

Forord

Jeg vil først og fremst takke min meget dyktige veileder Hildegunn Ekroll Stokke for all hjelp jeg har fått i løpet av dette semesteret. Hennes hjelp, råd og tilbakemeldinger har vært upåklagelig.

Jeg vil også takke mine venner og medstudenter for noen flotte år i Trondheim. Takk til Stian for all hjelp jeg har fått med matematikken og takk til Tom André for gjennomlesningen av oppgaven. Til slutt vil jeg takke mine fantastiske foreldre for all støtte jeg har fått i løpet av studietiden. Uten dere hadde jeg ikke fått det til.

Øyvind Skjervold
Desember 2012

1. Innledning	1
2. Teoretisk rammeverk	3
2.1.1 Antagelser	3
2.1.2 Spatial Likevekt	3
2.1.3 Produktivitetssjokk i by b.....	6
2.2.1 Inkludering av agglomerasjonseffekter	8
2.2.2 Produktivitetssjokk i by b.....	9
3. Tidligere forskning	11
3.1.1 Agglomerasjon	11
3.1.2 Ulike mål på produktivitet.....	12
3.1.3 Løsningen på endogenitetsproblemet	12
3.1.4 Sortering.....	13
3.1.2 Urbane kostnader:	14
4. Økonometriske utfordringer	17
4.1 Eksogenitetsbetingelsen	18
4.2 Brudd på eksogenitetsbetingelsen	18
4.2.1 Utelatt relevant forklaringsvariabel.....	18
4.2.2 Målefeil	19
4.2.3 Simultanitet	19
4.3 Håndtering av endogenitetsproblemet	20
4.3.1 Instrumentmetoden (IV).....	20
5. Data	23
5.1 Avhengig variabler	23
5.1.1 Nominell inntektsnivå i norske kommuner.....	23
5.1.2 Boligpriser i norske kommuner	23
5.1.3 Reallønn i norske kommuner	24
5.2 Forklaringsvariabler	24
5.2.1 Befolkningstetthet	24
5.2.2 Vintertemperatur.....	24
5.2.3 Kyst.....	25
5.2.4 Utdanningsnivå	25
5.2.5 Befolkningsvekst.....	25
5.3 Instrumentvariabler	26
5.3.1 Befolkningstetthet 1875	26
5.3.2 Befolkningsvekst 1825 - 1900	26
5.4 Deskriptiv statistikk	27
6. Empirisk analyse	29
6.1 Agglomerasjon	29
6.1.1 Effekten av befolkningstetthet på inntektsnivå, OLS	29
6.1.2 Instrumentmetoden første steg.....	31
6.1.3 Instrumentmetoden andre steg, 2SLS.....	32
6.2 Urbane kostnader	34
6.2.1 Effekten av befolkningstetthet på boligpris, OLS	34
6.2.2 Instrumentmetoden første steg.....	36
6.2.2 Instrumentmetoden andre steg, 2SLS.....	37
6.3 Reallønn	38
6.3.1 Effekten av befolkningstetthet på reallønn, OLS.....	38
6.3.2 Instrumentmetoden andre steg	40
6.4 Robusthetstest	41

6.4.1	Alternativt instrument: Endring i befolkningstetthet 1825-1900	41
6.4.2	Effekten av befolkningstetthet på reallønn for enkeltår, 2SLS	42
6.5	Oppsummering	43
7	Konklusjon	45
	Referanser	47
	Appendix	50
	A Matematiske utregninger	50
	A.1 Førsteordensbetingelse for arbeidskraft.....	50
	A.2 Førsteordensbetingelsen for kapital	50
	A.3 Produktivitetssjokk	51
	A.4 Agglomerasjon.....	53
	A.4.1 Effekt på lønn i by b	55
	A.4.2 Effekt på boligpris.....	56
	A.4.3 Endring i reallønn	57
	B Tabeller.....	58
	B.1 Alternativt instrument.....	58
	B.2 2SLS-estimeringer for 2002, 2004 og 2007	58

1. Innledning

Et begrep som har fått stadig mer oppmerksomhet innen økonomisk forskning de siste tiår er agglomerasjon. Utviklingen viser at byer vokser seg større, både i form av økonomisk aktivitet og befolkning. Agglomerasjon er samlokalisering av bedrifter og konsentrasjon av mennesker innenfor et spesifikt geografisk område. Agglomerasjon kan observeres i mange ulike næringer. I California, USA, kan det observeres ved filmindustrien i Hollywood eller teknologiindustrien i Silicon Valley. Vi ser det også i vår egen hovedstad ved for eksempel varehandel på Alnabru.

Det finnes mye litteratur om fordeler og ulemper ved å bo i by kontra distriktene. Fordelene kalles agglomerasjonseffekter, og empiriske analyser viser at gjennomsnittslønnen er høyere i områder med høyt befolkningstall og høy befolkningstetthet. Denne lønnsgevinsten ved å bo i by oppstår på grunn av produktivitetsfordeler i de mest befolkede og tettbebygde områdene. Dersom det ikke fantes slike agglomerasjonseffekter så ville det ikke finnes storbyer (Combes et al. 2011). Motstykket til disse fordelene kalles urbane kostnader. Dette er kostnader knyttet til å bo i byer som for eksempel høyere boligpris, mer forurensning, mer trafikk og høyere kriminalitetsrate. Det er lett å observere trafikkaos i storbyer som Manhattan og London, samtidig som det ikke er noen hemmelighet at boligprisene er høyere i disse byene (Combes et al. 2012). Dersom slike urbane kostnader ikke fantes ville alle bodd i byene.

I denne oppgaven skal jeg studere det Fujita og Thisse (2002) døpte *"the fundamental tradeoff between agglomeration economies and urban cost"*. Inspirert av Combes et al. (2010, 2012) skal jeg utføre en teoretisk og en empirisk analyse av bytteforholdet mellom agglomerasjonseffekter og urbane kostnader. Agglomerasjonseffektene blir målt i gjennomsnittlig nominelt lønnsnivå, mens de urbane kostnadene måles i boligpriser. Jeg skal finne et estimat på elasticiteten til nominell lønn og boligpris med hensyn på befolkningstetthet. Hovedmålet er å finne ut hvordan økt befolkningstetthet påvirker reallønnen. Dette avhenger av hvilken effekt som dominerer av lønnseffekt og boligpriseffekt. Jeg finner at elasticiteten til

gjennomsnittlig nominelt lønnsnivå med hensyn på befolkningstetthet er på 2,34% etter å ha kontrollert for utdanningsnivå.

Mitt foretrukne estimat på elasticiteten til boligpris med hensyn på befolkningstetthet er på 4,42%, når jeg har kontrollert for befolkningsvekst og utdanningsnivå.

Dette resulterer til at reallønnen har en elasticitet med hensyn på befolkningstetthet på -2,40%, altså at boligpriseffekten dominerer.

.

2. Teoretisk rammeverk

I dette kapitlet skal jeg gjennomgå en relevant teoretisk modell som jeg legger til grunn for min empiriske analyse. Jeg bruker Enrico Moretti sin modell fra *Handbook of labor economics, volume 4b* fra 2011 som modellerer spatial likevekt i arbeidsmarkedet og boligmarkedet. Den er bygget på Rosen (1979) og Roback (1982) sin modell og viser langsiktig migrasjonslikevekt, hvor den marginale arbeider er indifferent mellom to byer. Høyere lønn skal motsvares av høyere boligpriser. Dette vises teoretisk i dette kapitlet og empirisk i kapittel 6.

2.1.1 Antagelser

Modellen omhandler to byer, kalt by a og by b, hvor det finnes kun én type arbeid. Begge byene er konkurransedyktig og produserer én type vare, y , som selges på det internasjonale markedet hvor prisen er lik 1. Teknologi har konstant skalaravkastning. Arbeiderenes indirekte nytte avhenger av nominell lønn, boligkostnader og lokale fasiliteter. Fasiliteter betraktes som et positivt gode man får tilgang på ved å bo i byen. Arbeid er homogen i ferdighet og smak og hver arbeider produserer én enhet arbeid. Arbeid har perfekt mobilitet slik at lokalt arbeid er uendelig elastisk. Land er det eneste i modellen som anses som urokkelig og antas fast. Ettersom arbeidere og bedrifter er mobile vil de lokalisere seg der hvor nytten og profitten blir maksimert. Én arbeider sørger for én enhet arbeid slik at tilbud av arbeidskraft bestemmes av arbeiderenes valg av lokalisering.

2.1.2 Spatial Likevekt

Den indirekte nyttefunksjonen til arbeidere i by c er

$$U_{ic} = w_c - r_c + A_c + e_{ic} \quad (c = a, b) \quad (1)$$

hvor c er enten by a eller by b og ligningen er lik for begge byene.

w_c er nominell lønn i by c , r_c er boligkostnaden og A_c er et mål på fasilitetene i byen. Man ser at høyere lønn og bedre fasiliteter fører til en økning nytten mens økte boligkostnader reduserer nytten. Restleddet e_{ic} representerer arbeider i sin tilknytting til byen.

Det kan for eksempel være at man er oppvokst i den byen, har familie der, at man er supporter av det lokale idrettslaget eller lignende. En høy e_{ic} betyr at tilknytningen til by c er sterk slik at byen er mer attraktiv uavhengig av reallønn og fasiliteter i by c. Arbeider i sin preferanse av by a i forhold til by b er

$$e_{ia} - e_{ib} \sim U[-s, s] \quad (2)$$

hvor s er en parameter som forteller hvor viktig tilknytningen til en by er for valget av bosted og forteller derfor hvor mobil arbeideren er. En høy s betyr at følelsen av tilknytning til byen er meget viktig. En lav s betyr at tilknytningen til byen er lite viktig og arbeideren vil være villig til å flytte for å oppnå en høyere reallønn eller det man betrakter som bedre fasiliteter. Hvis s er lik null er det perfekt mobilitet blant arbeiderne.

Arbeiderne velger by a kun hvis $e_{ia} - e_{ib} > (w_b - r_b) - (w_a - r_a) + (A_b - A_a)$.

I likevekt må arbeideren være likegyldig til valg av by.

Denne likevektsbetingelsen medfører at tilbudet av arbeidskraft er stigende og avhenger av s . Tilbud av arbeidskraft i by b er gitt ved

$$w_b = w_a + (r_b - r_a) + (A_a - A_b) + s \frac{N_b - N_a}{N} \quad (3)$$

hvor $N = N_a + N_b$ og er skrevet på logaritmeform. Det antas at samlet tilbud av arbeidskraft for begge byene, N , er fast. Det viktige med ligning (3) er at elastisiteten til tilbud av arbeidskraft i byene avhenger av arbeiderenes personlige preferanse av byen, altså grad av tilknytning til byen. Ved høy s er arbeiderne lite mobile og elastisiteten til tilbud av arbeidskraft er lav. Det vil tilsa en bratt kurve for tilbud av arbeidskraft. Det må finnes store lønnsgevinster for at arbeideren skal være villig til å flytte på seg. Det samme vil gjelde for fasilitetene, slik at de må bli markant bedre relativt til den andre byen for at arbeideren skal flytte. Det vil være omvendt ved en lav s , hvor arbeiderne er svært mobile og elastisiteten til tilbudet av arbeidskraft vil være høy. Dette tilsier en slakk kurve for arbeidskraftstilbudet og arbeideren vil være villig til å flytte på seg hvis det gir en lønnsøkning eller forbedring av fasilitetene. Et spesialtilfellet vil være om s er lik null som vil bety perfekt arbeidskraftmobilitet. Da vil kurven for tilbud av arbeidskraft være vannrett.

Ligningen forteller altså at for en gitt kurve for tilbud av arbeidskraft vil en økning i reallønnen eller forbedring av fasilitetene i by a føre til at arbeidere flytter fra by b til fordel for by a.

Vi har en Cobb-Douglas produktfunksjon med konstant skalaavkasting

$$y_c = \bar{X}_c \bar{N}_c^h \bar{K}_c^{1-h}$$

Ved å skrive om produktfunksjonen ovenfor på logaritmeform får vi

$$\ln y_c = X_c + hN_c + (1 - h)K_c \quad (4)$$

hvor $X_c = \ln \bar{X}_c$, $N_c = \ln \bar{N}_c$, $K_c = \ln \bar{K}_c$.

X_c og K_c er henholdsvis produktivitet og kapital i by c. Det antas at bedriftene har perfekt mobilitet. For å finne et uttrykk for lønn må man ta førsteordensbetingelsen for arbeidskraftkraft. Man deriverer produktfunksjonen y_c med hensyn på arbeidskraft, \bar{N}_c , og setter det lik lønnen

$$w_c = X_c - (1 - h)N_c + (1 - h)K_c + \ln h \quad (5)$$

hvor ligning (5) er skrevet på logaritmeform og vist i appendix A.1.

Man antar videre at det er et internasjonalt kapitalmarked hvor kapital er uendelig for en gitt pris i . Ligning (6) finner man ved å ta førsteordenbetingelsen for kapital, og deriverer produktfunksjonen med hensyn på \bar{K}_c

$$K_c = \frac{X_c}{h} + N_c + \frac{\ln(1 - h)}{h} - \frac{\ln i}{h} \quad (6)$$

hvor også ligning (6) er skrevet på logaritmeform og vist i appendix A.2.

Det antas at hver arbeider forbruker én enhet bolig. For å finne etterspørsel etter bolig så omorganiserer man ligning (3) og løser med hensyn på boligpris

$$r_b = (w_b - w_a) + r_a + (A_b - A_a) - s \frac{(N_b - N_a)}{N} \quad (7)$$

Tilbud av bolig er gitt ved

$$r_c = z + k_c N_c \quad (8)$$

hvor antall boliger i by c antas å være lik antall arbeidere. Parameteren k_c er elastisiteten til tilbud av bolig og avhenger av lokale lover og forskrifter.

Dersom det er lett å bygge nye hus og boliger i et område er k_c liten. Finnes det ingen regler eller restriksjoner mot bygging av nye boliger er den lik null. I motsatt fall så vil k_c være stor hvis det er slik at bygging av nye boliger er svært vanskelig å få til. Dersom det er umulig å bygge nye boliger i området vil parameteren være uendelig.

Ligning (3) og (5) gir oss likevekt i arbeidsmarkedet, mens ligning (7) og (8) gir oss likevekt i boligmarkedet.

Modellen består av 9 endogene variabler;

Produksjon (y_c), sysselsetting (N_a, N_b), lønn (w_a, w_b), kapital (K_a, K_b) og boligkostnad (r_a, r_b). Dette gir 9 ligninger for å løse modellen:

$$w_b = w_a + (r_b - r_a) + (A_a - A_b) + s \frac{N_b - N_a}{N} \quad (3)$$

$$\ln y_c = X_c + hN_c + (1 - h)K_c \quad (4)$$

$$w_a = X_a - (1 - h)N_a + (1 - h)K_a + \ln h \quad (5a)$$

$$w_b = X_b - (1 - h)N_b + (1 - h)K_b + \ln h \quad (5b)$$

$$K_a = \frac{X_a}{h} + N_a + \frac{\ln(1 - h)}{h} - \frac{\ln i}{h} \quad (6a)$$

$$K_b = \frac{X_b}{h} + N_b + \frac{\ln(1 - h)}{h} - \frac{\ln i}{h} \quad (6b)$$

$$r_a = z + k_a N_a \quad (8a)$$

$$r_b = z + k_b N_b \quad (8b)$$

$$N = N_a + N_b \quad (*)$$

2.1.3 Produktivitetssjokk i by b

Vi tar i bruk en to-periode modell hvor man antar at by b opplever et eksogent produktivitetssjokk av størrelsen Δ i periode 2, slik at $x_{b2} = x_{b1} + \Delta$.

