

Serban, Iulian Vlad

**Working Paper**

## Fattigdom i Danmark: En socioøkonomisk fattigdomsgrænse

LIS Working Paper Series, No. 562

**Provided in Cooperation with:**

Luxembourg Income Study (LIS)

*Suggested Citation:* Serban, Iulian Vlad (2011) : Fattigdom i Danmark: En socioøkonomisk fattigdomsgrænse, LIS Working Paper Series, No. 562, Luxembourg Income Study (LIS), Luxembourg

This Version is available at:

<https://hdl.handle.net/10419/95472>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# LIS

## Working Paper Series

No. 562

Poverty in Denmark:  
A Socioeconomic Poverty Threshold

Iulian Vlad Serban

May 2011

Paper in Danish – Executive Summary in English



CROSS-NATIONAL  
DATA CENTER  
*in Luxembourg*

---

**Luxembourg Income Study (LIS), asbl**

---

# Poverty in Denmark

A socioeconomic poverty threshold

Executive Summary

In association with the Think Tank Cevea ([www.cevea.dk](http://www.cevea.dk)).

See Danish paper for details.

10.05.11

Contact:

Iulian Vlad Serban

Telephone: +45 30 20 00 74

E-mail: [julianserban@gmail.com](mailto:julianserban@gmail.com)

## Purpose

The purpose of this project is to develop a quantitative model for poverty measurement in Denmark which takes into account social mobility and social heritage. As such the main goal is to define, measure and forecast the poverty development in Denmark, as well as identify and analyse the factors influencing poverty. Furthermore it is the purpose of this paper to inspire the academic fields working with poverty to use sophisticated quantitative statistical models to gain a more detailed picture of poverty.

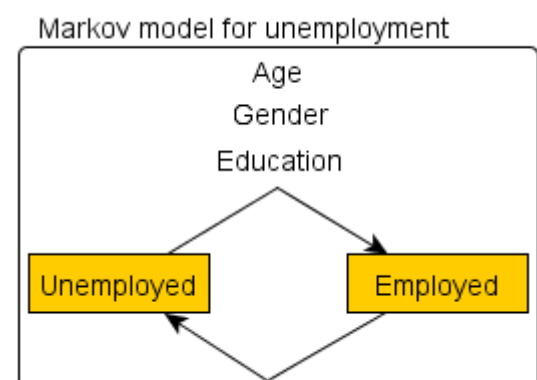
## Definition of poverty

The model for our poverty threshold is named SIM (Social Income Mobility Model). SIM classifies all households below the 50% median income as being poor by definition. Meanwhile households between the 50% and 70% median income are classified as poor if and only if after a 5 year period they cannot gain an income above the 70% median income. Finally all the households above the 70% median are classified as non-poor households by definition. With this definition we have explicitly established a poverty threshold which takes into account social mobility, and in effect also social heritage. From a conceptual point of view SIM is extremely simple, however to model it statistically a list of models and assumptions are needed to carry out the calculations for poverty. This is the compromise we make to ensure that the poverty threshold has sociological and political applications in Denmark. Only a conceptually simple poverty threshold, different from the traditionally used median incomes, can have wide influence and application.

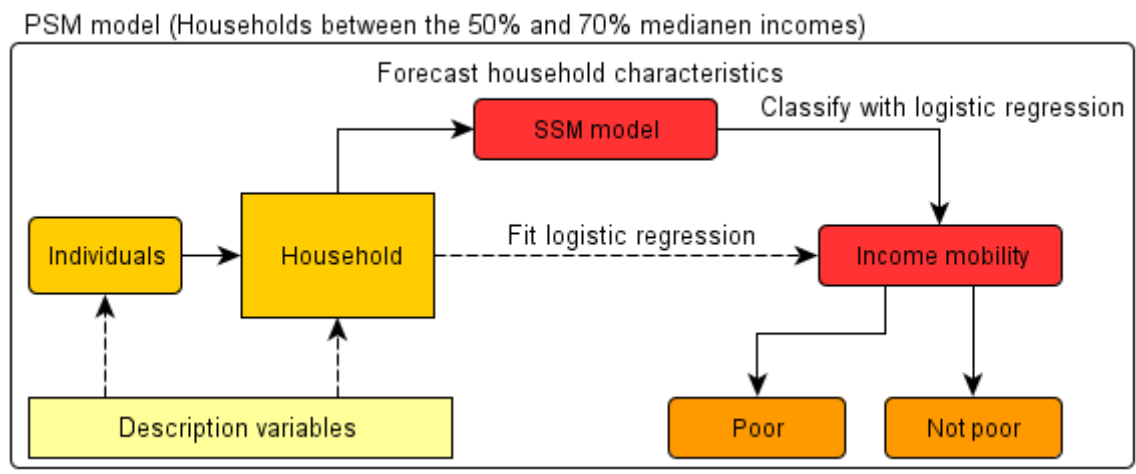
## Theoretical definition and assumptions

From our definition of poverty we define two sub models: SSM (*Stochastic Social Mobility Model*) and PSM (*Positive Social Mobility Model*). The SSM model forecasts household characteristics 5 years into the future. That is, given a household with a number of children, number of adults, age of each adult, education of each adult and so forth, it will give a vector of plausible future characteristics of this household with their estimated probabilities. There is only one important exception to this. The SSM model disregards the income of the household and all the individuals. In other words it uses only social characteristics to predict the future state of the household. This is an important aspect of our poverty definition, which ensures that we regard households in the bottom of the income fractiles in just as highly as the households on the verge to earning more than the 70% median income. So the social characteristics determine explicitly whether or not a household is in poverty.

The SSM model is modeled as different markov processes forecasting respectively unemployment status, handicap, child births and education. The first 3 markov models are continuous processes, while the last one for education is discrete. They each control for different relevant factors including age, gender, a priori education status, a priori number of children. So for each of our four characteristics that we are modeling, we have a list of markov models modeling these. To do this we need to assume a range of properties of the poverty processes, mainly that the social processes can be split into categorical states, that the transition between these states follow an exponential distribution, and that the population is homogenous. With these assumptions our estimates are central.



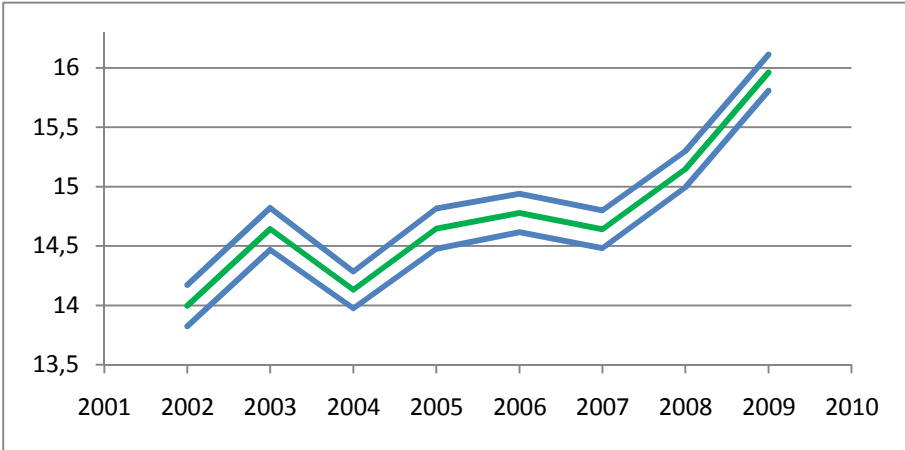
The PSM model is a logistic regression modeling whether or not a household, from its social characteristics, is below or above the 70% median income. It can be realized as the following diagram:



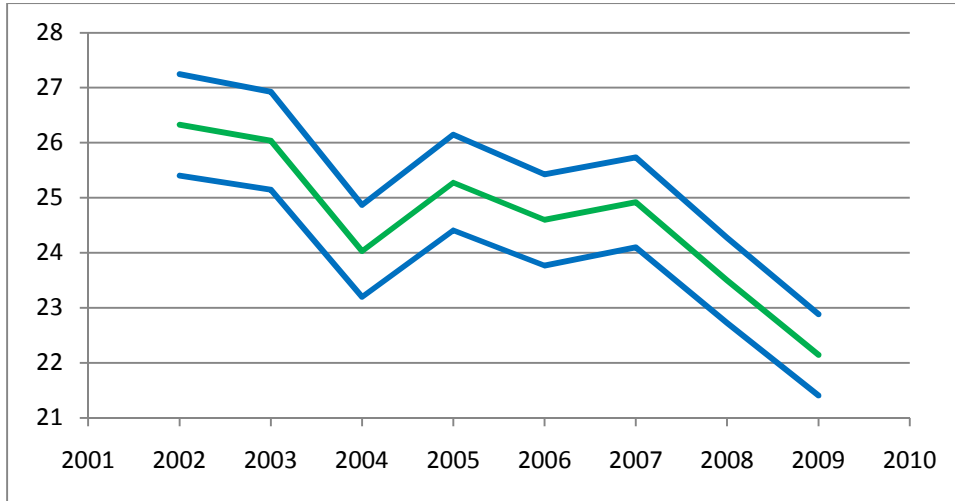
As the diagram shows the logistic regression is fitted for the households before they are forecasted 5 years into the future. This can be seen as a further development of the theoretical work by Chaudhuri et al. <sup>[1]</sup>, who shows that with panel data and linear models, one can successfully model the risk of poverty. This is essentially what we are doing with the PSM model. The output for a certain household will be a probability of whether or not that household is in poverty after 5 years. Our final poverty threshold, SIM, will then be the sum of all the probabilities added with the number of households living below the 50% median income. To gain the most applicability we choose to fit it for our data for the year 2009. This should be viewed as reference point in our definition of poverty. However we expect not to have any considerable influence on the final results, as the characteristics influencing poverty are believed to change, but still overlap extensively throughout the years. Among the variables measured in the PSM model are gender, educational level, activity (working, unemployed, student, retired), medically handicapped, ethnic background, fathers education, mothers education, number of people living in household, number of children living in household. Of which the first four variables also apply to the household spouse.

**Results**

We estimate our SSM model from Eurostat, OECD, LABORTA and Denmark Statistics <sup>[4], [5], [6], [7]</sup>, while the PSM model is estimated from both the Luxemburg Income Study data <sup>[3]</sup> and the European Social Survey data <sup>[2]</sup>, from 2002, 2004, 2006 and 2008. From the Luxemburg Income study we are able to make deeper analysis of poverty for the single year 2004, while for the European Social Survey data we are able to estimate the poverty development from 2002 to 2009. We now present the main results:



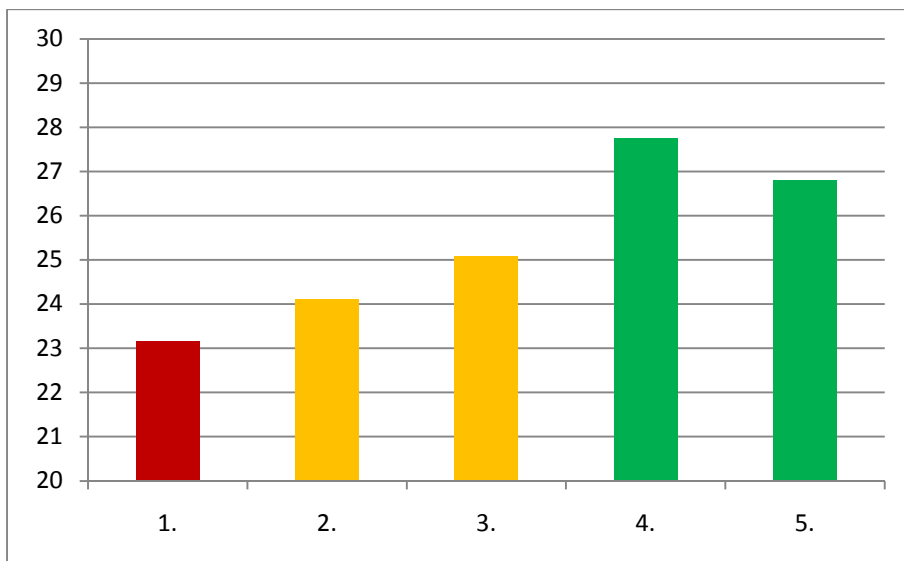
*SIM poverty threshold for Denmark from 2002 to 2009 with 80% simultaneous confidence intervals for comparing two different years (ESS)*



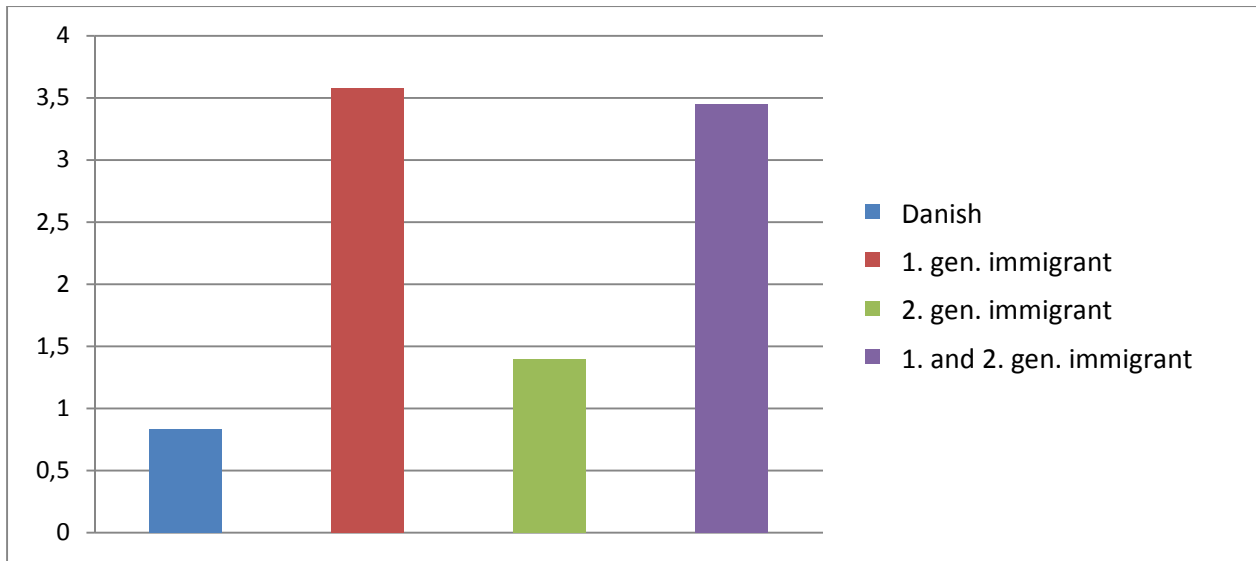
Percentage individuals below 70% median income who escape poverty from 2002 to 2009 with 80% simultaneous confidence intervals for comparing two different years (ESS)



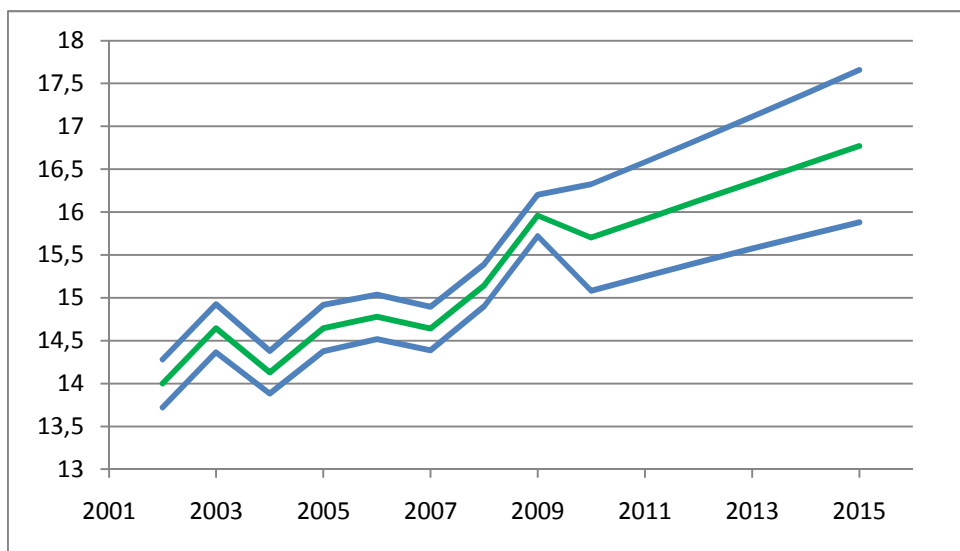
Percentage individuals who gain an income over 70% median income based on household head education for 2009 (ESS)



Percentage individuals who gain an income over 70% median income based on household income quantile for 2004 (LIS)



*Ratio of poor people compared to the whole population based on ethnic background of household head for 2004 (LIS)*



*Linear prediction of SIM poverty threshold for Denmark from 2009 to 2014 with 80% simultaneous confidence intervals for comparing two different years (ESS)*

## Conclusion

From our new definition of poverty we find that the poverty in Denmark, just like the poverty threshold based on the 50% median income, has increased extensively and statistically significant over the last decade, with an increase of at least 10 percent. This is a huge increase in poverty, with strong indications of further increases, and may therefore require immediate action from the elected political representatives of the Danish government.

Furthermore we have shown that social heritage, educational level, unemployment and ethnicity play a dominant statistically significant role in determining if a family lives in poverty or not, and whether or not they can ever escape it. The poor people tend to be concentrated among families with many children and renters. Lastly the picture gets somewhat more complicated. Currently fewer women than men are living below the 50% median, but our results show that once they are in poverty, their chances of escaping it are far less than that of men. This could be a potentially serious poverty trap.

The project has developed a quantitative poverty threshold which takes into account social mobility and social heritage. With a simple definition of poverty and a list of statistical and sociological assumptions, we have managed to model this poverty threshold successfully with the use of markov models and logistic regression and prior scientific research. Through this model we have managed to measure poverty in Denmark over the last decade, and have so been able to make extensive analysis of the factors influencing poverty. Moreover with our developed statistical theory and the applications of our work, we have created a project which may inspire the academic fields working with poverty to use sophisticated quantitative statistical models to measure poverty in deeper detail.

## References

[1] Shubham Chaudhuri, Jyotsna Jalan og Asep Suryahadi, 2002, "Assessing Household Vulnerability to Poverty from Cross-sectional Data: A Methodology and Estimates from Indonesia".

[2] 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.

[3] The Luxembourg Income Study

[4] OECD Statistics

[5] LABORTA Statistics

[6] Eurostat

[7] Denmark Statistics



# Fattigdom i Danmark

En socioøkonomisk fattigdomsgrænse

Detaljeret analyse og modelbeskrivelse,

Mårettet sociologer, økonomer og statistikere.

10.05.11

Kontaktperson:

Iulian Vlad Serban

Tlf: +45 30 20 00 74

E-mail: [julianserban@gmail.com](mailto:julianserban@gmail.com)

## 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, uddannelsesniveau, 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](http://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 <sup>English</sup> .....	2
Resumé <sup>Dansk</sup> .....	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 handikappede, 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 husstands niveau. 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.<sup>[3][5]</sup> 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<sup>[4]</sup>, 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<sup>[6]</sup> 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.

Læseren bør bemærke, at størstedelen af den matematik 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.

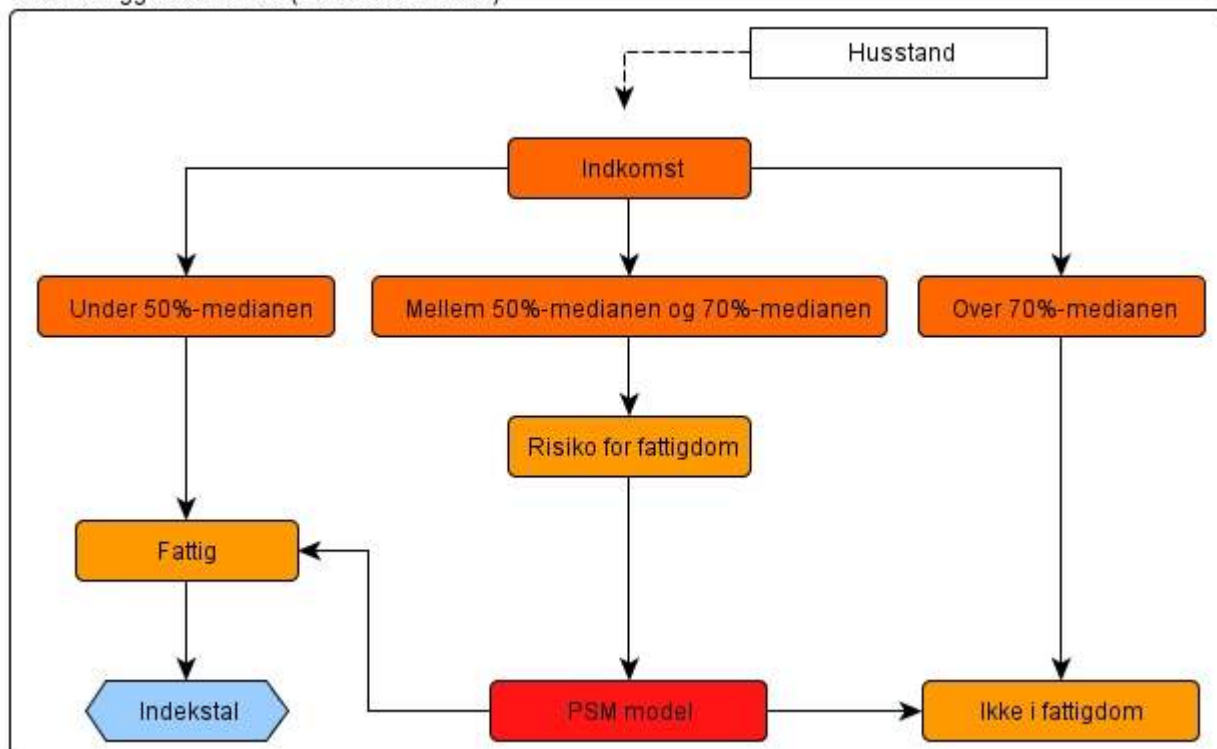


## Grundlæggende model

Ud fra ovenstående definition kan modellen opstilles som nedenstående diagram.

Grundlæggende model (Husstandsniveau)

(A.1)

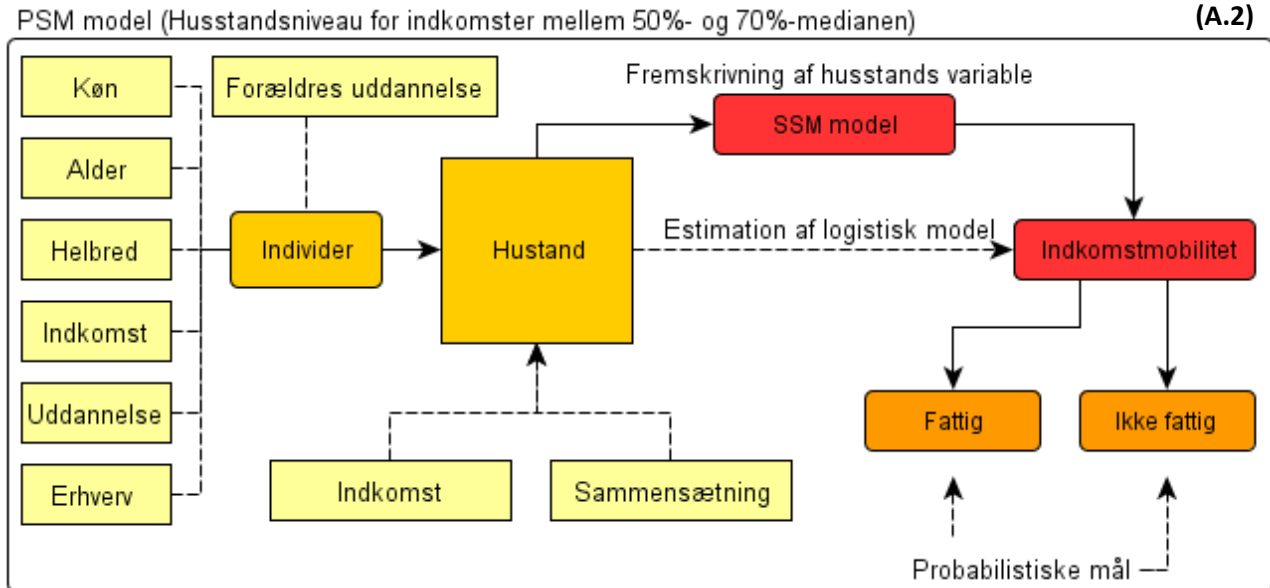


Alle udregninger foretages på husstands niveau. Som en analogi kan man anskue modellen som et rullebånd. Husstande kommer ind øverst i diagrammet og deres indkomstniveau bliver vurderet. Såfremt deres disponible indkomst er lavere end 50%-medianen, bliver de klassificeret som fattige, og hvis deres disponible 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)<sup>1)</sup>. 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  $\beta$  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](http://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 år<sup>9)</sup>. 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 uddannelsesniveau

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<sup>[10]</sup>. Det er næsten en fordobling i antallet af arbejdsløse, med kun 10 år til forskel. Derfor inddrager vi også alder i analysen.

Uddannelsesniveau 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<sup>[13]</sup>) 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)<sup>[10][11]</sup>. Der er således næsten dobbelt så mange arbejdsløse blandt lavt uddannede proportionelt til højt uddannede. Derfor er uddannelsesniveauet 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 uddannelsesniveauet 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 <sup>[12]</sup>. 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 <sup>[10] [9] [11] [7]</sup>.

