidige generationer vil kunne finansiere en fortsat faldende andel af de reale offentlige udgifter med afkastet fra petroleumsfonden, hvis de offentlige udgifter vokser i takt med væksten i økonomien. Det øger den norske offentlige sektors følsomhed over for vækst, fordi en betydelig del af grundlaget for fremtidige offentlige indkomster bliver mere uafhængigt af væksten i økonomien.<sup>2</sup>

### Rentediskussionen

I fravær af retfærdighedsargumenter for en opsparingsstrategi kommer diskussionen om opsparingsstrategi kontra tilpasningsstrategi i høj grad til at være en afvejning af gevinsten ved opsparing over for den fordelingsmæssige byrde, som opsparingsstrategien lægger på nuværende generationer. Hvis realrenten korrigeret for vækst er høj, trækker det alt andet lige i retning af opsparingsstrategien, fordi det forventede afkast af opsparingen er høj.<sup>3</sup>

VFK har i sin fremskrivning anvendt en realrente korrigeret for vækst på 0,9 for obligationer. Det er lavere end andre tilsvarende fremskrivninger fra f.eks. Finansministeriet og Det Økonomiske Råd. Der er således med VFKs antagelser mindre stærke argumenter for en opsparingsstrategi end i de øvrige fremskrivninger.

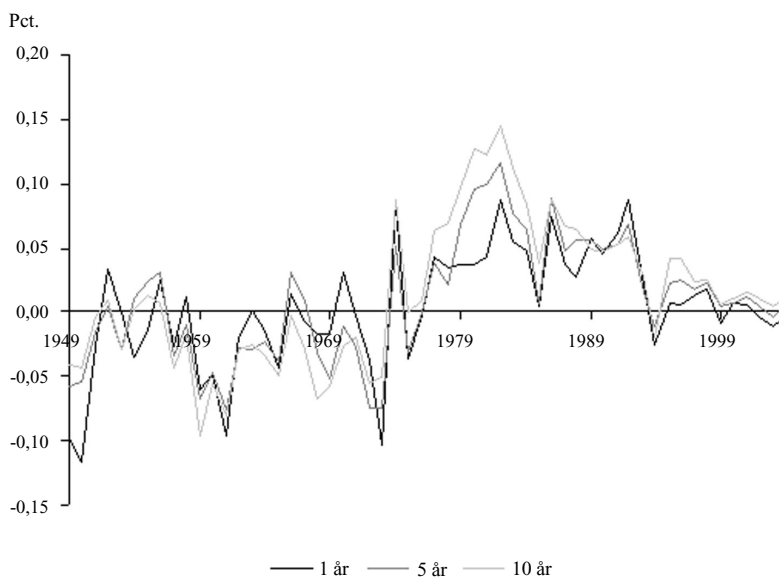
Med hensyn til valget af niveauet for den vækstkorrigerede realrente har VFK argumenteret for, at det er mere sandsynligt, at den fremtidige realrente ligger under det historiske gennemsnit end over. Årsagen er, at det historiske gennemsnit i meget høj grad er påvirket af perioden fra 1970 til 1990, hvor den internationale realrente var høj, og hvor Danmark havde et meget betydeligt positivt rentespænd til verdensmarkedsrenten på grund af devalueringspolitikken op gennem 1970'erne og frem til 1982. Med en antagelse om, at de førende nationalbanker fastholder en pengepolitik baseret på en inflationsmålsætning, og at Danmark fastholder fastkurspolitikken, der blev etableret i 1982, må det derfor forventes, at realrenten bliver lavere end det historiske gennemsnit, der lægges til grund for VFKs fremskrivning.

---

1. Der kan argumenteres for at der også er ikke ubetydelige indtægter fra den danske del af Nordsøen og ekstraordinære indtægter, som skyldes, at de nuværende erhvervsaktive generationer er store relativt til både de kommende og de forrige.

2. Argumentet forudsætter, at vækstændringer ikke fuldt ud modsvares i ændringer i realrenten.

3. Følsomhedsberegningen af en stigning i renten på 1 procentpoint viser, at den permanente skattestigning kan reduceres fra 8,7 til 5,0 procentpoint på grund af det højere afkast af den offentlige opsparing. På baggrund af JJs kommentar skal det bemærkes, at denne beregning tager højde for de afledte aktivitetsvirkninger af rentestigningen, som reducerer investeringerne permanent og det private forbrug på kort sigt.



Figur 1. Den vækstkorrigerede realrente for en obligation med 1, 5 og 10 års løbetid.

Kilde: Egne beregninger.

Disse forhold er alle medvirkende til, at VFK foreslår en tilpasningsstrategi, hvor der indgår et mere begrænset element af opsparing (gældsnedbringelse), mens den resterende del af tilpasningen sker ved gradvise på forhånd annoncerede politikændringer, eksemplificeret ved kommissionens forslag om levetidsindeksering af pensionsalderen. Det er med glæde, at vi noterer, at CS accepterer VFKs argumentation, fordi det kan betyde, at debatten fremadrettet i højere grad kommer til at dreje sig om, hvordan tilpasningsstrategien bedst udformes.<sup>4</sup>

### Vækstdiskussionen

Spørgsmålet om størrelsen af den fremtidige vækst spiller en tilsvarende rolle for diskussionen om opsparings- kontra tilpasningsstrategi. For det første betyder en højere økonomisk vækst – alt andet lige – en lavere vækstkorrigeret realrente. Som ovenfor betyder dette, at afkastet af en given opsparing vil veje mindre i forhold til den fremtidige produktion. For konstant udgiftstryk skal der således en større opsparing til at finansiere en given andel af de fremtidige udgifter. Høj vækst betyder således, at opsparingsstrategien bliver mindre relevant.

4. Det giver dog anledning til en vis bekymring, at CS på trods af den erklærede enighed om renteutviklingen alligevel fastholder tabellen, hvor han korrigerer VFKs resultater mht. den finanspolitiske holdbarhed for bl.a. en stigning i renten, som er langt den største korrektion i tabellen.

For det andet er øget vækst selv ved en tilpasningsstrategi med løbende finansiering af de årlige udgifter ikke helt neutral for det offentlige budget. Indretningen af det danske velfærdssamfund gør, at øget vækst ikke alene øger skatteindtægterne, men også de offentlige udgifter. Det sidste følger af, at offentlige lønninger følger lønudviklingen på det private arbejdsmarked, og at reguleringen af overførselsindkomster er baseret på den private lønudvikling, jf. nedenfor om satsreguleringen. Disse egenskaber betyder, at såvel offentlige udgifter som indtægter vokser parallelt med den økonomiske vækst. Som udgangspunkt fører vækst således hverken til en forbedring eller en forværring af det offentlige finansieringsproblem. Det er imidlertid ikke alle offentlige indtægter, der udvikler sig proportionalt med den generelle indkomstudvikling i samfundet. Den andel af indtægterne, der stammer fra beskatning af udbetalte pensionsmidler, er i sagens natur baseret på indbetalinger ud af tidligere perioders indkomster. Derfor vil denne offentlige indtægtskilde ikke vokse, hvis den fremtidige vækst i samfundet stiger. Af denne grund betyder øget vækst en begrænset forværring af det offentlige finansieringsproblem selv ved en løbende finansiering.

Både CS og JJ hæfter sig ved, at en permanent forhøjelse af den årlige vækst fører til, at kravet til den permanente skattestigning vokser. CS kalder resultatet paradoksalt, mens JJ tilskriver det særlige egenskaber ved DREAM modellen. Begge dele er forkerte. Resultatet følger af anvendelsen af simple budgetbetingelser, jf. Velfærdskommissionen (2005c). Pointen er, at fremtidige underskud vokser stort set proportionalt med vækstraten. Med en stigning i væksten vokser den tilbagediskonterede værdi af fremtidige underskud for en given rente. Hvis nuværende generationer skal finansiere en uændret andel af disse underskud via det nuværende overskud, må dette overskud stige. Kravet til skattestigningen er derfor større.

CS finder, at pointeringen af tilpasningsstrategien er en betydelig ændring i forhold til VFKs første delrapport. Det er beklageligt, hvis kommissionens publikationer har kunnet bibringe det indtryk. Det er ikke vores (og heller ikke kommissionens) opfattelse. I den nævnte rapport, Velfærdskommissionen (2004), afsluttes sammenfatningen om tilrettelæggelsen af den økonomiske politik med en diskussion af de to strategier (s. 28): *»Som alternativ til opsparingsstrategien kan man vælge en strategi, som kunne kaldes »regulering med levetid«. Denne strategi er baseret på, at væksten i midlellelivetiden er hovedforklaringen på holdbarhedsproblemet ... En sådan reform vil betyde, at behovet for, at den offentlige sektor opbygger formue reduceres betydeligt«* og videre: *»Herudover er der et spørgsmål om rimeligheden ved, at nuværende generationer, der både er relativt fattigere end fremtidige generationer og har en kortere forventet levetid, skal spare op til, at fremtidige generationer kan have en større andel af deres liv som tilbagetrukket fra arbejdsmarkedet.«*

Synspunktet er endvidere uddybet i debatoplægget »Fremtidens velfærd – og globaliseringen« fra marts 2005, der afsluttes med et afsnit som diskuterer opsparingsstrategi eller tilpasningsstrategi. Heri konkluderes »*En opsparingsstrategi kan derfor ikke stå alene. Alternativet er en tilpasningsstrategi, som Velfærdskommissionen tidligere har fremført*«, jf. Velfærdskommissionen (2005b).

### **Terminalproblemet**

CS rejser endvidere spørgsmål om det, der betegnes »terminalproblemet«. Som vi forstår begrebet, er pointen, at nuværende generationer ud over en finansiell gæld også efterlader realkapital i form af f.eks. offentlige bygninger og andet kapitalapparat til fremtidige generationer. Hverken VFK eller andre institutioner medregner dette i forbindelse med beregninger af den finanspolitiske holdbarhed. Det skyldes, at beregninger af den finanspolitiske holdbarhed tager udgangspunkt i, at den offentlige sektor i fremtiden stiller den samme type offentlige ydelser til rådighed for borgerne. Den offentlige sektor kunne derfor vælge at frasælge de reale aktiver og leje nogle tilsvarende i den private sektor og således producere fremtidige ydelser ved hjælp af de lejede aktiver. Når der derfor kan ses bort fra aktiverne i holdbarhedsberegningen, skyldes det en forudsætning om, at prisen på markedet for aktiverne afspejler værdien af den fremtidige indtægt, der kunne opnås ved udlejning af aktiverne.<sup>5</sup> »Terminalproblemet« er derfor ikke relevant for holdbarhedsberegningen.

»Terminalproblemet« er imidlertid relevant i forbindelse med diskussionen af den intergenerationelle fordeling. Jo større værdien af de reale offentlige aktiver er, jo større er den fremtidige »besparelse på lejeudgifter«. Det vil derfor typisk være tilfældet, at der er en positiv »arv« i form af offentlige aktiver fra en generation til den næste – også selvom den offentlige finansielle gæld er vokset som procent af BNP. Hvis man ønskede at finde en »fair« fordeling mellem generationer ud fra et givent kriterium, ville man ikke kunne se bort fra de reale offentlige aktiver. Som beskrevet i den anden artikel er VFK afstået fra sådanne overvejelser og har i stedet anlagt det mere pragmatiske synspunkt, at der ikke hidtil har været gjort forsøg på systematiske omfordelinger mellem generationer, der ikke umiddelbart overlapper. VFK har derfor lagt til grund, at det heller ikke i fremtiden vil indgå i politikernes prioritering, i modsætning til hvad der har været tilfældet i Norge.

### **Velfærdskommissionens fremskrivning og fordelingen mellem offentlige ydelser og overførselsindkomst**

VFK forsøgte med sin første rapport at beskrive det fremtidige pres på den offentlige sektors finansiering dels gennem lavere skattebetalinger pga. færre personer i den

---

5. Det følger af ovenstående, at hvis en offentlig ydelse ikke længere produceres (dvs. privatisering), da vil der optræde en engangsgevinst ved frasalg af de til produktionen knyttede reale aktiver.



erhvervsaktive alder og dels gennem øget pres for udgifter til både flere overførselsindkomstmodtagere og flere personer i de aldersgrupper, der er hoveddaftagere af offentlige ydelser. Samtidig forsøgte kommissionen via følsomhedsberegninger at vise, hvilke forudsætninger, der er kritiske for resultatet, og hvilke, der betyder mindre. Efterfølgende har der været en betydelig debat, som i høj grad har drejet sig om betydningen af forudsætninger i VFKs fremskrivning.

### **Arbejdsmarkedspensioner og private pensioner**

Denne debat videreføres i nogen grad af JGA. Hovedsynspunktet er her, at den demografiske udfordring er lille i Danmark, og at indførelsen af arbejdsmarkedspensioner har løst størstedelen af dette problem »fordi man i ualmindelig høj grad har op hævet denne (sociale (vores tilføjelse)) kontrakt«. Dvs. at fremtidige generationer af erhvervsaktive kun i begrænset omfang overfører midler til fremtidige generationer af pensionister.

Dette synspunkt er vi *helt* uenige i.

For det første er den demografiske udfordring noget større end JGA giver udtryk for. I VFKs befolkningsfremskrivning stiger ældrebrøken (forholdet mellem 65+-årige og 15-64-årige) med 70 pct. fra 2000 til 2040. Fra 2040 og til 2050 falder ældrebrøken lidt, således at stigningen til 2050 »kun« er på 62 procent.

For det andet og væsentligere har JGA helt urealistiske forestillinger om betydningen af de fremtidige udbetalinger fra arbejdsmarkedspensionsordninger og privattegnede pensioner.

Opbygningen af arbejdsmarkedspensionerne har betydelig effekt på det nuværende og forventede fremtidige pres på pensionssystemet, men det skyldes først og fremmest, at disse pensioner løser et strukturelt problem i folkepensionssystemet: Den universalistiske folkepension gives til alle uanset tidligere arbejdsmarkedstilknytning. Det betød historisk, at en familie, hvis levestandard hovedsageligt var baseret på én indkomst i den erhvervsaktive alder, kun ville opleve et mindre fald i den disponible indkomst ved overgang til pensionsalderen, fordi indkomsten kom til at bestå af to pensioner. I takt med at også ældre kvinder fik stigende erhvervsfrekvens i løbet af 1980erne og 1990erne, betød overgangen til pensionsalderen, at familierne i disse generationer kunne se frem til en betydelig nedgang i den disponible indkomst, fordi også indkomsten i de erhvervsaktive år hovedsageligt var baseret på to indkomster. Uden opbygning af arbejdsmarkedspensionssystemet ville denne udvikling have ført til et pres for højere dækning af den offentlige folkepension.

Bortset fra denne – meget væsentlige – effekt har arbejdsmarkedspensioner begrænset positiv effekt på den offentlige sektors finansiering. Effekten bliver begræn-

set, fordi der er en del modsatrettede virkninger. Det er rigtigt, som JGA nævner, at der betales skat af udbetalinger fra arbejdsmarkedspensioner og privattegnede pensioner, men modsvaret er, at indbetalinger til arbejdsmarkedspensioner er fradragsberettigede i marginals-katten. I gennemsnit betyder det, at skattefradraget i procent af indbetalingerne er større end skattebetalingerne i procent af udbetalingerne, fordi indkomsten som erhvervsaktiv er højere end indkomsten som pensionist. Hertil kommer, at pensionsopsparing er skattesubsidieret, fordi pensionsafkastbeskatningen er på 15 pct., mens almindelig kapitalindkomst beskattes efter en progressiv skala, der starter med 33 pct. I modsat retning trækker, at udbetalinger fra arbejdsmarkedspensioner fører til reduktion i den indkomstafhængige del af folkepensionen og i visse særydelser til pensionister, først og fremmest boligydelse. Den sidste effekt af arbejdsmarkedspensionerne er, at satsreguleringen er baseret på lønudviklingen efter indbetalinger til arbejdsmarkedspensioner. I de år, hvor der aftales en højere indbetalingssats til arbejdsmarkedspensionsordninger, vil satsreguleringen derfor blive reduceret, hvilket reducerer de offentlige udgifter på langt sigt.

I Velfærdskommissionen (2004) er der set på betydningen af at øge indbetalingsprocenten med 3 pct. af lønnen for alle beskæftigede. Det fører til, at den nødvendige skattestigning ved holdbar skattesats reduceres fra 8,7 til 7,6 pct., svarende til en reduktion i den holdbare skattesats på 1,1 pct. Som et overslag kan det på basis af denne beregning anslås, at hvis der ikke havde været arbejdsmarkedspensioner, ville det samlede holdbarhedsproblem have krævet en stigning i bundskattesatsen, som ville være ca. 4-5 procentpoint højere, end den VFKs fremskrivning fører til. Det er således ikke nogen ubetydelig effekt, men altså ikke tilstrækkeligt til at løse det samlede problem.

Mest uenige er vi dog med udsagnet: »Velfærdskommissionens beregnede manko på ca. 4 pct. af BNP... (er) da heller ikke nogen gyser. Det er ikke mere end Finansministeriets »budgetbisser« kunne finde over en 10-årig periode«.

Her tales om en permanent årlig besparelse på ca. 50 mia. kr. Det svarer til 20 procent af de samlede årlige brutto overførselsindkomster. VFK er klart af den opfattelse, at ændringer i velfærdssamfundet i dette omfang bør gennemgå en grundig demokratisk proces og indføres med en lang indfasning, så befolkningen kan nå at tilpasse sig de ændringer, som denne proces leder frem til.

### **Arbejdsstyrkens størrelse**

Både CS og JGA diskuterer udviklingen i den fremtidige arbejdsstyrke. Begge finder, at VFKs fremskrivninger leder til en undervurdering af den fremtidige arbejdsstyrke, og dermed til en undervurdering af grundlaget for de fremtidige skatteindtægter. Tilsammen peger de på forudsætninger vedrørende erhvervsdeltagelsen for ind-

vandrere (JGA) og efterkommere fra mindre udviklede lande (CS), arbejdsstyrkens fremtidige uddannelsesniveau (CS, JGA) og incitament til senere tilbagetrækning (JGA).

#### *Indvandreres erhvervsfrekvens*

JGA påpeger, at indvandreres erhvervsdeltagelse er stigende med varigheden af opholdet i Danmark, og at den nuværende stramme indvandringspolitik har reduceret den årlige indvandring, således at der er en tendens til, at indvandreres gennemsnitlige erhvervsfrekvenser stiger i takt med, at herboende indvandreres gennemsnitlige opholdstid vokser. Det er fuldstændig rigtigt, og det er da også sådan, at DREAMs fremskrivning af erhvervsdeltagelsen for indvandrere er lavet. Der henvises til en forholdsvis grundig beskrivelse af dette i Velfærdskommissionen (2005a) s. 100-105.

Det fremgår også af ovenstående reference, at der er tendens til, at denne metode leder til en systematisk overvurdering af den fremtidige udvikling i erhvervsfrekvensen. Det skyldes, at de alders- og opholdstidsbestemte erhvervsfrekvenser er baseret på tværsnitoplysninger fra Danmarks Statistiks RAS-statistik. Imidlertid er sammensætningen af indvandrere på oprindelseslande gradvist ændret fra 1980 til i dag. Ændringen betyder, at jo kortere ophold indvandrere har i landet, jo større andel kommer fra lande, hvor uddannelsesniveau mv. afviger fra det danske. Ved anvendelse af tværsnitsdata bliver disse forskelle i sammensætningen af oprindelseslande omsat til effekter af opholdets længde. Schultz-Nielsen (2000) og Schultz-Nielsen og Constant (2004) dokumenterer dette ved registeranalyser af kohorter af indvandrere, mens DA (2001) betragter effekten af at korrigere landesammensætningen. Ændringen i indvandrerpolitikken i 2002 kan trække i modsat retning, fordi sammensætningen på typer af opholdsgrundlag er ændret. Det er endnu for tidligt at vurdere størrelsen af denne effekt.

#### *Efterkommeres erhvervsfrekvens*

CS fremhæver, at VFK anvender en for pessimistisk vurdering af erhvervsfrekvensen for indvandrere fra mindre udviklede lande. Der gives ingen begrundelser for dette synspunkt bortset fra, at tidligere fremskrivninger har en højere erhvervsdeltagelse. Disse fremskrivninger baseres på videreudviklinger af Pedersen (2002), hvor herkomstafhængig adfærd introduceres i danske makroøkonomiske fremskrivninger. Baggrunden for VFKs antagelse er, at en række nyere undersøgelser entydigt peger på, at gruppen af efterkommere fra mindre udviklede lande klarer sig dårligt i skolesystemet, hvor halvdelen ifølge PISA undersøgelserne har begrænsede funktionelle læsefærdigheder, jf. Egelund og Schindler (2005), ligesom der er 65 procent, der ikke gennemfører en ungdomsuddannelse, jf. Tænk tanken og AKF (2004). Udgangspunktet for denne gruppe unges uddannelsesniveau er derfor bekymrende, og den store andel uden kompetence-

givende uddannelse har negativ konsekvens for den fremtidige erhvervsfrekvens, jf. nedenfor. Senest har LO (2005) analyseret udviklingen i erhvervsfrekvenser for efterkommere fra ikke-vestlige lande og fundet en betydelig reduktion i erhvervstilknytningen i perioden fra 2001-04. LOs analyse peger ligeledes på, at efterkommernes manglende uddannelse giver et væsentligt forklaringsbidrag til den lave erhvervsdeltagelse.

For at kunne vurdere ældre aldersgrupper af efterkommere fra mindre udviklede lande er det nødvendigt at foretage internationale sammenligninger. Rockwoolfondens Forskningsenhed har foretaget en sammenlignende analyse med forholdene i Tyskland, hvor indvandringen fra mindre udviklede lande startede tidligere end i Danmark. Derfor er de ældste efterkommere fra mindre udviklede lande ca. 10 år ældre end i Danmark. De tyske erfaringer tyder på, at nettobidragene til den offentlige sektor fra efterkommere fra mindre udviklede lande ikke har samme tendens til at stige med alderen, som tilfældet er i Danmark for personer af dansk oprindelse. Niveaue for nettobidragene for gruppen af 30 – 36-årige efterkommere fra ikke-vestlige lande i Tyskland er på mellem 0 og 5.000 euro pr. år. I Rockwools opgørelse er det tilsvarende tal for danskere og efterkommere fra vestlige lande omkring 10 – 15.000 euro, jf. Wadensjö og Gerdes (2004). Der er ikke noget, der tyder på, at Danmark skulle være bedre til at løse integrationsopgaven end Tyskland.<sup>6</sup> De tyske erfaringer peger således i retning af, at arbejdsmarkedstilknytningen forbliver forholdsvis lav for efterkommere efter indvandrere fra mindre udviklede lande. I VFKs analyse er det antaget, at gruppens arbejdsmarkedstilknytning ligger højere end for indvandrere fra mindre udviklede lande, men lavere end arbejdsmarkedstilknytningen for efterkommere fra mere udviklede lande.

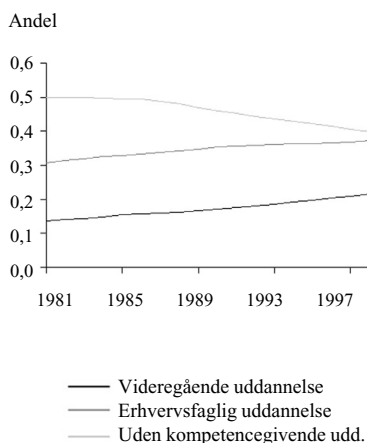
På grund af usikkerheden er der i Velfærdskommissionen (2005a) lavet en række følsomhedsanalyser af dette forhold. De viser, at hvis denne gruppes erhvervstilknytning forbedres i forhold til forudsætningerne, svarer det til en reduktion i den permanente stigning i bundskattesatsen på mellem 1 og 2 procentpoint. I disse beregninger er der set bort fra, at en sådan stigning næppe kan opnås uden forøget ressourceindsats i uddannelsessektoren.

#### *Øget uddannelse – flere på arbejdsmarkedet?*

Både CS og JGA påpeger, at et øget uddannelsesniveau efter deres opfattelse vil trække i retning af øget gennemsnitlig erhvervsdeltagelse. Årsagen er, at det gennemsnitlige uddannelsesniveau i arbejdsstyrken forventes at stige i takt med, at nye generationer med relativt højt uddannelsesniveau erstatter ældre med lavere gennemsnitligt uddannelsesniveau i de kommende år, og at personer med højere uddannelsesniveau har højere erhvervsdeltagelse. Begge de udsagn er isoleret set korrekte.

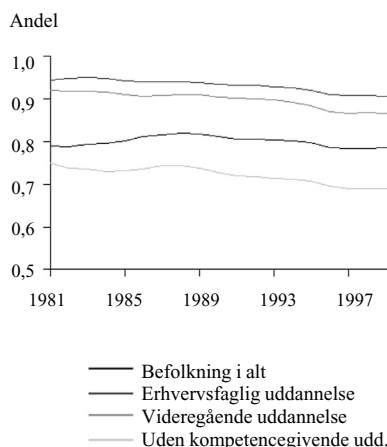
---

6. Jf. også Velfærdskommissionen (2005b) for en sammenligning af integrationen i forskellige lande.



Figur 2. Uddannelsesfordeling 1981-2003.

Kilde: Danmarks Statistik.



Figur 3. Uddannelsesbetingede erhvervsfrekvenser 1981-2003.

Kilde: Danmarks Statistik.

VFK indregner en positiv effekt af stigende uddannelsesniveau på erhvervsfrekvensen. Denne er på 35.000 personer på langt sigt. I lyset af det historiske forløb vurderes dette imidlertid som en optimistisk effekt, jf. nedenfor.

Arbejdsstyrkens gennemsnitlige uddannelsesniveau er stigende. Det kan illustreres ved at betragte udviklingen siden 1981 i sammensætningen af arbejdsstyrken opdelt i tre uddannelsesniveauer: Uden kompetencegivende uddannelse, Erhvervsfaglig uddannelse og Videregående uddannelse,<sup>7</sup> jf. figur 2. Det ses af figuren, at der er en betydelig stigning i uddannelsesniveaet for personer i den erhvervsaktive alder (15-64-årige) i perioden 1981 til 2003. Den andel af personer i den erhvervsaktive alder, der har en videregående uddannelse, vokser fra 13 til 25 pct. Også andelen af personer med erhvervsfaglig uddannelse er voksende i perioden. I 1981 var det 31 procent af personerne i de erhvervsaktive aldre, der havde en erhvervsfaglig uddannelse, mens andelen var vokset til 38 procent i 2003. Denne udvikling med en gradvist voksende andel af de to uddannelseskategorier betyder, at andelen af personer uden kompetencegivende uddannelse falder fra 50 procent til 36 procent af personerne i den erhvervsaktive alder. Der er i høj grad tale om en generationseffekt, hvor de unge generationer, der kommer til, er bedre uddannet end de generationer, der forlader arbejdsstyrken på grund af alder. Der er således grund til at forvente, at dette fortsætter, selvom omfanget kan diskuteres.

Der er imidlertid to vigtige forhold, der skal tages i betragtning for at vurdere betyd-

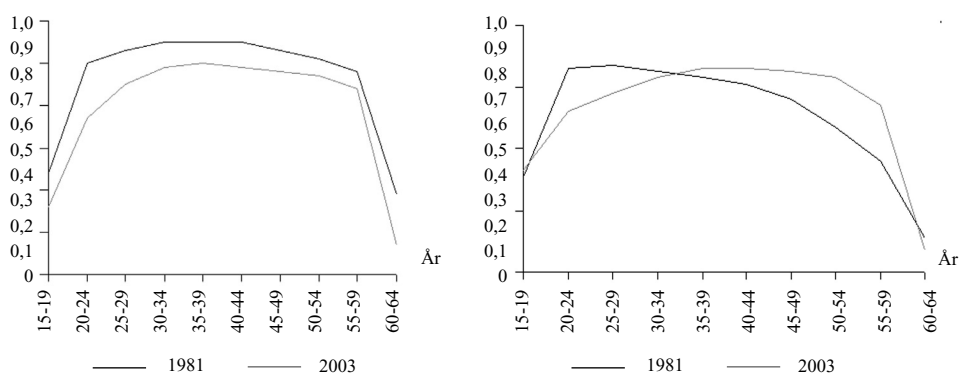
7. Kortvarig, mellemlang og lang videregående uddannelse, bachelorer, erhvervsbachelorer og forskeruddannelse.

ningen af et øget uddannelsesniveau for erhvervsdeltagelsen. Det første – som umiddelbart kan aflæses af figur 3 – er, at selvom erhvervsdeltagelsen er voksende i uddannelsesniveaue, er der for alle grupper et betydeligt fald over tid i erhvervsfrekvenserne. For personer uden kompetencegivende uddannelse er faldet i erhvervsdeltagelsen størst, idet det er på 8 procentpoint over perioden, mens de tilsvarende fald for de to andre uddannelseskategorier er 6 procentpoint for erhvervsfagligt uddannede og 5 procentpoint for videregående uddannede. En del af forklaringen på, at der er en faldende erhvervsdeltagelse for alle tre grupper kan være velstandsstigningen, jf. diskussionen i vores andet indlæg i dette nummer.

Årsagen til, at det har været muligt at fastholde en nogenlunde konstant gennemsnitlig erhvervsdeltagelse for alle uddannelsesgrupper under ét, er således, at stigningen i uddannelsesniveaue har ført til, at der er blevet flere personer i de grupper, der har den højeste erhvervsdeltagelse. Det kræver altså en fortsat stigning i uddannelsesniveaue for at kunne fastholde en konstant erhvervsfrekvens i fremtiden, hvis tendensen til faldende erhvervsfrekvens inden for den enkelte uddannelsesgruppe fortsætter. Det kan blive vanskeligt at fastholde samme vækst i uddannelsesniveaue fremadrettet, fordi de generationer, som forlader arbejdsmarkedet, efterhånden også vil være bedre uddannet i gennemsnit.

For det andet er det ikke oplagt, at en given ændring i uddannelsessammensætningen i fremtiden vil få samme effekt. Det skyldes, at et stigende uddannelsesniveau ikke alene vil påvirke gruppen, der får en given uddannelse, men også gruppen af personer uden kompetencegivende uddannelse. Sidstnævnte gruppe indeholder personer med f.eks. en studentereksamen som højeste uddannelsesniveau. Samtidig indeholder gruppen også personer som pga. fysiske eller psykiske handicaps ikke er i stand til at gennemføre en uddannelse og derfor heller ikke har mulighed for en høj arbejdsmarkedstilknytning. Når gruppen af personer uden kompetencegivende uddannelse reduceres, sker dette hovedsageligt ved, at de bedst kvalificerede i gruppen sikres en uddannelse. Det betyder, at den gennemsnitlige erhvervstilknytning falder for de, der er »tilbage« i gruppen uden kompetencegivende uddannelse. Man kan således ikke realistisk antage uændrede erhvervsfrekvenser for alle grupper i takt med, at der sker et uddannelsesløft.

Dette problem ses endnu mere markant, når man betragter udviklingen i antallet af førtidspensionister som følge af en ændret uddannelsessammensætning. Tager man udgangspunkt i konstante alders-, køns- og uddannelsesbetingede førtidspensionsfrekvenser for perioden 1980 til 2001 og benytter den faktiske udvikling i uddannelsessammensætningen, vil man forvente et faldende antal førtidspensionister. Årsagen er, at de historiske erfaringer udviser den sammenhæng, at hyppigheden for førtidspension er faldende med uddannelsesniveaue. Faktisk steg antallet af førtidspensionister over



Figur 4. Aldersbetingede erhvervsfrekvenser for henholdsvis mænd (venstre figur) og kvinder (højre figur).

denne periode. Problemet opstår, fordi man ikke tager højde for, at de bagvedliggende faktorer – fysiske eller psykiske – der begrundet tildeling af førtidspensioner også i mange tilfælde er årsagen til, at de pågældende ikke får en uddannelse.<sup>8</sup>

Samlet vurderet er det således ikke oplagt, at øget uddannelse vil øge erhvervsdeltagelsen. Det niveau for stigningen i erhvervsdeltagelsen, der er indarbejdet i VFk's fremskrivning, kan derfor vise sig at være for optimistisk.

#### *Incitament til tilbagetrækning*

De historiske aldersbetingede erhvervsfrekvenser for mænd ligger lavere for alle aldersgrupper i 2003 end i 1981, jf. figur 4. Faldet er særligt markant for gruppen af 60-64 årige. For kvinder gælder, at erhvervsfrekvensen for yngre aldersgrupper er lavere i 2003 end i 1981, mens det for ældre aldersgrupper er omvendt. Der er således en betydelig generationseffekt i kvindernes erhvervsdeltagelse. Det er bemærkelsesværdigt, at selvom erhvervsfrekvensen for 55-59 årige kvinder er 18 procentpoint højere i 2003 end i 1981, er den 4 procentpoint lavere for aldersgruppen 60-64 år. Der er således for begge køn en tendens til, at en større andel trækker sig tilbage ved 60 års alderen end tidligere. Den historiske udvikling taler derfor for, at en antagelse om konstant tilbagetrækningsalder har en bias i retning af for høj fremtidig erhvervsdeltagelse.

JGA argumenterer for, at arbejdsmarkedspensionerne kan betyde en tilskyndelse til senere tilbagetrækning, fordi arbejdsmarkedspensionsudbetalingen pr. år er lavere, jo tidligere man trækker sig tilbage (substitutionseffekt). Hvis arbejdsmarkedspensions-

8. I CS's papir i dette nummer af tidsskriftet er de af Arbejderbevægelsens Erhvervsråd beregnede effekter på beskæftigelsen af øget uddannelse reduceret, således at førtidspensionsfrekvensen ikke påvirkes ved uddannelsesændringen.

ordningerne på den anden side fører til en stigning i nettoopsparingen, betyder ordningerne også, at den samlede formue på tilbagetrækningstidspunktet er højere (indkomst/formueeffekt). Derfor vil formueeffekten trække i retning af, at arbejdsmarkedspensionerne leder til tidligere tilbagetrækning. JGA sammenligner med effekten af ændringen i det svenske pensionssystem, men denne sammenligning er ikke perfekt, idet ændringen i det svenske pensionssystem kun indebærer introduktionen af en substitutionseffekt, men ingen formue/indkomst effekt, idet de svenske pensioner ikke vokser som følge af reformen.

### **Satsregulering – skjulte besparelser?**

I VFKs fremskrivning er det antaget, at satserne for overførselsindkomster følger med lønudviklingen. En tilsvarende antagelse anvendes i Finansministeriets og Det Økonomiske Råds fremskrivninger. Satsreguleringen indeholder den såkaldte satspulje, hvortil en del af midlerne overføres. Det betyder, at den årlige automatiske regulering af satserne for overførslerne som hovedregel er 0,3 procent lavere end lønudviklingen.<sup>9</sup>

Der har været en betydelig diskussion om denne forudsætning. Indledningsvist er der grund til at fremhæve to almindelige misforståelser om satspuljen, som hver især kan føre til den fejlagtige konklusion, at satspuljen indebærer en automatisk, gradvis forbedring af de offentlige finanser, som Velfærdskommissionen ikke har taget højde for.

*Misforståelse 1:* »Satspuljepenge er kun 1-årige, mens forhøjelserne af satserne for overførselsindkomsterne er permanente«.

*Misforståelse 2:* »Satspuljepengene udhules, fordi de beløb, der går via satsreguleringspuljen, ikke reguleres årligt«.

Faktum er, at de penge, der hvert år føres over i satspuljen, er permanente bevillinger. Puljemidlerne akkumulerer således over tid. Hvis overførslen hvert år udgør f.eks. 1/2 mia. kr., så vil der i år 2 være 1 mia. nutidskroner, 2,5 mia. nutidskroner i år 5, 5 mia. nutidskroner i år 10 osv.

Satspuljen finansierer aktuelt i størrelsesordenen 200 initiativer. F.eks. kan nævnes førtidspensionsreformen fra 2003, som i 2020 forventes at koste over 2 mia. kr. Hovedparten af initiativerne er permanente, dvs. at initiativerne vil fortsætte i al fremtid, medmindre det politisk besluttes at ændre reglerne/grundlaget for initiativet. En vigtig pointe er i den forbindelse, at pengene ikke falder bort, hvis et initiativ ophører eller reduceres i omfang. Pengene lægges i så fald tilbage i satspuljen til genudmøntning

---

9. I 2005 er det dog kun 0,2 pct., fordi den opgjorte lønstigning kun var 2,2 pct. I alle tidligere år siden ordningens start i 1993 er der dog blevet afsat 0,3 pct.



med »renters rente«. De initiativer, som er finansieret af satspuljen, reguleres i øvrigt hvert år med relevante pris- og lønindeks. Pengene »forrentes« altså også, mens de er i spil.

Satsreguleringspuljens midler skal i al væsentlighed bruges på initiativer, der henhører under overførselsindkomster eller velfærdsservice. VFKs antagelse svarer til, at alle de afsatte midler anvendes til forbedringer af overførselsindkomsterne. I det omfang satspuljens midler i stedet bruges til forbedringer af velfærdsservice, så vil VFKs fremskrivning overvurdere udgifterne til overførselsindkomster. Til gengæld undervurderes udgifterne til velfærdsservice tilsvarende.

JGA argumenterer dels for, at en del af de midler, der anvendes fra satspuljen, fortrænger andre bevillinger på dette felt. Implicit gælder, at der reelt er tale om besparelser. Det kan ikke afvises, men argumentet er symmetrisk: Hvis satsreguleringen ikke sikrer fuld kompensation for den gennemsnitlige indkomststigning, øges bevillingerne til de grupper, hvor der er et fald, jf. ældrecheck og Dansk Folkepartis aktuelle forslag om stigninger i førtidspensionen finansieret uden for satspuljen. Fordelingen på service eller overførsler påvirker derfor næppe det samlede udgiftstryk.

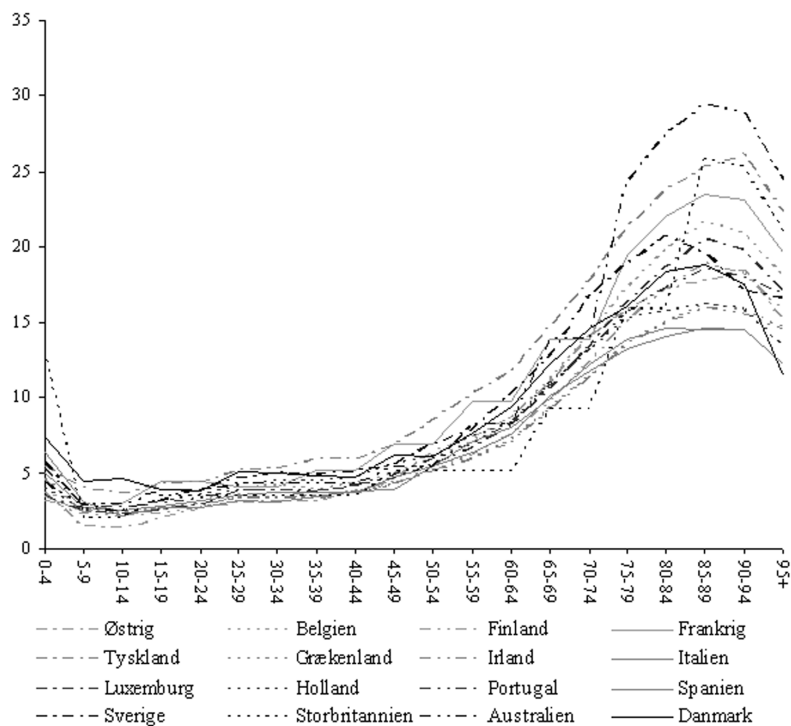
I forhold til vurderingen af den finanspolitiske holdbarhed skal man være opmærksom på, at bevillinger til serviceydelser ikke beskattes i første led, hvilket er tilfældet for de fleste overførselsindkomster. Ved at antage, at hele reguleringen tilfalder overførselsindkomst, overvurderes skatteeffekten, mens det finanspolitiske holdbarhedsproblem undervurderes.

Derimod er det rigtigt, som JGA påpeger, at udformningen af satsreguleringen, hvor stigningen baseres på faste vægte, generelt tenderer mod, at stigningen i den gennemsnitlige lønindkomst er højere end satsreguleringen, fordi gennemsnittet er påvirket af en gradvis ændret uddannelsessammensætning i arbejdsstyrken på DA området. Denne effekt er ikke med i VFKs fremskrivning.

### **Længere levetid og ny teknologi – mindre eller mere behov for sundhedsvæsenet?**

Sundheds- og plejeområdet er centralt for den fremtidige udvikling af velfærdssamfundet. Det er en af de vigtigste velfærdsopgaver at sikre et tidssvarende og velfungerende sundhedsvæsen. Samtidig er det et spørgsmål om, hvor markant den ændrede demografi vil påvirke sundhedsvæsenet. Flere ældre vil umiddelbart betyde et øget pres på sundhedsvæsenet, da der er en klar sammenhæng mellem alder og brug af sundhedsvæsenet, jf. nedenfor. Da hovedkilden til den ændrede demografi er en stigende middellevetid, er det et spørgsmål, i hvilket omfang »flere sunde leveår« vil medvirke til at mindske presset på sundhedsvæsenet. Såvel CS som JGA rejser dette spørgsmål i forhold til VFKs fremskrivning.

VFK har taget udgangspunkt i det faktum, at flere ældre vil betyde et øget behand-

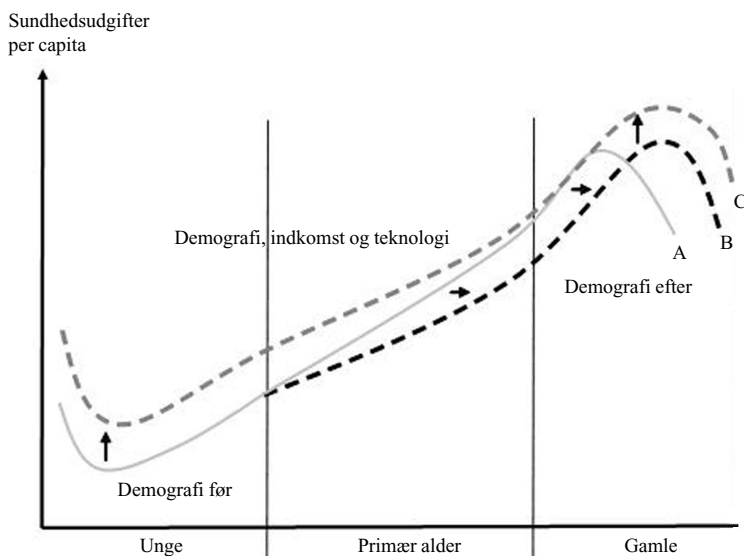


Figur 5. Aldersfordelte sundhedsomkostninger i EU-15 og Australien målt som pct. af BNP pr. capita.

Kilde: OECD Health Data.

lingsbehov i sundhedsvæsenet og dermed i sig selv trække i retning af et øget udgiftsbehov. I forhold til dette grundforløb er der to væsentlige usikkerhedsfaktorer. Den ene er den nævnte mulighed for, at stigende levetid giver sig udslag i »flere sunde leveår« og dermed en forskydning i den aldersmæssige brug af sundhedsvæsenet. Dette kan betyde en mindre udgiftsstigning end i VFKs fremskrivning. Den anden faktor er den teknologiske udvikling inden for sundhedsområdet, hvor der – heldigvis – løbende opdages nye og bedre behandlingsformer, som gør lægevidenskaben i stand til at behandle sygdomme, der tidligere ikke kunne behandles, eller forbedre den eksisterende behandling af andre sygdomme.

Forenklet kan man sige, at usikkerheden er knyttet til, om den længere middellevetid er drevet en udvidelse af den raske periode i livet eller en udvidelse af den periode, hvor man ved lægelige behandlinger holdes længere i live. Der findes empiriske undersøgelser, der understøtter begge effekter. De fleste undersøgelser finder, at i hvert fald en del af levetidsforlængelsen finder sted som en forlængelse af den raske del af



Figur 6. Skematisk illustration af udviklingen i den reale fordeling af sundhedsudgifterne over livet.

livet. Samtidig finder undersøgelser også typisk, at der er en tendens til, at forskellen i sundhedsudgifter mellem de forholdsvis billige unge aldersgrupper og de forholdsvis dyre ældre aldersgrupper vokser, således at aldersfordelingen over tid bliver mere stejl.

Den teknologiske udvikling på sundhedsområdet er central, og i kraft af landvindinger gjort inden for lægevidenskaben er der i sundhedsvæsenet sket en kraftig forbedring i de behandlinger, befolkningen kan tilbydes. Dette er en væsentlig del af fremgangen i velfærden. Der er en stigende erkendelse af, at den teknologiske udvikling på det lægevidenskabelige område har afgørende betydning for udviklingen i sundhedsudgifterne. Teknologiuudviklingen betyder på den ene side, at der kan tilbydes behandlinger på sygdomsområder, der ikke tidligere var mulige at behandle og på den anden side, at behandlingen med eksisterende teknologier bliver billigere over tid. I en række lande viser undersøgelser, at effekten af flere og bedre behandlingsformer er klart dominerende og hovedforklaringen på, at sundhedsudgifterne i de fleste lande udgør en voksende andel af BNP. En række undersøgelser i forskellige lande leder til det resultat, at teknologiuudviklingen kan forklare en mervækst i sundhedsudgifterne pr. person ud over den generelle økonomiske vækst på typisk mellem 0,5 til 1,5 pct. pr. år, jf. Productivity Commission (2005).

Spørgsmålet omkring betydningen af længere levetid og ændrede og forbedrede behandlingsformer og -metoder i sundhedsvæsenet kan hensigtsmæssigt diskuteres med

udgangspunkt i figur 5. Figuren viser de aldersafhængige sundhedsudgifter for Danmark og en række andre lande. Det fremgår, at der er den forventede aldersafhængighed i brugen af sundhedsvæsenet. Sammenhængen findes for alle lande, og formen af kurven er påfaldende ens på tværs af landene givet forskellige middellevetider m.m.

I figur 6 er udgangspunktet bestemt af de historiske aldersbetingede sundhedsudgifter givet ved den fuldt optrukne linie mærket A. Effekten af øget levetid drevet af »flere sunde leveår« og dermed en tilknytning af sundhedsudgifterne til forventet restlevetid frem for alder vil svare til, at den aldersfordelte kurve forskydes mod højre i takt med væksten i middellevetiden, f.eks. til den stiplede kurve mærket B. Det vil føre til en begrænsning i stigningen i sundhedsudgifterne i procent af BNP, men ikke forhindre, at den ændrede befolkningsfordeling med flere pensionister og færre i den erhvervsaktive alder fører til en stigning i sundhedsudgifternes andel af BNP.

Den teknologiske udvikling har en tendens til at gøre behandlingen dyrere og dermed forskyde de aldersfordelte sundhedsudgifter opad, jf. den stiplede kurve mærket C i figur 6.<sup>10</sup>

For at analysere de to effekters relative betydning er de begge indlagt i VFKs fremskrivning. Effekten af, at levetidsstigningen fører til flere raske leveår, er medtaget ved at antage, at de aldersfordelte sundhedsudgifter reguleres delvist med restlevetiden, således at stigninger i restlevetiden slår igennem med 50 procent på de aldersafhængige udgifter. Det svarer til, at de aldersfordelte sundhedsomkostninger i figur 5 rykker mod højre med stigende middellevetid, men kun halvt så hurtigt som stigningen i restlevetiden, dvs. fra kurven A til kurven B. Derved fås en antagelse om, at den forlængede levetid både giver anledning til flere raske år og flere år med høj sygdomsfrekvens. Isoleret set betyder dette, at sundhedsudgifterne ikke stiger så meget i procent af BNP som i VFKs fremskrivning. Indregnes kun denne effekt skal bundskattesatsen stige med 6,5 procentpoint for at sikre den finanspolitiske holdbarhed i stedet for de 8,7 procentpoint, som er tilfældet i fremskrivningen.

Indregnes både den delvise restlevetidsafhængighed og en mervækst i de aldersafhængige sundhedsomkostninger på 0,5 pct. om året, således at der tages højde for begge effekter i figur 6 – dvs. effekten af at gå fra kurven A til kurven C – fås, at stigningen i bundskattesatsen skal være på 22,1 procentpoint for at sikre finanspolitisk holdbarhed. Det svarer til hele 13,4 procentpoint mere end i VFKs fremskrivning.

Ved at sammenligne de to forhold – flere sunde leveår, og forbedre teknologi – fremgår det af beregningerne, at usikkerhed knyttet til effekten af »flere sunde leveår« kun giver en begrænset effekt på finansieringsproblemet, set i forhold til effekten af

---

10. Teknologiuudviklingen er i figuren antaget at påvirke alle aldersgrupper ens. Det er mere i overensstemmelse med eksisterende undersøgelser, at effekten er størst for de ældste generationer, således at fordelingen bliver mere stejl.

ændret teknologi. Selv en meget begrænset årlig merstigning i sundhedsudgifterne som følge af den teknologiske udvikling er tilstrækkelig til at eliminere de positive effekter, der forhåbentlig kommer som følge af, at antallet af raske leveår stiger med den stigende levetid. Set i dette perspektiv må grundforløbet i VFKs fremskrivninger, for så vidt angår sundhedsområdet, siges at ligge til en optimistisk side.

I Danmark er sundhedsudgifterne overvejende offentligt styret, og derfor er det ultimativt et politisk spørgsmål, hvor mange ressourcer, der skal anvendes på sundhedsområdet. En stram styring af området er mulig, men det vil betyde et stigende spænd mellem, hvad det offentlige sundhedsvæsen kan tilbyde befolkningen og både sundhedsvæsenets muligheder og befolkningens ønsker i kraft af den teknologiske udvikling. De problemer dette kan afstedkomme, er baggrunden for det, vi har kaldt velstandsdilemmaet i vores første artikel i dette nummer af tidsskriftet.

### **Metodemæssige problemstillinger**

I dette afsnit ser vi på en række ret forskellige diskussioner af det metodiske grundlag for VFKs fremskrivning, rejst af hhv. JJ og CV. Diskussionerne har kun i begrænset omfang fælles berøringspunkter.

Den sociale kontrakt knyttet til velfærdssamfundet indebærer en klar aldersafhængighed i den enkeltes nettobidrag til de offentlige finanser. Dette rejser en generationsproblematik, og derfor er det empirisk relevante udgangspunkt at vælge en analysemetode, der eksplicit tager højde for overlappende generationer. Dette er ryggraden i DREAM-modellen, som er anvendt ved beregningerne.

Det er velkendt fra litteraturen, at OLG-modeller kan indebære et rationale for velfærdsordninger. I fravær af offentlige velfærdsordninger vil allokeringen ikke nødvendigvis være Parteto-optimal (selv under fuldkommen konkurrence), og alle agenter (unge og gamle) kan have fordel af f.eks. et PAYG-pensionssystem med overførsel fra unge til gamle, se f.eks. Blanchard og Fischer (1989).

Det er også velkendt, at udviklingen i de offentlige finanser er af stor betydning, da det fastlægger enten de løbende bidrag fra de unge (erhvervsaktive) eller det, der kan stilles til rådighed for de gamle. Dette gælder hvad enten velfærdsordningerne er af en eksplicit PAYG-form eller muliggøre budgetvariationer over tid.

JJ argumenterer for, at den offentlige sektors budgetrestriktion ikke er en relevant restriktion, men at det derimod er den samlede økonomis restriktion summeret i landets betalingsbalance, der er det relevante kriterium. Det er således summen af den diskonterede værdi af det private og det offentlige forbrug, der skal kunne finansieres af de samlede indtægter i økonomien. Det er selvfølgelig rigtigt, at den samlede budgetrestriktion skal være opfyldt, men det implicerer ikke, at man kan se bort fra den offentlige budgetrestriktion, med mindre man mener, at den offentlige sektor til en-

hver tid kan beslaglægge privat formue uden omkostninger af nogen art, således at den offentlige budgetrestriktion er opfyldt. Det forekommer på denne baggrund at være en lovlig flot bemærkning, når JJ skriver, at »det økonomisk-politiske problem i grundforløbet er således ikke finansiering af det offentlige underskud ...«.

Bemærk endvidere, at finanspolitik (øget offentligt forbrug finansieret ved en skat-testigning) påvirker den samlede efterspørgsels sammensætning og dermed den reale valutakurs. Dette er en standard transmissionskanal for finanspolitikken i alle nyere »åbne-makromodeller«. Det er således fejlagtigt, når JJ fremfører, at finanspolitik i DREAM-modellen kun har effekt via en forvridende effekt af skatter på arbejdsudbuddet.

Endelig argumenterer JJ for, at anvendelsen af en generel ligevægtsmodel (her DREAM-modellen) fører til, at man begår »den atomistiske fejlslutning«. Vi er ikke sikre på, hvad han mener med dette. I Elster (1979) defineres den atomistiske fejlslutning som den situation der optræder, når det gælder at: »hvis det for hvilken som helst  $x$  gælder, at den kan være  $F$ , så kan alle  $x$  (samtidig) være  $F$ «. Det er præcis denne type fejltagelser man undgår ved eksplicit at opstille en generel ligevægtsmodel. Det klassiske eksempel fra økonomi er, at selvom den enkelte tager f.eks. en pris for givet og anser den for uafhængig af egen adfærd, vil der fremkomme prisreaktioner, når tilstrækkelig mange agenter ændrer adfærd. Anvendelsen af en generel ligevægtsmodel vil tage højde for disse afledede eller aggregerede effekter. Hvis man derfor mener, at den atomistiske fejlslutning er central for fastlæggelsen af den økonomiske politik, skal fastlæggelsen af denne netop ske ved anvendelse af en generel ligevægtsmodel. Generelle ligevægtsmodeller har derfor den fordel, at de kan tage højde for forskellige former for interdependens. For god ordens skyld skal det bemærkes, at en analyse baseret på en generel ligevægtsmodel ikke kræver antagelser om fuldkommen konkurrence eller nødvendigvis implicerer fravær af ledighed eller gunstige effekter af velfærdsordninger – heller ikke på langt sigt.

På den økonomisk-teoretiske side rejser CV spørgsmålet om en konsistent modellering af arvemotiv. Er det rimeligt at antage, at politikerne har et arvemotiv (som er det rationelle grundlag for at vælge en opsparingsstrategi), hvis de private agenter modelleres med udgangspunkt i en overlappende generationsstruktur (OLG)? CV underforstår, at der i en OLG-model ikke er arv af privat formue mellem generationerne. Dette gælder for nogle formuleringer af OLG-modeller, men dette er ikke en generel egenskab og er heller ikke egenskaben ved DREAM-modellen.

Udgangspunktet for CVs kommentar er, at hvis generationerne har (to-vejete) altruisme, dvs. fremtidige generationers nytte indgår i deres egen nytteafvejning, da kan OLG-modellen »kollapse« til en model med uendeligt levende agenter (en Ramsey-model). Dette er først vist af Barro (1974). Det er velkendt, at en Ramsey-model har

både andre empiriske og teoretiske egenskaber end den overlappende generationsmodel.

For at bevare OLG-strukturen, men inkludere et empirisk funderet arvemotiv, er arv i DREAM-modelleret som et nyttiggivende »forbrug«, der ligger i perioden efter sidste periodes forbrug. Denne periodes »forbrug« vægtes anderledes end andre perioders forbrug i nyttefunktionen, men bortset fra denne vægt fastlægges »den optimale arv« på samme måde som husholdningens optimale forbrug i en aktiv periode. Størrelsen af arven fastlægges med andre ord ud fra Keynes-Ramsey reglen for den optimale forbrugsudvikling. Vægten, som lægges på arv i nyttefunktionen, fastlægges ud fra data, idet formuefordelingen over livsforløbet kalibreres til en gennemsnitlig formueprofil fra Lovmodellen.<sup>11</sup>

Der er således ingen grundlæggende modstrid mellem den antagelse, der ligger til grund for husholdningernes optimering og en antagelse om, at politikerne har et arvemotiv ved fastlæggelsen af den økonomiske politik.