For å finne effekten av dette sjokket må vi løse ut ligningene i modellen slik at de endogene variablene kun blir uttrykt ved eksogene variabler og parametere.

Fullstendig utregninger for produktivitetssjokket finnes i appendix A.3.

Først setter man uttrykket for kapital inn i ligningen for nominell lønn for begge byene og får:

$$w_a = \frac{1}{h} X_a + \frac{1 - h}{h} \ln(1 - h) - \frac{1 - h}{h} \ln i + \ln h \quad (5a)'$$

$$w_b = \frac{1}{h} X_b + \frac{1 - h}{h} \ln(1 - h) - \frac{1 - h}{h} \ln i + \ln h \quad (5b)'$$

Man må først finne et uttrykk for tilbud av arbeidskraft i by b

Vi skriver om ligning (3) og løser med hensyn på N_b

$$N_b = \frac{N}{s} \left((w_b - w_a) - (r_b - r_a) - (A_a - A_b) \right) + N_a \quad (3)'$$

Vi setter inn uttrykkene for lønn og boligpris (w_b, w_a, r_a og r_b)

$$N_b = \frac{N}{s} \left[\frac{1}{h} (X_b - X_a) - (k_b N_b + k_a N_a) - A_a + A_b \right] + N_a$$

Ved å omformulere ligning (*) og bruke at $N_a = N - N_b$ får man

$$N_b = \frac{\frac{1}{h} (X_b - X_a) + k_a N - A_a + A_b + s}{\frac{2s}{N} + k_b + k_a} \quad (3)''$$

som er et uttrykk for tilbud av arbeidskraft i by b.

Sjokket gjør at produktiviteten er høyere i by b enn i by a. Dette tiltrekker arbeidere fra by a til by b. Vi setter inn sjokket i ligningen for N_b og trekker fra periode 1 for å finne endringen

$$N_{b2} - N_{b1} = \frac{N}{h[(k_a + k_b)N + 2s]} \Delta \geq 0 \quad (9)$$

Vi ser fra ligning (9) at antallet som migrerer fra a til b er større jo høyere elastisiteten til tilbud av arbeidskraft er (mindre s) og jo høyere elastisiteten til tilbud av bolig er (mindre k_b). Dette er lite overraskende ettersom en lav s indikerer at tilknytningen til en by har lite å si for lokaliseringen til arbeiderne, mens en lav k_b betyr at det er enkelt å bygge flere boliger. Det gir flere arbeidere muligheten til å flytte på seg. På grunn av produktivitetsfordeler vil den økte arbeidskraften i by b føre til økt lønnsnivå i by b. Økningen i nominell lønn i by b finner man ved å trekke den gamle lønnen i periode 1 fra den nye lønnen som oppstår ved produktivitetssjokket

$$w_{b2} - w_{b1} = \frac{1}{h} \Delta \quad (10)$$

Den økte innflyttingen fra by a til by b vil medføre at også boligkostnadene øker.

Omfanget av de økte boligkostnadene avhenger av elastisiteten til tilbud av bolig i by b i forhold til by a

$$r_{b2} - r_{b1} = \frac{k_b N}{h[N(k_a + k_b) + 2s]} \Delta \geq 0 \quad (11)$$

Økningen i boligkostnaden er større jo lavere elastisiteten til boligtilbudet i b er i forhold til by a (stor k_b). Vi ser fra ligning (10) og (11) at økningen i nominell lønn er større enn økningen i boligkostnad. Dette betyr at reallønnen i by b øker

$$(w_{b2} - w_{b1}) - (r_{b2} - r_{b1}) = \frac{k_a N + 2s}{h[N(k_a + k_b) + 2s]} \Delta \geq 0 \quad (12)$$

Selv om produktivitetssjokket kun er i by b, vil by a også bli påvirket av sjokket.

Ettersom folk flytter ut fra byen, vil boligkostnadene bli redusert

$$r_{a2} - r_{a1} = \frac{k_a N}{h[N(k_a + k_b) + 2s]} \Delta \leq 0 \quad (13)$$

Nominell lønn i a endrer seg ikke på grunn av bedriftene i by a nå bruker mer kapital.

Det betyr at produktivitetssjokket i b vil bidra til høyere reallønn også i by a

$$(w_{a2} - w_{a1}) - (r_{a2} - r_{a1}) = \frac{k_a N}{h[N(k_a + k_b) + 2s]} \Delta \geq 0 \quad (14)$$

Vi ser fra ligning (12) og (14) at reallønnen for periode 2 i by b er høyere enn i by a.

Ved perfekt mobilitet ($s=0$) ville endringen vært den samme for begge byene.

2.2.1 Inkludering av agglomerasjonseffekter

I ligning (4) var X_c eksogen gitt. Nå inkluderer vi muligheten for agglomerasjonseffekter slik at X_c er en funksjon av nivået av økonomisk aktivitet i by c. Man antar at produktiviteten i en by er en funksjon av antall arbeidere i den bestemte byen, slik at $X_c = f(N_c)$ med $f' > 0$.

Arbeiderenes valg av bosettelse genererer en positiv ekstern påvirkning på produktiviteten. Økt produktivitet resulterer i økt migrasjon i det området og er vist ved ligning (9). Befolkningsveksten gir opphav for økt produktivitet, som igjen øker befolkningen, og slik vil det fortsette helt til bolig-og tomteprisene er høye nok til at den marginale arbeider og bedrifter er likegyldig til hvor de vil være.

Man antar homogen arbeidskraft og at de positive effektene på produktivitet ved agglomerasjon er gitt ved $X_c = x_c + \gamma N_c$, hvor parameteren γ er styrken på agglomerasjonseffektene. Vi får nå en ny ligning for nominell lønn i by c

$$w_c = x_c + (\gamma - (1 - h))N_c + (1 - h)K_c + \ln h \quad (15)$$

En økning i antall arbeidere sysselsatt i en by har to motstridende effekter.

På den ene siden vil en økning av arbeidere redusere det marginale produktet av arbeidskraft, på den annen side vil en befolkningsvekst øke produktiviteten.

Hvis agglomerasjonseffektene er sterke nok, altså $\gamma > (1 - h)$, vil etterspørselen etter arbeidskraft være en stigende kurve.

2.2.2 Produktivitetssjokk i by b

På samme måte som uten agglomerasjon ovenfor antar vi at by b opplever et produktivitetssjokk slik at $x_{b2} = x_{b1} + \Delta$.

Produktivitetssjokket vil presse nominell lønn opp. Høy nominell lønn tiltrekker flere arbeidere til by b. Befolkningsveksten som by b nå opplever genererer økt produktivitet slik at produktivitetssjokket blir større i omfang enn hva sjokket i utgangspunktet var. Et produktivitetssjokk på Δ gir en større økning i nominell lønn

$$w_{b2} - w_{b1} = \frac{h[N(k_b + k_a) + 2s] - \gamma N}{h[h(N(k_b + k_a) + 2s) - 2\gamma N]} \Delta \geq \Delta \geq 0 \quad (16)$$

På bakgrunn av agglomerasjonseffektene blir økningen i nominell lønn større enn det faktiske sjokket. Uten slike effekter ville effekten på nominell lønn vært som ved ligning (10). Jo større parameteren γ er, jo større vil effekten på nominell likevektslønn være i by b:

$$\frac{\partial(w_{b2} - w_{b1})}{\partial\gamma} = \frac{N\Delta h^2(N(k_a + k_b) + 2s)}{\{h[h(N(k_b + k_a) + 2s) - 2\gamma N]\}^2} \geq 0 \quad (17)$$

Utrekningene for ligning (16) og (17) er vist i appendix A.4.1.

Høyere produktivitet tiltrekker flere arbeidere. Det vil resultere i en økning av arbeidsstyrken i by b. Full utregning i appendix A.4.

$$N_{b2} - N_{b1} = \frac{N}{h[(N(k_a + k_b) + 2s) - 2\gamma N]} \Delta \geq 0 \quad (18)$$

Det antas at tilbudet av bolig ikke er uendelig elastisk. Da vil den økte befolkningen i by b føre til økte boligkostnader

$$r_{b2} - r_{b1} = \frac{k_b N}{h[N(k_a + k_b) + 2s] - 2\gamma N} \Delta \geq 0 \quad (19)$$

Hvor stor økning man får i ligning (16) og (19) avhenger av størrelsen til γ , og utregningene vises i appendix A.4.2.

Ettersom både nominell lønn og boligkostnadene øker i by b er det uklart hvilken effekt dette har på reallønnen. Endring i reallønnen i likevekt er:

$$(w_{b2} - w_{b1}) - (r_{b2} - r_{b1}) = \frac{h(K_a N + 2s) - \gamma N}{h\{h[N(k_a + k_b) + 2s - 2\gamma N]\}} \Delta \quad (20)$$

som forteller at endringen i reallønnen avhenger av størrelsen til agglomerasjonseffektene (appendix A.4.3). Vi deriverer reallønnen med hensyn på γ

$$\frac{\partial((w_{b2} - w_{b1}) - (r_{b2} - r_{b1}))}{\partial \gamma} = \frac{\Delta N h^2 [N(k_a + k_b) + 2s]}{\{h[N(k_a + k_b) + 2s - 2\gamma N]\}^2} \quad (21)$$

Uttrykket gitt ved ligning (21) kan både være positivt og negativt (appendix A.4.3). Hvis elasticiteten til boligtilbud i by b er større eller lik elasticiteten i a, er den deriverte positiv (lav k_b). I motsatt fall vil den deriverte være negativ hvis elasticiteten til boligtilbudet i by b er mindre enn elasticiteten i by a. Dersom elasticiteten til boligtilbudet i by b er null ($k_b = \infty$) vil det ikke være noen endring i reallønnen. Dette er lett å forstå ettersom hvis elasticiteten til boligtilbud er null vil det ikke være mulig å bygge nye boliger.

Kort oppsummert sier modellen til Moretti at positiv befolkningsvekst vil resultere i en økning i både inntektsnivå og boligpris, men effekten på reallønnen er uklar. Effekten avhenger av styrken på agglomerasjonseffektene og elasticiteten til boligtilbud vist ovenfor ved ligning (21). Jeg skal teste disse konklusjonene empirisk. Hypotesene mine er:

1. Befolkningsvekst vil gi økt inntektsnivå
2. Befolkningsvekst vil gi økte boligpriser

Jeg vil også finne ut om det er inntektseffekten eller boligpriseffekten som dominerer, slik at jeg kan finne effekten av økt befolkning på reallønnen.

Dersom inntektseffekten er sterkest, vil effekten på reallønn være positiv, men hvis boligpriseffekten dominerer er effekten på reallønn negativ. Resultatene vises i kapittel 6.

3. Tidligere forskning

I dette kapitlet skal jeg gjennomgå tidligere forskning som er gjort på dette området. Agglomerasjonseffekter har blitt grundig utforsket. At det eksisterer slike effekter hersker det liten tvil om og det har blitt godtatt blant økonomer. Det er bred enighet om at det er en høyere produktivitet i store byer. Det har derimot vært vanskelig å identifisere slike effekter, finne årsaken til hvorfor de finnes og finne empiri som kan bekrefte det. Motstykket til slike effekter, såkalte urbane kostnader, har det derimot blitt forsket relativt lite på.

3.1.1 Agglomerasjon

Det er bred enighet blant forskere om at det finnes fordeler ved at bedrifter lokaliserer seg innenfor samme område. Man kan gå helt tilbake til Alfred Marshall sin analyse av lokaliseringfenomenet i 1890 som starten på slik forskning.

Han la vekt på såkalt teknologiske *spillovers*, og hevdet at bedrifter adopterte ny teknologi fra bedriften i nærheten (Marshall, 1920). Flere forskere har forsøkt å finne årsaken til disse positive effektene ved at folk hoper seg opp på samme sted. Duranton og Puga (2004) argumenterte for at fordelene ved et stort arbeidsmarked gjorde så flere kunne dele på kostnaden ved fasilitetene knyttet til et område. Matchingen mellom arbeidsgiver og arbeidstaker ble også bedre, ettersom sannsynligheten for at de finner hverandre øker med befolkningsveksten. Det vil også kunne føre til at man lærer av hverandre i større grad. Costa og Kahn (2000) støttet argumentet deres angående matching ved at større arbeidsmarked hjelper husholdninger hvor flere er i arbeid. Ved større arbeidsmarked er sannsynligheten større for at eksempelvis både mor og far finner seg arbeid. Wheeler (2001) konkluderer med at større marked har høyere lønn og produktivitet ved bruk av data fra USA.

Det er i utgangspunktet lett å tenke seg at det er naturlige fortrinn som er bakgrunnen for produktivetsfordelene i ulike områder. Men Ellison og Glaeser (1999) finner derimot at selv om naturlige fortrinn spiller en rolle, så forklarer det bare 20% av industrienes lokalisering. Henderson (1997) finner lignende resultat, hvor han bruker paneldata med faste effekter for området og konkluderer med at naturlige fortrinn ikke forklarer en tilstrekkelig del av agglomerasjonseffektene. Derfor er ikke naturlige fortrinn ved produksjon svaret på hvorfor store byer har høyere produktivitet.

Ellison og Glaeser (1997) bruker en "dart-board" stil metode når de ser på grad av geografisk konsentrasjon mellom industriene. De finner at de fleste industrier er lokalisert sammen til en viss grad, men i mange tilfeller er lokalisering liten. Industrier med høy lokalisering var industrier innen tobakk, tekstil og lær, mens industrier med lav lokalisering var innen papir, gummi og plastikk

3.1.2 Ulike mål på produktivitet

Produktivitet kan være vanskelig å måle, og det finnes forskjellige teorier om hvilket mål som er best egnet for å måle produktivitet. Glaeser et al (1992) og Henderson et al (1995) brukte sysselsettingsvekst som mål på produktivitet for data fra USA. Carlton (1983) og Rosenthal og Strange (2003) brukte etablering av nye bedrifter som mål for produktivitet. Argumentet var at gründere ville lokalisere seg i områder hvor produktiviteten var størst, ettersom de vil maksimere profitten. Ciccone og Hall (1996) brukte produksjon per arbeider som mål for produktivitet. Et fjerde alternativ for måling på produktiviteten er å bruke nominell lønn. Dette er en metode brukt av blant andre Glaeser og Mare (2001), Wheaton og Lewis (2002) og Carlsen et al (2012). Fordelen ved å måle produktivitet i form av nominell lønnsnivå er at det er svært enkelt å få tilgang på data. Resultatene deres viser en signifikant sammenheng mellom lønninger og bystørrelse. Dette er hovedårsaken for at jeg velger å bruke det som mål i denne analysen. Det siste målet er husleie for bedrifter. Dekle og Eaton (1999) brukte dette som mål i deres analyse av agglomerasjonseffekter for Japan. I motsetning til for eksempel lønn er data for leiekostnader vanskeligere å finne. Uansett hvilket av målene som brukes så er resultatet en positiv sammenheng mellom produktivitet og størrelsen på byen. Utfordringen er å finne ut om det er en kausal sammenheng.

