

Forord

Denne masteroppgåva er den avsluttande delen av eit fem-årig studium i samfunnsøkonomi på NTNU.

Først og fremst ønskjer eg å takka veiledaren min, Torberg Falch, for grundig og konstruktiv rettleiing.

Eg vil også takka alle studievenner og -venninner for eit godt studiemiljø og for alle seriøse og useriøse diskusjonar i løpet av desse åra.

Trondheim

30. mai 2012

Anders Linga

Innhold

1	Innleiing	1
1.1	Oppbygging av oppgåva	3
2	Teori og tidlegare empiriske studiar	5
2.1	Mikroøkonomisk teori om utdanning og grunnar til fråfall	5
2.2	Økonomisk vekstteori	7
2.3	Tidlegare empiriske studiar om humankapital og utvikling	8
3	Empirisk tilnærming	13
3.1	Minste kvadraters metode	13
3.2	Estimering med faste regioneffektar	15
3.3	Estimering med tilfeldige effektar	17
3.4	Testmetodar	18
3.4.1	t-test	18
3.4.2	F-test	19
3.4.3	Hausman-testing	19
4	Datamateriale, deskriptiv statistikk og empirisk spesifikasjon	21
4.1	Definisjon av regionar	21
4.2	Fråfallsstatistikk	22
4.3	Folketalsstatistikk	24
4.4	Inntektsstatistikk	25
4.5	Andre variablar	28
4.6	Empirisk spesifikasjon	29
5	Resultat	33
5.1	Inntektsvekst	33
5.2	Folketalsutvikling	39
6	Oppsummering og konklusjon	45
	Referansar	47
A	Appendiks	49
A.1	Deskriptiv statistikk	49
A.2	Tabell 7. Fullføringsstatistikk per år	53

Figurer

1	Gjennomsnittleg fullføringsgrad sortert på år	23
2	Gjennomsnittleg fullføringsgrad i dei ulike regionane	23
3	Folketalet i regionane 1.1.1990, målt i tusen innbyggjarar	24
4	Gjennomsnittleg folketalsendring i regionane	25
5	Gjennomsnittleg inntekt i regionane i år 2000, målt i tusen kroner	27
6	Gjennomsnittleg inntektsvekst i regionane	27
7	Effekt av fullføringsgrad på inntektsutvikling	38
8	Effekt av fullføringsgrad på folketalsutvikling	43

Tabeller

1	Oppsummering av kontrollvariablane i 1991	28
2	Oppsummering av kontrollvariablane i 2008	29
3	Estimering av inntektsnivå	33
4	Estimering av inntektsvekst	36
5	Estimering av folketalsutvikling	40
6	Regionvis deskriptiv statistikk	51
7	Årleg oppsummering av fullføringsstatistikken	53

1 Innleiing

Denne masteroppgåva handlar om samanhengen mellom fullføring i vidaregåande opplæring og den økonomiske utviklinga i norske regionar. I Noreg er skulesystemet bygd opp med ein obligatorisk 10-årig grunnskule og deretter vidaregåande opplæring. Den vidaregåande utdanninga kvalifiserer enten til opptak i høgare utdanning, gjennom at dei som fullfører får studiekompetanse, eller til yrkeskompetanse. På grunn av dette er det rimeleg å seia at den vidaregåande opplæringa i praksis er obligatorisk den også. Likevel fell rundt 1 av 3 frå, og fullfører aldri den vidaregåande opplæringa. Denne fullføringsgraden har halde seg nokolunde stabil sidan byrjinga på 90-talet, før dette var fråfallet enda større.

Fullføringsgraden viser i denne oppgåva til andelen elevar som fullfører vidaregåande opplæring innan 5 år etter at dei gjekk ut av ungdomsskulen. Det er store geografiske skilnader på fullføringsnivået innad i Noreg. Særleg Finnmark og Nord-Troms har vedvarande låg fullføringsgrad, om lag 50 prosent mot ca. 2/3 i resten av landet. Nordland og resten av Troms har også noko lågare fullføring enn landsgjennomsnittet. Det er også store skilnader i andelen som fullfører mellom dei ulike utdanningsprogramma. Programma som gjev studiekompetanse, Studiespesialiserande, Musikk, dans og drama og Idrettsfag, har ein vesentleg høgare fullføringsgrad enn dei meir yrkesretta programma (med Kjemi og prosessfag og Medie og kommunikasjon som unntak).¹

Dei siste åra har det vore eit stort offentleg fokus på fullføringsgraden i den vidaregåande skulen. Mellom anna handlar Stortingsmelding 44 (2008-2009) «Utdanningslinja» i stor grad om denne problematikken. I ei pressemelding med overskrift «Kamp mot frafall er jobb nr. 1» sa dåverande kunnskapsminister Bård Vegar Solhjell om denne meldinga at «Det store frafallet i videregående opplæring er en av de faktorene som skaper mest sosial ulikhet i vårt samfunn og de personlige omkostningene er store. Vi må gå løs på frafallutfordringen med nye grep og vilje til å tenke nytt.»² Dette illustrerer at dei siste åra har det å få fleire til å fullføra vidaregåande opplæring vorte eit av dei store skulepolitiske fokusområda i Noreg, både blant politikarar og skulebyråkratar. Fleire ulike løysingar har blitt foreslått og prøvd ut for å få opp fullføringa.

Målet om auka fullføring er basert på ei oppfatning om at det er viktig for å knytte folk til arbeidsmarknaden, for kampen mot sosial ulikskap og for den generelle økonomiske utviklinga i samfunnet. I eit samfunnsøkonomisk perspektiv er det openbart interessant å studera desse effektane av auka fullføring. Kor store økonomiske og menneskelege ressursar ein legg ned i arbeidet mot fråfall bør nødvendigvis avhenga av kor stor innverknad fråfallet faktisk har, både for enkeltmenneska som fell frå og for samfunnet som heilheit.

¹Falch et al. (2010)

²<http://www.regjeringen.no/nb/dep/kd/pressepenter/pressemeldinger/2009/utdanningslinja-kamp-mot-fracfall-er-jobb.html?id=566373>

Det er fylkeskommunane som er ansvarleg for den vidaregåande opplæringa, etter læreplanar og læremål frå statleg hald. Dette betyr i praksis at fylkeskommunane står meir eller mindre fritt til å sjølv velge skulestruktur og kor store ressursar dei ønskjer å bruka på den vidaregåande opplæringa, gitt visse minimumsrammer. I eit økonomisk perspektiv er det svært relevant for ein fylkeskommune om høgare fullføringsgrad i vidaregåande opplæring bidreg til utvikling eller ikkje. Utanom dei lovpålagde oppgåvene til fylkeskommunen, som deler av vegnettet og vidaregåande opplæring, ser mange regional utvikling som ei av dei viktigaste oppgåvene fylkeskommunen har. Til dømes har dei fleste, om ikkje alle, fylkeskommunane eigne næringsavdelingar og planar for regionale utvikling.³ Difor er det særleg interessant å studera samanhengen mellom fullføringsgrad i vidaregåande opplæring og regional utvikling. Dersom det viser seg at høgare fullføring ikkje bidreg til inntektsvekst, eller at høgare fullføring ikkje bidreg til befolkningsauke, er det i eit utviklingsperspektiv mindre effektivt for fylkeskommunane å bruka store ressursar på den vidaregåande opplæringa enn det elles ville vore. Viss høgare fullføring faktisk bidreg til regional økonomisk vekst, vil det vera eit insentiv for fylkeskommunane til å bruka enda større ressursar for å auka fullføringsgraden.

Nullhypotesen i estimeringane i denne oppgåva vil likevel vera at det ikkje er ein samheng mellom fullføringsgraden og økonomisk utvikling. Alternativhypotesane er at fullføringsgraden faktisk har ein påverknad på den regionale utviklinga.

Det finst fleire argument både for og mot at fullføringsgraden skal påverka den økonomiske utviklinga i norske regionar. I henhold til økonomisk teori om at auka humankapitalnivå gjev større utvikling, burde ein tru at høgare fullføringsgrad har ein positiv påverknad på den regionale økonomiske utviklinga. Det er ikkje urimeleg å tenkja seg at dersom humankapitalnivået påverkar den generelle økonomiske situasjonen, bør dette også gje til dømes større inntektsvekst. Eit argument som kan strida mot at fullføring gjev positiv folketalsendring er at forskning viser at høgare utdanningsnivå også gjev høgare migrasjon.⁴ I mange regionar må ein flytta til ein større by eller tettstad om ein ønskjer å ta høgare utdanning. Det er også rimeleg å tru at ein del av personane dette gjeld ikkje flyttar tilbake til heimregionen etter endt utdanning. Difor kan det vera at for enkelte regionar vil større fullføringsgrad faktisk føra til at fleire flyttar ut av regionen og at auka fullføring dermed fører til negativ folketalsutvikling.

Som eg skal skriva litt om i teoridelen av oppgåva, viser forskning at det å ikkje fullføra vidaregåande opplæring i snitt har ein del andre negative konsekvensar for dei som ikkje fullfører. Slik sett vil det for samfunnet som heilskap vera både lønnsamt og hensiktsmessig å jobba for å få flest mogleg til å fullføra den vidaregåande opplæringa. Sidan den

³To eksempel er Akershus og Hordaland, www.akershus.no/tema/regionalutvikling/ og <http://www.hordaland.no/Hordaland-fylkeskommune/Naring/Regionalt-utviklingsprogram-RUP/>

⁴Sjå til dømes Carlsen et al. (2010)

vidaregåande opplæringa i stor grad også blir drifta lokalt, kan eventuelle negative konsekvensar av høgare fullføring føra til mindre økonomisk satsing på vidaregåande opplæring, og dermed lågare fullføringsgrad enn det som er optimalt samfunnet sett under eitt. Å studera dei regionale vekstkonsekvensane av høgare fullføringsgrad kan difor vera svært interessant, både i eit samfunnsøkonomisk og politisk perspektiv.

1.1 Oppbygging av oppgåva

I kapittel 2 tek eg for meg ein del økonomiske teoriar og empiriske studiar som er relevante for denne oppgåva. Kapittel 3 handlar om dei estimeringsmetodane eg brukar i oppgåva. Kapittel 4 inneheld deskriptiv statistikk, ei presentasjon av datasettet og den empiriske spesifikasjonen eg brukar i estimeringane. I kapittel 5 vert resultata av estimeringane presenteret, mens kapittel 6 gjev ei oppsummering av dei viktigaste funna i denne oppgåva.

2 Teori og tidlegare empiriske studiar

I dette kapittelet skal eg gå gjennom enkelte økonomiske teoriar og tidlegare empiriske studiar som er relevante for denne oppgåva. I den første delen av dette kapittelet ser eg på mikroøkonomisk forskning om mellom anna utdanning og løn, samt ulike grunnar til at elevar ikkje fullfører vidaregåande opplæring. I del to tek eg for meg teoriar om økonomisk vekst, og då med fokus på utdanningsnivå og humankapital. Til slutt ser eg på nokre empiriske studiar som handlar om samanhengen mellom økonomisk utvikling og humankapital.

2.1 Mikroøkonomisk teori om utdanning og grunnar til fråfall

I utgangspunktet går mikroøkonomisk teori ut frå avgjersler på bakgrunn av forventet marginalnytte og -kostnad. Når ein i økonomisk samanheng estimerer nettonytte av utdanning, ser ein med andre ord på forventet marginalnytte og -kostnad av å ta utdanning. Becker (1964) la grunnlaget for mykje av dagens økonomiske forskning knytta til utdanning, lønn og humankapital.

Til dømes presenterer Ehrenberg og Smith (2003), kapittel 9, ein enkel modell for val av utdanningslengde. Dei deler kostnadene ved utdanning i tre hovuddelar. Direkte kostnader, som til dømes kjøp av bøker og anna læremateriell, tap av inntekter ein elles ville fått ved å vera i jobb og ubehag som følgje av at undervisninga kan vera vanskeleg og keisam. Forventa avkastning av utdanning kjem mellom anna som følgje av større framtidig løn og betra jobbmoglegheiter. Ehrenberg og Smith går ut frå at alle personar er nyttemaksimerande over eit livstidsløp. Dette betyr i praksis at personar vil ta ekstra utdanning så lengje noverdien av all framtidig nytte den ekstra utdanninga gjev er større enn kostnadene. Med andre ord vil folk tilpassa seg der marginalkostnadene er lik marginalnytta av utdanning.

Ashenfelter og Rouse (1998) set opp to stiliserte fakta dei meiner alle mikroteoriar som omhandlar samanhengen mellom utdanning og lønnsnivå bør ta omsyn til. Det eine er at skuleprestasjonar blir sterkt påverka av familiebakgrunnen til elevane, det andre er at det i praksis er ein log-lineær samanheng mellom inntekt og antal års utdanning.

Eit viktig spørsmål er om samanhengen mellom inntekt og løn berre speglar dugleikar som er felles for dei som utdannar seg, slik at utdanning berre blir ein måte å signalisere ibuande kompetanse på, eller om det auka lønnsnivået er ein kompensasjon for ei kompetanseheving auka utdanning i seg sjølv gjev. Dette vil nødvendigvis også ha mykje å seia for korleis utdanningssystemet bør leggast opp. Om det er slik at utdanning berre synleggjer allereie ibuande kompetanse, vil utdanningssystemet i praksis ha mindre å bety

i eit samfunnsøkonomisk perspektiv. Viss utdanning i seg sjølv gjev høgare kompetanse, og som følgje av dette også auka lønn, vil eit godt utdanningssystem vera svært viktig for å sikra innbyggjarane i ein region gode inntekter.

I tråd med Ehrenberg og Smith (2003) kan det å enten fullføra vidaregåande eller å dropa ut sjåast på som eit val den enkelte elev gjer. Dei som droppar ut gjer det fordi dei trur nytten ved å gjera dette vil vera høgare enn framtidig nytte ved å fullføra. Grovt sett kan ein seia at dette skjer enten fordi eleven vurderer den framtidige marginalnytten ved å fullføra som låg eller at marginalkostnadene ved å fullføra er høge.

Ehrenberg og Smith antek i sin enkle modell at alle personar har eit livstidsperspektiv når dei tek avgjersler om å utdanna seg eller ikkje. Med ei låg diskonteringsrate vil framtidig inntekt bety mykje når noverdien av nytten reknast ut. Med ei høg diskonteringsrate vil inntekt i nærmaste framtid bety relativt meir i forhold til framtidig inntekt. Då vil vala ein gjer bli svært påverka av kortsiktige inntektsmoglegheiter. Det er ikkje urimeleg å tru at denne diskonteringsfaktoren ofte endrar seg med alderen, også undervegs i den vidaregåande opplæringa. Dette kan påverka andelen som bytter linje undervegs i den vidaregåande opplæringa eller som vel å ikkje fullføra i det heile tatt.

I praksis vil forventningane ein elev har til nytten av utdanning svært ofte endra seg over tid. Det som er relevant for fullføringa er dersom forventningane endrar seg vesentleg i løpet av åra i vidaregåande opplæring. Dette kan til dømes vera fordi marginalnytten av studiane vart overvurdert då ein valde program, eller at marginalkostnadene vart undervurdert. For mange vil det naturlege valet i slike tilfelle vera å bytta linje, og dermed sinka fullføringa. For enkelte endar ei slik endring i forventningane at eleven droppar ut og ikkje fullfører ei vidaregåande opplæring i det heile tatt.

At nytten ved å fullføra er låg kan til dømes vera på grunn av dårlege utsikter til jobbar eller at utdanning betyr lite for framtidig løn. Ut frå dette er det rimeleg å tru at arbeidsmarknadssituasjonen på tidspunktet påverkar om ein elev fullfører vidaregåande eller ønskjer å gå rett ut i jobb. I periodar med stor arbeidsløyse vil ein kunne sjå at mange tek meir utdanning, enten for å oppnå meir jobbrelatert kompetanse eller for å bruka fleire år på utdanninga i håp om at arbeidsmarknadssituasjonen betrar seg i løpet av dei åra. På same måte vil fleire gå rett ut i arbeidslivet utan å ha fullført vidaregåande opplæring i periodar med mange ledige jobbar og få arbeidsledige først.

Truleg er det i praksis mange ulike grunnar til at folk ikkje fullfører vidaregåande opplæring. Falch et al. (2010), tabell 3.7, viser at i Noreg skil karakternivået frå grunnskulen seg ut som den klart viktigaste enkeltfaktoren for om ein elev fullfører vidaregåande opplæring eller ikkje. Ein studie frå Reiling og Strøm (2012) viser også at effekten av arbeidsmarknadssituasjonen på fullføringsgraden er til stades i det norske skulesystemet. I periodar med lågkonjunktursituasjonar vel fleire å fullføringa vidaregåande opplæring, og motsatt

er det færre som fullfører i høgkonjunktursituasjonar. Reiling og Strøm viser også til fleire andre studiar frå andre land som finn den same tendensen.