Et andet specifikt økonomisk teoretisk problem, der rejses i CVs indlæg, er spørgsmålet om eksplicit aggregering over generationer. I teoretisk orienterede artikler anvendes ofte en eksplicit aggregering baseret på Blanchard (1985) eller efterfølgende mere avancerede udvidelser. I empirisk baserede analyser (herunder analyser med DREAM) anvendes Auerbach og Kottlikoff (1987) typisk som udgangspunkt. I denne type modeller løses optimeringsproblemet for hver type af husholdning, og den aggregerede adfærd fremkommer som summen af de individuelle husholdningers adfærd. Der foretages således ingen analytisk aggregering. Ulempen ved denne teknik er, at det giver anledning til meget store økonomiske modeller.<sup>12</sup> Fordelen er, at man kan afvige fra de forholdsvis restriktive krav, der stilles for at kunne opnå analytiske løsninger på aggregeringen. Det betyder f.eks., at de empirisk observerede dødeligheder og de estimerede fremskrivninger heraf kan anvendes direkte.

Et tredje økonomisk teoretisk problem, der rejses af CV, er, om modeller for små åbne økonomier i et verdensmarked med perfekt kapitalmarked og varehandel vil føre til uplausible hurtige tilpasninger i reale størrelser. Igen findes problemet i mindre analytisk løsbare modeller. I anvendte generelle ligevægtsmodeller for åbne økonomier anvendes som standard den såkaldte Armington modellering af udenrigshandlen, Armington (1969), for at imødegå denne problemstilling. Modelleringen betyder, at det reale bytteforhold er endogent, og at der derfor ikke er en øjeblikkelig tilpasning af reale størrelser. I DREAM er der endvidere en antaget træg tilpasning af det fysiske

11. Pr. konstruktion er al arv i DREAM deterministisk. Der forekommer ikke »tilfældig« arv i modellen. Bemærk at denne modellering af arv giver anledning til, at både giver og modtager har nytte af arv.

12. Til VFKs afsluttende rapport anvendes en version af DREAM på 1-årstrin. Denne model består af ca. 1,5 mio. ligninger.

kapitalapparat, således at investeringerne bestemmes ved det marginale  $q$ , jf. Hayashi (1982).

CVs hovedindvendig mod analysen er, at der ikke er nogen empirisk begrundelse for, at udgangsforløbet og de to typer tilpasninger udgør de realistiske grænser for den politiske reaktionsfunktion. Det er vi ikke uenige i. Udgangspunktet for analysen er at skabe grundlaget for at diskutere økonomisk politiske reaktioner og ikke at forudsige sådanne. Formålet med vores artikel er derfor primært at diskutere effekterne af at vælge en strategi med gradvis tilpasning i den økonomiske politik over for en engangsændring i politikken. For at illustrere denne forskel har vi valgt at fokusere på en situation, hvor de nuværende velfærdsordninger og den nuværende beskatning fastholdes.

Udgangspunktet for artiklen er den nuværende finanspolitiske strategi for det mellem-lange sigt, som denne er formuleret i 2010-planen. Ideen bag 2010-planen er grundlæggende, at der foretages tilpasninger i den økonomiske politik frem til 2010, således at finanspolitikken givet de nuværende forventninger er holdbar, hvis velfærdsordninger mv. fastholdes fra 2010. I 2010-planen er kravet om den langsigtede holdbarhed omsat til krav til de årlige overskud, krav til den årlige udgiftsvækst og krav til gældnedbringelsen frem til 2010.

Den mest restriktive fortolkning af 2010 planen er derfor, at alle tilpasninger i velfærdsordningerne skal være gennemført inden 2010. I vores beregninger er dette fortolket som, at tilpasningerne skal være gennemført i 2021. Det er således givet den nuværende mellemfristede strategi, at vi opfatter engangsændringen frem til 2021 som en overgrænse for, hvor hurtigt tilpasningen af velfærdsordninger kan ske.<sup>13</sup> Det er også rigtigt, at vores andet yderpunkt, som er, at der ikke sker ændringer i velfærdsordningerne, ikke er baseret på en politikreaktionsfunktion. Vi opfatter dette forløb som et »naturligt« benchmark, hvor der ikke er taget stilling til, hvordan holdbarhed opnås.

2010-planen er imidlertid ikke så præcist formuleret, at den mest restriktive fortolkning er den eneste fortolkning (selvom meget af diskussionen om 2010-planen tager dette udgangspunkt). En mere lempelig fortolkning er, at der frem til 2010 skal påbegyndes reformer, som i løbet af deres gradvise gennemførelse efter 2010 leder til, at finanspolitikken bliver holdbar. Med henblik på at præcisere en fremtidig mellem-fristet plan er forløbet med PAYG-politik valgt som et alternativt målepunkt med den egenskab, at der ikke omfordeles mellem generationer via den offentlige budgetrestriktion. Formålet med analysen er at give et mål for den hastighed, hvormed reformer skal indføres, hvis dette skal ske uden omfordelinger mellem generationer. En fortolkning af VFKs kommissorium er, at kommissionen skal angive forslag, der sikrer, at velfærdsydelse til nuværende generationer ikke finansieres af fremtidige generationer

---

13. Udskydelsen af tidspunktet for tilpasningen til 2021 er arbitrær, men også uden betydning for det principielle i diskussionen.



ved offentlig gældsætning.<sup>14</sup> Denne fortolkning indebærer, at PAYG-forløbet kan opfattes som den mest gradvise strategi, der ligger inden for VFKs kommissorium.

De valgte forløb afspejler således grænserne for, hvor hurtigt og hvor langsomt de af VFK foreslåede reformer skal kunne påvirke det offentlige budget for at ligge inden for de rammer, der via kommissoriet er givet kommissionen.

---

14. Kommissoriets præcise formulering er: »Kommissionens forslag skal respektere regeringens målsætning om, at de nuværende generationer ikke efterlader en ufinansieret regning til kommende generationer.«

#### Litteratur

- Andersen, J. G. 2005. Splinten og bjælken, *Nationaløkonomisk Tidsskrift* 143, s. 258-64.
- Armington, P. S. 1969. A Theory of Demand for Products Distinguished by Place of Production, *IMF-Staff Papers*, vol. 16, s. 159-78.
- Auerbach, A. og L. Kottlikoff. 1987. *Dynamic Fiscal Policy*, Cambridge University Press.
- Barro, R. J. 1974. Are Government bonds Net Wealth?, *Journal of Political Economy*, vol. 82, s. 1095-1117.
- Blanchard, O. J. 1985. Debt, Deficits, and Finite Horizons, *Journal of Political Economy* vol. 93, s. 223-247.
- Blanchard, O. J. og S. Fischer. 1989. *Lectures on Macroeconomics*, MIT-Press.
- Dansk Arbejdsgiverforening. 2001. *Arbejds-markedsrapport*, København.
- Egelund, N. og B. R. Schindler. 2005. *PISA-København – Kompetencer hos elever i 9. klasse i København*, AKF-Forlaget, København.
- Elster, J. 1979. *Forklaring og dialektikk*. Noen grunnbegreper i vitenskapsteorien, Forlaget Pax, Oslo.
- Hayashi, F. 1982. Tobin's Marginal and Average q: A Neoclassical Interpretation, *Econometrica*, vol. 50, s. 213-24.
- Holmøy, E. 2005. Velferdsstatens finansieringsproblem: Utsikter i »annerledeslandet« Norge, *Nationaløkonomisk Tidsskrift* 143, s. 265-74.
- Jespersen, J. 2005. Makroøkonomisk analyse, finanspolitisk holdbarhed og DREAM-modellen: en umage cocktail, *Nationaløkonomisk Tidsskrift* 143, s. 243-49.
- Landsorganisationen, LO. 2005. *Zoom på arbejdsmarkedet 2005*, København.
- Pedersen, L. H. 2002. *Befolkningsudvikling, integration og økonomisk politik*, København.
- Productivity Commission, Australian Government. 2005. *The Impact of Advances in Medical Technology on Health Expenditure in Australia*, Progress Report.
- Schultz-Nielsen. 2000. Integrationen på arbejdsmarkedet – og de samfundsøkonomiske forholds betydning. I Mogensen, G. V. og P. C. Matthiesen, *Integration i Danmark omkring årtusindeskiftet*, Rockwool Fondens Forskningsenhed og Aarhus Universitetsforlag.
- Schultz-Nielsen, M. L. og A. Constant. 2004. Immigrant Self-Employment and Economic Performance. I Tranæs, T. og K. F. Zimmermann, red., *Migrants, Work and the Welfare State*. Syddansk Universitetsforlag og Rockwool Fondens Forskningsenhed.
- Sørensen, C. 2005. Velfærdskommissionens analyse af finanspolitisk holdbarhed – før og nu, *Nationaløkonomisk Tidsskrift* 143, s. 230-42.
- Tænketanken om udfordringer for integrationsindsatsen i Danmark og AKF. 2004. *Udlændinges vej gennem uddannelsessystemet*. Ministeriet for Flygtninge, Indvandrere og Integration, København.
- Vastrup, C. 2005. Velfærdskommissionens makroøkonomiske analyser og anbefalinger, *Nationaløkonomisk Tidsskrift* 143, s. 250-57.
- Velfærdskommissionen. 2004. *Fremtidens velfærd kommer ikke af sig selv – Analyse-*

- rapport, København.
- Velfærdskommissionen. 2005a. *Fremtidens velfærd og globaliseringen* – Analyserapport, København.
- Velfærdskommissionen. 2005b. *Fremtidens velfærd og globaliseringen* – Debatoplæg, København.
- Velfærdskommissionen. 2005c. *Fremtidens velfærd – sådan gør andre lande* – Analyserapport.
- Velfærdskommissionen. 2005d. *Befolkningsudvikling, velstands dilemma og makroøkonomiske strategier* – Teknisk Analyserapport.
- Wadensjö, E. og C. Gerdes. 2004. Immigrants and the Public Sector in Denmark and Germany. I Tranæs, T. og K. F. Zimmermann, red., *Migrants, Work, and the Welfare State*, Syddansk Universitetsforlag og Rockwool Fondens Forskningsenhed.

# Debat og kommentarer

## Copenhagen Consensus eller Discontent? En kommentar til konferencen i København i maj 2004

Thorkil Casse

Institut for Geografi og Internationale Udviklingsstudier, Roskilde Universitets Center, E-mail: [casse@ruc.dk](mailto:casse@ruc.dk)

### Indledning

Under stor dansk og delvis udenlandsk opmærksomhed afholdt Institut for Miljøvurdering (IMV) i maj 2004 en konference under titlen Copenhagen Consensus (CC) med 8 inviterede økonomer/eksperter, hvis rolle var at prioritere mellem et større antal udfordringer for menneskeheden, specielt i de fattige lande. CC havde en ambition om at prioritere de centrale udfordringer for menneskeheden ud fra en rendyrket cost-benefit betragtning. Af en oprindelig liste på 32 udfordringer blev 10 udvalgt til nærmere undersøgelse. Et panel, som kommenterede på de oprindelige 32 udfordringer, udvalgte dem, som havde størst chance for højt samfundsmæssigt afkast, Lomborg (2004), s. 4. De 10 udfordringer blev genstand for en sektoranalyse, opdeling i indsatsområder og en beskrivelse af de eksisterende cost-benefit analyser for hver indsats. Efter fremlæggelse af sektorpapirerne fra specialister for de 10 udfordringer, havde CC ekspertpanelet, bestående af udenlandske økonomer, til formål at fortage en prioritering. Mere specifikt var spørgsmålet til dem: Hvis verden kunne blive enige om at spendere 50 mia. \$ de næste 5 år på at forbedre den, hvilke projekter ville så give de største afkast (benefits)? Udfordringerne dækkede mulige indsats på klimaændringer, sygdomme, underernæring og sult, sanitet og vand, migration, subsidier, finansiel ustabilitet, korruption, konflikter og endelig uddannelse. I alt 17 indsats blev prioriteret i fire grupper: virkelig gode projekter, gode projekter, rimelige projekter og dårlige projekter. Denne kommentar forholder sig til bogen, *Global Crises, Global Solutions*, som udkom i efteråret 2004 og er redigeret af Bjørn Lomborg.

I næste afsnit refereres konklusionerne fra konferencen og hovedargumentet er, at den endelige prioriteringsliste fra eksperternes side ikke er logisk opbygget. Et yderligere aspekt, som behandles senere i dette debatindlæg, er manglen på alternativomkostningsberegninger. De forsøges beregnet i to tilfælde: AIDS og vandforsyning. På baggrund af alternativberegningen foreslås en mere logisk opbygget prioriteringsliste, hvor modsat eksperternes liste kun kvantificerede udfordringer og indsats er

medtaget. I konklusionen diskuteres kort anvendeligheden af cost-benefit analyse i debatten om de globale prioriteringer.

### **Konklusionerne fra konferencen**

Man kan indvende, at det havde forøget transparensen, hvis CC havde været mere eksplicit om denne første frasortering. Det kan undre, at værdisætning af natur ikke kom med, når der eksisterer så mange kvantitative analyser af økosystemers ydelser (en oversigt findes hos Turner m.fl. (2003)). Fra analyser fra Zimbabwe ved vi, at skovprodukter kan udgøre op til 30% af de fattiges indkomst, Cavendish (2000). Med den kraftige betoning af benefit-cost ratioer (BCR), som var kriteriet i frasorteringen, havde man forestillet sig en slagkraftig endelig prioriteringsliste fra ekspertpanelet med de tilsvarende cost-benefit ratioer udregnet på hver indsats. Jeg skal i det følgende argumentere for, at den endelige prioritering langt fra var logisk argumenteret.

Ganske i overensstemmelse med almindelig praksis valgte CC ekspertpanelet at lade andre faktorer end benefit-cost ratioen for de 10 områder influere på denne endelige prioriteringsliste. Lomborg skriver, at kvantificeringen var vanskelig og med en kort tids-horisont ikke mulig for eksperterne (s. 7). De 10 papirer tilvejebragte ganske givet i varieret grad de ønskede BCR. Men prioriteringslisten fra eksperterne blev et besynderligt kompromis mellem indsatser med høje BCR og overvejelser om institutionelle betingelser og kompleksitet, således at den endelige dom, prioriteringslisten, fremstår som et meget subjektivt dokument (s. 605-08). Ufordringer, som delvis var kvantificeret (uddannelse og finansiell ustabilitet) røg ud pga. af disse argumenter, mens andre ikke kvantificerede indsatser kom med (prioritet 8, 9, 10 og 14: forskning i stigende vandproduktivitet, fjernelse af red-tape ved start af nye virksomheder, opblødning af migrationsregler for uddannede og andre indvandrere). To helt snævert forbundne indsatser blev klart adskilt (prioritet 5 og 8). Hvad angår klima, satte ekspertpanelet direkte spørgsmålstegn ved beregningerne. Eksperterne mente, at omkostninger sandsynligvis i alle tilfælde, modsat klimapapirets påstand, ville overstige gevinsterne (s. 607). Uenighed har der også været mellem eksperterne, og det synes at være grunden til at finansiell ustabilitet ikke kom med på listen (se Bhagwatis bemærkning s. 610). Endeligt valgte eksperterne ikke at fremlægge eksplicit de tilhørende BCR for de prioriterede indsatser. BCR skal man selv identificere ved at læse sektorpapirerne.

Ekspertpanelet understregede, at de alene forholdt sig til de til konferencen skrevne 10 papirer. En lavere karbonskat koblet til forskning i CO<sub>2</sub> reducerende teknologi kunne være svaret på klimaområdet. Ikke totalt at afvise al tale om global klimapolitik, men at fortage nye beregninger, ifølge eksperterne. Så en ting var eksperternes kommentarer til prioritetslisten, en anden var den måde, hvorpå IMV præsenterede konklusionerne efterfølgende, ikke mindst hvad angår klimaområdet.

*Tabel 1. Eksperternes prioritering af globale indsatser fra Copenhagen Consensus konference, maj 2004.*

<i>Bedømmelse</i>	<i>Udfordring</i>	<i>Indsats</i>	<i>BCR<sup>(a)</sup></i>
Meget god	1. Sygdomme	Kontrol af AIDS udbredelse	14,2 (mikroundersøgelse) <sup>(b)</sup> 49,9 (makroundersøgelse)
	2. Underernæring	Mikroernæringstilskud	15-520
	3. Subsidier	Handelsliberalisering	21,9-29,2
	4. Sygdomme	Kontrol af malaria-udbredelse	1,9-4,7 (makroundersøgelser) 17,1 (malariapakke)
God	5. Underernæring	Udvikling af nye landbrugsteknikker	8,5-19
	6. Vand og sanitet	Lokalbaseret vandforsyning	4,9
	7. Vand og sanitet	Små kunstvandingsanlæg	7
	8. Vand og sanitet	Forskning i vandproduktivitet	Ingen tal, bortset fra henvisning til prioritet 5 (s. 509, note 35)
	9. Regeringsførelse og korrupsion	Reduktion af red-tape ved start af nye virksomheder	Ingen tal
Acceptabelt	10. Migration	Opblødning i migrationsregler for uddannet arbejdskraft	Ingen tal
	11. Underernæring	Forbedring af småbørnskost	1,4-67,1
	12. Sygdomme	Udvidelse af basale sundhedssystemer	2,1-2,6
	13. Underernæring	Reduktion af tilfælde med lav fødevægt	0,6-35,2
Dårlig	14. Migration	Handlingsplaner for ikke faglærte indvandrere	Ingen tal
	15. Klima	Optimal CO <sub>2</sub> skat	2,1
	16. Klima	Kyoto protokollen	1,8
	17. Klima	Forsigtighedsprincippet omsat i CO <sub>2</sub> skat	3,8

*Noter:* (a) BCR: Benefit-Cost Ratios. (b) For enkeltinterventioner varierer BCR fra 0,3 til 466! (s. 107).

*Kilde:* Lomborg, red. (2004), s. 606 samt BCR citeret fra de enkelte papirer.

Man sidder tilbage med en uklar konklusion. Hvis kvantificeringen nu er så vanskelig, og her er jeg enig med Lomborg, og hvis alle i praksis anvender andre kriterier end

blot cost-benefit kolde tal, hvorfor ikke erkende dette faktum og komme med et mere ydmygt udspil? Sammenlignes eksperternes prioriteringsliste med de tilsvarende BCR (ikke gengivet i bogen), er det tydeligt, at prioriteringen ikke slavisk følger en cost-benefit logik (jf. tabel 1). En yderligere komplikation er det faktum, at specialisterne for de 10 udfordringer ikke anvendte samme diskonteringsrate. Det sidstnævnte argument skal ikke forfølges nærmere her.

### **Hvis nu cost-benefit analysen var den eneste ene?**

Alternativt kunne bogen have forsøgt at vise, hvad en prioritering baseret alene på cost-benefit beregninger ville have nået frem til. Den sidste mulighed, nemlig at eksperterne omregnede specialisternes fremlagde estimater, kom ikke på tale. En ny beregning bør omfatte alle de kvantificerede indsatser og desuden tager hensyn til, at de 10 papirer i deres oprindelige udgave fra bogen ikke er helt sammenlignelige. Papiret om sygdomme anvender PPP tal ved beregning af gevinsterne (antal færre døde multipliceret købekraftspariteten), andre analyser bruger bruttonationalindkomsten, og et enkelt papir minimumslønnen i industrien.

Der er ikke international enighed om at anvende PPP tal ved indkomstsammenligning mellem lande, se f.eks. Milanovic (2003) og Radetzki og Jonsson (2002). Et af hovedproblemerne er manglende opdaterede oplysninger fra husholdningsundersøgelser i Kina og Indien. Men enighed eksisterer i hvert fald om, at man ved sammenligning mellem lande og sektorer (eller udfordringer som hos CC) skal anvende en og samme metode. Her svigtede eksperterne ved ikke at gøre opmærksom på, at der var anvendt forskelligt grundlag for beregning af benefits i sektorpapirerne. Papirerne om sygdomme og vand tager endvidere ikke hensyn til alternativomkostninger.

Man kan indledningsvis konstatere, at en fuldstændig cost-benefit analyse burde være gennemført som en marginalanalyse. Gevinsterne ved nye AIDS kampagner f.eks. bør forholde sig til de allerede eksisterende programmer og beregningen alene tage hensyn til marginalgevinsten. Det er yderst vanskeligt i praksis, når vi taler om globale indsatser, og jeg skal ikke komme tilbage hertil senere. Blot konstatere, at her er et uløst problem for næsten alle papirer. Klimapapiret kommer muligvis tættest på at gennemføre en sådan marginalanalyse.

### **Alternativomkostninger**

Sygdomme og vand omfatter indsatser, som er prioriteret i fem ud af de bedste ni tilfælde ifølge eksperternes liste. Problemet er, at en bekæmpelse af AIDS/HIV og malaria kræver et udbygget sundhedssystem. I Mills og Shillcutt om sygdomme er dette en implicit forudsætning. På trods af, at netop udbygning af sundhedssektoren indgår som en anden indsats i papiret om sygdomme. Udbygning af sundhedssektoren

figurerer på listen som prioritet 12. Hertil kommer nedjustering af benefitsiden ved at gå fra PPP tal til bruttonationalindkomsten per capita for at muliggøre sammenligning med andre udfordringer.

Vandpapiret er lidt vanskeligere at gennemregne på ny. Rijsberman argumenterer for, vandforsyning ikke står over for andre brugere. Konkurrence om vand finder kun sted mellem vand til kunstvanding og vand til økosystemer (s. 499-500). Dette virker ikke overbevisende, hvilket også pointeres af en af opponenterne (Vaux s. 536). Han henviser til eksempler fra USA, Spanien og Mexico, hvor kamp om vandressourcer lægger begrænsninger for byudviklingen. Et andet eksempel er fra Jordan. Efter 6-dages krigen i 1967, blev Jordan afskåret fra at bruge vand fra Jordan floden. Vandforsyning til bla. Aman kommer i dag fra et drænet, vådområde Azraq i det østlige Jordan. Vådområdet er forsvundet og med det et vigtigt overvintringssted for trækfugle i Mellemøsten, Andrews (1997). Konflikter er selvfølgelig geografisk specifikke og i hvor høj grad konflikten mellem vandforsyning og vådområder er et globalt problem er et andet spørgsmål. Tit er problemstillingen følgende: hvordan omallokeres vand fra landbrug til menneskelig vandforsyning, jf. Rosegrant, Cai og Cline (2002)? Selv i vigtige landbrugsproducerende lande som Thailand, kan betalingsvilje fra urbane områder overstige betalingsvilje og -evne (ikke ens) fra risbønder, Nordthailand, jf. Tiwari (2000)).

Uanset disse indvendinger, kan Rijsberman have ret i, at værdien af vandforsyning målt i betalingsvilje sikkert overstiger værdi af vand til andre formål (måske fordi økosystemers økonomiske værdi fortsat er debatteret, jf. Turner og m.fl. (2003)). Men der er al mulig grund til at inkludere alternativomkostningerne, uanset om vi taler om vandforsyning eller små kunstvandingsanlæg.

Rijsbermans analyse lider af en anden skavank. For så vidt angår vandforsyning er der tale om en global analyse og den kan derfor uden store problemer sammenlignes med AIDS/HIV indsatsen. Hvad angår små kunstvandingsanlæg er der tale om ekstrapolering af et casestudie fra Asien. Betingelserne er her ideelle med et grundvandspejl kun 1,5-3,5 m. nede i jorden. Højtydende afgrøder produceres i dette område ved anvendelse af lavteknologiske kunstvandingsanlæg (Ganges-Brahmaputra-Meghna flodbassin, jf. s. 518). Rijsberman anvender den gennemsnitlige ekstrairndtægt fra casestudiet multipliceret med antallet af potentielle familier i Afrika og Asien, som vil nyde godt af tilvejebringelse af små kunstvandingsanlæg, uden at tage hensyn til om vand overhovedet findes i samme omfang andre steder i verden, om der er konkurrence på vand fra andre brugere (vådområder f.eks.), eller om udbud af vand påvirkes af klimaændringer, se f.eks. Frederick (2002). Jeg vil i det følgende koncentrere mig om vandforsyning, eftersom den globale beregning virker meget mere overbevisende end beregningen af gevinsten ved udbygning af små kunstvandingsanlæg.

Blandt de højest prioriterede indsatser finder man også underernæring og frihandel. Af hensyn til plads og mit manglende kendskab til disse felter, vil jeg ikke komme ind på dem senere.

### **Hvis nu CC ekspertpanelet alene havde fokuseret på cost-benefit analyser?**

Jeg vil nu forsøge kort at argumentere for, at prioriteringslisten havde set helt anderledes ud, hvis

- (1) kun de kvantificerede projekter er medtaget (uddannelse og finansiell ustabilitet tilføjes, prioritet 8, 9, 10 og 14 fjernes fra den oprindelige liste)
- (2) man tager hensyn til en mindre justering af BCR for sundhed (AIDS) og vand (vandforsyning).

#### *AIDS:*

AIDS beregningen stammer oprindeligt fra Stover m.fl. (2002), som har gennemført en undersøgelse omfattende fattige og mellemindkomst lande, af omkostninger og gevinster (reduktion i tab af dødsfald) ved et udvidet præventions- og mindre behandlingsprogram. Resultatet er, at omkostningerne beløber sig til 1000 US\$ per undgået infektionstilfælde. De totale omkostninger udgør 4,8 milliarder US\$ per år fra 2005, og antallet af undgåede infektionstilfælde er skønnet til at nå 28,5 millioner per år. Ifølge Mills og Shillcutt (s. 95) er middellevealderen per infektionstilfælde estimeret til 23,08 år (YLL – Yearly Lost Lives).

Gevinst: 28,5 millioner x 23,08 x 4.460 US\$ (købekraftsparitet). Omkostninger: 7,3 milliarder US\$ (omkostninger omsat til købekraftsparitet).

Sammenligningen mellem udfordringer (sektorer) kan kun lade sig gøre, hvis man anvender bruttofaktor- eller bruttonationalindkomst per capita. Ingen andre analyser anvender PPP tal, bortset fra et af opponentpapirerne på uddannelse, som anvender begge beregningsmetoder (Wössmann s. 246, note 8). En implicit forudsætning hos Mills og Shillcutt synes at være, at sundhedssystemet er fuldt udbygget. Selv nævner de kun den politiske vilje som en forudsætning for gennemførelse af AIDS programmet (s. 97).

Beregningen af alternativomkostningerne bør i sagens natur være geografisk specifik, da betingelserne varierer fra land til land. Det kan ikke lade sig gøre, så der er nu to andre muligheder. Enten at addere investeringsomkostninger til de løbende omkostninger fra Stover studiet, reference til et studie i Sydafrika, se Geffen m.fl. (2003), eller, mere realistisk, at forvente, at en kraftig satsning på AIDS kommer til at have negativ betydning for andre sundhedsprogrammer. Oplysninger fra AIDS medarbejdere taler om, at AIDS programmer i flere afrikanske lande i dag udgør 30-40% af det sam-



*Tabel 2. Alternativ beregning af fordelene ved en global AIDS indsats. Millioner US\$.*

Forudsætning (3% diskonteringsrate, 8 år)	Årlige indtægter	Årlige udgifter	BCR
Alternativ 1	95.378	4.026	24 (23,7)
Alternativ 2	95.378	5.527	17 (17,3)

*Anm.:* Total gevinst: 28,5 millioner\*23,08 middelleveraldere per infektionstilfælde\*1160 US\$ (bruttofaktorindkomst per capita, gennemsnit for fattige og mellemindkomstlande, Mills og Shillcutt, s. 71). Total omkostninger (A1): 28, 5 millioner \* 1130 US\$ (beregnet gennemsnit fra Sydafrika studie, Greffen, (2003), inklusiv investeringsomkostninger). Total omkostninger (A2): 28, 5 millioner US\$ + 1.964 millioner US\$ årligt (10% af nettogevinsten ved udbygning af basale sundhedsydelser, Mills og Shillcutt, s. 100).

lede sundhedsbudget (Tanzania og Rwanda, baseret på personlig information). Mills og Shillcutt har beregnet den globale gevinst ved udbygning af de basale sundhedsydelser. Vi kunne forsigtigt forudsætte, at AIDS programmet i ulandene reducerer gevinsten ved udbygningen af basale sundhedsydelser med 10%, hvis vi realistisk forventer, at indsatsen på AIDS området ikke bliver additionel fra donorside.

Den sidstnævnte BCR er på linie med BCR fra et Thai case studie med uddeling af kondomer (BRC = 14,2, Mills and Shillcutt s. 95). Man kan med rette indvende, at Mills og Shillcutt ikke anvender Stover studiet korrekt. Forfatterne regner med en gennemsnitindkomst for fattige og mellemindkomstlande (PPP tal), men i virkeligheden burde beregningen været foretaget lande- eller regionsvis. Det kunne give betydelige udsving i den endelige BCR. I min kommentar har jeg begået samme fejl.

#### *Malaria:*

Af tabel 2.22 (Mills og Shillcutt, s. 104) fremgår det, at AIDS pakken er betydelig mere profitabel end malariapakken. Tre gange så høj BCR (jf. også min tabel 1). Af deres sektorrapport fremgår det imidlertid, at anvendes samme PPP tal for de to pakker (3.830 US \$) er der ingen forskel (BCR, malaria = 41,9 (s. 85) og BCR, AIDS = 42,9 (s. 95)).

Det eneste argument for at anvende et højere PPP tal for AIDS end for malaria skulle være, at malaria er mindre udbredt i lande med højere bruttonationalindkomst per capita. Der synes at være en positiv korrelation i Afrika (ikke i Asien og Latinamerika) mellem høj nationalindkomst per capita og høj tilstedeværelse af AIDS tilfælde. Men malaria forekommer stadig i næsten alle ulande og også i mellemindkomstlandene (med undtagelse af visse asiatiske lande). Jeg har anvendt Mills og Shillcutts udgangspunkt på 1.470 US\$, over for 1.160 US \$ for AIDS. Resultatet er, at der ingen forskel er i samfundsmæssig gevinst mellem AIDS indsats og malariabekæmpelse (samme konklusion som hos Mills og Shillcutt hvis man anvender samme PPP tal, jf. ovenfor).

*Vandforsyning:*

Et af indsatsområderne for Rijsberman er gevinsten ved at opfylde FNs 2015 mål for vandforsyning (halvering af jordens befolkning med manglende rent vand og sanitet). Han henviser til et studie af Evans m.fl. (2004), som igen baseres på to af forfatterens baggrundspapir til WHO, Hutton og Haller (2004). Den centrale beregning er opsummeret i tabel 9.4 (s. 514), hvoraf det fremgår, at den globale BCR er estimeret til 4,9 (perioden 2004-2015, diskonteringsrate på 5%).

I et bistandspolitisk perspektiv er der indledningsvis en vigtig pointe. Den største gevinst indtræffer ved reduceret tidstab pga. lettere adgang til toiletter og mindre ventetid ved offentlige vandboringer.<sup>1</sup> Tidsgevinsten udgør mellem 61% og 96% af den samlede gevinst ved lokalforvaltet vandforsyning.<sup>2</sup> De umiddelbare sundhedsgevinster ved færre sygdomstilfælde kan altså ikke retfærdiggøre en investering i den målestok som analysen fra Rijsberman og i sidste ende Hutton og Haller lægger op til. Denne pointe har dog selvfølgelig ingen betydning for beregningen.

Det store problem med beregningen af opfyldelse af FNs 2015 mål er det manglende hensyn til alternativomkostninger. Alle er enige om stigende kamp om vandet i fremtiden. Men her stopper enigheden. Hydrologer tenderer til at male »skræk« scenarier, økonomer tenderer til at fokusere på stigende vandproduktivitet i landbruget som svar. Forskning i vandproduktivitet indgår som prioritet 8. Igen et CC papir med manglende kobling mellem de forskellige indsatser.

Oki m.fl. (2001) anvender en global model for total vandafstrømning og når frem til, at vandstress allerede ved slutningen af forrige årtusind ramte 2,2 milliarder mennesker modsat FNs beregninger, hvor vandstress berørte 0,5 milliarder. Vand stress indekset defineres som  $R_{ws} = W \cdot S / Q$ , hvor  $W$  = vandabstraktion,  $S$  = afsaltet havvand og  $Q$  = vandmængde til rådighed (udbud). Hvis  $R_{ws} > 0,4$  er der tale om vandstress. Problemet er, at indekset er statisk og ikke tager højde for, at produktivitetstilvækst kan sænke vandaftaget i fremtiden. Lignende metoder anvendt af World Resources Institute tager ikke engang højde for afsaltning af havvand. I Mellemøsten er brøken i nogle lande over 1000, WRI (2003). I den anden ende af skalaen finder man dokumenter som Hutton og Haller (2004), som ikke engang nævner hvilken stigning i vandabstraktion en opfyldelse af FN 2015 mål afstedkommer.

I Hutton og Haller omtales en følsomhedsanalyse. Hvis man anvender bruttonationalindkomsten i stedet for minimumsløn i industrien og samtidig har et pessimistisk skøn til omkostningsudviklingen, opnås meget lave BCR. I bedste tilfælde for det øst-

1. »This shows that the greatest proportion of time gain is from sanitation interventions - i.e. the closer proximity of toilets or less waiting time for public facilities«, Hutton og Haller (2004), s. 32.

2. Beregnet som andelen af »annual time savings« i forhold til »total economic benefits« for de vigtigste 6 delregioner i verden (hvor gevinsten er størst). Tabel A 2. 16 i forhold til tabel A 2. 18 (ibid.: s. 61 og s. 63), intervention 2 (Millenium targets with sanitation targets).

Tabel 3. Forskellige beregninger af fordelene ved en global vandforsyningsindsats (opfyldelse af FNs 2015 mål). Millioner US\$.

Forudsætning (5% diskonteringsrate, 12 år)	Årlige indtægter	Årlige udgifter	BCR
Rijsberman	4550	921	4,9
Ved forventet fald i landbrugsproduktion (50% af fald i 2025)	4550	4864	0,9
Ved forventet fald i landbrugsproduktion (25% af fald i 2025)	4550	2893	1,6

*Anm.:* Forventet fald i landbrugsproduktion i ulandene når i 2025 op på 7.500 millioner US\$ for korn (30 millioner tons x 250 US\$), 238 millioner US\$ for sojabønder (0,5 i ulandene (skøn) x 1,9 millioner tons x 250 US\$) + 150 millioner US\$ for sødkartofler, jf. Rosegrant m.fl. (2002), s. 160-62).

*Kilde:* Rijsberman, i Lomborg (2004), s. 514.

lige Afrika en BCR på 2,5, mens andre regioner (ikke alle regioner er nævnt i følsomhedsanalysen) viser en negativ ratio (Hutton og Haller s. 37-38). Argumentet skal ikke forfølges her (udregning ikke foretaget), men en del tyder på, at tages der højde for arbejdsløshed (anvendelse af bruttonationalindkomsten), så falder den samlede ratio under nul.

Mit argument er, at selv i tilfælde af anvendelse af minimumsløn, falder den samlede BCR for vandforsyning kraftigt, hvis man tager hensyn til alternativomkostninger. I en analyse af behov for vand til landbrug, industri, kvæg, vandforsyning og vådområder, frem til 2021-25 forventes et betydeligt fald i vand til landbrugsformål. I scenariet bæredygtig vandudnyttelse regnes med, om end urealistisk, at alle får adgang til rent vand i 2025 (Rosegrant m.fl. s. 56). Dette kunne sammenlignes med FNs 2015 mål (50% af målet i 2025). Beregningen kan kun opfattes som et skøn, pga. stor uenighed om hvor meget af verdens vandressourcer, som bliver genstand for overefterspørgsel i fremtiden. Ifølge International Water Management Institute (2004 s. 2) er Rosegrants skøn optimistisk i forhold til andres analyser.

#### *Uddannelse:*

Ifølge Wössmann (s. 246) skønnes BCR for fjernelse af skolepenge fra Uganda til at være mellem 10-37. Der er tale om et casestudie, men dette var også tilfældet for små kunstvandingsanlæg. Lidt vanskeligere er det at udregne alternativomkostningerne i dette tilfælde. Jeg vil derfor ud fra et forsigtighedsprincip anvende en ratio på 10.

#### *Finansiel ustabilitet:*

Eichengreens analyse indeholder en opgørelse af gevinster og omkostninger ved at

*Tabel 4. Revurdering af eksperternes prioritering af globale indsatser fra Copenhagen Consensus konference, maj 2004 (kun konferencepapirer er medtaget).*

<i>Bedømmelse</i>	<i>Udfordring</i>	<i>Indsats</i>	<i>BCR</i>
Ikke omvurderet	1. Underernæring	Mikroernæringstilskud	15-520 <sup>(a)</sup>
Kommenteret, ikke omvurderet	2. Finansiell ustabilitet	Lån udstedt i udviklingslandes egen valuta	>100
Ikke omvurderet	3. Underernæring	Forbedring af småbørnskost	1,4-67,1 <sup>(a)</sup>
Ikke omvurderet	4. Underernæring	Reduktion af tilfælde med lav fødevægt	0,6-35,2 <sup>(a)</sup>
Ikke omvurderet	5. Subsidier	Handelsliberalisering	21,9-29,2
Ikke omvurderet	6. Underernæring	Udvikling af nye landbrugs-teknikker	8,5-19 <sup>(a)</sup>
	7. Sygdomme	Kontrol af AIDS udbredelse	17,3
Kommenteret, ikke omvurderet	8. Sygdomme	Kontrol af malaria udbredelse	17,1
Kommenteret, ikke omvurderet	9. Uddannelse	Fjernelse af skolepenge	10
Ikke omvurderet	10. Vand og sanitet	Små kunstvandingsanlæg	7
Ikke omvurderet	11. Klima	Forsigtighedsprincippet omsat i CO <sub>2</sub> skat	3,8
Ikke omvurderet	12. Klima	Optimal CO <sub>2</sub> skat	2,1
	13. Sygdomme	Udvidelse af basale sundheds-ydelser	1,9 <sup>(b)</sup>
Ikke omvurderet	14. Klima	Kyoto protokollen	1,8
	15. Vand og sanitet	Vandforsyning	1,6

*Noter:* <sup>(a)</sup> Højeste BCR anvendt. <sup>(b)</sup> Den oprindelige beregning minus 10% af den totale gevinst (figurerer som alternativomkostninger under AIDS pakken).

reducere den finansielle ustabilitet og dens effekt på ulandenes gældsbeholdning. Ideen er, at internationale finansielle institutioner bør udlåne i landenes egen valuta. Herved undgås misforhold mellem et lån optaget i US\$ og virkningerne på gældsbyrden af en pludselig opstået devaluering af landets egen valuta i kriseperioder. Ved devaluering stiger udgiften til afbetaling omregnet i landets egen valuta. Dette problem kunne imødegås ved, at lånene optages i et mix af forskellige valutaer.

BCR kan estimeres til over 200 ved 5% diskonteringsrate (s. 267). Det virker som et urimeligt højt estimat. En af opponenterne, Wyplosz (s. 288-289) indvender også, at

Eichengreen overvurderer gevinsterne (der vil stadig være bankkriser) og undervurderer omkostningerne (stor risiko for økonomisk kaos i visse lande, og derfor vil disse nye sammensatte valutalån bære høj rente). Skal jeg følge princippet om at komme så tæt på CC papirernes egen vurdering så muligt, kunne man anvende en BCR på 100 (synes dog stadig meget høj). Alternativomkostningerne er svære at vurdere i dette tilfælde.

### **En omvurdering af CC prioriteringslisten**

Den nye liste bærer selvfølgelig præg af, at jeg ikke har revurderet underernæring og liberalisering af handel. I praksis er det umuligt at undgå at anlægge betragtninger ud over den rene cost-benefit analyse i en prioriteringsdebat, specielt når der er stor spredning i opgjorte BCR (gælder primært for underernæring). Muligvis burde papiret om underernæring således revurderes. Den store spredning skyldes fokuseringen i papiret på et større antal casestudier, hvor BCR varierer betydeligt (såkaldt mesa analyse).

### **Konklusion: Kan vi anvende cost-benefit analyser globalt?**

Lomborgs bog om de globale kriser indeholder et væld af informationer og mange spændende diskussioner inden for 10 områder/udfordringer, som er vigtige for menneskeheden. Men logikken brister, da et ekspertpanel forsøger at etablere en videnskabelig begrundelse for en bestemt og entydig prioritering af globale udfordringer. For flere af indsatsområderne er der ikke tale om globale løsninger. Casestudier som ophøjes til globale undersøgelser (kunstvandring, underernæring, uddannelse), manglende hensyn til kompleksiteten (finansiell ustabilitet), manglende kobling til andre indsatser inden for samme eller anden udfordring (inden for sygdomme og vand og mellem vand og underernæring), fravær af hensyn til vigtige områder af de fattiges levevilkår (afhængighed af naturressourcer) og manglende hensyn til alternativomkostningerne (AIDS og vandforsyning). Ekspertpanelet kunne sagtens have valgt at indtage en mere ydmyg holdning i dets konklusion. I stedet fik vi en meget subjektiv og ikke særlig logisk opbygget prioriteringsliste. Benefit-cost ratioen for de prioriterede indsatser var generelt højere i den øverste halvdel af dets prioriteringsliste end i den anden halvdel, men den indbyrdes fordeling mellem indsatser var som sagt subjektiv fastsat.

Formålet med min kommentar har blot været at påpege, at selv enkelte små justeringer i udvalgte papirer kan medføre en helt anden prioritering. Jeg har ikke undersøgt alle papirer, og derfor er min alternative liste kun en indikation af de problemer, de konkrete sektorpapirer indeholder. At nå frem til en endelig videnskabelig prioritering synes illusorisk.

Lad os hellere få en bredere diskussion af, om cost-benefit analyser, hvis de bliver globale en gang i fremtiden, overhovedet kan anvendes i en prioriteringsdiskussion for

hele kloden. Flere af de ovenstående problemer forsvinder ikke, selvom vi kan komme af med afhængighed af casestudier ved globale undersøgelser. Endvidere, kan vi overhovedet sammenligne et projekt på 8 år (AIDS pakken) med indsatser på klimaområdet (300 års tidshorisont)? Her er det et problem med fremtidige generationers tidspræferencer. En af eksperterne mener ligefrem, at eftersom fremtidige generationer vil blive meget rigere, så skal vi ikke tage hensyn i dag (Schelling s. 627). Som omtalt virker klimaeffekten tværgående blandt andet ved påvirkning på biodiversitet, vand, ændringer af økosystemer og sundhed. Kan det opgøres fornuftigt i økonomiske termer? Hvis ikke, hvilke alternativer findes der da?

#### Litteratur

- Andrews, I. J. 1997. Birding in Jordan, *Dutch Birding*, nr. 19(2), s. 49-60.
- Cavendish, W. 2000. Empirical regularities in the poverty-environment relationship of rural households: Evidence from Zimbabwe, *World Development*, Vol. 28, 11, s. 1979-2003.
- Evans, B., G. Hutton og L. Haller. 2004. Closing the sanitation gap – the case for better public funding of sanitation and hygiene, Round Table on Sustainable Development, OECD, 9-10 March ([www.oecd.org](http://www.oecd.org)).
- Frederick, K. D. 2002. Water resources and climate change, *The Management of Water Resources*, Vol. 2, Edgar Reference Collection, Cheltenham, UK.
- Geffen m.fl. 2003. The cost of HIV prevention and treatment interventions in South Africa, *CSSR Working Paper*, no. 28, University of Cape Town, South Africa.
- Hutton, G. og L. Haller. 2004. Evaluation of the costs and benefits of water and sanitation improvements at the global level, WHO/SDE/WSH/04.04 ([www.who.org](http://www.who.org)).
- International Water Management Institute, Sri Lanka. 2004. Investing in water for food, ecosystems and livelihoods, Blue Paper, Stockholm Conference.
- Lomborg, B., red., 2004. *Global crises, global solutions*, Cambridge University Press, Cambridge UK.
- Milanovic, B. 2003. The two faces of globalization: against globalization as we know it, *World Development*, Vol. 31,4, s. 667-83.
- Oki, T. m.fl. 2001. Global assessment of current water resources using total runoff integrating pathways, *Hydrological Sciences*, Vol. 46, 6, s. 983-95.
- Rosegrant, M. W., Ximing Cai og Sarah A. Cline. 2002. *World water and food to 2025. Dealing with scarcity*. Washington D.C., USA.
- Radetzki, M. og B. Jonsson. 2002. The expanding global income gap: How reliable is the evidence?, *The European Journal of Development Research*, Vol. 14, 1, s. 244-63.
- Stover, J. m.fl. 2002. Can we reverse the HIV/AIDS pandemic with an expanded response?, *The Lancet*, Vol. 360, s. 73-77.
- Tiwari, D. N. 2000. Sustainability criteria and cost-benefit analysis: an analytical framework for environmental-economic decision making at the project level, *Environment and Development Economics*, Vol. 5, s. 259-88.
- Turner, K. R. m.fl. 2003. Valuing nature: lessons learned and future research directions, *Ecological Economics*, Vol. 46, s. 493-510.
- World Resources Institute. 2003. *World resources 2002-2004*, Washington D.C., USA.

## Replik

### Copenhagen Consensus. Elfenbenstårn eller virkelighed?

Bjørn Lomborg

Copenhagen Business School, E-mail: [bjorn@lomborg.com](mailto:bjorn@lomborg.com)

Henrik Meyer

IBT – Center for Anvendt Datalogi, E-mail: [henrik.meyer@ibt.ku.dk](mailto:henrik.meyer@ibt.ku.dk)

Torkil Casse rejser en række vigtige spørgsmål i forbindelse med Copenhagen Consensus og bogen »Global Crises, Global Solutions«, resultatet af over 40 økonomers indsats. Det er godt, at der også opstår en faglig debat. Enkelte af Casses kritikpunkter peger på reelle metodiske problemer, men desværre er artiklen også udtryk for nogle misforståelser. Vi vil afholde os fra at diskutere de meget specifikke kritikpunkter Casse fremfører på de enkelte papirer. Det rækker pladsen ikke til og gøres bedre af forfatterne selv.

Det er vigtigt at klargøre, hvad formålet med Copenhagen Consensus først og fremmest var, nemlig at skabe opmærksomhed om økonomisk prioritering som en daglig politisk realitet samt at finde nogle af de bedste løsninger på verdens problemer. Copenhagen Consensus har derfor fra starten taget det for nogle ubehagelige skridt ud af elfenbenstårnet. Noget af Casses kritik bunder i dette. Copenhagen Consensus blev heldigvis heller ikke en akademisk eksercits i økonomisk teori. Derimod baserede Copenhagen Consensus sig på empirisk økonomi. Casse kritiserer dog heller ikke den økonomiske børnelærdom om at prioritere knappe ressourcer – det er der nu heller ikke mange, der gør. Derimod kritiserer Casse især den konkrete prioritering som panelet foretog. Den er »subjektiv«. Men Copenhagen Consensus og »Global Crises, Global Solutions« er ikke dogmatisk, og selvom cost-benefit analyse ikke er svaret på alt, er det dog et meget bedre redskab, end hvad der hidtil har foreligget af sammenlignende økonomisk information til beslutningstagerne. Og hvad er Casses alternativ til en økonomisk analyse og sammenligning?

Lad os starte med en af Casses afsluttende konklusioner: »Endvidere, kan vi overhovedet sammenligne et projekt på 8 år (AIDS pakken) med indsatser på klimaområdet (300 års tidshorisont)? Her er det et problem med fremtidige generationers tidspræferencer. En af eksperterne mener ligefrem, at eftersom fremtidige generationer vil blive meget rigere, så skal vi ikke tage hensyn i dag (Schelling s. 627).« Casse har naturligvis en principiel pointe, men problemet er, at han taler fra elfenbenstårnet. Det er mu-

ligt, at verden burde have en anden tidspræference, og at vi burde kunne forudsige bedre, hvordan fremtidige generationers præferencer vil se ud (her er et oplagt arbejde til Casse, som kunne få indflydelse på fremtidige prioriteringer), men indtil da tager Copenhagen Consensus udgangspunkt i den empirisk konstaterede diskonteringsfaktor. Copenhagen Consensus er bestemt ikke perfekt, men prøver at forholde sig til verden som den ser ud, ikke som den burde se ud. Og i den verden træffer politikerne beslutninger mellem klima, AIDS og operahuse. Casse leverer en principielt korrekt kritik, som er så uanvendelig i virkelighedens verden.

Det betyder ikke, at der ikke er mulighed for forbedringer. Det er der, og de erfaringer, der er kommet fra Copenhagen Consensus i 2004, vil også blive indarbejdet i det næste Copenhagen Consensus i 2008. Her tænkes bl.a. på at inddrage endnu større forskerkredse over en længere periode og nærmere undersøgelser af uddannelse, lokale konflikter og finansiell ustabilitet. Det var også ærgerligt, at bl.a. naturressourcer ikke var med, men spørgsmålet til Casse er, hvilket af de 10 andre problemer der i så fald skulle have været taget af listen?

En, desværre udokumenteret, påstand i artiklen er det angivelige misbrug som Institut for Miljøvurdering har gjort sig skyldig i. Det skulle især være i forhold til klimaændringer. Vi vil her blot konstatere, at panelet enstemmigt afviste de løsninger, de blev præsenteret for. Det betyder jo ikke, at der ikke kan tænkes bedre løsninger, men blot at de ikke ligger lige for og vigtigt i forhold til hele prioriteringen, at de næppe vil havne i toppen under alle omstændigheder. Casse fremfører beklageligvis, at: »Eksperterne mente, at omkostninger sandsynligvis i alle tilfælde, modsat klimapapirets påstand, ville overstige gevinsterne«. Dette er forkert. Klimapapiret af Cline er beregnet med diskonteringsrater fra 0,5 til 4,5%. Kun ved rater under 2% er benefit-cost ratioen over 1 og altså favorabel (s. 37). Panelet valgte at lytte til opponenternes argumenter om, at Clines tiltag var for drastiske, samt, og det er nok så vigtigt, at en diskonteringsrate på under 2% er urealistisk lav.

Copenhagen Consensus blev en enestående kombination af nogle af verdens bedste økonomer, som leverede grundlæggende empirisk indsigt i 10 af verdens største problemer. Med panelets endelige vurderinger – som også fremgår individuelt – leverede Copenhagen Consensus en økonomisk grundlæggende anbefaling af, hvor verden skulle bruge ekstra \$50 milliarder over de næste 4 år. I stedet for at stille spørgsmål uden at give svar, gav Copenhagen Consensus en række svar på løsninger af verdens største problemer.

#### *Litteratur*

Lomborg, B., red. 2004. *Global Crises, Global Solutions*, Cambridge University Press, Cambridge UK.



## Boganmeldelser

Torben Pedersen. *Dansk erhvervslivs planering i globaliseringen*. Magtudredningen, Aarhus. 2004. 95 s. Anmeldt af Carsten Kowalczyk. The Fletcher School, Tufts University og Institut for Økonomi, Aarhus Universitet.

Danmark, en af verdens mest åbne økonomier, har en lang tradition for den internationale handel af varer og tjenesteydelser, og bevægelse af kapital, arbejdskraft, og teknologi, som økonomisk globalisering repræsenterer. Professor ved Handelshøjskolen i København, Torben Pedersen, foreslår i indledningen til dette bidrag til magtudredningen, at »globaliseringen ikke er en naturlov – det er ikke en selvkørende proces, der vil fortsætte, uanset hvilke prioriteringer og politiske beslutninger der træffes. Globaliseringen er tværtimod i høj grad bestemt af de prioriteter og politiske beslutninger, der træffes af forbrugere, vælgere, virksomheder og politikere.« (s. 8). Han vælger i dette lille bind at fokusere på, hvordan multinationale selskaber forbinder den danske økonomi til udlandet. Af hvilke årsager vælger virksomheder at blive multinationale eller på anden måde at flytte nogle af deres aktiviteter til udlandet? Hvordan vælger de at gøre det? Og hvad er effekterne på vigtige økonomiske variable i Danmark af virksomheders internationale aktiviteter?

Hans svar, kort sammenfattet, er, at danske virksomheder vælger at producere i udlandet, enten selv eller ved outsourcing, primært for at få adgang til udenlandske markeder; at danske virksomheder opererer i de nære østeuropæiske lande, men ikke har megen tilstedeværelse i de store og hurtigt voksende lande, Indien og Kina; og at danske virksomheders aktiviteter i udlandet ikke synes at have forårsaget nogen væsentlig reduktion af beskæftigelsen i Danmark. Disse resultater bygger

især på spørgeskemaundersøgelser af danske industrivirksomheder.

Bogen har seks kapitler. I et kort introduktionskapitel fremhæver Torben Pedersen lavere forbrugspriser som en af gevinsterne ved globalisering, og han nævner i kapitel 2 faldende handelsomkostninger og -barrierer som vigtige årsager til den nuværende globaliseringsproces. Det er et interessant spørgsmål, om ikke de faldende omkostninger og barrierer også har et endogent element – at de falder, fordi politiske beslutningstagere og investorer arbejder på at nedbringe dem i den forventning, at der er gevinster ved øget økonomisk samkvem. Sådanne forventninger støttes af den økonomiske forskning: Afhængig af hvorfra et land og dets partnere starter processen, og af hvor omfattende liberaliseringen er, er der estimeret statistiske gevinster i nationalindkomsten på fra en til ti procent om året fra bedre resourceallokering og øget konkurrence, og dynamiske gevinster på over ti procent om året, hvis liberaliseringen skaber økonomisk vækst gennem investeringer og teknologiske innovationer. Når et lands borgere erfarer, hvilke forbedringer i forbrug, helbred og uddannelse, sådanne tal medfører, bliver det næppe let for dets politikere at standse liberaliseringsprocessen – eller at sætte den i bagegear. Hvad der dog er ganske klart, ikke mindst fra erfaringen indtil for nylig i de øst-europæiske lande og fra situationen i mange af verdens fattigste lande i dag er, at globalisering – og liberalisering – er processer, som kan være svære at starte.

Kapitel 3 giver en meget detaljeret indførelse i begrebet »direkte investeringer« og nævner andre former for internationalt økonomisk samarbejde mellem virksomheder såsom strategiske alliancer, joint ventures, og outsourcing. Kapitlet dokumenterer også de multinationales stigende betydning for den globale økonomi. Torben Pedersen viser, at investeringer til og fra Danmark er af samme størrelsesorden og er relativt balanceret i for-

hold til individuelle handelspartnere bortset fra i forhold til Sverige og USA, som begge har placeret betydelig mere kapital i Danmark end danskere har i hvert af de to lande. Han viser også, at dansk kapital i de baltiske og østeuropæiske lande langt overstiger dansk kapital i Indien, Kina og det øvrige Asien. Torben Pedersen er først og fremmest beskrivende i denne bog, og han afholder sig stort set fra normative betragtninger. Han udtrykker dog nogen bekymring over, at danske virksomheder ikke er stærkere repræsenteret i Asien givet regionens store vækstpotentiale.

Kapitel 4 er spændende læsning. Her rapporterer Torben Pedersen resultater fra de tre etableringsundersøgelser som Dansk Industri og Handelshøjskolen i København gennemførte i 1991, 1998 og i 2003. De adspurgte industrivirksomheder udviser en svagt faldende eksport men en eksplosion i omsætningen i deres datterselskaber i udlandet, og en konstant beskæftigelse i Danmark men en kraftigt stigende beskæftigelse i udlandet. Virksomhederne angiver endvidere, at markedsadgang er langt den vigtigste årsag til etablering i udlandet, at omkostningsfordele spiller en begrænset rolle, og at læringsmæssige fordele næsten ingen rolle spiller.

