

Er der et dansk produktivitetskollaps?

Carl-Johan Dalgaard

Økonomisk Institut, Københavns Universitet, E-mail: carl.johan.dalgaard@econ.ku.dk

Henrik Hansen

Fødevarerøkonomisk Institut, Københavns Universitet, E-mail: henrik.hansen@foi.dk

SUMMARY: Recently the Danish economy has been described as having serious structural problems showing up as a considerable decrease in productivity growth. In this study we question if this is so. We analyze the Danish growth process using formal statistical tests. The tests show that the annual growth rates of GDP per capita and of GDP per employee – from the early 1950s until the late 2000s – are well described as random noise around a constant mean. In contrast, the average annual growth rate of GDP per work hour appears to have declined during the period 1960-2007. But, the significant decline occurred in the 1970s, while the latter period has seen a fairly constant average annual growth rate. Hence, we find no statistically significant indication of a productivity slowdown after 1995 – or later. We also discuss the productivity growth in a cross-country perspective comparing the Danish growth to the growth rates in other OECD countries. We ask if the Danish experience gives cause for alarm. Our answer challenges most of the recent claims of a Danish growth collapse as we argue that the observed outcomes are well in line with expectations once conditional convergence effects are considered.

1. Indledning

Væksten i arbejdskraftproduktiviteten er central for ethvert land, da denne i praksis er eneansvarlig for, at forbrugsmulighederne, privat som offentligt, kan øges over tid. Det er derfor foruroligende, hvis produktivitetsvæksten udviser en tendens til at løje af.

I November 2009 præsenterede Økonomi- og Erhvervsministeriet (ØE) rapporten »Den danske produktivitetsudvikling«, hvori det blandt andet fremhæves, at produktivitetsvæksten har været faldende fra midten af 1990'erne og frem. Lignende betragtninger finder man i CEPOS (2008), som finder det bekymrende, at »Danmark er på

Dette indlæg bygger på Dalgaard og Hansen (2010). Der har været en overraskende stor interesse for dette EPRU analysepapir, og vi har modtaget mange konstruktive kommentarer fra en lang række personer. Specielt medarbejdere fra Økonomi- og Erhvervsministeriet, Arbejderbevægelsens Erhvervsråd og Det Økonomiske Råds Sekretariat har kommenteret, eller på anden vis berørt arbejdsrapporten i såvel skrift som tale. Vi er meget taknemmelige for de mange kommentarer, som har ført til, at dette indlæg er revideret substantielt i forhold til arbejdsrapporten.

vej mod en midterplacering i OECD«. Arbejderbevægelses Erhvervsråd (2009) (AE) er delvist enig; her påpeges produktivetsproblemet ligeledes, om end tidspunktet for nedgangen synes at være fastlagt til omkring år 2000.

Det er desværre ikke helt klart, hvorledes ØE, CEPOS og AE når frem til, at væksten er »kollapset«. Specielt er det ikke oplagt, hvordan de tre institutioner er nået frem til tidspunktet for det ekstraordinære fald i produktivitetsvæksten. ØE antager, at processen begyndte at køre af sporet omkring 1995; CEPOS fremhæver 1998, mens AE fokuserer på 2000. Denne usikkerhed i dateringen af nedgangen i produktivitetsvæksten betyder, at man bør undersøge mere præcist, hvorvidt og hvornår produktivitetsvæksten begyndte at falde, idet timingen bør have afgørende betydning for en nærmere analyse af problemstillingen.

Nedenfor adresserer vi to spørgsmål:

1. Er den danske vækst *trend* i aftagende?
2. Er den danske vækst urovækkende lav i *forhold* til andre (OECD) lande?

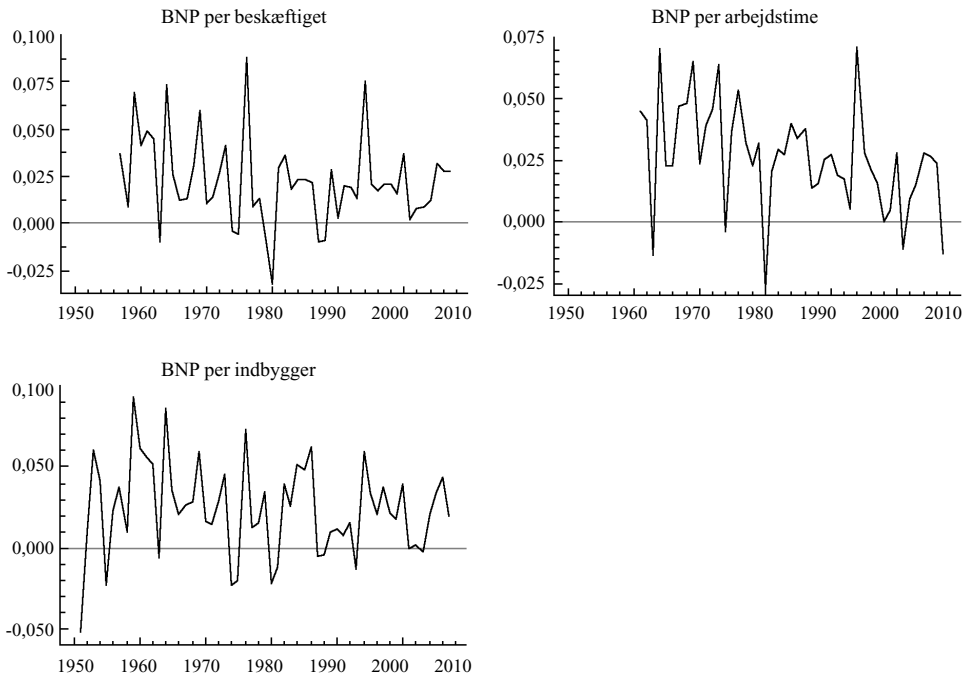
I analysen undlader vi at fastlægge en dato for et eventuelt skift i væksten *a priori*. Baggrunden er det velkendte faktum, at man altid kan finde »systematiske« ændringer i tidsrækker, hvis man bare stirrer længe nok – og det sete afhænger som bekendt af øjnene der ser. Vi vil derfor primært »se« på data med statistiske metoder og test.

Vi starter med at analysere den danske vækst i næste afsnit, hvorpå vi vender os mod spørgsmålet om komparativ udvikling i Afsnit 3. Afsnit 4 konkluderer.

2. Den danske vækst 1951-2007

I dette afsnit ser vi nærmere på udviklingen i produktivitetsvæksten fra 1950erne og 1960erne og frem til i dag (2007). Formålet er at belyse, i hvilket omfang man med statistiske metoder kan se tegn på fald i vækstraten i produktiviteten. I analysen benytter vi som udgangspunkt bruttonationalproduktet beregnet i faste priser med kædevægte. Data for BNP er fra Penn World Table Version 6.3 (PWT), hvilket er den foretrukne database ved internationale sammenligninger af det langsigtede vækstforløb. Disse BNP data sættes herefter i forhold til udviklingen i beskæftigelsen samt udviklingen i det totale antal arbejdstimer, idet specielt timeproduktiviteten har været i fokus i den danske debat.¹ Endelig medtager vi også for fuldstændighedens skyld BNP per indbygger for at belyse, i hvilken grad eventuelle skift i produktivetsudviklingen har påvirket udviklingen i det hyppigst anvendte velstandsmål. Om end analysen hovedsagelig trækker på data fra Penn World Tables inddrager vi også BNP data, der stammer direkte fra Danmarks Statistik i tidsrækkeanalysen, for derved at tjekke re-

1. Data for det totale antal arbejdstimer er hentet fra det Hollandske projekt »The Conference Board Total Economy Database« (CB).



Figur 1. Årlige vækstrater i BNP per beskæftiget, BNP per arbejdstime samt BNP per indbygger.

Note: Vækstraterne er beregnet som de logaritmiske årlige ændringer. Kilde: Datakilder og transformationer er beskrevet i Appendiks.

sultaternes robusthed. De præcise datakilder og datatransformationer er beskrevet i Appendiks.

De årlige vækstrater i BNP per beskæftiget, per arbejdstime samt per indbygger er afbildet i figur 1. Vækstraterne er beregnet som 100 gange de logaritmiske årlige ændringer. Data for BNP og befolkningen findes tilbage til 1950 i PWT, mens der desværre kun er data fra 1956 for antallet af beskæftigede og fra 1960 for antallet af arbejdstimer.

Figur 1 viser at vækstraterne i BNP per beskæftiget og per indbygger ikke stiger eller falder systematisk over tiden. Visuelt fremstår de to serier som stationære omkring konstante gennemsnitlige vækstrater, og dette bekræftes af formelle Dickey-Fuller test for stationaritet, som er angivet øverst i Tabel 1. For timeproduktiviteten kan man, hjulpet af den verserende debat, visuelt spore en faldende tendens fra 1970erne til i dag. Den trendmæssige udvikling er dog ikke tilstrækkelig til at manifestere sig i en accept af hypotesen om, at vækstraten i timeproduktiviteten er en stokastisk trend, når

Tabel 1. Statistisk analyse af vækstraten i BNP per beskæftiget, per arbejdstime samt per indbygger med data fra Penn World Table Version 6.3 samt The Conference Board Total Economy Database.

	Vækst i BNP per beskæftiget		Vækst i BNP per arbejdstime		Vækst i BNP per indbygger	
	57-07	57-93	61-07	61-93	51-07	51-93
Estimationsperiode	57-07	57-93	61-07	61-93	51-07	51-93
Dickey-Fuller <i>t</i> -test	-7,33	-6,12	-6,43	-5,90	-6,44	-5,40
<i>Regressionsmodel</i>						
Konstant	2,26	2,23	2,62	2,98	2,43	2,41
Standardafvigelse	(0,33)	(0,41)	(0,31)	(0,37)	(0,38)	(0,48)
Modelkontrol (<i>p</i> -værdier)						
Førsteordens autokorrelation	0,72	0,79	0,99	0,63	0,61	0,59
Førsteordens ARCH	0,90	0,97	0,94	0,89	0,65	0,93
Normalitet	0,11	0,31	0,53	0,19	0,73	0,97
<i>Parameterstabilitetstest</i>						
Stabilitet af konstant	0,30	0,51*	0,78*	0,41	0,15	0,23
Test af trendpolynomium (<i>p</i> -værdier)	0,19	0,17	0,06	0,28	0,26	0,14
<i>Forudsigelsestest 1994-2007</i>						
Forudsigelsestest (<i>p</i> -værdier)		0,95		0,24		0,99
CUSUM-test (<i>p</i> -værdier)		0,95		0,07		0,98
Positive/negative forudsigelsesfejl		5/9		1/13*		6/8

Note: De årlige vækstrater er beregnet som $100 * dlog(Y)$, hvor $dlog$ er den logaritmiske ændring og Y er henholdsvis, BNP per beskæftiget, per arbejdstime og per indbygger. Den 5% kritiske værdi for Dickey-Fuller *t*-testet er -2,9. Den 5% kritiske værdi for stabilitet af konstant er 0,47. Det 5% kritiske antal positive/negative forudsigelsesfejl er 2 af de 14 forudsigelser.

Kilde: Datakilder og transformationer er beskrevet i Appendiks.

dette testes med Dickey-Fuller testet. Testet afviser ganske klart nul-hypotesen om at vækstraten er integreret af første orden.

På trods af den tilsyneladende klare stationaritet i de tre serier vil vi se nærmere på, hvorvidt der statistisk kan findes tegn på hurtige eller mere langsomt snigende ændringer i vækstraterne. Vi har derfor estimeret univariate tidsrækkemodeller for hver af serierne.²

Vi startede først med en autoregressiv model af 4. orden. Denne model kunne uden problemer reduceres til en simpel model uden lags. Herefter undersøgte vi en ARMA(1,1)-model, der ligeledes kunne testes ned til den simple model uden lags. Vi benytter derfor den simple model, hvor vækstraterne alene modelleres med et konstant-

2. Alle estimationer og test er foretaget i PcGive 13, se Doornik og Hendry (2009).

led (den gennemsnitlige årlige vækstrate) til at teste, hvorvidt der er strukturelle brud i produktivitetsvæksten.³

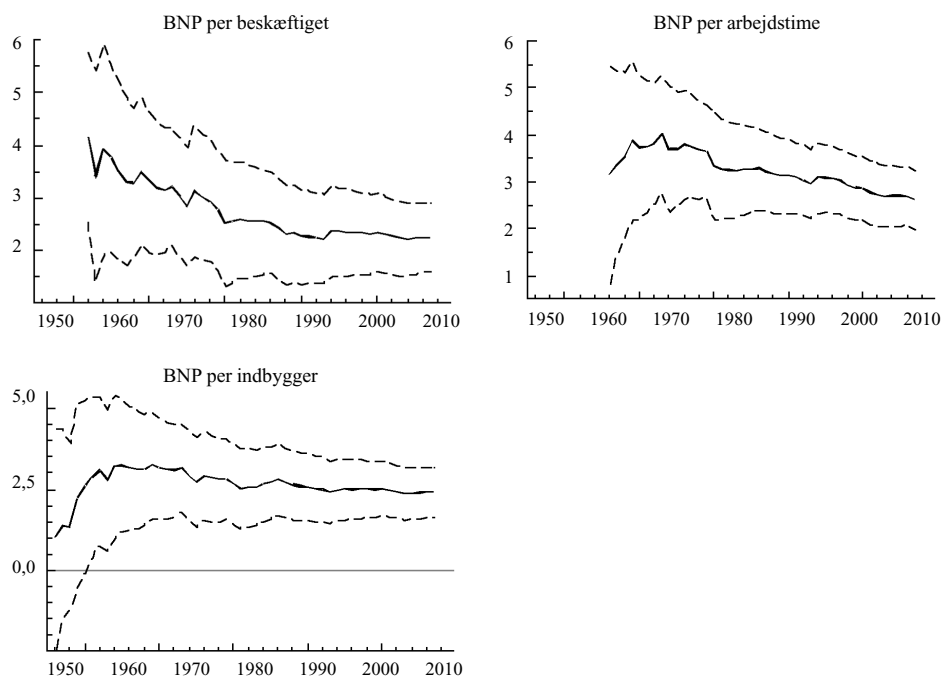
Tabel 1 viser, at den gennemsnitlige årlige vækstrate i BNP per beskæftiget fra 1957 til 2007 var lige over 2,2%, mens vækstraten i BNP per arbejdstime var lidt højere (2,6%). Væksten i BNP per indbygger lå pænt mellem disse med et gennemsnit lige over 2,4% fra 1951 til 2007. Spredningerne i vækstraterne var relativt lave, hvorved spredningen i gennemsnittene kommer under 0,5% for alle tre serier, hvilket indikerer at eventuelle skift i de gennemsnitlige vækstrater skal være tæt på et procent point før de er statistisk signifikante.

I Tabel 1 rapporteres tre standardtest af modelspecifikationen, hvilket med det simple modelvalg egentlig er test af de tre seriers egenskaber, idet regressionsmodellen blot trækker gennemsnittene ud af serierne. Specifikationstestene viser, at de årlige vækstrater uden problemer kan antages at være normalfordelte, og der er ingen tegn på autokorrelation, hverken i serierne som sådan eller i deres betingede varianser (såkaldte ARCH-effekter). At tænke på vækstraterne i BNP per beskæftiget, per arbejdstime samt per indbygger som tilfældigt varierende omkring et fast gennemsnit lige over 2% (lige under 3% for væksten per arbejdstime) synes derfor ikke umiddelbart urimeligt.

I Tabel 1 vises også to test for parameterstabilitet. Med det første («Stabilitet af konstant») tester vi, hvorvidt middelværdien med rimelighed kan antages at være konstant over tiden over for alternativet, at middelværdien bør beskrives som en 'random walk'. Testet er udviklet i Nyblom (1989) og beskrevet generelt i Hansen (1992). I de simple modeller vi bruger her, kan testet ses som et alternativ eller måske nærmere et supplement til Dickey-Fuller testet for stationaritet. I tabellen angives teststørrelserne for nulhypoteserne, at konstantleddene ikke varierer over tid. Den kritiske værdi for testet på 5% er 0,47. Det fremgår, at nulhypoteserne generelt ikke kan afvises; med to væsentlige undtagelser. Teststørrelsen er lige over den kritiske værdi for væksten i BNP per beskæftiget, når gennemsnittet estimeres på den korte periode fra 1957 til 1993, og vi finder en klart signifikant teststørrelse for væksten i BNP per arbejdstime, når modellen estimeres på den fulde periode fra 1957 til 2007.

Med det andet parameterstabilitetstest, som er foreslået af Lin og Terasvirta (1994), undersøger vi, hvorvidt et 3. grads polynomium i tiden (en fleksibel trend) giver en signifikant bedre forklaring af vækstraterne end blot konstantleddet. Dette test har ligeledes styrke over for relativt bløde skift i serierne. Det fremgår af tabellen, at den fleksible trend er insignifikant for alle serier, men det bemærkes også at nulhypotesen kun lige kan afvises på et 5% testniveau for vækstraten i BNP per arbejdstime. Samlet

3. For vækstraten i BNP per indbygger var ARMA(1,1) modellen statistisk signifikant, men vi analyserer alligevel denne serie med den simple model for at have ens modeller for de tre serier. Konklusionerne vedrørende nedgang i vækstraterne påvirkes ikke af dette valg.

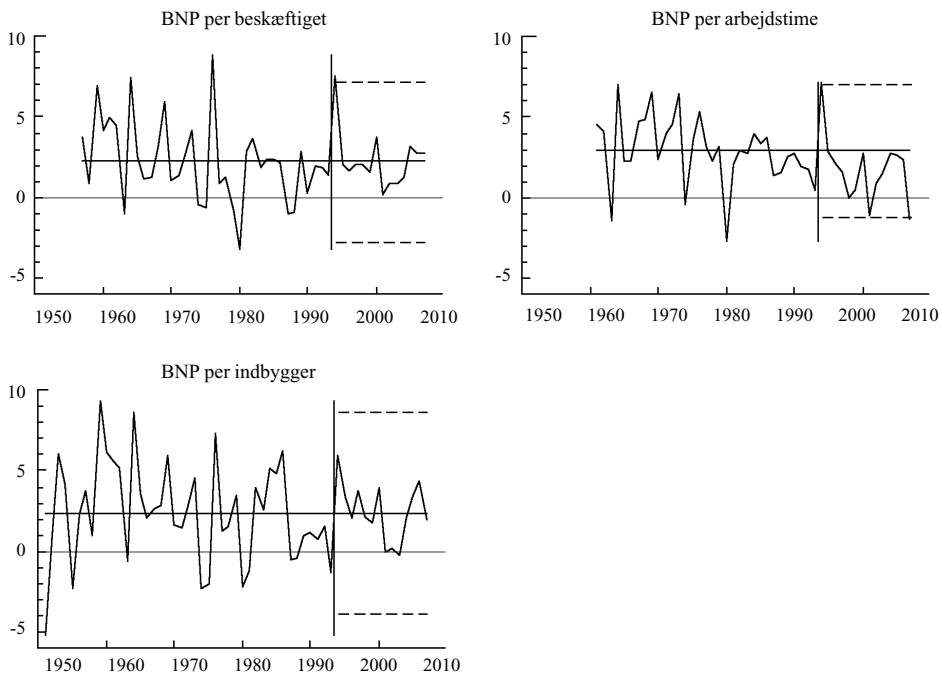


Figur 2. Estimerede konstantled (gennemsnitlige årlige vækstrater), når modellerne for de tre vækstrater estimeres rekursivt fra hhv. 1961, 1966 samt 1956 frem til 2007.

Note: Figuren viser de estimerede konstantled samt approksimative 95% punktvisse konfidensintervaller beregnet som ± 2 standardafvigelsen til hvert tidspunkt.

er det dog umiddelbart vanskeligt at påvise trendbrud i vækstraten i arbejds- og timeproduktiviteten i Danmark, når man ser på disse serier og bruger de valgte formelle test.

Figur 2 giver en tredje, lidt mindre formel, metode til at undersøge, hvorvidt der er systematiske ændringer i vækstraternes middelværdier. I figuren plottes de estimerede gennemsnit (med approksimative 95% punktvisse konfidensintervaller) for serierne, når de estimeres rekursivt fra fem år efter første observation og frem til 2007. For BNP per indbygger og per arbejdstime er det tydeligt, at der først sker en gradvis stigning i de gennemsnitlige vækstrater frem til omkring 1970. Denne stigning afløses af et gradvist fald fra starten af 1970'erne og frem. For vækstraten i BNP per beskæftiget og per indbygger stabiliseres de estimerede gennemsnit fra slutningen af 1980'erne og frem, mens gennemsnittet for vækstraten i BNP per arbejdstime ikke på samme måde stabiliseres når estimationsperioden forlænges. Ændringerne i gennemsnittene er dog



Figur 3. Faktiske, estimerede og forudsagte værdier for de årlige vækstrater i BNP per beskæftiget, per arbejdstime samt per indbygger.

Note: Forudsigelserne bygger på gennemsnittene estimeret på delperioderne 1957-1993, 1961-1993 samt 1951-1993, for hhv. BNP per beskæftiget, per arbejdstime og per indbygger.

ikke voldsomme sammenlignet med den statistiske usikkerhed. Konfidensintervallerne omkring de gennemsnitlige vækstrater indeholder altid henholdsvis 2,2-2,3% (for BNP per beskæftiget), 2,6-2,8% (for BNP per arbejdstime) samt 2,4% (for BNP per indbygger).

I nederste del af tabel 1 samt i figur 3 fokuserer vi på modellernes forudsigelsesegenskaber. I tabellen vises resultater for de tre serier, når de gennemsnitlige vækstrater er estimeret for perioden frem til 1993, hvorefter disse gennemsnit benyttes til forudsigelser af vækstraterne fra 1994 til 2007. I figur 3 er forudsigelserne plottet med tilhørende forudsigelsesintervaller sammen med de faktiske observationer. Da regressionsmodellen alene indeholder et konstantled er forudsigelserne naturligvis konstante i hele forudsigelsesperioden. Det fremgår tydeligt af figur 3, at 1994 var et særligt år med store positive forudsigelsesfejl i vækstraterne i BNP per beskæftiget og per arbejdstime. Samtidig ses det, at de efterfølgende forudsigelsesfejl typisk er ganske små sammenlignet med udsvingene i vækstraterne i estimationsperioden.

Egentlige test af parameterstabilitet med udgangspunkt i modellernes forudsigelsesegenskaber for perioden fra 1994 til 2007 er givet nederst i tabel 1. Det første test er et klassisk forudsigelsestest, hvor det undersøges, hvorvidt variansen i forudsigelsesfejlene er forskellig fra variansen i estimationsperioden. Nulhypotesen om konstante parametre afvises, hvis enten gennemsnittet eller variansen i vækstraterne ændres over tid. Det ses, at dette ikke er tilfældet – i en sådan grad at det kan spores med dette test. Det andet test (CUSUM-test) undersøger, hvorvidt middelværdien af forudsigelsesfejlene kan antages at være nul. Hvis der er trend i vækstraterne vil forudsigelserne systematisk blive stadig mere positive eller negative over tid, og dette vil lede til en afvisning af nul-hypotesen. Det fremgår, at hypotesen klart ikke kan afvises for vækstraterne i BNP per beskæftiget og per indbygger, mens konklusionen er mere tvivlsom for vækstraten i BNP per arbejdstime. Konklusionen understøttes af det tredje test, som blot angiver antallet af positive og negative forudsigelsesfejl. Hvis middelværdierne er konstante over tid, vil vi forvente, at forudsigelsesfejlene har en fordeling, som er symmetrisk omkring nul. Hvis dette er rigtigt, bør antallet af positive og negative forudsigelsesfejl være binomialfordelt med en sandsynlighed på 0,5. Under denne antagelse vil vi med ca. 95% sandsynlighed forvente at få mindst 3 og højest 11 positive (negative) forudsigelsesfejl ud af de 14 forudsigelser. Det fremgår af tabel 1, at vi ikke kan afvise hypotesen om en symmetrisk fordeling af forudsigelsesfejlene for vækstraterne i BNP per beskæftiget og per indbygger, mens der med 1 positiv og 13 negative forudsigelsesfejl for vækstraten i BNP per arbejdstime er tegn på en skæv fordeling, som kan skyldes et skift i middelværdien.

På baggrund af disse estimationer og statistiske test konkluderer vi, at de årlige vækstrater i BNP per beskæftiget samt i BNP per indbygger må regnes for at være tilfældige udsving omkring en konstant middelværdi. For vækstraterne i BNP per arbejdstime kan man spore en signifikant parameterstabilitet, som muligvis kan henføres til et systematisk fald i den gennemsnitlige vækstrate over tid.

I flere kommentarer til arbejdspapirversionen af vores analyse, Dalgaard og Hansen (2010) er det fremført, at resultaterne ovenfor er stærkt afhængige af datavalget. Vi benytter Penn World Table som kilde til det kædevægtede BNP, dels fordi denne datakilde i årevis har været det naturlige valg til internationale vækstsammenligninger, og dels fordi PWT giver længere tidsrækker end nationalregnskabstallene fra Danmarks Statistik, som kun findes tilbage til 1966. For fuldstændighedens skyld rapporterer vi derfor i tabel 2 samt figurerne 3 og 4 resultater af en analyse af tidsrækkerne fra Danmarks Statistik. Alle data er taget direkte fra Danmarks Statistiks statistikbank, og de bygger på november 2009 versionen af nationalregnskabet.⁴

4. Appendix indeholder en tabel med direkte sammenligninger af de gennemsnitlige vækstrater og korrelationerne mellem de årlige vækstrater estimeret på baggrund af de to datakilder.

Tabel 2. Statistisk analyse af vækstraten i BNP per beskæftiget, per arbejdstime samt per indbygger med data fra Danmarks Statistik.

	Vækst i BNP per beskæftiget		Vækst i BNP per arbejdstime		Vækst i BNP per indbygger	
	67-07	67-93	67-07	67-93	67-07	67-93
Estimationsperiode	67-07	67-93	67-07	67-93	67-07	67-93
Dickey-Fuller <i>t</i> -test	-5,42	-4,57	-4,24	-3,73	-5,47	-4,06
<i>Regressionsmodel</i>						
Konstant	1,91	2,12	2,52	3,08	2,07	2,09
Standardafvigelse	(0, 24)	(0,29)	(0,30)	(0,35)	(0,30)	(0,42)
<i>Modelkontrol (p-værdier)</i>						
Førsteordens autokorrelation	0,40	0,38	0,01*	0,09	0,39	0,37
Førsteordens ARCH	0,30	0,00*	0,00*	0,03*	0,00*	0,06
Normalitet	0,35	0,65	0,27	0,45	0,67	0,33
<i>Parameterstabilitetstest</i>						
Stabilitet af konstant	0,49*	0,37	1,57*	1,06*	0,21	0,31
Test af trendpolynomium (<i>p</i> -værdier)	0,01*	0,04*	0,00*	0,00*	0,26	0,09
<i>Forudsigelsestest 1994-2007</i>						
Forudsigelsestest (<i>p</i> -værdier)		0,22		0,06		0,99
CUSUM-test (<i>p</i> -værdier)		0,16		0,01*		0,85
Positive/negative forudsigelsesfejl		3/11		1/13*		8/6

Note: De årlige vækstrater er beregnet som $100 \cdot dlog(Y)$, hvor $dlog$ er den logaritmiske ændring og Y er henholdsvis, BNP per beskæftiget, per arbejdstime og per indbygger. Den 5% kritiske værdi for Dickey-Fuller *t*-testet er -2,9. Den 5% kritiske værdi for stabilitet af konstant er 0,47. Det 5% kritiske antal positive/negative forudsigelsesfejl er 2 af de 14 forudsigelser.

Kilde: Datakilder og beregninger er beskrevet i Appendiks 2.

En væsentlig forskel mellem resultaterne i tabel 1 og tabel 2 er, at der er meget klare tegn på skift i varianserne i vækstraterne i tabel 2. Den vigtigste forskel er dog nok de meget klare tegn på parameterstabilitet i vækstraten i BNP per arbejdstime i tabel 2. Både Nyblom-testet og testet for en fleksibel trend leder til en afvisning af nullhypotesen om en konstant middelværdi, og de tre egentlige forudsigelsestest leder også til en afvisning af antagelsen om konstante parametre. De formelle test bekræftes af en visuel inspektion af de rekursivt beregnede gennemsnitlige vækstrater i BNP per arbejdstime i figur 4 samt af de faktiske og forudsagte værdier i figur 5.

Den overraskende forskel i resultaterne for BNP per arbejdstime er bekymrende. Derfor ser vi i tabel 3 nærmere på et par udvidede modeller for vækstraten i BNP per arbejdstime. De to første kolonner i tabel 3 viser resultater for to modeludvidelser med PWT data, mens de to sidste kolonner viser resultaterne for de to modeludvidelser med DST data.

Tabel 3. En nærmere analyse af vækstraten i BNP per arbejdstime.

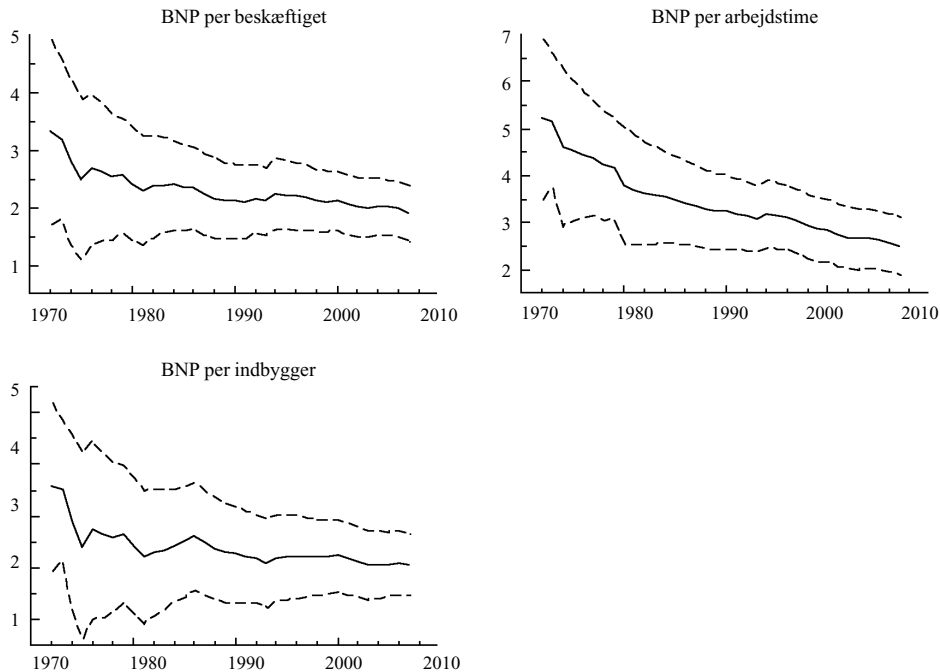
Specifikation	<i>PWT</i>	<i>data</i>	<i>DST</i>	<i>data</i>
	Lineær trend	Invers trend	Lineær trend	Invers trend
<i>Estimationsperiode</i>	61-07	61-07	67-07	67-07
<i>Regressionsmodel</i>				
Konstant	1,19	1,15	0,46	1,40
Standardafvigelse	(0, 57)	(0, 68)	(0,47)	(0,31)
Trend/invers trend	-0,06	42,5	-0,10	13,8
Standardafvigelse	(0,02)	(17,5)	(0,02)	(2,48)
R^2	0,16	0,12	0,40	0,44
<i>Modelkontrol (p-værdier)</i>				
Førsteordens autokorrelation	0,31	0,48	0,67	0,86
Førsteordens ARCH	0,91	0,87	0,61	0,51
Normalitet	0,02*	0,06	0,00*	0,07
<i>Parameterstabilitetstest</i>				
Stabilitet af konstant	0,03	0,10	0,11	0,24
Stabilitet af trend/invers trend	0,03	0,06	0,15	0,03
Test af trendpolynomium (p-værdier)	0,94	0,32	0,05*	0,15
<i>Forudsigelsestest 1994-2007</i>				
Forudsigelsestest (p-værdier)	0,51	0,47	0,03*	0,03*
CUSUM-test (p-værdier)	0,93	0,40	0,30	0,10
Positive/negative forudsigelsesfejl	9/5	6/8	11/3	3/11
<i>Encompassing test</i>				
Reduktion fra fælles model (p-værdier)	0,66	0,13	0,01*	0,06

Note: De årlige vækstrater er beregnet som $100 \cdot d\log(\text{BNP per arbejdstime})$, hvor $d\log$ er den logaritmiske ændring. Den 5% kritiske værdi for Dickey-Fuller t -testet er -2,9. Den 5% kritiske værdi for stabilitet af konstant er 0,47. Det 5% kritiske antal positive/negative forudsigelsesfejl er 2 af de 14 forudsigelser.

Kilde: Datakilder og beregninger er beskrevet i Appendiks.

I kolonne 1 og 3 udvider vi den simple model fra tabel 1 og 2 til at indeholde en lineær trend ($t = -T, \dots, 0$), mens modeludvidelsen i kolonne 2 og 4 er inklusion af en invers trend ($1/t, t = 1, \dots, T$). Begge modelformuleringer følger naturligt af, at vi finder et marginalt signifikant trendpolynomium i tabel 1 og et klart signifikant trendpolynomium i tabel 2.

Den lineære trend er specificeret, så den slutter med værdien 0 i 2007. Konstantleddene i regressionerne angiver dermed de estimerede vækstrater i 2007. I regressionerne med en invers trend angiver konstantleddene den estimerede fremtidige konstante vækstrate. Det fremgår, at der med PWT data er en meget lille forskel mellem de estimerede konstantled, mens forskellen er meget markant for regressionerne baseret på DST data.

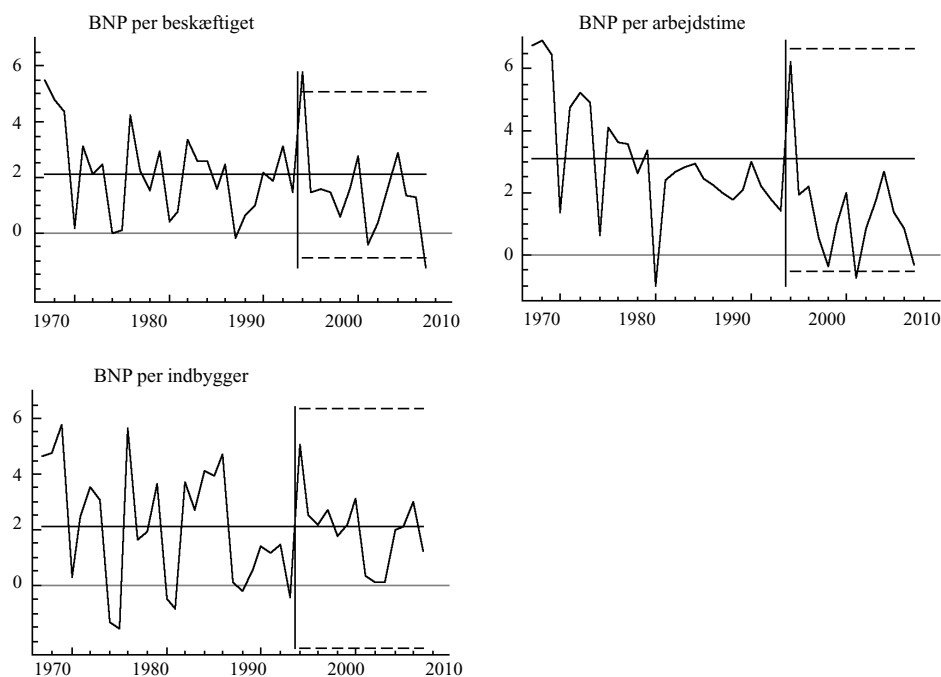


Figur 4. Estimerede konstantled (gennemsnitlige årlige vækstrater) når modellerne for de tre vækstrater estimeres rekursivt fra 1971 til 2007.

Note: Figuren viser de estimerede konstantled samt approksimative 95% punktvisse konfidensintervaller beregnet som $\pm 2 \cdot \text{standardafvigelsen}$ til hvert tidspunkt.

Det er ligeledes klart, at de estimerede trender er statistisk signifikante i alle fire regressionser, og at modellerne ikke fremstår fejlspecificerede. Der er dog tegn på specifikationsproblemer i modellerne med den lineære trend, idet normalitetsantagelsen afvises, og der er tegn på parameterustabilitet i modellen baseret på DST data.

De to modelformuleringer, henholdsvis en lineært aftagende vækstrate og en hyperbolsk aftagende vækstrate repræsenterer to meget forskellige historiske forløb for vækstraterne i BNP per arbejdstime. I modellen med den lineære trend forventes vækstraten at falde med en konstant værdi (henholdsvis ca. 0,05 og 0,1 procent point om året afhængig af om modellerne estimeres på PWT eller DST data) mens faldet forventes at aftage i modellen med den inverse trend, således at de store fald forekom tidligt i estimationsperioden, mens faldene i slutningen af estimationsperioden er klart mindre (gående mod nul). De to forløb er illustreret i figur 6. Øverste plot i figur 6 er vækstraten i BNP per arbejdstime med de to fittede forløb indlagt, baseret på data fra PWT. Nederste plot viser resultaterne for DST data.

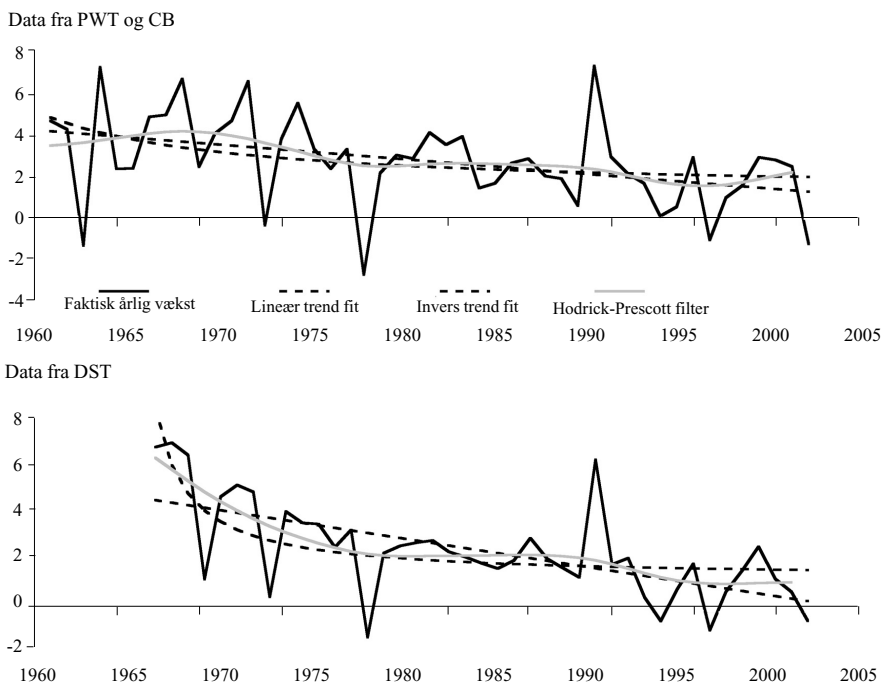


Figur 5. Faktiske, estimerede og forudsagte værdier for de årlige vækstrater i BNP per beskæftiget, per arbejdstime samt per indbygger.

Note: Forudsigelserne bygger på gennemsnittene estimeret på delperioden 1967-1993.

Det er oplagt at den ene model ikke er et specialtilfælde af den anden (modellerne er ikke »nestede«), men man kan formulere en fælles model, som indeholder både en lineær trend og en invers trend. Vi kan dermed teste, hvorvidt den ene model »indeholder« (encompasser) den anden ved at teste, hvorvidt den fælles model kan reduceres til enten den ene eller den anden formulering. Sådanne test er rapporteret i sidste række i tabel 3. Det fremgår, at modellerne for de PWT-baserede vækstrater er meget ens. Der er ikke statistisk belæg for at foretrække den ene modelformulering frem for den anden. For de DST-baserede modeller er der derimod en entydig afgørelse, idet modellen med den inverse trend foretrækkes frem for modellen med den lineære trend, da den lineære trend ikke er signifikant på et 5% testniveau, mens den inverse trend er statistisk signifikant.

Ser man på figur 6, er det tydeligt, at de to modeller giver ens resultater for de PWT baserede vækstrater, idet faldet i vækstraterne er meget lille og jævnt fordelt over tiden. For vækstraterne baseret på DST-data observerer vi de store fald i vækstraten i BNP per arbejdstime i 1970'erne, mens forløbet fra starten af 1980'erne og frem er væsentligt mere fladt.



Figur 6. Faktiske og estimerede værdier for de årlige vækstrater i BNP per arbejdstime.

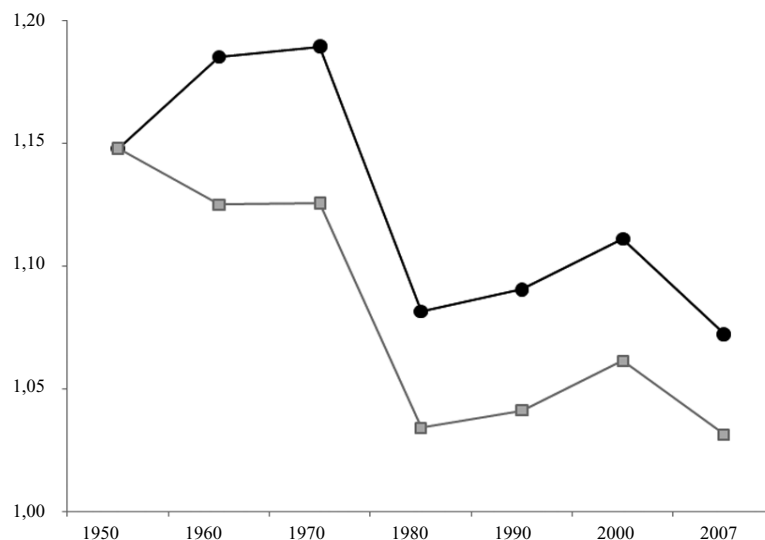
Note: Der er tre estimerede værdier: To modelbaserede, svarende til modellerne i tabel 3, samt en ikke-parametrisk beregning baseret på Hodrick-Prescott filtret (med en udglatningsparameter på 100).

For at understøtte konklusionerne fra regressionsanalyserne har vi indtegnet udglattede værdier af vækstraterne i figur 5. De udglattede værdier er beregnet med Hodrick-Prescott filtret (med en udglatningsparameter på 100). Det ses, at der er god overensstemmelse mellem de semiparametriske og de parametriske fit, idet de HP-filtrede værdier overalt ligger ganske tæt på de beregnede værdier fra modellen med den inverse trend. Dette understøtter, at de store ændringer i vækstraterne forekom tidligt i estimationsperioden, samt at den senere periode er langt mere stabil.

Baseret på den historiske udvikling er det derfor vanskeligt at slutte, at vi har oplevet et signifikant fald i den gennemsnitlige vækstrate i BNP per arbejdstime på et eller andet tidspunkt efter 1995.

3. Komparativ udvikling

Selvom den danske produktivitetsvækst altså ikke har udvist tegn på at aftage i nyere tid, kan det godt være tilfældet, at Danmark klarer sig skidt i et komparativt perspektiv. I ØE's rapport fremhæves det netop, at den danske levestandard (BNP per indbygger) er i aftagende relativt til det gennemsnitlige niveau i OECD.



Figur 7. Danmarks BNP per indbygger relativt til gennemsnittet i OECD: 1950-2007.

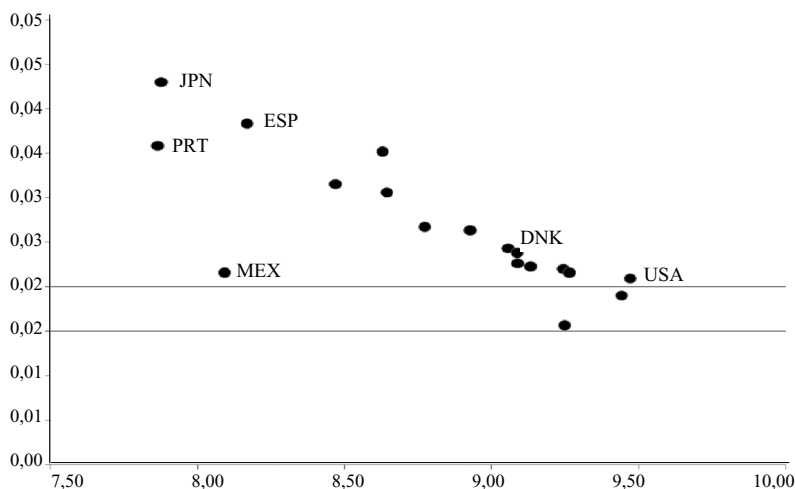
Note: (1) Her følger vi Økonomi- og Erhvervsministeriet og udelader Polen, Ungarn, Tjekkiet og Luxemborg samt Norge siden landets BNP per indbygger er følsomt over for udsving i olieprisen. (2) For den øverste kurve skifter antallet af underliggende observationer, da der ikke er data for Tyskland i 1950 og 1960 samt for Grækenland og Korea i 1950. Den nederste kurve holder antallet af lande konstant, og udelader dermed Tyskland, Grækenland og Korea i alle år.

Kilde: Penn World Table Version 6.3 og egne beregninger.

Danmarks BNP per indbygger relativt til det gennemsnitlige niveau i OECD er faktisk faldet over de sidste 50-60 år. Dette faktum er illustreret i figur 7 (se også ØE, 2009, figur 1). Samtidig er klart, når udviklingen betragtes årti for årti, at der er tale om visse svingninger. Således erstattes faldet i 70'erne af en stigning frem mod 2000, hvorefter den relative indkomst viger en smule. Men samlet set er der tale om et fald over hele perioden; Danmarks relative indkomst viger med ca. 10 procent point set over hele perioden. Der er imidlertid en simpel årsag til, at denne udvikling ikke bør anses for at være specielt urovækkende: Betinget konvergens.

Det er en veletableret empirisk regularitet, at der kan spores en negativ sammenhæng mellem vækst og initial BNP per indbygger blandt lande på forskellige indkomstniveauer, når disse er tilpas ens i forhold til (dvs. »betinget af«) væsentlige vækstdeterminanter såsom investeringsrater i fysisk og humankapital, teknologi adoptionsrater etc.⁵ Noget sådant gør sig tydeligvis gældende i klubben af OECD lande

5. Den klassiske reference er Barro (1991); se Dalgaard og Kreiner (1999) for en basal indføring.



Figur 8. Konvergens i OECD 1950-2007.

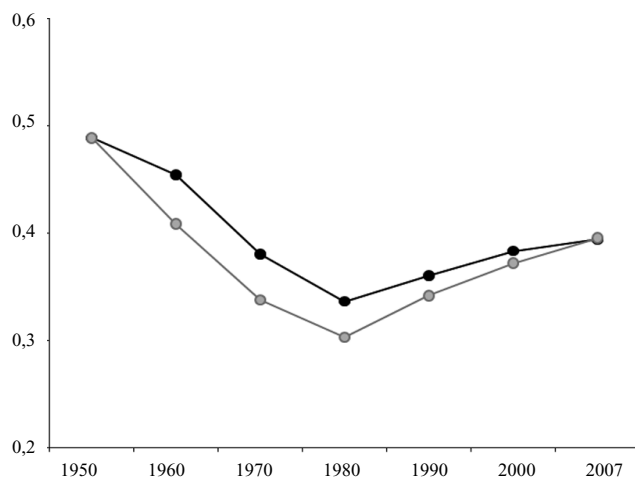
Note: Figuren viser sammenhængen mellem den gennemsnitlige vækstrate 1950-2007 vs. (log) BNP per indbygger i 1950. (1) Her følger vi Økonomi- og Erhvervsministeriet og udelader Polen, Ungarn, Tjekkiet og Luxemborg samt Norge siden landets BNP per indbygger er følsomt over for udsving i olieprisen. Desuden mangler der data for Tyskland, Grækenland og Korea i 1950.

Kilde: Penn World Table Version 6.3 og egne beregninger.

(der jo i alt fald er ens nok til at være del af den samme »klub«), hvilket figur 8 illustrerer for perioden 1950-2007.

Der er flere forskellige teoretiske forklaringer på fænomenet »betinget konvergens«, herunder aftagende grænseproduktivitet samt de fordele, der måtte bestå i at være initialt teknologisk relativt tilbagestående; teknologisk »catch-up« opstår, fordi det er enklere at adoptere eksisterende teknologier fra omverdenen end i det væsentlige at skulle forlade sig på den mere langsommelige forskning og udviklingsproces. Den præcise årsag desuagtet, så er konsekvensen, at de relativt fattigere landes indkomst øges hurtigere end de initialt riges (forudsat, som nævnt, at landene deler strukturelle karakteristika). Resultatet er, at de relative indkomstforskelle skrumpet ind over tid. Det er denne proces, vi ser resultat af i figur 8 ovenfor. En anden illustration af konvergensprocessen er at finde i figur 9, der viser udviklingen i indkomstspredningen for gruppen af OECD lande; figuren dokumenterer, at fra 1950 og frem er der tale om aftagende relative indkomstforskelle OECD landene i mellem.

Det er således forventeligt, at Danmarks BNP per indbygger viger relativt til gennemsnittet i OECD; faktisk vil vi forvente, at denne proces fortsætter i de kommende år. Midlertidige afvigelser fra denne tendens kan dog utvivlsomt ventes, som figur 7



Figur 9. Variationen i BNP per indbygger, OECD 1950-2007.

Note: Figuren viser udviklingen i variationskoefficienten (spredning divideret med gennemsnit) på udvalgte tidspunkter 1950-2007. (1) Her følger vi Økonomi- og Erhvervsministeriet og udelader Polen, Ungarn, Tjekkiet og Luxemborg samt Norge siden landets BNP per indbygger er følsomt over for udsving i olieprisen. (2) For det øverste kurve skifter antallet af underliggende observationer da der ikke er data for Tyskland i 1950 og 1960, samt for Grækenland og Korea i 1950. Den nederste kurve holder antallet af lande konstant, og udelader dermed Tyskland, Grækenland og Korea i alle år.

Kilde: Penn World Table Version 6.3 og egne beregninger.

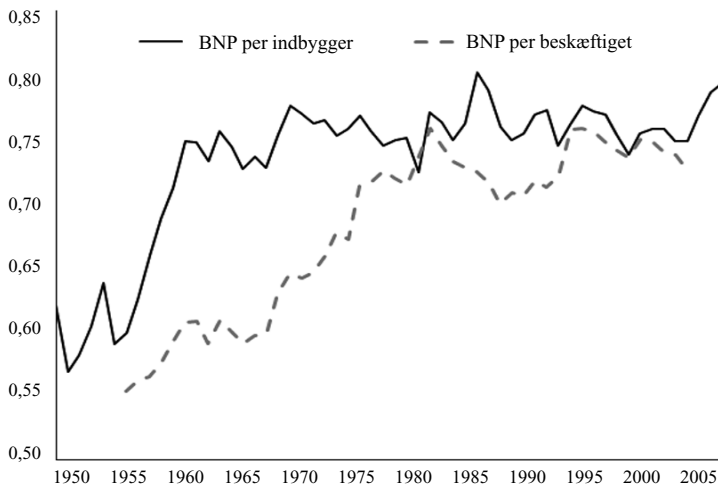
da også antyder. Men disse afvigelser skal primært tilskrives asynkrone konjunkturforløb snarere end langsigtede tendenser.

Konvergenseffekten betyder selvfølgelig også, at Danmark burde indhente (i en relativ forstand) nationer, der i 1950 var rigere end os. Figur 10 afbilleder udviklingen i BNP per indbygger og per beskæftiget i Danmark relativt til USA fra 1950 og frem til i dag.

Set over hele perioden 1950-2007 er tendensen til konvergens tydelig: over tid er Danmark blevet rigere og mere produktivt, relativt til USA. Hvis man ser nøjere efter, står det imidlertid også klart, at tempoet på konvergensten er aftaget i den sidste del af perioden sammenlignet med den første del af perioden.

En væsentlig forklaring på denne tendens til afmatning i konvergensten i arbejds-kraftproduktiviteten er at finde i figur 11, der viser udviklingen i det relative timetal per person i arbejdsstyrken i de to lande. I 1950 udbød danskeren i gennemsnit 7% flere timer per år; i 2007 er det amerikaneren der arbejder omkring 10% mere per år.

En stor del af forklaringen på det produktivitetsgab, vi ser i figur 10, skyldes dermed, at danskerne i gennemsnit lægger færre timer på arbejdsmarkedet; og det er i sin tur utvivlsomt tildels forårsaget af en hastigere tilbagetrækning. I fald det relative timetal havde været uændret over perioden, ville gabet imellem BNP per beskæftiget i



Figur 10. Forholdet mellem BNP per beskæftiget samt per indbygger i Danmark og USA.

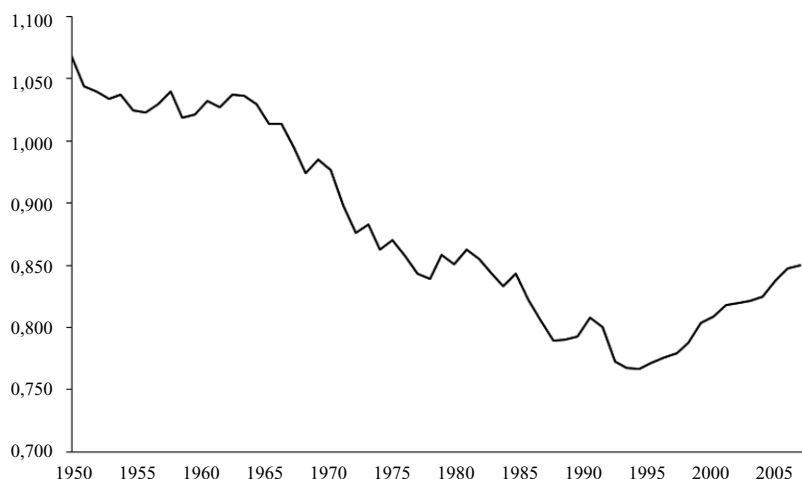
Note: BNP per beskæftiget er kun til rådighed for perioden 1956-2004 for USA.

Kilde: Penn World Table Version 6.3.

Danmark og USA i dag være mindre. Mere konkret ville forholdet i BNP per beskæftiget være 0,85, fremfor 0,73 i 2004 (som er sidste år vi har data for BNP per beskæftiget). Dette »løft« tydeliggør, at den del af konvergensprocessen mellem Danmark og USA, der vedrører kapitalintensitet (fysisk, human) og total faktor produktivitet, har været en persistent del af virkeligheden over perioden 1950-2004.

Umiddelbart kan det forekomme overraskende, at den komparative analyse af dansk vækst antyder, at konvergenseffekter har spillet en vigtig rolle, mens tidsrækkeanalysen peger på, at væksten i BNP per indbygger bedst ses som konstant. Teoretisk vil man nemlig vente, at økonomiens vækstrate kun er »konstant«, når fuld tilpasning til den langsigtede steady state er nået; i tilpasningen (dvs. i konvergensprocessen) bør vækstraten derimod være aftagende, se f.eks. Dalgaard og Kreiner (1999). Så hvordan forener vi konstant vækst med konvergens?

Svaret er givet i en elegant analyse af Jones (2002). Ganske som Danmark har væksten i BNP per capita været meget stabil i USA. Men, som Jones viser, er dette ikke ensbetydende med, at den amerikanske økonomi er i steady state. Stabiliteten kan nemlig også ses som udtryk for modsatrettede påvirkninger af væksttrenden. På den ene side vil fysisk kapital akkumulation ventes at lede til aftagende grænseudbytte og derigennem gradvist aftagende vækst. På den anden side vil (bl.a.) humankapital ak-



Figur 11. Udvikling i arbejdstiden i Danmark relativt til USA.

Note: Figuren viser udviklingen i det årlige antal arbejdstimer per person i arbejdsstyrken i Danmark relativt til USA.
Kilde: The Conference Board Total Economy Database, January 2010.

kumulation ventes at trække modsat. I praksis viser det sig, at disse mekanismer netop er gået (nogenlunde) lige op for USA's vedkommende. Derved genereres en »konstant vækststi«, selvom økonomien ikke er i steady state. Noget tilsvarende må således gøre sig gældende for Danmark. Det kunne være interessant at gennemføre Jones' analyse på danske data; men dette ligger uden for sigtet med nærværende analyse.

4. Konkluderende bemærkninger

Den forudgående analyse leder til tre overordnede konklusioner:

1. Der er ingen trend i væksten i BNP per beskæftiget eller i væksten i BNP per indbygger
2. Der er muligvis en aftagende deterministisk trend (men ikke en stokastisk trend) i væksten i timeproduktiviteten. Men vækstfaldet var primært i 1970'erne. Der er ingen statistisk signifikant tendens til, at trenden i den danske produktivitetsvækst har været aftagende i de senere år. Dette udelukker naturligvis ikke, at man (mere eller mindre) arbitrært kan vælge delperioder ud, hvor væksten synes at være markant forskellig.

Set lidt fra fugleperspektiv fortæller tidsrækkeanalysen altså, at Danmark (ganske som store dele af resten af verden) oplevede et fald i produktivitetsvæksten i 1970'erne (kendt som »The Productivity Slowdown«). Dette fald har Danmark tilsyneladende aldrig rigtigt kommet sig over.

Den tredje konklusion er:

3. Over perioden 1950-2007 har den danske vækst ikke været usædvanlig lav i et komparativt OECD perspektiv.

Der er intet specielt uset eller urovækkende i, at Danmarks BNP per indbygger (eller per beskæftiget, per arbejdstime) aftager i forhold til *gennemsnittet* i OECD. Dette mønster skal tilskrives konvergensprocessen, hvorved fattigere, velfungerende lande udviser en tendens til »catch-up« i forhold til de initialt rigere lande. Denne proces må ventes at fortsætte i de kommende år. Det er værd at bemærke, at selvsamme konvergensproces har manifesteret sig i, at Danmark over de sidste ca. 60 år er blevet rigere og mere produktivt i forhold til USA, som var det rigeste land i OECD-klubben i 1950.

Endelig bør man nok som konsekvens af de to første konklusioner se nærmere på Danmarks Statistiks opgørelse af det samlede antal arbejdstimer. Der er et databrud i 1995, idet datakilderne ændres, og man kan se, at udviklingen i antallet af arbejdstimer per beskæftiget i perioden fra ca. 1995/6 til 2000/01 er ganske usædvanlig i et historisk perspektiv. Dette *kan* skyldes de meget lange konjunkturopsving fra 1994/5 og frem, men det *kan* også være et opgørelsesproblem, som påvirker beregningen af vækstraten i timeproduktiviteten.

Litteratur

- Arbejderbevægelsens Erhvervsråd. 2009. Sløj produktivitet bremser dansk velstand. <http://www.ae.dk/analyse/sloj-produktivitets-bremser-dansk-velstand>.
- Barro, R. 1991. Economic Growth in a Cross-Section of Countries. *Quarterly Journal of Economics*, 106, 407-43.
- CEPOS. 2008. Velstand blandt OECD-lande: Danmark på vej mod midterplacering«. <http://www.cepos.dk/publikationer/analyser-notater/analyse/single/artikel/velstand-blandt-oecd-lande-danmark-paa-vej-mod-midterplacering/>
- Dalgaard, C-J. og H. Hansen. 2010. Er der et dansk produktivetskollaps? *EPRU Analyse*, 2010/27. Økonomisk Institut, Københavns Universitet.
- Dalgaard, C-J. og C. T. Kreiner. 1999. Global indkomstulighed i et vækstteoretisk perspektiv *Nationaløkonomisk Tidsskrift*, 137, 305-32
- Doornik, J. A. and D. F. Hendry. 2009. *Empirical Econometric Modelling – PcGive 13: Volume 1*, Timberlake Consultants Ltd., London.
- Hansen, B. E. 1992. Testing for parameter instability in linear models, *Journal of Policy Modeling*, 14(4), 517-33.
- Heston, A., R. Summers og B. Aten. 2009. *Penn World Table Version 6.3*, Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania, August 2009.
- Jones, C. I. 2002. Sources of US economic growth in a world of ideas. *American Economic Review*, 92, 220-39
- Lin, C.-F. J. and T. Terasvirta. 1994. Testing the constancy of regression parameters against continuous structural change, *Journal of Econometrics*, 62(2), 211-28.
- Nyblom, J. 1989. Testing for the constancy of parameters over time, *Journal of the American Statistical Association*, 84, 223- 30.
- Økonomi- og Erhvervsministeriet. 2009. Den danske produktivetsudvikling. *Økonomisk Tema* nr. 8.

Appendiks: Datadokumentation

Tabel A1. Sammenligning af vækstraterne baseret på henholdsvis Penn World Table Version 6.3 og Danmarks Statistik.

Vækst i	Gennemsnit 1967-2007		t-test af forskel	Korrelation
	PWT	DST		
BNP per beskæftiget	1,96	1,91	0,17	0,57
BNP per arbejdstime	2,54	2,52	0,17	0,89
BNP per indbygger	2,14	2,08	0,45	0,93

Vi benytter 3 datakilder:

1. Alan Heston, Robert Summers and Bettina Aten (2009) Penn World Table Version 6.3, Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania, August 2009. <http://pwt.econ.upenn.edu>
2. The Conference Board Total Economy Database, January 2010, <http://www.conference-board.org/economics/database.cfm>
3. Statistikbanken, <http://www.statistikbanken.dk>

Data fra Penn World Table Version 6.3 (PWT) er: *rgdpch* (Real GDP per capita, chain index), *rgdpl2* (Real GDP per capita, Laspeyre index) og *rgdpl2te* (Real GDP per total employed, Laspeyres index).

Data fra the Conference Board on Total Economy Database (CB) er *thwdnk* (Total hours worked, Denmark).

Data fra Statistikbanken (Danmarks Statistik):

BNP: NAT02: B1*g Bruttonational produkt, BNP, 2000-priser kædede værdier

Befolkning: NAT17: Gennemsnitsbefolkning

Beskæftigelse: NAT18: Samlet antal beskæftigede ekskl. orlov, i alt

Arbejdstimer: NAT18: Samlede præsterede timer, i alt

Følgende BNP mål beregnes ud fra PWT samt CB :

BNP per beskæftiget: $rgdpch * rgdpl2te / rgdpl2$

BNP per arbejdstime: $rgdpch * pop / thwdnk$,

BNP per indbygger: $rgdpch$

Følgende BNP-mål beregnes ud fra Statistikbanken:

BNP per beskæftiget: BNP/Beskæftigelse

BNP per arbejdstime: BNP/Arbejdstimer

BNP per indbygger: BNP/Befolkning

Sundhedsudgifter og finanspolitisk holdbarhed

Marianne Frank Hansen

Danish Rational Economic Agents Model, DREAM, E-mail: mfr@dreammodel.dk

Lars Haagen Pedersen

De Økonomiske Råds Sekretariat, DØRS, E-mail: lh@dots.dk

SUMMARY: Average growth in Danish healthcare expenditures relative to GDP has been among the lowest in the OECD area since 1970, but growth rates are rapidly increasing. The analysis separates expenditure growth according to demographic and non-demographic factors. As an innovation the effects of so-called healthy ageing are included into the demographic effects. Annual demographic and non-demographic real growth in publicly financed health expenditures is estimated to be 0.4 pct. and 2.0 pct. respectively for the period 1993-2008. Average non-demographic growth in health-care expenditures exceeds average annual growth rate in real productivity per working hour by 0.9 pct. in the period. Fiscal sustainability is shown to be very sensitive to non-demographic cost increases in excess of productivity growth: A sustained non-demographic expenditure growth of 0.3 pct. in excess of the productivity growth increases the fiscal sustainability problem by 2.1 pct. of GDP. Therefore current growth in non-demographic healthcare expenditures cannot be maintained for a longer period without challenging the public financing of healthcare expenditures in Denmark. On the other hand fiscal sustainability is shown to be robust with respect to growth in healthcare expenditures due to future increases in life expectancy. This is a consequence of healthy ageing and the indexation of the statutory retirement age to life expectancy.

Indledning

Sundhedsydelser er en central del af produktionen i alle udviklede lande. Der er en tendens til, at sundhedsudgifternes andel af den samlede produktion er højere jo mere velstående landet er. I Danmark udgør sundhedsydelser omkring 10 pct. af produktio-

Resultaterne, der præsenteres i denne analyse, er sammenfaldende med resultaterne i Det Økonomiske Råd: *Dansk Økonomi efterår 2009*. Synspunkterne i dette papir er forfatterens og er ikke nødvendigvis sammenfaldende med vurderinger fra Det Økonomiske Råds formandskab. Søren Arnberg, Niels Henning Bjørn, Thomas Bue Bjørner, Anne Kristine Høj og Peter Stephensen takkes for konstruktive bidrag og kommentarer.

nen afhængigt af den konkrete afgrænsning. Finansieringen af de danske sundhedsydelse er hovedsagelig offentlig, hvilket har baggrund i sundhedsloven, der, som et vigtigt formål, angiver, at der skal være »let og lige adgang til ydelser fra sundhedsvæsenet«. Hermed betragtes sundhedsydelser som en ydelse, der ikke skal være afhængig af den enkeltes økonomiske formåen, men bestemt af det medicinske behov.

Hvis der tages udgangspunkt i sundhedslovens formål, bliver såvel sundhedsydelserne som deres finansiering centrale områder for velfærdssamfundet. I denne artikel søger vi at vurdere den fremtidige udvikling i sundhedsudgifterne på basis af dels de historiske tendenser og dels den forventede udvikling i befolkningens alderssammensætning. Under antagelse af, at udgifterne også i fremtiden finansieres primært via de offentlige budgetter, gives en vurdering af betydningen af udviklingen i sundhedsudgifterne for det finanspolitiske holdbarhedsproblem.

Analysen af den forventede udvikling i sundhedsudgifterne baseres på estimationer af stigningen i de aldersbetingede sundhedsudgifter i den historiske periode. Udgangspunktet er, at den forventede fremtidige vækst i sundhedsudgifterne analytisk kan opdeles i en demografisk betinget del og en resterende del, hvor sidstnævnte dels kan tilskrives en højere efterspørgsel som følge af øget økonomisk velstand og dels et større udbud som følge af udviklingen i lægevidenskaben, som betyder, at stadig flere sygdomme kan behandles, jf. Hall og Jones (2007) for en teoretisk og empirisk analyse af dette fænomen. I den historiske udvikling siden 1970 har det demografiske bidrag været begrænset, da perioden er præget af en meget lav vækst i restlevetiden for ældre frem til 1995 samt et forholdsvist stabilt antal af ældre. Begge fænomener er ændret i de seneste år således, at restlevetiden for ældre stiger markant og de generationer, der i fremtiden trækker sig tilbage vil være væsentlig større end de nuværende generationer af ældre.

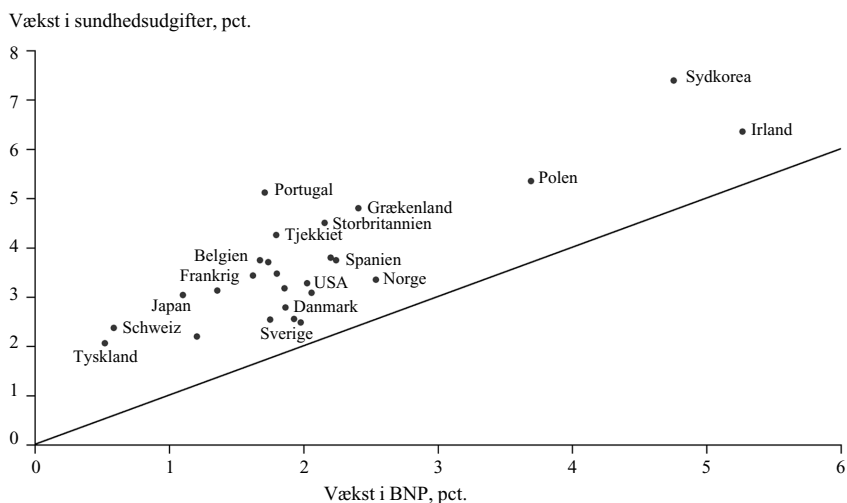
Udviklingen betyder, at det forventes, at den demografisk betingede del af udgiftsvæksten vil være stigende, fordi der er udsigt til et stigende antal ældre i befolkningen, og fordi sundhedsudgifterne er voksende med alderen. En del af stigningen i antallet af ældre skyldes den forventede stigende restlevetid for ældre, mens den resterende del af stigningen er midlertidig og skyldes de historisk store fødselsårsgange i 1940'erne. Der er en betydelig litteratur, som diskuterer sammenhængen mellem væksten i midlertidigheden og sundhedsudgifternes aldersafhængighed, jf. Fuchs (1984, 1990), Lublitz og Riley (1993), Cutler og Sheiner (1998), Lee og Miller (2002), Breyer og Felder (2004), Seshamani og Grey (2004a, 2004b) samt Arnberg og Bjørner (2009). Udgangspunktet for litteraturen er en opdeling af sundhedsudgifterne i udgifter til såkaldt terminalbehandling og øvrige sundhedsudgifter, hvor terminaludgifter fastlægges som sundhedsudgifter til patienter, der dør inden for en given periode. En betydelig del af sundhedsudgifterne til ældre bliver med denne opdeling karakteriseret som terminaludgifter.

Ud af de samlede offentlige sundhedsudgifter til personer på 60 år eller derover i år 2000 blev 51 pct. anvendt på personer, der var afgået ved døden inden udgangen af år 2005, dvs. personer, hvor restlevetiden udgjorde mindre end 6 år. I det omfang disse udgifter er forbundet med behandling af den lidelse, som fører til døden, kan man argumentere for, at en længere forventet restlevetid vil føre til, at en betydelig del af de pågældende udgifter skal afholdes senere i personens liv. Dermed vil terminaludgifterne ikke afhænge af personens alder, men primært af personens resterende levetid. Denne effekt er afgørende, når restlevetiden for personer i en given alder vokser. En effekt af denne type betyder, at de samlede sundhedsudgifter pr. person i en given aldersgruppe alt andet lige vil falde, når restlevetiden stiger, idet afholdelsen af terminaludgifterne udskydes i takt med forøgelsen af restlevetiden. Dette fænomen kaldes »sund aldring« og indebærer en reduktion i den del af den demografisk betingede forventede vækst i sundhedsudgifterne, som skyldes forøget restlevetid for ældre. Arnberg og Bjørner (2009) estimerer terminaludgifternes omfang i Danmark og vurderer betydningen for den forventede fremtidige demografisk betingede udvikling i sundhedsudgifterne. Analysen i dette papir tager udgangspunkt i resultaterne fra Arnberg og Bjørner.

Den del af den forventede fremtidige vækst i sundhedsudgifterne, der ikke er demografisk betinget, baseres på en estimation af den historiske udvikling i denne del af sundhedsudgifterne. Analysen er en videreudvikling af metoden præsenteret i Pedersen og Hansen (2006). Mens Pedersen og Hansen (2006) opdeler væksten i sundhedsudgifterne i en del, der skyldes ændret alderssammensætning og en resterende del, tages der i nærværende analyse herudover højde for den i Arnberg og Bjørner (2009) estimerede effekt af »sund aldring«, således at den historiske vækst i sundhedsudgifterne opdeles i en demografisk betinget vækst inklusiv »sund aldring« og en resterende del.

Med udgangspunkt i DREAMs befolkningsfremskrivning, de initiale aldersbetingede sundhedsudgifter og estimationerne af henholdsvis »sund aldring« og den resterende vækst i sundhedsudgifterne kan udviklingen i sundhedsudgifterne fremskrives. Denne fremskrivning anvendes som input i den generelle ligevægtsmodel DREAM, og udviklingens betydning for den finanspolitiske holdbarhed vurderes på basis af beregninger med modellen.

Den resterende del af papiret er organiseret på følgende måde: I første afsnit præsenteres de overordnede træk af udviklingen i sundhedsudgifterne i en række OECD-lande herunder Danmark i perioden 1970-2006. Andet afsnit præsenterer metode og resultater af estimationen af den ikke-demografisk betingede vækst i sundhedsudgifterne, når der er taget højde for »sund aldring«. I tredje afsnit fremskrives sundhedsudgifternes andel af BNP, og i fjerde afsnit vurderes sundhedsudgifternes betydning for



Figur 1. Gennemsnitlig årlig realvækst i sundhedsudgifter pr. capita og BNP pr. capita fra 1970 til 2006.

Anmærkning: Landenavne er kun angivet for udvalgte lande. Den optrukne linie er 45-graders linien, der angiver samme vækstrate i sundhedsudgifter og BNP.

Kilde: OECD Health Data.

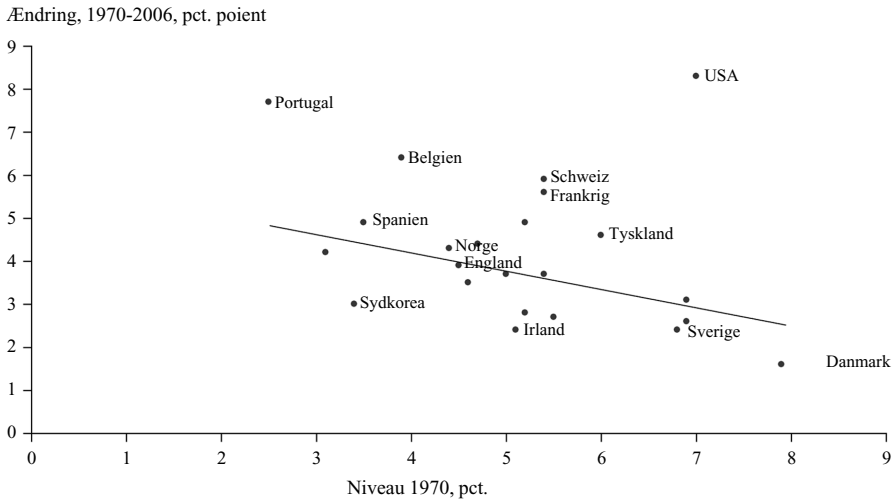
den finanspolitiske holdbarhed. Analysens hovedkonklusioner sammenfattes i det afsluttende afsnit.

Udviklingen i sundhedsudgifterne i de seneste årtier

Sundhedsudgifternes andel af BNP har været voksende i alle OECD-lande i den periode, hvor der foreligger data, dvs. fra 1970 til 2006. I 1970 udgjorde sundhedsudgifterne i gennemsnit 5 pct. af BNP i OECD-landene, mens de i 2006 udgjorde 9 pct. i gennemsnit.¹ Opgjort ved den gennemsnitlige årlige reale vækstrate i henholdsvis sundhedsudgifter og BNP pr. capita har alle lande en mervækst i de reale sundhedsudgifter i forhold til BNP. Størst forskel er der for Portugal, hvor den gennemsnitlige årlige realvækst i sundhedsudgifterne pr. capita er på godt 5 pct., mens realvæksten i BNP pr. capita er på knap 2 pct. i samme periode. I den anden ende af skalaen ligger lande som Sverige og Danmark, hvor den gennemsnitlige realvækst i sundhedsudgifterne pr. capita kun er 1/2-1 pct. point højere end væksten i BNP pr. capita, jf. figur 1.

Selvom alle lande har en betydelig mervækst i sundhedsudgifterne i forhold til væksten i BNP, er der også en stor spredning, som kunne pege på, at sundhedsudgifterne i en række OECD-lande »catcher up« i forhold til lande, som i udgangspunktet havde de højeste sundhedsudgifter. Dette er i et vist omfang tilfældet, men spredningen

1. OECDs opgørelse af sundhedsudgifter inkluderer flere udgifter til ældrepleje end Danmarks Statistiks opgørelse. For Danmarks vedkommende udgør forskellen ca. 25 mia. kr. i 2006, jf. DØR (2009).



Figur 2. Vækst i sundhedsudgifternes andel af BNP sat i forhold BNP-niveaueet i 1970.

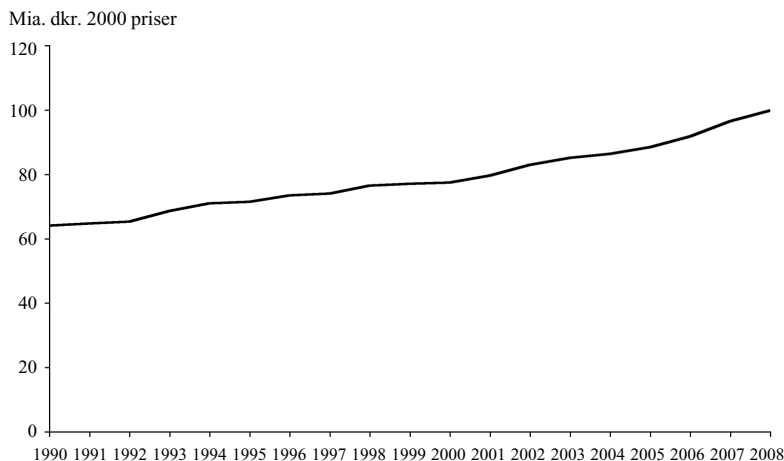
Anmærkning: Landenavne er kun angivet for udvalgte lande.

Kilde: OECD Health Data.

mellem landene er stor, jf. figur 2. USA skiller sig således ud som det land, der har haft den største vækst i sundhedsudgifternes andel af BNP i perioden, selvom USA også i udgangspunktet i 1970 var blandt de lande, som havde de højeste sundhedsudgifter. USA's sundhedsvæsen er i højere grad efterspørgselsstyret end de fleste andre OECD-lande, mens lande som Danmark og Sverige, der dels havde et højt udgangsniveau for sundhedsudgifterne og dels er blandt de lande, der har haft den laveste mervækst i perioden, begge er relativt udbudsstyrede (via generelle rammebevillinger, der ikke afhænger af efterspørgslen) i størstedelen af den pågældende periode.

Udviklingen i de danske sundhedsudgifter afviger fra den generelle internationale udvikling. Danmark var i 1970 det OECD-land, som havde de højeste sundhedsudgifter i pct. af BNP på ca. 8 pct. Frem til starten af 1980'erne steg sundhedsudgifterne til godt 9 pct. af BNP, hvorefter der skete en betydelig opbremsning i udgiftsvæksten således, at sundhedsudgifternes andel af BNP gradvist faldt frem mod 1995, hvor den udgjorde godt 8 pct. af BNP. Fra omkring 1995 følger udviklingen i de danske sundhedsudgifter stort set de øvrige OECD-landes og særligt i det seneste årti har stigningen i sundhedsudgifterne været betydelig.²

2. Der er et bemærkelsesværdigt sammenfald mellem den relative udvikling i sundhedsudgifterne og den relative udvikling i midlellelivetiden. I perioden fra ca. 1980 til 1995 steg den danske midlellelevetid med ca. 1 md. om året, hvilket er væsentligt langsommere end i de øvrige OECD-lande, hvor stigningen var omkring 2-3 mdr. om året. Fra 1995 begyndte den danske midlellelevetid at stige i samme omfang som midlellelevetiden i de øvrige OECD-lande.



Figur 3. *Udviklingen i de reale individuelle offentlige sundhedsudgifter, mia. dkr. (2000-priser).*

Kilde: Danmarks Statistik samt egne beregninger.

Der er to hovedforklaringer på de seneste års forøgede stigningstakt i sundhedsudgifterne. For det første har der været forøget politisk opmærksomhed på sundhedssektoren blandt andet som følge af fokus på at nedbringe den danske merdødelighed for centrale sygdomsområder som hjerte-/karområdet og senest kræftområdet. For det andet er de danske sundhedsudgifter i højere grad blevet efterspørgselsstyret gennem indførelsen af frit sygehusvalg og udvidet frit sygehusvalg. Det frie sygehusvalg indebærer, at patienter, der er blevet henvist til et offentligt sygehus har mulighed for at vælge blandt de offentlige sygehuse. Det udvidede frie sygehusvalg indebærer, at hvis de offentlige sygehuse ikke kan foretage behandlingen inden for de fastlagte tidsfrister, kan patienten vælge at blive behandlet på et privat sygehus eller klinik, som regionen har en aftale med. Den ændrede styring af sundhedssektoren, der betyder, at sundhedsproduktionen i højere grad styres gennem ventetidsgarantier, har primært ført til stigninger i regionernes udgifter til bl.a. sygehuse samt statens udgifter til sygesikring og medicin. Udviklingen i de reale individuelle offentlige sundhedsudgifter er illustreret i figur 3. Samlet har den gennemsnitlige årlige realvækst i de danske sundhedsudgifter været på 2,4 pct. i perioden fra 1993-2008 og på 3,0 pct. i perioden 1999-2008.³

Mervækst i sundhedsudgifter korrigeret demografi og »sund aldring«

Med det formål at belyse årsagerne til udviklingen i sundhedsudgifterne udføres i

3. Reale vækstrater er beregnet ud fra udviklingen i de individuelle offentlige udgifter til sundhed deflateret med udviklingen i BNP kædeindekset, Statistikbanken, Danmarks Statistik.

nærværende afsnit en analyse af væksten i sundhedsudgifterne i Danmark gennem perioden 1993-2008. Analysen resulterer i en dekomponering af den reale vækst i sundhedsudgifterne i bidrag, der kan tilskrives henholdsvis demografiske og ikke-demografiske faktorer. Realvæksten i de ikke-demografiske faktorer vurderes i forhold til udviklingen i timeproduktiviteten i samfundet. Dette gøres med henblik på at vurdere om væksten i de aldersbetingede sundhedsudgifter pr. person i en given alder vokser hurtigere end den generelle indkomstudvikling. Hvis antallet af arbejdstimer er uændret, stiger den samlede indkomst med samme vækstrate som timeproduktiviteten. Det betyder, at skatteindtægterne stiger med samme vækstrate som timeproduktiviteten, hvis det yderligere antages, at skattesatserne er konstante og progressionsgrænserne reguleres med samme vækstrate som indkomsten. Hvis de ikke-demografisk betingede sundhedsudgifter pr. person vokser med samme vækstrate som produktiviteten, kan denne stigning derfor finansieres uden, at den enkelte persons sundhedsudgifter lægger beslag på en større andel af det samlede skatteprovenu end før væksten. Hvis de ikke-demografisk betingede sundhedsudgifter pr. person derimod udviser en mer-vækst i forhold til timeproduktiviteten, indebærer dette umiddelbart en øget finansieringsbyrde, som enten kræver forøgede skattebetalinger eller tilsvarende reduktioner i andre udgiftstyper.

Udgangspunktet for dekomponeringen er fremgangsmåden beskrevet i Hagist og Kotlikoff (2005), der på danske data tidligere har været anvendt til analyse af væksten i sundhedsudgifterne gennem perioden 1993-2002, jf. Pedersen og Hansen (2006). Den her udførte analyse vil således bidrage til at belyse udviklingen gennem en længere periode, der specielt inkluderer perioden med øget efterspørgselsstyring af sundhedsudgifterne. Samtidig tages der i modsætning til, hvad der tidligere har været tilfældet, højde for »sund aldring« i vurderingen af forholdet mellem det demografiske og ikke-demografiske vækstbidrag.⁴

Med det formål at opdele den samlede vækst i sundhedsudgifterne i en demografisk del og en del, der vedrører udviklingen i det reale ydelsesniveau pr. person, tages afsæt i, at de samlede sundhedsudgifter på tidspunkt t , E_t , kan skrives som summen over alder i af produktet af den gennemsnitlige aldersfordelte sundhedsudgift pr. person, $\varepsilon_{i,t}$, og antallet af personer fordelt på alder, $P_{i,t}$,

$$E_t = \sum_i \varepsilon_{i,t} P_{i,t} \quad (1)$$

4. Herudover tager nærværende analyse udgangspunkt i Danmarks Statistiks afgrænsning af sundhedsudgifterne, mens Pedersen og Hansen (2006) tager udgangspunkt i OECDs sundhedsdata, der som nævnt ovenfor inkluderer flere udgifter end den nationale opgørelse i Danmarks Statistik.

Generelt er det vanskeligt at tilvejebringe data, der for en længere historisk periode på en konsistent måde beskriver udviklingen i de gennemsnitlige aldersfordelte sundhedsudgifter pr. person, $\varepsilon_{i,t}$. Dette motiverer umiddelbart anvendelsen af en aldersfordeling, der er konstant over tid. Sædvanligvis er det kun muligt at anvende en aldersprofil for det gennemsnitlige træk fra et tidspunkt i slutningen af den observerede periode, jf. Pedersen og Hansen (2006). Hvis der i den historiske periode er betydelige stigninger i restlevetiden for givne aldersgrupper, vil denne metode have tendens til at indebære en undervurdering af de aldersbetingede udgifter i starten af den historiske periode, hvis der samtidig har været tendens til »sund aldring«. Dette giver en bias i estimaterne for den demografiskbetingede stigning i sundhedsudgifterne, som indebærer en overvurdering af denne komponent. Arnberg og Bjørner (2009) estimerer imidlertid den historiske effekt af »sund aldring« og muliggør dermed, at de konstante gennemsnitlige aldersfordelte udgifter erstattes af tidsafhængige estimater for samme. Dette gøres ved at kombinere de estimerede effekter af »sund aldring« med den faktiske vækst i den aldersbetingede restlevetid i perioden. De derved dannede estimater for de aldersbetingede udgifter vil sammen med udviklingen i den aldersfordelte befolkning og i de samlede reale sundhedsudgifter (deflateret med udviklingen i BNP kædeindekset) nedenfor danne udgangspunkt for dekomponeringen af væksten i de samlede sundhedsudgifter i en demografisk del, der som noget nyt indregner effekten af »sund aldring« og i en resterende del, der redegør for udviklingen i serviceniveauet.

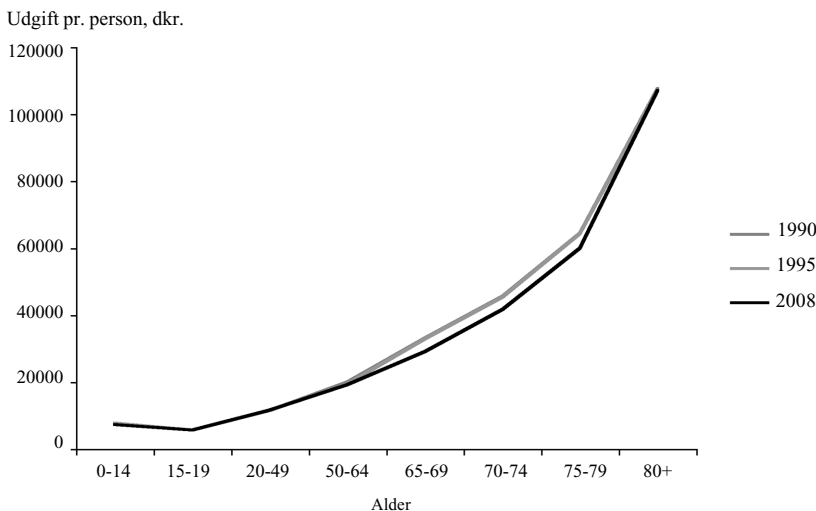
Til estimationsformål opdeles befolkningen i følgende aldersgrupper: 0-14 år, 15-19 år, 20-49 år, 50-64 år, 65-69 år, 70-74 år, 75-79 år og 80-120 år. Det antages, at den reale udvikling i de samlede sundhedsudgifter kan udtrykkes som

$$E_t = \varepsilon_{50-64,b} (1 + \lambda)^{t-b} \sum_i \alpha_{i,t} P_{i,t} \quad (2)$$

hvor $\alpha_{i,t} = \frac{\varepsilon_{i,t}}{\varepsilon_{50-64,b}}$ og $\varepsilon_{50-64,b}$ er sundhedsudgifterne pr. person for aldersgruppen 50-64 år i basisåret b og λ er den gennemsnitlige årlige vækstrate i de gennemsnitlige sundhedsudgifter pr. person. Den normaliserede aldersprofil $\alpha_{i,t}$ afhænger som noget nyt af t og kan betegnes som den tidligere anvendte konstante profil $\alpha_i = \frac{\varepsilon_{i,b}}{\varepsilon_{50-64,b}}$ korri-geret for »sund aldring«. Antages det, at E_t i (2) måles med fejl, kan estimation af det ikke-demografiske vækstbidrag udføres med udgangspunkt i relationen

$$\ln(E_t) - \ln\left(\sum_i \alpha_{i,t} P_{i,t}\right) = \ln(\varepsilon_{50-64,b}) + (t-b) \ln(1 + \lambda) + v_t \quad (3)$$

hvor v_t er et støjled. Her angiver venstresiden den del af de samlede sundhedsudgifter, der ikke kan tilskrives demografiske faktorer.

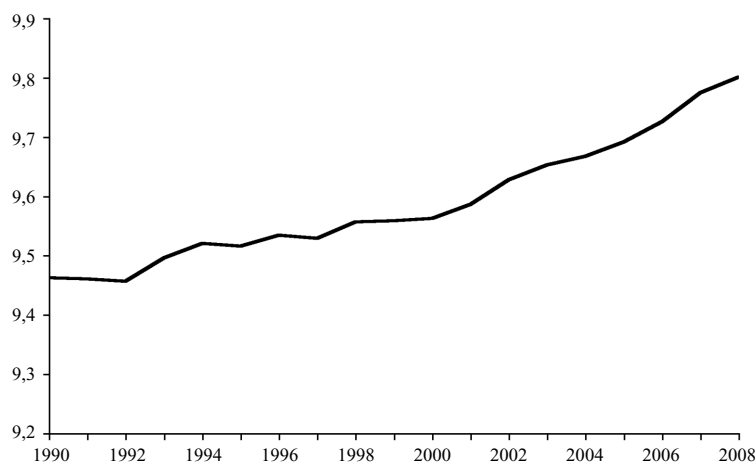


Figur 4. *Udvikling i aldersfordelte sundhedsudgifter pr. person korrejeret for ikke-demografiske effekter (2006-priser), udvalgte år.*

Kilde: Arnberg og Bjørner (2009) samt egne beregninger.

I nærværende analyse stammer data for sundhedsudgifterne fra Danmarks Statistik, jf. figur 3. Udviklingen i de gennemsnitlige aldersfordelte sundhedsudgifter pr. person korrejeret for ikke-demografiske effekter er for udvalgte historiske år illustreret i figur 4. De illustrerede serier danner udgangspunkt for beregning af den normaliserede aldersprofil, $\alpha_{i,t}$, jf. ovenfor.

Over tid ses en tendens til, at de aldersbetingede gennemsnitlige reale sundhedsudgifter pr. person korrejeret for ikke-demografiske effekter stiger mindre med alderen. Fra 1990-1995 er der stort set ikke nogen ændring i aldersfordelingen af sundhedsudgifterne, hvilket hænger sammen med, at der stort set ikke var nogen vækst i restlevetiden for ældre i de pågældende år. Fra 1995 sker der en markant stigning i restlevetiden for de ældre, som indebærer, at den aldersbetingede vækst i sundhedsudgifterne sætter ind ved en højere alder. Fra 1990-95 falder de gennemsnitlige reale sundhedsudgifter pr. person korrejeret for ikke-demografiske effekter med blot 0,6 pct. og 0,4 pct. for henholdsvis 65-69- og 70-74-årige. For personer i aldersgruppen 75-79 år noteres et fald på 0,3 pct. I perioden fra 1995-2008 er de gennemsnitlige reale sundhedsudgifter pr. person korrejeret for ikke-demografiske effekter i alderen 65-69 år faldet med hele 11,5 pct. For henholdsvis 70-74-årige og 75-79-årige observeres fald på 8,2 pct. og 6,7 pct. Generelt er aldersprofilen for sundhedsudgifterne faldende fra aldersgruppen 0-14 år til aldersgruppen omfattende de 15-19-årige. Herefter er den stigende. Dette



Figur 5. Udvikling i de samlede ikke-demografisk betingede sundhedsudgifter, logaritmisk.

Anm: Figuren afbilder venstresiden i relationen (3).

Kilde: Egne beregninger.

svarer til resultatet for de 10 lande, der indgår i undersøgelsen i Hagist og Kotlikoff (2005). Opdeles i kortere aldersintervaller fremgår det, at faldet i udgifterne fra den første til den anden aldersgruppe skyldes høje sundhedsudgifter til helt små børn og er således knyttet til fødsel, barnepleje og vaccinationer.

Relationen (3) indebærer, at venstresiden kan skrives som en lineær funktion af tiden. På danske data synes der imidlertid at være et markant niveauskifte startende i 1993 således, at venstresiden i ligningen frem til dette år kan beskrives ved en meget flad udvikling, mens der efterfølgende kan observeres en stigende udvikling over tid, jf. figur 5. Analysen tager derfor udgangspunkt i perioden 1993-2008.

Estimationen viser, at ud af en realvækst på 2,4 pct. årligt i perioden 1993-2008, kan omkring 2,0 pct. point tilskrives ikke-demografiske forhold. Således udgør det demografiske vækstbidrag, når der tages højde for »sund aldring«, mindre end 20 pct. Indregnes effekten fra »sund aldring« ikke, vil betydningen af det ikke-demografiske bidrag reduceres til 1,8 pct. point. Udeladelse af effekten af »sund aldring« afstedkommer altså en ikke ubetydelig overvurdering af det demografiske vækstbidrag. Stigningen i de reale gennemsnitlige sundhedsudgifter pr. person i en given alder udtrykker den andel af den samlede vækst, der kan tilskrives velstands- eller relative priseffekter.

Det fundne ikke-demografiske bidrag på 2,0 pct. point kan sammenholdes med den faktiske udvikling i den gennemsnitlige årlige vækstrate i timeproduktiviteten i sam-

me periode. Baseret på de foreløbige nationalregnskabstal (november 2009) er denne vækst på 1,1 pct. point. Tallet er dog i betydeligt omfang påvirket af de foreløbige værdier for 2007-08, som er bemærkelsesværdigt lave. Den gennemsnitlige vækstrate i timeproduktiviteten for perioden 1993-2006 er til sammenligning 1,6 pct. point. Den gennemsnitlige årlige mervækst i den seneste 15 års periode kan dermed anslås til mellem 0,4 pct. point og 0,9 pct. point, afhængigt af om de seneste års produktivitetsudvikling medtages.

Restrangeres estimationen til den seneste 10-års periode 1999-2008, hvor udviklingen i de reale sundhedsudgifter som nævnt har været præget af en relativ markant stigning, vil betydningen af det ikke-demografiske bidrag stige, mens det demografiske vækstbidrag reduceres yderligere. Ud af en gennemsnitlig årlig vækst på 3,0 pct. i de samlede sundhedsudgifter skal blot 0,2 pct. point tilskrives demografiske faktorer. Da der i samme periode har været en relativ lav vækst i den gennemsnitlige timeproduktivitet på $1\frac{1}{4}$ pct. i perioden 1999-2006 og på kun $\frac{1}{2}$ pct. i perioden 1999-2008, er der tale om en meget betydelig mervækst i sundhedsudgifterne pr. person på mellem 1,5 og 2,3 pct. point. Mervækstrater i dette omfang kan ikke fastholdes i længere perioder og kan formentlig tilskrives en uheldig kombination af ekstraordinær høj (engangs)vækst i sundhedsudgifterne som følge af overgang til delvist efterspørgselsstyret sundhedsvæsen baseret på overgangen til det udvidede frie sygehusvalg og en ekstraordinært svag udvikling i timeproduktiviteten.

Fremskrivning af sundhedsudgifter

Udgangspunktet for fremskrivningen af de offentlige sundhedsudgifter er de aldersbetingede sundhedsudgifter pr. person i 2006, som er justeret med realvæksten i de samlede sundhedsudgifter frem til 2008 svarende til, at stigningen har påvirket de aldersafhængige udgifter proportionalt.

Væksten i sundhedsudgifterne opdeles i tre elementer: En årgangseffekt, en effekt af stigende restlevetid for ældre, som inkluderer effekten af »sund aldring« og en effekt af mervæksten i de reale sundhedsudgifter pr. person. Alle effekter måles i dette afsnit med udgangspunkt i 2008 pris- og lønniveau.

Årgangseffekten angiver den stigning i sundhedsudgifterne, som ville komme alene som følge af, at de fremtidige årgange af ældre består af større fødselsårgange end de nuværende generationer af ældre. Denne effekt findes ved at fremskrive befolkningen med uændrede aldersbetingede dødeligheder og ved at fastholde de aldersbetingede sundhedsudgifter. Bemærk, at i dette tilfælde er der ingen effekt af »sund aldring«, fordi der ikke er nogen stigning i restlevetiden. Årgangseffekten alene indebærer en gradvis stigning i sundhedsudgifterne på 18 pct. over de 25 år fra 2008-2033. Herefter er der udsigt til et svagt fald, således at stigningen i 2050 er på 17 pct. i forhold til 2008.

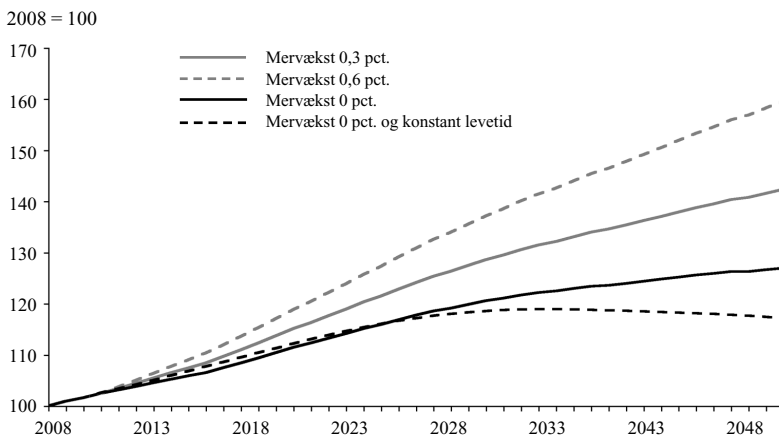
Levetidsforøgelseeffekten angiver stigningen i sundhedsudgifterne som følge af den længere forventede levetid, som er inkluderet i DREAMs befolkningsfremskrivning fra 2009, idet der tages højde for, at stigningen i restlevetiden betyder, at terminaludgifterne indtræder på et senere tidspunkt – dvs. det antages, at der er »sund aldring«. Frem mod 2030 betyder antagelsen om »sund aldring«, at der stort set ikke er nogen effekter på de offentlige sundhedsudgifter af den forøgede levetid, fordi »sund aldring« betyder en udskydelse af terminaludgifterne. I perioden efter 2030 betyder levetidsforøgelsen en gradvis merstigning i sundhedsudgifterne, som i 2050 har ført til en isoleret stigning på 10 pct. point i forhold til 2008.

Den samlede effekt af årgangs- og levetidsforøgelseeffekten indebærer således en gradvis stigning i sundhedsudgifterne, som i 2050 tilsammen udgør 27 pct. i forhold til år 2008.

Mervæksteffekten angiver effekten af, at de aldersbetingede reale sundhedsudgifter vokser hurtigere end produktivitetstigningen i økonomien. Som påpeget ovenfor har der været en markant tendens til stigning i mervæksten i sundhedsudgifterne i Danmark i de senere år. Dette skyldes både en tendens til højere vækst i sundhedsudgifterne, når der korrigeres for den demografiske udvikling og »sund aldring« og en tendens til reduktion i timeproduktiviteten i økonomien. Som udgangspunkt anvendes en årlig mervækstrate på 0,3 pct. point, hvilket stort set svarer til underkantsskønnet for mervæksten i den seneste 15 års periode.⁵ Som det fremgår af afsnittet ovenfor, indebærer denne forudsætning en markant opbremsning i den mervækst, som har været gældende i den seneste 10-års periode. Baseret på de historiske erfaringer er en antagelse om 0,3 pct. point mervækst i de aldersbetingede sundhedsudgifter et meget forsigtigt skøn. Selv med dette forsigtige udgangspunkt indebærer en forsat mervækst en betydelig vækst i sundhedsudgifterne. Frem mod 2050 indebærer mervæksten isoleret en stigning i sundhedsudgifterne på 15 pct. point. Den samlede vækst i sundhedsudgifterne frem mod 2050 bliver således på i alt 42 pct., når der tages højde for såvel mervækst- som levetidsforøgelses- og årgangseffekter. Et alternativ, hvor der anvendes en mervækstrate på 0,6 pct. point for de ikke-demografisk betingede sundhedsudgifter indebærer, at sundhedsudgifterne i 2050 er vokset med 60 pct. når der tages højde for alle tre effekter, jf. figur 6. Med en mervækst på 0,6 pct. point vil det således være mere end halvdelen af den samlede vækst i sundhedsudgifterne (målt i pris- og produktivetskorrigerede enheder) frem til 2050, der kan tilskrives ikke-demografiske effekter, jf. figur 6.

Kombinationen af de demografiske og de ikke-demografiske effekter giver anledning til en meget betydelig stigning i sundhedsudgifterne (målt i pris- og produktivets-

5. Som nævnt er der set bort fra 2007 og 2008 ved beregning af timeproduktiviteten, da disse tal er foreløbige og traditionelt korrigeres ved overgang til endelige tal.



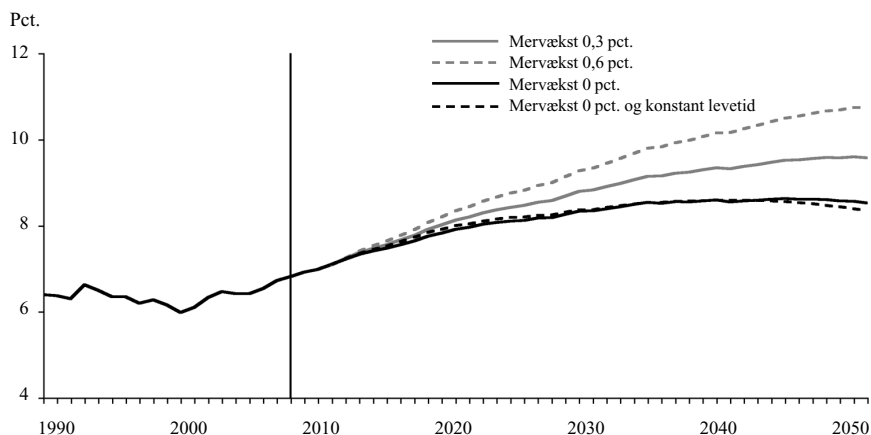
Figur 6. Sundhedsudgifter og mervækst.

Ann.: Figuren viser effekten af mervæksten (givet »sund aldring«) for udviklingen i de offentlige sundhedsudgifter til individuelt konsum. Fast lønniveau (opgjort ved de pris- og produktivitetskorrigerede sundhedsudgifter i DREAM modellen).

Kilde: DREAM modellen og egne beregninger.

korrigerede enheder), selv ved mervækstrater i den ikke-demografisk betingede del af sundhedsudgifterne, som ligger et godt stykke under mervækstraten i disse udgifter gennem de seneste 10 år. Denne udvikling peger isoleret på, at en forlængelse af den historiske tendens vil give anledning til et betydeligt offentligt prioriteringsproblem, hvis den hidtidige offentlige finansiering af sundhedsydelse fastholdes.

Det er imidlertid ikke alene de demografisk betingede sundhedsudgifter, der stiger, når den forventede levetid forøges. Udviklingen i BNP afhænger ligeledes af befolkningsudviklingen i det omfang, at denne fører til en ændring i arbejdsstyrken og dermed beskæftigelsen. Arbejdsstyrken vokser i takt med stigningen i levetiden af to grunde. For det første betyder den generelle tendens til lavere dødelighed sammen med en antagelse om uændret aldersbetinget erhvervsdeltagelse, at antallet af personer i arbejdsstyrken stiger. Hertil kommer, at stigningen i restlevetiden for 60-årige fører til stigninger i efterløns- og folkepensionsalderen fra 2025 og frem som en følge af velfærdsreformen. Disse reguleringer forventes at medføre stigninger i tilbagetrækningsalderen og dermed også stigninger i arbejdsstyrken. Frem til 2025 giver levetidsudviklingen anledning til, at arbejdsstyrken er godt 14.000 personer større i fremskrivningen, end tilfældet ville være under en antagelse om uændret levetid. På grund af reguleringerne i efterløns- og folkepensionsalderen stiger denne forskel meget hurtigere i den efterfølgende periode, således at forskellen i arbejdsstyrken er på knap 31.000 personer i 2030 og hele 145.000 personer i 2050, jf. DREAM (2009).



Figur 7. Offentlige sundhedsudgifter i pct. af BNP

Anm.: Den lodrette streg angiver overgangen til fremskrivning. Graferne angiver de offentlige sundhedsudgifter til individuelt konsum.

Kilde: DREAM modelberegninger.

Den hurtigere vækst i arbejdsstyrken i forløbet med stigninger i levetiden indebærer for konstant strukturel ledighedsprocent, at produktionen målt ved BNP også vokser hurtigere end i alternativet med uændret levetid. Det betyder, at stigningen i sundhedsudgifterne i forløbene med voksende restlevetid i mindre grad giver anledning til et stigende udgiftstryk målt som udgifternes andel af BNP, jf. figur 7.

Udviklingen i de offentlige sundhedsudgifters andel af BNP i alternativet uden mervækst og uden vækst i restlevetiden fra 2011 (sort stiplede kurve) indebærer, at de offentlige sundhedsudgifters andel af BNP vokser fra 6,8 pct. i 2008 til 8,0 pct. i 2020 og til et toppunkt på 8,6 pct. i 2040. I takt med, at de store årganges børn uddør efter 2040 og herefter erstattes af mindre generationer af pensionister, falder sundhedsudgifternes andel af BNP gradvist. Faldet er dog begrænset i forhold til den forudgående stigning. Det afspejler, at ældreandelen i befolkningen var lav igennem 1990'erne, og at dette lave niveau ikke nås i den fremtidige udvikling. Målt i forhold til udgiftstrykket i 2008 er der således en stigning i sundhedsudgifternes andel af BNP på ca. 1½ pct. point fra 2008 til 2050 uden fremtidig vækst i restlevetiden og uden mervækst i de ikke-demografiske sundhedsudgifter.

Udgiftstrykket for de offentlige sundhedsudgifter med stigende restlevetid og en mervækst på 0 pct. pr. person i en given alder (sort kurve) udvikler sig omtrent på samme måde som udgiftstrykket uden vækst i restlevetiden. Med andre ord er velfærdsreformen i kombination med »sund aldring« tilstrækkelig til at sikre, at sundhedsud-

gifterne i en situation uden mervækst ikke vokser med en højere vækstrate end stigningen i arbejdsstyrken som følge af den forlængede levetid frem mod 2050.

I scenariet med 0,3 pct. point mervækst i de ikke-demografisk betingede sundhedsudgifter (grå kurve) vokser de offentlige sundhedsudgifters andel af BNP fra 6,8 pct. i 2008 til 9,6 pct. i 2050. Det svarer til en ekstra vækst i udgiftstrykket på 1,1 pct. point i forhold til situationen uden mervækst i sundhedsudgifterne. Hvis mervæksten i sundhedsudgifterne i de kommende år i højere grad afspejler de seneste års mervækst, er en mervækst på 0,3 pct. point som nævnt langt fra en realistisk vurdering. I scenariet, hvor mervæksten er på 0,6 pct. point pr. år, udgør sundhedsudgifterne 10,8 pct. af BNP i 2050. Det svarer til en ekstra vækst i udgiftstrykket på 2,3 pct. point i forhold til situationen uden mervækst i sundhedsudgifterne. Målt i forhold til 2008 stiger sundhedsudgifternes andel af BNP med 4 pct. point frem til 2050.

Finanspolitisk holdbarhed

For at vurdere de forskellige scenariers betydning for udviklingen i de offentlige sundhedsudgifter og dermed for det langsigtede finansieringsproblem i den offentlige sektor anvendes den dynamiske CGE model DREAM, som netop er udviklet med henblik på at vurdere de langsigtede konsekvenser af finanspolitikken i Danmark.

DREAM er baseret på en overlappende generationsstruktur og har fokus på den demografiske udvikling og den danske offentlige sektor. Overordnet modelleres økonomien i DREAM som en lille åben økonomi med fast valutakurs, perfekt kapitalmobilitet og bopælsbaseret beskatning. Danske og udenlandske varer er imperfekte substitutter således, at det reale bytteforhold er endogent, og indenlandske priser og lønniveau påvirkes af det indenlandske aktivitetsniveau. Modellens arbejdsmarked er karakteriseret ved imperfekt konkurrence i form af fagforeninger, som giver anledning til strukturel arbejdsløshed, jf. DREAM (2008).

Modellens kerne er modelleringen af husholdningerne, som baseres på en detaljeret befolkningsfremskrivning. Den voksne befolkning i den erhvervsaktive alder er opdelt i repræsentative husholdninger for hver årgang. Størrelsen af årgangen fastlægges, så den svarer til befolkningsfremskrivningen. Børn fordeles på de repræsentative husholdninger svarende til befolkningsfremskrivningens aldersspecifikke fertilitetskvo-tienter. Husholdningerne inkluderer den forventede udvikling i husstandens størrelse i den dynamiske fastlæggelse af forbrug, arbejdsudbud og opsparing, som vælges givet en forudsætning om perfekt forudseenhed. Husholdningernes opsparing kan placeres i ejerbolig, finansielle aktiver (aktier og obligationer) og pensionsopsparing i form af såvel arbejdsmarkedspensioner som frivillige private pensionsopsparinger. Centralt for analysen i dette papir er, at det individuelle nominelle offentlige forbrug pr. individ (herunder sundhedsudgifter) tager udgangspunkt i den nuværende aldersfordeling

som fremskrives for hver aldersgruppe (der er underopdelt på køn og etnisk herkomst) med inflation og produktivitetsudvikling. Den fremtidige aldersfordeling af sundhedsudgifterne korrigeres herefter for betydningen af »sund aldring« og en eventuel årlig mervækstrate i sundhedsudgifterne tillægges.

Finanspolitisk holdbarhed defineres som en situation, hvor den tilbagediskonterede værdi af de fremtidige offentlige primære budgetoverskud mindst modsvarer den offentlige sektors nettogæld i udgangsåret. Størrelsen af et givet finanspolitisk holdbarhedsproblem præsenteres som den faste årlige forbedring af det offentlige primære budget i pct. af årets BNP, som skal til for at sikre den finanspolitiske holdbarhed. Størrelsen af holdbarhedsindikatoren inkluderer ikke det forvriddningstab, som ville opstå ved en faktisk finansiering af holdbarhedsproblemet gennem et stigende skatte-tryk, med mindre finansieringen eksplicit er angivet. Dette følger den etablerede fremgangsmåde, som også anvendes af Finansministeriet.

Principielt gennemføres analyser af den finanspolitiske holdbarhed med en uendelig tidshorisont. Det spørgsmål, der ønskes besvaret med beregningen er, om velfærdssystemet er robust over for den forventede udvikling i centrale størrelser som for eksempel stigende relativ efterspørgsel efter offentlige sundhedsydelser i en situation med øget velstand og voksende middellevetid. Fremskrivningen i dette papir følger standarden i DREAM-beregninger og betragter en befolkningsfremskrivning frem til år 2100, hvorefter befolkningen antages konstant – det vil blandt andet sige uden yderligere vækst i middellevetiden. Tilsvarende antages, at eventuelle antagelser om mervækst i den ikke-demografiske del af sundhedsudgifterne fastholdes til år 2100, hvorefter mervækstraten er 0. Dermed vokser både offentlige overførsler og offentligt forbrug med samme rate som den generelle vækst i økonomien efter år 2100. Antagelserne betyder, at økonomien efter år 2100 nærmer sig en tilstand, hvor alle økonomiske størrelser vokser med samme rate – dvs. en steady state. Fortolkningen af beregningerne er derfor, at de vurderer robustheden af velfærdssystemet givet den forventede udvikling i sundhedsudgifterne i de enkelte scenarier frem til år 2100.

Der tages udgangspunkt i DREAMs seneste langsigtede fremskrivning, jf. DREAM (2009). I forhold til denne ændres antagelserne om sundhedsudgifternes udvikling således, at der ses på de fire scenarier, som er beskrevet ovenfor. Således betragtes først et scenario uden vækst i middellevetiden og uden mervækst i den ikke-demografiske del af sundhedsudgifterne. Dernæst betragtes et scenario med vækst i middellevetiden og »sund aldring«, men uden mervækst i den ikke-demografiske del af sundhedsudgifterne og endelig to scenarier med vækst i middellevetiden, »sund aldring« og en mervækstrate på henholdsvis 0,3 pct. point og 0,6 pct. point i den ikke-demografiske del af sundhedsudgifterne.

Fremskrivningen af sundhedsudgifterne med konstant levetid og uden en mervækst i sundhedsudgifterne pr. person i en given alder indebærer et finanspolitisk holdbarhedsproblem på 1,0 pct. af BNP. For at finansiere de fremtidige offentlige udgifter uden, at det senere bliver nødvendigt at øge beskatningen, er det således nødvendigt at opnå en permanent forbedring af de offentlige nettoindtægter på 1,0 pct. af BNP. I 2009 svarer det til en nettoforbedring på ca. 18 mia. kr. I dette scenario svarer holdbarhedsproblemet stort set til den permanente stigning i sundhedsudgifternes andel af BNP.

Med vækst i middellevetiden, »sund aldring« og ingen mervækst i sundhedsudgifterne pr. person i en given alder bliver det finanspolitiske holdbarhedsproblem 0,9 pct. af BNP (svarende til ca. 16 mia. kr. i 2009 niveau). Også i dette tilfælde gælder, at holdbarhedsproblemet stort set modsvarer stigningen i sundhedsudgifternes andel af BNP. Resultatet er bemærkelsesværdigt, fordi det betyder, at velfærdsreformen i kombination med »sund aldring« netop er tilstrækkelig til at gøre velfærdssystemet robust over for ændringer i middellevetiden. Velfærdskommissionen præsenterede netop robustheden over for fremtidige stigninger i levetiden som en af de helt centrale demografiske udfordringer for velfærdsamfundet, jf. Velfærdskommissionen (2006). Selvom velfærdsreformen ikke løser det fulde holdbarhedsproblem, viser beregningerne således, at givet de forventede effekter af velfærdsreformen på arbejdsudbuddet og de estimerede effekter af »sund aldrings« begrænsning i væksten i sundhedsudgifterne, er denne centrale robusthed af velfærdsordningerne opnået med velfærdsreformen.

Velfærdskommissionen pegede endvidere på, at velstandsdilemmaet i form af ønsket om en mervækst i primært offentlige sundhedsudgifter i kombination med et ønske om lavere arbejdstid over livet kunne udgøre et langt alvorligere problem for robustheden af velfærdssamfundet, jf. Andersen og Pedersen (2005). Med de seneste års betydelige mervækst i sundhedsudgifterne er denne del af velstandsdilemmaet forstærket langt mere end forventet.

En mervækst i den ikke-demografiske del af sundhedsudgifterne på 0,3 pct. point pr. år fører til, at holdbarhedsproblemet forøges således, at det permanente finansieringsbehov bliver på 3,0 pct. af BNP svarende til ca. 54 mia. kr. i år 2009, mens en mervækst på 0,6 pct. point giver anledning til et holdbarhedsproblem på ikke mindre end 5,7 pct. af BNP, svarende til et permanent finansieringsbehov på ca. 102 mia. kr. i år 2009. Der er således i begge tilfælde tale om meget betydelige holdbarhedsproblemer. Den isolerede effekt på den finanspolitiske holdbarhed af en mervækst på 0,3 pct. point er en forværring på 2,1 pct. af BNP. Øges mervæksten med yderligere 0,3 pct. point opnås en ekstra forværring af holdbarheden på 2,7 pct. point. At forværringen af holdbarheden således ikke er lineær i ændringen af mervæksten skal tilskrives en mekanisk effekt, der følger af, at stigningen i den tilbagediskonterede værdi af sundhedsudgifterne

Tabel 1. Hovedresultater af forskelle i antagelser om mervækstrate og levetid.

	Ingen stigning i midlelevetid og 0 pct. mervækst i sundhedsudgifter	Stigning i middel- levetid og 0 pct. mervækst i sund- hedsudgifter	Stigning i middel- levetid og 0,3 pct. mervækst i sund- hedsudgifter	Stigning i middel- levetid og 0,6 pct. mervækst i sund- hedsudgifter
Realvækst i sundhedsudgifter fra 2008 til 2050 ^(a)	17,0 pct.	27,0 pct.	42,0 pct.	60,0 pct.
Sundhedsudgifters andel af BNP i 2050	8,3 pct.	8,5 pct.	9,6 pct.	10,8 pct.
Finanspolitisk holdbarheds- problem i pct. af BNP	1,0 pct.	0,9 pct.	3,0 pct.	5,7 pct.

Note: ^(a) Stigningen måles som stigningen i fast lønniveau, dvs. opgjort ved pris- og produktivitetskorrigerede sundhedsudgifter.

som følge af en årlig mervækst ikke er lineær i mervækstbidraget. Det er imidlertid den tilbagediskonterede værdi af de fremtidige sundhedsudgifter, der er afgørende for holdbarhedseffekten, jf. Appendiks for en uddybning.

Den samlede betydning af de forskellige alternativer vedr. vækst i restlevetid og mervækstrate i sundhedsudgifterne pr. person fremgår af tabel 1.

De meget markante stigninger i det finanspolitiske holdbarhedsproblem, som følger af disse begrænsede mervækstrater viser, at de seneste 10 års udvikling i sundhedsudgifterne – med mervækstrater på 1½ – 2 pct. point – er fuldstændig uholdbar, hvis vækstraterne fastholdes over bare lidt længere perioder. I fravær af reduktioner i andre offentlige udgifter eller betydelige skattestigninger er det ikke muligt over en længere periode at fastholde selv en mervækstrate på niveau med gennemsnittet for de seneste 15 år, der ligger på mellem 0,4 og 0,9 pct. Dette vil jf. beregningerne ovenfor føre til så store behov for forbedringer af de offentlige finanser, at dette næppe kan løses ved reformer, der øger beskæftigelsesgraden og dermed skattegrundlaget tilstrækkeligt til at sikre finansieringen uden skattestigninger og/eller udgiftsreduktioner andre steder.

Det bemærkes, at beregningen af holdbarhedsproblemet, hvor der beregningsmæssigt opgøres et permanent nettofinansieringsbehov fra »start til slut«, skal betragtes som en illustration af, hvor betydningsfuld selv en mindre, men længerevarende merstigning i sundhedsudgifterne er for den langsigtede finansiering – og ikke som en anbefaling om en makroøkonomisk strategi: En merstigning i sundhedsudgifterne på 0,3 pct. point om året kræver en varig reduktion af de offentlige nettoudgifter på (3,0-0,9 =)

2,1 pct. af BNP (svarende til ca. 38 mia. kr. i 2009 niveau) for at sikre finansieringen af de ekstra sundhedsudgifter, der følger af mervæksten. Hvis man gennemfører en politik, som indebærer, at de offentlige nettoudgifter permanent reduceres med 2,1 pct. af BNP, vil det føre til en markant intergenerational omfordeling til ugunst for nuværende generationer og til fordel for kommende generationer. Det skyldes naturligvis, at mervæksten fører til en gradvis stigning i sundhedsudgifterne, mens finansieringen kommer som en permanent ændring i de offentlige nettoindtægter.

Som eksempel på en holdbar finanspolitik er der gennemført en beregning, der finansierer den samlede stigning i sundhedsudgifterne med en permanent stigning i bundskattesatsen fra 2011. Den nødvendige permanente stigning i bundskattesatsen er på 7 pct. point. Finansieringen dækker således både den manglende finansiering uden mervækst og den stigning, der følger af en begrænset mervækst på 0,3 pct. point i sundhedsudgifterne pr. person i en given alder.

Imidlertid får de første generationer kun i begrænset omfang stigende sundhedsudgifter pr. person, fordi stigningen kommer gradvist. Forskellen mellem stigningen i de offentlige indtægter som følge af den højere skattebetaling og stigningen i sundhedsudgifterne fører til en gradvis stigning i den offentlige formue. Først i 2085 er sundhedsudgifterne steget med 2,1 pct. point af BNP, og merudgiften til sundhed er dermed lige så høj som merindtægten fra skattestigningen. I perioden op til 2085 er der opbygget en ekstra offentlig formue, og renteafkastet af denne anvendes til at finansiere de varige, ekstra sundhedsudgifter, der følger af den fortsatte merstigning i sundhedsudgifterne fra 2085 og frem til år 2100, samt en permanent fastholdelse af dette niveau. Det er næppe hverken hensigtsmæssigt eller realistisk at basere den offentlige udgiftsplanlægning på et sådant opsparingsforløb.

Sammenfatning

Sundhedsudgifternes ikke-demografiske del er vokset meget markant i de senere år, hvilket til dels kan forklares ved en øget politisk opmærksomhed på området og en overgang til en mere efterspørgselsstyret udvikling i sundhedssektoren gennem indførelse af det frie sygehusvalg og det udvidede frie sygehusudvalg. Udviklingen har ført til mervækstrater i den ikke-demografiske del af sundhedsudgifterne på 1,5 – 2,3 pct. point pr. år i gennemsnit over den seneste 10-års periode. Mervækstrater i dette omfang kan ikke fastholdes i en længere periode uden at føre til omfattende finansieringsproblemer for den offentlige sektor. Selv hvis den varige mervækstrate reduceres til 0,6 pct. point pr. år vil denne mervækst isoleret bidrage til et holdbarhedsproblem i størrelsesordenen omkring 4,8 pct. af BNP, hvilket er betydeligt mere end, hvad der kan forventes at kunne opnås gennem strukturreformer, som udvider arbejdsudbuddet.

Der er derfor et markant politisk prioriteringsproblem, som indebærer enten en betydelig opbremsning af udgiftsudviklingen på sundhedsområdet, besparelser på andre offentlige udgiftsområder, skattestigninger eller indførelse af alternative finansieringsmuligheder for (visse af) sundhedsydelserne. Det er naturligvis også muligt at anvende flere af disse elementer i sammenhæng med strukturreformer på arbejdsmarkedet.

Det er dog et lyspunkt, at resultaterne i denne analyse peger på, at de markante prioriteringsproblemer alene vedrører spørgsmålet om mervæksten i de ikke-demografisk betingede sundhedsudgifter, mens velfærdssamfundets finansiering er blevet robust overfor effekten af den forventede stigende restlevetid. Årsagen til dette er en kombination af på den ene side den forventede sammenhæng mellem levetidsstigninger og stigninger i arbejdsstyrken (og reduktion i antallet af pensionister) som følger af velfærdsreformen fra 2006 og på den anden side effekten af »sund aldring«, som reducerer væksten i sundhedsudgifterne ved en stigende levetid.

Litteratur

- Andersen, T. M. og L. H. Pedersen. 2005. Demografi, velstandsdilemma og makroøkonomiske strategier, *Nationaløkonomisk Tidsskrift* bd. 143, nr. 2, s. 189-229.
- Arnberg, S. og T. B. Bjørner. 2009. Estimation af sundhedsudgifternes afhængighed af alder og afstand til død, Working paper, De Økonomiske Råds Sekretariat.
- Breyer, F. og S. Felder. 2004. Life Expectancy and Health Care Expenditures in the 21st Century: A New Calculation for Germany Using the Costs of Dying, *Discussion paper* 452, German institute for Economic Research, Berlin.
- Cutler, D. M. og L. Sheiner. 1998. Demographics and Medical Care Spending: Standard and Non-standard Effects, *NBER Working paper* no. 6866.
- De Økonomiske Råd. 2009. *Dansk Økonomi efterår 2009*. København.
- DREAM. 2008. Dokumentation (foreløbig version), kapitel 1-10, www.dreammodel.dk
- DREAM. 2009. Langsigtet økonomisk fremskrivning 2009.
- Fuchs, V. 1984. Though Much Is Taken: Reflections on Ageing, Health, and Medical Care, *Milbank Quarterly* 62: 143-66.
- Fuchs, V. 1990. The Health Sector's share of Gross National Product, *Science* vol. 247, 534-38.
- Gornick, M., A. McMillan, og J. Lublitz. 1993. Data Watch: A Logitudinal Perspective On Patterns Of Medicare Payments, *Health Affairs* 141-150.
- Hagist, C. og L. J. Kotlikoff. 2005. Who's Going Broke? Comparing Healthcare Cost in Ten OECD Countries, *NBER Working Paper* no. 11833.
- Hall, R. E. og C. Jones. 2007. The Value of Life and the Rise in Health Spending, *Quarterly Journal of Economics*, vol. 122, s. 39-72.
- Lee, R. og T. Miller. 2002. An Approach to Forecasting Health Expenditures, with Application to the U.S. Medicare System, *Health Services Research* vol. 37:5, s. 1365-85.
- Lublitz, J. D. og G. F. Riley. 1993. Trends in Medicare Payments in the Last Year of Life, *New England Journal of Medicine*, 1092-1096.
- Seshamani, M. og A. M. Gray. 2004a. A longitudinal study of the effects of age and time to death on hospital costs, *Journal of Health Economics* vol. 23. 217-235.

Seshamani, M. og A. Gray. 2004b. Time to death and health expenditure: an improved model for the impact of demographic change on health care costs, *Age and Ageing* vol. 33 no. 6, 556-561.

Pedersen, L. H. og M. F. Hansen. 2006. Sti-

gende sundhedsudgifter, et aldrings- eller et velstandsfænomen?, *Nationaløkonomisk Tidsskrift* bd. 144, s. 304-325.

Velfærdskommissionen. 2006. *Fremtidens velfærd – vores valg*. København.

Appendiks

Ikke-lineariteten mellem det finanspolitiske holdbarhedsproblem og mervækstraten i sundhedsudgifterne kan indses ved følgende ræsonnement. Lad sundhedsudgifterne i hver periode være 1 i udgangspunktet og lad g angive mervæksten. I dette tilfælde kan den med raten r tilbagediskonterede værdi af ændringen i de samlede udgifter dE over en uendelig tidshorisont, udtrykkes som:

$$dE = \sum_{t=0}^{\infty} ((1+g)^t - 1) \left(\frac{1}{1+r}\right)^t \quad (4)$$

hvilket kan omskrives til

$$dE = (1+r) \left(\frac{1}{r-g} - \frac{1}{r} \right) \quad (5)$$

hvor det antages, at den betragtede mervækst er begrænset således at $0 < g < r$.

Det fremgår, at hvis mervækstbidraget, g går imod tilbagediskonteringsraten, r vil den tilbagediskonterede værdi af ændringen i sundhedsudgifterne gå mod uendelig. Det indses let, at den første afledte af dE mht. mervæksten g er positiv med den givne parameterrestriktion. Dette indikerer blot, at ændringen i den tilbagediskonterede værdi af sundhedsudgifterne som forventeligt stiger, når mervækstbidraget forøges. Den andenafledte af dE mht. g er positiv givet parameterrestriktionen og viser dermed, at stigningen i den tilbagediskonterede ændring i sundhedsudgifterne vil være voksende i mervækstbidraget, hvilket netop fandtes i fremskrivningerne.

$$dE_g'' = (1+r) \left(\frac{2}{(r-g)^3} \right) \quad (6)$$

Sundhedsudgifter og levetid

Søren Arnberg

De Økonomiske Råds Sekretariat, E-mail: sar@dors.dk

Thomas Bue Bjørner

De Økonomiske Råds Sekretariat, E-mail: tbb@dors.dk

SUMMARY: An empirical analysis is presented separating the individual public health expenditures of the Danish population into the individual effects of age and proximity to death (reflecting terminal costs of dying). The analysis is based on micro data from the period 2000 to 2007, covering a random sample of 10 percent of the Danish population. Health expenditures include treatment in hospitals, subsidies to prescribed medication and health care provided by general practitioners and specialists. The results show that proximity to death has a significant impact on the health costs of the individual. This has important implications for the future health expenditures. When life expectancy increases, the terminal costs are postponed and the increases in health expenditure that follow from longer life expectancy are not as large as the increase in the number of elderly persons would suggest.

1. Indledning

I de kommende årtier vil der blive flere ældre i befolkningen, og dette ventes at lede til et øget pres på en række offentlige serviceydelser inden for sundhed og ældrepleje, hvor udgifterne pr. person er højere for ældre. Forventet vækst i sundhedsudgifterne er således blevet fremhævet som en af de største udfordringer for den finanspolitiske holdbarhed, jf. DREAM (2009).

En række nyere bidrag fra den sundhedsøkonomiske litteratur peger imidlertid på, at det demografiske pres på sundhedsudgifterne ikke bliver så markant, som stigningen i antallet af ældre umiddelbart tilsiger. Dette skyldes, at en stor del af sundhedsudgifterne afholdes i de sidste leveår. Når levetiden forlænges, vil disse såkaldte »terminaludgifter« også blive udskudt. Forlængelser af levetiden kan således give sig udslag

Resultaterne i artiklen er anvendt i forbindelse med fremskrivninger af de offentlige sundhedsudgifter i sundhedskapitlet i De Økonomiske Råd (2009). I den forbindelse vil forfatterne gerne takke formandskab og sekretariat for sparring i forbindelse med de præsenterede analyser. Endvidere har Mette Møller Jørgensen ydet et stort bidrag i forbindelse med konstruktion af data. Marianne Frank Hansen fra DREAM gruppen har bistået med konsekvensberegningerne i afsnit 6. Eventuelle fejl kan dog alene tilskrives forfatterne.

i flere raske leveår uden større behov for sundhedsydelse. Dette betegnes som »sund« aldring. Alternativt kan forlængelser af levetiden lede til flere behandlingskrævende leveår med høje sundhedsudgifter. Dette betegnes ofte som »usund« aldring. Det vil have stor betydning for udviklingen i de fremtidige sundhedsudgifter, hvorvidt forlænget levetid leder til sund eller usund aldring.

I denne artikel præsenteres en mikroøkonometrisk analyse af betydningen af alder og andre demografiske effekter for de offentlige sundhedsudgifter. Formålet er at bestemme størrelsen af terminaludgifterne i Danmark og deres betydning for de fremtidige offentlige sundhedsudgifter, når levetiden forlænges.

I analyserne fokuseres på betydningen af demografiske faktorer for sundhedsudgifterne. Det er i andre sammenhænge dokumenteret, at en væsentlig del af den hidtidige historiske vækst i sundhedsudgifterne skyldes ikke-demografiske faktorer, som f.eks. en høj indkomstelasticitet for sundhedsydelse og fremkomsten af nye behandlinger, som gør det muligt at behandle flere sygdomme, jf. f.eks. Pedersen og Hansen (2006), OECD (2006) og Fogel (2008). I de kommende årtier vil forlængelser af levetiden og de store efterkrigsårgange imidlertid give en væsentlig stigning i andelen af ældre. Det vil øge betydningen af demografiske effekter på de fremtidige offentlige sundhedsudgifter.

Vores analyser viser, at terminaludgifterne udgør en væsentlig del af de offentlige sundhedsudgifter, og at dette giver sig udslag i faldende gennemsnitlige sundhedsudgifter for ældre, når levetiden forlænges. Dette svarer til sund aldring. De faldende aldersfordelte sundhedsudgifter har stor betydning for udviklingen i de samlede sundhedsudgifter. Når der tages højde for sund aldring, halveres stigningen i de reale sundhedsudgifter som følge af forlængelser af levetiden sammenlignet med en beregning, hvor der ikke tages højde for sund aldring.

I det følgende afsnit præsenteres litteraturen om sammenhængen mellem alder, forlænget levetid og sundhedsudgifter. I afsnit 3 beskrives de anvendte data. I afsnit 4 præsenteres den anvendte økonometriske model. Estimationsresultater beskrives i afsnit 5, og på baggrund heraf opgøres betydningen af sund aldring for de fremtidige offentlige sundhedsudgifter i afsnit 6. I afsnit 7 diskuteres de potentielle endogenitetsproblemer i forbindelse med at medtage afstand til død som en eksogen variabel til forklaring af sundhedsudgifterne.

2. Aldring og udgifter til sundhed

Sundhedsudgifterne for ældre er væsentligt højere end for yngre, og længere levetid betyder, at der kommer flere ældre. Umiddelbart leder dette til en forventning om, at der kommer en drastisk stigning i sundhedsudgifterne, når levetiden forlænges. En stor

del af sundhedsudgifterne har imidlertid karakter af terminaludgifter, som afholdes i de sidste leveår, jf. f.eks. Felder m.fl. (2000) og Hoover m.fl. (2002). Den observerede korrelation mellem alder og sundhedsudgifter er således ikke (nødvendigvis) en egentlig alders effekt, men kan skyldes, at dødeligheden stiger med alderen. Tilstedeværelsen af terminaludgifter er vigtige for udviklingen i sundhedsudgifterne i forbindelse med forlængelser af levetiden. Terminaludgifterne har karakter af en engangsudgift, som udskydes, når levetiden forlænges. Det er derfor vigtigt, at adskille alders effekten fra terminaleffekten i forbindelse med fremskrivninger af sundhedsudgifterne ved ændringer i levetiden.

Et af de første mere formaliserede forsøg på at adskille effekterne af alder og terminaludgifter blev præsenteret i Zweifel m.fl. (1999). For at adskille alders effekter fra terminaludgifter medtager Zweifel m.fl. variabler for senere afdødes afstand til død i estimation af individers sundhedsudgifter. Når variabler for afstand til død medtages finder Zweifel m.fl., at aldersvariablen bliver insignifikant. På den baggrund konkluderer Zweifel m.fl., at alder overhovedet ikke har nogen direkte effekt på sundhedsudgifterne (ud over den indirekte effekt af alder som skyldes, at dødeligheden stiger med alderen). Sammenhængen mellem sundhedsudgifter og alder blev betegnet som en »red herring«, dvs. noget der ikke findes. I litteraturen refereres til red-herring-hypotesen som hypotesen om, at alder ikke har betydning for den enkeltes sundhedsudgifter efter, at der er taget højde for afstand til døden. Der har været kritik af den konkrete økonometriske specifikation anvendt i Zweifel m.fl. (1999), fordi denne kan under-vurdere den eventuelle alders effekt, jf. Salas og Raftery (2001). Senere bidrag tyder da også på, at selvom afstand til død er en væsentlig faktor i beskrivelsen af sundhedsudgifterne, så spiller alder også en rolle i sig selv, jf. f.eks. Häkkinen m.fl. (2008), Zweifel m.fl. (2004), Seshamani og Gray (2004a) og Seshamani og Gray (2004b).

Der har også været danske undersøgelser på området, men disse tager kun højde for afstand til død i et enkelt år inden dødstidspunktet, jf. Madsen m.fl. (2002) og Serup-Hansen m.fl. (2002). Hvis afstand til død har betydning i flere år inden dødstidspunktet vil det give en forkert vurdering af de samlede terminaludgifter og rene alders effekter. I DREAM's fremskrivning af de offentlige udgifter og holdbarhedsproblemet tages ikke højde for muligheden for sund aldring, jf. DREAM (2009). Finansministeriet tager højde for afstand til død i 3 år før dødstidspunktet, men synes ikke at have udført en egentlig økonometrisk analyse for at adskille effekten af alder og afstand til død, jf. Finansministeriet (2007).

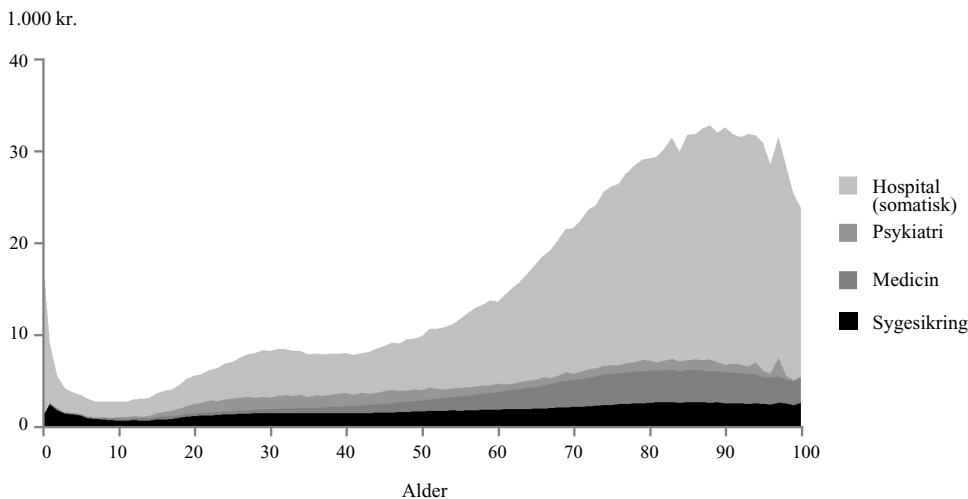
Et grundlæggende problem er spørgsmålet om kausaliteten mellem sundhedsudgifter og alder. I analyser baseret på ovenstående tilgang betragtes sundhedsudgifterne som en funktion af afstand til død. Behandling i sundhedsvæsenet har imidlertid til

formål at forbedre helbredstilstanden og forlænge livet. I andre dele af den sundhedsøkonomiske litteratur betragtes sundhedsudgifter som et instrument til at opnå forlængelser af levetiden i stedet for noget, som er en funktion af (eksogene) stigninger i levetiden, jf. f.eks. Hall og Jones (2007). Afstand til død kan således være en endogen variabel i forhold til sundhedsudgifterne. Der er ikke taget højde for dette endogenitetsproblem i de ovennævnte empiriske undersøgelser. Tilsyneladende er Felder m.fl. (2008) den eneste undersøgelse af terminaleffekter, som eksplicit søger at tage højde for endogenitet. Dette gøres ved at instrumentere afstand til død ud fra en model for dødeligheden fra en tidligere periode. Selv om endogenitet af afstand til død ikke formelt kan afvises finder Felder m.fl., at størrelsen af endogenitetsbias er beskednen og ikke ændrer ved, at man bør tage højde for terminaludgifter. I vores analyser baseret på danske data foretages ikke et egentlig test for endogenitet, men en form for robusthedscheck i afsnit 7 peger i retning af, at potentiel endogenitet ikke har afgørende indflydelse på hovedkonklusionen.

