

Fattigdom i Danmark

En socioøkonomisk fattigdomsgrænse

Detaljeret analyse og modelbeskrivelse,

Mårettet sociologer, økonomer og statistikere.

31.05.11

Kontaktperson:

Iulian Vlad Serban

Abstract English

The purpose of this project is to develop a quantitative statistical model over poverty in Denmark, which takes into account social mobility and social heritage. The project defines a new poverty threshold and carries out a discussion upon the pros and cons of this new poverty threshold and its applications to Danish policy making. The necessary statistical theory is deployed and a set of models and assumptions established to model the poverty explicitly and in a statistically central method. With these models the project measures the poverty development in Denmark over the past decade and the many factors contributing and influenced by poverty using data from the Luxembourg Income Study and the European Social Survey. Overall the project finds a major increase of poverty over the last decade, with strong indication of further increases, and further identifies a range of risk factors including unemployment, education and ethnicity.

Resumé Dansk

Projektets formål er at udvikle en kvantitativ statistisk model over fattigdom i Danmark som inddrager social mobilitet og social arv. Denne model benævnes SIM (*Social Indkomstmobilitets Model*). SIM definerer alle husstande under 50% medianen som værende fattige, mens husstande mellem 50 og 70% medianen bliver defineret som værende fattige hvis og kun hvis, de ikke kan opnå en indkomst over 70% medianen efter en 5-årig periode. Ud fra denne definition har vi opstillet en række antagelser, og har derved kunne modellere fattigdommen i Danmark ved brug af markov modeller og logistisk regression. Hertil er konstruktionen af samtlige modeller argumenteret ud fra tidligere forskning, sociologisk teori og statistisk teori for i sidste ende, at kunne tolke vores resultater nøjagtigt og undgå bias. Data fra European Social Survey og Luxembourg Income Study er blevet anvendt, således at modellerne er centrale, og resultater for fattigdommen er blevet beregnet. Det er således lykkedes projektet, at måle både sociale og økonomiske faktorer i forbindelse med fattigdom samt, at inddrage og analysere betydningen af social mobilitet og social arv i det danske samfund. De beregnede resultater har vist os, at fattigdommen i Danmark har været kraftigt stigende siden 2002, og at også den sociale mobilitet har været kraftigt faldende siden 2002. Vi er desuden kommet frem til, at social arv, uddannelsesniveaue, arbejdsløshed og handicap er statistisk centrale faktorer for fattigdom, hvorfor de bør inddrages i enhver analyse af fattigdom. Særligt er kvinder og husstande med mange børn udsatte for fattigdom. Dertil har vi også fremskrevet fattigdomsudviklingen fem år fremad og kan deraf konkludere, at fattigdommen forsat vil sige, mens den sociale mobilitet for de fattigste i samfundet forsat vil falde. Dette sidste resultat har, sammen med vores analyse af årsagerne bag fattigdom og de traditionelle fattigdomsgrænser, indikeret skarpe ændringer i befolkningssammensætningen for fremtidens Danmark. Netop derfor er projektets forhåbning, at man fra politisk side bliver mere bevist om disse store ændringer, samt ikke kun baserer sig på de traditionelle 50% og 60%-medianer.

Projektet er udarbejdet i et samarbejde mellem Iulian Vlad Serban (IMF, Københavns Universitet) og Tænketanken Cevea, med feedback fra Nicolai Nygaard Petersen (Europæiske Studier, Copenhagen Business School). I forlængelse af denne rapport udarbejdes der også kortere analyser, med henblik på specifikke underemner. Disse vil blive publiceret gennem Tænketanken Cevea (www.cevea.dk).

Nøgleord: fattigdom, social mobilitet, social arv, ulighed, kvantitativ prognose, markov model, markov proces, logistisk model, logistisk regression, lineær regression, arbejdsløshed, fødsler, handicap, uddannelse, demografisk data, luxembourg income study, european social survey

Indholdsfortegnelse

Abstract ^{English}	2
Resumé ^{Dansk}	2
Formål	4
Teori og definitioner	5
Positiv social mobilitet	8
Grundlæggende model	9
PSM model	10
Spatial-korrelation og kompensation af data	11
Data syntetisering	12
Analyse	14
Data	14
SSM Model	15
Arbejdsløshed	15
Handicap	19
Uddannelse	20
Fødsler	20
PSM Model	22
Opstilling af model	22
Anvendte data syntetisering	24
Estimation af variabler	25
Resultater: ESS	27
Resultater: LIS	30
Fremskrivning	32
Diskussion af resultater	34
Nye veje og videre arbejde	35
Proportional hazard model	35
Målefejl	35
Makroøkonomisk latente variable	35
Økonomisk udvidelse til modellen	35
Konklusion	36
Referencer	37

Formål

Projektets formål er at udvikle en kvantitativ statistisk model over fattigdom i Danmark. Formålet med denne model er primært at kunne definere, måle og fremskrive fattigdommen i Danmark. Det er endvidere projektets formål at sætte fokus på nødvendigheden af en officiel fattigdomsgrænse i Danmark og bidrage akademisk til sådan. Det er projektets sekundære formål, gennem modellen at kunne udpege og analysere de faktorer der har indflydelse på fattigdom, for derved at bidrage med nye løsninger og et mere nuanceret syn på fattigdom.

På et højere plan er det projektets formål, at præsentere nye metoder og modeller til modellering af fattigdom, samt en helt ny kvantitativ tankegang, der kan inspirere det fattigdomsfaglige miljø i Danmark. Formålet er særligt at illustrere hvordan avanceret statistik og kvantitativ tankegang kan give mere præcise og frugtbare resultater end traditionelt antaget, samt med uddybende forklaring vise hvordan de mest hyppige problemstillinger i denne sammenhæng kan løses.

Teori og definitioner

Projektets formål er at inddrage social mobilitet når man måler fattigdom i Danmark. Der er mange vigtige aspekter af social mobilitet og et utal af måder hvorpå disse kan inddrages. Afhængig de aspekter man vælger opfanger man forskellige grupper af befolkningen, og blandt dem man opfanger vægtes disse også forskelligt. Uanset hvilken metode man vælger risikerer man således at udelukke bestemte former for fattigdom. Det er projektets opfattelse at nyskabelse, anvendelse og gennemsigtighed af vores nye definition har højest prioritet. Definition skal således kunne anvendes separat til måling af fattigdom af fagfolk som ikke har deltaget i udarbejdelsen af denne. Definitionen skal endvidere være så gennemsigtig af den på et simplere niveau kan trække konklusioner vedrørende fattigdom og fattigdomsudvikling i Danmark. Den skal bl.a. øge viden i nuværende politiske debatter.

Alle disse krav til udformningen har ført os til at definere vores fattigdomsgrænse på en økonomisk baggrund. Fattigdomsgrænsen defineres derfor ud fra 50% og 70% medianen for disponibel indkomst i den danske befolkning. Medianen er bredt anvendt både i Danmark og internationalt regi. Den anvendes dog i sin mest rå format, nemlig som andelen af befolkningen der har en disponibel indkomst under enten 50% eller 60% under medianen. Denne primitive indikator har sine fordele og ulemper. En af fordelene er, at den er nem at måle og fortolke. Det er svært at diskutere resultaterne, og så længe man acceptere dens begrænsninger, kan man nå en række forholdsvis fornuftige konklusioner på baggrund af denne. Fattigdomsgrænsen har dog også en række ulemper. En af de meget diskuterede ulemper er dens relativitet. Relativiteten medfører nemlig, at fattigdomsgrænsen bliver påvirket af den øvrige udvikling i samfundet, selvom virkeligheden for den fattige befolkningsgruppe forbliver den samme. Grænsen har også svært ved at følge prisændringer. Eksempelvis vil et økonomisk opsving for middel- og overklassen kunne presse priserne langt opad proportionelt for essentielle varer, som f.eks. mad og bolig. Derudover ville grænsen også faktisk udvide den nedre befolkningsgruppe væsentlige mere end den økonomiske vækst umiddelbart påpeger. Et sidste og lige så væsentligt problem er, at økonomer og politikere, som vil reducere fattigdommen ud fra en relativ grænse, kun bliver stillet med et enkelt valg: at tage fra overklassen og fordele blandt underklassen.

Men den største ulempe i vores kontekst er uden tvivl, at den ikke måler sociale faktorer, og derfor særligt ikke måler social mobilitet i samfundet. Ved udelukkende at kigge på udviklingen af fattigdom i et økonomisk perspektiv med 50% eller 60% medianen, er det ikke muligt på nogen måde at udtale sig om social mobilitet. Der findes nemlig et væld af forskellige grupper i den fattigste gruppe. Der findes mennesker som kun er fattige i en kort periode, og mennesker som er fattige hele deres liv. Der findes fattige som ikke kan få job, fordi de er handicappede, og fattige som lever i udsatte områder og omgås ustabile mennesker. Særligt har antallet af ufaglærte arbejdspladser faldet hurtigt over det sidste årti og således yderligere marginaliseret en stor befolkningsgruppe i Danmark. Dertil er en betydelig del af de fattige mennesker ikke permanent fattige, men kun midlertidigt fattige. De fluktuere over og under medianen 50% og 60% medianen. Mens andre kun med nød og næppe forbliver over den.

Ved udelukkende at kigge på sådan en indkomstmedian mister vi hele denne underliggende dynamik, og bliver altså stillet med et urealistisk og i høj grad rigid syn på fattigdommen.

Netop af disse årsager har vi valgt, at inddrage social mobilitet i vores fattigdomsgrænse. Som nævnt ovenfor defineres vor fattigdomsgrænse ud fra 50% og 70% medianen. Vi definerer, at et individ lever i fattigdom hvis en af følgende to betingelser gør sig gældende:

1. Han/hun har en disponibel indkomst under 50%-medianen
2. Han/hun har en disponibel indkomst mellem 50%- og 70%-medianen
Og han/hun kan ikke over en 5-årig periode opnå over en indkomst over 70% medianen

Disse to betingelser skal ses i lyset af, at størstedelen af den danske befolkning bor sammen med andre, og således i mere eller mindre grad har fællesøkonomi. Så selvom vores fattigdomsgrænse går på individer, så måles fattigdom reelt for hver enkelt husstand i Danmark. Det vil dog også være forkert at sige, at vi kun måler fattigdom på husstandsniveau. Det kan f.eks. være at et ægteskab går i opløsning, og at det der før var en husstand med to voksne og børn, som resultat kun har en voksen tilbage. Den gamle husstand bliver altså delt op i to nye husstande, hvoraf fattigdommen nu skal måles for dem hver især. Vi måler altså reelt fattigdommen for hver enkelt husstand, men tager hensyn til nye dannelser og separation af husstande. Dette uddybes senere.

Hvilke fordele og ulemper arver vores nye fattigdomsgrænse så fra 70% medianen? Vores fattigdomsgrænse er givetvis stadig relativ, men strukturen har ændret sig grundlæggende. Kort sagt udjævner vi den største relativitet der findes i medianen. En procentvis ændring i medianen vil direkte afspejles i den primitive 50% eller 60% median. Men en procentvis stigning eller fald i vores fattigdomsgrænse ændrer primært kun på de individer vi analyserer. Vi ville altså få enten færre eller flere individer vi analyserer, men det afgørende er stadigvæk hvorvidt de kan opnå en indkomst over 70% medianen eller ej over en 5-årig periode. Det er klart, at de individer som ligger lige under medianen indkomstmæssigt har en særdeles høj sandsynlighed for at arbejde sig over den i en 5-årig periode, i modsætning til dem som ligger langt under medianen. En stigning i 70% medianen, vil altså få en masse flere mennesker med i vores analyse, men størstedelen af disse vil sandsynligvis kunne arbejde sig over den igen over en 5-årig periode, hvorfor den relative effekt er meget mindre end hvis vi blot havde kigget på 50%-medianen. Hertil skal det også tages i betragtning, at ændringer i medianen hovedsagligt skyldes ændringer i lønninger. Større ændringer her vil således også medføre ændringer for de individer der er på arbejdspladsen, og vi får således differentieret stærkere mellem individer på arbejdsmarkedet og udenfor. Det betyder ikke at individer udenfor arbejdsmarkedet vil blive vægtet unødvendigt højere. Men netop fordi vi arbejder med en 5-årig periode, vil individer som kun lige er blevet arbejdsløse (f.eks. grundet den økonomiske krise) ikke vægte mere. De individer vores definition ender med at vægte højt, er de individer der over længere tid ikke kan komme i arbejde, og risikerer dermed at leve en længere del af deres liv i fattigdom, hvilket præcis stemmer overens med vores intention om at inddrage social mobilitet. Vores fattigdomsgrænse er derfor på en række punkter en markant mere nøjagtig indikator for social og strukturel fattigdom, og kan derfor på et reelt og nyt niveau anvendes til at analysere og skabe ændringer i samfundet.

Vores fattigdomsgrænse er dog stadig følsom overfor prisændringer, eftersom disse ikke afspejles i den sociale mobilitet. Så selvom den i højere grad fanger social og strukturel fattigdom, kan den stadig påvirkes af det øvrige samfund, hvorfor den også skal anvendes med agtpågivenhed.

En sidste og væsentlig ulempe er dog, at vores fattigdomsgrænse er mere kompliceret end 50% og 60% mediangrænsen. Den er naturligvis forsat gennemsigtig, og fra et akademisk synspunkt simpel, men det kræver væsentligt mere energi at forstå hver eneste underkomponent i sociologisk forstand. Dette skyldes, at der mange måder man kan vurdere hvorvidt en husstand statistisk set kan opnå en indkomst over 70% medianen. Hvilke variabler vælger man, f.eks. at inddrage? Hvordan skal sammensætningen i husstanden vægtes? Dertil er det også mere omfattende at foretage ændringer og udvidelser. Men det er et kompromis vi tilfredsstillende foretager for at høste vores største fordel: at måle social mobilitet, og derved få et mere nuanceret syn på fattigdom.

Som tidligere diskuteret opfanger vi væsentlige aspekter af social mobilitet fra et økonomisk perspektiv. Vores definition opfanger i langt højere grad den sociale og strukturelle fattigdom i Danmark. Richard Layte et al.^{[3][5]} støtter vores formodning om, at fattigdom skal ses i et dynamisk perspektiv, da antallet af fattige er langt højere end paneldata umiddelbart indikerer. De kommer frem til, at der estimeres omkring dobbelt så mange fattige ud fra longitudinale data end ud fra paneldata i forbindelse med 50% indkomstmedianen, hvilket er en enorm forskel. Da longitudinale data er mest detaljerede, er der derfor ingen tvivl om, at den reelle fattigdom underestimeres som en simpel indkomstmedian af paneldata. De og en række andre forskere, heriblandt Didier Fourage^[4], kommer frem til, at det er utrolig vigtigt, at man fra politisk side arbejder med fattigdom på en nuanceret plan, og særligt tager hensyn til de underliggende processer såvel som dynamikken. De efterviser hertil, at man udelukkende ud fra de dominerende politiske tendenser (regime-types), som f.eks. størrelsen på velfærdsstaten og liberaliseringen i samfundet, kan modellere en række af de vigtigste dynamiske faktorer som påvirker fattigdom.

Vi vil gerne understøtte vores fattigdomsgrænse med mere grundlæggende sociologisk teori. Deepa Narayan og Patti L. Petesch^[6] præsenterer en bred sociologisk teori for fattigdom, det såkaldte "Empowerment Framework", som bygger på det fattige individs evne til at påvirke egen situation og opnå højere indkomst. Teorien har en antagelse om, at de fattigste mennesker gerne vil forbedre deres egne levevilkår, men på grund af både personlige og strukturelle årsager i samfundet ikke kan gøre dette. Det er præcis denne evne til at forbedre egen økonomisk stilling vi måler på fra et kvantitativt synspunkt. Derudover argumenterer de, bl.a. ud fra sociologisk teori og tidligere forskning, at social arv og uddannelse er vigtige faktorer i disse processer. Netop denne teori vil kunne lægge et godt fundament for vores fattigdomsgrænse. At fattige mennesker har incentiv på personligt plan til, at påvirke egen situation understøttes også af Richard Layte et al. Den interesserede læser bør derfor selv finde vores reference og sætte sig ind i materialet. Det skal dog bemærkes, at adskillige sociologiske teorier vil kunne understøtte vores definition af fattigdom, og at man altid bør huske på at fattigdom er et meget kompliceret emne.

