# Forord

Denne oppgaven markerer avslutningen på min mastergrad i samfunnsøkonomi ved NTNU. Arbeidet med denne oppgaven har vært en spennende og lærerik prosess. Jeg vil først og fremst rette en stor takk til veilederen min, Hildegunn Ekroll Stokke for meget god og grundig veiledning. Stokke har igjennom hele prosessen gitt konstruktive og inspirerende tilbakemeldinger.

Videre vil jeg takke Vegard og Anders for gode diskusjoner og innspill om problemer som har dukket opp underveis i arbeidet med oppgaven. Jeg ønsker også å rette en stor takk til mamma og pappa for korrekturlesing av oppgaven. Til slutt vil jeg takke mine medstudenter for en fantastisk tid i Trondheim.

Jeg er selv ansvarlig for alle feil og mangler

Trondheim, 1. juni 2012

Erlend Karlsen

# Innhold

**1 Innledning** **1**

**2 Teori** **3**

2.1 Neoklassisk vekst med mobil arbeidskraft . . . 3

2.2 Betydningen av heterogen arbeidskraft . . . 13

2.3 Agglomerasjonseffekter . . . 14

2.4 Oppsummering . . . 15

**3 Tidligere forskning** **17**

3.1 Forskning gjort med en tverrsnittsmodell . . . 17

3.2 Forskning gjort med en paneldatamodell . . . 19

3.3 Oppsummering . . . 20

**4 Databeskrivelse** **21**

4.1 EU samarbeidet og regionene i EU . . . 21

4.2 Datasettet . . . 23

4.2.1 Datasett 1: EU-27 1995 - 2008 . . . 23

4.2.2 Datasett 2: EU-12 1980 - 2009 . . . 24

4.2.3 Kontrollvariabler . . . 25

4.3 Oppsummering og deskriptiv statistikk . . . 25

**5 Økonometrisk spesifikasjon** **27**

5.1 Tverrsnittsmodellen . . . 27

5.2 Paneldatamodellen . . . 29

5.3 Økonometriske utfordringer . . . 31

5.4 Oppsummering . . . 32

**6 Resultater** **35**

6.1 Tverrsnittsmodellen . . . 35

6.2 Paneldatamodellen . . . 39

6.3 Oppsummering og sammenligning mellom de to modellene . . . 44

**7 Konklusjon** **47**

**Referanser** **50**

**A Appendiks** **51**

A.1 Steady state løsningen for velferden til arbeiderne . . . 51

A.2 Dynamikken til arbeidernes velferd . . . 51

A.3 Fullt navn på NUTS regionene som er droppet fra analysen . . . 52

# Figurer


1 Fasediagram $\mu = 0$ . . . . . 10
2 Fasediagram $\mu > 0$ . . . . . 11
3 Resultater Rappaport . . . . . 12


# Tabeller


1 Minimum og maksimum terskler for gjennomsnittlig befolkningsstørrelse på NUTS regioner . . . . . 22
2 Deskriptiv statistikk . . . . . 26
3 Tverrsnittsmodell hvor jeg bruker datasett 1 . . . . . 37
4 Tverrsnittsmodell hvor jeg bruker datasett 2 . . . . . 39
5 Paneldatamodell med årlige og 3 års intervaller . . . . . 41
6 Paneldatamodell med 5 års intervaller . . . . . 43
7 Fullt navn på NUTS regionene som er droppet fra analysen . . . . . 52

# 1 Innledning

Økonomisk vekst har lenge vært et sentralt forskningsområde innenfor økonomi. Det er flere grunner til det. Økonomisk vekst er et av de viktigste målene vi har på utvikling i samfunnet. I dag består verden av store inntektsforskjeller mellom land og regioner. Europa, som blir sett på som et velutviklet og rikt kontinent, har også store forskjeller mellom regionene. Et av de politiske målene i EU er at gapene mellom disse regionene skal bli mindre. For at disse gapene skal bli mindre er den økonomiske veksten til en region viktig. Et viktig spørsmål er dermed om fattige regioner vil ta inn på rike regioner, eller vil disse forskjellene bare bli større over tid. Dette er viktig sett ut ifra et fordelingsperspektiv, hvor vi ønsker et mer rettferdig samfunn. Det er dermed viktig å se på hvilke faktorer som påvirker denne veksten.

I Europa har EU lagt mest mulig til rette for at økonomiske vilkår skal være best mulig. Dette har de gjort ved å fjerne tollbarrierer, legge til rette for handel og sørge for at arbeidskraften kan flyte fritt mellom land. Denne arbeidskraften er en viktig faktor som kan påvirke den økonomiske veksten til regioner. En slik fri flyt av arbeidskraft kan være med på å påvirke trenden til urbanisering av regionene vi ser i dag. Det er en tendens til at arbeidskraft flytter fra mer perifere og fattige regioner til mer sentrale og rikere regioner. Hvordan påvirker denne befolkningsendringen veksten i disse regionene? Blir inntektsforskjellene mellom regioner i Europa mindre over tid? Min problemstilling er om det er konvergens mellom europeiske regioner og hvordan befolkningsvekst i en region påvirker inntektsveksten.

Økonomisk vekst er blitt mye forsket på. Det sentrale begrepet i disse analysene har vært å undersøke om inntektsforskjellene i land og regioner har blitt mindre eller større over tid. Det har blitt sett på som viktig for å undersøke om land og regioner tar igjen rike regioner av seg selv, eller om det må tas eventuelle grep. EU prøver i dag å minske noe av denne forskjellen og øke veksten i fattige regioner med strukturfond. Jeg ønsker å kombinere en slik standard vekstundersøkelse med å se på hvordan eventuell befolkningsvekst påvirker konvergensen. Fattige regioner kan oppleve en stor utflytning som kan forverre de økonomiske forutsetningene en region har for å ta igjen en rikere region. Dersom de økonomiske forutsetningene blir forverret vil det være viktig for EU at de finner måter å hjelpe de fattige regionene, slik at ikke forskjellene øker.

I den empiriske analysen vil jeg ta i bruk to forskjellige modeller. De to modellene jeg har valgt å bruke er en tverrsnittsmodell og en paneldatamodell(feilkorrigeringsmodell). Ved hjelp av disse to modellene vil jeg undersøke problemstillingen min om inntektsforskjellene i Europa øker eller minker over tid, og om befolkningsveksten i en region påvirker den økonomiske veksten. Resultatet fra den empiriske analysen er at europeiske regioner opplever konvergens, slik at fattige regioner tar innpå rikere regioner. Farten på denne

konvergensen varierer en god del ut ifra hvilken modell jeg brukte. Jeg får også klare positive effekter av befolkningsvekst på inntektsveksten til en region. Siden trenden i Europa i dag er at folk flytter fra fattige regioner vil effekten av befolkningsvekst på inntektsveksten isolert sett bidra til inntektsdivergens. Men siden jeg totalt sett finner konvergens virker den positive effekten av befolkningsvekst å bli dominert av konvergensmekanismen.

Oppgaven vil bli organisert på følgende måte. I kapittel 2 vil jeg ta for meg en teorimodell som tar utgangspunkt i å forklare hvordan arbeidskraftsmobilitet påvirker konvergensen til en økonomi. Jeg vil også se på om betydningen av arbeidskraft vil påvirke dette resultatet. Til slutt i dette kapittelet ser jeg på eventuelle agglomerasjonseffekter som oppstår når mennesker og bedrifter klynger seg sammen. I kapittel 3 vil jeg vise til tidligere forskning som er gjort på konvergens og effekt av befolkningsvekst på inntektsvekst. I kapittel 4 vil jeg diskuterer hvilket datamateriale jeg bruker i den empiriske analysen min. I kapittel 5 vil jeg beskrive hvilke estimeringsmetoder jeg velger å bruke på dette datasettet og diskuterer hvilke empiriske utfordringer jeg står overfor. Mine egne empiriske resultater vil bli presentert i kapittel 6. I kapittel 7 vil jeg runde det av med en konklusjon som er basert på de empiriske resultatene mine.

# 2 Teori

I dette kapittelet skal jeg gjøre rede for den teoretiske delen av analysen. Jeg starter her med å se på om migrasjon vil ha betydning for inntektsutviklingen. Her tar jeg utgangspunkt i en neoklassisk vekstmodell hvor arbeidskraften er mobil. Denne modellen ble utviklet av Rappaport (2005) og forklarer hvordan veksten i en region blir påvirket av mobil arbeidskraft. I del to av dette kapittelet diskuterer jeg betydningen av heterogen arbeidskraft og dersom denne migrerer. Det vil si at jeg skiller mellom faglært og ufaglært arbeidskraft. Forskjellen mellom faglært og ufaglært migrasjon er at dersom faglært arbeidskraft migrerer vekk, vil dette få konsekvenser for produktiviteten i regionen. Her tar jeg utangspunkt i en modell utviklet av Rattsø og Stokke (2009). Denne modellen prøver å forklare hvorfor perifere regioner i stor grad er preget av langvarige stagnasjoner. I den siste delen av kapittelet diskuterer jeg eventuelle agglomerasjonseffekter. Ideen bak agglomerasjon er at når mennesker og firmaer klynger seg sammen for å oppnå flere stordriftsfordeler. Dette vil være av stor betydning for en region som opplever befolkningsvekst.

## 2.1 Neoklassisk vekst med mobil arbeidskraft

I en standard neoklassisk vekstmodell uten mobil arbeidskraft vil et negativt kapitalsjokk ha en rask konvergens tilbake til likevekten. Det er fordi lønnsnivået vil synke i denne økonomien og dermed gjøre det mer attraktivt med kapitalinvesteringer på grunn av økt utbytte av kapital. Resultatet av dette er relativ rask konvergens tilbake til likevekten. Slik EU/EØS fungerer i dag med fri flyt av arbeidskraft mellom regioner og land kan et kapitalsjokk fort få konsekvenser for arbeidsstyrken. Dersom arbeidsledige i en region finner det lettere å skaffe seg jobb i en annen region, eventuelt at arbeidere får en høyere lønn i en annen region vil flere velge å flytte. Rappaport (2005) utvider dermed den neoklassiske vekstmodellen ved å la arbeidskraften være mobil. Intensjonen bak denne modellen er at dersom en region blir utsatt for et negativt kapitalsjokk vil arbeidere forlate regionen til fordel for en annen. Dette fører til at lønnen blir høyere enn den ville vært uten mobilitet for arbeiderne som blir igjen på grunn av avtakende utbytte av arbeidskraft. Dette har igjen en positiv effekt på regionen, og trekker i retning av hurtigere fart tilbake til likevekten. Emigrasjon av arbeidskraft fra en region til en annen fører også til at marginalproduktiviteten til kapitalen går ned i den regionen som mister arbeidskraften. Skyggeverdien på installert kapital i denne regionen blir således lavere. Skyggeverdien på kapital er det maksimale en agent er villig til å betale for en ekstra enhet av kapital. Dette gjør investeringer i regionen mindre attraktivt og investeringer i forhold til kapitalmengden synker. Denne effekten fungerer negativt på økonomien og trekker i retning av tregere fart tilbake til likevekten.

Jeg anvender modellen til Rappaport (2005) for å undersøke betydningen av migrasjon på inntektsutviklingen og hvilken effekt som vil dominere. Før jeg setter opp modellen er jeg nødt til å gjøre flere antagelser: Verden er delt opp i to økonomier hvor man har en stor og en liten. Den store økonomien er i steady state mens den lille økonomien er i en tilstand hvor kapitalintensitet er lavere enn i steady state.[1] Ettersom den lille økonomien har vært utsatt for et negativt kapitalsjokk, og er på vei tilbake til steady state, vil det være mer lukrativt å være arbeider i den store økonomien. Dermed vil den store økonomien tiltrekke seg arbeidere fra den lille økonomien som følge av høyere lønnsnivå. Alt annet i de to økonomiene antas å være likt.

**Bedriftene**

$$Y_i(t) = AK_i(t)^{\alpha} L_i(t)^{1-\alpha} \tag{2.1}$$

$$\dot{K}_i(t) = I_i(t) - \delta K_i(t) \tag{2.2}$$

Likning (2.1) er produktfunksjonen for den aggregerte økonomien. Innsatsfaktorene er kapital, K, og arbeidskraft, L. $\alpha$ måler graden av inntekt som tilfaller eierne av kapital og er $0 < \alpha < 1$. Modellen har en standard Cobb-Douglas funksjon med konstant skalaavkastning. Konstant skalaavkastning vil si at, dersom en øker alle innsatsfaktorene med en konstant, så vil produksjonen øke med denne konstanten. Den totale faktorproduktiviteten i økonomien er A, og A er større en null. A er også eksogent gitt i denne modellen. Likning (2.2) viser hvordan kapitalen vokser over tid. Den relaterer investeringer, I, til endringer i kapitalbeholdningen. $\delta$ er kapitalens depresieringsrate. Bedriften ønsker å velge en bane for produksjonen som maksimerer profitten beregnet i nåverdi. Med en diskonteringsrate r, som er antatt å være konstant vil netto nåverdi av fremtidige kontantstrømmer være gitt ved:

$$V_i(t) = \int_t^{\infty} \left( Y_i(s) - w_i(s)L_i(s) - I_i(s) - \frac{b}{2}(\frac{I_i(s)}{K_i(s)})^2 K_i(s) \right) e^{-r(s-t)} ds \tag{2.3}$$

Lønnen til arbeiderne er gitt ved w, og b er kapitalinstallasjonsfriksjoner. Formålet til b er å bremse opp kapitalmobiliteten slik at lønnen i den lille økonomien ikke øyeblikkelig konvergerer tilbake til steady state nivået. Det fjerde leddet i (2.3) er kostnadsfunksjonen til bedriften. Dette er en kvadratisk funksjon som er argumentert for av Abel (1982) og Hayashi (1982). Her vil kostnaden av å installere ny kapital avhenge av forholdet mellom I og K. Ettersom både første- og andrederiverte er positiv vil det være en økende og konveks funksjon. Dette gjør at kostnadene for hver investeringsenhet vil øke med investeringsraten

[1] Steady state er en stabil likevekt

til en gitt K. Dersom bedriftene ønsker et nytt nivå på kapital, vil det være ekstra kostbart dersom de ønsker å hoppe rett til det nye nivået, isteden for en mer jevnere overgang på grunn av en kvadratisk kostnadsfunksjon. Bedriften ønsker å maksimere likning (2.3) under bibetingelsen gitt ved likning (2.2).

$$H = \left( Y_i(t) - w_i(t)L_i(t) - I_i(t) - \frac{b}{2}(\frac{I_i(t)}{K_i(t)})^2 K_i(t) + q_i(t)(I_i(t) - \delta K_i(t)) \right) e^{-r(s-t)} \quad (2.4)$$

Her har jeg satt opp optimeringsproblemet til bedriften som er en Hamilton funksjon. $I_i(t)$ og $L_i(t)$ er kontrollvariablene som bedriften kan velge fritt. $K_i(t)$ er en tilstandsvariabel og er bestemt ut ifra tidligere valg firmaet har tatt. q er Costate-variabelen som er skyggeverdien til kapital. Denne representerer verdien for firmaet av en ekstra enhet med kapital på tidspunkt t. Jeg finner førsteordensbetingelsene for kontrollvariablene $I_i(t)$ og $L_i(t)$ som jeg løser ut for $I_i(t)$ og $w_i(t)$:

$$\begin{aligned}
\frac{\partial H}{\partial I_i(t)} &= \left( -1 - 2 * \frac{b}{2}(\frac{I_i(t)}{K_i(t)}) * K_i(t) * (\frac{K_i(t)}{K_i(t)^2}) + q_i(t) \right) e^{-r(s-t)} = 0 \\
\Rightarrow \; 1 + b\frac{I_i(t)}{K_i(t)} &= q_i(t) && (2.5) \\
\Rightarrow \quad \frac{I_i(t)}{K_i(t)} &= \frac{q_i(t) - 1}{b} && (2.6) \\
\Rightarrow \quad I_i(t) &= \frac{q_i(t) - 1}{b} K_i(t) && (2.7)
\end{aligned}$$

Likning (2.5) viser at nivået på bedriftenes marginale investeringskostnader bør være lik verdien på en ekstra kapitalenhet. Ved å investere i dag øker verdien på bedriften. Den økte verdien er målt ved $q_i(t)$. Vi kan omformulere denne til (2.6) som sier at bedriften bør investere opp til punktet hvor marginalkostnaden av ny kapital er lik den neddiskonterte verdien av inntektene.

$$\begin{aligned}
\frac{\partial H}{\partial L_i(t)} &= \left( (1-\alpha)AK_i(t)^\alpha L_i(t)^{-\alpha} - w_i(t) \right) e^{-r(s-t)} = 0 \\
\Rightarrow \quad w_i(t) &= (1-\alpha)A\frac{K_i(t)^\alpha}{L_i(t)^\alpha} \\
\Rightarrow \quad w_i(t) &= (1-\alpha)Ak_i(t)^\alpha && (2.8)
\end{aligned}$$

I likning (2.8) er $k_i(t)$ kapital per arbeider. Ved hjelp av denne likningen kan vi se ved hvilket lønnsnivå bedriftene er villig til å ha full sysselsetting. Optimalt for bedriften vil være når marginal arbeidskraftsproduktivitet er lik reallønn. Vi har nå funnet førsteor-

densbetingelsene til L og I. Videre skal jeg nå se på hvordan kapitalraten endrer seg over tid. For å finne denne bruker jeg definisjonen på kapitalveksten per arbeider som er gitt ved likning (2.9). Jeg bruker også likning (2.2) hvor jeg har satt inn for $I_i(t)$. Jeg får dermed disse to likningene:

$$\dot{k}_i(t) = \frac{\dot{K}_i(t)}{L_i(t)} + \frac{K_i(t)}{L_i(t)^2}\dot{L}_i(t) \tag{2.9}$$

$$\dot{K}_i(t) = \frac{q_i(t)-1}{b}K_i(t) - \delta K_i(t) \tag{2.10}$$

For å finne hvordan kapitalen per arbeider endrer seg over tid setter jeg (2.10) inn i (2.9).

$$\begin{aligned}
\dot{k}_i(t) &= \frac{\frac{q_i(t)-1}{b}K_i(t) - \delta K_i(t)}{L_i(t)} - \frac{K_i(t)}{L_i(t)^2}\dot{L_i}(t) \\
\Rightarrow\ \dot{k}_i(t) &= \frac{q_i(t)-1}{b}k_i(t) - \delta k_i(t) - \frac{\dot{L}_i(t)}{L_i(t)}k_i(t) \\
\Rightarrow\ \dot{k}_i(t) &= (\frac{q_i(t)-1}{b} - \delta - \frac{\dot{L}_i(t)}{L_i(t)})k_i(t) \qquad (2.11)
\end{aligned}$$

Likning (2.11) viser kapitalveksten per arbeider. For å finne den optimale banen for skyggeprisen må en se nærmere på førsteordensbetingelsen for tilstandsvariabelen $K_i(t)$.

$$\begin{aligned}
\frac{\partial H}{\partial K_i(t)} &= A\alpha K_i(t)^{-(1-\alpha)}L_i(t)^{1-\alpha} + b\frac{I_i(t)}{K_i(t)}K_i(t)\frac{I_i(t)}{K_i(t)^2} - \frac{b}{2}(\frac{I_i(t)}{K_i(t)})^2 - \delta q_i(t) = rq - \dot{q}_i(t) \\
\Rightarrow \quad -\dot{q}_i(t) &= A\alpha\frac{K_i(t)^{-(1-\alpha)}}{L_i(t)^{-(1-\alpha)}} + b(\frac{I_i(t)}{K_i(t)})^2 - \frac{b}{2}(\frac{I_i(t)}{K_i(t)})^2 - \delta q_i(t) - rq \\
\Rightarrow \quad \dot{q}_i(t) &= (r+\delta)q_i(t) - \alpha A k_i(t)^{-(1-\alpha)} - \frac{b}{2}(\frac{I_i(t)}{K_i(t)})^2 \qquad (2.12)
\end{aligned}$$

Setter inn for $I_i(t)$ fra likning (2.7)

$$\begin{aligned}
\dot{q}_i(t) &= (r+\delta)q_i(t) - \alpha A k_i(t)^{-(1-\alpha)} - \frac{b}{2}(\frac{\frac{q_i(t)-1}{b}K_i(t)}{K_i(t)})^2 \\
\Rightarrow \dot{q}_i(t) &= (r+\delta)q_i(t) - \alpha A k_i(t)^{-(1-\alpha)} - \frac{(q_i(t)-1)^2}{2b} \qquad (2.13)
\end{aligned}$$

Likning (2.13) viser kapitalgevinstens utvikling over tid. Denne likningen sammen med likning (2.11) har et todimensjonalt system. Dette systemet med differensiallikninger kan senere brukes til å lage et fasediagram. Med et fasediagram kan en se på likevekten til

den lille økonomien og se på overgangen tilbake til likevekt.