3. Data

De empiriske analyser er baseret på en database, som kobler oplysninger om udgifterne ved personers brug af offentlige sundhedsydelser for perioden 2000 til 2007. Databasen er baseret på oplysninger fra en række registre: *Landspatientregisteret* indeholder oplysninger om indlæggelser og ambulante behandlinger på somatiske (dvs. ikke psykiatriske) hospitalsafdelinger. Omkostningen ved hospitalsbehandlinger er beregnet ud fra de såkaldte DRG og DAGS takster, som er mål for driftsomkostningen ved forskellige typer af henholdsvis indlæggelser og ambulante behandlinger. *Det psykiatriske centrale forskningsregister* indeholder oplysninger om behandling på psykiatriske hospitalsafdelinger bl.a. antal indlæggelsesdage og antal ambulante psykiatriske behandlinger. For at beregne udgiften anvendes en standard sengedagstakst og en standard takst for ambulante psykiatriske behandling.¹ *Lægemiddeldatabasen* indeholder oplysninger om salgspriser og offentligt medicintilskud for receptpligtig medicin. I de præsenterede analyser indgår kun det offentlige medicintilskud (ikke patienters egenbetaling). *Sygesikringsstatistikregisteret* indeholder oplysninger om udgiften ved besøg hos læge, speciallæge og andre ambulante sygesikringsydelser. Endelig er der fra *dødsårsagsregisteret* indhentet oplysninger om dødstidspunkt (indtil primo 2009).

1. Beregningen af udgiften til almindelig (somatisk) og psykiatriske hospitalsbehandling er således beregnet ud fra gennemsnitstakster for typiske behandlingsforløb. Disse gennemsnitstakster tager ikke højde for, at der kan være systematiske aldersbetingede forskelle i behandlingsforløbene. Hvis en given behandling systematisk er dyrere for ældre end for yngre vil det lede til en endnu større forskel i sundhedsudgifterne mellem unge og gamle end det, der kommer til udtryk i artiklen.



Figur 1. Gennemsnitlige offentlige sundhedsudgifter pr. person i 2000.

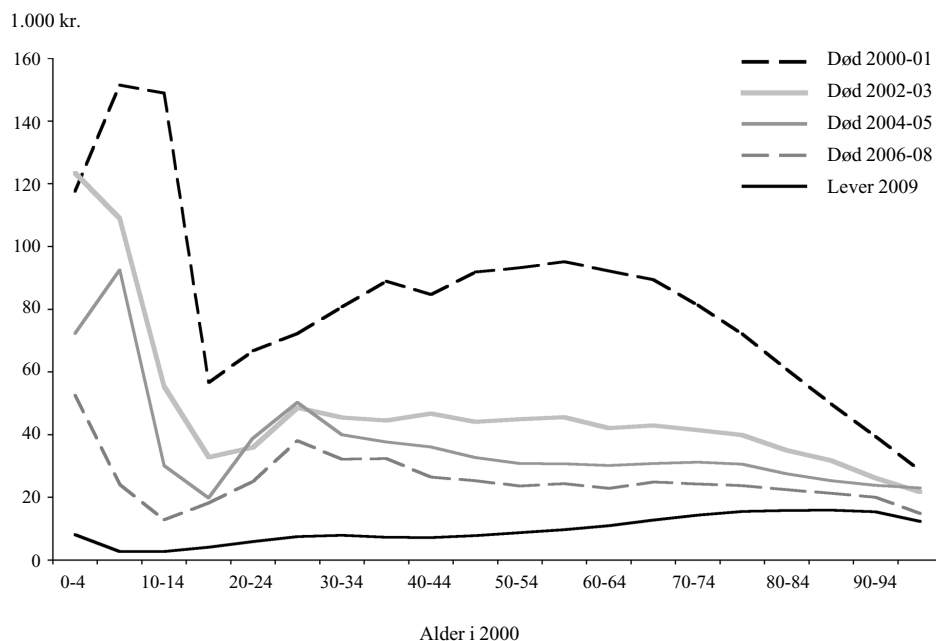
Anm: 2007 priser.

Databasen indeholder individuelt fordelte sundhedsudgifter svarende til knap 80 pct. af det samlede offentlige forbrug på individuelle sundhedsydelser.² De manglende godt 20 pct. skyldes især, at de individuelt fordelte hospitalsudgifter er mindre end de samlede offentlige hospitalsudgifter, bl.a. fordi DRG og DAGS taksterne kun medregner driftsomkostninger, og således ikke medtager udgifter til hospitalsudstyr og bygninger mv. Endvidere har der ikke været tilgængelige registeroplysninger for en række sundhedsydelser som f.eks. kommunal børnetandpleje.

De gennemsnitlige årlige sundhedsudgifter pr. person varierer betydeligt over livsforløbet, jf. figur 1. Således er sundhedsudgifterne kraftigt stigende fra ca. 60-års alderen frem til omkring 85-års alderen. For yngre er sundhedsudgifterne relativt høje omkring 30-års alderen. Dette skyldes høje sundhedsudgifter for kvinder i de år, hvor de føder børn.

Sundhedsudgifterne i de sidste leveår (terminaludgifterne) er særligt høje. Dette er illustreret i figur 2, som viser de gennemsnitlige aldersfordelte sundhedsudgifter i 2000 opdelt for personer med forskellig afstand til død. Den øverste, stiplede kurve angiver således de aldersfordelte sundhedsudgifter for personer, der dør i samme år eller året efter (dvs. i 2000 og 2001). Den lysegrå kurve er udgiften for personer, der dør

2. I Danmarks Statistiks opgørelse af sundhedsudgifter indgår også udgifter til plejehjem og dele af de øvrige ældreplejeudgifter. Der har ikke været registeroplysninger til rådighed om disse ældreplejeudgifter. Databasen indeholder individuelt fordelte sundhedsudgifter svarende til knap 80 pct. af de samlede sundhedsudgifter, når man ikke medregner de sundhedsudgifter, som har karakter af ældrepleje.



Figur 2. Gennemsnitlige sundhedsudgifter i 2000 for personer med forskellig afstand til død.

Anm: Gennemsnitlige offentlige sundhedsudgifter pr. person i 2000 i 2007 priser. Angivet for 5-års aldersintervaller.

2-3 år efter (i 2002 og 2003) osv. Den nederste, mørke kurve er sundhedsudgiften i 2000 for personer, som endnu levede primo 2009.

Det fremgår, at sundhedsudgifterne i 2000 er betydeligt højere for dem, der er tæt på at dø sammenlignet med personer, som ikke var døde primo 2009. Den nederste, mørke kurve i figur 2 viser også, at alderen i sig selv har relativt begrænset indflydelse på sundhedsudgifterne for personer, som har mange leveår tilbage. De med alderen stigende sundhedsudgifter i figur 1 er således et udtryk for en sammenblanding af alderseffekter og terminaleffekter, idet dødeligheden stiger kraftigt med alderen.

Terminaludgiften begynder at aftage fra 60-70 års alderen, jf. figur 2. Aftagende terminaludgifter for de ældste er også fundet i amerikanske og europæiske undersøgelser, jf. f.eks. Cutler og Sheiner (1998), Cutler og Meara (1999) og OECD (2006). I modsætning hertil er terminaludgifterne for børn meget høje. Der kan være forskellige forklaringer på de aftagende terminaludgifter for de ældste. En forklaring er, at det kan være vanskeligt at behandle ældre svækkede mennesker med succes, eller at det er andre sygdomme, som rammer de ældre. En anden mulig forklaring er, at de ældre »sover stille ind« uden at blive behandlet i nævneværdig grad. Endelig kan de faldende

terminaludgifter være udtryk for, at sundhedsvæsenet prioriterer unges liv højt i forhold til de ældstes liv.³

Den anvendte database dækker hele befolkningen i perioden 2000 til 2007. I estimationerne præsenteret i afsnit 5 er anvendt et udtræk på 10 pct. af befolkningen for perioden 2000 til 2003. Årene 2004 til 2007 er ikke medtaget, fordi vi gerne vil have mulighed for, at identificere effekten af afstand til død i en årrække inden dødstidspunktet. I afsnit 7 præsenteres supplerende estimationer baseret på data for hele perioden 2000 til 2007. Deflatoren for offentligt forbrug fra ADAM's databank er anvendt til at beregne sundhedsudgifterne i faste priser.

4. Empirisk model

Lad ε_{it} angive de offentlige sundhedsudgifter i år t for person i . Der ønskes en model, der bestemmer ε_{it} som en funktion af observerede demografiske karakteristika.⁴ Omkring 6 pct. af den danske befolkning har ingen offentlige sundhedsudgifter i et givet år. Det kan give en alvorlig bias at udelade disse personer med relativt godt helbred. Derfor estimeres en såkaldt two-part model, hvor første del er en probit-model, som angiver sandsynligheden for at have positive sundhedsudgifter, mens anden del bestemmer ε_{it} givet $\varepsilon_{it} > 0$.

Variablen ε_{it} er meget højreskæv, da der er store sundhedsudgifter for personer, som rammes af alvorlig sygdom. En højreskæv afhængig variabel håndteres ofte ved at foretage en logtransformation og lave en OLS-regression på den transformerede variabel. Nedenfor estimeres anden del af two part modellen dog som en generaliseret lineær model (GLM) med en loglink relation, jf. f.eks. Manning og Mullahy (2001) og Seshamani og Gray (2004a). Two-part modellen kan opskrives således (hvor $E[\bullet]$ er forventningsoperatoren, X_{it} er en vektor af forklarende variable og Φ er standard normalfordelingsfunktionen (kumuleret)):

$$P(\varepsilon_{it} > 0) = P(X'_{it} \tilde{\gamma}_{it} + v_{it}) = \Phi(X'_{it} \tilde{\gamma}_{it}) \quad (1)$$

$$\ln[E[\varepsilon_{it} | X_{it}; \varepsilon_{it} > 0]] = X_{it}' \hat{\beta}, \quad \varepsilon_{it} \sim Poisson \quad (2)$$

3. Selv om terminaludgifterne til mindre børn er høje pr. barn, så er de samlede terminaludgifter til børn små i forhold til de samlede sundhedsudgifter, fordi dødeligheden for børn er lille. Terminaludgifterne har således størst samlet betydning for ældre. I 2000 var det i alt 14 pct. af de samlede sundhedsudgifter, som blev anvendt til personer, der døde samme år eller året efter og 34 pct. til personer, der døde inden for ni år. Ses alene på sundhedsudgifterne for de 65+ årige, så er de tilsvarende tal 28 pct. og 68 pct.

4. I den empiriske model medtages kun demografiske variabler, som også er tilgængelige i DREAM's befolkningsfremskrivning (alder, køn og variabler relateret til forventet restlevetid). Dette er gjort for at kunne fremskrive de fremtidige sundhedsudgifter ved demografiske ændringer. Socioøkonomiske variabler, som f.eks. uddannelse og ægteskabelig status er derfor ikke medtaget i modellen.

GLM-modellen med log-link relation er et alternativ til OLS-regression af den log-transformerede model. Denne model giver konsistente estimater af de forventede niveauer for sundhedsudgiften selv med heteroscedastiske residualer. Omvendt kan OLS-regression af den log-transformerede model give problemer i forbindelse med prediktion af sundhedsudgifterne i oprindelige niveauer. Det er velkendt, at det ikke er nok at retransformere ved at tage den inverse logaritme, idet $\exp[E[\ln(\varepsilon_{it}) | X_{it}; \varepsilon_{it} > 0]] \neq E[\varepsilon_{it} | X_{it}; \varepsilon_{it} > 0]$. Den gængse metode til at korrigere for dette er at multiplicere med et korrektionsled: $\exp[X_{it}' \hat{\beta}] \times E[\exp(u_{it})]$, hvor $E[\exp(u_{it})]$ er beregnet som gennemsnittet af residualerne i estimationssamplet, jf. Duan (1983). Denne metode forudsætter imidlertid homoscedastiske residualer. Hvis residualerne i stedet er heteroscedastiske, vil OLS-regression resultere i skæve estimater af de forventede niveauer for sundhedsudgiften, jf. Manning og Mullahy (2001). Et test for homoscedasticitet ved estimation af en logtransformeret model afviser kraftigt nulhypotesen (om homoscedasticitet).⁵

Two-part modellen er også anvendt i andre nylige bidrag af terminaludgifternes betydning for sundhedsudgifterne, jf. f.eks. Häkkinen m.fl. (2008) og Seshamani og Gray (2004a). Nogle af de første bidrag på området (f.eks. Zweifel m.fl. (1999)) anvendte en Heckman selektionsmodel, hvor der tages højde for potentiel korrelation mellem restledene fra ligning (1) og ligning (2) ved i ligning (2) at medtage Mills ratios beregnet efter estimation af ligning (1). Denne metode er efterfølgende blevet kritiseret, fordi det i praksis er vanskeligt at finde information, som forklarer, at man har positive sundhedsudgifter, og som ikke samtidig forklarer sundhedsudgifternes størrelse. I dette tilfælde vil inklusion af Mills ratios i estimationen af ε_{it} (eller $\ln(\varepsilon_{it})$) i andet trin medføre meget høj multikolaritet, som f.eks. kan medføre at alder bliver insignifikant, jf. f.eks. Salas og Raftery (2001). I two-part modellen antages, at restledene i ligning (1) og (2) er uafhængige, således at de kan estimeres hver for sig, hvilket også letter prediktion i modellen jf. nedenfor.⁶

Lidt forenklet kan (2) specificeres som følger, hvor fodtegn for i og t er undertrykt på højresiden :

$$\begin{aligned} \ln[E[\varepsilon_{it} | X_{it}; \varepsilon_{it} > 0]] = \\ \beta_0 + \beta_1 A + \beta_2 A^2 + \beta_3 A^3 + \beta_4 S + \beta_5 S \cdot A + \beta_6 S \cdot A^2 + \beta_7 S \cdot A^3 + \end{aligned} \quad (3)$$

5. Det er ganske vigtigt i forhold til forudsigelsen i modellen at anvende GLM modellen i stedet for den traditionelle model med logtransformation. Den logtransformerede model giver en god prediktion af $\ln(\varepsilon)$ for alle aldersgrupper, men ved retransformation til ε fås en meget kraftig overvurdering af de predikterede sundhedsudgifter for de ældste.

6. A priori antagelsen om uafhængighed kan naturligvis kritiseres, men omvendt synes estimation af covariancen mellem restledene også at være meget afhængig af estimationsmetode (to trins selektionsmodel over for f.eks. maksimum likelihood model), specielt når der ikke er forskellige forklarende variable til rådighed, jf. Cameron og Trivedi (2005).

$$\sum_{j=0}^4 \beta_{8+j} DB(j) + \sum_{c=1}^4 \beta_{12+c} DF(c) + \sum_{k=1}^6 \beta_{16+k} AD(k-1, k)$$

$$\beta_{23} D2000 + \beta_{24} D2001 + \beta_{25} D2002 + u_{it} \quad \varepsilon_{it} \sim \text{Poisson}$$

Her er A alder, mens S er en dummyvariabel for køn ($S = 1$ for kvinder). Første linje i (3) angiver således sundhedsudgifterne som en funktion af personernes alder, hvor alderseffekterne er tilladt at variere for mænd og kvinder. De to næste led er medtaget for at tage højde for, at sundhedsudgifterne for små børn er meget høje ($DB(j)$) og at sundhedsudgifterne for kvinder i fødedygtige aldersklasser er relativt høje ($DF(c)$), jf. figur 1. Således er $DB(j)$ dummyvariabler for mindre børn i alderen 0 til 4 år, mens $DF(c)$ er dummyvariabler for kvinder i aldersintervallerne 20-24 år, 25-29 år, 30-34 år og 35-39 år. Betydningen af afstand til død måles ved parametrene til dummyvariablerne $AD(0,1)$.. $AD(5,6)$. Dummyvariablen $AD(0,1)$ angiver således, om individet dør i løbet af samme år, mens dummyvariablen $AD(1,2)$ angiver, om individet dør året efter osv. Hvis alle »afstand til død« dummyvariablerne har værdien 0, betyder det, at pågældende individ ikke dør inden for seks år efter årsobservationen for sundhedsudgiften. Endelig er $D2000$ til $D2002$ årsummyer, som er medtaget for at korrigere for effekten af årseffekter, som påvirker sundhedsudgiften i et givet år på tværs af individer. Det kan f.eks. være ændringer i det samlede offentlige ydelsesniveau pr. person eller ændringer som skyldes særlige behov i et givet år, f.eks. på grund af influenzaår. Året 2003 er base case.

Medtagelsen af dummyvariablerne for afstand til død svarer til den metode, der er anvendt i tidligere udenlandske undersøgelser, som har søgt at adskille alderseffekter og terminaleffekter. I formuleringen i ligning (3) er det antaget, at betydningen af afstand til død er ens på tværs af aldersgrupper. Dette harmonerer ikke med figur 2, hvor der er et »knæk« terminaludgifterne, som aftager for de ældre. Ligning (3) bør modificeres for at tage højde for dette knæk.

Ved fremskrivning af de fremtidige sundhedsudgifter i forhold til forøget levetid bør det overvejes, om terminaludgifterne også fremover vil aftage ved 60-70 års alderen, eller om »knækket« i terminaludgifterne forskydes i takt med, at levetiden forlænges. I førstnævnte tilfælde bør knækket modelleres som en funktion af personers alder, mens knækket bør beskrives som en funktion af forventet restlevetid for personer i en given aldersklasse, hvis det forventes, at knækket udskydes i takt med stigninger i levetiden. Empirisk kan dette kun afklares, hvis man har individ data for en længere periode, hvor der sker ændringer i de ældres restlevetid. Det forekommer dog mest plausibelt, at forlængelser af levetiden også udskyder faldet i terminaludgifterne. Det-

te opnås i den empiriske model ved at lade parametrene til terminaludgifterne afhænge af den aldersbetingede restlevetid. Konkret implementeres dette ved at medtage fire dummyvariabler for forventet restlevetid (*DR0-3*, *DR3-6*, *DR6-11* og *DR11-16*). Her angiver f.eks. *DR0-3*, at restlevetiden er mellem 0-3 år for personer i en given aldersklasse. Disse fire dummyvariabler for restlevetid indgår som interaktionseffekter med de seks dummyvariabler for afstand til død, hvilket giver i alt yderligere 28 yderligere parametre i modellen.⁷

Ligning (3) estimeres på individer med positive sundhedsudgifter. Probitmodellen i ligning (1) estimeres med samme variabler for alle individer for at bestemme sandsynligheden for at $\varepsilon_{it} > 0$.

Beregning af forventede sundhedsudgifter

De forventede sundhedsudgifter kan herefter beregnes som følger:

$$E[\varepsilon_{it} | \mathbf{X}_{it}] = \exp\left[\mathbf{X}_{it}' \hat{\boldsymbol{\beta}}\right] \times E[P[\varepsilon_{it} > 0 | \mathbf{X}_{it}]] \quad (4)$$

Her er de forventede sundhedsudgifter beregnet ved at multiplicere den forventede sandsynlighed for positive sundhedsudgifter ud fra probit estimationen med de forventede sundhedsudgifter i niveauer beregnet for alle i samplet.

5. Estimationsresultater

I dette afsnit præsenteres resultater ved estimation af ligning (3) og probit modellen. Estimation er foretaget med STATA version 10.

Selve estimationsresultaterne er vist i tabel A1 i bilag A. Det er vanskeligt direkte at tolke de estimerede parametre, dels fordi den afhængige variabel i GLM modellen er i logaritmer, dels fordi nogle af de forklarende variabler indgår som polynomier med flere led, og endelig fordi den samlede effekt afhænger af parametrene fra både GLM og probit modellerne. Da andelen af personer med positive sundhedsudgifter er meget høj (94 pct.), er det dog især parametrene i GLM modellen, som har betydning for de samlede resultater.

Det fremgår af tabel A1, at alle »afstand til død« dummyvariablerne er signifikante i GLM modellen. Som forventet er parametrene aftagende med afstand til død. Det fremgår dog også, at betydningen af afstand til død stadig er væsentlig større end nul

7. Bemærk at disse dummyvariabler er ens på tværs af individer med samme alder og køn (man kan således godt have en forventet restlevetid mellem 3 til 6 år selv om man dør året efter). I 2003 var restlevetiden knap 16 år for en henholdsvis 65-årig mand eller 69-årig kvinde, mens restlevetiden er knap 3 år for en 92-årig mand eller 95-årig kvinde, jf. Hansen og Barington (2009). Medtagelsen af interaktionsledene mellem faktisk afstand til død og intervaller for forventet restlevetid er medtaget for at få »knækket« i terminaludgifterne til at udskydes i takt med at levetiden forlænges (i forbindelse med forudsigelser på baggrund af modellen).

Tabel 1. Predikterede sundhedsudgifter i forhold til alder, køn og afstand til død.

Alder	Mænd			1000 kr.	Kvinder		
	Dør 0-1 år	Dør 3-4 år	Ikke død efter 5-6 år		Dør 0-1 år	Dør 3-4 år	Ikke død efter 5-6 år
10	39,5	14,5	3,5		37,3	13,9	3,3
20	48,6	18,0	4,3		76,3	30,6	7,2
30	61,9	23,3	5,6		113,7	45,7	10,8
40	79,0	30,2	7,2		91,8	36,5	8,6
50	98,2	38,0	9,0		110,9	44,3	10,5
60	115,9	45,1	10,7		124,4	49,7	11,7
70	102,0	40,1	16,4		102,3	41,2	16,9
80	82,7	38,8	19,5		78,2	37,0	18,7
90	58,0	30,3	19,4		52,5	27,6	17,7
100	33,7	21,0	14,6		30,0	18,7	13,0

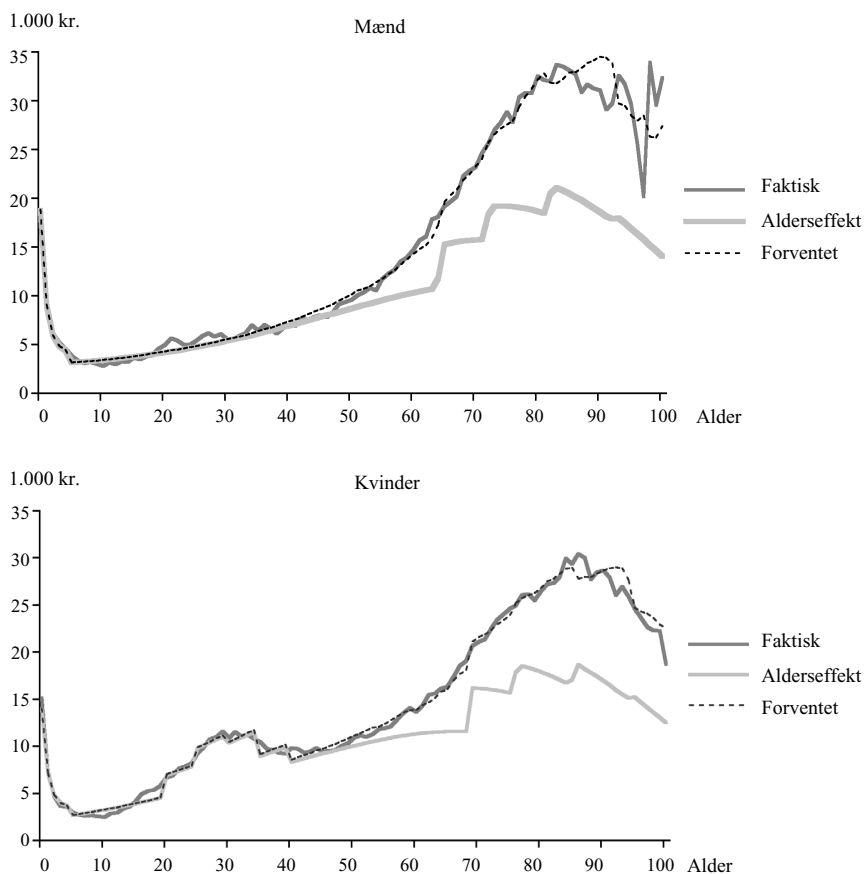
Anm.: Effekten af andre aldersrelaterede variabler (f.eks. dummier for kvinder i fødegytige aldersklasser og dummier for restlevetid) indgår også i prediktionerne i tabel 1. Sundhedsudgifter målt i faste 2006 priser beregnet på baggrund af serviceniveau i sundhedsudgifterne i 2003.

Kilde: Egne beregninger på baggrund af registerdata.

ved afstand til død 5-6 år. Dette tyder på, at dummier for afstand til død i mere end 6 år også burde inkluderes i modellen, hvis der var oplysninger om sundhedsudgifterne længere tid tilbage. I Seshamani og Gray (2004a) og Seshamani og Gray (2004b) findes også effekter af afstand til død i længere end 6 år, mens andre undersøgelser finder effekter af afstand til død i færre år, jf. f.eks. Zweifel m.fl. (2004).

Interaktionsledene mellem »afstand til død« dummyvariablerne og intervallerne for forventet restlevetid for alle i samme aldersklasse giver generelt de forventede resultater. Således har afstand til død stor betydning for sundhedsudgifterne, men for de ældste med kort forventet restlevetid er der en betydelig reduktion i effekten af afstand til død, som afspejler knækket i betydningen af afstand til død, jf. figur 2. Som eksempel er parameteren til afstand til død fra 0 til 1 år på 2,34 i GLM modellen, jf. bilags-tabel A1. For de ældste med restlevetid 0-3 år er der dog et fradrag på 1,56 (parameteren til interaktionseffekten mellem afstand til død 0 til 1 år og restlevetid 0-3 år), således at nettoparameteren af afstand til død 0 til 1 år for de ældste er på 0,78.

For at lette tolkningen af de estimerede parametre er beregnet predikterede værdier for sundhedsudgifterne for personer med forskellig alder, køn og afstand til død ved brug af ligning (6), dvs. den samlede effekt fra GLM og probit modellen, jf. tabel 1. Kolonnerne med »ikke død efter 5-6 år« er de predikterede sundhedsudgifter for personer, som ikke var døde i op til 6 år efter observationen for sundhedsudgiften. Med den givne specifikation kan dette tolkes som den »rene« alderseffekt. Det fremgår, at alder fortsat har væsentlig betydning for sundhedsudgifterne. For eksempel er den for-



Figur 3. Faktisk og forventet sundhedsudgift.

Anm: Sundhedsudgifter på baggrund niveau for sundhedsudgifterne i 2003 målt i 2006 priser. Alderseffekten er beregnet i prediktion ved at sætte alle afstand til død dummy variable til nul. »Hakkene« i alderseffekterne er udtryk for, at de tre dummyvariable for forskellige niveauer af restlevetid er en lidt ufleksibel beskrivelse af betydningen af alder for de ældste.

ventede sundhedsudgift for en 30-årig mand på ca. 5.600 kr. pr. år, mens det tilsvarende tal for en 80-årig er på 19.500 kr. pr. år. Analysen på danske data støtter således ikke »red herring« hypotesen om, at sundhedsudgifterne slet ikke afhænger af alder, når der tages højde for terminaleffekter. Aldersafhængigheden i sundhedsudgifterne er dog væsentlig mindre end det, der kommer til udtryk i figur 1.

Det fremgår endvidere, at for givne aldersklasser indtil 60-års alderen er de predikterede sundhedsudgifter ca. 11 gange højere for personer, der dør i løbet af året. For ældre mindskes den relative forskel i sundhedsudgifterne mellem personer, der dør i løbet af året, og overleverne.

Der er god overensstemmelse mellem de faktiske gennemsnitlige aldersfordelte sundhedsudgifter og de predikterede aldersfordelte sundhedsudgifter, jf. figur 3. For de ældste ses nogle forskelle mellem de observerede og forventede sundhedsudgifter, men dette kan tilskrives det mindre antal personer (gælder især for de ældste mænd).

Endvidere viser den nederste kurve, hvor meget af den predikterede sundhedsudgift, som kan tilskrives alderseffekten, hvor bidraget fra terminaludgifter sættes til nul. Mere præcist er alderseffekten de predikterede sundhedsudgifter for personer, som ikke dør inden for 6 år. Frem til omkring 40-års alderen består sundhedsudgifterne stort set udelukkende af alderseffekter, men herefter udgør terminaludgifterne en stigende andel af sundhedsudgifterne. Dette afspejler, at dødeligheden er lille for børn og yngre. Selv om terminaludgifterne er høje for børn og yngre, så er dødeligheden trods alt så beskedne for disse aldersgrupper, at terminaludgifterne gennemsnitligt har meget lille betydning.

De analyserede data for offentlige sundhedsudgifter medtager ikke oplysninger om udgifter til plejehjem og andre ældreplejeudgifter, jf. afsnit 3. Ud fra Danmarks Statistiks definition af sundhedsudgifter henregnes udgifter til plejehjem og dele af andre ældreplejeudgifter til sundhedsudgifter. Der er udført en række supplerende analyser af ældreplejeudgifter baseret på data for en kortere årrække fra Københavns Kommune, jf. De Økonomiske Råd (2009). Datagrundlaget for disse ældreplejeudgifter er ikke så godt som for de her analyserede udgifter til sundhed, men en sammenligning tyder på, at terminaludgifterne har større betydning i forhold til ældrepleje end for sundhedsudgifter. I en sammenligning af betydningen af alderseffekter og terminal-effekter for forskellige typer sundhedsudgifter finder Werblow m.fl. (2007), at terminaludgifterne også er af væsentlig betydning for sundhedsudgifter, som har karakter af ældrepleje.⁸ Med hensyn til andre kommunale offentlige sundhedsudgifter spiller afstand til død formentlig mindre rolle (f.eks. skoletandpleje!), men ældreplejeudgifterne udgør langt den største del af de samlede kommunale sundhedsudgifter (inklusive ældrepleje). Skønmæssigt ville det næppe give et markant andet billede af betydningen af terminaludgifter, hvis der var tilgængelige data til rådighed for alle de offentlig sundhedsudgifter. Betydningen af ældreplejeudgifter vil formentlig trække i retning af, at terminaludgifterne kan have endnu større vægt end fundet ud fra de her analyserede data.⁹

8. Analysen i Werblow m.fl. (2007) vedrører udgifter til »long term care«. Werblow m.fl. finder dog også, at alder i sig selv har betydning for udgifter til long term care (ligesom vi finder for sundhedsudgifter generelt).

9. Artiklen fokuserer på demografiske faktorerens betydning for de offentlige sundhedsudgifter. Omkring 15 pct. af de samlede sundhedsudgifter betales direkte af den enkelte bruger af sundhedsydelser (f.eks. egenbetaling til medicin, udgifter til tandpleje, briller mv.). Disse privat betalte sundhedsudgifter er formentlig mindre afhængige af afstand til død end de offentlige sundhedsudgifter.

I de empiriske analyser er sundhedsudgifterne bestemt som en funktion af afstand til død, som således antages eksogen. Der kan sættes spørgsmålstegn ved det rimelige i en sådan antagelse. Dette diskuteres nærmere i afsnit 7.

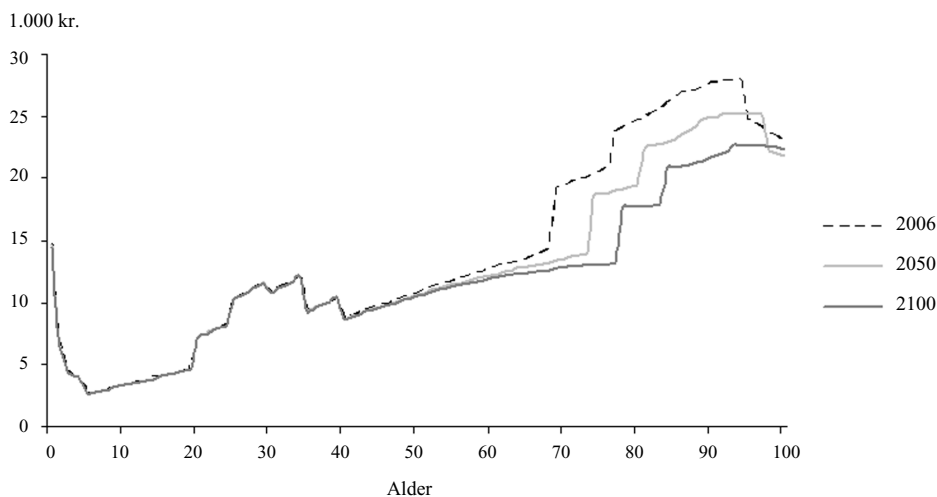
6. Udviklingen i de aldersfordelte sundhedsudgifter

Sundhedsudgifterne blev i afsnit 5 estimeret som en funktion af alder, køn og indikatorer for den enkeltes restlevetid, som tager højde for terminaludgifterne. Det fremgik, at en væsentlig del af de gennemsnitlige aldersfordelte sundhedsudgifter for de ældre kan tilskrives terminaludgifterne. Dette vil påvirke de gennemsnitlige aldersfordelte sundhedsudgifter ved forlængelser i levetiden og i sidste ende udviklingen i de samlede sundhedsudgifter. I dette afsnit vurderes den kvantitative betydning for de fremtidige sundhedsudgifter af at tage højde for terminaludgifter.

En vurdering af det fremtidige demografiske pres på sundhedsudgifterne kan fås ved at kombinere estimationsresultaterne med befolkningsfremskrivninger, som indeholder en prognose for udviklingen i de aldersbetingede dødeligheder. Ved beregningen erstattes de observerede »afstand til død« dummier med de forventede sandsynligheder for, at en person i en given aldersklasse vil dø samme år, året efter osv. Konkret anvendes DREAM's befolkningsfremskrivning, hvor der fremover ventes en fortsat stigning i levetiden som følge af fald i de aldersbetingede dødeligheder, jf. Hansen og Barington (2009). For eksempel havde en 85-årig kvinde i 2006 en dødsrisiko på 8,3 pct. inden for ét år. I 2050 er den forventede dødsrisiko ifølge fremskrivningen reduceret til 5,3 pct. for denne aldersgruppe, mens den er 3,2 pct. i 2100. De tilsvarende tal for en mand er henholdsvis 11,5 pct., 7,7 pct. og 4,6 pct. Derved reduceres de forventede terminaludgifter for en fremtidig 85-årig. Dvs. efterhånden som restlevetiden for ældre øges, vil terminaldelen af sundhedsudgifterne forskydes.¹⁰

Prediktion af aldersbetingede sundhedsudgifter på baggrund af estimationsresultaterne og den anvendte fremskrivning af udviklingen i de aldersbetingede dødeligheder giver en udvikling i aldersprofilen for sundhedsudgifterne, som er illustreret i figur 4. Sund aldring afspejler, at stigningen i sundhedsudgifterne afhænger af nærhed til død i stedet for alder i sig selv. Sund aldring giver sig derfor udslag i, at de gennemsnitlige aldersfordelte sundhedsudgifter pr. person falder i takt med, at de aldersfordelte dødeligheder over tid falder. Det ses, at sund aldring især har betydning for de gennemsnitlige aldersfordelte reale sundhedsudgifter for personer over 65 år. Som eksempel falder de forventede årlige sundhedsudgifter for en 85-årig kvinde med 13 pct. fra 2006 til 2050.

10. Beregningerne af de fremtidige aldersfordelte sundhedsudgifter afhænger således af udviklingen i de aldersfordelte dødeligheder. Der er naturligvis usikkerhed forbundet med fremskrivning af udviklingen i dødelighederne og befolkningsfremskrivningen mere generelt.



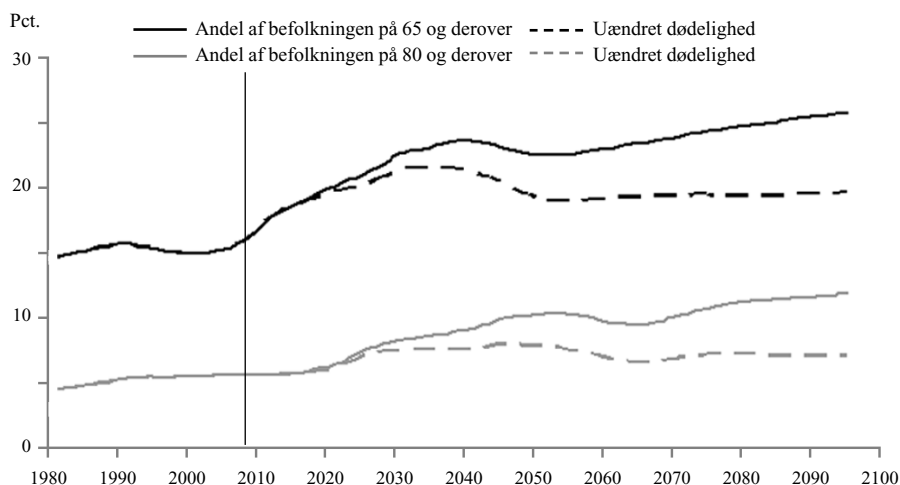
Figur 4. Fremtidige årlige aldersfordelte sundhedsudgifter for kvinder

Anm: Sundhedsudgifter beregnet på baggrund af niveau for sundhedsudgifterne i 2003 målt i faste 2006 priser.

Udviklingen i de fremtidige aldersfordelte sundhedsudgifter pr. person kan kombineres med udviklingen i befolkningens sammensætning med henblik på at vurdere de fremtidige sundhedsudgifter. Der vil i de kommende år være en kraftig stigning i andelen af ældre, jf. figur 5. En del af denne stigning kan tilskrives en årgangseffekt, som skyldes de store efterkrigsårgange. I figur 5 er årgangseffekten illustreret ved udviklingen i andelen af ældre givet, at restlevetiden er konstant (uændret aldersbetinget dødelighed). Den øvrige stigning i andelen af ældre kan tilskrives den forventede øgede restlevetid.

Betydningen af demografiske ændringer og sund aldring kan bedst illustreres ved at skelne mellem tre forskellige demografiske effekter, som påvirker sundhedsudgifterne fremover. Dels er der årgangseffekten, som vil bidrage til flere ældre frem mod midten af dette århundrede. Dels forventes middellevetiden at stige, således at de ældre trækker på sundhedssystemet i flere år. Endelig vil effekten af længere levetid delvist blive modvirket af sund aldring.

Størrelsen af disse tre effekter er illustreret i figur 6. Hvis det hypotetisk antages, at levetiden er uændret, følger de offentlige sundhedsudgifter den nederste (mørke optrukne) kurve. Væksten i sundhedsudgifterne er her alene bestemt af årgangseffekterne. Ud over årgangseffekter vil den forlængede levetid bidrage til, at der kommer flere ældre. Den samlede effekt af dette er vist ved den øverste (stiplede) kurve under antagelse af, at de gennemsnitlige aldersfordelte sundhedsudgifter pr. person ikke ændres med forlængelser af levetiden (dvs. en antagelse om usund aldring). Når der tages højde for, at



Figur 5. Udviklingen i andelen af ældre.

Kilde: DREAM's befolkningsfremskrivning jf. Hansen og Barington (2009). Befolkningsudvikling givet uændret dødelighed beregnet af Hansen.

forlængelser af levetiden leder til sund aldring (grå kurve) vil dette omtrent halvere stigningen i sundhedsudgifterne, som skyldes forlænget levetid.

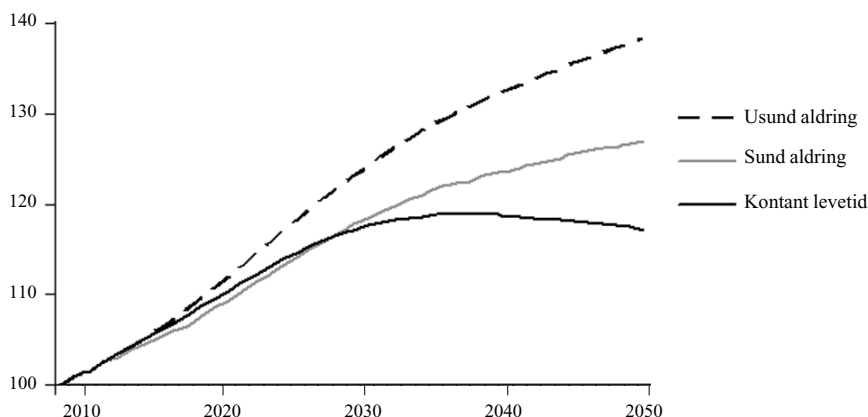
Frem mod 2050 vil der være en demografisk betinget vækst i de reale sundhedsudgifter på omkring 27 pct. Heraf kan ca. 17 pct. point tilskrives årgangseffekter, mens de resterende 10 pct. point kan tilskrives forlænget levetid (efter der er korrigeret for sund aldring).

Ud over de demografisk betingede stigninger i sundhedsudgifterne kan der i fremtiden også ventes en yderligere stigning, som skyldes, at sundhedsudgifterne historisk er vokset mere end indkomsten selv efter, at der er korrigeret for demografiske ændringer, jf. f.eks. De Økonomiske Råd (2009), Pedersen og Hansen (2006) og OECD (2006).

7. Endogenitet af afstand til død

I den estimerede økonometriske model er det antaget, at sundhedsudgifterne kan beskrives som en funktion af afstand til død. Der er således ikke taget højde for, at dødstidspunktet selv kan afhænge af behandlingsomfanget, dvs. at afstand til død potentielt er en endogen variabel (jævnfør afsnit 2). I den tidligere litteratur om terminaludgifter er endogenitetsproblemet noteret, jf. Salas og Raftery (2001), men i de hidtidige empiriske undersøgelser er endogenitetsproblemet ignoreret. Eneste undtagelse er Felder m.fl. (2008), som instrumenterer afstand til død ud fra en model for dødelig-

2008 = 100



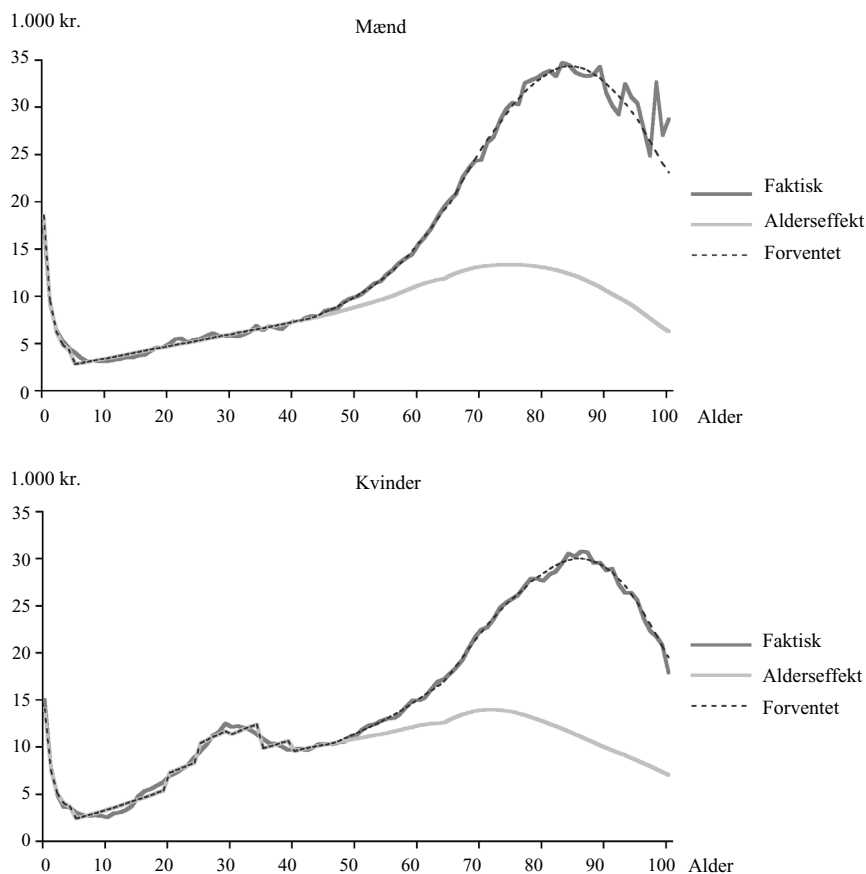
Figur 6. Demografiske effekters betydning for sundhedsudgifterne.

Anm: Beregninger foretaget med DREAM model og befolkningsfremskrivninger på baggrund af egne estimationsresultater. Sundhedsudgifter opgjort i fast lønniveau (opgjort ved pris- og produktivitetskorrigerede offentlige sundhedsudgifter).

heden for en tidligere periode. Felder m.fl. konkluderer, at det ikke formelt kan afvises, at afstand til død er endogen. Størrelsen af endogenitetsbias er imidlertid beskedne, og ændrer ikke ved, at man bør tage højde for afstand til døden.

Endogenitet af afstand til død opstår, hvis høje (lave) sundhedsudgifter til en given person er med til at forlænge (forkorte) restlevetiden. Der kan korrigeres for endogenitetsbias, hvis afstanden til død kan instrumenteres med en anden variabel, som er korreleret med afstand til død, men som samtidig er uafhængig af de sundhedsudgifter, den enkelte person modtager. Den forventede restlevetid for personer opdelt på alder, køn og år opfylder disse kriterier. Udviklingen i restlevetiden for en given kohorte er således korreleret med udviklingen i afstand til død for den enkelte og samtidig vil udviklingen i restlevetiden ikke være påvirket af sundhedsudgiften til enkeltpersoner. Anvendelse af den forventede restlevetid for personer i en given aldersklasse i stedet for den faktiske afstand til død har endvidere den fordel, at det er muligt at medtage sundhedsudgifter for en længere (og mere aktuel) tidsperiode. Estimationer med restlevetid er derfor foretages med data for perioden 2000 til 2007.

Der er imidlertid også ulemper ved at bruge forventet restlevetid i stedet for observeret afstand til død. Der er således langt mindre variation i forventet restlevetid – som i et givet år er ens for alle med samme alder og køn – sammenlignet med faktisk afstand til død. Endvidere er alder og forventet restlevetid meget korrelerede, hvilket i praksis kan gøre det vanskeligt at adskille effekten af de to. Der har dog i den pågæl-

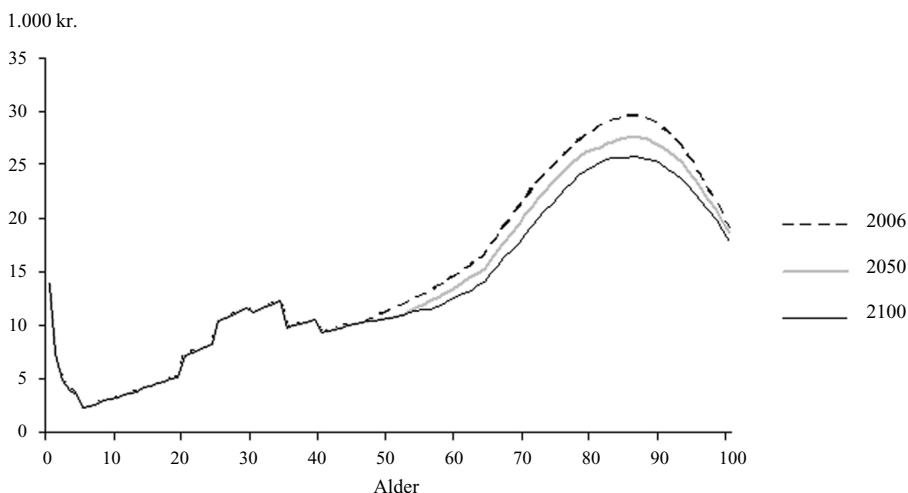


Figur 7. Faktisk og forventet sundhedsudgift (restlevetidsmodel).

Anm: Alderseffekten er i dette tilfælde beregnet ved at sætte restlevetiden til 35 år for alle, som har en faktisk restlevetid mindre end 35 år. Det bemærkes, at alders effekten ikke direkte kan sammenlignes med alders effekten i figur 3, som også indeholdt effekten af dummyvariabler relateret til restlevetid.

dende periode været en forbedring i restlevetiden, som betyder, at der er variation i restlevetiden for personer i givne aldersklasser. Det skal bemærkes, at de præsenterede estimationer med restlevetid har karakter af robusthedscheck og ikke er et formelt test af endogenitet.

I estimationerne med restlevetid i stedet for observeret afstand til død er det antaget, at restlevetid kun har betydning for personer med restlevetid under 35 år. I 2003 svarer dette til mænd over 42 år og til kvinder over 45 år. Denne begrænsning er lagt ind, fordi terminaludgifter ikke har nogen praktisk betydning for yngre, hvor dødeligheden er beskednen, jf. figur 3. Restlevetider (under 35) indgår både lineært og kvadreret. I mo-



Figur 8. Fremtidige årlige aldersfordelte sundhedsudgifter for kvinder (restlevedidsmodel).

dellen er endvidere medtaget en dummyvariabel for, hvorvidt restlevetiden er under 35. For at få en relativ fleksibel beskrivelse af effekten af restlevetid og alder, er det endvidere tilladt, at parametrene til restlevetiden kan ændre sig, når restlevetiden er mindre end 10 år samt at parametre til aldersvariablerne også kan ændre sig for personer på 75 år eller derover.¹¹

Estimationsresultater fra en sådan model er gengivet i tabel A2 i bilag A. Det fremgår, at også denne model giver en høj grad af overensstemmelse mellem de faktiske og de predikterede aldersfordelte offentlige sundhedsudgifter, jf. figur 7. Endvidere er den rene alderseffekt illustreret. Det fremgår, at alderseffekten har et forløb, som svarer nogenlunde til alderseffekten i modellen med observeret afstand til død, jf. figur 3.

Endelig viser figur 8 udviklingen i de fremtidige aldersfordelte sundhedsudgifter betinget af de forventede ændringer i restlevetiden. Det fremgår, at der også i denne model vil ske et fald i de forventede sundhedsudgifter for ældre i takt med at levetiden forlænges. Aldersafhængigheden og betydningen af ændringer i restlevetiden er dog relativt følsomme over for den konkrete modelspecifikation, hvilket afspejler, at korrelationen mellem forventet restlevetid og alder gør det svært at adskille effekterne af de to variabler.

Alt i alt giver modellen med forventet restlevetid resultater, som er sammenlignelige med modellen med observeret afstand til død. Da forventet restlevetid kan antages

11. Det er implementeret ved splines i parametrene til restlevetiden, når denne er mindre end 10 år og splines i parametrene til alder, når alderen er over 65 år.

for eksogen i forhold til den enkeltes sundhedsudgifter, tyder dette på, at modellen med observeret afstand til død ikke er behæftet med alvorlig endogenitetsbias. De to modeller er dog ikke helt analoge, og der er derfor ikke udført et egentligt test af eksogenitet af afstand til død. En enkelt udenlandsk undersøgelse tyder ligeledes på, at endogenitetsbias er beskeden. Endvidere tyder analyserne på, at det vil være behæftet med en meget betydelig fejl i fremskrivningen af sundhedsudgifter, hvis man ikke søger at tage højde for betydningen af terminaludgifter og sund aldring. Den potentielle endogenitetsbias vurderes således som beskeden i forhold til den fejl i fremskrivningen af sundhedsudgifterne, der vil være resultatet, hvis man ikke korrigerer alders-effekterne for terminaludgifter.

8. Konklusion

I de kommende år bliver der flere ældre på grund af forlænget levetid og de store årgange født lige efter anden verdenskrig. Sundhedsudgifterne er generelt højere for ældre end for yngre. Umiddelbart giver dette en forventning om et meget stærkt demografisk pres for øgede sundhedsudgifter. Analyserne af de danske offentlige sundhedsudgifter viser imidlertid, at dette pres ikke er så stærkt, som man umiddelbart forestiller sig. Det skyldes, at de offentlige sundhedsudgifter er høje i årene op til døden, og disse terminaludgifter udgør en væsentlig del af de gennemsnitlige offentlige sundhedsudgifter for de ældre. Ved forlænget levetid vil disse terminaludgifter blive udskudt. Dette vil dæmpe væksten i sundhedsudgifterne som følge af forlængelser af levetiden.

Analyserne viser, at det er vigtigt at tage højde for tilstedeværelsen af terminaludgifter ved beregning af den fremtidige demografisk betingede vækst i de offentlige sundhedsudgifter. Frem mod 2050 halveres væksten i sundhedsudgifterne som følge af forlænget levetid, når der tages højde for terminaludgifternes betydning.

Selv om analyserne bekræfter tilstedeværelsen af såkaldt sund aldring, så viser analysen også, at sund aldring ikke fjerner, men kun lægger en dæmper på væksten i sundhedsudgifterne som følge af forlænget levetid. Længere levetid leder til flere behandlingskrævende leveår, og højere alder bidrager også i sig selv til højere sundhedsudgifter til den enkelte, selv når der er taget højde for, at en væsentlig del af sundhedsudgifterne til ældre har karakter af terminaludgifter. Forlænget levetid vil således bidrage til et betydeligt pres på de offentlige sundhedsudgifter. I de kommende årtier vil der være et særlig stort pres på sundhedsudgifterne som følge af de store ældreårgange.

Litteratur

- Cameron, A. Colin and Trivedi, K. Pravin. 2005. *Microeconometrics – Methods and Applications*, Cambridge University Press.
- Cutler, D. M. og E. Meara. 1999. The Concentration of Medical Spending: An update. *NBER Working Paper 7279*.
- Cutler, D. M. og L. Sheiner. 1998. Demographics and Medical Care Spending: Standard and Non-Standard Effects. *NBER Working Paper No. 6866*.
- De Økonomiske Råd. 2009. *Dansk Økonomi, efterår 2009*. De Økonomiske Råd.

- DREAM. 2009. *Langsigtet økonomisk frem-skrivning 2009*. DREAM.
- Duan, N. 1983. Smearing estimate: A nonparametric retransformation method. *Journal of the American Statistical Association*, 78, s. 605-10.
- Felder, S., M. Meiers og H. Schmitt. 2000. Health care expenditure in the last months of life. *Journal of Health Economics*, 19 (5), s. 679 652-95.
- Felder, S., A. Werblow og P. Zweifel. 2008. Do Red Herrings Swim in Circles? *Ruhr Economic Papers* 73.
- Finansministeriet. 2007. *Mod nye mål – Danmark 2015: Teknisk baggrundsrapport*. Finansministeriet.
- Fogel, R. W. 2008. Forecasting the Cost of U.S. Health Care in 2040. *NBER Working Paper* 14361.
- Häkkinen, U., P. Martikainen, A. Noro, E. Nihtilä og M. Peltola. 2008. Aging, Health Expenditure, Proximity to Death, and Income in Finland. *Health Economics, Policy and Law*, 3, s. 165-95.
- Hall, R. E. og C. I. Jones. 2007. The Value of Life and the Rise in Health Spending. *Quarterly Journal of Economics*, 122 (1), s. 39 652-72.
- Hansen, M. F. og M. L. Barington. 2009. Danmarks fremtidige befolkning – Befolkningsfremskrivning 2009. DREAM.
- Hoover, D. R., S. Crystal, R. Kumar, U. Sambomorthi og J. C. Cantor. 2002. Medical Expenditures during the Last Year of Life: Findings from the 1992-1996 Medicare Current Beneficiary Survey. *Health Services Research*, 37 (6), s. 1625 652-1642.
- Madsen, J., N. Serup-Hansen, J. Kragstrup og I. Sønbo Kristiansen. 2002. Ageing may have limited impact on future costs of primary care providers. *Scandinavian journal of primary health care*, 20 (3), s. 169 652-173.
- Manning, W. G. og J. Mullahy. 2001. Estimating log models: to transform or not to transform? *Journal of Health Economics*, 20 (4), s. 461 652-494.
- OECD. 2006. Projecting OECD Health and Long-term Care Expenditures: What are the main Drivers, *Economics Department Working Papers*, 477.
- Pedersen, L. H. og M. F. Hansen. 2006. Stigende sundhedsudgifter – et aldrings- eller et velstandsfænomen? *Nationaløkonomisk Tidsskrift*, 144 (3), s. 304-25.
- Salas, C. og J. P. Raftery. 2001. Econometric issues in testing the age neutrality of health care expenditure. *Health Economics*, 10, s. 669 652-671.
- Serup-Hansen, N., J. Wickstrøm og I. Sønbo Kristiansen. 2002. Future health care costs – do health care costs during the last year of life matter? *Health Policy*, 62, s. 161 652-172.
- Seshamani, M. og A. Gray. 2004a. Ageing and health-care expenditure: The red herring argument revisited. *Health Economics*, 13 (4), s. 303-14.
- Seshamani, M. og A. Gray. 2004b. Time to death and health expenditure: an improved model for the impact of demographic change on health care costs. *Age and Ageing*, 33 (6), s. 556-61.
- Werblow, A., S. Felder og P. Zweifel. 2007. Population Ageing and Health Care Expenditure: A School of »red herrings«? *Health Economics*, 146, s. 1109-27.
- Zweifel, P., S. Felder og M. Meiers. 1999. Ageing of Population and Health Care Expenditure: A Red Herring? *Health Economics*, 8 (6), s. 485-96.
- Zweifel, P., S. Felder og A. Werblow. 2004. Population Ageing and Health Care Expenditure: New Evidence on the Red Herring. *The Geneva Papers on Risk and Insurance*, 29 (4), s. 652-66.

Bilag A Estimationsresultater

Table A1. Estimation af sundhedsudgifter for perioden 2000 til 2003.

Variabel	GLM model		Probit	
	Estimat	<i>t</i> -værdi	Estimat	<i>t</i> -værdi
Alder	0,01**	5,7	-0,01**	-13,0
Alder ² /1000	0,39**	7,6	0,78**	25,1
Alder ³ /1000000	-4,15**	-11,0	-7,38**	-25,7
Kvinde	-0,27**	-12,2	-0,20**	-17,3
Kvinde×alder	0,02**	11,4	0,03**	26,2
Kvinde×alder ² /1000	-0,44**	-8,3	-0,55**	-15,6
Kvinde×alder ³ /1000000	2,13**	5,9	2,30**	8,6
Nul-årig	1,81**	66,1	0,30**	21,9
Et-årig	0,98**	32,0	1,83**	40,9
To-årig	0,55**	20,1	1,60**	47,3
Tre-årig	0,34**	6,5	1,19**	53,2
Fire-årig	0,24**	5,6	1,00**	52,9
Kvinde 20-24 år	0,35**	16,9	0,62**	44,7
Kvinde 25-29 år	0,54**	31,3	0,54**	40,9
Kvinde 30-34 år	0,45**	28,9	0,42**	33,5
Kvinde 35-39 år	0,18**	12,0	0,27**	23,6
Afstand til død 0-1 år	2,34**	123,0	0,49**	10,7
Afstand til død 1-2 år	2,21**	91,9	0,11**	3,2
Afstand til død 2-3 år	1,71**	62,9	-0,02	-0,7
Afstand til død 3-4 år	1,45**	47,9	-0,08**	-2,8
Afstand til død 4-5 år	1,28**	41,9	-0,11**	-4,1
Afstand til død 5-6 år	1,15**	35,5	-0,02	-0,8
Forventet restlevetid 0-3 år	0,63**	6,2	1,27**	10,7
Forventet restlevetid 3-6 år	0,62**	20,3	0,83**	21,8
Forventet restlevetid 6-11 år	0,50**	28,5	0,39**	19,5
Forventet restlevetid 11-16 år	0,33**	27,0	0,08**	7,2
Afstand til død 0-1 år×Restlevetid 0-3 år	-1,56**	-15,2		
Afstand til død 0-1 år×Restlevetid 3-6 år	-1,29**	-45,5		
Afstand til død 0-1 år×Restlevetid 0-6 år			0,58**	5,7
Afstand til død 0-1 år×Restlevetid 6-11 år	-0,94**	-38,6	0,38**	4,7
Afstand til død 0-1 år×Restlevetid 11-16 år	-0,56**	-20,2	0,33**	3,5
Afstand til død 1-2 år×Restlevetid 0-3 år	-1,66**	-15,5	0,35*	2,2
Afstand til død 1-2 år×Restlevetid 3-6 år	-1,44**	-41,6	0,44**	6,7
Afstand til død 1-2 år×Restlevetid 6-11 år	-1,06**	-34,8	0,28**	5,5
Afstand til død 1-2 år×Restlevetid 11-16 år	-0,68**	-19,2	0,13*	2,4
Afstand til død 2-3 år×Restlevetid 0-3 år	-1,32**	-11,3	0,37*	2,3
Afstand til død 2-3 år×Restlevetid 3-6 år	-1,17**	-31,0	0,30**	5,4
Afstand til død 2-3 år×Restlevetid 6-11 år	-0,88**	-25,7	0,24**	5,2
Afstand til død 2-3 år×Restlevetid 11-16 år	-0,61**	-14,9	0,10*	2,0
Afstand til død 3-4 år×Restlevetid 0-3 år	-1,11**	-8,9	0,37*	2,1
Afstand til død 3-4 år×Restlevetid 3-6 år	-1,01**	-24,8	0,21**	4,0

fortsættes næste side ...

fortsat ...

Afstand til død 3-4 år×Restlevetid 6-11 år	-0,77**	-20,5	0,24**	5,6
Afstand til død 3-4 år×Restlevetid 11-16 år	-0,55**	-12,6	0,07	1,4
Afstand til død 4-5 år×Restlevetid 0-3 år	-1,00**	-7,4	0,20	1,1
Afstand til død 4-5 år×Restlevetid 3-6 år	-0,94**	-22,0	0,21**	4,0
Afstand til død 4-5 år×Restlevetid 6-11 år	-0,74**	-19,5	0,21**	5,1
Afstand til død 4-5 år×Restlevetid 11-16 år	-0,45**	-9,9	0,09*	2,1
Afstand til død 5-6 år×Restlevetid 0-3 år	-0,94**	-6,6	0,18	0,8
Afstand til død 5-6 år×Restlevetid 3-6 år	-0,92**	-20,7	0,05	0,9
Afstand til død 5-6 år×Restlevetid 6-11 år	-0,71**	-17,9	0,06	1,5
Afstand til død 5-6 år×Restlevetid 11-16 år	-0,49**	-10,3	0,02	0,4
År 2000	-0,10**	-15,3	-0,02**	-5,5
År 2001	-0,07**	-10,4	-4,27E-03	-1,1
År 2002	-0,03**	-4,9	-8,38E-03*	-2,1
Konstant	8,15**	360,5	1,16**	123,5
Antal observationer		1.996.332		2.122.823
Pseudo R^2		0,201		0,071

Anm.: »Afstand til død > 5-6 år« er referencekategorien for »afstand til død«-dummiene. Restlevetid > 16 år er referencekategorien for restlevetid-dummiene. År 2003 er referencekategori for årsummiene. ** angiver signifikans på 1 pct. niveau, mens * angiver signifikans på 5 pct. niveau. *T*-værdier er beregnet ud fra robuste standardafvigelse (tilnærmede *t*-værdier ud fra ML).

Tabel A2. Estimation af sundhedsudgifter i model med forventet restlevetid i stedet for afstand til død for perioden 2000 til 2007.

Variabel	GLM model		Probit	
	Estimat	<i>t</i> -værdi	Estimat	<i>t</i> -værdi
Alder	0,06**	15,6	-0,06**	-37,5
Alder ² /1000	-1,34**	-8,5	2,64**	38,5
Alder ³ /1000000	12,70**	6,5	-30,40**	-31,7
Kvinde	-0,26**	-13,9	-0,13**	-13,4
Kvinde×alder	0,02**	8,3	0,02**	15,6
Kvinde×alder ² /1000	-0,26*	-2,5	0,01	0,2
Kvinde×alder ³ /1000000	-1,37	-1,4	-2,72**	-5,4
Nul-årig	2,13**	66,3	0,16**	13,1
Et-årig	1,34**	42,8	1,60**	49,2
To-årig	0,86**	27,4	1,30**	59,1
Tre-årig	0,56**	14,9	0,89**	61,6
Fire-årig	0,40**	11,9	0,74**	59,9
Kvinde 20-24 år	0,20**	10,6	0,65**	65,5
Kvinde 25-29 år	0,40**	24,6	0,52**	51,9
Kvinde 30-34 år	0,35**	24,8	0,36**	36,0
Kvinde 35-39 år	0,12**	9,5	0,14**	15,2
65 år og derover (dummy)	-14,14**	-2,6	31,94**	4,9
Alder over 65	0,43*	2,1	-1,09**	-4,4
Alder over 65 ² /1000	-3,24	-1,3	10,59**	3,3
Alder over 65 ³ /1000000	-0,49	0,0	-20,40	-1,5
Kvinde over 65 (dummy)	10,74*	2,3	-25,56**	-3,9
Kvinde×alder over 65	-0,43*	-2,4	1,07**	4,2
Kvinde×alder over 65 ² /1000	5,37*	2,4	-14,92**	-4,5
Kvinde×alder over 65 ³ /1000000	-19,90*	-2,1	68,60**	4,9
Restlevetid under 35 år (dummy)	0,42	1,3	3,20**	16,9
Restlevetid (i år hvis under 35)	-0,01	-0,5	-0,18**	-17,2
Restlevetid ² /1000 (i år hvis under 35)	-0,10	-0,4	2,49**	16,4
Restlevetid under 10 år (dummy)	0,43*	2,3	-1,17**	-4,1
Restlevetid (i år hvis under 10)	-0,08*	-2,1	0,18**	3,1
Restlevetid ² /1000 (i år hvis under 10)	3,38	1,7	-7,19*	-2,2
År 2000	-0,09**	-13,0	-0,03**	-7,2
År 2001	-0,06**	-9,0	-0,01*	-2,5
År 2002	-0,03**	-4,2	-0,01*	-2,2
År 2004	0,02**	2,9	-0,01	-1,4
År 2005	0,06**	8,4	0,02**	4,0
År 2006	0,06**	9,3	0,04**	9,9
År 2007	0,08**	12,1	0,04**	10,0
Konstant	7,74	255,1	1,40	141,9
Antal observationer	3.994.637		4.243.342	
Pseudo R ²	0,1232		0,0762	

Anm.: År 2003 er referencekategori for årsummierne. ** angiver signifikans på 1 pct. niveau, mens * angiver signifikans på 5 pct. niveau. *T*-værdier er beregnet ud fra robuste standardafvigelses (tilnærmede *t*-værdier ud fra ML).

Approaching the value of a life year – Empirical evidence from a Danish contingent valuation survey

Jytte Seested Nielsen

Department of Health Economics, University of Southern Denmark, E-mail: jns@sam.sdu.dk

SUMMARY: The accepted standard for dealing with changes in mortality in cost-benefit analyses has been to estimate the value of a statistical life. However, when valuing premature mortality caused by a permanent change in air pollution it has been suggested to value a gain in life expectancy instead and these values should then be aggregated into the value of a life year. Theoretical and empirical research in this area is, however, scarce. This paper presents results from a Danish web-based contingent valuation survey in an air pollution context and finds the VOLY to be in the range of EUR 30,000-90,000.

1. Introduction

The harmful effects of particulate air pollution on health has long been recognized. Both in Denmark and within the European Union considerable efforts have been devoted to investigating the different aspects of air pollution. In the economic analysis of the impacts caused by a change in air pollution a variety of ways to handle the valuation of a change in mortality has emerged. The »Value of a Statistical life« (VSL) approach is based on the valuation of a one-period change in the probability of dying and was developed for cases in which the valuation of an »acute death« (which takes place instantly after an accident) is needed, e.g. in the traffic sector, Mishan (1971); Schelling (1968). However, an air-pollution related change in mortality often takes place some time after the individuals have been affected by the change in air pollution or as a consequence of long-term exposure, Pope et al. (2002) – and is often termed »chronic death«. This cannot be integrated in the one-period model that underlies the original VSL approach. Moreover, instead of calculating number of premature deaths, epidemiologists prefer to compute the environment-related change in mortality in terms of life-years lost, Brunekreef (1997). This has led to the idea of the value of a life-year (VOLY).

Generally, in order to find the willingness-to-pay (*WTP*) for a non-market good, data can be found either indirectly from observing people acting in the market or directly from people's responses to hypothetical questions. Hence, *WTP* can be estimated indirectly based on the revealed preference methodology (hedonic pricing or averting behaviour studies) or directly via the stated preference technique (contingent valuation (*CVM*) or discrete choice methods). There have been only a few empirical attempts to value a life-year. One empirical method applied is to find the *WTP* for a change in life expectancy directly, by means of a stated preference technique. Subsequently, the value of a »statistical life-year« can be found by aggregating across individuals. This will not be the value of one specific individual's life-year. Instead, it will reflect, for instance, 12 individuals' valuation of a change in their life expectancy of 1 month.

The aim of this paper is to estimate a Danish value of a gain in life expectancy directly in an air pollution context. As such, the paper offers a contribution to the limited number of studies in the area of valuing changes in life expectancy. The results are based on data from the first web-based *CVM* survey on the value of a change in life expectancy. The paper consists of a brief discussion of the background for the valuation of a change in life expectancy in an air pollution context. As part of the background section three different approaches to valuing a change in life expectancy are outlined. Section 3 describes the method and the *CVM* survey. The next section presents the key results and tests the validity of the results. The results are discussed in section 5 and the final section concludes.

2. Background

2.1. The Impact Pathway Method

For investigating the different aspects of air pollution the so-called »impact-pathway« approach has evolved through a range of EU projects named ExternE¹ (Externalities of Energy). This method has been developed in order to be able to link the measurement, identification and valuation of the impacts of air pollution. It is a »bottom-up« approach in which the point of departure is a particular emission from, for example, a power plant followed by measurement of the way in which the emission in question is dispersed. The impacts are subsequently estimated, for example in the form of dose-response functions, that estimate the physical effects on human morbidity and mortality separately. Finally, economic methods are employed to assign values to each of the effects that can be avoided by reducing the pollution. The approach has been adapted directly to the Danish context in Andersen et al. (2004) and further developed

1. ExternE's studies are published in a range of reports, *ExternE Externalities of Energy Vol 1-10*. See also <http://www.externe.info/>

in a Danish model system named EVA (Economic Valuation of Air pollution), Andersen et al. (2008).

Given that the value of a change in mortality can comprise over 80% of the health damage estimates, Andersen et al. (2004), the discussion on how to value a change in mortality is of great importance in relation to the impact-pathway method. As such, the results from this paper and similar studies on the valuation of a change in morbidity conducted within the EU, Ready et al. (2004) form the basis for further increasing the validity of the results from the impact-pathway method.

2.2. Estimating the value of a life year (VOLY)

A number of different approaches have emerged for valuation of a change in mortality risk. The original framework for dealing with physical risks was developed for the use in cost-benefit analyses in the traffic sector and here the so-called *VSL* model has evolved, Jones-Lee (1976); Mishan (1971). *VSL* is defined as the population mean of the marginal rate of substitution between wealth and risk, Jones-Lee (1991), and describes the aggregate amount that a large group of people would be prepared to pay for small individual risk reductions which, when applied to the whole group, will reduce the expected number of fatalities during the forthcoming period by precisely one. This single-period framework for dealing with changes in the probability of dying was formalized in the expected utility framework in Drèze (1962). The framework was further elaborated in Jones-Lee (1976); (1989) and has formed the basis for numerous empirical surveys. However, the need for a theoretical economic model to value lifetime extensions instead of prevented fatalities originates in the approach whereby epidemiologists calculate the change in mortality from a change in air pollution. Some attempts to develop a theoretical set-up for valuation of changes in life expectancy have been presented in Jones-Lee (1989) and in connection with life-cycle consumption models, Rosen (1988); Cropper & Portney (1990); Shepard & Zeckhauser (1984). However, considerable challenges, both theoretical and empirical, need to be overcome before a clear-cut approach to valuing changes in life expectancy can emerge.

Within the field of empirical estimation of the value of a life year lost, three different approaches have evolved; two indirect and one direct. The two indirect approaches are both based on results from *VSL* surveys. They will be explained in the following. In the direct approach applied in the present survey, people are directly asked to value a change in life expectancy in a *CVM* survey.

The first indirect approach uses the mathematical relationship² between a change in

2. An individual's life expectancy is the sum of the individual's conditional survival probabilities; the area below the survival curve. A change in the probability of dying is also a change in life expectancy, since a change in the probability of dying one year will affect the whole survival curve.

the probability of dying and a change in life expectancy for an individual directly. The method uses the *WTP* values from a conventional *VSL* survey in which the respondents are asked to value a one-period change in the probability of dying, Alberini et al. (2006). Following the stated values and taking the respondents' age into account it is then possible to calculate the change in life expectancy that the respondents, have implicitly valued (the average in the sample being 1.23 months). Finally, respondents' valuations of a change in life expectancy are aggregated to the value of a life-year lost. The result of the survey in Alberini et al. (2006) is a value of a life-year equal to a median of EUR 54,000, or a mean of EUR 163,000. Hence this indirect method uses the empirically well-established *VSL* framework. However, by transferring the stated values to a change in life expectancy, this method assumes comparability of the scale used for valuing changes in the probability of dying and changes in life expectancy. This assumption has been questioned in empirical work by McNeil et al. (1982) and Morris & Hammitt (2001). Also, since this indirect method is based on the single-period *VSL* model, the method cannot take into account that a change in life expectancy can arise as a consequence of various changes in the survival curve. Accordingly, the model cannot take into account that the *WTP* assigned to a change in life expectancy could be dependent on the specific change in the survival curve. This introduces the need for a time dimension and the need for a multi-period setting as incorporated in the life-cycle consumption models, Rosen (1988); Cropper & Portney (1990); Shepard and Zeckhauser (1984) or the multi-period *VSL* model (Hammitt & Liu (2004); Jones-Lee (1989).

The second indirect approach is very similar and has been applied within the ExternE framework and the starting point for this procedure is an existing aggregated empirical estimate of the *VSL* at age a ($VSL(a)$).³ With theoretical reference to the life-cycle consumption models, Cropper & Sussman (1990), the method converts *VSL* into the value of a life-year. The method applied is described in *EQ* 1, but has been applied in various ways, differing with respect to discount rate and inclusion of conditional survival probabilities.

$$VSL(a) = VOLY_r \sum_{i=a+1}^{i=T} P_i (1+r)^{-(i-a-1)} \quad (\text{Holland et al. (1998)}) \quad (1)$$

3. Though not directly comparable, this indirect method has also been applied within health economics attempting to find the *WTP* for a Quality Adjusted Life-Year (*QALY*), Hirth et al. (2000); Johannesson & Meltzer (1998); Abelson (2003). For instance, in Hirth et al. (2000) a value for a Quality Adjusted Life-year based on 35 *VSL* estimates is derived. The overall median value is equal to USD 265,345 (1997) (approximately EUR 330,000).

Table 1. Values of changes in life expectancy. The direct approach. The values reflect the risk reduction valued without any aggregation and are given in EUR (2004) corrected for inflation and purchasing power parity.

<i>Author</i>	<i>Context</i>	<i>Risk Reduction</i>	<i>Mean WTP</i>	<i>Median WTP</i>	<i>Payment</i>	<i>N</i>
Johannesson & Johannesson (1996)	private	1 year	1472		One time	2013
Morris & Hammitt (2001)	private	11; 5 months		456;294	One time	
Chilton et al., (2004)	public	1; 3 ; 6 months (normal health)	74; 84; 100	31;37;49	Household WTP per year for the rest of life	517
Desaigues et al., (2007a)	No	1; 3 ; 12 months (among others)	140;234;377		Every year for the next ten years	59

${}_aP_i$ is the conditional probability of surviving from age a to age i , conditional on surviving to age a , and r is the discount rate.⁴ Again, this method applies the theoretically as well as empirically well-documented *VSL* framework. The disadvantages with this method are that on top of the assumptions given in the life-cycle consumption literature, Cropper & Sussman (1990), additional assumptions are required in order to justify assigning each life-year equal value. According to Krupnick et al. (1999), a requirement for using *EQ1* is constant utility of consumption for all t . Moreover the implicit implication of using a constant *VOLY* is that the value of a prevented fatality will decrease monotonically with age. However, this contradicts with the theoretical result that *VSL* is constant across the entire life-cycle when consumption is constant⁵ (and time preferences and hazard rates are constant as well), Johannesson (2002).

The approach used in this survey is the direct approach in which the respondents are asked directly about their *WTP* for a change in life expectancy. Up until the initiation of the *NEEDS* questionnaire there have been only four published surveys on the direct valuation of a change in life expectancy. Table 1 presents an overview of these.

4. The discount rate is not discussed further in this application. However, in the author's opinion the discount rate must be the individual's discount rate for consumption.

5. The author is very grateful to Prof. Per-Olov Johannesson (Stockholm School of Economics) for making this point.

By valuing a change in life expectancy the direct method acknowledges the need for a multi-period setting and compared to the indirect approaches, the direct approach has a major advantage since we do not have to rely on the comparability of the scale between valuing a change in life expectancy and a change in the probability of dying.

3. Method

3.1. The questionnaire

This paper examines the direct valuation of a change in life expectancy in the context of air pollution using the *CVM*. Development of the questionnaire took place in the realm of the NEEDS project (New Energy Externalities Development for Sustainability), refer to Desaignes et al. (2007b) for a full description of the development of the questionnaire. The study was conducted using an internet panel. First, general information about the background to the project and its rationale was provided. Attention was then devoted to the respondents' knowledge of the effects of air pollution on health, air pollution-related illnesses in their family and their general health and lifestyle. Given a respondent's age at the time of the survey, information about the average life expectancy was presented and subsequently, respondents were informed about factors that affect individual life expectancy (genetic factors etc.).

In the next stage of the questionnaire, two potential policies that could reduce air pollution and hence generate gains in life expectancy were outlined. The remaining information focused on presentation of a schematic graph developed specifically for the NEEDS-project. This diagram is presented in the Appendix. The explanations were split into three steps on the screen; the survival curve, the change in the survival curve and an emphasis of that the survey was concerned with a change in life expectancy in normal health. By showing the survival curve, weight was given to the fact that a gain in life expectancy is not a certain »add-on« at the end of an individual's life. To justify the payment vehicle the respondents were told that all polluters will have to respect stricter emission limits and that as a consequence of the increased costs to industry there will be a general increase in the cost of living.

The respondents were all asked to value two different changes; firstly an average gain in life expectancy equal to three months and secondly an average gain equal to six months.⁶ In order to identify »protest votes« from those who legitimately did not value the benefit, each respondent was asked first of all whether they would in principle be willing to pay something in the form of higher prices to gain an increase in their life expectancy. If they responded »no« to this question they were asked to note down their reasons. For those who said »yes« the respondent would be reminded at this stage of their budget constraint. A randomized card sorting procedure, Chilton et al. (2004)

6. In the original NEEDS questionnaire, the respondents were given the questions in reverse order.

was applied as the payment mechanism followed by an open-ended *WTP* question. Accordingly, if the respondents indicated a *WTP* above 0 they were presented with 11 different payment cards that each showed a monthly and an annual amount which they were going to pay for the rest of their lives. The first card was picked randomly by the computer.⁷ They were then asked to sort the cards into three piles (would definitely pay, would definitely not pay, or unsure) and the card-sorting was followed by an open-ended *WTP* question, which is shown in the Appendix. Finally in the debriefing section, individuals were asked some questions to better understand their answers and they were offered an opportunity to change their stated values.

The respondents to the web-based survey were all resident in the Copenhagen Area. The data collection was carried out in the period from December 2006 to January 2007. The respondents were all members of a Danish internet panel with 36,000 active panel participants. The panel is representative of individuals over 15 years old who have access to the Internet.⁸ All participants were recruited actively by use of interviews and no panel participants are self-recruited. The questionnaire was emailed to the respondent via a link in an email and the respondents answered the questionnaires on their personal computer. Two follow-up emails with the same link were emailed to the respondents as reminders (one after 4 days and one after a week). All respondents were entered into a draw for a voucher worth EUR 40 (one voucher per 100 respondents).

3.2. Estimating the value of a life-year lost (*VOLY*)

Using the *WTP* estimates from the survey, the net present value $NPV_j(3_months)$ ($NPV_j(3)$) of individual j 's monthly WTP_i for a gain in life expectancy of three months can be calculated in the following way:

$$NPV_j = \sum_{t=0}^{12-LE_j} \frac{WTP_j(3)}{(1 + r_j(t))^t} \quad (2)$$

in which LE_j is individual j 's remaining expected life expectancy measured in years, $r_j(t)$ is individual j 's discount rate for month t and $WTP(3)$ is the monthly amount that the respondent is willing to pay for the rest of his/her life for a gain in life expectancy equal to three months.

Subsequently, the »value of a one-year increase in aggregate life expectancy« can

7. However, the same card was picked for the two valuation questions (a gain in six and three months) for each respondent.

8. According to Statistics Denmark, in 2007 86% of the Danish population had access to the Internet, Statistics Denmark (2007).

be calculated by increasing 4(2) individuals' life expectancy by 3(6) months. Accordingly we could estimate the mean value of a life year as follows:

$$VOLY(3) = \frac{1}{N} \cdot 4 \sum_{j=1}^N NPV_j(3) = \frac{1}{N} \cdot 4 \sum_{j=1}^N \sum_{t=0}^{12-LE_j} \frac{WTP_j(3)}{(1+r_j(t))^t} \quad (3)$$

There is a large body of literature on discounting and which discount rate to apply, see, Møller (2009); Gyrd-Hansen & Søgaaard (1998); Gafni (1984). In Chilton et al. (2004) the *VOLY* results are presented both with a pure utility rate and without discounting. The present paper will not discuss discounting in detail yet in order to demonstrate sensitivity to choice of discount rate, the results will be presented without a discount rate, following Chilton et al. (2004) and with two different discount rates; 3% and 6%. The chosen discount rates of 3% and 6% represent the recommended rate for environmental policies in Denmark, Møller et al. (2000) and the recommended rate from the Danish Ministry of Finance (Finansministeriet 1999) respectively.

3.3. Description of key variables and the hypothesis

In Table 2, the variables used in the analyses are described and the proposed hypothesis in a test of internal validity of the *WTP* estimates is presented. Internal validity in this paper will be examined in Section 4.3 in an econometric analysis by examining whether the signs and sizes of the estimated coefficients are in accordance with the theoretical expectations. Only variables with an *a priori* expectation are included in the analysis. In addition, validity will also be examined in a test for scope sensitivity.

With regards to income, the theoretical predictions from the one-period model by Jones-Lee are that the marginal value of a decrease in risk is an increasing function of initial wealth and that a safety improvement can be characterized as a normal good, Jones-Lee (1976). Two proxies for wealth are included in the analyses; household income and a dummy variable for whether or not the respondent owns a house. Considering health, predictions are not as clear-cut. In the static one-period model, whether $VSL(\text{good health}) \gtrless VSL(\text{bad health})$ depends on whether the effects of health on the marginal value of wealth outweighs its effects on the total utility of survival, Hammitt (2000). A dummy variable for whether or not the respondent has or has had any serious symptoms is included in the analyses. Since it has been claimed that a change in the risk of dying in relation to a change in air pollution mainly affects the elderly⁹ it is very important to investigate how the value of a change in the probability of dying

9. Although this has been shown only in time-series studies and not yet in cohort studies, there appears to be consensus about this assumption, due to the specific mechanisms of air pollution death.

varies with age. A relationship between the *WTP* for a risk reduction and the age of the individual pictured by an inverse U-form peaking at about age 40 has been found in some life-cycle consumption analyses, Shepard & Zeckhauser (1982).¹⁰ However, it is important to emphasise that the inverse U-form found in the life-cycle models is based on a time-invariant utility function, and it has not yet been possible to establish the theoretical expectation about the value of a change in the probability of dying when taking into account that the marginal utility of wealth could change with age.

As an indication of potential altruistic concern two proxies have been included. Altruism arises if the *WTP* of an individual reflects not only concern for his own consumption but concern for the consumption of others as well, Bateman et al. (2002). Hence, the presence of altruism can have consequences for the summation of the individualistic valuation of the change in the probability of dying. The consequences depend on the nature of the altruism. Different kinds of altruism have been given much attention in the literature on welfare economics, Sen (1977); Jones-Lee (1992) and a variety of definitions have developed. In Sen (1977) a distinction is made between *sympathy* and *commitment*, where the former corresponds to the case in which the concern for others directly affects one's own welfare, whereas an example of commitment could be a case that does not make you feel personally worse off but since you think it is wrong you are ready to do something about it. It can be argued that sympathy enters the personal utility function directly whereas it is not as clear-cut with commitment, which makes the case of commitment difficult to incorporate in the utilitarian framework. Whether or not the respondent has children has been used as a proxy for sympathy and »interest in environmental issues« has been included as a proxy for commitment. For the latter, the respondent's location in the so-called Gallup-Kompas (»Gallup Compass«) has been used.¹¹ The GallupKompas is a segmentation tool and in this analysis a dummy variable has been coded based on whether or not the respondents belong to the »social-oriented« segments in this compass. The »social-oriented« segments are characterized among other things by indicating an »interest in environmental issues«. The variable »interest in environmental issues« has been taken as an indication of commitment based on this association between being »social-oriented« and showing an »interest in environmental issues«. Finally, the expectation with regards to the individuals' educational level is mixed, since a control for income will be incorporated into the regressions.

10. The inverse U-form is found in both the so-called 'Robinson Crusoe' case as well as in the case of perfect markets.

11. When the respondents enter the panel they are grouped into one of nine homogenous segments. The segmentation is accordingly independent of the topic of this specific questionnaire and information given to the respondents.

Table 2. Results for variables used in the regression analyses ($N=1064$). The income variable is in EUR (2007).

Variable	Variable	Mean (s.e.)	Hypothesis
Income	Household income per month, after tax	2513 (38)	+
House	Dummy equal to 1 if respondent owns a house	0.25	+
Serious health symptoms	Dummy equal to 1 if the respondent has or has had one of the following symptoms: bronchitis, asthma, emphysema, cardiovascular disease, cancer or other chronic diseases	0.20	+/-
Age	Age in years	49.5 (10.8)	+/-
Age-group 30-39	Dummy equal to 1 if the respondent is between 30 and 39 years old	0.22	
Age-group 40-49	Dummy equal to 1 if the respondent is between 40 and 49 years old	0.26	
Age-group 50-59	Dummy equal to 1 if the respondent is between 50 and 59 years old	0.28	
Age-group 60-69	Dummy equal to 1 if the respondent is between 60 and 69 years old	0.24	
Children	Dummy equal to 1 if the respondent has children in household	0.41	+
Concern for environment	Dummy equal to 1 if the respondent has indicated concern for the environment in the GallupKompas ¹	0.46	+
Male	Dummy equal to 1 if the respondent is male	0.49	+/-
Higher education	Dummy equal to 1 if the respondent has higher education	0.46	+/-

Note: (1) Refer to <http://www.gallup.dk/vores-markedsfokus/medier/printmedier/gallupkompas.aspx> for a description of each segment /Assesed 16 March 2010, in Danish).

The test of scope is a validity test in the sense that one would expect an increase in *WTP* when the scope of the good in question increases, Carson (1997); Carson et al. (2001); Hanemann (1994). Whether the scope test is fulfilled or not can be tested either externally (by varying the size of the good between samples) or internally (by asking the respondent to value different amounts of the good) as in the current survey, in which the respondents are asked to value a three-month gain in life expectancy as the first question and then asked to value a six-month gain.

4. Results

Out of 3003 invited panel members, 1,064 responded to the questionnaire, resulting

Table 3. Willingness-to-pay results (*WTP*) in EUR per month.

	Mean <i>WTP</i> (<i>s.e.</i>)	<i>N</i>
<i>LE=3 (EUR (2006/2007))</i>		
Total sample (excl. protesters)	74(11)	1017
Total sample (excl. protesters and <i>WTP</i> /household income > 1)	63(3)	1016
Total sample (excl. protesters and <i>WTP</i> /household income > 0.25)	63(3)	1014
Total sample (excl. protesters and <i>WTP</i> /household income > 0.10)	58(2)	994
<i>LE=6 (EUR (2006/2007))</i>		
Total sample (excl. protesters)	86(11)	1017
Total sample (excl. protesters and <i>WTP</i> /household income > 1)	76(3)	1016
Total sample (excl. protesters and <i>WTP</i> /household income > 0.25)	75(3)	1014
Total sample (excl. protesters and <i>WTP</i> / household income > 0.10)	70(3)	994

in a response rate of 35%. The mean duration of the interview was 26 minutes. For this survey the agency sampled based on quotas on age (30-70), sex and education. The quotas are representative for the Copenhagen area. However, it is apparent that despite the quotas, the sample has a higher educational level than that found in the Copenhagen area population. Also, comparing the mean income level in Denmark with the mean level of income in this sample reveals a higher mean income level than the Danish mean income level.¹² Table 2 summarizes the hypothesis from above and the findings in the survey with regards to the variables used in the regression analyses presented later.

4.1. Willingness-to-pay results

47 (4%) respondents were classified as protesters¹³ and 42 (4%) respondents had a *WTP* equal to zero. It is evident that some respondents indicated a very high *WTP*. This raises the question whether these can be defined as outliers. According to Collett (1991, p. 141) outliers are »... observations that are surprisingly distant from the remaining observations in the sample...«. Naturally, it is difficult to establish what »surprisingly far away« means. However, in Table 3 below results are reported with exclusions where there was a high proportion of *WTP* for a three-month gain in life expectancy compared to household income. By excluding only one observation with a *WTP* estimate above the household income, the *WTP* estimate is reduced by 15% and the std. error is reduced by 73%. Excluding more observations does not have the same substantial effect and it is thereby inferred that this observation is surprisingly far away from the remaining observations and, as such, is excluded from the sample.

12. According to Statistics Denmark, the overall household income level was EUR 4,000 per month before tax (Statistics Denmark 2004).

13. Protest responses are classified as respondents who refused the scenario or stated that another source should pay.

Table 4. VOLY results without outliers and protesters, in EUR.

	VOLY(6 months question) <i>n</i> = 1016		VOLY(3 months question) <i>n</i> = 1016	
	Median	Mean	Median	Mean
No discounting	28,930	56,640	45,600	93,910
Discounting (<i>r</i> = 0.03)	18,960	35,390	29,470	58,800
Discounting (<i>r</i> = 0.06)	12,980	24,470	20,810	40,726

4.2. VOLY estimates

As mentioned, following the *WTP* estimates from the survey, the »value of a year's increase in aggregate life expectancy« could be calculated by increasing 4(2) individuals' life expectancy by 3(6) months. The results of applying formulas 2-3, are shown in Table 4 above. To show sensitivity to discounting, the results are presented without a discount rate as well as with a discount rate equal to 3% (the recommended rate for environmental policies in Denmark) and a discount rate equal to 6% (the recommended rate from the Danish Ministry of Finance).

4.3. Testing for internal validity

Validity will be examined in a test for scope sensitivity as well as in an econometric analysis of whether the signs and sizes of the estimated coefficients are in accordance with the theoretical expectations. According to a box-cox regression there seems to be some non-linearity issues and therefore a log-log regression has been carried out in order to examine whether the signs and sizes of the estimated coefficients are in accordance with the theoretical expectations. In addition to the control characteristics from the quota sampling only variables with an *a priori* expectation as to the sign of the relationships are included in the regressions. Two independent regressions are carried out; for the first question on a gain in life expectancy equal to three months and for the second question on a gain equal to six months. The results are presented in Table 5.

Since the error terms across equations may potentially be correlated, the appropriate model choice would be the SUR (seemingly unrelated regression) model. However, since the same regressors are included in the two regressions the SUR estimation and ordinary least-squares estimation are identical, Pindyck and Rubinfeld (1998).

The results of the regressions show that the proxies for wealth (owning a house and the income variable) have the expected positive significant effect. The education variable is positive as well.¹⁴ No significant relationship was found when regressing on age and age squared. However, when constructing dummy variables for 10-year age

14. There is a positive correlation between income and education – but leaving the education variable out of the regression does not change the overall results in terms of significant variables.

Table 5. Regressions results (log-log regression).

	Expected sign	lnWTP3		lnWTP6	
		Coeff.	Std. error	Coeff.	Std. error
Male	+/-	-0.004	0.089	0.031	0.088
Age-group 30-39	+/-	0.171	0.126	0.258*	0.126
Age-group 50-59	+/-	0.056	0.125	0.078	0.124
Age-group 60-69	+/-	-0.085	0.140	-0.094	0.139
Serious health symptoms	+/-	0.099	0.110	0.093	0.109
House	+	0.285***	0.107	0.333***	0.106
Children	+	-0.077	0.103	-0.102	0.103
Concern for environment	+	0.535***	0.091	0.557***	0.091
Higher education	+/-	0.328***	0.090	0.355***	0.090
log(incomeh)	+	0.400***	0.095	0.400***	0.094
Constant		-1.52	0.954	-1.52	0.954
<i>N</i>		1010		1010	
<i>R</i> ²		0.089		0.101	

*, **, *** Significant at 0.1, 0.05 and 0.01 levels, respectively. Income is defined as household or individual income depending on the answer to the debriefing question.

groups the individuals in the age group 30-39 state a significantly higher (at a 10% level) *WTP* for a six-month gain than the individuals aged 40 to 49. Since the respondents in the 30-39 year old group will be paying for a longer life-time, this relationship will be strengthened when aggregating the values into the total *WTP* for the respondents' remaining life-time. With regards to health different proxies for health status have been included but no significant relationship was found. For altruism, the dummy variable for children has a negative insignificant sign, whereas the dummy variable for environmental concern is highly significant with a relatively large co-efficient. Finally, in this survey the internal scope test is passed since a paired one-sided *t*-test rejects that the *WTP* for three months and six months are identical ($p = 0.00$).

5. Discussion

The mean *WTP* values for a change in life expectancy of three and six months, respectively have, in this study, been found to be in the range of EUR 60-75 per month and the corresponding value for a life-year lost has been calculated and found to be in the region of EUR 30,000-90,000. The results fulfil the expected theoretical relationships, the internal scope test is fulfilled and the values are within the range of the values found in other published studies (the direct and indirect approaches).

The VOLYs found in this study are lower than the values found by using the indirect approach in Alberini et al. (2006) (EUR 163,000) and some explanations for this discrepancy will be offered here. In the present study a three-month and a six-month gain

in life expectancy has been valued whereas in the study by Alberini et al. (2006) the risk reduction corresponds to a calculated gain in life expectancy of 1.23 months. Given the size of the good in the present survey one could suspect that the benefit is no longer considered to be marginal and hence the budget constraint could constrain the respondent when assigning value to the six-month gain. Secondly, a non-linear relationship between *WTP* and the magnitude of the reduction in the probability of dying (i.e. a gain in life expectancy) could be expected due to the decreasing marginal value of a change in the probability of dying.

The internal scope test is fulfilled and the ratio between *WTP* for a three and six-month gain is equal to 1.2. The arguments presented above concerning decreasing marginal value and non-marginal goods could potentially also offer explanations for a scope ratio different from two. Still, a scope ratio different from two entails that the aggregated *VOLY* will be very dependent on whether the *WTP* for three months or six months will be taken as the starting point. The author refers to Nielsen (2008) for a further discussion on scope and order biases in the present survey.

In the piloting for the NEEDS survey as well as in the study by Chilton et al. (2004) it was demonstrated that according to a substantial number of respondents, three months is the smallest good »worth buying« (accordingly, the utility function for these respondents does not go through the origin). Since a smaller good often has been valued in *VSL* surveys this illustrates that the two concepts (a change in the probability of dying and a change in life expectancy) are perceived differently. This has also been illustrated in studies by McNeil et al. (1982); Morris & Hammitt (2001). The fact that the two concepts are perceived differently highlights the problem inherent in the indirect approaches mentioned in section 2.2. Moreover, a theoretical issue is that the change in the survival curve would have a very different shape according to whether the risk reduction is a one-year change in the probability of dying (the *VSL* framework) or a permanent change (the framework applied in this survey), Johannesson et al. (1997). Therefore, even with an identical gain in life expectancy, the distribution of the gain (the changes in the conditional survival probabilities) would be very different in the two frameworks. The respondent's response to different distributions of the life expectancy gain is a much underexposed area in the theoretical literature as well as in the empirical surveys, and the author wishes to draw attention to this point. More attention here could be of assistance in theoretical as well as empirical investigations of the relationship between the indirect and direct approaches.

Even though the values found in this survey are lower than values found by means of the indirect approach, the values seem relatively large compared with the values found using the direct approach. However, there are some important differences between the studies that limit their comparability and hence the *WTP* estimates in

Table 1 have not been converted into *VOLYs*. The study by Chilton et al. (2004) is the most obvious choice for comparison because of the similar timing of the payment (the rest of the individual's life) and the same context. Still, there is some dissimilarity since in the UK survey respondents are asked to come up with a household *WTP* and they are asked to value a gain in life expectancy as part of a bundle of other goods. Taking into account that the UK values are *WTP* for the household as a whole, the Danish values are much higher than the UK values.

In this particular survey there could be some potential issues with regards to the contextual dependence. The respondents in this survey were asked to value only a gain in their personal life expectancy and not to think about the change in air pollution when stating their *WTP*. The values found in this survey are meant to form the basis for further increasing the validity of the results from the impact-pathway method. Hence, inclusion of the context is important since potential preferences for context need to be taken into account. It is crucial to establish whether the respondents have succeeded in isolating the change in life expectancy from the benefit of improved air quality in their valuation task. Since one could expect that people expressing an interest for the environment would find this distinction more difficult the variable »environmental interest« was included in the regressions as an indicator of commitment. This is a highly significant variable with a high coefficient and, as argued in section 3.3., could prove to be a considerable problem for the utilitarian framework since instead of indicating a real value for the different changes in life expectancy, respondents may indicate their »moral satisfaction« from contributing to a public good (air pollution), Kahneman & Knetsch (1992). This way of reasoning is supported by the fact that the segmentation is based on the GallupKompas and a distinction between the individualistically oriented and the socially orientated segments. Interest in the environment is found in the socially orientated segments who, among other things, are very interested in social welfare and collective behaviour. Accordingly, one could have some doubts as to whether the results in this survey are actually *WTP* for a gain in life expectancy or the *WTP* for a change in air pollution or a combination of the two.¹⁵ Future research should also seek to investigate the possibilities for valuing a gain in life expectancy in a context-free setting.

6. Conclusion

This paper adds to the limited number of empirical surveys on the topic of *WTP* for

15. However, another argument could be that the proxy of interest in the environment is an index of the strength of the individual's environmental preferences, which correlates with the environmental benefits that a reduction in air pollution gives the individual. Hence, the reason for the higher values from the individuals indicating 'environmental interest' is not commitment but that these people would feel personally better off if life expectancy was increased in the context of air pollution.

a gain in life expectancy. The paper presents results from a Danish web-based CVM survey in an air pollution context. The mean *WTP* values found for a gain in life expectancy of three and six months, respectively have been found to lie in the range of EUR 60-75 per month. The values have been tested with regards to theoretical validity and the corresponding values for a life year lost have been calculated and found to lie in the area of EUR 30,000-90,000. These values have been compared with other values from different approaches and have been found to be within the range of previous published studies. However, caution is needed in interpreting the results due to potential contextual issues. Three approaches have been proposed for the estimation of VOLYS; two indirect and one direct. This paper uses the direct approach and the author would argue that this is the best approach to use, at least until the indirect approaches have been validated in terms of theory and tested empirically. Given that in the existing analyses of externalities of energy production changes in mortality constitute by far the largest share of health damage estimates in terms of morbidity and mortality, Andersen et al. (2004), research is urgent in order to investigate comparability and the theoretical link between VOLY and VSL further.

Acknowledgment

The author has benefited greatly from discussions with Trine Kjær and Kjeld Møller Pedersen (University of Southern Denmark) and Mikael Skou Andersen and Fleming Møller (National Environmental Research Institute, University of Aarhus). This work was supported by the Health Insurance Foundation in Denmark and the Danish Centre of Excellence, AIRPOLIFE (Air Pollution in a Life Time Health Perspective). The development of the questionnaire was financed by the EU Commission under the 6th Framework Programme. The author is grateful for helpful comments and suggestions from two anonymous referees.

Literature

- Abelson, P. 2003. The Value of Life and Health for Public Policy. *Economic Record*, 79, S2-S13.
- Alberini, A., A. Hunt, and A. Markandya. 2006. Willingness to pay to reduce mortality risks: Evidence from a three-country contingent valuation study. *Environmental & Resource Economics*, 33, 251-64.
- Andersen, M. S., L. M. Frohn, S. S. Jensen, J. S. Nielsen, P. B. Sørensen, O. Hertel, J. Brandt and J. Christensen. 2004. Sundhedseffekter af luftforurening – beregningspriser. *Faglig rapport fra DMU* nr. 507.
- Andersen, M. S., L. M. Frohn, J. S. Nielsen, M. Nielsen, J. B. Jensen, S. S. Jensen, J. Christensen and J. Brandt. 2008. A non-linear Eulerian approach for assessment of health-cost externalities of air pollution. 16th annual conference of the European association of environmental and resource Economists (EAERE), June 25-28.
- Bateman, I. J., R. T. Carson, B. Day, M. Hanemann, N. Hanley, N., T. Hett, M. Jones-Lee, G. Loomes, S. Mourato, E. Özdemiroglu, D. W. Pearce, R. Sugden and J. Swanson. 2002. *Economic Valuation with Stated Preference Techniques. A Manual*,

- Edward Elgar, Cheltenham.
- Brunekreef, B. 1997. Air pollution and life expectancy: is there a relation? *Occupational and environmental medicine*, 54, 781-84.
- Carson, R. T. 1997. Contingent Valuation Surveys and Tests of Insensitivity to Scope. In: R. J. Kopp, W. Pommerhene, and N. Schwartz (Editors). *Determining the Value of Non-Marketed Goods: Economic, Psychological, and Policy. Relevant Aspects of Contingent Valuation Methods*. Kluwer, Boston.
- Carson, R. T., N. E. Flores and N. F. Meade. 2001. Contingent Valuation: Controversies and Evidence. *Environmental and Resource Economics*, 19, 173-210.
- Chilton, S., J. Covey, M. W. Jones-Lee, G. Loomes and H. Metcalf. 2004. *Valuation of health benefits associated with reductions in air pollution*. Final Report, Department for Environment, Food and Rural Affairs.
- Collett, D. 1991. *Modelling Binary Data*, Chapman and Hall, London, UK.
- Cropper, M. L. and P. R. Portney 1990. Discounting and the Evaluation of Lifesaving Programs. *Journal of Risk and Uncertainty*, 3, 369-79.
- Cropper, M., and F. G. Sussman. 1990. Valuing future risks to life. *Journal of Environmental Economics and Management*, 160-74.
- Desaigues, B., A. Rabl, D. Ami, B. M. Kene, S. Masson, M.-A. Salomon, L. Santoni. 2007a. Monetary value of a life expectancy gain due to reduced air pollution: Lessons from a contingent valuation in France, *Revue d'economie Politique*, vol. 117, no. 5, pp. 675-98.
- Desaigues, B., D. Ami, M. Hutchinson, A. Rabl, S. Chilton, H. Metcalf, A. Hunt, R. Ortiz, S. Navrud, P. Kaderjak, R. Szanto, J. S. Nielsen, C. Jeanrenaud, S. Pelligrini, M. B. Kohlova, M. Scasny, V. Maca, J. Urban, M. E. Stoeckel, A. Bartczak, O. Markiewicz, P. Riera, V. Farreras. 2007b. *Final Report on the monetary valuation of mortality and morbidity risks from air pollution*. NEEDS RS1b. Deliverable D6.7
- Drèze, J. H., 1962. L'Utilite Sociale D'une vie Humaine. *Revue française de recherche opérationnelle*, 6:93-118.
- Finansministeriet, 1999. *Vejledning i udarbejdelse af samfundsøkonomiske konsekvensvurderinger*.
- Gafni, A. and G. W. Torrance. 1984. Risk attitude and time preference in health. *Management Science*, 30,4, 440-51.
- Gyrd-Hansen, D. and J. Sogaard. 1998. Discounting life-years: Whither time preference? *Health Economics*, 7, 121-27
- Hammit, J. K. 2000. Valuing mortality risk: Theory and practice. *Environmental Science and Technology*, 34, 1396-1400.
- Hammit, J. K. and J.-T. Liu. 2004. Effect of Disease Type and Latency on the Value of Mortality Risk. *Journal of Risk and Uncertainty*, 28, 73-95.
- Hanemann, M. W., 1994. Valuing the Environment Through Contingent Valuation. *The Journal of Economic Perspectives*, 8:19-43.
- Hirth, R. A., M. E. Chernew, E. Miller, A. M. Fendrick and W. G. Weissert, 2000. Willingness to Pay for a Quality-adjusted Life Year: In Search of a Standard. *Medical Decision Making*, 20, 332-42.
- Holland, M., J. Berry and D. Forster, 1998. ExternE Externalities of Energy. Volume 7: Methodology Update. European Commission Directorate-General XII Science, Research and Development.
- Johannesson, M. and P. O. Johannsson, 1996, To Be, or Not to Be, that is the Question: An Empirical Study of the WTP for an increased Life Expectancy at an advanced Age: *Journal of Risk and Uncertainty*, 13:163-74.
- Johannesson, M., P. O. Johannsson, and K.-G. Löfgren. 1997. On the Value of Changes in Life Expectancy: Blips Versus Parametric Changes. *Journal of Risk and Uncertainty*, 15:221-39.
- Johannesson, M. and D. Meltzer. 1998. Some reflections on cost-effectiveness analysis. *Health Economics*, 7:1-7.
- Johansson, P. O. 2002. On the definition and Age-Dependency of the Value of a Statistical Life. *Journal of Risk and Uncertainty*, 25, 251-63.
- Jones-Lee, M. W. 1976. *The Value of Life. An economic analysis*, Martin Robertson &

- Co. Ltd, London.
- Jones-Lee, M. W. 1989. *The Economics of Safety and Physical Risk*, Basil Blackwell, Oxford.
- Jones-Lee, M. W. 1991. Altruism and the Value of Other People's Safety. *Journal of Risk and Uncertainty*, 4, 213-19.
- Jones-Lee, M. W. 1992. Paternalistic altruism and the value of statistical life. *Economic Journal*, 102, 80-90.
- Kahneman, D. and J. Knetsch. 1992. Valuing Public Goods, The Purchase of Moral Satisfaction. *Journal of Environmental Economics and Management*, 22, 57-70.
- Krupnick, A., A., Alberini and M. Cropper. 1999. *Mortality Risk Valuation for Environmental policy. Resources for the Future*.
- McNeil, B. J., S. G. Pauker, H. C. Sox and A. Tversky. 1982. On the elicitation of preferences for alternative therapies. *New England Journal of Medicine*, 306, 1259-62.
- Mishan, E. J, 1971. Evaluation of Life and Limb: A Theoretical Approach. *The Journal of Political Economy*, 79:687-705.
- Morris, J. and J. K. Hammitt. 2001. Using Life Expectancy to Communicate Benefits of Health Care Programs in Contingent Valuation Studies. *Medical Decision Making*, 21, 468-78.
- Møller, F., S. P. Andersen, P. Grau, H. Huusom, T. Madsen, J. Nielsen, L. Strandmark. 2000. *Samfundsøkonomisk vurdering af miljøprojekter*. Danmarks Miljøundersøgelser, Miljøstyrelsen og Skov- og Naturstyrelsen
- Møller, F. 2009. *Velfærd nu eller i fremtiden. Velfærdsøkonomisk og nytteetisk vurdering over tid*. Aarhus Universitetsforlag
- Nielsen, J. S. 2008. Valuing gains in life expectancy. Theoretical and empirical issues. *PhD thesis*, University of Southern Denmark & National Environmental Institute, University of Aarhus.
- Pindyck, R. S. and D. L. Rubinfeld. 1998. *Econometric models and economic forecasts*, 4th edition. Irwin McGraw-Hill, Singapore.
- Pope, C. A., R. T. Burnett, M. J. Thun, E. F. Calle, D. Krewski, K. Ito and G. D. Thurston. 2002. Lung Cancer, Cardiopulmonary Mortality, and Long-term Exposure to Fine Particulate Air Pollution, *Journal Of the American Medical Association*, 2002, 1132-41.
- Ready, R., S. Navrud, B. Day, R. Duborg, F. Machado, S. Mourato and M. Rodriguez. 2004. How reliable are transfers between countries? *Environmental and Resource Economics*, 29, 67-82.
- Rosen, S. 1988. The Value of Changes in Life Expectancy. *Journal of Risk and Uncertainty*, 1,285-304.
- Schelling, T. C. 1968. The Life You Save May Be Your Own. In: SB Chase (Editor), *Problems in Public expenditure Analysis*. The Brookings Institution, Washington D.C, 127-66.
- Sen, A. 1977. Rational Fools: A Critique of the Behavioral Foundations of Economic Theory. *Philosophy and Public Affairs*, 6, 4, 317-44.
- Shepard D. S. and R. Zeckhauser. 1982. Lifecycle consumption and willingness to pay for increased survival. In: M.W.Jones-Lee (Editor), *The Value of Life and Safety*. Collection of Papers presented at the Geneva conference on The Value of Life and Safety held at the University of Geneva 30th, 31st March and 1st April, 1981. North-Holland Publishing Company, 95-145.
- Shepard, D. S. and R. Zeckhauser. 1984. Survival versus consumption. *Management Science*, 30, 423-39.
- Statistics Denmark. 2007. Statbank Denmark.
- Statistics Denmark. 2004. *Statistical Yearbook 2004* (Statistisk Årbog).

Appendix

Ability to survive

(Example of someone who is aged 40 now)

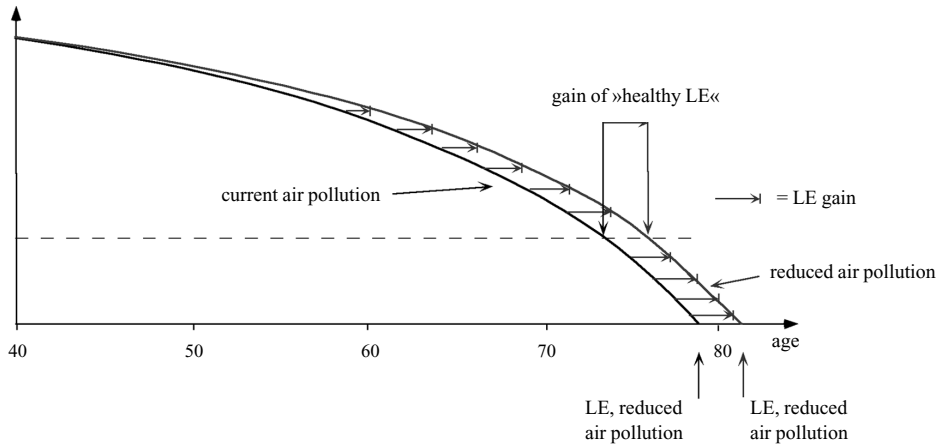


Figure A1. The survival curve shown to the respondents in the questionnaire.

Question number 11 and 12 in the questionnaire

Now, please think about the maximum amount you would be willing to pay each year. It may be one of the figures on the cards or something in between. Please consider only your personal gain in life expectancy:

What is the maximum amount you are willing to pay in the form of higher expenses to gain an average of 3(6) months in your life expectancy?

Maximum amount I would be willing to pay each year DDK. _____

International Trade in Services – Evidence from Danish Micro Data

Vibeke Borchsenius

Centre for Economic and Business Research (CEBR), Copenhagen Business School,
E-mail: vbo@forsikringogpension.dk

Nikolaj Malchow-Møller

Department of Business and Economics, University of Southern Denmark, E-mail: nmm@sam.sdu.dk

Jakob Roland Munch

Department of Economics, University of Copenhagen, E-mail: Jacob.Roland.Munch@econ.ku.dk

Jan Rose Skaksen

Department of Economics, Copenhagen Business School, E-mail: jrs.eco@cbs.dk

SUMMARY: International trade in services is a growing phenomenon, but data in this area trail far behind those for international trade in goods. This is partly due to the fact that trade in services is less tangible and more diversified than trade in goods. This article exploits unique Danish micro data from different sources to construct a detailed picture of which services are currently exported from Denmark. Furthermore, we compare those services that are currently exported to those that are potentially tradable. This comparison provides an indication of the unexploited potential for export of services, and hence also indicates where different »barriers« (such as trade costs, trade restrictions and in practice also differences in demand) to international trade in services currently exist.

1. Introduction

International trade in services is becoming increasingly important. According to the IMF Balance of Payments Statistics, global trade in services has grown more rapidly than international trade in manufactured goods in recent decades. However, data on international trade in services are still much scarcer and less detailed than data on international trade in manufactured goods; see, e.g., Lipsey (2006). This has partly to do with the intangible nature of many services, and the different ways in which services can be traded between countries. As opposed to trade in goods, the export of services

Funding for this project from the Economic Policy Research Network (EPRN) is gratefully acknowledged.

rarely requires physical goods to cross the customs border between countries. Instead, some services can be delivered by e-mail or through telephones while other services require the supplier to produce the service at the location of the customer.

The lack of information on international trade in services is probably the main reason why there has only been very little empirical research on international trade in services; see, *e.g.* Lipsey (2006). With respect to trade in goods, there is a long tradition for empirical research ranging from the extent and implications of trade policy, see, *e.g.* Feenstra (1995), to the productivity of exporting firms relative to non-exporting firms; see, *e.g.*, Bernard and Jensen (1995). The lack of knowledge about international trade in services is an increasing problem for at least two reasons. First, in developed countries an increasing share of production takes place within services, and an increasing share of international trade occurs within services. Second, there are only few restrictions on international trade in goods. If substantial further gains are to be realised from increasing international trade, these gains are therefore most likely to arise from trade in services. To achieve this, a better knowledge of the implications of different types of restrictions on international trade in services is required.

In this paper, we exploit information in Danish registers to construct unique micro data on services exported (and imported) by Danish firms in recent years. We exploit from the VAT register in Statistics Denmark, where we have information on total exports of goods and services at the firm level, and from the External Trade Statistics and the VAT register, we can also construct detailed measures of exports of goods at the firm level. Combining these data sources allows us to derive firms' exports of services as the residual.

To the best of our knowledge only one other study, Breinlich and Criscuolo (2009) is able to apply information on firm-level trade in services. Using data from various sources for a large number of British firms in the period 2000-2005, Breinlich and Criscuolo (2009) focus on how services traders differ from non-traders with respect to outcomes such as size, productivity, capital intensity and wages. As in the literature on firm-level trade in goods, they find significant exporter premiums in these dimensions; see also Bernard and Jensen (1995), Wagner (2007) and Munch and Skaksen (2008).

In the present paper, we use register information to construct firm-level measures of exports for the entire population of Danish firms. This in turns allows us to construct a detailed picture of which services are currently exported from Denmark, and to identify sectors with unexploited export potential.

More precisely, we show in this paper that our firm-level measures of trade in services are consistent with the more aggregate data from the balance-of-payments (BoP) statistics. This is a strong indication that the data provide relatively precise measures of international trade in services at the firm level. We also illustrate how these data al-

low us to calculate exports of services at, *e.g.*, a detailed industry level, whereas existing aggregate statistics only report exports of services on relatively aggregate categories, and not for separate industries. Our calculations reveal, among other things, that while there is a lot of international trade in goods within service industries, there is only very little trade in services within manufacturing industries.

Finally, we apply the data on exports of services to perform a simple analysis to identify service industries in which Danish producers do not (yet) fully exploit the potentials for export, and hence where potential »barriers« to international trade in services still exist. In this context, barriers should be interpreted relatively broadly, and may cover different impediments to international trade such as high trade (transportation) costs, formal or informal trade restrictions (*e.g.* standards), and even differences in demand can be interpreted as a kind of »barrier«, as some services may not be demanded in other countries/cultures.

The character of trade costs associated with services is very different from the character of those associated with manufactured goods, and trade costs also vary considerably across different types of services. Services that can be shipped by e-mail, such as software, are thus associated with very low trade costs, while other services are associated with extremely high trade costs as they require the temporary movement of persons across borders. This in turn makes some services intrinsically more tradable than others, although technological innovations, such as the internet, have lowered the trade costs of some services significantly over time.

International trade in services may also be hampered by trade restrictions which may be very different from trade restrictions on manufactured goods. Thus, international trade in services has been limited mostly by non-tariff barriers such as local standards and restrictions on temporary migration. Moreover, it is very complex to identify these barriers, as they often serve other purposes than just restricting import competition; see, *e.g.*, Copeland and Mattoo (2008).

One way of identifying existing barriers to international trade in services is to compare those services that are currently exported to those that are potentially tradable. In this paper, we identify the potentially tradable service industries using a method developed by Jensen and Kletzer (2006) which compares the geographical concentration of production and demand for an industry within a country. A related method was applied by Malchow-Møller *et al.* (2007) in an analysis of the tradability of construction services within Denmark. Comparing the tradability of different service industries with their actual exports of services derived from our firm-level data, provides us with an indication of the unexploited potential for export of services and hence indicates where barriers to international trade in services currently exist.

Our approach is also related to a couple of recent studies in the literature which use gravity regressions to estimate the potential trade flows for the more aggregate service

categories from the balance-of-payments statistics using aggregate bilateral trade data. By comparing these potential flows with actual trade flows, estimates of (the tariff equivalent of) barriers to trade can be constructed; see Francois and Hoekman (2009) or Brandicourt *et al.* (2008) for more details.¹

Our analysis reveals a robust relationship between the tradability of a service industry and its actual international exports of services. However, within »IT-services (excluding software)« and to a smaller extent within »software« and »advertising« we find signs of an unexploited export potential. For comparison, Francois *et al.* (2008) report – based on the gravity-regression approach – that within the EU, the largest barriers to trade are to be found within communication services.

The rest of the paper is structured as follows. In Section 2, we first discuss the distinction between goods and services, and we explain the different ways in which international trade in services can take place. In Section 3, we consider the aggregate development in the Danish international trade in services according to the official balance-of-payments statistics. In Section 4, we derive our firm-level measures of trade in services that allow for a more detailed account of the Danish export of services. In Section 5, we identify the potentially tradable service industries and compare those to the numbers for actual exports from Section 4. Finally, Section 6 concludes.

2. Basic Concepts

In this section, we first discuss how services are defined and distinguished from manufactured goods both at a conceptual level and in practice, *i.e.*, in the statistics. We then present the different modes through which international trade in services may occur according to the General Agreement on Trade in Services (GATS), and we discuss how this corresponds to the way in which international trade in services is currently measured in the statistics.

2.1 Services vs. goods

Services are difficult to define precisely. The »Manual on Statistics of International Trade in Services« (MSITS) – a manual developed and published in 2002 jointly by the United Nations (UN), the European Commission (EC), the International Monetary Fund (IMF), the Organisation for Economic Co-operation and Development (OECD), the United Nations Conference on Trade and Development (UNCTAD), and the World Trade Organization (WTO) – uses the following definition of services:²

1. Another approach used to identify barriers to international trade in services has been to construct different openness indices through collection of information on applied policies; see Francois and Hoekman (2009).

2. This definition is basically taken from the System of National Accounts 1993, which was produced jointly by the OECD, the United Nations Statistical Division, the International Monetary Fund, the World Bank and the Commission of the European Communities, and is a conceptual framework that sets the international statistical standard for the measurement of the market economy; see <http://unstats.un.org/unsd/sna1993>.

»Services are not separate entities over which ownership rights can be established. They cannot be traded separately from their production. Services are heterogeneous outputs produced to order and typically consist of changes in the condition of the consuming units realised by the activities of the producers at the demand of the customers. By the time their production is completed they must have been provided to the consumers.« [MSITS, p. 7].

As opposed to goods, services thus cover a heterogeneous group of products and activities, which are often intangible, invisible and perishable, requiring simultaneous production and consumption, and hence cannot be traded like goods; see also Copeland and Mattoo (2008). Examples include: cleaning, transport, and education.

In practice, there are a number of exceptions from this definition. One example is the production of software which is considered a service activity despite the fact that software may be stored on, *e.g.*, harddisks and CDs and traded like goods. Other examples are movies and music. More generally, the MSITS classifies all output in sections 5-9 of the Central Product Classification (CPC), version 1 as services.³

Instead of distinguishing at the product/activity level, a distinction between goods and services can be made at the firm or industry level. Productive units are thus typically allocated to an industry in the International Standard Industrial Classification of all Economic Activities (ISIC) or the Classification of Economic Activities in the European Community (NACE) according to their main economic activity.⁴ Hence, firms belonging to industries considered as service industries in the ISIC/NACE can then be considered as service producers. However, it is important to bear in mind that many firms in the manufacturing industries produce both goods and services; and similarly for firms in the service industries.

2.2 Trade in Services

International trade in goods requires the shipment of physical goods across borders. International trade in services is different for at least two reasons. First, some services require physical proximity between the supplier and the consumer, as in the case of, *e.g.*, haircuts, cleaning and construction services, and hence cannot be shipped across borders. Second, among the services that can be shipped from the supplier to the customer, many of these are intangible, as in the case of many financial services. Hence, the trade does not involve physical units passing the borders.

3. The CPC can be found on: <http://unstats.un.org/unsd/cr/registry/>. Sections 5-9 are: (5) Intangible assets; land; constructions; construction services; (6) Distributive trade services; lodging; food and beverage serving services; transport services; and utilities distribution services; (7) Financial and related services; real estate services; and rental and leasing services; (8) Business and production services; and (9) Community, social and personal services.

4. The ISIC can be found on: <http://unstats.un.org/unsd/cr/registry/>, and the NACE on <http://ec.europa.eu/eurostat>.

As a consequence, GATS has defined four different *modes* of trade in services – based on the location of both the supplier and the consumer of the traded service:

- Mode-1 trade (*cross-border supply*) takes place when the supplier and the consumer remain in their respective countries of residence, and the service is shipped across national borders. Although, this is the type of trade in services that resembles trade in goods most closely, there is often no package that passes the customs frontier. Instead, the shipment can take place over the internet, via telephones, via fax machines, *etc.* Examples include financial services, computer software, and call-centre services.
- Mode-2 trade (*consumption abroad*) takes place when the consumer moves temporarily to the country of residence of the supplier to consume the service. This is the case of, *e.g.*, tourism and education services, but the repair of ships in foreign ports also constitutes an example of mode-2 trade.
- Mode-3 trade (*commercial presence*) covers trade in services through a commercial presence of the supplier in the country of the customer. It thus covers the cases where it is necessary to establish a commercial presence abroad, *e.g.*, a locally established subsidiary or branch of the company, through which the supplier can deliver her services. Consultancy services sold through a local division of a multinational firm thus fall within this category.
- Mode-4 trade (*presence of natural persons*) takes place when an individual resident in the exporting country moves temporarily to the country of the consumer to provide the service. The service can be provided in two conceptually different ways. First, the individual can deliver a service to a customer through a *service transaction* on behalf of the individual's employer in the exporting country or on behalf of the individual if she is herself. This is the case for many consultancy and construction services. Second, the individual can take up *temporary employment* in a (service) firm in the host country and deliver the service through her employment in that firm. In that case, the payment for the service is the wage received by the individual.

Thus, according to the GATS definition, international trade in services is not so much about units passing the borders, but about service transactions between customers resident in one country and suppliers resident in (or controlled by residents in) another country.

Although the distinction between the four modes may seem conceptually clear, reality is less so. First, many service transactions in practice involve more than one mode. Thus, some business services may mainly be delivered through mode 1, but may

sometimes require the company to temporarily visit the customer's country, in which case mode 4 is also involved. Second, other transactions may involve both goods and services, as when buying a new IT system that also includes on-site installation by the supplier. Third, while services delivered through temporary movement of persons belong to mode 4, it is unclear what it takes before this turns into a commercial presence and instead becomes mode-3 trade. In GATS, a commercial presence can thus also cover entities created for a short period. Finally, mode-4 trade through employment covers only non-permanent employment in the country of the consumer, but again it is not clear where to draw the line between »non-permanent« and »permanent« employment.

Thus, determining actual trade flows within each of the four modes – and in total – is difficult, and the data required are therefore not available at the moment. The MSITS provides a detailed discussion of the difficulties involved using currently available data. It also provides a number of recommendations for future data collection.

The Balance of Payments (BoP) is currently the main source of data on international trade in services in most countries. In the BoP, international trade in services is split into 11 different service categories (Transportation, Travel, Communication services, Construction services, Insurance services, Financial services, Computer and information services, Royalties and license fees, Other business services, Personal, cultural and recreational services, and Governmental services) following the recommendations in IMF's Balance of Payments Manual (BPM5).⁵ Together, these 11 categories should cover modes 1, 2 and that part of mode 4 that takes place through service transactions. But a given transaction can still belong to more than one mode, *cf.* the examples above. In the MSITS, an attempt is made to allocate the first 10 categories across the four different modes of trade, with some categories belonging to more than one mode. As an example, »Construction services« are partly mode 3 and partly mode 4, depending on whether an office is established in the customer's country.

Mode-4 trade through temporary employment in firms in the host countries is not accounted for by these 11 categories. However, if individuals work for less than a year in a country in which they are not residents, their income is recorded in the Balance of Payments under »Compensation of employees«. Hence, one can use this as an approximate measure of the part of mode-4 trade that takes place through temporary employment.

Finally, services traded under mode 3, *i.e.*, services sold by a local subsidiary or a branch of the company, are not counted as part of the BoP. Information about this type of trade typically requires rather detailed information about the activities of foreign affiliates – information which is currently not available.

5. See <http://www.imf.org/external/pubs/ft/bopman/bopman.pdf>.

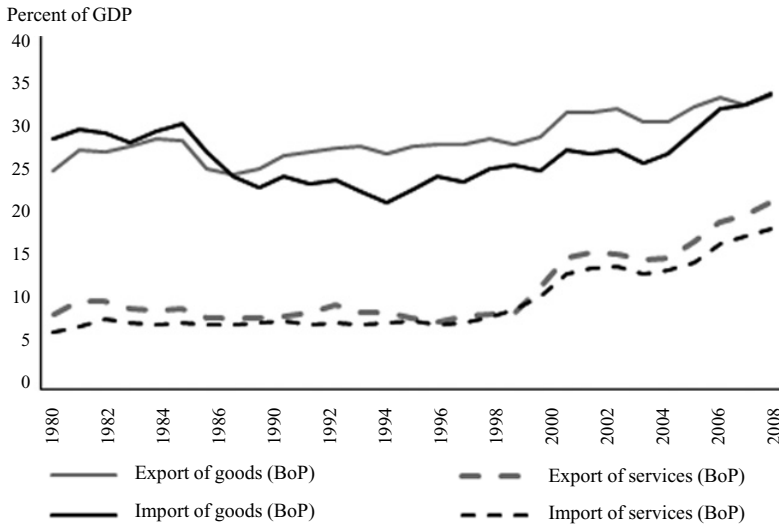


Figure 1. The Danish International Trade in Goods and Services (Balance of Payments Statistics), 1980-2008.

Source: The Balance of Payments Statistics (BET6 and BOP6) and the national accounts (NAT01) from Stat Bank Denmark, Statistics Denmark.

Because of the lack of data on mode-3 trade, we shall focus exclusively on modes 1, 2 and 4 in the rest of the paper. In the next section, we briefly present the aggregate numbers according to the BoP. When we turn to micro-level evidence in section 4, the focus is on mode-1 trade and the part of mode-4 trade that takes place through service transactions, as this is the type of service export that is recorded at the firm level.

3. The Aggregate Development

In this section, we briefly present the aggregate development in trade in services for Denmark using data from the Balance-of-Payments Statistics from Statistics Denmark. Like other countries, Denmark follows the guidelines in the MSITS and uses the Extended Balance of Payments Services Classification (EBOPS), which is an elaboration of the BPM5 classification from above. Within the 11 main categories of BPM5, trade in services is divided into subcategories. As mentioned above, these categories cover Modes 1, 2 and the part of 4 that takes place through service transactions.

The data are collected through reporting from the population of firms. Each month, approximately 350 of the largest service firms report to Statistics Denmark. Together these firms cover around 70% of the international trade in services. This information is supplemented by quarterly information from 1,200 randomly selected small and medium-sized firms, representing the more than 20,000 firms that are responsible for

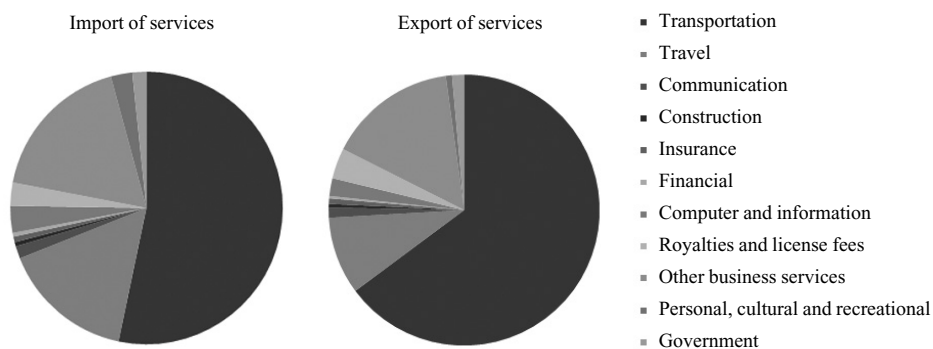


Figure 2. The Composition of Danish Exports and Imports of Services (Balance of Payments Statistics), 2008.

Source: The Balance of Payments Statistics (BOP6) from Stat Bank Denmark, Statistics Denmark.

the final 30% of the trade. With respect to the »Travel« category, the firm information is supplemented by information from, *e.g.*, interviews with foreigners visiting Denmark and Danish households travelling abroad. See Statistics Denmark (2007) for more details.

The aggregate development in Denmark's international trade in goods and services is illustrated in Figure 1. The figure shows that over the last 10-15 years, exports and imports of services have doubled their share in GDP. However, Lipsey (2006) argues that part of the observed increases in the importance of trade in services for many countries may reflect better data collection rather than new activities.⁶

Data on the distribution of Danish exports and imports by the 11 main categories from BPM5 are only available from 2005 onwards. Figure 2 shows the distribution in 2008. It is seen that the distribution is more or less the same for exports and imports of services. »Transportation« constitutes the largest shares in both cases. In fact, 65 percent of the total export of services is due to transportation services, while 53 percent of the total import of services is due to transport. »Travel« and »other business services« (including, *e.g.*, consulting services) are responsible for 9 percent and 15 percent, respectively, of exports of services, while they each account for 16-18 percent of the imports. »Financial services« and »construction services« are responsible for the smallest shares of both exports and imports of services as they each account for less than 0.5 percent of total exports and imports of services.

6. Note that in 2006, there are sizeable increases in both exports and imports of services. These are partly due to the acquisition of the British-Dutch company P.O. Nedlloyd by A.P. Møller – Mærsk. This increases exports, as the services of P.O. Nedlloyd are now credited Denmark. It also increases imports, however, as many of the ships are registered abroad, and hence the charter rent (the payment for the ships) is registered as an import of services; see DØRS (2006).

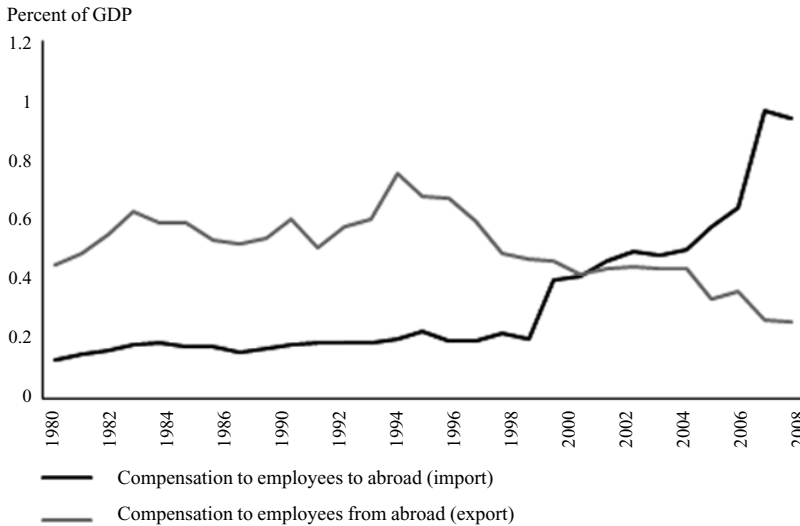


Figure 3. Compensation to employees (*Balance of Payments Statistics*), 1980-2008.

Source: The Balance of Payments Statistics (BET6 and BOP6) and the national accounts (NAT01) from Stat Bank Denmark, Statistics Denmark.

Figure 3 shows the development in the compensation to employees according to the BoP statistics. This measure is a proxy for the part of mode-4 trade that takes place through temporary employment in foreign firms. While compensation to abroad (import) is more or less constant from 1988 to 1998, it increases rapidly after 1998, rising to close to 1 percent of GDP in 2007-8. The compensation to employees from abroad, on the other hand, has exhibited a declining trend since the mid 1990s. Quantitatively, this type of trade in services, is thus less important than the trade depicted in Figure 2, *i.e.*, the trade through mode 1, mode 2 and the transactions part of mode 4.

4. Micro Evidence

In this section, we derive firm-level measures of exports (and imports) of services from micro data. The firm-level measures in turn allow us to calculate exports of services at, *e.g.*, a detailed industry level, whereas the aggregate statistics only report exports of services for relatively aggregate types of services and not by industries.

Our firm-level measure of exports of services will be derived as the difference between the firm's total exports (including both goods and services) and the firm's exports of goods only. The total exports are recorded at the firm-level for taxation purposes in the VAT register in Statistics Denmark. The data for trade in goods are based on information from the Danish External Trade Statistics at Statistics Denmark and the VAT register. In the same way, it is possible to derive imports of services at the firm level, by subtracting imports of goods from total imports.

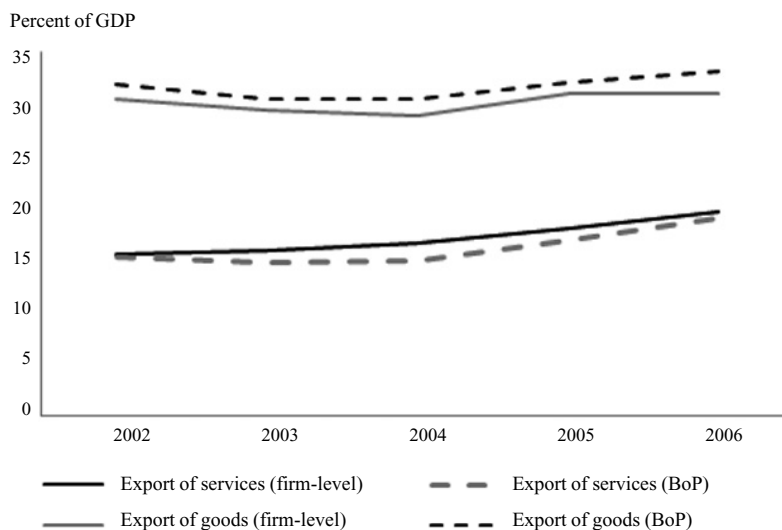


Figure 4a. The Danish International Export of Goods and Services (firm-level data and BoP data), 2002-2006.

Source: Own calculations on the External Trade Statistics and the VAT register from Statistics Denmark.

A couple of additional remarks about the construction of the service trade measure are in order. The External Trade Statistics are compiled in two systems: Intrastat (trade with EU countries) and Extrastat (trade with non-EU countries). The level of detail in this register is very high as trade flows are recorded by destination/origin country and an eight-digit product code. Trade flows in Extrastat are recorded by customs authorities, and the coverage rate is therefore close to complete. In contrast, the coverage rate in Intrastat is lower (around 90%), because some predominantly small firms appear not to report data to the system. Also, data on intra-EU trade is censored in such a way that only firms exporting goods with a total annual value exceeding a certain threshold are recorded in the files. Fortunately, the VAT register also records the total export of goods to EU countries and the total import of goods from EU countries from 2002 onwards, and the coverage rate here is higher. Hence, we should be able to calculate trade in services at the firm level fairly accurately for the years 2002 through 2006.

The numbers for trade in services derived in this way will in principle cover all mode-1 exports as well as the part of mode-4 exports that takes place via firms, *i.e.*, through service transactions. The import numbers will, of course, only cover that part of mode-1 and mode-4 imports that takes place via the firms.

Figures 4a and 4b show the aggregate development in exports and imports, respectively, of goods and services when calculated from firm-level information. For compa-

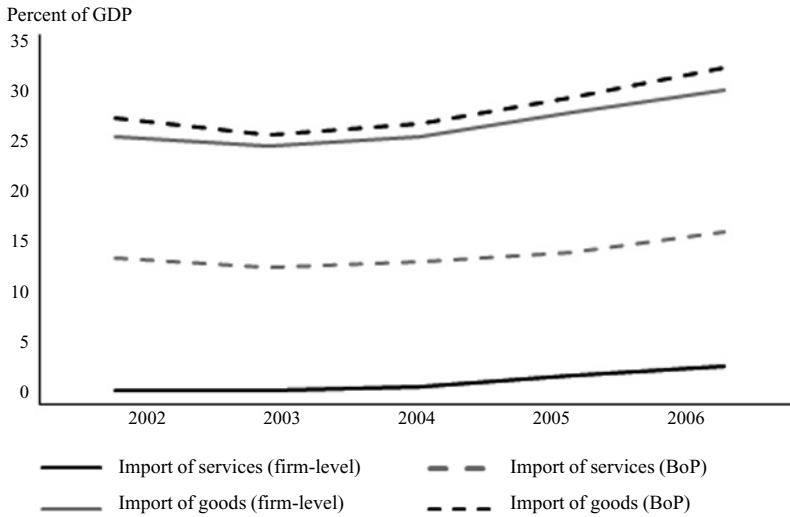


Figure 4b. The Danish International Import of Goods and Services (firm-level data and BoP data), 2002-2006.