3.1.3 Løsningen på endogenitetsproblemet

En utfordring som dukker opp i de fleste forskningsarbeidene om agglomerasjon er at dersom et område har en underliggende produktivitetsfordel vil det tiltrekke seg både flere bedrifter og arbeidere og dermed bli større på grunn av dette. Dette blir i litteraturen kalt et endogenitetsproblem. For å takle dette problemet bruker Ciccone og Hall (1996) historisk befolkningsmønster som instrument. Tanken er at drivkreftene bak høyere produktivitet i dag ikke er de samme som i gamle dager. Ciccone og Hall (1996) viser at en økning i befolkningstetthet er korrelert med en

signifikant økning i produksjon per arbeider. De finner ved bruk av data fra USA at en dobling av befolkningstettheten vil øke produktiviteten øke med ca. 6%. Glaeser et al. (2011) finner at befolkningsstørrelsen blant amerikanske fylker i dag er sterkt korrelert med befolkningstallene fra midten av 1900-tallet. Eaton og Eckstein (1997) kommer frem til det samme ved bruk av tall fra Frankrike og Japan. Ciccone (2002) finner ved bruk av tall fra Europa en produktivitetsøkning på 4,5%. De fleste forskningsarbeidene finner at en dobling av befolkningsstørrelse øker produktiviteten med mellom 3-8 prosent (Puga 2010).

3.1.4 Sortering

Puga (2010) hevder at høyere lønn i tettbebygde områder kan ses på som bevis på høyere produktivitet. For arbeidere vil høyere pendlingskostnader og boligkostnader være motvekten til høyere lønn. Høyere lønn og høyere leiekostnader for bedriftene ville ført til at de ville lokalisert seg i andre områder om det ikke fantes signifikante produktivitetsfordeler. Store byer kan ha høyere produktivitet fordi innbyggerne er produktive, ikke fordi byene i seg selv gjør de mer produktive (Carlsen et al. 2012). Det kan være slik at flere dyktig arbeidere samler seg i store byer, såkalt sortering. Da vil lønnsgevinsten ved å bo i by reflektere de bedre egenskapene arbeiderne i byen besitter istedenfor noen virkelig fordeler ved urban lokalisering.

Nyere og mer avanserte analyser tar høyde for denne sorteringen ved bruk av individdata. Man kontrollerer for observerbar karakteristik som alder, kjønn og utdanning samtidig som man kontrollerer for uobserverbar karakteristik som ferdigheter. For å kunne kontrollere for slik uobserverbar karakteristik utnytter de paneldata og inkludere arbeider faste effekter. Glaeser og Mare (2001), Combes et al (2010) og Carlsen et al (2012) gjør alle dette. Glaeser og Mare (2001) kontrollerer for observerbare ferdigheter, instrumenterer for urban beboelse ved bruk av foreldres bakgrunn og utnytter paneldata ved å inkluderer arbeider faste effekter for data for USA. De finner en signifikant lønnspremie ved å bo og arbeide i tettbebygde byer, men omfanget blir mindre når de kontrollerer for disse variablene. Combes et al (2010) bruker også slike arbeider faste effekter, og kontrollerte for fruktbar jord ved bruk av fransk data.

De argumenterte for at fruktbar jord var en viktig grunn til bosettelse når landbruk var en større næring. De finner at størrelsen på agglomerasjon reduseres med en halv prosent. Deres estimerte elasticitet på lønn med hensyn på befolkningstetthet er på 2%. I likhet med de to nevnte finner også Carlsen et al. (2012) en lavere elasticitet når de kontrollerer for arbeider faste effekter. Deres estimat er en elasticitet på 1,7% ved bruk av norsk data. Konklusjonen fra de ulike forskningsarbeidene er at selv om de er gjort på forskjellig land finnes det sterke bevis på at det finnes sortering av arbeidere med gode egenskaper i de urbane områdene som bidrar til høyere lønnsnivå.

Men ikke alle gir agglomerasjon æren for den relativt høye produktiviteten til storbyene. Melitz og Ottaviano (2008) argumenterte for at byer med mange bedrifter bidro til høyere konkurranse slik at de mindre produktive bedriftene ble utkonkurrert. Dette gjør at snittet for produktivitet i de store byene er høyere enn i de små. Combes et al (2010) motbeviser dette i sin forskning og finner at produktivitetsskjellene i franske byer i all hovedsak forklares av agglomerasjon.

3.1.2 Urbane kostnader:

I motsetning til agglomerasjonseffekter har det blitt forsket lite på kostnadene knyttet til økt befolkning. Det faktum at det er finnes urbane kostnader kan alle være enig om, men det har blitt gjort lite empirisk forskning på dette området.

Det mest omfattende arbeidet rundt dette temaet er gjort av Combes et al (2012) som tar i bruk et avansert datasett for franske kommuner. De kontrollerer for landområdet, befolkningsvekst, temperatur, geografiske karakteristikk som tilgang til sjø og hvor lang kommunen er i forhold til de andre. De kontrollerer også for geologisk karakteristikk som fruktbar jord, og sosioøkonomiske variabler som inntekt per innbygger og utdanningsnivå. Som mål på hvor bra fasiliteter en kommune har bruker de andel 1-stjerners hotellrom. De finner en elasticitet på urbane kostnader med hensyn på befolkning på 4,1%. Kahn (2010) ser på det han kaller offentlige onder i USA. Disse ondene er pendling, kriminalitet og forurensning. Kahn finner at pendlingstiden er lengre for arbeiderne i de store urbane strøkene, men at elasticiteten til pendling med hensyn på befolkningsstørrelse er lav med 13%.

Kriminalitet per innbygger er også større i de store byene, men tallene er på vei ned slik at frykten for økt kriminalitet ved å bo i storby blir mindre over tid.

Det samme gjelder forurensing som i de store byene har redusert de siste årene.

Han konkluderer med at disse ondene var mer fremtredende før i tiden, og at storbyenes offentlige onder nærmer seg samme tall som utenfor de store byene.

Koster (2012) finner at agglomerasjon er en viktig determinant for leiekostnader ved kontorplass. Ifølge Gyourko et al (2010) kan forskjellen i boligpris også komme av at det finnes sortering av rike husholdninger i urbane områder som utbyr andre for boligene, slik at prisen presses opp.

4. Økonometriske utfordringer

Når man skal gjøre en slik empirisk forskning som jeg nå skal gjøre dukker det om en rekke utfordringer. Metoden jeg bruker er minste kvadraters metode (OLS), fra Wooldridge, 2009. Det går ut på å velge estimatorene som gjør slik at summen av de kvadrerte avvik (residual) minimeres. Utfordringen er å få forventingsrette og konsistente estimater ved bruk av OLS. For å få til dette må en rekke forutsetninger oppfylles. I dette kapittelet skal jeg gå igjennom disse, og forklare hvordan jeg kan ta høyde for disse slik at mine estimat blir forventingsrette og konsistente.

I min tversnittsanalyse for både agglomerasjonseffekter og urbane kostnader vil hovedregresjonene se slik ut ved bruk av gjennomsnittsdata

$$\ln \text{inntekt}_i = \beta_0 + \beta_1 \ln \text{tetthet}_i + v_i \quad (i)$$

$$\ln \text{boligpris}_i = \alpha_0 + \alpha_1 \ln \text{tetthet}_i + u_i \quad (ii)$$

hvor jeg til slutt lager en regresjon for reallønn

$$\ln \text{reallønn}_i = \delta_0 + \delta_1 \ln \text{tetthet}_i + \varepsilon_i \quad (iii)$$

Fotnote *i* forteller hvilken kommune vi ser på. α_0 , β_0 og δ_0 er konstantledd, α_1 og β_1 og δ_1 er parameterne vi ønsker å estimere og u_i , v_i og ε_i er et stokastiske restledd. Restleddene i ligning (i) og (ii) forteller oss at hverken sammenhengen mellom inntekt og befolkningstetthet eller boligpris og befolkningstetthet er helt eksakt.

Ettersom reallønnen er nominell lønn minus boligpris vil heller ikke sammenhengen i ligning (iii) være eksakt. Restleddene i ligning (i) – (iii) fanger opp andre effekter som kan påvirke boligpris og lønn. Videre i dette kapittelet bruker jeg en generell og enkel regresjonsmodellen for å forklare hvilke økonometriske utfordringer man støter på ved en slik analyse som jeg skal foreta meg. Modellen blir da slik

$$y_i = \alpha_0 + \alpha_1 x_i + u_i \quad (*)$$

og unngår å spesifisere hvilke variabler av y_i og x_i som gjelder i mitt eksakte tilfelle, ettersom utfordringene er like for ligningene (i) - (iii)

4.1 Eksogenitetsbetingelsen

En forutsetning som må være tilstede for restleddet er

$$E(u_i) = 0$$

som betyr at forventet verdi på restleddet er lik null, og

$$\text{Cov}(x_i, u_i) = 0,$$

som betyr at det ikke er kovarians mellom forklaringsvariabelen x_i og restleddet.

Jeg slår sammen disse to forutsetningene og skriver

$$E(u_i|x_i) = 0$$

Hvis disse forutsetningene er oppfylt så har vi estimat som er konsistente og forventingsrette. Dette kalles eksogenitetsbetingelsen og er et krav om at variabel x_i er strengt eksogen. Det antas også at variansen til restleddet er konstant, såkalt homoskedastisitet

$$\text{var}(u_i|x_i) = \sigma^2$$

4.2 Brudd på eksogenitetsbetingelsen

Ved brudd på betingelsene nevnt ovenfor vil det oppstå skjevhet i estimatene.

For at mine estimater skal bli så gode og presise som mulig, må jeg være obs på en rekke økonometriske problemer som kan være til stede.

4.2.1 Utelatt relevant forklaringsvariabel

Det kan for eksempel være at tversnittmodellene (i) – (iii) er underspesifisert.

Det betyr at det finnes én eller flere forklaringsvariabler som ikke er spesifisert i modellen, men som påvirker boligpris og inntekt. Dersom en kommune har gode offentlige tjenester og fasiliteter kan det være at folk flytter dit på tross av at boligprisen er høy, slik at effekten er overestimert. I regresjonen for boligpris kontrollerer jeg for utdanning, befolkningsvekst, vintertemperatur og en dummyvariabel for om kommunen ligger langs kysten. I regresjonen for inntekt kontrolleres det også for utdanning, vintertemperatur og dummy for kystlinje.

Modellene ser da slik ut

$$\ln \text{inntekt}_i = \beta_0 + \beta_1 \text{tett}_i + \beta_2 \text{utd}_i + \beta_3 \text{temp}_i + \beta_4 \text{kyst}_i + v_i \quad (i)'$$

$$\ln \text{boligpris}_i = \alpha_0 + \alpha_1 \ln \text{tett}_i + \alpha_2 \text{utd}_i + \alpha_3 \text{vekst}_i + \alpha_4 \text{temp}_i + \alpha_5 \text{kyst}_i + u_i \quad (ii)'$$

$$\ln \text{reallønn}_i = \delta_0 + \delta_1 \ln \text{tett}_i + \delta_2 \text{utd}_i + \delta_4 \text{vekst}_i + \varepsilon_i \quad (iii)'$$

hvor utd_i er et mål på utdanningsnivået for personer over 16 år, oppgitt i prosentandel av befolkningen og variabelen $vekst_i$ er årlig befolkningsvekst.

Dette betyr at koeffisientene i ligning (i) – (iii) blir feil hvis ligning (i)' – (iii)' er de sanne, riktige modellene. Selv om jeg kontrollerer for disse variablene kan utelatt forklaringsvariabel fortsatt være et problem. Det kan være andre geografiske karakteristikk som for eksempel tilgang til annen natur som skog og fjell, infrastruktur som godt utarbeidet veinett eller nærhet til flyplass og jernbane. Disse variablene kan føre til produktivtetsfordeler som påvirker inntektsnivå, eller bidra til økt etterspørsel etter bolig i den kommunen som vil drive boligprisen opp. Mangelfull informasjon kan derfor gi feilaktige og skjeve estimater.

4.2.2 Målefeil

I de fleste økonometriske analyser er variablene til en viss grad beheftet med målefeil. Med målefeil menes generelle avvik mellom en teoretisk korrekt definert variabel og den variabelen vi faktisk observerer (Wooldridge, 2009). Det kan gjelde både for den avhengig variabelen og forklaringsvariabelen. Målefeil i den avhengig variabelen vil normalt ikke gi opphav til skjevhet i OLS-estimatorene gitt at korrelasjonen mellom restleddet og forklaringsvariabelen(e) er lik null, men det kan resultere i en høyere varians og standardavvik. Målefeil i forklaringsvariabelen kan gi skjevhet hvis variabelen er korrelert med restleddet. Det som betyr noe for størrelsen på skjevheten er variansen til målefeilen i forhold til variansen i den sanne forklaringsvariabelen.

4.2.3 Simultanitet

Det siste potensielle problemet som kan gi brudd på eksogenitetsforutsetningen er om den avhengig variabelen og forklaringsvariabelen er simultant bestemt, slik at vi har motsatt kausalitet. Det er for eksempel ikke utenkelig at folk flytter dit hvor boligprisen er lav. Samtidig kan det tenkes at det er lav befolkning der hvor boligprisene er høye. Det kan også være sannsynlig at arbeidere flytter dit hvor lønnsnivået er høyt. Det samme kan gjelde for lønnsregresjonen. Dersom folk flytter til en kommune på grunn av høyere lønn vil økt befolkning gi økt lønn slik at effekten av befolkning blir overestimert.

Jeg illustrerer dette med hjelp av ligning (*) ovenfor hvor

$$y_i = \alpha_0 + \alpha_1 x_i + u_i \quad (*)$$

men vi antar at

$$x_i = \beta_0 + \beta_1 y_i + \gamma_1 z_1 + v$$

Her er y_i og x_i begge endogene variabler, mens z_1 er eksogen. Det blir problematisk å undersøke kausaleffekten av x_i på y_i når x_i avhenger av y_i . Så lenge disse to variablene er korrelerte, så vil man få skjeve estimater ved bruk av OLS. Dersom vi bruker boligpris og befolkningstetthet for henholdsvis y_i og x_i , så ser vi at det ikke er utenkelig at lokaliseringen av folk er en funksjon som avhenger av boligprisen i den kommunen. Det samme vil gjelde ved lønnsnivå som y_i , slik at bosettelsen vil avhenge av lønnsnivået i kommunen. Dette virker å være rimelig antagelser, og vi har derfor et simultanitetsproblem i denne analysen.

4.3 Håndtering av endogenitetsproblemet

4.3.1 Instrumentmetoden (IV)

Ovenfor er det nevnt ulike årsaker til at man kan få skjeve estimater ved bruk av OLS. En mulig løsning på et slikt problem er å ta i bruk instrumentmetoden (IV-metoden). Det er denne metoden som er den viktigste i denne analysen, og det er resultatene ved bruk av denne metoden som blir lagt til grunn når jeg til slutt skal gi en tolkning på elasticiteten til inntekt og boligpris med hensyn på befolkningstetthet.

Vi antar at eksogenitetsforutsetningen ikke holder og vi har skjeve og inkonsistente estimater, slik at $Cov(x, u) \neq 0$.

Før man kan ta i bruk IV-metoden må man finne en variabel som er korrelert med forklaringsvariabelen men som ikke er tatt med i strukturligningen

$$y_i = \alpha_0 + \alpha_1 x_i + u_i \quad (*)$$

Om det finnes en variabel som z_i som oppfyller forutsetningene

$$(i) Cov(u_i, z_i) = 0$$

$$(ii) Cov(x_i, z_i) \neq 0$$

vil variabelen z_i oppfylle kravene for å fungere som et instrument for x_i .

