

Hvem emigrerer fra Norge?

En empirisk analyse

Kristian Vigsø

Masteroppgave

Masteroppgaven er levert for å fullføre graden

Master i samfunnsøkonomi

Universitetet i Bergen, Institutt for økonomi

[Juni 2021]



UNIVERSITETET I BERGEN

Forord

Denne oppgaven markerer slutten på fem fine år ved Institutt for økonomi ved UiB. Jeg vil rette en stor takk til min veileder Espen Bratberg for tips til oppgave, tilrettelegging av data og gode tilbakemeldinger underveis i skriveprosessen. Videre vil jeg takke alle mine medstudenter for en flott og minnerik studietid.

Sammendrag

Norge har vært karakterisert ved en sammenpresset lønnsfordeling som impliserer at avkastningen av utdanning og annen kompetanse er relativt lav i Norge. Det betyr at det i hovedsak er individer med høyt kompetansenivå som vil tjene på å emigrere fra Norge. Med utgangspunkt i norske registerdata fra perioden 1993-2003 undersøker denne oppgaven hva som kjennetegner norske emigranter når det gjelder utdanning og annen kompetanse.

Roy-modellen predikerer at det er avkastningen av utdanning og annen kompetanse som driver migrasjonsstrømmene over landegrensene. I et land som Norge, hvor avkastningen av kompetanse er relativt lav, vil Roy-modellen predikere at emigrasjonsstrømmen hovedsakelig er positivt selektert. Positiv seleksjon betyr at emigrantene tilhører den øvre delen av kompetansefordelingen. Oppgaven tester prediksjonene til Roy-modellen med utgangspunkt i rammeverket til Borjas et al. (2019). De har gjort et tilsvarende arbeid på danske data.

Oppgaven benytter metoder for å undersøke hvordan emigrantene er selektert på observert og uobservert kompetanse. I oppgaven blir utdanning brukt som et mål på observert kompetanse, og residualer fra lønnslikninger blir brukt som et mål på uobservert kompetanse. Inntekten til et individ avhenger av både observert og uobservert kompetanse. Hvordan emigrantene er selektert på inntekt avhenger derfor av begge kompetansetyperne.

Resultatene til oppgaven viser at emigrantene har de forventede kjennetegnene. Tilbøyeligheten til å emigrere stiger med utdanningsnivået. Videre viser resultatene at emigrantene har et høyere inntektsnivå og uobservert kompetansenivå enn ikke-emigrantene. Resultatene viser derfor at emigrantene er positivt selektert på observert og uobservert kompetanse. Til den statistiske analysen har STATA/SE 16.1 blitt brukt.

Innholdsfortegnelse

1	Innledning	1
2	Litteraturoversikt	3
	2.1 Immigrasjon til Norge	3
	2.2 Emigrasjon ut av Norge og Norden	4
	2.3 Emigrasjon andre steder i verden	6
	2.4 Oppsummering av litteraturoversikten	7
3	Teoretisk rammeverk	9
	3.1 Positiv og negativ seleksjon	9
	3.2 Roy-modellen	10
4	Data og utvalgsdesign	16
	4.1 Data	16
	4.2 Utvalgsdesign	17
	4.3 Deskriptiv statistikk	19
5	Metode	24
	5.1 Seleksjon på observerbar kompetanse	24
	5.2 Seleksjon på uobserverbar kompetanse	25
	5.3 Metode for å undersøke om fordelingene er ulike	27
6	Resultater og diskusjon	29
	6.1 Seleksjon på observerbar kompetanse	29
	6.2 Analyse av stokastisk dominans prediksjonen til Roy-modellen	31
	6.2.1 Seleksjon på uobserverbar kompetanse	33
	6.2.2 Seleksjon på inntekt	39
	6.3 Begrensninger i analysen	44
	6.4 Sammenligning med Borjas et al. (2019) og andre studier	45
7	Oppsummering og avsluttende kommentar	47
	Litteraturliste	49
	Appendiks	52
	A.1 Negativ seleksjon	52
	A.2 Inndeling av utdanningskategorier	52
	A.3 Metode for å undersøke om fordelingene er ulike	54
	A.4 Kolmogorov-Smirnov-test	56
	A.5 Lineær sannsynlighetsmodell	57
	A.6 Kumulative fordelingsfunksjoner	58
	A.7 Sensitivitetsanalyse for kvinner	61
	A.8 Deskriptiv statistikk for emigranter med kjent destinasjonsområde	64

Tabelliste

Tabell 1: Deskriptiv statistikk for menn.....	20
Tabell 2: Deskriptiv statistikk for kvinner.....	21
Tabell 3: Oversikt over panelet for menn.....	23
Tabell 4: Oversikt over panelet for kvinner.....	23
Tabell 5: Gjennomsnittlige marginaleffekter fra logistisk regresjon.....	29
Tabell 6: Lønnslikninger.....	33
Tabell 7: Lønnslikninger.....	34
Tabell A.1: Nivåinndeling i NUS2000.....	53
Tabell A.2: Nivåinndeling av utdanning i denne oppgaven.....	53
Tabell A.3: Lineær sannsynlighetsmodell.....	57
Tabell A.4: Deskriptiv statistikk for kvinner (inntektsgrense på 150 000 (2015-NOK))	61
Tabell A.5: Deskriptiv statistikk for emigranter med kjent destinasjonsområde.....	64

Figurliste

Figur 1: Positiv seleksjon.....	12
Figur 2: Kumulative fordelinger av residuallønninger (standardisert inntekt).....	36
Figur 3: Estimert differanse mellom kumulative fordelinger med konfidensintervall.....	37
Figur 4: Kumulative fordelinger av standardisert inntekt.....	40
Figur 5: Estimert differanse mellom kumulative fordelinger med konfidensintervall.....	41
Figur 6: Kumulative fordelinger av standardisert inntekt (korttids- og langtidsemigranter).....	43
Figur A.1: Negativ seleksjon.....	52
Figur A.2: Kumulative fordelinger av residuallønninger (ikke-standardisert inntekt).....	58
Figur A.3: Kumulative fordelinger av residuallønninger (ikke-standardisert inntekt).....	59
Figur A.4: Kumulative fordelinger av residuallønninger (standardisert inntekt).....	60
Figur A.5: Sensitivitetsanalyse for kvinner (seleksjon på inntekt).....	62
Figur A.6: Sensitivitetsanalyse for kvinner (seleksjon på uobservert kompetanse).....	63

1 Innledning

Den internasjonale migrasjonen har økt over tid, både i absolutte tall og som prosent av verdens befolkning. I 2019 var det estimert at verden hadde ca. 272 millioner internasjonale migranter, som er 3,5 % av verdens befolkning.¹ Til en sammenligning var det ca. 153 millioner internasjonale migranter i 1990, som var 2,9 % av verdens befolkning den gangen. Det som får størst oppmerksomhet er migrasjonen inn til vestlige rike land. Migrasjonsstrømmen inn til Europa er dominert av migranter fra Afrika og Asia. Nord-Amerika mottar særlig mange migranter fra Latin- og Sør-Amerika. Rundt 50 % av verdens internasjonale migranter i 2019 bodde i Europa og Nord-Amerika (IOM 2020).

Migrasjonen innad i Europa har også økt over tid. Utvidelsen av EU østover har medført at en rekke øst-europeere har flyttet til Vest-Europa. I 2019 bodde 3,3 % av EUs befolkning, i aldersgruppen 20-66, i et annet EU-land enn hvor deres statsborgerskap tilhører. Dette er en økning fra 2009 hvor det var 2,4 % for samme gruppe. Det er stor variasjon mellom medlemslandene. Mange østeuropeiske land har høy andel av befolkningen, i aldersgruppen 20-66, som bor utenlands i et annet EU-land, mens andelen er lav for de fleste vesteuropeiske og nordiske landene. For samme alderspopulasjon bodde 1,6 % av norske statsborgere i et europeisk EU-land i 2009, hvor andelen ble redusert til 1,4 % i 2019 (Eurostat 2020). Siden januar 1994 har nordmenn hatt like rettigheter som andre EØS-borgere til å ta arbeid, studere og bo i alle EØS-land (Regjeringen 2021).

Krig og konflikt, politiske årsaker eller et ønske om å oppnå en høyere inntekt er eksempler på motiver på hvorfor et individ velger å emigrere fra hjemlandet sitt. Det vil være mange faktorer som påvirker hvilket destinasjonsland migranter velger, men inntektsmulighetene i destinasjonslandet vil trolig være en viktig faktor for mange migranter uansett hovedmotiv.

Norge har relativt lav avkastning av kompetanse (Bratsberg et al. 2014). Det impliserer at Roy-modellen predikerer at emigrasjonsstrømmen ut av Norge hovedsakelig vil være positivt

¹ Migrant er en samlebetegnelse for emigrant og immigrant. Som en migrant er man både en utvandrer (emigrant) fra landet man forlater og innvandrer (immigrant) til landet man flytter til (Tønnesen 2021). Fra kapittel 3 og ut vil jeg hovedsakelig bruke ordet emigrant. I kapittel 1 og 2 blir alle tre ordene jevnlig brukt.

selektert. Det vil si at emigrantene tilhører den øvre delen av kompetansefordelingen. På den andre siden vil immigrantere til Norge i hovedsak være negativt selektert.²

Det har nylig blitt publisert en vitenskapelig studie som analyserer emigrasjonsstrømmen ut av Danmark, hvor studien tester prediksjonene til Roy-modellen. I likhet med Norge er Danmark et land med relativt lav avkastning av kompetanse (Borjas et al. 2019). Roy-modellen predikerer derfor at danske emigranter i hovedsak vil være positivt selektert. De empiriske funnene til Borjas et al. (2019) bekrefter denne prediksjonen. Det eksisterer studier som har sett på emigrasjonsstrømmen ut av Norge (Schrøder 1996; Pedersen et al. 2003; Pedersen et al. 2008). Kort oppsummert finner de at norske emigranter har mer utdanning og inntekt enn ikke-emigrantene. De tester ikke prediksjonene til Roy-modellen i samme grad som Borjas et al. (2019) gjør. Motivasjonen for denne oppgaven vil derfor være å undersøke om atferden til norske emigranter samsvarer med prediksjonene til Roy-modellen. Oppgaven vil, så langt det lar seg gjøre, replikere analysen til Borjas et al. (2019) ved å bruke norske data.

Oppgaven har følgende struktur: I kapittel 2 gjør jeg rede for relevant litteratur. I kapittel 3 presenterer jeg prediksjonene til Roy-modellen som jeg skal teste empirisk. I kapittel 4 blir data, utvalgsdesign og deskriptiv statistikk presentert. I kapittel 5 gjør jeg rede for metodene jeg benytter i oppgaven. I kapittel 6 blir resultatene formidlet og drøftet. Til slutt vil jeg oppsummere og gi en avsluttende kommentar i kapittel 7.

² Litteraturen bruker begrepene positiv og negativ seleksjon til å beskrive migrasjonsstrømmer. At migranter er negativt selektert er ikke ensbetydende med at migrantene er et problem. Hvis et land har stort behov for arbeidskraft med lav formell kompetanse, kan det være en styrke at immigrasjonsstrømmen til landet hovedsakelig er negativt selektert.

2 Litteraturoversikt

Litteraturen på migrasjonsfeltet dekker mange aspekter ved migrasjon. Litteraturen ser blant annet på assimilasjon av immigranter, effekten av immigrasjon på lønninger og sysselsetting og kartlegging hva som kjennetegner immigrantene/emigrantene når det gjelder kompetanse. Denne oppgaven skal analysere seleksjonsprosessen til norske emigranter. Altså hva som kjennetegner emigrantene når det gjelder kompetansenivå. Det er derfor vesentlig å se hva andre studier har funnet ut ved å studere emigrasjonen ut av Norge og Norden. Litteratur som studerer seleksjonsprosessen til emigranter flere steder i verden vil også være relevant, slik at en kan få et overblikk over hvor valid prediksjonene til Roy-modellen er. Jeg har også valgt å presentere litteratur som tar opp immigrasjon til Norge. Disse studiene har ikke fokus på å identifisere om immigrantene er en positivt eller negativt selektert gruppe. Disse studiene fokuserer hovedsakelig på hva arbeidsmarkeds- og trygdeutfallet til immigrantene er, og hvordan immigrasjonen påvirker lønningene til utsatte yrkesgrupper. Det eksisterer også offentlige utredninger som har undersøkt innvandringen til Norge.

2.1 Immigrasjon til Norge

Bratsberg et al. (2010) studerer arbeidsmarkedsutfallet til arbeidsimmigranter fra utviklingsland som ankom Norge på 1970-tallet. De finner at sysselsettingen blant denne gruppen faller bratt over tid. Bratsberg et al. (2010) mener det er to mekanismer som driver utviklingen i sysselsettingsraten blant immigrantene. Den første mekanismen er at immigrantene arbeidet i nedadgående industrier, og den andre er den generøse norske velferdsstaten.

I likhet med Bratsberg et al. (2010) studerer Bratsberg et al. (2014) arbeidsmarkedsutfallet til immigranter. Forskjellen er hovedsakelig at Bratsberg et al. (2014) studerer flere immigrantgrupper, og inkluderer flere tidsperioder. Studien til Bratsberg et al. (2014) finner at immigranter som kommer fra høy-inntektsland presterer godt i det norske arbeidsmarkedet og på nivå med etniske nordmenn. Videre finner de at immigranter fra lavinntektsland har lav sysselsettingsrate og høy andel på trygd sammenlignet med etniske nordmenn med like bakgrunnsvariabler.

Bratsberg og Raaum (2012) studerer hvordan lønningene til bygningsarbeidere ble påvirket av immigranter i perioden 1998-2005. De finner at en økning i andelen immigranter sysselsatt i bygningssektoren har en negativ effekt på lønningene til etniske nordmenn som jobber i

sektoren. Studien viser at lønnseffekten er ujevnt fordelt, hvor den største effekten er på arbeidere med lavest formell kompetanse.

Hoen (2020) studerer hvordan etniske nordmenn sine lønninger har blitt påvirket av den store immigrasjonsbølgen av øst-europeiske arbeidsimmigranter som ankom på 2000-tallet. Hoen (2020) finner at en økning i andelen med østeuropeiske immigranter i et spesifikk yrke har en negativ effekt på etniske nordmenn sine lønninger i det spesifikke yrket, relativt til etniske nordmenn som er i yrker hvor østeuropeiske immigranter ikke er særlig representert i.

Det eksisterer offentlige utredninger som har undersøkt innvandringen til Norge. De er kjent som Brochmann 1 (NOU 2011:7) og Brochmann 2 (NOU 2017:2). NOU (2011:7) undersøker spesielt hvordan innvandringen til Norge påvirker bærekraften til den norske velferdsmodellen. NOU (2011:7) kommer med en rekke forslag på hvordan velferdsordningene kan tilpasses en situasjon med mer innvandring til Norge. NOU (2017:2) undersøker hvilke langsiktige konsekvenser høy innvandring kan skape med særlig vekt på flyktninger. NOU (2017:2) beskriver mulige strategier for hvordan Norge skal klare å integrere flyktninger på en god måte. Videre viser NOU (2017:2) at immigranter som kommer fra land i Asia, Afrika og Øst-Europa er sterkt representert i mange næringer hvor det typisk er mange arbeidere med lav formell kompetanse. Det kan være en indikasjon på at mange immigranter som kommer til Norge har lav formell kompetanse. Det må presiseres at mange av immigrantene ikke har kommet som arbeidsimmigranter, men som flyktninger eller via familiegjenforening.

2.2 Emigrasjon ut av Norge og Norden

Abramitzky et al. (2012) tar for seg emigrasjonsstrømmen fra Norge til USA som foregikk i perioden 1850-1913. Abramitzky et al. (2012) finner at emigrantene fra urbane områder i Norge var negativt selektert fra populasjonen. De finner tvetydige resultater når de analyserer seleksjonsprosessen blant emigrantene fra rurale områder. Emigrantene fra de urbane områdene tilhørte den nedre delen av inntektsfordelingen. USA hadde på den tiden en mer sammenpresset inntektsfordeling enn Norge (Abramitzky et al. 2012). Negativ seleksjon fra de urbane områdene er i tråd med Roy-modellen.

Schröder (1996) viser at tilbøyeligheten til å emigrere økte med utdanningsnivået i de skandinaviske landene på 1980-tallet. Videre viser Schröder (1996) at relativt lavt utdannede

hovedsakelig emigrerte til andre nordiske land, mens de relativt høyt utdannede emigrerte til ikke-nordiske land i Europa eller andre rike OECD-land.

Pedersen et al. (2003) undersøker emigrasjonsstrømmen ut av Skandinavia i årene 1981, 1989 (1991 for Norge) og 1998. Emigrantutvalget til Pedersen et al. (2003) består av alle individer over 19 år som emigrerte fra de skandinaviske landene i de respektive årene. Ikke-emigrantutvalget består av et tilfeldig trukket utvalg fra populasjonen (individer over 19 år). Pedersen et al. (2003) finner at sannsynligheten for å emigrere ut av Skandinavia øker med utdanningsnivået. Videre finner de en U-formet sammenheng mellom sannsynligheten for å emigrere og nivået på inntekten. Det er individer, som er utenfor eller marginalt i arbeidsstyrken, som dominerer i det laveste inntektsintervallet. Resultatene til Pedersen et al. (2003) viser tydelig at sannsynligheten for å emigrere stiger med inntekten for dem som er veletablert i arbeidsstyrken.

