

Arbejdsrapport nr.: 22/2010

En panelmodel for timeløn i Danmark: Ny metode til imputering af skyggelønninger i Finansministeriets forskelsbeløbsmodel

Lars Grønvall Foldspang og Mads Vej Andersen* **

Resumé:

En vurdering af tilskyndelsen til at arbejde kan foretages ved beregning af forskelsbeløb og kompensationsgrader. Beregninger forudsætter oplysninger om lønindkomst. Ca. 8 pct. har ingen lønoplysninger, bla. som følge af ledighed. For at give et retvisende billede imputeres oplysninger for disse personer. Dette papir præsenterer en nyudviklet panelmodel som via en såkaldt *random effects* estimation og efterfølgende simulation kan anvendes til at estimere en alternativ arbejdsindkomst for de personer, hvor der ikke foreligger registeroplysninger. Metoden forbedrer kvaliteten af de imputerede lønoplysninger, som anvendes i de videre beregninger af tilskyndelsen til at arbejde.

* Forfatterne ønsker at rette en meget stor tak til chefkonsulent Hans Bækgaard for hjælp med simulationer, udfærdigelse af appendiks B og uundværlig sparring gennem hele forløbet. Udledningerne i appendiks B er baseret på Bækgaard, H. (2010) *Work incentives and latent wages for the unemployed: estimating a dynamic panel model for wage rates with missing observations*, arbejdsrapport præsenteret ved det 9. Nordiske Mikrosimulationsseminar i København d. 16. og 17. juni 2010.

** Synspunkter udtrykt i dette papir er forfatterens egne og stemmer ikke nødvendigvis overens med Finansministeriets.

1. Indledning

Dette papir beskriver den nye metode, Finansministeriet benytter til at estimere og imputere den såkaldte *skyggeløn* for personer, hvor der ikke foreligger tilgængelige timelønsoplysninger. For ca. 8 pct. af den personkreds, som indgår i Finansministeriets beregninger af forskelsbeløb, findes ingen anvendelige lønoplysninger.

Skyggeløn er et udtryk for den lønindkomst, en ledig ville have fået, hvis vedkommende var i job. Skyggelønnen kan derfor benyttes til at beregne, hvor stor en ændring af rådighedsbeløbet – dvs. hvor stort et forskelsbeløb – ledige personer ville opnå, hvis de kom i job.

Metodeændringen har konsekvenser for de beregnede forskelsbeløb og kompensationsgrader, og derfor også for de skøn for arbejdsudbudsvirkninger, der udarbejdes på baggrund af modellen, eksempelvis til vurdering af politiske tiltag.

Afsnit 2 og 3 giver en introduktion af Finansministeriets Forskelsbeløbsmodel samt en beskrivelse af konsekvenserne af metodeændringen for forskelsbeløb mv. Afsnit 4 beskriver den nye metode til at estimere timelønninger baseret på paneldata fra 1995 til 2006. Data beskrives i afsnit 5, mens afsnit 6 viser nogle grundlæggende egenskaber ved metoden. Afsnit 7 og 8 diskuterer den anvendte metode, mens metoden til simulering af skyggelønnen gennemgås i afsnit 9.

2. Skyggelønningernes anvendelse: Forskelsbeløbsmodellen

Finansministeriet anvender den såkaldte Forskelsbeløbsmodel baseret på registerdata fra Danmarks Statistik til at anslå, hvor store ændringer i rådighedsbeløbet, beskæftigelse frem for ledighed medfører for forskellige persongrupper.

Forskelsbeløb og kompensationsgrader opgøres med udgangspunkt i en sammenligning af personernes disponible indkomst i to situationer. I den ene situation er personerne i fuldtidsbeskæftigelse, og i den anden situation modtager personerne indkomsterstattende overførselsindkomst, eksempelvis arbejdsløshedsdagpenge eller kontant-/starthjælp. For beskæftigede kan lønindkomst opgøres ved hjælp af registeroplysninger, og potentiel overførselsindkomst kan beregnes på baggrund af, hvorvidt personen er berettiget til dagpenge, kontanthjælp mv. (herunder forsikringsstatus). Ligeledes er det muligt for visse ledige at opgøre den potentielle lønindkomst via dagpengeregistre mv. For disse personer kan den økonomiske tilskyndelse til beskæftigelse derfor opgøres relativt præcist. For de personer, hvor der ikke forefindes oplysninger om lønindkomst, *jf. boks 1*, er det nødvendigt at fastsætte et bedste skøn for lønnen, også kaldet en *imputeret skyggeløn*.

Boks 1**Manglende oplysninger i lønregistrene**

Der kan være flere årsager til, at en person ikke har en lønoplysning fra lønregistrene i det relevante år. Hovedårsagen er, at vedkommende er ledig (eller på efterløn) hele året, og derfor ikke har haft lønarbejde. En anden årsag kan være, at arbejdsgiver ikke indberetter til lønregistret. Hvis virksomheden består af mindre end 10 ansatte, er der ikke indberetningspligt. En del af arbejdsstyrken har derfor ikke tilgængelige lønoplysninger i lønregistrene. Derudover forefindes ingen oplysninger fra landbrug eller fiskeri.

I disse tilfælde er det nødvendigt at søge oplysninger andre steder, eksempelvis i sygedagpenge- eller dagpengeregisteret, eller i indkomstregisteret, som indeholder alle modtagne indkomster i løbet af året. Indkomstregisteret er dog den dårligste kilde til at fastlægge en normeret årsindkomst ved fuldtidsbeskæftigelse, da der ikke er nøjagtige oplysninger om arbejdstiden i registret.

Metodeændringen er gennemført for at forbedre kvaliteten af den imputerede løn for personer, som ikke har en registreret timeløn i året (timelønnen anvendes i beregningerne til at bestemme en samlet årsindkomst ved beskæftigelse). Det drejer sig om godt 65 pct. af de fuldt ledige, omtrent 35 pct. af efterlønsmodtagerne samt en lille restgruppe af beskæftigede, *jf. tabel 1*, hvor der af forskellige årsager ikke findes en registreret løn. Samlet udgør disse personer knap 8 pct. (ca. 6.400 observationer) af den population, der indgår i beregningerne (82.000 observationer, svarende til gruppen af 18-64 årige fraregnet selvstændige, studerende og førtidspensionister).

Tabel 1**Andel med lønoplysninger fra de forskellige kilder til timeløn, fordelt på beskæftigelsesstatus, 2005**

	Løn- registre	Sygedag- penge	Dag- penge	Tidligere IR	Tidligere år*	Ingen løn- oplysning	Alle
	Andel, pct.						
Fuldt besk.	72,0	5,3	0,6	21,2	.	0,9	100
Delvist besk.	55,2	13,3	17,4	9,3	.	4,8	100
Fuldt ledige	2,1	6,2	25,0	.	.	66,7	100
Efterlønsmodtagere	0,3	.	.	.	64,0	35,7	100
Alle	61,3	6,2	4,5	17,1	3,1	7,8	100

Anm.: "Lønregistre" svarer til Danmarks Statistiks lønregistre for den private-, kommunale og statslige sektor. "IR" = Indkomstregisteret.

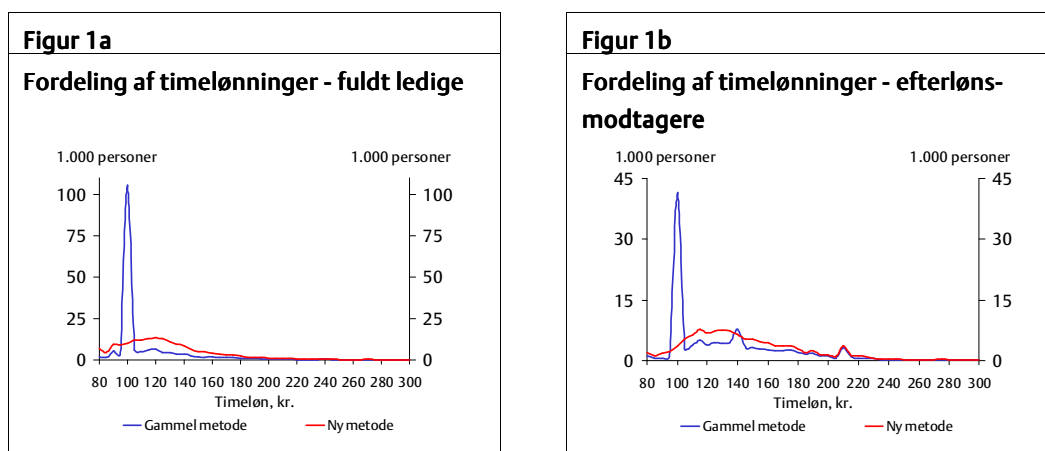
* De efterlønsmodtagere, der får tildelt tidligere års løn (fremskrevet til det relevante år) er overgået fra beskæftigelse til efterløn.

Kilde: Egne beregninger på baggrund af en stikprøve på 3,3 pct. af befolkningen.

Metoden beskrevet i afsnit 4 og de følgende afsnit anvendes til at imputere en skyggeløn for hele denne gruppe.

3. Konsekvenser af metodeændringen

De imputerede timelønninger beregnet ved hjælp af den nye metode giver betydeligt mere variation i lønfordelingen for ledige og efterlønsmodtagere end den hidtidige metode, jf. figur 1a og 1b. Det skyldes, at der tidligere har været anvendt en fast lav løn svarende til 5 pct.-fraktilen i de beskæftigedes lønfordeling (100,30 kr. i 2005). Derfor fås en ”sammenklumpning” af forskelsbeløb og kompensationsgrader, og således potentielt et fejlskøn over antal personer med lave forskelsbeløb og høje kompensationsgrader. De to toppe i figurerne viser, at der med den ”gamle metode” er et stort antal personer, der tildeles samme løn svarende til 5 pct.-fraktilen i de beskæftigedes lønfordeling.



