

Ida Kristine Stakston

# Agglomerasjonseffekter i norske regioner

En empirisk analyse av heterogeniteter i de urbane lønnspremiene blant individer med ulik innvandringsbakgrunn

Masteroppgave i Samfunnsøkonomi

Veileder: Jørn Rattsø

Juni 2022



Ida Kristine Stakston

# **Agglomerasjonseffekter i norske regioner**

En empirisk analyse av heterogeniteter i de urbane lønnspremiene blant individer med ulik innvandringsbakgrunn

Masteroppgave i Samfunnsøkonomi  
Veileder: Jørn Rattsø  
Juni 2022

Norges teknisk-naturvitenskapelige universitet  
Fakultet for økonomi  
Institutt for samfunnsøkonomi



Kunnskap for en bedre verden



## Sammendrag

Formålet med denne oppgaven var å analysere agglomerasjonseffektene i de økonomiske regionene i Norge, og mulige heterogeniteter i disse blant individer med ulik innvandringsbakgrunn. Oppgaven har tatt i bruk en individuell fixed effects modell som hovedstrategi, for å estimere de urbane lønnspremiene som et empirisk mål på agglomerasjonseffektene.

Hovedanalysen til oppgaven fant bekreftelse på signifikante lønnspremier av å bo i de store økonomiske regionene, definert som de med over 150 000 innbyggere, hvor vi fant en premie på rundt 1%. I tillegg til dette fant vi antydning til at innvandrere fra lavinntektsland hadde signifikant lavere premier. Dette indikerte at de vil ha mindre gevinster fra agglomerasjonen, sammenliknet med resten av befolkningen.

Disse resultatene viste seg dog og kun være drevet av Stor-Oslo regionen. En robusthetssjekk som trakk ut Stor-Oslo som en egen kategori, resulterte i at vi ikke fant signifikante premier i de resterende store regionene lenger. Et hovedfunn fra analysen er at Stor-Oslo forventes å ha en gjennomsnittlig lønnspremie på 2.3% over de andre regionene i Norge, og denne drives i stor grad av de yngre med høyere utdanning. Det andre hovedfunnet er at det viser seg og ikke være noen signifikante heterogeniteter i lønnspremiene blant individer med ulik innvandringsbakgrunn.

Denne analysen finner alt i alt at samlokaliseringen av økonomisk aktivitet kun tilsynelatende er stor nok i Stor-Oslo regionen. Der vi finner tegn på tilstedeværelsen av agglomerasjonseffekter målt av en betydelig statistisk, og økonomisk signifikant positiv lønnspremie. Produktivitetsgevinstene fra agglomerasjonen viser seg å være universell for alle fulltids-arbeiderne, og det finnes ingen tegn på betydelige forskjeller blant de med forskjellige innvandringsbakgrunner.

## **Abstract**

The purpose of this paper is to investigate the presence of agglomeration economies in Norwegian economic-worker regions, and the heterogeneity with respect to different types of immigrants. An individual fixed effects model is used as the main strategy for estimating the urban wage premiums, which is used as the empirical measurement of agglomeration economies.

The main analysis in this paper finds evidence of a significant urban wage premium estimated to be around 1% in the larger economic regions, defined as those with above 150 000 inhabitants. Additionally, we also find some indications that the immigrants from low-income countries are the only group with significant differences in this premium. We find that they appear to have statistically significantly lower economic gains from the agglomeration compared to the rest of the population.

These results turn out to mainly be driven by the Stor-Oslo region. Where a robustness check on the above-mentioned results where we extract Stor-Oslo as its own economic region leads to it becoming the only region with a significant wage premium. One of the main results from this paper is that the workers in Stor-Oslo is predicted to on average have a wage premium of about 2.3%, over the workers in the other regions. This wage premium is in large driven by the younger individuals with higher education. The other main result from this analysis is that in the end there appears to be no significant heterogeneities in the agglomeration economies for individuals with various immigrant backgrounds.

This paper thus concludes with that the collocation of economic activity only appears to be large enough in the Stor-Oslo region in Norway. It is only within this region that we find evidence of the presence of agglomeration economies, through a statistically and economic significant estimated positive wage premium. The productivity advantages from the agglomeration appears to be universal for all the fulltime workers in Norway, where we find no support for heterogeneities for individuals with different immigration backgrounds.

## **Forord**

Denne masteroppgaven markerer avslutningen på et 5-årig langt studieløp innenfor samfunnsøkonomi i Trondheim ved NTNU. Arbeidet med oppgaven har vært ekstremt tidkrevende og frustrerende, men aller mest lærerikt. Først og fremst vil jeg takke veilederen min, Jørn Rattsø, for god hjelp og oppfølging gjennom hele prosessen. Videre vil jeg rette en enormt stor takk til mine medstudenter for alle gode diskusjoner, og for å ha gjort studietiden min helt uforglemmelig. Jeg vil også takke familie og venner for korrekturlesing og motiverende samtaler.

Ida Kristine Stakston

Trondheim, 15.juni 2022

# Innholdsfortegnelse

1. Innledning .....	1
1.1 Motivasjon og hypoteser .....	1
1.2 Disposisjon .....	2
2. Teori og tidligere litteratur .....	3
2.1 Den monosentriske bymodellen .....	3
2.2 Tidligere litteratur .....	5
2.3 Oppsummerende forventninger .....	10
3. Datagrunnlag .....	11
3.1 Microdata.no .....	11
3.2 Avgrensing av utvalget .....	12
3.3 Variabler .....	13
3.4 utfordringer og begrensninger med datasettet .....	20
4. Metode .....	23
4.1 Modellspesifikasjon .....	23
4.2 Estimeringsmetoder .....	24
4.3 Økonometriske utfordringer .....	25
5. Resultater .....	29
5.1 Førsteintrykk med OLS .....	29
5.2 Hovedmodellen FE .....	32
5.3 Variasjons-grunnlaget; Flytterne .....	34
6. Robusthet .....	36
6.1 Ekstremverdier .....	36
6.2 Andre heterogeniteter .....	38
6.2 Omdefinering av regionene .....	40
7. Oppsummering og diskusjon .....	43
7.1 Drøfting av resultatene .....	43
7.2 utfordringer .....	45
7.3 Videre forskning .....	46
8. Konklusjon .....	48
Referanser .....	49
Appendiks .....	52



## 1. Innledning

Til tross for enklere, billigere og friere flyt av varer, tjenester og mennesker, er tilstedeværelsen av så kalte agglomerasjonseffekter fortsatt ansett som en avgjørende forklaringsfaktor for hvorfor noen byer vokser seg større og raskere enn andre.

Agglomerasjonseffekter omfatter gevinster og kostnader som oppstår fra samlokaliseringen av individer og bedrifter innenfor et geografisk område (Glaeser, 2010). Det bygger på teorien om stor-driftsfordeler, og kan fra litteraturen oppsummeres med å oppstå gjennom tre overordnede mekanismer; deling, matching og kunnskapsutveksling. Når komplementære bedrifter er lokalisert i nærheten av hverandre vil de kunne få lavere transportkostnader og kostnadsfordeler ved at markedene blir mer integrerte, og de f.eks. kan dele på infrastruktur. Næringsklynger vil som et resultat lettere kunne oppstå og føre til mer spesialiserte arbeidsmarkeder som sikrer en bedre matching mellom arbeidstakere og arbeidsgivere (Bolter & Robey, 2020). Vanlige eksempler på slike næringsklynger er f.eks. teknologi i Silicon Valley, finansnæringen i New York og oljeindustrien i Stavanger. Disse samlokaliseringseffektene fører til at det oppstår selvforsterkende mekanismer der byer som opplever økt økonomisk aktivitet blir desto mer attraktive, og vil derfor fortsette å vokse over tid. Prisene og lønningene i byene vil som et resultat presses opp, og nominelle lønninger brukes derfor ofte som et empirisk mål på agglomerasjonseffektene.

### 1.1 Motivasjon og hypoteser

Studier som forsøker å tallfeste omfanget av agglomerasjonseffekter anses som svært betydningsfulle blant annet for utredning av politiske tiltak og for å få fullverdige kostnads-nytte analyser av prosjekter (Combes, Duranton, & Gobillon, 2010). Denne oppgaven ønsker derfor å analysere tilstedeværelsen av agglomerasjonseffekter i Norge, som i likhet med resten av verden preges av stadig mer sentraliserte bosettingsmønstre. Mer spesifikt ønsker vi å undersøke mulige heterogeniteter i disse effektene blant individer med ulik innvandringsbakgrunn.

Det er lite empiri på agglomerasjonseffekter blant innvandrere, og denne oppgaven ønsker å være et bidrag til dette. I tillegg har vi her valgt å ta i bruk mikropaneldata på individnivå for å forsøke å estimere de individuelle agglomerasjonseffektene og mulige heterogeniteter blant disse med tanke på innvandringsbakgrunn. Dette er i noe kontrast til mye tidligere litteratur som i stor grad

tar i bruk regiondata. Det har medført flere utfordringer og resultert i noen svakheter ved analysen, men er likevel vurdert som hensiktsmessig for å forsøke å estimere mer presise individuelle effekter av agglomerasjonen.

Integreringsarbeidet både i Norge, og generelt i resten av verden, vektlegger innvandreres tilknytning til arbeidsmarkedet som en avgjørende arena for å knytte sosiale bånd og skape tilhørighet i samfunnet (Olsen, 2017). Det vil derfor være nærliggende å tenke seg at innvandrere vil ha desto større gevinster fra agglomerasjonen sammenliknet med den øvrige befolkningen. Fra klassisk økonomisk teori kan det dog sies å ville avhenge av hvorvidt innvandrerne er komplementære eller substitutter for den lokale arbeidskraften. Denne oppgaven ser spesifikt på tre ulike innvandringsbakgrunner; norskfødte med innvandrerforeldre og innvandrere fra høy- og lavinntektsland. Vi forventer å finne positive ekstra premier blant de norskfødte med innvandrerforeldre og innvandrerne fra høyinntektsland, som kan anses som mer komplementære til etnisk norske arbeidere. Mens innvandrerne fra lavinntektsland kan argumenteres for og i større grad være substitutter, og dermed forventes å ha lavere agglomerasjonseffekter sammenliknet med resten.

### **1.2 Disposisjon**

Oppgaven er totalt delt inn i 8 forskjellige kapitler, som er strukturert som følger. Kapittel 2 vil presentere det teoretiske rammeverket og gjennomgår i kortere trekk et utvalg av tidligere empiri. Kapittel 3 gjennomgår datagrunnlaget for analysen, med presentasjon av variablene for analysen og deres tilhørende deskriptiv statistikk. Den empiriske strategien, valg av metode og utfordringene med disse presenteres i kapittel 4, før resultatene fra analysen gjennomgås i kapittel 5. Kapittel 6 omfatter robusthetssjekker og alle funnene fra analysen diskuteres så oppsummerende i kapittel 7. Til slutt er det en konklusjon av oppgaven i kapittel 8.

## 2. Teori og tidligere litteratur

Dette kapittelet presenterer først et teoretisk rammeverk for byvekst og agglomerasjonseffekter. Deretter gjennomgås et utvalg av tidligere empiri, og de empiriske effektene av innvandrere på lokale arbeidsmarkeder i kortere trekk.

### 2.1 Den monosentriske bymodellen

Agglomerasjonseffekter sitt forhold til byvekst og hvordan det påvirker økonomiske størrelser kan tydeliggjøres i et forenklet format av den monosentriske bymodellen. Denne oppgaven bruker i likhet med artikkelen til Duranton og Puga (2013) denne modellen som grunnlag for analysen, og den vil her gjennomgås i oppsummerende trekk.

Modellen bygger på antakelser om at all økonomisk aktivitet og spesielt arbeidsmarkedet er konsentrert innenfor et spesifikt geografisk område. Dette er svært strenge antakelser som kan argumenteres for å gjenspeile dagens bybilde i liten grad, hvilket heller kan karakteriseres som polysentrisk, og preges av storbyer med opptil flere bykjerner. Modellen fanger likevel opp viktige grunnleggende sammenhenger og dynamikk som denne oppgaven bruker som motivasjon for sin analyse.

Det er kun et mindre utvalg av likningene fra artikkelen som er spesielt relevante for å forstå sammenhengene mellom byvekst og agglomerasjonseffektene. Her indekserer underbetegnelsen  $i$  byen og  $j$  sektoren.

$$(1) U\left(A_i, v\left(c\left(\bar{R} + \tau N_i\right), w\right)\right) = \bar{U}, \quad \frac{\partial U}{\partial A} > 0, \frac{\partial U}{\partial v} > 0, \frac{\partial v}{\partial N_i} < 0, \frac{\partial U}{\partial w} > 0$$

$A_i$  - Fellesnivå på fasiliteter og bekvemmeligheter i byen, eks. klima, kystlinje o.l.

$\bar{R}$  - Boligprisene

$c$  - Funksjon for de urbane kostnadene, knyttet til transport og trengsel

$N_i$  - Befolkningen/arbeidsstyrken, antar full sysselsetting i økonomien

$w$  - Nominell lønnsrate

Likning (1) betegner den spatiale likevektsbetingelsen til den monosentriske bymodellen som tar hensyn til tilstedeværelsen av felles fasiliteter (Duranton & Puga, 2013, s. 17). Den sier at for at en skal ha en stabil likevekt i økonomien må nytten av å bo i byene være lik. Men denne likevekten alene forklarer ikke byvekst og oppstandelsen av nye storbyer over tid. Den oppgir kun en negativ effekt av befolkningsstørrelsen på nytten til innbyggerne gjennom økte urbane kostnader. Mens en større befolkning også vil kunne påvirke innbyggernes nytte positivt gjennom agglomerasjonseffekter.

En vanlig måte å inkludere agglomerasjonseffekter i slike urbane modeller på er å anerkjenne i tråd med empirien at lønnen i byene avhenger positivt av befolkningen  $w_i = w(N_i)$ ,  $\frac{\partial w_i}{\partial N_i} > 0$  (Glaeser & Maré, 2001, s. 22). Det betyr at når befolkningen øker er det også en positiv effekt gjennom agglomerasjonseffekter i modellen, der lønnsnivået øker.

Vi ser nærmere på hva som ligger bak denne lønnslikningen ut ifra et mikrogrunnlag utviklet av Abdel-Rahman og Fujita (1990), hvilket forklarer produktivetsfordelene som oppstår i større byer. Modellen bygger på et sett med likninger og teori rundt produksjonen av mellom- og endelige varer som er spesifikke for sektorer, der en full utledning av modellen er vedlagt i appendiks A1.

To av de mest relevante likningene som utledes i modellen er:

$$(2) Y_i^j = B^j (N_i^j)^{1+\sigma^j}$$

$$(3) w_i^j = P^j B^j (N_i^j)^{\sigma^j}$$

$Y_i^j$  - Aggregert produksjon av endelige varer

$B^j$  - Teknologiutvikling, ansett som et mål på produktivitet i sektoren

$\sigma^j$  - Agglomerasjonseffekter, hvor  $0 < \sigma^j < 1$

$P^j$  - Varepriser

Den aggregerte produksjonen og lønnen innenfor en sektor i byen avhenger begge positivt av produktiviteten og befolkningsstørrelsen, hvor agglomerasjonseffektene så forsterker disse kreftene.

$$\frac{\partial Y_i^j}{\partial N_i^j} = B^j(1 + \sigma^j)N_i^{j\sigma^j} > 0, \quad \frac{\partial^2 Y_i^j}{\partial N_i^{j^2}} = B^j\sigma^j(1 + \sigma^j)N_i^{j\sigma^j-1} > 0$$

$$\frac{\partial w_i^j}{\partial N_i^j} = P^j B^j \sigma^j N_i^{j\sigma^j-1} > 0, \quad \frac{\partial^2 w_i^j}{\partial N_i^{j^2}} = P^j B^j \sigma^j (\sigma^j - 1) N_i^{j\sigma^j-2} < 0$$

Befolkningsvekst i byene gir tilgang på en større arbeidsstyrke og ifølge denne modellen har det tiltakende positive skalaeffekter på produksjonen, mens effekten på lønnen er positiv og avtakende. Den samlede produksjonen har altså tiltakende skalaavkastning på både by- og sektornivå som følge av tilstedeværelsen av agglomerasjonseffektene. Noe som gjør at bedriftene kan øke produksjonen med mer enn proporsjonalt av økningen i arbeidsstyrken (Duranton & Puga, 2013, s. 23). Lønnen påvirkes også positivt av agglomerasjonseffektene, men i avtakende grad.

Her ser en altså hvilken rolle agglomerasjonseffektene spiller inn på bybildet ut ifra den monosentriske bymodellen, som fremhever avveiningen mellom økt nytte og kostnader som følge av befolkningsvekst. På den ene siden bidrar en større befolkning  $N_i^j \uparrow$  til en eksponentielt høyere produksjon hos bedriftene  $Y_i^j \uparrow\uparrow$  og høyere lønninger til arbeiderne  $w_i^j \uparrow$ , men det fører også til økte urbane kostnader  $c \uparrow$ . Denne oppgaven fokuserer på agglomerasjonseffektene målt av lønningene, ved å forsøke å tallfeste de urbane lønnspremiene på lik linje med flere empiriske studier.

## 2.2 Tidligere litteratur

Det er et stort omfang av litteratur som analyserer agglomerasjonseffekter og forsøker å estimere de urbane lønnspremiene i flere forskjellige land og områder med ulike empiriske strategier. Heterogeniteten til lønnspremiene er bredt undersøkt, spesielt med fokus på betydningen av utdanning og yrker. Det er derimot som nevnt innledningsvis lite empiri på den relative produktiviteten og agglomerasjonseffektene blant ulike innvandringsgrupper. Oppgaven baseres derfor til dels på teori på innvandreres empiriske oppførsel på arbeidsmarkedet. Hvorvidt de anses som substitutter eller komplementære til den lokale arbeidskraften vil avgjøre hva vi kan forvente av forskjeller i deres urbane lønnspremier.

### 2.2.1 Agglomerasjonseffekter

Nominelle lønninger er fra klassisk økonomisk teori ansett som et mål på arbeidernes marginalproduktivitet og empirisk observeres det en svært sterk korrelasjon mellom befolkningstetthet og høye lønninger (Glaeser, 2010, s. 3). Noen av de første som beviste at befolkningstettheten forklarer en betydelig andel av variasjonen i arbeideres produktivitet var Ciccone og Hall (1996). De fant at en dobling i sysselsettingstettheten økte produktiviteten med omtrent 6% i USA, ved bruk av mikropaneldata fra 1990.