Pedersen et al. (2008) studerer den norske emigrasjonsstrømmen ut av Norge i 1991 og 1998, hvor de skiller mellom emigranter som flytter til nordiske land og rike OECD-land. Utvalget deres omfatter individer med positiv inntekt som befant seg i aldersgruppen 20-65 år, og som var født i Norge. Pedersen et al. (2008) finner at sannsynligheten for å emigrere til et OECD-land utenfor Norden er stigende med inntektsnivået, mens sannsynligheten for å emigrere til et nordisk land er synkende med inntektsnivået. Videre viser de at norske emigranter er yngre, har færre barn og har sjeldnere en partner. Studien viser at det er en positiv korrelasjon mellom utdanningsnivå og emigrasjon, hvor korrelasjonen er sterkest for emigranter som flytter til ikke-nordiske land.

Pirttilä (2004) studerer emigrasjonsstrømmen ut av Finland på 1990-tallet. Når det kontrolleres for en rekke bakgrunnsvariabler viser studien at høyt utdannede individer har fem ganger større sannsynlighet for å emigrere ut av Finland enn individer med bare videregående skole.

Borjas et al. (2019) studerer emigrasjonsstrømmen ut av Danmark i perioden 1996-2005.³ Utvalget deres omfatter etniske dansker som jobbet fulltid, og som tilhørte aldersgruppen 25-54 år. Studien utforsker hvordan emigranter og ikke-emigranter skiller seg fra hverandre på

³ Borjas et al. (2019) bruker registerdata fra perioden 1995-2010, men deler emigrantene inn i gruppene korttids- og langtidsemigranter. Langtidsemigranter, etter Borjas et al. (2019) definisjon, er emigranter som bor i utlandet i minst fem år. For å kunne dele emigrantene inn i de to gruppene er bare individer som emigrerer til og med 2005 inkludert i analysen deres.

observerbare og uobserverbare karakteristikkene.⁴ Når det gjelder observerbare karakteristikkene fokuserer Borjas et al. (2019) på inntekten til individene, og de benytter residualer fra lønnslikninger (residuallønninger) til å undersøke hvorvidt emigrantene og ikke-emigrantene har forskjellig uobservert kompetanse. Det nye med denne studien er at de tester stokastisk dominans prediksjonen til Roy-modellen.⁵ Danmark har relativt lav avkastning av kompetanse, og prediksjonen til Roy-modellen er at Danmark hovedsakelig vil ha en positivt selektert emigrasjonsstrøm. De empiriske funnene til Borjas et al. (2019) viser at danske emigranter er positivt selektert på observerbare og uobserverbare karakteristikkene. I analysen skiller Borjas et al. (2019) mellom emigranter som flytter til nordiske land og ikke-nordiske land, og mellom kjønn. Borjas et al. (2019) finner at begge emigranttypene er positivt selektert, men de finner en sterkere positiv seleksjon blant emigrantene som flytter til ikke-nordiske land.

2.3 Emigrasjon andre steder i verden

Parey et al. (2017) utforsker seleksjonsprosessen blant tyske emigranter med høy utdanning ved å bruke predikert inntekt. De finner resultater som er konsistent med Roy-modellen. Emigranter som flytter til land med mindre ulikhet enn Tyskland er negativt selektert, og emigranter som flytter til land med mer ulikhet er positivt selektert.

Gould og Moav (2016) analyserer emigrasjonsstrømmen ut av Israel, hvor de avgrensede analysen til kun å studere menn. De finner at sannsynligheten for å emigrere øker med utdanningsnivået. Videre finner de en invers u-formet sammenheng mellom emigrasjon og uobservert kompetanse. De benytter, i likhet med Borjas et al. (2019), residuallønninger som mål på uobservert kompetanse. De fleste emigrantene i studien flytter til USA, hvor avkastningen av kompetanse er høyere enn i Israel.

Det er flere studier som har analysert emigrasjonsstrømmen fra Mexico til USA. Chiquair og Hanson (2005) predikerer hvilken lønn meksikanere hadde i Mexico, før de emigrerte til USA, basert på deres observerbare karakteristikkene i USA. Mexico har høyere avkastning av kompetanse enn USA, og Roy-modellen predikerer derfor en negativt selektert emigrasjonsstrøm fra Mexico til USA. Chiquair og Hanson (2005) estimerer at de meksikanske emigrantene ikke tilhører den nedre delen av lønnsfordelingen i Mexico, men har en medium-

⁴ Hovedanalysen til Borjas et al. (2019) omfatter kun langtidsemigranter. En separat analyse av korttidsemigrantene finnes i appendikset til Borjas et al. (2019).

⁵ Stokastisk dominans prediksjonen til Roy-modellen vil bli gjort rede for i kapittel 3.

høy plassering i lønnsfordelingen. Fernandez-Huertas Moraga (2011) og Kaestner og Malamud (2014) bruker spørreskjema-data, som potensielt rapporterer den faktiske lønnen individene hadde i Mexico før de emigrerte til USA. De finner empiriske bevis på at emigrantene er negativt selektert på lønn fra den meksikanske populasjonen. Patt et al. (2021) undersøker hvordan meksikanere som emigrerer i USA er selektert på manuelle og kognitive ferdigheter.⁶ De finner at meksikanske emigranter har høyere manuelle og lavere kognitive ferdigheter enn meksikanske ikke-emigranter. Funnene er i tråd med det teoretiske rammeverket ettersom avkastningen av manuelle (kognitive) ferdigheter er høyere (lavere) i USA enn i Mexico (Patt et al. 2021).

Rosso (2019) studerer emigrasjonsstrømmen fra Polen til Storbritannia og Tyskland. Hun undersøker hvordan polske emigranter er selektert på observerbar (utdanningsnivå) og uobserverbar kompetanse (residuallønninger). Rosso (2019) estimerer at Polen har høyere avkastning på observerbar kompetanse enn Tyskland og Storbritannia, men finner at sannsynligheten for å emigrere fra Polen ikke faller med utdanningsnivået som Roy-modellen predikerer. Når det gjelder uobserverbar kompetanse finner Rosso (2019) funn som delvis samsvarer med prediksjonene til Roy-modellen.⁷

2.4 Oppsummering av litteraturoversikten

Litteraturen om immigrasjonen til Norge viser at ikke-vestlige immigranter presterer dårligere på arbeidsmarkedet enn innfødte nordmenn med like observerbare karakteristikk. Videre finner de at vestlige immigranter presterer godt på det norske arbeidsmarkedet (Bratsberg et al. 2010; Bratsberg et al. 2014). Andre studier studerer hvordan immigrasjonen påvirker innfødte sine lønninger. Hoen (2020) og Bratsberg og Raum (2012) viser at immigrasjonen har en negativ effekt på enkelte yrkesgrupper sine lønninger.

Litteraturen, som utforsker emigrasjonen ut av Norden, får konsistente resultater. Oppsummert finner litteraturen at nordiske emigranter har mer utdanning og høyere inntekt enn ikke-emigrantene. Borjas et al. (2019) viser at danske emigranter har et høyere uobservert

⁶ Manuelle ferdigheter er for eksempel praktiske ferdigheter og fysisk styrke. Kognitive ferdigheter er for eksempel kreativitet og evne til problemløsning.

⁷ Rosso (2019) finner at seleksjonen på uobserverbar kompetanse til Storbritannia er konsistent med modellen, men ikke til Tyskland. Sannsynligheten for å emigrere til Storbritannia faller med nivået på den uobserverte kompetansen, mens sannsynligheten for å emigrere til Tyskland ikke øker med nivået på den uobserverte kompetansen. Rosso (2019) estimerer at Tyskland har høyere og Storbritannia har lavere avkastning av uobserverbar kompetanse enn Polen.

kompetansenivå enn ikke-emigrantene. Funnene til studiene er i tråd med prediksjonene til Roy-modellen, siden nordiske land har relativt lav avkastning av kompetanse.

Litteraturen om seleksjonsprosessen til emigranter andre steder i verden finner stort sett funn som er i tråd med prediksjonene til Roy-modellen. To av studiene, som ikke får resultater i tråd med det teoretiske rammeverket, studerer emigrasjonsstrømmer hvor teorien predikerer negativ seleksjon (Chiquair og Hanson 2005; Rosso 2019).

3 Teoretisk rammeverk

3.1 Positiv og negativ seleksjon

Vi fokuserer på den økonomiske beslutningen om å emigrere og i så fall hvor. Inntekten til et individ bestemmes av kompetansenivået og den avkastningen individet får av den i bostedslandet. En person vil ønske å flytte til utlandet hvis det gir høyere avkastning av kompetansen og flyttekostnadene blir dekket. Avkastningen av kompetanse er ulik mellom land. Det er derfor ikke tilfeldig hvem som tjener på å emigrere. Av den grunn vil vi på populasjonsnivå se positiv eller negativ seleksjon. Hvis emigrantene er positivt selektert tilhører de den øvre delen av kompetansefordelingen i hjemlandet, og den nedre delen av kompetansefordelingen hvis de er negativt selektert. To betingelser må være oppfylt for at positiv seleksjon skal skje: (a) Det må være en positiv korrelasjon mellom inntektsmulighetene i hjemlandet og destinasjonslandet. (b) Destinasjonslandet må ha en mer spredt inntektsfordeling enn hjemlandet. Dersom betingelse (a) er oppfylt, men ikke (b) vil negativ seleksjon være tilfellet (Borjas 1987). Inntektsulikheten i et land er en proxy på avkastningen av kompetanse (Borjas 2020).⁸ I dette rammeverket er det altså avkastningen av kompetanse som avgjør om emigrantene er positivt eller negativt selektert fra populasjonen, gitt at kompetanse er tilstrekkelig overførbart mellom land.

Norge har vært karakterisert ved en sammenpresset lønnsfordeling som impliserer at Norge har relativt lav avkastning av kompetanse (Bratsberg et al. 2014). Hanushek et al. (2015) viser også at Norge har lavere avkastning av kognitiv kompetanse enn en rekke OECD-land. Det teoretiske rammeverket predikerer dermed at emigrasjonsstrømmen ut av Norge hovedsakelig vil være positivt selektert. Det vil ikke nødvendigvis være lik seleksjon til alle land. I dette rammeverket er det forskjellen i avkastningen av kompetanse mellom hjemlandet og destinasjonslandet som avgjør hvilken seleksjon teorien predikerer. Hvis et land har høyere avkastning av kompetanse enn Norge vil teorien predikere positiv seleksjon, og negativ seleksjon hvis landet har lavere avkastning av kompetanse enn Norge. I USA er dette tydeligere. USA har en mer spredt

⁸ Når et land har liten spredning i inntektsfordelingen vil det indikere at landet har lav avkastning av kompetanse, siden det er små forskjeller mellom individer med ulikt kompetansenivå. Når inntektsulikheten øker vil det bety at inntektsforskjellen mellom individer med ulikt kompetansenivå stiger. Det taler for at inntektsulikhet kan brukes som en proxy på avkastningen av kompetanse i et land.

inntektsfordeling enn nordiske land. Nordiske emigranter som flytter til USA vil derfor typisk være positivt selektert. På den andre siden mottar USA negativt selekterte immigranter fra Latin- og Sør-Amerika, hvor inntektsulikheten er større enn i USA. I neste delkapittel gjør jeg rede for en formell framstilling av modellen for seleksjon av migranter, basert på Borjas et al. (2019).

3.2 Roy-modellen

Borjas (1987) var den første til å anvende modellen til Roy (1951) til å analysere selv-seleksjon av migranter, og Borjas et al. (2019) tar utgangspunkt i denne. Roy-modellen beskriver positiv og negativ seleksjon. En implikasjon av modellen er at positivt selekterte emigranter vil ha høyere gjennomsnittlig kompetanse enn populasjonen i landet de emigrerer fra. Det nye i Borjas et al. (2019) er at de også viser at modellen impliserer at emigrantene sin kompetansefordeling vil førsteordens stokastisk dominere kompetansefordelingen til ikke-emigrantene når emigrantene er positivt selektert.

I modellen er det to land. Innbyggerne i hjemlandet (land 0) kan velge å bli værende eller emigrere til utlandet (land 1).⁹ Vi antar at beslutningen om å emigrere er irreversibelt. Videre ser vi bort fra flyttekostnader, og at det er forventet livtidsinntekt individene vil vurdere. Innbyggerne i hjemlandet står overfor følgende inntektsfordeling¹⁰:

$$\log w_0 = \alpha_0 + r_0 s + \varepsilon_0 \quad (1)$$

Hvor w_0 er lønnen i hjemlandet, r_0 er avkastningen på den observerte kompetansen og den tilfeldige variabelen ε_0 måler uobserverbare forhold som er antatt å være normalfordelt med forventning lik null og varians σ_0^2 . Hvis det er liten spredning i ε_0 vil det bety at lønnen til individene hovedsakelig avhenger av avkastningen på den observerte kompetansen. Jo større spredning det er i ε_0 desto mer avhenger lønnen til individene av uobserverbare forhold. Vi kan derfor tolke spredningen til den tilfeldige variabelen (residualen) ε_0 som prisen på uobservert kompetanse i land 0. Fordelingen av den observerbare kompetansen til populasjonen i hjemlandet er gitt ved $s = \mu_s + \varepsilon_s$. Hvor den tilfeldige variabelen ε_s er antatt å være normalfordelt med forventning lik null og varians σ_s^2 .

⁹ Roy (1951) studerer hvordan individer selekterer seg til ulike yrker.

¹⁰ For å forenkle modellen har Borjas et al. (2019) valgt å ikke ha med fotskrift i i modellen. Jeg velger å gjøre det samme.

Hvis hele populasjonen i hjemlandet (land 0) emigrerte til utlandet (land 1) vil denne populasjonen stå overfor følgende inntektsfordeling:

$$\log w_1 = \alpha_1 + r_1 s + \varepsilon_1 \quad (2)$$

Hvor den tilfeldige variabelen ε_1 er antatt å være normalfordelt med forventning lik null og varians σ_1^2 . Vi antar at $\text{Cov}(\varepsilon_0, \varepsilon_s) = \text{Cov}(\varepsilon_1, \varepsilon_s) = 0$. Det betyr at uobservert kompetanse er ukorrelert med observert kompetanse.¹¹ Korrelasjonskoeffisienten mellom ε_0 og ε_1 er ρ_{01} . Den tilfeldige variabelen ε_s er individspesifikk og har samme verdi for et individ i begge landene. ε_0 og ε_1 er både individ- og landspesifikk.

Vi antar at ligning (1) og (2) beskriver inntektsmulighetene til et individ født i hjemlandet (land 0) fullstendig. Videre antar vi at en beslutning om å emigrere avhenger av forskjellen i avkastningen av observert og uobservert kompetanse mellom de to landene.¹² Basert på ligning (1) og (2) definerer vi indeksfunksjonen:

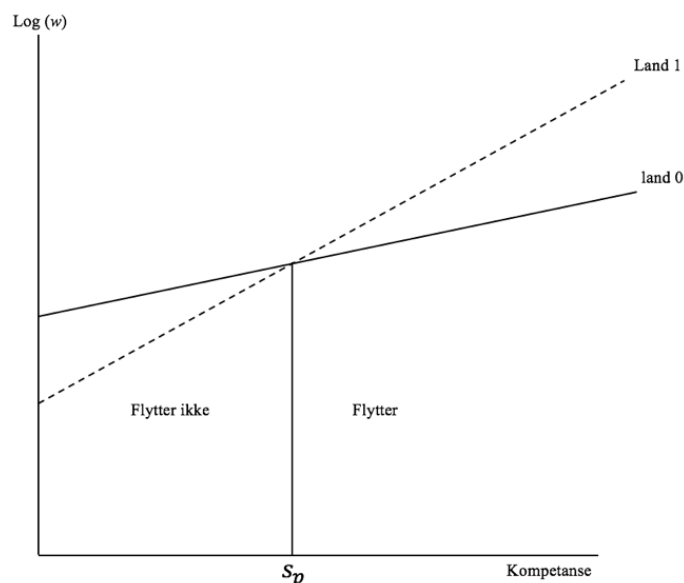
$$\begin{aligned} I = \log\left(\frac{w_1}{w_0}\right) &\approx [(\alpha_1 - \alpha_0) + (r_1 - r_0)\mu_s] + [(r_1\varepsilon_s + \varepsilon_1) - (r_0\varepsilon_s + \varepsilon_0)] \\ &= \Delta\mu + (v_1 - v_0) \end{aligned} \quad (3)$$

Inntektsforskjellen mellom de to landene for et individ med gjennomsnittlig observert og uobservert kompetanse er gitt ved $\Delta\mu = [(\alpha_1 - \alpha_0) + (r_1 - r_0)\mu_s]$. Forskjellen i inntekt som skyldes individuelle avvik fra gjennomsnittlig kompetanse er gitt ved $(v_1 - v_0)$, hvor $v_i = (r_i\varepsilon_s + \varepsilon_i)$ for $i \in \{0, 1\}$. En person emigrerer hvis, og bare hvis, indeksen $I > 0$.