Source: Own calculations on the External Trade Statistics and the VAT register from Statistics Denmark.

risson, we have also included the official aggregate numbers from the extended balance of payments statistics from Figure 1.

Consider first the exports in Figure 4a. It appears that the total level of exports of goods is slightly lower (1-2% of GDP) when using our firm-level information instead of the official aggregate numbers from the extended balance of payments statistics in Figure 2. This is probably due to the incomplete coverage rate in the firm-level statistics on goods trade as discussed above, and other technical corrections made in the aggregate numbers; see Statistics Denmark (2007).

When comparing the two series for exports of services, we should bear in mind that trade with mode-2 services is not included in the firm-level numbers. Since exports of travel services (mode 2) amount to approximately 1.5% of GDP in the aggregate numbers, we should expect (all else equal) the firm-level export numbers to undershoot the BoP numbers by this amount.

However, as the two series are in fact close to identical, another bias (of a similar size) must work in the opposite direction (provided that the BoP numbers measure the aggregate export of services correctly). This bias is likely to come from the fact that total exports at the firm-level also include some domestic sales that are not subject to VAT, *e.g.*, newspapers and transport of persons. As exports of services at the firm level are calculated as a residual by subtracting exports of goods from total exports, these domestic sales will be included there.

Turning to the import numbers in Figure 4b, the differences are larger. The total level of imports of goods is thus lower by 1.2-2.2% of GDP when using firm-level information instead of the official aggregate numbers from the extended balance of payments statistics in Figure 2. While this may again be due to incomplete coverage in the firm-level statistics, it may also be partly explained by the fact that imports of goods directly by consumers are not included in the firm-level numbers.

When it comes to the numbers for imports of services, the divergence between the two sources is substantial. There are several potential explanations for this. First, imports of mode-2 services (which are more important than exports) are not included in the firm numbers. Second, import directly by consumers is not included in the firm numbers either. Third, some mode-4 imports of services may not be registered as imports in the firms if the foreign supplier is required to pay Danish VAT. Fourth, imported transportation services may be registered as imports of goods at the firm level. The reason for this is that imports of goods are recorded as CIF (including Cost, Insurance, and Freight) in the trade statistics, thereby including the value of, *e.g.*, insurance and transportation services in the total imports of goods. In the BoP statistics the value of these services is estimated and subtracted from the total imports of goods, and added to the total imports of services, Statistics Denmark (2007). Exports of goods, on the other hand, are recorded as FOB (Free On Board), and insurance and transportation services will thus appear as export of services in this case – both at the firm level and in the BoP statistics.

In summary, we conclude that while there is a rather large divergence between the firm-level numbers and the BoP numbers on imports of services, the biases affecting the service export numbers appear to be relatively small (although not negligible). Due to this, we shall focus on the export part of the international trade in services in most of the subsequent analyses. We will be using data from 2005 as it turns out that the match between the VAT register, the External Trade Statistics and the General Firm Statistics – a match which is required in the following analyses – is much better in 2005 than in 2006.

Table 1 illustrates how much of the international trade in goods and services takes place within firms belonging to the manufacturing and service industries. Manufacturing industries are defined here as industries with NACE codes in the interval (150000, 400000), whereas service industries belong to the interval (400000, 750000). As can be seen from the table, international trade in goods is not limited to manufacturing firms. In fact, over 40 percent of total exports of goods and close to 75 percent of total imports of goods in these sectors are traded through firms in the service industry. We suspect that wholesale firms are responsible for a large part of that trade; but this is an issue that deserves closer analysis in future work. In the case of trade in services,

Table 1. Exports and imports of goods and services by manufacturing and service firms, 2005 (mil. DKK).

	Industries	
	Manufacturing	Service
Export of services	24,465	234,352
Export of goods	264,596	178,650
Import of services	6,296	25,521
Import of goods	101,448	295,264

Source: Own calculations on the General Firm Statistics, the External Trade Statistics and the VAT register from Statistics Denmark.

however, 80 percent of imports and over 90 percent of exports take place through firms in the service industries; making it reasonable to focus on these industries when considering the international trade in services.

In Table 2, the five service industries (defined at the 2-digit NACE level) with the highest service-export intensities (calculated as exports of services relative to total sales) are listed. The picture that emerges is consistent with the previous findings. The three industries with the highest service-export intensities are all related to the transport sector – with water transport as a clear outlier.⁷ Remember that transportation was responsible for around 65 percent of the total export of services according to the aggregate numbers in Figure 2. However, the »Research and development« industry also exhibits considerable service export.

Using a more detailed industry classification, *e.g.*, at the 4-digit NACE level, yields a very similar picture with the ten most service-export intensive industries found to be within the five 2-digit industries from Table 6 plus »Renting of machinery and equipment without operator and of personal and household goods« (NACE code 71) and »Wholesale trade and commission trade« (NACE code 51).

5. Tradable Services vs. Exported Services

As mentioned in the introduction, the trade costs associated with services are very different from those associated with manufactured goods, and they vary considerably across different types of services. Services traded through mode 1, which can be shipped by e-mail or via downloads from the internet, such as software, are typically easier to trade than most manufactured goods which require a physical package to be shipped. On the other hand, mode-4 trade in services is typically associated with higher

7. Remember, however, that the export intensities in the transportation sector may be biased upwards because domestic sales that are not subject to VAT (this includes transportation of passengers) are included in the measure of total exports at the firm level.

Table 2. Top-5 Service-export industries (2-digit NACE level), 2005.

NACE code	Industry	Service export intensity	Total sales (mil. DKK)
61	Water transport	0.86	125,013
62	Air transport	0.40	18,023
63	Supporting and auxiliary transport activities; activities of travel agencies	0.36	81,502
73	Research and development	0.33	5,342
72	Computer and related activities	0.14	54,914

Note: Financial and insurance industries are omitted due to missing information on total sales.

Source: Own calculations on the General Firm Statistics, the External Trade Statistics and the VAT register from Statistics Denmark.

costs than trade in manufactured goods as it involves the movement of persons across borders. Hence, we can think of some services as being intrinsically more tradable than others, although technological innovations, such as the internet, have lowered the trade costs of some services significantly over time.

However, international trade in services may also be hampered by trade restrictions which may be very different from trade restrictions on manufactured goods. Furthermore, these may not be as easy to spot as barriers to trade in manufactured goods. International trade in services has thus been limited mostly by non-tariff barriers such as local standards and restrictions on temporary migration. Yet in practice another »barrier« to international trade may be differences in demand across countries, implying that some services are not demanded (or less demanded) in other countries for, *e.g.*, cultural or geographic reasons.

One way of identifying existing barriers to international trade in services is to compare those services that are currently exported to those that are potentially tradable. The potentially tradable services can be identified using a method developed by Jensen and Kletzer (2006). A related method was applied by Malchow-Møller *et al.* (2007) in an analysis of the tradability of construction services in Denmark.

Jensen and Kletzer (2006) use the geographical concentration of industries within the US to find the potential tradability of services produced in different industries. The idea is to compare the demand for services and goods in a given industry (by other industries and consumers) in a given area of the country with the production or employment of the same industry in the same area. Goods and services produced by industries that are geographically more concentrated than their demand must be tradable within the country. Without any restrictions or impediments to international trade, we must expect the same goods and services to be tradable internationally.

Formally, the share of total domestic demand for industry i in area p is calculated as:

$$IDS_{i,p} = \sum_{j=1}^J \left(\frac{Y_{i,j}}{Y_i} \cdot \frac{EMPL_{j,p}}{EMPL_j} \right)$$

where Y_i is the total output of industry i used domestically, and $Y_{i,j}$ is the output of industry i used in industry j . Hence, $Y_{i,j}/Y_i$ is the share of industry i 's production used by industry j . Similarly, $EMPL_{j,p}$ is the employment in industry j in area p , and $EMPL_j$ is the total employment in industry j . Thus, $EMPL_{j,p}/EMPL_j$ is a measure of the share of industry j located in region p . Multiplying the two shares therefore gives us a measure of the share of industry i 's total output used by industry j in area p . And summing over industries in area p then gives us the share of industry i 's total output demanded by area p .

Note that investment and consumption by the government and private households is included as an extra »industry« among the J industries in the summation above, where the »local employment share«, $EMPL_{j,p}/EMPL_j$, of this »industry« is given by the share of residents in area p . Hence, $IDS_{i,p}$ measures the share of total output (excluding exports) from industry i used by other industries, consumers or the government in area p .

We then define $\sigma EMPL_{i,p}$ as the cumulative share of industry i employment in regions $1, \dots, p$, and $\sigma IDS_{i,p}$ as the cumulative share of demand for industry i in regions $1, \dots, p$, when regions are sorted in ascending order by their employment shares in industry i . Plotting these cumulative shares against each other gives us industry i 's »Lorentz curve«. If production shares match the demand shares perfectly, the Lorentz curve is a straight 45-degree line with an area of 0.5 below the curve. In all other cases, the Lorentz curve will be convex with an area less than 0.5 below it. The smaller the area, the worse is the match between the geographical distribution of demand and the geographical distribution of production (employment).

The Gini-coefficient, which lies between 0 and 1, is calculated as 1 minus two times the area under the Lorentz curve. It is therefore equal to zero for industries with a perfect geographical match between demand and production. On the other hand, an industry with production concentrated in one area and demand spread across areas will have a high Gini-coefficient, and is therefore considered tradable (within the country). Technically, the Gini is approximated using the following formula:

$$G_i = 1 - \sum_p (\sigma EMPL_{i,p} + \sigma EMPL_{i,p-1}) \cdot (\sigma IDS_{i,p} - \sigma IDS_{i,p-1})$$

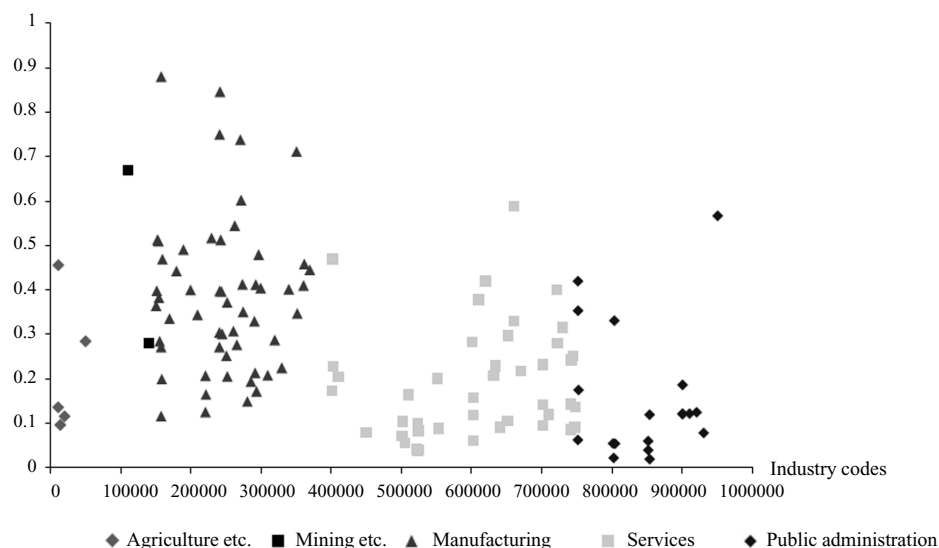


Figure 5. Gini coefficients for 124 different industries, 2005.

Source: Own calculations on input-output tables and matched employer-employee data from Statistics Denmark.

To calculate the demand shares we use information from input-output tables and matched employer-employee data (IDA) from Statistics Denmark. We use 124 different industries in the analysis as the input-output tables »only« distinguish between 130 different industries and since some of these have to be aggregated as they »share« one or more NACE codes. Alternatively, we could use a more detailed industry classification, *e.g.*, at the 4-digit NACE level as in Jensen and Kletzer (2006), but would then have to assume identical demand shares, $IDS_{i,p}$, for industries in the 4-digit classification that belong to the same industry in the 130-classification.

Furthermore, we use 9 different geographical areas which are somewhat smaller than the geographical areas used by Jensen and Kletzer (2006) for the US.⁸

Figure 5 illustrates the distribution of the calculated Gini-coefficients for the 124 industries for 2005. Not surprisingly, the manufacturing industries are on average more tradable than other industries. However, a number of service industries also appear relatively tradable with Ginis exceeding 0.3 – at least more tradable than many manufacturing industries. Four of the seven service industries with Ginis above 0.3 correspond closely to four of the industries from Table 2 (excluding »supporting and auxiliary transport activities«). The last three industries with Ginis exceeding 0.3 are two insurance industries plus a »gas supply« industry.

8. The nine areas are: Copenhagen, Bornholm, Zealand, Lolland-Falster, Funen, Southern Jutland, Eastern Jutland, Western Jutland and Northern Jutland.

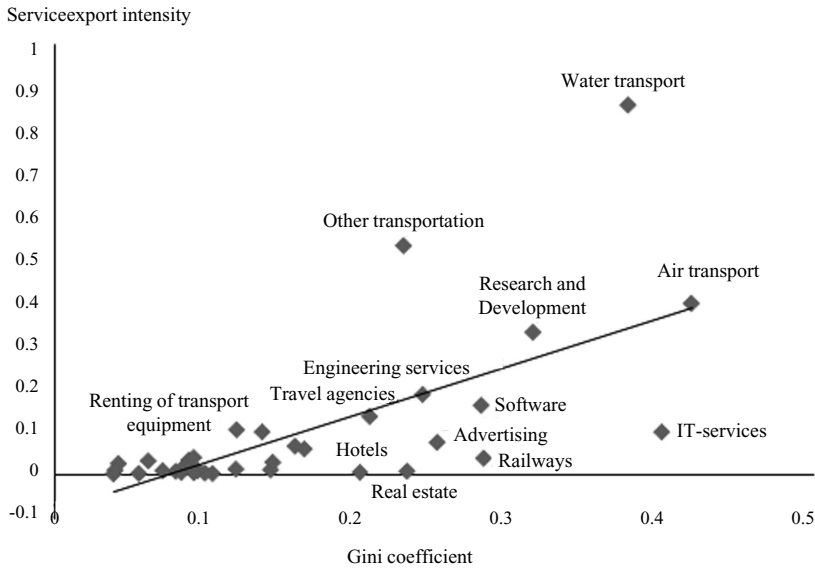


Figure 6. Tradability and service-export intensity, services industries, 2005.

Source: Own calculations on input-output tables, matched employer-employee data, the General Firm Statistics, the External Trade Statistics and the VAT register from Statistics Denmark.

Jensen and Kletzer (2006) use a Gini-value of 0.1 as the cut-off between tradable and non-tradable industries. Because of the smaller geographical areas used in the present study, a similar value cannot be applied here. The use of smaller geographical areas will automatically increase the Gini-coefficients – making industries look more tradable than in Jensen and Kletzer (2006).

Another problem with the Ginis as indicators of tradability is that industries that are tradable may still sometimes be geographically dispersed, *e.g.*, if there are no economies of scale or if some factors of production cannot be concentrated as in, *e.g.*, agriculture. The low Ginis for agriculture in Figure 5 are likely to be caused by the distribution of arable land which automatically leads to a geographically dispersed production, despite very tradable output. Similarly, a geographical concentration of industries need not imply international tradability but may be due to domestic specialization in more concentrated markets, *e.g.*, theatres and hotels in larger cities.

Despite these limitations of the concentration indices, it still seems reasonable to assume that industries that are concentrated in one part of Denmark are (all else equal) more tradeable than industries spread equally across the country. Hence, the Gini-coefficient seems to be a good indicator of the trade potential for an industry. Partly to check this, we analyse the relationship between tradable services and actually traded services.

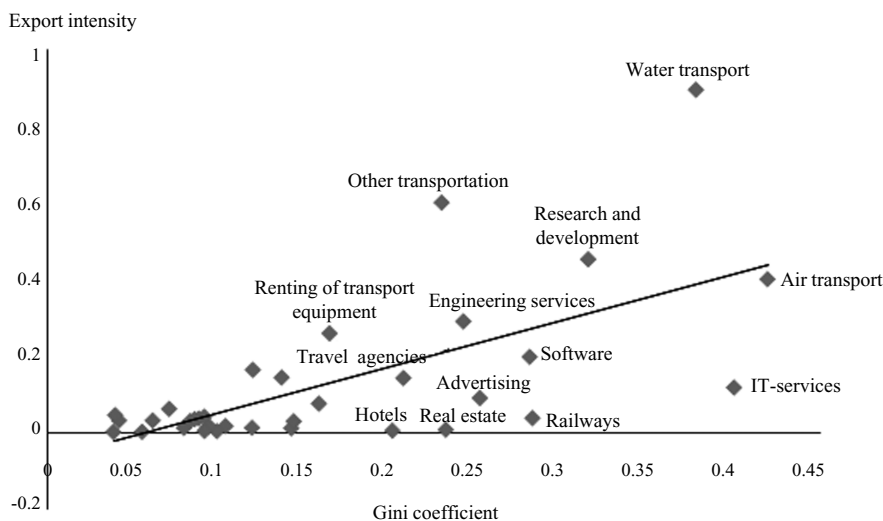


Figure 7. Tradability and total-export intensity, service industries, 2005.

Source: Own calculations on input-output tables, matched employer-employee data, the General Firm Statistics, the External Trade Statistics and the VAT register from Statistics Denmark.

Figure 6 plots the relationship between the calculated Gini-coefficient for a service industry – which indicates the potential for international trade – and the actual service-export intensity in the industry. Similarly, Figure 7 plots the relationship between the Gini-coefficient for a service industry and the total (services and goods) export intensity in the industry. There are 43 service industries in the 124-classification used. However, information about actual export intensity is missing for 9 of these industries (the finance/insurance industries and the industries responsible for water, gas and electricity supply).

As expected there is a clear and significantly positive relationship between the Gini-coefficient and the actual export intensity of an industry – both when using the service-export intensity and when using the total-export intensity. In both cases, the t -statistics exceed 4. This should not be taken to reflect any causal relationship, but it supports the idea that the Gini is a useful indicator of the international trade potential.⁹

The figures also show that most of the transport industries are characterised by both high tradability of their services and high actual export intensities. However, the figures also reveal some interesting outliers that exhibit high tradability but little actual export activity (observations below the regression line), and hence an apparent un-

9. Although both the Gini and the export intensity range between 0 and 1, there is not a one-to-one relationship between the two measures, and hence we should not expect a 45-degree line to emerge.

exploited potential for international trade. This is the case for, *e.g.*, »IT-services (excluding software)«, »advertising«, »railways«, »real estate« and »hotels«. These industries all have Gini-coefficients above 0.2 (which is relatively high, *cf.* Figure 8), while their service-export intensities are all lower than 0.1 (which is low compared to other service industries). There might be an obvious explanation for this in the case of the »hotel« industry: Hotels are typically concentrated in larger cities, and are therefore found to deliver potentially tradable services, when using the approach from Jensen and Kletzer (2006). However, their international export of services takes place through mode-2 and is therefore not recorded in the firm statistics. Hence, the difference between tradability and actual trade in this case is likely to be an artifact of the data. A similar explanation may apply (at least partly) to »railways« and »real estate«.

When it comes to »IT-services« and »advertising«, other explanations of the observed difference between tradability and actual trade could be at work. While services from these industries are traded within the country (as reflected in the Ginis), differences in languages, cultural patterns and market structures may work as impediments to international trade in these industries. The fact that the »software« industry enjoys higher export intensity than the »IT-services« industry perhaps reflects that software can more easily be shipped across borders than other IT-services.¹⁰

As a robustness check, we also estimated the relationships in Figures 6 and 7, excluding the four industries with service-export intensities exceeding 0.3. There is still a significantly positive relationship between an industry's Gini-coefficient and its actual service-export (or total-export) intensity. Furthermore, »IT-services (excluding software)« and to some extent also the »advertising« and »software« industries are still found below the regression line, *i.e.*, they are characterised by relatively high Ginis compared to their actual export intensities.

In summary, apart from »IT-services« and to a smaller extent »software« and »advertising«, the analysis reveals a rather robust relationship between a service industry's tradability and its actual international export of services.

6. Conclusion

In this paper, we have exploited information from the Danish statistical registers to construct a unique micro data set which contains firm-level measures of services exported (and imported) by Danish firms in recent years. These data allow for a much more detailed picture of service-export activities than existing studies which rely on aggregate BoP data.

10. The two outliers above the line »other transportation« and »water transport« may be due do some domestic sales being recorded as exports (as some of the services in these industries are not subject to VAT).

We showed that these firm-level measures are consistent with the more aggregate BoP data, which strongly indicates that the data provide relatively precise measures of exports of services at the firm level. We also illustrated how the data allow us to calculate exports or services at, *e.g.*, a detailed industry level. We found that while service industries are responsible for a large part of the trade in manufactured goods, most of the trade in services does in fact take place within the service industries. Among the services industries, high service-export intensities are found within some transport industries – in particular »water transport« – but also within R&D and computer services.

We also compared the actual export intensity of a service industry with its potential tradability using a method developed by Jensen and Kletzer (2006). The analysis revealed a robust relationship between a service industry's tradability and its actual international export of services. Only within »IT-services (excluding software)« and to a smaller extent within »software« and »advertising« do we find signs of an unexploited export potential, and hence potential barriers to international trade.

The data set constructed and used in the present paper also provides interesting possibilities for future research. A substantial literature has documented that exporters of manufactured goods tend to be more productive and pay higher wages than non-exporting firms; see, *e.g.*, Wagner (2007) for a recent survey. This evidence has highlighted a new potential source of gains from trade, as trade liberalization will reallocate resources from the least productive to the most productive firms. A similar analysis of the heterogeneity among service exporters can now be performed by exploiting the panel dimension of the current data. This is an obvious priority for future research.

Literature

- Bernard, A. and J. B. Jensen. 1995. Exporters, Jobs and Wages in US Manufacturing: 1976-1987, *Brookings Papers on Economic Activity, Microeconomics*, 1995, 67-119.
- Brandicourt, V., C. Schwellnus, and J. Wörz. 2008. Austria's Potential for Trade in Services, *FIW Research Report* No. 002, June 2008.
- Breinlich, H. and C. Criscuolo. 2009. International Trade in Services: A Portrait of Importers and Exporters, *Working Paper*, University of Essex.
- Copeland, B. and A. Mattoo. 2008. The Basic Economics of Services Trade, in A. Mattoo, R. M. Stern and G. Zanini (eds.), *A Handbook of International Trade in Services*, Oxford University Press, 84-129.
- DØRS. 2006. *Dansk økonomi – efterår 2006*. Det Økonomiske Råd, København, 2006.
- Feenstra, R. C. 1995. Estimating the Effects of Trade Policy, in G.M. Grossman and K. Rogoff (eds.), *Handbook of International Economics* 3, Elsevier, 1553-1595.
- Francois, J., O. Pindyuk, and J. Wörz. 2008. Trade Effects of Services Trade Liberalization in the EU, *FIW Research Report* No. 004, June 2008.
- Francois, J. and B. Hoekman. 2009. Services Trade and Policy, *Working Paper* No. 0903, Department of Economics, Johannes Kepler University of Linz.
- Jensen, J. B. and L. G. Kletzer. 2006. Tradable Services: Understanding the Scope and Impact of Services Offshoring, in S.M. Collins and L. Brainard (eds.), *Brookings*

- Trade Forum 2005, Offshoring White-Collar Work*, Brookings Institution, Washington DC, 75-134.
- Lipsey, R. E. 2006. Measuring International Trade in Services, *NBER Working Paper* No. 12271.
- Malchow-Møller, N., J. R. Munch, and J.R. Skaksen. 2007. Mobilitet af håndværksydelse i Danmark, *Nationaløkonomisk Tidsskrift*, 145, 113-43.
- Munch, J. R., and J. R. Skaksen. 2008. Human Capital and Wages in Exporting Firms, *Journal of International Economics* 75, 363-72.
- Statistics Denmark. 2007. *External Trade of Denmark 2006 – Goods and Services (Udenrigshandel 2006 – Varer og tjenester)*. Statistics Denmark, Copenhagen, www.dst.dk/ebog/udenrigsvarer.
- United Nations, European Commission, International Monetary Fund, Organisation for Economic Co-operation and Development, United Nations Conference on Trade and Development, and World Trade Organization (2002): *Manual on Statistics of International Trade in Services*. United Nations, Geneva, Luxembourg, New York, Paris, Washington D.C.
- Wagner, J. 2007. Exports and Productivity: a Survey of the Evidence from Firm-level Data, *The World Economy* 30, 60-82.

Debat og kommentarer

De økonomiske konsekvenser af familiegenerationsskiftet i danske virksomheder.

Morten Bennedsen

André and Rosalie Hoffman, Chair in Family Business Management, INSEAD, E-mail: bennedsen@insead.edu

Kasper Meisner Nielsen

Hong Kong University of Science and Technology, E-mail: nielsen@ust.hk

SUMMARY: In this article we analyze succession and exit in Danish family firms from 1995-2003. Focusing on managerial succession we identify 6219 changes in CEO position. Using difference-in-differences estimates and instrumental variables we find the cost of family succession in the CEO position to be around 70.000 DKK per 1 million DKK in assets. The cost of succession is even higher in firms where active leadership is more valuable and for incoming CEOs with lower measurable management skills.

1. Indledning

Der findes over 50.000 små og mellemstore virksomheder i Danmark, der er etableret enten som anparts- eller aktieselskab.¹ Langt de fleste af disse virksomheder ejes og kontrolleres af enkeltpersoner og familier. Familievirksomhederne spænder fra sværvægtede som Danfoss, Grundfos og LEGO til de fleste små og mellemstore virksomheder. Vores beregninger viser, at der i dag er op mod en million danskere, der har familiemæssig berøring med ejerskab og ledelse i små og mellemstore virksomheder.

Fra et samfundsmæssigt synspunkt leverer familievirksomhederne et afgørende bidrag til beskæftigelsen og produktionen i Danmark. Det er også blandt familievirksomhederne, at mange af de mest vækstorienterede virksomheder befinder sig, og det er derfor ofte blandt disse virksomheder fremtidens danske industri lokomotiver fin-

Denne artikel bygger på rapporten »Vækst via generationsskifte« offentliggjort af CEBR og Erhvervs- og Boligstyrelsen og indeholder resultater fra Bennedsen, Nielsen, Perez-González og Wolfenzon (2007), »Inside the Family Firm: The Role of Families in Succession Decisions and Performance«, The Quarterly Journal of Economics, May 2007.

1. Vi medtager alene virksomheder på konsolideret niveau, dvs. et moderselskab med to datterselskaber tæller kun som en enkelt virksomhed i opgørelsen over antallet af små og mellemstore virksomheder.

des. I den nyere corporate governance litteratur har de relative fordele og ulemper ved familieejerskab været intensivt studeret både teoretisk, Bennesen og Wolfenzon (2000), Burkart m.fl. (2002), Casselly og Gennaiolo (2003) og Chami (2001) og empirisk, Anderson og Reeb (2003), Anderson, Mansi og Reeb (2003), Villalonga og Amit (2004) og Pérez-González (2006).

Ejerskabsstrukturerne i små og mellemstore danske virksomheder er under en voldsom forandring. På den ene side står tusindvis af familievirksomheder over for et generationsskifte eller en anden form for exit i løbet af de næste 10 år. På den anden side bliver det traditionelle familiegenerationsskifte udfordret af strukturelle opbrydninger i de traditionelle familiemønstre. Set over det lange perspektiv vil fremtidens familievirksomheder blive påvirket af de sidste generationers stigende skilsmisserater, ændrede samlivsformer, faldende børnetal og et stadigt mere globaliseret arbejdsmarked, se f.eks. Bennesen m.fl. (2005), Bertrand m.fl. (2008), Gersick m.fl. (1997), Lansberg m.fl. (1999) og Ward (1987).

I denne lille oversigtsartikel analyserer vi generationsskiftet i de danske familievirksomheder. Vi fokuserer på at estimere de økonomiske konsekvenser af generationsskifte inden for familien og på, hvad kan der gøres for at få familiegenerationsskiftet til at klare sig økonomisk bedre i fremtiden. Vi identificerer generationsskifte både i ledelsen og ejerskabet, men fokuserer analysen på direktørskifter, eftersom det er metodisk vanskeligere at identificere veldefinerede generationsskifter mellem kontrollerende ejere.

2. Data

Udgangspunktet for analysen er alle danske aktie- og anpartsselskaber. Vi har data for generationsskifter i perioden 1995 til 2003 og regnskabstal for perioden 1992 til 2006. Disse virksomheder er jævnfør årsregnskabs- og selskabslovgivningen pålagt at indsende deres årsregnskab til Erhvervs- og Selskabsstyrelsen. Årsregnskabet skal foruden en række stamoplysninger indeholde finansielle oplysninger (en årsopgørelse med årets resultat og en balance ved regnskabsårets afslutning), informationer om virksomhedens ledelse (direktion og bestyrelse), samt en liste over alle aktionærer, som besidder mere end 5 procent af aktiekapitalen.

Foruden indsendelse af årsregnskabet, er alle aktie- og anpartsselskaber forpligtet til at underrette Erhvervs- og Selskabsstyrelsen om ændringer i virksomhedens ledelse (direktion og bestyrelse) senest inden for 2 uger efter ændringen. Vi har fået adgang til Erhvervs- og Selskabsstyrelsens database over nuværende og tidligere direktører og bestyrelsesmedlemmer. Endelig har vi suppleret dette datasæt med oplysninger om familierelationer fra CPR-registret, som gør os i stand til at identificere de familiære bånd mellem medlemmerne i de virksomhedsdrivende familier. CPR-registret inde-

holder en række personlige oplysninger om et givent individs kernefamilie (forældre, søskende, ægtefælle og børn).

Dette omfattende datamateriale giver mulighed for at analysere to typer af generationsskifter: direktør- og ejerskifter og den efterfølgende konsekvens for virksomhedens drift. Eftersom selskabslovgivningen giver mulighed for at have op til 3 direktører, fokuserer vi alene på virksomheder med en enkelt direktør, hvor vi med sikkerhed ved, hvornår generationsskiftet er foregået. Desuden kræver vi, at både den afgangende og den nye direktør har siddet på posten i mindst 2 år henholdsvis før og efter generationsskiftet. Dermed sikrer vi, at udviklingen af virksomhedens drift kan henføres til generationsskiftet, og at den nye direktør har besiddet posten tilstrækkelig længe til at kunne have en målbar effekt på driften. Samtidig udelukker vores valg, at midlertidige og kortvarige skift i direktørposten har indflydelse på resultaterne.

For ejerskifter er identifikationen af generationsskifter mere kompliceret, eftersom lovgivningen ikke stiller krav om, at den faktiske fordeling af ejerskabet skal offentliggøres i årsrapporten. For at sikre os, at vi analyserer faktiske generationsskifter og ikke tilfældige udsving i data, har vi været nødt til at træffe en række afgrænsende valg. Som konsekvens heraf begrænses antallet af ejerskifter, hvor vi med sikkerhed kan analysere effekten af generationsskiftet på virksomhedens performance. I enkelthed kræver vi, at vi kan observere et veldefineret skift i kontrollen med virksomheden. Med andre ord kræver vi, at den afgangende og den nye ejer er kontrollerende (dvs. ejer mere end 50 procent af aktiekapitalen), og at den afgangende ejer forlader virksomhedens ejerkreds (dvs. sælger helt ud). Derudover kræver vi, i lighed med direktørskifterne, at både den gamle og nye ejer har kontrol med virksomheden i 2 år henholdsvis før og efter generationsskiftet.

For direktørskifter klassificerer vi generationsskiftet i tre typer, som adskiller sig med hensyn til den nye direktørs relation til den gamle direktør og virksomheden i øvrigt. De tre typer af generationsskifte er familiedirektør, intern direktør og ekstern direktør. Familiedirektører er kendetegnet ved at være i kernefamilie med den afgangende direktør (dvs. forældre, børn, ægtefælle eller søskende). For ikke-familiedirektører skelner vi mellem at være intern eller ekstern, hvor en intern direktør modsat den eksterne er kendetegnet ved at have arbejdet i virksomheden umiddelbart før forfremmelsen.

Vi fokuserer på denne opdeling, fordi de to typer af ikke-familiedirektører som udgangspunkt må formodes at være meget forskellige. Når den nye direktør har arbejdet i virksomheden, er der tale om en ny ledelse, der som udgangspunkt kender virksomheden godt, og hvis ideer omkring fremtidig virksomhedsledelse må formodes at tage udgangspunkt i virksomhedens nuværende organisations og ledelsesform. I nogle tilfælde vil den nye ledelse på sigt også overtage ejerskabet af virksomheden, det vil sige foretage en management buy out (MBO). Det er værd at understrege, at vi ikke

kræver, at den nye familiedirektør har arbejdet i virksomheden umiddelbart før ansættelsen som direktør, men vi kan notere, at langt de fleste af familiedirektørerne rent faktisk arbejder i virksomheden før forfremmelsen.

Den eksterne direktør vil derimod ikke være opfostret i virksomhedens nuværende drift. Det er således muligt, at eksterne direktører vil være mere forandringsvillige med hensyn til virksomhedens fremtidige drift og organisering. Dette vil formodentlig gøre sig endnu mere gældende i de tilfælde, hvor den eksterne direktør også overtager ejerskabet med virksomheden, det vil sige foretager en management buy in (MBI).

Når vi analyserer ejerskifter, klassificerer vi generationsskiftet i fire kategorier, nemlig familieejer, intern ejer, ekstern ejer og finansiel ejer. Som i tilfældet med direktørskiftet er nye ejere fra familien kendetegnet ved at være i kernefamilie med den afgående ejer. For ikke-familie ejere skelner vi i tilfældet med ejerskab mellem interne, eksterne og finansielle købere. De interne og eksterne ejere har henholdsvis arbejdet og ikke arbejdet i virksomheden. Den nye kategori under ejerskabstransition er den finansielle køber, der enten er en anden virksomhed eller en eller flere personer, der ejer andre virksomheder i forvejen.

Panel A i tabel 1 viser, at vi har identificeret 6219 direktørskifter i danske virksomheder i perioden fra 1995 til 2003. Det årlige antal ligger på mellem 420 og 850. Fordelt på typen af exit er der foretaget en familietransition i 1744 virksomheder, hvilket svarer til 28 procent af alle direktørskifter. I 2381 tilfælde (38,3 procent) var den nye direktør ansat i virksomheden, mens de resterende 2094 (33,7 procent) var eksternt rekrutterede.

Panel B i tabel 1 fokuserer på ejerskifter i perioden 1995 til 2003. For ejerskifter er identifikationen af generationsskifter mere kompliceret, eftersom lovgivningen ikke stiller krav om, at den faktiske fordeling af ejerskabet skal offentliggøres i årsrapporten, og identifikation af kontrolskifter i ejerskabskredsen er derfor ikke altid muligt. I alt har vi identificeret 819 ejerskifter, hvor der er sket et observerbart skifte af kontrollen med virksomheden. Eftersom vi forventer, at disse kun udgør en lille andel af de reelle skift, der er sket i perioden fra 1995 til 2003, fokuserer resten af denne oversigtsartikel på direktørskifter.

Det er klart, at vi har foretaget en række afgørende valg for at identificere disse event virksomheder, hvor det er muligt at analysere generationsskiftet. På grund af disse valg adskiller vores event virksomheder sig på visse områder fra gennemsnittet af danske virksomheder. De to væsentlige forskelle er, at de er ældre og større end gennemsnitsvirksomheden. De fleste virksomheder overlever ikke de første 10 år som selvstændig virksomhed. Derfor er det naturligt, at de virksomheder, der når til at skifte direktør eller ejere, er både større og ældre end gennemsnitsvirksomheden.

Det er også værd at notere, at der ofte er en sammenhæng mellem direktørskift og ejerskift. For det meste vil skift i ejerskabet være sammenfaldende med et direk-

Tabel 1. Antal generationsskifter fordelt på generationsskiftetype.

	Antal (#)	Familie		Generationsskiftetype				Finansiel	
		#	%	Intern #	%	Ekstern #	%	#	%
Panel A: Direktørskifte									
Alle	6219	1744	28,0	2381	38,3	2094	33,7	n.a.	n.a.
Panel B: Ejerskabsskifte									
Alle	819	509	62,1	67	8,2	133	16,2	110	13,4

tørskift. Vi kan ikke præcist dokumentere, at det modsatte er tilfældet for direktørskifter. Igen er problemet, at vi for mange direktørskifter ikke kan se, om der har været et skift i den de facto kontrollerende ejer.

3. De økonomiske konsekvenser af generationsskiftet

Økonomisk intuition giver to modsatrettede bud på de økonomiske konsekvenser af at lade direktørstolen gå i arv: På den positive side vil man få en ny direktør, der gennem et fokuseret langsigtet aktivt ejerskab beskytter virksomhedens ve og vel og som derudover besidder en stor relationsspecifik viden om virksomheden allerede inden ansættelsen.

På den mere negative side er det ikke givet, at den bedst egnede blandt virksomhedsejerens børn kan matche den bedst mulige udefrakommende professionelle direktør. Sandsynligheden for, at den bedst kvalificerede kandidat til lederposten kommer fra familien, påvirkes af det relative antal af familie- og professionelle kandidater. Hvis markedet for ledelsestalent er velfungerende, må det med rimelighed kunne formodes, at antallet af professionelle kandidater vil overstige antallet af kvalificerede og interesserede kandidater inden for familien. I så tilfælde vil et generationsskifte inden for familien kunne føre til en svagere og mindre kompetent ledelse, end hvis virksomheden vælger en professionel ledelse udefra.

3.1 *Difference-in-differences* estimat

Spørgsmålet om, hvordan virksomhedens resultat påvirkes af beslutningen om at overdrage ledelsen til et familiemedlem, er metodisk set vanskeligt at besvare. Den mest simple metode til at analysere effekten er at beregne ændringer i virksomheders performance før og efter direktørskiftet. Man sammenligner så disse ændringer i virksomheder, hvor den nye direktør vælges indefra med ændringerne i virksomheder, der vælger en udefrakommende direktør. Denne metode kaldes for et »*difference-in-differences*« estimat, og resultatet er præsenteret i tabel 2.

Tabel 2. Difference-in-differences estimation af effekten af generationsskifter på virksomhedens performance.

	Generationsskifte			Difference
	Alle	Inden for familien	Uden for familien	
Før	-0,003** (0,002) [6219]	0,005* (0,003) [1744]	-0,006*** (0,002) [4475]	0,011*** (0,003)
Efter	-0,004** (0,002) [6219]	-0,001 (0,003) [1744]	-0,004** (0,002) [4475]	0,003 (0,004)
Difference	-0,001 (0,002)	-0,006* (0,002)	0,001 (0,002)	-0,008** (0,004)

Vi klassificerer generationsskifter i to grupper: Inden for familien, når den nye direktør eller ejer er i familie med den afgående direktør og uden for familien, når dette ikke er tilfældet. Tabellen rapporterer den gennemsnitlige branchejusterede overskudsgrad i tre år før og efter generationsskiftet. Vi udelader eventåret i alle tilfælde. Branchejusteringen foretages på et 4-cifret DB93 brancheniveau ved at fratække den gennemsnitlige overskudsgrad inden for branchen. Overskudsgraden er defineret ved primært resultat over den bogførte værdi af aktiverne. Standardafvigelse er rapporteret i parenteser, mens antallet af observationer er angivet i klammer. ***,** og * angiver signifikans på henholdsvis 1, 5 og 10 procents niveau.

For at sikre, at vi isolerer effekten af generationsskiftet sammenligner tabellen den gennemsnitlige overskudsgrad, defineret ved primært resultat over den bogførte værdi af aktiverne, i de 3 år forud for generationsskiftet med de 3 efterfølgende år. Selve året med generationsskiftet anvendes således ikke i analysen af udviklingen i virksomhedens performance. For at gøre sammenligningen meningsfyldt er overskudsgraden justeret på brancheniveau.

Tabel 2 viser, at virksomheder, som foretager et direktørskifte inden for familien, er kendetegnet ved et bedre driftsresultat i de 3 år som går forud for generationsskiftet. I gennemsnit har virksomheder, som vælger en familiedirektør en 1,1 procentpoints højere overskudsgrad end virksomheden som vælger en udefrakommende direktør. Denne forskel er signifikant på et 1 procentsniveau.

Efter generationsskiftet ændres dette billede. Virksomheder med et generationsskifte inden for familien klarer sig nu ikke længere signifikant bedre end virksomheder, der udpegede en udefrakommende direktør. Virksomhederne med familietransition oplever således en nedgang i den branchejusterede overskudsgrad på 0,6 procentpoint samtidig med at de udefrakommende direktører forbedrer resultatet marginalt med 0,1 procentpoint. Forskellen på 0,8 procentpoint i udviklingen i overskudsgraden er signifikant på et 5 procentsniveau. Familiedirektører klarer sig altså i gennemsnit dårligere end udefrakommende direktører, når vi brugere difference-in-differences metoden.

Gennemsnitsvirksomheden har et ordinært resultat over aktiver på omkring 5,5 procentpoint, hvilket svarer til 55.000 kroner per million kr. i aktiver. Det betyder, at en virksomhed, der vælger en direktør fra familien i gennemsnit, taber ca. 6.000 kroner per million kr. i aktiver, når vi opgør det ved difference-in-differences metoden. Omvendt oplever virksomheder, der vælger en udefrakommende direktør, en resultatforbedring på 0,2 procentpoint, svarende til 2.000 kroner pr. million kroner i aktiver. Det vil sige, at forskellen mellem at vælge en direktør indefra eller udefra i denne simple måling ligger på omkring 8.000 kroner per million kr. i aktiver. Alt andet lige er det værd at notere, at disse effekter må betragtes som økonomisk små. Selv om difference-in-differences estimationen viser, at familiegenerationsskiftet klarer sig signifikant dårligere i statistisk forstand er denne forskel ikke særlig stor relativt til virksomhedernes samlede resultat.

For ejerskifter finder vi ligeledes en negativ effekt på virksomhedens drift af at lade virksomheden gå i arv. Igen oplever virksomheder med et generationsskifte inden for familien en nedgang i overskudsgraden svarende til 1,5 procent point eller 15.000 kroner per million kr. aktiver. Eftersom virksomheder med ejerskifter uden for familien ikke ændrer overskudsgrad, udgør dette fald den samlede forskel mellem de to typer af ejerskiftet. Denne forskel er altså økonomisk set større end for direktørskifterne, men igen må vi konkludere, at den ikke er alarmerende.

3.2 Instrument variabel estimation

Ideen bag difference-in-differences estimationen er intuitivt appellerende. Hvis vores analyse byggede på et kontrolleret eksperiment, hvor vi tilfældigt udvalgte hvilke virksomheder, der skulle vælge et familiemedlem som ny direktør og hvilke der skulle ansætte en udefrakommende direktør, så ville det være korrekt at konkludere, at en direktør udefra vil skabe et resultat, der er 8.000 kr. bedre per million i aktiver end en direktør indefra. Tilsvarende ville et ejerskifte uden for familien forbedre virksomhedens drift med 15.000 kr. per million kr. i aktiver.

Det er metodisk vigtigt at understrege, at det ikke er tilfældigt hvilke virksomheder, der vælger generationsskifte inden for og uden for familien. Så vi kan ikke betragte generationsskiftet som et kontrolleret eksperiment. Derfor eksisterer der en række økonometriske problemer i anvendelsen af den simple difference-in-differences metode, som kan have afgørende indflydelse på resultaterne.

Tabel 2 viser, at virksomheder, hvor ledelsen bliver inden for familien, generelt er i bedre finansiell stand. Det er i sig selv ikke et problem for beregningen af effekten, da vi kan kontrollere for målbare forskelle i en regressionsanalyse. Problemet opstår, når der er ex-ante forskelle, som ikke kan måles. Oplagte eksempler herpå er, hvis virksomheder, der gives videre inden for familien, har bedre fremtidsudsigter, fordi der er

flere potentielle investeringsmuligheder, nye uopdyrkede markeder eller bare gode ideer til fremtidig produktudvikling.

Disse potentielle problemer ved anvendelsen af difference-in-differences estimatoren kan illustreres ved hjælp af følgende eksempel: Antag at virksomhedens performance efter et generationsskifte er givet ved $Q + I$, hvor Q er kvaliteten af virksomhedens projekt og I er identiteten af den nye direktør. Den nye direktør kan enten være et familiemedlem eller en udefrakommende person. Lad $I = f$ og $I = u$ karakterisere henholdsvis et generationsskifte inden for og uden for familien. Med andre ord er vi interesserede i at estimere $f - u$, som er forskellen i performance mellem generationsskifter foretaget inden for og uden for familien. Antag yderligere at kvaliteten af virksomhedens projekt kan variere, og at projektets kvalitet ikke kan måles, men er velkendt for alle familiens medlemmer. For simpelthedens skyld antager vi derudover, at projektet kan have enten høj (qH), mellem (qM) eller lav kvalitet (qL), hvor hvert udfald har sandsynligheden en tredjedel.

For at gøre eksemplet interessant antager vi, at jo bedre kvaliteten af virksomhedens investeringsmuligheder er, jo større interesse er der blandt familiens børn for at overtage virksomheden, fordi fremtidsudsigterne er lysere. Derudover vil vi også antage, at drenge generelt set er mere villige til at fortsætte virksomhederne, end piger er. Det er et generelt resultat, at jo flere drenge der er i de danske virksomhedsdrivende familier, desto større er tendensen til at virksomhederne går i arv. Tabel 3 illustrerer en situation, som er konsistent med de ovenstående antagelser.

I dette eksempel vil generationsskifter inden for (uden for) familien aldrig ske, hvis kvaliteten af virksomhedens produkt er lav (*høj*). I dette tilfælde vil difference-in-differences estimatet ikke blot inkludere den sande effekt, men også den selektions bias som fejlagtigt bliver tillagt direktørens talent. Rent illustrativt vil difference-in-differences estimatoren sammenligne ændringer i performance af generationsskifter inden for familien, som har en tendens til at ske i den øverste del af tabel 3, med generationsskifter uden for familien, som har en tendens til at ske i nederste del af tabellen. Som følge heraf vil difference-in-differences estimatoren have en positiv bias, idet vi vil tillægge forskellen i kvaliteten af investeringsmuligheder ($q_H - q_L$) til familiendirektørens evner.

Vi kan undgå denne bias ved at bruge et såkaldt instrument for generationsskiftet. Vi vælger at bruge kønnet på den førstefødte i familien som vores instrument. Vi forklarer nedenfor, hvorfor dette er et godt instrument. Når vi estimerer de økonomiske konsekvenser ved hjælp af et velegnet instrument, undgår vi den bias, som er indbygget i difference-in-differences estimatoren.

Instrument estimationen (IV-estimation) sammenligner alle generationsskifter betinget på kønnet af det første barn. Intuitivt svarer dette til at udregne forskellen i per-

Tabel 3. Illustration af økonometriske problemer forbundet med difference-in-differences estimation.

Investeringsmuligheder	Kønnet af førstefødte barn	
	Dreng	Pige
Høje	Generationsskifte inden for familien $q_H^+ f$	Generationsskifte inden for familien $q_H^+ f$
Mellem	Generationsskifte inden for familien $q_M^+ f$	Generationsskifte uden for familien $q_M^+ u$
Lave	Generationsskifte uden for familien $q_L^+ u$	Generationsskifte uden for familien $q_L^+ u$

formance på tværs af kolonerne i tabel 3. Bemærk, at i det ovenstående eksempel valgte alle virksomhederne et generationsskifte inden for familien (uden for) uanset kønnet af det førstefødte barn, hvis kvaliteten af virksomhedens projekt var *høj* (lav). I tilfældet hvor kvaliteten er *mellem*, vælger virksomheder, hvor kønnet på det førstefødte barn er en dreng, en ny direktør inden for familien, hvorimod en udefrakommen direktør vælges, hvis kønnet er en pige. IV-estimatoren bruger med andre ord kun variationen inden for gruppen med »mellem« investeringsmuligheder, hvor virksomhederne responderer til variationen i instrumentet, se Imbens og Angrist (1994).

Ovenstående eksempel påviser fordelene ved at anvende IV-estimation, idet enhver selektions bias i difference-in-differences estimatoren vil bestå, så længe generationsskiftevalget er afhængigt af andre ukontrollerbare determinanter af virksomhedens performance. IV-estimatet bliver derimod ikke påvirket af forskelle mellem virksomheder, idet instrumentet er uafhængigt af andre determinanter af virksomhedens performance. Endelig er det værd at bemærke, at IV-estimatet kun er baseret på den gruppe af virksomheder, hvor generationsskiftet påvirkes af instrumentet, hvilket i sig selv kan være et problem for fortolkningen af de resultater, vi opnår ved brug af instrumentet. I denne sammenhæng er dette problem mindre, eftersom vi ikke finder signifikante forskelle i de målbare ledelseskompetencer (tidligere direktør- og bestyrelseserfaring, uddannelseslængde og universitetsgrad) mellem den gennemsnitlige nye familiedirektør og de nye familiedirektører, som vælges på grund af kønnet på det førstefødte barn.

Et godt instrument skal opfylde to væsentlige egenskaber i vores kontekst: For det første skal det være korreleret med, om der vælges ny direktør (ejer) fra familien. For det andet skal det kun påvirke virksomhedens resultat gennem direktørskiftet og ikke via andre uobserverbare kanaler.

Vi anvender kønnet på den afgående direktørs førstefødte barn som vores instru-

ment. Vores instrument er intuitivt appellerende, idet barnets køn i praksis er eksogent, og der samtidig er en systematisk tendens til at videregive familievirksomheden til en mandlig arvtager.² Selvom denne tendens må formodes at være aftagende, er der stadig empirisk belæg for at hævde, at sønner enten er foretrukket eller har en større interesse end døtre i at være direktør. Blandt de virksomheder, som vælger en familiedirektør, er der således en klar overvægt af sønner.³ Kønnen af det førstefødte barn er desuden eksogent i forhold til virksomheds- eller familiekarakteristika, som kunne have betydning for virksomhedens overskudsgrad. Der er således ingen signifikant forskel i virksomhedens størrelse og alder, den afgangende direktørs familie størrelse (antal børn) og civilstand mellem virksomheder med en førstefødt søn eller datter.

Vi laver en IV-2SLS estimation med brug af dette instrument i tabel 4. For at lette præsentationen har vi valgt udelukkende at præsentere anden del af regressionen, som analyserer den kausale effekt af familietransition på virksomhedens økonomiske performance. Den afhængige variabel er således ændringen i virksomhedens drift fra de 3 år før generationsskiftet til de 3 år efter.

Tabel 4 viser overraskende, at den negative effekt på virksomhedens drift økonomisk set er væsentligt større, når vi tager højde for forskelle i virksomhedens fremtidsudsigter. Vi finder således i model IV, at det gennemsnitlige afkast falder med op mod 7,7 procentpoint, når virksomheden går i arv. Dette er økonomisk set en meget stor negativ effekt. Med andre ord så forsvinder alt overskuddet fra den primære drift i gennemsnit (dvs. 70-80.000 kroner på bundlinjen for hver million kr. i aktiver), når en virksomhed vælger et generationsskifte inden for familien.

Forskellen mellem OLS og IV-estimatet skyldes formentlig, at afgangende direktører i høj grad tager højde for virksomhedens fremtidsudsigter, når de vælger den nye direktør. Den afgangende direktør er forventeligt mere tilbøjelig til at vælge et generationsskifte inden for familien, hvis fremtidsudsigterne er gode, end hvis de er dårlige. Ønsket om afkommets succes, giver dermed en positiv bias til OLS estimatet i stil med den illustrerede effekt i Tabel 3. IV estimatet indtificerer derfor en betydelig større effekt af generationsskifter end OLS estimatet.

4. Årsager til omkostningerne ved familiegenerationsskiftet

Vi analyserer nu, hvorfor familiegenerationsskiftet klarer sig økonomisk dårligere

2. I artiklen Bennedsen, Nielsen, Pérez-González og Wolfenzon (2007) viser vi ligeledes, at kønnen på den afgangende direktørs første barn ikke påvirker virksomhedens resultat gennem andre kanaler end valget af den nye direktør. Denne eksklusiv egenskab er afgørende for at instrumentet er validt.

3. En måde at illustrere denne tendens på er at kigge på virksomheder, der vælger en direktør fra familien. I familier med f.eks. 2 børn fordelt på 1 dreng og 1 pige, udgør andelen af mandlige arvtagere 2/3. Hvis kønnet ikke spillede en rolle var den forventede andel af mandlige arvtagere 1/2. Lignende »overrepræsentation« af mandlige arvtagere eksisterer for andre kombinationer af antal børn og kønsfordeling.

Tabel 4. OLS- og IV-estimation af effekten af generationsskifter på virksomhedens performance.

Model	(I)	(II)	(III)	(IV)
Estimationsmetode	OLS	IV	OLS	IV
Familiedirektør	-0,0077** (0,0038)	-0,0082** (0,0040)	-0,0035 (0,0039)	-0,0767** (0,0338)
Virksomhedens størrelse			-0,0030*** (0,0010)	-0,0038*** (0,0017)
Branchejusteret overskudsgrad, $t-1$			-0,3066*** (0,0142)	-0,2950*** (0,0170)
Års-dummies	Nej	Nej	Ja	Ja

Den afhængige variabel er ændringen i virksomhedens branchejusterede overskudsgrad omkring generationsskifter. Virksomhedens overskudsgrad er defineret ved primært resultat over den bogførte værdi af aktiverne. Branchejusteringen foretages på 4-cifret DB93 brancheniveau. Ændringen i performance er udregnet ved differencen mellem den gennemsnitlige performance tre år før og efter generationsskiftet. Selve året for generationsskiftet er udeladt. Instrumentet for generationsskiftebeslutningen er kønnet af det førstefødte barn, som antager værdien 1 for drenge. Vi kontrollerer for virksomhedens størrelse (log. til den bogførte værdi af aktiverne) samt virksomhedens overskudsgrad i årene umiddelbart før generationsskiftet. Robuste standardafvigelse er rapporteret i parenteser, mens ***, ** og * angiver signifikans på henholdsvis 1, 5 og 10 procents niveau.

end andre former for exits. Vi fokuserer på de målbare kompetencer, som direktørerne har opnået gennem uddannelse og erhvervs erfaring samt på, i hvor høj grad virksomhedernes har brug for en dynamisk ledelse.

Tabellen sammenligner hyppigheden af en kompetenceegenskab hos en familiedirektør med hyppigheden af den samme kompetenceegenskab hos en udefrakommende direktør. Blandt familiedirektører har kun godt 16 procent tidligere erfaring som direktør, blandt de professionelle er det godt 24 procent. Hvis vi derimod fokuserer på de direktører, som har ledelseserfaring, så har både familie- og ikke-familiedirektører omkring 4 års ledelsesanciennitet, før de tiltræder i forbindelse med generationsskiftet.

Måske mere overraskende er det, at de udefrakommende direktører også har mere bestyrelseserfaring (35 procent) end familiedirektørerne (31 procent). Fokuserer vi på bestyrelsesancienniteten blandt de nye direktører med bestyrelseserfaring, så viser det sig, at familiedirektører har mere erfaring fra bestyrelsesarbejde end både de interne og eksterne direktører. I gennemsnit har familiedirektøren 2,5 års bestyrelsesanciennitet mod omkring 1,5 år for ikke-familiedirektører.

Den sidste parameter, som tabel 4 medtager, er, om den afgående direktør forsætter i virksomheden som bestyrelsesmedlem. I virksomheder med et direktørskifte inden

Tabel 5. De nye direktørers målbare ledelseskompetencer fordelt på generationsskiftetyper.

	Generationsskiftetype			
	Alle	Familie	Intern	Ekstern
A. Baggrund				
Alder (år)	44,2	40,9	45,1	45,9
Mand (%)	87,0	76,7	91,0	91,1
Kvinde (%)	13,0	23,3	9,0	8,9
B. Uddannelse				
Gennemsnitlig længde (år)	13,4	12,7	13,5	13,8
Universitetsgrad (%)	34,1	22,7	37,7	39,9
C. Ledelseserfaring				
Erfaring som direktør (%)	22,2	16,4	24,8	24,2
Gennemsnitlig erfaring for tidligere direktører (år)	4,2	4,2	4,4	4,1
D. Bestyrelseserfaring				
Erfaring som bestyrelsesmedlem	34,1	31,5	37,4	32,4
Erfaring for tidl. bestyrelsesmedlemmer (år)	1,7	2,5	1,7	0,9
E. Glidende generationsskifte				
Afgående direktør forsætter i bestyrelsen (%)	26,9	38,6	26,4	17,5

for familien bliver den afgangende direktør i bestyrelsen i næsten 39 procent af tilfældene, mens det kun sker i omkring 20 procent af tilfældene, hvis det er en ikke-familiedirektør. Interessant nok bliver den afgangende direktør i bestyrelsen oftere hvis den nye direktør er intern end ekstern. For internt rekrutterede direktører bliver den afgangende direktør i bestyrelsen i 26 procent af tilfældene mod 18 procent for de eksterne rekrutterede direktører. En forskel som formentlig skal tillægges, at ejerskabet skifter i færre af de tilfælde, hvor den nye direktør er internt frem for eksternt rekrutteret.

Indtil videre har vi kun dokumenteret, at der er færre målbare kompetencer hos nye direktører og ejerledere når disse kommer fra familien. Vi har ikke dokumenteret, at de økonomiske konsekvenser af generationsskiftet også kan henføres til disse forskelle. Derfor sammenligner vi i tabel 6 forskellene i de målbare kompetencer mellem de 20 bedste og dårligste virksomheder inden for hver generationsskiftetype.

Generelt set viser tabel 6, at de virksomheder som har klaret sig godt i gennem generationsskiftet, har udpeget en direktør, som har bedre målbare kompetencer end de virksomheder, som har klaret sig mindre godt. Denne forskel kommer specielt til udtryk i gruppen af virksomheder, som vælger en ny direktør fra familien. De 20 bedste virksomheder med familietransition er således kendetegnet ved, at de nye direktører

Tabel 6. Forskelle i målbare kompetencer blandt de 20 bedste og 20 dårligste virksomheder inden for hver generationsskiftetype.

	Generationsskiftetype		
	Familie	Intern	Ekstern
A. Uddannelse			
Gennemsnitlig længde (år)			
– Dårligste 20 virksomheder	11,8	12,8	12,6
– Bedste 20 virksomheder	12,5	13,1	12,6
Universitetsgrad (%)			
– Dårligste 20 virksomheder	10,0	25,0	15,0
– Bedste 20 virksomheder	23,8	23,8	14,3
B. Ledelseserfaring			
Erfaring som direktør (%)			
– Dårligste 20 virksomheder	5,0	40,0	5,0
– Bedste 20 virksomheder	38,1	33,3	14,3
C. Bestyrelseserfaring			
Erfaring som bestyrelsesmedlem			
– Dårligste 20 virksomheder	20,0	45,0	15,0
– Bedste 20 virksomheder	28,6	14,3	4,8

har markant længere uddannelse, og at en langt større andel af de nye direktører har en relevant grad fra universiteterne eller handelshøjskolerne.

Ligeledes er det slående, at andelen af nye direktører, der allerede har ledelseserfaring før de bliver ansat, er 38 pct. for de bedste virksomheder, der valgte generationsskifte inden for familien, hvorimod det kun er ca 5 pct. for de dårligste virksomheder.

Tabel 6 illustrerer således, at hvis familievirksomhederne ønsker et generationsskifte inden for familien, så er de målbare kompetencer meget afgørende for, om generationsskiftet bliver en økonomisk succes eller fiasko. For interne og eksterne er tendensen mindre klar, hvilket formentlig skyldes at f.eks. branchekendskab kan kompensere for den manglende ledelseserfaring.

Hvis det er de nye familiedirektørers lave kompetenceniveau, der er årsag til at generationsskiftet inden for familien ikke klarer sig så godt som andre former, må det formodes, at omkostningerne er højere i de virksomheder, der kræver aktiv ledelse, f.eks. fordi der er behov for kontinuerligt at ændre på organisationen og produktionen. I Bennedsen m.fl. (2007) dokumenterer vi netop, at dette er tilfældet: Vi viser, at omkostningen ved familiegenerationsskiftet er højere i konkurrence udsatte virksomheder, i virksomheder med meget forskning og udvikling og i virksomheder, der generelt har et højt niveau af humankapital.

5. Fremtidens generationsskiftebehov og politikovervejelser

Ved at kombinere virksomhedernes ejerskabsform med alderen på direktørerne kan vi tegne et billede af, hvor mange familievirksomheder, der står foran generationsskifte i løbet af de næste 10-20 år. Vi vurderer, at der er mere end 27.000 virksomheder, der er familieejede og har en direktør over 55 år. Heraf er direktørens alder mellem 55 og 65 år i ca 18.500 og over 65 år i ca 8.500 virksomheder. Det konkluderes således, at der vil blive et betydeligt antal exits i små og mellemstore danske virksomheder i løbet af de næste par årtier. Hvis vi går lidt ned i disse tal, så viser det sig, at behovet for generationsskifte er presserende inden for alle størrelser af virksomheder, i alle brancher og i alle fem regioner i landet, (se Bennedsen og Nielsen (2008) for uddybende analyse af fremtidige generationsskifter).

Vi har vist, at familiedirektørerne mangler målbare kompetencer relativt til de udefrakommende direktører, og at dette i høj grad gælder de virksomheder, der klarer sig dårligst i vores undersøgelser. Vi anerkender, at der kan være andre faktorer, der medvirker til de observerbare forskelle. F.eks. kan man med rimelighed forestille sig, at familiedirektøren i højere grad kan være nødt til at tage højde for de enkelte familiedlemmers interesser og følelsesmæssige koblinger til virksomheden – velvidende de negative effekter på bundlinjen. Men overordnet set vurderer vi, at disse faktorer ikke rører ved et grundlæggende problem i uddannelsen og forberedelsen af næste generations familiedirektører.

Vi konkluderer derfor, at et væsentligt indsatsområde er at give de potentielle kandidater i de virksomhedsdrivende familier flere målbare kompetencer. Traditionelt set har uddannelsen af nye familieledere foregået efter et slags mesterlæresystem, hvor mange kompetencer er erhvervet gennem det tætte relationspecifikke forhold til virksomhed og til virksomhedens stifter. Vi anerkender værdien af relationsspecifik humankapital, men anbefaler derudover, at dette suppleres med mere målrettede lederuddannelser og målsætninger om at opnå ledelseserfaring fra andre virksomheder.

Litteratur

- Anderson, R. C. og D. M. Reeb. 2003. Founding-family ownership and firm performance: Evidence from the S&P 500, *Journal of Finance* 58, 1301-28.
- Anderson, R. C., S. A. Mansi og D.M. Reeb. 2003. Founding family ownership and the agency cost of debt, *Journal of Financial Economics* 68, 263-85.
- Bennedsen, M. og K. M. Nielsen. 2008. Vækst via generationsskifte, Rapport offentliggjort af CEBR og Erhvervs- og Boligstyrelsen.
- Bennedsen, M., K. M. Nielsen, F. Pérez-González og D. Wolfenzon. 2007. Inside the family firm: The role of families in succession decisions and performance, *The Quarterly Journal of Economics* 122, 647-91.
- Bennedsen, M., K. M. Nielsen og D. Wolfenzon. 2005. The family behind the family firm: Evidence from successions in Danish firms, *CEBR Discussion Paper* 2005-15.
- Bennedsen, M. og D. Wolfenzon. 2000. The balance of power in closely held corporations, *Journal of Financial Economics* 58, 113-39.

- Bertrand, M., S. Johnson, K. Samphantarak og A. Schoar. 2008. Mixing family with business: A study of Thai business groups and the families behind them, *Journal of Financial Economics* 88, 466-98.
- Burkart, M., F. Panunzi og A. Shleifer. 2003. Family firms, *Journal of Finance* 58, 2167-2202.
- Casselli, F. og N. Gennaioli. 2003. Dynastic management, *NBER Working Paper* No. 9442.
- Chami, R. 2001. What is different about family businesses?, *IMF Working Papers* 01/70, International Monetary Fund.
- Gersick, K. E., J. A. Davis, M. M. Hampton og I. Lansberg. 1997. *Generation to generation: Life cycles of the family business*, Harvard Business School Press.
- Imbens G. W. og J. D. Angrist. 1994. Identification and estimation of local average treatment effects, *Econometrica*, 62, 467-75.
- Lansberg, I. S. 1999. *Succeeding generations: Realizing the dream of families in business*, Harvard Business School Press.
- Pérez-González, F. 2006. Inherited control and firm performance, *The American Economic Review* 96, 1559-88.
- Villalonga, B. og R. Amit. 2004. How do family ownership, management and control affect firm value?, *Journal of Financial Economics* 80, 385-417.
- Ward, J. L. 1987. *Keeping the family business healthy: How to plan for continuing growth, profitability and family leadership*, Jossey-Bass, San Francisco, USA.

DET DANSKE VELFÆRDS- SAMFUNDS HISTORIE 1-2

Tiden efter 1970

Af Gunnar Viby Mogensen

En økonomisk, demografisk og socialpolitisk kortlægning af velfærdssamfundets udvikling og udfordringer

»Bogens indholdsrige, vel dokumenterede sider sætter socialpolitikken ind i en meget større sammenhæng, som giver læseren rig lejlighed til at glæde sig over opnåede resultater, ærgre sig over spildte muligheder og bemærke de advarselsslamper, der blinker.«

ERLING OLSEN, JYLLANDS-POSTEN

»En grundig og velunderbygget analyse.«

POUL CHR. MATTHIESSEN,
KRONIK I BERLINGSKE TIDENDE

»Veldokumenterede værk ...
En tankevækkende analyse ...
Imponerende sammenfatning.«

FREDE VESTERGAARD, WEEKENDAVISEN

»Nyt pragtværk om
velfærdssamfundet.«

OLE OLESEN, KRISTELIGT DAGBLAD



GUNNAR VIBY MOGENSEN
Det danske velfærds-
samfunds historie 1-2
kr. 499,- (vejl.)

KØB DEM HOS DIN BOGHANDLER · twitter.com/gyldendal

GYLDENDAL 
www.gyldendal.dk

Det danske vækstmirakel

Carl-Johan Dalgaard

Økonomisk Institut og EPRU, Københavns Universitet, E-mail: Carl.Johan.Dalgaard@econ.ku.dk

SUMMARY: Why is Denmark today among the richest nations in the world? The present article develops a theory of »the Danish Growth Miracle«. It is hypothesized that the exceptional coastal orientation of Denmark led to cultural openness to new ideas, which generated rapid adoption of ideas and technologies, and ultimately enabled a comparatively early growth take-off in the 19th century. An important implication of the theory is that services was the key sector driving Danish growth in the very long run. The central mechanisms of the proposed theory receives support in cross-country data.

1. Introduktion

Danmark hører til blandt verdens rigeste lande. I 2007 – året før finanskrisen brød ud – var Danmark det 17. rigeste land i verden målt på (købekraftsjusteret) BNP per indbygger, jf. tabel 1.

Et kig på top 20 i sin helhed gør det klart, at der er tale om en broget skare af lande der udgør de mest velstående: Nogle af landene kan tilskrive deres velstand til olie og naturgas (f.eks. Qatar, Forenede Arabiske Emirater), andre finansiel service (f.eks. Bermuda), mens endnu andre lande måske snarere er drevet af industrien. Der er tydeligvis »mange veje der leder til Rom«; der er mange opskrifter på velstand.

Men overordnet er der dog et gennemgående tema at spore: hovedparten af verdens rigeste lande i 2007 var, de fremhævede nationer til trods, europæiske. Mere præcist er 11 ud af de 20 lande europæiske; 14 hvis man også anser USA, Canada og Australien for at være af »europæisk oprindelse«. Blandt disse nationer finder vi altså Danmark.

Men kan Danmark så bevare sin placering i toppen af den globale indkomstfordeling i tiden fremover? Hvor skal væksten komme fra? Det er spørgsmål der p.t. debatteres intenst. Nærværende artikel søger at bidrage til denne debat. Dog ikke ved at gætte på fremtiden, men ved at skue bagud. For ganske som det er svært at forstå nu-

Dette manuskript dannede baggrund for min professortiltrædelsesforelæsning, ved Økonomisk Institut, Københavns Universitet. Jeg vil gerne takke Thomas Barnebeck Andersen, Jacob Gyntelberg, Henrik Hansen, Ingrid Henriksen, Nikolai Kaarsen, Jakob Brøchner Madsen, Karl Gunnar Persson, Paul Sharp og Annette Højen Sørensen for nyttige kommentarer til forskellige udkast til nærværende manuskript. Redaktøren af *Nationaløkonomisk Tidsskrift*, Christian Hjorth-Andersen, takkes også for kommentarer. Fejl og mangler er alene mit ansvar.

Tabel 1. Top 20: BNP per capita.

Nr.	Land	BNP per capita 2007
1	Qatar	88320,78
2	Luxembourg	77766,19
3	Forenede Arabiske Emirater	51342,68
4	Brunei	50568,99
5	Bermuda	48867,42
6	Norge	48391,18
7	Singapore	44598,62
8	USA	42897,42
9	Kuwait	42061,18
10	Irland	41635,31
11	Island	38196,59
12	Schweiz	37308,82
13	Australien	36302,66
14	Canada	36166,12
15	Østrig	36026,75
16	Holland	34391,35
17	Danmark	34287,45
18	Belgien	33794,14
19	Sverige	32952,04
20	Finland	32481,02

Ann: (i) Data er fra Penn World Tables version 6.3, og er dermed givet i købekraft justerede US\$. (ii) Macao og Hong Kong optræder som selvstændige territorier i PWT, men er her udeladt, da de begge udgør kinesisk territorium. Begge ville dog have optrådt i top 20.

tiden, hvis man ikke forstår fortiden, forekommer det at være vanskeligt at gisne om den fremtidige danske vækstproces uden en vis fornemmelse for den proces, der gik forud. Af den årsag søger nærværende analyse at konstruere et samlet billede af den danske historiske vækstproces, der danner basis for nutiden og dermed repræsenterer udgangspunktet for fremtiden.

Hvordan kan man så forklare, at Danmark hører til verdens rigeste lande? På overfladen ligner det nærmest et lille mirakel. Det historiske Danmark er et land af begrænset størrelse; området har ikke været rigt på betydningsfulde naturressourcer, og Danmark har altid været omsluttet af magtfulde samfund der mellem år og dag har haft rig mulighed for at tage vores selvstændighed fra os (og har prøvet). Hvordan har Danmark formået at blive rigt under disse omstændigheder?

Inden for den økonomiske vækstofforskning er der en indflydelsesrig litteraturgren, der søger at give svaret på denne type af spørgsmål. Præmissen er, at man med fordel kan analysere komparative forskelle i (meget) langsigtede udviklingstemi ud fra landeforskelle i tre dimensioner: geografi, kultur og institutioner. Disse faktorer skal op-

fattes som »fundamentale« for et lands udvikling, idet de afgørende påvirker hvilke produktionsformer, der er (fysisk) mulige at implementere og disses effektivitet; de øver vigtig indflydelse på, hvor meget »smag« befolkningen har for f.eks. at spare op, investere i uddannelse, og holde fri; de berammer »spillereglerne« inden for hvilken produktion, investering og handel finder sted, og de fastlægger de transaktionsomkostninger, der er forbundet med økonomisk aktivitet. Ultimativt, lyder argumentet, er det disse underliggende faktorer, der i fællesskab forklarer innovationstempo og investeringsniveau, hvilket i sin tur bestemmer indkomsten per indbygger via landets forsyning af kapital (fysisk og human) samt den makroøkonomiske effektivitet (»allokeringsmæssig effektivitet« og teknologisk stade).¹ Af de tre nævnte fundamentale determinanter er det naturligvis alene geografi, der kan anses for at være udefra given, mens kultur og institutioner er menneskeskabte og dermed afledt af historiske processer.

Så spørgsmålet er, om Danmarks økonomiske succes ultimativt har sin basis i geografiske forhold? Nedenfor argumenterer jeg (indirekte) for, at svaret er bekræftende. Jeg vil i al fald forsøge at skitsere en mulig årsag-virknings kæde med afsæt i geografiske forhold, der kan have ledt Danmark til velstand.

Processen jeg beskriver nedenfor spænder ca. 7.000 år; fra tidspunktet hvor der er rimelig sikre indikationer på (varig) beboelse af det område, vi i dag kalder »Danmark«, og op til nutiden. Så det giver sig selv, at jeg oversimplificerer ud i det ekstreme. Der er mange faktorer, som jeg her ser bort fra, der kan have øvet indvirkning på den danske vækstproces op igennem historien. Men det er nu engang nødvendigt at forsimplere for at kunne skitsere hovedtrækkene i en udviklingsproces, der måske også peger fremad. Og fra tid til anden kan det være nyttigt at studere de (meget) brede udviklingstræk frem for alene at fokusere på detaljen. Ikke mindst i en tid hvor rammerne for den fremtidige danske vækstproces søges afstukket.²

I korte træk argumenteres der som nævnt for, at »det danske vækstmirakel« ultimativt har sit udspring i geografiske forhold. Primært dette, at Danmark – komparativt set – er en udsædvanlig kystvendt nation. Denne initialbetingelse var afgørende for, at Danmark blev en handelsnation; og at landet blev en handelsnation ledte til et samfund med indbyggere, der var åbne over for nye ideer. Eller sagt på en anden måde: afkastet på (ny) viden blev anset, i Danmark, for at være højt. Den kulturelle værdsættelse af

1. Se Acemoglu (2009, kapitel 4) for en koncis oversigt over denne litteratur. Opsplitningen af den makroøkonomiske effektivitet (»totalfaktorproduktivitet«) på allokeringsmæssig effektivitet og teknologisk stade er foreslået i Weil (2005, Kapitel 10), og kvantificeret i Chanda og Dalgaard (2008) samt Vollrath (2009).

2. Set over 7 årtusinder er »Danmark« naturligvis ikke noget specielt veldefineret geografisk begreb. Det er derfor værd at præcisere, at når der refereres til »Danmark« i det følgende, da er det »kerneterritoriet« der sigtes til. Altså dét område, der falder inden for de nuværende territorialgrænser. Grønland og Færøerne (såvel som Norge, Island, brudstykker af Baltikum og nutidens Sverige) regnes altså ikke som »danske«.

viden har i sin tur været af stor betydning for *timing*en for adoptionen af obligatorisk skolegang; en idé der generelt må antages at have sit udspring i den Protestantiske Reformation såvel som i Oplysningstiden (derfor: adoption). Endelig, dette at Danmarks befolkning mod slutningen af det 19. århundrede således var forholdsvis veluddannet, gjorde det muligt at drage nytte – i stor skala – af de innovationer, der så dagens lys i forbindelse med den industrielle revolution, der havde sit epicenter i landene omkring Danmark. Konsekvensen var, at Danmark kom med »på bagsmækken« da industrialiseringen tog fart. Teknologiadoptionen fandt altså sted fordi den kulturelle værdsættelse af nye ideer indebar en *villighed* til at lære udefra, mens den (delvist afledte) formelle uddannelse gjorde det *muligt* at afkode de nye teknologier og derved implementere dem i produktionen.

Det tidlige »afsæt« til økonomisk vækst, der altså fandt sted i Danmark og i en række andre europæiske lande udgør hovedforklaringen på, at europæiske lande (heriblandt Danmark) i dag (endnu) udgør toppen af den globale indkomstfordeling (jf. tabel 1).³

I et forsøg på at afdække om teorien har en vis ekstern validitet (altså kan motivere mere og andet end Danmarks relativt høje nutidige velstand), undersøges centrale mekanismer og implikationer af den foreslåede teori ved brug af tværlandedata. Det fremgår bl.a., konsistent med teorien, at lande, der geografisk er udstyret med en lang kystlinje (i forhold til samlet areal) i dag har befolkninger, der kulturelt værdsætter viden og nye ideer, tidligt investerede betydeligt i uddannelse og i dag er mere velstående end lande med mindre adgang til havet. Sammenhængen mellem relativ kyststrækning og arbejdskraftproduktivitet synes at være forklaret ved førnævntes indflydelse på uddannelse.

Analysen er struktureret som følger. I afsnit 2 fremlægges en teori for den danske (meget langsigtede) økonomiske udvikling. Denne beskriver, hvordan dette, at det historiske Danmark er kystvendt, er den underliggende årsag til vores nuværende velstand. I afsnit 3 diskuteres teoriens eksterne validitet, mens afsnit 4 afrunder ved kort at ridse implikationerne af historien op for den igangværende debat om den danske vækstproces.

2. Det danske vækstmirakel: en teori

Det er velkendt for alle, at Danmark består af øer, samt en halvø. Hvad der måske er mindre velkendt er, at bortset fra Singapore er Danmark (i sin nuværende form) det land i verden der har den længste kystlinje relativt til samlet areal; tabel 2 opregner de 10 mest kystvendte nutidige nationer i verden, målt ved relativ kyststrækning.⁴

3. Se Dalgaard og Strulik (2010) for en empirisk analyse af betydningen af det historiske afsæt til vækst for den nuværende globale indkomstfordeling.

4. Her tillader jeg mig at se bort fra miniput stater med et indbyggertal på under 1 million.

Tabel 2. Top 10: Længst kyststrækning ifht areal.

	Land	Kyst/areal	Befolkning i 2000 (1000er)
1	Singapore	0,27	4036
2	Danmark	0,17	5337
3	Filippinerne	0,12	81222
4	Kroatien	0,10	4410
5	Grækenland	0,10	10559
6	Jamaica	0,09	2615
7	Estland	0,09	1379
8	Mauritius	0,09	1185
9	Norge	0,08	4492
10	Japan	0,08	126729

Kilder: CIA factbook, World Development Indicators.

Anm: Listen opregner de 10 mest kystvendte nationer med mere end 1 million indbyggere i 2000.

I forhistorisk tid var forholdet mellem kystlinje og areal faktisk endnu større, f.eks. Enghoff (1994). Det er dermed næsten en triviell konstatering, at menneskene der har beboet det område, der blev til Danmark, har haft let adgang til havet. Triviell observation eller ej, så er det en vigtig initialbetingelse. For det betyder, at befolkningen på disse breddegrader har færdedes til søs i årtusinder.

Initialt har motivet utvivlsomt været jagt og fiskeri. Det er der i al fald meget der tyder på var tilfældet. Ved brug af kulstof 13 analyser af knogler fra tidlige beboere af det danske område er det nemlig muligt at klarlægge, hvordan diæten var sammensat helt tilbage til jægerstenalderen. Herved er det blevet fastslået, at før landbrugets indtog bestod diæten i overvejende grad af fisk fra havet, Tauber (1981); Enghoff (1994).⁵ Betydelige fund af forskellige former for fiskeudstyr, såvel som efterladte køkkenmøddinger, understøtter disse analyser. Selv efter den neolitiske revolution lader det til, at fisk blev ved med at være en vigtig del af diæten, om end det præcise omfang er omstridt, f.eks. Fischer (2007).

Denne færdsel på havet har gradvist ledt til akkumulation af viden om skibskonstruktion og navigation. En god illustration af en sådan underliggende innovationsproces i forhold til bådebyggeri er den gradvise udvikling af maritime fartøjer fra en spæd begyndelse ca. 5000 f.Kr. i form af relativt primitive kanoer til de frygtede dragehovedbesatte langskibe (en ægte skandinavisk opfindelse) fra omkring år 1000 e.Kr., f.eks. Hale (1998). Disse innovationer og den underliggende viden satte vikingerne i stand til – som de første europæere – at sætte deres fod på (nord)amerikansk jord århundredere før Columbus, f.eks. Roesdahl (1998).

5. Jægerstenalderen, eller »Ertebøllekulturen«, spænder perioden fra ca. 5400 f.Kr. – 3900 f.Kr. Den neolitiske revolution – overgangen fra jæger-samler samfundet til landbrugssamfundet – finder efter alt at dømme gradvist sted i Danmark over perioden 3900-3000 f.Kr. Se f.eks. Richards m.fl. (2003).

Med skibe og navigation følger muligheden for handel; også handel uden for de nuværende grænser for landet. Anførte rationale og de arkæologiske fund tyder dermed på, at vi som nation har haft et betydeligt *potentiale* for at blive en nation drevet af handel.

Det turde være klart, at dette potentiale blev realiseret. I al fald er der rigelig evidens for handel mellem mennesker i Danmark og omverdenen, selv fra de tidligste tider. Helt tilbage til jægerstenalderen er der i al fald keramiske fund der tyder på import fra Centraleuropa, Richards m.fl. (2003). Andre arkæologiske fund, fra samme periode dokumenterer eksport af flintesten; fortrinsvis til det øvrige Skandinavien. Men dansk flintesten er nu også fundet i udgravninger syd for den nuværende grænse i f.eks. Holland, f.eks. Becker (1959); Sarauw (2008). Meget af denne handel har dog nok været af indirekte natur.

Da vi når Bronzealderen i det 2. årtusinde f.Kr. udveksles særligt forarbejdede bronzevarer mod lokalt rav; en vareudveksling der udvikler sig støt. Omkring 1500 f.Kr. kan man således påvise, at rav fra danske egne har fundet vej helt ned til Middelhavet, f.eks. Jensen (1983). En række andre fund tyder på videns- og vareudveksling med den omkringliggende verden. Det er således i samme periode, at man ser introduktionen af tohjulede hestevogne, der er identiske med typen, der blev fundet i Tutankhamuns grav; ligeledes er der i denne periode fundet særlige klapstole, der er identiske med fund i bronzealderens Ægypten og Grækenland, Randsborg (2009, s. 21-23).

Udviklingen tager efter alt at dømme fart op igennem vikingetiden, Ulriksen (1994); Hybel (2003), og det er nok velkendt for de fleste, at handel og søfart har spillet en betydelig rolle i Danmark i historisk tid. Den dag i dag udgør samhandlen med omverdenen (import plus eksport) et omfang, der rundt regnet svarer til Danmarks samlede BNP. Udenrigshandlen er altså fortsat uhyre vigtig for Danmark.

Men hvad danner baggrunden for denne tingenes tilstand? Hvad er årsagen til, at potentialet for samhandel, der altså tænkelig er en konsekvens af at landet er kystvendt, blev realiseret?

En del af forklaringen ligger utvivlsomt i et andet geografisk karakteristiskum, der på overfladen virker som et handicap: vores mangel på naturressourcer (da først flintesten var udrangeret). Denne mangel på vigtige råstoffer må historisk have gjort det nødvendigt at interagere med andre lande, hvis danskerne skulle holde trit med sine naboer; nød lærer som bekendt nøgen kvinde at spinde.⁶ Det er i al fald veldokumenteret, at befolkningen i det, der blev Danmark, importerede tin, kobber, bronze og jern af

6. Eller som Henriksen (1992, s. 156) udtrykker det i en kontekst af dansk handelspolitik, på et meget senere tidspunkt i historien: »[by] maintaining free trade, the Danes adhered to a national tradition of liberalism, a reflection of a small economy without any domestic mineral resources«.

høj kvalitet, Randsborg (2009). Behovet for at erhverve sig naturressourcer må have været en motivation for samhandlen.

En anden del af forklaringen er, at Danmark er behændigt placeret i en region, der meget tidligt viste sig særlig dynamisk: Europa. Der var derfor også noget at rejse ud efter.⁷

Endelig, dette at Danmark er en forholdsvis lille nation med et deraf afledt begrænset hjemmemarked, har formodentlig forstærket incitamentet til at rejse ud. Der er således flere årsager til, at Danmark realiserede sit handelspotentiale.

I en vis forstand har Danmark dermed nok altid har været »globaliseret«. Men hvad der nok er vigtigere for udviklingshistorien er, at siden landets befolkning blev handlende, måtte det danske samfund nødvendigvis blive *vidensorienteret*.

At Danmark var vidensorienteret skal naturligvis forstås i sin historiske kontekst; den danske befolkning var selvsagt ikke uddannet i en moderne forstand. Men informationsindsamling og dette at værdsætte ny viden, må nødvendigvis være blevet centralt for overlevelsen. For (specielt) set i et *førindustrielt* perspektiv er handelserhvervet uhyre videns intensivt. Det kræver viden at overskue, hvad andre mennesker ønsker at forbruge; hvordan man tilvejebringer varer, man ønsker at bytte for andre varer, som tredjepart har interesse i; hvordan man sikrer sig imod at blive snydt, alt imens man interagerer på fremmedsprog og i en kontekst af en anden kultur (om end, måske ikke altid én, der var milevidt fra vores egen). Hertil kommer behovet for visse formelle færdigheder, såsom basal algebra til brug for fremstilling af f.eks. regnskaber.⁸

Det er værd kort at dvæle ved postulatet om, at befolkningen i Danmark var vidensorienteret, tænkeligt som en konsekvens af det geografisk betingede behov for at handle. Ved vi, at det er rigtigt? Nej, men det er muligt at finde indikationer på, at der kunne være noget om snakken.

En sådan tidlig indikation er at finde i tiden omkring overgangen til landbruget; 4000-3000 f.Kr. Allerede for denne periode har arkæologerne gjort fund, der tyder på en ganske kompleks samfundsstruktur: imponerende gravhøje der har krævet betydelig ingeniør viden for at kunne lade sig konstruere, såvel som rituelle middelstore bebyggelser af forskellig art. Disse funds beskaffenhed og formål antyder for arkæo-

7. Dette at det europæiske kontinent historisk var dynamisk kan i øvrigt også gives en »bio-geografisk« forklaring; se Diamonds (1997) fængslende bog om emnet. Se også Persson (2010) for en interessant teori for den europæiske udvikling fra 600 og frem.

8. Se Dittmar (2010) for en interessant diskussion af denne pointe. Dittmar viser, at bogtrykkerkunsten i betydeligt omfang initialt blev brugt til at sprede viden om basal algebra til glæde for det spirende handelserhverv. Og konsistent med formodningen om værdien af bogtrykkerkunsten for handelserhvervet finder Dittmar, at introduktionen af trykkemaskinerne havde størst effekt på væksten i bybefolkningerne, når denne fandt sted i havnebyer, hvor handlen må ventes at have foregået mest intenst.

logerne, at i det daværende »danske« samfund var det tilfældet at: »*Leadership must have been linked to knowledge and technological performance*«, Randsborg (2009, s. 15).

For at opnå en yderligere indikation på appetitten for ny viden i Danmark må vi rejse til historisk tid, hvor det bliver muligt at studere teknologiadoption i bl.a. Danmark. I et nyere studie udvikler Comin m.fl. (2010) således mål for graden af teknologi adoption før industrialiseringen. Konkret kompilerer forfatterne en liste af teknologier, der kan dokumenteres at være opfundet, bl.a. i år 1500, hvorefter de undersøger, hvor stor en andel af disse teknologier, der var adopteret i verdens lande i samme periode. Som forfatterne selv gør opmærksom på, er der selvsagt en række problemer forbundet med denne slags forehavender, hvorfor det absolutte niveau for det resulterende adoptionindeks måske ikke skal tages alt for håndfast. Men det bør dog være tilstrækkeligt retvisende til at kunne give en vis indikation af, hvor Danmark tidligt var placeret, relativt til andre (europæiske) lande.

Den centrale pointe man kan trække ud af Comin et al's studie er (i nærværende sammenhæng), at Danmark allerede i 1500 synes at være langt fremme teknologisk, vurderet på nævnte indeks; stort set kun overgået af stormagterne på daværende tidspunkt: Spanien, Portugal og England. Dette støtter den grundlæggende hypotese, udviklet ovenfor, at der i Danmark tidligt var en forståelse for det vigtige i at tilegne sig ny viden. Og en villighed til at tage den i anvendelse.

Ovenstående teori vil altså vide, at fundamentale geografiske karakteristika for Danmark ledte til søfart, handel og derved øgede det oplevede afkast på viden. Dette sidstnævnte – kulturelle – karakteristikum har utvivlsomt spillet en central rolle i forhold til indførelsen af obligatorisk skolegang.

Ideen om det attraktive i en uddannet befolkning er naturligvis ikke på noget måde af dansk oprindelse. Den skal nok snarere tilskrives idémæssige strømninger udefra: den Protestantiske Reformation (gennemløbes i Danmark omkring 1530), samt Oplysningstiden (1690-1800).

At protestantismens indførelse har spillet en rolle i forhold til uddannelse er påpeget af Landes (1999). Det centrale aspekt ved den Protestantiske Reformation (i nærværende sammenhæng) er, at den troende skulle være i stand til at læse biblen på sit eget sprog. Dette religiøse dogme har tænkeligt haft en (utilsigtet) gunstig effekt på vækstprocessen, da disse formelle færdigheder senere viste sig værdifulde i forbindelse med den »industrielle revolution«. I et nyere studie finder Becker og Woessmann (2009) empirisk støtte for denne hypotese ved at sammenligne væksten i tyske regioner, der gennemløb Reformationen, med de, der blev ved med at være katolske. De pro-

testamentiske regioner er i dag rigere end de katolske, tilsyneladende pga. en systematisk forskel i uddannelsesniveaet i protestanternes favør.⁹

Udover Reformationen har de idealer, der voksede frem i forbindelse med Oplysningstiden i verden omkring Danmark, utvivlsomt medvirket til et øget fokus på uddannelse, f.eks. Easterlin (1981).¹⁰

Disse strømninger har naturligvis påvirket snart sagt alle de europæiske lande. Alligevel vandt uddannelsesreformerne ikke indpas lige hurtigt alle steder, f.eks. Easterlin (1981). Når det sker i Danmark, forekommer det sandsynligt, at det kan forklares ved det skitserede historiske forløb, som ledte til, at afkastet af vidensakkumulation blev anset af danskerne (ganske tidligt) for at være højt. Det er i al fald et faktum, at uddannelsesreformerne i Danmark synes at følge hastigt i kølvandet på såvel Reformationen som Oplysningstiden.

Den gradvise udbygning af et egentligt skolesystem er naturligvis en langsommelig proces, der dog kan siges at tage sin begyndelse med »Kirkeordinansen« fra 1537 (lige efter Reformationen). Denne forordning bød, at der i alle købstæder skulle oprettes skoler hvor børn (læs: drenge) kunne modtage en teologisk grunduddannelse. Det næste store fremskridt finder sted i 1700 tallet (midt under Oplysningstiden), hvor man begynder at se tegningen til et sekulært skolesystem. Særligt er »Forordningen om undervisningspligt på landet« fra 1739 værd at bemærke; herefter skulle alle børn fra 5-6 års alderen modtage undervisning. Den danske folkeskole anses dog først for at have sin spæde oprindelse i 1814, med undervisningspligt såvel på land som i by frem til konfirmationsalderen.

Den historisk betingede åbenhed over for ny viden, der altså kan have været medvirkende til udbygningen af det formelle skolesystem, er utvivlsomt central for en forståelse af, hvorfor Danmark kunne blive blandt de lande, hvor vækstprocessen startede tidligst. Hvor det første karakteristikum indebærer en *villighed* til at adoptere nye ideer, betyder det andet (måske delvist afledte) karakteristikum, at der var *mulighed* for at få nye teknologier implementeret bredt i samfundet, da verden for alvor forandres i forbindelse med den industrielle revolution. I sidstnævnte henseende var det selvsagt ikke danskerne, der bidrog med de nye innovationer. Men siden Danmark iboende var åbent over for nye ideer, og i tilgift forholdsvis veluddannet, vandt de nye ideer og tek-

9. Ifølge Weber (1905) skulle Reformationen også have ledt til en højere værdsættelse af hårdt arbejde og opsparing; »Den protestantiske etik«. Som vist i Andersen, Bentzen, Dalgaard og Sharp (2010) har disse værdier imidlertid et dybere fundament i Kristendommen, hvorfor Reformationen tænkeligt ikke har haft nogen kataklysmisk effekt af dén årsag; konsistent med Becker and Woessman's (2009) fund. Det skal samtidig understreges, at de værdier, Weber fremhævede synes at have en effekt på væksten, men at værdiforandringen altså synes at begynde tidligere end, hvad Weber antog, Andersen m.fl. (2010).

10. Se Mokyr (2005) for en grundig analyse af Oplysningstidens betydning for vækst processen i almindelighed.

nologier forholdsvis hurtigt indpas. De nye teknologier vandt først vej til landbruget, der gradvist blev »automatiseret«; senere vandt industrien også indpas.

Dette bringer os frem til nutiden. Ifølge den foreslåede teori er Danmark altså rigt, siden landet tidligt begyndte at uddanne befolkningen, hvilket tillod hastig teknologi-adoption. Den tidlige satsning på uddannelse kan i sin tur tilskrives langsigtede endogene processer; specielt er fremhævet det faktum at Danmark er særligt kystvendt, lille og fattigt på naturressourcer, hvilket affødte (international) handel og international interaktion, hvilket ledte til en kulturel værdsættelse af, og åbenhed over for ny viden. Skematisk kan hovedpunkterne i den fremsatte teori opsummeres således:

- Geografi (kystlinje/areal) → handel
- Kulturel åbenhed overfor nye ideer (/højt afkast på viden)
- Tidlig (formel) humankapital akkumulation
- Høj indkomst i dag,

hvor teknologiadoption er implicit i den sidste pil.

Inden vi vender os mod teoriens eksterne validitet, er det nok værd kort at berøre landbruget. For det kan måske virke underligt, at dette erhverv ikke er blevet tildelt nogen væsentlig rolle i fortællingen ovenfor. Dels fordi landbruget uomtvistelig har spillet en væsentlig rolle for beskæftigelsen og BNP, og dels fordi der trods alt ikke findes et land i verden, der i dag er rigt uden at have gennemløbet den neolitiske revolution (altså overgangen fra jæger-samlersamfund til landbrug). Den sidste observation synes således at tilsiige, at landbrugsproduktion er en slags nødvendig historisk betingelse for nutidig velstand.

Men samtidig er det nok også velkendt, at der er mange lande, der har gennemgået den neolitiske revolution, uden disse (endnu) er blevet velstående på et niveau svarende til det danske af i dag, hvorfor landbrugsproduktion ikke kan være en *tilstrækkelig* betingelse for nutidig velstand. Set i dét lys giver det derfor en vis mening, at arbejde ud fra den grundhypotese, at landbruget ikke var den afgørende drivende kraft bag det danske vækstmirakel.

Det betyder selvfølgelig ikke, at landbruget var uvæsentligt; landbruget har været en vigtig medvirkende, supplerende kraft bag den danske økonomiske udvikling. Men det forekommer altså rimeligt at tro, at det ikke var landbrugsproduktion, som sådan, der gjorde landet velstående. Der må have været andre faktorer, der drev udviklingen og var medvirkende til produktivitetsudviklingen også i landbruget. Udviklingsteorien ovenfor peger på en mekanisme, der løber fra geografi via handel og kulturel åbenhed over for nye ideer til uddannelse. Under denne teori vil landbrugets succes i Danmark derfor netop skulle tilskrives, at den formelle uddannelse vandt frem; en hypotese der

finder støtte i forskningen inden for økonomisk historie, se Henriksen, (1999); Henriksen m.fl. (2010).

3. Ekstern validitet

Det er selvsagt altid muligt at konstruere en teori for et lands udvikling, der kan hævdes at være logisk sammenhængende og dermed er »internt valid«. Lakmusprøven for en sådan teori er, om den også kan siges at have forklaringskraft i forhold til andet, end at netop Danmark blev velstående. Har teorien nogen ekstern validitet? Dette spørgsmål adresseres i dette afsnit, der falder i tre underafsnit.

Ifølge den foreslåede teori skal Danmarks velstand ultimativt tilskrives, at landet er usædvanligt kystvendt. Men hvordan er det gået med de andre særligt kystvendte nationer fra tabel 2? Er disse nationer, som Danmark, økonomisk succesfulde i dag? Dette spørgsmål undersøges i afsnit 3.1.

Hernæst, i afsnit 3.2, udstrækkes analysen til at omfatte et langt større antal lande, hvorigennem de centrale fremhævede mekanismer kan undersøges ved brug af regressionsanalyse.

Endelig diskuteres en vigtig bi-implikation af udviklingsteorien. Nemlig den, at den danske økonomiske vækst i betydeligt omfang må tilskrives *servicesektoren*. Søtransport og »handel« er centrale mellemliggende determinanter for den økonomiske vækst, der er drevet af geografi og selv driver værdsættelsen af ny viden. Siden disse erhverv er ledende serviceerhverv, tilsiger teorien, at Danmark (og andre kystvendte samfund) bør have en relativt stor servicesektor; denne prædiction undersøges i afsnit 3.3.

3.1. Er de mest kystvendte nationer økonomisk succesfulde?

Ved at sammenligne tabel 1 og 2 står ét klart, at tre af verdens mest kystvendte nationer også er at finde i top 20 i den globale indkomstfordeling: Singapore, Danmark og Norge (der som bekendt var særdeles velstående allerede, da landet fandt olie i 1969). Andre kystvendte stater fra tabel 2 er tæt på top 20 i indkomstfordelingen. Japan var således nr. 24 i 2007, mens Grækenland nåede en 27. plads i samme år. Det er her værd at betænke, at i denne del af indkomstfordelingen er indkomstforskellene ikke voldsomt store. I 2007 var Japans BNP per capita således blot 6% lavere end Finlands, som indtager 20. pladsen i tabel 1. Tilsvarende havde Grækenland 12% lavere BNP per indbygger end Finland i 2007. Alle disse samfund må siges at være særdeles økonomisk succesfulde i en global kontekst.

Andre lande fra tabel 2 er imidlertid længere væk fra top 20 i indkomstfordelingen. Afrikanske Mauritius har verdens 8. længste relative kystlængde, og kan ikke siges at være velstående som Danmark, Grækenland, Japan, Norge og Singapore. Men Mauri-

tius er faktisk det næsthøjest placerede afrikanske land i den globale indkomst fordeling; I 2007 indtog landet en 45. plads, med et indkomstniveau, der er ca. 60% af Finlands.¹¹ Til sammenligning ligger Sydafrika på en 73. plads, mens Botswana – der ofte fremhæves som en sjælden afrikansk »tiger økonomi« – er placeret som nr. 85 med et (købekraftsjusteret) BNP per capita på mindre end halvdelen af Mauritius. Det forholdsvise høje indkomst niveau på Mauritius modsvarer også af en relativt veluddannet befolkning: i 2005 kunne en nyfødt på Mauritius således forvente 14 års skolegang. Det tilsvarende tal er 17 år i Danmark.¹² Givet Mauritius regionale placering, i den tropiske klimazone langt fra verdens innovative centre (både historisk og i dag), har det klaret sig særdeles godt. Flertallet af de kystvendte nationer i tabel 2 er dermed indkomstmæssigt succesfulde i dag.

Umiddelbart kan de resterende lande i tabel 2 dog kun vanskeligt anses for at være specielt succesfulde: Estland (nr. 47 i indkomstfordelingen), Kroatien (nr. 62), Jamaica (nr. 88) og Filippinerne (nr. 121). De to første landes placering er måske specielt bemærkelsesværdig, siden de begge er placeret i Europa. Ved nærmere øjesyn er der imidlertid noget, der binder disse fire lande sammen: de er alle blevet ugunstigt påvirket af fremmede magter.

Estland var således under skiftende dansk, svensk, tysk og russisk overherredømme i århundreder før landet – efter en kort selvstændighedsperiode 1918-40 – blev indlemmet i Sovjetunionen indtil murens fald. I de ca. 50 år landet var underlagt Sovjetunionen var det en del af et planøkonomisk system, der – i al fald set i bakspejlet – ikke var gavnligt. Kroatien var, frem til afslutningen af 1. Verdenskrig, en del af det Østrig-Ungarske rige. Efter 1. Verdenskrig blev landet en del af Jugoslavien, der ligesom Estland var en del af det planøkonomiske eksperiment. Efter Sovjetunionens kollaps endte Jugoslavien som bekendt i en bitter borgerkrig, der i 1995 førte frem til Kroatiens uafhængighed. På trods af besættelse, krig og et behov for at restrukturere hele økonomien efter Sovjetunionens sammenbrud, har både Estland og Kroatien trods alt opnået status af mellemindkomstlande, med indkomstniveauer på hhv. 60% og godt 40% af Finlands i 2007. Historien taget i betragtning er det ikke det ringeste resultat at kunne fremvise, indtil videre.

De sidste to lande, Jamaica og Filippinerne, må anses for at være decideret fattige lande. Er de dermed modeksempler på, at kystvendte nationer har en udviklingsmæssig fordel?

Jamaica blev opdaget af Columbus i 1494, hvilket ledte til, at *alle* oprindelige indbyggere på øen blev udryddet. Befolkningen blev i stedet erstattet af afrikanske slaver,

11. Det rigeste afrikanske land i 2007, i al fald blandt de der er data for i Penn World Tables 6.3., var Ækvatorialguinea (indtager 39. pladsen i den globale indkomstfordeling). Også dette lille land, med ca. 600.000 indbyggere, er kystvendt.

12. Kilden til disse uddannelsesdata er CIA factbook <https://www.cia.gov/>.

og landet opnåede først sin uafhængighed i 1962. Man kan dårligt forestille sig en mere dramatisk (og ugunstig) udenlandsk indflydelse end en, der leder til, at hele den oprindelige befolkning, og det samfund de måtte have opbygget, udslettes fra jordens overflade. At landet derfor har udviklet sig temmelig meget anderledes, sammenlignet med mange af de andre særligt kystvendte nationer, er dermed knap så stort et mysterium.

Det sidste – og fattigste – land på listen fra tabel 2 er Filippinerne, der blev underlagt Spanien i 1562. Landet opnåede selvstændighed efter afslutningen på 2. Verdenskrig. Siden den tid har Filippinerne været plaget af interne stridigheder og politisk ustabilitet.

For at forstå denne tingenes tilstand er det værd at observere, at Filippinerne, forud for vesteuropæernes ankomst slet ikke var en politisk enhed. Dette er udmærket illustreret ved den kulturelle mangfoldighed, man den dag i dag finder på Filippinerne. Øriget huser således intet mindre end 171 forskellige sprog, hvor iblandt engelsk i dag er det officielle sprog – et levn fra tiden under amerikansk overherredømme. Denne etnisk-lingvistiske heterogenitet, der altså er en konsekvens af grænsetrækningen fra kolonitiden, kan være en (del)forklaring på den observerede politiske ustabilitet, der utvivlsomt har bidraget til at fastholde landet på et lavt per capita indkomstniveau. Et meget anvendt mål for etnisk-lingvistisk heterogenitet er sandsynligheden for, at to tilfældigt udvalgte borgere tilhører *forskellige* etniske grupper. På Filippinerne er denne sandsynlighed hele 87%; til sammenligning er den ca. 5% i Danmark. Easterly og Levine (1997) var blandt de første til at påvise empirisk, at stor etnisk-lingvistisk heterogenitet (således kvantificeret) kan lede til politisk ustabilitet og derigennem være vækstskaadelig.

Samlet set synes teorien derfor at have et vist mål af ekstern validitet. Når man betragter de mest kystvendte nationer i verden, er de fleste relativt succesfulde. Og i de tilfælde, hvor det ikke er tilfældet, er der tale om lande der på markant og ugunstig vis er blevet »forstyrret« udefra.

3.2. Et konsistentstjek ved brug af tværsnitsdata

Det er selvsagt umuligt at bevise empirisk, at den fremlagte teori er »korrekt« for Danmark. Sikker viden vil kun kunne opnås ved at lave (alene) Danmarks geografiske beliggenhed og beskaffenhed om, og derpå lade historien udspille sig igen. Det lader sig selvsagt ikke gøre.

At (de fleste) andre nationer med særlig lang relativ kyststrækning har udviklet sig økonomisk gunstigt leverer måske en vis støtte for en reduceret form sammenhæng mellem velstand og kyststrækning. Men er der mere generelt en sammenhæng mellem kyststrækning og velstand? Hvad med de anførte mekanismer, hvorved relativ kyststrækning – ifølge teorien – bør lede til økonomisk udvikling?

I praksis er den eneste reelle mulighed for mere formelt at adressere disse spørgsmål at tage regressionsanalyse i anvendelse. Denne tilgang, hvor information om udviklingen af mange lande inddrages på én gang, giver en vis mulighed for tjekke den eksterne validitet. Det skal dog medgives, at der er tale om et forholdsvis svagt test. I fald man ikke kan afvise de fremhævede mekanismer, er det selvsagt *ikke* ensbetydende med, at teorien er korrekt (for Danmark). Men teoriens natur taget i betragtning er det dog nok imidlertid det stærkeste test, der lader sig gennemføre for nærværende.

I dette afsnit vil jeg derfor se nærmere på tre centrale spørgsmål ved brug af tværsnitsdata:

1. *Kyststrækning og kulturel værdsættelse af nye ideer.* Er der noget der tyder på, at kystlængde relativt til areal leder til kulturel værdsættelse af ny viden?
2. *Kyststrækning og uddannelse.* Er der noget der tyder på, at kyststrækning relativt til areal øger *tidlig* humankapital investering?
3. *Kyststrækning og nutidig indkomst.* Er der noget der tyder på, at relativ kyststrækning har ledt til højere nutidig indkomst? *Via* tidlig humankapital akkumulation?

Man kunne måske også ønske sig et test af, om adgang til havet har en betydning for international handel og rejseaktivitet; mekanismen der tænkes at forbinde kystlinje med kulturel værdsættelse af ideer. Når et sådan ikke foretages, har det en simpel forklaring. Inden for udenrigshandelsteorien har man siden Tinbergen (1962) arbejdet med såkaldte »gravitationsmodeller«, der beskriver bilateral samhandel primært ud fra den geografiske afstand mellem lande, men også ud fra andre geografiske karakteristika, herunder landenes areal og deres adgang til havet. Via estimation af gravitationsmodellerne er det således blevet afklaret, at store lande uden kystlinje handler (og indbyggerne: rejser) ved *lavere* intensitet end små lande med adgang til havet, se også Frankel og Romer (1999); Andersen og Dalgaard (2010). Det er således interessant at notere sig, at i en verden, hvor internettet har stadig større betydning, da er handelsstrømme altså fortsat stærkt påvirket af rene geografiske forhold. Men mere fundamentalt betyder eksistensen af »gravitationslitteraturen«, at sammenhængen mellem kystlinje og handelsintensitet (og rejseintensitet) kan anses for at være veldokumenteret, hvorfor test af denne mekanisme ikke foretages nedenfor.

3.2.1 *Kyststrækning og kultur*

En central del af teorien består i det postulat, at dette at være kystvendt ultimativt manifesterer sig i en kulturel værdsættelse af nye ideer og ny viden. Givet landets geo-

grafi påvirker handelsintensiteten, er forudsigelsen, at samfund med længere kystlængde/areal forhold burde udvikle en særlig værdsættelse af viden og nye ideer.

Denne hypotese er selvsagt vanskelig at undersøge på historiske data. Men mulighederne er en smule bedre, når det gælder samtiden. *World Value Survey* bliver jævnligt gennemført på tværs af et større udsnit af verdens lande, hvor mere end 1000 individer interviewes i hvert af de deltagende lande. Formålet er at søge at afdække forskelle i normer, værdier og præferencer på tværs af verdens lande. Den resulterende database bliver flittigt brugt i forskningen, hvorfor den må anses for at være af tilstrækkelig kvalitet i nærværende sammenhæng.

I de surveys, der blev gennemført i perioden 2005-2008, er der indføjet spørgsmål, der taler til den kulturelle værdsættelse af kreativitet og dette at producere nye ideer, hvilket synes at være en tålelig proxy for »kulturel værdsættelse af ny viden«, som fremhævet i den udviklede teori.

I interviewet præsenteres respondenterne konkret for en tænkt person der »lægger stor vægt på at være kreativ og få nye ideer«. Dernæst bedes den enkelte placere sig på en 6 punkt skala, hvor det ene yderpunkt er, at respondenterne »helt og aldeles« identificerer sig med den tænkte person, mens det andet yderpunkt er, at fornævnte ser sig selv som modsætningen til den kreative og ideproducerende person. Aggregeret får man dermed en fordeling af, hvordan individerne i stikprøven ser sig selv i forhold til den tænkte persons egenskaber (se Appendix for uddybning). I nærværende sammenhæng synes det at være rimeligt at antage, at i samfund, hvor folk identificerer sig med »kreativitet og produktion af nye ideer«, da har man udviklet en kulturel værdsættelse af disse ting. Hvorfor skulle man ellers ønske at identificere sig med slige karakteristika?

Der er imidlertid et problem med denne type af selvrapporterede værdier. Dette at være »kreativ og ideproducerende« er utvivlsomt et positivt karakteristika i de fleste samfund, hvorfor der vil være en tendens til, at respondenterne erklærer, at de er kreative og iderige, uagtet om dette er en præcis beskrivelse af tingenes tilstand eller ej.

For at søge at afkode, lidt mere klart, hvordan befolkningerne i realiteten stiller sig, inddrages derfor endnu et spørgsmål i målingen af kulturel værdsættelse af viden. Respondenterne bliver nemlig også spurgt om hvorvidt »traditioner« er vigtige. Igen er svarerne fordelt på en 6 punkt skala. Dette at »være traditionerne« er utvivlsomt også positivt værdiladet i de fleste samfund. Men på mange stræk er det modpolen til at være kreativ og dermed gøre tingene på en *ny* måde. Det er derfor sandsynligt at man kan afdække kulturel præference for kreativitet og iderigdom ved at bruge *forholdet* mellem hvor mange respondenter der identificerer sig med hhv. kreativitet og tradition. Konkret er målet for »kulturel værdsættelse af ny viden« givet ved

Tabel 3. Kultur og kyststrækning.

	-1 Kreativ/ Tradition	-2 Kreativ/ Tradition	-3 Kreativ/ Tradition	-4 Kreativ/ Tradition	-5 Kreativ/ Tradition
Log kyststrækning	0,076 (3,11)**		0,082 (2,91)**	0,072 (2,92)**	0,084 (2,85)**
Log areal	-0,075 (2,94)**		-0,073 (2,52)*	-0,072 (2,81)**	-0,077 (2,24)*
Kyststrækning		0,006 (2,88)**			
Areal		-0,008 (1,45)			
Uden Kyst (=1)		0,036 (0,45)			
Breddegrad				0,003 (1,25)	0,001 (0,34)
Regioner	Nej	Nej	Ja	Nej	Ja
Observationer	41	48	41	41	41
Kystkoefficient = - Arealkoefficient (p-værdi)	0,98	0,73	0,69	0,97	0,78
R ²	0,17	0,05	0,31	0,20	0,31

Robuste *t* værdier i parentes + signifikant, 10%; * signifikant, 5%; ** signifikant, 1%.

$$C = \frac{k1 + k2 + k3}{t1 + t2 + t3}$$

hvor *k1* til *k3* er procentandelen af respondenter, der falder i de tre kategorier, hvor man mestendels identificerer sig med kreativitet og ide rigdom, mens *t1* til *t3* er tilsvarende i forhold til dette at ære traditionerne.¹³

Med disse data i hænde kan et (groft) tjek af den første mekanisme gennemføres. Spørgsmålet er, om meget kystvendte samfund i højere grad end mindre kystvendte samfund scorer højere værdier på »C indekset«, som defineret ovenfor.

Konkret estimeres følgende model:

$$C_i = a + b \cdot \log(\text{kyststrækning}_i) + c \cdot \log(\text{areal}_i) + d \cdot x_i + \varepsilon_i,$$

hvor forventningen er $b > 0$, $c < 0$; og ideelt set $b = -c$, hvilket vil indikere, at det er for-

13. Resultaterne er kvalitativt de samme hvis man i stedet anvender summen af *k1* og *k2*, hhv *t1* og *t2*.

holdet mellem kyststrækning og areal der er af relevans for kulturel værdisættelse af viden og nye ideer. Variablen x fanger alternative determinanter af C .

Da alle regressionerne, der følger, er simple OLS regressioner, kan man *ikke* nødvendigvis tillægge estimerne en kausal fortolkning. Men de partielle korrelationer bør dog give en indikation om, hvorvidt den foreslåede teori er plausibel eller ej.

Tabel 3 rapporterer resultaterne. I kolonne 1 undersøges den simple sammenhæng mellem C og hhv. kyststrækning og areal. Konsistent med den foreslåede teori er der en positiv sammenhæng mellem kyst/areal forholdet og kulturel værdisættelse af kreativitet og nye ideer, relativt til traditioner. At det er kyst/areal *forholdet*, der synes at være afgørende, ses nederst i tabellen, hvor restriktionen $b = -c$ testes; denne hypotese kan tydeligvis ikke afvises.

Idet modellen er specificeret som log-lineær er konsekvensen, at alle lande *uden* kyststrækning falder ud af samplet. Som et tjek undersøges det derfor i kolonne 2, om dette at være helt uden kyststrækning bidrager til en forklaring af C . Fittet er tydeligvis dårligere, når de rå »niveauer« for kyststrækning og areal inddrages. Men det væsentlige resultat er, at dummyvariablen for dette at være uden kyststrækning ikke holder forklaringskraft.

I kolonne 3 introduceres dernæst regionsindikatorer, og i kolonne 4 inddrages breddegrad som en forklarende variable. Motivationen for at inddrage regionsdummyer er selvsagt at kontrollere for uobserverbar heterogenitet, mens breddegraden udgør en (eksogen) kontrol for indkomst per capita. Det er velkendt, at der er nær korrelation mellem breddegrad og indkomst per capita, hvorfor kolonne 4 undersøger, om korrelationen mellem relativ kyststrækning blot kan tilskrives en underliggende sammenhæng mellem kulturel værdisættelse af viden og indkomst.

I sidste kolonne inddrages breddegrad og regionsindikatorerne samtidigt. Det overordnede resultat er, at kyststrækning og areal synes at være forholdsvis robust korreleret med C , konsistent med den foreslåede teori. Ydermere: I alle kolonnerne er det ikke muligt at afvise hypotesen $b = -c$, hvilket indikerer, at det er forholdet mellem kyststrækning og areal, der synes at være af betydning.

En indvending over for ovenstående kan være, at disse »kultur data« muligvis ikke siger noget reelt omkring, hvorvidt befolkning i et land værdisætter viden og er ide søgende. Som et groft tjek af om » C « reelt fanger hvad den tænkes at fange, undersøges det, om variabelen synes at være forbundet med *ideoutput*; konkret målt ved patenter per capita. I tabel 4 ses resultaterne af dette tjek.

I tabellens første kolonne tjekkes den reducerede form. Heraf fremgår det, at der er en systematisk tendens til, at samfund der er karakteriseret ved længere relativ kyststrækning også *producerer* flere patenter per indbygger. I kolonne 2 sammenholdes kulturvariablen med patenter; det fremgår, at lande hvor kreativitet og nye idéer står

Tabel 4. Vidensoutput, kyststrækning og kultur.

	-1	-2	-3	-4
	log patenter p.c.	log patenter p.c.	log patenter p.c.	log patenter p.c.
log kyststrækning	0,501 (2,78)**		0,311 (1,04)	
log areal	0,111 (0,62)		0,177 (0,46)	
Kreativ/Tradition		3,556 (2,72)**	3,468 (2,60)*	2,598 (1,73)+
Regioner	Nej	Nej	Nej	Ja
Observationer	70	44	36	40
R ²	0,17	0,11	0,29	0,55

Robuste *t* værdier i parentes + signifikant, 10%; * signifikant, 5%; ** signifikant, 1%.

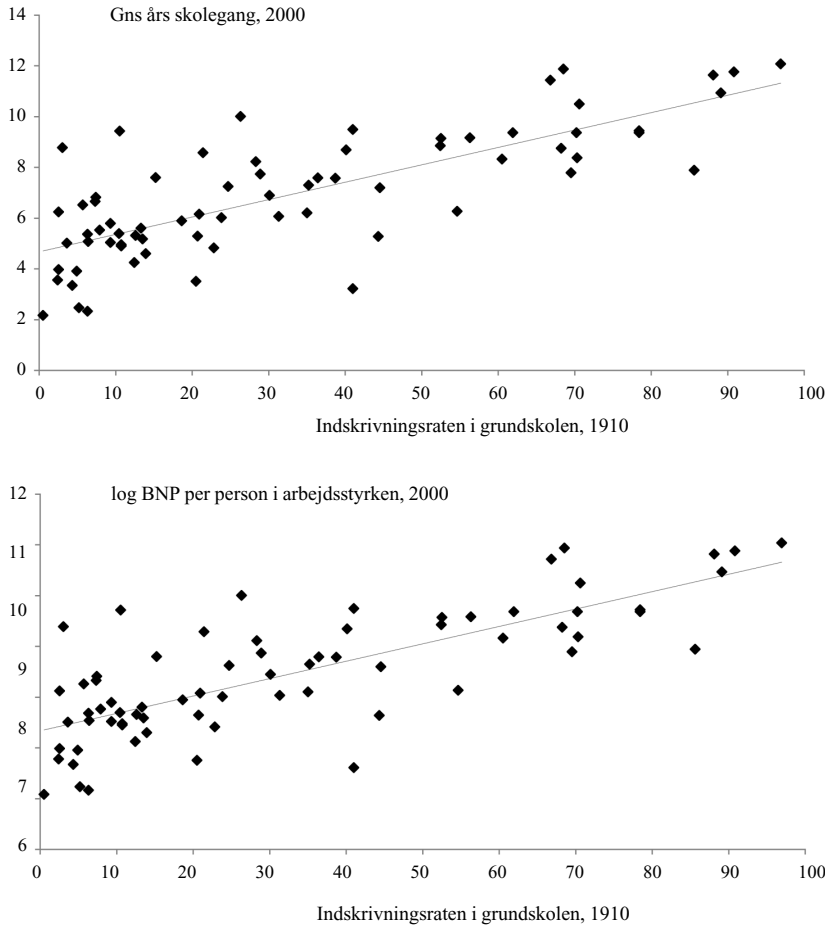
mere centralt i befolkningens bevidsthed end dét at følge traditionerne, da produceres der også flere patenter per indbygger. Kolonne 3 tjekker om indflydelsen fra kystlinje og areal forsvinder, dersom der betinges på *C*. Det viser sig at være tilfældet, hvilket er konsistent med teorien om, at areal og kystlængde manifesterer sig i kulturel værdsættelse af ny viden, hvilket i sin tur burde komme til udtryk i flere producerede (nye) ideer. Endelig, i kolonne 5, tjekkes det om kulturvariablen forbliver signifikant dersom der betinges på regionsindikatorer. Om end sammenhængen svækkes, er kulturvariablen dog fortsat signifikant på 10% niveau.

Det kan bemærkes, at arealvariablen gennemgående ikke er signifikant i regressio-
nerne i tabel 4. En simpel forklaring kan være, at et lands areal generelt kan være af betydning af årsager, der ligger ud over den foreslåede teori for »det danske vækstmi-
raket«. For eksempel er det sandsynliggjort, at landets størrelse spiller en (separat) rolle for den institutionelle udvikling, f.eks. Olsson og Hansson (2010). Desuden vil et lands areal være en proxy for, hvor mange naturressourcer der er til rådighed, hvilket igen tildeler landets areal en separat betydning udover dette at »normalisere« kystlinjen. Da disse faktorer let kan påvirke innovationstempoet i en økonomi, behøver man ikke opfatte det som et mysterium, at landets areal udviser en mindre robust korrelation med *patent*intensiteten.

Overordnet viser resultaterne i denne sektion dog, at der synes at være en systematisk tendens til, at lande med længere kyststrækning (i forhold til areal) kulturelt tillægger nye ideer og kreativitet mere vægt end traditioner, konsistent med den præsenterede teori.

3.2.2. Kyststrækning og tidlig humankapital akkumulation

Formelt estimeres følgende regressionsmodel nedenfor



Figur 1. Humankapital investering i 1910, Uddannelsestid og arbejdskraftproduktivitet i 2000.

Anm: Figuren til venstre viser den partielle korrelation mellem indskrivningsraten på grundskoleniveau (andel af relevant aldersgruppe i %) i år 1910 vs. den gennemsnitlige uddannelsestid i 2000. Til højre: indskrivningsraten i år 1910 vs. arbejdskraftproduktivitet i 2000. Linjerne er estimeret ved OLS, og der betinges kun på én konstant.

$$y_i = a + b \log(\text{kystlinie}_i) + c \log(\text{areal}_i) + X_i' \gamma + \varepsilon_i \quad (1.1)$$

Hvor y_i er humankapital investeringer i 1910, mens X_i er en vektor af alternative determinanter; parametrene a , b , c og vektoren γ estimeres ved OLS. Interessen samler sig imidlertid særligt om parametrene b og c .

Af årsager netop diskuteret ovenfor kan man ikke *a priori* vente at $b = -c$. Og igen er det værd at erindre, at de OLS regressioner der følger ikke nødvendigvis kan tillægges

Table 5. Tidlig humankapital akkumulation og kyststrækning.

	-1	-2	-3	-4	-5	-6	-7	-8	-9
	hk1910	hk1910	hk1910	hk1910	hk1910	hk1910	hk1910	hk1910	hk1910
log kyststrækning	9,435 (3,97)**	9,747 (3,99)**		6,756 (3,05)**	8,247 (2,99)**	9,281 (3,77)**	9,008 (3,58)**	4,929 (1,95)+	5,378 (3,02)**
Kyststrækning			1,155 (3,96)**						
log areal	-6,024 (5,05)**	-5,466 (3,63)**	-2,090 (1,57)	-5,343 (3,83)**	-3,704 (1,71)+	-5,706 (3,71)**	-4,094 (2,43)*	-2,424 (1,04)	-1,277 (0,69)
Britisk koloni							14,863 (1,92)+	22,387 (3,19)**	
Fransk koloni							-1,285 (0,27)	9,433 (2,07)*	
Protestant								0,149 (0,89)	
ELF					-0,226 (1,60)	0,294 (1,94)+		-0,115 (0,70)	
Breddegrad				0,766 (3,74)**				0,769 (2,72)**	
Uden kyst			-0,353 (0,04)						
BNP per capita, 1900									0,019 (8,98)**
Regioner	Nej	Ja	Ja	Ja	Ja	Ja	Ja	Ja	Ja
Observationer	72	72	75	70	65	70	71	64	29
R ²	0,23	0,36	0,27	0,49	0,42	0,42	0,40	0,58	0,91

Robuste *t* værdier i parentes + signifikant, 10%; * signifikant, 5%; ** signifikant, 1%.

en kausal fortolkning. Den uundgåelige specificationsusikkerhed vil næsten uvilkårligt lede til at OLS estimaterne ikke er middeltrette. Når dét er sagt, bør det dog også bemærkes, at der *a priori* ikke er nogen særlig grund til at tro, at denne usikkerhed gør det sværere at afvise de teoretiserede sammenhænge; udeladte variable kan lige så godt »styre« OLS estimaterne mod nul.

Som »opvarmning« viser figur 1 sammenhængen mellem uddannelsesinvesteringer (indskrivningsraten af en årgang til grundskolen) i 1910 og hhv. gennemsnitlig uddannelsestid i befolkningen og arbejdskraftproduktivitet i 2000. Det er tydeligt, at lande der allerede i 1910 uddannede sine borgere i stor stil, også gør det ca. 100 år senere. Det fremgår også, at indkomsten i lande, der tidligt investerede intenst i uddannelse, er signifikant højere et århundrede senere. Det rejser så spørgsmålet, om man kan finde systematiske karakteristika ved lande, der valgte at implementere formel uddannelse tidligt.

Tabel 5 rapporterer derfor den partielle korrelation mellem uddannelse i 1910 og længden af kystlinjen, betinget af landets samlede areal. I første kolonne ses blot på den »rå« sammenhæng, hvoraf det fremgår at (små) lande med lang kyststrækning i gennemsnit investerede mere i uddannelse i 1910 end (geografisk store) lande med kort kyststrækning.

At disse korrelationer skulle være udtryk for en tilfældighed kan afvises med over 99% sandsynlighed. I kolonne 2 betinges der på et fuldt sæt af regionsindikatorer (Afrika syd for Sahara, Asien osv.). Den partielle korrelation svækkes ikke nævneværdigt.

I de første to modeller er alle lande uden kystlinje udeladt som konsekvens af »log specificationen«. I kolonne tre tjekkes derfor, om dette at være aldeles uden kystlinje spiller en særlig rolle. Der indføjedes derfor en dummyvariabel, der andrager værdien 1, hvis landet er uden adgang til havet. Det fremgår, at den indføjede dummyvariabel ikke tilføjer væsentlig information.

I de næste fire kolonner undersøges, om kyststrækningens betydning forsvinder i fald der kontrolleres for klimatiske forhold (approximeret ved breddegraden for landets »centrum«), den etnisk-lingvistiske heterogenitet (fremmer politisk ustabilitet og sænker indkomsten), protestantandelen i befolkningen (hvilket taler til ideen om, at Reformationen stimulerer humankapital akkumulation), samt identiteten af en evt. kolonimagt. Endelig introduceres alle kontrollerne samtidigt. Klimatiske forhold (breddegrad) er stærkt korreleret med tidlig humankapital akkumulation, som forventet. Ligeledes er kontrollen for protestantisme og koloni dummyen for England (marginalt) statistisk signifikante. Det vigtigste budskab er dog i nærværende sammenhæng, at den partielle korrelation med kystlinjen forbliver signifikant.

Endelig tjekkes i kolonne 9, om indkomst per capita i 1900 nedbryder den partielle korrelation med kyststrækning. Med sidstnævnte kontrol reduceres antallet af observationer dramatisk til blot 29 lande. Interessant nok er der dog fortsat en statistisk signifikant (om end blot på 10% niveau) sammenhæng mellem kyststrækning og uddannelse i 1910.

Indflydelsen fra kystlinjen må også siges at være økonomisk signifikant. Hvis man tager estimatet i kolonne 9 for pålydende (når der altså betinges på indkomst), og dermed som udtryk for en kausal effekt, vil en forøgelse af kyststrækningen på 1% lede til en stigning i indskrivningsraten på grundskoleniveau i 1910 med ca. 5 procent point.

Disse regressioner er, som nævnt ovenfor, blot udtryk for et (sandt at sige: lidt løst) konsistenstjek. Men de viser dog, at der synes at være en systematisk sammenhæng mellem landes kyststrækning og tidlig humankapital akkumulation. Vi kan også konkludere, at denne sammenhæng ikke kan bortforklares ved, at der er mange lande uden adgang til havet i Afrika syd for Sahara; en region, hvor uddannelse ikke var udbredt i 1910 af årsager, der nok har meget lidt med kulturelle præferencer for uddannelse, eller mangel på samme. Ej heller kan korrelationen tilskrives, at regioner med lang kyststrækning tidligt var rige, og korrelationen er heller ikke frembragt af en mulig sammenhæng mellem kyststrækning og andre geografiske eller klimatiske karakteristika, der systematisk varierer henover kloden med landes fysiske placering (i.e., breddegrad) endside af kolonihistorie eller etnisk-lingvistisk diversitet.

3.2.3 *Kyststrækning og Arbejdskraftproduktivitet*

I dette afsnit estimeres ligning (1.1) igen, blot udskiftes tidlig humankapital med (log) BNP per person i arbejdsstyrken i 2000. Tabel 6 rapporterer resultaterne.

Som det fremgår følges mønsteret fra tabel 5 i forhold til kontroller (bortset fra, at indkomst i 1900 ikke inddrages i analysen her).

Flere resultater er værd at bemærke. Et interessant resultat er, at dette at være helt uden kystlinje er produktivitetshæmmende, hvilket er i tråd med tidligere studier, f.eks. Sachs og Warner (1997). Dog skal det bemærkes at *kystlængden* har en større *økonomisk* »effekt« på de observerede indkomstforskelle end dummyen for at være »landomsluttet«: de standardiserede regressionskoefficienter for kyststrækning og »uden kyst dummyen« er hhv 0,3 og -0,004 (ikke vist).

Et andet markant resultat er, at etnisk-lingvistisk diversitet synes at være forbundet med lavere arbejdskraftproduktivitet i 2000. Dette er konsistent med Easterly og Levines (1997) analyse, der knytter etnisk-lingvistisk diversitet til lav vækst (via politisk ustabilitet).

Ligeledes ses det, at der er en klar klima »gradient« i data: breddegrad er signifikant. Endelig er også koloni dummyerne signifikante. Igen er det centrale resultat

Tabel 6. Arbejdskraftproduktiviteten og kyststrækning.

	-1	-2	-3	-4	-5	-6	-7	-8
	logy	logy	logy	logy	logy	logy	logy	logy
log kyststrækning	0,282 (4,23)**	0,372 (6,07)**		0,184 (3,58)**	0,298 (4,74)**	0,376 (6,03)**	0,333 (5,15)**	0,175 (3,70)**
log areal	-0,176 (4,14)**	-0,220 (5,64)**	-0,103 (2,55)*	-0,168 (4,67)**	-0,161 (3,24)**	-0,216 (5,33)**	-0,205 (4,85)**	-0,154 (3,36)**
Kyststrækning			0,046 (2,72)**					
Britisk koloni							-0,001 (0,00)	0,265 (1,16)
Fransk koloni							-0,715 (2,96)**	-0,482 (2,05)*
Protestanter						0,003 (0,96)		-0,005 (1,73)+
ELF					-0,011 (2,75)**			-0,003 (0,61)
Breddegrad				0,037 (6,56)**				0,036 (4,23)**
Uden kyst			-0,697 (3,06)**					
Regioner	Nej	Ja	Ja	Ja	Ja	Ja	Ja	Ja
Observations	134	134	155	133	128	131	132	124
R-squared	0,14	0,25	0,17	0,44	0,31	0,27	0,31	0,52

Robuste t værdier i parentes + signifikant, 10%; * signifikant, 5%; ** signifikant, 1%.

Tabel 7. 2SLS regressionser, tidlig humankapital akkumulation og nutidig produktivitet.

	-1	-2	-3	-4	-5	-6
	logy	logy	logy	logy	logy	logy
humankapital, 1910	0,038 (5,00)**	0,028 (3,61)**	0,038 (4,52)**	0,037 (4,80)**	0,033 (3,56)**	0,033 (1,98)*
Breddegrad		0,021 (2,22)*				0,011 (0,61)
Protestanter			-0,007 (0,92)			-0,008 (1,27)
Britisk koloni				-0,714 (2,50)*		-0,422 (0,73)
Fransk koloni				-1,035 (2,72)**		-0,838 (1,83)+
ELF					-0,010 (1,43)	-0,003 (0,45)
Regioner	Ja	Ja	Ja	Ja	Ja	Ja
Observationer	68	67	67	67	64	63
Anderson-Rubin (p-værdi)	0,0001	0,0006	0,0001	0,0004	0,0133	0,07
Hansen J-Static (p-værdi)	0,51	0,86	0,31	0,74	0,48	0,84

Robuste *t* værdier i parentes + signifikant, 10%; * signifikant, 5%; ** signifikant, 1%.

imidlertid, at kyststrækning forbliver robust korreleret med indkomst per capita, også når samtlige kontroller indføres samtidigt.

Ifølge den foreslåede teori skal den reducerede form sammenhæng mellem kyststrækning og nutidig arbejdskraftproduktivitet tilskrives tidlig humankapital akkumulation. Men udover kyststrækning og areal faktisk deres indflydelse *via* tidlig humankapital akkumulation? For at undersøge dette spørgsmål tages der afslutningsvis instrumentvariabelestimation i brug.

Den foreslåede teori etablerer, at kystlængde og landeareal er kandidater til at udgøre instrumenter for tidlig humankapital akkumulation, siden teorien tilsiger at disse karakteristika bør påvirke nutidig indkomst *via* en tidlig start på formelle uddannelsesinvesteringer. Med to potentielle instrumenter (kyststrækning hhv. areal) og én endogen variabel er det muligt at teste udeladelsesrestriktionen; altså om kyststrækning og areal synes at bidrage til forklaringen af nutidig indkomst af årsager, der ligger *udover* disse variables indflydelse *via* humankapital. I fald det ikke kan afvises, at udeladelsesrestriktionen er opfyldt, vil 2SLS estimation også tillade et skøn over betydningen af den kausale effekt af tidlig uddannelse på nutidige produktivetsniveauer.

Tabel 7 viser resultaterne af at tage kystlinje og areal i anvendelse som instrumenter for humankapital investering i 1910.¹⁴ I alle kolonnerne kontrolleres der for regions-

14. Afpladshensyn undertrykkes »first stage resultaterne«; linket med uddannelse og kyst/areal forholdet er jo i øvrigt også undersøgt i tabel 5.

effekter, og desuden inddrages kontrollerne fra tabel 6. Nederst i tabellen testes udeladelsesrestriktionen; intet sted er denne inkonsistent med data. Man kan dermed ikke afvise, at kystlinje og areal udøver sin indflydelse på nutidig produktivitet via tidlig humankapital akkumulation.¹⁵ I rækken ovenover testes signifikansen af tidlig humankapital vha. Anderson-Rubin testet, der er robust overfor svage instrumenter. I alle kolonner kan det konkluderes, at tidlig humankapital akkumulation har en signifikant positiv *effekt* på nutidig BNP per person i arbejdsstyrken. Disse resultater er igen konsistente med den foreslåede teori for den danske økonomiske udvikling.¹⁶

Men resultaterne i tabel 7 åbner også muligheden for at undersøge den økonomiske signifikans af humankapital mekanismen ved brug af 2SLS estimatet for effekten af tidlig humankapital akkumulation på produktiviteten. Målsætningen er naturligvis at undersøge, hvor tæt humankapital mekanismen er på at motivere eksisterende produktivitetsforskelle.

Men hvor store er produktivitetsforskellene? I datamaterialet der ligger til grund for resultaterne i tabel 7 finder vi en produktivitetsvariation på ca. en faktor 26; dette er forholdet mellem BNP per person i arbejdsstyrken i toppen af »indkomstfordelingen« (konkret: den 95. percentil) og BNP per person i arbejdsstyrken i bunden (den 5. percentil).

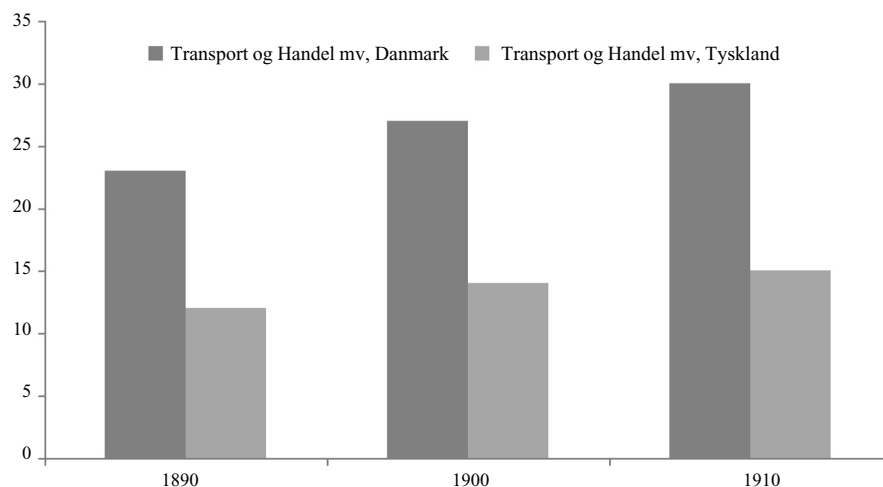
Hernæst kan man passende spørge, hvor stor forskellen var i relation til uddannelsesinvesteringerne i 1910. Her finder vi ved den 95. percentil, at indskrivningsprocenten var 88,2%; ved den 5. percentil var den blot 3,7%. Spørgsmålet er altså, hvor stor en andel af det observerede »produktivitetsgab« (faktor 26) der kan forklares ved de observerede forskelle i humankapital investeringerne i 1910 ($84,5\% = 88,2\% - 3,7\%$).

Der er tydeligvis variation i parameterestimaterne i tabel 7, hvorfor det virker tilrådeligt at udregne et interval for humankapital mekanismens styrke. Givet de observerede forskelle i uddannelsesinvesteringerne i data, finder vi, at humankapital mekanismen kan motivere produktivitetsforskelle på mellem en faktor 10 ($= \exp[0,028 * (88,2 - 3,7)]$) og en faktor 24 ($= \exp[0,0328 * (88,2 - 3,7)]$).

Med andre ord: tidlig humankapital akkumulation kan groft sagt motivere mellem 40% og 100% af det observerede produktivitetsgab mellem »rig og fattig« i den globale indkomstfordeling anno år 2000. Disse fund understøtter den grundlæggende

15. Det skal gerne medgives, at OID test af denne type er forholdsvis svage. Igen kan man måske derfor bedst anskue resultaterne som konsistentstjek.

16. I alle regressionerne, bortset fra i kolonne 1, er instrumenterne »svage« i en statistisk forstand, Steiger og Stock (1997); derfor anvendes Anderson-Rubin testet konsekvent til test af nulhypotesen om fravær af forklaringskraft i tidlig humankapital vis-a-vis nutidig arbejdskraftproduktivitet. Instrumenterne er dog altid signifikante (i sidste kolonne dog kun på 10% niveau) i »first stage«.



Figur 2. Transport og Detailhandel mv. bidrag til BNP i Danmark og Tyskland, 1890-1910.

Kilde: Krantz (1994), tabel 2.

præmis, at en tidlig start på uddannelsesinvesteringerne udgjorde en førsteordensforklaring på, at Danmark i dag er at finde blandt verdens rigeste nationer.¹⁷

Det er værd at betone, at dette estimat næppe skal anses for at være udtryk for den direkte produktivitetseffekt af formel uddannelse. Altså dette, at højere uddannelse øger produktiviteten for den enkelte og derigennem gennemsnitsproduktiviteten. For hertil kommer en vigtig indirekte effekt, via teknologisk innovation og adoption. I fald uddannelse anses for at være et centralt »input« bag teknologisk fremskridt, vil estimaterne også fange den inducerede produktivitetseffekt af uddannelse via teknologiske fremskridt.