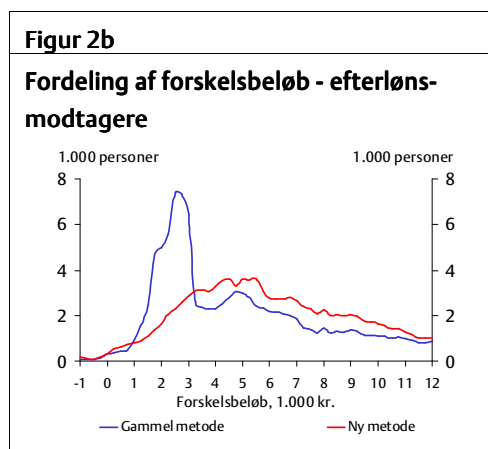
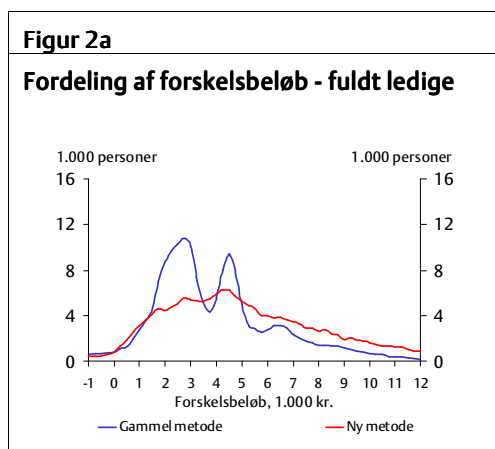
Kilde: Egne beregninger på baggrund af en stikprøve på 3,3 pct. af befolkningen.

Den nye metode giver en lønspredning, der er større for gruppen af ledige end den tidligere metode. Således ligger de midterste 50 pct. inden for et interval på 15 kr. med den tidligere metode, mens intervallet er 38 kr. med den nye metode. For efterlønsmodtagerne er spredningen omtrent den samme i de to opgørelser. Den gennemsnitlige løn for ledige øges fra 113 til 128 kr. og for efterlønsmodtagere fra 136 til 148 kr. Det svarer til en stigning på mellem 9 og 13 pct. for de to grupper og skyldes en mere realistisk fordeling af de imputerede lønninger sammenlignet med den tidligere metode.

Konsekvens for forskelsbeløb

For gruppen af fuldt ledige og gruppen af efterlønsmodtagere medfører den nye metode en mere jævn fordeling af forskelsbeløb end den hidtidige metode, jf. figur 2a og 2b. Fordelingerne for fuldt og delvist beskæftigede er omtrent uændrede, da en meget lille andel tildeles en imputeret løn (det vil sige, at lønnen i stedet er fastlagt ud fra registeroplysninger).

Den to-puklede fordeling for de fuldt ledige, som man finder med den ”gamle metode”, kan henføres til, at der med den hidtidige metode imputeres en løn svarende til 5 pct.-fraktilen i de beskæftigedes lønfordeling for en stor del af de fuldt ledige samtidig med, at nogle af de fuldt ledige er berettiget til dagpenge (puklen med forskelsbeløb omkring 2.500 kr.), mens andre alene er berettiget til kontanthjælp (puklen med forskelsbeløb omkring 4.500 kr.). Med den nye metode fjernes puklerne som følge af et ændret skøn for skyggelønnen.



Kilde: Egne beregninger på baggrund af en stikprøve på 3,3 pct. af befolkningen.

Det gennemsnitlige forskelsbeløb for gruppen af ledige samt gruppen af efterlønsmodtagere stiger som følge af metodeændringen. Stigningen er på 1.100-1.200 kr. i snit, *jf. tabel 2*.

For gruppen af ledige stiger forskelsbeløbet fra 4.100 til 5.300 kr., hvilket svarer til en stigning på knap 30 pct. Stigningen skal ses i lyset af, at lønnen imputeres med udgangspunkt i de enkelte personers evt. tidligere løn samt andre karakteristika, herunder uddannelse, som har en vis spredning. De hidtil anvendte lønninger, anslået som 5 pct.-fraktile i lønfordelingen for de beskæftigede, er derimod ikke baseret på personlige karakteristika.

Tabel 2							
Gennemsnitligt forskelsbeløb og kompensationsgrad i 2009-niveau							
	Antal	Forskelsbeløb			Kompensationsgrad		
		Ny	Gammel	Forskel	Ny	Gammel	Forskel
	Personer	-----Kroner-----			-----Pct.-----		Pct.-enh.
Fuldt besk.	1.795.000	8.000	7.900	100	57,8	58,0	-0,2
Delvist besk.	316.000	5.000	5.000	0	66,8	67,1	-0,3
Fuldt ledig	173.000	5.300	4.100	1.200	63,3	68,0	-4,7
Efterløn	117.000	7.100	6.000	1.100	58,7	63,6	-4,9
I alt	2.402.000	7.400	7.200	200	59,4	60,2	-0,8

Kilde: Egne beregninger på baggrund af en stikprøve på 3,3 pct. af befolkningen.

Med den hidtil anvendte metode var de fuldt lediges gennemsnitlige forskelsbeløb (4.100 kr.) omkring 20 pct. lavere end de delvist beskæftigedes gennemsnitlige forskelsbeløb (5.000 kr.). Med den nye metode er de fuldt lediges gennemsnitlige forskelsbeløb omkring 6 pct. højere end de delvist beskæftigedes gennemsnitlige forskelsbeløb.

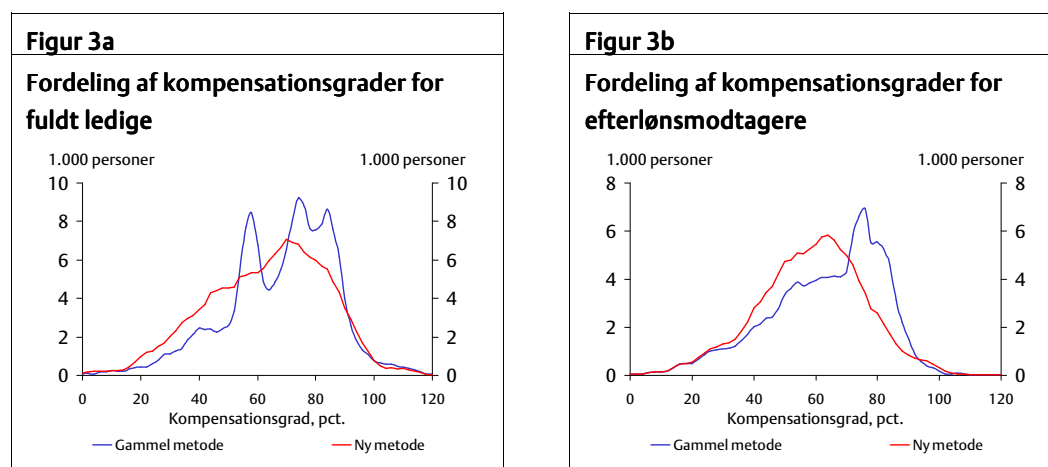
Det gennemsnitlige forskelsbeløb for efterlønsmodtagere er efter implementering af den nye metode tættere på forskelsbeløbet for de fuldt beskæftigede i gennem-

snit. Det skyldes, at den tildelte timeløn for efterlønsmodtagere stiger med ca. 9 pct. som følge af metodeændringen.

Konsekvens for kompensationsgrader

Den højere imputerede timeløn for de fuldt ledige medfører, at den gennemsnitlige kompensationsgrad for gruppen reduceres fra 68,0 til 63,3 pct., *jf. tabel 1*. Dermed bliver den noget lavere end kompensationsgraden for delvist beskæftigede. Det skal ses i sammenhæng med, at under halvdelen af de fuldt ledige er forsikrede mod ledighed, mens mere end 80 pct. af de delvist beskæftigede er medlem af en a-kasse. Dermed bliver den gennemsnitlige overførselsydelse – dagpenge for forsikrede og kontanthjælp for ikke-forsikrede – lavere for de fuldt ledige end for de delvist beskæftigede. Det medfører isoleret set en lavere kompensationsgrad for fuldt ledige end for delvist beskæftigede.

Kompensationsgraden for efterlønsmodtagere reduceres med knap 5 pct.-enheder og er i gennemsnit omtrent 1 pct.-enhed højere end for de fuldt beskæftigede. Fordelingen af kompensationsgraderne er således ændret noget for fuldt ledige og efterlønsmodtagere, *jf. figur 3a-3b*. For beskæftigede er fordelingerne stort set uændrede.



Kilde: Egne beregninger på baggrund af en stikprøve på 3,3 pct. af befolkningen.

Lave forskelsbeløb og høje kompensationsgrader

Andelen af personer med forskelsbeløb under 1.000 kr. om måneden reduceres samlet fra 3,3 til 2,9 pct. af de personer, som indgår i beregningen, *jf. tabel 3*.

Andelen af ledige og efterlønsmodtagere med forskelsbeløb under 1.000 kr. øges en smule. Det afspejler, at flere tildeles en imputeret timeløn i den helt lave ende af fordelingen med den nye metode end med den hidtil anvendte metode. Andelen af fuldt ledige og efterlønsmodtagere med forskelsbeløb under 2.000 kr. eller kompensationsgrader over 80 pct. er reduceret betydeligt med den nye metode – for efterlønsmodtagere er andelen omtrent halveret. Det skyldes, at timeløn svarende til 5 pct.-fraktilen i højere grad medfører et forskelsbeløb på under 2.000 kr. sammenlignet med de nye, mere spredte lønninger, *jf. figur 2a og figur 2b*, som viser fordelingen af forskelsbeløb for fuldt ledige og efterlønsmodtagere.

Tabel 3				
Andele med lave forskelsbeløb/høje kompensationsgrader				
	< 1000 kr.	< 2000 kr.	> 90 pct.	> 80 pct.
	----- Pct. -----			
Gammel metode				
Fuldt besk.	2,1	5,7	3,0	10,0
Delvist besk.	9,2	21,3	11,2	29,9
Fuldt ledig	5,7	17,8	7,2	28,6
Efterløn	1,9	12,4	2,7	19,4
<i>I alt</i>	<i>3,3</i>	<i>8,9</i>	<i>4,3</i>	<i>14,4</i>
Ny metode				
Fuldt besk.	1,7	4,9	2,5	9,4
Delvist besk.	8,4	20,5	0,5	29,3
Fuldt ledig	6,2	15,7	7,6	22,2
Efterløn	2,7	6,6	3,2	9,9
<i>I alt</i>	<i>2,9</i>	<i>7,9</i>	<i>3,9</i>	<i>12,9</i>

Kilde: Egne beregninger på baggrund af en stikprøve på 3,3 pct. af befolkningen.

Arbejdsudbud ved Forårspakke 2.0

Samlet fører metodeændringen til et skøn over den samlede arbejdsudbudseffekt af Forårspakke 2.0 på ca. 19.400 personer med udgangspunkt i den oprindelige vurdering af timeeffekten. Det svarer omtrent til det oprindelige skøn på 19.300 personer.