### 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.

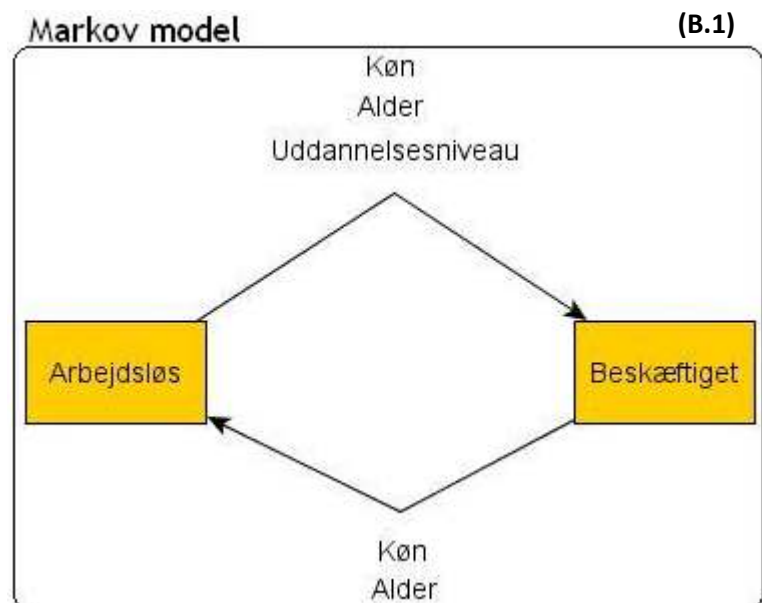
Antagelsen om vores model som markov proces betyder, at vores proces ingen hukommelse har (den såkaldte memoryless property). Det vil sige, at hvis vi betragter to individer, med samme alder, køn og uddannelsesniveau, hvor den første blev arbejdsløs for et år siden og den næste blev arbejdsløs i dag, 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<sup>[11]</sup>. 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<sup>[12]</sup>. 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<sup>[15]</sup>, Gupta et al<sup>[14]</sup> og Abernethy<sup>[17]</sup>, 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 acceptabel at udelukke dette fra vores model.

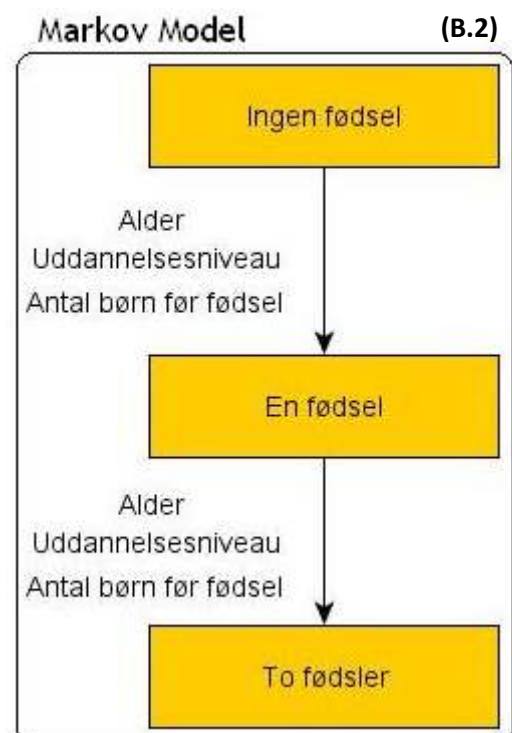
Statistikken viser, at der er betydelig forskel på antallet af børn individer føder. Langt de fleste føder 1-2 børn, mens en mindre andel af befolkningen føder flere. For at kontrollere antallet af fødsler i en 5-årig periode i forhold til tidligere fødsler, antager vi at antallet af 2. gang, 3.gang 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 <sup>[16]</sup>. 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:

<b>Analysis of Maximum Likelihood Estimates</b>				
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

<b>Analysis of Maximum Likelihood Estimates</b>				
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

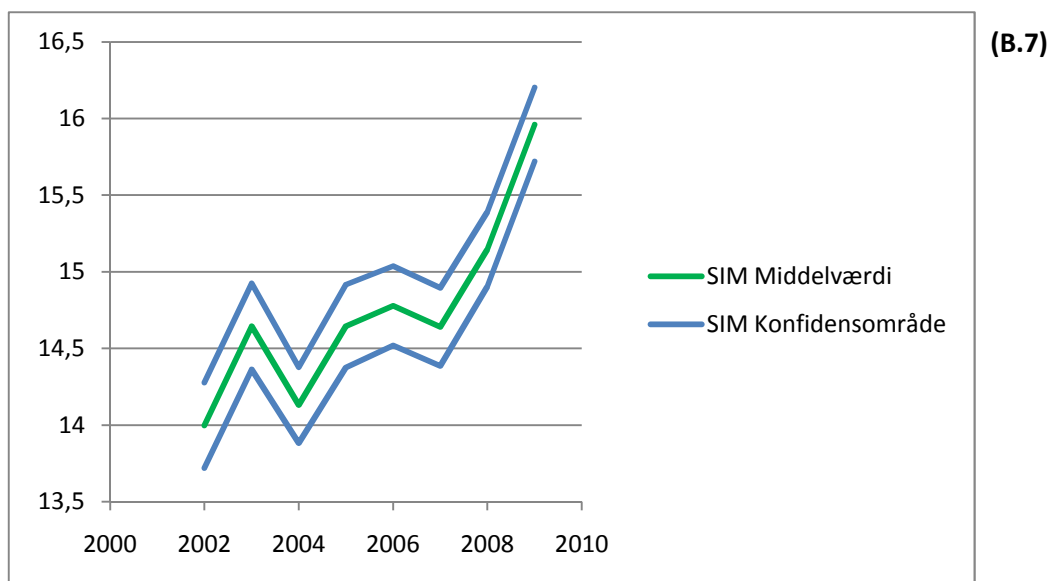
Vi har indtil videre opstillet og estimeret både SSM og PSM modellerne og er dermed klar til, at sammensætte modellerne med vægtningsmetoden og beregne vores fattigdomsindeks. Dette gøres nøjagtig som beskrevet i teori afsnittet. Vi præsenterer her hovedresultaterne for vores fattigdomsgrænse:

(B.6)

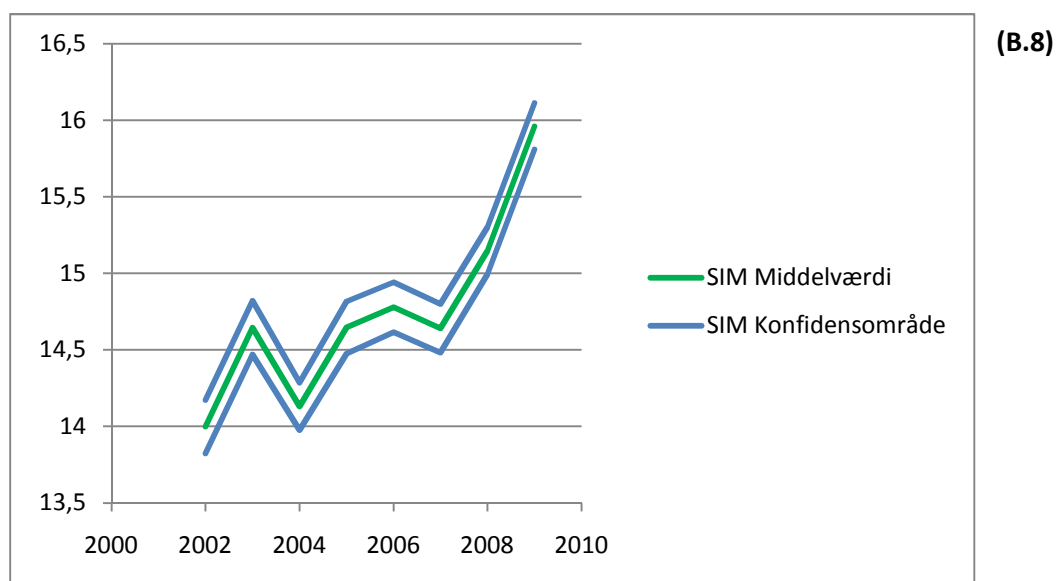
År	Befolkning under 50%-median	Befolkning under 70%-median	Pct. som er fanget i fattigdom (mellem 50-70% median)	Social-mobilitets fattigdomsgrænse
2009	7,2	20,5	65,86918898	15,96060213
2008	6,2	19,8	65,78646356	15,14695904
2007	5,9	19,5	64,27177263	14,64096108
2006	5,8	19,6	65,06434075	14,77887902
2005	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.

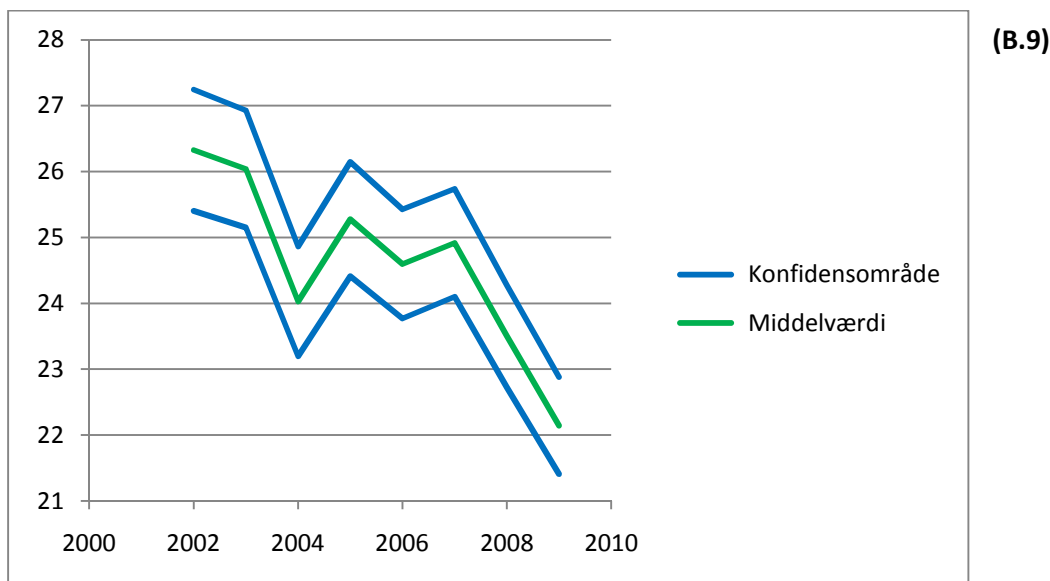


Nedenstående diagram viser SIM sammen med de tilhørende 80% simultane konfidensintervaller for T-fordelingen. Diagrammet skal anvendes til at sammenligne to punkter, og se om der er signifikant forskel.



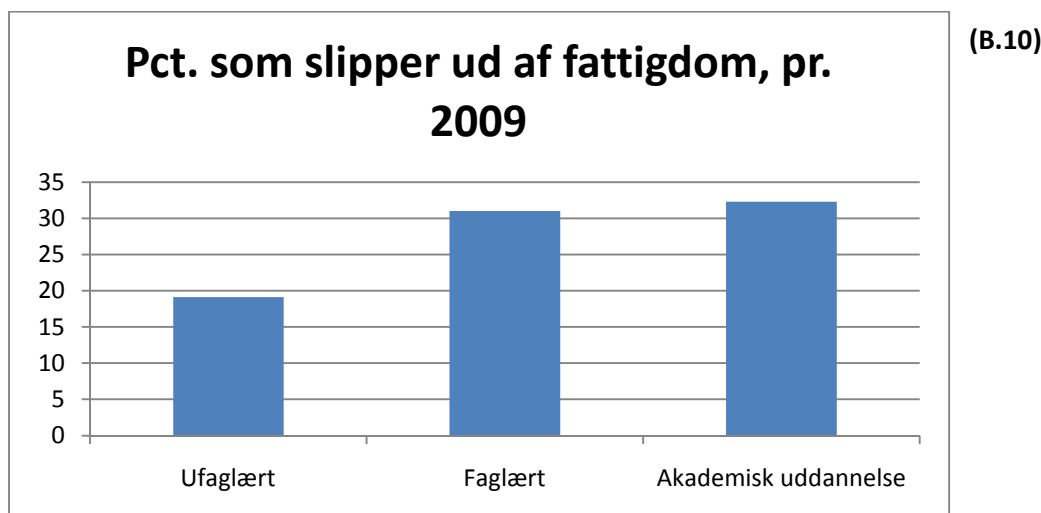
Fra diagrammerne observerer vi, at der er en signifikant opadgående tendens fra 2004 til 2005, og fra 2007 til 2009, hvor der, kun mellem 2003 og 2004 var en nedadgående tendens. Konklusionen bliver altså, at fattigdommen over det sidste årti er vokset signifikant. Vi har her angivet 80% konfidensintervaller,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



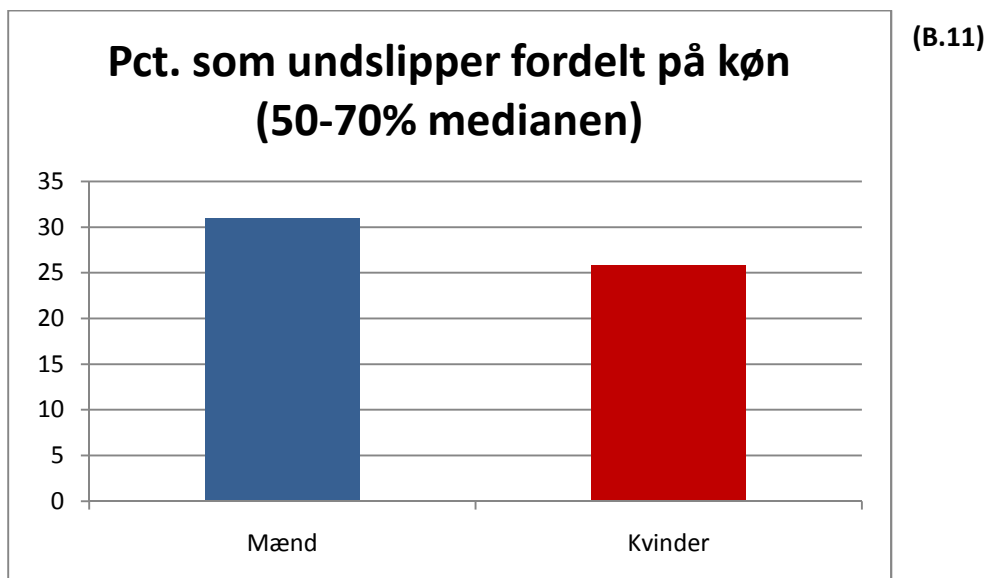
Her ser vi, at chancen for at slippe ud af fattigdom er faldet signifikant siden 2002. Vi oplever altså en særdeles dystre udvikling. På den ene side er fattigdommen stærkt voksende, og på den anden side, så bliver det væsentligt sværere at slippe ud af fattigdommen.

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:



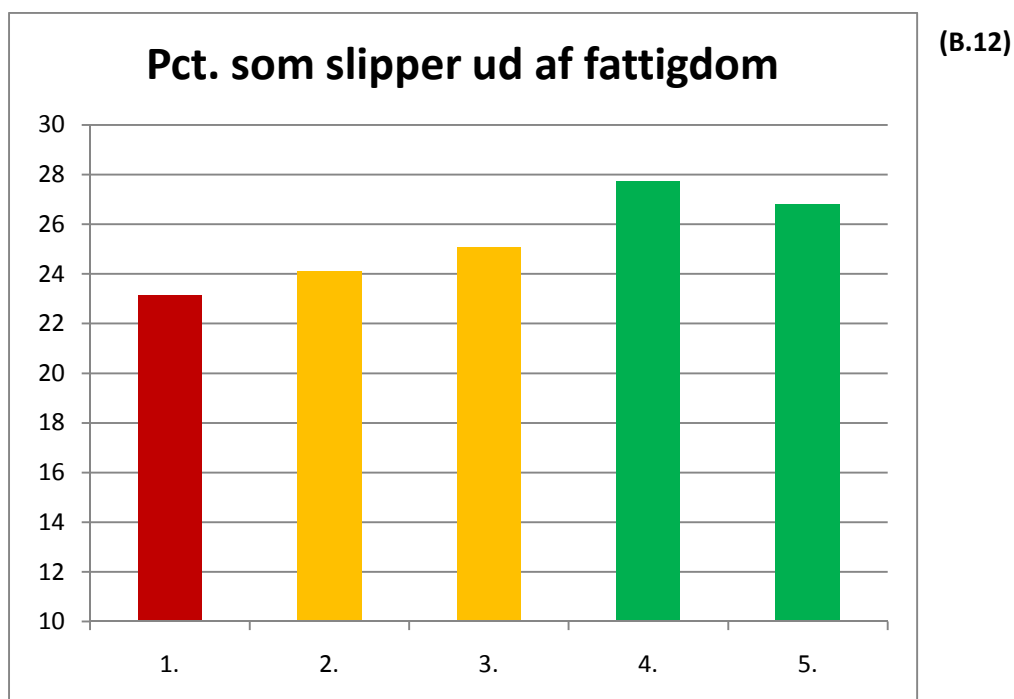
Ufaglærte mennesker har altså en markant dårligere chance for, at undslippe fattigdom. Det er måske ikke i sig selv så overraskende, men forskellen på uddannelsesniveauerne er enorm, og det er et nyt resultat. En faglært eller akademiker vil have den dobbelte chance for at undslippe fattigdom.

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.

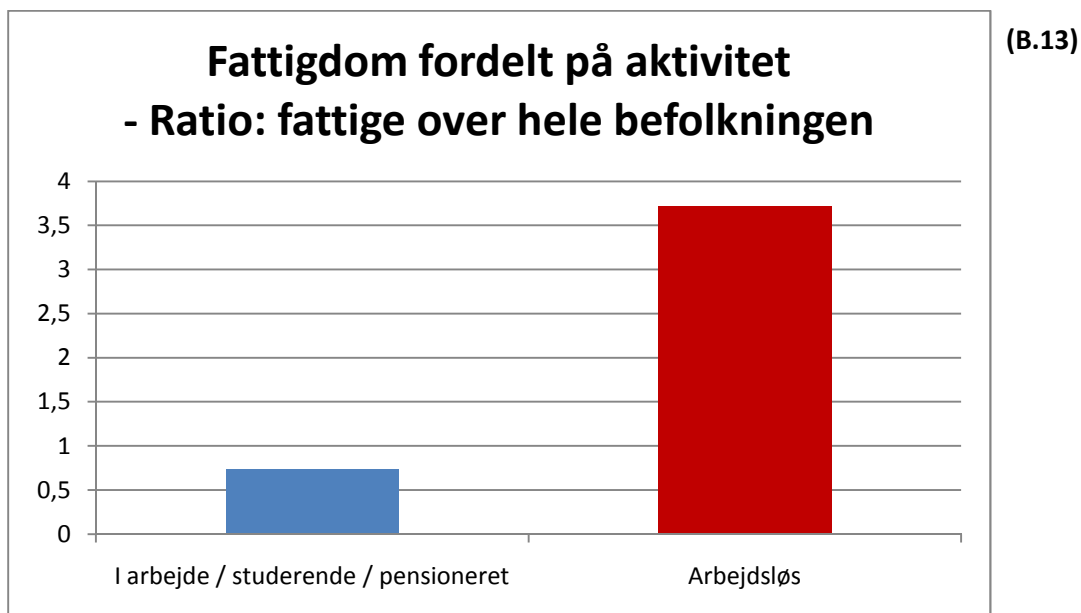


### Resultater: LIS

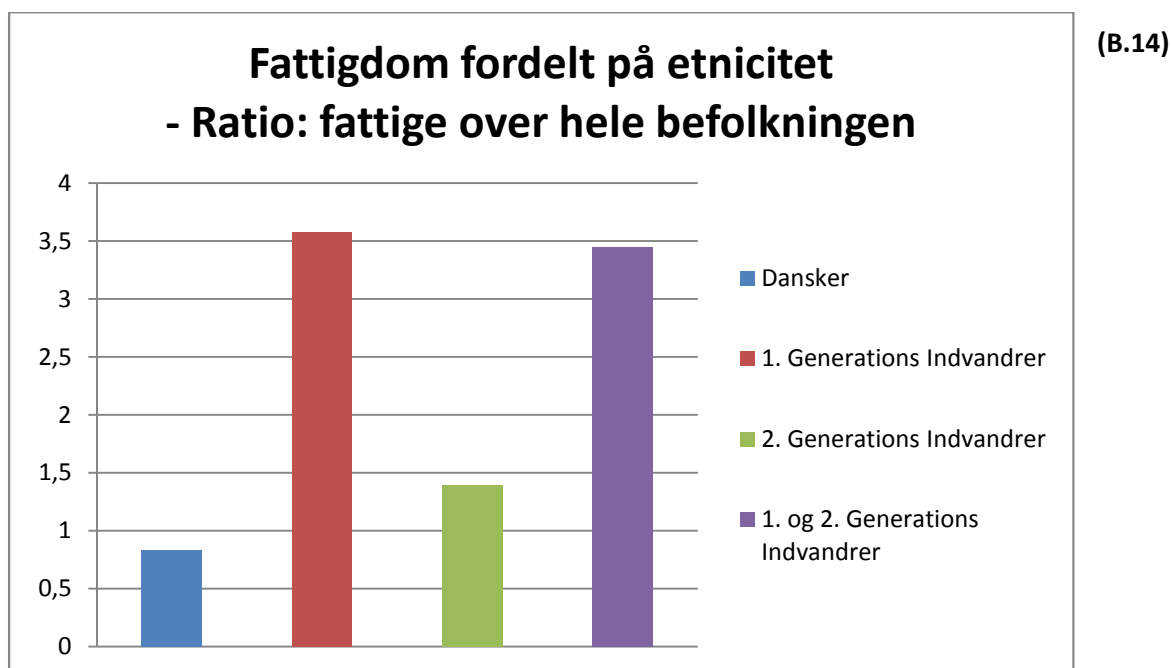
Vi præsenterer nu resultaterne for LIS datasættet. De går alle sammen på året 2007, og det bør igen bemærkes, at der forefindes et større udvalg af resultater i appendikset. Følgende opsummerer således blot de mest centrale resultater og konklusioner. Nedenstående diagram viser den sandsynlighed for, 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.

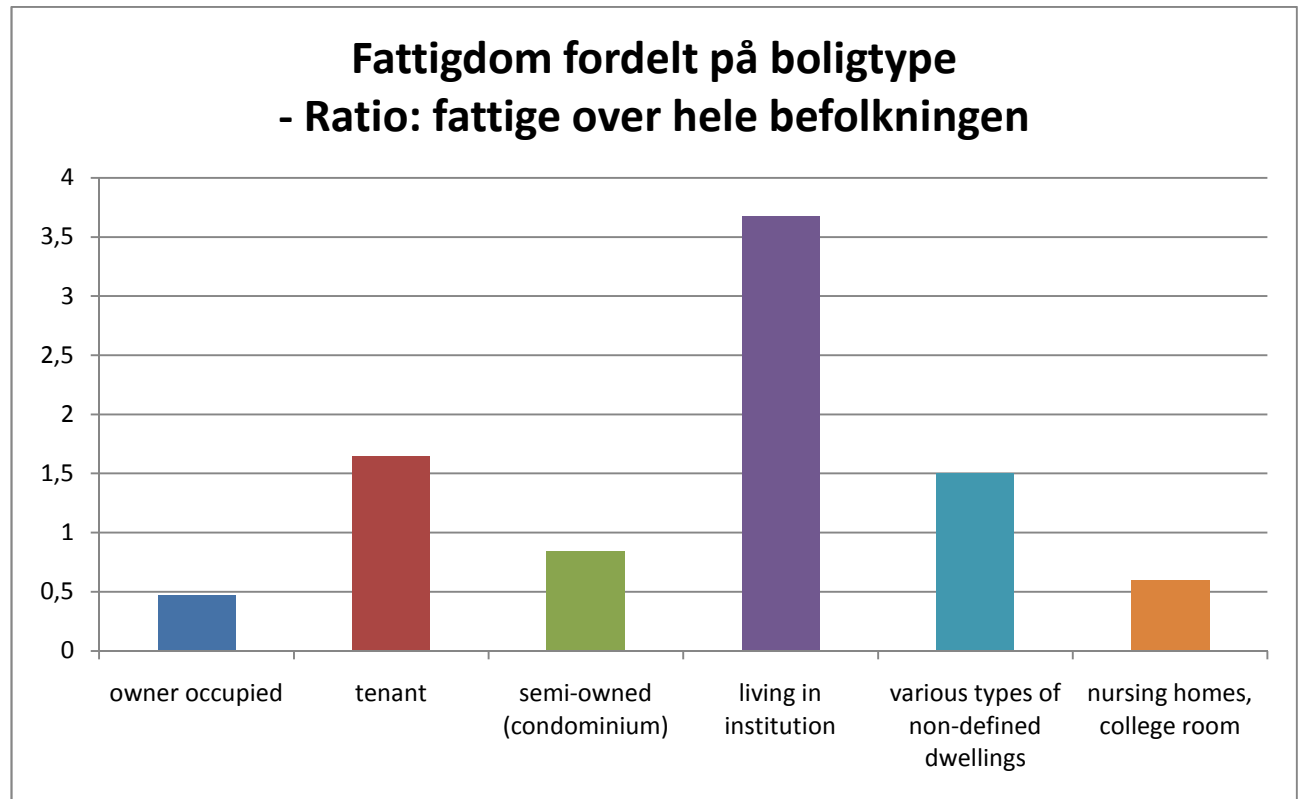


Nedenstående diagram viser en tilsvarende ratiofordelingen af fattige fordelt på etnicitet.



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.