3.3. Kyststrækning og sektorsammensætning

Den foreslåede teori, der altså også synes at have en vis forklaringskraft i forhold til andre lande end Danmarks udvikling, holder også en prædiction for økonomiens struktur, om end dette naturligvis ikke er sigtet med teorien. Denne prædiction er således et udmærket – yderligere – udgangspunkt for et tjek af teoriens eksterne validitet.

Konkret implicerer teorien, at den veludbyggede danske servicesektor har dybe historiske rødder. Såvel (detail- og engros-) handel som søtransport er ledende service-

17. Se Galor og Weil (2000) for en model, der motiverer, at »afsættet« til vækst er snævert knyttet til igangsættelsen af formelle uddannelsesinvesteringer. Dette studie startede et helt forskningsfelt, der sigter i mod at forstå nutidige indkomstforskelle som udkomme af meget langstrakte historiske processer; Se Galor (2005) for en oversigt.

Tabel 8. Økonomiens sammensætning og kyststrækning.

	-1	-2	-3	-4	-5	-6	-7	-8	-9	-10
	Service/ BNP	Service/ BNP	Service/ BNP	Service/ BNP	Service/ BNP	Service/ BNP	Service/ BNP	Service/ BNP	Service/ BNP	Service/ BNP
log kyststrækning	3,773 (3,95)**	5,051 (5,73)**	3,123 (2,24)*	2,821 (2,64)**	3,167 (3,46)**	4,620 (5,09)**	4,735 (5,14)**	5,038 (5,17)**	3,075 (2,90)**	
Kyststrækning				0,493 (2,42)*						
log areal	-3,218 (6,20)**	-3,792 (7,60)**	-2,088 (2,69)**	-2,52 (3,92)**	-3,432 (6,26)**	-3,465 (4,87)**	-3,517 (6,53)**	-3,875 (6,47)**	-3,509 (4,18)**	
Britisk koloni								-1,605 (0,47)		2,465 (0,64)
Fransk koloni								0,749 (0,22)		5,567 (1,83)+
Protestanter							0,104 (2,26)*			-0,007 (0,12)
ELF							-0,068 (1,52)			0,073 (1,38)
Breddegrad					0,402 (4,82)**					0,541 (4,19)**
Uden kyst					-6,691 (2,87)**					
log(BNP/Arbejdsstyrke), 2000				4,62 (2,89)**						
humankapital, 1910			0,197 (3,11)**							
Regioner	Nej	Ja	Ja	ja	Ja	Ja	Ja	Ja	Ja	Ja
Observations	129	129	68	121	147	128	118	126	127	114
R-squared	0,20	0,31	0,53	0,34	0,25	0,44	0,29	0,34	0,31	0,44

Robuste t værdier i parentes + signifikant, 10%; * signifikant, 5%; ** signifikant, 1%.

erhverv i dag, og har – ifølge teorien – spillet en afgørende rolle i den danske økonomiske vækst proces. Dette illustreres ganske godt af figur 2, der viser estimater for bidraget fra disse sektorer til Danmarks BNP 1890-1910. Allerede på tidspunktet, hvor den økonomiske vækst i BNP per capita »sætter af«, er nævnte serviceerhverv ganske store: da vi når 1910 udgør de til sammen ca. 1/3 af den danske økonomi. Til sammenligning spillede disse erhverv – konsistent med den fremlagte teori – en noget mere tilbagetrukket rolle i (nabolandet) Tyskland, der er betydeligt mindre kystvendt.

I fald man regner alle serviceerhvervene sammen (inklusive offentlig service) udgjorde servicesektoren 45% af BNP i 1910 i Danmark (ikke vist). At servicesektoren er »stor« i Danmark, er således ikke noget nyt. Men der skal ikke herske tvivl om, at sektoren er blevet større med tiden. I 2007 udgjorde service sektoren således ca. 70% af BNP. En god del af væksten skal naturligvis tilskrives offentlig service, der i 1910 kun udgjorde 5% af BNP mod 21% i 2007. Til den ende har systematiske humankapital investeringer (sundhed såvel som uddannelse) gennem et århundrede naturligvis bidraget til væksten i servicesektoren.

Men i hvilket omfang er dette et unikt karakteristikum for Danmark? Dette spørgsmål lader sig adressere ved tværlande data. Konkret vender vi os afslutningsvis mod dette:

4. *Kyststrækning og økonomiens sammensætning.* Er der noget der tyder på at servicesektoren er relativt større i lande med relativt lange kyststrækninger?

Tabel 8 indeholder regressionsresultaterne.

Som det fremgår af kolonne 1 i tabel 8 er den relative kyststrækning systematisk positivt korreleret med servicesektorens BNP bidrag. Hvis man tager punktestimatet for kyststrækning for pålydende, vil en (betinget) stigning i kyststrækningen med 1 procent forøge service sektoren med knap 4%.

I de efterfølgende kolonner betinges der som i afsnittet ovenfor på hhv. regioner, protestantandelen i befolkningen, klimatiske karakteristika m.m. Særligt værd at bemærke er kolonne 3, hvor der introduceres tidlig humankapital akkumulation. Siden den relative kyststrækning synes at være forbundet med tidlige uddannelsesinvesteringer kunne det være tilfældet, at denne variabel forklarer korrelationen mellem førnævnte og servicesektorens relative størrelse. Dette ville imidlertid være *inkonsistent* med den foreslåede teori for den danske økonomis udvikling: teoretisk set bør effekten af kystlængde på sektorsammensætningen være *direkte* af natur – frembragt af geografis betydning for udvikling af handel og søfart. Det er derfor betryggende, at den relative kyststrækning forbliver signifikant, når tidlig uddannelse introduceres.

Et andet vigtigt resultat er, at kyststrækning forbliver signifikant, når indkomst indføjes. Det er velkendt, at servicesektorens relative størrelse erfaringsmæssigt øges, når

indkomsten stiger, f.eks. Kuznets (1957); Chenery (1960). Siden kyststrækning og indkomst er positivt korreleret (tabel 6), kunne sammenhængen mellem kyststrækning og servicesektorens størrelse være frembragt via indkomst. Dette ville også være inkonsistent med den foreslåede teori, idet kyststrækning bør påvirke sektorsammensætningen direkte, uden henvisning til det forhold, at efterspørgslen efter serviceydelser er indkomstelastisk. Betyggende nok ses det, konsistent med den fremlagte teori, at den partielle korrelation mellem serviceandelen og den relative kystlængde fortsat er signifikant, når der betinges på indkomst. Faktisk viser det sig, når man udregner de standardiserede regressionskoefficienter, at kystlinjen har en *stærkere* »effekt« på servicesektorens størrelse end indkomsten har; de standardiserede parameterestimer for kystlængde og indkomst er hhv. 0,28 og 0,26 (ikke vist).

Samlet set antyder disse fund, at den veludbyggede danske servicesektor må anses for at have særdeles dybe rødder. I lyset af det faktum, at service sektorens bidrag til BNP fortsat er stort, efter et århundrede med historisk kraftig vækst i BNP, må det konkluderes, at denne sektors andel i »det danske vækstmirakel« har været umådelig stor.

3.4 Opsummering

Den empiriske analyse har sandsynliggjort at: samfund med længere kyst-areal forhold er (i) kulturelt mere videnskabsfokuserede; (ii) begyndte at investere tidligere i formel humankapital; (iii) er i dag velstående, af årsager der synes nært relateret til (ii). Endelig (iv), kystvendte samfund er typisk karakteriseret ved en større servicesektor.

Den økonomiske effekt af tidlig humankapital akkumulation er markant: humankapital mekanismen kan motivere produktivitetforskelle i år 2000 på mellem en faktor 10 og en faktor 24. Det er her vigtigt at forstå, at dette estimat er udtryk for en *akkumuleret* effekt af tidlig humankapital akkumulation. Teoretisk vil man vente, at intensiverede uddannelsesinvesteringer på et tidligt tidspunkt har stimuleret den økonomiske vækst i det fremadrettede på flere forskellige måder. Dels vil humankapital akkumulation sædvanligvis være ledsaget af en nedgang i fertiliteten, hvilket stimulerer væksten direkte via mindre »kapitalfortrængning«; dels bør humankapital opbygning stimulere teknologiadoption og innovation, og endelig må man vente at humankapital investeringer gør det mere attraktivt at investere også i fysisk kapital. At tidlig humankapital investering har så markant en effekt på nutidige produktivitetniveauer, skal derfor ses i lyset af de *indirekte* effekter, uddannelsesinvesteringerne efterfølgende har »kastet af sig«.¹⁸

Det skal gerne erkendes, at disse fund ikke beviser, at den fremlagte teori er korrekt. Men de viser ret klart, at den ikke indlysende er forkert.

18. Se Galor (2010) for en uddybning af disse pointer; se også Dalgaard og Strulik (2010).

4. Afrunding

I fald ovenstående teori holder vand, er der et sæt af potentielt nyttige implikationer for den nuværende debat om, hvor den danske vækst skal komme fra i fremtiden.

1. *Teknologiadoption vs. teknologiopfindelse*

Et nøgleord for den danske historiske vækstproces har været: *Teknologiadoption*.

Historisk har Danmark med succes formået (hurtigt) at tage nye ideer til sig, som kunne øge produktiviteten eller udgøre basis for indtjening. Automatiseringen af landbruget fra slutningen af det 19. århundrede og frem er måske det mest fundamentale eksempel på, at den historiske danske vækst har været baseret på hastig teknologiadoption, snarere end af *teknologiopfindelse*. Et andet, senere, eksempel er udbredelsen af Internettet. I 2007 var Danmark i den globale top 3 over Internetbrugere per capita (81/100 indbyggere); kun overgået af Island og Holland.¹⁹ I fald denne vækstopskrift skal følges også i tiden fremover, bør et centralt pejlemærke for den økonomiske politik være, at udrydde (eller undgå at rejse) barrierer for *teknologiadoption*. Set med udgangspunkt i historien er hensynet til hastig teknologiadoption vigtigere end hensynet til »opfindelse« af nye teknologier.

2. *Den danske vækst havde ikke (og har ikke) sit primære udspring i industrien*

Historisk set har servicesektoren, ifølge den fremlagte teori, været den centrale drivkraft for den danske vækstproces. Sektoren har stået for en stor del af beskæftigelsen op igennem det 20. århundrede; og en stadig stigende del i de seneste årtier, hvor beskæftigelsen i industrien er aftaget. Fra 1990-2007 har stort set hele beskæftigelsesfremgangen fundet sted i den private servicesektor, alt i mens beskæftigelsen i industrien og landbruget er aftaget.

Denne gradvise beskæftigelsestransition har fundet sted alt imens den gennemsnitlige økonomiske vækst har holdt sig forholdsvis stabil i samme periode, Dalgaard og Hansen (2010). Det må forventes at denne proces vil fortsætte i årene fremover.

Alligevel er der god grund til at se nærmere på servicesektoren. For mens den danske økonomiske vækst ikke er løjet af *for nylig*, da var der et fald i timeproduktiviteten (i lighed med næsten alle vestlige økonomier)??? *fra begyndelsen af 1970erne*, Dalgaard og Hansen (2010). Dette »slowdown« har bl.a. den danske økonomi tilsyneladende ikke rigtig »kommet sig over«. ²⁰ Men det har den amerikanske imidlertid. Fra

19. Kilden til disse tal er World Development Indicators (2010).

20. Dog bør man holde sig et forbehold in mente: datakvaliteten på antallet af det årlige antal leverede timer er dårligt belyst i Danmarks Statistik, og kan være en forklaring på at timeproduktiviteten opfører sig anderledes en BNP per indbygger, eller per beskæftiget; sidstnævnte serier udviser stationær vækst fra 1950 og frem til 2007. Se Dalgaard og Hansen (2010) for uddybning.

midten af 1990'erne har USA oplevet et »growth revival« (frem til finanskrisens udbrud, forstås), som også har ledt til en vis indkomst*divergens* mellem USA og Europa som en helhed. Til den ende er det vigtigt at notere sig, at servicesektoren i USA også udviste en markant produktivitetsacceleration i 1990'erne og derigennem bidrog til den aggregerede acceleration, Tripplet og Bosworth (2003). IT investeringer forekommer ikke at være grundlaget for denne acceleration; hele stigningen i væksten 1995 og frem i den amerikanske servicesektor synes at udspringe af andre årsager, der er opsummeret i TFP væksten. Årsagen til, at et tilsvarende forløb ikke synes at være forekommet i Danmark er ukendt og derfor værd at studere i lyset af service sektoren store (historiske) betydning for det danske BNP og beskæftigelsen.

3. Uddannelse og vidensfokusering har været helt central for den danske vækstproces

Humankapital har spillet en central rolle i den danske vækstproces ifølge den fremlagte teori. Det er derfor velbegrunderet, at uddannelse sædvanligvis står centralt i den politisk-økonomiske debat.

Alligevel er det måske værd at overveje, om uddannelsessystemet – som det fremstår i dag – er optimalt indrettet. Specielt i lyset af hensynet til hastig teknologiadoption. Således er der nyere forskning der peger på, at uddannelsesinstitutionerne kan have noget med vækstforskellen mellem Europa og USA at gøre (fra midten af 1990'erne og frem). I Danmark sikrer universiteterne uddannelsen af specialister. Når man starter på universitetet, specialiserer man sig med det samme; modsat i USA hvor universitetsuddannelsen tillader de studerende i højere grad, at »snuse« til andre felter end dét, de ultimativt vælger at specialisere sig indenfor. En interessant hypotese er, at den europæiske uddannelsesmæssige specialisering kan udgøre en barriere for at drage nytte af nye teknologier, hvis den hæmmer kandidaternes muligheder for at se *breddere* anvendelser af de seneste innovationer, Krueger og Perri (2003; 2004). Kan dette være en del af baggrunden for, at den produktivitetsvæksten i privat service ikke er accelereret i Danmark, mens det altså har været tilfældet i USA? Også dette perspektiv kunne det være en værd at dykke dybere ned i, når rammerne for den fremtidige danske vækstproces søges afstukket.

Litteratur

- Acemoglu, D. 2009. *Introduction to modern economic growth*. Princeton University Press.
- Andersen, T. B. og C-J. Dalgaard. 2009. Cross-Border Flows of People, Technology Diffusion and Aggregate Productivity. Arbejdspapir (Københavns Universitet).
- Andersen T. B., C-J. Dalgaard og P. Selaya. 2010. Eye disease and development. Arbejdspapir (Københavns Universitet).
- Andersen, T. B., J. Bentzen, C-J. Dalgaard og P. Sharp. 2010. Religious Orders and Growth through Cultural Change in Pre-Industrial England. Arbejdspapir (Københavns Universitet).
- Becker, C. J. 1959. Flint Mining in Denmark.

- Antiquity* 33, 87-92.
- Becker, S. and L. Woessmann. 2009. Was Weber Wrong? A Human Capital Theory of Protestant Economic History. *Quarterly Journal of Economics* 124, 531-96.
- Benavot, A. and P. Riddle. 1988. The expansion of primary education, 1870-1940: Trends and Issues. *Sociology of Education* 61, 191-210.
- Comin, D., W. Easterly and E. Gong. 2010. Was the Wealth of Nations Determined in 1000 B.C.? *American Economic Journal: Macroeconomics* 2, 65-97.
- Chanda, A og C-J. Dalgaard. 2008. Dual Economies and International Total Factor Productivity Differences: Channeling the Impact from Institutions, Trade and Geography. *Economica*, 75, 2008, 629-61.
- Chenery, H. B. 1960. Patterns of Industrial Growth. *American Economic Review* 50, 624-54.
- Dalgaard, C-J. og H. Hansen. 2010. Er der er et dansk væskollaps? *Nationaløkonomisk Tidsskrift* 148, 1-20.
- Dalgaard, C-J. og H. Strulik. 2010. The History Augmented Solow Model. Arbejdsrapport (Hannover Universitet).
- Diamond, J. 1997. *Guns, Germs and Steel*. Penguin Books.
- Dittmar, J. 2010. Information Technology and Economic Change: The Impact of the Printing Press. Working paper (American University).
- Easterlin. 1981. Why isn't the whole world developed? *Journal of Economic History* 41, 1-17.
- Easterly, W. and R. Levine. 1997. Africa's Growth Tragedy: Policies and Ethnic Divisions. *Quarterly Journal of Economics* 112, 1203-50.
- Engeloff, I. B. 1994. Fishing in Denmark during the Ertebølle period. *International Journal of Osteoarchaeology* 4, 65-96.
- Frankel, J. og D. Romer. 1999. Does trade cause growth? *American Economic Review* 89, 379-99.
- Fisher, A. 2007. Coastal Fishing in Stone Age Denmark – Evidence from below and above the present sea level and from human bones. I: N. Milner, G. Bailey og O. Craig (eds.), *Shell Middens in Atlantic Europe*. Oxbow Books.
- Galor, O. 2005. From Stagnation to Growth: Unified Growth Theory, I *Handbook of Economic Growth*, P. Aghion and S. Durlauf (red.), Elsevier.
- Galor, O. 2010. The Lawrence Klein Lecture: Comparative economic development: insights from unified growth theory. *International Economic Review* 51, 1-44.
- Galor O. og D. Weil. 2000. Population, technology, and growth: From Malthusian stagnation to the demographic transition and beyond. *American Economic Review* 90, 806-28.
- Hale, J. R. 1998. The Viking Longship. *Scientific American* (februar).
- Hencken, H. 1950. Herzsprung Shields and Greek Trade. *American Journal of Archaeology* 54, 294-309.
- Henriksen, I. 1992. The Transformation of Danish Agriculture 1870-1914. I K. G. Persson (red.) *The Economic Development of Denmark and Norway since 1870*. Aldershot, UK: Edward Elgar Publishing Ltd.
- Henriksen, I. 1999. Avoiding lock-in: Cooperative Creameries in Denmark, 1882-1903. *European Review of Economic History* 3, 57-78.
- Henriksen, I., M. Lampe and P. Sharp. 2010. The Strange Birth of Liberal Denmark: Danish trade protection and the growth of the dairy industry since the mid-nineteenth century. *Discussion Paper* no. 10-04. Department of Economics, University of Copenhagen.
- Hybel, N. 2003. *Danmark i Europa 750-1300*. Museum Tusulanums Forlag.
- Jensen, J. 1983. *The Prehistory of Denmark*. Routledge.
- Jones, G. 1968. *A History of the Vikings*. Oxford University Press.
- Kratz, O. 1994. Service production in historical national accounts. *Review of Income and Wealth* 40, 19-41.
- Krueger, D. og F. Perri. 2003. US-Europe Differences in Technology Driven Growth: Quantifying the role of Education. *Journal of Monetary Economics* 51, 161-90.
- Krueger, D. og F. Perri. 2004. Skill specific rather than general education: A reason for

- US-Europe Growth Difference? *Journal of Economic Growth* 9, 167-2007.
- Kuznets, S. 1957. Quantitative Aspects of the Economic Growth of Nations: II. Industrial Distribution of National Product and Labor Force. *Economic Development and Cultural Change* 5, 1-111.
- Landes, D. 1999. *The Wealth and Poverty of Nations*. Abacus.
- Maddison, A. 2001. *The world economy: a millennial perspective*.
- Mokyr, J. 2005. *The gifts of Athena: Historical origins of the knowledge economy*. Princeton University Press.
- Olsson, O. og G. Hansson. 2010. Country Size and the Rule of Law: Resuscitating Montesquieu. Udkommer i: *European Economic Review*.
- Persson, K. G. 2010. *An Economic history of Europe – Knowledge, Institutions and Growth 600 to present*. Oxford University Press (Cambridge).
- Randsborg, K. 2009. *The Anatomy of Denmark – Archaeology and History from the Ice Age to the Present*. Duckworth.
- Richards, M. P., TD. Price og E. Koch. 2003. Mesolithic and Neolithic Subsistence in Denmark: New Stable Isotope Data. *Current Anthropology* 44, p. 288-95.
- Roesdahl, E. 1998. *Vikingernes Verden*. Gyldendal.
- Sachs, J., A. Warner. 1997. Fundamental sources of long-run growth. *American Economic Review* 87, 184-88
- Sarauw, T. 2008. Danish Bell Beaker Pottery and Flint Daggers- the Display of Social Identities? *European Journal of Archaeology*, 11, 23-47.
- Staiger, D. and J. Stock. 1997. Instrumental variables with Weak Instruments. *Econometrica* 65, 557-86.
- Timbergen, Jan. 1962. *Shaping the World Economy: Suggestions for an International Economic Policy*. New York: The Twentieth Century Fund.
- Tripplert, J. and B. Bosworth. 2003. Productivity Measurement Issues in Services Industries: Baumol's Disease Has Been Cured. *FRBNY Economic Policy Review*.
- Ulriksen, J. 1994. Danish sites and settlements with a maritime context, AD 200-1200. *Antiquity* 68, 797-811.
- Tauber, H. 1981. C-evidence for dietary habits of prehistoric man in Denmark. *Nature* 292, 332-33.
- Vollrath, D. 2009. How Important are Dual Economy Effects for Aggregate Productivity? *Journal of Development Economics*, 88, pp. 325-34.
- Weber, M. 1905. *The Protestant Ethic and the Spirit of Capitalism*. Allen & Unwin
- Weil, D. N. 2005. *Economic Growth*. Pearson.

Appendix

Data anvendt i Sektion 3

Kystlinjens længde er hentet fra CIA factbook

< <https://www.cia.gov/library/publications/the-world-factbook/>>, og henviser altså til landenes nuværende kystlinje.

Værdsættelse af nye ideer. Kilden til data er World Value Survey

< <http://www.worldvaluessurvey.org/>>

I forhold til »værdsættelse af nye ideer« får respondenterne læst følgende højt:

»V80.- Now I will briefly describe some people. Using this card, would you please indicate for each description whether that person is very much like you, like you, somewhat like you, not like you, or not at all like you? (Code one answer for each description): It is important to this person to think up new ideas and be creative; to do things one's own way.

Possible answers:

- 1 Very much like me
- 2 Like me
- 3 Somewhat like me
- 4 A little like me
- 5 Not like me
- 6 Not at all like me «

Mht. »traditioner« lyder spørgsmålet:

»V89.- Now I will briefly describe some people. Using this card, would you please indicate for each description whether that person is very much like you, like you, somewhat like you, not like you, or not at all like you? (Code one answer for each description): Tradition is important to this person; to follow the customs handed down by one's religion or family.« Listen af svar muligheder er som ovenfor.

I Sektion 3,2 anvendes den procentvise andel af respondenterne der svarer (1)-(3) i situationen, der handler om kreativitet relativt til andelen af respondenter, der svarer (1)-(3) i situationen, der handler om traditioner.

Data for uddannelsesinvesteringer i 1910 (»primary school enrolment rates«) er hentet fra Benavot and Riddle (1988). *Data for BNP per capita i 1900* er fra Maddison (2001). *Servicesektorens størrelse og patenter per capita* er hentet fra World Development Indicators 2010. < <http://data.worldbank.org/>>. *Øvrige data* er »standard data« der også anvendes i Andersen, Dalgaard og Selaya (2010); der henvises til dette papir for kildeangivelser.

Den samfundsøkonomiske kalkulationsrente

Anders Chr. Hansen

Department of Environmental, Social and Spatial Change (ENSPAC), Roskilde Universitet,

E-mail: anders@ruc.dk

SUMMARY: The discount rate prescribed by the Danish Ministry of Finance for the economic appraisal of politically driven investments is relatively high and this has given rise to concerns for sub-optimal ranking of these. The paper addresses the theoretical and empirical basis for the social discount rate and recommends a level of 3%+/-2% for time horizons within 30 years. For longer time horizons a gradually declining discount rate is recommended (hyperbolic discounting). It is shown that hyperbolic discounting does not give rise to time inconsistency in the appraisal of classical (putty-clay/sunk cost, inseparable) investment plans.

Den politisk-administrativt fastsatte kalkulationsrente

Finansministeriets krav til samfundsøkonomisk vurdering af offentlige investeringer og visse andre politisk drevne investeringer har siden 1999 indebåret anvendelse af en kalkulationsrente på 6%, Finansministeriet (1999). Nogle planlægningsområder har dog fra 2009 kunnet anvende en kalkulationsrente på 5%. Både 5% og 6% er relativt høje kalkulationsrentesatser, og det kan medføre, at kortsigtede projekter prioriteres over langsigtede, samt at projekter, hvor omkostningerne er jævnt fordelt over projektperioden, prioriteres over projekter, hvor omkostningerne ligger tidligt i forløbet. For projekter og planer med virkninger for fremtidige generationer har diskontering desuden etiske aspekter.

Udgangspunktet for denne artikel er, at staten eller samfundet ligesom husholdninger og virksomheder har gode og rationelle grunde til at tillægge fremtidige værdier mindre vægt end nutidige, altså at diskontere. Spørgsmålet er, *hvor meget* staten bør diskontere. Ikke alle grunde er lige gode, og besvarelsen af spørgsmålet må derfor starte med, hvilke gode grunde staten eller samfundet har til at diskontere.

Der er ikke nogen konsensus blandt økonomer om den samfundsøkonomiske kalkulationsrentes størrelse. Alligevel er det nødvendigt at lægge sig fast på et bestemt niveau, hvis de økonomiske kalkuler skal spille nogen rolle for beslutningerne om offentlige investeringer og andre investeringer, der er drevet af politiske beslutninger. Ellers er de økonomiske besluningsgrundlag for dem ikke sammenlignelige og desuden for lette at manipulere.

Det er således en almindelig praksis, også i udlandet, at den myndighed, der står for planlægningen af den offentlige økonomi, udmelder en kalkulationsrente, der skal bruges i den økonomiske vurdering af skattefinansierede projekter og af private investeringer, der forsøges drevet i en samfundsønskelig retning af handlingsplaner mm.

I denne artikel drøftes de økonomiske argumenter bag valget af kalkulationsrente. Artiklen kan ikke komme omkring hele diskussionen, men det anbefales at læse Møller (2009), som med stor indsigt forklarer stort set alle begreber og metoder, der er behandlet i nærværende artikel og konfronterer efficiens-problemet med det etiske problem.

Andre bud på kalkulationsrenten

Finansministeriets kalkulationsrente på 6% er langt højere end den kalkulationsrente, som tilsvarende myndigheder i andre europæiske lande foreskriver.

I Norge, Det Kongelige Finansdepartement (2005), Sverige, Vägverket (2006) og Frankrig, Commissariat général du Plan (2005) foreskriver man 4%. Det tyske finansministerium, Bundesministerium der Finanzen (2010, 2009) har i de seneste par år foreskrevet 2,2%.

En vejledning fra Europa-Kommissionen, Commission of the European Communities, DGRP (2008) beregner samfundsøkonomiske kalkulationsrenter på 2,8-4,1% for de gamle EU-lande (3,5% for Danmark) og 5,3-8,1% for de nye, som har højere vækstrater og alternativafkast. Disse kalkulationsrenter anbefales til vurdering af projekter, der finansieres af EU's social- og samhørighedsfonde.

Der findes dog også europæiske lande der anvender kalkulationsrenter på 5%. Bl.a. anvender det finske transport- og regionsministerium 5% i hvert fald på transportrelaterede investeringer, Liikenne- ja aluehallinto (2004).

Man kunne overveje, om der ikke burde være en kalkulationsrente for hele EU i betragtning af den konvergerende udvikling af medlemslandenes økonomier. Konvergenen er dog næppe så vidt fremskreden i dag, at dette er aktuelt. Især har finanskrisen tydeliggjort de betydelige forskelle imellem Nord- og Sydeuropa, imellem samhørighedslande og det øvrige EU samt imellem de nye og de gamle medlemslande. Forskellene imellem de ovennævnte landes og den danske kalkulationsrente bør dog give anledning til at genoverveje grundene til at have så højt et niveau i Danmark.

Blandt danske økonomer er der langt fra enighed om Finansministeriets forskrift. Miljøministeriets eksperter på området anbefaler en kalkulationsrente på 3% +/-2%, Møller m.fl. (2000).

De Økonomiske Råd har igennem årene set på mange typer af langsigtede investeringer i eksempelvis miljø, energi og uddannelse og har i den forbindelse brugt diskonteringsrenter på 3%, Det Økonomiske Råd (2008, 2007, 2006) og 4%, Det Økono-

miske Råd (1999). Rådets beregninger af økonomien i 1990ernes energiinvesteringer viste i øvrigt, at anvendelsen af en diskonteringsrente på 6% i stedet for 3-4% er alt, der skal til for, at udviklingen af Danmarks vindmøllesektor fremstår som en økonomisk fiasko, Det Økonomiske Råd (2002).

Virkninger af at bruge en høj kalkulationsrente

En for høj kalkulationsrente kan ligesom en for lav kalkulationsrente føre til en sub-optimal prioritering.

Økonomien i de vurderede projekter består i reglen i at balancere en stor investering nu over for en strøm af fordele i fremtiden. For eksempel investeringer i fjernvarmeanlæg eller vindmøller til gengæld for lavere (eller ingen) udgifter til og forurening fra brændsel i fremtiden. En høj kalkulationsrente reducerer nutidsværdien af de fremtidige fordele og gør dermed projekterne mindre lønsomme. Det er i højere grad tilfældet, jo længere projektets driftsperiode er. Derfor kan en for høj kalkulationsrente føre til, at der investeres for lidt i de projekter, der har en lang tidshorisont. Projekter med særlig lang tidshorisont findes for eksempel inden for vandsektoren, naturgenopretning, transmissionsnet og geotermisk varme.

Desuden vil en højere kalkulationsrente føre til lavere nutidsværdi af projekters omkostninger, hvis de er jævnt fordelt over eller ligger sent i projektforløbet, end hvis de er koncentreret tidligt i forløbet.

Kalkulationsrenten er således udslagsgivende for den samfundsøkonomiske vurdering af projekter, handlingsplaner og strategiplaner, men ikke for deres faktiske rentabilitet og ikke nødvendigvis for deres faktiske gennemførelse.

Den samfundsøkonomiske vurdering er ofte kun et af flere hensyn i beslutningsprocessen. Nogle samfundsøkonomiske aspekter som innovation, forsyningssikkerhed og importbelastning indgår ikke i de samfundsøkonomiske vurderinger, men spiller alligevel en stor rolle for projekternes samfundsøkonomiske betydning. Hertil kommer helt andre hensyn jf. det grundlæggende politologiske problem for politiske beslutningstagere om »stemmeoptimering vs. samfundsøkonomisk optimering«.

Det er dog vanskeligt, at afgøre præcis hvilken rolle de enkelte hensyn har spillet for de enkelte projekters og planers realisering. Eksempelvis fandt Trafikministeriet (2001) at elektrificering af banenettet i Jylland ville have intern rente på -2%. Der er tydeligvis mange andre årsager til, at projektet 10 år efter den planlagte færdiggørelse kun er nået halvvejen, men den samfundsøkonomiske vurdering må antages at have en vis betydning.

Kalkulationsrenten har dog næppe en stor indflydelse på den samlede størrelse af offentlige investeringer i Danmark, da denne bestemmes af en langsigtet makroøkonomisk ramme. Efter 1970ernes massive offentlige investeringer på 3-5% af BNP blev

deres omfang reduceret betydeligt, og det har siden varieret omkring 1,8% af BNP. Skulle man i dag løfte denne ramme til f.eks. 3% af BNP som i det øvrige Skandinavien, ville det forudsætte en politisk beslutning herom snarere end en lavere kalkulationsrente. Der er ganske vist politologiske forklaringsmodeller, ifølge hvilke en højere kalkulationsrente skulle være attraktiv, fordi den dæmper presset på regeringen for at gennemføre offentlige investeringer, men den type af politologiske mekanismer er meget vanskelige at verificere.

Finansministeriets kalkulationsrente spiller derimod ikke nogen rolle for den faktiske finansiering og dermed for projekternes faktiske rentabilitet. I praksis finansieres den offentlige sektors udgifter til en rente, som ikke adskiller sig væsentligt fra den internationale rente, og som anvendes i finansielle eller budgetøkonomiske analyser.

En stor del af investeringerne i infrastruktur og lignende foretages af offentlige selskaber, som typisk kan finansiere deres investeringer med statsgaranterede lån eller statslige genudlån til omtrent samme rente. De omfatter bl.a. DR, DSB, selskaberne, der står for de faste forbindelser og Københavns havne-Ørestads-Metro-udvikling samt Danmarks Skibskreditfond.

Nogle af de større projekter, der udføres af disse selskaber, underkastes også en samfundsøkonomisk vurdering, som kan få betydning for deres gennemførelse. Det samme er tilfældet for større handlingsplaner og strategiske planer om f.eks. klimapolitik, vedvarende energi og biobrændstof. På disse områder får kalkulationsrenten betydning for meget store private investeringer, der ikke er omfattet af rammen for offentlige investeringer, og hvis gennemførelse er drevet af politisk besluttede incitamenter og krav.

På klima- og energiområdet vil en høj kalkulationsrente føre til en mere positiv samfundsøkonomisk vurdering af løsninger, der indebærer store omkostninger sent i perioden, som eksempelvis atomaffalds- og geologisk CO₂-deponering. Disse teknologier spiller dog ikke den store rolle i de foreliggende danske energi- og klimaplaner. Omvendt vil løsninger, der indebærer store investeringer tidligt i perioden til gengæld fordele meget lang tid ud i fremtiden vurderes mere negativt. Det kunne være eksempelvis være geotermisk varme eller internationale el-transmissionsnetværk.

Kommunerne, som har ansvaret for godkendelse af investeringer i kollektive varmforsyningsanlæg (fjernvarme- og kraftvarmeanlæg) skal godkende det »ud fra en konkret vurdering ... samfundsøkonomisk mest fordelagtige projekt« (BEK nr. 1295 af 13/12/2005, Godkendelse af kollektive varmforsyningsanlæg (2005, §6). Det skal ske på basis af Energistyrelsens vejledning, Energistyrelsen (2007), ifølge hvilken analyser af projekter, der gennemføres som led i den vedtagne energipolitik, skal gennemføres med den af Finansministeriet foreskrevne kalkulationsrente. Der findes dog muligheder for at dispensere for godkendelseskravet i energisektoren, såfremt der ek-

sempelvis er tale om et udviklingsprojekt eller andet, der bibringer projektet værdi, som ikke er medregnet i den samfundsøkonomiske analyse.

Tilsvarende vejledninger i samfundsøkonomiske analyser er udarbejdet af andre planlægningsmyndigheder, og de foreskriver anvendelse af Finansministeriets diskonteringsrente på deres respektive planlægningsområder, Energistyrelsen (2007, 2008), Miljøministeriet (2010), Kystdirektoratet (2009).

Nedprioritering af offentlige investeringer er ikke ensbetydende med, at de aldrig gennemføres, men udskydelsen af investeringer kan også medføre store samfundsøkonomiske tab, jf. f.eks. udskydelsen af udskiftningen af banenettets signalsystem.

Rationaler for diskontering

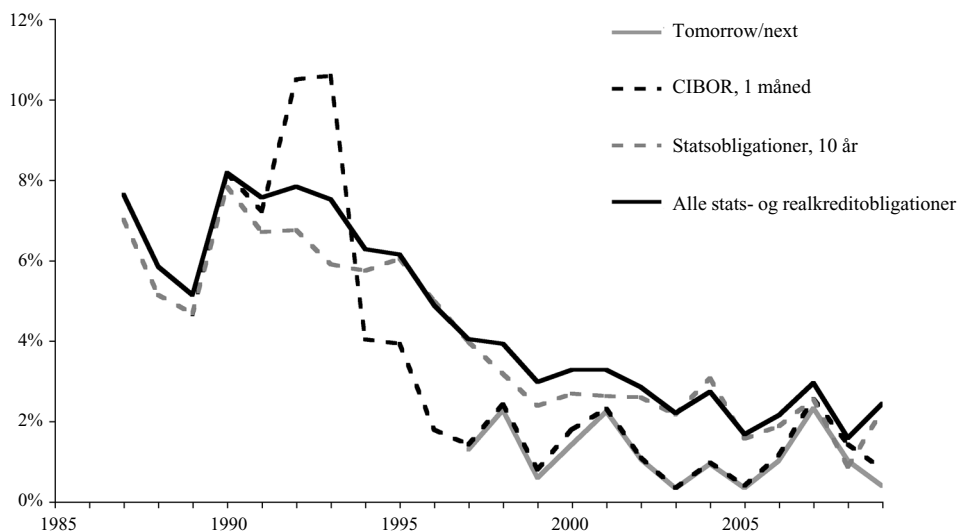
Der er to økonomiske rationaler bag diskontering. Som forbrugere (og dermed også opsparende) er vi tilbøjelige til at foretrække fremskyndet frem for udskudt forbrug. Som investorer forlanger vi et afkast af den investerede kapital, der er mindst lige så stort som det afkast, vi kunne have fået ved at investere i alternative kapitalanbringelser. Disse rationaler afspejles i begge de to hovedtilgange til empirisk indkredsning af kalkulationsrenten »social rate of time-preference« og »social opportunity cost«, Lind (1982), der kan karakteriseres som henholdsvis »prescriptive« og »descriptive«, Arrow m.fl. (1996). Der er også tilgange, der medregner efficienstab ved forvridende skatter i kalkulationsrenten, men det kan med fordel analyseres særskilt, som det foreskrives i den danske vejledning.

Den økonomiske debat om kalkulationsrenten handler i høj grad om, hvilken af disse to tilgange, man bør vælge, og om man bør lægge vægt på forbruger- eller investorrationalet. En stor del af litteraturen konkluderer imidlertid, at begge rationaler må spille en rolle for den samfundsøkonomiske kalkulationsrente. I praksis bliver det til enten et vejet gennemsnit eller en fællesmængde af de mulige kalkulationsrenter, som kan udledes til at repræsentere de to rationaler.

Det er også tilfældet i Finansministeriets vejledning, Finansministeriet (1999). Her giver man afkald på at beregne den samfundsøkonomiske tidspræferencerate, idet der er for stor usikkerhed forbundet med at estimere parametrene. I stedet estimeres forbrugerdiskonteringsrenten som den disponible realrente for bankernes indlån fra og udlån til forbrugerne, der historisk har svinget imellem 2% og 15%.

Investordiskonteringsrenten beregnedes som en gennemsnitlig omkostning ved kapital investeret i selskabssektoren (Weighted Average Cost of Capital – WACC) til 6-11%. Det baseredes på historiske realafkast af aktier på 6-13%, en tilsvarende historisk realrente på fremmedkapital på 4-6% og en forudsætning om 40% egenkapital.

Disse ræsonnementer fører frem til en kalkulationsrente på 6%, men rummer en række metodiske problemer, som vil blive diskuteret i det følgende.



Figur 1. Reale korte pengemarkedsrenter og effektive obligationsrenter 1987-2009.

Kilde: Danmarks Statistik, Statistikbanken (Danmarks Statistik 2010).

Investorrationalitet

Investeringer i markedsmæssige aktiviteter indebærer en risiko for ikke alene at miste sin fortjeneste, men også at miste sin kapital eller noget heraf. Denne risiko må investor nødvendigvis tage hensyn til i sit afkastkrav, som er den diskonteringsrente, en investor diskonterer en aktivitet's fremtidige pengestrømme med i vurderingen af, om den er rentabel.

Afkastet af aktier er derfor typisk højere end afkastet af obligationer. Det gennemsnitlige realafkast af aktier er for eksempel opgjort til 7,1% om året over perioden 1922-1999, Nielsen, Risager (2001). Afkastraten svinger dog meget, og gennemsnittet er meget afhængigt af periodeafgrænsningen.

En standardmetode til analyse af afkast og risici er the Capital Asset Pricing Model (CAPM), Sharpe (1964), Lintner (1965, 1969), ifølge hvilken afkastkravet for investeringer i markedsmæssige aktiviteter kan opdeles i en risikofri rente og en risikopræmie.

Den risikofri rente er den rente, investorerne kræver for blot og bart at stille kapital til rådighed med fuld sikkerhed for at få den tilbage efter en periode. Den er en kompensation for ikke at kunne bruge sin kapital til andet i det tidsrum. Den kan ikke direkte aflæses af markedsdata, men man kan bruge renten af en meget kort kredit som tilnærmelse.

I Danmark kan den af Nationalbanken siden 1997 opgjorte tomorrow/next (dag til

dag) pengemarkedsrente bruges som tilnærmelse til den risikofri rente. Den reale T/N-rente er vist i figur 1 og den er ikke så forskellige fra andre korte renter som CIBOR med 1 måneds løbetid. Man skal dog ikke langt tilbage i historien før de korte penge-markedsrenter kommer til at afspejle pengepolitik og valutakriser i en grad, som næppe kan lægges til grund for den forventede fremtidige risikofri rente. Gennemsnittet for den reale T/N-rente 1997-2008 er 1,2%.

I tillæg til den risikofri rente vil afkastkravet inkludere en risikopræmie, som afspejler de systematiske, men ikke de aktiv-specifikke (ikke-systematiske) risici. Er porteføljen tilstrækkeligt diversificeret, kan man regne med, at de ikke-systematiske risici udligner hinanden. Det er derimod ikke muligt at diversificere sig ud af de risici, der rammer hele aktiemarkedet og dermed hele porteføljen. Disse såkaldt systematiske risici omfatter blandt andet risici for recession og energikriser, rente og valutakursændringer, beskatning og anden lovgivning, finanssektor- og boblekollaps, osv med de realøkonomiske og finansielle kædereaktioner, som disse stød fører med sig. Finans-krisen gav en påmindelse om, at selv virksomheder uden ikke-systematiske tab kan risikere konkurs, når efterspørgslen forsvinder, betalingerne for leverede varer og tjenester udebliver, og det er umuligt at låne til nye aktiviteter.

Diskonteringsrenten eller det reale afkastkrav for ansvarlig kapital kan således beskrives ved en risikofri rente plus en risikopræmie, som afspejler i hvilket omfang, afkastet samvarierer med hele aktiemarkedets afkast. Alle investeringer i industri og andre markedsmæssige aktiviteter løber denne risiko uanset, om det er staten eller private, der investerer. Det er helt afgørende for den økonomiske logik, at man sammenligner afkastet af alternative investeringer med samme risici. Ellers ville staten kunne overflødiggøre en stor del af skatterne ved at finansiere aktieopkøb med salg af statsobligationer. Det er navnlig her, Finansministeriets analyse, der ligger til grund for kalkulationsrenten på 6%, er for unuanceret.

I flere europæiske lande var der tidligere en tradition for relativt store statslige investeringer i industri og service, og der var god grund til at vurdere disse ud fra deres markedsrisiko. Når man i Frankrig nedsatte kalkulationsrenten fra 8% til 4% (2% efter 30 år) var det blandt andet med henvisning til, at den franske stat, som hidtil har haft et stærkt engagement i industri og bankvæsen, i vidt omfang har trukket sig og fortsat trækker sig ud af markedssektorerne, Commissariat général du Plan (2005).

Den danske stat har imidlertid ikke på samme måde investeret offentlige midler som ansvarlig kapital i industri og bankvæsen og har derfor ikke haft samme grunde til en høj kalkulationsrente ved investering af offentlige midler. Der er dog eksempler, og mange af disse blev netop omdannet til aktieselskaber i 1990erne. Staten har aktier i SAS og DONG af strategiske årsager og desuden i selskaber, der står for infrastruktur (transport, el-transmission, kommunikation) og byudvikling.

Til forskel fra andre lande med et stærkt statsligt engagement i industri og service, har Danmark i højere grad koncentreret sig om investeringer i produktion af offentlige og fælles goder. Risiciene forbundet hermed er af en helt anderledes karakter end risiciene forbundet med investeringer i markeds-mæssige aktiviteter. Markedsfinansierede projekter har både en risiko for at efterspørgslen svigter, og at betalingsforpligtelserne for den faktiske efterspørgsel misligholdes (kreditrisiko). Disse risici har skattefinansierede projekter ikke. Tjenesterne er så at sige købt på forhånd og skatteyderne hæfter kollektivt for betalingen, så enkeltpersoners misligholdelse ikke er en risiko.

Skattefinansierede investeringer i produktion af offentlige og fælles goder adskiller sig således fra privat finansierede investeringer ved, at alle tab bæres ligeligt – og lovpligtigt – af alle borgere, som også deler alle gevinster. Der er således ikke brug for den risikopræmie, der er nødvendig for, at mobilisere risikovillig kapital i den private sektor. Arrow og Lind (1970, 1972) når på denne baggrund frem til, at kalkulationsrenten for offentlige projekter af denne type bør være den risikofri rente.

Der er dog også systematiske risici forbundet med obligationsfinansiering. De udgør nødvendige omkostninger ved finansiering af offentlige investeringer. De omfatter risici som inflation og valutakursændringer, der afspejles i en højere effektiv rente på statspapirer, jo længere løbetid ("rentekurven"). For lande med høje gældskvoter og små vækstudsigter afspejles den lavere kreditværdighed ligeledes i renten. Hvis man medregner disse, bliver konklusionen, at offentlige investeringer, som ikke har nogen markedsrisiko, bør diskonteres med den effektive realrente af statsobligationer med en passende løbetid, f.eks. 10 år.

Derved ville den samfundsøkonomiske kalkulationsrente blive tæt på den faktiske rente ved finansiering af offentlige investeringer. Det ville imidlertid efterlade et problem med sammenligning af private og offentlige enheders løsning af offentlige opgaver. Der er dokumenteret en betydelig forekomst af *forhåndsoptimisme* («*optimism bias*») i planlægning af offentlige investeringsprojekter både i Danmark og i udlandet, Bruzelius, Flyvbjerg & Rothengatter (2002) og andre arbejder af Bent Flyvbjerg. Det medfører budgetoverskridelser, som dækkes ind af offentlige budgetter i øvrigt.

Udlisitering og de mere komplekse offentligt private partnerskaber (OPP) anvendes i stigende omfang med det formål at undgå budgetoverskridelser og i det hele taget at effektivisere opgavevaretagelsen. Hvis projektet udføres på kontrakt med et selskab adskilt fra den ansvarlige myndighed, vil disse overskridelser skulle dækkes af selskabet. Det vil i så fald skulle indregne risikoen for budgetoverskridelser i sit bud – enten som højere forventede udgifter eller højere diskonteringsrente. Det private bud vil i så fald ikke kunne sammenlignes direkte med den offentlige kalkule, hvis denne hverken medregner risikoen for budgetoverskridelser i de forventede udgifter eller i kalkulationsrenten.

I Norge har man forsøgt at løse problemet ved at benytte en risikojusteret diskontingsrente. Kalkulationsrenten på 4% er således sammensat af en risikofri rente på 2% og et tillæg for moderat systematisk risiko på 2%. Projekter med høj systematisk risiko bliver tillagt 4%, så kalkulationsrenten bliver 6%. Projekter, der konkurrerer med den private sektor, kalkuleres med den risikofri rente på 2% plus et risikotillæg, der svarer til det, de private investorer står overfor, Det Kongelige Finansdepartement (2005).

Det principielle problem ved at håndtere systematisk risiko på denne måde er imidlertid, at man implicit forudsætter, at sandsynligheden for, at det lavere afkast indtræffer, kan beskrives ved en eksponentialfunktion ligesom diskonteringsfunktionen. Hvis dette ikke er tilfældet, vil man kunne få en bedre vurdering af projektets risici ved at omregne de enkelte usikre elementer i analysen til sikkerhedsækvivalenter og/eller anvende et udtømmende sæt af scenarieanalyser. Det er en tilgang, man nu synes at hælde til i Norge, Finansdepartementet (2005) og anbefaler i UK, Her Majesty's Treasury (HMT) (2004) og Frankrig, Commissariat général du Plan (2005). En udførlig diskussion af problemet fører også Price (1993) til samme konklusion.

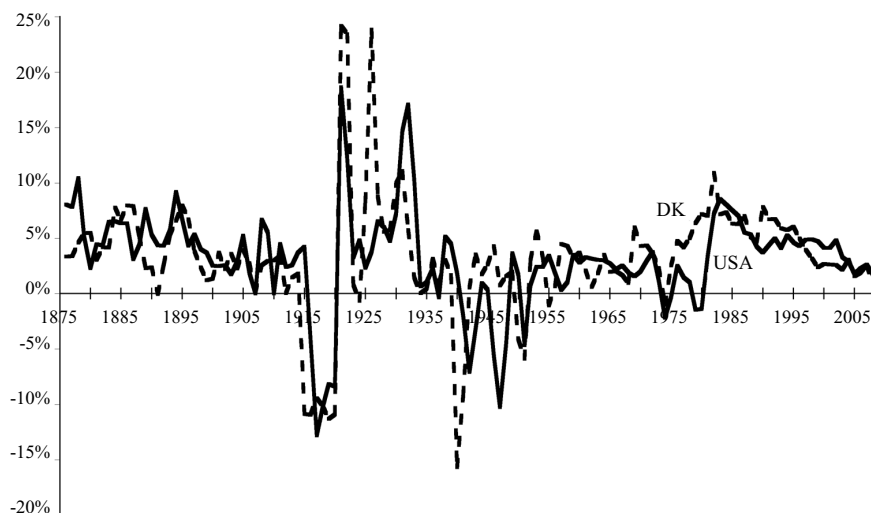
En sådan detaljeret analyse og kvantificering af de involverede risici frem for en ureflekteret anvendelse af aktiemarkedets risikopræmie er også hensigtsmæssig set i lyset af, at risici ved et i sidste ende skattefinansieret projekt adskiller sig principielt fra risici ved markedsfinansierede projekter. Det skattefinansierede projekt indebærer begrænsede risici for efterspørgselssvigt og for at blive overhalet af konkurrenter og ingen risici for misligholdelse af betalingsforpligtelser, partners konkurs og andre effekter af finansielle og reale kædereaktioner.

Det empiriske grundlag for det alternative kapitalafkast

Finansministeriets vejledning baserer som nævnt de 6% på et vejet gennemsnit af aktieafkast og renteniveauet for fremmedkapital. Derved kommer man til at indregne risikopræmier for en række risici, der typisk ikke er relevante for offentlige investeringer eller andre politisk drevne investeringsplaner. Hvis man i stedet tager udgangspunkt i den effektive rente af statsobligationer med en passende løbetid, får man automatisk den risikofri realrente plus en præmie for risikoen for inflation, valutakursændringer osv.

Finansministeriets vejledning anvender et gennemsnit af de historiske renter i en uspecificeret periode, mens det tyske finansministerium anvender et løbende gennemsnit af den reale effektive obligationsrente over de seneste 5 år, Bundesministerium der Finanzen (2010). Som det ses af figur 2 og 3, skal man imidlertid være meget opmærksom på, hvilken periodeafgrænsning, man anvender.

Figur 2 viser den effektive realrente af 10-årige statsobligationer i USA og Danmark i en periode fra 1876 til 2008, beregnet på grundlag af den effektive rente og de respektive forbrugerprisindeks.



Figur 2. Den effektive realrente for 10-årige statsobligationer i Danmark og USA. 1876-2008.

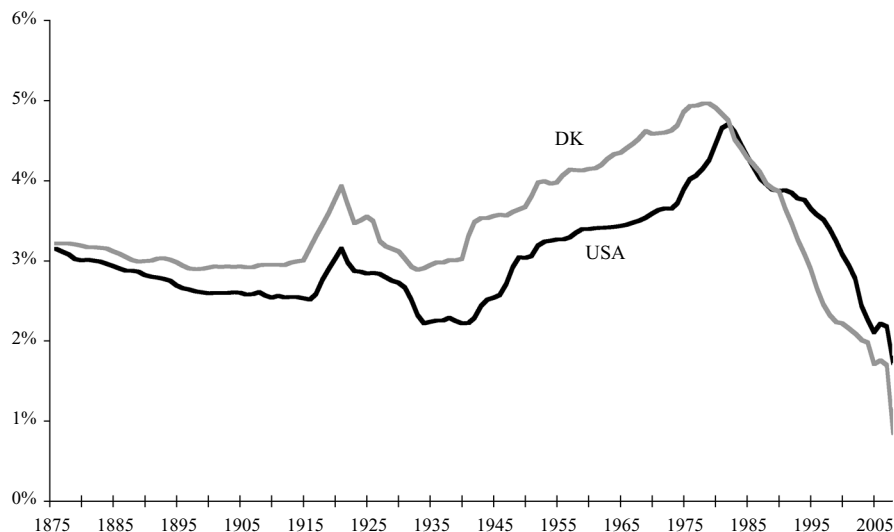
Kilde: Egne beregninger med $(1 + i_t)/(1 + p_t) - 1$ på rente- og forbrugerprisdata, Abildgren (2005), Danmarks Statistik (2009), Officer (2008a, 2008b).

Det fremgår af figur 2, at den danske realrente, som man kunne forvente, i de store linjer har fulgt den internationale repræsenteret ved amerikanske statsobligationer ret tæt fra den sidste del af det 19. århundrede og frem til begyndelsen af det 21. århundrede. Den danske rente afveg dog i perioder markant fra renten i USA, som i slutningen af 1970'erne og begyndelsen af 1980'erne, da der var stor usikkerhed om kronekursen.

Renten er i høj grad bestemt af penge- og valutakurspolitikken i de toneangivende økonomier. Realrenten var internationalt påvirket af en stram pengepolitik i perioderne omkring 1930 og 1980. Verdenskrigene trak den modsatte vej på grund af pris- og rentekontrol som modsvar på den inflation, som vareknapheden forårsagede.

Den gennemsnitlige realrente for den danske 10-årige statsobligation var 3,2% i perioden 1876-2008. Som det fremgår af figur 3, har det stor betydning for gennemsnittet, hvor langt man går tilbage i historien for at tage det.

Figur 3 viser ikke overraskende, at det historiske gennemsnit er højere, jo større en del højrenteperioderne udgør af den periode, som gennemsnittet tages over. Afgrænses perioden fra slutningen 1970'erne, kommer perioden med valutakursusikkerhed og høj rente til at fylde meget i gennemsnittet, som kommer op på 5%. Går vi et århundrede eller mere tilbage, så de ekstreme perioder udjævnes, bliver gennemsnittet om-



Figur 3. Den gennemsnitlige effektive realrente for 10-årige statsobligationer i Danmark og USA over perioden fra året til 2008.

Kilde: Som figur 2.

kring 3%. Tager vi kun gennemsnittet fra for få år siden, kommer vi ned omkring 1%. Ingen historiske gennemsnit over perioder, der ender i 2008, kommer op på 6%.

De særligt høje og særligt lave realrenter fandt sted i perioder med henholdsvis meget stram pengepolitik og verdenskrige, og de udligner i nogen grad hinanden i århundredets gennemsnit. Forventninger til det 21. århundredes renteniveauer må således rumme et betydeligt element af subjektive vurderinger af, i hvilket omfang denne type perioder med ekstreme realrenter vil optræde, og i hvilket omfang de vil udligne hinanden. Nedenfor introduceres en mere udviklet statistisk ramme, der kan håndtere denne fundamentale usikkerhed om den fremtidige udvikling.

Forbrugerrationalet

Det andet rationale er vore præferencer som forbrugere for tidligere hellere end senere forbrug.

Finansministeriet (1999) anvender pengeinstitutternes udlåns- og indlånsrenter som indikator for forbrugernes tidspræferencer og finder et interval på 2-15%. Også her skal man imidlertid være varsom med at sætte lighedstegn imellem markedrenten og den samfundsøkonomiske kalkulationsrente. Dels styres de observerede renter af pengemarkedsrenten, som igen er styret af pengepolitikken, og dels reflekterer markedrenter meget andet end de kollektive tidspræferencer.

Det er velkendt, at forbrugere ofte er mere kortsynede i deres økonomiske dispositioner, end det egentligt er økonomisk rationelt. Den irrationelle kortsigtethed kom historisk ikke mindst til udtryk i folks manglende pensionsopsparing, som tidligt nødvendiggjorde skattefinansieret aldersrente i en række europæiske lande.

Som eksempler på kortsigtet adfærd i dag kunne man pege på unges fravalg af uddannelse, beskæftigedes fravalg af a-kasse og pensionsordninger, impuls køb med strakskredit til meget høj rente for ikke at tale om varierende grader af shopaholics og bruger af korte lån til midlertidigt lav rente til finansiering af langsigtede investeringer. Irrationel diskontering imødegås i Danmark og andre europæiske lande med en vis grad af paternalisme (eller meritgode-regulering) i form af skattefinansieret folkepension, mere eller mindre pånødte arbejdsmarkedspensioner, uddannelse, a-kasse, restriktioner på realkreditfinansiering osv.

Fattige husholdninger og husholdninger uden adgang til kredit har typisk en højere kalkulationsrente og må handle mere kortsigtet end mere velstående husholdninger. Det er især velkendt i udviklingslande, hvor selv små investeringer med udsigt til gode afkast må fravælges på grund af risikoen for at komme under eksistensminimum i en periode.

I det omfang denne kortsigtethed og dermed de høje individuelle kalkulationsrenter sætter aftryk i markedsdata, vil en ukritisk anvendelse af disse som empirisk grundlag for den samfundsøkonomiske kalkulationsrente være problematisk.

Pigou (1950) fremhævede, at når man uden videre sætter lighedstegn imellem de høje renter og afkastkrav, man kan finde på markederne, og samfundets kalkulationsrente, bliver der simpelthen investeret for lidt i det langsigtede: Infrastruktur, vandforsyning, skovrejsning osv. Udnyttelsen af naturressourcerne bliver for generøs over for nutiden på bekostning af fremtiden med velkendte følger som overfiskning, udpining af landbrugsjord og hurtig udtømmning af mineralreserver.

Pigou (1950) kaldte forbrugernes hang til overdreven diskontering for »*irrational discounting*«. Der er ifølge Pigou ikke nogen grund til at lade statens diskonteringsrente styre af forbruger-diskonteringsrenter, der er for høje. Statens diskonteringsrente bør styres af et anderledes fornuftspræget syn på værdien af fremtidige værdier.

Alternativet er at identificere de forbrugerrationaler for diskontering, der egner sig til at blive løftet op som princip for kollektive beslutninger om investeringer.

Ramsey-modellen repræsenterer en sådan tilgang til fastsættelse af kalkulationsrenten, Ramsey (1928). Forbrugers rationale for at tillægge fremtidigt forbrug mindre vægt end nutidigt bygger dels på den klassiske stigende, men marginalt aftagende nyttefunktion og dels på en præference for fremskyndet forbrug uanset indkomstniveauet (*ren tidspræference eller utålmodighed*). Disse rationaler er beskrevet i Ramsey-formlen:

$$r = \mu g + q, (1)$$

hvor r er diskonteringsrenten, g er per capita indkomstvækstraten, μ er forbrugets eller forbrugsmulighedernes grænsenytteelasticitet (μg er forbrugsdiskontering) og q er den rene tidspræferencerate (nyttediskontering). Med denne opdeling kan man skelne imellem de tidspræferencer, der er baseret på indkomstens vækst over tiden og de, der alene er baseret på tidens gang.

Forbrugsdiskontering

Den forventede vækst i forbrugsmulighederne baseres ofte direkte på historiske gennemsnit, men disse afhænger af både det valgte indkomstbegreb og den valgte periode. Da forbrugsdiskonteringen logisk afhænger af indkomstniveauet, anvendes indkomsten per capita.

I princippet er nationalindkomsten det generelle udtryk for de samlede forbrugsmuligheder, men den afhænger meget af nettorenteindtægterne fra udlandet, som igen afhænger af både udlandsgæld og det internationale renteniveau. Det bidrag til NNI-væksten, der i de sidste par årtier er kommet fra nedbringelse af udlandsgælden og renteniveauet, kan næppe forventes at fortsætte permanent, og det relevante indkomstbegreb bliver derfor nationalproduktet.

En del af BNP er forbrug af fast realkapital og er således allerede brugt. NNP er derfor et bedre begreb for forbrugsmulighederne. Den del af NNP, der er ressource-rente fra olie og naturgas, kan imidlertid også betragtes som kapitalforbrug på linje med forbruget af fast realkapital, Hansen (1995), Hansen (1997), Det Økonomiske Råd (1998). Salg af olie- og gas modsvares af en tilsvarende nedgang i værdien af reserverne. Der er således tale om en engangs-indkomst fra et salg af aktiver og ikke en evne til værdiskabelse, der kan gentages og vokse år for år ligesom resten af økonomien. I de seneste par årtier har realiseringen af ressourcekapitalen bidraget til NNP-væksten, og det bidrag vil forsvinde gradvist over de kommende par årtier. En egnet indikator for væksten i forbrugsmuligheder uanset disse i et længere perspektiv forbigående bidrag vil derfor være NNP fratrukket nettooverskuddet ved udvinding af olie og naturgas.

Til gengæld kan forbrugsmulighederne forhøjes med et løbende afkast af den hidtil realiserede ressource-rente, »Golden-rule«-princip beskrevet i Hansen (2001) og parallelt til princippet bag den norske Petroleumsfond, Finansdepartementet (2001). Forudsættes forbrugsmulighederne fra den realiserede olie- og gaskapital fordelt på fremtidige år med en afkastrate på 3% af det kumulerede reale bruttooverskud, kan NNP uden olie og gasproduktion forøges med 1,2% i 2008. Det løfter også vækstraten i forbrugsmuligheder så længe olie- og gasproduktionen står på, men ikke varigt.

De gennemsnitlige vækstrater, der er vist i tabel 1, viser sig at være lavere, jo mere man forsøger, at indkredse de faktiske forbrugsmuligheder. Den reale vækst i værdi-

Tabel 1. Gennemsnitlige vækstrater for BNI, BNP, NNP og NNP uden olie og gasproduktion.

Gennemsnit til 2008 siden	1998	1988	1978	1968
<i>Antal år</i>	<i>10</i>	<i>20</i>	<i>30</i>	<i>40</i>
Bruttonationalindkomst	1,9%	2,0%	1,9%	2,1%
Bruttonationalprodukt, BNP	1,3%	1,6%	1,8%	1,9%
Nettonationalprodukt, NNP	1,0%	1,3%	1,5%	1,6%
NNP uden olie- og gasproduktion	0,7%	1,1%	1,4%	1,5%

Kilder: Danmarks Statistik (2010) og egne beregninger.

Anm.: NNP uden olie- og gasproduktion i løbende priser er beregnet som BNP – bruttooverskud i udvinding af olie og naturgas – forbrug af fast realkapital i alt + forbrug af fast realkapital i udvinding af olie og naturgas. Bruttooverskud i olie og naturgas for 2007 og 2007 er skønnet på baggrund af bruttooverskud i råstofudvinding. Serien er deflateret med nationalregnskabs implicitte NNP deflator.

skabelse (netto) per indbygger i økonomien bortset fra olie og gas har kun været 1,1% over de sidste 20 år, og det er kun godt halvdelen af BNI-væksten.

Gennemsnittene tages over perioder a 10, 20, 30 og 40 år, alle med slutår i 2008. De er ikke alene følsomme over for periodens længde, men også valg af slutår. Samme perioder med slutår i 2009 ville således give markant lavere vækstrater som følge af finanskrisen.

Forudsigelse af fremtidens vækstrater på basis af fortidens giver, som det er tilfældet med renten, anledning til overvejelser om, hvilke scenarier, der er realistiske. Går man f.eks. tilbage til 1966, medregner man den historisk enestående vækstperiode i 1960erne og først i 1970erne. Det er muligt, at fremtiden byder på tilsvarende perioder i forbindelse med den forestående »energirevolution« og andre dramatiske teknologiske skift. Men det er også muligt, at den ikke gør, og at vi må afse en voksende del af forbrugsmulighederne til tilpasning til klimaforandringer.

Lægger vi imidlertid de gennemsnitlige vækstrater i forbrugsmulighederne over de seneste årtier til grund for en vurdering af de fremtidige, kan g antage en værdi i intervallet [0,7%;1,5%] eller [0,7%;1,1%], hvis gennemsnittet kun tages over de seneste 10-20 år. Man kunne yderligere korrigere for det tab af miljøværdier, der også er fulgt med den økonomiske vækst og derved nå frem til et mål for »grønt NNP«, men det vil dels indebære en meget stor usikkerhed omkring tallene og dels forudsætte, at den snævre kobling imellem økonomisk vækst og miljøbelastning, som karakteriserede det 20. århundrede, vil blive gentaget i det 21. århundrede.

Usikkerheden i bestemmelsen af parameteren μ er ikke mindre. Den udtrykker nyttefunktionens krumning og kvantificeres i en nytte- eller velfærdsmaksimerende model. En høj værdi af μ indebærer en antagelse om at et givet beløb gør større nytte i

mindre budgetter end i større. I det omfang, der forventes positiv vækst, vil fremtiden være rigere end nutiden. Med en krum nyttefunktion vil der således være velfærdsgvinster, at hente ved at flytte forbrug fra fremtiden til nutiden, og det vil koste nytte at flytte forbrugsmuligheder den anden vej. Dette ræsonnement svarer til ræsonnementet bag en progressiv indkomstskat, og en nærliggende empirisk strategi er derfor, at gøre parameteren konsistent med den tilsvarende parameter, der kan udledes af skattestrukturen.

En undersøgelse af den nytteelasticitet, der ville ligge til grund for skattestrukturen i europæiske lande, såfremt denne antages maksimere nytten af den samlede indkomst, er udarbejdet på baggrund af OECDs statistik over gennemsnitlig løn i produktionen (APW), Evans (2005), Evans, Sezer (2005). Hvis den faktiske skattebelastning af forskellige indkomstgrupper er udtryk for, at enhvers bidrag repræsenterer samme nytte tab (*equal absolute sacrifice*), kan man estimere en marginal nytteelasticitet. Herpå kan man så antage, at denne tværsnits-nyttefunktion også gælder intertemporalt.

For Danmarks vedkommende fås en værdi på 1,2-1,3. Estimatet afhænger dog meget af datasættets tidsmæssige afgrænsning samt af den vægt, de forskellige indkomstgrupper tillægges. Hertil kommer problemet med realismen i de grundlæggende antagelser om, at den faktiske skattebelastning af de forskellige indkomstgrupper er ideel ud fra *equal absolute sacrifice*-kriteriet. Det er en temmelig krævende forudsætning. Endelig er det ikke uden problemer at sætte lighedstegn imellem skattesystemets omfordeling af samtidige indkomster og den inter-temporale allokering af indkomst.

Man kan også tolke parameteren ud fra en model for indkomstafhængig risikoaversion. Det giver en meget større spredning på de mulige parameterverdier, Cowell, Gardiner (1999).

Den marginale nytteelasticitet kræver også en etisk stillingtagen, hvis fremtidshorisonten er så lang, at nye generationer afløser de nulevende. Så er kalkulationsrenten ikke blot et spørgsmål om intertemporal arbitrage, men om intergenerational fordeling. Så er det ikke blot et spørgsmål om at flytte en generations eget forbrug til andre tidspunkter, men et spørgsmål om at flytte forbrugsmuligheder imellem helt forskellige mennesker.

Det udgør næppe et problem for vurderinger inden for en tidshorisont på omkring 30 år, men det kan være et på længere sigt. Derfor kunne man overveje at anvende en mere fordelingsneutral parameterværdi på det meget lange sigt. Hvis det forudsættes, at nutidsværdien af en procent af 2070-indkomsten skulle være den samme som nutidsværdien af en procent af 2090-indkomsten, måtte parameteren μ have værdien 1, og forbrugsdiskonteringsraten ville på det meget lange sigt svare til den forventede vækstrate.

I UK lægger man parameter værdien 1 til grund med henvisning til, at de fleste estimater i litteraturen ligger omkring 1, Her Majesty's Treasury (HMT) (2004). I Frankrig anvender man imidlertid en parameter værdi på 2, Commissariat général du Plan (2005).

Nyttediskontering

Parameteren q afspejler en præference for at fremskynde forbruget uanset den forventede indkomstudvikling (ren tidspræference eller utålmodighed). Pigous »irrationelle diskontering« er i høj grad sammenfaldende med Ramseys q , nyttediskontering. På dette punkt var der ikke nogen forskel til Fisher (1930), som også begrundede individuelle præferencer for nutidigt frem for fremtidigt forbrug med størrelsen, formen og sandsynligheden af den fremtidige indkomst i forhold til den nutidige. Herudover identificerede han en række faktorer, som også spiller en rolle: (1) kortsigtethed, (2) svag viljestyrke, (3) flothed i forbrugsvaner, (4) læggen vægt på livets korthed og usikkerhed, (5) selvskhed og fravær af ønsker om at sørge for efterladte, (6) underkassen sig modens luner. Det er denne type af motiver til diskontering, der afspejles i q . Fatigdom kan også medføre en høj kalkulationsrente jf. ovenfor.

Økonomer har igennem tiden vedholdende advaret imod at gøre disse grunde til at diskontere til samfundets. Ramsey (1928) fandt nyttediskontering, forstået som en positiv værdi af q , »ethically indefensible and arises merely from the weakness of the imagination« (s. 543). Harrod (1948) betegnede det som »a polite expression for rapacity and the conquest of reason by passion« (s. 40). For Pigou (1950) er nyttediskontering et udtryk for, at »our telescopic faculty is defective« og endda »perverted« (s. 25f). Solow (1974) siger om nyttediskontering: »In social decision making there is ... no excuse for treating generations unequally, and the time horizon is, or should be, very long. In solemn conclave assembled, so to speak, we ought to act as if the social rate of time preference were zero« (s. 9).

Det uetiske består i, at nulevende generationer træffer dispositioner både på egne og fremtidige generationers vegne, men gør det til egen fordel. Nyttediskontering betyder, at man udnytter denne position til at vurdere egen nytte højere end nytten for dem, der kommer senere. Den utilitaristiske etik er derimod, at al nytte er lige meget værd.

På det individuelle niveau prioriterer de fleste nok nytten for nære efterkommere højere end nytten for fjernere efterkommere. Det er der næppe noget etisk forkert i at gøre, men det betyder ikke nødvendigvis, at samfundet skal gøre det samme.

På samfundsniveau afvejes fremtid og nutid i den politiske proces. Derfor er det vore prioriteringer som borgere snarere end vore præferencer som forbrugere, der ligger til grund for afvejningen. Sen (1967) (og senere værker) kalder det isolationsparadokset, når individer i deres egenskab af borgere er villige til at investere mere af hensyn

til fremtiden, end de er som forbrugere. Der er gode grunde til, at paradokset findes, for når man handler isoleret, kan man ikke løse fremtidige samfundsproblemer ved at investere mere. Det kan kun ske ved, at alle gør det samme (»the assurance problem«) eller ved, at staten gør det på vegne af alle. Det kan ses som et »fangernes dilemma«-problem, hvor en optimal løsning forudsætter Kooperation.

På samfundsniveau er der således grund til at formode, at der vil være en større vilighed til at investere af hensyn til »landets fremtid«, »fremtidige generationer« osv., end man kan udlede af de tidspræferencer, der afsløres i forbrugernes adfærd. Vores afvejning af fremtid over for nutid må antages at være anderledes i vores egenskab af forbrugere, der handler isoleret fra hinanden, end i vores egenskab af borgere, der handler i fællesskab.

Koopmans (1960), som ellers også bekendte sig til en etisk præference for neutralitet imellem generationers velfærd, leverede et anerkendt argument for, at den utilitaristiske model med en uendelig tidshorizont logisk må føre til, at der findes en ren tidspræference, $q > 0$. Lidt forsimplet kan man forestille sig, at man i år 0 har mulighed for at foretage en investering, som vil give et konstant afkast i al evighed. Hvis man endvidere forestiller sig, at den fremtidige vækstrate er 0 bortset fra investeringens løft fra år 0 til år 1, ville det optimale være at investere hele den (endelige) forbrugsmulighed i år 0 for at opnå den uendelige forbrugsmulighed i fremtiden, såfremt der ikke var nogen nyttediskontering overhovedet. Implikationen heraf er, at q har en positiv om end ikke nødvendigvis en høj værdi.

Parameteren q afspejler på det individuelle niveau ikke *kun* en i Fishers forstand irrationel utålmodighed, men også en individuel risiko for at miste muligheden for at nyde frugterne af at udskyde forbruget til senere. Vi kan dø, miste evnen til at nyde materielle goder eller slet og ret berøves vore aktiver. Det kan således være rationelt for individer, at tage højde for denne type af risici, mens samfundet ikke står over for de samme risici.

Man kan selvfølgelig ikke afvise muligheden af, at menneskeheden eller blot det enkelte samfund kollektivt går til grunde i løbet af eksempelvis det 21. århundrede. Dele af Bangladesh og hele små østater er begyndt at regne på tiden for den endelige evaluering som følge af havstokkens stigning. Tredje Verdenskrig er nok meget usandsynlig, men næppe umulig ligesom terrorangreb med masseødelæggelsesvåben. Hvis man skal regne med økonomiens undergang, må det i hvert fald blive med en meget lille parameterværdi.

Både Her Majesty's Treasury (2004) og Stern (2007) bruger små, men positive værdier for denne parameter med henvisning til denne type af risici. Den franske planlægningsmyndighed finder imidlertid ikke, at der er grund til nyttediskontering overhovedet på det samfundsmæssige niveau, Commissariat général du Plan (2005).

En kalkulationsrente for den nære fremtid

De disponible realrenter, som forbrugere (og dermed opsparende) står overfor, afspejler både diskontering, som er relevant og diskontering, som er irrelevant for prioriteringen på samfundsøkonomisk niveau.

Ramsey-formlens parametre til bestemmelse af kalkulationsrenten giver også en stor spændvidde i mulige estimater af samfundsøkonomiske kalkulationsrenter. Det er vanskeligt at sige med sikkerhed, hvilke der kan forsvares økonomisk. Med de ovenfor beskrevne mulige parameterverdier ($g = 0,7-2,1\%$; $\mu = 1-2$; $0 > q > 1$), kan der argumenteres for anvendelse af kalkulationsrenter i intervallet $0,8-5,2\%$. Hvis vi regner med olie- og gaskorrigeret VNP-vækst og en meget lille værdi af q , vil den samfundsøkonomiske diskonteringsrente snarere være omkring $1-2\%$.

Det empiriske grundlag peger altså på både en risikofri realrente og en samfundsmæssig tidspræference omkring $1-2\%$. I ræsonnementerne bag disse indgår imidlertid ikke risici og transaktionsomkostninger og andre af den virkelige økonomis realiteter. Som investor er der en risiko for ikke at få den investerede kapital tilbage, og som forbruger/borger er der en risiko for ikke at kunne nyde servicen eller for, at præferencer/prioritering har ændret sig når vi når så langt.

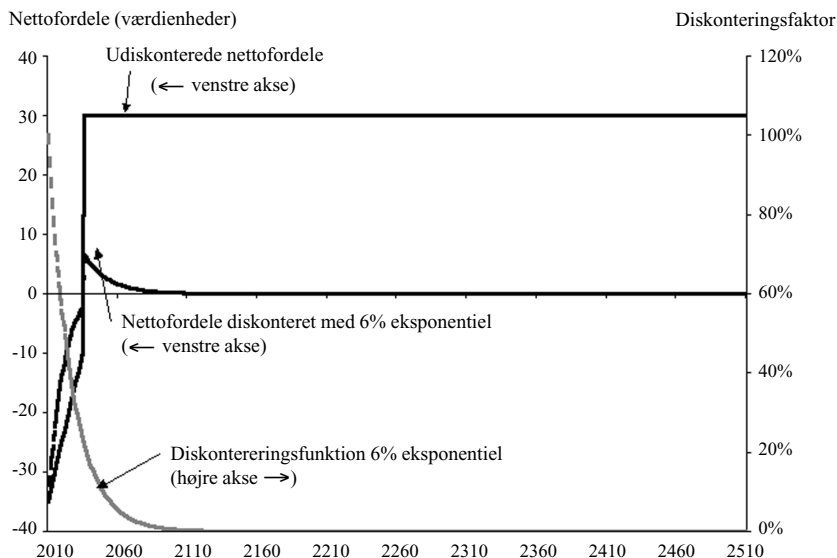
På baggrund af ovenstående kan den effektive realrente af en 10-årig statsobligation være en passende tilnærmelse til det alternative afkast, som ud fra et investorperspektiv bør danne grundlag for en samfundsøkonomisk kalkulationsrente. Både det reale aktieafkast og renten af fremmedkapital inkluderer kompensation for risici, der ikke er relevante for offentlige og lignende investeringer.

Det vil dog ikke være hensigtsmæssigt at basere den på et gennemsnit af for eksempel de seneste 5 års realrente. Det ville give en variabel kalkulationsrente, som kunne ændre prioriteringen imellem kort- og langsigtede projekter med få års mellemrum. I stedet bør kalkulationsrenten baseres på en mere varig vurdering af det forventede realrenteniveau.

Den gennemsnitlige effektive realrente varierer imellem 1% og 5% afhængigt af hvor lang en periode, gennemsnittet tages over op til 130 år. Der er således en betydelig variation, og subjektive skøn er nødvendige i vurderingen af, om der vil være tilsvarende udsving i fremtidige århundreder, og om de vil udligne hinanden som det 20. århundredes to højrente-perioder og to lavrente-perioder.

På denne baggrund kunne man anbefale at bruge et bredt spænd af kalkulationsrenter. En standard med $3\% \pm 2\%$ vil være dækkende ud fra investorperspektivet, men $2,5\%$ eller $3,5\%$ kunne også vælges på det ovenfor beskrevne empiriske grundlag.

Hensyn til både forbruger og investorperspektivet giver visse indbyggede konsistensproblemer både med hensyn til efficiens og etik. Forudsætter man, at fremtiden domineres af højrenteperioder som i 1980'erne, må man samtidig forudsætte, at den



Figur 4. Nutidsværdi af tænkt eksempel på en langsigtet handlingsplan med virkninger i den fjerne fremtid. Diskonteret med 6% kalkulationsrente.

også domineres af lav økonomisk vækst. En høj kalkulationsrente fra et investorperspektiv kan således give en lav kalkulationsrente baseret på et forbrugerperspektiv.

Det kan også tænkes, at man, efterhånden som stadig flere projekter bliver EU-støttede og i øvrigt indgår i tværnationale samarbejder, vil nærme sig en fælles EU kalkulationsrente, jf. EU-Kommissionen vejledning, Commission of the European Communities, DGRP (2008).

Faldende kalkulationsrente i den fjerne fremtid

Det kan være gavnligt, at illustrere virkningerne af en høj kalkulationsrente med et tænkt eksempel på en handlingsplan for langsigtet omstilling. Danmark har erfaring med denne type af omstillinger fra eksempelvis omstillingen af spildevandshåndteringen fra 1970'erne og frem, omstilling af landbrugets næringsstofgennemstrømning og omstillingen til vedvarende energi og energieffektivitet. Handlingsplaner for denne type af omstillinger vurderes også under anvendelse af Finansministeriets kalkulationsrente.

Figur 4 viser en typisk handlingsplan, som starter i år 2010 med løbende investeringer i 25 år, der udløser fordele efterhånden som planen gennemføres og store konstante fordele, som høstes i en uendelig tidshorisont, når planen som helhed er gennemført.

De udiskonterede nettofordele i figur 4 multipliceres med diskonteringsfunktionen for at få nutidsværdien af nettofordelene. Planens samlede nutidsværdi er arealet imellem den kurve, der viser nutidsværdien af nettofordelene, og den vandrette akse. Med en 6% kalkulationsrente og planlægningshorisonten 2010-2080, er planens totale nutidsværdi i år 2010 -246 værdienheder.

Det gør desuden næsten ingen forskel om omstillingens fordele faktisk ophører i 2080 eller om de kan vare ved i flere århundreder herefter. Selvom planlægningshorisonten forlænges til 2310 eller endnu længere, således at flere århundreders ekstra nettofordele medregnes, kommer planens samlede værdi ikke over -233 værdienheder. Det er især et problem i det omfang, omstillingerne gennemføres for at fremtidige generationer kan råde over en økonomi, der kan vokse og udvikle sig uden at sætte sit naturgrundlag over styr.

Denne egenskab ved diskonteringen er ikke konsistent med den samfundsøkonomiske prioritering, der udtrykkes i princippet om bæredygtig udvikling. Den samfundsmæssige afvejning af fremtid og nutid indebærer i dag eksplicit en hensyntagen til fremtidige generationer, en inter-generationel etik. Allerede da man med bæredygtigheds-princippet begyndte at stille sig økonomiske spørgsmål, der spændte over flere generationer, stod det klart, at eksponentiel diskontering ikke er særlig velegnet til at blive strakt så langt. Som *The Economist* bemærkede: »*There is also something awkward about discounting benefits that arise a century hence. For, even at a modest discount rate, no investment will look worthwhile*«, *The Economist* (1991 s. 73). Det giver naturligvis også anledning til etiske dilemmaer, hvis løsning af nutidens udfordringer medfører tab af store værdier for fremtidige generationer, og disse bliver diskonteret ud af regnskabet.

Eksponentialfunktionen er imidlertid ikke den eneste mulige matematiske form for diskonteringsfunktionen. Eksponentiel diskontering har vist sig at være et ganske robust matematisk redskab til analyse af investeringer med tidshorisonter på f.eks. 5, 10, 20 eller 30 år. I takt med at politiske beslutninger med meget længere tidshorisonter er kommet på den politiske dagsorden, er der imidlertid sat spørgsmålstejn ved eksponentiel diskontering over så lange tidshorisonter.

Chichilnisky (1996) karakteriserer eksponentiel diskontering som nutidens diktatur over fremtiden, medens ingen diskontering ville være fremtidens diktatur over nutiden. I forlængelse heraf foreslår Heal (1998) en logaritmisk diskonteringsfunktion.

En logaritmisk diskonteringsfunktion er en af flere muligheder for matematisk repræsentation af en kalkulationsrente, der falder over tid i modsætning til eksponentiel diskontering med en konstant kalkulationsrente. De kaldes under et for hyperbolsk diskontering.

Hyperbolsk diskontering observeres også på det individuelle niveau. Allerede Thaler (1981) påpegede, at hvis en person har valgt imellem et æble i dag (A1) eller to i

morgen (A2) og ligeledes imellem et æble om 1 år (B1) eller to æbler om 1 år + 1 dag (B2), kunne en meget utålmodig person da godt finde på at vælge A1 frem for A2 på trods af den gode forrentning, men ingen kunne finde på at vælge B1 frem for B2. Hyperbolsk diskontering på det individuelle niveau er siden grundigt dokumenteret, f.eks. Loewenstein, Prelec (1992), Bateman (1995). Thaler anså hyperbolsk diskontering for at være en dynamisk inkonsistent adfærd, og der er næppe tvivl om, at den kortsigtethed og utålmodighed, der kan konstateres i individers eksponentielle diskontering også vil observeres i samme individers hyperbolske diskontering. Det udelukker imidlertid ikke, at der kan være gode grunde til at bruge hyperbolsk diskontering i samfundsøkonomiske vurderinger.

Hyperbolsk versus eksponentiel diskontering er imidlertid ikke kun et spørgsmål om inter-generationel etik. Weitzman (1998, 2001) fokuserer i stedet på det forhold, at usikkerheden om hvilken vækstbane, en økonomi bevæger sig ind i, vokser, jo længere tidshorizont, man opererer med. Dermed vokser også usikkerheden om kapitalens alternative afkast. Når man tager højde for denne usikkerhed om den fremtidige vækst, betyder det, at den sikkerhedsækvivalente diskonteringsfaktor og den deraf følgende kalkulationsrente nødvendigvis må være faldende i den fjernere fremtid. Med andre ord giver en hyperbolsk diskonteringsfunktion en bedre afspejling af diskonteringen, når man tager i betragtning, at der er usikkerhed om økonomiens fremtidige vækstbaner.

Den empiriske tilgang med at udlede middelværdien af de mulige alternativafkast eller vækstrater giver ikke den ønskede sikkerhedsækvivalente diskonteringsfaktor. Den sikkerhedsækvivalente diskonteringsfaktor for en fremtid med tre mulige scenarier, der kan beskrives med kalkulationsrenterne $r_1 < r_2 < r_3$ og sandsynlighederne hhv. $\alpha_1 + \alpha_2 + \alpha_3 = 1$, er

$$D_t = \alpha_1(1 + r_1)^{-t} + \alpha_2(1 + r_2)^{-t} + \alpha_3(1 + r_3)^{-t} \quad (2)$$

Den er derimod *ikke*

$$D_t = (1 + (\alpha_1 r_1 + \alpha_2 r_2 + \alpha_3 r_3))^{-t}, \quad (3)$$

som man ville få ved at benytte middelværdierne af de mulige kalkulationsrenter. Fejlen er ikke så stor for den nære fremtid, men for længere tidshorisonter bliver den meget mærkbar.

Gollier (2002a, 2002b) beskæftiger sig med den rationelle investors reaktioner på stigende usikkerhed. Stigende usikkerhed vil give anledning til større forsigtighed og således til mere opsparing og lavere kalkulationsrente. Derimod er Gollier (2004) kri-

tisk over for Weitzmans metode, idet fremaddiskontering til et tidspunkt i fremtiden vil give det modsatte resultat af tilbagediskontering til nutiden. Det rører dog ikke ved, at tilbagediskontering må ske ved en over tid aftagende kalkulationsrente. I en fælles artikel bilægges striden da også med konklusionen: »*The long run discount rate declines over time toward its lowest possible value*«, Gollier, Weitzman (2010).

Det afgørende problem er – ifølge sagens natur – at beskrive usikkerheden om de fremtidige vækstrater og alternativafkast af risikofri kapital. Det kan gøres på utallige måder, som undersøges af bl.a. Weitzman (2007a, 2007b) og Gollier, Koundouri & Pantelidis (2008) og Gollier (2008). Usikkerheden involverer bl.a. eksterne støds varige effekter på vækstrater og afkast, altså ikke blot bevægelser omkring en trend. Den overordnede konklusion røres der dog ikke ved: Usikkerhed om fremtidens vækstrater og alternativafkast fører til kalkulationsrenter, der falder over tid.

Spørgsmålet er så, hvordan de mulige fremtidige rater for vækst og alternativt afkast kan bestemmes på nogenlunde sikker grund. Weitzman (2001) selv spørger ganske enkelt et repræsentativt udsnit af verdens økonomer. Det kan forekomme lidt elitært kun at spørge økonomerne, men der er nok ikke så mange andre, der har et redbont bud på det fremtidige alternativafkast af kapital. På baggrund af de således indsamlede observationer beregnedes middelværdi og varians i en gammafordeling, som anvendes til beregning af den sikkerhedsækvivalente diskonteringsfunktion.

En alternativ empirisk strategi har brugt observationerne fra meget lange historiske serier for obligationsrenter som i figur 2 til at beregne en hyperbolsk kalkulationsrente for USA, Newell, Pizer (2003, 2004). Resultatet ligger imidlertid ikke så langt fra Weitzmans.

Man kan også anlægge et mere politisk perspektiv på alternative scenarier, som Li og Lofgren (2000) gør ved at sammenveje scenarier, der prioriterer henholdsvis vækst og miljøbeskyttelse.

Både Weitzman (2001) og andre har påvist, at en trinvis faldende kalkulationsrente kan være en brugbar tilnærmelse til den kontinuert faldende og den løsning er også forfulgt i nogle lande. I dag, hvor analyserne næppe udføres med lommeregner, men med regneark eller statistikprogrammer, må man dog stille spørgsmål, om det egentlig gør beregningerne lettere. Det medfører i hvert fald, at resultaterne bliver afhængige af trinenes arbitrære placering.

Her Majesty's Treasury (2004) anbefaler en forskrift, der lader kalkulationsrenten være konstant 3,5% for tidshorisoner op til 30 år for derefter at lade den falde til 1% over flere trin (31-75 år: 3%, 56-125 år: 2,5%, 126-200: 2%, 301-300 år: 1,5%, 301 + år: 1%). Den franske planlægningsmyndighed anbefaler en kalkulationsrente, som er 4% i de første 30 år og 2% herefter, Commissariat général du Plan (2005). I den ministerielle vejledning, blev kalkulationsrenten dog sat til 3,5% for 30-50 år og 3% for ef-

fekter længere end 50 år ude i fremtiden, Comité des directeurs transports (2005). Diskonteringsforskriften skal revideres hvert femte år.

Den norske planlægningsmyndighed er skeptisk over for anvendelsen af flere forskellige kalkulationsrenter og påpeger, at de store værdier, som har betydning for fremtidige generationer – f.eks. biodiversitet – alligevel ikke kan opgøres i penge, Finansdepartementet (2005). Derfor er det økonomiske spørgsmål, der skal analyseres, snarere at vurdere omkostningseffektive planer i den nærmere fremtid for at bevare eller genskabe en vis biodiversitet.

Det er en vigtig pointe, der stiller spørgsmålstejn ved, om fordelene ved f.eks., at have en økonomi, der kan vokse og udvikle sig uden at sætte klimastabiliteten på spil, overhovedet kan opgøres med den grad af sikkerhed, der må kræves i økonomiske analyser af denne type. Uagtet svaret herpå, gennemføres dog også langsigtede investeringer, der indebærer relativt sikre effekter på indkomster og omkostninger på 50 års sigt og længere.

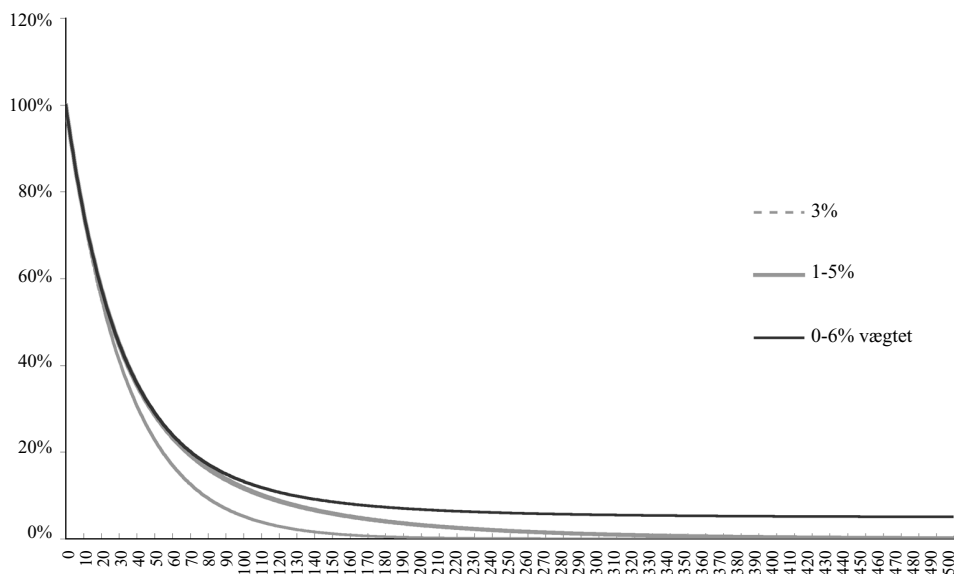
Selv om eksponentiel diskontering har vist sig som et glimrende redskab til at håndtere virkninger inden for en periode på omkring 25 eller 30 år, kommer metoden til kort over for den voksende usikkerhed, når vi inddrager længere tidshorisonter i analysen. Statistisk set vil usikkerheden omkring det fremtidige alternativafkast og den fremtidige økonomiske vækstrate komme til udtryk i en faldende diskonteringsrente i fremtiden. Derfor er der gode grunde til at anvende hyperbolsk diskontering til samfundsøkonomisk vurdering af projekter og planer med længere tidshorisonter end 30 år.

Den hyperbolske diskonteringsfunktions parametre

Det er langt fra klart, hvilken af de foreslåede metoder til at bestemme en hyperbolsk diskonteringsfunktion, der er bedst. Derfor vil en pragmatisk fremgangsmåde med identifikation af et antal mulige diskonteringsrenter og en skønsræssig vægtning af disse være at foretrække. Hvis der eksempelvis indgår 7 mulige scenarier svarende til 7 mulige kalkulationsrenter, skal man tilsvarende fastsætte 7 sandsynligheder.

I figur 5 er vist et eksempel, hvor vi står over for 5 mulige vækstbaner, som er karakteriserede ved kalkulationsrenterne 1%, 2%, 3%, 4%, og 5% og vægtet med *lige store* sandsynligheder. Der er også vist et eksempel, hvor scenarier med 0% og 6% er tilføjet og hvor de forskellige fremtider er tildelt forskellige sandsynligheder ($\alpha_{0\%} = 5\%$, $\alpha_{1\%} = 10\%$, $\alpha_{2\%} = 20\%$, $\alpha_{3\%} = 40\%$, $\alpha_{4\%} = 20\%$, $\alpha_{5\%} = 10\%$, $\alpha_{6\%} = 5\%$). I begge tilfælde er middelværdien af kalkulationsrenterne 3%, men den kalkulationsrente, der kan afledes af de sikkerhedsækvivalente diskonteringsfunktioner, aftager over tid.

Figur 5 viser, at så længe tidshorisonter er under 30 år, er forskellen imellem den eksponentielle og de hyperbolske diskonteringsfunktioner ikke særlig stor. Den viser imidlertid også, at den eksponentielle diskonteringsfunktion selv med den lavere kal-



Figur 5. To hyperbolske og en eksponentiel diskonteringsfunktion.

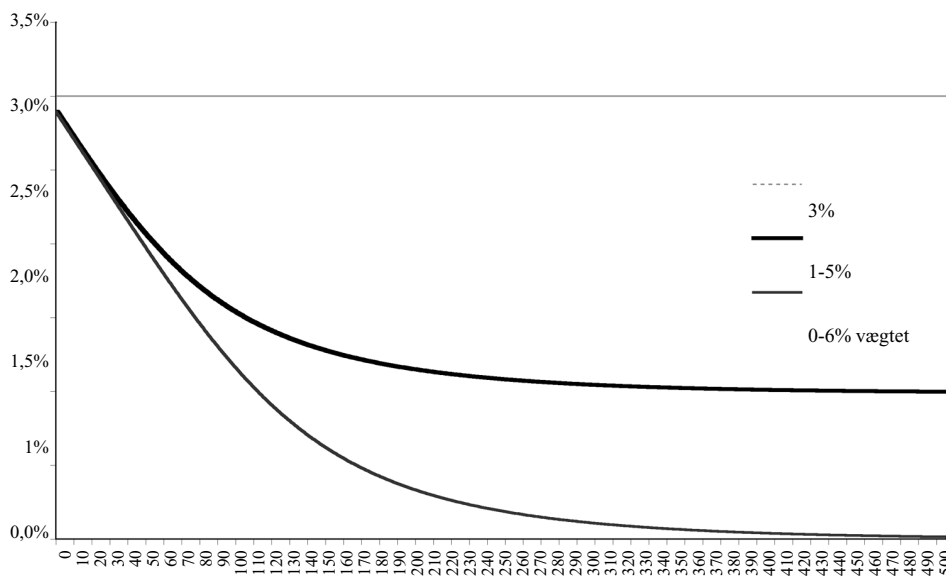
kalkulationsrente på 3%, ikke kan opfange meget af det, der sker i den fjerne fremtid. Den sikkerhedsækvivalente diskonteringsfunktion for mulige scenarier karakteriserede ved 1-5% kalkulationsrenter mere end fordobler denne horisont, mens det inkluderer al fremtid at medtage 0% scenariet selv om 6% scenariet inkluderes samtidigt.

De kalkulationsrenter, der kan udledes af de i figur 5 skitserede diskonteringsfunktioner, er vist i figur 6.

Når den sikkerhedsækvivalente diskonteringsfunktion baseret på scenarier fra 0% til 6% i figur 5 er næsten konstant i den meget fjerne fremtid, afspejler det, at kalkulationsrenten går imod 0, som vist i figur 6. Når 6%-scenariet ikke opvejer 0%-scenariet, som man måske umiddelbart skulle tro, er det fordi scenarier med høje kalkulationsrenter så at sige diskonterer sig selv ud af gennemsnittet, så der til sidst kun er scenarierne med de laveste kalkulationsrenter tilbage. Vægtningen bestemmer, hvor hurtigt kalkulationsrenten falder over tid. Hvis 0%-scenariet havde haft større vægt, ville kalkulationsrenten gå hurtigere mod 0.

Man kan diskutere om et 0%-scenarium ikke er for usandsynligt til at tage med, men Weitzman (1998) insisterer på, at det principielt er muligt, omend med en meget lille sandsynlighed.

Til støtte for en positiv sandsynlighed af et 0% scenarium kan anføres, at der ikke er noget fysisk, der vokser eksponentielt i uendelighed. Den naturlige vækstform er organisk, dvs. med en S-formet vækstkurve, Mesarovic, Pestel (1974), som i hvert fald



Figur 6. Kalkulationsrenter svarende til diskonteringsfunktionerne i figur 5.

asymptotisk nærmer sig nul-vækst. Økonomisk vækstteori behandler også i mange sammenhænge økonomisk vækst som S-formet, Duijn (1983), så økonomiens vækst i hvert fald på delstrækninger ender i en stationær tilstand eller en *steady state*. Selv i en *steady state*, må man dog forvente innovation og dermed noget afkast af investeringer.

De demografiske strukturændringer, vi står overfor, betyder også, at en del af væksten i arbejdstimeproduktivitet kan blive modvirket af en negativ vækst i arbejdstimer per indbygger. Det er dog ikke et argument for en varig nul-vækst i per capita-indkomsten, idet den negative vækst i arbejdstimer per indbygger kun vil være ved til strukturforandringen er tilendebragt, måske i 2040'erne.

De omfattende omstillinger af økonomierne fra fossil til ikke-fossil energi, som vil finde sted i de kommende årtier, har også givet anledning til bekymring, om det vil medføre nul-vækst. De økonomiske analyser viser imidlertid, at den økonomiske vækstrate ikke behøver at påvirkes væsentligt af denne omstilling (Organisation for Economic Co-operation and Development (OECD) 2009, International Energy Agency (IEA) 2009, European Climate Foundation (ECF) 2009, Ingeniørforeningen i Danmark (IDA) 2009).

Til gengæld er der en lille, men ikke desto mindre reel risiko for, at klimaforandringer, der allerede er sat i gang, kan være selvforstærkende, hurtige og voldsomme selv om det lykkes at få begrænset atmosfærens indhold af drivhusgasser til 450 ppm. Der er kun en meget lille sandsynlighed herfor, men det udelukker ikke, at det kan ske.

Et nul-diskonteringsrente-scenarium indebærer dog ikke, at al vækst indstilles i al fremtid. Det forudsætter blot, at det er en mulighed for den periode, der er relevant for en given samfundsøkonomisk vurdering. Derfor kan man nok ikke helt udelukke muligheden af et nul-diskonterings-scenarium. Tilsvarende kan man heller ikke udelukke en fremtid med et alternativt kapitalafkast på 6%, selv om det er højst usandsynligt. Derfor bør den hyperbolske diskonteringsforskrift inkludere begge disse scenarier, men med meget små vægte.

Hvis vi nu ser på det samme eksempel, som anvendtes i figur 4, og anvender dels en eksponentiel diskonteringsfunktion baseret på 3% kalkulationsrente og den hyperbolske diskonteringsfunktion baseret på forskelligt vægtede 0-6% scenarier, vist i figur 5, får vi nogle helt andre nutidsværdier.

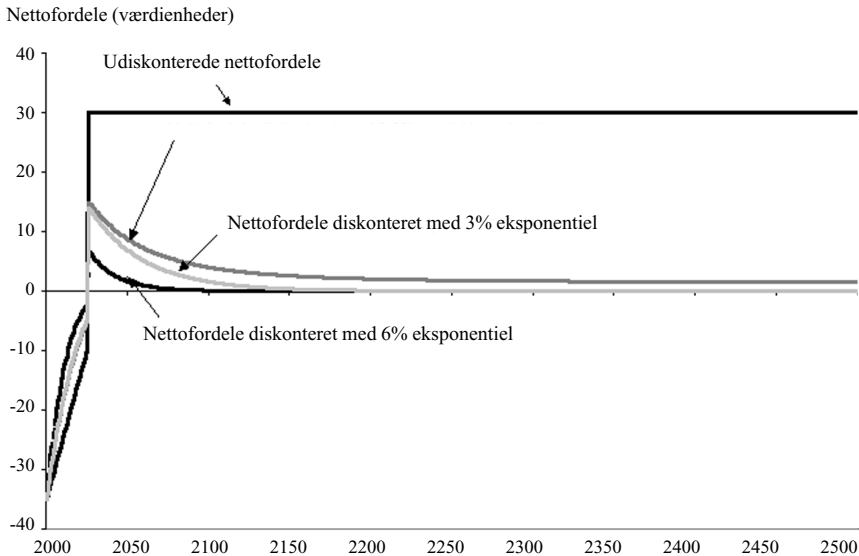
Det er næppe overraskende, at planens nutidsværdi ændrer sig markant ved anvendelse af 3% i stedet for 6% eksponentiel diskontering. Ved 3% er nutidsværdien for en planlægningshorisont på 100 år -18 værdienheder. Ved hyperbolsk diskontering er den imidlertid +125 værdienheder, og denne værdi stiger, jo længere planlægningshorisonten udstrækkes.

Denne tilgang til hyperbolsk diskontering er ikke udledt af et etisk princip, men af en mere reflekteret håndtering af usikkerhed. Derfor løser den ikke den grundlæggende konflikt imellem allokering af samfundets økonomiske goder over tid og fordelingen af dem imellem generationer. Denne konflikt består fortsat som et grundvilkår når man planlægger med meget lang tidshorisont. Men konflikten bliver meget mindre. Det gør i reglen økonomiske valg lettere, når der ikke er så stor forskel på det, der bedst kan betale sig, og det, der er det rigtige at gøre.

Til gengæld er det slet ikke klart, hvordan vægtene bør fastsættes. Man kan for eksempel anvende Weitzmans metode og indsamle svar fra et repræsentativt udsnit af Nationaløkonomisk Forenings medlemmer. Alternativt kunne man indsamle historiske rente-, pris- og vækstdata og på den baggrund estimere fremtidige alternativafkast. Resultaterne vil dog være forskellige, selv om de er udført med videnskabeligt anerkendte metoder. En pragmatisk løsning kunne være, at få et regeringsuafhængigt økonomisk forum til at identificere et antal scenarier med tilhørende vægte.

Tidsinkonsistens

Anvendelse af hyperbolsk diskontering bliver ofte mødt med en vis skepsis på grund af den tidsinkonsistens, der kan optræde, fordi en kalkulationsrente imellem f.eks. år 40 og år 41 ser mindre ud på planlægnings tidspunktet (år 0) end, når man faktisk står i år 40. »Using a discount rate that depends on the period over which the analysis is being conducted is not without problems. For one thing, it leads to time-inconsistent decisions: plans that people will not follow if given the opportunity to recon-



Figur 7. Nutidsværdi af tænkt eksempel på en langsigtet handlingsplan med virkninger i den fjerne fremtid. Diskonteret med eksponentiel (6% og 3% kalkulationsrente) og hyperbolsk (0-6%, vægtet) diskonteringsfunktion.

sider their actions. This property of hyperbolic discounting functions makes many people uneasy about their use in benefit-cost analysis.» , Portney, Weyant (1999).

Det er naturligvis et problem, hvis diskonteringsmetoden i sig selv får den økonomiske rådgivning til at skifte fortegn, efterhånden som et projekt skrider fremad. Der er imidlertid meget stor forskel på arten af investeringer. En ikke-separabel investeringsplan, som kan repræsenteres af en *putty-clay substituerbarhed* (investeringsudgifterne er *sunk cost*) og er *ikke-separabel*, vil ikke være plaget af tidsinkonsistens.

Putty-clay¹ substituerbarhed betyder, at kapital og arbejdskraft er substituerbare på langt sigt, dvs før investeringen, men komplementære på kort sigt, altså når investeringen er foretaget. Denne mangel på substituerbarhed betyder også, at samfundet ikke uden videre kan realisere værdien af den faste realkapital. Hvis det viser sig, at en motorvej ikke kaster så store gevinster af sig, som man havde regnet med, men at lokomotiver til gengæld er meget lukrative investeringer, kan man ikke blot flytte kapitalen fra motorvejene til lokomotiver.

En plan omhandler i reglen en række handlinger, det er nødvendigt at gennemføre for at få et bestemt resultat. Hvis alle handlingerne er nødvendige for at opnå resulta-

1. Putty-clay = kit-ler. Kit kan formes og omformes relativt frit, mens ler er vanskeligere at omforme, især efter nogen tid. Det kan dog lade sig gøre i nogle tilfælde. Eksempelvis kan bygninger ofte bruges til mange forskellige formål, mens lokomotiver ikke kan.

tet, kan man betegne planen som ikke-separabel. Man kan for eksempel ikke få 90% af fordelene ved at bygge en bro ud af kun at bygge 9 ud af de 10 brofag.

Finansiell kapital er derimod putty-putty og separabel. Investeringerne er langt fra *sunk cost* og den daglige porteføljepleje består netop i at veksle en type papirer til andre, som nu synes mere fordelagtige. Hvis man kun sparer halvt så meget op, får man kun halvt så stor pension. Noget tilsvarende gælder fornybare ressourcer som tømmerreserver. Det vender vi tilbage til nedenfor. Det følgende bygger på Hansen (2006).

En handlingsplan eller et projekt, som har putty-clay karakter og er ikke-separabel, og som har først en investeringsfase med konstante omkostninger c i perioden $(0;T)$ og herefter en driftsperiode $(T;\infty)$, hvor omkostningerne er 0 og de konstante fordele er b , vil ved starttidspunktet have nutidsværdien

$$NPV_0 = - \int_{t=0}^{\infty} c_t \Phi_t dt + \int_{t=0}^{\infty} b_t \Phi_t dt \quad (4)$$

Φ er diskonteringsfunktionen, defineret over $(0;\infty)$ hvis værdi til tidspunkt t er diskonteringsfaktoren. Den skal blot være faldende, således at fremtiden vægtes lavere end nutiden, men kan i øvrigt være eksponentiel eller hyperbolsk. Når en del af planen er gennemført, og vi befinder os på tidspunkt τ , vil nutidsværdien af den resterende del af projektet være

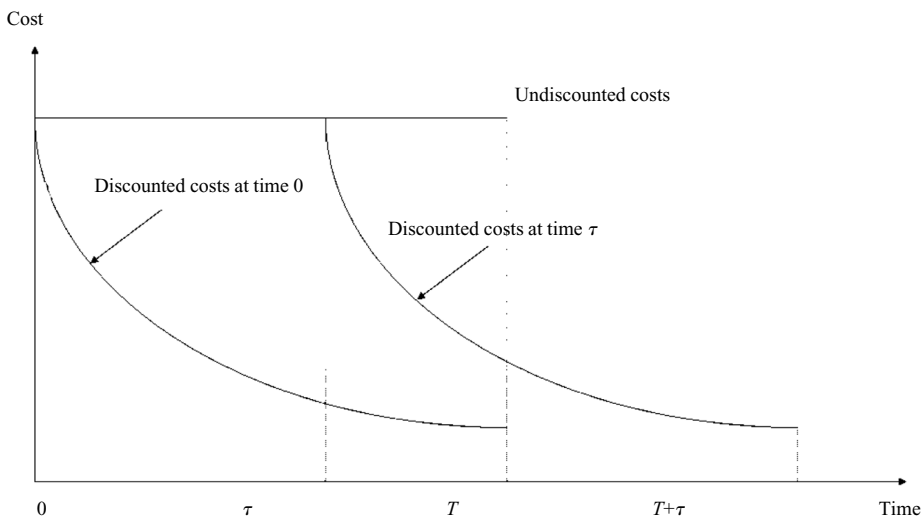
$$NPV_{\tau} = - \int_{t=\tau}^{\infty} c_t \Phi_{t-\tau} dt + \int_{t=\tau}^{\infty} b_t \Phi_{t-\tau} dt \quad (5)$$

Det er let at se, at nutidsværdien af den på tidspunkt τ resterende del af projektet må være større end nutidsværdien af planen på tidspunkt 0, hvis planen er realiseret som planlagt i perioden 0 til τ . Nutidsværdien af de resterende *omkostninger* må nemlig være mindre end nutidsværdien af hele projektets omkostninger på starttidspunktet:

$$\int_{t=0}^T c_t \Phi_t dt = \int_{t=\tau}^{\tau+T} c_{t-\tau} \Phi_{t-\tau} dt \Rightarrow \int_{t=\tau}^T c_t \Phi_{t-\tau} dt < \int_{t=0}^T c_t \Phi_t dt \quad (6)$$

Hertil kommer, at nutidsværdien af fordelene må være større, da de er nærmere i tid:

$$\int_{t=\tau}^{\infty} b_t \Phi_{t-\tau} dt > \int_{t=0}^{\infty} b_t \Phi_t dt \quad (7)$$



Figur 8. Nutidsværdi af investeringsplanens omkostninger.

Det kan også illustreres grafisk som i figur 8.

Resultatet er, at tidens fremadskriden og planens realisering ikke kan gøre planen mindre værdifuld at fuldføre, tværtimod. Jo mere man nærmer sig tidspunktet for at høste fordelene, jo større er deres nutidsværdi og jo flere omkostninger, man allerede har betalt, jo mindre er nutidsværdien af de resterende. Dette gælder uanset diskonteringsfunktionens form og diskonteringsrentens størrelse. Ved en klassisk investeringsplan karakteriseret af putty-clay substituerbarhed og ikke-separabilitet er der altså ikke grund til at være nervøs for tidsinkonsistens.

Investeringer i finansielle aktiver og fornybare ressourcer karakteriseres som nævnt af putty-putty substituerbarhed og separabilitet. Profitmaksimerende adfærd for ejere af denne type af aktiver er at veksle aktiver med udsigt til beskeden værdiforøgelse til aktiver med udsigt til stor værdiforøgelse – en løbende porteføljepleje. Der er intet i den hyperbolske diskontering, der ændrer ved sandsynligheden for på et givet tidspunkt at kunne opnå et alternativt afkast, der er højere.

Ekspontiel diskontering (i kontinuert tid) har den matematiske bekvemme egenskab, at den på et givet tidspunkt forventede alternative afkastrate er identisk med den øjeblikkelige diskonteringsrente (defineret som den relative ændring i diskonteringsfaktoren på tidspunktet). Derfor kan man anvende samme rentesats i den normative analyse af den bedste investering som i den deskriptive forudsigtelse af den fremtidige forvaltning af aktivet. Det kan man imidlertid ikke, når man anvender hyperbolsk diskontering.

Et simplificeret eksempel herpå kan være optimering af en beplantning med henblik på salg af tømmer. En beplantning af en bestemt årgang vokser med en vækstrate, som på et tidspunkt begynder at aftage. Tømmerreservens værdi er rodværdien, hvis vækstrate aftager i takt med den biologiske vækst. Den optimale rotationsperiode bestemmes af forholdet imellem rodværdiens vækstrate og det alternative afkast, der kan opnås ved at realisere rodværdien og investere den i andre aktiver. Den optimale periode afsluttes, når det alternative afkast overstiger rodværdiens vækstrate.

Sandsynligheden for at den alternative vækstrate overstiger en given værdi i et bestemt år og dermed for optimalitetskriteriet, er imidlertid helt uafhængig af den sikkerhedsækvivalente diskonteringsfaktor beskrevet i ligning (2). Derfor vil der kun optræde tidsinkonsistens i forbindelse med diskontering af denne type investeringer, hvis man på planlægningstidspunktet fejlagtigt forudsætter, at de fremtidige ressourceforvaltere vil realisere tømmerreserven på det tidspunkt, hvor dens vækstrate kommer under diskonteringsrenten i stedet for på det tidspunkt, hvor den kommer under det forventede alternative afkast.

En anden type af tidsinkonsistens kan optræde, når et investeringsprojekt med en profil, hvor driftsperioden afsluttes med en betydelig omkostning. Det kan for eksempel være grusgrave og anden råstofudvinding, hvor der skal foretages oprydning og naturgenopretning efter driftsperioden eller atomkraftværker, som sammen med atomaffaldet skal demonteres og deponeres. I disse tilfælde vil nutidsværdien af den afsluttende udgift vokse, efterhånden som driftsperioden nærmer sig sin afslutning. Det er imidlertid ikke noget specielt for hyperbolsk diskontering. Problemet er helt det samme ved eksponentiel diskontering.

Selve problemet med tidsinkonsistens er dog i højeste grad relevant. I Danmark har langsigtede planer som elektrificeringen af jernbanerne, omstillingen af landbruget til en miljømæssigt bæredygtig udledning af næringsstoffer til naturen og omstilling til ikke-fossil energi været præget af stop-go politik. De økonomiske og miljømæssige omkostninger ved denne tidsinkonsistens er betydelige. Det kræver dog en integreret økonomisk-politologisk analyse at forstå baggrunden for denne type af tidsinkonsistens, og hvordan man kan undgå den. Det har ikke så meget at gøre med diskonteringsmetoder.

Konklusion

Anvendelsen af en høj kalkulationsrente i den økonomiske vurdering af offentlige planer og projekter, som den af Finansministeriet foreskrevne kalkulationsrente på 5-6%, kan institutionalisere en overdreven kortsigtethed som rettesnor for politiske prioriteringer. Det kan gå ud over økonomiens udviklingsmuligheder, såfremt de økonomiske anbefalinger, der er baseret på denne kortsigtethed, faktisk følges.

I et investorperspektiv må den samfundsøkonomiske diskonteringsrente afspejle de rene omkostninger ved kapital, den risikofri rente. Det ville give en kalkulationsrentesats på omkring et par procent.

Hertil må man imidlertid lægge de risici, som samfundet løber, for at kapitalen skal gå tabt, og de er ikke helt de samme for aktieinvestorer. Man kan enten indregne dem som risikopræmie i kalkulationsrenten eller som justeringer af de forventede indtægter/fordele og omkostninger ved projektet. Sidstnævnte anbefales.

Risici for valutakurs- og renteændringer, inflation, misligholdelse osv. afspejles i obligationsrenterne. Disse risici repræsenterer nødvendige omkostninger ved at finansiere offentlige investeringer, så de bør regnes med, selv om de nu er meget små i Danmark. Den effektive realrente af en 10-årig statsobligation kan derfor tjene som empirisk grundlag for den politisk-administrativt foreskrevne kalkulationsrente.

I et forbrugerperspektiv må kalkulationsrenten afspejle de tidspræferencer, der er relevante for borgerne som helhed, den kollektive forbruger. De er ikke helt de samme som dem, der kommer til udtryk i de individuelle forbrugeres adfærd.

Diskonteringsrenten må baseres på en fremtidig vækstrate i forbrugsmulighederne (uden olie- og gasindtægter), som skal estimeres i lyset af, at den gennemsnitlige vækstrate i forbrugsmulighederne over de seneste par årtier har været omkring 1% om året.

Et bredt interval af diskonteringsrenter kan være konsistente med det empiriske grundlag, og et vist element af subjektiv vurdering af fremtidige alternativ-afkast og vækstrater er nødvendigt. For projekter og planer med en tidshorisont under 30 år, anbefales eksponentiel diskontering med en kalkulationsrente på 3% +/-2%, forudsat at den forventede værdi af risici kan indarbejdes i projekternes omkostninger og indtægter/fordele.

Hyperbolsk diskontering afspejler mere korrekt den sikkerhedsækvivalente nutidsværdi af fremtidige diskonteringsfaktorer. Den vægtede 0-6% hyperbolske diskonteringsfunktion, der er vist i figur 5 ville give et mere retvisende billede af økonomien i projekter og planer, hvis horisont ligger i en fjernere fremtid.

En pragmatisk løsning kunne være – parallelt til den engelske – at diskontere med 3% +/-2% fra år 0 til år 30 og med en hyperbolsk diskonteringsfunktion fra år 30.

Den hyperbolske diskonteringsfunktion indebærer dels valg af økonomiens mulige fremtider, karakteriseret ved deres eksponentielle kalkulationsrenter og dels en vægtning af disse fremtider med deres sandsynlighed. Et sæt af mulige fremtider med diskonteringsrenter fra 0% til 6% vil i hvert fald dække alle eventualiteter ind.

Vægtene må imidlertid i høj grad bygge på subjektive vurderinger af sandsynligheden for, at disse fremtider realiseres. Det er en udfordring at udvikle en model for fastsættelse af disse vægte.

En hyperbolsk diskonteringsfunktion løser ikke det etiske problem ved at diskontere på tværs af generationer, men det formindsker det væsentligt. Diskontering indebærer en risiko for modstrid imellem diskonteringen af fremtidige værdier til nutidsværdi og den værdi, disse værdier repræsenterer for de fremtidige generationer. Derfor er en inter-generational etik nødvendig også ved hyperbolsk diskontering.

Litteratur

- Abildgren, K. 2005. *A historical perspective on interest rates in Denmark 1875-2003*.
- Arrow, K. J., W. Cline, K. G. Mähler, M. Munasinghe og J. Stiglitz. 1996. Intertemporal equity, discounting and economic efficiency in *Climate Change 1995. Economic and Social Dimensions of Climate Change*, red. I.P.o.C.C. (IPCC), B. J. P., L. H. og H. E. F., Cambridge University Press, Cambridge, MA.
- Arrow, K. J. og R. C. Lind. 1972. Uncertainty and the Evaluation of Public Investment Decisions: Reply, *The American Economic Review*, vol. 62, no. 1/2, pp. 71-172.
- Arrow, K. J. og R. C. Lind. 1970. Uncertainty and the Evaluation of Public Investment Decisions, *The American Economic Review*, vol. 60, no. 3, pp. 364-78.
- Bateman, H. N. 1995. Empirical and Public Choice Evidence for Hyperbolic Social Discount Rates and the Implications for Intergenerational Discounting. *Environmental and Resource Economics*, vol. 5, pp. 413-23.
- BEK nr. 1295 af 13/12/2005 (Godkendelse af kollektive varmforsyningsanlæg) (2005): *Bekendtgørelse om godkendelse af projekter for kollektive varmforsyningsanlæg. BEK nr. 1295 af 13/12/2005*, Klima- og Energiministeriet.
- Bruzélius, N., B. Flyvbjerg og W. Rothengatter. 2002. Big decisions, big risks. Improving accountability in mega projects, *Transport Policy*, vol. 9, no. 2, pp. 143-54.
- Bundesministerium der Finanzen. 2010. *Personalkostenzätze, Sachkostenpauschale und Kalkulationszinssätze für Wirtschaftslieitsuntersuchungen*. RdSchr. d. BMF.
- Bundesministerium der Finanzen. 2009. *Personalkostenzätze, Sachkostenpauschale und Kalkulationszinssätze für Wirtschaftslieitsuntersuchungen*. RdSchr. d. BMF.
- Chichilnisky, G. 1996. An axiomatic approach to sustainable development, *Social Choice and Welfare*, vol. 13, no. 2, pp. 231-57.
- Comité des directeurs transports. 2005. *Instruction cadre relative aux méthodes d'évaluation économique des grands projets d'infrastructures de transport*.
- Commissariat général du Plan. 2005. *Le prix du temps et la décision publique. Révision du taux d'actualisation public*.
- Commission of the European Communities, DGRP. 2008. *Guide to cost-benefit analysis of investment projects. Structural Funds, Cohesion Fund and Instrument for Pre-Accession*.
- Cowell, F. A. og K. Gardiner. 1999. *Welfare Weights*, STICERD, London School of Economics, London.
- Danmarks Statistik. 2010. *Danmarks Statistiskbank*. Available: <http://statistikbanken.dk> [2010, 13.08.2010].
- Danmarks Statistik. 2009. *Danmarks Statistiskbank*.
- Det Kongelige Finansdepartement. 2005. *Behandling av kalkulasjonsrente, risiko, kalkulasjonspriser og skattekostnad i samfunnsøkonomiske analyser*.
- Det Økonomiske Råd. 2008. *Dansk Økonomi efterår 2008*.
- Det Økonomiske Råd. 2007. *Dansk Økonomi forår 2007*.
- Det Økonomiske Råd. 2006. *Dansk Økonomi efterår 2006*.
- Det Økonomiske Råd. 2002. *Dansk Økonomi forår 2002*.
- Det Økonomiske Råd. 1999. *Dansk Økonomi forår 1999*.
- Det Økonomiske Råd. 1998. *Dansk økonomi, efterår 1998*.

- Duijn, J. J. v. 1983. *The long wave in economic life*, George Allen og Unwin, London.
- Energistyrelsen. 2008. *Forudsætninger for samfundsøkonomiske analyser på energiområdet*.
- Energistyrelsen. 2007. *Vejledning i samfundsøkonomiske analyser på energiområdet*.
- European Climate Foundation (ECF). 2009. Roadmap 2050. *A practical guide to a prosperous, low carbon Europe. Technical analysis*.
- Evans, D. J. 2005. The elasticity of marginal utility of consumption: Estimates for 20 OECD countries, *Fiscal Studies*, vol. 26, no. 2, pp. 197-224.
- Evans, D. J. og H. Sezer. 2005. Social discount rates for member countries of the European Union, *Journal of Economic Studies*, vol. 32, no. 1, pp. 47-59.
- Finansdepartementet. 2005. *Veileder i samfundsøkonomiske analyser*.
- Finansdepartementet. 2001. *Retningslinjer for den økonomiske politikken*, Oslo.
- Finansministeriet. 1999. *Vejledning i udarbejdelse af samfundsøkonomiske konsekvensvurderinger*.
- Fisher, I. 1930. *The Theory of Interest. As Determined by Impatience to Spend Income and Opportunity to Invest It*, The Macmillan Co, New York.
- Gollier, C. 2008. Discounting with fat-tailed economic growth, *Journal of Risk and Uncertainty*, vol. 37, no. 2-3, pp. 171-86.
- Gollier, C. 2004. Maximizing the expected net future value as an alternative strategy to gamma discounting, *Finance Research Letters*, vol. 1, no. 2, pp. 85-89.
- Gollier, C. 2002a. Discounting an uncertain future, *Journal of Public Economics*, vol. 85, no. 2, pp. 149-66.
- Gollier, C. 2002b. Time horizon and the discount rate, *Journal of Economic Theory*, vol. 107, no. 2, pp. 463-73.
- Gollier, C., P. Koundouri og T. Pantelidis. 2008. Declining discount rates: Economic justifications and implications for long-run policy, *Economic Policy*, vol. 23, no. 56, pp. 757-95.
- Gollier, C. og M. L. Weitzman. 2010. How should the distant future be discounted when discount rates are uncertain?, *Economics Letters*, vol. 107, no. 3, pp. 350-53.
- Hansen, A. C. 2006. Do declining discount rates lead to time-inconsistent economic advice?, *Ecological Economics* 60(1): 138-44.
- Hansen, A. C. 2001. Estimating Non-Renewable Resource Capital Consumption in *The Sustainability of Long-term Growth*, red. M. Munashinghe, O. Sunkel og C.d. Miguel, Edward Elgar Publishing, Cheltenham, pp. 397-421.
- Hansen, A. C. 1997. Ægte opsparing – et operationelt bæredygtighedskriterium?, *Økonomi og Politik*, vol. 70, no. 3, pp. 34-48.
- Hansen, A. C. 1995. Bæredygtig opsparing og dansk økonomi i *Miljøet, markedet og velfærdsstaten*, red. P. Lübcke.
- Harrod, R. 1948. *Towards a dynamic economics: some recent developments of economic theory and their application to policy*, Macmillan.
- Heal, G. 1998. *Valuing the Future: Economic Theory and Sustainability*, Columbia University Press.
- Her Majesty's Treasury (HMT). 2004. *The Green Book. Appraisal and Evaluation in Central Government*, The Stationery Office, London.
- Ingeniørforeningen i Danmark (IDA). 2009. *IDAs klimaplan 2050*.
- International Energy Agency (IEA). 2009. World Energy Outlook 2009, <http://www.oecdilibrary.org/oecd/content/book/weo-2009-en> edn, OECD/IEA.
- Koopmans, T. C. 1960. Stationary Ordinal Utility and Impatience, *Econometrica*, vol. 28, no. 2, pp. 287-309.
- Kystdirektoratet. 2009. *Manual for udarbejdelse af bestemmelser knyttet til den samfundsøkonomiske analyse samt oversvømmelsesdirektivet*.
- Li, C. og K. Lofgren. 2000. Renewable Resources and Economic Sustainability: A Dynamic Analysis with Heterogeneous Time Preferences, *Journal of Environmental Economics and Management*, vol. 40, no. 3, pp. 236-50.
- Liikenne- ja aluehallinto. 2004. *Tiehankkeiden arviointiohje. Suunnitteluvaiheen ohjaus*, Tiehallinto, Helsinki.
- Lind, R. C. 1982. *Discounting for Time and*

- Risk in Energy Policy*, Resources For the Future, Johns Hopkins University Press-Lind, R.C, Washington.
- Lintner, J. 1969. The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets: A Reply, *The review of economics and statistics*, vol. 51, no. 2, pp. 222-24.
- Lintner, J. 1965. The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets, *The review of economics and statistics*, vol. 47, no. 1, pp. 13-37.
- Loewenstein, G. og D. Prelec. 1992. Anomalies in Intertemporal Choice. Evidence and an Interpretation, *Quarterly Journal of Economics*, vol. 107, no. 2, pp. 573-97.
- Mesarovic, M. og E. Pestel. 1974. *Mankind at the turning point: the second report to the Club of Rome*, E. P. Dutton, New York.
- Miljøministeriet. 2010. *Samfundsøkonomisk vurdering af miljøprojekter*.
- Møller, F. 2009. *Velfærd nu eller i fremtiden. Velfærdsøkonomisk og nytteetisk vurdering over tid*. Århus Universitetsforlag.
- Møller, F., S. P. Andersen, P. Grau, H. Huusom, T. Madsen, J. Nielsen og L. Strandmark. 2000. *Samfundsøkonomisk vurdering af miljøprojekter*.
- Newell, R. G. og W. A. Pizer. 2004. Uncertain discount rates in climate policy analysis, *Energy Policy*, vol. 32, no. 4, pp. 519-29.
- Newell, R. G. og W. A. Pizer. 2003. Discounting the distant future: How much do uncertain rates increase valuations?, *Journal of Environmental Economics and Management*, vol. 46, no. 1, pp. 52-71.
- Nielsen, S. og O. Risager. 2001. *Stock returns and bond yields in Denmark, 1922-99*.
- Officer, L. H. 2008a. *What Was the Consumer Price Index Then? A Data Study*.
- Officer, L. H. 2008b. *What Was the Interest Rate Then? A Data Study*.
- Organisation for Economic Co-operation and Development (OECD) 2009, *The Economics of Climate Change Mitigation Policies and Options for Global Action beyond 2012*.
- Pigou, A. C. 1950. *The Economics of Welfare*, MacMillan, London.
- Portney, P. R. og J. P. Weyant. 1999. *Discounting and Intergenerational Equity*, RFF Press.
- Price, C. 1993. *Time, discounting og value*, Blackwell Publishers.
- Ramsey, F. P. 1928. A Mathematical Theory of Saving, *The Economic Journal*, vol. 38, no. 152, pp. 543-59.
- Sen, A. K. 1967. Isolation, Assurance and the Social Rate of Discount, *The Quarterly Journal of Economics*, vol. 81, no. 1, pp. 112-24.
- Sharpe, W. F. 1964. Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk, *The Journal of Finance*, vol. 19, no. 3, pp. 425-42.
- Solow, R. M. 1974. The Economics of Resources or the Resources of Economics, *The American Economic Review*, vol. 64, no. 2, pp. 1-14.
- Stern, N. 2007. *The Economics of Climate Change. The Stern Review*. Cambridge University Press, New York.
- Thaler, R. 1981. Some empirical evidence on dynamic inconsistency, *Economics Letters*, vol. 8, no. 3, pp. 201-07.
- The Economist. 1991. What price posterity?, *Economist*, no. 23 Mar 91, pp. 97.
- Trafikministeriet. 2001. *Notat til trafikudvalgets høring om elektrificering af bane-nettet*.
- Vägverket. 2006. *Vägverkets samhällsekonomiska kalkylvärden*, Vägverket.
- Weitzman, M. L. 2007a. A review of the Stern Review on the economics of climate change, *Journal of Economic Literature*, vol. 45, no. 3, pp. 703-24.
- Weitzman, M. L. 2007b. Subjective expectations and asset-return puzzles, *American Economic Review*, vol. 97, no. 4, pp. 1102-30.
- Weitzman, M. L. 2001. Gamma Discounting, *The American Economic Review*, vol. 91, no. 1, pp. 260-71.
- Weitzman, M. L. 1998. Why the Far-Distant Future Should Be Discounted at its Lowest Possible Rate, *Journal of Environmental Economics and Management*, vol. 36, pp. 201-08.

A neglected FDI contribution – Arne Lund (1944)

Kurt Pedersen

Aarhus School of Business, and Centre for Business History, University of Aarhus, E-mail: kur@asb.dk

Jesper Strandskov

Faculty of Social Science, University of South Denmark, E-mail: jst@sam.sdu.dk

SUMMARY: From the 1960s a number of theories developed dealing with the concepts of Foreign Direct Investment and Multi-National Enterprises. Much of the new theorising was subsequently integrated into John Dunning's »eclectic paradigm« which still today takes a central position in the business theory of international production. Significant parts of the contributions to the theories of foreign direct investments had, however, been presented in a coherent and systematic way by the Danish economist Arne Lund in an article in Nationaløkonomisk Tidsskrift in 1944. The purpose of this paper is to present Lund's argumentation and give Lund's article, and not least its understanding of entrepreneurship and business dynamics, the recognition it deserves as an early and sophisticated contribution to the theory of foreign international investments.

Introduction – the significance of multinational activity

The rise of multinational enterprises in the second half of the 19th century radically changed the pattern of international production, and by implication international business and economics. Dunning (1988; chapter 3) offers a comprehensive survey of the period 1870-1985 and identifies the main features of the development. In 1914, for instance the accumulated foreign direct investments have been estimated to USD 14 bn. – about 35 per cent of the estimated total long-term international debt at the time. The interwar period saw a decline in multinational foreign investments, but liberalisation of the world economy from the late 1950s once more turned the tide. Radical politico-economic changes in the last 20 years have further propelled the global economy and the role of multinational enterprises. During the last decade UNs yearly *World Investment Report* has become an institution. A few facts from the *WIR 2010* will serve to illustrate the significance and level of foreign direct investments. The share of foreign

The authors thank Mark Casson for valuable comments to an earlier draft; we also are most grateful for a number of very useful remarks from two anonymous referees.

Table 1. Selected figures, 1990-2009, bn. USD.

	1990	2005	2008	2009
FDI inflows	208	986	1771	1,114
FDI inward stock	2,082	11,525	15,491	17,743
Sales of foreign affiliates	6,026	21,721	31,069	29,298
Exports of foreign affiliates	1,498	4,319	6,663	5,186
Employment in foreign affiliates (000)	24,476	57,799	78,957	79,825

Source: World Investment Report 2010, p. 16.

affiliates in global GDP amounted to 11 per cent in 2009, while employment in such units increased to 80 million workers (WIR 2010: xviii).

Half of global foreign direct investments now target developing and transitional economies, and China tops the employment statistics with 16 mill. employed in foreign affiliates (2008). It is remarkable that among the top-5000 multinational enterprises, those headquartered in developing and transition economies in 2009 accounted for roughly 10 per cent of foreign sales and foreign assets – up from less than 2 per cent in 1995 (WIR: 17-18). A few facts would hint to the level of multinational activity in today's world economy.

WTO, in its World Trade Statistics (2010), calculates world trade in merchandise to 15,717 bn. USD, and in commercial services to USD 3,778 bn. It readily appears from the table – where headquarters activity is not included – that multinational activity accounts for a significant part of international trade. It further appears that multinational activity was quite sensitive to the financial crisis of 2007/08.

The development of FDI theory

Almost one hundred years passed from the recognition of multinational enterprises as a historical fact until the concept was subjected to scientific investigation. The distinctive aspect of a *multinational* as compared to other enterprises is its operation and control of business activities across national borders through Foreign Direct Investments. For present purposes, we draw no distinction between categories of foreign direct investments, such as green-field FDI versus mergers and acquisitions, nor do we discuss the exact definition of the concept in terms of the ownership share required for a subsidiary to exist.

Over the years, scholars have repeatedly referred to Canadian economist Stephen Hymer as the founder of the theory of foreign direct investments and, consequently, the one who first grasped the nature of the multinational enterprise, see for instance Forsgren, 2008. Hymer defended his Ph.D thesis in 1960 but it was not published until 16 years later as Hymer (1976). Charles Kindleberger's somewhat apologetic foreword is quite understandable.

Hymer's starting point was the absence of »firms« in received theories on international economics, and particularly in the theory of international capital movements. Even if Hymer's goal was to build a theory of the multinational enterprise, direct investments played a crucial part, witness the words of Forsgren (2008: 15): »*Nowadays, the common view of international business researchers is that Hymer's thesis was a path-breaking achievement that triggered the development of a more comprehensive theory on foreign direct investments. It broke with the traditional view of perceiving foreign direct investment as a phenomenon that could be explained at the macro level*«. At the time, received theory did not distinguish between varieties of capital movements, notably portfolio investments versus direct ones. Neither the causes nor the effects of direct international investments were visible in the literature.

The massive wave of American direct investments into Europe in particular was recognised as a major macro-economic influence on employment and on external balances. Received doctrine argued that capital movements were triggered by interest rate differentials, a misconception which did not end until Hymer's thesis became known. His distinction was based on an investigation of the nature of direct investments and here business motives were crucial. The observation that establishing a subsidiary implied competition with local firms that were knowledgeable about the local market led Hymer to the formulation of two concepts which have been essential to the literature: the Liability of Foreignness and the (Asset of) Firm Specific Advantages. At the time, the performance of American multinationals around the world indicated that FSA tended to outbalance the Liability of Foreignness.

After Hymer, the multinational enterprise was no longer a black box in economic and business theory. It had objectives; it had tools at its disposal; and it had »advantages« that would explain its international investment behaviour. Overseas direct investments had now become *activities* that were related to the investing company's assets. It is well known that, as time went on, Hymer's political attitude became increasingly dominant and the broader social effects of multinational activities became the focus of his work. Suffice it to note briefly that the basic point in this later work was that multinational firm specific advantages spilled over into market dominance; in this respect Hymer was clearly a child of the industrial economics school which had emerged from Harvard University, Bain (1956). Industrial structure determined the conduct of firms and, in turn, the performance of firms and led to supra-normal profits. The mental and theoretical framework was applied micro-economic theory in combination with the distinction in capital movements mentioned above.

While Hymer's argument, in agreement with industrial economics of the Structure-Conduct-Performance thinking, rests on external drivers and firm assets at the core of its analysis, this approach was soon challenged by a number of »competing« theories.

Like Hymer, the Resource Based View took its starting point in the assets of the company, Wernerfelt (1984), Barney (1984), and thus followed a path laid out by Penrose (1959). Unlike Hymer, however, this view focused on the internal shaping and development of company resources. A set of characteristics – such as uniqueness, value, etc. – were defined as necessary for a competence to be »core«, but more importantly the internationalisation of the firm was understood as a development over time of assets (or, core competences). This again gave rise to the concept of »dynamic competences«, i.e. the ability of the enterprise to move towards ever new bases of competitiveness, Strandskov and Pedersen (2008).

Related to transaction cost theory, Coase (1937), Williamson (1975, 1985), theories of internalisation have combined the activity aspect with the resource perspective. Like the resource based view, this approach focuses on the multinational corporation rather than its environment, and argues that, by combining comparative advantages of different countries, the MNE is able to minimise the costs of transactions. This approach has a link to entrepreneurial functions although not quite in the market-related way that one finds in Austrian economics, Hayek (1937), Kirzner (1973). Central to the entrepreneurial aspects of internalisation theory are the building of know-how and the localisation of intangible knowledge across borders (Buckley and Casson 1985). In one important way this approach challenges the theorising of Hymer: as in the resource based approach, firm specific advantages are built, which means that market power is not a »free lunch«. It was paid for in the past when sustainable competitive advantages were created. For obvious reasons, attempts have been made to combine internalisation and RBV theories.

Even if less focused on the multinational aspect of business, Michael Porter (1980, 1985, 1986) contributed to the understanding of FDI and MNE. As in the case of Hymer, Porter's approach is rooted in the Harvard S-C-P philosophy but, unlike Hymer, Porter views activities and not assets as centrepieces. Positioning is seen as the strategic challenge in the competitive industry, Porter (1980), and activities must be coordinated in order to minimise costs in the value chain (system), Porter (1985). It follows that multinational enterprises are forced to navigate along two dimensions, focusing on configuration along the geographical dimension and on the degree of coordination between business units (Porter 1986). In this sense, the approach is a precursor to the corporate network theory that was developed by Bartlett and Ghoshal (1988) and Nohria and Ghoshal (1997), which emphasises the ways in which subsidiaries interact, playing different roles in the MNE business system.

Over the years, John Dunning took central position in the development of the theory of international production. In a vast number of books and articles he has developed and expanded what he termed the »eclectic paradigm«, see for instance Dunning

	Unit of analysis	
	Assets	Activities
External drivers	Hymer (1976) <i>Liability of Foreignness</i> <i>Firm Specific Advantages</i>	Porter (1986) <i>Configuration,</i> <i>Coordination</i>
Internal drivers	Barney (1984), Wernerfelt (1984) Wernerfelt (1984) <i>Resources</i> <i>Dynamic capabilities</i>	John Dunning (1976) <i>O-, L-, I- advantages</i> Buckley & Casson (1985) <i>Entrepreneurship</i> <i>Cost minimisation</i>

Figure 1. Economic theories of FDI (and MNE), 1976-1988.