¹¹ En mer realistisk antakelse ville vært å anta positiv korrelasjon mellom observert og uobservert kompetanse. En antakelse om positiv korrelasjon ville ikke endret de kvalitative prediksjonene til modellen.

¹² Flyttekostnaden til et individ er med på å avgjøre flyttebeslutningen til individet, men flyttekostnaden påvirker ikke seleksjonen modellen predikerer på populasjonsnivå hvis flyttekostnaden er konstant mellom individene eller ukorrelert med kompetanse (Borjas (1987); Chiquiar og Hanson (2005)). Flyttekostnaden er tatt ut av modellen for enkelhets skyld.

Figur 1



Figur 1 er en enkel illustrasjon av forholdet mellom inntekt og kompetanse i henholdsvis land 0 og 1. Figuren illustrerer et scenario der land 1 har høyest avkastning av kompetanse (inntektsulikheten er større). Individer i land 0 med kompetanse til venstre for S_p får lavere inntekt i land 1, og individer i land 0 med kompetanse til høyre for S_p får høyere inntekt i land 1. Individene i land 0 som tjener på å emigrere til land 1 tilhører den øvre delen av kompetansefordelingen. Det betyr at de er positivt selektert. En illustrasjon av negativ emigrantseleksjon er vist i appendikset.

Modellen skiller mellom seleksjon på observert og uobservert kompetanse. La $F_M(z)$ og $F_N(z)$ representere den kumulative fordelingsfunksjonen av kompetanse for emigranter og ikke-emigranter i hjemlandet (land 0), hvor z er et mål på kompetanse. Den kumulative fordelingsfunksjonen til emigrantene $F_M(z)$ førsteordens stokastisk dominerer fordelingen til ikke-emigrantene $F_N(z)$ hvis:

$$F_M(z) \leq F_N(z) \quad \forall z \tag{4}$$

og det er minst en verdi av z hvor streng ulikhet holder. Ligning (4) impliserer at en større andel av emigrantene har kompetanse over enhver terskelverdi z^* . Med andre ord, for ethvert nivå med kompetanse z^* , så har emigrantene alltid en større andel med kompetanse over terskelverdien z^* enn ikke-emigrantene. Grafisk vil dette si at den kumulative

fordelingsfunksjonen til emigrantene ligger til høyre for fordelingen til ikke-emigrantene over hele fordelingen. Vi kan si at emigrantene er positivt selektert uansett hva en setter som terskelverdi. Dette er intuisjonen bak stokastisk dominans prediksjonen til Roy-modellen. Negativ seleksjon ville vært tilfellet dersom $F_N(z) \leq F_M(z) \forall z$, hvor streng ulikhet holder for minst en verdi av z . Det kan vises at:

$$\begin{aligned} F_M(z) &< F_N(z) \text{ hvis } \rho > 0 \\ F_M(z) &> F_N(z) \text{ hvis } \rho < 0 \end{aligned} \quad (5)$$

hvor ρ er den relevante korrelasjonskoeffisienten. Emigrantene er positivt selektert når $\rho > 0$, og er negativt selektert når $\rho < 0$. Borjas et al. (2019) viser at når det gjelder observerbar kompetanse er den relevante korrelasjonskoeffisienten for den stokastiske rangeringen av kompetansefordelingen til emigrantene og ikke-emigrantene gitt ved:

$$\rho = \text{Corr}(\varepsilon_s, v_1 - v_0) = \frac{r_0 \sigma_s}{\sigma_v} \left(\frac{r_1}{r_0} - 1 \right) \quad (6)^{13}$$

Ligning (6) viser at den stokastiske rangeringen av fordelingene kun avhenger av de internasjonale forskjellene i avkastningen av observerbar kompetanse. Den kumulative fordelingsfunksjonen av observert kompetanse til emigrantene førsteordens stokastisk dominerer fordelingen til ikke-emigrantene når avkastningen av observerbar kompetanse er høyere i utlandet. Dersom avkastningen er høyere hjemme, vil det motsatte være tilfellet.

Når det gjelder uobserverbar kompetanse kan det vises at den relevante korrelasjonskoeffisienten for den stokastiske rangeringen av kompetansefordelingen til emigrantene og ikke-emigrantene er gitt ved:

$$\rho = \text{Corr}(\varepsilon_0, v_1 - v_0) = \frac{\sigma_0}{\sigma_v} \left(\rho_{01} \frac{\sigma_1}{\sigma_0} - 1 \right) \quad (7)^{14}$$

Den kumulative fordelingsfunksjonen av uobserverbar kompetanse til emigrantene førsteordens stokastisk dominerer fordelingen til ikke-emigrantene når $\rho_{01} \frac{\sigma_1}{\sigma_0} > 1$. Hvis denne

¹³ $\sigma_v = \sqrt{\text{var}(v_1 - v_0)}$. r_0 og r_1 er avkastningen av observert kompetanse i henholdsvis land 0 og 1. σ_s er standardavviket til ε_s .

¹⁴ σ_0 og σ_1 er standardavviket til henholdsvis ε_0 og ε_1 , og ρ_{01} er korrelasjonskoeffisienten mellom ε_0 og ε_1 .

betingelsen er oppfylt vil emigrantene være positivt selektert på uobservert kompetanse. Betingelsen har to komponenter. For det første må uobserverbar kompetanse være tilstrekkelig overførbart mellom landene, slik at ρ_{01} er høy. For det andre må variansen til ε_1 være større enn variansen til ε_0 . Når variansen til ε_1 er større enn variansen til ε_0 vil det indikere at avkastningen av uobserverbar kompetanse er høyere i utlandet. Gitt at avkastningen av uobservert kompetanse er høyere i utlandet vil emigrantene være positivt selektert på uobservert kompetanse hvis korrelasjonskoeffisienten ρ_{01} er tilstrekkelig høy.

Inntektsfordelingen avhenger av både observert og uobservert kompetanse, slik at stokastisk dominans i inntektsfordelingen kombinerer de to elementene vi har sett på til nå. Inntektsfordelingen i hjemlandet som er gitt ved ligning (1) kan skrives om til:

$$\log w_0 = (\alpha_0 + r_0\mu_s) + (r_0\varepsilon_s + \varepsilon_0) = (\alpha_0 + r_0\mu_s) + v_0 \quad (8)$$

Hvor den tilfeldige variabelen v_0 er antatt å være normalfordelt med forventning lik null og varians $\sigma_{v_0}^2$.

Det kan vises at den relevante korrelasjonskoeffisienten, når det gjelder den stokastiske rangeringen av inntektsfordelingen til emigrantene og ikke-emigrantene, er gitt ved:

$$\rho = \text{Corr}(v_0, v_1 - v_0) = \frac{\sigma_{v_0}}{\sigma_v} \left[\gamma \left(\frac{r_1}{r_0} - 1 \right) + (1 - \gamma) \left(\rho_{01} \frac{\sigma_1}{\sigma_0} - 1 \right) \right] \quad (9)$$

$$\text{hvor } \gamma = \frac{r_0^2 \sigma_s^2}{\sigma_{v_0}^2} \text{ og } 1 - \gamma = \frac{\sigma_0^2}{\sigma_{v_0}^2}$$

Fortegnet på korrelasjonskoeffisienten i ligning (9), som bestemmer seleksjonen på inntekt, avhenger av fortegnet på et vektet gjennomsnitt av seleksjonen som forekommer på observert og uobservert kompetanse. Vektingen bestemmes av hvilken grad de to kompetansetyperne bidrar til å skape varians i inntektsfordelingen. Dersom det er positiv (negativ) seleksjon på både observerbar og uobserverbar kompetanse vil det også være positiv (negativ) seleksjon på inntekt. Hvis det derimot er forskjellig type seleksjon i de to typene med kompetanse, blir seleksjonen på inntekt bestemt av det vektete gjennomsnittet av de to kompetansetyperne. Observerbare egenskaper forklarer stort sett lite av variansen i inntektsfordelingen (Borjas et

al. 2019; Lemieux 2006).¹⁵ Det impliserer at seleksjonen på uobserverbar kompetanse må være viktig for hvordan emigrantene er selektert på inntekt.

Modellen inneholder mange ukjente parametere, som bestemmer seleksjon og stokastisk dominans. For å undersøke alle sider av modellen vil vi trenge data om både avsender- og destinasjonsland. Det har vi typisk ikke. Modellen gir likevel noen prediksjoner som kan sjekkes for å undersøke om modellen er god. Ved positiv seleksjon skal emigrasjon være positivt korrelert med utdanning og inntekt i avsenderlandet. Det kan sjekkes ved å se på gjennomsnittsverdier for emigranter og ikke-emigranter. I tillegg kommer stokastisk dominans prediksjonen til modellen. Resten av oppgaven vil hovedsakelig handle om å undersøke disse prediksjonene på norske data.

¹⁵ I en lønnslikning vil variasjon i observert kompetanse (og andre observerte egenskaper) stort sett forklare lite av variasjonen i inntekt.

4 Data og utvalgsdesign

4.1 Data

Oppgaven bygger på registerdata fra Statistisk sentralbyrå (SSB) sin database FD-trygd. Statistikkloven pålegger offentlige institusjoner til å sende inn statistikk til SSB (Statistikkloven, § 10, 2019). Det medfører at inntektsopplysninger, informasjon om bakgrunnsvariabler og flyttebevegelser blir rapportert inn til SSB. For at denne oppgaven skal kunne rapportere troverdige resultater, er jeg avhengig av at emigrasjon ut av Norge blir rapportert inn til myndighetene. Det er mulig å emigrere ut av Norge uten å melde ifra. Det er grunn til å tro at det gjelder få individer, ettersom det er lovpliktig å melde inn flyttebevegelser til norske skattemyndigheter når utenlandsoppholdet har en varighet på seks måneder eller mer. En er også pliktig til å melde ifra til myndighetene når man flytter tilbake til Norge (Folkeregisterloven, § 8, 1970; Folkeregisterloven, § 6-3, 2016).

Individene er identifisert med en personlig ID i registerdataene. Det medfører at statistikk fra mange ulike registre kan kobles sammen. Registerdataenes viktigste fortrinn sammenlignet med utvalgsundersøkelser er at registerdataene ikke har frafall, de fanger opp hele befolkningen og opplysningene er objektive. Registerdata er dermed et svært godt utgangspunkt for statistisk analyse. Registerdataene er ikke nødvendigvis helt feilfrie. Det kan være problemer med manglende registreringer, og kvaliteten på registreringene kan variere med anvendelsesgraden hos dataeierne (Raaum et al. 2012).

Registerdataene jeg har til rådighet var utlevert til Institutt for økonomi ved UiB i et tidligere prosjekt og er siden anonymisert slik at studenter kan benytte dataene. Flere variabler har blitt aggregert slik at individer ikke kan bli gjenkjent. Variablene i datasettet som er aggregert er variabler for inntekt, utdanning og destinasjonsland. I det anonymiserte datasettet oppgis årlig inntekt i enheter, hvor 50 000 kr er en enhet, og inntekten er trunkert nedover.¹⁶ Destinasjonslandene til emigrantene er slått sammen til ni geografiske områder.¹⁷ Videre

¹⁶ Under aggregeringen har Institutt for økonomi ved UiB trunkert inntektsvariablene nedover. Årlig inntekt under 50 000 kr blir registrert som 0 kr, årlig inntekt fra 50 000 kr til 99 999 kr blir registrert som 50 000 kr og årlig inntekt fra 100 000 kr til 149 999 kr blir registrert som 100 000 kr. Slik er gangen i trunkeringen av inntektsvariablene. Inntektsvariablene har en toppkoding på 1 million (før deflatering). Det innebærer at observasjoner med årlig inntekt på over 1 million blir registrert som at de har 1 million i årlig inntekt (før deflatering).

¹⁷ De ni geografiske områdene er Norden, Vest-Europa, Øst-Europa, Midtøsten (inkludert Tyrkia), Sentral-Asia, Afrika, Nord-Amerika, Sør-Amerika (inkludert Mexico og Puerto Rico) og Oseania.

medfører anonymiseringen at jeg ikke kan se den spesifikke utdanningen til individet, men utdanningsnivå og utdanningsgruppe på et overordnet nivå (av SSBs 6-sifrede kode er de 3 første beholdt). Det er også gjort en del aggregering av datovariabler.

Det viste seg at en del emigranter har ukjent destinasjonsland i datasettet. Det er totalt 18 551 emigranter i utvalget, hvor rundt 25 % av disse har ukjent destinasjonsland. Årsaken til denne mangelfulle kvaliteten i datasettet kjenner jeg ikke. Det medfører at jeg ikke kan bruke variabelen som angir hvilket geografisk område emigrantene flytter til når jeg skal analysere seleksjonsprosessen til dem.

4.2 Utvalgsdesign

Datasettet, oppgaven benytter, inneholder årlige observasjoner fra perioden 1993-2003 av norskfødte individer uten innvandrerbakgrunn i alderen 25-54 år.¹⁸ Utvalgsdesignet mitt er forholdsvis likt det Borjas et al. (2019) anvender. Innvandrere utelukkes for å unngå at analysen blir preget av innvandrere som tilbakevendte til sine hjemland. Innvandrere har ofte en sterk tilknytning til et annet land. Motivet for å emigrere kan derfor være knyttet til andre grunner enn for å oppnå høyere avkastning av sin kompetanse.

Roy-modellen predikerer at det er avkastningen av kompetanse som avgjør seleksjonen blant emigrantene. Det er derfor vesentlig at individene som inkluderes i analysen hadde en tilknytning til arbeidsmarkedet, og helst ikke en for svak tilknytning. Inntektsopplysningene til individer som jobber deltid viser ikke nødvendigvis det sanne inntekspotensialet til disse individene (Borjas et al. 2019). Av den grunn ville det vært ideelt om jeg kunne avgrense analysen til individer som jobbet fulltid, eller bortimot fulltid. Jeg har ikke data på arbeidstiden til individene. Av den grunn utføres en inntektsavgrensning for å luke ut individer (observasjoner) som hadde svak tilknytning til arbeidsmarkedet. Kapitalinntekter og noen typer overføringer er ikke direkte knyttet til hvilken inntekt et individ oppnår på arbeidsmarkedet. Lønnsinntekten til et individ omfatter ikke næringsinntekter individet har. Derfor bruker jeg yrkesinntekt som mål på individenes inntekt i denne oppgaven. Yrkesinntekten er summen av lønnsinntekter og netto næringsinntekter i løpet av et kalenderår. Sykepenger og fødselspenger er inkludert (SSB 2021b). Videre i oppgaven vil yrkesinntekt bli omtalt som inntekt.

¹⁸ Datasettet jeg fikk utdelt av Institutt for økonomi ved UiB strekker seg over perioden 1993-2008. Jeg skiller mellom korttids- og langtidsemigranter i deler av analysen. En langtidsemigrant bor i utlandet i minst fem år. For å kunne klassifisere en emigrant som korttids- eller langtidsemigrant, må jeg ha kontroll på om emigrantene har tilbakevendt før det har gått fem år. Av den grunn kan jeg kun bruke emigrasjonsdata til og med 2003.

For å få et rimelig kriterium for en inntekt som indikerer fulltidsinntekt, tar jeg utgangspunkt i SSBs gjennomsnittsinntekt for fulltidsarbeidere i første desil i lønnsfordelingen i 1997, som er det første året jeg fant denne gjennomsnittsinntekten i SSBs statistikkbank (SSB 2021a).¹⁹ I 2015-kroner er dette 222 043. Under anonymiseringen ble inntektsvariablene aggregert og trunkert nedover. På grunn av denne trunkeringen velger jeg å runde av inntektskriteriet nedover. Jeg inkluderer derfor individer i det enkelte året hvis de har inntekt over 200 000 i 2015-kroner.²⁰

I likhet med Borjas et al. (2019) avgrenses utvalget til individer som tilhører aldersspennet 25-54 år. Inklusjonskriteriet fungerer på den måten at observasjonene til et individ er inkludert i utvalget i de årene individet befant seg i aldersspennet 25-54 år. Dette aldersspennet blir omtalt av OECD som «gullaldersårene» når det gjelder deltakelse på arbeidsmarkedet. Individer under 25 år er på vei inn i arbeidsstyrken og individer over 54 år har gjerne oppnådd toppen av karrieren og nærmer seg pensjonsalderen (OECD 2021). I denne sammenhengen vil jeg si det er viktigst at en ekskluderer unge mennesker i tidlig 20-årene. Mange unge mennesker er under utdanning, eller planlegger å ta en utdanning. En beslutning om å emigrere kan derfor være drevet av andre faktorer enn for å oppnå høyere avkastning på sin nåværende kompetanse. For eksempel at unge mennesker flytter til utlandet for å studere. Til slutt vil jeg presisere at emigranter som tilbakevendte til Norge ikke blir inkludert i utvalget igjen.