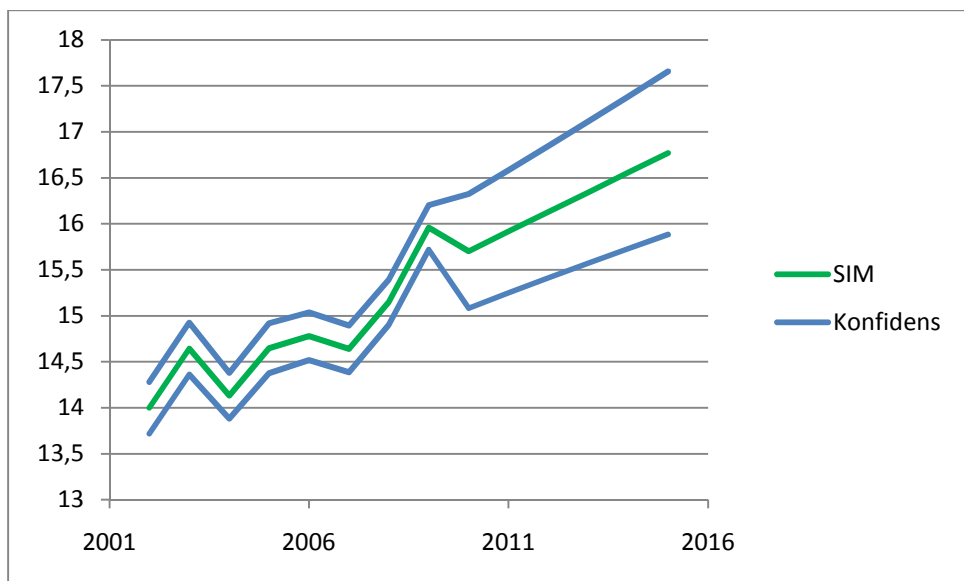


(B.15)

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**

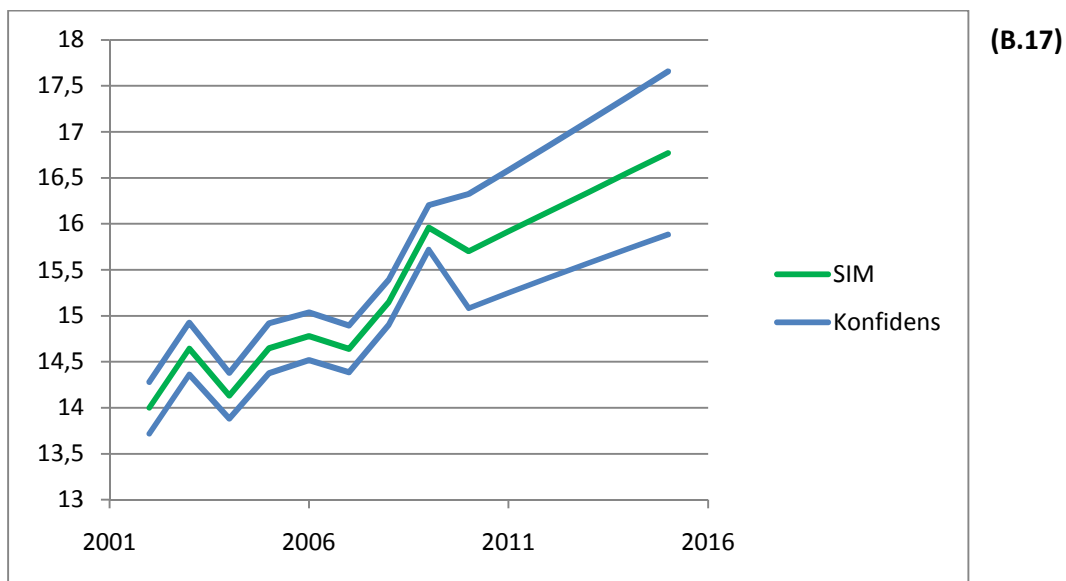
Vi fremskriver SIM 5 år ind i fremtiden ud fra variabelen år som en lineær regression. Diagrammet illustrerer denne fremskrivning samt 80% profil konfidensintervaller. Se appendiks for detaljerede resultater.



(B.16)



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 <sup>[19]</sup>. Se forsat appendikset for detaljerede resultater.

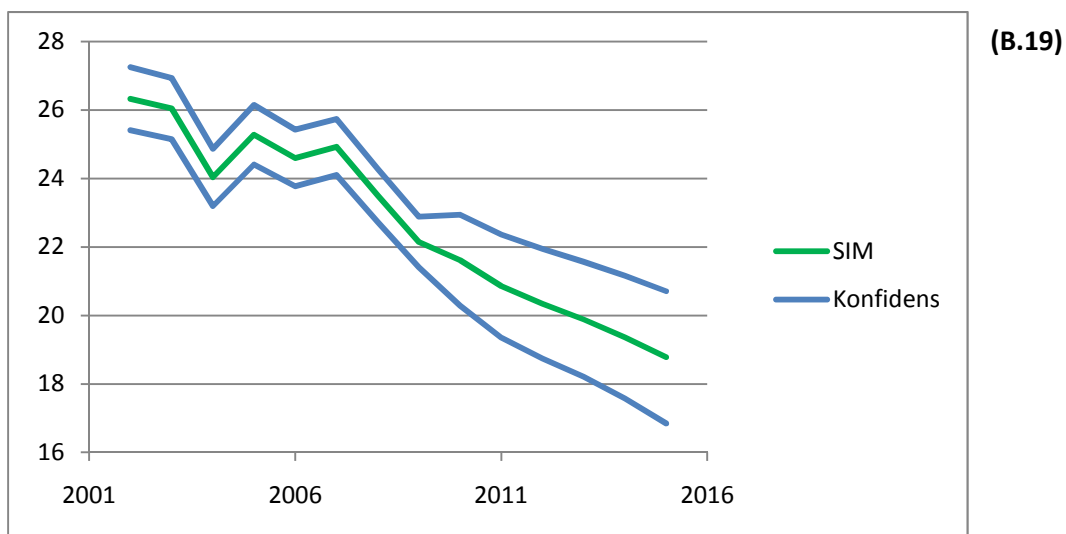


For den anden model er 90%-konfidensintervallerne:

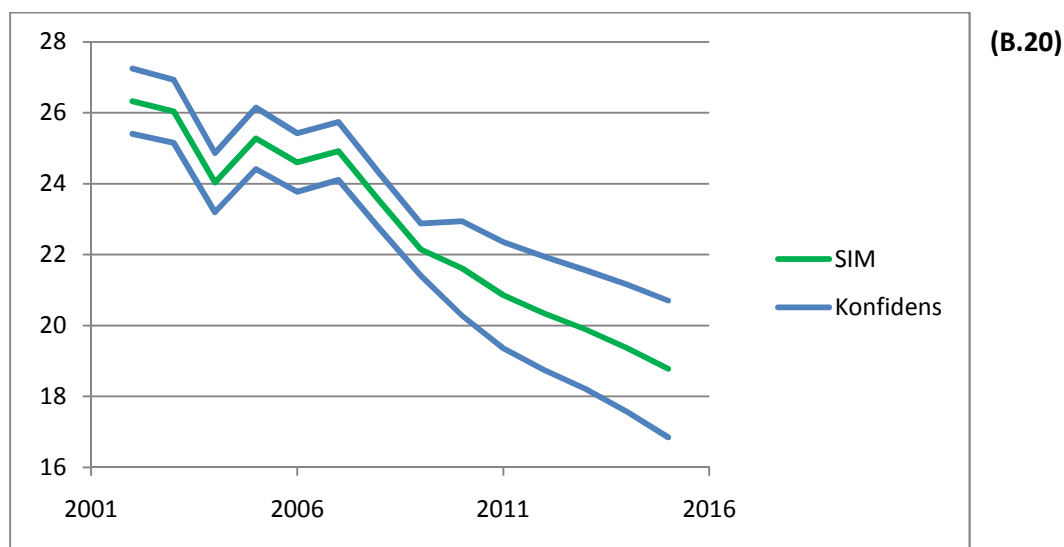
(B.18)

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 <sup>[19]</sup>.



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 <sup>[4]</sup> 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 <sup>[20]</sup>, 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 <sup>[2]</sup> 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 <sup>[1]</sup> 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 Moisiu, 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

## Appendikset

### Om appendikset

Appendikset indeholder overordnet fire afsnit:

- Teoretisk udledning og definition af modeller,
- Supplerende statistiske resultater
- Deskriptiv statistik
- Udeladte undermodeller

De teoretiske afsnit i appendikset er særdeles vigtige. Enhver læser med kendskab til statistik bør, efter at have læst rapporten, læse disse teoriafsnit. De lægger et fundament for de antagelser rapporten bygger på og retfærdiggør meget af det senere arbejde. Dertil viser de også hvordan man kan håndtere en række statistisk teoretiske problemer, samt hvordan den endelige model bør tolkes

De supplerende statistiske resultater er vigtige for forskere og eksperter der arbejder med fattigdom. Med dem kan man trække langt flere konklusioner end den oprindelige rapport ligger op til. Desuden får man også et indblik i hvad man ikke kan udtale sig om, altså hvor det statistiske materiale tyder på insignifikans. Det tilskyndes derfor alle fattigdomsfaglige mennesker, at betragte graferne hertil, og se om de finder vigtigt materiale.

Den deskriptive statistik er anført for, at de statistiske eksperimenter kan gentages. Hertil er det muligt også at få adgang til SAS koden fra projektet - man skal blot sende en email til [julianserban@gmail.com](mailto:julianserban@gmail.com).

Der er to udeladte SSM undermodeller. Den første er modellen for dødsfald, da den giver et definitions-mæssigt absurd syn på fattigdom. F.eks. hvis blot flere fattige mennesker døde, ville fattigdommen i Danmark være faldende, hvilket er uacceptabelt. Den anden model er for pardannelse og separation. Den er udeladt af to lige vigtige årsager, at den tilføjer en enorm kompleksitet til den overordnede model, og at dens indflydelse ville være relativ lille. Begge undermodeller er dog i sig selv udmærkede, og læsere som med særlig interesse for SSM modellen, bør også kigge på disse.

## Indhold

Appendikset .....	1
Om appendikset.....	1
Teori.....	3
Logistisk regression.....	10
Markov model.....	3
SSM modellen .....	11
SSM: Arbejdsløshedsmodel .....	11
SSM: Handikap .....	13
SSM: Uddannelse .....	14
SSM: Fødsler .....	15
PSM modellen.....	17
PSM: Variabel estimator .....	17
PSM: Variabel estimator .....	19
Gennemsnitlig vægtning.....	21
Resultater for data .....	22
ESS: Logistisk Regression .....	22
ESS: Resultater .....	32
LIS: Resultater .....	38
ESS: Fremskrivning.....	44
Deskriptiv statistik.....	51
ESS Data .....	51
ESS Uddannelsesniveauer.....	58
SSM - Udeladte undermodeller .....	59
Pardannelse og separation .....	60
Dødsfald .....	63

## Teori

### Markov model

Igennem vores arbejde, særligt SSM modellen, bruger vi de såkaldte endelige stationære tidskontinuerte markov kæder. At de er endelige antages herfra og diskuteres ikke videre. Modellen defineres formelt matematisk som en mængde  $S$  af  $N$  stadier:

$$\{1, 2, 3, \dots, N\} \in S$$

Mellem to stadier  $i, j \in S, i \neq j$  hører en transitionsintensitet,  $q_{i,j} \geq 0$ . Dertil defineres  $q_{i,i} = -\sum q_{i,j} \forall i \in S$ . Disse intensiteter opskrives typisk som en matrix benævnt intensitetsmatricen (eller infinitesimal generatoren) for en given stationær tidkontinuerte markov kæde:

$$Q = \begin{pmatrix} q_{1,1} & \cdots & q_{1,N} \\ \vdots & \ddots & \vdots \\ q_{N,1} & \cdots & q_{N,N} \end{pmatrix}$$

Til et givent tidspunkt  $t \geq 0$  vil vores markov kæde befinde sig i nøjagtigt ét af vores  $N$  stadier. Fra dette stadie vil den med bevæge sig til et nyt stadie  $j \neq i$ , med eksponentiel sandsynlighedsfordeling  $\exp(q_{i,j})$ . Med andre ord bestemmer intensiteterne hvilken sandsynligheden markov kæden har til at hoppe til et andet stadie.

Givet at kæden i tiden  $t + s$  har foretaget nøjagtigt et hop, kan vi udregne sandsynligheden for at den befinder sig i stadiet  $j$  som værende:

$$P(X(t+s) = j | X(t) = i) = \frac{q_{i,j}}{-q_{i,i}} \quad \forall j \neq i$$

Hvor  $X(t)$  og  $X(t+s)$  definerer stadiet kæden befinder sig i til tiden  $t$  og  $t+s$ . Endvidere kan vi udregne følgende matrice:

$$P(t) = \exp(Qt)$$

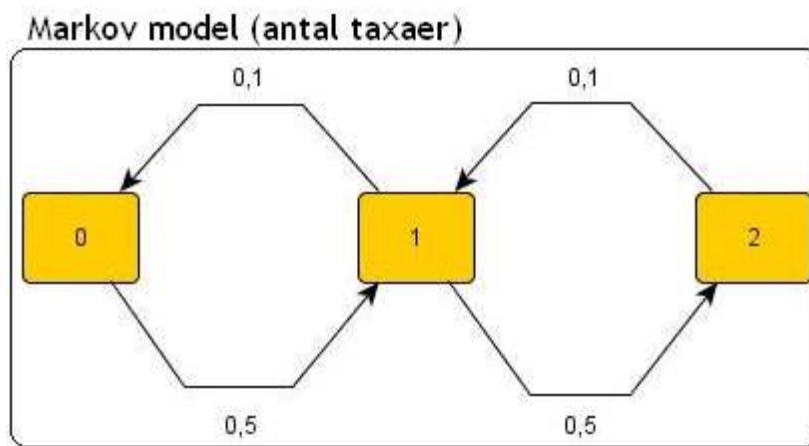
Hvor indgangen  $P(t+s)_{i,j}$  er sandsynligheden for kæden befinder sig i stadiet  $j$  til tiden  $t+s$ , givet at kæden befandt sig i stadiet  $i$  til tiden  $t$ .



Lad os illustrere modellen med et konkret eksempel: antallet af taxaer holdende ved hovedbanegården. Lad os antage at vi kun kigger på tidsrummet 12 - 15 på hverdage, og at der er så meget konstruktionsarbejde at der maksimalt kan holde 2 taxaer. Vi har således tre stadier i vores markov kæde. Lad os gå ud fra at alle taxaer i gennemsnit venter 10 min før de får en passager, men at der samtidig er så meget konkurrence om de få taxaplads, at der i gennemsnit kun går 2 min før en tom taxaplads bliver optaget. Da vi arbejder med eksponential fordelinger, og disse har middelværdi  $1/\lambda$ , hvor  $\lambda$  er fordelingsparameteren for eksponentiel fordelingen kan vi udregne intensiteterne således:

$$q_{0,1} = q_{1,2} = \frac{1}{2} = 0,5, \quad q_{2,1} = q_{1,0} = \frac{1}{10} = 0,1, \quad q_{0,0} = -0,5, \quad q_{2,2} = -0,1, \quad q_{1,1} = -(0,5 + 0,1) = -0,6$$

Hvor alle andre intensiteter er nul. Modellen kan illustreres som nedenstående diagram:



Vi kan desuden udregne vores matrice over sandsynligheder som:

$$Q = \begin{pmatrix} -0,5 & 0,5 & 0 \\ 0,1 & -0,6 & 0,5 \\ 0 & \dots & 0,1 & -0,1 \end{pmatrix}$$

$$P(t) = \exp(Qt)$$

Hvis der f.eks. kl. 15.00 er 2 taxaer, så er sandsynligheden for at der kl 15:15 stadig er 2 taxaer:

$$P(15)_{2,2} = 0.8069 = 80,69\%$$

Det behøver dog ikke at være de samme taxaer. Eksempelvis kunne en af de to taxaer have samlet en passager op, hvorefter en ny taxa har optaget dens plads.

Givet at der kun er en taxa på taxaholdepladsen, kan vi også udregne sandsynligheden for at denne ene taxa skal samle en passager op før en ny taxa ankommer. Det svarer til at vi befinder os i stadiet 1 og går til stadiet 0 før vi går til stadiet 2, og kan udregnes som:

$$\frac{0,1}{0,6} = 0,1667 = 16,67\%$$

Vores intention er at benytte disse markov modeller til at modellere sociologiske processer. Endvidere er det eneste data vi har til rådighed aggregeret data til SSM modellen. Det vil sige at vi prøver at udtale om individers adfærd ud fra kollektiv adfærd. For at dette skal kunne lade sig gøre arbejder vi under en række antagelser:

1. Diskret udfaldsrum
2. Markov-egenskaben
3. Stationær proces (Tidshomogenitet)
4. Processen foregår i kontinuert tid
5. Eksponentielt fordelte spring mellem stadier
6. Homogenitet af befolkningen

Disse antagelser er meget almindelige og udbredte i litteraturen, om end de sjældent bliver så udforsket. Nedenfor diskuterer vi disse antagelser i kontekst om social mobilitet generelt. Når vi i analysen udarbejder de enkelte modeller bliver de mest væsentlige aspekter diskuteret videre.

### Diskret udfaldsrum

Vores første antagelse, som alt vores senere arbejde bygger på, er at vi kan partitionere vores udfaldsrum ind i et endeligt antal stadier. Vi arbejder således med kategorier i vores data, som eksempelvis:

- Indkomstgrupper (0 – 9.999, 10.000 – 19.990 ... )
  - Aldersgrupper (10-14, 15-19 ... )
  - Fattig / Ikke fattig
  - Uddannelsesniveauer
- osv..

Det er klart at man altid vil miste nogle detaljer på denne måde. Et individ kan f.eks. godt være 14,99, i gang med en uddannelse eller ligge i en gråzone mellem to indkomstgrupper. Det er i virkeligheden ikke så væsentligt, da vi arbejder med et stort antal mennesker og derfor vil eventuelle fejl blive fuldstændigt udjævnet. Dertil skal det tilføjes at en overvældende mængde psykologisk teori peger på, at mennesker selv forstår verden i kategorier, eller "kasser" om man vil. Som konsekvens af dette bør måling af fattigdom også konsekvent arbejde med brede kategorier og ikke præcise tal som den rene statistik helst så.

### Markov-egenskaben

Markov-egenskaben er antagelsen om, at sandsynligheden for at bevæge sig til et nyt stadie kun afhænger af det nuværende stadie. I taxa-eksemplet fra før, betyder dette at antallet af taxaer på taxaholdepladsen kun afhænger af hvornår den sidste taxa ankom eller afgik. Vi arbejder i kontinuert tid, så det er umuligt at udtale sig om præcist hvilket tidspunkt der er afgørende. Hvis vi ved at der kl. 12.00 var to taxaer, men kl

13:00 kun en taxa, og vi gerne vil udtale os om hvordan situationen er kl 13:15, så afhænger dette udelukkende af den ene taxa kl 13.00.

Dette giver fin mening i taxa-eksemplet, men når vi skal til at kigge på sociologiske processer kan det blive mere omstændigt og usikkert. Hvis vi f.eks. gerne vil udtale os om hvorvidt en lavtuddannet mand med arbejde i 2008, er blevet arbejdsløs i 2009, vil dette højst sandsynligt også afhænge af hans anciennitet på arbejdspladsen. Hvis han for eksempel i årene 2000-2007 har været arbejdsløs, og netop i 2008 har fået arbejde, er der en væsentlig større sandsynlighed for at han igen i 2009 er arbejdsløs, end hvis han i stedet havde haft dette arbejde igennem 5 år.

McFarland argumenterer for at markov-egenskaben i sociologiske processer, og særligt social mobilitet, har en forvrængede effekt på resultaterne fra modellen. Sociologiske teorier viser at tiden et individ befinder sig i et givent stadie (eksempelvis på en arbejdsplads eller i et ægteskab) påvirker individets billede af verden og derfor hans/hendes valg for at ændre stadie (som eksempelvis at skifte arbejde eller blive separeret). McFarland argumenterer og referer til empiriske studier, som til gengæld viser, at markov-egenskaben bliver mere og mere pålidelig jo mindre tidsrum man arbejder i. Det er f.eks. fuldstændig absurd hvis man kigger på social mobilitet over generationer, men kan godt være acceptabel over en kort årrække.

Vores håb er, at vi ved at anvende disse markov modeller til SSM modellen over en 5-årig periode, har reduceret problemerne ved markov-egenskaben. Vi vil blandt andet argumentere ud fra de resultater vi opnår og andre empiriske studier for at godtgøre denne antagelse. Det kan dog ikke lade sig gøre at måle usikkerheden direkte på denne antagelse – dette ville kræve longitudinal registerdata.

### Stationær proces

Antagelsen om at vi har en stationær proces kan ses som en parallel til markov-egenskaben. Stationær proces betyder at, sandsynlighederne at skifte stadier ikke ændrer sig over tid. Det vil sige, at to identiske individer, hvoraf den første blev arbejdsløs i 2004 og den anden i 2008, begge har samme sandsynlighed for at opnå finde arbejde indenfor en 2-årig periode. I en vis forstand antager vi at resten af verden står stille – dog ikke hele verden, kun den del af verden som kan påvirke vores lille system.

Ligesom markov-egenskaben, betyder det, at vi ved at arbejde med en 5-årig periode har begrænset eventuelle fejl på denne antagelse væsentligt. Overordnet set er denne antagelse ikke væsentlig, da samtlige aggregeret data vi anvender selv stærkt udstråler stationaritet. Eventuelle t-test kan udføres for at eftervise hvor lille en fejl denne antagelse udøver på vores markov kæde. I afsnittet om arbejde og arbejdsløshed fører vid og en kortere diskussion, for at vi i perioden 2002-2008 har et stationært arbejdsmarkedet – da det er klart at en økonomisk krise kan påvirke denne væsentligt. Heldigvis er konsekvenserne af den økonomiske krise ikke begyndt før 2009, og vores antagelse er derfor acceptabel.

### Processen foregår i kontinuert tid

Denne antagelse betyder kort sagt, at spring mellem forskellige stadier kan ske i kontinuert tid. Fra taxa-eksemplet betyder det, at vi kan være i to forskellige stadier for tiderne 12:01 og 12:02, og måske i et helt tredje stadie i 12:01:30. For mange sociologiske processer er dette en fornuftig antagelse, da der ingen begrænsninger er på hvornår de formes. Et ægteskab kan således indgås 1 minut efter to mennesker har mødt hinanden, såvel som 10 år efter at de har mødt hinanden. Det vil naturligvis være højst usandsynligt at to mennesker ville indgå et ægteskab efter 1 minut, men dette bliver korrigeret for af vores eksponentiel fordeling, som givet data vil tvinge denne sandsynlighed til næsten nul.

Man kan dog diskutere hvorvidt kontinuitet altid er fornuftigt i sig selv. Eksempelvis vil vi også gerne modellere individer som forsøger at opnå en højere uddannelse, men uddannelsesinstitutionerne arbejder i hovedreglen udelukkende med faste datoer for indgang og afgang. Ligeledes er det (næsten) umuligt at finde arbejde på en søndag, da samtlige virksomheder holder lukket.

Antagelsen er dog meget mild, da modellen vil blive korrigeret af de data vi benytter. Data på uddannelser, parforhold, dødsfald etc... vil alle sammen afspejle de underliggende brud i den kontinuerte tid. Når vi opstiller modellerne for 5-årige perioder opnår vi derfor minimale fejl, som alle sammen er bør være jævnt fordelt omkring vores resultater.

Grunden til at vi vælger at arbejde med kontinuert tid, og ikke diskret tid, er først og fremmest fordi en opdeling af tiden i diskrete mængder vil medføre væsentlige fejl på aggregeret niveau, og desuden komplicere modellen unødvendigt.

### Eksponentielt fordelte spring mellem stadier

Eksponentiefordelingen er en sandsynlighedsfordeling som har en konstant rate. Det vil sige, at der er lige sandsynligheden er konstant for alle intervaller af en bestemt længde. Denne konstante rate medfører den særlige egenskab at den er hukommelsesløs. Netop derfor anvendes markov modeller hovedsagligt med denne fordeling. At den er hukommelsesløs er nemlig ensbetydende med at den overholder markov-egenskaben. Derudover er eksponentiefordeling bredt anvendt til at modellere en række sociologiske processer, og kan opfange mange af de underliggende processer som ikke kan observeres direkte.

Da vi stadig arbejder i en 5-årige periode, og desuden kontrollerer for en række forhold som køn, alder og uddannelse, virker det som en rimelig antagelse at vi har en konstant rate af sandsynlighed for hop i mellem stadier.

### Homogenitet af befolkningen

At vi antager homogenitet af befolkningen dækker over to ting. Først og fremmest antager vi, at aggregeret data er proportionelt til individuelt data. Det vil sige, at den gennemsnitlige intensitet for at hoppe mellem to stadier for er den samme for hele befolkningen såvel for individet. Sekundært antager vi at de individer

som ikke handler proportionelt til befolkningen er normaltfordelte omkring gennemsnittet på tværs af de parametre vi ikke kontrollerer.

Der er muligvis mange ting man ikke opfanger igennem aggregeret data, især når vi ikke kan kontrollere for indkomstniveauet i husstanden som et individ lever i. Det er klart, at der foregår to forskellige sociologiske processer for individer med samme køn, alder og uddannelse. Men hvis det ene individ hører til en økonomisk stærk husstand og den anden til en økonomisk svag husstand, vil deres handlinger afspejle dette. Dette kan vi ikke differentiere i mellem og vil derfor forvente at begge individer opfører sig ens igennem vores analyse.

Dette er således måske den mest omstridte antagelse vi foretager. Derfor vil vi løbende igennem vores arbejde benytte andre empiriske analyser til at begrænse fejl og forskubbe eventuelle fejl, således at hvis vi har fejl i vores resultater, da vil de fremstille et mere positivt mål på social mobilitet end det der reelt eksisterer. Denne tilgang retfærdiggøres af vores hypotese: at social mobilitet er langt ringere for husstande i de laveste indkomstgrupper end for alle andre. Ved at fremstille den sociale mobilitet mere positivt vil eventuelle resultater som bekræfter vores hypotese, ikke kun være tilstrækkelige for at bekræfte vores hypotese, men også skabe stærk formodning om at den sociale mobilitet er endnu værre. Vi vælger så at sige, kun at se toppen af isbjerget. Men det vil uden tvivl være det rigtige isbjerg vi kigger på.