(1978, 1988). The crux of his argumentation is that in order for a foreign direct investment to take place, some conditions must be met. Broadly he defined three sets of conditions. He changed Hymer's firm strategic advantages into *Ownership advantages*, introduced internalisation of business activity as *Internalisation advantages*, and added *Location advantages*, which were specifically related to the host country. Or briefly: the foreign firm must be competitive, the location must be in some way attractive, and the best way to do business is to keep it in-house. The eclectic paradigm, or OLI paradigm, summarises much of the theorising of the period, embracing assets as well as activities and internal as well as external drivers of MNE activities. Even today the three dimensions are being further developed, Dunning and Lundan (2008), but among them they still define the three necessary conditions for an FDI to take place.¹

Figure 1 provides an overview of the various theoretical approaches to foreign direct investments (and MNE) which have characterised the period from when Hymer's thesis became known and the following dozen years. The figure is not complete, however, as it exclusively presents theories of an economic nature.

1. Although John Dunning's »eclectic paradigm« has remained at the centre of FDI theory until well into the twenty-first century, it continues to be vulnerable to criticism due to its eclectic approach. A number of scholars have argued that the *O*, *L* and *I* concepts are poorly defined and that the relationships between them are far from obvious. Itaka (1991) argues that the borderline between *O* and *I* is blurred since *O* advantages may be best understood as composed of historically accumulated acts of internalisation. Casson (1985) is close to finding *O* and *L* factors redundant, and strongly emphasises the internalisation aspect. Cantwell and Narula (2001) argue that globalisation increases the interactive dynamics among the three dimensions, thus increasing the complexity between *O* and *I* variables. They tend to see the eclectic framework as a tool to »choose between potentially competing theories« and as a vehicle for synthesising complementary theories.

Interestingly, some of the questions dealt with in the business literature on multinational enterprises and foreign direct investments have been picked up in international trade theory. Common to the two strands of theory is the distinction between horizontal and vertical integration and (transaction cost like) considerations on the choice of entry mode in a new location; Navaratti and Venables (2004); Markusen (2002); Antras et al. (2009). It will be evident from the following section that the integration of foreign direct investments into a broader theory of trade would have been much to Lund's taste.

Arne Lund's contribution

The article *Kapitalbevægelses-Teorien og de direkte Investeringer* [The Theory of Capital Movements and Direct Investments] by the Danish economist Arne Lund was published in *Nationaløkonomisk Tidsskrift* [Danish Journal of Economics] – then as now the leading economic journal in Denmark – in 1944.² The article was very much an isolated case, not only in Lund's scientific production³ but also in the history of the journal.⁴

2. Arne Lund graduated from the University of Copenhagen in 1940 and was immediately employed by *Industriraadet* (The Confederation of Danish Industry) as a business analyst. From 1946-1951 he was a research assistant to the leading Danish economist, Jørgen Pedersen, at the University of Aarhus but never settled as a scholar. Later, he worked as a business analyst with the European Council in Strassbourg, and from 1954 until his retirement he was employed by the Danish Federation of Employers, serving as its managing director from 1962-79. He later stated that the five years as a university teacher were the happiest in his life, but he was irresistibly attracted to »real life and all its conflicts«.

The 1944 publication was unique in Lund's production in the sense that it was path-breaking. His other publications did not approach this level. In 1940 he published an article in *Nationaløkonomisk Tidsskrift* based on his master thesis; the topic was French economic theory after World War I. It shows that he was well-informed on economic thinking of the nineteen-twenties and nineteen-thirties, including Austrian theory. He published a book in 1949 on Danish Manufacturing during the German occupation (1940-45), one on industrial calculation in 1943, and a considerable number of pamphlets related to his professional work.

Arne Lund seems to have been a rather complex personality with overwhelming self-confidence. As a student he was an organised Social Democrat with excellent relations to Jens Otto Krag, the later prime minister of Denmark. Nevertheless, he spent most of his working life serving organisations on the employers' side and became a Member of the Danish Parliament for the Conservative Party. After his retirement, he became a prolific columnist in the social-liberal Danish newspaper »Politiken«.

3. As pointed out by one referee, Lund reproduced his main points one year later. Spanning the period 13.12.1944 – 22.08.1945 the weekly *Finanstidende* (»Financial Times«) ran a series of 17 brief articles on »Vor Tids Nationaløkonomi« (Modern Economic Theory) authored by young and talented Danish economists. Lund contributed two articles on »Short-run international capital movements« and »Long-run international capital movements« on June 13 and June 20, 1945. Between them, the articles highlighted the theory of international capital movements before WWI and in the interwar period. The latter article briefly sketched main points in the 1944 article, but did not add to it. The 17 articles were put together in a booklet, *Vor Tids nationaløkonomi*, published by Finanstidendes Press in 1945.

4. The journal was started in 1876; we have checked the first 70 years. Lund's article was the first to deal with FDI, and even neighbouring topics such as Balance of Payments problems, industry location or International Capital Movements were dealt with only sporadically – in 1916, 1927 and 1937, respectively. In 1970, professor H. Winding-Petersen contributed a survey article on FDI to the journal with no reference to Lund (1944). Indeed there was little inspiration to be found in the national environment.

The main points

The article provides a coherent and systematic analysis of direct international investments. It takes its point of departure in perceived deficiencies in the received doctrine of international capital movements. Lund's aim was to remedy the theoretical shortcomings, and in the process he established a theory of foreign direct investments. In Lund (1944), they are defined in a way which reflects the author's penetrating understanding of business behaviour – an understanding which Lund may well have built during his four years of service to *Industriraadet* [the Confederation of Danish Industry]. It is noteworthy that the concept of entrepreneurship holds a central position. Entrepreneurship is seen as a »wandering combination« of (business) competences and (financial) capital. The interaction of the two sub-functions is analysed and their international mobility pointed out.

The article makes clear that the motives for direct investments are complex and multifaceted. In the process, the author develops an argument quite similar to that of Austrian economics: business opportunities are spotted and defined and then lead to action. This is in line with the earlier thinking of Menger and von Hayek and has been more coherently set forth in Kirzner (1973). As regards decisions on foreign direct investments, the author sees investment calculus and net present value of the project as a central instrument for business decisions.⁵

In a sense the article may be interpreted as an alternative version of the eclectic paradigm. The *why*, *how* and *where* are not simultaneous conditions but form a sequence. First, opportunities are spotted (the *why*); next, the wandering combination (entrepreneurship) is mobilised (the *how*); and, finally, investment calculus is used to decide on the particular opportunities to choose (the *where*). The choice of location follows from the decision process. Another aspect worth noticing is the obvious congruence with Hymer's argumentation. Based on a distinction between portfolio and direct investments, a business-oriented understanding of the drivers of foreign direct investments is developed in Lund's article, comprising motives as well as behaviour including adaptation to local business environments.⁶ The main purpose of the present article is

5. This, in fact, is a unique feature which distinguishes Lund's contribution from later theory. During the 1920-1940s micro and business theory was in focus in Danish economic journals, such as *Nationaløkonomisk Tidsskrift* and *Nordisk Tidsskrift for Teknisk Økonomi*. Brilliant authors such as I. Janzen and E. Schneider contributed. No traces, however, are found of NPV calculus, even if the technique was generally known, at least since Fisher 1907. It has been argued that the thinking may be traced back at least to Edmond Halley's »Of compound Interest«, (posthumously 1742). Halley is better known for discovering the comet named after him.

6. It is widely believed, also by scholars, that Hymer was the first one to make the distinction between direct and portfolio investments across borders. See, for instance, Ietto-Gillies (2004:63), which states: »Hymer's theory was path-breaking in both a backward and a forward looking way. In a backward looking way, because it developed a fully coherent approach and theory in a field where nothing existed, and indeed foreign direct investment was not even considered in the economic literature as an autonomous category in need of explanation«.

to relate Lund's publication to international business theory as it developed a few decades later, cf. the presentation above.

At the time when Lund wrote his article, the memory of the Ohlin-Keynes debate on the nature and effects of reparation payments was still fresh. Lund (1944) relates direct investments, as opposed to portfolio investments, specifically to this famous exchange of views. This aspect, however, is outside the scope of the present article and is not dealt with below.

The argument

The contents and argumentation of Lund (1944) are outlined below. For the sake of simplicity, they have been subdivided into eight items which closely match the structure of the article.

1. The basis of the article was the observation that the state of the theory of international capital movements, some of the leading contributors to which were Bertil Ohlin and Carl Iversen, was deficient in the sense that it did not recognise the different characters of direct investments and portfolio ones and that it had been formulated as if all movements of capital were of the portfolio type.^{7, 8} Lund was puzzled by the fact that, unlike what might be expected, active investments into business activities were not singled out although both the underlying motives and the mechanisms employed differed significantly from such characteristic of portfolio investments. One source of inspiration might have been Remer's distinction between portfolio investments and »investments in kind«, see the following section. So he set out to develop some minor corrections to the main stream thinking of the time – just as David Ricardo wanted to add a few corrections to Adam Smith's value theory and in the process revolutionised economic reasoning.⁹
2. Lund argued that international capital theory did in fact distinguish between different kinds of capital transfers, for instance between long-run and short-run move-

7. John Dunning (1988) makes the same observation. It is interesting that, while John Williams (1929) approached the topic from the theory of international trade, which, at the time, was still dominated by Ricardo and Mill, Lund's starting point in establishing an FDI theory was that of international capital movements.

8. It is most likely that Carl Iversen had taught Lund at the University of Copenhagen. Carl Iversen's doctoral dissertation was published and very positively reviewed by Bertil Ohlin in *Nationaløkonomisk Tidsskrift* in 1937. Ohlin's main contribution lay in his dispute with Keynes over the German reparation payments following World War I (Ohlin 1929). Arne Lund's article from 1940 surveys French contributions to economic theory from the interwar period. This article particularly emphasises the contributions of Aftalion and others in the field of monetary theory. As one referee remarked, in 1944 Iversen served as editor of *Nationaløkonomisk Tidsskrift* and in this capacity accepted the correctives of the young rebel.

9. Incidentally, this was the approach that John Williams (1929) applied to Ricardian theory.

ments and between real and absorbing movements. This led him to consider the motives for such movements. He argued that capital is heterogeneous and that the particular shape of the transfer depends on the motive. But motives were not dealt with in conventional theory »...*tacitly assuming the international bond based loans as the prototype of capital movements*«. ¹⁰ The type of loan tacitly assumed was one involving fixed interest rates and well-defined pay-back schedules which suited governmental and municipal construction projects perfectly. Other major infrastructure construction projects undertaken by, for example, transportation companies were equally well served by this type of loan. In such cases the cash-flow must be considered fairly stable and independent of the business cycle but, in the case of FDI, matters were different, he argued:

Direct investments (are) investments in foreign corporations or enterprises which are controlled by a person or a small group of persons domiciled in the United States or in the management of which such group or person has an important voice.

In Lund's view, the absence of a fixed interest rate differentiates a direct investment from a portfolio investment. Direct investments result in highly variable profit rates depending on business opportunities and the nature of the investment. The crucial difference between the two forms lies in the fact that direct investors acquire not only a share of the profits but also ownership rights and, consequently, influence on the management of the subsidiary. On this point Lund probably found inspiration in Liefman-Keil (1938) who debated various entry-modes a multinational enterprise could choose between when entering a new market.

At this point, Lund made what clearly looks like a Schumpeterian move and adopted the excellent term *entrepreneurial function*:¹¹ portfolio investments do not involve entrepreneurs, whereas direct investments are the work of entrepreneurs.¹² Thus Lund drew a constructive conclusion from his discussion of motives.

3. On this basis, Lund developed some further differences between direct and portfolio investments. First, the investor carries the risk of the arrangement and, consequently, no specialised or risk-distributing capital market is necessary for this

10. This quotation from Lund (1944) and the ones that follow were translated from the Danish original into English by Kurt Pedersen.

11. J. A. Schumpeter, *Capitalism, Socialism and Democracy* (1942) gives a broad discussion of entrepreneurship in a broader perspective, but the concept was introduced in *The Theory of Economic Development* (1934 (1911)). The entrepreneur is seen as the prime mover of economic activity.

12. At this point, Lund refers to professor C. F. Remer, who, according to Lund, first of all pointed to this characteristic feature of FDI, and who actually used the term entrepreneur investments.

kind of capital movement. Second, the initiative to invest originates in the home country and not in the host country. At this point Lund was now ready to offer his own definition of a foreign direct investment (p. 34):

A direct investment implies that a person, group of persons, a company or another legal person places disposable capital in a foreign business and at the same time acquires a decisive influence or at least a voice in the economic dispositions and profitability of the foreign business unit and thus in the use of the investment.

In this way, direct investments became another expression for the geographical dispersion of production. Having company activities in various locations bore witness to a single concept: the localisation of production conditioned by business opportunities and, in particular, by the mobility of factors of production. Of course, the geographical distribution of production might be viewed from a macro-economic standpoint or from a business one.

Lund (1944) emphasised that direct investments are based on a number of factors which, depending on the specific investment situation, may assume various degrees of importance. His ambition was, probably inspired by Conolly (1939), to »extract the underlying economic complex of motives«; in other words, to formulate a theory of foreign direct investments. The conventional theory of the time held that investments across borders were governed by differential interest rates and capital movements were assumed to serve the purpose of reducing such differences. In Lund's opinion, this was far from satisfactory, and, posing the (rhetorical) question whether the entrepreneurial function might be internationally mobile, he started examining FDI further.

4. In order to answer the question, he analysed, in more detail, the entrepreneurial function, which serves a double function: it assumes management based on talent and knowledge, and it assumes risk-taking. Lund argued that these two sub-functions are, by their very nature, extremely international. Establishing knowledge of a new business environment or a new industry is invariably time-consuming but, even so, entrepreneurs are far more mobile than the work force.

At this point, Lund presents his clinching argument: the two sub-functions cannot be dealt with independently because, by its very nature, a direct investment is *a wandering combination of capital and know-how or business ability in command of capital*.¹³ Lund even assumes that the combination is more mobile than any of the

13. The first of the definitions is credited to the Danish economist Warming and the second to Alfred Marshall. No further references are given by Lund.

two in isolation, arguing that *»resistance to two empty hands is more intense than resistance to a pioneer who brings the capital along.«* Similarly, there will be no friction in local capital markets if the entrepreneur brings capital along.

5. Lund then identified the driver behind direct international investments: *»Competence and saving capacity is unequally distributed over nations, and this causes a division of labour between creditor nations and debtor nations that are related by international capital markets.«*¹⁴ Lund was a keen observer of real world phenomena and pointed to some features that at the time tended to stimulate direct investments, namely, the combination of competence and capital. Most industries faced increasing firm sizes and concentration rates, and this shaped opportunities for corporate financing of direct investments. This was an independent reason that capital markets in the 1930s were suffering, even if the main reason might have been the economic-political developments.¹⁵ To Lund this was a signal that the wandering combination was in no way subject to conventional economic wisdom and that direct investments were indeed a very different type of investment. On pp. 41-42 he offers *»a plausible theory«* of foreign direct investments:

The technical and commercial development has caused entrepreneurs in industrialised countries to acquire power and monopoly advantages which can be capitalised and which therefore represent a value. This relates not only to monopolistic advantages based on the (increasing) concentration in industries but also to specialised know-how and technology. The tendency is most clearly observed when innovations are patented or acquire a trade mark, but idiosyncratic technical or commercial insight will, in its own right and even without protection, offer advantages when applied in a nation which is technically and commercially less developed. In such cases, the initial investment may be quite limited but the advantage guarantees an income which is accumulated and capitalised and thus contributes to financing the further expansion of the foreign business unit.

In other words, foreign companies may finance activities, through loans or emissions, in nations with well-developed capital markets. In less developed capital

14. This is exemplified by the creditor nation France and the debtor nation Denmark. While French farmers had a preference for early retirement and a peaceful life, Danish farmers combined competence with quite substantial indebtedness.

15. Authoritative theoreticians like Ohlin and Iversen regretted that corporate savings bypassed the general capital market. They saw it as a dysfunction that harmed the role of the market, i.e. equalisation between interest rates. Lund disagreed as he clearly interpreted this segment of the international capital market as having quite different functions.

markets, however, they often have to bring capital with them. This insight reflects a clear understanding of the diversity of capital and the segmentation of capital markets.¹⁶ In fact, the FDI segment is so different from other segments that a »modification« of international capital theory was necessary (p. 44):

From the above considerations on the supply of capital and competence it is obvious that the driving impulse of direct investments cannot be interest rate differentials. Capital export is initiated by an entrepreneur who remains intimately connected to his capital outlay, and the impulse or motive of the direct investment will be found in his profit expectations of the »debtor country«.

Although the received doctrine occasionally referred to supernormal expected profit, this was seen only as a possibility and as an exemption to the general theory of international capital movements. Ohlin and Iversen had observed that, in rare cases, capital would move towards nations with low interest rates. To Lund this was not an exception but, rather, an expectation. Or, more precisely, he argued that capital would move towards countries with a decreasing interest rate because a drop in the interest rate would have a beneficial effect on the investment calculus and, in consequence, on expected profit rates. Lund stated this explicitly, »Each of the two motives, interest rate differentials and profit expectations, is related to its particular field«.

6. Lund (1944) subsequently poses the question whether it is legitimate to hypothesise that direct investments are dependent on differences in profit expectations between countries. His answer is »no« – there is no way to conceive a uniform profit rate across nations or regions. It would be far too abstract as industries are »non-competing groups«. The mode of Lund's thoughts on »the modern theories of entrepreneurial profits« is distinctively Austrian (p. 45):¹⁷

In a society where economic quantities constantly change more or less speedily, the entrepreneurial function is that of anticipating. The circumstance that, due to his command of capital, the entrepreneur combines factors of production in a certain business has the consequence that, in a world characterised by imperfect divisibility in time and space, he has the opportunity to make the business profitable. Entre-

16. Also the 1930s witnessed a lack of risk-taking capital. In Great Britain, the MacMillan committee strongly recommended that British banks should support outgoing FDI rather than finance foreign governments and municipalities.

17. Hayek's spontaneous order springs to mind, but this thinking has most precisely been portrayed in Kirzner (1973).

preneurial profit, in turn, is on the one hand dynamic, i.e. a consequence of uncertainty about the future and the derived difference between prices over time and, on the other hand, a phenomenon of power in the most general sense, i.e. a consequence of the friction in price formation which gives an advantage due to large scale production, particular location, special knowledge or patents, financial relations, etc.

7. The article then addresses the »spatial dimension« of the investment calculus, in other words, the choice of location of business (units). In this respect, Lund (1944) emphasised the importance of national borders. Again, the thinking seems fairly contemporary to a modern reader. The author observes that regulations or other obstacles to trade play an important role and that currency problems do so as well. He argues, however, that the essential factor in FDI decisions relates to the quantity and prices of factors of production.¹⁸ It is a definite advantage of the presentation that investment calculus is explicitly included in the argument.

8. Towards the end of the article, the author introduces firm infrastructure and links it to the importance of foreign direct investments for the flows of goods. As a consequence of FDI, links between the horizontal and vertical structures of the corporation are changed. Lund observed that relative factor prices were important in determining the horizontal, i.e. geographical, scope. At the time, this was a highly innovative use of the Heckscher-Ohlin theory of international trade, highlighting the interplay of factor price differentials and internalisation. The observation is followed by some considerations on a theme which has assumed importance in international economics today, namely, the interrelationship between international trade and investments:

Direct investments exercise an immediate effect on international trade flows and the directions of their movements. If trade flows are blocked, this will increase the opportunities of factor movements – and particularly capital movements in the shape of direct investments. Entrepreneurs in the blocked export businesses locate plants behind foreign tariff walls or other trade restrictions in order to earn a share of the monopoly gains due to such trade barriers.

As regards location decisions, the author argues that not only does trade policy play a role but so does the broader concept of business policy, including competition law,

18. At this point, the influence from Bertil Ohlin becomes very obvious as Ohlin had founded the factor proportions theory of international trade in the previous decade.

trade mark law, company law, etc., and that, in some cases, free zones or corporate taxation laws give rise to location decisions focusing predominantly on »tax holidays«.

Inspirations

Arne Lund's article may be appreciated as a protest – even a revolt – against the received doctrine of his (excellent) teachers. After four years of work for the Confederation of Danish Industry, he was able to formulate a business-oriented critique of the theory of international capital movements. He obviously looked around for support and, in the article, he explicitly mentioned four sources of inspiration. Two were by the American economist C. F. Remer, who had travelled for more than two decades, mapping FDI in China. His voluminous book (1933) and an article in the 1936 *Festschrift* to Taussig had introduced the concept of »investment in kind«, suggesting that this differed from portfolio investments, and it had further employed the concept of entrepreneurs. Remer's primary motivation, however, lay in trade balance consequences.

A third source was that of Conolly (1939), who, on the basis of analyses of U.S. statistics on FDI, had suggested that distinctions should be made between investment motives. He had taken a broad view of FDI, including exploitation of monopoly positions, technology transfer, risk-taking, and profit transfers. Conolly had found economic reasons for a future role of FDI but viewed the political developments of the late 1930s as a threat to cross-border investments.

Finally, the fourth source was Liefmann-Keil (1938), who identified foreign direct investments as a new and important mode of international capital movements. Liefmann-Keil had examined various forms of FDI, including *Beteiligung*, i.e. cooperation between local and foreign business entities. The article dealt with relative factor prices, trade-marks and patents and – in a Dunning perspective – introduced both *O* and *L* variables. The author was close to employing transaction costs as decisive for the choice of operation. She also very nearly formulated a stage model of internationalisation.

It is reasonable to conclude that the four contributions probably provided Lund with inspiration. None of them, however, had the clear business orientation of Lund (1944) and only Liefmann-Keil's analyses and perspective bordered on FDI thinking in a modern sense.

Discussion and conclusions

Even though Arne Lund wrote his article more than 60 years ago, it still holds its own, in our view, and represents a fine piece of deductive economic thinking. It provides the reader with a *coherent understanding* of international direct investments, and

with an outline of investment decisions in a multinational company. In some respects it deals with problems that were taken up much later. The article applies economic reasoning, but the nucleus is business thinking. Its focus on entrepreneurship was not deeply original, but it was missing in international business literature until introduced by Buckley and Casson (1985).

A crucial point is Lund's penetrating understanding of competences and capital as an internationally wandering combination. The concept echoes the thinking of Marshall as well as Schumpeter and other Austrians. The »combination« implies an (*I-cum-O*) nexus in a Dunning framework. In Lund's view, both (capitalists') capital and (entrepreneurs') ideas are necessary conditions for explaining direct foreign investments; neither of them is sufficient, however. The foreign investment, then, consists in the transfer of the combination of entrepreneurial ideas and capital. This incidentally indicates the direction of FDIs, which move *from* countries rich in ideas (and in most cases capital).

The emphasis on entrepreneurial dynamics was strengthened by the introduction of relative factor prices in the choice among host countries. This observation later appeared in Vernon (1966) and in the literature on technology transfer. Finally, the emphasis on market power owing to superior skills plays an important role in Lund's argument – as it does in that of Hymer (1960); for example, reflections on (sustainable) profit differentials between industries point to later industrial economics, for example, Bain (1956), which was crucial in Hymer's book.

In terms of *perspective*, Lund's article differs from the OLI approach. Unlike what is the case in the eclectic paradigm, *O*, *L* and *I* are not presented as separate (sets of) variables with a certain degree of overlap. On the contrary, Lund's approach takes its starting point *inside* the firm. He identified the »soul« of the company in terms of entrepreneurship and developed the business argument against this backdrop: any competitive advantage was ultimately extracted from the wandering combination of competence and capital. In this sense, Lund places himself firmly in the tradition of Schumpeter. In line with the argumentation, the transfer is one of ideas-*cum*-capital for the purpose of reaping business opportunities as conceived by the entrepreneur.

Arguably, the article links up with Edith Penrose (1959) as her work on growth and growth processes of the firm centred around the entrepreneur and limited organisational resources. Even if she was less specific in terms of the international perspective, the reasoning is a similar one. Crucial to Penrose as well as Lund are the possibilities for business development: the wandering combination pervades both publications. Where the eclectic framework is of a static nature, Lund seems to suggest a dynamic one. This ultimately reflects the fact that his main interest is in flow concepts and not in stocks.

The forward link to modern theories on entrepreneurship is obvious, and the backward link to the Austrians is equally clear. Menger, von Hayek and Schumpeter – who was probably less than half Austrian – saw market processes under uncertainty as the true sense of business. In recent years, Kirzner (1973) has reflected similar conceptions.

Methodologically the article represents a classic Ricardian approach to economic analysis, much like John Williams (1929). Like Hymer (1960), it takes as its starting point an anomaly in received theory (which treated FDI as if they were portfolio investments) and investigates it by focusing on the narrowest set of variables. The new contribution made by Hymer, as well as Lund, was that, in the case of FDI, the entrepreneur maintains managerial control and interest rate differentials are unable to explain investments. With its focus on entrepreneurship and uncertainty and its insight into the foundations of competitive advantages, the article is simple and straightforward.

An exciting detail is the application of the investment calculus (which does not, of course, require certainty about future prices or quantities). This allows for a ranking of FDI projects and, implicitly, the sequence of the location of business activities, given scarce capital and management resources to decide which ones to pursue. *Necessary* conditions of FDI are supplied by a criterion which offers the sufficient condition in terms of, for example, the internal rate of interest or net present value (NPV). Even if rudimentary, Lund offers a model for decision making in relation to international direct investments – the decision maker being the entrepreneur who at the same time spots business opportunities and carries the risk.¹⁹

Lund (1944) exudes *business relevance*. The essential problem in (big) business that the article identifies is the configuration of business units in an optimal pattern, given relative factor prices across nations. In fact, the article appears as relevant today as it was 60 years ago. The right configuration of business units is a long step towards identifying the business model of a company, which, in turn, reminds us of Porter (1986), who conceived of the value system as a combination of geographic configuration and managerial coordination.

The international business perspective

Rather than to argue that Lund is the father of any perspective on IB in particular, the ambition of the present article is to present the thinking, from an IB perspective, of

19. In fact, the rather narrow coupling of the two sub-functions is not in line with Austrian thinking which distinctly made a dividing line between the entrepreneur function and other functions, including finance, see, for example, Kirzner (1973).

an innovative, early thinker whose work was published at a point in time when IB had not yet been constituted as an academic field in its own right. Links to later theories are quite obvious, however. Lund (1944) can most clearly be seen as a precursor to the work of Stephen Hymer: they share both the crucial distinction between portfolio and direct investments and the presence of industrial economics considerations. While Hymer introduced aspects of Coasean thinking in later works, entrepreneurial aspects were essential in Lund's article. One notable difference is the absence of political attitudes from Lund's article.

The entrepreneurial aspect of Lund (1944) links it to Buckley and Casson (1985), as well as to Austrian theory in general. Like the one-third Austrian Joseph Schumpeter, he saw the entrepreneur (»captain of industry«) as a key figure and decisions under uncertainty as his function in business. Bridging FSA, a concept clearly comprehended in the article, and market opportunities across national borders, Lund created a framework for business dynamics which was handled by the picturesque concept of the wandering (OI) combination.

The article touches sporadically on the positioning school of Michael Porter, partly by industrial economics fragments and partly by presenting the L factor as a set of opportunities which could be exploited by way of NPV calculus. This would at least implicitly determine an MNE structure in terms of coordination and (particularly) configuration. Foreign direct investment and trade were seen in a broader perspective, even if, in contrast to Porter, Lund emphasised the investment side rather than the organisational. But he undoubtedly can be seen as an early contributor to the new trade-com-investment theory as developed by Venables, Markusen and others.

The Lund article seems to have little in common with the resource based view, although the concept of firm-specific advantages, of course, carries a certain resemblance with assets or resources (*O* factors). This takes us to Lund and the eclectic paradigm. *O*, *L* and *I* were, clearly, all present in the article, but the perspective remained firmly rooted at company level. In consequence, Lund's article is superior to later theories in two important ways: first, he added causality in the sense that $(OI) \rightarrow L$ after comparison of projects, and, second, in applying the NPV calculus, he introduced the possibility of choice between host countries as a rational business procedure.

In our opinion, *Kapitalbevægelses-Teorien og de direkte Investeringer*, i.e. Lund (1944), is a decent and sober piece of economic reasoning. Had it appeared in another language, and perhaps at a slightly different point in time, it might well have become the starting point of an qualified FDI theorising a quarter of a century before Hymer.

Literature

- Antras, P. et al. 2009. Multinational Firms, FDI Flows, and Imperfect Capital Markets. *The Quarterly Journal of Economics*, vol. 124, issue 3, pp. 1171-1219.
- Bain, J. S. 1956. *Barriers to New Competition*. Harvard University Press.
- Barney, J. 1984. Firm resources and sustained competitive advantage. *Journal of Management*, 17, pp. 99-120.
- Bartlett, C. A. and S. Ghoshal. 1988. *Managing across Borders. The Transnational Solution*. Harvard Business School Press.
- Buckley, P. and M. Casson. 1985. *The Economic Theory of the Multinational Enterprise*. MacMillan.
- Cantwell, J. and R. Narula. 2001. The Eclectic Paradigm in the Global Economy. *International Journal of the Economics of Business* 8, no. 2, pp. 155-72.
- Casson, M. 1987. *The Firm and the Market*. Oxford, MIT Press.
- Coase, R. 1937. The Nature of the Firm. *Economica*, 4, pp. 386-405.
- Conolly, F. G. 1939. Die Bedeutung der Kapitalbeteiligung an ausländischen Unternehmungen. *Weltwirtschaftliches Archiv*, Bd. XLIX, pp. 213-40.
- Dunning, J. H. and S. M. Lundan (2008). *Multinational Enterprises and the Global Economy*. Edward Elgar.
- Dunning, J. H. 2000. The eclectic paradigm as an envelope for economic and business theories of MNE activity. *International Business Review*, 9, pp. 163-90.
- Dunning, J. H. 1988. *Explaining International Production*. Unwin Hyman.
- Dunning, J. H. 1977. Trade, location of economic activity and the multinational enterprise: a search for an eclectic approach. In: B. Ohlin, P. O. Hesselborn and P. J. Wijkman (eds.). *The International Allocation of Economic Activity*. London, McMillan pp. 398-418.
- Fisher, I. 1907. *The Rate of Interest. Its Nature, Determination and Relation to Economic Phenomena*. McMillan, N.Y.
- Forsgren, M. 2008. *Theories of the Multinational Firm*. Edward Elgar.
- Hayek, F. A. von. 1937. Economics and Knowledge. *Economica*, IV (new series), pp. 33-54.
- Hymers, S. H. 1960/1976. *The International Operations of National Firms: A Study of Direct Foreign Investment*. PhD dissertation, published posthumously. The MIT Press 1976.
- Itaki, M. 1991. A critical assessment of the eclectic theory of the multinational enterprise. *Journal of International Business Studies*. II
- Kirzner, I. 1973. *Competition and Entrepreneurship*. Chicago University Press.
- Iversen, C. 1935. *Aspects of the Theory of International Capital Movements*. Levin and Munksgaard.
- Liefmann-Keil, E. 1938. Die direkte Auslandsinvestierung. *Jahrbücher für Nationalökonomie und Statistik*, Nov., pp. 525-55.
- Lund, A. 1944. Kapitalbevægelses-Teorien og de direkte Investeringer. *Nationaløkonomisk Tidsskrift*, pp. 28-57.
- Madhok, A. and A. Phene. 2001. The Co-evolutionary Advantage: Strategic Management Theory and the Eclectic Paradigm. *International Journal of the Economics of Business*, 8, no. 2, pp. 243-56.
- Markusen, J. 2002. *Multinational firms and the theory of international trade*. MIT Press.
- Navaretti, G. B. and A. J. Venables. 2004. *Multinational Firms in the World Economy*. Princeton University Press.
- Nohria, N. and S. Ghoshal. 1997. *The Differentiated Network. Organizing Multinational Corporations for Value Creation*. Jossey-Bass publishers.
- Ohlin, B. 1929. The Reparation Problem: a discussion. *Economic Journal*, 39, June, pp. 172-83.
- Penrose, E. 1959. *The Theory of the Growth of the Firm*. Basil Blackwell.
- Porter. 1980. *Competitive Strategy*. The Free Press.
- Porter, M. 1985. *Competitive Advantage*. The Free Press.

- Porter, M. 1986. *Competition in Global Industries*, Ch. 2, Harvard.
- Remer, C. F. 1933. *Foreign Investments in China*. McGraw-Hill.
- Remer, C. F. 1938. Investments in Kind. Chapter IX in *Explorations in Economics. In Honor of F. W. Taussig*. McGraw-Hill.
- Schumpeter, J. A. 1934. 1912. *The Theory of Economic Development*. Harvard University Press.
- Schumpeter, J. A. 1942. *Capitalism, Socialism and Democracy*. George Allen and Unwin.
- Strandskov, J. and K. Pedersen. 2008. The foreign expansion of a service company: The case of ISS A/S. *Business history*, 50, 1, pp. 40-61.
- Tolentino, P. E. 2001. From a Theory to a Paradigm: Examining the Eclectic Paradigm as a Framework in International Economics. *International Journal of the Economics of Business*, 8, no. 2, pp. 191-209.
- Vernon, R. 1966. International Investment and International Trade in the Product Cycle. *Quarterly Journal of Economics*, 80, pp. 190-207.
- Wernerfelt, B. 1984. A Resource-based View of the Firm. *Strategic Management Journal*, 5, pp. 171-86.
- Williams, John (1929). The Theory of International Trade Reconsidered. *The Economic Journal*, 39, June, pp. 195-209.
- Williamson, O. 1975. *Markets and Hierarchies*. Boston: The Free Press.
- Williamson, O. 1985. *The Economic Institutions of Capitalism: Firms, Markets, Relational Contracts*. The Free Press.

The theory of optimal taxation: New developments and policy relevance

Peter Birch Sørensen

Danmarks Nationalbank, E-mail: pebs@nationalbanken.dk

SUMMARY: The theory of optimal taxation has often been criticized for being of little practical policy relevance, due to a lack of robust theoretical results. This paper argues that recent advances in optimal tax theory have made that theory easier to apply and may help to explain some current trends in international tax policy. Covering the taxation of labour income and capital income as well as indirect taxation, the paper also illustrates how some of the key results in optimal tax theory may be derived in a simple, heuristic manner.

In the tradition established by the classical political economists, normative analysis of tax policy tended to follow a principles-oriented approach according to which a good tax system should satisfy certain desirable criteria. For example, Lord Overstone, who served as President of Britain's Royal Statistical Society from 1851 to 1853, thought that a tax should be »productive, computable, divisible, frugal, non-interferent, unannoyant, equal, popular, and uncorruptive«, see the discussion by O'Brien (2009).

The classical economists rarely discussed the trade-offs between the various goals of tax policy. In particular, they did not pay much attention to the trade-off between redistribution and economic efficiency, since they typically ruled out redistributive progressive taxation as a matter of principle, seeing it as a fundamental threat to property rights. The denouncement of any deviation from proportional taxation was vividly expressed by McCulloch (1845) who argued that »The moment you abandon the cardinal principle of exacting from all individuals the same proportion of their income or of their property, you are at sea without rudder or compass, and there is no amount of injustice and folly you may not commit«, see Creedy (2009, p. 2).

This paper is a revised and extended version of an invited plenary lecture given at the 20th Scientific Meeting of the Società italiana di economia pubblica (SIEP) in Pavia on 25-26 September, 2008. Without implicating her in any remaining shortcomings, I thank my discussant Lisa Grazzini for her comments on my lecture. I am also grateful to Claus Thustrup Kreiner, Etienne Lehmann and an anonymous referee for valuable comments on an earlier draft. The views expressed in this paper do not necessarily reflect those of Danmarks Nationalbank.

Following the neoclassical »marginalist« revolution in economic theory, Edgeworth (1897) argued that taxation should involve an equal marginal sacrifice (of utility) for each individual taxpayer in order to minimise the aggregate utility loss imposed on taxpayers. When combined with the neoclassical assumption of declining marginal utility of income, this utilitarian principle of equal marginal (i.e. minimum total) sacrifice did provide a rationale for progressive income taxation.

Edgeworth was aware that redistributive taxation involves a trade-off between equity and efficiency, but the development of a rigorous coherent framework for analysing this trade-off had to await the seminal work by Mirrlees (1971). However, much of the optimal tax literature building on Mirrlees' contribution has been highly technical and abstract, and for many years this body of theory seemed to offer few robust results. For these reasons many policy makers have tended to see the theory of optimal taxation as being of little practical relevance. In this paper I shall argue that recent advances in optimal tax theory have made that theory easier to apply and may help to explain some current trends in international tax policy. I shall also illustrate how some of the key results in optimal tax theory may be derived in a simple, heuristic manner.

The first part of the paper deals with the theory of optimal taxation of labour income. In Part 2 I focus on optimal indirect taxation, while Part 3 discusses the optimal taxation of income from capital. Part 4 summarises the main conclusions.

1. Optimal taxation of labour income

The Mirrlees model

In the canonical model of optimal income taxation set up by Mirrlees (1971) consumers are assumed to maximise a utility function of the general form

$$U = U(C, L), \tag{1}$$

subject to the budget constraint

$$C = wL - T(wL), \tag{2}$$

where C is consumption, L is labour supply, w is the real wage, and $T(wL)$ is a non-linear tax-transfer schedule. The solution to the consumer's problem yields his indirect utility function $V(w)$. In the Mirrlees model the pre-tax real wage rates are treated as exogenous and taken to reflect the different non-observable ability levels of individual taxpayers. With wage rates being distributed over the interval $[w, \bar{w}]$, $0 \leq w \leq \bar{w} \leq \infty$, Mirrlees assumed that the benevolent policy maker wishes to maximise an individualistic Bergson-Samuelson social welfare function of the form

$$W = \int_{\underline{w}}^{\bar{w}} \psi (V (w)) f (w) dw, \quad \psi' > 0, \quad \psi'' \leq 0, \quad (3)$$

Here $f(w)$ indicates the density of taxpayers earning the wage rate w and the (numerical) magnitude of the second derivative ψ'' reflects the strength of the policy maker's preference for equity. The maximisation of (3) takes place subject to the constraint that the government must raise an exogenous amount of revenue R :

$$\int_{\underline{w}}^{\bar{w}} T (wL (w)) f (w) dw = R. \quad (4)$$

The solution to the above optimal tax problem is technically demanding and does not yield very clear-cut results regarding the shape of the optimal income tax schedule. To get a feel for the likely shape of this schedule, Mirrlees carried out simulations assuming Cobb-Douglas utility functions, a classical utilitarian social welfare function (with $\psi' = 1$ and $\psi'' = 0$) and a log-normal wage distribution. On these assumptions he found that the optimal tax schedule was approximately linear, with an exemption level below which positive net transfers are payable.

Had this early result been robust, it would have had great practical policy relevance, since a linear labour income tax is fairly simple to administer. In particular, because a linear income tax features a constant marginal tax rate, it does not require information on individual incomes, since it can be implemented as a proportional payroll tax combined with a flat transfer to all taxpayers. However, subsequent work by Tuomala (1984) and others revealed that the near-optimality of a linear income tax is not a robust result once one allows for plausible respecifications of utility functions and of the shape of the wage distribution. Atkinson and Stiglitz (1980, ch. 13) also found that the optimal tax schedule deviates substantially from linearity when the social planner has more egalitarian preferences than those implied by classical utilitarianism.

With these discouraging findings, it seemed for a while that optimal tax theory could offer little guidance on income tax design. But building on earlier contributions by Revesz (1989), Piketty (1997), Diamond (1998), and Roberts (2000), Saez (2001, 2002a) showed how a formula for the optimal marginal tax rate at every income level can be derived in terms of the relevant elasticities of taxable income and the properties of the wage distribution. Since these parameters can in principle be observed or estimated empirically, the work of Saez has greatly enhanced the practical applicability of optimal income tax theory. Another important contribution by Saez (2002a) was the

explicit allowance for tax distortions to the extensive as well as the intensive margin of labour supply. This is highly relevant since many empirical studies have shown that labour supply is often more elastic on the extensive margin (where workers decide whether or not to participate in the labour market) than on the intensive margin (where people make decisions on their hours of work, given that they are already employed).

The Saez (2002a) model

Given its importance, it is worth restating the key contribution of Saez (2002a) in a heuristic, intuitive manner.¹

Following Piketty (1997), Saez assumes that employed workers allocate themselves across a range of different occupations involving different levels of effort and income. Workers can decide to participate or to stay outside the labour market. In the latter case they receive the public transfer S_0 , whereas a worker employed in job category i earns the after-tax income c_i . As the net gain $c_i - S_0$ from employment in occupation i increases, more workers move from non-employment into this occupation. This is the extensive labour supply response. Once employed, workers can move one step up or down the job ladder by adjusting their effort. In case of an increase in the gap $c_{i+1} - c_i$ between the net incomes obtainable in job categories $i + 1$ and i , respectively, some workers previously employed in occupation i will therefore be induced to upgrade themselves to the higher-earning occupation $i + 1$ by increasing their »effort«. Similarly, an increase in the net earnings differential $c_i - c_{i-1}$ will spur some workers to upgrade themselves from job category $i - 1$ to category i . These movements between job categories involving different earnings levels represent the intensive labour supply response. With these assumptions, the fraction h_i of the work force employed in occupation i can be specified as

$$h_i = h_i(c_i - S_0, c_{i+1} - c_i, c_i - c_{i-1}), \quad (5)$$

where $c_i \equiv z_i - T_i(z_i)$ is the difference between the pre-tax earnings z_i in occupation i and the net tax $T_i(z_i)$ payable on that income. By assuming that labour supply decisions depend only on the *differences* between the net incomes obtainable in different labour market states, equation (5) implicitly abstracts from income effects (since otherwise labour supply would also depend on the *level* of income). In recent years most empirical studies have indeed tended to find rather small income effects on individual labour supply. Thus there is a good case for considering the benchmark situation with zero income effects.

1. The present heuristic exposition follows the »perturbation method« adopted in the main text of Saez' article but is slightly more elaborate. The rigorous formal proofs of Saez's results may be found in the appendix to his paper.

From (5) we can define the *participation elasticity*

$$\eta_i \equiv \frac{\partial h_i}{\partial(c_i - S_0)} \cdot \frac{c_i - S_0}{h_i}, \quad (6)$$

and the *intensive labour supply elasticity*

$$s_i \equiv \frac{\partial h_i}{\partial(c_i - c_{i-1})} \cdot \frac{c_i - c_{i-1}}{h_i}. \quad (7)$$

As shown by Saez (2002a, p. 1070), $s_i \equiv \frac{\varepsilon_i - z_i}{z_i - z_{i-1}}$, where $\varepsilon_i \equiv \frac{\partial z_i}{\partial(1-T'(z_i))} \cdot \frac{(1-T'(z_i))}{z_i}$ is the elasticity of taxable income with respect to one minus the marginal tax rate, estimated in numerous recent empirical studies.

Optimal taxation in the Saez model

Consider now the welfare effects of increasing the tax liability by the (small) amount dT for all of the occupations $i, i + 1, i + 2, \dots$ up to the highest-earning occupation J . Measured in money metric units, the impact on the welfare of an individual in one of these occupations equals the direct impact on disposable income, $-dT$, since any change in disposable income stemming from a change in labour supply has no first-order effect on welfare, assuming that the taxpayer has optimised his/her labour supply in the initial equilibrium (this is just an application of the Envelope Theorem). Suppose now that the policy maker's evaluation of the marginal social utility of net income for individuals in occupation j is $g(c_j)$. The social evaluation of the impact of the tax increase on aggregate private welfare is then given by

$$dW(i) = - \sum_{j=i}^J g(c_j) h_j dT. \quad (8)$$

Against this private welfare loss one must balance the rise in public revenue. Abstracting from behavioural changes, the direct («mechanical») revenue gain is

$$dM = \sum_{j=i}^J h_j dT. \quad (9)$$

However, public revenue also changes because of the tax-induced changes in labour supply. The tax hike reduces the net income differential $c_i - c_{i-1} \equiv z_i - T_i - (z_{i-1} - T_{i-1})$ by the amount dT , whereas it has no impact on the income differential $c_j - c_{j+1}$ be-

tween any other two »neighbouring« occupations. With dh_i^I denoting the change in h_i resulting from the *intensive* labour supply response, the revenue loss implied by the change in behaviour on the intensive margin (dB^I) is

$$dB^I = dh_i^I \cdot (T_i - T_{i-1}) = -s_i \cdot \frac{h_i dT(T_i - T_{i-1})}{c_i - c_{i-1}} \quad (10)$$

where the last equality in (10) follows from (7). The tax hike also reduces the net income gain $c_j - S_0$ by the amount dT for all persons in job category i and above. If dh_j^E indicates the change in h_j stemming from labour supply responses on the *extensive* margin, the drop in revenue occurring as a result of lower labour force participation becomes

$$dB^E = \sum_{j=i}^J dh_j^E \cdot (T_j + S_0) = -dT \sum_{j=i}^J \frac{\eta_j h_j (T_j + S_0)}{c_j - S_0}, \quad (11)$$

where we have used the definition (6) to arrive at the last equality. Under the optimal tax policy, the tax burden on any occupation from i and upwards is increased up to the point where the social valuation of the resulting private utility loss is just matched by the government's net revenue gain, i.e. until

$$dW(i) + dM + dB^I + dB^E = 0. \quad (12)$$

Inserting (8) through (11) into (12), we get the optimal tax rule

$$\frac{T_i - T_{i-1}}{c_i - c_{i-1}} = \frac{1}{s_i h_i} \sum_{j=i}^J h_j \left[1 - g_j - \eta_j \left(\frac{T_j + S_0}{c_j - S_0} \right) \right]. \quad (13)$$

The marginal tax rate at the income level z_i may be defined as

$$m_i \equiv \frac{T_i - T_{i-1}}{z_i - z_{i-1}}, \quad (14)$$

from which it follows that

$$1 - m_i \equiv \frac{(z_i - T_i) - (z_{i-1} - T_{i-1})}{z_i - z_{i-1}} = \frac{c_i - c_{i-1}}{z_i - z_{i-1}} \Rightarrow \frac{m_i}{1 - m_i} = \frac{T_i - T_{i-1}}{c_i - c_{i-1}}. \quad (15)$$

Further, the »participation tax rate« measuring the increase in net taxes imposed when a person moves from non-employment to employment may be defined as

$$t_j \equiv \frac{T_j + S_0}{z_j}, \quad (16)$$

implying that

$$\frac{T_j + S_0}{c_j - S_0} \equiv \frac{T_j + S_0}{z_j - (T_j + S_0)} = \frac{t_j}{1 - t_j}. \quad (17)$$

Substituting (15) and (17) into (13), we may write the optimal tax rule as

$$\frac{m_i}{1 - m_i} = \frac{1}{s_i h_i} \sum_{j=i}^J h_j \left[1 - g_j - \eta_j \left(\frac{t_j}{1 - t_j} \right) \right]. \quad (18)$$

Implications of the optimal tax rule

Equation (18) has a number of implications for tax policy: (1) The optimal marginal tax rate at the income level z_i is lower the higher the intensive labour supply elasticity (s_i) and the larger the number of taxpayers (h_i) at that income level. This is intuitive, since the efficiency loss from a rise in the marginal tax rate will be greater the more taxpayers who are affected by it and the stronger their labour supply responds to a change in the net gain from additional effort. (2) Since a rise in the marginal tax rate at income level z_i reduces the net labour income of all taxpayers above that earnings level, it induces some of them to exit the labour market. The strength of this extensive labour supply response is larger the higher the participation elasticities η_j , $j = i, i + 1, \dots, J$, and the resulting loss in net public revenue is greater the higher are the initial participation tax rates t_j of the affected groups and the greater the number of people in these groups. The labour supply response at the extensive margin therefore reduces the optimal marginal tax rate at income level z_i to a larger extent the higher are the values of η_j , t_j and h_j above that income level. (3) Because a higher marginal tax rate at income level z_i cuts into the disposable income of all taxpayers above that level, the optimal marginal tax rate is lower the higher is the social valuation of income for taxpayers above the income level considered; i.e. the larger the values of the welfare weights g_j , $j = i, i + 1, \dots, J$, and the greater the number of people carrying these weights (h_j).

While the factors mentioned in (1) and (2) reflect how concerns about economic efficiency shape the optimal tax schedule, the parameters in (3) obviously reflect equity concerns. However, note from (18) that allowing for labour supply responses at the extensive margin (an efficiency concern) is equivalent to attaching a higher social welfare

weight to groups with high participation elasticities and/or groups with high participation tax rates.

One further point is worth emphasizing: to apply formula (18), no assumptions about individual preferences are needed. The policy maker »only« has to specify his/her distributional value judgements in terms of the social welfare weights g_j and to obtain estimates of the parameters s_j , η_j , t_j , and h_j .²

To illustrate the applicability of formula (18), we may consider some instructive special cases. As already mentioned, the analysis above abstracts from income effects on labour supply since such effects have typically been estimated to be small. With zero income effects a lump-sum transfer to an individual taxpayer has no impact on his/her labour supply and hence does not generate any indirect change in public revenue via this channel. A marginal euro of public funds will then be valued exactly as much as an additional euro distributed evenly across all taxpayers, implying that

$$\sum_{i=0}^J h_i g_i = 1. \quad (19)$$

Traditionally the optimal tax literature has abstracted from labour supply responses on the extensive margin, implicitly assuming $\eta_j = 0$. It then follows from (13) that the optimal marginal tax rate at the bottom of the income ladder is

$$\frac{T_1 - T_0}{c_1 - c_0} = \frac{1}{s_1 h_1} \sum_{j=1}^J h_j (1 - g_j) = \frac{1}{s_1 h_1} \left(\sum_{j=0}^J h_j - h_0 - \sum_{j=0}^J h_j g_j + h_0 g_0 \right), \quad (20)$$

where individuals outside the labour market have been categorized as group zero (indicated by subscript 0), while the lowest-paid workers are categorized as group 1. By definition, we have $\sum_{j=0}^J h_j = 1$ which may be inserted into (20) along with (19) to give

$$\frac{T_1 - T_0}{c_1 - c_0} = \frac{(g_0 - 1)h_0}{s_1 h_1}. \quad (21)$$

If policy makers have a strong preference for redistribution, the value of the social welfare weight g_0 will tend to be far greater than one, since the weighted average value of g_j across all taxpayers is unity (cf. (19)).³ According to (21) the marginal tax rate at

2. One difficulty is that estimation of the elasticities s_j and η_j has to be based on the individual labour supplies actually observed whereas formula (18) requires knowledge of the elasticities that would prevail under the *optimal* tax system where labour supply might differ from its current level.

3. If policy makers believe that people outside the labour market are unemployed mostly because they are »lazy«, it is conceivable that the welfare weight g_0 could be smaller than one, but this possibility is disregarded here.

the bottom of the earnings distribution will then be high. This is a well-known implication of the Mirrleesian model of optimal income taxation when the least productive individuals are outside the labour force. In other words, the much-debated »poverty traps« caused by the phase-out of social transfers as people belonging to the poorest section of the population move from unemployment into employment are in fact part of an optimal policy when the labour supply of the rest of the work force does not respond at the extensive margin.

However, suppose instead that labour supply responds only at the extensive margin, i.e. $s_i = 0, \forall i$. This alternative benchmark case is of some interest since empirical studies tend to find that labour supply is indeed much more elastic at the extensive than at the intensive margin, at least for the low-skilled and for females, see Heckmann, (1993). With $s_i = 0$ it follows from (18) that optimal tax policy requires

$$\frac{t_j}{1-t_j} = \frac{1-g_j}{\eta_j}. \quad (22)$$

Thus the optimal participation tax rate for income group j is lower the higher the participation elasticity of that group (efficiency concern) and the higher the social valuation of income for members of the group (equity concern). Since $S_0 \equiv T_0$, we have

$$t_1 \equiv \frac{T_1 + S_0}{z_1} \equiv \frac{T_1 - T_0}{z_1} \Rightarrow \frac{t_1}{1-t_1} = \frac{T_1 - T_0}{z_1 - (T_1 - T_0)} = \frac{T_1 - T_0}{c_1 - c_0} \quad (23)$$

According to (22) and (23) the optimal marginal tax rate at the bottom is therefore given by

$$\frac{T_1 - T_0}{c_1 - c_0} = \frac{1-g_1}{\eta_1} \quad (24)$$

Given that the weighted average value of the social welfare weights is unity (see (19)), a policy preference for redistribution will almost surely imply $g_1 > 1$ except in the extreme Rawlsian case where policy makers care only about the poorest group (so that $g_0 = 1$ and $g_j = 0$ for all $j \geq 1$). Thus (24) implies that, when labour supply only responds at the extensive margin, the marginal tax rate on the lowest-paid workers should generally be *negative*. This policy could be implemented by granting a sufficiently large Earned Income Tax Credit which is phased out with rising levels of labour income (to reflect the fact that g_j varies negatively with income). The result in (24) is in stark contrast to the more conventional result reported in (21), and it highlights the importance of allowing for labour supply responses at the extensive margin.

Saez (2002a) applies the general formula (18) to simulate the optimal tax schedule, using data on the U.S. wage distribution plus alternative assumptions about labour supply elasticities and the government's tastes for redistribution. The latter are specified as

$$g_j = g(c_j) = \frac{1}{p \cdot c_j^\nu}, \quad 0 \leq \nu \leq +\infty, \quad (25)$$

where the parameter ν measures the strength of the preference for redistribution, and p is the marginal value of public funds, calibrated to satisfy (19). For plausible values of the intensive labour supply elasticities, Saez finds that it takes fairly high participation elasticities to rationalize negative marginal tax rates at the bottom, especially if the preference for redistribution is strong. However, with realistic participation elasticities, the lowest-paid workers should face rather low marginal tax rates in order not to discourage their participation, and this could still provide a role for some form of Earned Income Tax Credit (EITC). The new focus in optimal tax theory on the importance of the extensive margin of labour supply thus offers a rationale for the recent trend in many OECD countries towards the introduction of various in-work benefits (such as an EITC) that are intended to »make work pay«.

An alternative application of optimal tax theory: deriving implicit social welfare weights

One potential obstacle to the applicability of optimal tax theory is that policy makers may not be able or willing to explicitly specify the social valuation weights g_j in the optimal tax formula (18). However, by »turning the formula on its head«, the researcher can ask: what are the magnitudes of the social welfare weights for the various income groups that would make the existing tax system optimal, given realistic assumptions on labour supply elasticities and the distribution of pre-tax earnings? Having estimated the implicit social welfare weights embodied in the current tax-transfer system, the researcher may then confront policy makers and ask them: do these social welfare weights provide a reasonable representation of your actual distributional preferences? If the answer is negative, perhaps policy makers can be induced to reconsider whether the existing tax-transfer system represents a rational trade-off between equity and efficiency. For example, suppose the researcher could point out that the social welfare weights implied by the current tax system are not monotonically decreasing with the taxpayer's level of income. Presumably the policy maker would find it hard to defend such a system and would therefore be willing to consider proposals for reforming it.

To illustrate how such a procedure might work, note that equation (13) may be inverted to give (using $S_0 \equiv -T_0 = c_0$):

$$g_J = 1 - \eta_J \frac{T_J - T_0}{c_J - c_0} - \varsigma_J \frac{T_J - T_{J-1}}{c_J - c_{J-1}}, \quad (26)$$

$$g_i = 1 - \eta_i \frac{T_i - T_0}{c_i - c_0} - \varsigma_i \frac{T_i - T_{i-1}}{c_i - c_{i-1}} + \frac{1}{h_i} \sum_{j=i+1}^J h_j \left[1 - g_j - \eta_j \left(\frac{T_j - T_0}{c_j - c_0} \right) \right]. \quad (27)$$

Equation (26) may be used to estimate the implicit social welfare weight for the top income group (group J), given estimates of the labour supply elasticities η_J and ς_J plus data on the net tax payments T_J , T_{J-1} and T_0 and the disposable incomes c_J , c_{J-1} and c_0 . Once this has been done, one can apply (27) to calculate the implicit welfare weights for groups $J-1$, $J-2$, ... etc. in a recursive manner, all the way down to group 1, using data on incomes and tax payments plus estimated elasticities. The welfare weight g_0 may then finally be calculated from (19).

Such an exercise has recently been undertaken by Spadaro (2008) for seven western European welfare states. Spadaro uses data from the tax-benefit calculator EUROMOD which groups the population into ten income deciles. Interestingly, he finds that the implicit social welfare weights are non-monotonic in all the countries considered.⁴

Optimal taxation in imperfect labour markets

The bulk of the literature on optimal income taxation abstracts from labour market distortions other than those caused by the tax system. In competitive labour markets, a switch from proportional to progressive income taxation increases the deadweight loss from taxation, except under highly implausible assumptions regarding the income effects of taxation, see Sandmo (1983). But once one allows for non-tax distortions due to labour market imperfections, some amount of tax progressivity can actually be defended on pure efficiency grounds, see Sørensen (1999). For example, in unionised labour markets union wage setting tends to generate involuntary unemployment. If the government raises the marginal tax rate while holding the average tax rate constant (e.g. by raising the personal exemption level), unions can be motivated to moderate

4. Implicit social welfare weights may also be derived from the condition that the marginal social cost of raising an additional euro of tax revenue should be equalized across all income groups under the optimal tax system. Applying this method to Danish data, Petersen (2007) likewise finds that the Danish tax-transfer system implies non-monotonic social welfare weights, assigning higher welfare weights to the middle income groups than to the poorest groups.

their wage claims, thus paving the way for lower unemployment. The reason is that, with a higher marginal tax rate, it becomes less costly for unions to »buy« more jobs through wage moderation, since a given fall in the pre-tax wage rate will now lead to a smaller drop in the after-tax wage. Stronger tax progressivity also generates wage moderation and lower involuntary unemployment in efficiency wage models where employers pay wages above the market-clearing level as a means of inducing higher productivity of their workers. The explanation is that a rise in the marginal tax rate reduces the effectiveness of a high (pre-tax) wage rate as an instrument for encouraging high labour productivity, given that workers care about after-tax rather than pre-tax wages.

These observations do not imply that tax progressivity is a »free lunch« in imperfect labour markets, since higher marginal tax rates also have distortionary effects, e.g. by inducing unions to bargain for fewer working hours, and by reducing work efforts in an efficiency wage setting. However, even when these tax distortions are accounted for, the numerical analysis in Sørensen (1999) suggests that a substantial degree of tax progressivity can be rationalised on pure efficiency grounds, especially when unemployment benefits are generous.

A challenging task is to analyse optimal taxation when the tax system must serve the goal of redistributing income while at the same time accounting for non-tax labour market frictions. This problem was recently addressed by Boone and Bovenberg (2004) and Hungerbühler et al. (2006) for an economy with labour market frictions stemming from imperfect information that gives rise to job search. When an employer with a job vacancy has been matched with an unemployed job seeker, the wage bargain between the two parties determines the distribution of the rent from the job match.⁵ The worker's share of the rent represents the return to his job search effort, and the employer's share is the return to his investment in searching for a worker. To generate an efficient labour market equilibrium, the employer's share of the surplus from the match should correspond to the increase in the probability of a match occurring when he posts an extra vacancy. This efficiency condition is known as the Hosios condition, Hosios (1990). However, if the worker's bargaining power is »too« strong (weak), his share of the rent will be larger (smaller) than the efficient share, thus weakening (strengthening) the incentive for employers to post vacancies and thereby generating too much (little) labour market slack and hence too much (little) unemployment.

Hungerbühler et al. (2006) analyse the optimal non-linear tax-transfer system in the benchmark case where the Hosios condition is met so that the labour market equilibrium would be efficient in the absence of tax. The introduction of a positive marginal labour income tax rate for redistributive purposes induces firms and workers to bargain

5. The rent arises from the fact that neither party can immediately and costlessly find another match.

for lower gross wages, because it raises the cost to the employer of providing the worker with any given increase in the after-tax wage while at the same time reducing the cost to the worker of conceding more profit to the employer by accepting a lower pre-tax wage rate. Hence the optimal tax system involves a level of wages and unemployment *below* the efficient level when the Hosios condition is met, since it is optimal for the government to accept this labour market distortion in return for the redistributive gain from progressive taxation. Indeed, since progressivity implies that the average tax rate increases with income, and since a rise in the average (as opposed to the marginal) tax rate generates upward wage pressure in the Nash bargaining set-up considered by Hungerbühler et al., they find that it is optimal to set marginal tax rates above the average tax rates to moderate wages. Simulating their model for plausible parameter values, they show that the optimal marginal tax rates in their economy with job search frictions tend to be substantially higher than the optimal tax rates in the traditional Mirrleesian optimal tax model with a competitive labour market.

The study of optimal taxation in imperfect labour market is still in its infancy, but the work discussed in this section suggests that it may generate new interesting insights with important implications for tax policy.

2. Optimal indirect taxation

Does optimal tax theory offer any useful guidance for indirect taxation? This is the topic addressed in the present part of the paper.

It is well known that a uniform indirect ad valorem tax on all goods and services is equivalent to a proportional tax on labour income.⁶ From a theoretical perspective, the main issue is thus whether there is a case for introducing a system of differentiated commodity taxes? And if so, which commodities should bear the highest (lowest) rates of tax? There are two strands of optimal tax literature dealing with this issue.

6. We can be a little more precise here. Suppose the consumer's life cycle is divided into two periods, indicated by subscripts 1 and 2. Let C_i , W_i , Π_i and T_i denote consumption, wage income, rent income ('pure profit') and transfer income in period i , respectively, and suppose the consumer receives an inheritance I in period 1 and leaves a bequest B at the end of period 2. If S is the saving undertaken in period 1, r is the interest rate, and t is the uniform indirect tax on consumption, assumed constant over time, the consumer's budget constraints in the two periods are:

$$\text{Period 1: } S = W_1 + \Pi_1 + T_1 + I - (1+t)C_1, \quad \text{Period 2: } (1+t)C_2 + B = (1+r)S + W_2 + \Pi_2 + T_2.$$

Eliminating S and consolidating, one obtains the lifetime budget constraint:

$$C_1 + \frac{C_2}{1+r} = (1-\tau) \left(W_1 + \Pi_1 + T_1 + \frac{W_2 + \Pi_2 + T_2}{1+r} + I - \frac{B}{1+r} \right), \quad \tau \equiv \frac{t}{1+t}.$$

This shows that a uniform ad valorem commodity tax levied at the rate t is equivalent to a proportional tax levied at the rate $\tau = t / (1+t)$ on the sum of wages, rents, transfers and the present value of net bequests received $(I - \frac{B}{1+r})$.

The Ramsey approach to indirect taxation

The first one, building on the classical contribution by Ramsey (1927), assumes that the government has to raise some given amount of revenue via indirect taxes. When one abstracts from consumer heterogeneity, the optimal tax problem boils down to raising the required amount of revenue in a manner that minimises the total deadweight loss. This leads to the famous Ramsey rule which says that indirect taxes should be designed so as to cause an equi-proportionate reduction in the compensated demands for all commodities. Thus the optimal indirect tax system seeks to avoid distorting the quantitative pattern of consumption, but since the own price and cross price elasticities will generally differ across commodities, commodity taxes (as a percentage of the total consumer price) should generally be differentiated to induce the same relative reduction in all quantities demanded.

The well-known inverse elasticity rule – stating that the optimal commodity tax rates are inversely related to the (compensated) own-price elasticity of demand – is a special case of the Ramsey rule, holding only when the cross price elasticities of demand for the taxed commodities are zero. The inverse elasticity rule essentially seeks to minimise tax distortions to labour supply, for when cross price elasticities in commodity demands are zero, the only way a tax-induced rise in the consumer price can reduce the demand for some commodity is by causing substitution from material consumption towards leisure. In such a setting a low own price elasticity of demand means that a commodity tax has little discouraging effect on labour supply.

The optimal tax rule discovered by Corlett and Hague (1953) is another special case of the Ramsey rule derived in a context with only two taxed commodities plus (untaxed) leisure. The Corlett-Hague rule states that the commodity which is more complementary to (less substitutable for) leisure should carry a relatively high tax rate to offset the tendency of the tax system to induce substitution towards leisure. Thus uniform taxation is optimal only in the special case where both commodities are equally substitutable for (complementary to) leisure. Again, we see that optimal indirect tax design seeks to minimise the tax distortions to labour supply that inevitably occur when only commodities (but not leisure) can be taxed.

More generally, if utility is generated by consumption of the goods bundle $x_0, x_1, x_2, \dots, x_n$, where x_0 is untaxed leisure, it can be shown, see Sandmo (1974) or Sadka (1977), that uniform commodity taxation is optimal if the utility function takes the form

$$U(x_0, x_1, \dots, x_n) = U(x_0, v(x_1, \dots, x_n)) \quad (28)$$

where the sub-utility function $v(x_1, \dots, x_n)$ is homothetic. In other words, uniform

commodity taxation is optimal if preferences are separable in leisure and commodities – so that all commodities are equally substitutable for leisure – and if all commodities have the same income elasticity of demand. The intuition is that when income elasticities are identical, the equi-proportionate reduction of all compensated commodity demands prescribed by the Ramsey rule also requires the same relative reduction of the *uncompensated* demands for all commodities. Since a uniform ad-valorem tax on all commodities is equivalent to a proportional labour income tax, it will indeed generate the same relative fall in the consumption of all commodities when they all have the same income elasticity and are all equally substitutable for leisure (labour).

The classical Ramsey rule focuses on the pure efficiency aspects of indirect taxation by abstracting from consumer heterogeneity. Diamond (1975) showed how the Ramsey rule is modified in a world of heterogeneous consumers where policy makers trade off efficiency against their redistributive goals. His analysis indicates that while efficiency may call for relatively high commodity tax rates on leisure complements, concerns about equity call for relatively low tax rates on commodities that weigh heavily in the budgets of low-income families.

The Mirrleesian approach to indirect taxation and the Atkinson-Stiglitz theorem

Although Diamond (op.cit.) introduced equity concerns into the theory of optimal indirect taxation, he did not account for the role that a non-linear income tax could play as a means of achieving the policy maker's distributive goals. Another strand of literature on indirect taxation – which might be termed »the Mirrleesian approach« – asks the question: which (if any) commodity taxes should supplement the income tax in order to attain an optimal trade-off between equity and efficiency when the government has to raise a given amount of total revenue?

This literature assumes that consumers have different abilities, reflected in their wage rates. The government cannot observe individual wage rates, but it observes individual incomes and is hence able to levy a non-linear income tax. At the same time, the government does not observe the individual taxpayer's consumption of a particular good, so commodity taxes must be impersonal and hence linear. Note that since uniform commodity taxation is equivalent to a proportional wage income tax, the issue for indirect tax policy is whether there is any need for differentiated commodity taxes when the government can levy a personal tax on labour income.

The benchmark result in this line of research was established early on by Atkinson and Stiglitz (1976) who showed that if preferences are weakly separable in leisure and all other goods taken together, it is inoptimal to differentiate taxes across commodities when the government optimises the non-linear labour income tax. Weak separability of preferences implies that the utility function takes the form (28), but the Atkinson-

Stiglitz theorem does *not* require that the subutility function $v(x_1, \dots, x_n)$ be homothetic.

The intuition behind the Atkinson-Stiglitz theorem may be explained as follows. The nonlinear income tax is deployed to achieve the optimal amount of redistribution. If the tax system becomes too progressive (relative to the optimum), an individual with a high wage rate will choose to work less so as to »mimic« the income level of an individual with a lower wage rate. In that case the two persons will pay the same amount of income tax and have the same disposable income, but the person with the higher wage rate will enjoy more leisure. However, the government obviously cannot use the income tax to achieve a further redistribution of welfare from high-ability to low-ability individuals when the former persons mimic the incomes of the latter. If differentiated commodity taxes could reduce the incentive for mimicking, the scope for redistribution via the income tax would increase. But with separability between goods and leisure, the government cannot use differentiated commodity taxes to impose a higher tax burden on high-ability persons by exploiting any relationship between leisure and the consumption of taxable commodities, since the high-ability individuals choose the same commodity bundle as the low-ability persons whom they mimic, even though they consume more leisure. Hence differentiated commodity taxes cannot play any useful role by relaxing the non-mimicking constraints that limit the government's ability to redistribute income. Differentiated taxation will only introduce distortions in commodity choices and hence it is not optimal.

Atkinson and Stiglitz proved their theorem by assuming that the government optimises the non-linear income tax schedule. More recently, Laroque (2005) and Kaplow (2006) have shown that as long as preferences are weakly separable in leisure and commodities, differentiated commodity taxes are undesirable even when the non-linear income tax schedule is *not* optimal. Since this strong result is not so well known, and since Kaplow's proof strengthens the intuition for the Atkinson-Stiglitz theorem, it is worth providing a sketch of his proof.⁷ Suppose each consumer has a utility function of the separable form (28), with leisure x_0 being equal to $E - L$, where E is the time endowment and L is labour time. An individual with the pre-tax wage rate w (reflecting his exogenous productivity) maximises the utility function (28) subject to the budget constraint

$$\sum_{i=1}^n (p_i + t_i) x_i = wL - T(wL), \quad (29)$$

7. The sketch offered here is a condensed version of the heuristic exposition in Kaplow (2008, pp. 125-133). The rigorous mathematical proof is given in Kaplow (2006). Working independently of Kaplow, Laroque (2005) also provided a rigorous proof of the redundancy of indirect taxation under the assumptions stated above.

where p_i is the (fixed) producer price of commodity i , t_i is the tax rate on that commodity, and $T(wL)$ is the consumer's labour income tax bill, given the non-linear tax schedule $T(\cdot)$. With \mathbf{t} denoting the vector of commodity tax rates, we may now introduce the indirect sub-utility function $V(\mathbf{t}, T(wL), wL)$, defined as the value of sub-utility $v(x_1, \dots, x_n)$ maximised over all the commodities x_1, \dots, x_n , where the commodity tax vector \mathbf{t} , the income tax schedule $T(wL)$, and pre-tax income wL are taken as given. Note that the indirect sub-utility depends only on income, wL , but not on labour supply L . This is a consequence of the separable utility function (28) which implies that the marginal rate of substitution between any two commodities is independent of the amount of labour.

Starting from an initial situation with differentiated commodity taxation (different ad valorem tax rates) where there exists i, j such that $(p_i + t_i) / (p_j + t_j) \neq p_i / p_j$, consider now a move towards a uniform commodity tax vector \mathbf{t}^* satisfying $(p_i + t_i^*) / (p_j + t_j^*) = p_i / p_j$ for all i, j . By going through the following steps, Kaplow (2006) shows that this reform allows the government to generate a Pareto improvement. First, suppose the government introduces a new nonlinear income tax schedule $T^*(wL)$ which has the property that, at all income levels, the utility of all taxpayers will be the same after the reform as before the reform, provided they do not change their labour supply. If the pre-reform tax schedule is $T(wL)$, the post-reform tax schedule will thus satisfy $V(\mathbf{t}, T(wL), wL) = V(\mathbf{t}^*, T^*(wL), wL)$ for all wL , since in this case all taxpayers will then also enjoy the same total utility $U(E - L, V(\cdot))$ before and after the reform, assuming that their labour supply L does not change. And in fact no taxpayer will have any incentive to change his/her labour supply. The reason is that the adjustment of the income tax schedule is undertaken for all income levels, so when a taxpayer with wage rate w varies his labour supply, thereby varying wL , he will find that the equality $V(\mathbf{t}, T(wL), wL) = V(\mathbf{t}^*, T^*(wL), wL)$ holds for all choices of L . Therefore, if L^o was the optimal labour supply that maximised total utility $U(E - L, V(\cdot))$ before the reform, it must still be the optimal labour supply after the reform, given that the maximum attainable subutility $V(\cdot)$ is unchanged for any choice of L .

The next step in the proof is to show that the move towards uniform commodity taxation combined with the utility-preserving change in the income tax schedule will generate an increase in government revenue. Suppose for a moment that taxpayers do not respond to the change in relative commodity prices, continuing to consume the same amounts of each individual commodity as before. With an unchanged labour supply in all income groups, the utility-preserving change in the income tax schedule would then imply that the change in indirect tax payments would be exactly offset by the change in income tax payments so as to leave real disposable incomes unchanged at all income levels. As a consequence, total tax revenue would also be unchanged. In

reality, consumers would, of course, react to the commodity tax reform and the ensuing change in relative consumer prices by substituting towards those commodities that have become relatively cheaper. Since this substitution can only make consumers better off, it allows a further increase in the income tax bill at each income level without reducing consumer utility below the pre-reform level. At the same time, the substitution across commodities does not change the total revenue from indirect taxation, since all commodities bear the same ad valorem tax rate after the reform. The utility-preserving reform of direct and indirect taxation must, therefore, increase total public revenue, allowing the government to create a Pareto improvement by distributing the extra revenue to all taxpayers.

Although popular and important as a benchmark case, the assumption that preferences are separable in leisure and commodities is, of course, rather restrictive. Christiansen (1984) analysed which commodity taxes should supplement an optimal nonlinear income tax when preferences are not separable. He found that a commodity should be taxed (subsidised) if it is positively (negatively) related to leisure in the sense that more (less) of the good is consumed if more leisure is obtained at a constant income. This result has the same flavour as the Corlett-Hague rule: the indirect tax system should discourage the purchase of commodities that tend to be consumed jointly with leisure. The intuition for Christiansen's result follows directly from the intuition underlying the Atkinson-Stiglitz theorem: by taxing complements to leisure, the government makes it less attractive for the more productive individuals (who must sacrifice less leisure to earn a given amount of income) to mimic the income of the less productive individuals, so in this way commodity taxes relax the incentive-compatibility constraints that restrict the government's ability to redistribute income via the income tax. Saez (2002b) has extended Christiansen's analysis to a setting with heterogeneous consumer tastes, showing that the optimal nonlinear labour income tax should be supplemented not only by excises on commodities that are consumed jointly with leisure, but also by taxes on commodities for which high-income earners tend to have a relatively strong taste.

Further arguments for uniform indirect taxation

The empirical work by Browning and Meghir (1991), as well as casual observation, strongly indicates that some commodities are better substitutes for leisure than others, i.e. consumer preferences are in fact not weakly separable in leisure and commodities. This would seem to call for a system of differentiated commodity taxes with the characteristics suggested by Christiansen (1984) and Saez (2002b). A number of practical arguments nevertheless speak in favour of uniform taxation.

First of all, the government may simply lack the solid evidence on the compensated cross-price elasticities with leisure that would be needed to implement the optimal differentiation of commodity taxes. Second, a commodity tax system differentiated according to the principles of optimal tax theory would require frequent changes in tax rates in response to changes in tastes and technologies. This would introduce a potentially welfare-reducing element of risk and uncertainty into the tax system. Third, a uniform VAT is easier to administer and less susceptible to fraud than a VAT system with several differentiated rates, since a uniform VAT does not require any borderlines to be drawn between different categories of goods. Fourth, acceptance of differentiated taxation as a general principle might invite special interest groups to lobby for low tax rates on particular economic activities, so adherence to a principle of uniformity may provide a stronger bulwark against wasteful lobbying.

All of this suggests that uniformity should be the guiding principle for indirect taxation, except where consumption of specific commodities generates obvious externalities that need to be corrected through high excises.⁸ However, recent contributions to optimal tax theory by Kleven, Richter and Sørensen (2000)⁹ and Kleven (2004), allowing for home production along with market production, have pointed to some specific areas where deviations from uniform taxation may be warranted.

Optimal indirect taxation with household production: a simple example

To illustrate the importance of home production for optimal taxation, suppose the representative consumer produces services within the household (S^h) subject to the concave household production function

$$S^h = h(H), \quad h' > 0, \quad h'' < 0, \quad (30)$$

where H denotes hours spent working in the home. Suppose further that services can also be purchased in the market place so that total service consumption (S) is the sum of services bought in the market (S^m) and services produced at home:

$$S = S^m + S^h. \quad (31)$$

If the consumer spends L hours working in the market and his total time endowment is $E=1$, his consumption of leisure (ℓ) will be

$$\ell = 1 - L - H. \quad (32)$$

8. Sandmo (1975) provides the classical treatment of optimal commodity taxation in the presence of externalities.

9. The paper by Kleven et al. (2000) built on the earlier work by Sandmo (1990).

Apart from leisure and services, the consumer also consumes »goods« (G). For concreteness, let utility be given by

$$U = U(\ell, v(G, S)) = \ell^\alpha (G^\beta S^{1-\beta})^{1-\alpha}, \quad 0 < \alpha < 1, \quad 0 < \beta < 1. \quad (33)$$

Thus preferences are weakly separable in leisure and commodities, and the sub-utility function $G^\beta S^{1-\beta}$ is homothetic. As mentioned earlier, with such preferences uniform commodity taxation would be optimal in the absence of home production, but as we shall see, this conclusion no longer holds when home production is allowed for.

Without loss of generality we may assume that it takes one unit of market work to produce one unit of each of the two commodities G and S . Choosing leisure as the numeraire good (i.e. normalising the wage rate at unity) and assuming that labour is the only factor of production, the producer prices of G and S will then both be equal to one. With t_G and t_S denoting the unit commodity taxes imposed on G and S , respectively, the consumer's budget constraint therefore becomes

$$P_G G + P_S S^m = L, \quad P_G = 1 + t_G, \quad P_S = 1 + t_S, \quad (34)$$

where P_G and P_S are consumer prices, and $L (= wL)$ is the consumer's market income. Using (30) through (32), we may rewrite (34) as

$$P_G G + P_S S + \ell = Y, \quad Y \equiv 1 + P_S h(H) - H. \quad (35)$$

The variable Y is the consumer's »full income«, consisting of his potential market income ($=1$) plus the »profit« from home production, $P_S h(H) - H$. According to (35) this full income may be spent on acquiring commodities or on »buying« leisure (by abstaining from work).

The consumer maximises utility (33) subject to the budget constraint (35). The solution to this problem can be shown to imply that

$$\ell = \alpha Y, \quad P_G G = \beta(1 - \alpha)Y, \quad P_S S = (1 - \beta)(1 - \alpha)Y, \quad (36)$$

$$P_S h'(H) = 1 \quad \Rightarrow \quad H = H(t_S), \quad H' = -\frac{h'}{P_S h''} > 0. \quad (37)$$

According to (37) the consumer engages in home production until the resulting marginal saving on services bought in the market, $P_S h'(H)$, equals the marginal opportunity cost of working at home rather than in the market ($= w = 1$). Since $P_S = 1 + t_S$, we see that this behaviour implies that the time spent on home production rises with

the tax rate imposed on services delivered from the market. Using the results in (37) and recalling that $S^h = h(H)$, we note that full income depends on t_S in the following way:

$$Y(t_S) \equiv 1 + (1 + t_S)h(H(t_S)) - H(t_S), \quad Y' = S^h. \quad (38)$$

Inserting the solutions in (36) into the direct utility function (33), we obtain the indirect utility function

$$V = c \cdot \left[(1 + t_G)^\beta (1 + t_S)^{1-\beta} \right]^{\alpha-1} \cdot Y(t_S), \quad (39)$$

where c is a constant depending solely on the taste parameters α and β . We may now prove that, starting from an initial situation with uniform taxation ($t_G = t_S$), the government can increase welfare by moving towards a situation where $t_G > t_S$. We do so by showing that if the government raises t_G and lowers t_S in a way that maintains constant utility, it will earn additional revenue. Clearly the government will then be able to raise consumer welfare by transferring the additional revenue back to the consumer. Differentiating (39) and recalling from (38) that $Y' = S^h$, one finds that

$$dV = 0 \quad \Rightarrow \quad dt_G = \left(\frac{P_S S^h - (1 - \beta)(1 - \alpha)Y}{\beta(1 - \alpha)Y} \right) \cdot dt_S. \quad (40)$$

Note that since $S = S^m + S^h$ and $(1 - \beta)(1 - \alpha)Y = P_S S$ (cf. (36)), the numerator on the right-hand side of (40) equals $-P_S S^m$, so as long as the consumer buys some services from the market ($S^m > 0$), the maintenance of a constant utility level will indeed require a rise in t_G when t_S is lowered.

Consider now how this utility-preserving perturbation of tax rates will affect government revenue. Using (31) and inserting the consumer's optimal demands for G and S stated in (36), we may write total revenue as

$$\begin{aligned} R &= t_G G + t_S S^m = t_G G + t_G (S - S^h) = \\ & \left[\beta \left(\frac{t_G}{1 + t_G} \right) + (1 - \beta) \left(\frac{t_S}{1 + t_S} \right) \right] Y(t_S) - t_S h(H(t_S)), \end{aligned} \quad (41)$$

which may be differentiated to give

$$dR = \frac{\beta(1-\alpha)Y}{(1+t)^2} \cdot dt_G + \left\{ \frac{(1-\beta)(1-\alpha)Y}{(1+t)^2} + \left[(1-\alpha) \left(\frac{t}{1+t} \right) - 1 \right] S^h - \left(\frac{t}{1+t} \right) H' \right\} \cdot dt_S, \quad (42)$$

where we have used (30), (37), (38) and the fact that $t_G = t_S = t$ initially. Substituting (40) into (42), we obtain:

$$dR = - \left(\frac{t}{1+t} \right) \left[\alpha S^h + H'(t_S) \right] \cdot dt_S > 0 \quad \text{for } dt_S < 0. \quad (43)$$

Thus government revenue does in fact increase, enabling the government to raise consumer welfare by recycling the extra revenue.¹⁰ Note that in the absence of home production we would have $S^h = H' = 0$, so in that case there would be no revenue gain and hence no welfare gain by deviating from uniform taxation, according to (43). In the absence of home production our model thus reproduces the standard result that uniform commodity taxation is optimal when preferences are separable in leisure and commodities and utility is homothetic in commodities. But once home production is allowed for, it follows from our analysis that commodities which can be produced at home as well as in the market economy should be taxed more lightly than commodities which cannot be produced within the household.

The practical importance of this result is that it is easy to think of commodities that would be candidates for reduced taxation under this principle. For example, housing repair and repair of other consumer durables, child care, cleaning and window-cleaning, garden care, cooking etc. are all consumer services that can either be produced at home or be delivered from the market. According to the analysis above such services should be taxed more lightly than, say, manufactured goods that cannot realistically be produced within the household.

Kleven, Richter and Sørensen (2000) show that this conclusion will almost surely hold also when preferences are not separable and homothetic. In fact they find that even if the consumer services which can be produced at home as well as in the market might be complementary to leisure, it may still be optimal to impose a relatively low

10. There are two reasons for the rise in revenue. First, even if home production were unchanged, the fall in t_S reduces the revenue loss caused by the fact that services produced in the home cannot be taxed. This effect is captured by the first term in the square bracket in (43). In addition, the fall in t_S induces consumers to substitute market-produced services for home production, thereby increasing the tax base. This is captured by the last term in the square bracket in (43).

tax rate on these services. The intuition for this modification to the Corlett-Hague rule is that a high tax on complements to leisure is not an efficient way of stimulating tax-discouraged labour supply to the market if such a tax induces a shift from market production to home production. The point is that the optimal tax system must minimise the distortionary substitution away from market activities towards untaxed activities. Taxes should distort the pattern of *market* activity as little as possible, and this calls for lenient taxation of those market activities that can most easily be replaced by home production. These activities would typically include the consumer services mentioned above.

The analysis above uses a conventional specification of utility which abstracts from the fact that consumption is a time-consuming activity. Kleven (2004) analyses optimal commodity taxation in the generalised household production framework proposed by Becker (1965) where all utility-generating consumption activities require the combination of some good or service with household time spent on the act of consumption (or on acquiring the commodity).¹¹ In this framework households combine a »commodity input« (the purchase of some good or service) with an input of time into utility-generating household activities (termed »household production«). Kleven (op.cit.) shows that the optimal commodity tax system imposes relatively high tax rates on commodities whose consumption requires a large input of household time. In this way the optimal tax system minimises the amount of time that is diverted from market work to consumption activity within the household sector. Indeed, when all utility-generating activities require a positive input of commodities as well as time, and when all commodities can be taxed, Kleven finds that the optimal commodity tax rate on some commodity j depends only (and inversely) on the ratio of the value of the commodity input to the sum of the values of the commodity input and the time input in the consumption activity in which commodity j is used. In other words, it is not necessary for policy makers to know all the compensated own price and cross price elasticities of demand to implement the optimal tax system. In principle, all that is needed is a combined survey of consumption expenditures and household time allocation. This information requirement is clearly less daunting than the information on compensated cross price elasticities with leisure needed to implement the classical Ramsey tax rule.

According to Kleven's analysis any type of consumption activity which requires little time, or even saves time, should carry a relatively low tax rate. The types of consumer services mentioned earlier in this section typically have this property: hiring somebody in the market to supply a service rather than engaging in do-it-yourself activities saves household time, so such services should be favoured by the tax system, just as implied

11. Boadway and Gahvari (2006) also analyse optimal commodity taxation when consumption is time-consuming.

by the analysis based on the more traditional model of household production presented above.

It is interesting to note that several European countries have experimented with reduced rates of tax or direct subsidies to a number of labour-intensive services that are easily substitutable for home-produced services, and the EU has recently allowed its member states to apply reduced rates of VAT on such services. The contributions to optimal tax theory discussed in this section suggest that there is indeed a rationale for this policy, and the numerical simulations in Sørensen (1997) indicate that the efficiency gains from reduced tax rates on this type of service could be significant.

3. Optimal taxation of income from capital

Should capital income be taxed?

The analysis of indirect taxation has important implications for the long-standing debate on the taxation of income from capital. The fundamental issues in this debate are whether capital income should be taxed at all, and if so, whether all returns to capital should be taxed at the same rate, i.e. whether capital income taxation should be »neutral«?

The Atkinson-Stiglitz theorem discussed in Part 2 provides a useful starting point for addressing the first issue: should we tax income from capital? Consider the standard life cycle model where the consumer supplies labour L during young age and enjoys total consumption C_1 and C_2 during young and old age, respectively. Suppose that labour income is subject to the nonlinear tax schedule $T(wL)$, where w is the pre-tax wage rate, and assume further that the return to capital (r) is taxed at the rate t_r . Denoting savings during young age by S , we may then write the consumer's budget constraints as

$$\begin{aligned}
 S = wL - T(wL) - C_1, \quad C_2 &= [1 + r(1 - t_r)]S, \\
 &\Downarrow \\
 C_1 + \frac{C_2}{1 + r(1 - t_r)} &= wL - T(wL) \Leftrightarrow C_1 + \\
 \left[\frac{1}{1 + r} + \frac{rt_r}{(1 - r)(1 + r(1 - t_r))} \right] C_2 &= wL - T(wL). \tag{44}
 \end{aligned}$$

As shown by the second term in the square bracket in (44), a capital income tax may be seen as an excise tax on future consumption, since it raises the relative price of future consumption above the level $1/(1 + r)$ that would prevail in the absence of capital

income taxation. If we normalize the consumer's time endowment at unity, his consumption of leisure during young age is $1-L$. Suppose now that his lifetime utility is given by

$$U = U(1 - L, C_1, C_2) = U(1 - L, v(C_1, C_2)). \quad (45)$$

These preferences are separable in leisure and consumption, so if we translate the analysis of Atkinson and Stiglitz (1976) to an intertemporal setting, interpreting C_1 and C_2 as two different commodities, we may conclude that these two goods should be uniformly taxed, given that the government can pursue its distributional goals via the nonlinear labour income tax. This was also the conclusion reached by Ordober and Phelps (1979) in an explicit overlapping generations setting, provided the government can use debt policy or public investment to steer the capital stock towards its Golden Rule level.

The policy relevance of this result is that, in the absence of firm knowledge about the degree of substitutability or complementarity between leisure and consumption in the two periods, a natural benchmark is to assume that present and future consumption are equally substitutable for leisure. As we have just seen, the marginal effective tax rate on capital income should then be set to zero. In practice this could be achieved by exempting capital income from tax, or by introducing a cash flow expenditure tax which is known to imply a zero marginal effective capital income tax rate (when the tax rate is constant over time).

However, as argued by Erosa and Gervais (2002), the leisure taken by the typical consumer does in fact tend to increase with age, suggesting that leisure and future consumption are complements. In that case the theory of optimal commodity taxation suggests that a positive capital income tax could be part of an optimal tax system, assuming that the life cycle model provides an adequate description of intertemporal consumer behaviour.

An alternative vision of consumer behaviour is embodied in the popular infinite horizon model originally suggested by Ramsey (1928) and usually defended by the assumption that generations are linked via an altruistic bequest motive so that consumers effectively behave as if they had an infinite horizon. Within this framework Chamley (1986) and Judd (1985) found that the optimal steady state tax rate on capital income is zero even if the alternative to capital income taxation is a distortionary labour income tax. One way of explaining this result is to note that in the Ramsey model the steady-state after-tax interest rate is closely tied to the consumer's exogenous utility discount rate. With a constant long-run equilibrium after-tax interest rate, a capital income tax gets fully shifted onto the pre-tax interest rate, i.e. the supply of capital is in effect in-

finitely elastic in the long run. Clearly it is not optimal to tax a factor with an infinite elasticity of supply. Further, regardless of the relevance of the Ramsey infinite horizon model, a small open economy with perfect capital mobility also faces an infinitely elastic supply of capital from the world capital market and hence should levy no source-based taxes on capital, as argued by Razin and Sadka (1991).

However, for a number of reasons discussed at length in Sørensen (2007a), the result that the optimal capital income tax rate is zero is not robust. For example, if pure profits cannot be fully taxed away, a source-based capital income tax may be a second-best means of taxing location-specific rents generated by domestic economic activity, see Stiglitz and Dasgupta (1971) and Huizinga and Nielsen (1997). The source-based corporate income tax may also be a necessary backstop to the personal income tax, since otherwise income could be accumulated free of tax in the corporate sector. Further, political constraints on the use of other sources of revenue may simply force the government to rely, to some extent, on the revenue from capital income taxes.

Should capital income taxes be neutral?

It is usually argued that a capital income tax should be neutral in the sense that all returns to capital should face the same marginal effective tax rate. In other words, capital income taxes should not be differentiated across different sectors or economic activities. This may be seen as an application of the famous production efficiency theorem of Diamond and Mirrlees (1971) which says that the second-best optimal tax system avoids production distortions provided the government can tax away pure profits and can tax all transactions between households and firms.

In an open economy production efficiency requires that the government can tax all returns to capital received by domestic residents from foreign as well as domestic sources. But enforcing a residence-based tax on worldwide income requires an effective system of information exchange between national tax administrations, and such a system does not exist in practice, not even in the European Union.¹² Tax administrators therefore find it hard to monitor the taxpayers' foreign source capital income, so in practice capital income taxes tend to be source-based. The domestic capital income tax thus becomes a selective tax on domestic investment as opposed to investment abroad. This violates the assumptions underlying the Diamond-Mirrlees production efficiency theorem and hence there is no longer any presumption that all domestic activities should

12. The EU Savings Tax Directive does require exchange of information on interest income received by EU personal taxpayers from other EU countries (and from a number of co-operating jurisdictions), but there are a number of ways in which taxpayers can in practice escape information exchange, e.g. by placing a legal person or a trust or fiduciary between the beneficial owner and the interest-generating asset. See the European Commission (2008) for details on this.

face the same marginal effective capital income tax rate. Instead, there is a case for differentiating the capital income tax in accordance with Ramsey principles.

A Ramsey rule for optimal source-based capital income taxation

To see this, consider a small open economy where total output Y is given by the sum of the outputs Y_1 and Y_2 from two production sectors using capital inputs K_1 and K_2 plus a fixed factor.¹³ The outputs are given by the concave production functions

$$Y_1 = f(K_1), \quad f' > 0, \quad f'' < 0, \quad Y_2 = F(K_2), \quad F' > 0, \quad F'' < 0. \quad (46)$$

Capital is perfectly mobile across borders, and investors can earn a fixed after-tax return r by investing abroad. If t_i is the unit tax on capital invested in domestic sector i , the marginal pre-tax return on investment in that sector must therefore equal $r + t_i$ to ensure an after-tax return equal to that obtainable by investing abroad. Hence we have the capital market equilibrium conditions

$$f'(K_1) = r + t_1, \quad F'(K_2) = r + t_2, \quad (47)$$

which imply that the capital stocks invested in the two sectors vary negatively with the cost of capital:

$$K_1 = K_1(r + t_1), \quad K_1' = \frac{1}{f''} < 0, \quad K_2 = K_2(r + t_2), \quad K_2' = \frac{1}{F''} < 0. \quad (48)$$

Assuming a given national wealth endowment \bar{K} , the amount of domestically-owned capital invested abroad is $\bar{K} - K_1 - K_2$, so total national income is

$$\begin{aligned} Y &= Y_1 + Y_2 + r(K - K_1 - K_2) \\ &= f(K_1(r + t_1)) + F(K_2(r + t_2)) + r[\bar{K} - K_1(r + t_1) - K_2(r + t_2)]. \end{aligned} \quad (49)$$

The optimal tax system maximises national income (49) with respect to t_1 and t_2 subject to the constraint that the government must raise a given amount of revenue R :

$$t_1 \cdot K_1(r + t_1) + t_2 \cdot K_2(r + t_2) = R. \quad (50)$$

13. The model is inspired by the seminal general equilibrium framework developed by Harberger (1962). A more general version of the model, incorporating three factors of production, is analysed in Sørensen (2007b); however, the qualitative conclusions remain the same as those reported below.

The solution to this problem can be written in either of the following ways:

$$\varepsilon_1 = \varepsilon_2, \quad \varepsilon_i \equiv \frac{dK_i}{dt_i} \frac{t_i}{K_i}, \quad i = 1, 2 \quad (51)$$

$$\frac{t_i}{r + t_i} = \frac{\lambda}{\eta_i}, \quad \eta_i \equiv - \frac{dK_i}{d(r + t_i)} \frac{r + t_i}{K_i}, \quad i = 1, 2. \quad (52)$$

Equation (51) is a *Ramsey rule for optimal capital taxation* stating that, at the margin, the tax system should cause the same relative reduction of investment in the different production sectors. In general, this rule calls for *differentiated* capital taxation since the elasticity of capital demand with respect to the tax rate (ε_i) will generally differ across sectors. The optimal degree of tax differentiation is given by the *inverse elasticity rule* (52) where λ is the shadow price of public funds. This rule says that the marginal effective capital income tax rate $t_i / (r + t_i)$ on a given sector should be inversely proportional to that sector's elasticity of capital demand with respect to the cost of capital (η_i). Obviously this policy rule is analogous to Ramsey's famous inverse elasticity rule for optimal commodity taxation.

If production functions take the constant-elasticity forms $Y_1 = AK_1^\alpha$, $0 < \alpha < 1$, and $Y_2 = BK_2^\beta$, $0 < \beta < 1$, it follows from (47) that $\eta_1 = 1/(1 - \alpha)$ and $\eta_2 = 1/(1 - \beta)$, implying that the optimal relative tax rates given by (52) become

$$\frac{t_1 / (r + t_1)}{t_2 / (r + t_2)} = \frac{1 - \alpha}{1 - \beta} \quad (53)$$

In competitive markets the parameters α and β are equal to the capital income shares of total sectoral income which can be estimated from the national income accounts. We see that the more capital-intensive sector (measured by the capital income share) should carry a lower relative tax rate.

Optimal capital taxation and capital mobility

With the constant elasticity production functions assumed above, the solutions for capital stocks given in (48) take the form

$$K_1 = \left(\frac{r + t_1}{\alpha} \right)^{\frac{1}{\alpha-1}}, \quad K_2 = \left(\frac{r + t_2}{\beta} \right)^{\frac{1}{\beta-1}}, \quad (54)$$

so that the tax elasticities in (51) become

$$\varepsilon_1 = -\left(\frac{1}{1-\alpha}\right)\left(\frac{t_1}{r+t_1}\right), \quad \varepsilon_2 = -\left(\frac{1}{1-\beta}\right)\left(\frac{t_2}{r+t_2}\right). \quad (55)$$

From (54) it follows that a higher tax rate on one sector does not channel more capital into the other domestic sector. Hence the elasticities in (55) arise only from the international mobility of capital. We see that the greater the capital intensity of production (measured by the size of α and β), the larger is the capital outflow generated by a given tax rate, so the lower is the optimal sectoral tax rate.

This is consistent with the observation that governments often tend to offer relatively generous tax treatment of activities that are perceived to be particularly mobile across borders. For example, in practically all countries the shipping industry benefits from large tax concessions. Firms in this sector are indeed highly capital intensive and internationally mobile, in accordance with the analysis above.

It should be stressed once again that, in the model framework described here, the optimal policy for the world as a whole would be a system of automatic international information exchange enabling governments to enforce a uniform residence-based capital income tax that would preserve production efficiency. But in the absence of the worldwide co-operation needed to implement such a regime, governments are exposed to the forces of international tax competition unleashed by source-based taxation. From a national perspective standard Ramsey principles familiar from the theory of optimal commodity taxation may then provide a rationale for differentiated capital income taxation in order to minimise distortions to the domestic pattern of investment.

As in the case of indirect taxation, administrative and political economy concerns suggest that tax economists should be cautious when advising departures from uniform capital income taxation. For example, extensive differentiation of capital income tax rates would invite tax avoidance through transfer-pricing within groups of affiliated companies. The work of Auerbach (1989) also indicates that the efficiency gains from an optimal system of differentiated capital income taxes may generally be small. The analysis in this paper nevertheless suggests that in the case of highly mobile activities, deviations from »neutrality« may be warranted from a national perspective as long as national governments cannot effectively coordinate their tax policies.

4. Concluding remarks

This paper has offered some examples from the recent literature on optimal taxation in the hope of convincing the reader that this body of theory can, in fact, offer useful guidance for practical tax policy. The first part of the paper illustrated how an optimal non-linear labour income tax schedule may be derived from estimates of labour supply

elasticities on the intensive and the extensive margin combined with data on the wage distribution plus assumptions on the relative marginal social evaluation of income for different income groups.

The second part of the paper dealt with the optimal design of indirect taxes. We noted that governments typically lack the information on cross-price elasticities with leisure needed to implement the differentiated commodity taxes prescribed by conventional optimal tax theory. This fact, combined with administrative and political economy considerations, suggests that uniform taxation should be the main guideline for indirect taxation. However, based on recent analyses of optimal taxation with household production, we argued that certain consumer services that can easily be produced as a do-it-yourself activity within the household should probably be subject to relatively low indirect tax rates in order to distort the pattern of market activities as little as possible.

The third part of the paper discussed the optimal taxation of income from capital, focusing on the issue of whether capital income taxes should be »neutral«, i.e. uniform across all types of investment. We found that neutral taxation should indeed be the norm if international information exchange enables governments to enforce residence-based capital income taxation. However, in the absence of an effective system of automatic information exchange, governments are typically forced to rely on source-based taxes on capital that discourage domestic investment as opposed to investment abroad. In that case the optimal system of capital income taxation secures an equal proportional reduction of investment in the different domestic production sectors. This principle generally calls for differentiated taxation in accordance with a Ramsey-type rule requiring that the marginal effective tax rate on capital income from a given sector be inversely proportional to that sector's elasticity of capital demand with respect to the cost of capital. We also saw that the elasticity of capital demand reflects the degree to which domestic taxation induces a capital export, so the Ramsey rule for optimal capital income taxation provides a rationale for relatively low capital tax rates on activities that are highly mobile across borders.

The recent advances in the theory of optimal labour income taxation have made that theory much more operational. As we explained in Section 1, the optimal labour income tax schedule is partly determined by the elasticities of taxable income for the various income groups. In recent years a large number of empirical studies have provided estimates of these elasticities for various countries.¹⁴ The research by Emmanuel Saez surveyed in Section 1 has also highlighted how the labour force participation elasticities of the different income groups influence the optimal labour income tax sche-

14. Kleven and Schultz (2009) offer a survey of these studies and present estimates of the elasticities of taxable income in Denmark for different income groups and various types of income.

dule. While there is still considerable uncertainty regarding the size of these elasticities, many recent empirical studies indicate that they could be quite large at the bottom of the income distribution. As we have seen, modern optimal tax theory would then suggest that some form of Earned Income Tax Credit or similar in-work benefits are part of an optimal tax system. Recent theoretical and empirical research on tax policy may thus help to explain why policies of »making work pay« have become increasingly popular in OECD countries during the last decades.

Some countries, including Finland, Sweden and several other EU member states, have also introduced tax benefits or subsidies to certain labour-intensive services competing directly with household and »underground« production. As our survey has shown, recent contributions to the theory of optimal indirect taxation suggest that there is indeed a rationale for this type of policy. Similarly, we saw how optimal tax theory could help to explain why national governments often find ways to reduce the effective capital income tax rates on investments that are particularly mobile internationally. At the same time the theory also provides a rationale for international co-operation by showing how countries could gain from an international exchange of information that would allow them to implement a neutral uniform capital income tax on domestic and foreign investment.

In summary, recent developments in the theory of optimal taxation have moved that body of theory closer to the real world of policy-making.

Literature

- Atkinson, A. B. and J. Stiglitz. 1976. The design of tax structure: direct versus indirect taxation. *Journal of Public Economics* 6, 55-75.
- Atkinson, A. B. and J. Stiglitz. 1980. *Lectures on Public Economics*. New York: McGraw- Hill.
- Auerbach, A. J. 1989. The deadweight loss from »non-neutral« capital income taxation. *Journal of Public Economics* 40, 1-36.
- Becker, G. S. 1965. A theory of the allocation of time. *Economic Journal* 75, 493-517.
- Boadway, R. and F. Gahvari. 2006. Optimal taxation with consumption time as a leisure or labor substitute. *Journal of Public Economics* 90, 1851-78.
- Boone, J. and A. L. Bovenberg. 2004. The optimal taxation of unskilled labor with job search and social assistance. *Journal of Public Economics* 88, 2227-58.
- Browning, M. and C. Meghir. 1991. The effects of male and female labour supply on commodity demands. *Econometrica* 51, 925-51.
- Christiansen, V. (1984). Which commodity taxes should supplement the income tax? *Journal of Public Economics* 24, 195-220.
- Chamley, C. 1986. Optimal taxation of capital income in general equilibrium with infinite lives. *Econometrica* 54, 607-22.
- Corlett, W. J. and D. C. Hague. 1953. Complementarity and the excess burden of taxation. *Review of Economic Studies* 21, 21-30.
- Creedy, J. 2009. The personal income tax structure: theory and policy. Paper presented at the conference on New Zealand Tax Reform – Where to Next? at the Victoria University of Wellington, 11-13 February, 2009.

- Diamond, P. A. 1975. A many-person Ramsey tax rule. *Journal of Public Economics* 4, 335-42.
- Diamond, P. A. 1998. Optimal income taxation: an example with a U-shaped pattern of optimal marginal tax rates. *American Economic Review* 88, 83-95.
- Diamond, P. A. and J. A. Mirrlees. 1971. Optimal taxation and public production (I). *American Economic Review* 61, 8-27.
- Edgeworth, F. Y. 1897. The pure theory of taxation. *Economic Journal* 7, 550-71.
- Erosa, A. and M. Gervais. 2002. Optimal taxation in life-cycle economies. *Journal of Economic Theory* 105, 338-69.
- European Commission. 2008. Report from the Commission to the Council, COM (2008) 552 final.
- Harberger, A. C. (1962). The incidence of the corporation tax. *Journal of Political Economy* 70, 215-40.
- Heckman, J. J. 1993. What has been learned about labor supply in the past twenty years? *American Economic Review* (AEA Papers and Proceedings) 83, 116-21.
- Hosios, A. 1990. On the efficiency of matching and related models of search and unemployment. *Review of Economic Studies* 57, 279-98.
- Huizinga, H. and S. B. Nielsen. 1997. Capital income and profits taxation with foreign ownership of firms. *Journal of International Economics* 42, 149-65.
- Hungerbühler, M., E. Lehmann, A. Parmentier and B. van der Linden. 2006. Optimal redistributive taxation in a search equilibrium model. *Review of Economic Studies* 73, 743-67.
- Judd, K. 1985. Redistributive taxation in a simple perfect foresight model. *Journal of Public Economics* 28, 59-83.
- Kaplow, L. 2006. On the undesirability of commodity taxation even when income taxation is not optimal. *Journal of Public Economics* 90, 1235-50.
- Kaplow, L. 2008. *The Theory of Taxation and Public Economics*. Princeton and Oxford: Princeton University Press.
- Kleven, H. J. 2004. Optimum taxation and the allocation of time. *Journal of Public Economics* 88, 545-57.
- Kleven, H. J., W. Richter and P. B. Sørensen. 2000. Optimal taxation with household production. *Oxford Economic Papers* 52, 584-94.
- Kleven, H. J. and E. A. Schultz. 2009. Estimating taxable income responses using Danish tax reforms. *Working paper*, March 2009.
- Laroque, G. 2005. Indirect taxation is superfluous under separability and taste homogeneity: A simple proof. *Economics Letters* 87, 141-44.
- McCulloch, J. R. 1845. *A Treatise on the Principles and Practical Influence of Taxation and the Funding System*. (Edited with introduction of D. P. O'Brien, 1975). London: Longman.
- Mirrlees, J. A. 1971. An exploration in the theory of optimum income taxation. *Review of Economic Studies* 38, 175-208.
- O'Brien, D. P. 2009. Introduction. In *Taxation and the Promotion of Human Happiness: An Essay by George Warde Norman* (Edited by D. P. O'Brien with J. Creedy). Cheltenham: Edward Elgar.
- Ordovery, J. and E. Phelps. 1979. The concept of optimal taxation in the overlapping generations model of capital and wealth. *Journal of Public Economics* 12, 1-26.
- Petersen, L. B. 2007. Velfærdsvægte for forskellige indkomstgrupper i Danmark. *Nationaløkonomisk Tidsskrift (Danish Journal of Economics)* 145, 265-87.
- Piketty, T. 1997. La redistribution fiscale face au chômage. *Revue Française d'Economie*, 12, 157-201.
- Ramsey, F. 1927. A contribution to the theory of taxation. *Economic Journal* 37, 47-61.
- Ramsey, F. 1928. A mathematical theory of saving. *Economic Journal*, 38, 543-59.
- Razin, A. and E. Sadka (1991). International tax competition and gains from tax harmonization. *Economics Letters* 37, 69-76.
- Revesz, J. T. 1989. The optimal taxation of labour income. *Public Finance* 44, 53-475.
- Roberts, K. 2000. A reconsideration of the optimal income tax. In Hammond, P. J. and Myles, G. D. (eds.), *Incentives and Organization: Essays in Honour of Sir James Mirrlees*. Oxford: Oxford University Press.

- Sadka, E. 1977. A theorem on uniform commodity taxation. *Journal of Public Economics* 7, 387-91.
- Saez, E. 2001. Using elasticities to derive optimal income tax rates. *Review of Economic Studies* 68, 205-29.
- Saez, E. 2002a. Optimal income transfer programs: intensive versus extensive labour supply responses. *The Quarterly Journal of Economics* 117, 1039-73.
- Saez, E. 2002b. The desirability of commodity taxation under non-linear income taxation and heterogeneous tastes. *Journal of Public Economics* 83, 217-30.
- Sandmo, A. 1974. A note on the structure of optimal taxation. *American Economic Review* 64, 701-06.
- Sandmo, A. 1975. Optimal taxation in the presence of externalities. *Swedish Journal of Economics* 77, 86-98.
- Sandmo, A. 1983. Progressive taxation, redistribution and labor supply. *Scandinavian Journal of Economics* 85, 311-23.
- Sandmo, A. 1990. Tax distortions and household production. *Oxford Economic Papers* 42, 78-90.
- Spadaro, A. 2008. Optimal taxation, social contract and the four worlds of welfare capitalism. Paris School of Economics, *Working Paper* No. 2008-38.
- Stiglitz, J. and P. Dasgupta. 1971. Differential taxation, public goods, and economic efficiency. *Review of Economic Studies* 38, 151-74.
- Sørensen, P. B. 1997. Public finance solutions to the European unemployment problem? *Economic Policy* 25, 223-64.
- Sørensen, P. B. 1999. Optimal tax progressivity in imperfect labour markets. *Labour Economics* 6, 435-52.
- Sørensen, P. B. 2007a. Can capital income taxes survive? And should they? *CESifo Economic Studies* 53, 172-228.
- Sørensen, P. B. 2007b. The theory of optimal taxation: what is the policy relevance? *International Tax and Public Finance* 14, 383-406.
- Tuomala, M. 1984. On the optimal income taxation: some further numerical results. *Journal of Public Economics* 23, 351-66.

Debat og kommentarer

Kampen om klimaet – historien om et topmøde der løb løbsk

Per Callesen

IMF for den Nordisk-Baltiske valgkreds, E-mail: pcallesen@imf.org

Journalist Per Meilstrup har skrevet en knaldroman om klimatopmødet i København, december 2009. Den er spændende og velskrevet. Den kommer også på en række punkter godt rundt i taktiske og storpolitiske spørgsmål. Han får formidlet stemningen af, hvad der var dramatisk, kaotisk og historisk. Og han får afdækket forhold, som selv for en der var der var tæt på igennem hele forløbet, stod dunkelt og uafklaret både under og efter topmødet. Den kan anbefales som spændende læsning om et vigtigt emne.

Journalistik fokuserer på konflikter, for det fænger. Forfatteren vælger at sætte interne danske konflikter i centrum. Hans udlægning fremmer enkeltheden i fremstillingen, men der skal gives det gode råd at læse bogen som et partsindlæg om en sag, hvorom der er meget forskellige opfattelser. Man kunne godt have ønsket en kritisk tilgang til holdbarheden af den formidlede version, herunder udvælgelsen af helte og skurke, under vurderingen af hvilken rolle faglighed betyder i ministerierne samt parallelliseringen af Statsministeriet med andre ministerier. Især kunne ønskes en bredere og dybere tilgang til de store internationale spørgsmål og Danmarks arbejde med disse. Der manglede ikke internationale konflikter, og uenigheder mellem danske ministerier er en dimension, som nok er mere interessant i danske medier end i resten af verden.

Eftertiden kommer formentlig til at leve videre med den skævhed i mediedækningen, der stammer fra ministeriekulturernes (og UNFCCC-sekretariatets) forskellige lyst og behov for at engagere medierne i deres samarbejdsforhold. Hvis ønsket er at komme dybere i den sag (hvor Meilstrups version på store strækninger er uigenkendelig), vil også det følgende i denne anmeldelse være ret skuffende. Måske kan anmeldelsen til gengæld bidrage lidt på de internationale dimensioner. De var faktisk interessante. Der er tale om personlige vurderinger.

Efter min vurdering har Meilstrup rigtigt godt fat i, hvad der skete på selve topmødet. Hvad der er opfanget af det lange forløb op til topmødet, er noget mindre vellykket. Bogens styrke er den minutiøse og dramatiske beskrivelse af handlingsforløbet i Bella Centeret, en god analyse af de centrale lande og institutioners ageren og holdninger, en god analyse af betydningen af den »danske tekst« og en interessant afdækning af den dramatiske lækage af et ældre diskussionsoplæg på en af konferencens første dage.

Svaghederne er fraværet af klimafaglige vinkler (hvad der skal til, for at en aftale har effekt), en nedspillet betydning af forhandlerkulturen i UNFCCC og en svag analyse af formandskabets internationale forarbejde før konferencen. De omfattende diskussioner og det nødvendige store danske arbejde for at sikre enighed i EU-systemet er stort set fraværende. Forarbejdet i forhold til f.eks. G20, Verdensbanken, OECD, IEA og IMF er slet ikke nævnt. Betydningen af MEF (Major Economies Forum) er nævnt, men ikke analyseret. De bilaterale kontakter til andre landes statschefer er nævnt, men uden mange detaljer.

Ambitionen for det danske formandskab var jo netop at brede klimadiskussionen ud og sikre bred international interesse og opbakning. Det skulle forebygge, at den meget brede og dybt økonomisk-politiske sag endnu engang gik i stå i den noget isole-rede og specialiserede boble af det internationale system, som udgøres af UNFCCC. De faglige og politiske udfordringer på klimaområdet rækker jo langt uden for det juridiske fokus i UNFCCC, og UNFCCC har ikke arbejdet med praktisk implementering. (Det blev som tilsigtet hverken statschefer eller finansministre, der bremsede fremskridt på klimatopmødet). Og der savnes en analyse af olieproducenternes interesser og ageren, i flere tilfælde vælger Meilstrup endda at fortolke deres lyrik som et idealistisk forsvar for lande, der er fattige og følsomme over for klimaforandringer.

Nogle af de mest interessante storpolitiske spørgsmål står stadig ubesvarede. Hvorfor skiftede BASIC-forhandlerne (fra Brasilien, Sydafrika, Indien, Kina) fra at være besværlige medspillere på det uformelle formøde den 1.-2. december til at være så afvisende under konferencen (indtil de sidste 12 timer)? Hvordan kunne der være så stor afstand mellem regeringschefernes positive meldinger om en politisk aftale i optakten til konferencen og forhandlernes modstand (indtil de sidste 12 timer)? Hvad var det, der om onsdagen fik BASIC-forhandlerne – på dette kritiske tidspunkt var det dem og ikke olielandene – til at angribe Statsministerens overtagelse af mødeledelsen og muligheden for at omdele en kompromistekst?

»Den danske tekst«. Den har Per Meilstrup ganske godt fat om betydningen af. Det var fatalt, at formandskabet aldrig fik omdelt et kompromisforslag. Forklaringen var til dels, at modstanden mod at se det var så stor. Men det skyldes også, at formandskabet efter svære diskussioner valgte at give efter for modstanden frem for, på trods af modstanden, at omdele et papir som formandskabets tanker på baggrund af tre års forarbejde og konsultationer. Efter min mening var det en fejl. Det ville have skabt stor ståhej, men ståhej kom der alligevel, det kunne ikke vælges fra. Meilstrup har ret i, at det kuldkastede strategien, og at formandskabet derefter måtte overgå til improvisering i Bellacentret. En lækage af en halvferdig tekst kan ikke ties ihjel, men omdeling af en egentlig tekst kan stoppe rygtedannelser.

Til gengæld er Meilstrup vist også tæt på at konkludere, at der ikke skulle være arbejdet med kompromismuligheder før selve topmødet. Hvis det er holdningen, må det un-

dre. Det havde været uansvarligt at have forberedt sig på et sådant topmøde uden et stykke papir og uden at have været i intensive drøftelser med nøgleparter om indholdet. Mange samtaler gør ingen aftaler. Det er først, når der kommer tekst på et stykke papir, man opdager, hvor meget arbejde der er i at forene mange parter med vidt forskellig kultur og holdninger. Forarbejdet med tekst var forsøget på at tilføre UNFCCC nogle af de nødvendige arbejdsredskaber, der får det internationale samarbejde til at virke på så mange andre felter.

Kun formandskabet havde kombineret en gennearbejdet vision for, hvad der faktisk kunne virke på klimaet i praksis, og hvad der ville være det rimelige kompromis. Problemet er, at der ikke kan sikres tilstrækkeligt ejerskab og indsigt i et sådant papir, om det fremlagdes på selve konferencen, endsige få dage inden dens afslutning. Det egentlige problem ift. regeringschefernes deltagelse blev, at de skulle forhandle en tekst, som de ikke havde set på forhånd, og som de ikke havde haft mulighed for at forberede sig på.

Forestillingen om at et kompromis alternativt kunne bygges op »nede fra« i UNFCCC-systemets arbejdsgrupper forekom og forekommer urealistisk. Internationalt samarbejde er altid hårdt og tidskrævende, men det virker ved at arbejde med papir, fokusere uenigheder på de vigtigste spørgsmål og stille kompromisforslag op i et overskueligt antal kantede parenteser. Sådan foregår det ikke i UNFCCC. Frem for at arbejde med kompromisforslag, så oplistede alle parter også taktisk motiverede alternativer efter hinanden i hundredevis af kantede parenteser. Frem for at nogen holder samling på, at de komplicerede elementer hænger sammen, arbejdes der med separate brudstykker.

Argumentet, der blev brugt imod teksten, var at det ville være in-transparent, hvis forhandlerne fik at se, hvad formandskabet havde tænkt. In-transparent at vise en tekst? Realiteten var nok snarere en mere eller mindre klart tænkt strategi om, at det var mere bekvemt at skyde en mulig aftale ned på procedurespørgsmål frem for at komme i reel dialog på de substantielle spørgsmål. Og i FN-systemet skal der ikke alene være enstemmighed mellem de 190 lande om indholdet af en beslutning, men også om man må omdele et forslag til overvejelse.

Lækagen? Meilstrup er tæt på at konkludere, at Danmark kun havde sig selv at takke for, at teksten ikke kom rundt og for, at et tidligt diskussionsoplæg blev lækket. Det er ikke rimeligt. Hvordan skulle nogen være i stand til at formulere et virkningsfuldt kompromis om noget så stort, kompliceret og politisk betændt, hvis man ikke kan vise nogen et stykke tekst og indhente vurderinger, bemærkninger og forslag? Hvordan skulle man sikre ejerskab til indholdet, hvis ingen måtte se noget på forhånd? Det ville være interessant med mere analyse af, hvorfor nogle ønskede at miskreditere formandskabet ved at lække teksten. Og interessant hvis Meilstrup også havde interviewet journalisten John

Vidal fra The Guardian om hans refleksioner på, hvad han egentlig fik sat i gang. Mest interessant havde det været med en analyse af, hvad formandskabet burde have gjort anderledes, da teksten blev lækket. Efter min mening burde formandskabet hurtigst muligt efter lækagen have omdelt sine egentlige og gennearbejdede overvejelser. Hvis det ikke kunne gøres som et forslag, så som et refleksionspapir.

Juridisk bindende aftale? Meilstrup fremhæver – helt korrekt – at dette spørgsmål og status for Kyoto-protokollen fra først til sidst var konfliktspørgsmål nummer ét. Først 10 timer inden Obama og kollegers ankomst var der grønt lys til at arbejde med en egentlig tekst – de 2^{1/2} side, der endte med at blive Copenhagen Accord. Af de 8 nattetimer, der var til gennemgangen med klimaministre og ambassadører, blev mellem 6 og 7 timer atter brugt på et spørgsmål, der ikke kunne føre nogen vegne. Af de 2^{1/2} side gik de egentlige 2 sider med substans derfor ubearbejdet direkte til regeringscheferne fredag morgen kl. 9.

Der er to formål med en juridisk bindende aftale. Det ene er teknisk og vedrører en række meget vigtige spørgsmål som f.eks. enighed om, hvordan man måler et ton CO₂ (mindre trivielt end man skulle tro). Det andet er imidlertid forestillingen om, at klimaproblemet skal løses ved, at verdens lande forpligter sig, juridisk, til at gennemføre nogle stramme målsætninger for deres aggregerede CO₂-udslip. Så kan de jo ikke rende fra løftet igen? Eller kan de? Og opnår man med den tilgang nogensinde at få de ønskede løfter?

Kyoto-protokollen er en sådan bindende aftale. Den dækker en – faldende – 30 pct. andel af verdens CO₂-udslip, i praksis EU, Japan og nogle få andre. USA ville ikke skrive under. Resten af verden er slet ikke med. En række underskrivere har ikke levet op til de påtagne mål. Den blev til i en tid, hvor OECD-landene udledte 80 pct. af verdens CO₂. Nu dækker de 50 pct. Om 30 år er andelen formentlig nede på 20 pct. De tre egentlige problemer ved begrebet juridisk bindende aftale er: (1) USA ville ikke skrive under på en sådan, med mindre Kina ville. Kina ville ikke, med henvisning til de vestlige landes forpligtelser, og Kina er næppe klar til den form for internationale forpligtelser. Hvis USA ikke ville, ville Japan ikke (længere). EU har i forvejen sit ETS-system, så her gør det ikke megen forskel. (2) Effekten af en juridisk binding afhænger af, at der også er nogle stærke sanktioner, hvis målene ikke opfyldes. Hvem vil både påtage sig stramme forpligtelser og samtidigt skrive under på stærke sanktioner? Hvem kan sælge en sådan underskrift til sit politiske bagland? (3) Kan det internationale system overhovedet håndhæve sådanne sanktioner (idet vi ser bort fra militære virkemidler)? På de store linier er en juridisk bindende aftale reelt en politisk aftale, der er lidt mere politisk bindende end hvis den »bare var en politisk aftale«, men denne erkendelse er ikke gjort i klimaforhandlingsuniverset.

Dette var og blev kernepunktet, også selv om målsætninger og finansiering var svært.

Udviklingslandene holder af Kyoto-protokollen, fordi den forpligter de rige lande, ikke dem selv. Og på trods af at verden ser anderledes ud end dengang, Kina, Korea og Kenya alle blev defineret som fattige udviklingslande. Ikke mindst USA finder omvendt, at hvis det kun handler om at forpligte de rigeste lande, så er det ikke en forhandling, men spil til ét mål. Så kan de rige lande aftale med sig selv, hvad de alligevel vil gøre, men hvor er begrundelsen for, at andre skal medvirke, hvis de ikke har forpligtelser?

Det var denne gordiske knude formandskabet søgte at løse ved at søge tilslutning til at lave en politisk aftale med to typer af forpligtelser, egentlige mål for de rige lande, og intensitetsmål, afvigelser fra en baseline eller nogle bestemte partielle politiktiltag for udviklingslandene. Så kunne man ønske fremtidige COP-formandskaber held og lykke med at omsætte en sådan aftale til en juridisk bindende tekst. Altså i stedet for at starte med juraen, så starte med substansen. Gevinsten ville være, at alle lande faktisk forpligtede sig, ikke til det samme, men med noget, der virkede og kunne virke med samme. Forpligtelsen ville ligge i national lovgivning. Måske – det var der mange signaler om – ville landene endda påtage sig mere ambitiøse mål, end de ville, hvis de også skulle skrive under på sanktioner. Muligheden lå i, at så mange regeringschefer på forhånd nikkede til tanken. Samtidigt var der en bred accept af, at forpligtelsernes karakter godt kunne være forskellig. Misvisende blev strategien i medierne beskrevet som »mindre ambitiøs«. I praksis blev det så besværligt, fordi UNFCCC-forhandlerne, godt presset af NGO'er i Bella Centeret, faldt tilbage til synspunktet om Kyoto-protokollen som omdrejningspunktet.

Var det en omhyggeligt planlagt strategi med henblik på at forhindre en sådan aftale? Næppe, og ikke nødvendigvis. Selv om de havde været intellektuelt parate, så skal BASIC-landenes muligheder for at påvirke forhandlere fra over 100 udviklingslande ikke overvurderes, også selv om et undervurderet stort antal klimafølsomme udviklingslande, herunder små ø-stater, afrikanere og lavtliggende lande, var meget konstruktive. Man skal heller ikke gøre sig illusioner om, at alle andre lande ankom til Bellacenteret med fuldt afklarede og indbyrdes enige delegationer. Problemet er snarere iboende og kulturelt forankret i forhandlerkulturen. Mange af forhandlerne har opereret i systemet i 17 år. De kender alle paragraffer i konventionen og Kyoto-protokollen forfra og bagfra. Og det er dygtige jurister. Nye formuleringer og metoder falder ikke i god jord hos så stærke tekstoperatører, og der var ikke lydhørhed over for indvendinger om sparsomheden i fremskridt ved blot at gentage velkendt tekst. Dertil kommer indsatsen fra en mindre, men kyndig, gruppe af olieproducenter, der ikke ønsker nogen aftale. Samtidig er der jo enstemmighedskrav i forhandlingerne, og en mindste fællesnævner-effekt drevet af følelsen af behov for indbyrdes solidaritet over for de rige lande. Det er ikke nemt at vende stemningen i en sådan forsamling i en uoverskuelig situation præget af stress og søvnmangel.

Pengeproblemer? Aftalte og afstemte reduktionsmål er den ene sag, en klimaaftale kan levere. Den anden er finansieringsbidrag til klimaindsats i fattigere lande og – undervurderet – et »governance«-system, der kan få en sådan finansiering til at virke efter hensigten. Pengespørgsmålet fik en dominerende pressedækning før konferencen. Det ser på overfladen enkelt ud: de rige lande skal presses til at betale, og når de ikke uden videre gør det, så må det være betalingsviljen, der blokerer fremskridt i forhandlingerne? Virkeligheden er noget mere kompliceret.

Hvordan sikrer man sig, at der kommer en klimaeffekt ud i den anden ende, når man finansierer en indsats? I hvert fald ikke uden at måle, rapportere og kontrollere. I sig selv kontroversielt for lande, der ikke har været vant til samarbejde som i EU og OECD. Et ellers så velkendt og harmløst begreb som »peer review« blev af nogle fortolket som skræmmende. Hvem skal administrere pengene og dermed sikre den ønskede klimaeffekt af indsatsen. UNFCCC? (Det synes klimaforhandlerne, hvorimod finansministerier og udenrigsministerier verden over er yderst skeptiske over for at opbygge et nyt bi-standslignende parallelsystem og skeptiske over for UNFCCC's manglende operationelle erfaring). Verdensbanken og regionale udviklingsbanker? (De har papir på at gøre den slags bedst, men har ikke et stort system klar endnu). Eller bilaterale nationale donorer i sammenhæng med den udviklingsbistand, det er svært at sondre indsatsen fra? (det synes mange udenrigsministerier). Kan man forestille sig – hvordan »løfter« nu en gang læses eller ej – at en klimafinansiering har en holdbar fremtid, med mindre finansministerier i de donerende lande føler sig trygge ved, at de også kan rapportere til deres parlamenter om gunstige effekter ved brug af skatteydernes penge?

Disse spørgsmål kommer Meilstrup ikke ind på, men de var afgørende i finansieringsdiskussionen. Der findes ikke et internationalt system i dag, der kan håndtere den nødvendige skala og effekt af klimafinansiering. Her havde det måske været muligt at komme længere i København, kompromiser var inden for rækkevidde (om end alle detaljer skabte kontroverser), men det blev overdøvet af de større spørgsmål. Det er bl.a. undervurderet, hvor vigtigt det vil være at få etableret et integreret statistiksystem for klimafinansiering og klimatiltag. Det findes ikke i dag, hvilket er en grobund for mistillid mellem parterne.

Alt i alt kom klimafinansierings spørgsmål ikke til at spille den helt store rolle i København. Ud over at »governance«-strukturen druknede i større spørgsmål og tidspres, så skyldes det, at de rige lande inkl. USA havde nogle store tal med til konferencen. Ikke ubetingede løfter, men store tal. Det var et smart træk, idet fokus dermed blev flyttet fra det mediepopulære pengespørgsmål til reduktionsforpligtelser, og ikke mindst hvad de store udviklingslande var klar til.

Må man tale om subsidierne? Pengespørgsmålet hænger en del sammen med en interessant alliance – strækkende over NGO'er og olieproducenter til det amerikanske se-

nat og Bjørn Lomborg – der er enige om at fortælle udviklingslandene, at det er meget dyrt at gøre noget ved klimaproblemet. Olieproducenterne forsømmer ingen lejlighed til at forklare, hvor vigtigt det er for væksten i de fattige lande, at de får købt og brugt en masse fossile brændstoffer. For udviklingslandene er det en praktisk støtte at have til forhandlingspositionen om at skaffe penge fra de rige lande.

Hvad er virkeligheden? Erkendt, vi løser ikke klimaproblemet (i fravær af endnu u-sete opfindelser) uden incitamenter til substitution (til energieffektivisering som isole-ring og vedvarende energi som vindmøller), som rækker pænt ud over, hvad rene markedspriser ville tilsige. Med den substitution skal verden formentlig give afkald på en eller flere pct. af den fremtidige velstandsstigning. Men vi kan komme meget langt bare ved at bruge markedspriser. Realiteten er – lad os kalde det »the second inconvenient truth« – at forbruget af fossile brændstoffer overalt i verden (EU og Japan er undtagelser), og især i mange udviklingslande, tilskyndes af massiv subsidiering og CO₂-frem-mende reguleringspolitik¹. Alene de direkte subsidier til at holde kunstigt lave priser – altså under verdensmarkedsprisen – er skønnet til mellem 500 og 1000 mia. dollars, altså 5-10 gange den årlige globale udviklingsbistand og langt over noget skøn for, hvad klimaindsatsen i U-landene behøver at koste. Med andre ord: den implicite globale skat på CO₂ er i dag stærkt negativ. Hertil kommer, måske endnu vigtigere, forbrugstilskyn-dende regulering som f.eks. betaling af energiforbrug over huslejen (hvilket indebærer en marginalpris på nul).

Denne subsidieringspolitik har intet at gøre med ønsker om økonomisk vækst. Tværtimod er det en forvridding, som fører til et ineffektivt overforbrug af fossile brændstoffer til skade for levestandarden (samt offentlige finanser og betalingsbalance, nogle lande bruger 5-10 pct. af nationalproduktet på sådanne subsidier), og det presser verdensmarkedspriserne på især olie i vejret. Dermed skal de politiske forhindringer for reformer på området (dvs. overgå fra at støtte produktet til at støtte indkomst) ikke undervurderes, men det er en stor og undervurderet del af problemet.

Meilstrup citerer en ukendt kilde for undren over, at Finansministeriet pressede på (hvilket er helt korrekt) for at få nævnt udfasning af disse subsidier i aftalen, da »det jo ikke er noget forhandlingerne har berørt«. Det afsnit i bogen fortæller noget om, hvor frakoblet klimaforhandlingerne er fra praktisk klimapolitik og fra andet internationalt samarbejde. Subsidieringen er genstand for omfattende analyser og rapporter fra både OECD, IEA, IMF og Verdensbanken og er ikke mere fremmede for politikerne, end at statscheferne fra samtlige G20-lande (dækker 75 pct. af klodens BNP og CO₂-udslip) få måneder før klimatopmødet i efteråret 2009 skrev under på at udfase disse subsidier, og det forsøger de internationale organisationer at følge op på.

1. Uddybet i »Økonomiske mekanismer i den globale klimapolitik«, Per Callesen, Nationalbankens Kvar-taloversigt, 3, oktober 2009.

Man bør spørge, om det giver mening at finansiere klimaprojekter i lande, der med den anden hånd bruger store dele af nationalproduktet på at holde kunstigt lave priser på brug af fossile brændstoffer frem for at give sin befolkning penge til selv at prioritere sit forbrug. Man savner erkendelsen af, at klimaproblemet er uløseligt med en fortsat praksis med subsidiering af fossile brændstoffer. Og tanken om at lokke nogen til at skrive under på nogle overordnede mål, fordi de ikke kan gennemskue de praktisk-politiske konsekvenser af det – og at de så alligevel skulle leve op til målet, når konsekvenserne opdages senere – er noget idealistisk.

Når alle de taktiske overvejelser og besværet med at vurdere, hvad de mange forhandlingsparter egentlig ønsker og kan gå med til, er gjort, kan fokus med rette inddrage, hvad der i praksis kan virke gunstigt på klimaet. Frem for at diskutere om en tekst er U-landsvenlig eller I-landsvenlig kunne den udsættes for et mere fagligt kriterium: er den klimavenlig? Modstanden mod at inddrage subsidieringen af fossile brændsler er kun det vigtigste eksempel på, hvor svært det var at holde fokus på substansen.

Kunne Danmark have gjort det anderledes og bedre? Givetvis. F.eks. skulle tekstlækagen være håndteret bedre, og der skulle være fundet en måde at bringe en kompromistekst på banen i tide. Alle både observatører og deltagere i processen har vist mange meninger om, hvad der kunne være gjort anderledes. (Sidstnævnte vil sikkert spare mere på de offentlige kommentarer end førstnævnte, og det havde været spændende at læse Meilstrups samlede vurdering af, hvad der burde have været gjort). Gælder det også på de helt store spørgsmål? Det er mere tvivlsomt. Jo den kunstneriske udførelse i Bella Centeret havde været bedre, hvis vi havde gjort ét af to.

Vi kunne have kørt det juridisk bindende spor til ende og fremlagt en tekst med en separat forpligtelse for de rige lande på Kyoto-protokollen og ingen forpligtelser for udviklingslandene (inkl. de hurtigt voksende). Det var der ikke kommet nogen aftale ud af, men det havde givet klapsalver i Bella Centeret og verden havde fået nogle andre skurke, bl.a. den formentlig mest klimavenlige amerikanske regering, vi får chancen for at forhandle med.

Vi kunne også have lænet os tilbage og ladet UNFCCC-systemets arbejdsgrupper fortsætte indtil fredag morgen kl. 9 med deres mange hundrede kantede parenteser. Det var der heller ikke kommet nogen aftale ud af, men Danmark var sluppet af kroge, skylden havde klarere været FN-systemets, og der var nok blevet vedtaget en køreplan for nye møder frem til COP16 i Mexico i december 2010.

I stedet valgte vi at bruge tre år på en omfattende faglig, politisk og international proces med henblik på at samle og gøde grunden for et egentligt kompromis inkl. de operationelle detaljer. Det fik vi ikke lov til at fremlægge og forhandle i slutfasen, og indtil de sidste 24 timer blev der talt procedure og ikke substans i Bella Centeret. Men vi fik for første gang i verdenshistorien samlet statschefer til en forhandling om klimaspørgsmål,

som vedrørte substansen, som de havde sat sig ind i (det havde de næppe gjort, hvis ikke de skulle til mødet) og som førte til en aftale med global dækning, hvor summarisk den end er, og hvor usikkert det endnu er, hvad den fører til i praksis.

Invitationen af statscheferne var en stor beslutning, og den var helt rigtig. Det var den eneste mulighed for et gennembrud. Det var noget rystende at opleve forargelsen blandt forhandlere fra tvivlsomme regimer over, at der var inviteret folkevalgte politikere til at forholde sig til sagen. Som Meilstrup helt korrekt beskriver, var intentionen, at statslederne skulle have forhandlet de egentlige politiske tal om målsætninger og finansiering, men så skulle de operationelle strukturer omkring det have været handlet af først, og det var ikke muligt. Aftalen mangler derfor dybde og det nødvendige afsæt for handling på det internationale plan. Men den kan stadig være en ramme om nationale handlingsplaner og en vifte af mere decentrale internationale samarbejdsinitiativer. Det er første gang, der er lavet en global klimaaftale i den forstand, at alle store lande har forpligtet sig.

Der er to parter, der grundlæggende afgør, om vi får løst klimaproblemet. Den ene er USA, og der er endnu ikke noget rigtigt godt svar på, hvor meget USA vil rykke. For USA var det i slutspillet næppe helt uvelkomment, at BASIC-landene insisterede på et lavere ambitionsniveau. (Meilstrup gør meget ud af dramatikken om Obamas møde med lederne fra BASIC-landene, men ikke hvad de egentlig forhandlede om: Det var reelt en svækkelse af de talmæssige målsætninger til gengæld for at opfylde USA (og EU's) krav om måling af, at lande lever op til de påtagne forpligtelser). De politiske problemer med klimaindsatsen i USA er velkendte, men til gengæld er potentialet for ret nemme energibesparelser så stort, at man ikke skal undervurdere, hvad det stigende fokus på sikker energiforsyning og en bedre privatøkonomi kan føre til. Den anden er de store udviklingslande. Spørgsmålet er, om de er parat til at tage globalt medansvar på sagen, og i København var der ikke et klart svar. Problemet er formentlig bl.a., at de bliver bedt om at tage globalt medansvar på et tidspunkt i udviklingen, hvor de stadig har rigeligt at gøre med hjemlige problemer og andre spørgsmål. Det er imidlertid helt afgørende, at de gør det, og også for dem, ikke mindst Kina, kan energisikkerheden måske være det, der giver udslaget. Kineserne ved godt, at kloden er for lille til, at de kan kopiere det amerikanske eksperiment.

Selvfølger blev man ikke, som engageret i sagen, i perlehumør af at læse Meilstrups – og det skal være gentaget – spændende og relevante bog. Det skyldes ikke, at han gør sig til part i danske ministeriediskussioner (og rammer disse skævt). Men derimod erkendelsen af, at Danmark i en del internationale sammenhænge udlægges som uforberedt, når vi var alt andet (så er vi tilbage til kompromisforslaget og de forgrenede internationale forberedelser igen).

Men man skal ikke undervurdere de afledte effekter af den proces, der har været omkring forberedelsen. Interessen er vakt hos en masse aktører verden over, der er fokuserede på praktiske resultater mere end medieomtalen. Den politiske fokus er skabt, selv om den økonomiske krise naturligt nok overskygger meget. Håbet må være, at der igangsættes en masse initiativer i national lovgivning, i private virksomheder og internationalt samarbejde i andre sammenhænge, som fører til praktiske resultater og gradvis overbeviser om, at det her faktisk kan lade sig gøre. Mulighederne falder især inden for tre områder: Reduktion af subsidierne til fossile brændsler, energipolitik drevet af flere hensyn end klima (energisikkerhed og økonomi) samt stadig flere praktisk gode erfaringer med klimafinansiering gennem de eksisterende organisationer som f.eks. Verdensbanken. Så må vi have den globale masterplanlægning til gode til en anden gang.

Vismænd, valutakurser, finanskrisen og euroen

Claus Vastrup

Institut for Økonomi, Aarhus Universitet, E-mail: cvastrup@econ.au.dk

Indledning

Med jævne mellemrum tager vismandsinstitutionen dansk valutakurspolitik og specielt Danmarks forhold til Euroen op til drøftelse. Det skete sidst i DØR (F2009), hvor der bl.a. redegøres for den faktuelle udvikling i Eurozonen og Danmark og for en ny empirisk undersøgelse af handelseffekter af Euroen og fastkurspolitik i almindelighed. På begge områder følger rapporten den nyere litteratur på området; samtidig er det for tjenstfuldt, at vismandsinstitutionen nu offentliggør en analyse af handelseffekter, som også indbefatter tal for Danmark. På grundlag af tal for 1995-2006 konkluderer vismændene, at ØMU'en har haft en væsentlig og positiv betydning for samhandlen, at virkningen af fastkurspolitikken for Danmark også har været positiv, og at de sidstnævnte virkninger ville have været større, hvis Danmark havde været medlem af Euroen.