Det opprinnelige datasettet inneholdt 17 047 869 observasjoner, og 2 166 474 individer, for den relevante aldersgruppen.²¹ Alle observasjoner med ugyldig informasjon om bakgrunnsvariabler ble fjernet fra datasettet. Observasjonene som ble fjernet var de med negativ yrkesinntekt, manglende data om familiestatus og de som er registrert med null års utdanning. 1939-kullet ble også fjernet, siden individer fra det årskullet viste seg å mangle data om inntekt. Antall observasjoner ble redusert til 16 393 464 (2 070 587 individer) på grunn av dette. Videre reduserte inntektsavgrensningen antall observasjoner til 10 819 802 (1 673 276 individer). Før

¹⁹ En for høy inntektsavgrensning kan medføre at individer som jobbet fulltid, og tilhørte den nedre delen av lønnsfordelingen, blir ekskludert fra utvalget. På den andre siden vil en for lav inntektsavgrensning medføre at individer som hadde svak tilknytning til arbeidsmarkedet blir inkludert i utvalget. Jeg mener det er mest problematisk å ende opp med å ekskludere individer med svært lave lønninger, men som jobbet fulltid eller bortimot fulltid. Derfor tar jeg utgangspunkt i første desil av lønnsfordelingen.

²⁰ Ettersom jeg runder av nedover har det ikke noe å si om jeg har en felles eller separat inntektsgrense for menn og kvinner, siden det er liten forskjell mellom menn og kvinner i første desil av lønnsfordelingen i 1997.

²¹ Datasettet var allerede avgrenset på alder og innvandrerbakgrunn da jeg mottok datasettet fra Institutt for økonomi ved UiB.

inntektsavgrensningen er 50,91 % av observasjonene menn, og 49,09 % er kvinner. Etter inntektsavgrensninger er 61,73 % av observasjonene menn og 38,27 % er kvinner.²²

Oppsummert består mitt utvalg av observasjoner med en alder i spennet 25-54 år, med en inntekt på 200 000 eller mer (2015-kroner), og uten en innvandrerbakgrunn. Det blir totalt 10 819 802 observasjoner fordelt over årene 1993-2002. Variabelen som angir om et individ emigrerer angir emigrasjon påfølgende år, fordi vi ikke har fullstendige inntektsdata det året individet emigrerer. Emigrasjonsvariabelen viser derfor emigrasjon fra og med 1994 til og med 2003. Andre variabler er målt fra og med 1993 til og med 2002.

4.3 Deskriptiv statistikk

Ettersom en del emigranter har ukjent destinasjonsland vil hovedanalysen fokusere på forskjellene mellom emigranter og ikke-emigranter. Jeg undersøker også hvorvidt korttids- og langtidsemigranter skiller seg fra hverandre.²³ Jeg følger Borjas et al. (2019) når jeg deler emigrantene inn i de to gruppene. En korttidsemigrant bor i utlandet i mindre enn fem år, og en langtidsemigrant bor i utlandet i fem år eller mer. Ifølge NOU (2008:6) er norsk arbeidsliv preget av å være kjønnssegregert, det vil si at menn og kvinner arbeider i ulike sektorer og i ulike yrker. Med bakgrunn i at menn og kvinner har ulik atferd på arbeidsmarkedet vil det være naturlig å studere dem separat. Jeg bruker standardisert inntekt i deler av analysen for å ta hensyn til års- og alderseffekter. Den standardiserte inntekten er forholdstallet mellom et individ sin inntekt og gjennomsnittlig inntekt til individer med lik alder og kjønn i et kalenderår. Standardisert inntekt = $\frac{w_{itkj}}{\bar{w}_{tkj}}$. w_{itkj} er inntekt til individ i i kalenderår t som tilhører alder k og kjønn j . \bar{w}_{tkj} er gjennomsnittlig inntekt til individer med alder k og kjønn j i kalenderår t . Standardiseringen av inntekten ble utført etter at observasjoner med ugyldig informasjon om bakgrunnsvariabler ble fjernet fra datasettet. Inntektsavgrensningen ble gjennomført etter standardiseringen av inntekten.

²² I kapittel A.7 vises en sensitivitetsanalyse for kvinner hvor inntektsgrensen er satt til 150 000 2015-kroner. Resultatene til sensitivitetsanalysen viser samme mønster som hovedanalysen.

²³ Borjas og Bratsberg (1996) viser empirisk og teoretisk at kompetansekomposisjonen til emigranter som tilbakevender til sine hjemland avhenger av hvilken type seleksjon som var tilfellet i første omgang. Det kan derfor være interessant å undersøke om det er forskjeller mellom korttids- og langtidsemigranter i denne oppgaven.

Tabell 1

Deskriptiv statistikk for menn

	Alle	Ikke- emigranter	Emigranter	Korttids- emigranter	Langtids- emigranter
Observasjoner	6 678 597	6 666 868	11 729	7 654	4 075
Alder					
Gjennomsnitt	39,56	39,57	36,97	36,91	37,09
Median	40	40	36	36	35
Årlig inntekt					
Gjennomsnitt	8,31	8,30	10,82	11,03	10,42
Median	7,33	7,33	9,27	10,01	8,89
Standardisert årlig inntekt					
Gjennomsnitt	1,20	1,20	1,61	1,65	1,54
Median	1,07	1,07	1,46	1,52	1,35
Utdanning (år)					
Gjennomsnitt	13,06	13,06	15	15,42	14,23
Median	13	13	15	16	14
Utdanningsgrupper (%)					
Grunnskole	19,70	19,71	9,34	7,15	13,45
VGS, studieforbereidende	8,46	8,46	8,39	7,73	9,62
VGS, yrkesfag	39,53	39,57	18,52	15,34	24,49
Fagskole	4,76	4,76	5	4,64	5,67
Bachelor og tilsvarende	19,08	19,06	31,69	33,17	28,91
Master og tilsvarende	7,93	7,90	24,19	28,26	16,54
Forskerutdanning (PhD)	0,55	0,55	2,88	3,71	1,33

Merknad: Årlig inntekt måles i enheter, hvor en enhet er 50 000 kr. Årlig inntekt er deflatert til 2015-kroner. De to første kolonnene angir observasjoner, og de tre siste angir individer. En beskrivelse av inndelingen av utdanningskategoriene finnes i appendikset.

Tabell 1 viser deskriptiv statistikk for menn. De to første kolonnene angir observasjoner, og de tre siste kolonnene angir individer. Datasettet er satt opp slik at et individ blir registrert som en ikke-emigrant helt til individet eventuelt emigrerer. Dummyvariabelen som viser om et individ emigrerer tar verdien 1 året før individet emigrerer, og 0 alle tidligere år. Det vil si at datagrunnlaget for emigrantene er basert på data fra året før emigrantene emigrerer, mens datagrunnlaget for ikke-emigrantene er basert på data for alle individer som ikke emigrerer påfølgende år. Det er totalt 946 659 mannlige individer i utvalget. Basert på inklusjonskriteriene finner jeg at 11 729 av disse emigrerer i løpet av perioden. 7 654 av emigrantene er korttidsemigranter, og 4 075 er langtids- emigranter.

De mannlige emigrantene er yngre enn ikke-emigrantene. Det passer godt overens med økonomisk teori at emigrantene er yngre. Yngre mennesker vil i større grad få en positiv netto migrasjonsfortjeneste av å emigrere siden de har et lengre yrkesliv foran seg. Dermed har de flere år å høste gevinst av og dele flyttekostnadene på (Boeri og Van Ours 2013). Yngre mennesker vil trolig ha lavere flyttekostnader ettersom de har et mindre etablert privatliv. Videre viser tabell 1 at mannlige emigranter har høyere inntekt, både standardisert og ikke-standardisert, enn ikke-emigrantene. Det samme mønsteret gjelder for utdanning, hvor emigrantene har i gjennomsnitt nesten to år lengre utdanning enn ikke-emigrantene. Tabellen presenterer også utdanningsfordelingen hvor det fremkommer at 27,51 % av ikke-emigrantene har bachelorgrad eller høyere. For emigrantene samlet er andelen henholdsvis 58,76 %.

Tabell 2

Deskriptiv statistikk for kvinner

	Alle	Ikke- emigranter	Emigranter	Korttids- emigranter	Langtids- emigranter
Observasjoner	4 141 205	4 134 383	6 822	4 571	2 251
Alder					
Gjennomsnitt	39,77	39,78	35,6	35,67	35,47
Median	40	40	34	34	33
Årlig inntekt					
Gjennomsnitt	6,16	6,16	6,86	6,87	6,85
Median	5,72	5,72	6,09	6,09	6,09
Standardisert årlig inntekt					
Gjennomsnitt	1,57	1,57	1,84	1,85	1,80
Median	1,45	1,45	1,67	1,69	1,61
Utdanning (år)					
Gjennomsnitt	13,40	13,40	14,9	15,22	14,26
Median	13	13	15	16	14
Utdanningsgrupper (%)					
Grunnskole	15,84	15,85	7,93	5,95	11,95
VGS, studieforbereende	13,75	13,75	10,39	8,95	13,33
VGS, yrkesfag	27,99	28,01	14,86	12,89	18,88
Fagskole	3,04	3,03	3,34	3,15	3,73
Bachelor og tilsvarende	33,84	33,82	47,52	50,91	40,65
Master og tilsvarende	5,31	5,30	15,02	17,00	11,02
Forskerutdanning (PhD)	0,23	0,23	0,92	1,16	0,44

Merknad: Årlig inntekt måles i enheter, hvor en enhet er 50 000 kr. Årlig inntekt er deflatert til 2015 kroner. De to første kolonnene angir observasjoner, og de tre siste angir individer. En beskrivelse av inndelingen av utdanningskategoriene finnes i appendikset.

Tabell 2 viser deskriptiv statistikk for kvinner. De to første kolonnene angir observasjoner, og de tre siste kolonnene angir individer. Det er totalt 726 617 kvinnelige individer i utvalget. Basert på inklusjonskriteriene finner jeg at 6 822 av disse emigrerer i løpet av perioden. 4 571 av emigrantene er korttidsemigranter, og 2 251 er langtidsemigranter.

Den deskriptive statistikken for kvinnene har samme mønster som den til mennene. Kvinnelige emigranter er yngre og har høyere inntekt og utdanningsnivå enn ikke-emigranter. 39,35 % av de kvinnelige ikke-emigrantene har utdanning på bachelornivå eller høyere. Det er noe høyere enn for mennene hvor 27,51 % av ikke-emigrantene hadde bachelorgrad eller høyere. 63,46 % av de kvinnelige emigrantene har bachelorgrad eller høyere.

Kvinnene er på et høyere nivå på den standardiserte inntekten enn mennene når en sammenligner samme emigrantkategori mellom kjønnene. Grunnen til disse forskjellene skyldes inntektsavgrensningen. Gjennomsnittsinntekten (ikke-standardisert inntekt) til kvinnene er lavere enn gjennomsnittsinntekten til mennene. Inntektsavgrensningen medfører derfor at flere kvinner enn menn blir fjernet fra utvalget. Gjennomsnittsinntekten som er brukt til å regne ut kvinnenens standardiserte inntekt er lavere enn den som er brukt til å regne ut mennenes standardiserte inntekt. En felles inntektsavgrensning fører derfor til at kvinnene sin standardiserte inntekt blir «høyere» enn mennene sin.

Oppsummert viser den deskriptive statistikken i tabell 1 og 2 at emigrantene har mer utdanning og inntekt og lavere gjennomsnittsalder enn ikke-emigrantene. Korttidsemigrantene skiller seg ut som gruppen med høyest inntekts- og utdanningsnivå. Den deskriptive statistikken gir en klar indikasjon på at emigrantene er positivt selektert fra populasjonen.

Datasettet oppgaven benytter er et ubalansert panel. Tabell 3 og 4 gir en oversikt over panelet for henholdsvis menn og kvinner. I tabellene fremkommer det hvor mange individer som er med i panelet i hvert år, hvor det skilles mellom om individet er en ikke-emigrant eller emigrant. I tillegg gir tabellene en oversikt over andelen av individene som emigrerte i et år. Andelen emigranter avtar noe over tid for begge kjønn.

Tabell 3

Oversikt over panelet for menn

År	Antall individer	Antall individer, ikke-emigranter	Antall individer, emigranter	Emigranter %
1993	601 031	599 937	1 094	0,18
1994	635 355	634 224	1 131	0,18
1995	656 939	655 684	1 255	0,19
1996	674 199	672 814	1 385	0,21
1997	693 327	692 025	1 302	0,19
1998	709 978	708 806	1 172	0,17
1999	711 013	709 644	1 369	0,19
2000	667 097	665 977	1 120	0,17
2001	667 244	666 226	1 018	0,15
2002	662 414	661 531	883	0,13

Merknad: Variabelen som angir om et individ emigrerer angir emigrasjon påfølgende år. For eksempel de som er registrert som en emigrant i 1993 emigrerer i 1994.

Tabell 4

Oversikt over panelet for kvinner

År	Antall individer	Antall individer, ikke-emigranter	Antall individer, emigranter	Emigranter %
1993	324 622	323 982	640	0,20
1994	356 589	355 962	627	0,18
1995	382 436	381 765	671	0,18
1996	412 363	411 559	804	0,19
1997	441 477	440 646	831	0,19
1998	478 531	477 821	710	0,15
1999	501 398	500 546	852	0,17
2000	387 629	387 019	610	0,16
2001	415 378	414 808	570	0,14
2002	440 782	440 275	507	0,12

Merknad: Variabelen som angir om et individ emigrerer angir emigrasjon påfølgende år. For eksempel de som er registrert som en emigrant i 1993 emigrerer i 1994.

5 Metode

I kapittel 3 ble stokastisk dominans prediksjonen til Roy-modellen gjort rede for. Modellen skiller mellom observert og uobservert kompetanse. Modellen viser at seleksjonen på en kompetansetype alene bestemmes av forskjeller i avkastningen av den kompetansetypen mellom landene. Det kan derfor være ulik seleksjon på observert og uobservert kompetanse. Dersom emigrantene er positivt selektert på en kompetansetype predikerer modellen at emigrantene sin kumulative kompetansefordeling skal førsteordens stokastisk dominere fordelingen til ikke-emigrantene. Inntektsfordelingen avhenger av både observert og uobservert kompetanse, slik at stokastisk dominans i inntektsfordelingen kombinerer de to kompetansetypene.

Jeg skal undersøke hvordan emigrantene er selektert på observert og uobservert kompetanse samt inntekt. Antall år med utdanning rapporteres i hele tall, og det er store opphopninger på enkelte år. En kumulativ fordelingsfunksjon med antall år med utdanning som variabel blir derfor vanskelig å tolke siden vi ikke får s-formede fordelinger. Jeg vil derfor fokusere på inntekt og uobservert kompetanse når jeg tester stokastisk dominans prediksjonen til Roy-modellen. For å analysere seleksjon på observert kompetanse bruker jeg en regresjonsmodell til å utforske hvordan utdanningsnivå og emigrasjon henger sammen. Det er bare i analysen av observert og uobservert kompetanse jeg bruker regresjonsmodeller. Til analysen av seleksjon på inntekt vises de kumulative inntektsfordelingene til emigrantene og ikke-emigrantene for å undersøke hvordan emigrantene er selektert på inntekt.

5.1 Seleksjon på observerbar kompetanse

Jeg ønsker å undersøke hvordan utdanningsnivå og tilbøyeligheten til å emigrere henger sammen. En regresjonsmodell med emigrasjonsstatus som avhengig variabel vil være en diskret valgmodell, ettersom emigrasjonsstatus er en binær variabel som tar verdien 0 eller 1. Jeg velger å estimere denne med logistisk regresjon. Koeffisientene til logit-modellen er vanskelig å tolke direkte, derfor regnes koeffisientene om til marginaleffekter når resultatene rapporteres. Jeg estimerer følgende ligning:

$$Pr(Y_{it} = 1 | \mathbf{X}_{it}, \sum_{k=25}^{54} \alpha_{itk}, \eta_t) = F(\gamma \mathbf{X}_{it} + \sum_{k=25}^{54} \alpha_{itk} + \eta_t) \quad (10)$$

Antar at sannsynlighetsfordelingen til F er en logistisk fordeling. Den avhengige variabelen er en dummyvariabel som er lik 1 dersom et individ emigrerer ut av Norge påfølgende år, og 0 ellers. \mathbf{X}_{it} er en vektor med dummyvariabler for de ulike utdanningsgruppene og bakgrunnsvariabler. Dummyvariabelen for en utdanningskategori tar verdien 1 dersom individet tilhører den utdanningskategorien, og 0 ellers. Det er inkludert en slik dummyvariabel for alle utdanningskategoriene som ble introdusert under databeskrivelsen, hvor grunnskoleutdanning er valgt som referansekategori. Det er inkludert bakgrunnsvariabler som viser om individet har barn i husstanden og om individet er gift i det enkelte året.²⁴ Barn er en dummyvariabel som er lik 1 dersom individet har ett eller flere barn under 18 år i husstanden, og 0 ellers. Gift er en dummyvariabel som er lik 1 dersom individet er gift, og 0 ellers. Disse bakgrunnsvariablene er inkludert fordi de er korrelert med utdanning og trolig har selvstendig forklaringskraft på utfallsvariabler som inntekt og emigrasjonsstatus.