Undersøgelserne synes også at vise, at andelen af højtuddannede (og faglærte) i danske udenlandske datterselskaber er større end andelen af ufaglærte og også større end i Danmark for produktionen. En mulighed er, at resultatet afspejler et målingsfænomen, da det ofte er svært at sammenligne uddannelseskategorier og kvalifikationer internationalt. Man kunne således forestille sig, at en dygtig dansk faglært har erhvervet lige så meget uddannelse og viden om sit fag og sin produktion som en kandidat fra et udenlandsk teknisk universitet, hvor sidstnævnte måske ville blive kategoriseret som ingeniør og ikke som faglært. Men det er også muligt, at danske aktiviteter i udlandet er intensive i højtuddannet arbejdskraft i sammenligning med danske aktiviteter i Danmark: For eksempel har de østeuropæiske lande, hvor Torben Pedersen viser, at danske virksomheder er særligt akti-

ve, i mange år satset på videnskabelige og tekniske uddannelser, som har medført et relativt stort udbud af højtuddannede (ingeniører, etc.). Endelig kan resultatet skyldes, at danske multinationale virksomheder udfører forskellige aktiviteter – og derfor har forskellige kvalifikationsbehov – i Danmark og i udlandet. Torben Pedersen viser for eksempel, at en forholdsvis stor andel af danske virksomheders beskæftigelse i udlandet er i salg og markedsføring – aktiviteter, som ofte kræver et højt uddannet personale sammenlignet med produktion, hvor især arbejdskraftintensive processer ofte kan udføres af ufaglært arbejdskraft.

Kapitel 5 præsenterer en diskussion af begreberne outsourcing – en virksomheds udlicitering af interne opgaver – og offshoring, som forekommer, når outsourcing sker til udlandet. Torben Pedersen foreslår, at omkostningsbesparelserne for danske virksomheder er lige så store ved at gå til de østeuropæiske lande som til Asien, især hvis geografiske og kulturelle faktorer tages i betragtning. Noget overraskende nævner han først i dette kapitel, at serviceydelser kan handles internationalt. Betragtningerne i dette kapitel er vigtige og burde måske have været præsenteret i et af bogens første kapitler. Et kort kapitel 6 sammenfatter og konkluderer.

Bogen er velskrevet og kan med udbytte læses af enhver med interesse i dansk industris placering i globaliseringen. Bogen inspirerer også til yderligere forskning:

Der er et klart behov for at udvide analysen til at omfatte Danmarks internationale handel og investeringer i serviceaktiviteter. For eksempel er skibsfart og ledelse af arbejdskraftintensive serviceaktiviteter vigtige danske eksporterhverv og bør indgå i en beskrivelse af dansk erhvervsliv og globaliseringen. En målrettet satsning på, at Danmark forbliver et førende videnssamfund, vil medføre, at servicesektoren bliver endnu vigtigere i det fremtidige danske eksport- og investeringsbillede.

Der er også behov for yderligere analyse af udenlandske investeringer i Danmark og af

insourcing – og af begges effekter på danske indkomster, priser og beskæftigelse. Bogens fokusering på udgående investeringer og outsourcing kan give et skævt billede af konsekvenserne af globalisering. Globalisering giver jo også forøgede muligheder for udenlandske aktiviteter i Danmark, dels på grund af yderligere dansk liberalisering, dels på grund af at udenlandsk liberalisering forøger det internationale udbud af udenlandske varer, tjenesteydelser og kapital.<sup>1</sup> For eksempel har Kinas åbning til international handel og kapitalbevægelser ikke blot medført store udenlandske investeringer i Kina, men også allerede nu kinesiske investeringer i selv de rige OECD lande.

De mest interessante resultater i bogen er baseret på spørgeskemaundersøgelser. Men sådanne undersøgelser bygger ofte på selvrapportering, hvilket kan medføre systematisk bias i resultaterne. Hvor mange eksportvirksomheder i Danmark – et samfund, som

1. Der opstod i de amerikanske medier i 2004 en debat om effekten af amerikansk outsourcing på beskæftigelsen i USA. Debatten forstummede brat, da det blev påvist, at der skabes flere jobs ved insourcing til USA, end der tabes ved outsourcing fra USA.

Sørensen, Peter Birch and Hans-Jørgen Whitta-Jacobsen, »*Introducing Advanced Macroeconomics: Growth and Business Cycles*«. McGraw-Hill 2004. Reviewed by Neil Rankin. University of Warwick.

This is an advanced undergraduate/introductory graduate macroeconomics textbook. The authors argue in their preface that there is a wide gap between the level of the typical undergraduate »intermediate macroeconomics« textbook and that of the typical graduate macroeconomics textbook, and a lack of textbooks which help students bridge this gap. I

har flotte traditioner i viden, kreativitet, og design – vil for eksempel villigt erklære, at de søger til udlandet for at lære af dygtige udlændinge, selv hvis det var en vigtig årsag? Der er flere måder at arbejde videre på det spørgsmål: For eksempel kan analysen udvides til at undersøge, hvorfor de danske virksomheder, som ikke går til udlandet, vælger en helt indenlandsk strategi. Man kan også undersøge, om udenlandske virksomheder investerer i Danmark for at få adgang til dansk viden. Endelig er der den internationale litteratur om viden-spillovers i international handel og investeringer, som kan bidrage til at forstå de danske resultater.

Torben Pedersen motiverer bogen med, at der er et behov for at forstå samspillet mellem globaliseringens økonomiske og politiske kræfter, og hans diskussion af effekterne på dansk arbejdskraft af danske investeringer i udlandet og outsourcing er en god begyndelse. En fuld analyse kræver, at effekterne af globalisering på andre grupper, herunder forbrugerne, også undersøges, og at deres muligheder for politisk at påvirke forløbet belyses. Der er behov for yderligere forskning i den politiske økonomi af globalisering og det danske samfund.

agree that there is a big gap in the market here. The existing books which occupy this sort of territory tend either to attempt only partial coverage of the field, being a selection of topics according to the personal preferences of the author, or to be devoted to more specialised fields such as growth, monetary or international economics. With their new book, Sørensen and Whitta-Jacobsen (SWJ, henceforth) have not merely inserted a helpful stepping-stone in this gap, they have filled it in almost completely. »Introducing Advanced Macroeconomics« is a thumping great contribution to this less-visited part of the textbook market. It is substantial in every way: more than 800 pages long, with 26 chap-

ters, and a comprehensive coverage of all the topics which one would expect to find in a complete macroeconomics course.

The sub-title indicates the organisation of the book. It is divided into two halves. Book 1 deals with »The Long Run«, and consists of 9 chapters on growth and 3 on structural unemployment, while Book 2 deals with »The Short Run« and consists of 9 chapters on the closed economy and 4 on the open economy. The inclusion of structural unemployment alongside growth as part of »The Long Run« should help students to appreciate better this fundamental division within macroeconomics, something which is not clear to many undergraduates: that is, the division between models which are concerned with trends rather than cycles and in which only real factors matter, and models which are concerned with fluctuations about the trend and in which monetary factors and nominal rigidities play a key role. Nevertheless, the fact that it is possible to devote as much as half the book to »The Long Run« is a sign of how important the subject of growth has become in contemporary macroeconomics. Not long ago growth would quite commonly have been the last chapter out of 10 in an undergraduate macro text. In SWJ growth not only comes first (the new orthodoxy in macro textbook layout), but also takes up more than one third of the book.

Another distinctive feature of SWJ's approach is the attempt to cover all the usual topics in macroeconomics using a minimum number of different models. It is true that macroeconomics, to a much greater extent than microeconomics, often appears to students to consist of a huge variety of models, each using specialised assumptions in order to address the issue at hand, with no apparent consistency of approach. This makes it hard for students to understand how the different models and issues are related. SWJ aim to overcome this by using as few models as possible and by preserving the maximum number of common elements between them. Of course, there are underlying reasons for this

tendency to variety which relate to the nature of macroeconomics as a subject. It is not just due to self-indulgence or intellectual untidiness on the part of macroeconomists. Macroeconomics by definition has to deal with the whole economy – with a general equilibrium system – so that macroeconomists have in principle to model more of the economy than microeconomists. There are certain elements which they cannot leave out or else their models would be mere partial equilibrium models. This extra burden means that sacrifices have to be made somewhere in order for the analysis not to be overwhelmed by complexity. It hence results in shortcuts being taken with parts of the model, with different shortcuts being adopted depending on the particular macroeconomic phenomenon one is trying to study. It is largely from this that the »non-nested« nature of so many macro-models stems. The potential cost, then, of trying to achieve model consistency across different topics is that the model may become unnecessarily cumbersome for the modelling task in hand.

Do SWJ avoid this cost? The answer is that they make a very good attempt to avoid it. As a test of how well they succeed, consider their treatment of the AD-AS model. Some version of AD-AS lies at the heart of most undergraduate macroeconomics texts. For their baseline model, SWJ choose a set-up in which short-run nominal rigidity arises from money wages being set one period in advance. Inflation expectations, on which wage setting is based, are modelled as »static«: that is, equal to the lagged value of actual inflation (at least, in the closed economy; SWJ use a slightly different assumption in the open economy). Such a choice means that the majority of the short-run analysis is therefore based on »backward-looking«, »non-rational«, expectations. It further means that – since SWJ analyse the dynamic implications of all their models thoroughly – the dynamics (of the closed-economy model, anyway) are completely driven by the static expectations assumption. Possibly rather too much there-

fore rests on this assumption, which is used repeatedly across several chapters. All the usual criticisms of the rational expectations school of course apply to it, and indeed these are laid out in Chapter 21. On the other hand it serves as a convenient and simple vehicle through which SWJ can give, for example, a thorough introduction to the principles of business cycle analysis, viewing such cycles as the outcome of random shocks and propagation mechanisms (the propagation mechanism in this case, therefore, being the gradual adjustment of inflation expectations).

Another aspect of SWJ's exposition of AD-AS is that from the outset their basic AD-AS diagram uses inflation rather than the price level on the vertical axis. The short-run AS curve is hence really a short-run Phillips curve. The advantage of this is that it better matches discussions of real-world macroeconomic problems, which are usually in terms of inflation rather than the price level; and that it avoids the tedious repetition which would be necessary if AD-AS were first presented using the price level as the nominal variable and then converted to an analysis using the inflation rate. On the other hand, students who are already familiar with the price-level version of AD-AS (which is the traditional one, and still by far the most common in the textbooks) may be puzzled by the difference, and may want to know how the two are related. It might have been a good idea at least to have mentioned that there exists another version which students are likely to encounter. An example of where the consistent use of the »inflation-rate« AD-AS model throughout the book arguably has a cost, in terms of making things harder to understand, is in the fixed-exchange rate case of the open economy – on which more below.

As regards the AD curve (whose form of course depends on the way monetary policy is specified), a notable feature is that SWJ have opted to dispense entirely with the traditional fixed-money-supply representation of monetary policy, and instead to use a version of the »Taylor Rule« for interest-rate setting through-

out. Some other textbook authors have also recently adopted this approach, notably DeLong. One advantage is that it brings the modelling of monetary policy much closer to current central bank practice. Another advantage is that, if one is willing to distort the idea of the Taylor Rule slightly, presenting it as a rule for the real, rather than the nominal, interest rate (as SWJ and DeLong do), then, together with the usual »IS« relationship, it generates a very convenient AD relationship in output-inflation space. This can neatly be combined with the Phillips curve for the purposes of comparative static and dynamic analysis. Such is SWJ's basic AD-AS set-up, which they use repeatedly. This set-up is elegant and, to those readers more familiar with the traditional analysis based on a fixed money supply and hence an »LM« curve (which has no role under Taylor-rule analysis), the fresh perspective which it provides is stimulating.

Having said this, the ubiquitous use of a Taylor-rule monetary policy regime does have drawbacks. One is that there is no obvious parameter to represent a simple discretionary monetary policy expansion or contraction. This means that comparisons of the relative effectiveness of monetary and fiscal policy (which are standard fare in most undergraduate macro texts) cannot easily be made. Another is that all shocks have effects which are to some extent influenced by the values chosen for the feedback coefficients in the Taylor Rule, so it is not clear what an »inactive« monetary policy is. When the model is studied under rational expectations, it is also rather disorientating – though novel – to see the Policy Ineffectiveness Proposition presented through the use of a Taylor rule and not a monetary feedback rule. By omitting altogether an analysis under the alternative assumption of a fixed money supply, SWJ furthermore fail to communicate some of the subtler issues associated with conducting monetary policy through interest-rate control, such as price-level indeterminacy under rational expectations, or path-dependence of the price level under backward-looking expect-

tations. An understanding of this latter could be helpful when it comes to studying the fixed exchange rate regime. Like a fixed money-supply regime but unlike a Taylor-rule regime, a pegged exchange rate regime implies a long-run price level which is independent of the initial price level. Under such conditions the AD curve shifts over time, since its location now turns out to depend on a lagged variable (output or the real exchange rate). In SWJ's open-economy models, in fact, the source of dynamics is sharply different from that in their closed-economy models. In the latter, as noted, the »static« inflation expectations assumption drives all the dynamics; whereas in the former, this source of dynamics is removed by using the assumption that expected inflation equals the (exogenous) foreign inflation rate, and instead the presence of the lagged real exchange rate in the AD curve drives all the dynamics. This contrast does not of course reflect a general difference between closed- and open-economy models. However, since the reason for it is not discussed by SWJ, there is a risk that students might obtain the contrary impression.

The quality of exposition throughout »Introducing Advanced Macroeconomics« is uniformly high. The clarity of the verbal argument and of the English is excellent, and there is a pleasing lack of typos. Above all, there are no »fudges« in the reasoning when the argument becomes difficult, unlike in many undergraduate macro texts. Once the assumptions are in place, their implications are pursued rigorously and fully – one might almost say »relentlessly«. The book has full respect for the intelligence and maturity of the reader – there is no irritating »chattiness« or resort to hyperbole in an attempt to provide entertainment or artificial excitement. Mathematics is used freely, but the techniques needed are modest: algebra and differentiation, »yes«; expectations operators, »yes«; solution of a first-order linear difference equation, »yes«; but »forward-looking« solution of rational expectations models, »no«; and dynamic optimisation, »no«.

The overall flavour of SWJ's book is distinctly »European« rather than »American«. This is manifested, first, in the attention paid to market imperfections when dealing with the microeconomic foundations of aggregate supply. Firms are generally treated as monopolistic competitors faced by constant-elasticity demand curves, so that prices are a fixed markup over marginal costs. As regards wage setting, two alternative stories are given roughly equal prominence. They are first introduced in Book 1 (»The Long Run«), without expectations errors, before being taken up again in Book 2, with expectations errors. One is a model of efficiency wages, and the other is a model of optimising trade unions. By contrast the »classical« story of aggregate supply, based on a competitive and (by implication) symmetric-information model of the labour market, makes only the briefest of appearances, in Chapter 18. Second, the book contains numerous discussions of empirical evidence, several being so up-to-date that they are taken from yet-to-be-published sources such as working papers. While these are by no means confined to European examples, such examples do tend to dominate. For both these reasons, and also because, as previously noted, the analytical level of the book is intentionally pitched above the typical American »intermediate macroeconomics« level, it is a book which one would expect to be more attractive to European rather than American students and their teachers.

Finally, even in 800 pages, some things still have to be left out. One such topic is political economy questions, such as political business cycles. Chapter 22 deals with the time-consistency, inflation-bias problem in monetary policy, but it is not extended to cover political economy more broadly. Another topic not included is staggered prices or wages, of the Taylor or Calvo variety. It is necessary to teach »forward-looking« solutions of rational expectations models in order to treat this properly, so the decision not to introduce this more advanced technique presumably lies behind their exclusion. The idea of staggered

prices might nevertheless have deserved a mention, especially since the ballooning literature on the »New Keynesian Phillips Curve« is one which students are increasingly likely to stumble across. In terms of supporting material, the book is extremely thorough in its provision of exercises at the end of every chapter – sometimes up to 6 or 7 pages' worth. On the other hand, it lacks »suggestions for further reading« sections, and even lacks a list of references at the end, references being pro-

vided through footnotes. Maybe the authors felt they had been so exhaustive in their coverage that students would have no need or desire to read more; something which is quite possibly true!

Overall, then, this is a major contribution to the advanced undergraduate macroeconomics textbook market, one of impressive quality and thoroughness, and it is unlikely to have any serious rivals for a long time to come.





# Labor Supply Behavior and the Design of Tax and Transfer Policy

Henrik Jacobsen Kleven

University of Copenhagen, EPRU, and CEPR, E-mail: [henrik.kleven@econ.ku.dk](mailto:henrik.kleven@econ.ku.dk)

Claus Thustrup Kreiner

University of Copenhagen, EPRU, and CESifo, E-mail: [ctk@econ.ku.dk](mailto:ctk@econ.ku.dk)

*SUMMARY: This paper argues that recent empirical evidence on labor supply behavior – showing stronger participation responses than hours-of-work responses – has important implications for the design of tax and transfer policy. Based on a review of recent research in this field, we conclude the following: (i) Conventional ways of evaluating the impact of tax-transfer reforms on economic efficiency have to be revised. (ii) Optimal redistributive policies may well involve negative marginal tax rates at the bottom of the earnings distribution. (iii) The introduction of the Earned Income Tax Credit (EITC) in the United States in the mid-1970s and the large expansions of the credit during the past two decades did not involve a trade-off between efficiency and equality as suggested by previous estimates. Instead, the EITC has improved both equality and efficiency. (iv) For most European countries, redistribution to the working poor involves a substantially lower trade-off between efficiency and equality than traditional redistributive policies targeted to those out of work.*

---

## 1. Introduction

Redistribution to low-income individuals has grown significantly in European countries and in the United States since World War II. Today, most of these countries devote a sizeable amount of public spending to various low-income support programs. These programs include unemployment insurance, social assistance, in-work benefits, health insurance, food programs, housing programs, child care support, and disability benefits for the disabled.

These transfer programs have been motivated by a desire to alleviate poverty and to make the income distribution more equal. However, the presence of generous low-income support provided by the government may have adverse effects on the labor

---

This paper has been written on the occasion of Claus Thustrup Kreiner's inaugural lecture as full Professor at the University of Copenhagen on the 4th of February 2005. Parts of the paper draw heavily on recent, joint work with Nada Eissa, Herwig Immervoll and Emmanuel Saez.

market. In particular, since eligibility in most welfare programs is directly related to the size of income, individuals have an incentive to reduce or under-report earnings so as to qualify for welfare payments. For example, some people may want to reduce hours of work or effort on the job, or they may choose to opt out of the labor market entirely. There is convincing empirical evidence that welfare programs do, in fact, induce labor supply responses of this sort, Moffitt (2002). Moreover, taxes collected from middle- and high-income earners to finance low-income support leads to additional negative effects on labor supply and employment, Blundell and MaCurdy (1999).

The negative effects on labor supply implied by taxes and transfers create a loss of efficiency in the economy. In other words, policy makers face a trade-off between efficiency and equality, since more distributional equity generated by government programs implies lower efficiency. The size of this trade-off depends on the generosity and on the design of welfare programs. If the programs are inappropriately designed, they will be costly in terms of lost efficiency implying that the feasible amount of redistribution is lower. By contrast, if the programs are designed to minimize the adverse effects on labor supply, then we can redistribute more.

An important choice in the design of low-income support is whether it should apply to the poor or to the working poor. This policy choice has varied over time and across countries. In continental Europe and in Scandinavia, welfare programs tend to be aimed at guaranteeing a minimum level of consumption for all individuals, including those without a job (redistribution to the poor). By contrast, Anglo-Saxon countries tend to spend large amounts on in-work benefit programs targeted to the working poor. In the United States, for example, the so-called Earned Income Tax Credit (EITC) has become the single largest cash transfer program for lower-income families at the federal level. Eligibility for the EITC is conditional on having a low income and on having a job. The program attempts to achieve the dual goal of redistributing income to low-income earners and avoiding some of the adverse effects on the labor market mentioned above. In particular, by making the transfer conditional on employment, the EITC aims at increasing labor market participation for the eligible population.

A program similar to the EITC was introduced in the United Kingdom in the 1988 and expanded substantially in the late 1990's. Moreover, a number of continental European countries (including Denmark) have recently experimented with in-work benefit programs, although on a very small scale. In other words, the introduction and further expansion of in-work benefits targeted to the working poor is high on the agenda when discussing welfare reforms in Europe. Since the expansion of such programs – at the expense of traditional welfare programs – would represent a fundamental change in the approach to income redistribution, it is important to understand their implications for distribution and the labor market.

In this paper, we provide an overview of some of the central issues in the evaluation of welfare and tax reforms, and we discuss the appropriate design of redistributive policies. The paper pays special attention to the importance of modeling labor supply decisions correctly when considering policy reforms affecting the labor market. Until recently, a common feature of the public finance literature was an almost exclusive focus on hours of work for those who are working (the intensive margin of labor supply). However, an emerging consensus in the empirical labor market literature is that most of the observed variation in labor supply reflects changes in labor force participation (the extensive margin of labor supply), whereas changes in hours worked are much smaller, Heckman (1993); Blundell and MaCurdy (1999). Consistent with these findings, we outline a simple theoretical framework accounting for both the intensive and extensive margins of labor supply response, and we review a number of recent empirical applications showing that the composition of the response on the two margins is crucial for the evaluation of tax and transfer policy.

Considerable space will be devoted to discussing the desirability of policies targeted at the working poor versus traditional welfare policies aimed at the poorest (non-employed) segments of the population. Based in part on the experience in the US and the UK, we argue that there is a promising scope for in-work benefit reforms in continental Europe. In fact, in-work benefit reforms seem to be particularly desirable in a country such as Denmark where effective tax rates are very high at the bottom of the income distribution resulting from generous support to those out of work combined with high tax rates on the working population.

In Scandinavia, policy proposals of this sort are usually met by two critiques. First, it is pointed out that in-work benefit reforms tend to be less desirable in a country characterized by a highly compressed earnings distribution. We discuss this argument and find that it is not strong enough to overturn the case for in-work benefit reform in Denmark. Second, it is pointed out that most of the evidence on strong participation responses has been based on countries characterized by relatively low female labor force participation rates. While this argument is not without merit, we argue that working poor policies will be desirable in Denmark even at participation elasticities far below those estimated for the United States and the United Kingdom. This is because the design of the Danish welfare and tax systems creates exceptionally high effective tax rates at the bottom, implying that even very small elasticities involves substantial efficiency losses on the margin.

The paper is organized as follows. Section 2 provides a short review of the empirical labor supply literature and discusses the appropriate theoretical modeling of labor supply behavior. Section 3 provides a simple introduction to the measurement of the welfare costs of taxation. Section 4 shows theoretically and empirically why it is impor-

tant to distinguish explicitly between intensive and extensive labor supply responses when evaluating the impact of tax reforms. Section 5 discusses different issues involved in the optimal design of redistributive policy and provides empirical estimates of the trade-off between equality and efficiency for 15 European countries. Section 6 concludes.

## **2. Labor supply behavior at the intensive and extensive margins**

### *2.1 A brief review of the empirical literature*

A central finding in the empirical literature is that labor supply elasticities are low at the intensive margin, Heckman (1993); Blundell and MaCurdy (1999). It is notable that this finding is consistent across different methods and different samples, and that it holds for both males and females. The old findings of high elasticities for women (married women and single mothers) were based on censored specifications including non-participating individuals, thereby conflating extensive and intensive responses in the estimated elasticity. Once labor supply is estimated conditional on labor force participation, it turns out that the female hours-of-work elasticity is close to that of males, Mroz (1987); Triest (1990); Blundell et al. (1992).

Hence, a strong degree of labor supply responsiveness would have to come from the margin of entry and exit in the labor market. Indeed, there is an emerging consensus that extensive labor supply responses may be much stronger than intensive responses, Heckman (1993). In particular, participation elasticities seem to be very high for certain subgroups of the population, typically people in the lower end of the earnings distribution. Let us briefly review some evidence from both the United States and Europe.

One body of empirical work exploits evidence from recent tax reforms and expansions to tax-based transfers in the United States and the United Kingdom. In particular, a number of in-work benefit reforms targeted at lower-income families with children has provided an ideal opportunity to estimate labor supply behavior. For the United States, Eissa and Liebman (1996) use a quasi-experimental approach to show that the 1986 expansion of the Earned Income Tax Credit (EITC) had large effects on the labor force participation of single mothers. This was especially the case for single mothers with low education, where the Eissa-Liebman study implies an elasticity around 0.6. Meyer and Rosenbaum (2001) use a more structural approach and data from 1985 to 1997, a period of time where three large tax reforms were enacted in the US (in 1986, 1990, and 1993). Their finding, that the EITC accounts for about 60 percent of the increase in the employment of single mothers over the period implies a participation elasticity of about 0.7. A study by Blundell et al. (2000) considers the labor market impact of the recently implemented Working Families Tax Credit (WFTC) in the UK.

Like the EITC, this program was designed to induce low-income people with children, typically single mothers, from welfare into work. Their results indicate that the reform was quite effective in achieving this goal, increasing the participation rate of single women with children by 2.2 percentage points (5 percent).

While the literature on labor supply in Anglo-Saxon countries is extensive, there are fewer studies for continental European and Scandinavian countries. One might expect that labor supply elasticities are smaller in the more rigid labor markets of continental Europe and Scandinavia. However, several recent studies suggest that this is not the case. A number of structural studies of female labor supply are surveyed in Blundell and MaCurdy (1999). These studies found high elasticities – typically between 0.5 and 1 – across a number of European countries such as Germany, the Netherlands, France, Italy, Sweden, and the UK.<sup>1</sup> Since these results were based on samples containing non-employed individuals, the participation response is included in the estimated elasticities. Indeed, the results probably reflect an underlying labor supply curve where elasticities are large around the point of participation but fall off rapidly with increases in working hours, Blundell (1995).

The finding that tax incentives may have substantial effects on labor force participation is consistent with another stream of empirical literature estimating the effect of out-of-work benefits on unemployment. Krueger and Meyer (2002) surveys the evidence from a number of OECD countries. They conclude that benefits raise the incidence and the duration of unemployment, and that the elasticity of lost work time with respect to benefits tends to be around one. Since the risk of unemployment is largest among low-skilled workers, this evidence also indicates that strong participation responses tend to be concentrated at the bottom of the wage distribution.

### *2.2 Modeling entry-exit behavior in the labor market*

Having established that extensive labor supply responses are empirically important, we wish to explore their implications for the welfare evaluation of tax and transfer policy. In order to do so, we have to think about how to model participation behavior. As emphasized by Eissa et al. (2004), a realistic model must be consistent with empirical distributions of hours worked showing almost no workers at low annual or weekly hours of work. In other words, if individuals decide to work at all, they tend to work a substantial number of hours (say, 30 or 40 hours per week). To explain this discreteness of participation behavior, we need a framework featuring non-convexities in

---

1. We are not aware of studies estimating the elasticity of labor force participation with respect to taxes for Denmark. While the results for Sweden may give some indication of labor supply behavior in Denmark, it would be interesting and important to carry out econometric studies based on Danish data and reforms in order to discuss the design of future tax and welfare reforms in this country.

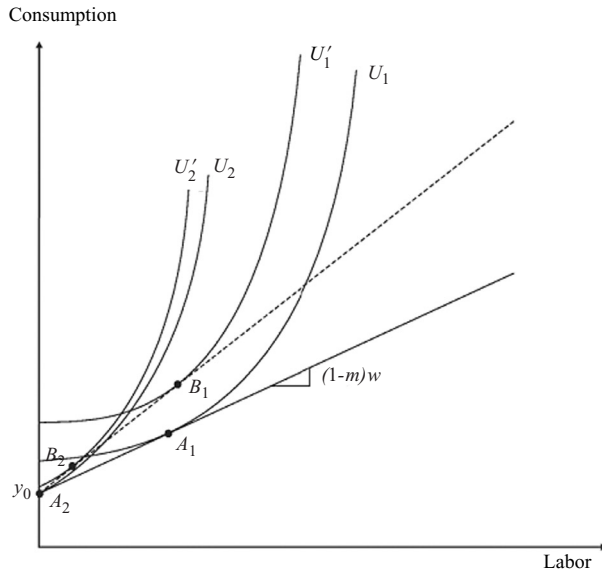


Figure 1. Intensive versus Extensive Responses in the Convex Model.

preferences and/or budget sets. The observed discrete responses along the extensive margin cannot be captured within a standard convex labor supply model.<sup>2</sup>

To see the point, Figure 1 illustrates the choice of labor supply in the standard convex model. In the figure, we consider two individuals facing the same budget constraint, where  $y_0$  is non-labor income,  $w$  is the wage rate, and  $(1 - m)w$  is the marginal net-of-tax wage. The indifference curves of the two individuals are drawn such that individual 1 has a relatively low valuation of leisure, while individual 2 has a relatively high valuation of leisure. Consequently, it is optimal for the first individual to work many hours (at point  $A_1$ ), whereas the second individual chooses to stay outside the labor market (at point  $A_2$ ) since there is no point of tangency at positive hours. As we have drawn the figure, the point of tangency for individual 2 is exactly at zero hours worked.

Consider now a small reduction in the tax rate increasing the slope of the budget line a little bit. In principle, we want to consider a marginal (infinitesimal) tax change, but we have exaggerated the size of the change in the figure for the point of illustration. The increased slope of the budget line induces individual 1 to increase his hours of work a little bit from point  $A_1$  to point  $B_1$  (an intensive response). For individual 2, on the other hand, the higher net wage gives rise to an interior solution so that he decides to enter the labor market at point  $B_2$  (an extensive response). Notice that this extensive

2. See Eissa, Kleven and Kreiner (2005) for a more detailed discussion of the modeling of labor supply.

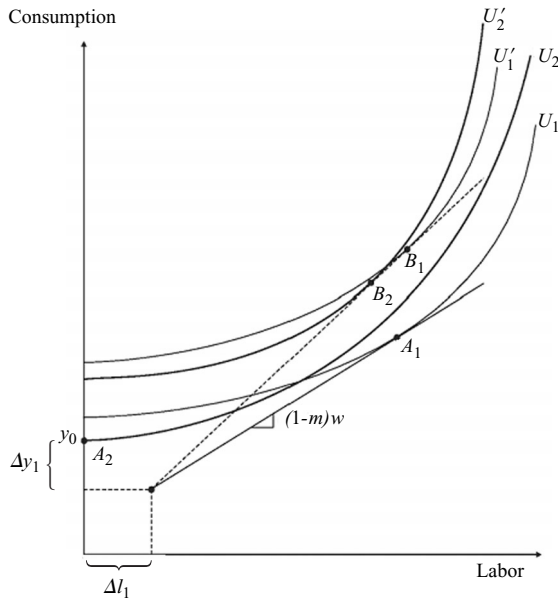


Figure 2. Intensive versus Extensive Responses with Fixed Costs of Working.

response involves a change in labor supply from zero hours to some small (infinitesimal) number of hours. Hence, the type of participation response predicted by this framework is a marginal one, just like the change in hours of work for those who are working. This conflicts with the empirical evidence described above showing that people do not enter the labor market at infinitesimal hours of work but that they do so at, say, 30 or 40 hours. Therefore, to be able to capture participation behavior in a realistic way, we need to adopt a framework incorporating some type of non-convexity making very low hours of work non-optimal.

In the empirical literature, non-convexities are typically introduced by way of fixed work costs, e.g. Cogan (1981); Blundell et al. (1987). In Cogan's (1981) analysis, the fixed costs may be monetary costs (say child care and transportation expenses) or they may take the form of a loss of time (e.g., commuting time). In Figure 2, we extend the analysis of labor supply choice along these lines. An individual who chooses to stay outside the labor market receives non-labor income  $y_0$ . If he decides to enter the labor market, he loses  $\Delta y_f$  in income and  $\Delta l_f$  in leisure time upon entry, thereby creating a discontinuity in the budget set. In the initial situation, it is still the case that individual 1 works relatively long hours (at point  $A_1$ ), while individual 2 does not work at all (at point  $A_2$ ). Now, if we reduce the tax rate a little bit, individual 1 responds again with a marginal change in working hours from  $A_1$  to  $B_1$ . By contrast, individual 2 now

reacts by making a discrete jump from not working at all to working nearly as many hours as individual 1 (at point  $B_2$ ). This discreteness of the participation choice is consistent with empirical distributions of hours worked. Hence, the incorporation of fixed work costs allows for a more realistic model of participation behavior.<sup>3</sup>

To summarize, we note that tax reforms entail intensive as well as extensive labor supply responses, and to account for the observed discreteness of extensive responses, they have to be modeled by introducing non-convexities into the standard framework.

### 3. The excess burden of taxation

In general, the tax system affects labor supply along both the intensive and extensive margins. These distortionary effects on labor supply affect government revenue and give rise to a loss of economic efficiency. This is what we call the excess burden or the deadweight loss of taxation. The aim of this section is to explain why it is important to distinguish explicitly between the intensive margin and the extensive margin when trying to measure the excess burden of taxation. Notice that it is not obvious a priori that we need to distinguish between the two margins in order to quantify the efficiency cost of the tax system. Taxes distort labor supply along both dimensions, but maybe this could be captured simply by looking at the aggregate labor supply curve thereby incorporating both the intensive margin effect and the extensive margin effect in a single relationship. Although this reasoning may sound plausible, it turns out to be wrong, as we will show below.

#### 3.1 *The excess burden of taxation with intensive labor supply responses*

We start with a review of the standard deadweight loss of taxation, which focuses only on the intensive labor supply margin. Consider an individual with an hourly productivity/wage of  $w$ . For simplicity, assume that the individual faces a simple proportional tax scheme where  $m$  denotes the marginal tax rate. The individual is willing to work more hours at a higher after-tax wage rate thereby giving rise to the upward-sloping labor supply curve illustrated in Figure 3a. With no taxes, the labor market would be at point  $A$ . The impact of the tax system is to reduce the hourly take-home wage from  $w$  to  $(1 - m)w$ , thereby giving rise to an equilibrium at point  $B$ .

At point  $B$ , the individual obtains a consumer surplus given by the area  $CS$ , the government revenue is given by the area labeled  $GR$ , while the triangle  $DWL$  captures the deadweight loss from the tax. The deadweight loss measures the amount that is

---

3. In addition to fixed work costs, the presence of low-income support programs featuring gradual phase-out and possibly discrete earnings and work tests create non-convexities making low hours of work very unattractive. In fact, Figure 2 can easily be reinterpreted to illustrate the case of non-convex tax-transfer programs by thinking of the cost  $\Delta y_f$  as an out-of-work welfare benefit, which is lost upon entry.



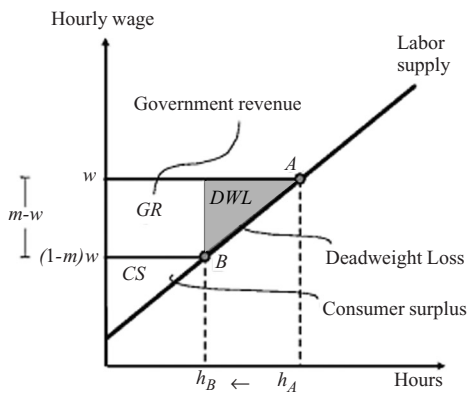


Figure 3a. Welfare cost of taxation with intensive labor supply responses.

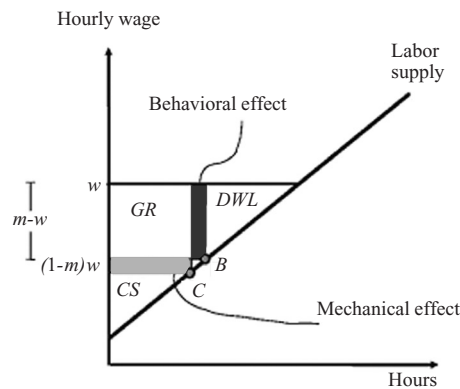


Figure 3b. Marginal welfare cost of taxation with intensive labor supply responses.

lost in excess of what the governments collects. If the same tax revenue were to be collected by a lump sum tax – a tax which is independent of earnings – the individual would continue to supply labor at point *A*, and obtain a consumer surplus equal to *CS* + *DWL*.<sup>4</sup> Hence, the difference between the income tax and the lump sum tax is exactly equal to the area *DWL*. This area therefore measures the efficiency cost of having a tax system that depends on income.

Formally, the deadweight loss may be derived by first calculating the area *GR* + *DWL* using integration and afterwards subtracting the rectangle *GR*. This gives

$$DWL = \int_{(1-m)w}^w h(x) dx - m \cdot w \cdot h((1 - m) w), \tag{1}$$

where *h*(·) is the labor supply function of the individual, while *h*((1 - *m*) *w*) denotes hours worked at point *B* in Figure 3a.

Equation (1) reflects the *total* deadweight loss of the income tax system, but this is typically not very interesting for tax policy. The reason is that practical tax reforms tend to involve relatively small changes within the existing tax schedule rather than abolishing income taxation altogether. For tax policy analysis, it is therefore more interesting to look at the efficiency impact of small adjustments in the existing system, i.e. the *marginal* deadweight loss.

4. The tax system influences hours worked through a substitution effect and an income effect. The income effect is, however, irrelevant for the measurement of excess burden. Notice that the excess burden is derived by comparing a given tax system with a lump sum tax system that generates the same tax revenue. Accordingly, the net-income of the individual does not change when going from one system to the other.

Figure 3b illustrates the impact of a small increase in the marginal tax rate  $m$ . Point  $B$  is the choice of the individual at the pre-reform tax system (equivalent to point  $B$  in Figure 3a). The reform reduces the marginal net-of-tax wage and moves the optimal choice from  $B$  to  $C$ . It can be seen in the figure that the increase in the deadweight loss triangle following the reform is given (approximately) by the black rectangle.<sup>5</sup> To understand the source of this efficiency loss, it is useful to relate it to the effects on government revenue. In general, tax reforms affect government revenue through two channels – the mechanical effect and the behavioral effect. The mechanical revenue effect reflects the impact of the increase in the tax rate at given behavior, and it corresponds to the grey rectangle in the figure. This effect reflects a transfer of income from the individual to the government (the consumer surplus is reduced by the same area), and it has no consequences for efficiency. In other words, although the mechanical effect represents a welfare loss to the taxpayer (a loss of income), it entails no loss of resources for the economy and hence no efficiency loss.

The behavioral revenue effect reflects the impact of changed behavior at the given tax rate. Because the increase in the marginal tax rate reduces hours worked from point  $B$  to  $C$ , taxable earnings go down leading to a negative feed-back effect on government revenue. This behavioral revenue effect from the tax reform – given by the black area in the figure – is exactly identical to the marginal deadweight loss of taxation. The insight that the marginal deadweight burden is equivalent to the behavioral revenue effect is not specific to the model adopted here. It is an insight that follows from any model where individuals optimize and markets are efficient, a point that we come back to below.

From the graph, or by differentiation of expression (1), we see that the change in the deadweight loss becomes<sup>6</sup>

$$\Delta DWL = m \cdot w \cdot h'((1-m)w) \cdot w \cdot \Delta m, \quad (2)$$

where  $h'(\cdot)$  denotes the derivative of the individual labor supply with respect to the after-tax wage rate. A small increase in the tax rate,  $\Delta m$ , induces the individual to reduce the number of hours by  $h'(\cdot) \cdot w \cdot \Delta m$ . Since a one-hour reduction in labor supply reduces the tax payment by  $m \cdot w$ , it follows that the change in the deadweight loss is exactly equal to the behavioral feedback effect on government revenue. This insight,

5. This is only approximately true, since the black rectangle does not include the small white triangle immediately below it. However, this triangle is a second-order effect arising because the reform in the figure is not exactly marginal (infinitesimal). For a marginal reform, this effect is zero.

6. Notice that the marginal tax rate also enters in the lower limit of the integral in equation (1). Using Leibnitz's rule, the derivative of this term with respect to  $m$  becomes  $w \cdot h((1-m)w)$ .

that the welfare loss is given by the revenue loss created by changed behavior, is not related to the specific model adopted here but holds in general for tax policy analysis.

For empirical applications, it is useful to rewrite the above result. Let  $\varepsilon$  denote the elasticity of hours worked with respect to the net-of-tax wage rate.<sup>7</sup> Then the change in the deadweight loss in proportion to the aggregate wage income may be written as

$$\frac{\Delta DWL}{wh} = \frac{m}{1-m} \cdot \varepsilon \cdot \Delta m \quad (3)$$

This expression is a classic Harberger-type formula for the marginal deadweight burden of taxation.<sup>8</sup> It shows that the marginal welfare cost depends on the initial level of the marginal tax rate, the increase in the marginal tax rate, and the hours-of-work elasticity. Notice that the marginal loss is equal to zero when the marginal tax rate is zero. That is, initially the welfare cost of raising tax revenue is zero. As the tax rate is increased, the marginal welfare cost of taxation also increases at a given labor supply elasticity. In other words, it becomes more and more costly to raise additional government revenue as the tax rate goes up. In fact, welfare effects can be substantial even for very small labor supply elasticities provided that initial marginal tax rates are high. Finally, since the empirical evidence does not give any indication of large, significant differences in hours-of-work responses across individuals (conditional on labor force participation), we may conclude that the largest efficiency gains of tax rate reductions are to be found where marginal tax rates are relatively high.

### 3.2 *The excess burden of taxation with extensive labor supply responses*

The modeling of tax distortions along the extensive margin is a bit more complicated. As noted in Section 2, an appropriate theory requires some type of non-convexity in order to account for the observed discreteness in labor supply behavior.<sup>9</sup> In addition, the theory has to incorporate heterogeneity across individuals. With no heterogeneity in preferences or productivity, either everybody participates or nobody participates in the labor market. In reality, some fraction of those in their working age

7. To be precise,  $\varepsilon$  is the compensated hours-of-work elasticity. As described in footnote 3, income effects are irrelevant for the measurement of the excess burden of taxation.

8. The result is more general than it appears. Specifically, the result is not restricted to a proportional tax scheme although this is the case for Figure 3a, b and for expression (1). The relationship in (3) holds for any tax system as long as the marginal tax rate is locally constant. The result also holds for other definitions of the deadweight loss. Here, we have derived expression (3) from a simple Marshallian definition but the result may also be derived from deadweight loss measures based on compensating variation, equivalent variation, or compensating surplus.

9. A general framework for analyzing welfare effects with discrete choice is provided by Small and Rosen (1981).

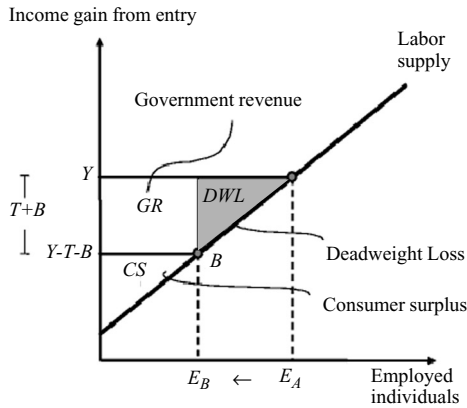


Figure 4a: Welfare cost of taxation with extensive labor supply responses.

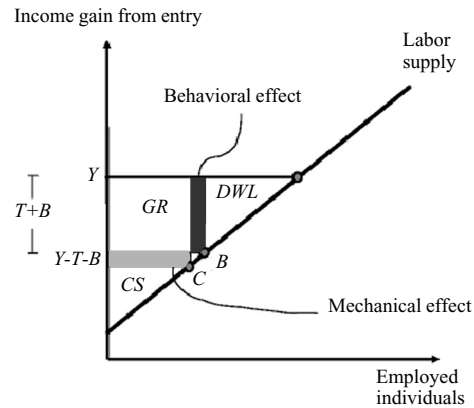


Figure 4b: Marginal welfare cost of taxation with extensive labor supply responses.

participate, and small changes in labor market incentives will typically create small adjustments in the number of employed workers. The combination of non-convexities and heterogeneity enables us to explain discrete entry-exit behavior at the individual level together with smooth changes in employment at the macro level. Here we present a very simple framework, which illuminates the main insights. A more general analysis is presented in Eissa et al. (2004).

Consider a group of individuals who all have the same hourly productivity/wage  $w$ . For simplification, we disregard intensive labor supply responses and assume that all individuals work some fixed number of hours,  $\bar{h}$ , if they work at all. Those who are working receive the same pre-tax earnings denoted by  $Y \equiv w\bar{h}$ . The individuals face fixed work costs, which may reflect monetary costs, time costs or simply a fixed disutility of working. These work costs are assumed to vary (smoothly) across individuals, thereby giving rise to the upward sloping labor supply curve in Figure 4a. This curve displays the number of employed individuals as a function of the net income gain of working (i.e., earnings net of taxes paid and transfers lost). At a low net income gain, many individuals stay outside the labor market because their work costs outweigh the gain from entry. As the net income gain increases, more and more individuals find it worthwhile to enter the labor market so as to generate a positively sloped extensive labor supply curve.

With no tax-transfer system, the equilibrium would be at point  $A$ . Assume now that an employed who is earning  $Y$  has to pay an amount  $T$  in taxes (net of any benefits), while a non-employed individual receives the amount  $B$  in welfare benefits (net of any taxes). This implies that the net income gain of working equals  $Y - T - B$ , thereby

moving the equilibrium number of employed from point  $A$  to point  $B$ . Those who work at point  $B$  obtain consumer surplus  $CS$  and generate government revenue (taxes paid plus benefits saved) equal to  $GR$ . The tax-transfer system drives a wedge,  $T + B$ , in between the social return to work and the private return, thereby creating a deadweight loss equal to the area  $DWL$ .

Formally, the deadweight loss may be written as

$$DWL = \int_{Y-T-B}^Y E(x) dx - (T+B) \cdot E(Y-T-B), \quad (4)$$

where  $E(\cdot)$  denotes the number of employed people as function of the net income gain of working. The first term measures the size of the area  $GR + DWL$  in Figure 4a, while the second term equals the area  $GR$ .

Let us now consider the impact of a reform making a small adjustment to the existing tax system. Figure 4b displays the consequences of a small increase in the tax burden on the employed, which moves the equilibrium from point  $B$  to point  $C$ . As in the previous model, it is useful to distinguish between the mechanical effect and the behavioral effect on government revenue. The mechanical effect is the increase in government revenue at given labor market participation, which is shown by the grey area in Figure 4b. This effect corresponds to a redistribution of income from the employed to the government and involves no deadweight loss. The distortionary effect of the tax change comes entirely from the behavioral effect on government revenue. The tax change reduces the number of employed individuals and this reduces the net-payment to the government by  $T + B$  per person. Accordingly, the change in the deadweight loss becomes

$$\Delta DWL = (T+B) \cdot E'(Y-T-B) \cdot \Delta T, \quad (5)$$

where  $E'(\cdot)$  denotes the derivative of the employment function. This result may also be derived from differentiation of expression (4).

The sensitivity of employment with respect to changes in the net income gain from entry may be measured by the extensive labor supply elasticity  $\eta \equiv \frac{E'(\cdot)}{E} \cdot (Y-T-B)$ . In addition, we define an average tax rate  $a \equiv T/Y$  and a benefit rate  $b \equiv B/Y$ . With these definitions, expression (5) may be rewritten to

$$\frac{\Delta DWL}{Y \cdot E} = \frac{a+b}{1-a-b} \cdot \eta \cdot \Delta a, \quad (6)$$

where  $\Delta a = \Delta T/Y$  denotes the reform-induced change in the average tax rate. This deadweight loss formula reflects the same basic form as the traditional one in (3), but it is related to different policy parameters and a different elasticity. In particular, the welfare cost is no longer related to the marginal tax rate. It is instead related to the tax rate applying at the extensive margin – the *participation* tax rate – calculated as the sum of the average tax rate and the out-of-work benefit rate,  $a + b$ . The participation tax rate measures the fraction of earnings that is lost in taxes and forgone benefits upon labor market entry. Finally, the welfare effect depends on the sensitivity of entry-exit behavior as measured by the elasticity of labor force participation with respect to the net-of-tax income gain from entry,  $\eta$ .

In general, a tax reform affects labor supply along both the intensive margin and the extensive margin. The strength of the responses are determined by the sensitivity of each margin with respect to economic incentives, captured by the two elasticities  $\varepsilon$  and  $\eta$ . The intensive and extensive adjustments create a change in aggregate labor supply, defined as the total number of hours worked of all employed people ( $E \cdot h$ ). It is therefore natural to ask if it is possible to construct a measure of the deadweight burden which relies only on the sensitivity of aggregate labor supply with respect to economic incentives. The above analysis demonstrates that, in general, it is not feasible to construct such a measure, since labor force participation is related to a different tax wedge than are working hours.

Having said that, it should be noted that there is one special case where tax rates on the two labor supply margins are identical. This is the case, where the entire tax-transfer system is characterized by a *linear* Negative Income Tax (NIT). This type of system provides a guaranteed minimum income  $B$  to all individuals in the economy (participants and non-participants) financed by a constant marginal tax rate  $m$  on earnings. In this special case, the participation tax rate  $a + b$  is exactly identical to the marginal tax rate  $m$ .<sup>10</sup> We may then sum the right-hand sides of (3) and (6) so as to obtain the total welfare effect in proportion to aggregate labor income as

$$\frac{\Delta DWL}{Y \cdot E} = \frac{m}{1 - m} \cdot (\varepsilon + \eta) \cdot \Delta m, \quad (7)$$

where  $\varepsilon + \eta$  is the aggregate labor supply elasticity. Note, that the only difference between this formula and the pure intensive welfare effect in formula (3) is the presence of the aggregate labor supply elasticity instead of the hours-of-work elasticity.

An approach based on (7) would be simple and has been used in some empirical

---

10. To see this, notice that the net tax payment of each individual under a linear NIT is given by  $T = -B + mY$ . By dividing this relationship with earnings  $Y$ , we obtain  $a + b = m$ .

work, but it is unlikely to yield accurate results. It requires that the entire welfare system is a linear NIT, which is far from being true in empirical applications. Public benefits tend to be non-universal, targeted to low and middle income classes through earnings or work tests (creating discontinuities) or through gradual phase-outs. Consider as examples low-income support, in-work benefits, housing- and education subsidies, child benefits, medical aid, food stamps, and public pensions. While some benefits may be universal in some countries, never are they collectively so. Moreover, income tax systems are generally characterized by increasing marginal tax rates further conflicting with the NIT assumption.<sup>11</sup> In Section 4, we demonstrate the potential large errors from not accounting properly from the welfare effects along the extensive margin.

### *3.3 Excess burden, government revenue, and other dimensions of behavioral response*

In the previous sections, we set up two simple models to derive formal expressions for the marginal excess burden of taxation with intensive and extensive responses, respectively. We emphasized the equivalence of the marginal excess burden with the impact of behavioral responses on government revenue. The insight, that the efficiency effect of a policy reform is given simply by the behavioral revenue effect, is not specific to the simple models adopted above. It is an insight, that follows from any model where individuals optimize and markets are efficient, i.e., except for the inefficiencies introduced by the existence of distortionary taxation.

To see that the result holds in general, consider a small (marginal) tax reform changing government revenue by  $\Delta GR$ . In general, this change in government revenue reflects a mechanical revenue effect,  $\Delta M$ , and a behavioral revenue effect,  $\Delta B$ , such that  $\Delta GR = \Delta M + \Delta B$ . In any model where agents optimize and markets are efficient, the total utility loss (in monetary units) on individuals will be exactly equal to the mechanical tax increase  $\Delta M$ . To see this, start by noting that – in the absence of behavioral responses – a higher tax burden involves a transfer of income from individuals to the government given by  $\Delta M$ , creating a utility loss for the individual equal to  $\Delta M$ . But of course behavioral responses are not absent: the individual has an incentive to re-optimize behavior so as to avoid some of the tax increase. However, as long as the reform is small, this behavioral adjustment entails no first-order effect on utility, since individuals were initially in their optimum. This is an application of the envelope theorem.

The marginal deadweight loss from the reform,  $\Delta DWL$ , is defined as the marginal utility loss to individuals in excess of the additional revenue collected. By exploiting

---

11. A description of the actual tax-transfer systems prevailing in a number of European countries is provided by Immervoll et al. (2005), while the US tax-transfer system is described in Eissa et al. (2004).

the above relationships, we immediately obtain  $\Delta DWL = \Delta M - \Delta R = -\Delta B$ , showing that the marginal deadweight loss is exactly identical to the behavioral revenue loss. Notice that, under the assumptions of rational individuals and efficient markets, the result relies simply on an application of the envelope theorem and the definition of the deadweight loss. In particular, it does not rely on the form of preferences and technology, nor does it depend on the framework being one of continuous choice. For example, the result encompasses the case of discrete participation behavior due to non-convexities in budget sets or preferences.

There are several points to make in relation to this result. First, behavioral responses are central to the creation of efficiency losses, even though these responses entail no first-order utility effects on individuals due to the envelope theorem. The point is that behavioral responses create efficiency losses indirectly in the form of an externality operating through the government budget. For example, although a marginal labor supply reduction involves no first-order effect on the optimizing worker (envelope theorem), it implies a loss of government revenue and therefore less money for the financing of, say, public goods, creating a negative externality on everybody else.

Second, the derivation did not specify the type of behavioral response we are considering. Following most of the literature, this paper has focused on labor supply behavior in its standard dimensions: hours worked and labor force participation. But there are many other dimensions of labor supply that may be responsive to taxation. For example, higher taxes may lead to lower effort on the job, and they can have adverse effects on the incentives to improve skills by taking education or engaging in on-the-job training. Taxation can also affect the type of jobs that people take (in terms of profession, geographical location, etc.) and it may affect search and labor mobility. All of these different types of labor supply responses can affect taxable earnings and government revenue, thereby creating efficiency losses in the economy. Although there has been a tremendous focus in the empirical literature on hours worked, the other dimensions of labor supply response are just as important from the point of view of economic efficiency.

Third, the relationship between excess burden and behavioral revenue effects implies that we should go beyond thinking just about labor supply in its various dimensions. Instead, we should be thinking about changes in taxable income more generally. Taxable earnings may respond to taxation through several other margins than labor supply. For example, individuals may be able to change the form of payment for labor services into something, that is more leniently taxed. This includes fringe benefits such as corporate cars, in-house sports facilities, free lunch and dining, laptop computers, etc. It also includes turning labor income into capital income (typically carrying a lower rate of tax) by being paid in stock options instead of a traditional wage income. Alterna-



tively, taxable income can respond if higher taxes lead to a more aggressive interpretation of tax rules (e.g., claiming questionable deductions) or tax evasion (understating income, claiming unjustified deductions). Tax-induced changes in avoidance or evasion affect government revenue, hence creating efficiency losses in the same way that labor supply responses do.

This type of reasoning underlies the recent literature on the elasticity of taxable income, starting with the important paper by Feldstein (1995). While taxation may affect behavior in a myriad of ways – as suggested by the above comments – the responses matter for efficiency only to the extent that they change taxable income. While it may be extremely difficult to estimate each of the responses separately, it is possible to estimate the elasticity of taxable income based data from individual tax returns. A focus on this parameter might make intensive responses more important than portrayed in our review of the empirical labor supply literature.

Indeed, for the United States, the literature has shown that the (intensive) elasticity of taxable income can be very high at the top of the income distribution (see Saez, 2004 for a recent survey). However, this is a phenomenon occurring only at the extreme top of the income distribution (top 1%). We are not aware of studies estimating the elasticity of taxable income for European countries.

Finally, based on the discussion in this section, one might be tempted to conclude that we do not need to distinguish between different margins of behavioral response. To estimate the revenue (and hence efficiency) effects, it seems that all we need to know is the elasticity of taxable income and then apply a marginal tax wedge. However, this is not entirely correct. An obvious point is that the tax system may involve several different bases subject to selective tax rates. In this case, different margins of response can relate to different bases and tax rates, and we then have to distinguish between different taxable income elasticities to get the revenue and efficiency effects right. A more subtle point is that, under non-linear tax systems, marginal and discrete changes in taxable income are taxed differently. As shown in the previous sections on intensive and extensive labor supply effects, it is important to distinguish between marginal and discrete effects in the measurement of excess burden. Hence, the discussion in this section is entirely consistent with the previous conclusions.