**Mobiliteten til arbeidskraften**

$$\frac{\dot{L}_{liten}(t)}{L_{liten}(t)} = \mu(log(h_{liten}(t)) - log(h_{stor}(t))) \tag{2.14}$$

$$h_i(t) \equiv \int_t^{\infty} w_i(s)e^{-r(s-t)}ds \tag{2.15}$$

Likning (2.14) representerer mobiliteten til arbeidskraften i den lille økonomien. Vi antar her at den store økonomien er så stor at migrasjon ut eller inn i denne økonomien ikke vil ha noen effekter. I den lille økonomien vil arbeidskraften øke eller avta med en rate som er proporsjonal til den logaritmiske netto nåverdiforskjellen i lønnsinntektene mellom den store og lille økonomien. Mobiliteten til arbeidskraften blir målt med konstanten $\mu$. Dersom $\mu$ er høy vil arbeidskraften reagere mer på forskjellene i velferd enn dersom den er lav. På generell basis vil det være store forskjeller i størrelsen på $\mu$ ettersom hvor i verden man ser på denne. Den virker å være større mellom stater i USA enn den er for regioner i Europa. Dette er viktig å tenke på i den empiriske analysen. Likning (2.15) representerer velferden til arbeiderne som er gitt med $h_i(t)$.

**Store økonomiens steady state løsning**

For å finne steady state løsning til den store økonomien forenkler jeg modellen. Antar at naturlig befolkningsvekst er lik null. Dette fører til at befolkningsstørrelsen i den store økonomien er konstant. Den vil også ha full sysselsetting. Vi finner steady state verdien til q og k ved å sette henholdsvis likning (2.11) , (2.13) og $\frac{\dot{L}_i(t)}{L_i(t)}$ lik null.

$$\begin{aligned} \dot{k} &= 0 \\ \Rightarrow \quad \frac{q^* - 1}{b} - \delta &= 0 \\ \Rightarrow \quad q^* &= \delta b + 1 \end{aligned} \tag{2.16}$$

I likning (2.16) er q* steady state verdien til q, som er større enn 1. Det er fordi bedriften må reinvestere for å veie opp for depresieringsraten til kapitalen og på grunn av tilpasningskostnader (b > 0). Steady state verdien til kapitalintensiteten er gitt ved:

$$\begin{aligned} &\dot{q} = 0 \\ &\Rightarrow 0 = (r + \delta)q - \alpha A k^{-(1-\alpha)} - \frac{(q-1)^2}{2b} \end{aligned}$$

Setter inn for q* og løser ut for k*

$$\begin{aligned}
\alpha A k^{*-(1-\alpha)} &= (r+\delta)(1+\delta b) + \frac{(\delta b + 1 - 1)^2}{2b} \\
\Rightarrow k^{*-(1-\alpha)} &= \frac{2(r+\delta)(\delta b+1) - \delta^2 b}{2\alpha A} \\
\Rightarrow \qquad k^* &= A^{\frac{1}{1-\alpha}} (\frac{2\alpha}{\tilde{b}})^{\frac{1}{1-\alpha}} \qquad (2.17)
\end{aligned}$$

I likning (2.17) er k* steady state verdien til kapitalen per arbeider. $\tilde{b}$ er gitt ved $\tilde{b} \equiv 2(\delta + r) + b(\delta^2 + 2\delta r)$. For å finne velferden til arbeiderne har jeg brukt likning (2.15) og integrert denne. Se appendiks 1 for mer utledning av dette. Velferden vil være lik for de to økonomiene i steady state. Ender dermed opp med dette utrykket for velferden til arbeiderne:

$$h^* = \frac{w^*}{r} \qquad (2.18)$$

**Dynamikken og steady state for den lille økonomien**

Jeg ser først på dynamikken til den lille økonomien. Dette gjøres ved å se på hvordan de fire variablene L, k, q og h utvikler seg over tid. For q og L gjelder henholdsvis likning (2.13) og (2.14). For k gjelder likning (2.11) hvor jeg har satt inn for arbeidskraftsmobiliteten $\frac{\dot{L}}{L}$. Jeg ønsker også å ha en likning som beskriver utviklingen til h over tid. Denne finner jeg ved å derivere med hensyn på t i likning (2.15). Se appendiks A.2 for hvordan jeg får dette utrykket. Jeg får da de fire likningene nedenfor som representerer dynamikken til den lille økonomien på vei tilbake til likevekt:

$$\dot{q}(t) = (r+\delta)q(t) - \alpha A k(t)^{-(1-\alpha)} - \frac{(q(t)-1)^2}{2b} \qquad (2.13)$$

$$\frac{\dot{L}(t)}{L_(t)} = \mu(log(h_{liten}(t)) - log(h_{stor}(t))) \qquad (2.14)$$

$$\dot{k}(t) = (\frac{q(t)-1}{b} - \delta - \mu(log(h_{liten}(t)) - log(h_{stor}(t))))k(t) \qquad (2.11')$$

$$\dot{h}(t) = rh(t) - w(t) \qquad (2.19)$$

Disse fire likningene viser hvordan variablene påvirkes for den lille økonomien på vei tilbake til likevekt. Steady state løsningene vil være lik som for den store økonomien ettersom befolkningsveksten er null i steady state. Vi får da steady state løsningene q, k og h gitt ved likning (2.16), (2.17) og (2.18).

For å vise hvordan dynamikken i lille økonomien fungerer tar jeg i bruk en faseplananalyse.

Setter først opp diagrammet for en vanlig standard neoklassisk vekstmodell før jeg lar arbeidskraften være mobil. På denne måten får jeg sammenlignet forskjellene mellom disse to teoriene. I den neoklassiske vekstmodellen er $\mu = 0$ som gjør at $\frac{\dot{L}}{L} = 0$. Videre utvider jeg modellen med mobil arbeidskraft, slik at jeg fjerner antagelsen om at $\mu = 0$. Da er $\mu > 0$, noe som fører til $\frac{\dot{L}}{L} < 0$.

I fasediagrammet er fokuset på de to variablene k og q, og det er disse jeg må finne isoklinene til. En isoklin er en linje i (q, k) planet hvor differensiallikningen er satt lik en konstant. For å finne k isoklinen ser jeg på likning (2.11) hvor jeg har satt $\frac{\dot{L}}{L} = 0$. Når $q = 1 + \delta b$ er $\dot{k} = 0$, som er når markedsverdien til en ekstra enhet med kapital er lik erstatningskostnadene. Vi får da den horisontale $\dot{k} = 0$ isoklinen, hvor skjæringspunktet med q aksen er når $q = 1 + \delta b$ som vist i figur 1. Dersom $q > 1 + \delta b$ ser man fra likning (2.11) at $\dot{k} > 0$, slik at i hele området over $\dot{k} = 0$ vil pilene gå mot høyre. Dette er fordi bedriftene vil finne det lønnsomt å investere når marginalproduktiviteten er høy. Dette er i motsetning til når $q < 1 + \delta b$, for da vil marginalproduktiviteten på kapital være lav. Dermed vil pilene gå mot venstre i hele området under $\dot{k} = 0$. Det å finne q isoklinen er en vanskeligere oppgave. Setter likning (2.13) lik null og forenkler utrykket ved å multiplisere med 2b:

$$2b(r + \delta)q - 2b\alpha Ak(t)^{-(1-\alpha)} - (q - 1)^2 = 0 \tag{2.20}$$

Har her en implisitt funksjon mellom q og k. Denne deriverer jeg med hensyn på k, hvor jeg tar hensyn til at q er en funksjon av k:

$$(2b(r + \delta) - 2(q - 1))\frac{dq}{dk} + 2b\alpha A(1 - \alpha)k(t)^{\alpha-2} = 0 \tag{2.21}$$

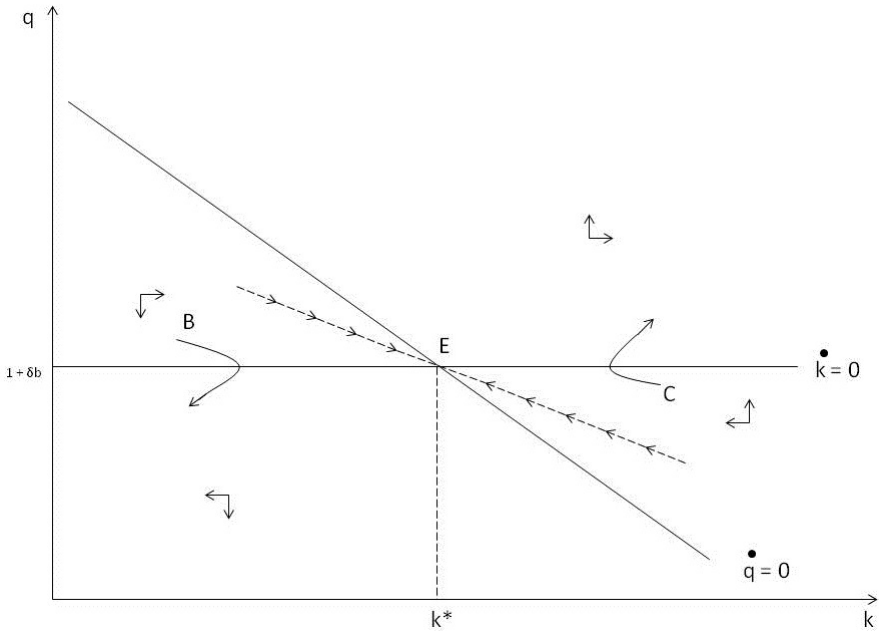
Løser ut for $\frac{dq}{dk}$:

$$\frac{dq}{dk} = \frac{b\alpha A(1 - \alpha)k^{\alpha-2}}{q - 1 - b(r + \delta)} \tag{2.22}$$

I likning (2.22) er telleren positiv og nevneren negativ. Det er enkelt å se at telleren alltid vil være positiv siden $\alpha$ alltid vil være $1 > \alpha > 0$. Nevneren er negativ dersom $q < 1 - b(r + \delta)$, noe vi vet den vil være i likevekt når $q = 1 + b\delta$ siden r $> 0$. Dette fører til at $\dot{q} = 0$ har en negativ helning som vist ved figur 1. Til venstre for $\dot{q} = 0$ synker skyggeprisen på kapitalen slik at $\dot{q} < 0$. Dette fører til piler som peker nedover på venstre side av q isoklinen. På høyreside av $\dot{q} = 0$ vil skyggeprisen på kapitalen stige slik at $\dot{q} > 0$ og pilene peker oppover. Disse bevegelsene er illustrert i figur 1. Ut ifra denne figuren kan en nå se dynamikken til den lille økonomien. Her finnes det en unik løsning av q og k som gir en sadelbane som leder til likevektsløsningen E. Dersom en starter under dette nivået

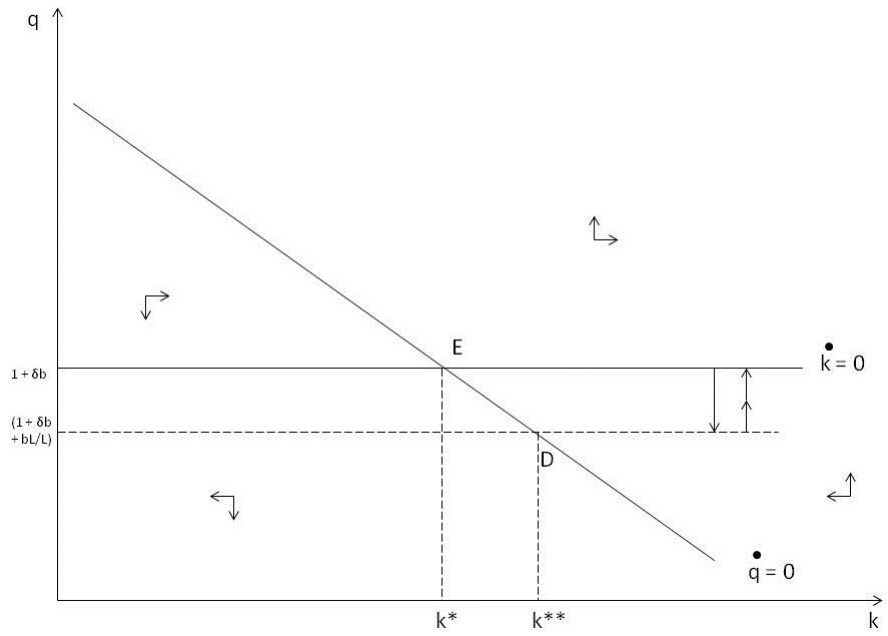
vist ved B, vil industrien etter hvert havne i det området hvor q og k faller uendelig, og en vil aldri havne i E. Det samme gjelder dersom q er for høy som i punkt C. Da vil en havne i det området hvor q og k stiger uendelig. I disse to tilfellene vil ikke bedriftene maksimere profitten og de vil ikke havne i likevekten E. I dette likevektspunktet er $q = 1 + \delta b$ og $k = k^*$.

Figur 1: Fasediagram $\mu = 0$




Dette er tilfellet i en standard neoklassisk vekstmodell med $\mu = 0$. Jeg skal nå se på et tilfelle hvor jeg lar arbeidskraften kunne flyte fritt mellom de to økonomiene. Da vil $\mu > 0$ noe som fører til at $\frac{\dot{L}}{L} < 0$. Siden $\frac{\dot{L}}{L}$ inngår i likning (2.11) vil $\dot{k}$ isoklinen endres, men den vil fortsatt være horisontal. Siden $\frac{\dot{L}}{L} < 0$ så er $1 + \delta b + \frac{\dot{L}}{L} b < 1 + \delta b$, og q er dermed lavere enn tidligere når $\dot{k} = 0$. Dette er illustrert i figur 2 ved et negativt skift i k isoklinen. q isoklinen vil være den samme som tidligere. Den vil fortsatt ha en negativ helning siden jeg antar at $r > \frac{\dot{L}}{L}$. Ved hjelp av denne antagelsen vet jeg at nevneren fortsatt vil være negativ når $q = 1 + \delta b + \frac{\dot{L}}{L} b$ som den er i den midlertidige løsningen. Punktet D er en midlertidig løsning hvor kapitalen er k** og $q = 1 + \delta b + \frac{\dot{L}}{L} b$. Når den lille økonomien har et utgangspunkt med $k < k^*$ vil sadelbanen til den lille økonomien gå mot punktet D. Etter hvert som $\frac{\dot{L}}{L}$ går tilbake mot null vil sadelbanen skifte oppover. Når $\frac{\dot{L}}{L}$ har nådd null vil sadelbanen ende som en sadelbane mot punktet E. Akkurat hvordan en slik utvikling foregår er vanskelig å bestemme fordi den avhenger størrelsen på forskjellige parametre. Rappaport prøver dermed å gjøre simuleringer av denne modellen for å undersøke hvordan en liten økonomi tilpasser seg med en $k < k^*$. Dette ser jeg kort på i neste del hvor jeg gir en oversikt over hovedresultatene til Rappaport.

Figur 2: Fasediagram $\mu > 0$




## Rappaports resultater

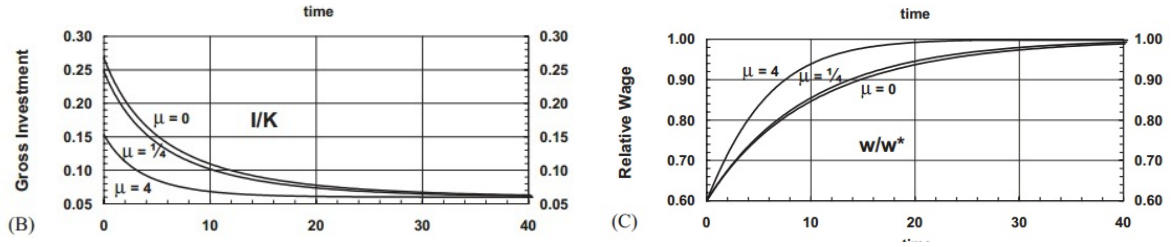
Rappaports resultater viser at et sjokk i den lille økonomien får to effekter. Den første er at arbeidere migrerer til den store økonomien, fordi de får høyere lønninger der. Dette fører til hurtigere lønnsutjevning i den lille økonomien på grunn av avtakende utbytte av arbeidskraft som er positivt for den lille økonomien. Migrasjonen ut av den lille økonomien fører også til at marginalproduktiviteten på kapital blir lavere. Dette fører til at skyggeverdien på installert kapital går ned, noe som igjen fører til lavere investeringer. Denne effekten er negativ for konvergensfarten tilbake til steady state for den lille økonomien. Hvilken effekt som dominerer er ikke lett å avgjøre fordi den er avhengig av størrelsen på parametrene i modellen. Ettersom disse tallene er ukjente kjører Rappaport simuleringer av modellen hvor han bruker benchmark verdier på disse parametrene: $\alpha = \frac{1}{3}$ , $\delta = 0.06$ og $r = 0.03$. Videre setter han en grunnverdi for steady state skyggeverdien til kapital på $q* = 1.48$ som er i et en til en forhold med b. Når det gjelder graden av arbeidskraftsmobilitet vil den variere mellom land og regioner. Det virker å være lettere å flytte mellom stater i USA enn det vil være mellom land i Europa, selv om det i Europa er fri flyt av arbeidskraft. Dette tyder på at preferansene for å flytte er forskjellig fra land til land selv om forutsetningene er noenlunde like. Rappaport ser på forskjellige tidligere undersøkelser som er på dette område og der varierer $\mu$ fra $\frac{1}{25}$ til 2. Så for å få fram poenget om eventuelle forskjeller med mobil arbeidskraft velger han å gjøre analysen med tre forskjellige verdier av $\mu$. Han velger verdiene: $\mu = 0$ hvor arbeidsstyrken er immobil, $\mu = \frac{1}{4}$ hvor mobiliteten er middels og $\mu = 4$ hvor mobiliteten er høy. Ved et sjokk vil arbeidsstyrken falle drastisk i startfasen dersom arbeiderne er mobile, men den vil flate ut ganske fort og nå sin nye steady state verdi. Når dette inntreffer varierer av hvor mobil arbeidskraften er. Etter hvert som økonomien har havnet tilbake i steady state vil verdien på k være lik

som før sjokket.