Samanhengen mellom utdanningsnivå og lønn er også mykje behandla i den empiriske litteraturen. Dei aller fleste studiar på feltet finn ein klar positiv korrelasjon mellom desse to. Ashenfelter og Rouse (1998) studerte einegga tvillingar, som genetisk er heilt identiske, og prøvde på denne måten å finna ut om auken i lønn hang saman med genetiske variablar (signalteorien), eller om ulik utdanning også gav ulik løn mellom genetisk like tvillingar. Dei fann at eit år ekstra utdanning gav om lag 9 prosent auka løn for genetisk identiske tvillingar. Norske studiar, mellom anna Falch et al. (2009), argumenterer for at gjennomsnittleg inntektseffekt truleg ligg på rundt 4-5 prosent per år ekstra utdanning. Dette skulle også tilseia at dei som ikkje fullfører vidaregåande utdanning i snitt skal tapa framtidig inntekt som følge fråfallet.

2.2 Økonomisk vekstteori

I dei tidlege vekstteoriane, også kalla neoklassisk vekstteori, er det vekst i fysisk kapital og sparing som i stor grad driv den økonomiske veksten. Solow (1956) og Swan (1956) lanserte to av dei tidlegaste teoriane innan denne typen vekstteori. Desse to teoriane omhandlar mykje av det same og vart skrivne praktisk talt samtidig, og denne typen modellering blir difor ofte kalla Solow-Swan-vekstmodellar. For ein diskusjon av denne teorien, sjå til dømes Aghion og Howitt (2009), kapittel 1. I neoklassiske vekstteoriar kan økonomisk vekst påverkast ved at sparinga i økonomien aukar, og på den måten går kapitalnivået opp. Dette vil likevel ikkje endra veksten på lang sikt, der er den økonomiske veksten lik den teknologiske framgangen. I denne typen vekstmodellering blir humankapital og teknologinivå behandla som eksogene variablar. Endringar i teknologinivået skjer utan påverknad frå økonomiske variablar, og er dermed ikkje mogleg å påverka ved hjelp av økonomisk politikk. I desse er det modellane difor ikkje mogleg for økonomar og politikarar å endra den økonomiske veksten på lang sikt.

I nyare tid har såkalla endogene vekstmodellar tatt stadig meir over som dei leiande innan vekstteoriar. Også her er teknologiutvikling, som kjem som følge av mellom anna humankapital, ein viktig faktor som driv økonomisk vekst. Den store skilnaden er at teknologinivået blir forklart innad i modellen, altså ei endogen forklaring av teknologinivået, derav namnet endogene vekstmodellar. Teknologinivået blir ofte forklart ut frå økonomiske faktorar, som satsinga på utdanning, forskinga til profittsøkande bedrifter og statleg økonomisk forskingsstøtte. Ein av pionerane innan endogen vekstteori var Joseph Schumpeter. Han lanserte begrepet kreativ destruksjon i 1942 (sjå Aghion og Howitt (2009)). Her er det konkurrerande verksemder som driv innovasjonen, og ny teknologi avløyser kontinuerleg gamal teknologi. Denne teknologiske utviklinga er i hovudsak det som driv

den økonomiske veksten i følge Schumpeters teori.

Ein viktig drivfaktor for teknologisk utvikling i mange vekstmodellar er humankapitalnivået. Humankapitalen i ein økonomi kan definerast som befolkningas kunnskap og ferdigheiter. Humankapitalen kan aukast gjennom trening og utdanning, og å bruke ressursar på utdanning kan sjåast på som ein investering i humankapital. Økonomisk teori peikar ofte på at humankapitalnivået er viktig for økonomisk utvikling. Eit mykje sitert døme i nyare tid er Lucas (1988). Han utvikla ein av dei mest kjente teoriene som peikar på samanhengen mellom desse to variablane.

Vandenbussche et al. (2006) argumenterer for at i allereie godt utvikla økonomiar er det ikkje det totale nivået på humankapitalen, men nivået på høgt utdanna arbeidskraft som er relevant for den økonomiske utviklinga. Dei meiner utrent arbeidskraft er relativt sett viktigare for økonomiar som heng etter i den økonomiske utviklinga, sidan desse økonomiane i stor grad baserer seg på imitasjon av eksisterande teknologi og mange arbeidsintensive næringer. I økonomiar som er godt utvikla vil den økonomiske og teknologiske utviklinga i stor grad vera avhengig av innovasjon, og til dette trengst det mykje høgt utdanna arbeidskraft. I følge deira teoretiske modell er det dermed andelen med høgare utdanning som betyr noko, ikkje samfunnets totale humankapital.

2.3 Tidlegare empiriske studiar om humankapital og utvikling

I den empiriske litteraturen om humankapital og økonomisk utvikling studerer ein vanlegvis samanhengen mellom befolkningsandelen med høgare utdanning og økonomisk vekst. Sjølv om teorien seier at humankapital er svært viktig for økonomisk utvikling, har fleire empiriske studiar ikkje klart å finna nokon klar samheng mellom desse to variablane i vestlege land. Eit døme kan vera Krueger og Lindahl (2001), der eit av hovudfunna er at: «The results consistently indicated that education was statistically significantly and positively associated with subsequent growth only for the countries with the lowest level of education.» (side 30).

Pritchett (2001) finn i sin studie ingen positiv korrelasjon mellom utdanning og økonomisk utvikling i det heile tatt. Han listar opp tre moglege forklaringar på dette. Den første forklaringa er at i enkelte land kan institusjonelle svakheiter gjera at ein ikkje får utnytta humankapitalen til å skapa økonomisk vekst. Forklaring nummer to er at marginalavkastninga på utdanning kan ha vorte redusert etter kvart som utdanningsnivået har auka. Den tredje forklaringa han legg fram er at i enkelte land er utdanningssystema så dårlege at lengre utdanning ikkje fører til ein vesentleg auke i humankapitalen, og dermed bidreg lite til auka økonomisk vekst.

Hanushek og Woessmann (2011) og de la Fuente (2011) argumenterer for at måla ein del tidlegare empiriske studiar har brukt på humankapital, mellom anna antal års utdanning i befolkninga, er for snevre og ikkje klarar å fanga opp det reelle nivået på eit lands humankapital. Dei meiner også at det i tidlegare studiar ligg store feilkjelder i målefeil og dårlege datasett. Antal års utdanning seier mykje om kvantiteten på utdanningsnivået i eit land, men lite om kvaliteten på utdanninga, og er difor i seg sjølv ikkje eit tilfredsstillande mål på humankapital. Hanushek og Woessmann utvidar humankapitalbegrepet til også å inkludera fleire mål, mellom anna testresultat på internasjonale testar som PISA-undersøkinga. I deira studie finn dei ein klar samanheng mellom humankapital og økonomisk vekst. Når ein ser på regionale skilnader i Noreg, og ikkje skilnader på tvers av land, vil mykje av problema knytt til måling av humankapital forsvinna. Det er naturleg å anta at kvaliteten på det norske utdanningssystemet er rimeleg jamn på tvers av regionane. Det norske utdanningssystemet er bygd opp på same måte i heile landet, og alle offentlege skular baserer seg på dei same læreplanane og læremåla. Dermed vil alle som fullfører den same vidaregåande utdanninga, uavhengig av om det er i Finnmark eller Vest-Agder, oppnå tilnærma den same kompetansen. Andelen i ein region som fullfører den vidaregåande opplæringa bør dermed kunne vera eit rimeleg mål på regionens humankapital.

Vandenbussche et al. (2006) argumenterer, som vist i førre delkapittel, for at det er andelen med høgare utdanning som er avgjerande for økonomisk vekst i relativt rike økonomiar. I ein empirisk studie av spanske regionar finn Manca (2012) ein motsatt samanheng. Han finn at andelen som tek sekundær opplæring (i Noreg: vidaregåande opplæring) vert viktigare jo betre utvikla ein region i utgangspunktet er. Forklaringa hans på denne observasjonen er at regionar som ligg langt framme i den økonomiske utviklinga treng stor grad av ulike utdanningsnivå som utfyller kvarandres kompetanse. Manca argumenterer for at det er viktig at mange fullfører sekundær opplæring for å fullt ut kunne nytta kompetansen til dei i regionen som har teke høgare utdanning.

Glaeser og Saiz (2004) studerer samanhengen mellom utdanningsnivå og befolkningsvekst for byar i USA og finn ein positiv korrelasjon. Hovudforklaringsvariabelen i artikkelen deira er befolkningsandelen med høgare utdanning (minst college-grad). Dei meiner humankapitalen først og fremst påverkar befolkningsvekst på to ulike måtar, enten ved at høg humankapital gjev produktivitetsvekst, eller ved at større humankapital vert sett på som noko som i seg sjølv gjer eit område attraktivt å busetja seg i. I enkelte av regresjonane deira vert andelen som ikkje fullfører det som i USA tilsvarar vidaregåande opplæring, andelen high school-dropouts, brukt som kontrollvariabel. For andelen high school-dropouts finn dei ingen signifikant korrelasjon med befolkningsvekst for storbyområda (metropolitan areas), men dei finn ein signifikant negativ korrelasjon mellom befolkningsvekst og high school-dropouts på bynivå (city). For storbyområda er det andelen med college-utdanning

som betyr noko for befolkningsveksten. Forklaringa deira er i stor grad at ein høg andel som ikkje har fullført vidaregåande gjev stor arbeidsløyse. Dette gjev vidare andre negative sosiale konsekvensar som stor fattigdom og mykje kriminalitet. I sum gjer dette ein bydel mindre attraktiv å flytta til og påverkar dermed folketalsutviklinga negativt. Dette har mindre å seia for storbyområde, sidan det kan vera store variasjonar innan eit slikt område. For eit slikt område under eitt er det andelen med høgare utdanning som driv produktivitetsveksten, og på den måten skapar befolkningsvekst.

Som tidlegare nemnt er det godt dokumentert at utdanningsnivå og inntekt heng saman hos enkeltmenneske. Dei som ikkje fullfører vidaregåande opplæring er også overrepresentert på fleire negative statistikkar, som kriminalitetsstatistikkar, arbeidsledighetsstatistikkar og statistikkar for kven som mottak økonomiske stønader og trygder. For Noreg er dette mellom anna vist i Falch et al. (2010). Å måla dei kausale effektane ved fråfall, både inntektseffektar og den totale samfunnsøkonomiske effekten, er nødvendigvis vanskeleg i praksis. I mange tilfelle vil det vera vanskeleg å veta om det er fråfallet i seg sjølv eller andre faktorar som er hovudgrunnen til at det er ein korrelasjon mellom fråfall og til dømes stønadsmottak. Eit døme kan vera at personar med psykiske lidningar vil ha både lågare sannsynlighet til å fullføra vidaregåande og større sannsynlighet for å falla utanfor arbeidsmarknaden. I slike tilfelle er det enkelt å finna ein samheng mellom fullføring og offentlege stønader, sjølv om det i røynda kan vera dei psykiske lidingane som står bak begge deler.

Falch et al. (2009) argumenterer for at det totale samfunnsøkonomiske sparepotensialet i Noreg av auka fullføring likevel er stort. Dei argumenterer for at den samfunnsøkonomiske vinsten av at fråfallet går ned med ein tredjedel truleg ligg ein plass mellom 1,1 og 8,8 milliardar kroner per årskull. Det er verdt å merka seg at dette er på nasjonalt nivå. Ein god del av desse kostnadene er trygde- og stønadsutgifter som staten dekkjer. Akkurat desse ekstrakostnadene vert difor ikkje regionane direkte ramma av. Samstundes er mykje av det samfunnsøkonomiske tapet reine inntektseffektar, og dette vil påvirka regionane på fleire måtar. Mellom anna vil lågare gjennomsnittleg personinntekt for innbyggjarar i ein region føra til mindre skatteinntekter for kommunane i regionen og fylkeskommunen regionen høyrer til i, noko som alt anna likt fører til eit lågare offentleg tilbod. Dette kan vidare gjera desse regionane mindre attraktive å busetja seg i og motarbeida det regionale utviklingsarbeidet til kommunane og fylkeskommunane.

Som nemnt i innleiinga er nullhypotesen i denne oppgåva at fullføringa ikkje påverkar den regionale utviklinga. Men som vist peikar økonomisk teori på at det er ein positiv samheng mellom økonomisk utvikling og humankapital. Dersom det er slik skulle dette også tyde på at høg fullføringsgrad gjev fleire arbeidsplassar, som igjen kan føra til både befolkningsvekst og høgare inntektsnivå i regionane. Dette kan støtta ein alternativhypotese om at fullføringsgraden påverkar den regionale utviklinga positivt.

Fullføringsgraden kan også vera ein mekanisme i negativ retning ved at folk flyttar ut av ein region etter å ha fullført vidaregåande opplæring. Mange regionar manglar eit høgare utdanningstilbod, og folk må midlertidig flytte ut av desse regionane om ein ønskjer å ta høgare utdanning. Viss desse personane etter endt utdanning ikkje flyttar attende til regionen dei vaks opp i, vil regionar som får mange til å fullføra vidaregåande opplæring bidra til vekst i humankapitalen i andre regionar i staden for å få ein auke i eigen human-kapital. Eit anna argument for nullhypotesen er teorien i Vandenbussche et al. (2006), det er andelen med høgare utdanning som reelt har noko å seia. Sjølv om det er økonomiske variasjonar mellom norske regionar er det rimeleg å skildra alle regionane som godt økonomisk utvikla. Dersom vekst må basera seg på innovasjon og ikkje imitasjon, vil det vera andelen med høgare utdanning i regionen som er relevant for utviklinga og ikkje andelen som har fullført vidaregåande opplæring.

3 Empirisk tilnærming

I dette kapitlet skal eg gjera greie for dei økonometriske metodane eg brukar i analysane i kapittel 5. Eg kjem til å gå gjennom hovudtrekka i dei estimeringsmetodane som vert brukt for å studera samanhengen mellom fullføringsgrad i vidaregåande opplæring og regional utvikling og kva som skil dei ulike metodane frå kvarandre.

Det er blitt gjort svært få, om nokon, empiriske studiar på samanhengen mellom fullføringsgrad i vidaregåande opplæring og regional utvikling. I estimeringsdelen av denne oppgåva kjem eg difor å prøva ut fleire ulike økonometriske metodar og framgangsmåtar. I tillegg vil eg variera spesifikasjonane i dei ulike estimeringane for å sjå om det har nokon betydning for resultatane.

Dette datasettet er eit typisk døme på eit paneldatasett. Det vil seia at datasettet inneheld observasjonar som har både ein tidsdimensjon og ein tverrsnittsdimensjon. Tverrsnittsdimensjonen vil i denne oppgåva i praksis bety regionane. Alternativa til paneldata vil vera data der ein ser på observasjonar for ein enkelt region, men der ein likevel har ein tidsdimensjon eller observasjonar som varierer på tverrsnittsnivå, men ikkje over tid. I eit paneldatasett ser ein på det same tverrsnittet over ein tidsperiode og ser korleis eigenskapane endrar seg. Det finst fleire ulike økonometriske metodar å bruka når ein skal estimera med eit paneldatasett. I dei neste delkapittela skal eg ta for meg desse. Kjeldene til dette kapitlet er Brooks (2008), Woolridge (2009) og Verbeek (2008). Å læra seg økonometri går i stor grad ut på å forstå når ulike metodar bør brukast, og i slike tilfeller er det vanskeleg å sitera direkte. I enkelte konkrete tilfelle rundt til dømes spesifikke likningar kjem eg likevel til å henvisa direkte til ei eller fleire av desse bøkene.

3.1 Minste kvadraters metode

Ein av dei mest grunnleggjande og mest brukte estimeringsteknikkane er minste kvadraters metode (MKM). Minste kvadraters metode er ein lineær estimeringsmetode der målet er å gje ein best mogleg lineær tilnærming til variabelen ein ønskjer å estimera. I metoden prøver ein å minimera skilnadene mellom den faktiske verdien på variabelen ein ønskjer å forklara, og den estimerte verdien på den same variabelen.