Det oppsummeres videre av flere studier, deriblant av Combes et al. (2010) at en empirisk står overfor to overordnede endogenitetsutfordringer ved identifiseringen av det kausale forholdet mellom befolkningstetthet og produktiviteten i byene. En kan ha endogent bestemt kvalitet på arbeidskraften der mer produktive arbeidere sorterer seg til de mer urbane områdene, empirisk referert til som individuell endogenitet. I tillegg til at en kan ha endogent bestemt kvantum av arbeidskraft, hvor byer med naturlige fortrinn har produktivitetsfordeler som tiltrekker seg desto flere arbeidere, empirisk referert til som regional endogenitet.

#### Løsninger på individuell endogenitet

Empirisk er det vurdert som viktig å kontrollere for individuelle heterogeniteter mellom individene for å ta hensyn til den endogene årsakssammenhengen mellom agglomerasjonen og produktiviteten i byene. Det argumenteres for at individene i storbyene kan ha heterogene egenskaper og ferdigheter som gjør dem mer produktive og når disse ikke kontrolleres for, overestimeres de urbane lønnspremiene. Bruken av mikrodata og en individuell fixed effects modell er derfor empirisk mye brukt for å løse for dette problemet.

Dette bekreftes blant annet av Glaeser og Maré (2001) som underbygger funnene til Ciccone og Hall og finner en betydelig høy urban lønnspremie på mellom 25-29% i USA ved bruk av mikropaneldata fra 1990-tallet. Estimatenes deres nedjusteres deretter kraftig til rundt 4.5% når de kontrollerer for individuelle heterogeniteter ved bruk av en individuell fixed effects modell. Liknende resultater finnes også i en noe nyere og mer nærliggende studie til denne oppgaven, der Meeke (2021) er den første til å analysere agglomerasjonseffekter og estimere de urbane lønnspremiene i Australia. Ved bruk av mikropaneldata finner Meeke en noe mer moderat urban lønnspremie på

2.7% ved bruk av OLS, men som så nedjusteres til 1.6% ved kontrollering for individuelle fixed effects.

Disse funnene indikerer at det er viktig å kontrollere for individuelle heterogeniteter for å forsøke å ta hensyn til den individuelle endogeniteten der mer produktiv arbeidskraft sorterer seg til de mer urbane områdene. Dette underbygges også av Glaeser og Ellison (1999) som fant at diverse naturlige fortrinn kun forklarte omtrent 20% av agglomerasjonen av industrier i USA i 1987. Det er derfor viktig å ta i bruk mikrodata og kontrollere for individuelle heterogeniteter blant arbeiderne.

Dette utfordres derimot av blant annet De la Roca og Puga (2014) som argumenterer for at viktigheten av den individuelle endogene sorteringen empirisk overvurderes. De viser til funn som tilsier at fordelingen av arbeidere i de store og små byene i USA ikke er spesielt forskjellig med tanke på heterogene uobserverbare egenskaper. De la Roca og Puga sier at den empiriske nedgangen i de estimerte lønnspremiene ved kontrollering for individuelle fixed effects kan forklares av viktige gevinster som kommer fra delingen og tilgangen på akkumulert kunnskap i de mer urbane områdene (Duranton & Puga, 2013, s. 32). Ifølge dem er det gevinstene fra akkumuleringen av kunnskap, kombinert med at dyktigere arbeidere har desto større gevinster av å jobbe i de større byene som forklarer de urbane lønnspremiene.

### **Løsninger på regional endogenitet**

Den andre forklaringen til den simultane årsakssammenhengen mellom befolkningstetthet og produktivitet i de mer urbane områdene, er at områdene i seg selv er mer produktive og derfor tiltrekker seg desto mer arbeidskraft. Dagens storbyer kan ha regional karakteristikkk hvilket gjør dem mer produktive i produksjonen av ulike varer og tjenester, slik at når disse ikke tas hensyn til vil de urbane lønnspremiene overestimere. En empirisk mye brukt løsning på dette er å finne såkalte instrumentelle variabler som argumenteres for å kunne predikere dagens befolkningsmønstre uten å være korrelert med produktivitetsmål som lønn.

Historiske befolkningsmønstre er et veldig populært brukt instrument for dagens befolkningsmønstre fordi det er vist til en sterk korrelasjon mellom tidligere og dagens bosetting, som blant annet kan forklares av holdbarheten til boliger og annen infrastruktur (Combes, Duranton, & Gobillon, 2010, s. 260). Denne strategien er blant annet tatt i bruk av Ciccone og Hall (1996) som nevnt ovenfor, og Combes et al. (2008) som har kopiert deres strategi på franske data og finner

liknende resultater med positiv sammenheng mellom tetthet og produktivitet. Glaeser og Maré (2001) tar i bruk et tilsvarende instrument for å estimere sine urbane lønnspremier, som er foreldrenes fødselssted. De kombinerer bruken av en individuell fixed effects modell med en instrumentell variabel og løser dermed både for den individuelle og regionale endogeniteten. Og de finner som nevnt tidligere en betydelig urban lønnspremie som bekrefter tilstedeværelsen av agglomerasjonseffekter.

Andre nærliknende instrumenter som er mye brukt i empirien er geografisk karakteristik, der en av fordelene med disse er at de er eksogent bestemt, og i stor grad er uavhengig av menneskelig atferd. Fruktbar jord har blitt brukt av blant annet Combes et al. (2010) på franske data, hvor det historisk har vært en stor kilde til økonomisk aktivitet når jordbruk var mer sentralt i økonomien. Historiske gruver er et annet liknende instrument som har blitt brukt av blant annet Glaeser et al. (2008, 2012) på data fra USA, og av Leksnes (2014) på norske data. Sistnevnte brukte historiske gruver mellom det 12- til 19-århundre for å predikere dagens befolkningsmønstre i sin analyse av befolkningstetthet sin påvirkning på livskvalitet. Strategien til Leksnes har blant annet blitt kopiert av Carlsen et al. (2016) som analyserer de statiske og dynamiske agglomerasjonseffektene i norske regioner, og heterogenitetene blant ulike utdanningsnivåer. De finner blant annet at agglomerasjonseffektene er tiltakende med utdanningsnivået til norske arbeidere. I tillegg motbeviser Carlsen et al. de øvrige utsagnene til De la Roca og Puga, ved å vise til at det er betydelige forskjeller i egenskapene til arbeiderne i de norske regionene.

En mulig svakhet ved kontrollering for regional karakteristik er at det kan begrense variasjonsgrunnlaget til analysen betydelig. Meeke (2021) finner at bruken av historiske befolkningsmønstre som instrumentell variabel og kontrollering for regionale forskjeller kan føre til at en får større bias i estimatene. På grunn av begrenset mobilitet og at innbyggerne ikke flytter ofte vil identifiseringen av premiene baseres på desto færre individer og en kan dermed ende opp med å overestimere de urbane lønnspremiene. Han finner at å kontrollere for den individuelle endogeniteten er viktigere enn å kontrollere for den regionale som er i tråd med tidligere litteratur av Combes og Gobillon (2015).



### 2.2.2 Innvandrere på arbeidsmarkedet

Det er større uenighet i tidligere empiri over hvorvidt innvandrere bidrar positivt eller negativt til de lokale arbeidsmarkedsforholdene. Diskusjonen handler i stor grad om hvorvidt de er komplementære eller substitutter for den lokale arbeidskraften.

Økt innvandring har tidligere i stor grad blitt argumentert for å redusere lønningene og begrense sysselsettingsmulighetene til lokalbefolkningen. Dette følger blant annet fra klassisk Keynesiansk teori om at en eksogen økning i arbeidstilbudet naturlig vil øke ledigheten og sette et negativt press på lønningene. Denne negative effekten bekreftes blant annet av Altonji og Card (1991) som fant at en økning i innvandrerandelen med 1% i amerikanske byer, reduserer lønningene til den ufaglærte arbeidskraften med 1.2% på sitt høyeste. Altonji og Card brukte tidligere bosettingsmønstre for innvandrerne som et instrument for å ta hensyn til den endogene sorteringen til byer med høyere lønninger, i likhet med empirien nevnt ovenfor. Dette viser til at innvandrere er ansett som substitutter, der de erstatter den eksisterende lokale arbeidskraften og presser ned lønningene. Hvilket insinuerer at innvandrere ikke bidrar til skalaeffekter i produksjonen i tråd med den monosentriske bymodellen, og vil ikke bidra positivt til agglomerasjonseffektene.

Det er dog satt spørsmål til hvorvidt disse funnene er generaliserbare for andre land med forskjellige arbeidsmarkedsforhold. Pischke og Velling (1994) bruker den samme metoden på tyske data og Friedberg (2001) på Israelske data, hvor begge konkluderer med at det ikke er noen signifikant effekt av økt innvandring på de lokale arbeidsmarkedsforholdene, deriblant lønningene. Disse funnene er i tråd med Friedberg og Hunt (1995) som var noen av de første til å oppsummere tidligere litteratur på temaet. De konkluderte med at effekten av innvandring på de lokale er små, hvor på sitt høyeste finner de at en økning i andelen innvandrere med 10% kun vil redusere de lokale lønningene med 1%. Friedberg og Hunt avslutter med å nevne at retningen og størrelsen på effekten avgjøres av nivået på humankapitalen blant innvandrerne. Dette viser til at innvandrere ikke nødvendigvis er perfekte substitutter til den lokale arbeidskraften, men kanskje vil ha samme innvirkning på agglomerasjonseffektene.

Mye av den nyere litteraturen har satt fokus på komplementariteten innvandrere har til de lokale arbeiderne. Ottaviano og Peri (2006) finner blant annet at innvandrere har en positiv effekt på gjennomsnittslønnen til amerikanske arbeidere i perioden 1990-2004, både på kort- og lang sikt.

De viser til at innvandrere er uperfekte substitutter til den lokale arbeidskraften med samme utdanning, kjønn og erfaring. Dette bekreftes også av Brücker og Jahn (2011) som finner at den lokale arbeidsstyrken tenderer til å dra nytte av økt innvandring gjennom høyere lønninger og lavere ledighet på lang sikt. Grogger og Hanson (2011) viser til at det er de med høyest utdanning som har den høyeste sannsynligheten for å emigrere. Dette bruker Peri (2016) til å argumentere for at innvandrere bidrar til en høyere produktivitet på de lokale arbeidsmarkedene gjennom flere ulike kanaler. Han finner at innvandrerbefolkningen i USA er desto mer konsentrerte enn den øvrige befolkningen. Videre viser han til at det er en betydelig større andel av innvandrere som har høyere utdanning og jobber i yrker som ingeniører og med forskning. Dette er mer høyt betalte yrker som er med på å promotere både innovasjon og læring. I tillegg nevner han at innvandrere tilfører verdi til lokalsamfunnet gjennom økt variasjon av varer og ideer som følger fra ulike kulturelle bakgrunner. Dette kan anses som å øke nivået på fasilitetene og gjøre visse områder desto mer attraktive for bosettelse i tråd med den monosentriske bymodellen. Peri mener at alle disse funnene samlet insinuerer at innvandrere vil kunne bidrar desto mer til agglomerasjonseffektene i økonomien sammenliknet med den øvrige befolkningen.

### **2.3 Oppsummerende forventninger**

Basert på den monosentriske bymodellen forventer vi å finne positive urbane lønnspremier i de større norske regionene sammenliknet med de små, i tråd med tidligere empiri. Videre er det noe usikkert hvordan premiene som indikerer størrelsen på agglomerasjonseffektene, varierer mellom de ulike innvandringsgruppene. Dersom innvandrerne er tilnærmet perfekte substitutter for de lokale arbeiderne vil de ikke bidra positivt til agglomerasjonseffektene i økonomien, og det vil reflekteres gjennom lavere lønnspremier. Er de derimot mer komplementære, betyr det at de vil kunne bidra desto mer til agglomerasjonseffektene i økonomien, og vi vil finne positive ekstra lønnspremier blant de ulike innvandringskategoriene. Vi antar at det er rimelig å tenke seg at innvandrere fra høyinntektsland og norskfødte med innvandrerforeldre i større grad vil være komplementære til de etnisk norske arbeiderne. Innvandrere fra lavinntektsland derimot antas å i større grad typisk å ville ha et lavere nivå på humankapitalen, og i større grad virke mer som substitutter til den ufaglærte lokale arbeidskraften. Dette legger til grunn for at vi forventer positive ekstra lønnspremier blant innvandrerne fra høyinntektsland og de norskfødte med innvandrerforeldre, mens de vil være negative blant innvandrere fra lavinntektsland.

### 3. Datagrunnlag

Dette kapittelet går igjennom dataen som brukes for å undersøke agglomerasjonseffektene i de norske regionene blant etnisk norske og individer med ulik innvandringsbakgrunn. Grunnlaget for analysen er registerdata fra microdata.no, og det brukes et paneldatasett som følger et utvalg på totalt litt over 930 000 individer over 5 år, fra 2013 til 2017. Det vil her i kortere trekk skildres hvordan microdata.no som plattform har blitt brukt, og noen utfordringer det har medført. Deretter gjennomgås variablene som blir brukt i analysen, og tilhørende deskriptiv statistikk.

#### 3.1 Microdata.no

Microdata.no er en noe nyere forskningsplattform som driftes og utvikles i et samarbeid mellom Statistisk sentralbyrå (SSB) og Norsk senter for forskningsdata (NSD). Plattformen gir tilgang til årlige demografiske og sosioøkonomiske skarpe data om befolkningen, deriblant utdanning, inntekt, sysselsetting og trygd på individnivå. Det inkluderer alle individer som har hatt permanent eller midlertidig norsk fødselsnummer, men denne direkte personidentifikasjonen er fjernet fra databasen (Johansen, 2020). For å sikre personvernet til individene er systemet metadatatyrkt, der en hverken kan hente ut eller se direkte på rådataen. Det er kun mulig å få beskrivelser av dataen gjennom diverse funksjonaliteter som deskriptiv statistikk og regresjonsanalyser.

##### 3.1.1 Konfidensialitetstiltak

Personvern og konfidensialitet opprettholdes videre gjennom flere ulike tiltak. Det er blant annet ikke mulig å definere en populasjonsstørrelse med færre enn 1000 personer og det benyttes en 2% winsorisering på rådataen. Det betyr at de høyeste verdiene settes lik verdien på 99-prosentilen, mens de laveste verdiene settes lik 1-prosentilen. Dette fører til at de mest ekstreme observasjonene ikke får like stor innvirkning på de deskriptive resultatene. Vi får en mindre skjevfordeling, men som et resultat vil gjennomsnitt og standardavvik bli noe feilestimert, typisk underestimert (SSB, 2022).

Winsoriseringen vil ikke påvirke regresjonsanalysene som kun benytter de underliggende ikke-winsoriserte dataene, fordi resultatene fra disse analysene ikke betraktes som personidentifiserbar informasjon. I tillegg til winsoriseringen støylegges den deskriptive dataen med  $\pm 5$  individer og

denne støyleggingen er konstant og stokastisk med en forventning lik null. Dette vil ikke påvirke analysen betydelig ettersom antall observasjoner i denne analysen er såpass høyt.

### **3.1.2 Begrensninger med programmet**

Det er som nevnt ikke mulig å se selve datamatrixene i programmet og en kan heller ikke laste ned dataen, hvilket betyr at all analyse må gjøres direkte i programmet. Resultatene fra disse er deretter eksportert til Excel for videre manuell bearbeiding.

Utvalget av relevante variabler tilgjengelig i programmet som er aktuelle for denne analysen er noe begrenset. Det hadde for eksempel vært ønskelig å se på befolkningsstørrelse eller tettheten som kontinuerlige variabler, for å kunne beregne spesifikke elastisiteter i tråd med tidligere empiri. Disse variablene er hverken tilgjengelige eller mulig å generere i programmet per dags dato. Variabler som karakteriserer arbeidsforholdene til individet, er det også vanskelig å kontrollere for i et paneldatasett i programmet.

Det er heller ikke mulig å koble på ekstern data i programmet enda og dette kombinert med den begrensede tilgjengeligheten på relevante variabler gjør at det er tilnærmet umulig å finne gode proxy- og instrumentelle variabler for analysen. Dette er en av de største utfordringene oppgaven står overfor. Et resultat av dette er at det ikke er mulig å løse for den mulige regionale endogenitet i tråd med tidligere empiri i denne oppgaven og de økonometriske utfordringene dette medfører samt flere gjennomgås mer i detalj i neste kapittel.

Den siste nevneverdige svakheten med programmet er at det er begrenset med funksjonaliteter, spesielt knyttet til bruken av paneldata. Det er blant annet ikke mulig å generere nye variabler i paneldata, men kun kode om de som er tilgjengelige fra databanken og her er det også flere begrensninger. Blant annet er det foreløpig ikke implementert funksjonaliteter knyttet til lag/lead av variabler med hensyn på tid, hvilket gjør det vanskelig å sammenlikne på tvers av observasjoner som vil diskuteres nærmere senere i oppgaven.

## **3.2 Avgrensing av utvalget**

I microdata.no anbefales det å avgrense utvalgspopulasjonen til maks rundt 1 million individer før en går over til et paneldatasett for å ikke overbelaste programmet. Dette har jeg gjort ved å ta utgangspunkt i en hedonisk lønnslikning i tråd med artikkelen til Carlsen et al. (2016). I

utgangspunktet var hele populasjonen på litt over 5.3 millioner individer. Avgrensning til kun de mellom 25 til 67 år ekskluderer de fleste studenter og de i pensjonsalder, og dette fjerner litt over 2.1 millioner fra utvalget.