Kilder:

INTRAGENERATIONAL SOCIAL MOBILITY AS A MARKOV  
PROCESS: INCLUDING A TIME-STATIONARY MARKOVIAN  
MODEL THAT EXPLAINS OBSERVED  
DECLINES IN MOBILITY RATES  
OVER TIME, David McFarland, 1970

On Estimating Transition Intensities of a Markov  
Process with Aggregate Data of a Certain Type:  
"Occurrences but no Exposures", 1986 Richard D. Gill

## Logistisk regression

Vi benytter logistisk regression til vores PSM model. Formelt definer vi en række binomial fordelte variable:

$$Y_i \sim \text{Bin}(n_i, p_i) \quad i \in \{1, 2, 3, \dots, N\}$$

Hvor antalsparameteren  $n_i$  er velkendt, mens  $p_i$  antages ukendt. Vi opstiller nu:

$$p_i = E\left(\frac{Y_i}{n_i} \mid X_i\right)$$

$$\text{logit}(p_i) = \ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_{1,i} + \dots + \beta_N x_{N,i}$$

Hvor  $X_i$  er en stokastisk variabel vektor, som indeholder de beskrivende kategoriske variable, og  $x_{j,i}$  er den  $j$ 'te indgang til  $X_i$ . Her er  $\beta_0, \dots, \beta_N$  parametre til modellen som estimeres gennem maximum likelihood estimation.

For en observeret stokastisk variabel vektor  $Z$ , i tilknytning til en uobserveret stokastisk bernoulli-fordelt variabel  $W$ , kan vi udregne sandsynligheden:

$$P(W = 1) = \frac{1}{1 + \exp(\beta_0 + \beta_1 z_1 + \dots + \beta_N z_N)}$$

Hvor  $z_j$  er den  $j$ 'te indgang til  $Z$ .

Da vores  $Y_i$ 'er er binomialt fordelte, vil fejlen på vores logistiske regression også være binomialt fordelte. Det giver således ikke mening at uddybe fordelingen af vores fejl teoretisk. I stedet skal denne betragtes for den konkrete model.

Vores intention er at benytte logistisk regression til at modellere sociologiske processer, som beskrevet tidligere. For at dette skal kunne lade sig gøre arbejder vi under en række antagelser, hovedsagligt:

1. Logistisk (odds-lineær) sammenhæng mellem beskrivende variable og indkomst-mobilitet
2. Stationær proces (Tidshomogenitet)
3. Homogenitet af befolkningen (Se afsnittet for markov modeller)

Disse antagelser er alle sammen beskrevet i afsnittet for markov modeller, på nær den første om den logistiske sammenhæng mellem vores variable. Denne antagelse bygger som beskrevet i ligningen ovenfor på, en eksponentiel lineær sammenhæng af vores beskrivende variable. Men da vores variable i både definitionen og praksis er kategoriske, så reduceres antagelsen betragteligt. Vi får nemlig tildelt en konstant til hvert eneste udfald i hver eneste beskrivende variabel, for derefter at maksimere vores likelihood. Antagelsen bliver derfor en antagelse på tværs af de beskrivende variable, og ikke imellem dem. Da vi stadigvæk arbejder med spatielt data, bør modellens antagelser være relativt milde. Modellen er derfor, næsten udelukkende, kun begrænset af de beskrivende variable vi har til rådighed.

## SSM modellen

### SSM: Arbejdsløshedsmodel

Vores statistikker udregnet ud fra Laborta Data og OECD Statistics:

Uddannelsesniveau (EU Standard)			Level 0-2	Level 3-4	Level 5-,6	
Sandsynlighed for at have et job 5 år senere, givet at individet ikke har job på nuværende tidspunkt	Mænd	18 til 24	0,660410653	0,662508037	0,668728841	
		25 til 29	0,670406395	0,683673419	0,718053177	
		25 til 54	0,781171494	0,79225	0,819765784	
		30 til 34	0,739878626	0,75187909	0,782228502	
		35 til 39	0,772348602	0,783634133	0,81178118	
		40 til 44	0,815269181	0,825498349	0,850448349	
		45 til 49	0,838830313	0,848428447	0,871492858	
		50 til 54	0,862221046	0,871158507	0,89226406	
		55 til 59	0,777597353	0,81500439	0,880483842	
		55 til 64	0,792264943	0,829182887	0,892593372	
		60 til 64	0,821114528	0,857005323	0,916197684	
		65 til 69	0,778728347	0,816098489	0,881420277	
		70 til 74	0,90152219	0,934129246	0,980613021	
		70+	0,90152219	0,934129246	0,980613021	
		75+	0,90152219	0,934129246	0,980613021	
		alle	0,757682031	0,770051343	0,800745471	
		Kvinder	20 til 24	0,614715318	0,60862797	0,620019079
			25 til 29	0,663691984	0,63801438	0,683003878
	25 til 54		0,782436246	0,761170498	0,797823888	
	30 til 34		0,753961157	0,731426563	0,770447549	
	35 til 39		0,772783478	0,751073333	0,788554181	
	40 til 44		0,814466452	0,794777627	0,828507316	
	45 til 49		0,83683202	0,818335095	0,84986555	
	50 til 54		0,861134612	0,844014706	0,87301345	
	55 til 59		0,832886304	0,73589756	0,876047377	
	55 til 64		0,844310131	0,748137802	0,886299542	
		60 til 64	0,878526197	0,785025579	0,91685373	
		65 til 69	0,874439191	0,780601979	0,913215854	
	70 til 74	0,945695619	0,858371104	0,976212561		
	70+	0,945695619	0,858371104	0,976212561		
	75+	0,945695619	0,858371104	0,976212561		
	alle	0,755921205	0,730072082	0,774400436		



Uddannelsesniveau (EU Standard)			Level 0-2	Level 3-4	Level 5-,6
Sandsynlighed for at have et job 5 år senere, givet at individet HAR job på nuværende tidspunkt	Mænd	18 til 24	0,661217676	0,663280414	0,669404721
		25 til 29	0,67539426	0,687699613	0,720151998
		25 til 54	0,792035845	0,801019679	0,824337345
		30 til 34	0,748217315	0,758610061	0,785737302
		35 til 39	0,782638586	0,791940184	0,816111056
		40 til 44	0,828534366	0,836205976	0,856030146
		45 til 49	0,853925055	0,860612889	0,877844506
		50 til 54	0,879272073	0,884922058	0,899438885
		55 til 59	0,838436237	0,855698869	0,892460979
		55 til 64	0,857092259	0,872545186	0,905355697
		60 til 64	0,89415755	0,905863019	0,930577406
		65 til 69	0,839870217	0,856995632	0,893457062
		70 til 74	1	1	1
		70+	1	1	1
	75+	1	1	1	
	alle		0,767882777	0,778209823	0,804880858
	Kvinder	20 til 24	0,615164665	0,609134277	0,62021167
		25 til 29	0,665879531	0,641361631	0,684520533
		25 til 54	0,788078644	0,769804153	0,801735838
		30 til 34	0,758575719	0,738487486	0,773646887
		35 til 39	0,7780624	0,75915082	0,792214128
		40 til 44	0,821428844	0,805431059	0,833334436
		45 til 49	0,844822782	0,830562077	0,855405653
		50 til 54	0,870345188	0,858108173	0,879399267
		55 til 59	0,866989063	0,824839486	0,890985717
		55 til 64	0,880232789	0,841826126	0,902035069
		60 til 64	0,920205767	0,893728255	0,935111009
		65 til 69	0,915407046	0,88744846	0,931161373
		70 til 74	1	1	1
		70+	1	1	1
		75+	1	1	1
		alle		0,761246416	0,738585203

**SSM: Handikap**

De statistiske resultater er givet i nedenstående tabeller:

Sandsynligheden for at blive alvorligt fysisk hæmmet efter 5-årig periode			
Uddannelsesniveau (EU Standard)	Alder	Mænd	Kvinder
Level 0-2	15-24	0,010217	0,014832
Level 0-2	25-34	0,034599	0,058494
Level 0-2	35-44	0,026284	0,03933
Level 0-2	45-54	0,04006	0,025564
Level 0-2	55-64	0,040791	0,058122
Level 0-2	65-74	0,042986	0,020183
Level 0-2	75-84	0,018753	0,043719
Level 3-4	15-24	0,004215	0,024845
Level 3-4	25-34	0,010217	0,017682
Level 3-4	35-44	0,020183	0,031336
Level 3-4	45-54	0,023767	0,031336
Level 3-4	55-64	0,016969	0,055892
Level 3-4	65-74	0,024485	0,023049
Level 3-4	75-84	0,026284	0,054779
Level 5-6	15-24	0,007388	0,009509
Level 5-6	25-34	0,023767	0,014476
Level 5-6	35-44	0,012344	0,019825
Level 5-6	45-54	0,043352	0,037143
Level 5-6	55-64	0,031698	0,037143
Level 5-6	65-74	0,039695	0,031698

### SSM: Uddannelse

Der foreligger ingen pålidelige data for årene 18-24 blandt tilgang af elver til Level 5-6 uddannelser.

Årsagen til dette er formentlig at der er høj tilgang, høj frafald og et stort antal forkerte registreringer. For at sikre os at vi ikke skaber et negativt bias, vil vi igen foretage en antagelse. Vi antager at alle mellem 18-24, som har en Level 3-4 uddannelse, påbegynder en Level 5-6 uddannelse.

De statistiske resultater er givet i nedenstående tabel:

Sandsynlighed for at opnå højere uddannelsesniveau

ISCED97	Alder	Mænd	Kvinder
Level 3-4	18	0,10677693	0,092898615
Level 3-4	19	0,026442231	0,029656838
Level 3-4	20 - 24	0,019476158	0,024620518
Level 3-4	25 - 29	0,012815545	0,021013027
Level 3-4	30 - 34	0,006582434	0,019426392
Level 3-4	35 - 39	0,004195033	0,014449773
Level 3-4	40+	0,001611137	0,002557356
Level 5-6	18	1	1
Level 5-6	19	1	1
Level 5-6	20 - 24	1	1
Level 5-6	25 - 29	0,032509261	0,039962625
Level 5-6	30 - 34	0,012097579	0,017124927
Level 5-6	35 - 39	0,00485257	0,009438437
Level 5-6	40+	0,001091841	0,002520304

De 3 rækker markeret med rød er estimeret til 100% her, da der kun foreligger fejlfyldt data om disses sandsynlighed for at starte på en uddannelse. Antagelsen bør udelukkende skabe et mere positivt syn på den sociale mobilitet, og vi opnår således stadigvæk solide resultater ved at bruge disse data senere hen.

**SSM: Fødsler**

Sandsynlighed for at føde levende antal børn efter 5-årig periode, givet at moderen IKKE har nogen børn, afhængig af moderens alder og højeste uddannelsesniveau (EU Standard).

Level 0-2 (ISCED 1997)	Moderens alder	Et barn	To børn
	20-24	0,216257303	0,028383
	25-29	0,344414868	0,492621
	30-34	0,280109497	0,62157
	35-39	0,382376802	0,169704
	40-44	0,100024918	0,004927
	45-49	0,004023603	6,99E-06
	50+	1,51212E-05	9,83E-11
Level 3-4 (ISCED 1997)	20-24	0,077620808	0,00287
	25-29	0,087129778	0,003667
	30-34	0,058972609	0,001614
	35-39	0,028102696	0,000352
	40-44	0,005246676	1,19E-05
	45-49	0,000387455	6,46E-08
	50+	6,97184E-06	2,09E-11
Level 5-6 (ISCED 1997)	20-24	0,311326311	0,076239
	25-29	0,232796141	0,034128
	30-34	0,18917928	0,020543
	35-39	0,100463101	0,004974
	40-44	0,021501671	0,000204
	45-49	0,000913324	3,59E-07
	50+	1,18985E-05	6,09E-11

Sandsynlighed for at føde levende antal børn efter 5 årig periode, givet at individet KUN har et barn, afhængig af moderens alder og højeste uddannelsesniveau (EU Standard).

Level 0-2 (ISCED 1997)	Antal Børn:	Et barn	To børn
	20-24	0,203778122	0,009766
	25-29	0,568052408	0,212425
	30-34	0,567024735	0,286593
	35-39	0,431309997	0,06325
	40-44	0,089126793	0,001651
	45-49	0,003463338	2,3E-06
	50+	1,29989E-05	3,24E-11
Level 3-4 (ISCED 1997)	20-24	0,068559535	0,000958
	25-29	0,077240941	0,001226
	30-34	0,051727677	0,000537
	35-39	0,024383699	0,000116
	40-44	0,004517896	3,93E-06
	45-49	0,000333113	2,13E-08
	50+	5,99328E-06	6,88E-12
Level 5-6 (ISCED 1997)	20-24	0,314887277	0,027095
	25-29	0,221566219	0,011797
	30-34	0,175606501	0,007019
	35-39	0,089533045	0,001667
	40-44	0,018614643	6,75E-05
	45-49	0,00078536	1,18E-07
	50+	1,02284E-05	2E-11

Sandsynlighed for at føde levende antal børn efter 5 årig periode, givet at individet HAR to børn, afhængig af moderens alder og højeste uddannelsesniveau (EU Standard).

Level 0-2 (ISCED 1997)	Antal Børn:	Et barn	To børn
	20-24	0,085658873	0,00103
	25-29	0,381902741	0,02835
	30-34	0,438169295	0,040814
	35-39	0,214348353	0,007276
	40-44	0,03543	0,000169
	45-49	0,001328731	2,31E-07
	50+	4,98041E-06	3,25E-12
Level 3-4 (ISCED 1997)	20-24	0,027011576	9,76E-05
	25-29	0,030545893	0,000125
	30-34	0,020235656	5,45E-05
	35-39	0,00943269	1,17E-05
	40-44	0,001734032	3,94E-07
	45-49	0,000127645	2,13E-09
	50+	2,29627E-06	6,91E-13
Level 5-6 (ISCED 1997)	20-24	0,141685443	0,002961
	25-29	0,094051766	0,001251
	30-34	0,072732059	0,000734
	35-39	0,035597879	0,000171
	40-44	0,007184381	6,8E-06
	45-49	0,000300994	1,19E-08
	50+	3,91893E-06	2,01E-12

## PSM modellen

### PSM: Variabel estimater

Nedenstående er vores maximum likelihood parametre med deres værdier og tilhørende 90%-konfidensintervaller for den primære PSM model:

#### Profile Likelihood Confidence Interval for Parameters Primær PSM, husstande uden partner

Parameter		Estimate	90% Confidence Limits	
Intercept		5.0963	0.2834	.
GNDR	1	-0.00364	-0.0923	0.0852
BLGETMG	-1	8.0934	-0.7834	.
BLGETMG	0	-3.8298	.	0.6690
EDULVL	0	0.3059	0.1640	0.4482
EDULVL	1	-0.2258	-0.3499	-0.1024
EDULVLM	-1	-2.4209	-5.8171	-0.8199
EDULVLM	0	0.7676	0.2137	1.9048
EDULVLM	1	0.6407	0.0837	1.7790
EDULVLF	-1	0.7282	-0.0247	1.6383
EDULVLF	0	-0.1022	-0.4351	0.1897
EDULVLF	1	-0.2974	-0.6219	-0.0164
EDCTN	0	-1.0408	-1.1941	-0.8920
Head_Unemployed	0	-0.0323	-0.1971	0.1309
DSBLD	0	0.1344	-0.0559	0.3239
RTRD	0	-0.1011	-0.2171	0.0149
Number_Of_Children_S	0	-0.4353	-0.8565	-0.0156
Number_Of_Children_S	1	0.5575	0.2321	0.8936
HHMMB	1	0.3304	-0.0131	0.6685
HHMMB	2	0.0126	-0.2879	0.3063

**Profile Likelihood Confidence Interval for Parameters**  
Primær PSM, husstande med partner

<b>Parameter</b>		<b>Estimate</b>	<b>90% Confidence Limits</b>	
Intercept		-4.0219	.	1.7710
GNDR	1	-0.2205	-0.4221	-0.0216
BLGETMG	-1	0.8291	-0.3633	2.6197
BLGETMG	0	-1.0550	-2.1189	-0.1691
EDULVL	0	-0.0640	-0.4333	0.2973
EDULVL	1	-0.2930	-0.6072	0.0110
EDULVLP	-1	-0.8552	-10.4814	3.1506
EDULVLP	0	0.7821	-0.5792	3.9987
EDULVLP	1	0.3219	-1.0348	3.5367
EDULVLM	-1	-0.8377	-2.5651	0.7723
EDULVLM	0	0.3898	-0.2348	1.0505
EDULVLM	1	0.7689	0.0973	1.4801
EDULVLF	-1	0.5917	-0.6382	2.1776
EDULVLF	0	-0.4036	-0.9988	0.1044
EDULVLF	1	-0.3283	-0.9346	0.1941
EDCTN	0	-0.0242	-0.6716	0.5731
Head_Unemployed	0	0.1620	-0.3985	0.6969
DSBLD	0	0.2617	-0.2635	0.7588
RTRD	0	-0.0543	-0.4128	0.3135
Number_Of_Children_S	0	-0.6832	-2.0593	0.6275
Number_Of_Children_S	1	0.2816	-0.5756	1.1745
Spouse_Unemployed	0	-0.4125	-1.0216	0.1268
DSBLDP	0	-0.2317	-1.0798	0.4670
RTRDP	0	-0.0547	-0.4294	0.3273
EDCTNP	0	0.4380	-0.1659	1.0574
HHMMB	1	-8.6673	.	6.8719
HHMMB	2	4.6067	-3.1879	.

**PSM: Variabel estimer**

Vi præsenterer nu estimerne for den sekundære PSM model, altså estimerne 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

**Maximum Likelihood Analysis of Variance**

Sekundær PSM, husstande uden partner

Source	DF	Chi-Square	Pr > ChiSq
-----	-----	-----	-----
Intercept	1	73,38	<,0001
D4	2	374,4	<,0001
D27	3	290,31	<,0001
MARRIED	0*	,	,
Hpeduc	2	806,93	<,0001
Hpdisabl	1	1006,09	<,0001
hpumas	1	130,47	<,0001
hpethnat	2	110,98	<,0001
Likelihood Ratio	147	1637,02	<,0001

**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

**Maximum Likelihood Analysis of Variance**

Sekundær PSM, husstande med partner

Source	DF	Chi-Square	Pr > ChiSq
-----	-----	-----	-----
Intercept	1	47,2	<,0001
D4	1	17,11	<,0001
D27	3	33,04	<,0001
MARRIED	1	10,05	0,0015
Hpeduc	2	93,21	<,0001
Hpdisabl	1	19,04	<,0001
Hpumas	1	33,01	<,0001
Hpethnat	2	35,93	<,0001
Speduc	2	22,22	<,0001
Spdisabl	1	115,11	<,0001
Spumas	1	50,99	<,0001
Spethnat	2	24,15	<,0001
Likelihood Ratio	1,00E+03	1719,27	<,0001

**Gennemsnitlig vægtning**

Datasæt	Kategori / År	Gennemsnitlig vægtning for alle husstande							
		2009	2008	2007	2006	2005	2004	2003	2002
2008	Under 70%-medianen	2,58	2,55	2,47	2,48	2,47	2,33	2,48	2,37
	Over 70%-medianen	0,82	0,83	0,84	0,84	0,84	0,85	0,84	0,85
2006	Under 70%-medianen	2,46	2,42	2,35	2,37	2,36	2,23	2,37	2,28
	Over 70%-medianen	0,76	0,77	0,78	0,78	0,78	0,79	0,78	0,79
2004	Under 70%-medianen	2,43	2,39	2,32	2,34	2,33	2,20	2,34	2,25
	Over 70%-medianen	0,76	0,76	0,77	0,77	0,77	0,78	0,77	0,78
2002	Under 70%-medianen	2,46	2,42	2,35	2,37	2,36	2,23	2,37	2,28
	Over 70%-medianen	0,76	0,77	0,78	0,78	0,78	0,79	0,78	0,79

For at vurdere konsekvensen af vores vægtning er ovenstående skema af vægtenes gennemsnit angivet. Der er et klart system i, at husstande under 70% medianen skal vægtes 2,5, og husstande over 70% medianen kun skal vægtes omkring 0,8. Det indikerer, at vores samples er pænt fordelte, og at der ikke sker uigennemskuelige beregninger.

## Resultater for data

### ESS: Logistisk Regression

#### Logistic Regression Results (Husstande uden partner)

##### The LOGISTIC Procedure

##### Model Information

Data Set	SASUSER.FILTER_FOR_CON_DATA_FORMATT_0005
Response Variable	In_Poverty
Number of Response Levels	2
Weight Variable	Weight_Household_2009
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	849
Number of Observations Used	849
Sum of Weights Read	1956.889
Sum of Weights Used	1956.889

##### Response Profile

Ordered Value	In_Poverty	Total Frequency	Total Weight
1	0	425	641.2671
2	1	424	1315.6223

Probability modeled is In\_Poverty='1'.

##### Class Level Information

Class	Value	Design	Variables	
GNDR	1	1		
	2	-1		
BLGETMG	-1	1	0	
	0	0	1	
EDULVL	1	-1	-1	
	0	1	0	
EDULVLM	1	0	1	
	2	-1	-1	
EDULVLM	-1	1	0	0
	0	0	1	0
EDULVLF	1	0	0	1
	2	-1	-1	-1
EDULVLF	-1	1	0	0

	0	0	1	0
	1	0	0	1
	2	-1	-1	-1
EDCTN	0	1		
	1	-1		
Head_Unemployed	0	1		
	1	-1		
DSBLD	0	1		
	1	-1		
RTRD	0	1		
	1	-1		
Number_Of_Children_SEC	0	1	0	
	1	0	1	
	2	-1	-1	
HHMMB	1	1	0	
	2	0	1	
	3	-1	-1	

### Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

### Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	2477.605	2259.832
SC	2482.349	2354.713
-2 Log L	2475.605	2219.832

### Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	255.7732	19	<.0001
Score	230.4672	19	<.0001
Wald	193.4027	19	<.0001

**Type 3 Analysis of Effects**

Effect	DF	Wald	
		Chi-Square	Pr > ChiSq
GNDR	1	0.0046	0.9462
BLGETMG	2	1.4570	0.4826
EDULVL	2	19.6619	<.0001
EDULVLM	3	7.6953	0.0527
EDULVLF	3	4.8543	0.1828
EDCTN	1	128.6529	<.0001
Head_Unemployed	1	0.1055	0.7454
DSBLD	1	1.3617	0.2433
RTRD	1	2.0562	0.1516
Number_Of_Children_S	2	10.7154	0.0047
HHMMB	2	2.9117	0.2332

**Analysis of Maximum Likelihood Estimates**

Parameter	DF	Estimate	Standard Error	Wald	
				Chi-Square	Pr > ChiSq
Intercept	1	5.0963	155.4	0.0011	0.9738
GNDR	1	-0.00364	0.0539	0.0046	0.9462
BLGETMG	-1	8.0934	310.8	0.0007	0.9792
BLGETMG	0	-3.8298	155.4	0.0006	0.9803
EDULVL	0	0.3059	0.0863	12.5473	0.0004
EDULVL	1	-0.2258	0.0752	9.0153	0.0027
EDULVLM	-1	-2.4209	1.2826	3.5626	0.0591
EDULVLM	0	0.7676	0.4352	3.1112	0.0778
EDULVLM	1	0.6407	0.4367	2.1524	0.1423
EDULVLF	-1	0.7282	0.4948	2.1658	0.1411
EDULVLF	0	-0.1022	0.1870	0.2986	0.5847
EDULVLF	1	-0.2974	0.1809	2.7020	0.1002
EDCTN	0	-1.0408	0.0918	128.6529	<.0001
Head_Unemployed	0	-0.0323	0.0996	0.1055	0.7454
DSBLD	0	0.1344	0.1152	1.3617	0.2433
RTRD	0	-0.1011	0.0705	2.0562	0.1516
Number_Of_Children_S0	1	-0.4353	0.2551	2.9124	0.0879
Number_Of_Children_S1	1	0.5575	0.2005	7.7348	0.0054
HHMMB	1	0.3304	0.2066	2.5573	0.1098
HHMMB	2	0.0126	0.1799	0.0049	0.9443

**Odds Ratio Estimates**

<b>Effect</b>	<b>90% Wald</b>		
	<b>Point Estimate</b>	<b>Confidence Limits</b>	
GNDR 1 vs 2	0.993	0.831	1.185
BLGETMG -1 vs 1	>999.999	<0.001	>999.999
BLGETMG 0 vs 1	1.543	0.854	2.788
EDULVL 0 vs 2	1.471	1.091	1.983
EDULVL 1 vs 2	0.864	0.658	1.136
EDULVLM -1 vs 2	0.032	0.002	0.543
EDULVLM 0 vs 2	0.783	0.575	1.065
EDULVLM 1 vs 2	0.689	0.511	0.930
EDULVLF -1 vs 2	2.877	0.959	8.631
EDULVLF 0 vs 2	1.254	0.940	1.673
EDULVLF 1 vs 2	1.032	0.799	1.333
EDCTN 0 vs 1	0.125	0.092	0.169
Head_Unemployed 0 vs 1	0.937	0.675	1.301
DSBLD 0 vs 1	1.309	0.896	1.912
RTRD 0 vs 1	0.817	0.648	1.030
Number_Of_Children_S 0 vs 2	0.731	0.296	1.805
Number_Of_Children_S 1 vs 2	1.973	0.901	4.323
HHMMB 1 vs 3	1.961	0.847	4.542
HHMMB 2 vs 3	1.427	0.649	3.139

**Association of Predicted Probabilities and Observed Responses**

Percent Concordant	64.5	Somers' D	0.310
Percent Discordant	33.5	Gamma	0.317
Percent Tied	2.0	Tau-a	0.155
Pairs	180200	c	0.655

**Profile Likelihood Confidence Interval for Parameters**

<b>Parameter</b>		<b>Estimate</b>	<b>90% Confidence Limits</b>	
Intercept		5.0963	0.2834	.
GNDR	1	-0.00364	-0.0923	0.0852
BLGETMG	-1	8.0934	-0.7834	.
BLGETMG	0	-3.8298	.	0.6690
EDULVL	0	0.3059	0.1640	0.4482
EDULVL	1	-0.2258	-0.3499	-0.1024
EDULVLM	-1	-2.4209	-5.8171	-0.8199
EDULVLM	0	0.7676	0.2137	1.9048
EDULVLM	1	0.6407	0.0837	1.7790
EDULVLF	-1	0.7282	-0.0247	1.6383
EDULVLF	0	-0.1022	-0.4351	0.1897
EDULVLF	1	-0.2974	-0.6219	-0.0164
EDCTN	0	-1.0408	-1.1941	-0.8920
Head_Unemployed	0	-0.0323	-0.1971	0.1309
DSBLD	0	0.1344	-0.0559	0.3239
RTRD	0	-0.1011	-0.2171	0.0149
Number_Of_Children_S	0	-0.4353	-0.8565	-0.0156
Number_Of_Children_S	1	0.5575	0.2321	0.8936
HHMMB	1	0.3304	-0.0131	0.6685
HHMMB	2	0.0126	-0.2879	0.3063

Year: 2009.