Den første forutsetningen krever at instrumentet ikke skal være korrelert med restleddet, altså en sann eksogen variabel. Dette er lett å forstå, ettersom det er akkurat det som var problemet med x_i . Den andre forutsetningen sier at instrumentet må være korrelert med x_i . Instrumentet z_i må altså kunne gi informasjon om x_i .

I denne oppgaven vil befolkningstetthet fra 1875 fungere som instrument for befolkningstetthet. Det er rimelig å anta at befolkningstettheten fra 1875 ikke er korrelert med de produktivitetsfaktorene som bidrar til skjevhet i estimatene ved bruk av OLS. Samtidig vil bosettelsen i 1875 gi informasjon om hvor folk bor i dag, og korrelasjonen er høy mellom disse variablene. Bruk av historisk data på befolkning har etter Ciccone og Hall resultater fra 1996 blitt anerkjent og godkjent som instrument. Combes, Duranton og Gobillon bruker også denne metoden i sin forskning, både av agglomerasjon (2010) og urbane kostnader (2012).

I denne oppgaven brukes det en variant av IV-metoden kalt 2SLS, eller to-steps OLS. Første steget består i å estimere sammenhengen mellom den endogene variabelen og instrumentet, og skal "rense ut" endogenitet i variabelen.

Man bruker vanlig OLS, men har befolkningstetthet som avhengig variabel og instrumentet som høyresidevariabel i regresjonen. Dette gir predikert verdi som skal erstatte de faktiske verdiene i den endogene forklaringsvariabelen i andre steg.

5. Data

Grunnlaget for denne analysen er i hovedsak data for boligpris, inntekt og befolkningstetthet i de største kommunene i Norge. Det er mange av de små kommunene som ikke har offentlig tilgjengelig tall på boligpriser og inntektsnivå for tidligere år, så jeg tar i bruk tall fra de 162 største kommunene i Norge. Mesteparten av datasettet er hentet fra SSB sin statistikkbank og NSD sin kommunedatabase. Analysen på agglomerasjon er gjort med gjennomsnittstall fra 2002-2008 mens analysen på urbane kostnader er gjort med gjennomsnittstall fra 2002-20011.

5.1 Avhengig variabler

5.1.1 Nominell inntektsnivå i norske kommuner

Den avhengige variabelen for agglomerasjonseffekter er samlet allmenn inntekt per kommune fordelt på antall innbyggere. Tallene er hentet fra SSB sin statistikkbank, konsumprisindeks tabell 03014. Jeg bruker den naturlige logaritmen til inntekt i analysen. For å ha best mulig sammenligningsgrunnlag for de ulike årene tar jeg høyde for inflasjon og måler inntekt per år i konstante 1998-priser. Inflasjonstallene er hentet fra SSB sin statistikkbank, konsumprisindeks tabell 03014. Gjennomsnittlig nominell lønn varierer stort mellom kommunene. Kommunen med høyest lønn er Bærum med kr 226528,39 mens kommunen med det laveste lønnsnivået er Eidskog med kr 119339,93.

5.1.2 Boligpriser i norske kommuner

Den avhengig variabelen i regresjonen for urbane kostnader er den naturlige logaritmen til gjennomsnittlig kvadratmeterpris for eneboliger. Tallene er fra SSB sin statistikkbank tabell 06035, hvor prisen er *per kvadratmeter primær rom*.

I likhet med tallene for lønn er boligprisene omregnet til konstante 1998-priser med inflasjonstall fra samme tabell fra SSB. Tallene viser en jevn stigning i boligprisene for hvert år. Gjennomsnittlig boligpris for årene 2002-2011 varierer fra kr 5786,64 i Våler, Hedmark til kr 25434,61 per kvadratmeter i Oslo.

Det er kommune i nærheten av hovedstaden som har et høyest snitt i boligpris, som Bærum, Asker, Nesodden og Frogn. Man kan også se at det er et jevnt over høyere snitt i store byer som Stavanger, Trondheim og Tromsø.

5.1.3 Reallønn i norske kommuner

Den siste avhengig variabelen i analysen er reallønnen, som er lønnen etter at man har tatt høyde for boligprisen. Den finner jeg ved å ta nominell inntekt minus boligpris for hver kommune.

5.2 Forklaringsvariabler

5.2.1 Befolkningstetthet

Tallene er hentet fra NSD sin kommunedatabase. Siden 1946 har hver kommune hatt sitt eget folkeregister, hvor det blir registrert antall fødte, døde og ut-og innflytninger til kommunene. Antall kommuner og hvor de grenser har endret seg over tid hvor mange små kommuner blir slått sammen til én. Derfor bruker jeg kommunene slik de var i 2002. Befolkningstettheten finner man ved å ta total befolkning i kommunen fordelt på areal i kvadratkilometer. Jeg har valgt totalt flateinnhold for kommunen som også inkluderer vannareal. Befolkningstettheten varierer stort mellom kommunene. Kommunene med høyest tetthet er Stavanger hvor det bor 1648,809 personer per kvadratkilometer. Kommunen med lavest tetthet er Forsand med 1,42 personer per kvadratkilometer.

5.2.2 Vintertemperatur

Tallene er hentet fra klimavakten på Meteorologisk institutt og er såkalte normaltemperaturer, altså gjennomsnittverdier, over en bestemt 30-års periode (her 1961-1990). Selv om det er store variasjoner i været, har det et hovedmønster som gjør at det kan lages en statistisk beskrivelse av det. Ettersom noen kommuner har flere målingsstasjoner, bruker jeg snittet av disse. Jeg bruker kun normaltemperatur for januar fordi det gir best inntrykk av vintertemperaturen samt det er ved vintertemperatur det størst variasjon mellom kommunene. Bakgrunnen for at jeg tar med temperatur i regresjonen er at det finnes sterke bevis på at vær er en viktig faktor ved lokaliseringvalg i Europa (Cheshire og Magrini, 2006)

5.2.3 Kyst

Jeg inkluderer en kystdummy som forteller om kommunen har en kystlinje eller ikke. Variabelen er 1 hvis kommunen har kystlinje, 0 ellers. Kystlinje tolkes her som en positiv tilleggseffekt ved å bo i kommunen, og kan tolkes som en fasilitet nevnt i kapittel 2. Det kan for eksempel påvirke produktiviteten ved transportkostnader knyttet til sjøfart og boligpris om tilgang til sjø oppfattes som en positiv gode. Tallene er registrert i 2003.

5.2.4 Utdanningsnivå

Tallene er fra SSB sin statistikkbank tabell 09429, og beskriver utdanningsnivået i prosent av befolkning. Dette er personer over 16 år med høyere utdanning og er delt inn i to grupper, kort og lang utdanning. Jeg samler disse to begrepene til ett mål for utdanningsnivå. Utdanningsnivå inkluderes i agglomerasjonsregresjonen fordi høyere utdanning påvirker lønnsnivået. Ettersom høyere utdanning er positivt korrelert med lønn, kontrolleres det også for utdanning i boligprisregresjonen. Dersom det er slik at velutdannede personer samler seg i kommunen med høy befolkningstetthet vil rike husholdninger utby andre og presse boligprisen opp. (Gyourko et al. 2010). Rosenthal og Strange (2008) finner at agglomerasjonseffektene er sterke for de med høyere utdanning. Det samme gjør Wheeler (2001) som konkluderer med at lønnsgevinsten øker med antall år med høyere utdanning. Kommunene Bærum, Asker og Oslo er de tre kommunene med høyest andel med høyere utdanning. Bærum hadde en andel på 44%. Tvedestrand har lavest med 2,71%.

5.2.5 Befolkningsvekst

Tallene er fra NSD sin database og er befolkningsveksten fra 1991 til 2001. Jeg har tatt gjennomsnittlig vekst i befolkning fra denne tiårsperioden. Vekstraten er relativt lik fra år til år ettersom befolkningsvekst har en høy seriekorrelasjon (Combes et al. 2012)

5.3 Instrumentvariabler

5.3.1 Befolkningstetthet 1875

Tallene er hentet fra NSD sin database. De er tatt fra folketellingen 31. desember 1875. For å være et gyldig instrument må befolkningstettheten i 1875 påvirke nåtidens befolkningsmønster. Bakgrunnen for produktivitetsfordelene i dag kan ikke være de som i 1875. Naturlige fortrinn ved produksjon som kan ha tiltrukket folk i 1875 burde ikke være representert. Det vil påvirke produksjonen og bryte med ekskluderingsrestriksjonen.

5.3.2 Befolkningsvekst 1825 - 1900

Som et alternativt instrument brukes veksten i befolkningstetthet fra 1825 til 1900. Dette instrumentet brukes som robustsjekk for analysen.

Tallene er hentet fra NSD sin database for folketellingene fra 1825-1900.

Tallene er samlet hjemmehørende folkemengde, som inkluderer midlertidig tilstedeværende utlendinger. Under denne tellingen ble folk registrert der de bodde fast, ikke der de oppholdt seg på tellingstidspunktet. NSD opplyser at folketellingene ikke nødvendigvis er helt fullstendige.

5.4 Deskriptiv statistikk

Variabler	Obs	Gj. Snitt	St.avvik	Min	Maks
Befolkningstetthet 2002	162	147.0047	156.3291	.0027204	1088.16
Befolkningstetthet 2003	162	148.2692	157.6738	.0027438	1097.52
Befolkningstetthet 2004	162	148.4097	157.8232	.0027464	1098.56
Befolkningstetthet 2005	162	149.0474	158.5013	.0027582	1103.28
Befolkningstetthet 2006	162	149.8039	159.3058	.0027722	1108.88
Befolkningstetthet 2007	162	150.4145	159.9552	.0027835	1113.4
Befolkningstetthet 2008	162	151.8033	161.4321	.0028092	1123.68
Befolkningstetthet 2009	162	153.4082	163.1388	.0028389	1135.56
Befolkningstetthet 2010	162	155.4995	165.3627	.0028776	1151.04
Befolkningstetthet 2011	162	157.8988	167.9142	.002922	1168.8
Befolkningstetthet1825	162	14.30381	14.222	.0002767	85.01408
Nominell lønnsnivå 2002	162	128992.2	16537.24	105732.9	196796.4
Nominell lønnsnivå 2003	162	134613.7	16673.05	106157.3	207468.8
Nominell lønnsnivå 2004	162	142121.8	18830.17	111997.4	223703.4
Nominell lønnsnivå 2005	162	154116.5	28621.52	120889.2	377303
Nominell lønnsnivå 2006	162	146101.4	18825.38	117989.3	235214.5
Nominell lønnsnivå 2007	162	158364.2	20591.39	127130.6	246383.9
Nominell lønnsnivå 2008	162	164803.4	19857.39	133064.8	248574.5
Boligpris 2002	162	8749.896	2551.264	4348.257	18635.26
Boligpris 2003	162	8824.387	2531.259	4818.476	18174.87
Boligpris 2004	162	9713.98	2790.99	4839.666	19688.31
Boligpris 2005	162	10463.88	3057.845	5402.958	22863.29
Boligpris 2006	162	11662.54	3551.743	5568.578	25140.02
Boligpris 2007	162	13182.48	4117.323	6362.417	27754.34
Boligpris 2008	162	13285.42	4117.849	6012.822	28139.08
Boligpris 2009	162	13457.24	4061.241	5421.516	28814.96
Boligpris 2010	162	14290.14	4446.533	6499.823	31094.97
Boligpris 2011	162	15085.75	4779.786	6763.977	34041.05
Utdanningsnivå 2002	162	15.55123	3.85909	8.9	28.7
Utdanningsnivå 2003	162	15.98519	3.830674	9.2	28.8
Utdanningsnivå 2004	162	16.47469	3.864945	9.1	29.2
Utdanningsnivå 2005	162	16.96049	3.881896	9.5	29.6
Utdanningsnivå 2006	162	17.30123	3.891941	10	29.8
Utdanningsnivå 2007	162	17.61358	3.891071	10.4	29.9
Utdanningsnivå 2008	162	17.95123	3.901286	10.8	30
Januartemperatur	162	-4.146296	3.527446	-11.5	2.1

6. Empirisk analyse

6.1 Agglomerasjon

6.1.1 Effekten av befolkningstetthet på inntektsnivå, OLS

Tabell 1: Modell for agglomerasjon med OLS for 2002-2008

	(1) OLS Ln lønn	(2) OLS Ln lønn	(3) OLS Ln lønn	(4) OLS Ln lønn	(5) OLS Ln lønn
Ln tetthet	0.0619*** (0.00504)	0.0321*** (0.00552)	0.0322*** (0.00590)	0.0353*** (0.00610)	0.0349*** (0.00612)
Utdanning		0.0103*** (0.00123)	0.0103*** (0.00124)	0.0100*** (0.00126)	0.00980*** (0.00128)
Kyst			-0.000753 (0.0128)		0.0156 (0.0166)
Vintertemperatur				-0.00222 (0.00180)	-0.00362 (0.00233)
Konstant	11.65*** (0.0205)	11.56*** (0.0206)	11.56*** (0.0209)	11.54*** (0.0241)	11.54*** (0.0250)
Observasjoner	162	162	162	162	162
R ²	0.485	0.643	0.643	0.646	0.648

Standardavvik oppgitt i parentes.. *** Signifikant på 1%nivå ** Signifikant på 5% nivå, * Signifikant på 10% nivå

Tabell 1 viser resultatene ved bruk av OLS på regresjonen for agglomerasjon (i)' fra kapittel 4.2.1, hvor logaritmen til nominell lønnsnivå er den avhengige variabelen.

Kolonne (1) viser elastisiteten til gjennomsnittlig lønn med hensyn på befolkningstetthet når ingenting annet blir kontrollert for. Modellen viser en koeffisient på 0,0619 og befolkningstettheten er signifikant på ettprosentnivå.

Ved en dobling av befolkningstettheten vil nominell lønn øke med 6,19%.

Koeffisienten er positiv, og indikerer at det er høyere lønn i områder med høyere befolkningstetthet. Dette høres riktig ut om man ser på hvilke byer som har det høyeste lønnsnivået i Norge. Resultatet er i tråd med Combes et al. (2010) som finner en elastisitet på 4,8% ved bruk av fransk data. Dette er også nære funnene til Carlsen et al. (2012) som fant en elastisitet på 7%. Forskjellen fra min analyse og den gjort av Carlsen et al. (2012) er at jeg bruker kommuner mens deres arbeid er gjort på arbeidsmarkedsregioner, slik at resultatene ikke vil være helt like men fremdeles delvis sammenlignbart.

Arbeidsmarkedsregioner er en sammenslåing av flere kommuner ut ifra kriterier til arbeidsmarked og andre økonomiske forhold, og tar høyde for pendling på tvers av kommunene. Samtidig bruker Carlsen et al. (2012) befolkningsstørrelse istedenfor befolkningstetthet. Ved bruk av befolkningstetthet finner de en elastisitet på 6,2%. Nederst i kolonne (1) ser man at modellen har en forklaringskraft på 0,485 som vil si at befolkningstetthet forklarer 48,% av variasjonen i lønnsnivået.