Deretter ble utvalget avgrenset til kun de som er etablerte i en 100% fulltidsstilling under hele observasjonsperioden. Alle som ikke har en fulltidsstilling under hele perioden ekskluderes, og dette fjerner litt over 2 millioner individer fra utvalget, men microdata.no har kun registerdata på individenes sysselsettingsandelen fra og med 2015. For å ekskludere flest mulig som ikke har en fulltidsstilling i årene 2013 og 2014 fjernes alle som har en registrert årslønn på under 286 000NOK disse to årene og dette ekskluderer litt over 73 000 individer. Denne grensen er satt basert på det regjeringen oppgir som det laveste lønnstrinnet i offentlig sektor i 2013, der lønnstrinn 19 hadde en brutto årslønn på 286 200NOK (Kommunal- og distriktsdepartementet, 2013). Dette vil ekskludere de aller fleste som kun har en deltidsstilling disse to årene, men det er ikke gjort noen avgrensning for de som jobber mye overtid disse to årene fordi det er vanskeligere å luke ut. Analysen ser altså kun på de som er etablerte på arbeidsmarkedet, og ser bort ifra nyutdannede og andre som kommer inn på markedet i tillegg til de som forlater det i løpet av perioden.

Disse avgrensningene antas som hensiktsmessige for å kunne estimere de mest presise urbane lønnspremiene, men det gjøres på bekostningen av generaliserbarheten til resultatene. De kan føre til at det oppstår skjevheter i utvalget som vil gjør det mindre representativt for resten av befolkningen, og utfordringene med dette diskuteres og undersøkes nøyere i slutten av dette kapittelet.

### **3.3 Variabler**

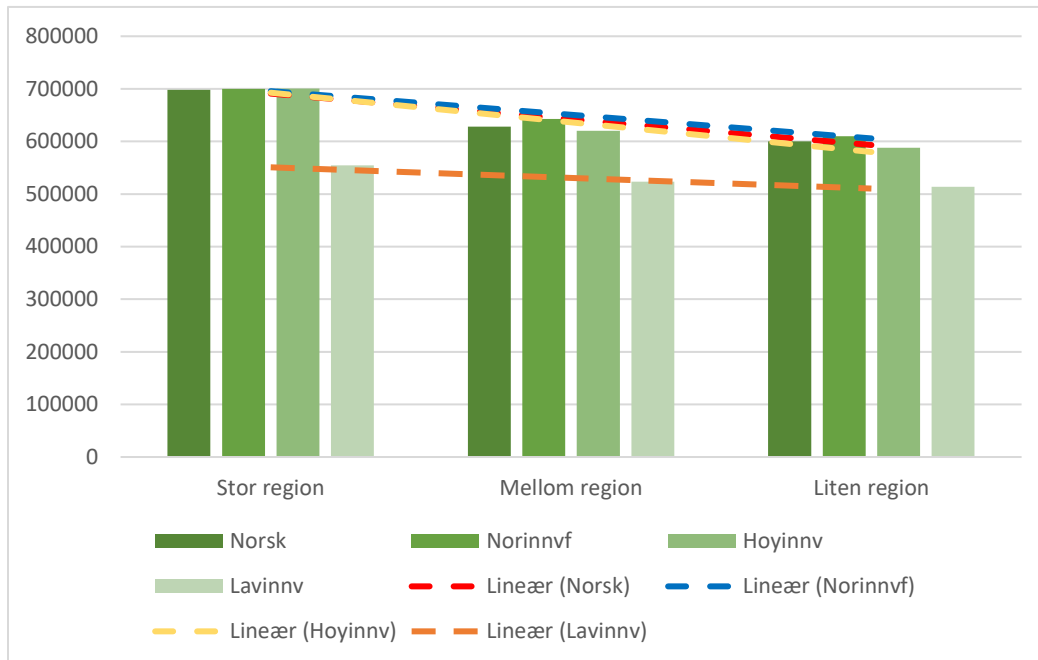
Hovedfokuset til analysen er å se på heterogeniteten til de urbane lønnspremiene, et mål på agglomerasjonseffektene, blant individer med ulik innvandringsbakgrunn. Oppgaven tar inspirasjon fra en nyere artikkel av Røed et al. (2021) som analyserer hvordan innvandrere påvirker den økonomiske mobiliteten og de lokale arbeidsmarkedsforholdene i Norge. De skiller mellom innvandrere fra høy- og lavinntektsland, og vi ser i tråd med dette på tre ulike innvandringskategorier sammenliknet med de etnisk norske blir referansekategori til analysen.

<b>Innvandringskategori</b>	<b>Variabelnavn</b>
- Etnisk norske	<i>norsk</i>
- Norskfødte med innvandrereforeldre	<i>norinnvf</i>
- Innvandrere fra lavinntektsland <ul style="list-style-type: none"><li>○ Østeuropeiskeland, Afrika, Asia m/ Tyrkia, Sør- og Mellom-Amerika</li></ul>	<i>lavinnv</i>
- Innvandrere fra høyinntektsland <ul style="list-style-type: none"><li>○ De resterende EU/EØS-landene, Nord-Amerika, Oseania</li></ul>	<i>hoyinnv</i>

Samlet vil de tre siste kategoriene refereres til som personer med innvandringsbakgrunn, når de sammenliknes med de etnisk norske som er den øvrige befolkningen.

### **3.3.1 Avhengig variabel**

Den avhengige variabelen for analysen er nominell årslønn, der lønn som nevnt tidligere er et vanlig brukt mål på agglomerasjonseffekter, blant annet illustrert av likning (3) i forrige kapittel. Lønnsinntektene omfatter kontantlønn, skattepliktige naturytelser og syke- og fødselspenger (Microdata.no, u.d.). En log-transformering av variabelen brukes slik at den blir mindre venstreskjev og mer normalfordelt, illustrert grafisk i appendiks A2. Dette sikrer at variabelen bedre vil oppfylle kravene til den lineære regresjonsmodellen, samt at estimatene blir mindre sensitive for eventuelle ekstremverdier (Wooldridge, 2016, s. 172).

**Figur 1:** Gjennomsnittlig årslønn for innvandringskategoriene i regionene

Fra figur 1 ser en i tråd med empirien at både etnisk norske og norskfødte med innvandrerforeldre har en gjennomgående høyere gjennomsnittslønn enn innvandrere, både fra høy- og lavinntektsland, i alle de økonomiske regionene. I tillegg til dette ser vi det første tegnet på tilstedeværelsen av agglomerasjonseffekter. Gjennomsnittslønnen for alle individene er avtakende med befolkningsstørrelsen til de økonomiske regionene, illustrert av de fallende trendlinjene i figuren. Dette bekreftes også fra de oppgitte gjennomsnittsverdiene nedenfor i tabell 1, der en ser at gjennomsnittslønnen i de små regionene er nesten 100 000NOK lavere enn i de store. Dette gir et førsteinntrykk av at vi muligens har positive urbane lønnspremier av å bo i de større økonomiske regionene i Norge.

**Tabell 1:** Deskriptiv statistikk for årslønn hos individene og regionene

ÅRSLØNN	Obs.	Gj.Snitt	Std.avvik	1%	99%
Hele utvalget	4 622 302	647 247	269 981	321 807	1 856 267
Norske	4 041 302	651 035	269 399	326 629	1 858 097
Norinnvf	183 342	670 090	323 099	332 994	2 015 332
Høyinnv	154 817	660 393	301 088	308 633	2 036 825
Lavinnv	190 218	543 265	206 794	294 151	1 425 406
Store regioner	2 264 124	691 774	315 048	333 881	2 151 780
Mellom regioner	1 119 228	626 227	238 445	331 764	1 628 713
Små regioner	1 326 058	599 443	210 021	328 302	1 477 667

Fra tabellen får vi også bekreftet muligheten for større heterogeniteter mellom de ulike innvandringskategoriene. Norskfødte med innvandrerforeldre i utvalget rapporteres å ha en snittlønn som er nesten 20 000NOK høyere enn de etnisk norske i året. En gjennomsnittlig innvandrer fra et høyinntektsland i utvalget vil i snitt tjene over 117 000NOK mer i året enn en gjennomsnittlig innvandrer fra et lavinntektsland. Følgelig ser vi også fra standardavvikene i tabellen at lønnsvariasjonen naturlig nok øker med gjennomsnittslønnen til de ulike kategoriene.

### 3.3.2 Forklaringsvariabler

Forklaringsvariablene av hovedinteresse for oppgaven er de økonomiske regionenes befolkningsstørrelse og interaksjoner av disse med individers innvandringsbakgrunn. Oppgaven bruker de økonomiske regionene fordi de er definert etter kriterier om arbeidsmarkeds- og økonomiske forhold (Longva & Tønder, 2000).

Regionene er definert av SSB som et nivå mellom fylke og kommune, og omfatter totalt 85 enheter. Denne oppgaven tar utgangspunkt i en tredeling, igjen i tråd med artikkelen til Carlsen et al. (2016), der de har definert tre kategorier for regionstørrelsene:

- Store regioner; med over 150 000 innbyggere
- Mellomstore regioner; med mellom 65-150 000 innbyggere
- Små regioner; med mindre enn 65 000 innbyggere



Definisjonene på regionene kombinert med data på innbyggertall fra 2013 hentet fra SSBs databank gjør at vi totalt definerer 7 store-, 14 mellomstore- og 64 små regioner. Det er dog ikke en egen variabel for de økonomiske regionene i microdata.no. Istedenfor summeres det over kommunene i tråd med kommunefordelingen i de økonomiske regionene fra SSB sin rapport «Standarder for økonomiske regioner» fra år 2000. En samlet oversikt over inndelingen av kommunene og kategoriseringen av regionene er vedlagt i appendiks A3.

**Tabell 2:** Gj.snitt bosettingsandeler til utvalget

FORDELING UTVALGET	Totalt	Norske	Norinnvf	Hoyinnv	Lavinnv
Stor region	49%	47%	60%	60%	70%
Mellom region	23%	24%	23%	20%	17%
Liten region	28%	29%	17%	20%	13%

Fra tabell 2 oppgis det at i løpet av perioden er omtrent halvparten av utvalget bosatt i de store økonomiske regionene, mens den andre halvparten er noe likt fordelt mellom de mellomstore- og små regionene. Det er derimot en desto større andel av alle individene med ulik innvandringsbakgrunn som er bosatt i de store økonomiske regionene sammenliknet med de etnisk norske. Der 60% av alle norskfødte med innvandrerforeldre og innvandrere fra høyinntektsland bor i de store økonomiske regionene, mens hele 70% av alle innvandrere fra lavinntektsland er bosatt der. Dette er et gjenkjennbart mønster som er som forventet i tråd med empirien, og funnene til Peri (2016).

### 3.3.3 Kontrollvariabler

Analysen inkluderer et sett med kontrollvariabler for å minimere problemer med utelatte relevante variabler som fører til endogenitet dersom de påvirker individers lønn, samtidig som de påvirker inkluderte forklaringsvariabler.

**Tabell 3:** Deskriptiv statistikk kontrollvariablene

GJ.SNITT	Alder	Hoyutd	Kvinne	Gift	Barn
Totalt	46	0.50	0.40	0.52	1.12
Norske	46	0.49	0.49	0.51	1.12
Norinnvf	43	0.56	0.56	0.49	1.10
Hoyinnv	44	0.65	0.39	0.53	1.03
Lavinnv	42	0.50	0.41	0.68	1.15
Stor region	45	0.56	0.42	0.42	1.07
Mellom region	46	0.45	0.39	0.55	1.15
Liten region	47	0.42	0.39	0.52	1.18

### Alder

Ansiennitet og erfaring er en viktig forklaringsfaktor for individers lønn, og alder er ansett som en god proxy variabel for dette. Alder er anerkjent positivt korrelert med lønn, der oppbyggingen av ansiennitet og erfaring fører til at en kan kvalifisere seg til mer høytlønnede stillinger (Askvik, 2020). Snittalderen til det samlede utvalget er på 46 år og det er ikke store avvik fra dette mellom regionene, men et noe gjennomgående lavere snitt på mellom 42-44 år blant individene med ulik innvandringsbakgrunn.

### Utdanning

Empirisk har en sett at utdanning spiller en avgjørende rolle spesielt for agglomerasjonseffektene fordi de høyere utdannede individene bidrar til å forsterke gevinstene fra agglomerasjonen og det er direkte korrelert med deres lønnsnivå (Bolter & Robey, 2020). Oppgaven kontrollerer derfor for individenes høyeste fullførte utdanningsnivå. Opplysningene om norsk utdanning bygger på informasjon fra det norske utdanningssystemet og Nasjonal vitnemålsdatabase (NVB), mens informasjon om utdanning fullført i utlandet suppleres fra Lånkassen (Rustad Holseter, 2019). Variabelen for individenes utdanningsnivå i microdata.no er gruppert etter norsk standard for utdanningsgruppering (NUS).

Den alfanumeriske variabelen er her kodet om til en dummy som indikerer hvorvidt individet har høy utdanning, her definert som bachelornivå eller høyere. I tabell 3 oppgis det at litt under halvparten av utvalget har høyere utdanning, men det er noe større variasjon mellom de ulike innvandringskategoriene. Opp til 56% av alle norskfødte med innvandrerforeldre, og 65% av alle innvandrere fra høyinntektsland oppgis å ha høyere utdanning. Dette er også noe som forventet i tråd med empirien som nevnt tidligere av Grogger og Hanson (2011) med at det er de med høyere utdanning som har en høyere sannsynlighet for å emigrere. I tillegg ser en fra tabellen at andelen med høyere utdanning avtar med befolkningsstørrelsen på de økonomiske regionene. Dette er som forventet og insinuerer tegn på at vi har en endogen sortering av «flinkere» arbeidere til de mer urbane områdene.

Det er dog et mulig problem med denne variabelen at kravene til hva høyere utdanning innebærer kan være noe forskjellig for norske og utenlandske utdanningsinstitusjoner. Kravene for en mastergrad her i Norge kan være vidt forskjellig fra kravene i særlig mindre utviklede land, hvor utdanningsgrunnlaget generelt er lavere f.eks. Dette kan være en kilde til målefeil og denne utfordringen diskuteres noe videre senere i oppgaven.

#### **Kjønn**

Empirisk tjener menn mer enn kvinner og det vil derfor kontrolleres for kjønn i analysen, gjennom en dummyvariabel som indikerer hvorvidt et individ identifiserer seg som kvinne eller ikke. Tabell 3 oppgir en noe gjennomgående overvekt av menn i utvalget, men igjen med noe større variasjon mellom innvandringskategoriene. Blant de norskfødte med innvandrerforeldre er det en liten overvekt av kvinner, mens blant innvandrere fra høy- og lavinntektsland er det betydelig flere menn. Disse skjevfordelingene i forhold til det vi ser i populasjonen vil gjøre at sannsynligheten er høyere for at de estimerte premiene er mer presise for det overrepresenterte kjønn i gruppen.

#### **Sivilstatus og barn**

Til slutt kontrolleres det også for individenes sivilstatus og antall barn i husholdningen. Sivilstatusen er omkodet til en dummyvariabel for hvorvidt individet er gift eller ikke, og antall barn er en kontinuerlig variabel. Omtrent halvparten av utvalget oppgis å være gift og i snitt ha 1 barn, uten noen betydelig stor variasjon mellom innvandringskategoriene eller regionene. Unntaket er innvandrere fra lavinntektsland der en betydelig større andel nærmere 70% oppgis å være gift.

### 3.4 utfordringer og begrensninger med datasettet

Det er flere utfordringer med datasettet som må tas med i betraktning under den videre analysen og tolkningen av resultatene. Manglende data og avgrensingen av utvalget fører til betydelige skjevheter i utvalget som må tas høyde for ved tolkningen og generaliserbarheten til analysen.

#### 3.4.1 Manglende data

Datagrunnlaget for utdanningsnivået til individene er noe svakt for innvandrere der det mangler opplysninger om et betydelig antall. Det mangler informasjon om utdanningen til totalt nesten 9 000 individer, hvorav 86% av disse er innvandrere, oppgitt i tabell 4.

**Tabell 4:** Andelen savnede utdanningsverdier

	Antall	Andel av totalen	Andel av egen gruppe
Totalt	8 955	1 %	-
Norsk	533	6 %	0 %
Norinnvf	55	1 %	0 %
Hoyinnv	6 397	71 %	14 %
Lavinnv	1 314	15 %	5 %

De manglende utdanningsverdiene til individene utgjør en noe ubetydelig andel av sine egne grupper for de etnisk norske og norskfødte med innvandrerforeldre, mens 14% av alle innvandrere fra høyinntektsland mangler informasjon om utdanningsnivået sitt.

Individene med manglende utdanningsverdier blir fjernet fra utvalget, og dette blir derfor en kilde til at vi får skjevheter som gjør utvalget mindre representativt for befolkningen.

#### 3.4.2 Seleksjonsskjevhet

Avgrensingen av utvalget fører til at det er en høyere sannsynlighet for at vi får seleksjonsskjevheter i utvalget, deriblant spesielt i forhold til hvor representativt det vil være for hele befolkningen.

Ekskluderingen av alle individer som ikke har en fulltidsstilling slår spesielt skjevt ut på andelen innvandrere som forblir i utvalget. Det er en betydelig mindre andel blant innvandrerne, spesielt de fra lavinntektsland som er ansatt i en fulltidsstilling hele perioden sammenliknet med de etnisk norske og norske med innvandrerforeldre.

**Tabell 5:** Andelen av utvalget med fulltidsstilling hele perioden

FULLTID	Andel av totalen	Andel av egen gruppe
Norske	85%	38%
Norinnvf	4%	35%
Hoyinnv	6%	24%
Lavinnv	3%	15%

Etter å ha avgrenset utvalget til de innenfor det arbeidsdyktige aldersintervallet er det kun 35% av disse som har en fulltidsstilling gjennom hele perioden fra 2013 til 2017. Fra tabell 5 ser en at denne andelen er noe lik innad blant de etnisk norske og de norskfødte med innvandrerforeldre. Andelen er derimot betydelig lavere blant innvandrerne, hvor kun 24% og 15% av hhv. alle innvandrere fra høy- og lavinntektsland har en fulltidsstilling under hele perioden.