Det er imidlertid en skuffelse, at finanskrisen 2007-2009 og valutakursteorien ikke har fået en mere adækvat behandling. Finanskrisen betragtes alene som en valuta- og ikke en generel finanskrisen, og dens betydning antages alene at vedrøre forskelle i rente mellem lande og Danmarks Nationalbanks intervention på valutamarkedet. Der er ikke mange andre spor af det største finansielle sammenbrud siden verdensdepressionen i 1930'erne. Noget tilsvarende gælder valutakurser og nyere økonomisk teori.

På det analytiske område har vismændene, Peter Birch Sørensen, Michael Rosholm, Hans Jørgen Whitta-Jacobsen og Erik Schrøder Amundsen således begrænset deres indsats til en fremstilling af og vurdering på grundlag af økonomisk litteratur, der stopper for 10-15 år siden. Det gælder både fremstillingen af årsager og virkninger af finanskriser og virkninger af valutakursændringer. På det grundlag konkluderer vismændene, at Danmark vil miste en positiv optionsværdi ved at deltage i ØMU-samarbejdet og herved fraskrive sig muligheden for at anvende valutakursen og dermed en selvstændig pengepolitik til stabiliseringspolitiske formål. Denne ulempe angives at være næsten på størrelse med de strukturelle gevinster ved en tilslutning til ØMU'en og således at reducere den danske økonomiske nettogevinst til næsten nul.

Mine indvendinger mod fremstillingen (og derfor også vismændenes konklusion) er først og fremmest, at vi i dette årtusinde har fået væsentlige erfaringer med finanskriser samt ny teoridannelse om både valutakurser og betydningen af finansielle markeder i valutarisk sammenhæng, som DØR (F2009) ikke tager i betragtning. Men fremstillingen ville også have vundet ved at tage nyere litteratur om valutaunioner og ØMUen som sådan i betragtning, se f.eks. Vastrup (2009).

Valutakurser

Vismændene (DØR, E2009, s. 198) anfører om valutakursen, at »Ved at devaluere sin valuta sænker man (...) øjeblikkeligt prisen på alle eksportvarer og hæver samtidig prisen på importerede varer«. Det er en opfattelse, som almindeligvis tillægges Friedman (1953) og andre af de tidlige forkæmpere for fleksible valutakurser. Siden er forudsætningen anvendt i modeller i Mundell-Flemming traditionen, der foruden vareimport og -eksport også medtager internationale kapitalbevægelser. Fra midten af 1960'erne og indtil først i 2000-tallet var denne modeltradition arbejdshesten i spørgsmål om international økonomisk politik og valutakurser. Forudsætningen om valutakursændringers pris-mæssige virkninger har sammen med en forudsætning om intern prisstivhed på kort sigt været helt afgørende for tidligere tiders og dermed vismændenes vurdering af de stabiliseringspolitiske virkninger af faste versus fleksible valutakurser.

Siden har den mikro- og dermed pristeoretisk baserede »New Open Economy Macroeconomy« delvis taget over, og herved er man blevet opmærksom på, f.eks. Devereaux og Engel (2003), at det for så vidt angår prisstivhed i udenrigshandelen er nødvendigt at skelne mellem på den ene side Producer Currency Pricing (PCP), hvor eksportør-producenten forud fastsætter sin pris på eksportmarkedet i egen valuta lige-som ved salg til hjemmemarkedet, og på den anden side Local Currency Pricing (LCP – tidligere også kaldet Pricing to Market), hvor eksportør-producenten forud fastsætter sin pris på eksportmarkedet i importlandets valuta modsat hans pris ved salg til hjemmemarkedet, som stadig forud fastsættes i hjemlandets valuta. Som følge af denne (beret-tigede) interesse for prisfastsættelsen forudsætter de generelle ligevægtsmodeller med profitmaksimerende virksomheder og nyttemaksimerende husholdninger, at der ikke er fuldkommen konkurrence på varemarkederne, og at det er muligt for producenterne at prisfastsætte forskelligt og derfor at prisdiskriminere mellem eksport- og hjemmemarkeder.

Kun i tilfælde af PCP vil en ændret valutakurs medføre ændrede importpriser i importlandet og derfor en ændring i dets efterspørgsel efter import og dermed en ændring i eksport, produktion og beskæftigelse i eksportlandet. I en situation med LCP vil importpriser og efterspørgsel i importlandet være uændrede efter en ændring i valutakursen, og det samme vil derfor gælde eksportlandets eksport, produktion og beskæftigelse. Virkningen af en valutakursændring vil her blive opfanget og inddækket af en ændring i virksomhedernes dækningsbidrag (mark up), idet produktionen som følge af de monopolistiske forhold og de forud fastsatte priser er bestemt af efterspørgselen (inden for visse grænser).

Den umiddelbare forskel mellem de to forudsætninger om prisfastsættelse viser sig således i valutakursændringernes gennemslag (»pass through«) i priserne. Med PCP er der øjeblikkeligt og fuldt gennemslag, mens det ikke er tilfældet med LCP. Det følger

heraf og det anførte citat, at DØR (F 2009) uden diskussion forudsætter PCP og dermed hurtige realøkonomiske konsekvenser af valutakursændringer. Der er imidlertid gode argumenter – i form af den lille producent i forhold til det store marked – for at antage, at LCP er den relevante forudsætning for en væsentlig del af dansk eksport. Distinktionen mellem PCP og LCP er derfor både væsentlig og relevant i en diskussion af mulige valutakursregimer i Danmark.

Forskelle i prisfastsættelsen medfører, at et givet stød resulterer i forskelle for både efterspørgsel, produktion, beskæftigelse og i de endogent bestemte udsving i en fleksibel valutakurs. Variabiliteten i de nævnte variable er forskellig i tilfælde af PCP henholdsvis LCP, fordi valutakursens »expenditure switching« virkning kun er til stede i tilfælde af PCP. Derfor er der i en sammenligning mellem virkningerne af PCP henholdsvis LCP også forskelle både i samvariationen mellem forbruget på tværs af lande og i samvariationen mellem produktionen på tværs af lande og dermed den indbyrdes konjunkturpåvirkning. Analytisk er distinktionen således væsentlig mellem den tidligere enerådende forudsætning om PCP og den nye og til tider plausible forudsætning om LCP.

Det gælder også pengepolitikens virkninger, hvor PCP i både import- og eksporthandel gennem valutakursændringer isolerer et land fra udefrakommende stød, og den optimale pengepolitik derfor alene skal tage hensyn til indenlandske stød og søge at stabilisere de indenlandske producentpriser og det reale outputgab. Med PCP er der således kvalitativ overensstemmelse mellem pengepolitik i en lukket økonomi og en åben økonomi, Benigno og Benigno (2003). I tilfælde af LCP skal den hjemlige pengepolitik derimod tage hensyn til priser fastsat også på baggrund af udenlandske vilkår og den forventede valutakurs. Den optimale pengepolitik bliver derfor delvis eksternt orienteret og giver selv med formelt variable valutakurser en vis indbyrdes afhængighed mellem ind- og udland og fremkalder naturligt et ønske om koordination mellem de involverede pengepolitikker. Samtidig vil endogent bestemte ændringer i pengepolitikken på tværs af lande som følge af den indbyrdes afhængighed alt andet lige tendere mod at være positivt korrelerede og valutakursvariabiliteten derfor lavere med LCP end i tilfælde af PCP. DØR (F2009) omtaler i sin fremstilling af dansk pengepolitik i en tænkt situation med variabel valutakurs ikke denne problemstilling, men henviser alene til, at pengepolitikken i en sådan situation skal være internt orienteret og forudsætter derfor (igen) PCP.

For at illustrere betydningen af LCP og dermed den eksterne orientering af pengepolitikken har flere forfattere vist, at hvis alle varer handles, og importerede og hjemmeproducerede varer indgår med samme vægt i forbrugernes nyttefunktion, er det (i en to-lande model) med LCP optimalt med en fast valutakurs. Forudsætningerne medfører, at de velfærdsbaserede og dermed optimale pengepolitikker (begge) skal stabilisere for-

bruget af en identisk kurv af varer. Alle pengepolitikker skal derfor reagere på identisk vis i tilfælde af både efterspørgselsstød og symmetriske og asymmetriske produktivitetstød. Valutakurserne vil som følge heraf være stabile og dermed faste, se f.eks. Corsetti (2008 b).

Som følge af de forskellige virkninger af PCP henholdsvis LCP er det selvfølgelig interessant, hvilke forudsætningerne om prisfastsættelsen i udenrigshandelen der i praksis er opfyldt, og derfor hvilke karakteristika som er forbundet med de to former for prisfastsættelse. Som forventet er der stærke indicier for LCP, hvis eksportpriserne er faktureret i importlandets valuta, Giovannini (1988), Fitzgerald og Haller (2009). Men ellers er man empirisk indtil videre især henvist til undersøgelser af graden af valutakursændringers gennemslag (»pass through«) på import- og forbrugerpriser, hvor LCP modsat PCP som nævnt medfører ingen eller en lav gennemslagskraft, som dog også kan have andre årsager. Flere empiriske undersøgelser viser, at valutakursændringers gennemslag langt fra er øjeblikkelige og fuldstændige, og derfor at LCP snarere end PCP i hvert fald i en vis udstrækning må formodes at være den relevante forudsætning. Litteraturen diskuterer også, om f.eks. det empirisk beregnede omfang af »pass through« har været faldende over tid.

Undersøgelser i forhold til det store amerikanske valutaområde tyder tillige på, at i forholdet mellem små eksportlande og store importlande vil en asymmetrisk prisfastsættelse i det store lands valuta være fremherskende, Atkeson og Burstein (2008). Det er i overensstemmelse med empiri, f.eks. Goldberg og Tille (2008), der viser, at periferilande (som Danmark) og centerlande (som Eurozonen) fakturerer i centerlandets valuta, og dertil en formodning om, at prisfastsættelsen sker i den fakturerede valuta. Det er således rimeligt at antage, at periferilandes eksport overvejende prisfastsættes som LCP og centerlandes eksport som PCP. Den nævnte asymmetri i prisfastsættelsen kan muligvis forklares ved, at eksportør-producenter i lande med relativt ustabil pengepolitik foretrækker prisfastsættelse i valutaer fra importlande med relativt stabile pengepolitiske forhold, Devereaux, Engel og Storgaard (2004), og at store lande (som Eurozonen) pengepolitisk ofte er mere stabile end deres periferilande (som Danmark).

Der er således gode grunde til at antage, at lande i og omkring Euro-zonen prisfastsætter i samme valuta (Euroen). Med en sådan forudsætning vil en ændring i periferilandenenes (f.eks. Danmarks) valutakurs i forhold til centerlandet ikke ændre de relative priser og således heller ikke have »expenditure switching« effekter, der ændrer efterspørgselen efter periferilandets eksport og den heraf afledte produktion og beskæftigelse. Derfor vil valutakursændringer ikke være et relevant instrument i tilfælde af stød til den eksterne eller interne efterspørgsel og en fast valutakurs derfor et naturligt valg for periferilandet. Periferilande vil dog foretrække en variabel valutakurs i tilfælde af et aktuelt asymmetrisk stød til produktiviteten, mens et forventet stød af tilsvarende karakter med

intertemporalt optimerende husholdninger vil være at sammenligne med et aktuelt efterspørgselsstød og det optimale valg derfor igen en fast valutakurs. Fra en stabiliseringspolitisk synsvinkel vil valget af valutakursregime herefter som sædvanligt afhænge af, hvilken type af stød der er fremherskende, og hvilke fordele f.eks. om troværdighed der i øvrigt måtte være forbundet med det ene og ikke det andet regime.

Da valget af valutakurs for givet form for stød med den traditionelle forudsætning om PCP er forskellig fra ovennævnte vil valget af valutakursregime i visse situationer – afhængig af fordelingen af stød – også afhænge af om eksportør-producentens prisfastsættelse er LCP eller PCP. Det gælder for nogle forudsætninger også uanset denne fordeling af stød. Som nævnt er det nemlig i den mikro- og pristeoretisk baserede valutakursteori muligt at konkretisere, under hvilke betingelserne og LCP en fast valutakurs altid vil være optimal uanset stødenes karakter. I den udstrækning disse betingelser ikke er opfyldt, kan det som også nævnt vises, at LCP på grund af den afledte eksterne dimension i den optimale pengepolitik vil reducere den optimale grad af variabilitet i valutakursen. LDC vil derfor (altid) reducere og eventuelt helt eliminere de traditionelle stabiliseringspolitiske omkostninger ved et fastkurs regime (og muligvis mere end det). For valget af valutakursregime er det derfor væsentligt om eksportør-producenterne er PCP- eller LCP prisfastsættere.

Ved at se bort fra nyere valutakursteori og en stærk formodning om forekomsten af LCP overvurderer DØR (F2009) eventuelle stabiliseringspolitiske omkostninger ved, at Danmark opgiver en eller anden grad af selvstændighed i sin valutakurspolitik for at deltage i ØMU-samarbejdet. Sådanne omkostninger skal – i det omfang de eksisterer – som bekendt sammenholdes med de strukturelle gevinster ved deltagelse i den fælles mønt. For en nærmere opgørelse af de stabiliseringspolitiske omkostninger og gevinster ved valget mellem en fast og variabel valutakurs har den nyere pris- og mikrobaserede valutakursteori klarlagt, hvilke forhold der skal belyses empirisk.

Finansielle markeder

Når produktionsstrukturen ikke er medtaget ovenfor, skyldes det, at de omtalte modeller forudsætter fuld risikospredning hos husholdningerne og derfor et komplet sæt af (finansielle) markeder. På tilsvarende vis medfører forudsætningerne i en del modeller om valutakurser, at saldoen på betalingsbalancens løbende poster (og dermed de udenlandske nettoaktiver) antages at være identisk lig med nul. Det tilsiger forsigtighed med ensidig anvendelse af de konkrete resultater, idet man herved bortser fra forskelle i finansielle muligheder og problemer vedrørende betalingsbalance og udlandsgæld i valget mellem valutakursregimer.

Den manglende specifikation af finansielle markeder gjorde sig også gældende i Friedmans (1953) argumentation til fordel for fleksible valutakurser. Efterfølgende

medtog Mundell-Flemming modellen internationale kapitalbevægelser, hvilket førte til en bedre forståelse af den økonomiske politik og valutakursernes rolle i et åbent samfund. Både de analytiske begrænsninger, de teoriehistoriske erfaringer og erkendelsen af de finansielle forholds konkrete betydning i forbindelse med f.eks. finanskrisen 2007-09 tilsiger derfor, at de finansielle markeder inddrages i modeller og overvejelser om valutakurspolitik, herunder om Danmarks tilknytning til Euroen. Det sker kun i begrænset omfang i DØR (F2009).

Spørgsmålet i dette afsnit går derfor videre end ovenfor, hvor det især drejede sig om (skepsis vedrørende) valutakursernes stødabsorberende evne. I forhold til de finansielle markeder er det et spørgsmål, om valutakurser formindsker husholdningernes risikospredning, og om de skaber og/eller videresender stød. Fra en penge- og valutakurspolitisk vinkel er det også et væsentligt (og slet ikke omtalt) spørgsmål, om pengepolitikken skal korrigere prisbobler og andre anomalier i finansielle priser, herunder valutakurser, som måtte afvige fra ligevægtskurser bestemt af reale faktorer.

Nyere mikro- og pristeoretiske modeller for åbne økonomier har også bidraget til en belysning af finansielle markeders betydning for valget af valutakursregime, idet det der er vist (og indlysende), at den finansielle struktur i form af det konkrete sæt af finansielle markeder har betydning for agenternes porteføljeallokering og deres velværd. Inspirationen stammer fra Mundell (1973), der anfører, at i et regime med uafhængige nationale valutaer (og derfor med et begrænset sæt af finansielle markeder) skal et negativt stød absorberes inden for det ramte land, mens virkningerne af et tilsvarende stød i en valutaunion (med en fælles mønt og derfor et større sæt af finansielle markeder) kan forsikres i og derfor spredes til andre medlemslande. En fælles mønt med integrerede finansielle markeder tillader således, at et (internt) betalingsbalanceunderskud opstår og finansieres ved et finansielt træk på andre medlemslande.

Herefter har først Neumeyer (1998) forholdvis generelt og siden Ching og Devereux (2003) i en model med PCP vist, at der både er omkostninger og fordele for et land ved at deltage i en valutaunion i tilfælde af landespecifikke produktivitetsstød. Nettofordelen kan være både negativ og positiv, men dog således, at en forøget grad af prisstivhed forøger nettotabet/reducerer nettogevinsten ved en valutaunion. Det er samtidig sådan, at en forøgelse af variabiliteten i de landespecifikke stød ikke nødvendigvis forøger nettotabet/reducerer nettogevinsten ved en valutaunion. Det skyldes, at den fleksible valutakurs stødabsorberende effekter (i en PCP-model) måske mere end opvejes af den bedre risikospredning gennem de finansielle markeder i en valutaunion. Det centrale budskab er således, at asymmetriske produktivitetsstød som følge af en større grad af finansiell integration i en valutaunion ikke nødvendigvis er ensbetydende med nettofordele ved en fleksibel valutakurs. Nettofordelen kan være i favør af en valutaunion.

Et lignende synspunktet er udviklet af Devereux (2004) for en situation med stød til både efterspørgsel (præferencer) og produktivitet. Det antages her, at den fleksible valutakurs stabiliserer den produktion, der ellers ville afhænge af den indenlandske prisstivhed, på et niveau svarende til den naturlige produktion. Sidstnævnte kan samtidig ændres som følge af stød til produktiviteten. På trods af lønstivhed (eller PCP-stivhed) eliminerer valutakursændringer virkningerne af prisstivheden og absorberer dermed de produktivitetsskud, som forekommer i modellen. Imidlertid er den naturlige produktion ikke efficient som følge af fraværet af et komplet sæt af internationale finansielle markeder. Det skyldes ikke fraværet af internationale kapitalbevægelser, men at produktionen ikke reagerer tilstrækkelig kraftigt på stødene til efterspørgselen, idet den fuldt variable valutakurs i sådanne situationer fuldstændig og udelukkende lukker produktionsgabene, uanset at efterspørgselen og præferencerne er ændret.

Større fasthed i valutakurs vil have procykliske virkninger og således forøge variabiliteten i produktionen, hvilket kan give en bedre sammensætning i forbruget af indenlandsk og udenlandsk producerede varer og derfor højere velfærd. Flexibiliteten i valutakursen og traditionel absorption af produktivitetsskud kan derfor være uønsket, mens en fast valutakurs målt i velfærdstermer kan være optimal, hvilket tenderer at være tilfældet, når elasticiteten i arbejdsudbuddet er høj. En fast valutakurs vil i hvert fald være at foretrække, hvis arbejdets grænseulyst er konstant. Et tilsvarende resultat om faste valutakurser som stødabsorberende i velfærdsmæssig forstand kan opstilles for tilfælde med fuldt fleksible priser.

På baggrund af sådanne analytiske resultater om betydningen af integration og risikospredning via de finansielle markeder har der i ØMU-litteraturen i de senere år været en del studier af de finansielle markeder i Eurozonen, se f.eks. Lane (2008 a og b). Betydningen af risikospredning via de finansielle markeder omtales ikke i DØR (F2009), hvor den finansielle integration alene omtales og måles ved et rentespænd, som henføres til en større eller mindre grad af usikkerhed om valutakursen. Vismændene forbigår således væsentlige synspunkter om virkningen af bl.a. asymmetriske stød og muligheden for velfærdsgevinster som følge af finansiell integration i en monetær union. I det realistiske tilfælde med fravær af et komplet sæt af markeder kan både en forbedret finansiell struktur og en fast valutakurs i en monetær union forøge graden af risikodeling i forhold til en situation med separate finansielle markeder og en fleksibel valutakurs.

Med mellemrum opstår der både finansielle kriser og valutakriser. Det er mindre interessant, om finansielle kriser defineres inklusiv eller eksklusiv valutakriser, men det er bemærkelsesværdigt, at DØR (F2009, s. 259-77) først lader sin definition af finansielle kriser omfatte både finansielle kriser og valutakriser og herefter næsten udelukkende omtaler valutakriser og rentespænd. Det er imidlertid et interessant og relevant

spørgsmål, om en større grad af finansiel integration som følge af dansk medlemskab af ØMU'en ville have reduceret de problemer, som Danmark løb ind i som følge af f.eks. den amerikanske initierede finanskrisen 2007-09, der i sin oprindelse og størstedelen af sit forløb var en likviditetskrise og ikke en valutakrise.

Til illustration af samspillet mellem finans- og valutakriser har Aghion m.fl. (2001) opstillet en model af en lille, åben økonomi, hvor virksomheder både er kreditrationerede, har gæld (også) i udenlandsk mønt og eksporterer til forud fastsatte, hjemlige priser (PCP). I situationer med en variabel valutakurs kan en sådan økonomi som resultat af flere forskellige former for stød have i en recession med en devalueret valuta. På trods af en forbedret konkurrenceevne i eksporten skyldes den mindre produktion en forøget gæld i indenlandsk mønt, en mindre egenkapital og som følge af kreditrationeringen også et mindre kapitalapparat og derfor en mindre produktion. En fast valutakurs kan i sådanne situationer stabilisere værdien af gælden i indenlandsk mønt og dermed produktionen, men det kræver selvfølgelig, at de relevante stød ikke er så store, at den resulterende valutaudstrømning udtømmer landets valutaeserver og fremtvinger en devaluering.

Modellen er en såkaldt tredje generationsmodel for valutakriser og har været anvendt til at belyse den finansielle krise i Asien 1997-98 og hvorfor faste valutakurs i en sådan situation havde været at foretrække frem for deprecieringer af fleksible valutakurser. Spørgsmålet er, om forudsætningerne er relevante for Danmark f.eks. under den internationale finanskrisen 2007-09, og hvad et Euro-medlemskab i så fald havde betydet? Der er næppe tvivl om, at en del danske virksomheder var kreditrationeret under krisen, og at udenlandsk gæld ville have reduceret deres egenkapital og dermed produktion i tilfælde af en dansk devaluering. Når Danmark ikke blev tvunget til at devaluere, skyldes det, at valutaeserven netop var tilstrækkelig til at modstå den spekulation mod den danske krone, som opstod i oktober-november 2008. Men da det ikke var muligt for den danske stat og nationalbank på det pågældende tidspunkt at låne på de internationale kapitalmarkeder, ville en lidt større eller længerevarende spekulation mod den danske krone have tømt valutabeholdningen og fremtvunget en devaluering med mulighed for en kraftigere recession til følge – med mindre ECB og dermed Eurozonen havde støttet kronen, om de faktiske forhold se Bernstein (2010).

Dansk deltagelse i Euro-samarbejdet ville i en sådan situation have forhindret, at de finansielle (og reale) stød havde ført til en devaluering, idet den danske kronkurs ville have været sikret af den fælles mønt. En anvendelse af Aghion m.fl. (2001) model på den danske situation under finanskrisen i 2008 kan således illustrere nogle af fordelene også for produktion og beskæftigelse ved deltagelse i Euro-samarbejdet. Selvom DØR (F2009) nævner den islandske finanskrisen tages finanskrisens betydning for dansk deltagelse i Eurosamarbejdet ikke op til behandling, muligvis på grund af den (for) snævre

afgræsning af finansielle kriser til alene at være valutakriser med tilhørende rentespænd. Det er formentlig også baggrunden for, at DØR (F2009, s. 260) fejlagtigt angiver, at lande med fleksible valutakurser ikke kan blive ramt af valutakriser defineret som en kraftig depreciering over kort tid. Aghion m.fl. (2001) viser, at det godt kan være tilfældet f.eks. i forbindelse med en finanskris.

Konklusion

DØR (F 2009) sammenligner Danmarks nuværende fastkurspolitik med dansk medlemskab af Eurozonen ved at omskrive differencen mellem de almindelige udtryk for den forventede nytte af henholdsvis det ene og det andet valutakursregime til et udtryk i bl.a. optionsværdien af at kunne devaluere med det nuværende valutakursregime. Det er selvfølgelig tilladeligt, men fremstillingsmæssigt næppe hensigtsmæssigt, idet mange læsere formentlig ikke vil opfatte, at optionsværdien ved at kunne devaluere kan være både positiv og negativ – og ikke kun positiv, således som også vismændene synes at forudsætte. Hvorfor ikke vedblive med at sammenligne de forventede værdier af nytten ved forskellige valutakursregimer?

Mine kritik af fremstillingen i DØR (F2009) vedrører dog især den manglende anvendelse af nyere analyser og empiri om valutakurser. Som følge heraf overvurderer vismændene formentlig valutakursændringers muligheder for at stabilisere velfærd, betalingsbalance, produktion og beskæftigelse især for så vidt angår periferilande (inkl. Danmark) i forhold til Eurozonen og USA. I nyere litteratur findes i hvert fald gode analytiske og empiriske begrundelser for at antage, at valutakursændringers stødabsorberende virkninger er begrænsede i forhold til den tidligere almindelige opfattelse. Den nyere opfattelse tager vismændene ikke stilling til og undlader at skelne mellem »producer currency pricing« (PCP) og »local currency pricing« (LCP).

På tilsvarende vis undlader vismændene at inddrage nyere litteraturs påvisning af en større grad af risikospredning og derfor højere velfærd som følge af en større grad af finansiel integration mellem lande i en valutaunion, herunder ØMU'en, end i lande med selvstændige valutaer. De finansielle og valutariske aspekters betydning for valget af valutakursregime i tilfælde af finansielle kriser er også underbelyst og nogle helt udeladt. De forhold, som ikke er behandlet eller underbelyst, taler til fordel for Danmarks deltagelse i Euro-samarbejdet.

Selvom kapitlet om handelseffekter som følge af Danmarks eventuelle deltagelse i Euro-samarbejdet er et vægtigt bidrag til debatten, er DØR (F2009) ikke det bedst mulige grundlag for en diskussion af Danmarks stilling i forhold til Euroen. For mange undladelser og manglende anvendelse af nyere litteratur og empiri medfører, at vismændenes fremstillingen af de økonomiske forhold er mindre fyldestgørende end nødvendigt.

Litteratur

- Atkeson, A. og A. Burstein. 2008. Pricing-to-Market, Trade Costs, and International Relative Prices. *American Economic Review* 98:1998-2031.
- Aghion, P., P. Bacchetta og A. Banerjee. 2001. Currency crises and monetary policy in an economy with credit constraints. *European Economic Review* 45:1121-50.
- Benigno, G. og Benigno, P. 2003. Price stability in open economy. *Review of Economic Studies* 70:743-64.
- Bernstein, N. 2010. Outside the euro in the financial crisis – the Danish experience. Speech 22 March 2010. www.natioanlbanken.dk.
- Ching, S. og M. B. Devereux. 2003. Mundell Revisited: a Simple Approach to the Costs and Benefits of a Single Currency Area. *Review of International Economics* 11: 674-91.
- Corsetti, G. 2008 a. Openness and the case for flexible exchange rates. *Research in Economics* 60:1-21.
- Corsetti, G. 2008 b. A modern reconsideration of the theory of Optimal Currency Areas. *Economic Papers* No. 308. European Commission.
- Corsetti, G. og P. Pesenti. 2005. International dimensions of optimal monetary policy. *Journal of Monetary Economics* 52:281-305.
- Devereux, M. B. 2004. Should the exchange rate be a shock absorber? *Journal of International Economics* 62:359-77.
- Devereaux, M. B. of C. Engel. 2003. Monetary policy in the open economy revisited: exchange rate flexibility and price setting behavior. *Review of Economic Studies* 70:765-83.
- Devereaux, M. B., C. Engel og P. Storgaard. 2004. Endogenous exchange rate pass-through when nominal prices are set in advance. *Journal of International Economics* 63:263-91.
- DØR. F2009. Danmark og euroen. Kapitel II i *Dansk Økonomi. Forår 2009*. København p.153-329.
- Fitzgerald, D. og S. Haller. 2009. Pricing-to-Market; Evidence From Producer Prices. Arbejdsrapport.
- Friedman, M. 1953. The case for flexible exchange rates. I Milton Friedman (red.), *Essays in Positive Economics*. Chicago.
- Giovannini, A. 1988. Exchange Rates and Traded Goods Prices. *Journal of International Economics* 24:45-68.
- Goldberg, L.S. og C. Tille. 2008. Macroeconomic Interdependence and the International Role of the Dollar. *NBER Working Paper* 13820.
- Lane P. R. 2008 a. The macroeconomics of financial integration: A European perspective. *IIIS Discussion Paper* 265.
- Lane, P. R. 2008 b. EMU and financial integration. *IIIS Discussion Paper* 272.
- Neumeier, P. A. 1998. Currencies and the Allocation of Risk: The Welfare Effects of a Monetary Union. *American Economic Review* 88:246-59.
- Mundell, R. A. 1973. Uncommon Arguments for Common Currencies. I H. G. Johnson og A. Swoboda (red.), *The economics of common currencies*. Cambridge, MA.
- Vastrup, C. 2009. Økonomiske konsekvenser af ØMU'en. *DIIS Working Paper* 2009:1.

A Treatise on Money – om debatten mellem Hayek og Keynes

Finn Olesen

Institut for Miljø- og Erhvervsøkonomi, Syddansk Universitet, E-mail: finn@sam.sdu.dk

*SUMMARY: Throughout all of his writings Keynes tried to make economic theory to become better in accordance with empirical evidence. In doing so, he often challenged the economic orthodoxy of his days. For many years he struggled to be able to deliver the right theoretical answer as to why the macro economy could be put in a position of disequilibrium for a rather long period of time with many involuntary unemployed workers. As we all know, he finally succeeded in doing so with the publication of *The General Theory* in 1936. On the road to this new macroeconomic understanding Keynes had hoped that his *A Treatise on Money* from 1930 could deliver some important theoretical elements that made him able to explain the problems of the economic crisis in a more convincing way than that of the neoclassical mainstream. However, his macroeconomic reasoning in *Treatise* was still rather old-fashioned as his analysis was restricted to the understanding of the quantity theory of money. Be this as it may, following the publication of *Treatise* an interesting debate unfold itself between Hayek and Keynes. In many ways this debate illustrates the differences in opinion between two theoretical »modes of thought«. In what follows the focus of the present article would be partly on highlighting some of the important elements of *Treatise* and partly on the debate of Hayek and Keynes.*

1. Indledning

I den økonomiske teorihistorie er det af indlysende årsager oftest *The General Theory* fra 1936, som underkastes en belysning. Men Keynes' forfatterskab indeholder naturligvis også andre interessante bidrag, der i sin samtid bidrog til den økonomisk teoretiske debat og til den generelle teoriudvikling. Og på sin vis kan hele Keynes forfatterskab ses som en langstrakt kamp for at få skabt en ny og mere relevant forståelse af især de makroøkonomiske fænomener, jf. fremstillingen i Olesen (2010). Som bekendt kom dette gennembrud, teoretisk såvel som også metodologisk, endelig i 1936; jf. eksempelvis Olesen (2007). En mulig vigtig brik i denne intellektuelle befrielseskamp fra den herskende ortodoksi er *A Treatise on Money*, der efter flere år med en fortløbende udvidelses- og revisionsproces endelig så dagens lys den 31. oktober

Jeg skal takke Henrik Preben Perregaard og tidsskriftets referees for kommentarer til en tidligere version af denne artikel.

1930. Og under den lange skriveproces havde Keynes ændret sin opfattelse så meget, at han ikke rigtigt var helt tilfreds med indholdet af bogen ved dens udgivelse.¹ Som han skriver i forordet til bogen:

The ideas with which I have finished up are widely different from those with which I began. The result is, I am afraid, that there is a good deal in this book which represents the process of getting rid of the ideas which I used to have and of finding my way to those which I now have; Keynes [1930 vol. 1:xvii].

Værket består af to bind: en teoretisk del med titlen *The pure theory of money* og en mere praktisk orienteret del *The applied theory of money*, som på basis af især en empirisk belysning af samtidige britiske som amerikanske monetære forhold har til hensigt at give et kvantitativt orienteret fundament for den teoretiske udledning, som Keynes foretager i det første bind. Af interesse for Keynes' gradvise udvikling af sin makroøkonomiske forståelse er især fremstillingen i Bog 3 og 4 i det første bind; jf. Olesen (2009a). Det er heri, at Keynes præsenterer sine såkaldte »fundamental equations«, der udgør værkets økonomiske forståelsesramme. En kort præsentation af denne følger i det efterfølgende afsnit. Som den internationalt anerkendte økonom Keynes var, havde samtiden set frem til udgivelsen af *Treatise* med stor interesse. Den blev derfor også efter udgivelsen anmeldt i alle de førende tidsskrifter. Især et af disse bidrag står centralt i belysningen af *Treatise*. Værket blev mest grundigt anmeldt og kritisk diskuteret af en anden af samtidens markante økonomer: Friedrich Hayek. Derudover gav denne anledning til en berømt debat mellem Hayek og Keynes, der på sin vis repræsenterer en »kamp« mellem to alternative økonomiske verdensanskuelser (»modes of thought«). I afsnit 3 forsøges denne debat mellem de to gengivet og perspektiveret. Endelig afsluttes artiklen med nogle afrundende og opsamlende kommentarer.

2. Om *Treatise*

I forordet til værket pointerer Keynes, at det alt afgørende formål med bogen er at forsøge at opnå en bedre økonomisk teoretisk forståelse af en makroøkonomi i bevægelse:

1. Som Moggridge (1992, s. 484) sammenfattende skriver om *Treatise*; »He produced a Janus-like book: it looked back to his Cambridge inheritance; it looked forward to some of the concerns of *The General Theory of Employment, Interest and Money*«. For et indblik i hvorledes skriveprocessen forløb for Keynes, kan der henvises til Skidelsky (1992, s. 281-85). Et bud på, hvordan og hvorfor Keynes ændrede sin teoretiske opfattelse af de monetære sammenhænge fra sin *A Tract on Monetary Reform* fra 1923 til arbejdet med *Treatise* endelig var færdigt, er eksempelvis givet i Hirai (2007). For et ældre bidrag omhandler den samme transformationsproces hos Keynes byggende især på vidnesbyrd fra Richard Kahn og Joan Robinson kan der henvises til Lambert (1969). Endelig gives der i Moggridge (1973), og senere også i Patinkin (1976), en detaljeret belysning af, hvorledes Keynes gradvist fra 1931 og fremefter udviklede konturerne til det, der i 1936 skulle blive til *The General Theory*.

My object has been to find a method which is useful in describing, not merely the characteristics of static equilibrium, but also those of disequilibrium, and to discover the dynamical laws governing the passage of a monetary system from one position of equilibrium to another; Keynes [1930 vol. 1:xvii].

Det er altså de økonomiske processer i en dynamisk makroøkonomi, som Keynes ønsker at belyse. Belært af den økonomiske historie vidste Keynes om nogen, at det makroøkonomiske univers er af en foranderlig karakter. Og også selvfølgelig, at det makroøkonomiske udfald kunne være præget af disharmoni og mangel på stabilitet. En tilfredsstillende makroøkonomisk ligevægt kunne man ikke være sikker på ville etablere sig automatisk. Det var i høj grad en af de vigtige erfaringer, der fulgte i kølvandet på afslutningen af den 1. Verdenskrig. Økonomiske krisetendenser havde været et af 1920ernes kendetegn. Det er derfor værkets centrale opgave inden for rammerne af et reformuleret kvantitetsteoretisk set up at forsøge at besvare de følgende to spørgsmål: hvad bestemmer/påvirker prisniveauet i samfundet, og hvilke muligheder har centralbanken for at påvirke denne proces? Til besvarelse af disse spørgsmål er især fremstillingen i Bog 3 og 4 i Treatise's 1. bind central, hvorfor der i det følgende gives en kortfattet belysning heraf.

Således får vi i kapitel 10 fastlagt opgaven for pengepolitikken. Den er givet ved et ønske om at styre udviklingen i prisniveauet eller inflationstakten. Målet for pengepolitikken er dermed ikke, som det senere blev tilfældet for analysen i *The General Theory*, at forsøge at styre udviklingen i reale størrelser (i en likviditetspræferenceforståelse: gennem ændringer i renteniveauet at forsøge at påvirke investeringsdannelsen og dermed den effektive efterspørgsel og den samlede produktion i samfundet). Opfattelsen hos Keynes er på dette punkt således fuldstændig i god samklang med samtidens herskende mainstream forståelse.

Byggende på kvantitetsteoretiske antagelser foretager Keynes dernæst nogle ligningsmæssige manipulationer, som giver den følgende ligning og konklusion.

$$P = E/O + (I - S)/O \quad (1)$$

hvor P er det generelle prisindeks, E/O er fortjenesten pr. produceret enhed (earnings/output), og I og S er henholdsvis investeringsomfanget og den samlede opsparing. Af ligningen ses det, at et stabilt prisniveau (og en stabil inflationstakt) er betinget af to forhold:

It follows that the stability of the purchasing power of money involves the two conditions – that efficiency earnings should be constant and that the cost of new investment should be equal to the volume of current savings; op. cit., s. 122-23.

Denne kvantitetsteoretiske forståelse bekræftes blot yderligere, når Keynes senere konkluderende skriver:

If we start from a position of equilibrium, then – provided that efficiency earnings are stable – the condition for the continued stability of price levels is that the total volume of money should vary in such a way that the effect of the corresponding volume of bank lending on the market rate of interest is to keep the value of new investment at an equality with current savings; op. cit., s. 197.

Og en sådan konklusion vil næppe nogen af Keynes' samtidige kolleger være uenig med ham i. Det er klassisk lærdom. Men en sådan form for stabilitet kan kun forventes at findes i en situation, hvor den statiske ligevægt er opretholdt, påpeger Keynes noget mere moderne. I en mere uligevægtig makroøkonomisk situation holder ækvivalensen mellem opsparingsomfanget og de foretagne investeringer ikke nødvendigvis automatisk.² I en sådan situation er kausaliteten bagved prisudviklingen anderledes, må Keynes' opfattelse være. Her er det en anden forklaring end den traditionelle kvantitetsteoretiske, der må tage over. Derfor må denne forståelsesramme reformuleres. Og det er netop, hvad Keynes med sin fremstilling har til hensigt at gøre. Men det er næppe en opgave, som Keynes kan lykkes med inden for den makroøkonomiske forståelsesramme, der er givet ved Treatise, idet accepten af de fundamentale ligninger hos Keynes faktisk forudsætter en økonomisk situation med fuld beskæftigelse. Og dette forhold er netop en fundamental svaghed ved værket. Med en sådan forståelsesramme får uligevægtsprægede forløb ikke rigtig lov til at udfolde sig og udspille sine fulde konsekvenser. En mere relevant belysning af den økonomiske krises opståen og fastholdelse kræver en fundamentalt anderledes makroøkonomisk modelramme. Som bekendt så en sådan først dagens lys i kapitel 3 i *The General Theory*.

Kapitel 11 omhandler mere præcist de betingelser, som skal være gældende for at sikre en makroøkonomisk ligevægt fremkaldt og fastholdt. Heri får vi at vide, at entreprenorerne reagerer på deres profitudvikling. Er profitten positiv (og med profitten mener Keynes den overnormale profit, som ligger udover en normal aflønning af kapitalen), vil de investere og udvide deres produktion. Og omvendt indskrænkes udbuddet af varer, hvis profitten er negativ. Skal en ligevægtssituation etableres kræver dette, helt klassisk, at profitten er nul. I en sådan situation er der samtidig naturligvis også en ækvivalens mellem investeringsomfanget og den faktiske opsparing i samfundet. I det efterfølgende kapitel 12 uddyber og nuancerer Keynes forståelsen af sammenhængene mellem investeringsomfanget og den samlede opsparing. Eksempelvis får vi det føl-

2. »If the volume of saving becomes unequal to the cost of new investment ... then the fundamental price levels can depart from their equilibrium values without any change having occurred in the quantity of money or in the velocities of circulation«; Keynes (1930 vol., s. 132).

gende at vide om tilpasningsprocessens forløb, der terminalt følgende en klassisk argumentation repræsenterende validiteten af Say's lov skal sikre den makroøkonomiske ligevægts etablering:

That saving can occur without any corresponding investment is obvious, if we consider what happens when an individual refrains from spending his money income on consumption. It does not matter what he does with the surplus – whether he deposits it at his bank, pays off a loan or buys a house or a security – provided it is not accompanied by an additional act of investment by an entrepreneur. There is now in the market one purchaser less for consumption goods, with the result that their prices fall. This fall of prices increases the purchasing power of the money incomes of the rest of the community and they are able, therefore, to increase their consumption by the amount which the saver has forgone, whilst spending the same amount of money as before ... An act of saving by an individual may result either in increased investment or in increased consumption by the individuals who make up the rest of the community; op. cit., s. 156 and 158.

Og dette er som nævnt en ræsonnementskæde, som er helt i overensstemmelse med den klassiske tænkning. En effektiv prismekanisme har til opgave at sikre, at ligevægten i økonomien genetableres.

Fastholdelse af en sådan harmonisk ligevægt stiller også krav til banksektoren, påpeger Keynes. Denne sektor er en helt essentiel spiller i den makroøkonomiske tilpasningsproces. Som Skidelsky (1992, s. 317) har påpeget: »The Treatise is a glorification of the banker as economic therapist«. Bankerne skal nemlig: »so regulate its rate of lending that the value of investment is equal to savings«; Keynes (1930, vol., s. 137). Er dette ikke tilfældet, får vi konjunktursvingninger, idet der sammenfattende konkluderes: »Booms and slumps are simply the expression of the results of an oscillation of the terms of credit about their equilibrium position«; op. cit., s. 165.

Samtidig skelner Keynes, med inspiration fra Wicksell, mellem en markeds- og en naturlig rente.³ Er de to renter sammenfaldende, er økonomien i en stabil makroøkonomisk ligevægt. Er de forskellige fra hinanden, er økonomien i bevægelse; enten i form af, at investeringsomfanget overstiger henholdsvis er mindre end opsparingen, hvorfor profitudviklingen i denne situation er positiv henholdsvis negativ. Og da produktion tager tid, er det i realiteten de forventede frem for de faktiske profitgevinster henholdsvis -tab, som entreprenorerne reagerer på, når de planlægger at øge eller sænke deres investeringsomfang, påpeges det. En vis form for underliggende dynamik kan dog godt tænkes i en sådan makroøkonomisk set stabil situation. Enkelte entreprenører kan have en positiv profit, mens andre kan have en negativ. Keynes pointerer

3. Som påpeget i det næste afsnit er det dog diskutabelt, i hvilken udstrækning Keynes egentlig har forstået den wicksellianske argumentation korrekt.

derfor, at det er den aggregerede profit, som skal være nul for at sikre en makroøkonomisk ligevægt fremkaldt og fastholdt med et stabilt prisniveau. Men økonomien kan godt, i hvert tilfælde på det kortere sigt, befinde sig uden for en sådan ligevægt.

Overordnet set er den økonomiske tænkning i *Treatise* derfor fundamentalt set forankret i en ligevægtsanalyse helt i tråd med samtidens mainstream tænkning. Endnu havde Keynes altså ikke påbegyndt den teoretiske som metodologiske nytænkning, som skulle komme til at bryde igennem med udgivelsen af *The General Theory*, om end fundamentet til en metodologisk revolution egentlig havde været til stede længe. Som megen Keynes-forskning vil fremhæve, skete det med udgivelsen af *A Treatise on Probability* i 1921.⁴ Derfor indtager forventningsdannelsen i *Treatise* ikke nær den helt afgørende hovedrolle, som denne efterfølgende skulle få for Keynes' makroøkonomiske tænkning. Nok kunne det økonomiske system i *Treatise* være i bevægelse og uden for en makroøkonomisk stabil ligevægt, men det var slet ikke med den samme grad af foranderlighed, som Keynes senere i 1936 argumenterede for kunne være tilfældet. Som påpeget af Olesen (2009) fulgte en sådan betydelig mere markant dynamisk grad af foranderlighed over tid i det økonomiske system af introduktionen af en fundamental type af usikkerhed. Det er først med inddragelsen af sådanne forventninger, at det økonomiske system hos Keynes bliver at ligne med et åbent og socialt foranderligt system, hvori de enkelte agenter planlægger, beslutter og handler under usik-

4. Generelt er det dog kendetegnende for Keynes' forfatterskab, at det ikke indeholder nogen samlet metodologisk redegørelse, men hist og her kan der ses nogle brudstykker til en overordnet metodologi. Eksempelvis henvises der i denne sammenhæng ofte til hans brevsveksling med Roy Harrod i sommeren 1938, hvor Keynes blandt andet var optaget af et review af Tinbergens tidlige økonomiske pionerarbejde; jf. CW (XIV, s. 295-301). Heri skriver Keynes blandt andet til Harrod: »It seems to me that economics is a branch of logic, a way of thinking ... *Progress* in economics consists almost entirely in a progressive improvement in the choice of models ... Economics is a science in terms of models joined to the art of choosing models which are relevant to the contemporary world. It is compelled to be this, because, unlike the typical natural science, the material to which it is applied is, in too many respects, not homogeneous through time ... In chemistry and physics and other natural sciences the object of experiment is to fill in the actual values of the various quantities and factors appearing in an equation or formula; and the work when done is once and for all. In economics that is not the case, and to convert a model into a quantitative formula is to destroy its usefulness as an instrument of thought«; CW (XIV, s. 296 and 299). Men Keynes' grundlæggende metodologiske opfattelse var tilsyneladende på plads længe før end 1938 (som tidligere nævnt formentlig med udgangspunkt i *Treatise on Probability*). Således kan man læse det følgende, som er hentet fra hans introduktion til *Cambridge Economic Handbooks* i 1922: »The theory of economics does not furnish a body of settled conclusions immediately applicable to policy. It is a method rather than a doctrine, an apparatus of the mind, a technique of thinking, which helps its possessor to draw correct conclusions. It is not difficult in the sense in which mathematical and scientific techniques are difficult; but the fact that its modes of expression are much less precise than these, renders decidedly difficult the task of conveying it correctly to the minds of learners«. At dette syn på økonomi må have haft en ganske fremtrædende rolle i Keynes' metodologiske forståelse, fremgår alene deraf, at det ovenstående citat gentages ordret i årbøgerne fra 1923 og 1928-36; jf. CW (XII, s. 856-59). Men nogen afgørende betydning synes denne opfattelse ikke at have haft for fremstillingen i *Treatise*. Måske også derfor var Keynes ikke rigtig tilfreds med værket; jævnfør de gjorte bemærkninger herom i det indledende afsnit.

kerhed på basis af forventninger til en principiel ukendt fremtid; jf. Chick (2003) og Chick og Dow (2005).

I de efterfølgende kapitler 16-18 forsøger Keynes at afdække, hvad der kan forstyrre den makroøkonomiske prisstabilitet. Her får han fastlagt tre årsager. Et uligevægtspræget forløb i økonomien kan sættes i gang af: (1) monetære forhold, som øver indflydelse på »the effective supply of money for income purposes«, altså forstyrrelser fremkaldt af ændringer i M^S henholdsvis i M^D ; (2) ændringer i investeringsomfanget som følge af en afvigelse mellem markedsrenten og den naturlige rente; samt (3) ændringer i den samlede produktion. Af disse forhold er det nok især fluktuationer i investeringsomfanget, som er af væsentlig betydning. Keynes påpeger således:

When there is disequilibrium between savings and investment, this is much more often due to fluctuations in the rate of investment than to sudden changes in the rate of savings, which is, in normal circumstances, of a fairly steady character; Keynes [1930: vol., s. 85].

En anerkendelse af, at det er en forskel mellem investeringsomfanget og den samlede opsparing, som skal forklare ændringerne i priserne (inflationstakten), synes at være i god overensstemmelse med argumentationen hos Wicksell. Således påpeger eksempelvis Stein (1969), Fischer (1972) og Flaschel (2000), at i Keynes-Wicksell inspirerede monetære modeller med en indbygget antagelse om fuld beskæftigelse kan en steady state inflationstakt forklares ved en permanent situation med overskudsefterspørgsel på varemarkederne.⁵ I en sådan modelmæssig forståelsesramme har vi derfor, at »money is superneutral, where therefore inflation is solely due to excessive monetary growth«; Flaschel (2000, s. 453). En sådan opfattelse af penge og monetære forholds betydning er dermed i klar modstrid med argumentationen i *The General Theory*. På dette punkt har Keynes virkelig foretaget et teoretisk nybrud i forhold til den makroøkonomiske forståelsesramme, der kommer til udtryk i *Treatise*. Et af kerneudsagnene i *The General Theory* er netop at penge ikke er neutrale hverken på kort eller på lang sigt. En ændring af de monetære forhold i økonomien har dermed klare realøkonomiske konsekvenser.⁶

Kendetegnende for Keynes' gennemgang i *Treatise* af, hvad der kan forstyrre stabiliteten i økonomien, er, at uligevægtsforklaringerne ikke rigtigt synes at komme ud af den spændetrøje, som den kvantitetsteoretiske forankring påfører den makroøkonomiske analyse. Dertil er »the credit cycle«-forklaring for dominerende i den keyneske

5. Som Stein (1969, s. 170) formulerer det: »According to Walras' Law, the excess demand for goods must equal the excess flow supply of real balances, since the other markets are assumed to be in equilibrium«.

6. Som Fontane (2007, s. 58) sammenfatter det: »Aggregate demand matters for the long-run level of output and employment in a country. This is why money and monetary policy really matter«.

fortælling. Og så alligevel måske ikke. Senere hen i bind 2 forsøger Keynes i sin kriseforklaring nemlig tilsyneladende at give denne et pift af noget nyt, der virker mere teoretisk overbevisende, og som sætter et spørgsmålstejn ved den klassiske teoris accept af eksistensen af tilpasningsprocesser i økonomien, som automatisk fører til genetableringen af den makroøkonomiske ligevægt. Han skriver således det følgende om, hvorfor en svag krisetendens nærmest pr. automatik kan forstærkes af iboende kræfter i det økonomiske system:

When the rate of saving is exceeding the rate of investment, goods are coming forward at a faster rate than that at which they can be purchased at prices corresponding to their costs of production. Now it might be that a very moderate fall in prices would restore equilibrium ... But the foregoing argument proves that the fall in prices must be substantial and must proceed until it is accompanied by a fall of production. Owing, however, to the length of the period of production, the falling off in the rate of input will begin by aggravating matters. For it will cause the rates of employment and of earnings to fall off before the rate of output declines. Thus the stock of working capital will fall away and the disparity between saving and investment will be still further aggravated ... We have, therefore, an adequate theoretical explanation of the violence and rapidity of the slump once it has begun; op. cit., s. 131.⁷

Som tidligere påpeget var det Keynes' hensigt med Treatise at forsøge at besvare to helt centrale spørgsmål. Dels hvad der bestemmer/påvirker prisniveauet i samfundet, og dels hvilke muligheder centralbanken har for at kunne påvirke denne proces? Om svarene herpå skriver Keynes konkluderende og sammenfattende i bind 2, at:

We have claimed to prove in this treatise that the price level of output depends on the level of money income relatively to efficiency, on the volume of investment ... relatively to saving, and on the »bearish« or »bullish« sentiment of capitalists relatively to the supply of savings deposits available in the banking system. We have claimed, further, that the banking system can control the supply of savings deposits ... that it can by the terms of credit influence to any required extent the volume of investment ... and that the indirect effects of its influence on the value and the volume of investment determine the money offers which entrepreneurs make to the factors of production ... But we have not claimed that the banking system can produce any of these effects instantaneously; or that it can be expected always to foresee the operation of non-monetary factors in time to take measures in advance to counteract their influence on prices ... or that a central bank, which is member of an international system, can preserve domestic stability irrespective of the behavior of other central banks; op. cit., s. 309-10.

7. Jf. også Amadeo og Dutt (1994) der påpeger, at det ikke er givet, at prisfald (deflation) nødvendigvis vil bidrage til at genetablere ligevægten på varemarkederne i en sådan modelmæssig forståelsesramme.

Ifølge Keynes er banksystemet dermed bedre gearret til at forsøge at fastholde en given makroøkonomisk ligevægt, end systemet er til at forsøge at få bragt en given uligevægt tvunget hen imod et mere stabilt og stationært udfald i økonomien. Det er derfor nok forståeligt, at der kan være en måske berettet utilfredshed i befolkningen med styringen af de monetære forhold i økonomien. Men som Keynes påpeger: »We do badly; but we do not know how to do better«; op. cit., s. 364. I hvert tilfælde indtil Treatise så dagens lys, må man antage, at Keynes' opfattelse er. Og så måske alligevel ikke. Afslutningsvist påpeger Keynes nemlig, at kvantitetsteorien som den afgørende teoretiske referenceramme er »difficult to comprehend«. Så som man allerede kan læse ud af forordet til værket, så er Keynes et eller andet sted godt klar over, at opgaven med at formulere den afgørende nye teoretiske forståelsesramme er mislykkedes for ham. Hans reformulering af kvantitetsteorien leverer så at sige ikke varen: den tilfredsstillende forklaring af, hvorfor en makroøkonomi kan låses fast i en tilsyneladende langvarig uligevægtssituation. Kvantitetsteorien er den forkerte teoretiske referenceramme. Og jf. fremstillingen i Moggridge (1973) indså Keynes først for alvor dette efter kritik fra de unge økonomer i Cambridge (The Cirkus), der i perioden januar til maj 1931 dissekerede indholdet i Treatise med den hensigt at afdække værkets teoretiske mangler og fejltagelser. At Keynes således hurtigt efter udgivelsen af Treatise var på vej til en ny erkendelse ses eksempelvis af indholdet i nogle forelæsningsrækker – The Harris Foundation lecture series – som han afholdt i Chicago i juni 1931. Titlen for forelæsningsrækken var *An Economic Analysis of Unemployment*. Heri er fokus i de makroøkonomiske tilpasningsprocesser mere markant på en mængdemæssig frem for på en prismæssig tilpasning; jf. CW (XIII, s. 343-67). Der var altså brug for et alternativ. Og dette blev først givet i The General Theory i kapitel 3, hvor Keynes' nye makromodel så dagens lys i form af det effektive efterspørgselsprincip. Med et sådant nyt makroteoretisk set up kunne der endelig gives en mere tilfredsstillende forklaring på eksistensen af en underbeskæftigelsesligevægt og den ufrivillige arbejdsløshedstype.

3. Samtidens reaktion og debatten mellem Hayek og Keynes

Om end Keynes' navn havde været internationalt berømt siden han i 1919 udgav The Economic Consequences of the Peace, og han efterfølgende havde udgivet flere værker, var der alligevel blevet set frem til udgivelsen af Treatise med spænding. Ville dette værk, som det banebrydende nye værk om en moderne monetær forståelse, give Keynes den endegyldige berømmelse også på det økonomisk teoretiske plan? Blandt de mange reaktioner på Treatise er den kendtes nok den, som Friedrich Hayek stod for. I sin samtid kom kritikken fra netop Hayek nok til at stå stærkest. Og i debatten mellem Hayek og Keynes kom fronterne mellem to alternative forklaringsmodeller – eller såkaldte »modes of thought«; jf. Dow (1996) – til at stå skarpt stillet op over for hinan-

den. I det følgende vil debatten mellem de to kombattanter, og hvad deraf fulgte, blive forsøgt belyst.

Først følger dog et afsnit, der har til hensigt kort at skitsere Hayeks teoretiske udgangspunkt, som dannede baggrund for hans vurdering og kritik af Treatise. Ifølge Caldwell (2004) var anvendelsen af ligevægtsteori et must for Hayek allerede fra 1928.⁸ Sine synspunkter m.h.t. hvorledes makroøkonomiske uligevægtsforløb burde belyses, redegjorde Hayek mere omfattende for i *Monetary Theory and Trade Cycle* fra 1933 (oprindeligt publiceret på tysk i 1929). Ifølge Caldwell tog Hayek afstand fra en simpel kvantitetsteoretisk forklaring af »the trade cycle«. Grundlæggende burde de følgende fire forbehold respekteres ved studier af konjunktursvingninger: (1) forklaringer måtte bygge på teori; (2) den fremsatte teori måtte være konsistent med standard ligevægtsteori; (3) teorien måtte indeholde monetære forhold til forklaring af krisens opståen; og (4) teorien måtte fokusere på de økonomiske konsekvenser, der følger af ændrede relative prisforhold.⁹ Og især det fjerde. forbehold fik Hayek til at tage afstand fra simple kvantitetsteoretiske forklaringer og analyser, som påpeget af Caldwell (2004 s. 164). Det, som var vigtigt for Hayek, var at forsøge at forklare ændringerne i produktionsstrukturen i samfundet, der følger af konjunkturopp- som konjunkturedgange.¹⁰ Dermed var Hayek, som påpeget af Caldwell (2010), helt klar over, at markedssystemet ikke altid fungerer perfekt. Eksistensen af konjunktursvingninger var for ham en naturlig konsekvens af »a creditusing economy«. For Hayek var recessioner nødvendige for på ny at bringe det økonomiske system tilbage til sin ligevægtsposition. I øvrigt skal Hayeks konjunkturteori ikke her nærmere beskrives, idet der fyldigt er redegjort for denne i Perregaard (2009).

I august 1931 kommer det første af i alt tre fyldige indlæg fra Hayeks side. Som repræsentant for den anden økonomiske skole i England (LSE) var en reaktion herfra på Cambridge økonomen Keynes' værk blot at forvente. I det første bidrag dels roses Keynes for sin indsats dels påpeges det, at værket ikke fremstår som et fuldstændigt sammenhængende helt værk, hvorfor »its appearance cannot be said to have that definitive significance which at one time was expected of it«. Derfor er det også, efter

8. Jf. Caldwell (2004, s. 156) der citerer Hayek for den følgende opfattelse af ligevægtsteoriens relevans: »its field of application is identical with that of economic theory, since only with its assistance is it possible to give a summary depiction of the very great number of different tendencies of movement which are operative in every economic system at every point in time«.

9. Som Hayek udtrykte det selv: »But general price changes are no essential feature of a monetary theory of the Trade Cycle; they are not only inessential, but they would be completely irrelevant if only they were completely »general« – that is, if they affected all prices at the same time and in the same proportion. The point of real interest to Trade Cycle theory is the existence of certain deviations in individual price-relations occurring because changes in the volume of money appear at certain individual points«.

10. Idet: »Changes in the structure of production are caused by changes in relative prices. As such, a monetary trade cycle theory must be capable of showing the origins and effects of cycle-producing relative price changes«; Caldwell (2004, s. 324).

Hayeks mening: »unfair to regard it as anything else but experimental ... to publish the work in what is avowedly an unfinished state«; Hayek (1931, s. 270 og 271). Og Hayek er straks helt med på, hvad der er det centrale i Keynes' fremstilling. Det er i Bog 3 og 4, at Keynes kommer med sit nye bud på en monetær forståelsesramme. Om denne del af Treatise siger Hayek (1931, s. 271):

It is here that all the force and all the weakness of the argument are concentrated, and it is here that the really original work is set forth. And here, unfortunately, the exposition is so difficult, unsystematic, and obscure, that it is extremely difficult for the fellow economist who disagrees with the conclusions to demonstrate the exact point of disagreement and to state his objections.

I forlængelse heraf gives der efterfølgende en kritik af Keynes' produktions- og profitforståelse. Efter Hayeks mening er begge behandlet for forenklet i Treatise. Synspunkterne herpå burde have været mere nuanceret, eksempelvis hvad angår tidsdimensionen i produktionsprocessen (»the successive stages of production«: processerne fra råvarer over halvfabrikata til færdigt produkt tager tid). Som Hayek (1931, s. 276) sammenfatter det:

profits cannot be explained as the difference between expenditure in one period and receipts in the *same* period or a period of equal length because *the result of the expenditure in one period will very often have to be sold in a period which is either longer or shorter than the first period.*

Ligeledes kritiseres Keynes for ikke at være gået videre i sin analyse end blot at stille sig tilfreds med at fokusere på begrebet »total profits«. Derudover har Hayek ligeledes et problem med præcist at forstå, hvad Keynes mener med begreberne »investment« og »capital«. Men et fokus på investeringsdannelsen er selvfølgelig væsentligt for forståelsen af de økonomiske processer; det er Hayek helt med på. Desværre synes han også, at Keynes på dette punkt i sin analyse af investeringernes fluktuationer byder på en forsimplet og mangelfuld forklaringsgrad. Denne analyse udgør: »the central and most obscure theme of the book«, og i denne sammenhæng har Keynes helt ignoreret det teoretiske fundament bagved Wicksells teori, som med held kunne have været inddraget, påpeges det. Hos Wicksell indtager banksektoren en afgørende rolle til bestemmelse af det faktiske investeringsomfang. Er denne sektors fastsatte udlånsrente forskellig fra investeringsprojektets reale afkast (den naturlige rente), er det samlede investeringsomfang forskelligt fra den samlede opsparing. Er udlånsrenten lavere end den naturlige rente, vokser investeringsomfanget. Og omvendt mindskes dette, hvis udlånsrenten er højere end den naturlige rente. Hvis investeringsomfanget ikke præ-

cist matcher den samlede opsparing, er økonomien i en uligevægtssituation med enten inflationære eller deflationære tendenser til følge. Og hos Wicksell er der ingen garanti for, at den af banksektoren manipulerbare udlånsrente netop skulle være identisk med det reale investeringsafkast.¹¹

I sin argumentation på investeringsdannelsen er Hayek her helt tro mod den traditionelle neoklassiske opfattelse om, at det kun er en ændring i de relative prisforhold i økonomien, som får agenterne til at reagere og til at ændre på deres adfærd:

But, surely, an explanation of the causes which make investment more or less attractive should form the basis of any analysis of investment. *Such an explanation can, however, only be reached by a close analysis of the factors determining the relative prices of capital goods in the different successive stages of production* – for the difference between these prices is the only source of interest. But this is excluded from the outset if only total profits are made the aim of investigation; op. cit., s. 277.

Afslutningsvist forholder Hayek sig kritisk til de mange manipulationer, som Keynes foretager på basis af sine fundamentale ligninger. Her kan det være svært at følge Keynes, og udbyttet af læserens anstrengelser på dette område er ganske begrænset, er synspunktet hos Hayek.

Alt i alt er Hayek, som man måske kunne have forventet, særdeles kritisk over for Keynes' vej væk fra den herskende ortodoksi. Han synes helt klart ikke om Keynes' bestræbelser på at reformulere (og dynamisere) den kvantitetsteoretiske forståelsesramme. Og dette var en anmeldelse fra en kollega, til hvem Keynes tidligere havde haft et ret venskabeligt forhold, ligesom begge i enighed før havde påpeget betydningen af forventningernes og usikkerhedens rolle for den økonomiske beslutningstagen, som sårede Keynes.¹² Det kan man læse af den sammenfattende kommentar, som Keynes skrev på sin kopi af Hayeks review:

Hayek has not read my book with that measure of »good will« which an author is entitled to expect of a reader. Until he can do so, he will not see what I mean or know whether I am right. He evidently has a passion which leads him to pick on me, but I am left wondering what this passion is; CW [vol. XIII, s. 243].¹³

11. Jf. Fontana (2007 s. 46): »This equilibrium case is the ideal outcome, though, as Wicksell argues, it is a rare case. The reason being that it is only by chance that the banking system is able to set the interest rate on loans (r) equal to the real return on the production process, the natural rate (ρ) ... The difference between investments (I) and savings (S) is thus a positive function of the gap or disequilibrium between the natural (ρ) and the loan (r) rates ... Importantly, Wicksell assumes that the normal case of the economy is characterized by full employment, with the result then that an increase in investment spending by entrepreneurs causes an increase in the price level«.

12. Ifølge Caldwell (2010) så Hayek som ung studerende Keynes som en af sine helte på baggrund af den analyse som Keynes havde givet af Versailles-freden i *The Economic Consequences of the Peace*. Med

I sit svar til Hayek er Keynes da også sikker på, at forskellen i opfattelsen imellem de to ikke bygger på hverken nogle »verbal contradictions« eller påståede »terminological inconsistencies«. Forskellen mellem de to er af en mere fundamental karakter, hvorfor Keynes da heller ikke indledningsvist tøver med at fastslå, at han med sine kommentarer ganske sikkert kun vil have »little or no progress towards convincing Dr. Hayek«; Keynes (1931, s. 387). Grundlæggende har de to en forskellig opfattelse af opsparings/investerings ligevægten. Dermed var deres kriseforklaringer, som Skidelsky (2010, s. 5) ser det, hinandens kausale modsætninger: »For Hayek the depression was threatened by »investment running ahead of saving«; for Keynes by saving running ahead of investment«. Hayek, i modsætning til Keynes, ifølge Keynes selv, mener, at en sådan ækvivalens mellem de to størrelser altid vil være til stede, når det makroøkonomiske system har tilpasset sig:

My analysis is quite different from this; as it necessarily must be, since, in my view, saving and investment (as I define them) can get out of gear without any change on the part of the banking system from »neutrality« as defined by Dr. Hayek, merely as a result of the public changing their rate of saving or the entrepreneurs changing their rate of investment, there being no automatic mechanism in the economic system (as Dr. Hayek's view would imply there must be) to keep the two rates equal, provided that the effective quantity of money is unchanged; op. cit., s. 393.

Keynes kan derfor kun konkludere, at »Dr. Hayek has seriously misapprehended the character of my conclusions«. Og så skal der også siges tak for sidst med en såret mands følelser i centrum. Hayek havde netop fået udgivet sin *Prices and Production*, og om dette værk skriver Keynes:

The book, as it stands, seems to me to be one of the most frightful muddles I have ever read, with scarcely a sound proposition in it ... It is an extraordinary example of how, starting with a mistake, a remorseless logician can end up in Bedlam; op. cit., s. 394.

Så meget for indbyrdes kollegial respekt og hensynstagen.

I sit gensvar anklager Hayek Keynes for ikke at have forsøgt at forholde sig til den oprindelige kritik på en konstruktiv måde. I stedet for at forsøge »to improve the basis for further discussion«, så indeholder Keynes' svar til Hayek kun »a sweeping accu-

fortsat ...

udbruddet af den 2. Verdenskrig blev relationerne mellem de to bedre. Således så Hayek eksempelvis anerkendende på *How to Pay for the War*, som udkom som et Keynes skrift i 1940. For en mere generel belystning af ligheder såvel som forskelle i opfattelsen af centrale økonomisk relevante aspekter hos de to berømte økonomer, udover den her belyste debat, kan der henvises til Carabelli og De Vecchi (1999).

13. Samme sted oplyses det, at kopien af review'et er en af de mest kommenterede artikler, som fandtes i Keynes' arkiv med 34 håndskrevne bemærkning på et i alt 26 siders langt review.

sation of confusion«, hvorfor Hayek »cannot see what possible end is served by an unproved condemnation of my views in general«; Hayek (1931a, s. 398). Og Keynes' mangel på en accept af eksistensen af en tilpasningsmekanisme, som sikrer overensstemmelse mellem den samlede opsparring og de gennemførte investeringer, forstår Hayek simpelt hen ikke. Som Hayek (1931a, s. 401) bemærker; en sådan mangel på forståelse svarer, ifølge ham, til at acceptere »the more general contention that there is no automatic mechanism in the economic system to adapt production to any other shift in demand«! Det må nok skyldes, som han i forlængelse heraf skriver, at Keynes aldrig har været interesseret i at studere de fundamentale ikke-monetære aspekter og problemer ved en kapitalistisk produktionsproces: »Mr. Keynes has never reflected upon the function of the rate of interest in a society where there is no banking system«.

I februar 1932 kommer så den tredje artikel, hvori Hayek på ny gør sig nogle refleksioner over Keynes monetære forståelse. I dette bidrag, som synes at være mindre polemisk,¹⁴ men bestemt ikke mindre kritisk, i sit indhold end de foregående to bidrag fra Hayeks hånd, tages især Keynes' mere monetære aspekter op til diskussion.¹⁵ Således påpeges det, at Keynes' sondring mellem en markedsrente og den naturlige rente, og disses betydning for bestemmelse og påvirkning af investeringsomfanget næppe kommer som en overraskelse for de, som er bekendt med Wicksells pengeteori. Derimod vil Keynes' konklusioner nok snarere volde dem problemer, mener Hayek. Dernæst gives der en mere generel redegørelse af rentens betydning og af banksystemets mulige påvirkning og/eller kontrol af denne. Endelig tager Hayek endnu engang Keynes' opfattelse af, at opsparings/investerings ligevægten ikke automatisk holder op til diskussion uden, at der egentlig kommer nye aspekter på banen i forhold til de tidligere bidrag fra Hayek.

Efter læsning af debatten mellem Hayek og Keynes kan man kun give Dostaler (1991, s. 90) ret, når han om denne siger: »One often gets a sense of a dialogue among

14. Nu havde der også forud for denne artikels publicering været en længere brevveksling mellem Keynes og Hayek, hvor de begge forsøgte på en mere kollegial vis at kommunikere deres indbyrdes forskellige forståelse til hinanden; jf. CW (vol. XIII, s. 257-66). Efter offentliggørelsen af Hayeks artikel skriver Keynes således i et brev til ham den 19. marts 1932: »Having been much occupied in other directions, I have not yet studied your *Economica* article as closely as I shall. But, unless it be on one or two points which can perhaps be dealt with in isolation from the main issue, I doubt if I shall return to the charge in *Economica*. I am trying to re-shape and improve my central position, and that is probably a better way to spend one's time than in controversy«; op. cit., s. 266. Keynes' vej mod *The General Theory* var således allerede begyndt.

15. Og som påpeget af Cochran og Glahe (1994) så blev forskellen mellem Hayek og Keynes, hvad angår deres monetære forståelse stadig mere uddybet, som tiden gik. Med fremkomsten af *The General Theory* blev det entydigt klart, at Keynes var af den overbevisning, at effekterne af monetære forstyrrelser også kunne være af en realøkonomisk karakter. Også derfor var deres syn på pengepolitikens muligheder hinandens modsætninger. Med Cochran og Glahe (1994, s. 83) om et forsøg på en aktiv indsats over for et økonomisk kriseforløb: »If Keynes is right, the effects of a Keynesian policy change are permanent; policy may be needed and will be effective. If Hayek is right, the effects of policy are not permanent; policy will not only be ineffective in the long run but also cause future instability«.

the deaf«. Som også fremhævet af Dostaler var Hayek dog overbevist om sine egne teories forklaringskraft.¹⁶ Og rigtigt er det da også, at debatten mellem de to fremtrædende økonomer i sin samtid blev set som kampen mellem to alternative systemer eller økonomiske verdenssyn.¹⁷ Som John Hicks – der så at sige havde en fod i begge lejre, i og med, at han i begyndelsen af 1930'erne var tilknyttet LSE og alligevel samtidig var under kraftig inspiration af Keynes' tænkning – skriver i et tilbageblik herom:

When the definitive history of economic analysis during the nineteen-thirties comes to be written, a leading character in the drama (it was quite a drama) will be Professor Hayek ... it is hardly remembered that there was a time when the new theories of Hayek were the principal rival of the new theories of Keynes. Which was right, Keynes or Hayek? There are many still living teachers of economics and practical economists, who have passed through a time when they had to make up their minds on that question; and there are many of them (including the present writer) who took quite a time to make up their minds; Hicks (1967, s. 203).

Eller som Caldwell (1995 s. 31) ser det:

In early 1932, Hayek appeared to be the victor in the theoretical battle. His model, though unfamiliar to his English audience, at least seemed the more coherent. But was it the right model, did it really constitute a general theory of the cycle? If it did not, were Hayek's pessimistic policy views then mistaken, and Keynes, whose model was insufficiently worked out, really right after all? These were the questions that taxed economists in London and Cambridge in the spring of 1932. Disagreement over the matter would not last for long.

Nok havde de to oprindelig det samme teoretiske udgangspunkt – nemlig Wicksells tænkning – men mens Hayeks loyalitet mod dette udgangspunkt var dybfølt, var Keynes mere troløs. Eller som Hicks (1967, s. 204) pointerede: »Wicksell plus Keynes said one

16. Jf. Dostaler (1991, s. 80): »As late as 1975, in a preface to the French translation of *Prices and Production*, Hayek wrote that he still believed that his 1931 theory was the correct explanation of the crisis and that the errors of the previous forty years would have been avoided if he – rather than Keynes – had been listened to«. Hos Caldwell (2004) tages der dog noget af et forbehold m.h.t. dette. Efter debatten med Keynes arbejdede Hayek videre med sine teorier. Resultatet heraf var *The Pure Theory of Capital* fra 1941. Hayek var dog ikke selv helt tilfreds hermed. Dertil var for mange problemer forblevet uløste. Tiden var derfor inde for en ændring af Hayeks fokus m.h.t. det fremtidige arbejde. Som Hayek selv bemærkede: »After a highly specialized effort of seven years with only half the task completed I must confess I had become so tired of it that I used the outbreak of the war as an excuse to publish the first part and then to turn to what had come to appear to me more pressing problems«; op. cit., s. 180.

17. For en belysning af, hvorvidt man ligefrem kan tale om to inkommensurable kuhnske paradigmer, kan der henvises til diskussionen i Tieben (1997). Det kan man dog næppe, er den konkluderende opfattelse heri, idet Hayek og Keynes i deres tænkning »account for different classes of phenomena which are not mutually exclusive, but refer to different aspects of the same process ... that of coordination between individual decisions in a decentralized economy«; op. cit., s. 121. Deres teorier supplerer således snarere hinanden frem for at være hinanden udelukkende.

thing, Wicksell plus Hayek said quite another«. Som hævdet af Fontana (2007) er den wicksellianske forståelse basalt set forankret i en udbudsforståelse. Dermed er denne opfattelse af penge og monetære forhold ikke i overensstemmelse med Keynes i hans General Theory-forståelse, hvor det effektive efterspørgselsprincip indtager den makroøkonomiske modelscene. Der er således sket en afgørende videreudvikling hos Keynes, når tænkningen i Treatise sammenlignes med tænkningen i The General Theory.

Som Foss (1994) fremhæver, så kunne Hayek med rette kritisere Keynes for at være for aggregeret i sin makroøkonomiske analyse. Alene af den grund må den østrigsk inspirerede økonomiske tankegang hos Hayek have voldt ham problemer at forstå til bunds. Dermed undertrykkes eksempelvis den substituerende effekt en ændring i de relative prisforhold kan afstedkomme. Og dette er især belastende for en makroøkonomisk analyse med en antagelse om fuld beskæftigelse, som den Keynes foretager i Treatise på basis af sine fundamentale ligninger.¹⁸ Netop derfor holder Keynes' berømte bananplantageeksempel heller ikke set med den hayekske optik.¹⁹ I dette eksempel forsøger Keynes, Keynes (1930 vol., s. 158-60), at beskrive en situation, hvor en stigning i opsparingstilbøjeligheden skævvrider opsparings/investeringsligevægten. Nu overstiger den samlede opsparing de faktisk gennemførte investeringer. Dette afstedkommer et fald i prisen på bananer, men da omkostningerne er de samme som før prisfaldet, bevirker dette blot, at entreprenurerne lider et tab. Der er hermed sket en omfordeling af velstand, påpeger Keynes, fra producenterne til forbrugerne. Som en reaktion på deres tab, vil de enkelte entreprenure nu forsøge at reducere deres omkostninger (presse lønnen ned) og/eller reducere antallet af ansatte. Som en konsekvens heraf oplever forbrugerne nu som gruppe betragtet en nedgang i deres købekraft, hvorfor efterspørgslen i samfundet efter bananer falder, hvilket blot på ny får prisen på disse til at falde yderligere. Og så kan processen tilsyneladende fortsætte i nærmest al uendelighed. Hvordan kan man nu komme ud af en sådan fælde? Herom skriver Keynes:

Thus there will be no position of equilibrium until either (a) all production ceases and the entire population starves to death; or (b) the thrift campaign is called off or peters out as a re-

18. Som Foss (1994, s. 27) påpeger, så kritiserede Hayek Keynes for at undertrykke: »the relative price between consumer goods and investment goods, or, more trivially, of overlooking that under conditions of full resource utilization, investment and consumption must vary inversely in the short run. In Hayek's own analysis, this relative price is the central one«.

19. Med Foss (1994, s. 31): »Keynes has suppressed the allocative function of the rate of interest and the relative price between consumption goods and investment goods, and it is this essentially arbitrary suppression which allows Keynes to produce his dire conclusion«.

sult of the growing poverty; or (c) investment is stimulated by some means of another so that its cost no longer lags behind the rate of saving; Keynes [1930 vol., s. 160].

Derudover kan Keynes også, igen med Foss (1994) med rette angribes for at behandle rentens betydning for den intertemporale allokering stedmoderligt. Måske undertrykte Keynes netop aspekterne ved en intertemporal koordination på grund af usikkerhedens betydning for de økonomiske transaktioner, som rækker langt ind i fremtiden. Agerer de økonomiske agenter i et økonomisk miljø, der er kendetegnet ved eksistensen af en fundamental form for usikkerhed, jf. Olesen (2009), hvor den fremtidige økonomiske udvikling på afgørende vis måske er ganske uforudsigelig – agenternes aktuelle handlinger er med til at forme udfaldsrummet for fremtidige hændelser eller udfald – mister den intertemporale koordineringsproces noget af sin relevans. Det er dog nok tvivlsomt om denne opfattelse af en sådan »sti-afhængighed« i den økonomiske beslutningstagen allerede er til stede hos Keynes i *Treatise*. Et sådant gennembrud bør nok snarere tilskrives den økonomiske forståelsesramme, der præsenteres med fremkomsten af *The General Theory*.

Som belyst i Zouache (2008) kan der dog også uddrages andre aspekter af debatten mellem Hayek og Keynes. Hans pointe i sin belysning af debatten imellem de to er således den, at Hayek med rette kritiserede Keynes' kapitalteori i *Treatise* for ikke at have et mikroøkonomisk fundament. Hayek insisterede på, at det var nødvendigt at studere konjunkturudsvingene ud fra et mikroøkonomisk funderet perspektiv. Og tankegangen hos Keynes fokuserede alene på funktionelle relationer mellem aggregerede størrelser. Ifølge Zouache tog Keynes denne kritik til sig. Hans efterfølgende tænkning i *The General Theory* er således meget mere fokuseret på, at den præsenterede makroøkonomiske analyse skal have en mikroøkonomisk basis. Mest udtalt kommer dette, ifølge Zouache, til udtryk i det nye MEC-begreb, hvor den enkelte producent danner sig nogle forventninger med hensyn til et givet investeringsprojekts fremtidige afkast, der sammenstilles med renten. Er afkastmulighederne gunstige (de overstiger renten), gennemføres investeringen.

4. En opsamlende afrunding

Som belyst i det ovenstående blev udgivelsen af *Treatise* ikke den succes, som både Keynes og flere af hans samtidige kolleger havde set frem til. Hensigten med bogen var anerkendelsesværdigt – hvorfor kommer en makroøkonomi uden for den stabile ligevægt, som den herskende mainstream tænkning påstod, at den for normale forholds vedkommende pr. automatik skulle befinde sig i? Og, hvordan bevæger den sig dynamisk i et uligevægtspræget forløb hen imod den nye stabile makroøkonomiske ligevægt? – men forklaringerne herpå var ikke rigtigt overbevisende. En reformulering af

kvantitetsteorien som den relevante forståelsesramme synes ikke at kunne levere de overbevisende svar på den oven for rejste problematik. Og egentlig er det vel på forhånd også en nytteløs opgave at forsøge at få en teoretisk forståelsesramme, der implicit indeholder en antagelse om etablering og fastholdelse af et harmonisk makroøkonomisk udfald med fuld beskæftigelse, til at kunne give relevante forklarende bud på, hvordan en økonomisk krise kan opstå og tilsyneladende fastholdes for en længerevarende periode i et ikke ubetydeligt omfang. Og i en sådan klassisk forståelsesramme har penge ingen essentiel rolle at spille udover den at virke som et hensigtsmæssigt smøremiddel, der fremmer de forskellige transaktioners gennemførelse. Realøkonomisk set er penge neutrale i en kvantitetsteoretisk sammenhæng: »The *Treatise* ... tended to assume that money was neutral – that changes in monetary variables could not affect the long run equilibrium position of the economy's real variables«; Møggridge (1992, s. 484). Det var først i *The General Theory*, at Keynes for alvor satte fokus på, at »money matters« på kort såvel som på lang sigt.

Men selvfølgelig er der også noget i *Treatise*, som peger fremad mod den nye forståelse, som skulle komme til at materialisere sig i *The General Theory*. Væsentligst er her Keynes' fokus på opsparings/investerings ligevægten. I 1930 er han helt med på, at en sådan ligevægt ikke nødvendigvis altid holder. Opsparings- og investeringsbeslutningerne foretages typisk af forskellige personer, ligesom motiverne bagved de to beslutningstyper er forskellige. Spørgsmålet bliver nu, om der findes en endogen tilpasningsmekanisme i det økonomiske system, som kan bringe en eventuel manglende ækvivalens mellem de to størrelser til ophør. Som Skidelsky (1992) ser det, forsøger Keynes i *Treatise* at fortælle to historier. Den ene – som fylder mest i værket – er en »credit cycle« historie. Den anden historie er en fortælling om, hvordan et deflationært chok kan fremkalde en selvforstærkende nedadgående konjunkturbevægelse med tilsyneladende »no point of return«; dette er Keynes' berømte bananplantage eksempel. Inden for den første histories rammer er man sikret en sådan endogen tilpasningsmekanisme, der om end den til tider kan være træg og noget langsom i sin funktionsduelighed, dog alligevel sikrer den makroøkonomiske ligevægt genoprettet på sigt. En sådan effektiv tilpasningsmekanisme er ifølge Skidelsky ikke givet inden for den anden histories rammer:

In the second story, which abstracts from banking policy, there is no adjustment mechanism. Keynes regarded the second story as his real contribution, the kernel of his message. It points to the *General Theory*. Did Keynes already feel, by the time he had finished the *Treatise*,

20. Som argumenteret for af Ertürk (1998) så forsøgte Keynes efter udgivelsen af *Treatise* faktisk at formulere et oplæg til en dynamisk tankegang, der havde til hensigt at vise, at en ligevægtsskabende tilpasningsmekanisme næppe var til stede i en moderne monetær produktionsøkonomi. Ifølge Ertürk (1998, s. 178 og

that cyclical phenomena were less significant than the fact that the cycle was taking place at such a reduced level of activity?; Skidelsky [1992, s. 315-16].²⁰

Og har Skidelsky ret, synes man heri at kunne ane spiren til den permanente underbeskæftigelsesligevægt, som var et af de væsentlige temaer i den fortælling, som Keynes præsenterede sine læsere for i *The General Theory*.

Og så er der også en helt anden forskel på de to værker. *The General Theory* omhandler primært et studie af en lukket økonomi, da det alt afgørende formål med bogen, som nævnt, er at udfordre den herskende ortodoksi på dennes kerneudsagn: hvordan bestemmes den samlede produktion og indkomst i samfundet, og hvorledes foreløber de makroøkonomiske tilpasningsprocesser. Dermed udgør *The General Theory* i forhold til Keynes' tidligere forfatterskab – eksemplificeret ved *A Tract on Monetary Reform* og *A Treatise on Money* – et klart skifte i den økonomiske fokusering, der tidligere ellers altid havde været helliget et studie af den internationale økonomis monetære forhold. Som Patinkin (1976 s. 8) påpeger, så var et sådant internationalt fokus »one of Keynes's major concerns throughout his life, and in these two books in particular«.

På sin vis kan man også, som eksempelvis Chase (1994) gør det, sige, at de to værker supplerer eller måske snarere komplementerer hinanden. Mens teorigrundlaget i *Treatise* i sit oplæg er mikroøkonomisk forankret med fokus på udbudsforhold, er *The General Theory* mere banebrydende makroøkonomisk forankret – dog naturligvis med et mikroøkonomisk fundament: det er i *General Theory* de enkelte agents for-

fortsat...

179) var Keynes' argumentation bygget op omkring de følgende fem forhold: (1) variationen i kapacitetsudnyttelsen er en funktion af overskudsefterspørgslen i samfundet; (2) enhver afvigelse fra den makroøkonomiske ligevægt kan være selvforstærkende; (3) tendensen til instabilitet i økonomien forsøges holdt i ave af en ikke-lineær opsparingsfunktion; (4) en given makroøkonomisk ligevægt er hverken automatisk entydig eller nødvendigvis optimal; samt (5) tendensen til en endogen omkostningstilpasning virker destabiliserende på det kortere sigt, men kan på det længere sigt have modsatrettede effekter i form af en forbedret konkurrenceevne. Senere forlod Keynes dog denne linie af argumentation og søgte tilbage til en mere statisk belysning af den effektive efterspørgsel og det økonomiske systems mangel på en automatisk makroøkonomisk stabilitet; jf. fremstillingen i *General Theory*. Ifølge Ertürk foretog Keynes dette markante skifte i sin teoretiske formulering, fordi han primært ønskede at angribe den herskende mainstream på dennes egne metodologiske betingelser.

21. Med Chase (1994, s. 850): »Thus, the transition from *TM* to *GT*, as well as the relationship between the two works, can be seen essentially as a methodological shift ... This was a shift from a focus on the disequilibrium, micromethodological, process-oriented framework of *TM* (AS) to the equilibrium macroeconomic adjustment mechanism of the *GT* (AD)«. Som argumenteret for i Olesen (2008 og 2009) så var den metodologiske tilgang, som Keynes anvendte i *The General Theory* dog nok mere revolutionerende i sit udgangspunkt som i sit indhold end som så. Det fundamentalt metodologisk set nye – oplægget til, hvordan det økonomiske system skal betragtes som et åbent, socialt og foranderligt, det vil sige som et ikke-repetitivt kontekst afhængigt system (non-ergodicity) – i denne approach tages dog ikke op til debat i Chases fremstilling.

ventningsbestemte adfærd til en principielt ukendt fremtid præget af en fundamental type af usikkerhed, som betinger det til enhver tid makroøkonomiske udfald – med en kraftig fokusering på efterspørgselsforhold. Og som Chase også rigtig påpeger, så var den metodologiske tilgang i de to værker af en forskellig karakter.²¹ Om der kunne være kommet noget frugtbart ud af at kombinere de to tilgange, er nok tvivlsomt. Under alle omstændigheder forsøgte Keynes sig ikke selv med et sådant forsøg.

Hvorom alting er; så kan læsning også af *Treatise* dog give en et indblik i, hvorledes Keynes' langstrakte intellektuelle befrielseskamp fra den herskende ortodoksi forløb. I denne forstand er værket en væsentlig brik i et større puslespil. Også selvom *Treatise*-Keynes generelt set ikke er nær så sproglig inspirerende og udfordrende at læse som *General Theory*-Keynes.²²

22. Som Patinkin (1976, s. 24) siger det: »Least enjoyable as a reading experience is the *Treatise*, whose generally heavy and constrained style reflects the stately scientific objective that Keynes set for himself in it. Indeed, when one reads the *Treatise* against the background of Keynes' other writings, one cannot escape the feeling that it represents a Keynes out of character, a Keynes attempting to act the role of a Professor, and a Germanic one at that«.

Litteratur

- Amadeo, Edward and Dutt, Amitava Krishna. 1994. The Wicksell-Keynes Connection: Dynamic analysis, loanable funds, and wage flexibility, *Australian Economic Papers*, December 1994, s. 253-71.
- Caldwell, B. 2010. Cambridge Talk on Hayek, konferenceindlæg på The Economic Crisis and Crisis in Economics, New Economic Thinking, King's College Cambridge, 8.-10. april 2010.
- Caldwell, B. 2004. *Hayek's Challenge. An Intellectual Biography of F. A. Hayek*, The University of Chicago Press 2004.
- Caldwell, B. 1995. Introduction fra bogen *Contra Keynes and Cambridge – Essays, Correspondence, The Collected Works of F. A. Hayek* Vol. IX, Routledge 1995, s. 1-48.
- Carabelli, A. and N. De Vecchi. 1999. Where to draw the line? Keynes versus Hayek on knowledge, ethics and economics, *The European Journal of the History of Economic Thought*, Summer 1999, s. 271-96.
- Chase, R. 1994. The Fatal Flaw of Classical Economics: Aspects of Keynes's Evolution from The *Treatise* to The *General Theory*, *Journal of Economic Issues*, September 1994, s. 847-75.
- Chick, V. 2003. Theory, method and mode of thought in Keynes's *General Theory*, *Journal of Economic Methodology* 10:3, September 2003, s. 307-27.
- Chick, V. and S. C. Dow. 2005. The meaning of open systems, *Journal of Economic Methodology* 12:3, September 2005, s. 363-81.
- Cochran, J. P. and F. R. Glahe. 1994. The Keynes-Hayek Debate: Lessons for Contemporary Business Cycle Theorists, *History of Political Economy* 26:1, pp. 69-94.
- CW (XII). *The Collected Writings of John Maynard Keynes, Vol. XII, Economic Articles and Correspondence – Investment and Editorial*, Macmillan Cambridge University Press 1983.
- CW (XIII). *The Collected Writings of John Maynard Keynes, Vol. XIV, The General Theory and After, Part I: Preparation*, Macmillan Cambridge University Press 1973.
- CW (XIV). *The Collected Writings of John Maynard Keynes, Vol. XIV, The General*

- Theory and After, Part II: Defence and Development*, Macmillan Cambridge University Press 1973.
- Dostaler, G. 1991. The debate between Hayek and Keynes, fra bogen *Themes in Keynesian Criticism and Supplementary Modern Topics*, Barber William (ed.), *Perspectives on the History of Economic Thought*, Vol. VI, Edward Elgar 1991, s. 77-101.
- Dow, S. 1996. *The Methodology of Macroeconomic Thought – A Conceptual Analysis of Schools of Thought in Economics*, Edward Elgar 1996.
- Ertürk, K. A. 1998. From the Treatise to the General Theory: the transformation of Keynes's theory of investment in working capital, *Cambridge Journal of Economics* 1998, s. 173-85.
- Fischer, S. 1972. Keynes-Wicksell and Neoclassical Models of Money and Growth, *American Economic Review*, December 1972, s. 880-90.
- Flaschel, P. 2000. Keynes-Marx and Keynes-Wicksell Models of Monetary Growth: a framework for future analysis, *Review of Political Economy* 2000, s. 453-68.
- Fontana, G. 2007. *Why money matters: Wicksell, Keynes, and the new consensus view on monetary policy*, *Journal of Post Keynesian Economics*, Fall 2007, s. 43-60.
- Foss, N. J. 1994. *Intertemporal Coordination of Economic Activities: the Early Debate between Hayek and Keynes* fra bogen *The Austrian School and Modern Economics: Essays in Reassessment*, Handelshøjskolen Forlag 1994, s. 13-38.
- Hayek, F. A. 1931. Reflections on the Pure Theory of Money of Mr. J. M. Keynes, *Economica*, August 1931, s. 270-95.
- Hayek, F. A. 1931a. A Rejoinder to Mr. Keynes, *Economica*, November 1931, s. 398-403.
- Hayek, F. A. 1932. Reflections on the Pure Theory of Money of Mr. J. M. Keynes (continued), *Economica*, February 1932, s. 22-44.
- Hicks, J. 1967. *The Hayek Story* fra bogen *Critical Essays in Monetary Theory*, Oxford at the Clarendon Press 1967, s. 203-15.
- Hirai, T. 2007. How did Keynes transform his theory from the Tract into the Treatise? – Consideration through primary material, *The European Journal of the History of Economic Thought*, June 2007, s. 325-48.
- Keynes, J. M. 1931. The Pure Theory of Money. A Reply to Dr. Hayek, *Economica*, November 1931, s. 387-97.
- Keynes, J. M. 1930. *A Treatise on Money, The Collected Writings of John Maynard Keynes Vol. V-VI*, Macmillan St. Martin's Press 1971.
- Lambert, P. 1969. The evolution of Keynes's thought from The Treatise on Money to The General Theory, *Annals of Public and Cooperative Economics*, July 1969, s. 243-64.
- Moggridge, D. E. 1992. *Maynard Keynes – An economist's biography*, Routledge 1992.
- Moggridge, D. E. 1973. From the Treatise to The General Theory: An Exercise in Chronology, *History of Political Economy* 1973, s. 72-88.
- Olesen, F. 2010. *John Maynard Keynes: The Making of a Revolution – The birth of modern macroeconomics*, VDM Verlag Dr. Müller, Saarbrücken 2010.
- Olesen, F. 2009. Keynes om fundamental usikkerhed, *Samfunnsøkonomen* 2009, nr. 3, s. 12-22.
- Olesen, F. 2009a. A Treatise on Money – et teoriehistorisk case studie, *Working Paper* 91/09, IME, SDU.
- Olesen, F. 2008. Keynes' metodologi og makroøkonomisk forskning – et bud på en belysning, *Økonomi & Politik* 2008, nr. 1, s. 46-65.
- Olesen, F. 2007. *Rundt om The General Theory – en teoriehistorisk belysning*, Syddansk Universitetsforlag 2007.
- Patinkin, D. 1976. Keynes' Monetary Thought. A study of its development, *History of Political Economy* 1976, s. 1-150.
- Perregaard, H. P. 2009. F. A. Hayeks konjunkturteori, *Nationaløkonomisk Tidsskrift* 2009, s. 318-36.
- Skidelsky, R. 2010. Interpreting the Great Depression: Hayek versus Keynes, konferenceindlæg på *The Economic Crisis and Crisis in Economics*, New Economic Thinking, King's College Cambridge, 8.-10. april 2010.

- Skidelsky, R. 1992. *John Maynard Keynes – The economist as saviour, 1920-1937*, Macmillan 1992.
- Stein, J. 1969. Neoclassical and Keynes-Wicksell Monetary Growth Models, *Journal of Money, Credit and Banking*, May 1969, s. 153-71.
- Tieben, B. 1997. Misunderstandings and other coordination failures in the Hayek-Keynes controversy fra bogen *Austrian Economics in Debate*, Keizer, Willem et al. (eds.), Routledge 1997, s. 95-126.
- Zouache, A. 2008. On the microeconomic foundations of macroeconomics in the Hayek-Keynes controversy, *The European Journal of the History of Economic Thought*, March 2008, s. 105-27.

Replik

»Er der et dansk produktivetskollaps?«

Frank Rasmussen

Økonomi- og Erhvervsministeriet, E-mail: fra@oem.dk

Dalgaard og Hansen (DH) bringer i artiklen »Er der et dansk produktivetskollaps?« en statistisk analyse af den danske produktivitetsvækst for at undersøge, om der er sket et brud i udviklingen. Artiklen er et bidrag til debatten om, hvorvidt Danmark har et produktivetsproblem, og sætter fokus på et vigtigt og aktuelt emne.

Økonomi- og Erhvervsministeriet udgav i efteråret 2009 en analyse af produktivetsudviklingen, som viser, at Danmark velstandsmæssigt har tabt terræn i forhold til andre lande siden midten af 1990'erne. Danmarks tab af terræn skyldes en særskilt svag produktivitetsvækst. Den danske produktivitetsvækst har i perioden siden midten af 1990'erne ligget på gennemsnitligt knap 0,8 pct. om året. Det er lavt i både et historisk perspektiv og i forhold til andre rige lande. OECD-landene har under ét haft en gennemsnitlig årlig vækst på knap 2 pct. om året.

DH rejser to spørgsmål i deres artikel. Dels om trenden i den danske produktivitetsvækst har været aftagende, dels om den danske vækst er urovækkende lav i forhold til andre lande.

Har trenden i den danske produktivitetsvækst været aftagende?

Væksten i produktiviteten svinger en hel del fra år til år. Dette er et veldokumenteret fænomen og også belyst i Økonomi- og Erhvervsministeriets analyse. Det betyder i sig selv, at det statistiske konfidensinterval om trendudviklingen bliver meget bredt.

En måde at komme ud over dette problem på er ved at se på et glidende gennemsnit af produktivitetsvæksten over en længere årrække. En anden måde er at sammenligne udviklingen i produktivetsniveauet i forhold til andre, sammenlignelige lande. I begge tilfælde fjernes de store årlige udsving, og der fås et bedre indtryk af den trendmæssige udvikling. Den sidste tilgang er f.eks. anvendt af vismændene, der har vist, at måler man udviklingen i Danmarks relative timeproduktivetsniveau over tid i forhold til Tysklands, kan det med betydelig statistisk margin påvises, at den danske mængdemæssige værditilvækst pr. time er faldet i forhold til Tysklands. Samme klare konklusion får man ikke, hvis man analyserer vækstraterne. Således kan statistiske tests ikke afvise, at timeproduktivetsvæksten har været den samme i Danmark og Tyskland i perioden, jf. De Økonomiske Råd (2010).

DH har valgt at tage afsæt i de voldsomt fluktuerende årlige vækstrater. Det heraf følgende brede konfindensinterval gør det statistisk vanskeligt at påvise skift i den trendmæssige udvikling i produktiviteten. Den statistiske usikkerhed i DH's tests øges ydermere af, at de i deres primære analyse af udviklingen i den danske produktivitetsvækst har valgt at analysere udviklingen i BNP pr. arbejdstime ved at anvende data fra Penn World Table (PWT) suppleret med data for antal arbejdstimer fra The Conference Board Total Economy Database. Timetallene i denne database er identiske med time-tallene i det danske nationalregnskabstal fra 1966 og frem.

De forskellige analyseresultater, som DH finder på baggrund af de internationale tal og tallene fra Danmarks Statistik, skal derfor findes i, at tallene for den mængdemæssige udvikling i BNP pr. indbygger i PWT afviger fra de officielle tal i det danske nationalregnskab. Den større variation i tallene fra PWT fremgår tydeligt af artiklens figur 6, hvor forskellene i de faktiske vækstrater alene kan henføres til forskelle i vækstraterne i BNP pr. indbygger.

Anvendelsen af de forskellige datasæt har betydning for konklusionen. Når DH anvender de officielle tal fra det danske nationalregnskab, kan de ikke afvise, at vækstraten i BNP pr. arbejdstime er aftaget (idet de finder »meget klare tegn på parameterustabilitet i vækstraten i BNP pr. arbejdstime«), mens konklusionen er noget mere svag, når data fra PWT anvendes. Det betyder, at de i deres samlede konklusion kun når frem til, at »der *muligvis* er en aftagende trend«.

Da analysen alene er på den danske produktivitetsudvikling, forekommer det ubegrundet at anvende de internationale tal. Derfor har Økonomi- og Erhvervsministeriet i tidligere statistiske analyser anvendt de officielle nationalregnskabstal og vist, at væksten i BNP pr. arbejdstime er faldet over tid (eller formelt, at produktivitetsvæksten ikke kan beskrives ved udsving omkring en konstant middelværdi, idet konstanten ikke er stabil over hele perioden), jf. Økonomi- og Erhvervsministeriet (2010).

Er den danske vækst urovækkende lav?

Det er velkendt, at fattige lande normalt har højere vækstrater end rige lande som følge af catch-up. DH finder i overensstemmelse med betinget konvergens, at Danmarks BNP pr. indbygger relativt til det gennemsnitlige niveau i OECD er faldet over de sidste 50-60 år. Ses der alene på perioden siden midten af 1990'erne, har Danmark imidlertid også mistet terræn til lande, der i forvejen var rigere end os. Denne udvikling er ikke i overensstemmelse med betinget konvergens, og denne pointe forsvinder i DH's lange tidshorisont siden 1950. En udvikling, der har stået på i mere end et årti, kan ikke tilskrives midlertidige afvigelser fra konvergensprocessen i form af asynkron konjunkturforløb som anført i artiklen.

Flere andre kilder har også peget på en særskilt svag dansk produktivitetsudvikling siden midten af 1990'erne, herunder OECD (2009) og senest vismændene. Vismændene finder blandt andet, at »fra midten af 1990'erne frem til 2008 er den danske værditilvækst pr. arbejdstime faldet fra at udgøre 96 pct. af Tysklands til at udgøre 88 pct. Tilsvarende er den relative, danske timeproduktivitet faldet fra 97 pct. til ca. 85 pct., når der sammenlignes med USA. I forhold til Sverige faldt den danske værditilvækst pr. arbejdstime fra 108 pct. til 95 pct. i samme periode.«, jf. De Økonomiske Råd (2010). Faldene er kommet efter, at Danmark frem til midten af 1990'erne har halet kraftigt ind på lande med et højere produktivetsniveau.

Samlet set er indtrykket fortsat, at trenden i den danske produktivitetsvækst er aftagende og tilmed aftagende i forhold til en række sammenlignelige lande siden midten af 1990'erne. Hvad der er den bagvedliggende årsag til denne bekymrende udvikling, er der endnu ikke fundet udtømmende svar på. En bedre forståelse for, hvad der driver udviklingen, er helt central i forhold til tiltag, der vil kunne løfte produktiviteten. Fremadrettet vil aldringen af befolkningen trække i retning af, at arbejdsindsatsen vil falde. En styrket produktivitetsvækst vil af den grund blive afgørende for den danske velstandsudvikling.

Litteratur

- Dalgaard, C.-J. og H. Hansen. 2010. Er der et dansk produktivitetskollaps? *Nationaløkonomisk Tidsskrift*, 148, 1-20.
- De Økonomiske Råd. 2010. *Dansk Økonomi – Efterår 2010*. København.
- OECD. 2009. *Economic Survey – Denmark*. Paris.
- Økonomi- og Erhvervsministeriet. 2009. Den danske produktivitetsudvikling. *Økonomisk Tema nr. 8*. November 2009.
- Økonomi- og Erhvervsministeriet. 2010. Bemærkninger til analysen: Er der et dansk produktivitetskollaps?

Duplik

Er der et dansk vækstkollaps?

Carl-Johan Dalgaard

Økonomisk Institut, Københavns Universitet, E-mail: Carl.Johan.Dalgaard@econ.ku.dk

Henrik Hansen

Fødevareøkonomisk Institut, Københavns Universitet, E-mail: henrik.hansen@foi.dk

Indledningsvis vil vi gerne takke Økonomi- og Erhvervsministeret (ØEM) for den interesse, de har udvist for vores analyse i flere omgange. Først via den kommentar ØEM cirkulerede via deres hjemmeside (ØEM, 2010), og som vi efterfølgende svarede på, Dalgaard og Hansen (2010b). Senest ved nærværende – nye – kommentar, som vi vil adressere i det følgende.

Vores analyse, Dalgaard og Hansen (2010a) adresserer to – separate – spørgsmål:

- (1) Er den danske produktivitetsvækst aftaget siden midten af 1990'erne?
- (2) Er den danske produktivitetsvækst aftaget i *forhold til andre lande* siden midten af 1990'erne?

Der er oplagt tale om to separate spørgsmål. Hvor et bekræftende svar på (1) kræver et decideret fald i væksten i Danmark (i 1990'erne), kan (2) udmærket være foreneligt med konstant vækst i Danmark; blot kræver det, at væksten er accelereret i verden omkring os.

I ØEM (2009) hævdes det, at man kan svare bekræftende på begge spørgsmål. Dette postulat er udgangspunktet for nærværende debat.

Spørgsmål 1. Er den danske vækst trend aftaget i 1990'erne?

For at svare på spørgsmål 1 forekommer det oplagt, at analysere udviklingen i den danske vækst ved brug af *formelle* økonometriske metoder, og derigennem teste om væksten i Danmark udviser trendbrud i 1990'erne.

Svaret, man opnår, er ganske entydigt: *Der er ikke tale om noget trendbrud i væksten i 1990'erne.* Denne konklusion er robust. Dels udsætter vi data for hele batteriet af formelle test for trendbrud (stokastisk såvel som deterministisk), og dels varierer vi data-kilderne. Modsat hvad man måske kunne forledes til at tro på basis af ØEMs kommentar, så opnås kvalitativt den samme konklusion, ligegyldigt, om vi trækker på Penn World Tables BNP data, eller direkte på Danmarks Statistisk. Det er også ligegyldigt, om vi betragter BNP per indbygger, per beskæftiget eller per arbejdstime.

Det betyder imidlertid ikke, at vi konkluderer, at væksten har været uændret over de sidste ca. 40 år. Tværtimod, vores analyse peger netop på, at der kan være tale om et fald i produktivitetsvæksten målt ved BNP per arbejdstime. Dette til trods for, at noget sådant åbenbart (jfr., ØEMs kommentar) slet ikke burde være muligt at påvise grundet »de voldsomt svingende vækstrater« og det deraf afledte »brede konfidensinterval«.

Men sagens kerne er, at dette fald indfinder sig hele 2 årtier, før ØEM (2009) hævder faldet indtræffer: altså allerede i 1970erne. Hvad mere er, denne afmatning er velkendt, internationalt udbredt, og kendt som »the productivity slowdown«, se f.eks. Grilliches (1988). Der er således næppe tale om et isoleret dansk fænomen.

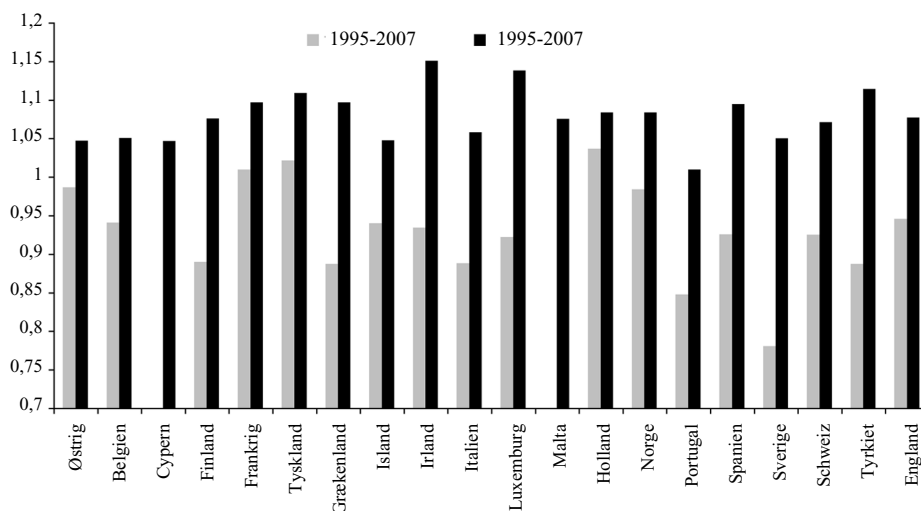
Når vi så, dette fund til trods, tager et mildt forbehold i vores konklusion (jfr., der »kan« være tale om et fald i væksten i timeproduktiviteten, med epicenter i 1970erne), så skyldes det ganske rigtigt dataforhold. Dog ikke data mht. BNP, som ØEM antyder i kommentaren; vi opnår jo kvalitativt de samme konklusioner BNP kilderne desuagtet. Derimod går vores bekymring på data mht. *timeindsatsen per beskæftiget*. Det er værd at præcisere, hvori vores bekymring består, siden dette (i lyset af ØEMs kommentar) åbenbart ikke er klart nok kommunikeret i Dalgaard og Hansen (2010a).

Indledningsvis er det værd at bide mærke i, at vores analyse viser, at væksten i produktiviteten, målt ved BNP *per beskæftiget*, er konstant igennem hele perioden, for hvilken vi har data (modsat timeproduktiviteten). Det turde være indlysende, at BNP per beskæftiget også rammes af »the productivity slowdown« (BNP indgår jo). Når man alligevel opnår et samlet billede af relativ stabilitet, så skal det tilskrives, at væksten i denne produktivitetsserie *revitaliseres* senere i perioden, hvorfor der overordnet set bliver tale om en stationær dansk vækstproces, når BNP per beskæftiget anvendes som produktivitetsmål.

Men denne revitalisering finder altså ikke sted, når det gælder timeproduktiviteten. Hvorfor ikke? Hvis timetallet per beskæftiget har været faldende (eller konstant) i Danmark over de seneste årtier, da bør man jo vente at væksten i time produktiviteten blot er lidt højere end væksten i BNP per beskæftiget. I så fald bør førnævnte dataserie udvise et mønster der i det store hele er parallelt med det for BNP per beskæftiget.

Når det ikke er tilfældet, så skyldes det, at timetallet per beskæftiget faktisk *stiger* i Danmark fra midten af 1990erne. Efter årtier med faldende timeindsats ses i denne serie et abrupt skift. I Dalgaard og Hansen (2010a) illustrerer vi dette fænomen afbilledet ved timetallet per beskæftiget i Danmark relativt til USA. Se Dalgaard og Hansen, (2010a, figur 11); timetallet stiger i Danmark, absolut såvel som relativt til USA.

Men det er nu ikke blot i forhold til USA, at den danske timeudvikling er bemærkelsesværdig over perioden 1995 og frem. Som figur 1 viser, stiger det danske timetal per beskæftiget i forhold til *samlige* europæiske OECD lande. I nogle tilfælde overordentlig markant. Således øges timetallet per beskæftiget i Danmark med ca. 10% i forhold til



Figur 1. Det danske timemirakel?

Figuren viser ændringen over perioderne 1970-1995 og 1995-2007 i det årlige timetal per beskæftiget i Danmark relativt til andre udvalgte OECD lande.

Kilde: The Conference Board Total Economy Database.

lande som Tyskland, Frankrig og England. Stigningen er endnu mere udtalt i forhold til Tyrkiet, og ganske spektakulær i forhold til f.eks. Irland (ca. 15%). Udviklingen i time-tallet per beskæftiget, 1995 og frem, er unik for Danmark. Kontrasten er tydelig til pe-rioden 1970-1995, hvor det danske timetal per beskæftiget falder, absolut og i forhold til snart sagt alle europæiske OECD lande.

Men hvorfor er disse tendenser »bekymrende«? Set i gennem vores optik er forløbet bekymrende *fordi datakilderne bag timeindsatsen skifter i Danmarks Statistisk, netop i midten af 1990'erne*. Det virker som et bemærkelsesværdigt sammentræf, at den danske timeindsats per beskæftiget udviser en markant andeles trendudvikling efter skiftet i datakilder. Ikke blot i en dansk intertemporal sammenhæng, men så sandelig også i et internationalt perspektiv.

Derfor tog vi et (mildt) forbehold i vores konklusion. For *hvis* der er sammenligningsproblemer med timeindsatsen før og efter kildeskiftet, da er det muligt konklusionen om faldende timeproduktivitetsvækst skal tilskrives disse dataproblemer (her skal det igen betænkes, at væksten i såvel BNP per beskæftiget som per indbygger er stabil over de sidste 5-6 årtier). Hvis den markante stigning i time tallet per beskæftiget er en dataillusion, da er væksten i timeproduktiviteten undervurderet fra 1995 og frem (og muligvis overvurderet i den foregående periode). Dette kan være med til at forklare, at væksten i timeproduktiviteten tilsyneladende ikke revitaliseres, sådan som det er tilfældet for BNP per beskæftiget (og per indbygger).

Omvendt er det naturligvis også muligt, at timeudviklingen er reel. I så fald følger to konklusioner:

A. *Den danske vækst i timeproduktiviteten er løjet af i 1970erne.* Dette »produktivitets slowdown« er den danske økonomi aldrig rigtigt kommet sig over.

Denne konklusion strider imod ØEM (2009), hvor det gøres gældende, at produktivtetsnedgangen opstår i 1990erne. Denne konklusion, som klart og tydeligt er fremført i ØEM (2009), er fagligt set *forkert*.

Der er intet i ØEMs kommentar, der rokker ved denne konklusion.¹ Når vi i Dalgaard og Hansen (2010a) opnår andre resultater end ØEM, så skyldes det slet og ret, at vores analyser baserer sig på formelle statistiske test, der sikrer imod subjektive og tilfældige vurderingerne af udviklingen, mens ØEMs analyse – i denne henseende – alene er af overraskende uformel karakter.

Den anden vigtige konklusion (som vi nok ikke betoner kraftigt nok i Dalgaard og Hansen (2010a) er denne:

B. *Over de sidste 15 år har vi set lidt af et dansk »timemirakel«.* Arbejdsindsatsen per beskæftiget er gradvist steget over tid fra midten af 1990erne.

Som figur 1 antyder, er udviklingen bemærkelsesværdig, særligt relativt til andre rige lande. Dette timefænomen er – i sig selv – et nærmere studie værd. Hvad er der sket i Danmark i 1990erne, som ikke har fundet sted i resten af Europa, og som kan motivere denne unikke danske timeudvikling? Dette forekommer at være et vigtigt spørgsmål for den dansk fokuserede anvendte forskning at afdække, ikke mindst i relation til den finanspolitiske holdbarhedsdiskussion.

Spørgsmål 2. Er den danske vækst aftaget i 1990erne relativt til »lande vi normalt sammenligner os med«?

I Dalgaard og Hansen (2010a) undersøger vi også den danske komparative vækststud-

1. Det skal gerne medgives, at ØEM i kommentaren fremfører følgende metodemæssige kritik af vores analyse: »Væksten i produktiviteten svinger en hel del fra år til år. Dette er et veldokumenteret fænomen og også belyst i Økonomi- og Erhvervsministeriets analyse. Det betyder i sig selv, at det statistiske konfidensinterval om trendudviklingen bliver meget bredt. [...] En måde at komme ud over dette problem på er ved at se på et glidende gennemsnit af produktivitetsvæksten over en længere årrække.« Fagligt set må man anse dette argument for at være svært forståeligt, siden et glidende gennemsnit selvfølgelig også er udtryk for et estimat for den gennemsnitlige udvikling i f.eks. timeproduktiviteten. For at vurdere om der er tale om »brud« vil man under alle omstændigheder skulle konstruere et konfidensinterval, se Dalgaard og Hansen (2010b). Problemet med ØEMs analyse er, at noget sådant ikke udfærdiges. Alle konklusioner er baseret på »øjemål«, hvilket er udtryk for et noget usædvanligt (og betænkeligt) metodevalg.

vikling. Præmissen for vores analyse er, at man bør betragte så lang en periode som muligt for at aflure mulige tendenser i relative væksts-tender. Det vil i praksis sige fra 1950 og frem til i dag. Den anden præmis er, at man bør fokusere sammenligningen på lande, der er strukturelt sammenlignelige med Danmark. Klubben af OECD lande er i den henseende et godt bud.

Med dette fokus kan det konstateres, at den danske vækst ikke har været urovækkende lav. Tværtom, den danske vækst er konsistent med, hvad man ville vente på basis af betinget konvergens, se f.eks. Dalgaard og Kreiner (1999). Ud fra ØEMs nye kommentar må man forstå, at der ikke ... længere; se imidlertid ØEM (2010) sættes spørgsmålstegn ved disse fakta og fortolkninger.²

I ØEM (2009, 2010) hævdes det imidlertid også, at den danske vækst har været lavere end lande der var rigere end Danmark, set over en temmelig meget kortere periode. Dette visualiseres ved udviklingen i Danmarks BNP per indbygger relativt til »Top 5« i OECD over de senere årtier. Desværre har ØEM valgt at analysere udviklingen i den løbende top 5, hvilket metodisk ikke giver nogen mening ud fra et konvergenssynspunkt, se Dalgaard og Hansen (2010b) for en uddybning.

Måske i erkendelse heraf henviser kommentaren i stedet til nylige analyser fra Det Økonomiske Råd (DØR). DØR viser således, at man kan konstruere *andre* figurer med fokus i relativ timeproduktivitet, på hvilken basis man uformelt kan argumentere for et ugunstigt forløb i den danske relative produktivitetsvækst, fra 1990erne og frem.

Som vi læser DØR, drages der nu ikke håndfaste konklusioner i denne henseende. Ganske vist kan man spore en tendens til, at den danske vækst har været lavere end f.eks. Sveriges fra midten af 1990erne og frem, hvilket har givet anledning til, at den danske timeproduktivitet er aftaget med godt 10 % i forhold til Sveriges. Men samtidigt har den danske produktivitetsvækst været hurtigere end andre OECD landes, hvorfor det samlede billede bliver en anelse »mudret«.

Det er værd igen minde om den bemærkelsesværdige udvikling i timeindsatsen (figur 1); ifald timevæksten er overvurderet kan de observerede komparative vækstforskelle i timeproduktiviteten muligvis være forklaret, eller i al fald markant mindsket.

Men lad os skyde denne mulighed til side, og i stedet antage, at timeudviklingen faktisk er, som data antyder. Hvordan skal man da opfatte vækstforskelle á la den danske versus den svenske fra midten af 1990erne og frem?

2. I den nye kommentar fremgår det således, at ØEM erkender, at »Det er velkendt, at fattige lande normalt har højere vækstrater end rige lande som følge af catch-up. DH finder i overensstemmelse med betinget konvergens, at Danmarks BNP pr. indbygger relativt til det gennemsnitlige niveau i OECD er faldet over de sidste 50-60 år.« På dette punkt har debatten altså affødt en gryende konsensus.

Vækstteoretisk vil man tænke som følger. Klubben af OECD lande udgøres hovedsageligt af lande, der er forholdsvis tæt på deres langsigtede steady state vækststi; konvergensprocessen er – disse lande i mellem – mere eller mindre tilendebragt. De observerede indkomstforskelle er beskedne i et bredt internationalt perspektiv og svinger som følge af asymmetriske konjunkturforløb. Eller som følge af asynkrone vækstcykler, om man vil.

Fra OEMs kommentar må man forstå, at ministeriet ikke mener, at væksthforskelle der opretholdes over et årti (eller mere) kan være udtryk for – midlertidige – afvigelser fra væksthrenden, hvorfor slige væksthforskelle må være udtryk for (urovækkende) trend »brud«.

Denne vurdering må alene stå for ministeriets egen regning. Der er *intet* forskningsmæssigt belæg for at konkludere noget sådant. Igen er der tale om en lidt bemærkelsesværdig lemfældig metodeomgang.

Som illustration kan nævnes, at Danmark voksede knap 1% hurtigere – i gennemsnit – end Sverige over perioden 1980-1995, målt ved timeproduktiviteten.³ Dette endskønt Danmarks timeproduktivitet var højere end den svenske i 1980. Denne udvikling indelbar, at den danske timeproduktivitet forøgedes, relativt til Sveriges, med ca. 12% over nævnte periode. I det seneste årti (godt og vel) har vi så set en modsatrettet proces. Det er svært at se, hvordan dette ikke udmærket kan være udtryk for cykliske bevægelser.

Det kan selvfølgelig være, at der er tungtvejende grunde til at lægge særlig vægt på den seneste væksthforskel (der altså er i Sveriges favør), fremfor forgangne væksthforskelle (i Danmarks favør). Men i så fald må man bede om at slige tungtvejende årsager bliver fremlagt. Vi kender dem ikke. Og de er ikke at finde i OEMs kommentar, endside i OEM (2009, 2010).

3. Kilden til dette tal er Penn World Tables 6.3. (serien, »rgdpl2th«).

Litteratur

- Dalgaard, C.-J. og H. Hansen. 2010a. Er der et dansk produktivitetskollaps? *Nationaløkonomisk Tidsskrift*, 148, 1-20.
- Dalgaard, C.-J. og H. Hansen. 2010b. Er der et dansk produktivitetskollaps? – Duplik. *EPRU analyse* nr. 29.
- Dalgaard, C.-J. og C. T. Kreiner. 1999. Global indkomstulighed i et vækstteoretisk perspektiv: En Introduktion til Konvergensdebatten. *Nationaløkonomisk Tidsskrift*, 137, 305-32.
- Grilliches, Z. 1988. Productivity Puzzles and R & D: Another Non-explanation. *Journal of Economic Perspectives* 2, 9-21.
- Økonomi- og Erhvervsministeriet. 2009. Den danske produktivitetsudvikling. *Økonomisk Tema* nr. 8. November 2009.
- Økonomi- og Erhvervsministeriet. 2010. Bemærkninger til analysen: Er der et dansk produktivitetskollaps? De Økonomiske Råd (2010). *Dansk Økonomi – Efterår 2010*. København.

Bog anmeldelser

Svend Erik Hougaard Jensen og Peter Møllgaard (red.). *Design og styring af institutioner*. Festskrift til ære for Niels Blomgren-Hansen i anledning af hans 70-års-fødselsdag. Anmeldt af Birgitte Sloth.

De videnskabelige medarbejdere fra Økonomisk Institut på CBS har udgivet et festskrift til deres kollega, professor Niels Blomgren-Hansen.

Forhåbentlig og formodentlig er Niels Blomgren-Hansen blevet glad og taknemmelig for publikationen. Det er en hyldest fra bidragsyderne til modtageren, der er velfortjent. Som det indledende kapitel af Svend E. Hougaard Jensen og Peter Møllgaard viser, så har Niels Blomgren-Hansen gennem sit 50-årige virke som økonom leveret interessante og relevante bidrag til økonomisk forskning og dansk økonomisk politik og praksis. Niels Blomgren-Hansen har også ydet en stor og anerkendelsesværdig indsats som formidler af økonomisk indsigt og som underviser, hvilket af en eller anden grund ikke omtales i det indledende portræt. Det er nok ikke udtryk for, at forfatterne mener, at undervisning og formidling er mindre betydelige opgaver for videnskabelige medarbejdere ved CBS, men skyldes måske, at alle festskriftets bidragsydere er universitetsuddannede økonomer, og at der alene af den grund ikke er nogen af dem, der selv er blevet undervist af Niels Blomgren-Hansen.

Festskriftet er udgivet af Handelshøjskolen Forlag og fremtræder som en bog med nydeligt omslag i grønt og orange, fortløbende paginering og nogenlunde ensartet typografi. En samlet indholdsfortegnelse opdeler de 15 bidrag i syv dele plus det indledende kapitel om Niels Blomgren-Hansen, der også giver en samlende oversigt over de efterfølgende bidrag.

Det har ikke været let at skrive denne samlende indledning, for det er svært at se bidragssamlingen som et sammenhængende værk. Ved gennemlæsning fremstår den som en samling tekster af blandet indhold og form og med en særdeles uklar målgruppe. Det forhindrer ikke, at der er tale om spændende og interessante bidrag, der viser den store bredde i forskningen på Økonomisk Institut på CBS. Af indledningen fremgår, at instituttets forskning betragter samfundet »som en samling økonomiske institutioner, som bestemmer virksomhedernes konkurrenceevne, borgernes velfærd og politikernes rammer«.

Nogle af bidragene fremstår som egentlige videnskabelige publikationer, der præsenterer ny teoretisk eller empirisk indsigt. Birthe Larsen og Gisela Waisman fra Institute of International Studies ved Stockholms Universitet viser på baggrund af en relevant og kompetent økonometrisk analyse, at indvandrere fra udviklingslande flytter væk fra de svenske kommuner, der er karakteriseret ved negative holdninger over for indvandrere. Det er klart et bidrag til vores viden, og supplerer de resultater, om hvad negative indvandrereholdninger betyder for lønninger, som de samme forfattere har skrevet om andre steder. Også det nye bidrag vil interessere især svenskere og forskere inden for feltet. Imidlertid er der næppe ret mange svenskere eller (internationale) immigrationsforskere, der kommer til at læse i dette festskrift. For festskriftlæserne som f.eks. Niels Blomgren-Hansen og under tegnede ville det have været interessant, hvis de to forskere havde tilføjet en vurdering af, om tilsvarende effekter kan forventes at gælde i Danmark.

Andre bidrag opsummerer og perspektiverer forskningsresultater, som forfatterne har publiceret andre steder. For eksempel skriver Morten Bennedsen og Kasper Meisner Nielsen om generationsskifter og det danske familieejerskab med udgangspunkt i en række yderst spændende forskningsresultater, der

fornuftigt nok er publiceret i *Quarterly Journal of Economics*. De argumenterer for, at der kan være problemer forbundet med generationsskifter inden for familien, men peger også på veje til at løse disse problemer. En tekst med spændende metodiske pointer, der vil være relevant for undervisning på CBS og mange andre steder.

Også andre af bogens kapitler lægger vægt på det formidlingsmæssige og forekommer egnede til studerende og andre medlemmer af den interesserede offentlighed. Eksempelvis Steffen Andersens bidrag om identifikation og eksperimenter i økonomisk forskning, og Anette Booms diskussion af udfordringer ved liberalisering af elektricitetsmarkeder i Storbritannien og Skandinavien. Om end begge disse bidrag ville vinde ved mere mundret og præcist sprogbrug. Det fremgår ikke, men jeg vil gætte på, at Anette Booms bidrag er et af de to, der er blevet oversat til bogen, men i så fald er der desværre ikke tale om nogen særlig fremragende oversættelse.

Redaktørernes arbejde har sikkert ikke været nemt. Det er altid besværligt at få så mange og travle bidragydere til at aflevere bidrag. I et festskrift til en bestemt lejlighed er det jo nødvendigt, at værket bliver færdigt til tiden. Men det ville have pyntet og også gjort bidragene det mere egnede til eksempelvis undervisningsbrug, hvis redaktørerne eller forlaget havde bidraget med en solid faglig og sproglig korrektur. I nogle af bidragene er argumentationen endog særdeles vanskelig at følge, og kunne utvivlsomt have været forbedret ved kritiske kollegiale gennemlæsninger. Specielt et bidrag af Bodil Olai Hansen og Hans Keiding om oligopolteori er tungt. Det er særligt ærgerligt, da det er gjort relevant også for en bredere læserkreds ved at benytte de teoretiske overvejelser til en diskussion af en konkret dansk virksomhedsfusion.

Bogen omfatter en række spændende bidrag af meget varierende art, form og levetid og er desuden rettet mod forskellige målgrupper, der næppe alle vil finde frem til dem i netop denne sammenhæng. Det mest ekstreme eksempel er et bidrag af Bjarne Sloth

Jensen, der synes henvendt til en endog meget international læserkreds af økonomiske specialister. Så – udover det gode formål at hylde Niels Blomgren-Hansen – er det svært at se, at antologien tjener et formål som bogudgivelse. Den kan fungere som en slags uvidet visitkort, der giver et godt og bredt indblik i forskningen ved Økonomisk Institut på CBS og sikrer så også en forholdsvis stor pointsum ved opgørelse af den bibliometriske forskningsindikator, der benyttes ved fordeling af midler mellem danske universiteter.

Selvom der ikke er noget der tyder på, at der har været benyttet systematisk peer review i forbindelse med bogen, udgør bidragene efter min vurdering omkring halvdelen af de publikationspoint, som disse i øvrigt anerkendte og produktive forskere har optjent til den såkaldte bibliometriske forskningsindikator til CBS i 2009.¹ Det siger noget om forskningsindikatoren, der i fremtiden vil blive brugt til at opgøre forskningsproduktionen for de danske universiteter, og som skal indgå i grundlaget for fordeling af forskningsmidler mellem universiteterne. Det er naturligvis ikke noget, man skal bebrejde hverken bidragydere eller redaktører, der blot har ønsket at hylde Niels Blomgren-Hansen. Bogen har utvivlsomt været planlagt, inden det blev klart, at forskningsindikatoren kom til at fungere på denne måde, og det er helt usandsynligt, at forskningsindikatoren har haft betydning i denne sammenhæng. Men hvis den bibliometriske forskningsindikator er kommet for at blive og i stadig højere grad bruges til

1. Information om forskningsindikatoren kan findes på Forskningsstyrelsens hjemmeside, <http://www.fi.dk/forskning/den-bibliometriske-forskningsindikator>.

Hverken Forskningsstyrelsen eller andre myndigheder offentliggør beregninger af indikatorværdierne. Min vurdering af bogens relative bidrag til forskningsindikatoren er baseret på de publikationsoplysninger for de enkelte bidragydere, der er tilgængelige på primært CBS' hjemmeside, og regler om niveaudeling, point, fraktionering m.v., som de fremgår af Forskningsstyrelsens hjemmeside.

fordeling af økonomiske midler måske også internt på universiteter, så har universiteter og institutter klare incitamentter til at lave (mange) flere af denne type publikationer i fremtiden. Det vil næppe være fremmende for forskningen.

SORT ARBEJDE

– sort på hvidt

Danskerne og det sorte arbejde

Camilla Hvidtfeldt, Bent Jensen og Claus Larsen

312 sider, rigt illustreret
i farver, indbundet

148,00 kr.



En samlet kortlægning af det sorte arbejde i Danmark, der bl.a. besvarer spørgsmålene:

Hvor stort er problemet? Hvem vinder? Hvem taber?

*"...den nok mest omfattende undersøgelse af sin art i verden".
Skrev Berlingske Tidende*

Rockwool Fondens Forskningsenhed – Syddansk Universitetsforlag

www.universitypress.dk



Anders Lebeck Ølgaard

5. september 1926 – 25. april 2009



Anders Lebeck Ølgaard

Anders Lebeck Ølgaard

5. september 1926 – 25. april 2009

I

1. Indledning

Til det sidste var professor dr. polit. Anders Ølgaard knyttet til Økonomisk Institut som professor emeritus, og her bibeholdt han sit kontor i næsten 13 år, efter at han som 70-årig gik på pension, og efter at han i sit arbejdsliv på skift havde udfyldt tre professorater:

Nationaløkonomi 1966-1973, samfundsbeskrivelse 1973-1979 og økonomi 1979-1996,

alle ved Økonomisk Institut og i Studiegården.

I sine sidste år arbejdede han blandt andet med at skrive sine erindringer, og godt et år før hans død udkom de i form af et værk på 445 s. med titlen: »*Den syngende vismand*« i det følgende omtalt som *Erindringerne*. At skrive mindeord om Anders Ølgaard og hans virke kan på denne baggrund forekomme overflødig. Hertil kommer de mange sider af hans liv, der allerede er belyst i det festskrift på 600 s., som han fik til sin 70 års dag. Her har mange af hans tidligere elever, kolleger og medarbejdere beskrevet hans pædagogiske metoder, hans forskning, hans virke som vismand og deltager i udredningsarbejde og kommissioner og som orienteringsløber.

Alligevel er tiden ikke inde til et brud med den efterhånden gamle tradition med at offentliggøre mindeord om universitetets professorer og doktorer i dette tidsskrift. Disse mindeord for Anders Ølgaard kan forhåbentlig belyse en mere end 50-årig epoke i politistudiets, Økonomisk Instituts og Studiegårdens historie. Denne epoke berøres først i afsnit 2. *Læreren* og i afsnit 3. *Forskeren*. Det er naturligt og nødvendigt at fremhæve Anders Ølgaards betydning som samfundsdebattør, som landets mest kendte økonom og som folkeoplyser, en indsats som især tog fart, efter han blev medlem af Det økonomiske Råds formandskab, først som almindelig vismand, siden som overvismand. Dette behandles i det følgende afsnit 4. *Vismanden*. Der skal også sættes lys på Anders Ølgaards betydning for Nationaløkonomisk Tidsskrift og Nationaløkonomisk Forening, og det sker i det følgende afsnit 5. *Redaktøren*. Disputatsen fra 1966 er

selvfølgelig et hovedværk i hans produktion. Men den er ikke typisk for kvantiteten af hans skriftlige arbejder. Udover nogle få, men langtidsholdbare undervisningsnoter er den del af hans skriftlige produktion, som han selv står som eneforfatter til, karakteriseret af to typer, nemlig dels artikler, ikke mindst i dette tidsskrift, og dels avis-kronikker. Hertil kommer betænkninger og rapporter samt kommissionsbetænkninger, som ikke nærmere skal omhandles her. De talrige kronikker anså han for at være en meget væsentlig del af sin indsats, og han sørgede for at holde styr på dem via en løbende bibliografering; de behandles i afsnit 6. *Kronikøren*. Endelig skal behandles hans centrale indsats som: 7. *Professoren i samfundsbeskrivelse*.

2. Læreren

Anders blev student fra Statsskolen i Aabenraa i 1944 og flyttede straks til København, hvor han begyndte at studere jura. Skønt han ved enhver given lejlighed fremhævede, at han var sønderjyde, levede han resten af sit liv i hovedstaden. Han tog 1. del på jura med et flot resultat, men var ikke tilfreds med sit studievalg; han fandt juraen provinsiel. Hans tid på jurastudiet skaffede ham dog mange venskaber med andre jurastuderende og senere professorer ved *Det rets- og statsvidenskabelige fakultet*. Og det gav ham også en særlig faglig ballast i de udvalg og kommissioner, som han senere deltog i med liv og sjæl. Han kunne sikkert være blevet en fremragende jurist, men var ikke enig i at:

»På den yderste dag, dommens dag, er det altid juristerne der har magten. Der sidder ingen økonomer i Højesteret eller Europa-domstolen i Luxembourg...«
[Mogens Koktvedgaard, *Festskrift til Anders Ølgaard*, Tillægsnummer til NT 1996 s. 215].

Søgte han magten, var det i hvert fald ikke den form for magt. Efter eksamen aftjente han sin værnepligt ved signaltropperne og endte to år senere som sekondløjtnant ved de danske brigader i det nordvestlige Tyskland. Her styrede han 70 mænd og deres militære køretøjer og udviklede sin evne til at råbe en større forsamling op. I 1948 begyndte så hans liv som polit:

»Polit har jeg været hele resten af mit liv. Bortset fra 1953 og første halvdel af 1954 er jeg regelmæssigt gået op og ned af Studiegårdens trapper, først som student senere som lærer. Siden 1956 har jeg haft mit skrivebord i det kontor, hvor jeg stadig sidder.«
[*Den syngende vismand*, 2008, s. 73].

Allerede efter et års studier tog han 1. del, fik studenterjob på Slotsholmen i Det økonomiske Sekretariat, og efter yderligere tre år blev han i 1953 cand. polit. og fik straks ansættelse i DØS. Her – siger han i *Erindringerne* – lærte han mere økonomi

end på hele studiet. Ja, som så mange andre kendte politter fra dengang og senere havde han travlt med at blive færdig med studiet. Han var meget glad for sit arbejde i DØS, men begyndte på opfordring fra sine tidligere lærere på en livslang undervisningskarriere, først som timelærer lørdag formiddag. Var begyndelsen beskeden, blev det dog hans fremtidige karrierevalg, og han arbejdede fra 1957 fuldtids på universitetet. Egentlig blev han først fastansat, da han blev professor i 1966, så han havde i lange perioder orlov fra sin tjenestemandstilling i centraladministrationen for at forske og undervise.

Hvorfor valget til sidst blev universitetet og ikke centraladministrationen, har han ikke fortalt os. En fornemmelse af, at der var en vis dualitet i hans interesser, illustreres dog – omend indirekte – af to tabelbilag, de eneste bilag i *Erindringerne*:

Bilag 1: Statsministre efter befrielsen og

Bilag 2: Statsvidenskabelige studerende 1950-1996 immatrikulerede, bestand og kandidater

Universitetsverdenen vandt over Slotsholmen. Det var nok afgørende, at universitetet og professorgerningen dengang stod for et frit liv og ikke mindst udstrakt talefrihed. Embedsmænd burde efter Ølgaards mening tie om deres sagsområde. Men på den tid var lektorater kun deltidstillinger, og vejen til at blive professor var lang og forudsatte, at man i en længere årrække skrev på en disputats. Inden Anders Ølgaard forsvandt i disputatsskrivningen, gjorde han en stor indsats for at reformere politstudiet.

Der var kommet en ny anordning for studiet i 1952, så fundamentet til en fornyelse var lagt. Fra efteråret 1957 overtog Anders Ølgaard som lektor undervisningen i nationaløkonomi for alle studerende på 1. del, jurister og politter sammen, nyimmatrikulerede såvel som studerende på 2. og 3. år, en højst utaknemmelig opgave, som han løste med megen entusiasme og kraftudfoldelse og i en munter og kammeratlig tone, usædvanlig for datidens universitetslærere. Hans interesse for studiet fremgår af hans første bidrag til dette tidsskrift. Artiklen: *Hvordan studerer stud. polit'er?* ((NT vol. 95, s. 73-88) refererer en undersøgelse, den første af mange, som han gennemførte af studiet. Han vidste jo fra egne oplevelser, hvor sløjt det stod til med studieaktiviteten for de fleste; erhvervsarbejdet tog megen tid. Det samlede resultat af (1) det usammenhængende pensum, (2) undervisning alene i form af store forelæsninger og (3) uforberedte tilhørere blev da også, at det faglige niveau pressesedes ned:

»Herved opstår der en alvorlig risiko for, at man får ulemperne ved det frie studium – upræcis tilegnelse af det elementære pensum – uden at opnå fordelene – opøvelse til selvstændigt arbejde med stoffet.« [NT 95, s. 87].

Der var behov for reformer. Og fra han i efteråret 1957 debuterede som forelæser i nationaløkonomi på 1.del, arbejdede han på at forbedre undervisningsforholdene, især ved at få indført holdundervisning og få ansat studenterinstruktører. Det lykkedes også at få finansieret en speciel økonomiundervisning for de jurastuderende, og med tiden blev de politstuderende niveauopdelt gennem årsprøver. Disse reformer indrettede i tide politstudiet på den massive vækst i tilgangen af studenter, som indtraf fra 1960, og som han har fremhævet ved bilagstabellen i *Erindringerne*. Stigningen i den årlige tilgang, som i sidste halvdel af 1950erne lå under 30, kan selvfølgelig ikke udelukkende tilskrives hans indsats, men vi var bedre forberedt end mange andre studier på at håndtere masseundervisning.

Han havde umiddelbart mindre held med at revolutionere pensum i den introducerende undervisning i nationaløkonomi. Hans interesse for dette pensum fremgår af nogle boganmeldelser, også i dette tidsskrift. Der er i øvrigt kun få anmeldelser fra hans hånd; som han senere udtrykte det:

»Han skrev, andre måtte læse.«

Citatet kunne passende udbygges således: Og når han talte, måtte andre lytte.

Først anmeldtes en meget formaliseret svensk lærebog, som ikke faldt i hans smag; man burde kunne læse den indledende økonomi *uden at kunne differentiere*, og det kunne ikke alle datiden politter; mange havde ikke haft matematik i gymnasiet, og selve studiet gav ikke matematikundervisning. På den anden side var han heller ikke tilhænger af meget bløde og verbale udredninger. Modeller og ikke mindste grafiske analyser var at foretrække. Det viser hans anmeldelse af den dominerende danske lærebog af P. Nyboe Andersen m.fl., hvor han roser 2. udgaven på bekostning af dens forgænger, som han for det meste havde undervist i. Han fik ikke umiddelbart held til udskifte det sammenstykkede pensum, hvoraf størstedelen var på dansk; det havde han ikke autoritet nok til som lektor. Han var glad for Erich Schneider: *Einführung in die Wirtschaftstheorie*, det var lige hans stil, og han bidrog selv til pensum med den meget holdbare note: *Efterspørgsel efter produktionsfaktorer. Memorandum nr. 1 Københavns Universitets Økonomiske Institut 1958*. Her lykkedes det ham på 45 sider at vise produktionsfunktionens egenskaber, substitution og optimal faktorkombination, sammenhængen til omkostningsfunktionen og dennes egenskaber.