Det endelige resultat af vores fattigdomsgrænse bliver forhåbentligt, at økonomer, sociologer og politikere vil kunne arbejde langt mere nuanceret med vores fattigdomsgrænse end det før har været muligt med 50% og 60% medianen. Hertil vil man både kunne analysere udviklingen af fattigdom over tid, men også de faktorer som har indflydelse på fattigdom, og hvordan de indbyrdes spiller sammen. Disse fordele træder mere markant frem senere hen og vil derfor blive diskuteret løbende, med opsummering til sidst. Vi vil fremover i denne rapport døbe vores fattigdomsgrænse: *Den Sociale Indkomstmobilitets Model (SIM)*.

Nedenfor føres den teoretiske og definitions-mæssige behandling af vores fattigdomsgrænse.

Læsere bør bemærke, at størstedelen af den matematisk-statistiske teori er blevet ført over i appendikset.

Positiv social mobilitet

Den *positive sociale mobilitet* (PSM) defineres som den procentvise sandsynlighed for, at en husstand over en 5-årig periode kan opnå en indkomst som overstiger 70%-medianen for den respektive husstands type. Denne sandsynlighed benævner vi fremover som *indkomstmobilitet*. Faktorerne, som indgår i udregning af PSM, opdeles i to kategorier. Husstandsvariablerne omfatter:

- Husstandstype
Antal voksne, børn og hvorvidt husstandshoved bor med partner
- Indkomstgruppe
0 - 110.000, 110.000 - 147.999, 148.000 - 185.999, 186.000 - 225.999, 226 000+

Individvariable forefindes kun for husstandens hoved og dennes partner. Variablerne omfatter:

- Køn
Mand, kvinde.
- Aldersgruppe
0-4, 5-9, 10-14, 15-19 ...
- Helbred
Hvorvidt vedkommende er fysisk hæmmet for at arbejde
- Uddannelsesniveau
Folkeskole uddannelse,
Gymnasial eller erhvervsfaglig uddannelse
Videre gående uddannelse
- Erhverv
Hvorvidt individet er i arbejde eller ej.
- Forældres uddannelsesniveau (kun for husstandshovedet)
Folkeskole uddannelse,
Gymnasial eller erhvervsfaglig uddannelse
Videre gående uddannelse

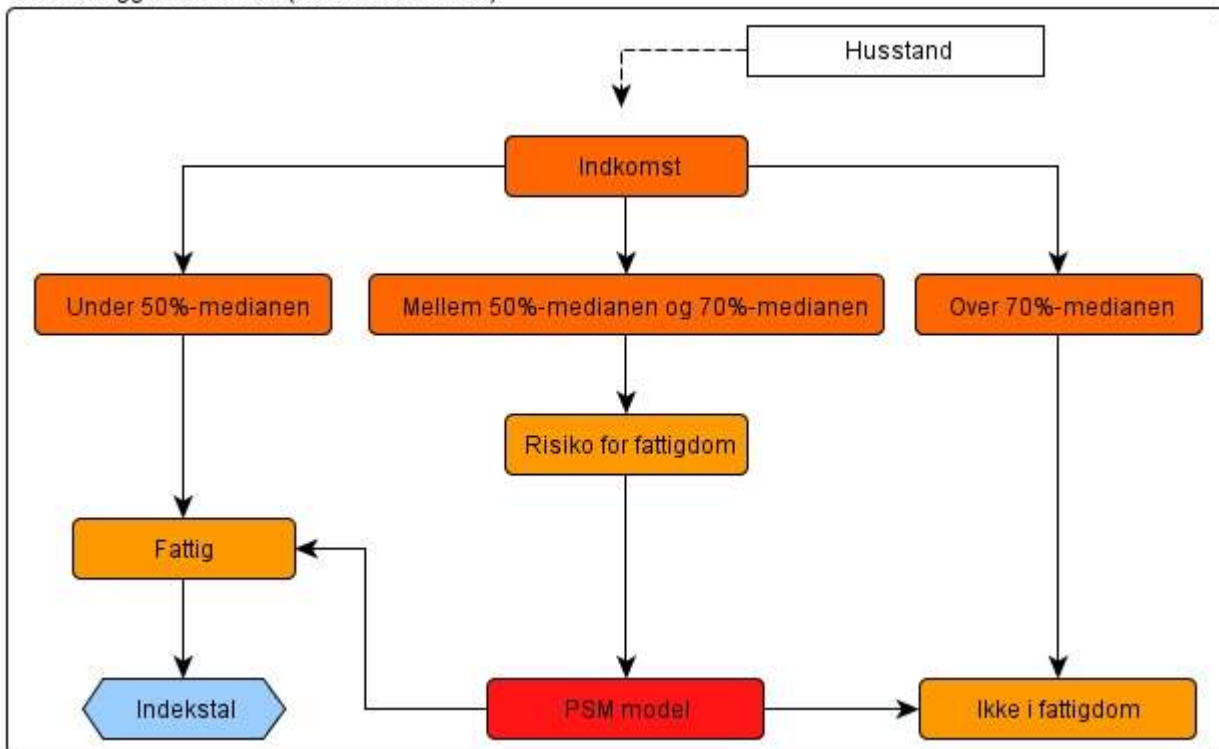
Man kunne naturligvis vælge at betragte en række andre variable også. Dette skyldes tre grunde. Den første årsag er den manglende tilgængelighed af data, både for projektets adgang og internt hos de institutter som ellers er ansvarlige for at udarbejde disse data. Den anden årsag er deres modellering. Da der ikke er særlig meget data til rådighed, ville for mange variable øge chancen for at fejlestimerer fattigdom. Den tredje årsag er korrelation. Ved at indføre variable, som ikke bidrager med ny væsentlig information, ender vi med at have et kolinearitetsproblem, hvilket fra statistisk side er uønsket. Denne tredje årsag støder vi dog igen ind i senere hen, men uden at den har væsentlig indflydelse på vores endelige resultater.

Grundlæggende model

Ud fra ovenstående definition kan modellen opstilles som nedenstående diagram.

Grundlæggende model (Husstandsniveau)

(A.1)

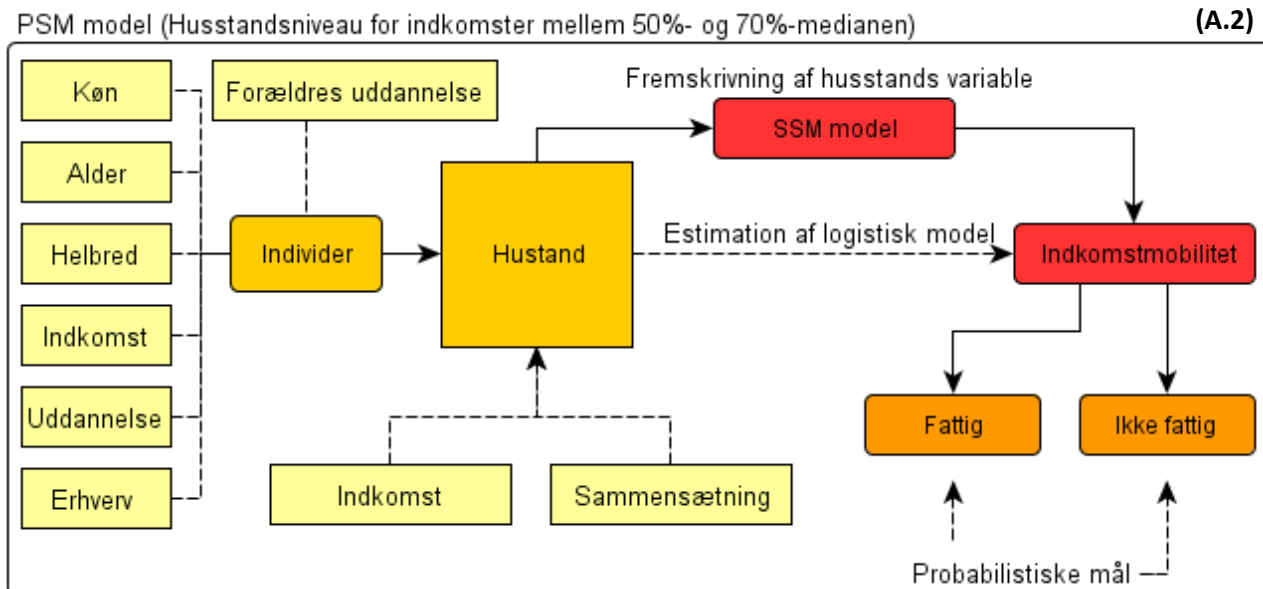


Alle udregninger foretages på husstands niveau. Som en analogi kan man anskue modellen som et rullebånd. Husstande kommer ind øverst i diagrammet og deres indkomstniveau bliver vurderet. Såfremt deres disponible indkomst er lavere end 50%-medianen, bliver de klassificeret som fattige, og hvis deres disponible indkomst er over 70%-medianen bliver de udeladt. De husstande som ligger mellem 50%- og 70%-medianen bliver analyseret med PSM modellen. PSM modellen, som bliver beskrevet i næste afsnit, udregner herefter en probabilistisk sats for at den givne husstand befinder sig under 70%-medianen efter 5 år. Denne sats kunne f.eks. være 10% for en bestemt husstand, hvilket betyder at der er 10% chance for at netop denne husstand lever under 70%-medianen efter 5 år. I eksemplet vil husstanden således tælle som 0,1 husstande i fattigdom. Til sidst vil summen af alle disse probabilistiske satser blive lagt sammen, adderet med antallet af husstande under 50%-medianen, og derved har vi vores nye fattigdomsgrænse *SIM* (*Social Indkomstmobilitets Model*).

For at lave klassificere husstande over og under henholdsvis 70%- og 50%-medianen, vil vi foretage den statistiske antagelse, at fejl i observationerne af disponible indkomst er normalfordelte og centreret omkring nul.

PSM model

Ud fra vores grundlæggende model vil vi opstille PSM modellen som nedenstående diagram:



Alle udregninger foretages forsat på husstands niveau, hvorunder individ variable også medtages samlet for husstanden. Denne model skal forstås som en input-output model. Modellen opstilles først, og samtlige variable bliver brugt til at beskrive *indkomstmobilitet*, det vil sige, sandsynligheden for at husstanden efter en 5-årig periode opnår en samlet disponibel indkomst som overstiger 70%-medianen. Efter modellen er opstillet anvendes den på den enkelte husstand.

Husstandens variable fremskrives af stokastiske processer for de enkelte variable, ved brug af *den stokastisk sociale mobilitet* (SSM) model. SSM modellen udvikles og beskrives i løbet af analysen. Den bygger hovedsagligt videre på tidligere arbejde og formaliseres statistisk gennem markov kæder. See appendikset for en teoretisk diskussion af markov kæder.

PSM modellen opstilles formelt med inspiration fra (Chaudhuri et al, 2002)¹⁾. Vi definerer husstand h 's disponible indkomst til år t som en reelt stokastisk variabel $C_{h,t}$. Definer endvidere $k_{n,t}$ til at være 70%-medianen til året t .

Med antagelse at prisindeks, lønindeks og den overordnede makroøkonomiske situation i Danmark ikke ændrer sig drastisk på under et år, vil vi betragte værdierne til året t , som værende et gennemsnitligt udtryk for hele året. Dette lader til at være en relativ mild antagelse.

Vi kan nu definere vores *indkomstmobilitet* som følgende sandsynlighed:

$$I_{h,t} = P(C_{h,t} \geq K_{n,t}) \quad (1.1)$$

Vi vil gerne beskrive den disponible indkomst ved brug af vores andre variable:

$$C_{h,t} = c(X_{h,t}, \alpha_{h,t}, e_{h,t}) \quad (1.2)$$

Her definerer $X_{h,t}$ en vektor bestående af samtlige variable for den enkelte husstand som beskrevet tidligere, men uden variabelen for disponibel indkomst. $\alpha_{h,t}$ definerer en vektor af parametre i tilknytning til $X_{h,t}$, og $e_{h,t}$ fejlobservationen for den enkelte husstand. Om fordelingen af $e_{h,t}$ antager vi, at fejlen er normalfordelt omkring nul. Funktionen c vil betragtes som ukendt, og bruges ikke direkte i projektet. I stedet er det antagelsen af, at der findes en sådan funktion c , som retfærdiggør vores model. (Chaudhuri et al, 2002) fører en længere diskussion og har en række referencer som støtter denne antagelse. Kort sagt bygger vores teori på, at der findes en underliggende sociologisk model, som ud fra vores observerede variable, kan bruges til akkurat at modellere indkomsten for den enkelte husstand over tid. Hvad denne sociologiske model præcis er, er irrelevant. Men dens eksistens berettiger følgende videre arbejde.

Det største og måske vigtigste skridt for hele projektet foretages nu. Vi antager at $I_{h,t}$ kan modelleres som en logistisk regression af samtlige variable i vektoren $X_{h,t}$. Det betyder, at $I_{h,t}$ får følgende udtryk:

$$I_{h,t} = 1 - P(C_{h,t} \geq K_{n,t}) = P(C_{h,t} < K_{n,t}) = \frac{1}{1 + \exp(\beta X_{h,t})} \quad (1.3)$$

Hvor β er vektoren indeholdende regressionskoefficienterne for den logistiske regression. Disse estimeres ud fra mikrodata over de enkelte husstande ved brug af maximum likelihood estimation (MLE).

Hvad betyder alle vores ligninger og antagelser for modellen? For reelt at opstille en ny fattigdomsgrænse skal den naturligvis formaliseres statistisk. For at gøre dette er vi nødt til at foretage en række antagelser. Funktionen c udtrykker de sociale og økonomiske processer en husstand gennemgår. I teorien afhænger funktionen kun af den enkelte husstand. Ideelt bør den afhænge af alle andre husstande, særligt husstande i samme nabolag, samt arbejdsmarkedet og den makro-økonomiske situation i Danmark. Det gør den ikke, og som konsekvens afhænger vores indkomstmobilitet også kun af den enkelte husstands variable. Vi tænker således, i baner af sociologiske processer. En husstand, som oftest består af en samlet familie, udvikler sig: individerne får højere uddannelse, skifter job, bliver handikappede, får børn osv. Givet at vores funktion c kan beskrive husstandens indkomst fuldstændigt, er det en naturlig udvidelse, at antage at husstandens indkomstmobilitet kan beskrives som en logistisk regression over de selv samme variable. Se appendikset for en længere diskussion af de antagelser og konsekvenser den logistiske regression implicerer.

Spatial-korrelation og kompensation af data

Da vi arbejder med demografiske oplysninger, og særligt afgrænser os til en mindre befolkningsgruppe i PSM, har vi spatial-korrelation. Spatial-korrelation vil i korte træk sige, at individer med små forskelle i indkomster vil have de samme sandsynlighedsfordelinger for andre variable som f.eks. boligforhold, uddannelsesniveau osv. Det betyder også, at der hvor der er mangel på mikrodata, kan vi ud fra spatial-korrelation syntetisere resultater fra andre studier. Ved eksempelvis at afgrænse sig husstande som lever under 70%-medianen, kan projektet konstruere betingede sandsynlighedsfordelinger for positiv social mobilitet. Dette vil blive anvendt i både SSM og PSM modellen. Hovedsagligt vil anvendelsen dog ligge i data syntetiseringen til PSM.