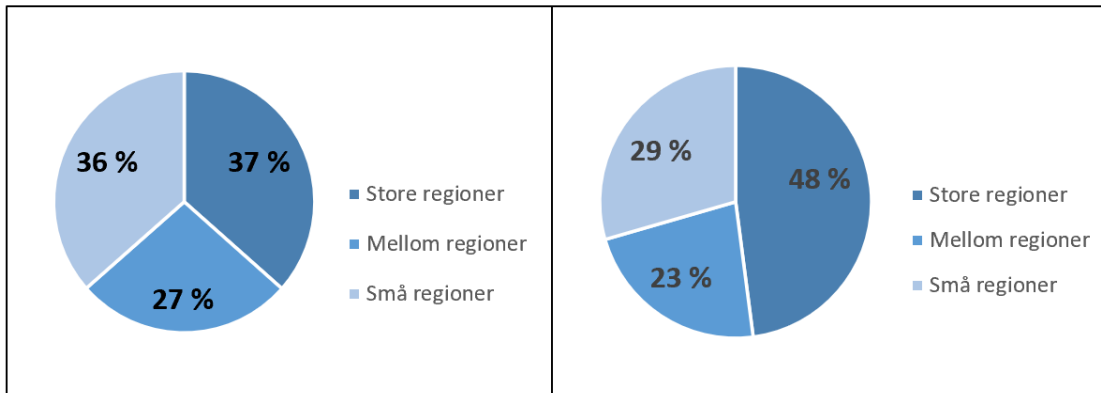
**Tabell 6:** Innvandringskategoriernes fordeling i populasjonen og utvalget

	Populasjonen	Utvalget
Norske	78%	89%
Norinnvf	8%	4%
Hoyinnv	6%	3%
Lavinnv	6%	4%

Denne avgrensingen i tillegg til de andre forklart tidligere, fører til en vesentlig overvekt av nordmenn i utvalget. Fra tabell 6 oppgis det at de etnisk norske utgjør 11% mer av utvalget sammenliknet med det de utgjør av hele befolkningen og dette går på bekostningen av de andre innvandringskategoriene. Individer med innvandringsbakgrunn utgjør omtrent en halvparten så stor andel av utvalget som de gjør av populasjonen, altså er de betydelig underrepresenterte i utvalget.

Utvalget preges også av større skjevheter i forhold til fordelingen mellom regionene. Avgrensingen av utvalget fører til en overvekt av individer som bor i de store økonomiske regionene illustrert i figur 2, som vi ser at i stor grad går på bekostningen av de som bor i de mellomstore. I figuren sammenliknes fordelingen mellom regionene til utvalget og befolkningen basert på bosettingen i 2013.

**Figur 2:** Andelen av befolkningen (V.S) og utvalget (H.S) i regionene



Disse overvektene av etniske nordmenn og individer bosatt i de store økonomiske regionene i utvalget i forhold til den faktiske fordelingen i befolkningen, resulterer i skjevheter som vil kunne gjøre resultatene mindre generaliserbare. De estimerte lønnspremiene basert på utvalget vil på grunn av dette muligens ikke være like representative for hele befolkningen, spesielt for de gruppene som er underrepresenterte i utvalget.

Likevel er disse avgrensningene ansett som nødvendige forenklinger for å kunne estimere de mest presise lønnspremiene. Resultatene fra analysen vil derfor kun regnes som relevante for de som er aktive og ansatt i en 100% stilling under hele perioden og ikke nødvendigvis kunne generaliseres til resten av befolkningen.

## 4. Metode

Denne delen av oppgaven presenterer modellspesifikasjonen, og går igjennom det metodiske rammeverket som brukes for analysen av agglomerasjonseffekter blant individer med ulik innvandringsbakgrunn. Til slutt diskuteres noen økonometriske utfordringer i kortere trekk.

### 4.1 Modellspesifikasjon

For å estimere de urbane lønnspremiene som et mål på agglomerasjonseffekter blant individer med ulik innvandringsbakgrunn tas det utgangspunkt i en standard Mincer-inspirert empirisk modell.

$$(4) \ln y_{ict} = \beta_0 + \beta_1 \text{storregion}_{ct} + \beta_2 \text{mellomregion}_{ct} + \beta_3 \text{innvbak}_i + \gamma_1 \text{innvbak}_i * \text{storregion}_{ct} + \gamma_2 \text{innvbak}_i * \text{mellomregion}_{ct} + \delta_1 \mathbf{X}'_{it} + D_t + \alpha_i + u_{ict}$$

Underbetegnelse  $i$ ,  $c$  og  $t$  refererer hhv. til individet, regionstørrelsen og året. Vektoren  $\mathbf{X}'$  består av kontrollvariablene for diverse individkarakteristikk som kjønn, alder o.l. som ble presentert i forrige kapittel. Fra den monosentriske bymodellen og tidligere empiri forventer vi positive estimater på koeffisientene  $\beta_1 > 0$  og  $\beta_2 > 0$ . Det betyr at vi har positive urbane lønnspremier i hhv. de store og mellomstore økonomiske regionene i Norge.

Hovedfokuset til denne oppgaven er identifiseringen av koeffisientene  $\gamma_1$  og  $\gamma_2$  som sier noe om heterogeniteten til de urbane lønnspremiene blant individer med ulik innvandringsbakgrunn. Her var teorien og empirien noe mer blandet avhengig av om innvandrerne er substitutter eller komplementære til den lokale arbeidsstyrken. Vi forventer som nevnt tidligere at innvandrere fra lavinntektsland vil kunne ha negative ekstra premier  $\gamma_1, \gamma_2 < 0$ , altså at de bidrar mindre til agglomerasjonseffektene sammenliknet med den øvrige befolkningen. Mens innvandrere fra høyinntektsland, samt norskfødte med innvandrerforeldre antas å være mer komplementære, og vil derfor kunne ha positive ekstra premier  $\gamma_1, \gamma_2 > 0$ .

Det kontrolleres også for årlige faste effekter gjennom  $D_t$ , som fanger opp all variasjonen i lønnen som følger fra eksempelvis konjunkturerendringer og inflasjon.

Ved bruk av paneldata har vi to restledd. Det idiosynkratiske restleddet  $u_{ict}$  fanger opp individspesifikke variabler og egenartede sjokk som endres over tid, mens  $\alpha_i$  fanger opp individspesifikke

uobserverbare heterogeniteter som er konstante over tid. Det sistnevnte omfatter blant annet det som kalles for abilities blant arbeiderne som kan anses som synonymt med medfødte egenskaper.

## 4.2 Estimeringsmetoder

Paneldata har en enhets- og en tidsdimensjon, og datasettet denne analysen bygger på observerer som nevnt tidligere individer som enhetene, over tiden målt i år. Denne oppgaven vil hovedsakelig ta i bruk to ulike estimeringsstrategier for å forsøke å estimere likning (4). Metodene og de økonomiske utfordringene kan i enklere trekk forklares med utgangspunkt i en generell multippel regresjonsmodell gitt ved:

$$(5) \quad y_{it} = \beta_0 + \beta_1 x_{it1} + \dots + \beta_k x_{itk} + v_{it} \quad , \quad v_{it} = a_i + u_{it}$$

Ved bruk av minstekvadraters metode (OLS), estimeres den kausale årsakssammenhengen mellom variablene ved å minimere summen av residualene som er de kvadrerte avvikene fra gjennomsnittsverdiene til de uavhengige variablene (Wooldridge, 2016, s. 28).

For at estimatoren skal være forventningsrett og konsistent, må blant annet det som kalles for eksogenitetsbetingelsen være oppfylt. Betingelsen sier at restleddet har en forventet verdi lik null og at det må være ukorrelet med alle de inkluderte forklaringsvariablene, altså  $E(v_{it}|x_j) = 0$ . Det er rimelig å anta at dette er en svært streng og urealistisk antakelse. Sannsynligheten er høy for at det er diverse uobserverbare heterogeniteter blant individene som er korrelert med én eller flere av de uavhengige variablene. I denne analysen spesifikt, kan individkarakteristikk som IQ og andre egenskaper påvirke bosettingsmønstrene og lønningene til individene som gjør at når disse ikke kontrolleres for får en heterogenitets bias i estimatene (Wooldridge, 2016, s. 413). Disse individuelle heterogenitetene er som nevnt tidligere i kapittel 2 empirisk en av hovedutfordringene for identifiseringen av de kausale agglomerasjonseffektene.

Fixed effects er hovedstrategien til oppgaven, og det er en estimeringsmetode på paneldata som transformerer regresjonslikningen til tidsgjennomsnittlige avvik for hver variabel:

$$(4.3) \quad y_{it} - \bar{y}_i = \beta_1(x_{it1} - \bar{x}_{i1}) + \dots + \beta_k(x_{itk} - \bar{x}_{ik}) + (u_{it} - \bar{u}_i)$$

Alle variabler som er konstante over tid faller dermed bort, og dette fører da også til at det indvidspesifikke restleddet  $a_i$  fjernes. Fixed effects tilnærmingen er som gjennomgått tidligere empirisk mye brukt fordi det eliminerer de fleste problemer som følger fra uobserverbare



heterogeniteter. Det er spesielt nyttig her fordi det i stor grad kontrollerer for problemene som følger fra den potensielle individuelle endogeniteten som ble forklart tidligere i kapittel 2.

Oppgaven tar derfor i tråd med tidligere empiri som Meeke (2021) i bruk en individuell fixed effects modell som hovedstrategi for å forsøke å estimere likning (4). Det innebærer her at vi vil estimere de urbane lønnspremiene basert på individer i utvalget som flytter i løpet av perioden. Personene i utvalget som ikke flytter mellom de tre regionkategoriene vil ha en konstant verdi og dermed falle bort ved bruk av fixed effects som metode. En utfordring med dette er at et av kravene for at fixed effects estimatoren skal være gyldig er at en har nok variasjon. Det betyr at ved bruk av individuell fixed effects modell må vi ha nok variasjon innenfor individene i løpet av perioden for at de estimerte lønnspremiene skal være forventningsrette og konsistente. Vi må ha et tilstrekkelig antall individer som flytter og som også opplever variasjon i lønningene sine i løpet av perioden. Det er dog ingen fast regel på hva nok variasjon er, men det kan kommenteres på i relative sammenhenger som vi kommer tilbake til senere i oppgaven.

### **4.3 Økonometriske utfordringer**

De tre overordnede teoretiske kildene til endogenitet og skjevheter i analysen gjennomgås her. Bruken av en individuell fixed effects modell i hoved analysen gjør at de økonometriske utfordringene i all hovedsak vil være knyttet til brudd på eksogenitetsbetingelsen koblet til det idiosynkratiske restleddet.

#### **4.3.1 Målefeil**

Dersom det er et avvik mellom de observerte og sanne verdiene til en variabel har en det som kalles for målefeil. I den avhengige variabelen gir målefeil normalt ikke opphav til skjevhet i estimatoren så lenge feilen er ukorrelert med de uavhengige variablene, men det kan føre til høyere standardavvik og varians (Wooldridge, 2016, s. 288). Det er generelt ansett som et større problem dersom en har målefeil i de uavhengige variablene, men det avhenger av hvorvidt målefeilen er avhengig av de observerte verdiene eller ikke. Hvis målefeilen er uavhengig av den observerte verdien vil vi i likhet med målefeil i den avhengige variabelen fortsatt ha en forventningsrett estimator, men med høyere varians. Dersom målefeilen er korrelert med de observerte verdiene derimot vil vi få partiske regresjonsresultater, hvor estimatoren vil være inkonsistent og partisk mot null (Wooldridge, 2016, s. 291).

Vi kan ha målefeil i den avhengige variabelen i analysen dersom det er avvik mellom innrapportert årslønn og faktisk årslønn. Individuer kan jobbe mye overtid eller det kan være andre forhold som gjør at lønnen avviker fra det som oppgis i arbeidskontrakten. Det er vurdert som rimelig å anta at dette ikke vil påvirke analysen betydelig så lenge ikke avvikene er systematiske og selv da vil det kun føre til at vi får høyere varians.

Systematiske målefeil i noen av de uavhengige variablene er som sagt mer problematisk og det mest bekymringsverdige for denne analysen er mulige målefeil i kategoriseringen av de økonomiske regionene. Kategoriseringen av regionene er som forklart i kapittel 3 gjort basert på innbyggertall fra statistikkbanken til SSB og disse innbyggertallene vil variere noe i løpet av perioden vi ser på. Det er spesielt én økonomiske regioner som er kategorisert med noe skjønn. Regionen Molde har blitt definert som en liten region i datasettet, fordi de kun har registrert rundt 64 000 innbyggere i perioden 2013-2016, men de bikker over grensen på 65 000 i første kvartal i 2016. Det er likevel ansett som mest hensiktsmessig å definere den som en liten region fordi den er under grensen over halvparten av tidsperioden vi ser på.

Målefeil vil prege de fleste datasett, men det er ikke noe som tilsier at det burde være av vesentlig stor bekymring her og vurderes derfor å i liten grad ville påvirke estimatorene til denne analysen..

### **4.3.2 Utelatte relevante variabler**

Utelatelse av relevante forklaringsvariabler fra modellen kan føre til forventningsskjev estimater, men kun dersom de er korrelert med noen av de inkluderte forklaringsvariablene (Wooldridge, 2016, s. 81). Dette bruddet på eksogenitetsbetingelsen fører til at estimatene vil over- eller underestimeres, avhengig av retningene på korrelasjonene.

Bruken av en individuell fixed effects modell reduserer som sagt omfanget til denne utelatte variabelskjevheten fordi det fjerner det heterogene restleddet, og alle tidsfastevriabler fra det idiosynkratiske restleddet. Til tross for dette er modellen fortsatt høyst sannsynlig underspesifisert. Tilgangen på et begrenset utvalg variabler fra microdata.no gjør at vi ikke får kontrollert for mye individuell karakteristikk, hvor det også er vanskelig å kontrollere for individkarakteristikk som sektorer og yrker i programmet. Slike variabler påvirker høyst sannsynlig både lønnsnivået og bosettingsmønsteret til individene, som betyr at utelatelsen fører til skjevheter. En mulig løsning på denne underspesifiseringen som oppgaven har forsøkt å undersøke er å inkludere laggede

versjoner av den avhengige variabelen i regresjonslikningen (4). Denne variabelen ville fanget opp variabler som påvirker tidligere lønninger og gjennom denne også dagens. Dette ville potensielt kunne rettet opp i noe av skjevheten, men funksjonen for å lagge variabler er som nevnt tidligere ikke tilgjengelig i microdata.no per dags dato.

Det er gjort et forsøk på å teste for potensiell feilspesifisering av modellen for analysen gjennom en RESET-test som er ansett som en generell test for å undersøke dette. Testen finner at modellen er korrekt spesifisert innenfor et 1% signifikansnivå, vedlagt i appendiks A4. Dette burde derimot kun tolkes som at funksjonsformen til modellen er korrekt spesifisert, og vi vil fortsatt høyst sannsynlig ha problemer med utelatte relevante variabler.

### 4.3.3 Simultanitet

Den simultane årsakssammenhengen mellom befolkningstettheten og lønn er en velanerkjent kilde til endogenitet og har empirisk blitt forsøkt løst på flere ulike måter som forklart tidligere i kapittel 2. Her er det en gjensidig påvirkning mellom avhengig og forklaringsvariablene av interesse som resulterer i at forklaringsvariablene er korrelert med restleddet og vi har brudd på eksogenitetsbetingelsen.

Endogeniteten som følger av denne toveisårsakssammenhengen kan som sagt forklares på både individ- og regionnivå. På individnivå har det blitt nevnt tidligere at det empirisk er identifiseringsproblemer som følger fra en mulig endogen sortering av mer produktive arbeidere til mer tettbygde strøk. Denne endogene sorteringen vil i tråd med tidligere empiri i stor grad kontrolleres for ved bruk av en individuell fixed effects modell. Endogeniteten på regionnivå skyldes derimot at det muligens er regionkarakteristikk som gir byene produktivetsfortrinn og tiltrekker seg desto flere arbeidere. Det er ingen aktuelle relevante variabler tilgjengelig i microdata.no per dagsdato som kunne blitt brukt for å instrumentere dagens befolkningsmønstre, og dermed løst for den regionale endogeniteten. Instrumentet måtte i tillegg til å være relevant for å forklare dagens befolkningsmønstre også oppfylt eksogenitetsbetingelsen. Den sier at variabelen ikke direkte påvirker produktivetsmål som lønningene, men kun påvirker de gjennom dagens bosettingsmønstre. Det er essensielt at denne betingelsen er oppfylt for at strategien skal være gyldig, men det er empirisk vanskelig å bevise. Typisk pleier det kun å argumenteres for at betingelsen med høy sannsynlighet er oppfylt ut ifra et teoretisk grunnlag.

Når den regionale endogeniteten ikke løses for vil de urbane lønnspremiene typisk overestimeres, og en kort analytisk utledning av dette er vedlagt i appendiks A5. Det er en stor svakhet ved analysen at det ikke er mulig å forsøke å løse for dette, og utfordringene med dette vil diskuteres noe videre mot slutten av oppgaven i kapittel 7.

#### 4.3.4 Multikollinearitet

Multikollinearitet betyr at en har høy, men ikke perfekt korrelasjon mellom to eller flere uavhengige variabler (Wooldridge, 2016, s. 84). Dersom analysen preges av multikollinearitet vil det ikke føre til skjevheter i estimatene, men kun til høyere varians som heller ikke er ønskelig fordi det kan føre til at estimatene blir mindre statistisk signifikante. Korrelasjonsmatrisen vedlagt i appendiks A6 gir et inntrykk av at det ikke er noen bekymringsverdige høye korrelasjoner mellom de inkluderte forklaringsvariablene i analysen. Dette bekreftes også av en VIF-test, vedlagt i appendiks A6 som oppgir en gjennomsnittlig VIF-verdi på 2.5 noe som er godt under den noe omdiskuterte grensen på 10. Alt i alt er det rimelig å anta at multikollinearitet ikke vil påvirke estimatene i denne analysen.

#### 4.3.5 Heteroskedastisitet og seriekorrelasjon

For at vi skal ha en forventningsrett og konsistent estimator er det også viktig at vi ikke har brudd på betingelsene om homoskedastisitet og ingen seriekorrelasjon i restleddet. Heteroskedastisitet betyr at restleddet ikke har en konstant varians,  $var(u_{it}|\mathbf{x}) \neq \sigma^2$ , og seriekorrelasjon innebærer at restleddet er korrelert med restledd fra andre perioder,  $corr(u_{it}, u_{is}|\mathbf{x}) \neq 0, \forall t \neq s$  (Wooldridge, 2016, s. 395). Brudd på disse betingelsene vil føre til at standardavvikene feilberegnes og statistisk inferens basert på disse vil dermed kunne gi feilaktige resultater.