Anta at ein ønskjer å estimera β_0 og β_1 i likninga

$$y_{it} = \beta_0 + \beta_1 x_{it} \tag{1}$$

Ved å nytta MKM i estimeringa vil β_0 og β_1 bli bestemt slik at

$$\sum_{i=1}^n \sum_{t=1}^T (y_{it} - \hat{\beta}_0 - \hat{\beta}_1 x_{it})^2 \quad (2)$$

blir minimert. På denne måten forsøker ein å få ei best mogleg lineær tilnærming til likninga ein ønskjer å estimera. Her er $\hat{\beta}_0$ og $\hat{\beta}_1$ det estimerte konstantleddet og den estimerte parameterverdien til forklaringsvariabel x_{it} . y_{it} viser faktisk verdi på variabelen som skal forklarast i region i og år t . \hat{y}_{it} er den estimerte verdien for den same variabelen. Den er definert ved $\hat{y}_{it} = \hat{\beta}_0 + \hat{\beta}_1 x_{it}$

Likninga inni parentesen i (2) kan derfor skrivast som

$$\hat{u}_{it} = y_{it} - \hat{y}_{it} \quad (3)$$

Målet er at variasjonen til u_{it} vert minimert. u_{it} vert ofte kalla residualen, eller restleddet i estimeringa. Residualen viser, som likningane ovanfor seier, kor stor skilnad det er på den reelle y -verdien og den estimerte y -verdien. Ut frå likning (3) ser ein også at ein positiv \hat{u}_{it} viser at y er underestimert, mens ein negativ \hat{u}_{it} viser at y er overestimert. Forventningsverdien til \hat{u}_{it} skal i ein MKM-estimering vera null, dette vil i praksis seia at gjennomsnittsverdien til \hat{u}_{it} over alle observasjonar er null.

Forklaringskrafta til ein estimering vert ofte målt ved verdien R^2 . Denne verdien er definert ved (sjå til dømes Woolridge (2009), side 81)

$$\frac{\sum_{i=1}^n \sum_{t=1}^T (\hat{y}_{it} - \bar{y})^2}{\sum_{i=1}^n \sum_{t=1}^T (y_{it} - \bar{y})^2}, \quad (4)$$

forklart variasjon i datasettet delt på dei totale variasjonane i datasettet. \bar{y} er gjennomsnittet av observasjonane av variabelen som skal forklarast. Dersom estimert verdi og faktisk verdi er lik for alle observasjonar vil teljaren og nemnaren vera like, og R^2 blir 1. R^2 ligg alltid mellom 0 og 1. Dersom estimeringa vert dårlegare og forklaringskrafta går ned vil R^2 gå mot 0. Ein R^2 -verdi på 0 betyr i praksis at likninga ein estimerer har inga forklaringskraft. Ved bruk av ein paneldatateknikk som estimerer tilfeldige effektar, forklart i delkapittel 3.3, vil ikkje R^2 som mål på forklaringskrafta vera samanliknbart med den verdien for R^2 ein oppnår ved hjelp av MKM-estimering. For å samanlikna estimeringane mellom dei to estimeringsmetodane må ein difor bruka andre målemetodar.

Ein svært viktig føresetnad for MKM er at restleddet, u_{it} , er uavhengig av alle forklaringsvariablane. Viss dette ikkje er tilfellet vil ein få skeive estimat. Slik korrelasjon mellom restleddet og ein eller fleire av forklaringsvariablane kan skuldast at det er variablar ein ikkje har tatt med i estimeringa, men som påvirkar estimeringsresultatet, eit såkalla ute-

latt variabel-problem. Ein måte å løysa dette på er ved å inkludera fleire variablar ein trur kan bidra til å forklara venstresidevariabelen i estimeringa. Om ein mistenkjer at det er faste effektar på tverrsnittsnivå som spelar inn kan ein bruka ei estimering med faste effektar, grundigare diskutert i kapittel 3.2. Ein anna vanleg måte å omgå dette problemet på er å bruka ein såkalla instrumentvariabelestimering. Eg kjem ikkje til å gå inn på den metoden i denne oppgåva, men kort fortalt består den i at ein brukar ein anna variabel, eit instrument, som er ukorrelert med restleddet til å predikera forklaringsvariabelen, og deretter estimerer den originale likninga.

I praktisk økonometri er det eit vanleg problem at to variablar påverkar kvarandre begge vegar og simultant. Dersom dette er tilfellet vil ein lett kunna trekkja feilaktige konklusjonar ved bruk av vanlege estimeringsmetodar. Dette kallast eit direkte simultanitetsproblem. Eit slikt problem kan til dømes løysast ved hjelp av instrumentvariabelestimering eller å tilbakedatera forklaringsvariabelen ein mistenkjer har ein tovegs-korrelasjon med venstresidevariabelen.

Homoskedastisitet ein viktig forutsetning for at MKM skal gje effisiente estimatorar. Homoskedastisitet vil seia at variansen til restleddet er konstant på tvers av alle individ. Om dette ikkje er oppfylt og variansen til restleddet varierer, har ein heteroskedastisitet. Dette kan justerast for i dei fleste estimeringsprogram, og dette kjem eg til å gjera når resultatata skal estimerast. Også dersom restledda er seriekorrelerte vil effisiensen til MKM-estimeringa bli påverka negativt. Seriekorrelerte restledd gjer at forventningsverdien til residualen ikkje er lik null.

Sjølv om MKM-metoden er ein av dei mest grunnleggjande og mest brukte metodane innanfor økonometri, har metoden likevel som nemnt ovanfor ein del potensielle svakheitane. Desse kan gjera MKM ueigna i ein del tilfeller. Difor er det også utvikla ei rekke andre estimeringsmetodar der ein forsøker å ta omsyn til og forbedra svakheitane i MKM-estimera.

3.2 Estimering med faste regioneffektar

Eit praktisk problem som fort kan dukka opp i ei paneldataestimering er at regionane har eigenskapar som ikkje er observerbare ut frå datasettet, men som er korrelerte med nokre av forklaringsvariablane. Dersom dette er tilfellet vil ein estimering med vanleg MKM-metode gje skeive estimat. Restleddet i estimeringa vil fanga opp dette, og dermed bli korrelert med dei aktuelle forklaringsvariablane. Som tidlegare nemnt er ein viktig forutsetning i MKM-estimering nettopp at restleddet er uavhengig av alle forklaringsvariablane, og dermed vil estimeringar i slike tilfelle gje skeive estimat, noko ein sjølv sagt bør forsøka å unngå. Eit eksempel på noko som kan oppstå i denne oppgåva er til dømes at

enkelte regionar med høg fullføringsgrad også kan ha stor inntektsvekst, men det kan vera uobserverbare variablar som driv både den høge fullføringsgraden og den store inntektsveksten i desse regionane. Tilsvarande kan det vera at regionar med låg fullføringsgrad også har låg økonomisk vekst, utan at det nødvendigvis er fullføringen som er grunnen til dette. I desse tilfella vil ein få skeive estimat ved bruk av vanleg MKM, fordi effekten av fullføringsgraden vil bli overvurdert. Motsatt vil det vera mogleg å tenkja seg at effekten av fullføringsgraden på økonomisk vekst også kan undervurderast. I slike tilfelle er det hensiktsmessig å nytta ein type estimering der ein tek omsyn til strukturelle skilnader mellom regionane.

Ein slik type estimering kallast estimering med faste effektar. Med ein slik metode vil alle forhold som er spesifikke for ein region, men som ikkje varierer over tid, bli transformert vekk. På den måten vil det i mange samanhengar vera mogleg å få eit betre bilde av dei kausale effektane av variablane ein ønskjer å studera, utan at uobserverbare effektar som er spesifikke for enkelte regionar spelar inn og fører til skeive estimat. Ein kan også løyse dette problemet ved å innføra dummyvariablar for alle regionane i estimeringa, men ofte vil det i praksis vera enklare å transformera vekk dei tidsuavhengige effektane.

Ein antek først at restleddet u_{it} består av to ulike deler. Den eine delen er ein tidsuavhengig del, a_i , som varierer berre på tvers av regionar. Den andre delen er ein restleddsdel som også varierer over tid, v_{it} . Delen som varierer over tid vert også kalla eit idiosynkratisk restledd. For kjelde, sjå til dømes Woolridge (2009) s.481-483. Anta samme modell som tidlegare:

$$y_{it} = \beta_1 x_{it} + a_i + v_{it} \quad (5)$$

Slår ein det idiosynkratiske og det tidsuavhengige restleddet saman vil ein få at $u_{it} = a_i + v_{it}$. Tek gjennomsnittet over tid for kvar region i og får

$$\bar{y}_i = \beta_1 \bar{x}_i + a_i + \bar{v}_i \quad (6)$$

Subtraherer (6) frå (5) og får

$$y_{it} - \bar{y}_i = \beta_1 (x_{it} - \bar{x}_i) + v_{it} - \bar{v}_i \quad (7)$$

Alternativt kan dette skrivast som

$$\ddot{y}_{it} = \beta_1 \ddot{x}_{it} + \ddot{v}_{it} \quad (8)$$

Det viktigaste som har skjedd her er at restleddet som berre varierer over tverrsnittet er transformert vekk. Dette gjer i praksis at alle uobserverbare variablar som ikkje varierer

over tid vert teke vekk i estimeringa. På denne måten er det mogleg å studera effektane av dei variablane av interesse, utan at tidsinvariate, uobserverbare regionspesifikke variablar ein ikkje kan måla ut frå datasettet gjev skeive estimat. Den nye likninga, (7), vert estimert med hjelp av MKM-metoden. På same måte som med vanleg MKM er det ein føresetnad at restleddet er ukorrelert med forklaringsvariablane, det idiosynkratiske restleddet må difor uansett vera uavhengig av forklaringsvariablane brukt i estimeringa.

Ein vanleg måte å testa om ein estimering med faste effektar er hensiktsmessig, er ved den såkalla Hausman-testen. Den vert forklart i kapittel 3.4.3.

3.3 Estimering med tilfeldige effektar

Ein mellomting mellom ein vanleg MKM-estimering og ein estimering med faste region-effektar er det som kallast estimering med tilfeldige effektar. Denne typen estimering brukast i staden for ei estimering med faste effektar viss ein trur dei uobserverte effektane er ukorrelerte med forklaringsvariablane. Dersom dette er tilfellet vil ein estimering der ein transformerer vekk tidsuavhengige effektar ikkje bli effisient, sidan desse effektane då vil bidra til informasjon i estimeringa utan at det gjev skeive estimat. Ved å transformera desse vekk vil ein dermed få til eit dårlegare estimat enn om dei vert tekne med. Estimering med tilfeldige effektar vert brukt i staden for MKM dersom ein har eit tidsuavhengig restledd i estimeringa som har ein varians ulik null. Eit slikt ledd vil vera uavhengig av tid, men avhengig av individ, og vil difor gje seriekorrelerte restledd, som skildra i kapittel 4.1. Ein måte å omgå dette problemet er ved å bruka ein estimering med tilfeldige effektar. Ein slik estimering vert også kalla generalisert minste kvadraters metode (GLS).

For å gjera ein slik estimering må ein først definera ein variabel θ . Denne er gitt ved (Brooks (2008) s. 499)

$$\theta = 1 - \frac{\sigma_v}{\sqrt{T\sigma_a^2 + \sigma_v^2}} \quad (9)$$

Her er σ_a^2 variansen til det tidsuavhengige restleddet, mens σ_v^2 er variansen til det idiosynkratiske restleddet. T er antal år i datasettet.

På samme måte som med ein estimering med faste effektar blir resultatet ein transformert modell, her lik

$$y_{it} - \theta \bar{y}_i = \beta_1(x_{it} - \theta \bar{x}_i) + u_{it} - \theta \bar{u}_i \quad (10)$$

Denne likninga kan også skrivast som

$$y'_{it} = \beta_1 x'_{it} + u'_{it} \quad (11)$$

Det som kan vera verdt å merka seg her er at dersom θ er lik 0 vil ein i likning (10) få ein vanleg MKM-estimering, mens dersom θ er lik 1 vil ein få ei estimering med faste effektar. Her ser ein altså matematisk at ei estimering med tilfeldige effektar i praksis er ei mellomløysing mellom dei to estimeringsteknikkane. Med ein slik estimering er det ein heilt nødvendig forutsetning at det tidsuavhengige restleddet er uavhengig av forklaringsvariablane og har forventningsverdi null. Dersom restleddet ikkje er det, vil ein få skeive estimat ved bruk av denne typen estimering, og ein bør heller nytta estimering med faste effektar.

3.4 Testmetodar

Det er mange ulike metodar for å testa resultatane av ei estimering. To av dei mest sentrale testmetodane er t-testing og F-testing. Desse er i utgangspunktet velkjende, men eg kjem til å gå kort gjennom hovudtanken bak dei. I tillegg kjem eg til å bruka den såkalla Hausman-testen. Den vert brukt for å samanlikna estimering med faste effektar og estimering med tilfeldige effektar.

3.4.1 t-test

t-testing vert gjort for å bestemma om ein parameterverdi er signifikant ulik ein viss verdi. I praksis testar ein som oftast om parameterverdien er signifikant ulik null. I praksis vil dei seia om ein variabel har ein signifikant påverknadskraft, eller om påverknaden i praksis er lik null.

Den generelle teststatistikken som gjev t-verdien er gitt ved

$$t = \frac{\hat{\beta}_j - \beta_j}{se(\hat{\beta})} \quad (12)$$

$se(\hat{\beta})$ er standardfeilen til den estimerte verdien.

Med ei nullhypotesen om at effekten av ein variabel er lik null, altså $H_0 : \beta_j = 0$, vil ein få at $t = \frac{\hat{\beta}_j}{se(\hat{\beta})}$.

Det er to måtar å utføra ein t-test på. Ein kan ta ein såkalla ein-sidig test, altså at alternativhypotesen er enten $H_1 : \beta_j < 0$ eller $H_1 : \beta_j > 0$, eller ein kan ta ein to-sidig test. Ved ein to-sidig test er alternativhypotesen $H_1 : \beta_j \neq 0$. Framgangsmåten er lik i desse to metodane, skilnaden er at den kritiske verdien vert noko høgare med ein to-sidig test. Signifikansnivåa oppgitt i resultatkapittelet er basert på ei to-sidig testing.

3.4.2 F-test

F-testing kan minna om t-testing, men her testar ein om effekten av fleire variablar er signifikant ulik ein viss verdi, som oftast null. Nullhypotesen er at fleire parameterverdier alle er lik denne verdien. Matematisk blir dette, til dømes med tre variablar som skal testast: $H_0 : \beta_j = \beta_k = \beta_l = 0$. Nullhypotesen seier at dei tre tilhøyrande variablane til desse parameterverdiane ikkje har nokon effekt på variabelen som skal forklarast, og dermed bør utelatast frå estimeringa. Alternativhypotesen er at $H_1 : H_0$ er feil. Dersom alternativhypotesen stemmer betyr det at minst ein av variablane som har blitt testa har ein signifikant påverknadskraft på variabelen som skal forklarast.

Teststatistikken for F-testing er definert ved (Woolridge (2009) s.145)

$$F \equiv \frac{(SSR_r - SSR_{ur})/q}{SSR_{ur}/(n - k - 1)} \quad (13)$$

SSR_r er summen av dei kvadrerte residualane i modellen med restriksjonar, altså ein modell der ein har teke vekk variablane ein ønskjer å testa for. SSR_{ur} er summen av dei kvadrerte residualane i modellen utan restriksjonar. Ved å fjerna variablar vil truleg denne summen gå opp, den kan umogleg gå ned og vil i beste fall vera uendra. I likning (13) er q antal variablar droppa i modellen med restriksjonar, n er antal observasjonar og k antal variablar i den opprinnelege estimeringa. Nullhypotesen vert avvist dersom F-verdien estimert ut frå (13) er større enn ein kritisk verdi som ein finn ved hjelp av ein F-fordeling og (q,n-k-1) frihetsgrader.

3.4.3 Hausman-testing

I ein Hausman-test samanliknar ein estimering med faste effektar og estimering med tilfeldige effektar (sjå til dømes Brooks (2008) s.273-274 eller Verbeek (2008) s. 368-369). Nullhypotesen er at det tidsuavhengige restleddet er uavhengig av forklaringsvariablane brukt i modellen. I dette tilfellet vil estimering med faste effektar og estimering med tilfeldige effektar gje tilnærma like parameterverdier, sidan ingen av metodane gjev skeive estimat. Om det skjer vil estimering med tilfeldige effektar vera å føretrekkja grunna effisiens. Alternativhypotesen vil nødvendigvis vera at dette restleddet ikkje er uavhengig av forklaringsvariablane. Då vil ein kunna sjå ein systematisk skilnad på parameterverdiane i dei to estimeringsmetodane og ei estimering med faste effektar vil vera å føretrekkja. Det Hausman-testen i praksis går ut på er difor å samanlikna parameterverdiane frå desse to estimeringsmetodane, og ut frå dette tolka kva metode som er mest hensiktsmessig. Det ein testar er om $\hat{\beta}_{FE} - \hat{\beta}_{RE} = 0$, som vil vera tilfellet dersom nullhypotesen stemmer.