**Logistic Regression Results (Husstande med partner)  
The LOGISTIC Procedure**

**Model Information**

Data Set SASUSER.FILTER\_FOR\_CON\_DATA\_FORMATT\_0004  
 Response Variable In\_Poverty  
 Number of Response Levels 2  
 Weight Variable Weight\_Household\_2009  
 Model binary logit  
 Optimization Technique Fisher's scoring

Number of Observations Read 500  
 Number of Observations Used 500  
 Sum of Weights Read 399.3754  
 Sum of Weights Used 399.3754

**Response Profile**

Ordered Value	In_Poverty	Frequency	Total Weight
1	0	247	125.52755
2	1	253	273.84783

Probability modeled is In\_Poverty='1'.

**Class Level Information**

Class	Value	Design	Variables	
GNDR	1	1		
	2	-1		
BLGETMG	-1	1	0	
	0	0	1	
	1	-1	-1	
EDULVL	0	1	0	
	1	0	1	
	2	-1	-1	
EDULVLP	-1	1	0	0
	0	0	1	0
	1	0	0	1
	2	-1	-1	-1
EDULVLM	-1	1	0	0
	0	0	1	0



	1	0	0	1
	2	-1	-1	-1
EDULVLF	-1	1	0	0
	0	0	1	0
	1	0	0	1
	2	-1	-1	-1
EDCTN	0	1		
	1	-1		
Head_Unemployed	0	1		
	1	-1		
DSBLD	0	1		
	1	-1		
RTRD	0	1		
	1	-1		
Number_Of_Children_SEC0		1	0	
	1	0	1	
	2	-1	-1	
Spouse_Unemployed	0	1		
	1	-1		
DSBLDP	0	1		
	1	-1		
RTRDP	0	1		
	1	-1		
EDCTNP	0	1		
	1	-1		
HHMMB	1	1	0	
	2	0	1	
	3	-1	-1	

**Model Convergence Status**

Convergence criterion (GCONV=1E-8) satisfied.

**Model Fit Statistics**

Criterion	Intercept and Covariates	
	Intercept Only	Intercept and Covariates
AIC	499.227	519.149
SC	503.441	632.944
-2 Log L	497.227	465.149

**Testing Global Null Hypothesis: BETA=0**

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	32.0774	26	0.1906
Score	31.5260	26	0.2093
Wald	23.4957	26	0.6048

**Type 3 Analysis of Effects**

Effect	DF	Wald	
		Chi-Square	Pr > ChiSq
GNDR	1	3.2908	0.0697
BLGETMG	2	4.3781	0.1120
EDULVL	2	2.5102	0.2851
EDULVLP	3	6.1129	0.1062
EDULVLM	3	5.2042	0.1574
EDULVLF	3	1.9856	0.5754
EDCTN	1	0.0042	0.9484
Head_Unemployed	1	0.2421	0.6227
DSBLD	1	0.7269	0.3939
RTRD	1	0.0609	0.8050
Number_Of_Children_S	2	0.8970	0.6386
Spouse_Unemployed	1	1.4288	0.2320
DSBLDP	1	0.2595	0.6105
RTRDP	1	0.0571	0.8112
EDCTNP	1	1.4266	0.2323
HHMMB	2	0.2240	0.8940

**Analysis of Maximum Likelihood Estimates**

Parameter	DF	Estimate	Standard	Wald	Pr > ChiSq
			Error	Chi-Square	
Intercept	1	-4.0219	295.3	0.0002	0.9891
GNDR	1	-0.2205	0.1216	3.2908	0.0697
BLGETMG	-1	0.8291	0.8355	0.9847	0.3211
BLGETMG	0	-1.0550	0.5693	3.4339	0.0639
EDULVL	0	-0.0640	0.2216	0.0834	0.7727
EDULVL	1	-0.2930	0.1874	2.4445	0.1179
EDULVLP	-1	-0.8552	2.5604	0.1116	0.7384
EDULVLP	0	0.7821	0.8728	0.8030	0.3702
EDULVLP	1	0.3219	0.8686	0.1374	0.7109
EDULVLM	-1	-0.8377	0.9566	0.7670	0.3812
EDULVLM	0	0.3898	0.3776	1.0658	0.3019
EDULVLM	1	0.7689	0.4096	3.5243	0.0605
EDULVLF	-1	0.5917	0.8189	0.5221	0.4700

EDULVLF	0	1	-0.4036	0.3266	1.5271	0.2165
EDULVLF	1	1	-0.3283	0.3345	0.9637	0.3262
EDCTN	0	1	-0.0242	0.3731	0.0042	0.9484
Head_Unemployed	0	1	0.1620	0.3293	0.2421	0.6227
DSBLD	0	1	0.2617	0.3069	0.7269	0.3939
RTRD	0	1	-0.0543	0.2198	0.0609	0.8050
Number_Of_Children_S0	0	1	-0.6832	0.7777	0.7718	0.3797
Number_Of_Children_S1	1	1	0.2816	0.5224	0.2906	0.5899
Spouse_Unemployed	0	1	-0.4125	0.3451	1.4288	0.2320
DSBLDP	0	1	-0.2317	0.4547	0.2595	0.6105
RTRDP	0	1	-0.0547	0.2289	0.0571	0.8112
EDCTNP	0	1	0.4380	0.3667	1.4266	0.2323
-HHMMB	1	1	-8.6673	590.7	0.0002	0.9883
HHMMB	2	1	4.6067	295.3	0.0002	0.9876

### Odds Ratio Estimates

Effect	90% Wald		
	Point Estimate	Confidence	Limits
GNDR 1 vs 2	0.643	0.431	0.960
BLGETMG -1 vs 1	1.828	0.237	14.072
BLGETMG 0 vs 1	0.278	0.096	0.804
EDULVL 0 vs 2	0.656	0.296	1.457
EDULVL 1 vs 2	0.522	0.253	1.075
EDULVLP -1 vs 2	0.545	0.002	154.375
EDULVLP 0 vs 2	2.804	1.361	5.778
EDULVLP 1 vs 2	1.770	0.916	3.419
EDULVLM -1 vs 2	0.596	0.069	5.193
EDULVLM 0 vs 2	2.036	0.913	4.538
EDULVLM 1 vs 2	2.974	1.273	6.948
EDULVLF -1 vs 2	1.571	0.234	10.560
EDULVLF 0 vs 2	0.581	0.266	1.267
EDULVLF 1 vs 2	0.626	0.281	1.394
EDCTN 0 vs 1	0.953	0.279	3.251
Head_Unemployed 0 vs 1	1.383	0.468	4.085
DSBLD 0 vs 1	1.688	0.615	4.632
RTRD 0 vs 1	0.897	0.435	1.849
Number_Of_Children_S 0 vs 2	0.338	0.051	2.261
Number_Of_Children_S 1 vs 2	0.887	0.340	2.312
Spouse_Unemployed 0 vs 1	0.438	0.141	1.364
DSBLDP 0 vs 1	0.629	0.141	2.809
RTRDP 0 vs 1	0.896	0.422	1.903
EDCTNP 0 vs 1	2.401	0.719	8.023

HHMMB 1 vs 3	<0.001	<0.001	>999.999
HHMMB 2 vs 3	1.727	0.259	11.531

### Association of Predicted Probabilities and Observed Responses

Percent Concordant	58.1	Somers' D	0.180
Percent Discordant	40.1	Gamma	0.183
Percent Tied	1.8	Tau-a	0.090
Pairs	62491	c	0.590

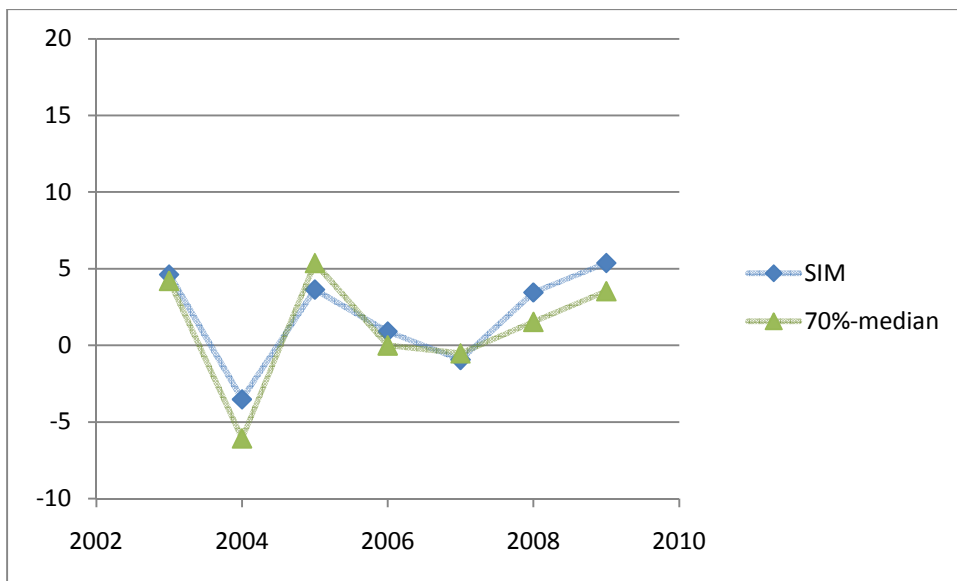
### Profile Likelihood Confidence Interval for Parameters

Parameter		Estimate	90% Confidence Limits	
Intercept		-4.0219	.	1.7710
GNDR	1	-0.2205	-0.4221	-0.0216
BLGETMG	-1	0.8291	-0.3633	2.6197
BLGETMG	0	-1.0550	-2.1189	-0.1691
EDULVL	0	-0.0640	-0.4333	0.2973
EDULVL	1	-0.2930	-0.6072	0.0110
EDULVLP	-1	-0.8552	-10.4814	3.1506
EDULVLP	0	0.7821	-0.5792	3.9987
EDULVLP	1	0.3219	-1.0348	3.5367
EDULVLM	-1	-0.8377	-2.5651	0.7723
EDULVLM	0	0.3898	-0.2348	1.0505
EDULVLM	1	0.7689	0.0973	1.4801
EDULVLF	-1	0.5917	-0.6382	2.1776
EDULVLF	0	-0.4036	-0.9988	0.1044
EDULVLF	1	-0.3283	-0.9346	0.1941
EDCTN	0	-0.0242	-0.6716	0.5731
Head_Unemployed	0	0.1620	-0.3985	0.6969
DSBLD	0	0.2617	-0.2635	0.7588
RTRD	0	-0.0543	-0.4128	0.3135
Number_Of_Children_S	0	-0.6832	-2.0593	0.6275
Number_Of_Children_S	1	0.2816	-0.5756	1.1745
Spouse_Unemployed	0	-0.4125	-1.0216	0.1268
DSBLDP	0	-0.2317	-1.0798	0.4670
RTRDP	0	-0.0547	-0.4294	0.3273
EDCTNP	0	0.4380	-0.1659	1.0574
HHMMB	1	-8.6673	.	6.8719
HHMMB	2	4.6067	-3.1879	.

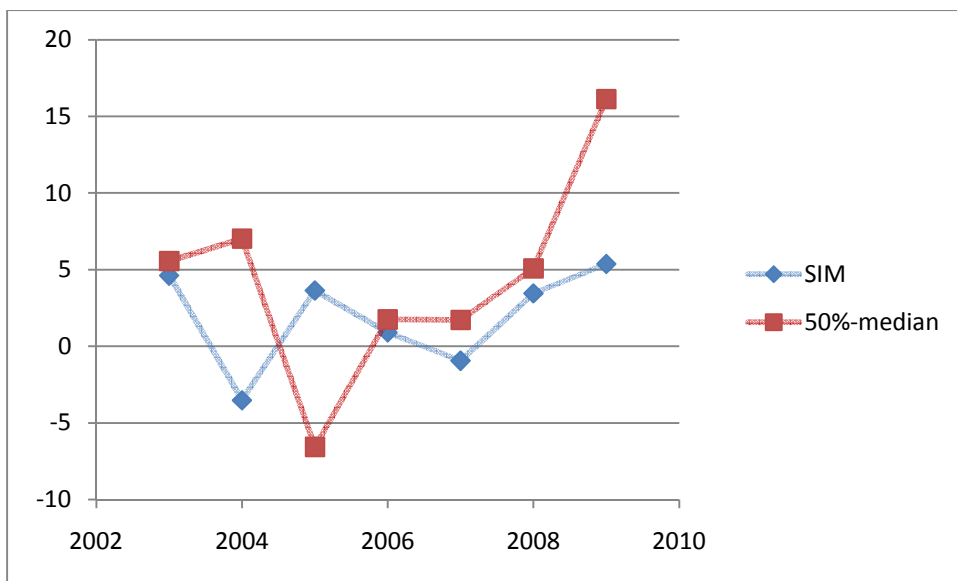
Year: 2009.

**ESS: Resultater**

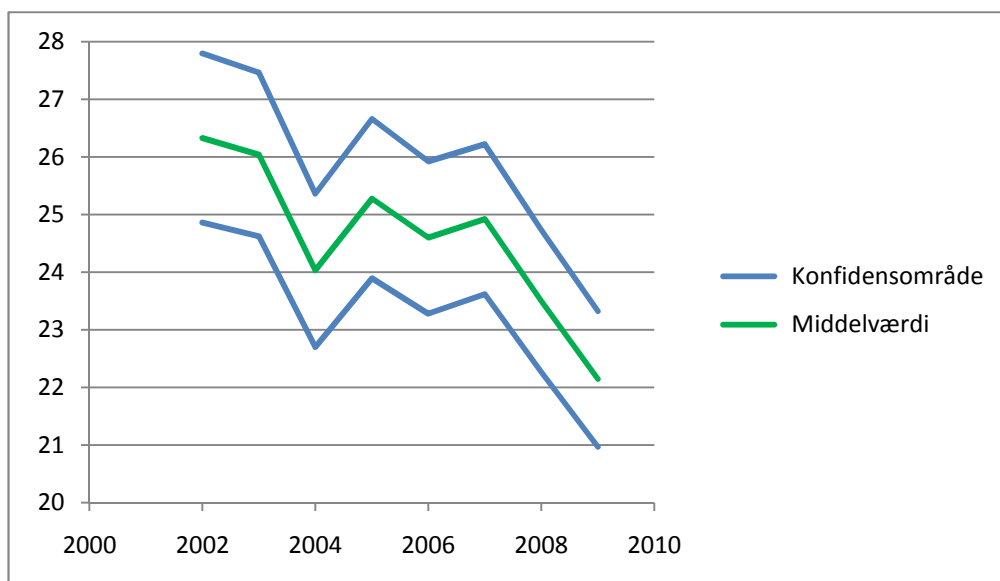
Nedenstående diagram illustrer de procentvise stigninger for SIM og 70%-medianen fra 2003 til 2009.



Nedenstående diagram illustrer de procentvise stigninger for SIM og 50%-medianen fra 2003 til 2009.

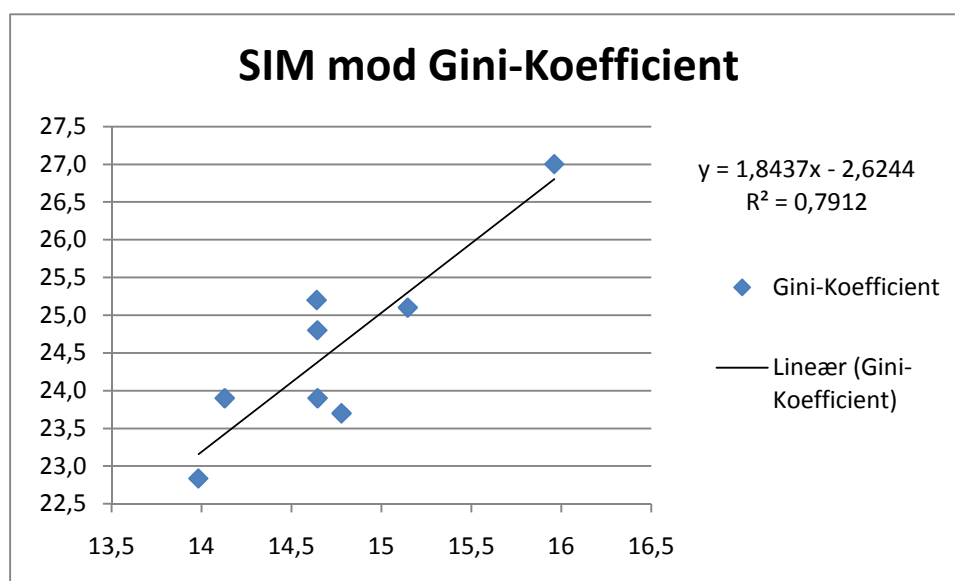


Nedenstående diagram viser procentdelen af fattige med de tilhørende 80% profil-konfidensintervaller for T-fordelingen. Denne skal anvendes til at forstå udviklingen af fattigdom over det sidste årti.

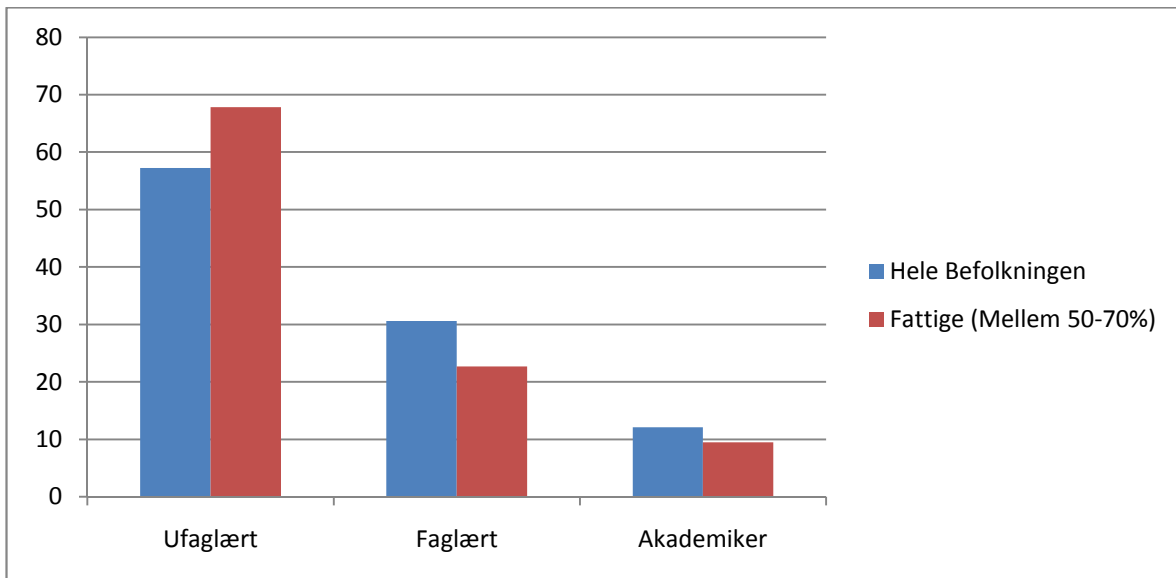


Nedenstående illustrer sammenhængen mellem SIM og Gini-koefficienten:

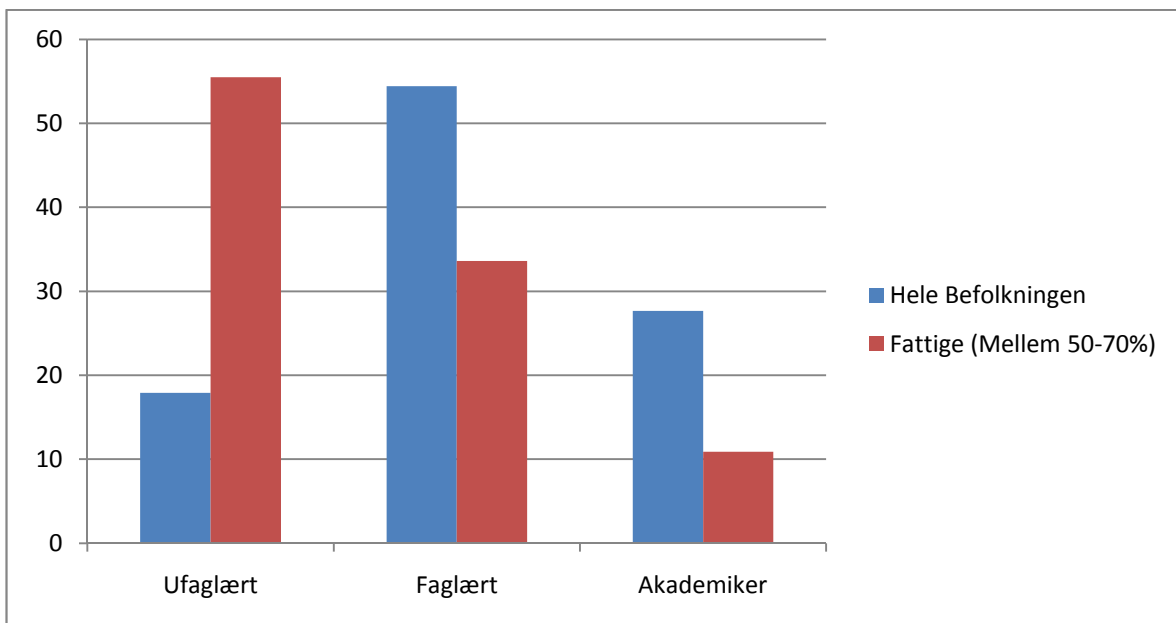
År	Social-mobilitets fattigdomsgrænse	Gini-Koefficient
2009	15,9606	27,0
2008	15,14696	25,1
2007	14,64096	25,2
2006	14,77888	23,7
2005	14,64579	23,9
2004	14,13051	23,9
2003	14,64446	24,8
2002	13,98392	22,8



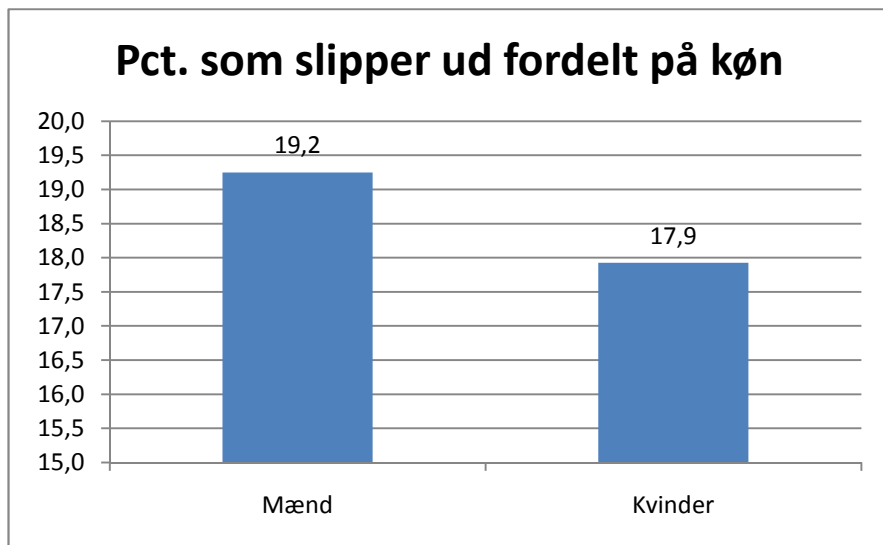
Nedenstående illustrer sammensætningen af befolkningen og de ifølge SIM fattige mellem 50-70% af indkomstmedianen:



Næste graf illustrer det samme, dog ud fra partnerens uddannelse:



Nedenstående viser procentdelen af henholdsvis mænd og kvinder der slipper fri. Resultatet er ikke signifikant.



Nedenstående diagram viser procentdelen af fattige som undslipper fattigdom med de tilhørende 80% profil-konfidensintervaller for T-fordelingen. Denne skal anvendes til at forstå udviklingen af fattigdom over det sidste årti.