Det er lite realistisk at antakelsen om homoskedastisitet er oppfylt for denne analysen. Individuer med ulik innvandringsbakgrunn vil høyst sannsynlig ha forskjellige interesser og jobbmuligheter, noe som indikerer at de vil ha systematisk ulik variasjon i lønningene deres. Dette fikk vi blant et inntrykk av fra den deskriptive statistikken i kapittel 3, der det var større forskjeller i standardavvikene til gjennomsnittslønnen til de ulike innvandringskategoriene. En Breusch-Pagan test, vedlagt i appendiks A4, bekrefter at vi har heteroskedastisitet innenfor alle signifikansnivå. I tillegg er sannsynligheten stor for at vi har seriekorrelasjon i restleddet, og for å løse for disse to problemene gjennomføres derfor hele analysen ved bruk av klynge robuste standardavvik.

## 5. Resultater

Dette kapitlet presenterer de empiriske resultatene fra analysen. Først gjennomgås resultatene fra OLS-estimeringen av ulike modellspesifikasjoner, som gir et førsteinntrykk av sammenhengene. Deretter presenteres resultatene fra hovedanalysen med en individuell fixed effects modell og variasjonsgrunnlaget for analysen gjennomgås i kortere trekk.

### 5.1 Førsteintrykk med OLS

De estimerte lønnspremiene ved bruk av OLS på paneldatasettet er presentert i tabell 7, der referanse-kategorien for premiene er etnisk norske som bor i de små økonomiske regionene.

**Tabell 7:** Estimerte agglomerasjonseffekter med OLS

VARIABLER	(1) OLS lnlonn	(2) OLS lnlonn	(3) OLS lnlonn
stor_region	0.0960*** (0.0004)	0.1049*** (0.0003)	0.1047*** (0.0003)
mellom_region	0.0125*** (0.0004)	0.0116*** (0.0004)	0.0108*** (0.0004)
innvbak		-0.0584*** (0.0010)	
innvbak_stor_region		-0.0257*** (0.0011)	
innvbak_mellom_region		-0.0015 (0.0013)	
norinnvf			0.0192*** (0.0015)
hoyinnv			-0.0502*** (0.0017)
lavinnv			-0.1473*** (0.0017)
norinnvf_stor_region			-0.0109*** (0.0018)
hoyinnv_stor_region			0.0251*** (0.0021)
lavinnv_stor_region			-0.0541*** (0.0019)
norinnvf_mellom_region			0.0035* (0.0021)
hoyinnv_mellom_region			0.0026 (0.0025)
lavinnv_mellom_region			-0.0112*** (0.0023)
Konstant	12.22***	12.22***	12.21***
Kontrollvariabler	Ja	Ja	Ja
Tidskontroller	Ja	Ja	Ja
Observasjoner	4609433	4609433	4609433
R-squared	0.122	0.234	0.240

Standardavvik i parenteser  
\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Førsteintrykket av resultatene er at vi får bekreftet at vi har gjennomgående positive urbane lønnspremier i både de store- og mellomstore økonomiske regionene, over de små. Dette er i tråd med forventningene fra den monosentriske bymodellen og forholdet presisert tidligere i likning (3) med et positivt forhold mellom befolkningsstørrelsen og lønningene. De store regionene får estimert en gjennomgående statistisk signifikant premie på rundt 10%. Denne er betydelig høyere enn premien i de mellomstore regionene som er estimert til å være betraktelig lavere, men fortsatt signifikant på litt over 1%.

Videre får vi bekreftet forventningene våre om at det er større heterogeniteter i lønnspremiene blant individer med ulik innvandringsbakgrunn sammenliknet med de etnisk norske. Fra modell (2) får vi oppgitt at individer med innvandringsbakgrunn i utvalget samlet forventes å ha betydelig lavere lønnspremier i de store regionene sammenliknet med de etniske norske. Mens premiene deres i de mellomstore regionene ikke vil være signifikant forskjellig.

Den fullspesifiserte modellen (3) viser deretter til at det forventes å være betydelige forskjeller innad mellom de ulike innvandringsbakgrunnene i utvalget. I tråd med forventningene finner vi blant annet at innvandrere fra lavinntektsland har de desidert laveste lønnspremiene, og det insinuerer at de er gruppen som drar minst økonomisk nytte av agglomerasjonen. Innvandrere fra høyinntektsland forventes å ha de desidert høyeste premiene i både de store regionene sammenliknet med resten. Det impliserer at de er gruppen som bidrar mest til agglomerasjonseffektene, noe som stemmer overens med forventningene våre om at de er mer komplementære til den lokale arbeidskraften. Noe mer overraskende er resultatene om at de norskfødte med innvandrerforeldre i utvalget tilsynelatende forventes å ha lavere premier i de store regionene, men svakt høyere premier i de mellomstore regionene.

Alt i alt får vi fra disse resultatene bekreftet tilstedeværelsen av agglomerasjonseffekter, gjennom positive urbane lønnspremier som avtar med befolkningsstørrelsen i regionene. Vi har også fått bekreftet antydninger til betydelige heterogeniteter i agglomerasjonseffektene, gjennom signifikante forskjeller i de ekstra lønnspremiene til individer med ulik innvandringsbakgrunn. Foreløpig virker det som at innvandrere fra lavinntektsland har lavere gevinster fra agglomerasjonseffektene, og innvandrere fra høyinntektsland har de største gevinstene. De norskfødte med innvandrerforeldre derimot har tilsynelatende ingen entydig virkning på agglomerasjonseffektene i regionene.

Disse resultatene ved bruk av OLS som estimeringsmetode vil som sagt høyst sannsynlig hverken være forventningsrette eller konsistente. De økonometriske utfordringene gjennomgått i kapittel 4, deriblant problemer med utelatte relevante variabler og simultanitet fra både en potensiell individuell- og regional endogenitet, vil høyst sannsynlig prege estimatene og gjør dem biased. Den analytisk utledningen av simultanitetsproblemet (appendiks A5), og resultatene fra tidligere empiri gjør at vi forventer at disse urbane lønnspremiene vil være overestimerte.

## 5.2 Hovedmodellen FE

**Tabell 8:** Estimerte urbane lønnspremier med FE

VARIABLER	(4) FE lnlønn	(5) FE lnlønn	(6) FE lnlønn
stor_region	0.0114*** (0.0012)	0.0117*** (0.0012)	0.0116*** (0.0012)
mellom_region	-0.0007 (0.0012)	-0.0004 (0.0013)	-0.0006 (0.0013)
innvbak_stor_region		-0.0021 (0.0034)	
innvbak mellom_region		-0.0018 (0.0037)	
norinnvf_stor_region			0.0095 (0.0059)
hoyinnv_stor_region			0.0021 (0.0058)
lavinnv_stor_region			-0.0165*** (0.0060)
norinnvf_mellom_region			0.0095 (0.0058)
hoyinnv_mellom_region			0.0024 (0.0056)
lavinnv_mellom_region			-0.0153** (0.0073)
Konstant	13.17***	13.17***	13.17***
Kontrollvariabler	Ja	Ja	Ja
Tidskontroller	Ja	Ja	Ja
Observasjoner	4609433	4609433	4609433
R-squared	0.158	0.158	0.158

Standardavvik i parenteser  
 \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

I tråd med tidligere empiri får vi en relativt kraftig nedjustering av de estimerte urbane lønnspremiene ved overgangen til en individuell fixed effects modell. Dette insinuerer det som ble forklart i kapittel 2, at når uobserverbare individuelle heterogeniteter og den mulige endogene sorteringen av arbeidere ikke kontrolleres for, overestimeres OLS-premiene. Mer spesifikt antyder det at vi har permanente betydelige forskjeller mellom individene i utvalget. De har medfødt individkarakteristikk som IQ og andre ferdigheter som ikke er direkte observerbare, og bidrar til at individene får høyere lønninger. Når disse individene av ulike grunner tenderer til å sortere seg til de større økonomiske regionene resulterer det i at vi får en overestimeringen av lønnspremiene ved bruk av OLS.



Fra tabell 8 får vi med en individuell fixed effects modell fortsatt robuste positive lønnspremier i de store økonomiske regionene som alle er statistisk signifikante innenfor et 1% signifikansnivå. Her er lønnspremien i de store regionene estimert til å ligge på mellom 1.14%-1.17%, hvilket er mer nærliggende de noe nylig estimerte premiene i Australia fra studien til Meekes (2021). I kontrast til de øvrige OLS-estimatene finner vi ikke lenger statistisk signifikante premier i de mellomstore regionene. Det viset til at det ikke er noen økonomiske gevinster av å bo i de mellomstore økonomiske regionene sammenliknet med i de små når vi kontrollerer for individuelle heterogeniteter.

Videre får vi her estimert færre heterogeniteter i lønnspremiene mellom de ulike innvandringskategoriene. Fra modell (5) predikeres det å ikke lenger være noen statistisk signifikante ekstra lønnspremier blant individer med innvandringsbakgrunn, i hverken de store- eller mellomstore regionene. Her forventes altså de med innvandringsbakgrunn å ha de samme lønnspremiene som de etnisk norske. Det viser seg dog fra modell (6) og fortsatt være hensiktsmessig å skille mellom de ulike innvandringskategoriene. Hverken norskfødte med innvandrerforeldre eller innvandrere fra høyinntektsland predikeres å ha signifikante ekstra premier i de store- eller mellomstore regionene. Når vi har kontrollert for de individuelle heterogenitetene finner vi altså at de forventes å ha de samme økonomiske gevinstene fra agglomerasjonen som de etnisk norske. Innvandrerne fra lavinntektsland er her de eneste med signifikante ekstra lønnspremier, både i de store- og mellomstore regionene. Premiene deres forventes å være litt over 1.5% lavere i både de store- og mellomstore regionene, og disse estimatene er signifikante innenfor hhv. et 1%- og 5% signifikansnivå.

Disse negative ekstra lønnspremiene til innvandrere fra lavinntektsland i utvalget er såpass store at det innebærer at de predikeres å ha svakt negative lønnspremier. Det betyr at de forventes og i snitt ha lavere lønninger i de store- og mellomstore regionene sammenliknet med referansekategorien som er etniske nordmenn i de små regionene. Innvandrere fra lavinntektsland i de store regionene forventes og i snitt tjene 0.49% mindre, og i de mellomstore regionene forventes de å i snitt tjene 1.59% mindre enn de etnisk norske i de små regionene i utvalget. Dette impliserer at de fortsatt har en positiv økonomisk trend av å bo i de større økonomiske regionene sammenliknet med de mellomstore, men som er betydelig lavere enn resten.

Kontrollvariablene inkludert er som nevnt tidligere i stor grad konstante over tid, men spesielt utdanning har som forventet en betydelig påvirkning på individenes lønninger i utvalget. Fra de

fullstendige regresjonsresultatene til modell (6) vedlagt i appendiks A7, oppgis det i tabell A7.1 at et individ som har høyere utdanning i snitt forventes å tjene litt over 12% mer enn de med lavere utdanning i utvalget. Vi vil derfor se nærmere på heterogenitetene i lønnspremiene blant innvandringskategoriene med hensyn på utdanning som en sensitivitetsanalyse i neste kapittel.

Alt i alt finner vi fra hovedanalysen og bruken av en individuell fixed effects modell, bekreftelse på tilstedeværelsen av agglomerasjonseffekter kun i de store økonomiske regionene i Norge. Vi finner positive urbane lønnspremier på rundt litt over 1% av å bo i de store regionene over de små, men ingen signifikante premier i de mellomstore regionene. Den monosentriske bymodellen impliserer ut ifra disse resultatene at det kun er i de store økonomiske regionene i Norge at samlokaliseringen av individer og økonomisk aktivitet tilsynelatende er stor nok til at det observeres agglomerasjonseffekter, målt av positive urbane lønnspremier.

Videre finner vi færre tegn på at det er betydelige heterogeniteter i agglomerasjonseffektene blant individer med ulike innvandringsbakgrunn, som er i noe strid med forventningene våre. Det stemmer dog overens med funnene til studiene til Friedberg (2001) og Pischke og Velling (1994), hvor begge fant at innvandrere ikke hadde noen signifikant påvirkning på de lokale arbeidsmarkedsforholdene. Fra denne analysen har vi funnet at det kun er innvandrere fra lavinntektsland som predikeres å ha signifikant lavere lønnspremier sammenliknet med resten, både i de store- og mellomstore regionene. Resultatene våre bygger altså ikke opp under argumentene til Peri (2016) om at innvandrere vil kunne bidra desto mer til agglomerasjonseffektene i økonomien. Her finner vi kun bevis på at innvandrere fra lavinntektsland kan oppfattes som substitutter for den lokale arbeidskraften, og har lavere agglomerasjonseffekter enn den øvrige befolkningen. Disse funnene indikerer at gevinstene av å være bosatt i de mer urbane områdene i Norge for de med innvandringsbakgrunn, tilsynelatende i større grad kun vil være sosiale og ikke nødvendigvis økonomiske.

### **5.3 Variasjons-grunnlaget; Flytterne**

Bruken av fixed effects som metode for hovedanalysen til oppgaven innebærer som forklart tidligere at lønnspremiene estimeres kun basert på individene i utvalget som flytter i løpet av observasjons perioden. Fra microdata.no kan en per dags dato kun se på antall flyttinger i løpet av hele perioden, og det er ikke mulig å fange opp de unike individene dette gjelder. Det er totalt observert litt over 51 000 flyttinger mellom kategoriene for de økonomiske regionene. Dette hadde utgjort

5% av utvalget dersom det hadde tilsvart unike individer, og vi ikke har individer som flytter mellom de økonomiske regionene flere ganger i løpet av tidsperioden.

Til tross for at det kun er mulig å kommentere på antall flyttinger og ikke antall unike flyttere er det en rimelig antakelse at disse mest sannsynlig ikke vil avvike alt for mye fra hverandre. Kategoriene for de økonomiske regionene er relativt bredt definerte og vi ser kun på en 7-årig tidsperiode. For å kunne kommentere på variasjons-grunnlaget som hovedmodellen bygger på, tolker jeg derfor de oppgitte flyttingene fra microdata.no som individuelle flyttere og den deskriptive statistikken for disse er presentert nedenfor i tabell 9.

**Tabell 9:** Fordelingen av antall flyttinger/flyttere

FLYTTINGER	Antall	Andel av totalen	Andel av eget utvalg
Totalt	51 022	5 %	-
Norsk	41 138	81 %	5 %
Norskinnvf	2 434	5 %	7 %
Hoyinnv	3 434	7 %	8 %
Lavinnev	6 788	13 %	24 %

Fra tabell 9 ser en at blant flytterne er det en stor overvekt av nordmenn, men som andeler av sin egen gruppe er det en desto større andel av de med innvandringsbakgrunn som flytter. Totalt flytter 7% av alle norskfødte med innvandrerforeldre og 8% av alle innvandrere fra høyinntektsland i løpet av perioden. Mens opptil 24% av alle innvandrere fra lavinntektsland flytter. Dette innebærer at innvandrere fra lavinntektsland vil ha det desidert største variasjonsgrunnlaget for den øvrige analysen sammenliknet med de andre gruppene. Desto mer variasjon innenfor en gruppe, som her er antall flyttere, fører til at det er en desto høyere sannsynlighet for at de estimerte premiene vil være forventningsrette.

Vi har alt i alt en relativt liten andel av hver av gruppene som flytter i løpet av perioden og det gjør at estimatene vil være mer utsatt for skjevheter. Dersom de få som flytter har karakteristikk som avviker stort fra resten av utvalget, vil ikke de estimerte premiene i tabell 8 være generaliserbare, men kan gi et feilinntrykk av de faktisk kausale effektene. Det vil reflekteres mer rundt under diskusjonen av analysen i kapittel 7.

## 6. Robusthet

I dette kapittelet gjennomgås det resultater fra noen robusthetssjekker av den øvrige hovedanalysen. De vil vise hvor sensitive resultatene er for andre spesifikasjoner og diverse avgrensinger av utvalget, samt noen kommentarer på andre heterogeniteter i premiene.

### 6.1 Ekstremverdier

Bruken av hele utvalget kan føre til at de urbane lønnspremiene feilestimeres på grunn av målefeil i den avhengige variabelen som nevnt tidligere i kapittel 4. Blant annet kan tilstedeværelsen av ekstremverdier føre til at vi får større skjevheter i premiene. For å undersøke dette er det gjennomført en sensitivitetsanalyse på resultatene fra tabell 8 der alle observasjoner innenfor 1- og 99-persentilen av registrert årslønn er ekskludert fra paneldatasettet. For å unngå at dette skulle føre til verre seleksjonsskjevheter i utvalget har de respektive ekstremverdiene fra hvert delutvalg med hensyn på innvandringskategoriene blitt kuttet. Denne avgrensingen fjerner alt i alt nesten 92 000 observasjoner fra datasettet og de nye regresjonsresultatene for hovedanalysen er presentert nedenfor i tabell 10.

**Tabell 10:** Estimerte lønnspremier ekskludert ekstremverdiene

VARIABLER	(4') FE lnlønn	(5') FE lnlønn	(6') FE lnlønn
stor_region	0.0097*** (0.0010)	0.0101*** (0.0011)	0.0100*** (0.0010)
mellom_region	0.0000 (0.0010)	0.0000 (0.0011)	-0.0002 (0.0011)
innbak_stor_region		-0.0022 (0.0029)	
innbak_mellom_region		-0.0006 (0.0030)	
norinnvf_stor_region			0.0076 (0.0051)
hoyinnv_stor_region			0.0011 (0.0048)
lavinnv_stor_region			-0.0133* (0.0049)
norinnvf_mellom_region			0.0073 (0.0049)
hoyinnv_mellom_region			0.0012 (0.0048)
lavinnv_mellom_region			-0.0066 (0.0053)
Konstant	13.18***	13.18***	13.18***
Kontrollvariabler	Ja	Ja	Ja
Tidskontroller	Ja	Ja	Ja
Observasjoner	4518815	4518815	4518815
R-squared	0.242	0.242	0.242

Standardavvik i parenteser  
\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Ekskludering av ekstremverdiene fører til at vi fortsatt har gjennomgående robuste og signifikante positive urbane lønnspremier i de store økonomiske regionene, men som er blitt noe nedjustert. På sitt høyeste er premien her estimert til 1.01% sammenliknet med 1.17% ved bruk av hele utvalget i hovedmodellen tidligere. Premien i de mellomstore regionene er fortsatt ikke signifikant forskjellig fra null.