#### **4. The evaluation of tax and transfer reforms**

Following Harberger (1964), a large numerical literature has attempted to evaluate the distortions to the labor-leisure choice induced by labor income taxation and the impact of different tax reforms. The evaluation methods include simple deadweight loss calculations, Browning (1987, 1995), Computable General Equilibrium models, Ballard et al., (1985); Ballard (1988), and micro-simulations, Browning and Johnson,

(1984); Triest (1994). A common feature of this literature is the assumption of a standard convex labor supply model, ruling out discrete participation responses. Although the models applied are typically more sophisticated than the theory presented above, the basic forces driving the results are the same as those underlying formula (3). Extensive labor supply responses are either ignored, or included implicitly by substituting the elasticity of aggregate labor supply for the hours-of-work elasticity.<sup>12</sup> Ignoring extensive responses, corresponds to completely disregarding the welfare effects in (6). The second approach corresponds to applying a formula like (7), thereby approximating the participation tax rate ( $\Delta a + b$ ) and the change in this rate ( $\Delta a$ ) with the level and change, respectively, in the marginal tax rate (i.e.  $m$  and  $\Delta m$ ). In this section, we demonstrate that these approaches, based on the traditional labor supply model, may be seriously misleading. As a case in point, we examine the welfare effects for single women with dependent children in the United States following four tax acts passed in 1986, 1990, 1993 and 2001.

The welfare effects of these reforms on single mothers are particularly interesting to study. The group of single mothers experienced large tax cuts from the reforms due to expansions of the Earned Income Tax Credit (EITC), and because of increases in the standard deduction, personal exemptions, and a more favorable tax rate schedule. The combination of all these elements implied substantial improvements in the incentive to supply labor along both the intensive and extensive margin. The labor supply responses to these changes in incentives have been studied extensively and, not surprisingly, all the evidence points to strong participation responses (cf. the discussion in Section 2). On the other hand, hours-of-work responses (conditional on participation) were very small.

#### *4.1 Effects on labor supply and efficiency of an Earned Income Tax Credit*

Let us first discuss theoretically how an EITC influences labor supply and economic efficiency. Figure 5 illustrates the impact on labor supply of introducing an EITC. The straight line represents the relationship between disposable income and earnings before the introduction of the EITC (assuming a linear tax system for simplicity). We assume that the individuals face a fixed cost of working implying that each person will have to work a certain number of hours before consumption is raised above the non-working level. The impact of the EITC depends on the earnings level of the individual.

---

12. Indeed, it is not uncommon that simulation studies based on the standard convex model employ total elasticities in their calibration. For example, this seems to be the case in Browning and Johnson (1984), Ballard et al. (1985), Ballard (1988), Browning (1995), and Bourguignon and Spadaro (2002a, 2002b). In all these studies, high female labor supply elasticities (around 0.5 – 1.0) are used in the calibration, although elasticity estimates of this magnitude tend to be based on censored specifications including observations with zero hours of work, Mroz (1987); Triest (1990). By implication, the above studies are lumping together extensive and intensive responses in the simulations.

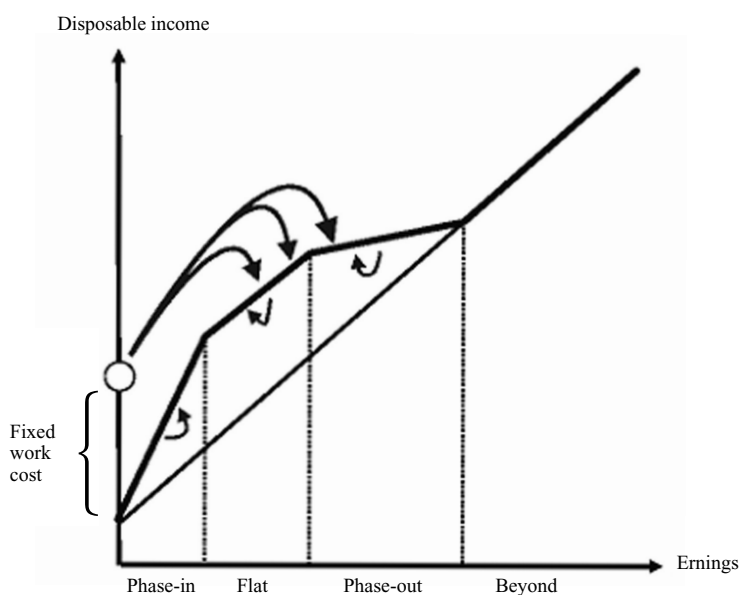


Figure 5. Impact of an Earned Income Tax Credit.

The credit is first phased in at low earnings, then stays constant, and is finally phased out at higher earnings.<sup>13</sup> The consequence for the relationship between disposable income and earnings is shown in Figure 5. The total tax burden on labor income is reduced in all three intervals of the EITC. The impact on the marginal tax rate depends on the initial earnings level of the person. In the phase-in interval, the marginal tax rate is reduced, in the plateau range it is unchanged, while in the phase-out range it is increased.

The labor supply responses along the intensive margin, illustrated by the small arrows in Figure 5, depends on substitution and income effects. In the phase-out range both of these effects reduce labor supply. In the flat range only the income effect is present which leads to a reduction in labor supply. Finally, in the phase-in range the substitution effect and income effect have opposite signs, so that hours increase as long as the substitution effect is stronger than the income effect. This is equivalent to the so-called uncompensated hours-of-work elasticity being positive in this range.

The sign of the intensive margin welfare effect depends on the direction of the

13. Notice that the EITC recently introduced in Denmark is different. The Danish EITC is given to everyone participating in the labor market, while the EITC in the United States is phased-out at high earnings. In addition, the size of the EITC in the US is related to the number of dependent children and, if married, on the income of the spouse.

change in the marginal tax rate (cf. (3)).<sup>14</sup> An individual in the phase-in region experiences a lower marginal tax rate ( $\Delta m < 0$ ), yielding a positive welfare effect. On the other hand, the welfare effect is zero for individuals in the plateau region (where  $\Delta m = 0$ ), while the welfare effect is negative for individuals in the phase-out region (where  $\Delta m > 0$ ). The total effect on economic efficiency from intensive responses may go either way but, in practice, it is quite likely that the negative effects will dominate. The reason is that the welfare effects have to be weighted by the aggregate labor income in the respective income segments (observe that the deadweight loss in (3) is measured in proportion to the earnings of the individual). Since earnings are higher in the phase-out range this gives, ceteris paribus, a reason to expect an overall welfare loss.

The welfare effects created along the extensive margin are unambiguously positive in all income ranges of the EITC. In all ranges, the total tax burden decreases (i.e.  $\Delta a < 0$  in all intervals), which yields a positive behavioral response on government revenue (cf. (6)).

The total welfare effect of introducing the EITC depends in the end on the strength of the different effects. However, if the tax wedge on participation is non-negligible, it is reasonable to expect a positive total effect since participation elasticities are very large compared to hours-of-work elasticities for the group of single mothers.

The potential errors of basing the evaluation on the standard, convex framework is now easy to see. A welfare analysis disregarding extensive responses is biased downwards, and the error may be substantial if participation elasticities are large. If we try to include extensive responses by substituting the aggregate labor supply elasticity for the hours-of-work elasticity in the traditional formula, the error may very well become even bigger. Assume, for example, that the intensive welfare effect is negative when calculated using the hours-of-work elasticity, while the total effect is positive after accounting for extensive effects in the proper way. If we now try to modify the standard approach by substituting a large aggregate labor supply elasticity for a small hours-of-work elasticity then the measured welfare effect becomes even more negative, thereby getting further away from the correct positive effect. This type of error seems to underlie the pessimistic results in Browning (1995) concerning the EITC. In the next subsection, we try to quantify these errors in an empirical application.

#### *4.2 Welfare effects on single mothers of four tax reforms in the United States*

In Eissa, Kleven and Kreiner (2004), we estimate the welfare effects on single mothers of a series of tax acts passed in the United States in 1986, 1990, 1993 and 2001. Our estimation is based on a micro-simulation approach using a general theore-

---

14. Notice that  $\varepsilon$  in (3) is the compensated hours-of-work elasticity. Thus, only the substitution effects matter for welfare.

tical framework, which incorporates intensive and extensive labor supply behavior as well as individual heterogeneity in productivities and preferences. The empirical implementation requires information about labor supply elasticities as well as various tax/transfer parameters and wage income shares. In order to get unbiased results, it is particularly important that we account for all elements in the tax and benefits system that drive a wedge in between the social return and the private return of supplying labor.

The labor supply elasticities are based on the empirical literature reviewed in Section 2. To generate the tax/transfer parameters, we estimate individual effective tax rates ( $m$  and  $a$ ) on a representative sample of single women with dependent children in the United States. The tax parameters are calculated using the Tax Simulation Model (TAXSIM) of the National Bureau of Economic Research (NBER). These tax rates include the federal, state and payroll tax components but do not include the transfer component.

In the United States, lower-income families are eligible to receive cash assistance from the Temporary Assistance to Needy Families (TANF) program, previously Aid to Families with Dependent Children (AFDC). In addition, families may be eligible for in-kind benefits in the form of food vouchers (Food Stamps) and health insurance (Medicaid). Since most single mothers are eligible for these types of programs, it is important to incorporate the benefit side in the calculation of effective tax rates. Accordingly, we develop a benefit calculator accounting for the differences in eligibility and benefit structures at the federal and state levels.

Figures 6a and b display tax rates along the intensive and extensive margins in the four pre-reform years. The graphs illustrate the dramatic changes in the taxation of single mothers over the period. Most notable is the decline in the effective tax rate on labor market participation going from 57 percent in 1986 to 32 percent in 2000. The marginal tax rate shows a more moderate and less systematic decline than the participation tax rate, which is not surprising given the large expansions of the EITC during this period. As discussed above, an EITC reduces unambiguously the tax burden, whereas its impact on the marginal tax rate depends on the distribution of individuals on the different income intervals of the EITC (phase-in, flat and phase-out). The two graphs also highlight the importance of accounting for the benefit system when calculating the relevant tax wedges along the two margins. Excluding benefits would seriously underestimate tax wedges. For example, the mean of the participation tax rates would be reduced by more than 35 percentage points if we did not account for the benefit system.

In Figures 6a and b, the changes in tax rates across the different pre-reform years reflect all changes in the tax and welfare system taking place at the federal and state

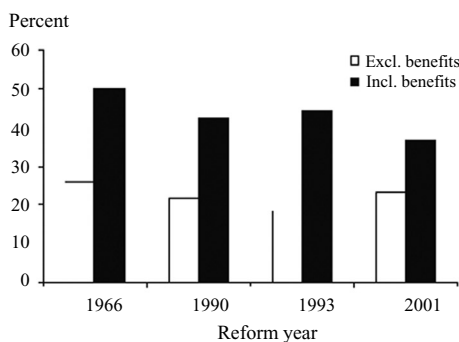


Figure 6a. Marginal tax rate.

Source: Elssa et al. (2004).

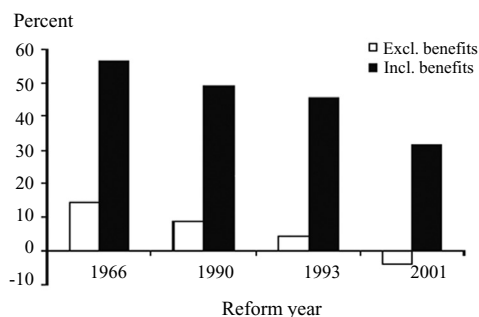


Figure 6b. Participation tax rate.

Source: Elssa et al. (2004).

levels over the period, and they also incorporate any behavioral responses to these tax and benefit changes along with macro/time effects on income and demographic variables. To isolate the direct impact of federal tax reforms, we calculate the changes in the tax rates that are attributed to the reforms alone. The results are illustrated in Figure 7, which confirms that the decrease in effective tax rates over the 15-year period has been driven to a large extent, by tax changes at the federal level. This was particularly the case for the 1986 and 1993 reforms, which reduced the participation tax rate by 8 and 13 percentage points, respectively.

There is a substantial variation in wage income and tax rates across the single mothers in the data. This heterogeneity highlights the need for using micro-simulations to evaluate the tax reforms. Large errors may occur in more aggregate studies because of the correlation between earnings, tax rates and tax changes. In Eissa, Kleven and Kreiner (2004), we carry out a number of simulations assuming different elasticity scenarios. The simulations point to substantial welfare gains for all the reforms. In particular, this is the case for the tax reform act of 1986. In our baseline scenario with a participation elasticity equal to 0.4 and a (compensated) hours-of-work elasticity equal to 0.1, we obtain a welfare gain for this reform equal to 7.3 percent of labor income.<sup>15</sup> The welfare effects are also high for the 1993 reform (2.3 percent of labor income), which contained the single largest expansion of the EITC. For the 1990 and 2001 reforms, the welfare gains are not quite as large.

For all four reforms, most of the total welfare gain is generated on the extensive margin. While  $\frac{3}{4}$  of the gain from the 1986 tax act is created by labor market entry, essentially all of the positive effect from the 1990 reform is driven by labor force par-

For this reform, it is interesting to note that a total labor supply elasticity of 0.6 – certainly not out of bound of the empirical estimates – would imply Laffer curve effects. In this scenario, the large tax reductions granted to single mothers are recouped entirely from the labor supply responses created by the reform.

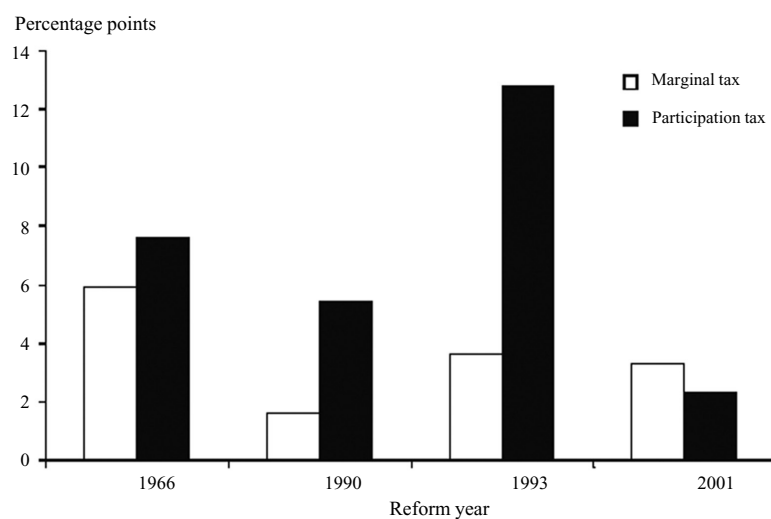


Figure 7. Reform-induced reductions in tax rates.

Source: Eissa et al. (2004).

ticipation. For this reform, the intensive welfare effect is around zero because negative effects in the phase-out region cancel out positive effects in the phase-in region of the EITC. For the 1993 reform, the large welfare gain is a result of the extensive margin strongly dominating welfare losses created on the intensive margin. Finally, for the recently enacted 2001 reform, the difference between the intensive and extensive welfare effects is less pronounced. This occurs for two reasons: First, the 2001 tax cuts reduced participation tax rates only slightly and, second, by the year 2000 the previous reforms had already eliminated much of the inefficiency along the extensive margin.<sup>16</sup>

Our finding that extensive responses drive almost all of the welfare effects created by the four reforms underpins the importance of accounting for this margin of labor supply. A simulation based on the traditional, convex labor supply model with only intensive responses would seriously underestimate the welfare effects. For the 1993 reform, even the sign would be wrong. Here, the intensive welfare effect in our base-

16. One might argue that our findings regarding the relative sizes of the extensive and intensive welfare effects were to be expected under the assumed elasticity scenario. Notice, however, that the difference between extensive and intensive welfare effects cannot be explained exclusively by elasticities, since the two kinds of welfare effects are related in different ways to the tax-transfer system. For example, for the tax act of 1990, increasing the hours-of-work elasticity would leave the intensive welfare effect more or less unchanged since the losses in the phase-out region would continue to cancel out the gains created in the phase-in region. For the 1993 reform, increasing the intensive elasticity would simply exacerbate the welfare loss along that margin, thereby reinforcing the point regarding the difference of welfare effects along the two margins.

line scenario equals -0.38, whereas the welfare effect equals 2.30 when including the extensive welfare effect. We may instead try to estimate the welfare effect from the traditional, convex model by substituting the aggregate labor supply elasticity for the hours-of-work elasticity, i.e., apply expression (7). This gives a welfare estimate of -1.88 which should be contrasted with the 2.30 obtained when accounting properly for the extensive welfare effect. These numbers demonstrate clearly the large errors that may arise from basing the evaluation of tax reforms on the traditional, convex labor supply model.

The errors reflect that the two margins of labor supply response depend on taxes and transfers in different ways. While the intensive margin depends on the effective marginal tax rate, the extensive margin is related to the average tax rate and the benefit rate. By implication, the size of the error depends crucially on the properties of the tax-transfer programs being analyzed.

### **5. The optimal design of redistributive policies**

The proper amount of redistribution and the design of transfer programs is an important and controversial issue in the political sphere. As is well known from the theory of optimal income taxation, redistribution gives rise to a trade-off between equity and efficiency. Redistribution from middle and high incomes to low incomes is desirable for equity reasons. On the other hand, redistributive programs tend to reduce labor supply incentives, thereby creating efficiency costs.

In this section, we first discuss the implications of the recent evidence on labor supply behavior for the theory of optimal income taxation. Secondly, we analyze empirically the tradeoff between equity and efficiency in redistribution policy across different EU countries. Thirdly, we discuss some implications for the tax treatment of married couples.

#### *5.1 The optimal income tax*

Following the seminal contribution by James A. Mirrlees (1971) most of the literature exploring the optimal income tax structure has applied the standard, convex labor supply model. One of the lessons from this literature is that marginal tax rates should be positive at all earnings levels.<sup>17</sup> By implication, a tax-transfer system involving an EITC creating negative tax rates at the bottom of the earnings distribution is never optimal. In a recent interesting contribution, Saez (2002) shows theoretically that this result breaks down if all (or most) of the variation in labor supply occurs along the extensive margin rather than on the intensive margin. His calibration of a general

---

17. Kleven and Kreiner (2004a) provides a short review of the standard results in the theory of optimal income taxation.



discrete model featuring both intensive and extensive responses indicates that an EITC is in fact optimal for the United States.

It is possible to understand the gist of these results based on simple heuristic arguments. Let us start by discussing the original Mirrlees result based on the convex labor supply model. The theory takes as its point of departure that taxes and transfers must be observed income – abilities are private information and unobservable for the tax authorities. It is further assumed that society puts a higher value on the marginal consumption of those with low abilities/incomes than on the marginal consumption of the well-off. The society therefore wishes to have a relatively high taxation at the top of the earnings distribution and low net taxes (maybe even negative) at the bottom of the distribution. The drawback of this redistribution policy is that it involves positive marginal tax rates, leading to lower hours worked and thus generating an efficiency loss (cf. Section 3.1). If the hours-of-work elasticity is large, the efficiency cost will be high which, *ceteris paribus*, calls for less redistribution and lower marginal tax rates. But the effect can never be so strong that it calls for negative tax rates in some range. To see why this is so, consider for a moment a negative tax rate in some small earnings range  $[Y^*, Y^* + \Delta Y]$ . The negative marginal tax rate implies that an individual with income  $Y^* + \Delta Y$  pays less in taxes than an individual with income  $Y^*$ . Suppose now that a revenue-neutral tax reform increases the marginal tax rate a little bit in the interval  $[Y^*, Y^* + \Delta Y]$  while leaving marginal tax rates unchanged everywhere else. This small marginal tax rate increase allows the government to raise more money from all individuals with income above  $Y^* + \Delta Y$ . If the proceeds are then distributed uniformly to all individuals in the population, the reform clearly improves equity. But what about efficiency? Recall from Section 3.1 that the effect on economic efficiency is determined by the change in the marginal tax rate and that the effect is given by the behavioral effect on government revenue. Only individuals in the interval  $[Y^*, Y^* + \Delta Y]$  experience a change in the marginal tax rate. These individuals face a higher marginal tax rate and respond by reducing labor supply. This would normally reduce efficiency. However, since the marginal tax rate is initially negative in this interval, the behavioral response actually generates a positive effect on government revenue and thereby an efficiency gain. To conclude, if marginal tax rates are negative then it is possible to increase the tax rate and thereby improve both equity and efficiency, thus contradicting that negative tax rates can be part of an optimal policy.

In the standard theory described above, labor supply responds only along the intensive margin: hours of work change a little bit when the marginal tax rate is changed a little bit. This stands in contrast to the labor supply evidence reviewed in Section 2.1. In particular, this is the case at the bottom of the income distribution where the variation in labor supply is dominated by entry/exit behavior. Saez (2002) is the first paper

to incorporate entry/exit decisions into the theory of optimal income taxation. Below we explain his main result, namely that negative marginal tax rates may be optimal at the bottom of the income distribution.

Consider for a moment the opposite case where the marginal tax rate is positive at the bottom of the earnings distribution. Individuals outside the labor market pay  $T(0)$  in net taxes (presumably a negative number due to out-of-work transfers), while working individuals with very low earnings pay  $T(Y^*) > T(0)$ . Suppose now that a reform increases the net tax payment of all individuals in society by a small fixed amount and gives the proceeds to individuals at  $Y^*$ . This reduces the tax payment for the low-skilled, while the tax payment for everybody else goes up. If the social welfare weight of the low-skilled individuals at  $Y^*$  is higher than the average welfare weight of all individuals in the population, this redistributive scheme increases equity. In addition, the lower tax payment at the earnings level  $Y^*$  compared to the zero-earnings situation induces some low-skilled unemployed people to enter the labor force at  $Y^*$ . Since  $T(Y^*) > T(0)$ , this behavioral response raises government revenue and thereby economic efficiency.<sup>18</sup> Therefore the suggested reform improves both equity and efficiency. As a consequence, it cannot be optimal to have a positive marginal tax rate at the bottom of the earnings distribution. Instead, the tax-welfare system should subsidize labor market entry for low-skilled individuals.

### *5.2 The trade-off between efficiency and equality in European countries*

The traditional theory of optimal income taxation emphasized a simple trade-off between the distribution of income/welfare and the size of aggregate income/welfare. The more we redistribute from rich to poor, the higher are marginal tax rates, and the lower is labor supply, employment and the size of aggregate income. If hours-of-work elasticities are large, this tradeoff tends to be very unfavorable. Following this literature, the old debate on redistribution was a classical left-right debate, with the left arguing that redistribution is desirable and the right arguing that labor supply responses are large. By contrast, following Saez (2002) and others, the new debate on redistributive policies focuses to a smaller extent on the size of welfare programs and to a larger extent on the shape of tax-transfer programs and the incentives they create in the decision to enter or exit the labor market. The new debate asks whether it is desirable to increase the incentives to work at the bottom by redistributing from the middle- and high-income earners to the working poor, rather than to non-workers as in the old debate.

---

18. Note, that the economic incentive to enter/exit is unchanged for individuals with potential earnings above  $Y^*$ . The reform has increased their tax payment if they work, but the tax payment for those outside the labor market is increased by the same amount. Hence, the difference in tax payments between working and not working is not changed for individuals with potential earnings above  $Y^*$ .

In Immervoll et al. (2004) we try to cast light on the welfare reform debates, both the old debate on traditional welfare programs and the new debate on redistribution towards the working poor. We construct a simple and fully explicit model encompassing labor supply responses along both the intensive and extensive margins and we then apply the model to the analysis of a welfare reform for 15 European Union countries using the EUROMOD micro-simulation model that has recently become available.

The EUROMOD micro-simulation model combines a tax and benefits calculator with detailed country-specific, but partly harmonized, micro data on income, earnings, labor force participation, as well as many demographic variables. For any set of household characteristics and country, EUROMOD is able to calculate the amount of benefits the household is entitled to and the taxes it should pay. EUROMOD has been constructed to incorporate all relevant tax and transfer programs in place in all countries that were members of the European Union prior to May 1, 2004. It is therefore a unique tool for obtaining a complete picture of the incentives to work generated by these programs as well for analysing of a welfare reform.<sup>19</sup>

We consider two types of redistributive reforms: a traditional welfare policy with universal transfers and a working poor policy with transfers targeted to the working poor. The traditional welfare policy raises the tax rate on all units of income by 1 percentage point and returns the collected revenue as a lump sum to all individuals in the economy. The consequence on labor supply incentives is illustrated in Figure 8. The linear dashed curve represents the relationship between disposable income and earnings prior to the reform, while the solid curve illustrates the relationship after the reform. The reform raises the tax burden on high-income individuals and reduces the tax burden (net of transfers) on low-income workers as well as on non-workers. The marginal tax rate is increased at all earning levels, implying that workers reduce hours worked (illustrated by the small arrows in Figure 8). Since transfers are universal, non-workers gain more from the reform than workers do. The reform therefore increases the participation tax rate, leading to a reduction in the population wanting to participate in the labor market (illustrated by the large arrows in Figure 8).

The working poor policy differs from the traditional policy in one important aspect: the tax proceeds now finance a transfer given only to those who work. The policy therefore improves the standard-of-living for low-skilled individuals participating in the labor market compared to those out of work. By implication, non-participating individuals are induced to enter the labor market at low-earnings occupations.

The desirability of each type of welfare reform depends on several factors. In general, as discussed extensively above, the efficiency effects of a tax-transfer reform de-

---

19. Detailed information on EUROMOD is provided by Sutherland (2001) and on the Internet at <http://www.econ.cam.ac.uk/dae/mu/emod.htm>.

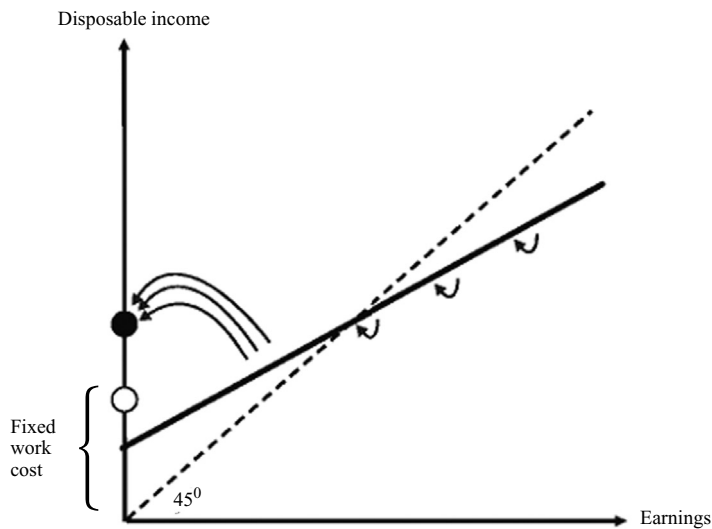


Figure 8. Traditional redistribution policy.

depends on the size of elasticities along each margin of labor supply response (hours-of-work and participation elasticities), and on the size of tax rates along each margin (marginal and participation tax rates). If effective marginal tax rates are initially high, the increase in taxes creates large distortions along the intensive margin. For both policies, this gives an unfavorable trade-off between equity and efficiency. By contrast, the tax on labor force participation affects the impact of the two policies in opposite directions. The presence of high participation tax rates makes the traditional welfare policy more costly, whereas the working poor policy tends to become more desirable. The reason is that the traditional policy reduces participation, while the working poor policy stimulates it. In particular, the working poor policy will be attractive when participation tax rates are high at the bottom of the earnings distribution, because this is where the additional entry takes place.

In addition to elasticities and tax rates, the distribution of earnings in the population will be important. For example, if the degree of earnings equality is already very high, the gain from additional redistribution will be relatively low. To see this, recall that tax and transfer payments for each individual are functions of earned income. If the income distribution is strongly compressed, the taxes paid and the benefits received tend to be of a relatively similar size for most individuals in the population. Hence, a reform that increases tax rates and benefits generates relatively little redistribution per dollar of deadweight burden, thereby creating an unfavorable equity-efficiency trade-off. Indeed,

*Table 1. Ranking of 15 EU countries according to earnings inequality and tax rates.*

<i>Equality of pre-tax earnings</i>															
Low															High
	UK	IR	LU	SP	AT	NL	BE	GE	PT	FR	GR	IT	DK	FI	SW
<i>Average effective marginal tax rate on labor income</i>															
Low															High
	SP	GR	UK	LU	PT	IR	NL	AU	IT	FR	GE	SW	DK	FI	BE
<i>Relative tax burden on labor market participation of low-skilled individuals</i>															
Low															High
	GR	LU	AT	UK	GE	SP	IT	NL	BE	PT	FI	FR	SW	IR	DK

*Source:* Authors' own calculations based on EUROMOD simulations reported in Immervoll et al. (2005).

*Note:* The calculations are based on a sample of individuals aged 18 to 59 who have been working the whole year. The tax and benefit rules we consider are those that were in place in 1998. The earnings equality is measured as earnings in the lowest quintile relative to earnings in the highest quintile. The average effective marginal tax rate equals the sample population average of the individual tax rates. The relative taxation on labor market participation of low-skilled individuals is calculated as the average tax rate on participation for those in the first 3 deciles of the earnings distribution relative to the sample average.

the presence of a high earning equality in Scandinavian countries is often mentioned as an argument against introducing working poor policies in this region. However, it is important to notice, that high income inequality tends to make both the working poor policy and the traditional welfare policy costly. If the presence of a high earning equality is used as an argument against working poor policies, it may also be used as an argument for rolling back traditional welfare programs targeted at those out of work. Still, the effect is somewhat less pronounced for the traditional welfare policy because it also gives transfers to those with zero earned income. We will come back to the implications of a high earnings equality for the design of a welfare reform below.

The EUROMOD calculations in Immervoll et al. (2004) provide detailed information on the distribution of earnings along with effective tax-benefit rates applying on the two margins of labor supply for the pre-2004-enlargement countries of the European Union. In Table 1, we briefly summarize some of the results from these calculations.

The upper panel in the table ranks countries according to the equality of pre-tax earnings as measured by income at the 80th percentile relative to income at the 20th percentile in the distribution. As one would expect, Nordic countries have the highest equality of earnings while Anglo-Saxon countries have the lowest. In the United Kingdom, earnings at the 80th percentile is more than 3 times as large as earnings at the 20th percentile, while the ratio in Sweden are only around 1.5. Moreover, the countries with a high degree of pre-tax earnings equality tend to be those with very high tax

rates.<sup>20</sup> Marginal tax rates tend to be highest in the Nordic countries, somewhat lower in central/continental Europe,<sup>21</sup> and lowest in Anglo-Saxon and Southern European countries. The variation in tax rates is substantial. At the low end, we have Spain where the average effective marginal tax rate is only 36 percent. At the other end, Belgium has a tax rate that is close to 70 percent.

The Nordic countries also have some of the highest tax burdens on labor force participation for the low-skilled. In countries such as Denmark and Sweden, participation tax rates are largest at the bottom, because the implicit tax on working created by generous earnings- and work-tested benefits, weigh more heavily on low-income people. Moreover, unemployment insurance benefits can be subject to a floor so that replacement rates for low-wage earners can be very high in some cases. By contrast, countries such as Greece, Luxembourg, Spain, and the United Kingdom have relatively lower tax rates at the bottom because minimum income programs do not exist or are modest relative to earnings, because tax rates on earned income tend to be small and/or because they operate in-work benefits which partly offset the loss of social assistance or unemployment benefits at the point of entry. Note, finally, that Ireland and France have relatively high participation tax rates compared to their marginal tax rates.

Table 2 shows estimates of the equity-efficiency trade-off for the two kinds of welfare policy. We define this trade-off as the euro value of the welfare loss for those who lose from the reform (the rich) in proportion to the euro value of the welfare gain for those who gain (the poor or working poor). In other words, the trade-off gives the welfare cost to the rich from the transfer of one additional dollar of welfare to the poor. If a given redistributive reform reduces economic efficiency, the welfare cost on the rich from transferring an additional euro to the poor will be larger than one euro. That is, in the process of redistributing income from rich to poor something is lost due to behavioral responses, giving rise to a trade-off larger than one.<sup>22</sup> The benchmark case

---

20. This is contrary to the policy recommendation in the optimal tax literature. In this literature, a high degree of equality in abilities/productivities tends to make redistribution more costly, so that optimal marginal tax rates are lower. Of course, the pre-tax earnings distribution is in itself affected by the tax system, and does not measure the ability distribution relevant for optimal taxation. However, the presence of high marginal tax rates would normally lead to labor supply responses making the pre-tax earnings distribution (with the tax system in place) more unequal than the ability distribution without a tax system. Hence, this point serves to reinforce the contradiction between optimal and practical policy. Of course, this gap between optimality and practice might be explained by other differences across countries such as different social preferences for distributional equity. It should also be kept in mind that the optimal tax model is not really a model to explain actual government behavior, but one to analyze the best possible behavior under a given set of government preferences.

21. Luxembourg is an exception. As other smaller and very wealthy European countries or principalities such as Liechtenstein or Switzerland, tax rates are significantly lower in Luxembourg than in other larger continental European countries.

22. Okun (1975) used the metaphor of a leaky bucket to illustrate the efficiency loss, taking place in income redistribution.

*Table 2. Trade-off between equity and efficiency across 15 EU countries.*

<i>Traditional welfare policy</i>	Trade-off	Countries
	1.5 – 2	GR LU SP UK
	2 – 4	AT IR IT NL PT
	4 – 25	BE DK FI FR GE SW
<i>Working poor policy</i>	Trade-off	Countries
	0 – 1	DK FR IR PT SP
	1 – 1.5	AT GR LU NL UK
	1.5 – 5	BE FI GE IT SW

*Source:* Table 2 in Immervoll et al. (2005).

*Note:* Trade-off denotes the ratio of the welfare loss of losers to the welfare gains of gainers from the reform. We report the results from the baseline elasticity scenario considered in Immervoll et al. (2005). In this scenario the hours-of-work elasticity equals 0.1, while the participation elasticity for the aggregate economy is equal to 0.2. The calculations are based on tax and benefit rules that were in place in 1998.

where the trade-off is equal to 1, so that the welfare gain of the poor is exactly equal to the welfare loss of the rich, reflects the situation where a reform involves no efficiency cost (or gain). Finally, if the trade-off were to be less than one, there is no conflict between distributional equity and economic efficiency for the reform under consideration. The trade-off measure used, was originally proposed by Browning and Johnson (1984).

Redistributing income to the poor by increasing the level of universal transfers leads to efficiency losses in all countries, implying a trade-off larger than one. For most European countries, expanding the generosity of traditional welfare programs creates large efficiency costs: for 11 out of 15 countries, redistributing one additional euro to low-income individuals requires a reduction in the welfare of high-income individuals by 2 euros or more. This is due to the fact that most European countries already impose quite large marginal tax rates on earnings. The largest trade-offs are found in the Nordic and Continental European countries where taxes and benefits are high. In fact, the 6 countries with the highest trade-off between equity and efficiency are identical to the countries having the highest marginal tax rates in Table I. Likewise the lowest trade-offs are associated with the countries with the lowest marginal tax rates.

A completely different picture emerges once we turn to the working poor policy. For all countries, the equity-efficiency trade-off is substantially lower for this type of reform. In fact, for Denmark, France, Ireland, Portugal, and Spain, the policy would create an aggregate welfare gain such that the trade-off is lower than 1. Hence, in these countries it would cost less than 1 euro for the rich to transfer 1 euro to the working poor. In the case of Denmark, the trade-off is very close to zero – it costs close to

nothing for the rich to give money to the working poor – implying that the reform is (almost) Pareto improving. For a number of other countries, the working poor policy creates very small efficiency losses such that the trade-off is quite close to one. This applies to countries such as Austria, Greece, Luxembourg, the Netherlands, and United Kingdom. In these countries there is no significant trade-off between efficiency and equality when we consider redistribution from the rich to the working poor. Only in the case of Finland and Sweden does the working poor policy involve an unfavorable equity-efficiency trade-off, which reflects in part the extremely equal earnings distribution in these two countries.

A comparison of Table 1 and 2 reveals that countries with relatively high participation tax rates at the bottom of the earnings distribution, such as Denmark, Ireland and France, tend to gain more by choosing a working poor policy rather than a traditional welfare policy. The working poor policy creates higher incentives for participation in the labor force. Moreover, participation increases mainly at the bottom deciles where participation elasticities are large. If participation tax rates are very large at the bottom deciles, the increase in labor market participation creates a large increase in government revenue (through reduced benefit expenditures and a higher tax take) and hence in economic efficiency.

*Ceteris paribus*, the presence of a highly equal pre-tax earnings distribution tends to make redistribution very costly. This applies to redistributive policies in general, and to the working poor policy in particular because it attempts to redistribute only within the group of workers who are more equal than the entire population of workers and non-workers. The presence of strongly compressed wage distributions in Scandinavian countries (cf. Table 1) is often mentioned as an argument against pursuing in-work benefit reforms in this region. It is therefore interesting that the working poor policy seems to be extremely desirable in Denmark, more so than in any other country. The explanation is that participation tax rates are very high in Denmark, especially in the bottom deciles where participation elasticities are also high. In the simulations, this effect strongly dominates the effect of a compressed earnings distribution. On the other hand, the working poor policy is less desirable in Finland and Sweden, because participation tax rates at the bottom are somewhat lower there and because the earnings distribution is even more equal than in Denmark. It is important to realize, however, that the equity-efficiency trade-off is even worse for the traditional welfare policy in Finland and Sweden. The bottom line is that both kinds of redistribution are very costly in these two countries, at least on the margin, due to the combination of an already equal distribution and high tax rates.

Besides the presence of an equal earnings distribution, a potential argument against working poor policies in Scandinavia is that participation elasticities might be signifi-



cantly lower in this region than in, say, Anglo-Saxon countries. It is true that most of the estimates of high participation-elasticities were based on the United States and the United Kingdom where female labor force participation rates have been lagging behind those of Scandinavia, creating more room to increase participation through in-work benefits targeted at low-income females. This argument is potentially important and points to the need for good empirical studies on the responsiveness of labor market participation. Still, there are several counter-arguments that one can make at this point. Firstly, as mentioned in Section 2.1, there are some empirical studies suggesting that female participation elasticities may be high even in Northern European countries (Germany, Sweden, and the Netherlands). Second, while female labor force participation is relatively high in Scandinavia, there is in fact substantial non-participation among a number of demographic subgroups (e.g. immigrants, single mothers, and the young). In Denmark, around 1/4 of those in the working age (15-64) are outside the labor market. Thirdly, and perhaps most importantly, even if participation elasticities were to be substantially lower in Scandinavian countries, a working poor policy may still be desirable due to the fact that the existing welfare system is a lot more distortionary. This argument is very important for Denmark where participation tax rates, in an international perspective, are extremely high at the bottom of the income distribution. In Immervoll et al. (2005), we find that the equity-efficiency trade-off for the working poor policy is exactly equal to 1 in Denmark – implying no efficiency loss for this type of redistribution – at an average participation elasticity for the population equal to 0.13, a number far below the existing international estimates.

To conclude this section, expanding traditional welfare programs have very different implications than introducing in-work benefits. Increasing redistribution through traditional welfare can lead to significant negative labor supply responses along both the intensive and the extensive margins. By contrast, in-work benefit reforms generates positive labor supply responses along the extensive margin (along with a negative response along the intensive margin). As a result, the efficiency cost of redistribution through this type of reform is much smaller.<sup>23</sup>

## 6. Discussion

In this survey, we discussed some recent research on the effects and design of redistributional policies, emphasizing the distortionary effects of taxes and transfers on

---

23. It needs to be emphasized, however, that the groups who benefit from redistribution in the two reforms are different. In the traditional welfare case, those who benefit the most are people with no earnings at all, presumably those who are the most in need of support. In the in-work benefit case, people with no earnings receive no additional support and redistribution benefits only the working poor. Finally, notice that the implementation of one policy does not exclude the other policy. Hence, the two policies should be seen as complements rather than substitutes in the design of redistribution policy.

labor supply behavior, employment and economic efficiency. We paid special attention to the modern labor market literature showing that labor force participation is more responsive to taxes and transfers than is hours worked. It was shown that the presence of significant participation responses tend to make an in-work benefit reform inducing individuals from welfare into work a desirable policy option. By contrast, traditional welfare policies providing either universal or earnings-tested benefits are very costly in terms of government revenue and efficiency resulting from adverse effects on labor force participation. This point is reinforced by the fact that earnings- and work-tested benefits generate the highest effective tax rates on labor force participation at the bottom of the earnings distribution where participation elasticities are particularly high.

It needs to be emphasized, however, that the individuals who gain from in-work benefit reforms are different from those gaining from traditional welfare policies. In the latter case, those who benefit the most are people who are not working and have zero earned income, presumably those in the strongest need of support. With in-work benefit policies, people with no earnings receive no additional support and redistribution helps only the working poor. As a result, if the government has extremely strong redistributive tastes and puts a much higher welfare weight on those with no earnings than on the working poor (such as in the case of a Rawlsian welfare criterion), it is possible that expanding traditional welfare would be more desirable than introducing in-work benefits. By contrast, if the government puts lower welfare weights on those with no earnings than on the working poor, the case for in-work benefits would be even stronger. It seems that conservative governments (especially in Anglo-Saxon societies) tend to hold the latter view: those not working are seen as »lazy«, whereas the working poor are seen as »deserving«.

While the implementation of benefits targeted to the working poor was pioneered by the United States (starting in the 1970s), in-work benefit reforms has recently been high on the agenda in political discussions of welfare reforms in Europe. The first European countries to introduce in-work benefit programs were the United Kingdom (in 1988, subsequently expanded in 1999) and Ireland (in the early 1990s). In the past 10 years, a number of other European countries have introduced programs of a similar kind. France introduced an in-work benefit program as of 2001 (»Prime Pour l'Emploi« or premium for employment). The Netherlands introduced an Employment Tax Credit in 2001, while Belgium phased-in an Earned Income Tax Credit program between 2002 and 2004. Germany introduced the so-called »Mainzer Modell« program in 2002 which, more recently, has been replaced by SSC reductions available to low-wage earners. Finland has recently introduced and then expanded an Earned Income Tax Allowance, and Denmark introduced a similar type of program on a small scale in 2004. Finally, while the Italian family benefit has not normally been conside-

red an in-work benefit, it increases with the number of days worked. However, with the exception of Ireland and the United Kingdom, the European in-work benefit programs are still modest in size with maximum annual benefits between 300 and 1000 Euros, see OECD (2004), for the most recent and systematic description of these programs. Therefore, the »small reform« methodology adopted in this paper would seem to provide a good approximation of the effect of introducing these programs. Our results lend support to the conjecture that these policies constitute a very efficient type of redistribution, and that the expansion of such policies should perhaps be part of future welfare reforms.

The simple labor supply model underlying the analysis in this paper abstracts from a number of issues, which we would like to discuss briefly. First, and perhaps most importantly, we have assumed that the labor market is perfectly competitive. This might be a poor approximation of European labor markets, where minimum wages tend to be substantial, and where wage rates are often the result of bargaining between unions and employers. Minimum wages prevent employers from paying wages, that are below a defined minimum, thereby eliminating jobs with very low productivities and potentially creating involuntary unemployment among the low-skilled. Likewise, union bargaining models, efficiency wage models, and search models imply that a fraction of individuals become involuntarily unemployed.

The effects of taxation in imperfect labor markets have been explored in a number of papers, see, e.g., Hansen et al. (1996). The introduction of imperfections will not change the most important mechanisms at work in our analysis. Firstly, variation in aggregate employment is still the result of behavioral responses along the intensive and the extensive margins. For example, Sørensen (1999) considers optimal taxation in three different models of involuntary employment (unions, efficiency wages, and search) where both intensive and extensive responses are present. Secondly, in all imperfect labor market models, a reduction of average tax rates leads to higher employment, where the effect is channeled through lower equilibrium wages. Accordingly, a working poor policy would lead to increased job opportunities, while a demo grant policy would reduce the chances of finding a job.<sup>24</sup>

While the most important employment effects would survive the introduction of labor market imperfections, the welfare implications of changed labor force partici-

---

24. In one respect the imperfect labor market models do involve different comparative statics than the competitive model. This difference relates to the effect of changes in the marginal tax rate. A higher marginal tax rate (for a given value of the average tax rate) may lead to a lower equilibrium wage rate which, *ceteris paribus*, implies higher employment. At the same time, higher marginal tax rates give rise to lower working hours for those who are working as in the standard competitive model. However, the important point to note is that these effects would apply to both the demo grant and the working poor policies, since both types of reforms increase the marginal tax rate. The main difference between the reforms lies in their implications for the average effective tax rates at the bottom of the earnings distribution.

pation would be affected by the presence of involuntary unemployment. Following the introduction of in-work benefits, those who obtain jobs would experience a discrete (as opposed to an infinitesimal) increase in utility because some of them were previously involuntarily unemployed. This reinforces the positive effect of the working poor policy on welfare. Increasing traditional welfare programs, on the other hand, creates higher unemployment. To the extent that people lose their jobs involuntarily, the welfare loss is exacerbated relative to the case of voluntary unemployment. In conclusion, a model with labor market imperfections generating involuntary unemployment would most likely increase the attractiveness of redistributing to the working poor and reduce the attractiveness of increasing traditional welfare programs, thereby reinforcing the main conclusion of this paper.

Second, there might be issues related to the presence of segmented labor markets. A well-known hypothesis is that labor markets tend to have a dual structure, being segmented into a perfectly competitive sector offering low-paying, low-productivity jobs and an imperfectly competitive sector offering high-paying, high-productivity jobs. Indeed, labor economists have gathered considerable evidence in favor of the dual labor market hypothesis, see, e.g., the survey by Saint-Paul (1996) pp. 62-8. In the dual labor market model, there is a distortion in the allocation of employment in favor of the perfectly competitive sector offering low-paying jobs. As pointed out by Bulow and Summers (1986), this implies that the government ought to use industrial policy to shift resources away from the low-productivity sector. In the context of a tax reform, Kleven and Sørensen (2003) show that such sectoral distortions tend to make policies aimed at the working poor less attractive, because they promote bad jobs at the expense of good jobs. A policy which succeeds in increasing aggregate employment by promoting low paying jobs may, in theory, reduce welfare as it creates a deterioration in the sectoral mix of employment. It seems unlikely, however, that a composition effect of this sort, would be able to dominate the large gains from a higher level of employment at the bottom of the distribution. Even with segmented labor markets it remains the case that higher employment at the bottom creates large effects on government revenue (and hence efficiency), as participation tax rates are particularly high at the bottom.

Labor force participation may also generate externalities. Positive externalities of working would make the introduction of in-work benefits even more attractive relative to traditional welfare, while negative externalities would make in-work benefits less attractive. Some of these externalities take the form of fiscal externalities, where higher employment rates affect the demand for certain commodities that are initially taxed or subsidized by the government. For example, higher employment may generate more demand for childcare, which would then create positive or negative externalities depending on whether this commodity carries a positive or negative tax rate (in the Nordic countries, for example, child care is heavily subsidized). Externalities

could also come in the form of social externalities. Positive social externalities would be reduced crime (as working individuals have less need and time to resort to criminal activities), newly employed parents being better role models for their children (which could increase the incentives of children to do well at school, etc.). Negative externalities are also a possibility if working reduces the time that parents can devote to their children and therefore worsens the quality of parental education.

Finally, a large body of work in behavioral economics has shown that individuals are not always able to make the best decisions for themselves, especially when those decisions involve intertemporal trade-offs. In the case of labor supply, it is conceivable that some individuals may not perceive the full future benefits of starting to work, or procrastinate in the decision to leave welfare and start working. Such models with inconsistent time preferences generate so-called internalities, Herrnstein et al. (1993), that are conceptually close to externalities: an individual may not internalize fully the utility of future selves and hence decide to work too little today. As a result, in-work benefits may be desirable to induce people to work more and help correcting such internalities.

#### References

- Ballard, C. L. 1988. The Marginal Efficiency Cost of Redistribution. *American Economic Review* 78, 1019-133.
- Ballard, C. L., J. B. Shoven, and J. Whalley. 1985. General Equilibrium Computations of the Marginal Welfare Costs of Taxes in the United States. *American Economic Review* 75, 128-38.
- Blundell, R. 1995. The Impact of Taxation on Labor Force Participation and Labor Supply, Chapter 3, OECD Job Study: *Taxation, Employment, and Unemployment*, OECD: Paris.
- Blundell, R. W. and T. MaCurdy (1999). Labor Supply: A Review of Alternative Approaches, in O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics* vol. 3A. Elsevier Science B.V.: Amsterdam.
- Blundell, R. W., A. Duncan, J. McCrae and C. Meghir. 2000. The Labour Market Impact of the Working Families Tax Credit. *Fiscal Studies* 21(1), 75-104.
- Blundell, R. W., A. Duncan, and C. Meghir. 1992. Taxation and Empirical Labour Supply Models: Lone Parents in the UK. *Economic Journal* 102, 265-78.
- Blundell, R. W., J. Ham and C. Meghir. 1987. Unemployment and Female Labour Supply. *Economic Journal* 97, 44-64.
- Browning, E. K. 1987. On the Marginal Welfare Cost of Taxation. *American Economic Review* 77, 11-23.
- Browning, E. K. 1995. Effects of the Earned Income Tax Credit on Income and Welfare. *National Tax Journal* 48, 23-43.
- Browning, E. K. and W. R. Johnson. 1984. The Trade-Off between Equality and Efficiency. *Journal of Political Economy* 92(2), 175-203.
- Bulow, J. I. and Summers, L. H. 1986. A Theory of Dual Labor Markets with Application to Industrial Policy.
- Cogan, J. F. 1981. Fixed Costs and Labor Supply. *Econometrica* 49, 945-63.
- Eissa, N. 1995. Taxation and Labor Supply of Married Women: The Tax Reform Act of 1986 as a Natural Experiment. *NBER Working Paper* No. 5023.
- Eissa, N. and H. Hoynes. 2004. Taxes and the Labor Market Participation of Married Couples: The Earned Income Tax Credit. *Journal of Public Economics* 88, 1931-1958.
- Eissa, N., H. J. Kleven, and C. T. Kreiner. 2004. Evaluation of Four Tax Reforms in the United States: Labor Supply and Welfare Effects for Single Mothers *NBER Working Paper* no. w10935.

- Eissa, N., H. J. Kleven, and C. T. Kreiner. 2005. Welfare Effects of Tax Reform, and Labor Supply at the Intensive and Extensive Margins, in J. Agell and P. B. Sørensen (eds), *Tax Policy and Labour Market Performance*, forthcoming at MIT Press, Cambridge, MA.
- Eissa, N. and J. Liebman. 1996. Labor Supply Response to the Earned Income Tax Credit. *Quarterly Journal of Economics* 61, 605-37.
- Feldstein, M. 1995. The Effect of Marginal Tax Rates on Taxable Income: A Panel Study of the 1986 Tax Reform Act. *Journal of Political Economy* 103, 551-72.
- Hansen, C. T., L. H. Pedersen, and T. Sløk. 1996. Danske resultater om sammenhængen mellem marginalskat og løn. *Nationaløkonomisk Tidsskrift* 134, 153-74.
- Harberger, A. C. 1964. Taxation, Resource Allocation, and Welfare. in John F. Due (ed.), *The Role of Direct and Indirect Taxes in the Federal Revenue System*, Princeton University Press, pp. 25-70.
- Heckman, J. 1993. What Has Been Learned About Labor Supply in the Past Twenty Years? *American Economic Review Papers and Proceedings* 83(2), 116-21.
- Herrnstein, R. J., Loewenstein, G. E., Prelec, D. and Vaughan Jr, W. 1993. Utility maximization and melioration: Internalities in individual choice, *Journal of Behavioral Decision Making*, vol. 6(3) (September), 149-85.
- Immervoll, H., H. J. Kleven, C. T. Kreiner, and E. Saez. 2005. Welfare Reform in European Countries: A Micro-Simulation Analysis. *Economic Journal* (forthcoming).
- Kleven, H. J. and C. T. Kreiner. 2004a. A Note on Optimal Income Taxation. Lecture Note, University of Copenhagen.
- Kleven, H. J. and Sørensen, P. B. 2004. Labor Tax Reform, The Good Jobs and The Bad Jobs, *Scandinavian Journal of Economics* vol. 106(1) (March), pp. 45-64.
- Krueger, A. B. and B. D. Meyer. 2002. Labor Supply Effects of Social Insurance, in Alan J. Auerbach and Martin Feldstein (eds.), *Handbook of Public Economics*, vol. 4. Amsterdam: North-Holland.
- Meyer, B. and D. Rosenbaum. 2001. Welfare, the Earned Income Tax Credit, and the Labor Supply of Single Mothers. *Quarterly Journal of Economics* 66, 1063-114.
- Mirrlees, J. A. 1971. An Exploration in the Theory of Optimal Income Taxation. *Review of Economic Studies*, 38, 175-208.
- Moffitt, R. A. 2002. Welfare Programs and Labor Supply, in A. J. Auerbach and M. Feldstein (eds), *Handbook of Public Economics* vol. 4, Elsevier: Amsterdam.
- Mroz, T. A. 1987. The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions. *Econometrica* 55, 765-99.
- OECD (2004) *Benefits and Wages*, OECD: Paris.
- Okun, A. M. 1975. *Equality and Efficiency: The Big Tradeoff*. Washington: Brookings Institution.
- Saez, E. 2002. Optimal Income Transfer Programs: Intensive Versus Extensive Labor Supply Responses. *Quarterly Journal of Economics* 117(3), 1039-073.
- Saez, E. 2004. Reported Incomes and Marginal Tax Rates, 1960-2000: Evidence and Policy Implications, in J. Poterba, ed., *Tax Policy and the Economy*, Volume 18, Cambridge: MIT Press.
- Saint-Paul, G. 1996. *Dual Labor Markets: A Macroeconomic Perspective*, Cambridge, Massachusetts: MIT Press.
- Small, K. and H. S. Rosen. 1981. Applied Welfare Economics with Discrete Choice Models. *Econometrica* 49, 105-30.
- Sutherland, H. 2001. Final Report EURO-MOD: An Integrated European Benefit-Tax Model. *EUROMOD Working Paper* No. EM9/01.
- Sørensen, P. B. 1999. Optimal Tax Progressivity in Imperfect Labour Markets. *Labour Economics* 6, 435-52.
- Triest, R. K. 1990. The Effect of Income Taxation on Labor Supply in the United States. *Journal of Human Resources* 25, 491-516.
- Triest, R. K. 1994. The Efficiency Cost of Increased Progressivity, in Joel Slemrod (ed.), *Tax Progressivity and Income Inequality*. Cambridge: Cambridge University Press.

# What affects students' performance? An investigation of the importance of admission characteristics

Lisbeth la Cour

Department of Economics, CBS, E-mail: llc.eco@cbs.dk

Pascalis Raimondos-Møller

Department of Economics, CBS, E-mail: prm.eco@cbs.dk

*SUMMARY: The present study relates student's performance, as measured by the grades the students achieve in individual course exams, and the admission characteristics. We focus on the group of students who are enrolled in the Bachelor of Science in International Business (B.Sc.IB) at Copenhagen Business School (CBS) – a program that bases 20% of its student admission on kvote-1 (quota-1) students and 80% on kvote-2 (quota-2 students). This unique admission system for Denmark allows us to examine what characteristics matters for the quota-2 group of students. Our results emphasize the importance of high-school GPAs, while all other variables are either insignificant or significant for the student performance in a couple of individual courses.*

---

## 1. Introduction

Information about student academic performance is an increasingly important issue for universities and other agents related to the educational sector, e.g. potential university applicants, prospective employers of graduates, and public-policy makers in government. While the demand for such information has been high in countries with private or quasi-private education markets, see Blundell et al., (1997); and Smith et al., (2000), in Denmark such information, if collected, has rarely been processed and analysed. There are reasons, however, to believe that this may change in the future. For example the newly introduced admission fees for students from countries outside the European Economic Area opens up a new market for universities in Denmark that necessitates a careful design of admission selection criteria. Statistics on what admission

---

We acknowledge comments by two referees, Karsten Albæk and seminar participants at CBS and at the 2005 Danish Applied Statistics Meeting. We also thank Ulla Hansen, Tanja Christiansen, and Katrine Dam for assistance with the data.

characteristics students should have will become even more important for the success of particular university programs in attracting foreign students.

The present paper is one of the few that collects and analyses information about student performance and their admission characteristics. In particular the paper investigates the relation between students' performance during studies and their characteristics at the time of admission. We focus on the group of students who are enrolled in the Bachelor of Science in International Business (B.Sc.IB) at Copenhagen Business School (CBS); a program that for Denmark has a unique, admission procedure, which allows an interesting analysis of the above issues.