Hausman-teststatistikken er gitt ved (Verbeek s.368)

$$H = \frac{(\hat{\beta}_{FE} - \hat{\beta}_{RE})'(\hat{\beta}_{FE} - \hat{\beta}_{RE})}{[\hat{V}(\hat{\beta}_{FE}) - \hat{V}(\hat{\beta}_{RE})]} \quad (14)$$

\hat{V} viser eit estimat av kovariansmatrisene. Ein avviser nullhypotesen dersom den estimate teststatistikken er større enn ein grenseverdi basert på ein χ^2 -fordeling med antal frihetsgrader basert på antal observasjonar.

4 Datamateriale, deskriptiv statistikk og empirisk spesifikasjon

Store deler av datamaterialet til denne oppgåva er henta frå kommunedatabasen til Norsk samfunnsvitenskapelig database (NSD).

Sidan det ikkje finst tal på bruttonasjonalprodukt for norske regionar må eg bruka andre indikatorar for regional utvikling i denne oppgåva. Eg har valt å nytta vekst i inntekt per innbyggjar og folketalsutvikling som mål på regional utvikling. Dette er variablar som er lett identifiserbare og enkle å finna data for, og begge deler kan sjåast på som mål på den økonomiske situasjonen i ein region. Å ha størst mogleg inntektsvekst for innbyggjarane vil alltid vera eit naturleg mål for ein region. Samstundes slit mange norske kommunar og regionar med synkande folketal. I desse kommunane og regionane vert difor nettopp folkevekst, eller kampen mot fråflytting, vurdert som ei av dei viktigaste oppgåvene i åra framover. Grunnane til at enkelte område slit med folketalet er truleg fleire, men tendensen er at byområda i snitt har vesentleg høgare folkevekst enn distrikta. Dette bør også sjåast i samheng med økonomiske faktorar, det er truleg at folkevekst heng saman med tilgangen på attraktive jobbar.

Tabell 6 i appendiks A.1 gjev ei nærare skildring av tala for kvar enkelt region. Tala og figraune presentert i dei neste delkapittela er meint som illustrasjonar og oversiktsbilete for korleis situasjonen er totalt sett.

I kapittel 4.1 definerer eg regionnivået eg estimerer på. I 4.2 presenter eg fullføringsstatistikken. 4.3 og 4.4 vert brukt til å presentera utviklingsmåla, 4.5 kontrollvariablane og 4.6 den empiriske spesifikasjonen av modellane eg brukar i estimeringane.

4.1 Definisjon av regionar

Ei praktisk utfordring ved å studera norske regionar er å finna rett regionnivå å estimera på. Mange norske kommunar har ikkje ein vidaregåande skule lokalisert i kommunen, slik at ved å bruka kommunar ville det truleg bli vanskeleg å finna nokon kausale effektar av fråfall. Samstundes er det store geografiske, demografiske og økonomiske variasjonar innad i enkelte fylke, slik at ved å bruka ei fylkesinndeling vil ein risikera å gå glipp av mykje informasjon.

Løysinga eg har valt er å bruka ei regional inndeling foreslått av Statistisk sentralbyrå i Hustoft et al. (1999). Denne baserar seg på NUTS-inndelinga definert av EU. NUTS, Nomenclature des Unités Territoriales Statistiques, er EUs standardardindeling for regionar. Hustoft et al. (1999) definerer norske regionar som tilsvarar NUTS-4, dette vil

i Noreg bety ei inndeling som ligg mellom fylkesnivå og kommunenivå. SSB sjølv kallar inndelingssystemet REGIN. Skalaen går frå 1 til 5, der REGIN-1 er heile landet og REGIN-5 er kommunane. Regioninndelinga er laga på bakgrunn av kva kommunar SSB meiner heng naturleg saman i ein region, basert på mellom anna pendlarstatistikkar. Dette betyr at regionane laga av SSB i praksis viser dei regionale arbeidsmarknadsområda. Sidan arbeidsmarknadssituasjonen heng tett saman med den økonomiske utviklinga, vil det vera gunstig å visa den økonomiske utviklinga for område som tilsvarar korleis arbeidsmarknaden i praksis er inndelt. Med regioninndelinga eg har valt i denne oppgåva består Noreg av 90 økonomiske regionar.

Eit av kriteriana SSB har satt er at kommunane må vera i same fylke. I enkelte tilfelle gjer dette at kommunar som i praksis naturleg høyrer saman, ikkje kjem i same region. Dette kan vera litt uheldig i enkelte tilfelle. Samstundes blir den vidaregåande opplæringa administrert av fylkeskommunane, og norske elevar gjennomfører difor i svært stor grad den vidaregåande opplæringa i heimfylket sitt. Unntaka er landsdekkjande linjer, som til dømes toppidrettslinjer og nokre særavtalar som finst i enkelte grensekommunar. Dette gjeld likevel ein svært liten andel av norske elevar.

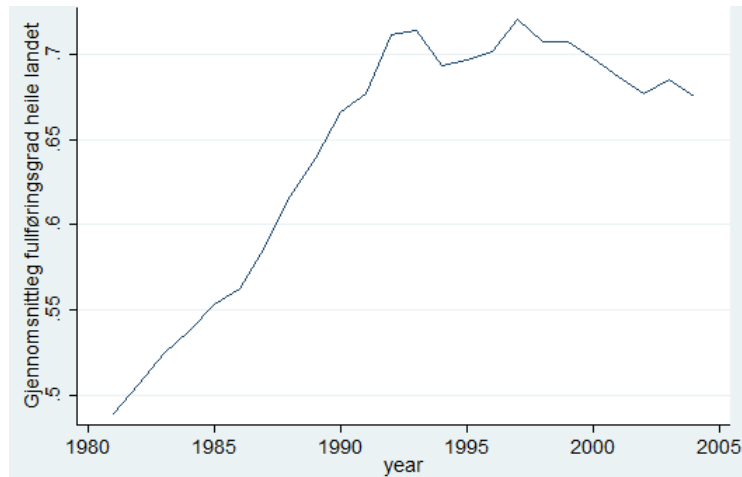
4.2 Fråfallsstatistikk

Fullføringsstatistikkane er knytta til og henta frå prosjektet «Governance, management and performance in the Norwegian educational system» ved Senter for økonomisk forskning, finansiert av Norges Forskningsråd.

Definisjonen på fullføring i dette datasettet er andelen i ein region som har fullført vidaregåande opplæring fem år etter at dei gjekk ut av ungdomsskulen. Datasettet tek altså ikkje omsyn til om fullføringa har skjedd på normert tid eller ikkje, så lenge ein elev har fullført i løpet av fem år. Difor vil det heller ikkje bli teke omsyn til dei som har fullført vidaregåande opplæring seinare enn fem år etter at dei gjekk ut av grunnskulen, desse vil bli registrert som at dei ikkje har fullført i denne oppgåva. Årstala for fullføring i datasettet viser til det året ein elev gjekk ut av grunnskulen. Datasettet byrjar med dei som var ferdig med grunnskulen i 1981. Difor går datamaterialet for fullføringsandelar til 2004.

Skilnaden i landsgjennomsnittet på fullføringsgraden frå år til år vert illustrert i figur 1. I kapittel A.2, tabell 7, finst det meir utfyllande tal for kvart enkelt år.

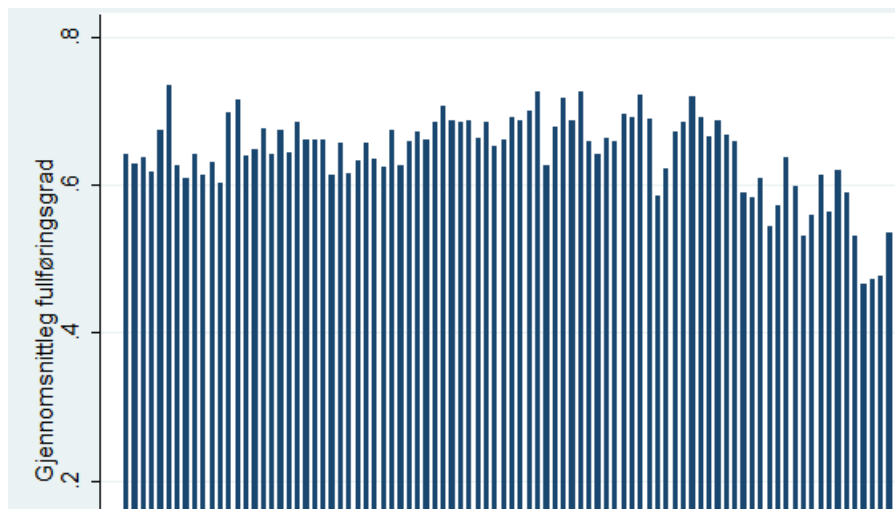
Figur 1 viser at det var ein jamn vekst i den gjennomsnittlege fullføringsgraden frå 1981 til 1992. Etter dette har fullføringsgraden svinga litt opp og ned, med ein svak nedgang frå 1997 til 2004. Appendiks 2, tabell 7, viser at standardavviket for det årlege landsgjennomsnittet har halde seg rimeleg stabilt gjennom heile tidsperioden. Dette kan tolkast i retning av at endringane i fullføringsgraden har skjedd jamnt over heile landet, og at det



Figur 1: Gjennomsnittleg fullføringsgrad sortert på år

dermed ikkje er enkeltregionar som driv endringane i gjennomsnittleg fullføringsgrad.

Variasjonen i gjennomsnittleg fullføringsgrad mellom regionane kan også illustrerast i ein figur, figur 2.



Figur 2: Gjennomsnittleg fullføringsgrad i dei ulike regionane

Som ein ser av figur 2 er det ganske store skilnader i fullføringsgrad mellom dei ulike regionane. Søylen er ikkje namngjevne, men region 1 ligg heilt til venstre, og så følgjer dei andre regionane i stigande rekkefølge, med region 90 til slutt. Regionane er nummerert geografisk, slik at region nr. 1 er Halden og region nr. 90 er Kirkenes.

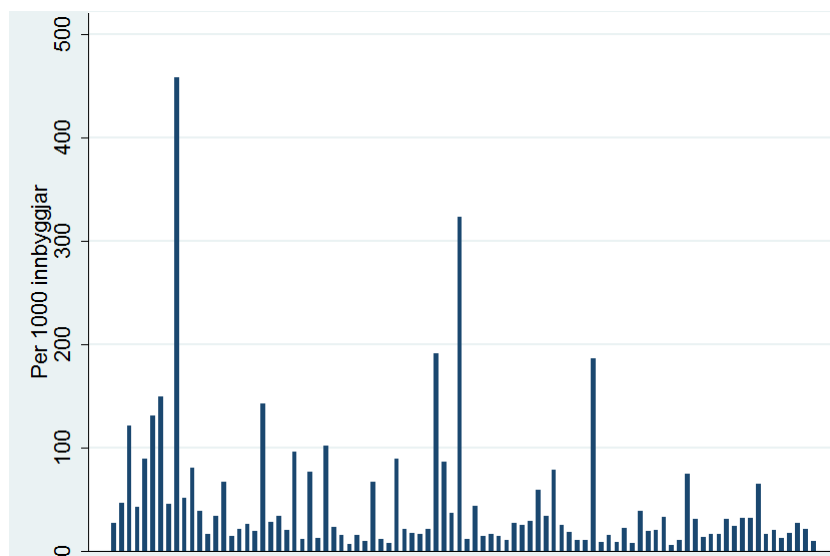
Ut frå figur 2 er det ganske enkelt å sjå at den gjennomsnittlege fullføringsgraden i snitt er vesentleg lågare i Nord-Noreg, og då særleg i enkelte regionar i Finnmark. Dei tre regionane med klart lågast fullføringsgrad gjennom heile datasettet er region 87,88 og 89 (Vadsø, Alta og Hammerfest) som alle ligg i Finnmark. I snitt har region 6, Bærum og Asker, den høgaste fullføringsgraden.

Noko denne figuren ikkje fangar opp er at det vil vera store variasjonar i fullføringsgrad også innad i regionane frå år til år, slik at det faktiske fråfallet i enkeltår kan variera mykje frå det totale gjennomsnittet i ein region.

4.3 Folketalsstatistikk

Folketalsutviklinga er eit av to mål på regional utvikling i denne oppgåva. Data for folketal har blitt funne ved å ta ut data for kommunanes folketal, og deretter aggregere desse tala opp til regionsnivå. Tala for folketal er henta frå NSDs kommunedatabase. Folketala er målt per 1. januar kvart år. Den minste regionen målt i folketal gjennom heile datasettet er region 72, Grong. Denne regionen består av kommunane Grong, Lierne, Røyrvik og Namsskogan, med Grong som eit sentrum i regionen. Den største regionen gjennom heile perioden er region 9, Oslo, som utelukkande består av Oslo kommune.

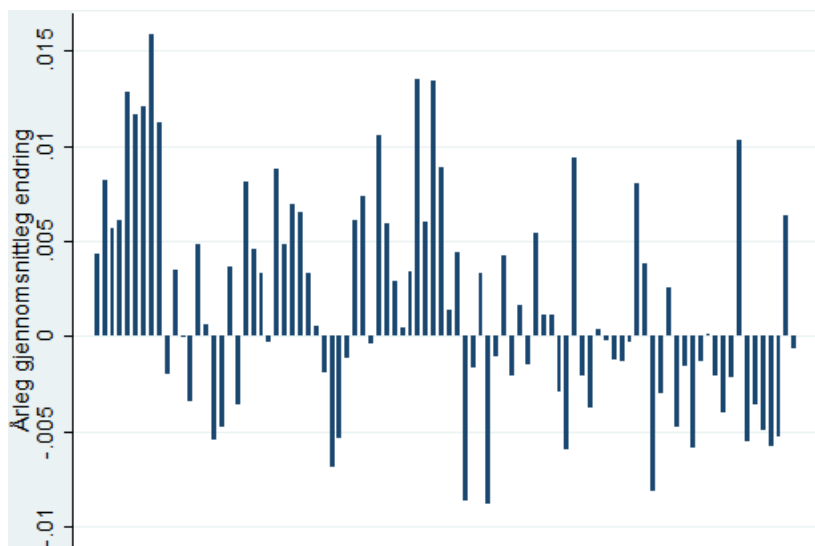
Skilnadene i folketal mellom dei ulike regionane er illustrert grafisk i figur 3.



Figur 3: Folketalet i regionane 1.1.1990, målt i tusen innbyggjarar

Det som er mest openbart ut frå figur 3 er at det er svært store skilnader i folketalet mellom dei ulike regionane. Som ein ser har dei største regionane fleire hundre tusen innbyggjarar, mens dei fleste regionane i 1990 hadde godt under 50 000 innbyggjarar. Gjennomsnittsregionen hadde i 1990 om lag 47 000 innbyggjarar, medan medianregionen hadde om lag 25 000 innbyggjarar. Dette fortel det samme som figuren viser, Noreg har mange små regionar og nokre få relativt store som dreg opp snittet. Hadde ein laga ein tilsvarende figur for 2008 ville bildet vore ganske likt.

Også folketalsutviklinga i regionane kan illustrerast grafisk.



Figur 4: Årleg gjennomsnittleg folketalsendring frå og med 1991 til og med 2008 i dei ulike regionane

Figur 4 illustrerer først og fremst at regionane har ulike utfordringar knytta til folketalsutviklinga. Enkelte regionar er i sterk vekst, medan andre har i perioden 1991 til 2008 hatt stor folketalsnedgang. Regionen med størst folketalsvekst i perioden er region 8, Ulensaker/Eidsvoll, med ein årleg gjennomsnittleg endring på 1,59 prosent. Regionen med størst folketalsnedgang er region 51, Høyanger, med ein årleg gjennomsnittleg endring på -0,88 prosent. Ut frå figuren kan ein sjå at det er ein geografisk trend på på folketalsutviklinga i regionane. Dei ni første regionane, altså alle regionane i Østfold, Akershus og Oslo, har positiv gjennomsnittleg folketalsendring. Samstundes har eit stort fleirtal av regionane i Nord-Noreg i samme periode hatt negativ folketalsnedgang.

4.4 Inntektsstatistikk

Det andre målet på regional utvikling som eg brukar i denne oppgåva er vekst i inntekt per innbyggjar. Inntektstala er også henta frå NSDs kommunedatabase.

Inntekt per innbyggjar er rekna ut ved å først finna den samla inntekta for personlege skatteytarar i kvar kommune. Desse tala er aggregert opp til regionnivå, og deretter rekna ut i gjennomsnitt per innbyggjar i regionane. Alle inntektstal er i nominelle verdar.

Spesifikasjonane for inntekt har endra seg noko gjennom datasettet. Frå 1981 til 1995 brukar NSD skattbar inntekt for skattepliktige forskotsytarar. I 1996 og 1997 brukar dei nemninga allminneleg inntekt for skattepliktige forskotsytarar. Frå 1998 til 2001 oppgjev NSD allminneleg inntekt pluss særfradrag for personlege skatteytarar, mens etter 2002 er det allminneleg inntekt for personlege skatteytarar som vert rapportert. Dette har ført til enkelte store årlege svingingar i regionanes inntektsmål, men svingingane vert justert for i estimeringane.