3. Forskeren

Mens hans undervisning blev betalt af universitetet, forskede han for midler fra det amerikanske samfundsvidenskabelige forskningsråd under ledelse af den senere no-

belpristager i økonomi Simon Kuznets. Det drejede sig om at opstille lange makroøkonomiske tidsserier, og det danske bidrag til det internationale forskningsprojekt blev serier for danske im- og eksportpriser offentliggjort som: *Danmarks udenrigshandel 1874-1958*, (med Ole Bus Henriksen) *Studier nr. 2, Københavns Universitets Økonomiske Institut 1960*. Dette arbejde var også tænkt som et forstudie til disputatsen, hvor ændringer i bytteforhold skulle være et hovedtema. Det blev det dog ikke, men kun en tredjedel af værket. Studiet af prisbevægelser og bytteforhold vendte han dog tilbage til senere ved flere lejligheder således i bidrag til tre festskrifter til kolleger publiceret af Nationaløkonomisk Forening, nemlig i *Festskrift til Frederik Zeuthen* i 1958, i *Festskrift til Carl Iversen* i 1969 og i *Essays in Honour of Jørgen H. Gelting* i 1982.

Hovedtemaet blev et andet, nemlig vækstteori, som udgjorde den centrale og første tredjedel af bogen, *Growth, Productivity and Relative Prices*, og som bar hele disputatsen. Hans interesse for økonomisk vækst afspejles i artiklen fra 1958, hvor han blandt andet gennemgik Domars model, men han havde et problem:

»Den danske terminologi, der leder tanken hen på drivhuse, forekommer ikke særligt vellykket. Måske kan man få en ide til en bedre betegnelse fra den danske økonom, der hidtil har interesseret sig mest for disse problemer, professor Hans Brems (hans ansvar i denne forbindelse er ikke blevet mindre efter at han i Nationaløkonomisk Tidsskrift anvendte betegnelsen: Ligevægtsvækstprocenten).« [Problemstillinger i vækstteorien, NT vol. 96, s. 52].

Det er måske vanskeligt i dag at forestille sig, hvor mange uløste problemer man dengang havde, når emnet dynamisk analyse pressede sig på. I en samtidig anmeldelse af den engelske udgave af Frederik Zeuthen, *Økonomisk teori og metode*, finder han under emnet dynamik ikke meget andet end spindelvæv og indkomstmultiplikator. Ølgaards indgangsvinkel er behovet for at se på investeringer som kapitalakkumulation og ikke blot som en komponent i efterspørgslen.

På et Rockefeller stipendium tilbragte han det meste af året 1960 sammen med familien i USA, hvor han forskede ved Massachusetts Institute of Technology og Harvard University. Hans forskning og dermed indholdet i hans disputats blev her stærkt påvirket af tidens ensidige optagethed af vækstmodeller: Som lærere fik han Domar og Solow, hvis vækstmodeller han havde beskrevet i 1958, og hans forskning skiftede retning fra det empiriske til det abstrakte og teoretiske. Efter et år vendte han hjem til Danmark og skrev så i de følgende år på disputatsen, som blev antaget i 1965 og forsvaret i 1966.

Anders Ølgaard var tilfreds med resultatet, han var stolt af disputatsen, som udkom i flere oplag på North-Holland, og som blev anmeldt positivt i en række af de førende,

internationale tidsskrifter. Den blev også indført som pensum på 2. del af politstudiet, men her var den ikke populær. Skrevet som den var med et helt andet formål end det pædagogiske, var den vanskelig læsning. Alligevel sidder der tilbage hos de fleste, der måtte sætte sig ind i den, et nyttig tankeskema til at beskrive produktivitetsudviklingen i bestande over tid. Hvad enten det drejer sig om maskiner, bygninger, boliger eller arbejdere kan produktivitetsudviklingen over tid spaltes i tre effekter: *års*, *årgangs-* og *alderseffekter*.

Han blev nu udnævnt til professor i 1966, og disputatsen tjente derfor sit formål. Mens selve udarbejdelsen af monografier over lange perioder næppe tjente til at opnå de største fremskridt inden for forskningen, så holdt det i hvert fald antallet af ansøgere til professorater nede. Og det var kun hans første professorat.

4. Vismanden

Samtidig med, at han omsider kunne forsvare sin disputats, blev Anders Ølgaard også vismand efter Poul Winding, som kun nåede at være det i et år, før han pludselig døde. Anders Ølgaard var vismand fra 1966 til 1968, derefter var han rådgiver for regeringen i Malaysia, og ved sin hjemkomst blev han overvismand i periode 1970-1975. Det var uden tvivl en meget vigtig periode i hans tilværelse. Han kom tæt på, hvad der var oppe i tiden af økonomisk-politiske temaer. Han kunne påvirke den økonomiske politik, måske ikke så meget via politikerne som via befolkningen, idet han havde fået en platform, hvorfra han kunne udbrede sin pædagogik og forklare indviklede økonomiske problemer lige så underholdende, som når han underviste politter. Desuden slap han fra disputatsarbejdets krav om enkeltmandspræstationer. Han fik et sekretariat bestykket med de mest energiske og ambitiøse unge økonomer, og i modsætning til traditionen i Det økonomiske Sekretariat, hvor han var opdraget, havde sekretariatet i Det økonomiske Råd dengang en meget flad struktur uden hierarki, hvor yngste sekretær havde lige taleret og forventedes at bidrage på lige fod med de ældre ud fra den tanke, at de yngste måske var de mest veluddannede.

Rapporterne og det øvrige arbejde i Det økonomiske Råd har han beskrevet i *Festskrift til H. Winding Pedersen* (1977 NT vol. 115 s. 122-156) samt i Rådets 25 års jubilæumsskrift, som er supplement til *Dansk økonomi december 1987*, her sammen med en række andre bidrag, som beskriver vismandsinstitutionens historie.

Desuden har han jo også behandlet denne tid i *Erindringerne*. Her fremhæver han arbejdet med SMEC I og ikke mindste det interessante arbejde med SMEC II. Vi er nu ved den første energikrise i 1973, og vækstteoriene er uaktuelle. Men traditionelle keynesianske tanker og empiriske modeller giver sekretariatet et trumfkort som den første institution, der forudsiger arbejdsløshedens tilbagekomst i 1974 efter den lange periode med vækst og faldende ledighed siden 1956. I relation til modelarbejde var det

ikke økonometrien, der interesserede ham, men modellernes anvendelsesmuligheder. Han var glad i tal, men man skulle helst kunne se argumentet i en figur uden brug af mere avancerede statistiske metoder.

5. Redaktøren

Nationaløkonomisk Tidsskrift har i sin lange levetid kun haft lidt mere end en håndfuld redaktører. Anders Ølgaard har sænket gennemsnittet for redaktørernes funktionstid, men var hans periode end kortvarig, så satte han tydelige spor på tidsskriftet, og sporene var holdbare. I 1963 afløste han Poul Milhøj som redaktør, og blev selv afløst af Thorkild Davidsen ultimo 1966, altså lige efter at han var blevet vismand. Ligesom i forhold til reformen af politstudiet målte han også sin succes som tidsskriftsredaktør på deltagerkredsens størrelse, her målt på antal abonnenter, som det lykkedes ham at forøge med 200 per år i sin redaktørperiode. Reformen af indholdet byggede på to mål. For det første skulle manuskripterne være mere gennearbejdede, og for det andet skulle indholdet kunne læses af selv lidt ældre økonomer. Læser man i dag artiklerne fra 1950'erne, så er der mange lange, som kunne have vundet ved en forkortelse; dette gælder også hans egne. Det er klart, at forfatterne havde overtaget i forhold til redaktøren før Ølgaards redaktørtid; professorer og ministre var vænnet til at publicere deres taler i ret uredigeret form. Det satte han en stopper for. I *Erindringerne* fremhæver han de tilfælde, hvor han tabte kampe om ortografi og tegnsætning, men det var kun over for sine ældre kolleger. Vi unge debuterende forfattere fik sat skik på vores bidrag, selv om det skete under mismod og tårer.

Som redaktør var han med helt nede i trykkeriet, og han bekæmpede hele livet trykfejl.

I 1982 blev Anders Ølgaard medlem af bestyrelsen for Nationaløkonomisk Forening og året efter blev han foreningens formand. I hans formandsperiode, der varede til 1988, havde han et tæt samarbejde med Thorkild Davidsen, blandt andet fik de udarbejdet den meget nyttige *Indholdsfortegnelse 1873-1989*, til tidsskriftet, som gav lettelser i arbejdet for mange.

Thorkild Davidsen, som var ansat på det kongelige bibliotek, arbejdede længe på et manuskript om det statsvidenskabelige studiums historie, som skulle indgå i en serie af jubilæumsskrifter til fejring af, at Københavns Universitet blev 500 år i 1979. Men Thorkild Davidsen døde pludseligt i 1999 midt i arbejdet med færdiggørelsen af dette værk. Straks påtog Anders Ølgaard sig det betydelige og også vanskelige arbejde med at færdiggøre manuskriptet sammen med en mindre gruppe af kolleger. Derfor kunne omsider universitetshistoriens bind VI.B, *Statsvidenskab*, endelig udkomme i 2001.

Anders Ølgaard var glad for og stolt af sit arbejde for Nationaløkonomisk Forening. I *Erindringerne* konstaterer han også med tilfredshed en fortsat stigning i antal-

let af foreningens medlemmer i hans periode. Så helt fortjent blev han ved foreningens 125 års jubilæum i 1998 udnævnt til *æresmedlem*.

6. Kronikøren

En vigtigt side ved det at være vismand var, at det gav Anders Ølgaard materiale til mange kronikker. Rapporternes indhold skulle udbredes og yderligere pædagogiseres, og sekretariatets medarbejdere skrev med, og så omsatte man i fællesskab honoraret til middage. En meget vigtig medspiller var Arne Mikkelsen, som ligesom Anders skrev sine udkast med fyldepen, men mens Arne Mikkelsen brugte gult folio conceptpapir med bred margin, skrev Anders altid på kvadreret A4. Ingen af dem gik ind for mere moderne skriveredskaber eller -metoder, dog blev Anders som pensionist glad for sin PC, for så kunne han sende sine kronikudkast direkte ned i kronikredaktøernes e-mailpostkasser.

Han var også for så vidt angår kronikker interesseret i at tælle dem. I *Erindringerne* beskriver han Kjeld Phillips politiske karriere noget kritisk men:

»Derimod var han en populær kronikør. I Politikens 'scrapbog'... i anledning af Politikens 100 års jubilæum i 1984, optræder han i hitlisten over kronikører med 108 kronikker. Selv sneg jeg mig lige akkurat ind med 26.« [*Erindringerne* s. 188].

Vil man have en oversigt over hans egne kronikker, har han forsynet os med materialet op til 1991, da han skrev: »*Hvad jeg skrev, det skrev Anne Lise og Else*« Blå memoserie 184, Økonomisk Institut, Københavns Universitet 1991. Også for årene efter 1991 har han efterladt os en liste i form af en udskrift fra universitetets publikationsarkiv. Udover bidrag til bøger og fire artikler i Nationaløkonomisk tidsskrift har han skrevet over 50 kronikker 1991-2005 i de store dagblade, flest i Politiken og Berlingske Tidende. Emnerne lader sig ikke her opregne, dog er den sidste del af perioden meget præget af temaet: *Indvandring*.

7. Professoren i samfundsbeskrivelse

I 1973 skiftede Anders Ølgaard fra nationaløkonomi til samfundsbeskrivelse, og igen i 1979 skiftede han, nu tilbage til økonomi. Man skal ikke være blind for, at hans skift mellem lærestolene på Økonomisk Institut i nogen grad bundede i strategiske overvejelser. Han forsøgte at påvirke professorforsamlingen ved selv at byde ind på jobbene.

Skiftet mellem hans andet og hans tredje professorat havde minimal betydning for hans aktiviteter. Hverken hans undervisning eller hans skriftlige produktion ændredes. Derimod tog han udnævnelsen som professor i samfundsbeskrivelse helt bogstaveligt,

og hans undervisning blev i hele resten af hans tid som professor koncentreret om samfundsbeskrivelse på 1. årsprøve. Der var ikke megen konkurrence i lærerforsamlingen om at passe den elementære undervisning, så han fik lov, så længe han ville. Anders Ølgaard var den store forelæser og nød det; jo flere tilhørere, jo bedre. Han ville have sine budskaber ud, det gjaldt om at fange det store publikum, og på politstudiet var det derfor altid de nyeste studenter, som fik fornøjelsen af hans undervisning. I samfundsbeskrivelse sørgede han med omhu for vedligeholdelsen af pensum og udarbejdelse af opgaver. Lyn- og hjemmeopgaver såvel som eksamensopgaver tog han et personligt ansvar for. Da han blev institutbestyrer for Økonomisk Institut i 1985 delte han undervisningen i samfundsbeskrivelse med P.C. Matthiessen – et semester til hver – men overlod i øvrigt ikke sit fag til andre.

Han ansatte selv sine undervisningsassistenter i samfundsbeskrivelse, dog ikke uden en del spektakel. Da studenteroprøret rullede over universitet i 1968 var han i Malaysia som ulandsekspert og rådgiver, og da han kom hjem var tingene meget forandrede. Gennem *Politrådet* havde de studerende fået magt og indflydelse, og ligesom mange andre jævnaldrende professorer var han dybt bekymret over tingenes udvikling. I nationaløkonomi var det ret let at holde stand mod den marxistiske bevægelse – dog fik Joan Robinsons bog om *Marxistisk økonomi* en fod indenfor på 1. årsprøve uden at blive populær. Samfundsbeskrivelse, derimod, blev en kampplads, hvor krigen stod om pensum og ansættelsen af undervisningsassistenter, en krig som Anders Ølgaard gik mere op i end hans kolleger, hvor specielt de nu mange yngre, ikke-professorale lærere tog tingene mere afslappet.

8. Afrunding

Kampene med studenterne i slagene om faget samfundsbeskrivelse er udførligt beskrevet i *Erindringerne*. Heri er som helhed ikke afsat plads til ret mange krige eller personopgør. Der er ikke noget, der skal hævnnes, og kun få personer er beskrevet lidt hårdt. Derimod fortæller Anders Ølgaard om sit lange og indholdsrige arbejds- og familieliv, hvor politstudiet, Studiegården og Nationaløkonomisk Tidsskrift var de faste holdepunkter i arbejdslivet, mens professorstillinger gav talefrihed og frihed til at deltage i samfundsdebatter.

Man kan overveje, om hans livsbane var blevet meget anderledes, såfremt hans moster Bodil, som husede den nybagte student, havde boet i Aarhus i stedet for på Østerbro, og såfremt Anders var begyndt at studere statskundskab i stedet for jura og statsvidenskab. Hans store og vedholdende interesse for politik og samfundsproblemer ufortalt, så var han dog i bund og grund økonom, når man hermed mener en, der tænker og arbejder i tankeskemaer, modeller om man vil. Og blandingen af interesse for

praktiske problemstillinger og anvendelse af teorier har altid været den sande statsvidenskabers kendetegn.

Men statskundskaben har for længe overhalet statsvidenskaben, i hvert fald kvantitativt, hvad enten man ser på universiteterne eller på Slotsholmen. Universitetslærere har fået andre, og mindre frie arbejdsvilkår. Nu er også Studiegården tømt for økonomer, og Anders Ølgaard er her heller ikke mere. En epoke er forbi.

Ellen Andersen

Del II påbegyndes næste side ...

II

Med Anders Lebeck Ølgaards død d. 25. april 2009 sluttede en sjældent helstøbt karriere begyndende med studieophold på Harvard og professor som 36-årig og sluttende med velanmeldte erindringer, der udkom kun et årstid før hans død som 82-årig. Han kombinerede en akademisk karriere, hvor han nåede de højeste poster inden for dansk økonomi (leder af Københavns Universitets Økonomiske Institut 1985-1993, formand for Nationaløkonomisk Forening 1983-1988, et af denne forenings meget få æresmedlemmer fra 1998 og overvismand 1970-1976) med en folkelig popularitet, der ikke er set siden L.V. Birck spillede en tilsvarende rolle i begyndelsen af det 20. århundrede.

Hvis man ikke havde fulgt karrieren nøjere, kunne man tro, at dette var nået ved hårdt arbejde og håndfast karriereplanlægning. Og der er ingen tvivl om, at han i mange perioder arbejdede hårdt, men »karriereplanlægning« synes at være et helt forkert ord at bruge i forbindelse med Anders. Han var snarere præget af en engageret kamp for de idealer, han havde fået indpodet med sin sønderjyske opvækst. Få akademikere har i samme grad som ham været præget af idealer og rødder. Når han diskuterede EU, Færøerne eller Grønland, så var det hele tiden med opvæksten i Sønderjylland som baggrund.

Han vil af adskillige tusinder af cand.polit'ere først og fremmest blive husket som læreren. Han dyrkede ikke avanceret undervisning for særligt interesserede. Det er begrænset, hvor mange speciale- og ph.d.-studerende, han har været vejleder for. Men han gav næsten alle polit-studerende fra 1950erne til op i 1990erne en introduktion til de økonomiske grundbegreber, som de aldrig glemte, dels som forelæser i introduktionskurser i nationaløkonomi og dels som forelæser i samfundsbeskrivelse.

Hans forelæsninger var langt fra afpuddede og strømlinjede, og afrundede lærebøger var heller ikke hans stil. Han har næppe nogensinde været på et pædagogisk kursus. Hans disputats, der gennem en årrække blev brugt som lærebog i vækstteori, var måske den bog fra hans hånd, som flest studerende husker, men den var langt fra typisk, og han underviste næppe nogensinde selv i den. Den var ikke skrevet som lærebog og heller ikke oplagt egnet som sådan. Det, der viste ham som den store pædagog, han var, var kortere notater og geniale pædagogiske tiltag. Mange fik et blik for sammenhænge i mikroøkonomien ud fra hans lille hæfte om »Efterspørgsel efter produktionsfaktorer«, og første gang, de studerende i 1960erne mødte makroøkonomien og Keynes-modellen

på algebraisk form, var i hans stencilerede notat om multiplikator-teori, der var en supplerende note til P. Nyboe Andersen, Bjarke Fogs og Poul Windings grundbog i nationaløkonomi.

For økonomistuderede uden matematik og kendskab til funktioner af flere variable var produktionsfunktioner, der var homogene af første grad og de af den udledte isokvanter, dyb mystik, indtil Ølgaard fik en stukkatør til at fremstille en 3-dimensionel gipsmodel af en produktionsfunktion med 2 produktionsfaktorer og produktet ud af den 3. akse. »Gipsmodellen« blev et centralt pædagogisk værktøj ved polit-studiet i flere årtier, indtil obligatoriske matematikkurser gjorde den overflødig.

En tilsvarende visualisering var en kasse fremstillet til Nordisk Nationaløkonomisk møde i 1958. Her illustreres udviklingen i produktion, arbejde og kapital i en åben kasse, hvor de observerede punkter var bordtennisbolde holdt på plads af ståltråd. Man så tydeligt multikollinearitetsproblemet. Punkterne bestemte ikke en 3-dimensionel flade, som de skulle, hvis man skulle kunne fastlægge produktionsfunktionen. Bordtennisboldene lå på en ret linje, fordi kapital og arbejdskraft voksede i samme takt.

Hans uortodokse pædagogik bestod også ofte i at udlove guldbajere til de studerende, der kunne komme nærmere til en afklaring af uklarheder og paradokser i pensum. Næppe en metode, der ville vinde anerkendelse i vore dages mere puritanske og politisk konkrete undervisningsmiljø. På den måde blev respekten for teori, præcise begreber og virkelighedens størrelsesordener banket ind i hovedet på de nystartede økonomistuderede. Der, hvor han var i sit es, var ikke i små specialiserede studiekredse, men ved de store forelæsninger, hvor han med sin stemmepragt (dels optrænet som befalingsmand for en lastbilafdeling i den danske brigade i Tyskland lige efter krigen, hvor ordrene skulle kunne høres selv om lastbilerne startede, og dels udviklet fordi han selv var lidt tunghør) og sit engagement kunne fastholde interessen hos hele årgangens flere hundrede studerende.

Det betød ikke, at han var en helhjertet talsmand for forelæsninger. Det var i høj grad ham, der fik indført holdundervisning parallelt med forelæsningerne både i nationaløkonomi og samfundsbeskrivelse. Og valget af de rigtige holdlærere var noget, der lå ham meget på sinde. I nationaløkonomi skulle det være de bedste ældre studerende og i samfundsbeskrivelse yngre fuldmægtige fra de statistikproducerende og statistikanvendende institutioner.

Det gav anledning til intense kampe med studenterrådet op gennem 1970'erne, hvor vi sad sammen i studienævnet. Han var den, der stod for idealerne; jeg skulle så lave kompromiserne. Studienævnet ved polit-studiet bestod på den tid af 3 »sorte« lærere (dvs. fra SFs pragmatiske fløj og ud mod højre) og en »rød« lærer, og så 3 »røde« (dvs. marxistiske) studerende og en moderat. Kunsten var så at få den røde lærer til at føle sig mere som lærer end som rød, og den moderate studerende som mere moderat end studerende.

Kampene stod om både pensum (der var betydelige klumper af både marxisme og teorihistorie i nationaløkonomi og mere socialt farvede fremstillinger i samfundsbeskrivelse) og læreransættelserne (de røde studerende kæmpede for marxistiske lærere). Når det lykkedes at holde de »progressive« elementer på et rimeligt niveau, var det i høj grad Ølgaards fortjeneste – det var ham, der holdt os andre fast på, hvor der ikke kunne gås på kompromis med fagligheden.

Universitetspolitik spillede i det hele taget en central rolle i Anders Ølgaards liv fra 1950erne til hans afgang som institutleder i 1993. Da studentoprøret nåede polit-studiet i 1968, og der begyndte forhandlinger med aktivistiske studerende rundt omkring på de forskellige studier, fastholdt Ølgaard, at man ikke kunne forhandle med aktivister uden et veldokumenteret mandat. Der kom derfor, modsat ved alle andre studier, et ekstraordinært valg til studenterrådet ved polit-studiet i foråret 1968, hvor aktivisterne (med Christian Hjorth-Andersen og jeg selv som spidskandidater) vandt en overlegen sejr over det siddende studenterråd. Episoden er karakteristisk for Ølgaards krav om legitimitet og overholdelse af demokratiske spilleregler. Han krævede, at det vælgermæssige mandat var i orden. I de følgende forhandlinger kom Ølgaard ikke til at spille nogen rolle, da han fra maj 1968 til 1969 var i Malaysia.

Efter hjemkomsten var han i begyndelse af 1970erne økonomernes repræsentant i konsistorium, hvor han sammen med andre professorer førte en stærkt kritisk linje over for Mogens Fogs studentervenlige politik. Det, han først og fremmest vil blive husket for her, er dog nok en langt mere aggressiv forhandlingsstrategi end tidligere ved de årlige budgetforhandlinger i konsistorium. Varmt støttet af de yngre lektorer Birgit Grodal og Nils Groes kæmpede han med ord og tal for flere penge til polit-studiet, modsat de ældre økonomiske professorer der tidligere havde haft de offentlige udgifters vækst og den økonomiske overophedning i baghovedet, og dermed overlod de store stigninger i budgetterne til andre og mindre ansvarsbevidste fag. De fleste af min generation af københavnske universitetsøkonomer fik ansættelse på KU i nye stillinger, som var et resultat af Ølgaards kampe i konsistorium. I 1975, hvor jeg blev lektor, besatte Økonomisk Institut en 5-6 nye lektorstillinger.

Økonomisk Institut ledede han 1985-1993 som institutbestyrer. For mange stod han som talsmand for en mere anvendelsesorienteret og mindre international linie, end hvad der blev moderne. Det er langt fra den fulde sandhed. Han var i hele den periode, hvor Økonomisk Institut blev et internationalt orienteret forskningsinstitut med hovedvægten på teoretisk økonomi, en magtfuld professor og i en stor del af perioden institutbestyrer. Det er klart, at han lagde vægt på relevans og på, at man ikke glemte virkeligheden. Men han havde også sans for kvalitet og for, at man skulle vælge den bedste ansøger uafhængigt af forskningspolitiske forskelligheder. Når han flere gange skiftede professorat, var det utvivlsomt bl.a. led i personalepolitiske kabaler, eller for

at sige det rent ud for at blokere for andre nærliggende kandidater. Men disse andre kandidater, der derved blev undgået, var aldrig estimerede teoretikere. Hans ledelsesstil var langt fra den bulldrende stil, som han af nogen blev anset for at have. Han var i høj grad en kompromissøgende pragmatiker. »Jeg er jeres leder, derfor følger jeg jer« sagde han som institutleder engang, hvor han var uenig i en beslutning, der var flertal for.

Rent videnskabeligt er Ølgaards produktion præget af tidens krav, som var nogle ganske andre end i dag. Perioden sidst i 1950erne og først i 1960erne var præget af fremkomsten af den neoklassiske vækstteori med bl.a. Solows artikel fra 1956. Vækstteori var det, der greb alle periodens makroteoretikere som en vækkelse. Og Ølgaard var med fra starten; i 1958 skrev han en artikel om vækstteori i *Nationaløkonomisk Tidsskrift* med den ofte senere citerede formulering om denne teori:

Alle de problemer af mere kortsigtet karakter, som nødvendigvis opstår samtidig, er ladet ude af betragtning. Imidlertid kan man jo ikke samtidigt stange ål og gå på harejagt, og spørgsmålet er vel, om man ikke tilstrækkeligt længe har stanget de Keyneske ål, så det snart kunne være på tide at prøve at jage harer på vækstens udstrakte vidder; ellers risikerer man måske at glemme, hvordan det her ser ud. [Ølgaard, 1958].

I 1960 kom han på studieophold på MIT og Harvard med Evsey Domar og Robert Solow som lærere, kolleger og venner (Solow bidrog f.eks. med en artikel til festskriftet i anledning af Ølgaards 70-årsdag). Han var altså i makroteoriudviklingens brændpunkt. Havde det været i dag, vil det utvivlsomt have resulteret i en række tidsskriftsartikler i førende internationale tidsskrifter. Men det var ikke den danske tradition i 1960erne – da var det store mål helt indiskutabelt en doktordisputats. Og en sådan skrev han på frem til indleveringen i 1965. Det blev bogen *Growth, Productivity and Relative Prices* på 552 sider, der udkom på North Holland i 1966. Som danske disputatser ofte er, var det adskillige års arbejde med flere forskellige ideer, og for en udlænding uden kendskab til det danske disputatssystem måtte en sådan bog forekomme ikke så lidt underlig, men den var betydningsfuld nok til at blive anmeldt i de førende internationale tidsskrifter. Men altså uden forståelse for, hvordan en dansk disputats måtte komme til at se ud. Christopher Bliss starter sin anmeldelse i *Economic Journal*:

It is not at all clear what Dr. Ølgaard is aiming to do in this volume. It is divided into three sections: one-sector models, two-sector models; and relative prices and terms of trade. The connection between the first two sections and the last is, while not existent, at least very slight. [Bliss, 1968, s. 416].

Hans Brems peger meget præcist på problemstillingen i sin anmeldelse i *Econometrica*:

In general, one might perhaps be pardoned for saying that Scandinavian writers would be well advised to submit themselves to the normal screening and editing procedures of the better journals and in return reach the readership they really are aiming at, rather than clinging to an outmoded dissertation form. Precisely because Scandinavian dissertation standards are higher than American or German ones, why risk confusion with the latter and why not make it easier for the rest of us to enjoy the results? [Brems, 1972, s. 217].

Brems har givetvis ret i, at Ølgaards resultater kunne have resulteret i en række bemærkelsesværdige tidsskriftsartikler, der, hvis de var kommet i perioden 1960-65, ville have vakt international opmærksomhed. Oversigtsartikler vedrørende én- og to-sektorvækstmodeller, opdelingen af økonomiske variable i års-, årgangs- og alderseffekt og bytteforholdsanalyserne i part III kunne hver for sig have båret vægtige tidsskriftsartikler, men nu var det alt sammen skjult i udvalgte afsnit blandt de 552 sider, hvor kun få af de interesserede specialister fandt dem. Ølgaards internationale ry som vækstteoretiker og nationalregnskabsanalytiker kunne have været et helt andet, hvis han ikke havde været bundet op på den danske disputatstradition. Nu kom kun dele af part III i en artikel i *Review of Income and Wealth* i 1981; i øvrigt Ølgaards eneste artikel i et etableret internationalt tidsskrift.

Selv om Ølgaard således var den, der mere end nogen anden introducerede den neoklassiske vækstteori i Danmark, var han ingen inkarneret neoklassiker. Som de fleste andre danske økonomer var hans forhold til de internationale skoler lidt som sommerfuglens til blomsterne. De har flakket fra skole til skole og har taget det, de mente at kunne bruge. Næsten umiddelbart efter, at Ølgaard skrev disputats om neoklassisk vækstteori, blev han økonomisk vismand, og dette virke skete i høj grad med udgangspunkt i keynesianske synspunkter, hvilket kom til udtryk i en lang række policy-anbefalinger og i strukturen i de økonometriske modeller, der i hans formandstid blev bygget i Det økonomiske Råd, se Hansen og Paldam (1973), og Rosted, Schaumann og Sørensen (1975).

Ølgaards andet videnskabelige hovedværk kom også langt fra den videnskabelige alfarvej. Det var den 255 sider lange bog *The Danish Economy*, der blev udgivet af Europakommissionen som nr. 14 i serien *Studies. Economic and Financial Series* i 1980. Ølgaard var i 1973 blevet professor i samfundsbeskrivelse, og samtidig var Danmark blevet medlem af EF, som EU hed dengang. Der var udkommet et stort værk med beskrivelse af de forskellige medlemslande, og nu skulle Danmark så også have et bind. Det blev Ølgaards opgave at skrive det. Ølgaards mål med bogen var at vise, hvad samfundsbeskrivelse »efter min mening skulle være, og hvordan det skulle realiseres«.

Ølgaard, (2008, s. 340). De andre nye medlemslande (UK og Irland) kom med deres bind i 1973, men havde heller ikke Ølgaards ambitioner. De lod f.eks. forskellige forfattere skrive om de forskellige erhverv.

Ølgaard var fuldt beskæftiget med alt muligt andet (overvismand, universitetspolitik, ny undervisning og diverse udvalg) og havde altså ambitioner om at gøre bogen til et slags programskrift for sit nye fag samfundsbeskrivelse, der gjorde op med fagets tradition. Tidligere havde faget heddet Danmarks Statistik og havde levet tæt op til navnet. Ølgaards mål, som han gennemførte i sin undervisning, var snarere at gøre faget til beskrivende økonomi. Bogen overholdt derfor ingen af de givne deadlines; det var en yndet joke i sidste halvdel af 1970'erne at fortælle, at Anders var på vej i EF-fængsel for kontraktbrud.

Da bogen kom, var det igen en bog med mange ideer, der kunne være blevet til interessante forskningsbidrag – problemstillingerne fra disputatsens part III blev igen taget op, og en anden af Ølgaards kæpheste, definitionen af pengebasen og pengemængdebegreberne, blev dyrket. Bl.a. baseret på arbejdet i Det Økonomiske Råd var det også en oversigt over Danmarks økonomiske historie i efterkrigstiden. Viby Mogens karakteriserer således fuldt fortjent bogen på følgende måde i sin oversigt over dansk økonomisk-historisk forskning:

Dette arbejde lægger ganske vist i sit udgangspunkt hovedvægten på den danske økonomis aktuelle strukturproblemer, men i sin interesse for netop at afdække de mere robuste problemer går bogen ofte flere tiår tilbage i tiden. Den bliver derved i virkeligheden den hidtil nyeste generelle oversigt over dansk økonomis historie i det sidste halve århundrede. [Mogens, 1987, s. 119].

Et afsnit af bogen er da også optrykt i den bog, Gunnar Persson udgav i 1993 med artikler om Danmarks og Norges økonomiske historie, (i øvrigt sammen med et afsnit fra Ølgaards disputats).

Alle bogens fordele til trods blev den naturligvis ikke nogen succes. Man kigger jo ikke efter videnskabelige begrebsdiskussioner eller dyberegående historiske analyser i en bog, der fra bestillerens side er tænkt som en let tilgængelig oversigt over et nyt EU-medlemslands økonomiske struktur. Og de, der så efter en hurtig oversigt over dansk økonomis struktur, blev givetvis forvirrede over at møde lange analyser af bytteforholdsmålinger og dobbeltdeflateringsproblemer.

Ølgaard har skrevet mange videnskabelige artikler bl.a. i *Nationaløkonomisk Tidsskrift* (17 større artikler i dette tidsskrift samt nogenlunde lige så mange boganmeldelser og kommentarer) med teoretiske analyser af f.eks. beholdningsefterspørgsel (1965) og med analyser af udvalgte sektorer, bl.a. landbruget (1987) og boligsektoren (1969, 1977, 1979, 1983, 1988, 1989), men man undervurderer hans videnskabelige

indsats, hvis man overser, at hans to hovedværker ikke fik nogen videnskabelig set optimal fremstilling, fordi det ene skulle falde inden for den danske disputatstraditions rammer, og det andet passe til en bestilling fra EU-kommissionen. Formidlingen af hans videnskabelige resultater blev derfor i mindre grad direkte til den videnskabelige verden, men snarere til hans studerende og den danske økonomiverden, der læste rapporter, betænkninger og faglige debatindlæg, hvor han trak på værkerne.

Og langt den største del af Anders' produktion var faglige artikler, betænkninger, rapporter og kronikker. Som overvismand i 6 år havde han ansvaret for 12 halvårsrapporter, desuden var han medlem af eller formand for en lang række udvalg og kommissioner. Både nationale og internationale, f.eks. Verdensbankens Kuwait kommission i 1961, Det rådgivende udvalg for Færøerne 1964-76 (som formand), udvalget om inflationens årsager 1965-66, Harvard Development Advisory Service i Malaysia 1968-1969, udvalget om samfundsvidenskabelig basisuddannelse 1975 (formand), støttegruppen for Grønlands Landsråds Erhvervsudviklingsråd 1976-1979 (formand), besparelser på lægemiddelområdet 1980 (formand), kommissionsdomstolen vedrørende P&T 1982-85, udvalg om monopolloven 1984-1986, udvalg om boligmarkedet 1987-1993, (formand). Dette affødte ud over betænkningerne en lang række kronikker og artikler i fagblade. Han har i alt skrevet omkring 100 kronikker i de store danske blade.

De er ingen tvivl om, at han først og fremmest vil blive husket som overvismand. Han var overvismand 1970-1976, dvs. i en periode med oliekrise, EF-afstemning og et monopol-fjernsyn, der så tungere på public service forpligtelsen end senere tiders konkurrenceudsatte medier. Derfor er han utvivlsomt den af samtlige de 29 vismænd, der har været i Det Økonomiske Råds historie, der har været mest kendt i den brede befolkning. Selv 25 år efter han var holdt op som vismand, kunne man opleve som siddende formandskab, at når man gik med Anders på gaden, vendte folk sig om og bemærkede »jamen er det ikke vismanden, der går der« og så var det altid Ølgaard, hvis udseende eller stemme de genkendte, ikke det siddende formandskabs.

Han fik utvivlsomt også helt afgørende indflydelse på vismændenes virke. Det Økonomiske Råd blev startet i 1962, men arbejdet det første årti foregik på en noget anden måde end i dag. I starten arbejdede man en hel del på de indre linjer med f.eks. lukkede møder med ministre. Det fremgår bl.a. af Arne Mikkelsens beretning om aflysningen af militærets efterårsmanøvre 1964, Mikkelsen (1987 s. 191-192). Hvert år blev alle danske hjemsendte værnepligtige indkaldt til øvelse. Derved blev de jo i disse uger trukket ud af arbejdsstyrken, og der var i 1964 arbejdskraftmangel. Derfor foreslog regeringen i en pressemeddelelse d. 23. juni »bl.a. tilskyndet af samtaler med Det Økonomiske Råds formandskab« en række tiltag bl.a. en aflysning af efterårsmanøvren det år. Denne tradition er helt forladt i dag. Vismændene holder stadig møder med finansministeren og økonomiministeren, men nu typisk efter det halvårslige råds-

møde, hvor de synspunkter, der er kommet til offentlighedens kendskab i forbindelse med rådsmødet, diskuteres. I dag kommer der ikke intern rådgivning til regeringen uden om rådet og offentligheden. Denne politik har ført til, at de økonomiske vismænd i dag opfattes som troværdige og uafhængige af regering og politik. Dette skift fra at være en delvis intern regeringsrådgivende institution til at være en fuldt uafhængig institution uden kontakt med regering og politiske partier blev i høj grad gennemført i Anders Ølgaards tid som overvismand.

Ud over de faglige poster han faktisk kom til at bestride, så gjorde hans folkelige yndest og klare synspunkter ham til en oplagt kandidat til poster uden for den økonomiske verden. Han var kandidat til posten som generaldirektør for Danmarks Radio i 1976 mod TV- direktør Laurits Bindsløv, i valgkamp som formandskandidat for FDB i 1986 (mange mente, at han i denne valgkamp »førte med et mulehår«, Vendsyssel Tidende 5/3 1983, men han endte med at tabe til Bjarne Møgelhøj). I 1980 søgte Jyllandsposten at forudsige, hvordan ministerlisten for en VK-regering, der kunne afløse Anker Jørgensen, ville komme til at se ud. De havde Henning Christophersen som statsminister, Poul Schlüter som udenrigs- og økonomiminister og Anders Fogh Rasmussen som boligminister. Avisen skriver om finansministerposten, at

Knud Engaard (V) og Palle Simonsen (K) er som finansudvalgsmedlemmer og erfarne parlamentarikere gode bud på finansministerposten, men det erfarer at tidligere overvismand, professor Anders Ølgaard, forsøgsvis skal overtale til at indtræde i en VK-regering. [Jyllandsposten 10/8 1980].

Anders har imidlertid altid afvist at gå ind i politik. I et stort interview i anledning af sin 50-års fødselsdag med overskriften »Jeg vil have lov at sige, hvad jeg har lyst til« i Jyllandsposten specificerede Ølgaard sit forhold til politik:

Jeg er meget interesseret i politik og mener, det er vigtigt, at der sidder gode folk som partipolitikere. Men udover at de har et forbandet slæb, bliver de jo nødt til at rette sig ind efter hinanden, bliver bundet, fordi der af loyalitetsgrunde er grænser for, hvad de kan sige. Samtidig er jeg tilhænger af, at der bruges udenomsparlamentariske metoder, forstået på den måde, at det efter min mening er af stor betydning at der er nogle mennesker, der kan kaste synspunkter ud til debat, uden at deres uafhængighed af enhver interesse kan betvivles. [Jyllandsposten, 5/9 1976].

Denne rolle, som en der kaster synspunkter ud til debat, uden at uafhængigheden kan betvivles, beholdt Anders til sin død. Og han holdt sig ikke fra kontroversielle emner; han skrev kronikker om indvandrerpolitik og imod skattestoppet efter sin 75- års fødselsdag, f.eks. kronikken »Velfærd trods skattedrop?« i *Berlingske Tidende* d. 20. januar 2006.

Da Anders Fogh Rasmussen talte til Nationaløkonomisk Forenings store Koldingfjord-møde som nykåret statsminister i 2002, var der ikke mange spørgsmål; stort set kun ét, og det var fra Anders, der udtrykte sin betænkelighed ved kontraktpolitikken. Ølgaard frygtede, at det ville blive besværligt at få gennemført nødvendige reformer, hvis man var bundet af de ofte letsindige løfter, der blev givet i forbindelse med en valgkamp. Man kan ikke sige andet end, at udviklingen siden har vist, at Ølgaard havde en pointe.

Ølgaard store styrke var, at han igennem mere end en menneskealder – hans første kronik kom i 1959 og hans sidste i 2006 – deltog i den offentlige debat om de mest brandfarlige emner, og samtidig kandiderede til politisk stærkt kontroversielle poster uden at miste sit ry som en uafhængig og saglig ekspert. Der er adskillige økonomer, der har fået en rolle som debattører i samfundsdebatten, og også mange der har haft en fast position som anerkendte økonomiske eksperter. Men det er meget få, der har kunnet kombinere de to roller.

Når det lykkedes for Anders, skyldes det nok hans integritet og høje moral, der kunne nærme sig puritanisme. Han var den, der påtalte vismænds og økonomers fejltrin uden barmhjertelighed. Vismændenes »uskrevne regler« om, at de altid skal optræde i enighed, og at de aldrig må være i nærheden af inhabilitet og f.eks. ikke have bestyrelsesposter, har i høj grad deres rod i debatter, hvor Ølgaard havde en central rolle. Han var ikke kun sine venners ven; han var også deres revser. Det mest oplagte eksempel herpå – men langt fra det eneste – er det i erindringerne deltaljeret skildrede opgør med venen Mogens Glistrup, da denne efter Ølgaards opfattelse ikke følte sig forpligtet til at bruge sine rige evner til samfundets bedste, men tværtimod brugte dem til omgåelse af lovgivningens intentioner.

En anden ting, der medvirkede til, at Ølgaards position som saglig ekspert blev opretholdt på trods af hans livlige deltagelse i den offentlige debat, var også hans permanente seriøsitet. Selv om han villigt deltog i underholdningen ved diverse økonomarrangementer – sidste gang var, da han optrådte som »syngende vismand« til Nationaløkonomisk Forenings Koldingfjordmøde i 2008 med et af sine gamle glansnumre visen »Josephine« – holdt han sig langt fra letbenede virkemidler i det offentlige rum. Han optrådte ikke i billedbladene og viste ikke familie og privatliv frem. I et interview med de økonomistuderendes blad P i 1989 siger han, hvad han er villig til at sige om sit privatliv:

Jeg er gift og har tre børn. Og så bor jeg i Virum, hvor – hvad enhver politstuderende, der har fulgt undervisningen på 1. år i samfundsbeskrivelse vil vide – vi i 1955 købte et rækkehus for 66.000 kr. Det står i dag i ca. 800.000. I min fritid interesserer jeg mig for og løber orienteringsløb. Det har jeg gjort siden 1950. [P, marts 1989].

I samme interview angiver Ølgaard sit motto som »*Man har sine ører for at få dem i maskinen!*«. Det motto levede han efter fra sin tidligste ungdom, til han var over 80. Men han gjorde det på en sådan måde, at han aldrig blev betragtet som provokatør eller polemiker, men tværtimod som en uafhængig autoritet i den offentlige debat.

Kilder:

Andersen, L. J. 1976. Jeg vil have lov at sige, hvad jeg har lyst til, *Jyllandsposten*, den 5. september, s. 10.

Bliss, C. 1968. Anmeldelse af Growth, Productivity and Relative Prices by Anders Ølgaard, *Economic Journal*, vol. 78, s. 416-17.

Brems, H. 1972. Anmeldelse af Growth, Productivity and Relative Prices by Anders Ølgaard, *Econometrica*, vol. 40.1, s. 216-17.

Erhvervslivet. 1983. Ølgaard synes at føre med »mulehår« i FDB-opløbet, *Vendsyssel Tidende*, den 5. marts, s. 5.

Larsen, B. og K. Kristensen. 1989. Anders Ølgaard, *P – et blad for politter*, 22. årgang nr. 2, s. 6-7.

Mikkelsen, A. 1987. Plusser og minusser gennem 25 år: Elementer til en vurdering, i Det Økonomiske Råds Formandskab: *Dansk Økonomi – Råd og realiteter 1962-1987*, Det Økonomiske råd, København.

Mogensen, G. Viby. 1987. *Historie og økonomi*, Akademisk Forlag, København.

Paldam, M. og J. Hansen. 1973. *SMEC, en kvartalsmodel af den danske økonomi*, København.

Persson, K. G., red. 1993. *The Economic development of Denmark and Norway since 1870*, Edward Elgar Publishing, Aldershot.

Rosted, J., A. Schaumann og C. Sørensen. 1974. *SMEC II – måling af Finanspolitikens aktivitetsevirkninger*, København.

Wagner, J. 1980. Her er skyggerne fra højre, *Jyllandsposten*, den 10. august, s. 15.

Niels Kærgård
Fødevareøkonomisk Institut
Københavns Universitet

A. KRONOLOGISK PUBLIKATIONSLISTE

(ekskl. kronikker og andre avisartikler samt tilsvarende bidrag til tidsskrifter m.v.)

1956

1. Anmeldelse af Arne Rasmussen, Pristeori eller parameterteori. *Erhvervsøkonomisk Tidsskrift* nr. 1, 1956, 20. årgang s. 48-51.
2. Anmeldelse af P. Nyboe Andersen, Udenrigsøkonomi. *Det danske Marked*, hæfte, 1956, 15. årgang s. 123-26.
3. »Hvorledes gennemfører de statsvidenskabelige studerende deres studium?« København 1956, stencileret. 63 s.

1957

4. »Hvordan studerer stud.polit.'er?« *Nationaløkonomisk Tidsskrift*, 1.-2. hæfte 1957, 95. bind, s. 73-88.
5. Anmeldelse af Harald Dickson, Nationalekonomisk teori. *Nationaløkonomisk Tidsskrift*, 1.-2. hæfte 1957, 95. bind s. 121-22.
6. Anmeldelse af Olof Lindahl, Sveriges Nationalprodukt 1861-1951. *Nationaløkonomisk Tidsskrift*, 1.-2. hæfte 1957, 95. bind s. 122-23.
7. Anmeldelse af Poul Winding, Some Aspects of the Acceleration Principle. *Finanstidende* nr. 38, 1957, s. 1097-1099.

1958

8. »Problemstillinger i vækst-teorien«, *Nationaløkonomisk Tidsskrift*, 1.-2. hæfte 1958, 96. bind s. 51-75.
9. Anmeldelse af Bjarke Fog, Priskalkulation og Prispolitik. *Finanstidende* nr. 25, 1958, s. 798-99.
10. »Afgrænsningsproblemer ved anvendelse af prisindex«. *Det danske Marked*. 2. hæfte 1958, 17. årgang, s. 97-106.
11. »Efterspørgselen efter produktionsfaktorer«, *Memorandum fra Københavns Universitets Økonomiske Institut* nr. 1, København 1958, 45 s. 2. udgave 1962, 58 s. Genoptrykt af Akademisk Forlag 1974.
12. »Det danske bytteforhold i udenrigshandelen 1875-1955«, *Festskrift til Frederik Zeuthen*, København 1958, s. 383-400.
13. »Dynamiken« i »Omkring Zeuthens Economic Theory and Method«, oversigtsartikel i *Nationaløkonomisk Tidsskrift*, 5.-6. hæfte 1958, 96. bind s. 213-16.

1959

14. Indlæg i tilknytning til Lars Wahlbecks foredrag »Arbetskraftutveckling, kapital-

behov och ekonomisk tillväxt», *Forhandlingerne ved det nordiske national-økonomiske møde i København 1958*, København 1959, s. 94-106.

15. »Long-Term Trends in Prices and in Internal and External Terms of Trade in Denmark from 1875 to 1955«, International Association for Research in Income and Wealth's 6. europæiske konference i Portoroz, Jugoslavien. Stencileret. 65+12 s.
16. *Danmarks økonomi og udlandet*, København 1959, 96 s.
17. »The Cobb-Douglas Function and the Economic Growth of Norway 1900-1955«, *Productivity Measurement Review* nr. 18, August 1959, s. 49-53.
18. Tre kapitler i Arne Jensen og Anders Ølgaard, red. Indeksproblemer, *Memorandum fra Københavns Universitets Økonomiske Institut* nr. 4, København 1959, s. 24-68, 69-96 og 127-32.

1960

19. Sammen med Ole Bus Henriksen, Danmarks Udenrigshandel 1874-1958, *Studier fra Københavns Universitets Økonomiske Institut* nr. 2, København 1960, 133 s.
20. »Det danske nationalprodukt siden 1870«, oversigtsartikel vedr. Kjeld Bjerke og Niels Ussing, Studier over Danmarks Nationalprodukt 1870-1950, *Nationaløkonomisk Tidsskrift*, 5.-6. hæfte 1960, 98. bind s. 254-59.

1961

21. »En klassiker i konfirmationsalderen«, oversigtsartikel vedr. P. Nyboe Andersen, Bjarke Fog og Poul Winding, Nationaløkonomi, 2. udgave, *Nationaløkonomisk Tidsskrift*, 3.-4. hæfte 1961, 99. bind s. 173-85.

1964

22. »Studieaktiviteten ved Københavns Universitet d. 2.-7. marts 1964«, København 1964. Rotaprint, 27 sider. Genoptrykt, *Blåt memo fra Københavns Universitets Økonomiske Institut* nr. 12, Københavns Universitets Økonomiske Institut 1973, 23 s.

1965

23. »Om beholdningsefterspørgsel«, *Nationaløkonomisk Tidsskrift*, 1.-2. hæfte 1965, 103. bind s. 31-48.

1966

24. »Growth, Productivity and Relative Prices«, *Studier fra Københavns Universitets Økonomiske Institut* nr. 10, Amsterdam og København 1966, 317 s.

1967

25. Indledning til Stabiliseringspolitik under fuld beskæftigelse, konference afholdt af Danske Økonomers Forening, København 1967, s. 9-11.

1968

26. »Økonomisk-politiske relationer« i Anders Ølgaard, red., *Færingers-Frænder*, København 1968, s. 145-81.
27. »What is the Best Way of Diversifying Malaysian Exports and Increasing Malaysian Export Earnings?«, *Journal of the Economics Society, University of Malaya*, vol. 9. december 1968, s. S9-S16. Genoptrykt i Malaysia, Ministry of Commerce and Industry, Journal nr. 4 1969, bind 2 s. 8-14.

1969

28. Sammen med P. Nørregaard Rasmussen, »The Role of Projections in Planning and Projection Techniques«, Sectoral Aspects of Projections for the World Economy, *First Interregional Seminar on Long-Term Economic Projections*, United Nations, New York 1969, Vol. II s.1-16.
29. »Modeller vedrørende økonomisk vækst«. Niels Thygesen og P. Nørregaard Rasmussen, red., Udviklingslinjer i makroøkonomisk teori, *Studier fra Københavns Universitets Økonomiske Institut* nr. 13, København 1969, s. 144-78.
30. »Ændringer i det danske produktions- og prismønster siden 1950«, i *Aktuelle økonomiske problemer. Festskrift til Carl Iversen*, København 1969, s. 233-48.
31. »Monetary Policy in Malaysia as Part of Policies towards Economic Development«, *UMBC Economic Review, Kuala Lumpur*, vol. V, nr. 2, 1969, s. 15-29. Genoptrykt i David Lim, red., Readings on Malaysian Economic Development, Oxford University Press, Kuala Lumpur 1975, s. 362-374.
32. »Aktuelle boligproblemer«, *Nationaløkonomisk Tidsskrift*, 5.-6. hæfte 1969, 107. bind, s. 215-25.

1970

33. Anmeldelse af Kjeld Bjerke, Indkomst- og lønfordelinger. *Weltwirtschaftliches Archiv*, 1970 I, 104. bind, s. 60x-63x.
34. »Demografi og kapitalteori«, *Nationaløkonomisk Tidsskrift*, 3.-4. hæfte 1970, 108. bind s. 212-14.

1971

35. Anmeldelse af Niels Thygesen, The Sources and the Impact of Monetary Changes. *Finanstidende* nr. 37 1971, s. 1122-23.

36. »Om Lorenzkurver m.m.« i *Økonomien bag socialpolitikken* (Festskrift til Jørgen S. Dich), København 1971, s. 160-75. Optrykt i *Socialt Tidsskrift*, nr. 6-7 1971, 47. årgang, s. 290-305.
37. »En disputats om dansk pengepolitik 1951-68«, *Nationaløkonomisk Tidsskrift*, 1.-2. hæfte 1971, 109. bind, s. 76-96.

1972

38. »Om kalendereffekt og årgangseffekt i en guldalder«, s. 3.1-3.24, og »Kommentar til Buch-Hansen«, s. 5.1-5.7, begge i Jørgen Birk Mortensen, red., *Aspekter af økonomisk vækstteori, Memorandum fra Københavns Universitets Økonomiske Institut* nr. 28, København 1972.

1973

39. »Føres der pengepolitik i Danmark?«, *Nationaløkonomisk Tidsskrift* nr. 1 1973, 111. bind s. 9-24.

1974

40. »Finanspolitikens aktivitetsvirkninger – resultater fra SMEC II«, *Nationaløkonomisk Tidsskrift* nr. 2 1974, 112. bind s. 249-66.
41. »The Problems of the Balance of Payments of the United States and Proposals for Improvement of the Situation« i Nasrollah S. Fatemi, red., *Problems of Balance of Payment and Trade*, New Jersey 1974, s. 70-78.

1975

42. »Finansiering under inflation – indeksslån og andre løsningsmuligheder«, *Nationaløkonomisk Tidsskrift* nr. 1 1975, 113. bind s. 55-69.
43. »Hvilken utbygging har vi råd til?« i Jan F. Reymert, red., *Referat Nordisk Byggedag XII*, Oslo 1975, s. 65-70.
44. »Nielsen contra Glistrup« *Weekendavisen Berlingske Aften* 1975, d. 13. juni, d. 20. juni og d. 11. juli.

1976

45. »The Role of External Advisers – Some Danish Experience«, *UMBC Economic Review*, Kuala Lumpur, vol. XII, no. 1, 1976, s. 12-17.
46. »Arbejdslønnen i byerne og i landbruget i Danmark 1875-1955«. *Blåt memo fra Københavns Universitets Økonomiske Institut* nr. 49, København 1976, 54 s.

1977

47. »Om politiserende økonomer – med særligt henblik på den danske »vismands« institution« i Festskrift til H. Winding Pedersen, *Nationaløkonomisk Tidsskrift* nr. 1, 1977, 115. bind s. 122-56.

48. Seks kapitler i Erik Gørtz og Jørgen Drud Hansen, red., *Indeksteori*, Odense 1977, s. 10-20, 37-43, 102-5 og 129-55.
49. Redaktør af *Bolig- og byggeøkonomi i de nordiske lande*, Nordisk Byggedag XIII og Samfundsvidenskabeligt forlag, København 1977, 204 s.
50. »Om Grønland, grønlændere og danskere«, *Tidsskriftet Grønland* nr. 7-8 1977, 25. årgang s. 197-244.

1978

51. »*Boligmasse og byggeri, Økonomi og miljø*«. Sammenfatning af forhandlingerne på Nordiske Byggedage XIII, Referatbog, København 1978, s. 81-112.
52. »Ansvar skal følge kompetence«. *Grønlands erhvervsliv*, Rolighedskonference, København 1978, s. 45-49.
53. »Om hjemmestyrets muligheder for at styre den del af den grønlandske produktionssektor, som foreslås overtaget fra KGH« (sammen med Knud Overø), bilag 22 til *Betænkning afgivet af Kommissionen om Hjemmestyre i Grønland, Betænkning nr. 837*, København 1978. Bind 2, s. 85-87.
54. »*Kalâtdlit – nunanik, kalâtdlinik danskinigdlo*« (grønlandsk oversættelse af nr. 50). Institut for eskimologi, København 1978, 62 s.

1979

55. »*Den fremtidige erhvervsudvikling i Grønland set i lyset af råstofproblemerne*«. Konferencen vedrørende Mineralske Råstoffer. Grønland, Julianehåb, Juni 1978, København 1979, 15 s.
56. »*Samfundet, boligen – og skænderierne mellem beboergrupperne. Hvordan bor vi i 1990*«? Boligfonden Bikuben, København 1979, s. 3-11.
57. Concepts and problems of consistency in terms-of-trade analysis. *Blåt memo fra Københavns Universitets Økonomiske Institut* nr. 79, København 1979, 40 s.
58. »Beslutningerne i den økonomiske politik«, *Nationaløkonomisk Tidsskrift* nr. 3, 1979, 117. bind s. 385-94.

1980

59. »The patteredn of Danish prices during the 1970s and Danish economic policy.« *Blåt memo fra Københavns Universitets Økonomiske Institut* nr. 87, København 1980, 27 s.
60. »*Økonomi og boliger*« og »*Nye vilkår for fremtidens ejere og lejere*«, Byggeriet i en urolig tid, Rolighedskonference, København 1980, s. 8-27 og 76-82.
61. »The Danish Economy. Collection studies«. *Economic and Financial Series* no. 14. Commission of the European Communities, 1980, 255 s.

62. »Åbenrå-dreng i 1930erne«, *Sprogforeningens Almanak* 1981, Sønderborg 1980, s. 57-65.

1981

63. »Sundhedssektoren i samfundet«, *Sundhedspolitik?* Amtskommunernes og kommunernes Forskningsinstitut, København 1981, s. 8-26.
64. »Concepts and Problems of Consistency in Terms of Trade Analysis«, *The Review of Income and Wealth* no. 2, June 1981, series 27, s. 165-193.
65. »Indekslån – et supplement til traditionelle fordringer?«, *Nationaløkonomisk Tidsskrift* nr. 3, 1981, 119. bind s. 309-329.
66. »Arstruktúra és gazdaságpolitika a hetvenes éveken Dániában«, *Gazdaság, Budapest*, nr. 3, 1981, XV bind pp 137-152. (Udarbejdet på grundlag af nr. 59).

1982

67. »Indekslån«, Boligfinansiering og skat, *Kreditforeningen Danmarks debatserie*, København 1982, s. 35-47.
68. »Changes of Relative Prices as Based on National-Accounts Figures«, *Essays in Honour of Jørgen H. Gelting*, Copenhagen 1982, s. 147-56.

1983

69. »Hvad påvirker omfanget af boligbyggeriet?« *Blåt memo fra Københavns Universitets Økonomiske Institut* nr. 114, København 1983, 25 s.

1984

70. »Realprodukt og realindkomst fordelt på erhverv 1974-76«. *Blåt memo fra Københavns Universitets Økonomiske Institut* nr. 121, København 1984, VII + 15 s.
71. »Ejendomspriser og boligbyggeri – en nostalgisk jubilæumsnote om stockflow«. *Økonomiske Essays, Studier fra Københavns Universitets Økonomiske Institut* nr. 28, København 1984, s. 403-416.
72. »Hvad jeg skrev, det skrev Else«. (Bibliografi). *Blåt memo fra Københavns Universitets Økonomiske Institut* nr. 122, København 1984, 21 s.
73. »Comments on New Technology and Relative Prices« i Csikós-Nagy, Hague and Hall, red., *The Economics of Relative Prices*, Proceedings of an IEA Conference in Athens, Greece, in September 1981. Macmillan 1984, s. 140-50.
74. »Det 25. Nordiske Nationaløkonomiske Møde«, *Nationaløkonomisk Tidsskrift* nr. 3, 1984, 122. bind s. 305-06.

1986

75. »Uligevægte på bolig- og byggemarkederne« i *Økonomi og Landbrug (Festskrift til Holger Gad)*. Aarhus 1986, s. 327-42.

76. »Finansierings- og fordelingsproblemer i lys af balanceproblemerne«, *Nationaløkonomisk Tidsskrift* nr. 2, 1986, 124. bind s. 179-88.
77. Stabilization Policies and Planning Problems of the Danish Economy. *Blåt memo fra Københavns Universitets Økonomiske Institut* nr. 150, København 1986, 30 s.

1987

78. »Dansk landbrugs fremtid«, *Nationaløkonomisk Tidsskrift* nr. 2, 1987, 125. bind s. 171-84.
79. »A dán gazdaság stabilizációs politikája«, *Gazdaság, Budapest* nr. 3, 1987, XXI. bind s. 130-140. (Udarbejdet på grundlag af nr. 77).
80. »Perioden 1972-75: Selvglæde og gode råd« i Råd og realiteter. *Dansk økonomi 1962-1987*. Udgivet i anledning af 25 året for oprettelsen af Det økonomiske Råd. København 1987, s. 223-228.

1988

81. »Pensionsdebatten«, *Nationaløkonomisk Tidsskrift* nr. 2, 1988, 126. bind s. 119-123.
82. »Den forkalkede boligpolitik«, *Nationaløkonomisk Tidsskrift* nr. 3, 1988, 126. bind s. 267-285.

1989

83. Sammen med Ellen Andersen og Poul Christian Matthiessen, »Boligmasse og boligbyggeri i første halvdel af dette århundrede«, i Festskrift til Svend Aage Hansen, *Økonomi og Politik* nr. 2, 1989, 62. årgang s. 6-19. 1991.

1991

84. »ØMU, den økonomiske politik og institutionerne«, *Økonomi & Politik* nr. 2, 1991, 64. årgang s. 4-14.
85. »Hvad jeg skrev, det skrev Anne Lise og Else«. Københavns Universitet. 1991, *Blåt memo fra Københavns Universitets Økonomiske Institut* nr. 184.
86. »Mætning på boligmarkedet – og konsekvenserne«. *Aabenraa Andelsboligforening: 50 års jubilæum 12. september 1991: en bog der handler om fortid, nutid og fremtid for lejeboligerne*. Aabenraa Andelsboligforening. 1991. s. 42-44.

1992

87. »Europæisk samarbejde i flere hastigheder«? *Nationaløkonomisk Tidsskrift*, nr. 4, 1992, 130 bind s. 551-59

88. »Om udlands-, offentlig og privat gæld«, *Nationaløkonomisk Tidsskrift*, nr. 1, 1992, 130 bind s. 19-29

1993

89. (Med Ellen Andersen og P. C. Matthiessen) »Boligmætning og huslejespænd – et debatoplæg«: *Betænkning afgivet den 1. juli 1993 af udvalget til belysning af udviklingen på boligmarkedet i de kommende år*, Kbh.: C. A. Reitzel, 1993. 144 s.

1994

90. »Kommissionsdomstolen vedr. P&T«. I: Andersen O. S., red., *Beretning fra Advokatrådets konference om kommissionsdomstole og undersøgelsesretter*. Djøf/Jurist- og Økonomforbundet. 1994. s. 45-47.

1995

91. »Pengepolitik, valutakurspolitik, monetær politik og pengemængde-definitio-ner«, *Nationaløkonomisk Tidsskrift*, nr. 2, 1995, 133 bind 2 s. 189-195.

1998

92. »Simon Kuznets«. I: *Den Store Danske Encyklopædi*. Kbh.: Gyldendal. 1998. s. 411.
93. Monetary prospects for the European Union. *Social Sciences – Socialiniai*. 1998; 1(14):82-84.

1999

94. »Dansk valutakurspolitik siden Første Verdenskrig«. *Socialøkonomisk debat gennem 25 år*. Kbh.: Djøf/Jurist- og Økonomforbundet. 1999. s. 27-35.
95. »Prospects and Perspectives for the European Union«. *Social Sciences – Socialiniai*. 1999;(1 (18):90-92.

2001

96. (med Niels Kærgård) »H. Winding Pedersen 19. marts 1907 – 24. marts 1999«, *Nationaløkonomisk Tidsskrift*, nr. 1, 2001, bind 139 s. 1-16.

2002

97. »Vismændene og den offentlige debat«, *Samfundsøkonomen*. 2002;(8):8-14.

2003

98. »Economic Policies in the European Union«, *Inzinerine Ekonomika*. 2003;3 (34): 121-125.
99. »Et nyt boligforlig«. I: *Salg af almene boliger – vejen frem?: en antologi om dansk boligpolitik*. Kbh.: AKF-Forlaget. 2003. s. 133-138.

100. (med Niels Kærgård og Peter Birch Sørensen) »A. Hans Brems 16. oktober 1915 – 16. september 2000«, *Nationaløkonomisk Tidsskrift*, nr. 2, 2003, bind 141 s. 139-53.

2005

101. »Den strandede ITP-reform i 1967: en personlig reformhistorie«. I: *13 reformer af den danske velfærdsstat*. Odense: Syddansk Universitetsforlag, 2005. s. 147-156.

2008

102. *Den syngende vismand: biografi*. Kbh.: Nyt Juridisk Forlag, 2008. 445 s.

B. KRONIKKER M.V.

1959

1. »Landbrugets priser, valutakurserne og pristalsreguleringen«, *Jydske Tidende*, 7. og 29. april 1959.

1962

2. »Fællesmarkedet og vort grænseproblem«, *Information*, 1. februar 1962.
3. »Landbruget og Monopoltilsynet«, *Politiken*, 12. maj 1962.
4. »Økonomerne og Fællesmarkedet«, *Kristeligt Dagblad*, 26. juni 1962.
5. »Myter om valutakursen«, *Kristeligt Dagblad*, 10. september 1962.

1963

6. »Lån og legater«, *Politiken*, 23. november 1963.
7. »Udflytning, der haster«, *Politiken*, 2. oktober 1964.

1964

8. »Tidsmæssig forskydning af skatten – led i konjunkturstabilisering«, *Information*, 15. oktober 1965.

1966

9. »Huslejepolitikken og boligbyggeriet«, *Politiken*, 14. januar 1966.
10. »Færøerne, Danmark og Grønland«, *Kristeligt Dagblad*, 15. februar 1966.

1967

11. »Forskellen er større mellem rig og fattig«, *Aktuelt*, 30. maj 1967.
12. »Efter ATP kommer ITP«, *Politiken*, 4. september 1967.
13. »Ønsker vi at få ITP?«, *Politiken*, 5. september 1967.
14. »ATP-problemer«, *Kristeligt Dagblad*, 1. november 1967.
15. »Uden for Fællesmarkedet«, *Ingeniørens Ugeblad*, 17. november 1967.

1968

16. »Den nye renteaftale og konkurrencen mellem pengeinstitutterne«, *Information*, 26. februar 1968.
17. »Færøerne og Danmark«, *Politiken*, 17. oktober 1968.

1969

18. »Økonomiske rådgivere, økonomisk råd og offentligheden«, *Politiken*, 20. marts 1969.
19. »Racefordomme – samme sag i Malaysia og Grønland?«, *Aktuelt*, 5. november 1969.
20. »De fede 1960ere – de magre 1970ere?«, *Politiken*, 31. december 1969.

1970

21. »Grundejerne, lejerne og Folketinget«, *Kristeligt Dagblad*, 22. januar 1970.
22. »50 år efter Genforeningen«, *Berlingske Tidende*, 9. februar 1970.
23. »Om vatnisser«, *Politiken*, 10. april 1970.
24. »Skal Grønland have fjernsyn?«, *Politiken*, 29. april 1970.
25. »Basic Trends in the Danish Economy«, *Danish Journal* (udg. af Udenrigsmin.) maj 1970.
26. »Færøernes fremtid«, *Finanstidende*, 15. maj 1970.
27. »Lagtingsvalg og Færø-problemer«, *Berlingske Tidende*, 6. november 1970.
28. »Problemer om erhvervslivets investeringer«, *Erhvervsbladet*, 10. december 1970.

1971

29. »Skab ikke nye uddannelses-svælg i befolkningen«, *Aktuelt*, 1. januar 1971.
30. »Udsigt til arbejdsløshed«, *Berlingske Tidende*, 10. januar 1971.
31. »Færøerne skal nå meget i 70erne«, *Sparekassetidende*, februar 1971.
32. »Støtte til idrætten – på hvilke betingelser?«, *Politiken*, 14. februar 1971.
33. »En brosten mod boligpolitikken«, *Information*, 8. marts 1971.
34. »The Faroese economy«, *Danish Journal* (udg. af Udenrigsmin.), 1971.
35. »Om at partere mammutter«, *Politiken*, 2. november 1971.
36. »Vismændene« lægger EF-fakta på bordet, nu politikernes tur«, *Aktuelt*, 9. november 1971.

1972

37. »Virksomhedernes selvstændighed kan komme i fare«, *Dansk Industri*, januar 1972.
38. »Hvad er det varige tab ved at stå uden for EF?«, *Aktuelt*, 24. februar 1972.
39. »Tjenestemand-status«, *Berlingske Tidende*, 26. marts 1972.

40. »Fordelings-problemerne ændres: Nye skel i indkomstdannelsen«, *Aktuelt*, 30. marts 1972.
41. »Dybfrys mineralerne på Grønland«, *Politiken*, 4. maj 1972.
42. »Er det nye dagpengeforslag grydeklart?«, *Politiken*, 19. maj 1972.
43. »Hvad koster dagpengereformen?«, *Politiken*, 26. maj 1972.
44. »Forbruger – herre eller træl?«, *Tænk*, nr. 4, 1972.
45. »Forbrugersagens to roller«, *Tænk*, nr. 5, 1972.
46. »EF, »vismændene« og mig«, *Aktuelt* samt en række provinsblade, september 1972.
47. »Større privat og offentligt forbrug? Næppe i 1970erne«, *Aktuelt*, 15. december 1972.
48. »Hvor central bliver den centrale fond, når det kommer til stykket?«, *Aktuelt*, 19. december 1972.
49. »Den offentlige sektors (manglende) mekanisering«, *Kooperationen*, nr. 9/10, 1972.
50. »Hvad lærte vi af rektorvalget?«, *Universitetsavisen*, nr. 1, 1973.

1973

51. »Giv os en chance for at købe indeksobligationer«, *Information*, 20.-21. januar 1973.
52. »Hvad skal vi med indeksslån?«, *Weekendavisen*, 26. januar 1973.
53. »Biltogstunnel eller bro over Storebælt?«, *Politiken*, 29. maj 1973.
54. »Vismændene myrder et misfoster. Længe leve "finanseffekten"«, *Aktuelt*, 7. juli 1973.
55. »Wilhjelm, »vismændene« og finanseffekten«, *Information*, 12. juli 1973.
56. »Studentergårdens økonomi«, *Gårdbroderen ved Studentergårdens 50 års jubilæum* 1973.
57. »Tømmermænd og protest«, *Politiken*, 11. december 1973.
58. »Vatnisser med tømmermænd?«, *Politiken*, 12. december 1973.
59. »Vismands dystre spådom – beskæftigelsen trues, selv om vi får olie nok«, *Aktuelt*, 24. januar 1974.

1974

60. »Glistrup og skattelovene«, *Politiken*, 17. marts 1974.
61. »Disponibel realløn – og sådan noget«, *Berlingske Tidende*, 29. marts 1974.
62. »Om H-planens analyse af registrerede studieforbøjer«, *Universitetsavisen*, nr. 6, 1974.
63. »Disponibel realløn – med noget oveni«, *Berlingske Tidende*, 15. maj 1974.
64. »Skruestikken vi sidder i«, *Politiken*, 6. november 1974.

65. »Lønstop som led i en beskæftigelsesløsning«, *Politiken*, 7. november 1974.

1975

66. »Nielsen contra Glistrup«, *Weekendavisen*, 13. juni 1975.
67. »I lommen på sin advokat«, *Weekendavisen*, 20. juni 1975.
68. »Nielsen contra Glistrup – en afsluttende randbemærkning«, *Weekendavisen*, 11. juli 1975.
69. »Politikerne skal vælge, deres valg vil påvirke vore arbejdsløshedstal«, *Aktuelt*, 27. juni 1975.

1977

70. »Ejer-, lejer- og andelsboliger«, *Politiken*, 31. august 1977.
71. »Boligselskaberne og socialpolitikken«, *Politiken*, 16. september 1977.

1978

72. »Indekslån vil ligestille sparer og låntager«, *Børsen*, 16. januar 1978.
73. »Hvem kan sidde i Folketinget?«, *Berlingske Tidende*, 17. februar 1978.
(Genoptrykt som bilag til betænkning fra udvalget til prøvelse af valgene, Folketingstidende, februar 1984.)
74. »Kronen stiger – hent raflebægeret«, *Politiken*, 30. maj 1978.
75. »Tunnel til biltog den bedste løsning«, *Politiken*, 12. juni 1978.

1979

76. »Om almennyttige boliger og uvidende professorer«, *Boligen*, nr. 9, 1979.
77. »Om »blålys« og amerikansk olie«, *Berlingske Tidende*, 2. december 1979.

1980

78. »Beskatning af renter«, *Politiken*, 17. marts 1980.
79. »EF-tilhængere, gå i tænkeboks!«, *Politiken*, 3. maj 1980.
80. »Den offentlige sektors vækst«, *Berlingske Tidende*, 21. maj 1980.
81. »Konkurrenceevnen og samfundsøkonomien«, *Provinsbankens temablade*, nr. 13, 1980.

1981

82. »Olietater, forbrugerlande og olieselskaber«, *Synspunkt fra Esso*, nr. 3, 1981.
83. »Pensioner, rentegab og indeksslån«, *Berlingske Tidende*, 12. november 1981.

1982

84. »Good-turns«, *Pluk fra forskning i Sønderjylland*, nr. 1, 1982.

85. »Indekslån vil gøre huskøb overskueligt«, *Aalborg Søndags Stiftstidende*, 14. februar 1982.
86. »The Danish economy after 1945«, *Factsheet Denmark* (udg. af Udenrigsmin.), 1982, 2. reviderede udgave 1983.
87. »Plagerne i mandtal – og hvad der bør gøres«, *Weekendavisen*, 1. oktober 1982.
88. »Danmarks svar på Billy Graham«, *Politiken*, 14. oktober 1982.

1983

89. »Det handler om tillid til Brugsen«, *Aktuelt*, 25. februar 1983.

1984

90. »Man må ikke forveksle kompetence med habilitet«, *Børsen*, 5. december 1984.
91. »Meddelelser fra Nationaløkonomisk Forening« (med Thorkild Davidsen), Supplement til *Nationaløkonomisk Tidsskrift* 1984.

1985

92. »Realkredit og skattereformen«, *Jyllands-Posten*, 27. august 1985.
93. »Hvordan skal vi pensioneres?«, *Berlingske Tidende*, 16. oktober 1985.
94. »Meddelelser fra Nationaløkonomisk Forening« (med Thorkild Davidsen), Supplement til *Nationaløkonomisk Tidsskrift* 1985.

1986

95. »Om tvivl og om vetoret«, *Fyens Stiftstidende*, 30. januar 1986 samt *Aarhus Stiftstidende*, 1. februar 1986 og *Aalborg Stiftstidende*, 3. februar 1986.
96. »Den private opsparing har svigtet«, *Børsen*, 10. februar 1986.
97. »De offentligt ansattes løn«, *Børsen*, 28. august 1986.
98. »Meddelelser fra Nationaløkonomisk Forening« (med Thorkild Davidsen), Supplement til *Nationaløkonomisk Tidsskrift* 1986.

1987

99. »Dansk landbrug – fremtiden og problemerne«, *Weekendavisen*, 27. februar 1987.
100. »Deltidslandmænd« kan blive et betydeligt aktiv«, *Weekendavisen*, 6. marts 1987.
101. »Om cheføkonomer, prognoser og nye indgreb«, *Jyllands-Posten*, 10. marts 1987.
102. »Fortæl os hvad vi skal« (om ministerielt udpegede medlemmer af bankbestyrelser), *Berlingske Tidende*, 14. maj 1987.

103. »Meddelelser fra Nationaløkonomisk Forening« (med Thorkild Davidsen), Supplement til *Nationaløkonomisk Tidsskrift* 1987.

1988

104. »Skattetrykket er stærkt undervurderet«, *Børsen*, 12. februar 1988.
105. »Ansvaret for boligbyggeriet – og debatten«, *Weekendavisen*, 8. juli 1988.
106. »Meddelelser fra Nationaløkonomisk Forening« (med Thorkild Davidsen), Supplement til *Nationaløkonomisk Tidsskrift* 1988.

1989

107. »Der er lotterigevinster«, *Berlingske Tidende*, 1. marts 1989.
108. »Turorientering på norsk«, *O-posten*, Dansk Orienterings-forbund, 38. årgang 1989 nr. 2.
109. »Magten over kronerne (I)«, *Politiken*, 2. oktober 1989.
110. »Magten over kronerne (II)«, *Politiken*, 3. oktober 1989.
111. »Delors-rapporten og den økonomiske union«, *Weekendavisen*, 3. november 1989.

1990

112. »Det fastlåste boligmarked«, *Aktionærnyt, Den Danske Bank*, nr. 1. September 1990.
113. »ØMU-udemokratisk eller utroværdig?«, *Information*, 27. november 1990.

1991

114. »Priserne til landbruget skal ned«, *Politiken*, 11. februar 1991.
115. »ØMU-planerne og det politiske demokrati«, *Den grimme ælling*, 4. årgang nr. 1. Februar 1991.
116. »Lottoet om boligen«, *Politiken*, 30. maj 1991.
117. Blomgren, vismændene og Knud Kristensen. *Politiken*. 1991, 16. oktober, 2. sektion.
118. Om frie bask med halen. *Politiken*. 3. november, 2. sektion.

1992

119. Den europæiske vævning. *Børsen* 21. maj.
120. Efter unionsafstemningen: et nyt troværdighedsproblem. *Berlingske Tidende*. 11. juni.
121. Kampen om stemmerne. *Politiken*. 17. marts.
122. Maastricht som EF's tagetage. *Politiken*. 12. september.
123. Stabiliteten i Europa afgørende. *Politiken*. 3. april.

1993

124. Europas netværk. *Politiken*. 27. januar.
125. Unødig frygt for EF. *Politiken*. 10. april.

1994

126. Nok af boliger. *Politiken*. 19. december.
127. Stop eksperimentet med EU-mønt. *Politiken*. 18. juli.
128. ØMU med de store. *Politiken*. 14. august.

1995

129. 600.000 ledige? *Børsen*. 8. december.
130. Borgerløn ad bagdøren. *Børsen*. 2. februar.
131. EU-debat – men om hvad? *Berlingske Tidende*. 6. oktober.
132. Europæisk union i flere hastigheder. *Berlingske Tidende*. 23. januar.
133. Hvordan redder vi velfærdsstaten? *Børsen*. 3. februar.
134. Klar tendens trods tvivlsomme tal. *Børsen*. 22. februar.
135. ØMU: Utopi eller vild(m)and. *Morgenavisen Jyllands-Posten*. 11. marts.

1996

136. Arbejdsmarkedet i en stagnationsøkonomi. *Børsen*. 5. juli.
137. Danmark bliver en stagnationsøkonomi. *Børsen*. 3. juni.
138. ÆldreForum uden cola – men med pop? *Berlingske Tidende*. 6. juli.

1997

139. ØMU på katastrofekurs. *Børsen*. 5. juni.

1998

140. Glæden ved de fremmede. *Børsen*. 5. marts.
141. Noget-for-noget princippet breder sig. *Børsen*. 16. januar.
142. Organdonation og dødsproces. *Berlingske Tidende*. 15. juli.
143. Åbne døre kan ende som den rene inkvisitation. *Politiken*. 10. september.

1999

144. EU og mig. *Berlingske Tidende*. 26. februar.
145. Konverterings-fænomenet. *Fyens Stiftstidende*. 3. februar.

2000

146. EU: tillid eller berøringsangst? *Berlingske Tidende*. 8. august.
147. En helt urimelig skat på statistiske nøgletal. *Politiken*. 6. maj.

148. Hvor er ØMU-entusiasmen? *Børsen*. 2000. 27. marts.
149. Vismænd, euro og røverhistorier. *Berlingske Tidende*. 8. december.
150. ØMU-tillid og selvtillid. *Politiken*. 2000. 24. marts.

2001

151. Boligmarkedets vilkår. *Berlingske Tidende*. 29. januar.
152. Den danske indvandrerpolitik. *Politiken*. 13. november.
153. Fra bund til top. *Berlingske Tidende*. 24. oktober.
154. Færinger – frænder I. *Politiken*. 2001. 7. maj.
155. Færinger – frænder II. *Politiken*. 8. maj.
156. Indvandring og statistik. *Politiken*. 7. august.

2002

157. Forenkling frem for millimeterretfærdighed. *Berlingske Tidende*. 27. januar.
158. Antisemitisme og ytringsfrihed. *Politiken*. 3. maj.
159. Ytringsfrihedens grænser. *Politiken*. 26. maj.
160. Kun en åben debat giver pote. *Weekendavisen*. 7. juni.
161. Ølgaard A. Vitaminer til debatten. *Berlingske Tidende*. 27. juni.
162. Et nyt boligforlig. *Berlingske Tidende*. 29. juli.
163. Sæt pris på lejeboliger. *Berlingske Tidende*. 19. oktober.

2003

164. Femern-bro er gamle mænds ide. *Weekendavisen*. 6. juni.
165. Overlad boligpolitikken til kommunerne. *Morgenavisen Jyllands-Posten*. 2. december.
166. Sveriges umulige EU-debat. *Berlingske Tidende*. 7. september.

2004

167. Femernbroen: et levn fra Jerntæppetiden. *Politiken*. 14. juli.

2005

168. Sælg ghetto-boligerne. *Berlingske Tidende*. 26. april.

2006

168. Trafikprioriteter: hovedreparer banenettet – og glem Femernbroen. *Politiken*. 2. august.
170. Velfærd trods skattedrop? *Berlingske Tidende*. 20. januar.

Økonomisk afkast af ph.d. uddannelse: Livsindkomst og humankapital

Per Vejrup-Hansen

Copenhagen Business School, E-mail: pvh.ino@cbs.dk

SUMMARY: Only a few studies on the economic returns to Ph.D. education have been undertaken. Based on a new kind of data, lifetime earnings and human capital by Ph.D.'s are estimated separately by identical regression models. Both measures are relative to holders of a M.Sc./M.A. degree. Based on crosssection data, the overall finding is that for most subject fields and sectors (academia, government, and business) lifetime earnings by Ph.D. graduates taken from the time of their master graduation are lower than by master graduates. As an indicator of relative human capital by Ph.D. education, accumulated salaries by Ph.D. holders are 8 percent higher than by masters.

1. Indledning

En person med en lang videregående uddannelse, altså en kandidatgrad, har i princippet et valg mellem at gå ud på arbejdsmarkedet som kandidat, eller at fortsætte med et ph.d. studium og opnå en ph.d. grad efter 3-4 år. Hvis valget af en ph.d. uddannelse vurderes rent økonomisk, dvs. som monetært afkast, er forskellen i livsindkomst mellem ph.d.'er og kandidater et relevant grundlag. Ph.d.'ers relative livsindkomst er indkomstforskellen regnet fra kandidattidspunktet for begge grupper, og er et udtryk for det private afkast af en ph.d. uddannelse.

Hvis man derimod sammenligner ph.d.'ers løn i job efter ph.d. dimission med kandidaters løn fra kandidatdimission, altså mellem nyuddannede, får man (som udgangspunkt) et mål for afkastet af den humankapital, der erhverves via en ph.d. uddannelse. Her sammenlignes således årlige lønninger over erfaring efter afsluttet uddannelse, modsat over kandidatålder. Det kan tænkes, at den enkelte kandidats valg af ph.d. uddannelse også er motiveret af en løngevinst efter endt uddannelse. Men uanset dette er forskellen i humankapital en komponent i relativ livsindkomst.

En fuldgældig beregning af livsindkomst sker via den aritmetiske metode, hvor den konkrete omkostning ved uddannelsen opgøres og sammenholdes med indkomststrømmen. Estimationen af humankapitalafkastet foretages typisk via regressionsana-

lyse baseret på en Mincer-lønfunktion. Under nogle restriktive forudsætninger, f.eks. at der ikke regnes med en forskel i længden af arbejdslivet for personer med forskellige uddannelseslængder, Checchi (2006, s. 195ff), kan resultatet af regressionen også fortolkes som forskelle i livsindkomst.

I nærværende undersøgelse vil livsindkomst også blive estimeret via regressionsanalyse. Det sker i kraft af, at data omfatter personer under uddannelse. I vores sammenhæng vil det sige, at ph.d. studerende indgår sammen med de personer, der har afsluttet ph.d. uddannelsen. Derved fås hele indkomstforløbet fra kandidatdimission parallelt med kandidater, hvilket svarer til en aritmetisk opgørelse. Endvidere muliggøres, at estimationen af livsindkomst og humankapital foretages via et ensartet datagrundlag og med samme regressionsmodeller.

I relation til ph.d.'ers livsindkomst er det interessante spørgsmål, om der er overensstemmelse med det generelle mønster, at livsindkomst er jo større des længere uddannelse. Det skal ses i lyset af, at kun ganske få undersøgelser har inddraget ph.d. uddannelsen. I dansk sammenhæng har man således i generelle undersøgelser kun betragtet uddannelsesniveauer til og med kandidatniveauet.

Det økonomiske Råd har med års mellemrum beregnet livsindkomster for uddannelseskategorier, senest i Det økonomiske Råd (2003). Her kumuleres lønninger eller indkomster over arbejdslivet fra det typiske tidspunkt for afslutningen af de forskellige uddannelser, og der laves et skøn for indkomster (SU, elevløn o.l.) under uddannelse. Der er tale om en aritmetisk beregning, hvor der regnes med forskelle i længden af arbejdslivet, idet pensionsalderen for alle grupper sættes til 70 år. Resultatet er, at livsindkomsten vokser med højere uddannelsesniveau. Oversigter over resultater af tilsvarende generelle, udenlandske analyser findes i Hartog m.fl. (2007), Checchi (2006) og Johnes (1993).

En eksplicit undersøgelse af ph.d.'ers livsindkomst er Siegfried (1971). Det drejer sig om ph.d.'er i økonomi, der sammenlignes med M.A. graden i økonomi. Her beregnes interne renter (rates of return) ved at sammenholde skønnede omkostninger ved ph.d. uddannelsen (mistet indkomst og direkte studieudgifter) og afkastet i form af højere løn i job som ph.d. Der foretages en opdeling efter ansættelsessector, dvs. education, government og business. Det interessante er, at når man anvender det simple gennemsnit af M.A. uddannedes løn i alle erhverv under ét som mål for mistet indkomst under ph.d. uddannelsen, bliver den interne rente meget lav i education og negativ i government. Det svarer til, at man i en regressionsanalyse udelader variabelen for sector, dvs. at sector er endogen. Siegfried diskuterer tilstedeværelse af ikke-monetære afkast ved arbejde inden for uddannelse.

I Stock og Siegfried (2001) undersøges igen ph.d.'er i økonomi, men her ses på lønforskelle internt i gruppen af ph.d.'er, nemlig job og startløn set i relation til ph.d. afhandlingens emne og ansættelsessector.

Hvis vi herefter ser på undersøgelser af forskelle i afkastet af humankapital, foretages en analyse på danske løndata i Christensen og Westergaard-Nielsen (2001). Via en Mincer-lønfunktion estimeres det gennemsnitlige lønniveau som funktion af uddannelseslængde til og med antal år svarende til kandidatuddannelse. En forøgelse af uddannelse med ét år resulterer i en højere humankapital på omkring 5 procent.

Eksempler på undersøgelser af forskelle i humankapital hvori ph.d. niveauet indgår, er Taubman og Wales (1973), Ashenfelter og Mooney (1968) og Wilson (1987). Taubman og Wales (1973) finder, at lønniveauet for Ph.d.'er generelt set er lavt, når der sammenlignes med kandidater (M.A.). Når der imidlertid kontrolleres for ansættelse inden for college uddannelser (teaching), er ph.d.'ers lønniveau relativt højere. Det fører til antagelser om ikke-monetære afkast ved arbejde inden for uddannelse. Ashenfelter og Mooney (1968) sammenligner bachelorer og ph.d.'er inden for uddannelsesretninger (humaniora, samfundsvidenskab og naturvidenskab). Wilson (1987) sammenligner gennemsnitslønnen for forskellige niveauer af ingeniøruddannelser i England, hvor resultatet er, at ph.d.'ers gennemsnitlige lønniveau er 8-10 procent højere end M.Sc.'ers.

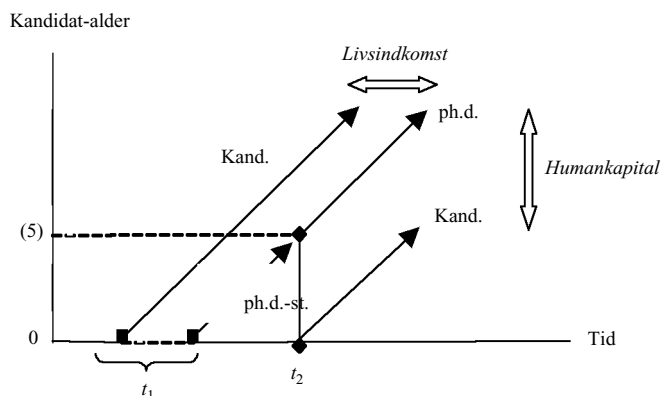
De foreliggende undersøgelser peger på, at en opdeling på sektorer og fagområder er central for analysen af relativ livsindkomst for ph.d.'er.

Artiklen er opbygget på følgende måde. I afsnit 2 illustreres og uddybes de to begreber livsindkomst og humankapital, herunder i relation til data. I afsnit 3 redegøres for regressionsmodellen til analyse af begge begreber samt for variable og kendetegn ved analysepopulationen. Analyseresultater præsenteres i afsnit 4, og afsnit 5 indeholder nogle afsluttende bemærkninger.

2. Begreber og data

De to begreber livsindkomst og humankapital illustreres via tidslinier i figur 1. Livsindkomsten for ph.d.'er versus kandidater tager udgangspunkt i samme år for kandidatdimission (t_1), og slutter ved samme kandidatalder for tilbagetrækning fra arbejdsstyrken. På tidslinjen for ph.d. indgår år for ph.d. studium (»ph.d.-st.«), for hvilket der fiktivt er angivet en kandidatalder på 5 år ved ph.d. dimission, som svarer til medianværdien af kandidatalderen (jf. tabel 1). Af illustrative grunde er de to tidslinjer vist adskilt, selv om de reelt skulle ligge oveni hinanden fra tidspunkt t_1 .

Det relative afkast af humankapital er forholdet mellem ph.d.'ers og kandidaters løn pr. hvert års erfaring efter afsluttet uddannelse, dvs. med start ved henholdsvis nyuddannede kandidater og nyuddannede ph.d.'er ved tidspunkt t_2 . På dette tidspunkt har ph.d.'er som nævnt en kandidatalder på omkring fem år, hvilket også må antages at være forskellen i biologisk alder. Men da humankapitalen er den gennemsnitlige lønforskel over erfaring, forudsættes implicit, at længden af arbejdslivet efter dimission er den samme.



Figur 1. Begreberne livsindkomst og humankapital: Illustration via tidslinjer.

Datagrundlaget er IDA databasen, hvorfra er udtrukket den totale bestand af personer med afsluttet kandidat- eller ph.d. uddannelse i hvert af årene 1980-2006. Via IDA kan timeløn anvendes som mål for indkomst. Som supplement til standarddata i IDA er inddraget karakter ved studentereksamen o.l. samt reference til forældre.

Umiddelbart viser figur 1 forløbet for to kohorter med henholdsvis samme kandidatår (t_1) og samme dimissionsår (t_2). I analysen vil vi imidlertid anvende tværsnitsdata for år 2006. Personer med en given kandidat-alder/erfaring i 2006 tilhører en given kohorte, men det er kun lønnen mv. ved dette »slut-punkt« i 2006, der registreres. Metodemæssigt er analyser af kohortedata at foretrække, men det forudsætter omfattende data.¹

Med hensyn til *livsindkomst* (livsløn) er det afgørende kendetegn ved data, at man via oplysninger om uddannelse kan identificere personer, der er under uddannelse – her kandidater, som er under ph.d. uddannelse. Derved kan der dannes en bruttogrube af ph.d.'er, der også omfatter ph.d. studerende. Når grundlaget er tværsnitsdata fås en beskrivelse af et gennemsnitligt livsforløb med hensyn til løn mv. Der er tale om faktiske, gennemsnitlige forløb med hensyn til starttidspunkt og længde af ph.d. uddannelser.

Mistet indkomst under uddannelsen bliver således opgjort direkte og »realistisk« med hensyn til hvilke kandidat-aldre, der skal lægges til grund for sammenligning af

1. IDA databasen består af successive, årlige tværsnitsdata over hele befolkningen (og alle arbejdssteder), med start i 1980. Via uddannelsesdata kan identificeres årlige kohorter af nyuddannede, og i kraft af det fortløbende personnummer er der samtidig tale om med årlige paneldata. Med IDA data for f.eks. 2006 kan personer, der afsluttede en lang videregående uddannelse i 1980, følges frem til en kandidat-alder på 26 år. Dvs. at der om ca. 15 år vil begynde at foreligge data om kohorter (årgange) gennem hele arbejdslivet. For at analysere kohorteeffekter skal man dog have data om kohorter over flere år.

ph.d.'ers løn (stipendium) med kandidaters løn. For hele gruppen af ph.d.'ere vil der for kandidataldre omkring fire-fem år indgå både færdiguddannede ph.d.'er og ph.d. studerende, men det har ingen betydning for opgørelsen af en gennemsnitlig livsindkomst. Hvis man fortolker det gennemsnitlige uddannelsesforløb som det sandsynlige eller realistiske forløb, kan det også være et relevant grundlag for det enkelte individs beslutning om at erhverve en ph.d. grad.

Det antages implicit, at mistet indkomst er den eneste private omkostning ved uddannelsen, hvilket stemmer overens med, at den direkte udgift ved et ph.d. studium betales af universiteter mv. Den direkte omkostning ved uddannelsen skulle derimod fratrækkes ved en opgørelse af det direkte samfundsmæssige afkast af uddannelsen (social returns). Men også med hensyn til social returns gælder, at de faktiske, gennemsnitlige ph.d. studieforløb er det relevante grundlag.

I den uddannelsesøkonomiske litteratur diskuteres forskellige faktorer, som kan give en skævhed ved at anvende indkomst eller løn som mål for den humankapital, der er en følge af uddannelse. En lønforskel kan være påvirket af andre faktorer end selve uddannelsen. Det drejer sig for det første om forskelle i »medfødte« evner og familiebaggrund. Forskelle i evner kunne i vores sammenhæng indebære, at personer med en ph.d. grad alligevel, dvs. som kandidater, ville få en højere løn. Forældres uddannelse kan i kraft af sociale netværk have betydning for, om man får et højtlønnet job, Taubman og Wales (1973). I analysen kontrolleres for disse forhold via data om studenterksamenskarakter og forældres uddannelsesniveau samt fagområde. Endvidere kan der være en signalerings- eller screeningseffekt af uddannelse. Arbejdsgivere kan opfatte en højere uddannelse i sig selv som et signal om større produktivitet, og kan også være grundlag for screening af ansøgere. Denne effekt udgør imidlertid ikke en skævhed i relation til den private livsindkomst. Endelig kan der opstå en ændring i relativ løn som følge af en relativ ændring i udbuddet af uddannelseskategorier.

I relation til livsindkomst gælder, at målet principielt set skal være forskellen i løn efter skat. Vi vil imidlertid anvende bruttolønnen før skat, da kandidater og ph.d.'er typisk befinder sig på samme, højeste skattetrin og dermed har samme marginalskat. Det betyder, at den relative forskel i nettoløn er lig med forskellen i bruttolønnen reduceret med den marginale skattesats. Ud fra en incitamentssynsvinkel svarer det til, at de enkelte personer kun betragter forskellen i bruttoløn.

3. Regressionsmodel, variabler og population

Der anvendes en klassisk Mincer regressionsmodel med en række kontrolvariabler, Checchi (2006, s. 164); modellen bruges til estimation af både livsindkomst og afkast af humankapital:

$$\ln(w_{it}) = b_0 + b_1S_i + b_2Exp_{it} + b_3Exp_{it} \cdot S_i + b_4Exp_{it}^2 + b_5Exp_{it}^2 \cdot S_i + b_6A_i + b_7F_i + b_8X_{it} + v_i$$

Den afhængige variabel er den naturlige logaritme til timelønnen w for individet i på tidspunktet t . S er uddannelse, her en dummy variabel for ph.d. eller kandidat. Exp er erfaring, dvs. antal år efter dimission (jf. nedenfor). Exp^2 giver et 2. grads polynomium af erfaring, og interaktionen mellem S og erfaring frembringer separate lønprofiler over erfaring for ph.d.'er og kandidater. A er en indikator for evner (studentereksamenskarakter), F er familiebaggrund (forældres uddannelse), og X er yderligere personkarakteristika, der kontrollerer for generelle lønforskelle. Centrale variable er fagområde (naturvidenskab, samfundsvidenskab osv.) og sektor (forskning, offentlig sektor og privat sektor). Andre kontrolvariable er køn, virksomhedsstørrelse og arbejdsstedets geografiske beliggenhed.

Regressionsmodellen resulterer i ordinære mindste kvadraters estimater (OLS). En oversigt over skævheder ved OLS estimater gives i Checchi (2006, s. 165f, 200ff). Vores data har fordelagtige egenskaber i relation til skævheder. Med hensyn til udeladte variable har vi data om evner og forældres uddannelse. Men der kan være andre forhold, der har betydning for selektion af uddannelse, og som dermed kan indebære, at grupperne har forskellige karakteristika. Det drejer sig specielt om forskelle i interesser og præferencer. En løsning kunne bestå i anvendelse af instrumentvariablestimation IV. Problemet er imidlertid at finde en relevant instrumentvariabel. En nærliggende variabel er forældres uddannelsesniveau, men forældrebaggrund kan også direkte påvirke den afhængige variabel, dvs. lønnen; det kan ske via forældres større eller mindre netværk i relation til arbejdsgivere. Det er netop i denne betydning, at forældres uddannelse indgår som forklarende variabel i vores regressionsmodel. En delvis løsning på forskelle i karakteristika såsom interesser kan bestå i, at selvstændige (f.eks. praktiserende læger) ikke indgår i analysen af relativ indkomst, jf. senere.

Problemer med pålidelig registrering af uddannelse er ikke til stede, da data om personers uddannelse er baseret på løbende administrative indberetninger. Med hensyn til selektion af personer i uddannelsesgrupperne skal der her tages i betragtning, at der her er tale om personer med høje uddannelser. Følgelig burde der ikke forekomme selektion af individer i tilgangen til beskæftigelse baseret på forskelle i evner.

Variabler

Løn (w) er timelønnen i personens hovedbeskæftigelse pr. den første uge af november måned. I afkastet af uddannelse indgår således ikke supplerende løn i bijob, eller anden indkomst end arbejdsindkomst (af formue o.l.). Med hensyn til afkastet af uddannelse kunne den samlede lønindkomst være et alternativ, men inddragelse af løn i bijob ville være ensbetydende med at indregne (brutto)afkastet af flere arbejdstimer;

samtidig vil der ikke være tale om et mål for afkast af humankapitalen i betydningen større produktivitet. Bl.a. Christensen og Westergaard-Nielsen (2001) anvender også timeløn som mål.

Teknisk set er timelønnen i IDA databasen beregnet som en lønsam divideret med et skøn for antal arbejdstimer (via ATP bidraget). For ansatte med mindre end halv arbejdstid eller sæsonarbejde er antal arbejdstimer behæftet med usikkerhed, og disse personer udelades af populationen. For fuldtidsansatte er timelønnen pålidelig, og da de fleste højtuddannede arbejder på fuld tid, er dækningen høj. Pensionsbidrag er ikke inkluderet, men for vores uddannelsesgrupper må bidragsprocenten antages at være ensartet (15-17 procent).

Uddannelse (S) er en dummy-variabel for ph.d. (= 1) og kandidat (= 0). Uddannelsesmæssig status for hver person, inklusive afsluttet uddannelse, rapporteres årligt pr. 1. oktober fra samtlige uddannelsesinstitutioner. For de personer, der har afsluttede en uddannelse før 1970, er arten af uddannelse registreret ved folketællingen i 1970, men dimittendåret er uoplyst.

Erfaring (Exp). Ved estimation af livsindkomst er erfaring lig med antal år efter kandidatdimission (kandidatalder) for både ph.d.'er og kandidater, og ved estimation af humankapital er erfaring antal år efter henholdsvis ph.d. dimission og kandidatdimission. Erfaring tager ikke højde for perioder med f.eks. sabbat eller arbejdsløshed. I tværnsnitsdata for 2006 er erfaring trunckeret ved 36 år, svarende til antal år efter folketællingen i 1970.

Evner (A) operationaliseres ved hjælp af gennemsnitskarakteren ved studentereksamen. I princippet burde karakteren være oplyst for alle med en (nyere) studentereksamen o.l., men den er uoplyst for ca. 10 procent på grund af mangelfulde indberetninger fra uddannelsesinstitutioner. Da der ikke er en systematik heri, synes der ikke at foreligge et bortfaldsproblem. Dog skyldes en mindre del af bortfaldet, at nogle personer ikke har en dansk studentereksamen

Familiebaggrund (F) måles ved begge forældres uddannelse med hensyn til dels niveau og dels fagområde. Niveau er en kandidatgrad eller en ph.d. grad versus en ikke-akademisk uddannelse. For akademikerforældre sammenlignes forældrenes fagområde med fagområdet for den pågældende person (qua enhed i populationen). Sammenkædningen med forældre sker via referencen til forældrenes personnumre.² CPR registrats historik indebærer, at forældre ikke »findes« for personer, der fik en kandidatgrad før omkring 1985, svarende til at familiebaggrund mangler for »ældre« årgange med en erfaring på over ca. 20 år. For immigranter mangler oplysninger om familiebaggrund helt.

2. Teknisk set spores forældrene i hele populationen af kandidater og ph.d.'er år for år fra 1980 og frem. Hvis en forælder findes i grundpopulationen, er forælderen 'akademiker', ellers ikke.

Forældres uddannelse afspejler sociale netværk, Checchi (2006, s. 166); Taubman og Wales (1973). Hvis børn af akademikerforældre i højere grad vælger et ph.d. studium,³ og højere uddannede har et større, relevant netværk, kan ph.d.'er opnå højere aflønnede job. Denne effekt må formodes at være stærkere, når forældre er uddannet inden for samme fagretning. For en bredere fortolkning, se Bowles, Gintis and Osborne (2001, s. 1138).

Fagområde (universitetsfag) består i 7 hovedområder (naturvidenskab, samfundsvidenskab, teknik osv.).

På detaljeret niveau er der for nogle kandidatuddannelser foretaget et specifikt valg af tilhørsforhold, f.eks. mellem humaniora og samfundsvidenskab.

Sektor er en fordeling af erhverv på tre hovedgrupper: Forskning, offentlig administration mv. og privat sektor. Ved forskning forstås offentlig forskning, der omfatter højere læreanstalter og sektorforskningsinstitutioner. Da afgrænsningen er organisationer, omfatter forskning også akademikere i administrative job.

Øvrige kontrolvariabler er køn, virksomhedsstørrelse (over/under 200 ansatte i alt) og arbejdsstedets geografiske beliggenhed (hovedstaden/provinsen).

Population

Når indkomst defineres som løn medfører det, at populationen kun omfatter *lønmodtagere*. Arbejdsgivere og selvstændige er således ikke inkluderet, hvilket der også er et rationale bag. I vores tilfælde indebærer det, at eksempelvis praktiserende læger, tandlæger m.fl. ikke indgår. En ph.d.-grad i medicin retter sig ikke mod privat praksis, eller sagt med andre ord vil en kandidat, som forestiller sig at blive praktiserende læge, ikke vælge en ph.d. Populationen af lønmodtagere medvirker således til, at ph.d'er sammenlignes med kandidater, der i højere grad har lignende interesser og præferencer (og som potentielt kunne vælge et ph.d. studium). Et andet aspekt er, at der ved inklusion af selvstændige skulle anvendes samlet indkomst, som imidlertid kan indeholde anden indkomst end arbejdsindkomst.

Personer, der har fuldført en korterevarende »master« uddannelse, medtages ikke. Begrundelsen er, at en kandidatgrad (efter 5-6 års stadium) er den normale forudsætning for at blive immatrikuleret som ph.d. studerende.

Studentereksamenskarakter og forældreoplysninger vil ofte mangle for (nye) indvandrere. Følgelig indgår kun personer med *dansk oprindelse*, dvs. individer født i Danmark og hvis forældre er danske statsborgere. Man kan også anlægge det synspunkt, at der kan være kendetegn ved indvandrere, for hvilke der mangler variabler (f.eks. med hensyn til sproglige færdigheder).

3. Jf. Mattsson og Munk (2008, s. 15).

4. Analyser

Deskriptiv statistik vises i tabel 1 for henholdsvis kandidater, ph.d. dimittender og ph.d. studerende. For det første ses tal for de centrale numeriske variabler timeløn og erfaring. Ekstreme værdier for timeløn er ekskluderet ved at udelade observationer med en timeløn på under 1. percentil eller over 99. percentil. For ph.d. dimittender vises specielt, at ph.d. graden typisk opnås 5 år efter kandidatdimission (median observationen). For det andet fokuseres på de to personkendetegn, som kunne forklare lønforskelle uafhængigt af uddannelse, dvs. karakter i gymnasium og forældres uddannelse. Det viser sig, at karakterniveauet er højere for ph.d.'er end for kandidater, idet forskellen er ca. et halv point på 13-skalaen. Ph.d.'er har i højere grad end kandidater forældre med en lang videregående uddannelse. 28 procent af ph.d. dimittenderne har en far med en lang videregående uddannelse, mens det gælder for 20 procent af kandidaterne.

Regressionsanalyser af timeløn rapporteres i tabel 2. Der indgår to forskellige populationer, som adskiller sig fra hinanden med hensyn til definitionen af variabelen erfaring, jf. tidligere. Populationen vedrørende livsløn findes i de to første kolonner, og populationen vedrørende humankapital i den tredje kolonne. For livsløn indgår erfaring (kandidatalder) over 0-36 år,⁴ mens erfaring vedrørende humankapital omfatter 0-30 år.⁵

Sektor er kritisk for estimation af den relative livsløn for ph.d.'er. Det skyldes kombinationen af den markante generelle lønforskel mellem specielt forskning og den private sektor (på 13 procent) og det forhold, at ph.d.'er i langt større omfang end kandidater er beskæftiget inden for offentlig forskning. Som følge heraf indgår sektor både som eksogen og endogen variabel. I første tilfælde, dvs. hvor variabelen sektor indgår, bliver lønforskellen mellem ph.d.'er og kandidater baseret på en ligelig fordeling på sektorer; lønforskellen er gennemsnittet af lønforskellen pr. sektor. Når sektor derimod er endogen, dvs. at variabelen sektor ikke indgår, er fordelingen på sektorer af henholdsvis ph.d.'er og kandidater lig med den faktiske fordeling; dermed indgår det lavere lønniveau i forskning med en forholdsvis stor vægt i ph.d.'ers lønniveau. Den faktiske fordeling kan fortolkes som den sandsynlige eller forventelige fordeling på sektorer, svarende til at sektor er forudbestemt. For estimationen af humankapital indgår sektor kun som eksogen variabel.

Regressionsanalyserne i tabel 2 viser, at koefficienter til ph.d. versus kandidat (b_1 - b_5) er signifikante på 1 procent niveau, og at det samme gælder for kontrolvariabler in-

4. Erfaring på 36 år medtages, selv om der her er en trunkering, idet alle som er dimitteret før 1970 får en erfaring (kandidatalder) på 36 år, jf. tidligere. En inspektion af data viser imidlertid, at lønnen for 36 år ligger på samme niveau som lønnen for 35 års erfaring, og at standardafvigelsen også ligger på samme niveau.

5. Valget af 0-30 år beror på, at det absolutte antal ph.d.'er i 2006 med mere end 30 år fra ph.d. dimission er lavt. Kravet har været, at antal ph.d.'er udgør mindst fem pr. celle (kategori/erfaring).

Tabel 1. Deskriptiv statistikk, 2006.

<i>Kandidater</i>										
	N	N Miss.	Gns.	Median	N. Kvartil	Ø. Kvartil	Std. afv.	Min.	Maks.	
Timeløn	140.906	0	289	262	218	330	109	123	870	
Antal år siden kandidateksamen	140.906	0	13	11	4	21	10	0	36	
Karakter ved studentereksamen	86.327	54.579	8,8	8,9	8,3	9,4	0,8	5,8	13,0	
Far: Lang videreg. uddannelse	114.612	26.294	0,20	0	0	0	0,40	0	1	
Mor: Lang videreg. uddannelse	116.368	24.538	0,07	0	0	0	0,26	0	1	
Far: Samme fagområde	23.299	117.607	0,30	0	0	1	0,46	0	1	
Mor: Samme fagområde	8.414	132.492	0,27	0	0	1	0,44	0	1	
<i>ph.d. dimittender</i>										
	N	N Miss.	Gns.	Median	N. Kvartil	Ø. Kvartil	Std. afv.	Min.	Maks.	
Timeløn	10.140	0	308	281	244	350	97	123	864	
Antal år siden kandidateksamen	9.415	725	14	14	9	18	7	1	36	
Antal år fra kandidat til Ph.d.	9.415	725	6	5	4	8	4	1	26	
Karakter ved studentereksamen	7.260	2.880	9,3	9,3	8,7	9,8	0,8	5,6	13,0	
Far: Lang videreg. uddannelse	8.914	1.226	0,28	0	0	1	0,45	0	1	
Mor: Lang videreg. uddannelse	9.067	1.073	0,10	0	0	0	0,30	0	1	
Far: Samme fagområde	2.457	7.683	0,30	0	0	1	0,46	0	1	
Mor: Samme fagområde	926	9.214	0,25	0	0	1	0,43	0	1	
<i>ph.d. studerende</i>										
	N	N Miss.	Gns.	Median	N. Kvartil	Ø. Kvartil	Std. afv.	Min.	Maks.	
Timeløn	3.262	0	207	190	170	231	56	125	756	
Antal år siden kandidateksamen	3.261	1	4	3	2	6	4	0	31	
Karakter ved studentereksamen	2.888	374	9,3	9,4	8,8	9,9	0,8	6,1	12,0	
Far: Lang videreg. uddannelse	3.189	73	0,31	0	0	1	0,46	0	1	
Mor: Lang videreg. uddannelse	3.220	42	0,15	0	0	0	0,36	0	1	
Far: Samme fagområde	993	2269	0,36	0	0	1	0,48	0	1	
Mor: Samme fagområde	482	2780	0,28	0	0	1	0,45	0	1	

Tabel 2. Timeløn over erfaring for ph.d.'er versus kandidater, tværsnitsdata 2006. Populationer vedrørende henholdsvis livsløn og afkast af humankapital. Afhængig variabel log timeløn.

	Erfaring: Antal år kandidat dimission (livsløn) ¹		Erfaring: Antal år fra fra ph.d. dimission hhv. kandidat dimission (humankapital) ²
	Sektor eksogen	Sektor endogen	
Intercept (b_0)	5,28***	5,38***	5,24***
S = ph.d. (b_1)	-0,19***	-0,26***	0,17***
Erfaring (b_2)	0,038***	0,037***	0,042***
Erfaring*ph.d. (b_3)	0,019***	0,023***	-,010***
Erfaring ² (b_4)	-0,001***	-0,001***	-,001***
Erfaring ² *ph.d. (b_5)	-0,000***	-0,001***	0,000***
Kandidat = 0			
Humaniora	-0,08***	-0,11***	-0,08***
Naturvidenskab	-0,01***	-0,04***	-0,02***
Pædagogik	0,11***	0,06***	0,11***
Samfundsvidenskab	0,08***	0,06***	0,08***
Sundhedsvidenskab	0,20***	0,17***	0,22***
Jordbrug, veterinær	-0,06***	-0,06***	-0,05***
– Teknik = 0			
Offentlig adm. mv.	0,04***	.	0,04***
Privat sektor	0,13***	.	0,13***
Sektor u.n.a.	0,08	.	0,11**
– Forskning = 0			
Karakter, høj (9,0+)	0,04***	0,03***	0,04***
Karakter, lav (5,5-6,9)	-0,03***	-0,03***	-0,04***
Karakter u.n.a.	-0,01***	-0,01***	0,00**
– Middelniveau (7,0-8,9) = 0			
Far: lang videreg. udd.	0,00	0,00	0,00*
Fars uddannelse u.n.a.	-0,02***	-0,03***	-0,02***
Mor: lang videreg. udd.	0,00	0,00	0,00
Mors uddannelse u.n.a.	-0,02***	-0,02***	-0,02***
Far: Samme fagområde	0,01**	0,01	0,01**
Mor: Samme fagområde	-0,01	0,00	0,00
<i>N</i>	153.582	153.582	140.052
<i>R</i> ²	0,36	0,34	0,37

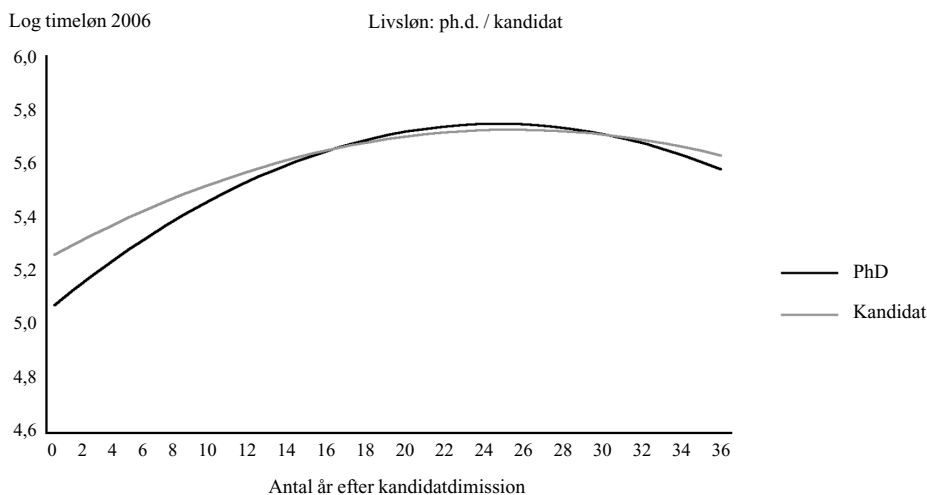
(1) ph.d. populationen inkluderer ph.d. studerende. Erfaring (kandidatalder) 0-36 år.

(2) Sektor eksogen. Erfaring 0-30 år.

• Stjerner ***, **, * angiver signifikansniveauer på henholdsvis 1, 5, and 10 procent.

• Yderligere kontrolvariabler er køn, arbejdsstedets beliggenhed og virksomhedsstørrelse.

• Referencekategorier for dummyvariabler: S (uddannelse): Kandidat; Fagretning: Teknik; Sektor: Forskning; Karakter: Middelniveau (7,0-8,9); Uddannelsesniveau for far/mor: Ikke lang videregående uddannelse; Fars/mors fagområde: Ikke samme; Køn: Mand; Arbejdsstedets beliggenhed: Hovedstaden; Virksomhedsstørrelse: SMV (under 200 ansatte).



Figur 2. Lønprofiler over kandidatalder: Relativ livsløn. Gennemsnit af alle fagområder og sektorer.

klusive karakterniveau (med positive koefficienter). For forældres uddannelse er koefficienter til lang videregående uddannelse ikke signifikante, set i forhold til personer uden en lang videregående uddannelse.⁶ Dog er koefficienten til 'far: samme fagområde' signifikant og positiv.

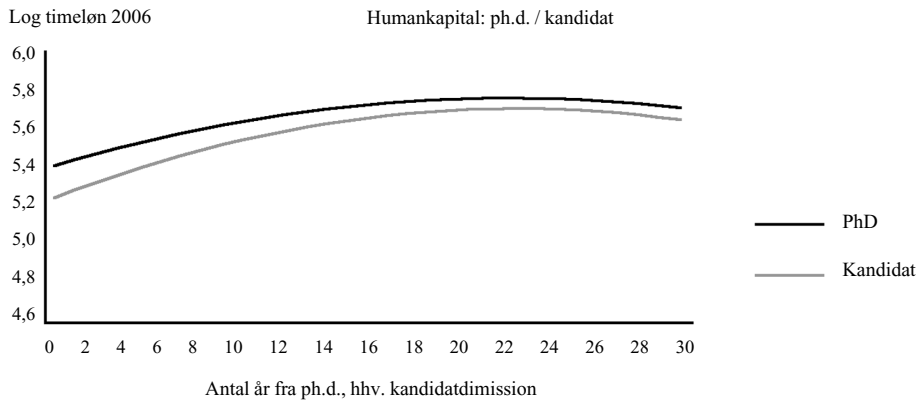
Lønprofiler over erfaring

Via tabel 2 kan man beregne lønprofiler over erfaring baseret på koefficienterne b_1 - b_5 . Det illustreres i figur 2 og 3, hvor sektor i begge tilfælde er eksogen.

Figur 2 viser lønprofiler efter kandidatdimission, dvs. at arealet under kurverne udtrykker den relative livsløn. Figuren illustrerer tydeligt den mistede indkomst under ph.d. studiet, og at ph.d.'er desuden har et lønefterslæb i nogle år efter ph.d. dimission, hvor kandidater har opnået en længere job erfaring efter færdiggørelse af uddannelsen. Ved en kandidatalder på omkring 12 år bliver ph.d.'ers lønniveau højere, hvilket kan betragtes som bruttoafkastet af ph.d. uddannelse. Denne lønforskel bliver mindre ved en høj kandidatalder.

Figur 3 viser lønudviklingen efter færdiggørelse af henholdsvis ph.d.- og kandidatuddannelse, og her udtrykker afstanden mellem kurverne den relative humankapital. Når en nyuddannet ph.d.'er sættes i relation til en nyuddannet kandidat fås et ganske

6. Koefficienter til uoplyst uddannelsesniveau er imidlertid signifikante og negative. Denne kategori vil omfatte de ældre årgange, og den negative koefficient stemmer overens med den aftagende relative løn for ph.d.'er med voksende erfaring, jf. figur 2 og 3.



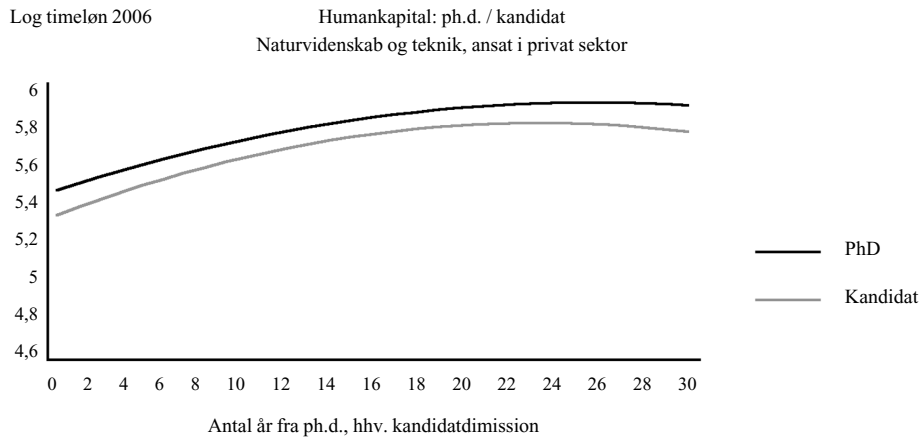
Figur 3. Lønprofiler fra afslutning af kandidat- eller ph.d. uddannelse: Relativ humankapital. Gennemsnit af alle fagområder og sektorer.

anderledes billede end i figur 2. Startlønnen for ph.d.'er er omkring 17 procent højere end for kandidater (jf. tabel 2), men væksten i ph.d.'ers løn er mindre. Det betyder, at lønningerne konvergerer, hvilket umiddelbart kunne være et tegn på en signalerings- eller screeningseffekt. Her er der dog tale om et summarisk grundlag bestående af alle fagområder og sektorer under ét. Derfor vises som supplement i figur 4 lønprofiler for dimittender inden for natur- og teknisk videnskab beskæftiget i den private sektor. For denne specifikke gruppe er lønprofilerne i mindre grad konvergerende, jf. også de lave koefficienter til interaktionen mellem erfaring og ph.d. angivet i noten til figur 4. En nærmere analyse måtte baseres på kohortedata.

Livsløn og humankapital

Den relative livsløn og humankapital af ph.d. uddannelser beregnes via summen af de estimerede timelønninger pr. år over erfaring, idet summen for ph.d.'er sættes i relation til summen for kandidater. Til sammenligning med andre beregningsmetoder gælder, at hvis denne simple sum af timelønninger multipliceres med antal arbejdstimer pr. år (2006), vil man få nutidsværdien af årlige samlede lønninger over arbejdslivet; her ville det være en implicit forudsætning, at antal arbejdstimer er ens for ph.d.'er og kandidater.

I litteraturen inden for uddannelsesøkonomi er det sædvane at diskontere indkomststrømmen, idet der typisk beregnes en intern rente af omkostningen ved uddannelsen. Her vil vi imidlertid følge Det økonomiske Råd, som for en del år siden ophørte med at diskontere indkomsterne via en (arbitrær) rentesats. Argumentet er, at reallønsstigninger korresponderer med den reale diskonteringsrate, Hovard and Knust (1997). Endvi-



Figur 4. Lønprofiler fra afslutning af kandidat- eller ph.d. uddannelse: Relativ humankapital. For personer inden for fagområder naturvidenskab og teknik ansat i den private sektor:

Regressionsligning: $\ln(w) = 5,378^{***} + 0,134^{***} \cdot s + (0,044^{***} + -0,006^{**} \cdot s) x + (-0,0010^{***} + 0,0002^{**} \cdot s) x^2$, hvor s er lig med ph.d. for $s = 1$ og kandidat for $s = 0$; x er erfaring 0-30 år; $N = 24.979$ / $R^2 = 0,32$.
Kumuleret løn 0-30 år, ph.d. over kandidat = 11,3 procent (jf. tabel 3).

dere finder man, at beregning af en intern rente er meget følsom over for opgørelsen af uddannelsesomkostningen, Det økonomiske Råd (2003).

Formlen til beregning af de kumulerede lønninger $Kum(w)$ er:

$$Kum(w) = \sum_{x=0}^n \exp(b_0 + b_1 \cdot S + (b_2 + b_3 \cdot S) x + (b_4 + b_5 \cdot S) x^2)$$

x er erfaring i år 0, 1, ..., n . Da $S = 0$ for kandidat, indgår kun koefficienterne b_0 , b_2 , og b_4 i summen for kandidater. For livsløn er erfaring 0-36 år og for humankapital 0-30 år. Den relative livsløn og humankapital af ph.d. uddannelser er lig med relationen $Kum(w)_{ph.d.} / Kum(w)_{Kand.}$.

Resultaterne vises i tabel 3, hvor der opdeles i henholdsvis seks fagområder og de tre sektorer. For alle fagområder og sektorer under ét er resultatet, at ph.d.'ers livsløn er lavere end kandidaters, i størrelsesordenen 2 eller 6 procent lavere afhængigt af, om sektor er henholdsvis eksogen eller endogen. Humankapitalen er ca. 8 procent højere, når sektor er eksogen (jf. også figur 3); det er i samme størrelsesorden som i Wilson (1987).

Generelt set er der tale om nogenlunde ensartede størrelsesordener på tværs af fag og specielt over sektorer. Da humankapital er estimeret med eksogen sektor, skal der

Tabel 3. Relativ livsløn og humankapital af ph.d. uddannelse, tværsnitsdata 2006. Separate regressionsanalyser pr. henholdsvis fagområde og sektor identisk med modellen i tabel 2.

	Livsløn		Humankapital
	Sektor eksogen	Sektor endogen	Sektor eksogen
	----- ph.d. over kandidat, procent -----		
Alle fagområder og sektorer	-2,4	-6,3	8,4
Humaniora	-5,0	-6,2	10,8
Naturvidenskab	-2,8	-6,9	5,8
Samfundsvidenskab	-2,5	-10,5	10,1
Teknik	0,8	-2,5	11,3
Jordbrug, veterinær	1,1	-1,1	12,6
Sundhedsvidenskab	-5,9	-9,5	5,7
<i>Sektor 2006, alle fagområder:</i>			
Forskning	-3,3	.	9,3
Offentlig adm. mv.	-3,9	.	7,6
Privat sektor	1,8	.	11,1

Pædagogik er ikke medtaget på grund af et lille antal ph.d.'er.

her sammenlignes med tilsvarende livsløn i kolonne 1. Ved forklaring af afvigelser fra det overordnede mønster gælder, at der er tale om flere faktorer. Den mistede indkomst under ph.d. studiet og lønefterslæbet efter ph.d. dimissionen afhænger dels af placeringen af ph.d. uddannelsen i forhold til kandidattidspunktet og dels af den samtidige vækst i lønnen for kandidater. Lønnen (stipendiet) til ph.d. studerende er imidlertid den samme for fagene, dog kan lønnen for Erhvervs-ph.d.'er være højere end den overenskomstmæssige løn på universiteter.

Forskellen i livsløn mellem eksogen sektor og endogen sektor kan forklares mere entydigt af den erhvervmæssige placering af ph.d.'er i forhold til kandidater. For humaniora vil fordelingen af ph.d.'er og kandidater på sektorer være nogenlunde ens (en stor andel af begge i den offentlige sektor), mens denne fordeling er meget forskellig for samfundsvidenskab (ph.d.'er inden for forskning, kandidater i den private sektor).

5. Afsluttende bemærkninger

Vi har fundet, at der er en negativ livsløn ved at tage en ph.d. grad frem for at gå ud i beskæftigelse som kandidat. Når den faktiske fordeling af ph.d.'er på sektorer lægges til grund, dvs. at en ph.d. grad prædeterminerer ansættelsesområdet, er den relative livsløn lavest. Sidstnævnte må antages at være det mest adækvate mål, idet en ph.d. grad overvejende retter sig mod offentlig forskning.

Den mindre livsløn må indebære, at ikke-monetære afkast af en ph.d. grad har en stor betydning for valget af ph.d. uddannelse. Det kan bestå i en prioritering af arbejde med teoretiske og metodiske problemstillinger eller med en høj grad af autonomi.

Vi finder, at timelønnen efter færdiggørelse af ph.d. uddannelsen er højere end nyuddannede kandidaters, svarende til en højere humankapital. Hvis monetære afkast også spiller ind på valget af en ph.d. uddannelse, kunne den højere timeløn som ph.d. dimittend være et forhold, som i større eller mindre grad kompenserer for den lavere livsløn.

Litteratur

- Ashenfelter, O. og Mooney, J. D. 1968. Graduate Education, Ability, and Earnings, *Review of Economics and Statistics*, vol. 50, nr. 1, s. 78.
- Bowles, S., Gintis, H. og Osborne, M. 2001. The Determinants of Earnings: A Behavioral Approach, *Journal of Economic Literature*, vol. 39, nr. 4, s. 1137.
- Checchi, D. 2006. *The Economics of Education. Human Capital, Family Background and Inequality*, Cambridge University Press.
- Christensen, J. J. og Westergård-Nielsen, N. 2001. Afkast af human kapital i Danmark, 1981-1995, *Nationaløkonomisk Tidsskrift*, vol. 139, nr. 2, s. 117.
- Det økonomiske Råd. 2003. *Dansk økonomi*, efterår 2003.
- Hartog, J. og van den Brink, Henriëtte Maassen (eds). 2007. *Human Capital. Advances in Theory and Evidence*, Cambridge University Press.
- Hovard, T. og Knust, P. L. 1997. *Stiliserede livsindkomster for 1995*. Det økonomiske Råd, Arbejdspapir 1997: 1.
- Johnes, G. 1993. *The Economics of Education*, MacMillan.
- Mattsson, C. og Munk, M. D. 2008. *Social uddannelsesmobilitet på kandidat- og forskeruddannelser*, Det Nationale Forskningscenter for Velfærd (SFI), publikation 08:33.
- Siegfried, J. J. 1971. Rate of Return to the Ph.D. in Economics, *Industrial and labor relations review*, vol. 24, nr. 3, s. 420-31.
- Stock, W. A. og Siegfried, J. J. 2001. So you want to earn a Ph.D. in Economics: How much do you think you'll make, *Economic inquiry*, vol. 39, nr. 2, s. 320.
- Taubman, P. J. og Wales, T. J. 1973. Higher Education, Mental Ability, and Screening, *Journal of Political Economy*, vol. 81, nr. 1, s. 28.
- Wilson, R. A. 1987. The determinants of the earnings of professional engineers in Great Britain in 1981, *Applied Economics*, vol. 19, nr. 7, s. 983.

Den økonomiske vækst og miljøet – nogle centrale forskningsresultater i det 21. århundrede

Poul Schou

De Økonomiske Råds Sekretariat, E-mail: ps@dors.dk

SUMMARY: The paper surveys the main theoretical research and empirical findings from the last decade about the connection between economic growth and the environment. These relate to whether continued economic growth is feasible and how it may be achieved, given environmental constraints due either to resource scarcity or waste accumulation. Recent research has clarified a number of issues, but also pointed out new research questions, e.g. the so-called Green Paradox, which highlights the danger that some climate policies may actually aggravate the greenhouse problem. One of the key issues in this area of research is still the assumptions made related to the substitution possibilities between man-made and natural (renewable or non-renewable) capital. At least unitary elasticities of substitution are necessary conditions for sustainable growth in the long run. Empirical research offers little help settling this issue because of mixed evidence and methodological problems. The paper also discusses issues related to economic growth, climate change and population growth.

1. Indledning og grundlæggende problemstillinger

Sammenhængen mellem økonomisk vækst og forskellige former for miljøproblemer er en problemstilling, der til stadighed optager både brede dele af befolkningen og også dele af den økonomiske sagkundskab. Er det ønskeligt at blive ved med at fremstille et stadig større BNP, hvis produktionen samtidig skaber et biprodukt i form af diverse miljøproblemer? Er det overhovedet muligt at forestille sig en permanent positiv økonomisk vækst i fremtiden, uden at vi løber ind i en miljømæssig katastrofe, som også før eller siden vil formindske den økonomiske velstand? Eller kan man tværtimod forestille sig, at den økonomiske – og herunder også den teknologiske – udvikling vil skabe nye muligheder for at få et renere miljø, sådan at fornyet økonomisk fremgang ligefrem også i sig selv kan medføre en miljømæssig gevinst? En række økonomiske forskere inden for områderne miljøøkonomi og makroøkonomisk vækstteori har be-

Christian Groth og to anonyme skøns mænd takkes for værdifulde kommentarer.

skæftiget sig med disse spørgsmål i tidens løb. De seneste par årtier har forskningen især taget udgangspunkt i den omfattende litteratur om endogen økonomisk vækst, som fra slutningen af 1980'erne fornyede de grundlæggende økonomiske vækstteorier. Artiklen her forsøger at give et overblik over den udvikling i den fagøkonomiske tankegang om disse spørgsmål, som har fundet sted i den økonomiske forskning i løbet af den seneste halve snes år. Den kan dermed også ses som en slags fortsættelse af Schou (2000a), som omtaler en række tidligere bidrag.

Der er to grundlæggende problemstillinger, som hver især på forskellige tidspunkter har domineret såvel den offentlige som den faglige debat om, i hvilket omfang det omgivende miljø sætter grænser for den økonomiske vækst. For det første anvender vi direkte produktionsfaktorer fra naturen i form af naturressurser, og disse midler er begrænsede, hvilket potentielt begrænser produktionen. For det andet udleder produktionen i en række tilfælde forskellige former for forurening, som opsamles i det omgivende miljø, og dette kan også påvirke såvel produktionsprocessen som vores velfærd generelt skadeligt på længere sigt. På engelsk snakkes om naturen som henholdsvis »source« (kilde) og »sink« (afløb) for produktionen. En alternativ skillelinje er opdelingen i ikke-fornybare naturressurser som fossile brændstoffer og mineraler og fornybare resurser som fisk, planter, husdyr mv. Forureningsspørgsmål kan i mange situationer betragtes og analyseres som udnyttelse af en fornybar resurse i form af naturens regenererende kapacitet – den omgivende natur kan absorbere en vis mængde forureningsstoffer uden større skade, men bliver udledningen for stor, vil det medføre skadevirkninger af forskellig slags. Man kan dermed analysere betydningen af såvel source- som sink-problemer for den langsigtede vækst ved at inddrage henholdsvis ikkefornybare og fornybare naturressurser i de modeller, som i øvrigt udgør state of the art inden for moderne vækstteori. Dette uddybes nærmere i de følgende afsnit.

2. Ikkefornybare naturressurser

Lad os først se på spørgsmålet om de ikkefornybare naturressurser. Her har forskningen i dette årtusinde leveret en række nye bidrag om mulighederne for positiv økonomisk vækst på lang sigt, selvom vi er nødt til at bruge disse resurser i stadig mindre omfang fremover.

Antag, at makroproduktionsfunktionen ser ud på følgende måde:

$$Y = f(K, L, V, R), \quad (1)$$

hvor Y er den samlede produktion, K beholdningen af fysisk kapital, L arbejdskraften (og dermed beholdningen af humankapital), V det teknologiske niveau (som også kan betragtes som en kapitalbeholdning, nemlig videnskapital) og R er forbruget af en

ikkefornybar resurse som eksempelvis jern eller fossile brændstoffer. Allerede i 1970'erne viste man, at substitutionseleasticiteten mellem naturresursen og den menneskeskabte kapital er afgørende for, om man på lang sigt kan opretholde positiv produktionsvækst (eller blot et konstant produktionsniveau), selvom det vil være nødvendigt i stadig stigende omfang at spare på resurseforbruget, der jo nødvendigvis må blive mindre og mindre, efterhånden som tiden går. Hvis substitutionseleasticiteten mellem R og den menneskeskabte kapital asymptotisk er større end én, efterhånden som resurseforbruget går mod nul, er det muligt at producere uden overhovedet at anvende resursen som input, og den lægger dermed ingen begrænsninger på langsigtvæksten. Hvis denne substitutionseleasticitet omvendt er lavere end én asymptotisk, er det nødvendigt at anvende naturresursen for at kunne producere overhovedet, og samtidig er dens grænseprodukt begrænset opadtil. I dette tilfælde får man det nedslående resultat, at ikke alene er økonomisk vækst umulig på lang sigt, men det vil være umuligt at opretholde et hvilket som helst positivt produktionsniveau i det lange løb. Forbrugsmulighederne vil uvægerligt gå mod nul før eller senere. Grænsetilfældet med en substitutionseleasticitet på netop én (Cobb-Douglas-tilfældet) er derfor det, der oftest er blevet studeret i litteraturen, fordi det er den eneste situation, hvor konklusionen om muligheden for positiv økonomisk vækst på lang sigt ikke på forhånd er indlysende. Lærebogen af Dasgupta og Heal (1979), som er en klassiker på området, giver en grundig indføring i disse problemstillinger og de tidlige bidrag.

Her er det dog nødvendigt at indskyde en præciserende bemærkning om, *hvilke* to produktionsfaktorer det er, som denne betingelse mht. substitutionseleasticiteten gælder for. I 1970'ernes litteratur opfattede man almindeligvis den teknologiske udvikling (V 's rolle i ovennævnte produktionsfunktion) som noget eksogent givet og tænkte på menneskeskabt kapital som et relativt snævert begreb (opfanget af K ovenfor), jf. eksempelvis Stiglitz (1974). Derfor opfattede man muligheden for resursebesparende teknologiske fremskridt som en selvstændig alternativ løsning på grænser for vækstproblemet: Selvom substitutionseleasticiteten mellem K og R er mindre end én, kan permanent positiv vækst alligevel være mulig, hvis V stiger – og substitutionseleasticiteten mellem V og R er mindst én. Her har den nyere litteratur om endogen vækst været med til at udvide den måde, vi tænker om begrebet »kapital« på. I nærværende sammenhæng er den afgørende skelnen mellem menneskeskabt kapital på den ene side og naturkapital på den anden side. Menneskeskabt og dermed reproducerbar kapital kan være indeholdt i hvert af de områder, man traditionelt opfatter som K , L og V i produktionsfunktioner som den ovenstående. Den nye vækstteori har dermed præciseret, at den nødvendige betingelse om en substitutionseleasticitet på mindst én mellem resursen og menneskeskabt kapital stadig er afgørende, men at det gælder et kapitalbegreb i ret bred betydning, som inkluderer menneskelig opfindsomhed og teknologiske frem-

skridt. Bidrag som Bretschger og Smulders (2003) og di Maria og Valente (2008) demonstrerer, at man godt kan få positiv vækst, selvom substitutionselasticiteten mellem ressourcen og fysisk kapital i snævrere forstand er mindre end én. I deres modeller er der til gengæld særlige endogene resursebesparende teknologiske fremskridt. Kravet om en enheds-substitutionselasticitet dukker dermed i denne modeltradition op i sammenhængen mellem denne nye form for menneskeskabt (teknologi-)kapital og ressourcen, idet deres antagelse svarer til det specialtilfælde, hvor (1) får udseendet

$$Y = f(K, L, V) R. \quad (2)$$

Her er V altså specifikt resursebesparende tekniske fremskridt, og da variabelen her indgår en multiplikativ forbindelse med ressourcen R , kan produktionen godt vokse permanent, selvom substitutionselasticiteten mellem K og R er lavere end 1, blot V vokser i tilstrækkeligt omfang. Om man ønsker at omtale dette tilfælde som et alternativ til kravet om en enheds-substitutionselasticitet eller tværtimod betragter det som en bekræftelse, hvor den afgørende menneskeskabte kapital nu blot er videnskapitalen V i stedet for den fysiske kapitalbeholdning K , er vel først og fremmest en smagssag. Når man inddrager empiriske vidnesbyrd, er det dog vigtigt at være opmærksom på, om de empiriske og teoretiske begreber modsvarer hinanden korrekt, jf. næste afsnit.

