

John Olav Skjønsberg
Sigrid Kornberg Skjeflo

Absolutt og betinget befolkningsdivergens i Norge

En empirisk analyse av drivkrefter til regional
befolkningsvekst

Masteroppgave i Samfunnsøkonomi

Veileder: Jørn Rattsø

Juni 2023

John Olav Skjønsberg
Sigrid Kornberg Skjeflo

Absolutt og betinget befolkningsdivergens i Norge

En empirisk analyse av drivkrefter til regional
befolkningsvekst

Masteroppgave i Samfunnsøkonomi
Veileder: Jørn Rattsø
Juni 2023

Norges teknisk-naturvitenskapelige universitet
Fakultet for økonomi
Institutt for samfunnsøkonomi



Kunnskap for en bedre verden

Sammendrag

Formålet med denne oppgaven er å analysere hvordan ulike faktorer driver befolkningsvekst i de økonomiske regionene i Norge, og hvordan forskjeller i befolkningsvekst i regionene forklarer befolkningsdivergens. Absolutt divergens blir undersøkt for perioden 1910-2013, men vi fokuserer på perioden 1995-2013 når vi undersøker for betinget divergens. Resultatene viser både absolutt og betinget divergens i befolkningsnivået i regionene, og at divergensprosessen har akselerert i nyere tid.

Spline-regresjoner og kernel-tettheter blir estimert for å kunne forklare befolkningsmønstre. Resultatene indikerer at det er de mellomstore- og periferiregionene som har drevet befolkningsdivergensen. I utvidelsen av analysen er resultatene for divergens robuste for perioden 1995-2013 når vi kontrollerer for humankapital og sysselsetting i ulike næringssektorer. Resultatene forteller oss at den gjennomsnittlige årlige befolkningsveksten er på 0,4 prosentpoeng i denne perioden.

Høyere humankapitalnivå i en region og høy sysselsetting i business- og produksjonssektoren har en positiv effekt på befolkningsveksten, blant annet gjennom agglomerasjonseffekter. Servicesektor-yrker bidrar ikke til de samme agglomerasjonseffektene, og vi finner en negativ effekt av sysselsetting i disse yrkene på befolkningsveksten. Dette er i tråd med tidligere litteratur på feltet. Bartik-instrumenter brukes som en robusthetssjekk av sysselsetting i de ulike næringssektorene, og vi finner at instrumentene gir en mer kausal effekt på befolkningsveksten.

Abstract

The purpose of this thesis is to analyze how different factors drive population growth in the economic regions of Norway, and how differences in population growth in the regions explain population divergence. Absolute divergence is examined for the period 1910-2013, but we focus on the period 1995-2013 when examining conditional divergence. The results show both absolute and conditional divergence in population levels in the regions, and that the divergence process has accelerated in recent years.

Spline regressions and kernel densities are estimated to explain population patterns. The results indicate that it is the medium-sized and peripheral regions that have driven population divergence. In the extension of the analysis, the results for divergence are robust for the period 1995-2013 when we control for human capital and employment in different industry sectors. The results tell us that the average annual population growth is 0.4 percentage points in this period.

Higher human capital levels in a region and high employment in the business and manufacturing sectors have a positive effect on population growth, partly through agglomeration effects. Service sector occupations do not contribute to the same agglomeration effects, and we find a negative effect of employment in these occupations on population growth. This is in line with previous literature in the field. Bartik instruments are used as a robustness check of employment in the different industry sectors, and we find that the instruments have a more causal effect on population growth.

Forord

Denne oppgaven er utarbeidet ved Institutt for samfunnsøkonomi ved NTNU som en avsluttende del av et 5-årig integrert masterprogram i samfunnsøkonomi. I løpet av disse årene har vi fått testet samarbeidsevnene våre, og avslutter samarbeidet på topp. Lunsjpausene oppe i 5. etasje med våre medstudenter har vært en viktig motivasjonsfaktor. Videre vil vi takke Gilles Durantan for innspill til oppgaven. Takk til Jørn Rattsø for god hjelp og konstruktive tilbakemeldinger. Til tross for skeptisk mottakelse fra vår side, har tilbakemeldingene uten tvil styrket oppgaven.

John Olav Skjøsberg
Sigrid Kornberg Skjeflo
Trondheim, mai 2023

Innholdsfortegnelse

1 Innledning	1
2 Tidligere empiriske studier	3
2.1 Konvergens og divergens	3
2.2 Humankapital og agglomerasjonseffekter	5
2.3 Næringssektorer og agglomerasjonseffekter	7
2.4 Bartik-instrumentet	9
2.5 Oppsummering og hypoteser	11
3 Datamateriale	13
3.1 Avhengig variabel og befolkningstall	14
3.2 Forklaringsvariabel	15
3.3 Kontrollvariabler	16
4 Metode	18
4.1 Barroregresjoner	18
4.2 Spline-regresjoner	20
4.3 Kernel-tettheter	21
4.4 Lokal polynomglatting (ikke-lineært kernelplott)	22
4.5 Bartik-instrumenter	23
4.6 Økonometriske utfordringer	26
4.6.1 Simultanitet	26
4.6.2 Utelatte variable	26
4.6.3 Målefeil	26
4.6.4 Multikollinearitet	27
4.6.5 Heteroskedastisitet og seriekorrelasjon	27
4.6.6 Metodediskusjon	28
5 Analyse	31
5.1 Befolkningsutvikling	31
5.2 Ubetingede barroregresjoner	32
5.3 Spline-regresjoner	36
5.4 Lokal polynomglatting (ikke-lineært kernelplott)	37
5.5 Kernel-tettheter	38
5.6 Barroregresjoner og humankapital	39
5.7 Barroregresjoner og næringssektorer	43
5.8 Barroregresjoner og Bartik-instrumenter	45
5.9 Bartik-instrumenter med tidligere befolkningstall	49
6 Konklusjon og videre forskning	52
7 Kilder	54

8	Appendiks.....	59
A1	Sektorer fordelt etter industri.....	59
A2	Kernel-tetthetsfunksjoner for 1910 og 2013.....	59
A3	Uten tidligere befolkningsnivåer	60
A4	Annet mål på humankapital.....	60
A5	Korrelesasjonsmatriser	61
A5.1	Korrelasjonsmatrise utdanning og næringssektorer	61
A5.2	Korrelasjonsmatrise befolkning.....	61
A5.3	Korrelasjonsmatrise utdanning og Bartik-instrumenter	61
A6	Endring i sysselsetting i olje- og gassnæringen og bygg- og anleggsvirksomhet	62

1 Innledning

Forskning innenfor urban økonomi og økonomisk geografi fokuserer vanligvis på å forstå organiseringen av økonomisk aktivitet, faktorer som driver vekst og utvikling av urbane områder, og implikasjonene disse har for regionale og nasjonale økonomier. Ved å studere ulike aspekter ved urban økonomi og økonomisk geografi ønsker forskere å utvikle en bedre forståelse av komplekse prosesser som former vekst og utvikling av byer og regioner, samt informere offentlige beslutningstakere for å fremme mer bærekraftige, rettferdige og økonomisk fremgangsrike urbane miljøer.

Norge har gjennomgått store endringer de siste 100 årene. Den økonomiske veksten har vært betydelig, befolkningen har økt og endringer blant annet i næringsstruktur, utdanningsnivå og transportmuligheter har bidratt til at en stadig økende andel av befolkningen bor i byer og tettsteder. Befolkningsendringer skjer imidlertid i ulikt tempo avhengig av en rekke regionspesifikke forhold. For eksempel er veksten i olje- og gassnæringen i Stavanger/Sandnes-regionen et eksempel på nærhet til naturressurser som en regionspesifikk driver for befolkningsvekst.

I denne oppgaven studerer vi divergens i befolkningsnivået i 90 norske økonomiske regioner ved bruk av barroregresjoner. Befolkningsdivergens refererer til prosessen der befolkningen i regionene utvikler seg i forskjellige retninger. Vi vil undersøke om det har vært absolutt divergens i lengre og kortere perioder, men fokuserer på perioden 1995-2013 når vi undersøker for betinget divergens. Betinget divergens betyr i vår sammenheng at kontrollvariablene vi inkluderer, er med på å forklare befolkningsvekst. Vi belyser hvordan sysselsetting, fordelt på tre næringssektorer, og utdanningsnivå påvirker befolkningsvekst. Sosiale og kulturelle faktorer vil også kunne spille inn, men er ikke analysert i denne oppgaven. Inspirasjon til variabler og metode er hentet fra Duranton (2016).

Ved å undersøke om det har vært absolutt divergens i norske regioner kan vi observere hvordan regioner endrer seg i størrelse over tid kun basert på deres initielle størrelse. Hvordan har befolkningsnivåene variert blant Norges 90 økonomiske regioner i perioden 1995-2013 – har det vært absolutt konvergens eller absolutt divergens i befolkningsnivåene? For å forstå befolkningsmønstre bedre, inkluderes spline-regresjoner og kernel-tetthetsfunksjoner. Ved å se

på ulike underutvalg kan man se om det er divergens blant spesifikke befolkningsgrupper. I så fall er det interessant å se hvilke grupper som driver divergensen for hele landet.

Videre ønsker vi å se på faktorer som kan forklare den eventuelle divergensen. Hva har humankapital og sysselsetting i produksjons-, business-, og servicesektoren å si for befolkningsveksten i regionene? Dersom lokal arbeidskraftetterspørsel i disse næringssektorene drives av arbeidstilbudssjokk, hva skjer med befolkningsveksten hvis vi fjerner den delen av arbeidstilbudssjokket som retter seg mot den spesifikke regionen? Bartik-instrumenter fjerner potensielt slike endogenitetsproblemer ved å isolere den delen av arbeidskraftetterspørsel som er drevet av nasjonale sjokk, og dermed fjerner regionspesifikke effekter av arbeidstilbudssjokk. Resultatene av denne studien kan bidra til en bedre forståelse av urban økonomi og regional utvikling, både i Norge og i en bredere global sammenheng. Å bruke Bartik-instrumenter som annet mål på sysselsetting i de ulike sektorene, vil kunne tilføre et nytt perspektiv til feltet.

Oppgaven er strukturert i fem påfølgende kapitler. I kapittel 2 presenteres sentrale begreper og det teoretiske fundamentet for denne oppgaven. Vi vil undersøke den eksisterende litteraturen på determinanter som driver befolkningsvekst, og vise deres relevans for vår analyse. I kapittel 3 redegjøres det for datamaterialet som er brukt i oppgaven. I kapittel 4 utledes metodene som er benyttet i den empiriske analysen og utfordringer knyttet til dette. Selve analysen og diskusjon rundt denne gjøres i kapittel 5. Avslutningsvis, i kapittel 6, kommer konklusjonen hvor vi oppsummerer analysen, knytter empirien opp mot teorien og foreslår videre forskning.

2 Tidligere empiriske studier

2.1 Konvergens og divergens

Konvergens og divergens er to konsepter som belyser hvordan ulikheter endres over tid. Begrepene gir oss mulighet til å studere endringene i fordelingen av verdier, som i vårt tilfelle er befolkningsstørrelser i ulike regioner.

Robert J. Barro var en pioner innen tverssnittsanalyser mellom land (Barro, 1991). Han baserte sitt arbeid på den neoklassiske vekstmodellen til Solow (1956), for å undersøke om inntektsveksten i fattige og rike land konvergente mot en langsiktig likevekt. I denne likevektstilstanden vil bruttonasjonalprodukt (BNP) per innbygger (inntekt per innbygger) vokse med en konstant rate, og forskjellene mellom land vil gradvis avta over tid. Det finnes to hovedtyper av inntektskonvergens; absolutt konvergens og betinget konvergens. Absolutt konvergens innebærer at alle land, uavhengig av deres utgangspunkt, vil konvergere mot samme inntektsnivå i det lange løp. Absolutt konvergens forutsetter at alle land har lik tilgang til teknologi, kapital og andre produksjonsfaktorer. Betinget konvergens innebærer at land vil konvergere mot deres eget langsiktige inntektsnivå, som er bestemt av deres spesifikke egenskaper, som humankapital, teknologisk nivå, naturressurser og andre strukturelle faktorer. Betinget konvergens anerkjenner at det kan være vedvarende forskjeller mellom land, men at land med lavere inntektsnivå vil vokse raskere og gradvis redusere gapet til rikere land.

Barro og Sala-i-Martin (1992) brukte videre den neoklassiske vekstmodellen til å studere inntektskonvergens mellom 48 amerikanske stater. De fant at fattigere stater har en tendens til å vokse raskere enn rike, målt i BNP per innbygger, fordi «loven» om avtagende avkastning på kapital som en egenskap ved økonomisk vekst, inntreffer saktere for fattigere stater enn rike. Fattigere stater har større potensial for vekst i BNP fordi de har lavere inntektsnivåer enn rike stater, og kan dermed oppnå høyere vekstrater raskere gitt at statene er relativt like i preferanser og teknologi. De trakk frem at forskjellene i amerikanske stater er så små at det trolig ikke er noe poeng å skille mellom absolutt og betinget konvergens innenfor et enkeltland. Hastigheten på konvergens viste seg å være omtrent lik, uansett tidsperiode de studerte og uansett hvilke variabler de kontrollerte for.

Divergens er det motsatte av konvergens, dvs. det innebærer en økning i ulikheter og at verdier beveger seg bort fra en likevekt. I lys av studien til Duranton (2016) kan befolkningsdivergens i Colombia forklares ved at byer med høyere initielle befolkningsstørrelser enn distriktsområder vokser med en høyere vekstrate i perioden 1993-2010. Sterkere urbanisering forklarer befolkningsdivergens ved at agglomerasjonseffektene i større geografiske områder er sterkere, og dermed tiltrekker seg flere folk. Disse menneskene må komme fra et sted. Om man ser bort fra fødselsoverskudd og innvandring fra andre land må noen regioner bli mindre for at andre skal kunne vokse. Duranton og Puga (2013) forklarte hvordan infrastruktur og boligtilbud kan tiltrekke seg folk. Når disse menneskene har kommet til et område, vil de bidra til agglomerasjonseffekter ved at de tar høyere utdanning eller bidrar i kunnskapsintensive yrker. Dette gjør området mer attraktivt for folk utenfra, som dermed kan oppleve befolkningsvekst.

Regionene i Norge er delt opp i henhold til Hustoft et al. (1999). Ved å se på alle regioner i Norge, kan man lettere forstå befolkningsmobilitet fordi man ikke bare studerer hvordan befolkningsveksten i store byer utvikler seg, men også hvordan befolkningsveksten utvikler seg i mindre urbane områder. Forskning innen urban økonomi og økonomisk geografi er ikke like vanlig på regionnivå, men blant annet Ayuda et al. (2010) og Rattsø og Stokke (2014) fant regional divergens i Spania og Norge.

Ayuda et al. (2010) fant absolutt befolkningsdivergens i regioner i ni vest-europeiske land i perioden 1850-2000. De regionale befolkningsforskjellene varierer i nesten alle tilfeller mellom 3,3 og 4,2 prosentpoeng gjennomsnittlig, årlig befolkningsvekst. Den eneste påfallende forskjellen gjelder hastigheten på divergensprosessen i Frankrike (på 9,5 prosentpoeng årlig). Ayuda et al. (2010) argumenterte for at de viktigste grunnene til disse regionale befolkningsforskjellene er de demografiske forskjellene som eksisterte i Europa før industrialiseringen som tok av på midten av 1800-tallet.

Rattsø og Stokke (2014) fant at det har vært absolutt befolkningsdivergens, men inntektskonvergens i norske regioner i perioden 1972-2008. Befolkningsvekst er positivt korrelert med initielt befolkningsnivå, og ved å se på spline-regresjoner og kernel-tetthetsestimering fant Rattsø og Stokke (2014) at veksten har vært drevet av utflytting fra mellomstore regioner til mindre og større regioner. Michaels et al. (2012) fant også en positiv sammenheng mellom mellomstore regioner og befolkningsvekst når de så på regioner i USA

fra 1880 til 2000. Glaeser og Shapiro (2003) fant derimot at byveksten i USA på 1990-tallet ble drevet av de største byene.

2.2 Humankapital og agglomerasjonseffekter

Tidligere empiriske studier har vist en sterk sammenheng mellom befolkningsvekst og humankapital. Glaeser et al. (1995) brukte forskjellige typer utdanningsvariabler som mål på humankapital, mens i nyere litteratur har fokuset i stedet vært på andelen sysselsatte som er høyskole- eller universitetsutdannede (Glaeser og Saiz, 2004, Moretti, 2004, Duranton, 2016). Dette er et mer diskriminerende mål på utdanning, og tidligere litteratur har vist at det har en sterk kausal sammenheng med befolkningsvekst (Duranton og Puga, 2013).

Glaeser og Saiz (2004) undersøkte forholdet mellom humankapital og byvekst i amerikanske storbyområder fra 1970 til 2000 og fant at en økning på 1 prosentpoeng i andelen universitetsutdannede var assosiert med 0,5 prosentpoeng økning i befolkningen i en gjennomsnittlig by i det påfølgende tiåret. Den sterke korrelasjonen mellom humankapital og befolkningsvekst innebærer imidlertid ikke nødvendigvis kausalitet. Korrelasjonen kan skyldes at høyt utdannede arbeidere kan være mer tilbøyelige til å flytte til voksende byer før lavere utdannede arbeidere gjør det, enten fordi de er mer mobile eller har sterkere insentiver til å flytte. Som et resultat kan høyt utdannede blir overrepresentert i hurtigvoksende byer, noe som kan føre til at den observerte effekten er sterkere enn den virkelige effekten (Duranton, 2016).

For å undersøke om det eksisterer en kausal sammenheng mellom utdanning og befolkningsvekst, utførte Glaeser og Saiz (2004) flere robusthetstester. De oppdaget at utdanningsnivået hadde en innvirkning på befolkningsveksten når de kontrollerte for byspesifikke egenskaper og variasjon innen byer over tid. Resultatet viser at befolkningsveksten i en gitt by øker når flere innbyggere har høyere utdanning. De argumenterte for at forholdet mellom humankapital og befolkningsvekst ikke drives av uobserverbare permanente faktorer som gjør at byer vokser raskere og tiltrekker seg mer utdannede arbeidstakere. Et problem kan være at utdanning og byvekst bestemmes samtidig. For å løse dette simultanitetsproblemet brukte Moretti (2004) en instrumentvariabel-metode. Han fant at resultatene fra IV-estimeringen var ganske like minste kvadrats metode (OLS)-resultatene.

Glaeser (2010) understrekte betydningen av agglomerasjonsøkonomier for økonomisk vekst. Blant annet kan byer med høyere humankapitalnivå forvente positive agglomerasjonseffekter. Byer i USA med høyere tetthet av dyktige, utdannede arbeidstakere er ofte mer produktive. Samlingen av humankapital i byer gir høyere kunnskapsoverføring og økt økonomisk dynamikk. Ringvirkninger fra akkumulering av humankapital fungerer også som positive eksternaliteter for resten av befolkningen som ikke nødvendigvis tar høyere utdanning selv. Flere forskere støtter argumentene til Glaeser (2010). Moretti (2004) fant at en økning i andelen med høyere utdanning i en amerikansk by, ikke bare øker lønnen for de høyere utdannede, men også for dem uten høyere utdanning. 1 prosentpoeng økning i andelen sysselsatte med høyskoleutdanning, øker lønnen for dem som dropper ut av videregående med 1,9 prosentpoeng og med 1,6 prosentpoeng for dem som kun fullfører ungdomsskolen. Duranton (2008) forklarte de samme mekanismene. Han presenterte en modell som forklarer de positive effektene størrelsen på regionen har på humankapital i regionene, og hvordan denne positive effekten bidrar til økt lønn i neste periode for både dem med økt humankapitalnivå, og for dem med lavere humankapitalnivå. Det skyldes at urbanisering og agglomerering bidrar til langsiktig økonomisk vekst utover hva som kan observeres gjennom lønnskurven alene.