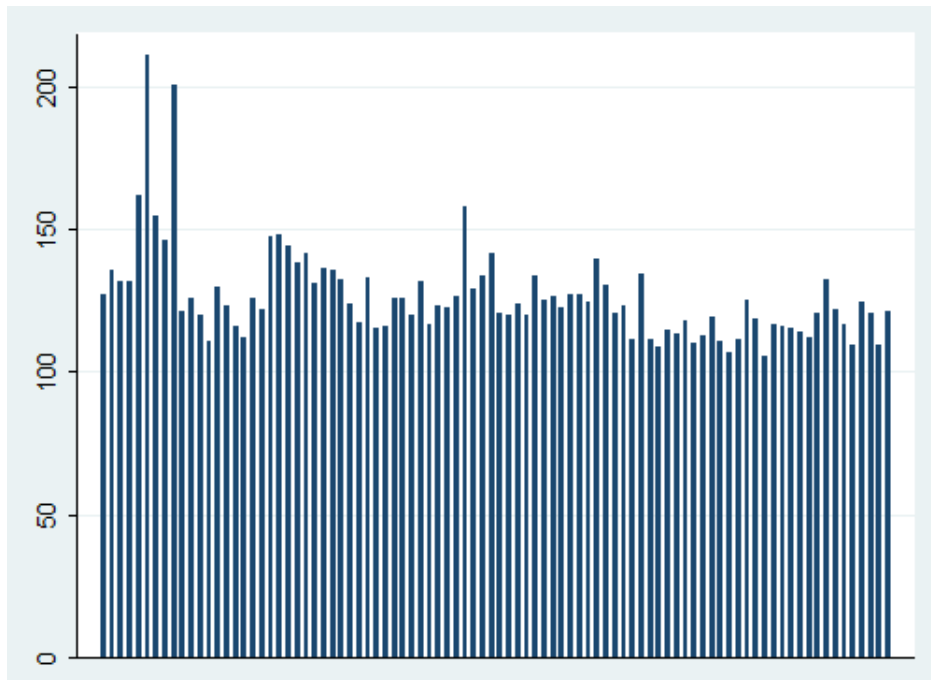
I utgangspunktet var planen å bruka antal personlege skatteytarar for å finna inntekt per skatteytar, men desse tala var ikkje heilt konsekvente. Mellom anna hadde alle regionar eit kraftig fall i antal skatteytarar i 2005, utan at spesifikasjonane av datasettet gav nokon logisk forklaring. I tillegg virka nokre av tala på skatteytarar for ein del regionar kunstig høge. For fleire regionar i Nord-Noreg var det enkelte år registrert fleire personlege skatteytarar enn innbyggjarar. Eg vel difor heller å bruka reine folketalsstatistikkar i utrekninga av gjennomsnittsinntekt.

Å bruka folketal i staden for antal skatteytarar kan ha enkelte bakdelar. Mellom anna vil demografiske skilnader mellom regionane kunne slå ut når inntekt per innbyggjar skal rekna ut. For regionar med ein stor andel innbyggjarar som ikkje er ein del av arbeidslivet, til dømes mange eldre eller unge, eller mange som mottek trygd, vil gjennomsnittsinntekta blir relativt låg. Dette skjer fordi det då vil vera færre som har lønsinntekt, utan at løna for dei som faktisk er i arbeidslivet nødvendigvis er lågare enn i andre regionar. For regionar med ein stor andel yrkesaktive vil dette slå ut på den måten at gjennomsnittsinntekta blir relativt høgare enn andre regionar. Samstundes er det eit mål for alle regionar i landet å ha flest mogleg i arbeid. Slik sett vil bruk av folketal vera hensiktsmessig. Viss ein større andel enn normalt står utanfor arbeidslivet i ein region, vil dette bli fanga opp i variabelen inntekt per innbyggjar. Som eg skal visa nærare i kapittel 4.5 kjem eg også til å bruka andelen i arbeidsfør alder som ein kontrollvariabel i regresjonen, slik at alderssamansetninga blir justert for i estimeringane.

Det manglar ein del observasjonar for inntekt enkelte år. Dette har skjedd fordi desse tala manglar i NSDs database. Då eg aggregerte inntektstalet opp til regionnivå, gjorde eg det slik at inntekt ikkje blei rekna ut for dei regionane der ein eller fleire kommunar manglar tal i det aktuelle året. Eit alternativ kunne til dømes vore å setja inn estimerte verdiar for dei kommunane der tal mangla og på den måten få verdiar for alle regionar i alle år. Hovudgrunnen til at eg ikkje har gjort dette er fordi eg meiner det ikkje ville tilført resultat i masteroppgava vesentleg informasjon. I estimeringane ser eg på eit tverrsnitt av alle regionane. Då er det trenden på tvers av regionane som er interessant, ikkje tala for kvar enkelt region i seg sjølv. Totalt sett har dette ført til at det manglar inntektsinformasjon i om lag fem prosent av observasjonane. Å mangla informasjon i eit datasett kan gje skeive estimat dersom informasjonsmanglane ikkje kjem tilfeldig. I datasettet eg brukar ser det ut som informasjonsmanglane kjem tilfeldig, og difor er det liten grunn til å tru at desse manglane vil gje skeive estimat.

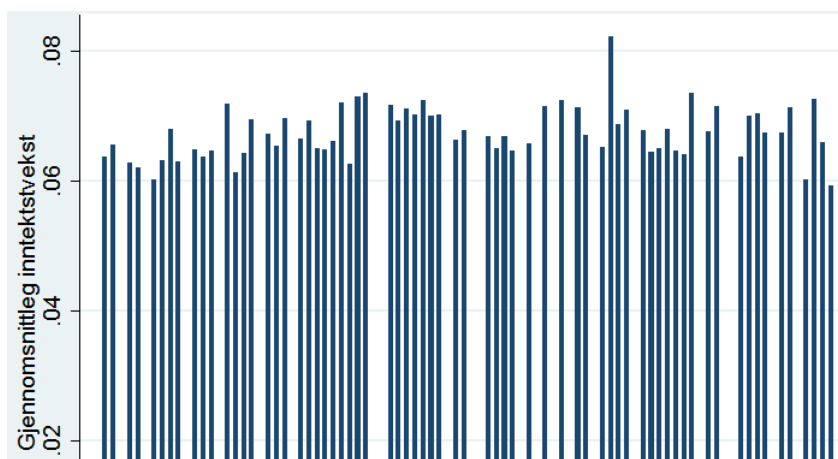
Det er til dels store skilnader mellom regionane når det gjeld gjennomsnittleg inntekt, men figur 5 viser ikkje ein klar geografisk trend. Til dømes ser det ikkje ut som om Nord-Noreg har systematisk lågare inntekt enn resten av landet. Sjølv om inntektsnivået er rimeleg jamnt over heile landet, skil region 6, Bærum og Asker, og region 9, Oslo, seg ut med ein vesentleg høgare gjennomsnittsinntekt enn resten av landet. Som vist i appendiks

A.1, tabell 7 har Bærum og Asker også den høgaste fullføringsgraden i den vidaregåande opplæringa.



Figur 5: Gjennomsnittleg inntekt i regionane i år 2000, målt i tusen kroner

Figur 6 illustrerer skilnaden i den økonomiske veksten mellom dei ulike regionane frå og med 1991 til og med 2008 for regionane med observert inntekt alle desse åra.



Figur 6: Årleg gjennomsnittleg inntektsvekst i regionane frå og med 1991 til og med 2008

I figur 6 er alle regionar der det manglar ein eller fleire observasjonar for inntektsvekst i perioden fjerna. Inntektsstatistikk for desse regionane finst likevel i appendiks A.1. For mange av desse regionane manglar det berre tal for inntektsveksten i eitt enkelt år. Dette vil likevel kunna påvirka gjennomsnittet, slik at resultatet for regionen ikkje direkte

er samanliknbart med andre regionar. Som tidlegare argumentert for er dei manglande observasjonane totalt sett såpass få at det ikkje bør påvirke resultatane i kapittel 5.1.

Hovudtrekket som visast i figur 6 er at alle regionane har opplevd ein positiv og relativt stor nominell inntektsvekst i denne perioden. Figur 6 viser at dei fleste regionane har mellom 6 og 7,5 prosent årleg inntektsvekst. Dette er vesentleg større enn den nominelle inntektsveksten reelt sett har vore, og denne skilnaden kjem som følgje av endringane av spesifikasjonane i datasettet. Det viktigaste er likevel å sjå på inntektsveksten for regionane relativt til kvarandre. I motsetnad til på fullføringsgraden og folketalsutviklinga, er det ikkje mogleg å spora ein klar geografisk trend på veksten i gjennomsnittleg inntekt. Regionen som klart skil seg ut med høgast inntektsvekst er region 63, Hitra/Frøya.

Samanliknar ein figur 4 og figur 6 ser ein også klare skilnader. Det ein kan tolka ut frå dette er at det ikkje nødvendigvis er regionane med størst inntektsvekst som opplever ein positiv folketalsutvikling. Dette er også i samsvar med ein studie gjort av Rattsø og Stokke (2011), som ikkje finn ein korrelasjon mellom desse to måla på regional utvikling. Ut frå dette er det også rimeleg å tenkja seg at fullføringsgraden kan påvirke folketalsutviklinga og inntektsveksten i regionane ulikt.

4.5 Andre variablar

I fleire av estimeringane kjem eg også til å inkludera kontrollvariablar. Kontrollvariablane eg har valt å bruka i denne oppgåva er andelen i ein region med høgare utdanning, andelen som bur i spredtbygde strøk, andelen i arbeidsfør alder og folketallet i regionen. Dette er alle variablar som er enkle å finne data for, samtidig som dei alle kan tenkjast å ha påverknad på den regionale utviklinga. Arbeidsfør alder vert i denne oppgåva definert frå 16 til 66. Data for kontrollvariablane er henta frå NSDs kommunedatabase.

I tabell 1 og 2 presenterer eg ei kort oppsummering av desse kontrollvariablane. Tabell 1 viser nøkkeltal for 1991, altså starten av estimeringsperioden, mens tabell 2 viser nøkkeltal for 2008, slutten av estimeringsperioden. Folketalet i regionane er grundig oppsummert tidlegare, eg kjem difor ikkje til å visa det i tabell 1 og 2.

Variabel	Observasjonar	Gjennomsnitt	Min	Maks
Andelen med høgare utdanning	90	0,124 (0,038)	0,071	0,339
Andelen som bur i spredtbygde strøk	90	0,408 (0,174)	0,006	0,790
Andelen i arbeidsfør alder	90	0,641 (0,023)	0,590	0,692

Tabell 1: Ei kort oppsummering av kontrollvariablane i 1991. Standardavvik i parentes

Frå tabell 1 ser ein at det for desse variablane er til dels store skilnader mellom regionane. Særleg gjeld dette andelen som bur i spredtbygde strøk. For andelen med høgare utdanning og andelen i arbeidsfør alder er skilnadene noko mindre.

Variabel	Observasjonar	Gjennomsnitt	Min	Maks
Andelen med høgare utdanning	88	0,206 (0,049)	0,119	0,424
Andelen som bur i spredtbygde strøk	88	0,355 (0,167)	0,003	0,761
Andelen i arbeidsfør alder	88	0,651 (0,016)	0,619	0,713

Tabell 2: Ei kort oppsummering av kontrollvariablane i 2008. Standardavvik i parentes

Som ein ser av tabell 2 er det to regionar som manglar observasjonar i 2008. Dette er region 43 (Haugesund) og region 46 (Søndre Sunnhordland). Frå 1991 til 2008 har i snitt andelen med høgare utdanning gått opp, andelen som bur i spredtbygde strøk gått ned og andelen i arbeidsfør alder halde seg stabil. For alle tre variablane har standardavvika halde seg forholdsvis stabile. Dette betyr i praksis at endringane i andelen med høgare utdanning og andelen som bur i spredtbygde strøk truleg har skjedd jamt over i alle regionar.

4.6 Empirisk spesifikasjon

I dette delkapittelet skal eg presentera den modellen eg set opp for å undersøka samanhengen mellom fullføring i vidaregåande opplæring og regional utvikling.

Modellen eg brukar i estimeringa av inntektsveksten utan faste effektar ser slik ut:

$$\Delta Inpc_{it} = \beta_1 fullforing_{i,t-10} + \beta_2 x_{it} + \sum_{t=1992}^{2008} \gamma_t D_t + u_{it} \quad (15)$$

$i = 1, 2, \dots, 90$ og $t = 1991, 1992, \dots, 2008$

På same måte vil tilsvarande likning for folkevekst bli:

$$\Delta Folketal_{it} = \alpha_1 fullforing_{i,t-10} + \alpha_2 x_{it} + \sum_{t=1992}^{2008} \psi_t D_t + v_{it} \quad (16)$$

$i = 1, 2, \dots, 90$ og $t = 1991, 1992, \dots, 2008$

Dei to variablane som skal forklarast er definert på følgjande måte:

$$\Delta Folketal_{it} = \frac{Folketal_{it} - Folketal_{i,t-1}}{Folketal_{i,t-1}} \quad \text{og} \quad \Delta Inpc_{it} = \frac{Inpc_{it} - Inpc_{i,t-1}}{Inpc_{i,t-1}}.$$

Som tidlegare nemnt er det altså årleg endring i gjennomsnittleg inntekt og folketal i regionane som skal estimerast.

Hovudforklaringsvariabelen i denne oppgåva er fullføringsgrad i vidaregåande skule, $fullforing_{i,t-10}$. Som skildra i kapittelet om deskriptiv statistikk, vert fullføringsgraden registrert som andelen som har fullført vidaregåande opplæring fem år etter at dei gjekk ut av ungdomsskulen. I Noreg fullfører folk vanlegvis ungdomsskulen det året dei fyller 16 år. Det er lite truleg at 16-åringar vil ha ein vesentleg påverknad på ein regions økonomiske utvikling. Difor har eg valt å tilbakedatere fullføringsgraden i forhold til folketals- og inntektsutviklinga. Eg har valt å tilbakedatera fullføringsvariabelen 10 år. Då vil dei fleste ha nådd å komma ut i arbeidslivet og det vil vera eit naturleg val med tanke på tidsspennet i datasettet. Dersom tilbakedateringa blir for stor, vil ein i estimeringa mista mange observasjonar, noko som kan gå ut over kvaliteten på estimeringsresultata.

Ein anna fordel med å tilbakedatera variabelen er at ein transformerer vekk eventuelle direkte simultanitetsproblem. Ved å bruka 10-års tilbakedatering vil eit slikt potensielt problem høgst truleg ikkje vera relevant. Den økonomiske utviklinga i dag kan nødvendigvis ikkje påverka fråfallet for 5-10 år sidan, og det er liten grunn til å tru at utelatne variablar er korrelerte med tilbakedatert fullføring i ein modell med faste effektar. Dermed er det høgst truleg i dette tilfellet ein einvegskorrelasjon mellom hovudforklaringsvariabelen og venstresidevariabelen.

Eg kjem også til å variere fullføringsvariabelen noko, for å sjå om resultata er følsame for om ein tilbakedaterer variabelen fleire eller færre år i staden for akkurat 10 år. Hypotesen er at dersom fullføring i vidaregåande opplæring har påverknad på økonomisk vekst 10 år etter at elevane starta i den vidaregåande opplæringa, bør den også ha det dersom tilbakedateringa vert større. Det vil også vera interessant å sjå om dei eventuelle effektane forsterkar seg etter kvart som tida går, eller om dei vert svekka. Eit minus med å gjera tilbakedateringa større er at datasettet vert mindre, sidan det vert færre tilgjengelige observasjonar per region. For å sikra eit stort tal observasjonar i regresjonen har eg valt å utvida tilbakedateringa til maksimalt 15 år. I tillegg kjem eg til å gå ned til 8 år, for å sjå om det påverkar resultata dersom tilbakedateringa vert mindre.

Variabelen x_{it} er ein vektor for kontrollvariablane eg brukar i estimeringane. Eg kjem til å variere antal kontrollvariablar som vert brukt for å sjå om dette har nokon påverknad på resultata i estimeringa. Det vil seia at enkelte estimeringar ikkje kjem til å ha kontrollvariablar i det heile tatt, mens andre estimeringar har fleire. Hovudpoenget med slike kontrollvariablar er å i størst mogleg grad kunna avdekkja andre variablar utanom hovudforklaringsvariabelen som kan tenkja seg å påverka den økonomiske veksten i enkelte regionar. Desse vil dermed kunna gje eit rikare bilete av kva som påvirkar den økonomiske veksten og utan desse kan ein få feilaktige utslag når ein forsøker å finna samanhengen

mellom fullføringsgraden og regional økonomisk utvikling. Disse kontrollvariablene skal med andre ord brukast til å minska problemet med eventuelle uteletne variablar. Resultata frå kontrollvariablene er ikkje det eg først og fremst ønskjer å undersøka i denne oppgåva, og er dermed i seg sjølv ikkje det viktigaste å få fram i resultatdelen.