Hovedresultatene til Rappaport for de to effektene kan vises i figur 3.[2] I figur 3a ser man at investeringene er høyere jo lavere arbeidskraftsmobiliteten er. Dette er fordi jo flere arbeidere som flytter ut, jo lavere blir marginal produktiviteten til kapitalen for den regionen som blir fraflyttet. Slik at investeringer i forhold til kapitalmengden synker som jeg har argumentert for tidligere. Dette er den negative effekten av migrasjon. Når arbeidskraftsmobiliteten er null er initiell investeringsrate på 26.8% i motsetning til 15.3% for høy mobilitet. Dersom arbeidskraftsmobiliteten er middels vil den være på 24.8%. Figur 3a viser et klart bilde av at investeringer synker jo lavere arbeidskraftsmobiliteten er. I figur 3b ser en lønnseffektene av migrasjon. I denne figuren ser vi at lønnsøkningen skjer en god del hurtigere i regionen med høy arbeidskraftsmobilitet. Dette er fordi når arbeidere flytter ut av en region vil produktiviteten til de som er igjen være høyere. Dette fører til at lønnsøkningen vil gå fortere enn dersom mobiliteten var lik null. Dette er den positive effekten av migrasjon. Her starter reallønnen ved 60% av resten av økonomien uavhengig av hvilket nivå mobiliteten er på. Den initielle vekstraten til lønnen er 6.9%, 7.0% og 10.6% for henholdsvis lav, middels og høy arbeidskraftsmobilitet. Det kommer tydelig fram i figur 3b hvor reallønnen er nesten tilbake på det opprinnelig nivået allerede etter 20 år ved høy mobilitet i motsetning til tilfellet ved lavere, og ingen mobilitet som bruker nesten dobbelt så lang tid.

Figur 3: Resultater Rappaport

(a) I/K (b) W/W*




Rappaport konkluderer med at hvilken effekt som dominerer, og hvilke resultater man får avhenger av nivået på parametrene og hvor langt unna økonomien er fra steady state. Han finner at dersom økonomien er nærme steady state vil høy arbeidskraftmobilitet føre til økt inntektskonvergens. Dersom økonomien er langt unna steady state vil høy arbeidskraftmobilitet føre til redusert inntektskonvergens. Høy arbeidskraftmobiliteten vil også føre til lavere inntektskonvergens dersom økonomien har et relativt lavt inntektsnivå. Rappaport finner også at ved høy kapitalmobilitet vil konvergensfarten være mindre påvirket av arbeidskraftsmobiliteten enn dersom kapitalmobiliteten er lav.

[2]Figuren er hentet fra side 573 Rappaport (2005)

## 2.2 Betydningen av heterogen arbeidskraft

I Europa i dag virker det å være en urbaniseringsprosess. EEA (2009) argumenterer for at andelen av den europeiske befolkningen som bor i urbane strøk vil øke fra 75% til at 20% i løpet av de neste 10 årene. Denne økningen vil skje på grunn av flytting fra perifere områder til urbane områder siden de finner at befolkningstallet i Europa vil holde seg tilnærmet konstant. Det meste av denne migrasjonen til urbane områder foregår innad i hvert enkelt land, men det er en tendens til en økning i urbaniseringen mellom land også. En slik urbanisering kan være en grunn til at perifere og fattige regioner sliter med langvarig stagnasjon som Rattsø og Stokke (2009) diskuterer i artikkelen sin. Den standard neoklassiske vekstmodellen og den neoklassiske vekstmodellen utvidet med mobil arbeidskraft(Rappaport (2005)) kunne ikke forklare dette. At fattige og perifere regioner sliter med langvarige stagnasjoner strider med tidligere neoklassisk forskning. Nelson og Phelps (1966) mener at fattige regioner bør ha en høyere økonomisk vekst enn rike regioner med høy produktivitet. Nelson og Phelps mener at det vil være lettere for fattige regioner å kopiere oppfunnet teknologi og andre effektive produksjonsmetoder enn å utvikle dem selv. En slik *catch-up* effekt bør føre til at slike perifere regioner bør ha en økende produktivitetsvekst i forhold til urbane regioner. En slik effekt kan bli svekket dersom faglært arbeidskraft migrerer fra fattige regioner til rike regioner. Aghion et al. (2005) undersøkte stater i USA og fant at fraflytting av faglært arbeidskraft førte til en reduksjon i veksteffekten av forskjellen i teknologigapet mellom fattige og rike regioner. En av årsakene kan være at humankapitalen påvirker produktiviteten og at dette ikke blir tatt hensyn til. I artikkelen til Rattsø og Stokke (2009) utvider de modellen til Rappaport ved å skille mellom ufaglært og faglært arbeidskraft og endogeniterer produktivitetsvekst. Slik at de spesifisere produktivitetsveksten i den lille økonomien som følger:

$$\hat{A}_t = (\frac{Ls_t}{L_t})^{\theta_1} + \lambda(\frac{Ls_t}{L_t})^{\theta_2}(\frac{A_t^*}{A_t} - 1) \tag{2.23}$$

I likning (2.23) er det første leddet på høyre side bidrag fra innovative aktiviteter. Andre leddet representerer den teknologiske adopsjonsfunksjonen. Det siste leddet som består av $A_t$ og $A^*$ er henholdsvis det regionale produktivitetsnivå og resten av verdens produktivitetsnivå. Nivået på produktiviteten vil ha en eksogen vekstrate på lang sikt. $\lambda$ , $\theta_1$ og $\theta_2$ er konstante parametre. Nytt i denne modellen er, at dersom faglært arbeidskraft nå migrerer fra en region til en annen vil produktiviteten til den opprinnelige regionen bli lavere. Dette er fordi menneskelig kapital er viktig for teknologioverføring og nyskapning, slik at tap av faglært arbeidskraft reduserer den menneskelige kapitalen. For å undersøke betydningen av heterogen arbeidskraft gjøres mange av de samme antagelsene som i den neoklassiske vekstmodellen med mobil arbeidskraft. En stor verden bestående av en liten og stor økonomi. Den store er i steady state og er upåvirket av migrasjon, og den lille blir

utsatt for et negativt sjokk og er på vei tilbake til steady state. For å kunne tydeliggjøre produktivitetseffekten antar de at det er kun faglært arbeidskraft som migrerer til den store økonomien. Siden en ser bort fra naturlig vekst vil andelen ufaglærte holde seg konstant. De faglærte står overfor et valg hvor de kan velge å bli igjen i den lille økonomien, eller flytte til den store.

For å undersøke hvilke effekter faglært migrasjon har på regional produksjon per innbygger utsetter Stokke og Rattsø den lille økonomien for et negativt kapitalsjokk. Dette negative sjokket fører til at initiell lønnsrate og produksjon per innbygger faller til 60 % av resten av økonomien. Resultatene av dette er at i en standard neoklassisk vekstmodell vil produksjon per innbygger nå 95 % etter 23 år. Dersom man tillater migrasjon ut av regionen slik som jeg la fram i seksjon 2.1, finner Rattsø og Stokke ut at det vil ta 26 år før produksjon per innbygger er tilbake til 95 % av resten av økonomien. Her er det ikke tatt hensyn til at det er faglært arbeidskraft som migrerer. Disse resultatene tyder på at arbeidskraftsmobiliteten har liten påvirkning på konvergensen. Dersom en skiller mellom faglært og ufaglært arbeidskraft, og den faglærte arbeidskraften migrerer vil produksjonen per innbygger være på 93 % av resten av økonomien etter 50 år. Her vil økonomien bruke mye lenger tid enn i de to andre tilfellene. Dette tyder på at produktivitetseffekten ved tap av faglært arbeidskraft er stor. Denne effekten kan være med på å forklare hvorfor perifere regioner opplever langvarige stagnasjoner og sliter med å komme seg ut av dem.

## 2.3 Agglomerasjonseffekter

Når individer flytter fra perifere til mer urbane regioner kan det oppstå såkalte agglomerasjonseffekter. Agglomerasjonseffekter kan være med på å forklare hvorfor produktiviteten er høyere i urbane regioner i forhold til de perifere. Slike effekter oppstår når bedrifter og mennesker klynger seg sammen og blir ekstra produktive som følge av flere positive eksternaliteter. Moretti (2010) mener at det er tre effekter som skjer ved agglomerasjon. Den første er at et stort arbeidsmarkedet er mer attraktivt enn et lite. Stort arbeidsmarked fører til at det er lettere for en bedrift å finne passende arbeidskraft som igjen fører til lavere risiko knyttet til arbeidskraft. Det er to teorier på hvorfor dette er tilfellet. Et stort arbeidsmarked skaper en god match mellom firmaet og arbeideren. Den gode matchen fører til høyere produktivitet og lønninger. En annen fordel med et stort arbeidsmarked et at bedrifter og arbeidere er bedre rustet mot sjokk. Når det finnes mange firmaer vil det være lettere for en arbeider å finne seg ny jobb dersom bedriften en jobber i blir utsatt for et negativt sjokk og må sparke arbeidere. Det samme gjelder dersom bedriften opplever et positivt sjokk, fordi det vil være enklere å finne kvalifisert arbeidskraft med et stort arbeidsmarked i forhold til et lite. Den andre effekten som blir diskutert av Moretti er at det vil være et stort marked for innsatsfaktorer. Det vil dermed være lettere og billigere for en bedrift å skaffe seg spesialiserte innsatsfaktorer som bedriften trenger i produksjonen.

Dette gjør at bedriftene som opererer i et slikt marked klynger seg sammen for å få til en mer effektiv prosess. Bedriftene kan oppnå synergi både når det gjelder transport og logistikk etc. Den tredje effekten er teknologisk spillover. Når bedrifter og menneskelig kapital klynger seg sammen fører det til informasjonsflyt og legger til rette for diffusjon av teknologi.

Forskning på området tyder på at agglomerasjonseffekter vil variere mellom forskjellige industrisektorer. Melo et al. (2009) viser at urban agglomerasjon har mye større effekter på servicesektoren enn den har på industri. Årsaker til dette kan være noe som Glaeser og Gottlieb (2009) har undersøkt. De argumenterer og finner bevis for at servicesektoren er mer avhengig av direkte menneskelig kontakt og vil lokalisere seg i befolkede og tette regioner. Lokalisering i disse områdene medfører økte kostnader i form av bo- og leiepriser. For industrisektoren vil ikke dette nødvendigvis være lønnsomt og de lokaliserer seg utenfor disse områdene. Bosker (2007) argumenterer for at de negative effektene ved for eksempel trafikkorker, økning i pris på hus, og lokal konkurranse kan overgå de positive. Bosker instrumenterer agglomerasjon med tettheten til arbeidsstyrken og estimerer en paneldatamodell. Resultatene av dette er at agglomerasjon varierer fra positiv til negativ ved forskjellige spesifikasjoner av modellen. Bosker finner en negativ effekt av agglomerasjon ved spesifikasjonen han la mest vekt på.

Når en skal analysere agglomerasjonseffekter står en ovenfor flere empiriske utfordringer. Det kanskje største problemet er å bestemme kausaliteten mellom agglomerasjon og produktivitet. Ciccone (2001) har en forklaring på hvordan denne kausaliteten kan gjelde. Er produktiviteten høy i urbane regioner på grunn av agglomerasjonseffekter? Eller er agglomerasjonseffektene en konsekvens av høy produktivitet? Han løser dette problemet med å bruke instrumentvariabelmetoden. Ved bruk av denne metoden finner han signifikante agglomerasjonseffekter i USA og europeiske land. Han finner ikke signifikante forskjeller i agglomerasjonseffektene mellom de europeiske landene. En slik kausalitet er en utfordring som jeg må ta hensyn til i min empiriske analyse.

## 2.4 Oppsummering

I denne teoridelen har jeg tatt utgangspunkt i en vekstmodell utviklet av Rappaport (2005) som undersøker betydningen av migrasjon på inntektsutviklingen i en økonomi. Denne vekstmodellen skilte ikke mellom faglært og ufaglært arbeidskraft. Derfor diskuterer jeg i del to eventuelle konsekvenser av å skille mellom migrasjon av faglært og ufaglært arbeidskraft. Her hadde faglært arbeidskraft betydning for produktivitetsveksten. Til slutt så jeg på eventuelle agglomerasjonseffekter som oppstår når bedrifter og mennesker klynger seg sammen. Disse tre delene tar utgangspunkt i å forklare effekten av befolkningsvekst på den inntektsveksten i en økonomi. Resultatet av disse teoriene er at befolkningsvekst har både positiv og negativ innvirkning på inntektsveksten. Jeg gir en

kort oppsummering nedenfor av de forskjellige effektene for en region med utflytting:

- Positiv effekt gjennom hurtigere lønnsutjevning, som skjer på grunn av avtakende utbytte av arbeidskraft. En hurtigere lønnsutjevning fører til økt inntektskonvergens.
- Negativ effekt gjennom lavere kapitalinvesteringer, på grunn av redusert avkastning på kapital.
- Negativ effekt hvis det er faglært arbeidskraft som migrerer ut av regionen. Dette fører til en nedgang i produktiviteten.
- Negativ effekt hvis det er agglomerasjonseffekter. Regioner som opplever en sentralisering vil vokse fortere.

I effektene listet opp ovenfor er det fire negative og en positiv effekt ved utflytting på regionens inntektsvekst. Kan standardeffekten via avtakende utbytte av arbeidskraft kunne dominere de andre effektene, slik at migrasjon fra utkant til by bidrar til inntektskonvergens? Dette ønsker jeg å undersøke i den empiriske analysen.

# 3 Tidligere forskning

I dette kapittelet skal jeg undersøke tidligere forskning som er gjort på konvergensfart og effekt av befolkningsvekst på BNP per innbygger. Jeg skiller mellom forskning som er gjort ved bruk av en tverrsnittsmodell og paneldatamodell, slik at jeg deler denne tidligere forskningen opp i to deler. Jeg starter med å se på tidligere undersøkelser som er gjort ved bruk av en tverrsnittsmodell. Siste del av kapittelet ser jeg på tidligere forskning som er gjort ved bruk av paneldata.

## 3.1 Forskning gjort med en tverrsnittsmodell

Når det kommer til forskning på økonomisk vekst ved bruk av en tverrsnittsmodell er dette undersøkt av flere. Barro og Sala-i Martin (1992) finner i en undersøkelse for stater i USA en ubetinget konvergens på rundt 2 prosent i perioden 1840 til 1988. I samme undersøkelse finner de for 98 land en betinget konvergens på rundt 1.8 prosent. Senere gjør Sala-i Martin (1996) en liknende undersøkelse og finner en ubetinget konvergens på 1.5 prosent for 90 regioner i Europa over tidsperioden 1950-90. Sala-I-Martin får for de samme 90 regionene en betinget konvergens på 1.5 prosent. Cuadrado-Roura et al. (2000) gjør en liknende analyse og undersøker for konvergens i Europa over perioden 1977-94 ved hjelp av tverrsnitt og paneldata. For tverrsnittsanalysen finner de en ubetinget konvergensfart i nærheten av 2 prosent. De estimerer også for betinget konvergens ved å bruke faste effekter i en paneldatamodell. Resultatet av dette ble en konvergensfart i nærheten av 20 prosent. Fingleton (1999) undersøker ubetinget og betinget konvergensfart for regioner i Europa i perioden 1975-95 ved bruk av *maximum likelihood* og markov matriser. Konklusjonen han trekker av dette var at konvergensfarten for begge disse estimeringene var en god del lavere, og kanskje til og med ubetydelig dersom den samme trenden fortsatte framover.

Dette var undersøkelser som undersøkte konvergensfarten til land og regioner. Jeg ønsker også å undersøke hvordan befolkningsvekst påvirker disse parametrene. Braun (1993) undersøkte effekten av nettomigrasjon på vekstraten til BNP per innbygger. Nettomigrasjon er en viktig del av befolkningsveksten og dermed relevant for min undersøkelse. Braun var spesielt interessert i å se på hvordan nettomigrasjon påvirket konvergensfarten. Braun tok utgangspunkt i en standard tverrsnittsmodell med gjennomsnittlig vekst i BNP per innbygger som en avhengig variabel og initell verdi på BNP per innbygger som måler konvergensfarten. Deretter utvidet han denne modellen ved å inkludere nettomigrasjon. Utvalget hans består av regioner i Europa, USA og Japan, men han undersøkte også spesielt Tyskland, Storbritannia, Italia, Frankrike og Spania. Ved å først undersøke standardmodellen før man inkluderer nettomigrasjon får man mulighet til å undersøke hvordan den påvirker konvergensfarten og veksten til BNP per innbygger. Ved standardmodellen får han en betinget konvergensfart på 1.97 prosent for Europa over hele dataperioden

som er 1950 - 1985. Når han videre inkluderer nettomigrasjon er migrasjonskoeffisienten for Europa på 0.022, men den er ikke signifikant. Dette vil si at dersom nettomigrasjon til en region i Europa øker med 10 prosentpoeng vil veksten til BNP per innbygger øke med 0.22 prosentpoeng. Denne var ganske lik for de andre observasjonene bortsett fra for Frankrike og Spania som hadde en negativ koeffisient. Inkluderingen av nettomigrasjon endret konvergensfarten fra 1.97 prosent til 1.88 prosent slik at den hadde en liten negativ effekt. Storbritannia skilte seg klart ut fra utvalget med en migrasjonsparameter på 4.22 og klart signifikant. Her vil økning i nettomigrasjon på 10 prosentpoeng føre til en økning veksten til BNP per innbygger på 40.22 prosentpoeng.

Braun instrumenterte også nettomigrasjon med befolkningstetthet og klima m.m. Dette gjorde Braun fordi han mente det var mulig at migrasjon korrelerte med restleddet. Ved å gjøre dette får migrasjonsparameteren i de fleste tilfellene negativt fortegn bortsett fra for Storbritiania, Italia og USA. For Europa er migrasjonsparameteren nå -0.193 og en konvergensfart på rundt 2 prosent, hvor kun konvergensen er signifikant. Til slutt estimerer han modellen hvor han legger til flere kontrollvariable som tetthet og vær. Resultatet av dette var en migrasjonskoeffisient på 0.04 som er sterkt usignifikant og en konvergensfart på rundt 2 prosent.