Duranton (2016) undersøkte hva som driver befolkningsveksten i colombianske byer mellom 1993 og 2010. Han trakk frem at den høyere befolkningsveksten i enkelte colombianske byer kan knyttes til høyere lønninger. Lønnsfordelene i enkelte byer kan delvis spores tilbake til humankapitalnivå og industrisjokk. Ifølge Duranton (2016) er standard byvekstregresjoner vanligvis basert på byens befolkningsvekst over tid, samt karakteristika fra den initiale perioden. Duranton (2016) mente også at humankapitalnivået bidrar til økte agglomerasjonseffekter. Han mente at det var informativt å se på lønn og humankapital sammen i byvekstregresjoner, men også hver for seg. Fordi utdanning i byer kan føre til agglomerasjonseffekter som kan bety høyere lønn, vil høyere humankapitalnivå i en by etterhvert føre til en tilstrømming av nyankommere.

Duranton og Puga (2013) forklarte sammenhengen mellom entreprenørskap og humankapital. De fokuserte på entreprenører når de snakket om agglomerasjonseffekter, fordi entreprenører ofte har en høyere utdanningsbakgrunn og dermed kan være overrepresentert blant utdannede arbeidstakere. Hvis dette er tilfellet, vil byer med høyere utdanningsnivå være mer entreprenørskapsorienterte, med flere nye firmaer som opprettes, og eksisterende firmaer som vokser raskere. Dette fører til sterkere befolkningsvekst. Chinitz (1961) sammenliknet New

York og Pittsburgh og fant at entreprenørskap og antall småbedrifter er nøkkelfaktorer for å forklare byers utvikling.

2.3 Næringssektorer og agglomerasjonseffekter

Sysselsetting i ulike næringer har vist seg å ha betydelig innvirkning på befolkningsvekst i ulike studier. Disse studiene har analysert hvordan sektorfordeling, næringsklynger og spesialisering påvirker befolkningsveksten. Duranton og Puga (2013) fant at næringsklynger, der bedrifter og organisasjoner innen en bestemt sektor samles i et område, skaper positive agglomerasjonseffekter som fører til økt produktivitet, kunnskapsoverføring og innovasjon. Dette kan tiltrekke seg mer arbeidskraft og bidra til befolkningsvekst.

Glaeser et al. (1992) studerte forskjellen i byvekst mellom byer som består av spesialiserte næringer og byer som består av diversifiserte næringer i USA mellom 1956 og 1987. Diversifisering refererer til en by med en bred rekke næringer, som er mer motstandsdyktig mot økonomiske sjokk ettersom den er mindre avhengig av én enkelt sektor. Spesialisering, derimot, innebærer at en by blir et sentrum for ekspertise innen en næring eller sektor. Det kan føre til agglomerasjonseffekter ved at produktiviteten og veksten innen den spesialiserte sektoren øker, fordi arbeidstakere drar nytte av de respektive kunnskaps- og kompetansebasene. Imidlertid kan for mye spesialisering øke sårbarheten for økonomiske sjokk relatert til den spesifikke næringen. Glaeser et al. (1992) finner en negativ koeffisient foran initiell spesialisering, og en sterk positiv koeffisient foran initiell diversifisering. Disse resultatene har blitt bekreftet av flere studier (Combes, 2000, Cingano og Schivardi, 2004). Mindre regioner har en tendens til å være spesialiserte, mens store regioner er ofte mer diversifiserte (Black og Henderson, 2003).

Duranton (2016) uttrykte bekymring for at den estimerte elastisiteten av lønn med hensyn til befolkning trolig er nedadgående skjev. Dette skyldes at arbeidstakere baserer sine flyttebeslutninger på forventninger om fremtidig lønn, snarere enn lønnen i det initiale året. Et annet problem er at lønn og befolkningsvekst sannsynligvis henger sammen, ettersom befolkningsinnvandring vil redusere lokale lønninger på kort sikt. Duranton (2016) påpekte at det er naturlig å forvente at høyere lønninger er et resultat av økt lokal etterspørsel etter arbeidskraft, og motsatt.

Duranton og Puga (2013) presenterte den monosentriske bymodellen, som ofte tilskrives arbeidet til Alonso (1964), Mills (1967) og Muth (1969), som gir en forenklet fremstilling av byens romlige struktur, der et bysentrum eller et sentralt forretningsdistrikt (CBD) dominerer byens aktiviteter. I denne modellen er all sysselsetting konsentrert i CBD, mens boligene brer seg utover fra sentrum. Innbyggerne pendler til sentrum for å jobbe. Pendlingskostnadene øker med avstanden fra sentrum, og dermed synker tomteprisene og boligtettheten vanligvis etterhvert som man beveger seg lenger bort fra sentrum. Dermed avveier innbyggerne høyere pendlingskostnader mot lavere boligkostnader når de bestemmer seg for hvor de skal bo.

Duranton og Puga (2013) brukte denne monosentriske bymodellen som utgangspunkt for å analysere den romlige organiseringen av byer. De fremhevet hvordan disse nyere modellene kan ivareta et bredere spekter av fenomener som observeres i virkelige byer, til å forklare bystrukturen, for eksempel eksistensen av flere sysselsettingssentre i en by, variasjonen i pendlingsmønstre og rollen til fasiliteter og lokale offentlige goder. De brukte blant annetmodellen til å forklare agglomerasjonseffekter. Ved å anta at lønnen i en by avhenger positivt av befolkningen, vil en befolkningsøkning både ha en positiv og en negativ effekt på innbyggernes samlede nytte. På kort sikt dominerer agglomerasjonseffektene trengselseffektene og nytten øker med bystørrelse. Over tid fører høyere bolig- og pendlingskostnader til at nytten avtar med bystørrelse. De fant altså positiv, men avtagende avkastning av agglomerasjonseffekter på befolkningsvekst.

Businessyrker er typiske yrker som krever høyere utdanning, og er lokalisert der de høyere utdannede er. Duranton og Puga (2004) forklarte at det er en tendens til at høyere utdannede individer konsentrerer seg i bestemte områder, som ofte er større regioner eller metropolregioner. Agglomerasjonseffektene som følge av denne konsentrasjonen kan i følge Glaeser et al. (1992) føre til økt innovasjon og økonomisk vekst, som kan tiltrekke seg enda flere høyt utdannede individer og innovative virksomheter.

Serviceyrker er karakterisert ved lavere lønninger og lavere utdanningskrav. Ifølge enkelte studier, som for eksempel Baum-Snow og Pavan (2013), kan denne typen jobber være mer spredt geografisk, siden de ofte er direkte knyttet til befolkningstettheten og behovene til lokalbefolkningen. Derfor er det ikke nødvendigvis de samme agglomerasjonseffektene av slike yrker som i businessyrker. Med andre ord, det er mulig at en økning i sysselsettingen innen serviceyrker kan være korrelert med lavere befolkningsvekst, særlig hvis disse jobbene ikke

skaper de samme agglomerasjonseffektene eller økonomiske incentiver for folk til å flytte til regionen.

Produksjonssektoryrker har ofte en variert utdanningsbakgrunn, avhengig av den spesifikke jobben. Selv om mange produksjonssektoryrker kanskje ikke krever høyere utdanning, kan de likevel være knyttet til høy produktivitet og innovasjon, spesielt innenfor industri og bygg- og anleggsvirksomhet. Denne typen jobber kan skape positive agglomerasjonseffekter, spesielt i regioner med sterke næringsklynger. Porter (1998) forklarte at når bedrifter i lignende eller relaterte næringer konsentrerer seg geografisk, kan de dra nytte av delte ressurser, som bidrar til positiv befolkningsvekst fordi slike områder er produktive og har positiv økonomisk vekst.

Porter (1998) trakk frem Silicon Valley og Hollywood som de mest kjente samlingene av økonomisk aktivitet. Silicon Valley er et globalt sentrum for høyteknologi og innovasjon. I denne regionen danner teknologiselskaper, entreprenører og utdanningsinstitusjoner et tett nettverk som samhandler og konkurrerer med hverandre.

Hollywood, som er et sentrum for film- og underholdningsindustrien, har høy konsentrasjon av filmskapere, skuespillere, teknisk personale, studioer og andre bransjerelaterte selskaper og fagfolk. Nettverket av konkurrerende og samarbeidende enheter fremmer innovasjon og økonomisk vekst i området. Både i Silicon Valley og Hollywood er det produksjonssektoren og businesssektoren som bidrar til agglomerasjonseffekter og vekst, – både i økonomien og befolkningen. Serviceyrkene oppstår som følge av økt befolkning i disse regionene, men bidrar ikke til agglomerasjonseffekter.

2.4 Bartik-instrumentet

Bartik-instrumentet er en statistisk teknikk som brukes i økonometri for å estimere den kausale effekten av en spesifikk politikk eller et sjokk på et bestemt resultat av interesse. Instrumentet ble introdusert av Bartik (1991), og senere popularisert av Blanchard og Katz (1992). Siden har det blitt mye brukt på ulike områder innen økonomi, deriblant arbeidsmarkeds-, offentlig, internasjonal og regional økonomi. Bartik-instrumentet er utformet for å løse endogenitetsproblemer som oppstår når politikken, eller sjokket av interesse, er korrelert med andre faktorer som påvirker utfallet som studeres. Hvis en region for eksempel innfører et nytt skattefradrag for å oppmuntre bedrifter til å lokalisere seg i regionen, kan det være vanskelig å

avgjøre om skattefradraget isolert er årsak til eventuelle påfølgende endringer i sysselsetting. For å løse dette problemet bruker Bartik-instrumentet variasjonen i politikken eller sjokket på tvers av ulike geografiske områder som en kilde til eksogen variasjon som kun baserer seg på den delen av sjokket som ikke retter seg mot en spesifikk region. Nærmere bestemt er det bygd opp ved interaksjon mellom andel i lokal næring og nasjonal næringsvekstrate.

I over to tiår har instrumentet blitt brukt i empiriske studier (Blanchard og Katz, 1992, Card 2009, Autro et al. 2013), og nylig har et lignende instrument blitt brukt til å undersøke påvirkningen av pandemi-relatert politikk (Granja et al., 2022). Bruken av instrumentet i Granja et. al (2022) baserer seg på antakelsen om at bankfilialandeler før politikken i en gitt region ikke er korrelert med resultatene de studerer. Konstruksjonen til instrumentet er fleksibelt og kan anvendes til flere empiriske studier innenfor økonomi.

Blanchard og Katz (1992) utforsket årsaker til hvorfor visse amerikanske stater i etterkrigstiden opplevde en raskere eller tregere sysselsettingsvekst enn det nasjonale gjennomsnittet. I den estimerte vekstregresjon var den avhengige variabelen forskjellen mellom en regions andel sysselsatte i forhold til arbeidsstyrken. Den regionale vekstregresjonen ble så sammenliknet med vekstregresjonen for hele USA. Den uavhengige variabelen, som de instrumenterte for, var innovasjoner i sysselsetting, antatt å gjenspeile etterspørselen etter arbeidskraft. Ifølge forfatterne hypotese skulle det være mulig å estimere ligningen ved hjelp av OLS hvis deres antagelser holdt. Imidlertid uttrykte de bekymring for at et immigrasjonssjokk kunne påvirke både sysselsatte og arbeidsledige samtidig i et gitt år, noe som ville føre til feilaktig estimering av sysselsettingskoeffisienten. På grunn av dette valgte forfatterne å estimere to forskjellige instrumentvariabler for innovasjoner i sysselsetting. En av de to instrumentvariablene var veksten i sysselsettingen i en stat, predikert av industriens nasjonale vekst, i tråd med Bartik (1991). Blanchard og Katz (1992) argumenterte for at instrumentet var eksogent, og varierte tilstrekkelig til å påvirke etterspørselen etter arbeidskraft. Resultatet fra IV-estimeringen viste seg å gi en noe større sysselsettingskoeffisient enn ved OLS.

Klein og Crafts (2020) benyttet Bartik-instrumentet i sin studie der de undersøkte betydningen av industriell struktur for arbeidsproduktivitetsveksten i amerikanske byers produksjonssektor under den såkalte «andre industrielle revolusjonen». Den avhengige variabelen i deres vekstregresjon var vekstraten for verdiskapning per arbeider i en gitt næring, som en funksjon av spesialisering og diversifisering i en sektor, i en gitt by i det opprinnelige året. Ved bruk av

OLS-estimering, fant Klein og Crafts (2020) sterke bevis for at spesialiserte byer hadde en positiv sammenheng med produktivitetsveksten i produksjonssektoren i det utvalgte settet av byer på tidlig 2000-tallet. Imidlertid var de mer bekymret for at den andre variabelen, diversifisering, kunne være endogen. Det skyldes at det kan være en sammenheng mellom faktorer som påvirker den initielle næringsstrukturen og de generelle forbedringene som påvirker fremtidig industriproduktivitet. For å løse dette potensielle endogenitetsproblemet, valgte Klein og Crafts (2020) å bruke totrinns minste kvadraters metode (2SLS), der de erstattet spesialisering og diversifisering med et Bartik-instrument. Ved å sammenligne resultater fra både OLS-estimeringen og 2SLS, oppnådde de statistisk signifikante resultater ved denne metoden. Endogenitetstestene de utførte ga overbevisende resultater om at spesialisering- og diversifiseringsvariablene er eksogene i de fleste spesifikasjoner med Bartik-instrumentet.

De fleste forskere har brukt Bartik-instrumentet som et instrument i paneldataanalyse. Duranton (2016) brukte det derimot i barroregresjoner for å måle lokale sjokk i etterspørselen etter arbeidskraft. Bartik-instrumentet måler sysselsettingsvekst i byene basert på byenes opprinnelige sektorsammensetning av sysselsetting i samspill med den nasjonale veksten i sektorer. De er beregnet for næringsdefinisjoner både på to-sifret og fire-sifret nivå. Tosifrede næringer er et bredere aggregat av undernæringer, mens fire-sifrede næringer er en mer detaljert definisjon.

2.5 Oppsummering og hypoteser

Studiene vi har presentert belyser hva som driver befolkningsvekst, og trekker i stor grad frem agglomerasjonseffekter som en viktig årsak til hvorfor store geografiske områder innenfor land fortsetter å vokse. Duranton og Puga (2013) trakk frem at næringsklynger og økt humankapitalnivå bidrar til disse agglomerasjonseffektene i amerikanske byer. For Norge, kan vi forvente at større urbane regioner, preget av næringsklynger og høye nivåer av humankapital, vil ha høyere befolkningsvekst.

Business- og produksjonssektoryrker forventes å ha en positiv effekt på befolkningsveksten fordi de inkluderer yrker som typisk krever høyere utdanning, særlig i businesssektoren, og derfor bidrar til agglomerasjonseffekter. I tillegg er flere av de typiske produksjonssektoryrkene lokalisert der mye økonomisk aktivitet er. For eksempel har petroleumsindustrien i Norge, som har sitt hovedsenter i Stavanger/Sandnes-regionen, trolig bidratt til mye økonomisk aktivitet og

jobbskaping, utover de direkte jobbene i olje- og gassnæringen. Serviceyrker har derimot ikke nødvendigvis den samme ringvirkningseffekten på befolkningsvekst, fordi disse yrkene trolig heller oppstår som følge av at det er befolkningsvekst.

Vi baserer vår hypotese om absolutt divergens i norske regioner for perioden 1995-2013 på arbeidet til Rattsø og Stokke (2014). Hypotesen om absolutt divergens innebærer at regioner med høyt initielt befolkningsnivå vokser raskere enn regioner med lavere initielt befolkningsnivå, uavhengig av andre initialbetingelser. Når vi senere utvider analysen til å se på noen av de underliggende faktorene til agglomerasjonseffekter, – nemlig humankapital og sysselsetting i ulike næringssektorer, i tråd med Duranton (2016), forventer vi at disse kan bidra til å forklare befolkningsvekst. Vår hypotese for de betingede barroregresjonene er at økt humankapitalnivå, samt sysselsetting i produksjons- og businesssektoren bidrar positivt til befolkningsvekst, og derfor kan være med på å forklare divergens i norske regioner. Vi forventer at servicesektor-yrker har en negativ effekt på befolkningsveksten fordi de ikke bidrar i like stor grad til agglomerasjonseffektene. Hypotesen om Bartik-instrumenter for de tre ulike næringssektorene går ut på at fortegnene er de samme ved å bruke dette målet på sysselsetting, men at effekten er mer kausal.

3 Datamateriale

Datamaterialet som brukes i analysen er i hovedsak hentet fra Kommunedatabasen og Statistisk Sentralbyrå (SSB). All data er hentet på kommunenivå, men aggregert opp i henhold til økonomiske regioner fra SSB (Hustoft et al., 1999). Denne inndelingen er gjort for å oppfylle EUs krav om NUTS-4-nivå. Regionene må bestå av hele, sammenhengende kommuner, og kan ikke krysse fylkesgrenser. De er definert utfra pendlingsdata fra SSB, og ved å aggregere dem opp fra kommunenivå tar man hensyn til at mye økonomisk aktivitet krysser kommunegrenser. På denne måten avgrenses lokale økonomier.

Befolkningsveksten i byer eller regioner kan være enklere å studere enn vekstprosesser for hele land. Den store litteraturen om nasjonal vekst som bygger på Barros (1991) arbeid, er preget av grunnleggende data- og landheterogenitetsproblemer som er mye mindre viktige i studier av byer eller regioner innenfor et land enn i studier av forskjeller mellom land. Det skyldes at bedrifter og husholdninger i ulike regioner innenfor et spesifikt land har mer lik tilgang til den samme teknologien og har likere preferanser og kultur. Dessuten deler regionene samme regjering og derfor har likere institusjonelle oppsett og rettssystemer. Videre er nasjonale vekstregresjoner preget av endogenitetsproblemer som ofte er svært vanskelig å håndtere i tverrnasjonal sammenheng (Durlauf et al., 2005). Den relative homogeniteten gjør det lettere å undersøke årsaker til divergens i regioner innad i et land. Å se på tverrsnitt av regioner innenfor land bør kunne forbedre muligheten for å finne løsninger på identifikasjonsproblemer.

Et problem ved å aggregere kommunedata til større regioner, er kommunesammenslåinger. Perioden 1995-2013 var preget av sterk økonomisk vekst trolig på grunn av forsinkede reaksjoner på store oljefunn, samt forsinkede reaksjoner på EØS-avtalen. EØS-avtalen som trådte i verk 1. januar 1994, gir borgere i Norge samme rettigheter og plikter som i andre EØS-land, når det gjelder varehandel, investeringer, bank- og forsikring, kjøp og salg av tjenester og rett til å ta arbeid, studere og bo i EØS-land (regjeringen.no, 2021). Kommuner i Norge fra 2014 og til i dag har vært preget av sammenslåinger, som gjør det vanskelig å aggregere dem opp til regionnivå. I tillegg til mangelfulle data fra perioden før 1995, har vi derfor valgt kun å se på betingende barroregresjoner for perioden 1995 til 2013, men vil undersøke ubetinget divergens helt tilbake til 1910.