Data syntetisering

I analysen har vi data fra 4 surveys, udført i årene 2002, 2004, 2006 og 2008. Før vi reelt kan analysere fattigdomsudviklingen er vi nødt til, at sammensætte disse på meningsfuld og statistisk korrekt måde. Vi vil derfor lave to centrale antagelser. Først og fremmest vil vi antage, at der i den virkelige verden, for et hvilket som helst år, findes grupper af fattige mennesker. Disse kan være alt lige fra etniske minoriteter til bolig kvarterer i København. Her er det igen eksistensen der er vigtig, mens de enkelte gruppers sammensætning er irrelevante. Hertil antager vi at man kan måle antallet af fattige mennesker i sådan en gruppe:

$$O_{g,t} = F_{g,t} + e_{g,t} \quad (2.1)$$

Hvor $O_{g,t}$ er antallet af observerede fattige til en given gruppe g og tid t , $F_{g,t}$ er det reelle antal fattige, og $e_{g,t}$ er en normalfordelt fejl. Vi skriver det op på denne måde for, at understrege at vi kun antager måle fejlen er normalfordelt. I en informationsperfekt verden ville der ikke være nogen måle fejl, og vi ville kende antallet af fattige præcist. Normalfordelingsantagelsen er altså ikke altafgørende, men et praktisk redskab til at gribe usikkerheden an med. Hvis vi nu vil måle alle fattige i Danmark kan vi skrive:

$$\sum_{g \in G} O_{g,t} = \sum_{g \in G} F_{g,t} + e_{g,t} = \sum_{g \in G} F_{g,t} + \sum_{g \in G} e_{g,t} \quad (2.2)$$

Altså bliver det endelige antal fattige normalfordelt med middelværdi som det reelle antal fattige mennesker og en varians der afhænger af $e_{g,t}$ 'erne. Men yderligere kan vi splitte resultatet op:

$$\sum_{g \in G} O_{g,t} = \sum_{g \in G_1} O_{g,t} + \dots + \sum_{g \in G_n} O_{g,t} \quad (2.3)$$

Hvor G_1, \dots, G_n er indbyrdes disjunkte, og deres forening er hele G , det vil sige hele befolkningen. Men vi kan nu vælge G_1, \dots, G_n til at være præcis vores surveys. Det vil sige, at vi fra hver eneste survey kan udregne antallet af fattige mennesker, og ved at enhver fejl er normalfordelt med middelværdi nul og en ukendt varians. Desuden ved vi, at vores surveys er udarbejdet uafhængigt, at også deres normalfordelte fejl må være uafhængige.

Yderligere vil vi gerne syntetisere vores data på tværs af tid. Her kommer den næste vigtige antagelse ind i billedet. Givet, at vi har makro-socioøkonomiske variable til rådighed for alle årene, vil vi antage at vi kan vægte data mht. til disse, og at enhver fejl stadig vil hver normalfordelt. Formelt lader vi F være den funktion som transformerer vores data, og G være den aggregerede funktion fra (1.3). Antagelsen bliver:

$$G \circ F(O_{g,t}, \varpi_{t,T}, T) = O_{g,T} + E_{g,t,T} = F_{g,T} + e_{g,T} + E_{g,t,T} \quad (2.4)$$

Hvor $\varpi_{t,T}$ er vores makro-socioøkonomiske variable, der knytter tidsperioden t til T , og $E_{g,t,T}$ er en ny fejl, som er normalfordelt med middelværdi nul og ukendt varians. Antagelsen er således, at vi ud fra en enkelt måling, kan fremskrive resultatet på tværs af tid, såfremt vi har disse makro-socioøkonomiske variable til rådighed. Vi skal blot tage højde for at der er en ny varians på fejlen, som formodentligt er større.

Ved at bruge denne teori, har vi sammen med vores makro-socioøkonomiske variable til hvert eneste år en række samples fra en normal fordeling. Givetvis vil der forventes at være en stor varians på dette, men dette bliver der taget højde for i variansen vi udregner til vores normalfordeling. Hvis vores vægtning mht.

makro-socioøkonomiske variable er usikker, vil variansen tilsvarende vokse, og udvide vores sikkerhedsintervaller. Desuden har vi per antagelse og teoretiske udregninger, at normalfordelingen er central for det reelle antal fattige for *SIM*. Vi har dermed en central og solid statistisk model, og kan nu ud fra vores survey, udregne variansen for den overordnede normalfordeling, og har dermed et udtryk for hvor korrekt vores fremskrivning af fattigdom på tværs af tid er, og hvor meget usikkerhed der overordnet er i modellen. Da vi kun har få samples, vil usikkerheden blive estimeret ved brug af t-fordelingen:

$$\left(X_t - z_{1-\alpha} \frac{S_n}{\sqrt{n}}, X_t - z_{1-\alpha} \frac{S_n}{\sqrt{n}}\right) \quad (2.5)$$

Hvor X_t er den estimeret gennemsnit for fattigdommen, $z_{1-\alpha}$ er $1 - \alpha$ fraktilen i t-fordelingen, og S_n er den estimerede standardafvigelse. Det reelle antal fattige mennesker vil altså forventes at ligge i dette interval.

Analyse

Data

Projektet anvender nedenstående data kilder:

European Social Survey

European Social Survey (ESS, <http://ess.nsd.uib.no>) er et dataarkiv som EU udgiver hvert andet år, med fokus på de sociale forhold i de europæiske lande. Her forefindes mikrodata fra 2002, 2004, 2006 og 2008 om Danmark, med variabler på husstands niveau, som omfatter indkomst, uddannelse, alder, social arv, etnisk baggrund, sociale værdier og politiske værdier. Variablerne i dette datasæt er langt mere fyldestgørende end LIS datasættet, men disse data indeholder kun data for ca. 1200 husstande per datasæt. Som konsekvens af dette vil hvert datasæt kun repræsentere ca. 150 individer som lever under 70%-medianen fattigdomsgrænsen i de respektive år. Vores analyse kommer samlet set til, at bygge på ca. 1200 husstande for hvert år, hvoraf 600 bliver analyseret og vurderet hvorvidt de kan opnå en indkomst over 70%-medianen over en 5-årig periode. Dette medfører en stor usikkerhed, men det vil der blive taget hensyn til både teoretisk og praktisk. Udover dette er datasættene af høj kvalitet og vil derfor være velegnede til vores formål. Data er ikke korrigeret. Se appendikset for data karakteristika.

The Luxembourg Income Study

The Luxembourg Income Study (LIS, www.lisproject.org) er et internationalt data arkiv og forskningsinstitut i Luxembourg. LIS har adskillige databaser, heriblandt to med mikrodata: LIS Databasen og LWS Databasen. Projektet fået godkendt adgang til begge databaser, men vil kun benytte LIS Databasen. LIS Databasen indeholder mikrodata fra Danmark fra årene 2004, 2000, 1995, 1992 og 1987 som beskriver husstande (og enkeltpersoner) i forhold til indkomst, uddannelse, alder og en række andre variable. Databasen indeholder dog ingen variabler relevant for social arv, hvorfor projektet ikke kan anvende dette datasæt til at beskrive fattigdom i forhold til social arv. Data som er tilgængeligt strækker sig imidlertid over 1/30 af den danske befolkning og bygger på registerdata. Det har sine ulemper og fordele. At datasættet er så stort, medfører at vi har en minimal usikkerhed på vores beregninger. På den anden side er der en enorm systematisk fejl i datasættet. Ca. 35% af alle husstandene har negative indkomster, hvorfor de er blevet kasseret under analysen. Det kan have en stor indflydelse på vores fattigdomsgrænse SIM, hvorfor vi kun bruger LIS data til at estimere og analysere faktorernes indflydelse på fattigdom. Datasættene er dog ikke yderligere korrigeret, og forventes derudover, at være af acceptabel kvalitet. Se appendikset for data karakteristika.

Danmarks Statistik (DST), OECD Statistics, Eurostat og LABORTA

Hverken af de ovenstående organisationer stiller mikrodata til rådighed for projektets vedkommende. De bliver dog alle sammen anvendt til, at estimere SSM modellerne, der med nogle antagelser kun kræver data på makroniveau. Endvidere bliver Eurostat data brugt til at vægte vores data over tid. Data fra samtlige af disse organisationer er af høj kvalitet, og altid i et langt højere antal samples end vi reelt har behov for. Den eneste kritik for, at anvende disse datakilder er, at de har forskellige ækvivalensskalaer og definitioner af fattigdom, arbejdsløshed, sygdom osv. Projektet har undervejs så vidt muligt korrigeret og antaget sig ud af dette, men den kritiske læser bør læse appendikset og så vidt muligt selv gennemgå datakildernes dokumentation.

SSM Model

I SSM modellen fremskriver vi følgende variabler stokastisk for hver husstand:

- Arbejdsløshed
- Uddannelse
- Handicappet
- Børn

Samtlige variabler bør have en stor indflydelse på indkomsten i en given husstand, og vi analyserer dem derfor nøje. Af hensyn til anvendeligheden af projektet starter vi med en udførlig gennemgang af modellen for arbejdsløshed. Her vil den overordnede fremgangsmåde og samtlige vigtige antagelser blive diskuteret med tilhørende eksempler. Efter dette vil de resterende tre modeller kun blive beskrevet i korte træk, da de følger samme fremgangsmåde og i mere eller mindre grad samme antagelser.

Hertil vælger vi, at opstille SSM modellen som udgangspunkt for 2009. Reelt kunne vi godt have opstillet SSM modellen for alle årene 2002 – 2009, som vi planlægger at analysere, men sådan et arbejde ville være enormt, gøre vores resultater uigennemskuelige, og samtidig gøre det sværere at sammenligne årene indbyrdes. Derfor har vi valgt, at opstille SSM modellen kun for året 2009. Det betyder selvfølgelig, at der er mere usikkerhed på fattigdommen omkring årene 2002 – 2004, men denne usikkerhed forventes at være langt mindre end de egentlige resultater.

Se også appendikset for en behandling af dødsfald, separation og ægteskab blandt husstande med laveste indkomster og uddannelser.

Arbejdsløshed

Arbejdstilstand

Da projektet ikke handler om processer på det danske arbejdsmarkedet, men i stedet deres tilknytning til husstande og fattigdom, vil vi forsimple billedet. Vi vælger derfor at se bort fra alle former for flex-arbejde og tilsvarende. Vi deler altså verden op i sort og hvid, og definerer at et individ enten er i arbejde eller er arbejdsløs. Er dette en fornuftig antagelse? Det statistiske problem i denne antagelse ville være i form af bias. Det kunne f.eks. tænkes at flex-arbejde var mere udbredt blandt de laveste indkomstgrupper, og at en sådan sort-hvid antagelse vil skabe et bias i forhold til hvor mange fattige vores model opfanger.

For at undersøge dette benytter vi OECD Data. OECD klassificere fuldtidsarbejde som arbejde over 30 timer om ugen og resten bliver klassificeres som deltid. Ifølge OECD Data var der i 2008 1.902.574 mennesker i fuldtidsarbejde i Danmark og kun 208.147 i deltid når man kigger på aldersgruppen 30 – 64 år⁹⁾. Dette betyder at der kun er 10,94% mellem 30 - 64 år som arbejder deltid. Derudover kontrollerer vi også for handicap gennem vores data. Betragt nu et scenario med et individ som har en meget lav indkomst, samt ikke er registreret som havende fysiske eller psykiske handicap, hvad er da chancerne for at hans lave indkomst skyldes at han har et flex-arbejde? Fra intuitiv logik er individet enten entreprenør, har lige fået børn, eller af anden grund ikke har haft tid til at arbejde fuldtid. Uanset hvordan så vil de i langt de fleste tilfælde være et valg individet selv har foretaget eller som er blevet tvunget ned over individet i en midlertid kontekst. Derfor viser vores logiske resonans kombineret med de givne data, at andelen af fattige mennesker på deltidsarbejde må være ubetydeligt lille, og derfor ikke bør påvirke vores analyse med bias.

For at arbejde med variablen arbejde stiller vi to spørgsmål:

- Givet at en person er arbejdsløs, hvad er da hans/hendes sandsynlighed for at være i arbejde efter en given tidsperiode?
- Givet at en person er i arbejde, hvad er da hans/hendes sandsynlighed for efter en given tidsperiode at være arbejdsløs?

Læseren bør passe på med at læse disse spørgsmål for hurtigt. Spørgsmålene er nemlig meget præcise. Vores data fra ESS fortæller os ikke hvor længe et individ har været i arbejde eller har været arbejdsløs.

Vi er derfor tvunget til at tage udgangspunkt i et stationært punkt tidsmæssigt. Derudover sætter vi ingen begrænsninger på individets muligheder. Det kan f.eks. godt være at et individ efter en periode med arbejdsløshed kun finder et midlertidigt arbejde, eller at individet finder et permanent arbejde i en helt anden branche. Det kan endda være at individet finder et arbejde, mister dette og finder et nyt arbejde. Det essentielle er kun hvorvidt individet er i arbejde igen den givne tidsperiode i statistisk forstand. Alle andre problemer bliver så at sige "udjævnet" af de statistiske vægtninger vi foretager senere.

For at besvare disse spørgsmål kontrollerer vi for tre vigtige faktorer:

- Køn
- Alder
- Højeste uddannelsesniveaue

Selvom kønsdiskrimination i Danmark er lav i forhold til mange andre lande, er den stadig betydelig. Derfor kontrollerer vi for køn. Bemærk dog at vores indkomstgrænser ikke er forskellige for køn. Det vil sige at vi tager hensyn til køn i vores model, men til sidst når vi skal til at vurdere antallet af fattige i Danmark, så vurderes mænd og kvinder på samme vilkår, i forhold til deres evne at leve for mere end 70%-medianen.

Alder er også en vigtig faktor når man snakker arbejdsløshed. Laborta Data bruger Danmarks definition af arbejdsløshed og finder at der i 2008 er 15.705 arbejdsløse mellem 20 og 24, 11.085 arbejdsløse mellem 30 og 34 og 6.036 arbejdsløse mellem 45-49^[10]. Det er næsten en fordobling i antallet af arbejdsløse, med kun 10 år til forskel. Derfor inddrager vi også alder i analysen.

Uddannelsesniveaue er måske den største faktor indenfor arbejdsløshed. I Danmark er der kommet flere og flere højt uddannede, mens deres ledighed forsat har været lav. Der sker altså en substitution af arbejdspladser som favoriserer de højt uddannede (Groes et al, 2004^[13]) over de sidste to årtier. Derudover viser statistiske tal også, at arbejdsløsheden er betydelig større blandt lavt uddannede. Laborta Data og Eurostat Data finder at der i 2008 er 3,31% arbejdsløshed blandt lavt uddannede (Level 0-2 EU Standard) og 1,84% blandt højt uddannede (Level 5-6 EU Standard)^{[10][11]}. Der er således næsten dobbelt så mange arbejdsløse blandt lavt uddannede proportionelt til højt uddannede. Derfor er uddannelsesniveaue for et individ en vigtig faktor at medtage når man skal analysere arbejdsløshed. Samtlige argumenter indtil videre har dog gået på selve arbejdsløsheden, det vil sige den periode et individ tilbringer som ledighed. Hvis vi betragter individer som allerede er i job, og undersøger sandsynligheden for at de skulle miste dette job får vi et helt andet billede. Der er nemlig ingen umiddelbar grund til at uddannelsesniveaue skal have indflydelse på hvor længe man er i beskæftigelse. Derfor vælger vi kun at kontrollere det andet spørgsmål i forhold til alder og køn. Det kan virke unødvendigt at foretage denne begrænsning, men eftersom kontrol

med uddannelsesniveaue ville kræve en syntetisering af data, kunne en alt for svag korrelation resultere i en unødvendig stor bias.

Yderligere når vi betragter arbejdsløshed vil vi arbejde med et gennemsnit af perioden 2004 - 2008. Det danske arbejdsmarked ikke har været påvirket væsentligt af krisen i forhold til de variabler vi måler på i dette tidsrum. Se eventuel Danmarks Statistik om arbejdsløshed ^[12]. Men et arbejdsmarked er stadig kontant i bevægelse. Arbejdsløshed, arbejdsstyrke og efterspørgsel udvikler sig hele tiden. Blot fordi vi undersøger 2008 kan vi ikke begrænse os til data i denne periode. Det vi leder efter, er det mest optimale gennemsnit for året 2009 – 2014, netop fordi vi opstiller vores model for året 2009. Men hvor langt skal vi gå for at få tilstrækkelig data? Det er klart at hvis man går for langt får man også forkerte resultater, netop på grund af den økonomiske krise. Derfor har vi valgt at begrænse os til årene 2004 - 2008. Arbejdsmarkedet har været nogenlunde konstant i forhold til de parametre som interesserer os, samt har vi både Laborta, OECD, Eurostat og ESS data til rådighed for denne periode ^{[10] [9] [11] [7]}.