Samme type undersøkelse ble gjort av Barro og Sala-i Martin (2004). De har tatt for seg de samme landene som Braun, med unntak av en felles undersøkelse for hele Europa. For USA estimeres effekten av migrasjon til å være 0.0931 og signifikant. Konvergensfarten de fant var på 2.31 prosent som også er signifikant. Begge er tilnærmet lik estimatene for USA under Braun. De får også tilnærmet like resultater for de andre landene bortsett fra Storbritannia. De fikk en nettomigrasjonskoeffisient for Storbritannia på 0.116 som ikke er signifikant og en konvergensfart på 2.2 prosent i motsetning til 4.220 og 3.48 prosent som Braun fikk. Dette er betydelig mindre enn hva Braun kom fram til. Tyskland har også en negativ migrasjonskoeffisient her i motsetning til Braun hvor den var positiv. Barro og Sala-i-Martin instrumenterer også migrasjon. Som i Braun skifter også her migrasjonsparameteren og blir stort sett negativ. Storbritannia er den eneste med positiv koeffisient på 0.222, og en konvergensfart på 2.61 prosent, hvor ingen av de er signifikante. Tyskland har den laveste med -0.542 og en konvergensfart på 1.81 prosent. Her er kun konvergensfarten signifikant. Barro og Sala-i Martin (2004) synes det er overraskende at migrasjonskoeffisienten ikke er signifikant for de Europeiske landene. Årsaken til dette mener de kan være at regional nettomigrasjon for europeiske regioner ikke er presist nok målt. De mener også at migrasjon kan ha et endogenitetsproblem og derfor instrumenterer de nettomigrasjon. Ved å gjøre dette får de negativ effekt av nettomigrasjon på vekstraten til BNP per innbygger for alle landene bortsett fra Storbritannia.

## 3.2 Forskning gjort med en paneldatamodell

Analyser som er gjort med en tverrsnittsmodell finner stort sett en konvergensfart på rundt 2 prosent som er en forholdsvis lav konvergensfart. Dette er ikke nødvendigvis tilfellet når en bruker paneldata. Caselli et al. (1996) undersøker konvergensfarten til land ved å bruke *Generalized method of moments*(GMM) estimering på en paneldatamodell. Ved bruk av denne metoden fant de en konvergensfart på 10 prosent. Dowrick og Rogers (2002) bruker GMM estimering på 57 land over perioden 1970 - 1990. De finner, ved å dele opp dataene i 5 års intervaller, en konvergensfart på rundt 5 prosent. Badinger et al. (2002) estimerer også en paneldatamodell med GMM metoden og kontrollerer denne for geografisk beliggenhet. Utvalget deres er NUTS 2 regioner i Europa over en periode på 1985 til 1999. Resultatene av dette er at før de tar hensyn til den geografiske beliggenheten, er konvergensfarten ved GMM metoden på opp mot 20 prosent. Med kontrollering for den geografiske beliggenheten er den på rundt 7 prosent.

Bosker (2007) undersøker europeiske regioner på NUTS 2 nivå i perioden 1977 - 2002 som han deler opp i 5 års intervaller. Her estimerer han konvergensfarten og betydningen av gjennomsnittlig befolkningsvekst ved bruk av faster effekter og periodedummier. Resultatene til Bosker varierer ut ifra hvilken estimeringsmetode han bruker. Ved å estimere modellen med faste effekter får han en konvergensfart på 2.33 prosent. Estimerer han modellen med faste effekter og periodedummier får han en konvergensfart på 12.7 prosent, noe som er en god del høyere. Når det gjelder effekten av befolkningsvekst på inntektsveksten får han en estimert verdi på henholdsvis -0.107 med faste effekter og -0.016 med faste effekter og periodedummier, hvor -0.016 ikke er signifikant. Bosker får dermed en negativ effekt av befolkningsvekst på inntektsveksten. En økning i gjennomsnittlig befolkningsvekst på 10 prosentpoeng vil føre til en reduksjon i gjennomsnittlig BNP per innbygger på 1.07 prosentpoeng for estimeringen med faste effekter. Bosker undersøker også effekten av agglomerasjon på veksten til BNP per innbygger ved å bruke yrkesbefolkningstetthet som proxy for agglomerasjon. Ved å gjøre dette får han både positive og negative effekter av agglomerasjon på veksten til BNP per innbygger. Estimeringen med faste effekter og periodedummier får en agglomerasjonskoeffisient på -0.080. Siden det er estimeringen med faste effekter og periodedummier han legger mest vekt på argumenterer han for at de negative eksternalitetene av agglomerasjon er større enn de positive. Resultatene til Bosker endrer seg mye ut i fra hvilken estimeringsmetode han bruker. Dette kan tyde på at uobserverbare region- og periodespesifikke effekter påvirker resultatet. Det vil derfor være viktig å kontrollere for disse noe jeg gjør i paneldatamodellen min.

Ved bruk av tverrsnittsmodell er det kommet fram til en norm på konvergensfarten på rundt 2 prosent. Bruker man paneldata viser tidligere forskning at man får en mye høyere konvergensfart. Argumentet for denne høye konvergensfarten er man ikke får kontrollert for forskjellige steady state nivå i tverrsnittsmodellen. Shioji (1997) har rettet kritikk mot

denne type estimering og mener at selv med paneldata vil man få den samme konvergensfarten som ved en tverrsnittsmodell dersom man gjør analysen riktig. Shioji mener problemet ligger i at tidshorisonten på de eksisterende paneldatasettene er alt for korte, spesielt når man deler datasettet inn i 5 års intervaller. Ved å bruke et langt datasett for USA finner han en konvergensfart på nærmere 2 prosent med en paneldatamodell. For å sammenligne med korte datasett, korter han ned sitt eget datasett til en 60 års periode. Da finner han en mye høyere konvergensfart. Dette mener han er bevis på at analyser med en paneldatamodell ville gitt samme resultater som en tverrsnittsmodell dersom tidsperioden var lang nok. Shioji mener at estimeringer med paneldatamodell vil være beheftet med en skjevhet oppover på grunn av for kort tidsserie. Dette kan være et problem for min analyse ettersom tidshorisonten min kun er på 30 år.

## 3.3 Oppsummering

Empiriske analyser ved bruk av tverrsnitt- og paneldatamodeller gir forskjellige resultater. Variasjonen i konvergensfarten mellom disse forskjellige modellene er forholdsvis stor. Bruker man en tverrsnittsmodell ser konvergensfarten ut til å ligge på rundt 2 prosent, noe som ikke er en veldig hurtig konvergensfart. Med en konvergensfart på 2 prosent vil en fattig region bruke 34 år på halvere forskjellen til en rik region, noe som er forholdsvis lang tid.

Ved bruk av paneldata finner man en god del høyere konvergensfart. Gjennom forskjellige analyser i en paneldatamodell varierte konvergensfarten fra 5 prosent til godt opp mot 20 prosent, noe som er en mye raskere konvergensfart enn ved en tverrsnittsmodell. Effektene av å undersøke befolkningsvekst var også skiftende. Braun fant en positiv effekt av nettomigrasjon som ble til en negativ og usignifikant effekt når han instrumenterte denne i sin tverrsnittsmodell. Braun konkluderte med at nettomigrasjon ikke var signifikant og hadde liten påvirkningskraft på inntektsveksten. Bosker fant en negativ effekt av befolkningsvekst på vekstraten til BNP per innbygger.

# 4 Databeskrivelse

I dette kapittelet skal jeg beskrive datamaterialet jeg skal bruke i den empiriske analysen. For å svare på problemstillingen min om befolkningsvekst påvirker BNP per innbygger, og om inntektsforskjellene blir mindre over tid bruker jeg to forskjellige modeller. Den ene er en tverrsnittsmodell og, den andre er en paneldatamodell (feilkorrigeringsmodell). Jeg kommer til å forklare disse nærmere i kapittel 5. For å gjøre en slik empirisk analyse trenger jeg i hovedsak regional data på BNP per innbygger og befolkningsutvikling. Dataene er stort sett hentet eller tilsendt fra Eurostat sine sider. Jeg vil starte med å gi en kort forklaring til opprettelsen av den Europeiske Union hvor jeg videre forklarer Eurostat sin rolle. Til slutt forklarer jeg de to forskjellige datasettene og hvilke variable jeg velger å bruke i analysen.

## 4.1 EU samarbeidet og regionene i EU

Den europeiske union er en økonomisk og politisk union som kom i etterkant av den andre verdenskrig. Det hele startet med det europeiske kull og stållfelleskap (EKSF) som ble opprettet i 1950. Landene som signerte denne avtalen var Frankrike, Vest - Tyskland, Belgia, Nederland og Luxembourg. Denne avtalen skulle forene det Vestlige Europa og gjøre det vanskelig å starte en ny krig. Den knyttet sammen tungindustrien ved å lage et felles marked for kull og stål. I 1957 signerte de samme landene Roma-traktaten og Eurotam som førte til opprettelsen av det europeiske økonomiske felleskap (EØF). Målet til EØF gikk ut på at det skulle være fri flyt av varer, tjenester, personer og kapital på tvers av medlemslandene. Dette er sentralt for analysen jeg skal gjøre senere i oppgaven, at arbeidere kan flytte fritt mellom medlemslandene. I 1973 ble Storbritannia, Danmark og Irland medlemmer av denne avtalen. I 1993 kom Maastricht-traktaten og navnet på unionen ble byttet til den Europeiske Union(EU) som den heter i dag. EU bestod da av 12 medlemsland. Siden den gang har medlemstallet økt jevnlig. I 1995 ble tre land til medlem av unionen. Valutakursen euro kom for 12 av medlemslandene i 2002 og nå har hele 17 land denne valutakursen. Den største utvidelsen for EU kom i 2004 da ti nye land tegnet medlemskap. Senere fulgte enda flere land etter, og nå består EU av 27 land, hvor enda flere ønsker å bli medlem.

For å finne data for BNP per innbygger, samt befolkningsvekst for disse landene, har jeg brukt Eurostat sine internettsider. Eurostat er Europas statistiske senter. Senteret får tilsendt data fra medlemslandene, bearbeider dem, og publiserer sammenlignbar informasjon for de europeiske landene. Denne samlingen av informasjon er viktig og brukes på forskjellige områder. Først og fremst er statistikkene viktig for lokale myndigheter og næringsliv når det skal tas beslutninger. Statistikkene er også viktig for offentligheten og media, slik at resultatene kan evalueres for politikere og andre beslutningstagere. For å

kunne produsere regional statistikk for EU, deler Eurostat Europa opp i tre forskjellige nivåer. Dette blir gjort for å lage et enkelt og oversiktlig system over regionene i Europa. Disse nivåene blir kalt NUTS 1, 2 og 3.[3] Systemet går ut på at et medlemsland blir først delt opp i NUTS 1 regioner. Disse blir igjen delt opp i NUTS 2 regioner, hvor disse igjen blir delt opp i NUTS 3 regioner. NUTS reguleringene har visse maksimum og minimums terskler for gjennomsnittlig befolkningsstørrelse for nivåene som vi ser i tabell 1.[4] Det er noen regioner som viker fra disse tersklene. Hver region blir tilegnet en spesiell kode som forteller hvilken region det er. Koden begynner med to bokstaver som referer til landet. For eks. BE står for Belgia. Deretter kommer det et tall som sier hvilken NUTS 1 region det er. Slik at BE2 er regionen Vlaams Gewest i Belgia. NUTS 2 regioner blir tilegnet et tall til som gjør at Province Antwerpen har kode BE21 og ligger i Vlams Gewest. For en NUTS 3 region blir det lagt til enda et tall slik at Arondissement Antwerpen har kode BE211 som ligger i Province Antwerpen. I den nåværende NUTS klassifiseringen er det 97 NUTS 1 regioner, 271 NUTS 2 regioner og 1303 NUTS 3 regioner.

Tabell 1: Minimum og maksimum terskler for gjennomsnittlig befolkningsstørrelse på NUTS regioner

| **Level** | **Minimum** | **Maksimum** |
|---|---|---|
| NUTS 1 | 3 million | 7 million |
| NUTS 2 | 800 000 | 3 million |
| NUTS 3 | 150 000 | 800 000 |

Alle de tre NUTS-nivåene kan brukes til å gjøre empiriske analyser. Hvilken som er mest effektiv å bruke kommer an på hva man skal undersøke, og på hvilket nivå man ønsker å undersøke. Som det er diskutert i Eurostat (2007) ble det i Brussel 1961 avholdt en regional økonomikonferanse. Der kom de fram til at NUTS 2 regioner var det mest hensiktsmessige å bruke, dersom en skulle analysere regionale og nasjonale problemer. Dette er det nivået medlemslandene bruker når de driver med politikk på regionalt nivå. Skal man undersøke effekter av toll og økonomisk integrasjon er NUTS 1 regioner bedre. På mer spesifikke undersøkelser for å se på om eventuelle tiltak er nødvendig vil NUTS 3 være mest passende. Jeg velger derfor å bruke NUTS 2 regioner i analysen min. Noen mener at bruken av NUTS 2 regioner er uheldig. Boldrin og Canova (2001) mener at for mange av NUTS 2 regionene ikke vil være store nok, eller ha de nødvendige heterogene faktorene som gjør at de kan behandles som selvstendige økonomiske områder. De mener også at det er et for lite forhold mellom aktivitetene som skjer i en NUTS 2 region, og hva som er rapportert av de offisielle statistikkene. I tillegg mener de at det er for stor forskjell mellom NUTS 2 regioner ved at noen har et bredt sett med faktorer, mens andre er små. Det finnes også andre som mener NUTS 2 er det beste nivået for å gjøre disse analysene. Arbia et al. (2005) mener at NUTS 2 er det mest hensiktsmessige nivået for Europeiske analyser. De mener også at regionene er tilstrekkelig små, slik at en får fanget opp nasjonale variasjoner innad i

[3]Nomenclature des Unités territoriales statistiques
[4]Tabellen er hentet fra side 10 Eurostat (2007)

medlemslandene. Her får man også studert hendelser på et meso nivå. De argumenterer også for at det er det beste nivået å bruke for politiske formål.

## 4.2 Datasettet

For å gjøre en så presis analyse som mulig ønsker jeg å ha flest mulig observasjoner. Derfor har jeg fått tak i to datasett fra Eurostat, som er ESA79 og ESA95. Begge disse har både positive og negative sider. ESA79 går fra 1979 til 1996, men er dessverre noe mangelfullt på grunn av at det er få regioner med. ESA 95 er en oppgradert versjon av ESA79 og går fra 1995 til 2008. Den inneholder mange flere regioner enn ESA79, men her er tidshorisonten kort.

### 4.2.1 Datasett 1: EU-27 1995 - 2008

Dette datasettet består av BNP per innbygger og befolkningen for EU-27 landene på NUTS 2 nivå som har tilgjengelig data. Disse er hentet fra Eurostat sine sider. Befolkningen i regionen er regnet ut fra et gjennomsnitt fra begynnelsen og slutten av året, og er gitt i per 1000 innbygger. BNP per innbygger er gitt i euro og var oppgitt i dagens markedspriser. Når dette er tilfelle kan inflasjon påvirke resultatene mine, og dette vil være uheldig. Derfor benyttet jeg en BNP deflator fra WDI.[5] En BNP deflator måler prisnivået på innenlandske varer og tjenester for hvert land.[6] BNP deflator og konsumprisindeksen (KPI) er ganske like, men fordelen med BNP deflatoren er, at den tar hensyn til alle varer og tjenester produsert, i landet og ikke kun en fast "handlekurv". Dette gjør det mulig å sammenligne BNP over flere tidsperioder i konstante priser. Basisåret for de fleste landene er satt til 2005, mens noen land bruker andre basisår. Jeg har normalisert alle basisår til 2005. For å få BNP per innbygger i konstante priser deler jeg nominelt BNP på BNP-deflatoren og multipliserer med 100. Da vil jeg få reelt BNP per innbygger som er BNP per innbygger i konstante priser. Det er reelt BNP per innbygger jeg bruker i analysen min.

Det er ikke alle land som er like gode til å rapportere inn sine data for alle regioner. Derfor har jeg vært nødt til å fjerne noen regioner. Danmark sine data for befolkningsvekst er ikke tilgjengelig før 2004, og de fem regionene droppes derfor helt fra analysen. Det samme gjelder for to regioner fra Storbritannia, som ikke er tilgjengelig før 2001. Se appendiks A.3 for hvem som er ekskludert. Datasettet består dermed av 264 regioner over en tidsperiode på 14 år. Dette fører til 3696 observasjoner for dette datasettet.

[5] World Development Indicators som er den primære verdens bank databasen for utviklings data. http://databank.worldbank.org

[6] For Romania og Bulgaria var BNP deflatoren ubrukelig. Derfor har jeg brukt BNP deflatoren på Ungarn fordi det er nabolandet

### 4.2.2 Datasett 2: EU-12 1980 - 2009

Får å få flest mulig observasjoner har jeg fått tilsendt et datasett fra Eurostat basert på det gamle ESA 79 systemet. Dette består av data på NUTS 2 nivå for BNP per innbygger som er angitt i PPS for EU-12. Tidsperioden er fra 1979 til 1996, men siden mange observasjoner fra 1979 mangler, begynner jeg analysen i 1980. PPS er en fiktiv valuta som eliminerer forskjellen mellom forskjellige prisnivåer mellom landene. Denne er oppnådd ved å dividere den nasjonale valutaen på landets egen kjøpekraftsparitiet (KKP). KKP er et vektet gjennomsnitt av bestemte homogene varer og tjenester. Siden denne tidsperioden går kun fram til 1996 bruker jeg ESA 95 for å få en sammenhengende tidsserie fram til 2008. ESA 95 er også gitt i enheten PPS for at det skal være likt som ved ESA 79. Jeg har også her brukt en BNP deflator for å få konstante priser. Befolkningstall er tilgjengelig på eurostat sine sider helt tilbake til 1990. Fra 1980 til 1990 har jeg brukt gamle årbøker som er i pdf format på Eurostat sine sider.[7]

Jeg benytter data kun for regioner fra EU-12 ettersom det ikke finnes data for landene som ble med i EU etter dette så tidlig. Dessverre er det også mye mangler i dette datasettet, slik at en del regioner må fjernes.[8] Når det gjelder Italia og Irland finnes det ikke data for BNP per innbygger, slik at de to landene fjernes helt fra analysen. For Tyskland mangler det data på BNP per innbygger fra 10 regioner, slik at disse regionene også blir utelatt fra analysen. I Spania og Frankrike mangler det tall på to og fire regioner, slik at disse også må fjernes. For Nederland mangler det tre, og for Portugal mangler det fem regioner, som også ekskluderes fra analysen. Når det gjelder Storbritannia mangler store deler av datasettet på NUTS 2, at der har jeg valgt å bruke NUTS 1 regioner. Når det gjelder disse NUTS 1 regionene er det også en del mangler, slik at jeg ender opp med å fjerne fem regioner. Totalt sett står jeg igjen med 112 regioner. Av disse er 7 NUTS 1 regioner og 105 NUTS 2 regioner. Dette gir totalt 3360 observasjoner for datasett 2.

Selv om jeg har fjernet en del regioner som har mye mangler er datasettet fortsatt litt mangelfullt. For de to gjenværende portugisiske regionene PT11 og PT15 mangler befolkningstall for 1980. Det samme gjelder for alle de greske regionene. For de to belgiske regionene BE24 og BE31 mangler befolkningsdata for 1980 til 1984. Dette er løst ved å benytte tallet fra året etter der det mangler informasjon. For Nederland mangler BNP per innbygger data for 1980, det samme gjelder for franske regioner i 1981 og regioner i Storbritannia for 1990. For å løse dette har jeg gjort som jeg gjorde med befolkningsdata, og satt inn for året etter.