De estimerte heterogenitetene fra den øvrige hovedanalysen i lønnspremiene mellom innvandringskategoriene vedvarer også noe. Norskfødte med innvandrerforeldre og innvandrere fra høyinntektsland forventes fortsatt ikke å ha signifikant forskjellige premier fra de etnisk norske i noen av de økonomiske regionene. Mens innvandrere fra lavinntektsland fortsatt får estimert en negativ ekstra premie i de store regionene, men denne er også noe nedjustert, samt svakere statistisk signifikant. Tidligere hadde de en premie som var omtrent 2% lavere enn resten og signifikant innenfor et 5% signifikansnivå, mens her er premien estimert til -1.2% og kun innenfor et 10% signifikansnivå.

Det eneste estimatet som endres noe drastisk under denne sensitivitetsanalysen er den ekstra premien til innvandrere fra lavinntektsland i de mellomstore regionene. Den ble tidligere i tabell 8 estimert å være nærmere -2%, mens her er den ikke lenger signifikant forskjellig fra null. Dette indikerer at den tidligere estimerte premien i stor grad kun var drevet av såkalt støy. Fjerningen av ekstremverdiene fører til at det ikke lenger er noen av innvandringskategoriene som har en signifikant forskjellig premie i de mellomstore regionene. Dette betyr at alle individene i utvalget her forventes å ha de samme økonomiske gevinstene av å bo i de mellomstore- som i de små økonomiske regionene i Norge.

Alt i alt finner vi at ekstremverdiene ikke påvirker de øvrige resultatene fra hovedanalysen betydelig. Ekskluderingen av disse fører kun til en noe nedjustering i lønnspremiene i de store regionene, og den ekstra premien til innvandrere fra lavinntektsland i disse. Men det fører dog til at den ekstra premien deres i de mellomstore regionene forsvinner helt.

## 6.2 Andre heterogeniteter

En annen robusthetssjekk går ut på å undersøke betydningen av andre heterogeniteter, deriblant av utdanning, alder og kjønn. Dette er gjort ved å estimere den fullspesifiserte modellen (6) på flere delutvalg. Vi ser på delutvalg for de med lav- og høy utdanning, de under 45år og de som er 45år eller eldre, i tillegg til et separat for kvinner og menn.

**Tabell 11:** Heterogenitetssjekk mhp. utdanning og alder

VARIABLER	(6) FE lnlonn	(7) FE lnlonn	(8) FE lnlonn	(9) FE lnlonn	(10) FE lnlonn
stor_region	0.0116*** (0.0012)	0.0164*** (0.0016)	0.0051*** (0.0018)	0.0096*** (0.0022)	0.0092*** (0.0022)
mellom_region	-0.0006 (0.0013)	-0.0001 (0.0018)	-0.0005 (0.0018)	-0.0036** (0.0016)	0.0015 (0.0021)
norinnvf_stor_region	0.0095 (0.0059)	0.0093 (0.0079)	0.0113 (0.0080)	0.0108 (0.0068)	0.0135 (0.0108)
hoyinnv_stor_region	0.0021 (0.0058)	-0.0018 (0.0075)	0.0067 (0.0088)	0.0068 (0.0072)	0.0033 (0.0096)
lavinnv_stor_region	-0.0165*** (0.0060)	-0.0287*** (0.0081)	-0.0029 (0.0086)	-0.0160** (0.0066)	-0.0173 (0.0137)
norinnvf_mellom_region	0.0095 (0.0058)	0.0078 (0.0082)	0.0137* (0.0075)	0.0114* (0.0070)	0.0089 (0.0108)
hoyinnv_mellom_region	0.0024 (0.0056)	0.0033 (0.0074)	-0.0023 (0.0075)	0.0036 (0.0069)	0.0010 (0.0095)
lavinnv_mellom_region	-0.0153** (0.0073)	-0.0236** (0.0106)	-0.0063 (0.0092)	-0.0142* (0.0073)	-0.0281 (0.0208)
Konstant	13.17***	13.31***	13.16***	13.13***	13.26***
Kontrollvariabler	Ja	Ja	Ja	Ja	Ja
Tidskontroller	Ja	Ja	Ja	Ja	Ja
Utvalg	Hele	Høy utdanning	Lav utdanning	Under 45år	Over eller lik 45år
Observasjoner	4609433	2300075	2309358	2104921	2504512
R-squared	0.158	0.202	0.115	0.232	0.093

Standardavvik i parenteser  
 \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Resultatene fra tabell 11 viser til at vi fortsatt har gjennomgående positive urbane lønnspremier i de store økonomiske regionene som er robuste og statistisk signifikante over alle delutvalgene. Premien varierer her mellom 0.51% på sitt laveste for de med lav utdanning, og opp til 1.64% på sitt høyeste for de med høy utdanning. Dette er i tråd med forventningene våre og funnene til Carlsen et al. (2016) om tiltakende agglomerasjonseffekter med utdanningsnivået. Premiene i de mellomstore regionene er fortsatt ikke statistisk signifikant forskjellig fra null over de fleste

delutvalgene, med unntak av for de under 45år, der den er svakt negativ. I modell (9) forventes de yngre i utvalget og i snitt tjene 0.36% mindre i de mellomstore regionene sammenliknet med i de små, og estimatet er signifikant innenfor et 5% signifikansnivå.

Resultatene om ikke-signifikante forskjeller fra de etnisk norske i lønnspremiene til de norskfødte med innvandrerforeldre og innvandrere fra høyinntektsland, ser fortsatt ut til å gjelde over de fleste delutvalgene. Unntaket er norskfødte med innvandrerforeldre som har lav utdanning og eller er yngre enn 45år som rapporteres hhv. i modell (8) og (9) å ha svakt signifikant positive premier i de mellomstore regionene.

Det er videre flere avvik og større variasjon i signifikansen til de tidligere resultatene om betydelig lavere premier blant innvandrere fra lavinntektsland i hele utvalget. De får kun bekreftet negative ekstra premier i de store- og mellomstore regionene for de som har høy utdanning og for de under 45år. Eldre innvandrere fra lavinntektsland og de med lav utdanning estimeres og ikke ha signifikante forskjeller i premiene sine fra resten av utvalget i de store- og mellomstore regionene.

**Tabell 12:** Heterogenitetssjekk mhp. kjønn

VARIABLER	(6) FE lnlonn	(11) FE lnlonn	(12) FE lnlonn
stor_region	0.0116*** (0.0012)	0.0105*** (0.0016)	0.0133*** (0.0019)
mellom_region	-0.0006 (0.0013)	-0.0011 (0.0017)	0.0006 (0.0020)
norinnvf_stor_region	0.0095 (0.0059)	0.0088 (0.0075)	0.0117 (0.0094)
hoyinnv_stor_region	0.0021 (0.0058)	0.0078 (0.0069)	-0.0094 (0.0103)
lavinnv_stor_region	-0.0165*** (0.0060)	-0.0214*** (0.0075)	-0.0066 (0.0098)
norinnvf_mellom_region	0.0095 (0.0058)	0.0110 (0.0073)	0.0079 (0.0095)
hoyinnv_mellom_region	0.0024 (0.0056)	0.0047 (0.0069)	-0.0020 (0.0094)
lavinnv_mellom_region	-0.0153** (0.0073)	-0.0170* (0.0096)	-0.0112 (0.0103)
Konstant	13.17***	13.25***	13.06***
Kontrollvariabler	Ja	Ja	Ja
Tidskontroller	Ja	Ja	Ja
Utvalg	Hele	Menn	Kvinner
Observasjoner	4609433	2751502	1857931
R-squared	0.158	0.131	0.211

Standardavvik i parenteser  
 \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Heterogenitetssjekken på kjønn presentert i tabell 12 bekrefter også vedvarende positive urbane lønnspremier i de store regionene, men som er betydelig høyere for kvinnene i utvalget. Fra modell (12) finner vi at kvinner i de store regionene forventes å ha en premie på 1.33%, mens mennene i utvalget kun har en estimert premie på 1.05%. Dette insinuerer at kvinner har desto større økonomiske gevinster fra agglomerasjonseffektene enn menn i de store økonomiske regionene i Norge.

Videre får vi bekreftelse på resultatene om ingen signifikante premier i de mellomstore regionene, og heller ingen signifikante forskjeller i premiene blant norskfødte med innvandrerforeldre og innvandrere fra høyinntektsland. De signifikant lavere premiene til innvandrere fra lavinntektsland oppgis i tabellen og kun gjelde for mennene i utvalget, mens kvinnene ikke har signifikant forskjellige premier fra resten. Disse resultatene antyder at spesielt innvandrerkvinner fra lavinntektsland forventes å ha desto større gevinster fra agglomerasjonseffektene sammenliknet med mennene.

Alt i alt finner vi fra denne sensitivitetsanalysen med avgrensning til ulike delutvalg, at funnet om signifikante positive urbane lønnspremier i de store økonomiske regionene i Norge er svært robuste. Videre avdekker undersøkelsen av andre heterogeniteter i premiene med hensyn på utdanning, kjønn og alder, at de tidligere resultatene for forskjellene i premiene blant innvandrere fra lavinntektsland, i stor grad drives av menn med høy utdanning og de yngre under 45 år i utvalget.

## **6.2 Omdefinering av regionene**

Den siste sensitivitetsanalysen er en omdefinering av kategoriene for de økonomiske regionene som er noen av forklaringsvariablene av hovedinteresse. Her trekkes Stor-Oslo regionen ut som en egen kategori, fordi det har med årene blitt desto mer pendling innenfor denne regionen. Det vil derfor være interessant å undersøke hvorvidt denne står for den største andelen av de øvrige estimerte premiene, spesielt i de store økonomiske regionene. Dummyvariabelen som indikerer hvorvidt et individ bor i Stor-Oslo eller ikke er kodet om ifølge definisjonen fra regjeringens stortingsmelding «Storbymeldingen 2002-2003». Stor-Oslo regionen omfatter totalt 46 kommuner, hvorav 17 tidligere har inngått i de store regionene, 4 i de mellomstore og 25 har vært kategorisert under de små. En fullstendig oversikt over hvilke dette gjelder er vedlagt i appendiks A8.



**Tabell 13:** Regresjonsresultater robusthetssjekk med Stor-Oslo

VARIABLER	(13)	(14)
	FE lnlønn	FE lnlønn
stor_oslo	0.0221*** (0.0014)	0.0231*** (0.0015)
stor_region_o	0.0025** (0.0013)	0.0023 (0.0015)
mellom_region_o	-0.0011 (0.0012)	-0.0012 (0.0013)
norinnvf_stor_oslo		-0.0063 (0.0063)
hoyinnv_stor_oslo		-0.0023 (0.0065)
lavinnv_stor_oslo		-0.0127 (0.0081)
norinnvf_stor_region_o		0.0102 (0.0065)
hoyinnv_stor_region_o		0.0036 (0.0059)
lavinnv_stor_region_o		-0.0088 (0.0055)
norinnvf_mellom_region		0.0044 (0.0055)
hoyinnv_mellom_region		0.0034 (0.0055)
lavinnv_mellom_region		-0.0087 (0.0061)
Konstant	13.17***	13.17***
Kontrollvariabler	Ja	Ja
Tidskontroller	Ja	Ja
Observasjoner	4609433	4609433
R-squared	0.159	0.159

Standardavvik i parenteser  
 \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Resultatene fra tabell 13 er modellene (4) og (6) fra hovedanalysen estimert på nytt med omdefineringen av regionene, der Stor-Oslo er definert som en egen region. Det er kun i den enkle regresjonsmodellen (13) uten kontroller for innvandringsbakgrunn, at vi fortsatt får en signifikant, men betraktelig svakere positiv premie i de store økonomiske regionene uten Stor-Oslo. Her estimeres premien i de store regionene til kun 0.25% sammenliknet med estimatene fra hovedanalysen på over 1% tidligere. Det er derimot ingen av de tidligere signifikante resultatene fra hovedanalysen som vedvarer i den fullspesifiserte modellen (14). Hverken den positive lønnspremien i de store

økonomiske regionene, eller de negative ekstra lønnspremiene til innvandrere fra lavinntektsland i de store- og mellomstore regionene er statistisk signifikante lenger.

Det eneste signifikante estimatet i den fullspesifiserte modellen (14) er en sterkt positiv lønnspremie av å bo i Stor-Oslo regionen. Der individer som er bosatt i Stor-Oslo forventes og i snitt tjene 2.31% mer enn de andre i utvalget, og dette estimatet er signifikant innenfor et 1% signifikansnivå. Videre er det tilsynelatende ingen signifikante heterogeniteter i lønnspremien i Stor-Oslo mellom de ulike innvandringskategoriene.

Denne sensitivitetsanalysen avdekker at det er Stor-Oslo regionen som står bak alle de tidligere estimatene av positive urbane lønnspremiene i de store økonomiske regionene i Norge. Når Stor-Oslo trekkes ut som en egen region, finnes det ingen signifikante lønnspremier for utvalget i de store regionene lenger. Disse resultatene indikerer at vi kun har agglomerasjonseffekter i Stor-Oslo regionen, og ikke i de andre økonomiske regionene i Norge. Vi finner også at de tidligere funnene om heterogeniteter i lønnspremien blant innvandrere fra lavinntektsland, heller ikke er robuste i denne nye modellspesifikasjonen.

## 7. Oppsummering og diskusjon

Dette kapittelet oppsummerer og diskuterer resultatene. Utfordringene og svakhetene ved analysen vil drøftes, og det vil til slutt gis noen oppfordringer til videre forskning.

### 7.1 Drøfting av resultatene

#### Funnene fra hovedanalysen

Oppgaven fant bevis på tilstedeværelsen av agglomerasjonseffekter i de norske økonomiske regionene, men med varierende resultater for heterogeniteten blant individer med ulik innvandringsbakgrunn. Vi fant signifikante positive lønnspremier i de store økonomiske regionene i Norge, i likhet med det som ble insinuert av den deskriptive statistikken illustrert grafisk i figur 1. Premiene varierte mellom 10-11% når de ble estimert med minste kvadraters metode, men ble kraftig nedjustert til rundt 1% ved bruk av en individuell fixed effects modell. Den sistnevnte premien virket spesielt troverdig, ettersom den er svært nærliggende premiene som nylig ble estimert i Australia av studien til Meekes (2021).

Bruken av fixed effects som metode gjør at vi har løst for den individuelle endogeniteten der flinkere arbeidere sorterer seg til de mer tettbefolkede regionene. Nedgangen i de estimerte lønnspremiene fra OLS til FE indikerer at arbeiderne har uobserverbare egenskaper som gjør dem mer produktive og hvilket er positivt korrelert med bosetting i de større norske regionene. Dersom vi ikke hadde tatt hensyn til dette ville vi fått en overestimering av premiene ved kun bruk av OLS som metode. Disse resultatene er troverdige og er i likhet med funnene fra tidligere empiri. De stemmer spesielt godt overens med noen av funnene til Carlsen et al. (2016) som viste til at det var betydelige forskjeller i egenskapene til arbeiderne på tvers av regionene i Norge.

Resultatene av hovedinteresse for oppgaven var de ekstra lønnspremiene til de ulike innvandringskategoriene, som sier hvorvidt de har signifikante forskjeller av agglomerasjonseffektene i de norske regionene. Fra hovedanalysen fant vi kun bevis på at innvandrere fra lavinntektsland hadde signifikante ekstra premier, som indikerte at de hadde lavere gevinster fra agglomerasjonen sammenliknet med resten av utvalget. Disse resultatene kan ut ifra noe av teorien presentert tidligere forklares av at de er innvandringsgruppen med det laveste nivået på humankapitalen. Dette gjør at de vil dra desto lavere økonomiske gevinster fra agglomerasjonen i de norske regionene. De vil

tilsynelatende ikke dra like stor nytte av kunnskapsutvekslingen eller den bedre matchingen på arbeidsmarkedet i de mer befolkningstette regionene, sammenliknet med resten av befolkningen. De norskfødte med innvandrerforeldre og innvandrere fra høyinntektsland forventes derimot å ha de samme produktivetsgevinstene av å bo i de større økonomiske regionene som de etnisk norske. Disse innvandringsbakgrunnene gir altså ingen videre produktivetsgevinster til agglomerasjons-effektene som initielt forventet ut ifra utsagnene til Peri (2016).

### **Sensitivitetsanalysene**

Den siste sensitivitetsanalysen gjennomgått i kapittel 6 viste dog til at det kun er agglomerasjons-effekter i Stor-Oslo regionen, målt av en høyt statistisk og økonomisk signifikant premie på 2.3%. Når Stor-Oslo blir trukket ut som en egen kategori finnes det ikke lenger signifikante positive premier i de resterende store økonomiske regionene i Norge. Dette virker noe rimelig ettersom Stor-Oslo er anerkjent for å være et metropol for økonomisk aktivitet, men det kan argumenteres for og likevel være noe overraskende. Noen av de største norske storbyene Bergen, Trondheim og spesielt Stavanger inngår fortsatt i kategorien for de store regionene, og gjør at en kunne forventet å finne positive lønnspremier selv uten Stor-Oslo.

Et mer overraskende resultat er at når modellspesifikasjonen endres til å se på Stor-Oslo regionen som en egen kategori forsvinner absolutt alle heterogeniteter i lønnspremiene blant de ulike innvandringskategoriene. Det finnes ikke lenger noen statistisk signifikant negative ekstra premie blant innvandrene fra lavinntektsland. Disse resultatene var ikke robuste over en omdefinering av modellen, og virker tilsynelatende å være drevet av de som var bosatt i de store- og mellomstore regionene som nå er definert innenfor Stor-Oslo.

En sensitivitetsanalyse i likhet med den gjennomgått i underkapittel 6.2 er gjennomført på modell (14), og resultatene er vedlagt i appendiks A9. Her finner vi at det fortsatt er de yngre med høyere utdanning som er de som i større grad driver agglomerasjonseffektene, også i Stor-Oslo regionen. De oppgis å ha over dobbelt så store premier sammenliknet med de eldre med lav utdanning. I tillegg vi får fra denne analysen bekreftet at det heller ikke er noen tegn til heterogeniteter i lønnspremiene blant de ulike innvandringskategoriene, selv blant de mindre heterogene delutvalgene.