The B.Sc.IB program has been accepting around 90 students every year since 1996. These approximately 90 students are admitted through a 20% »quota-1«, i.e. based on the high-school Grade Point Average (GPA), and 80% »quota-2« system, i.e. based on points that students collect through a number of activities that are not necessarily linked to their high-school GPAs. In 1995 it was argued that this unique system was necessary to ensure a 50% Danish/50% non-Danish student participation – a cornerstone of the Program's international orientation. Given this 20-80 rule, applicants who apply directly from high-school compete for only 18 places. These few »quota-1« places ensure that the admission GPAs have been as high as 10.1 on the Danish scale.<sup>1</sup> For the rest of the 72 students, a point system that puts more weight on other things than the GPA level is used.<sup>2</sup> In brief, while a GPA is rewarded by a maximum of 12 points, living abroad is rewarded by a maximum of 12 points, extra-curricular activities are rewarded by a maximum of 16 points, other education is rewarded by a maximum of 12 points, working experience abroad is rewarded by a maximum of 36 points, and finally a motivational essay (which is carefully and independently read and evaluated by two program managers) is rewarded by a maximum of 20 points. It is clear from the above that the admission process of the quota-2 students exhibits a very strong bias against GPAs and the qualifications they represent. The question is whether such a bias is well founded.

It is important to mention that the B.Sc. IB program is, judging by the demand for the program a success story: the number of applications has been steadily increasing every year, topping with 440 applications for the »quota-2« admission group in 2004 (a 20% growth from the previous year). From these 440 applications approximately

---

1. This high average admission GPA has been the highest among all university programs in Denmark for several years now, with a considerable media attention, and thus with considerable positive externalities (for 2004 see »Real Businessclass« Politiken, 05.09.2004, 6. section, p. 31). This high admission average remained at the same level (10.1) when the Ministry of Education increased the quota-1 percentage to 40% in the 2005/2006 academic year. Note that the Danish grading scale is the so-called »13-scale«, where the following grades exist 0, 3, 5, 6, 7, 8, 9, 10, 11, 13 and where the passing grade is 6.

2. See Appendix 1 for details for the admissions procedure. This procedure has changed as of the 2005/2006 academic year.



20% did not fulfil the minimum requirements for entry into the program, leaving approximately 350 students to compete for the 72 places. This intensive competition for entering the program implies that, over the years, the IB program has indeed had the opportunity to select among applicants. Such an opportunity does not exist in other Danish programs where 80% of applicants are accepted within the »quota-1« group and only 20% are accepted within the »quota-2« group.<sup>3</sup> In that sense, the B.Sc. IB program is the only one in Denmark that has had a long experience in selecting among quota-2 students. The present analysis attempts to draw some lessons from this experience. These lessons are topical given the recent interest the Ministry of Education has shown on what should be accounted for in the »quota-2« admission process.<sup>4</sup>

Aiming thus to contribute to the debate about admission rules, the present study relates students' performance, as it is measured by the grades the students achieve in individual course exams, and the admission characteristics they have. We connect data from CBS's central register system (HSAS) that records all grades students get during their education, with their admission characteristics data the B.Sc. IB office has recorded for 2002 and 2003.<sup>5</sup> A priori, one should not expect a high explanatory validity of the analysis (i.e. a low  $R^2$  is expected); after all, becoming more mature through education, the social and learning environment that students face, the quality of the teachers, etc., affects students' performance. Thus, our hope is to find statistically significant relations, i.e. relations that the data reveal as important in determining the likelihood of having high or low grades, and not finding that admission characteristics explain a large percent of the existent grade's variation (one could also argue that the latter could be interpreted as a defeat for the program's learning potential).<sup>6</sup>

To this end, the following two caveats apply. (i) Some will argue that »performance« is not only described by a grade. In particular, we are aware that a grade captures a particular ability to reproduce knowledge. Still, this is the only »fact« that is recorded in the data, and it is something widely used in our education system. (ii) The large number of group project exams introduces ambiguities as for the interpretation of the grade

3. Exceptions do exist, but they are either too small in student numbers, or too young to be considered in a statistical analysis.

4. While the »quota-1/quota-2« rule is determined by the Ministry of Education, the weights given under each non-GPA activity in the »quota-2« admission system are at the discretion of the particular program.

5. Even if the B.Sc. IB program has been running since 1995, a detailed database that records the points an admitted student gets has unfortunately only been collected in a systematic way for the last 3 years, i.e. 2002, 2003 and 2004.

6. A priori one may expect that the explanatory power of the equations would have increased by introducing of a year dummy to account for quality of the teacher and the learning environment. We have tried to include such year dummies but these dummies only became significant for two courses and only with the result that the  $R$  square was increased. Size and sign of the other coefficients were hardly affected. Hence with no qualitatively new conclusion from these dummies we decided to exclude them from the models.

given,<sup>7</sup> and in avoiding these ambiguities, we have decided to abstract from group exams and focus on individual written exams. In this sense, we are linking admission characteristics to performance that does not capture learning externalities (as these may be represented in group work).<sup>8</sup>

Before we proceed we should note that our analysis is not unique in the literature. First of all there exists an elaborative literature on these issues in international journals – see, e.g. the study by Naylor and Smith (2004) for universities in the UK and their references within. Clearly such an aggregation among universities introduces some noise to the data, a problem that is avoided in, e.g., Tay (1994) who looks at performance and admission characteristics of a particular program in Singapore. For the case of Denmark we are only aware of two other studies that ask questions similar to ours, viz. Albæk (2001) and Andersen and Østergaard (2005). Albæk analyses the probability of passing the first year of the economics programme at the University of Copenhagen. As one possible explanatory factor Albæk uses high-school GPAs. The focus in his study is the effects of the admission rules on the economics programme and on how the probability of passing first year is affected by various background variables. His results indicate that there is a significantly increased probability of passing first year for students with mathematics at level A from high-school. In addition to this, the passing probability also seems to increase for students who have spent a couple of years doing other things other than university studies before they begin the course. Both of these findings are supported by the study by Andersen and Østergaard (op.cit).<sup>9</sup>

## 2. Description of the data

As mentioned above, the data have been collected for students who started the B.Sc. IB in 2002 and 2003, as these are the years where admission characteristics data have been systematically registered. CBS's central student register (HSAS) provided the data on grades from individual written exams with the latest entry being August 2004. These data are confidential, as they are constructed on the basis of the individual civil

---

7. The main question here is: »Has the student earned the grade as a result, or in spite, of the abilities of the other group members?«

8. One should note, however, that many students prepare their individual written exams together with other students, and thus some form of learning externality cannot be precluded.

9. Focusing on performance is also the topic in Lorentzen (2001) and Gabrielsen (1996a,b). The former paper analyses the background factors that leads to a good student performance at the English-language masters programmes at CBS. The latter papers investigate whether there is a »teacher effect« on the grades in the statistics course at CBS' HD programme. While the latter paper does not examine the influence of admission characteristics (the focus of our paper), the former paper looks only at Bachelor GPA and country of origin as admission characteristics.

registration numbers (cpr.nr.). Based on this database, we collected data on grades from the following 11 courses:<sup>10</sup>

- 111: The Company in an International Setting
- 112: Microeconomics
- 114E: International Business Environment
- 122E: Macroeconomics
- 123: Intercultural Communication
- 131: International Economics
- 132E: Principles of Marketing
- 212E: Management Information Systems (MIS)
- 222: Corporate Finance
- 234: Statistics for Business
- 241: International Business law

For 1<sup>st</sup> year courses the number of observations (grades per course) ranges from 72 to 146, indicating that while in some courses students get pre-approval from having studied similar courses previously, in other courses extra students attend first-year exams (second year students who want to improve their grades).<sup>11</sup> For 2<sup>nd</sup> year courses the number of observations ranges from 32 to 51, indicating in turn that many students have postponed taking the exam.

It should be noted that due to the B.Sc. IB's intense quarter structure, it is customary for the students to hand-in a blank answer just in order to get permission to a re-take exam (as it is required by the rules and regulations).<sup>12</sup> However, given the fact that a blank answer was (until mid 2004) awarded the grade, »0« we have chosen to abstract from observations with grade »0«. <sup>13</sup> The total number of students (i.e. cpr. numbers) in our data set is 156. However, not all 156 students took the same courses, as some of them may have had credit transfers from a previous education. Thus, for a single course the maximum number of students is 146.

Moving to what determines a student's grade, we focus on admission characteristics. More generally, however, a student's grade is a function of his or her: (*i*) intellectual

---

10. The first number in the code indicates the year the course is taught, while the »E« code denotes written exam, indicating that the course has other forms of evaluations too, e.g. a project.

11. Note that the first year has to be passed with an average of at least 6, and not have any »0« entries, in order to be allowed to take second year exams.

12. Each semester has two 7-week quarters. At the end of each quarter there is a week of exams. Two courses are taught intensively in each quarter. Responding to this exam pressure, students tend to postpone the exam by handing-in blank answers.

13. Clearly this may bias upwards the performance averages.

ability, *(ii)* academic background, *(iii)* personal factors, and *(iv)* instructional input. As we will see, our admission characteristics cover (albeit partially) intellectual ability, academic background, and personal factors, leaving instructional input uncovered. However, since we focus on grades in particular courses – which, note, have had the same teachers through the period of our analysis – we do take into account that different lecturers may have created different learning environments in the classes. This is clearly one of the advantages of focusing on course grades rather than on all grades together.

From all the admission characteristics we have access to, we start by including the following:

- (i)* whether a student is »quota-1« or »quota-2«,
- (ii)* if »quota-2« student, the level of the student's GPA in the admission process,
- (iii)* if »quota-2« student, the quality of the motivational essay (as that is represented by the average of the points two program managers award independently),
- (iv)* the age of the student upon admission,
- (v)* if »quota-2« student, whether the student is from Scandinavia or not.

Variable *(i)* is obvious. Variable *(ii)* represents the »hard« facts about quota-2 students, while variable *(iii)* represents the »soft« facts about »quota-2« students. The reason for not including more »soft« variables from the admission characteristics of a »quota-2« student is that we expect these variables to be inter-related.<sup>14</sup> Finally, we include a couple of other variables: *(iv)* age, since it can be argued that delaying studies has an ambiguous effect on performance (ability to learn vs. maturity and understanding), and *(v)* Scandinavian origin, since there can be an effect of knowing the country, the education system, or having a similar high-school education to Danish students have.<sup>15</sup>

Extending this simple model section 5 presents a sensitivity analysis (including the rest of the admission characteristics) that shows the appropriateness of our simple model. There we also introduce the grade and level of high-school Mathematics, and for explaining the performance in the second year exams, we introduce as an explanatory variable the first year GPA. The math grade is a very important variable in Albæk (2001) and Andersen and Østergaard (2005) and here we test whether the same is true

---

14. An additional analysis including all admission variables can be found in section 5.

15. A variable that is usually included in estimating the education production function is »Male«, see e.g. Tay (1994). We tried to include such a dummy at an early stage of the analysis but it wasn't significant and hence it was omitted.

for our data. The first year GPA is introduced to test whether the data exhibit a »progression in learning«, i.e. whether doing well in the first year implies good performance in the second year.

Before moving on to the estimations, some data averages are presented. As mentioned, 20% of the students are admitted as quota-1 students. Of the remaining 80%, roughly 50% come from a Scandinavian country. In total, approximately 50% of the students are Danes.<sup>16</sup> The average age of the quota-1 student is 20.38 years, while the average age of the quota-2 student is 23.78 years. Finally, we note that the average admission GPA of the quota-1 students is 10.1, while the average admission GPA for the quota-2 student is 8.87 (Scandinavian) and 8.46 (non-Scandinavian).

### 3. The estimated equation

We shall now describe the equation that we estimate. Our dependent variable is the grades in each 1<sup>st</sup> and 2<sup>nd</sup> year course that has a written examination. The independent variables are the following: a dummy variable for whether the students have gained access by quota-1 or quota-2 and a variable for the age of the students when admitted to the program. For quota-2 students in particular we include the GPA from high-school, the motivational average and an indicator (dummy) for whether or not the student comes from a Scandinavian country.

For each course, we run regressions of the following type:

$$\begin{aligned} \text{GRADE}_i = & \beta_0 + \beta_1 DQ_i + \beta_2 \text{GPA}_i * DQ_i + \beta_3 \text{MOTAVR}_i * DQ_i \\ & + \beta_4 \text{DSCAND}_i * DQ_i + \beta_5 \text{AGEADMIT}_i + \text{error}_i \end{aligned} \quad (1)$$

where *GRADE* is the grade for a student *i* who has participated in the exam under study; *DQ* is the dummy for admission quota, taking the value 1 if the student is admitted via quota-2, and 0 if the student is admitted via quota-1; *GPA* is the high-school grade-point average for a quota-2 applicant (this is why it is multiplied by *DQ* in the model equation); *MOTAVR* is the figure for the average given to the motivational essay of the student (multiplied by *DQ* for same reasons as above); *DSCAND* is a dummy for home country of the student (*DSCAND*=1 if the student comes from a Scandinavian country and 0 otherwise). Again, this background variable is of interest for quota-2 applicants only, and this is why it has been multiplied by *DQ* in the model equation; *AGEADMIT*

---

16. Some discretion may also enter in the admission process as, other things equal, country diversity matters to the program: the IB office does attempt to put together a diversified, international class.

is the students' age when admitted. The 'error' is a stochastic error term assumed to follow a normal distribution with zero mean and a constant variance.

The interpretation of the model is as follows:

- The expected grade of a student who has been admitted via quota-1 is:  
 $\beta_0 + \beta_5 * AGEADMIT$ ,  
 i.e. the intercept of the model plus the coefficient of »ageadmit« times the age of the student at the time of the admission. If this age is 20.3 years, we have:  $\beta_0 + \beta_5 * 20.3$ .
- Similarly, the expected grade of a non-Scandinavian quota-2 student is:  
 $\beta_0 + \beta_1 + \beta_2 * GPA + \beta_3 * MOTAVR + \beta_5 * AGEADMIT$   
 For example, if the student has a GPA = 8.46 (the average), has MOTAVR = 12.80 (the average) and is 23.78 years old when admitted, we have:  $\beta_0 + \beta_1 + \beta_2 * 8.46 + \beta_3 * 12.80 + \beta_5 * 23.78$ .
- The expected grade of a Scandinavian quota-2 student is:  
 $\beta_0 + \beta_1 + \beta_2 * GPA + \beta_3 * MOTAVR + \beta_4 + \beta_5 * AGEADMIT$   
 For example, if the student has GPA = 8.87 (the average), has MOTAVR = 12.80 and is 23.7 years old when admitted, we have:  $\beta_0 + \beta_1 + \beta_2 * 8.87 + \beta_3 * 12.8 + \beta_4 + \beta_5 * 23.7$

#### 4. The results

The results are presented in the following two tables; table 1 for the first year courses and table 2 for the second year courses.<sup>17</sup> Note that the tables present a parsimonious model for each course. The »-« indicate the variables that are omitted compared to equation (1) due to acceptance of the null hypothesis that the coefficients of these variables were simultaneously zero (F-test results are reported in the final rows of the tables).<sup>18</sup>

Two misspecification tests are performed: the White's general test for heteroscedasticity and the Jarque-Bera (JB) test for normality of the error term. In both cases a *p*-value smaller than 0.05 would indicate that the model was mis-specified. As can be seen, this is not the case.

Our model has some explanatory value for all but two courses, viz. Intercultural Communication (123) and International Business Law (241). For these two courses none of the variables (that we examined) have an effect on grades. In these two courses, the expected average grade is 9.17 and 8.66, respectively. For all other courses the

17. Instead of running regressions for each course (which reduces the number of observations available), we could have run one pooled regression that includes a dummy for each course. This would have increased the number of observations and presumably reduced the standard deviations and improved the fit. We decided, however, not to follow this procedure as that will remove the valuable information that one gets from running individual regressions – as will be discussed below, some variables work for some courses but not for others. Running one regression would blur that picture.

18. For completeness we report the estimation results of the full models based on equation (1) in appendix 2.

Table 1. Results of the estimation, 1<sup>st</sup> year courses.

Variable	Course 111	Course 112	Course 114E	Course 122E	Course 123	Course 131	Course 132E
Intercept term	10.61*** (0.48)	8.81*** (0.41)	8.17*** (0.27)	9.2*** (0.50)	9.17*** (0.08)	9.0*** (0.32)	9.8*** (0.59)
<i>DQ</i>	–	-1.83*** (0.54)	-2.84*** (0.66)	-3.57*** (0.72)	–	-3.47*** (0.85)	–
<i>GPA*DQ</i>	–	0.11** (0.05)	–	0.18*** (0.06)	–	0.09** (0.04)	–
<i>MOTAVR*DQ</i>	–	–	0.18*** (0.05)	–	–	0.15** (0.06)	–
<i>DSCAND*DQ</i>	–	–	–	–	–	–	–
<i>AGEADMIT</i>	-0.06*** (0.02)	–	–	–	–	–	-0.07*** (0.03)
Number of observations	145	120	146	76	–	114	116
<i>R</i> <sup>2</sup>	0.06***	0.09***	0.12***	0.26***	–	0.17***	0.06***
Whites hetero <i>p</i> -value	0.21 0.90	3.73 0.29	1.74 0.63	3.14 0.37	–	12.91 0.04	0.10 0.95
JB <i>p</i> -value	3.21 0.20	0.87 0.65	0.56 0.75	0.78 0.68	–	3.45 0.18	2.98 0.23
Over identifying restrictions <i>F</i>	1.74	0.24	0.86	0.26	–	2.03	0.25
<i>p</i> -value	0.15	0.87	0.46	0.85	–	0.14	0.91

Note: \* significant at the 10% level, \*\* significant at the 5% level, \*\*\* significant at the 1% level. Figures in parentheses are standard errors.

regressions have an explanatory power between 6% (The company in an international setting) and 39% (Business Statistics), which, as mentioned in the introduction, was expected.<sup>19</sup> What these numbers indicate is that the admission characteristics used can explain at most 39% of the variation in the grade in individual B.Sc.IB courses – the rest of the variation is explained by other factors that affect performance and learning in general.

Focusing on which variables turn out to be significant, it is seen that whether a student has been admitted through quota-1 or quota-2 affects the grade quite a lot in courses like Microeconomics, International Business Environment, Macroeconomics, International Economics, MIS, Corporate Finance and Business Statistics. Moreover,

19. In general, however, small R-square values are not unusual with this type of micro-data.

Table 2. Results of the estimation, 2<sup>nd</sup> year courses.

Variable	Course 212E	Course 222	Course 234	Course 241
Intercept term	8.00*** (0.49)	10.13** (0.59)	8.50*** (0.62)	8.66*** (0.21)
<i>DQ</i>	-1.31** (0.62)	-2.59*** (0.79)	-3.60*** (0.82)	–
<i>GPA*DQ</i>	0.15*** (0.05)	0.18** (0.07)	0.15** (0.06)	–
<i>MOTAVR*DQ</i>	–	–	–	–
<i>DSCAND*DQ</i>	–	–	2.00*** (0.61)	–
<i>AGEADMIT</i>	–	–	–	–
Number of observations	51	31	46	–
$R^2$	0.17**	0.30***	0.39***	–
Whites test	2.06	2.50	4.53	–
<i>p</i> -value	0.56	0.48	0.48	–
JB test	1.85	1.09	7.69	–
<i>p</i> -value	0.40	0.58	0.02**	–
Overidentifying Restrictions <i>F</i>	1.68	1.20	0.85	–
<i>p</i> -value	0.18	0.33	0.44	–

Note: \* significant at the 10% level, \*\* significant at the 5% level, \*\*\* significant at the 1% level. Figures in parentheses are standard errors.

the effect is negative, i.e. compared to quota-1 student a quota-2 student performs comparatively less well in these courses. The age variable turns out to be significant only in two courses (The Company in an International setting and Principles of Marketing) and in both of these courses it enters with a negative effect on grades, i.e. older students perform worse than younger students.<sup>20</sup>

Looking now at what is important for the performance of the quota-2 students, we notice that a good *GPA* upon admission has a positive effect on grades in all courses,

20. In other studies (Albæk (2001)) a variable like *AGEADMIT* enters the model equation non-linearly. Whether our equations are subject to misspecification of this kind is tested generally by Ramsey's RESET test. Only in one case does this test point toward a functional misspecification of the equation at the 5% level. At the 1% level no misspecification is detected hence we do not feel a need to include non-linear terms in the models.



Table 3. Expected grades for HA-IB students.

Course	Quota-1 student	Quota-2/Scand.	Quota-2/Non-Scand.
1. YEAR			
The Company in an International Setting (111)	9.39	9.18	9.18
Microeconomics (112)	8.81	7.64	7.64
International Business Environment (144E)	8.17	7.62	7.62
Macroeconomics (122E)	9.20	6.80	6.80
Intercultural Communication (123)	9.17	9.17	9.17
International Economics (131)	9.00	8.01	8.01
Principles of Marketing (132E)	8.37	8.19	8.19
2. YEAR			
MIS (212E)	8.00	7.62	7.62
Corporate finance (222)	10.13	8.64	8.64
Statistics for Business (234)	8.50	7.80	5.80
International Business law (241)	8.66	8.66	8.66

while a good motivational essay improves the students' performance only in two courses (International Business Environment and International Economics)<sup>21</sup> – for the rest of the courses the variable is not statistically significant. Finally, whether the student is Scandinavian or not, matters only in Statistics for Business, where non-Scandinavians perform considerably worse (some further analysis of this result is presented later on).

In order to provide some quantitative feeling of the above results, we can use the formulas described in the interpretation of the model (see p. 8), and we can calculate the expected (average) grades in the different courses. The results are described in table 3.

Consistent with the previous presentation of the results, a quota-1 vs. quota-2 admission most evidently affects the performance in Microeconomics, International Business Environment, Macroeconomics, International Economics, MIS, Corporate Finance and Business Statistics. For two of the courses, Intercultural Communications and International Business Law, we find no significant differences what so ever. Notice that for some courses (111, 132E) the difference in performance stems exclusively from the average age differences that can be observed amongst quota-1 (20.3 years of age) and quota-2 students (23.78 years of age). The Scandinavian vs. Non-Scandinavian characteristic seems to affect only (but significantly) the expected average grade in Statistics for Business.

### 5. Sensitivity analysis

In order to determine the robustness of our conclusions, and to see whether we can raise the explanatory power of our models we extended our work in three directions.

21. The coefficient of correlation between *GPA* and *MOTAVR* for quota-2 students is 0.144. Hence there is no indication in the data that these two variables measure »the same characteristic« of the students.

Table 4. Results of testing down to the models of table 1 & 2 from the extended models.

Course	F-stat for over identifying restriction towards the models found in table 1 or 2	p-value of F-test	Competing model – if any
111	0.94	0.50	
112	0.79	0.61	
114E	1.10	0.37	
122E	0.51	0.84	
123	1.09	0.37	
131	2.38	0.03**	Extended by Eduhome
132E	0.64	0.72	DQ2, GPA*DQ2 Livabr*DQ2
212E	0.77	0.63	
222	0.76	0.64	
234	0.44	0.87	
241	0.79	0.64	

Note: \*\* significant at the 5% level.

First, we extend our initial model with data for the additional point-giving activities of a quota-2 student, in order to see whether we were right in omitting these variables in the first place. Our additional variables are: *Eduhome*: months spent on additional education in the home country; *Eduout*: months spent on additional education abroad; *Workhome*: months spent on work in the home country; *Workout*: months spent on work abroad; *Livabr*: months spent living abroad. As all of these variables are relevant for quota-2 students only, they enter the regression equation multiplied by *DQ* (the quota dummy) just like the *GPA* variable, the motivational average and the *DSCAND* dummy.

The result of this exercise can be found in table 4. Instead of presenting all the regressions with the new coefficient estimates,<sup>22</sup> table 4 presents the results of testing whether the simple parsimonious models of tables 1 and 2 are still appropriate. Again, if the column showing the *p*-value has values larger than 0.05 this hypothesis is not rejected. The general conclusion from this exercise is that not much is changed. For a couple of models, however, competing equations with one or two of these variables arise. These competing models are also appropriate compared to the extended model when tested by an *F*-test. Hence the suggestion that important explanatory variables were missing from our original equation is, in general, not supported by this sensitivity analysis.

Secondly, inspired by the result in Albæk (2001) and Andersen and Østergaard (2005) that the high-school grade and level in Mathematics are important explanatory factors for performance, we manually collected such data for as many of the students in

22. Appendix 3 present in detail all the regression and the derived estimates.

Table 5. Results of the estimation, 1<sup>st</sup> year courses with math-variables.

Variable	Course 111	Course 112	Course 114E	Course 122E	Course 123	Course 131	Course 132E
Intercept term	7.88***	1.69	8.13***	-0.07	7.82***	6.89***	8.24***
<i>DQ</i>	–	–	-2.52***	–	–	-3.78***	–
<i>GPA*DQ</i>	–	–	–	–	–	–	–
<i>MOTAVR*DQ</i>	–	–	0.17***	–	–	0.26***	–
<i>DSCAND*DQ</i>	–	–	–	–	–	–	–
<i>AGEADMIT</i>	–	–	–	–	–	–	–
<i>MGRADE</i>	0.17***	0.67***	–	0.74***	0.15**	0.21(*)	–
<i>ALEVEL</i>	–	–	–	1.74***	–	–	–
Number of observations	79	71	80	52	80	69	82
<i>R</i> <sup>2</sup>	0.08***	0.21***	0.12***	0.37***	0.07**	0.23***	–
<i>OLD</i>	–	–	+	+	–	–	–

Note: \*\* significant at the 5% level. \*\*\* significant at the 1% level.

our original data set as possible.<sup>23</sup> It turned out that this was not an easy task: the data set we ended up with contains a reduced number of observations compared to the original data set. In addition, the data set is biased in the sense that it was easier to get the math-grade data for Danish students than for the foreign students – in the new data set, about two thirds of the students are Danish, while only one third are foreign.<sup>24</sup>

In line with the results in Albæk (2001) and Andersen & Østergaard (2005), we decided not to combine the math characteristics with the quota-1/2 variable. The procedure of the analysis was then as follows: First, for each of the courses we run the original (small) model of table 1 and table 2 to see whether it is still valid for the now smaller data set. Next, we include both the math grade (*MGRADE*) and a dummy for the math level (*ALEVEL* = 1 if math has been taken at the A-level and 0 otherwise) as additional explanatory variables and again reduced the model when testing of sets of zero restrictions allow us to do so. The results of this procedure are found in tables 5 and 6.

For the first year courses the reduction of the size of the data set means that only in two courses (114, International Business Environment and 122, Macroeconomics) were we able to obtain models with the same explanatory variables as before – for all other courses one or more of the original explanatory variables have become insignifi-

23. Note that the IB's minimum admission requirement is mathematics at high-school level B.

24. Due to these problems, we have chosen to report the results of the analysis in the sensitivity section of this paper.

Table 6. Results of the estimation, 2<sup>nd</sup> year courses with math-variables.

Variable	Course 212E	Course 222	Course 234	Course 241
Intercept term	4.53***	9.62***	8.50***	8.71***
<i>DQ</i>	–	–	-5.39***	–
<i>GPA*DQ</i>	–	–	0.27**	–
<i>MOTAVR*DQ</i>	–	–	–	–
<i>DSCAND*DQ</i>	–	–	2.77***	–
<i>AGEADMIT</i>	–	–	–	–
<i>MGRADE</i>	0.36*	–	–	–
<i>ALEVEL</i>	–	–	–	–
Number of observations	32	21	31	–
<i>R</i> <sup>2</sup>	0.12*	–	0.35***	–
<i>OLD</i>	–	–	+	+

Note: *OLD*: a plus means that a model with similar characteristics as the one in table 1 still applies to the smaller data set. A minus means that this is not the case. \* significant at the 10% level, \*\* significant at the 5% level, \*\*\* significant at the 1% level.

cant. In most cases we have been able to find new models and very often the math grade enters the new equations significantly and with a positive sign. Also the explanatory power is increased for all courses. Furthermore, it seems that the importance of the *GPA* and the quota-variable has been reduced: only in two of the new models is the quota-variable still significant. For the second year courses the inclusion of the math grade data does not seem to add much – actually, very few observations are left for analysis here. However, due to the reduced number of observations and the bias they introduce, we should be a bit careful in our interpretations. We shall therefore tentatively conclude that the math grade is an important factor to include in future models on these issues.

The last extension that we investigate here has to do with something that we mentioned in the very beginning of our paper, viz. that the students' performance may improve considerably in a good learning environment. Thus, and in addition to our primary goal of determining the importance of student background characteristics to their performance, we test here whether the first year's *GPA* has any explanatory value for the second year performance. To investigate this hypothesis on our data, we ran some additional regressions for the 2<sup>nd</sup> year courses to see if adding the 1<sup>st</sup> year *GPA* to

the regressions (1YAVR) would be of any value.<sup>25</sup> It turns out, that for 3 out of the 4 courses this is in fact the case:

- For Management Information Systems (212) the  $R^2$  increases from 17% to 43% and the inclusion of 1YAVR makes both the quota- and *GPA*-variables unimportant.
- For Corporate Finance (222) the  $R^2$  is increased from 30% to 65% and again quota- and *GPA*-variables become unimportant.
- For Statistics for Business (234) the  $R^2$  increases from 39% to 76% while both 1YAVR and the original explanatory variables (quota, *GPA* and being Scandinavian) are significant in the model.
- For International Business Law (241) we still do not find any significant explanatory variable, although we do find  $R^2 = 10\%$  and 1YAVR significant at the 10% level in a model based on only that explanatory variable.

Thus, the results indicate that the first year *GPA* is an important variable in explaining the students' performance in the second year. Furthermore, and as perhaps it was expected, the first year *GPA* turns out to be correlated with the admission *GPA* and »quota« characteristics, rendering these variables insignificant in some courses. These additional regressions seem promising for the hypothesis of progression, even though we have to keep in mind that the data series available for these regressions only contain a limited number of observations.

### 5. Concluding remarks

The overall picture that emerges from the above analysis is clear on two issues:

- (i) a quota-2 student is generally performs worse than a quota-1 student; and
- (ii) a quota-2 student performs better the higher her/his entry-*GPA* is.

Thus, the study has by and large confirmed our prior expectations concerning the importance of the *GPA* variable for performance of the quota-2 students at the B.Sc.IB programme at CBS. Also the prior belief that the admission group of the student was important has been confirmed. The size of the effects varies from course to course but the direction is the same: the influence of »hard« admission characteristics is strong, while the influence of »soft« admission characteristics matters only for a couple of courses.

---

25. I.e. the parsimonious equations of table 2 are in each case extended by the additional explanatory variable 1YAVR and for e.g. the course 212 the equation will be:  $GRADE_i = \beta_0 + \beta_1 DQ_i + \beta_2 GPA_i * DQ_i + \beta_3 1YAVR_i + error_i$ .

Compared to the existing literature, our results do not, to the same extent, confirm the importance of mathematics grades and levels, and we do not confirm any positive effect of delaying entering the Bachelor education (if nothing, we find a negative relation).<sup>26</sup> Clearly this is not surprising if one considers the nature of the different programs (the IB program is definitely not a comparable program to the economics program at university).

Concluding, we admit that for some of our models the explanatory power – although significant – is low. To reinforce the results of the study it will therefore be desirable to obtain more observations in order to decrease the standard errors and possibly increase the significance of the explanatory variables. Nevertheless, we believe that our present study is quite unique for Danish universities, as it identifies a number of novel hypotheses and confronts them with the data. We hope that such analysis of the data would become standard for those with responsibilities for the admission selection and the design of selection criteria in universities.

---

26. This, however, may not be a contradiction, since Albæk (2001) investigates the determinants of passing the first year, and not the students' performance in their studies. In other words, both Albæk's and our results can be true simultaneously.

#### References

- Albæk, K. 2001. Hvem består på politstudiet?, *Nationaløkonomisk Tidsskrift* 139, 208-33 (in Danish).
- Andersen, S. and P. Østergaard. 2005. Fra elev i gymnasiet til studerende ved en videregående uddannelse – erfaringer med præstationerne i statistik fra uddannelserne ved Handelshøjskolen i Århus, i *Symposium i Anvendt Statistik 2005*, Økonomisk Institut, Syddansk Universitet og Danmarks Statistik, 397-421 (in Danish).
- Blundel, R., L. Dearden, A. Goodman, and H. Reed. 1997. *Higher Education, Employment and Earnings in Britain*, London: Institute of Fiscal Studies.
- Gabrielsen, G. 1996a. Evalueringsrapport I – lærerevaluering, *Internal working paper* 16/1, Department of Mathematical Business Economics and Statistics, CBS (in Danish).
- Gabrielsen, G. 1996b. Evalueringsrapport I – semesterenquêter, *Internal working paper* 16/2, Department of Mathematical Business Economics and Statistics, CBS (in Danish).
- Lorentzen, J. 2001. In the Ditch: English-Language Masters Programmes at CBS, mimeo, Department of International Economics and Management, CBS.
- Naylor, R. and J. Smith. 2004. Degree performance of economics students in UK universities: absolute and relative performance in prior qualifications, *Scottish Journal of Political Economy* 51, 250-65.
- Politiken. 2004. Real Businessclass. *Politiken* 05.09.2004, 6. section, p. 31.
- Smith, J., A. McKnight, and R. Naylor. 2000. Graduate employability: policy and performance in higher education in the UK. *Economic Journal* 110, F382-411.
- Tay, R. S. 1994. Students' performance in economics: does the norm hold across cultural and institutional settings? *The Journal of Economic Education* 25, 291-301.

### Appendix 1

Until the 2005/2006 study year, »quota-2« applicants were assessed according to the following Point System:

<i>Grade Point Average</i>	8.1-8.4	3 points
	8.5-8.9	6 points
	9.0-9.4	9 points
	9.5-over	12 points
<i>Other education and post-secondary courses</i> (min. 3 months – max. 1 academic year per category)	Point per month or course:	
	In home country:	0.5
	Abroad:	0.5
<i>Work experience</i> (min. 3 months – max. 24 months per category)	Point per month:	
	In home country:	0.5
	Abroad:	1
<i>Living abroad</i> (min. 3 months – max. 24 months)	0.5 point per month	
<i>Extra-curricular activities</i> (max. one page)	2-3 points per personal qualification	
	3-5 points per professional qualification	
	(max. 16 points)	
<i>Motivational essay</i> (max. two pages)	Possible points:	0-20

Work experience, time spent abroad, and post-secondary courses can only be included in the assessment if the activity:

- took place after the applicant turned 15 years old
- occurred within 5 years prior to 1 July in the year of application
- is included in detail on the application form
- is documented

*Post-secondary courses* Has to be full-time study to count as months, otherwise it will count as courses. Courses must be successfully passed.

*Work experience:* Only work experience with more than 15 hours per week will be recognised. Part-time work will be recalculated to full-time work (30 hours per week).

*Extra-curricular activities:* Any activities you have participated in that you feel gave you particular qualifications. Such as student organisations, volunteer work in political parties, grass root's organisations, etc. The special qualifications that were needed or gained should be stated, e.g. independence, leadership, maturity, cooperation, the practical use of a foreign language, first-hand knowledge gained of another culture.

*Motivational essay:* These pages should allow CBS to get to know the person in a different way from objective data. The essay should tell us more about who you are and what you value.

**Appendix 2**

Estimation results of the full models given in equation (1).

*Table A2.1. Results of the estimation, 1<sup>st</sup> year courses.*

Variable	Course 111	Course 112	Course 114E	Course 122E	Course 123	Course 131	Course 132E
Intercept term	10.90*** (0.56)	10.00*** (1.55)	8.60*** (0.83)	9.35*** (2.10)	10.49*** (0.57)	9.06*** (1.20)	9.63*** (0.68)
<i>DQ</i>	0.52 (0.50)	-1.13 (1.23)	-2.76*** (0.73)	-4.61** (1.77)	-0.29 (0.49)	-3.44*** (0.94)	-0.19 (0.54)
<i>GPA*DQ</i>	-0.00 (0.02)	0.10** (0.05)	-0.00 (0.03)	0.17** (0.07)	-0.01 (0.02)	0.08** (0.04)	0.02 (0.02)
<i>MOTAVR*DQ</i>	-0.06* (0.03)	-0.03 (0.08)	0.17*** (0.05)	0.09 (0.11)	0.03 (0.03)	0.13** (0.06)	-0.00 (0.03)
<i>DSCAND*DQ</i>	-0.15 (0.22)	-0.23 (0.51)	0.33 (0.32)	-0.03 (0.74)	-0.23 (0.21)	0.72 (0.39)	-0.02 (0.23)
<i>AGEADMIT</i>	-0.06** (0.03)	-0.06 (0.07)	-0.02 (0.04)	-0.01 (0.10)	-0.05** (0.03)	-0.00 (0.06)	-0.06* (0.03)
Number of observations	145	120	146	76	142	114	116
<i>R</i> <sup>2</sup>	0.10**	0.10**	0.13***	0.26***	0.05	0.20***	0.07

*Note:* \* significant at the 10% level, \*\* significant at the 5% level, \*\*\* significant at the 1% level. Figures in parentheses are standard errors.



*Table A2.2. Results of the estimation, 2<sup>nd</sup> year courses.*

Variable	Course 212E	Course 222	Course 234	Course 241
Intercept term	11.99*** (1.86)	14.06*** (2.42)	11.14*** (2.41)	10.10*** (2.17)
<i>DQ</i>	-0.33 (2.00)	-1.34 (2.46)	-4.32** (2.02)	0.34 (1.94)
<i>GPA*DQ</i>	0.12** (0.06)	0.11 (0.08)	0.10 (0.07)	0.07 (0.05)
<i>MOTAVR*DQ</i>	0.01 (0.13)	-0.01 (0.16)	0.10 (0.13)	-0.05 (0.13)
<i>DSCAND*DQ</i>	-0.59 (0.52)	0.07 (0.88)	1.75** (0.66)	-0.14 (0.53)
<i>AGEADMIT</i>	-0.20 (0.09)**	-0.20 (0.12)	-0.13 (0.12)	-0.07 (0.10)
Number of observations	51	31	46	41
<i>R</i> <sup>2</sup>	0.26**	0.40**	0.42***	0.07

*Note:* \* significant at the 10% level, \*\* significant at the 5% level, \*\*\* significant at the 1% level. Figures in parentheses are standard errors.

### Appendix 3

Estimation results of equation (1) with *EDUHOME*, *EDUOUT*, *WORKHOME*, *WORKOUT*, and *LIVABR* as additional regressors.

Table A3.1. Results of the estimation, 1<sup>st</sup> year courses.

Variable	Course 111	Course 112	Course 114E	Course 122E	Course 123	Course 131	Course 132E
Intercept term	10.51*** (0.66)	9.97*** (1.92)	8.78*** (0.98)	8.35*** (2.78)	10.73*** (0.68)	9.77*** (1.35)	9.06*** (0.76)
<i>DQ</i>	0.92 (0.59)	-0.62 (1.59)	-2.79*** (0.85)	-4.66** (2.18)	-0.39 (0.59)	-2.97** (1.16)	-0.33 (0.67)
<i>GPA*DQ</i>	-0.01 (0.02)	0.08 (0.06)	-0.01 (0.03)	0.16* (0.08)	-0.00 (0.02)	0.07* (0.04)	0.03 (0.02)
<i>MOTAVR*DQ</i>	-0.06* (0.03)	-0.04 (0.08)	0.15*** (0.05)	0.13 (0.12)	0.02 (0.03)	0.11** (0.06)	-0.01 (0.03)
<i>DSCAND*DQ</i>	-0.19 (0.22)	-0.30 (0.56)	0.33 (0.33)	-0.50 (0.84)	-0.28 (0.22)	0.56 (0.40)	-0.05 (0.24)
<i>AGEADMIT</i>	-0.04 (0.03)	-0.06 (0.09)	-0.03 (0.05)	0.04 (0.14)	-0.07** (0.03)	-0.04 (0.06)	-0.03 (0.04)
<i>EDUOUT</i>	-0.06 (0.05)	0.09 (0.16)	-0.08 (0.08)	-0.38 (0.24)	-0.02 (0.05)	-0.03 (0.12)	0.06 (0.06)
<i>EDUHOME</i>	-0.03 (0.05)	-0.12 (0.12)	0.03 (0.07)	-0.09 (0.21)	-0.05 (0.04)	-0.17* (0.09)	-0.05 (0.05)
<i>WORKOUT</i>	-0.01 (0.01)	-0.03 (0.04)	-0.01 (0.02)	-0.03 (0.06)	0.01 (0.01)	-0.02 (0.03)	-0.01 (0.02)
<i>WORKHOME</i>	-0.02 (0.03)	0.05 (0.08)	0.03 (0.04)	0.00 (0.11)	0.04 (0.03)	0.08 (0.06)	0.00 (0.03)
<i>LIVABR</i>	-0.01 (0.02)	-0.02 (0.06)	0.05 (0.03)	0.04 (0.08)	0.01 (0.02)	-0.00 (0.04)	0.04 (0.03)
Number of observations	141	118	143	75	140	112	113
<i>R</i> <sup>2</sup>	0.12*	0.14*	0.17***	0.30***	0.08	0.28***	0.13

Note: \* significant at the 10% level, \*\* significant at the 5% level, \*\*\* significant at the 1% level. Figures in parentheses are standard errors.

*Table A3.2. Results of the estimation, 2<sup>nd</sup> year courses.*

Variable	Course 212E	Course 222	Course 234	Course 241
Intercept term	11.10*** (2.77)	15.96*** (3.81)	12.38*** (3.47)	13.31*** (3.11)
<i>DQ</i>	0.82 (3.09)	-3.64 (3.98)	-5.64* (3.09)	-2.06 (2.63)
<i>GPA*DQ</i>	0.11 (0.07)	0.14 (0.11)	0.12 (0.09)	0.11 (0.07)
<i>MOTAVR*DQ</i>	-0.01 (0.17)	0.06 (0.22)	0.13 (0.16)	-0.01 (0.14)
<i>DSCAND*DQ</i>	-0.92 (0.63)	-0.29 (1.05)	1.75** (0.77)	-0.16 (0.60)
<i>AGEADMIT</i>	-0.16 (0.14)	-0.29 (0.19)	-0.19 (0.17)	-0.23 (0.15)
<i>EDUOUT</i>	-0.11 (0.20)	0.05 (0.30)	0.05 (0.25)	0.41 (0.26)
<i>EDUHOME</i>	-0.13 (0.15)	-0.13 (0.33)	-0.07 (0.22)	0.06 (0.16)
<i>WORKOUT</i>	-0.01 (0.04)	0.04 (0.06)	0.01 (0.06)	0.04 (0.04)
<i>WORKHOME</i>	-0.04 (0.10)	0.19 (0.14)	0.11 (0.12)	0.16 (0.10)
<i>LIVABR</i>	-0.04 (0.07)	0.04 (0.11)	0.06 (0.09)	0.06 (0.08)
Number of observations	50	31	45	40
<i>R</i> <sup>2</sup>	0.27	0.46	0.43**	0.21

*Note:* \* significant at the 10% level, \*\* significant at the 5% level, \*\*\* significant at the 1% level. Figures in parentheses are standard errors.

# Skattekonkurrence og økonomisk geografi

Niels Johannesen

E-Mail: nielsjohannesen@gmail.com

*SUMMARY: Whilst simple theoretical models of tax competition suggest that international capital mobility reduces the capital tax rate in equilibrium, empirical studies conducted in the field have produced ambiguous results. This paper argues that the puzzling empirical findings may be due to an omitted variables bias and mitigates the bias by estimating a comprehensive model for capital taxation controlling for tax export effects, productivity effects of public goods as well as agglomeration effects. The estimated impact of increased capital mobility in the OECD-countries between 1975 and 2000 was a reduction of average capital tax rates of around 10 percentage points. In the same period, however, decreasing trade barriers in the goods markets boosted economic rents in capital agglomerations triggering an increase in capital tax rates of a similar magnitude. In brief, this paper finds strong evidence of tax competition between OECD-countries and demonstrates that its effects have so far been neutralized by spill-over effects from the goods markets. The policy implications of these results are remarkable. Since agglomeration rents can be shown to be a hump-shaped function of trade costs, further integration of goods markets will arguably reinforce the effects of tax competition rather than counterbalancing these effects as hitherto observed. Consequently, there would be an increased need for international policy coordination in the field of capital taxation.*

---

## 1. Indledning

Helt centralt i den teoretiske litteratur om skattekonkurrence står påvisningen af en negativ sammenhæng mellem international kapitalmobilitet og kapitalbeskatning: I tilstedeværelsen af kapitalmobilitet er kapitalbeskatning forbundet med en omkostning i form af kapitaludstrømning, og dette resulterer i et nedadgående pres på kapital-skattegraden. Argumentet synes overbevisende, men ikke desto mindre har det vist sig vanskeligt at etablere en sådan sammenhæng empirisk. Adskillige studier finder i modstrid med den teoretiske skattekonkurrencemodell en positiv empirisk sammenhæng mellem kapitalmobilitet og kapitalbeskatning.

I denne artikel søges en forklaring på uoverensstemmelsen mellem litteraturens teoretiske og empiriske resultater: I et teoretisk afsnit identificeres yderligere tre faktorer,

som potentielt påvirker den optimale kapitalbeskatning: (i) skatteeksport, (ii) produktivitetseffekter af offentlige goder og (iii) internationale handelsomkostninger. Dernæst etablerer et empirisk afsnit en negativ sammenhæng mellem kapitalmobilitet og kapitalbeskatning, idet der kontrolleres for (i)-(iii). Resultaterne implicerer, at den gennemsnitlige kapitalskatterate i OECD-landene i de sidste årtier alt andet lige er reduceret med omkring 10 procentpoint som følge af stigende kapitalmobilitet. Dette er imidlertid blevet mere end opvejet af faldende handelsomkostninger, der har øget den økonomiske rente (»economic rent«) forbundet med produktion i kapitalklynger og dermed muliggjort en højere kapitalbeskatning. Ved at inkludere faktorerne (i)-(iii) i modellen for kapitalbeskatning synes nærværende studie således at eliminere en *omitted variables bias*, der formentlig er årsag til de kontra-intuitive resultater i tidligere empiriske studier.

Resultaterne har særdeles vigtige policy-implikationer. Mens den empiriske model peger på, at faldende handelsomkostninger hidtil har modvirket tendensen til skattekonkurrence, så forudsiger den teoretiske økonomisk geografi model (ØG), at fald i handelsomkostningerne under et kritisk niveau i stedet vil forstærke denne tendens. Dette tyder på, at kapitalmarkedsintegration og varemarkedsintegration vil komme til at udøve et fælles nedadgående pres på kapitalbeskatningen, og at der følgelig vil opstå et langt større behov for international koordinering af skattepolitikken end tilfældet er i dag.

Artiklen leverer yderligere en række bidrag til litteraturen om skattekonkurrence: Til brug for analysen af handelsomkostningerne præsenteres en forenklet ØG-model for det internationale kapitalmarked, som tillader en eksplicit analytisk løsning for økonomiens opbruds- og bevaringspunkter. Modellen muliggør desuden, at velfærdsimplikationerne af skattekonkurrence sammenlignes direkte med resultaterne fra den sædvanlige neoklassiske kapitalmarkedsmodel. Med henblik på den empiriske analyse udvikles et avanceret mål for internationale handelsomkostninger, og det lykkes med udgangspunkt i modellen for kapitalbeskatning at estimere et udtryk for OECD-økonomiernes opbrudspunkt.

Artiklen har følgende struktur: Afsnit 2 præsenterer den traditionelle skattekonkurrencemodell og beskriver kort betydningen af skatteeksport og produktivitetseffekter af offentlige goder. Afsnit 3 forklarer de vigtigste mekanismer i ØG-modellen og analyserer skattekonkurrence i denne. Afsnit 4 præsenterer de fundne resultater og afsnit 5 konkluderer.

## 2. Skattekonkurrence i neoklassisk kapitalmarkedsmodel

I dette afsnit præsenteres først en simpel skattekonkurrencemodell baseret på Eijffinger og Wagner (2001). Dernæst skitseres betydningen af henholdsvis skatteeksport

og produktivitetseffekter af offentlige goder, og endelig udledes tre testbare hypoteser med henblik på den senere empiriske analyse.

Der betragtes en lille, åben økonomi, hvor den repræsentative forbruger er udstyret med en vis mængde kapital  $s$ . Kapital er den eneste variable produktionsfaktor, og økonomiens aggregerede produktion er dermed givet ved  $f(n)$ , hvor  $n$  er den anvendte mængde kapital. Kapitalmarkedet antages at være kompetitivt, så kapital i ligevægt aflønnes med sit marginalprodukt  $f'(n)$ , og en arbitragebetingelse for det internationale kapitalmarked kan således opskrives som:

$$f'(n) - t = r^* \quad (1.1)$$

hvor  $t$  er den indenlandske kapitalskatterate, og  $r^*$  er det eksogent givne nettoafkast af kapital på verdensmarkedet.

Kapital beskattes ved indkomstkilden, og provenuet anvendes til offentligt forbrug. Det antages, at regeringen er godgørende (»benevolent«) og sætter den for forbrugeren optimale skatterate, idet forbrugerens præferencer over privat og offentligt forbrug beskrives ved følgende kvasilineære nyttefunktion:

$$U(y, g) = y + \theta(g) \quad (1.2)$$

hvor  $y$  er privat forbrug,  $g$  er offentligt forbrug, og  $U(\cdot)$  og  $\theta(\cdot)$  er strengt konkave nyttefunktioner. Idet regeringen antages omkostningsfrit at kunne transformere en enhed af det private gode  $y$  til en enhed af det offentlige gode  $g$ , får regeringens problem formen:

$$\max_t U(y, g)$$

under bibetingelserne:

$$g \leq t \cdot n \quad (1.3)$$

$$y \leq f(n) - f'(n) \cdot n + r^* \cdot s \quad (1.4)$$

Bibetingelsen (1.3) siger, at det offentlige forbrug ikke må overstige skatteprovenuet og kan altså forstås som et krav om budgetbalance. Tilsvarende siger (1.4), at det private forbrug ikke må være større end den disponible indkomst bestående af virksomhedsprofitter og kapitalindkomst.

Eijffinger og Wagner (2001) viser, at regeringens problem har følgende førsteordensbetingelse:

$$\theta'(g^*) = \frac{1}{1 + \varepsilon} \quad (1.5)$$

idet den optimale mængde af det offentlige gode kaldes  $g^*$ , og  $\varepsilon$  er elasticiteten af det indenlandske kapitaludbud  $n$  med hensyn til skatteraten  $t$ . Mens ligningens venstre side er et udtryk for forbrugers marginales substitutionsforhold mellem offentligt og privat forbrug, kan ligningens højreside fortolkes som den effektive marginalomkostning ved offentligt forbrug målt i enheder af privat forbrug. Da  $\varepsilon$  er entydigt negativ, må  $g^*$  findes, hvor marginalnyttens af offentligt forbrug er større end en.

Heraf fremgår det, at den godgørende regering fra et globalt synspunkt ikke sikrer en efficient allokering af ressourcer mellem den private og den offentlige sektor. Idet marginalnyttens af offentligt forbrug overstiger en, vil en marginal overførsel af ressourcer fra den private til den offentlige sektor alt andet lige udgøre en velfærdsforbedring for forbrugeren, og følgelig er  $g^*$  mindre end den paretooptimale mængde. Dette resultat kan forekomme overraskende i betragtning af, at regeringen er godgørende og dermed per definition sætter den for forbrugeren optimale skatterate. Forklaringen er, at kapitalbeskatning i den åbne økonomi er forbundet med en positiv international eksternalitet. Når et land hæver sin kapitalskatterate giver det anledning til en udstrømning af kapital, som er skadelig for landets egne borgere, men som gavner de lande, der modtager kapitalstrømmene. Idet regeringerne kun tager højde for deres egne borgers velfærd, får disse internationale effekter af kapitalbeskatning karakter af en positiv eksternalitet, og skatteraten bliver i ligevægt lavere end det globalt set optimale.

Ovenstående model illustrerer det traditionelle skattekonkurrenceargument. Visse forfattere har imidlertid identificeret mekanismer, der modvirker tendensen til skattekonkurrence.<sup>1</sup> Huizinga og Nielsen (1997) viser, at når en del af økonomiens virksomheder ejes af udlændinge, opstår et fænomen kaldet skatteeksport, som trækker i retning af en inefficent høj kapitalbeskatning. I den lille, åbne økonomi bæres skattebyrden nemlig ikke af kapitalejerne, som aflønnes med det eksogent givne verdensmarkedsnettoafkast  $r^*$ , men af virksomhedsejerne, som modtager virksomhedernes rene profitter. I tilstedeværelsen af udenlandsk ejerskab er kapitalbeskatning således forbundet med en negativ international eksternalitet, fordi de udenlandske ejere bærer en del af omkostningen ved beskatningen. Idet det ikke a priori kan bestemmes, om den positive skattebaseeksternalitet eller den negative skatteeksporteksternalitet do-

1. Se Johannesen (2004) for en formel behandling af disse forfatteres bidrag i en konsistent modelramme.

minerer, kan kapitalskatteraten i ligevægt være enten højere eller lavere end det socialt efficiente.

Wooders, Zissimos og Dhillon (2001) analyserer skattekonkurrence under antagelse af, at offentligt forbrug øger produktiviteten i den private sektor, og finder, at den godgørende regering under visse omstændigheder vælger en inefficent høj kapitalbeskatning. Dette kan indses ved at betragte kapitalmarkedsligevægten (1.1) for en given værdi af  $n$ . Som sædvanlig vil øget kapitalbeskatning – et højere  $t$  – alt andet lige føre til et lavere nettoafkast af kapital, men dette modvirkes nu af, at det øgede offentlige forbrug har produktivitetseffekter i den private sektor – et højere  $f'(n)$ . Den samlede effekt på kapitalens nettoafkast  $f'(n) - t$  af en stigning i kapitalbeskatningen er således ubestemt. Hvis det offentlige forbrugs produktivitetseffekter i den private sektor er tilstrækkelig stærke, vil øget kapitalbeskatning derfor – i modsætning til det traditionelle skattekonkurrenceargument – give anledning til kapitalindstrømning fra resten af verden. I sådanne tilfælde vil kapitalbeskatningen være forbundet med en negativ international eksternalitet og kapitalskatteraten være højere end det socialt efficiente.

Analysen af skattekonkurrence i en neoklassisk kapitalmarkedsmodel kan således opsummeres i følgende ikke-entydige normative konklusion:

- *N1: I tilstedeværelsen af international kapitalmobilitet, udenlandsk ejerskab og produktivitetseffekter af offentligt forbrug, kan beskatningen af kapital være højere eller lavere end det socialt efficiente niveau.*

Til brug for den empiriske analyse i afsnit 4, kan der desuden med udgangspunkt i den præsenterede teori opstilles tre testbare hypoteser:

- *H1: Øget kapitalmobilitet medfører lavere kapitalskatterater*
- *H2: Mere udenlandsk ejerskab medfører højere kapitalskatterater*
- *H3: Større produktivitetseffekter af offentligt forbrug medfører højere kapitalskatterater.*

### **3. Skattekonkurrence og økonomisk geografi**

I denne del analyseres skattekonkurrence i en udvidet kapitalmarkedsmodel inspireret af den økonomiske geografi. Afsnit 3.1 præsenterer den udvidede kapitalmarkedsmodel, som først og fremmest adskiller sig fra den hidtil benyttede neoklassiske model ved at tillade kapitalklynger. Afsnit 3.2 og 3.3 analyserer skattekonkurrence i kapitalmarkedsligevægte henholdsvis med og uden klynger.



### 3.1 Kapitalmarkedsmodel

Den økonomiske geografi (ØG) omfatter en stor gruppe af modeller, som beskriver den geografiske placering af mobile produktionsfaktorer. Uanset om der fokuseres på migrationer af faglærte arbejdere som i Krugman (1991) eller på internationale kapitalbevægelser som i Robert-Nicoud (2002), er det karakteristisk for ØG-modellerne, at allokeringen af de mobile produktionsfaktorer bestemmes af styrkeforholdet mellem to modsatte kræfter: *spredningskræfter*, som trækker i retning af en geografisk spredning af produktionsfaktorerne og *samlingskræfter*, der tenderer mod at samle produktionsfaktorerne i en klynge. Når samlingskræfterne dominerer, kan der således opstå ligevægte, hvor produktionsfaktorerne er stærkt ulige fordelt mellem i øvrigt identiske lande. I dette afsnit gives en uformel beskrivelse af mekanismerne i en ØG-model for det internationale kapitalmarked, idet en formel præsentation af modellen er henvist til appendiks A.