[7] Her varierte det litt når landene målte befolkningstallet sitt, men det var små forskjeller i forhold til 1990 og utover.

[8] Som ved datasett 1. Se appendiks A.3 for hvem som er ekskludert fra analysen.

### 4.2.3 Kontrollvariabler

Det er viktig å kontrollere for variabler som kan være med å påvirke resultatet i den empiriske analysen. Derfor har jeg innhentet data på to kontrollvariable. En kontrollvariabel er en variabel som holdes konstant og er gitt i modellen. Kontrollvariabler er viktig for å vite at det ikke kun er tilfeldig variasjon som påvirker resultatet. Videre følger en forklaring på de kontrollvariablene jeg har tilgang til, og mener er viktig for analysen.

En variabel som kan ha stor betydning for veksten til en region er utdanningsnivået. For denne kontrollvariabelen har jeg valgt å bruke to forskjellige utdanningsdata avhengig av hvilke metode jeg bruker. Den ene består av den yrkesaktive befolkningen over 15 år med *Tertiary education level 5 og 6* på NUTS 2 regionnivå, noe som tilsvarer høyere utdanning. Disse tallene har jeg funnet på Eurostat sine sider under *Regional labour market statistics*. Denne har jeg valgt å bruke i tverrsnittsmodellen min og fordelen er at denne er gitt for hver region. Et lite problem med denne er at den ikke går lenger tilbake enn til 1999. Året 1999 inneholder også en del mangler for regionene i analysen min, slik at jeg har valgt 2000 som startår. Dette kan by på problemer noe som jeg diskuterer i kapittel 5.3.

Optimalt sett burde jeg bruke slike utdanningsdata i paneldatamodellen også, men siden den ikke går lenger tilbake enn til 2000 vil den ikke være til stor nytte der. Derfor har jeg valgt å bruke et annet mål på utdanning i den modellen. For paneldatamodellen har jeg tatt i bruk *The Barro-Lee Dataset(2011)*.[9] Dette datasettet inneholder utdanningsnivå på landsbasis hvor jeg har brukt de landene som er i analysen min. Disse dataene er gitt i 5 års intervaller og inneholder andel av befolkning over 15 år med fullført høyere utdanning. Jeg har også valgt å korrigere for andel av befolkningen med fullført høyere utdanning, og de som kun har fullført videregående utdanning. Dette gjør jeg for å sjekke om resultatene forholder seg stabile.

En annen variabel som kan være med å påvirke veksten til en region er befolkningstetthet. Befolkningstetthet blir betegnet som antallet innbyggere per kvadratkilometer i en region. Dataene har jeg hentet fra Eurostat sin demografiske regions statistikkside. Denne effekten kan fungere som et mål på eventuelle agglomerasjonseffekter.

## 4.3 Oppsummering og deskriptiv statistikk

For å gjøre den empiriske analysen min bruker jeg to datasett. Datasett 1 er for regionene i EU-27. Dette datasettet har en relativ kort tidsserie, men mange regioner. Datasett 2 består av regionene i EU-12. Dette datasettet har ikke så mange regioner, men går over en lang tidsserie. Disse to datasettene inneholder observasjoner både over tid og over regioner. Dette fører til at jeg kan bruke både en tverrsnittsmodell og en paneldatamodell.

[9] Hentet fra www.barrolee.com

Paneldata blir betegnet som datasett med observasjoner over både tid og individ. Når jeg gjør estimeringene mine vil jeg ha mulighet til å kontrollere både for befolkningstetthet og utdanning. I tabell 2 er den deskriptive statistikken for variablene jeg skal bruke i den empiriske analysen. Grunnen til at det er så stor forskjell på initielle verdier i datasett 1 og datasett 2, er på grunn av i datasett 2 har jeg brukt NUTS 1 regionene til Storbritannia.

Tabell 2: Deskriptiv statistikk

| Variabler | Gjennomsnitt | Standardavvik | Min | Maks |
|---|---|---|---|---|
| **Tverrsnittsmodellen** | | | | |
| Datasett 1: | | | | |
| Gjennomsnittlig årlig inntektsvekst | 0.023 | 0.016 | -0.016 | 0.095 |
| BNP per innbygger 1995 | 17471 | 8552 | 2697 | 51182 |
| ln BNP per innbygger 1995 | 9.6 | 0.64 | 7.9 | 10.8 |
| Befolkningstetthet | 371 | 874 | 1.6 | 8210 |
| Befolkningstetthet (ln) | 7.18 | 0.86 | 3.23 | 9.29 |
| Andel høyere utd 2000 | 0.09 | 0.04 | 0.02 | 0.22 |
| Gjennomsnittlig årlig befolkningsvekst | 0.003 | 0.006 | 0.016 | 0.037 |
| | | | | |
| Datasett 2: | | | | |
| Gjennomsnittlig årlig inntektsvekst | 0.011 | 0.023 | -0.055 | 0.051 |
| BNP per innbygger 1980 | 39426 | 44265 | 2640 | 294151 |
| ln BNP per innbygger 1980 | 10.18 | 0.89 | 7.88 | 12.59 |
| Befolkningstetthet | 296 | 648 | 20 | 6236 |
| Befolkningstetthet (ln) | 5 | 1.04 | 3 | 8.74 |
| Andel høyere utd 2000 | 0.101 | 0.031 | 0.033 | 0.18 |
| Gjennomsnittlig årlig befolkningsvekst | 0.004 | 0.004 | -0.002 | 0.018 |
| **Paneldatamodellen** | | | | |
| 1 års intervaller: | | | | |
| Vekstraten til BNP per innbygger | 0.009 | 0.053 | -0.276 | 0.294 |
| BNP per innbygger(ln) | 10.3 | 0.92 | 7.84 | 13.1 |
| Befolkningstørrelse (ln) | 7.27 | 0.87 | 5.19 | 9.37 |
| Vekstraten til befolkningen | 0.004 | 0.008 | -0.113 | |
| | | | | |
| 3 års intervaller: | | | | |
| Vekstraten til BNP per innbygger | 0.04 | 0.12 | -0.578 | 0.322 |
| BNP per innbygger(ln) | 10.27 | 0.91 | 7.88 | 12.96 |
| Befolkningstørrelse (ln) | 7.26 | 0.87 | 5.20 | 9.34 |
| Vekstraten til befolkningen | 0.013 | 0.017 | -0.059 | 0.092 |
| | | | | |
| 5 års intervaller: | | | | |
| Vekstraten til BNP per innbygger | 0.046 | 0.180 | -0.723 | 0.478 |
| BNP per innbygger(ln) | 10.29 | 0.92 | 7.88 | 13 |
| Befolkningstørrelse (ln) | 7.26 | 0.87 | 5.20 | 9.35 |
| Vekstraten til befolkningen | 0.021 | 0.026 | -0.104 | 0.155 |
| Andelen høyere utdanning | 0.086 | 0.038 | 0.014 | 0.187 |
| Endring andel høyere utdanning | 0.015 | 0.014 | -0.01 | 0.07 |
| Andelen videregående og høyere utdanning | 0.239 | 0.114 | 0.034 | 0.512 |
| Endring andel videregående og høyere utdanning | 0.04 | 0.04 | -0.041 | 0.157 |

*Merk:* Deskriptiv statistikk for de to modellene. Datasett 1 er for perioden 1995 - 2008. Datasett 2 er for perioden 1980 - 2009. I paneldatamodellen gjelder kun datasett 2.

# 5 Økonometrisk spesifikasjon

I dette kapittelet skal jeg gjøre rede for hvilke metoder jeg bruker i analysen min. Jeg ønsker å undersøke veksten til BNP per innbygger og hvordan denne blir påvirket av befolkningsvekst. For å undersøke dette velger jeg å starte med en tverrsnittsmodell, som vil være den første delen av analysen min. En slik modell har vært mye brukt i tidligere vekstundersøkelser. Siden jeg har tilgang på et paneldatasett utvider jeg analysen med en paneldatamodell som jeg estimerer med faste effekter. Paneldatamodellen vil være den siste delen av analysen min.

## 5.1 Tverrsnittsmodellen

Jeg starter den empiriske analysen min ved å ta i bruk en tverrsnittsmodell. Denne modellen går ut på å studere regioner i Europa over en bestemt periode. En slik modell gir meg mulighet til å undersøke konvergensfart og hvordan befolkningsvekst påvirker BNP per innbygger. Konvergens blir betegnet som at fattige regioner vokser raskere enn rike regioner, en såkalt *catch-up* effekt. Dersom rike regioner vokser raskere kaller vi tilfellet divergens. Jeg kommer til å skille mellom to typer konvergens, henholdsvis ubetinget og betinget konvergens. Ubetinget konvergens er, ifølge Sala-i Martin (1996), at fattige økonomier vokser fortere enn rike økonomier, og alle økonomiene vokser mot et likt steady state nivå. For å undersøke ubetinget konvergens tar jeg utgangspunkt i modellen gitt ved denne likningen:

$$\bar{y}_i = b_0 + b_1 lny_i + u_i \tag{5.1}$$

I likning (5.1) er $\bar{y}_i$ gjennomsnittlig årlig vekst i BNP per innbygger for region i over hele perioden. $lny_i$ er initiellt BNP per innbygger for region i. Det initielle året vil være det første året i det respektive datasettet. Dette vil være 1995 for datasett 1 og 1980 for datasett 2. $u_i$ er restleddet som representerer andre uobserverbare faktorer utenom forklaringsvariablene som påvirker veksten. Dersom parameteren $b_1$ er negativ vil det være ubetinget konvergens, i motsetning til divergens dersom $b_1$ er positiv. For å finne konvergensfarten bruker jeg formelen: $CV = -ln(1 + b_1 T)/T$ hvor T er antall år i perioden. For å estimere parametrene i denne modellen bruker jeg en vanlig minste kvadraters metode (MKM). Som forklart i Wooldridge (2009) går MKM ut på å minimere summen av kvadrerte avvik. I Wooldridge er det en grundig gjennomgang av hvordan denne estimeringen foregår.[10] For at MKM skal gi forventningsrette og konsistente estimater må flere forutsetninger være oppfylt. Disse forutsetningene er også grundig forklart i Wooldridge. To av disse forutsetningne er at $E(u_i) = 0$ som sier at restleddet ikke er korrelert med

[10] Se Wooldridge side 22 - 67

forklaringsvariablene. $Var(u_i|lny_i) = \sigma^2$ vil si at det er ingen heteroskedastisitet, noe som fører til at variansen er lik for alle forklaringsvariablene.

Det er utfordringer med en slik modell som undersøker om det er ubetinget konvergens. I følge Sala-i Martin (1996) antar man i denne spesifikasjonen at den eneste forskjellen i verden er det initielle nivået på kapitalen. Dette er en streng antagelse fordi, det er grunn til å tro at økonomier vil ha en varierende grad av teknologi, sparerate og vekstrate etc. Dette vil føre til at økonomiene vokser mot en forskjellig steady state og likning (5.1) vil være feilaktig. Derfor har begrepet betinget konvergens blitt innført. Betinget konvergens blir definert som at fattige regioner vokser fortere enn rike regioner betinget på at andre variabler holdes konstant. Dette kan undersøkes på to måter: En metode er å begrense konvergensstudien til et sett med økonomier hvor steady state nivået er forholdsvis likt(noe som ikke er utenkelig for regioner i Europa). Den andre metoden er å innføre variabler som fungerer som en proxy for steady state. Dette gjør at en får justert for forskjellene mellom regionene. Derfor utvider jeg modellen med variablene befolkningstetthet, gjennomsnittlig befolkningsvekst og utdanning. Disse variablene vil jeg legge til en etter en for å kunne undersøke om de er signifikante og betydelige for modellen. Ved å legge til gjennomsnittlig befolkningsvekst gir det meg muligheten til å undersøke om problemstillingen om befolkningsveksten påvirker vekstraten til økonomien og konvergensfarten. Ved å inkludere disse variablene ender jeg opp med å estimere forskjellige spesifikasjoner av denne modellen:

$$\bar{y}_i = b_0 + b_1 lny_i + b_2 lnBeftetthet_i + b_3 Andutd_i + b_4 \bar{Bef}_i + u_i \tag{5.2}$$

I likning (5.2) har jeg utvidet modellen med tre variabler. $Beftetthet_i$ som er hvor mange innbyggere som bor per kvadratkilometer i regionen. Dette er initiell verdi som også her vil variere med hvilket datasett jeg bruker(1995 og 1980). $Andutd_i$ som andelen av befolkningen i region i som har høyere utdanning i 2000. $\bar{Bef}_i$ er gjennomsnittlig årlig befolkningsvekst for region i over hele perioden. I tillegg til å se på effekten av befolkningsvekst på inntektstveksten får jeg mulighet til å undersøke om befolkningsveksten påvirker konvergensen. Dersom den påvirker konvergensen vil dette være inkludert i $b_1$ estimatet i likning (5.1). Barro og Sala-i Martin (2004) og Braun (1993) har gjort noe lignende hvor de setter nettomigrasjon som en eksogen variabel. Ettersom datasett 1 inneholder størst tverrsnitt tar jeg utgangspunkt i dette datasettet i tverrsnittsmodellen. Jeg vil deretter sjekke om jeg får de samme resultatene ved å bruke datasett 2. Ved å gjøre det i denne rekkefølgen får jeg også en flytende overgang til paneldatamodellen hvor jeg bruker datasett 2. Det er fordi datasett 2 er mest effektiv å bruke i paneldatamodellen siden den har lengst tidsserie.

Tverrsnittsmodellen er en populær estimeringsmetode når det kommer til økonomisk vekst, men det finnes svakheter med en slik analyse. Her antar man at den aggregerte

produksjonsfunksjonen til hvert land er lik. Dette trenger ikke nødvendigvis å være tilfellet, og flere mener denne type estimeringsmetoder vil gi skjeve estimater. Islam (1995) argumenterer for at det regionspesifikke elementet i aggregert produksjon korrelerer med de inkluderte forklaringsvariablene. Dette kan føre til et utelatt variabelproblem. Han argumenterer videre for at med en paneldatatilnærming, vil det være mulig å korrigere for denne skjevheten. Ettersom jeg har tilgang på paneldata utvider jeg derfor analysen min med en paneldatamodell.

## 5.2 Paneldatamodellen

Jeg skal her bruke en paneldatamodell. Paneldatamodellen jeg har valgt å bruke er en dynamisk modell som estimerer hvilken fart en avhengig variabel vil ha tilbake til likevekt. En slik paneldatamodell blir også ofte kalt feilkorrigeringsmodell. Fordelen med denne type estimering er at jeg får muligheten til å bruke all variasjon i datasettet mitt. Som Canova (2001) argumenterer for, gir denne tilnærmingen mulighet til å estimere forskjellig konvergensfart til forskjellige steady state likevekter. Ved denne modell spesifiseringen får jeg mulighet til å undersøke effekten av konvergensfart, inntektsnivå, befolkningsnivå og befolkningsvekst på BNP per innbygger. Jeg tar utgangspunkt i denne paneldatamodellen:

$$\triangle\, lny_{it} = a_0 + a_1 lny_{it-1} + a_2 lnBef_{it-1} + a_3 \triangle\, lnBef_{it} + \eta_i + \gamma_t + e_{it} \tag{5.3}$$

I likning (5.3) er $\triangle lny_{it}$ vekstraten til BNP per innbygger for region gjennom perioden t. $lny_{it-1}$ er forrige periodes verdi på BNP per innbygger for region i. $lnBef_{it-1}$ er befolkningsstørrelsen til region i, den forrige perioden. $\triangle lnBef_{it}$ er vekstraten til befolkningen for region i gjennom perioden t. $e_{it}$ er restleddet som representerer andre uobserverbare faktorer. $\gamma_t$ er de periodespesifikke effektene i modellen. $\eta_i$ er de regionspesifikke effektene i modellen.

De periodespesifikke og regionspesifikke effektene føre til skjeve estimater i estimeringen. De regionspesifikke effektene kan representere forskjeller som for eksempel klima, institusjoner og geografi. Bosker (2007) mener det er gode grunner til å tro at disse regionspesifikke effektene kan representere systematiske forskjeller i steady state verdiene. Dersom disse blir utelatt fra estimeringen kan det skape et utelatt variabelproblem som igjen kan føre til skjevheter i estimatene. Jeg estimerer derfor denne modellen med faste effekter. Faste effekter er en estimeringsmetode som blir brukt på paneldatasett dersom det er fare for uobserverbar heterogenitet. Det vil si at regionene er forskjellige fra hverandre. Bosker (2007) argumenterer for at faste effekter er den mest hensiktsmessige metoden å bruke til fordel for tilfeldige effekter ved en slik spesifikasjon. Det er nettopp på grunn av at de uobserverbare regionspesifikke effektene med stor sannsynlighet er korrelert med variablene i modellen. De regionspesifikke effektene får man transformert bort ved faste

effekter estimering. Ulempen med transformasjonen er at vi kun får utnyttet variasjonen over tid ettersom all regionspesifikk variasjon er transformert bort. Analytisk utledning av faste effekter er nøye gjort i læreboka Wooldridge (2009).[11].

Det kan også være problemer med uobserverbare periodespesifikke effekter. Dette kan for eksempel være makroøkonomiske sjokk som ikke blir plukket opp og kan føre til skjeve estimater. Løsningen på dette er å inkludere periodedummier for hver periode som vil plukke opp den periodespesifikke effekten. Ved å gjøre dette vil all variasjon over tid også bli transformert bort. Da vil jeg eventuelt sitte igjen med variasjonen som kun er over tid og region(dvs. ingen regionspesifikke eller periodespesifikke effekter). Jeg velger derfor å estimere to versjoner av modellen: En uten periodedummier og en med periodedummier.

Ved hjelp av denne paneldatamodellen kan jeg undersøke effekten av inntektsnivået forrige periode på vekstraten til BNP per innbygger som vil være gitt ved $a_1$. $a_1$ blir også brukt til å beregne den betingede konvergensfarten. Parametrene $a_2$ og $a_3$ gir meg muligheten til å undersøke hvordan befolkningsnivået og vekstraten til befolkningen påvirker vekstraten til BNP per innbygger. Til slutt kan jeg se på effekten av befolkningsnivået på BNP per innbygger. Dette er mulig fordi på lang sikt vil alle variablene ha konvergert mot sine langsiktige løsninger. Dette fører til at $lny_{it} = lny_{it-1}$, som vil si at $\triangle lny_{it} = 0$, hvor det samme gjelder for $\triangle lnBef_{it}$. Ettersom de har nådd sine langsiktige løsninger kan jeg fjerne tidsperspektivet og derivere med hensyn på befolkningen:

$$\begin{aligned} 0 &= a_0 + a_1 lny + a_2 lnbef \\ \Rightarrow \quad -a_1 lny &= a_0 + a_2 lnbef \\ \Rightarrow \quad \frac{\partial lny}{\partial lnBef} &= -\frac{a_2}{a_1} \end{aligned} \tag{5.4}$$

Det vises i likning (5.4) at langtidseffekten av befolkningsnivå på BNP per innbygger er $-a_2/a_1$. Siden begge disse variablene er oppgitt på ln form kan disse betraktes som den langsiktige elastisiteten til BNP per innbygger med hensyn på befolkningsnivå.