Arbejdsløshed som model

Vi benytter OECD Data for arbejdsløshedens længde i Danmark. OECD har som udgangspunkt kun angivet aldersgrupper og køn. Begge disse er som tidligere anført meget vigtige variabler. Men når vi kigger på muligheden for at få et job har vi også konkluderet at uddannelsesniveaue er vigtigt.

Laborta har data på arbejdsløshed i Danmark for perioden 2004 - 2008 delt op på uddannelsesniveauer. For at indføre dette data i vores model skal vi foretage yderligere antagelser. Hvis et individ er arbejdsløs og ikke handicappet, da vil individet uanset sit uddannelsesniveaue have incitament for at søge arbejde. Hvis vi antager at alle individer, uanset uddannelsesniveaue, med samme ihærdighed søger arbejde, da vil det i sidste ende være efterspørgslen af arbejdskraft der afgør hvorvidt individer kan få arbejde. Med andre ord, hvis der er flere lavt uddannede arbejdsløse end der er højt uddannede, er det udtryk for at der ikke er nok efterspørgsel efter deres arbejdskraft. Således vil en højt uddannet og en lavt uddannet have forskellige sandsynligheder for at komme i job igen, og disse sandsynligheder vil være proportionelle til arbejdsløsheden i hver kategori. Lad os vende tilbage til vores eksempel fra tidligere. I 2008 havde vi 3,31% arbejdsløshed blandt lavt uddannede (Level 0-2 EU Standard) og 1,84% blandt højt uddannede (Level 5-6 EU Standard). Det vil sige at der er 1,799 (3,31 / 1,84) så stor chance for at en højt uddannet får arbejde modsat en lavt uddannet i det samme tidsinterval. Grunden til at vi kan foretage dette er at arbejdsløsheden fordelt på uddannelsesniveauer har været nogenlunde konstant for hele perioden 2004 - 2008. Hvis vi så således antog at dette ikke var tilfældet, men i stedet f.eks. antog at en højt uddannet havde samme sandsynlighed for at finde arbejde som en lavt uddannet, da ville markedet i løbet af få år konvergere mod en arbejdsløshed som var ens uanset uddannelsesniveaue. Men arbejdsmarkedet har været konstant, hvilket netop betyder at vi kan uddrage udbudsforskellen på denne måde. Dette forudsætter naturligvis vores tidligere antagelse om at beskæftigelsestiden er uafhængig af uddannelsesniveaue.

For at kunne opstille en statistisk model er vi nødt til at indføre endnu to antagelser. Vi indfører hovedsagligt disse af mangel på mere præcis data og for at undgå bias. Resultaterne af disse antagelser vil i gennemsnit stemme overens med det vi måler på. Antagelserne er at vi kan beskrive vores model over arbejdsløshed som en stationær (time-homogeneous) kontinuert markov proces.

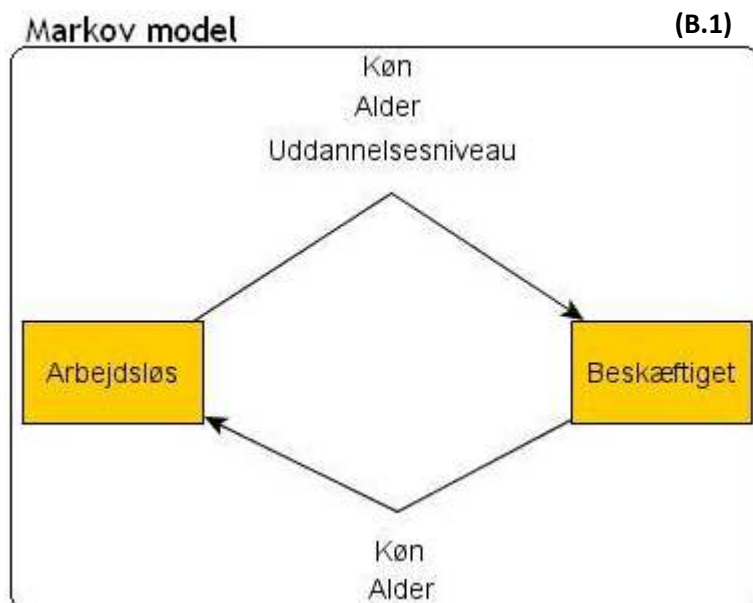
Antagelsen om vores model som markov proces betyder, at vores proces ingen hukommelse har (den såkaldte memoryless property). Det vil sige, at hvis vi betragter to individer, med samme alder, køn og uddannelsesniveau, hvor den første blev arbejdsløs for et år siden og den næste blev arbejdsløs i dag, givet at de begge stadig er arbejdsløse vil de begge have en lige sandsynlighed for at finde arbejde indenfor det næste år. Der er således ikke plads til, at individer går i depression grundet manglende arbejde, eller at de begynder at søge lavere betalte jobs. Denne antagelse er dog ikke så streng som den umiddelbart virker. For det første benytter vi data over arbejdsløshed som dækker årlige intervaller, og eventuelle ændrede jobsøgningsvaner bør derfor være afspejlet efter et år i arbejdsløshed. For det andet vil eventuelle ændringer i den måde individer søger arbejde på alligevel udtrykke sig i vores data, blot gennemsnitligt. Tilsvarende gælder for individer som er i arbejde. Et individ som har været i arbejde 10 år har lige så stor sandsynlighed for at miste sit arbejde et givent år som et individ der kun har været i arbejde i 5 år. Det skaber dog et bias med hensyn til anciennitet på arbejdspladsen. Denne bias vil dog være til fordel for de laveste indkomstgrupper. De laveste indkomstgrupper har svært ved at finde arbejde og er oftere arbejdsløse. Med denne antagelse bliver de sat i samme "bås" som resten af befolkningen, der generelt har nemmere ved både at finde og holde på et arbejde. Her skal læseren dog huske, at antagelsen ikke er så skarp som umiddelbart redegjort. Vores data spænder stadig over årlige intervaller og vi regner forsat på gennemsnittet.

Antagelsen om at vores markov proces er stationær bygger blot på et særligt tilfælde af vores antagelse om at arbejdsmarkedet er nogenlunde konstant for årene 2004 - 2008. I teoretisk sammenhæng betyder den simpelthen, at et individ med et givet køn, alder og uddannelsesniveau vil have samme sandsynlighed for at finde job i 2006 som i 2008. Tilsvarende vil et individ som har job, have samme sandsynlighed for at miste dette job i 2006 som i 2008.

Udover samtlige overstående antagelser, antager vi også at alle andre faktorer ikke hænger sammen med arbejdsløshed. Det udelukker f.eks. graviditet, ulykker, uddannelseskvalifikationer vi ikke kan måle osv. Vores formodning er at disse spiller en ringe rolle i det overordnede billede og ikke påvirker vores analyse i nogen væsentlig retning.

Arbejdsløshed analyse

Givet vores tidligere antagelser og diskussioner kan vi udregne sandsynligheder for arbejdsløshed og beskæftigelse som en stationær kontinuert markov proces. Diagrammet til højre illustrer modellen og de faktorer som indgår. Se appendiks for ligninger til denne model. Da vi ikke har adgang til mikrodata, kan vi ikke udregne konfidens-intervaller for vores resultater. Disse ville alligevel være underordnede senere hen, da vi er nødt til at antage, at hele SSM modellen er korrekt. De statistiske resultater er givet i appendikset.



Handicap

Faktorer og model

Vi har kun begrænset data til rådighed vedrørende handicap som omfatter hele den danske befolkning. Derfor kan vi ikke kontrollere sandsynligheden for handicap i forhold til indkomsten. I stedet kontrollerer vi sandsynlighederne i forhold til uddannelsesniveau. Den underliggende resonans er at, vi gennem uddannelsesvariablen i højere grad fanger de individer som har en lav indkomst. Det er ingen tvivl om at lav indkomst og lav uddannelse er stærkt korreleret. Derudover vil dette ikke medføre en bias, eftersom vi i ESS data stadig har uddannelse som variabel. Giver det så mening statistisk? Ifølge Eurostat data har lavt uddannede mellem 15-54 år 8,7% chance for at være alvorligt handicappet ("severely hampered"), mens kun 4,1% af højt uddannede mellem 15-54 år er alvorligt handicappet^[11]. Handicap og uddannelsesniveau er altså stærkt korrelerede. Eurostat bruger som udgangspunkt mindst 3.500 samples til sådan et datasæt. Den kritiske læser kan her selv opstille en binomial fordeling og se, at der er signifikant forskel på de to grupper.

For reelt at kunne udnytte vores data er vi nødt til at opstille en række antagelser igen. Det mest pålidelige data vi har til rådighed kommer fra Eurostat. Disse data udtrykker kun procentdele af befolkningen som er alvorligt handicappet i året 2008. Vi antager derfor, at procentdelen af handicappede i en given aldersgruppe, køn og uddannelse er proportionelt til den reelle sandsynlighed for at forblive syg. Vi kan illustrere dette ved, at bruge vores eksempel fra før. Da der er 8,7% lavt uddannede mellem 15-54 som er alvorligt handicappede, men kun 4,1% alvorligt handicappede blandt de højt uddannede mellem 15-54, får vi at sandsynligheden for at en lavt uddannet bliver handicappet er 2,12 ($8,7/4,1$) så stor som en højt uddannet. Det er kort sagt en faktor vi tillader os, at regne med.

Vi opstiller desuden den antagelse, at de mennesker som er handikappede i alderen 15-24, forsat er handikappede i alderen 25-34 blandt de lavt uddannede. Da der er næsten lige mange mennesker i hver aldersgruppe kan vi trække de handikappede fra i begge perioder, og får derved sandsynligheden for at et individ mellem 15-24 bliver handicappet i alderen 25-34. Vi vælger netop at gøre dette for de lavt uddannede fordi vi arbejder med de laveste indkomstgrupper, samt vil det give det mest stabile resultat, da en række mennesker endnu ikke har gennemført deres uddannelse i denne alder.

Vi opstiller sidst den antagelse, at handicap er uafhængigt af andre faktorer, samt at vi kan modellere den nøjagtigt med en stationær kontinuert markov proces. Det er klart at menneskers biologi arbejder i kontinuert tid, og derfor vil kontinuitet være den bedste approksimation. Derudover har vi også splittet sandsynlighederne ind i aldersgrupper, hvorfor kravet om stationærhed (time-homogeneous) kun er afgrænset til disse aldersgrupper, hvilket er en meget mild antagelse. Derudover har vi igen forsimplet situationen og skabt en sort-hvid verden: enten er individet ved godt helbred eller er individet handicappet. Hvis der skulle være en bias, ville den sandsynligvis påvirke den sociale mobilitet positivt for de laveste indkomstgrupper sådan, at vi har overvurderet den sociale mobilitet.

Handicap analyse

Givet vores tidligere antagelser og diskussioner kan vi udregne sandsynligheder for, at blive handicappet som en stationær kontinuert markov proces. Denne proces vil afhænge udelukkende af køn, alder og uddannelsesniveau. Se appendiks for ligninger til denne model. Da vi ikke har adgang til mikrodata, kan vi

ikke udregne konfidensintervaller for vores resultater. Disse ville alligevel være underordnede senere hen, da vi er nødt til at antage, at hele SSM modellen er korrekt. De statistiske resultater er givet i appendikset.

Uddannelse

Faktorer og model

Vi har kun begrænset data til rådighed vedrørende uddannelse, som omfatter hele den danske befolkning. Derfor vælger vi at begrænse os til, at kigge på uddannelse som en diskret proces. Individer har, afhængig af køn, alder og tidligere uddannelse, en given chance for at indskrive sig på en uddannelse og fuldføre denne over 5 år. Der foreligger til vores viden ingen pålidelige data eller rapporter, som belyser hvordan individer i lave indkomstgrupper gennemfører uddannelse sent i livet. Men da dette er vores største målgruppe i ESS data, vælger vi at antage at der ikke er nogen som falder fra. Det er en urealistisk antagelse, men eftersom almindelige frafaldsrater ligger omkring 10-14% (Danmarks Statistik, Gymnasielle uddannelser (1992-2008)), som er startet på en Level 5-6 uddannelse (EU Standard), og vi blandt individer over 24 har under 5% tilgang^[12]. Derudover risikerer vi ikke et negativt bias for lave indkomst grupper. I stedet bør enhver systematisk fejl, skabe en positiv bias for de laveste indkomstgrupper.

Derudover definerer vi også, at det nye uddannelsesniveau kun er relevant, såfremt det er startet samme år som vi fremskriver fra. Det indskrænker den sociale mobilitet overordnet, men vores bias vil stadig vægte de laveste indkomstgrupper positivt, da deres ringere mulighed for at opnå en uddannelse ikke bliver skaleret op.

Uddannelse analyse

Givet vores tidligere antagelser og diskussioner kan vi udregne sandsynlighederne for, at et individ opnår en højere uddannelse som en stationær diskret markov proces. Denne proces vil afhænge udelukkende af køn, alder og tidligere uddannelsesniveau. Se appendiks for ligninger til denne model. Da vi ikke har adgang til mikrodata, kan vi ikke udregne konfidensintervaller for vores resultater. Disse ville alligevel være underordnede senere hen, da vi er nødt til at antage, at hele SSM modellen er korrekt. De statistiske resultater er givet i appendikset.

Fødsler

Faktorer og model

Vi har kun begrænset data til rådighed vedrørende fødsler som omfatter hele den danske befolkning. Derfor kan vi ikke kontrollere sandsynligheden for fødsler i forhold til indkomsten. I stedet kontrollere vi sandsynligheden i forhold til uddannelsesniveau og alder. Derudover kontrollere vi også modellen for antallet af forrige fødsler moderen har haft (også kaldt birth order). Se Brodmann et al^[15], Gupta et al^[14] og Abernethy^[17], for de faktorer og processer som påvirker fødsler i husstanden. Vi forsætter vores analyse med udgangspunkt i disse, og regner derfor med at et eventuelt bias, hvis der skulle eksistere et, vil påvirke den sociale mobilitet i en positiv retning, og dermed undervurdere den reelle fattigdom.

Vi opstiller den antagelse, at fødsler uafhængig af andre faktorer, samt at vi kan modellere den nøjagtigt med en stationær kontinuert poisson proces. Det er klart at menneskers biologi arbejder i kontinuert tid, og derfor vil kontinuitet være den bedste approksimation. Derudover har vi også delt sandsynlighederne ind i aldersgrupper, hvorfor kravet om stationærhed (time-homogeneous) kun er afgrænset til disse aldersgrupper, hvilket er en meget mild antagelse.

Vi har igen forsimplet situationen og skabt en sort-hvid verden: homoseksuelle kan ikke få børn, børn bliver ikke født ind i komma tilstand osv. Derudover har vi også implicit antaget, at børnedødeligheden er uafhængig af alder, indkomst og uddannelse. Dette er urealistisk, men effekten er ubetydelig og det er derfor acceptabelt at udelukke dette fra vores model.

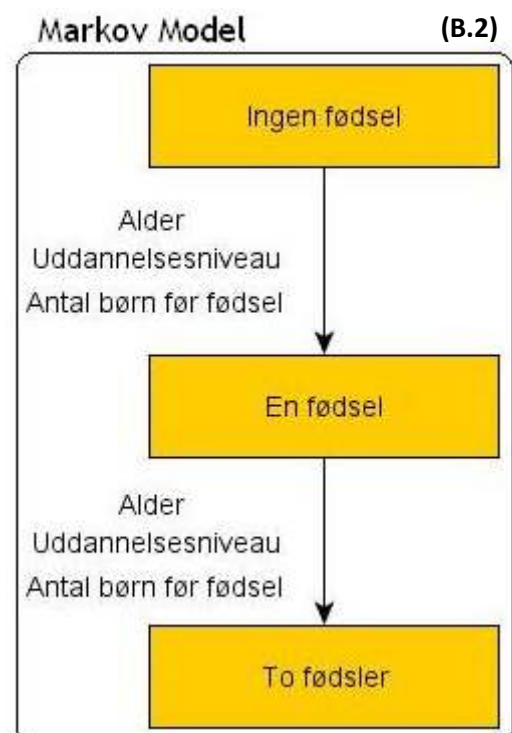
Statistikken viser, at der er betydelig forskel på antallet af børn individer føder. Langt de fleste føder 1-2 børn, mens en mindre andel af befolkningen føder flere. For at kontrollere antallet af fødsler i en 5-årig periode i forhold til tidligere fødsler, antager vi at antallet af 2. gang, 3.gang og 4.gangs fødsler er proportionelle til sandsynligheden for at føde et tilsvarende antal, samt at denne er uniformt fordelt over uddannelsesniveauerne.