Den skønnede arbejdsudbudsvirkning som følge af deltagelseeffekten af skatteelementerne i Forårspakke 2.0 vurderes med den nye metode at svare til godt 1.050 personer, mens arbejdsudbudsvirkningen med den tidligere metode vurderes til godt 950 personer. Ændringen skyldes, at lediges potentielle arbejdsindkomst er forøget med den nye metode. Derved forøges gevinsten ved at komme i arbejde ved eksempelvis en afskaffelse af mellemskatten, hvilket alt andet lige øger arbejdsudbuddet.

4. Metode: En panelmodel for timeløn i Danmark 1995-2006

Den nye metode til imputering af skyggeløn er todelt:

1. *Estimation* af en model for timeløn ved brug af paneldata, hvor de samme individer følges over tid. Modellen udnytter information om individuelle karakteristika og timelønnen i hvert af de år, hvor der foreligger oplysninger. Lønnen fra Danmarks Statistiks lønregistre lægges til grund for estimationen.
2. *Simulation* af timelønnen for personer, der ikke har tilgængelige lønoplysninger i året.

Modellen estimeres på en stikprøve af personer, der har en registreret løn i lønstatistikkerne i minimum ét af årene. Personer, der ikke har en løn indgår selvsagt ikke. Den resulterende mulige selektionsbias er søgt håndteret gennem medtagelse af arbejdsmarkedshistorik som forklarende variabel (se endvidere afsnit 6).

Der estimeres en ligning på formen:

$$(1) \quad y_{it} = x_{it}\beta + u_i + \varepsilon_{it} \quad i=1, \dots, N \text{ og } t=1, \dots, T_i$$

$$\varepsilon_{it} = \rho\varepsilon_{it-1} + e_{it} \text{ hvor } -1 < \rho < 1,$$

$$e_{it} \stackrel{iid}{\sim} N(0, \sigma^2) \text{ og } u_i \stackrel{iid}{\sim} N(0, \sigma_u^2)$$

hvor indekset i angiver individet, mens indekset t angiver året, og x_{it} er en vektor af variable, der kan variere både på tværs af individer og over tid. u_i er en stokastisk, uobserverbar individspecifik effekt – en såkaldt *random effect*. Modellen tillader autokorrelation i fejleddet.

Modellen anvender 17 forskellige personlige karakteristika som forklarende variable til at forudsige lønnen (se tabel A3 i appendiks A for en oversigt). Alle relevante, tilgængelige oplysninger, som kan bidrage til at øge forklaringskraften, er således medtaget. Det er gjort som følge af analysens formål om at kunne forudsige timelønnen – og altså simulere den – og i mindre grad at fastsætte de enkelte bidrag fra de forskellige karakteristika.

Estimationen er foretaget ved anvendelse af paneldata af især to årsager. *For det første* giver en panelmodel – i sammenligning med en model baseret udelukkende på tværssnitsdata – væsentlig ekstra information, der kan udnyttes til en mere præcis imputation af de manglende timelønninger.

For det andet muliggør denne form for estimation identifikation af individspecifikke effekter, der ikke er observerbare. Det er således eksempelvis vanskeligt at måle effekten af evner og flid, men der kan ikke være meget tvivl om, at disse karakteristika spiller en rolle for lønnen. Med en panelmodel er det muligt at kvantificere disse uobserverbare individspecifikke effekter i en slags samlet *black box* (u_i -termen i ligning (1)). Det muliggør en mere præcis simulation af lønnen, end hvad der ellers ville have været muligt.

Modellens parametre estimeres ved anvendelse af Baltagi og Wus random effects-estimator¹ som implementeret i STATA 10.

Den endelige model opdeler data i 14 grupper, defineret ved køn samt syv kategorier efter højst fuldførte uddannelse. Det bidrager til at sikre en mere nøjagtig simulation af lønnen, blandt andet fordi fejlleddet kan simuleres mere præcist for de resulterende mere homogene grupper. Samtidig imødegås problemer omkring heteroskedasticitet på tværs af køn og uddannelsesgrupper. Parameterestimerer for de 14 estimationer fremgår af appendiks A.

5. Data

Der er anvendt et paneldatasæt med ca. 525.000 observationer fordelt på de 12 år i perioden 1995-2006. Lønoplysningerne til estimationen stammer fra Danmarks Statistiks lønregistre for den private, kommunale og statslige sektor.

Panelet tager udgangspunkt i en stikprøve på 3,3 pct. af Danmarks befolkning. Stikprøven er første gang udtaget pr. ultimo 1993, og den samme personkreds er herefter fulgt hvert år. For at bevare repræsentativiteten suppleres hvert år op med 1/30 af alle nyfødte og indvandrede, mens døde og udvandrede personer naturligt udgår.²

De lønoplysninger, der ligger til grund for estimationen, er først tilgængelige fra 1995, hvilket har afgrænset panelet tilbage i tid. Der er ligeledes foretaget en aldersafgrænsning, så kun 18-67 årige indgår. Derudover er personer, som ændrer uddannelsesniveau i perioden kun medtaget i de år, hvor de har opnået det seneste uddannelsesniveau (se afsnit 4 for en uddybning).

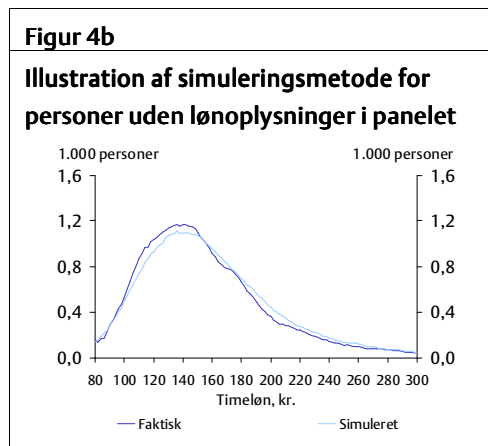
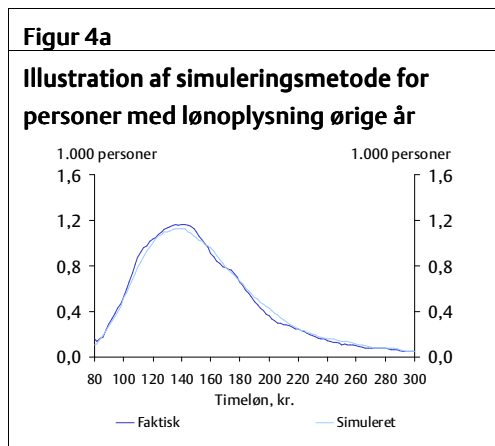
Panelet er ubalanceret, dvs. at der ikke er oplysninger om alle personer i alle 12 år. Samlet indgår godt 900.000 observationer i panelet. Heraf har 58 pct. – svarende til 525.000 observationer – en lønoplysning. Antal observationer med en lønoplysning varierer fra knap 29.000 i 1995 til knap 59.000 i 2006. Stigningen skyldes fortrinsvis, at lønoplysningerne er blevet udbygget i løbet af perioden.

6. Grundlæggende egenskaber

Modellen fører til en simuleret lønfordeling, der lægger sig tæt op ad den faktiske fordeling, jf. figur 4a og 4b, hvor fordelingerne sammenlignes for de personer, der har lønoplysninger i det aktuelle år. Figur 4a illustrerer metoden til at forudsige lønnen for personer med lønoplysninger i øvrige år end det aktuelle, mens figur 4b illustrerer den metode, der benyttes for personer, som ingen lønoplysninger har i panelet.

¹ Baltagi, H. Badi & Ping X. Wu (1999). Unequally Spaced Panel Data Regressions with AR(1) Disturbances. *Econometric Theory* 15, 814-823.

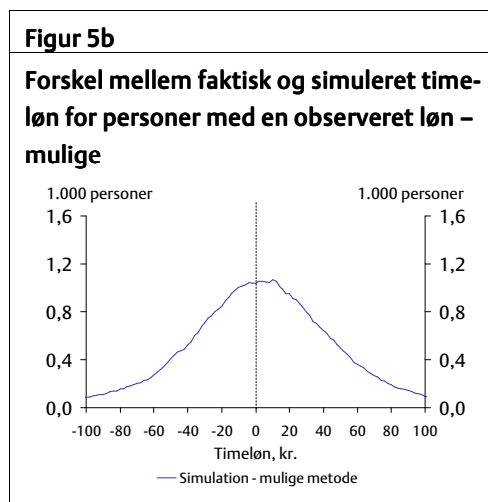
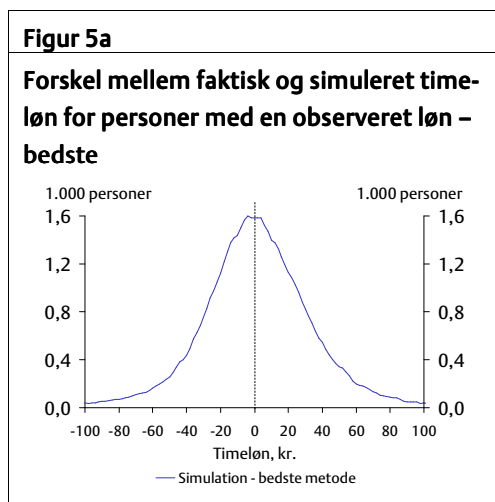
² Udvandrede individer, som senere genindvandrer, kommer automatisk med i stikprøven igen, hvis de oprindeligt er udtaget.



Kilde: Egne beregninger på baggrund af en stikprøve på 3,3 pct. af befolkningen.

Anm.: I figur 1a er vist den simulerede lønfordeling, når der anvendes den betingede fordeling for det individspecifikke heterogenitetsled. Det er den bedste simulerede fordeling, givet den anvendte metode og data. I figur 1b er vist den simulerede lønfordeling, når den ubetingede fordeling for det individspecifikke heterogenitetsled anvendes. Denne metode benyttes for personer, der ikke har lønoplysninger.

Residualerne fordeler sig pænt omkring nul ved metoden benyttet til personer med lønoplysninger i øvrige år, *jf. figur 5a*, hvor omkring 70 pct. af de individuelle afvigelser er mindre end 30 kr. Ved at anvende den bedst mulige metode for personer uden nogen lønoplysninger i perioden, øges spredningen noget, *jf. figur 5b*, men afvigelserne fordeler sig stadig jævnt omkring nul. Godt 50 pct. af individerne ligger her inden for en afvigelse på 30 kr.



Kilde: Egne beregninger på baggrund af en stikprøve på 3,3 pct. af befolkningen.