I kolonne (2) kontrolleres det for utdanning som har en koeffisient på 0,0103 og er signifikant på ettprosentnivå. Utdanningsvariabelen er en prosentstørrelse og er derfor en semi-elastisitet. Koeffisienten er positiv og betyr at jo høyere utdanningsnivået i en kommune er jo høyere vil gjennomsnittlig nominell lønn være. Dersom utdanningsnivået til personer over 16 år øker med ett prosent, vil nominell lønn øke med 1,03%. Det faktum at utdanning har en positiv påvirkning på lønn er ikke spesielt overraskende, ettersom utdanning gjør en person mer ettertraktet på arbeidsmarkedet og sannsynligheten for høyere lønn er større enn ved ingen utdanning. Det er også ofte slik at utsiktene for potensielt høyere lønn er drivkreftene bak nettopp valg av høyere utdanning. Mitt resultat er at en fordobling av befolkningstettheten vil gi en økning i gjennomsnittlig lønnsnivå på 3,21%. Inkludering av utdanningsnivå i regresjonen reduserer altså elastisiteten til nominell lønn med hensyn på befolkningstetthet med nesten 3 prosent. Årsaken til at elastisiteten til befolkningstetthet reduseres når man kontrollerer for utdanning er at det finnes en betydelig sorteringseffekt. Det kan være at velutdannede mennesker vil foretrekke tettbebygde områder, eller at produktivetsfordelene ved å bo i tettbebygde områder er høyere for folk med høy utdanning (Combes et al. 2010). Befolkningstetthet og utdanning forklarer 64,3% av variasjonen i gjennomsnittlig nominell lønn. Man kan sammenligne resultatene i kolonne (2) med Carlsen et al. (2012) som finner en elastisitet på 6,5% etter at de har kontrollert for arbeiderkarakteristikker som inkluderer utdanning. I kolonne (3) og (4) kontrolleres det for henholdsvis kyst og vintertemperatur, uten at disse er signifikant. Kolonne (5) inkluderer alle forklaringsvariablene, og vi ser at koeffisienten til befolkningstetthet og utdanning endrer seg svært lite. Det betyr at mine estimater på befolkningstetthet og utdanning er stabile. Ettersom det er en risiko for at effekten mellom lønnsnivå og befolkningstetthet går begge veier, altså lønnsnivået også påvirker befolkningstettheten, må man bruke instrumentmetoden forklart i kapittel 4 i håp om å få fjernet dette endogenitesproblemet.

6.1.2 Instrumentmetoden første steg

Tabell 2: Modell for agglomerasjon med IV for 2002-2008, første steg

	(1) IV Ln tetthet	(2) IV Ln tetthet	(3) IV Ln tetthet	(4) IV Ln tetthet	(5) IV Ln tetthet
Ln tetthet 1875	1.1238*** (0.0578)	0.9257*** (0.0492)	0.8978*** (0.0519)	0.8650*** (0.0509)	0.869*** (0.0514)
Utdanning		0.0834*** (0.0082)	0.0810*** (0.008)	0.0840*** (0.008)	0.0854*** (0.0082)
Kyst			0.1638 (0.101)		-0.088 (0.1294)
Vintertemperatur				0.045*** (0.0134)	0.0525*** (0.0176)
Konstant	0.6683*** (0.1727)	-0.478** (0.1757)	-0.424** (0.0212)	-0.131 (0.1999)	-0.1004 (0.2049)
Observasjoner	162	162	162	162	162
R ²	0.702	0.819	0.822	0.832	0.832
F-verdi	377.05	361.05	244.03	260.20	194.60

Standardavvik oppgitt i parentes.. *** Signifikant på 1% nivå ** Signifikant på 5% nivå, * Signifikant på 10% nivå

Tabell 2 viser resultatene ved bruk av førstestegs instrumentmetode (IV) for agglomerasjonseffekter. Instrumentet som brukes er befolkningstettheten fra 1875, og variabelen antas å si noe om urbaniseringen uten at den blir påvirket av hverken lønn eller produktivitet. Årsaken til at det første steget i IV-metoden er viktig er fordi det sier noe om i hvilken grad instrumentet er relevant. Dersom instrumentet ikke er signifikant i førstesteget har man ikke identifisert modellen og man vil lære lite av IV-resultatene. I førstesteget IV blir endogeniteten fra forklaringsvariabelen befolkningstetthet "renset ut" og man finner predikert verdi som skal brukes i andreteget. Tabell 2 viser at instrumentet er sterkt og relevant, og korrelasjonen mellom befolkningstetthet i dag og tettheten som var i 1875 er sterk. Vi ser fra kolonne (1) at koeffisienten til tettheten fra 1875 er på 1,1238 og er signifikant på ettprosentnivå. Befolkningstettheten i 1875 forklarer 70,2% av variasjonen i befolkningstetthet i dag. Dette er sammenlignbart med Carlsen et al. (2012) som finner en koeffisient på 1,08 ved bruk av befolkningsstørrelse fra 1875 som instrument. Koeffisienten varierer fra 1,238 til 0,8650 avhengig av hvor mange variabler man kontrollerer for. F-verdien på 377,05 er tilstrekkelig høy og er godt

over den kritiske verdien til Stock et al. (2002). Dette betyr at instrumentet er veldig sterkt og relevant.

6.1.3 Instrumentmetoden andre steg, 2SLS

Tabell 3: Modell for agglomerasjon med IV for 2002-2008, andre steg

	(1) 2SLS Ln lønn	(2) 2SLS Ln lønn	(3) 2SLS Ln lønn	(4) 2SLS Ln lønn	(5) 2SLS Ln lønn
Ln tetthet	0.0478*** (0.00612)	0.0234*** (0.00664)	0.0221*** (0.00727)	0.0246*** (0.00757)	0.0235*** (0.00758)
Utdanning		0.0116*** (0.00134)	0.0116*** (0.00135)	0.0115*** (0.00139)	0.0113*** (0.00141)
Kyst			0.00675 (0.0132)		0.0179 (0.0165)
Vintertemperatur				-0.000870 (0.00188)	-0.00240 (0.00237)
Konstant	11.71*** (0.0244)	11.56*** (0.0208)	11.57*** (0.0212)	11.56*** (0.0251)	11.55*** (0.0258)
Observasjoner	162	162	162	162	162
R ²	0.460	0.637	0.636	0.639	0.641

Standardavvik i parentes.. *** Signifikant på 1%nivå ** Signifikant på 5% nivå, * Signifikant på 10% nivå

I tabell 3 har den endogene forklaringsvariabelen befolkningstetthet blitt erstattet med predikert verdi fra førstesteget i tabell 2. Denne metoden kalles to-steps OLS (2SLS). Man ser fra tabell 3 at ved å fjerne endogeniteten i variabelen for befolkningstetthet så får man en lavere koeffisient. Kolonne (1) er en regresjon med kun befolkningstetthet som forklaringsvariabel. Man ser at koeffisienten til befolkningstetthet har blitt redusert fra 0,0619 før instrumentering til 0,0478 etter instrumentering.

Det betyr at man har lyktes med å fjerne noe av endogenitetsproblemet knyttet til variabelen. Koeffisienten er signifikant på ettprosentnivå, og er sammenlignbar med Combes et al. (2010) som finner en elastisitet på 4% når de instrumenterer for befolkningstetthet i 1831 ved bruk av fransk data. Carlsen et al. (2012) finner også en reduksjon i elastisiteten etter instrumentering med en elastisitet på 6,6%.

I kolonne (2) kontrolleres det for utdanningsnivå, og var min foretrukne modellspesifisering ved OLS i tabell 1. Kontroll for utdanningsnivå reduserer koeffisienten til befolkningstetthet til 0,0234, altså over halvparten. Koeffisienten til

utdanningsnivå øker fra 0,0103 ved OLS til 0,0116 ved IV. Befolkningstetthet og utdanningsnivå forklarer 63,7% av variasjonen i inntektsnivå.

Sammenlignet med mine resultater fra OLS blir koeffisienten til befolkningstetthet når det kontrolleres for utdanningsnivå redusert fra 3,21% til 2,34% etter at det instrumenteres for befolkningstetthet. Carlsen et al. (2012) finner en reduksjon i elastisiteten fra 6,5% ved bruk av OLS til 6% ved bruk av IV når de kontrollerer for utdanning. Dette betyr at man har skjevhet i OLS-estimatoren. Skjevheten skyldes antagelig både mangel på relevant forklaringsvariabel og simultanitet. Reduksjonen i elastisiteten kan forklares av simultanitet ved at noe av den positive korrelasjonen mellom befolkningstetthet og lønn er på grunn av at folk velger å flytte til kommuner med høyere lønn. Da går effekten fra lønn til befolkningstetthet. Denne blir renset bort i tabell 3 og gir kun effekten fra befolkningstetthet til lønn, slik at elastisiteten derfor blir lavere. Det kan være at de som bor i de største kommunene har bedre ferdigheter, slik at de tjener mer enn andre med samme utdanningsnivå.

Dette er rimelig å anta, og skulle helst ha vært kontrollert for.

Carlsen et al. (2012) kontrollerer for slik uobserverbar karakteristik og finner en elastisitet på 3,4%. Selv om de bruker befolkningsstørrelse som mål på urbanisering og ikke befolkningstetthet, så viser funnene deres at slik skjevhet finnes og vil sannsynligvis gi opphav for skjevhet i mine estimater.

Videre i kolonne (3) og (4) kontrolleres det for henholdsvis kyst og vintertemperatur. Vi ser at elastisiteten til befolkningstetthet ligger stabilt på rundt 2,3%.

I kolonne (5) inkluderes alle forklaringsvariablene. Selv om det kun er befolkningstetthet og utdanning som er signifikant er det verdt å merke seg at koeffisienten til befolkningstetthet er stabil og er tilnærmet lik som i kolonne (2). Selv om man inkluderer disse variablene er det tilnærmet ingen endring i forklaringskraften til modellen

6.2 Urbane kostnader

6.2.1 Effekten av befolkningstetthet på boligpris, OLS

Tabell 4: Modell for urbane kostnader med OLS, 2002-2011

	(1) OLS Ln boligpris	(2) OLS Ln boligpris	(3) OLS Ln boligpris	(4) OLS Ln boligpris	(5) OLS Ln boligpris
Ln tetthet	0.154*** (0.0115)	0.0985*** (0.0122)	0.0912*** (0.0131)	0.0597*** (0.0124)	0.0608*** (0.0133)
Vekst		0.0193*** (0.00248)		0.0156*** (0.00229)	0.0152*** (0.00237)
Utdanning			0.0431*** (0.00580)	0.0340*** (0.00530)	0.0354*** (0.00561)
Kyst					-0.0356 (0.0349)
Vintertemperatur					0.00356 (0.00503)
Konstant	8.751*** (0.0467)	8.835*** (0.0413)	8.546*** (0.0489)	8.657*** (0.0462)	8.672*** (0.0534)
Observasjoner	162	162	162	162	162
R ²	0.528	0.658	0.650	0.729	0.731

Standardavvik i parentes.. *** Signifikant på 1% nivå ** Signifikant på 5% nivå, * Signifikant på 10% nivå

Tabell 4 viser resultatene ved bruk av OLS på regresjonen for urbane kostnader vist ved (ii)' i kapittel 4.2.1, med logaritmen til boligpris som avhengig variabel.

I kolonne (1) er det kun én forklaringsvariabel, logaritmen til befolkningstetthet.

Koeffisienten til befolkningstetthet er 0,154 og er signifikant på ettprosentnivå.

Det betyr at en fordobling av befolkningstettheten vil føre til en økning i boligprisen på 15,4%. Dette virker noe høyt og kan tyde på skjevhet i estimatene.

Forklaringskraften til denne modellen er 52,8%, som betyr at befolkningstetthet forklarer litt over halvparten av variasjonen i boligprisen. I kolonne (2) kontrolleres det for total befolkningsvekst fra 1991-2001. Befolkningsveksten er oppgitt som en vekstrate i prosent, og er derfor en semi-elastisitet. Befolkningsvekst er seriekorrelert, og det er rimelig å anta at befolkningsveksten er tilnærmet lik fra år til år.

Koeffisienten til befolkningsvekst er 0,0193 og er signifikant på ettprosentnivå.

Det betyr at hvis årlig befolkningsvekst øker med 0,1 prosentpoeng vil boligprisen øke med 1,93%. Vi ser at koeffisienten til befolkningstetthet blir redusert til 0,0985 når man kontrollerer for befolkningsvekst.

Begge forklaringsvariablene er signifikant på ettprosentnivå. I kolonne (3) kontrolleres det for utdanningsnivå, men ikke for befolkningsvekst. Utdanningsvariabelen er den samme som ble brukt i modellen for agglomerasjon. Argumentet for å inkludere utdanningsnivå i regresjonen er en sortering av velutdannede personer i de store kommunene. Høy utdanning er sterkt korrelert med høyere lønn. Ifølge Gyourko (2010) kan rike husholdninger utby andre i de store kommune slik at boligprisen presses opp. Vi ser at koeffisienten til befolkningstetthet endrer seg nesten like mye som ved kontroll for befolkningsvekst, med en koeffisient på 0,0912. Utdanning har en koeffisient på 0,0431 og er signifikant på ettprosentnivå. Det betyr at hvis utdanningsnivået for folk over 16 år øker med ett prosentpoeng, vil boligprisen øke med 4,31%. I kolonne (4) kontrolleres det både for befolkningsvekst og utdanningsnivå. Koeffisient til befolkningstetthet er på 0,0597. Det vil si at en fordobling av tettheten gir en økning i boligpris på 5,97%. En 0,1-prosentpoengs økning i befolkningsvekst gir en økning i boligprisen på 1,56%, mens den økning i utdanningsnivå på ett prosentpoeng gir en økning på 3,4%. Begge variablene har en positiv virkning på boligprisen. Befolkningstetthet, befolkningsvekst og utdanningsnivå forklarer 72,9% av variasjonen i boligprisen. I kolonne (5) kontrolleres det i tillegg for kystlinje og vintertemperatur. Dette påvirker koeffisienten til befolkningstetthet i svært liten grad, hvor den endres fra 0,059 til 0,0608. Påvirkningen på koeffisientene til befolkningsvekst og utdanning er også svært lav. Variablene kyst og vintertemperatur er heller ikke signifikant. Dette gjør at modellen i kolonne (4) er min fortrukne modell

6.2.2 Instrumentmetoden første steg

Tabell 5: modell for urbane kostnader med IV for 2002-2011, første steg

	(1) IV Ln tetthet	(2) IV Ln tetthet	(3) IV Ln tetthet	(4) IV Ln tetthet	(5) IV Ln tetthet
Ln tetthet 1875	1.1238*** (0.0578)	0.9691*** (0.049)	0.925*** (0.4920)	0.8828*** (0.445)	0.857*** (0.047)
Befolkningsvekst		0.715*** (0.0074)		0.0471*** (0.007)	0.0427*** (0.007)
Utdanning			0.0834*** (0.008)	0.0585*** (0.008)	0.0615*** (0.087)
Kyst					-0.0217 (0.1193)
Vintertemperatur					0.0248 (0.017)
Konstant	0.6683*** (0.1727)	0.6258*** (0.1379)	-0.477*** (0.1757)	-0.1643*** (0.1643)	-0.0106 (0.1888)
Observasjoner	162	162	162	162	162
R ²	0.702	0.811	0.820	0.856	0.859
F-verdi	377	342.26	361.05	314	190.88

Standardavvik i parentes.. *** Signifikant på 1%nivå ** Signifikant på 5% nivå, * Signifikant på 10% nivå

Tabell 5 viser resultatene fra førstestegs IV-metode.