Der betragtes en verdensøkonomi med to lande – land H og F – som er helt identiske med hensyn til præferencer, teknologi og faktorudrustning. Kapital antages at kunne flyttes omkostningsfrit mellem lande, og idet kapitalejerne søger at maksimere deres kapitalafkast, vil der findes to typer af ligevægt på det internationale kapitalmarked: *spredt ligevægt* – når kapitalafkastet er ens i de to lande – og *koncentreret ligevægt* – når al kapital befinder sig i et land, og kapitalafkastet er højest i dette land.

Kapitalafkastet i et land bestemmes på varemarkedet, hvor den anvendte teknologi fordrer, at hver virksomhed råder over netop en kapitalenhed. Idet nye virksomheder antages at have fri adgang til markedet, vil kapitallejen blive budt op indtil hele virksomhedernes variable profit går til aflønning af kapital. Varemarkedet antages desuden at være karakteriseret ved følgende: (i) *imperfekt konkurrence* – virksomhederne er mængdesættende oligopolister og tjener derfor en variabel profit, (ii) *handelsomkostninger* – virksomhederne afsætter på både hjemmemarked og eksportmarked, men må afholde en omkostning ved salg på eksportmarkedet, (iii) *koblinger* – virksomhederne producerer udover forbrugsgodet et differentieret halvfabrikata, som indgår i produktionsprocessen hos de øvrige producenter. Det bemærkes, at antagelserne (i)-(ii) tilsammen giver anledning til en *trængselseffekt* på varemarkedet. Når antallet af virksomheder i et land stiger, udhules *alt andet lige* disse virksomheders profit, fordi deres konkurrencefordel på det profitable hjemmemarked reduceres. Tilsvarende skaber antagelserne (ii)-(iii) en *klyngeeffekt* på varemarkedet. Når antallet af virksomheder i et land stiger, øges *alt andet lige* disse virksomheders profit, fordi færre halvfabrikata må importeres, og omkostningerne ved produktion af forbrugsgodet dermed reduceres. Virksomheder placeret i samme land påvirker altså hinandens profitabilitet negativt gennem trængselseffekten og positivt gennem klyngeeffekten. Idet aflønningen af ka-

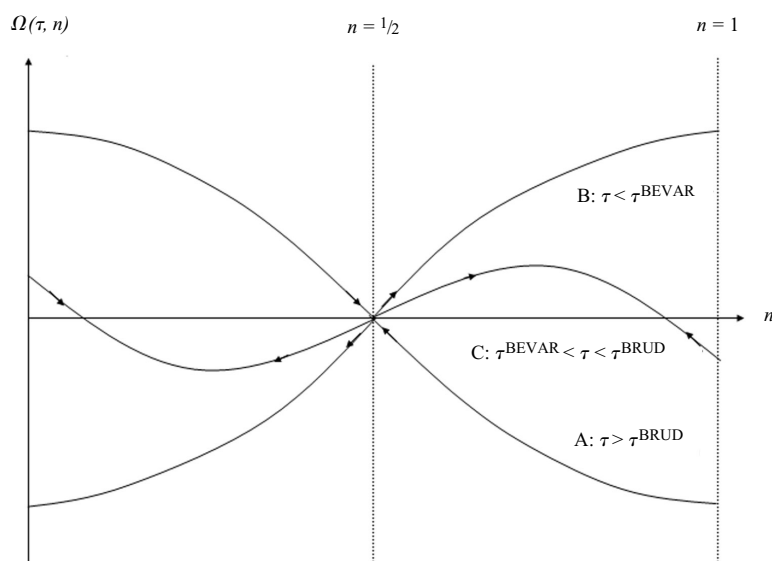
pital afhænger af virksomhedernes profitabilitet, er der på kapitalmarkedet identificeret både en spredningskraft – trængselseffekten – og en samlingskraft – klyngeeffekten.

Hvorvidt der på kapitalmarkedet opstår en spredt ligevægt eller en koncentreret ligevægt afhænger altså af om trængselseffekten eller klyngeeffekten dominerer, og her viser det sig, at handelsomkostningerne spiller en afgørende rolle. Handelsomkostningerne udtrykker i hvor høj grad virksomhederne er afskærmet fra udenlandsk konkurrence og har dermed betydning for såvel trængselseffekten – det er mere værdifuldt at befinde sig på et marked med relativt få konkurrenter, når handelsbarriererne er store – som klyngeeffekten – det er mere omkostningsbesparende at befinde sig tæt på ens inputleverandører, når handelsbarriererne er store. Det er et standardresultat i ØG-litteraturen og formelt demonstreret i appendiks A, at trængselseffekten dominerer, når handelsomkostningerne er store, mens klyngeeffekten er stærkest, når handelsomkostningerne er små.

Det kan nu bestemmes, hvilken type af ligevægt der opstår for forskellige intervaller af handelsomkostninger. Det bemærkes indledningsvist, at den *symmetriske kapitalallokering* er en ligevægt uanset handelsomkostningernes størrelse. Dette skyldes, at de to lande er fuldstændig identiske med hensyn til teknologi og præferencer, hvorfor virksomheder i de to lande må være lige profitable, når produktionsfaktorerne er ligeligt fordelt. Den symmetriske ligevægt er imidlertid ikke nødvendigvis stabil. For handelsomkostninger lavere end et såkaldt *opbrudspunkt*, dominerer klyngeeffekten trængselseffekten, hvorfor små stød til kapitalallokeringen er selvforstærkende – eksempelvis vil en lille kapitalstrøm fra land F til land H gøre virksomhederne i land H mere profitable end deres konkurrenter i land F og dermed udløse nye kapitalstrømme fra land F til land H. Omvendt vil den symmetriske ligevægt være stabil, når handelsomkostningerne er højere end opbrudspunktet, fordi trængselseffekten dominerer klyngeeffekten, og små stød til kapitalallokeringen derfor er selvophævende – en lille kapitalstrøm fra land F til land H vil reducere profitabiliteten i land H og øge den i land F, og udløse kapitalstrømme tilbage til land F.

Ligevægte kan ligeledes opstå, når al kapital er placeret i et af de to lande. I sådanne *koncentrerede kapitalallokeringer* vil klyngeeffekten dominere trængselseffekten for handelsomkostninger lavere end et såkaldt *bevarelsespunkt*. Dermed vil virksomheder placeret i land H (eller F) være mere profitable end potentielle konkurrenter i land F (eller H), og den koncentrerede allokering udgør en stabil ligevægt. Omvendt er trængselseffekten stærkest for handelsomkostninger højere end bevarelsespunktet, og den koncentrerede allokering udgør ikke en ligevægt.

Kapitalmarkedets dynamik illustreres af figur 1, der viser hvordan det relative kapitalafkast  $\Omega$  – forskellen mellem kapitalafkastet i land H og i land F – afhænger af kapitalallokeringen  $n$  – andelen af verdensøkonomiens kapital placeret i land H – for for-

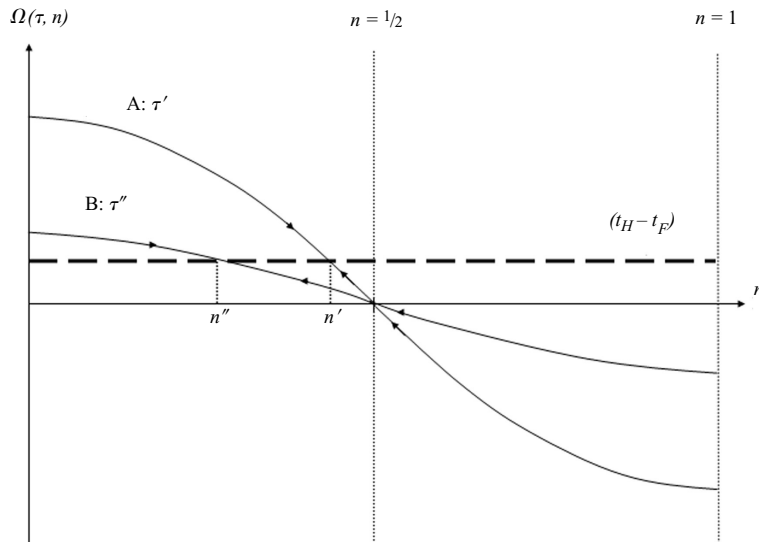


Figur 1. Kapitalmarkedets dynamik.

skellige intervaller af handelsomkostningerne  $\tau$ . Kapitalejerne placerer deres kapital i landet med det højeste afkast, og  $n$  er derfor voksende når  $\Omega > 0$  og aftagende når  $\Omega < 0$  som illustreret af figurens pile. Der findes tre typer af ligevægt: (i)  $\Omega = 0$  – spredt ligevægt, (ii)  $\Omega > 0$  og  $n = 1$  – koncentreret ligevægt med klynge i land H og (iii)  $\Omega < 0$  og  $n = 0$  – koncentreret ligevægt med klynge i land F.

Når handelsomkostningerne er højere end opbrudspunktet  $\tau^{BRUD}$  er det relative kapitalafkast givet ved den negativt hældende kurve A. Fra ethvert udgangspunkt vil kapitalmarkedet bevæge sig mod den symmetriske allokering, som derfor er en stabil ligevægt. For handelsomkostninger lavere end bevarelsespunktet  $\tau^{BEVAR}$  er det relative kapitalafkast givet ved den positivt hældende kurve B, og i dette interval af  $\tau$  vil kapitalmarkedet bevæge sig mod en af de to koncentrerede allokeringer, som således er lokalt stabile ligevægte. Idet  $\tau^{BRUD} > \tau^{BEVAR}$ , findes der et interval af handelsomkostninger, hvor den symmetriske ligevægt er ustabil, men den koncentrerede allokering endnu ikke er en ligevægt. Her er det relative kapitalafkast givet ved kurven C, og der vil således opstå spredte, men asymmetriske ligevægte. Der henvises til Johannesen (2004) for en nærmere analyse af denne type ligevægte.

Den ovenfor beskrevne ØG-model er formelt præsenteret i appendiks. Modellen adskiller sig fra eksisterende kapitalmarkedsmodeller i ØG-litteraturen ved at antage oligopolistisk konkurrence på varemarkedet i stedet for monopolistisk konkurrence. Idet der må modelleres et selvstændigt marked for inputs, er modellen mindre elegant



Figur 2. Skattekonkurrence i symmetrisk ligevægt.

end eksempelvis Robert-Nicoud (2002). Til gengæld findes der eksplicitte løsningsudtryk for opbrudspunkt og bevarelsespunkt, hvor modeller, der antager monopolistisk varemarkedskonkurrence, ofte må nøjes med implicite løsninger. Modellernes kvalitative egenskaber er dog identiske med den undtagelse, at spredte, asymmetriske ligevægte ikke forekommer i typiske ØG-modeller, hvor det er et standardresultat at  $\tau^{BRUD} < \tau^{BEVAR}$ .

### 3.2 Skattekonkurrence i symmetrisk ligevægt

For at kunne analysere skattekonkurrence i den opstillede kapitalmarkedsmodel, må den udvides med en offentlig sektor, som udbyder et skattefinansieret offentligt gode. Kapitalejerne søger nu at maksimere nettoafkastet af deres kapital. Er der forskel på nettoafkastet i de to lande, flyttes kapital til landet med det højeste nettoafkast, hvilket enten fører til at nettoafkastene udlignes – spredt ligevægt – eller til at kapital koncentrerer sig i et land – koncentreret ligevægt. Kapitalmarkedet fungerer således nøjagtig som beskrevet i afsnit 3.1 med den ene forskel, at regeringerne direkte påvirker kapitalafkastet gennem deres valg af skattesats.

I dette afsnit belyses skattekonkurrence mellem to lande under antagelse af, at handelsomkostningerne er tilstrækkelig høje til, at den symmetriske ligevægt er stabil ( $\tau > \tau^{BRUD}$ ). Regeringernes optimale valg af skattesat i symmetrisk ligevægt er illustreret i figur 2 og den grafiske analyse følger Baldwin m.fl. (2003).

Der tages udgangspunkt i en situation, hvor landene har identiske skatterater, og kapitalen derfor er fordelt ligeligt mellem landene. Hvis regeringen i land H hæver skatteraten marginalt, reduceres nettoafkastet af kapital i dette land tilsvarende. Følgelig strømmer kapital til land F, hvilket øger profitabiliteten i land H og reducerer den i land F, fordi trængselseffekten dominerer klyngeeffekten. Når kapitalbevægelsen er tilstrækkelig stor til, at forskellen i profitabilitet netop kompenserer forskellen i skatterater, nås en ny ligevægt.

Hvor meget kapitalapparatet må reduceres i land H, før kapitalmarkedet igen er i ligevægt, afhænger imidlertid af styrkeforholdet mellem samlingskræfter og spredningskræfter og dermed af handelsomkostningernes niveau. For meget høje handelsomkostninger  $\tau'$  er spredningskræfterne langt stærkere end samlingskræfterne og ligevægten genfindes ved  $n'$ . For lavere handelsomkostninger  $\tau''$  afbalanceres spredningskræfterne i højere grad af samlingskræfterne, og en større mængde kapital må flytte, før kapitalmarkedet vender tilbage til ligevægt ved  $n''$ . I ligevægt vil de to identiske lande naturligvis vælge den samme skatterate, hvilket udligner kapitalbevægelserne og gør  $n = 1/2$  til ligevægt.

I modsætning til tidligere studier løser Johannesen (2004) regeringens problem formelt og udleder førsteordensbetingelsen på en form, som er direkte sammenlignelig med resultaterne i den traditionelle skattekonkurrencelitteratur.

$$\theta'(g^*) = \frac{\overbrace{\frac{1}{2}}^{\text{Skatteeksport}}}{1} + \frac{\overbrace{\frac{\partial P}{\partial t_H} \frac{y}{n}}^{\text{Priseffekt}}}{\underbrace{1 + \varepsilon}_{\text{Handelsomk. effekt}}} \quad (2.1)$$

hvor  $\theta'(g)$  er den repræsentative forbrugers marginalnytte af det offentlige gode,  $y$  er det aggregerede forbrug,  $P$  er prisniveauet som afhænger af skatteraten gennem dens effekt på kapitalallokeringen,  $\varepsilon$  er elasticiteten af kapitaludbuddet  $n$  med hensyn til skatteraten  $t_H$ . Ligning (2.1) afviger fra førsteordensbetingelsen i den simple skattekonkurrencemodell (1.5) på tre måder:

(1) *Skatteeksporteksternalitet*: Kapitalen aflønnes med hele virksomhedernes overskud, og kapitalejerne er således de facto ejere af virksomhederne. Kapitalmarkederne antages at være fuldstændig integrerede, hvorfor halvdelen af den anvendte kapital er

udenlandsk ejet. I ligning (2.1) optræder derfor  $1/2$  i tælleren i stedet for 1 som i standardløsningen (1.5). Der er tale om en skatteeksporteffekt, der trækker i retning af en inefficent høj kapitalskatterate (som beskrevet i afsnit 2).

(2) *Priseksternalitet*: Kapitalbeskatning har en omkostning gennem varepriserne, som ikke optræder i standardmodellen, og som ikke synes eksplicit beskrevet i tidligere studier. Når kapitalskatteraten hæves, falder nettoafkastet af kapital, hvilket reducerer antallet af indenlandske virksomheder. Dermed øges de tilbageværende virksomheders markedsmagt og vareprisen stiger. Resultatet er en udhuling af forbrugernes realindkomst, som modsvares af en tilsvarende realindkomststigning i udlandet, hvor tilstrømningen af virksomheder får priserne til at falde. Priseffekten trækker således i retning af en inefficent lav skatterate, fordi effekten på det udenlandske prisniveau ikke tages i betragtning af de skattesættende regeringer.

(3) *Handelsomkostninger* optræder ikke direkte i (2.1) men påvirker den optimale skatterate gennem elasticiteten  $\varepsilon$ . Når handelsomkostningerne er høje, er spredningskræfterne relativt stærke, og et lille fald i  $n$  er derfor tilstrækkeligt til at genskabe ligevægten på kapitalmarkedet efter en stigning i  $t_H$ . Dette svarer til en værdi af  $\varepsilon$  nær nul. Omvendt giver en skattestigning anledning til en stor kapitaludstrømning, når handelsomkostningerne er små, hvilket implicerer, at  $\varepsilon$  er stærkt negativ.<sup>2</sup> Altså intensiveres skattekonkurrencen, når handelsomkostningerne falder, og den optimale skatterate  $t_H^*$  varierer positivt med  $\tau$ .

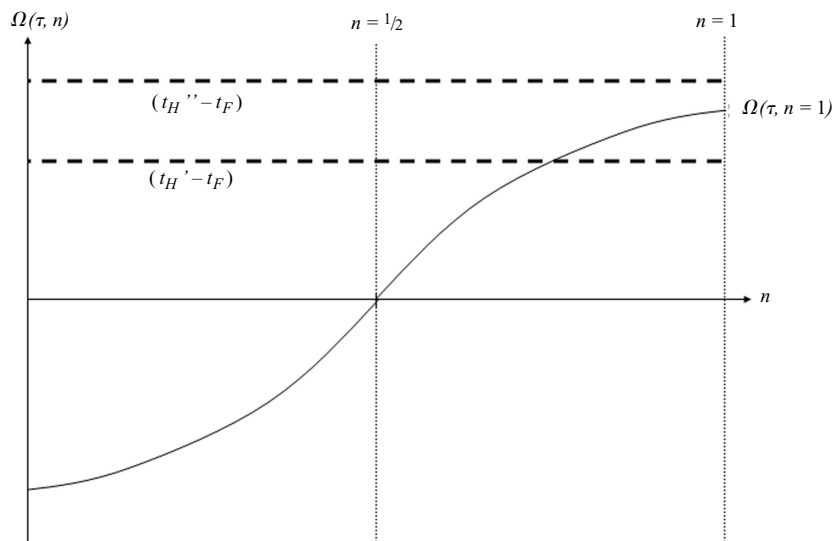
Det kan således konstateres, at analysen af skattekonkurrence i symmetrisk ligevægt ikke giver anledning til entydige normative konklusioner. Den sædvanlige skattebaseeffekt samt en priseffekt trækker i retning af en inefficent lav skatterate, men modvirkes af en skatteeksporteffekt, og det kan ikke a priori afgøres, hvilke af disse effekter, der er dominerende. Udover den tidligere identificerede sammenhæng mellem kapitalmobilitet og kapitalbeskatning peger modellen på en positiv sammenhæng mellem handelsomkostninger og kapitalskatterater i det givne interval af  $\tau$ .

### 3.3 Skattekonkurrence i koncentreret ligevægt

I det følgende analyseres skattekonkurrence i koncentreret ligevægt, idet det antages, at alle virksomheder initialt befinder sig i land H, og at handelsomkostningerne er tilstrækkelig lave til, at dette er en ligevægt. I det følgende antages det desuden, at hældningen på  $\Omega(\tau, n = 1)$  er positiv, idet der henvises til Johannesen (2004) for en analyse af det mere komplicerede tilfælde, hvor hældningen på  $\Omega(\tau, n = 1)$  er negativ.

Afsnittet er inspireret af Baldwin og Krugman (2000), men analysen er modificeret betydeligt, dels på grund af forskelle i den underliggende kapitalmarkedsmodel og dels for at gøre sammenligneligheden med den simple skattekonkurrencemodel så stor

2. For en formel demonstration af at  $\varepsilon$  er negativ og afhænger positivt af  $\tau$ , se Johannesen (2004).



Figur 3. Alt-eller-intet skattekonkurrence.

som muligt. Eksempelvis antager Baldwin og Krugman (2000), at land H vælger sin skatterate før land F, og finder dermed en Stackelberg-ligevægt i skatespillet. I nærværende afsnit fastholdes den sædvanlige antagelse om et simultant skatespil, og der bestemmes således en Nash-ligevægt.

Der betragtes altså en ligevægt, hvor hældningen på  $\Omega(\tau, n = 1)$  er positiv, og al kapital er placeret i land H. Som tidligere opfatter regeringerne hinandens skatterater som givne og vælger den for forbrugeren optimale beskatning af kapital. Kapitalmarkedets dynamik er illustreret i figur 3. Vælges en skatterate som  $t_H'$ , vil al kapital forblive i land H, mens en skatterate som  $t_H''$  vil få al kapital til at flytte til land F. Det er en væsentlig forskel i forhold til tidligere, at perfekt kapitalmobilitet ikke nødvendigvis giver anledning til skattekonkurrence. Hvis forbrugeren i land H har en relativ svag præference for det offentlige gode, således at regeringen i land H i fravær af kapitalmobilitet ville vælge en skatterate mindre end  $\Omega(\tau, n = 1)$ , da vil selv ikke en skatterate på nul i land F være begrænsende for skattevalget i land H. Den sædvanlige fortolkning i ØG-litteraturen er, at økonomiens samlingskræfter skaber en stedbunden økonomisk rente i land H, som kan beskattes, uden at det udløser kapitalbevægelser. Hvis denne rente er stor nok, vil skattekonkurrence ikke opstå, fordi land H kan vælge sin foretrukne skatterate og beholde kapitalklyngen uanset skattepolitikken i land F.

Kaldes den optimale skatterate i land H i fravær af kapitalmobilitet for  $t_H^{LUK}$ , må følgende betingelse altså være opfyldt, for at landene interagerer strategisk:

$$t_H^{LUK} > \Omega(\tau, n = 1) \quad (2.2)$$

I det følgende antages (2.2) at være opfyldt, således at skattekonkurrencen fra land F i det mindste er bindende for land H, når  $t_F$  er nul.

Det bemærkes, at den sædvanlige metode til bestemmelse af den optimale skatterate ikke umiddelbart kan benyttes, fordi  $\varepsilon$  er stærkt diskontinuert. For skatterater som  $t_H'$  og  $t_H''$  vil en marginal skattestigning ikke påvirke allokeringen af kapital ( $\varepsilon = 0$ ). Omvendt hvis nettokapitalafkastene i de to lande i udgangspunktet er identiske, vil en lille stigning i kapitalskatteraten få hele kapitalkoncentrationen til at flytte ( $\varepsilon = \infty$ ). Når  $\varepsilon$  ikke er kontinuert, er det ikke meningsfuldt at løse regeringens problem med differentialregning.

I stedet søges Nash-ligevægten bestemt med følgende verbale ræsonnement: Uanset skatteraten i land F, vil regeringen i land H aldrig sætte skatteraten højere end  $t_H^{LUK}$ , som er den optimale kapitalbeskatning i den lukkede økonomi. Regeringen i land H vil heller aldrig vælge en skatterate, som er så høj, at kapitalkoncentrationen flytter til land F. Regeringen vælger den højeste skatterate, som opfylder disse to betingelser, og reaktionsfunktionen i land H bliver dermed:

$$t_H^*(t_F) = \min \{t_H^{LUK}; \Omega(\tau, n = 1) + t_F\} \quad (2.3)$$

I land F vil regeringen søge at overtage kapitalkoncentrationen, så snart dette er muligt. Er skatteraten i land H større end  $\Omega(\tau, n = 1)$ , vil regeringen i land F vælge den højeste skatterate, som får kapitalkoncentrationen til at flytte. Er skatteraten i land H derimod mindre end  $\Omega(\tau, n = 1)$ , har skatteraten i land F ingen betydning, fordi der ikke er nogen skattebase at beskattes. I sådanne situationer, antages regeringen i land F at vælge en skatterate på nul, og reaktionsfunktionen i land F bliver således:

$$t_F^*(t_H) = \max \{0; t_H - \Omega(\tau, n = 1)\} \quad (2.4)$$

Med udgangspunkt i disse to reaktionsfunktioner kan den entydige Nash-ligevægt bestemmes som:

$$[t_H^N; t_F^N] = [\Omega(\tau, n = 1); 0]$$

Givet  $t_F^N = 0$  er  $t_H^N = \Omega(\tau, n = 1)$  optimal for land H, fordi en højere skatterate vil få al industri til at flytte til land F. Givet  $t_H^N = \Omega(\tau, n = 1)$  er enhver skatterate  $t_F$  optimal for land F, idet  $t_F$  hverken påvirker priser eller indkomster. Ligevægtens entydighed



kan vises på følgende måde: Antages det, at en skatterate  $t_F' > 0$  er optimal for land F, vil regeringen i land H hæve sin skatterate over  $\Omega(\tau, n = 1)$  uden at allokeringen af kapital ændres. Dette modsiger, at  $t_F'$  er optimal, idet  $t_F = 0$  da ville få al industri til at flytte til land F.

Det er et standardresultat i ØG-litteraturen og formelt demonstreret i appendiks A, at den stedbundne økonomiske rente  $\Omega(\tau, n = 1)$  er en først stigende og siden faldende funktion af handelsomkostningerne  $\tau$ . Eftersom land H i koncentreret ligevægt sætter sin skatterate lig  $\Omega(\tau, n = 1)$  kan analysen i afsnit 3.2 og 3.3 opsummeres på følgende måde:

- *H4: Reducerede handelsomkostninger medfører: (1) lavere kapitalskatterater når handelsomkostningerne er store, (2) højere kapitalskatterater, når handelsomkostningerne er mellemstore, og (3) lavere kapitalskatterater, når handelsomkostningerne er små.*

#### 4. Skattekonkurrence og empiri

I dette afsnit efterprøves de opstillede hypoteser H1-H4 i en empirisk model for kapitalbeskatning. Afsnit 4.1 gennemgår den eksisterende empiriske litteratur, hvilket leder til opstillingen af en empirisk model i afsnit 4.2. I det følgende afsnit 4.3 beskrives de i modellen anvendte variable, og de fundne resultater præsenteres i afsnit 4.4.

I modsætning til tidligere empiriske studier efterprøves hypoteserne H1-H4 i en samlet model i et forsøg på at reducere risikoen for *omitted variables bias*, og der eksperimenteres med en specifikation, der muliggør test for et ikke-monotont forhold mellem kapitalbeskatning og handelsomkostninger som forudsagt af H4. Desuden forbedres målingen af flere af modellens vigtigste variable – i særdeleshed udvikles med udgangspunkt i en gravitetsmodel en ny metode til bestemmelse af handelsomkostningerne ved international handel.

##### 4.1 Litteraturoversigt

Den empiriske litteratur om skattekonkurrence er relativt ny og endnu forholdsvis sparsom. De fleste studier efterprøver skattekonkurrencehypotesen, H1, ved at estimere en simpel sammenhæng mellem kapitalbeskatning og kapitalmobilitet, idet der typisk kontrolleres for enkelte politiske, konjunkturelle og demografiske variable. Både Quinn (1997) og Garrett og Mitchell (2001) finder en signifikant positiv sammenhæng mellem kapitalbeskatning og kapitalmobilitet, hvilket er det modsatte af, hvad man måtte forvente under hypotesen om skattekonkurrence. Krogstrup (2003) opnår mere tvetydige resultater, idet øget kapitalmobilitet vises at være forbundet med højere eller lavere kapitalbeskatning afhængig af det specifikke mål, der anvendes for de to

variable. Bretschger og Hettich (2002) etablerer en signifikant negativ empirisk sammenhæng mellem effektiv kapitalbeskatning og et udtryk for varemarkedsintegration, som imidlertid i bedste fald er et meget indirekte mål for international kapitalmobilitet. I litteraturen synes kun Devereux m.fl. (2003) at levere overbevisende empirisk støtte for H1, idet de finder en signifikant negativ sammenhæng mellem kapitalmobilitet og et mål for skattebyrden baseret på beskatningen af hypotetiske investeringsprojekter.

Meget få studier estimerer sammenhængen mellem kapitalbeskatning og henholdsvis udenlandsk ejerskab, produktivitetseffekter af offentlige udgifter og handelsomkostninger (H2-H4). En undtagelse udgøres af Hansson og Olofsdotter (2003), som introducerer et mål for handelsomkostningerne i en model af den ovenfor beskrevne type (simultant test af H1 og H4), uden at handelsomkostningerne dog har en signifikant effekt på den effektive kapitalskatterate. På tilsvarende måde anvender Krogstrup (2003) andre variable inspireret af den økonomiske geografi til at kontrollere for klyngeeffekter i en empirisk skattekonkurrencemodel. Endelig finder Huizinga og Nicodème (2003) en signifikant positiv sammenhæng mellem graden af udenlandsk ejerskab og kapitalbeskatning i en model, som dog ikke samtidig kontrollerer for international kapitalmobilitet (test af H2).

Denne læsning af den eksisterende empiriske litteratur giver anledning til følgende kommentarer: Først og fremmest er det bemærkelsesværdigt, i hvor høj grad tidligere studier fokuserer på H1. Kun Hansson og Olofsdotter (2003) og Krogstrup (2003) kombinerer forsøgsvist test af H1 med test af andre hypoteser inspireret af den teoretiske litteratur om skattekonkurrence. For det andet er det påfaldende, hvor vanskeligt det har vist sig at underbygge H1 empirisk. Kun Devereux m.fl. (2003) og til dels Krogstrup (2003) finder tegn på, at øget kapitalmobilitet signifikant har reduceret kapitalbeskatningen i løbet af de sidste årtier.

En mulig fortolkning er, at den ensidige fokusering på H1 giver anledning til en omitted variables bias, som forklarer vanskelighederne ved at eftervise H1 empirisk. Idet (i) kapitalmobilitet formentlig er korreleret med (ii) udenlandsk ejerskab, (iii) produktivitetseffekter af offentlige udgifter og (iv) handelsomkostninger, vil (i) – i modeller hvor (ii)-(iv) er udeladte – delvist opsamle effekten af disse variable, hvorfor koefficienten på (i) generelt vil være misvisende. For at undgå denne type bias, opstilles i det følgende en model, hvor kapitalbeskatningen forklares af alle variablene (i)-(iv) samt af diverse kontrolvariable.

#### 4.2 Økonometriske overvejelser og model

Inden der opstilles en empirisk model, må visse økonometriske vanskeligheder diskuteres. Det drejer sig særligt om *stationaritet*, der er relateret til anvendelsen af dy-

namiske variable, såvel som mere generelle problemstillinger som *endogenitet* og *kontrolvariable*.

Det er velkendt, at regressioner med tidsserier af makroøkonomiske variable kan give misvisende resultater, fordi sådanne variable hyppigt er ikke-stationære. Der findes i litteraturen adskillige tilgange til løsning af dette problem, og her følges i alt væsentligt en metode anvendt af Krogstrup (2003): For hvert land testes alle forklarende variable for eksistensen af en *unit root*. Variablen inkluderes i førsteordens ændringer, hvor unit roots forekommer hyppigt, og i niveauer, hvor der ikke forekommer unit roots. Metodens største svaghed synes at være, at der ikke findes klare retningslinjer for, om en variabel skal medtages i niveauer eller førsteordens ændringer, hvis der findes unit roots for nogle lande og ikke for andre.

En anden problemstilling relateret til regressioner med makroøkonomiske størrelser er endogenitet af de forklarende variable. En potentielt endogen forklarende variabel er kapitalmobilitet, som må formodes til en vis grad at være bestemt af politiske beslutninger om liberalisering af kapitalmarkedet, og tilsvarende synes de fleste andre forklarende variable i en vis udstrækning at være endogene. Igen benyttes en tilgang, som er inspireret af Krogstrup (2003): Potentielt endogene forklarende variable inkluderes med en periodes *lag*, hvilket sikrer at variablene er eksogene. En yderligere fordel ved denne fremgangsmåde er, at forsinkelsen også kan afspejle politikprocessens tidsdimension, og således forbedre modellens specifikation.

Endelig må der kontrolleres for makroøkonomiske og demografiske faktorer, som ikke er direkte relateret til skattekonkurrence, men som kan have betydning for regeringernes valg af kapitalskatterate. Der inkluderes udtryk for (1) økonomisk vækst, (2) inflation og (3) erhvervsfrekvens i den estimerede ligning, idet der henvises til Johannesen (2004) for en diskussion af disse variable. Den empiriske basismodel får således følgende udseende:

$$\Delta KAPSKAT_{i,t} = \alpha + \beta \cdot \begin{pmatrix} \Delta KAPMOB_{i,t-1} \\ \Delta UDLEJER_{i,t-1} \\ \Delta OFFPROD_{i,t-1} \\ \Delta HANDOMK_{i,t-1} \end{pmatrix} + \gamma \cdot \begin{pmatrix} \Delta VÆKST_{i,t-1} \\ \Delta ARBDYG_{i,t-1} \\ \Delta INFLA_{i,t} \end{pmatrix} + v_i + \varepsilon_{i,t} \quad (4.1)$$

hvor  $KAPMOB_{i,t}$  er mål for kapitalmobiliteten i land  $i$  på tidspunkt  $t$ ,  $UDLEJER_{i,t}$  er udtryk for et udenlandsk ejerskab,  $OFFPROD_{i,t}$  udtrykker produktivitetseffekter af offentlige udgifter,  $HANDOMK_{i,t}$  er mål for omkostningerne ved international handel,  $VÆKST_{i,t}$  er væksten i det reale bruttonationalprodukt,  $ARBDYG_{i,t}$  er andelen af befolkningen i den arbejdsdygtige alder og  $INFLA_{i,t}$  er væksten i forbrugerprisindekset.

Den første matrix af forklarende variable repræsenterer hypoteserne H1-H4, mens den anden matrix består af kontrolvariable.

Baseret på tests for enhedsrødder i Johannesen (2004) formuleres modellen i førsteordens ændringer, mens analysen af endogenitet i Johannesen (2004) leder til, at alle forklarende variable undtagen *INFLA* inkluderes med en periodes lag.

Den opstillede basismodel udvides nu på to forskellige måder. Først introduceres et interaktionsled mellem *HANDOMK* i niveauer og *HANDOMK* i førsteordensforskelle. Dermed undersøges det, om handelsomkostningernes effekt på kapitalskatteraten afhænger af handelsomkostningernes niveau, og der kan således testes for et ikke-monotont forhold mellem kapitalskatterate og handelsomkostninger, som forudsagt af H4:

$$\Delta KAPSKAT_{i,t} = \alpha + \beta \cdot \begin{pmatrix} \Delta KAPMOB_{i,t-1} \\ \Delta UDLEJER_{i,t-1} \\ \Delta OFFPROD_{i,t-1} \\ \Delta HANDOMK_{i,t-1} \\ (HANDOMK_{i,t-1} \times \\ \Delta HANDOMK_{i,t-1}) \end{pmatrix} + \gamma \cdot \begin{pmatrix} \Delta VÆKST_{i,t-1} \\ \Delta ARBDYG_{i,t-1} \\ \Delta INFLA_{i,t} \end{pmatrix} + v_i + \varepsilon_{i,t} \quad (4.2)$$

Denne type modeludvidelse forekommer ikke i tidligere studier, selvom ØG-litteraturen klart forudsiger en sådan ikke-monoton sammenhæng mellem kapitalbeskatning og handelsomkostninger.

Endelig udvides basismodellen med to variable for at tage højde for: (1) strategisk interaktion mellem regeringerne – skattestigninger i et land medfører kapitaludstrømning, som kan påvirke den optimale skatterate i andre lande – og (2) eksterne effekter af kapitalmarkedliberaliseringer – den faktiske mobilitet af kapital afhænger ikke bare af barrierer i det land, hvor kapitalen befinder sig, men også af barrierer i de lande, hvor kapitalen potentielt kan flyttes hen. Disse modeludvidelser er foreslået af Devereux m.fl. (2003):

$$\Delta KAPSKAT_{i,t} = \alpha + \lambda \cdot \begin{pmatrix} \Delta KAPSKAT_{i,t-1} \\ \Delta KAPMOB_{i,t-1} \end{pmatrix} + \beta \cdot \begin{pmatrix} \Delta KAPMOB_{i,t-1} \\ \Delta UDLEJER_{i,t-1} \\ \Delta OFFPROD_{i,t-1} \\ \Delta HANDOMK_{i,t-1} \end{pmatrix} + \gamma \cdot \begin{pmatrix} \Delta VÆKST_{i,t-1} \\ \Delta ARDYG_{i,t-1} \\ \Delta INFLA_{i,t} \end{pmatrix} + v_i + \varepsilon_{i,t} \quad (4.3)$$

hvor  $KAPSKAT_{-i,t}$  er den gennemsnitlige kapitalskatterate i andre lande end det betragtede land  $i$  på tidspunkt  $t$  og  $KAPMOB_{-i,t}$  er den gennemsnitlige kapitalmobilitet udenfor land  $i$ .

#### 4.3 Databeskrivelse

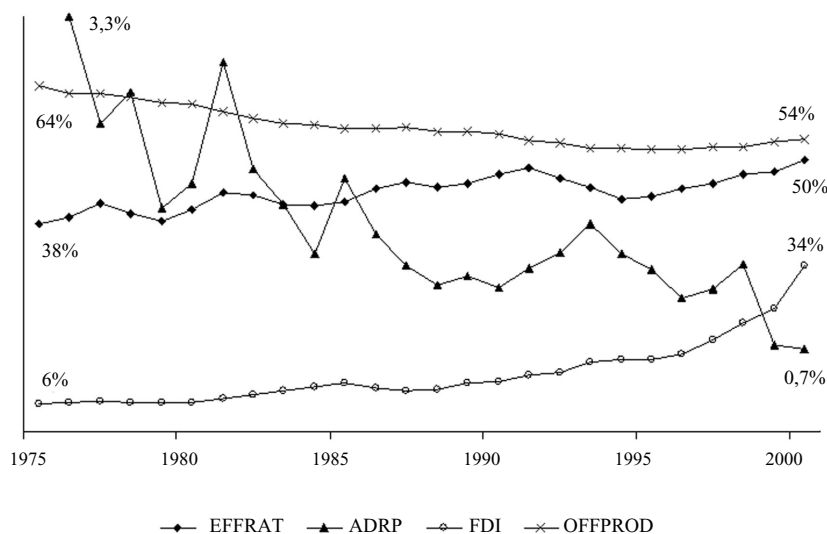
I dette afsnit beskrives de variable, som anvendes til estimation af ligningerne (4.1)-(4.3). Der er i princippet søgt data for så mange lande og så lang en tidsperiode som muligt. I praksis benyttes dog udelukkende tal for OECD-området, dels fordi data er langt mere tilgængelige for denne landegruppe end for andre lande, og dels fordi OECD-landene nogenlunde opfylder skattekonkurrencemodellens forudsætninger om homogenitet med hensyn til teknologi, faktorudrustninger og præferencer.

I litteraturen anvendes flere forskellige mål for skattebyrden på kapital,  $KAPSKAT$ , hvoraf de vigtigste er: den nominelle kapitalskatterate,  $NOMRAT$ , selskabsskatteprovenuet som andel af BNP,  $SELPRO$  og den effektive kapitalskatterate defineret som skatteprovenuet fra kapitalbeskatning som andel af den samlede kapitalindkomst,  $EFFRAT$ . Sidstnævnte mål er udviklet af Mendoza m.fl. (1994) og Carey og Tchilinguirian (2000) og diskuteres indgående af Krogstrup (2003).  $EFFRAT$  benyttes som mål for kapitalbeskatningen i de fleste nyere studier og er ligeledes anvendt i nærværende regressioner. Som det fremgår af figur 4 har  $EFFRAT$  været jævnt stigende gennem perioden 1975-2000 fra et niveau omkring 40 pct. i 1975 til omkring 50 pct. i 2000.<sup>3</sup>

Tre mål for international kapitalmobilitet,  $KAPMOB$ , har vundet udbredelse i litteraturen: beholdninger og strømme af direkte udenlandske investeringer,  $FDI$ , indeks som opsummerer graden af liberalisering af kapitalmarkedet,  $LIDX$  og afvigelse fra dækket renteparitet,  $ADRP$ , der kan opfattes som et mål for barriererne for kapitalmobilitet. I de følgende regressioner anvendes udelukkende  $ADRP$ , idet der henvises til Johannesen (2004) for en nærmere redegørelse for beregningsmetoder og forudsætninger. I figur 4 vises den gennemsnitlige afvigelse fra dækket renteparitet med Tyskland i OECD-landene i perioden 1976-2000. Kapitalmobiliteten i OECD-området er steget markant i løbet af perioden med særlig bemærkelsesværdige stigninger indtil midten af 1980'erne.

Af de omtalte studier anvender kun Huizinga og Nicodème (2003) et mål for graden af udenlandsk ejerskab  $UDLEJER$ , men idet dette er baseret på vanskeligt tilgængelige ejerskabsdata på virksomhedsniveau, må der her findes en alternativ fremgangsmåde. I de følgende regressioner anvendes estimerede beholdninger af indgående direkte udenlandske investeringer,  $FDI$ , som beregnet af Lane og Milesi-Ferretti (2001).

3. De effektive skatterater på kapital er ikke beregnet af forfatteren, men er venligst udlånt af David Carey, OECD.

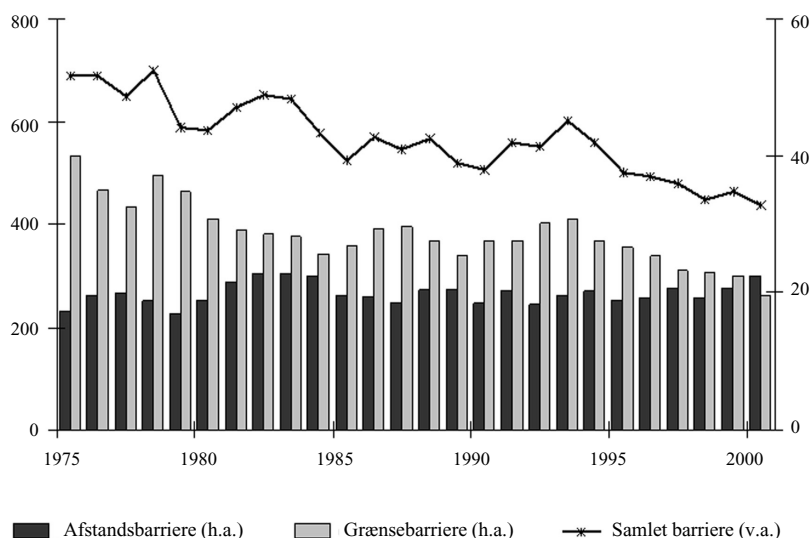


Figur 4. Anvendte variable.

Ideelt set ønskes beholdningen af udenlandsk ejede aktiver udtrykt som en andel af alle indenlandske aktiver, men i mangel på konsistente og sammenlignelige mål for disse, er *FDI* her normeret med det indenlandske BNP. I figur 4 rapporteres den gennemsnitlige beholdning af indgående direkte udenlandske investeringer i pct. af BNP i OECD-landene i perioden 1975-2000. Udenlandsk ejerskab er særligt efter 1990 vokset meget kraftigt. Fra et niveau på omkring 5-10 pct. af BNP i perioden 1975-1990 steg det til mere end 30 pct. af BNP i 2000.

Der synes ikke i den empiriske litteratur om skattekonkurrence at eksistere studier, der kvantificerer produktivitetseffekten af de offentlige udgifter, og der må således her konstrueres en ny fremgangsmåde. En simpel tilgang består i at kategorisere de forskellige typer af offentlige udgifter i henhold til om de må formodes at øge produktiviteten af privat kapital eller ej. I Johannesen (2004) forsøges en sådan kategorisering, og det foretrukne mål for produktivitetseffekten af de offentlige udgifter viser sig at være summen af offentlige investeringer, erhvervssubsidier og offentligt forbrug som andel af de totale offentlige udgifter. I figur 4 er OECD-gennemsnittet af dette mål rapporteret for perioden 1975-2000, og det er tydeligt, at andelen af offentlige udgifter med produktivitetseffekter i den private sektor er faldet jævnt i løbet af perioden.

Kun Hansson og Olofsdotter (2003) og Dreher (2003) inddrager internationale handelsomkostninger i analysen af skattekonkurrence, og begge anvender implicitte transportomkostninger, *IMPTRA*, – beregnet som værdien af et lands samlede import i



Figur 5. Estimerede effekter af handelsbarrierer.

c.i.f. priser divideret med værdien af importen i f.o.b. priser – som udtryk for handelsomkostningerne. Johannesen (2004) påpeger imidlertid en række problemer ved dette mål, og der udvikles derfor i appendiks B en alternativ metode til bestemmelse af økonomiens handelsomkostninger. Med udgangspunkt i en gravitetsmodel for international handel estimeres et meget bredt udtryk for handelsbarrierer omfattende:

- *Grænsebarrierer* – handelsbarrierer som favoriserer indenlandske producenter i forhold til udenlandske producenter. Mere præcist defineres denne variabel som det estimerede forhold mellem indenlandsk salg og import fra et land med samme udbudspotentiale i fraværet af afstandsbarrrierer. Det måles således, hvor meget en repræsentativ indenlandsk producents salg overstiger en repræsentativ udenlandsk producents salg som følge af faktorer, der ikke er relateret til markedsafstanden.
- *Afstandsbarrierer* – handelsbarrierer som favoriserer nære producenter i forhold til fjerne producenter. Variablen defineres som det estimerede forhold mellem indenlandsk salg og import fra et land med samme udbudspotentiale i fraværet af grænsebarrierer. Derved måles, hvor meget en repræsentativ indenlandsk producents salg overstiger en repræsentativ udenlandsk producents salg som følge af kortere afstand til markedet.

Summen af de estimerede grænsebarrierer og afstandsbarrrierer udgør den samlede handelsbarriere. Den gennemsnitlige samlede handelsbarriere for OECD-området er

illustreret i figur 5. Det bemærkes, at grænsebarrierer er den vigtigste barriere for international handel, og at denne type barriere har været kraftigt faldende gennem hele perioden. Omvendt synes afstandsbarriererne at have været nogenlunde stabile over perioden.

#### 4.4 Resultater

I dette afsnit estimeres modellerne opstillet i afsnit 4.2 ved hjælp af variablene dannet i afsnit 4.3. Først estimeres basismodellen, og der præsenteres en række robusthedstests. Dernæst estimeres de udvidede modeller, og resultaterne fortolkes. Resultatet af estimationen af basismodellen er rapporteret i første kolonne af tabel 1.

Det bemærkes at resultaterne støtter flere af de indledningsvist opstillede hypoteser. Koefficienten på *KAPMOB* er signifikant positiv, således at reducerede afvigelser fra dækket renteparitet i overensstemmelse med H1 er forbundet med lavere kapitalskatterater. Den signifikant negative koefficient på *HANDOMK* betyder, at lavere handelsomkostninger resulterer i højere kapitalskatter, hvilket passer med H4 i det tilfælde, hvor handelsomkostningerne er mellemstore. Koefficienterne på *UDLEJER* og *OFFPROD* er begge, som forudsagt af H2 og H3, positive uden dog at være signifikante. Kontrolvariablene *INFLA*, *ARBDYG* og *VÆKST* har de forventede fortegn, idet dog kun *INFLA* er signifikant. Resultaterne implicerer en relativt beskedne effekt på kapitalskatteraterne af den øgede kapitalmobilitet i OECD-området. Den gennemsnitlige afvigelse fra dækket renteparitet faldt med omkring 3 procentpoint i perioden 1975-2000, hvilket ifølge basismodellen alt andet lige skulle have medført et gennemsnitligt fald i kapitalskatteraten på under 2 procentpoint.

Johannesen (2004) udfører en række specifications- og robusthedstest af basismodellen, og estimaterne viser sig særdeles robuste over for ændringer i sample og estimationsmetode, ligesom der findes meget beskedne autokorrelation i fejleddet. Over en lang række regressioner findes der konsekvent en positiv koefficient på *KAPMOB* signifikant på mindst 10% signifikansniveau, samt en negativ koefficient på *HANDOMK* signifikant på mindst 1% signifikansniveau. På baggrund af disse resultater synes det rimeligt at konkludere, at basismodellen yder meget solid støtte til skattekonkurrencehypotesen H1, og at handelsomkostningerne som forudsagt af H4 spiller en væsentlig rolle for kapitalbeskatningen. Omvendt er *UDLEJER* ikke signifikant i nogen regressioner, og der findes således ingen empirisk støtte for H2 med det anvendte mål for udenlandsk ejerskab. Endelig er koefficienten på *OFFPROD* altid positiv og i visse regressioner signifikant på 10% signifikansniveau, hvilket yder en begrænset støtte til H3.

Dernæst testes for en ikke-monoton sammenhæng mellem handelsomkostninger og kapitalskatterate, idet der introduceres et interaktionsled mellem ændringer og niveauer af *HANDOMK* (ligning 4.2). Resultaterne er rapporteret i anden søjle af tabel 1.



Tabel 1. Regressioner.

	Basismodel (3.1)	Fleksibel funktionel form for handels- omkostninger (3.2)	Strategisk interaktion og symmetrisk kapitalmobilitet (3.3)
$\Delta KAPMOB_{i,t-1}$	0,572 (2,35)*	0,567 (2,34)*	0,310 (1,25)
$\Delta UDLEJER_{i,t-1}$	0,110 (0,61)	0,069 (0,38)	0,070 (0,39)
$\Delta OFFPROD_{i,t-1}$	0,356 (1,57)	0,347 (1,54)	0,365 (1,64)
$\Delta HANDOMK_{i,t-1}$	-0,018 (-2,66)**	-0,064 (-3,00)**	-0,028 (-3,95)**
$HANDOMK_{i,t-1} \times$ $\Delta HANDOMK_{i,t-1}$		0,058 (2,26)*	
$\Delta KAPSKAT_{-i,t-1}$			0,366 (1,97)*
$\Delta KAPMOB_{-i,t-1}$			1,865 (3,60)**
$\Delta VÆKST_{i,t-1}$	-0,069 (-0,52)	-0,082 (-0,63)	-0,047 (-0,36)
$\Delta ARBDYG_{i,t-1}$	-0,424 (-0,38)	-0,456 (-0,41)	-0,785 (-0,71)
$\Delta INFLA_{i,t}$	0,591 (3,84)**	0,531 (3,42)**	-0,728 (4,70)**
Konstant	0,546 (1,59)	0,403 (1,16)	0,655 (1,91)
$R^2$	0,095	0,112	0,135
$P(F)$	0,000	0,000	0,000
Antal observationer	313	313	323

Note: *t*-værdier i parenteser: \*\* signifikans på 1% niveau; \* signifikans på 5% niveau.

Det bemærkes, at koefficienten på *HANDOMK* forbliver signifikant negativ, mens det introducerede interaktionsled får en signifikant positiv koefficient. Dette tyder på en ikke-monoton sammenhæng, idet et lille fald i handelsomkostningerne påvirker skatteraten positivt, hvis handelsomkostningerne er små, mens det har en negativ effekt på skatteraten, hvis handelsomkostningerne er store. Det bemærkes, at den mere fleksible funktionelle form forbedrer modellens forklaringskraft betydeligt, idet  $R^2$  stiger med omkring 20 pct. i forhold til basismodellen.

Den udvidede model kan benyttes til at bestemme OECD-området opbrudspunkt. Det er indres fra ØG-modellen, at når  $\tau > \tau^{BRUD}$ , har fald i handelsomkostningerne en negativ effekt på skatteraten, fordi kapitalens følsomhed over for forskelle i nettoafkast vokser. Når omvendt  $\tau < \tau^{BRUD}$ , betyder yderligere fald i handelsomkostningerne, at skatteraten stiger, fordi den økonomiske rente ved at befinde sig i agglomerationen vokser. Opbrudspunktet bestemmes dermed som det niveau af handelsomkostningerne, hvor effekten af faldende handelsomkostninger skifter fra at påvirke skatteraten negativt til at påvirke den positivt. Dette sker netop, når det lineære led og interaktionsleddet ophæver hinanden, altså når *HANDOMK* er omkring 1100.<sup>4</sup>

Resultat er bemærkelsesværdigt givet de vanskeligheder, der er forbundet med empirisk at kvantificere økonomiens opbrudspunkt. Problemet er, som fremhævet af blandt andet Head og Mayer (2003), at ØG-modellerne forudsiger en stærkt diskontinueret sammenhæng mellem handelsomkostninger og allokeringen af virksomheder, som det ikke er muligt at teste med almindelig regressionsanalyse. Dette problem er her løst ved at estimere en indirekte, men kontinuert sammenhæng mellem kapital-skatte-rater og handelsomkostninger udledt fra en ØG-model med kapitalbeskatning. Som det fremgår af figur 5 havde et gennemsnitligt OECD-land allerede i 1975 passeret opbrudspunktet, mens få lande som f.eks. Sydkorea og New Zealand i dette år endnu havde totale handelsbarrierer større end 1100. Siden 1984 har alle OECD-lande haft handelsomkostninger lavere end det estimerede opbrudspunkt.

Endelig estimeres en udvidelse af basismodellen, som kontrollerer for strategisk interaktion og eksterne effekter af kapitalmarkedsliberaliseringer (ligning 4.3). Resultaterne er præsenteret i tredje kolonne af tabel 1. Koefficienten på  $KAPSKAT_{-i}$  er signifikant positiv og der er således klare tegn på strategisk interaktion mellem OECD-landene. Et typisk OECD-land reagerer på et 1 procentpoints fald i kapital-skatte-raten i de øvrige OECD-lande ved at sænke sin egen skatteme med omkring 0,4 procentpoint. Dette betyder, at den negative eksternalitet, et lands naboer udøver ved at sænke deres kapital-skatte-rater, delvist imødekommes med en sænkning af egen kapital-skatte-rate og delvist udmøntes i udstrømning af kapital. Det fremgår desuden af tabellen, at også koefficienten på  $KAPMOB_{-i}$  er positiv og stærkt signifikant, idet koefficienten på  $KAPMOB_i$  samtidig reduceres til omkring halvdelen. Dette resultat udgør en stærk støtte til Devereux m.fl. (2003), der påpeger, at øget kapitalmobilitet i et land har en effekt på skatteraten i alle lande.

4. Opbrudspunktet findes hvor:

$$\frac{\partial KAPSKAT}{\partial HANDOMK} = 0 \Leftrightarrow -0,064 \cdot \Delta HANDOMK + 0,058 \cdot \frac{HANDOMK}{1000} \cdot \Delta HANDOMK = 0$$

$$\Leftrightarrow HANDOMK \approx 1100$$

Det bemærkes sluttelig, at ovenstående resultater implicerer en langt større skattekonkurrenceeffekt end tidligere beregnet med udgangspunkt i resultaterne fra basismodellen. Det gennemsnitlige fald i afvigelsen fra dækket renteparitet på omkring 3 procentpoints siden 1975 har ifølge denne model resulteret i et gennemsnitligt fald i kapitalskatteraten på omkring 7 procentpoints, fordi øget kapitalmobilitet i et land medfører et nedadgående pres på kapitalskatteraten i alle andre lande.<sup>5</sup> Desuden forstærkes denne effekt af strategisk interaktion, fordi hvert land ud over unilateralt at sænke sin skatterate om følge af den øgede kapitalmobilitet også sænker sin skatterate som reaktion på de faldende skatterater i de øvrige lande. Den samlede effekt på skatteraten af øget kapitalmobilitet i perioden 1975-2000 kan således alt andet lige beregnes til omkring 10 procent.<sup>6</sup> Dette er en langt større effekt af skattekonkurrence end fundet i tidligere studier.

### 5. Konklusion

Mens den teoretiske litteratur om skattekonkurrence forudser en negativ sammenhæng mellem kapitalmobilitet og kapitalbeskatning synes en overfladisk empirisk analyse at afkræfte denne hypotese. Den internationale kapitalmobilitet er blevet stærkt forøget gennem de sidste årtier, men denne kapitalmarkedsintegration har i modstrid med teorien om skattekonkurrence været ledsaget af en betydelig stigning i den effektive kapitalskatterate (som defineret ovenfor).