Den kanskje viktigaste formen for kontrollvariablar i denne oppgåva er årsummyane representert ved uttrykket $\sum_{t=1992}^{2008} \psi_t D_t$ i likningane (15) og (16). Observasjonane for fullføring i vidaregåande opplæring går i datasettet frå 1981 til 2004. Sidan eg brukar 10-års tilbakedatering på desse, vil i praksis åra som blir inkludert i estimeringsmodellen vera frå og med 1991 til og med 2008. Dette gjev 18 ulike tidsobservasjonar for kvar region. (Enkelte regionar manglar observasjonar for enkelte av desse åra, men dei aller fleste vil få 18 observasjonar). I desse estimeringane vil det difor bli 17 årsummyar, sidan 1991 blir brukt som base.

Eg kjem til å ha med årsummyar i alle regresjonar. Dette er for å hindra at årlege svingningar i datasettet spelar inn på resultatet. Sidan inntekt er i nominell verdi er dette særleg viktig når inntektsveksten skal estimerast, men eg kjem til å ha med årsummyar også i alle estimeringar av folketalsutviklinga. Utan slike årsummyar vil det kunne vera fleire feilkjelder. Eit døme er at spesifikasjonane for personleg inntekt har endra seg gjennom datasettet, som nemnt i 4.4. Makroøkonomiske svingingar vil også ha stor påverknad på inntektsveksten i enkelte år. I tillegg har også gjennomsnittleg fullføringsgrad blitt høgare, som vist i kapittel 4.2. Dersom ein estimerer utan å ta med årsummyar vil ein på grunn av dette truleg kunna finna ein positiv samanheng mellom fullføring og inntektsvekst, utan at det nødvendigvis er fullføringa som spelar inn. Parameterverdiane for årsummyane er i seg sjølv ikkje interessante, og eg kjem difor ikkje til å rapportera dei i resultatkapittelet. Om påverknaden årsummyane har på inntektsveksten og folketalsutviklinga er signifikant ulik null er derimot vesentleg for om dei bør inkluderast i estimeringane eller ikkje. I resultatdelen kjem eg difor til å utføra F-testar for å sjekka dette.

5 Resultat

I dette kapitlet presenterer eg resultatane frå estimeringane av samanhengen mellom fullføringsgraden i den vidaregåande opplæringa i 90 norske regionar og utviklinga i desse regionane. I det første delkapitlet presenterer eg resultatane frå estimeringane av inntektsveksten. I kapittel 5.2 viser eg resultatane frå estimeringane av kva påverknad fullføringsgraden har på folketalsutviklinga i regionane. I dei to neste delkapittela vert det mykje diskutert om effekten av ein variabel er signifikant forskjellig frå null eller ikkje. Signifikanstesten og -nivået er basert på ein to-sidig t-test. For å utføra estimeringane har eg brukt dataprogrammet STATA, versjon IC 10.

5.1 Inntektsvekst

I dette delkapitlet viser eg først ei estimering av kva som kjenneteiknar regionar med høg gjennomsnittsinntekt (tabell 3). Deretter viser eg resultatane i estimeringane av kva som påverkar inntektsveksten i dei 90 regionane brukt i denne oppgåva. Desse vert presentert i tabell 4, og hovudtrekka derfrå vert kommentert i teksten.

Tabell 3. Estimering av inntektsnivå				
Variabel	MKM		FE	
	Modell 1	Modell 2	Modell 1	Modell 2
Fullført10	30,8*** (6,17)	16,3*** (3,68)	-7,55 (-1,19)	-3,25 (-0,91)
Spredtandel	-	-18,5*** (-9,21)		-11,6 (-0,79)
Folketal	-	0,047*** (6,18)		0,705*** (7,12)
Høg.utd.	-	162,4*** (8,43)		6,76 (0,09)
Arbeidsalder	-	6,25 (0,24)		78,5 (1,62)
Årsdummyar	Ja	Ja	Ja	Ja
R^2	0,87	0,95	0,98	0,99
Obs	1549	1549	1549	1549

Tabell 3: robuste t-verdiar i parentes. Inntektsnivået vert målt per innbyggjar og i tusen kroner.

*Signifikant på eit ti-prosentsnivå

**Signifikant på eit fem-prosentsnivå

***Signifikant på eit eitt-prosentsnivå

Inntektsnivået er i seg sjølv ikkje det eg ønskjer å måla i denne oppgåva. Samstundes

kan det å finna ut kva som kjenneteikner regionar med høg gjennomsnittleg inntekt vera interessant bakgrunnsinformasjon for den vidare diskusjonen om økonomisk utvikling i norske regionar. I tabell 3 har eg estimert inntektsnivået ved hjelp av MKM og med faste effektar (FE).

Modell 1 viser ei estimering med berre årsummyar og fullføringsgraden som forklaringsvariabel. Her ser ein at høg gjennomsnittleg inntekt og høg fullføringsgrad heng saman. Dersom fullføringsgraden aukar med 10 prosentpoeng impliserer modellen at gjennomsnittleg inntekt aukar med om lag 3000 kroner.

I modell 2 er også dei fire kontrollvariablane inkludert. Desse er skildra i kapittel 4.5. Framleis er fullføringsgraden positivt korrelert med gjennomsnittleg inntekt, men både parameterverdi og t-verdi har gått ned noko. I tillegg ser ein at regionar med høg gjennomsnittleg inntekt også har ein tendens til å vera folkerike, ha ein stor andel høgt utdanna og ha få som bur i spredtbygde strøk. Eit kanskje litt overraskande resultat er at effekten av andelen i arbeidsfør alder ikkje er signifikant ulik null i denne estimeringa.

Ei bekymring i desse estimeringane er at det er eit såkalla utelatt variabel-problem, drøfta i kapittel 3.1. For å løysa dette estimerer eg også inntektsnivået med faste effektar. Det fører til at effekten av fullføring vert identifisert ved å utnytta variasjon i inntekt og fullføring innan regionar. Det er altså endringar i fullføring som forklarar endringar i inntekt.

I estimeringa av modell 1 med faste effektar endrar resultatet seg vesentleg frå estimeringa med MKM. No har ikkje lenger fullføringsgraden ein signifikant påverknad på inntekta. Ved å estimera modell 2 med faste effektar finn ein den same tendensen. Heller ikkje her er påverknaden frå fullføringsgraden signifikant ulik null. I tillegg vert effekten av andelen som bur spredtbygde og effekten av andelen med høgare utdanning ikkje lenger signifikant ulik null. Påverknaden folketalet har på gjennomsnittsinntekta vert større ved bruk av estimering med faste effektar. Heller ikkje her er effekten av andelen i arbeidsfør alder signifikant ulik null.

Når ein skal estimera økonomisk utvikling er det naturleg å bruka ein endringsform. Difor er det mest interessante i dette delkapittelet påverknaden fullføringsgraden har på inntektsutviklinga, ikkje inntekt i seg sjølv. Likninga brukt i estimeringa av inntektsveksten er likning (15) frå kapittel 4.6. Resultata frå estimeringane av inntektsveksten er vist i tabell 4. Som vist i tabell 4 har eg variert både estimeringsmetode og kva variablar som har vorte tekne med i estimeringa. Metodane eg har brukt er minste kvadraters metode (MKM), estimering med tilfeldige effektar (RE) og estimering med faste effektar (FE). Eg har valt å estimera to modellar med kvar metode, ein med berre årsummyar og fullføring som forklaringsvariablar, og ein med alle kontrollvariablar inkludert. I den siste kolonna i tabell 4 har eg også teke med ei estimering med fullføringsgraden kvadrert (modell 3).

Seinare i teksten presenterer eg ein figur, figur 7, som viser påverknaden ulike tilbakedateringar av fullføringsgraden har på inntektsveksten. Desse resultatata vert ikkje presentert i tabell 4.

I modell 1 har eg estimert inntektsendringa berre ved hjelp av årsummyar og fullføringsgraden for dei som gjekk ut av ungdomsskulen 10 år tidlegare med MKM. Fullføringsgraden er i denne estimeringa ikkje signifikant ulik null, med ein t-verdi på 0,55. Likevel er R^2 såpass høg som 0,89. Dette tyder på at årsummyane har ein klart signifikant effekt på inntektsveksten. Ein F-test stadfestar også dette. Ved å testa nullhypotesen om at ingen av årsummyane har ein signifikant påverknad blir F-verdien 414,7. Den kritiske verdien med (17,1516) frihetsgrader på eit fem-prosentsnivå vert om lag 1,6, og dermed kan ein klart avvise denne nullhypotesen. Sidan F-verdien er såpass høg testar eg ikkje signifikansen til årsummyane i dei andre estimeringane.

Modell 2 er, som i estimeringa av inntektsnivået, inkludert dei fire kontrollvariablane. Heller ikkje her er påverknaden til fullføringsgraden er signifikant ulik null. Ingen av parameterverdiane til kontrollvariablane viser seg å vera signifikant ulik null i denne estimeringa. R^2 endrar seg ikkje noko, dette kan tolkast i retning av at tidsdummyane også i denne estimeringa har svært mykje å bety.

I kolonne 4 og 5 har eg estimert modell 1 og 2 med tilfeldige effektar, skildra i kapittel 3.3. Tendensen er den same som tidlegare, fullføringsgraden har ikkje ein signifikant påverknad på inntektsveksten. Parameterverdiane og t-verdiane er identiske med tilsvarande estimeringar med MKM, både for fullføringsgraden og kontrollvariablane. Dette betyr at heller ikkje her er effekten av verken fullføringsgraden eller kontrollvariablane signifikant ulik null.

Dei tre siste kolonnene viser estimeringane med faste effektar. Grunnen til at eg brukar denne estimeringa mykje er fordi eg antar det kan liggja eit utelatt variabel-problem i dei andre estimeringane, altså at det finst utelatne variablar som kan påvirka estimeringane. Ved å bruka faste effektar vil ein fjerna dei regionspesifikke effektane som ikkje varierer over tid. Denne metoden er skildra nærmare i kapittel 3.2.

Ved estimering av modell 1 med faste effektar går både parameterverdien og t-verdien til fullføringa noko opp, men effekten er framleis ikkje signifikant. Estimeringa av modell 2 gjev same tendens, og heller ikkje her er effekten signifikant ulik null. Å bruka faste effektar endrar også den estimerte påverknaden frå kontrollvariablane. I denne estimeringa har folketalet ein parameterverdi på $-4,86 * 10^{-4}$, og er signifikant ulik null på eit eitt-prosentsnivå. Dette betyr i praksis at dersom folketalet i ein region aukar med 1000 innbyggjarar, vil inntektsveksten gå ned med i underkant av 0,05 prosent.

Andelen i arbeidsfør alder vert i denne estimeringa signifikant ulikt null, men berre med

Variabel	Tabell 4. Fullføringsgradens effekt på inntektsvekst						
	MKM		RE		FE		
	Modell 1	Modell 2	Modell 1	Modell 2	Modell 1	Modell 2	
Fullført (10)	0,006 (0,55)	0,001 (0,05)	0,006 (0,55)	0,001 (0,05)	0,018 (1,11)	0,017 (0,99)	-0,056 (-0,70)
Fullført (10)sq	-	-	-	-	-	-	0,060 (0,97)
Spredtandel	-	0,004 (0,62)	-	0,004 (0,62)	-	0,046 (1,21)	0,051 (1,35)
Folketal	-	$1,01 * 10^{-5}$ (0,57)	-	$1,01 * 10^{-5}$ (0,57)	-	$-4,86 * 10^{-4}$ ** (-4,98)	$-4,83 * 10^{-4}$ ** (-4,99)
Høg.utd.	-	-0,010 (-0,31)	-	-0,010 (-0,31)	-	0,151 (1,32)	0,135 (1,21)
Arbeidsalder	-	-0,066 (-1,12)	-	-0,066 (-1,12)	-	-0,156* (-1,83)	-0,186** (-2,03)
Årsdummyar	Ja	Ja	Ja	Ja	Ja	Ja	Ja
R^2	0,89	0,89			0,89	0,89	0,89
Obs	1535	1535	1535	1535	1535	1535	1535

Tabell 4: t-verdier i parentes. Disse er justert for mogleg heteroskedastisitet

*Signifikkant på eit ti-prosentsnivå

**Signifikkant på eit fem-prosentsnivå

***Signifikkant på eit eit-prosentsnivå

på eit ti-prosentsnivå. Kanskje noko overraskande er parameterverdien negativ, $-0,156$. Koeffisienten impliserer at dersom andelen i regionen i arbeidsfør alder aukar med ti prosentpoeng, vil inntektsveksten i regionen gå ned med $1,56$ prosentpoeng. Dei estimerte effektane av dei andre kontrollvariablane er framleis ikkje signifikant ulik null.

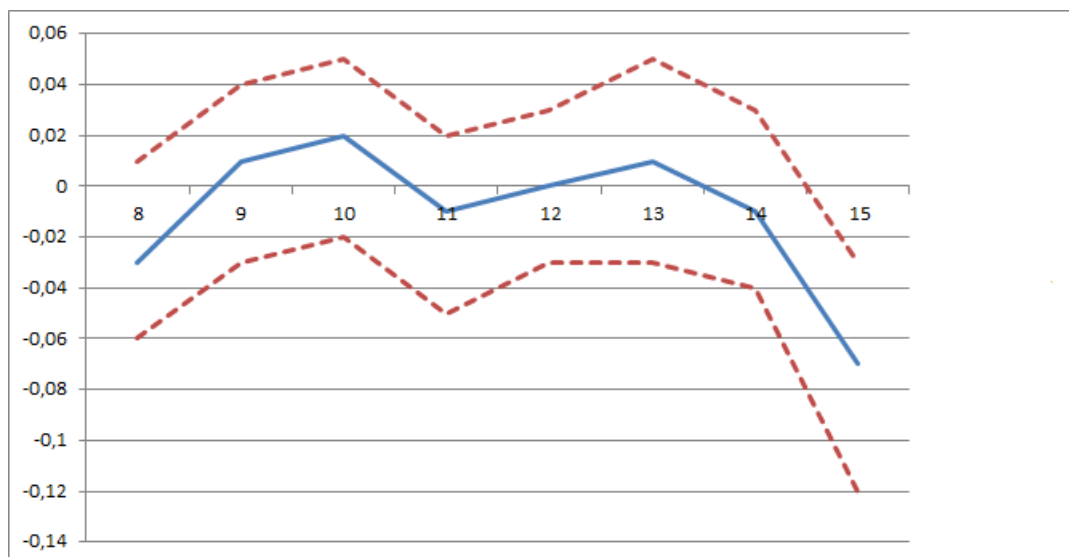
Ein Hausman-test for å samanlikna estimeringane utført på modell 1 og modell 2 med tilfeldige og faste effektar viser ikkje at estimering med faste effektar er å føretrekkja framfor estimering med tilfeldige effektar. χ^2 -verdiane i desse to testane er låge, $1,71$ for modell 1 og $11,0$ for modell 2. Modell 1 gjev 18 frihetsgrader, mens modell 2 gjev 21 frihetsgrader. Kritisk verdi for ein χ^2 -test med 18 frihetsgrader og eit fem-prosents signifikansnivå er $28,9$. Med 21 frihetsgrader og likt signifikansnivå vert den kritiske verdien $32,7$. Dermed kan ein ikkje avvisa nullhypotesen om at tilfeldige effektar og faste effektar begge gjev konsistente estimat. Dersom restledda i estimeringane er ukorrelerte, vil ein estimering med tilfeldige og faste effektar gje like parameterverdiar. I desse tilfella er estimering med tilfeldige effektar å foretrekka grunna effisiens. F-test på uteletne variablar viser heller ikkje at estimering med faste effektar er å foretrekka. Ein F-test på regiondummyane i modell 2 (gjort med MKM) gjev ein F-verdi på desse på $0,59$. Med $(88,1424)$ frihetsgrader og eit fem-prosents signifikansnivå vil den kritiske verdien vera i underkant av $1,3$. Med andre ord peikar heller ikkje dette av at det er uobserverbare regioneffektar som ikkje endrar seg over tid i estimeringa.

I modell 3 inkluderer eg fullføring kvadrert for å sjå om dette kan ha påverknad på effekten av fullføring på inntektsvekst. Som ein kan sjå av tabellen har verken fullføringsgraden i seg sjølv eller fullføringsgraden kvadrert ein signifikant påverknad. Heller ikkje kontrollvariablane endrar seg vesentleg frå den tidlegare estimeringa med faste effektar, dei estimerte effektane av kontrollvariablane vert i praksis uendra.