7. Ledighedsperioder, selektionsproblemet og endogenitet

Arbejdsmarkedshistorik er medtaget som forklarende variabel. Arbejdsmarkedshistorik medtages for at kontrollere for den mulige effekt, at personer, der ligner hinanden på alle øvrige faktorer i modellen, ikke vil få samme løn, hvis de ikke har samme arbejdsmarkedshistorik. Hvis én person eksempelvis er ledig i en længere periode, vil vedkommende sandsynligvis ikke efterfølgende kunne opnå samme løn som en person, der i øvrigt er identisk, med den undtagelse, at vedkommende

har været i job i perioden. Dette er også kendt som selektionsproblemet – altså det, at der estimeres løn for personer uden job på grundlag af personer i job. Medtagelsen af arbejdsmarkedshistorikken har bl.a. til formål at imødegå denne problematik ved at estimere betydningen af ledighed.

Estimatet for arbejdsmarkedshistorik bygger på oplysninger om personer, der har været beskæftiget mindst 1 uge i perioden og således har lønoplysninger registeret. Det er argumentet, at personer, der har meget lav beskæftigelsesgrad ligner personer med en beskæftigelsesgrad på 0 tilstrækkeligt til, at der kan estimeres meningsfulde effekter.

Effekten estimeres endvidere på ugebasis med udgangspunkt i oplysninger om antal ugers ledighed i DREAM-registret. Der er altså tale om en relativt ”finmasket” kvantificering af effekten af perioder uden for beskæftigelse.

Medtagelse af arbejdsmarkedshistorik betyder, at ledighedsperioders længde anvendes til at forudsige timelønnen. Kausaliteten går ikke nødvendigvis kun én vej mellem ledighedsperioder og timeløn. Derfor er der en risiko for, at estimaterne overvurderer effekten af arbejdsmarkedshistorik. Denne problematik er kendt som endogenitetsproblemet. Det vurderes dog, at medtagelse af arbejdsmarkedshistorik indfanger en vigtig sammenhæng, som bør medtages i modellen.

8. Panelmodeller og simulering

Hvis formålet var ren estimation (og altså ikke efterfølgende simulation), ville en dynamisk model, hvor eksempelvis den ”laggede” løn ($y_{i,t}$) indgår som forklarende variabel, være at foretrække. Det skyldes, at lønnen i et givet år vil have en nær tilknytning til lønnen det foregående år.

Det er til gengæld ikke hensigtsmæssigt at medtage den ”laggede” løn, når formålet er at anvende estimaterne til at simulere lønnen blandt andre for personer, der ikke har en lønoplysning i et eneste af de forudgående år (i de 12 år, panelet strækker sig over).

Det er imidlertid nødvendigt at tage højde for autoregression, fordi lønnen i et givet år som nævnt har nær tilknytning til lønnen i det foregående år. Baltagi og Wus estimator tager netop højde for en autoregressiv proces i fejleddet (AR[1]).

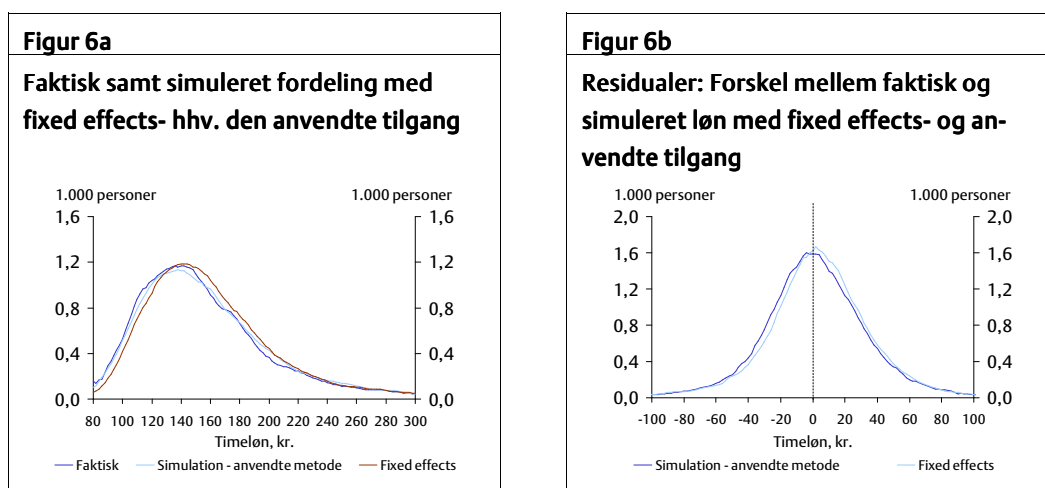
Inden for ”main stream” paneldataestimation anvendes især en *fixed effects*- og en *random effects*-tilgang, når der ikke anvendes dynamisk modellering. Til estimationerne er anvendt en random effects-estimator af især to nært knyttede årsager.

For det første gør random effects-tilgangen brug af såvel tværsnitsdimensionen som af tidsdimensionen i panelet. Til sammenligning gør fixed effects-tilgangen alene brug af information fra faktorer, der varierer over tid. Effekten af tidsinvariante faktorer som eksempelvis køn og karakterniveau estimeres således ikke eksplicit ved anvendelse af en fixed effects-estimator. Derfor vil der som udgangspunkt være nogle faktorer, som fixed effects-estimation ikke giver parameterestimater for, selvom de har en oplagt betydning for løndannelsen. Effekten af de tidsinvariante faktorer håndteres derimod samlet som uobserverbar individualspecifik heterogenitet med fixed effects-tilgangen

For det andet estimeres den uobserverbare individualspecifikke heterogenitet i fixed effects-tilgangen som noget fast, der afhænger af den givne estimation. Derfor kan den ikke simuleres uafhængigt af den givne estimation – og det kræver, at det enkelte individ indgår i estimationen. Nogle af de individer, der skal simuleres skyggeløn for, har ingen lønoplysninger. De indgår naturligvis ikke i estimationen – og værdien af u_i er ikke mulig at fastlægge. For gruppen af personer, der slet ikke har en lønoplysning i nogle af de 12 år i panelet, vil det derfor ikke være optimalt at simulere løn på baggrund af fixed effects-tilgangen.

Baltagi & Wus estimator har i tillæg den fordel, at den kan håndtere ubalancerede paneler. Der kan derfor tages højde for, at alle individer ikke indgår i alle år.

Der er ikke store forskelle på den resulterende simulerede lønfordeling, hvis en fixed effects-model ligger til grund for simulationen i stedet for den anvendte random effects-model, *jf. figur 6a*. Den store lighed mellem de resulterende fordelinger for de to tilgange medfører, at den anvendte metode står stærkere. Det skyldes, at fixed effects-estimation som udgangspunkt er mere robust end random effects-estimation over for eksempelvis manglende variabler eller observationer.



Kilde: Egne beregninger på baggrund af en stikprøve på 3,3 pct. af befolkningen.

Den anvendte random effects-model rammer den faktiske fordeling noget bedre for de laveste indkomster. Det har betydning i forhold til den videre anvendelse af de simulerede lønninger, fordi der er flest personer uden lønoplysninger i den lavere del af indkomstfordelingen. Residualerne fordeler sig omtrent ens for random effects-tilgangen og fixed effects-tilgangen, *jf. figur 6b*.

9. Simulering af skyggeløn

Simulering af skyggeløn i 2005 tager udgangspunkt i for hvilke år, der er anvendelige timelønninger i lønregistrene. I det følgende beskrives simulering af manglende 2005-timelønninger for personer, der har timeløn i mindst ét år i perioden fra 1995 til 2006.

Simuleringen foretages ved tilfældig udtrækning af modellens to stokastiske elementer, nemlig e_{it} og u_i betinget på modellens estimerede parametre. Modellens specifikation er

$$(1.1) \quad y_{it} = x_{it}\beta + u_i + \varepsilon_{it} \quad i=1, \dots, N \text{ og } t=1, \dots, T_i$$

$$(1.2) \quad \varepsilon_{it} = \rho\varepsilon_{it-1} + e_{it} \text{ hvor } -1 < \rho < 1$$

$$(1.3) \quad e_{it} \stackrel{iid}{\sim} N(0, \sigma^2)$$

$$(1.4) \quad u_i \stackrel{iid}{\sim} N(0, \sigma_u^2)$$

Simuleringen er konsistent med modellens struktur og foretages med følgende trin:

1. Udtrækning af den individuelle effekt (random effect) u_i enten fra dennes betingede eller ubetingede fordeling.
2. Udtrækning af $y_{it=2005}$ fra den betingede fordeling, dvs. betinget på u_i og modellens parametre. Den betingede fordeling for en persons $y_{it=2005}$ afhænger af for hvilke år, der er timelønsoplysninger for personen.

De to trin er beskrevet i detaljer i appendiks B.

Appendiks A. Variable samt parameterestimer

Tabel A3 indeholder en variabeliste med beskrivelse af de baggrundsoplysninger, der indgår i estimationen. Der er gjort visse valg ved dannelse af nogle af de øvrige parametre, der indgår i estimationen. Antagelserne gennemgås nedenfor.

Det er valgt at medtage såvel erfaring som alder, selvom de udtrykker omtrent det samme. Det skyldes formålet med imputeringen, hvor den forudsagte timeløn er i fokus frem for de enkelte parameterestimer. Til andre formål kunne argumenteres for at udelade alder i estimationen, da erfaring ofte indgår i fastsættelse af løn (eksempelvis i offentlige overenskomster), hvorimod der er forbud mod at aldersdiskriminere. For at ramme en eventuel udfladning af effekten fra erfaring er det derudover valgt at medtage kvadratroden af erfaring.

Det ville være oplagt i en løn estimation at medtage oplysninger om branche, som formentlig vil have en god forklaringskraft. Der er imidlertid samme problem med brancheoplysninger som med løn fra året før (y_{t-1}). Den personkreds, som i sidste ende er interessant, har ingen lønoplysninger i året og muligvis heller ikke i årene før. Som en konsekvens af dette er en del af personerne heller ikke tilknyttet nogen branche, hvilket umuliggør medtagelse af branche som forklarende faktor.

Inddeling af a-kasser

Der er sket en del fusioner og opsplitninger i løbet af perioden fra 1995-2006, og det er valgt at samle de a-kasser, som har været knyttet sammen i dele af perioden. Tabel A1 gengiver inddelingen.