Et annet problem med regioninndelingen til Hustoft et al. (1999), er at mye økonomisk aktivitet faktisk krysser fylkesgrenser. For eksempel er det stor forskjell på antall som jobber i Oslo, og de som jobber og bor i Oslo. Det skyldes mye innpendling fra kommunene rundt som tilhører andre fylker (f.eks. Lillestrøm eller Bærum i Viken fylke). Fordi kommunene rundt Oslo er store pendlerkommuner, kunne vi valgt å aggregere kommunene opp til 46 storregioner i henhold til Bhuller (2009). Denne regioninndelingen tillater økonomisk aktivitet å krysse fylkesgrenser. Men fordi det er større forskjeller innad i storregionene som gjør det utfordrende å isolere viktige lokale faktorer som påvirker befolkningsdivergensen, velger vi å forholde oss til de 90 økonomiske regionene.

Duranton (2016) brukte en rekke determinanter for å forklare hva som drev befolkningsvekst i colombianske byer. Denne grundige tilnærmingen er av betydelig inspirasjon for vårt arbeid, og alle variablene vi har anvendt i vår egen analyse er inspirert av dem han tok i bruk. Imidlertid er det viktig å merke seg at Duranton (2016) sitt arbeid var mer omfattende, og involverte et langt større antall variabler enn det vi har kunnet medta i vår analyse. For det første er tilgangen til tilsvarende data på kommunenivå i Norge begrenset, noe som gjør det vanskelig å omgjøre dataene til å gjelde for regioner. For det andre, kan for mange kontrollvariabler i en modell føre til overtilpasning av modellen. Problemer med overtilpasning er ytterligere drøftet i delkapittel 4.6.6.

3.1 Avhengig variabel og befolkningstall

Avhengig variabel er gjennomsnittlig, årlig vekstrate i befolkningen på logaritmisk form, og er konstruert på følgende måte:

$$\frac{1}{T} \log \left(\frac{p_{rT}}{p_{r0}} \right)$$

Her er T antall år. p_{rT} er befolkningsstørrelsen i sluttåret i region r , og p_{r0} er befolkningsstørrelsen i startåret i region r . I utgangspunktet studeres perioden 1995-2013, men også kortere og lengre perioder vil bli studert for å se på robustheten av analysen. Målingene for befolkningen er gjort 1. januar samme år. Folketellingen for befolkningen i 2013, er dermed registrert 1. januar 2014.

Den deskriptive statistikken i tabell 1 viser at befolkningen i landet har økt. Regionene har i gjennomsnitt blitt større. En gjennomsnittlig region har økt fra 25918 innbyggere i 1910, til

48300 innbyggere i 1995 og 56266 innbyggere i 2013. Det er også tydelig at forskjellen på størrelsene har blitt større, med lavere minimumsverdi og høyere maksimumsverdi for befolkningen i 2013 enn i 1930, 1954, 1969 og 1995. Det gir en indikasjon på større spredning i befolkningen i 2013 enn tidligere perioder.

Den gjennomsnittlige årlige befolkningsveksten for hele perioden 1995-2013 er på 0,35 prosentpoeng. 31 regioner har hatt en negativ befolkningsvekst, med Vadsø helt i bunnen med gjennomsnittlig, årlig befolkningsvekst på nesten $-0,8$ prosentpoeng. Ullensaker/Eidsvoll er den raskest voksende regionen med gjennomsnittlig, årlig befolkningsvekst på 2,2 prosentpoeng. Standardavviket tilsier at det er store forskjeller i veksten i regionene.

Tabell 1: Deskriptiv statistikk

Variabler	Obs	Gj.snitt	Std.avvik	Min	Max
Befolkning1910 90		25 918	34 538	3570	273 229
Befolkning1930 90		30 647	42 891	6464	342 285
Befolkning1954 90		37 159	55 162	6956	444 072
Befolkning1969 90		42 404	63 152	7082	488 377
Befolkning1995 90		48 300	69 255	5981	483 401
Befolkning2013 90		56 266	88 524	5271	623 966
Vekst ₁₉₉₅₋₂₀₁₃	90	0.00350	0.006512	-0.00798	0.021629

3.2 Forklaringsvariabel

Analysen går i hovedsak ut på å beskrive hvordan befolkningsstrømmene i Norge har beveget seg. Hovedforklaringsvariabelen er logaritmisk tilnærming til initielt befolkningsnivå. I utgangspunktet ser vi på logaritmisk initielt befolkningsnivå i 1995 (*log befolkning 95*). For delperioder og lengre periode, vil den logaritmiske tilnærmingen til befolkningsnivået for startåret i perioden være forklaringsvariabel. Ved å kjøre ubetingede regresjoner av forklaringsvariabelen på veksten, kan vi se om det er divergens eller konvergens i befolkningen i regionene.

3.3 Kontrollvariabler

Et sett kontrollvariabler inkluderes for å forklare mer av variasjonen i barroregresjonene. Som mål på humankapitalnivået, brukes andelen av befolkningen i arbeidsdyktig alder i hver region som er sysselsatt med minimum 1 års høyere utdanning i 1995. Arbeidsdyktig alder gjelder for dem mellom 20 og 67 år som er bosatt i regionen i vår oppgave. Variabelen er konstruert på følgende måte:

$$\text{andel utdannet}_r = \frac{\text{unikort95}_r + \text{unilang95}_r}{\text{arbeidsdyktige95}_r},$$

hvor unikort95 er sysselsatte med 1-4 års universitet- eller høyskolebakgrunn i region r, og unilang95 er sysselsatte med minimum 5 års universitet- eller høyskolebakgrunn i region r. En forutsetning for at estimatene skal bli riktig, er at alle tallene gjelder for dem som er bosatt i regionen. Tabell 2 viser deskriptiv statistikk for kontrollvariablene våre. I gjennomsnitt er ca. 14 prosent av arbeidsdyktig befolkning høyere utdannet i norske regioner, med et standardavvik på ca. 3,5 prosent.

For å sjekke robustheten av utdanningsvariabelen, prøver vi også en annen definisjon hvor vi deler på totalt antall sysselsatte i regionen i 1995 som i Duranton (2016). Denne måten å måle humankapital på er populær i urban vekstteori, og brukes blant annet også i Glaeser et al. (1995), Simon og Nardinelli (2002) og Glaeser og Saiz (2004). Med dataene vi har, er andelen av befolkningen i arbeidsdyktig alder som er sysselsatt med høyere utdanning, et foretrukket mål på humankapital i denne oppgaven.

I utvidelsen av modellen vil vi skille mellom tre ulike næringssektorer som i Duranton (2016). Hentet fra initialåret, er variablene for produksjon-, business-, og servicesektoren definert ved å dele sysselsatte i de ulike sektorene (s) i regionen på sysselsettingen i regionen. Da får vi andelen som jobber i hver sektor. Hvilke næringer som inngår i hver sektor, er til dels bestemt ut fra egen intuisjon, samt nummereringen til SSB.¹ Variablene er definert slik:

$$\text{andelsektor}_{r_s} = \frac{\text{sysselsatte i sektor95}_{r_s}}{\text{tot antall sysselsatte95}_{r_s}}$$

¹ Se A1 i appendiks for hva som inngår i de tre næringssektorene i appendiks.

Tabell 2 viser deskriptiv statistikk for våre kontrollvariabler. I 1995 var servicesektoren desidert den sektoren med flest sysselsatte. I gjennomsnitt jobbet 66 prosent av befolkningen som var i arbeidsdyktig i alder, i serviceyrker i norske regioner. Lavere standardavvik enn gjennomsnittet for alle kontrollvariablene tilsier at de fleste regioner var konsentrert rundt gjennomsnittet i 1995. Altså har andelene ikke variert så mye fra region til region. Dette kan tyde på en viss homogenitet i økonomisk aktivitet og sektorfordeling av arbeidsstyrken på tvers av Norges regioner i 1995.

Tabell 2: deskriptiv statistikk for kontrollvariabler

Variabler	Obs	Gj.snitt	Std.avvik	Min	Max
Andel høyere utdannet 90		0.1423	0.0353	0.0938	0.3282
Andel i produksjon	90	0.2760	0.0782	0.1385	0.4594
Andel i business	90	0.0615	0.0281	0.0254	0.1923
Andel i service	90	0.6625	0.0735	0.5042	0.8210

4 Metode

4.1 Barroregresjoner

Barroregresjoner er en vanlig metode for å undersøke om det har vært konvergens eller divergens i underliggende data. Metoden ble presentert av Barro og Sala-i-Martin (1992) for å studere sammenhengen mellom økonomisk vekst og initielt inntektsnivå mellom land, og undersøke om fattige land vokser raskere enn rike land, - og dermed om land konvergerer i inntekt over tid. Metoden har i senere tid blitt populær til å studere demografisk dynamikk. Den estimerte regresjonslinjen er lineær, og beregnes ved å bruke minste kvadrats metode (OLS). Barro og Sala-i Martin (1992) introduserte følgende likning i den absolutte barroregresjonen:

$$\frac{1}{T} \log \left(\frac{p_{rT}}{p_{r0}} \right) = \beta_0 + \beta_1 \log(p_{r0}) + u_r \quad (1)$$

I vår analyse er årlig vekstrate beregnet for perioden 1995-2013 for region r . I likningen er den avhengige variabelen gjennomsnittlig, årlig prosentvekst i befolkningen. Logaritmiske verdier brukes for å gjøre analysen mindre sensitiv for utliggere (Woolridge, 2015, s. 171-173). β_1 representerer vår interessevariabel, nemlig log befolkning i startåret av analysen. En negativ β_1 indikerer absolutt konvergens, dvs. at regioner med lav befolkning i 1995 opplever høyere befolkningsvekst enn regioner med høy befolkning i 1995, uavhengig av andre faktorer. Hvis dette stemmer, vil regioner over tid nå en felles likevekt hvor regionene er av samme befolkningsstørrelse. Regioner med høy initiell befolkning er forventet å ha høyest befolkningsvekst. Argumentet støttes av FN (2019) som finner bevis for raskere økning i befolkning i urbane områder i verden, sammenliknet med rurale områder. Vi forventer derfor å finne befolkningsdivergens i analysen vår.

Quah (1993) er blant dem som har vært kritiske mot barroregresjoner. Han påpekte at fortegnet til regresjonskoeffisienten foran det initiale nivået ikke forteller noe om hvorvidt det forekommer konvergens eller divergens. Dette illustrerte Quah (1993) ved å anvende Galton's feiltakelsesteori, også kalt regresjoner mot snittet. Galton observerte at høye fedre ofte fikk sønner som var kortere enn dem, noe som var i strid med den observerte kontinuerlige spredningen. Quah (1993) argumenterte for at Galton's teori kan anvendes på vekstregresjoner. Han hevdet at koeffisienten for den initiale tilstanden, enten den er positiv eller negativ, kan antyde om tverrsnittfordelingen vil konvergere eller divergere. Men pågående forstyrrelser kan

trekke i motsatt retning, slik at koeffisienten for den initielle tilstanden ikke nødvendigvis trenger å være positiv eller negativ.

Artikkelen fokuserer primært på konvergenhypotesen gjennom Galtons feiltagelsesteori, men Quah (1993) bemerket at det samme problemet også gjelder for divergens. Han konkluderte med at det er vanskelig å trekke dynamiske implikasjoner fra tverrsnittsbevis. Kritikken til Quah (1993) impliserer at barroregresjoner antar at en regions befolkningsnivå i en periode er uavhengig av befolkningens nivå i forrige periode, - en antagelse som er urealistisk, men som benyttes for å illustrere et poeng.

Kritikken mot barroregresjoner gjør at kernel-tetthetsfunksjoner også estimeres for bedre å kunne studere dynamikken i data. Quahs (1993) arbeid understreker behovet for en kritisk gjennomgang av barroregresjoner og viktigheten av å være klar over potensielle svakheter i modellene som benyttes. Han går over på markov-kjedeanalyse, som i vårt tilfelle kunne vært utført for å studere hvordan fordelingen av befolkningen endres over tid, og hvordan dynamikken mellom regionene er. En slik metode er omfattende, og i likhet med Duranton (2016) ønsker vi heller å se på faktorer som påvirker forskjellene i befolkningsvekst.

I utvidelsen av de absolutte barroregresjonene vil vi inkludere andelen sysselsatte med høyere utdanningsnivå, samt skille mellom andelen sysselsatte i de tre næringssektorene. Når vi utvider den absolutte barroregresjonen til å inkludere et sett kontrollvariabler, ser modellen slik ut:

$$\frac{1}{T} \log \left(\frac{p_{rT}}{p_{r0}} \right) = \beta_0 + \beta_1 \log(p_{r0}) + \alpha X_{r0} + u_r,$$

hvor X_{r0} er et sett kontrollvariabler, med tilhørende koeffisient α . Dette er likningen for den betingede barroregresjonen. Dersom det initielle befolkningsnivået er uavhengig av utdanningsnivå og sysselsetting, vil inkludering av disse variablene i regresjonen ikke påvirke estimatet av β_1 . Dersom det initielle befolkningsnivået derimot er avhengig av utdanningsnivå og sysselsetting, vil en positiv β_1 indikere at regionenes divergens forklares av disse faktorene. Derfor gir betingede barroregresjoner oss muligheten til å vurdere hvordan endringer i befolkningsnivået påvirkes av disse andre faktorene, samtidig som vi kontrollerer for initieell befolkning.

4.2 Spline-regresjoner

Spline-regresjoner er en utvidelse av barroregresjoner, der man kan studere sammenhenger mellom vekstrate og initielt befolkningsnivå når man tar hensyn til det initielle nivået. Ved å dele logaritmisk befolkningsnivå i 1995 inn i tre intervaller med like mange observasjoner, får man en egen koeffisient for hver av de tre undergruppene. Med ubetingede spline-regresjoner i hver av de tre intervallene kan en se om det har vært divergens innad i ulike regionsstørrelser. Fra Dupont (2013) gir dette følgende likning:

$$\frac{1}{T} \log \left(\frac{p_{rT}}{p_{r0}} \right) = \beta_0 + \beta_1 \log(p_{r0}) + \beta_2 \log(p_{r0}) + \beta_3(p_{r0}) + u_r$$

hvor:

$$\frac{\partial \frac{1}{T} \log \left(\frac{p_{rT}}{p_{r0}} \right)}{\partial p_{r0}} = \begin{cases} \beta_1 & \text{hvis } p_{r0} < x_1 \\ \beta_2 & \text{hvis } x_1 \leq p_{r0} \leq x_2 \\ \beta_3 & \text{hvis } p_{r0} > x_2 \end{cases}$$

hvor x_1 og x_2 er grensene som deler $\log(p_{r0})$ inn i ulike grupper etter befolkningsnivå. β_1 ser på divergens for gruppen med lavest befolkningsnivå, β_2 ser på divergens innad i de mellomstore regionene, og β_3 ser på divergens innad i de største regionene. En positiv koeffisient for en eller flere av gruppene betyr at det er befolkningsdivergens innad i gruppen(e). Hvis koeffisientene foran initielt befolkningsnivå i de tre gruppene er lavere enn ved den generelle barroregresjonen for alle 90 regioner, betyr det at divergensprosessen er langsommere for disse gruppene. Langsommere divergensprosess kan oppstå innad i de ulike gruppene fordi de er likere i størrelse enn alle 90 regionene i sin helhet. I hvert intervall kan regionene være likere i karakteristika. For eksempel hvis vi tenker oss at regionene i det øverste intervallet består av relativt høyt utdannede mennesker, kan vi tro at disse er mer mobile i denne gruppa, og dermed kan dempe veksten. I tillegg kan regionene være relativt like når det gjelder institusjonelle forhold, og samling av bedrifter, som også demper veksten.

Med andre ord, ved å se på gruppene separat, tar man hensyn til at regionene innenfor hvert intervall er mer homogene, noe som kan resultere i langsommere divergensprosess sammenliknet med den generelle barroregresjonen for alle 90 regionene. Dette gir et mer detaljert bilde av befolkningsdynamikken i de ulike gruppene av regioner.

4.3 Kernel-tettheter

Kernel-tetthetsestimering (KDE) er en populær ikke-parametrisk metode for å estimere sannsynlighetstetthetsfunksjonen til en tilfeldig variabel basert på et sett av observasjoner (Parzen, 1962). Vi finner et estimat for sannsynlighetstettheten $p(x)$ for utvalg n :

$$\hat{p}(x) = \frac{1}{nh} \sum_{r=1}^n K\left(\frac{x - x_r}{h}\right),$$

hvor $K(x)$ er kernel-funksjonen. Denne funksjonen er jevn og symmetrisk slik som Gaussian, med en båndbredde $h > 0$. I hovedsak jevner KDE ut hvert datapunkt x_r til små tetthetsbuler, og deretter summeres alle disse små bulene for å oppnå den endelige tetthetsestimeringen. Arealet under Kernel-funksjonen må være 1, altså: $\int k(x)dx = 1$. Med en Gaussisk normalfordeling, vil formen på kernel-funksjonen være, gitt av Zucchini (2003):

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Valget av båndbredde h spiller en viktig rolle for kvaliteten av KDE. Når h er liten blir det en «undersmoothing» av tetthetsfunksjonen. Utjevningen er for liten, slik at estimeringen blir mer følsom for tilfeldige variasjoner i dataene. Resultatet blir en «støyende» tetthetsestimering som kan inneholde falske strukturer og topper som egentlig bare er forårsaket av tilfeldigheter i dataene. På den andre siden, når h er stor, har vi en situasjon med «oversmoothing» - utjevningen er for stor. Dette fører til at estimeringen blir mindre følsom for faktiske strukturer i dataene, og resultatet blir en for glatt tetthetsestimering som kan skjule viktige trekk ved de underliggende dataene.

Valget av riktig båndbredde er viktig for å få en nøyaktig, informativ tetthetsestimering. «Silvermans rule of thumb» (Silverman, 1986) er en tommelfingerregel for å velge en passende båndbredde i kernel-tetthetsestimeringen, og er som følger:

$$h = 1.06\sigma n^{-0.2},$$

hvor σ er standardavviket. Regelen er basert på antagelsen om at dataene er normalfordelte.

Når vi studerer relativ befolkningsstørrelse i norske regioner, kan kernel-tetthetsestimering brukes til å utforske fordelingen av befolkningsstørrelser, oppdage eventuelle mønstre og avdekke mulige klynger av regioner med liknende befolkningsstørrelser. Hvis befolkningen øker jevnt over hele spekteret av regioner, vil tetthetsfunksjonene for 1995 og 2013 ha omtrent samme form, men 2013-funksjonen vil være forskjøvet mot høyre, noe som indikerer økte befolkningstall i regionene. Hvis det har skjedd divergens i befolkningsveksten, med noen regioner som opplever betydelig vekst, og andre som opplever liten, ingen eller negativ vekst, vil tetthetsfunksjonene vise en bredere spredning. Hvis det har skjedd konvergens i befolkningsveksten, der tidligere små regioner vokser raskere enn større regioner, vil tetthetsfunksjonene vise en innsnevring av spredningen og en tendens mot en mer ensartet fordeling av befolkningsstørrelser. Hvis befolkningsveksten har vært konsentrert i spesifikke regioner, vil tetthetsfunksjonene vise en økning i tetthet rundt de berørte regionene, noe som indikerer at befolkningsveksten har vært ujevn og konsentrert seg på bestemte geografiske områder.