En slik modell bør ha en så lang tidsserie som mulig. Derfor bruker jeg kun datasett 2 i denne modellen. Datasett 2 er en lang tidsserie som gir meg mulighet til å dele den opp i intervaller og fortsatt sitte igjen med relativt mange observasjoner. Derfor vil ikke datasett 1 fungere like bra. Jeg har derfor valgt å estimere tre forskjellige kombinasjoner av modellen. En hvor datasettet består av årlige intervaller som gjør at jeg ender opp med 29 perioder. Deretter har jeg delt det opp slik at det blir beregnet med 3 års intervaller som gjør at jeg ender opp med 9 perioder. Til slutt har jeg gjort det samme for 5 års intervaller som gjør at jeg ender opp med 6 observasjoner per gruppe. Jeg brukte da 2009

[11] På sidene 481 - 484 er det gitt en grundig forklaring av estimeringsmetoden faste effekter.

som siste år for å få flest mulig perioder. Siden jeg deler opp datasettet slik og bruker en paneldatamodell må jeg modifisere konvergensformelen til $CV = -ln(1 + a_1)/k$ hvor k er lengden på intervallet(1, 3 og 5). Disse intervallene har hver sine fordeler og ulemper. Fordelen med årlige intervaller er at jeg får mange observasjoner, noe som fører til mer presise estimater. Ulempen er som Bosker (2007) begrunner i artikkelen sin, at ved bruk av årlige intervaller kan mye av veksten mellom periodene være som følge av konjunktursykluser. Bosker mener at dersom man bruker 5 års intervaller istedenfor årlige, vil ikke disse konjunktursyklusene påvirke estimatene for den langsiktige vekstraten til økonomien. Bosker mener 5 års intervaller vil også gjøre et eventuelt seriekorrelasjonsproblem mindre sannsynlig. Seriekorrelasjon er når restleddet i to forskjellige perioder korrelerer med hverandre. Dette vil si at dersom en overestimerer effekten av befolkning i forrige periode vil den overestimere effekten i fremtidige perioder. Konsekvensene av seriekorrelasjon er at standardavvikene blir mindre. Dette kan føre til at en går ut fra at estimatene er mer presise enn de egentlig er. Derfor har jeg valgt å estimere modellen med årlige, 3 års og 5 års intervaller. En annen fordel med 5 års intervaller, er at her kan jeg bruke utdanningsdataene mine og kontrollere for utdanning. Så når jeg estimerer modellen med 5 års intervaller utvider jeg modellen til:

$$\triangle lny_{it} = a_0 + a_1 lny_{it-1} + a_2 lnBef_{it-1} + a_3 \triangle lnBef_{it} + a_4 Utd_{it-1} + a_5 \triangle Utd_{it} + \eta_i + \gamma_t + e_{it} \tag{5.5}$$

I likning (5.5) har jeg utvidet modellen gitt ved (5.3) med utdanning. $Utd_{it-1}$ er andelen av befolkningen i region i med høyere utdanning forrige periode. $\triangle Utd_{it}$ er endringen i andelen som har høyere utdanning for region i periode t. Utdanningstallene er basert på nasjonale data. Jeg velger også å kjøre samme modell hvor utdanningsnivået er for videregående og høyere utdanning. Modellen gitt ved likning (5.5) vil også bli estimert med og uten periodedummier.

## 5.3 Økonometriske utfordringer

Estimering av disse to modellene er ikke uproblematisk. Jeg starter med å se på tverrsnittsmodellen. I likning (5.2) kan det være et mulig endogenitetsproblem. Et endogenitetsproblem oppstår når de uavhengige variablene er korrelert med restleddet. Det er flere årsaker som kan gjøre at dette problemet oppstår. De mest tenkelige årsakene i modell (5.2) er utelatt variabelproblem og simultanitet. Et utelatt variabel problem er at jeg kan ha utelatt relevante forklaringsvariabler fra estimeringen, noe som kan føre til skjevhet i estimatene. Det er dette som blir sett på som det største problemet ifølge Islam (1995) som jeg argumenterte for tidligere. Simultanitet oppstår når en av forklaringsvariablene i modellen er simultant bestemt med den avhengige variabelen. Det vil si at de to variablene

påvirker hverandre. Dette er ikke et utenkelig problem i tverrsnittsmodellen min. Simultanitetsproblemet kan oppstå med variablene utdanning og befolkningsvekst. Utdanning kan være påvirket av veksten i en region. Det er fordi økt inntektsvekst kan være med å stimulere til økt utdanning. Disse kan dermed være simultant bestemt fordi utdanning påvirker inntektsveksten og inntektsveksten påvirker andelen som tar utdanning. Dersom inntektsveksten påvirker andelen som tar høyere utdanning kan dette føre til skjeve estimater. Dette problemet oppstår på grunn av at andelen som har høyere utdanning er gitt i 2000. Denne andelen kan dermed være påvirket av tidligere inntektsvekst i regionene og simultanitetsproblemet oppstår. Dette gjelder for begge datasettene men sannsynligheten er større for at dette kan skje i datasett 2. Hadde andelen som tar utdanning vært gitt i det første året i datasettet ville problemet vært mindre relevant. Det samme problemet kan gjelde for gjennomsnittlig befolkningsvekst. Det er på grunn av en region som opplever sterk inntektsvekst kan være en stimuli til økt migrasjon til denne regionen, siden flere ønsker å ta del i denne inntektsveksten. Dette fører til at inntektsveksten kan være med på å påvirke befolkningsveksten gjennom migrasjon. Siden målet på befolkningsveksten er gjennom hele perioden, kan disse være simultant bestemt ved at befolkningsvekst og BNP vekst påvirker hverandre. En løsning på simultanitetsproblemet ville vært å instrumentere disse to variablene, men slike instrumenter er vanskelig å finne. Et mulig instrument for gjennomsnittlig befolkningsvekst kunne vært å bruke historiske befolkningsdata, noe jeg ikke har tilgang til.

I paneldatamodellen vil også simultanitetsproblemet oppstå. En fordel her er at ved å bruke faste effekter og eventuelle periodedummier får jeg transformert bort de regionspesifikke og periodespesifikke effektene som kan være med på å føre til et utelatt variabelproblem. Jeg slipper dermed dette problemet her. Befolkningsveksten i denne modellen er også over hele perioden, slik at også her vil det kunne oppstå et simultanitetsproblem som jeg argumenterte for i tverrsnittsmodellen. Det samme vil gjelde for utdanning. For utdanning oppstår også et annet problem. Utdanningsdataene jeg har er kun på nasjonalt nivå. Dette fører til at alle regionene i samme land har et likt utdanningsnivå. Siden det også er færre land med i dette datasettet kan dette føre til variasjonen mellom de forskjellige regionene blir for små. Optimalt sett skulle jeg hatt regionale data for hele perioden. Disse utfordringene kan det være mulig å se på for fremtidig forskning.

## 5.4 Oppsummering

Analysedelen min består av en tverrsnittsmodell og en paneldatamodell. Jeg starter med tverrsnittsmodellen hvor jeg tar utgangspunkt i en enkel modell hvor jeg får undersøkt ubetinget konvergens. Denne modellen utvider jeg med variablene utdanning, befolkningstetthet og gjennomsnittlig befolkningsvekst. Ved å utvide denne modellen får jeg undersøkt betinget konvergens og hvordan befolkningsvekst påvirker veksten i BNP per innbyg-

ger. For å få en mer komplett analyse tar jeg i del to i bruk en paneldatamodell som jeg estimerer med faste effekter. I denne modellen tar jeg i bruk forskjellige kombinasjoner ved å dele datasettet inn i årlige, 3 års og 5 års intervaller siden disse har forskjellige egenskaper. For 5 års intervaller kontrollerer jeg også for utdanning.

En slik analyse som jeg gjør her har økonometriske utfordringer. Det er mulige endogenitetsproblemer forbundet med analysen. Dette er et vanskelig problem å løse empirisk med de dataene som er tilgjengelig. Jeg har derfor brukt forskjellige spesifikasjoner av modellene for å løse problemet på best mulig måte.

# 6 Resultater

I dette kapittelet presenteres resultatene av modellene jeg har estimert. Analysen min består av en tverrsnittsmodell og en paneldatamodell. Ved hjelp av disse vil jeg forsøke å gi svar på om inntektsforskjellene i Europa blir mindre over tid og om befolkningsvekst har en positiv eller negativ innvirkning på veksten til BNP per innbygger. Først presenterer jeg tverrsnittsmodellen før jeg utvider analysen med en paneldatamodell og sammenligner resultatene.

## 6.1 Tverrsnittsmodellen

I dette delkapittelet vil jeg se på resultatene av tverrsnittsmodellen. Jeg starter med datasett 1 ettersom den har størst tverrsnitt og egner seg bra for en slik analyse. Deretter estimerer jeg datasett 2 på samme måte. Datasett 2 gir jeg meg muligheten til å undersøke om jeg får samme effekter ved en lengre tidsserie og færre regioner. Ved å estimere modellen med to forskjellige datasett får jeg muligheten til å sjekke om jeg får samme effekter med begge to.

Jeg starter med å undersøke om det er ubetinget konvergens, og tar utgangspunkt i modellen gitt ved likning (5.1). Resultatene fra denne regresjonen kan man se i tabell 3, kolonne 1. Ved denne estimeringen får jeg her en parameterverdi på -0.012 som er klart signifikant. Denne har et negativt fortegn som vil si at det er ubetinget konvergens. Tolkningen av denne parameteren er, at dersom initiell inntekt øker med 10 prosent, vil gjennomsnittlig årlig vekstrate synke med 0.12 prosentpoeng. Konvergensfarten beregner jeg ved hjelp av formelen forklart i 5.1.[12] Den ubetingede konvergensen er på 1.33 prosent med en tilsvarende *half life* på 53.32 år. *Half life* er at i løpet av 53 år vil forskjellene i inntekt mellom regionene være halvert, heretter kaller jeg denne hl.[13] En konvergensfart på 1.33 prosent vil si at europeiske regioner konvergerer med 1.33 prosent årlig. Dette er forholdsvis lav konvergensfart, og den er litt lavere enn det Barro og Sala-i-Martin har funnet med sine studier. Sala-i Martin (1996) fant for regioner i Europa en ubetinget konvergensfart på 1.5 prosent. Sala-i Martin (1996) fant for stater i USA en ubetinget konvergensfart på 2 prosent. $R^2$ er definert av Wooldridge (2009) som hvor mye av den totale variasjonen i den avhengige variabelen som blir forklart av forklaringsvariablene. Denne modellen har en $R^2$ på 0.225, noe som vil si at nesten en fjerdedel av variasjonen i modellen er forklart. Det kan være grunn til å tro at det er flere variabler som påvirker gjennomsnittlig vekst i BNP per innbygger, og at steady state nivået til regionene er forskjellig. Spesielt ettersom regioner i øst- og vest-Europa er inkludert i datasett 1, og det er tenkelig at disse har et forskjellig steady state nivå.

[12] Konvergenseformelen som er $CV = -ln(1 + b_1T)/T$

[13] $hl = ln(2)/CV$

For å undersøke for betinget konvergens utvider jeg modellen med variablene befolkningstetthet, utdanning og gjennomsnittlig befolkningsvekst. Jeg estimerer dermed forskjellige spesifikasjoner av modellen gitt ved likning (5.2). Starter med å inkludere befolkningstetthet og utdanning hver for seg. Ved å gjøre dette får jeg noe nedgang i den initelle verdien på BNP per innbygger, som vist ved kolonne (2) og (3) i tabell 3. Variablene er også klart signifikante på 1 prosentnivå. Jeg inkluderer dermed begge disse i samme modell og resultatet av dette er gitt i kolonne (4). Koeffisienten til befolkningstettheten er på 0.001. Dermed vil en økning i den initielle befolkningstettheten på 10 prosent føre til en økning i vekstraten til BNP per innbygger på 0.01 prosentpoeng. En positiv effekt av befolkningstetthet på inntektsveksten kan tyde på positive agglomerasjonseffekter. Dette står i sterk kontrast med Bosker (2007) som fikk negativ effekt av agglomerasjon på inntektsveksten i sin paneldatamodell. Effekten av utdanning er også positiv og har en koeffisient på 0.102. Dersom andelen som har høyere utdanning øker med 10 prosentpoeng øker vekstraten til BNP per innbygger med ca 1 prosentpoeng. Både befolkningstetthet og utdanning er fortsatt signifikante ved henholdsvis 5 og 1 prosentnivå. Den betingede konvergensfarten er på 1.8 prosent med tilsvarende hl på 38.5 år. Når jeg nå undersøkte for betinget konvergens har konvergensfarten økt. Sammenlignet med tidligere forskning er dette tilnærmet likt resultatene i andre studier. Braun (1993) får en betinget konvergens på 1.97 prosent før han inkluderte nettomigrasjon. Sala-i Martin (1996) fikk en betinget konvergens på 1.5 prosent. Konvergensfarten til begge disse var for europeiske regioner. Den betingede konvergensfarten jeg finner er omtrent midt imellom disse to. Jeg utvider nå modellen med gjennomsnittlig befolkningsvekst som gir meg muligheten til å se hvordan denne påvirker inntektsveksten og konvergensfarten. Dette er også gjort av Barro og Sala-i Martin (2004), og Braun (1993). Forskjellen er at de bruker nettomigrasjon. Nettomigrasjon er en viktig faktor i befolkningsveksten til en region og har dermed likhetstrekk med analysen min.

I tabell 3, kolonne (5) er likning (5.2) estimert med alle variablene. Ser fra tabellen at gjennomsnittlig befolkningsvekst kun er signifikant ved et 10 prosent signifikansnivå. Den er dermed ikke like signifikant som de andre variablene i modellen. Effekten av gjennomsnittlig befolkningsvekst på gjennomsnittlig vekst i BNP per innbygger er positiv og på 0.25. Dette vil si at dersom befolkningsveksten øker med 10 prosentpoeng vil gjennomsnittlig årlig vekstrate til BNP per innbygger øke med 2.5 prosentpoeng. Dette er en positiv og betydningsfull effekt og tyder på at befolkningsveksten til en region bidrar til inntektsvekst. Ved inkludering av gjennomsnittlig befolkningsvekst har parameteren til initiell BNP per innbygger blitt redusert ytterligere til -0.017. Dersom initiellt BNP øker med 10 prosent vil det føre til en nedgang i veksten på 0.17 prosentpoeng. Konvergensfarten er nå på 1.9 prosent. Denne konvergensfarten er tilnærmet lik konvergensfarten jeg fikk i tilfellet uten gjennomsnittlig befolkningsvekst. Gjennomsnittlig befolkningsvekst påvirket dermed konvergensfarten i liten grad. Sammenlignet med tidligere forskning er den rimelig

Tabell 3: Tverrsnittsmodell hvor jeg bruker datasett 1

| | MKM | | | | |
|---|---|---|---|---|---|
| | Avhengig variabel er $\bar{y}_i$ | | | | |
| | Sett 1 | Sett 1 | Sett 1 | Sett 1 | Sett 1 |
| Variabler | (1) | (2) | (3) | (4) | (5) |
| $lny_{i,1995}$ | -0.012*** | -0.013*** | -0.015*** | -0.016*** | -0.017*** |
| | (-9.34) | (-9.93) | (-10.99) | (-11.30) | (-11.08) |
| $lnBeftetthet_{i,1995}$ | | 0.002*** | | 0.001** | 0.002** |
| | | (3,01) | | (2.24) | (2.37) |
| $Andutd_{i,2000}$ | | | 0.111*** | 0.102*** | 0.103*** |
| | | | (5.07) | (4.62) | (4.65) |
| $\bar{Bef}_i$ | | | | | 0.25* |
| | | | | | (1.76) |
| $Konstant$ | 0.140*** | 0.139*** | 0.158*** | 0.156*** | 0.165*** |
| | (11.14) | (11.25) | (12.62) | (12.52) | (12.29) |
| Oservasjoner | 264 | 264 | 264 | 264 | 264 |
| $R^2$ | 0.250 | 0.275 | 0.317 | 0.330 | 0.338 |

*Merk:* t-verdier er oppgitt i parentes. Sett 1 er datasettet 1995 - 2008. Dersom *** er den signifikant på 1 % nivå, ** signifikant på 5 % nivå og ved * er den signifikant på 10% nivå.

lik Braun (1993) når han inkluderte nettomigrasjon. Braun fikk en konvergensfart på 1.88 prosent for europeiske regioner. Denne konvergensfarten er nesten identisk med min konvergensfart. For effekten av nettomigrasjon fikk Braun en verdi på 0.023 og usignifikant, noe som er betydelig mindre enn effekten jeg fikk av gjennomsnittlig befolkningsvekst i min modell. Dette kan tyde på at effekten av befolkningsvekst er større. Det er også viktig å bemerke seg at datasettet til Braun er forholdsvis langt tilbake i tid og besto kun av de største landene i Europa. Dette datasettet ligner dermed mer på datasett 2. Ser fra $R^2$ at forklaringskraften har økt med en god del til 0.338 noe som er naturlig ettersom flere variabler er blitt til inkludert i modellen. Dette vil si at rundt en tredjedel av variasjonen i modellen er forklart.

Resultatene fra tverrsnittsmodellen med datasett 1 tyder på inntektskonvergens blant regionene i Europa. Dette fører til at inntektsforskjellene mellom regionene vil bli mindre over tid. Jeg fikk også en signifikant og positiv effekt av gjennomsnittlig befolkningsvekst på gjennomsnittlig inntektsvekst. Dette tyder på at dersom en region opplever økt befolkningsvekst vil dette føre til at inntektsveksten i regionen vil øke. Jeg kan nå sjekke om jeg får samme resultater ved å bruke datasett 2. Her er tidsserien lenger og det færre regioner med, ettersom jeg kun ser på EU-12.

Jeg tar igjen utgangspunkt i modellen gitt ved likning (5.1) og starter med å undersøke ubetinget konvergens, men nå med datasett 2. Resultatet fra denne estimeringen kan man se i tabell 4, kolonne 1. Nå vil en økning i initiell inntekt på 10 prosent føre til en reduksjon i den gjennomsnittlige årlige vekstraten til BNP per innbygger på 0.07 prosentpoeng. Denne effekten er mindre enn i datasett 1 som fikk en reduksjon på 0.12 prosentpoeng. Den ubetingede konvergensfarten er på 0.78 prosent. Denne konvergensfarten er en god del lavere enn i datasett 1. Dette tilsvarer en hl på 87.7 år, noe som er en lang tid. Denne

konvergensfarten er lav i forhold til tidligere slike studier. Sala-i Martin (1996) fant en ubetinget konvergensfart på 1.5 prosent. Forklaringskraften til modellen er på lave 0.088, noe som ikke er spesielt høyt.