Tabel A1	
Inddeling af a-kasser – forskellige navne på a-kasser i løbet af perioden	
Værdi	A-kasser
0	Øvrige
1	Ikke forsikret
2	3F, Specialarbejde (SID), Kvindelige arbejdere, Restaurations og bryggeriarbejdere, Beklædning og tekstil
3	FOA, offentligt ansatte, pædagogmedhjælperne, Børne- og ungdomspædagoger
4	Grafikere (splittes op i tre i løbet af perioden)
5	Handels- og kontorfunktionærerne (HK)
6	Kristelige
7	TIB, Snedker- og Tømrerfaget, Træindustriarbejdere
8	Lederne, Danske teknikere, Ingeniørerne, Civiløkonomerne
9	Danmarks aktive handelsrejsende, Danske sælgere, Business
10	Journalistik, Kommunikation og Sprog, Journalisterne, Erhvervsproglig
11	Funktionærer og Tjenestemænd, Folkeskolelærere, Danske sygeplejersker, Forsvaret

Urbaniseringsgrad

Der er foretaget en inddeling i kommuner for at tage højde for, at urbaniseringsgraden kan påvirke lønniveauet. Inddelingen er foretaget på baggrund af de gamle

kommuner og amter, idet de nye regioner og kommuner først blev dannet efter panelets sidste år, nemlig 1. januar 2007.

Tabel A2	
Urbaniseringsinddeling	
Værdi	Område
0	Øvrig provins
1	Odense, Esbjerg, Aalborg
2	København + omegn (København, Frederiksberg, Københavns Amt, Frederiksborg Amt minus Hundested, Frederiksværk, Frederikssund, Jægerspris, Skibby, Skævinge, Slangerup, Ølstykke) samt Århus kommune

Tabel A3	
Variabelliste	
Variabel	Indhold
ald0-ald10	Alder opdelt i 5-års intervaller
alks0-alks11	A-kasse opdelt, se tekst
barn, b0002, b0306, b0717, b0317	Barn i husstanden, øvrige variable betinget på barnets alder i intervaller 0-2,3-6,7-17,3-17
by0-by2	Opdelt på baggrund af geografisk beliggenhed, se tabel A2
d1995-d2006	Årsdummy
Efterløn	Overgang til efterløn i løbet af året
Enlig	Dummy for om individet er enlig
Erfaring	Antal års erhvervs erfaring
Sqrterfaring	Kvadratrod af antal års erhvervs erfaring
kar0-kar4	Gymnasiekarakter i intervaller 0-7,7-9,9-10,10+ samt ukendt
land	Vestligt (inkl. DK)/ Ikke-vestligt
d_dagpenge, d_føp, d_kth, d_revalidering, d_sdp	Dummy for mindst én uges dagpenge, førtidspension, kontanthjælp, revalidering og sygedagpenge i året
ldagpenge, lføp, lkth, lrevalidering, lsdp	Logaritmen til antal ugers modtagelse af dagpenge, førtidspension, kontanthjælp, revalidering og sygedagpenge i året
lkum_dagpenge, lkum_føp, lkum_kth, lkum_revalidering, lkum_sdp	Logaritmen til kumuleret antal uger på den respektive overførsel 3 år tilbage i forhold til det aktuelle år (revalidering dog kun 2 år)
Int_samlet	Dummy for mindst 26 ugers modtagelse af samme ydelse i året (én af de fem ovenstående)
Nyuddannet	Dummy for første år med ny uddannelse
uddf0-uddf6	Faders højest fuldførte uddannelsesniveau
uddm0-uddm6	Moders højest fuldførte uddannelsesniveau
Anm.: Dagpenge indeholder også ledighedsydelse. Kontanthjælp indeholder også starthjælp og introduktionsydelse.	

Tabel 4

Parameterestimer (standardafvigelser) for de 14 estimationer

	<i>Mænd</i>							<i>Kvinder</i>						
<i>Uddannelsesniveau (højeste)</i>	<i>Ukendt</i>	<i>Folkeskole</i>	<i>Gymnasial</i>	<i>EVU</i>	<i>KVU</i>	<i>MVU</i>	<i>LVU</i>	<i>Ukendt</i>	<i>Folkeskole</i>	<i>Gymnasial</i>	<i>EVU</i>	<i>KVU</i>	<i>MVU</i>	<i>LVU</i>
Antal observationer	3.454	56.784	15.755	108.699	15.609	36.017	24.560	2.523	55.883	17.189	97.559	12.205	63.082	15.975
R² – within	0,231	0,308	0,355	0,353	0,451	0,405	0,475	0,939	0,326	0,314	0,408	0,382	0,470	0,480
R² – between	0,256	0,239	0,433	0,229	0,247	0,320	0,309	0,265	0,254	0,319	0,259	0,285	0,326	0,293
R¹ – overall	0,261	0,251	0,429	0,269	0,292	0,347	0,336	0,317	0,306	0,393	0,326	0,309	0,410	0,363
Konstant	4,887 (0,047)	4,448 (0,014)	4,44 (0,021)	4,529 (0,013)	4,44 (0,039)	4,67 (0,025)	4,661 (0,026)	4,607 (0,043)	4,436 (0,009)	4,545 (0,026)	4,497 (0,012)	4,374 (0,035)	4,612 (0,014)	4,567 (0,168)
ald1	-0,368 (0,046)	-0,011 (0,012)	. (.)	. (.)	. (.)	-0,294 (0,046)	. (.)	-0,208 (0,043)	. (.)	-0,155 (0,023)	-0,011 (0,009)	-0,096 (0,037)	0,151 (0,038)	. (.)
ald2	-0,313 (0,041)	0,046 (0,011)	0,051 (0,007)	0,041 (0,005)	0,117 (0,031)	-0,101 (0,017)	. (.)	-0,15 (0,038)	0,036 (0,005)	-0,104 (0,022)	0,014 (0,008)	0,052 (0,025)	-0,095 (0,01)	0,121 (0,166)
ald3	-0,192 (0,039)	0,06 (0,01)	0,152 (0,01)	0,075 (0,005)	0,152 (0,032)	-0,017 (0,015)	0,123 (0,009)	-0,078 (0,037)	0,054 (0,006)	-0,024 (0,021)	0,043 (0,008)	0,104 (0,023)	-0,066 (0,009)	0,187 (0,166)
ald4	-0,127 (0,037)	0,064 (0,01)	0,209 (0,012)	0,1 (0,006)	0,181 (0,033)	0,05 (0,014)	0,228 (0,01)	-0,091 (0,036)	0,056 (0,007)	0,006 (0,02)	0,054 (0,007)	0,113 (0,022)	-0,042 (0,009)	0,248 (0,166)
ald5	-0,095 (0,035)	0,071 (0,009)	0,23 (0,015)	0,112 (0,006)	0,192 (0,033)	0,102 (0,013)	0,292 (0,012)	-0,079 (0,034)	0,053 (0,007)	0,025 (0,019)	0,054 (0,007)	0,115 (0,021)	-0,025 (0,008)	0,279 (0,167)

ald6	-0,031 (0,033)	0,069 (0,008)	0,222 (0,017)	0,108 (0,007)	0,193 (0,034)	0,113 (0,012)	0,309 (0,013)	-0,075 (0,032)	0,054 (0,007)	0,033 (0,017)	0,048 (0,006)	0,099 (0,02)	-0,012 (0,008)	0,289 (0,167)
ald7	-0,03 (0,03)	0,06 (0,007)	0,182 (0,02)	0,098 (0,008)	0,186 (0,035)	0,114 (0,01)	0,298 (0,015)	-0,086 (0,03)	0,046 (0,008)	0,033 (0,016)	0,036 (0,006)	0,082 (0,018)	-0,012 (0,007)	0,279 (0,167)
ald8	-0,013 (0,027)	0,04 (0,007)	0,152 (0,024)	0,076 (0,008)	0,14 (0,036)	0,088 (0,009)	0,255 (0,017)	-0,078 (0,028)	0,037 (0,008)	0,017 (0,014)	0,019 (0,006)	0,051 (0,018)	-0,014 (0,007)	0,239 (0,167)
ald9	-0,03 (0,024)	0,027 (0,006)	0,093 (0,029)	0,042 (0,009)	0,083 (0,037)	0,046 (0,008)	0,196 (0,019)	-0,069 (0,027)	0,035 (0,009)	.	0,009 (0,005)	0,026 (0,017)	-0,02 (0,007)	0,169 (0,167)
ald10	.	.	0,041 (0,041)	0,013 (0,01)	0,032 (0,04)	.	0,128 (0,021)	.	0,037 (0,01)	-0,063 (0,035)	.	.	.	0,126 (0,168)
alks0	-0,02 (0,019)	0,051 (0,004)	0,038 (0,01)	-0,006 (0,003)	0,002 (0,008)	-0,045 (0,008)	-0,091 (0,007)	0,055 (0,019)	0,051 (0,004)	0,07 (0,009)	0,03 (0,004)	0,025 (0,011)	-0,007 (0,004)	-0,05 (0,009)
alks2	-0,082 (0,018)	0,011 (0,004)	-0,038 (0,01)	-0,034 (0,004)	-0,061 (0,013)	-0,09 (0,016)	-0,204 (0,028)	-0,035 (0,016)	-0,005 (0,004)	0,033 (0,011)	-0,024 (0,004)	-0,063 (0,018)	-0,113 (0,013)	-0,236 (0,041)
alks3	-0,143 (0,039)	-0,055 (0,008)	-0,091 (0,02)	-0,095 (0,007)	-0,17 (0,024)	-0,129 (0,01)	-0,238 (0,043)	-0,031 (0,022)	-0,005 (0,004)	-0,049 (0,011)	-0,026 (0,003)	-0,085 (0,015)	-0,069 (0,004)	-0,28 (0,031)
alks4	-0,184 (0,121)	0,079 (0,013)	-0,012 (0,053)	0,078 (0,009)	-0,062 (0,045)	0,011 (0,079)	.	0,001 (0,089)	0,047 (0,014)	-0,151 (0,101)	0,084 (0,013)	0,22 (0,077)	-0,019 (0,047)	-0,222 (0,143)
alks5	-0,155 (0,049)	0,026 (0,008)	-0,021 (0,012)	-0,021 (0,004)	-0,009 (0,011)	-0,03 (0,017)	-0,269 (0,032)	-0,022 (0,025)	0,02 (0,004)	0,016 (0,009)	-0,015 (0,003)	-0,045 (0,011)	0,002 (0,007)	-0,165 (0,025)
alks6	-0,085 (0,026)	0,001 (0,006)	0,005 (0,013)	-0,022 (0,005)	0,007 (0,012)	-0,014 (0,015)	-0,183 (0,032)	-0,078 (0,024)	0,004 (0,005)	0,013 (0,011)	-0,011 (0,004)	0,005 (0,014)	-0,034 (0,009)	-0,142 (0,033)
alks7	-0,105 (0,037)	-0,002 (0,007)	-0,086 (0,034)	-0,018 (0,005)	-0,018 (0,03)	-0,007 (0,034)	-0,237 (0,089)	-0,076 (0,149)	0,021 (0,01)	-0,003 (0,044)	-0,004 (0,012)	0,22 (0,064)	-0,186 (0,067)	.
alks8	-0,071	0,166	0,122	0,123	0,07	0,135	0,013	0,157	0,139	0,188	0,099	0,053	0,187	0,015