Aldersfaste effekter (α_{itk}) er inkludert ved å bruke dummyvariabler. Det er en dummyvariabel for hvert aldersår individene i datasettet har. En slik dummyvariabel tar verdien 1 dersom individ i har alder k i år t , og 0 ellers. Tidsfaste effekter (η_t) er også inkludert i ligningen ved å bruke dummyvariabler, hvor det er en dummyvariabel for hvert kalenderår som er med i datasettet. En slik dummyvariabel tar verdien 1 dersom året er år t , og 0 ellers.²⁵ Aldersfaste effekter er inkludert for å kontrollere for arbeidserfaring. Tidsfaste effekter er inkludert for å kontrollere for uobserverte forhold som er konstant på tvers av individene, men som endrer seg over tid. Ettersom datasettet er paneldata vil et individ potensielt bli observert flere ganger i panelet. Observasjonene til et individ er trolig korrelert over tid. For å ta høyde for det benytter jeg klyngerobuste standardfeil som er gruppert på individ. Observasjonene for et individ blir da definert som en klynge. Det kan være problematisk å bruke klyngerobuste standardfeil hvis det er få klynger (Angrist og Pischke 2009). Det er ikke et problem i denne oppgaven.

5.2 Seleksjon på uobserverbar kompetanse

Når det gjelder testing av stokastisk dominans prediksjonen til Roy-modellen vedrørende uobservert kompetanse bruker jeg residualer fra lønnslikninger (residuallønninger) som mål på uobservert kompetanse. Jeg estimerer to lønnslikninger. Den første lønnslikningen, som er ligning (11), har ikke-standardisert inntekt på logaritmeform som avhengig variabel. Den andre

²⁴ Siden jeg i stor grad replikerer analysen til Borjas et al. (2019) velger jeg å inkludere de samme bakgrunnsvariablene som dem.

²⁵ Ett aldersår og ett kalenderår blir valgt ut som referansekategori av Stata.

lønnslikningen, som er ligning (12), bruker standardisert inntekt på logaritmeform som avhengig variabel. Begge lønnslikningene estimeres ved å bruke minste kvadraters metode (MKM). Jeg estimerer følgende lønnslikning når ikke-standardisert inntekt på logaritmeform er avhengig variabel:

$$\text{Log}(Y_{it}) = \beta \mathbf{X}_{it} + \sum_{k=25}^{54} \alpha_{itk} + \eta_t + \varepsilon_{it} \quad (11)$$

Ligning (11) har samme oppsett som ligning (10). Jeg estimerer følgende lønnslikning når standardisert inntekt på logaritmeform er avhengig variabel:

$$\text{Log}(\tilde{Y}_{it}) = \tilde{\beta}_0 + \tilde{\beta}_1 \mathbf{X}_{it} + \varepsilon_{it} \quad (12)$$

Ligning (12) har samme oppsett som ligning (10), men aldersfaste og tidsfaste effekter er ikke inkludert. Jeg har to lønnslikninger, fordi det er to ulike måter å kontrollere for arbeidserfaring og årseffekter. Ligning (11) kontrollerer for arbeidserfaring og årseffekter ved hjelp av faste effekter, mens ligning (12) kontrollerer for det ved hjelp av den standardiserte inntekten. I tillegg benytter jeg den standardiserte inntekten når jeg tester stokastisk dominans prediksjonen til Roy-modellen på inntekt.

Når bakgrunnsvariablene (og eventuelt de faste effektene) er inkludert i regresjonen av inntekt på utdanning, vil det som står igjen som uforklart variasjon i den avhengige variabelen være variasjon som ikke kan forklares av variasjon i utdanning og de andre kontrollvariablene. Residuallønningen måler da avviket mellom den faktiske inntekten et individ oppnår og gjennomsnittsinntekten til individer med de samme observerbare karakteristikene (Joonas og Wadensjö 2013). I teorikapittelet ble det gjort rede for at seleksjonen på uobservert kompetanse avhenger av hvor stor grad uobservert kompetanse er overførbart over landegrensene og spredningen til fordelingen av residuallønningene. Hvis residuallønningene var lik null ville alle individer med like observerte karakteristikker oppnådd samme inntekt. Jo større spredning det er i residuallønningene desto mer variasjon er det i inntektsfordelingen til individer med like observerte karakteristikker. Uobserverte egenskaper er med å skape denne variasjonen i inntektsfordelingen. I likhet med Borjas et al. (2019) tolker jeg residualvariansen som «prisen» på uobservert kompetanse. Høyere avkastning av uobservert kompetanse vil bidra til større ulikhet i residuallønningene. Med bakgrunn i dette blir residuallønningene brukt som et mål på individenes uobserverte kompetanse.

Det er også mulig å tolke residuallønningene på andre måter. For eksempel bruker Gielen og Van Ours (2010) residuallønningene som mål på «jobb-matchen» mellom arbeidsgivere og arbeidstakere. Hvis residuallønningene bare reflekterer «jobb-matchen» mellom arbeidsgivere og arbeidstakere, vil vi forvente at arbeidere med høye residuallønninger vil være mindre tilbøyelig til å skifte jobb og emigrere (Borjas et al. 2019). Forklaringen på det er at arbeidstakeren i stor grad vil være «låst» til sin nåværende arbeidsgiver på grunn av «jobb-matchen». Det blir altså vanskeligere å emigrere, og skifte jobb i hjemlandet, hvis det er et krav om at lønnsnivået skal opprettholdes. Flaks har også en innflytelse på hvilken residuallønning individet har. At individet er på rett sted til rett tid. Alle de tre nevnte komponentene, uobserverte egenskaper, «jobb-match» og flaks, har nok en påvirkning på hvilken residuallønning et individ har. Roy-modellen predikerer at Norge i hovedsak vil ha en positivt selektert emigrasjonsstrøm på grunn av relativt lav avkastning av kompetanse. Dersom jeg får resultater som viser at den kumulative fordelingsfunksjonen av residuallønninger til emigrantene førsteordens stokastisk dominerer fordelingen til ikke-emigrantene, er det rimelig å tro at det er komponenten med uobserverte egenskaper i residuallønningene som er årsaken til at jeg får slike resultater. Hvis residuallønningene hovedsakelig representerer «jobb-match» vil vi se negativ seleksjon på residuallønningene.²⁶ Flakskomponenten i residuallønningen er med stor sannsynlighet tilfeldig fordelt mellom ikke-emigranter og emigranter. Det betyr at flakskomponenten med liten sannsynlighet vil bidra til positiv seleksjon på residuallønningene.

5.3 Metode for å undersøke om fordelingene er ulike

Jeg følger metoden Borjas et al. (2019) bruker til å undersøke om avstanden mellom kumulative fordelingsfunksjoner er statistisk signifikant. Metoden benytter statapakken til Araar og Duclos (2013). De tekniske detaljene til metoden er presentert i appendikset. Roy-modellen predikerer at den kumulative fordelingsfunksjonen til emigrantene skal førsteordens stokastisk dominere fordelingen til ikke-emigrantene dersom emigrantene er positivt selektert. Dersom jeg får resultater som tyder på at emigrantene er positivt selektert, vil formålet med denne metoden være å identifisere intervallet hvor verdiene til den kumulative fordelingsfunksjonen til ikke-

²⁶ I kapittel 3 ble det vist at Roy-modellen predikerer at emigrantene vil være positivt selektert hvis residuallønningene er tilstrekkelig overførbart mellom landene (høy korrelasjonskoeffisient) og spredningen til residuallønningene er større i utlandet. Hvis residuallønningene hovedsakelig reflekterer jobb-matchen vil korrelasjonskoeffisienten være liten. Når korrelasjonskoeffisienten er liten vil seleksjonen på residuallønningene være negativ fordi individer med en god jobb-match, og dermed en høy residuallønning, ikke vil emigrere (Borjas et al. (2019).

emigrantene er statistisk signifikant større enn verdiene til fordelingen til emigrantene.²⁷ Med andre ord, vil metoden vise estimert differanse mellom kumulative fordelingsfunksjoner med 95 % konfidensintervall for å undersøke om fordelingene er ulike.

²⁷ Dersom en får resultater som indikerer at emigrantene er negativt selektert vil formålet med metoden være å identifisere intervallet hvor verdiene til den kumulative fordelingsfunksjons til emigrantene er statistisk signifikant større enn verdiene til fordelingen til ikke-emigrantene.

6 Resultater og diskusjon

Dette kapittelet presenterer hovedfunnene fra den empiriske analysen. I den første delen presenteres analysen av seleksjon på observert kompetanse. I den analysen vil formålet være å identifisere hvordan utdanningsnivå og emigrasjon henger sammen. I den andre delen presenteres analysen av stokastisk dominans prediksjonen til Roy-modellen. Formålet med den analysen er å kartlegge om inntektsfordelingen og den uobserverte kompetansefordelingen til emigrantene førsteordens stokastisk dominerer fordelingen til ikke-emigrantene. Det predikerer Roy-modellen hvis emigrantene er positivt selektert på inntekt og uobservert kompetanse.

6.1 Seleksjon på observerbar kompetanse

Tabell 5
Gjennomsnittlige marginaleffekter fra logistisk regresjon

	(1) Menn	(2) Kvinner	(3) Menn	(4) Kvinner
Avhengig variabel: Emigrasjonsstatus				
Barn			-0,0009*** (0,0001)	-0,0010*** (0,0001)
Gift			-0,0001*** (0,0000)	-0,0002*** (0,0001)
VGS, studiefor.	0,0014*** (0,0001)	0,0008*** (0,0001)	0,0015*** (0,0001)	0,0008*** (0,0001)
VGS, yrkesfag	0,0000 (0,0001)	0,0000 (0,0001)	0,0001 (0,0001)	0,0000 (0,0001)
Fagskole	0,0014*** (0,0001)	0,0011*** (0,0001)	0,0014*** (0,0001)	0,0011*** (0,0001)
Bachelor	0,0023*** (0,0001)	0,0015*** (0,0001)	0,0023*** (0,0001)	0,0016*** (0,0001)
Master	0,0034*** (0,0001)	0,0026*** (0,0001)	0,0035*** (0,0001)	0,0026*** (0,0001)
PhD	0,0046*** (0,0001)	0,0037*** (0,0002)	0,0047*** (0,0001)	0,0036*** (0,0002)
N	6 678 597	4 141 205	6 678 597	4 141 205
Alders. effekter	Ja	Ja	Ja	Ja
Tids. effekter	Ja	Ja	Ja	Ja

Merknad: Tabellen rapporterer gjennomsnittlige marginaleffekter fra logistisk regresjon. Den avhengige variabelen er emigrasjonsstatus. Det er en dummyvariabel som tar verdien 1 dersom individet emigrerer ut av Norge påfølgende år, og 0 ellers. Analysen inkluderer både korttids- og langtidsemigranter. Referansekategori er grunnskoleutdanning. Klyngerobuste standardfeil i parentes som er gruppert på individ. * $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

Tabell 5 rapporterer gjennomsnittlige marginaleffekter for utdannings- og bakgrunnsvariabler fra en logistisk regresjon med emigrasjonsstatus som avhengig variabel. Emigrasjonsstatus er en dummyvariabel som tar verdien 1 dersom et individ emigrerer påfølgende år, og 0 ellers. Både korttids- og langtidsemigranter er inkludert. Modellspesifikasjon 1 og 2 er en grunnmodell uten familiefaktorer, mens modellspesifikasjon 3 og 4 følger hele oppsettet til ligning (10). Koeffisientene til utdanningsvariablene endrer seg lite når det kontrolleres for familiefaktorene.

I delkapittel 4.3 ble deskriptiv statistikk for utvalget presentert. Der ble det belyst at emigrantene har vesentlig mer utdannelse enn ikke-emigrantene.²⁸ Resultatene som presenteres i tabell 5 viser en tydelig sammenheng mellom utdanningsnivå og emigrasjonsstatus.²⁹ Sannsynligheten for å emigrere øker monotont med utdanningsnivået, og marginaleffektene er gjennomgående statistisk signifikant på 0,1 prosentnivå. Individene oppfører seg akkurat slik det teoretiske rammeverket predikerer. Når avkastningen av kompetanse er relativt lav i Norge, vil det gi individer med høyt kompetansenivå insentiver til å emigrere for å få økt avkastning av egen kompetanse. For like flyttekostnader vil insentivene til å emigrere øke monotont med utdanningsnivået siden netto migrasjonsfortjeneste stiger med utdanningsnivået.³⁰ At sannsynligheten for å emigrere stiger monotont med utdanningsnivået samsvarer derfor godt med teorien. Resultatene fra den deskriptive statistikken og tabell 5 gir grunn til å hevde at emigrantene er positivt selektert på observert kompetanse.

Det som skiller seg ut i tabell 5 er at marginaleffekten for individer med yrkesfaglig videregående utdanning ikke er statistisk signifikant og størrelsen på marginaleffekten er svært liten for denne gruppen. Vi kan dermed ikke si at individer med yrkesfaglig videregående utdanning er mer tilbøyelig til å emigrere enn individer med grunnskoleutdanning. Det er ikke et overraskende funn, siden begge utdanningskategoriene inneholder individer med et lavt utdanningsnivå.

Modellspesifikasjon 3 og 4 inkluderer bakgrunnsvariabler for om individet har barn i husstanden og er gift. Resultatene viser at bakgrunnsvariablene reduserer sannsynligheten for

²⁸ I tillegg viser den deskriptive statistikken at utdanningsfordelingen til emigrantene er høyreskjev, noe fordelingen til ikke-emigrantene ikke er. Det gjelder både for menn og kvinner. Slik utdanningsfordelingen til emigrantene er i forhold til fordelingen til ikke-emigrantene kan det, uten å vise kumulative fordelingsfunksjoner, indikere at fordelingen til emigrantene førsteordens stokastisk dominerer fordelingen til ikke-emigrantene.

²⁹ I kapittel A.5 i appendikset har jeg estimert ligning (10) ved å bruke en lineær sannsynlighetsmodell. Den lineære sannsynlighetsmodellen gir ganske like marginaleffekter som logit-modellen. Forskjellen er hovedsakelig at den lineære sannsynlighetsmodellen har større marginaleffekter for master og PhD.

³⁰ Figur 1 illustrerer dette poenget godt.

å emigrere. Det er intuitivt. Å ha barn og være gift øker trolig flyttekostnaden til individene. Alt annet likt vil høyere flyttekostnad gjøre det mindre attraktivt å emigrere.

Gjennomgående er margineffektene ikke særlig store i tabell 5. For eksempel i kolonne 4, som er den fulle modellen for kvinnene, estimeres det at en PhD-grad vil øke sannsynligheten for å emigrere med 0,36 prosentpoeng relativt til individer med bare grunnskoleutdanning. Margineffektene er små fordi det totalt sett er få som emigrerer. Gjennomsnittlig sannsynlighet for å emigrere er 0,18 % for menn, og 0,16 % for kvinner. Dermed er de relative effektene på sannsynlighetene likevel store. En kvinne med PhD har $3,25 \left(\frac{0,16 + 0,36}{0,16} \right)$ ganger så stor sannsynlighet for å emigrere relativt til den gjennomsnittlige sannsynligheten for kvinner.

6.2 Analyse av stokastisk dominans prediksjonen til Roy-modellen

Funnene i den deskriptive statistikken og logitanalysen tyder på at emigrantene er positivt selektert. Formålet med denne delen av analysen vil være å identifisere om stokastisk dominans prediksjonen til Roy-modellen, når det gjelder positiv emigrantseleksjon, er oppfylt. I delkapittel 3.2 ble stokastisk dominans prediksjonen til Roy-modellen gjort rede for. Der ble det belyst at emigrantene sin kumulative fordelingsfunksjon skal førsteordens stokastisk dominere fordelingen til ikke-emigrantene, dersom emigrantene er positivt selektert. Det ble belyst hvilke betingelser som må være oppfylt for at stokastisk dominans resultatet skal være oppfylt, når det gjelder positiv seleksjon på observert og uobservert kompetanse samt inntekt. Jeg kommer til å fokusere på inntekt og uobservert kompetanse i denne analysen. Observerbare egenskaper forklarer stort sett lite av variansen i inntektsfordelingen (Borjas et al. (2019; Lemieux 2006). Med andre ord betyr det at en lønnslikning ofte har lav forklaringskraft. Mesteparten av variasjonen i inntekt skyldes da variasjon i uobserverbare egenskaper. I delkapittel 3.2 ble det vist teoretisk at seleksjon på inntekt avhenger av et vektet gjennomsnitt av seleksjonen som forekommer på observert og uobservert kompetanse. Vektingen bestemmes av kompetansetypens bidrag til å skape varians i inntektsfordelingen. Når uobserverte egenskaper ofte forklarer den største delen av variasjonen i inntekt, vil det sentrale være å avdekke hvordan emigrantene er selektert på uobservert kompetanse.

Den deskriptive statistikken viser at emigrantene har høyere gjennomsnittlig inntekt enn ikke-emigrantene. Det tyder på at emigrantene er positivt selektert på inntekt. Selv om emigrantene har høyere gjennomsnittlig inntekt, kan et stort antall individer oppføre seg i strid med stokastisk dominans prediksjonen til Roy-modellen (Borjas et al. 2019). Intuisjonen er at

dersom stokastisk dominans resultatet holder, så er emigrantene positivt selektert uansett hvilket inntektsnivå en setter som terskelverdi. Emigrantene har da alltid en større andel som har inntekt over den gitte terskelverdien enn ikke-emigrantene. Det er ikke nødvendigvis tilfellet når en bare sammenligner gjennomsnittlig inntekt mellom de to gruppene. Stokastisk dominans resultatet er dermed en mye sterkere prediksjon av modellen enn at emigrantene har høyere/lavere gjennomsnittsinntekt, fordi det viser at emigrantene har lik seleksjon uansett hva en setter som terskelverdi. Det gjør at Borjas et al. (2019) hevder at et bevis på at stokastisk dominans resultatet holder, demonstrerer at forskjeller i pris på kompetanse spiller en meget stor rolle i migrasjonsbeslutninger.