Prosessen er lik som i delkapittelet for agglomerasjon, hvor hovedmålet er å finne predikert verdi av befolkningstetthet, for så å erstatte det med den endogene variabelen for befolkningstetthet i andresteg senere. Tabell 5 viser at instrumentet er både sterkt og relevant med en F-verdi høyere enn kritisk verdi. Den predikerte verdien av befolkningstetthet funnet i tabell 5 brukes nå som forklaringsvariabel i tabell 6.

6.2.2 Instrumentmetoden andre steg, 2SLS

Tabell 6: Modell for urbane kostnader med IV for 2002-2011, andre steg

	(1) 2SLS Ln boligpris	(2) 2SLS Ln boligpris	(3) 2SLS Ln boligpris	(4) 2SLS Ln boligpris	(5) 2SLS Ln boligpris
Ln tetthet	0.116*** (0.0141)	0.0726*** (0.0145)	0.0606*** (0.0158)	0.0442*** (0.0146)	0.0433*** (0.0160)
befolkningsvekst		0.0224*** (0.00266)		0.0167*** (0.00234)	0.0161*** (0.00238)
Utdanning			0.0261*** (0.00321)	0.0187*** (0.00277)	0.0196*** (0.00294)
Kyst					-0.0311 (0.0346)
Vintertemperatur					0.00498 (0.00502)
Konstant	8.898*** (0.0563)	8.914*** (0.0477)	8.577*** (0.0498)	8.680*** (0.0467)	8.703*** (0.0545)
Observasjoner	162	162	162	162	162
R ²	0.495	0.648	0.637	0.726	0.727

Standardavvik i parentes.. *** Signifikant på 1% nivå ** Signifikant på 5% nivå, * Signifikant på 10% nivå

Tabell 6 viser resultatene ved bruk av 2SLS-metoden.

I likhet med delkapittelet om agglomerasjon har man nå erstattet den endogene forklaringsvariabelen med predikert verdi fra førstesteget i tabell 5 ovenfor .

Som antatt ser vi fra kolonne (1) at koeffisienten til befolkningstetthet nå er lavere enn ved bruk av OLS. Koeffisienten har blitt redusert fra 0,154 med OLS til 0,116 med 2SLS. Det betyr nå at en fordobling av befolkningstettheten gir en økning i boligprisen på 11,6%. Videre i kolonne (2) og (3) kontrolleres det for henholdsvis befolkningsvekst og utdanning slik som ved OLS metoden. I kolonne (4), som var min foretrukne modellspesifikasjon fra OLS-estimeringen, kontrolleres det både for befolkningsvekst og utdanning. Koeffisienten til befolkningstetthet som før var 0,0597 ved OLS i tabell 4 kolonne (4) er nå på 0,0442. En reduksjon på over halvannen prosent tyder på at det fantes et endogenitetsproblem som ga opphav til skjevhet i OLS-estimatene. Det er slik at folk velger å flytte til kommuner på tross av høy boligpris. Årsaken kan være andre positive effekter i denne kommunen, som for eksempel gode offentlige tjenester. Resultatet blir en elastisitet på boligpris med hensyn på befolkningstetthet på 4,42%. Dette er veldig nære hva Combes et al. (2012) finner ved bruk av fransk data, hvor de ender opp med en elastisitet på boligpris med

hensyn på befolkningstetthet på 4,1%. Grunnen for at jeg får noe høyere elastisitet kan være på grunn av Combes et al. (2012) har tilgang på et langt større og mer komplisert datasett enn hva som inngår i denne analysen. I kolonne (5) inkluderes kyst og vintertemperatur, og vi ser at dette har liten påvirkning på estimatene ved min fortrukne modell i kolonne (4) og estimatet for befolkningstetthet er stabil.

6.3 Reallønn

6.3.1 Effekten av befolkningstetthet på reallønn, OLS

Tabell 7: Modell for reallønn med OLS

	(1) OLS Ln reallønn	(2) OLS Ln reallønn	(3) OLS Ln reallønn	(4) OLS Ln reallønn
Ln tetthet	-0.0921*** (0.00969)	-0.0504*** (0.0106)	-0.0596*** (0.0121)	-0.0335*** (0.0119)
Befolkningsvekst		-0.0145*** (0.00217)		-0.0128*** (0.00219)
Utdanning			-0.0113*** (0.00271)	-0.00750*** (0.00254)
Konstant	2.903*** (0.0393)	2.840*** (0.0361)	3.008*** (0.0452)	2.917*** (0.0439)
Observasjoner	162	162	162	162
R ²	0.361	0.501	0.424	0.527

Standardavvik i parentes.. *** Signifikant på 1% nivå ** Signifikant på 5% nivå, * Signifikant på 10% nivå

Tabell 7 viser resultatene ved bruk av OLS når man har reallønnen som den avhengige variabelen og er en estimering av ligning (iii)' fra kapittel 4.2.1 Reallønnen er funnet ved å ta logaritmen til gjennomsnittlig lønn minus logaritmen til gjennomsnittlig boligpris. Kolonne (1) viser elastisiteten til reallønnen når kun befolkningstetthet inkluderes. Vi ser at koeffisienten er negativ og signifikant på ett prosentnivå. Dette skyldes at elastisiteten til boligpris med hensyn på befolkningstetthet er høyere enn elastisiteten til lønnen. En fordobling av befolkningstettheten vil gi en reduksjon i reallønnen på 9,21%. I kolonne (2) og (3) kontrolleres det for henholdsvis befolkningsvekst og utdanningsnivå som nærmest halverer elastisiteten.

I kolonne (4) kontrolleres det for både vekst og utdanning og er min foretrukne modell. Koeffisienten til befolkningstetthet er $-0,0335$, som betyr at en dobling av befolkningstettheten reduserer reallønnen med $-3,35\%$. Resultatet sier og at en økning i befolkningsvekst og utdanningsnivå vil ha en negativ påvirkning på reallønnen. Dersom befolkningsveksten øker med $0,1$ prosentpoeng vil reallønnen reduseres med $-1,28\%$. Ved en økning i utdanningsnivå med ett prosentpoeng vil reallønnen reduseres med $-0,75\%$. Bakgrunnen for reduksjonen i reallønnen er fordi boligprisen er mer elastisk enn lønnen, noe vi ser fra tabell 1 og 4 ved OLS og tabell 3 og 6 ved IV. Boligprisen reagere raskere enn nominell lønn. I Norge er lønnsdannelser langsomme og kompliserte, og de vil ikke reagere like fort og med samme størrelse som boligprisen. Etersom det finnes knapphet av antall boliger og jord man kan bygge hus på, vil prisen på bolig reagere med et større omfang hvis befolkningen øker, og derav økning i etterspørsel etter bolig.

6.3.2 Instrumentmetoden andre steg

Tabell 8: Modell for reallønn med IV, andre steg

	(1) 2SLS Ln reallønn	(2) 2SLS Ln reallønn	(3) 2SLS Ln reallønn	(4) 2SLS Ln reallønn
Ln tetthet	-0.0678*** (0.0117)	-0.0369*** (0.0125)	-0.0372** (0.0146)	-0.0240* (0.0139)
Befolkningsvekst		-0.0161*** (0.00231)		-0.0135*** (0.00223)
Utdanning			-0.0145*** (0.00295)	-0.00848*** (0.00263)
Konstant	2.810*** (0.0468)	2.798*** (0.0413)	2.988*** (0.0458)	2.905*** (0.0444)
Observasjoner	162	162	162	162
R ²	0.336	0.495	0.411	0.525

Standardavvik i parentes.. *** Signifikant på 1% nivå ** Signifikant på 5% nivå, * Signifikant på 10% nivå

Tabell 8 viser andresteg IV for reallønn. Vi ser fra kolonne (1) tabell 8 at koeffisienten til befolkningstetthet øker til minus 6,78%, hvor den ved OLS var på minus 9,21%. Dette er ikke overraskende når vi vet at elastisiteten til boligpris med hensyn på tetthet falt med cirka 3,8% etter instrumentering, og elastisiteten til inntekt med hensyn på befolkningstetthet falt med cirka 1,40% etter instrumenteringen. Kolonne (2) kontrolleres det for befolkningsvekst og i kolonne (3) for utdanning. I kolonne (4) kontrolleres det for begge. Vi ser fra kolonne (4) at koeffisienten til befolkningstetthet er på -0,0240, som betyr at en fordobling av tettheten vil redusere reallønnen med 2,40%.

Dette er ikke overraskende, og bakgrunnen for det er de samme som ved OLS. Befolkningstettheten er signifikant på ti prosentnivå. Befolkningsvekst og utdanning er begge signifikant på ett prosentnivå. Forklaringskraften til modellen er så og si den samme med 52,5%.

6.4 Robusthetstest

Jeg bekrefter mine funn ved å utføre noen robusthetstester. Jeg estimerer en modell hvor jeg instrumenterer for endringen i befolkningstetthet fra 1825-1900 istedenfor befolkningstettheten i 1875 som gjort tidligere. Deretter skal jeg vise resultatene fra år 2002, 2004 og 2007 både ved bruk av 2SLS for å se om resultatene er avhengig av konjunktursituasjonen i landet.

6.4.1 Alternativt instrument: Endring i befolkningstetthet 1825-1900

Tabell 9: Modell for reallønn med IV andre steg, alternativt instrument

	(1) 2SLS Ln reallønn	(2) 2SLS Ln reallønn	(3) 2SLS Ln reallønn	(4) 2SLS Ln reallønn
Ln tetthet	-0.0763*** (0.0121)	-0.0489*** (0.0131)	-0.0390** (0.0164)	-0.0321** (0.0153)
Befolkningsvekst		-0.0146*** (0.00234)		-0.0129*** (0.00227)
Utdanning			-0.0142*** (0.00314)	-0.00764*** (0.00271)
Konstant	2.842*** (0.0483)	2.835*** (0.0428)	2.990*** (0.0462)	2.915*** (0.0451)
Observasjoner	162	162	162	162
R ²	0.350	0.501	0.413	0.527

Standardavvik i parentes.. *** Signifikant på 1%nivå ** Signifikant på 5% nivå, * Signifikant på 10% nivå

Resultatene fra første steget er vist i appendix , og viser en koeffisient på 0,703 ved kun inkludering av endring i befolkningstetthet fra 1825-1900. Korrelasjonen mellom befolkningstetthet i dag og endringen fra 1825-1900 er høy, og testen har en F-verdi på over 200, uavhengig av hvor mange av forklaringsvariablene som inkluderes.

Instrumentet er derfor gyldig og relevant. Vi ser fra tabell 9 kolonne (1) at koeffisienten til befolkningstetthet etter instrumentering er høyere enn når man bruker befolkningstetthet i 1875 som instrument. Koeffisienten er redusert fra -0,0678 i tabell 8 til -0,0763. Resultatene viser at endringen i tetthet fra 1825-1900 ikke fjerner like mye av endogeniteten til variabelen befolkningstetthet.

Resultatene fra robusthetstesten bekrefter at mine funn fra tabell 8, hvor jeg finner at økt befolkningstetthet gir en reduksjon i reallønnen.

6.4.2 Effekten av befolkningstetthet på reallønn for enkeltår, 2SLS

Tabell 10: Modell for reallønn per år med 2SLS

	(1) 2002 2SLS ln Reallønn	(2) 2004 2SLS ln Reallønn	(3) 2007 2SLS ln Reallønn
ln tetthet	-0.0247* (0.0150)	-0.0211 (0.0126)	-0.0409** (0.0179)
befolkningsvekst	-0,0990*** (0.0236)	-0.141*** (0.0248)	-0.142*** (0.0283)
Utdanning	-0.0109*** (0.00270)	-0.00635** (0.00287)	-0.00743** (0.00307)
Konstant	3.085*** (0.0418)	3.012*** (0.0478)	2.933*** (0.0530)
Observasjoner	162	162	162
R ²	0.483	0.442	0.488

Standardavvik i parentes.. *** Signifikant på 1%nivå ** Signifikant på 5% nivå, * Signifikant på 10% nivå

Tabell 10 viser resultatene ved bruk av 2SLS for enkeltår, hvor jeg valgte ut 2002, 2004 og 2007. Tabellen viser kun resultatene fra den foretrukne modellspesifikasjonen lik kolonne (4) i tabell 8, hvor det kontrolleres for befolkningstetthet, befolkningsvekst og utdanningsnivå. Modellene i sin helhet er vist i appendix. Koeffisienten til befolkningstetthet uten kontrollvariabler for de tre årene - 0.0630 til -0.0868. Dette stemmer bra overens med resultatene fra IV-regresjonen med gjennomsnittsverdier fra tabell 8. Når man kontrollerer for befolkningsvekst og utdanning faller koeffisienten til befolkningstetthet betraktelig for hvert av årene. Vi ser fra tabell 10 at koeffisienten til befolkningstetthet hvor hvert år ligger nærme snittverdiene fra tabell 8, hvor den i tabell 8 var på -0,0240. Den høyere elastisiteten i 2007 kan skyldes de høye boligprisene i forkant av finanskrisen i 2008. Resultatene viser at elastisiteten til reallønn med hensyn på befolkningstetthet varierer over tid.

6.5 Oppsummering

Jeg har brukt data for gjennomsnittlig nominelt inntektsnivå og boligpris for 162 norske kommuner for å undersøke om disse påvirkes av økt befolkningstetthet. Deretter har jeg sett på hvordan befolkningstettheten påvirker reallønnen. Jeg estimerer modellene utformet i kapittel 4 ved bruk av OLS og finner at denne metoden overestimerer effektene av befolkningstetthet. Skjevheten skyldes problemer knyttet til utelatt relevant forklaringsvariabler og simultanitet. Jeg tar i bruk historisk befolkningstetthet som instrument i en to-steps OLS (2SLS) for å løse endogenitetsproblemet. Instrumentet som brukes er befolkningstetthet fra 1875 og det fremstår som både gyldig og relevant. Jeg finner at elastisiteten til nominelt inntektsnivå med hensyn på befolkningstetthet er 2,34% ved bruk av 2SLS når det kontrolleres for utdanningsnivå. Resultatene fra tabell 6 viser at elastisiteten til boligpris med hensyn på befolkningstetthet er 4,42% ved 2SLS når det kontrolleres for befolkningsvekst og utdanningsnivå. Ettersom elastisiteten til boligprisen er høyere enn nominell lønn, viser regresjonene i tabell 7 og 8 en negativ sammenheng mellom reallønn og befolkningstetthet. Etter instrumentering viser tabell 8 at reallønn har en elastisitet med hensyn på befolkningstetthet på -2,40% når det kontrolleres for utdanning og befolkningsvekst. Jeg utfører noen robusthetstester for å undersøke mine resultatene. Ved å bruke et alternativ instrument som endring i befolkningstetthet fra 1825 til 1900 og deretter vise regresjoner for tre ulike år, finner jeg tilfredsstillende resultater som støtter opp under mine estimeringer av reallønn og viser at resultatene er solide.