For å undersøke for betinget konvergens gjør jeg som ved datasett 1 og utvider modellen med befolkningstetthet, utdanning og gjennomsnittlig befolkningsvekst. Jeg estimerer dermed forskjellig spesifikasjoner av modellen gitt ved likning (5.2). Jeg starter med å inkludere befolkningstetthet og utdanning hver for seg. Både befolkningstetthet og utdanning er klart signifikant ved 1 prosentnivå. Jeg inkluderer dermed begge disse samtidig og får modellen gitt ved kolonne (4). Ved denne estimeringen er koeffisienten til befolkningstettheten nå på 0.008. Nå vil en 10 prosent økning i befolkningstettheten føre til en økning i vekstraten til BNP per innbygger på 0.08 prosentpoeng. Effekten av befolkningstettheten er dermed en god del høyere enn effekten ved datasett 1. Dette tyder på en høyere agglomerasjonseffekt i datasett 2 enn i datasett 1. Effekten av høyere utdanning har nå en koeffisient på 0.229. Dersom andelen høyere utdanning øker med 10 prosentpoeng fører dette til en økning i veksten til BNP per innbygger på 2.29 prosentpoeng som også er en god del høyere enn ved datasett 1. Denne effekten bør man være litt forsiktig med å legge alt for mye vekt på ettersom den initielle verdien er gitt i 2000. Både befolkningstetthet og høyere utdanning er signifikant ved 1 prosentnivå. Initiell verdi på BNP per innbygger har nå sunket til -0.01 slik at 10 prosent økning i BNP per innbygger i 1980 fører til 0.1 prosentpoengs reduksjon i vekstraten til BNP per innbygger. Betinget konvergensfart er på 1.19 prosent. Denne konvergensen er en god del lavere enn ved datasett 1. Konvergensen er også lavere enn resultatet til Braun (1993) som fikk en konvergensfart på 1.5 prosent før han inkluderte nettomigrasjon. Denne spesifikasjonen har også fått en solid økning i forklaringskraften, slik at $R^2 = 0.456$ som vil si at nesten halve variasjonen i modellen er forklart.

Jeg inkluderer nå gjennomsnittlig befolkningsvekst og estimerer modellen gitt ved likning (5.2). Resultatene av denne estimeringen er gitt i tabell 4, kolonne (5). I motsetning til ved datasett 1 er gjennomsnittlig befolkningsvekst kun signifikant ved et 20 prosentnivå. Koeffisienten til gjennomsnittlig befolkningsvekst har i denne modellen økt til 0.742, noe som er en markant økning i forhold til datasett 1. Nå vil en økning på 10 prosentpoeng i gjennomsnittlig befolkningsvekst øke vekstraten til BNP per innbygger med 7.42 prosentpoeng. Dette er mer enn en dobling i forhold til effekten gitt ved datasett 1. Vi ser at befolkningsveksten ikke påvirker konvergensen til regionene ettersom parameteren fortsatt er på -0.01. Den betingede konvergensfarten er dermed fortsatt på 1.19 prosent. Sammenlignet med tidligere forskning fikk Braun (1993) en usignifikant effekt av nettomigrasjon for regionene i Europa, og liten påvirkning på konvergensfarten. Konvergensfarten til Braun er på 1.88 prosent som er en god del høyere enn min konvergensfart i datasett 2.

Tabell 4: Tverrsnittsmodell hvor jeg bruker datasett 2

| | MKM | | | | |
|---|---|---|---|---|---|
| | Avhengig variabel er $\bar{y}_i$ | | | | |
| | Sett 2 | Sett 2 | sett 2 | Sett 2 | Sett 2 |
| Variabler | (1) | (2) | (3) | (4) | (5) |
| $lny_{i,1980}$ | -0.007*** | -0.010*** | -0.009*** | -0.010*** | -0.010*** |
| | (-3.25) | (-5.23) | (-4.58) | (-5.61) | (-5.62) |
| $lnBeftetthet_{i,1980}$ | | 0.012*** | | 0.008*** | 0.009*** |
| | | (7.28) | | (4.60) | (4.85) |
| $Andutd_{i,2000}$ | | | 0.377*** | 0.229*** | 0.21*** |
| | | | (6.63) | (3.73) | (3.45) |
| $\bar{Bef}_i$ | | | | | 0.742 |
| | | | | | (1.62) |
| $Konstant$ | 0.088*** | 0.053*** | 0.065*** | 0.050*** | 0.045*** |
| | (3.69) | (2.66) | (3.19) | (2.63) | (2.37) |
| Oservasjoner | 112 | 112 | 112 | 112 | 112 |
| $R^2$ | 0.088 | 0.386 | 0.350 | 0.456 | 0.469 |

*Merk:* t-verdier er oppgitt i parentes. Sett 2 er datasettet 1980 - 2009. Dersom *** er den signifikant på 1 % nivå, ** signifikant på 5 % nivå og ved * er den signifikant på 10% nivå.

Sammenligner en disse datasettene er det viktig å tenke på de to store forskjellene mellom datasettene. Den ene er lengden på tidsserien og den andre er hvilke regioner som er med. Når det kommer til resultatene fant begge disse datasettene både ubetinget og betinget konvergens. Dette tyder på at Europa opplever konvergens mellom regioner. Når det kommer til farten på denne konvergensen lå den betingede konvergensen på rundt 2 prosent og litt mindre for ubetinget konvergens. Den var generelt sett lavere for datasett 2 i forhold til datasett 1. Det kan dermed virke som at konvergensen fungerer hurtigere blant de nye landene som har kommet inn i EU senere tid. En annen forklaring kan være at forholdene ligger mer til rette for konvergens senere i tidsperioden og dette kommer mer fram i datasett 1 som kun er 1995 - 2008. Når det gjelder effekten av befolkningsvekst er den også positiv i begge datasettene. Dette tyder på, ifølge tverrsnittsmodellen, at økt befolkningsvekst i en region fører til inntektsvekst i denne regionen. Effekten virket å være høyere i datasett 2, men her var ikke denne effekten like signifikant. Det virker også å være fordeler ved at mennesker klynger seg sammen, noe som kan tyde på positive agglomerasjonseffekter. Når jeg nå utvider analysen med en paneldatamodell kan jeg undersøke om jeg får samme resultater for konvergensfart og effekt av befolkningsvekst.

## 6.2 Paneldatamodellen

I denne delen ser jeg nærmere på resultatene fra estimeringen av paneldatamodellen. Her får jeg anledning til å undersøke både konvergensfart og effekten av befolkningsvekst og befolkningsnivå på vekstraten til BNP per innbygger. Jeg har tatt utgangspunkt i datasett 2, som er den lengste tidsserien, og estimert denne med faste effekter med og uten periodedummier. Dataene er delt opp i årlige, 3 års og 5 års intervaller. Ved alle intervallene fører

inkludering av periodedummier til at effekten av befolkningsnivå på vekstraten til BNP per innbygger blir kraftig redusert. Effekten av befolkningsvekst på vekstraten til BNP per innbygger varierer også en god del med inkludering av periodedummier. Dette kan tyde på at periodedummiene i modellen fanger opp en periodespesifikk trend som gjelder befolkningen. Parametrene i modellen uten periodedummier kan dermed være forbundet med skjeve estimater. Jeg velger derfor å legge mest fokus på estimatene med periodedummier. Jeg får ikke mulighet til å undersøke den ubetingede konvergensfarten i denne paneldatamodellen, slik at konvergensfarten jeg undersøker er betinget.

Jeg tar utgangspunkt i modellen gitt ved likning (5.3) og starter med å estimere modellen med årlige intervaller på dataene. Resultatene av denne estimeringen er gitt i tabell 5, kolonne (1) og (2). Ved årlige data vil en økning i inntektsnivået forrige periode på 10 prosent føre til en reduksjon i vekstraten til BNP per innbygger på 0.92 prosentpoeng. Jeg bruker nå den modifiserte konvergensformelen og finner en konvergensfart på 9.65 prosent.[14] Denne konvergensfarten er atskillig høyere enn ved tverrsnittsmodellen. Sammenlignet med tidligere studier på paneldata får jeg en konvergensfart som er ganske lik Caselli et al. (1996) som fikk en konvergensfart på 10 prosent ved bruk av paneldatamodell og GMM estimering for land. Den er litt lenger unna Badinger et al. (2002) som fikk en konvergensfart med paneldatamodell og GMM estimering på 7 prosent med korrigering for geografisk beliggenhet. Det samme gjelder for Bosker (2007) som fikk en konvergensfart på 12.7 prosent med bruk av faste effekter og periodedummier for regioner i Europa. En konvergensfart på 9.65 prosent vil ha en hl på litt over 7 år noe som er nokså hurtig. Som jeg argumenterte for i kapittel 5.2, kan det være at denne fanger opp konjunktursykluser og konvergensfarten blir dermed påvirket av dette. Jeg forsøker å løse dette problemet med å estimere samme modell med data for 3 års intervaller. Når jeg estimerer modellen med data for 3 års intervaller ser man at effekten av initiellt inntektsnivå forrige periode på vekstraten reduseres til -0.261 med periodedummier. Det er naturlig at disse er betraktelig lavere fordi effekten vil spille seg over en treårs periode. Dersom nivået på BNP per innbygger forrige periode nå øker med 10 prosent vil dette føre til en reduksjon i vekstraten til BNP per innbygger på 2.61 prosentpoeng gjennom perioden. Dette vil tilsvare en årlig effekt på 0.87 prosentpoeng som er forholdsvis likt resultatet med årlige intervaller. Konvergensfarten vil nå være på 10.1 prosent som er noe høyere enn ved årlige intervaller. Sammenlignet med tidligere studier er den fortsatt forholdsvis lik de jeg kommenterte tidligere. Denne konvergensfarten vil ha en hl på nesten 7 år. Både med årlige og 3 års intervaller er effekten av inntektsnivået forrige periode på vekstraten til BNP per innbygger klart signifikant på 1 prosentnivå.

Ved årlige intervaller er koeffisienten til vekstraten til befolkningen på 0.179 med periodedummier. Effekten av en økning i vekstraten til befolkning på 10 prosentpoeng vil føre til en økning i vekstraten til BNP per innbygger på 1.79 prosentpoeng. Dersom en ser på

[14] Konvergensformelen som nå er gitt ved $CV = -ln(1 + a_1)/k$ hvor $k = 1$. Senere er $k = 3$ og $k = 5$.

3 års intervaller vil effekten av en økning på 10 prosentpoeng i vekstraten til befolkningen føre til en økning på 4.85 prosentpoeng i vekstraten til BNP per innbygger over perioden. Denne effekten er en god del høyere enn den var ved årlige. Disse effektene er signifikant ved henholdsvis 5 og 1 prosentnivå og viser en positiv effekt av en økning i befolknings-veksten på inntektsveksten. Denne effekten er motstridene i forhold til hva Bosker (2007) fikk når han estimerte sin modell med faste effekter og periodedummier for regioner i Europa.

Tabell 5: Paneldatamodell med årlige og 3 års intervaller

| | Faste effekter | | | |
|---|---|---|---|---|
| | $\triangle lny_{it}$ er forklaringsvariabelen | | | |
| | 1 år | 1 år | 3 år | 3 år |
| Variabler | (1) | (2) | (3) | (4) |
| $lny_{it-1}$ | -0.090*** | -0.092*** | -0.254*** | -0.261*** |
| | (-26.01) | (-32.60) | (-24.06) | (-31.56) |
| $lnBef_{it-1}$ | 0.194*** | 0.083*** | 0.924*** | 0.221*** |
| | (11.40) | (4.52) | (15.69) | (3.41) |
| $\triangle lnBef_{it}$ | .321*** | 0.179** | 0.626*** | 0.485*** |
| | (2.98) | (2,06) | (3.43) | (3.31) |
| *Konstant* | -0.476*** | 0.340*** | -4.071*** | 1.187** |
| | (-4.26) | (2.6) | (-10.23) | (2.55) |
| *Periodedummier* | Nei | Ja | Nei | Ja |
| Oservasjoner | 3248 | 3248 | 1008 | 1008 |
| $R^2$ | 0.178 | 0.498 | 0.413 | 0.660 |

*Merk:* Her er (1) og (2) årlige intervaller, mens (3) og (4) er 3 års intervaller. I parentes er t-verdier oppgitt. Dersom *** er den signifikant på 1 % nivå, ** signifikant på 5 % nivå og ved * er den signifikant på 1% nivå. Dersom det står ja ved periodedummier er disse inkludert. Jeg rapporterer ikke periodedummiene.

I denne modellen får jeg også muligheten til å undersøke effekten av befolkningsnivå på BNP per innbygger. Med årlige intervaller vil en økning på 10 prosent i befolkningsnivå forrige periode føre til økning i vekstraten til BNP per innbygger på 0.83 prosentpoeng. Sammenlignet med 3 års intervaller vil en økning på 10 prosent forrige periode føre til en økning i vekstraten til BNP per innbygger på 2.21 prosentpoeng over perioden. Dette vil tilsvare en årlig økning på 0.74 prosentpoeng som er litt lavere enn tilfellet med årlige intervaller. Begge viser positive effekter av befolkningsnivå på vekstraten til BNP per innbygger. Jeg kan nå se på den langsiktige effekten av befolkningsnivå på BNP per innbygger.[15] For årlige data vil denne være på henholdsvis 2.15 og 0.9 uten og med periodedummier. Disse kan tolkes som den langsiktige elastisiteten av BNP per innbygger med hensyn på befolkningsnivå. Elastisiteten er en prosentvis endring i befolkningsnivå som fører til en prosentvis endring i BNP per innbygger. Slik at effekten med periodedummier av en økning i befolkningsnivå på 10 prosent vil føre til en økning i BNP per innbygger på 9 prosent. Dersom modellen estimeres med 3 års intervaller er elastisiteten til BNP per innbygger med hensyn på befolkningsnivå på 3.64 uten periodedummier 0.84 med periodedummier. Elastisiteten med periodedummier er tilnærmet lik som for årlige data.

[15] Som jeg forklarte i kapittel 5.2 er denne $-a_3/a_2$

Sammenligner man effektene i resultatene ved årlige og 3 års intervaller er resultatene tilnærmet like, bortsett fra effekten av vekstraten til befolkning på vekstraten til BNP per innbygger. Dette tyder på at resultatene jeg har fått er robuste. Jeg ser nå videre på hvordan resultatene blir påvirket når jeg estimerer modellen med 5 års intervaller. En fordel med 5 års intervaller er at jeg også får muligheten til kontrollere for utdanning.

Når jeg skal estimere modellen med data for 5 års intervaller tar jeg først utgangspunkt likning (5.3) og estimerer denne med og uten periodedummier. Resultatene av denne estimeringen er gitt i tabell 6, kolonne (1) og (2). Nå vil en økning på 10 prosent i inntektsnivået forrige periode føre til en reduksjon i vekstraten til BNP per innbygger på 4.17 prosentpoeng over perioden. Dette vil tilsvare en årlig reduksjon på 0.833 prosentpoeng som er tilnærmet likt tilfellene ved årlige og 3 års intervaller. Konvergensen vil her være på 10.79 prosent. Denne har økt litt i forhold til årlige og 3 års intervaller, men er i nærheten av forskningen som jeg argumenterte for tidligere. En konvergensfart på 10.79 prosent vil ha en hl på 6.42 år.

Ved å anvende data med 5 års intervaller vil en økning på 10 prosentpoeng i vekstraten til befolkningen nå føre til en økning i vekstraten til BNP per innbygger på 4.96 prosentpoeng. Denne effekten er tilnærmet lik effekten for 3 års intervaller. Siden begge disse er positive og tilnærmet like bygger det opp under resultatet om at befolkningsvekst i en region har en positiv effekt på veksten i BNP per innbygger. Ser jeg på effekten av befolkningsnivå vil en økning i befolkningsnivået forrige periode på 10 prosent føre til en økning i vekstraten til BNP per innbygger på 3.63 prosentpoeng over perioden. Dette vil tilsvare en årlig inntektsvekst på 0.726 prosentpoeng som er tilnærmet likt estimatene for årlige og 3 års intervaller. Elastisiteten til BNP per innbygger med hensyn på befolkningsnivå er nå på 2.05 uten periodedummier og 0.87 med periodedummier. Elastisiteten med periodedummier er tilnærmet likt årlige og 3 års intervaller.

Jeg ser at ved å gjøre samme estimering med 5 års intervaller er resultatet fortsatt tilnærmet likt resultatene for årlige og 3 års intervaller. Effekten av vekstraten til befolkningen på vekstraten til BNP per innbygger var her veldig lik effekten ved 3 års intervaller. Dette er et bra tegn, og tyder på at resultatene jeg har fått virker å være gyldige. Fordelen med 5 års intervaller er at jeg får muligheten til kontrollere for utdanning. Dermed utvider jeg nå modellen med andelen utdanning. Jeg skiller da på henholdsvis andel av befolkning med høyere utdanning, og andel av befolkningen med videregående eller høyere utdanning samlet sett. Dette vil fungere som en form for robusthetssjekk, for å sjekke om estimatene fortsatt er stabile.

Modellen gitt ved likning (5.3) utvider jeg nå med utdanning. Jeg estimerer dermed modellen gitt ved likning (5.5). Resultatene av denne estimeringen er gitt i tabell 6 kolonne (3), (4), (5) og (6). Kolonne (3) og (4) er kun med høyere utdanning, mens (5) og (6) er med videregående og høyere utdanning. Estimering av denne modellen gir ikke store

utslag på de allerede inkluderte variablene i modellen. Konvergensfarten er på 10.86 med kontrollering for høyere utdanning, og 10.48 med kontrollering for videregående og høyere utdanning. Denne konvergensfarten er så godt som identisk med estimering uten kontrollering for utdanning. Konvergensfarten virker dermed å være robust på rundt 10 prosent i denne paneldatamodellen.

Tabell 6: Paneldatamodell med 5 års intervaller

| | Faste effekter | | | | | |
|---|---|---|---|---|---|---|
| | $\triangle lny_{it}$ er forklaringsvariabelen | | | | | |
| | 5 år | 5 år | 5 år | 5 år | 5 år | 5 år |
| Variabler | (1) | (2) | (3) | (4) | (5) | (6) |
| $lny_{it-1}$ | -0.438*** | -0.417*** | -.433*** | -0.419*** | -0.476*** | -0.408*** |
| | (-28.52) | (-38.43) | (-28.35) | (-37.75) | (-28.28) | (-33.08) |
| $lnBef_{it-1}$ | 0.898*** | 0.363*** | 0.567*** | 0.390*** | 0.665*** | 0.356*** |
| | (10.85) | (4.45) | (5.11) | (4.82) | (6.92) | (4.36) |
| $\triangle lnBef_{it}$ | 0.786*** | 0.496*** | 0.458** | 0.594*** | 0.847*** | 0.435*** |
| | (4.16) | (3.64) | (2.35) | (4.17) | (4.56) | (3.06) |
| $Utd_{it-1}$ | | | 0.837*** | -0.819*** | | |
| | | | (4.60) | (-3.16) | | |
| $\triangle Utd_{it}$ | | | 1.070*** | 0.252 | | |
| | | | (3.48) | (1.08) | | |
| $Utd2_{it-1}$ | | | | | 0.309*** | -0.093 |
| | | | | | (5.02) | (-1.53) |
| $\triangle Utd2_{it}$ | | | | | 0.382*** | -0.074*** |
| | | | | | (3.06) | (-11.20) |
| $Konstant$ | -1.987*** | 1.699*** | 0.285 | 1.572*** | 0.005 | 1.687*** |
| | (-3.57) | (2.92) | (0.37) | (2.72) | (0.01) | (2.90) |
| *Periodedummier* | Nei | Ja | Nei | Ja | Nei | Ja |
| Oservasjoner | 672 | 672 | 672 | 672 | 672 | 672 |
| $R^2$ | 0.594 | 0.810 | 0.614 | 0.816 | 0.612 | 0.811 |

*Merk:* Her er datasettet delt opp i 5 års intervaller. I parentes er t-verdier oppgitt. Dersom *** er den signifikant på 1 % nivå, ** signifikant på 5 % nivå og ved * er den signifikant på 1% nivå. Utd er andelen av befolkningen med fullført høyere udtanning. Utd2 er andel av befolkningen over 15 med fullført videregående eller høyere utdanning. Jeg rapporterer ikke periodedummiene.