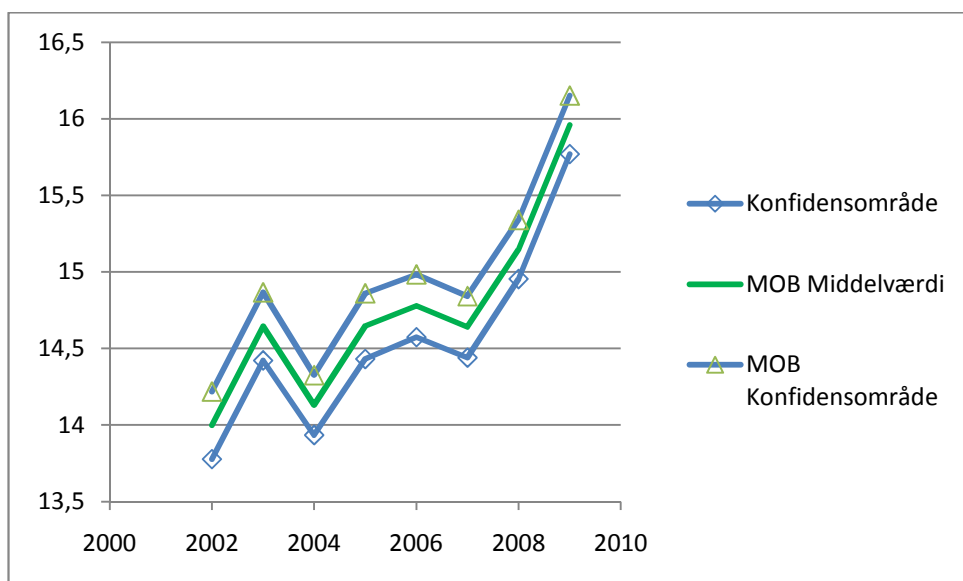




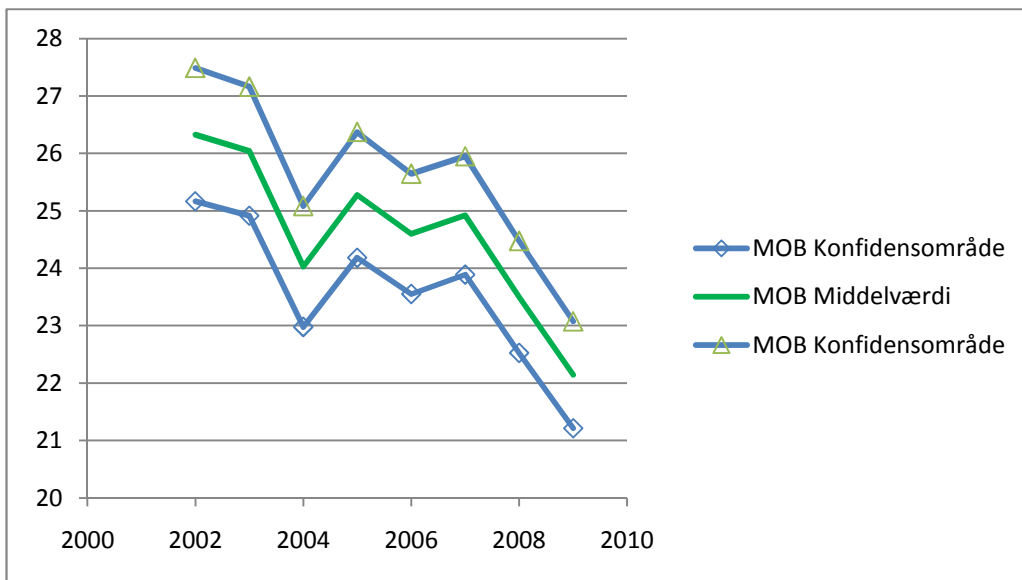
Nedenstående diagram viser procentdelen af fattige med de tilhørende 80% de simultane-konfidensintervaller for T-fordelingen. Denne skal anvendes til at sammenligne årene.



Vi viser yderligere 2 resultater, under normalfordelingsantagelsen, for at understrege robustheden af vores data. Nedenstående diagram viser SIM med tilhørende 80% simultane-konfidensintervaller for normalfordelingen. Denne kan anvendes til at sammenligne enkelte år.

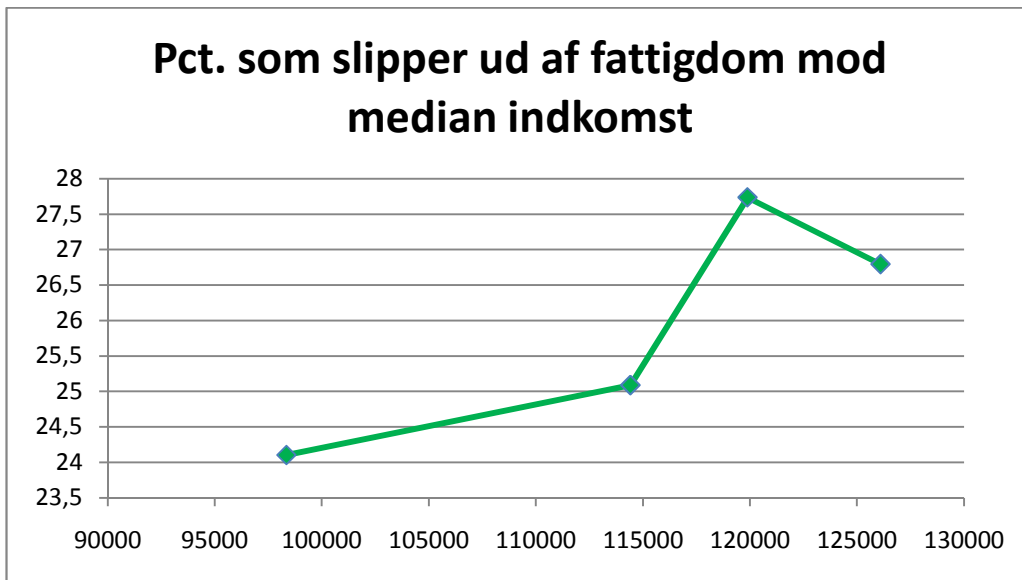


Nedenstående diagram viser procentdelen af fattige med de tilhørende 80% simultane-konfidensintervaller for normal-fordelingen. Denne kan anvendes til at sammenligne enkelte år.

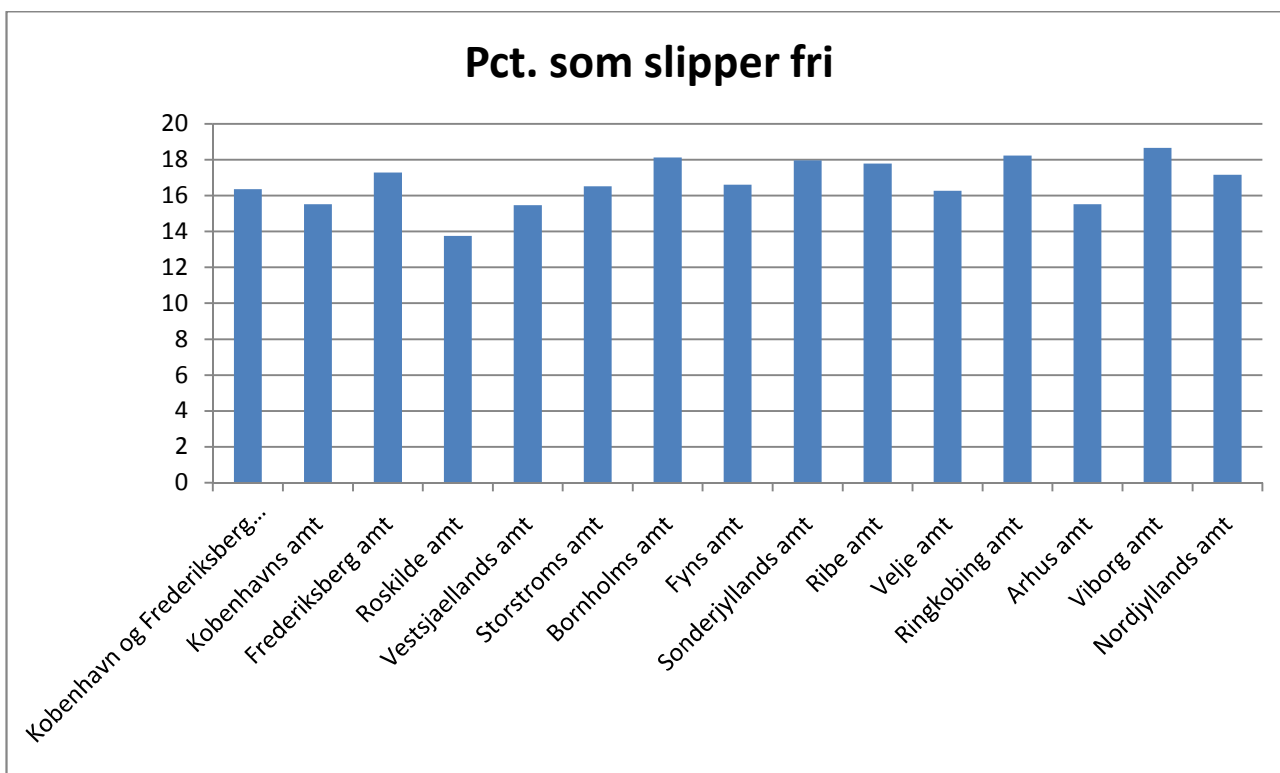


**LIS: Resultater**

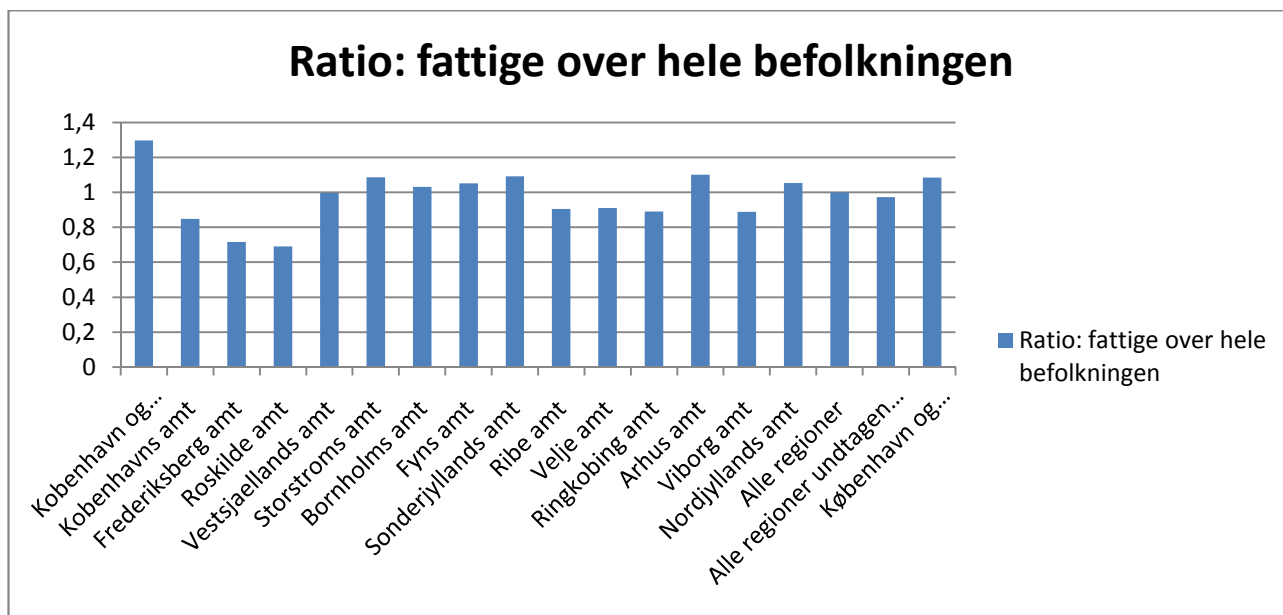
Nedenstående graf viser kvartilerne for alle fattige under 70% medianen, deres chance for at undslippe fattigdom mod deres årlige medianindkomst.



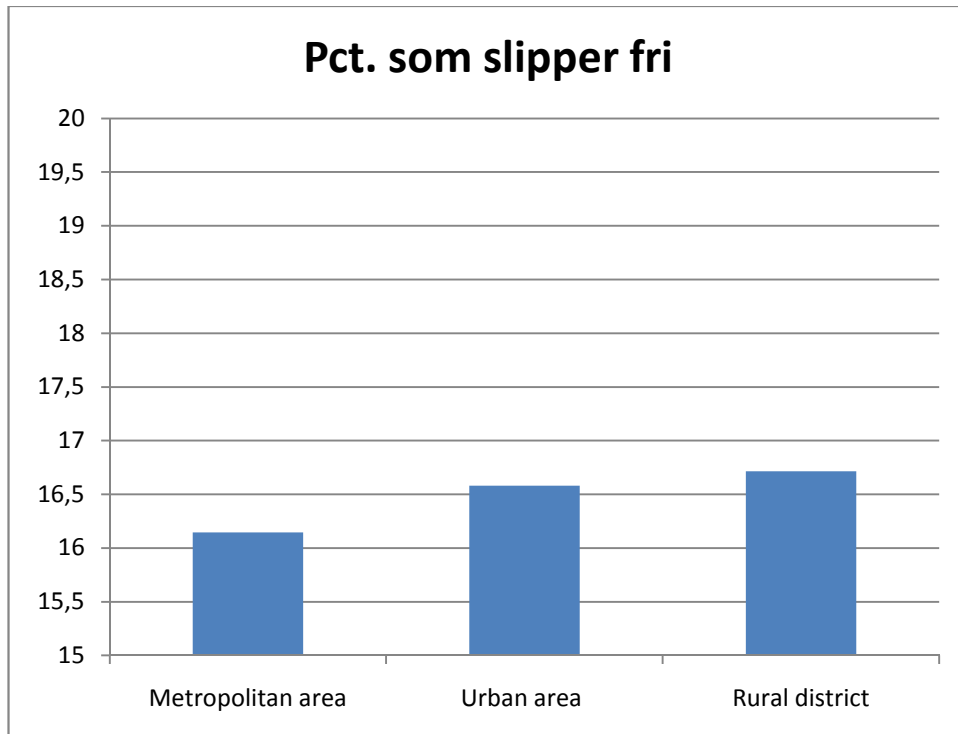
Nedenstående viser procentdelen af mennesker under 70% medianen som slipper fri fordelt på amter.



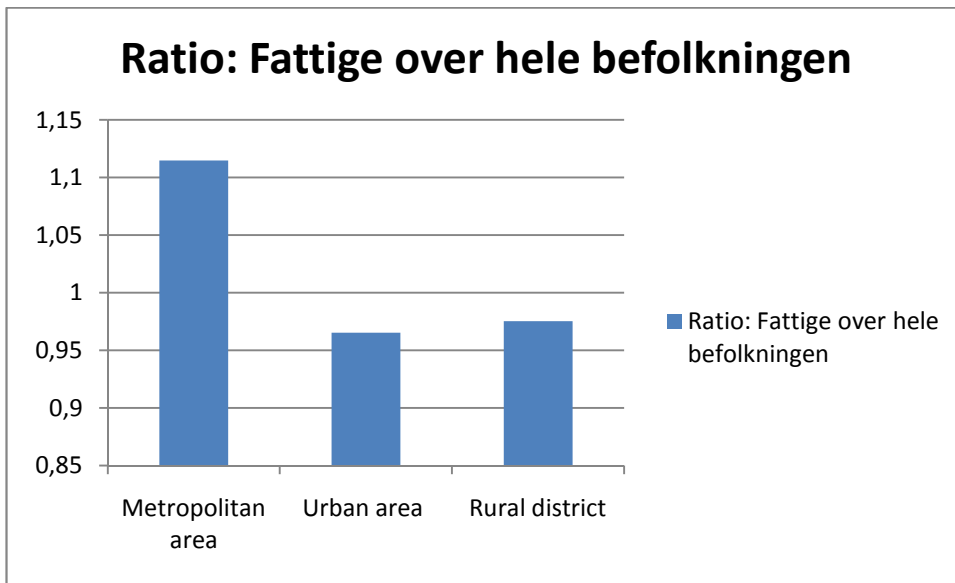
Nedenstående viser ratioen af fattige mennesker fordelt på amter.



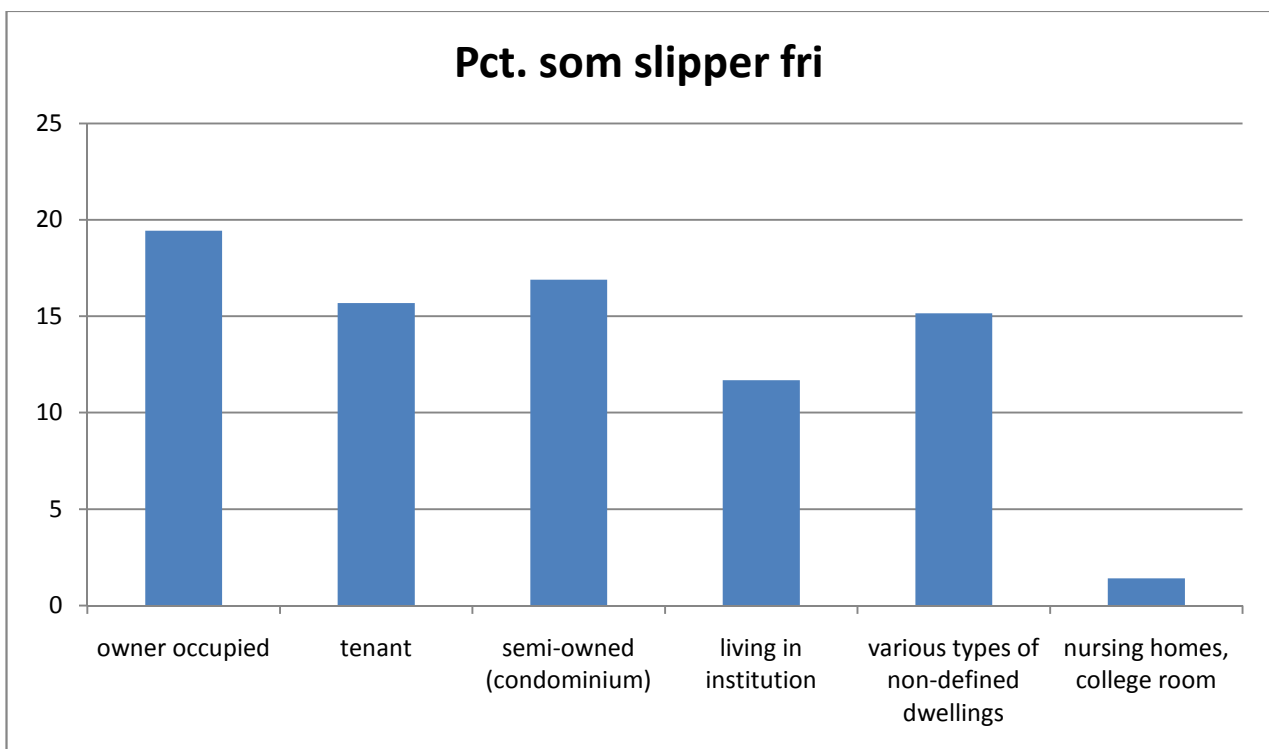
Nedenstående viser procentdelen af mennesker under 70% medianen som slipper fri fordelt på beboelsesområder.



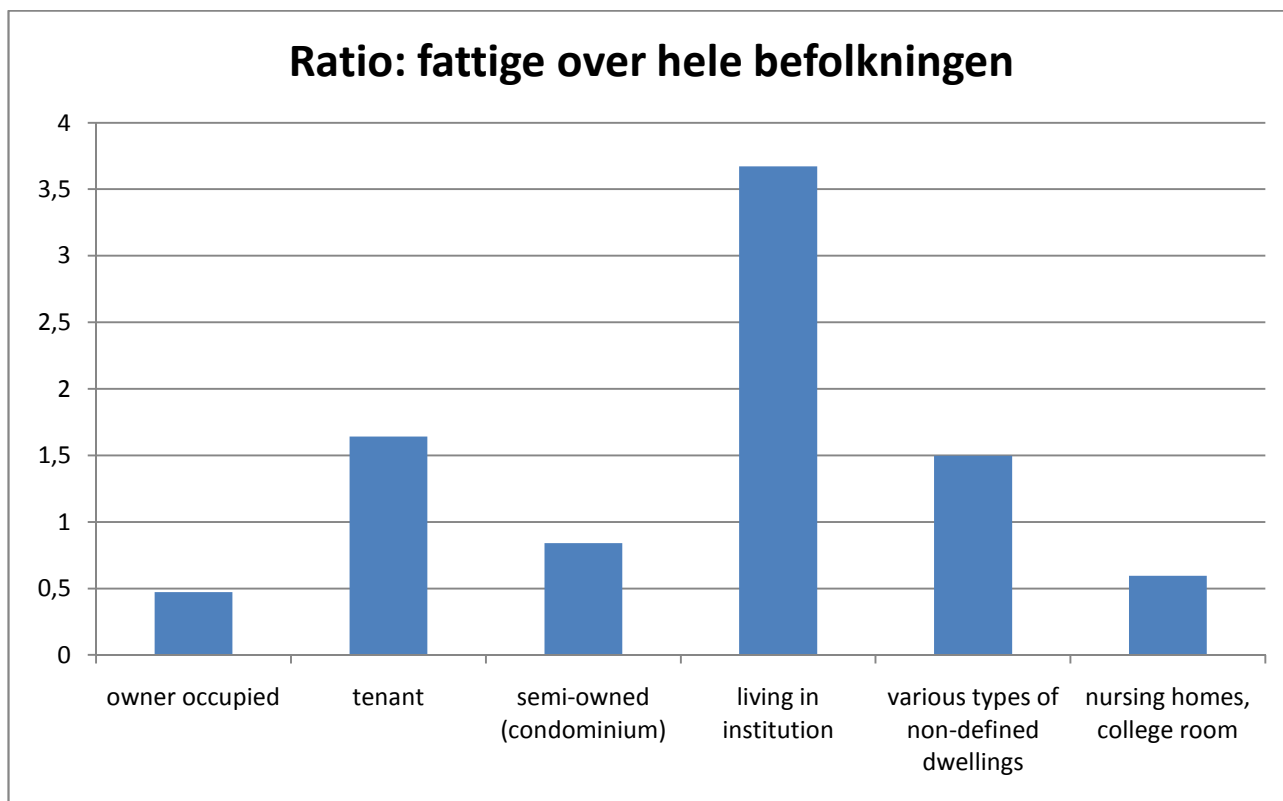
Nedenstående viser ratioen af fattige mennesker fordelt på beboelsesområder.



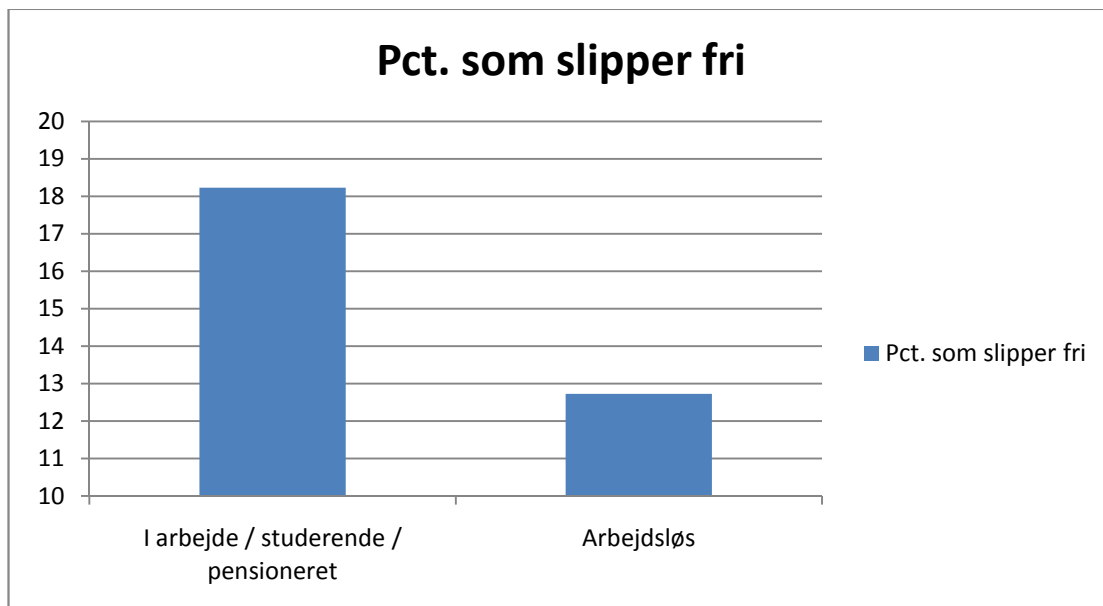
Nedenstående viser procentdelen af mennesker under 70% medianen som slipper fri fordelt på husstandstype.



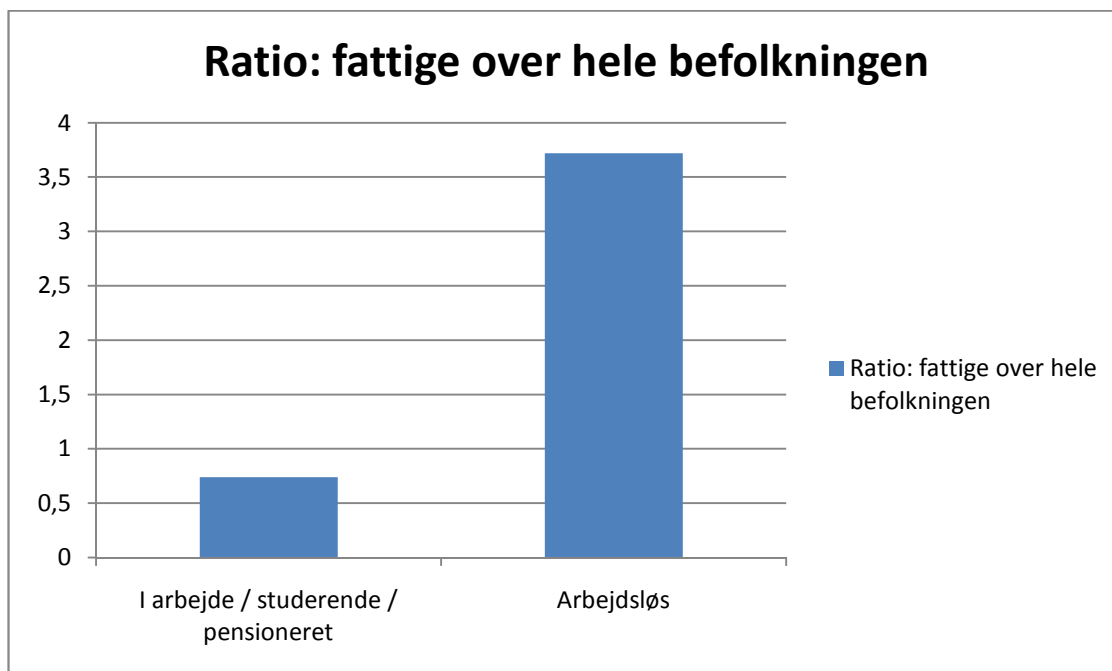
Nedenstående viser ratioen af fattige mennesker fordelt på husstandstype.