7 Konklusjon

I denne oppgaven har jeg utført en empirisk analyse av Enrico Moretti (2011) sin teoretiske modell som sier at dersom det finnes agglomerasjonseffekter vil positiv befolkningsvekst føre til høyere nominelt lønnsnivå og høyere boligpriser. Effekten på reallønnen er derimot uklar. Jeg har testet hans modell ved bruk av tverrsnittsdata for norske kommuner for å finne elasticiteten til lønn og boligpriser med hensyn på befolkningstetthet. I tråd med tidligere forskning gjort av blant andre Ciccone og Hall (1996), Combes et al (2010) og Carlsen et al (2012) undersøker jeg om lønnsforskjellene mellom kommunene i Norge kan forklares av befolkningstetthet. Med bakgrunn i Combes et al (2012) sin forskning av urbane kostnader undersøker jeg også om høyere boligpriser er forårsaket av høyere befolkningstetthet. I analysen støter jeg på en rekke økonometriske utfordringer. For å komme frem til riktige og troverdige resultater er det viktig å indentifisere kausaleffekten. Det må ikke være tilfelle at høyere befolkningstetthet er et resultat av høyere inntektsnivå eller lavere boligpris. Høyere befolkningstetthet skal påvirke inntekt og boligpris men ikke omvendt. Mine resultater tyder på en betydelig skjevhet i estimatene ved bruk av OLS, og effektene av befolkningstetthet blir overestimert ved bruk av denne metoden. Ved å kontrollere for utdanningsnivå finner jeg en elasticitet på 3,21% for lønn. Dette tyder på at det finnes en sortering av velutdannede mennesker i de store kommunene. Jeg finner at elasticiteten til boligpris er på 5,97%. I frykt for et simultanitetsproblem tar jeg i bruk instrumentmetoden for å få fjernet eventuelle endogenitetsproblemer. Inspirert av Ciccone og Hall (1996) tar jeg i bruk historisk befolkningstetthet som instrument. Resultatene fra instrumentmetoden viser at elasticiteten til nominell lønn med hensyn på befolkningstetthet er 2,34% når det kontrolleres for utdanningsnivå. Annen forskning gjort av Carlsen et al. (2012) viser at agglomerasjonseffektene reduseres når det kontrolleres for individuell uobserverbar heterogenitet som for eksempel ferdigheter slik at dette burde vært kontrollert for. Elasticiteten til boligpris er på 4,42% når det kontrolleres for befolkningsvekst og utdanningsnivå. Resultatene forteller at folk er villig til å flytte til kommuner på tross av at boligprisen er høy. Dette tyder på at kommunen har andre positive effekter som for eksempel gode offentlige eller flott natur.

Jeg finner videre at økt befolkningstetthet har en negativ påvirkning på reallønnen. Etter instrumentering og kontroll for befolkningsvekst og utdanningsnivå finner jeg at elastisiteten til reallønnen med hensyn på befolkningstetthet er -2,40%.

Dette kan være en konsekvens av at lønnsforhandlingene i Norge er langsomme, mens boligprisen kan reagere raskere økning av befolkningstettheten. Ut ifra mine resultater må husholdninger ta til takke med lavere reallønn hvis de skal flytte til kommuner med høy befolkningstetthet. Men disse kommunene kan ha flotte fasiliteter som veier opp for reduksjonen i reallønnen.

Referanser

- Carlsen, F., J. Rattsø og H.E. Stokke (2012): "Urban wage premium increasing with education level: Identification of agglomeration effects for Norway". Norges tekniske-naturvitenskapelig universitet.
- Carlton, D. W (1983): "The Location and Employment Choices of New Firms: An Econometric Model with Discrete and Continuous Endogenous Variables" *Review of Economics and Statistics*, 14, 209–220.
- Cheshire, P. og S. Magrini (2006): "Population Growth in European Cities: weather matters - but only nationally", *Regional studies*, 40(1):23-37
- Ciccone, A og R. Hall (1996), "Productivity and the density of economic activity", *American Economic Review* 86,(1): 54-70
- Ciccone, A (2002): "Agglomeration effects in Europe". *European Economic Review* 46(2): 213-227
- Combes, P-P, G. Duranton, L. Gobillon og S. Roux (2010): "Estimating Agglomeration Economies with History, Geology and Worker Effects". Glaeser, E.L.(ed) *Agglomeration Economics* 15-66. Chicago, US: University of Chicago Press.
- Combes, P-P, G. Duranton og L. Gobillon (2011). "The identification of Agglomeration Economies", *Journal of Economic Geography* 11: 253-266
- Combes, P-P. G. Duranton og L. Gobillon (2012). "The Cost of Agglomeration: Land Prices in French Cities.
- Costa, Dora L. og Matthew E. Kahn (2000). "Power couples: changes in the locational choice of the college educated 1940-1990", *Quarterly Journal of Economics* 115(4): 1287-1315.
- Dekle, R. og J. Eaton (1999): "Agglomeration and land rents: Evidence from the prefectures" *Journal of Urban Economics* 46(2): 200-214
- Duranton, G og D. Puga (2004): "Micro-foundations of urban agglomeration economies", *Handbook of Urban and Regional Economics* vol.14 Geography and Cities. Ed. Henderson, J.V., Thisse J-F
- Eaton, J. og Z. Eckstein (1997): "Cities and Growth: Theory and Evidence from France and Japan", *NBER Working paper* 4612
- Ellison, Glenn og Edward L. Glaeser (1997): "Geographic concentration in US manufacturing industries: a dartboard approach". *Journal of Political Economy* 105(5): 889-927

- Ellison, Glenn og Edward L. Glaeser (1999): "The geographic concentration of industry: does natural advantage explain agglomeration?" *American Economic Review Papers and Proceedings* 89(2): 311-316.
- Fujita og Thisse (2002). *Economics og Agglomeration: Cities, Industrial Location, and Regional Growth*. Cambridge: Cambridge University Press
- Gibbons, S., H.G. Overman og G. Resende (2011): "Real Earnings Disparities in Britain", Spatial Economics Research Centre, London School of Economics.
- Glaeser, E.L., H.D. Kallal, J. A. Scheinkman og A. Shleifer (1992): "Growth in cities" *Journal of Political Economy* 100, 1126-1152
- Glaeser, E.L og C. Maré (2001): "Cities and Skills", *Journal of Labour Economics* 19(2): 316-342
- Glaeser, E. L., G. Ponzetto og K. Tobio (2011): "Cities, skills and regional change", *Working Paper 16934*
- Gyourko, J. C. Mayer og T. Sinai (2010): "Dispersion in House Price and Income Growth across Markets: Facts and Theories", Glaeser, E.L (ed) *Agglomeration Economics*. Chicago, US: University of Chicago Press.
- Henderson, J. V., A. Kuncoro og M. Turner (1995): "Industrial Development in Cities", *Journal of Political Economy* 103(5): 1067-1090
- Henderson, J. V (1997): "Externalities and industrial development". *Journal of Urban Economics* 42(3): 449-470
- Kahn, M.E. (2010): "New Evidence on Trends in the Cost of Urban Agglomeration", Glaeser, E.L.(ed) *Agglomeration Economics*. Chicago, US: University of Chicago Press.
- Koster, H.R.A (2012): "Rocketing rents: The magnitude and attenuation of agglomeration economies"
- Marshall, A (1920). *Principles of Economics* (Revised Edition ed.). London: Macmillan and Co.
- Melitz, M. og G. I. P. Ottaviano (2008): "Market size, trade and productivity". *Reveiw of Economic Studies* 75(1): 295-316
- Moretti, E (2010): "Local Labor Markets". *Handbook of Labor Economics, Volume 4(2): 1254-1296*
- Puga, D. (2010): "The magnitude and causes of Agglomeration economies", *Journal of Regional Science* 50(1): 203-219.
- Rosenthal, S. S og W. C. Strange (2003): " Geography, Industrial organization and agglomeration". *Reveiw of Economics and Statistics* 85(2): 377-393

- Rosenthal, S. S. og W. C Strange (2008): "The attenuation of human capital spillovers". *Journal of Urban Economics* 64, 373-389
- Stock, J.H. Wright og M. Yogo (2002): "A survey of Weak Instruments and Weak Identification in Generalized Methods of Moments". *Journal of Business and Economics Statistics* 20(4): 518-529
- Sveikauskas, L. (1975): "Productivity of cities. *Quarterly Journal of Economics* 89(3):393-413
- Wheaton, W. C og M. J. Lewis (2002): "Urban wage and labor market agglomeration". *Journal of Urban Economics* 51(3): 542-562
- Wheeler, C.H. (2001): "Search, Sorting and Urban Agglomeration", *Journal of Labor Economics* 19(4): 879-899
- Wooldridge, J.M. (2009). *Introductory Econometrics: A Modern Approach* (4.utgave.) Canada: South-Western

Appendix

A Matematiske utregninger

A.1 Førsteordensbetingelse for arbeidskraft

Cobb-Douglas produktfunksjon er gitt ved

$$y_c = \bar{X} \bar{N}_c^h \bar{K}_c^{1-h} \quad (4)$$

Førsteordensbetingelsen for arbeidskraft

$$\frac{\partial y_c}{\partial \bar{N}_c} = \bar{w}_c$$

($\ln \bar{w} = w$, $\ln \bar{N} = N$, $\ln \bar{K} = K$, $\ln \bar{X} = X$)

Deriverer produktfunksjonen med hensyn på \bar{N}_c

$$\bar{X}_c h \bar{N}_c^{h-1} \bar{K}_c^{1-h} = \bar{w}_c$$

skriver om på logaritmeform

$$\begin{aligned} \ln \bar{X}_c + \ln h + (h-1) \ln \bar{N}_c + (1-h) \ln \bar{K}_c &= \ln \bar{w}_c \\ x_c + \ln h - (1-h) N_c + (1-h) K_c &= w_c \end{aligned} \quad (5)$$

A.2 Førsteordensbetingelsen for kapital

Deriverer Cobb-Douglas produktfunksjon med hensyn på kapital, \bar{K}_c

$$\frac{\partial y_c}{\partial \bar{K}_c} = i \Rightarrow \bar{X}_c * (1-h) \bar{N}_c^h \bar{K}_c^{-h} = i$$

Hvor i er prisen på kapital, og er eksogen

Skriver på logaritmeform:

$$\ln \bar{X}_c + \ln(1-h) + h * \ln \bar{N}_c - h * \ln \bar{K}_c = \ln i$$

Skriver om

$$X_c + \ln(1-h) + h * N_c - h * K_c = \ln i$$

Løser for K_c

$$h K_c = X_c + h N_c + \ln(1-h) - \ln i$$

$$K_c = \frac{X_c}{h} + N_c + \frac{\ln(1-h)}{h} - \frac{\ln i}{h} \quad (6)$$

A.3 Produktivitetssjokk

Må finne nye uttrykk for nominell lønn. Setter inn uttrykkene for kapital, K_a og K_b fra ligning (6) inn i ligning (5):

$$w_a = X_a - (1-h)N_a + (1-h) \left[\frac{X_a}{h} + N_b + \frac{\ln(1-h)}{h} - \frac{\ln i}{h} \right] + \ln h$$

$$w_a = \left(1 + \frac{1-h}{h}\right) X_a + \frac{1-h}{h} \ln(1-h) - \frac{1-h}{h} \ln i + \ln h$$

$$w_a = \frac{1}{h} X_a + \frac{1-h}{h} \ln(1-h) - \frac{1-h}{h} \ln i + \ln h \quad (5a)'$$

Og tilsvarende for b:

$$w_b = \frac{1}{h} X_b + \frac{1-h}{h} \ln(1-h) - \frac{1-h}{h} \ln i + \ln h \quad (5b)'$$

Løsning for N_b

$$N_b = \frac{N}{s} ((w_b - w_a) - (r_b - r_a) - (A_a - A_b)) + N_a \quad (3)'$$

Setter inn de nye uttrykkene for nominell lønn fra ligning (5a)' og (5b)' og tilbud-og etterspørselsligningene for bolig, (8a) og (8b):

$$N_b = \frac{N}{s} \left[\frac{1}{h} X_b + \frac{1-h}{h} \ln(1-h) - \frac{1-h}{h} \ln i + \ln h - \frac{1}{h} X_a - \frac{1-h}{h} \ln(1-h) + \frac{1-h}{h} \ln i - \ln h - z - k_b N_b + z + k_a N_a - A_a + A_b \right] + N_a$$

$$N_b = \frac{N}{s} \left[\frac{1}{h} (X_b - X_a) - k_b N_b + k_a N_a - A_a + A_b \right] + N_a$$

Bruker at $N_a = N - N_b$ og setter inn for N_a

$$N_b = \frac{N}{s} \left[\frac{1}{h} (X_b - X_a) - k_b N_b + k_a (N - N_b) - A_a + A_b \right] + N - N_b$$

Løser for N_b :

$$2N_b + \frac{N}{s} (k_b + k_b) N_b = \frac{N}{s} \left[\frac{1}{h} (X_b - X_a) + k_a N - A_a + A_b \right] + N$$

$$\frac{N}{s} \left\{ \frac{2s}{N} N_b + (k_b + k_b) N_b \right\} = \frac{N}{s} \left[\frac{1}{h} (X_b - X_a) + k_a N - A_a + A_b + s \right]$$

$$\begin{aligned} \left(\frac{2s}{N} + k_b + k_a\right) N_b &= \frac{1}{h}(X_b - X_a) + k_a N - A_a + A_b + s \\ N_b &= \frac{\frac{1}{h}(X_b - X_a) + k_a N - A_a + A_b + s}{\frac{2s}{N} + k_b + k_a} \end{aligned} \quad (3)''$$

Setter produktivitetssjokket $X_{b2} = X_{b1} + \Delta$ inn i uttrykket for N_b og finner endringen fra periode 1 til periode 2

$$\begin{aligned} N_{b2} - N_{b1} &= \frac{\frac{1}{h}(X_{b1} + \Delta - X_a) + k_a N - A_a + A_b + s}{\frac{2s}{N} + k_b + k_a} \\ &\quad - \frac{\frac{1}{h}(X_{b1} - X_a) + k_a N - A_a + A_b + s}{\frac{2s}{N} + k_b + k_a} = \frac{\frac{1}{h}\Delta}{\frac{2s}{N} + k_b + k_a} \end{aligned}$$

Ganger med hN i teller og nevner

$$(N_{b2} - N_{b1}) \frac{N}{h[(k_a + k_b)N + 2s]} \Delta \geq 0 \quad (9)$$

For å finne endringen i nominell lønn i by b må man først sette inn for produktivitetssjokket i periode 2 i ligning (5b) så trekker ifra periode 1

$$\begin{aligned} (w_{b2} - w_{b1}) &= \frac{1}{h}(X_{b1} + \Delta) + \frac{1-h}{h} \ln(1-h) - \frac{1-h}{h} \ln i + \ln h - \frac{1}{h}(X_{b1}) \\ &\quad - \frac{1-h}{h} \ln(1-h) + \frac{1-h}{h} \ln i - \ln h \\ (w_{b2} - w_{b1}) &= \frac{1}{h}\Delta \end{aligned} \quad (10)$$

Må finne endringen i boligkostnad for by b etter sjokket. Setter inn N_b fra ligning (3)'' inn i r_b fra ligning (8b)

$$\begin{aligned} r_b &= z + k_b N_b \quad (8b) \\ r_b &= z + k_b \frac{\frac{1}{h}(X_b - X_a) + k_a N - A_a + A_b + s}{\frac{2s}{N} + k_b + k_a} \end{aligned}$$

$$(r_{b2} - r_{b1}) = z + k_b \frac{\frac{1}{h}(X_{b1} + \Delta - X_a) + k_a N - A_a + A_b + s}{\frac{2s}{N} + k_b + k_a} - z$$

$$+ k_b \frac{\frac{1}{h}(X_b - X_a) + k_a N - A_a + A_b + s}{\frac{2s}{N} + k_b + k_a} = \frac{k_b \frac{1}{h} \Delta}{\frac{2s}{N} + k_b + k_a}$$