Vi undlader at kontrollere for ægteskab/partnerstatus. Dette skyldes, at data på dette område er utilstrækkelig og fejlfyldt, og vi ville derved risikere et bias. For det andet indikere den tilgængelige statistik, at der ingen markant forskel er på om en kvinde er i ægteskab, parforhold eller single når det kommer til antallet af fødsler. Ifølge Danmarks Statistik bor 53% af den samlede befolkning over 25 år sammen med deres ægtefælle. 11% af befolkningen indgår i andre typer parhold og 33% af befolkningen er single ^[16]. Eurostat giver desuden at der i 2008 var 35,019 antal fødsler af kvinder i ægteskab, og 30,019 antal fødsler af kvinder udenfor ægteskab. Proportionelt til befolkningen har vi afrundet 660,74 (35,019 /53) og 682,25 (30,019 / (11+33)) antal fødsler per procentdel af befolkningen i henholdsvis ægteskab og udenfor ægteskab. Kort sagt er der 3% flere fødsler udenfor ægteskab, end der er fødsler i ægteskab. Forskellen er minimal og vi kan derfor uden problemer konkludere, at vores antagelse om ikke at kontrollere for ægteskab og partnerstatus er mild, hvis ikke helt ubetydelig.

Fødsel analyse

Givet vores tidligere antagelser og diskussioner kan vi udregne sandsynlighederne for, at individ får en højere uddannelse som en stationær diskret stokastisk proces. Diagrammet til højre illustrerer modellen og de faktorer som indgår. Se appendiks for ligninger til denne model.

Da vi ikke har adgang til mikrodata, kan vi ikke udregne konfidensintervaller for vores resultater. Disse ville alligevel være underordnede senere hen, da vi er nød til at antage, at hele SSM modellen er korrekt. De statistiske resultater er givet i appendikset.



PSM Model

Da der har været to forskellige datasæt til rådighed har projektet valgt at opstille to PSM Modeller. Den første, og primære, PSM model bygger på ESS data fra årene 2002, 2004, 2006 og 2008. Hertil har der været ca. 1600 samples (dvs. husstande) per survey, hvoraf ca. 400 af disse har været anvendt til opstilling af PSM, og kun ca. 200 har befundet sig under 70% medianen. Det er denne primære PSM model der hovedsagligt anvendes og beskrives i den videre rapport. Den anden, og sekundære, PSM model bygger på LIS data fra året 2004. Her har der været ca. 80.000 samples (dvs. husstande) i alt, ca. 40.000 har været anvendt til opstilling af PSM, og ca. 15.000 har været under 70% medianen. LIS datasættet er fra et kvantitativt standpunkt klart at fortrække fra ESS datasættet. Men denne kvantitative fordel falder tilbage set i lyset af sine to mangler, som vi også var inde på tidligere. Ved at benytte ESS datasættet kan vi opstille et statistisk (næsten) centralt estimat for årene 2002 - 2008, selvom det betyder at vi har en større usikkerhed på tallene. Fordelen ved dette er betydelig. Resultaterne bliver pålidelige og kan med rette anvendes til at analysere fattigdomsudviklingen for årene 2002 - 2008. Mens vi ved, at bruge LIS datasættet fra 2004, kan analysere de konkrete faktorer som påvirker fattigdom. I LIS datasættet fremkommer også en række interessante nye variable, som f.eks. geografisk placering og boligtype, hvilket tillader os at foretage længere og mere nøjagtige analyser.

Opstilling af model

Primær model

Den primære PSM model deles op i to dele. En for husstande, hvor husstandshovedet bor sammen med partner, og en for husstande, hvor husstandshovedet ikke bor sammen med partner. Grunden til, at vi splitter op i de to tilfælde er fordi der er forskellig information til rådighed for dem. Hvis vi også har en partner i husstanden, vil vi have næsten dobbelt så mange variable at måle på.

Efter denne opdeling er vi kommet til valg af variable. I tråd med vores tidligere analyser for SSM modellen, og afsnittet Positiv social mobilitet, har vi udvalgt følgende variable fra ESS datasættet:

- GNDR: Køn (0=Mand, 1=Kvinde)
- BLGETMG: Tilhører etnisk gruppe (0=Nej, 1=Ja)
- EDULVL: Husstandshoveds højeste uddannelsesniveau (0=EU ILSC 0-2, 1= EU ILSC 3-4, 2= EU ILSC 5+)
- EDULVLM: Husstandshoveds moders højeste uddannelsesniveau (Samme som EDULVL)
- EDULVLF: Husstandshoveds faders højeste uddannelsesniveau (Samme som EDULVL)
- EDCTN: Husstandshoved under uddannelse (0=Nej, 1=Ja)
- Head_Unemployed: Husstandshoved arbejdsløs, dvs. aktivt søger arbejde (0=Nej, 1=Ja)
- DSBLD: Husstandshoved erklæret handikappet (0=Nej, 1=Ja)
- RTRD: Husstandshoved pensioneret, også førtidspensioneret (0=Nej, 1=Ja)
- Number_Of_Children_S: Antal børn under 18 (0=Ingen, 1=Et barn, 2=to eller flere børn)
- HHMMB: Antal beboere i husstanden (1=En, 2=to eller flere børn)

Hvoraf der yderligere inkluderes følgende såfremt husstandshovedet bor sammen med partner:

- EDULVLP: Partners højeste uddannelsesniveau (Samme som EDULVL)
- Spouse_Unemployed: Partner arbejdsløs (Samme som Head_Unemployed)
- DSBLDP: Partner erklæret handikappet (Samme som DSBLD)

- RTRDP: Partner pensioneret, også førtidspensioneret (Samme som RTRD)
- EDCTNP: Partner under uddannelse (Samme som EDCTN)

Værdier med "-1" er ukendte værdier eller hvor respondenter har nægtet at svare. Disse har alle sammen været behandlet som en kategori for sig selv. Det vil sige, vi får mere usikkerhed på vores data, men vi undgår at foretage yderligere antagelser, så som at f.eks. skulle have tilhørt den mest udbredte gruppe.

Det er klart, at adskillige af disse variable er korrelerede. F.eks. har lavt uddannede personer, mere end 6 gange så stor sandsynlighed for at blive handikappede over en 5-årig periode, end akademikere. Der må derfor forventes en stærk sammenhæng mellem variabelen DSBLD og EDULVL. Samtidig må handicap forventes, at have en helt anden betydning for husstanden, end en lavt uddannet husstandshoved, hvorfor vi opnår betydelig mere information ved at inddrage denne variabel. Denne fremgangsmåde er foregået ved udvælgelsen af samtlige variable i analysen og derfor er vi endt i et statistisk multikolaritetsproblem.

Multikolaritet betyder netop, at vores variable er indbyrdes afhængige. Da vi maksimerer vores likelihood under modellen i logistisk regression, betyder multikolaritet ikke noget for den overordnede model. De endelige estimater er forsat maximum likelihood estimater i statistisk forstand, og derfor stadig de mest fornuftige estimater vi kan opnå. Hvis vi havde valgt at fjerne en række variable, kunne vi reducere denne multikolaritet, men så ville den overordnede model være mere usikker, netop fordi vi ville have mindre information til rådighed. Dertil er logistisk regression konsistent for antallet af samples gående mod uendelig. Se Amemiya et al. (1976). Vi vil derfor tillade os, at betragte den overordnede model som værende central for vores problemstilling, men samtidig betragte de enkelte estimater som værende biased. Det betyder kort sagt, at vi ikke skal lægge vægt i selve estimaterne af vores variable, men i stedet fokusere på den overordnede models resultater og teststørrelser.

Sekundær model

For den sekundære model har vi, med samme fremgangsmåde, valgt følgende variable fra LIS datasættet:

- D7: Antal beboere i husstanden (1=En, 2=To, 3=Tre eller flere børn)
- D27: Antal beboere i husstanden (1=En, 2=To, 3=Tre eller flere børn)
- hpeduc: Husstandshoved højeste uddannelsesniveau (0=EU ILSC 0-2, 1= EU ILSC 3-4, 2= EU ILSC 5+)
- hpdisabl: Husstandshoved handikappet, og modtager handikappension, (0=Nej, 1=Ja)
- hpumas: Husstandshoved i arbejde, dvs. aktivt søger arbejde (0=Nej, 1=Ja)
- hpethnat: Tilhører etnisk gruppe (0=Nej, 1=Ja)

Hvoraf der yderligere inkluderes følgende såfremt husstandshovedet bor sammen med partner:

- speduc: Partner højeste uddannelsesniveau (0=EU ILSC 0-2, 1= EU ILSC 3-4, 2= EU ILSC 5+)
- spdisabl: Partner handikappet, og modtager handikappension, (0=Nej, 1=Ja)
- spumas: Partner i arbejde, dvs. aktivt søger arbejde (0=Nej, 1=Ja)
- spethnat: Tilhører etnisk gruppe (0=Nej, 1=Ja)

Værdier med "-1" er ukendte værdier eller hvor respondenter har nægtet at svare. Disse bliver, ligesom i den primære model, behandlet som sin helt egen kategori.

Her har vi dog fjernet enkelte variable, som f.eks. køn, da deres effekt har været minimal eller ikke signifikante. Ligeledes har vi ikke haft mulighed for at inddrage enkelte variable, som f.eks. forældres uddannelsesniveau, da disse ikke har været tilgængelige i datasættet. Dette har samlet set reduceret multikolariteten væsentligt, og øget nøjagtigheden af vores resultater væsentligt.

Anvendte data syntetisering

Som beskrevet i teori afsnittet Data syntetisering, vil vi syntetisere vores data på tværs af tid. Den funktion F vi opstillede i afsnittet, vil vi definere som en funktion der uniformt fremskriver sammensætningen af befolkningen for husstande under 70%-medianen i forhold til komponenter: husstandstype og hvorvidt de er over eller under 70% medianen. Husstandstyperne op i følgende kategorier:

- Enlig voksen under 65 år
- Enlig voksen over eller lig 65 år
- Enlig voksen med børn
- To voksne, begge under 65 år
- To voksne, mindst den ene er over eller lig 65 år
- To voksne, med et barn
- To voksne med to børn
- To voksne med tre eller flere børn
- Tre eller flere voksne
- Tre eller flere voksne med børn

Samt om den enkelte husstand er over eller under 70% medianen. Funktionen F tager derefter en hvilken som helst husstand, og fremskriver den til et nyt år sådan, at det samlede antal husstande i dens kategori, til det fremskrevne år, stemmer overens med det virkelige antal husstande i denne kategori. Fremskrivning foregår således proportionelt. Hertil anvendes Eurostat Data, som giver os netop procentvise statistikker på disse husstandstyper.

Tag som eksempel en husstand, med en enlig voksen under 65 år fra 2004, og lad os fremskrive denne til 2009. I vores survey fra 2004 er det kun 12,78% af husstandene, som består af enlige voksne under 65 år. Men ifølge Eurostat er 31,9% af alle husstande i Danmark per 2009 af typen enlige husstande. Det vil sige, at vi er nødt til at vægte disse husstande 2,497 ($31,9/12,78$) gange så højt for at disse er proportionelle til hele befolkningen. Dertil skal vi også vægte husstandene i forhold til om de befinder sig under eller over 70% medianen. I 2009 var 34,5% af enlige husstande under 70% medianen. Men samtidig udgør de enlige husstande under 70% medianen kun 22,63% i vores survey fra 2004, hvilket betyder en faktor $1/0,2263$. Vores endelige vægt for husstanden bliver $3,80$ ($0,345 * 2,497 / 0,2263$). Enhver husstand med enlig voksen under 65 år, som befinder sig under 70%-medianen, vil altså blive vægtet som 3,8 husstande i alt.

Hele vores fremgangsmåde bygger på, at ændringerne i befolknings sammensætningen er uniform overfor alle de andre parametre vi ikke observerer. Det er en urealistisk antagelse, men dens usikkerhed bliver målt igennem vores normalfordelingsteori. Jo mere fejlagtig denne antagelse er, jo større vil sikkerhedsintervallet blive i vores model. Ligeledes er der ingen sociologisk begrundelse for, at dette skulle skabe bias. Med vores formål for øje, og manglende data, er dette den bedst mulige og mest pålidelige løsning.

For at vurdere effekten af denne vægtning er der i appendikset anført et skema over gennemsnitlige vægtninger. Skemaet viser at de fleste vægte holder sig enten mellem 0,7-0,8 eller 2,3-2,6, hvilket indikerer, at vores samples er pænt fordelte, og at der ikke sker uigennemskuelige beregninger.

Estimation af variable

I appendikset anfører vi vores estimater for maximum likelihood parametrene med deres værdier og tilhørende 90%-konfidensintervaller for den primære PSM model. Disse estimater skal alle sammen ses i lyset af vores multikolaritetsproblem som diskuteret tidligere. De der derfor tilstrækkelige og unbiased for vores ordnede model, men for usikre til, at betragte selvstændigt. Det bør dog nævnes at adskillige af vores variable, på trods af multikolaritet, har en signifikant effekt på hvorvidt husstande klassificeres som værende fattige eller ej. Uddannelsesniveau, hvorvidt husstandshovedet er i uddannelse eller ej og antallet af børn har alle sammen en absolut signifikant indflydelse for husstande uden partner. Lavt uddannelsesniveau og mange børn øger begge sandsynligheden for at være i fattigdom, mens at være under uddannelses formindsker den. At tilhører en etnisk minoritet øger også sandsynligheden for at være i fattigdom betydeligt for husstande uden partner. Ligeledes kan man fra et sociologisk standpunkt se, at en række andre variable, som f.eks. arbejdsløshed, også øger chancen for at være fattig, men dette er vores statistiske resultater ikke tilstrækkelige til at konkludere ud fra.

Nedenstående er vores test statistikker for de primære PSM modeller:

(B.3)

Testing Global Null Hypothesis: BETA=0 Primær PSM, husstande uden partner				Testing Global Null Hypothesis: BETA=0 Primær PSM, husstande med partner			
Test	Chi-Square	DF	Pr > ChiSq	Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	255.7732	19	<.0001	Likelihood Ratio	320.774	26	0.1906
Score	230.4672	19	<.0001	Score	315.260	26	0.2093
Wald	193.4027	19	<.0001	Wald	234.957	26	0.6048
Association of Predicted Probabilities and Observed Responses Primær PSM, husstande uden partner				Association of Predicted Probabilities and Observed Responses Primær PSM, husstande med partner			
Percent Concordant	64.5	Somers' D	0.310	Percent Concordant	58.1	Somers' D	0.180
Percent Discordant	33.5	Gamma	0.317	Percent Discordant	40.1	Gamma	0.183
Percent Tied	2.0	Tau-a	0.155	Percent Tied	1.8	Tau-a	0.090
Pairs	180200	c	0.655	Pairs	62491	c	0.590

Tabellerne viser, at vores primære PSM modeller kan klassificere henholdsvis 64,5% og 58,1% af husstande uden partner og husstande med partner korrekt. For at undersøge hvor signifikant vores modeller er, vil vi gerne opstille 95% konfidensintervaller for klassifikationskorrektheden af begge modeller. Vi har 64,5% korrekte klassifikationer med 849 samples i det første tilfælde, hvilket modelleret som en binomial fordeling giver os 95%-sikkerhedsintervallet (0,6124853, 0,6772674). For den anden model har vi 58,1% korrekte klassifikationer med 500 samples, hvilket giver os 95%-sikkerhedsintervallet (0,538, 0,624). Begge modeller opfanger således, en række væsentlige aspekter og mønstre bag fattigdommen. Set i lyset af, at vi ikke benytter økonomiske variable, kan vi altså konkludere, at vores udvalgte variable har en væsentlig indflydelse på fattigdom, og dermed at sociale og demografiske forhold spiller en central rolle for fattigdom.