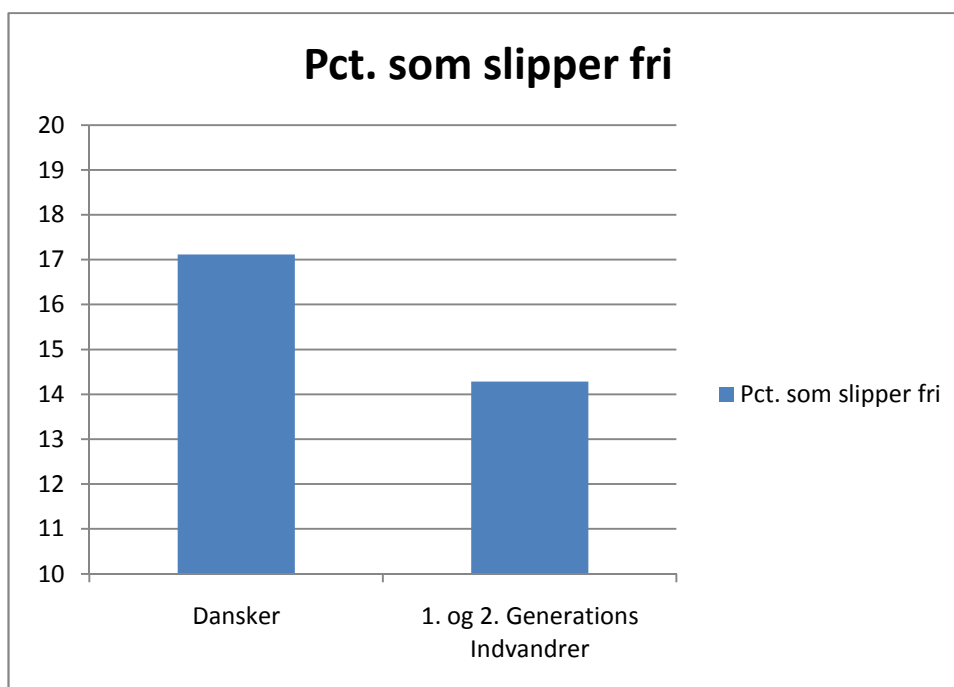
Nedenstående viser procentdelen af mennesker under 70% medianen som slipper fri fordelt på aktivitetstype.



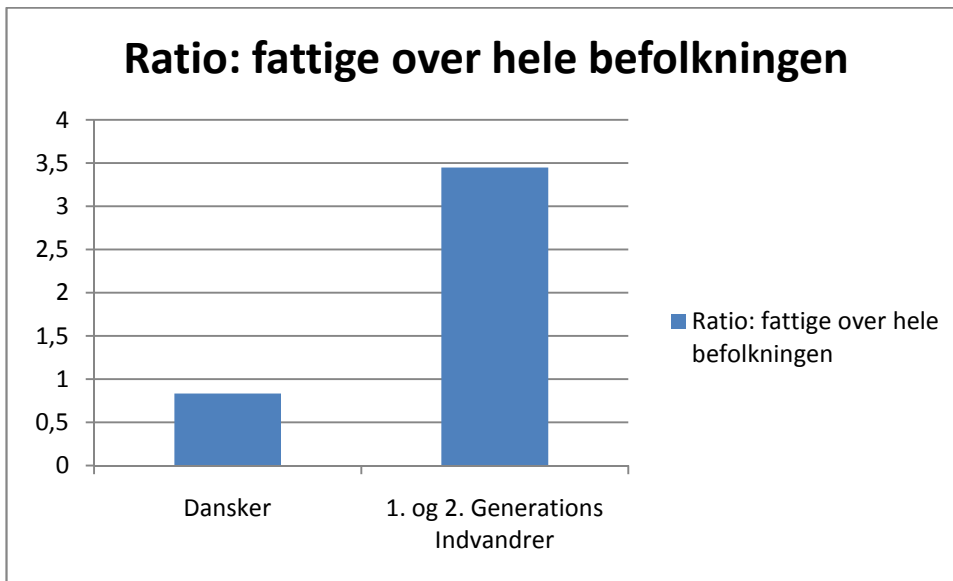
Nedenstående viser ratioen af fattige mennesker ordelt på aktivitetstype.



Nedenstående viser procentdelen af mennesker under 70% medianen som slipper fri fordelt på etnicitet.



Nedenstående viser ratioen af fattige fordelt på etnicitet.





**ESS: Fremskrivning**

Vi starter med at fremskrive SIM, udelukkende ved brug af variabelen år. Nedenstående viser de statistisk resultater for dette.

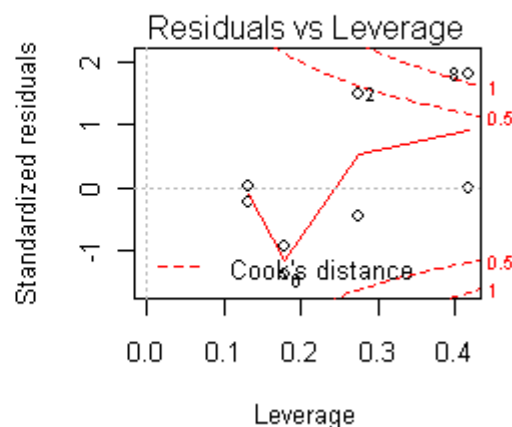
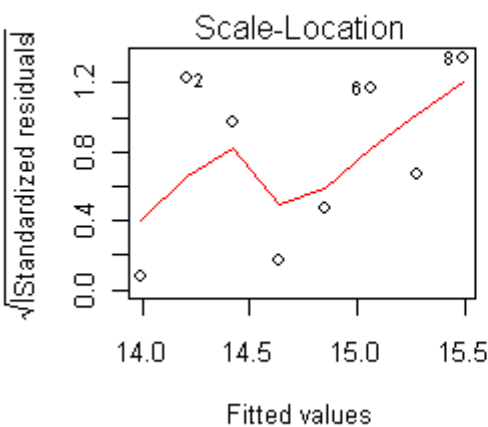
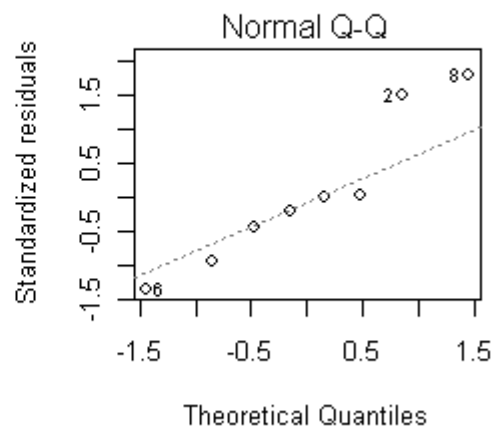
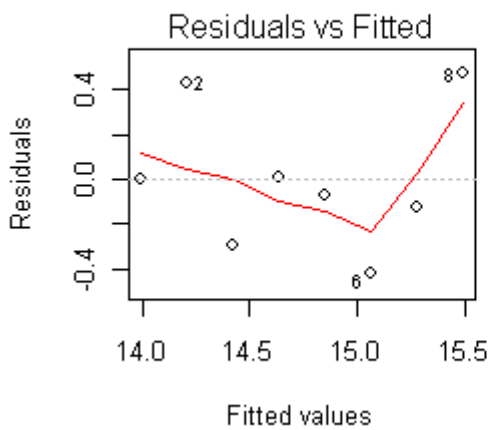
```

Call:
lm(formula = SIM ~ ÅR, data = mydata)

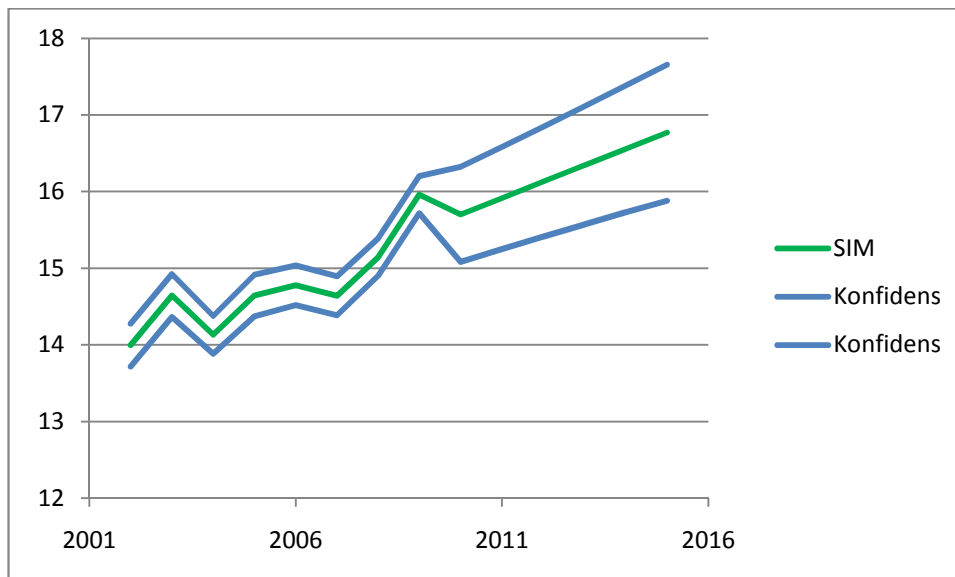
Residuals:
    Min     1Q   Median     3Q     Max
-0.42222 -0.17034 -0.03487  0.11542  0.47090

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -412.94642  105.42214  -3.917  0.00783 **
    ÅR         0.21326   0.05257   4.057  0.00667 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3407 on 6 degrees of freedom
Multiple R-squared: 0.7328,    Adjusted R-squared: 0.6883
F-statistic: 16.46 on 1 and 6 DF, p-value: 0.006674
    
```



	5 %	95 %
(Intercept)	-617.8006475	-208.0921904
ÅR	0.1111123	0.3154045



Herefter fremskrives SIM ud fra både ÅR og procent af BNP som anvendes til indkomstoverførsler. Nedenstående viser de statistisk resultater for dette.

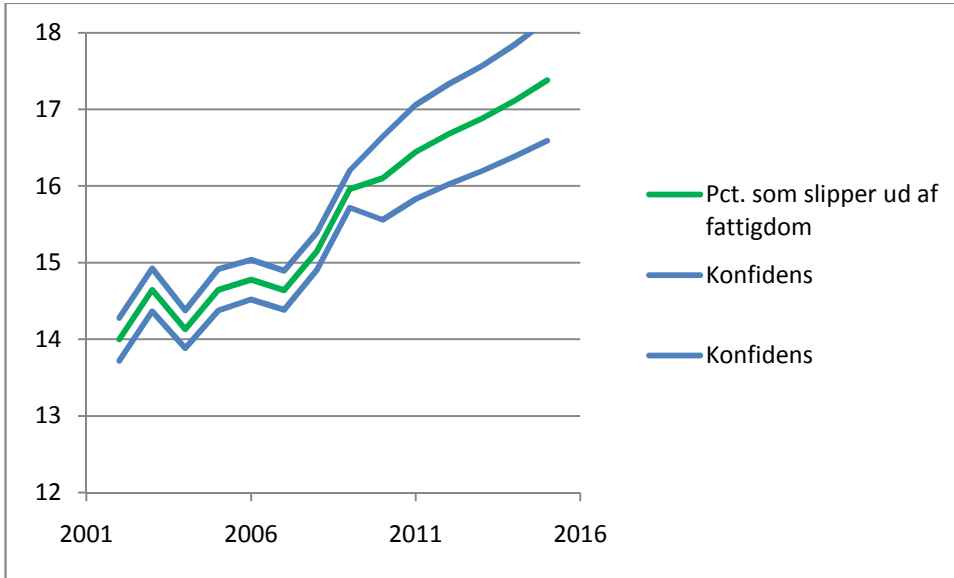
```

Call:
lm(formula = SIM ~ ÅR + TY, data = mydata)

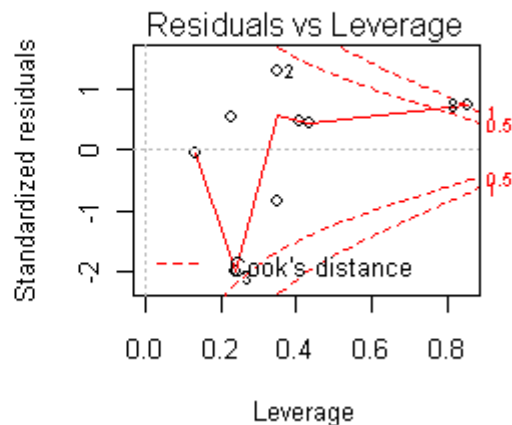
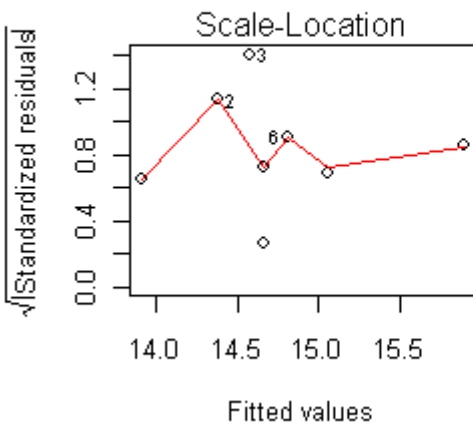
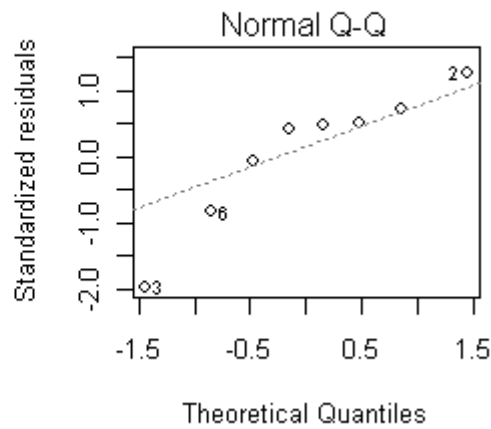
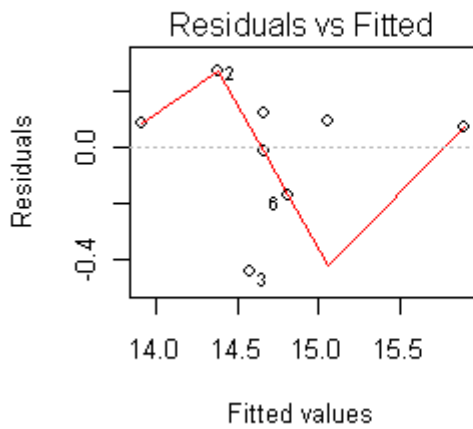
Residuals:
    1     2     3     4     5     6     7     8 
0.08272 0.26568 -0.44249 -0.01612 0.11872 -0.17246 0.09253 0.07141

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -520.31003   91.80207  -5.668 0.00238 **
      ÅR         0.26441    0.04528   5.840 0.00208 **
      TY         0.29249    0.12450   2.349 0.06562 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2573 on 5 degrees of freedom
Multiple R-squared: 0.873,    Adjusted R-squared: 0.8222
F-statistic: 17.19 on 2 and 5 DF, p-value: 0.005746
    
```



	5 %	95 %
(Intercept)	-705.29565047	-335.3244103
ÅR	0.17317057	0.3556474
TY	0.04161561	0.5433606



Herefter fremskrives pct. delen af mennesker som slipper ud af fattigdom i forhold til år. Nedenstående viser de statistisk resultater for dette.

Call:

```
lm(formula = PCT ~ ÅR, data = mydata)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.2879	-0.1566	0.1591	0.2923	1.0284

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	979.0130	241.7318	4.050	0.00673 **
ÅR	-0.4759	0.1205	-3.948	0.00755 **

---

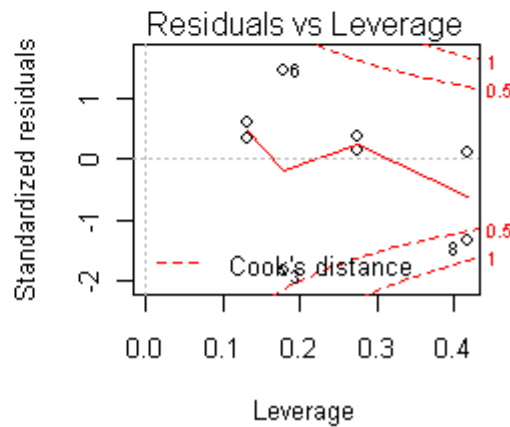
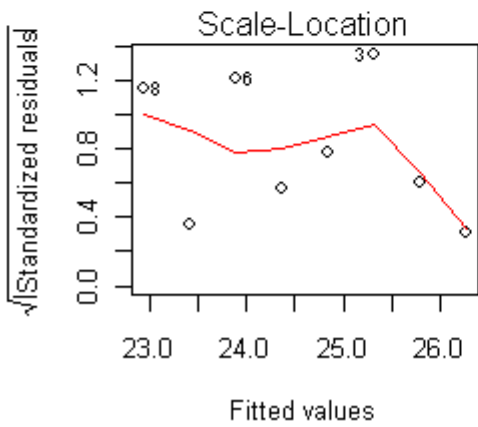
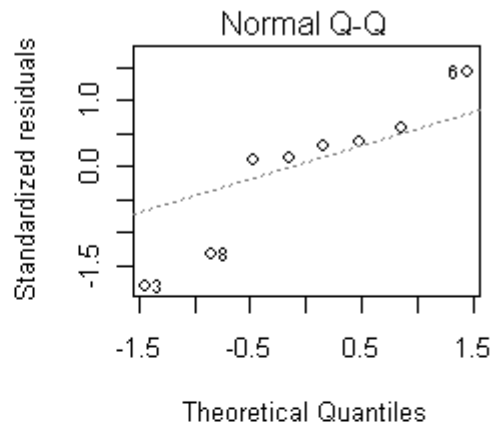
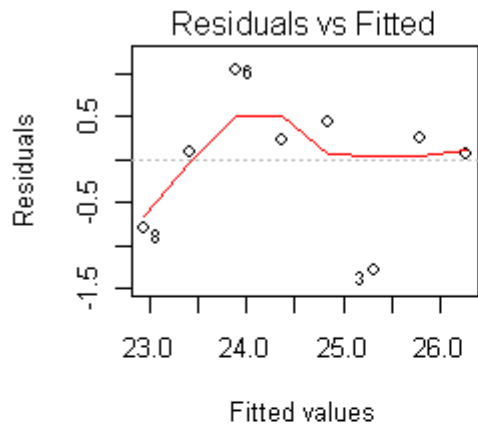
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

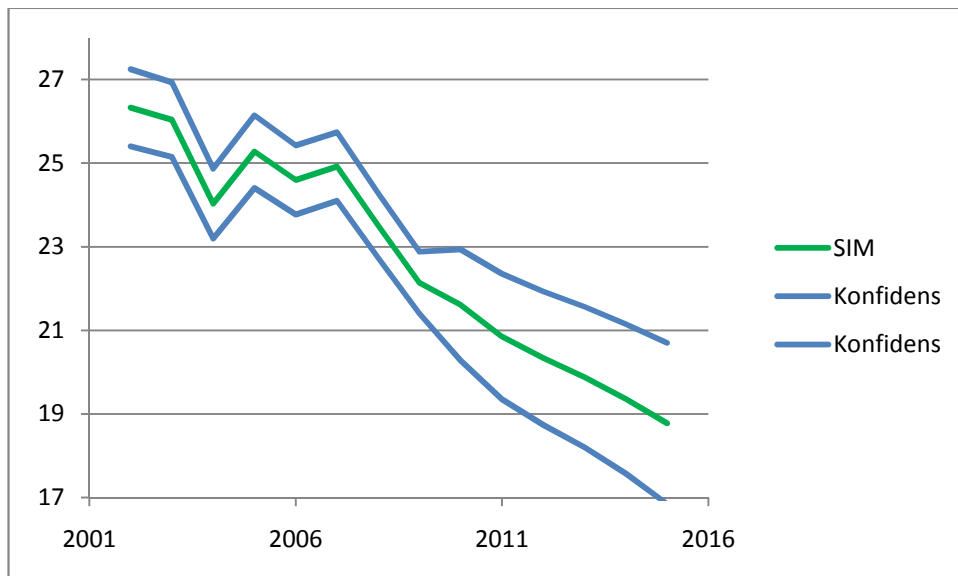
Residual standard error: 0.7812 on 6 degrees of freedom

Multiple R-squared: 0.7221, Adjusted R-squared: 0.6758

F-statistic: 15.59 on 1 and 6 DF, p-value: 0.007553

	5 %	95 %
(Intercept)	509.284573	1448.7415260
ÅR	-0.710116	-0.2416760





Til sidst fremskrives pct.delen af mennesker som slipper ud af fattigdom i forhold til år og procentdel af bruttonationalprodukt brugt på indkomstoverførsler. Nedenstående viser de statistisk resultater for dette.

```

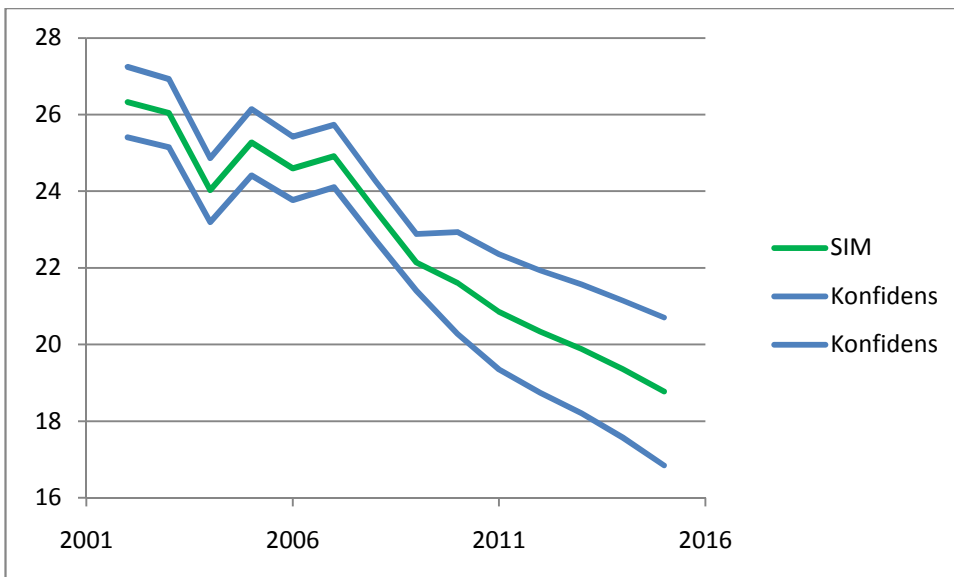
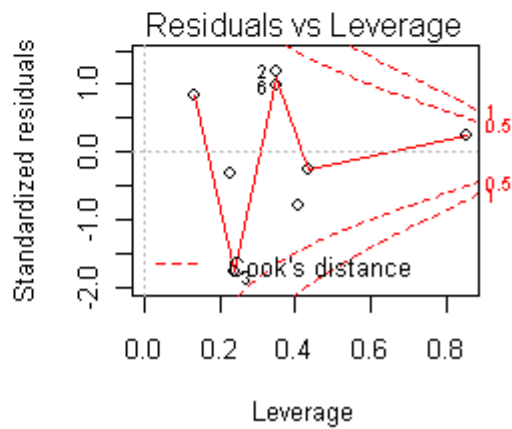
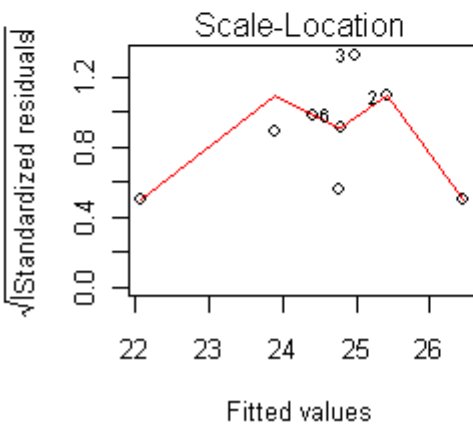
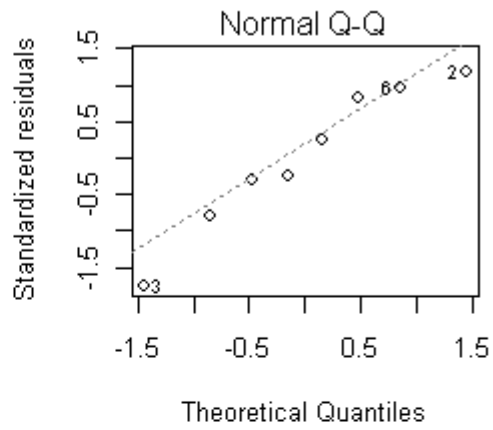
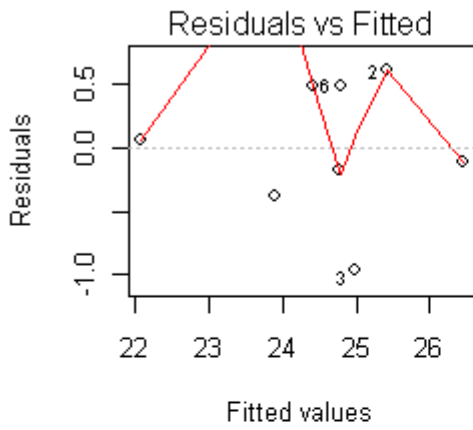
Call:
lm(formula = PCT ~ ÅR + TY, data = mydata)

Residuals:
    1    2    3    4    5    6    7    8
-0.11834  0.60587 -0.96757  0.48912 -0.17448  0.49353 -0.38909  0.06096