Der er i dette studie blevet præsenteret teoretiske modeller for kapitalbeskatning, som medtager en række andre faktorer end blot kapitalmobilitet: Udenlandsk ejerskab af virksomheder muliggør skatteeksport og giver dermed anledning til en højere skatterate. Produktivitetseffekter af offentligt forbrug reducerer omkostningerne ved beskatning, og øger ligeledes den optimale skatterate. Handelsomkostninger har en kompleks betydning for kapitalbeskatningen, idet fald i handelsomkostningerne (1) intensiverer skattekonkurrencen og reducerer skatteraten, når handelsomkostningerne er store, (2) forstørrelser den økonomiske rente forbundet med en placering i virksomhedskoncentration og øger skatteraten, når handelsomkostningerne er mellemstore og (3) udhuler samme økonomiske rente og reducerer skatteraten, når handelsomkostningerne er små.

De teoretiske sammenhænge er desuden blevet testet i en empirisk model, og resultaterne viser, at øget international kapitalmobilitet faktisk har reduceret kapitalbeskatningen ganske væsentligt. Tages der højde for strategisk interaktion mellem landene,

5. Denne samlede effekt er simpelthen beregnet som koefficienten på  $KAPMOB_i$  plus koefficienten på  $KAPMOB_i$  – i alt omkring 2,2 – multipliceret med det gennemsnitlige fald i  $KAPMOB_i$  – omkring 3 points.

6. Resultatet fremkommer som værdien af en uendelig række  $6,6 * [1 + 0,37 + 0,37^2 + 0,37^3 \dots] = (6,6) / (1 - 0,37)$  hvor 6,6 er den »initiale« effekt på skatteraten forårsaget af øget kapitalmobilitet, og 0,37 er effekten af strategisk interaktion.

er den gennemsnitlige kapitalsskatterate i OECD-landene i de sidste årtier blevet reduceret med omkring 10 procentpoint som følge af skattekonkurrence. I samme periode har faldende handelsomkostninger imidlertid mere end opvejet dette fald, således at den samlede effekt på kapitalsskatteraten har været positiv.

De fundne resultater har særdeles vigtige policy-implikationer. Indtil nu har faldende handelsomkostninger muliggjort en øget kapitalbeskatning, fordi den økonomiske rente forbundet med at producere i en virksomhedskoncentration er vokset. Det forudses imidlertid af H4, at yderligere fald i handelsomkostningerne på et tidspunkt vil reducere den økonomiske rente i virksomhedskoncentrationen og dermed forstærke tendensen til skattekonkurrence snarere end at modvirke den. Mens international koordinering af skattepolitikken således indtil nu har været overflødig, fordi stadigt stærkere klyngeeffekter neutraliserede virkningerne af skattekonkurrence, forudsiger den teoretiske model, at klyngeeffekterne på et tidspunkt vil aftage. Når det sker, vil begge den økonomiske globaliserings to vigtigste elementer – kapitalmarkedsintegration og varemarkedsintegration – udøve et nedadgående pres på kapitalbeskatningen, og internationalt samarbejde vil blive af afgørende betydning, hvis en udhuling af landenes skatteprovenu skal undgås.

#### Litteratur

- Baldwin, R. R. Forslid, P. Martin, G. Ottaviano og F. Robert-Nicoud. 2003. *Economic Geography and Public Policy*, Princeton University Press.
- Baldwin, R. og P. Krugman. 2000. Agglomeration, integration and tax harmonisation, *CEPR Discussion Paper* nr. 2630.
- Bretschger, L. og F. Hettich. 2002. Globalisation, Capital Mobility and Tax Competition: Theory and Evidence for OECD Countries, *European Journal of Political Economy*, November 2002, 18(4), s. 695-716.
- Carey, D. og H. Tchilinguirian. 2000. Average effective tax rates on Capital, Labour and Consumption, *OECD Economics Department Working Paper* nr. 258.
- Devereux, M., B. Lockwood og M. Redoano. 2003. Capital account liberalization and corporate taxes, *Working Paper*, University of Warwick.
- Dreher, A. 2003. The Influence of Globalization on Taxes and Social Policy – an Empirical Analysis for OECD Countries, *Working Paper*, University of Exeter, nr.???
- Eijffinger, S. og W. Wagner. 2001. Taxation if capital is not perfectly mobile: Tax Competition versus tax Exportation, *CEPR Discussion Paper*, nr. 3084.
- Garrett, G. og D. Mitchell. 2001. Globalization, government spending and taxation in the OECD, *European Journal of Political Research* 39, s. 145-77.
- Hansson, Aa. og K. Olofsdotter. 2003. The effects of tax competition and new geography on taxation in OECD countries, Mimeo, Lund University.
- Head, K. og T. Mayer. 2003. The empirics of agglomeration and trade, kommer som kapitel i *Handbook of Regional and Urban Economics*.
- Huizinga, H. og G. Nicodème. 2003. *Foreign Ownership and Corporate Income Taxation: An Empirical Evaluation*, ....
- Huizinga, H. og S. B. Nielsen. 1997. Capital Income and Profit Taxation with Foreign Ownership of Firms, *Journal of International Economics*.
- Johannesen, N. 2004. Skattekonkurrence og Økonomisk Geografi, Specialeopgave,

- Københavns Universitet.
- Krogstrup, S. 2003. Are Capital Taxes Racing to the Bottom in the European Union?, *Working Paper* 01/2003, HEI Geneva.
- Krugman, P. 1991. Increasing returns and economic geography, *Journal of Political Economy* 99, s. 483-99.
- Lane, P. og G. M. Milesi-Ferretti. 2001. The External Wealth of Nations: Measures of Foreign Assets and Liabilities for Industrial and Developing Countries, *Journal of International Economics* 55, s. 263-94.
- Mendoza, E., A. Rafin og L. Tesar. 1994. Effective Tax Rates in Macroeconomics: Cross Country Estimates of Tax Rates on Factor Incomes and Consumption, *Journal of Monetary Economics*.
- Nitsch, V. 2000. National Borders and International Trade: Evidence from the European Union, *Canadian Journal of Economics*, 33(4): s. 1091-1105.
- Quinn, D. 1997. The Correlates of Change in International Financial Regulation, *The American Political Science Review*, bd. 91, nr. 3, s. 531-51.
- Robert-Nicoud, F. 2001. A simple geography model with vertical linkages and capital mobility, Mimeo, London School of Economics.
- Wei, S.-J. 1996. Intra-national versus International Trade: How Stubborn Are Nations in International Integration, *NBER Working Paper* nr. 5531.
- Wooders, M., B. Zissimos og A. Dhillon. 2001. Tax Competition Reconsidered, *Warwick Economic Research Papers*, nr. 622.

## Appendiks A: En kapitalmarkedsmodel med vertikale koblinger

### *A.1 Overblik:*

To identiske lande H og F – Samme teknologi, præferencer og faktorbeholdning

To sektorer A og B:

- Sektor A: Produktion af vare A samt et differentieret input til produktion af vare A
- Sektor B: Produktion af vare B

### *To varer A og B:*

- Vare A produceres med faste omkostninger til én kapitalenhed og variable omkostninger til differentierede inputs
- Vare B produceres med variable omkostninger til arbejdskraft

### *To varemarkeder A og B:*

- Varemarked A: (i) Oligopolistisk konkurrence, (ii) Transportomkostninger ved eksport
- Varemarked B: (i) Fuldkommen konkurrence, (ii) Ingen transportomkostninger

### *Stort antal faktormarkeder:*

- Arbejdsmarkeder: (i) Arbejdskraft immobil mellem lande, (ii) Fleksibel løn sikrer markedsclearing
- Kapitalmarked: (i) Perfekt kapitalmobilitet, (ii) Ligevægt enten når kapitalafkastet er ens i de to lande, eller når al kapital er koncentreret i et land og kapitalafkastet er højest i denne klynge
- Inputmarkeder: (i) Antal inputs = Antal virksomheder i sektor A, (ii) Transportomkostninger ved eksport, (iii) Simplificerende antagelse: virksomhederne afsætter det producerede input til en pris lig produktionsomkostningerne og opfører sig således kompetitivt på trods af de facto monopolstatus.

### *A.2 Notation:*

Når der skelnes mellem landene H og F, vil toptegn generelt indikere afsætnings- eller forbrugssted, mens fodtegn refererer til produktionsforhold. Priser vil derfor typisk have toptegn, fordi de er fælles for alle sælgere og købere på et givet marked, mens omkostninger har fodtegn, fordi de er fælles for producenterne på et marked. Mængder har derimod både toptegn og fodtegn, idet fodtegnet refererer til produktionsstedet, og toptegnet indikerer afsætningsstedet.

### A.3 Forbrugeren:

Forbrugerenes præferencer er givet ved nyttefunktionen  $U(\cdot)$ , hvor den funktionelle form på  $\Psi(\cdot)$  er valgt for at sikre en simpel lineær efterspørgsel efter  $y$ .

$$U(y, z, g) = \Psi(y) + x + \theta(g) \quad (\text{A.1})$$

$$\text{hvor: } \Psi(y) = \left(\alpha - \frac{y}{2}\right) \cdot y \quad (\text{A.1a})$$

og  $y$  er den repræsentative forbrugers forbrug af vare A,  $x$  er sammes forbrug af vare B, og  $g$  er den af regeringen valgte mængde af det offentlige gode. Forbrugerenes problem kan dermed opskrives som:

$$\max_{y,z} U(y, x, g)$$

$$\text{u.b. } I = P \cdot y + V \cdot x \quad (\text{A.2})$$

hvor  $I$  er den repræsentative forbrugers indkomst hidrørende fra et uelastisk arbejdsudbud  $m$  og et uelastisk kapitaludbud  $s$ ,  $P$  er prisen på vare A, og  $V$  er prisen på vare B. Løsning af forbrugerenes problem giver:

$$y^* = \alpha - \frac{P}{V} \quad (\text{A.3})$$

$$x^* = \frac{I}{V} - \left(\frac{P}{V}\right) \cdot \alpha + \left(\frac{P}{V}\right)^2 \quad (\text{A.4})$$

Antallet af forbrugere normeres til en, således at (A.3)-(A.4) kan opfattes som aggregerede efterspørgselsfunktioner.

### A.4 Sektor B

Produktionsteknologien i sektor B er givet ved:

$$\beta(l) = l$$

hvor  $\beta(l)$  er produktionsfunktionen for vare B og  $l$  er input af arbejdskraft. Antagelsen om fuldkommen konkurrence giver en fuldstændig elastisk udbudskurve:

$$V = W$$

hvor  $W$  er aflønning af arbejdskraft. Prisen på vare B normeres til en. Grundet antagelsen om omkostningsfri international handel i sektor B må  $V=1$  gælde i både land H og F.

#### A.5 Sektor A

Produktionsteknologien i sektor A fordrer, at hver virksomhed har netop en enhed fysisk kapital. Altså bestemmes antallet af virksomheder i den globale økonomi af mængden af kapital:

$$2s = n^W \quad (\text{A.5})$$

hvor  $n^W$  er antal virksomheder i sektor A i den samlede økonomi.

#### Omkostningsdel:

Virksomheder i såvel land H som land F producerer ved anvendelse af en enhed fysisk kapital vare A i henhold til produktionsfunktionen:

$$\eta(z) = \frac{\min\{z_1, z_2, \dots, z_{n^W}\}}{\lambda}, \lambda < 1 \quad (\text{A.6})$$

hvor  $\eta(z)$  er produktionsfunktion for vare A, og  $z_i$  er mængden af input  $i$  (for  $i = 1, 2, 3, \dots, n^W$ ). Den marginale enhedsomkostning  $c$  fremkommer således som løsningen til følgende omkostningsminimeringsproblem:

$$\min \sum_{j=1}^{n^W} z_j \cdot D_j$$

u. b.  $\eta(z) = 1$

hvor  $D_j$  er prisen på input  $i$  (for  $i = 1, 2, 3, \dots, n^W$ ). Løsningen til minimeringsproblemet er:

$$c = \sum_{j=1}^{n^W} D_j \cdot \lambda \quad (\text{A.7})$$

Virksomhederne producerer således med konstante marginalomkostninger. Idet inputpriserne afhænger af produktionsstedet, må der i det følgende skelnes mellem  $c_F$  og  $c_H$ .



NB: De samme kvalitative resultater kunne være opnået med eksempelvis en Cobb-Douglas produktionsfunktion. De afgørende egenskaber ved produktionsfunktionen er (1) konstant skalaafkast og (2) at de  $n^W$  inputs ikke er perfekte substitutter, således at en prisstigning på et input resulterer i højere enhedsomkostninger ved produktion af output.

*Outputdel:*

De  $n^W$  virksomheder udgør på såvel marked H som F et oligopol, og der findes en Nash-ligevægt på hvert af disse markeder. En tilfældig virksomhed  $i$  placeret i land H har på marked H følgende profitmaksimeringsproblem (bibetingelsen er en omskrivning af A.3):

$$\max_{q_i^H} \pi_i^H = q_i^H \cdot (P^H - c_{H_i}) \quad (\text{A.8})$$

$$\text{u.b. } P^H = \alpha - \sum_{j=1}^{n^W} q_j$$

hvor  $q_i^H$  er den mængde virksomhed  $i$  afsætter på marked H. Problemet har førsteordensbetingelsen:

$$\frac{\partial \pi_i^H}{\partial q_i^H} = 0 \Leftrightarrow q_i^H = \frac{(\alpha - \sum_{j \neq i} q_j^H - c_{H_i})}{2} \quad (\text{A.9})$$

Tilsvarende er førsteordensbetingelsen for profitmaksimering på marked H for en tilfældig virksomhed  $k$  placeret i land F:

$$\frac{\partial \pi_k^H}{\partial q_k^H} = 0 \Leftrightarrow q_k^H = \frac{(\alpha - \sum_{j \neq k} q_j^H - c_{Fk} - \tau)}{2} \quad (\text{A.10})$$

hvor  $\tau$  er enhedsomkostning ved international handel. Alle virksomheder placeret i land H vil producere samme mængde til marked H som virksomhed  $i$  – denne mængde kaldes  $q_H^H$  – og alle virksomheder i land F producerer samme mængde til marked H som virksomhed  $k$  – denne mængde kaldes  $q_F^H$ . Dermed kan (2.7) og (2.8) opfattes

som et system af to ligninger med to ubekendte, fra hvilket de to virksomhedstypers ligevægtsmængder kan udledes:

$$q_H^H = \frac{\alpha - c_H}{2} + \frac{(1-n)}{2} \cdot (c_F + \tau - c_H) \quad (\text{A.11})$$

$$q_F^H = \frac{\alpha - c_F - \tau}{2} + \frac{n}{2} \cdot (c_H - \tau - c_F) \quad (\text{A.12})$$

hvor  $n$  er antallet af virksomheder i land H og  $n^W$  er normeret til en. Ved at aggregere de fundne mængder og indsætte i efterspørgselsfunktionen findes prisen på vare A på marked H:

$$P^H = \frac{\alpha + n \cdot c_H + (1-n) \cdot (c_F + \tau)}{2} \quad (\text{A.13})$$

Tilsvarende findes for marked F ligevægtsmængder og pris:

$$q_H^F = \frac{\alpha - c_H - \tau}{2} + \frac{(1-n)}{2} \cdot (c_F - \tau - c_H) \quad (\text{A.14})$$

$$q_F^F = \frac{\alpha - c_F}{2} + \frac{n}{2} \cdot (c_H + \tau - c_F) \quad (\text{A.15})$$

$$P^F = \frac{\alpha + n \cdot (c_H + \tau) + (1-n) \cdot c_F}{2} \quad (\text{A.16})$$

*Faktormarked:*

Virksomhederne producerer hver et input i henhold til produktionsfunktionen:

$$\varphi_i(l) = l_i \quad \text{for } i = 1, 2, \dots, n^W$$

hvor  $l_i$  er input af arbejdskraft. Inputtene prisfastsættes kompetitivt, og en virksomhed  $j$  placeret i land H vil således sætte følgende inputpriser i henholdsvis land H og land F:

$$D_j^H = w$$

$$D_j^F = w + \tau$$

Dermed kan virksomhedernes enhedsomkostninger opskrives på følgende måde:

$$c_H = \sum_{j=1}^{n^W} D_j^H \cdot \lambda = w \cdot \lambda + (1 - n) \cdot \tau \cdot \lambda \quad (\text{A.17})$$

$$c_F = \sum_{j=1}^{n^W} D_j^F \cdot \lambda = w \cdot \lambda + n \cdot \tau \cdot \lambda \quad (\text{A.18})$$

NB: Heller ikke antagelsen om kompetitive inputmarkeder er afgørende for de kvalitative resultater. Den reducerer imidlertid modellens kompleksitet, idet den reducerer profitten fra inputproduktionen til nul. Dette svarer til at eliminere ØG-litteraturens markedsadgangseffekt.

#### A.6 Arbejdsmarked

Arbejdsmarkedet antages at være kompetitivt, således at en fleksibel nominel løn  $W$  skaber ligevægt mellem det inelastiske arbejdsudbud og arbejdsefterspørgslen som kommer fra såvel outputproduktionen i sektor B som inputproduktionen i sektor A.

#### A.7 Kapitalmarked

Forbrugerne udlejer kapital til virksomhederne i sektor A, og idet der antages at være fri adgang til varemarkedet, vil denne leje blive budt op, indtil hele virksomhedernes variable profit går til aflønning af kapital. Da hver virksomhed anvender en enhed kapital, er kapitalens nominelle afkast derfor identisk med hele virksomhedens variable profit. Forbrugerne placerer deres kapital, hvor de får det højeste nominelle afkast, og der kan derfor forekomme to typer af ligevægte på kapitalmarkedet.

(I) En *spredt ligevægt* opstår, når virksomheder placeret i land H og land F er lige profitable

$$\pi_H^{TOT} = \pi_F^{TOT} \quad (\text{A.19})$$

hvor  $\pi_H^{TOT}$  og  $\pi_F^{TOT}$  er den totale profit for en virksomhed placeret i henholdsvis land H og F.

(II) En koncentreret ligevægt opstår, hvis alle virksomheder er samlet i et land, og deres profit er højere end den profit, en virksomhed potentielt kunne tjene i landet uden virksomheder:

$$n = 1 \text{ og } \pi_H^{TOT} > \pi_F^{TOT} \quad (\text{A.20A})$$

eller:

$$n = 0 \text{ og } \pi_H^{TOT} < \pi_F^{TOT} \quad (\text{A.20b})$$

#### *Symmetrisk ligevægt*

Da landene er identiske, må den symmetriske kapitalallokering  $n = 1/2$  være en spredt ligevægt som defineret ovenfor. Ifølge Baldwin m.fl. (2003) kan en sådan ligevægt karakteriseres som stabil, hvis små stød til kapitalallokeringen er selvophævede, mens den er ustabil, hvis små stød er selvforstærkende. Der udføres et sådant stabilitetstest, idet det beregnes, hvordan en marginal kapitalstrøm fra land F til land H påvirker den relative profitabilitet – identisk med det relative kapitalafkast – i de to lande. Efter megen vanskelig algebra fås:

$$\left( \frac{\partial (\pi_H^{TOT} - \pi_F^{TOT})}{\partial n} \Big|_{n = \frac{1}{2}} \right) = [-2\lambda^2 - 2\lambda - 2] \cdot \tau^2 + [4\lambda \cdot (\alpha - \lambda)] \cdot \tau \quad (\text{A.21})$$

Dette andengradspolynomium i  $\tau$  er negativt for store værdier af  $\tau$ , og positivt for små værdier af  $\tau$ . Nulpunkterne er  $\tau = 0$  og:

$$\tau^{BRUD} = \frac{2\lambda \cdot (\alpha - \lambda)}{\lambda^2 + \lambda + 1}$$

Den symmetriske ligevægt er således stabil for  $\tau > \tau^{BRUD}$  og ustabil for  $\tau < \tau^{BRUD}$

#### *Koncentreret ligevægt*

Kapitalallokeringen  $n = 1$  (alle virksomheder placeret i land H) er en ligevægt, hvis  $\pi_H^{TOT} > \pi_F^{TOT}$ , og betingelserne for dette undersøges. Endnu engang kan det med vanskelig algebra vises at:

$$(\pi_H^{TOT} - \pi_F^{TOT} | n = 1) = [-4\lambda^2 - 2\lambda - 2] \cdot \tau^2 + [4\lambda \cdot (\alpha - \lambda)] \cdot \tau \quad (\text{A.22})$$

Også dette polynomium er negativt for store værdier af  $t$  og positivt for små værdier. Nulpunkterne er  $\tau = 0$  og:

$$\tau^{BEVAR} = \frac{2\lambda \cdot (\alpha - \lambda)}{2\lambda^2 + \lambda + 1}$$

Koncentrerede ligevægte er altså mulige for  $\tau < \tau^{BEVAR}$ .

#### *Asymmetriske, spredte ligevægte*

Af ovenstående følger, at der findes et interval mellem  $\tau^{BEVAR}$  og  $\tau^{BRUD}$  hvor den symmetriske ligevægt er ustabil, og koncentrerede ligevægte ikke kan opstå. I dette interval vil kapitalen være placeret asymmetrisk uden at være fuldt koncentreret.

#### *Dynamik*

Kapitalmarkedets dynamik illustreres af figur 1 i teksten.

### Appendiks B: En gravitetsmodel til bestemmelse af internationale handelsomkostninger

Grænsebarrierer og afstandsbarrierer for international handel estimeres ved hjælp af en udvidet gravitetsmodel som præsenteret i Wei (1996). I modsætning til simple gravitetsmodeller inkluderes såvel internationale som intra-nationale handelsstrømme, hvor sidstnævnte beregnes som et lands samlede produktion fratrukket landets eksport:

$$SALG_{i,j,t} = \frac{\alpha \cdot BNP_{i,t}^{\beta_{1,t}} \cdot BNP_{j,t}^{\beta_{2,t}} \cdot HJEMMAR_{i,j}^{\beta_{4,t}}}{AFST_{i,j}^{\beta_{3,t}}} \quad (B.1)$$

hvor  $BNP_i$  og  $BNP_j$  er bruttonationalproduktet i henholdsvis land  $i$  og land  $j$ ,  $SALG_{i,j,t}$  er salg fra land  $i$  til land  $j$  på tidspunkt  $t$ , hvilket vil sige eksport når  $i \neq j$  og produktion til indenlandsk anvendelse når  $i = j$ ,  $HJEMMAR$  er en variabel, som antager værdien  $e$ , når der er tale om intra-nationale handelsstrømme og værdien 1, når handelsstrømme er internationale, og  $AFST_{i,j}$  er afstanden mellem land  $i$  og land  $j$ .

Internationale handelsafstande måles her som afstanden mellem de handlende landes hovedstæder, mens intra-nationale handelsafstande i lighed med Nitsch (2000) tilnærmes ved radius i en cirkel med samme areal som landets overflade.

Estimationen af (4.4) foretages med bilaterale handelsdata fra OECD's *Monthly International Trade*, indkomstdata fra OECD's *Main Economic Indicators*, og afstandsdata. I praksis anvendes naturligvis en logaritmisk omskrivning, og der estimeres en separat ligning for hvert  $t$ , så der opnås tidsspecifikke koefficienter. Resultaterne af regressionerne er rapporteret i Johannesen (2004). Alle variable er stærkt signifikante og  $R^2$  er i alle regressioner mellem 0,8 og 0,9. De estimerede koefficienter er i overensstemmelse med resultaterne fra den eksisterende litteratur om udvidede gravitetsmodeller.

Grænsebarrierer bestemmes som det estimerede forhold mellem indenlandsk salg og import fra et land med samme udbudspotentiale i fraværet af afstandsbarriere. Dette mål er identisk med begrebet *homebias* introduceret af Wei (1996).

$$GRÆNSEBAR_t = e^{\hat{\beta}_{4,t}} \quad (B.2)$$

Der synes ikke i litteraturen at være beskrevet en måde, hvorpå der kan konstrueres et tilsvarende mål for effekten af afstandsbarrierer. Det virker imidlertid naturligt at bestemme afstandsbarrierer som det estimerede forhold mellem indenlandsk salg og

import fra et land med samme udbudspotentiale i fraværet af grænsebarrierer. Simplest er det at måle afstandsbarrieren mellem to lande  $i$  og  $j$ :

$$AFSTANDSBAR_{i,j,t}^{BILAT} = \left( \frac{AFST_{i,i}}{AFST_{i,i}} \right)^{-\hat{\beta}_{3,t}} \quad (B.3)$$

For at opnå et udtryk for den gennemsnitlige effekt af afstandsbarrierer i land  $i$ , må der findes en gennemsnitlig international afstand til land  $i$ . Det er oplagt, at sammenveje afstandene til land  $i$  med brug af BNP i land  $j$  som vægte, idet et sådant mål netop afspejler afstanden fra en gennemsnitlig udenlandsk producent til land  $i$ .

$$AFSTANDSBAR_{i,t}^{NAT} = \left( \frac{AFST_{i,i}}{\sum_{j \neq i} \alpha_{j,t} AFST_{i,j}} \right)^{-\hat{\beta}_{3,t}} \quad \text{hvor } \alpha_{j,t} = \frac{BNP_{j,t}}{\sum_{j \neq i} BNP_{j,t}} \quad (B.4)$$

Endelig kan effekten af den samlede handelsbarriere beregnes, det vil sige det estimerede forhold mellem indenlandsk salg og import fra et land med samme udbudspotentiale som følge af grænse- og afstandsbarrierer:

$$TOTBAR_{i,t} = GRÆNSEBAR_t \cdot AFSTANDSBAR_{i,t} \quad (B.5)$$

# Two Tests of Divorce Behaviour on Danish Marriage Market Data

Michael Svarer

Department of Economics, University of Aarhus, E-mail: [msvarer@econ.au.dk](mailto:msvarer@econ.au.dk)

*SUMMARY: In this paper I provide a test of the two dominating theories of divorce. The first theory suggests that deviations between expected and realised utility of marriage drives divorce. This hypothesis is tested in terms of earnings and the results show interesting asymmetries between men and women. The second theory suggests that learning about match quality is crucial. This theory is tested by investigating whether couples that cohabited prior to marriage are more likely to endure. The presented results suggest they are.*

---

## 1. Introduction

There are a number of reasons why individuals should be better off in unions than alone. A (presumably non exhaustive) subset of these reasons are economic by nature. A union allows, among other things, for division of labour, economies of scale, coordination of investments, sharing of collective goods, and risk pooling, Weiss (1997). Given that individuals acknowledge these advantages and coordinate actions accordingly it is relevant to investigate why individuals leave unions (henceforth denoted divorce).

In the economics literature two main ideas have been followed to explain why couples divorce. The first, initially formulated by Becker et al. (1977), argues that it is deviations between expected and realized output from marriage that trigger divorce. Following this line of argument, unexpected events are crucial determinants of divorce. The second idea argues that a reason for divorce is attributed to uncertainty about the quality of the current match and other union related characteristics of the partner. At the time of union formation the two spouses have only limited information on the determinants of the gains from the union. As time passes, new information on the quality of the match and on the outside options of each partner is accumulated, and the couple then decides whether to dissolve the partnership or to continue the marriage (see e.g. Weiss (1997),

---

I thank the Danish National Research Foundation for support through its grant to Center for Applied Macroeconometrics (CAM). I thank Birgitte Højklint for reading the manuscript.



Burdett and Coles (1997), and Brien et al. (2001)). This idea of learning about match-specific quality is closely related to Jovanovic's (1979) model of job turnover. Thus, the implications overlap in the sense that the divorce risk – like the exit-rate out of employment – (eventually) exhibits negative duration dependence. The longer a relationship has existed the lower the risk that a break-up will occur. In addition, couples who started out as cohabitators have the advantage of gathering information about their partner before marrying. Hence, only good matches should evolve into marriage and marriages preceded by cohabitation would have a lower divorce risk.

In the present paper I conduct an empirical test of the two theories of divorce discussed above. First, I investigate whether couples in which one of the partners experience deviations from expected earnings are more likely to dissolve. Second, I investigate whether marriages where the couple cohabited prior to marriage are more likely to endure. Support for the latter has already been found in Svarer (2004). The contribution of the current analysis is to test whether this result still holds when the decision to cohabit is endogenized in the empirical analysis.

The investigation is based on a register based data set collected by Statistics Denmark covering a subset of all individuals born between January 1, 1955 and January 1, 1965. The individuals are followed from 1980 to 1995 on an annual basis. I find interesting asymmetries between men and women in terms of the effect of deviations in actual and expected earnings on divorce risk. Specifically, I find that the divorce risk is lower (higher) if the husband experience higher (lower) earnings than expected (where the expectation is based on the outcome of a Mincer earnings regression). On the other hand, I find that deviations between expected and realised earnings of the wife destabilize a marriage. With respect to premarital cohabitation, I find that premarital cohabitation decreases the subsequent divorce hazard. This result is also found in Svarer (2004). What is new to this study is that it also identifies a positive selection effect. The selection effect suggests that couples who are more prone to premarital cohabitation are also more prone to divorce. However, the learning effects, which enable the partners to learn about the quality of the partnership before marriage, dominate in terms of sustaining marriage.

The paper is organized as follows. In Section 2 I briefly discuss some theoretical issues concerning divorce. The data are presented in Section 3. Section 4 contains the econometric model. In Section 5 I present the results, and in Section 6 I conclude.

## **2. Theoretical considerations**

The theoretical question facing researchers who want to analyse separations is how it can be optimal to enter a partnership at one time and then again optimal to leave this partnership at some later date. In this article I are specifically interested in investigating this question in the context of the marriage market. Following Burdett and Coles

(1998), the causes of divorce are discussed under three headlines: learning effects, variations in match productivity, and variations in outside options values.

Learning effects relate to the situation where the individuals in the marriage at the start of the relationship are incompletely informed about the quality of their partner. As the marriage evolves more and more information about the partner is revealed, and sufficient unfavorable information may lead to divorce. The quality of the match is in the terminology of Jovanovic (1979) *a pure experience good*. Both partners receive, as the marriage evolves, noisy signals about each other, and when new information arrives they choose whether to resolve the marriage or not. Initially, the stock of information is low, and bad shocks will have a larger effect than later in the marriage when more experience has been collected. On the other hand, the variance on the match quality is higher in the beginning giving more scope for a big positive realization. This implies that the – theoretically derived – divorce hazard is increasing initially, whereafter it declines with marriage duration. Brien et al. (2001) extend Jovanovic's model to the marriage market. Both models assume from the outset that the characteristics of the two spouses are known and constant over time, and that information of the initially unobserved match-specific quality component evolves over time. In this set-up it is clear that low realizations of the match-specific component will lead to divorce as only the inside option value decreases. An interesting prediction from these models is that premarital cohabitation increases the survival probability of marriage, since the cohabitation period enables information gathering, and only good realized matches will evolve into marriage.

Although individuals are fully informed about the quality of the match,<sup>1</sup> the flow payoff to the match may change over time. This mechanism has been the building block of divorce models by Becker et al. (1977), Weiss and Willis (1997), and Ermisch & Böheim (2001). Here it is argued that it is the deviation between expected and realized utilities/payoffs that triggers divorce. Since partners were matched based on predicted characteristics both positive and negative surprises could lead to divorce. If the gains to marriage are substantial, small shocks will not lead to divorce. Individuals, of course, acknowledge that the probability of divorce is higher the lower the expected gain from marriage, and sort on characteristics that enhance the expected gain from marriage. Becker (1973, 1974) provides an extensive analysis of optimal marital sorting that explains the predominance of positive assortative mating with respect to personal characteristics such as education, intelligence, age etc. Positive assortative mating is optimal when the traits are complements in the production function. When the traits are substitutes, production is increased through negative assortative mating (see also Burdett and Coles (1997) and Smith and Shimer (2000)).

---

1. This amounts to assuming that the couple at the time of partnership formation immediately obtains all information about the unobserved match component.

Even though individuals are fully informed about the match quality and the flow payoff to the match is constant over time, the payoff to separating may be time varying. The accumulation of marriage-specific capital, like children, property and information about the spouse during the marriage, clearly decrease the payoff to separating. The accumulation of marriage-specific capital increases the inside option value to marriage and decreases the outside options.

In sum, the theoretical models provide several testable implications. My main interest in this paper is to test the first two theories. A test of learning effects follows Svarer (2004) and is based on a test of whether people who premaritally cohabit are more likely to have longer marriages. To test whether deviations between expected and realised utility trigger divorce I investigate the association between divorce risk and the gap between expected and realised earnings. There is no direct information on expected earnings in the data sets, so in order to assess this point I estimate the expected earnings in a given year for each individual based on a Mincer earnings regression. Based on the expected earnings I investigate whether differences are associated with divorce behaviour.

### 3. Data

The data used in this study come from IDA (Integrated Database for Labour Market Research) created by Statistics Denmark. The information comes from various administrative registers that are merged in Statistics Denmark. The IDA sample used here contains (among other things) information on marriage market conditions for a randomly drawn sub-sample of all individuals born between January 1, 1955 and January 1, 1965. The individuals are followed from 1980 to 1995. The data set enables us to identify individual transitions between different states on the marriage market on a yearly basis. The information about civil status is based on the individual's situation on December 31 each year and is derived from household information. This means that only individuals sharing the same address are identified as cohabiting or married. If two individuals are sharing a flat, say, without being a couple it will still count as cohabitation in the data. The only way I can ascertain that individuals living together actually are partners is to consider married couples only. In this study I therefore restrict focus to marriages. Of course, married couples who are not living together will be registered as single, but this type of relationships are likely to be low in number. If there is a break in a marriage, e.g. if I observe a couple to be married in 1987, to live as single individuals in 1988, and then as a married couple again in 1989, I disregard the break, and contribute the intervening spell to measurement error.<sup>2</sup>

---

2. For these couples the intervening spell could imply that they actually divorce and then remarry (what Bracher et al. (1993) label »The Elisabeth Taylor syndrome«). The data, however, do not allow for correct identification of this type of behavior.

Table 1. Number of marriages.

Number of marriages	Number of persons	Percentage
1	10409	95.7
2	459	4.2
3	12	0.1
4	2	0.0

The information used in the analysis is gathered in the following way: I observe the individuals in 1980 where I have information about various personal characteristics and marriage market status. For each subsequent year I observe a new stream of data for the individuals. If the individual enters a relationship I also observe the personal characteristics of the partner. Since I am interested in marriages, and especially the personal characteristics during the marriage, I disregard left censored marriages. The individuals were between 15 and 25 years of age in 1980, and since I disregard individuals already married, the first marriage of the unit of observation is observed for the majority of the individuals (some may, of course, already be divorced). The partners are not restricted by age, and it is therefore more likely that they have experienced more marriages at the time I observe them.

In the sample there are 10822 marriages during the period. The number of marriages per person is distributed as presented in Table 1.

Only around 4 percent experience more than one marriage during the sample period.

### 3.1. Explanatory variables

Table 2 contains descriptive statistics of the variables used in the analysis. The mean and standard deviation of the explanatory variables are the calculated at start of the marriage. All variables are, however, allowed to vary during the marriage spell.

The main variables of interest are a dummy indicating whether the couple premaritally cohabited and four variables representing differences in expected and realised earnings. The latter variables are constructed in the following manner. First, an earnings equation is estimated by OLS. In this first step, earnings are regressed on education, age, geographical belonging, number of kids, level of unemployment and sickness. In the second step, the predicted earnings in a given year are compared to the actual earnings to define four variables:

$$\text{Positive deviation}_i = (\text{Earnings}_{i,t} - \widehat{\text{Earnings}}_{i,t}) \quad \text{if } \text{Earnings}_{i,t} > \widehat{\text{Earnings}}_{i,t},$$

$$\text{Negative deviation}_i = (\text{Earnings}_{i,t} - \widehat{\text{Earnings}}_{i,t}) \quad \text{if } \text{Earnings}_{i,t} < \widehat{\text{Earnings}}_{i,t}$$

where  $i$  = husband, wife.

*Table 2. Descriptive statistics of variables in sample by type of marriage.*

	Mean	Std. dev.
<i>Education of Husband</i>		
Vocational	0.27	
Short	0.04	
Medium	0.04	
Long	0.05	
<i>Differences if expected and realised earnings (in DKK 1980)</i>		
Positive deviation, wife	793	7711
Negative deviation, wife	399	6156
Positive deviation, husband	1374	22108
Negative deviation, husband	644	7187
<i>Age at marriage</i>		
Wife < 21	0.20	
Wife between 21-25	0.23	
Wife between 26-30	0.08	
Husband < 21	0.10	
Husband between 21-25	0.23	
Husband between 26-30	0.12	
Husband more than 4 years older	0.23	
Wife more than 4 years older	0.03	
<i>Unemployment and sickness</i>		
Wife's yearly degree of unemployment	0.13	0.23
Husband's yearly degree of unemployment	0.07	0.18
Sickness, wife	0.18	
Sickness, husband	0.12	
<i>Children</i>		
Own child younger than 2	0.26	
Own child between 3 and 6	0.15	
Own child between 7 and 9	0.05	
Own child between 10 and 17	0.04	
Presence of step children	0.13	
Premarital birth	0.26	
<i>Other characteristics</i>		
Province	0.53	
Duration of pre-marital cohabitation	2.82	2.78
Mean duration of marriage (in years)	6.97	3.90
Number of observations		10882

In addition, we have dummies for educational attainment. The reference category is no education beyond primary school. Since some individuals may still be studying, I observe the current education at the time of observation. The educational variables are therefore also allowed to be time-varying. Next, I have six variables on children; an indicator variable for the presence of a child born to the couple younger than 3 years of

age, between 3 and 6 years of age, between 7 and 9 years of age, between 10 and 17 years of age, an indicator variable for the presence of stepchildren, and finally an indicator variable that takes the value 1 if the couple had a child together before marriage. I also include variables measuring the age of the couple and the age difference. The variable, sickness, is an indicator variable taking the value 1 if the individual receives sickness benefits during the year. As a general rule sickness benefits are received if a person has a spell of illness for more than 13 weeks. I distinguish between individuals living in the Copenhagen metropolitan area and individuals living in the provinces by the indicator variable province. Each individual's degree of unemployment during the year is defined as the number of hours of unemployment divided by the number of potential supplied working hours.

#### 4. Econometric specification

Our empirical model closely follows Lillard et al. (1995). That is, I specify a model of the risk of divorce which explicitly incorporates both the direct effect of premarital cohabitation and a potential self-selection of more divorce-prone individuals into premarital cohabitation. Our model differs slightly from Lillard et al.'s as our data are observed on a yearly basis. Hence, I model the transition out of marriage as a grouped duration model. In addition, I impose a different distributional assumption on the unobserved heterogeneity components. The correlation structure between the two processes; the transition out of marriage and the decision to cohabit before marriage, determines the endogeneity of cohabitation, i.e. the self-selection effect. Therefore, if unobserved factors both influence the individual risk of divorce and the decision to cohabit, then they are correlated and cohabitation is endogenous. Below I briefly sketch the empirical model employed in this study.

##### 4.1. Empirical model

The hazard function is specified as a mixed proportional hazard. That is, it is a product of a function of time spent being married (the baseline hazard), a function of observed time-varying characteristics,  $x^d$ , and a function of unobserved characteristics,  $v^d$ :

$$h(t|x_t^d, v^d) = \lambda(t) \cdot \varphi(x_t^d, v^d), \quad (4.1)$$

where  $\lambda(t)$  is the baseline hazard and  $\varphi(x_t^d, v^d)$  is the scaling function specified as  $\exp(x_t^{d'}\beta^d + v^d)$ . I follow Lillard et al. (1995) and assume that the unobserved component,  $v^d$ , is individual-specific to the unit of observation. The unit of observation is the person in our sample that is followed throughout the period from 1980 to 1995.

Since I only observe the transitions on the marriage market on a yearly basis, I specify a model for grouped duration data (see e.g. Kiefer (1990)). The marriage duration  $T$  is observed to lie in one of  $K$  intervals, with the  $k$ 'th interval being  $(t_{k-1}, t_k]$  and the convention  $t_0 = 0$  for  $k = 1, \dots, 15$ . The probability that the duration  $T$  for an individual with explanatory variables  $x_t$  is greater than  $t_k$  given that the duration is greater than  $t_{k-1}$  is given by:

$$\begin{aligned} P(T > t_k | T > t_{k-1}, x_k^d, v^d) &= \exp \left[ - \int_{t_{k-1}}^{t_k} h(t | x_t^d, v^d) dt \right] \\ &= \exp \left[ - \exp [x_k^d \beta^d + v^d] \cdot \Lambda_k \right] \end{aligned} \quad (4.2)$$

where  $\Lambda_{i,k} = \int_{t_{k-1}}^{t_k} \lambda_i(t) dt$ . The interval-specific survivor expression (4.2) is henceforth denoted  $\alpha_k$ . The probability of observing an exit out of marriage in interval  $k$ , conditional on survival until  $T > t_{k-1}$ , is consequently  $1 - \alpha_k$ . If I do not specify a functional form for the baseline hazard, the  $\Lambda_{i,k}$ s are just parameters to be estimated.

As argued by Bennett et al. (1988), the decision to cohabit might be correlated with the divorce risk. The notion of self-selection is modelled by specifying a simultaneous model of the selection process into cohabitation and the transition rate out of marriage. I define an indicator for the selection,  $C$ , taking the value 1 when an individual has cohabitated prior to marriage and 0 otherwise. The selection may depend on explanatory variables,  $x^c$ , and an unobserved component,  $v^c$ . The selection process is specified as a logit model, i.e.

$$P = \Pr(C = 1 | x^c, v^c) = \frac{\exp [x^c \beta^c + v^c]}{1 + \exp [x^c \beta^c + v^c]}.$$

The individual contribution to the likelihood function is then

$$\mathcal{L} = \int \int P^c (1 - P)^{1-c} (1 - \alpha_k)^j \alpha_k^{1-j} \prod_{l=1}^{k-1} \alpha_l g(v^d, v^c) dv^d dv^c, \quad (4.3)$$

where  $g(v^d, v^c)$  is the joint probability density function of the unobservables and where  $j = 1$  if the marriage is not right censored and 0 otherwise.  $g(v^c, v^d)$  is assumed to follow a  $2 \times 2$  discrete distribution. Uncompleted durations therefore only contribute with the survivor probabilities.

I normalize one of the support points in each of the cause-specific hazard functions to zero, since the baseline hazard acts as a constant term.

Identification of the correlation between the unobservable components is obtained through multiple occurrences of both marriage spells and premarital cohabitation status for the individuals, as suggested by identification results for duration models with multiple spells, cf. Honoré (1993). This identification approach has been used in a series of papers by Lillard and co-authors (see e.g., Panis and Lillard (1994), Lillard et al. (1995), and Upchurch et al. (2002)). Identification requires that I – for at least a subset of individuals – observe marriage spells both when the individual premaritally cohabited and when the individual did not. The intuition for identification is spelled out in Panis (2004). In terms of our application, his argument goes as follows: suppose I observe only one respondent over a long period of time during which she switches premarital cohabitation status. With a sample of one, there is no heterogeneity and no correlation across equations, such that the equations are independent. The effect of premarital cohabitation on exit rates from marriage is identified because of repeated observations on marriage spells and variations in premarital cohabitation status. More generally, conditional on heterogeneity, the equations are independent and identification rests on repeated outcomes with intraperson variation in premarital cohabitation status. In terms of intraperson variation in premarital cohabitation status 4% of the individuals in the sample are observed both as premarital cohabitators and not. Obviously, basing identification on 4% of the sample is not the most attractive nor powerful identification strategy. The author of this article were, however, not able to come up with an instrument that could explain premarital cohabitation and at the same time not influence divorce decisions.

## 5. Results

In Table 5.1 I report the maximum likelihood estimates for the parameters of the estimated model. Since I only observe that a divorce has occurred sometime within a given year, I use explanatory variables or time  $t - 1$  to explain the divorce hazard at time  $t$  in order to avoid that the value of a given characteristic is influenced by the divorce event.

Compared to Svarer (2004) I pay more attention to the possible endogeneity of premarital cohabitation in the divorce hazard. This is, however, not the only potential endogenous variable. Several of the included explanatory variables face the same issue of endogeneity. Weiss and Willis (1997) provide evidence of this with respect to earnings, Johnson and Skinner (1986) with respect to female labour supply, Lillard and Waite (1993) with respect to children, and Lillard and Panis (1996) with respect to health. The endogeneity problem of other variables than cohabitation will not be explicitly addressed in the empirical model. However, I keep the endogeneity issues in mind when drawing inference.



Table 3. Results from econometric model.<sup>a</sup>

	Cohabitation decision		Hazard of divorce	
	Coeff.	Std dev.	Coeff.	Std dev.
<i>Cohabitation</i>				
Couple has cohabited			-0.6970 <sup>b</sup>	0.1418
Duration of cohabitation			-0.1284 <sup>b</sup>	0.0181
<i>Deviation between predicted and realised earnings</i>				
Positive deviation, wife	0.6092 <sup>b</sup>	0.1254	0.2677 <sup>b</sup>	0.1032
Negative deviation, wife	-0.1647 <sup>b</sup>	0.0220	0.0611 <sup>b</sup>	0.0144
Positive deviation, husband	0.0518	0.1202	-0.3267 <sup>b</sup>	0.1070
Negative deviation, husband	-0.2525 <sup>b</sup>	0.0280	0.0386 <sup>b</sup>	0.0168
<i>Husband's education</i>				
Vocational	0.1255 <sup>b</sup>	0.0889	-0.3562 <sup>b</sup>	0.0723
Short	0.0233	0.1618	-0.1637	0.1361
Medium	-0.1380	0.1540	-0.4874 <sup>b</sup>	0.1522
Long	0.0158	0.1413	-0.7911 <sup>b</sup>	0.1588
<i>Children</i>				
Own children younger than 2	0.5469 <sup>b</sup>	0.1760	-0.4613 <sup>b</sup>	0.0545
Own children between 3 and 6	1.8239 <sup>b</sup>	0.1909	-0.0968 <sup>b</sup>	0.0583
Own children between 7 and 9	1.4347 <sup>b</sup>	0.1793	-0.1181 <sup>b</sup>	0.0692
Own children between 10 and 17	1.4310 <sup>b</sup>	0.1820	0.0243	0.0887
Step children	-1.3673 <sup>b</sup>	0.2042	0.5263 <sup>b</sup>	0.0906
Premarital birth to the couple	6.5812	8.8726	0.2771 <sup>b</sup>	0.0838
<i>Age</i>				
Wife between 15-20	-1.1765 <sup>b</sup>	0.1737	0.1272	0.1470
Wife between 21-25	-0.9160 <sup>b</sup>	0.1476	0.2637 <sup>b</sup>	0.1073
Wife between 26-30	-1.1010 <sup>b</sup>	0.1500	0.0634	0.0908
Husband between 15-20	0.8542 <sup>b</sup>	0.1748	0.4790 <sup>b</sup>	0.1775
Husband between 21-25	1.0460 <sup>b</sup>	0.1453	0.2819 <sup>b</sup>	0.1165
Husband between 26-30	0.7899 <sup>b</sup>	0.1334	0.2180 <sup>b</sup>	0.0911
Wife more than 4 years older	-0.8507 <sup>b</sup>	0.1481	0.1062	0.1363
Husband more than 4 years older	-0.0959 <sup>b</sup>	0.0798	0.2636 <sup>b</sup>	0.1073
<i>Sickness and unemployment</i>				
Sickness, wife	-0.0682	0.0930	0.2104 <sup>b</sup>	0.0631
Sickness, husband	-0.2924 <sup>b</sup>	0.0951	0.2617 <sup>b</sup>	0.0662
Unemployment degree, wife	-0.3725 <sup>b</sup>	0.1214	0.3900 <sup>b</sup>	0.0944
Unemployment degree, husband	-0.4902 <sup>b</sup>	0.1578	0.9638 <sup>b</sup>	0.1266
<i>Other characteristics</i>				
2 <sup>nd</sup> or higher order marriage	-1.0919 <sup>b</sup>	0.1612	0.0344	0.1722
Province	0.3462 <sup>b</sup>	0.0586	-0.3628 <sup>b</sup>	0.0623
$v_2^d$	-1.6469 <sup>b</sup>	0.3431		
$v_1^c$	1.7575 <sup>b</sup>	0.2132		
$v_2^c$	0.3664	0.2461		
$P1(v^d = 0, v^c = v_1^c)$	0.4960	0.3431		
$P2(v^d = v_2^d, v^c = v_2^c)$	0.5040 <sup>b</sup>	0.1150		
Number of observations			10882	

Notes: (a) The reference categories are: no education above mandatory education and over 30 years old. (b) significant at .05. (c) significant at .1.

### *5.1. Do deviations between expected and realised earnings trigger divorce?*

In the present data set there is no information that provides a clear cut answer to whether realised earnings were actually expected or not. As an approximation to this I use the predicted earnings from a Mincer earnings regression as a proxy for expected earnings and subsequently test whether larger differences between the predicted earnings and the realised earnings are related to divorce risk. The results in Table 3 suggest they are. Marriages where the husband experiences a positive income shock are more likely to survive. The opposite is the case if 1) the husband experiences a negative income shock or 2) if there are changes in the wife's earnings in either a positive or negative direction. For wives the effect of positive income differences have the highest impact on the divorce hazard. A mayor drawback of this analysis is that I treat earnings as an exogenous variable. There are good reasons to expect that earnings are endogeneous to divorce risk. Weiss (1997) argues that spouses' earnings and children are potentially endogenous to the divorce decision from a dynamic model of *defensive investments*. In the model, fertility and child quality requires inputs of time and money by parents in the first period, and child quality requires inputs of time and money by parents in the second period. Also, parents' wages in the second period depend on how much they worked in the first period due to learning-by-doing in the labour market. In the second period, new information is available which could induce the dissolution of the marriage. Hence, fertility in the first period, wages in the second period and the probability of divorce are jointly determined.

Weiss and Willis (1997) address the endogeneity of earnings explicitly. After endogenising earnings they find that if the husband experiences a positive surprise then the divorce probability decreases, whereas the divorce probability increases if the wife experiences a positive surprise in earnings. I do not endogenize earnings, but the results are similar.

That asymmetries exist in terms of earnings and behaviour of husbands and wives is also found by Lentz and Tranæs (2004). They find that the job finding effort of married men and women is affected by the income of their spouses in opposite directions. For women, spouse income influences job finding negatively, just as own wealth does: the more the man earns and the wealthier the woman is, the longer it takes for her to find a job. The contrary is the case for men, where spouse income affects job finding positively: the more the wife earns, the faster the husband finds a job. This is so despite the fact that greater own wealth also prolongs unemployment spells for men.

### *5.2. Premarital cohabitation and divorce risk*

I find that premarital cohabitation is negatively correlated with divorce risk. This result is not new (see Svarer (2004)). What is new, however, is that the finding still pre-

vails after endogenising the premarital cohabitation decision. As argued previously, our identification strategy is perhaps not the most powerful. Still the empirical model suggests that endogenising premarital cohabitation improves the fit of the model. A likelihood ratio test of the preferred model (Table 3) against a model where premarital cohabitation is not endogenised supports the first model. The LR statistics has a value of 13.96, the 95% critical value in the  $\chi^2$ -distribution with 4 degrees of freedom is 9.48. In addition the correlation between the two processes is estimated to be positive,<sup>3</sup> suggesting that individuals who cohabit prior to marriage have unobserved characteristics that make them more prone to divorce. This is in accordance with the selection effect suggested previously. However, the learning effects dominate the selection effect and premarital cohabitation is supporting longer marriages. Also, the longer the premarital cohabitation period has lasted the more stable is the subsequent marriage.<sup>4</sup> The results differ from practically all other studies in this field (see e.g. Blanc (1985) on Norwegian data, Balakrishnan et al. (1987) on Canadian data, Bennett et al. (1988), Trussell et al. (1992), and Hoem and Hoem (1992) all on Swedish data, Bracher et al. (1993) on Australian data, Lillard et al. (1995), Weiss and Willis (1997), and Brien et al. (2001) all on US data).<sup>5</sup> One of the main reasons for the divergence from the other studies is presumably the period which the data sets covers. This study is based on more recent data and in the present context this could be quite important. According to Kiernan (2000), cohabitation was largely statistically invisible prior to the 1970s. However, the form of cohabitation that we consider in this paper, i.e. cohabitation as a prelude to marriage, »...came to the fore during the 1960s in Sweden and Denmark, and during the 1970s in other Northern and Western European countries, North America, and Australia and New Zealand...« Kiernan (2000, p. 42). Contemporarily premarital cohabitation is the norm in a number of countries like the US, the UK, Norway,

---

3. It should be noted that the correlation is perfect. Due to the small amount of individuals with repeated marriage spells the empirical identification of the correlation structure is rather weak. This implies that in a discrete mixture model with two points of support in each equation I was not able to obtain valid estimates without restricting the correlation to be either 1, -1 or 0. The correlation of 1 is however not sensitive to different starting values and in that sense is a robust finding conditional on the identifying power of the model.

4. A referee suggested to endogenise time spent as cohabitators. This is clearly a very good point. An obvious empirical approach would be to model the period of premarital cohabitation as an additional duration process and then allow for correlation between this and the marriage equation. The main empirical obstacle preventing me from following this path is that many premarital cohabitation periods are left censored, i.e. they start before the observation period. Left censored duration data are quite complicated to model. This complexity increases with the inclusion of unobserved heterogeneity, which is required to identify the correlation between the two processes of interest. Accordingly, it is only the decision to cohabit that is endogenised in the model.

5. One exception is Georgellis (1996). He finds, based on British data from the General Household Survey collected in 1990-91, that the duration of premarital cohabitation and subsequent divorce risk is negatively correlated. He does not take the possible endogeneity of cohabitation into account.

Sweden, Finland, France, Austria, Switzerland, West Germany, the Netherlands, New Zealand, and Estonia. I would therefore expect that studies based on more up-dated data sets in these countries would support the finding of the present paper.

### 5.3. *Other results*

In this section the association between the other explanatory variables and the divorce risk is briefly discussed.

The educational level of the husband is negatively correlated with the divorce risk. That is, a higher degree of education monotonically decreases the instantaneous probability of divorce. This is in accordance with a number of other studies (e.g. Weiss and Willis (1997) and Georgellis (1996)) and confirms the simple correlation between divorce rates and education that is found to be negative Becker et al. (1977).

As Weiss (1997) argues, the decision to invest in children is presumably not independent of the probability of divorce. Lillard (1993) and Lillard and Waite (1993) specify and estimate a simultaneous equation system of the two processes. They confirm that the two decisions are negatively correlated. However, they still find that the first child born to the couple stabilizes the marriage, which corroborates Becker et al.'s (1977) notion of children constituting positive capital in the household. I find that own children born to the marriage have a stabilizing effect when they are younger than 3 years of age (for related results see Waite and Lillard (1991) and Weiss and Willis (1997)). This result could be driven by the lacking endogenization of fertility since couples that decide to have children might be those couples who experience higher match quality and therefore are less likely to split up anyway (cf. Lillard (1993)). In Svarer and Verner (2004) fertility is treated as endogenous in the divorce equation. This turns out to be crucial for the results. Svarer and Verner (2004) find that the couples who are more prone to divorce are less likely to have children. Once they treat fertility as an endogenous process to divorce the stabilizing effect of children disappears.

Whereas own children are considered to be positive capital in the household, stepchildren, however, represent negative capital in the household according to Becker et al. (1977). Our results support this since I find that stepchildren destabilize marriages. Similar results are found by White and Booth (1985).

Children born to the couple but out-of-wedlock increase the divorce risk. This correlation could be driven by the observation that premarital cohabitation increases the risk of unwanted pregnancy, which again could lead to an otherwise unwanted – and therefore more fragile – marriage.

The marriage market is characterized by search frictions. It takes time to locate the most suitable match. Empirically, this is supported by the result that *age at marriage* is found to be negatively correlated with divorce risk in many studies (e.g. Becker et al.

(1977) and Weiss and Willis (1997)). The inclusion of age at marriage did not give any significant results. Instead, I followed the suggestion by Lillard (1993) and included a time varying age variable. Of course, the duration of marriage and the age of the individual each change perfectly collinear with time. However, their separate effects can be distinguished because they begin at different initial values at the beginning of the current marriage. I find that the younger the husband is the more unstable the marriage. Hence, I confirm the idea that it takes time to locate a good partner.