Ganger med hN i både teller og nevner

$$(r_{b2} - r_{b1}) = \frac{k_b N}{h[N(k_b + k_a) + 2s]} \Delta \geq 0 \quad (11)$$

For å finne endringen i reallønnen tar man ligning (10) og trekker ifra ligning (11)

$$(w_{b2} - w_{b1}) - (r_{b2} - r_{b1}) = \frac{1}{h} \Delta - \frac{k_b N}{[2s + N(k_a + k_b)] h} \Delta$$

$$(w_{b2} - w_{b1}) - (r_{b2} - r_{b1}) = \frac{\frac{1}{h} \Delta [2s + N(k_a + k_b)] h - k_b N \Delta}{[2s + N(k_a + k_b)] h}$$

$$(w_{b2} - w_{b1}) - (r_{b2} - r_{b1}) = \frac{k_a N + 2s}{h[N(k_a + k_b) + 2s]} \Delta \geq 0 \quad (12)$$

A.4 Agglomerasjon

Vi får nye uttrykk for lønn. Setter inn det nye uttrykket for $X_c = x_c + \gamma N_c$ inn i uttrykket for k_a og k_b og får nye uttrykk for kapital

$$K_a = \frac{x_c + \gamma N_a}{h} + N_a + \frac{\ln(1-h)}{h} - \frac{\ln i}{h}$$

$$K_a = \frac{x_a}{h} + \frac{\gamma}{h} N_a + N_a + \frac{\ln(1-h)}{h} - \frac{\ln i}{h}$$

$$K_a = \frac{x_a}{h} + (1 + \frac{\gamma}{h}) N_a + \frac{\ln(1-h)}{h} - \frac{\ln i}{h}$$

Og likt for k_b

$$K_b = \frac{x_b}{h} + (1 + \frac{\gamma}{h}) N_b + \frac{\ln(1-h)}{h} - \frac{\ln i}{h}$$

Setter inn i ligning (15)

$$w_a = x_a + (\gamma - (1-h)) N_a + (1-h) \left[\frac{x_a}{h} + (1 + \frac{\gamma}{h}) N_a + \frac{\ln(1-h)}{h} - \frac{\ln i}{h} \right] + \ln h$$

$$w_a = \left(1 + \frac{1-h}{h}\right) x_a + \left(\gamma - (1-h) + (1-h)\left(1 + \frac{\gamma}{h}\right)\right) N_a + \frac{1-h}{h} (\ln(1-h) - \ln i) + \ln h$$

$$w_a = \frac{x_a}{h} + \left[\gamma + (1-h)\left(1 + \frac{\gamma}{h} - 1\right)\right] N_a + \frac{1-h}{h} (\ln(1-h) - \ln i) \ln h$$

$$w_a = \frac{x_a}{h} + \gamma \left(1 + \frac{1-h}{h}\right) N_a + \frac{1-h}{h} (\ln(1-h) - \ln i) + \ln h$$

$$w_a = \frac{x_a}{h} + \frac{\gamma}{h} N_a + \frac{1-h}{h} (\ln(1-h) - \ln i) + \ln h$$

Tilsvarende for b:

$$w_b = \frac{x_b}{h} + \frac{\gamma}{h} N_b + \frac{1-h}{h} (\ln(1-h) - \ln i) + \ln h$$

Må finne N_b . Løser (3) med hensyn på N_b

$$N_b = \frac{N}{s} [w_b - w_a - (r_b - r_a) - (A_a - A_b)] + N_a$$

Setter inn for w_a, w_b, r_b og r_a

$$N_b = \frac{N}{s} \left[\frac{1}{h} (x_b - x_a) + \frac{\gamma}{h} (N_b - N_a) - k_b N_b + k_a N_a - A_a + A_b \right] + N_a$$

Setter inn for $N_a = N - N_b$

$$N_b = \frac{N}{s} \left[\frac{1}{h} (x_b - x_a) + \frac{\gamma}{h} N_b - \frac{\gamma}{h} (N - N_b) + k_a (N - N_b) - k_b N_b - A_a + A_b \right] + (N - N_b)$$

Løser med hensyn på N_b

$$2N_b + \frac{N}{s} \left[(k_a + k_b) - \frac{2\gamma}{h} \right] N_b = \frac{N}{s} \left[\frac{1}{h} (x_b - x_a) + \frac{\gamma}{h} N + k_a N - (A_a - A_b) + s \right]$$

$$N_b = \frac{\frac{1}{h} (x_b - x_a) - \frac{\gamma}{h} N + k_a N - (A_a - A_b) + s}{\frac{2s}{N} + (k_b + k_a) - 2\frac{\gamma}{h}} \quad (3^*)$$

Setter inn for produktivitetssjokket $x_{b2} = x_{b1} + \Delta$

$$(N_{b2} - N_{b1}) = \frac{\frac{1}{h}(x_{b1} + \Delta - x_a) - \frac{\gamma}{h}N + k_a N - (A_a - A_b) + s}{\frac{2s}{N} + (k_b + k_a) - 2\frac{\gamma}{h}} - \frac{\frac{1}{h}(x_{b1} - x_a) - \frac{\gamma}{h}N + k_a N - (A_a - A_b) + s}{\frac{2s}{N} + (k_b + k_a) - 2\frac{\gamma}{h}}$$

$$(N_{b2} - N_{b1}) = \frac{\frac{1}{h}\Delta}{\frac{2s}{N} + (k_b + k_a) - 2\frac{\gamma}{h}}$$

Ganger med hN i teller og nevner

$$(N_{b2} - N_{b1}) = \frac{N\Delta}{h[N(k_b + k_a) + 2s] - 2\gamma N} \quad (18)$$

A.4.1 Effekt på lønn i by b

Setter inn sjokket i ligningen for lønn

$$(w_{b2} - w_{b1}) = \frac{1}{h}(x_{b1} + \Delta) + \frac{\gamma}{h}N_{b2} + (1-h)\frac{\ln(1-h)}{h} - (1-h)\frac{\ln i}{h} + \ln h - \frac{1}{h}x_{b1} - \frac{\gamma}{h}N_{b1} - (1-h)\frac{\ln(1-h)}{h} - (1-h)\frac{\ln i}{h} - \ln h$$

$$(w_{b2} - w_{b1}) = \frac{1}{h}\Delta + \frac{\gamma}{h} * \frac{N\Delta}{h[N(k_a + k_b) + 2s] - 2\gamma N}$$

$$(w_{b2} - w_{b1}) = \frac{h[N(k_a + k_b) + 2s] - 2\gamma N}{h^2[N(k_a + k_b) + 2s] - 2\gamma N h} \Delta$$

$$\gg (w_{b2} - w_{b1}) = \frac{h[N(k_b + k_a) + 2s] - \gamma N}{h[h(N(k_b + k_a) + 2s) - 2\gamma N]} \Delta \quad (16)$$

Deriverer med hensyn på γ

$$u' = -\Delta N$$

$$v' = -2hN$$

$$u = \{h[N(k_b + k_a) + 2s] - \gamma N\}\Delta$$

$$v = h[h(N(k_b + k_a) + 2s) - 2\gamma N]$$

$$\frac{u'v - uv'}{v^2}$$

$$\begin{aligned}
& \frac{\partial(w_{b2} - w_{b1})}{\partial\gamma} \\
&= \frac{\Delta N h [h(N(k_a + k_b) + 2s - 2\gamma N) + 2N h [h(N(k_a + k_b) + 2s - \gamma N) \Delta]}{\{h[h(N(k_b + k_a) + 2s) - 2\gamma N]\}^2} \\
& \frac{\partial(w_{b2} - w_{b1})}{\partial\gamma} = \frac{N \Delta h \{2\gamma N - h(N(k_a + k_b) + 2s) + 2h(N(k_a + k_b) - 2\gamma N)\}}{\{h[h(N(k_b + k_a) + 2s) - 2\gamma N]\}^2} \\
& \frac{\partial(w_{b2} - w_{b1})}{\partial\gamma} = \frac{\Delta N h [N(k_a + k_b)]}{\{h[h(N(k_b + k_a) + 2s) - 2\gamma N]\}^2} \\
& \frac{\partial(w_{b2} - w_{b1})}{\partial\gamma} = \frac{N \Delta h^2 (N(k_a + k_b) + 2s)}{\{h[h(N(k_b + k_a) + 2s) - 2\gamma N]\}^2} \tag{17}
\end{aligned}$$

A.4.2 Effekt på boligpris

Setter inn det nye uttrykket for N_b inn i uttrykket for boligkostnad

$$r_b = z + k_b N_b \tag{8b}$$

$$r_b = z + k_b \frac{\frac{1}{h}(x_b - x_a) - \frac{\gamma}{h}N + k_a N - A_a + A_b + s}{\frac{2s}{N} + (k_b + k_a) - 2\frac{\gamma}{h}}$$

Setter inn sjokket og finner endringen fra periode 1 til periode 2

$$\begin{aligned}
(r_{b2} - r_{b1}) &= z + k_b \frac{\frac{1}{h}(x_{b1} + \Delta - x_a) - \frac{\gamma}{h}N + k_a N - A_a + A_s + s}{\frac{2s}{N} + (k_b + k_a) - 2\frac{\gamma}{h}} - z \\
&\quad - k_b \frac{\frac{1}{h}(x_{b1} - x_a) - \frac{\gamma}{h}N + k_a N - A_a + A_b + s}{\frac{2s}{N} + (k_b + k_a) - 2\frac{\gamma}{h}} \\
(r_{b2} - r_{b1}) &= \frac{k_b \frac{1}{h} \Delta}{\frac{2s}{N} + (k_b + k_a) - 2\frac{\gamma}{h}}
\end{aligned}$$

Ganger med hN i teller og nevner

$$(r_{b2} - r_{b1}) = \frac{k_b N \Delta}{h[N(k_a + k_b) + 2s] - 2\gamma N} \tag{19}$$

A.4.3 Endring i reallønn

Endringen i reallønnen finner man ved å trekke ligning (19) fra (16)

$$(w_{b2} - w_{b1}) - (r_{b2} - r_{b1}) = \frac{h(K_a N + 2s) - \gamma N}{h\{h[N(k_a + k_b) + 2s - 2\gamma N]\}} \Delta \quad (20)$$

Dervierer ligningen (20) med hensyn på γ

$$\begin{aligned} u' &= -\Delta N \\ v' &= -2hN \end{aligned}$$

$$\begin{aligned} &\frac{\partial((w_{b2} - w_{b1}) - (r_{b2} - r_{b1}))}{\partial \gamma} \\ &= \frac{-\Delta N h [h(N(k_a + k_b) + 2s) - 2\gamma N] + [N k_a + 2s] h - \gamma N \Delta * 2N h}{\{h[N(k_a + k_b) + 2s - 2\gamma N]\}^2} \end{aligned}$$

Viser mellomregningen i teller

$$\begin{aligned} &= \Delta N h [2\gamma N - h(N(k_a + k_b) + 2s) + 2h(N k_a + 2s) - 2\gamma N] \\ &= \Delta N h [2h(N k_a + 2s) - h(N(k_a + k_b) + 2s)] \\ &= \Delta N h [2h N k_a + 4hs - N h(k_a + k_b) - 2sh] \\ &= \Delta N h [2N h k_a + 2hs - N h k_a - N h k_b] \end{aligned}$$

$$\frac{\partial((w_{b2} - w_{b1}) - (r_{b2} - r_{b1}))}{\partial \gamma} = \frac{\Delta N h^2 [N(k_a + k_b) + 2s]}{\{h[N(k_a + k_b) + 2s - 2\gamma N]\}^2} \quad (21)$$

B Tabeller

B.1 Alternativt instrument

Tabell B.1.1: IV-estimering, første steg. Endring i befolkningstetthetfer 1825-1900 som instrument

	(1) IV Ln tetthet	(2) IV Ln tetthet	(3) IV Ln tetthet	(4) IV Ln tetthet
Ln endring tetthet (1825-1900)	0.703*** (0.041)	0.601*** (0.035)	0.559*** (0.041)	0.5421*** (0.036)
Befolkningsvekst		0.077*** (0.008)		0.0597*** (0.0087)
Utdanning			0.776*** (0.010)	0.044*** (0.0103)
Konstant	2.447*** (0.1037)	2.13*** (0.0895)	1.14*** (0.195)	1.452*** (0.1781)
Observasjoner	162	162	162	162
R ²	0.640	0.770	0.734	0.794
F-verdi	284.42	266.69	219.37	203.84

Standardavvik oppgitt i parentes..

*** Signifikant på 1%nivå ** Signifikant på 5% nivå, * Signifikant på 10% nivå

B.2 2SLS-estimeringer for 2002, 2004 og 2007

Tabell B.2.1: IV-estimering, andre steget, 2002

	(1) 2SLS Ln reallønn	(2) 2SLS Ln reallønn	(3) 2SLS Ln reallønn	(4) 2SLS Ln reallønn
Ln tetthet	-0.0699*** (0.0117)	-0.0474*** (0.0133)	-0.0311** (0.0152)	-0.0247* (0.0150)
befolkningsvekst		-0.124*** (0.0251)		-0.0990*** (0.0236)
Utdanning			-0.0154*** (0.00297)	-0.0109*** (0.00270)
Konstant	2.987*** (0.0463)	2.981*** (0.0431)	3.133*** (0.0425)	3.085*** (0.0432)
Observasjoner	162	162	162	162
R ²	0.322	0.426	0.417	0.483

Standardavvik oppgitt i parentes

*** Signifikant på 1%nivå ** Signifikant på 5% nivå, * Signifikant på 10% nivå

Tabell B.2.2: IV-estimering, andre steget, 2004

	(1) reg1 Ln reallønn	(2) reg2 Ln reallønn	(3) reg3 Ln reallønn	(4) reg4 Ln reallønn
Ln tetthet	-0.0630*** (0.0126)	-0.0349** (0.0138)	-0.0291* (0.0172)	-0.0211 (0.0162)
befolkningsvekst		-0.156*** (0.0262)		-0.141*** (0.0257)
Utdanning			-0.0130*** (0.00330)	-0.00635** (0.00287)
Konstant	2.954*** (0.0497)	2.946*** (0.0447)	3.088*** (0.0485)	3.012*** (0.0478)
Observasjoner	162	162	162	162
R-squared	0.276	0.426	0.320	0.442

Standardavvik oppgitt i parentes..

*** Signifikant på 1%nivå ** Signifikant på 5% nivå, * Signifikant på 10% nivå

Tabell B.2.3: IV-estimering, andre steget, 2007

	(1) reg1 Ln reallønn	(2) reg2 Ln reallønn	(3) reg3 Ln reallønn	(4) reg4 Ln reallønn
Ln tetthet	-0.0868*** (0.0137)	-0.0582*** (0.0152)	-0.0479** (0.0189)	-0.0409** (0.0179)
befolkningsvekst		-0.160*** (0.0292)		-0.142*** (0.0283)
Utdanning			-0.0141*** (0.00350)	-0.00743** (0.00307)
Konstant	2.857*** (0.0546)	2.849*** (0.0495)	3.014*** (0.0532)	2.933*** (0.0530)
Observasjoner	162	162	162	162
R ²	0.328	0.462	0.390	0.488

Standardavvik oppgitt i parentes.

*** Signifikant på 1%nivå ** Signifikant på 5% nivå, * Signifikant på 10% nivå