Det kan også vera interessant å sjå kva som skjer dersom tilbakedateringane av fullføringsgraden vert endra. Eg har brukt tilbakedateringar frå og med 8 år til og med 15 år. Desse estimeringane vart gjort med faste effektar. Resultatet av desse estimeringane er ikkje presentert i tabell 2, men i figur 7. Her viser den heiltrukne linja dei estimerte parameterverdiane for ulike tilbakedateringar, mens dei to stipla linjene markerer eit 95-prosents konfidensintervall.

Tala på x-aksen i figur 7 viser til antal år fullføringsgraden er tilbakedatert. Som ein ser av figuren ligg den estimerte effekten av fullføringsgraden, med eit unntak for fullføringsgraden tilbakedatert 15 år, rundt null. Figur 7 viser at det ikkje er mogleg å spora ein gjennomgåande påverknad frå fullføringsgraden på inntektsveksten i dei norske regionane.

For å summera opp synast det som at estimeringane ikkje er i stand til å gje ein god forklaring på regionale skilnader i inntektsveksten. Det er få og til dels ingen signifikante samanhengar mellom dei nytta variablane og inntektsveksten. Som ein ser av tabell 4 og



Figur 7: Effekt av fullføringsgrad på inntektsutvikling med eit 95-prosents konfidensintervall

figur 7 er det vanskeleg å konkludera med at det er ein samanheng mellom fullføringsgraden og regional inntektsvekst. Ved bruk av faste effektar viser estimeringane ein negativ samanheng mellom endring i dei to variablane folketal og andelen i arbeidsfør alder og endring i inntektsveksten. At inntektsveksten skal gå ned i ein region dersom andelen i arbeidsfør alder aukar, er kanskje eit noko overraskande resultat. Føyingsmålet R^2 varierer i praksis ikkje mellom dei ulike estimeringane. Dette tyder på at årsummyane forklarar variasjonen i inntektsvekst, og at fullføringsgraden og kontrollvariablane bidreg med lite ny informasjon.

At fullføringsgraden ikkje klarar å forklara variasjonen i regional inntektsvekst er i tråd med nullhypotesen, som var at fullføringa ikkje påverkar regional utvikling. Samstundes er det også i tråd med funna til mellom anna Krueger og Lindahl (2001), som ikkje finn nokon samanheng mellom humankapital og økonomisk vekst i OECD-land, og Pritchett (2001).

Eit viktig spørsmål rundt estimeringane av inntektsveksten vil nødvendigvis vera kvifor forklaringsvariablane i såpass liten grad er i stand til å forklara skilnadene mellom regionane i inntektsveksten. Eit viktig poeng er at inntektsveksten i seg sjølv i stor grad kan forklarast ved hjelp av årsummyane. Dette er ikkje spesielt overraskande. Inntektsvekst henge ofte nøye saman med konjunktursvingingar, og slike svingingar er eksogent gitt i denne modellen. I tillegg har, som eg skildra i kapittel 4.3, spesifikasjonane for inntekt endra seg enkelte gangar gjennom datasettet. Desse spesifikasjonsendringane har ført til unormalt stor endring i gjennomsnittleg inntektsvekst enkelte år, noko som vert fanga opp av årsummyane. Noreg har i tillegg ein sentralisert lønnsstruktur, særleg i offentleg sektor, som gjer at lønnsforskjellane innad i mange yrke er liten. Dette fører også til at mykje av lønnsauken ein observerer nasjonalt, finn ein også igjen i alle dei 90 regionane.

Årsdummyane forklarar ikkje variasjonen på tvers av regionane. Som vist i appendiks A ligg inntektsveksten i regionane stort sett mellom 6 og 7,5 prosent, noko som betyr at det er til dels store skilnader i gjennomsnittleg årleg inntektsvekst mellom regionane. Som ein ser av tabell 4 er mange av variablane i estimeringa ikkje i stand til å forklara desse skilnadene. Eg kjem ikkje til å gå grundig inn på kva anna som kan påvirka inntektsutviklinga, men det er grunn til å tru at det er fleire ulike faktorar som kan spela inn. Mellom anna er det ikkje urimeleg å anta at regionar med gode naturgitte forhold for visse typar næringsverksemd, som til dømes nærleik til oljeindustri og -utvinning eller oppdrettsnæringa, ofte har ein større inntektsvekst enn regionar utan desse forholda.

5.2 Folketalsutvikling

I dette delkapittelet skal eg presentera dei viktigaste resultata i estimeringa av kva som påverkar folketalsendringa i dei 90 regionane brukt i denne oppgåva. Heller ikkje her vert parameterverdiane for årsdummyane eller konstantleddet i regresjonen presentert. Likninga som vert brukt i denne estimeringa er likning (16) frå kapittel 4.6. Tabell 5 viser resultata frå estimeringane av folketalsutviklinga.

Estimeringane av folketalsendringa har gjeve vesentleg ulike resultat enn estimeringane i førre delkapittel. Først og fremst er det her i mykje større grad ein tydelegare tendens til signifikante effektar. I tillegg varierer R^2 ein god del mellom dei ulike modellane, i motsetnad til estimeringane av inntektsveksten. Sånn umiddelbart kan dette tyde på at årsdummyane ikkje har så stor påverknad på folketalsendringa som på inntektsveksten, og at dei andre variablane brukt i estimeringa spelar ein relativt sett større rolle på folketalsendringa i regionane. Dette kjem eg tilbake til i diskusjonen nedanfor.

Den første kolonna i tabell 5 viser resultatet frå estimeringa av modell 1 med MKM, altså ei estimering med berre fullføringsgraden og årsdummyar. Her viser fullføringsgraden seg å vera signifikant forskjellig frå null, med ein t-verdi på 5,16. Estimeringa gjev ein parameterverdi på 0,014, dette vil seia at ein auke i fullføringsgraden på 10 prosentpoeng gjev ein positiv endring i folketalsutviklinga på 0,14 prosentpoeng. Som i førre delkapittel testar eg også om årsdummyane er signifikante ved hjelp av ein F-test. F-verdien vert 3,90, som er vesentleg lågare enn tilsvarande test i estimeringa for inntektsvekst. Med (17,1596) frihetsgrader og eit fem-prosents signifikansnivå vert den kritiske verdien også her om lag 1,6, og dermed kan nullhypotesen om ingen effekt av årsdummyane forkastast. Eg kjem difor til å ta med årsdummyar i alle estimeringane også for kva som påverkar folketalsutviklinga. R^2 er i denne estimeringa 0,05, noko som betyr at fullføringsgraden og årsdummyane er i stand til å forklara 5 prosent av variasjonen i folketalsutviklinga mellom regionane.

Variabel	Tabell 5. Fullføringsgradens påverknad på folketalsutviklinga						
	MKM		RE		FE		
	Modell 1	Modell 2	Modell 1	Modell 2	Modell 1	Modell 2	
Fullført (10)	0,014*** (5,16)	0,021*** (7,70)	0,008*** (2,92)	0,009*** (3,61)	0,008** (2,55)	0,009*** (2,62)	-0,038** (-2,09)
Fullført (10)sq	-	-	-	-	-	-	0,038*** (2,72)
Spredtandel	-	-0,011*** (-7,79)	-	-0,008*** (-2,72)	-	-0,003 (-0,20)	-0,001 (-0,01)
Folketal	-	9,92 * 10 ⁻⁶ *** (3,55)	-	1,95 * 10 ⁻⁵ *** (2,72)	-	1,98 * 10 ⁻⁴ *** (3,00)	1,95 * 10 ⁻⁴ *** (3,09)
Høg.utd.	-	0,014*** (2,96)	-	0,027** (2,39)	-	-0,020 (-0,31)	-0,028 (-0,45)
Arbeidsalder	-	0,104*** (7,82)	-	0,059*** (3,43)	-	0,045 (1,09)	0,028 (0,73)
Årstdummyar	Ja	Ja	Ja	Ja	Ja	Ja	Ja
R ²	0,05	0,38			0,62	0,63	0,64
Obs	1614	1614	1614	1614	1614	1614	1614

Tabell 5: t-verdier i parentes. Disse er justert for mogleg heteroskedastisitet

Signifikkant på eit ti-prosentsnivå

**Signifikkant på eit fem-prosentsnivå

***Signifikkant på eit eit-prosentsnivå

I modell 2 utvidar eg med dei fire kontrollvariablane. Også i denne estimeringa er effekten av fullføringsgraden signifikant ulik null. Parameterverdien har auka frå 0,014 til 0,021. Det betyr at ein auke i fullføringsgraden på 10 prosentpoeng gjev ein positiv endring i folketalsutviklinga på 0,21 prosentpoeng. Effekten av kontrollvariablane er signifikante i denne estimeringa og R^2 har auka til 0,38. Andelen i regionen som bur i spredtbygde strøk har her ein parameterverdi på -0,011. Dette vil seia at dersom andelen i ein region som bur i spredtbygde strøk går ned med 10 prosentpoeng, vil folketalsutviklinga betra seg med 0,11 prosentpoeng per år. For variabelen folketal betyr resultatet at ein eksogen auke i folketalet på 10 000 gjev ein auke i folketalsutviklinga på om lag 0,01 prosentpoeng per år. Også andelen med høgare utdanning og andelen i arbeidsfør alder har positive parameterverdiar, henholdsvis 0,014 og 0,104. Resultata frå estimeringen med desse kontrollvariablane er ikkje overraskande. Veldig generelt viser dette ein tendens der flyttestraumen går frå spredtbygde strøk og til regionar med stort folketal frå før, der befolkninga i stor grad er i arbeidsfør alder og har høgare utdanning.

Kolonne tre og fire viser resultata frå estimeringa med tilfeldige effektar (RE). Ved å estimera modell 1 med tilfeldige effektar går den estimerte effekten av fullføringsgraden ned til 0,008. t-verdien viser at effekten framleis er signifikant ulik null. Ved estimering av modell 2 med tilfeldige effektar endrar parameterverdien for fullføringsgraden seg til 0,009. Også for kontrollvariablane har parameterverdiane endra seg noko, men kvalitativt er resultata dei same. t-verdien har gått ned for alle kontrollvariablane samanlikna med MKM-estimeringa, men alle parameterverdiane er framleis signifikant forskjellig frå null på eit fem-prosentsnivå

Ved å estimera modell 1 med faste effektar held parameterverdien seg på 0,008, som i estimeringa av tilsvarande modell med tilfeldige effektar. t-verdien har gått ned, men effekten er framleis signifikant på eit fem-prosentsnivå.

Estimeringa av modell 2 med kontrollvariablar gjev litt andre resultat enn dei tidlegare estimeringane av denne modellen. Effekten av fullføringsgraden er framleis signifikant null, denne gangen også på eit eitt-prosentsnivå. Ut frå estimeringane som er gjort så langt ser det ut til at fullføringa har ein signifikant positiv påverknad på folketalsutviklinga. Ser ein på estimeringane med tilfeldige og faste effektar gjev ein auke i fullføringsgraden på ti prosentpoeng ein positiv endring i folketalsutviklinga på i underkant av 0,10 prosentpoeng per år. Som tabellen viser har t-verdien for enkelte av kontrollvariablane endra seg ganske mykje. Ved bruk av estimering med faste effektar er ikkje effekten av andelen i arbeidsfør alder, andelen i spredtbygde strøk eller andelen med høgare utdanning lenger signifikant ulik null. At desse effektane er signifikant ulik null ved bruk av MKM og tilfeldige effektar kan tyda på at det er ein korrelasjon mellom desse variablane og folketalsutviklinga, men at det er uobserverbare regionspesifikke variablar som står bak denne korrelasjonen. Folketalet er framleis signifikant på eit eitt-prosentsnivå, og parameterverdien har

auka samanlikna med verdiane estimert med MKM og tilfeldige effektar. Resultatet her viser at ein eksogen auke i folketalet på 10 000 innbyggjarar gjev ein positiv endring i folketalsutviklinga på i underkant av 0,20 prosentpoeng per år.

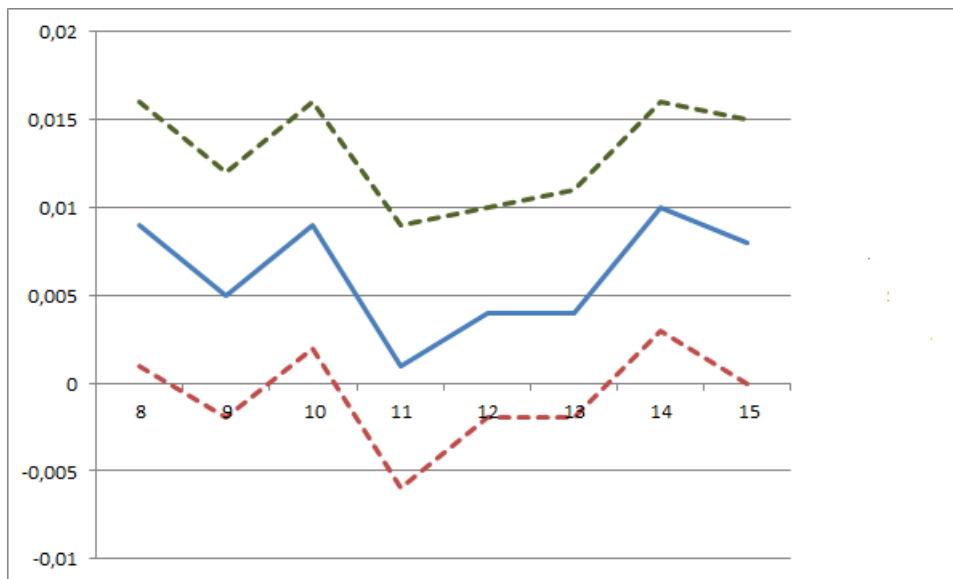
For å diskriminera mellom estimering med tilfeldige effektar og estimering med faste effektar går det også i desse estimeringane an å bruka ein Hausman-test. χ^2 -verdien er 0,91 for modell 1 og 36,3 for modell 2. Frihetsgradene er henholdsvis 18 og 21, og på eit fem-prosents signifikansnivå gjev dette kritiske verdiar på 28,9 for modell 1 og 32,7 for modell 2. I modell 1 avviser altså ikkje testen bruk av tilfeldige effektar til fordel for faste effektar, mens i modell 2 kan Hausman-testen tyda på at estimering med faste effektar er å foretrekka. Ein F-test på regiondummyar brukt i ein MKM-estimering stadfestar dette bildet. Ved å utføra ein F-test på regiondummyar i modell 2 blir F-verdien 11,7. På same måte som i 5.1 er den kritiske verdien i underkant av 1,3. Dette gjer at ein kan avvise nullhypotesen om at effekten av alle regiondummyane er lik null. I praksis kan ein dermed konkludera med at i estimeringa av folketalsutviklinga finst det uobserverte regionspesifikke variablar som spelar inn.

I modell 3 har eg, på same måte som i førre delkapittel, inkludert fullføringsgraden kvadrert. Ved å leggja til eit kvadrert ledd ser ein at effekten av både det opprinnelege fullføringsleddet og det kvadrerte fullføringsleddet er signifikant forskjellig frå null. Det opprinnelege leddet har ein negativ parameterverdi, mens det kvadrerte leddet har ein positiv parameterverdi. Dette er ein indikasjon på at fullføringa ikkje påverkar folketalsutviklinga lineært. Eit interessant moment er at desse parameterverdiane er identiske i absoluttverdi. Dermed vil nullpunktet av påverknaden ei endring i fullføringsgraden har på endringa i folketalsutviklinga vera 0,5. For fullføringsgrader lågare enn 0,5 vil ein marginal auke i fullføringsgraden føra til ei marginal negativ endring i folketalsutviklinga. Ved ein fullføringsgrad høgare enn 0,5 vil derimot ein marginal auke i fullføringsgraden gje ein marginal positiv endring i folketalsutviklinga. Denne positive effekten er aukande i fullføringsgraden. Som ein ser av appendiks A.1, tabell 6, har dei fleste regionane i snitt ein høgare fullføringsgrad enn 0,5.

Den estimerte effekten av fullføringsgraden på folketalsutviklinga i modell 2 med faste effektar er 0,009. For å få ein effekt på 0,009 med modell 3 må fullføringsgraden vera 0,61. Dette er ganske nært den gjennomsnittlege fullføringsgraden i datasettet.

Også i estimeringa av folketalsutviklinga har eg variert tilbakedateringslengda for å sjå om det påverkar effekten av fullføringsgraden funne i tabell 5. I figur 8 viser eg påverknaden fullføringsgraden har på folketalsutviklinga med tilbakedateringar frå og med 8 års tilbakedatering til og med 15 års tilbakedatering. Den heiltrukne linja viser dei estimerte effektane, mens dei stipla linjene markerer eit 95-prosents konfidensintervall.