I find that sickness and unemployment (insignificantly at the 5% level for wives) increase the divorce hazard. The results are not surprising if the Becker et al. (1977) theory prevails. As noted in Section 2, they argue that it is unexpected events (which sickness and unemployment could be considered to be) that lead to divorce. From the viewpoint of marriage as an institution this is rather unfortunate, since Weiss (1997) argues that one of the advantages of marriage is risk pooling, i.e. the partner provides insurance in case of bad shocks. The evidence in this paper (and in related papers e.g. Jensen and Smith (1990), Bracher et al. (1993), Kiernan and Mueller (1998), and Böheim (2000)) seems to suggest that the former effect is operating.

Lillard and Panis (1996) also notice that marriage and health are positively correlated in the sense that married persons (especially men) have substantially lower mortality rates. As possible explanations for this correlation they suggest that marriage protects its incumbents or that healthier men select themselves into marriage. The former idea should encourage men in poor health to marry sooner and stay married longer. They specify and estimate a simultaneous equations model for health and marital dissolution for men. They find that healthier men have a reduced hazard of divorce, and that men who for unmeasured reasons tend to be in good health also tend to have longer marriages.

There is, at the 5 percent level, no significant effect of being in 2nd or higher order marriage. A destabilizing effect were to be expected if divorce was stigmatizing. Becker (1998, p. 337) argues that individuals have a higher probability of divorce in second or higher order marriages, simply because »...the average divorced person can be presumed to be more quarrelsome and in other ways less pleasant than the average person remaining married, because an unpleasant temperament is one cause of divorce...«. Actually, in a model without unobserved heterogeneity the prior marriage indicator came out significantly positive. That is, however, an inconsistent estimate of the prior occurrence effect. Including heterogeneity gives a consistent estimate, and in this case, an insignificant effect on the divorce risk.

I have also included a rough measure on regional belonging. In most international studies it is found that individuals living in heavily populated areas are more likely to

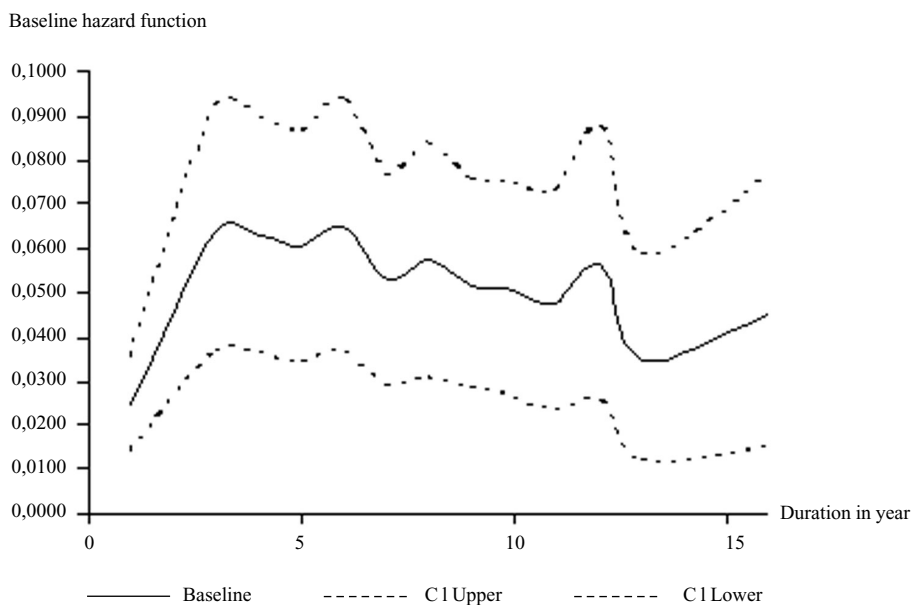


Figure 1. Baseline hazard.

experience divorce than individuals living in less populated areas. The same pattern is confirmed in the Danish data. There is for both men and women a significant negative effect on the divorce hazard of living outside the Copenhagen area.

In Figure 1 the baseline hazard function is depicted. The figure reveals that the divorce hazard increases sharply in the first couple of years after marriage. Hereafter, the hazard flattens with an insignificantly decreasing tendency.

Based on the discussion in Section 2 I would have expected the decreasing part of the hazard to be significant, since time spent in marriage increases the accumulation of marriage-specific capital, like children, property,<sup>6</sup> and information about the spouse, and consequently reduces the incentives to divorce. Likewise, I would expect that the presence of unobserved heterogeneity would lead to significant negative duration dependence, since divorce-prone couples exit earlier.

The same pattern as in Figures 1 is found by Lillard (1993) and Weiss and Willis (1997). The results suggest that something works counter the theoretical prediction. A likely explanation is that the correction for some of the marriage-specific capital accumulation by the inclusion of children tends to remove some of the expected duration dependence.

6. We were not able to control for property.

## 6. Conclusion

In this paper I have tested the two dominating theories of divorce on Danish marriage market data. First, I find that premarital cohabitation prolongs marriages. This result is identical to Svarer (2004). In addition this study identifies the presence of a selection effect suggesting that couples who premaritally cohabit are more prone to divorce. This effect is however dominated by – presumably – a learning effect.

The paper also shows that deviations between expected (or at least predicted) earnings and realised earnings are associated with divorce risk. For husbands who experience positive earnings shocks marriages are more stable, whereas the opposite is the case in situations with negative earnings shock. For wives both positive and negative shocks are associated with higher divorce risks. With the magnitude of positive earnings shocks being the highest.

Although both theories of divorce are supported by the empirical investigations presented in this paper, the results also indicate that there is room for improvements both in terms of treating endogeneous variables more seriously and in terms of finding better data to test whether shocks to the marriage are important for divorce decisions.

## References

- Balakrishnan, T. R., K. V. Rao, E. Lapierre-Adameyk and K. J. Krotki. 1987. A Hazard Model Analysis of the Covariates of Marriage Dissolution in Canada, *Demography*, 24(3), 395-406.
- Becker, G. 1973. A Theory of Marriage: Part I, *Journal of Political Economy*, 82 (4), 813-46.
- Becker, G. 1974. A Theory of Marriage: Part II, *Journal of Political Economy*, 82 (2, pt. 2), S11-S26.
- Becker, G. 1998. A Treatise on the Family, Harvard University Press.
- Becker, G., E. M. Landes and R. T. Michael. 1977. An Economic Analysis of Marital Instability, *Journal of Political Economy*, 85(6), 1141-87.
- Bennett, N. G, A. K. Blanc and D. E. Bloom. 1988. Commitment and the Modern Union: Assessing the Link Between Premarital Cohabitation and Subsequent Marital Stability, *American Sociological Review*, 53, 127-38.
- Blanc, A. K. 1985. The Effect of Nonmarital Cohabitation on Family Formation and Dissolution: A Comparative Analysis of Sweden and Norway, Ph.D. dissertation, Princeton University.
- Bracher, M., G. Santow, S. P. Morgan and J. Trussell. 1993. Marriage Dissolution in Australia: Models and Explanations, *Population Studies*, 47, 403-25.
- Brien, M. J., L. A. Lillard and S. Stern. 2001. Cohabitation, Marriage, and Divorce in a Model of Match Quality, Unpublished manuscript, University of Virginia.
- Böheim, R. 2000. Financial Surprises, Unemployment and Divorce, Draft, University of Essex.
- Burdett, K. and M. Coles. 1997. Marriage and Class, *Quarterly Journal of Economics*, 112, 141-68.
- Burdett, K. and M. Coles. 1998. Separating Cycles, *Journal of Economic Dynamics and Control*, 22, 1069-91.
- Burgess, S., C. Propper & A. Aassve. 1997. 'I Wont to be Alone'. Transitions to Independent Living, Marriage and Divorce among Young Americans, *Discussion Paper No. 1715*, University of Bristol.
- Ermisch, J. and M. Francesconi. 2000. Cohabitation in Great Britain: Not for Long, But Here to Stay, *Journal of the Royal Statistical Society*, 163 (2), pp 153-71.
- Ermisch, J. and R. Böheim. 2001. Partnership dissolution in the UK: the Role of Econo-

- mic Circumstances, *Oxford Bulletin of Economics and Statistics*, Vol. 63(2), 197-208.
- Georgellis, Y. 1996. Duration of First Marriage: Does Pre-Marital Cohabitation Matter?, *Applied Economic Letters*, 3, 217-19.
- Hoem, B. and J. Hoem. 1992. The Disruption of Marital and Nonmarital Unions in Contemporary Sweden In *Demographic Applications of Event History Analysis* eds. James Trussell, Richard Hankinson, and Judith Tilton, 61-93. Oxford England: Clarendon Press.
- Honoré, B. E. 1993. Identification Results for Duration Models with Multiple Spells, *Review of Economic Studies*, 60, 241-46.
- Jensen, P. and N. Smith. 1990. Unemployment and Marital Dissolution, *Journal of Population Economics*, 3, 215-29.
- Johnson, W. R. & J. Skinner. 1986. Labor Supply and Marital Separation, *American Economic Review* 76 (June), 455-69.
- Jovanovic, B. 1979. Job Matching and the Theory of Turnover, *Journal of Political Economy*, 87 (5 pt. 1), 972-90.
- Kiefer, N. M. 1990. Econometric Methods for Grouped Duration Data. In: *Panel Data and Labour Market Studies*, North-Holland.
- Kiernan, K. and G. Mueller. 1998. The Divorced and Who Divorces?, CASEpaper 7, LSE.
- Kiernan, K. 2000. European Perspectives on Union Formation, In: *The Ties That Bind*, (ed. Waite, L.), 40-58.
- Lentz, R. and T. Tranæs. 2004. Marriage, Wealth, and Unemployment Duration: A Gender Asymmetry Puzzle, Manuscript, Boston University.
- Lillard, L. A. and L. J. Waite. 1991. A Joint Model of Marital Childbearing and Marital Disruption', *Demography*, 4, 653-81.
- Lillard, L. A. 1993. Simultaneous Equations for Hazards – Marriage Duration and Fertility Timing, *Journal of Econometrics*, 56, 189-217.
- Lillard, L. A. and L. J. Waite. 1993. A Joint Model of Marital Childbearing and Marital Disruption, *Demography*, 4, 653-81.
- Lillard, L. A., M. J. Brien and L. J. Waite. 1995. Premarital Cohabitation and Subsequent Marital Dissolution: A Matter of Self-Selection?, *Demography*, 32 (3), 437-57.
- Lillard, L. A. and C. W. A. Panis. 1996. Marital Status and Mortality: the Role of Health, *Demography*, 33 (3), 313-27.
- Mare, R. D. 1991. Five Decades of Educational Assortative Mating, *American Journal of Sociology*, 97, 496-523.
- Panis, C. and L. Lillard. 1994. Health Inputs and Child Mortality: Malaysia, *Journal of Health Economics*, 13, 455-89.
- Panis, C. 2004. Microsimulations in the Presence of Unobserved Heterogeneity, Draft, RAND Corporation.
- Peters, E. H. 1986. Marriage and Divorce: Informational Constraints and Private Contracting, *American Economic Review*, 76 (3), 437-54.
- Shimer, R. and L. Smith. 2000. Assortative Matching and Search, *Econometrica*, 68 (2), 743-70.
- Svarer, M. 2004. Is Your Love in Vain? Another Look at Premarital Cohabitation and Divorce, *Journal of Human Resources*, 39 (2), 523-36.
- Svarer, M. and M. Verner. 2004. Do Children Stabilize Danish Marriages?, *CAM Working Paper* 2004-16.
- Trussell, J., G. Rodríguez, and B. Vaughan. 1992. Union dissolution in Sweden. In *Demographic Applications of Event History Analysis* eds. J. Trussell, R. Hankinson, and J. Tilton, 38-60. Oxford England: Clarendon Press.
- Upchurch, D. M., L. Lillard and C. Panis. 2002. Nonmarital Childbearing: Influences of Education, Marriage, and Fertility, *Demography*, 39 (2), 311-29.
- Weiss, Y. 1997. The Formation and Dissolution of Families: Why Marry? Who Marries Whom? And What Happens Upon Divorce, *Handbook of Population and Family Economics*, Elsevier Science B.V.
- Weiss, Y. and R. J. Willis. 1997. Match Quality, New Information, and Marital Dissolution, *Journal of Labor Economics*, 15(1), S293-S329.
- White, L. K. and A. Booth. 1985. The Quality and Stability of Remarriages: The Role of Stepchildren, *American Sociology Review*, 50, 689-98.



# Taxation of shareholder income and the cost of capital in an open economy: theory and applications to the Nordic countries

Peter Birch Sørensen

Department of Economics, University of Copenhagen, EPRU and CESifo,

E-mail: [peter.birch.sorensen@econ.ku.dk](mailto:peter.birch.sorensen@econ.ku.dk)

*SUMMARY: This paper discusses whether double taxation of corporate equity income distorts corporate investment in an open economy. The integration of capital markets has popularized the view that domestic personal taxes on dividends and capital gains have no impact on real investment incentives in a small open economy. The present paper argues that a personal tax on equity income in an open economy will in fact distort the required returns on shares in small companies which are not traded in the international stock market. However, if shareholders are granted a deduction for a risk-free rate of return, in line with the recent Norwegian tax reform, the tax will be neutral towards real investment when investors are well diversified. If they are not, a personal tax on the equity premium will tend to reduce the required return on shares in small companies, because of the income insurance effect of a tax with full loss offsets. This tendency will be amplified if the rate-of-return allowance is granted only to holders of shares in small companies, in line with current practice in Sweden.*

---

## **1. Capital market integration and double taxation of corporate income**

Equity income from the corporate sector is subject to corporation tax and to personal taxes on dividends and capital gains. Should this double taxation be alleviated to avoid serious distortions to corporate investment incentives? For decades academics and policy makers have debated this perennial issue. Advocates of the »old view« developed by Harberger (1962, 1966) and others stress the need to relieve the double taxation of dividends. This recommendation has had a profound influence on public policy. For example, Denmark taxes dividends at a lower rate than other capital income, and Finland and Germany only include a fraction of dividends in taxable income.

---

I am grateful to Diderik Lund, Vidar Christiansen and two anonymous referees for insightful comments. Any remaining errors or shortcomings are my own responsibility.

From 2006 Norway will only impose personal tax on dividends exceeding an imputed rate of return to the share, and Sweden has already implemented a similar tax rule for dividends from unquoted shares. Several other countries are granting full or partial credit for the underlying corporation tax against the personal tax on dividends.

However, proponents of the »new view« of dividend taxation have argued that there is little need to alleviate the double taxation of dividends. They point out that retained earnings are the main source of equity finance, and that the returns to shareholders accrue as capital gains when corporate investment is financed by retentions. Hence they conclude that double taxation is hardly a serious problem in a world where effective tax rates on capital gains on shares are typically quite low, see King (1974), Auerbach (1979), Bradford (1981), and Sinn (1987).

More recently, the international integration of capital markets has popularized the view that neither personal dividend taxes nor personal taxes on capital gains have any significant impact on corporate investment incentives in a small open economy, see, e.g., Boadway and Bruce (1992), Sørensen (1995), Andersson et al. (1998), and Fuest and Huber (2000). According to this view the marginal shareholder in a small economy with an open stock market is likely to be a foreign investor whose required return on shares is unaffected by personal taxes on domestic residents. If a residence-based personal tax on equity income makes shareholding less attractive to domestic investors, they will sell off (some of) their domestic shares to foreign investors who stand ready to buy domestic shares at prices determined on the world stock market. Thus, although they will influence the pattern of ownership, personal taxes on equity income will have no effect on the cost of equity finance for domestic corporations.

Many policy makers have found it hard to accept the postulate that personal taxes on equity income are irrelevant for real investment incentives in a small open economy. They argue that although this claim may be relevant for large firms that are listed on the stock exchange, the smaller unquoted firms do not, in practice, have access to the international equity market. If shares in small unquoted companies are imperfect substitutes for shares in large public corporations, say, because they have different risk characteristics, one might therefore think that personal taxes on dividends and capital gains will drive up the cost of equity for small companies.

In a recent paper Apel and Södersten (1999) have tried to take this concern seriously. They set up a mean-variance model of portfolio choice in a small open economy where domestic investors may hold three assets: risk-free bonds; shares in small companies which are not traded internationally; and shares in large corporations which are traded in the international stock market. Within this framework they study the effects of personal taxes on domestic holdings of traded shares and on the required rate of return on non-traded shares. Their most striking result is that, for plausible parameter values,

a residence-based personal tax on dividends and capital gains on shares may actually *reduce* the required rate of return on shares in small unquoted companies, because of the risk-reducing insurance effect of a tax with full loss offsets. This might seem to suggest that policy concern over the double taxation of corporate equity income is indeed unwarranted.

In this paper I extend the Apel-Södersten model to include an additional risky asset. Within this extended model I show that although personal taxes on shareholder income may well, on average, reduce the cost of capital for small companies, they will also distort the pattern of risk-taking and investment *within* the group of small companies by favouring investment in unquoted shares whose returns are highly correlated with the return on the market portfolio. Presumably this distortion is undesirable, so there is a case for designing tax rules that will be neutral towards the allocation of risk and investment. This paper shows that if shareholders are allowed a deduction for a risk-free rate of return on their shares so that tax is only levied on the equity premium, the tax system will be *neutral* provided investors are well-diversified. If shareholders in small companies are not well diversified, I show that a symmetric tax on the equity premium will tend to *reduce* the required rate of return on unquoted shares, and I argue that this non-neutrality is likely to be socially desirable. Finally, I demonstrate the non-neutrality of a tax system that allows a deduction for a risk-free rate of return only to holders of shares in unquoted companies. These results are relevant for the current Nordic tax policy debate, since Norway is about to introduce a rate-of-return allowance for all resident personal shareholders, while Sweden has already introduced such an allowance for holders of unquoted shares.

The next section sets up a model of portfolio choice in a small open economy. Section 3 uses the model to derive and discuss the distortionary effect of a personal tax on the full return to equity. Section 4 demonstrates the neutrality of an equity income tax which allows a deduction for a risk-free rate of return, assuming well-diversified investors. Section 5 analyzes the effect of a tax on the equity premium under incomplete diversification, and section 6 studies the effect of a selective rate-of-return allowance targeted at the holders of unquoted shares. The final section 7 discusses some limitations of the model.

## 2. A model of portfolio choice in a small open economy

Like Apel and Södersten (1999), I distinguish between shares that are traded in the international stock market («quoted shares») and shares that are only traded domestically («unquoted shares»). However, in order to study the effects of taxation on the allocation of risk within the sector of small companies, I extend the Apel-Södersten framework to include two types of unquoted shares. Thus the representative investor invests a fraction  $v_1$  of his initial wealth  $V_0$  in unquoted shares of type 1, yielding an

uncertain after-tax rate of return  $R_1$ ; a fraction  $v_2$  of his wealth in unquoted shares of type 2 with a risky net rate of return  $R_2$ , and a fraction  $v_3$  of his wealth in quoted shares generating an uncertain after-tax return  $R_3$ . The remaining fraction of initial wealth is invested in risk-free internationally traded bonds paying an after-tax interest rate  $i$ . The investor's wealth  $V$  at the end of the period will then be

$$V = [1 + v_1 R_1 + v_2 R_2 + v_3 R_3 + (1 - v_1 - v_2 - v_3)i]V_0 \quad (1)$$

I assume that the returns to the three risky assets are normally distributed and that the investor has a constant absolute risk aversion. As demonstrated in e.g., Silberberg and Suen (2001, p. 408) maximization of the investor's *expected* utility is equivalent to maximizing the utility function

$$U = E[V] - \frac{\hat{\rho}}{2} \cdot \text{var}[V] \quad (2)$$

where  $E[V] \equiv V^e$  is the expected level of wealth at the end of the period,  $\text{var}[V] \equiv E[(V - V^e)^2]$  is the variance of terminal wealth, and  $\hat{\rho}$  is the constant coefficient of absolute risk aversion. Denoting expected returns by the superscript 'e', it follows from (1) that

$$E[V] = [1 + v_1 R_1^e + v_2 R_2^e + v_3 R_3^e + (1 - v_1 - v_2 - v_3)i]V_0 \quad (3)$$

$$\text{var}[V] = (v_1^2 \hat{\sigma}_1^2 + v_2^2 \hat{\sigma}_2^2 + v_3^2 \hat{\sigma}_3^2 + 2v_1 v_2 \hat{\sigma}_{12} + 2v_1 v_3 \hat{\sigma}_{13} + 2v_2 v_3 \hat{\sigma}_{23})V_0^2 \quad (4)$$

where  $\hat{\sigma}_j^2$  is the variance of the rate of return on asset  $j$ , and  $\hat{\sigma}_{ij}$  is the covariance between the returns on assets  $i$  and  $j$ . Given the expected rates of return and the risk characteristics of the four asset types, the investor chooses his portfolio shares  $v_j$ ,  $j = 1, 2, 3$ , to maximize the utility function (2). The first-order conditions for the solution to this problem may be written as

$$R_1^e = i + \rho (v_1 \hat{\sigma}_1^2 + v_2 \hat{\sigma}_{12} + v_3 \hat{\sigma}_{13}) \quad (5)$$

$$R_2^e = i + \rho (v_2 \hat{\sigma}_2^2 + v_1 \hat{\sigma}_{12} + v_3 \hat{\sigma}_{23}) \quad (6)$$

$$R_3^e = i + \rho (v_3 \hat{\sigma}_3^2 + v_1 \hat{\sigma}_{13} + v_2 \hat{\sigma}_{23}) \quad (7)$$

where  $\rho \equiv \hat{\rho} V_0$  is the coefficient of *relative* risk aversion, measured at the investor's predetermined level of initial wealth.

**3. The distortionary effect of a personal tax on the full return to equity**

Suppose the government levies a symmetric residence-based personal tax at the rate  $t$  on all returns to shares (since we focus on shareholder income taxes, we do not explicitly model other taxes). If  $r_j$  and  $r_j^e$  are the actual and expected pre-tax returns to shares of type  $j$ , and if taxpayers are granted full loss offsets, we then have

$$R_j = (1 - t) r_j, \quad R_j^e = (1 - t) r_j^e, \quad j = 1, 2, 3 \tag{8}$$

$$\hat{\sigma}_j^2 \equiv E[(R_j - R_j^e)^2] = (1 - t)^2 \sigma_j^2, \quad j = 1, 2, 3 \tag{9}$$

$$\hat{\sigma}_{ij} \equiv E[(R_i - R_i^e)(R_j - R_j^e)] = (1 - t)^2 \sigma_{ij}, \quad j = 1, 2 \quad j = 2, 3 \tag{10}$$

where  $\sigma_j^2$  and  $\sigma_{ij}$  are the variances and covariances of the *pre-tax* rates of return, respectively. Inserting (8) through (10) into (5) through (7), we find

$$r_1^e = \frac{i}{1 - t} + \rho (1 - t) (v_1 \sigma_1^2 + v_2 \sigma_{12} + v_3 \sigma_{13}) \tag{11}$$

$$r_2^e = \frac{i}{1 - t} + \rho (1 - t) (v_2 \sigma_2^2 + v_1 \sigma_{12} + v_3 \sigma_{23}) \tag{12}$$

$$r_3^e = \frac{i}{1 - t} + \rho (1 - t) (v_3 \sigma_3^2 + v_1 \sigma_{13} + v_2 \sigma_{23}) \tag{13}$$

The interest rate  $i$  and the required expected rate of return  $r_3^e$  on quoted shares are exogenously given from the world market and hence unaffected by the shareholder income tax.<sup>1</sup> To see how the tax affects the required expected returns on unquoted shares, we solve (13) for  $\rho (1 - t)$  and insert the result into (11) and (12) to get

---

1. Strictly speaking, since  $i \equiv i^p (1 - \tau)$  is an after-tax interest rate, it is determined by the pre-tax interest rate  $i^p$  given from the world capital market and by the domestic personal tax rate on interest income,  $\tau$ . If the latter were always equal to the tax rate  $t$  on shareholder income, the terms  $i / (1 - t)$  in (14) and (15) would be replaced by  $i^p$ , and the shareholder income tax would not distort the cost of corporate finance (although the corporate income tax would still drive up the required pre-tax return on the company's real investment). However, since I wish to study the effect of granting double tax relief at the shareholder level, the focus of the present paper is on the isolated effect of personal taxes on income from shares, so the personal tax rate on interest income is taken as given and independent of  $t$ .

$$r_1^e = \frac{i}{1-t} + \beta_1 \left( r_3^e - \frac{i}{1-t} \right), \quad \beta_1 \equiv \frac{v_1 \sigma_1^2 + v_2 \sigma_{12} + v_3 \sigma_{13}}{v_3 \sigma_3^2 + v_1 \sigma_{13} + v_2 \sigma_{23}} \quad (14)$$

$$r_2^e = \frac{i}{1-t} + \beta_2 \left( r_3^e - \frac{i}{1-t} \right), \quad \beta_2 \equiv \frac{v_2 \sigma_2^2 + v_1 \sigma_{12} + v_3 \sigma_{23}}{v_3 \sigma_3^2 + v_1 \sigma_{13} + v_2 \sigma_{23}} \quad (15)$$

The variables  $\beta_1$  and  $\beta_2$  in (14) and (15) may be either smaller or greater than unity, so a personal tax on the return to equity will not necessarily drive up the required returns on unquoted shares, as already pointed out by Apel and Södersten (1999). To provide some intuition for this result, consider an investor with a well-diversified portfolio where unquoted shares carry a relatively small weight. In this case where  $v_1 \rightarrow 0$  and  $v_2 \rightarrow 0$ , it follows from (14) and (15) that

$$r_1^e = \frac{i}{1-t} + \beta_1 \left( r_3^e - \frac{i}{1-t} \right), \quad \beta_1 \approx \frac{\sigma_{13}}{\sigma_3^2} \quad (16)$$

$$r_2^e = \frac{i}{1-t} + \beta_2 \left( r_3^e - \frac{i}{1-t} \right), \quad \beta_2 \approx \frac{\sigma_{23}}{\sigma_3^2} \quad (17)$$

The variance of the return on the »market portfolio« tends towards  $\sigma_3^2$  when  $v_1 \rightarrow 0$  and  $v_2 \rightarrow 0$ . Our variables  $\beta_1$  and  $\beta_2$  then become roughly equal to the »beta« known from the conventional Capital Asset Pricing Model, as indicated in (16) and (17). For unquoted shares with a relatively high covariance with the market portfolio ( $\beta_j > 1$ ), the required risk premium will be high, because the holding of such shares adds to the aggregate risk on the investor's portfolio. For such relatively risky shares the benefit from the income insurance offered by a symmetric shareholder income tax (with full loss offset) outweighs the fact that the tax reduces the average net rate of return relative to the net return on safe assets. Accordingly, it follows from (16) and (17) that the tax will *reduce* the required return on unquoted shares when  $\beta_j > 1$ .<sup>2</sup> Conversely, for unquoted shares with a relatively low  $\beta_j (< 1)$ , the tax will drive up the required return. A personal tax on the full return to shares will thus distort the pattern of risk-taking and the cost of capital within the sector of small companies whose shares are not traded on the international stock market.

---

2. The mean-variance model of Apel and Södersten (1999, p. 85) also implies that  $\sigma_{j3}/\sigma_3^2 > 1$  ( $j = 1, 2$ ) is a necessary and sufficient condition for  $dr_j^e/dt < 0$ , when unquoted shares have a negligible weight in the investor's portfolio.

This distortionary effect of taxation will also prevail in the more likely case where investors are less than perfectly diversified. From (14) and (15) we see that unquoted shares with different risk characteristics will have different  $\beta$ -values, in which case the personal shareholder income tax  $t$  will have a different impact on the cost of capital for different types of unquoted companies.

**4. A personal tax on the equity premium: Neutrality of the Norwegian tax regime in a world of well-diversified investors**

I will now show that this distortion may be eliminated if the shareholder income tax is levied only on the *equity premium*, defined as the excess of the rate of return on shares over the after-tax interest rate on risk-free bonds. Such a tax was proposed in Sørensen (2005a) and will be introduced in Norway from 2006.

When shareholders are allowed a deduction for their opportunity cost of investment, given by the risk-free rate of return, the actual and expected after-tax returns on shares will be

$$R_j = r_j - t (r_j - i), \quad R_j^e = r_j^e - t (r_j^e - i), \quad j = 1, 2, 3 \tag{18}$$

Note that the tax system in (18) is assumed to be *symmetric*: if shareholders fail to obtain a return equal to the risk free rate, they can deduct their »loss«  $i - r_j$  against income from other shares (or they can carry their tax loss forward with interest so that the deduction maintains its present value). Under the tax regime (18) the variances and covariances of the net rates of return will still be given by (9) and (10), as the reader may easily verify. Substitution of (9), (10) and (18) into (5) through (7) then yields

$$r_1^e = i + \rho (1 - t) (v_1 \sigma_1^2 + v_2 \sigma_{12} + v_3 \sigma_{13}) \tag{19}$$

$$r_2^e = i + \rho (1 - t) (v_2 \sigma_2^2 + v_1 \sigma_{12} + v_3 \sigma_{23}) \tag{20}$$

$$r_3^e = i + \rho (1 - t) (v_3 \sigma_3^2 + v_1 \sigma_{13} + v_2 \sigma_{23}) \tag{21}$$

Solving (21) for  $\rho (1 - t)$  and substituting the result into (19) and (20), we find that

$$r_1^e = i + \beta_1 (r_3^e - i) \tag{22}$$

$$r_2^e = i + \beta_2 (r_3^e - i) \tag{23}$$

where  $\beta_1$  and  $\beta_2$  are defined in (14) and (15). Recall that when portfolios are well-diversified, we have  $\beta_j \rightarrow \sigma_{j3} / \sigma_3^2$ . Thus it follows from (22) and (23) that a shareholder

income tax with a rate-of-return allowance will have a negligible impact on the required return on unquoted shares when these shares only carry a small weight in investor portfolios. This is the basis for the claim that the tax will be approximately neutral when investors are well-diversified.<sup>3</sup>

The idea that a capital income tax which allows a deduction for the opportunity cost of finance will be neutral has previously been elaborated by King (1975), Boadway and Bruce (1984), Fane (1987), Devereux and Freeman (1991), Bond and Devereux (1995), and Lund (2002), among others. However, all of these authors assumed that the deduction for the imputed return on equity would be granted at the *corporate* level. But as Bond (2000) points out, such a corporation tax on »above-normal« profits would raise the relative tax burden on the most profitable companies. While this might not distort investment in a closed economy, in open economies a tax on supernormal profits would affect the international location decisions of multinational companies earning mobile rents. The analysis in this section suggests that this distortion may be avoided if the deduction for an imputed return on equity is instead granted at the *shareholder* level.

Although he does not explicitly analyse a tax system with a rate-of-return allowance, the closed-economy model developed by Gordon (1985) also implies that a personal capital income tax which exempts the risk-free rate of return will be neutral. However, in the Gordon model tax neutrality requires that tax revenues be transferred back to taxpayers in a lump sum fashion. In practice government transfers are not lump sum, and a substantial part of tax revenue serves to finance public goods provision. While some publicly provided goods are substitutes for private consumption, other public goods are complementary to private goods. As an intermediate benchmark, the model in this paper implicitly assumes that tax revenues finance public goods which enter investor utility functions in an additively separable manner.

### **5. The Norwegian tax regime in general equilibrium with incomplete portfolio diversification**

In practice the owners of small unquoted companies often invest a large fraction of their wealth in their own company. This does not necessarily invalidate the neutrality result derived in the previous section as long as the *marginal* shareholder in the firm (whose required return determines the cost of equity finance) is well diversified. Nevertheless, it is also of interest to analyse the case where investors are less well di-

---

3. When  $v_1$  and/or  $v_2$  are significantly above zero, the tax will no longer be neutral, since it will affect  $\beta_1$  and  $\beta_2$  via its impact on the portfolio shares  $v_j$ . Hence the neutrality result holds only in the case of well-diversified investors where  $v_1$  and  $v_2$  tend to zero, since in this case the portfolio shares drop out of the expressions for  $\beta_j$ . The next section will deal with the case of incomplete portfolio diversification.



verified. For this purpose, it is useful to move from the level of the individual investor to the macro level. If  $K_j$  is the aggregate capital stock invested in an unquoted company of type  $j$  at the start of the period;  $\pi_j$  is the rate of profit on this capital stock (after payment of corporation tax), and  $q_j$  is the ratio of the initial market value to the replacement value of the capital stock (Tobin's average  $q$  for unquoted companies), we have

$$r_j = \frac{\pi_j}{q_j}, \quad r_j^e = \frac{\pi_j^e}{q_j}, \quad j = 1, 2 \quad (24)$$

$$\sigma_j^2 \equiv E[(r_j - r_j^e)^2] = \frac{\tilde{\sigma}_j^2}{q_j^2}, \quad \tilde{\sigma}_j^2 \equiv E[(\pi_j - \pi_j^e)^2], \quad j = 1, 2 \quad (25)$$

$$\sigma_{j3} \equiv E[(r_j - r_j^e)(r_3 - r_3^e)] = \frac{\tilde{\sigma}_{j3}}{q_j}, \quad \tilde{\sigma}_{j3} \equiv E[(\pi_j - \pi_j^e)(r_3 - r_3^e)], \quad j = 1, 2 \quad (26)$$

$$\sigma_{12} \equiv E[(r_1 - r_1^e)(r_2 - r_2^e)] = \frac{\tilde{\sigma}_{12}}{q_1 q_2}, \quad \tilde{\sigma}_{12} \equiv E[(\pi_1 - \pi_1^e)(\pi_2 - \pi_2^e)] \quad (27)$$

where  $\pi_j^e$  is the expected profit rate. At the macro level  $v_1 \equiv q_1 K_1 / V_0$ ,  $v_2 \equiv q_2 K_2 / V_0$  and  $v_3 \equiv S_3 / V_0$ , where  $S_3$  denotes the aggregate stock of quoted shares held by domestic residents. Using these definitions along with (24) through (27), and remembering that  $\rho \equiv \hat{\rho} V_0$ , we may rewrite (19) through (21) as

$$\pi_1^e = i q_1 + \hat{\rho} (1 - t) (K_1 \tilde{\sigma}_1^2 + K_2 \tilde{\sigma}_{12} + S_3 \tilde{\sigma}_{13}) \quad (28)$$

$$\pi_2^e = i q_2 + \hat{\rho} (1 - t) (K_2 \tilde{\sigma}_2^2 + K_1 \tilde{\sigma}_{12} + S_3 \tilde{\sigma}_{23}) \quad (29)$$

$$r_3^e = i + \hat{\rho} (1 - t) (S_3 \sigma_3^2 + K_1 \tilde{\sigma}_{13} + K_2 \tilde{\sigma}_{23}) \quad (30)$$

In the short run considered here, the initial capital stocks in unquoted companies are predetermined. Moreover,  $\pi_1^e$  and  $\pi_2^e$  are given by the state of expectations, and  $i$  and  $r_3^e$  are exogenously given from abroad. However, even in the short run investors can adjust their aggregate holding of quoted shares by buying or selling on the international stock market. Equation (30) then determines the equilibrium value of  $S_3$  which

may be inserted into (28) and (29) to find the short run equilibrium values of  $q_1$  and  $q_2$  (and hence the values of  $r_1^e \equiv \pi_1^e/q_1$  and  $r_2^e \equiv \pi_2^e/q_2$ ). If an increase in the tax rate  $t$  reduces the value of  $q_j$ , it will drive up  $r_j^e$ , thereby reducing the real investment activity of unquoted companies over time. Using (24) through (30) and the formula for the coefficient of correlation  $c_{ij} = \sigma_{ij}/\sigma_i\sigma_j$ , one can show that

$$\frac{dr_1^e}{dt} = \left( \frac{\rho\sigma_1^2 r_1^e}{i} \right) \left[ v_1(c_{13}^2 - 1) + v_2 \left( \frac{\sigma_2}{\sigma_1} \right) (c_{13}c_{23} - c_{12}) \right] \quad (31)$$

$$\frac{dr_2^e}{dt} = \left( \frac{\rho\sigma_2^2 r_2^e}{i} \right) \left[ v_2(c_{23}^2 - 1) + v_1 \left( \frac{\sigma_1}{\sigma_2} \right) (c_{13}c_{23} - c_{12}) \right] \quad (32)$$

Since the coefficient of correlation  $c_{ij}$  cannot exceed 1 numerically, the terms  $(c_{ij}^2 - 1)$  in (31) and (32) are generally negative, whereas the term  $(c_{13}c_{23} - c_{12})$  cannot be signed a priori. However, if all share returns are positively correlated, a *sufficient* (but not necessary) condition for negativity of the derivatives in (31) and (32) is that

$$c_{12} \geq c_{13} \quad \text{or} \quad c_{12} \geq c_{23} \quad (33)$$

In other words, if the returns to the two unquoted shares are at least as correlated as the returns to quoted and (one of the) unquoted shares, we can be sure that a symmetric tax on the equity premium will *reduce* the cost of capital for unquoted companies when marginal investors are not fully diversified. Again, this is due to the insurance effect of the tax which reduces the riskiness of investment from the viewpoint of private investors.

Note that (31) and (32) are consistent with our earlier neutrality result: if the market value of the assets invested in unquoted firms is small relative to aggregate investor wealth so that  $v_1$  and  $v_2$  are close to zero,  $dr_j^e/dt$  will also be close to zero, and the shareholder income tax will be approximately neutral towards real investment decisions.

The tendency for a symmetric tax on the equity premium (with full loss offsets) to stimulate risk-taking is well known from the literature; see, e.g., Sandmo (1977 and 1989). This non-neutrality may well be socially desirable, since one would expect that the owners of small companies are inclined to take too little risk from a social perspective when they have failed to diversify their portfolios. However, to make reliable welfare judgements on this issue, one should explicitly model the factors (such

as asymmetric information and adverse selection) which make risk sharing with outside investors difficult.<sup>4</sup>

### 6. The Swedish tax regime: A rate-of-return allowance for unquoted shares

A popular view in the Nordic tax policy debate is that double taxation of corporate equity income drives up the cost of capital for small companies, whereas it does not affect capital costs for large companies with access to the international stock market. To limit the revenue loss from the rate-of-return allowance considered above, it might then be natural to offer the allowance only to holders of *unquoted* domestic shares. Motivated by this line of reasoning, the Swedish government has introduced a deduction for an imputed risk-free rate of return for holders of Swedish shares that are not listed on the stock exchange. In terms of our model, the Swedish tax rules thus imply that

$$R_j = r_j - t(r_j - i) \Rightarrow R_j^e = r_j^e - t(r_j^e - i), \quad j = 1, 2 \quad (34)$$

$$R_3 = (1 - t)r_3 \Rightarrow R_3^e = r_3^e(1 - t) \quad (35)$$

Our previous analysis suggests that if a shareholder income tax with a general rate-of-return allowance for all shares is neutral (when investors are well-diversified), a selective allowance granted only to holders of unquoted shares will be non-neutral. To demonstrate that this is indeed the case, we insert (34) and (35) into (5) through (7) and find that the investor's first-order conditions are now given by (19), (20) and (13). Using (24) through (27), equations (19) and (20) may be rewritten as (28) and (29), and (13) may be written as

$$r_3^e = \frac{i}{1 - t} + \hat{\rho}(1 - t)(S_3\sigma_3^2 + K_1\tilde{\sigma}_{13} + K_2\tilde{\sigma}_{23}) \quad (36)$$

Solving (36) for  $S_3$ , inserting the result into (28) and (29), and using (24) through (27), we find that

---

4. For an interesting contribution along these lines, see Konrad and Richter (1995). In a model where asymmetric information and adverse selection induce entrepreneurs to bear risks which could otherwise have been spread in the capital market, they find that a symmetric capital income tax on the full rate of return will not necessarily improve the allocation of risk. However, they do not analyze the effects of a tax which only falls on the equity premium.

$$\frac{dr_1^e}{dt} = \left( \frac{\rho\sigma_1^2 r_1^e}{i} \right) \left[ v_1(c_{13}^2 - 1) + v_2 \left( \frac{\sigma_2}{\sigma_1} \right) (c_{13}c_{23} - c_{12}) \right] - \left( \frac{\sigma_{13}}{\sigma_3^2} \right) \frac{r_1^e}{(1-t)^2} \quad (37)$$

$$\frac{dr_2^e}{dt} = \left( \frac{\rho\sigma_2^2 r_2^e}{i} \right) \left[ v_2(c_{23}^2 - 1) + v_1 \left( \frac{\sigma_1}{\sigma_2} \right) (c_{13}c_{23} - c_{12}) \right] - \left( \frac{\sigma_{23}}{\sigma_3^2} \right) \frac{r_2^e}{(1-t)^2} \quad (38)$$

These results show that even if investors are well-diversified ( $v_1 \rightarrow 0$  and  $v_2 \rightarrow 0$ ), the shareholder income tax will distort the required return to unquoted shares. In the normal case where  $\sigma_{j3} > 0$ , a higher shareholder income tax rate is seen to *reduce* the cost of equity finance for unquoted companies, because the tax falls more heavily on quoted shares, thereby inducing substitution towards unquoted shares.

In summary, the Swedish policy of granting double tax relief only to holders of unquoted shares is not neutral towards real investment incentives. To achieve (approximate) neutrality in the case of well-diversified investors, it is necessary to grant a general rate-of-return allowance for *all* shareholdings, as we saw in section 4. Alternatively, neutrality with well-diversified investors can be achieved if the shareholder income tax with a rate-of-return allowance is levied only on returns to unquoted shares while returns to quoted shares are left free of tax, as proposed in Sørensen (2005b). In such a regime one can show that  $dr_1^e/dt$  and  $dr_2^e/dt$  will again be given by (31) and (32), implying  $dr_j^e/dt \rightarrow 0$  for  $v_j \rightarrow 0$ ,  $j = 1, 2$ . Imposing the tax only on unquoted shares might make sense if the main purpose of the tax is to prevent the owners of small companies from transforming heavily taxed management salaries into lightly-taxed dividends or capital gains on shares, and if policy makers do not wish to give domestic shareholders a tax incentive to sell their shares in quoted companies to foreign investors.

Assuming plausibly that  $\sigma_{13} > 0$  and  $\sigma_{23} > 0$ , a comparison of (31) and (32) to (37) and (38) reveals that the current Swedish tax system provides a stronger stimulus to risk-taking in small companies compared to a general shareholder income tax with a rate-of-return allowance for all shares (»the uniform tax«), or compared to a tax which is levied only on the equity premium on unquoted shares (»the selective tax«). The fact that the Swedish tax system tends to reduce the cost of capital for small companies even when investors are fully diversified suggests that it is too favourable to investment in this sector. By contrast, »the uniform tax« and »the selective tax« will only tend to stimulate risk-taking in unquoted companies when investors are not well diversified, indicating that these tax systems have better efficiency properties than the current Swedish rules.

### 7. Concluding remarks

The literature on the economics of taxation has been keen on identifying »neutral« taxes that do not distort the allocation of resources in well-functioning markets. This paper started with the question: if private markets in a globalized economy have the potential to ensure an efficient allocation of risk, could we design a system of shareholder income taxation that does not distort this allocation? Our first result was that, in an open economy with well-diversified investors, a conventional personal income tax on the return to shares (with full loss offset) will distort the pattern of risk-taking and investment, by favouring investment in relatively risky shares. Our main result was that if shareholders are granted a deduction for a risk-free rate of return on their shares, a symmetric tax on the remaining income from shares will indeed be neutral when investors are fully diversified. Proceeding to the more realistic case with incomplete diversification, the analysis showed that a symmetric personal tax on the equity premium will most likely tend to reduce the cost of capital for unquoted companies. Presumably this is socially desirable, since lack of diversification is likely to imply too little risk-taking. In effect, a symmetric tax on the equity premium allows a pooling of risk through the tax system as a (partial) compensation for the imperfect risk pooling via the market. Moreover, in countries where capital income is taxed at relatively low rates, a tax on the equity premium may counter tax avoidance by eliminating the ability of corporate owner-managers to transform heavily-taxed labour income into lightly-taxed capital income. Thus a shareholder income tax with a general deduction for a risk-free rate of return may be a promising model for corporate-personal tax integration in small open economies. As already mentioned, such a tax regime will in fact be introduced in Norway from 2006.

Given the assumption of normally distributed returns, the results derived in this paper are identical to those that would follow from a conventional Capital Asset Pricing Model (CAPM) with three risky assets. Because of the limited empirical success of the CAPM, the results and policy conclusions presented above should be interpreted with care. For example, the CAPM cannot explain why shares in small companies tend to yield higher risk-adjusted (beta-adjusted) rates of return than shares in large companies. This may be because investors are less well informed about shares in small companies and/or because the market for these shares is less liquid than assumed in the model. Hence the CAPM probably cannot uncover all of the effects of taxation on required returns. Moreover, the recent literature on Behavioral Finance (surveyed by Barberis and Thaler (2003)) has cast doubts on the assumption of expected utility maximization and has pointed to the psychological phenomenon of loss aversion which suggests that an asymmetric tax treatment of gains and losses might be a relevant poli-

cy if the aim is to stimulate investment.<sup>5</sup> More generally, if loss aversion is widespread, the insurance properties of a shareholder income tax with loss offsets are likely to be even more important than emphasized in this paper.

---

5. I am grateful to a referee for drawing my attention to these points.

### References

- Andersson, K., V. Kanninen, J. Södersten and P. B. Sørensen. 1998. Corporate tax policy in the Nordic countries, in P. B. Sørensen, red., *Tax Policy in the Nordic Countries*, Macmillan Press.
- Apel, M. and J. Södersten. 1999. Personal taxation and investment incentives in a small open economy, *International Tax and Public Finance*, 6, 79-88.
- Auerbach, A. 1979. Wealth maximization and the cost of capital, *Quarterly Journal of Economics*, 94, 433-46.
- Barberis, N. C. and R. H. Thaler. 2003. A survey of Behavioral Finance, ch. 18 in G. M. Constantinides, M. Harris and R.M. Stulz, red., *Handbook of the Economics of Finance*, Volume 1B, North-Holland.
- Boadway, R. and N. Bruce. 1984. A general proposition on the design of a neutral business tax, *Journal of Public Economics*, 58, 57-71.
- Boadway, R. and N. Bruce. 1992. Problems with integrating corporate and personal income taxes in an open economy, *Journal of Public Economics*, 48, 39-66.
- Bond, S. (2000), Levelling up or levelling down? Some reflections on the ACE and CBIT proposals, and the future of the corporate tax base, in S. Cnossen, red. *Taxing Capital Income in the European Union – Issues and Options for Reform*, Oxford University Press.
- Bond, S. and M. P. Devereux. 1995. On the design of a neutral business tax under uncertainty, *Journal of Public Economics*, 58, 57-71.
- Bradford, D. 1981. The incidence and allocation effects of a tax on corporate distributions, *Journal of Public Economics*, 15, 1-22.
- Devereux, M. P. and H. Freeman. 1991. A general neutral profits tax, *Fiscal Studies*, 12, 1-15.
- Fane, G. 1987. Neutral taxation under uncertainty, *Journal of Public Economics*, 33, 95-105.
- Fuest, C. and B. Huber. 2000. Can corporate-personal tax integration survive in open economies, *FinanzArchiv*, 57, 514-24.
- Gordon, R. H. 1985. Taxation of corporate capital income: tax revenues versus tax distortions, *Quarterly Journal of Economics*, 100, 1-27.
- Harberger, A. C. 1962. The incidence of the corporation income tax, *Journal of Political Economy*, 70, 215-40.
- Harberger, A. C. 1966. Efficiency effects of taxes on income from capital, in M. Krzyzaniak, red., *Effects of Corporation Income Tax*, Detroit: Wayne State University Press.
- King, M.A. 1974. Taxation and the cost of capital, *Review of Economic Studies*, 41, 21-35.
- King, M. A. 1975. Taxation, corporate financial policy and the cost of capital: a comment, *Journal of Public Economics*, 4, 271-79.
- Konrad, K. and W. Richter. 1995. Capital income taxation and risk spreading with adverse selection, *The Canadian Journal of Economics*, 28, 617-30.
- Lund, D. 2002. Petroleum tax reform proposals in Norway and Denmark, *Energy Journal*, 23, 37-56.
- Sandmo, A. 1977. Portfolio theory, asset demand and taxation: Comparative statics with many assets, *Review of Economic Studies*, 44, 369-79.
- Sandmo, A. 1989. Differential taxation and the encouragement of risk taking, *Economics Letters*, 31, 55-59.
- Sinn, H.-W. 1987. *Capital Income Taxation and Resource Allocation*, North-Holland.
- Silberberg, E. and W. Suen. 2001. *The Structure of Economics – A Mathematical Ana-*

- lysis*, Third Edition, McGraw-Hill.
- Sørensen, P. B. 1995. Changing views of the corporate income tax, *National Tax Journal*, vol. XLVIII, 279-94.
- Sørensen, P. B. 2005a. Neutral taxation of shareholder income, forthcoming in *International Tax and Public Finance*.
- Sørensen, P. B. 2005b. Dual income taxation – Why and how?, forthcoming in *Finanz-Archiv*.

## Bog anmeldelse

Michael Møller og Niels Christian Nielsen.  
»Din økonomi« – en bog om tid og penge.  
Anmeldt af Kim Valentin, Cand. polit. Fi-  
nanshuset i Fredensborg A/S.

Før du køber denne bog, skal du overveje, om den kan bibringe dig viden, som vil gøre dig i stand til at tjene pengene hjem igen. Du skal tilmed regne med at bruge lang tid på at forstå indholdet, og i den tid kunne du jo have tjent penge på andre ting.

Du skal også regne med, at dersom du ikke forstår bogen – eller i det mindste ikke forstår at bruge den – så har du blot bragt dig selv i den situation, at du nu er bevidst uvidende, og det er trods alt lidt mere ubehageligt, end at være ubevidst uvidende.

Er du indstillet på at tage denne risiko, så læs blot videre.

Bogen er fyldt med denne form for logik, og til tider kan den være befriende og hjælpe dig til at prioritere, hvad du vil bruge tid på, og hvad du ikke vil bruge tid på, men til andre tider vil den slags logik ikke give dig værdi, og i sidste ende er en af pointerne i bogen, at det, du bruger din tid på, skal give dig værdi.

Bogen er udarbejdet i tre hovedafsnit, som beskæftiger sig med privatøkonomi med hver sin indgangsvinkel. Endvidere er de enkelte hovedafsnit pædagogisk opdelt i mindre kapitler. Denne opdeling giver læseren fordelene af at kunne afbryde læsningen løbende og måske bruge bogen som opslag til forståelse af enkelte privatøkonomiske emner på et senere tidspunkt. Men til gengæld vil den privatøkonomiske metodeindlæring halte lidt efter. Det er forståeligt at gøre dette valg, om end det så bliver sværere at bruge bogen til personlig privatøkonomisk planlægning.

Til tider får man forståelsen af, at forfatterne ønsker, at læserne skal afholde sig fra at gøre valg omkring deres økonomi, de ikke forstår til fulde, hvorfor det generelt også an-

befales, at læserne gør det simpleste i stedet for at gøre det optimale.

Et hovedbudskab i bogen er, at privatøkonomiske beslutninger skal ses i en sammenhæng over hele livet og måles på formueudviklingen. Det kan dog ikke forventes, at læserne kan opnå denne viden ved at læse bogen, således at man kan tage de rigtige beslutninger på de rigtige tidspunkter gennem hele livet. Bogen anviser ikke metoden til at finde denne sammenhæng. Bogen gør læserne i stand til at træffe mere rigtige beslutninger på delområder i privatøkonomien.

Den første del er således koncentreret om de beslutningsprocesser, som foregår i en privatøkonomi, og er godt opbygget med referencer til, hvordan du kan lære at forstå, hvorfor du opfører dig, som du gør, når du gør beslutninger i/forvalter din privatøkonomi.

Ligeledes er der gode eksempler på, hvordan du tilrettelægger dit forbrug, og hvordan du eksempelvis forholder dig til »problematikken« omkring arvinger, forsikringsdækning, deltagelse i spil mv.

Bogen kommer ganske pædagogisk ind på en forklaring omkring risikobegreberne i din privatøkonomi, og man fristes til at citere Mark Twain i den sammenhæng – »Kun to gange i sit liv bør et menneske ikke spekulere, når han har råd til det, og når han ikke har råd til det«.

Bogen siger på den måde en masse rigtige ting omkring forholdene i ens økonomi, og første del af bogen forbinder de enkelte økonomiske områder i privatøkonomien med det, som psykologer siger, vi går rundt og gør, når vi tager beslutninger, der både har kortsigtet og langsigtet effekt.

På den måde kan første del være en ganske god »øjenåbner« til rigtig mange mennesker.

Denne del af bogen er også ganske glimrende at læse, før du forsøger at planlægge en del af dit liv, og for at kunne stille de rigtige spørgsmål til de rigtige mennesker – nemlig rådgiverne.



Første del er endvidere ganske let læselig og vil for de fleste kunne bruges til at skabe den ovenfor nævnte værdi i deres økonomi. Alene af denne årsag kan det anbefales at købe bogen.

Med hensyn til anden del af bogen, som omhandler de finansielle omgivelser, må læseren forvente, at det til tider kan være vanskeligt at forstå den viden, som forfatterne forsøger at overgive til læseren. Områder som rate- og kapitalpension, gennemgangen af de centrale danske skattesatser samt tilgangen til anbringelse af de frie midler beskrives godt hver for sig og skal i bogen forstås hver for sig. Opsparingsrækkefølgen kan imidlertid kun ses i en sammenhæng, og den får du ikke anvist ved at læse bogen.

Af de enkelte områder mener jeg, at de frie midler, som generelt er en forholdsmæssig lille del af den langsigtede opsparing, i gennemgangen bliver prioriteret for højt. Læseren ville også kunne have fordel af samme detaljeringsgrad i gennemgangen af de muligheder, rate- og kapitalpensioner samt ikke mindst boligkøb kan skabe for læserens privatøkonomi.

Specielt kunne forholdene omkring boligkøb med fordel være indlagt i denne del af bogen.

Uanset disse forhold er denne del af bogen også rigtig god at gennemgå, før man beslutter sig for, hvad man vil gøre med sin opsparing.

Således er man tilbøjelig til at huske på *John Maynard Keynes'* citat – »Der findes intet så katastrofalt, som en rationel investeringspolitik i en irrationel verden«. Dette citat vil i høj grad være erindringsværdigt, når du læser anden del af bogen.

Den sidste del af bogen omhandler nogle

udvalgte privatøkonomiske problemstillinger, som er ganske gode at huske på, når man står i situationen. Denne del kan derfor bruges som opslagsværk for særskilte beslutninger. Forfatterens personlige privatøkonomiske aldre er dog vægtet for kraftigt. Det er eksempelvis vanskeligt at tro på, at afsnittet omkring – »hvordan du bør forholde dig til gavegivning/støtte af forskellige humanitære foreninger« – kan være andet end dette.

Det kunne ønskes, at man i den tredje del af bogen havde vægtet gennemgangen af de forvaltningsmæssige mekanismer i en økonomi højere i eksempelvis områder som forbrugsstyring gennem planlægning i forhold til et ønsket rådighedsbeløb eller igen her detaljviden omkring opsparingsrækkefølgen, altså valget mellem opsparing i frie midler, pensionsmidler eller eksempelvis afdrag på lån/køb af fast ejendom.

På den måde kunne man have sikret læseren at prioritere planlægningen af økonomien, før man går i markedet og vælger produkter.

Bogen er således udtryk for en relativ negativ holdning til planlægning for langt frem i tiden, og man kommer derfor til at tænke på, hvad én af vores egne forfattere og digtere skrev på et tidspunkt – »Én af de få fordele ved at være fattig er, at man aldrig tillader sig selv at se for langt ud i fremtiden« (Tove Ditlevsen). Denne form for logik er dog i dagens Danmark svær at forstå, og de fleste mennesker vil derfor, efter at have læst denne bog, være stillet i en situation, hvor man har fået en masse viden, men man har svært ved at bruge den.

Som helhed kan bogen imidlertid stærkt anbefales, og der er en rigtig god chance for, at dine penge og din tid er godt brugt.