	(0,036)	(0,008)	(0,015)	(0,004)	(0,008)	(0,006)	(0,009)	(0,045)	(0,009)	(0,017)	(0,006)	(0,013)	(0,007)	(0,013)
alks9	0,227 (0,09)	0,189 (0,015)	0,082 (0,025)	0,134 (0,007)	0,134 (0,018)	0,065 (0,034)	-0,078 (0,09)	0,823 (0,129)	0,288 (0,027)	0,205 (0,033)	0,142 (0,013)	0,16 (0,043)	0,331 (0,025)	0 (0)
alks10	0,846 (0,299)	0,235 (0,044)	0,176 (0,041)	0,128 (0,028)	-0,022 (0,054)	0,138 (0,017)	-0,129 (0,041)	0,219 (0,108)	0,185 (0,044)	0,34 (0,028)	0,186 (0,037)	0,076 (0,017)	0,102 (0,009)	-0,105 (0,023)
alks11	0,01 (0,042)	0,08 (0,008)	0,017 (0,013)	0,07 (0,005)	0,009 (0,01)	-0,045 (0,007)	-0,154 (0,019)	0,053 (0,027)	0,114 (0,007)	0,11 (0,011)	0,059 (0,004)	0,02 (0,011)	0,036 (0,004)	-0,106 (0,017)
b0002	-0,003 (0,019)	0,008 (0,004)	0,019 (0,01)	0,007 (0,003)	0,016 (0,006)	0,002 (0,005)	-0,006 (0,006)	-0,014 (0,022)	-0,007 (0,005)	-0,019 (0,009)	-0,014 (0,003)	-0,016 (0,008)	-0,025 (0,003)	-0,018 (0,007)
b0306	-0,015 (0,017)	0,001 (0,004)	0,008 (0,009)	0,003 (0,003)	0,003 (0,006)	0,004 (0,004)	0,01 (0,005)	0,002 (0,017)	0,001 (0,004)	-0,004 (0,007)	-0,003 (0,002)	-0,005 (0,007)	-0,004 (0,002)	-0,005 (0,006)
b0317	0,009 (0,03)	-0,007 (0,006)	0,022 (0,014)	0,005 (0,004)	0,019 (0,009)	0,016 (0,007)	0,006 (0,008)	-0,063 (0,034)	-0,02 (0,007)	-0,005 (0,012)	0,003 (0,004)	-0,007 (0,011)	-0,004 (0,004)	-0,008 (0,009)
b0717	0,014 (0,024)	0,009 (0,005)	0,029 (0,011)	0,001 (0,003)	0,007 (0,007)	-0,005 (0,005)	0 (0,007)	0,033 (0,024)	0,002 (0,005)	-0,011 (0,009)	-0,004 (0,003)	-0,005 (0,009)	-0,016 (0,003)	-0,001 (0,007)
barn	-0,021 (0,028)	-0,001 (0,006)	0,018 (0,013)	0,004 (0,004)	-0,018 (0,009)	0,011 (0,007)	0,027 (0,008)	0,024 (0,033)	0,017 (0,006)	0,005 (0,011)	-0,007 (0,004)	0,014 (0,011)	0,017 (0,004)	0,029 (0,009)
by1	0,042 (0,026)	0,004 (0,006)	-0,031 (0,01)	-0,001 (0,004)	-0,023 (0,01)	-0,015 (0,008)	-0,002 (0,01)	0,009 (0,026)	0,01 (0,005)	-0,027 (0,008)	-0,007 (0,004)	-0,002 (0,012)	0,005 (0,004)	-0,024 (0,012)
by2	0,063 (0,016)	0,023 (0,004)	0,014 (0,006)	0,045 (0,003)	0,031 (0,006)	0,022 (0,005)	0,026 (0,006)	0,016 (0,013)	0,031 (0,003)	0,028 (0,005)	0,041 (0,002)	0,067 (0,007)	0,03 (0,002)	0,023 (0,006)
d1996	0,029 (0,017)	0,045 (0,003)	0,035 (0,01)	0,04 (0,002)	0,043 (0,006)	0,044 (0,004)	0,05 (0,006)	0,04 (0,015)	0,031 (0,003)	0,035 (0,009)	0,038 (0,002)	0,039 (0,007)	0,052 (0,003)	0,039 (0,008)
d1997	0,076 (0,017)	0,073 (0,003)	0,059 (0,01)	0,068 (0,002)	0,08 (0,007)	0,069 (0,005)	0,081 (0,006)	0,087 (0,016)	0,062 (0,003)	0,064 (0,009)	0,064 (0,002)	0,068 (0,008)	0,085 (0,003)	0,077 (0,008)

d1998	0,081 (0,017)	0,1 (0,003)	0,071 (0,01)	0,094 (0,003)	0,103 (0,007)	0,097 (0,005)	0,099 (0,006)	0,091 (0,016)	0,076 (0,003)	0,072 (0,009)	0,085 (0,002)	0,099 (0,008)	0,115 (0,003)	0,09 (0,008)
d1999	0,105 (0,018)	0,13 (0,004)	0,107 (0,01)	0,133 (0,003)	0,156 (0,007)	0,137 (0,005)	0,152 (0,006)	0,126 (0,017)	0,103 (0,003)	0,113 (0,009)	0,12 (0,002)	0,139 (0,008)	0,15 (0,003)	0,141 (0,008)
d2000	0,133 (0,017)	0,152 (0,004)	0,138 (0,01)	0,16 (0,003)	0,184 (0,007)	0,177 (0,005)	0,188 (0,006)	0,148 (0,017)	0,129 (0,003)	0,126 (0,009)	0,145 (0,002)	0,162 (0,008)	0,192 (0,003)	0,178 (0,008)
d2001	0,141 (0,018)	0,179 (0,004)	0,164 (0,01)	0,189 (0,003)	0,219 (0,007)	0,203 (0,005)	0,226 (0,007)	0,173 (0,017)	0,145 (0,003)	0,15 (0,009)	0,166 (0,003)	0,192 (0,008)	0,199 (0,003)	0,208 (0,008)
d2002	0,145 (0,018)	0,195 (0,004)	0,191 (0,01)	0,209 (0,003)	0,244 (0,007)	0,217 (0,005)	0,245 (0,007)	0,181 (0,017)	0,169 (0,003)	0,174 (0,009)	0,188 (0,003)	0,217 (0,008)	0,218 (0,003)	0,237 (0,008)
d2003	0,165 (0,018)	0,216 (0,004)	0,189 (0,01)	0,227 (0,003)	0,268 (0,008)	0,232 (0,005)	0,263 (0,007)	0,195 (0,017)	0,181 (0,003)	0,2 (0,009)	0,214 (0,003)	0,243 (0,008)	0,238 (0,003)	0,259 (0,009)
d2004	0,196 (0,019)	0,231 (0,004)	0,205 (0,01)	0,245 (0,003)	0,283 (0,008)	0,259 (0,006)	0,288 (0,007)	0,23 (0,018)	0,205 (0,004)	0,211 (0,009)	0,238 (0,003)	0,263 (0,009)	0,275 (0,003)	0,29 (0,009)
d2005	0,226 (0,018)	0,266 (0,004)	0,227 (0,01)	0,274 (0,003)	0,316 (0,008)	0,288 (0,006)	0,322 (0,007)	0,264 (0,017)	0,232 (0,004)	0,231 (0,009)	0,263 (0,003)	0,288 (0,009)	0,303 (0,003)	0,313 (0,009)
d2006	0,239 (0,018)	0,294 (0,004)	0,251 (0,01)	0,308 (0,003)	0,347 (0,008)	0,315 (0,006)	0,347 (0,008)	0,306 (0,017)	0,253 (0,004)	0,246 (0,009)	0,287 (0,003)	0,317 (0,009)	0,319 (0,003)	0,343 (0,009)
d_dagpenge	-0,03 (0,024)	-0,001 (0,004)	-0,005 (0,018)	0,004 (0,003)	-0,013 (0,01)	0,001 (0,008)	-0,017 (0,011)	0,008 (0,02)	0,01 (0,004)	-0,004 (0,011)	-0,008 (0,003)	-0,006 (0,009)	0,014 (0,003)	-0,011 (0,008)
d_føp	0,109 (0,364)	-0,044 (0,083)	-0,336 (0,708)	0,048 (0,061)	-0,421 (0,127)	0,144 (0,114)	1,122 (0,476)	-9,828 (21,221)	0,086 (0,056)	0,131 (0,208)	-0,023 (0,057)	0,086 (0,232)	0,014 (0,087)	0,051 (0,228)
d_kth	-0,034 (0,047)	-0,011 (0,011)	0,01 (0,027)	0,019 (0,013)	-0,061 (0,042)	0,011 (0,03)	0,064 (0,043)	0,052 (0,039)	-0,016 (0,011)	-0,067 (0,021)	0,009 (0,014)	-0,011 (0,047)	0,051 (0,018)	0,011 (0,047)
d_revalidering	0,253	-0,065	0,107	-0,007	0,222	0,004	0,38	0,298	-0,022	-0,098	0,005	-0,065	0,003	0,078