3. Empiriske resultater om substitutionselasticitetens størrelse

På grund af den afgørende betydning af disse substitutionselasticiteter er det naturligt at se nærmere på, hvad økonometrikere har fundet frem til mht. den historiske størrelsesorden heraf. Der er i tidens løb foretaget mange undersøgelser i forskellige sammenhænge af partielle substitutionselasticiteter på makroniveau mellem f.eks. energi og (fysisk) kapital. Neumayer (2000) er en udmærket oversigtsartikel over bl.a. en række af disse tidligere empiriske undersøgelser. De konkluderer efter en diskussion af forskellige metodiske problemer, at vi ikke har noget troværdigt svar på spørgsmålet om, hvorvidt energi og menneskeskabt kapital er komplementær eller substitut. I en nyere oversigt, Markandya og Pedroso-Galinato (2007), bringes dels en tilsvarende tabel over forskellige tidligere studier af aggregerede substitutionselasticiteter; dels præsenterer forfatterne et nyt sæt estimater, som bygger på et verdensbank-studie. Baseret på data for en aggregeret produktionsfunktion med fire input (fysisk kapital, humankapital, ikkefornybare energiresurser og jordresurser) for godt 200 lande finder de frem til substitutionselasticiteter mellem energi og fysisk kapital, der strækker sig fra 0,37 til 1,57, afhængigt af specifikationen. Også de tidligere studier, der bringes i en oversigt i artiklen, har estimeret substitutionselasticiteter mellem forskellige former for ikkefornybare naturressurser på den ene side og et bestemt kapitalbegreb på

den anden side, der ligger både over og under den afgørende værdi én. En nyere undersøgelse, der giver et eksempel på det sidste, er van der Werf (2008). Her foretages en systematisk analyse af substitutionselasticiteterne mellem kapital, arbejdskraft og energi for en række brancher i 12 OECD-lande. Forfatterens formål er at undersøge, hvilket belæg der er for de forskellige antagelser om substitutionselasticiteten i en række anvendte klimamodeller som WITCH, GREEN, MIT EPPA, WIAGEM, MERGE og ENTICE. I dette studie bliver Cobb-Douglas-egenskaberne generelt afvist. I stedet findes substitutionselasticiteter mellem et KL -aggregat og energi i størrelsesordenen 0,1 til 0,6 – og mellem K og L i størrelsesordenen 0,2 til 0,6.

De forskellige undersøgelser kommer altså med forskellige svar på spørgsmålet om substitutionselasticiteternes størrelse historisk, og da der i det hele taget er data- og metodemæssige problemer forbundet med beregning af substitutionselasticiteter i denne sammenhæng, kan man næppe sige, at vi indtil videre har noget særligt fyldestgørende svar på spørgsmålet. Således giver tværsnits- og tidsserieanalyser typisk estimater, der afviger fra hinanden, og i praksis begrænses man ofte til at anvende et ret snævert afgrænset kapitalbegreb i analyserne, i modsætning til, hvad de moderne teoretiske vækstmodeller egentlig tilsiger. En yderligere problemstilling er den følgende: De foretagne estimationer (og de fleste parameteriserede teoretiske modeller) antager ofte uden videre, at substitutionselasticiteten er konstant over tid. Det kan man dog ikke regne med – fleksibiliteten i produktionsfunktionen for moderate værdier af resurse/kapitalforholdet kan godt tænkes at være anderledes end den asymptotiske tilstand, når dette forhold går mod nul, og det er denne sidste asymptotiske værdi, der er den interessante, når vi beskæftiger os med spørgsmålet, om endeligheden af ressourcerne i sidste ende vil sætte grænser for væksten. Jones (2005) viser, at hvis de ideer, som ligger bag de teknologiske fremskridt i en vidensdrevet vækstmodel, er Paretofordelte, så fører det til, at den langsigtede globale produktionsteknologi faktisk er en Cobb-Douglas-produktionsfunktion. Dette forskyder dog blot problemet til et spørgsmål om, hvorvidt der er grund til at tro, at ny teknologi faktisk er fordelt på denne måde. Selvom Jones henviser til forskellige empiriske vidnesbyrd om, at bl.a. patenter og citater i videnskabelige tidsskrifter i nogle konkrete sammenhænge har været Paretofordelte, er der igen ingen tvingende grunde til at tro, at dette forhold altid vil være konstant.

Man kan dermed foreløbig konkludere, at der indtil videre ikke er nogen empirisk konsensus om, hvilken størrelsesorden den relevante substitutionselasticitet mellem ikkefornybare (især energi-)ressurser og menneskeskabt kapital har været historisk; og at sådan et resultat, selv hvis det havde foreligget, ikke ville give noget klart svar på, hvad den *fremtidige* elasticitet så vil blive, efterhånden som en større andel af den nuværende resursebeholdning bruges op. Med denne indsigt vil vi i næste afsnit se på, hvad den teoretiske litteratur siger om de langsigtede vækstmuligheder, givet at vi er i

Cobb-Douglas-tilfældet med en substitutionselasticitet på netop én – altså som nævnt det eneste tilfælde, hvor det langsigtede resultat ikke på forhånd er givet.

4. Cobb-Douglas-tilfældet

Der er efterhånden en lang række bidrag i moderne vækstteori, hvor man på den ene side eksplicit forklarer den teknologiske udvikling inden for modellens rammer og på den anden side anerkender, at produktionen af forbrugsvarer også afhænger af en essentiel ikkefornybar naturresurse. Schou (2000b) anvender en model, hvor vækstmotoren er uddannelse (akkumulation af humankapital) i stil med Lucas (1988). Scholz og Ziemes (1999), og Schou (2002b) bruger en model med forskning og udvikling, der fører til stigende specialisering i form af stadig flere former for halvfabrikata à la Romer (1990) (horizontal differentiering), mens Grimaud og Rougé (2003) som basis har en schumpetersk model med stadig stigende vertikal differentiering i lighed med Aghion og Howitt (1992). Disse modeller har alle det tilfælles, at forbrugsgoderne såvel som investeringerne i fysisk kapital produceres i en Cobb-Douglas-produktionsfunktion, hvor såvel naturressurser som humankapitalen henholdsvis vidensniveauet påvirker produktionen. Selve vækstmotoren – altså humankapital eller teknologisk knowhow – produceres imidlertid i en særskilt sektor, hvor resursen IKKE indgår. I dette Cobb-Douglas-tilfælde kan den langsigtsvækstrate, som markedsøkonomien frembringer, både være positiv og negativ. Hvorvidt den bliver det ene eller det andet, afhænger af såvel produktionsteknologien som aktørernes præferencer – hvor utålmodige de er – og også af, i hvilket omfang økonomien formår at korrigere de forskellige markedsfejl, der typisk er til stede, eksempelvis i forsknings- og uddannelsessektoren. Den sidste pointe er nok den vigtigste indsigt, som denne modelgeneration har ført med sig: I modsætning til den tidligere litteratur om vækst og naturressurser fra 1970'erne viser f.eks. Schou (2002b), at selv i et parameteriseret tilfælde, hvor positiv langsigtsvækst er både mulig og optimal, kan væksten i en ufuldkomment reguleret markedsøkonomi godt blive negativ på lang sigt alligevel.

At enten uddannelse eller forskningsideer produceres uden anvendelse af ikkefornybare naturressurser virker imidlertid ikke som nogen realistisk antagelse. Forskere i laboratorier eller universitetskontorer bruger eksempelvis computere og andre redskaber, der typisk er fremstillet af metaller og olieprodukter (plastic), og de bruger energi, som i høj grad kommer fra fossile brændstoffer, til opvarmning, transport og elektronisk udstyr osv. Og de bruger resurserne indirekte i form af anvendelsen af mange former for fysisk kapital, som i sig selv er fremstillet ved hjælp af indsatsen fra mineraler og fossilt brændsel. Det forekommer derfor mere realistisk, også at lade forskningssektoren være afhængig – direkte eller indirekte – af den ikkefornybare resurse. Dette gøres i Groth og Schou (2002) i en Cobb-Douglas-ensektormodel, hvor der kun er en

form for reproducerbar kapital. Såvel forbrugs- som investeringsgoder (inklusive investeringer i ny viden, som også indgår i det udvidede kapitalbegreb i denne model) kræver dermed et vist resurseinput for at kunne fremstilles overhovedet. Det viser sig her, at ud over substitutionselasticiteten på én kræver en stabil positiv langsigtsvækst, at der er dels positiv befolkningsvækst, dels en passende grad af stigende skalaafkast mht. den reproducerbare kapital og arbejdskraften under ét.

Dette resultat afviger på flere måder fra de forudsigelser, der sædvanligvis kommer fra den moderne endogene vækstteori – inklusive den ovenfor citerede afdeling, hvor naturresser indgår i produktionen af forbrugsvarer, men ikke i selve den langsigtede vækstmotor. En implikation er, at en positiv befolkningsvækst på lang sigt nu er en nødvendig betingelse for stabil langsigtsvækst.

En anden implikation er, at (asymptotisk) konstant skalaafkast mht. den reproducerbare kapital nu ikke længere er den afgørende betingelse for stabil positiv vækst. Det er ellers den »knivsæg«-betingelse, som ligger implicit i alle standard-endogene vækstmodeller uden naturresser. Tager man hensyn til essentielle ikkefornybare resurser, bliver nu en eller anden grad af stigende skalaafkast – til gengæld med hensyn til kapital OG arbejdskraft – i stedet det afgørende.

Man kan diskutere, om denne nye betingelse er stærkere eller svagere end den gamle, men inddragelsen af naturresser får altså flere af de problemer, som den endogene vækstteori i almindelighed har sloges med, til at forsvinde. Det gælder dels dette knivsægproblem, dels skalaproblemet, som går ud på, at befolkningsvækst er uforenelig med en steady state i en række forskningsbaserede vækstmodeller. Hermed er resultatet også interessant i diskussionen om endogene over for semiendogene vækstmodeller, hvor sidstnævnte er kendetegnet ved, at den teknologiske vækst er endogen, men afhængig af stadig vækst i en anden eksogen variabel, typisk befolkningen. I den glimrende oversigt i Groth (2007) diskuteres disse problemstillinger nærmere.

5. Fornybare naturresser og forurening

Lad os dernæst forlade de ikkefornybare resurser og se på det andet hovedspørgsmål: Under hvilke omstændigheder vil forurening forårsaget af produktionen på forskellig vis føre til ophør af væksten, måske endda en nødvendig nedgang i produktionen, på lang sigt?

De grundlæggende sammenhænge kan stadig i dag præsenteres ved hjælp af den tre-ligningsmodel for hele verdensøkonomien inklusive vekselvirkningen med det omgivende miljø, som er opstillet i Schou (2000a), og som igen er en modificeret version af Smulders (1995). De tre ligninger beskriver henholdsvis systemets økologiske dimension, den aggregerede økonomiske aktivitet (produktionen) og efterspørgselsiden (præferencerne). Variablen N er beholdningen af naturkapital, som er en aggre-

geret variabel, der i fysiske enheder måler mængden af naturressurser eller miljøets overordnede kvalitet. Tilvæksten i N pr. tidsenhed beskrives af ligningen

$$\dot{N} = E(N, S) - R, \quad \text{hvor } \dot{N} = \frac{dN}{dt}, E_S > 0. \quad (3)$$

R repræsenterer den forurening, de forskellige produktionsprocesser medfører som biprodukt, og som fører til en forringelse af det omgivende miljø. Det dækker f.eks. udledningen af drivhusgasser eller skadelige partikler i atmosfæren. Forurening nedbryder naturkapitalen, men funktionen E angiver, at denne naturkapital omvendt kan regenereres. Hastigheden, hvormed dette sker, afhænger dels af N selv og dels af tilførslen af den eksogene (og konstante) solenergi, S .¹ Når denne variabel indgår eksplicit, skyldes det, at den spiller en vigtig rolle i diskussionen om, hvordan termodynamikken påvirker mulighederne for langsigtet vækst, jf. Schou (2000a). E_S angiver den partielle afledede af funktionen med hensyn til S . Formuleringen i (3) svarer til den måde, som en fornybar resurse som f.eks. en fiskebestand normalt behandles i resurseøkonomiske lærebøger og kan altså også gives en alternativ fortolkning, hvor R opfattes som den strøm af *fornybare* naturressurser (tømmer, fisk, afgrøder mv.), som på ethvert tidspunkt anvendes i økonomien.

Produktionen foregår i en enkelt sektor, hvor output dels bliver anvendt som forbrug C , dels investeres i opbygningen af større viden V . Den afgørende menneskeskabte kapitalbeholdning er altså viden. Ændringen heri over tid kan angives ved følgende akkumulationsligning:

$$\dot{V} = Y(N, R, V) - C, \quad Y_N \geq 0, Y_R > 0, Y_V > 0. \quad (4)$$

Produktionen Y afhænger dels af mængden af forurening R , dels af den eksisterende viden på det pågældende tidspunkt. Desuden indgår også N som et ikke-rivaliserende element i produktionsfunktionen. Det afspejler de forskellige produktive serviceydelser, som naturen stiller til rådighed som et offentligt gode: Eksistensen af ozonlaget, der beskytter mod skadelige virkninger af solens stråling, biodiversitet, hvis genetiske mangfoldighed er en kilde til produktion af ny viden, klimatiske betingelser for at opretholde en sund og produktiv arbejdsstyrke osv.

Samfundets velfærdsfunktion antages at være et integral over de tilbagediskonterede nyttebidrag pr. tidsenhed, hvor ρ er den konstante tidspræferencerate:

$$W = \int U(C, N) e^{-\rho t} dt, \quad U_C > 0, U_N \geq 0. \quad (5)$$

1. Funktionen E antages normalt at være konkav i N , dvs. den har form som et omvendt U .

De tre ligninger udgør tilsammen en sammenhængende økonomisk vækstmodel, der er i stand til at udtrykke de vigtigste vekselvirkninger mellem den økonomiske aktivitet og den omgivende natur. Naturkapitalen N er begrænset opadtil, eftersom den måles i fysiske enheder, og de termodynamiske love og solens indstrømning sætter en øvre grænse for den mængde energi, der kan ophobes på jorden. I en langsigtslige vægt må den være konstant, og ifølge ligning (3) medfører det, at også R så må være konstant. Vi kan altså ikke permanent forøge vores forurening. Dette er én grundlæggende betingelse for bæredygtig vækst. En anden betingelse er, at den partielle substitutionselasticitet mellem viden og hver af de andre input i produktionsfunktionen asymptotisk skal være mindst én. Ellers er det nemlig ikke muligt for produktionen at vokse ubegrænset samtidig med, at N og R har et konstant niveau, da N og R 's grænseprodukt så vil være begrænset opadtil, uanset hvor stor V bliver. Samtidig kan substitutionselasticiteten ikke asymptotisk være større end én, da fysiske input er uundværlige i produktionen. Dermed får man det resultat, at en substitutionselasticitet på netop én er nødvendig for, at verdensøkonomien kan opleve vedvarende vækst på helt lang sigt. Parallellen til de grundlæggende resultater fra modellerne med ikkefornybare resurser er åbenlys.

Som tidligere nævnt er det desuden et centralt resultat fra den moderne vækstteori, at *balanceret* vækst, altså en situation, hvor produktion og forbrug i det lange løb vokser med en konstant vækstrate, kræver, at der er konstant skalaafkast i produktionsfunktionen Y med hensyn til den menneskeskabte produktionsfaktor viden (det såkaldte knivsægproblem, som blev nævnt ovenfor). Selvom permanent bæredygtig økonomisk vækst er mulig uden at nedbryde naturkapitalen, er den ikke nødvendigvis optimal – det vil afhænge af præferencerne. Hvis forbruget vokser over tid, mens miljøkvaliteten forbliver konstant, vil grænsenyttens af de menneskefremstillede goder C falde i forhold til grænsenyttens af de serviceydelser, der kommer fra miljøet N . Det medfører to modsatrettede virkninger: En positiv indkomsteffekt (fordi vi i et vækstsamfund bliver rigere med tiden) forøger efterspørgslen efter serviceydelser fra miljøet, mens en substitutionseffekt omdirigerer efterspørgsel efter et godt miljø hen imod efterspørgsel efter menneskeskabte goder. Den stigende relative knaphed på miljøgoder vil nemlig forøge prisen på disse, målt i forbrugsvarer. Hvis den sidstnævnte effekt dominerer, vil naturkapitalen konstant blive mindre over tid, hvilket ikke kan være en langsigtslige vægt. Hvis den førstnævnte effekt dominerer, vil investeringer i miljøet stige indtil det punkt, hvor de fuldstændig fortrænger investeringer i produktionen af forbrugsvarer, sådan at den økonomiske vækst vil standse. Kun hvis de to effekter fuldstændig udligner hinanden, kan vedvarende økonomisk vækst sammen med et konstant niveau for miljøkvaliteten være optimal. Det vil ske, når de to goder hverken er substitutter eller komplementær for hinanden, eller sagt med andre ord: Der skal

være en substitutionselasticitet på nøjagtig én mellem C og N . Også den samfundsmæssige velfærdsfunktion skal altså have et Cobb-Douglas-lignende udseende.

Ovenstående formulering, hvor forurening er et input i produktionsfunktionen, kan virke unintuitiv for mange, der betragter forurening som et output (et biprodukt ved fremstillingen af forbrugs- og investeringsgoder). I Smulders (1995) stilles derfor en alternativ formulering op, som anvender denne fremstilling. Dette kan illustreres ved at erstatte (3) og (4) ovenfor med følgende ligninger:

$$\dot{B} = R - \delta(B, S). \quad (6)$$

$$\dot{V} = Y(B, V) - C - A, \quad Y_B \leq 0, Y_V > 0. \quad (7)$$

$$R = P(Y, A), \quad P_Y > 0, P_A > 0. \quad (8)$$

B er den samlede forurening, som er en beholdningsvariabel. R er udledningen på hvert enkelt tidspunkt, som stiger med produktionens størrelse (ligning 8), men kan afhjælpes ved at anvende midler til forureningsbekæmpelse/genopretning af miljøet (abatement, variabelen A). Ligesom funktionen E i ligning 3 var konkav, antages også funktionen δ at være det. Termodynamikken kræver, at B ikke vokser ud over alle grænser. I en langsigtligvægt må den være konstant, hvilket kræver, at også forureningen R er konstant. Det kræver, at forureningsbekæmpelsen A er tilstrækkelig effektiv, idet positiv langsigtsvækst i forbruget så vil kræve, at man kan holde forureningen i skak ved at bruge en konstant andel af produktionen ($A/Y < 1$) herpå. Er denne betingelse ikke opfyldt, vil forbrugsvæksten på lang sigt dø ud.

Denne indsigt har ikke ændret sig siden Smulders' artikel udkom. Den såkaldte børnehavemodell, som er den centrale model i oversigtsartiklen i Brock og Taylor (2005), kan således betragtes som en variant af ovenstående model (5)-(8). Brock og Taylor får dermed også som en af deres vigtigste konklusioner, at teknologiske fremskridt mht. forureningsbekæmpelse er en nøgleforudsætning for at få bæredygtig vækst i det lange løb.

6. Empiri om fornybare resurser/forurening og vækst

Det foregående afsnit viser, at også når spørgsmålet drejer sig om forureningens betydning for den økonomiske vækst, er størrelsen af substitutionselasticiteten mellem natur- og menneskeskabt kapital afgørende, på samme måde som tilfældet var med hensyn til ikkefornybare naturressurser. I modsætning til de resultater, der er omtalt i afsnit 4, foreligger der så vidt vides ikke nogen empiriske forsøg på at måle substitutionselasticiteten mellem natur- og videnskapital i en sammenhæng, som ligning 4 ud-

trykker, dvs. mellem naturen i bred forstand og menneskeskabt kapital. Den underliggende model er da også et noget abstrakt koncept, som mere tjener til at klare tanken end til at danne udgangspunkt for egentlige empiriske beregninger. Det er således en noget arbitrær opgave at skulle kvantificere begrebet »videnskapital«.

Til gengæld er der en anden problemstilling, som har tiltrukket sig meget stor opmærksomhed i den professionelle debat om forholdet mellem miljø og økonomisk vækst og også været genstand for mange empiriske afprøvninger, nemlig begrebet miljø-Kuznetskurven. Det er en hypotese om, at forholdet mellem forurening og økonomisk velstand kan udtrykkes som et omvendt *U*: Forureningen er relativt lille for såvel høje som lave velstandsniveauer, men stor for en økonomi på et mellemindkomst-stadie. Der findes en stor såvel teoretisk som empirisk litteratur om miljø-Kuznetskurven, der er refereret i en nylig oversigtsartikel, Kijima mfl. (2010). Her omtales således adskillige empiriske undersøgelser, der har påvist en sådan omvendt *U*-formet sammenhæng for en række – men ikke alle – forureningsproblemer og lande. Samtidig har en række teoretiske bidrag dokumenteret forskellige mulige begrundelser for, hvordan den kan opstå. En intuitivt enkel forklaring er, at forurening i sig selv vil have en tendens til at stige med produktionen, kombineret med, at ønsket om et godt miljø er et luksusgode. I et fattigt land, hvor der kun er en beskedent materiel produktion, er der så heller ikke nogen synderlig forurening. Efterhånden som landet bliver rigere, begynder det også at opleve større forureningsproblemer. Men efterhånden som disse bliver identificeret af befolkningen, fører det til større forureningsbekæmpende indgreb, og – forudsat at de teknologiske muligheder er til det – vil det føre til en tredje fase, hvor forureningen igen mindskes. En supplerende begrundelse er, at den økonomiske udvikling ofte fører til sektorforskydninger fra traditionelt drevet landbrug over industri til serviceerhverv, hvor den mellemliggende fase med en dominerende industriproduktion ofte opfattes som den mest forurenende. Ligeledes kan analysen udvides med udviklingen af teknologiske fremskridt inden for forureningsbekæmpelsen, som kan fremkomme som et resultat af den generelle teknologidrevne vækst og dermed forklare den gunstige miljøudvikling i miljø-Kuznetskurvens sidste fase.

Det er således ganske enkelt at forklare, hvordan en sådan omvendt *U*-formet sammenhæng kan tænkes at eksistere. I lyset heraf virker det faktisk en smule påfaldende, at spørgsmålet har tiltrukket sig så stor opmærksomhed som det har, såvel i den akademiske litteratur som ind imellem i den bredere offentlige debat. I den forbindelse bør man huske på, at der ikke i den foreliggende litteratur er nogle begrundelser for, at der skulle være tale om nogen lovmæssighed, som nødvendigt eller naturligt gælder for *alle* miljøproblemer. Hverken de teknologiske muligheder, betalingsviljen eller øvrige relevante faktorer behøver at være de samme i forskellige situationer med forskellige

typer af forurening. De hidtil fundne empiriske sammenhænge kan bruges til at påvise, at der ikke *nødvendigvis* behøver at være nogen negativ sammenhæng mellem økonomisk udvikling og et godt miljø på specifikke områder. Hverken empirien eller den ledsagende teoretiske litteratur kan imidlertid bruges til omvendt at postulere, at der *aldrig* vil være miljøomkostninger ved økonomisk vækst på længere sigt. Heller ikke litteraturen om miljø-Kuznetskurven kommer således afgørende nærmere på et egentligt definitivt svar på det grundlæggende spørgsmål om forholdet mellem miljø og økonomisk vækst.

Den ganske omfangsrige empiriske litteratur om miljø-Kuznetskurven omhandler udviklingstræk for en økonomi, der oplagt ikke er i steady state. En anden rolle, som økonometrikerne ofte spiller i den generelle økonomiske vækstlitteratur, er at afprøve forskellige steady state-antagelser, som kan udledes fra alternative teoretiske modeller for på denne måde at teste, hvilke af de forskellige teoretiske bud der bedst passer på virkeligheden. Som Brock og Taylor (2005) gør opmærksom på, glimrer sådanne empiriske undersøgelser af forskellige steady state-antagelsers validitet imidlertid ved deres fravær i litteraturen om vækst og miljøproblemer. Her er imidlertid også den særlige problemstilling, at det ikke er spor oplagt, at de vestlige lande rent faktisk i øjeblikket befinder sig i nærheden af en steady state-vækstbane i de modeller, som er omtalt i det foregående. Hvad vækstmodellerne, der bygger på inddragelse af ikkefornybare resurser angår, er det tværtimod oplagt, at vi IKKE befinder os tæt på en steady state. I en sådan vil resurseforbruget nemlig aftage eksponentielt hvert år, og det er ikke det billede, vi ser, hvad enten vi ser på forbruget af fossile brændsler eller de fleste vigtige mineraler, som anvendes i produktionen. De stylized facts, som normalt i vækstteori anvendes som argument for, at i hvert fald de mest udviklede lande i verden er tæt på en situation med balanceret vækst, holder altså oplagt ikke, når man introducerer ikkefornybare resurser i billedet. Hvad angår den anden gruppe af modeller, som omtales her, som omhandler forbruget af fornybare resurser/forurening, er billedet lidt mindre klart, men hvis man ser på et konkret forureningsspørgsmål som klimaproblemet (jf. næste afsnit), er det også klart, at de historiske udledninger af drivhusgasser ikke svarer til en steady state, der kan fortsætte uændret fremover. Hvad enten man vil lave empirisk arbejde for at teste de teoretiske modellens implikationer, eller man ønsker at bruge de teoretiske resultater direkte, må man derfor først af alt gøre sig klart, at verden netop ikke har været i en steady state i den periode, vi normalt har foreliggende data for.

7. Klimaproblemer

Det miljøproblem, som der er allermest fokus på i disse år, er uden tvivl de forventede klimaændringer, som følger af udledningen af drivhusgasser, først og fremmest

via afbrændingen af fossile brændstoffer. I den kontekst, som er emnet for nærværende artikel, er klimaproblemet først og fremmest bemærkelsesværdigt, fordi det forener de to ovennævnte hovedproblemstillinger: Det fremkommer i høj grad ved, at udnyttelsen af en ikkefornybar naturresurse (fossile brændstoffer) i sig selv skaber en forureningsmæssig eksternalitet, som kan påvirke såvel vores globale produktionsfunktion som den globale velfærd direkte, og potentielt i betydeligt omfang. En række bidrag har gennem de seneste årtier påvist, at den omstændighed, at forureningskilden er en ikkefornybar resurse, har en række vigtige implikationer for, hvordan forureningsproblemet optimalt kan tackles. Allerede Suzuki (1976) demonstrerer således (i en model med eksogene produktivitetsfremskridt) betydningen af, at det samlede produktivetsniveau kan afhænge negativt af udtømningen af en ikkefornybar resurse. Hoel og Kverndokk (1996) opsummerer den tidligere litteratur om drivhusproblemerne i en vækstsammenhæng, når man tager hensyn til forureningskildens ikkefornybare karakter. Et hovedresultat, som også Hoel og Kverndokk demonstrerer, er, at under forskellige antagelser skal den optimale skattesats på naturressourcen være *faldende* over tid, i hvert fald i et vist tidsrum. I traditionel resurseøkonomi er det velkendt, at en konstant skattesats på f.eks. værdien af olieudvinding ikke vil påvirke udvindingsprofilen. En skattesats, som formodes at blive lavere over tid, vil derimod få resurseejerne til at udskyde deres udvinding for at få gavn af den lavere beskatning – hvis vel at mærke skattesænkningen virker troværdig. Og en mindre udvinding er netop, hvad der kan være behov for, hvis udvindingen og forbruget af ressourcen har en negativ ekstern effekt i form af en drivhuseffekt, hvilket forklarer det umiddelbart overraskende resultat.

Schou (2002) demonstrerer i en endogen vækstmodel et specialtilfælde, hvor den omstændighed, at et forureningsproblem skyldes udtømningen af en ikkefornybar resurse, gør, at miljøpolitik bliver overflødig i den forstand, at en traditionel Pigouskat på forureningskilden ikke ændrer på allokeringen overhovedet i økonomien. En medvirkende årsag til dette resultat er, at det pågældende forureningsproblem er modelleret som en strøm, dvs. der er ikke nogen tidsmæssig afstand mellem udledningen og de problemer, den forårsager. Groth og Schou (2007) påviser omvendt, at det tidligere nævnte resultat om en optimalt faldende skattesats over tid også holder i en endogen vækstmodel, hvor den pågældende naturresurse indgår i økonomiens vækstmotor, og hvor der er andre mulige skatte- og subsidieinstrumenter til rådighed. Sinn (2008) gør i en policy-orienteret artikel opmærksom på den samme grundlæggende pointe: Skal tiltag dæmpe udledningerne af drivhusgas, skal de påvirke Hotelling-reglen som f.eks. en skattesats, der falder over tid, gør. De eksisterende klimaplaner, som ofte bygger på det modsatte: en gradvis »skrappere« klimapolitik over tid, risikerer derfor at opnå det modsatte af deres mål. En sådan tidsprofil kan nemlig få profitmaksimerende olieudbydere til at sætte olie- og kulproduktionen *op* og dermed fremskynde klimaforandrin-

gerne. Når visse både teoretiske og numeriske modeller får det modsatte resultat, skyldes det typisk, at de ikke tager hensyn til de særlige problemstillinger, som forureningskildens ikkefornybare karakter medfører. Denne problemstilling er blevet døbt »det grønne paradoks«. van der Ploeg og Withagen (2010) nuancerer denne problemstilling, dog uden at inddrage nogen egentlig makroøkonomisk vækstmodel. De gør således opmærksom på, at de forskellige resultater i debatten om det grønne paradoks også afhænger kritisk af, hvilke antagelser man gør sig om omkostningerne ved at udnytte en bagstopper-resurse, altså energikilder som atomkraft eller vedvarende energi, som alternativ til de fossile brændstoffer.

Klimaspørgsmålet har i øvrigt ført til en betydelig interesse for at opbygge anvendte generelle ligevægtsmodeller, der er kvantificerede computerudgaver af de teoretiske vækstmodeller, der beskæftiger sig med afvejningen mellem forbrug og forureningsproblemer. Disse modeller sætter dermed mere konkrete størrelsesordener på en række af de her nævnte problemstillinger. Foregangsmodellen er her William Nordhaus' DICE-model. Den er opbygget som en relativt enkel Ramsey-vækstmodel for en global økonomi med eksogent bestemte teknologiske fremskridt, jf. Nordhaus (2008). Ud over en forbrugs-/opsparingsbeslutning er den centrale problemstilling i DICE, hvor mange resurser der skal afsættes til at formindske drivhusproblemet på hvert enkelt tidspunkt. Produktionen mindskes isoleret set ved en ambitiøs klimapolitik, men samtidig påvirker en sådan endogent i modellen den globale middeltemperatur, og det kan give en afledt effekt i form af forbedrede produktionsmuligheder på længere sigt. DICE-modellen integrerer således nogle af de formodede naturvidenskabelige sammenhænge med samfundsøkonomien. Der er senere udviklet mange andre tilsvarende modeller, som giver et bud på de økonomiske konsekvenser af forskellige former for klimapolitik. Det ligger uden for denne artikel at beskrive disse nærmere, men den tidligere nævnte artikel af van der Werf (2008) refererer til en række af dem.

8. Befolkningsvækst

Befolkningsvækstens betydning for et eventuelt fremtidigt økologisk katastrofescenarie er et tema, som spiller en selvstændig rolle i den offentlige debat. Sammenhængen mellem befolkningsudvikling og vækst i materiel velstand pr. capita har optaget økonomer i århundreder, længe før miljøspørgsmål for alvor kom på den fagøkonomiske dagsorden. Ældre vækstteori (f.eks. Solowmodellen) har typisk peget på, at der er en negativ sammenhæng mellem befolkningsvækst og materiel pr. capita-velstand på grund af den udtynding af kapitalintensiteten, som opvoksende generationer giver anledning til. Moderne endogen vækstteori har derimod understreget en modsatrettet effekt: Kilden til vores velstand er opfindelser og ny viden, og ny viden skabes af mennesker. Jo større befolkning og dermed arbejdsstyrke og – alt andet lige

– flere forskere, jo flere ideer kan man derfor alt andet lige forestille sig, at der vil blive skabt. Da ideer er et ikke-rivaliserende gode, bliver denne form for kapital samtidig ikke udtyndet af en større befolkning, og man kan dermed få en positiv årsagsvirkning fra befolkningsvækst til økonomisk pr. capita-vækst, jf. eksempelvis Jones (1997). Det er årsagen til, at befolkningen somme tider udråbes til »vores ultimative resurse«. Selve denne tanke er dog ikke ny. I Danmark har således økonomiprofessor Mogens Boserup tidligere fremført samme synspunkt, jf. eksempelvis Boserup (1974).

I endogene vækstmodeller, hvor økologiske dimensioner inddrages, bevares disse modsatrettede muligheder typisk. Et eksempel, hvor højere befolkningsvækst påvirker pr. capita-vækstraten positivt, ses i Groth og Schou (2002) – og faktisk er der som tidligere nævnt i denne model kun mulighed for stabil positiv pr. capita-vækst i det lange løb, hvis befolkningen bliver ved med at vokse! En implikation heraf er således også, at stabil positiv vækst ikke vil være mulig på det helt lange sigt, hvis man tror, at der er en øvre grænse for, hvor stor verdens befolkning kan blive.

Denne og andre af de ovennævnte modeller tager imidlertid ikke hensyn til, at en større befolkning kunne have en direkte negativ effekt på miljøtilstanden, og heller ikke til, at der kunne være en årsagssammenhæng den anden vej – altså at ændringer i miljøet eller i den materielle velstand i sig selv kunne tænkes at påvirke befolkningsvæksten. Denne er nemlig eksogen i de nævnte modeller. For at kunne studere sådanne interaktioner kræves modeller, hvor selve befolkningsvæksten også er endogen. Det komplicerer dynamikken betragteligt, og den økonomiske teori er da også stadig ret mangelfuld på det felt, der hedder at undersøge den gensidige sammenhæng mellem økonomisk vækst, befolkningsvækst og miljøspørgsmål. Et forsøg herpå gøres i Schou (2002a), der demonstrerer, hvordan traditionelle konklusioner om optimal miljøpolitik modificeres, når man antager, at husholdningerne endogen vælger, hvor mange børn der fødes i hver generation. Heller ikke i denne model er det dog befolkningsvæksten direkte, som påvirker forureningen. Derimod afhænger miljøpåvirkningen af den samlede produktions størrelse og evt. af mulighederne for at afsætte ressourcer til forureningsbekæmpelse og miljøgenopretning. Afhængigt af, hvor effektive disse sidste muligheder er, kan højere befolkningsvækst i denne model således resultere i henholdsvis et forværret eller et forbedret miljø. Hvis der er stigende skalaafkast i produktionen af bedre miljøforhold, vil en større befolkning og dermed tilsvarende større produktion i optimum føre til miljøforbedringer, idet den øgede resursetilførsel til miljøgenopretningen mere end opvejer det forureningspres, som den større befolkning/-produktion umiddelbart medfører. I sådan en situation er det optimalt at give et subsidium til børnefødsler, fordi der ellers vil blive født for få børn. Omvendt hvis der er faldende skalaafkast.

Et andet bidrag, hvor befolkningsvæksten formelt er endogen, er Bretschger (2008).

Her er der dog ikke noget egentligt mikrofundament for befolkningsudviklingen, som antages at afhænge direkte af produktivitets- og reallønsvækst. Modellen og dens resultater er derfor meget lig eksempelvis modellerne i afsnit 5.2 i Groth (2007), der undersøger den samme situation, hvor forskning produceres med arbejdskraft og ikke-fornybare resurser som nødvendige produktionsfaktorer, men hvor befolkningsvæksten er eksogen.

9. Policy

Generelt er miljøproblemer ofte forårsaget af markedsfejl, og anbefalingerne til, hvad man fra politisk side bør gøre, er derfor overordnet ligetil: Der bør korrigeres for markedsfejlene på passende vis i overensstemmelse med traditionel mikroøkonomisk tankegang. Diskussionerne i de foregående afsnit om, hvorvidt langsigtet økonomisk vækst er mulig og ønskelig eller ej, påvirker ikke direkte dette generelle princip. Uanset om man er optimistisk eller ej med hensyn til de langsigtede vækstmuligheder, skal forureningsproblemerne stadig internaliseres i beslutningsprocesserne i en markedsøkonomi. I denne forstand har litteraturen om endogen vækst og miljøspørgsmål ikke noget grundlæggende nyt at byde på, hvad spørgsmålet om optimal miljøpolitik angår; det er dog generelt af interesse at få at vide, om en skrapere miljøpolitik i en given situation vil påvirke den økonomiske vækst på længere sigt positivt eller negativt. Dette er en variant af det mere generelle spørgsmål, om der kan være en dobbelt dividende ved en grøn afgiftsomlægning.

Det intuitive udgangspunkt er, at en større indsats for miljøet, hvis man som udgangspunkt er i en suboptimal situation med for store forureningsproblemer, nok vil forbedre den samlede velfærd, men det vil virke hæmmende for den økonomiske aktivitet. I en endogen vækstmodel kan dette virke hæmmende for vækstraten. Dette kan ske, hvis den skrapere miljøpolitik fører til, at der allokeres færre resurser til økonomiens vækstmotor. Man har dog identificeret forskellige situationer, hvor der kan være en positiv sammenhæng mellem skrapere miljøpolitik og en højere vækstrate i den materielle produktion. Det kan være tilfældet, hvis miljøet som ikke-ivaliserende faktor påvirker produktiviteten positivt (variablen N 's rolle i ligning 4), hvis der i forvejen i økonomien er andre markedsfejl end de miljørelaterede, OG miljøpolitikken også samtidig vil mindske disse markedsfejl, hvis befolkningens nytte afhænger af andre faktorer end blot materielle forbrugsgoder og et godt miljø, eller endelig – i en mere speciel situation – hvis forureningskilden er ikkefornybare naturressurser, jf. den uddybende gennemgang i Schou (2000a). Også Fullerton og Kim (2008) viser i en artikel, der også gennemgår en række af resultaterne i den tidligere litteratur, at skrapere miljøpolitik i form af højere forureningsafgifter både kan virke vækstoffremmende og -hæmmende afhængigt af de nærmere omstændigheder. Brock og Taylor (2005) de-

monstrerer med nogle simple regneeksempler, hvordan man kan forsøge at kvantificere den hæmsko (drag), som en optimal miljøpolitik kan være for den økonomiske vækst i forbrugsmulighederne fremadrettet. Resultaterne under de antagelser, som forfatterne oplever som rimelige, peger på et fradrag på mellem 0,06 og $1/2$ pct.point i den årlige amerikanske BNP pr. capita-vækst, som i sig selv har ligget på et niveau på ca. 2 pct. om året.

Hvad sammenhængen mellem langsigtet økonomisk vækst og beskatningen af ikkefornybare naturressurser angår, er det et emne, som i øvrigt ikke er særlig belyst i den teoretiske litteratur, jf. oversigtsartiklen i Lund (2009), der efterlyser flere undersøgelser af forskellige former for resursebeskatning i intertemporal ligevægt. En tidlig undtagelse her er Sinn (1982), som peger på, at en værdiskat på den udvundne resurse vil være mere hensigtsmæssig end en kapitalgevinstskat, som ellers anbefales i forskellige partielle resursemodeller. Groth og Schou (2007) undersøger i en Cobb-Douglas-model forskellige kapital- og resurseskatters indflydelse på steady state-vækstraterne i en model for en lukket økonomi (dvs. hele verden), hvor naturressourcen påvirker økonomiens vækstmotor. I modsætning til, hvad der er et almindeligt resultat i endogene vækstmodeller uden naturressurser, påvirker traditionelle kapitalindkomstskatter som en skat på renteindkomst ikke vækstraten i det lange løb. Det samme gælder for et subsidium til kapitalakkumuleringen. Derimod vil en kapitalgevinstskat på resursebeholdninger (som under visse betingelser kan oversættes til en skat på den rene profit hos resurseudvindende selskaber) påvirke den langsigtede vækst negativt, fordi den giver en tilskyndelse til en inoptimalt hurtig udtømmning af ressourcen. En skat på den løbende resurseudvinding, som falder over tid (altså skattesatsen), vil omvendt medføre en højere vækstrate i det lange løb, fordi den vil udskyde den optimale udvinding. Da udtømmningen i denne model i en ren markedsøkonomi vil ske inoptimalt hurtigt, vil en sådan skat være velfærdsforbedrende, jf. diskussionen ovenfor i klimafsnittet.

10. Konklusion

Når talen falder på sammenhængen mellem langsigtet økonomisk vækst og miljøproblemer, tænkes der typisk på én af to forskellige problemstillinger: For det første spørgsmålet, om væksten – eller bare et konstant forbrugsniveau – kan opretholdes i en verden, hvor forbruget af ikkefornybare naturressurser som mineraler og fossile brændstoffer nødvendigvis må falde på lang sigt. For det andet spørgsmålet, om forureningsproblemer, som typisk opstår som et biprodukt af vores produktion, i det lange løb vil medføre, at fortsat økonomisk vækst enten ikke er mulig eller i hvert fald forbundet med så store velfærdsmæssige omkostninger, at det vil være uønskeligt at opretholde den. Denne artikel har givet et overblik over, hvad den økonomiske vækst-

teori p.t. har at sige om disse spørgsmål. For hver af de to problemstillinger findes der en tradition for vækstmodeller med henholdsvis ikkefornybare og fornybare resurser (herunder naturens muligheder for at absorbere forurening) som et vigtigt element i makroproduktionsfunktionen.

I begge tilfælde er det en afgørende betingelse for vedvarende økonomisk vækst, at substitutionsmulighederne mellem naturressourcen (af den ene eller anden slags) på den ene side og menneskeskabt kapital (i sidste ende i form af viden) på den anden side er tilstrækkelig gode. Sagt mere teknisk skal substitutionselasticiteten mellem de to produktionsfaktorer være mindst én asymptotisk, efterhånden som tidshorizonten går mod uendeligheden. Samtidig eksisterer der også et argument for, at substitutionselasticiteten mellem den fornybare naturkapital (som ud fra termodynamiske begreber i sidste ende kan betragtes som den energi, vi får fra solen) ikke kan være større end én, idet produktion ikke kan finde sted fuldstændig uden forbrug af fysiske resurser af nogen art. Den grundlæggende makroproduktionsfunktion må altså have Cobb-Douglas-lignende egenskaber, for at fortsat vækst er mulig, idet Cobb-Douglas-produktionsfunktionen jo er kendetegnet ved at have en substitutionselasticitet på netop én mellem sine argumenter. Er substitutionselasticiteten mellem vores ikkefornybare resurser og menneskeskabt kapital i det lange løb mindre end én, vil det være umuligt at opretholde et hvilket som helst positivt produktionsniveau permanent: Vores forbrugsmuligheder må nødvendigvis gå mod nul før eller siden. Er substitutionselasticiteten mellem viden og naturkapital forstået som en fornybar resurse mindre end én, er konsekvenserne ikke tilsvarende dystre, idet det så stadig vil være muligt vedvarende at opretholde et givet forbrugsniveau; blot kan forbruget ikke permanent *øges*.

Ud over en substitutionselasticitet på mindst én er det også en nødvendig betingelse for vedvarende vækst, at der er i hvert fald konstant skalaafkast med hensyn til menneskeskabt kapital i den aggregerede produktionsfunktion. I specialtilfældet, hvor ikkefornybare resurser er et nødvendigt input (direkte eller indirekte) i produktionen af ny kapital, er det nødvendigt ENTEN med stigende skalaafkast mht. kapital ELLER med en grad af stigende skalaafkast mht. kapital og arbejdskraft tilsammen kombineret med positiv befolkningsvækst. Lægger man til grund, at positiv befolkningsvækst ikke er mulig i et uendeligt tidsperspektiv, bortfalder den sidste mulighed.

Selvom vedvarende økonomisk vækst er mulig, er det ikke sikkert, at det også vil være optimalt at opretholde væksten, hvis produktionen har bivirkninger, som påvirker vores velfærd negativt. Det vil afhænge af vores præferencer, og her er igen en substitutionselasticitet på netop én mellem menneskeskabte forbrugsvarer og naturkapital en nødvendig betingelse for, at positiv økonomisk vækst kan være optimal på helt lang sigt.

I lyset af den afgørende betydning, som substitutionselasticiteterne har for muligheden for permanent vækst, er det nærliggende at spørge, hvad empirien så siger om

disse størrelsesordener? Her er der desværre ikke noget klart svar, og det er i sagens natur også nærmest umuligt at få det. Historiske empiriske analyser af substitutionselasticiteten mellem f.eks. fossile brændstoffer og kapital har fundet værdier på både over og under én, men dels er disse resultater usikre af mange forskellige grunde, og dels siger de ikke nødvendigvis noget om den fremtidige asymptotiske substitutionselasticitet, som er den, der er interessant i vores tilfælde. Man må derfor konstatere, at vi grundlæggende ikke ved, hvorvidt de afgørende betingelser for evig økonomisk vækst er til stede, og selvom den økonomiske forskning fortsat gør fremskridt, vil det næppe heller være muligt at få et definitivt svar på spørgsmålet.

Heldigvis kan man sige, at disse betragtninger egentlig heller ikke er så afgørende for, i hvilket omfang man skal gribe regulerende ind på udviklingen, altså påvirke markedsøkonomiens mekanismer ved hjælp af økonomisk politik. Den overordnede politikanbefaling er den sædvanlige: Der bør korrigeres for de markedsfejl, som udviklingen skaber, i overensstemmelse med en traditionel mikroøkonomisk tilgang. Diskussionerne i de foregående afsnit om, hvorvidt langsigtet økonomisk vækst er mulig og ønskelig eller ej, påvirker ikke direkte dette generelle princip. Det er dog ikke en uvæsentlig pointe, at litteraturen demonstrerer, at selvom permanent positiv vækst måtte være både mulig og ønskværdig, er det ikke sikkert, at en markedsøkonomi, som er overladt til sig selv, rent faktisk vil medføre en sådan udvikling med bæredygtig vækst. Såvel samspillet mellem økonomi og miljø som de processer, der fører til udvikling af ny viden, er præget af typiske markedsfejl, og der vil derfor generelt være behov for en økonomisk politik, der aktivt søger at regulere udviklingen. En sådan indgriben kan generelt både føre til en højere eller en lavere vækstrate end den, som en helt ureguleret økonomi af sig selv ville frembringe.

Litteratur

- Aghion, P. og P. Howitt. 1992. A Model of Growth Through Creative Destruction. *Econometrica*, 60, s. 323-51.
- Bretschger, L. 2008. Population growth and natural resource scarcity: long-run development under seemingly unfavourable conditions. CER-ETH *Economics working paper series* No. 08/87.
- Bretschger, L. og S. Smulders. 2003. *Sustainability and substitution of exhaustible natural resources: how resource prices affect long-term R&D-investments*. CER-ETH *Economics working paper series* No. 03/26.
- Brock, W. A. og M. S. Taylor. 2005. Economic Growth and the Environment: A Review of Theory and Empirics. Kapitel 28 i:
- Aghion, P. og S. A. Durlauf: *Handbook of Economic Growth, volume 1B*. Elsevier.
- Boserup, M. 1974. *Vor voksende verden*. Forum, København.
- Dasgupta, P. S. og G. M. Heal. 1979. *Economic Theory and Exhaustible Resources*. Nisbet, Cambridge.
- di Maria, C. og S. Valente. 2008. Hicks Meets Hotelling: The Direction of Technical Change in Capital-Resource Economies. *Environment and Development Economics*, 13, s. 691-717.
- Fullerton, D. og S.-R. Kim. 2008. Environmental investment and policy with distortionary taxes, and endogenous growth. *Journal of Environmental Economics and Management*, 56, s. 141-54.

- Grimaud, A. og L. Rougé. 2003. Non-renewable Resources and Growth with Vertical Innovations: Optimum, Equilibrium and Economic Policies. *Journal of Environmental Economics and Management*, 45, s. 433-53.
- Groth, C. 2007. A New-Growth Perspective on Non-renewable Resources. I: Bretschger, L. og S. Smulders: *Sustainable Resource Use and Economic Dynamics*. Springer, Dordrecht.
- Groth, C. og P. Schou. 2002. Can nonrenewable resources alleviate the knife-edge character of endogenous growth? *Oxford Economic Papers*, 54, s. 386-411.
- Groth, C. og P. Schou. 2007. Growth and Non-renewable Resources: The Different Roles of Capital and Resource Taxes. *Journal of Environmental Economics and Management*, 53, s. 80-98.
- Hoel, M. og S. Kverndokk. 1996. Depletion of fossil fuels and the impact of global warming. *Resource and Energy Economics*, 18, s. 115-36.
- Jones, C. 1997. Population and Ideas: A Theory of Endogenous Growth. National Bureau of Economic Research, Inc, *NBER Working Papers*: 6285, 1997.
- Jones, C. 2005. The Shape of the Production Functions and the Direction of Technical Change. *Quarterly Journal of Economics*, 120 (2), s. 517-49.
- Kijima, M., K. Nishide og A. Ohyama. 2010. Economic Models for the environmental Kuznets curve: A survey. *Journal of Economic Dynamics and Control*, 34, s. 1187-1201.
- Lucas, R.E. 1988. On the mechanics of economic development. *Journal of Monetary Economics*, 22, s. 3-42.
- Lund, D. 2009. Rent Taxation for Nonrenewable Resources. *Annual Review of Resource Economics*, 1, 287-308.
- Markandya, A. og S. Pedroso-Galinato. 2007. How substitutable is natural capital? *Environmental and Resource Economics*, 37, s. 297-312.
- Neumayer, E. 2000. Scarce or Abundant? The economics of natural resource availability. *Journal of Economic Surveys*, 14, s. 307-35.
- Nordhaus, W. 2008. *A Question of Balance: Weighing the Options on Global Warming Policies*. Yale University Press.
- Romer, P. M. 1990. Endogenous technical change. *Journal of Political Economy*, 98, s. 71-102.
- Scholz, C. M. og G. Ziemes. 1999. Exhaustible resources, monopolistic competition, and endogenous growth. *Environmental and Resource Economics*, 13, s. 169-85.
- Schou, P. 2000a. Miljøproblemer og endogen vækst. *Nationaløkonomisk Tidsskrift*, 138, s. 301-13.
- Schou, P. 2000b. Polluting nonrenewable resources and growth. *Environmental and Resource Economics*, 13, s. 211-27.
- Schou, P. 2002a. Pollution Externalities in a Model of Endogenous Fertility and Growth. *International Tax and Public Finance*, 9 (709), s. 725.
- Schou, P. 2002b. When environmental policy is superfluous: Growth and polluting resources. *Scandinavian Journal of Economics*, 104, s. 605-20.
- Sinn, H.-W. 1982. Taxation, Growth and Resource Extraction. *European Economic Review*, 19, s. 357-86.
- Sinn, H.-W. 2008. Public policies against global warming: a supply side approach. *International Tax and Public Finance*, 15, s. 360-94.
- Smulders, S. 1995. Entropy, Environment, and Endogenous Economic Growth. *International Tax and Public Finance*, 2, s. 319-40.
- Stiglitz, J. 1974. Growth with exhaustible natural resources: Efficient and optimal growth paths. *Review of Economic Studies*, 41, Symposium Issue, s. 123-37.
- Suzuki, H. 1976. On the possibility of steadily growing per capita consumption in an economy with a wasting and non-replenishable resource. *Review of Economic Studies*, 43, s. 527-35.
- van der Ploeg, F. og C. Withagen. 2010. Is There Really a Green Paradox?, *CESifo Working Paper* 2963.
- van der Werf, E. 2008. Production functions for climate policy modeling: An empirical analysis. *Energy Economics*, 30, s. 2964-79.

Debat og kommentarer

Den sociale velfærd i Danmark på baggrund af to udgivelser om dansk velfærdshistorie

Bjarne Hastrup

Ældre Sagen, E-mail: bjarne.hastrup@aeldresagen.dk

SUMMARY: Two new publications on the Danish welfare society's history provides a range of new research on both Danish welfare's origin and evolution of the last generation. Before the year 1799 welfare was not enrolled neither in Denmark's birth certificate nor in medieval and Renaissance period. Apart from a short approach to the improvement for the poor economic conditions for people in the bottom of the Danish society are tough until the 1891. From this year initiated major social reformers time, when poor but especially sick, old and industrial accident victims achieve orderly and improved social conditions. These years also founded the Danish welfare design, without a great master plan but with many funding sources for mixture. This model proved to be an advantage for Denmark from 1980 to 2000, where funded pensions strengthened Denmark's alignment with the future demographic developments. This was followed up with major reforms in the 1990s to 2010 in labor, tax and retirement in which the interest of future economic structures played a prominent role in contemporary welfare design. History, present and future, in a comprehensive political synthesis. The incentive structure was thus significantly improved including lower replacement rates for unemployed people. At the same time Denmark could maintain its position as the country in the world that had the most equal income distribution, although the Gini ratio rose during the last 10 years.

1. Indledning

Danmark står som et fyrtårn, sammenlignet med den øvrige vestlige verden, når det drejer sig om velfærd både i udlandets og danskernes egne øjne. Dansk økonomi er samlet set lille, og udgør da også kun 0,5% af den samlede globale økonomi. Når det drejer sig om velfærdshistorie eller Danmarks økonomiske velfærdshistorie synes landet som en stormagt. Danmark skal i hvert fald nu rumme ikke et, men hele to værker om landets velfærdshistorie på i alt otte bind, fordelt med to bind på Gunnar Viby Mo-

gensens »Det danske velfærdssamfunds historie« (GVM10), der dækker tiden fra 1970erne, hvor det tidligere syv binds værk fra Gyldendal sluttede, til i dag; og seks bind på Jørn Henrik Petersen, Klaus Petersen og Niels Finn Christiansen: »Dansk velfærdshistorie« (CPP10), hvor første bind (BDI) »Frem mod socialhjælpsstaten«, der strækker sig over den længste af alle binds perioder nemlig 1536-1898, nu er udkommet. Det samlede værk fra (CPP) forventes færdigt i 2014. Værkerne følger sig naturligt ind i den lange række af fremstillinger af dansk velfærdshistorie. Der er tradition for at trække dansk velfærdshistorie ud af den generelle økonomiske historie som selvstændige værker, og både danske og internationale forskere har leveret betydelige bidrag. Således udkom i 1974-81 syv binds værket om dansk velfærdshistorie på forlaget Gyldendal, og (GVM10) er dette værks forlængelse frem til begyndelsen af dette århundrede. Men også andre har bidraget. Cand. Polit. Henrik Christoffersen udgav bogen: »Dansk velfærdspolitik efter 1945, i 1984 og i 1988 sammen med forfatteren til denne artikel: »Når få har for meget – en historie om de nordiske folk i moderne tid«. To år efter i 1990 kom Peter Baldwin's værk » The politics of social solidarity«, og værkerne (BH11), (BH08) og (AM02), indeholder tillige betydelige elementer af dansk velfærds økonomiske historie.

2. Samlet oversigt

Med udgangspunkt i de to nyeste værker, kan der tegnes en skitse af velfærdens udvikling i Danmark de sidste ca. 500 år. I begyndelsen af 1500 tallet er der betydelige forskelle mellem rige og fattige, den politiske og civile magt er stærk koncentreret, men den danske del af den katolske kirke udøver, ikke mindst gennem tiggermunkenes mellemkomst, en stor indsats over for syge, gamle, invalider, forældreløse børn og fattige.

Reformationen overfører både kirke- og klostergodset til kongens besiddelser og efterlader fattigforsorgen til frivillige bidrag, godkendte tiggere og sognets forgodtbefindende. Der er hele 198 godser, eller 20% af alle danske godser, der overgår fra kirken til kongen. Krongodset vokser fra 10% til 30% af alle danske godser.¹ Der kæmpes med mange midler det ikke godkendte tiggeri, men ikke fattigdommen, en kamp der runger op gennem to århundreder, inden velfærden lyser op i et kort glimt omkring 1799-1803, hvor oplysningstidens reformatorer bryder båndet mellem pligt og ret og etablerer en fattigpolitik, der kortvarigt tager hensyn til de fattiges behov snarere end pligten, deres forudgående bidrag eller betaling. I perioden efter napoleonskrigene frem til de store reformer er dansk socialpolitik præget af tankerne om ikke at bryde bånd og netværk i de lokale samfund, men at støtte individets egen selvopholdsdrift til livets opretholdelse og at støtte fattige gennem frivillige initiativer eller

1. Jfr. Svend Aage Hansen, Adelsvældens grundlag, G.E.C. Gads forlag 1964.

fra 1860erne at udbygge forsikringstanken på markedsvilkår. Det er tanken om opdeling i de værdigt og uværdigt trængende, der dominerer debatten i det konservativt/liberalt prægede samfund. Konsekvensen af denne udvikling blev oprettelse af fattighuse, arbejdsanstalter med straf for arbejdssky fattige. Først omkring 1890 kommer der skred i velfærdstankerne, drevet frem af nødvendigheden, på tværs af de herskende ideologier, men løftet af andre og højere politiske mål end den rene altruisme eller forsøg på at starte en ny socialpolitisk bølge. Selvom der er tale om en revolutionerende ny strategi, der senere bliver døbt den »skattefinansierede universelle model«, hvilket samtiden næppe ænsede, udformes socialpolitikken i praksis som et patchwork af mange finansieringskilder i skøn blanding, der bestemt ikke antyder, at der var en samlet masterplan bag de historiske danske reformer, i modsætning til de tilsvarende Preussiske sociale reformer i 1880erne.

Det er skatteyderne, gårdmændene, middelklassen på landet og i byerne og dens ledere, der sammen med de moderate højrekræfter udarbejder og gennemfører reformerne, dvs. ejere af fast ejendom og virksomhedsejerne, og det bærer reformerne præg af. Endelig er det praktiske og pragmatiske folk, som lovgiver. Reformerne kan langt fra betegnes socialisme, og heller ikke som i Tyskland socialkonservatisme. Der er snarere tale om en mild form for »middelklassesocialisme«, hvor statskonstitutionen opretholdes, middelklassens arbejdere bliver der taget socialt højde for, og middelklassen kan selv få andel i ydelserne, når tiden kommer dertil.² Herved undlader Rigsdagen at bygge nye administrative systemer op – som i Tyskland – men benytter de kendte administrative kanaler gennem kommunerne. Det var middelklassens ledere, der opbyggede strukturen i det sociale system, mens arbejderklassen knap var repræsenteret i Rigsdagen; men det blev Socialdemokratiet, der senere i 1930erne reformerede og forøgede bevillingerne hen mod det høje niveau vi kender i dag. For tredje gang i danmarkshistorien åbnede vinduet sig til sociale forbedringer (første gang i 1799-1803, anden gang i 1890erne), denne gang i 1930ernes økonomiske depression. Der blev ikke afgørende ændret i finansieringskildernes patchwork, i blandingsforholdet mellem offentligt og privat egenfinansiering, men nu løsnede man de stramme fattighjælpvirkninger for de, der modtog hjælpen (ikke valgbar og ingen stemmeret). Retsprincippet blev allerede i 1922 knæsat, adgangen til hjælp blev gjort lettere, og man forenkede lovgivningen i fire hovedlove. Efter en lovpause, på grund af krigsforholdene, besluttede Folketinget i 1956, selvom der næppe var økonomi dertil, at gennemføre den stort anlagte lov om folkepension, som sammen med det økonomiske opsving fra 1957 tegnede det store spring fremad i dansk velfærd 1956-74. Dette nye åbne vindue for forbedret velfærd blev den danske velfærdsstats glansperiode, mens guldalderen strækker sig ind i dette århundrede.

2. Bjarne Hastrup: The social contract between the generations, 2008 kap. 1.

Det var socialdemokratiet og fagbevægelsen, der opbyggede nye velfærdsinstitutioner og en helt ny ideologi for overførselsindkomster, hvor forbindelsen mellem ret og pligt igen definitivt blev brudt, så de sociale behov blev styrende for velfærdsstatens udvikling. Der tegner sig en forbindelse mellem enevælden og den moderne socialpolitik, hvor det i højere grad er behovene, der styrer politikken. I denne periode blev den fulde folkepension indført til alle uanset indtægt, mens ATP som tillægspension omfattede alle på arbejdsmarkedet (bortset fra selvstændige), selvom ordningen var tiltænkt lavindkomstgrupper. Ydelserne til arbejdsløse øgedes, og ydelser og service, der var udviklet og givet til svage grupper, blev udstrakt til alle (eks.: børneydelser, daginstitutioner), og nye områder af velfærdspolitikken blev inddraget. Den omfattende, universelle, skattefinansierede velfærdsstat var blevet en realitet. På dette grundlag udbyggedes velfærden fortsat, selvom der i 1980'erne politisk introduceres en begyndende stramning af den sociale disciplin.

På den ene side indføres efterlønsreformen i 1979, førtidspensionen fik fire niveauer og ældreboligloven fra 1987 repræsenterede en innovativ tænkning, der kombineret med gratis, varig, hjemmehjælp fra 1989, støttede bevægelsen hen imod hjemmeboende ældre. På den anden side strammedes op på grund af den belastning arbejdsløsheden udsatte ydelsessystemet for. Med bistandsloven (1976) introduceres indkomstbortfaldsprincippet, men udviklingen viste, at dette var økonomisk uholdbart. Der blev indført overgrænser for ydelserne og navnlig begrænsninger for unge, så studerede efter 1983 ikke mere kunne opnå kontanthjælp. I 1990'erne og 2000 årene tages skridtet fuldt ud. Det er ikke længere personens svagelighed og sociale forhold, der er i fokus ved hjælp og tildelinger, men erhvervs- og arbejdsevnen. Den aktive arbejdsmarkedspolitik og aktive socialpolitik, stramninger i 1999 af efterlønsordningen og førtidspensionen tegner den nye linje: så mange, der kan, skal i beskæftigelse, om nødvendigt med beskæftigelsestilskud. Som Axel Mossin³ udtrykker det: Danmark bevæger sig fra Welfare mod Workfare. Fra tiden 2003 til 2011 er megen velfærdstankegang koncentreret omkring ydre og indre udfordringer. Velfærdsforliget 2006 tager udgangspunkt i Velfærdskommissionens påpegning af udfordringer fra demografien, globaliseringen og individualiseringen. Derved begynder fremtiden at spille en aktiv rolle for nutidens udformning af velfærdsstatens politiske design. Reformpolitikkerne satser herefter på, at unge og ældre skal ud af de sociale systemer gennem hurtigere studieforløb og senere tilbagetrækningsalder. Hertil kommer med virkning fra 1. juli 2011, stramninger i dagpengeperioden (nu 2 år fra 7-10 år initialt hen mod slutningen af forrige århundrede) og en skattereform i 2009, der sænkede de høje marginale indkomstkattesatser, for at øge incitamenterne for den enkelte til at yde mere på arbejdsmarkedet.

3. AM02: Socialpolitikken og velfærdssamfundets teori s. 140.

3. Dansk Velfærds resultater

Ved indgangen til 2011 står Danmark således med en stjernestatus globalt, med en velfærdsstat med høje pr capita indkomster (2009 på 9. pladsen i OECD), i et veludviklet demokrati, med verdens mest ligelige indkomstfordeling, og en betydelig redistribution af indkomster og risici via staten (sygdom, arbejdsløshed, alderdom osv.) oven i købet i en fri markedsøkonomi. Velfærden er udvidet til store befolkningsgrupper og ikke bare lavindkomstgrupperne, men middelklassen nyder også godt af den sociale sikrings betryggelse, i en model, der er alt andet end enstrenget finansieret. Så godt som alle finansieringskilder (bortset fra obligatoriske sociale bidrag fra arbejdsgivere, hvor Danmark har noget nær bundrekord) er taget i brug, i et indbyrdes blandingsforhold, fra skat, bidragsfinansiering, tvungne socialforsikringsordninger, frivillige forsikringsordninger med skattefradrag, såvel gennem arbejdsmarkedsordninger som brede ydelser, der står til rådighed for alle. Resultatet er en betydelig offentlig sektor, men relativ lav ledighed, en økonomi, der har opnået balance i forhold til udlandet (med store betalingsbalanceoverskud og en voksende fordring på udlandet), men som alle vestlige lande siden 2009 med meget store underskud på de offentlige budgetter og en offentlig gældsætning, om end af mindre størrelsesorden sammenlignet med de fleste andre EU-lande. Det er på denne baggrund at (CPP10) søger efter oprindelsen til den danske velfærdsstat og (GVM10) forlænger det store værk om dansk velfærdshistorie fra 1970erne til 2010.

4. Den tidligste oprindelse

Oprindelsen er for (CPP10) fra 1536, fra reformationen. Men rødderne strækker sig endnu længere tilbage i tiden, og gør det bestemt ikke oplagt, at det blev Danmark, der indtog stjernestatus i velfærdspolitik. Velfærd stod ikke hugget i runer i Danmarks dåbsattest. Fra rigets oprindelse omkring år 800 begynder dansk fjernhandel og skaber indtægter til samfundets top, med betydelig akkumulation af den økonomiske og politiske magt i det hierarkisk opbyggede samfund af højættede og lavættede familier og trælle, men med en spirende middelklasse af frie handelsmænd og håndværkere, der organiserer sig i gilder og lav. Det er det slægtskabsorienterede samfund med slaver og store forskelle på rige og fattige.⁴ Der er ikke meget på dette tidspunkt, i samfundsstrukturer, økonomisk eller i den politiske magt eller for den sags skyld i danskernes gener, der tyder på en spirende velfærdsunderskov. Fra vikingetiden til middelalderen kommer byerne ind som et nyt element, hvor mellemlaget i samfundet: håndværkere, købmænd og skibsredere m.fl. holder til og, hvor gilder og lav udvikler spinkle netværk til indbyrdes hjælp. På landet derimod er den gængse produktionsform store godskomplekser, der drives af fæstebønder på udlejet jord, med fæstekontrakter i et

4. Se i øvrigt (DS81) BD 1.

feudalistisk, socialt forhold. I denne fastlåste struktur melder den katolske kirke sig på den sociale platform, der er så godt som helt fri. Klostrene med munkeordnerne, især tiggermunkene, virker midt blandt menneskene i byerne og på landet. Deres ideologiske og sociale bevægelse er pauperismen, en søgen mod en fattig tilværelse for at komme Kristus nær gennem godgørende gerninger over for landets fattige.⁵ Dette bliver udgangspunktet for kirkens arbejde for fattige, det vil sige mennesker, der kun havde sine hænder til at skaffe sig sit udkomme, for syge, ældre, hjemløse, forældreløse børn og invalider. Men der er mange fattige og nødlidende, jordløse husmænd, landarbejdere, vagabonderende tiggere, og en stor del af den danske befolkning lever på et niveau, der grænser til det egentlige residuum.

5. Fra reformationen 1536 til socialreformerne i 1890erne

Efter reformationen mister kirken sin stærke position i samfundet, kongen overtager kirkens og klostrenes godser, og fattigforsørgelsen sendes ned til lokalsamfundene. Fra dette tidspunkt til de store reformer 350 år senere i 1890erne præges fattigideologien, bortset fra få år under oplysningstiden (1799-1803), af Kirkereformatoren Martin Luthers feltråb: Bekæmp tiggeriet, et råb der genlød op gennem tre til fire århundreder. Herved forskydes indsatsen bort fra bekæmpelse af fattigdommen. Staten skulle kun, med støtte fra Martin Luthers ideologi, nøjes med at give generelle rammer og regler for, hvordan det lokale samfund skulle udforme hjælpen.

Dette er afsættet for CPP10' hovedværk om dansk velfærdshistorie. Værket, der er på 615 s., hviler ikke på en entydig teoretisk tilgang, hvis hypoteser skal eftersøges. Det er, mener forfatterne, velfærdshistorien alt for kompliceret til. I stedet tages der afsæt i en institutionel tankegang, hvor der er fokus på velfærdspolitikens indretning og udbygning. Forfatterne understreger, at deres analyser drejer sig om de principper, værdier og ideer, velfærden bygger på. Der undersøges ideer, deres typer (forgrunds eller baggrunds) og deres effekter på den politiske debat. Det er altså klangbunden i den politiske ideologi i samfundet og socialpolitikens praktiske gennemførelse i politiske kredse, herunder uden for politikken og i befolkningen, der afsøges i det store værk. Samtidig sættes fokus på de fire store klassiske sociallove frem mod 1898: for fattige, gamle, syge og ulykkesramte. Herved bliver velfærdspolitikken grundigt eftergået fire gange, med to ekstra gange, en for idestrømningerne og en afsluttende sammenfatning. Man kan sige, at materialet bliver vasket, skyllet, tørret og strøget, hele seks gange, så der ikke er mange sorte, ubelyste pletter overladt til læseren; ligesom når man skal implementere strategier i store organisationer: vi gentager i det ulidelige for at sikre forståelsen og eksekutionen.

Den gængse holdning fra 1500 tallet til, at det var sognepræstens og byernes opgave at sikre de virkelig trængende fattige og syge, at de blev indlagt på hospitaler, levede

5. Se i øvrigt (DS81) BD 2.

op til den Lutheranske ideologi. De værdigt trængende fattige fik tiggertegn, og uværdigt trængende tiggere bortjagedes, eller indfangedes og sendtes i tvangsarbejde (eks.: 1708-loven). Periodelvis knaphed på arbejdskraft bragte også disse tanker på banen for at løse et akut samfundsproblem. Først med oplysningstidens tro på, at øvrigheden kunne forbedre forholdene for den enkelte, skiftede samfundet holdningen til fattigpolitikken. På det økonomiske og sociale område blev landboreformerne 1788 skelsættende fordi reformerne gødede jorden for en blomstrende økonomi ud til langt flere befolkningsgrupper, og frisatte fremtidige politiske kræfter, der senere arbejdede for en demokratisering af landet. I denne atmosfære gennemførte reformpolitikkerne omkring kongemagten, på baggrund af forudgående kommissionsarbejde, en kovending i socialpolitikken. Forordningen for København, 1799 handlede ikke mere om at afskaffe det omsiggribende, farlige og ukontrollable tiggeri, men om at hjælpe det enkelte menneske i social nød og bekæmpe fattigdommens svøbe. Metoden blev etablering af distrikter med opsøgende virksomhed ud til de fattige, hjælp og støtte til livets ophold og fremskaffelse af arbejde til de arbejdsføre. Ønsket var at forebygge sygdom og fattigdom. Særligt var opmærksomheden på børn. De skulle undervises og bispises i skolerne, en ideologi der førte frem til skolereformen 1814. I 1802 og 1803 udvidedes disse fremsynede ordninger til at gælde resten af landet.

Men ambitionerne fra 1799 holdt ikke længe. Frem til 1890erne plages Danmark af en lang række tilbageslag for fattige, syge, ældre, forældreløse børn og invalider. Der er økonomisk dårlige tider i Danmark under og efter napoleonstiden, og derfor er det sparelinien og bondefedtetheden blandt skatteyderne, gårdmændene, der vinder overhånd, og skatteyderne klager over tyngende fattigudgifter. Selvom omgangsforpleje og udlicitering af pleje af børn forsvinder ud af den kommunale praksis, vinder fattiggårde og tvangsarbejdsanstalter indpas. Hvad er der sket? Oplysningstidens fremsynede embedsmænd er borte fra magten og de nye sociale strukturer repræsenteret af Godsejere og Gårdmænd hylder henh. konservatismen og liberalismen. Alt præget af de skræmmende demografiske fremskrivninger, udført i Storbritannien af præsten, økonomen og samfundskritikeren Robert Malthus.

Liberalister og konservative finder senere i århundredet sammen i den nye socialpolitik, men deres ideologier er væsentligt forskellige. Konservative ser samfundet som en statspagt, som en nødvendighed, menneskene må underlægge sig. Tingene kan forbedres, men kun for at bevare samfundsstrukturen. Hensynet til helheden tæller højt; vi er uadskilleligt forbundne som mennesker, og vi skal tage hensyn til historien og de nedarvede rettigheder og kun gradvist udvikle samfundet under hensyntagen også til de kommende generationer. Socialpolitikken på statsplan kunne derfor ikke komme på tale, fordi det opløste de bånd, der i de lokale samfund knyttede individerne sammen i gensidig forpligtelser. I virkeligheden står liberalisterne i stærk kontrast til denne ideologi. Deres tanke er næringsfrihed og markedskræfternes usynlige hånd, den pri-

vate ejendomsret og selvudfoldelsen. De frie kræfter kan sprænge de gammelkonservervative bånd og strukturer. For det enkelte individ var det vigtigt at lyse af selvværd, selvrespekt, og selvudfoldelse, og vilje til at yde det yderste for selvopretholdelsen. Uden dette ville samfundets fundament smuldre.

Disse tanker førte til skel mellem de værdigt trængende, syge, børn, gamle og invalider over for de uværdigt trængende, de arbejdsføre fattige, de arbejdsløse og de dovne og samfundsskadelige. Denne tankegang førte debatten i 1860erne i to retninger. Dels egen finansiering gennem private forsikringer og opsparing i sparekasser til opretholdelse af tilværelsen ved ulykkelige sociale begivenheder, eller i værste fald hjælp og støtte fra frivillige bidrag gennem filantropiske foreninger. Dels oprettelse af fattiggårde og arbejdsanstalter, med adskillelse af børn og voksne mænd og kvinder, ordensregler og arbejdstvang. Det skulle kunne mærkes, at man var fattig. De mange kommissioner hjalp ikke de fattige meget, samfundet bevarede et todelt fattigvæsen de værdigt og de uværdigt trængende, i et samfund med to nationer: de rige og de fattige.⁶

Giganternes kamp mellem ideologierne fik Karl Polanyi⁷ til at beskrive udviklingen i ideerne som en dobbeltbevægelse i et pendul. Et dominerende paradigme udfordres af et andet med andre ideer, og fra tid til anden vinder det ny paradigme indpas og paradigmet ændres. Som da den katolske tiggerideologi udfordres af Martin Luther og han igen af oplysningstiden, der atter provokeres af liberalisterne. En bevægelse udfordrer en modbevægelse, der fører til en dialektik og et samspil. Udviklingen fører i den første grundlov af 1849, stærkt præget af liberalismen, til at fattige ganske vidst kunne hjælpes, men kun hvis de underkastede sig de betingelser loven foreskrev, nemlig tab af valgbarhed og stemmeret og i nogle tilfælde tab af civile rettigheder som giftermål, eller opholdssted. Til gengæld skulle uvældigt trængende straffes og sættes til tvangsarbejde. Også løfteparagraffen om næringsfrihed understregede den hele liberale holdning, at statslige indgreb eller snærende bånd ville forstyrre de naturlige samfundsforhold. Den efterfølgende næringsfrihedslov fra 1857, med virkning fra 1862, var dog snarere en lovgivningsmæssig overbygning og konstatering af den lange bevægelse hen mod erhvervs-mæssig frihed, der havde fundet sted efter napoleonskrigene⁸ og som skabte en blidere overgang til markedsøkonomi uden laugsreguleringer. Selvom næringsfrihedsloven ophævede lav og gilder og udstedte forbud mod tvungne svendeprøver, indførte håndværkerne frivillige prøver, organiserede sig i frivillige foreninger, der videreførte det sociale arbejde som sygekasser, begravelses- og

6. Disraeli, Benjamin: (1804-1881) »Europa har to nationer: de fattige og de rige«.

7. Karl Polani. *The Great Transformation* 1944. *The Political and Economic origins of our time*, 1957.

8. Se: (5) (BH79): *Håndværkets økonomiske Historie, 1879-1979*, s. 62: Der var indført en række frie næringer, myndighederne havde tilladt ikke håndværkere at nedsætte sig som selvstændige, en stigende virksomheds-etablering uden for byerne, uden bevilling, flere kunne nedsætte sig som frimestre, kun sendrægtigt forfulgte myndighederne fuskere og bøn-hase, og gifte svendes hjemmearbejde ekspanderede.

enkehjælp og senere opførtes stiftelser til gamle. Så Polanyis pendul svingede noget hurtigere tilbage end det fremgår af fremstillingen i (CPP10).

6. Socialreformerne 1891-1898

Midt i denne lidet positive ideologiske baggrund for socialpolitikken gennemføres omfattende reformer i 1890'erne. Flere forhold trækker politikerne i Rigsdagen til forhandlinger om socialreformer trods deres ideologier. For det første er fattigdommen taget til, og flere kommissioner har forgæves søgt varige løsninger på problemet. For det andet er landbruget presset af stadig dårlige internationale konjunkturer, af et lønpres, og stigende salgspriser, samtidig med at det nye arbejderparti, Socialdemokratiet bider Venstre i haserne. Venstre er i en livsvigtig kamp mod Højre om konstitutionen og ser ikke gerne deres politiske position udfordret af andre. Det fører de reformvenlige moderate i Højre og Venstre til at finde sammen. Samtidig kan to andre formål bestræbes, at overvælde skattebyrden fra land til by og lægge op til det højeste mål: på længere sigt at styrte Konseilspræsident Estrup og indføre parlamentarismen i Danmark. Det økonomiske middel blev generelle skatter. Ølproduktionen fik en uventet strategisk vigtig plads i den danske socialpolitik, gennem en ølskat, der skulle finansiere den nye socialpolitik.

Fattigloven af 1891 var i princippet udtryk for en fortsættelse af traditionen snarere end fornyelse. De, der ikke kunne skaffe det nødvendige til livets ophold, kunne få hjælp mod at underkaste sig fattighjælpens virkninger. Man mistede stadig valgret og valgbarhed, fattigvæsenet kunne kræve erstatning for hjælpen eller eftergive den. Fattige var som rekrutter i hæren underlagt fattigvæsenets disciplin, beslutninger og tilsyn. Der var stadig bestemmelser om samtykke ved ægteskab og fattighjælpen kunne blokere for anden hjælp (eks.: alderdomsydelsen). Endelig kunne man risikere at blive hjemsendt til sin egen forsørgelseskommune, om muligt til fods. Hjælpen blev dog givet fortrinnsvis i hjemmet, og der var tale om et brud i gængs tankegang: man kunne uden betingelser få hjælp til bl.a. læge og jordmoder. Endelig fik kommunerne pligt til at skaffe husvilde tag over hovedet.

Samtidig trækkes ældre ud af fattigloven og over i en særlov om alderdomsunderstøttelse. Her opfandt man et figenblad for at følge samfundsordenens gennemgående princip om hjælp til selvhjælp, om forudgående bidrag inden støtte til ældre kunne finde sted. Personer over 60 år blev dækkede med små trangsbedømte ydelser, hvis de havde holdt sig fri af fattighjælp de sidste 10 år, havde boet i landet et tilsvarende antal år og ikke ved dom var fundet skyldig i en i den offentlige mening vanærende handling eller havde fået æresoprejsning. Personkredsen blev universel og tildelingskriterierne blev altså alder, trang og værdighed. Take up rate inden for målgruppen blev dog kun 20% omkring år 1900, men udgiftsniveauet ift. BNP blev større end den tyske ordning,

fordi der var en 70 års grænse, og ikke som i Danmark 60 år i en her og nu ordning.

Allerede året efter, 1892 gennemførtes støtte til sygekasserne. Der er gennem flere kommissioner fokus på hjælp til selvhjælpsprincippet i kasserne, selvom kasserne led af mangler, ønskede man ikke tvangsprincippet som i Tyskland, men at fortsætte det frivillige princip. Så den gennem kommunen og staten støttede personkreds blev de, der frivilligt og gennem egne bidrag, meldte sig ind i en sygekasse, som anerkendtes af sygekasseinspektører. Ydelserne blev fri lægehjælp, sygehusophold og sygedagpenge med 3 karensdage, og hjælp i max 6 måneder. Kassernes antal voksede og tredobledes fra 1893 til 1915, hvor medlemstallet var næsten 900.000 personer.

Tyskland fik en tvungen ulykkesforsikring i 1888. Lovgivningen blev dobbelt tvangsmæssig. Industrierbejderne var siden 1883 tvangsmæssigt medlem af en sygekasse, hvortil han betalte bidrag. Sygekassen dækkede de første 13 uger efter en ulykke, derefter kunne han, hvis lidelsen efter ulykken stadig ikke var overvundet, overgå til den tvungne arbejdsgiversammenslutnings forsikring. I Danmark medførte den spæde industrialisering flere ulykkeshændelser og landbrugets samtidige mekanisering i samklang på jernbanens fremdrift trak i samme retning. Også dansk bygningshåndværk indebar ulykkesrisici. Området herhjemme var imidlertid dækket af private forsikringer, hvis ledere og repræsentanter behændigt involverede sig i den politiske beslutningsproces. Det kom bl.a. frem, at antallet af arbejdsulykkesinvalidere pr. medarbejder i Tyskland var vokset til det dobbelte efter indførelsen af den tvungne tyske ulykkesforsikring og en eksplosion på 50% forøgelse i udgifterne til følge (fra 1888 til 1894) førte bl.a. til, at de danske lovgivere gik andre veje. I Danmark fastholdt man det liberale forsikringsprincip uden tvang. Der blev tale om en slags udvidet erstatningslovgivning, der medtog hændelige ulykker i en række farlige erhverv, hvorunder maskinulykker i landbruget også var inkluderet. Ulykkerne blev defineret som en pludselig ydre påvirkning, mens de mange ulykker, der ofte først medførte lidelser flere dage efter, ikke var dækkede. Forsikringen trådte til efter 13 uger, hvorefter man kunne få erstatning, men ikke livsvarig forsørgelse. Arbejdsgiverne og/eller forsikringsselskabet bar den økonomiske risiko. Man kan undre sig over, at der ikke gennemførtes en tvangsordning, fordi Danmark netop bestod af mange mindre og mellemstore virksomheder i industri, byggeri, handel og landbrug, hvor risiciene er store og forsikringsbehovene derfor størst.

7. Den danske velfærdsstats initiale design

Ved indgangen til det 20. århundrede stod Danmark således med en socialpolitik, der var sammenstykket af ikke en, men mange finansieringskilder, og ikke af enstrenge finansieringsordninger, men blandede. Fattigloven og loven om alderdomsunderstøttelse blev universelle, afhængig af statsborgerskabet og finansieret over de gene-

relle skatter, men administreret gennem kommunen. Sygekassesikringen derimod var finansieret af egne bidrag med tilskud fra det offentlige. Medlemskabet var frivilligt og administrationen blev afviklet gennem eksisterende eller nyoprettede private sygekasser. Ulykkesforsikringen derimod byggede på rene private forsikringsprincipper, og det var frivilligt for virksomhederne, om de ville forsikre sig.

Der synes ingen masterplan for socialpolitikens oprindelse, men derimod en platform støbt i beton, hvis principper, som det senere i århundrede ikke var til at vige fra, og senere generationer syede videre på dette velfærdstæppe i patchwork. Vigtigt var det dog, at platformen gav »her og nu« ydelser til de trængende, var velegnet til redistribution mellem alle borgerne, og at der kunne sys videre på systemet. Mens syge og gamle straks mærkede effekten af reformen i Danmark, var det tyske system på eks. ældreområdet en økonomisk billig omgang for Kansler Bismarck. Her skulle borgerne være 70 år og have været medlem i mange år, inden den fulde udbetaling fandt sted.

8. Tiden efter 1970

I den modsatte ende af århundredet 1970-2010 undersøges i to bind over 670 s. (GVM10) den danske velfærdsmodels modne periode. Metoden er ganske anderledes. Værket tager udgangspunkt i eksisterende forskning inden for velfærdsområdet i den omhandlende periode, ikke mindst Rockwoolfondens bidrag, og en særlig omnibusanalyse danner baggrund for en undersøgelse af befolkningens holdning til den danske velfærdsmodel. Vi er nået til velfærdssamfundets gyldne periode, hvor det danske samfund samtidig udfordres af flere kriser, der medfører stor arbejdsløshed (1974-82) og hvor udfordringerne fra fremtiden bringer politikerne ind på omfattende reformprogrammer for tilbagetrækningsordninger, dagpenge og skat, mens velfærdsmodellen stort set drives videre i uændret form. Hofhistorikeren Knud J.V. Jespersen citeres for, at den danske velfærds model historisk set har skabt:

...helt enestående social og økonomisk baggrund for de nuværende efterkommere af ikke så fjerne tiders fattige danske bondebefolkning, der bogstavelig talt levede fra hånden og i munden...⁹

I 1970erne var den store velfærdsfremgang i højt gear ved at være slut. De store engangsgevinster som, blandt meget andet, overflytningsgevinsterne fra landbrug til industri og den offentlige sektor og liberaliseringen af den danske handel ud mod verden, var ved at være opbrugte, og i stedet løber verden og den danske velfærdsstat ind i en serie af oliekriser, der udstiller velfærdsstatens svagheder. Den »Skandinaviske model« mistede sin uskyld: det var ikke muligt med det daværende velfærdsdesign og

9. (GVM10) BD 1, s. 89.

økonomisk politiske setup, at undgå arbejdsløshed. I Danmark steg ledigheden først til 5-6% af arbejdsstyrken fra 2-3% og senere i 1980'erne til 8-12%. (GVM10) påpeger, at prisstrukturen i Europa udvikledes forskelligt, og det gav vanskeligheder for dansk konkurrenceevne og for betalingsbalancen. Finanspolitisk betød dette at Polanyis pendul svingede væk fra John M. Keynes over mod Milton Friedman til en mindre interventionistisk og aktivistisk politik i 1980'erne for at skabe en mere stabil makroøkonomisk ramme omkring produktionslivet, hvilket f.eks. i USA betød at vægten omlagdes fra fuld beskæftigelse over til prisstabilitet. Danmark tiltrådte denne linje, efter at landet under den borgerlige regering fra 1982 knyttede den danske valuta til D-marken og derved stoppede årtiers devaluerings- og prisglidningspolitik.

Alligevel finder forfatteren til værket, at der sker en klar forbedring af ligheden i Danmark fra 1970'erne til midten af 1990'erne. Fordelingen er, som der bliver sagt, hjerteblodet i den socialpolitiske debat. Mange årsager fører frem til en større uligning af indkomsterne frem til midten af 1990'erne. Aldersforskydningen, så der er blevet flere personer i gruppen af 25-65 årige og relativt færre over 65 år, og en betydelig offentlig redistribution til ledige og til et stigende antal personer uden for arbejdsmarkedet. Samtidig peger værket på de gode muligheder, der er i Danmark for borgere med lavindkomster til at rykke op i indkomstgraderne. Den samlede indkomstmobilitet er ret høj. Når man over flere perioder måler fattigdom, er der i Danmark som i Holland kun permanent 2% af befolkningen fattige, mod 7% i USA, Australien, Portugal og Grækenland, ifølge 2008-studiet om lighed og ulighed i OECD-området.¹⁰ Det er en interessant observation, at (GVM) bruger megen plads til modsvarende at forklare opdriften i Gini fra 1990'erne til 2000-årene. Gini vokser fra ca. 20% til 25%, hvilket må siges at være bemærkelsesværdigt. Forklaringerne hentes fra demografien: vi importerer en underklasse af lavt betalte indvandrere, derimod har øvrige demografiske forskydninger, som ændringer i familiestrukturen, ingen mærkbare virkninger på fordelingen. Samtidig har flere gennemført langvarige uddannelser under lave indkomster i studietiden. Endelig har kapitalindkomsterne i aktieafkast og stigende boligpriser kraftigt øget uligheden. Man må antage at dette gælder frem til den finansielle krise, der derefter helt vendte dette billede på hovedet, selvom skattereformen må forventes at trække i modsat retning. Målt på de mere usikre livsindkomster konkluderer (GVM), at der har været tale om en svagt formindsket ulighed. Men internationalt står Danmark som det mest ligelige samfund: »Danmark og Sverige var stadigvæk OECD-kredsens, og dermed hele verdens, mest ligelige samfund (i 2005 red.)«. ¹¹

Bogen følger op på det syn, Socialforskningsinstituttet (SFI) tidligere fremsatte, at der synes at være opstået en vis konvergens mellem de forskellige typologier af velfærdsstater.¹² Den omfattende skandinaviske og hollandske model er ved at blive

10. Growing unequal? OECD 2008.

11. (GVM10) BD2, s. 492.

kopieret af de mindre omfattende modeller. Den skandinaviske model kopierer imidlertid også. Inden for pensionsområdet opbygges i 1980'erne i Danmark arbejdsmarkedsordninger med bidragsfinansiering, som supplement til det offentlige system. Dette giver gnidninger og samspilsproblemer over til det offentlige system. I EU foregår der imidlertid ikke en egentlig harmonisering på området, men kun en blød koordinering med svagere, usikre og mere langsigtede virkninger. Der peges dog på, at Vesteuropa har udvist en vis grad af catch up på børne- ældre- og arbejdsmarkedsområdet. Forfatteren måler også den samlede offentlige og private nettovelfærd i de forskellige velfærdstypologier. Mens den universelle model når 25,5% af BNP, eller kun knap 1% point mere end den residuale model, ligger både den liberale model (eks. England) og den kontinentale (eks. Tyskland og Frankrig), overraskende langt højere end den danske.

En af de helt centrale problemstillinger for velfærdsstaten i Danmark bliver i stigende grad incitamentsstrukturen på arbejdsmarkedet. Fra 1970 til 1993 øgedes arbejdsstyrken med omkring 600.000 personer og nåede i 2009 samlet 2,9 mill. Forskydningerne over perioden er velkendte fra bl.a. velfærdskommissionen: Ældre medarbejdere har ladet sig pensionere tidligere, arbejdsmarkedsordninger og sociale ordninger, som efterløn og førtidspension, har taget sin andel af den tidlige tilbagetrækning, men mens erhvervsfrekvenserne faldt frem til midten af 1990'erne var de kraftigt stigende frem til slutningen af 2000'erne. Den danske flexicurity model kan tage en del af æren herfor. OECD påpeger, at ansættelsesbeskyttelsen er lavest i Danmark og Storbritannien. Ancienniteten blandt medarbejderne er lige så lav i Danmark som USA. Mobiliteten er stor, tilpasningsevnen i erhvervslivet ligeså. Samtidig faldt kompensationsgraderne for en produktiv industriarbejder i forhold til arbejdsløshedsunderstøttelsen fra 65% i 1993 til 56% i 2007. Lavindkomstgruppen, der ved ledighed kan få øgede ydelser som boligsikring, opnår dog helt op til 75% compensation. Dette kombineret med at dagpengeperioden hele tiden justeredes nedad (for 1. juli 2011 at lande på 2 år), samtidig med at betingelserne for at stå til rådighed for arbejdsmarkedet strammedes, understregede den glidende bevægelse fra »Welfare« til »Workfare«. Flexicuritymodellen var dog dyr i drift, idet udgifterne til arbejdsmarkedspolitikken var den højeste i Danmark blandt OECD landene.

(GVM10) er lykkeligvis udrustet med et detaljeret navne- og sagsregister, hvad der er et krav for at kunne finde rundt i et så stort værk. Det er ikke indgået i (CPP10), og er formentlig først planlagt efter sidste bind, men det er en mangel i så stort et værk. Til gengæld holder (CPP10) sig helt til det, som historie kan: fortælle og analysere på de begivenheder og de fakta, der rent faktisk fandt sted, uden utidig debat, en leveregel (GVM10) ser stort på i allersidste del, s. 576, ff.

12. The Danish National Institute of Social Research, »Social Security in Europe, book no. 1-6, 1992-1998.

9. Afslutning

Velfærdssamfundet er en kompleks størrelse, og det er ikke muligt med enkelte greb at beskrive dens udvikling. Der er imidlertid træk helt tilbage fra oprindelsen, der bør fremhæves som vigtige for modellens udvikling eller dens indbyggede barrierer. Det har vist sig at være en uhyre vanskelig operation at modernisere modellen. En vis konvergens med Europas velfærdsmodeller er dog opnået gennem en fonds- eller forsikringsmæssig overbygning i eks. de bidragsfinansierede arbejdsmarkedsmarkeds-pensioner, som blev et af det mest markante træk ved innovationen i velfærdsmodellen i nyere tid; en idemæssig udvikling, der i princippet foregik uden for den offentlige sektors normale operationsområde. Det andet vigtige træk, der er bevaret siden oprindelsen, er den varierede sammensætning af finansieringskilder, der imidlertid viser sig at blive modellens styrke i tilpasningen til den fremtidige demografiske udvikling. Mens den universelle skattefinansierede del af velfærdsprogrammerne er velegnede til at redistribuere indkomster og de mange velfærdsrisici, er forsikringsordningernes styrke, at de i højere grad kan sikre mod demografiske forskydninger, eller ændringer i f.eks. tilbagetrækningsadfærd. Den danske model med mange forskellige finansieringskilder ligner således en prototype på Verdensbankens flersøjle-model for et pensionsystem. (jf. Worldbank SP discussion paper No. 0824 og BH08).

I begyndelse af det 21. århundrede er det opportunt at vende tilbage til udgangspunktet: Polanyis pendul og modbevægelse. Indtil videre har pendulet svinget ud mod neoklassikerne, og radikale liberalister som Robert Nozick og Milton Friedman. Når vi skal vurdere den underlægningsmusik, der kan tænkes at blive baggrund for de næste års velfærdspolitiske innovation, skal vi vurdere, hvor langt pendulet er svinget ud, og hvornår og hvordan det vender tilbage. Hænger Friedmans eller Keynes' efterkommere på pendulet, når det svinger den anden vej?

Literatur

- Baldwin, P. The Politics of social Solidarity. Class Bases of the European Welfare State 1875-1975. Cambridge University Press 1990. (PB90).
- Christiansen, N. F., J. H. Petersen og K. Petersen. Dansk velfærdshistorie, BD 1, 1536-1898. Frem mod socialhjæpsstaten. Syddansk Universitetsforlag, Odense 2010, (CPP 10).
- Christoffersen, H. Dansk velfærdspolitik efter 1945. Nyt Nordisk forlag, Arnold Busck, København 1984. (HC84).
- Hansen, S. Aa. m.fl. Dansk Socialhistorie, BD 1-7 Gyldendal, 1974-81 (DS81).
- Hastrup, B. Social welfare, the Danish model. Multivers, København 2011. (BH11).
- Hastrup, B. The social Contract between the Generations. Multivers København 2008. (BH08).
- Hastrup, B. Håndværkets Økonomiske Historie 1879-1979. Håndværksrådets forlag, København 1979. (BH79).
- Mogensen, G. Viby. Det danske velfærdssamfundets Historie. Tiden efter 1970. Gyldendal, København 2010, (Bind I og II) (GVM 10).
- Mossin, A. Socialpolitikken og velfærdssamfundets teori, note 86. Københavns Universitets økonomiske institut, 2002 (AM02).

Finanskrisen og den økonomiske videnskab

Peter Birch Sørensen

Danmarks Nationalbank, E-mail: pebs@nationalbanken.dk

Den internationale finanskrisen, der kulminerede i efteråret 2008, kom som en overraskelse for det store flertal af økonomer. Krisen var så voldsom, at aktiviteten på flere centrale finansielle markeder tørrede helt ud. Uden de omfattende indgreb fra stater og centralbanker ville vi utvivlsomt have oplevet et totalt sammenbrud af det globale finansielle og økonomiske system.

Denne alvorlige forskrækkelse har udløst dyb selvransagelse hos mange økonomer. Paul Krugman, der i 2008 modtog Nobelprisen i økonomi, stillede forrige år i New York Times spørgsmålet: »Hvordan kunne økonomerne tage så meget fejl?« Den hårde debat, som Krugmans artikel udløste, er blot ét blandt mange eksempler på de voldsomme faglige kontroverser blandt økonomer, som finanskrisen har igangsat. Debat-tørerne kan groft sagt inddeles i to lejre. På den ene side står de, der ser et behov for en fundamental nyorientering af økonomisk teori. På den anden side finder man dem, der mener, at de centrale træk ved finanskrisen udmærket kan forklares ved brug af gængs økonomisk teori. I det følgende vil jeg give mit eget bidrag til denne diskussion.

Finanskrisen begyndte så småt i foråret 2007, hvor nogle finansielle institutioner meldte om tab relateret til de såkaldte subprime-lån til mindre kreditværdige amerikanske boligejere. Mange af disse låntagere havde fået bevilget boliglån med afdragsfrihed og i nogle tilfælde også meget lave renter i de første år af lånets løbetid, da de i praksis ikke havde råd til at betale normale renter og afdrag. Subprime-låntagerne var derfor i stort omfang afhængige af, at de kunne optage nye billige lån med sikkerhed i boligen for at kunne tilbagebetale de gamle lån på det tidspunkt, hvor de ellers skulle have begyndt at betale normale renter og afdrag. Dette forløb nogenlunde glat, så længe de amerikanske boligpriser blev ved med at stige, hvilket de havde gjort, så længe de fleste kunne huske tilbage. Fra midten af 2006 begyndte huspriserne i USA imidlertid så småt at falde, og faldet tog til i løbet af 2007. Det satte en stopper for trafikken med at finansiere tilbagebetaling af gamle lån via optagelse af nye lån, da sikkerheden i form af friværdien i boligerne forsvandt. På grund af manglende betalingsevne hos låntagerne opstod der derfor tab på obligationer udstedt på basis af subprime-lån.

Indlægget er baseret på et foredrag i Videnskaberne Selskab den 28. april 2011. Jeg takker Kim Abildgren, Anders Møller Christensen og Ib Hansen for nyttige kommentarer til en tidligere version af dette manuskript. Alle tilbageværende mangler er mit eget ansvar. De fremførte vurderinger og synspunkter repræsenterer ikke nødvendigvis Nationalbankens holdning.

Vi ved alle, hvordan denne misere endte. Efter længere tids finansiell uro accelererede finanskrisen voldsomt den 15. september 2008, hvor den store amerikanske investeringsbank Lehman Brothers fik lov at gå ned. Efter Lehmans fallit frøs penge-markederne helt til. Bankerne turde ikke længere låne ud til hinanden, da ingen havde overblik over, hvilke banker der stod til at lide tab som følge af det finansielle kaos. Centralbanker og regeringer måtte derfor gennemføre drastiske indgreb for at undgå et fuldstændigt sammenbrud af tilliden til det finansielle system.

Det er ikke nogen stor intellektuel udfordring at forklare, hvorfor den uansvarlige långivning på det amerikanske subprime-marked før eller siden måtte gå galt. Udfordringen består i at forstå, hvorfor problemer i et forholdsvis lille hjørne af det internationale finansielle system kunne bringe hele systemet på randen af det totale sammenbrud. Udlånene til de amerikanske subprime-låntagere udgjorde trods alt kun en meget beskeden andel af de samlede aktiver i det globale finansielle system. Der må have været nogle faktorer, som gjorde systemet meget ustabil, siden den lille subprime-tue kunne vælte hele det store finansielle læs.

Spørgsmålet er, om man i gængs økonomisk teori kan finde gode forklaringer på, hvorfor det finansielle system var blevet så ustabil? Svaret er et klart »Ja«. En økonom vil fokusere på, at den finansielle sektor kan være præget af incitamenter til overdreven gældsætning og risikotagning, og at priserne på finansielle aktiver ikke altid afspejler den samfundsmæssige investeringsrisiko. Af forskellige grunde, som jeg vender tilbage til, fik disse uheldige incitamenter og tendenser lov at slå stærkere igennem i de seneste årtier end tidligere.

Lad os se nærmere på, hvilke økonomiske incitamenter, der kan være tale om. Langt de fleste finansielle virksomheder drives på aktieselskabsform. Som påpeget af bl.a. den tyske økonom Hans-Werner Sinn (2010, kap. 4) kan aktionærernes begrænsede økonomiske ansvar give en tilskyndelse til overdreven risikotagning og gældsætning, hvis aktieselskabets kreditorer ikke er tilstrækkeligt velinformede. Tabel 1 giver en forenklet illustration af Sinns pointe. I tabellen ses på to alternative investeringer, der begge indebærer en samlet investeringsudgift på 100, finansieret af en aktiekapital på 20 og et lån på 80, der skal forrentes med 5%. Den ene investeringsmulighed giver med 90 procents sandsynlighed et afkast på 7,2%, mens der er 10 procents sandsynlighed for at tabe 15% af den investerede kapital. Den anden investeringsmulighed er mere risikabel, da den ganske vist giver et afkast på 11,1% i den situation, hvor investeringen går godt, men til gengæld indebærer en 10 procents sandsynlighed for, at halvdelen af den investerede kapital går tabt. I den sidstnævnte situation går selskabet fallit, da både egenkapitalen og en del af fremmedkapitalen mistes. Tallene i tabellen viser de forventede investeringsafkast, beregnet som afkastet i de to situationer, hvor investeringen går henholdsvis godt og skidt, ganget med sandsynlig-

Tabel 1. Kapitalafkast og risikoprofil. (Samlet investering = 100, aktiekapital = 20, lånekapital = 80, lånerente = 5%).

	Investering med lav risiko: 90% sandsynlighed for afkast på 7,2% 10% sandsynlighed for afkast på -15%	Investering med høj risiko: 90% sandsynlighed for afkast på 11,1% 10% sandsynlighed for afkast på -50%
Forventet afkast af samlet investering	$\frac{0,9 \times 100 \times 1,072}{100} + \frac{0,1 \times 100 \times 0,85 - 100}{100} = 5\%$	$\frac{0,9 \times 100 \times 1,111 + 0,1 \times 100 \times 0,5 - 100}{100} = 5\%$
Forventet afkast til långivere	$\frac{0,9 \times 80 \times 1,5 + 0,1 \times 80 \times 0,05 - 80}{80} = 5\%$	$\frac{0,9 \times 80 \times 1,05 + 0,1 \times 50 - 80}{80} = 0,8\%$
Forventet afkast til aktionærer	$\frac{0,9 \times (100 \times 1,072 - 80 \times 1,05)}{20} + \frac{0,1 \times (100 \times 0,85 - 80 \times 1,05) - 20}{20} = 5\%$	$\frac{0,9 \times (100 \times 1,111 - 80 \times 1,05) - 20}{20} = 22\%$

hederne for de to udfald. Afkastraterne i de to projekter er valgt sådan, at både den mindre og den mere risikable investering giver et samlet forventet afkast på 5%. Ved gennemførelse af det mindre risikable projekt vil både långivere og aktionærer ligeledes få et forventet afkast på 5%, som det fremgår af første søjle i tabellen. Men hvor långiverne har fuld sikkerhed for at opnå et afkast på 5%, uanset om investeringen går godt eller skidt, så står aktionærerne til at tabe 19 ud af deres investerede kapital på 20, hvis projektet går dårligt. I så fald skal aktionærerne nemlig dels dække projektets tab på 15 og dels betale 4 i renter til långiverne. En risikoavers investor vil derfor foretrække en fredelig tilværelse som passiv kuponklipper frem for at stille risikovillig aktiekapital til rådighed for et projekt som det, der er beskrevet i første søjle i tabel 1.

Projektet i tabellens anden søjle er derimod langt mere interessant for aktionærerne. Det gælder på trods af, at projektet indebærer en større risiko uden at give et højere forventet samlet afkast. Det behagelige for aktionærerne er, at långiverne nu bærer størstedelen af tabet i den situation, hvor projektet slår fejl, da aktionærernes ansvar er begrænset til deres indskudte aktiekapital. Hvis projektet derimod går godt, får aktionærerne hele merafkastet ud over markedsrenten på de 5%. Aktionærerne kan dermed se frem til et respektabelt forventet afkast på 22%, mens de har en 10% risiko for at miste hele deres investerede kapital på 20. I det mindre risikable projekt har aktionærerne den samme risiko på 10% for at miste en kapital på 19, men her opnår de kun

et forventet afkast på 5%. Aktionærerne vil derfor helt klart foretrække at gennemføre det mere risikable projekt, med mindre de er ekstremt risikoaverse.

På grund af risikoen for tab får långiverne kun et forventet afkast på 0,8% i det mere risikofyldte projekt. En oplagt indvending mod eksemplet i tabellen er derfor, at långiverne vil kræve en højere rente for at finansiere den mere risikable investering frem for det mindre risikable projekt. Hvis långiverne f.eks. justerer deres rentekrav, så de opnår det samme forventede afkast på 5% i de to investeringssituationer, vil aktionærerne være slået tilbage til start, da de i så fald ikke kan hæve deres forventede afkast op over 5% ved at vælge det mere risikable projekt.

Moderne økonomisk teori fremhæver imidlertid, at mangelfuld og asymmetrisk information ofte forhindrer denne disciplinerende rentemekanisme i at virke. Specielt i den finansielle sektor vil indskydere og långivere typisk kun have begrænset viden om risikoprofilen i de mange forskellige investeringsmuligheder, som en moderne finansiell virksomhed står overfor. Bankerne låner jo ud til et stort antal virksomheder og husholdninger, hvis kreditværdighed de undertiden selv har svært ved at vurdere. Endvidere investerer bankerne ofte store beløb i komplicerede strukturerede værdipapirer og afledte finansielle instrumenter. Selv de mere sofistiskerede långivere og indskydere kan derfor have meget svært ved at vurdere, hvorvidt og i hvilket omfang en bank har omlagt sine aktiver i retning af mere risikable investeringer. Banken selv har ingen interesse i at informere om en sådan porteføljeomlægning, da det vil udløse krav om højere rente fra bankens långivere, hvilket medfører lavere afkast til bankens aktionærer.

Problemet med at sikre en passende disciplinerende risikopræmie i forbindelse med lån til bankerne forstærkes af de offentlige indskydergarantiordninger. I tidligere tider oplevede man ofte under finansielle kriser, at mange banker gik ned, fordi et stort antal indskydere i panik trak deres indlån ud på samme tid. Garantiordninger kan forebygge sådanne panikdrevne stormløb mod bankerne. Til gengæld betyder de også, at indskydere med indlån under den garanterede grænse ikke har nogen tilskyndelse til at tjekke, om banken forvalter indskuddene på ansvarlig vis, da indlånene jo under alle omstændigheder er sikrede.

Indskydere og långivere, der ikke er dækket af en eksplicit garantiordning, vil heller ikke have incitament til at indsamle information om risikoen ved bankernes udlån og investeringer, hvis der er udbredte forventninger om, at staten vil træde til og dække deres tilgodehavender, såfremt banken bliver insolvent. Det er det velkendte »too-big-to-fail« problem: Staten kan føle sig presset til at redde nødlidende banker, eller i hvert fald til at holde deres kreditorer skadesløse, hvis en konkurs i en bank truer den generelle finansielle og økonomiske stabilitet. Finanskrisen leverede utallige eksempler på dette.

Informationsproblemer og eksplicite eller implicite garantier kan også give banker et incitament til at arbejde med en uhensigtsmæssigt lav egenkapital. Det er illustreret i tabel 2. Forudsætningerne i tabellens første søjle er helt identiske med forudsætningerne i første søjle af tabel 1, og de forventede afkast er derfor også de samme. 80% af investeringen finansieres med lånekapital, mens 20% finansieres med aktiekapital, og aktionærene har kun udsigt til et pauvert forventet afkast på 5%, selvom de risikerer at miste 95% af deres kapital, hvis projektet slår fejl.

Finansieringsmodellen i anden søjle i tabel 2 er langt mere attraktiv for aktionærene. Her finansieres kun 5% af investeringen med aktiekapital, idet långiverne leverer de resterende 95%. Da långiverne nu har sat mere på spil, står de til et tab på 10, hvis investeringen går dårligt, og deres forventede afkast falder derfor fra 5% til 3,4%. I forhold til situationen med lav gearing taber aktionærene kun en lidt større andel af deres kapital, nemlig 100% i stedet for 95%, hvis investeringen går skidt, men deres forventede afkast når nu op på hele 34,5%.

Aktionærene har altså opnået et langt gunstigere forhold mellem forventet afkast og risiko og vil derfor foretrække finansieringsmodellen med høj gearing, dvs. den høje belåningsgrad. Långiverne står derimod med en risiko for et større tab, og derfor kan man argumentere, at de vil kræve en højere lånerente end de 5%, der er forudsat i taleksemplet med den høje gearing. Et berømt teorem formuleret af Modigliani og Miller (1958) siger netop, at aktionærene ikke kan tiltvinge sig en mere fordelagtig kombination af forventet afkast og risiko ved at øge selskabets gældsprocent, fordi långiverne vil tage sig betalt for den højere risiko for tab, som de løber.

Modigliani-Miller teoremet forudsætter imidlertid, at långiverne er fuldt lige så velinformerede om selskabets risikoprofil som aktionærer og ledelse, og at lånekapitalen ikke er omfattet af eksplicite eller implicite statsgarantier. Men som eksemplet i tabel 1 viste, kan banker have et incitament til at underdrive, hvor risikable deres investeringer er. I eksemplet i tabel 2 kan det derfor tænkes, at långiverne undervurderer størrelsen af tabet i det ugunstige investeringsudfald, eller at de undervurderer sandsynligheden for, at tabet indtræffer. Hvis långiverne vurderer tabsrisikoen til at være tilstrækkeligt lille, så vil de forvente, at de under alle omstændigheder vil få deres lån tilbagebetalt med rente, med mindre investeringens egenkapitalandel falder ned under et kritisk minimumsniveau. Selvom den sande tabsrisiko er givet ved tallene i tabel 2, kan långiverne altså fejlagtigt tro, at deres risiko ikke øges mærkbart ved, at gearingsgraden af investeringen øges fra 80 til 95%. I så fald vil de heller ikke se nogen grund til at kræve en højere rente.

Men selv hvis långiverne indser, at den højere gearing påfører dem væsentligt højere risiko for tab, vil de mangle incitament til at kræve en højere lånerente, hvis de er dækkede af eksplicite eller implicite statslige garantier, der kompenserer dem for

Tabel 2. Kapitalafkast og gearing. (Samlet investering = 100, 90% sandsynlighed for afkast på 7,2%, 10% sandsynlighed for afkast på -15%, lånerente = 5%).

	Lav gearing: Lånekapital = 80 Aktiekapital = 20	Høj gearing: Lånekapital = 95 Aktiekapital = 5
Forventet afkast af samlet investering	$\frac{0,9 \times 100 \times 1,072}{100} + \frac{0,1 \times 100 \times 0,85 - 100}{100} = 5\%$	$\frac{0,9 \times 100 \times 1,072 + 0,1 \times 100 \times 0,85 - 100}{100} = 5\%$
Forventet afkast til långivere	$\frac{0,9 \times 80 \times 1,05 + 0,1 \times 80 \times 1,05 - 80}{80} = 5\%$	$\frac{0,9 \times 95 \times 1,05 + 0,1 \times 85 - 95}{95} = 3,4\%$
Forventet afkast til aktionærer	$\frac{0,9 \times (100 \times 1,072 - 80 \times 1,05)}{20} + \frac{0,1 \times (100 \times 0,85 - 80 \times 1,05) - 20}{20} = 5\%$	$\frac{0,9 \times (100 \times 1,072 - 95 \times 1,05) - 5}{5} = 34,5\%$

eventuelle tab. Dermed er der ingen markedsmechanisme til at modvirke bankernes incitament til at vælge en meget høj gearing med deraf følgende høj risiko for fallit.

Myndighederne er naturligvis opmærksomme på dette problem og har derfor igennem mange år stillet krav til banker om, at deres aktiver er bakket op af en vis minimal egenkapitalandel. Spørgsmålet er derfor, hvordan balancen mellem de omtalte skæve incitamenter og bankreguleringen kunne tippe, så vi i sidste årti fik en finanskrisen af hidtil uset omfang. Forklaringen skal søges i et samspil af mange komplekse faktorer, men de fleste er enige om, at især tre forhold spillede en vigtig rolle.¹ For det første fremvoksede der især i USA et omfattende såkaldt skyggebanksystem, der ikke var underkastet den traditionelle bankregulering med minimumskrav til egenkapitalen. Mange banker oprettede nye investeringsselskaber eller stillede garantier for sådanne selskaber, der specialiserede sig i investering i forskellige nye typer af værdipapirer. Værdipapirerne var i stort omfang udstedt på basis af lån, som blev bevilget men efterfølgende frasolgt af de traditionelle banker. En stadig større del af den samlede kreditgivning blev således kanaliseret via værdipapirmarkederne uden om det traditionelle banksystem.

De store amerikanske investeringsbanker som f.eks. Lehman Brothers var en del af

1. En mere udførlig dansk fremstilling af årsagerne til finanskrisen kan findes i Berg og Bech (2009).

skyggebanksystemet og opererede med egenkapitalprocenter på 3-4 % af deres samlede aktiver. En egenkapitalprocent på 4 betyder, at hver gang der investeres for en dollar aktiekapital, bliver der investeret for 24 dollars af lånte midler. I den situation skal værdien af de erhvervede aktiver kun falde med 4%, før egenkapitalen er udraderet, og långiverne begynder at lide tab. Der blev således løbet store risici, tilskyndet af de meget høje forventede afkaststrater, som denne ekstreme grad af gearing typisk indebærer. Hvis investeringerne gik godt, scorede aktionærerne store gevinster, og hvis de gik skidt, var det primært långiverne eller skatteyderne, der måtte bære tabet.

For det andet opbyggede mange banker i årene op til finanskrisen et stigende indlånsunderskud, dvs. deres udlån og investeringer oversteg i stigende grad deres indlån. Indlånsunderskuddet var typisk finansieret via kortfristede lån og gældspapirer, mens provenuet af disse lån fortrinsvis var placeret i mere langfristede og dermed mindre likvide aktiver. Bankerne blev derved i stigende omfang afhængige af, at andre investorer og finansielle institutioner var villige til løbende at refinansiere deres kortfristede gæld. Da denne mulighed for refinansiering pludselig faldt bort, som det skete under finanskrisen, blev nogle banker tvunget til at foretage brandudsalg af deres aktiver. Det førte til drastiske kursfald på mange værdipapirer, hvorved andre banker blev tvunget til at nedskrive værdien af deres aktiver og dermed kom i vanskeligheder.

Bankernes grundlæggende forretningsmodel har altid været at omdanne likvide kortfristede indlån til mindre likvide langfristede udlån og investeringer. Denne transformation af korte indlån til lange udlån er overordentlig nyttig og vigtig for det økonomiske system, men den indebærer selvsagt også en stor risiko. Hvis tilliden til en bank svigter, og indskyderne ikke er dækket af en garantiordning, vil de løbe storm mod banken for at trække deres penge ud. Dermed går banken ned, da den ikke med øjeblikkeligt varsel kan omsætte sine langfristede aktiver til likvider. De førømtalte problemer med manglende information hos indskydere og långivere kan i ugunstige situationer betyde, at selv fornuftigt drevne solvente banker går fallit på grund af likviditetsmangel, hvis de udsættes for et stormløb, der udspringer af manglende tillid. Denne dyrekøbte erfaring førte efter 1930'ernes store depression til, at de fleste lande indførte generelle indskydergarantiordninger.

Men i årene op til finanskrisen var det, som om man havde glemte problemet med den iboende ustabilitet i bankernes forretningsmodel. De store indlånsunderskud i mange banker var jo netop finansieret af kortfristet gæld, der ikke var sikret af garantiordninger. Da tilliden svigtede under finanskrisen, oplevede banker med indlånsunderskud derfor en ny form for stormløb, idet de pludselig ikke kunne få fornyet de kortfristede lån, de var vant til at optage hos andre banker og finansielle institutioner. Hvor de klassiske bankkriser tidligere i historien typisk bestod i, at almindelige borgere og virksomheder trak deres penge ud af bankerne, så oplevede vi under finanskrisen

i efteråret 2008, at det finansielle system foranstaltede et stormløb mod sig selv. Alle finansielle institutioner holdt sig tilbage fra at udlåne til hinanden, da de frygtede selv at komme til at mangle likviditet. De grundlæggende mekanismer var imidlertid de samme som under et klassisk bank run. Mekanismerne er beskrevet i et fremragende videnskabeligt bidrag af Douglas Diamond og Philip Dybvig (1983), der både påviser de samfundsmæssige fordele ved bankernes transformation af korte indlån til lange udlån og samtidig demonstrerer, hvordan tillidssvigt kan føre til skadelige bank runs, som kan forebygges af indskydergarantiordninger. Da finanskrisen nåede sit klimaks, måtte staterne træde til med omfattende garantier for alle indskydere og kreditorer for at redde det finansielle system fra det totale sammenbrud. De udvidede garantiordninger var en nødvendig nødløsning i den konkrete situation, men de har også forstærket det såkaldte moral hazard problem, der består i, at garantier mod tab eliminerer indskydernes og långivernes tilskyndelse til at overvåge, om bankerne investerer pengene ansvarligt. Som fremhævet i et vigtigt bidrag fra John Kareken og Neil Wallace (1978) kan moral hazard problemet i sig selv gøre banksystemet mere risikobetonet og ustabil.

Et tredje væsentligt forhold, som viste sig at øge ustabiliteten i det finansielle system i årene op til finanskrisen, var fremkomsten af stadig mere komplicerede typer af værdipapirer, hvis risikoegenskaber var svære at gennemskue. De såkaldte asset-backed securities, dvs. værdipapirer udstedt med basis i boliglån og andre former for lån, blev samlet i puljer, som dannede grundlag for udstedelse af nye værdipapirer, der igen blev samlet i puljer som basis for en ny runde af udstedelser, og så fremdeles. Formålet med disse komplicerede konstruktioner var at mindske den risiko, der knyttede sig til det enkelte værdipapir. Puljerne af værdipapirer blev også inddelt i forskellige risikoklasser. Ideen var, at hvis der opstod tab på de oprindelige bagvedliggende udlån, så skulle tabene i første omfang bæres af ejerne af obligationerne i den mest udsatte risikoklasse, inden obligationsejerne i den næstmest udsatte risikoklasse kunne begynde at lide tab, osv. På den måde mente man, at obligationerne i de mindre udsatte risikoklasser var meget sikre papirer, selvom alle obligationerne f.eks. var udstedt på basis af subprime-lån til mindre kreditværdige boligejere med tvivlsom betalings-evne. Man forestillede sig ikke, at et meget stort antal låntagere kunne komme i vanskeligheder på samme tid, således at selv obligationsejerne i de mest gunstigt stillede risikoklasser ikke kunne få deres tilgodehavender. Men det var netop, hvad der skete, da boligpriserne vendte overalt i USA, så subprime-låntagere ikke længere kunne finansiere tilbagebetaling af gamle boliglån ved at optage nye lån.

Formodningen var altså, at man via de nye såkaldte strukturerede værdipapirer havde fået spredt risikoen mere effektivt ud i det finansielle system, så risikoen for den enkelte investor var blevet mindre. De nye syntetiske værdipapirer var imidlertid så

komplicerede, at de færreste var i stand til at gennemskue, hvilke pengestrømme de var baseret på, og hvor store sandsynlighederne var for, at de pågældende pengestrømme faktisk ville materialisere sig. Den vurdering overlod man til de store kreditvurderingsbureauer, hvilket skulle vise sig at være en skæbnesvanger fejl. Da finanskrisen brød ud, blev mange værdipapirer, som rating-bureauerne havde vurderet som helt sikre, pludseligt nærmest værdiløse. Det medvirkede selvsagt til at forstærke den finansielle panik. Hvordan kunne rating-bureauerne tage så meget fejl? En af forklaringerne er formentlig, at erfaringerne med og dermed det statistiske grundlag for at vurdere de mange nye komplicerede finansielle produkter var meget begrænsede.

Men en økonom vil også pege på, at den grundlæggende konstruktion omkring rating-bureauerne indebærer et incitamentsproblem, som kan tilskynde dem til at være overoptimistiske i deres vurderinger. Rating-bureauerne betales af de institutioner, der udsteder de værdipapirer, som skal vurderes. Disse institutioner er naturligvis interesserede i, at papirerne får en god rating, så de kan sælges til en høj pris. I forbindelse med konstruktionen af de mange nye strukturerede værdipapirer baseret på f.eks. subprime-lån skete det ofte, at udstederne af papirerne på forhånd betalte rating-bureauerne for at rådgive om, hvordan de bagvedliggende puljer af lån og aktiver skulle kombineres, således at det planlagte nye strukturerede værdipapir fik den højest mulige rating. Efterfølgende blev det selvsamme rating-bureau så bedt om – og betalt for – at afgive en rating af det nye værdipapir, det selv havde været med til at konstruere. Ikke overraskende viste denne rating sig typisk at være god.

Hermed være ikke sagt, at rating-bureauerne altid overvurderer de aktiver, de skal bedømme. Bureauerne har naturligvis også en interesse i at beskytte deres renommé og sikre deres troværdighed. Men det kræver evne og vilje til langsigtet tænkning, hvis denne disciplinerende mekanisme skal virke, og langsigtet tænkning var ikke i høj kurs i årene op til finanskrisen.

Ifølge den historie, jeg har fortalt, skyldtes finanskrisen grundlæggende en mangelfuld offentlig regulering af det finansielle system. Men hvorfor svigtede reguleringen? Det er der flere forklaringer på, men den væsentligste årsag er nok, at der også i det politiske system er et incitamentsproblem. Hvis et enkelt land strammer sin finansielle regulering, vil det mindske indtjeningsmulighederne i landets finanssektor. Da mange finansielle aktiviteter er særdeles mobile over landegrænserne, kan en strammere national regulering let føre til, at en masse gode højtlønnede job flytter ud af landet. Det betyder lavere indtjeningsmuligheder for andre lønmodtagere og virksomheder i landet og lavere skatteindtægter. Den internationale konkurrence om at tiltrække velaflønnede mobile finansielle aktiviteter medfører derfor, at det enkelte land vil være tilbageholdende med at indføre en stram finansiell regulering. Problemet skærpes i takt med, at globaliseringen skrider frem, og de finansielle aktiviteter bliver stadig mere

mobile. Resultatet kan blive en såkaldt »dårlig ligevægt«, hvor alle lande egentlig gerne ville have en strammere regulering, men hvor ingen tør gå forrest af frygt for at miste indtjeningsmuligheder til andre lande. Dette problem er velbeskrevet i den økonomiske litteratur, jf. f.eks. Sinn (2003). Det kræver stærke internationale, politiske institutioner at dæmme op for problemet og sikre en effektiv regulering.

Lad mig opsummere, hvad jeg ovenfor har sagt om de mikroøkonomiske årsager til finanskrisen. I en situation med ufuldstændig og asymmetrisk information indebærer aktieselskabsformen med dens begrænsede ansvar for aktionærerne et incitament til overdreven låntagning og risikotagning. Informationsproblemet og det deraf følgende skæve incitament er særligt udtalt i banksektoren på grund af den høje kompleksitet af de aktiver, som bankerne investerer i. Samtidig ligger der en stor indbygget risiko i bankernes grundlæggende forretningsmodel, som bygger på en potentielt ustabil ligevægt, hvor kortfristede indlån omdannes til langfristede udlån. Ustabiliteten har historisk manifesteret sig i gentagne ødelæggende bank runs, som har ført til indførelsen af indskydergarantiordninger. Garantiordninger fjerner beklageligvis indskydernes og långivernes tilskyndelse til at overvåge, at bankerne forvalter deres midler ansvarligt. Det har derfor været nødvendigt at stille regulatoriske minimumskrav til størrelsen af bankernes egenkapital. Disse krav reducerer imidlertid bankaktionærernes forventede afkast og giver dem et incitament til at kanalisere kreditgivning over i et skyggebanksystem, der ikke er omfattet af kapitalkravene. Det var netop, hvad der i stort omfang skete i årene op til finanskrisen. Den ekstreme gearing i det voksende skyggebanksystem kombineret med et stigende indlånsunderskud i mange banker betød sammen med den stigende kompleksitet og deraf følgende uoverskuelighed i det finansielle system, at systemet blev stadig mere sårbart over for ubehagelige overraskelser. Derfor kunne problemer på det forholdsvis begrænsede marked for amerikanske subprime-lån bringe hele det globale finansielle system på randen af det totale sammenbrud.

Denne skitse til forklaring af finanskrisen er på mange måder forenklet, men jeg håber at have illustreret, at den mikroøkonomiske teori med dens fokus på betydningen af informations- og incitamentsproblemer har gode bud på de grundlæggende årsager til, at krisen opstod og blev så alvorlig.

Også den gængse makroøkonomiske teori kan bidrage til forklaringen. I makroteorien fokuserer man især på årsagerne til, at kapitalistiske markedsøkonomier gennemløber tilbagevendende konjunkturcykler. I et berømt bidrag foreslog makroøkonomen John Taylor (1993) en regel for den amerikanske centralbanks rentepolitik, som ifølge hans analyse ville bidrage til at mindske udsvingene i den økonomiske aktivitet og inflationen. Ifølge denne såkaldte Taylor-regel bør centralbanken hæve realrenten op over sit normale niveau, hvis inflationen overstiger det tilstræbte niveau, eller hvis den

samlede produktion er højere end i en normal konjunktursituation. Omvendt bør centralbanken sænke realrenten under det normale niveau, når inflationen er mindre end tilstræbt, eller når produktionen er lavere end svarende til en normal konjunktursituation. Centralbankerne i USA og en række andre lande førte faktisk i mange år en rentepolitik, der var i god overensstemmelse med Taylor-reglen.

I en periode forud for finanskrisen holdt den amerikanske centralbank imidlertid den pengepolitiske rente noget under det niveau, som en konsekvent anvendelse af Taylor-reglen ville have tilsagt. Lavrentepolitikken blev indledt for at imødegå virkningerne af, at den såkaldte dot.com boble i aktiekurserne var bristet i år 2000, og af frygt for et tilbageslag i økonomien efter terrorangrebet den 11. september 2001. Det lave renteniveau bidrog imidlertid til at puste de amerikanske boligpriser op og stimulerede til et kraftigt byggeboom, som hen ad vejen øgede boligudbuddet og forstærkede det boligprisfald, der indtraf fra 2006. På basis af en nylig økonometrisk analyse har John Taylor (2009) hævdet, at udsvingene på det amerikanske boligmarked ville have været langt mindre, hvis Federal Reserve havde fulgt hans pengepolitiske regel. Andre iagttagere har lagt mere vægt på betydningen af de nye låneformer og den meget lemfældige kreditgivning forud for sammenbruddet, og atter andre har fremhævet, at de store opsparingsoverskud i Kina og en række andre asiatiske lande bidrog til at presse de lange renter usædvanligt langt ned i sidste årti. Der er dog næppe tvivl om, at den amerikanske centralbanks lavrentepolitik i perioden fra 2002 til 2005 bidrog til prisboblen på det amerikanske boligmarked og dermed til den efterfølgende finanskriser.

Både den mikroøkonomiske og den makroøkonomiske teori kan altså bidrage til at forklare finanskrisen. Jeg er derfor uenig med dem, der ser finanskrisen som et bevis for en fundamental krise i den økonomiske videnskab, jf. f.eks. Collander m.fl. (2009). Man kan tværtimod argumentere for, at finanskrisen i stort omfang skyldtes, at myndighederne ikke fulgte de principper for mikroøkonomisk regulering og makroøkonomisk stabiliseringspolitik, som hovedstrømmen af økonomisk teori foreskriver.

På den anden side er det naturligt og særdeles nyttigt, at finanskrisen har givet anledning til kritisk eftertanke i det fagøkonomiske miljø. Krisen har f.eks. intensiveret den faglige debat om relevansen af teorien om effektive finansielle markeder. Denne teori siger, at priserne på finansielle aktiver altid afspejler al relevant tilgængelig information om de fremtidige indtægtsstrømme, som aktiverne kan ventes at kaste af sig. Det betyder f.eks., at aktiekurserne altid afspejler den mest ædruelige vurdering af, hvor store fremtidige udbytter aktieselskaberne kan forventes at udbetale. Hypotesen indebærer, at den enkelte investor ikke skal gøre sig håb om at kunne »slå markedet«. Hvis man investerer ud fra en formodning om, at prisen på et aktiv er systematisk over- eller undervurderet, vil man blive skuffet: Aktiver har altid den pris, de skal have, givet det man ved på det pågældende tidspunkt.

Skeptikere såsom Robert Shiller (2003) har indvendt, at aktiekurserne synes at være langt mere volatile, end man skulle vente, hvis kurserne altid afspejler det bedste bud på de fremtidige aktieudbytter. Kan det virkelig være rigtigt, at forventningerne til de fremtidige udbytter svinger så meget som aktiekurserne, spørger han? Hertil svarer advokaterne for tesen om effektive markeder, at aktiekurserne udmærket kan svinge mere end de forventede udbytter, hvis der er hyppige udsving i investorernes villighed til at påtage sig risiko.

For en del år siden rejste Sanford Grossman og Joseph Stiglitz (1980) imidlertid en mere fundamental indvending mod teorien: Hvis de observerede aktivpriser altid korrekt afspejler al tilgængelig information, og ingen investor skal tro, at han er klogere end markedet, hvem vil så overhovedet påtage sig besværet med at indsamle den information, der er nødvendig for en korrekt værdiansættelse af de forskellige aktiver? Hvordan kan aktivpriserne afspejle al relevant information, når ingen har incitament til at fremskaffe den? Der må være nogle investorer, der arbejder på at skaffe sig information i håb om, at de kan tjene penge ved at købe aktiver, som forekommer at være undervurderede, og ved at sælge dem, der forekommer at være overvurderede.

Problemet for investorerne er, at fremtiden er fundamentalt usikker. Ingen kan derfor vide med sikkerhed, hvad den »sande« værdi af et aktiv er. Investorernes grad af tillid til fremtiden bliver dermed helt afgørende for aktivprisernes udvikling, som John Maynard Keynes (1936) allerede for mange år siden gjorde opmærksom på. Set i bagklogskabens ulideligt klare lys er det ikke svært at udpege perioder i den økonomiske historie, hvor udviklingen i f.eks. aktiekurser og boligpriser har været præget af en overdreven optimisme. Årene forud for den seneste finanskriser forekommer at være en sådan periode. Set i bakspejlet kan man også finde perioder, hvor aktivpriserne synes at have været drevet af urimeligt pessimistiske forventninger, jævnfør f.eks. de meget drastiske kursfald, der indtraf i kølvandet på den nylige krise. Det betyder dog ikke, at aktivprisernes udvikling drives af helt arbitrære forhold. Som Roman Frydman og Michael Goldberg (2011) har dokumenteret i en netop udkommen bog, påvirkes f.eks. aktiekurserne i meget høj grad af nye informationer, der må anses for relevante i en vurdering af virksomhedernes fremtidige indtjening. Disse forfattere påviser også, at selv om aktivpriserne gennemløber store og langvarige udsving, så har de en tendens til at vende tilbage til, hvad der opfattes som et normalt langtidsholdbart niveau. Frydman og Goldberg fremhæver dog samtidigt, at markedsdeltagerne over tid ændrer den vægt, som de tillægger de forskellige økonomiske forhold og nøgletal, når de danner deres forventninger til de fremtidige afkast af forskellige aktiver. Derfor skal man ikke gøre sig håb om at kunne beskrive forventningsdannelsen ved simple, mekaniske og uforanderlige beslutningsregler.

Det gør en væsentlig forskel, om man er tilhænger af tesen om effektive finansielle markeder, eller om man hælder til det mere pragmatiske syn på markederne, som f.eks. Frydman og Goldberg står for. Teorien om effektive markeder lægger op til, at myndighederne ikke skal bekymre sig om, hvorvidt markederne er over- eller undervurderede og dermed ikke bør søge at påvirke aktivpriserne i en bestemt retning. Hvis man er mere skeptisk over for markedernes evne til at sætte de »rigtige« aktivpriser, kan det være et argument for, at den økonomiske politik skal forsøge at læne sig mod vinden på f.eks. aktie- og boligmarkederne. Med andre ord kan det være et argument for, at den økonomiske politik søger at dæmpe de meget store udsving i aktivpriserne, som historisk har været med til at underminere den finansielle og makroøkonomiske stabilitet.

Teorien om effektive finansielle markeder er et eksempel på en anvendelse af den mere generelle hypotese om rationelle forventninger, som blev introduceret i makroøkonomisk teori af Robert Lucas (1972). Denne tese siger, at de økonomiske aktører ikke begår systematiske fejl, når de danner deres forventninger til den fremtidige udvikling i de økonomiske variable, der er relevante for dem. Når folk f.eks. gætter på den fremtidige inflation eller den fremtidige udvikling i boligpriserne, vil de altså ikke gennem længere tid systematisk overvurdere eller undervurdere prisudviklingen. I formelle økonomiske modelanalyser indebærer tesen om rationelle forventninger, at de økonomiske aktører danner deres forventninger, som om de kendte de økonomiske sammenhænge, der er indbygget i modellen.

Tesen om rationelle forventninger er den dominerende hypotese om forventningsdannelsen i moderne makroøkonomisk teori. Tesen kan være nyttig, når man vil analysere, om en bestemt økonomisk politik kan forventes at have den ønskede virkning. Hvis politikken kun virker, når husholdninger og virksomheder begår systematiske forventningsfejl, kan der være grund til at betvivle, hvor robuste dens effekter er. Mennesker kan jo lære af deres fejl, så en politik, der f.eks. kun virker, hvis folk systematisk undervurderer inflationen, er næppe langtidsholdbar.

Mere generelt kan tesen om rationelle forventninger være et nyttigt udgangspunkt i mange økonomiske analyser. Det har interesse at vide, hvordan det økonomiske system virker, når den økonomiske modelbygger ikke er klogere end de individer, hvis adfærd han søger at beskrive. På den anden side er det ikke nemt at forene tesen om rationelle forventninger med forløbet op til og under finanskrisen. I midten af sidste årti blev der f.eks. foretaget flere interviewundersøgelser, som viste, at de danske husholdninger forventede betydelige boligprisstigninger over en længere periode. Undersøgelserne pegede klart på, at forventningerne til den fremtidige udvikling var stærkt præget af de meget kraftige boligprisstigninger, der fandt sted i 2005 og 2006, jf. Dam m.fl. (2011). Disse prisstigninger blev i stort omfang ekstrapoleret ind i fremtiden,

selvom de fra en nøgtern betragtning måtte forekomme at være løsrevet fra de underliggende økonomiske forhold. Lignende undersøgelser fra udlandet peger ligeledes i retning af, at forventningerne til fremtiden ofte er domineret af de nærmest foregående års erfaringer.

I mange sammenhænge kan det derfor være vildledende at basere sig på tesen om rationelle forventninger. Selvom tesen altså ikke bør skrottes helt, er der et klart behov for at supplere den med mere realistiske, men samtidig operationelle teorier om forventningsdannelsen. Heldigvis arbejdes der på mange fronter med denne problemstilling i den aktuelle økonomiske forskning.

Finanskrisen har også ført til, at der arbejdes på livet løs med at indbygge mere realistiske beskrivelser af den finansielle sektor i de makroøkonomiske modeller. Et godt og pædagogisk bud på, hvordan dette kan gøres, er f.eks. for nylig givet af Michael Woodford (2010), men han suppleres af en hærskare af andre makroøkonomer.

Selv om krisen har ført til skærpede kontroverser inden for den økonomiske videnskab, er jeg optimistisk på mit fags vegne. Uenighederne vil vise sig at være produktive og vil føre til nye erkendelser, ligesom 1930'ernes Store Depression blev startskuddet til en usædvanligt frugtbar periode for den økonomiske forskning. Vi ser allerede nu, hvordan det vælter frem med nye videnskabelige bidrag inspireret af begivenhederne under finanskrisen.

Vi kunne sandt for dyden godt have undværet krisen, men når den nu kom, er det rart at vide, at der arbejdes på at uddrage, hvad vi kan lære af den.

Litteratur

- Adrian, T. og H. Song Shin. 2009. The shadow banking system: implications for financial regulation. *Federal Reserve Bank of New York Staff Report* No. 382.
- Berg, J. og M. L. Bech. 2009. *Finansernes Fald*. Gyldendal.
- Collander, D., M. Goldberg, A. Haas, K. Juselius, A. Kirman, T. Lux og B. Sloth. 2009. The financial crisis and the systemic failure of the economics profession. *Critical Review: A Journal of Politics and Society*, vol. 21, no. 2 og 3, 249-67.
- Dam, N. A., T. Saaby Hvolbøl, E. Haller Pedersen, P. Birch Sørensen og S. Hougaard Thamsborg. 2011. Udviklingen på ejerboligmarkedet i de senere år – Kan boligpriserne forklares? Danmarks Nationalbank, *Kvartalsoversigt*, 2. kvartal 2011, Del 2: 1-79.
- Diamond, D. og P. Dybvig. 1983. Bank runs, deposit insurance, and liquidity. *Journal of Political Economy*, vol. 91, no. 3, 401-19.
- Frydman, R. og M. Goldberg. 2011. *Beyond Mechanical Markets – Asset Price Swings, Risk, and the Role of the State*. Princeton University Press, Princeton and Oxford.
- Grossman, S. J. og J. E. Stiglitz. 1980. On the impossibility of informationally efficient markets. *American Economic Review*, vol. 70, 393-407.
- Kareken, J. H. and N. Wallace. 1978. Deposit insurance and bank regulation: a partial-equilibrium exposition. *Journal of Business*, vol. 51, 413-38.
- Keynes, J. Maynard. 1936. *The General Theory of Interest, Employment and Money*. Macmillan.
- Krugman, P. 2009. How did economists get it so wrong? *The New York Times*, September 6, 2009. (<http://www.nytimes.com/2009/09/06/magazine/06Economic-t.html>)

- Lucas, R. E. 1972. Expectations and the neutrality of money. *Journal of Economic Theory*, vol. 4, 103-24.
- Modigliani, F. og M. H. Miller. 1958. The cost of capital, corporation finance, and the theory of investment. *American Economic Review*, vol. 48, 261-97.
- Shiller, R. J. 2003. From efficient markets theory to behavioural finance. *Journal of Economic Perspectives*, vol. 17, Winter 2003, 83-104.
- Sinn, H.-W. 2010. *Casino Capitalism*. Oxford University Press.
- Sinn, H.-W. 2003. *The New Systems Competition*. Basil Blackwell, Oxford.
- Taylor, John B. (1993). Discretion versus policy rules in practice. *Carnegie-Rochester Conference Series on Public Policy*, vol. 39, 195-214.
- Taylor, J. B. 2009. The financial crisis and the policy responses: an empirical analysis of what went wrong. National Bureau of Economic Research Working Paper No. 14631, January 2009.
- Woodford, M. 2010. Financial intermediation and macroeconomic analysis. *Journal of Economic Perspectives*, vol. 24, No. 4, Fall 2010, 21-44.