4.4 Lokal polynomglatting (ikke-lineært kernelplott)

Lokal polynomglatting er en annen ikke-parametrisk metode for å tilpasse en glatt kurve til et sett med observerte data. Metoden fungerer ved å tilpasse polynomfunksjoner til små deler av dataene i stedet for å prøve og tilpasse en enkelt funksjon til alle dataene. Dette gjør at kurven kan tilpasse seg lokale trender og fange opp svingninger i dataene som en global modell ikke kan. Hentet fra Fan og Gijbels (1996) lar vi $\{(X_r, Y_r)\}_{r=1}^n$ være par av datapunkter slik at $Y_r = m(X_r) + \sigma(X_r)\epsilon_r$, hvor restleddet ϵ_r har forventning lik 0 og varians lik 1. m og ϵ er glattingsparametere, hvor $m(X_0)$ estimeres ved hjelp av et polynom av orden P på observasjonene (X_r, Y_r) . Ordenen på polynomet kan ha en standardverdi på null, som innebærer gjennomsnittsglatting. Når polynomets orden er null, tilpasser metoden en konstant funksjon til datapunktene, noe som resulterer i en lokal konstant tilpasning. Båndbredde-verdien følger også her «Silvermans rule of thumb». Standard kernelfunksjoner for lokal polynomglatting er ofte av Epanechnikov-formen, og benyttes i denne metoden.

Ved å anvende lokal polynomglatting kan vi evaluere korrelasjonen mellom initielt befolkningsnivå og befolkningsvekst for regioner mellom 1995 og 2013. Dersom det er en positiv korrelasjon, vil regresjonslinjen vise en positiv stigning. Dette indikerer at regioner med høyere befolkningsstørrelse i 1995 har opplevd høyere befolkningsvekst i perioden. På en annen

side, hvis det er en negativ korrelasjon, vil regresjonslinjen vise en negativ stigning. Dette indikerer at regioner med lavere befolkningsstørrelse i 1995 har hatt høyere befolkningsvekst i samme periode. Hvis det ikke er noen tydelig korrelasjon mellom initielt befolkningsnivå og befolkningsvekst, vil regresjonslinjen være relativt flat. Dette betyr at det ikke er noen systematisk sammenheng mellom befolkningsstørrelse i 1995 og befolkningsvekst mellom 1995 og 2013.

Lokal polynomglatting kan avsløre ikke-lineære forhold mellom initielt befolkningsnivå og befolkningsvekst, noe som kan gi et tydelig bilde på hvordan korrelasjonen endrer seg over ulike regionsstørrelser.

4.5 Bartik-instrumenter

Bartik-instrumentene, introdusert av Bartik (1991), er et annet mål på sysselsetting i de ulike sektorene som vi ønsker å konstruere for en robusthetssjekk av resultatene vi får med *andel sektor* som kontrollvariabel. Instrumentene brukes i Klein og Crafts (2020) som instrumentvariabler for sysselsetting i diversifiserte og spesialiserte næringer i paneldatanalyse for å kunne estimere effekten av disse sysselsettingsvariablene på verdiskapning per arbeider. I likhet med Duranton (2016), vil vi konsentrere oss om effekten av disse instrumentene på befolkningsveksten, når befolkningsvekst ofte blir brukt som forklaringsvariabel for noe annet. Bartik-instrumentet konstruerer en eksogen kilde til variasjon i sysselsettingen i en region basert på nasjonal sysselsettingsvekst på sektornivå. Ved å gjøre dette isoleres den eksogene variasjonen i lokal arbeidskraftetterspørsel som kun kommer fra nasjonale sjokk i hver sektor, og fjerner dermed potensiell endogen lokal arbeidskraftetterspørsel som kan drive regional sysselsetting. I Klein og Crafts (2020) hviler eksogeniteten til Bartik-instrumentene på antagelsen om at nasjonal sysselsettingsvekst på næringsnivå ikke påvirker produktivitetsveksten i regionnæringen.

Variabelen konstrueres i likhet med Klein og Crafts (2020) på følgende måte:

$$l_{rst+m}^{predikert} = l_{rst} \left(\frac{l_{st+m}}{l_{st}} \right),$$

hvor l_{rst} er sektorandeler i region r, i sektor s på tidspunkt t (t=1970), l_{st+m} er sysselsatte i sektor s på tidspunkt t+m (t+m=1990) og l_{st} er sysselsatte i sektor s på tidspunkt t.

Sysselsettingsvekstraten i sektor s i region r på tidspunkt i perioden 1970-1990 ser slik ut:

$$g_{rst,t+m} = g_{st,t+m} + \tilde{g}_{rst,t+m},$$

hvor $g_{st,t+m}$ er nasjonale sysselsettingsvekstrater i sektoren som utgjør den systematiske komponenten av sysselsettingsvekstraten. $\tilde{g}_{rst,t+m}$ er den idiosynkratiske vekstraten i en sektor som retter seg mot en spesifikk region. Et nasjonalt arbeidstilbudssjokk kan påvirke den idiosynkratiske vekstraten ved at lokale sysselsettingsmønstre ikke er tilfeldig. Sjokket kan påvirke regioner forskjellig avhengig av lokale forhold. Gauss Markov-antakelsene sier at det idiosynkratiske leddet i en modell skal ha en forventet verdi på 0. Når det idiosynkratiske leddet fjernes, tar man hensyn til uobserverbare, regionspesifikke faktorer som påvirker sysselsettingsveksten og vi får at predikert sysselsetting baserer seg på eksogeniteten av andelene i sektoren nasjonalt, altså: $\left(\frac{l_{st+m}}{l_{st}}\right) = g_{st,t+m}$.

Et eksempel på et arbeidstilbudssjokk som påvirket arbeidskraftetterspørselen i perioden 1970-1990 var inntoget av kvinner i arbeidsmarkedet som bidro til vekst i sysselsettingen på slutten av 70-tallet og begynnelsen av 80-tallet. I 1980 hadde 73 prosent av kvinner i alderen 20-54 år inntektsgivende arbeid, mot bare 44 prosent ti år tidligere (Kilden, 2005). Denne økningen av kvinner i arbeid skyldtes mange ting, blant annet kvinnefrigjøring og fremveksten av likestillingsideologi. Tilbudet av kvinnelig arbeidskraft økte som følge av dette. Arbeidstilbudssjokket bidro til økt sysselsettingsvekst fordi etterspørselen etter arbeidskraft også økte i de påfølgende årene som følge av økt befolkningsvekst i de regionene tilbudet rettet seg mot. En stor del av sysselsettingsøkningen skjedde i offentlig sektor, blant annet i kommunale og fylkeskommunale virksomheter innenfor helse- og utdanningsyrker. Ved å ta bort den delen av inntoget av kvinner på arbeidsmarkedet som retter seg mot spesifikke regioner, får man at lokal sysselsettingsvekst i de spesifikke næringene i regionene kun drives av den delen av inntoget av kvinner på arbeidsmarkedet som drives av det nasjonale sjokket. På denne måten fjernes den regionspesifikke delen av variasjonen i lokal arbeidskraftetterspørsel som driver befolkningsvekst i regionen, og gir mulighet til å isolere den kausale sammenhengen.

Det er en positiv sammenheng mellom befolkningsvekst og arbeidstilbud. Et problem er at befolkningsvekst ikke bare bidrar til økt etterspørsel etter arbeidskraft, men det driver også tilbudet av arbeidskraft. Dermed kan effekten av etterspørselen etter arbeidskraft på befolkningsvekst bli overestimert. Lokal arbeidskraftetterspørsel som påvirker veksten, isoleres dermed ved å fjerne det lokale arbeidstilbudssjokket. Variasjonen i befolkningsveksten er på regionnivå, og den eneste komponenten av Bartik-instrumentene som varierer på lokalt nivå er sektorandelene i 1970. Identifikasjonen av variabelen er best angitt i form av nasjonale vekstrater (Goldsmith-Pinkham et al., 2020).

Bartik-instrumentet er konstruert fra en periode før vår studieperiode for alle tre næringssektorer for å sikre at den er mest mulig uavhengig av variablene vi studerer. Hvis variabelen hadde vært konstruert for vår periode, ville variabelen lidd av toveisendogenitet. La oss se på hva som hadde skjedd hvis vi konstruerte Bartik-instrumentet for produksjonssektoren fra 1995 til 2013. Telleren i nasjonale sysselsettingsvekstrater ville i så fall vært sysselsettingen i produksjonssektoren i 2013. Simultanitetsproblemet kan da oppstå fordi: En økning i sysselsatte i produksjonssektoren i 2013 vil påvirke befolkningsveksten positivt. Dersom variablene påvirker hverandre, ville denne økningen i befolkningsvekst ha medført en ytterligere økning i sysselsettingen i produksjonssektoren i 2013. Dette vil i så fall gi en overestimert effekt av den sanne effekten av Bartik-instrumentet (for 1995-2013) på befolkningsveksten. Simultanitetsproblemet adresseres derfor i vår oppgave ved å konstruere Bartik-instrumentene for en periode før vår studieperiode.

Den største motivasjonen for å bruke Bartik-instrumentene som et annet mål på sysselsetting i de tre sektorene, er først og fremst at instrumentene muligens er mer eksogene enn *andel sektor*-variablene våre. For det første er Bartik-instrumentene uavhengig av lokale forhold som gjør at sysselsetting kan være endogent bestemt. For det andre er de vanlige *andel sektor*-variablene trolig mer endogene enn Bartik-instrumentene fordi restleddet i modellen kan inneholde uobserverbare faktorer som er korrelert med variabelen, og som påvirker lokal sysselsetting. Fordi dette «lokale» uobserverbare leddet tas bort i Bartik-instrumenter, og fordi det er konstruert fra en periode før vår studieperiode, kan trolig Bartik-instrumentet bidra til mer eksogen variasjon i sysselsettingen i de tre sektorene. For eksempel kan Bartik-instrumentet for produksjonssektoren i Stavanger/Sandnes regionen, som regnes som «oljehovedstaden», være basert på nasjonale trender i petroleumsindustrien, ikke lokale forhold i regionen.

4.6 Økonometriske utfordringer

Dersom en variabel i en regresjonsmodell er korrelert med restleddet i modellen, vil denne variabelen tolkes som endogen. Dette er et brudd på den såkalte eksogene variabelforutsetningen, som er en viktig forutsetning for at OLS skal gi forventningsrette estimater (Wooldrige, 2015, s. 76-77). Det er hovedsakelig tre kilder til endogenitet: Simultanitet, utelatte variable og målefeil.

4.6.1 Simultanitet

Simultanitetsproblemet oppstår når en eller flere av kontrollvariablene bestemmes på samme tidspunkt som den avhengige variabelen (Wooldrige, 2015, s. 499-504). Hvis for eksempel *andel utdannet* bestemmer *veksten* og *veksten* bestemmer *andel utdannet* samtidig, blir det en overestimering av effekten av *andel utdannet* på *veksten*. Fordi forklaringsvariabelen og kontrollvariablene våre er hentet ut fra startåret, reduserer vi sannsynligheten for simultanitetsproblemer.

4.6.2 Utelatte variable

Den ubetingede barroregresjonen er en indikasjon på om det har vært divergens eller ikke, men *log befolkning 95* kan lide av utelatte variableskjevheter (Wooldrige, 2015, s. 78-81). Når vi utelater humankapital som kontrollvariabel, er det grunn til å tro at effekten av initielt befolkningsnivå i 1995 er underestimert i den ubetingede barroregresjonen. Dersom variabelen for humankapital er korrelert med *log befolkning 95* og har en positiv effekt på befolkningsveksten, vil utelatelse av denne medføre en overestimering av *log befolkning 95* på befolkningsveksten.

4.6.3 Målefeil

Målefeil er feil som oppstår når den faktiske verdien til en variabel avviker fra den observerte verdien (Wooldrige, 2015, s. 287-292). Slike feil kan skje når dataene som brukes i analysen er feil eller mangelfulle, og kan føre til at en feilestimert forklaringsvariabel blir skjev mot null ved bruk av OLS-estimatoren. Alle data er hentet fra SSB og Kommunedatabasen som er anerkjente og seriøse miljøer. Vi antar derfor at dataene er korrekt, og at målefeil ikke vil påvirke analysen.

4.6.4 Multikollinearitet

Multikollinearitet betyr at det eksisterer høy, men ikke perfekt korrelasjon mellom to eller flere uavhengige variabler (Wooldridge, 2015, s. 83-86). For eksempel ser vi av korrelasjonsmatrise A5.1 i appendiks at korrelasjonen mellom *Andel business* og *andel utdannet* er nesten 0,83. Denne høye korrelasjonen kan føre til problemer i analysen, fordi det kan medføre høyere varians og mindre nøyaktige estimater. Multikollinearitet gjør det vanskelig å skille ut den faktiske påvirkningen av hver uavhengige variabel på avhengig variabel.

VIF-test kan gjennomføres for å undersøke om det er noe multikollinearitet i de estimerte regresjonene (Wooldridge, 2015, s. 86). VIF oppnås ved denne formelen: $VIF = 1/(1 - R^2)$. Tommelfingerregelen for VIF-tester sier at en verdi over 10 er for høy. Fordi R^2 er godt under 0.9 i alle våre estimerte regresjoner, vil VIF-testene aldri bli over 10. Multikollinearitet kan fortsatt være et problem til tross for lav R^2 .

4.6.5 Heteroskedastisitet og seriekorrelasjon

Siste leddet i likning (1) er restleddet i regresjonen. For at estimatorene i en regresjon skal være forventningsrette, må restleddet være uavhengig av forklaringsvariablene (Wooldridge, 2015, s. 60-104). White-test for å teste for heteroskedastisitet for alle regresjoner, gir signifikante resultater for å beholde nullhypotesen om homoskedastisitet. Til tross for manglende heteroskedastisitet i dataene, vil vi bruke robuste standardavvik for å beskytte mot potensielle feilspesifikasjoner og redusere innflytelsen av ekstreme observasjoner (utliggere), som gir mer pålitelige konfidensintervaller og hypotesetester i vår analyse (Cameron og Trivedi, 2005, s. 78-81).

Robuste standardavvik tar også høyde for seriekorrelasjon. Tidligere befolkningsnivåer er høyt korrelert med hverandre og initielt befolkningsnivå i 1995 (se A5.2). Dette kalles seriekorrelasjon fordi det er korrelasjon mellom restledd over flere tidsperioder (Wooldridge, 2015, s. 372). I likhet med multikollinearitet vil det føre til mindre nøyaktige estimater og varians.

4.6.6 Metodediskusjon

At forklaringsvariabelen og kontrollvariablene er forhåndsbestemt gjør dem ikke nødvendigvis eksogene. Avgjørelser blir tatt på grunn av forventninger om fremtidig vekst. For eksempel hvis høyere humankapitalnivå i en region øker veksten i en periode, kan man forvente at den også gjør det i neste periode (Duranton, 2008). Dermed er etterspørselen etter humankapital ikke tilfeldig i neste periode, fordi humankapital i neste periode er sterkt korrelert med befolkningsveksten fra tidligere perioder. Dette kalles omvendt årsakssammenheng fordi det er vanskelig å si om økt humankapitalnivå bidrar til befolkningsvekst eller om det er den høye befolkningsveksten som bidrar til etterspørsel etter humankapital. Det er dermed vanskelig å argumentere for at determinanter som sysselsetting i ulike næringssektorer og humankapital er eksogene (Duranton og Puga, 2013).

Homogenitet kan reflektere selvseleksjon som påpekt av De Long (1988). Han hevdet at Baumols (1986) konklusjon om at ubetinget β -konvergens skjedde i løpet av 1870-1979 blant et sett av velstående land (i 1979) er falske av den grunn. Landene som deltok i undersøkelsen kunne ha valgt å delta nettopp fordi de hadde høyere vekst enn gjennomsnittet blant de velstående landene. Dette ville føre til en overestimering av konvergens i den valgte gruppen av land. Slike land utgjør et suksessutvalg. Siden vi har data for alle regionene i Norge, er et slikt suksessutvalg ikke et problem i vår analyse.

Selv om barroregresjoner skal isolere ut mye eksogenitet, har metoden fortsatt svakheter. Dersom man tror kontrollvariablene er endogene, er en IV/2SLS-estimering en standard metode innen økonometri for å finne årsakssammenheng (Wooldrige, 2015, s. 461-498). Å finne et instrument som oppfyller relevans- og eksogenitetskriteriet er vanskelig. 2SLS kan føre til økt usikkerhet og høyere standardavvik enn ved OLS, spesielt hvis det er få instrumentvariabler eller instrumentene er svake. Duranton (2016) mente at det er bekymringsfullt for identifikasjonen av utdanning at raskt voksende byer tiltrekker seg flere utdannede arbeidere og at utdanningsnivået til en bys arbeidsstyrke bestemmes sammen med vekstraten for denne arbeidsstyrken. For å adressere denne bekymringen argumenterte Duranton (2016) for å bruke to instrumentvariabler som bestemmer byens utdanning, men som ikke er korrelert med byens befolkningsvekst. Et godt instrument for utdanning påvirker altså utdanningstilbudet i byen, men ikke etterspørselen etter utdanning. Duranton (2016) brukte antall høyere utdanningsinstitusjoner, hvorav de fleste ble etablert lenge før studieperioden, samt antall

SENA-seter som instrumenter.² Til tross for instrumentenes styrke, viser resultatene at koeffisientene er høyere enn ved OLS, med tilhørende høye standardavvik. Det gjør det vanskelig å sammenlikne IV-estimatene med OLS-estimatene. Andelen sysselsatte med høyere utdanning beholdes som kontrollvariabel og koeffisienten er ganske robust i alle spesifikasjoner. Fordi det er vanskelig å finne gode instrumenter, og fordi barroregresjoner i seg selv skal isolere eksogenitet, har vi valgt å se bort fra IV/2SLS-metoden.

Man kan bli fristet til å håndtere mulig simultanitetsproblem og utelatte variableskjevheter i barroregresjonen ved å bygge et paneldatasett, og kjøre regresjoner med region-faste effekter. Faste effekter fanger opp regionale forskjeller som ikke varierer over tid, og det kan antas at endringene i humankapital er uavhengige av disse faktorene. Hvis for eksempel en regions økonomiske velstand påvirker investeringen i humankapital, medfører dette en korrelasjon mellom endringer i humankapital og uobserverbare faktorer som ikke fanges opp av faste effekter. Estimer med faste effekter (og førstedifferanse-estimer ved 'difference-in-difference'-metoden) kan lide av større skjevheter enn enkle tverrsnittestimer hvis investeringer i humankapital gjøres på grunnlag av regionenes uobserverbare faktorer. Dette gjør at korrelasjonen mellom endringer i humankapital og restleddet i regresjonen er mye sterkere enn korrelasjonen mellom initielt humankapitalnivå og restleddet. For å illustrere dette, kan man tenke at en region opplever befolkningsvekst på grunn av en uobserverbar faktor, som for eksempel en gunstig politikk. Denne veksten fører til økt investering i humankapital, som igjen resulterer i en positiv korrelasjon mellom endringer i humankapital og den uobserverbare faktoren. Når man estimerer effekten av humankapital på befolkningsveksten ved hjelp av faste effekter, vil man ikke kunne fange opp denne korrelasjonen, noe som fører til en skjevhet i estimatet.