	(0,152)	(0,034)	(0,124)	(0,034)	(0,074)	(0,047)	(0,178)	(0,214)	(0,03)	(0,091)	(0,027)	(0,093)	(0,022)	(0,136)
d_sdp	-0,01 (0,023)	-0,026 (0,004)	-0,033 (0,013)	-0,02 (0,003)	-0,02 (0,009)	-0,015 (0,007)	-0,012 (0,01)	-0,011 (0,019)	-0,021 (0,004)	-0,049 (0,01)	-0,017 (0,003)	-0,003 (0,011)	-0,025 (0,003)	-0,005 (0,01)
efterløn	-0,08 (0,042)	0,012 (0,007)	0,048 (0,054)	-0,014 (0,006)	0,017 (0,021)	-0,021 (0,012)	-0,023 (0,022)	-0,022 (0,035)	-0,003 (0,006)	-0,011 (0,038)	-0,025 (0,005)	0,001 (0,019)	-0,015 (0,007)	-0,031 (0,03)
enlig	-0,011 (0,013)	-0,01 (0,003)	-0,006 (0,006)	-0,004 (0,002)	0,001 (0,005)	-0,015 (0,004)	-0,019 (0,005)	0,047 (0,011)	0,01 (0,002)	0,002 (0,004)	0,003 (0,002)	-0,004 (0,005)	0,001 (0,002)	-0,01 (0,005)
erfaring	-0,002 (0,002)	-0,001 (0,001)	0,013 (0,001)	0,002 (0,001)	-0,001 (0,001)	0,002 (0,001)	0,006 (0,001)	0,001 (0,002)	0,006 (0,001)	0,01 (0,001)	0,007 (0,001)	-0,002 (0,002)	0,001 (0,001)	0,005 (0,001)
int_samlet	-0,021 (0,02)	-0,005 (0,004)	0,017 (0,014)	-0,007 (0,004)	0 (0,011)	0,033 (0,009)	0,016 (0,012)	-0,019 (0,015)	0,006 (0,003)	-0,008 (0,009)	-0,025 (0,003)	-0,021 (0,008)	-0,02 (0,004)	-0,021 (0,008)
kar1	. (.)	0 (0)	-0,033 (0,015)	0,028 (0,018)	-0,022 (0,023)	-0,056 (0,018)	-0,121 (0,038)	. (.)	-0,198 (0,156)	-0,007 (0,01)	0,018 (0,008)	-0,018 (0,019)	0,014 (0,006)	-0,089 (0,044)
kar2	. (.)	0,069 (0,2)	0,018 0(0,01)	0,018 (0,01)	0,013 (0,013)	-0,022 (0,009)	-0,022 (0,012)	. (.)	0,043 (0,152)	0,005 (0,007)	0,048 (0,006)	0,015 (0,011)	0,025 (0,004)	-0,025 (0,012)
kar3	. (.)	. (.)	0,01 (0,017)	0,05 (0,033)	0,044 (0,038)	0,037 0(0,016)	0,037 (0,014)	. (.)	. (.)	0,021 (0,011)	0,076 (0,019)	0,035 (0,028)	-0,006 (0,007)	0,011 (0,013)
kar4	. (.)	. (.)	0,058 (0,036)	0,257 (0,087)	0,077 (0,105)	-0,077 (0,034)	0,036 (0,021)	. (.)	. (.)	-0,003 (0,023)	-0,202 (0,151)	0,094 (0,089)	0,017 (0,017)	0,022 (0,02)
land	0,128 (0,02)	0,008 (0,009)	0,043 (0,018)	0,03 (0,01)	0,055 (0,022)	0,045 (0,018)	0,136 (0,022)	0,078 (0,017)	-0,013 (0,007)	0,04 (0,013)	-0,015 (0,008)	0,019 (0,023)	-0,004 (0,01)	0,061 (0,021)
ldagpenge	-0,004 (0,01)	-0,008 (0,002)	-0,01 (0,007)	-0,013 (0,002)	-0,018 (0,005)	-0,022 (0,004)	-0,018 (0,005)	-0,011 (0,008)	-0,008 (0,002)	0 (0,005)	-0,003 (0,001)	-0,003 (0,004)	-0,012 (0,002)	-0,011 (0,004)
lføp	-0,085 (0,109)	-0,002 (0,026)	0,07 (0,192)	-0,003 (0,019)	0,169 (0,046)	-0,047 (0,034)	-0,364 (0,136)	2,513 (5,45)	-0,029 (0,018)	-0,041 (0,069)	0,024 (0,017)	-0,004 (0,073)	0,006 (0,026)	-0,059 (0,069)

lkth	-0,024 (0,017)	-0,006 (0,004)	-0,018 (0,011)	-0,025 (0,005)	0,001 (0,016)	-0,027 (0,013)	-0,047 (0,017)	-0,026 (0,013)	-0,002 (0,004)	0,02 (0,008)	-0,009 (0,005)	-0,019 (0,018)	-0,028 (0,008)	-0,021 (0,02)
lkum_dagpenge	-0,011 (0,004)	-0,009 (0,001)	-0,016 (0,002)	-0,013 (0,001)	-0,02 (0,002)	-0,021 (0,001)	-0,026 (0,002)	-0,01 (0,003)	-0,011 (0,001)	-0,012 (0,002)	-0,011 (0)	-0,015 (0,002)	-0,01 (0,001)	-0,018 (0,002)
lkum_føp	-0,028 (0,026)	-0,009 (0,007)	-0,013 (0,02)	-0,023 (0,006)	-0,002 (0,02)	-0,041 (0,012)	-0,019 (0,021)	-0,059 (0,102)	-0,004 (0,005)	0,002 (0,023)	-0,007 (0,004)	-0,044 (0,019)	-0,005 (0,006)	-0,011 (0,028)
lkum_kth	-0,013 (0,004)	-0,005 (0,001)	-0,01 (0,002)	-0,014 (0,001)	-0,019 (0,004)	-0,019 (0,003)	-0,021 (0,005)	-0,011 (0,004)	-0,002 (0,001)	-0,002 (0,002)	-0,006 (0,001)	-0,01 (0,005)	-0,006 (0,002)	-0,021 (0,005)
lkum_revalidering	0,003 (0,012)	-0,006 (0,003)	-0,005 (0,01)	-0,012 (0,003)	-0,011 (0,004)	-0,01 (0,004)	-0,013 (0,011)	-0,005 (0,009)	-0,006 (0,002)	-0,013 (0,006)	-0,005 (0,002)	-0,012 (0,006)	-0,006 (0,002)	-0,021 (0,009)
lkum_sdp	-0,003 (0,005)	-0,003 (0,001)	-0,01 (0,003)	-0,007 (0,001)	-0,011 (0,002)	-0,01 (0,002)	-0,005 (0,003)	0 (0,004)	-0,003 (0,001)	-0,005 (0,002)	-0,004 (0,001)	-0,004 (0,002)	-0,008 (0,001)	-0,004 (0,002)
lrevalidering	-0,08 (0,051)	0,004 (0,011)	-0,056 (0,042)	-0,017 (0,011)	-0,089 (0,023)	-0,022 (0,016)	-0,135 (0,056)	-0,102 (0,063)	0,006 (0,009)	0,03 (0,028)	0 (0,008)	0,018 (0,029)	-0,001 (0,008)	-0,047 (0,045)
lsdp	-0,002 (0,012)	0,014 (0,002)	0,01 (0,008)	0,009 (0,002)	0,008 (0,005)	0,009 (0,004)	0,001 (0,006)	0,007 (0,009)	0,018 (0,002)	0,019 (0,005)	0,015 (0,001)	-0,002 (0,005)	0,017 (0,002)	-0,004 (0,004)
nyuddannet	-0,053 (0,036)	-0,106 (0,049)	-0,012 (0,007)	-0,042 (0,004)	-0,041 (0,006)	-0,024 (0,005)	-0,054 (0,006)	-0,028 (0,031)	-0,037 (0,031)	-0,01 (0,006)	-0,02 (0,003)	-0,012 (0,008)	-0,015 (0,003)	-0,065 (0,007)
sqrterfaring	0,003 (0,013)	0,046 (0,004)	0,015 (0,008)	0,025 (0,005)	0,055 (0,01)	0,037 (0,007)	0,018 (0,008)	0,016 (0,012)	0,007 (0,004)	-0,001 (0,006)	0,005 (0,004)	0,068 (0,011)	0,038 (0,004)	0,041 (0,009)
uddf1	0,078 (0,039)	0,005 (0,006)	-0,006 (0,015)	-0,006 (0,005)	-0,031 (0,014)	-0,001 (0,012)	-0,005 (0,017)	-0,039 (0,046)	-0,001 (0,005)	-0,003 (0,01)	0 (0,004)	-0,014 (0,014)	-0,009 (0,005)	0,012 (0,015)
uddf2	0,188 (0,142)	-0,008 (0,023)	-0,012 (0,027)	0,007 (0,025)	0,051 (0,043)	0,001 (0,029)	-0,056 (0,035)	0,034 (0,13)	0,026 (0,02)	0,003 (0,02)	-0,002 (0,02)	-0,126 (0,059)	-0,008 (0,015)	0,072 (0,033)
uddf3	0,031	0,012	0,003	0,011	-0,013	0,002	-0,017	-0,043	0,003	0,001	0	-0,013	-0,004	0,009

	(0,052)	(0,007)	(0,014)	(0,005)	(0,013)	(0,011)	(0,014)	(0,07)	(0,006)	(0,01)	(0,004)	(0,013)	(0,004)	(0,014)
uddf4	0,125	0,033	0,023	0,004	-0,008	0,014	-0,033	-0,072	0,009	-0,006	0,014	0,051	-0,005	-0,009
	(0,299)	(0,016)	(0,023)	(0,012)	(0,024)	(0,022)	(0,028)	(0,115)	(0,014)	(0,015)	(0,011)	(0,025)	(0,009)	(0,029)
uddf5	0,08	-0,001	-0,009	0,03	0,017	0	0,003	0,118	0,009	0,011	0,006	-0,011	-0,006	0,011
	(0,125)	(0,014)	(0,016)	(0,01)	(0,02)	(0,014)	(0,016)	(0,144)	(0,012)	(0,012)	(0,009)	(0,02)	(0,006)	(0,014)
uddf6	0,588	0,008	-0,015	0,016	0,033	0,01	0,004	0,264	0,014	0,006	-0,012	-0,002	-0,004	0,02
	(0,213)	(0,021)	(0,018)	(0,019)	(0,028)	(0,018)	(0,016)	(0,152)	(0,021)	(0,013)	(0,016)	(0,026)	(0,008)	(0,015)
uddm1	-0,08	0,014	0,009	0,006	-0,006	0,014	0,001	-0,047	0,012	-0,008	-0,001	0,006	0,007	-0,006
	(0,035)	(0,006)	(0,015)	(0,005)	(0,013)	(0,011)	(0,014)	(0,045)	(0,005)	(0,011)	(0,004)	(0,013)	(0,004)	(0,014)
uddm2	-0,073	-0,009	-0,019	-0,015	-0,019	-0,009	0,025	-0,356	0,009	0,01	0,016	0,003	0,011	-0,036
	(0,108)	(0,023)	(0,029)	(0,024)	(0,046)	(0,033)	(0,03)	(0,135)	(0,022)	(0,02)	(0,021)	(0,041)	(0,014)	(0,027)
uddm3	0,036	0,026	0,03	0,028	0,026	0,018	0,021	-0,02	0,012	0,012	0,019	0,019	0,01	-0,018
	(0,06)	(0,008)	(0,015)	(0,006)	(0,014)	(0,011)	(0,014)	(0,067)	(0,007)	(0,011)	(0,005)	(0,014)	(0,005)	(0,014)
uddm4	0	0	0,055	0,052	-0,005	-0,024	0,084	-0,018	-0,007	0,008	0,014	-0,016	0,027	0
	(0)	(0,02)	(0,025)	(0,016)	(0,029)	(0,024)	(0,025)	(0,21)	(0,019)	(0,017)	(0,014)	(0,027)	(0,01)	(0,023)
uddm5	-0,102	-0,002	0,013	0,019	-0,003	-0,009	0,003	-0,146	0,024	0,016	0,025	0,021	0,006	-0,011
	(0,088)	(0,011)	(0,017)	(0,009)	(0,019)	(0,014)	(0,015)	(0,155)	(0,01)	(0,012)	(0,008)	(0,019)	(0,006)	(0,015)
uddm6	-0,108	-0,012	0,024	0,026	-0,005	-0,06	0,002	.	-0,014	-0,018	-0,011	0,07	-0,012	0,008
	(0,268)	(0,031)	(0,023)	(0,03)	(0,051)	(0,03)	(0,024)	(.)	(0,03)	(0,018)	(0,036)	(0,048)	(0,013)	(0,02)

Appendiks B. Detaljeret gennemgang af simulering af skyggeløn

Udtrækning af u_i

Udtrækning af u_i kan enten gøres vha. af dens ubetingede fordeling givet ved (1.4) eller vha. af dens betingede fordeling, der afhænger af modellens parametre og en persons karakteristika. Ved anvendelse af den betingede fordeling opnås de bedste prædiktioner for $y_{it=2005}$.

