

# Høyere utdanningsinstitusjoner og regional befolkningsvekst

Torstein Otterlei Fjørtoft

Institutt for samfunnsøkonomi (ISØ)

Norges teknisk-naturvitenskapelige universitet

Juni 2014

# Forord

Jeg ønsker å takke alle som har hjulpet meg med oppgaven. Spesielt vil jeg takke min veileder Jørn Rattsø. Rattsø satte meg på sporet av problemstillingen, og har gitt konstruktive tilbakemeldinger underveis. Takk også til de som har lest korrektur på oppgaven og forhåpentligvis rettet opp de fleste skrivefeilene mine.

Trondheim, 31. Mai 2014

Torstein Otterlei Fjørtoft

# Innhold

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>Innledning</b>  | <b>1</b>  |
| 1.1      | Problemstilling . . . . .  | 1         |
| 1.2      | Kort om norske forhold og motivasjon for oppgaven . . . . .            | 1         |
| 1.3      | Sammendrag . . . . .   | 2         |
| 1.4      | Oppgavens oppbygning . . . . .   | 2         |
| <b>2</b> | <b>Bakgrunn, teori og tidligere empirisk forskning</b>                 | <b>4</b>  |
| 2.1      | Tilbudseffekter . . . . .  | 5         |
| 2.2      | Etterspørselseffekter . . . . .  | 6         |
| 2.3      | Agglomerasjonseffekter . . . . .                                       | 7         |
| 2.3.1    | Humankapital og læring . . . . .                                       | 8         |
| 2.4      | Tidligere empirisk forskning . . . . .                                 | 9         |
| 2.4.1    | Høgskolers og universiteters påvirkning på regionaløkonomien . . . . . | 9         |
| 2.4.2    | Humankapital og vekst . . . . .  | 11        |
| 2.5      | Norske undersøkelser . . . . .   | 12        |
| <b>3</b> | <b>Datamateriale</b>   | <b>14</b> |
| 3.1      | Økonomiske regioner . . . . .  | 15        |
| 3.2      | Høyere utdanningsinstitusjoner og studenttall . . . . .                | 16        |
| 3.2.1    | Dummyvariabler for studiested . . . . .                                | 17        |
| 3.2.2    | Studenttall . . . . .  | 19        |
| 3.3      | Befolkning . . . . .   | 20        |
| 3.4      | Statiske kontrollvariabler . . . . .                                   | 21        |
| <b>4</b> | <b>Empirisk tilnærming</b>   | <b>23</b> |
| 4.1      | Spesifisering av paneldatamodeller . . . . .                           | 23        |
| 4.1.1    | Dummyvariabler for utdanningsinstitusjon . . . . .                     | 24        |
| 4.1.2    | Faste regionseffekter . . . . .  | 25        |
| 4.2      | Estimeringsmetoder . . . . .   | 26        |
| 4.2.1    | Sammenkoblet MKM . . . . .   | 27        |
| 4.2.2    | Faste effekter . . . . .   | 27        |

|          |   |           |
|----------|---|-----------|
| 4.3      | Økonometriske utfordringer . . . . .                                | 29        |
| 4.3.1    | Simultanitet . . . . .  | 29        |
| 4.3.2    | Utelatte variabler . . . . .  | 31        |
| 4.3.3    | Heteroskedastisitet og seriekorrelasjon . . . . .                   | 33        |
| 4.3.4    | Faste effekter og lagget avhengig variabel . . . . .                | 34        |
| <b>5</b> | <b>Empiriske resultater</b>   | <b>35</b> |
| 5.1      | Virkning av utdanningsinstitusjoner på befolkningsveksten . . . . . | 36        |
| 5.1.1    | Små høgskoleregioner sammenlignet mot andre små regioner . . . . .  | 39        |
| 5.2      | Effekt av studenttall på befolkning . . . . .                       | 40        |
| 5.2.1    | Faste effekter . . . . .  | 42        |
| 5.2.2    | Kort- og langtidseffekter . . . . .                                 | 44        |
| 5.2.3    | Regionale forskjeller . . . . .                                     | 46        |
| <b>6</b> | <b>Oppsummering</b>   | <b>50</b> |

# Tabeller

|     |  |    |
|-----|--|----|
| 3.1 | Utvikling i hovedvariablene for alle regioner over hele studieperioden . . .   | 17 |
| 3.2 | Studenttall etter regionstørrelse. . . . .   | 18 |
| 3.3 | Gjennomsnittlig befolkningsvekst i regioner med ny høgscole sett opp mot alle andre regioner . . . . .                     | 19 |
| 3.4 | Gjennomsnittlig befolkningsvekst i 12 små høgscoleregioner sammenlignet med andre små regioner . . . . .                   | 20 |
| 3.5 | Befolkningsvekst i prosent i regioner med og uten høyere utdanningsinstitusjoner . . . . .                                 | 21 |
| 3.6 | Befolkningsvekst i prosent i små, mellomstore og store regioner . . . . .  | 21 |
| 5.1 | Resultater: Effekter av ny høgscole. . . . .   | 37 |
| 5.2 | Resultater: Effekter av ny og utvidet høgscole. . . . .  | 38 |
| 5.3 | Resultater: Små regioner med høgscole sammenlignet med små regioner uten. . . . .  | 39 |
| 5.4 | Resultater: Sammenkoblet MKM - Grunnmodell . . . . .   | 41 |
| 5.5 | Resultater: Grunnmodell med faste effekter . . . . .   | 43 |
| 5.6 | Resultater: Faste effekter på endringsform . . . . .   | 44 |
| 5.7 | Resultater: Korttids- og langtidseffekter - MKM (1) og (2), Faste effekter (3) og (4) . . . . .                            | 45 |
| 5.8 | Resultater: Faste effekter med oppdelt datasett. (1) Små regioner, (2) mellomstore regioner, (3) store regioner. . . . .   | 47 |
| 5.9 | Resultater: Sammenkoblet MKM med oppdelt datasett. (1) Små regioner, (2) mellomstore regioner, (3) store regioner. . . . . | 48 |

# Kapittel 1

## Innledning

### 1.1 Problemstilling

Målet med oppgaven er å finne ut hvilken effekt høyere utdanningsinstitusjoner har på regional befolkningsvekst i Norge.

### 1.2 Kort om norske forhold og motivasjon for oppgaven

Norge har en desentralisert struktur for høyere utdanning sammenlignet med de fleste andre land. Et av hovedmålene i norsk utdanningspolitikk er at høyere utdanning skal være lett tilgjengelig for hele befolkningen, og et viktig virkemiddel er å ha institusjoner for høyere utdanning spredt rundt i hele landet (NOU, 2008:3).

En desentralisert struktur er også et mål i regionalpolitikken. Høgskolene skaper økonomisk vekst i distriktene de ligger i (Raymond Florax, 1992), og kan hjelpe regionene med å opprettholde økonomisk aktivitet og bosetning. Det finnes likevel argumenter for en mer sentralisert struktur. I følge NOU 2008:3 har små studiesteder ofte mindre fagmiljøer, som kan gi dårligere forskning og utdanningskvalitet. Ved å konsentrere utdanningsressursene på færre læresteder, er det sannsynlig at kvaliteten på forskning og utdanning kan bedres. Det vil da oppstå en avveining der distriktspolitikk og bred tilgang på utdanning må vurderes opp mot kvaliteten på utdanningen<sup>1</sup>. Hvor stor effekt utdanningsinstitusjonene har på den regionale økonomien kan derfor være viktig når ressursene skal allokeres.

Det er utført flere analyser i Norge som utforsker utdanningsinstitusjonenes påvirkning

---

<sup>1</sup>Diskuteres i NOU 2008:3, kapittel 11.

på regionaløkonomien. Såvidt meg bekjent har ingen tatt i bruk samme metoder som jeg benytter meg av. Masteroppgaven baserer seg på økonometriske strategier, tidligere benyttet i utenlandske empiriske studier om emnet. Datamaterialet hentes fra NSDs Kommunedatabase, som gir tilgang på data på kommunenivå. Ved å benytte data for folkemengde og studenttall i regionene, er målet at oppgaven skal se problemstillingen i et nytt lys, sammenlignet med de eksisterende norske analysene. Jeg benytter studenttallet i regionene til å finne ut hvorvidt regionene har høyere utdanningsinstitusjoner, og eventuelt når disse ble opprettet eller utvidet. Jeg gjør også analyser der jeg bruker studenttallet i regionen som et mål på størrelsen av utdanningsinstitusjonene.

Empirisk er oppgavens hovedutfordring å håndtere endogenitet i modellene på en tilfredsstillende måte, og samtidig utnytte nok variasjon i datamaterialet til å hente ut relevant informasjon. Det er opplagt en korrelasjon mellom folkemengde og studenttall i regionene. I hvilken grad høyere utdanningsinstitusjoner har en kausal effekt på befolkningsveksten er derimot mer usikkert. Jeg vil derfor forsøke å benytte metoder som reduserer skjevhet i estimatene, samtidig som tilstrekkelig variasjon beholdes til at eventuelle effekter kan identifiseres.

### 1.3 Sammendrag

Analysene kan deles inn i to hovedretninger. Den ene delen ser direkte på virkningen av utdanningsinstitusjoner på befolkningsveksten i regionen, og sammenligner hovedsakelig på tvers av regioner. Disse analysene gir generelt lite signifikante effekter. Jeg finner noen positive korttidseffekter som er noenlunde signifikante, men ingen signifikante langtidseffekter.

Den andre analyseformen bruker studenttallet som en proxy for omfanget av utdanningsinstitusjonene i regionen. Disse analysene gir mer signifikante effekter. De mest troverdige modellene indikerer at én ekstra student i en region, øker befolkningen i regionen med litt over én i neste periode. Begrensninger i datamaterialet gjør at selv om oppgaven ikke finner enkelte kausale sammenhenger, kan det ikke konkluderes med at disse ikke eksisterer.

### 1.4 Oppgavens oppbygning

Kapittel 2 gir en gjennomgang av relevant teori og tidligere studier. Teorien fokuserer på de regionale virkningene av utdanningsinstitusjoner. Vekstteori og agglomerasjonseffekter er sentrale. Etter den grunnleggende teorien, presenteres noen relevante tidligere

empiriske studier. Kapitlet avsluttes med en gjennomgang av en norsk analyse, for å gi et overblikk over norske forhold.

I kapittel 3 presenteres datagrunnlaget. Her går jeg igjennom variablene som studien bygger på, og hvordan de geografiske enhetene og tidsperiodene defineres.

Kapittel 4 inneholder metodedelen. Her blir den empiriske strategien forklart. Først spesifiserer jeg paneldatamodellene som danner utgangspunktet for de ulike analysene. Deretter tar jeg for meg estimeringsmetodene og de nødvendige forutsetningene for forventningsrette og effisiente estimater. Til slutt går jeg inn på de økonometriske utfordringene i studien, som legger grunnlaget for diskusjonen om hvorvidt resultatene er pålitelige.

Resultatene blir presentert og diskutert i kapittel 5. Kapitlet starter med analyser om hvordan tilstedeværelse av høyere utdanningsinstitusjoner påvirker veksten i en region. Disse analysene bruker hovedsakelig variasjon på tvers av regioner, og ser om opprettelser og utvidelser av høyskoler og universiteter endrer befolkningsveksten i regionen i forhold til andre regioner.

Resten av kapitlet fokuserer på analyser der studenttallet i en region er interessevariabel. Her presenteres først en grunnleggende analyse for å gi overblikk. Deretter benytter jeg mer avanserte modeller, for å utforske datamaterialet og øke presisjonen i estimatene. Analysene som benytter studenttallet i regionen som interessevariabel, vil i større grad utnytte variasjon over tid innenfor regioner. Oppgaven avsluttes med en oppsummering i kapittel 6.



## Kapittel 2

# Bakgrunn, teori og tidlige empirisk forskning

Høyere utdanningsinstitusjoner bidrar til økonomisk vekst i samfunnet ved at de gir kunnskap og ferdigheter til befolkningen. Raymond Florax (1992) skriver at kunnskap i stadig større grad blir sett på som en av de viktigste produksjonsfaktorene i moderne industrialiserte økonomier. Universitetene og høyskolene blir derfor sett på som viktige tilskudd i regionene. De kan øke humankapitalen i regionen ved at flere av innbyggerne får en utdanning.

Økt humankapital i regionen kalles tilbudseffekten av utdanningsinstitusjoner, fordi de virker inn på tilbudssiden i økonomien. Flere utdannede mennesker gir bedre tilgang på produksjonsfaktorer, som gjør produksjonen mer effektiv<sup>1</sup>.

Utdanningsinstitusjonene påvirker også etterspørselssiden i regionaløkonomien. Høgskoler og universiteter skaper etterspørsel direkte som statlig arbeidsplass, men også indirekte ved at de ansatte, studentene og institusjonen etterspør varer og tjenester i nærområdet<sup>2</sup>.

Det er gjennom disse to hovedkanalene utdanningsinstitusjonene påvirker regionaløkonomien (Florax, 1992). Når høgskoler og universiteter vurderes som virkemiddel for å stimulere regional vekst, er det hovedsakelig disse effektene som avgjør hvor godt det fungerer. Tilbudseffektene blir gjennomgått i delkapittel 2.1, der jeg fokuserer på forholdet mellom utdanning, humankapital og økonomisk vekst, og diskuterer hvorvidt virkningene er lokale nok til at de fanges opp i små geografiske inndelinger. Etterspørselseffektene diskuteres i delkapittel 2.2. I delkapittel 2.3 går jeg gjennom litt teori om agglomerasjonseffekter,

---

<sup>1</sup>Florax (1992) kaller tilbudseffektene for 'knowledge impacts'. De blir også referert til som langtidsvirkninger og indirekte virkninger.

<sup>2</sup>Florax (1992) kaller etterspørselseffekter for 'expenditure impacts'. Også referert til som korttidsvirkninger og direkte virkninger.

og diskuterer hvordan høyere utdanningsinstitusjoner kan påvirke regional vekst gjennom disse.

I delkapittel 2.4 presenterer jeg noen empiriske studier som er relevante for denne masteroppgaven. Jeg fokuserer mest på studier som ser etter effekter av utdanningsinstitusjoner på regionaløkonomien, men tar også med to studier som ser på sammenhengen mellom humankapital og befolkningsvekst. Kapitlet avsluttes med en gjennomgang av en norsk undersøkelse.

## 2.1 Tilbudseffekter

Utdanningsinstitusjonene påvirker tilbudssiden først og fremst ved at de utdanner studenter. Mange universiteter har i tillegg et forskningsmiljø som kan være verdifullt for nærliggende bedrifter. Duranton og Puga (2003) tar opp at læring i mange tilfeller involverer kontakt med andre mennesker. Dette innebærer at tilstedeværelsen av en høyere utdanningsinstitusjon også mer indirekte kan øke humankapitalen i en region.

Humankapital har lenge blitt sett på som sentralt for økonomisk vekst, og er inkludert i de fleste populære vekstmodeller i økonomisk teori. I Solow-Swan-modellen fra 1956, som er en av de mest grunnleggende modellene innen vekstteori, er det teknologisk utvikling som driver veksten på lang sikt. Teknologisk utvikling er her eksogent bestemt. Flere senere utvidelser har endogenisert teknologi og humankapital.

N. Gregory Mankiw, David Romer og David Weil kom med en utvidelse i 1992 som inkluderte humankapital som produksjonsfaktor, i tillegg til fysisk kapital og arbeidskraft:

$$Y = AK^\alpha H^\beta L^{1-\alpha-\beta} \quad (2.1)$$

I ligning 2.1 representerer H humankapitalen, A teknologisk utvikling, K fysisk kapital og L arbeidskraft. Økt humankapital gjør at regionene kan utnytte teknologien og den fysiske kapitalen bedre. Modeller som (2.1) benyttes ofte til å se på forskjeller mellom land, men teorien kan også benyttes på mindre regioner. Dersom høyere utdanningsinstitusjoner øker nivået på humankapital i en region, gir dette ifølge modellen høyere økonomisk vekst.

De økonomiske regionene i denne masteroppgaven er små, geografisk sett. Ettersom flytting innenriks er vanligere enn flytting mellom land, kan det spørres om i hvor stor grad høgskolene og universitetene øker humankapitalen i egen region og hvor mye som slipper ut til andre regioner. Tidligere forskning på hvorvidt nyutdannede blir værende i sin studieregion, viser blandede resultater. Ifølge Winters (2011) vokser befolkningen i regioner

med høy humankapital fordi noen av de tilflyttede studentene blir boende. Venhorst et al. (2010) peker på arbeidsmarkedet i en region som en viktig faktor. De finner at lokale bedrifter innen mange felt ofte lykkes i å holde på de beste studentene.

En rapport av Lasse Sigbjørn Stambøl (2011) ser på antallet av tilflyttede studenter ved høyskoler og universiteter i Norge i 1994 som fremdeles var bosatte i studieregionen i 2000 og 2005. Undersøkelsen viser at små høyskoleregioner, der høyskolene utgjør en stor del av regionaløkonomien<sup>3</sup>, opplever at få tilflyttede studenter som blir værende. I gjennomsnitt for disse regionene ble bare 2,9 prosent igjen. Det kan likevel tenkes at høyskolene har bidratt til å holde på unge fra regionen som ellers ville ha reist til et annet sted for å studere. Selv om rapporten ikke ser etter kausale sammenhenger, kan tallene indikere at påvirkningen på befolkningen gjennom denne kanalen er til stede, men ikke spesielt stor, i små høyskoleregioner i Norge.

En mulighet er at studiestedets påvirkning på humankapitalen i en region, kan avhenge av det eksisterende arbeidsmarkedet. Gitt at studiene som peker på arbeidsmarkedet som sentralt for nyutdannede stemmer, vil regioner med et sterkt næringsliv tjene mer på å drive med høyere utdanning enn regioner med svakere næringsliv og mindre arbeidsmarked. Regionene med lite arbeidsmarked vil i mindre grad lykkes i å holde på studentene, og vil også dra mindre nytte av et forskningsmiljø. Regionene med større arbeidsmarked vil tiltrekke seg nyutdannede fra andre regioner. I Norge er det for eksempel sannsynlig at en region med mange industribedrifter har mer nytte av å utdanne ingeniører enn en region med lite industri. Varga (2003) finner at lokale teknologioverføringer fra universiteter er størst i regioner med stor befolkning og et avansert næringsliv.

## 2.2 Etterspørselseffekter

Høyskoler og universiteter kan, i tillegg til rollen som utdanningsinstitusjoner, også bli regnet som egne virksomheter. Den økonomiske aktiviteten de genererer gjennom sine utgifter, kan være betydningsfull for regionen (Veugelers og Del Rey, 2014). Institusjonene er ofte store arbeidsgivere og genererer i tillegg mange arbeidsplasser i regionen indirekte. Ansatte, studenter og utdanningsinstitusjonen etterspør varer og tjenester i nærmiljøet og legger grunnlaget for flere arbeidsplasser. Omfanget av etterspørselseffektene kommer an på utdanningsinstitusjonen sin størrelse og hvor stor andel av etterspørselen den genererer som går til regionen den er i.

Det finnes en multiplikatoreffekt som påvirker omfanget av etterspørselseffektene. Den første etterspørselsøkningen skaper grunnlag for flere arbeidsplasser og økonomisk vekst.

---

<sup>3</sup>Stambøl definerer dette som regioner der mer enn fem prosent av de høyt utdannede i regionen arbeider ved utdanningsinstitusjonen.

Denne veksten skaper igjen grunnlag for nye arbeidsplasser og mer vekst. Bjørnar Sæther et al. (2000) bruker en PANDA-modell<sup>4</sup> for å studere blant annet multiplikatoreffekten i sju norske universitets- og høyskoleregioner. De finner at multiplikatoren ligger på mellom 1,73 og ca 2,5, noe som i følge rapporten er høyere enn ved andre typer arbeidsplassetableringer. Siegfried et al. (2007) argumenterer imidlertid for at 'impact-analyser' av denne typen ofte overestimerer virkningene.

Hoveddelen av forskning på etterspørselseffekter er studier som analyserer virkningene av en enkel utdanningsinstitusjon på lokaløkonomien. I følge Goldstein og Drucker (2006) er det vanlig å måle de direkte og indirekte virkningene av institusjonen ved å ta utgangspunkt i institusjonens forbruk, investeringer og sysselsetting, og se virkningene dette har på regionaløkonomien. De diskuterer to hovedproblemer ved denne metoden. For det første er det vanskelig å påvise kausale sammenhenger mellom studiestedets aktiviteter og økonomiske variabler som befolkningsvekst. Utfordringen er ofte å finne en god måte å kontrollere for hvordan utfallet i regionen ville ha blitt uten utdanningsinstitusjonen til stede. Siegfried et al. (2007) bruker som eksempel at det er vanskelig å forestille seg Boston uten Harvard. Det kan også være vanskelig å generalisere resultatene til å gjelde andre regioner og utdanningsinstitusjoner.

Der tilbudseffektene forventes å virke over en lengre tidshorison, er etterspørselseffektene mer umiddelbare. Økningen i humankapital kan legge grunnlag for nye arbeidsplasser på sikt, ved å legge til rette for innovasjon og næringsvirksomhet. Arbeidsplassene som kommer av økt forbruk eller investeringer fra utdanningsinstitusjonene, fungerer mer som eksogene sjokk i etterspørselen i en region. Multiplikatoreffekter, samt mulige agglomerasjonseffekter av den initiale befolkningsøkningen, kan likevel gjøre at etterspørselseffektene også har en virkning på lengre sikt (Florax, 1992).

## 2.3 Agglomerasjonseffekter

Gevinsten ved å samle mennesker og næringsvirksomhet i klynger kalles agglomerasjonseffekter. Alfred Marshall identifiserte effektene i 1890. I større byer får bedriftene god tilgang på arbeidskraft og kunder, og innbyggerne får god tilgang til arbeidsplasser, varer og tjenester. Fordelene som oppnås ved å konsentrere mennesker og virksomheter, kan føre til selvforsterkende vekst når byen får en viss størrelse (Krugman, 1991). Det finnes også kostnader ved å konsentrere befolkningen i små områder, som sørger for at byene ikke vokser uendelig<sup>5</sup>. Statlige arbeidsplassetableringer fra universiteter, høyskoler

---

<sup>4</sup>Plan- og analysemodell for næringsliv, demografi og arbeidsmarked. <http://pandagruppen.no/Om-Panda>

<sup>5</sup>Diseconomies of scale. For eksempel forurensning og trafikkproblemer.

og sykehus kan da stimulere til videre vekst i en region, dersom de positive effektene ved befolkningskonsentrasjonen er større enn de negative.

Venables (2005) oppsummerer viktige agglomerasjonseffekter i 'New Economic Geography'. For det første er interaksjoner over lang avstand kostbare. Han nevner lavere transportkostnader og enklere kommunikasjon som insentiver til å samle mennesker og næring i mindre geografiske områder. For det andre kan større byer gi stordriftsfordeler. Blant annet kan mange fellesgoder produseres mer effektivt til mange mennesker når de er samlet på et mindre område. Et stort og allsidig arbeidsmarked innenfor et konsentrert område er et annet eksempel. Både arbeidsgivere og arbeidstakere øker nytten når de får bedre mulighet til å velge en mer optimal arbeider eller arbeidsplass.

Opprettelser og utvidelser av høyskoler og universiteter bidrar til slik konsentrasjon i regionene det gjelder. I tillegg til de økonomiske faktorene jeg nevnte i forrige avsnitt, kan utdanningsinstitusjonene spille en viktig sosial og kulturell rolle. Med mange studenter samlet skapes det mer aktive og levende lokalmiljøer (NOU 2008:3). Tilbudet av kulturelle arrangementer, fritidsaktiviteter og andre tjenester styrkes ved økt etterspørsel. Dette kan resten av befolkningen også dra nytte av, og det kan bidra til økt nettoinnflytting til regionen.

### **2.3.1 Humankapital og læring**

Læring av andre mennesker er en svært viktig agglomerasjonseffekt. Byer gjør kontakt med andre mennesker enklere og kan på den måten i seg selv øke humankapitalen. Lucas (1988) peker på at i tillegg til innovasjoner, er også mer hverdagslige ferdigheter og kunnskap viktige for å øke humankapitalen og gjøre byene mer produktive. Disse ferdighetene kan læres ved å observere andre. Populære endogene vekstmodeller som Romer (1986, 1990), bygger på at slike læringseffekter eksisterer. En bedrift vil ikke greie å holde ny kunnskap for seg selv, slik at forskning har positive eksternaliteter.

En høyere utdanningsinstitusjon vil være en viktig arena for disse læringseffektene. Duranton og Puga (2003) har en modell som tar utgangspunkt i at nærhet til individer med høyere kunnskapsnivå eller bedre ferdigheter, øker læringsprosessen. Ettersom høyskoler og universiteter tiltrekker seg ansatte og studenter med høyt kunnskapsnivå, vil andre mennesker i regionen kunne tilegne seg mer kunnskap og ferdigheter. Jaffe (1989) sine funn er en indikasjon på dette. Han finner signifikante effekter av geografisk nærhet til et universitet på innovasjonen i en region.

Empiriske studier som forsker på tilbudseffektene av utdanningsinstitusjoner, bruker ofte læringseffekten som et argument for produktivitetsvekst, blant annet Persson og Regner (2004).

## 2.4 Tidligere empirisk forskning

To typer empirisk forskning blir sett på i dette kapitlet. Jeg gjennomgår noen studier som ser på hvordan omfanget av høyere utdanning i regionene påvirker regionaløkonomien. Disse studiene ligner på analysen i denne oppgaven med tanke på forklaringsvariablene, men venstresidevariablene er ulike mine.

Siden denne oppgaven har befolkningsvekst som avhengig variabel, tar jeg også med to studier som ser på hvordan nivået av humankapital påvirker befolkningsveksten i en by. Disse studiene fokuserer ikke direkte på høyere utdanningsinstitusjoner, men er relevante i den grad høyskoler og universiteter øker humankapitalen i en region.

Empiriske utfordringer som studiene står overfor, utdypes ikke i dette kapitlet. Jeg vil hovedsakelig nevne utfordringer som studiene har til felles med masteroppgaven. Disse blir gjennomgått i kapittel 4. Jeg velger å henvise dit i stedet for å forklare metodene og utfordringene her.

### 2.4.1 Høgskolers og universiteters påvirkning på regionaløkonomien

Jeg presenterer først to svenske studier som utnyttet en desentraliseringspolitikk som gav nye høyskoler til enkelte regioner omtrent samtidig. Videre ser jeg på en amerikansk studie som fokuserer på hvorvidt høyere utdanning i en region gir eksternaliteter som påvirker innovasjonen. Til slutt tar jeg med en studie av Raymond Florax (1992). Han bruker datamateriale fra Nederland, og fokuserer på universitetenes påvirkning på regionale investeringer.

#### Svenske studier

Sverige har gjort flere utvidelser i universitets- og høyskolestrukturen sin de siste tiårene. Utvidelsene har ofte vært mer systematiske enn tilsvarende utvidelser i Norge, og det har gitt gode muligheter for å studere den regionale effekten av høyere utdanningsinstitusjoner. I 1987 ble det igangsatt en desentraliseringspolitikk for høyere utdanning. De utvidet flere regionale høyskoler, og både studenttallene og forskningsressursene hos disse skolene økte betraktelig de påfølgende årene.

Andersson et al. (2004) bruker utvidelsene i 1987 i sin studie. De har gjennomsnittsproduktiviteten til arbeidere som avhengig variabel, og ser på denne i forhold til størrelsen på universitetene. For å måle universitetstørrelsen, bruker de både antall studenter og antall

forskere. De benytter et paneldatasett, noe som gjør at de kan bruke faste regionseffekter for å fjerne skjevhet fra permanente forskjeller mellom regionene. Faste tidseffekter blir også tatt i bruk. Den svenske studien benytter noe mer sofistikerte metoder, blant annet for å håndtere eksternaliteter fra universiteter som går over regiongrensene. Sett bort fra disse, er modellene ganske like studenttallanalysene i min masteroppgave. De bruker Sveriges 285 kommuner som regional inndeling, og finner i likhet med denne oppgaven svært lokale effekter.

Studien finner signifikante positive sammenhenger både for antall studenter og forskere. Effekten på produktiviteten er høyere for antall forskere. Nyetablerte institusjoner viser seg å ha større effekt enn eksisterende universiteter.

Persson og Regnér (2004) bruker arbeidsgiver-arbeidstaker-data for alle svenske foretak mellom 1985 og 1995. De utnytter en utbygging av regionale høyskoler på 1970-tallet til å sammenligne regioner med og uten ny høyskole<sup>6</sup>. Avhengige variabler er generell sysselsettingsvekst, sysselsettingsvekst i høyteknologiske industrier, overlevelsesrate for bedrifter og økning i personer med høyere utdanning. Som forklaringsvariabel benyttes en dummyvariabel for om en region har høyskole eller ikke. Dette ligner på analyseformen i denne oppgaven som ser på effekten av utdanningsinstitusjoner uten å ta hensyn til størrelsen.

Studien diskuterer utfordringen med at høyskolene ikke utplasseres tilfeldig i regionene. De gjør et forsøk på å ta hensyn til dette ved å konstruere kontrollgrupper bestående av relativt like regioner uten høyskole. I tillegg legger de på kontrollvariabler. Persson og Regnér finner ingen signifikante effekter av høyskoleetableringene.

### **Anselin, Varga og Acs**

Anselin, Varga og Acs (1997) undersøker i hvilken grad nærhet til universiteter påvirker innovasjonen i en region. I motsetning til denne studien fokuserer de på forskning i stedet for antall studenter. Som venstresidevariabel har de en proxy for innovasjon, blant annet antall patenter, og som interessevariabel bruker de forskning utført på universiteter. Geografisk fokuserer de på fysisk avstand fra universitetene i stedet for å dele inn i regioner. Dette skiller studien fra denne oppgaven og de to svenske studiene.

Studien finner at universitetsforskning gir positive eksternaliteter som øker innovasjonen opp til en radius på 50 miles rundt universitetene. De finner også at universitetene har en positiv regional påvirkning på forskning i privat sektor. Resultatene støtter opp under

---

<sup>6</sup>I 1977 var det kun seks universiteter i Sverige, i tillegg til fem store tekniske høyskoler og 14 regionale høyskoler. Samme år ble det opprettet 11 nye institusjoner, og flere eksisterende høyskoler ble oppgradert.

teoriene om tilbudseffektene av utdanningsinstitusjonene. Regioner med utdanningsinstitusjoner kan forventes å vokse på grunn av et styrket næringsliv.

### **The university: A Regional Booster?**

Raymond Florax (1992) ønsker å fange opp tilbudseffektene av utdanningsinstitusjonene. Han argumenterer for at regionale investeringer er en bedre avhengig variabel for dette formålet enn regional produksjon. Regional produksjon fanger i større grad opp etterspørselseffekter, mens regionale investeringer sier mer om humankapitalnivået.

Han bruker datamateriale fra Nederland mellom 1977 og 1984, og har blant annet geografisk avstand fra universiteter som interessevariabel. Resultatene indikerer at tilbudseffektene generelt er lite signifikante, spesielt for sentrale regioner. For perifere regioner finner han indikasjoner på at avstand fra universiteter kan være relevant med tanke på teknologisk utvikling, fordi avstanden har signifikant effekt på investeringer i produksjonsutstyr. Økt behov for nytt produksjonsutstyr, kan tyde på at regionene hurtigere tar i bruk ny teknologi.

Florax understreker at Nederland er et lite land med små avstander og høy befolkningstetthet. Det er usikkert om betydningen av resultatene kan overføres til et større og mindre folkerikt land som Norge.

## **2.4.2 Humankapital og vekst**

Den empiriske forskningen jeg har presentert til nå, har i hovedsak gått ut på i hvor stor grad høyere utdanningsinstitusjoner påvirker humankapital i regionen. Når en studie finner at et universitet øker produktiviteten eller innovasjonen i en region, tyder det på at humankapitalen øker. Siden denne studien har befolkning som avhengig variabel, er det relevant å se på noen studier som tar for seg hvordan humankapital påvirker befolkningsveksten.

### **Winters**

I følge Winters (2011) er det en populær hypotese at mennesker flytter til byer med mye humankapital fordi disse byene er mer produktive, og dermed har høyere lønninger. Studien ser på migrasjonsdata i USA, og bruker en modell som har innflytting, utflytting og netto innflytting som avhengige variabler. Interessevariabelen er andelen med høyere utdanning. Studien vil finne sammenheng mellom migrasjon og andelen med høyere utdanning i en by. Winters kaller byer med høy andel utdannede for 'smarte byer' ('smart



cities’). Han fokuserer spesielt på å finne ut hvilke typer mennesker som flytter inn og ut av disse byene, ved å kjøre separate estimeringer for personer med og uten en rekke karakteristikk.

Studien finner at en stor del av befolkningsveksten til ’smarte byer’ kommer av studenter som flytter til byen for å studere og blir igjen når de er ferdig utdannet. Denne migrasjonen gjør at byer med høy humankapital vokser raskere enn andre byer. Dersom resultatene til Winters (2011) kan overføres til norske forhold, kan regionene med studiested forventes å vokse på grunn av utdanningsinstitusjonene.

## Glaeser og Saiz

Glaeser og Saiz (2004) begynner med å presentere deskriptiv statistikk som forteller at byområder der under 10 prosent av befolkningen har høyere utdanning, vokste med 13 prosent mellom 1980 og 2000. Byområder med over 25 prosent med høyere utdanning vokste med 45 prosent i samme periode. Med tanke på dette, ser de etter årsaker til hvorfor ’smarte byer’ vokser raskere. De ser både på hele byregioner og områder innen byregionene for å teste svært lokale effekter.

Vekstraten til befolkningen er avhengig variabel, og andelen av befolkningen med høyere utdanning er interessevariabel. Glaeser og Saiz benytter både kontrollvariabler, faste byeffekter og faste regioneffekter for å øke påliteligheten av estimatene. De bruker også omfanget av høyere utdanning i regionen i 1940 som interessevariabel, fordi denne variabelen trolig er en mindre kilde til omvendt kausalitet og utelatt variabelskjevhet. Grunnen til at de ikke bruker den som instrumentvariabel, er at forutsetningene for et instrument trolig ikke er oppfylt<sup>7</sup>. Omvendt kausalitet og utelatt variabel skjevhet er sentrale utfordringer i denne masteroppgaven, og jeg tar i bruk flere av metodene som denne studien benytter seg av.

Studien indikerer at humankapitalen i regionen har en signifikant positiv effekt på befolkningsveksten. Resultatene tyder på at dersom utdanningsinstitusjonene klarer å påvirke humankapitalen i regionene, vil de ha en positiv effekt på befolkningsveksten.

## 2.5 Norske undersøkelser

Jeg velger også å presentere en norsk undersøkelse om emnet. Metodene brukt i undersøkelsen ligner ikke på metodene i denne oppgaven, men den kan likevel gi verdifull bakgrunnsinformasjon om forholdene i Norge.

---

<sup>7</sup>Forutsetningene diskuteres i kapittel 4.3.1

Bjørnar Sæther et al. (2000) ser på tilbudseffekter, etterspørselseffekter, og påvirkning på flytting og bosettingsmønster. I undersøkelsene av tilbudseffekter, har de brukt kvalitativ analyse for Tromsø-regionen og Østfold, der de har funnet tegn på kunnskapsoverføringer fra høgskole og universitet til næringslivet. For Halden bruker de framveksten av et IT-miljø i byen som eksempel. Der har etableringen av studieprogram i IT på Høgskolen i Halden gitt tilgang til kvalifisert arbeidskraft, og vært en viktig lokaliseringsfaktor for IT-bedrifter.

Etterspørselseffektene analyseres her ved å bruke regnskapstall. De bruker den regionaløkonomiske modellen Panda<sup>8</sup> for å beregne etterspørselsimpulsene som genereres via høgskolenes kjøp av varer og tjenester og de ansattes konsum. De ønsker å tallfeste hvor stor betydning høgskolene har for hver enkelt regions økonomi, og hvor stor multiplikatoreffekten er. Tallene indikerer at høgskoler i mindre regioner spiller en større rolle for regionaløkonomien enn høgskoler i større regioner. Dette sammenfaller til en viss grad med resultatene i tabell 5.9, som viser at høgskoler har størst effekt på befolkningsvekstraten i små regioner.

Slik som Winters (2011), har Sæther et al. (2000) også sett på migrasjon. Kommunene deles opp etter sentralitetsnivå og hvorvidt de har høgskole, universitet eller ingen av delene. Resultatene er basert på årskullene født i 1960-1962, som er fulgt fra de var 15 til 35 år. Tallene viser at universitets- og høgskoleregioner jevnt over beholder flere av de unge som studenter i egen region, tiltrekker seg flere studenter fra andre regioner, og beholder flere tilflyttede studenter. Samtidig kan det diskuteres hvor mye som kan trekkes ut av disse tallene, siden universitets- og høgskoleregionene gjerne deler andre karakteristika som påvirker migrasjon.

Med tanke på mulige kausale sammenhenger, vil de mest interessante tallene hentes fra sammenligninger av regioner med og uten høgskole, på samme sentralitetsnivå. Her er tallene mindre klare, men det er en tendens til at flere unge fra lite sentrale områder med høgskole tar høyere utdanning, enn unge fra tilsvarende områder uten høgskole. Når det kommer til totaleffekten - antall personer med høyere utdanning som bor i regionen som 35-åringer i forhold til antallet som vokste opp i regionen i samme årskull - spiller sentralitetsnivået den viktigste rollen. Det er likevel en forskjell på regioner med og uten høgskole på samme sentralitetsnivå, der regionene med høgskoler har høyere nettoinnflytting. Kort oppsummert kan tallene i studien gi noen indikasjoner på at høgskoler har hatt en positiv effekt på befolkningsveksten i sine regioner.

---

<sup>8</sup><http://pandagruppen.no/Om-Panda>

# Kapittel 3

## Datamateriale

Datamaterialet i studien er paneldata, med befolkning i regionene som avhengig variabel og studenttall i regionene som utgangspunkt for interessevariablene. Den ene analysetypen jeg bruker, benytter dummyvariabler som interessevariabler. Disse har verdien 1 i observasjoner der studenttallet indikerer at regionen har høgskole, og 0 ellers. Den andre typen bruker studenttallet i regionen som interessevariabel. Denne variabelen skal fungere som en proxy for størrelsen på utdanningsinstitusjonen. Ved å benytte studenttallet, kan variasjonen over tid innenfor regionene utnyttes bedre. Datamaterialet er hovedsakelig hentet fra NSDs Kommunedatabase <sup>1</sup>. Tallene fra Kommunedatabasen er i utgangspunktet på årsbasis, men jeg velger å slå dem sammen, slik at tidsenhetene i analysene er femårsperioder.

Norge deles inn i økonomiske regioner etter SSBs definisjon<sup>2</sup>. Data fra Kommunedatabasen er i utgangspunktet på kommunenivå, men jeg aggregerer tallene til regionnivå før estimeringene gjøres. Jeg ser dette som en mer hensiktsmessig geografisk inndeling. Utdanningsinstitusjonene har trolig en økonomisk effekt langt utover kommunegrensene, ettersom arbeidsmarkeder blant annet, ofte deles av mange kommuner. SSB deler Norge inn i 89 økonomiske regioner.

Data for antall studenter og befolkning er for årene 1976 til 2000. Det vil si at det i utgangspunktet er 25 tidsperioder i datasettet. Etter sammenslåingen til femårsperioder, har jeg bare fem tidsperioder. Jeg begrunner sammenslåingen med at endringer i bosetningsmønster kan ta lang tid, slik at data på årnivå trolig ikke vil være optimalt til å fange opp virkninger. Interessevariabelen i mange modeller er lagget studenttall. Det vil være mer interessant å finne ut hvordan gjennomsnittlig studenttall i regionen over en femårsperiode påvirker befolkningen i neste femårsperiode, enn hvordan studenttallet i ett år påvirker befolkningen neste år. For å teste effekter like langt bakover med ettårspe-

---

<sup>1</sup>[https://trygg.nsd.uib.no/kdbbin/kdb\\_start.exe](https://trygg.nsd.uib.no/kdbbin/kdb_start.exe)

<sup>2</sup><http://www4.ssb.no/ClassificationFrames.asp?ID=676393&Language=nb>

rioder, må det legges på svært mange lagg, og det blir en lite oversiktlig og dårlig løsning. Ulempen med femårsperioder er at det blir færre observasjoner i estimeringene. Dette er en liten ulempe i forhold til gevinsten. Etter at datamaterialet er slått sammen til fem tidsperioder og 89 regioner, er det 445 observasjoner totalt.

Jeg starter med å forklare inndelingen i økonomiske regioner i delkapittel 3.1, før jeg går over til å presentere variablene i studien. I delkapittel 3.2 tar jeg for meg interessevariablene. Her går jeg nærmere inn på de to typene interessevariabler som benyttes, og ser på den geografiske strukturen av høyere utdanning i Norge. I delkapittel 3.3 gir jeg et overblikk over befolkningsutviklingen, og presenterer blant annet statistikk over befolkningsveksten i regioner med og uten høyskole. Til slutt presenterer jeg to kontrollvariabler jeg benytter meg av i delkapittel 3.4.

## 3.1 Økonomiske regioner

Statistisk sentralbyrå deler Norge inn i 89 økonomiske regioner. Det er en regional inndeling på nivå mellom kommune og fylkeskommune, og bygger på kriterier som handel og arbeidsmarked. Jeg har tatt datamaterialet jeg hentet fra NSDs Kommunedatabase og slått det sammen til regionsnivå.

Siden det er vanlig at flere kommuner kan ha felles arbeidsmarked og økonomisk samhandling, er det mer hensiktsmessig å se på økonomiske regioner enn på kommuner i denne studien. Nye arbeidsplasser og økt aktivitet i en kommune har ofte positive ringvirkninger for nabokommuner. Utdanningsinstitusjonene har nok i noen tilfeller også eksternaliteter som går utover de økonomiske regionene, men i mindre grad enn over kommunegrensene. En større inndeling ville heller ikke vært hensiktsmessig, ettersom det er relativt lokale virkninger som er av interesse.

En utfordring med omgjøringen til økonomiske regioner, er målefeilproblemene som oppstår når datamaterialet er ufullstendig. Mange variabler i NSDs Kommunedatabase har manglende data. Dette er ikke nødvendigvis et stort problem når regresjonene er på kommunenivå, så lenge observasjonene med manglende data er noenlunde tilfeldig spredt og omfanget ikke er for stort. Observasjoner med manglende data blir sett bort fra i regresjonene. Problemet oppstår når tallene aggregeres fra kommunenivå til regionsnivå, og minst én, men ikke alle kommunene i en region har manglende data. Da vil en kommune med manglende data og en kommune der variabelen har tallet null bli regnet som identiske.

En interessant variabel med mye manglende data er bruttoinvesteringer for bedrifter med mer enn fem ansatte. Ifølge Raymond Florax (1992) er dette en egnet variabel til å måle

tilbudseffektene av utdanning. Denne variabelen har data som overlapper med antall studenter-variabelen i 20 år, men har samtidig en del manglende data. For eksempel mangler Bodø kommune data tidlig i perioden, fra 1976 og noen år utover, mens andre kommuner i Bodø-regionen har data for samme periode. Bodø-regionen vil da få et kunstig lavt tall på bruttoinvesteringer i de to første 5-årsperiodene. Dersom modellen i stedet var på kommunenivå, ville Bodø kommune ikke blitt medregnet. Selv om dette kunne vært en interessant variabel, ser jeg på problemene med målefeil som for omfattende.

Sammenslåingen til økonomiske regioner er en av hovedgrunnene til at bare befolkning, med svært solid datamateriale, ble valgt som resultatvariabel. Antall studenter i kommunene er også en variabel med svært få manglende observasjoner.

## 3.2 Høyere utdanningsinstitusjoner og studenttall

Interessevariablene i studien baserer seg på studenttallet i regionene, hentet fra NSDs Kommunedatabase. Antall studenter brukes enten som utgangspunkt for å lage dummy-variabler, eller som proxy for størrelsen på utdanningsinstitusjonen.

Fordelen med å bruke studenttallet som interessevariabel, er at det gir mer variasjon over tid innenfor regionene. Studiestedene varierer i størrelse over tid, og det kan være interessant å se hvordan befolkningen i regionen henger sammen med størrelsen på studiestedet. Ulempen er at det ikke alltid er en god proxy for omfanget av utdanningsinstitusjonen. Det kan også i noen tilfeller være mer interessant å se hvordan høyskoler påvirker befolkningsveksten, uten å ta størrelsen på skolen i betraktning. Av de to svenske studiene jeg presenterte i kapittel 2.4.1, bruker Persson og Regnér (2004) dummyvariabler, mens Andersson et al. (2004) bruker studenttall som interessevariabel.

Før jeg går inn på de to variabeltypene, diskuterer jeg kort strukturen for høyere utdanningsinstitusjoner i Norge. Studiestedene i Norge ligger hovedsakelig i byregioner, og de store byene har større universiteter og høyskoler enn de mindre byene. Det finnes ingen større byregioner uten høyere utdanningsinstitusjoner.

Mange distriktshøyskoler i Norge ble opprettet mellom 1969 og 1981. Enkelte mindre regioner, som for eksempel Volda, Sogndal, Bø i Telemark, Halden og Levanger, har relativt store studiesteder. Disse regionene er på mange områder mer like andre ikke-urbane regioner uten studiesteder, enn folkerike regioner som har studiesteder av lignende omfang. For sammenligninger mellom regioner i Norge, kan det av denne grunn være hensiktsmessig å ta mindre regioner med studiesteder og se disse i forhold til mindre regioner uten høyskole eller universitet. Store regioner har ulik forventet befolkningsvekst i forhold til mindre regioner av andre årsaker enn bare utdanningsinstitusjonen. Det er

en svakhet ved datamaterialet at data for studenttall ikke går lengre bak enn 1976. Dette kunne ha åpnet for bedre analyser av befolkningsutviklingen i disse små regionene.

I perioden 1976-2000, skjer det en stor økning i antall studenter for landet sett under ett. I 1976 var høyere utdanning mye mer sjeldent enn i 2000. Tabell 3.1 viser folkemengde og studenttall totalt for landet i hver tidsperiode. Tabellen viser at det skjer en stor økning i antall studenter, særlig fram til midten av 90-tallet. I de to siste periodene er andelen studenter i befolkningen over dobbelt så høy som i den første perioden. I 1976-80 er det 58 personer for hver student i befolkningen, mens tallet har gått ned til 23,9 personer for hver student i 1996-2000.

Tabell 3.1: Utvikling i hovedvariablene for alle regioner over hele studieperioden

|                  | Befolkningsmengde | Antall studenter | $\frac{\text{Befolkningsmengde}}{\text{Antallstudenter}}$ |
|------------------|-------------------|------------------|---|
| Gjnsnitt 1976-80 | 4 048 737         | 69766            | 58,0  |
| Gjnsnitt 1981-85 | 4 119 447         | 88190            | 46,7  |
| Gjnsnitt 1986-90 | 4 196 079         | 116148           | 36,1  |
| Gjnsnitt 1991-95 | 4 298 240         | 166455           | 25,8  |
| Gjnsnitt 1996-00 | 4 419 882         | 185091           | 23,9  |

Veksten i studenttall er stor i både små, mellomstore og store regioner. Tabell 3.2 viser at studenttallene prosentvis vokser mest i små og mellomstore regioner, mens den absolutte økningen er størst i regioner med over 100 000 innbyggere. Noen få regioner bytter kategori ettersom folketallet endrer seg mellom periodene. For eksempel går Tønsberg-regionen fra rett i underkant av 100 000 innbyggere i perioden 1991-1995, til rett i overkant av 100 000 i 1996-2000. Det betyr at studenttallet for mellomstore regioner ville ha vært ca 45 000 i den siste perioden om Tønsberg-regionen var medregnet. Tallene i tabellen bør derfor bare brukes til å vurdere den generelle trenden, som er at den prosentvise økningen i studenttall er svært stor i små og mellomstore regioner.

### 3.2.1 Dummyvariabler for studiested

I analysene der størrelsen på studiestedet ikke er av interesse, benytter jeg dummyvariabler for om en region har høyskole eller ikke. Den viktigste analysen med dummyvariabler, utforsker virkninger av nyetablerte og utvidede høyskoler. Her definerer jeg dummyvariablene etter en regel, som benytter studenttallet i regionene til å fange opp regioner som får nye eller sterkt utvidede studiesteder.

Dummyvariabelen for ny høyskole får verdi 1 dersom endringen i den naturlige logaritmen

Tabell 3.2: Studenttall etter regionstørrelse.

|                  | Over 100 000 | Mellom 30 000 og 100 000 | Under 30 000 |
|------------------|--------------|--------------------------|--------------|
| Gjnsnitt 1976-80 | 54 392       | 11 538                   | 3 835        |
| Gjnsnitt 1981-85 | 66 508       | 16 679                   | 5 004        |
| Gjnsnitt 1986-90 | 85 022       | 24 765                   | 6 360        |
| Gjnsnitt 1991-95 | 115 904      | 38 312                   | 12 240       |
| Gjnsnitt 1996-00 | 128 349      | 41 349                   | 15 393       |

til studenttallet øker med to eller mer. Dette vil være en prosentvis økning på i overkant av 600 prosent. For å unngå at svært små studiesteder medregnes, må studenttallet også øke med 100 eller mer mellom periodene.

$$DNyskole = \begin{cases} 1, & \text{hvis } \Delta \text{LogStud} \geq 2 \ \& \ \Delta \text{Studenter} \geq 100 \\ 0, & \text{ellers} \end{cases} \quad (3.1)$$

Det kan også være interessant å inkludere eksisterende studiesteder som har fått store utvidelser. For å få med disse, endres regelen slik at dummyvariabelen får verdi 1 dersom endringen i den naturlige logaritmen til studenttallet øker med én eller mer.

$$DUtvidet = \begin{cases} 1, & \text{hvis } \Delta \text{LogStud} \geq 1 \ \& \ \Delta \text{Studenter} \geq 100 \\ 0, & \text{ellers} \end{cases} \quad (3.2)$$

Variablene generert av ligning (3.1) og (3.2) får verdien 1 akkurat i den perioden åpningen eller utvidelsen skjer. Det vil si at de kun måler den umiddelbare effekten.

For å teste om disse utbyggingene gir langsiktige endringer i befolkningsveksten, lages to dummyvariabler som får verdien 1 også i periodene etter utbyggingen. Tanken bak dette er at dersom høgskoleutbyggingen også påvirker veksten i periodene etter utbyggingen, kan dette i teorien fanges opp av dummyvariabelene *DEtterskole* eller *DEtterutvidet*.

Det er sju regioner som etter reglene får ny høgskole, mens det skjer 18 signifikante utvidelser<sup>3</sup>. Slik variasjon over tid innenfor regionene åpner for å sammenligne befolkningsveksten i regionene før og etter utbygging, med befolkningsvekst i andre regioner uten utbygging.

<sup>3</sup>Haugesund, Drammen og Lillestrøm telles to ganger i *DUtvidet*. Haugesund får i følge reglene ny skole i perioden 1981-1985 og en stor utvidelse i 1986-1990. Drammen og Lillestrøm får store utvidelser i periodene 1981-1985 og 1986-1990. Det vil si at 15 ulike regioner kommer med i *DUtvidet*.

Regionene som i følge regel (3.1) har fått nytt studiested i løpet av perioden er Moss, Hadeland, Haugesund, Førde, Kristiansund, Steinkjær og Namsos<sup>4</sup>. Tabell 3.3 viser gjennomsnittlig befolkningsvekst i prosent i de sju regionene, sammenlignet med gjennomsnittsvæksten i alle andre regioner. Tabellen viser at den gjennomsnittlige befolkningsveksten i disse sju regionene var høyere enn gjennomsnittet for samtlige regioner. Dette gjelder alle periodene. Samtidig er forskjellen størst i de to periodene på 1980-tallet, før alle sju regionene har fått høyskole. I de to siste periodene, når langtidseffektene av høyskolene i teorien burde merkes i flere av regionene, er forskjellene mindre. Det må også nevnes at Moss-regionen går tilbake til null studenter i perioden 1996-2000, selv om den fremdeles regnes med som en 'ny skole-region' etter regelen som benyttes. Moss-regionen har sterk vekst i de siste periodene på tross av dette.

Tabell 3.3: Gjennomsnittlig befolkningsvekst i regioner med ny høyskole sett opp mot alle andre regioner

|                  | Ny høyskole | Andre regioner |
|------------------|-------------|----------------|
| Gjnsnitt 1981-85 | 3,11        | 1,63           |
| Gjnsnitt 1986-90 | 1,06        | 0,26           |
| Gjnsnitt 1991-95 | 1,19        | 0,76           |
| Gjnsnitt 1996-00 | 1,08        | 0,66           |

I den andre analysen der jeg bruker dummyvariabler, sammenligner jeg 12 små regioner som har relativt store høyskoler med en kontrollgruppe bestående av andre små regioner. Jeg har funnet 12 regioner med mindre enn 40 000 innbyggere, der studenttallene indikerer at de har store høyskoler i forhold til befolkningen. De økonomiske regionene er Halden, Elverum, Lillehammer, Notodden/Bø, Sogndal, Ørsta/Volda, Steinkjer, Levanger/Verdal, Narvik, Mo i Rana, Harstad og Alta. Som kontrollgruppe benyttes alle andre regioner under 40 000 innbyggere. Tabell 3.4 viser at den gjennomsnittlige befolkningsveksten i de små regionene med høyskole har vært større i tre av fire perioder. Selv om forskjellene ikke er spesielt store, er det et utgangspunkt for å gjøre en økonometrisk analyse.

### 3.2.2 Studenttall

Antall studenter skal fungere som et mål på omfanget av høyere utdanning i regionen. I forhold til dummyvariablene, vil denne variabelen gi mer variasjon over tid innenfor

<sup>4</sup>Merk at studiested ikke nødvendigvis betyr at regionen har egen høyskole. For eksempel er studiestedet i Kristiansund en del av Høyskolen i Molde, mens studiestedet i Førde er en avdeling av Høyskolen i Sogn og Fjordane. En region som Harstad, som fikk distriktshøyskole i 1986, kommer ikke med i denne regelen. Der var det et visst omfang av høyere utdanning i periodene før.



Tabell 3.4: Gjennomsnittlig befolkningsvekst i 12 små høyskoleregioner sammenlignet med andre små regioner

|                  | Små høyskoleregioner | Andre regioner under 40 000 |
|------------------|----------------------|-----------------------------|
| Gjnsnitt 1981-85 | 0,92                 | 1,42                        |
| Gjnsnitt 1986-90 | -0,58                | -0,78                       |
| Gjnsnitt 1991-95 | 0,90                 | -0,29                       |
| Gjnsnitt 1996-00 | 0,24                 | -0,57                       |

regionene. Det er samtidig litt usikkert hvor godt den fungerer som proxy for størrelsen på studiestedet. En faktor som taler for at den er godt egnet, er at studenter ofte ikke melder flytting til kommunen de studerer i. De kommer da i liten grad med i befolkningsstatistikken for regionen. En viktig del av effekten studenttallet har på befolkningen i regionen, er trolig da tilbuds- og etterspørselseffekter knyttet til utdanningsinstitusjonene. En økning i studenttallet innebærer som oftest at utdanningsinstitusjonen får flere ansatte, økt forbruk og utdanner flere studenter som kan bosette seg i området i når de er ferdig utdannet.

Denne variabelen benyttes hovedsakelig lagget, slik at det er effekten av studenttallet i forrige tidsperiode på befolkningsveksten i denne tidsperioden som måles. Dette vil trolig gjøre variabelen mer egnet som proxy. Ettersom tidsperiodene er delt inn i femårsperioder, vil en stor andel av studentene ha endt studiene sine til den neste perioden starter. Det betyr at effekten som fanges opp, i større grad er andelen av studenter som blir igjen etter endt studie, heller en andelen av nåværende studenter som melder flytting til regionen. Etterspørselseffektene fra studiestedet kan også forventes å bruke noen år på å slå ut for fullt. Økningen i arbeidsplasser på høyskolen eller universitetet på grunn av økt studenttall skjer trolig relativt raskt, men de mer indirekte virkningene kan ta lengre tid.

### 3.3 Befolkning

For folkemengden i kommunene finnes det svært gode data. Det er ingen manglende observasjoner, og trolig svært pålitelige målinger. Befolkningsutviklingen bestemmes i stor grad av summen av mange andre økonomiske variabler. Den påvirkes av mange andre variabler som kunne ha vært interessante å teste opp mot utdanningsinstitusjoner. Eksempler er sysselsetting, gjennomsnittsinntekt og nyskapning i næringslivet. Lite målefeil og høy forklaringsverdi med tanke på regionale økonomiske forhold gjør befolkning til en

egnet resultatvariabel.

Tabell 3.5: Befolkningsvekst i prosent i regioner med og uten høyere utdanningsinstitusjoner

|                   | Med høyere utdanning | Uten høyere utdanning |
|-------------------|----------------------|-----------------------|
| Vekstrate 1981-85 | 2,19                 | 1,36                  |
| Vekstrate 1986-90 | 1,58                 | -0,96                 |
| Vekstrate 1991-95 | 1,93                 | -0,43                 |
| Vekstrate 1996-00 | 2,04                 | -0,62                 |

Tabell 3.5 viser den prosentvise endringen i folketallet i en region. Tabellen viser at vekst-raten har vært mye større i regioner som har en høyere utdanningsinstitusjon. Regioner uten høgskole eller universitet har i gjennomsnitt hatt negativ befolkningsvekst siden midten av 80-tallet.

Ifølge tabell 3.6 har befolkningsveksten vært størst i regioner med over 100 000 innbyggere, og minst i regioner med under 30 000 innbyggere. Alle observasjoner av regioner med mer enn 100 000 innbyggere i datasettet, har et studenttall høyere enn null. For regionene med under 30 000 innbyggere, har over halvparten av observasjonene null studenter. Tallene i tabell 3.5 og 3.6 reflekterer at sentrale regioner har vokst, delvis på bekostning av ikke-sentrale regioner og delvis på grunn av generell befolkningsvekst. Ettersom høgskolene og universitetene hovedsakelig ligger i sentrale regioner, viser disse tallene at det kan være utfordrende å teste for kausale effekter av utdanningsinstitusjoner på befolkningsutviklingen.

Tabell 3.6: Befolkningsvekst i prosent i små, mellomstore og store regioner

|                   | Over 100 | Mellom 30 og 100 | Under 30 |
|-------------------|----------|------------------|----------|
| Vekstrate 1981-85 | 1,96     | 2,52             | 1,30     |
| Vekstrate 1986-90 | 3,67     | 1,40             | -0,84    |
| Vekstrate 1991-95 | 3,61     | 1,74             | -0,21    |
| Vekstrate 1996-00 | 4,44     | 1,90             | -0,61    |

### 3.4 Statistiske kontrollvariabler

Antall sysselsatte i industrien og antall personer med høyere utdanning i en region i 1970 inkluderes som kontrollvariabler i noen av modellene. Disse variablene kommer fra

folketellingen i 1970, og har ingen manglende data.

Sysselsatte i industrien i 1970 skal kontrollere for eksisterende næringsliv i regionen, mens antall med høyere utdanning skal kontrollere for det eksisterende nivået av humankapital. I teorien vil da endringer i befolkning som kommer av utdanningsinstitusjonenes effekt på humankapitalnivået fanges opp. Befolkningsendringer på grunn av det eksisterende nivået av humankapital og næringsliv, fanges opp av kontrollvariablene.

Det ville vært ideelt med kontrollvariabler som varierte over tid. Men statistikk fra folketellingen kommer kun hvert tiende år, og for disse to variablene var det vanskelig å finne gode tall for hvert år i perioden 1976 til 2000. På grunn av målefeilproblemene som oppstår når kommunene slås sammen til regioner, kombinert med dårlig datakvalitet på en del aktuelle kontrollvariabler, ble valget å kun benytte statiske kontrollvariabler med svært god datakvalitet.

# Kapittel 4

## Empirisk tilnærming

Målet for den empiriske tilnærmingen er å finne ut om datamaterialet indikerer kausale sammenhenger mellom høyere utdanningsinstitusjoner og befolkningsendringer i regionene. Jeg tester derfor flere forskjellige modellspesifikasjoner, og bruker ulike metoder for å håndtere endogenitet i materialet.

Oppgaven baserer seg på to hovedmodeller for befolkningsutvikling. Den ene typen har studenttallet i regionen som interessevariabel, mens den andre bruker dummyvariabler. Jeg spesifiserer disse modellene i delkapittel 4.1.

Modellene estimeres ved Minste Kvadraters Metode (MKM). I delkapittel 4.2 går jeg gjennom estimeringsmetodene, og diskuterer forutsetningene som må være på plass for at modellene skal produsere forventningsrette og effisiente estimater.

Deretter går jeg inn på de økonometriske utfordringene i delkapittel 4.3. Her diskuterer jeg også i hvilken grad metodene jeg benytter, klarer å ta hensyn til disse utfordringene og gi forventningsrette estimater.

### 4.1 Spesifisering av paneldatamodeller

De to modelltypene tar utgangspunkt i to enkle modeller, der befolkningen i en region kun avhenger av folkemengden i regionen i forrige periode og høyere utdanningsinstitusjoner i regionen.

$$Befolkning_{i,t} = \beta_0 + \beta_1 * DSkole_{it} + \beta_2 * Befolkning_{it-1} + \epsilon_{it} \quad (4.1)$$

$$Befolkning_{i,t} = \beta_0 + \beta_1 * Studenter_{it-1} + \beta_2 * Befolkning_{it-1} + \epsilon_{it} \quad (4.2)$$

I ligning (4.1) er interessevariabelen en dummyvariabel som er lik 1 hvis regionen har høgskole eller universitet i tidsperioden. Koeffisienten  $\beta_1$  måler effekten av utdanningsinstitusjonen på befolkning i regionen<sup>1</sup>. Ligning (4.2) har lagget studenttall i regionen som interessevariabel.  $\beta_1$  fanger opp effekten av studenttallet i regionen i forrige periode på befolkningen.

$$\text{LogBef}_{i,t} = \beta_0 + \beta_1 * DSkole_{it} + \beta_2 * \text{LogBef}_{it-1} + \epsilon_{it} \quad (4.3)$$

$$\text{LogBef}_{i,t} = \beta_0 + \beta_1 * \text{LogStud}_{it-1} + \beta_2 * \text{LogBef}_{it-1} + \epsilon_{it} \quad (4.4)$$

Ligning (4.3) og (4.4) har befolkning og studenttall på logaritmeform, slik at det er prosentvise endringer som måles. For å få befolkningsvekst i prosent som venstresidevariabel, kan  $\text{LogBef}_{it-1}$  flyttes over på venstresiden dersom  $\beta_2$  settes lik 1.

$$\Delta \text{LogBef}_{i,t} = \beta_0 + \beta_1 * DSkole_{it} + \epsilon_{it} \quad (4.5)$$

$$\Delta \text{LogBef}_{i,t} = \beta_0 + \beta_1 * \text{LogStud}_{it-1} + \epsilon_{it} \quad (4.6)$$

I (4.5) måler  $\beta_1$  hvordan en høyere utdanningsinstitusjon påvirker prosentvis befolkningsendring i regionen. I (4.6) måler  $\beta_1$  hvordan en prosentvis endring i studenttallet i regionen påvirker prosentvis befolkningsendring.

$$\Delta \text{LogBef}_{i,t} = \beta_0 + \beta_1 * DSkole_{it} + \gamma * DT_t + \epsilon_{it} \quad (4.7)$$

$$\Delta \text{LogBef}_{i,t} = \beta_0 + \beta_1 * \text{LogStud}_{it-1} + \gamma * DT_t + \epsilon_{it} \quad (4.8)$$

Alle modellene jeg tester, bruker faste tidseffekter. I ligning (4.7) og (4.8) er  $DT_t$  en dummyvariabel som er lik 1 for observasjonene i periode  $t$  og null ellers. Det betyr at  $\gamma$  gir hver tidsperiode en egen konstant. På denne måten blir variasjon over tid som er felles for alle regionene transformert bort.

### 4.1.1 Dummyvariabler for utdanningsinstitusjon

Målet med analysene som bruker dummyvariabler, er å finne ut om regioner som har utdanningsinstitusjoner vokser raskere enn regioner uten. Jeg utfører to analyser av denne typen. Den ene ser på effekten av nye eller utvidede høgskoler. Den andre sammenligner

---

<sup>1</sup>Koeffisientene kan testes for signifikans ved hjelp av en t-test.

$$t_{\beta_1} = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)}$$

Koeffisienten er signifikant forskjellig fra null når t-verdien er større enn kritisk verdi for valgt signifikansnivå (Woolridge, 2013).

små regioner som har store høyskoler med andre små regioner.

Den mest sentrale analysen ser på effekten av nye og utvidede høyskoler. Jeg bruker endringer i studenttall til å lage to dummyvariabler, ved hjelp av reglene vist i kapittel 3.2.1. Variabelen  $DNyskole$  får tallet 1 dersom endringer i studenttallene indikerer at regionen har fått ny høyskole.  $DEtterNS$  får verdien 1 i periodene etter denne utbyggingen, og 0 ellers.  $DUtvidet$  er 1 dersom studenttallene indikerer en stor utvidelse i regionen, og  $DEtterUS$  er 1 i periodene etter utvidelsen.

$$\Delta LogBef_{it} = \beta_0 + \beta_1 * DNyskole + \beta_2 * DEtterNS + \epsilon_{it} \quad (4.9)$$

I modell (4.9) vil  $\beta_1$  fange opp effektene en ny høyskole har på befolkningsveksten i den perioden utbyggingen skjer, mens  $\beta_2$  fanger opp effekter på befolkningsveksten i årene etter.

$$\Delta LogBef_{it} = \beta_0 + \beta_1 * DUtvidet + \beta_2 * DEtterUS + \epsilon_{it} \quad (4.10)$$

I modell (4.10) fanger  $\beta_1$  opp de umiddelbare effektene av både nye høyskoler og høyskoler med store utvidelser, mens  $\beta_2$  fanger opp effektene av disse på lengre sikt. Modellene inkluderer også faste tidseffekter, og bygges ut med flere kontrollvariabler i testene.

For at (4.9) og (4.10) skal gi korrekte estimater, forutsettes det at regionene har lik forventet befolkningsvekst sett bort ifra høyskoleutbyggingen. Koeffisientene vil da fange opp effektene av utbyggingene. I praksis er dette ikke tilfelle. Kontrollvariabler vil hjelpe, men det forventes ikke at disse fullt ut klarer å ta høyde for alle faktorer som gir ulik forventet befolkningsvekst i regionene. Dersom modellene ikke fanger opp effekter, kan det ikke utelukkes at effektene likevel eksisterer.

### 4.1.2 Faste regionseffekter

For å fjerne alle permanente forskjeller mellom regioner, kan modellen utvides til å inkludere faste regionseffekter.

$$Bef_{it} = \beta_0 + \beta_1 * Stud_{i,t-1} + \eta_i * DR_i + \gamma_t * DT_t + \epsilon_{it} \quad (4.11)$$

De uobserverte faste regionseffektene er notert som  $\eta_i$ .  $DR_i$  er en dummyvariabel som er lik 1 for region  $i$  og 0 ellers. Da gir  $\eta_i$  hver region en egen konstant, slik at variasjon på tvers av regioner transformeres bort. Modell (4.11) måler bare variasjon over tid innenfor hver region. Denne metoden er derfor kun interessant i analysene som bruker studenttall som interessevariabel. Analysene som benytter dummyvariabler som interessevariabler, er avhengige av variasjon på tvers av regionene.

## 4.2 Estimeringsmetoder

For at MKM-estimeringene av modellene spesifisert i forrige delkapittel skal være forventningsrette og effisiente, må visse forutsetninger være på plass. Woolridge (2013) forklarer Gauss-Markov-teoremet, som innebærer følgende forutsetninger for forventningsrett og effisient estimering av en modell:

- (i) Forventningsverdien til restleddet må være lik null, gitt forklaringsvariablene. Deresom restleddet er korrelert med forklaringsvariabler eller ulik null, vil modellen produsere forventningsskjeve estimater.
- (ii) Ingen perfekt multikolaritet. Det vil si at ingen uavhengige variabler er konstante eller en perfekt lineær kombinasjon av andre variabler. Ved perfekt korrelasjon mellom uavhengige variabler, vil det være umulig å separere effektene.
- (iii) Ingen seriekorrelasjon. Restleddet i en periode kan ikke være korrelert med restleddet i en annen periode. Brudd på denne vil i utgangspunktet ikke gi forventningsskjeve estimater, men standardavvikene blir feilberegnet, noe som påvirker testing for signifikans.
- (iv) Homoskedastisitet. Variansen til restleddet må være konstant for alle regioner og tidsperioder, gitt forklaringsvariablene. Brudd på denne innebærer heteroskedastisitet, og vil også føre til feilberegnet standardavvik.
- (v) Restleddet er uavhengig og normalfordelt, gitt forklaringsvariablene. Dette er nødvendig for å foreta t-testene som denne oppgaven benytter seg av ved hypotesetesting, fordi disse testene bygger på normalfordeling.

For denne oppgaven er den første forutsetningen, om forventningsverdien til restleddet, den mest aktuelle og problematiske. Dette blir hovedfokuset i kapittel 4.3 om de økonomiske utfordringene. Heteroskedastisitet og seriekorrelasjon blir også diskutert grundigere der.

De to estimeringsmetodene som benyttes, er vanlig sammenkoblet MKM og faste effekter. Metodene har fordeler og ulemper som må veies opp mot hverandre. Målet med modellen og hvilke type variasjon som er tilgjengelig i datamaterialet avgjør hvilken metode som er mest passende. Generelt er modeller med faste regioneffekter mer robuste, og holder i større grad forutsetningene for forventningsrett og effisient estimering. Samtidig utnytter de ikke en like stor del av variasjonen i datasettet, da tverrsnittsvariasjon transformeres bort.

### 4.2.1 Sammenkoblet MKM

Ved vanlig sammenkoblet MKM (Pooled cross section time series), blir alle observasjonene i hver periode og region sammenlignet med hverandre. I tillegg til noen studenttallmodeller, blir alle modellene med dummys som interessevariabler estimert ved sammenkoblet MKM. Siden regionens befolkning i en periode i svært stor grad er avhengig av regionens befolkning i tidligere perioder, er det nødvendig å kontrollere for befolkning i tidligere perioder. Modellene vil derfor inneholde enten lagget befolkningsmengde som kontrollvariabel, avhengig variabel på endringsform, eller begge deler.

I forhold til faste effekter, er det en svakhet ved metoden at den ikke kontrollerer for permanente forskjeller mellom regioner. Avhengig av måten modellen spesifiseres på, kan dette i større grad føre til at Gauss-Markov-forutsetningene brytes og gi mindre pålitelige resultater. Tilgang til gode kontrollvariabler er viktig for denne metoden. Med variabler som i stor grad kontrollerer for de permanente forskjellene mellom regionene, kan tverrsnittsvariasjonen i datamaterialet benyttes uten at modellen gir store skjevheter på grunn av de permanente forskjellene.

### 4.2.2 Faste effekter

Faste regionseffekter benyttes i modeller der variasjon innen regioner, over tid er av interesse. For å forklare metoden, tar jeg utgangspunkt i en forenklet modell (4.12), med studenttall i regionen som interessevariabel.

$$Bef_{it} = \beta_1 * Stud_{it} + \epsilon_{it} \quad (4.12)$$

Restleddet kan dekomponeres til en regionspesifikk komponent  $a_i$  og en idiosynkratisk komponent  $v_{it}$ .

$$\epsilon_{it} = a_i + v_{it} \quad (4.13)$$

Det vil si at 4.12 kan skrives som om til 4.14.

$$Bef_{it} = \beta_1 * Stud_{it} + a_i + v_{it} \quad (4.14)$$

Det regionspesifikke restleddet fanger opp ikke-observerbare effekter som er konstante over tid og kun varierer mellom regioner, mens det idiosynkratiske restleddet fanger opp effekter av utelatte variabler som varierer både over tid og mellom regioner. Jeg tar



gjennomsnittet over tid for hver region, og får ligning (4.15).

$$\overline{Bef}_i = \beta_1 * \overline{Stud}_i + a_i + \bar{v}_i \quad (4.15)$$

Ved å trekke (4.15) fra (4.14), kan det regionsspesifikke restleddet fjernes fra modellen.

$$(Bef_{it} - \overline{Bef}_i) = \beta_1(Stud_{it} - \overline{Stud}_i) + (\epsilon_{it} - \bar{v}_i) \quad (4.16)$$

Ligning (4.16) er faste effekter-modellen<sup>2</sup>. Den avhengige variabelen og forklaringsvariabelen måles som avvik fra gjennomsnittet i regionen, slik at ren tverrsnittsvariasjon transformeres bort.

Faste tidseffekter fungerer på samme måte, bortsett fra at det nå kun er variasjon over tid og ikke kun mellom regioner transformeres bort. Det betyr ren tidsvariasjon som er felles for alle kommunene, for eksempel at andelen studenter i befolkningen øker utover i studieperioden<sup>3</sup>. Med både faste tidseffekter og faste regionseffekter i modellen, benyttes bare variasjon i datamaterialet som varierer både over tid og mellom regioner.

Kort oppsummert er fordelene med faste regionseffekter at utelatt variabelskjevhet fra permanente forskjeller mellom regioner blir transformert bort. Ulempen er at dersom mye av variasjonen i datamaterialet er tverrsnitt og lite over tid, vil bare en liten del av variasjonen i materialet bli benyttet. I tilfeller der sammenligninger mellom regioner er nødvendig, kan ikke faste regionseffekter benyttes.

Faste regionseffekter fanger ikke opp skjevhet som kommer fra simultanitet og utelatte variabler som varierer både over tid og mellom regioner. Problemet med utelatte variabler kan i teorien løses ved å finne variablene som varierer over tid, og inkludere disse som kontrollvariabler. I praksis er dette vanskelig. Analysene i denne oppgaven bruker ofte lagget befolkning som kontrollvariabel. Denne variabelen kan fungere som en proxy for noen av de utelatte variablene. For å fjerne skjevhet fra simultanitet, kreves instrumentvariabler. Metodene som benyttes i studien vil derfor ikke forventes å produsere et fasitsvar. Målet med metodebruken har vært å løse de økonometriske utfordringene på best mulig måte, gitt datamaterialet.

---

<sup>2</sup>Metoden kalles 'within-groups'.

<sup>3</sup>Tabell 3.1.

## 4.3 Økonometriske utfordringer

De deskriptive tabellene i kapittel 3 viser at det finnes en sammenheng mellom befolkningsvekst og antall studenter. Jeg har tidligere forklart hvorfor sammenhengene ikke nødvendigvis er kausale. I dette delkapitlet går jeg nærmere inn på utfordringene knyttet til å finne kausalitet i sammenhengene.

For at en modell skal gi forventningsrette estimater, må som nevnt forventningsverdien til restleddet være lik null og ukorrelert med interessevariabelen. For å vise dette tar jeg utgangspunkt i en forenklet modell (4.17), med studenttallet i regionen som interessevariabel.

$$Bef_{it} = \beta_1 * Stud_{it} + \epsilon_{it} \quad (4.17)$$

Forventningsrette estimater innebærer at

$$E(\epsilon_{it} | Stud_{it}) = 0. \quad (4.18)$$

Ligning (4.18) sier at forklaringsvariabelen ikke avhenger systematisk av restleddet, hverken over tid eller på tvers av regioner. Dersom forutsetningen ikke holder, gir modellen forventningsskjevne estimater. De tre vanligste årsakene til endogenitet er målefeil, simultanitet og utelatt variabelskjevhet. Skjevhet på grunn av målefeil kan som tidligere nevnt bli relevant i forbindelse med sammenslåingen av data fra kommunenivå til regionsnivå. Siden de to hovedvariablene i studien ikke har manglende observasjoner og trolig inneholder få feil, antar jeg at målefeil er en ubetydelig kilde til skjevhet her. Målefeil vil av den grunn ikke utledes nærmere. Hovedkilden til endogenitet i forklaringsvariabelen i denne analysen er simultanitet og utelatt variabelskjevhet.

### 4.3.1 Simultanitet

I samfunnsøkonomi er det vanlig at to variabler har gjensidig påvirkning på hverandre. Det blir et problem når man vil finne effekten den ene av variablene har på den andre. I denne oppgaven er det modellene med studenttall i regionen som interessevariabel som er utsatt for simultanitet. Jeg tar derfor utgangspunkt i en forenklet toveis modell for studenttall og befolkning for å vise dette.

$$Bef_{it} = \beta_0 + \beta_1 * Stud_{it} + u_{it} \quad (4.19)$$

$$Stud_{it} = \alpha_0 + \alpha_1 * Bef_{it} + v_{it} \quad (4.20)$$

Ifølge ligning (4.19) og (4.20) er studenttall og befolkning i en region simultant bestemt. Det betyr at en endring i antall studenter påvirker befolkningen, samtidig som en end-

ring i befolkningen påvirker antall studenter. Settes (4.19) inn i (4.20), kan det vises at forutsetning (4.18) brytes dersom  $\alpha_1$  er ulik null (Woolridge, 2013).  $Stud_{it}$  er ikke lenger ukorrelert med restleddet  $u_{it}$ .

$\alpha_1$  er positiv hvis antall studenter i en by avhenger positivt av størrelsen på byen. Det betyr at den er positiv hvis studenter søker seg til byer som vokser. I praksis er det ikke usannsynlig at dette er tilfelle. En by blir trolig mer attraktiv for studenter hvis den blir større. Det er også rimelig å anta at  $\beta_1$  er positiv i praksis.

MKM-estimatene får skjevhet oppover dersom  $\alpha_1$  og  $\beta_1$  begge er positive. Hvis befolkningen i en region øker av andre årsaker enn studenttall, og dette igjen fører til at byen blir mer attraktiv for studenter, vil modellen fange opp årsak-virkning motsatt i forhold til virkeligheten. Den vil estimere at befolkningen øker fordi studenttallet øker, også i tilfeller der det egentlig er studenttallet som øker fordi befolkningen øker. Det betyr at den omvendte kausaliteten fører til at effekten av antall studenter på befolkningsmengde og befolkningsvekst trolig overestimeres.

Noen studier prøver å løse problemet med simultanitet ved å legge på en lagg på forklaringsvariabelen. Blant annet bruker Vergara (2010) og Aschhoff og Schmidt (2008) simultanitet som argument for å bruke lagg på forklaringsvariablene. Tanken bak er at forklaringsvariabelen ikke kan bli simultant bestemt med den avhengige variabelen når de ikke er i samme periode. Denne masteroppgaven bruker også lagget forklaringsvariabel i de fleste modellene. Ved å bruke denne logikken, kan man si at studentene ikke kjenner til regionens befolkningsvekst i neste periode, og derfor kan den ikke påvirke studenttallet. Med andre ord, dersom ligning (4.20) skrives om til

$$Stud_{it} = \alpha_0 + \alpha_1 * Bef_{it} + \alpha_2 * Bef_{it+1} + v_{it} \quad (4.20')$$

vil  $\alpha_2$  være lik null. Da er det ikke lenger skjevhet på grunn av simultanitet.

Problemet med denne metoden, er at seriekorrelasjon i forklaringsvariabelen og avhengig variabel kan gjøre at skjevheten består. Reed (2013) viser at dersom det er simultanitet mellom forklaringsvariabel og avhengig variabel samtidig som det er seriekorrelasjon, vil simultanitet fremdeles gi forventningsskjevne estimater i modellen. Det betyr at  $\alpha_2$  ikke nødvendigvis er lik null likevel. Dersom befolkning og antall studenter er simultant bestemt, og antall studenter er korrelert med antall studenter i forrige periode, vil modellene fremdeles overestimere effekten av antall studenter.

Det betyr at lagget forklaringsvariabel til en viss grad kan hjelpe mot omvendt kausalitet. Hvor mye skjevhet som blir igjen avhenger av graden av seriekorrelasjon. I praksis er nok forklaringsvariabelen og avhengig variabel relativt sterkt seriekorrelert i de fleste spesifikasjonene. Unntaket er trolig i modellene der alle variablene er differensierte, altså

i modellene som ser på endring i forhold til forrige periode. Differensiering av variablene reduserer seriekorrelasjon (Woolridge, 2013).

Motivasjonen bak å bruke lagget forklaringsvariabel er at det i hovedsak er mer interessant å se hvordan studenttallet påvirker befolkning i neste periode enn i samme periode. En eventuell reduksjon i skjevhet fra simultanitet er en bonus.

Det kan også tenkes at det er simultanitet i modellene som benytter dummyvariabler for høyere utdanningsinstitusjoner. Dersom studiestedene opprettes som tiltak for å øke veksten i regioner som sliter, kan en nedgang i befolkningen øke sannsynligheten for at en region får høgskole. Da vil modellene vise at høgskolene gir regionene svak vekst, mens årsak-virkning i realiteten er at svak vekst gir regionen høgskole. Selv om dette trolig er et mindre problem enn for modellene som benytter studenttall, kan det i teorien tenkes at modellene som ser på utbygging av høgskoler underestimerer virkningen av høgskolene på grunn av simultanitet.

### **Instrumentvariabler**

En potensielt mer effektiv måte å håndtere simultanitet på, er å bruke instrumentvariabler (Woolridge, 2013). En instrumentvariabel må være korrelert med forklaringsvariabelen (relevanskriteriet) og ukorrelert med restleddet i modellen (eksogenitetskriteriet). Dersom disse forutsetningene er oppfylt, er det mulig å skille ut deler av variasjonen i forklaringsvariabelen som er eksogen, og kun benytte seg av denne i estimeringene. Høyere korrelasjonen mellom forklaringsvariabel og instrumentvariabel, øker andelen av eksogen variasjon som kan benyttes.

I praksis vil utfordringen være å finne variabler som oppfyller begge kriteriene. Mangel på en slik variabel er årsaken til at jeg ikke benytter instrumentvariabler i denne oppgaven. Dersom oppgaven hadde hatt et større omfang, eller gode instrumentvariabler var enkelt tilgjengelig, ville instrumentvariabler ha vært en naturlig utvidelse. Fokuset har i stedet vært å bruke modellspesifikasjoner som i størst mulig grad reduserer, men trolig ikke eliminerer, skjevhet fra simultanitet og utelatte variabler.

### **4.3.2 Utelatte variabler**

Dersom en variabel som er korrelert med både forklaringsvariabel og avhengig variabel blir utelatt fra modellen, oppstår utelatt variabelskjevhet (Woolridge, 2013). Dette fører til at forutsetning (4.18) ikke holder, og estimatene blir forventningsskjeve.

For å vise dette, tar jeg utgangspunkt i en modell uten kontrollvariabler med lagget

studenttall som interessevariabel.

$$Bef_{it} = \beta_0 + \beta_1 * Stud_{it-1} + \epsilon_{it} \quad (4.21)$$

I (4.21) finnes det mange variabler som er korrelerte med både  $Stud_{it-1}$  og  $Bef_{it}$ . Dersom en utelatt variabel, for eksempel antall sysselsatte i industrien, er positivt korrelert med begge variablene, vil modellen overestimere  $\beta_1$ . Deler av effekten denne variabelen har på befolkningen vil bli plukket opp av forklaringsvariabelen den er korrelert med. Den mest åpenbare måten å løse problemet på, er å inkludere kontrollvariabler.

$$Bef_{it} = \beta_0 + \beta_1 * Stud_{it-1} + \beta_2 * SysInd1970_i + \beta_3 * Hogutd1970_i + \epsilon_{it} \quad (4.22)$$

Dersom antall sysselsatte i industrien og antall høyere utdannede i en region i 1970 var de eneste utelatte variablene, ville ligning (4.22) vært en tilstrekkelig løsning. Men i denne studien er det i praksis umulig å håndtere utelatt variabelskjevhet slik. Det finnes altfor mange variabler som må inkluderes.

Utelatt variabelskjevhet vil i hovedsak føre til en overestimering av effektene av utdanningsinstitusjonene på befolkningen. Dersom hele datasettet benyttes til å estimere en modell som (4.21), vil for eksempel historiske befolkningstall være sterkt positivt korrelert med både dummyvariabler for studiested, studenttall og folkemengde. Et mulig unntak kan igjen være i dummyvariabel-modellene som ser på små regioner. Dersom studiestedene blir opprettet på bakgrunn av politiske tiltak for å øke veksten i en region som sliter, kan dummyvariablene for høgskole være korrelerte med utelatte variabler som påvirker befolkningsveksten negativt.

Faste regioneffekter tar bort skjevhet fra permanente forskjeller mellom regioner, som for eksempel historiske befolkningstall og de statiske kontrollvariablene i (4.22). Dette minsker utelatt variabelskjevhet, og er en sentral strategi for å håndtere utelatt variabelskjevhet. Problemet er fremdeles til stede dersom det finnes variabler som varierer over tid og er korrelerte med både forklaringsvariabel og avhengig variabel. Et eksempel kan være befolkningsandelen mellom 15 og 35 år i en region, som kan tenkes å være positivt korrelert med både antall studenter og befolkningsvekst. Dersom denne andelen øker, blir effekten av studenttallet på befolkning overestimert, fordi studenttallet plukker opp deler av effekten som befolkningsandelen mellom 15 og 35 år har på befolkningsveksten.

Den andre hovedstrategien for å håndtere utelatt variabelskjevhet i oppgaven er å bruke lagget avhengig variabel som kontrollvariabel. I følge Woolridge (2013) kan lagget avhengig variabel fungere som en proxy for andre utelatte variabler som mangler data. Ideen er at dersom lagget avhengig variabel er korrelert med utelatte variabler som lager skjevhet, vil lagget avhengig variabel fange opp deler av effekten fra de utelatte variablene og

forbedre presisjonen på estimatene.

Et problem med denne strategien, er at seriekorrelasjon i restleddet kan føre til skjeve estimater ved bruk av en lagget avhengig variabel som kontrollvariabel. Aachen (2000) viser at med seriekorrelasjon i restleddet, kan en lagget avhengig variabel undertrykke estimatene til forklaringsvariablene og gjøre de mindre relevante enn de er i virkeligheten. Mer seriekorrelasjon i restleddet øker underestimeringen. Baker (2007) argumenterer for at dersom seriekorrelasjonen i restleddet skyldes en utelatt variabel som også er seriekorrelert, kan denne variabelen forårsake utelatt variabelskjevhet. En lagget uavhengig variabel kan da fungere som proxy og gi mer forventningsrette estimater enn en modell uten.

I denne studien har jeg vurdert at lagget uavhengig variabel trolig gjør større nytte enn skade i de fleste spesifikasjonene. Dessuten må det skilles mellom modeller der den brukes som proxy for utelatte kontrollvariabler, og i modeller der den hører hjemme som en del av spesifikasjonen. I de enkleste modellene uten faste effekter eller avhengig variabel på endringsform, er befolkningstallet i forrige periode opplagt en faktor som er relevant for befolkningstallet i nåværende periode. I modeller med avhengig variabel på endringsform eller med faste regioneffekter, er lagget befolkningsmengde inkludert som proxy for utelatte variabler. Estimeringene der lagget befolkningsmengde er inkludert på høyresiden er trolig mer forventningsrette, særlig i de enkleste modellene. Likevel må det tas hensyn til når resultatene tolkes at inkluderingen av lagget avhengig variabel muligens fører til en underestimering av effekten av utdanningsinstitusjonene.

### 4.3.3 Heteroskedastisitet og seriekorrelasjon

Heteroskedastisitet og seriekorrelasjon i restleddet kan som nevnt i starten av kapittel 4.2 føre til feilberegning av standardavviket, og vanskeliggjøre hypotesetesting. Seriekorrelasjon kan i tillegg problematisere bruk av lagget avhengig variabel som kontrollvariabel.

Ingen seriekorrelasjon i restleddet betyr at restleddet i en periode  $t$ , ikke er korrelert med restleddet i en annen periode  $s$ .

$$\text{cov}(\epsilon_{it}, \epsilon_{is} | X) = 0 \quad (4.23)$$

Dersom restleddet i en periode er korrelert med restleddet i forrige periode, holder ikke forutsetning (4.23).

Homoskedastisitet er at restleddvariansen er konstant for alle regioner og tidsperioder.

$$\text{Var}(\epsilon_{it}|X) = \sigma_{it}^2 \quad (4.24)$$

Dersom forutsetning (4.24) brytes, har modellen heteroskedastisitet.

Jeg antar at heteroskedastisitet er til stede i de fleste modellene. For å få standardavvik som muliggjør korrekt hypotesetesting under heteroskedastisitet, estimerer jeg modellene på en måte som gir robuste standardavvik. Ved å anta at restleddet varierer mellom regioner, kan standardavviket gjøres robust ved å bruke 'clustering' på regionnivå i estimeringen<sup>4</sup>.

### 4.3.4 Faste effekter og lagget avhengig variabel

På grunn av dårlig tilgang til gode kontrollvariabler som varierer både over tid og mellom regioner, er både faste effekter og lagget avhengig variabel viktige kontroller i analysen. Et problem er at faste effekter og lagget avhengig variabel i kombinasjon, fører til skjevhet dersom det er få tidsserier i datamaterialet. Nickell (1981) viser at når faste effekter innføres i en dynamisk modell med lagget avhengig variabel, blir det korrelasjon mellom restleddet og forklaringsvariabelen i modellen. Restleddet og lagget avhengig variabel blir målt som avvik fra gjennomsnittet for alle periodene når faste effekter innføres, og disse avvikene er korrelert.

Skjevheten blir mindre etterhvert som antall tidsperioder øker, men Beck og Katz (2011) nevner 20 tidsserier som et nivå der skjevheten er mindre betydelig. Denne oppgaven har bare fem perioder etter sammenslåingen til femårsperioder, slik at skjevhet fra denne kilden ikke kan neglisjeres. Det må vurderes om lagget avhengig variabel gjør mer skade enn nytte i modellene med faste regioneffekter. Jeg velger derfor å vise faste effekter-modeller både med og uten lagget avhengig variabel, og sammenligne resultatene.

---

<sup>4</sup>Kommandoene 'vce(robust)' eller 'cluster' brukes i STATA

# Kapittel 5

## Empiriske resultater

De empiriske studiene presentert i kapittel 2.4 har fått blandede resultater i forsøkene på å avdekke kausale sammenhenger mellom utdanningsinstitusjoner og regional økonomisk vekst. Norske undersøkelser (Lasse Stambøl (2011), Bjørnar Sæther et al. (2000)) har vist at det finnes en sammenheng mellom høyere utdanningsinstitusjoner og befolkning. Målet med analysene har vært å se om datamaterialet indikerer at denne sammenhengen er kausal, altså at utdanningsinstitusjonene påvirker befolkningsveksten, og ikke bare at det er en korrelasjon mellom disse to. De økonometriske utfordringene knyttet til dette ble diskutert i forrige kapittel.

Jeg deler resultatkapitlet inn i to hoveddeler. Delkapittel 5.1 bruker dummyvariabler som interessevariabler, og tester om tilstedeværelsen av utdanningsinstitusjoner påvirker befolkningsveksten. Hovedanalysen i delkapitlet leter etter effekt av høgskoler som etableres eller utvides i løpet av studieperioden. Jeg inkluderer også en analyse som undersøker om små regioner med relativt store høgskoler har hatt høyere befolkningsvekst enn andre små regioner.

I delkapittel 5.2 bruker jeg studenttallet som interessevariabel. Her vil jeg se om omfanget av utdanningsinstitusjonene påvirker befolkningsveksten, og utnytter variasjon over tid innenfor regionene i mye større grad. Jeg starter med en analyse for å gi overblikk, der hele datamaterialet brukes. I delkapittel 5.2.1 inkluderes faste regionseffekter for å fjerne skjevhet som kommer fra permanente forskjeller mellom regioner. Delkapittel 5.2.2 tar for seg langtids- og korttidseffekter. Til slutt ser jeg på om effekten varierer mellom regioner av ulik størrelse i delkapittel 5.2.3.



## 5.1 Virkning av utdanningsinstitusjoner på befolkningsveksten

I kapittel 3 nevnte jeg at noen regioner har fått nye eller sterkt utvidede studiesteder i løpet av studieperioden. Dette benytter jeg meg av i hovedanalysen med dummyvariabler.

Tabell 3.3 viste at de sju regionene med ny høyskole har hatt større befolkningsvekst enn gjennomsnittet. Veksten var imidlertid høyere i de første periodene, og mer lik de andre regionene i de to siste periodene. Ved å generere dummyvariabler etter reglene forklart i kapittel 3.2.1, gjør jeg et forsøk på å teste empirisk om utbyggingene har hatt en effekt på befolkningsveksten i disse regionene.

Det hadde vært ideelt å bruke en difference-in-differences-analyse for dette formålet (Wooldridge, 2013). Der sammenlignes befolkningsveksten i regionen før og etter den nye høyskolen med veksten før og etter i kontrollregioner uten ny høyskole. Dersom forskjellen mellom befolkningsveksten i regionene endrer seg før og etter høyskoleutbyggingen, kan dette være en effekt av høyskolen. En slikt type analyse blir tungvint å utføre, ettersom tidspunktene for høyskoleutbyggingene er forskjellige i de ulike regionene. Da blir det vanskelig å definere etter-perioden for kontrollregionene. Jeg velger derfor å bruke en enklere metode, som jeg beskriver i kapittel 4.1.1.

Variablene  $DNyskole$  og  $DUtvidet$ , generert av reglene (3.1) og (3.2), får verdien 1 i den perioden etableringen eller utvidelsen skjer. Det betyr at de måler den umiddelbare effekten.  $DEtterNS$  og  $DEtterUS$  har verdien 1 også i periodene etter etableringen, for å fange opp eventuelle langsiktige effekter.

For at variablene skal gi korrekte estimater, forutsettes det at regionene er like med tanke på forventet befolkningsvekst, gitt kontrollvariablene som er inkludert. Det er urealistisk å tro at dette er oppfylt. Dersom metoden ikke fanger opp effekter, betyr ikke det nødvendigvis at disse effektene ikke eksisterer. Dette forsterkes av at det er få tidsperioder etter etableringene, og relativt få regioner som har fått ny skole.

Modell (1) i tabell 5.1 er uten kontrollvariabler. Denne modellen viser at observasjonene med ny skole har hatt positiv vekst, men effekten er lite signifikant. I modell (2) legger jeg på sysselsetting i regionen i 1970 og antall personer med høyere utdanning i 1970 som kontrollvariabler. Dette gjør begge interessevariablene enda mindre signifikante. Resultatene for  $DNyskole$  er signifikante på 10 prosent signifikansnivå i modell (3), (4) og (5), når ytterligere kontrollvariabler legges til. I følge modell (4) vil en ny skole i en region gi en befolkningsvekst på 2,82 prosent i utbyggingsperioden.

Tabell 5.1: Resultater: Effekter av ny høyskole.

|                              | (1)                    | (2)                    | (3)                    | (4)                    | (5)                    |
|------------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
|                              | $\Delta \text{LogBef}$ | $\Delta \text{LogBef}$ | $\Delta \text{LogBef}$ | $\Delta \text{LogBef}$ | $\Delta \text{LogBef}$ |
| <i>DNyskole</i>              | 0.0121<br>(1.58)       | 0.0102<br>(1.20)       | 0.0279<br>(1.68)       | 0.0282<br>(1.76)       | 0.0262<br>(1.70)       |
| <i>DEtterNS</i>              | 0.00396<br>(0.54)      | 0.00175<br>(0.25)      | -0.00159<br>(-0.20)    | -0.00267<br>(-0.32)    | -0.00528<br>(-0.55)    |
| <i>LogSysind70</i>           |                        | 0.00282<br>(0.57)      | -0.0117<br>(-1.61)     | -0.0147<br>(-2.00)     | -0.0158<br>(-1.93)     |
| <i>LogHogutd70</i>           |                        | 0.00938<br>(1.65)      | 0.0223**<br>(2.90)     | 0.0144<br>(1.21)       | 0.0138<br>(1.12)       |
| <i>LogStud<sub>t-1</sub></i> |                        |                        | -0.000380<br>(-0.21)   | -0.0000921<br>(-0.05)  | -0.000665<br>(-0.23)   |
| $\Delta \text{LogStud}$      |                        |                        |                        |                        | -0.0000891<br>(-0.02)  |
| <i>LogBef<sub>t-1</sub></i>  |                        |                        |                        | 0.0138<br>(1.17)       | 0.0163<br>(1.38)       |
| Konstantledd                 | 0.0164***<br>(5.51)    | -0.0686***<br>(-4.59)  | -0.0522<br>(-1.64)     | -0.119*<br>(-2.04)     | -0.128*<br>(-2.17)     |
| <i>N</i>                     | 356                    | 356                    | 169                    | 169                    | 164                    |

*t*-verdier i parentes

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

I modell (5) kontrolleres det for endring i studenttallet. Dette innebærer at dummyvariablene måler effekter som kommer i tillegg til den vanlige effekten av økt studenttall. Koeffisienten og *t*-verdien for *DNyskole* endrer seg lite fra modell (4) til modell (5). Dette kan indikere at investeringene i forbindelse med utbyggingen er viktigere for befolkningsveksten enn en tilsvarende økning i studenttallet på et eksisterende studiested. Ingen av modellene finner signifikante effekter av høyskoleetableringene i periodene etter.

I tabell 5.2, som ser på effekten av både nye og utvidede skoler, finner modell (1) at *DUtvidet* er signifikant positiv uten kontrollvariabler inkludert. Alle signifikante effekter forsvinner når kontrollvariablene legges til, fra modell (2) og utover. Dette tyder på at regionene som får ny eller utvidet høyskole vokser i perioden utbyggingen skjer, men at veksten kanskje skyldes andre faktorer enn utbyggingen.

Korttidseffektene er mer signifikante enn langtidseffektene i begge modellene. Dette er som forventet. Investeringer, arbeidsplassetableringer og andre etterspørselssjokk i forbindelse med en etablering eller utvidelse av et studiested, har opplagt en positiv effekt

Tabell 5.2: Resultater: Effekter av ny og utvidet høyskole.

|                              | (1)                    | (2)                    | (3)                    | (4)                    | (5)                    |
|------------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
|                              | $\Delta \text{LogBef}$ | $\Delta \text{LogBef}$ | $\Delta \text{LogBef}$ | $\Delta \text{LogBef}$ | $\Delta \text{LogBef}$ |
| <i>DUtvidet</i>              | 0.0128*<br>(2.35)      | 0.00401<br>(0.66)      | 0.00922<br>(1.12)      | 0.00949<br>(1.15)      | 0.00401<br>(0.37)      |
| <i>DEtterUS</i>              | 0.0134<br>(1.57)       | 0.00421<br>(0.63)      | -0.000321<br>(-0.05)   | -0.000212<br>(-0.03)   | -0.000261<br>(-0.03)   |
| <i>LogHogutd70</i>           |                        | 0.00905<br>(1.60)      | 0.0224**<br>(2.75)     | 0.0148<br>(1.22)       | 0.0144<br>(1.11)       |
| <i>LogSysind70</i>           |                        | 0.00282<br>(0.58)      | -0.0118<br>(-1.61)     | -0.0148<br>(-1.98)     | -0.0150<br>(-1.83)     |
| <i>Logstud<sub>t-1</sub></i> |                        |                        | -0.000527<br>(-0.29)   | -0.000225<br>(-0.12)   | -0.000353<br>(-0.12)   |
| $\Delta \text{LogStud}$      |                        |                        |                        |                        | 0.00349<br>(0.43)      |
| <i>LogBef<sub>t-1</sub></i>  |                        |                        |                        | 0.0133<br>(1.16)       | 0.0147<br>(1.26)       |
| Konstantledd                 | 0.0158***<br>(5.25)    | -0.0663***<br>(-4.26)  | -0.0506<br>(-1.62)     | -0.115*<br>(-2.02)     | -0.125*<br>(-2.10)     |
| <i>N</i>                     | 356                    | 356                    | 169                    | 169                    | 164                    |

*t*-verdier i parantes

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

på den økonomiske aktiviteten i en region. Etter den initielle etterspørselsøkningen, er det tilbudseffekter og eventuelle multiplikatoreffekter av den opprinnelige etterspørselsøkningen som må øke veksten. Disse effektene kan eksistere, men vil være vanskeligere å fange opp empirisk med denne modellen.

Etter inkluderingen av kontrollvariabler, får *DNyskole* mer signifikante effekter enn *DUtvidet*. Dette kan skyldes at effekten fra nye skoler faktisk er mer signifikant, men også andre faktorer. *DNyskole* er lik 1 i bare sju observasjoner, mens *DUtvidet* er lik 1 i 18 observasjoner. Med 7 observasjoner kan én eller to regioner med sterk vekst av andre årsaker få stor påvirkning på koeffisienten. Med 18 observasjoner er det større sannsynlighet for at dette jevner seg ut.

Kort oppsummert finner modellene positive kortsiktige effekter av nye høyskoler som er noenlunde signifikante. Når store høyskoleutvidelser inkluderes, forsvinner signifikansen. Modellene finner ingen langsiktige effekter.

### 5.1.1 Små høgskoleregioner sammenlignet mot andre små regioner

I kapittel 3 om datamaterialet, nevner jeg at det finnes små regioner i Norge med relativt mange studenter. Ettersom befolkningsveksten i en region er korrelert med folketallet, kan det være en fordel å benytte regioner med noenlunde likt folketall for sammenligninger på tvers av regioner. Ved å velge ut små regioner med relativt store høgskoler og bruke andre mindre regioner som kontrollgruppe, gjør jeg en analyse som utnytter dette. De 12 utvalgte regionene ble presentert i kapittel 3.2.1. Tabell 3.4 viste at disse regionene har hatt en bedre gjennomsnittlig befolkningsvekst enn kontrollgruppen i tre av fire perioder.

Tabell 5.3: Resultater: Små regioner med høgskole sammenlignet med små regioner uten.

|                              | (1)                    | (2)                    | (3)                    | (4)                    |
|------------------------------|------------------------|------------------------|------------------------|------------------------|
|                              | $\Delta \text{LogBef}$ | $\Delta \text{LogBef}$ | $\Delta \text{LogBef}$ | $\Delta \text{LogBef}$ |
| <i>DLitenmedskole</i>        | -0.00121<br>(-0.14)    | -0.00861<br>(-0.87)    | -0.00874<br>(-0.84)    | -0.00108<br>(-0.09)    |
| <i>LogSysind70</i>           | 0.00782<br>(1.53)      | -0.00704<br>(-0.56)    | -0.00706<br>(-0.56)    | -0.0106<br>(-0.97)     |
| <i>LogHogutd70</i>           | -0.000164<br>(-0.02)   | -0.00604<br>(-0.28)    | -0.00948<br>(-0.29)    | -0.0171<br>(-0.58)     |
| <i>LogStud<sub>t-1</sub></i> |                        | 0.00210<br>(0.91)      | 0.00238<br>(0.86)      | 0.000768<br>(0.21)     |
| <i>LogBef<sub>t-1</sub></i>  |                        |                        | 0.00589<br>(0.19)      | 0.0186<br>(0.62)       |
| <i>NordNorge</i>             |                        |                        |                        | -0.0221<br>(-1.95)     |
| Konstantledd                 | -0.0411<br>(-1.09)     | 0.102<br>(1.14)        | 0.0638<br>(0.33)       | 0.0237<br>(0.12)       |
| <i>N</i>                     | 252                    | 77                     | 77                     | 77                     |

*t*-verdier i parentes

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Jeg lager en dummyvariabel *DLitenmedskole*, som har verdien 1 for alle observasjonene i de 12 regionene. For å minske forskjellene mellom regionene, utelukker jeg alle regioner som har mer enn 40 000 innbyggere fra estimeringene. Det betyr at kontrollgruppen blir alle de andre regionene med mindre enn 40 000 innbyggere. Tabell 5.3 viser at *DLitenmedskole* ikke har signifikant effekt på befolkningsveksten. I modell (4) inkluderer jeg en

dummyvariabel som er lik 1 dersom observasjonen er fra en region i Nord-Norge. Dette er den eneste variabelen som er signifikant på ti prosent signifikansnivå. Ved å inkludere denne, går koeffisienten til DLitenmedskole mot null, og mister mer signifikans. Alle modellene inkluderer tidsdummyer. Resultatene indikerer at høyskolene i små regioner ikke har ført til økt befolkningsvekst i perioden 1976 til 2000.

En slik studie kunne ha tjent på å ha befolkningsdata lengre bakover i tid. Dette ville gjort det mulig å sammenligne befolkningsveksten i høyskoleregionene før og etter høyskoleetableringene, med befolkningsutviklingen før og etter i lignende regioner. Også her ville det ha vært en utfordring å definere etter-perioden for kontrollgruppen, ettersom høyskolene opprettes på ulike tidspunkt. Etableringen av disse distriktshøyskolene skjer i hovedsak på 1970-tallet. En mulighet kunne ha vært å bruke perioden 1950-1969 som før-periode, og perioden 1980-1999 som 'etter-periode'. Dersom regionene som fikk distriktshøyskole hadde hatt svakere vekst enn kontrollgruppen i før-perioden, ville dette ha vært en indikasjon på at høyskolene ble plassert i regioner som hadde behov for vekstfremmende tiltak.

De begrensede effektene som dummyvariabel-analysene finner, trenger ikke å bety at effektene av høyskolene ikke eksisterer. Særlig gjelder dette langtidseffektene i tabell 5.1 og 5.2, som er vanskelige å fange opp. Få tidsperioder og store ulikheter mellom regionene er et problem. Det kan også tenkes at de små regionene med høyskoler deler fellestrekk som påvirker befolkningsveksten negativt. Små regioner kan ha fått høyskoler på bakgrunn av politiske beslutninger for å øke veksten i regioner som i utgangspunktet sliter. Disse fellestrekkene trenger ikke å bli fanget opp av kontrollvariablene. Det er mulig at de 12 små høyskoleregionene ville hatt mye svakere befolkningsutvikling i fravær av høyskolene. De ikke-signifikante funnene samsvarer forøvrig med Persson og Regner (2004), som også bruker dummyvariabler for høyskole som interessevariabler.

## 5.2 Effekt av studenttall på befolkning

Jeg finner få signifikante effekter i dummyvariabel-analysene. Samtidig er ikke datamaterialet godt nok til at jeg kan fastslå at effektene ikke eksisterer. For å utforske høyskoler og universiteter sin effekt på befolkning videre, går jeg over til å bruke antall studenter som forklaringsvariabel. Da vil jeg måle omfanget av høyere utdanning i regionen, i stedet for bare hvorvidt regionen har en høyere utdanningsinstitusjon eller ikke. Jeg starter analysene om studenttallet sin effekt på befolkningen i regionen med en grunnleggende modell som benytter hele datasettet.

Tabell 5.4 viser resultatet av en sammenkoblet MKM-modell med tidsdummyer, der all variasjon i datasettet utnyttes. Variablene er på logaritmeform, slik at den finner prosent-

vise endringer. Modell (1) i tabell 5.4 er uten kontrollvariabler. Modell (2) har to statistiske kontrollvariabler, antall sysselsatte i industrien i en region i 1970, og antall personer med høyere utdanning i regionen i 1970. Modell (3) bruker lagget avhengig variabel som kontroll, og modell (4) har både lagget avhengig variabel og de to statistiske variablene som kontrollvariabler.

Tabell 5.4: Resultater: Sammenkoblet MKM - Grunnmodell

|                                      | (1)                  | (2)                 | (3)                  | (4)                 |
|--------------------------------------|----------------------|---------------------|----------------------|---------------------|
|                                      | <i>LogBef</i>        | <i>LogBef</i>       | <i>LogBef</i>        | <i>LogBef</i>       |
| <i>LogStud</i> <sub><i>t</i>-1</sub> | 0.273***<br>(5.19)   | -0.0216<br>(-0.97)  | 0.000324<br>(0.17)   | -0.00114<br>(-0.60) |
| <i>Logsysind</i> 1970                |                      | 0.210*<br>(2.09)    |                      | -0.0147<br>(-1.97)  |
| <i>Loghogutd</i> 1970                |                      | 0.587***<br>(4.67)  |                      | 0.0161<br>(1.43)    |
| <i>LogBef</i> <sub><i>t</i>-1</sub>  |                      |                     | 1.014***<br>(234.19) | 1.013***<br>(88.32) |
| Dummy for periode 3                  | -0.112<br>(-1.73)    | 0.0303<br>(1.87)    | -0.00422<br>(-1.23)  | -0.00322<br>(-0.90) |
| Dummy for periode 4                  | -0.209**<br>(-2.87)  | 0.0627**<br>(2.78)  | 0.00112<br>(0.29)    | 0.00262<br>(0.70)   |
| Dummy for periode 5                  | -0.298***<br>(-4.02) | 0.0949**<br>(3.17)  | 0.000587<br>(0.13)   | 0.00235<br>(0.50)   |
| Konstant                             | 9.342***<br>(30.52)  | 4.792***<br>(18.05) | -0.136**<br>(-3.00)  | -0.115*<br>(-2.01)  |
| <i>N</i>                             | 169                  | 169                 | 169                  | 169                 |

*t*-verdier i parantes

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Resultatene viser som forventet at det er en klar sammenheng mellom studenttall og befolkning når kontrollvariabler ikke inkluderes. I følge (1) vil én prosent høyere studenttall gi 0,273 prosent større folketall i perioden etter. Hovedproblemet er at høgskolene og universitetene ikke er tilfeldig plasserte i landet. De største byene har som regel de største universitetene med plass til flest studenter. Det vil også være slik at flere studenter vil til store byer enn til små byer av andre årsaker. Med andre ord er studentene i byene fordi byene er store, heller enn at byene er store fordi det er mange studenter der.

Når sysselsetting i industrien i 1970 og antall personer med høyere utdanning i 1970 inkluderes som kontrollvariabler i (2), er studenttallet ikke lenger signifikant. Begge kon-

trollvariablene er derimot positive og signifikante. Med lagget befolkningsmengde som kontrollvariabel i modell (3), er det ingen signifikant effekt av antall studenter. Det samme gjelder modell (4), med alle kontrollvariablene inkludert. Heller ikke de to statiske kontrollvariablene er signifikante på 5 prosent signifikansnivå her. Svakheten ved bruk av lagget avhengig variabel under potensiell seriekorrelasjon, som jeg nevnte i kapittel 4.3.2, må kanskje tas i betraktning i denne modellen.

Selv om kontrollvariablene har svakheter, vil koeffisienten for interessevariabelen i de tre siste modellene i mye større grad vise om det er en kausal effekt av antall studenter på befolkning. At effekten forsvinner i modell (2), (3) og (4), tyder på at effekten som (1) finner, i stor grad er et resultat utelatt variabel skjevhet. Denne enkle MKM-spesifikasjonen greier altså ikke å skille ut effekter av studenttallet i forrige periode som er ukorrelert med kontrollvariablene.

Antall observasjoner i modellene i tabell 5.4 er bare 169 fordi variablene er på logaritmeform, og når studenttallet i en observasjon er null, blir logaritmen udefinert. Det vil si at med en slik spesifikasjon, vil kun observasjoner der studenttallet er ulik null bli inkludert i estimeringen.

Tabellen illustrerer utfordringen med studien på en bra måte. Korrelasjonen mellom størrelsen på høyere utdanningsinstitusjoner i regionen og befolkning er opplagt, men kausale sammenhenger er vanskeligere å avdekke. Etttersom jeg ikke har mulighet til å inkludere alle relevante kontrollvariabler, og heller ikke har gode instrumentvariabler, er det en mulig at kausale sammenhenger eksisterer i virkeligheten uten at modellene i studien klarer å fange dem opp. I de neste delkapitlene tester jeg ulike metoder for å undersøke hva som kan hentes ut av datamaterialet.

### 5.2.1 Faste effekter

Den største kilden til skjevhet i estimatene i den første modellen i tabell 5.4, er permanente forskjeller mellom regionene. Ved å legge til faste regioneffekter og bare benytte variasjon over tid innenfor regionene, vil disse forskjellene transformeres bort.

Tabell 5.5 viser resultatene for to modeller med befolkning i regionen som avhengig variabel og lagget studenttall i regionen som interessevariabel. Modell (1) er uten kontrollvariabler, mens (2) har lagget befolkningsmengde som kontrollvariabel.

Resultatene i begge estimeringene viser at antall studenter har en signifikant positiv effekt på befolkningen i neste periode, men verdiene på koeffisientene endrer seg kraftig i modell (2). Når lagget befolkningsmengde legges til som kontrollvariabel, går verdien ned fra 1,930 til 1,071. I følge modell (1) vil én ekstra student gi en befolkningsøkning på

Tabell 5.5: Resultater: Grunnmodell med faste effekter

|   | (1)                    | (2)                |
|---|------------------------|--------------------|
|   | <i>Befolkning</i>      | <i>Befolkning</i>  |
| <i>Studenter</i> <sub><i>t</i>-1</sub>  | 1.930***<br>(13.19)    | 1.071***<br>(4.34) |
| <i>Befolkning</i> <sub><i>t</i>-1</sub> |                        | 0.894***<br>(8.10) |
| Dummy for periode 3                     | 461.5*<br>(2.41)       | -71.16<br>(-0.37)  |
| Dummy for periode 4                     | 1003.1*<br>(2.67)      | -29.67<br>(-0.12)  |
| Dummy for periode 5                     | 1279.0*<br>(2.22)      | -294.7<br>(-1.73)  |
| Konstantledd                            | 44773.1***<br>(127.66) | 4763.2<br>(1.00)   |
| <i>N</i>                                | 356                    | 356                |

*t*-verdier i parantes

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

nærmere to i neste periode. I modell (2) er dette tallet nesten halvert.

Modell (1) kontrollerer ikke for utelatte variabler som varierer både over tid og mellom regioner, og som er korrelerte med både interessevariabel og avhengig variabel. Lagget befolkningsmengde er trolig korrelert med mange av disse variablene, og kan derfor fungere som en proxy for andre kontrollvariabler. Tidligere befolkningstall er i seg selv en viktig kontrollvariabel. Store regioner vokser generelt raskere, og har også hatt en større vekst i antall studenter. Ved å kontrollere for lagget befolkningsvekst, fanges denne virkningen opp, og effekten av antall studenter isoleres bedre. Samtidig må det taes i betraktning at lagget avhengig variabel fører med seg skjevhet i modeller med faste effekter og få tidsserier. Likevel vil jeg anta at lagget avhengig variabel gjør mer nytte enn skade, slik at modell (2) har det beste resultatet.

I tabell 5.6 brukes endring i befolkningmengde som avhengig variabel, og lagget endring i antall studenter som interessevariabel. Resultatene er som forventet svært like modell (2) i tabell 5.5 og effekten endres lite når befolkning i forrige periode legges til som kontrollvariabel. Begge modellene tyder på at én ekstra student øker befolkningen i neste periode med litt over én. Å bruke differensiert avhengig variabel, er som å sette koeffisienten for lagget avhengig variabel i en vanlig modell lik én, og flytte den over på venstresiden for å få variabelen på endringsform. I modell (2) i tabell 5.5 er koeffisienten nokså nærme én,



Tabell 5.6: Resultater: Faste effekter på endringsform

|                          | (1)                 | (2)                 |
|--------------------------|---------------------|---------------------|
|                          | $\Delta Befolkning$ | $\Delta Befolkning$ |
| $\Delta Studenter_{t-1}$ | 1.116**<br>(3.41)   | 1.048**<br>(3.15)   |
| $Befolkning_{t-1}$       |                     | 0.0547<br>(1.07)    |
| Dummy for periode 4      | 167.3<br>(1.51)     | 127.5<br>(1.40)     |
| Dummy for periode 5      | 105.8<br>(0.89)     | 20.46<br>(0.28)     |
| Konstantledd             | 629.9***<br>(5.49)  | -1885.6<br>(-0.79)  |
| $N$                      | 267                 | 267                 |

*t*-verdier i parantes

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

slik at det er lite overraskende at resultatene blir like. I modell (2) i tabell 5.6 legges lagget befolkning til som ytterligere kontroll, men påvirker koeffisienten i mindre grad.

Siden modellen på endringsform krever en ekstra periode for å måle variablene, går antall observasjoner ned fra 356 til 267. At modellene likevel gir relativt like resultater er en styrke. Selv om modellene med faste effekter gir mer presise estimater enn vanlig minste kvadraters metode, må det regnes med at det fremdeles finnes skjevhet. Jeg har nevnt både problemer knyttet til utelatte variabler som varierer over tid, og at lagget avhengig variabel og faste effekter gir skjevhet når de kombineres og datasettet har få tidsserier. Simultanitet vil heller ikke fjernes av faste effekter, og fører kanskje til en overestimering av resultatene.

## 5.2.2 Kort- og langtidseffekter

I teorikapitlet delte jeg inn virkningene av utdanningsinstitusjonene på regionaløkonomien i tilbuds- og etterspørselseffekter. Virkningen av tilbudseffektene vil generelt ta lengre tid, mens etterspørselseffektene i hovedsak har en mer umiddelbar virkning. I forsøk på å fange opp noe av denne dynamikken, deler jeg effektene i langtids- og korttidseffekter.

Først tester jeg noen modeller som bruker endring i befolkning som avhengig variabel, og som har både endring i studenttall og lagget studenttall som interessevariabler. Endring i studenttall representerer de kortsiktige virkningene. Ved å bruke feiljusteringsmodellen,

kan langsiktige effekter beregnes ved hjelp av lagget studenttall og lagget befolkningsmengde (Woolridge, 2013). Feiljusteringsmodellen (Error Correction Model) måler hvor raskt en avhengig variabel kommer tilbake til likevekt etter en endring i interessevariabelen. Korttidseffektene er koeffisienten for endring i interessevariabelen. Langtidseffektene er koeffisienten for lagget interessevariabel delt på lagget avhengig variabel, og så multiplisert med -1 for å få motsatt fortegn:

$$\frac{LogStud_{t-1}}{LogBef_{t-1}} \quad (5.1)$$

Dersom feiljusteringsmodellen er passende å bruke, vil lagget avhengig variabel være negativ og signifikant, mens lagget interessevariabel er signifikant i forventet retning.

Tabell 5.7: Resultater: Korttids- og langtidseffekter - MKM (1) og (2), Faste effekter (3) og (4)

|                    | (1)                 | (2)               | (3)                 | (4)                |
|--------------------|---------------------|-------------------|---------------------|--------------------|
|                    | $\Delta LogBef$     | $\Delta Bef$      | $\Delta LogBef$     | $\Delta Bef$       |
| $LogBef_{t-1}$     | 0.0129**<br>(2.88)  |                   | -0.0719<br>(-1.77)  |                    |
| $\Delta LogStud$   | 0.00739<br>(1.54)   |                   | -0.00459<br>(-0.92) |                    |
| $LogStud_{t-1}$    | 0.00176<br>(0.73)   |                   | -0.0111*<br>(-2.02) |                    |
| $Befolkning_{t-1}$ |                     | 0.0276<br>(1.69)  |                     | -0.0688<br>(-1.12) |
| $\Delta Studenter$ |                     | 0.0103<br>(0.06)  |                     | 0.680**<br>(2.90)  |
| $Studenter_{t-1}$  |                     | 0.0534<br>(0.38)  |                     | 1.102***<br>(4.43) |
| Konstantledd       | -0.138**<br>(-3.02) | -505.5<br>(-1.44) | 0.871<br>(1.96)     | 2919.6<br>(1.11)   |
| $N$                | 164                 | 356               | 164                 | 356                |

*t*-verdier i parantes

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Tabell 5.7 viser resultatene av to estimeringer med sammenkoblet MKM og to estimeringer med faste regionseffekter. Alle modellene har faste tidseffekter. De to modellene med sammenkoblet MKM-estimering viser ingen signifikante effekter på hverken endring i studenttall eller lagget studenttall. Det betyr at de ikke fanger opp kortsiktige eller

langsiktige effekter. De kortsiktige effektene er positive og signifikante på 15 prosent signifikansnivå i modell (1) på logaritmeform, men dette er et veldig tynt resultat.

Modell (3) og (4), som bruker faste regionseffekter, fanger opp mer signifikante effekter. Modell (3), som er på logaritmeform, finner ikke signifikante kortsiktige effekter, men fanger opp en effekt på lagget studenttall som er signifikant negativ på 5 prosent signifikansnivå. Dette resultatet er noe uventet, selv om t-verdien tilsvarer et signifikansnivå som bare akkurat er sterkere enn fem prosent. Lagget befolkningstall er ikke signifikant, så feiljusteringsmodellen kan ikke benyttes her for å beregne langsiktig effekt.

I modell (4) har lagget studenttall som forventet tilnærmet lik effekt som i tabell 5.6. En ekstra student gir en befolkningsøkning på ca 1,1 i neste periode. Lagget befolkningsmengde er ikke signifikant, slik at langsiktige effekter ikke kan beregnes. De kortsiktige effektene er signifikante på ett prosent signifikansnivå, og indikerer at en ekstra student øker befolkningen med 0,68 personer i samme periode.

Alle modellene sliter altså med å fange opp langtidseffekter, men dette trenger ikke å bety at disse ikke finnes. Feiljusteringsmodellen fungerer best når datasettet har mer enn fem tidsserier. De mest langsiktige effektene disse modellene fanger opp er av studenttallet i forrige femårsperiode.

### 5.2.3 Regionale forskjeller

Virkningene av høyere utdanningsinstitusjoner kan variere mellom ulike regioner. Det kan tenkes at effektene er kraftigere i små og mellomstore regioner, der institusjonene som virksomhet gjerne utgjør en større andel av regionaløkonomien. En annen mulighet kan være at effektene er mindre i små regioner, fordi de ikke klarer å holde like godt på ferdigutdannede studenter og går glipp av en større andel av etterspørselseffekten til andre regioner.

For å teste om det finnes forskjeller mellom regioner, deler jeg inn datasettet i tre, etter regionstørrelse. Regionene defineres som små hvis de har under 30 000 innbyggere, mellomstore hvis de har mellom 30 000 og 100 000, og store dersom de har over 100 000.

Jeg tester to modeller med oppdelt datasett. Tabell 5.8 viser resultatene av en modell med faste regionseffekter, der endring i befolkningsmengde brukes som avhengig variabel. Endring i studenttall og lagget studenttall er interessevariabler. Alle koeffisientene for studenttall blir signifikant positive, bortsett fra endring i studenttall i små regioner. Denne er kun signifikant på 10 prosent signifikansnivå.

Både lagget studenttall og endring i studenttall får størst koeffisient i de store regionene. Her viser modellen at en økning på én student øker befolkningen med 0,619 personer i

samme periode, og 1,096 personer i neste periode. For de minste regionene gir én ekstra student 0,269 personer ekstra i samme periode og 0,267 i neste periode. For de mellomstore regionene er tallene midt i mellom.

Feiljusteringsmodellen kan benyttes til å beregne langtidseffekter i de små regionene, siden befolkningstallet for forrige periode er statistisk signifikant. Fra formel (5.1):

$$-\frac{0,267}{-0,220} = 1,214$$

I følge feiljusteringsmodellen er virkningene i små regioner at én ekstra student øker befolkningen med 1,214 personer på lang sikt. Store og mellomstore regioner hadde ikke signifikante effekter for befolkningstallet i forrige periode, noe som hindrer bruk av feiljusteringsmodellen.

Tabell 5.8: Resultater: Faste effekter med oppdelt datasett. (1) Små regioner, (2) mellomstore regioner, (3) store regioner.

|                    | (1)                  | (2)                 | (3)                 |
|--------------------|----------------------|---------------------|---------------------|
|                    | $\Delta Befolkning$  | $\Delta Befolkning$ | $\Delta Befolkning$ |
| $\Delta Studenter$ | 0.269<br>(1.82)      | 0.437**<br>(3.91)   | 0.619*<br>(2.50)    |
| $Befolkning_{t-1}$ | -0.220***<br>(-6.08) | 0.0114<br>(0.34)    | -0.139<br>(-1.52)   |
| $Studenter_{t-1}$  | 0.267*<br>(2.54)     | 0.301**<br>(2.93)   | 1.096**<br>(4.01)   |
| Dummy periode 3    | -299.1***<br>(-7.31) | -517.5**<br>(-3.54) | 2279.6<br>(2.07)    |
| Dummy periode 4    | -231.2***<br>(-4.72) | -509.4**<br>(-3.22) | 1533.3<br>(0.97)    |
| Dummy periode 5    | -292.4***<br>(-5.45) | -464.0*<br>(-2.22)  | 2338.4<br>(1.15)    |
| Konstantledd       | 3920.6***<br>(6.36)  | 614.3<br>(0.35)     | 20843.9<br>(1.32)   |
| $N$                | 209                  | 110                 | 37                  |

$t$ -verdier i parentes

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Tabell 5.9 viser resultatene for en modell på logaritmeform, der jeg har benytte vanlig sammenkoblet MKM. Her får de små regionene de største og mest signifikante effektene.

Både kort- og langtidseffekter er ikke-signifikante i store og mellomstore regioner. De små regionene får en befolkningsvekst på 2,2 prosent i samme periode dersom studenttallet øker med hundre prosent, og en vekst på 0,7 prosent i neste periode. Koeffisienten for endring i studenttallet er signifikant på én prosent signifikansnivå, og for lagget studenttall på ti prosent signifikansnivå. Feiljusteringsmodellen passer ikke å bruke, siden den laggede avhengige variabelen ikke er signifikant.

Antall observasjoner reduseres i tabell 5.9, særlig i små regioner. Dette er som tidligere nevnt fordi logaritmeform gjør at de observasjonene som har null i studenttall blir undefinert, og det er flest små regioner uten studenter. Regionene med over 100 000 innbyggere har studenter i alle observasjonene, så denne gruppen har like mange observasjoner i begge modellene.

Tabell 5.9: Resultater: Sammenkoblet MKM med oppdelt datasett. (1) Små regioner, (2) mellomstore regioner, (3) store regioner.

|                         | (1)                    | (2)                    | (3)                    |
|-------------------------|------------------------|------------------------|------------------------|
|                         | $\Delta \text{LogBef}$ | $\Delta \text{LogBef}$ | $\Delta \text{LogBef}$ |
| $\text{LogBef}_{t-1}$   | -0.0532<br>(-1.64)     | 0.0271*<br>(2.66)      | -0.0103<br>(-0.57)     |
| $\Delta \text{LogStud}$ | 0.0217**<br>(4.23)     | 0.00257<br>(0.62)      | 0.00873<br>(1.00)      |
| $\text{LogStud}_{t-1}$  | 0.00700<br>(1.80)      | 0.00163<br>(0.52)      | 0.00311<br>(0.58)      |
| Dummy periode 3         | -0.00739<br>(-1.49)    | -0.00895*<br>(-2.59)   | 0.0163*<br>(2.43)      |
| Dummy periode 4         | -0.00156<br>(-0.22)    | -0.00423<br>(-0.89)    | 0.0173<br>(1.84)       |
| Dummy periode 5         | -0.00141<br>(-0.18)    | -0.00531<br>(-0.94)    | 0.0256*<br>(2.43)      |
| Konstantledd            | 0.491<br>(1.60)        | -0.282*<br>(-2.65)     | 0.115<br>(0.61)        |
| $N$                     | 38                     | 89                     | 37                     |

$t$ -verdier i parentes

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Forskjellen mellom de to metodene er at den avhengige variabelen er på log-form i tabell 5.9, og at tabell 5.8 bruker faste regionseffekter mens tabell 5.9 differensierer variablene og bruker sammenkoblet MKM. Dette kan forklare at resultatene i tabell 5.8 viser høyere tall for de større regionene enn de små, mens det er motsatt i neste tabell. De absolutte

endringene kan være større i de store regionene, men ettersom de små regionene har lavere folketall, kan den prosentvise endringen være større og mer signifikant der.

Det mest interessante funnet i dette delkapitlet er kanskje at modell (1) i tabell 5.9 viser signifikante positive effekter av studenttallet i små regioner. Dette står i kontrast til funnene i tabell 5.3, som bruker en dummy som interessevariabel. At begge modellene sammenligner små regioner og bare én finner signifikante effekter, kan tyde på at studenttall er en bedre forklaringsvariabel. Ved å bruke antall studenter, utnyttes variasjon over tid innen regionene på en annen måte enn ved dummyvariabler.

# Kapittel 6

## Oppsummering

Resultatene i studien indikerer at studenttallet i en region har en signifikant positiv effekt på befolkningen i kommende femårsperiode. Modellene med faste regionseffekter, tabell 5.5 og 5.6, får at én ekstra student i regionen øker folketallet med i overkant av én person i perioden etter. Av teoriene presentert i kapittel 2, kan disse resultatene blant annet forklares av tilbuds- og etterspørselseffekter ved utdanningsinstitusjonene som skaper økonomisk vekst i regionen og dermed gir grunnlag for en større befolkning. De fleste analysene som skiller korttids- og langtidseffekter, indikerer at de kortsiktige etterspørselseffektene er de mest signifikante. Jeg peker også på at det kan være vanskeligere å måle langtidseffekter, særlig når datasettet bare har fem tidsserier.

Funnene i tabell 5.5 og 5.6 kan også tyde på at regioner med utdanningsinstitusjoner greier å holde på noen av de tilflyttede studentene når de er ferdig utdannet, slik Winters (2011) antyder. Resultatene i disse tabellene er antagelig blant de mest pålitelige i studien, men må likevel taes med forbehold. Utelatte kontrollvariabler og simultanitet kan gi en viss overestimering av effektene. Samtidig kan skjevhet knyttet til lagget avhengig variabel i forbindelse med mulig seriekorrelasjon gi underestimering.

Analysene som ser på effekten av nye utdanningsinstitusjoner i regionene ved å bruke dummyvariabler, finner ikke spesielt signifikante effekter. Dette samsvarer med Persson og Regnér (2004), som også har dummyvariabler som interessevariabel. Den andre svenske studien jeg presenterte (Andersson et al. 2004), bruker blant annet antall studenter som interessevariabel, der de finner signifikante effekter. Dette samsvarer også med resultatene i denne oppgaven. Det kan tyde på at valg av interessevariabel er avgjørende for om studien finner signifikante effekter. Det kan være nødvendig å måle størrelsen på studiestedet, og ikke kun om regionen har et studiested eller ikke. Mye av variasjonen over tid innen regionene forsvinner når det bare måles hvorvidt en region har høyskole eller ikke.

Det kan også tenkes at resultatene underestimeres i analysene med dummyvariabler, mens det motsatte skjer med antall studenter som interessevariabel. I analysene hvor jeg ser på effekt av nye regionale høyskoler, sammenligner jeg i hovedsak mindre regioner med høyskole mot andre regioner. Dersom plasseringen av regionale høyskoler i mindre regioner er delvis motivert av å skape vekst i regioner som har behov for det, kan det gi skjevhet nedover. I modellene med antall studenter som interessevariabel, inkluderes også de største regionene. Her er trolig overestimering på grunn av simultanitet og utelatt variabelskjevhet tydeligst. De store regionene som har vokst mest i studenttall, ville trolig ha vokst mest i befolkningstall, selv uten studenttalløkningen. Det må likevel legges til at i tabell 5.9 er det i regionene med mindre enn 30 000 innbyggere at studenttallene har den mest signifikante effekten på befolkningsveksten i prosent. Dette er en indikasjon på at endringer i omfanget av høyere utdanning innen regionene er sentralt.

### **Mulige forbedringer**

For å forbedre analysene, er den mest åpenbare løsningen å utvide datamaterialet. Befolkning er en generell og viktig økonomisk variabel, men det kan også være av interesse å studere effekten utdanningsinstitusjonene har på mer spesifikke variabler, som for eksempel bruttoinvesteringer i regionen, arbeiderproduktivitet eller gjennomsnittsinntekt. Dette er variabler som også kunne ha vært nyttige som kontrollvariabler i en befolkningsanalyse. I tillegg kunne en god instrumentvariabel ha økt presisjonen og påliteligheten av estimatene.

Data lengre bakover i tid kunne også ha vært en interessant utvidelse. Mange distriktshøyskoler i Norge ble opprettet mellom 1969 og 1981. En analyse på tvers av små regioner, som i delkapittel 5.1.1, men med befolkningsdata fra for eksempel 1950, ville gitt muligheter for en mer sofistikert analyse. Denne kunne ha sett på endringer i befolkningsutviklingen før og etter opprettelsene av høyskolene, og sammenlignet med befolkningsutviklingen i lignende regioner uten høyskole.



# Referanseliste

Achen, Christopher H. (2000): 'Why Lagged Dependent Variables Can Surpress the Explanatory Power of Other Independent Variables', Paper Presented at the Annual Meeting of the Political Methodology Society, Los Angeles, CA, 20-22 Juli.

Andersson, Roland, John M. Quigley og Mats Wilhelmsson (2004): 'University Decentralization as Regional Policy: The Swedish Experiment', *Journal of Economic Geography*, 4-2004, 371-388.

Anselin, Luc, Atilla Varga og Zoltan Acs (1997): 'Local Geographic Spillovers between University Research and High Technology Innovations', *Journal of Urban Economics*, 42, 422-448.

Aschhoff, B. og T. Schmidt (2008): 'Empirical evidence on the success of R&D cooperation-Happy together?', *Review of Industrial Organization*, 33, 41-62.

Baker, Regina (2007): 'Lagged Dependent Variables and Reality: Did you really specify that autocorrelation 'a priori'?', Paper presented at the 2007 Annual Meeting of the American Political Science Association, 30. August-2. September.

Beck, Nathaniel og Jonathan N. Katz (2011): 'Modelling Dynamicsc in Time-Series-Cross-Section Political Economy Data', *Annual Review of Political Science*, 14, 331-352.

Duranton, Gilles og Diego Puga (2003): 'Micro-foundations of Urban Agglomeration Economies,' NBER Working Paper No. 9931.

Glaeser, Edward L. og Albert Saiz (2004): 'The Rise of the Skilled City', *Brookings-Wharton Papers on Urban Affairs*, 5-2004, 47-94.

Goldstein, Harvey og Joshua Drucker (2006): 'The Economic Development Impacts of Universities on Regions: Do Size and Distance Matter?', *Economic Development Quarterly*, 2006 20: 22.

Florax, Raymond (1992): *The University: A Regional Booster?* Ashgate Publishing Limited, Aldershot, Avebury.

- Jaffe, Adam B. (1989): 'Real Effects of Academic Research', *The American Economic Review*, 79,5, 957-970.
- Krugman, Paul (1991): 'Increasing returns and economic geography', *Journal of Political Economy*, 99: 483-499.
- Lucas, Robert E. (1988): 'On the mechanics of economic development,' *Journal of Monetary Economics* 22, 3-42.
- Mankiw, Gregory N., David Romer, og David N. Weil (1992): 'A Contribution to the Empirics of Economic Growth', *Quarterly Journal of Economics*, May 1992, 407-37.
- Marshall, A (1890/1920): *Principles of Economics*, Macmillan Press Ltd, London.
- Nickell, Stephen (1981): 'Biases in Dynamic Models with Fixed Effects', *Econometrica*, 49, 1417-1426.
- NOU, 2008:3: *Sett under ett*. Kunnskapsdepartementet, Oslo.
- NSDs Kommunedatabase.
- Persson, Helena og Håkan Regnér (2004): 'Universities in the regional economy. Evidence from Swedish employer-employee linked data', Swedish Institute for Social Research (SOFI), Working paper 7/2004.
- Reed, Robert W. (2013): 'A note on the practice of lagging variables to avoid simultaneity', University of Canterbury, New Zealand, Working Papers in Economics, 13/32.
- Romer, Paul (1986): 'Increasing returns and long run growth', *Journal of Political Economy*, 94(5): 1002-1037.
- Romer, Paul (1990): 'Endogenous technological change', *Journal of Political Economy*, 98, 71-102.
- Siegfried, John J., Allen R. Sanderson og Peter McHenry (2007): 'The Economic Impact of Colleges and Universities', Vanderbilt University Department of Economics Working Papers, NO 0612.
- Solow, Robert M. (1956): 'A Contribution to the Theory of Economic Growth', *Quarterly Journal of Economics*, 70, February 1956, 65-94.
- Stambøl, Lasse Sigbjørn (2011): 'Den regionale betydningen av universiteter og høyskoler'. Vedleggsrapport til NOU 2011:3. Kommunal- og regionaldepartementet.
- Statistisk Sentralbyrå (2013): Standard for økonomiske regioner (2002), <http://www4.ssb.no/ClassificationFrames.asp?ID=676393&Language=nb>.

- Sæther, Bjørnar, Jan Mønnesland, Knut Onsager, Kjetil Sørli og Peter Arbo (2000): *Høgskolenes regionale betydning*. NIBR prosjektrapport nr. 2000:9.
- Varga, Attila (2003): 'Agglomeration and the Role of Universities in Regional Economic Development', *Knowledge Transfer, Small and Medium-Sized Enterprises and Regional Development in Hungary*, JATEPRESS, Szeged, 15-31.
- Venables, Anthony J. (2008): 'New economic geography', *The New Palgrave Dictionary of Economics*, Second Edition.
- Venhorst, Viktor, Jouke van Dijk og Leo van Wissen (2010): 'Do the best graduates leave the peripheral areas of the Netherlands?' *Tijdschrift voor Economische en Sociale Geografie*, 101(5), 521-537.
- Vergara, R. (2010), 'Taxation and Private Investment: Evidence for Chile', *Applied Economics*, 42, 717-725.
- Veugelers, Reinhilde og Elena Del Rey (2014): 'The contributions of universities to innovation, (regional) growth and employment', EENEE Analytical Report NO. 18, Prepared for the European Commission.
- Winters, John V. (2011): 'Why are smart cities growing? Who moves and who stays?', *Journal of Regional Science*, 51, 253-270.
- Woolridge, Jeffrey M. (2013): *Introductory Econometrics: A Modern Approach*. (5th Edition) South-Western, CENGAGE Learning.