6.2.1 Seleksjon på uobserverbar kompetanse

6.2.1.2 Resultater fra lønnslikningene

For å kunne sammenligne med tidligere analyser av avkastningen av utdanning, rapporterer jeg først resultater fra regresjoner av «vanlige» lønnslikninger der utdanning er målt som antall år.

Tabell 6
Lønnslikninger

	(1) Menn	(2) Kvinner	(3) Menn	(4) Kvinner
	Log årlig inntekt		Log standardisert årlig inntekt	
Gift	0,0345*** (0,0006)	-0,0306*** (0,0006)	0,0308*** (0,0006)	-0,0315*** (0,0006)
Barn	0,0320*** (0,0006)	-0,0577*** (0,0006)	-0,0056*** (0,0006)	-0,0836*** (0,0005)
Utdanning	0,0465*** (0,0001)	0,0346*** (0,0001)	0,0462*** (0,0001)	0,0341*** (0,0001)
Konstantledd	1,1122*** (0,0018)	1,1574*** (0,0018)	-0,5078*** (0,0017)	0,0141*** (0,0016)
N	6 678 597	4 141 205	6 678 597	4 141 205
Aldersfaste effekter	Ja	Ja	Nei	Nei
Tidsfaste effekter	Ja	Ja	Nei	Nei
R ²	0,2020	0,2077	0,1144	0,1302

Merknad: I kolonne 1 og 2 er avhengig variabel årlig inntekt på logaritmeform. Inntekten i kolonne 1 og 2 er målt i enheter, hvor 50 000 kr er en enhet. I kolonne 3 og 4 er avhengig variabel standardisert årlig inntekt på logaritmeform. Barn er en dummyvariabel som er lik 1 dersom individet har ett eller flere barn under 18 år i husstanden, og 0 ellers. Gift er en dummyvariabel som tar verdien 1 dersom individet er gift, og 0 ellers. Referansekategori er grunnskoleutdanning. Klyngerobuste standardfeil i parentes som er gruppert på individ. Lønnslikningene er estimert med minste kvadraters metode. * $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

Modellspesifikasjonene i tabell 6 viser at det er en positiv statistisk signifikant sammenheng mellom utdanning og inntekt. Kolonne 1 viser den estimerte avkastningen for menn når ikke-standardisert inntekt er avhengig variabel. Denne modellspesifikasjonen estimerer at mennene kan forvente å få 4,65 % høyere årlig inntekt ved å ta ett ekstra år med utdannelse. Kolonne 2 viser samme modellspesifikasjon for kvinner, hvor estimatet henholdsvis er 3,46 %. Aakvik et al. (2010, fotnote 16) finner et MKM-estimat av avkastningen av ett ekstra år med utdanning på rundt 5 % for menn. Det samsvarer bra med mine resultater.

Tabell 7
Lønnslikninger

	(1)	(2)	(3)	(4)
	Menn	Kvinner	Menn	Kvinner
	Log årlig inntekt		Log standardisert årlig inntekt	
Gift	0,0346*** (0,0006)	-0,0319*** (0,0005)	0,0215*** (0,0006)	-0,0388*** (0,0005)
Barn	0,0349*** (0,0006)	-0,0524*** (0,0006)	0,0012* (0,0006)	-0,0745*** (0,0005)
VGS, studiefor.	0,1591*** (0,0014)	0,0842*** (0,0010)	0,1572*** (0,0014)	0,0878*** (0,0010)
VGS, yrkesfag	0,0744*** (0,0008)	0,0415*** (0,0008)	0,0780*** (0,0008)	0,0397*** (0,0008)
Fagskole	0,1988*** (0,0016)	0,1440*** (0,0019)	0,1975*** (0,0016)	0,1472*** (0,0019)
Bachelor	0,2542*** (0,0010)	0,1705*** (0,0008)	0,2523*** (0,0010)	0,1731*** (0,0008)
Master	0,4414*** (0,0015)	0,4062*** (0,0018)	0,4315*** (0,0015)	0,4038*** (0,0018)
PhD	0,5474*** (0,0046)	0,5683*** (0,0073)	0,5231*** (0,0046)	0,5494*** (0,0074)
Konstantledd	1,5927*** (0,0009)	1,5271*** (0,0009)	-0,0410*** (0,0007)	0,3620*** (0,0007)
N	6 678 597	4 141 205	6 678 597	4 141 205
Aldersfaste effekter	Ja	Ja	Nei	Nei
Tidsfaste effekter	Ja	Ja	Nei	Nei
R ²	0,2122	0,2253	0,1195	0,1476

Merknad: I kolonne 1 og 2 er avhengig variabel årlig inntekt på logaritmeform. Inntekten i kolonne 1 og 2 er målt i enheter, hvor 50 000 kr er en enhet. I kolonne 3 og 4 er avhengig variabel standardisert årlig inntekt på logaritmeform. Referansekategori er grunnskoleutdanning. Klyngerobuste standardfeil i parentes som er gruppert på individ. * $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

Tabell 7 rapporterer resultatene fra estimeringen av lønnslikningene (likning (11) og (12)). Resultatene viser tydelig at forventet inntekt er monotont stigende med utdanningsnivået. Dette samsvarer med resultatene i tabell 6, hvor ett ekstra år med utdanning gir økt forventet inntekt. Vi kan tolke koeffisientene til utdanningsdummyene som hva et individ kan forvente å få i ekstra inntekt ved å fullføre et spesifikt utdanningsnivå relativt til referansegruppen. For eksempel viser kolonne 1 at regresjonsmodellen estimerer at menn med mastergrad har ca. 44 % høyere årlig inntekt enn menn med kun grunnskoleutdanning. Hovedformålet med å estimere disse lønnslikningene er å bruke de estimerte residuallønningene som mål på uobservert kompetanse. Forklaringskraften til regresjonsmodellene er relativt lav. Det viser at mesteparten av variasjonen i inntekt skyldes uobserverte egenskaper. Det innebærer at det er spesielt viktig å undersøke hvordan emigrantene er selektert på uobservert kompetanse.

6.2.1.3 Kumulative fordelingsfunksjoner av residuallønninger

Det er residuallønningene fra likning (11) og (12) som benyttes til analysen av seleksjon på uobservert kompetanse. I denne oppgaven blir residuallønningene brukt som mål på uobservert kompetanse. I delkapittel 5.2 ble denne tolkningen diskutert. Med bakgrunn i den diskusjonen er det rimelig å tolke at emigrantene er positivt selektert, hvis den kumulative fordelingsfunksjonen av residuallønningene til emigrantene førsteordens stokastisk dominerer fordelingen til ikke-emigrantene. Hvis stokastisk dominans prediksjonen er oppfylt, vil emigrantene være positivt selektert uansett hvilket nivå på residuallønningene en setter som terskelverdi. Emigrantene har da alltid en større andel som har residuallønninger over den gitte terskelverdien enn ikke-emigrantene. Vi kan med andre ord si at emigrantene da oppnår en høyere avkastning av kompetanse enn ikke-emigrantene i Norge, når en sammenligner individer med like observerbare karakteristikk.

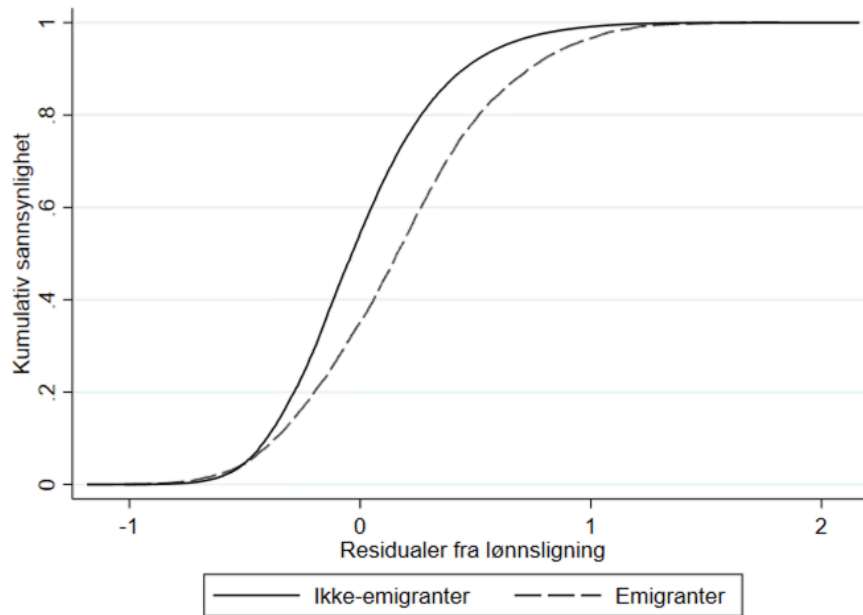
Figur 2 viser kumulative fordelingsfunksjoner av residuallønningene til mannlige (kvinnelige) ikke-emigranter og emigranter fra regresjon med standardisert inntekt som avhengig variabel. Figur A.2 i appendikset viser tilsvarende kumulative fordelingsfunksjoner av residuallønninger fra regresjon med ikke-standardisert inntekt som avhengig variabel. De visuelle resultatene i figur A.2 er nesten identisk med de visuelle resultatene i figur 2. Jeg har også undersøkt om det er forskjeller mellom langtids- og korttidsemigranter når det gjelder seleksjon på uobservert

kompetanse, men mine funn tyder på at det ikke er nevneverdige forskjeller mellom de to emigrantgruppene.³¹

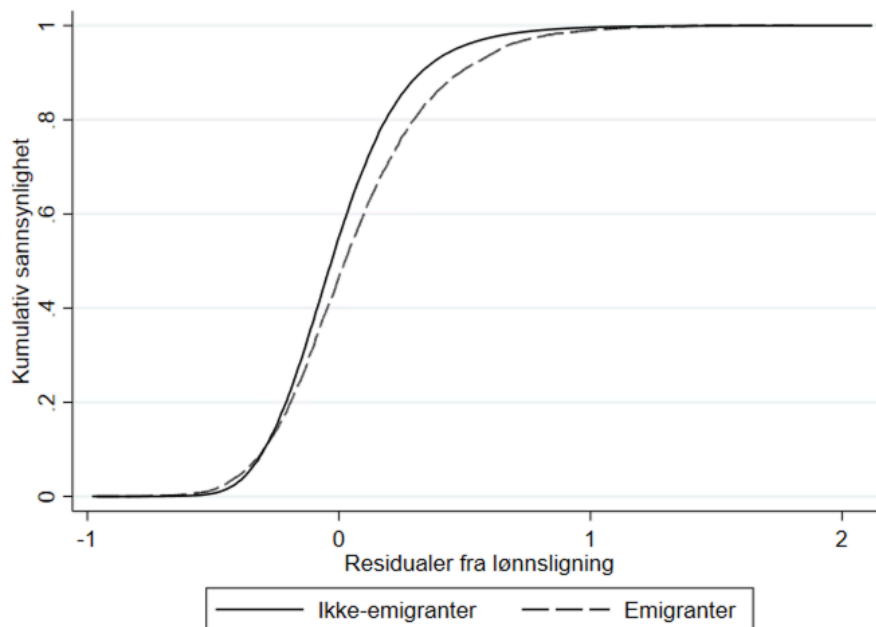
Figur 2

Kumulative fordelingsfunksjoner av residualer fra lønnslikning med standardisert inntekt som avhengig variabel. Forskjell mellom ikke-emigranter og emigranter

a) Menn



b) Kvinner

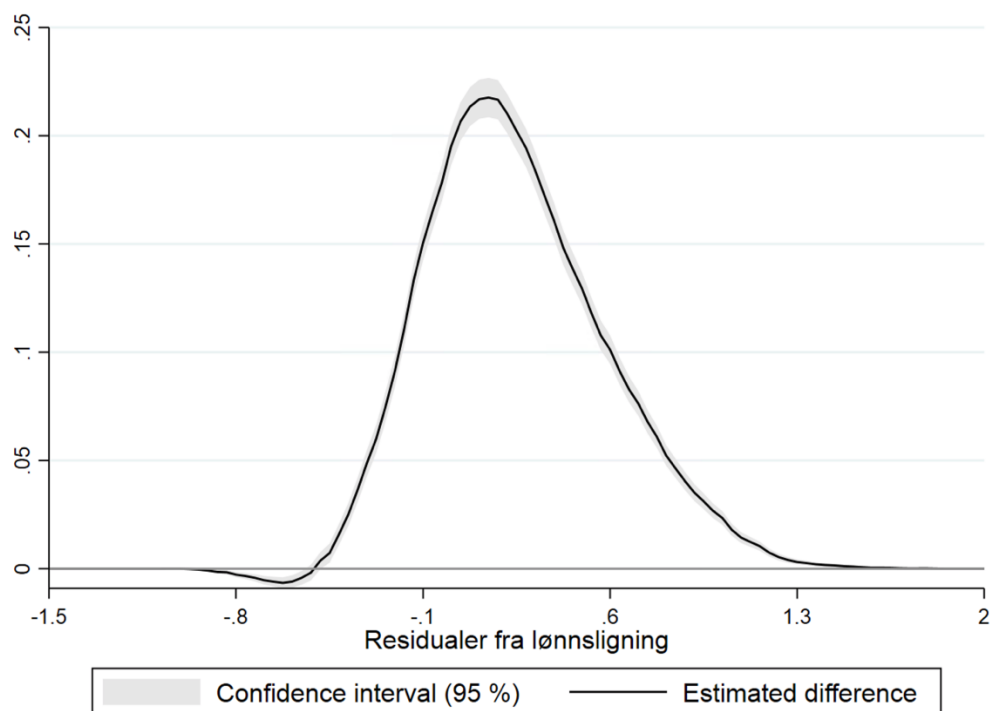


³¹ Resultatene fra den analysen kan finnes i kapittel A.6 i appendikset.

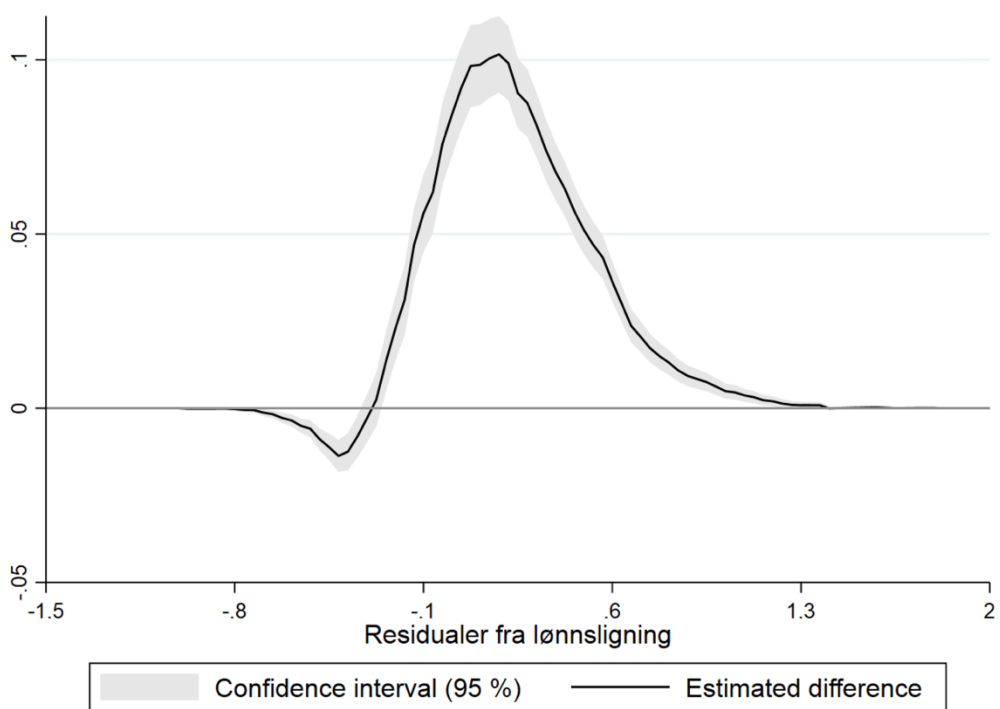
Figur 3

Estimert differanse mellom de kumulative fordelingsfunksjonene i figur 2 med 95 % konfidensintervall

a) Menn



b) Kvinner



Figur 2 (a) viser at den kumulative fordelingsfunksjonen til mannlige emigranter er til høyre for fordelingen til mannlige ikke-emigranter, men det er noe overlapp i den nedre delen av halen til fordelingene. Jeg har utført en Kolmogorov-Smirnov-test som tester om de to utvalgsfordelingene stammer fra den samme fordelingen i populasjonen. Testresultatene fra Kolmogorov-Smirnov-testen gir et sterkt grunnlag til å forkaste nullhypotesen om at fordelingene til mannlige ikke-emigranter og emigranter stammer fra samme fordeling.³² En utdypelse om Kolmogorov-Smirnov-testen finnes i kapittel A.4 i appendikset. Figur 3 (a) viser den estimerte differansen mellom de to kumulative fordelingene i figur 2 (a) med 95 % konfidensintervall. Figur 3 (a) viser at den kumulative fordelingen til ikke-emigrantene er statistisk signifikant over fordelingen til emigrantene over nesten hele fordelingen.