Effektene av befolkningsnivå og befolkningsvekst på vekstraten til BNP per innbygger endrer seg lite med inkludering av utdanning og virker dermed å være forholdsvis stabile. Dette gjelder estimeringen med periodedummier. Uten periodedummier varierer effekten mer. Den variabelen som har mest markant endring er effekten av vekstraten til befolkningen på vekstraten til BNP per innbygger. Nå vil en økning i vekstraten til befolkningen på 10 prosentpoeng forrige periode føre til en økning i vekstraten til BNP per innbygger på 5.94 prosentpoeng. Dette er nesten et prosentpoeng mer enn uten utdanningsvariablene. Dette gjelder modellen med kun høyere utdanning. Elastisiteten til BNP per innbygger med hensyn på befolkningsnivå er 1.31 uten periodedummier og 0.93 med periodedummier for høyere utdanning. Med videregående og høyere utdanning er denne elastisiteten 1.39 uten periodedummier og 0.87 med periodedummier. Elastisiteten til BNP per innbygger med hensyn på befolkningsvekst virker også dermed å være veldig robust med periodedummier siden elastisiteten er tilnærmet lik årlige og 3 års intervaller.

I denne modellen er det ikke like lett å få klare resultater av variabelen utdanning. En ville tro at denne var positiv slik som i tverrsnittsmodellen. Før jeg inkluderte periodedummier var koeffisienten til utdanningsnivået forrige periode positiv på 0.837. Denne var også klart signifikant, på 1 prosent signifikansnivå. Når jeg inkluderer periodedummier får denne plutselig motsatt fortegn med en koeffisient på -0.819 og fortsatt signifikant på 1 prosentnivå. Det samme skjer når videregående utdanningsnivå også tas med, men endringen blir ikke like stor her. Når en også ser på effekt av en endring i utdanningsnivået går denne fra å ha en stor effekt på vekstraten til BNP per innbygger med en koeffisient på 1.07 som er signifikant. Inkluderer jeg periodedummier blir denne effekten på 0.251 og usignifikant. I modellen hvor videregående også er tatt med, blir til og med denne effekten negativ. Dette gjør at utdanningsparametrene er vanskelig å tolke siden de endrer seg såpass mye. Som jeg argumenterte for i kapittel 5.3 kan det tyde på at nasjonale data ikke er bra nok. Ettersom det i datasett 2 ikke er så mange forskjellige land med og dataene er på nasjonalt nivå kan det bli for lite variasjon i utdanningsnivået mellom regionene.

I paneldatamodellen jeg har estimert her virker resultatene og være robuste. De endrer seg lite om jeg estimerer modellen med årlige, 3 års og 5 års intervaller. Dette er tydelige tegn på at resultatene jeg har fått er robuste. Konvergensen i denne modellen ligger på rundt 10 prosent noe som er forholdsvis likt lignende studier som Caselli et al. (1996), Badinger et al. (2002) og Bosker (2007). Jeg får også en positiv effekt av vekstraten til befolkningen på vekstraten til BNP per innbygger. Denne lå på rundt 0.4 for 3 års og 5 års intervaller. Den var en god del mindre for årlig. Dette var forskjellig fra Bosker (2007) som fikk en negativ effekt av vekstraten til befolkningen på vekstraten til BNP per innbygger. Jeg får også positive effekter av befolkningsnivå på vekstraten til BNP per innbygger. Så godt som alle resultatene jeg har fått er signifikante på 1 prosentnivå. Siden resultatene er forholdsvis like med de forskjellige spesifikasjonene av paneldatamodellen virker resultatene å være robuste.

## 6.3 Oppsummering og sammenligning mellom de to modellene

Tverrsnittsmodellen og paneldatamodellen ga forskjellige resultater når det kommer til konvergensfarten. Disse forskjellige resultatene var kanskje ikke så overraskende som en skulle tro. Tidligere forskning har fått tilnærmet like resultater ved de forskjellige modellene. Cuadrado-Roura et al. (2000) fikk også store forskjeller i konvergensfart med en tverrsnittsmodell og paneldatamodell for samme datasett. Fra en konvergensfart på rundt 2 prosent i tverrsnittsmodellen til rundt 10 prosent i paneldatamodellen er det vanskelig å trekke en slutning om hvilken av disse modellene som virker å være mest riktig. Det kan tyde på at tverrsnittsmodellen ikke fanger opp det forskjellige steady state nivået til regionene og dermed er beheftet med målefeil, slik som Islam (1995) argumenterte for. Når det gjelder paneldatamodellen min er den avhengig av lange tidsserier og det er mulig

at min tidsserie ikke er lang nok slik som Shioji (1997) argumenterte for i sin analyse. En konvergensfart på 2 prosent kan virke noe lavt, i motsetning til 10 prosent som kan virke noe høyt. Det vil dermed være vanskelig å konkludere med hva konvergensfarten er ut ifra resultatene til den empiriske analysen min.

Når det kommer til effekten av befolkningsvekst virker effekten å være noe større i tverrsnittsmodellen enn i paneldatamodellen. Dette gjelder særlig i datasett 2 selv om den var kun signifikant på et 20 prosentnivå. Begge modellene får en positiv effekt, noe som tyder på at økt befolkningsvekst i en region vil føre til økt inntektsvekst i denne regionen. Dette er motstridene i forhold til resultatene til Bosker (2007). Han fikk en negativ effekt av befolkningsvekst. Braun (1993) derimot fikk en positiv effekt av nettomigrasjon, selv om denne ikke var signifikant i hans tilfelle. Ettersom jeg fikk positiv effekt av befolkningstetthet, kan det tyde på at de positive effektene av agglomerasjon overgår de negative. Dette kan være med på å forklare den økte veksten i inntekten ved at folk flytter til en region. Denne effekten er også motstridene med Bosker (2007). Han fikk som nevnt tidligere både positiv og negativ effekt av agglomerasjon, avhengig av hvilken spesifikasjon han brukte. Han fikk negativ effekt ved den spesifikasjonen han la mest vekt på som var med faste effekter og periodedummier.

Det virker å være en tendens at folk flytter fra fattige og perifere regioner til mer urbane og rike regioner. Dette vil isolert sett føre til inntektsdivergens i de fattige regionene. Men både tverrsnittsmodellen og paneldatamodellen viste klare tendenser til konvergens. Det kan dermed virke som den positive effekten av befolkningsvekst blir dominert av konvergensmekanismen.

# 7 Konklusjon

I denne oppgaven har jeg undersøkt om det forekommer inntektskonvergens mellom regioner i Europa og hvordan befolkningsvekst påvirker inntektsveksten. Siden jeg ønsker å se på regionsnivå har jeg brukt NUTS 2 klassifiseringen til Eurostat og sett på perioden 1980 - 2009. For å undersøke denne problemstillingen har jeg brukt to forskjellige modeller, henholdsvis en tverrsnittmodell og paneldatamodell. I tverrsnittsmodellen har jeg brukt to datasett. Datasett 1 består av tilnærmet alle regionene i EU for perioden 1995 - 2008. Datasett 2 består av færre regioner for perioden 1980 - 2009. I paneldatamodellen har jeg kun brukt datasett 2. Siden jeg har brukt to forskjellige modeller får jeg mulighet til å gjøre en solid empirisk analyse av problemstillingen.

I analysen kom det tydelig frem at inntektsforskjellene i Europa ble mindre over tid. Farten på denne konvergensen varierte en god del avhengig av hvilken modell som ble benyttet. I tverrsnittsmodellen fikk jeg en ubetinget konvergens på rundt 1.3 prosent. Dette er en lav konvergensfart og noe mindre enn tidligere studier for ubetinget konvergens. Når jeg videre undersøkte for betinget konvergens, fikk jeg en konvergensfart på rundt 2 prosent. Konvergensfarten viste seg å være litt lavere i datasett 2 enn i datasett 1. En slik betinget konvergensfart stemte godt overens med tidligere forskning ved bruk av tverrsnittsmodell. Når jeg utvidet analysen min med paneldata økte konvergensfarten betraktelig. Konvergensfarten med paneldatamodellen ligger på rundt 10 prosent. Denne konvergensfarten virket å være robust, ettersom farten endret seg lite når jeg kontrollerte for utdanningsnivå og ved å beregne dataene med årlige, 3 års og 5 års intervaller. Dette resultatet er i tråd med resultater for tidligere paneldata undersøkelser. Begge modellene gir klare indikasjoner på at inntektsforskjellene mellom regioner i Europa blir mindre over tid. Hvor rask en slik konvergensfart egentlig er, blir vanskelig å avgjøre ettersom variasjonen mellom de to modellene er såpass stor.

I både tverrsnittsmodellen og paneldatamodellen gir befolkningsvekst en utelukkende positiv og signifikant effekt på veksten til BNP per innbygger i en region. Siden begge disse modellene gir en positiv effekt tyder dette på at dersom befolkningen i en region øker vil også inntektsveksten i regionen øke. I forhold til tidligere forskning var effekten av befolkningsvekst på inntektsvekst litt mere delt. I tverrsnittsmodellen fikk jeg også undersøkt effekten av befolkningstetthet på veksten til BNP per innbygger. Denne hadde en klar signifikant og positiv effekt. Det at befolkningstettheten er positiv kan tyde på at de positive effektene ved agglomerasjon overgår de negative effektene og bidrar til økt inntektsvekst. Dette kan være med på å forklare den positive effekten av befolkningsvekst på BNP per innbygger.

Resultatene fra den empiriske analysen åpner for diskusjon rundt en av hovedmålsetningene til EU. EU ønsker at inntektsforskjellene mellom regionene i Europa skal bli mindre

over tid. Ut ifra konvergensfarten fra den empiriske analysen virker denne målsetningen å bli oppnådd. Det som kan være problematisk er at dersom perifere regioner som i utgangspunktet er et stykke bak de rike regionene opplever fraflytting, noe som virker å være en trend i dag. Denne fraflyttingen vil ifølge resultatene mine, som virker å være robuste, føre til en nedgang i inntektsveksten i disse regionene. Siden jeg totalt sett finner konvergens, virker den positive effekten av befolkningsvekst å bli dominert av konvergensmekanismen. Dersom dette ikke er tilfellet kan det føre til at fattige regioner vil ha problemer med å tette gapet til de rike regionene. Det kan da være nødvendig for EU å ta i bruk virkemidler utover det som anvendes i dag for å motvirke dette gapet. Den empiriske analysen min virker å være robust, men som jeg har argumentert for står den overfor flere økonometriske utfordringer som blant annet simultanitet og utelatte variabler. Derfor vil det være behov for mer forskning på dette området. Bedre kunnskap om disse sammenhengene er av stor betydning for oppnå de ønskede politiske målsetningene i EU.

# Referanser


Abel, A. B. (1982): "Dynamic effects of permanent and temporary tax policies in a q model of investment." *Journal of Monetary Economics 9*, 353 – 373.

Aghion, P., Boustan, L., Hoxby, C. og Vandenbussche, J. (2005): "Exploiting states' mistakes to identify the causal impact of higher education on growth." Department of Economics, Harvard University.

Arbia, G., de Dominicis, L. og Piras, G. (2005): "Regional growth and regional inequality in transition countries: a spatial econometric approach." *ERSA Congress, Amsterdam, 23 - 27 August* .

Badinger, H., Müller, W. og Tondl, G. (2002): "Regional convergence in the European Union(1985- 1999)." *IEF Working Paper Nr. 47.*

Barro, R. J. og Sala-i Martin, X. (1992): "Convergence." *Journal of Political Economy 100 (2)*, 223 – 251.

Barro, R. J. og Sala-i Martin, X. (2004): *Economic Growth Second Edition.* The MIT Press, Cambridge, Massachusetts, London, England.

Boldrin, M. og Canova, F. (2001): "Inequality and convergence in Europe's regions: Reconsidering European regional policies." *Economic Policy 16*, 205 – 253.

Bosker, M. (2007): "Growth, agglomeration and convergence: A space-time analysis for European regions." *Spatial Economic Analysis 2:1*, 91–100.

Braun, J. (1993): "Essays on economic growth and migration." Ph.D Dissertation, Harvard University.

Canova, F. (2001): "Are EU policies fostering growth and reducing regional inequalities." *Els Opuscles del CREI 8*, 1 – 35.

Caselli, F., Esquivel, G. og Lefort, F. (1996): "Reopening the convergence debate: A new look at cross-country growth empirics." *Journal of Economic Growth 1*, 363 – 389.

Ciccone, A. (2001): "Agglomeration effects in Europe and the USA." *Els Opuscles del CREI 9.*

Cuadrado-Roura, J. R., Mancha-Navarro, T. og Garrido-Yserte, R. (2000): "Convergence and regional mobility in the European Union." *European Regional Science Association ersa00.*

Dowrick, S. og Rogers, M. (2002): "Classical and technological convergence: beyond the Solow-Swan growth model." *Oxford Economic Papers 54*, 369 – 385.

EEA (2009): "Ensuring quality of life in Europe's cities and towns." European Environment Agency Repor No *5*.

Eurostat (2007): "Regions in the European Union - Nomenclature of territorial units for stistics - NUTS 2006/ EU 27." *Eurostat Methodologies and working paper* .

Fingleton, B. (1999): "Estimates of time convergence : An analysis of regions of the European Union." *International Regional Science Review 1999 22:5*, 5 – 34.

Glaeser, E. L. og Gottlieb, J. D. (2009): "The wealth of cities: Agglomeration economien and Spatial equilibrium in the United States." *Journal of Economic Literature 2009 47:4*, 983 – 1028.

Hayashi, F. (1982): "Tobin's marginal q and average q: A neoclassical interpretation." *Econometrica 50 (1)*, 213 – 224.

Islam, N. (1995): "Growth empirics: A panel data approach." *The Quarterly Journal of Economics 110, (4)*, 1127–1170.

Sala-i Martin, X. (1996): "The classical approach to convergence analysis." *The Economic Journal 106*, 1019 – 1036.

Melo, C. M., Graham, D. J. og Noland, R. B. (2009): "A meta-analysis of estimates of urban agglomeration economies." *Regional Science and Urban Economics 39*, 332–342.

Moretti, E. (2010): "Local labor markets." *NBER Working Paper 15947* .

Nelson, R. R. og Phelps, E. S. (1966): "Investment in humans, technological diffusion, and economic growth." *American Economic Review 56(2)*, 69 – 75.

Rappaport, J. (2005): "How does labor mobility affect income convergence?" *Economic Dynamics & Control 29*, 567–581.

Rattsø, J. og Stokke, H. E. (2009): "Regional income convergence, skilled migration and productivity response: Explaining relative stagnation in the periphery*." *Working Paper* .

Shioji, E. (1997): "It's still 2%: evidence on convergence from 116 years of the US States panel data." *Economics Working Papers 236, Universitat Pompeu Fabra* .

Wooldridge, J. M. (2009): *Introductory econometrics - A modern approach.* Michigan State University: South-Western Cengage Learning.

# A Appendiks

## A.1 Steady state løsningen for velferden til arbeiderne

Siden jeg kun ser på en av økonomiene dropper jeg å ta med i.

$$
\begin{aligned}
h(t) &= \int_t^\infty w(s)e^{-r(s-t)}ds \\
h(t) &= \int_t^\infty w(s)e^{-rs}e^{rt}ds \\
h(t) &= e^{rt}\int_t^\infty w(s)e^{-rs}ds \\
h(t) &= e^{rt}[-\frac{1}{r}w(s)e^{-rs}]_t^\infty ds \\
h(t) &= e^{rt}[-\frac{1}{r}w(\infty)e^{-r\infty} - (-\frac{1}{r}w(t)e^{-rt})] \\
h(t) &= e^{rt}(\frac{1}{r}w(t)e^{-rt}) \\
h(t) &= \frac{w(t)}{r}
\end{aligned}
$$

Dersom jeg nå setter inn for inn for k* inn i likning 2.8 får jeg w*. Når jeg nå setter inn for w* får jeg h*. Slik at steady state verdien til arbeidsvelferden for de to økonomiene er gitt ved:

$$h^* = \frac{w^*}{r} \tag{2.18}$$

## A.2 Dynamikken til arbeidernes velferd

Har utrykket for velferden gitt ved likning 2.15. Siden jeg kun ser på den lille økonomien fjerner jeg i.

$$h(t) \equiv \int_t^\infty w(s)e^{-r(s-t)}ds \tag{2.15}$$

Deriverer nå denne gjennom med hensyn på t siden jeg ønsker å se på utviklingen over tid. Må her bruke Leibniz regel.

$$h(t)\frac{d}{dt} = \frac{d\infty}{dt}w(\infty)e^{-r(\infty-t)} - \frac{dt}{dt}w(t)e^{-r(t-t)} + \int_t^\infty \frac{\partial}{\partial t}w(s)e^{-r(s-t)}ds$$

$$\dot{h}(t) = -w(t) + -\int_t^\infty rw(s)e^{-r(s-t)}ds$$

Kan sette r utenfor integraltegnet og setter inn for $h(t) = \int_t^\infty w(s)e^{-r(s-t)}ds$. Får da:

$$\dot{h}(t) = rh(t) - w(t) \tag{2.19}$$

## A.3 Fullt navn på NUTS regionene som er droppet fra analysen

Tabell 7: Fullt navn på NUTS regionene som er droppet fra analysen

| NUTS kode | Fullt navn på regionen |
|---|---|
| Datasett 1: | |
| UKM5 | North Eastern Scotland |
| UKM6 | Highlands and Islands |
| Datasett 2: | |
| UKM5 | North Eastern Scotland |
| UKM6 | Highlands and Islands |
| DE30 | Berlin |
| DE41 | Brandenburg - Nurdost |
| DE42 | Brandenburg - Südwest |
| DE80 | Mecklenburg-Vorpommern |
| DED1 | Chemnitz |
| DED2 | Dresden |
| DED3 | Leipzig |
| DEE0 | Sachsen-Anhalt |
| DEG0 | Thüringen |
| ES63 | Ciudad Autónoma de Ceuta |
| ES64 | Ciudad Autónoma de Melilla |
| FR91 | Guadeloupe |
| FR92 | Martinique |
| FR93 | Guyane |
| FR94 | Réunion |
| NL21 | Overijssel |
| NL22 | Gelderland |
| NL23 | Flevoland |
| PT16 | Centro |
| PR17 | Lisboa |
| PT20 | Região Autonóma dos Açores |
| PT30 | Região Autonóma da Medeira |
| UKC | North East |
| UKD | North West |
| UKH | East of England |
| UKI | London |
| UKJ | South East |