Resultaterne er overordnet acceptable. Det skal dog pointeres, at likelihoodtesten for den anden PSM model er forholdsvis stor, men da antallet af korrekte klassifikationer er langt fra 50%, og vi stadig arbejder

med centrale estimater, kan vi tillade os at se bort fra dette. Vi vil derfor gå videre og anvende disse modeller til estimation af antallet af fattige.

Vi præsenterer nu estimaterne for den sekundære PSM model, altså estimaterne for LIS datasættet:

Analysis of Maximum Likelihood Estimates				
Sekundær PSM, husstande uden partner				
Parameter	Estimate	Standard Error	Chi- Square	Pr > ChiSq
Intercept	1	0,0846	73,38	<,0001
D4	1 0,3516	0,0483	52,94	<,0001
	2 -0,5914	0,0468	159,79	<,0001
D27	0 -0,3116	0,077	16,38	<,0001
	1 0,8011	0,074	117,26	<,0001
	2 0,0979	0,0804	1,48	0,2232
MARRIED	0 ,	,	,	,
hpeduc	0 -0,5428	0,0198	749,2	<,0001
	1 0,1605	0,0215	55,63	<,0001
hpdisabl	0 -0,9083	0,0286	1006,09	<,0001
hpumas	0 -0,2555	0,0224	130,47	<,0001
hpethnat	0 0,4219	0,0732	33,23	<,0001
	1 -0,1449	0,0792	3,35	0,0673

Analysis of Maximum Likelihood Estimates				
Sekundær PSM, husstande med partner				
Parameter	Estimate	Standard Error	Chi- Square	Pr > ChiSq
Intercept	1	0,1925	47,2	<,0001
D4	2 -0,4962	0,12	17,11	<,0001
D27	0 0,5542	0,1825	9,22	0,0024
	1 0,066	0,0897	0,54	0,4619
	2 -0,1445	0,0858	2,83	0,0923
MARRIED	2 -0,101	0,0319	10,05	0,0015
hpeduc	0 -0,2731	0,0384	50,55	<,0001
	1 0,2504	0,039	41,29	<,0001
hpdisabl	0 -0,2643	0,0606	19,04	<,0001
hpumas	0 -0,2358	0,041	33,01	<,0001
hpethnat	0 0,1336	0,1529	0,76	0,3822
	1 -0,4758	0,1554	9,38	0,0022
speduc	0 -0,1566	0,04	15,29	<,0001
	1 0,1077	0,0414	6,77	0,0093
spdisabl	0 -0,6267	0,0584	115,11	<,0001
spumas	0 -0,2616	0,0366	50,99	<,0001
spethnat	0 0,2808	0,1171	5,75	0,0165
	1 -0,2069	0,114	3,29	0,0696

Se appendikset for yderligere teststørrelser på disse variable.

Her er samtlige estimater signifikante for begge modeller, med undtagelse af enkelte underværdier af hpethnat, spethnat og D27. Begge sekundære modeller viser, at der er husstande med tre eller flere børn, har en markant højere chance for at være fattige, end husstande med færre end tre børn. Arbejdsløshed er også en signifikant faktor, som øger chancen for fattigdom. Ligeledes har husstande med lavt uddannede husstandshoveder og lavt uddannede partnere også markant større chance for at være fattige. Faktisk er effekten af uddannelsesniveau så voldsom, at uanset om vi kigger på husstande med partner eller ej, så har lavt uddannede husstandshoveder større chance for at være fattige, end mellem- eller højereuddannede arbejdsløse. Denne store effekt af uddannelsesniveau bør dertil ses i sammenhæng med uddannelseshomogamien, som yderligere forstærker konsekvenserne af at være lavt uddannet. Foruden disse variable er handicap og etnisk baggrund, for både husstandshoved og partner, også faktorer der betydeligt øger chance for at være fattig.

Samtlige resultater her er særdeles pålidelige. Vores Standard Error er i de fleste tilfælde mindre end 1/10 af selve estimaterne, og derfor kan vi også tillade os at sammenligne variable på tværs af hinanden. Dertil er vores likelihoodtest også særdeles stærk, og vores model kan derfor med rette anvendes til, at analysere fattigdomsudviklingen og årsagerne til fattigdom. Vi vil trække på disse resultater senere hen i vores analyse.

Resultater: ESS

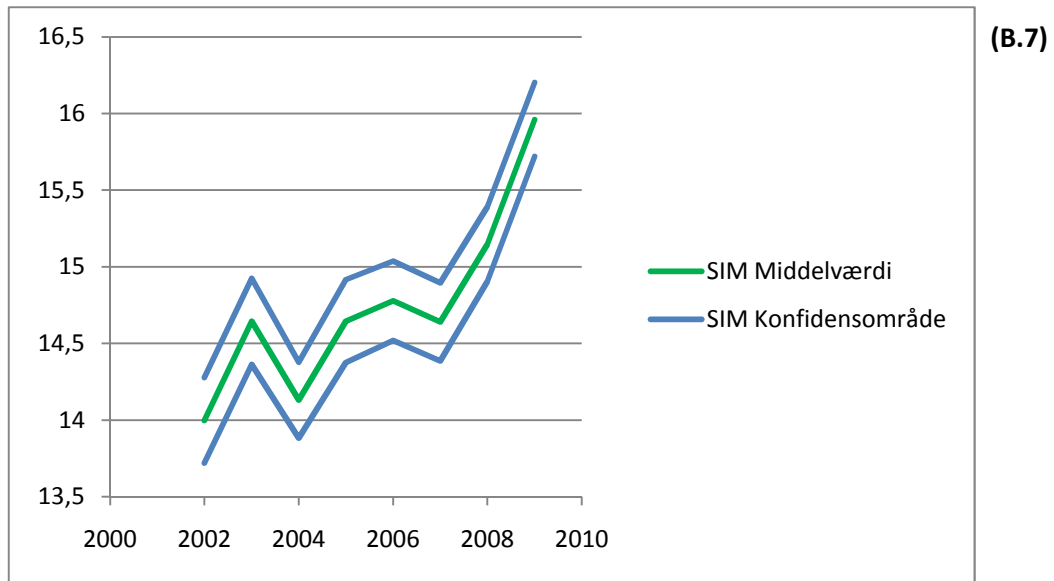
Vi har indtil videre opstillet og estimeret både SSM og PSM modellerne og er dermed klar til, at sammensætte modellerne med vægtningsmetoden og beregne vores fattigdomsindeks. Dette gøres nøjagtig som beskrevet i teori afsnittet. Vi præsenterer her hovedresultaterne for vores fattigdomsgrænse:

(B.6)

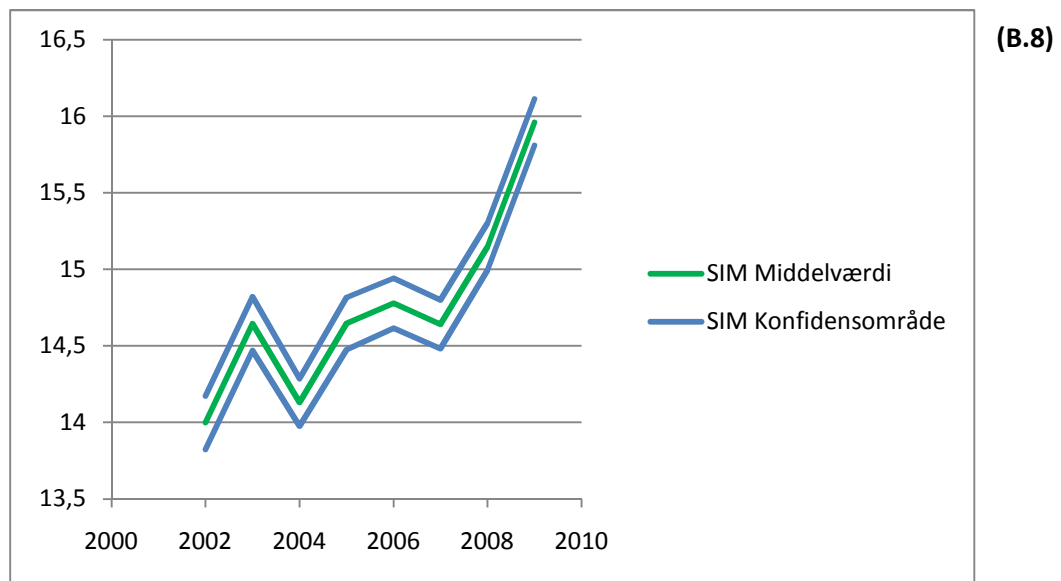
År	Befolkning under 50%-median	Befolkning under 70%-median	Pct. som er fanget i fattigdom (mellem 50-70% median)	Social-mobilitets fattigdomsgrænse
2009	7,2	20,5	65,86918898	15,96060213
2008	6,2	19,8	65,78646356	15,14695904
2007	5,9	19,5	64,27177263	14,64096108
2006	5,8	19,6	65,06434075	14,77887902
2005	5,7	19,6	64,35817665	14,64578655
2004	6,1	18,6	64,24406889	14,13050861
2003	5,7	19,8	63,43589543	14,64446126
2002	5,4	19	63,15576955	13,99820691

Fra tabellen ser vi, at alle tre fattigdomsgrænser har kraftige stiger siden 2002. SIM stiger med 14,1%, mens 50% og 70% medianerne stiger med henholdsvis 33,3% og 7,9%. Der er altså ingen tvivl her: fattigdommen i Danmark er kraftigt stigende. Men også den sociale mobilitet for de fattigste er faldende. I 2002 slap 36,8%, mellem 50-70% medianen, ud af fattigdom, mens tallet i 2009 kun var 34,1%. Det er fald på hele 7,9%, med en tendens der peger på, at faldet vil forsætte et godt stykke i fremtiden.

Vores fattigdomsindeks, SIM, illustreres nedenfor sammen med de tilhørende 80% profil konfidensintervaller for T-fordelingen. Dette diagram skal anvendes til, at forstå fattigdomsudviklingen over tid.

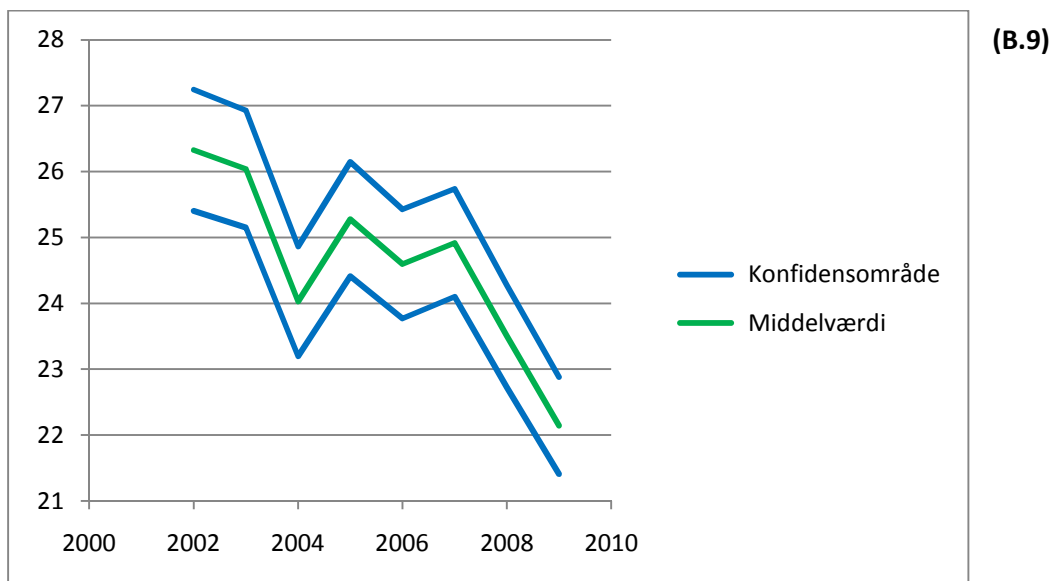


Nedenstående diagram viser SIM sammen med de tilhørende 80% simultane konfidensintervaller for T-fordelingen. Diagrammet skal anvendes til at sammenligne to punkter, og se om der er signifikant forskel.



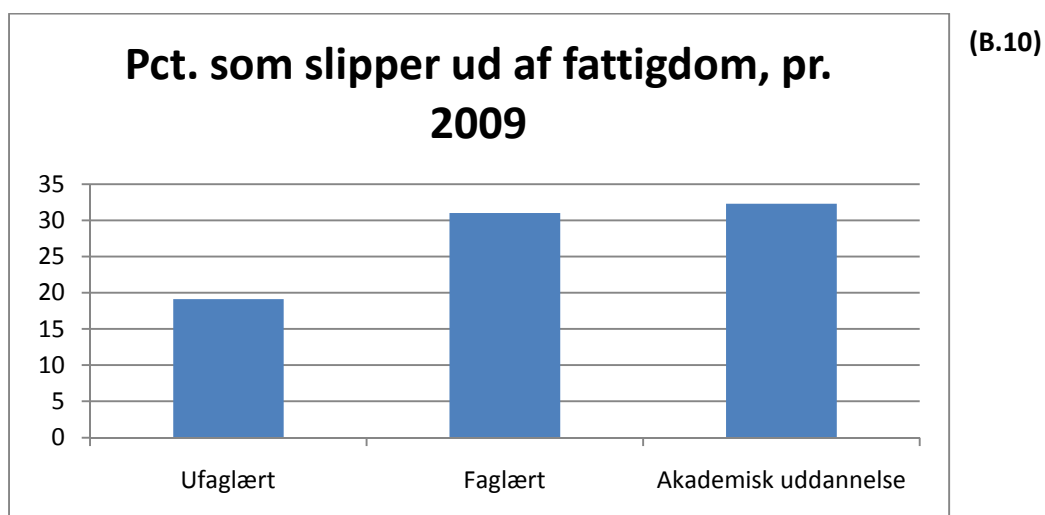
Fra diagrammerne observerer vi, at der er en signifikant opadgående tendens fra 2004 til 2005, og fra 2007 til 2009, hvor der, kun mellem 2003 og 2004 var en nedadgående tendens. Konklusionen bliver altså, at fattigdommen over det sidste årti er vokset signifikant. Vi har her angivet 80% konfidensintervaller, men i appendikset angives også 90% konfidensintervaller, der også signifikant bekræfter denne tendens.

Nedenstående diagram viser procentdelen af fattige, mellem 50-70% medianen, som undslipper fattigdom over en 5-årig periode med de tilhørende 80% simultane konfidensintervaller for T-fordelingen. Diagrammet skal anvendes til at sammenligne to punkter, og se om der er signifikant forskel.



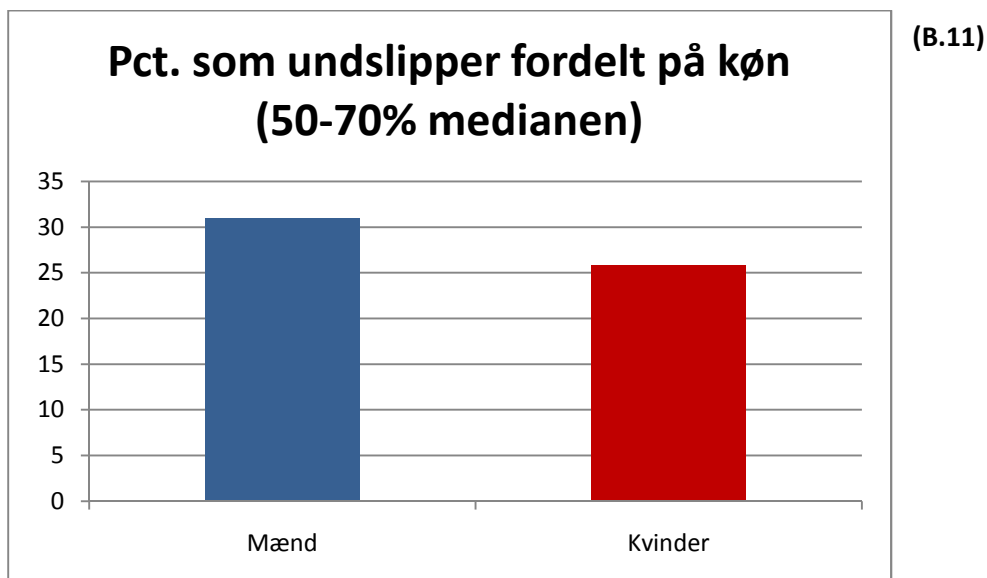
Her ser vi, at chancen for at slippe ud af fattigdom er faldet signifikant siden 2002. Vi oplever altså en særdeles dystre udvikling. På den ene side er fattigdommen stærkt voksende, og på den anden side, så bliver det væsentligt sværere at slippe ud af fattigdommen.