Resultatene til kvinnene, som er figur 2 (b), er i hovedsak ganske lik resultatene til mennene. Den kumulative fordelingsfunksjonen til de kvinnelige emigrantene er til høyre for fordelingen til ikke-emigrantene, men det er noe overlapp i den nedre delen av halen til fordelingene. Kolmogorov-Smirnov-testen gir samme resultat som den til mennene.³³ Figur 3 (b) viser den estimerte differansen mellom de to kumulative fordelingene i figur 2 (b) med 95 % konfidensintervall. Figur 3 (b) viser at den kumulative fordelingen til ikke-emigrantene er statistisk signifikant over fordelingen til emigrantene over nesten hele fordelingen.

Både for menn og kvinner er den kumulative fordelingsfunksjonen til emigrantene til høyre for fordelingen til ikke-emigrantene over nesten hele fordelingen, og differansen er statistisk signifikant der hvor fordelingen til emigrantene er til høyre for fordelingen til ikke-emigrantene. Det er i den nedre delen av halen til fordelingene det er overlapping for begge kjønnene. Tolkningen av resultatet blir derfor noe mer komplisert. Teoretisk skal fordelingen til emigrantene ikke krysse fordelingen til ikke-emigrantene for at stokastisk dominans skal være oppfylt. Det medfører at den intuitive tolkningen av stokastisk dominans resultatet (at emigrantene alltid er positivt selektert uansett hvilket nivå på residuallønningene man setter som terskelverdi) ikke blir fullstendig oppfylt. Likevel vil den intuitive tolkningen av stokastisk dominans resultatet være gjeldene over nesten hele fordelingen. Med tanke på at den kumulative fordelingsfunksjonen til emigrantene *nesten* førsteordens stokastisk dominerer fordelingen til ikke-emigrantene, velger jeg å tolke de visuelle resultatene i figur 2 og 3 som at emigrantene er positivt selektert på uobservert kompetanse. Det gjelder både for menn og

³² P-verdien fra testen er 0,000. Det gir en klar indikasjon om at vi kan forkaste nullhypotesen.

³³ P-verdi på 0,000.

kvinner. Når emigrantene er positivt selektert på uobservert kompetanse impliserer det positiv seleksjon innenfor utdanningsgrupper (Borjas et al. 2019).

6.2.2 Seleksjon på inntekt

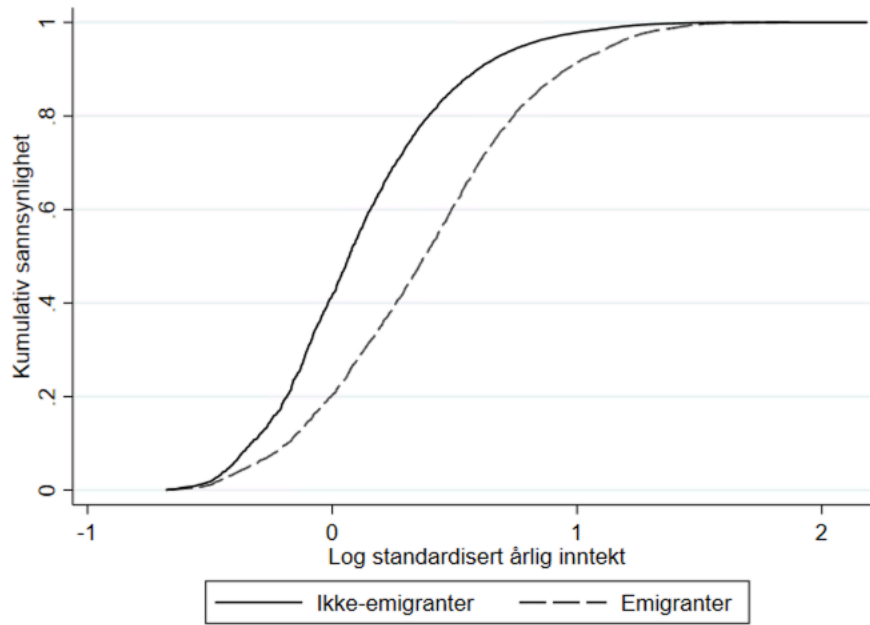
6.2.2.1 Analyse av forskjellen mellom ikke-emigranter og emigranter

I kapittel 3 ble det teoretisk gjort rede for at stokastisk dominans i inntektsfordelingen avhenger av både observert og uobservert kompetanse. I delkapittel 6.1 ble det vist at tilbøyeligheten til å emigrere stiger monotont med utdanningsnivået. Videre ble det i delkapittel 6.2.1 vist at emigrantene sin kumulative fordelingsfunksjon av residuallønninger nesten førsteordens stokastisk dominerer fordelingen til ikke-emigrantene. Basert på disse resultatene gir det grunnlag for å hevde at emigrantene er positivt selektert på observert og uobservert kompetanse. Roy-modellen predikerer at positiv seleksjon på begge kompetansetyper impliserer positiv seleksjon på inntekt. Ettersom jeg ikke benytter regresjoner i denne analysen, bruker jeg den standardiserte inntekten til å kontrollere for arbeidserfaring og årseffekter. For eksempel, vil en relativt nyutdannet person stort sett tjene mindre enn en erfaren arbeidstaker med tilsvarende utdanning, men den nyutdannede personen kan tjene mye sammenlignet med andre på sin alder. Jeg benytter et paneldatasett som strekker seg over 10 år. Inntekten til individene kan for eksempel variere med konjunktorene i økonomien. Av den grunn er den standardiserte inntekten regnet ut slik at den viser hva et individ tjener relativt til andre med lik alder og kjønn i et år.

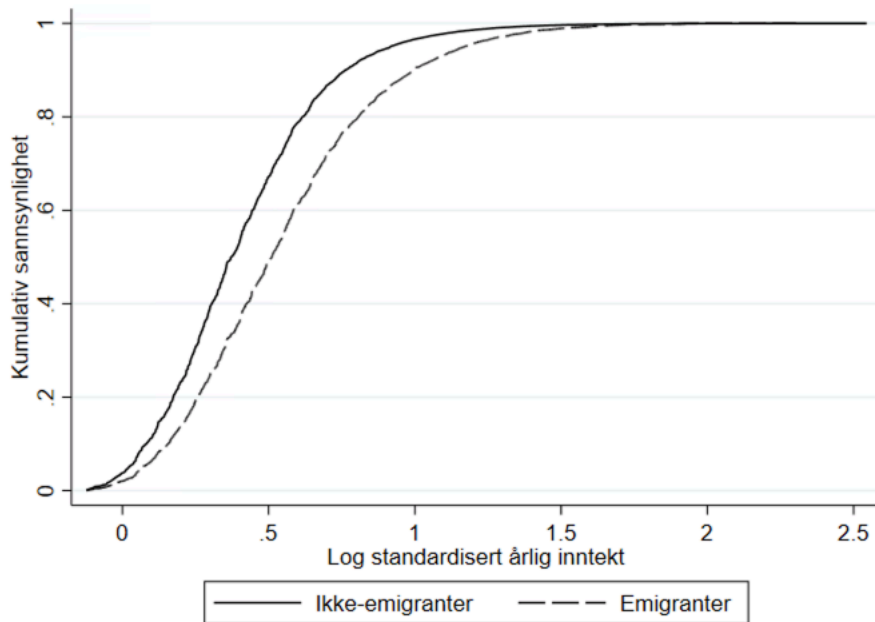
Figur 4

Kumulative fordelingsfunksjoner av standardisert inntekt for ikke-emigranter og emigranter

a) Menn



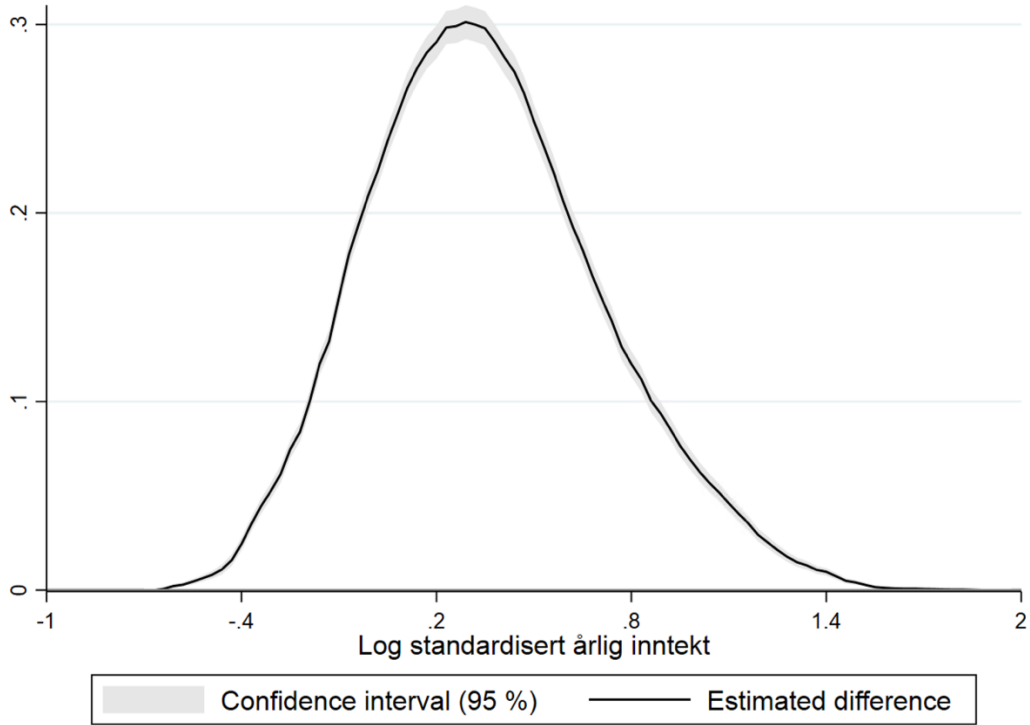
b) Kvinner



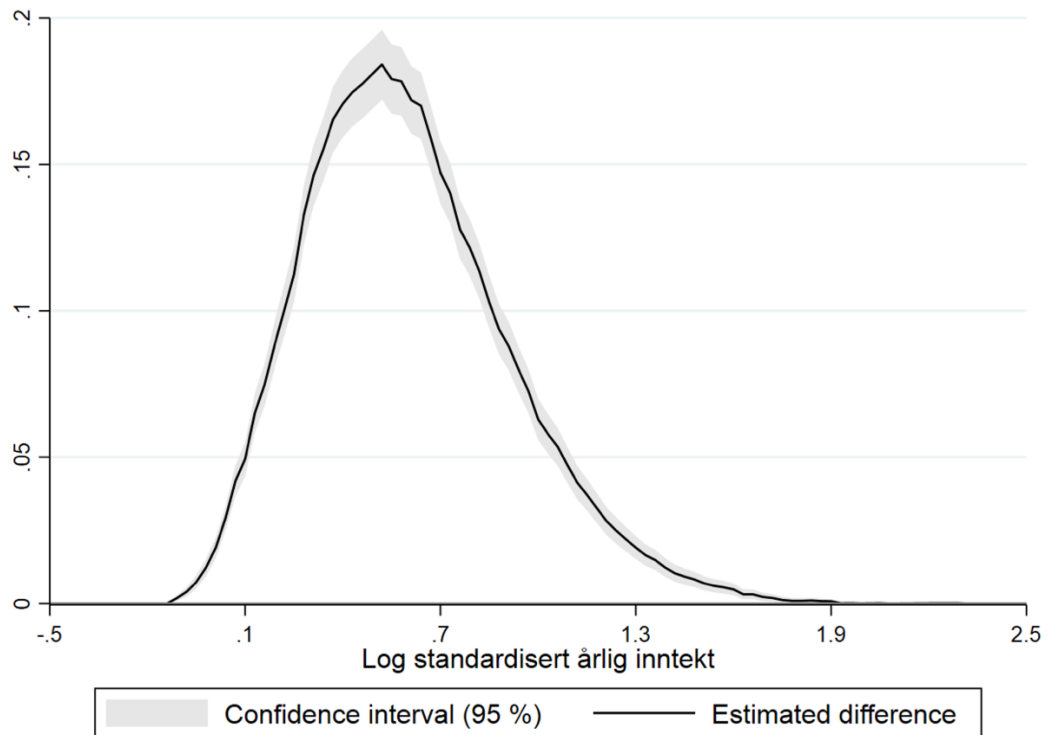
Figur 5

Estimert differanse mellom de kumulative fordelingsfunksjonene i figur 4 med 95 % konfidensintervall

a) Menn



b) Kvinner



Figur 4 (a) presenterer kumulative fordelingsfunksjoner av standardisert inntekt for mannlige ikke-emigranter og emigranter. Figur 4 (b) viser tilsvarende figur for kvinner. Figur 4 (a) og (b) viser tydelig at emigrantene sin fordeling ligger til høyre for fordelingen til ikke-emigrantene over hele fordelingen. Testresultatene fra en Kolmogorov-Smirnov-test gir et sterkt grunnlag til å forkaste nullhypotesen om at fordelingene til mannlige (kvinnelige) ikke-emigranter og emigranter stammer fra samme fordeling i populasjonen.³⁴ Figur 5 viser den estimerte differansen mellom de kumulative fordelingsfunksjonene til mannlige (kvinnelige) ikke-emigranter og emigranter med 95 % konfidensintervall. De visuelle resultatene i figur 5 demonstrerer at differansene er ikke-negativ over hele fordelingen, hvor differansene er statistisk signifikant. De visuelle funnene som fremkommer i figur 4 og 5 viser tydelig at fordelingen til emigrantene førsteordens stokastisk dominerer fordelingen til ikke-emigrantene. Det gir grunnlag for å konkludere at norske emigranter er positivt selektert på inntekt. Figur 2 og 4 viser at kvinner er noe svakere positivt selektert enn menn, fordi stokastisk dominansen blant kvinnene er svakere.³⁵ I en familie vil det være den samlede netto migrasjonsfortjenesten til familien som avgjør om det er lønnsomt å emigrere (Borjas og Bronars 1991). At gifte og samboende kvinner følger med mannen til utlandet kan være en forklaring på hvorfor kvinner er noe svakere selektert. Menn har ofte høyest inntekt i et parforhold. Gifte og samboende kvinner trenger derfor ikke nødvendigvis en så stor inntektsøkning for at den samlede netto migrasjonsfortjenesten til familien skal være positiv.

At emigrantene er positivt selektert på inntekt er et lite overraskende funn, ettersom mine resultater viser at emigrantene er positivt selektert på observert og uobservert kompetanse. Med tanke på at observerte egenskaper forklarer lite av variasjonen i inntekt, kan det derfor hevdes at uobserverte egenskaper er den viktigste driveren for at inntektsfordelingen til emigrantene førsteordens stokastisk dominerer fordelingen til ikke-emigrantene. I denne oppgaven er den observerbare kompetansen bare utdanningen til individet. Det innebærer at alt annet vil være uobserverte egenskaper. Personlige egenskaper er trolig en svært viktig faktor når det gjelder hvordan et individ blir lønnet på arbeidsmarkedet sammenlignet med andre med like observerte karakteristikk. Personlige egenskaper kan for eksempel være motivasjon, samarbeidsevner, gjennomføringsevne og hvor produktiv individet er.

³⁴ Jeg utførte en Kolmogorov-Smirnov-test separat for menn og kvinner, og p-verdien fra testen er 0,000 i begge testene.

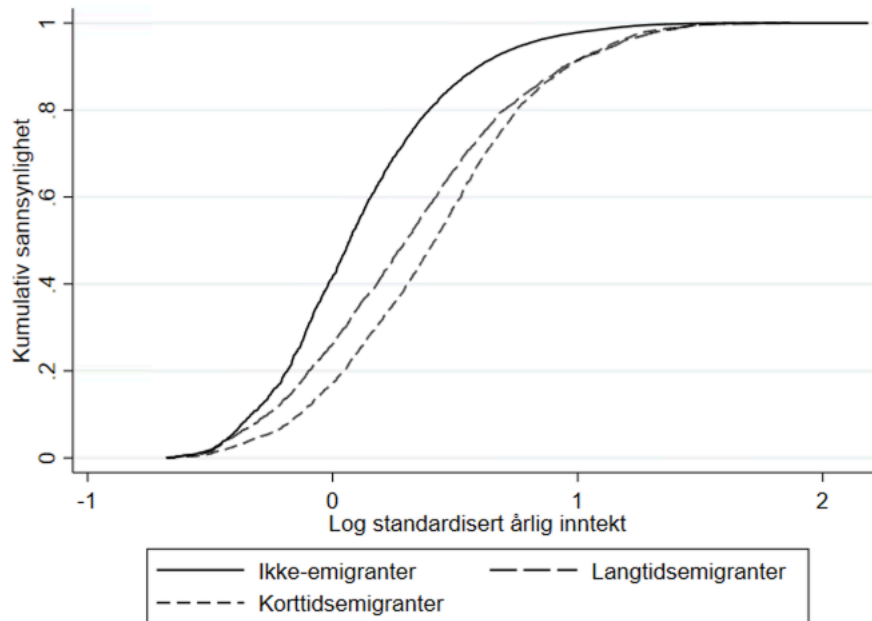
³⁵ Det er mindre avstand mellom kvinnelige emigranter og ikke-emigranter sine kumulative fordelingsfunksjoner enn for mannlige emigranter og ikke-emigranter.