## 7.2 utfordringer

### Variasjonsgrunnlaget

En av de største mulige svakhetene ved resultatene fra denne analysen, og valget av en individuell fixed effects modell som hovedstrategi, er variasjonsgrunnlaget det bygger på. Strategien vil som forklart tidligere i kapittel 4 kreve at vi har nok variasjon innenfor individene, for at estimatene skal være forventningsrette og konsistente. Her betyr det at vi må ha nok individer som flytter mellom de ulike regionkategoriene i løpet av periodene, og som opplever endringer i lønningene sine. Oppgaven har etter beste evne forsøkt å undersøke dette grunnlaget, men de begrensede funksjonalitetene i microdata.no har gjort det vanskelig. Vi har kun fått tilgang på tall for antall flyttinger i løpet av observasjonsperioden, og informasjon om fordelingen mellom de ulike innvandringskategoriene. Oversikten over antall flyttinger ble oppgitt i tabell 9, der vi fant at innvandrere fra lavinntektsland var innvandringskategorien med den største andelen flyttere innad sin egen gruppe. Dette sikrer som sagt at de har det beste variasjonsgrunnlaget for hovedanalysen, og det kan være en mulig forklaring på hvorfor de var den eneste kategorien med antydning til signifikante ekstra lønnspremier.

Det har som sagt ikke vært mulig for denne oppgaven å fange opp de unike flytterne og sjekke deres individkarakteristikken, eller finne informasjon om dem som kan brukes til å si noe om motivasjonen bak flyttingen deres. Karakteristikken til flytterne og motivasjonen bak flyttingen deres vil være avgjørende for validiteten og generaliserbarheten til resultatene fra analysen. I tråd med funnene til Grogger og Hanson (2011) vet vi at det typisk er de mer ressurssterke i samfunnet som har en høyere sannsynlighet for å flytte på seg. Dersom flytterne i utvalget i større grad har karakteristikk som gjør dem mer produktive sammenliknet med resten av utvalget, vil vi ha et positivt bias i premiene fra analysen vår. Bruken av en individuell fixed effects modell vil i stor grad kontrollere for dette så lenge det skyldes faktorer som er konstante over tid.

### Utelatte relevante variabler

Videre er det en større svakhet ved analysen at vi kan ha flere potensielle utelatte relevante variabler som gjør at resultatene fra analysen er skjeve og ikke er forventningsrette. Spesielt variabler for arbeidernes yrker og sektorer burde vært kontrollert for fordi de høyst sannsynlig er korrelert med både lønnen og bosettingsmønstrene deres. Det kan som nevnt tidligere forventes å være

systematiske forskjeller i yrkesvalg mellom de ulike innvandringskategoriene som gjør at når det ikke kontrolleres for vil de ekstra premiene deres være feilestimerte. Funnene fra hovedanalysen om negative ekstra premier hos innvandrere fra lavinntektsland kan for eksempel tenke seg og delvis kunne forklares av at de i større grad jobber innenfor mer lavtlønnede yrker typisk innenfor byggebransjen, service- og tjenestesektoren o.l. Når analysen ikke får kontrollert for dette vil det kunne føre til at de negative premiene overestimeres i dette tilfellet. Dette har ikke blitt forsøkt løst her i denne analysen fordi det innebærer mer omfattende arbeid i microdata.no som er vurdert som utenfor omfanget til denne oppgaven. Det kan derimot argumenteres for at å kontrollere for yrke- eller sektorfasteeffekter ville etterlatt oss med et desto mindre variasjonsgrunnlag. Sjansen for seleksjonsskjevheter ville økt og kunne dermed resultere i desto større skjevheter i estimatene. Denne utfordringen nevner som sagt tidligere både Meeke (2021) og Combes og Gobillon (2015) i sine studier, hvor begge sier at det kan resultere i overestimering av premiene.

### **Regional endogenitet**

Den siste store nevneverdige utfordringen og svakheten ved denne oppgaven er at den ikke har fått forsøkt å løse for den mulige regionale endogeniteten. Ifølge empirien er sannsynligheten høy for at vi har naturlige fortrinn eller annen karakteristik som vil gi norske regioner produktivetsfordeler, og gjør dem desto mer attraktive for bosetting. Det har typisk blitt argumentert for at når dette ikke tas hensyn til vil premiene typisk overestimeres som illustrert analytisk i appendiks A5. Dette baseres til dels på at det er antatt en positiv korrelasjon mellom lønninger og bosted, men det er verdt å sette spørsmålsteget ved denne antakelsen. Motivasjonen for flytting og valg av bosted kan argumenteres for og i større grad avhenge av familierelasjoner og emosjonelle tilknytninger til områdene, heller enn de økonomiske forholdene. Dette underbygger at den antatte simultaniteten ikke nødvendigvis vil være like problematisk som initielt antatt og fører ikke til at estimatene er overestimerte. Dette underbygges også av Meeke (2021) og Combes og Gobillon (2015) som viser til at det er desto viktigere å kontrollere for den individuelle endogeniteten enn den regionale og det har denne oppgaven gjort.

### **7.3 Videre forskning**

Denne oppgaven har valgt å ta i bruk mikrodata på individnivå fremfor regiondata for å forsøke å estimere de mest presise lønnspremiene, og tallfeste de individuelle agglomerasjonseffektene.

Forskningsplattformen microdata.no som er tatt i bruk for å gjennomføre analysen, er som sagt nyere og under stadig videreutvikling, hvilket har medført flere utfordringer og resultert i noen svakheter ved analysen. Spesielt mangelen på løsning for den regionale endogeniteten. Etter hvert som flere funksjonaliteter og variabler blir tilgjengelige i programmet vil en kunne undersøke dette nærmere, og muligens estimere mer forventningsrette og konsistente premier. Det vil være spesielt interessant og ta i bruk strategien til Leksnes (2014), og bruke historiske gruver som instrument for dagens bosettingsmønstre. Ved gjennomførelsen av dette vil en kunne si noe mer presist om hvorvidt det og ikke løse for den regionale endogeniteten faktisk fører til en overestimering av premiene, ettersom det fremstår noe uenighet i litteraturen.

Tidshorisonten til oppgaven er vesentlig kort, noe som er en spesielt stor svakhet for analysen på dette temaet og med strategien valgt, fordi det er med på å begrense variasjonsgrunnlaget. Flytting er generelt noe som ikke gjøres ofte, og spesielt ikke på tvers av de relativt bredt definerte regionkategoriene i denne analysen. Det vil være svært hensiktsmessig for videre analyser å se på en lengre tidshorisont for å sikre mer variasjon. I tillegg vil det kunne muliggjøre for å se nærmere på diskusjonen frembragt av Glaeser og Maré (2001) om hvorvidt det er et skille mellom en lønnsnivå- og lønnsvekst-effekt fra agglomerasjonen. Bosettelse og arbeid i de mer tettbefolkede områdene kan gi agglomerasjonseffekter gjennom en mer statisk økning i lønnsnivået, eller de kan være mer dynamiske og gi økt lønnsvekst over tid. Alternativt kan en annen måte å sikre et større variasjonsgrunnlag for en tilsvarende analyse på, være å bruke kontinuerlige mål på befolkningsstørrelsen/tettheten i de norske regionene.

Sist, men ikke minst gjelder resultatene fra denne analysen kun for individer som er etablerte på arbeidsmarkedet i en fulltidsstilling under hele perioden. Oppgaven har tatt utgangspunkt i et svært avgrenset utvalg, og det kan være interessant for videre forskning og se nærmere på hvordan agglomerasjonseffektene påvirker de som er ansatt i deltidstillinger for eksempel. Dette vil være spesielt relevant for å kunne si noe mer om heterogenitetene til agglomerasjonseffektene blant individer med ulik innvandringsbakgrunn. Det er rimelig å anta at de har en større tendens til å være ansatt i deltidstillinger over lengre tid enn de etnisk norske.

## 8. Konklusjon

Denne oppgaven har analysert tilstedeværelsen av agglomerasjonseffekter i de økonomiske regionene i Norge, og heterogeniteten til disse blant individer med ulik innvandringsbakgrunn. Dette er gjennomført ved bruk av et mikropaneldatasett fra den noe nyere forskningsplattformen microdata.no. Analysen er avgrenset til å følge litt over 930 000 fulltids-arbeidere fra 2013-2017, og er forankret i tidligere empiri på temaet samt teori fra den monosentriske bymodellen.

To av hovedutfordringene til identifiseringen av agglomerasjonseffektene er at en kan ha en endogen bestemt kvalitet og kvantum av arbeidskraft i regionene, empirisk oppsumert som hhv. individuell- og regional endogenitet. Oppgaven har løst for den individuelle endogeniteten ved bruk av en individuell fixed effects modell som hovedstrategi. Den regionale endogeniteten har derimot ikke blitt løst for på grunn av begrensninger med plattformen som er tatt i bruk, og det kan resultere i at de estimerte premiene i denne analysen er overestimerte.

Vi finner kun bevis på tilstedeværelsen av agglomerasjonseffekter i Stor-Oslo regionen, med en estimert lønnspremie på 2.3% over de andre regionene. Det viser til at det kun er i Stor-Oslo at samlokaliseringen av bedrifter og individer er stor nok til at vi observerer produktivitetsgevinster, gjennom en signifikant positiv lønnspremie. Disse agglomerasjonseffektene viser seg og i stor grad drives av de yngre med høy utdanning i befolkningen.

Hovedresultatet fra analysen er deretter at vi kun finner få antydninger til negative ekstra lønnspremier blant innvandrere fra lavinntektsland, men disse resultatene er ikke robuste over en omformulering av hovedmodellen. Vi ender opp med å konkludere med at det tilsynelatende ikke er noen signifikante heterogeniteter i de urbane lønnspremiene blant individer med ulik innvandringsbakgrunn. Dette indikerer at de står overfor de samme agglomerasjonseffektene som de etnisk norske.





- 
- Duranton, G., & Puga, D. (2013). The growth of cities. Hentet fra [https://real-faculty.wharton.upenn.edu/wp-content/uploads/~duranton/Duranton\\_Papers/Current\\_Research/urban\\_growth.pdf](https://real-faculty.wharton.upenn.edu/wp-content/uploads/~duranton/Duranton_Papers/Current_Research/urban_growth.pdf)
- Ellison, G., & Glaeser, E. L. (1999). The Geographic Concentration of Industry: Does Natural Advantage Explain Agglomeration? Hentet fra <https://pubs.aeaweb.org/doi/pdfplus/10.1257/aer.89.2.311>
- Friedberg, R. M. (2001). The Impact of Mass Migration on the Israeli Labor Market. *The Quarterly Journal of Economics*. Hentet fra [https://www.brown.edu/Departments/Economics/Faculty/Rachel\\_Friedberg/Links/Friedberg%20QJE.pdf](https://www.brown.edu/Departments/Economics/Faculty/Rachel_Friedberg/Links/Friedberg%20QJE.pdf)
- Friedberg, R. M., & Hunt, J. (1995). The Impact of Immigrants on host country wages, employment and growth. *Journal of Economic Perspectives*. Hentet fra <https://pubs.aeaweb.org/doi/pdfplus/10.1257/jep.9.2.23>
- Glaeser, E. L. (2010). Agglomeration Economics. ss. 1-3. Hentet fra <https://www.nber.org/system/files/chapters/c7977/c7977.pdf>
- Glaeser, E. L., & Maré, D. C. (2001). Cities and Skills. *Journal of Labour Economics*, ss. 1-10. Hentet fra <https://www.journals.uchicago.edu/doi/epdf/10.1086/319563>
- Grogger, J., & Hanson, G. H. (2011). Income maximization and the selection and sorting of international migrants. Hentet fra [https://gps.ucsd.edu/\\_files/faculty/hanson/hanson\\_publication\\_migration\\_income.pdf](https://gps.ucsd.edu/_files/faculty/hanson/hanson_publication_migration_income.pdf)
- Høydahl, E. (2013, April 30). Innvandrere i bygd og by. *SSB*. Hentet fra <https://www.ssb.no/befolkning/artikler-og-publikasjoner/innvandrere-i-bygd-og-by#:~:text=Generelt%20bor%20innvandrere%20mer%20sentralt,sentralitet%20og%20tet%2D%2Fspredtbygdhet.>
- Johansen, S. (2020). microdata.no – søknadsfri tilgang til registerdata om boforhold. *Tidsskrift for boligforskning*. Hentet fra <https://www.idunn.no/doi/10.18261/issn.2535-5988-2020-01-06>
- Jordheim, H. M. (2018, 05 26). Dette er byene med de høyeste lønningene: Oslo på topp 10. *E24*. Hentet fra <https://e24.no/naeringsliv/i/rL6lBa/dette-er-byene-med-de-hoeyeste-loenningene-oslo-paa-topp-10>
- Kommunal- og distriktsdepartementet. (2013). Lønn for ansatte i staten. Hentet fra <https://www.regjeringen.no/no/tema/arbeidsliv/Statlig-arbeidsgiverpolitikk/lonn-og-tariff-i-staten/lonnstabellen/id438643/>

- Leknes, S. (2020). *Voksende byer og aldrende bygder*. SSB. Hentet fra De fem kommunene med høyest framskrevet vekst tilhører alle Viken fylke, og har en vekst på over 30 prosent mot 2050.
- Leksnes, S. (2015). The more the merrier? Evidence on quality of life and population size using historical mines. *Regional science and urban economics*. Hentet fra <https://www.sciencedirect.com/science/article/pii/S0166046215000526>
- Longva, S., & Tønder, J.-K. (2000). *Standard for økonomiske regioner*. SSB. Hentet fra [https://www.ssb.no/a/publikasjoner/pdf/nos\\_c616/nos\\_c616.pdf](https://www.ssb.no/a/publikasjoner/pdf/nos_c616/nos_c616.pdf)
- Meekes, J. (2021). Agglomeration economies and the urban wage premium in Australia. Hentet fra [https://melbourneinstitute.unimelb.edu.au/\\_\\_data/assets/pdf\\_file/0004/3916984/wp2021n21.pdf](https://melbourneinstitute.unimelb.edu.au/__data/assets/pdf_file/0004/3916984/wp2021n21.pdf)
- Microdata.no. (u.d.). *Microdata.no*. Hentet fra Databankoversikt: <https://microdata.no/discovery/datastore/?datastore=no.ssb.fdb&version=16>
- Olsen, B. (2017). *Innvandrere i og utenfor arbeidsmarkedet*. SSB. Hentet fra <https://www.ssb.no/arbeid-og-lonn/artikler-og-publikasjoner/innvandrere-i-og-utenfor-arbeidsmarkedet>
- Ottaviano. (2006). Rethinking the effects of immigration on wages. *National bureau of economic research*. Hentet fra [https://www.nber.org/system/files/working\\_papers/w12497/w12497.pdf](https://www.nber.org/system/files/working_papers/w12497/w12497.pdf)
- Peri, G. (2016). Immigrants, Productivity, and Labor markets. Hentet fra <https://pubs.aeaweb.org/doi/pdfplus/10.1257/jep.30.4.3>
- Pischke, J.-S., & Velling, J. (1994). Wage and employment effects of immigration to Germany: an analysis based on local labor markets. Hentet fra <https://econpapers.repec.org/paper/zbwzewdip/9403.htm>
- Rustad Holseter, A. M. (2019). Hvordan klassifiseres en persons høyeste utdanningsnivå? SSB. Hentet fra <https://www.ssb.no/utdanning/artikler-og-publikasjoner/hvordan-klassifiseres-en-persons-hoyeste-utdanningsniva>
- SSB. (2022). *Brukermanual for microdata.no*. SSB. Hentet fra <https://www.microdata.no/wp-content/uploads/2022/04/brukermanual-no-1.pdf>
- Statistisk sentralbyrå. (2000). *Standarder for økonomiske regioner*. Hentet fra [https://www.ssb.no/a/publikasjoner/pdf/nos\\_c616/nos\\_c616.pdf](https://www.ssb.no/a/publikasjoner/pdf/nos_c616/nos_c616.pdf)
- Wooldridge, J. M. (2016). *Introductory Econometrics: A modern approach* (6. utg.). Cengage Learning.

## Appendiks

### A1 Utledning av agglomerasjonsmodellen

Antar at vi har flere sektorer i perfekt konkurranse som betegnes med indeksen  $j$ , og hver sektor produserer homogene endelige varer som kan handles fritt på tvers av byer.

Produksjonen i sektor  $j$  er gitt ved:

$$(34) Y_i^j = B^j \left\{ \int_0^{m_i^j} [y_i^j(h)]^{\frac{1}{1+\sigma^j}} dh \right\}^{1+\sigma^j}$$

Mellomprodukter er produsert av monopolbedrifter med teknologien:

$$(35) y_i^j(h) = \beta^j l_i^j(h) - \alpha^j$$

I likevekt vil alle bedriftene i enhver by og sektor sette den samme profitt maksimerende prisen  $q_i^j = w_i^j(1 + \sigma^j)/\beta^j$ . Fri inngang driver mellomproduktene til null profitt  $q_i^j y_i^j - w_i^j l_i^j = 0$ .