Utfordringene som er nevnt i kapittel 4, gjør at det er viktig å være forsiktig med å tolke resultatene som ren årsakssammenheng om hvordan befolkningsvekst påvirkes av kontrollvariablene. Å inkludere kontrollvariabler hjelper på å forklare variasjonen, men det kan fortsatt være en utfordring å kontrollere for alt som påvirker vekstraten. For det første øker sannsynligheten for multikollinearitet. For det andre, kan inkludering av for mange kontrollvariabler føre til overtilpasning av modellen, noe som betyr at modellen passer for godt

² SENA ('Servicio Nacional de Aprendizaje') er en offentlig institusjon som har som formål å fremme utdanning og yrkesopplæring i Colombia. SENA-seter referer til antall institusjoner.

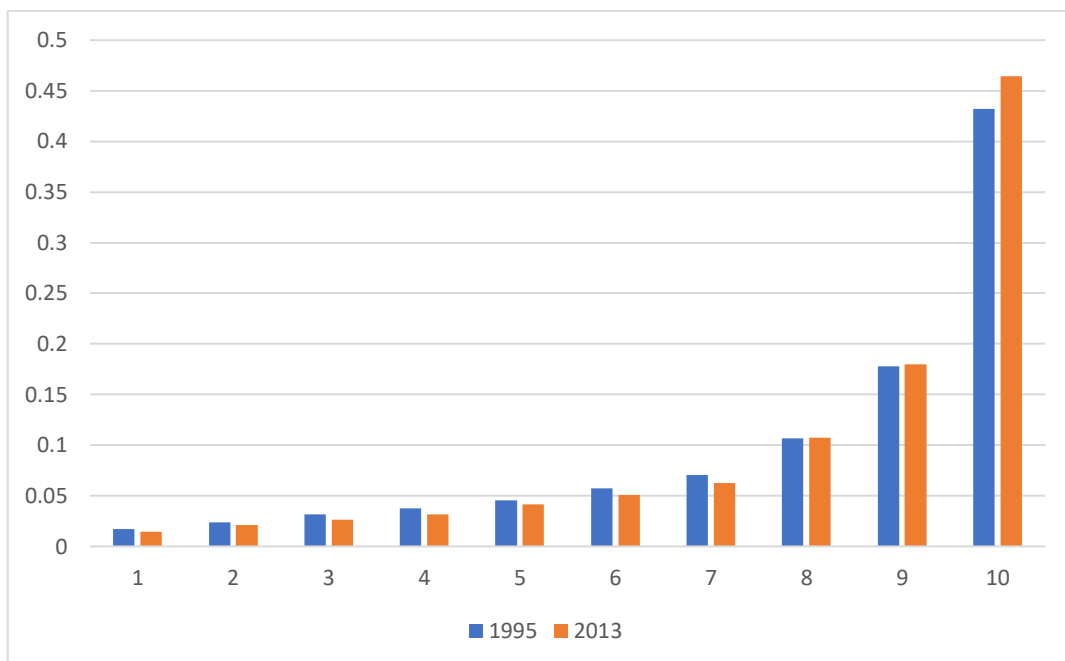
til datasettet som brukes og dermed blir mindre generaliserbar til ny data. I tillegg blir analysen mer kompleks jo flere variabler som inkluderes, noe som kan gjøre det vanskeligere å tolke og formidle resultatene.

Barroregresjoner brukes i vår oppgave til å adressere flere av de økonometriske utfordringene som er nevnt, og som kan være større ved andre empiriske metoder. Metoden er anerkjent for å studere økonomisk- og urbanvekst. Flere robusthetssjekker vil bidra til økt utsagnskraft for resultatene våre.

5 Analyse

5.1 Befolkningsutvikling

Figur 1 viser befolkningsandel for 1995 og 2013, der regionene er fordelt på 10 konstante desiler etter befolkningsnivået i 1995. I 70 prosent av de laveste befolkningsdesilene har befolkningen sunket fra 28,3 prosent i 1995 til 24,8 prosent i 2013. Motsatt har de 30 prosent største regionene økt fra 71,7 prosent i 1995 til 75,1 prosent i 2013. Den største økningen er i desil 10, hvor befolkningen har økt med 3,2 prosentpoeng fra 1995 til 2013. Figuren viser at regioner som allerede var tett befolket, har opplevd en vekst i befolkningsandelen.



Figur 1: Befolkningsandel fordelt i desiler

Sammenlignet med Rattsø og Stokke (2014) er det store likheter. De fant også at det var de tre øverste desilene som har hatt økning i befolkningsandelen i løpet av sin studieperiode. I deres tre høyeste desiler har befolkningen økt med 4,2 prosentpoeng, mens vi har en økning på 3,4 prosentpoeng i de tre høyeste desilene. Den lille forskjellen i Rattsø og Stokke (2014) sine tre høyeste desiler sammenliknet med våre, kan skyldes at vi ser på en kortere periode. De har også, i likhet med oss, en nedgang i befolkningstallet fra 1972 til 2008 for alle lavere desiler. Vårt resultat er ikke direkte sammenlignbart med Rattsø og Stokke (2014) fordi vi ser på en kortere periode, men i likhet med dem viser resultatene tegn til urbanisering. Beeson et al. (2001) fant at andelen av befolkningen i desil 4-9 i amerikanske regioner ble redusert i perioden

1840-1990, mens andelen av befolkningen i desil 1-3 økte. Rattsø og Stokke (2014) viste at Norge dermed har hatt en raskere urbaniseringsprosess enn USA, noe figur 1 også bekrefter. Resultatet gir en indikasjon på at det har vært divergens i befolkningsnivået mellom regionene i Norge.

5.2 Ubetingede barr regresjoner

Tabell 3 viser absolutte barr regresjoner for perioden vi studerer, men også for delperioder og utvidet periode. Kolonne 1 viser absolutt divergens for perioden 1995-2013. Den positive koeffisienten på 0,004 sier at en region med dobbelt så høyt befolkningsnivå som en annen i 1995, forventes å ha 0,4 prosentpoeng høyere befolkningsvekst i perioden, alt annet likt. Den positive koeffisienten for initielt befolkningsnivå indikerer ubetinget divergens for norske regioner i perioden 1995 til 2013 og er lavere enn i Rattsø og Stokke (2014) som fant at gjennomsnittlig, årlig vekstrate var på 0,6 prosentpoeng for perioden 1972-2008. Ayuda et al. (2010) fant omtrent like stor gjennomsnittlig, årlig vekstrate som oss for regioner i 8 av de 9 landene som ble studert i perioden. Den positive koeffisienten betyr at regioner som har hatt et høyt befolkningsnivå i 1995, har hatt en høyere befolkningsvekst enn regioner med et lavere befolkningsnivå i 1995. OLS-regresjonslinjen er illustrert i figur 2.

Absolutt divergens er funnet for alle delperioder, noe som innebærer at det har vært befolkningsdivergens i norske regioner helt tilbake til 1910. Denne forskjellen i utviklingen i befolkningsvekst har vist seg å øke i nyere tid. Dette betyr at divergensen i befolkningsnivået har akselerert i denne tidsperioden. Tilflytting til urbane områder har vært hyppigere i 1995-2013 enn tidligere perioder. Den raske veksten her er i tråd med tolkningen av figur 1.

Koeffisientene er derimot ikke signifikante for periodene 1910-1954 og 1910-1969. Disse delperiodene omfatter både første og andre verdenskrig, som kan ha hatt betydelig innvirkning på befolkningens bevegelse og fordeling i Norge. Krigene kan ha ført til interne migrasjonsbevegelser og endringer i befolknings sammensetningen i de ulike regionene. Dette, i tillegg til mulige målefeil på grunn av dårlig oppmøte på folketellingene, kan ha ført til midlertidige avvik fra den langsiktige trenden for befolkningsfordelingen (SSB, 2009).

Særlig var folketellingene i årene rett etter andre verdenskrig begrenset. Den planlagte folketellingen i 1940 ble avlyst, og i stedet ble det holdt en ekstraordinær telling i 1946. På

grunn av mangelfulle utfylte lister, ble denne tellingen kraftig forsinket og resultatene var derfor ikke klare før i 1950. Til tross for dette fødte norske kvinner som aldri før, og distriktsområder opplevde raskere vekst enn byområdene (SSB, 2009).

Etter flere år med synkende fødselsrater snudde trenden i 1935, og mellom 1941 og 1946 var fødselsraten tilbake på samme nivå som i 1920-årene. Dette, kombinert med lavere dødsrate og redusert emigrasjon, førte til at den fastboende befolkningen steg fra omtrent 2,7 millioner i 1930 til over 3,1 millioner i 1946 (SSB, 2009).

I løpet av denne perioden opplevde alle fylker en befolkningsvekst. Veksten var spesielt markant i Rogaland, Hordaland, Nordland og Troms. I Finnmark var mange nødt til å flykte under krigsårene, noe som resulterte i en betydelig svakere befolkningsvekst der (SSB, 2009). I 1946 bodde 72 prosent av befolkningen på landsbygda, mens 28 prosent holdt til i byene. I 1920 var dette forholdet 70 og 30 prosent. En viktig årsak til at byene opplevde lavere vekst sammenlignet med landsbygdene, var at befolkningstilveksten var betydelig sterkere i forstedene enn i bykjernene. Forstedene lå utenfor bygrensene, og bidro derfor ikke til befolkningsvekst i byene. Utviklingen av offentlig transport ble fremhevet som en vesentlig faktor for denne trenden. På grunn av den økte befolkningsveksten på landsbygda og i forstedene, relativt til befolkningsveksten i byområdene, er koeffisientene ikke så sterke i kolonne 3-6.

Det er også verdt å merke seg at etterkrigstiden var preget av stor befolkningsvekst i Norge, spesielt på grunn av økonomisk utvikling, økt levestandard og økende levealder. Dette førte til urbanisering fordi mange mennesker flyttet fra distriktsområder til byer og tettsteder. Migrasjonen kan ha bidratt til å utjevne forskjeller i befolkningsfordelingen som kan ha eksistert før krigen.

Fra 1970-tallet endret flyttemønsteret seg noe. Nettoinnflyttingen til sentrale områder avtok, noe som kan knyttes til lavere fødselstall og økt fokus på arbeidsplasser i distriktene. Innføringen av folketrygden i 1967 bidro også til å bremse utflyttingen fra distriktene (Utvik, 2022).

Utover 1980-tallet har Oslo og andre storregioner opplevd sterkere vekst igjen. Nettotilflyttingen til urbane områder skyldes i stor grad utflytting fra distriktene, men også

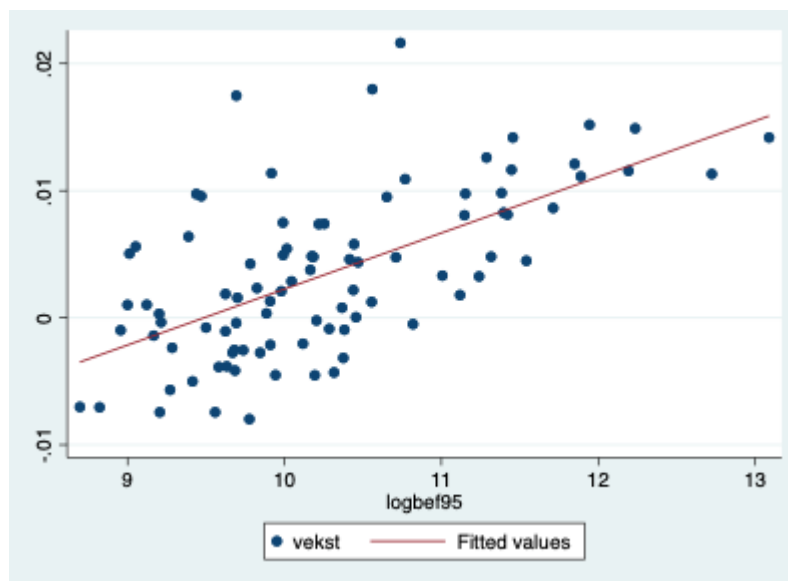
innvandring av folk som i stor grad bosetter seg i sentrale strøk (Utvik, 2022). Denne divergensen har vist seg å være sterkest i perioden vi studerer.

Duranton (2016) fant mye sterkere ubetinget divergens for colombianske byer i perioden 1938-2010. Koeffisienten er 0,08 foran initiell befolkning. Myndighetenes landsomfattende helsekampanjer rundt 1950, en økning i offentlige utgifter til utdanning fra 1956, og en økning i grunnskoleopptak bidro til en positiv endring i sosioøkonomiske forhold i Colombia. Imidlertid ble økte midler til helse og utdanning distribuert ujevnt og favoriserte urbane områder. Denne faktoren, kombinert med automatiseringen av jordbruket på 1950-tallet og veksten i industri og bygge- og anleggsvirksomhet på 1960- og 1970-tallet, førte til migrasjon fra distriktsområdene. Dette er med på å forklare den sterke urbaniseringen i colombianske byer (Mejía et al., 2008).

Tabell 3: Ubetinget divergens over ulike perioder.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
vekst	1995-2013	1910-2013	1910-1954	1910-1969	1910-1995	1930-1954	1930-1969	1930-1995	1954-1995
log bef95	0.004*** (0.000)								
log bef10		0.002*** (0.000)	0.000 (0.001)	0.001 (0.001)	0.001** (0.000)				
logbef30						0.001*** (0.001)	0.002*** (0.001)	0.002*** (0.001)	
logbef54									0.003*** (0.001)
Regioner	90	90	90	90	90	90	90	90	90
R ²	0.391	0.091	0.006	0.022	0.046	0.058	0.089	0.099	0.131

Robuste standardavvik i parentes. *, **, ***: signifikant ved 1%, 5% og 10%.



Figur 2: Ubetinget befolkningsdivergens for 90 regioner

5.3 Spline-regresjoner

Som beskrevet i metode-kapittelet deles regionene inn i tre grupper med like mange observasjoner i hver gruppe i tabell 4. Ved å sortere befolkningen i 1995 etter størrelse får vi en gruppe med lave befolkningsnivåer, en med mellomstore befolkningsnivåer og en med større befolkningsnivåer. Grensene for gruppeinndelingen er for den lille gruppa, et logaritmisk befolkningsnivå opp til 9,78. For den mellomstore gruppa er intervallet mellom 9,78 og 10,46 (dette tilsvarer befolkningstall mellom 17500 og 35000). Den største gruppa inkluderer regionene med logaritmisk befolkningsnivå over 10,46. Fra tabell 4 kan man lese at alle gruppens log befolkningsnivå i 1995 har hatt en positiv effekt på befolkningsveksten, men det er kun koeffisienten for den store gruppa som er signifikant.

Tabell 4: Splineregresjoner.

	(1) vekst	(2) vekst	(3) vekst
log befolkning 95	0.003 (0.003)	0.001 (0.004)	0.003** (0.001)
Konstant	-0.028 (0.032)	-0.013 (0.042)	-0.027 (0.016)
Regioner	30	30	30
Størrelse	liten	mellomstore	stor
logbef95	logbef95<9.78	9.78<logbef95<10.46	logbef95>10.46
R ²	0.024	0.005	0.149
F-stat	0.755	0.129	5.543

Robuste standardavvik i parentes. *, **, ***: signifikant ved 1%, 5% og 10%.

Rattsø og Stokke (2014) fant også positive effekter for alle tre gruppene, men nesten 10 ganger mindre effekt for den minste og største gruppa enn i vårt tilfelle, dog ikke signifikante resultater for disse gruppene. Resultatene for divergens for den mellomstore gruppa er betydelig større og signifikant i Rattsø og Stokke (2014).

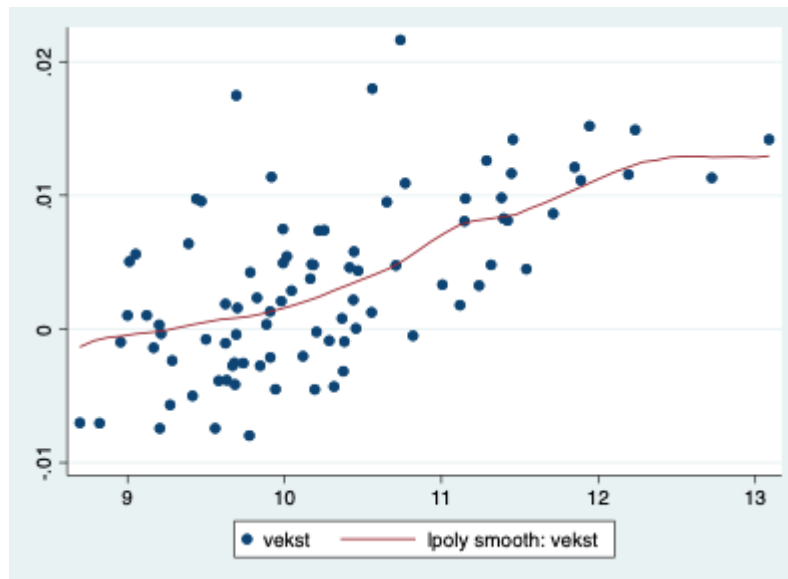
For den minste og mellomstore gruppa tyder standardavvikene på at det er større variasjoner i befolkningstallene innad i disse gruppene enn i den største gruppa, som gjør at resultatene for divergens ikke er signifikante. Samtidig tror vi på at koeffisientene er lavere enn i den generelle barroregresjonen for alle 90 regioner. Det skyldes at regionene innad i gruppene trolig er mer homogene når det gjelder institusjonelle forhold, utdanningsnivå og andre karakteristika. Koeffisienten for den største gruppa er trolig signifikant fordi de største regionene i Norge

inngår i denne gruppa, og har vesentlig «bedre» instiusjonelle forhold og andre faktorer som bidrar til befolkningsvekst enn de mindre regionene i denne gruppa.

5.4 Lokal polynomglatting (ikke-lineært kernelplott)

Lokal polynomglatting anvendes her for å se hvordan korrelasjonen mellom befolkningsvekst og initielt befolkningsnivå endrer seg etter regionsstørrelse. Figur 3 sammenlikner den ikke-lineære kernel-regresjonslinjen og spredningen i befolkningsvekst for de 90 økonomiske regionene. Regresjonslinjen bekrefter at alle gruppene har positiv gjennomsnittlig vekstrate, men den middels store gruppa og de minste regionene i den største gruppa har mer markant stigende helning. Det bratteste punktet er innad i den store gruppa, rett over cut-off på *log befolkning* 95 på 10,46. Ullensaker/Eidsvoll-regionen inngår blant annet her med den høyeste vekstraten i datamaterialet vårt (2,2 prosentpoeng). Med korte og gode transportforbindelser til Oslo og andre deler av landet, inkludert tog, veier og ikke minst flyplass (Gardermoen), har Ullensaker/Eidsvoll-regionen opplevd tilflyttingsstrømmer. Beslutningen om å etablere Gardermoen som hovedflyplass i 1993, og den offisielle åpningen i 1998, har nok spilt en viktig rolle for den høye befolkningsveksten i denne regionen i påfølgende år. Jæren, som ligger i utkanten av Stavanger/Sandnes, inngår også i det bratteste punktet her med en vekst på omtrent 1,8 prosentpoeng. Her har utviklingen i olje- og gassnæringen og leverandørindustrien hatt særlig stor betydning. For de fire største regionene har initelt befolkningsnivå hatt mindre og si for veksten. Dette gjelder regionene Oslo, Bergem, Stavanger/Sandnes og Trondheim, med en gjennomsnittlig, årlig vekst på litt over 1 prosentpoeng.