I appendikset er angivet en bred vifte af statistiske resultater som knytter SIM til ulighed, uddannelse, køn og befolknings sammensætning. Men for, at inddrage de mest centrale resultater, angiver vi kort to vigtige resultater. Nedenstående diagram viser sammenhæng mellem uddannelsesniveau af husstandshovedet og sandsynligheden for opnå en indkomst over 70% medianen over en 5-årig periode:



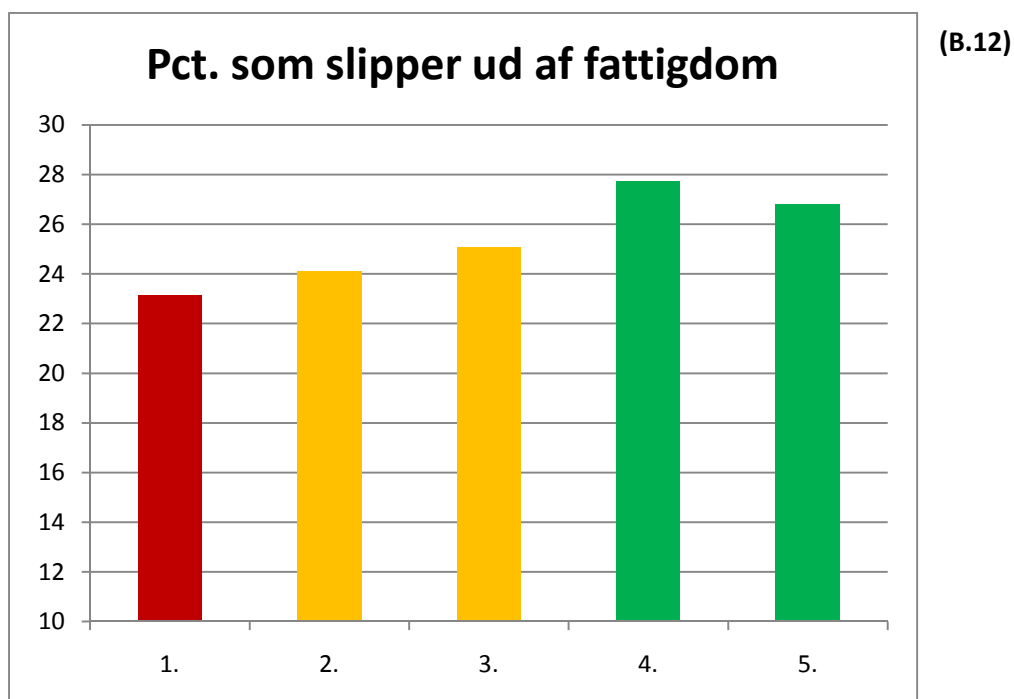
Ufaglærte mennesker har altså en markant dårligere chance for, at undslippe fattigdom. Det er måske ikke i sig selv så overraskende, men forskellen på uddannelsesniveauerne er enorm, og det er et nyt resultat. En faglært eller akademiker vil have den dobbelte chance for at undslippe fattigdom.

Nedenstående viser procentdelen af henholdsvis mænd og kvinder mellem 50-70% medianen der slipper ud af fattigdom. 80%-sikkerhedsintervallet for mænd, udregnet som binomialfordeling på 682 samples, bliver (0,2873900, 0,3328446), og for kvinder (0,2375367, 0,2800587). Der er altså væsentlig forskel på mænd og kvinder her.

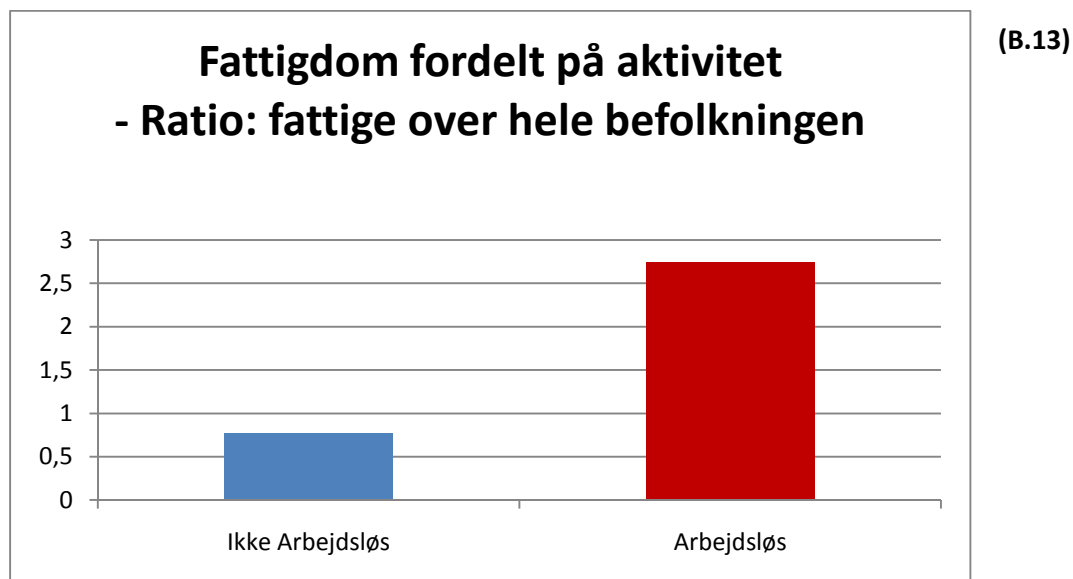


Resultater: LIS

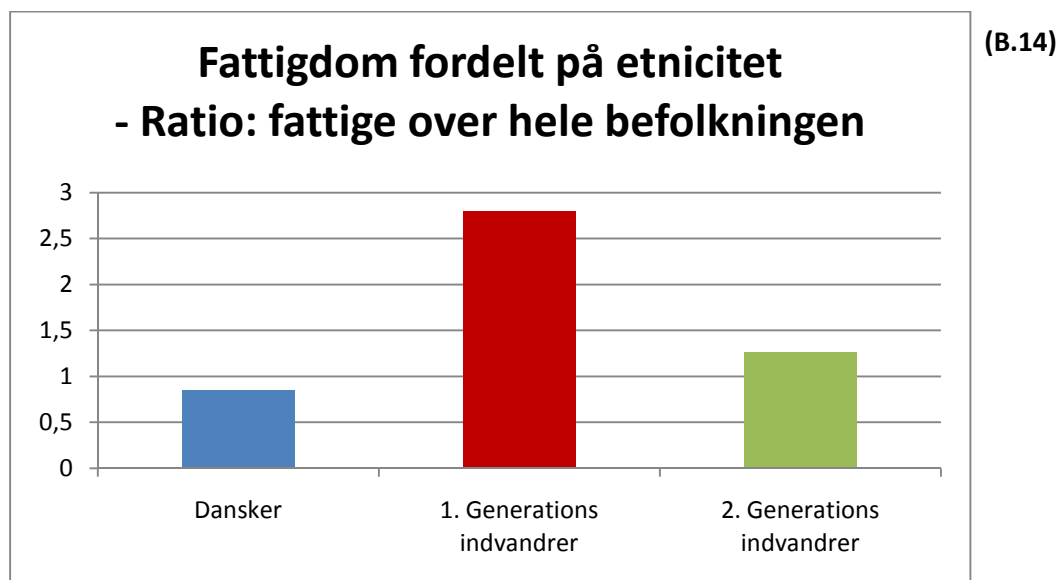
Vi præsenterer nu resultaterne for LIS datasættet. De går alle sammen på året 2007, og det bør igen bemærkes, at der forefindes et større udvalg af resultater i appendikset. Følgende opsummerer således blot de mest centrale resultater og konklusioner. Nedenstående diagram viser den sandsynlighed for, at slippe ud af fattigdom for alle husstande under 70% medianen, opdelt på 5 kvantiler. Som diagrammet viser, kan indkomsten for de fattigste have en altafgørende betydning for, at undslippe fattigdom.



Nedenstående diagram viser fordelingen af fattige fordelt på aktivitet. Diagrammet vises som et ratioforhold, hvilket matematisk udregnes som procentdelen af fattige der hører til en bestemt kategori divideret med procentdelen af mennesker i hele befolkningen der hører til samme kategori. Det betyder nedenfor, at en arbejdsløs har 3,7 gange så stor chance for at være fattige, end den gennemsnitlige dansker, hvilket er en ekstrem forskel.

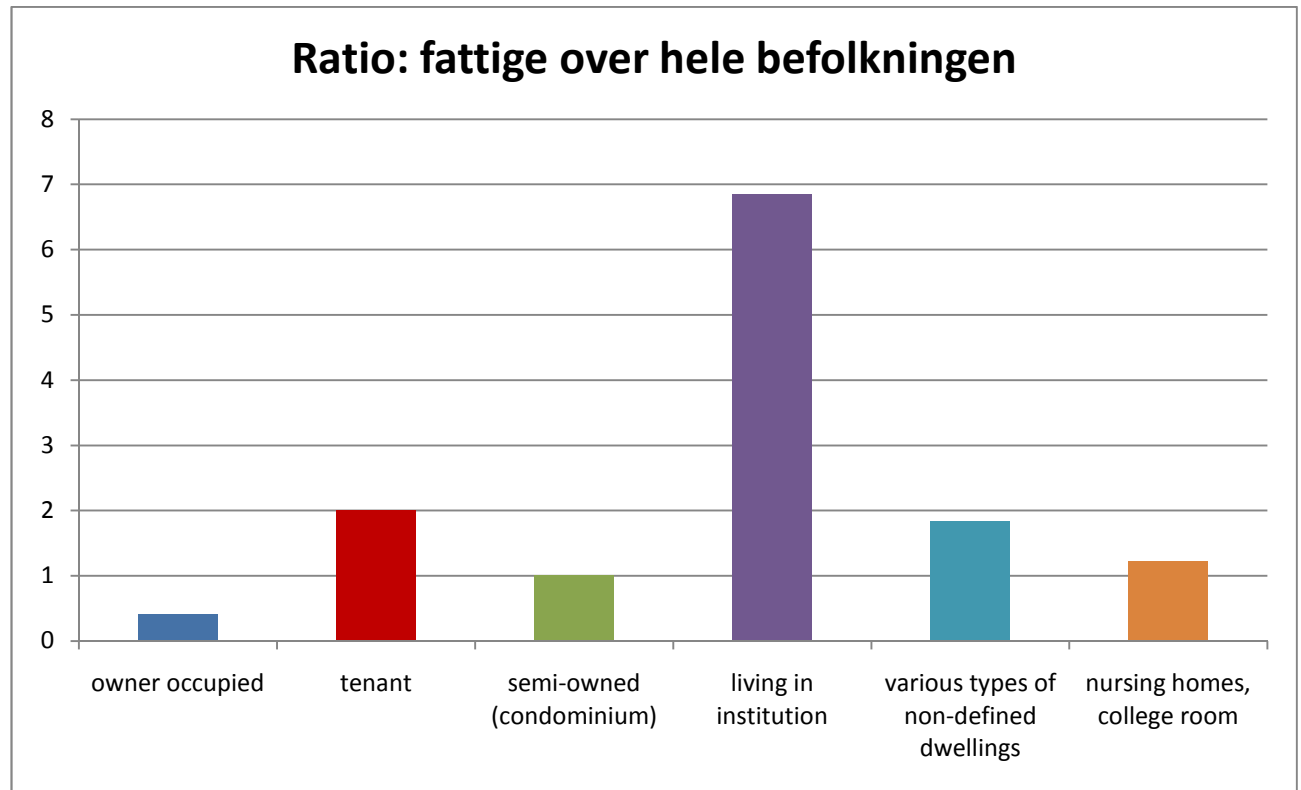


Nedenstående diagram viser en tilsvarende ratiofordelingen af fattige fordelt på etnicitet.



Dette diagram viser også en abnorm forskel på tværs af etnicitet i Danmark. Resultatet viser uden tvivl, at 1. generations indvandrere har over 3,5 gange så stor chance for at ende i fattigdom en den gennemsnitlige dansker, mens 2. generations kun har 1,5 gange så stor chance.

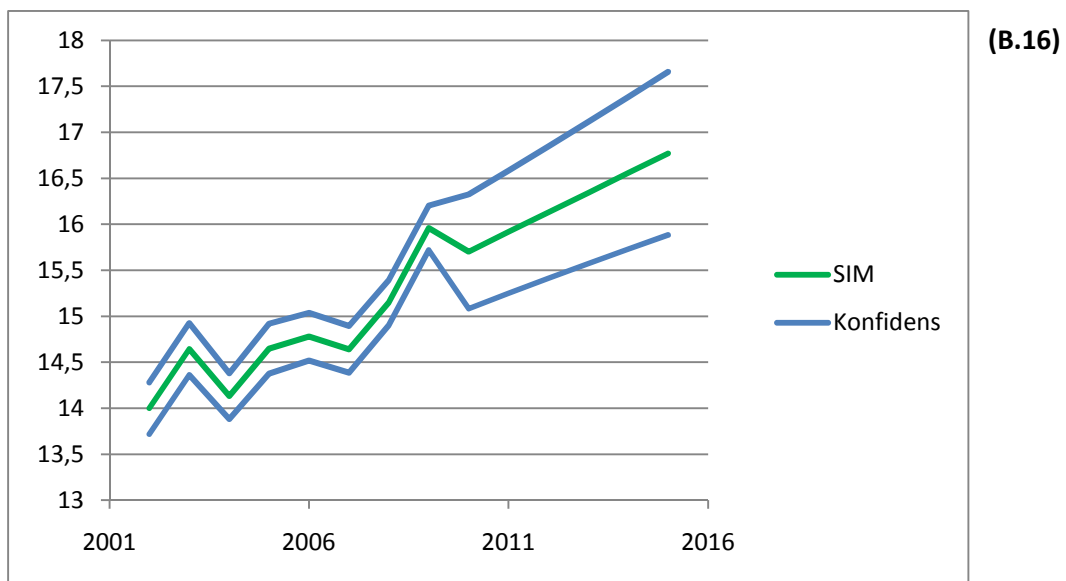
Nedenstående diagram viser igen en ratiofordelingen af fattige, som denne gang er fordelt på boligtype.



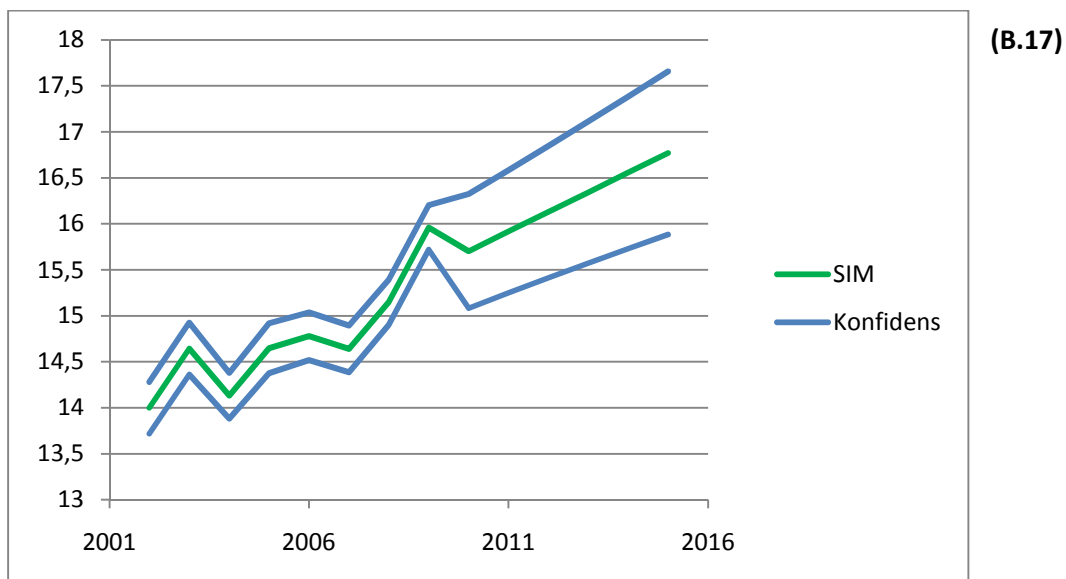
Her observerer vi endnu en gang, at visse faktorer har en kraftig sammenhæng til fattigdommen. Husstande som er bosiddende i en institution er klart dårligst stillet, med over 3,5 gange så mange fattige her, men også lejere har over 1,6 gange så stor chance for at være fattige end gennemsnitlige danskere.