Ved bruk av likning (35) og prisregelen og løser ut et uttrykk for  $y_i^j$  viser at det er et fast nivå på produksjonen av mellomprodukter konsistent med null profitt i hver sektor:

Setter inn for prisregelen inn i nullprofitt betingelsen:

$$\left( \frac{w_i^j(1+\sigma^j)}{\beta^j} \right) y_i^j - w_i^j l_i^j = 0$$

Skriver om til et uttrykk for  $\beta^j$

$$\beta^j w_i^j l_i^j = w_i^j (1 + \sigma^j) y_i^j$$

$$\beta^j = \frac{(1+\sigma^j)y_i^j}{l_i^j}$$

Setter dette inn i (35):

$$y_i^j = \left( \frac{(1+\sigma^j)y_i^j}{l_i^j} \right) l_i^j - \alpha^j$$

$$y_i^j - (1 + \sigma^j)y_i^j = -\alpha^j$$

$$(36) \quad y_i^j = \frac{\alpha^j}{\sigma^j}$$

Setter likningene (35) og (36) lik hverandre muligj or l osning for den konstante arbeidsstyrken av hver mellomprodukt tilbyder:  $l_i^j = \alpha^j(1 + \sigma^j)/(\beta^j \sigma^j)$ . Dermed er likevekts mengden av produsenter av mellomprodukter i sektor  $j$  i by  $i$  gitt ved:

$$(37) \quad m_i^j = \frac{N_i^j}{l_i^j} = \frac{\beta^j \sigma^j}{\alpha^j(1 + \sigma^j)} N_i^j$$

Ved valg av enheter for mellomproduksjon kan vi sette  $\beta^j = (1 + \sigma^j) \left(\frac{\alpha^j}{\sigma^j}\right)^{\sigma^j/(1+\sigma^j)}$ . Substituerer likningene (36) og (37) inn i (34) gir den aggregerte produksjonen i sektor  $j$  i by  $i$ :

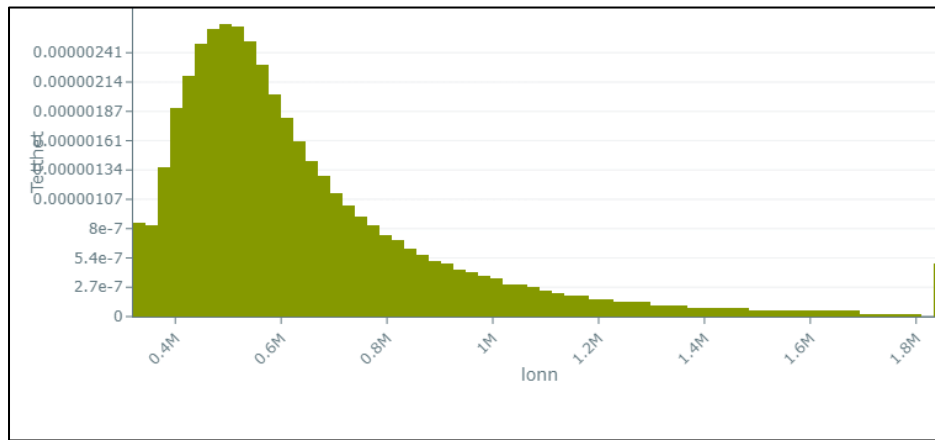
$$(38) \quad Y_i^j = B^j \left[ (y_i^j)^{\frac{1}{1+\sigma^j}} m_i^j \right]^{1+\sigma^j} = B^j (N_i^j)^{1+\sigma^j}$$

Nullprofitt i den endelige produksjonen impliserer at  $w_i^j N_i^j = P^j Y_i^j$  som betyr at l onningene er gitt ved:

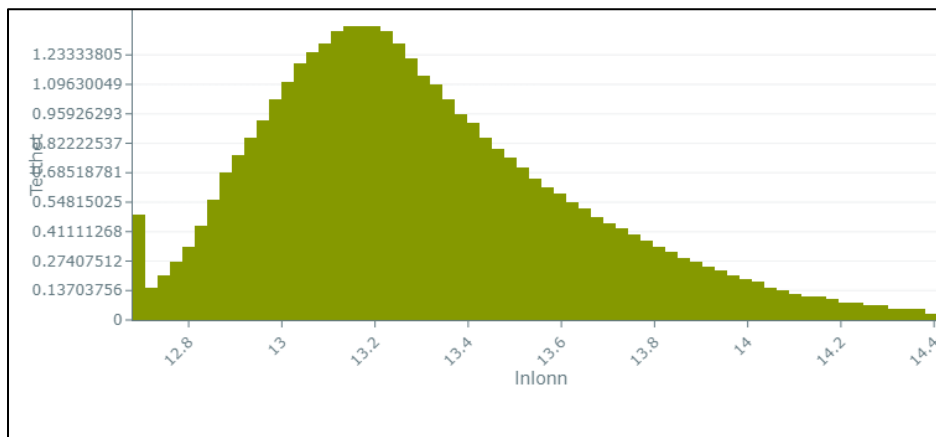
$$(39) \quad w_i^j = P^j B^j (N_i^j)^{\sigma^j}$$

## A2 Log-transformering av avhengig variabel

**Figur A2.1:** Fordelingen til den lineære lønnsvariabelen i paneldatasettet



**Figur A2.2:** Fordelingen til den log-transformerte lønnsvariabelen i paneldatasettet





## A4 Testresultater

**Figur A4.1:** Hausman-test som bekrefter valget av FE- over en RE-strategi

<b>Hausman-test</b>	
chi2(15): 2	3287.067
Prob > chi2:	0

**Figur A4.2:** RESET-test resultatet for feilspesifisering

<b>Ramseys RESET test</b>	
F(3, 4609409):	5796.690
Prob > F:	0

**Figur A4.3:** BP-test resultatene for heteroskedastisitet

<b>Breusch-Pagan</b>	
chi2(1):	253907.546
Prob > chi2:	0
<b>Breusch-Pagan, studentisert</b>	
chi2(1):	72665.209
Prob > chi2:	0
<b>Breusch-Pagan, f-test</b>	
F(1, 4609431):	73829.052
Prob > F:	0



## A5 Simultanitet analytisk

Ser nærmere på simultaniteten mellom befolkningsstørrelsen (N) og lønningene (W) innenfor to helt enkle regresjonsmodeller uten kontrollvariabler, og vi ser bort fra konstantene:

$$(1) W = \beta_1 N + u$$

$$(2) N = \alpha_1 W + v$$

- Setter inn for likning (1) inn i (2):

$$N = \alpha_1(\beta_1 N + u) + v$$

- Løser ut for N:

$$N - \alpha_1(\beta_1 N) = \alpha_1 u + v$$

$$N = \frac{\alpha_1 u}{1 - \alpha_1 \beta_1} + \frac{v}{1 - \alpha_1 \beta_1}$$

Setter dette inn i likning (1):

$$W = \beta_1 \frac{\alpha_1 u}{1 - \alpha_1 \beta_1} + \frac{v}{1 - \alpha_1 \beta_1} + u$$

Her ser vi at forklaringsvariabelen av hovedinteresse, befolkningsstørrelsen N er positivt korrelert med feilledet,  $cov(N, u) > 0$  som betyr at vi har brudd på eksogenitetsbetingelsen. Gitt rammeverket fra den monosentriske bymodellen og tidligere empiri har vi vist at  $\beta_1 > 0$  og  $\alpha_1 > 0$ , hvilket betyr at når denne simultaniteten ikke løses for vil OLS-estimatoren overestimere effekten av befolkningsstørrelsen på lønnen  $\hat{\beta}_1^{OLS} > \beta_1$ .

## A6 Multikollinearitet

**Figur A6.1:** Korrelasjonsmatrisen for alle forklaringsvariablene i den fullspesifiserte regresjonsmodellen

	stor_region	mellom_region	ainnv	vesten	finnv_v	D14	D15	D16	D17	kvinne	alder	hoyutd	gift
lnlonn	0.139680109	-0.044725139	0.016649423	-0.02002934	-0.06918726	-0.03384149	0.011975267	0.044798203	0.06514788	-0.2734656	0.115184857	0.235070613	0.13242314
stor_region		-0.428943764	0.044238287	0.044417241	0.086687916	0.001053534	0.001176649	-0.00080704	-0.00291339	0.030062752	-0.06953786	0.127240319	-0.02409706
mellom_region	-0.42894376		-0.00418669	-0.0154357	-0.03494231	-0.0024457	-0.00081715	0.000687329	0.00570096	-0.01642705	0.024072896	-0.05214628	0.033459665
ainnv	0.044238287	-0.004186693		-0.04321873	-0.03506802	-0.00035261	0.000121494	0.000407974	0.000778557	0.002639504	-0.04454731	0.025008617	-0.0118381
vesten	0.044417241	-0.015435695	-0.04321873		-0.03672611	0.000851929	0.000844876	0.000886218	-0.00074222	-0.009756	-0.05075835	0.04576141	0.01882706
finnv_v	0.086687916	-0.03494231	-0.03506802	-0.03672611		0.000208973	0.000183616	0.000198613	0.000170786	0.006301764	-0.04139564	0.002440223	0.057086852
D14	0.001053534	-0.002445702	-0.00035261	0.000851929	0.000208973		-0.25062196	-0.25094234	-0.25043308	0.000494732	-0.04244557	0.001230514	-0.00633288
D15	0.001176649	-0.000817155	0.000121494	0.000844876	0.000183616	-0.25062196		-0.25166296	-0.25115224	-0.000251	-0.00130995	0.000169589	0.000286783
D16	-0.00080704	0.000687329	0.000407974	0.000886218	0.000198613	-0.25094234	-0.25166296		-0.25147329	-0.00090996	0.041094679	-0.00097568	0.006186002
D17	-0.00291339	0.00570096	0.000778557	-0.00074222	0.000170786	-0.25043308	-0.25115224	-0.25147329		-0.00103773	0.083041683	-0.00163587	0.01121746
kvinne	0.030062752	-0.016427048	0.002639504	-0.009756	0.006301764	0.000494732	-0.000251	-0.00090996	-0.00103773		0.011652807	0.187989701	-0.03291481
alder	-0.06953786	0.024072896	-0.04454731	-0.05075835	-0.04139564	-0.04244557	-0.00130995	0.041094679	0.083041683	0.011652807		-0.09717044	0.251038912
hoyutd	0.127240319	-0.052146279	0.025008617	0.04576141	0.002440223	0.001230514	0.000169589	-0.00097568	-0.00163587	0.187989701	-0.09717044		0.044955046

**Figur A6.2:** Resultatene fra en VIF-test på nesten alle venstresidevariablene i den fullspesifiserte modellen (12)

Variance inflation factor	VIF	1/VIF
stor_region	1.3594	0.7356
mellom_region	1.3175	0.7590
ainnv	4.6655	0.2143
vesten	4.2678	0.2343
finnv_v	7.7531	0.1290
ainnv_stor_region	3.6905	0.2710
vest_stor_region	3.4769	0.2876
finnv_v_stor_region	6.6408	0.1506
ainnv_mellom_region	1.9274	0.5188
vest_mellom_region	1.7621	0.5675
finnv_v_mellom_region	2.0255	0.4937
D14	1.6141	0.6195
D15	1.6201	0.6172
D16	1.6266	0.6148
D17	1.6332	0.6123
kvinne	1.0399	0.9617
alder	1.1835	0.8450
hoyutd	1.0733	0.9317
gift	1.2150	0.8230
barn	1.1661	0.8576
Gj.snitt	2.5529	-

## A7 Fullstendig FE regresjon

Figur A7.1: FE regresjonen av den fullspesifiserte modellen (12) eksportert til Excel

MODELL 12						
R <sup>2</sup> i:	0.158418					
R <sup>2</sup> mellom:	0.063706					
R <sup>2</sup> total:	0.071877					
Corr(u <sub>i</sub> , X <sub>b</sub> ):	0.078774					
Sigma u:	0.341234					
Sigma e:	0.110819					
Rho:	0.904593					
Antall Obs:	4609433					
Antall grupper:	937231					
Min obs/grp:	0					
Snitt obs/grp:	4.91814					
Maks obs/grp:	5					
F(15,3672636):	46089					
Prob > F:	0					
Innlønn	Coef.	Std.feil	t	P> t	[95% Konf. intervall]	
stor_region	0.011584	0.001229	9.422648	0	0.009174	0.013993
mellom_region	-0.000563	0.001279	-0.440105	0.659861	-0.003071	0.001945
ainnv_stor_region	0.009482	0.00586	1.618187	0.105622	-0.002003	0.020966
vest_stor_region	0.002146	0.005763	0.372351	0.709632	-0.009149	0.013441
finnv_v_stor_region	-0.016463	0.006023	-2.733474	0.006267	-0.028268	-0.004659
ainnv_mellom_region	0.009491	0.005812	1.63311	0.102446	-0.0019	0.020882
vest_mellom_region	0.00241	0.005553	0.433932	0.664338	-0.008474	0.013293
finnv_v_mellom_region	-0.015275	0.007272	-2.10048	0.035687	-0.029529	-0.001022
D14	0.043294	0.000121	358.360643	0	0.043057	0.043531
D15	0.078317	0.00015	522.474259	0	0.078024	0.078611
D16	0.103319	0.000172	600.605978	0	0.102982	0.103656
D17	0.119162	0.000229	521.106015	0	0.118713	0.11961
hoyutd	0.121272	0.003713	32.663916	0	0.113996	0.128549
gift	0.010182	0.000508	20.031424	0	0.009186	0.011178
barn	0.001385	0.000172	8.069081	0	0.001048	0.001721
Konst	13.173015	0.002031	6484.756763	0	13.169033	13.176996

## A8 Stor-Oslo

Figur A8.1: Kommunene som inngår i Stor-Oslo regionen

Stor Oslo	
0301 - Oslo	
0213 - Ski	0234 - Gjerdrum
0216 - Nesodden	0235 - Ullensaker
0217 - Oppegård	0236 - Nes
0229 - Enebakk	0237 - Eidsvoll
0219 - Bærum	0238 - Nannestad
0228 - Rælingen	0239 - Hurdal
0230 - Lørenskog	0533 - Lunner
0231 - Skedsmo	0104 - Moss
0233 - Nittedal	0135 - Råde
0119 - Marker	0136 - Rygge
0122 - Trøgstad	0137 - Våler
0123 - Spydeberg	0602 - Drammen
0124 - Askim	0624 - Øvre Eiker
0125 - Eidsberg	0625 - Nedre Eiker
0127 - Skiptvet	0626 - Lier
0138 - Hobøl	0702 - Holmestrand
0211 - Vestby	0711 - Svelvik
0214 - Ås	0713 - Sande
0215 - Frogn	0714 - Hof
0220 - Asker	
0627 - Røyken	
0628 - Hurum	
0221 - Aurskog-Høland	
0226 - Sørum	
0227 - Fet	

## A9 Stor-Oslo heterogeniteter

Figur A9.1: Stor-Oslo heterogeniteter mhp. utdanning og alder

VARIABLER	(14) FE lnlonn	(15) FE lnlonn	(16) FE lnlonn	(17) FE lnlonn	(18) FE lnlonn
stor_oslo	0.0231*** (0.0015)	0.0274*** (0.0019)	0.0136*** (0.0025)	0.0221*** (0.0017)	0.0093*** (0.0031)
stor_region_o	0.0023 (0.0015)	0.0020 (0.0019)	0.0044** (0.0022)	0.0009 (0.0017)	0.0053* (0.0030)
mellom_region_o	-0.0012 (0.0013)	-0.0008 (0.0017)	-0.0016 (0.0018)	-0.0034** (0.0016)	-0.0003 (0.0021)
norinnvf_stor_oslo	-0.0063 (0.0063)	-0.0033 (0.0078)	-0.0158 (0.0103)	-0.0085 (0.0072)	0.0031 (0.0130)
hoyinnv_stor_oslo	-0.0023 (0.0065)	-0.0073 (0.0084)	0.0034 (0.0096)	-0.0042 (0.0080)	0.0078 (0.0109)
lavinnv_stor_oslo	-0.0127 (0.0081)	-0.0302 (0.0107)	0.0155 (0.0112)	-0.0135 (0.0083)	-0.0133 (0.0245)
norinnvf_stor_region_o	0.0102 (0.0065)	0.0170 (0.0085)	0.0022 (0.0093)	0.0103 (0.0072)	0.0077 (0.0137)
hoyinnv_stor_region_o	0.0036 (0.0059)	0.0039 (0.0074)	0.0051 (0.0100)	0.0059 (0.0070)	-0.0027 (0.0115)
lavinnv_stor_region_o	-0.0088 (0.0055)	-0.0095 (0.0074)	-0.0110 (0.0079)	-0.0097 (0.0062)	-0.0115 (0.0102)
norinnvf_mellom_region	0.0044 (0.0055)	0.0004 (0.0076)	0.0115 (0.0072)	0.0038 (0.0065)	0.0085 (0.0104)
hoyinnv_mellom_region	0.0034 (0.0055)	0.0033 (0.0070)	-0.0006 (0.0089)	0.0025 (0.0067)	-0.0030 (0.0092)
lavinnv_mellom_region	-0.0087 (0.0061)	-0.0153 (0.0083)	-0.0005 (0.0085)	-0.0106 (0.0069)	-0.0131 (0.0138)
Konstant	13.17***	13.31***	13.16***	13.13***	13.27***
Kontrollvariabler	Ja	Ja	Ja	Ja	Ja
Tidskontroller	Ja	Ja	Ja	Ja	Ja
Utvalg	Hele	Høy utdanning	Lav utdanning	Under 45år	Over eller lik 45år
Observasjoner	4609433	2300075	2309358	2104921	2504512
R-squared	0.159	0.202	0.115	0.232	0.093

Standardavvik i parenteser

\*\*\* p&lt;0.01, \*\* p&lt;0.05, \* p&lt;0.1

**Figur A9.2:** Stor-Oslo heterogeniteter mhp. kjønn

VARIABLER	(14) FE lnlonn	(19) FE lnlonn	(20) FE lnlonn
stor_oslo	0.0231*** (0.0015)	0.0233*** (0.0020)	0.0226*** (0.0023)
stor_region_o	0.0023 (0.0015)	0.0025 (0.0019)	0.0020 (0.0022)
mellom_region_o	-0.0012 (0.0013)	-0.0015 (0.0016)	-0.0005 (0.0020)
norinnvf_stor_oslo	-0.0063 (0.0063)	-0.0113 (0.0083)	0.0023 (0.0098)
hoyinnv_stor_oslo	-0.0023 (0.0065)	-0.0001 (0.0079)	-0.0052 (0.0116)
lavinnv_stor_oslo	-0.0127 (0.0081)	-0.0182* (0.0106)	-0.0021 (0.0122)
norinnvf_stor_region_o	0.0102 (0.0065)	0.0089 (0.0082)	0.0124 (0.0105)
hoyinnv_stor_region_o	0.0036 (0.0059)	0.0088 (0.0073)	-0.0072 (0.0100)
lavinnv_stor_region_o	-0.0088 (0.0055)	-0.0089 (0.0071)	-0.0090 (0.0085)
norinnvf_mellom_region	0.0044 (0.0055)	0.0063 (0.0070)	0.0015 (0.0087)
hoyinnv_mellom_region	0.0034 (0.0055)	0.0043 (0.0070)	0.0022 (0.0088)
lavinnv_mellom_region	-0.0087 (0.0061)	-0.0048 (0.0080)	-0.0150 (0.0092)
Konstant	13.17***	13.25***	13.06***
Kontrollvariabler	Ja	Ja	Ja
Tidskontroller	Ja	Ja	Ja
Utvalg	Hele	Menn	Kvinner
Observasjoner	4609433	2751502	1857931
R-squared	0.159	0.131	0.211

Standardavvik i parenteser

\*\*\* p&lt;0.01, \*\* p&lt;0.05, \* p&lt;0.1