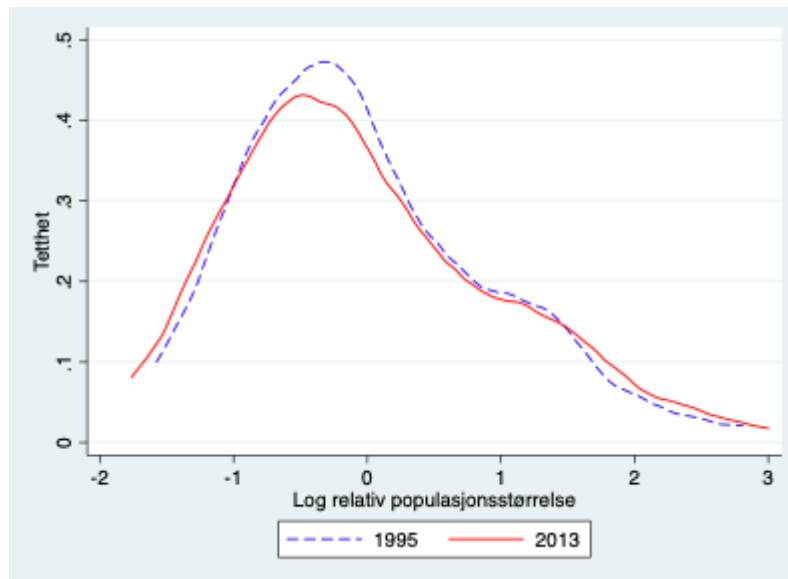
Selv om Jæren- og Ullensaker/Eidsvoll-regionene vokser raskest i den største gruppa, er det flere av de mindre regionene i den største gruppa som vokser saktere enn de fire største regionene, som demper den gjennomsnittlige årlige veksten. Resultatene fra spline-regresjonene og det ikke-lineære kernelplottet forklarer at absolutt divergens i Norge er drevet av mellomstore- og periferiregioner.



Figur 3: Lokal polynomglatting.

5.5 Kernel-tettheter

Kernel-tetthetsfunksjonene i figur 4 representerer de relative befolkningsstørrelsene i 1995 og 2013. Den horisontale akse er befolkningen i en gitt region relativt til gjennomsnittet for alle regionene, i logaritmisk skala. Den vertikale akse gjenspeiler tettheten som korresponderer med de varierende relative befolkningsstørrelsene. Begge funksjonene har en entoppet fordeling, med størst tetthet av regioner med befolkning rett under gjennomsnittet. Toppen ligger litt høyere for relativ befolkningsstørrelse i 1995. Differansen mellom de to toppene utgjør tapet av tetthet i 2013 relativt til 1995. Tettheten har vist en oppgang i begge ender av distribusjonen. Med andre ord var variasjonen i befolkningsstørrelse i norske regioner i 1995 mindre, med flere regioner som hadde lignende befolkningstall. På en annen side antyder den lavere toppen for 2013-funksjonen, samt at denne funksjonen ligger utenfor 1995-funksjonen for små og store regioner at befolkningstallene var mer spredt ut i 2013. Tolkningen er at mellomstore regioner med befolkningsnedgang øker tettheten av små regioner, mens mellomstore regioner med betydelig befolkningsvekst bidrar til den høye tettheten i den øvre delen av distribusjonen. Figur 4 viser at forskjellene i befolkningsstørrelse mellom regionene har økt, noe som tyder på befolkningsdivergens, samt at denne divergensen er drevet av de mellomstore regionene. Dette er konsistent med resultatene fra spline-regresjonene og det ikke-lineære kernelplottet.



Figur 4: Kernel-tettheter

Figur A2 i appendiks sammenlikner kernel-tetthetsfunksjoner for de relative befolkningsstørrelsene i 1910 og 2013. Tilpasningsmekanismene i befolkningen i perioden 1995-2013 er de samme som er funnet i perioden 1910-2013. For den lengre perioden er forskjellen i relativ befolkningsstørrelse i 2013 enda større enn i 1910. Dette bekrefter resultatene fra de ubetingede barroregresjonene om at divergensprosessen i Norge eksisterte lenge før vår studieperiode.

5.6 Barroregresjoner og humankapital

Humankapital er akseptert som en viktig driver til befolkningsvekst (Duranton, 2016). Vi utvider derfor den ubetingede barroregresjonen fra vår studieperiode til å inkludere ulike mål på humankapital. Tabell 5 viser effekten av andel sysselsatte med ulik utdanningsbakgrunn som del av arbeidsdyktig befolkning i regionen. Utdanningsvariabelen skiller mellom andelen sysselsatte av arbeidsdyktig befolkning med universitets- og høyskolebakgrunn, grunnskolebakgrunn og videregående bakgrunn. Tabell 5 viser resultatene fra barroregresjoner for de 90 økonomiske regionene i Norge over perioden 1995 til 2013.

Tabell 5: Deler utdanningsnivå på arbeidsdyktige i regionen i 1995. OLS-resultater.

	(1)	(2)	(3)	(4)	(5)	(6)
	vekst	vekst	vekst	vekst	vekst	vekst
log befolkning 95	0.004*** (0.001)	0.025*** (0.004)	0.025*** (0.004)	0.025*** (0.004)	0.025*** (0.004)	0.025*** (0.004)
andel utdannet	0.025 (0.015)	-0.005 (0.018)	-0.001 (0.003)	0.041 (0.037)	-0.009 (0.019)	-0.015 (0.060)
andel utdannet_sq						0.028 (0.152)
Konstant	-0.039*** (0.005)	-0.025*** (0.005)	-0.028*** (0.009)	-0.030*** (0.006)	-0.020** (0.010)	-0.024*** (0.006)
Tidligere befolkning	Nei	Ja	Ja	Ja	Ja	Ja
Utdanning	univ.	univ.	loguniv.	grunnskole	vgs.	univ.
Regioner	90	90	90	90	90	90
R ²	0.401	0.632	0.632	0.642	0.633	0.632

Robuste standardavvik i parentes. *, **, ***: signifikant ved 1%, 5% og 10%. Tidligere befolkning er log befolkning for 1910, 1930, 1954 og 1969.

Hittil har vi sett på ubetinget divergens mellom regioner, men disse regresjonene kontrollerer ikke for andre relevante faktorer som kan påvirke befolkningsveksten. Videre vil vi studere betinget divergens for perioden 1995-2013. Dersom utelatte variable er korrelert med befolkningsnivået i 1995, vil regresjonen gi en skjev koeffisient i tabell 3.

For å redusere sannsynligheten for utelatte variableskjevhet, inkluderes andel utdannet i kolonne 1-6. Når vi inkluderer andelen med minimum ett år høyere utdanning i kolonne 1, ser vi at forklaringsvariabelen vår ikke endrer seg. Det er konsistent med resultatet for absolutt divergens, og gir en indikasjon på en stabil kausal divergens. Dette kan forklares ved at initielt befolkningsnivå og initelt utdanningsnivå er uavhengige. Utdanningsvariabelen i kolonne 1 er positiv. Tolkningen av koeffisienten er at en dobling i andelen arbeidsdyktige med minimum ett års høyere utdanning som er sysselsatt forventer å øke den gjennomsnittlige, årlige befolkningsveksten med 2,5 prosentpoeng, alt annet likt. Koeffisienten er derimot ikke signifikant, og vi frykter at modellen lider av utelatte variableskjevhet.

Tidligere befolkningsnivåer inkluderes for å ta hensyn til seriekorrelasjon. Hvis restleddet er seriekorrelert, kan dette føre til et skjevt estimat av koeffisienten foran initielt befolkningsnivå. Forhåpentligvis vil skjevheten, om noen, bli dempet ved å legge til ytterligere tilbakedateringer av befolkningsnivåer som kontroller. I tillegg forventer vi at tidligere befolkningsnivåer

kontrollerer for mange uobserverbare variabler som driver befolkningsvekst og som er korrelert med befolkningsdemografi (Duranton, 2016).

Når vi kontrollerer for tidligere befolkning i kolonne 2-6, ser vi at effekten av ulikt utdanningsnivå er lav og ikke signifikant. Duranton (2016) har en mye større og signifikant effekt av utdanning på befolkningsveksten i Colombia når han kontrollerer for tidligere befolkningsnivåer. Dette skyldes trolig at tidligere befolkningsnivåer fanger opp effekter som drev vekst tidligere i Norge, som for eksempel utdanningsnivå og tilgang til ressurser og arbeidskraft. Når disse effektene fanges opp i tidligere befolkning, er det forventet at effekten av ulik utdanning i 1995 blir lavere og ikke signifikant. Det er fordi den tidligere befolkningsvariabelen er sterkt korrelert med befolkningsnivået i 1995. Tidligere befolkningsnivå forstyrrer forholdet mellom utdanning og befolkningsvekst, og gjør det vanskelig å isolere effekten av utdanning på befolkningsvekst. Vi kan derfor ikke si med sikkerhet at det er en kausal sammenheng mellom humankapital og befolkningsvekst.

I Duranton (2016), ved inkludering av tidligere befolkningsnivåer, øker koeffisienten for det initielle befolkningsnivået i studieperioden, og går fra å være ikke-signifikant til å bli signifikant og robust. Samtidig er koeffisientene for utdanningsvariablene signifikant og høyere enn i norske regioner ved inkludering av tidligere befolkningsnivåer. Det tyder på at i Colombia, i motsetning til i Norge, har utdanning en mer umiddelbar og betydelig innflytelse på befolkningsveksten, til og med når tidligere befolkningsnivåer er kontrollert for. Forskjellene i resultatene mellom Norge og Colombia kan skyldes flere ting. For det første er Colombia et utviklingsland, og tilgang til utdanning er ikke like opplagt som i Norge, og har derfor en mer direkte innvirkning på befolkningsvekst, siden det kan påvirke økonomiske muligheter betraktelig. For det andre er Norge generelt kjent for å ha høy levestandard og høyt utdanningsnivå, som betyr at ytterligere økninger i utdanningsnivå ikke nødvendigvis påvirker veksten like mye som i Colombia. For det tredje, kan det tenkes at Norge har en mer stabil befolkningsstruktur og mindre intern migrasjon sammenlignet med Colombia. Det kan også være at tidligere befolkningsnivåer har mer innflytelse i et mer stabilt miljø, mens i et mer ustabilt eller dynamisk miljø vil utdanning ha mer innflytelse.

Andelen sysselsatte av arbeidsdyktig befolkning med høyere utdanning kvadrert er inkludert i kolonne 6 (*andel utdannet_sq*) for å undersøke om det eksisterer en ikke-lineær sammenheng mellom andel høyere utdannede og befolkningsveksten. Duranton (2016) fant også en ikke-

signifikant negativ, og tiltakende avkastning av humankapital på befolkningsveksten når han kontrollerte for tidligere befolkningsnivåer. I A3 i appendiks, når vi ikke har kontrollert for tidligere befolkning, er effekten av høyere utdanning på veksten positiv, men avtagende. Dette er i tråd med loven om avtagende avkastning (Schultz, 1961). Loven om avtagende avkastning av humankapital på befolkningsveksten, kan sees i lys av Duranton (2008) sin modell. Den avtagende veksten kan blant annet skyldes mismatch i arbeidsmarkedet, men også at levekostnadene (trengselseffektene) øker dess sterkere agglomerasjonseffektene er av humankapital (Duranton, 2008, Duranton og Puga, 2013).

For å sjekke hvorvidt loven om avtagende avkastning av humankapital på befolkningsveksten gjelder i vårt tilfelle, utfører vi en F-test på koeffisientene for *andel utdannet* og *andel utdannet_sq*. Resultatene fra testen (F-stat=1.47) antyder at det er en lineær sammenheng mellom andel utdannede og befolkningsveksten. Vi får omtrent like lav F-stat uansett hvilket av de to målene på humankapital vi bruker (deler på sysselsatt befolkning eller arbeidsdyktig alder).³ Tolkningen av den lineære effekten av humankapital på befolkningsveksten er at investeringer i humankapital i Norge fortsatt har en høy avkastning til tross for høye initelle verdier. Dette kan skyldes høy kvalitet på utdanning, trening og læring, samt gode institusjonelle forhold som oppmuntrer og støtter livslang læring. På en annen side kan det ta tid for effekten å vise seg, og den fanges derfor ikke opp når vi henter humankapitalnivået kun fra det initialåret. Av disse grunner vil vi ikke fortsette med det kvadrerte leddet.

Selv om koeffisientene foran initielt befolkningsnivå i 1995 er signifikant i alle likningene, er utdanningsvariabelen ikke signifikant, og til og med lavere når vi inkluderer tidligere befolkningsnivå. Vi kan derfor ikke konkludere med at utdanning har en kausal effekt på befolkningsvekst, og vi trenger andre metoder for å finne den kausale sammenhengen. I appendiks A4 når vi deler utdanningsnivå på sysselsatt befolkning, er resultatene enda svakere enn i tabell 5. Selv om effekten av log initielt befolkningsnivå på veksten er den samme, er effekten av utdanning enda svakere. Å oppnå en ren signifikant effekt av utdanning i empirisk analyse byr på ulike utfordringer. Utdanningsnivå i en region avspeiler så mye annet som også er relevant for veksten. Det er dermed vanskelig å få rendyrket effekten av utdanningsnivå. For

³ Se appendiks A4: *andel utdannet* er sysselsatt befolkning med ulikt utdanningsnivå delt på sysselsatt befolkning.

eksempel er sysselsetting korrelert med utdanning og påvirker også befolkningsveksten. Å ikke kontrollere for sysselsetting kan føre til skjeve estimater av utdanning på befolkningsveksten.

Den viktige identifiseringsbekymringen i tabell 5 er at raskt voksende regioner primært kan tiltrekke seg mer utdannede arbeidere, eller at utdanningen i en regions arbeidsstyrke kan bestemmes av vekstraten til den arbeidsstyrken. For å ta tak i denne bekymringen, trenger vi en eller flere variable som bestemmer utdanningsnivået i regionen, men som ikke er korrelert med befolkningsvekst. Duranton (2016) går over til en IV/2SLS-estimeringsmetode for å ta tak i denne bekymringen i Colombia. Av grunner nevnt i delkapittel 4.6.6 velger vi å ikke gå videre med en slik metode.

Uansett hvilket mål vi bruker på utdanning i regresjonene, er resultatene for betinget divergens robust. Den større effekten av utdanning i A3 fanges opp i effekten av tidligere befolkning i tabell 5, og gjør det vanskelig å isolere effekten av utdanning. Videre vil vi se på sysselsatt andel av arbeidsdyktig alder i befolkningen med minimum ett år høyere utdanning i regionen som vår variabel for mål på humankapital.

5.7 Barroregresjoner og næringssektorer

Tabell 6 viser betinget divergens for alle spesifikasjoner når vi ser på næringssektorer. Tabellen undersøker hvordan sysselsetting i produksjons-, business-, og servicesektorene påvirker befolkningsveksten. Når vi ikke kontrollerer for tidligere befolkningsnivåer, ser vi at befolkningsveksten er den samme som i tabell 5, kolonne 1. Samtidig, er befolkningsveksten like stor i tabell 5 og 6 når det kontrolleres for tidligere befolkningsnivåer.

I kolonne 2 og 3 inkluderer vi også de som er sysselsatt med minimum ett år høyere utdanning som andel av arbeidsdyktig befolkning. Ser som i tabell 5, at denne kun er signifikant når vi ikke inkluderer tidligere befolkningsnivåer. Av samme intuisjon som i tabell 5, skyldes trolig den svakt negative og ikke-signifikante effekten av utdanning i kolonne 3 at tidligere befolkningsnivåer er sterkt korrelert med både utdanningsvariabelen og befolkningsveksten. Dette forklarer også hvorfor befolkningsveksten øker når vi inkluderer tidligere befolkning (se intuisjon i delkapittel 5.6). Av samme intuisjon ser vi at andelen sysselsatte i produksjonssektoren går fra å være signifikant i kolonne 1 og 2 til å reduseres og ikke være signifikant i kolonne 3. Dette skyldes trolig at tidligere sysselsettingsmønstre drev vekst før

studieperioden, som fanges opp av tidligere befolkningsnivåer. Den signifikante koeffisienten i kolonne 2 foran andelen i produksjonssektoren sier at en dobling av andelen sysselsatte i produksjonssektoren øker den årlige befolkningsveksten med 1,5 prosentpoeng, alt annet likt. Denne effekten er svært liten sammenliknet med i Duranton (2016).

Å inkludere sysselsetting i produksjon- og businesssektorene i kolonne 2 gjør at effekten av utdanning er høyere enn uten.⁴ Denne endringen antyder at det er et visst samspill mellom utdanning og sysselsetting, som kan påvirke den estimerte effekten av utdanning på befolkningsveksten. Effekten av utdanning på befolkningsveksten isoleres når vi kontrollerer for sysselsetting. Dersom multikollinearitet er et problem kan det være vanskelig å tolke de individuelle effektene av variablene. Vi skal se nærmere på problemene knyttet til multikollinearitet i det følgende.

Tabell 6: Næringssektorer

	(1) vekst	(2) vekst	(3) vekst
log befolkning 95	0.004*** (0.001)	0.004*** (0.001)	0.026*** (0.003)
andel utdannet		0.049** (0.019)	-0.004 (0.021)
andel produksjon	0.014** (0.007)	0.015** (0.007)	0.011 (0.008)
andel business	0.019 (0.027)	-0.027 (0.033)	0.015 (0.027)
Konstant	-0.046*** (0.007)	-0.049*** (0.006)	-0.027*** (0.006)
Regioner	90	90	90
Tidligere befolkning	Nei	Nei	Ja
R ²	0.471	0.494	0.764

Robuste standardavvik i parentes. *, **, ***: signifikant ved 1%, 5% og 10%. Tidligere befolkning er log befolkning for 1910, 1930, 1954 og 1969. Servicesektoren er brukt som referansekategori for næringssektorer.

Koeffisienten foran businesssektorvariabelen er ikke signifikant i noen av regresjonene, og til og med negativ i kolonne 2. Den sier at en dobling i andelen sysselsatte i businesssektoren er forventet å redusere den årlige befolkningsveksten med 2,7 prosentpoeng, alt annet likt. Dette er ikke en forventet effekt. Businesssektor-yrker krever ofte høyere utdanning (sterkt korrelert med utdanningsvariabelen i A5.1), og bedrifter i denne bransjen fungerer som positive goder

⁴ Se tabell 5, kolonne 1.

for omgivelsene, og øker befolkningsveksten (Porter, 1998). På grunn av den høye korrelasjonen mellom utdanning- og businessvariabelen er det vanskelig å isolere noen kausal sammenheng mellom businessvariabelen og befolkningsveksten.

Korrelasjonen (se A5.1) mellom utdanning og sysselsatte i produksjonssektoren er negativ. En stor del av jobbene i produksjonssektoren krever typisk ikke så mye utdanning, som forklarer hvorfor korrelasjonen mellom utdanningsvariabelen og produksjonssektorvariabelen er negativ. Effekten av utdanning i tabell 5, kolonne 1, er typisk feilestimert og lavere enn i tabell 6, kolonne 2 fordi den fanger opp den negative korrelasjonen utdanning og sysselsetting i produksjonssektoren har. Produksjonssektoren var sterkt voksende fra 1995 til 2013, primært som følge av nær en tredobling av sysselsatte i olje- og gassnæringen og i bygg- og anleggsbransjen fra 1995 til 2013 (se utregning i A6). Å kontrollere for denne «støyen» i tabell 6 fjerner dermed den negative effekten sysselsetting i produksjonssektoren har på utdanningen.