6.2.2.2 Analyse av forskjellen mellom korttids- og langtidsemigranter

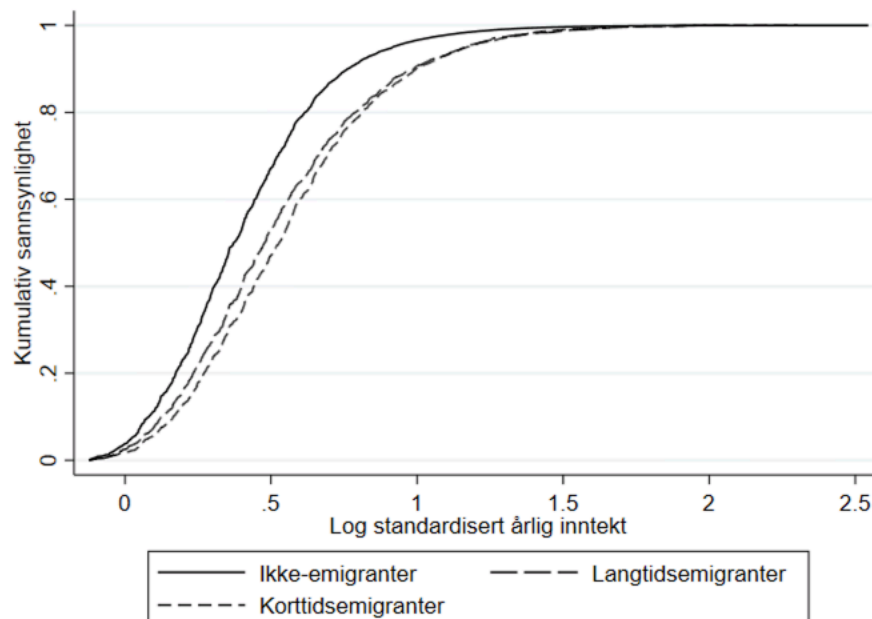
Figur 6

Kumulative fordelingsfunksjoner av standardisert inntekt for ikke-emigranter, langtids- og korttidsemigranter

a) Menn:



b) Kvinner:



Borjas og Bratsberg (1996) viser empirisk og teoretisk at kompetansekompisjonen til migranter som tilbakevender avhenger av hvilken type seleksjon som var tilfellet i første omgang. Datagrunnlaget er basert på migranter i USA på 1980-tallet. De viser at dersom migrantene var positivt selektert, vil migrantene som tilbakevender tilhøre den nedre delen av kompetansefordelingen til migrantene. Borjas et al. (2019) bekrefter resultatene til Borjas og Bratsbergs (1996) ved at de danske langtidsemigrantene er sterkere positivt selektert enn korttidsemigrantene når det gjelder emigranter som flytter til land utenfor Norden. For emigrantene som flytter til nordiske land er resultatene til Borjas et al. (2019) ikke i tråd med Borjas og Bratsberg (1996). Mine funn stemmer tilsynelatende ikke helt med Borjas og Bratsberg (1996) sine resultater. Den deskriptive statistikken viser at korttidsemigrantene, både for menn og kvinner, har høyere gjennomsnittlig inntekt og utdanning enn langtidsemigrantene. Figur 6 (a) og (b) viser at den kumulative fordelingsfunksjonen til korttidsemigrantene er til høyre for den til langtidsemigrantene, men det er overlapp i den øvre delen av fordelingene. Mine funn tyder derfor ikke på at emigrantene, som tilbakevender (først), tilhører den nedre delen av kompetansefordelingen til emigrantene.

Med tanke på at mange av emigrantene i utvalget har ukjent destinasjonsland, kan jeg ikke skille mellom emigranter som flytter til nordiske land og land utenfor Norden. Avkastningen av kompetanse er stort sett høyere utenfor Norden. Det kan derfor være at langtidsemigranter som flytter til ikke-nordiske land er sterkere positivt selektert enn korttidsemigranter som flytter til ikke-nordiske land i mitt utvalg også. Avkastningen av kompetanse er relativt lav i hele Norden. Hvis inntektsfordelingene i de nordiske landene er identisk, vil ikke Roy-modellen ha en klar prediksjon om seleksjonen (Borjas et al. 2019). Dersom emigrantene som flytter til nordiske land er positivt selektert kan det i større grad være tilfeldig om korttidsemigrantene er sterkere positivt selektert enn langtidsemigrantene. Jeg syns derfor vi må tolke disse resultatene med varsomhet siden mine data er mangelfulle på dette punktet.

6.3 Begrensninger i analysen

De mest sentrale begrensningene i analysen er at mange emigranter har ukjent destinasjonsområde, og en grovkornet indikator om individet jobber fulltid i et gitt år. Med tanke på at nordiske land, i likhet med Norge, har relativt lav avkastning av kompetanse ville det vært ønskelig å skille mellom emigranter som flytter til nordiske land og ikke-nordiske land

i analysen. Det var ikke mulig siden 25 % av emigrantene i utvalget har ukjent destinasjonsområde. Tabell A.5 i appendikset viser deskriptiv statistikk for de emigrantene som har kjent destinasjonsområde. Tallene viser at emigranter som flytter til ikke-nordiske land har høyere utdanningsnivå og inntekt enn emigranter som flytter til nordiske land. Det gir en indikasjon på at førstnevnte er sterkest positivt selektert. Det er konsistent med Roy-modellen. Jeg vil understreke at tallene i tabell A.5 må tolkes med varsomhet når mange emigranter i utvalget har ukjent destinasjonsområde.

Videre ville det vært ønskelig å avgrense utvalget til individer som jobbet fulltid. Inntektsopplysningene til et individ som jobber deltid viser ikke nødvendigvis det sanne inntektspotensialet til individet (Borjas et al. 2019). Jeg har ikke data på arbeidstiden til individene, og valgte derfor å utføre en inntektsavgrensning for å luke ut individer (observasjoner) som hadde svak tilknytning til arbeidsmarkedet. Problemet med en slik inntektsavgrensning er at det med stor sannsynlighet vil være individer i utvalget som jobbet deltid. Deltidsarbeid er langt mer utbredt blant kvinner enn menn (NOU 2008:6). Det er derfor hovedsakelig i analysen av kvinnene hvor dette kan være et problem. Likevel vil jeg påpeke at inntektsavgrensningen i det minste luker ut individene (observasjonene) med den svakeste tilknytningen til arbeidsmarkedet.

Som nevnt tidligere er registerdataene jeg har til rådighet anonymisert av Institutt for økonomi ved UiB. Flere variabler har blitt aggregert slik at individer ikke kan bli gjenkjent. Et datagrunnlag med opplysninger om den faktiske inntekten til individene ville medført at residuallønningene ble estimert mer presist, og inntektsfordelingen ville vært mer presis. Analysen viser tydelig at jeg har klart å skille emigrantene og ikke-emigrantene fra hverandre. Jeg tror derfor at uaggregerte inntektsdata ville gitt tilsvarende resultater som denne oppgaven har funnet.

6.4 Sammenligning med Borjas et al. (2019) og andre studier

Jeg har i stor grad fulgt Borjas et al. (2019) med bakgrunn i at begge studiene tester stokastisk dominans prediksjonen til Roy-modellen på emigrasjonsdata. Det er noen viktige forskjeller mellom denne studien og Borjas et al. (2019). Begge studiene inkluderer observasjoner i aldersspennet 25-54 år som ikke har en innvandrerbakgrunn. Borjas et al. (2019) avgrenser videre analysen til individer som jobber fulltid, mens denne studien utfører en inntektsavgrensning. Andre viktige forskjeller er at emigrantene i hovedanalysen til Borjas et

al. (2019) bare omfatter langtidsemigranter, og de skiller mellom emigranter som flytter til Norden og utenfor Norden. Jeg inkluderer både korttids- og langtidsemigranter i analysen, og har ikke mulighet til å skille emigrantene på hvor de flytter til på grunn av mangelfulle destinasjonsdata. På tross av noe ulikt utvalg og datagrunnlag er resultatene mine konsistente med Borjas et al. (2019) sine. Danmark er i likhet med Norge en velferdsstat med relativt lav avkastning av kompetanse. Det impliserer at Roy-modellen predikerer at emigrasjonsstrømmen hovedsakelig vil være positivt selektert i begge landene. Både denne studien og Borjas et al. (2019) finner empirisk støtte for stokastisk dominans prediksjonen til Roy-modellen når det gjelder positiv seleksjon. Til sammen finner Borjas et al. (2019) 30 632 korttids- og langtidsemigranter. Jeg finner totalt 18 551 korttids- og langtidsemigranter.

Borjas et al. (2019) viser at danske emigranter som flytter til nordiske land er positivt selektert. De deskriptive funnene i tabell A.5 gir en indikasjon på at norske emigranter, som flytter til nordiske land, er positivt selektert. Det er ikke konsistent med Roy-modellen hvis det er slik alle de nordiske landene har en positivt selektert emigrasjonsstrøm til de andre nordiske landene. Samme type seleksjon begge veier er ikke i samsvar med Roy-modellens prediksjoner. Ettersom mange emigranter i utvalget har manglende informasjon om destinasjonsland er det begrenset hvor mye vekt en kan legge på problemstillingen som tas opp i dette avsnittet.

Pedersen et al. (2003) og Pedersen et al. (2008) har i likhet med denne studien analysert emigrasjonsstrømmen ut av Norge på 1990-tallet. Det er noen vesentlige forskjeller mellom min studie og de to nevnte studiene. De har et bredere utvalg, hvor blant annet aldersspennet er bredere og de ekskluderer ikke individer med lav inntekt. Jeg studerer også emigrasjonsstrømmen over en lengre periode enn dem. Likevel samsvarer resultatene mine i stor grad med resultatene til Pedersen et al. (2003) og Pedersen et al. (2008).

7 Oppsummering og avsluttende kommentar

Formålet med denne oppgaven har vært å studere hvordan norske emigranter er selektert fra populasjonen. Norge har relativt lav avkastning av kompetanse. Roy-modellen predikerer derfor at norske emigranter hovedsakelig vil være positivt selektert. Positiv seleksjon betyr at emigrantene tilhører den øvre delen av kompetansefordelingen. Det eksisterer studier som har undersøkt hvordan norske emigranter er selektert. Denne oppgaven kommer med nye bidrag til den eksisterende litteraturen. Jeg har testet stokastisk dominans prediksjonen til Roy-modellen på norske emigrasjonsdata, og kartlagt hvordan norske emigranter er selektert på uobservert kompetanse. Av det jeg er kjent med er det ikke gjort på norske data før. Oppgaven bygger på anonymisert registerdata fra perioden 1993-2003.

Roy-modellen skiller mellom observert og uobservert kompetanse. Modellen viser at seleksjon på en kompetansetype alene bestemmes av forskjeller i avkastningen av den kompetansetypen mellom landene. Dersom emigrantene er positivt selektert på uobservert kompetanse predikerer modellen at emigrantene sin kumulative kompetansefordeling av uobservert kompetanse skal førsteordens stokastisk dominere fordelingen til ikke-emigrantene. Inntektsfordelingen avhenger av både observert og uobservert kompetanse, slik at førsteordens stokastisk dominans i inntektsfordelingen kombinerer de to kompetansetypene. Forklaringskraften til lønnslikningene i oppgaven er relativt lave. Det viser at mesteparten av variasjonen i inntekt skyldes uobserverte egenskaper. Det er derfor essensielt å avdekke hvordan emigrantene er selektert på uobservert kompetanse. Oppgaven benytter residualer fra lønnslikninger (residuallønninger) som mål på uobservert kompetanse.

Resultatene mine er i tråd med prediksjonene til modellen. Fordelingen av residuallønningene til emigrantene førsteordens stokastisk dominerer *nesten* fordelingen til ikke-emigrantene. Videre viser resultatene at inntektsfordelingen til emigrantene førsteordens stokastisk dominerer fordelingen til ikke-emigrantene. Ettersom variansen til inntektsfordelingen i hovedsak avhenger av uobservert kompetanse, er det grunn til å hevde at det er den uobserverte kompetansen som er den viktigste driveren til førsteordens stokastisk dominans i inntektsfordelingen. Oppgaven viser at tilbøyeligheten til å emigrere er monotont økende med utdanningsnivå. Resultatene viser tydelig at emigrantene er positivt selektert på observerbare og uobserverbare karakteristikk. Alt i alt tyder mine resultater på at Roy-modellen kan være et godt rammeverk for å studere migrasjonsstrømmer.

Utvalget mitt består av relativt få emigranter. Inntekts-, alders- og innvandreravgrensningen medfører at en rekke emigranter ikke kommer med i utvalget til oppgaven. Det vil derfor være misvisende å basere seg på mitt utvalg når en skal beskrive hvor mange individer som emigrerer ut av Norge i perioden som studeres. Jeg kan likevel konstatere at det er få individer uten innvandrerbakgrunn, i aldersgruppen 25-54 år, med relativt sterk tilknytning til arbeidsmarkedet som emigrerer. Flyttekostnaden er trolig stor for de fleste individer. Høy flyttekostnad gjør at emigrasjon blir ulønnsomt for et individ som i utgangspunktet ville fått høyere inntekt i utlandet. I tillegg er de fleste emigrantene i utvalget korttidsemigranter. Det innebærer at de aller fleste emigrantene flytter tilbake til Norge i løpet av relativt kort tid. Dette samsvarer med den vanlige oppfatningen at Norge ikke har et problem med høykompetanseflukt. Det er heller en styrke at folk med høy kompetanse høster arbeidserfaring i utlandet i en stadig mer globalisert verden.

Til videre forskning kan det være interessant å undersøke om seleksjonen endrer seg over tid. Norge går i retning mot større lønnsulikhet (SSB 2018). Den positive seleksjonen kan derfor bli svakere over tid.

Litteraturliste

- Aakvik, Arild, Kjell G. Salvanes, og Kjell Vaage. 2010. «Measuring heterogeneity in the returns to education using an education reform» *European Economic Review*, 54(4):483-500.
- Abramitzky, Ran, Leah Platt Bouston, og Katherine Erriksson. 2012. «Europe's Tired, Poor, Huddled Masses: Self-selection and Economic Outcomes in the Age of Mass Migration.» *American Economic Review*, 102(5):1832-1856.
- Angrist, Joshua D., og Jörn-Steffen Pischke. 2009. *Mostly harmless econometrics: An empiricist's companion*. Princeton, New Jersey: Princeton University Press.
- Araar, Abdelkrim, og Jean-Yves Duclos. 2013. *User manual for Stata package DASP; version 2.3*. Université Laval PEP og World Bank.
- Boeri, Tito, og Jan Van Ours. 2013. *The economics of imperfect labor markets*. Princeton University press.
- Borjas, George J. 1987. "Self-selection and the earnings of immigrants." *American Economic Review*, 77(4):531-553.
- . 2020. *Labor Economics*. New York: Mcgraw-Hill Education.
- Borjas, George J., Ilpo Kauppinen, og Panu Poutvaara. 2019. «Self-selection of emigrants: Theory and evidence on stochastic dominance in observable and unobservable characteristics.» *The Economic Journal*, 129(617):143-171.
- Borjas, George J., og Bernt Bratsberg. 1996. «Who leaves? The outmigration of the foreign-born.» *The Review of Economics and Statistics*, 78(1):165-176.
- Borjas, George J., og Stephen G. Bronars. 1991. «Immigration and the Family.» *Journal of Labor Economics*, 9(2):123-148.
- Bratsberg, Bernt, Oddbjørn Raaum, og Knut Røed. 2010. «When Minority Labor Migrants Meet the Welfare State.» *Journal of Labor Economics*, 28(3):633-676.
- . 2014. «Immigrants, Labour Market Performance and Social Insurance.» *The Economic Journal*, 124(580):644-683.
- Bratsberg, Bernt, og Oddbjørn Raun. 2012. «Immigration and wages: Evidence from construction.» *The Economic Journal*, 122(565):1177-1205.
- Chiquiar, Daniel, og Gordon H. Hanson. 2005. «International migration, self-selection, and the distribution of wages: Evidence from Mexico and the United States.» *Journal of Political Economy*, 113(2):239-281.
- Davidson, Russel, og Jean-Yves Duclos. 2000. «Statistical inference for stochastic dominance and for the measurement of poverty and inequality.» *Econometrica*, 68(6):1453-1464.
- Eurostat. 2020. *EU citizens living in another Member State - statistical overview*. Funnet 22. januar, 2021. https://ec.europa.eu/eurostat/statistics-explained/index.php/EU_citizens_living_in_another_Member_State_-_statistical_overview.
- Folkeregisterloven. 1970. *Lov om folkeregistering av 16. januar 1970 nr. 1*. <https://lovdata.no/pro/lov/1970-01-16-1>.
- . 2016. *Lov om folkeregistrering av 9. desember 2016 nr. 88*. <https://lovdata.no/lov/2016-12-09-88>.
- Gielen, