

Marte Olea Murud
Stella Margretardottir Hårstad

Høyskoleetablering og befolkningsstørrelse

En empirisk studie av høyskoleetableringer i
Norge fra 1950 til 1996

Masteroppgave i Samfunnsøkonomi
Veileder: Hildegunn Ekroll Stokke
Juni 2022

Marte Olea Murud
Stella Margretardottir Hårstad

Høyskoleetablering og befolkningsstørrelse

En empirisk studie av høyskoleetableringer i Norge
fra 1950 til 1996

Masteroppgave i Samfunnsøkonomi
Veileder: Hildegunn Ekroll Stokke
Juni 2022

Norges teknisk-naturvitenskapelige universitet
Fakultet for økonomi
Institutt for samfunnsøkonomi



Kunnskap for en bedre verden

Forord

Denne masteroppgaven er en del av vår mastergrad i samfunnsøkonomi ved Norges teknisk-naturvitenskapelige universitet. Vi ønsker å takke vår veileder Hildegunn Ekroll Stokke for betydningsfull og god veiledning. Hun har gitt utfyllende svar på spørsmål i skriveprosessen. Videre vil vi takke Henrik Furuheim for hjelp med håndtering av datasettet. Til slutt vil vi takke studentbyen Trondheim for mange gode minner fra studietiden!

Trondheim, juni 2022

Marte Olea Murud

Stella Margretardottir Hårstad

Sammendrag

I denne oppgaven undersøkes effekten av høyskoleetablering på befolkningsstørrelse. Oppgaven tar utgangspunkt i høyskoleetableringer i Norge fra 1950 til 1996. Analysen benytter et paneldatasett med event study oppsett for å kunne bruke Difference-in-Differences (DiD) metoden ved ulike etableringsår. Vi finner ingen signifikant effekt av høyskoleetablering i regresjoner hvor hele datasettet benyttes. Videre i analysen utnyttes heterogenitet i regionene. Resultatene viser at det er regioner med initielt antall virksomheter over 11 per 1000 innbyggere som får en signifikant effekt av høyskoleetablering på befolkningsstørrelse. Økningen i befolkning fra pre-reform til post-reform er 2,8 prosentpoeng høyere i treatmentregioner sammenlignet med kontrollregioner. Effekten slår ut 4 år etter etableringen og blir sterkere årene etter. Funnene tyder på at det er regioner med høy næringsaktivitet og gode jobbmuligheter fra før som får effekt av høyskoleetablering på befolkningsstørrelse.

Abstract

This thesis examines the effect of college establishment on population size. The thesis is based on college establishments in Norway from 1950 to 1996. The analysis uses a panel dataset with event study setup to be able to use the Difference-in-Differences (DiD) method for different years of establishment. We find no significant effect of college establishment in regressions where the entire dataset is used. Furthermore, heterogeneity in the regions are utilized. The results shows that there are regions with an initial number of firms over 11 per 1000 inhabitants, which have a significant effect of college establishment on population size. The increase in population from pre-reform to post-reform is 2.8 percentage points higher in treatment regions compared with control regions. The effect occurs 4 years after establishment and becomes stronger in the following years. These results indicate that only regions with high economic activity and good job opportunities from earlier, experience effect of college establishment on population size.

Innhold

1	Introduksjon	1
1.1	Problemstilling	1
1.2	Disposisjon	2
2	Litteratur	3
2.1	Tidligere empiriske studier	3
3	Metode	6
3.1	Difference-in-Differences metoden	6
3.2	Dekomponerer effekten for ulike år	9
3.3	Fixed Effect estimator	11
3.4	Event study	11
3.5	Økonometriske utfordringer	12
3.5.1	Utnytter heterogenitet	12
3.5.2	Forutsetningen om parallelle trender	12
3.5.3	Begrensninger og metodiske utfordringer ved DiD	14
4	Data	16
4.1	Avhengig variabel	17
4.2	Interessevariabel	18
4.3	Kontrollvariabler	18
4.4	Tidsspesifikke effekter	20
4.5	Deskriptiv statistikk	20
5	Analyse	22
5.1	Grunnmodell	22
5.2	Modellutvidelse 1: Årlige effekter	24
5.3	Modellutvidelse 2: Heterogenitet basert på alderssammensetning	27
5.4	Modellutvidelse 3: Heterogenitet basert på næringsaktivitet	29
5.4.1	Årlige effekter ved høy næringsaktivitet	32

6 Konklusjon	35
A Appendix	39
A.1 Oversikt over treatmentregioner og antall virksomheter per 1000 innbyggere	40
A.2 Oversikt over alle regioner og antall virksomheter per 1000 innbyggere	41
A.3 Oversikt over etableringsår for ulike regioner	44

1. Introduksjon

Hensikten med å etablere høyskoler har vært å heve befolkningens kunnskapsnivå. I dagens samfunn virker det å være stadig høyere krav til kunnskap. For å sikre økonomisk vekst og utvikling, er det viktig for Norge å ha en høyt utdannet befolkning.

Fraflytting fra mindre befolkede regioner har vært en utfordring for flere kommuner i Norge. Det har tidligere vært diskutert om det å sette opp en høyskole kan snu befolkningsutviklingen og rekruttere høyt utdannet arbeidskraft til distriktene (Aamodt and Lyby, 2019). En høyskole skaper arbeidsplasser og tiltrekker mange studenter til regionen. Med flere innbyggere kan man tenke seg en positiv effekt på det lokale næringslivet. Dette kan tenkes å gi ytterligere befolkningsøkning.

I denne analysen ser vi nærmere på om høyskoleetablering kan bidra til befolkningsvekst på regionalt nivå. Dette gjøres ved å analysere tall over befolkningsstørrelse og høyskoleetableringer i Norge under perioden 1950-1996. Når en høyskole etableres, kan det få ringvirkninger for aktivitetsnivået på tvers av kommunegrenser. For eksempel kan folk bo i en kommune og jobbe i en annen. Vi har derfor valgt å bruke Statistisk Sentralbyrå (SSB) sin økonomiske regioninndeling fra 2020. Resultatene fra denne analysen kan være av interesse for debatten vedrørende plassering av høyskoler i Norge.

1.1 Problemstilling

I denne oppgaven undersøker vi om etablering av høyskole i en region har effekt på befolkningsstørrelse. Vi har datasett for befolkningsstørrelse og høyskoleetableringer i Norge fra 1950-1996. Videre i analysen undersøkes det om effekten er ulik for regioner med forskjellig antall virksomheter per 1000 innbyggere før høyskoleetableringene. Den konkrete problemstillingen for vår oppgave er: Bidrar høyskoleetablering i Norge under perioden 1950-1996 til regional befolkningsvekst?

1.2 Disposisjon

Masteroppgaven er delt opp i seks kapitler. Kapittel 2 tar for seg tidligere litteratur. I dette kapitlet forklares det teoretiske rammeverket oppgaven baseres på. Vi ser på tidligere studier relatert til vår problemstilling, og hvilke funn disse studiene har gjort. I kapittel 3 gjøres det rede for den empiriske strategien benyttet i oppgaven, og hvilke utfordringer som er knyttet til metoden. Datamaterialet benyttet i analysen blir gjennomgått i kapittel 4. Dette inkluderer avhengig variabel, interessevariabel og kontrollvariabler. Videre i kapittel 5 gjennomføres analysen for å svare på problemstillingen. Det gjøres flere modellutvidelser for å øke troverdigheten til den estimerte effekten. Til slutt blir en konklusjon presentert i kapittel 6.

2. Litteratur

2.1 Tidligere empiriske studier

Det meste av forskning innen utdanningsøkonomi prøver å forklare sammenhengen mellom høyere utdanning og økonomisk vekst. I dette kapittelet vil vi ta for oss en rekke forskningsartikler som ser på sysselsetting, produksjonsnivå og økonomisk vekst i områder med høyskole- eller universitetsetableringer. Denne oppgaven ønsker å studere effekten av høyskoleetableringer i Norge på befolkningsstørrelse i regionen.

Lee (2021) ser på hvilke effekter åpningen av Ulsan National Institute of Science and Technology (UNIST) i Sør-Korea i 2009 har på den lokale industriproduksjonen. Forskningen viser at etablering av universitetet øker antall sysselsatte innen produksjon og varehandel. Dette skjer ved at nye bedrifter etableres i det lokale markedet under perioden 2009-2016. De nye bedriftene er i hovedsak teknologisk nært fagfeltet som universitetet forsker på. Antall bedrifter i teknologisk like industrier øker med mer enn 30 prosent. Tilsvarende økning for andre bedrifter i området, finnes imidlertid ikke. Et annet funn er at den industrielle spesialiseringen fra universitetet sprer seg til regioner som ligger geografisk nær det nye fagfeltet på universitetet. Det er også dokumentert av Kantor and Whalley (2019) og Li et al. (2020) at økonomiske effekter fra universiteter kan spre seg til nærliggende områder.

Nyere modeller viser at det kan være en positiv effekt av lokalisering av bedrifter på den regionale økonomien. Eksempelvis kan det være vanskelig å overføre forskningsfunn til en industri som ligger geografisk langt unna. Effektiv kunnskapsoverføring kan i større grad sikres ved å oppmuntre kommersielle startups til å lokalisere seg nære universitetets forskningssenter. Videre vil dette styrke antakelsen om at universiteter har betydning for den regionale økonomien (Moretti, 2010).

Carneiro et al. (2021) undersøker konsekvensene på arbeidsmarkedet av en eksogen økning i tilbudet av høyere utdannede. Den eksogene økningen av høyere utdannede kommer som følge av etablering av nye høyskoler i Norge på 1970-tallet. Resultatene

viser at lønningene til høyere utdannede øker som følge av etablering av nye høyskoler. Dette tyder på at sammen med en økning av høyere utdannede, øker også etterspørselen etter høyere utdannede på arbeidsmarkedet. Den økte etterspørselen kan gjøre at lønningene presses opp.

Bonander et al. (2016) studerer 2 høyskoler i Sverige som fikk universitetsstatus. Forskingen undersøker hvilken effekt endringen i utdanningsinstitusjonene hadde på økonomien i regionen. De analyserer utviklingen i regionene som gjennomgikk endring, og sammenlignet med ett sett av kontrollregioner ved bruk av syntetisk kontroll metode. Forskingen ser nærmere på den regionale innovasjonen og den økonomiske utviklingen ved å måle antall patenter, antall nye bedrifter og regional BNP. Resultatene viste ingen eller liten effekt på den lokale økonomien. Funnene gir grunn til tvil over effektiviteten av forskningsuniversiteter på regional vekst og utvikling. Det er verdt å nevne at Bonander ser på overgangen fra høyskole til universitet. Dette er annerledes enn overgangen fra å ikke ha høyskole til etablering av ny høyskole.

Bonander påpeker at mangel på effekt ikke betyr at høyere utdanning er ubetydelig for økonomien som en helhet. Migrering mellom regioner kan være grunnen for liten eller ingen regionale økonomiske effekter av universiteter, noe som gjør at fordelene er ikke- regionspesifikke. Pokarzhevskaya and Regnér (2015) finner at kun 30 % av studentene bor i regionen de studerte ved 10 år etter at de startet på universitetet. Dette tyder på at selv om studentkvaliteten øker vil ikke fordelene nødvendigvis tilfalle universitetsregionen.

Knutsen et al. (2020) studerer hvordan individers utdanningslengde og utdanningsvalg påvirkes av høyskolereformen i Norge. Det virker til at endringen skjedde på den intensive marginen ved at befolkningens utdannings sammensetning endret seg, mens utdanningslengden forble uendret. Funnene indikerer at kvinner eksempelvis velger sykepleie eller lærerutdanning, avhengig av hvilke alternativer som finnes i nærområdet. For menn finnes tilsvarende resultat for ingeniør versus teknologisk og økonomiske fag. Funnene gir ingen tegn til at personer med kortere reisevei til et høyere utdanningstilbud enn tidligere tar mer utdanning. De finner imidlertid at folk har høyere tilbøyelighet

til å velge høyskolegrader som tilbys ved de lokale høyskolene. På den måten kan det virke som høyskolereformen endret befolkningens utdannings sammensetning uten at individene nødvendigvis flyttet ut av regionene. Dette kan bety at en eventuell befolkningsvekst ikke nødvendigvis vil komme av at studenter flytter til regionen.

Drucker and Goldstein (2007) undersøker den aktuelle litteraturen innen regional utvikling. Dette inkluderer blant annet case-studier og tverrsnittsstudier. Evidensen er blandet, men de finner noe evidens for at regional utvikling blir forbedret når en høyere utdanningsinstitusjon finnes i området. Drucker (2016) revurderer de regionale økonomiske utviklingseffektene av høyere utdanningsinstitusjoner i USA. Studien modellerer forholdet mellom regional økonomiske resultater og utdanningsinstitusjoner i 2001-2011. Effekten av høyere utdanning varierer etter utfallsmål som produksjon og entreprenørskap, men effekten er mindre innflytelsesrik enn i tidligere studier. De økonomiske ringvirkningene er betydelige opptil 97 kilometer, noe som tyder på innflytelse over store avstander.

Det har blitt funnet evidens om at universitet tiltrekker nye bedrifter, som videre har ført til økt sysselsetting. Det er også funnet noe evidens om at produksjon og entreprenørskap får positiv effekt av at utdanningsinstitusjoner finnes i området. Det er sannsynlig at økt produksjon og sysselsetting vil føre til befolkningsvekst i regioner. Ifølge (Fylkeskommunen-Trøndelag, 2019) er sysselsetting og befolkningsvekst to faktorer som henger tett sammen.

3. Metode

I dette kapitlet forklares metoden som benyttes i analysen. Først blir metoden Difference-in-Differences (DiD) gjennomgått. Deretter ser vi nærmere på måten datasettet er bygget opp med event study oppsett. Videre blir økonometriske utfordringer drøftet. Merk at vi bruker flere engelske økonometriske uttrykk, da vi ikke finner noen god erstatning for disse begrepene på norsk.

3.1 Difference-in-Differences metoden

DiD metoden benytter en kvasi-eksperimentell tilnærming (Wing et al., 2018). Denne metoden skiller seg fra eksperimentell metode ved å ikke benytte randomisering til å velge kontrollgruppe. Til gjengjeld behøves flere forutsetninger om kontrollgruppen enn for eksperimentell metode. For eksempel forutsetter DiD parallelle trender. Denne strategien håndterer i hovedsak utelatte variabel problem/problemer med uobserverbar heterogenitet.

Metoden egner seg godt til vår oppgave. DiD gjør oss i stand til å møte problemstillingen på en god måte, da vi ønsker å finne effekten av en hendelse. Datagrunnlaget kan være enten sammenkoblede tverrsnitt (kan være ulike regioner) eller paneldata (følger samme regioner over tid). En fordel med paneldata er at vi kan følge de samme regionene over tid. Datasettet er utformet på en måte som tillater oss å se på endring i befolkningsstørrelse over tid for norske regioner. Siden vi har data for regioner både med og uten etablering, gjør det oss i stand til å gjennomføre en DiD-analyse hvor utviklingen i disse gruppene sammenlignes (Gertler et al., 2016).

Vi benytter DiD for å estimere den kausale effekten av en hendelse, politisk endring eller behandling på den avhengige variabelen, som beskrevet av Abadie and Cattaneo (2018). I vår oppgave forstås hendelsen som “etablering av høyskole” og den avhengige variabel som befolkningsstørrelse. Vi trenger data om befolkningsstørrelse for regioner som etablerer høyskole og regioner som ikke etablerer høyskole for å kunne se på forskjellene mellom dem. Treatmentgruppen er regioner i Norge som etablerer høyskole

i perioden 1950-1996. Kontrollgruppen er regioner i Norge som ikke har etablert høyskole i perioden 1950-1996.

For å kunne utføre DiD trenger vi data om befolkningsstørrelse før og etter hendelsen, for begge gruppene. Med DiD estimering ser vi på forskjellen i endring før og etter hendelsen for de to gruppene. Dersom det er en forskjell, tilegner vi denne hendelsen; etableringen av høyskole. Man ser på forskjellen i endring, og ikke bare forskjellen i gjennomsnittsverdi etter etableringen. Hvis ikke, ville vi ikke tatt hensyn til at treatmentregioner og kontrollregioner kan ha ulik befolkningsutvikling allerede før hendelsen.

En viktig forutsetning er at de to gruppene har lik utvikling i befolkningsstørrelse dersom hendelsen ikke hadde inntruffet. Det betyr at utviklingen i den avhengige variabelen for begge gruppene, i gjennomsnitt, ville vært det samme i fravær av hendelsen. Utviklingen til kontrollgruppen gir den kontrafaktiske situasjonen. Denne forutsetningen kalles parallelle trender.

Difference-in-Differences kan forstås via to steg. Det første steget er å se på forskjellen innad i hver region før og etter etableringstidspunktet. Dette er "den første forskjellen". Ved å sammenligne regioner med seg selv, kontrollerer "den første forskjellen" for faktorer som er konstante over tid innad i hver region. Dette gjøres for treatmentregioner (likning 3.2) og kontrollregioner (likning 3.3).

Det neste steget er å se på forskjellen i endring for treatment- og kontrollregionene, se (likning 3.4). Hver treatmentregion sammenlignes med hver kontrollregion, men regioner som tilhører samme gruppe blir ikke sammenlignet med hverandre. På den måten får vi "forskjellen i forskjellene", som er opphavet til metodens navn. I det andre steget blir tidsvarierende faktorer fanget opp ved at regionene blir sammenlignet med hverandre. Dette etterlater oss med delta (3.4), som tolkes som DiD estimatoren.

DiD estimator:

Definerer en dummy for treatment:

$D_T = 1$ for regioner som etablerer høyskole og $D_T = 0$ for kontrollgruppen

Definerer en dummy som separerer perioden før og etter:

$D_{AT} = 1$ for post-høyskoleetablering og $D_{AT} = 0$ for pre-høyskoleetablering.

Enkelt modelloppsett:

$$Y_{it} = \beta_0 + \beta_1 D_T + \beta_2 D_{AT} + \delta D_T D_{AT} + U_{it} \quad (3.1)$$

Ser at β_1 fanger opp forskjeller i gjennomsnittsverdier for de to gruppene. β_2 gir effekten av å være i tidsrommet etter etableringen (for både treatment- og kontrollgruppen), mens δ viser om denne effekten er annerledes for treatmentgruppen enn kontrollgruppen. Delta er koeffesienten til interaksjoledet som blir lik 1 hvis regionen etablerer høyskole, og er i perioden etter høyskoleetableringen skjedde. Delta gir gjennomsnittlig treatment-effekt, og kan gi den kausale effekten av etableringen. U_{it} er restleddet som fanger opp alle uobserverbare variabler. Summen av residualene i restleddet er forventet å være null.

Viser hvorfor delta kan tolkes som treatmenteffekten:

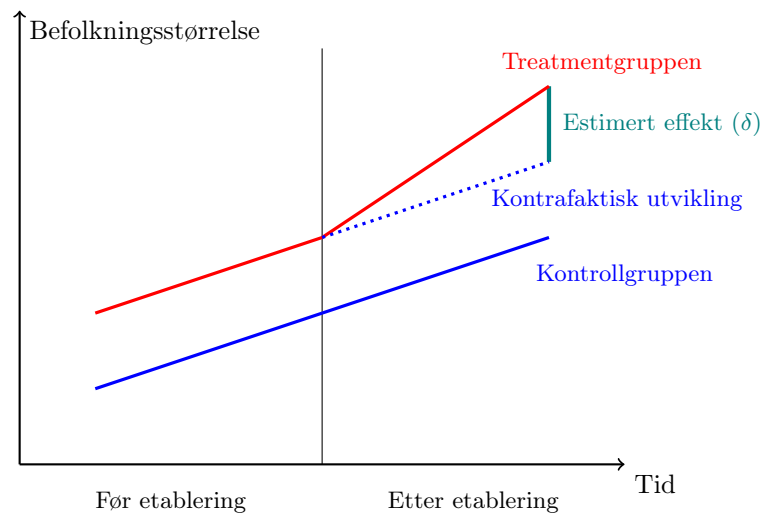
T = treatmentgruppe, C = kontrollgruppe, A = etter treatment, B = før treatment

$$\bar{Y}_{TA} - \bar{Y}_{TB} = \Delta \bar{Y}_T = [\beta_0 + \beta_1 + \beta_2 + \delta] - [\beta_0 + \beta_1] = \underline{\delta + \beta_2} \quad (3.2)$$

$$\bar{Y}_{CA} - \bar{Y}_{CB} = \Delta \bar{Y}_C = [\beta_0 + \beta_2] - [\beta_0] = \underline{\beta_2} \quad (3.3)$$

$$\Delta \bar{Y}_T - \Delta \bar{Y}_C = \delta + \beta_2 - \beta_2 = \delta \quad (3.4)$$

Ser fra (3.4) at delta er lik endringen i treatmentgruppen minus endringen i kontrollgruppen.

Figur 1: Difference-in-differences metoden

Figuren illustrerer identifikasjon ved DiD metoden (Gertler et al., 2016). Den blå linjen gir tidstrenden i kontrollgruppen. Den viser hvordan befolkningsstørrelse har utviklet seg for kontrollregionene. Treatment-dummyen gir forskjellen mellom de to gruppene før etableringen. I figuren er dette avstanden mellom den røde linjen og den blå linjen. Den estimerte effekten (δ) i figuren gir forskjellen i endringer (for de to gruppene) over tid. Den blå prikkete linjen viser utfallet for treatmentgruppen i fravær av treatment (kontrafaktiske situasjonen).

3.2 Dekomponerer effekten for ulike år

Det kan ta tid før en eventuell kausal effekt av høyskoleetableringen slår fullt ut. Individuer og bedrifters adferd krever tid for å endres (King and Behrman, 2009). Derfor kan effekten slå ut senere. Vi kan undersøke effekten av etablering ved å dekomponere den ned til hvert enkeltår etter etableringen. Vi har en post-treatment periode på 7 år hvor vi dekomponerer effekten ned til enkeltår. Med dette kan vi se hvilke år som eventuelt får signifikant effekt. Årene før etableringen er inkludert for å undersøke om forutsetningen om parallelle trender holder. En signifikant effekt i årene før etablering tyder på forskjellig utvikling for treatment- og kontrollgruppen før etablering, og følgelig en dårlig kontrollgruppe. Vi ser nærmere på denne forutsetningen i delkapittel 3.5.2.

Når man ser på gjennomsnittseffekten for hele perioden etter etableringen, er det ikke sikkert den viser noen signifikant sammenheng, men et av årene kan ha en effekt som ligger skjult. Dersom man kun får signifikant effekt i år 7 etter etableringen kan dette være tegn på at vi har for få år med i analysen. Det kan være mulig effekten kommer senere. Samtidig er det viktig å ikke inkludere for mange år etter etableringen da andre faktorer kan være med å påvirke utviklingen. Sannsynligheten for at det kontrafaktiske scenariet er gjeldende etter etablering reduseres for hvert år. Det kan komme flere faktorer inn som påvirker utviklingen i begge regioner. Dette kan svekke antakelsen om parallelle trender. Derfor bør det inkluderes så mange år som er nødvendig for å fange den kausale effekten, men ikke så mange år at den kontrafaktiske situasjonen ikke lenger holder. Dette er grunnen til at vi valgte å stoppe 7 år etter etableringen.

Alternativt modelloppsett: Årlige effekter

$$Y_{it} = \beta_0 + \beta_1 D_T +$$

$$\beta_2 \text{treatment} * \text{periode}_{-2} + \beta_3 \text{treatment} * \text{periode}_{-1} + \beta_4 \text{treatment} * \text{periode}_0 + \beta_5 \text{treatment} * \text{periode}_1 + \beta_6 \text{treatment} * \text{periode}_2 + \beta_7 \text{treatment} * \text{periode}_3 + \beta_8 \text{treatment} * \text{periode}_4 + \beta_9 \text{treatment} * \text{periode}_5 + \beta_{10} \text{treatment} * \text{periode}_6 + \beta_{11} \text{treatment} * \text{periode}_7 + U_{it}$$

(3.5)

I det alternative modelloppsettet inkluderes interaksjonsledd mellom treatment-dummyen og hver enkelt periode for å få de årlige effektene. Perioden -3 (3 år før etablering) er ikke inkludert da dette er referanseåret.

3.3 Fixed Effect estimator

Tidligere i metodekapitlet har vi sett på hvordan DiD fjerner konstante forskjeller mellom treatment- og kontrollgruppen som helhet. For å fange den isolerte effekten av høyskoleetablering er det viktig å kontrollere for ulikhetene mellom regionene i tillegg. Dette gjøres ved å benytte region-reformår Fixed Effect (FE), som tar ut alle permanente forskjeller mellom regioner.

FE estimering trekker fra variablenes gjennomsnittsverdi over tid, dermed utnyttes kun variasjonen i variablene som endrer seg over tid. Eksempelvis vil karakteristikka som natur transformerers bort. Vi får såkalt "within group variation" fordi variasjonen som står igjen etter FE transformeringen, er variasjon innad i hver region som varierer over tid (Wooldridge, 2015).

Dermed vil treatmentvariablen og kontrollvariablene bli utelatt ved FE estimeringen. Årsaken er at disse variablene ikke varierer over tidsperioden i datasettet vårt. Siden dette er et event study, er regionene inndelt etter reformår. Eksempel: Akser/Bærum telles som en annen region i 1958, enn i 1980. Med dette fjernes region-reformår spesifikke effekter.

3.4 Event study

DiD kan være en god løsning alene dersom alle høyskoleetableringene skjer det samme året. Eksempel; alle treatmentregioner får høyskole i 1960. Da kunne vi sammenlignet årene før og etter 1960 for de to gruppene ved å kun benytte DiD. I vårt tilfelle har høyskoleetableringen skjedd i ulike år for de ulike treatmentregionene. For å møte utfordringen med ulike etableringsår utvider vi DiD strategien ved å benytte event study. Vi tar hensyn til at etableringen kan skje i ulike år for regionene.

Hendelsen i vår event study skjer når en region går fra å ikke ha høyskole til å etablere høyskole. Eventet er altså det samme som "hendelsen" ved DiD, men vi åpner for at hendelsen kan skje på ulike tidspunkt.

For å kunne sammenligne forskjellige tidsperioder setter vi en standard lik 0 ved hvert etableringsår med 3 år før og 7 år etter. På denne måten blir det mulig å sammenligne eventet selv om det skjedde på forskjellige tidspunkter. Vi har laget 24 deldatasett som følger befolkningstallene over perioden -3 til 7 i hver enkelt treatmentregion med alle kontrollregioner. Når dette samles til et felles datasett, kan vi bruke event study metoden til å sammenligne alle treatmentregioner mot alle kontrollregioner. Når vi bruker DiD estimering på event study oppsettet, vil trenden for gruppene i periode -3,-2,-1 sammenlignes med trenden for de to gruppene i periode 0 til 7.

I tilfeller hvor det er vanskelig å identifisere tidspunktet eventet fant sted kan det være begrensninger eller mindre nyttig å bruke event study metoden (Chen, 2017). Siden vi har godt dokumentert dekning for etableringsårene, vil ikke dette gi betydelige utfordringer for vår oppgave.

3.5 Økonometriske utfordringer

3.5.1 Utnytter heterogenitet

Når vi inkluderer kontrollvariabler, blir ulikheter mellom regionene innad i treatment- og kontrollgruppen tatt hensyn til. Ved FE estimering vil kontrollvariablene bli overflødige, men de er fortsatt nyttige til å utnytte heterogenitet. Med heterogenitet kan vi se ulik effekt på befolkningsstørrelse for forskjellige delutvalg i datasettet (Gertler et al., 2016). Med kontrollvariablene har vi mulighet til å betrakte delutvalg av datasettet hvor regioner med visse kjennetegn blir utelatt. På den måten kan vi se om resultatene blir signifikante ved ulike avgrensninger i datasettet. Eksempelvis kan vi teste regresjoner med datasett som kun inneholder regioner med et høyt antall virksomheter per 1000 innbyggere.

3.5.2 Forutsetningen om parallelle trender

En forutsetning for å finne den kausale effekten med DiD estimering er at antakelsen om parallelle trender holder. For å sikre dette må kontroll- og treatmentgruppen

ha lik utvikling i avhengig variabel før hendelsen. I fravær av hendelsen vil begge grupper ha samme utvikling i avhengig variabel under forutsetning om parallelle trender (Gertler et al., 2016). Dette innebærer at vi trenger kontrollregioner som følger samme trend som treatmentgruppen. Det er et problem hvis de regionene som etablerer høyskole har en annen utvikling i befolkning enn kontrollgruppen i årene -3, -2, og -1.

Høyskolenes lokalisering ble valgt for å forbedre utdanning til alle landets innbyggere. Stortinget uttaler i St.prp.nr.136 (1969) at lokaliseringen ikke er valgt på grunnlag av favoriserte regioner. Beliggenheten ble valgt med hensikt å tilby innbyggerne i alle landets 19 fylker muligheten til å ta en grad innen sykepleie-, ingeniør-, lærer- og økonomi/administrasjon utdanning. De områdene i Norge som fikk høyskole ble ikke nødvendigvis valgt på grunnlag av lokal etterspørsel eller lokale ressurser. I St.prp.nr.136 (1969) ser vi at lokaliseringen av høyskolene er et resultat av et kompleks politisk samspill med hensikt å forbedre tilgangen til utdanning for landets innbyggere. På 1950-tallet var utdanningsnivået i Norge lavt, men utover 1960- og 1970-tallet endret dette seg, og Norge fikk en sterk vekst av høyere utdanningstilbud (Knutsen et al., 2020).

Ettersom treatmentregionene ligger spredt rundt om i landet, har vi valgt å inkludere alle regioner i Norge som ikke fikk høyskole under denne tidsperioden som kontrollregioner. Dette har vi valgt for å få like stor spredning og variasjon av regioner i landet i kontrollgruppen, slik som det er i treatmentgruppen.

Forutsetningen om parallelle trender kan testes ved å sjekke om interaksjonsledd mellom treatment og årene før etablering er signifikante. I likhet med (Gertler et al., 2016) kan vi utføre en "placebo-test" for å se om forutsetningen holder. Vi ser nærmere på om det er en treatment-effekt før hendelsen inntraff (falsk treatment-effekt). Dersom vi finner signifikant effekt av placebo-hendelsen, gir det direkte evidens imot parallelle trender. En signifikant effekt av høyskoleetablering på befolkningsstørrelse før etableringen fant sted, vil bety at det er andre faktorer som spiller inn. Det betyr at treatment- og kontrollregionene har ulik utvikling i befolkningsstørrelse allerede før etableringen skjer. Hvis det er tilfelle, har man sannsynligvis utelatt relevante

variabler, og estimatet på delta gir trolig ikke den kausale effekten. Hvis placebo regresjonen ikke finner noen signifikante effekter, vil det styrke troverdigheten om parallelle trender.

En annen mulighet er å benytte grupper som ikke risikerer å få treatment (Gertler et al., 2016). Med vår problemstilling er det imidlertid vanskelig å finne en region-gruppe som ikke er i stand til å motta treatment. Vi kunne muligens benyttet regioner i Sverige, men etableringer av høyskoler skjedde også i Sverige. Alternativt kunne vi sett på fattigere land som ikke etablerte høyskole i perioden, men regionene i disse landene vil trolig ha ulik utvikling i befolkningsstørrelse. Kontrastene mellom landene er såpass store at betydningen av utdanning kan være forskjellig. Selv med relativ lik befolkningsutvikling ville kontrastene vært så store at regioner i Afrika ikke ville gitt den kontrafaktiske situasjonen, og dermed vært dårlig egnet kontrollgruppe. Vi ser at denne fremgangsmåten ikke er hensiktsmessig i vår oppgave.

3.5.3 Begrensninger og metodiske utfordringer ved DiD

En begrensning ved DiD estimatoren er at det kan forekomme uoppdaget bias. Selv når befolkningstrenden i treatment- og kontrollgruppen er lik før etableringsåret, kan det forekomme bias i estimeringen. Skjevheten kan gå uoppdaget da DiD tilskriver hendelsen eventuelle forskjeller i trender mellom gruppene som oppstår fra tidspunktet hendelsen starter. Hvis det er andre faktorer tilstede som påvirker forskjellen i befolkningstrendene, og de ikke tas med i regresjonen, vil estimatet være ugyldig eller partisk (Gertler et al., 2016). Videre ser vi på metodiske utfordringer ved DiD estimatoren.

Imperfekt samsvar

Imperfekt samsvar oppstår når enheter tildelt treatmentgruppen ikke får "treatment", mens enheter tildelt kontrollgruppen faktisk får "treatment" (Gertler et al., 2016). I denne oppgaven vil imperfekt samsvar inntreffe hvis en eller flere regioner i treatmentgruppen ikke etablerer høyskole, mens en eller flere regioner som faktisk etablerer høyskole, blir oppført i kontrollgruppen. Våre data om etableringstidspunkt er basert på norsk utdanningshistorie (Knutsen et al., 2020). Vi betrakter kildegrunnlaget som

troverdig, og kan dermed argumentere for at det ikke er store problemer tilknyttet imperfekt samsvar i vår oppgave.

Attrition bias

Attrition bias skjer når deler av datasettet forsvinner over tid, og forskere ikke er i stand til å finne alle initielle data av treatment- og kontrollgruppen (Gertler et al., 2016). Vi har godt dokumenterte befolkningstall fra Statistisk Sentralbyrå (SSB) og Norsk Samfunnsvitenskapelig Datatjeneste (NSD) for alle regioner i tidsperioden (med unntak av Ofoten). Dette er troverdige og anerkjente kilder. Vi ser ingen tegn til systematiske mangler i befolkningstall for regionene i vårt datasett. Dette tyder på at attrition bias ikke er et betydelig problem i vår oppgave.

Spillover

Spillover skjer når en hendelse påvirker kontrollregioner enten positiv eller negativt (Gertler et al., 2016). I dette tilfellet vil ikke kontrollgruppen fungere som den kontrafaktiske situasjonen da den også blir påvirket av hendelsen. I denne oppgaven er spillover effekter høyst sannsynlig, da høyskoleetablering trolig vil ha effekt på befolkningsstørrelse i nabokommunene på grunn av pendling. For å angripe dette problemet har vi valgt å benytte SSB sin økonomiske regioninndeling som tar hensyn til dette.

Regionene kan imidlertid ikke krysse fylkesgrenser, noe Bhuller (2009) kritiserer. For å møte denne utfordringen kommer Bhuller med en alternativ inndeling med 46 regioner, hvor regioner også kan krysse fylkesgrenser. På den måten kan potensielle spillover effekter på tvers av fylkesgrenser også bli fanget opp. I vår oppgave har vi ikke gått nærmere inn på den alternative regioninndelingen for å begrense omfanget av denne oppgaven. For videre arbeid er dette en tilnærming som kan undersøkes nærmere.

4. Data

Datamaterialet om befolkning i Norge fra 1950-1996 er hentet fra Statistisk sentralbyrå (SSB). Tallene for befolkningsstørrelse gir befolkningsnivået i kommunene 1.januar. Treatment-dummyen er kodet etter norsk utdanningshistorie (Knutsen et al., 2020). Data om kontrollvariablene "andel personer i yrkesaktiv alder" og "antall personer med forretningsvirksomhet per 1000 innbygger" er hentet fra Norsk Samfunnsvitenskapelig Datatjeneste (NSD). Datamaterialet fra NSD er gjort tilgjengelig med spesiell tillatelse.

Data om befolkningsstørrelse er hentet på kommunenivå og ved hjelp av programmering aggregert til regionnivå. Det samme gjelder kontrollvariablene "yrkesaktiv alder" og "forretningsvirksomhet". Med regioner menes økonomiske regioner i henhold til SSB sin standard (Hustoft et al., 1999). Det er i dag (siden 2020) 85 økonomiske regioner i Norge. Regionene er inndelt på en måte som tilfredsstillende EU sine NUTS-4 krav. En av fordelene ved å benytte data på regionnivå sammenlignet med kommunenivå er at vi åpner opp for at den økonomiske aktiviteten kan krysse kommunegrensene. Et eksempel er arbeidskraft. Det er mange som bor i en annen kommune enn de er ansatt i (Berge, 2020). For å avgrense økonomiske regioner på en hensiktsmessig måte benytter SSB pendlingstall for innbyggerne. På den måten kan vi se på lokale økonomier, også på tvers av kommunegrensene.

Håndteringen av datamaterialet følger et event study oppsett. Vi konstruerer totalt 24 separate del-datasett med ett for hver region som etablerer høyskole med tilhørende kontrollregioner. Opprinnelig var det 25 høyskoleetableringer i tidsperioden vi studerer, men grunnet manglende tall har vi ikke kunnet inkludere Ofoten. Se appendix tabell A.3 for en oversikt over etableringsår for de ulike regionene. For å følge treatment- og kontrollregioner 3 år før og 7 år etter etableringene benyttes data fra 1952 til 1996.

For å separere del-datasettene benyttes en variabel kalt reformår. Reformår er lik etableringsåret for alle regioner i perioden -3 til 7 tilknyttet det spesifikke reformåret. Eksempelvis kan vi se på Arendal som etablerer høyskole i 1967. I dette tilfellet vil alle regioner tilknyttet denne etableringen, treatment og kontrollregioner, få

reformår-variabel lik 1967. Det er verdt å understreke at de ulike kontrollregionene er med i alle del-datasett. Det innebærer at de samme kontrollregionene er med ved alle etableringer, men med ulikt b-regionnummer. Hver region har sitt unike b-regionnummer i hvert deldatasett, som gjentas for de 11 periodene i et del-datasett. Et deldatasett vil bestå av 61 b-regioner, hvorav 1 av disse er en treatmentregion og resten er kontrollregioner. På den måten skiller datasettet mellom regionene for ulike tidsperioder knyttet til ulike etableringer. Eksempelvis blir Asker/Bærum sett på som en annen kontrollregion i 1960 enn i 1980.

4.1 Avhenging variabel

Den avhengige variabelen i vår oppgave er befolkningstørrelse. En utfordring med denne variabelen er forandringer i kommune- og regiongrenser over tid. For å møte utfordringen benytter vi 2020 regioner som konstante geografiske områder. I statistikkbanken til SSB er befolkningstall oppgitt i tre forskjellige kommuneinndelinger. Disse er kommuneinndelingen i 1951, 2018/2019 og 2020. Årsaken til at vi ikke valgte å hente ut befolkningstall fra 1951 og 2018/2019 er at mange kommuner har befolkningsstørrelse "0". Blant kommuner med befolkningsstørrelse "0" finner vi Kristiansund og Trondheim i tillegg til flere småkommuner. Bakgrunnen for manglende data om befolkningsstørrelsen er knyttet til endringer i kommunenummer. Eks: kommunenummeret til Trondheim er ikke det samme i 2018 som i 2020 ¹. Til forskjell har befolkningstallene med kommuneinndelingen fra 2020 sammenslåtte tidsserier. Dette vil si at tall fra eldre kommunenummer blir slått sammen med nye kommunenummer. SSB tar ikke hensyn til faste geografiske områder slik vi gjør. Følgelig blir det store hopp i befolkningstallene fra et år til det neste ved kommuneendringer.

Vi har kontrollert for kommunesammenslåinger- og splittelser ved å legge inn tall fra NSD. Med befolkningstall fra NSD dekker vi hoppene som skyldes kommuneendringer i perioden. Hos NSD er det mulig å hente ut befolkningstall fra tidligere kommuner som ikke lenger eksisterer. Med oversikt fra Regjeringen (2020) over

¹Kontaktet SSB vedrørende manglende data i 2018/2019 versjonen via epost. Ifølge SSB skyldes dette endringer i kommunenummer over tid.

kommuneendringer slår vi sammen tidligere kommuner manuelt, og legger til disse i vårt eksisterende datasett. Vi kan dermed tilpasse det opprinnelige datasettet fra SSB til vår oppgave som benytter faste geografiske områder.

For å kunne sammenligne endringer i befolkningsstørrelse på tvers av regioner må endringen bli oppgitt som prosent. Eksempelvis vil ikke en økning på 1000 mennesker i Oslo tilsvare det samme i Elverum, da de har forskjellig størrelse i utgangspunktet. Ved en logaritmisk transformasjon av den avhengige variabelen vil vi få endringer i befolkningsstørrelse oppgitt i prosent (Benoit, 2011).

4.2 Interessevariabel

Interessevariabelen i vår oppgave er etablering av høyskoler i Norge fra 1950-1996. For å fange interessevariabelen benytter vi tre dummyer, hvor den siste er et interaksjonsledd. Først har vi en "treatment" dummy som blir lik 1 hvis regionen har etablert en høyskole. Av tilsammen 84 regioner er 24 av disse treatmentregioner og 60 er kontrollregioner. Vi har så en "post reform" dummy som blir lik 1 dersom vi er i samme år eller i årene etter etableringen fant sted. Til slutt har vi et interaksjonsledd med disse to dummyene. Interaksjonsleddet blir lik 1 hvis vi er i en region med høyskoleetablering, og i perioden under eller etter etableringen.

Interessevariabelen er koeffisienten foran dette interaksjonsleddet, se δ i likning (3.1). Koeffisienten måler forskjellen i gjennomsnittlig befolkningsstørrelse fra perioden før til perioden etter etableringen, for treatmentregioner og kontrollregioner. I analysekapitlet vil vi se nærmere på om koeffisienten til interkasjonsleddet er ulik null. Det vil si om høyskoleetablering har en effekt på befolkningsstørrelse, og hvorvidt denne effekten er signifikant.

4.3 Kontrollvariabler

Vi har definert to kontrollvariabler i datasettet. "Andel yrkesaktiv alder" er definert som regionalt antall personer fra 20-64 år i 1950 relativt til regional befolkning i

1950. "Antall virksomheter per 1000 innbyggere" er definert som regionalt antall personer med selvstendig virksomhet i 1950 relativt til regional befolkning i 1950. Variablene er delt på regional befolkning i 1950 for å fjerne skalaeffekter, slik at kontrollvariablene kan sammenlignes på tvers av regioner med ulik befolkningsstørrelse. Datasettet til de to kontrollvariablene er hentet fra NSD. Vi har hentet ut tall fra 1950 med kommuneinndelingen fra 2019. Ved å bruke regjeringens oversikt over kommunesammenslåinger i 2020, har vi kunnet aggregere dataen om til SSB sin økonomiske regioninndeling fra 2020.

Begge kontrollvariabler kan bli påvirket av befolkningsvekst. For å unngå endogenitetsproblem bruker vi tall fra 1950 som initiell situasjon, før noen av regionene etablerer høyskole. Disse tallene holdes konstant over tid. Dummy for treatment tar ut alle permanente forskjeller mellom treatment- og kontrollregionene, mens initialdata på næringsstruktur og alder kontrollerer for forskjeller mellom regioner innad i treatment- og kontrollgruppen.

Regioner med en høyt antall virksomheter før etableringen, vil trolig ha større effekt av høyskoleetableringen enn regioner med relativt lite næringsliv. Høy andel i yrkesaktiv alder vil sannsynligvis bety flere folk i arbeid. Vi tenker det er mer lokkende for unge å flytte til en region med et yngre miljø og et mer attraktivt arbeidsmarked, enn til regioner med en høy andel pensjonister. I så fall vil en høyskoleetablering få større effekt i regioner hvor andel i yrkesaktive alder er høyere. Kontrollvariablene kan videre brukes til å separere datasettet inn til delgrupper av regioner. På den måten kan vi utnytte heterogenitet for aldersammensetning og næringsstruktur innad i treatment- og kontrollgruppen. I en robusthetsanalyse kan vi droppe regioner med spesifikke kjennetegn, og se om effekten på befolkningsstørrelse blir den samme.

Det kan tenkes at regioner med et høyt antall virksomheter og/eller høy andel yrkesaktive har høyere befolkningsnivå. For å undersøke dette ser vi etter om det er en trend mellom de 20 høyest befolkede regionene i 1950 og verdiene i kontrollvariablene. Variabelen andel yrkesaktiv alder viser ingen tydelige tegn til trend mellom høyere andel yrkesaktive og høyt befolkningsnivå. Ved antall virksomheter per 1000 innbygger

og befolkningsstørrelse er det en svak trend. Flere av de 20 mest befolkede regionene har høyt antall virksomheter, men det er fortsatt flere av disse som også har et lavt antall. Av de 20 høyest befolkede regionene har 8 av de en verdi på antall virksomheter per 1000 innbyggere under eller rett over gjennomsnittet. Følgelig vil vi ikke trekke noen konklusjon om at denne kontrollvariabelen har en korrelasjon med befolkningsstørrelse.

4.4 Tidsspesifikke effekter

For å fange opp effekten av makroøkonomiske trender benytter vi tidsspesifikke årsdummyer. Vi har en periode-variabel med verdier fra -3 til 7, som indikerer 3 år før til 7 år etter etableringen. I tillegg har vi variabelen reformår, som har verdi lik etableringsåret for alle observasjoner i perioden -3 til 7 tilknyttet det spesifikke reformåret. Ved å genere et interaksjonsledd mellom periode og reformår får vi en dummy for hver enkelt år i datasettet. Periode -3 (3 år før etablering) er brukt som referanseår, så effektene for de ulike årene er sammenlignet ut fra 3 år før etableringene. Dette gjør at vi kan kontrollere for effekter som er like for alle regioner i det spesifikke året. På denne måten kan tidsspesifikke årsdummy kontrollere for makroøkonomiske effekter.

4.5 Deskriptiv statistikk

Tabell 4.5: Deskriptiv statistikk

	Observasjoner	Gjennomsnitt	Std.av.	Min	Maks
Befolkning	10164	46566.83	65223.03	5226	488659
Andel i yrkesaktivalder	10164	0.5998961	0.054961	0.504	0.765
Antall virk. per 1000 innb.	10164	7.706795	2.937795	2.805635	17.43465
<i>N</i>	10164				

Tabell 4.5 viser at det er 10164 observasjoner for hver variabel. Etersom 9 av de 24 etableringene er overlappende, er det totalt 15 unike etableringsår. Det

blir feil å inkludere samme sett med kontrollregioner flere ganger for etableringer som skjedde ved samme år. Ved å multiplisere 60 b-regioner (kontrollregionene) med 15 etableringsår gir dette 900 observasjoner. Ved å legge til 24 b-regioner, som er treatmentregionene, får vi 924 observasjoner. Siden hver enkelt b-region multipliseres med 11 perioder gir dette 10164 observasjoner.

Befolkningsstørrelsen har et regionalt gjennomsnitt på 46567 personer. Det er stor variasjon i denne variabelen. Minste verdi finner vi i Midt-Gudbrandsdalen i 1971, hvor befolkningen er 5226 personer. Høyeste verdi finnes i Oslo i 1996 med et befolkningstall på 488659.

Andel yrkesaktiv alder har en gjennomsnittsverdi på 60 prosent. Regionen med laveste andel er Vest-Finnmark med 50 prosent, og regionen med høyest andel er Fosen med 76,5 prosent. Antall virksomheter per 1000 innbyggere har en gjennomsnittsverdi på 7,70. Regionen med lavest antall virksomheter per 1000 innbygger er Trondheim Forstad med omtrent 2,8, og regionen med høyest antall virksomheter per 1000 innbyggere er Oslo med omtrent 17,43.

5. Analyse

I dette kapitlet benyttes ulike modellspesifikasjoner for å undersøke effekten av høyskoleetablering på befolkningsstørrelse. Vi ser nærmere på etableringer for norske regioner i perioden 1950-1996.

5.1 Grunnmodell

Grunnmodellen består av tre ulike modellspesifikasjoner.

I den første kolonnen har vi kjørt DiD regresjon uten kontrollvariabler og uten FE. Dette er den enkleste regresjonen i hele analysen. Ser på effekten for henholdsvis kontroll- og treatmentregioner.

Kontrollregioner

$$\frac{\partial \ln(\text{befolkning})}{\partial \text{postreform}} = 0,054$$

Når dummy-variabelen postreform går fra 0 til 1 øker befolkningen med 5,4 prosent, men effekten er ikke signifikant. Det betyr at økningen i befolkning fra pre-reform til post-reform er 5,4 prosent i kontrollregionene.

Treatmentregioner

For å finne befolkningsveksten legger vi til den ytterligere effekten av å være en treatmentregion i perioden etter etableringen.

$$\frac{\partial \ln(\text{befolkning})}{\partial \text{postreform}} = 0,054 + 0,01 = 0,064$$

Økningen i befolkningen fra pre-reform til post-reform er på 6,4 prosent i treatment-regionene. Økningen i befolkning fra pre-reform til post-reform er 1 prosentpoeng høyere i treatmentregioner sammenlignet med kontrollregioner. Dette gir da effekten

av høyskoleetablering, men denne effekten er imidlertid ikke signifikant.

Videre i kolonne (2) inkluderes kontrollvariablene yrkesaktiv alder og antall virksomheter. Ser at kontrollvariablene er signifikante. Interessevariabelen har samme verdi som i kolonne (1), men denne effekten er fortsatt ikke signifikant.

Yrkesaktiv alder

$$\frac{\partial \ln(\text{befolkning})}{\partial \text{alder}} = 2,585$$

Dersom alder øker med en enhet, øker $\ln(\text{befolkning})$ med 2,585 enheter. Siden yrkesaktiv alder har verdier mellom 0 og 1, er en økning på en enhet mye (100 prosentpoeng). Ser derfor heller på en økning i yrkesaktiv alder på 0,01 enheter (= 1 prosentpoeng). Dette gir en økning i $\ln(\text{befolkning})$ på 0,02585 enheter, som betyr at befolkningen øker med 2,585 prosent. Med andre ord; en økning i andel yrkesaktiv alder med 1 prosentpoeng øker befolkningsstørrelse med 2,585 prosent.

Antall virksomheter per 1000 innbyggere

$$\frac{\partial \ln(\text{befolkning})}{\partial \text{virksomheter}} = 0,139$$

Dersom antall virksomheter per 1000 innbyggere øker med 1, øker befolkningen i regionen med 13,9 prosent.

I kolonne (3) benyttes FE på DiD regresjonen for å kontrollere for reformår-spesifikke regionfaste effekter. Kontrollvariablene inkluderes ikke da de uansett blir transformert bort med FE-estimeringen. Med FE vil all variasjon innad i regionen som ikke varierer over tid, bli transformert bort. Dette er også grunnen til at treatment variabelen blir utelatt. Post reform variabelen har økt og er signifikant, men ser at interessevariabelen forblir lik.

Interessevariabelen har samme verdi i de tre kolonnene i grunnmodellen, men ingen av

spesifikasjonene gir signifikant resultat. Dette tyder på at ingen av fremgangsmåtene gir den kausale effekten av interessevariabelen på befolkningsutviklingen. Følgelig kan ikke resultatet benyttes til å trekke noen konklusjoner. Kolonne (3) har lavest standardavvik, og en t verdi på $0,01/0,00771 = 1,30$. Ettersom FE kontrollerer for reformår-spesifikke regionfaste effekter vet vi at denne spesifikasjonen er mer til å stole på enn DiD uten FE. Ved en DiD analyse bør regionene være så like som mulig for å finne den isolerte effekten av treatment. Jo flere ulikheter som blir kontrollert for, jo mer presis estimering. På bakgrunn av dette vil vi benytte FE estimering i videre modellutvidelser.

Tabell 5.1: Grunnmodell

	DiD	DiD med kontrollvariabler	DiD med FE
	(1)	(2)	(3)
Avhengig variabel	ln_befolkning	ln_befolkning	ln_befolkning
Treatment	0.175* (0.0917)	0.0433 (0.0737)	
Postreform	0.0540 (0.139)	0.0540 (0.111)	0.0666*** (0.00994)
TreatmentxPostreform	0.0100 (0.108)	0.0100 (0.0863)	0.0100 (0.00771)
Yrkesaktiv alder		2.585*** (0.112)	
Antall virksomheter		0.139*** (0.00211)	
PeriodexReformår	Ja	Ja	Ja
FE	Nei	Nei	Ja
N	10164	10164	10164

Standardavvik i parentes

* $p < .10$, ** $p < .05$, *** $p < .01$

5.2 Modellutvidelse 1: Årlige effekter

Hensikten med modellutvidelse 1 er å dekomponere effekten år for år, for å undersøke om enkeltår har signifikant effekt. Selv om gjennomsnittseffekten ikke gir signifikant estimat kan det fortsatt være en effekt for enkeltår. Kolonne (1) gir gjennomsnittseffekten vi fant i kolonne (3) fra grunnmodellen. I kolonne (2) benyttes det alternative

modelloppsettet for å dekomponere effekten til enkeltår, hvor periode -3 fungerer som referanseår. Igjen har vi brukt FE estimering. De årlige effektene i Tabell 5.2 tolkes som endring i befolkning fra år -3 til aktuelt år, for treatment- versus kontrollregioner. I tabellen er endring i befolkning fra år -3 til -2 og fra -3 til -1 er ikke signifikant forskjellig for treatment- og kontrollregioner.

I et scenario med ikke-signifikante årlige effekter før treatment og signifikante effekter etter treatment, gir det tydelig signaler om at effekten på befolkningsstørrelse skyldes høyskoleetableringen. Et slikt resultat ville styrket antakelsen om parallelle trender. I denne modellutvidelsen ser vi at ingen av årene før eller etter etablering får signifikant effekt. Det at årene før periode 0 ikke er signifikante, styrker antakelsen om parallelle trender. Samtidig estimeres ingen signifikant effekt i periode 0 eller noen av årene etter etableringen.

Tabell 5.2: Årlige effekter

Avhengig variabel	Gjennomsnittseffekt	År for år
	(1)	(2)
	ln_befolkning	ln_befolkning
Postreform	0.0666*** (0.00994)	
TreatmentxPostreform	0.0100 (0.00771)	
TreatxPer-2		-0.00137 (0.0161)
TreatxPer-1		-0.000357 (0.0161)
TreatxPer0		-0.000129 (0.0161)
TreatxPer1		0.00394 (0.0161)
TreatxPer2		0.00626 (0.0161)
TreatxPer3		0.00957 (0.0161)
TreatxPer4		0.0108 (0.0161)
TreatxPer5		0.0130 (0.0161)
TreatxPer6		0.0132 (0.0161)
TreatxPer7		0.0191 (0.0161)
PeriodexReformår	Ja	Ja
FE	Ja	Ja
<i>N</i>	10164	10164

Standardavvik i parentes

* $p < .10$, ** $p < .05$, *** $p < .01$

5.3 Modellutvidelse 2: Heterogenitet basert på alderssammensetning

Vi utforsker datasettet ved å avgrense for regioner under ulike forutsetninger relatert til kontrollvariabelen andel yrkesaktiv alder. Vi vil utnytte eventuell heterogenitet og se om effekten kan være forskjellig for ulike grupperinger av regionene. Avgrensningene forteller noe om alderssammensetningen i regionene, og kan ha påvirkning på befolkningsutviklingen. Når avgrensningene i datasettet endres, vil også antall observasjoner endres. Dette medfører at vi har ulikt antall observasjoner i de 5 modellspesifikasjonene. Ved å ta totalt antall observasjoner i datasettet (10164) fratrukket droppede observasjoner¹ får vi antall observasjoner (N) i de ulike spesifikasjonene i Tabell 5.3.

Vi estimerer 5 ulike modellspesifikasjoner for å se nærmere på heterogenitet basert på alderssammensetning. I kolonne (1) og (2) droppes treatmentregioner med andel yrkesaktiv alder under henholdsvis 55 prosent og 60 prosent. Videre i kolonne (3) og (4) gjelder den samme avgrensningen for alle observasjoner. I kolonne (4) fjernes dermed alle regioner med andel yrkesaktiv alder lavere enn gjennomsnittet på 60 prosent. Ingen av kolonnene (1)-(4) gir signifikant effekt av interessevariabelen. T-verdien er imidlertid høyest i kolonne (1), og vi undersøker derfor denne spesifikasjonen ytterligere. I likhet med kolonne (1) betrakter kolonne (5) treatmentregioner med andel yrkesaktiv over 55 prosent. Dette gjøres for å se om det ligger en skjult effekt i enkelte år. Fra kolonne (5) ser vi at ingen enkelte år gir signifikant effekt. Oppsummert ser vi ingen signifikant effekt av interessevariabelen eller årlige effekter i modellutvidelse 2. Følgelig går vi ikke videre med denne variabelen.

¹Viser analytisk tilnærming for endring i antall observasjoner i modellutvidelse 2. I kolonne (1) fjernes 3 treatmentregioner, som utgjør 33 observasjoner (3 treatmentregioner * 11 perioder). I kolonne (2) fjernes 16 treatmentregioner, som utgjør 176 observasjoner (16 treatmentregioner * 11 perioder). I kolonne (3) fjernes totalt 11 regioner, hvor 3 av disse er treatmentregioner og 8 av de er kontrollregioner. Dette utgjør henholdsvis 33 (3 treatmentregioner * 11 perioder) og 1320 (8 b-regioner * 11 perioder * 15 etableringsår) observasjoner. I kolonne (4) fjernes totalt 53 regioner, hvor 16 av disse er treatmentregioner og 37 er kontrollregioner. Dette utgjør henholdsvis 176 (16 treatmentregioner * 11 perioder) og 6105 (37 b-regioner * 11 perioder * 15 etableringsår) observasjoner. Videre, i kolonne (5) gjelder samme regnestykke som i kolonne (1).

Tabell 5.3: Heterogenitet basert på alderssammensetning

Avhengig variabel	Kun treatment Alder>0.55	Kun treatment Alder>0.6	Alle Alder>0.55	Alle Alder>0.6	Kun treatment Alder>0.55
	(1)	(2)	(3)	(4)	(5)
	ln_befolkning	ln_befolkning	ln_befolkning	ln_befolkning	ln_befolkning
Postreform	0.0666*** (0.00995)	0.0666*** (0.0100)	0.0741*** (0.0113)	0.0935*** (0.0124)	
TreatmxPostref	0.00978 (0.00823)	0.0106 (0.0134)	0.00739 (0.00869)	-0.00325 (0.0106)	
TreatxPer-2					-0.00248 (0.0172)
TreatxPer-1					-0.00178 (0.0172)
TreatxPer0					-0.000816 (0.0172)
TreatxPer1					0.00380 (0.0172)
TreatxPer2					0.00617 (0.0172)
TreatxPer3					0.00896 (0.0172)
TreatxPer4					0.00977 (0.0172)
TreatxPer5					0.0114 (0.0172)
TreatxPer6					0.0117 (0.0172)
TreatxPer7					0.0158 (0.0172)
PeriodexReformår	Ja	Ja	Ja	Ja	Ja
FE	Ja	Ja	Ja	Ja	Ja
N	10131	9988	8811	3883	10131
<i>Antall regioner droppet:</i>					
Treatmentregioner	3	16	3	16	3
Kontrollregioner	0	0	8	37	0

Standardavvik i parentes

* $p < .10$, ** $p < .05$, *** $p < .01$

5.4 Modellutvidelse 3: Heterogenitet basert på næringsaktivitet

I modellutvidelse 3 undersøker vi om effekten kan være signifikant for regioner med høyere antall virksomheter per 1000 innbyggere. Avgrensningene forteller noe om næringslivet i regionen, og kan ha forskjellig påvirkning på befolkningsutviklingen.

Vi benytter en FE estimering i alle kolonnene i modellutvidelse 3. I kolonne (1) droppes alle treatmentregioner som har under 8 virksomheter per 1000 innbyggere. Videre i kolonne (2) og (3) innfører vi strengere begrensninger. Alle treatmentregioner som har henholdsvis under 10 og 11 virksomheter per 1000 innbyggere, blir droppet fra datasettet. Vi benytter de samme avgrensningene i kolonne (4)-(6), men nå gjelder avgrensningene for både kontroll- og treatmentregioner. I gjennomsnitt er det 7,71 virksomheter per 1000 innbyggere. Se tabell A.1 og A.2 i appendix for oversikt over hvilke regioner som blir droppet ved de 6 spesifikasjonene. Ved å ta totalt antall observasjoner i datasettet (10164) fratrukket droppede observasjoner² får vi antall observasjoner (N) i de ulike spesifikasjonene i Tabell 5.4.

I kolonne (1) får vi en effekt på 2,09 prosentpoeng. Effekten er signifikant ved et 10 prosent signifikansnivå. Videre i kolonne (2) er ikke effekten signifikant. Årsaken til dette kan være færre observasjoner, og/eller at de treatmentregionene som blir droppet ytterligere, har hatt vekst i befolkningsstørrelse. I kolonne (3) blir effekten igjen signifikant. For regioner med et høyt antall virksomheter per 1000 innbyggere, er økningen i befolkning fra pre-reform til post-reform 3,56 prosentpoeng høyere i treatmentregioner sammenlignet med kontrollregioner. Effekten er signifikant til 5 prosent signifikansnivå. Til forskjell fra kolonne (2) ekskluderes Halden i kolonne (3). Ved nærmere undersøkelse ser vi at Halden har negativ befolkningsvekst både før og etter etableringen. Haldens negative befolkningsutvikling kan dermed forklare

²Viser analytisk tilnærming for endring i antall observasjoner i modellutvidelse 3. I kolonne (1) fjernes 12 treatmentregioner, som utgjør 132 observasjoner (12 treatmentregioner * 11 perioder). I kolonne (2) fjernes 16 treatmentregioner, som utgjør 176 observasjoner (16 treatmentregioner * 11 perioder). I kolonne (3) fjernes 17 treatmentregioner, som utgjør 187 observasjoner (17 treatmentregioner * 11 perioder). I kolonne (4) fjernes totalt 49 regioner, hvor 12 av disse er treatmentregioner og 37 av de er kontrollregioner. Dette utgjør henholdsvis 132 (12 treatmentregioner * 11 perioder) og 6105 (37 b-regioner * 11 perioder * 15 etableringsår) observasjoner. I kolonne (5) fjernes totalt 62 regioner, hvor 16 av disse er treatmentregioner og 46 er kontrollregioner. Dette utgjør henholdsvis 176 (16 treatmentregioner * 11 perioder) og 7590 (46 b-regioner * 11 perioder * 15 etableringsår) observasjoner. I kolonne (6) fjernes totalt 69 regioner, hvor 17 av disse er treatmentregioner og 52 er kontrollregioner. Dette utgjør henholdsvis 187 (17 treatmentregioner * 11 perioder) og 8580 (52 b-regioner * 11 perioder * 15 etableringsår) observasjoner.

at effekten i kolonne (2) ikke blir signifikant. Hittil gir spesifikasjonen i kolonne (3) det mest signifikante resultatet. Det ser ut til at treatmentregioner som har høy næringsaktivitet opplever en effekt.

Vi skal nå undersøke tilfellet hvor vi avgrenser datasettet for alle regioner. Vi benytter de samme avgrensningene som tidligere, men nå gjelder restriksjonene for både kontroll- og treatmentregionene. I tabell A.2 gis en oversikt over hvilke regioner som blir droppet, når vi innskrenker datasettet for både kontroll- og treatmentregioner. Kolonnene (4) og (5) viser ingen signifikant effekt av interessevariabelen, mens effekten i kolonne (6) er signifikant til 1 prosent signifikansnivå. Effekten blir mer signifikant når vi dropper både treatment- og kontrollregioner med antall virksomheter under 11, og ikke kun treatmentregioner. Dette indikerer at i kolonne (6) har treatmentregionene som er med hatt en større befolkningsvekst etter etablering enn kontrollregionene som er med, i tillegg har treatment- og kontrollgruppen hatt lik befolkningsutvikling før etablering. Den strengeste avgrensningen gir mest signifikant effekt i denne analysen. Resultatene tyder på at kun regioner med svært høyt antall virksomheter får effekt av høyskoleetableringer.

Standardavviket har blitt betydelig lavere i kolonnene hvor restriksjonene gjelder alle regioner. Lavere standardavvik betyr at estimatene er mer presise. Det kan også ha betydning at treatment- og kontrollgruppen har likere type regioner i datasettet i kolonne (4)-(6) sammenlignet med kolonnen (1)-(3) ettersom begrensningen gjelder begge gruppene.

Effekten i kolonne (6) for henholdsvis kontroll- og treatmentregioner:

Kontrollregioner

$$\frac{\partial \ln(\text{befolkning})}{\partial \text{postreform}} = 0,125$$

Når dummy-variabelen postreform går fra 0 til 1, øker befolkningen med 12,5 prosent. Det betyr at økningen i befolkning fra pre-reform til post-reform er 12,5 prosent i

kontrollregionene.

Treatmentregioner

For å finne befolkningsveksten legger vi til den ytterligere effekten av å være en treatmentregion i perioden etter etableringen.

$$\frac{\partial \ln(\text{befolkning})}{\partial \text{postreform}} = 0,125 + 0,0280 = 0,153$$

Økningen i befolkningen fra pre-reform til post-reform er på 15,3 prosent i treatment-regionene. Økningen i befolkning fra pre-reform til post-reform er 2,8 prosentpoeng høyere i treatmentregioner sammenlignet med kontrollregioner. Denne effekten av høyskoleetablering på befolkningsstørrelse er signifikant til 1 % signifikansnivå.

Det er regioner med initielt høyt antall virksomheter per 1000 innbyggere som får signifikant effekt. Det er viktig å nevne at antall observasjoner har falt betydelig i kolonne (6). Likevel er det tilstrekkelig med observasjoner til at denne effekten blir signifikant. I kolonne (6) er regresjonen gjennomført med 8 kontrollregioner og 7 treatmentregioner i datasettet.

Tabell 5.4: Heterogenitet basert på næringsaktivitet

	Virk>8 Treatment	Virk>10 Treatment	Virk>11 Treatment	Virk>8 Alle regioner	Virk>10 Alle regioner	Virk>11 Alle regioner
	(1)	(2)	(3)	(4)	(5)	(6)
Avhengig variabel	ln_bef	ln_bef	ln_bef	ln_bef	ln_bef	ln_bef
Postreform	0.0664*** (0.0101)	0.0664*** (0.0101)	0.0664*** (0.0101)	0.0990*** (0.00915)	0.0958*** (0.0115)	0.125*** (0.0172)
TreatxPost	0.0209* (0.0109)	0.0220 (0.0134)	0.0356** (0.0143)	0.00625 (0.00625)	0.0114 (0.00765)	0.0280*** (0.00961)
PeriodexReformår	Ja	Ja	Ja	Ja	Ja	Ja
FE	Ja	Ja	Ja	Ja	Ja	Ja
N	10032	9988	9977	3927	2398	1397
<i>Antall regioner droppet:</i>						
Treatmentregioner	12	16	17	12	16	17
Kontrollregioner	0	0	0	37	46	52

Standardavvik i parentes

* $p < .10$, ** $p < .05$, *** $p < .01$

5.4.1 Årlige effekter ved høy næringsaktivitet

Det kan ta tid før effekten av høyskoleetableringen slår fullt ut på befolkningsstørrelse. Dette undersøkes ved å se om effekten av høyskoleetablering har ulikt signifikansnivå for enkeltperioder. Vi ser på årlige effekter med avgrensningen fra kolonne (6) i Tabell 5.4. Effektene er dekomponert for hvert enkelt periode med periode -3 som referanseår. Til forskjell fra modellutvidelse 1 finner vi signifikant effekt i årene etter høyskoleetablering.

De årlige effektene i Tabell 5.4.1 tolkes som endring i befolkning fra år -3 til aktuelt år, for treatment- versus kontrollregioner. I tabellen er endring i befolkning fra år -3 til -2 og fra -3 til -1 ikke signifikant forskjellig for treatment- og kontrollregioner, noe som styrker forutsetningen om parallelle trender. Dette signaliserer at gruppene har hatt lik befolkningsutvikling før høyskoleetablering, og følgelig en god kontrollgruppe.

I tabell 5.4.1 blir effekten av høyskoleetablering signifikant først i periode 4. Endring i befolkning fra år -3 til år 4 er 3,49 prosentpoeng høyere i treatmentregioner enn i kontrollregioner. Effekten blir sterkere etter 5, 6 og 7 år, og disse estimatene er mer signifikante. Dette tyder på at det tar noen år før effekten av høyskoleetablering slår ut på befolkningsstørrelsen i regionene. Det kunne vært interessant å legge til flere perioder i datasettet, og se om disse ville gitt større og mer signifikante effekter på befolkningsstørrelsen. På en annen side må vi passe på å ikke ta med for mange perioder etter etablering, da trenden i befolkningsstørrelse kan påvirkes av andre faktorer.

Tabell 5.4.1: Årlige effekter ved høy næringsaktivitet

Virksomheter	
Alle regioner	
(1)	
Avhengig variabel	ln_befolkning
TreatxPer-2	0.00535 (0.0201)
TreatxPer-1	0.00511 (0.0201)
TreatxPer0	0.0112 (0.0201)
TreatxPer1	0.0181 (0.0201)
TreatxPer2	0.0232 (0.0201)
TreatxPer3	0.0289 (0.0201)
TreatxPer4	0.0349* (0.0201)
TreatxPer5	0.0409** (0.0201)
TreatxPer6	0.0468** (0.0201)
TreatxPer7	0.0481** (0.0201)
PeriodexReformår	Ja
FE	Ja
N	1397
<i>Antall regioner droppet:</i>	
Treatmentregioner	17
Kontrollregioner	52

Standardavvik i parentes

* $p < .10$, ** $p < .05$, *** $p < .01$

Hovedresultatet fra analysen viser at det er regioner med høyt antall virksomheter per 1000 innbyggere initielt som får signifikant effekt. Dette forteller oss at det kreves en viss næringsaktivitet i regionen fra før for at høyskoleetablering skal ha effekt på befolkningsstørrelse. Effekten slår ikke ut før 4 år etter etableringen. Dette kan skyldes individer og bedrifters adferd som krever tid for å endres (King and Behrman, 2009).

Antall virksomheter per 1000 innbyggere kan tolkes som et mål på sysselsettingen i

regionen. Regioner med høyt antall virksomheter per 1000 innbyggere vil sannsynligvis ha flere arbeidsplasser tilgjengelig enn regioner med lavt antall. Det vil være mer attraktivt å flytte til regioner med større muligheter på arbeidsmarkedet enn regioner med lite jobber tilgjengelig. Eksempel; en lærer som vil flytte til en annen region på grunn av ny stilling på høyskolen har kanskje familiemedlemmer som også vil trenge jobb i regionen. Jobbmuligheter kan også ha innflytelse på om studentene flytter eller blir værende. Vi ser at effekten av høyskoleetablering kun er tilstede i regioner med høy næringsaktivitet.

Tidligere litteratur fant effekter av utdanningsinstitusjoner på regional økonomi og bedrifters lokalisering. Det kan være at regioner med høy næringsaktivitet fra før har bedre forutsetninger for at nye bedrifter velger å lokalisere seg i regionen. I tillegg kan eventuelle produksjons- og innovasjonsvirkninger i større grad feste seg i regioner med høy næringsaktivitet fra før. Dette er mulige faktorer som kan ha bidratt til den observerte befolkningsøkningen som følge av høyskoleetablering.

6. Konklusjon

Denne oppgaven undersøker effekten av høyskoleetablering på regional befolkningsstørrelse. Paneldatasettet inneholder observasjoner over norske regioner i perioden 1950-1996. Vi benytter metoden Difference in Differences med et event study dataoppsett for å svare på vår problemstilling. Ved regresjoner hvor hele datasettet er inkludert, finner vi ingen signifikant effekt av høyskoleetablering på befolkningsutvikling.

I den videre analysen undersøker vi om effekten kan være forskjellig for ulike regioner med bestemte karakteristikk. Resultatene viser ingen tegn til heterogenitet med hensyn på alderssammensetning i regionene. Når vi betrakter heterogenitet basert på næringsaktivitet, finner vi at den strengeste avgrensningen gir det mest signifikante resultatet i vår analyse. Ved denne avgrensningen fjernes regioner med antall virksomheter under 11 per 1000 innbyggere fra datasettet. Økningen i befolkning fra pre-reform til post-reform er 2,8 prosentpoeng høyere i treatmentregioner sammenlignet med kontrollregioner. Denne effekten av høyskoleetablering på befolkningsstørrelse er signifikant til 1 % signifikansnivå. Vi finner at effekten kommer først 4 år etter etableringen, og slår sterkere ut de tre neste årene. Årene før etablering gir ingen signifikante estimater. Dette styrker antakelsen om parallelle trender, og følgelig troverdigheten til hovedresultatet.

Resultatene i vår analyse viser at det er regioner med inititelt høyt antall virksomheter per 1000 innbygger som får signifikant effekt av høyskoleetablering. Vi kan tolke dette som et mål på jobbmuligheter i regionen. Det vil være mer attraktivt å flytte til regioner med gode muligheter på arbeidsmarkedet. Funnene tyder på at effekten av høyskoleetablering på befolkningsstørrelse kun er til stede i regioner hvor det er høy næringsaktivitet i regionen fra før av.

Referanser

- Aamodt, P. O. and Lyby, L. (2019). Instrument eller institusjon? skiftende politiske prioriteringer. *Geografi, kunnskap, vitenskap: Den regionale UH-sektorens framvekst og betydning*.
- Abadie, A. and Cattaneo, M. D. (2018). Econometric methods for program evaluation. *Annual Review of Economics*, 10(1):465–503.
- Benoit, K. (2011). Linear regression models with logarithmic transformations. *London School of Economics, London*, 22(1):23–36.
- Berge, C. (2020). Hvor mye pendling er det mellom kommuner? *Statistisk sentralbyrå, hentet fra: <https://www.ssb.no/arbeid-og-lonn/artikler-og-publikasjoner/hvor-mye-pendling-er-det-mellom-kommuner>*.
- Bhuller, M. S. (2009). Inndeling av norge i arbeidsmarkedsregioner. *Statistisk sentralbyrå*, (24).
- Bonander, C., Jakobsson, N., Podestà, F., and Svensson, M. (2016). Universities as engines for regional growth? using the synthetic control method to analyze the effects of research universities. *Regional Science and Urban Economics*, 60:198–207.
- Carneiro, P., Liu, K., and Salvanes, K. G. (2021). The supply of skill and endogenous technical change: Evidence from a college expansion reform. *Institutt for samfunnsøkonomi, NHH*.
- Chen, C. (2017). Limitations to event studies and how they apply. *Journal of Banking and Finance*, (2):1–3.
- Drucker, J. (2016). Reconsidering the regional economic development impacts of higher education institutions in the united states. *Regional Studies*, 50(7):1185–1202.
- Drucker, J. and Goldstein, H. (2007). Assessing the regional economic development impacts of universities: A review of current approaches. *International regional science review*, 30(1):20–46.

- Fylkeskommunen-Trøndelag (2019). Befolkningsutvikling og sysselsettingsutvikling. <https://www.trondelagfylke.no/vare-tjenester/plan-og-areal/kart-statistikk-og-analyse/nyhetsarkiv-kart-og-statistikk/faktafredag—befolkningsutvikling-og-sysselsettingsutvikling/>. Hentet 17. april 2022.
- Gertler, P. J., Martinez, S., Premand, P., Rawlings, L. B., and Vermeersch, C. M. (2016). *Impact evaluation in practice*. World Bank Publications.
- Hustoft, A. G., Hartvedt, H., Nymoene, E., Stålnacke, M., and Utne, H. (1999). Standard for økonomiske regioner. *Statistisk sentralbyrå*.
- Kantor, S. and Whalley, A. (2019). Research proximity and productivity: long-term evidence from agriculture. *Journal of Political Economy*, 127(2):819–854.
- King, E. M. and Behrman, J. R. (2009). Timing and duration of exposure in evaluations of social programs. *The World Bank Research Observer*, 24(1):55–82.
- Knutsen, T. K., Modalsli, J. H., and Rønning, M. (2020). Distance and choice of field: Evidence from a Norwegian college expansion reform. *Statistisk Sentralbyrå*.
- Lee, J. (2021). The role of a University in Cluster Formation: Evidence from a National Institute of Science and Technology in Korea. *Regional Science and Urban Economics*, 86:103617.
- Li, J., Liu, S., and Wu, Y. (2020). Identifying knowledge spillovers from universities: quasi-experimental evidence from urban China. *Available at SSRN 3621422*.
- Moretti, E. (2010). Local multipliers. *American Economic Review*, 100(2):373–77.
- Pokarzhevskaya, G. and Regnér, H. (2015). Val av högskoleort—regionala mönster bland kvinnor och män. *Report, Saco, Stockholm*. Available online: www.saco.se/globalassets/saco/dokument/rapporter/2015-val-av-hogskoleort.pdf.
- Regjeringen (2020). Oversikt over kommunesammenslåinger og nye kommunenavn. <https://www.regjeringen.no/no/tema/kommuner-og-regioner/kommunestruktur/nye-kommuner/id2470015/>. Hentet 4.mars 2022.

- St.prp.nr.136 (1969). Om prøvedrift med distriktshøgskolar.
<https://www.stortinget.no/nn/Saker-og-publikasjoner/Stortingsforhandlingar/Lesevisning/?p=1968-69paid=2wid=apsid=DIVL1579>. Hentet 28.april 2022.
- Wing, C., Simon, K., and Bello-Gomez, R. A. (2018). Designing difference in difference studies: best practices for public health policy research. *Annual review of public health*, 39.
- Wooldridge, J. M. (2015). *Introductory econometrics: A modern approach*. Cengage learning.

A. Appendix

A.1 Oversikt over treatmentregioner og antall virksomheter per 1000 innbyggere

Oransje = treatmentregioner med antall virksomheter under 8

Blå = treatmentregioner med antall virksomheter over 8 og under 10

Rød = treatmentregioner med antall virksomheter over 10 og under 11

Table A.1: Oversikt over alle treatmentregioner og antall virksomheter per 1000 innbyggere

Treatmentregion	Antall virksomheter per 1000 innbyggere
Ørsta/Volda	3,66
Steinkjer	4,06
Levanger/Verdalsøra	5,00
Vest-Finnmark	5,08
Sunnfjord og Ytre Sogn	5,38
Elverum	5,53
Indre sogn	6,02
Stord	6,28
Hønefoss	6,97
Gjøvik	7,40
Midt-Telemark	7,55
Sør-Troms	7,87
Lillehammer	8,38
Molde	8,66
Kongsberg	8,76
Midt-Finnmark	9,96
Halden	10,79
Fredrikstad/Sarpsborg	11,19
Arendal	11,43
Tromsø	11,80
Haugaland	12,31
Stavanger/Sandnes	13,65

Fortsetter på neste side

Table A.1 – Fortsettelse fra forrige side

Region	Antall virksomheter per 1000 innbyggere
Kristiansand	15,53
Ålesund	16,90

A.2 Oversikt over alle regioner og antall virksomheter per 1000 innbyggere

Oransje = regioner med antall virksomheter under 8

Blå = regioner med antall virksomheter over 8 og under 10

Rød = regioner med antall virksomheter over 10 og under 11

Table A.2: Oversikt over alle regioner og antall virksomheter per 1000 innbyggere

Regioner	Antall virksomheter per 1000 innbyggere
Trondheim Forstad	2,81
Nord-Gudbrandsdal	3,12
Trøndelag sør	3,23
Værnes	3,59
Ørsta/Volda	3,66
Steinkjer	4,06
Tynset	4,13
Fosen	4,23
Nord-Troms	4,4
Nordhordland	4,41
Midt-Gudbrandsdal	4,6
Jæren	4,93
Levanger/Verdalsøra	5
Fortsetter på neste side	

Table A.2 – Fortsettelse fra forrige side

Region	Antall virksomheter per 1000 innbyggere
Setesdal og Sirdal	5,03
Sunndal/Surnadal	5,08
Vest-Finnmark	5,08
Kongsvinger	5,2
Sunnfjord og Ytre Sogn	5,38
Salten	5,5
Elverum	5,53
Brønnøysund	5,54
Vesterålen	5,95
Midt-Troms	5,99
Indre sogn	6,02
Vest-Telemark	6,05
Holmestrand	6,12
Orkland og øyregionen	6,19
Namdal	6,2
Stord	6,28
Sandnessjøen	6,34
Voss	6,37
Midthordland	6,41
Hamar	6,55
Hallingdal	6,58
Valdres	6,69
Indre Ryfylke	6,7
Søre Sunnmøre	6,77
Ullensaker/Eidsvoll	6,78
Indre Hardanger	6,85
Hønefoss	6,97
Lillestrøm	7,2
Sunnhordland Aust	7,37
Gjøvik	7,4
Kristiansand Omland	7,53
Midt-Telemark	7,55
Mo i Rana	7,87
Fortsetter på neste side	

Table A.2 – Fortsettelse fra forrige side

Region	Antall virksomheter per 1000 innbyggere
Sør-Troms	7,87
Øst-Telemark	7,97
Ålesund Omland	7,99
Nordfjord og Kinn	8,19
Dalane	8,33
Lillehammer	8,38
Indre Østfold	8,44
Molde	8,66
Mandal	8,71
Kongsberg	8,76
Follo	8,81
Øst-Finnmark	9,12
Bodø	9,47
Østregionen	9,56
Mosjøen	9,65
Midt-Finnmark	9,96
Flekkefjord	10,13
Lyngdal/Farsund	10,2
Sandefjord/Larvik	10,27
Grenland	10,58
Moss	10,61
Halden	10,79
Trondheim	10,91
Fredrikstad/Sarpsborg	11,19
Arendal	11,43
Drammen	11,44
Tønsberg/Horten	11,47
Tromsø	11,8
Haugaland	12,31
Bergen	12,38
Kristiansund	12,48
Porsgrunn/Skien	12,53
Asker/Bærum	12,56
Fortsetter på neste side	

Table A.2 – Fortsettelse fra forrige side

Region	Antall virksomheter per 1000 innbyggere
Lofoten	13,24
Stavanger/Sandnes	13,65
Kristiansand	15,53
Ålesund	16,9
Oslo	17,43

A.3 Oversikt over etableringsår for ulike regioner

Region	Etableringsår
Hønefoss	1955
Molde	1958
Midt-Finnmark	1960
Fredrikstad/Sarpsborg	1965
Ålesund	1965
Arendal	1967
Gjøvik	1966
Kristiansand	1969
Stavanger/Sandnes	1969
Kongsberg	1970
Ørsta/Volda	1970
Lillehammer	1970
Midt-Telemark	1970
Indre Sogn	1975
Vest-Finnmark	1977
Halden	1977
Stord	1977
Sunnfjord og Ytre Sogn	1979
Elverum	1979
Haugaland	1980
Steinkjer	1980
Tromsø	1981
Sør-Troms	1982
Levanger/Verdalsøra	1989