Uavhengig av tilbakedateringslengde har fullføringsgraden ein positiv effekt på folketals-



Figur 8: Effekt av fullføringsgrad på folketalsutvikling med eit 95-prosents konfidensintervall

utviklinga. Kor stor den estimerte effekten er varierer noko etter kor stor tilbakedatering ein brukar. Ut frå figuren er det likevel rimeleg å anta at dersom fullføringsgraden endrar seg med 10 prosent, vil folketalsutviklinga betra seg med i underkant av 0,1 prosentpoeng per år.⁵

⁵Eg har også lagt til eit kvadrert fullføringsledd for fleire ulike tilbakedateringar. Ei slik estimering med 8 og 15 års tilbakedatering viser ikkje ein signifikant påverknad frå det kvadrerte leddet. Resultata med eit kvadrert fullføringsledd er dermed ikkje eintydige

6 Oppsummering og konklusjon

Denne masteroppgåva handlar om samanhengen mellom fullføring i den vidaregåande opplæringa og regional utvikling. Regiondefinisjonen som er brukt er laga av Statistisk sentralbyrå, og deler Noreg inn i 90 regionar. Grunnen til at akkurat dette nivået er valt er fordi desse regionane i stor grad representerer dei faktiske arbeidsmarknadane i Noreg i dag, og er såleis gode bilde på korleis Noreg faktisk er delt inn reint økonomisk. For å estimere denne samanhengen har eg utført ein empirisk paneldatanalyse, basert på observasjonar av fullføringsandelen for årskulla som har gått ut av grunnskulen etter 1981.

Vekst i gjennomsnittleg personinntekt og folketalsutvikling er brukt som mål på regional utvikling. Det er openbart i ein regions interesse å ha størst mogleg inntektsvekst for innbyggjarane. For dei fleste norske regionar er det også eit mål å få fleire innbyggjarar. I norsk samanheng er dermed folketalsutvikling eit adekvat mål på økonomisk utvikling.

Økonomisk vekstteori peikar ofte på at humankapital har ein positiv påverknad på den økonomisk utviklinga i eit område. Fullføringsgraden i vidaregåande må kunna sjåast på som eit av fleire mål på ein regions humankapitalnivå. Fleire empiriske studiar har likevel ikkje funne noko signifikant samsvar mellom humankapitalnivået og den økonomiske veksten ved å samanlikna mellom anna OECD-land. Enkelte forskarar peikar på at variablane som har blitt brukt for å skildra humankapitalnivået i desse estimeringane ikkje er tilfredsstillande, og at med gode nok mål på humankapital vil ein finna at humankapitalen påverkar økonomisk utvikling positivt. Nullhypotesa mi har likevel heile vegen vore at fullføringa ikkje har påverknad på den regionale utviklinga.

Estimeringsmodellane i denne oppgåva brukar fullføringsandel for elevane som avslutta grunnskulen ti år tilbake i tid som forklaringsvariabel av interesse. Resultata er ulike for dei to måla på regional utvikling. Eg finn ingen systematisk samanheng mellom fullføringsgraden og vekstraten i gjennomsnittleg inntekt. Ut frå estimeringane synest det altså ikkje å vera nokon effekt av auka fullføringsandel på regional inntektsvekst. Dette er i tråd med dei empiriske studiane som ikkje har funne ein samanheng mellom utdanningsnivået i ein økonomi og den økonomiske utviklinga.

Samanhengen mellom fullføringa og folketalsutviklinga er derimot meir eintydig positiv. Alle estimeringane viser at effekten av fullføringsgraden er signifikant og at å auka fullføringsgraden bidreg positivt på folketalsutviklinga i norske regionar. Desse resultata er stabile på tvers av dei ulike tilbakedateringslengdene brukt i estimeringane. Estimeringane viser at ved å auka fullføringsgraden med 10 prosentpoeng, vil folketalsutviklinga betra seg med i underkant av 0,1 prosentpoeng per år. At fullføringa ser ut til å påverka folketalsutviklinga i dei norske regionane positivt bør også vera interessant for fylkes-

kommunane som ønskjer å få folketalsauke i regionane sine. Desse fylkeskommunane vil ha eit ekstra insentiv til å satsa ytterlegare på å få fleire til å fullføra den vidaregåande opplæringa.

Ved å bruka ei tilbakedatering på fullføringa på 10 år, ser ein i praksis på kva påverknad 26-åringar har på den regionale utviklinga. Eit alternativ til å sjå på fullføringa for eitt og eitt årskull er å slå saman fleire årskull. Ved å slå saman fullføringa og bruka eit gjennomsnitt av fullføringa til dømes 10-15 år tilbake, vil ein studera påverknaden dei i aldersgruppa 26-31 år har på den regionale utviklinga.

Eit anna alternativ er å utvida tilbakedateringa på fullføringsgraden ytterlegare. Ved å ha lengre tilbakedatering vil ein sjå på påverknaden frå dei i høgare aldersgrupper enn det eg har gjort. Ved å til dømes utvida tilbakedateringa til 20 og 25 år vil ein sjå på påverknaden frå henholdsvis 36- og 41-åringar. Grunna datasettets storleik valte eg å ikkje gjera dette i denne oppgåva, men om nokre år vil det vera eit enda større tidsrom å studera. Då vil det vera rimeleg at også tilbakedateringa kan utvidast.

For framtidige studiar vil det også vera interessant å sjå om tendensane funne i denne oppgåva ser ut til å vedvara over tid eller om dette er effektar ein opplever akkurat no.

Referanser

- Aghion, P. og Howitt, P. (2009): *The Economics of Growth*. The MIT Press.
- Ashenfelter, O. og Rouse, C. (1998): "Income, Schooling and Ability: Evidence from a New Sample of Identical Twins." *The Quarterly Journal of Economics* 113(1), 253–284.
- Becker, G. (1964): *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. University of Chicago Press.
- Brooks, C. (2008): *Introductory Econometrics for Finance*. Cambridge University Press. 2. utgåve.
- Carlsen, F., Johansen, K. og Stambøl, L. (2010): "Effects of regional labour markets on migration flows, by education level." Versjon frå september 2010, working paper.
- Ehrenberg, R. og Smith, R. (2003): *Modern Labor Economics*. Pearson Education. 8. utgåve.
- Falch, T., Borge, L.-E., Lujala, P., Nyhus, O. og Strøm, B. (2010): "Konsekvenser av manglende videregående opplæring." Senter for økonomisk forskning AS, rapport nr. 03/10.
- Falch, T., Borge Johannesen, A. og Strøm, B. (2009): "Kostnader av frafall i videregående opplæring." Senter for økonomisk forskning AS, rapport nr. 08/09.
- de la Fuente, A. (2011): "Human capital and productivity." *Nordic Economic Policy Review* 2, 103–133.
- Glaeser, E. og Saiz, A. (2004): "The Rise of the Skilled City." *Brooking-Wharton Papers on Urban Affairs* 5, 47–94.
- Hanushek, E. og Woessmann, L. (2011): "How much do educational outcomes matter in OECD countries?" *Economic Policy, juli 2011* , 427–491.
- Hustoft, A. G., Hartvedt, H., Nymoene, E., Stålnacke, M. og Utne, H. (1999): "Standard for økonomiske regioner." Rapport 99/6, Statistisk sentralbyrå.
- Krueger, A. og Lindahl, M. (2001): "Education for Growth: Why and For Whom?" *Journal of Economic Literature* 24, 1101–1136.
- Kunnskapsdepartementet (2008-2009): "Stortingsmelding nr. 44 (2008-2009) "Utdanningslinja"."
- Lucas, R. (1988): "On the Mechanics of Economic Development." *Journal of Monetary Economics* 22, 8–42.

- Manca, F. (2012): "Human Capital Composition and Economic Growth at the Regional Level." *Journal of the Regional Studies Association* 46, 100–150.
- Pritchett, L. (2001): "Where Has All the Education Gone?" *The World Bank Economic Review* 15(3), 367–391.
- Rattsø, J. og Stokke, H. (2011): "Migration and dynamic agglomeration economies: Regional income growth in Norway." Versjon frå februar 2011, working paper.
- Reiling, R. og Strøm (2012): "Education outcomes and the business cycle." Versjon frå januar 2012, working paper.
- Vandenbussche, J., Aghion, P. og Meghir, C. (2006): "Growth, distance to frontier and composition of human capital." *Journal of Economic Growth* 11, 97–127.
- Verbeek, M. (2008): *A Guide to Modern Econometrics*. John Wiley and Sons. 3. utgåve.
- Wooldridge, J. (2009): *Introductory Econometrics - A Modern Approach*. South-Western Cengage Learning. 4. utgåve.

A Appendiks

A.1 Deskriptiv statistikk

I den første delen av dette appendikset kjem eg til å presentera litt nærmare statistikk for kvar enkelt region. I kapittelet Deskriptiv statistikk forsøker eg å gje ein oversikt over datasettet, i dette kapittelet skal eg presentera statistikkane for kvar enkelt region litt grundigare. For ei forklaring på kva region ulike kommunar held til i viser eg til Hustoft et al. (1999).

Tabell 6. Regionvis deskriptiv statistikk				
Region	Folketal 1990	Fullføringsgrad	Folketalsendring	Inntektsvekst
Halden	27309	0,642	0,43	6,38
Moss	46713	0,628	0,82	6,57
Fredrikstad/Sarpsborg	121514	0,636	0,57	6,87*
Askim/Mysen	42445	0,616	0,61	6,29
Follo	88992	0,674	1,28	6,21
Bærum/Asker	130716	0,734	1,17	5,52*
Lillestrøm	149575	0,627	1,20	6,03
Ullensaker/Eidsvoll	45285	0,609	1,59	6,32
Oslo	458364	0,642	1,12	6,80
Kongsvinger	51072	0,613	-0,20	6,30
Hamar	80990	0,630	0,35	6,79*
Elverum	38584	0,603	-0,01	6,50
Tynset	18254	0,697	-0,34	6,38
Lillehammer	33795	0,715	0,48	6,47
Gjøvik	67083	0,639	0,07	7,62*
Midt-Gudbrandsdalen	14878	0,647	-0,54	7,20
Nord-Gudbrandsdalen	21146	0,676	-0,48	6,14
Hadeland	26265	0,640	0,37	6,44
Valdres	19183	0,674	-0,36	7,00
Drammen	142327	0,643	0,81	5,39*
Kongsberg	28089	0,685	0,45	6,73
Hønefoss	34147	0,661	0,33	6,54
Hallingdal	20073	0,661	-0,03	6,96
Tønsberg/Horten	96175	0,661	0,88	8,85*
Holmestrand	11855	0,613	0,48	6,65
Sandefjord/Larvik	76313	0,656	0,69	6,92
Sande/Svelvik	12864	0,616	0,65	6,51
Skien/Porsgrunn	101826	0,633	0,33	6,48
Notodden/Bø	23330	0,655	0,05	6,61

Held fram på neste side

Regionvis deskriptiv statistikk				
Region	Folketal 1990	Fullføringsgrad	Folketalsendring	Inntektsvekst
Kragerø	15243	0,635	-0,19	7,20
Rjukan	6876	0,625	-0,69	6,27
Vest-Telemark	15706	0,674	-0,54	7,30
Risør	9599	0,626	-0,11	7,35
Arendal	67156	0,658	0,61	7,41*
Lillesand	12038	0,671	0,74	7,78*
Setesdal	8087	0,662	-0,04	7,18
Kristiansand	89006	0,684	1,05	6,93
Mandal	21261	0,706	0,59	7,12
Lyngdal/Farsund	17582	0,686	0,29	7,02
Flekkefjord	16177	0,684	0,04	7,24
Egersund	21319	0,686	0,34	7,00
Stavanger/Sandnes	191654	0,662	1,35	7,03
Haugesund	86029	0,684	0,60	7,66*
Jæren	36751	0,651	1,36	6,64
Bergen	323181	0,660	0,89	6,78
Søndre Sunnhordland	11754	0,690	0,14*	6,77*
Nordre Sunnhordland	43266	0,687	0,44	7,86*
Odda	14613	0,701	-0,86	6,69
Voss	16310	0,725	-0,16	6,52
Florø	14424	0,626	0,33	6,68
Høyanger	10529	0,678	-0,88	6,46
Sogndal/Årdal	26716	0,717	-0,11	6,51*
Førde	25406	0,687	0,42	6,57
Nordfjord	29465	0,726	-0,21	6,85*
Molde	59639	0,658	0,16	7,15
Kristiansund	34347	0,642	-0,14	7,92*
Ålesund	78737	0,663	0,54	7,24
Ulsteinvik	25552	0,656	0,11	10,4*
Ørsta/Volda	18192	0,696	0,11	7,14
Sunnalsøra	11016	0,691	-0,29	6,72
Surnadal	10863	0,721	-0,59	6,69*
Trondheim	186854	0,688	0,94	6,52
Frøya/Hitra	8591	0,584	-0,20	8,23
Brekstad	15946	0,622	-0,37	6,88
Oppdal	9102	0,672	0,04	7,11
Orkanger	21986	0,684	-0,02	8,23*
Røros	7865	0,719	-0,12	6,78
Steinkjer	38662	0,690	-0,13	6,46
Namsos	19320	0,666	-0,03	6,51
Stjørdalshalsen	20026	0,687	0,80	6,81
Levanger/Verdalsøra	32598	0,667	0,38	6,47

Held fram på neste side

Regionvis deskriptiv statistikk				
Region	Folketal 1990	Fullføringsgrad	Folketalsendring	Inntektsvekst
Grong	6112	0,658	-0,81	6,42
Rørvik	10180	0,589	-0,30	7,36
Bodø	74317	0,584	0,25	7,23*
Narvik	30558	0,609	-0,48	6,78
Brønnøysund	13374	0,543	-0,15	7,16
Sandnessjøen	16457	0,573	-0,59	8,40*
Mosjøen	16817	0,637	-0,13	7,16*
Mo i Rana	31277	0,597	0,02	6,38
Lofoten	24471	0,531	-0,21	7,00
Vesterålen	32261	0,559	-0,40	7,04
Harstad	32272	0,612	-0,22	6,74
Tromsø	65093	0,563	1,03	6,92*
Andselv	16528	0,619	-0,56	6,75
Finnsnes	20450	0,590	-0,36	7,14
Nord-Troms	12251	0,530	-0,49	7,41*
Vadsø	16920	0,466	-0,58	6,03
Hammerfest	26753	0,473	-0,52	7,26
Alta	20834	0,477	0,63	6,61
Kirkenes	9641	0,534	-0,07	5,94
Gjennomsnitt	47035	0,643	0,17	6,87

Tabell 6: Fullføringsgrad er eit gjennomsnitt over datasettet, folketalsendring og inntektsvekst er gjennomsnittleg prosentvis endring per år frå og med 1991 til og med 2008. Folketal 1990 er gitt per 1.1.1990. * betyr at regionen manglar observasjonar for enkelte år for gjeldande variabel. Dette vil nødvendigvis kunna påvirke det oppgitte resultatet, og gjera resultat ikkje direkte samanliknbart med resultatata for andre regionar. Særleg gjeld dette inntektsvekst, då denne avheng mellom anna av konjunktursvingingar, og vil variera sterkare frå år til år enn befolkningsendringa.

A.2 Tabell 7. Fullføringsstatistikk per år

Årstall	Observasjoner	Gjennomsnitt	Min	Maks
1981	90	0.489 (0.066)	0.323	0.683
1982	90	0.506 (0.070)	0.277	0.668
1983	90	0.524 (0.069)	0.328	0.692
1984	90	0.538 (0.073)	0.320	0.676
1985	90	0.554 (0.062)	0.364	0.698
1986	90	0.562 (0.070)	0.367	0.745
1987	90	0.588 (0.068)	0.378	0.732
1988	90	0.617 (0.065)	0.434	0.802
1989	90	0.639 (0.069)	0.441	0.782
1990	90	0.666 (0,065)	0.458	0.832
1991	90	0.676 (0.073)	0.344	0.799
1992	90	0.711 (0.062)	0.522	0.849
1993	90	0.714 (0.069)	0.470	0.869
1994	90	0.693 (0.088)	0.402	0.868
1995	90	0.696 (0.071)	0.490	0.839
1996	90	0.702 (0.067)	0.449	0.815
1997	90	0.721 (0.064)	0.560	0.840
1998	90	0.708 (0.074)	0.462	0.821
1999	90	0.707 (0.067)	0.516	0.847
2000	90	0.697 (0,076)	0.429	0.831
2001	90	0.687 (0.074)	0.439	0.818
2002	90	0.677 (0.074)	0.463	0.845
2003	90	0.685 (0.061)	0.466	0.825
2004	90	0.675 (0,067)	0.438	0.816

Tabell 7: Ei årleg oppsummering av fullføringsstatistikken. Standardavvik i parentes