Den betingede fordeling af u_i ($i=1, \dots, N$) er en normalfordeling, hvor middelværdien grundlæggende er gennemsnittet af de empiriske residualer for de år personen har observationer for den uafhængige variabel. Den betingede fordeling er vist i nedenstående resultat.

Resultat 1. Betinget fordeling for den individuelle effekt (random effect) i en AR(1) model

Der gælder, at den betingede fordeling for den individuelle 'random effect' u_i , givet y_{it}, x_{it} og de øvrige parametre er givet ved

$$(A1) \quad u_i | \beta, \rho, \sigma^2, \sigma_u^2, y_{it}, x_{it} \sim N \left(\left(\sum_{t \in \Omega_i} \frac{(1-\rho^{g_{it}})^2}{\hat{\sigma}_{g_{it}}^2} + \frac{1}{\hat{\sigma}_u^2} \right)^{-1} \sum_{t \in \Omega_i} \hat{\sigma}_{g_{it}}^{-1} (1-\rho^{g_{it}}) \hat{\eta}_{it}, \left(\sum_{t \in \Omega_i} \frac{(1-\rho^{g_{it}})^2}{\hat{\sigma}_{g_{it}}^2} + \frac{1}{\hat{\sigma}_u^2} \right)^{-1} \right)$$

$i=1, \dots, N$ og $\Omega_i = \{t > 1 \mid y_{it} \text{ findes}\}$

g_{it} er antallet af år indtil sidste observation før år t dvs., at $g_{it} = 1$ når der ikke er

'huller' umiddelbart før år t . Desuden defineres $\hat{\sigma}_{g_{it}} = \sigma \sum_{j=1}^{g_{it}} \rho^j$. De empiriske

normaliserede residualer er givet ved $\hat{\eta}_{it} = \sigma_{g_{it}}^{-1} y_{it}^{g_{it}} - \sigma_{g_{it}}^{-1} x_{it}^{g_{it}} \hat{\beta}$, hvor

$$y_{it}^{g_{it}} = y_{it} - \rho^{g_{it}} y_{it-g_{it}} \text{ og } x_{it}^{g_{it}} = x_{it} - \rho^{g_{it}} x_{it-g_{it}}$$

Bevis:

Observer først, at $\varepsilon_{it} - \rho \varepsilon_{it-g_{it}} = \sum_{j=1}^{g_{it}} \rho^j e_{it}$. For hvert t , hvor timelønnen findes i år

t og i mindst ét tidligere år, multipliceres (1.1) for den sidste tilgængelige observation med $\rho^{g_{it}}$, og fratrækkes (1.1). Derefter multiplicer med $\sigma_{g_{it}}^{-1}$:

$$y_{it} - \rho^{g_{it}} y_{it-g_{it}} = (x_{it} - \rho^{g_{it}} x_{it-g_{it}}) \beta + (1 - \rho^{g_{it}}) u_i + \varepsilon_{it} - \rho^{g_{it}} \varepsilon_{it-g_{it}}$$

$$y_{it}^g = x_{it}^g \beta + (1 - \rho^{g_{it}}) u_i + \sum_{j=1}^{g_{it}} \rho^j e_{it} \quad i=1, \dots, N \text{ og } t \in \Omega_i$$

$$(A2) \quad \sigma_{g_{it}}^{-1} (y_{it}^g - x_{it}^g \beta) = \sigma_{g_{it}}^{-1} (1 - \rho^{g_{it}}) u_i + v_{it} \quad \text{hvor } v_{it} \stackrel{iid}{\sim} N(0,1)$$

Dette angiver $T_i - \sum_{i=1}^{T_i} (g_{it} - 1)$ ligninger for hvert individ. Der tilføjes en observation for variansen af u_i i likelihood funktionen ($\hat{\sigma}_u^2 \sim \chi^2(N-1)$) hvilket svarer til regressionsligningen

$$(A3) \quad 0 = \sigma_u^{-1} u_i + \sigma_u^{-1} \omega_i$$

(A2) ligningerne omarrangeres og stakkes op med (A3):

$$\begin{Bmatrix} \cdot \\ y_{i,t}^u \\ \cdot \\ 0 \end{Bmatrix} = \begin{Bmatrix} \cdot \\ x_{i,t}^u \\ \cdot \\ \sigma_u^{-1} \end{Bmatrix} u_i + \begin{Bmatrix} \cdot \\ v_{i,t} \\ \cdot \\ v_{i,T_i+1} \end{Bmatrix} \quad i=1, \dots, N$$

Hvor $y_{i,t}^u = \sigma_{g_{it}}^{-1} (y_{it}^g - x_{it}^g \beta) = v_{it}$ og $x_{i,t}^u = \sigma_{g_{it}}^{-1} (1 - \rho^{g_{it}})$ eller

$$Y_i^u = X_i^u \tau_i + v_i \quad i=1, \dots, N$$

Dette er en standard OLS regression med kendt residualvarians og u_i er derfor

$$u_i \mid \beta, \rho, \sigma^2, \sigma_u^2, y_{it}, x_{it} \sim N\left((X_i^u{}' X_i^u)^{-1} X_i^u{}' Y_i^u, (X_i^u{}' X_i^u)^{-1}\right)$$

Middelværdien og variansen kan omskrives vha.

$$X_i^u{}' X_i^u = \sum_{t \in \Omega_i} \frac{(1 - \rho^{g_{it}})^2}{\hat{\sigma}_{g_{it}}^2} + \frac{1}{\hat{\sigma}_u^2}$$

$$X_i^u{}' Y_i^u = \sum_{t \in \Omega_i} \sigma_{g_{it}}^{-1} (1 - \rho^{g_{it}}) \hat{\eta}_{it}$$

Dette er identisk med det viste resultat. QED.

Udtrækning af $y_{it=2005}$

Udledning af fordelingen for y_{it} tager hensyn til, at residualen følger en AR(1)-proces, hvilket indebærer, at residualen i 2005 afhænger af residualerne i både tidligere og senere år.

Udgangspunktet er ekstrapolationer fra de nærmeste år, hvor der foreligger observation af den afhængige variabel. Der er her op to muligheder: 1) fremadrettet ekstrapolation fra det seneste tidligere år med en observeret timeløn (g_{it} år tilbage), og 2) bagudrettet ekstrapolation fra den næste efterfølgende år med en observeret timeløn (dvs. år 2006). På baggrund af 1) kan disse skrives som:

$$y_{it=2005} \mid t - g_i: \quad y_{it} = x_{it} \beta + u_i + \varepsilon_{it} \quad \text{med} \quad \varepsilon_{it} = \rho \varepsilon_{it-1} + e_{it}$$

$$y_{it} = x_{it} \beta + u_i + \rho^{g_i} \varepsilon_{it-g_i} + \sum_{j=1}^{g_i} \rho^j e_{it}$$

$$\text{Hvor} \quad \varepsilon_{it-g_i} = y_{it-g_i} - x_{it-g_i} \beta - u_i$$

$$y_{it=2005} | t+1: \quad y_{it} = x_{it} \boldsymbol{\beta} + u_i + \boldsymbol{\varepsilon}_{it} \text{ med } \boldsymbol{\varepsilon}_{it} = \rho \boldsymbol{\varepsilon}_{it+1} + e_{it+1}$$

$$y_{it} = x_{it} \boldsymbol{\beta} + u_i + \rho \boldsymbol{\varepsilon}_{it+1} + e_{it+1}$$

$$\text{Hvor } \boldsymbol{\varepsilon}_{it+1} = y_{it+1} - x_{it+1} \boldsymbol{\beta} - u_i$$

I de tilfælde hvor der kun er adgang til enten tidligere eller efterfølgende observationer for den afhængige variabel kan ovenstående ekstrapolationer for

$y_{it=2005} | t - g_i$ og $y_{it=2005} | t + 1$ anvendes direkte til udtrækning af $y_{it=2005}$. I tilfælde

hvor der både er tidligere og efterfølgende observationer kan fordelingen af

$y_{it=2005}$ udledes ved, for hver manglende observation, at betragte de to ovenstående

udtryk som en regression med to observationer og med y_{it} som parameter:

$$\begin{Bmatrix} x_{it} \boldsymbol{\beta} + u_i + \rho^{g_i} \boldsymbol{\varepsilon}_{it-g_i} \\ x_{it} \boldsymbol{\beta} + u_i + \rho \boldsymbol{\varepsilon}_{it+1} \end{Bmatrix} = \begin{Bmatrix} 1 \\ 1 \end{Bmatrix} y_{it} + \begin{Bmatrix} \sum_{j=1}^{g_i} \rho^j e_{it} \\ e_{it+1} \end{Bmatrix}$$

Det er herefter relativt ukompliceret at vise, at

$$y_{it=2005} \sim N \left(x_{it} \boldsymbol{\beta} + u_i + \frac{\rho \boldsymbol{\varepsilon}_{it+1} + \rho^{g_i} \left(\sum_{j=1}^{g_i} \rho^j \right)^{-2} \boldsymbol{\varepsilon}_{it-g_i}}{1 + \left(\sum_{j=1}^{g_i} \rho^j \right)^{-2}}, \frac{\sigma^2}{1 + \left(\sum_{j=1}^{g_i} \rho^j \right)^{-2}} \right)$$