På den annen side, tilfører businessvariabelen bare støy i regresjonen fordi korrelasjonen mellom utdanning og andel i businesssektoren er så høy.⁵ Utdanningskoeffisienten er derfor for høy i kolonne 2. Her kunne vi utført en VIF-test, men av grunner nevnt i delkapittel 4.6.4 vil denne foreslå at multikollinearitet ikke eksisterer fordi R^2 er altfor lav til å kunne gi VIF-verdi over 10. Businessvariabelen bør tas bort i regresjonen (se tabell 7, kolonne 1).

Å inkludere kontrollvariabler for tidligere befolkningsnivåer har økt forklaringskraften i modellen fra 47,1 prosent til 76,4 prosent, men av nevnte årsaker skyldes trolig den økte forklaringskraften at tidligere befolkningsnivåer er høyt korrelert med befolkningen i 1995 og befolkningsveksten. I delkapittel 5.8 vil vi prøve å isolere ut eksogenitet av sysselsetting på befolkningsveksten ved å bruke en annen definisjon på sysselsetting.

5.8 Barroregresjoner og Bartik-instrumenter

Vårt hovedfokus i tabell 7 er å undersøke robustheten av resultatene i tabell 6 ved å bruke Bartik-instrumentet som et annet mål på sysselsetting. Bartik-instrumentet fanger opp industri-spesifikk sysselsettingsvekst på nasjonalt nivå ved å multiplisere den lokale sysselsettingen i 1970 med den faktiske sysselsettingsveksten på nasjonalt, næringsnivå for alle sektorer fra 1970-1990. Dette kan hjelpe oss å identifisere den kausale påvirkningen av sysselsetting på

⁵ Appendiks A5.1

befolkningsvekst i perioden 1995-2013. Hensikten med Bartik-instrumenter er å fjerne noe av de lokale «vanene» eller «mønstrene» som gjør at folk velger hvilken sektor de vil jobbe i og i hvilken region. På denne måten kan vi fjerne den delen av arbeidstilbudssjokk, for eksempel inntog av kvinner, som retter seg mot den spesifikke regionen, og vi får at lokal arbeidskraftetterspørsel kun baserer seg på sektorandelene i 1970, og nasjonale sysselsettingsvekstrater fra 1970-1990. Dersom befolkningsveksten i norske regioner reagerer mer på tidligere sysselsettingsmønstre, enn initiell sysselsettingsstruktur, kan Bartik-instrumenter hjelpe oss å forklare sysselsettingsmønstre for vår studieperiode.

Selv om Bartik-instrumentene er konstruert fra en periode før vår studieperiode, antar vi at den er relevant for befolkningsveksten også i perioden 1995-2013, men at den er mindre korrelert med andre kontrollvariabler slik at det er enklere å finne en kausal sammenheng av sysselsetting på befolkningsveksten. I korrelasjonsmatrise A5.3 er korrelasjonen mellom Bartik i businesssektoren og utdanningsvariabelen mindre urovekkende enn korrelasjonen mellom andel i businesssektoren og utdanningsvariabelen. Det er lovende for identifiseringen av sysselsetting på befolkningsveksten.

Andel sysselsatte i produksjonssektoren sammenliknet med Bartik for denne sektoren har den samme positive koeffisienten, men Bartik er noe mer signifikant. En dobling i enten andel sysselsatte i produksjonssektoren i 1995 eller i den predikerte nasjonale veksten av andel sysselsatte i produksjonssektoren i 1990 vil øke den gjennomsnittlige, årlige befolkningsveksten med 1,6 prosentpoeng, alt annet likt. Å bruke predikert sysselsetting i produksjonssektoren i 1990 som ser bort fra arbeidstilbudssjokk i perioden 1970-1990, gir mer signifikant effekt på befolkningsveksten i 1995-2013 enn å bruke sysselsettingen i initialåret i produksjonssektoren. Det kan skyldes at denne variabelen har isolert ut mer eksogenitet enn andel sysselsatte-variabelen i initialåret har gjort. Andel sysselsatte kan lide av mer endogenitet sammenliknet med Bartik, selv om den er basert på data fra startåret av studieperioden, og dermed reduserer simultanitetsproblemer. Dette kan skyldes at kvantum av arbeidskraft kan være endogent bestemt av idiosynkratiske årsaker. Bartik er konstruert slik at denne endogeniteten, empirisk referert til som regional endogenitet, reduseres (Combes et al., 2010). Duranton (2016) brukte også Bartik-instrumentet for å studere befolkningsvekst, men konstruerte det på en annen måte enn oss som gjør det vanskelig å sammenlikne koeffisientene med våre.

En annen interessant observasjon er at ved å inkludere Bartik i businesssektoren, endrer både fortegnet seg og vi får en koeffisient som er signifikant på et 10-prosentnivå. Den negative koeffisienten foran andel i businesssektoren kan igjen skyldes at utdanningsvariabelen er høyt korrelert med andel i businesssektoren. Å bruke denne variabelen som mål på sysselsettingen tilfører bare støy i regresjonen, og den positive effekten av å være sysselsatt i businesssektoren på befolkningsveksten fanges dermed opp i utdanningsvariabelen. Resultatet av utdanning er, som i tabell 6, overestimert. Derimot er Bartik i businesssektoren mindre korrelert med utdanningsvariabelen, og vi kan isolere ut mer av effektene av disse to variablene på befolkningsveksten. Den høye korrelasjonen mellom utdanning og businesssektor-variabelen gjør at Bartik er et bedre mål på sysselsetting i denne sektoren. Det ser vi blant annet ved at utdanningsvariabelen er mer robust, og at effekten av Bartik på befolkningsveksten er positiv og signifikant.

Servicesektor-variabelen er negativ og signifikant på et 5 prosent signifikansnivå. Koeffisienten sier at en dobling i andelen sysselsatte i servicesektoren vil redusere den gjennomsnittlig, årlige befolkningsveksten med 1,5 prosentpoeng, alt annet likt. Koeffisienten er så og si lik når vi bruker Bartik som mål på sysselsetting istedenfor, men på et lavere signifikansnivå. Tolkningen er at serviceyrkene er lavproduktive og til dels lavlønte yrker i butikker, restauranter, og innenfor turisme osv. Dette kan bidra til at en høyere andel sysselsatte i servicesektoren fører til lavere befolkningsvekst, ettersom det kan være mindre attraktivt for folk å flytte til eller bli værende i områder med overveiende lavtlønnede jobber. Slike yrker har mangel på positive agglomerasjonseffekter (Baum-Snow og Pavan, 2013). I motsetning til kunnskapsintensive sektorer, som i produksjons- og businesssektoren, kan servicesektoren ha begrensede overføringseffekter. Dette kan resultere i mindre økonomisk vekst og lavere attraktivitet for potensielle innflyttere og investorer. For det andre er mange av jobbene i servicesektoren deltidsjobber, sesongarbeidsjobber eller midlertidige kontrakter. Dette kan gjøre det mindre attraktivt for arbeidstakere å etablere seg og slå seg ned i områder dominert av servicesektorarbeid. Områder med høy andel servicesektorarbeid kan også være preget av lavere utdanningsnivå, færre offentlige tjenester og mindre attraktive boforhold. Disse faktorene kan påvirke befolkningens ønske om å bo og arbeide i slike områder. For det tredje kan det argumenteres for at vekst i sysselsettingen i servicesektoren ofte er en konsekvens av befolkningsvekst, fremfor en drivkraft bak den. Dette skyldes at etterspørselen etter tjenester innenfor denne sektoren øker som respons på befolkningsvekst. I motsetning til yrker innenfor

business- og produksjonssektoren, bidrar derfor ikke serviceyrker i samme grad til agglomerasjonseffekter.

Vi skal prøve å forklare intuisjonen for at koeffisienten er den samme i produksjonssektoren og servicesektoren, men mer signifikant når vi bruker Bartik-instrumentene. Ved å fjerne sjokk i sysselsetningsveksten, tar man bort variasjon i den uavhengige variabelens evne til å forutsi utfallet. Den overordnede tilpassede modellen kan gi samme linje med samme helning (og derfor samme koeffisient), men i hvilken grad modellen passer forbedres ved å fjerne støy i prediksjonsevnen til den uavhengige variabelen. På grunn av bedre tilpasningsevne (lavere p-verdi) vil Bartik-instrumenter være mer informative for å forutsi utfallet.

En siste bemerkning er at divergensen er ganske robust i alle spesifikasjonene. Koeffisienten foran initielt befolkningsnivå er noe lavere når vi bruker Bartik-instrumenter som mål på sysselsetting, istedenfor den vanlige *andel sektor*-variabelen. Det kan skyldes at Bartik-instrumenter forklarer mer av variasjonen i befolkningsveksten. I produksjons- og servicesektoren er koeffisienten den samme uansett hvilket mål på sysselsetting vi bruker. Dette betyr ikke nødvendigvis at endogeniteten som Bartik-instrumentene forsøker å eliminere, ikke er et problem. Det kan heller tyde på at disse sektorene har stabile mønstre av sysselsetting som er mindre følsomme for idiosynkratiske faktorer. Bartik-målingene er generelt mer presise som kan skyldes måten variablene er konstruert på. *Andel sektor*-variablene kan være mer påvirket av historiske forhold og beslutninger, noe som kan introdusere endogenitet i analysen. Bartik-instrumentene er mindre sannsynlig å lide av dette problemet og kan derfor anses som mer eksogene.

Tabell 7: Næringssektorer og Bartik-instrumenter

	(1)	(2)	(3)	(4)	(5)	(6)
	vekst	vekst	vekst	vekst	vekst	vekst
log befolkning 95	0.004*** (0.001)	0.003*** (0.001)	0.004*** (0.001)	0.003*** (0.001)	0.004*** (0.001)	0.003*** (0.001)
andel utdannet	0.037** (0.016)	0.038** (0.015)	0.045** (0.020)	0.035* (0.020)	0.029* (0.015)	0.034** (0.015)
andel produksjon	0.016** (0.006)					
Bartik produksjon		0.016*** (0.005)				
andel business			-0.041 (0.032)			
Bartik business				0.073* (0.041)		
andel service					-0.015** (0.007)	
Bartik service						-0.016*** (0.005)
Konstant	-0.046*** (0.005)	-0.041*** (0.005)	-0.045*** (0.007)	-0.036*** (0.006)	-0.030*** (0.009)	-0.024*** (0.008)
Regioner	90	90	90	90	90	90
R ²	0.491	0.506	0.465	0.467	0.486	0.508

Robuste standardavvik i parentes *** p<0.01, ** p<0.05, * p<0.1. Uten tidligere befolkning.

5.9 Bartik-instrumenter med tidligere befolkningstall

I dette kapittelet sjekker vi robustheten av resultatene i tabell 7 ved å kontrollere for tidligere befolkningsnivåer. Fra tidligere tabeller i denne oppgaven har vi sett at effekten av utdanning og sysselsetting har forsvunnet eller blitt redusert når vi har inkludert tidligere befolkningsnivåer. Det samme skjer i tabell 8. Utdanningseffekten er så og si forsvunnet, eller til og med negativ. Sysselsetting i produksjon- og businesssektoren har en positiv koeffisient og sysselsetting i servicesektoren har en negativ koeffisient. Selv om fortegnene er som forventet for alle koeffisientene foran sysselsettingsvariablene, er ingen av dem signifikante, og med lavere verdier enn i tabell 7 kan vi ikke tolke dette som kausal sammenheng.

Divergensen er høyere enn i regresjonene vi har kjørt som ikke kontrollerer for tidligere befolkningsnivåer. Av samme intuisjon som i delkapittel 5.6 og 5.7, tyder dette på at tidligere befolkningsnivåer er sterkt korrelert med befolkningsnivået i 1995 (se A5.2). Tidligere befolkningsnivåer inneholder informasjon om utdanning og sysselsetting i de ulike sektorene

som gjør at initielt befolkningsnivå i 1995 forklarer variasjonen i sysselsetting og utdanning. På grunn av multikollinearitet blir det vanskelig å tolke den isolerte effekten av utdanning og sysselsetting på veksten.

Resultatet kan tolkes på samme måte som i delkapittel 5.6. I sammenheng med næringsstruktur har den tidligere lokale etterspørselen (1970-1990) og den initielle sysselsettingen i de forskjellige sektorene i regionen hatt lite påvirkningskraft på befolkningsvekst i regioner i 1995-2013 når det kontrolleres for tidligere befolkningsnivåer. Med andre ord, nordmenn endrer ikke nødvendigvis sin bosituasjon ut ifra hvor sysselsettingen er eller har vært. Siden etterspørsel etter arbeidskraft bestemmes etter lønnen, som nevnt i Duranton (2016), så kan vi tolke videre at lønnsforskjellene innad de forskjellige næringssektorene, mellom regionene i Norge, har vært minimale for hvor arbeidstakere bosetter seg.

Oppsummert kan vi tolke resultatene i tabell 8 som at befolkningsvekst i Norge reagerer mer på tidligere forhold enn initielle forhold. De forsinkede reaksjonene fanges opp i koeffisienten foran *log befolkning 95* i alle spesifikasjonene. Resultatene av inkludering av tidligere befolkningsnivåer i alle tabellene våre fremhever betydningen av å vurdere tidligere forhold i forståelsen av befolkningsvekst. Vi har så vidt illustrert at befolkningsvekst i perioden 1995-2013 reagerer sterkere på den nasjonale sysselsettingsveksten i perioden 1970-1990. Men å inkludere tidligere befolkningsnivåer fjerner også den sterke effekten av Bartik-instrumenter på befolkningsveksten. Å kontrollere for tidligere nasjonal sysselsettingsvekst, samt andre historiske forhold kan gi mer håp til å forklare befolkningsdivergens.

Tabell 8: Kontrollerer for tidligere befolkningsnivåer.

	(1)	(2)	(3)	(4)	(5)	(6)
	vekst	vekst	vekst	vekst	vekst	vekst
log befolkning 95	0.026*** (0.004)	0.025*** (0.004)	0.026*** (0.004)	0.025*** (0.004)	0.026*** (0.004)	0.025*** (0.004)
andel utdannet	0.000 (0.021)	0.001 (0.020)	-0.021 (0.018)	-0.008 (0.018)	-0.003 (0.019)	-0.000 (0.020)
andel produksjon	0.007 (0.007)					
Bartik produksjon		0.006 (0.004)				
andel business			0.025 (0.026)			
Bartik business				0.005 (0.048)		
andel service					-0.008 (0.007)	
Bartik service						-0.006 (0.004)
Konstant	-0.029*** (0.006)	-0.027*** (0.005)	-0.023*** (0.005)	-0.026*** (0.005)	-0.021*** (0.007)	-0.021*** (0.005)
Regioner	90	90	90	90	90	90
Tidligere befolkning	Ja	Ja	Ja	Ja	Ja	Ja
R ²	0.723	0.724	0.720	0.717	0.725	0.724

Robuste standardavvik i parentes. *, **, ***: signifikant ved 1%, 5% og 10%.

6 Konklusjon og videre forskning

I denne masteroppgaven har vi utforsket flere metoder for å undersøke migrasjonen mellom 90 økonomiske regioner i Norge i løpet av perioder langt tilbake i tid, men hovedsakelig perioden 1995 til 2013. Datamaterialet omfatter utdannelsesnivå og sysselsetting knyttet til tre næringssektorer. Ved å analysere endringene i befolkningstallet i disse regionene fant vi at regioner som allerede hadde stor andel av befolkningen i initialåret har økt sin befolkningsandel ytterligere i løpet av denne tidsperioden.

De absolutte barroregresjonene viser tegn til absolutt befolkningsdivergens. Resultatet for absolutt divergens er sterkest for perioden 1995-2013. Resultatene fra spline- og kernel-tetthetsestimeringene viser at det er de mellomstore- og periferiregionene som har drevet divergensen. Dette er i tråd med resultatene til Rattsø og Stokke (2014). Det har blitt flere små og store regioner. Homogeniteten innad i mindre befolkningsgrupper gjør at den gjennomsnittlige, årlige befolkningsveksten er lavere enn i den generelle barroregresjonen for alle 90 regionene. I Duranton (2016) er koeffisienten for absolutt divergens mye større enn i vårt tilfelle. Det kan forklares med at Colombia er et utviklingsland, hvor positive beslutninger i stor grad favoriserer de urbane områdene. I Norge derimot, er regionene generelt mer homogene enn i Colombia.

Vi utvider de absolutte barroregresjonene til å inkludere ulike kontrollvariabler for humankapital og sysselsetting i produksjons-, business-, og servicesektoren. Resultatene for divergens er robuste i alle utvidelsene når vi ikke inkluderer tidligere befolkningsnivåer. Når vi derimot inkluderer tidligere befolkningsnivåer, blir den estimerte gjennomsnittlige, årlige befolkningsveksten mye større, og effektene av de andre kontrollvariablene fjernes. Dette forklares med at tidligere befolkningsnivåer fanger opp faktorer som drev vekst tidligere i Norge, og er sterkt korrelert med initielt befolkningsnivå i studieperioden vår. Det er derfor vanskelig å isolere effekten av ulikt utdanningsnivå og sysselsetting i produksjons-, business-, og servicesektoren når vi inkluderer tidligere befolkningsnivåer. Når vi derimot ikke kontrollerer for tidligere befolkningsnivåer, finner vi at humankapital har en positiv effekt på befolkningsveksten. Det samme gjelder andel sysselsatte i produksjons- og businesssektoren. Disse tre faktorene bidrar til positive agglomerasjonseffekter som driver befolkningsvekst. Serviceyrker bidrar ikke i like stor grad til agglomerasjonseffekter, og har en negativ effekt på befolkningsveksten.

Lokal arbeidskraftetterspørsel som ikke drives av regionspesifikke arbeidstilbudssjokk i tråd med Bartik (1991), er et bedre mål på sysselsetting i de tre ulike sektorene. Selv om koeffisientene ikke forandrer seg i produksjons- og servicesektoren når vi bruker Bartik-instrumenter, er likevel resultatene mer signifikante. I tillegg er Bartik i businesssektoren mindre korrelert med utdanningsvariabelen enn den vanlige *andel sektor*-variabelen, og vi klarte dermed å isolere mer av effekten av utdanning og sysselsetting i businesssektoren på befolkningsveksten. Bartik-instrumenter bekrefter dermed resultatene fra tidligere litteratur som sier at sysselsetting i produksjons- og businesssektoren har positiv effekt på befolkningsveksten, men sysselsetting i servicesektoren ikke har det. At koeffisientene ikke forandret seg i produksjons- og servicesektoren tyder på at barroregresjoner er en bra metode innen urban vekstteori i Norge for å isolere effekten av kontrollvariabler på befolkningsveksten.

Denne masteroppgaven gir et nyttig bidrag til forståelsen av befolkningsvekst og migrasjon i norske regioner, og den åpner opp for mange spennende muligheter for videre forskning. Bartik-instrumenter i barroregresjoner har vært lite brukt i analyser av urban vekst. Vi har vist at denne variabelen bidrar til en eksogen kilde i variasjonen i lokal etterspørsel etter arbeidskraft. Tidligere befolkningsnivåer fjerner effekten av kontrollvariabler fra initialåret, og er høyt korrelert med initielt befolkningsnivå. Det kan dermed være verdifullt å gjennomføre ytterligere forskning på hvordan historiske befolkningstrender påvirker initielle faktorer. Den høye korrelasjonen mellom initielt befolkningsnivå og tidligere befolkningsnivåer, gjør at å kontrollere for historisk variasjon i dataene i barroregresjoner ikke er særlig informativt for å forklare divergensen i norske regioner. I kapittel 5.9 konkluderte vi med at befolkningsvekst i Norge reagerer mer på tidligere forhold enn initielle forhold. Med dette oppfordrer vi til videre forskning for å avdekke mer om den underliggende dynamikken til disse tidligere forholdene. I den forbindelse kan det være interessant også å vurdere sosiale og kulturelle faktorer. Forskning på dette området kan være informativt for å forstå hvordan tidligere forhold former fremtidig befolkningsvekst, og vil være et verdifullt bidrag til feltet om økonomisk- og urban vekst.

7 Kilder

- Alonso, W. (1964). *Location of Land Use; Toward a General Theory of Land Rent*.
Cambridge, MA: Harvard University Press.
<https://doi.org/10.4159/harvard.9780674730854>
- Autor, D. H., Dorn, D., & Hanson, G. H. (2013). The China Syndrome: Local Labor Market Effects of Import Competition in the United States. *The American Economic Review*, *103*(6), 2121–2168. DOI: 10.1257/aer.103.6.2121
- Ayuda, M. I., Collantes, F., & Pinilla, V. (2010). Long-run regional population disparities in Europe during modern economic growth: a case study of Spain. *The Annals of Regional Science*, *44*(2), 273–295. DOI 10.1007/s00168-008-0260-9
- Barro, R. J. (1991). Economic Growth in a Cross Section of Countries. *The Quarterly Journal of Economics*, *106*(2), 407–443. <https://doi.org/10.2307/2937943>
- Barro, R. J., & Sala-i-Martin, X. (1992). Convergence. *Journal of Political Economy*, *100*(2), 223–251. <https://doi.org/10.1086/261816>
- Bartik, T. J. (1991). Who benefits from state and local economic development policies?. DOI 10.17848/9780585223940
- Baum-Snow, N., & Pavan, R. (2013). Inequality and City Size. *The Review of Economics and Statistics*, *95*(5), 1535–1548. https://doi.org/10.1162/REST_a_00328
- Baumol, W. J. (1986). Productivity Growth, Convergence, and Welfare: What the Long-Run Data Show. *The American Economic Review*, *76*(5), 1072–1085.
- Beeson, P. E., DeJong, D. N., & Troesken, W. (2001). Population growth in U.S. counties, 1840–1990. *Regional Science and Urban Economics*, *31*(6), 669–699.
[https://doi.org/10.1016/s0166-0462\(01\)00065-5](https://doi.org/10.1016/s0166-0462(01)00065-5)
- Bhuller, M. S. (2009). Inndeling av Norge i arbeidsmarkedsregioner. *Statistisk Sentralbyrå(SSB)*.
https://www.ssb.no/a/publikasjoner/pdf/notat_200924/notat_200924.pdf
- Black, D., & Henderson, V. (2003). Urban evolution in the USA. *Journal of Economic Geography*, *3*(4), 343–372.
- Blanchard, O., & Katz, L. (1992). Regional Evolutions. *Brookings Papers on Economic Activity*, *23*(1), 1–76.
- Cameron, C. A., & Trivedi, P. K. (2005). *Microeconometrics : methods and applications*. Cambridge University Press.

- Card, D. (2009). Immigration and Inequality. *American Economic Review*, 99(2), 1–21.
<https://doi.org/10.1257/aer.99.2.1>
- Chinitz, B. (1961). Contrasts in Agglomeration: New York and Pittsburgh. *The American Economic Review*, 51(2), 279–289.
https://www.jstor.org/stable/pdf/1914493.pdf?refreqid=excelsior%3A89b04659c776907f53a8b9ff0e515eee&ab_segments=&origin=&initiator=
- Cingano, F., & Schivardi, F. (2004). Identifying the Sources of Local Productivity Growth. *Journal of the European Economic Association*, 2(4), 720–742.
- Combes, P. P. (2000). Economic structure and local growth: France, 1984–1993. *Journal of urban economics*, 47(3), 329–355.
- Combes, P.-P., Duranton, G., & Gobillon, L. (2010). The identification of agglomeration economies. *Journal of Economic Geography*, 11(2), 253–266.
<https://doi.org/10.1093/jeg/lbq038>
- De Long, J. B. (1988). Productivity Growth, Convergence, and Welfare: Comment. *The American Economic Review*, 78(5), 1138–1154.
<https://www.jstor.org/stable/pdf/1807174.pdf>
- Dupton, W. (2013). *Stata Manuel: mkspline*.
https://www.stata.com/manuals13/rmkspline.pdf?fbclid=IwAR05vgf-Et1P56xchUWEQvtAtbbIRxsEonPng_iTJrqLVJcpVwxb43QIRE
- Duranton, G. (2007). Urban Evolutions: The Fast, the Slow, and the Still. *American Economic Review*, 97(1), 197–221. <https://doi.org/10.1257/aer.97.1.197>
- Duranton, G. (2008). Viewpoint: From Cities to Productivity and Growth in Developing Countries. *The Canadian Journal of Economics / Revue Canadienne D'Économique*, 41(3), 689–736.
- Duranton, G. (2016). Determinants of city growth in Colombia. *Papers in Regional Science*, 95(1), 101–131. <https://doi.org/10.1111/pirs.12225>
- Duranton, G., & Puga, D. (2004). Micro-foundations of urban agglomeration economies. In *Handbook of regional and urban economics* (Vol. 4, pp. 2063–2117). Elsevier.
- Duranton, G., & Puga, D. (2013). The growth of U.S. cities. *Revue d'Économie Régionale & Urbaine*, 5, 857. <https://doi.org/10.3917/reru.135.0857>
- Durlauf, S. N., Johnson, P. A., & Temple, J. R. W. (2005). Growth econometrics. In *Handbook of economic growth* (Vol. 100, Issue 1, pp. 1–183). Amsterdam. North-Holland. [https://doi.org/10.1016/s0304-4076\(00\)00055-5](https://doi.org/10.1016/s0304-4076(00)00055-5)

- Fan, J., & Gijbels, I. (1996). *Monographs on Statistics and Applied Probability*. (Local Polynomial Modelling and Its Applications). Chapman & Hall/CRC.
- Forente Nasjoner . (2019). World Urbanization Prospects: The 2018 Revision. *Population and Development Review*, 24(4), 883. <https://doi.org/10.2307/2808041>
- Gilles Duranton, & Puga, D. (2004). *Micro-foundations of urban agglomeration economies*. National Bureau Of Economic Research.
- Glaeser, E. L., Kallal, H. D., Scheinkman, J. A., & Shleifer, A. (1992). Growth in Cities. *Journal of Political Economy*, 100(6).
- Glaeser, E. L., & Kerr, W. R. (2009). Local Industrial Conditions and Entrepreneurship: How Much of the Spatial Distribution Can We Explain? *Journal of Economics & Management Strategy*, 18(3), 623–663. <https://doi.org/10.1111/j.1530-9134.2009.00225.x>
- Glaeser, E. L., & Saiz, A. (2004). The rise of the skilled city. *Brookings-Wharton Papers on Urban Affairs*, 2004(1), 47–105. <https://doi.org/10.1353/urb.2004.0005>
- Glaeser, E. L., & Shapiro, J. M. (2003). Urban Growth in the 1990s: Is City Living Back? *Journal of Regional Science*, 43(1), 139–165. <https://doi.org/10.1111/1467-9787.00293>
- Glaeser, E., Scheinkman, J., & Shleifer, A. (1995). Economic Growth in a Cross-Section of Cities. *Journal of Monetary Economics*, 36(1), 117–143. [https://doi.org/10.1016/0304-3932\(95\)01206-2](https://doi.org/10.1016/0304-3932(95)01206-2)
- Goldsmith-Pinkham, P., Sorkin, I., & Swift, H. (2020). Bartik Instruments: What, When, Why, and How. *The American Economic Review*, 110(8), 2586–2624. DOI: 10.1257/aer.20181047
- Granja, J., Makridis, C., Yannelis, C., & Zwick, E. (2022). Did the paycheck protection program hit the target? *Journal of Financial Economics*, 145(3), 725–761. <https://doi.org/10.1016/j.jfineco.2022.05.006>
- Hustoft, A., Hartvedt, H., Nymoene, E., Stålnacke, M., & Utne, H. (1999). *Standard for økonomiske regioner*. Statistisk sentralbyrå.
- Kilden. (2005). *Rett til lønnet arbeid*. Kvinnehistorie.no. <https://www.kvinnehistorie.no/artikkel/t-3939>
- Klein, A., & Crafts, N. (2020). Agglomeration externalities and productivity growth: US cities, 1880–1930. *The Economic History Review*, 73(1), 209–232. <https://doi.org/10.1111/ehr.12786>

- Mejía, D., Ramírez, T., & Tamayo, J. (2008). The Demographic Transition in Colombia: Theory and Evidence. *Borradores de Economía*, 538.
- Michaels, G., Rauch, F., & Redding, S. J. (2012). Urbanization and Structural Transformation. *The Quarterly Journal of Economics*, 127(2), 535–586. <https://doi.org/10.1093/qje/qjs003>
- Mills, E. S. (1967). An Aggregative Model of Resource Allocation in a Metropolitan Area. *The American Economic Review*, 57(2), 197–210.
- Moretti, E. (2004). Estimating the social return to higher education: Evidence from longitudinal and repeated cross-sectional data. *Journal of Econometrics*, 121(1-2), 175–212. <https://doi.org/10.1016/j.jeconom.2003.10.015>
- Muth, R. F. (1969). Cities and housing: The spatial pattern of urban residential land use. *Chicago and London: University of Chicago Press*. DOI: 10.4236/jep.2020.118033
- Parzen, E. (1962). On Estimation of a Probability Density Function and Mode. *The Annals of Mathematical Statistics*, 33(3), 1065–1076.
- Pinilla, V., & Ayuda, M.-I. . (2010). Taking advantage of globalization? Spain and the building of the international market in Mediterranean horticultural products, 1850-1935. *European Review of Economic History*, 14(2), 239–274. <https://doi.org/10.1017/s136149161000002x>
- Porter, M. E. (1998). *Clusters and the New Economics of Competition*. Harvard Business Review. <https://hbr.org/1998/11/clusters-and-the-new-economics-of-competition>
- Quah, D. (1993). Galton’s Fallacy and Tests of the Convergence Hypothesis. *The Scandinavian Journal of Economics*, 95(4), 427. <https://doi.org/10.2307/3440905>
- Rattsø, J., & Stokke, H. E. (2014). Population Divergence and Income Convergence: Regional Distribution Dynamics for Norway. *Regional Studies*, 48(11), 1884–1895. <https://doi.org/10.1080/00343404.2013.799842>
- regjeringen.no. (2021). *Hva EØS-avtalen omfatter*. Regjeringen.no. <https://www.regjeringen.no/no/tema/europapolitikk/eos1/hva-avtalen-omfatter/id685024/>
- Schultz, T. W. (1961). Investment in Human Capital. *The American Economic Review*, 51(1), 1–17. <https://www.jstor.org/stable/1818907>
- Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. Chapman & Hall. <https://ned.ipac.caltech.edu/level5/March02/Silverman/paper.pdf>

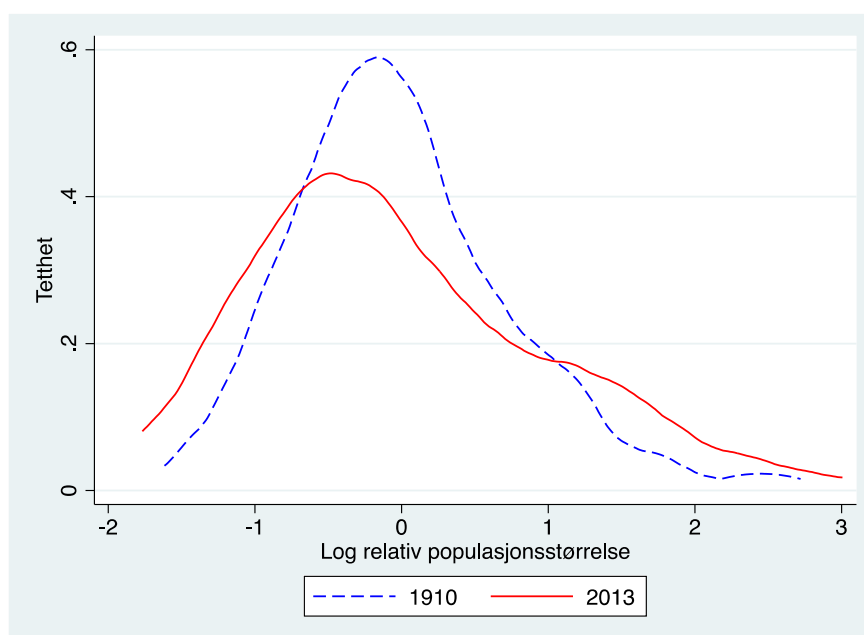
- Simon, C. J., & Nardinelli, C. (2002). Human capital and the rise of American cities, 1900–1990. *Regional Science and Urban Economics*, 32(1), 59–96.
[https://doi.org/10.1016/s0166-0462\(00\)00069-7](https://doi.org/10.1016/s0166-0462(00)00069-7)
- Solow, R. M. (1956). A Contribution to the Theory of Economic Growth. *The Quarterly Journal of Economics*, 70(1), 65–94. <https://doi.org/10.2307/1884513>
- SSB. (2009). *En krevende telling etter krigen*. SSB.no.
<https://www.ssb.no/befolkning/artikler-og-publikasjoner/en-krevende-telling-etter-krigen>
- Utvik, T. I. (2022). *Befolkningsutvikling og flytting i Norge på 1900-tallet - Historie (PB) - NDLA*. Ndl.no. <https://ndla.no/nb/subject:846a7552-ea6c-4174-89a4-85d6ba48c96e/topic:1c4d1853-9c9e-4320-ad19-1814090a911a/topic:ff9eb1ca-85aa-414c-9ddb-c23c41c89655/resource:abd567d5-5658-40a7-ba22-8b132ea5978b>
- Wooldridge, J. M. (2015). *Introductory Econometrics: A Modern Approach* (6th ed.). Cengage Learning.
- Zucchini, W. (2003). Applied smoothing techniques. *Part 1: Kernel density estimation*, 15, 1-20.

8 Appendiks

A1 Sektorer fordelt etter industri

Produksjonssektoren	01-03 Jordbruk, skogbruk og fiske 05-09 Bergverksdrift og utvinning 10-33 Industri 35-39 Elektrisitet, vann og renovasjon 41-43 Bygg- og anleggsvirksomhet
Businesssektoren	64-66 Finansiering og forsikring 68-75 Teknisk tjenesteyting, eiendomsdrift 77-82 Forretningsmessig tjenesteyting
Servicesektoren	45-47 Varehandel, reparasjon av motorvogner 49-53 Transport og lagring 55-56 Overnattings- og servicevirksomhet 58-63 Informasjon og kommunikasjon 84 Offentlig administrasjon, forsvar, sosial forsikring 85 Undervisning 86-88 Helse- og sosialtjenester 90-99 Personlig tjenesteyting

A2 Kernel-tetthetsfunksjoner for 1910 og 2013



A3 Uten tidligere befolkningsnivåer

Deler utdanningsnivå på arbeidsdyktige i regionen i 1995. OLS-resultater uten tidligere befolkningsnivåer.

	(1)	(2)	(3)	(4)	(5)	(6)
	vekst	vekst	vekst	vekst	vekst	vekst
log befolkning 95	0.004*** (0.000)	0.004*** (0.001)	0.004*** (0.001)	0.004*** (0.000)	0.005*** (0.000)	0.004*** (0.001)
andel utdannet		0.025 (0.015)	0.004 (0.003)	0.033 (0.042)	0.028 (0.021)	0.069 (0.057)
andel utdannet_sq						-0.118 (0.122)
Konstant	-0.042*** (0.005)	-0.039*** (0.005)	-0.026** (0.012)	-0.046*** (0.008)	-0.054*** (0.011)	-0.042*** (0.006)
Utdanning		univ.	loguniv.	grunnskole	vgs.	univ.
Regioner	90	90	90	90	90	90
R ²	0.391	0.401	0.404	0.398	0.410	0.404

Robuste standardavvik i parentes. *, **, ***: signifikant ved 1%, 5% og 10%.

A4 Annet mål på humankapital

Deler utdanningsnivå på sysselsatte i regionen i 1995. OLS-resultater.

	(1)	(2)	(3)	(4)	(5)	(6)
	vekst	vekst	vekst	vekst	vekst	vekst
log befolkning 95	0.004*** (0.001)	0.025*** (0.004)	0.025*** (0.004)	0.025*** (0.004)	0.025*** (0.004)	0.025*** (0.004)
andel utdannet	0.007 (0.014)	-0.003 (0.015)	-0.001 (0.004)	0.024 (0.025)	-0.020 (0.021)	-0.042 (0.069)
andel utdannet_sq						0.075 (0.124)
Konstant	-0.041*** (0.005)	-0.025*** (0.005)	-0.028*** (0.009)	-0.030*** (0.006)	-0.009 (0.017)	-0.021*** (0.008)
Tidligere befolkning	Nei	Ja	Ja	Ja	Ja	Ja
Utdanning	univ.	univ.	loguniv.	grunnskole	vgs.	univ.
Regioner	90	90	90	90	90	90
R ²	0.393	0.632	0.632	0.640	0.639	0.634

Robuste standardavvik i parentes. *, **, ***: signifikant ved 1%, 5% og 10%. Tidligere befolkning er log befolkning for 1910, 1930, 1954 og 1969.

A5 Korrelesasjonsmatriser

A5.1 Korrelasjonsmatrise utdanning og næringssektorer

	andel utdannet	andel produksjon	andel i business	andel i service
andel utdannet	1.0000			
andel produksjon	-0.3200	1.0000		
andel business	0.8272	-0.3427	1.0000	
andel service	0.0247	-0.9334	-0.0171	1.0000

A5.2 Korrelasjonsmatrise befolkning

	logbef10	logbef30	logbef54	logbef69	logbef95
logbef10	1.0000				
logbef30	0.9786	1.0000			
logbef54	0.9669	0.9916	1.0000		
logbef69	0.9379	0.9712	0.9910	1.0000	
logbef95	0.9156	0.9450	0.9688	0.9880	1.0000

A5.3 Korrelasjonsmatrise utdanning og Bartik-instrumenter

	andel utdannet	Bartik produksjon	Bartik business	Bartik service
andel utdannet	1.0000			
Bartik produksjon	-0.1120	1.0000		
Bartik business	0.7830	0.0441	1.0000	
Bartik service	-0.0171	-0.9868	-0.2052	1.0000

A6 Endring i sysselsetting i olje- og gassnæringen og bygg- og anleggsvirksomhet

	1995	2013	Endring
Utvinning av råolje og naturgass	16177	26684	65%
Tjenester tilknyttet utvinning av råolje og naturgass	4437	35676	704%
Rørtransport	527	365	-31%
SUM olje- og gass	21141	62725	197%
SUM bygg- og anleggsvirksomhet	99235	212354	114%

Tall hentet fra SSB.