Fremskrivning

Vi fremskriver SIM 5 år ind i fremtiden ud fra variabelen år som en lineær regression. Diagrammet illustrerer denne fremskrivning samt 80% profil konfidensintervaller. Se appendiks for detaljerede resultater.



Vi fremskriver SIM 5 år ind i fremtiden ud fra variablen år og variablen TY, procentdelen af BNP der anvendes til indkomstoverførsler, der er blevet fremskrevet af Det Økonomiske Råd ^[19]. Se forsat appendikset for detaljerede resultater.

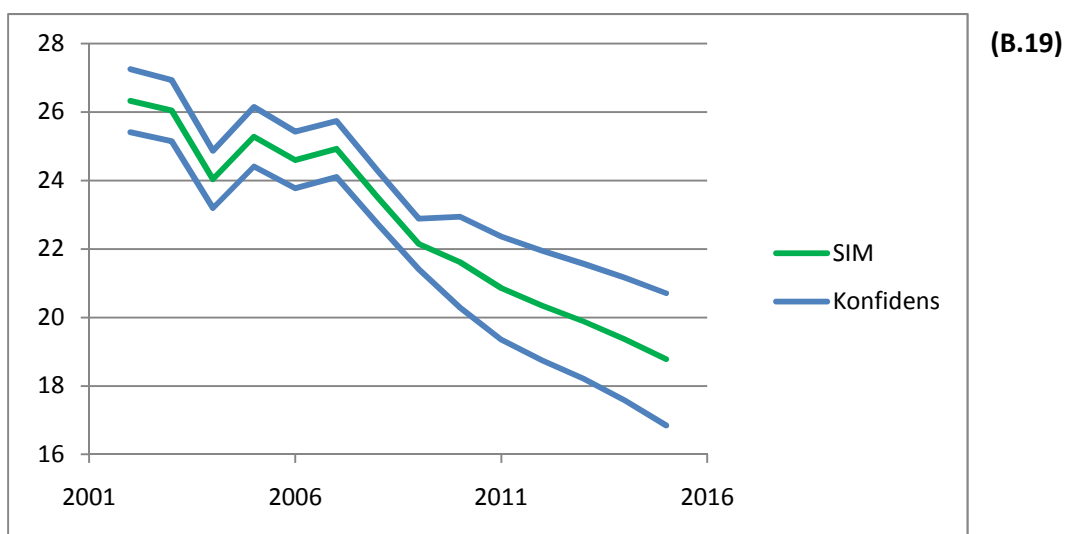


For den anden model er 90%-konfidensintervallerne:

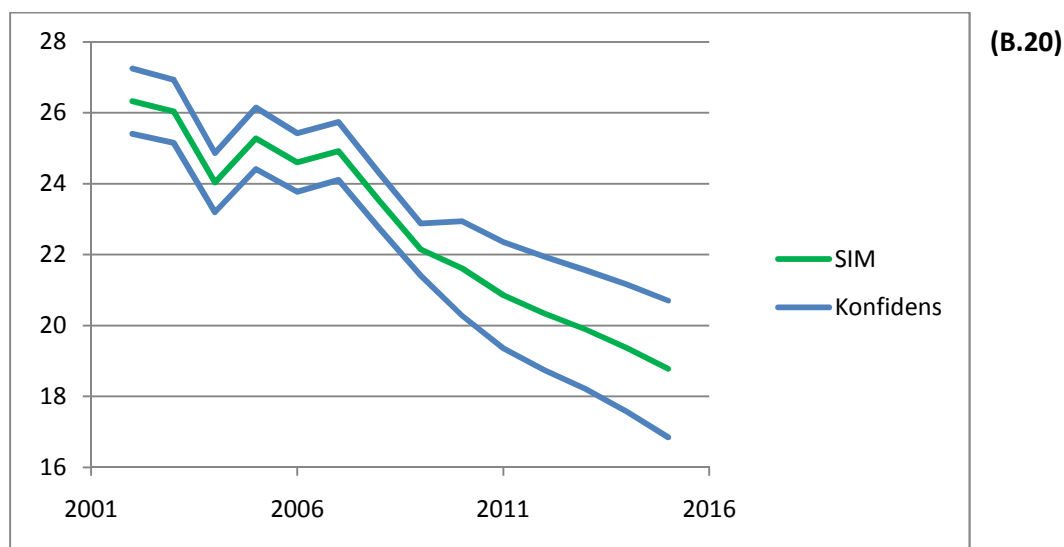
(B.18)

90%-konfidensintervaller	5%	95%
(Intercept)	-705,2956505	-335,324
ÅR	0,17317057	0,355647
TY	0,04161561	0,543361

Vi fremskriver procentdelen som undslipper fattigdom 5 år ind i fremtiden ud fra variablen år:



Vi fremskriver procentdelen som undslipper fattigdom over 5 år ind ud fra variabelen ÅR, og variabelen TY, procentdelen af BNP der anvendes til indkomstoverførsler, der er blevet fremskrevet af Det Økonomiske Råd ^[19].



For den anden model er 90%-konfidensintervallerne:

(B.21)

90%-konfidensintervaller	5%	95%
(Intercept)	755,8423604	1662,026
ÅR	-0,8089096	-0,36196
TY	-1,2408401	-0,0119

Diskussion af resultater

Vi har i de foregående afsnit præsenteret et hav af resultater. Det er dog udenfor denne rapport's omfang, at føre videregående diskussioner og konklusioner omkring fattigdom. Fattigdom er en langt mere kompliceret proces end vores definitioner og statistik siger, og dertil har forfatteren, som matematik studerende på Københavns Universitet, kun ringe erfaring med sociologisk arbejde. I stedet vil vi føre en kort opsummering af hvad vi er kommet frem til, og hvad det kraftigst og mest umiddelbart peger på.

Fattigdommen i Danmark har, ifølge vores nye definition, været stigende siden 2002. Samtidig er den sociale mobilitet for de fattigste faldet drastisk. Det vil sige, at vi får en større og dårligere stillet underklasse i Danmark. Men ikke nok med det, så er ændringerne på en så stor skala, at de kan have en afgørende betydning for befolkningssammensætningen i fremtidens Danmark. Tendensen er uangribelig, og viser kraftige stigninger af fattige mennesker, og endnu ringere mulighed for at de skulle undslippe fattigdommen, hvilket man fra politisk side omgående bør sætte fokus på.

Disse store ændringer skal ses i lyset af de faktorer der påvirker fattigdommen. Hertil har vi vist, at social arv, uddannelsesniveau, arbejdsløshed og etnicitet spiller en afgørende rolle, både for dem som ender i fattigdom, og for dem som slipper ud af fattigdom. Det lader desuden til, at de fattigste er stærkest koncentreret omkring boliglejere og familier med mange børn. Didier Fourage og Richard Layte ^[4] bekræfter dette, og finder ligeledes at også uddannelsesniveau er afgørende for hvorvidt en husstand kan undslippe fattigdom eller ej, uanset om de er arbejdsløse eller ej. Men billedet er ikke helt simpelt, da

kvinder som først er endt i fattigdom, har langt sværere ved at slippe ud af fattigdom igen, end deres mandlige modpart, på trods af at der til dags dato er færre kvinder i fattigdom end der er mænd i fattigdom.

Nye veje og videre arbejde

Proportional hazard model

Det Økonomiske Råd, har i en rapport fra 2006 vedrørende fattigdom ^[20], opstillet en proportional hazard model, over hvilke faktorer der påvirker fattigdom, og hvor længe et individ forventes at være i fattigdom. Dette er konceptuel anderledes model end vores modeller, men de underliggende faktorer og processer er de samme. Det Økonomiske Råd efterviste, at tiden et individ har været i fattigdom, dvs. under en hvis indkomstmedian, er afgørende for hvor længe de forventes at forblive i fattigdom. Kort sagt, jo længere tid et menneske har været i fattigdom, jo længere tid forventes han eller hun, at forblive i fattigdom. Det er derfor vigtigt, at man i videre arbejde med kvantitativ måling af fattigdom, aktivt inddrager tidsrummet et menneske eller en familie, har været i fattigdom, for reelt at vurdere fattigdommen. Det burde hertil være muligt, at koble denne proportional hazard model direkte sammen med vores SSM model for enten at forbedre den nuværende fattigdomsgrænse, ved inddrage tiden som faktor, eller udvikle en ny separat socioøkonomisk fattigdomsgrænse.

Målefejl

Breen og Moiso ^[2] viser hvordan indsamling af data omkring dynamisk fattigdom er fyldt af fejl. Særligt fokuserer de på EU, og viser bl.a. at målefejlen i Danmark er enorm. De kommer bl.a. frem til, at selvom Danmark er et foregående land i EU mht. fattigdom, så krakker det pæne billede når man tager hensyn til målefejl. Alle landene bliver langt mere ligestillet i forhold til hvor mange mennesker der ryger ind og ud af fattigdom hvert år (de såkaldte spells/exits), og således også i forhold til hvordan fattigdom udvikler sig. Det vil derfor være særdeles fordelagtigt for projektet, såfremt der opnås adgang til tilstrækkelig data, at implementere deres model og teste målefejl i projektets model over fattigdom.

Makroøkonomisk latente variable

Ved at indføre en (eller flere) makro-økonomiske latente (eller ikke-latente) variable i beregningen af PSM, vil man kunne beskrive *indkomstmobiliteten* for den enkelte husstand mere præcist, og samtidig tage den makro-økonomiske situation i Danmark i betragtning. For at kunne gøre dette, skal der imidlertid skaffes multidimensionelt data som knytter husstandens variable til makro-økonomiske variable. Chaudhuri et al ^[1] har to surveys til rådighed og benytter derfor også økonomiske variable relevante for den enkelte husstand. De indgår på samme vilkår i beregning af deres lineære model, med den forskel at de er fast for alle husstande, men forskellige for hver survey. Således vil vi ikke kun beskrive fattigdommen mere præcist, men også få en makroøkonomisk forklaring på fattigdommen, f.eks. i forhold til arbejdsmarkedet.

Økonomisk udvidelse til modellen

Enhver definition af fattigdom bunder i politiske årsager. Det vil for mange mennesker være helt oplagt. Det er derimod ikke oplagt, at den også kan bunde i sociale og samfundsøkonomiske årsager. Hvis man ud fra sin definition vil udtrykke samfundets økonomiske og sociale tab gennem en sådan definition, vil man have slået flere fluer med et smæk. Ikke blot ville man kunne sætte tal på udviklingen og samfundsmæssige gevinster ved at reducere fattigdom, men man vil også få et politisk og økonomisk talerør, hvorigennem

man kan skabe opmærksomhed og ændringer. Derfor foreslår projektet at udvikle en separat model, som anvender SIM indekstallet til, at estimerer samfundets omkostninger ved at have et voksende antal fattige mennesker, f.eks. ved at beregne den gennemsnitlig pris samfundet betaler per fattigt menneske per år. Hertil forventes *The Spirit Level*, af Richard Wilkinson and Kate Pickett, at etablere et teoretisk fundament.

Konklusion

Projektet har udviklet en kvantitativ statistisk model over fattigdom i Danmark ved navn *SIM (Social Indkomstmobilitets Model)*. Modellen har taget udgangspunkt i social mobilitet og social arv. Ved brug af denne model har vi defineret, målt og fremskrevet fattigdommen i Danmark. Vores beregnede resultater har vist os, at fattigdommen i Danmark har været kraftigt stigende siden 2002, og at også den sociale mobilitet har været kraftigt faldende siden 2002. Fremskrivning på disse data indikerer klart, at denne tendens vil forsætte, og at vi derfor skal forvente flere mennesker i fattigdom og færre som undslipper. Derudover har vi også analyseret de underliggende årsager og faktorer som hænger sammen med fattigdom. Vi er kommet frem til, at social arv, uddannelsesniveau, arbejdsløshed og handicap har en afgørende betydning for fattigdom, hvorfor de bør inddrages i enhver analyse af fattigdom. Det er således lykkedes projektet, at måle og fremskrive den overordnede fattigdom, samt at udpege og analysere de faktorer som har indflydelse på fattigdom for dermed, at bidrage med et mere nuanceret syn på fattigdom.

På det højere plan har projektet præsenteret nye metoder og modeller til modellering af fattigdom, herigennem markov modeller og logistiske regression, samt nye kvantitative tolkninger af modellerne. Desuden har projektet udviklet en solid vægtningsmetode, der kan anvendes på tværs af utilstrækkelig paneldata til, at forstå udviklingen af fattigdom. Projektet håber gennem teori og implementering af disse på, at have inspireret det fattigdomsfaglige miljø i Danmark.

Referencer

- [1] Shubham Chaudhuri, Jyotsna Jalan og Asep Suryahadi, 2002, "Assessing Household Vulnerability to Poverty from Cross-sectional Data: A Methodology and Estimates from Indonesia". Columbia Economics Discussion Paper
- [2] Richard Breen og Paso Moisiso, 2003, "POVERTY DYNAMICS FOR MEASUREMENT ERROR"
- [3] Richard Layte og Christopher T. Whelan, 2002, "Moving in and out of Poverty: the Impact of Welfare Regimes on Poverty Dynamics in the EU"
- [4] Didier Fouarge og Richard Layte, 2005, "Welfare Regimes and Poverty Dynamics: The Duration and Recurrence of Poverty Spells in Europe"
- [5] Richard Layte og Didier Fouarge, 2004, "The Dynamics of Income Poverty"
- [6] Deepa Narayan og Patti L. Petesch, 2007, "Moving Out of Poverty: Cross-disciplinary perspectives on mobility"
- [7] ESS Round 4: European Social Survey (2011): ESS-4 2008 Documentation Report. Edition 4.0. Bergen, European Social Survey Data Archive, Norwegian Social Science Data Services.
ESS Round 3: European Social Survey (2011): ESS-3 2006 Documentation Report. Edition 3.3. Bergen, European Social Survey Data Archive, Norwegian Social Science Data Services.
ESS Round 2: European Social Survey (2011): ESS-2 2004 Documentation Report. Edition 3.3. Bergen, European Social Survey Data Archive, Norwegian Social Science Data Services.
ESS Round 1: European Social Survey (2011): ESS-1 2002 Documentation Report. Edition 6.2. Bergen, European Social Survey Data Archive, Norwegian Social Science Data Services.
- [8] The Luxembourg Income Study
- [9] OECD Statistics
- [10] LABORTA Statistics
- [11] Eurostat
- [12] Danmarks Statistik
- [13] Nils Groes, Anders Holm, Fane Groes, Tina Honoré Kongsø og Kristian Brink, 2004, "Akademikernes arbejdsløshed og indtjening (2004) – brikker til et mønster"
- [14] Nabanita Datta Gupta og Nina Smith, "Children and Career Interruptions: The Family Gap in Denmark"
- [15] Stefanie Brodmann, Gosta Esping-Andersen og Maia Guell, "When Fertility is Bargained: Second Births in Denmark and Spain"
- [16] "Husstande og familier", Danmarks Statistik, Februar 2010
- [17] Virginia Deane Abernethy, "Population Dynamics: Poverty, Inequality, and Self-Regulating Fertility Rates"
- [18] Amemiya, T., 1976, "The maximum likelihood, the minimum chi-square and the nonlinear weighted least squares estimator in the general qualitative response model", Journal of the American Statistical Association 71, 347-351.
- [19] "Dansk Økonomi, forår 2010", Kapitel 2. Det Økonomiske Råd
- [20] "Dansk Økonomi, efterår 2006", Kapitel 2. Det Økonomiske Råd