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 1208.9343  224.8541  5.377 0.00300 **
ÅR          -0.5854   0.1109 -5.279 0.00325 **
TY          -0.6264   0.3049 -2.054 0.09515 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6302 on 5 degrees of freedom
Multiple R-squared:  0.8493,    Adjusted R-squared:  0.789
F-statistic: 14.09 on 2 and 5 DF,  p-value: 0.008821
    
```

	5 %	95 %
(Intercept)	755.8423604	1662.02629034
ÅR	-0.8089096	-0.36196245
TY	-1.2408401	-0.01189792



## Deskriptiv statistik

### ESS Data

Data anvendt er fra runderne 1, 2, 3 og 4 der svarer til henholdsvis årene 2002, 2004, 2006 og 2008.

#### Behandling af data for ESS 2008:

For at vurdere hvorvidt en husstand lever under 70% medianen eller ej tages udgangspunkt i dokumentation for husstandens indkomst, der er kodet som variablen der tager nedenstående værdier:

#### **ESS R4 Household net income 2008**

ESS R4	Strata	incom min corrected to 2008	incom max corrected to 2008	incom min corrected to 2008 rounded	incom max corrected to 2008 rounded
J	1	-233 889 185	109 966	0	109 999
R	2	109 966	147 886	110 000	147 999
C	3	147 887	186 306	148 000	185 999
M	4	186 307	226 272	186 000	225 999
F	5	226 273	277 349	226 000	276 999
S	6	277 350	337 747	277 000	337 999
K	7	337 748	393 429	338 000	392 999
P	8	393 430	450 200	393 000	449 999
D	9	450 202	533 573	450 000	533 999
H	10	533 574	426 161 831	534 000	eller derover
refuse	77				
dont now	88				

Grænse for at blive klassificeret som værende under 70% af medianen til året 2008.

Maks. Disponible Indkomstniveau	Min. antal mennesker	Min. antal børn under 14
J	1	0
R	1	1
R	2	0
C	1	4
C	2	2
C	3	0
M	1	5
M	2	3
M	3	1
M	4	0



Grænse for at blive taget med i betragtning til den logistiske regression (mht. 70%-medianen):

Maks. Disponible Indkomstniveau	Min. antal mennesker	Min. antal børn under 14
R	1	0
C	1	1
C	2	0
M	1	4
M	2	2
M	3	0
F	1	5
F	2	3
F	3	1
F	4	0

Grænse for at blive klassificeret som værende under 60% af medianen til året 2008.

Maks. Disponible Indkomstniveau	Min. antal mennesker	Min. antal børn under 14
J	1	0
R	1	1
R	2	0
C	1	2
C	2	2
C	3	0
M	1	5
M	2	2
M	3	1
M	4	0

Grænse for at blive taget med i betragtning til den logistiske regression (mht. 60%-medianen):

Maks. Disponible Indkomstniveau	Min. antal mennesker	Min. antal børn under 14
R	1	0
C	1	1
C	2	0
M	1	2
M	2	2
M	3	0
F	1	5
F	2	2
F	3	1
F	4	0

Behandling af data for ESS 2006, ESS 2004 og ESS 2002:

For at vurdere hvorvidt en husstand lever under 70% medianen eller ej tages udgangspunkt i dokumentation for husstandens indkomst, der er kodet som variablen der tager nedenstående værdier:

<b>Din husstands samlede nettoindkomst</b>			
	<b>Ca pr. uge</b>	<b>Ca. pr. måned/år</b>	<b>Ca. pr. år</b>
<b>J</b>	Mindre end 13.300 kr. ....	Mindre end 13.300 kr. ....	Mindre end 13.300 kr. ....
<b>R</b>	Mellem 13.300 og 26.600 kr. ....	Mellem 13.300 og 26.600 kr. ....	Mellem 13.300 og 26.600 kr. ....
<b>C</b>	Mellem 26.600 og 44.400 kr. ....	Mellem 26.600 og 44.400 kr. ....	Mellem 26.600 og 44.400 kr. ....
<b>M</b>	Mellem 44.400 og 88.800 kr. ....	Mellem 44.400 og 88.800 kr. ....	Mellem 44.400 og 88.800 kr. ....
<b>F</b>	Mellem 88.800 og 133.200 kr. ....	Mellem 88.800 og 133.200 kr. ....	Mellem 88.800 og 133.200 kr. ....
<b>S</b>	Mellem 133.200 og 177.600 kr. ....	Mellem 133.200 og 177.600 kr. ....	Mellem 133.200 og 177.600 kr. ....
<b>K</b>	Mellem 177.600 og 222.000 kr. ....	Mellem 177.600 og 222.000 kr. ....	Mellem 177.600 og 222.000 kr. ....
<b>P</b>	Mellem 222.000 og 266.400 kr. ....	Mellem 222.000 og 266.400 kr. ....	Mellem 222.000 og 266.400 kr. ....
<b>D</b>	Mellem 266.400 og 444.000 kr. ....	Mellem 266.400 og 444.000 kr. ....	Mellem 266.400 og 444.000 kr. ....
<b>H</b>	Mellem 444.000 og 666.000 kr. ....	Mellem 444.000 og 666.000 kr. ....	Mellem 444.000 og 666.000 kr. ....
<b>U</b>	Mellem 666.000 og 888.000 kr. ....	Mellem 666.000 og 888.000 kr. ....	Mellem 666.000 og 888.000 kr. ....
<b>N</b>	Over 888.000 kr. ....	Over 888.000 kr. ....	Over 888.000 kr. ....

Grænse for at blive klassificeret som værende under 70% af medianen:

Maks. Disponible Indkomstniveau	Min. antal mennesker	Min. antal børn under 14
J	1	0
R	1	0
C	1	0
M	1	0
F	1	1
F	2	0
S	1	4
S	2	2
S	3	0
K	1	5
K	2	3
K	3	1
K	4	0

Grænse for at blive taget med i betragtning til den logistiske regression (mht. 70%-medianen):

Maks. Disponible Indkomstniveau	Min. antal mennesker	Min. antal børn under 14
J	1	0
R	1	0
C	1	0
M	1	0
F	1	0
S	1	1
S	2	0
K	1	4
K	2	2
K	3	0
P	1	5
P	2	3
P	3	1
P	4	0

Grænse for at blive klassificeret som værende under 60% af medianen:

Maks. Disponible Indkomstniveau	Min. antal mennesker	Min. antal børn under 14
J	1	0
R	1	0
C	1	0
M	1	0
F	1	1
F	2	0
S	1	2
S	2	1
S	3	0
K	1	4
K	2	2
K	3	1
K	4	0

## Deskriptiv statistik af sammenlagte ESS Data efter behandling:

Husstandstype	Under 70%-Medianen				
	2002	2004	2006	2008	Alle år
En voksen under 65	32	28	43	31	134
En voksen over 65	36	27	27	20	110
En voksen med et eller flere børn	10	9	7	6	32
To voksne under 65	19	20	27	29	95
To voksne, hvoraf mindst den ene er over 65	47	41	30	57	175
To voksne med et barn	0	3	5	10	18
To voksne med to børn	3	7	3	6	19
To voksne med tre eller flere børn	3	8	4	8	23
Tre eller flere voksne	6	9	15	21	51
Tre eller flere voksne med børn	5	3	8	9	25
Alle husstande	161	155	169	197	682

Husstandstype	Anvendt til logistisk regression, mht. 70% medianen				
	2008	2006	2004	2002	Alle år
En voksen under 65	68	54	83	73	278
En voksen over 65	66	78	65	50	259
En voksen med et eller flere børn	17	17	13	13	60
To voksne under 65	33	30	48	48	159
To voksne, hvoraf mindst den ene er over 65	82	81	79	107	349
To voksne med et barn	3	7	6	14	30
To voksne med to børn	9	10	7	14	40
To voksne med tre eller flere børn	5	10	8	9	32
Tre eller flere voksne	13	14	28	37	92
Tre eller flere voksne med børn	11	10	20	20	61
Alle husstande	307	311	357	385	1360

Husstandstype	Anvendt til logistisk regression (mht. 60% medianen)				
	2008	2006	2004	2002	Alle år
En voksen under 65	32	54	83	73	169
En voksen over 65	36	78	65	50	179
En voksen med et eller flere børn	10	17	13	13	40
To voksne under 65	21	30	48	48	99
To voksne, hvoraf mindst den ene er over 65	41	81	79	107	201
To voksne med et barn	3	7	6	14	16
To voksne med to børn	9	10	7	14	26
To voksne med tre eller flere børn	5	10	8	9	23
Tre eller flere voksne	13	14	28	37	55
Tre eller flere voksne med børn	10	10	20	20	40
Alle husstande	180	311	357	385	848

Husstandstype	Under 60%-Medianen				
	2008	2006	2004	2002	Alle år
En voksen under 65	32	28	43	31	134
En voksen over 65	36	27	27	20	110
En voksen med et eller flere børn	14	14	9	10	47
To voksne under 65	19	20	27	29	95
To voksne, hvoraf mindst den ene er over 65	47	41	30	57	175
To voksne med et barn	0	7	6	14	27
To voksne med to børn	5	8	4	8	25
To voksne med tre eller flere børn	3	8	4	8	23
Tre eller flere voksne	6	9	15	21	51
Tre eller flere voksne med børn	5	3	8	9	25
Alle husstande	167	165	173	207	712

Husstandstype	Samples i alt				
	2002	2004	2006	2008	Alle år
En voksen under 65	160	176	190	168	694
En voksen over 65	124	121	114	102	461
En voksen med et eller flere børn	49	31	27	19	126
To voksne under 65	389	298	318	309	1314
To voksne, hvoraf mindst den ene er over 65	267	291	264	303	1125
To voksne med et barn	135	109	100	102	446
To voksne med to børn	169	155	111	90	525
To voksne med tre eller flere børn	67	56	44	35	202
Tre eller flere voksne	135	173	192	255	755
Tre eller flere voksne med børn	115	95	127	123	460
Alle husstande	1610	1505	1487	1506	6108

Grænse for at blive taget med i betragtning til den logistiske regression (mht. 60%-medianen):

Maks. Disponible Indkomstniveau	Min. antal mennesker	Min. antal børn under 14
J	1	0
R	1	0
C	1	0
M	1	0
F	1	0
S	1	1
S	2	0
K	1	2
K	2	1
K	3	0
P	1	4
P	2	2
P	3	1
P	4	0

**ESS Uddannelsesniveauer**

Side 52, POPULATION STATISTICS, ESS4-2008 ed. 3.0 (ESS4AppendixA1\_e03.0.pdf)

**Highest obtained education****Approx.****equivalence****with EDULVL****Education**

2	10 GRUNDSKOLE
3	20 ALMENGYMNASIAL UDDANNELSER
3	25 ERHVERVSGYMNASIAL UDDANNELSER
4	35 ERHVERVSUDDANNELSER
5	40 KORTE VIDEREGÅENDE UDDANNELSER
6	50 MELLEMLANGE VIDEREGÅENDE UDDANNELSER
6	60 BACHELOR
7	65 LANGE VIDEREGÅENDE UDDANNELSER
7	70 FORSKERUDDANNELSER
	90 Don't know

## **SSM - Udeladte undermodeller**

Nedenstående er analyser som er udeladt fra den originale rapport, da deres viden ikke er essentielt for ovenstående metoder og konklusioner. I stedet er de medtaget som appendiks for at underbygge forståelsen af hvordan de sociale processer for dødelighed og separation/parforhold fungerer for de lavt uddannede og økonomisk dårligst stillede danskere fra et statistisk synspunkt.



## Pardannelse og separation

### Motivation

Man kan spørge sig selv hvorfor pardannelse og separation er relevant for ulighed. Sociologiske teorier fortæller os at ægteskab og længere parforhold er et resultat af værdier hos de deltagende parter og deres bevidste handlingerne for at gå på kompromis og skabe en fremtid sammen. Sociologiske teorier fortæller os også ligeledes at status, indkomst, uddannelse og en lang række andre egenskaber kan være alt afgørende for denne proces. Smits et al og Hamplová bekræfter grundigt at der er stærk uddannelses homogami i Danmark (det vil sige at langt de fleste mennesker indgår i parforhold, hvor deres partner har samme uddannelsesniveau). Smits et al kommer blandt andet frem til at uddannelsesniveaut er langt mere afgørende end indkomsten for partner valg. Som konsekvens af dette vil individer med lavt uddannelsesniveau finde sammen med andre individer som også kun har et lavt uddannelsesniveau. Det forstærker naturligvis uligheden i samfundet. Forstærkelsen gælder både i forhold til antallet af børn parforholdet får, men også i forhold til indkomst og muligheden for at finde job i tilfælde af, at en part skulle blive arbejdsløs. Det er derfor afgørende at undersøge denne variabel så udstrakt vi kan.

### Faktorer og model

Baseret på ESS data fra 2002 udvikler Hamplová en solid statistisk model til at bestemme frekvenserne for parforhold der afspejler begge parter uddannelsesniveau. Vi vil benytte disse frekvenser til at udregne sandsynligheden for hvilket parforhold et individ ender i. Nedenstående viser frekvenstabellerne for de Skandinaviske lande:

Table 7: Estimated cell frequencies from Model 17

Man	Woman							
	Marriage				Cohabitation			
	primary	lower secondary	upper secondary	Higher	primary	lower secondary	upper secondary	higher
	<b>Scandinavia</b>							
Primary	46.00	31.61	16.29	11.11	24.00	24.76	22.24	14.99
Lower secondary	22.30	69.10	63.98	40.63	25.33	73.90	60.86	43.90
Upper secondary	11.75	65.43	177.24	97.58	19.72	52.74	203.76	96.78
Higher	5.95	30.87	72.50	237.68	8.95	25.60	65.13	237.32

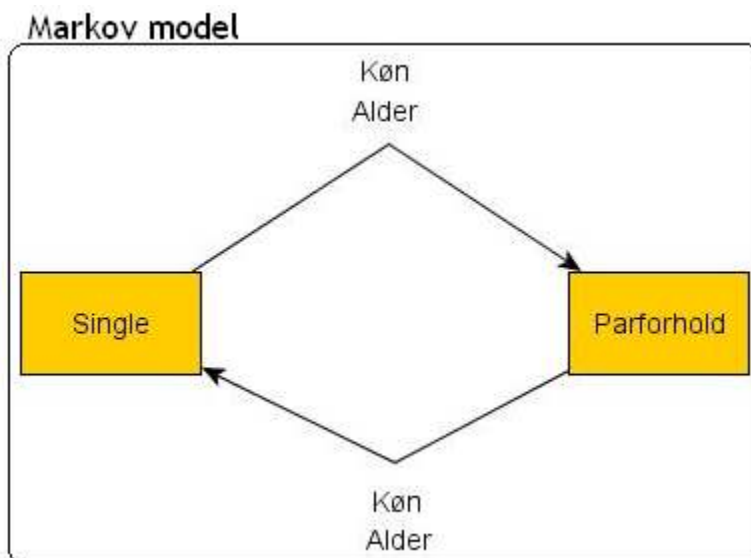
Vi antager at de sociale processer mellem mænd og kvinder, når det kommer til pardannelse og separation, er ens for alle de skandinaviske lander, og tillader derfor os selv at bruge modellens resultater. Det svarer i grove træk til at postulere at parforholds-"mentaliteten" er ens i alle de Skandinaviske lande. Det er klart at den aldrig helt vil være ens, men overordnet mener vi at dette er en mild antagelse. Man skal ikke forveksle denne antagelse om "ens mentalitet" med f.eks. løndiskrimination eller valg af uddannelses... Det er klart at både løndiskrimination og valg af uddannelse på baggrund af køn er forskellige i Skandinavien. Men vores antagelse holder ikke kønnene op mod hinanden, men går i stedet på det sammenspil der er mellem mand og kvinde.

Vi antager endvidere at der ikke kan dannes parforhold mellem samme køn. Denne antagelse er meget mild, men nødvendig, da tilstrækkelig data ikke forefindes for andre parforhold.

Vi antager endvidere at uddannelses homogamien ikke ændrer sig over tid. Smits et al bekræfter at uddannelses homogamien ændrer sig relativt lidt over både 70'erne og 80'erne, og faktisk at den eneste ændring de kan observere er forstærkelse af den. Visse denne tendens forsætter er vores antagelse til fordel for de laveste sociale klasser, da parforhold mellem individer i lave indkomstgrupper og højere indkomstgrupper skaber mindre fattigdom.

Pardannelse og separation analyse

Givet vores tidligere antagelser og diskussioner kan vi udregne sandsynligheder for pardannelse og separation som en stationær kontinuert markov proces. Diagrammet til højre illustrerer modellen og de faktorer som indgår. Se appendiks for ligninger til denne model. Bemærk: da alt data anvendt har været store samples fra survey data af høj kvalitet, må vi antage at det ikke kun er repræsentativt men også er præcist for hele den danske befolkning. Derudover har vi givet vores tidligere antagelser at samtlige resultater må være fuldstændige præcise. Således kan vi ikke udregne konfidens-intervaller eller på anden måde måle usikkerheden på disse data. For reelt at vurdere usikkerheden på disse data, skal man enten ind og kigge på usikkerheden af vores eller kigge på indsamlingen af data.



De statistiske resultater er givet i nedenstående tabeller:

Mænd	Er i ægteskab efter 5-årig periode, givet at vedkommende ikke var i ægteskab før.	Er ikke i ægteskab efter 5-årig periode, givet at vedkommende var det før.
20-24	0,023888993	0,045325814
25-29	0,105012409	0,043392166
30-34	0,134475389	0,042675214
35-39	0,072822155	0,045051213
40-44	0,03423445	0,031792718
45-49	0,019405333	0,023882661
50-54	0,013081387	0,014228183
55-59	0,006473214	0,01427589
60+	0,001549702	0,014311365

Kvinder	Er i ægteskab efter 5-årig periode, givet at vedkommende ikke var i ægteskab før.	Er ikke i ægteskab efter 5-årig periode, givet at vedkommende var det før.
20-24	0,053113288	0,044635757
25-29	0,143537734	0,042453009
30-34	0,118050996	0,043075901
35-39	0,052734292	0,04553941
40-44	0,024763919	0,031949178
45-49	0,01466451	0,023940632
50-54	0,009097246	0,014256959
55-59	0,003593564	0,014296646
60+	0,000601679	0,01431819

## Kilder:

More or Less Educational Homogamy? A Test of Different Versions of Modernization Theory Using Cross-Temporal Evidence for 60 Countries  
 Author(s): Jeroen Smits, Wout Ultee, Jan Lammers  
 Source:

Educational Homogamy in 65 Countries: An Explanation of Differences in Openness Using Country-Level Explanatory Variables  
 Author(s): Jeroen Smits, Wout Ultee, Jan Lammers  
 Source:

Educational Homogamy in Marriage and Cohabitation  
 in Selected European Countries **Dana Hamplová**

## Dødsfald

### Motivation

Man kan spørge sig selv hvorfor vi overhovedet gider bøvle med dødsfald. Svaret er, at dødeligheden er væsentlig højere for lave indkomstgrupper. I diskussion om social mobilitet er det derfor vigtigt at medtage dødelighed. Derudover arbejder vi med husstande i vores ESS data, hvoraf flere medlemmer som ikke er i arbejde sagtens være i en ældre aldersgruppe og således have stor dødelighed. Samtidig måler vi også fattigdommen i forhold til hvor mange individer der er i denne husstand, det er derfor vigtigt at sådanne dødsfald i husstande bliver taget i betragtning.

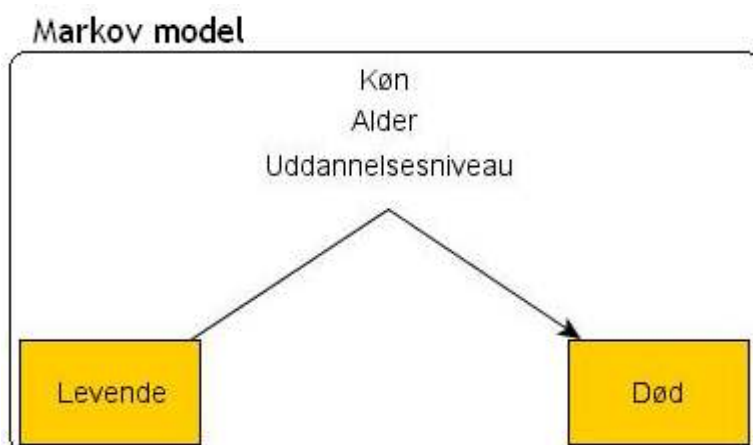
### Faktorer og model

Vi har dog kun begrænset data til rådighed vedrørende dødsfald som omfatter hele den danske befolkning. Derfor kan vi ikke kontrollere sandsynligheden for dødsfald i forhold til indkomsten. I stedet kontrollerer vi sandsynlighederne i forhold til uddannelsesniveau. Derudover kontrollerer vi også i forhold til alder og køn, som det er almen anerkendt spiller en stor rolle for dødeligheden.

Vi laver den antagelse at dødsfald er uafhængig af andre faktorer, samt at vi kan modellere den nøjagtigt med en stationær kontinuert markov proces. Det er klart at menneskers biologi er arbejder i kontinuert tid, og derfor vil kontinuitet være den bedste approksimation. Derudover har vi også splittet sandsynlighederne ind i aldersgrupper, hvorfor kravet om stationæritet (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 lever individet eller også er individet død. Der findes ingen mellem stadier, som f.eks. permanent hurtigt-dødelig sygdom. Samtlige antagelser er milde og skaber en langt mere præcis model end vi reelt har behov for. Hvis der skulle være en bias ville den påvirke den sociale mobilitet positivt for de laveste indkomstgrupper, således at vi har overvurderet den sociale mobilitet. Dette bekræftes også af en række forskere som finder at selvmords risikoen er op til 5 gange højere for lave indkomstgrupper (Qin et al, Ferda et al).

### Dødsfald analyse

Givet vores tidligere antagelser og diskussioner kan vi udregne sandsynligheder for dødsfald som en stationær kontinuert markov proces. Diagrammet til venstre illustrer modellen og de faktorer som indgår. Se appendiks for ligninger til denne model. Bemærk: da alt data anvendt har været register data af høj kvalitet, må vi antage at det ikke kun er repræsentativt men også er præcist for hele den danske befolkning. Derudover har vi givet vores tidligere antagelser at samtlige resultater må være



fuldstændige præcise. Således kan vi ikke udregne konfidens-intervaller eller på anden måde måle usikkerheden på disse data. For reelt at vurdere usikkerheden på disse data, skal man enten ind og kigge på usikkerheden af vores antagelser (som er meget milde) eller kigge på indsamlingen af data.

De statistiske resultater er givet i nedenstående tabeller:

ISCED97	Alder	Mænd	Kvinder
Pre-primary, primary and lower secondary education - levels 0-2 (ISCED 1997)	0-4	0	0
	5-9	0	0
	10-14	0,013414634	0,003996803
	15-19	0,000476716	0,000193907
	20-24	0,001192418	0,000329985
	25-29	0,001633082	0,000603415
	30-34	0,002147658	0,001165583
	35-39	0,002588935	0,001609036
	40-44	0,003540153	0,002339122
	45-49	0,004951225	0,003238638
	50-54	0,008172071	0,005352824
	55-59	0,011345102	0,00743473
	60-64	0,016582317	0,010264722
	65-69	0,024899296	0,015878353
	70-74	0,037680927	0,024575718
	75-79	0,056936648	0,040292579
	80-84	0,09626255	0,06482581
	85-89	0,171856465	0,119511636
	90-94	0,170731707	0,166153846
	95-99	0,2	0,393939394
100-104	1	0	
105-109	0	0	
110+	0	0	

ISCED97	Alder	Mænd	Kvinder
Upper secondary and post-secondary non-tertiary education - levels 3-4 (ISCED 1997)	0-4	0	0
	5-9	0	0
	10-14	0	0
	15-19	0	0
	20-24	0,000454534	6,83979E-05
	25-29	0,000327613	0,000186876
	30-34	0,000622497	0,000305802

35-39	0,000752598	0,000628439
40-44	0,001611517	0,001074013
45-49	0,003433466	0,00189122
50-54	0,005034235	0,003045167
55-59	0,007998253	0,004773237
60-64	0,013517382	0,008006522
65-69	0,022312088	0,012677947
70-74	0,035176166	0,021320446
75-79	0,064624006	0,039923476
80-84	0,107641939	0,062517986
85-89	0,307998013	0,208518519
90-94	0,142857143	0,342105263
95-99	1	0,2
100-104	0	1
105-109	0	0
110+	0	0

De sidste 6 rækker er markeret med rød da data er fejlfyldt her. Endvidere over 85 år antages at have samme dødelighed som individer mellem 80 og 84 år.

ISCED97	Alder	Mænd	Kvinder
Tertiary education - levels 5-6 (ISCED 1997)	0-4	0	0
	5-9	0	0
	10-14	0	0
	15-19	0	0
	20-24	0	7,5375E-05
	25-29	0,000183722	9,75499E-05
	30-34	0,00034493	0,000271904
	35-39	0,000587643	0,000519136
	40-44	0,000799753	0,000648138
	45-49	0,001642562	0,001201923
	50-54	0,003382286	0,002098636
	55-59	0,004566114	0,004140916
	60-64	0,011087533	0,007045107
	65-69	0,016849947	0,011020168
	70-74	0,028263607	0,017662469
	75-79	0,049421238	0,03305169
	80-84	0,098411602	0,063300402
	85-89	1,058394161	1,38
90-94	0,645833333	3,25	
95-99	1,5	0	

100-104	2	0
105-109	0	0
110+	0	0

Kilder:

Suicide Risk in Relation to Socioeconomic, Demographic,  
Psychiatric, and Familial Factors: A National Register–  
Based Study of All Suicides in Denmark, 1981–1997

Af Ping Qin, M.D., Ph.D.

Esben Agerbo, M.Sc.

Preben Bo Mortensen,

Dr.Med.Sc.

Determinants of Suicides in Denmark:

Evidence from Time Series Data

Af HALICIOGLU, Ferda and Andres, Antonio

Yeditepe University

20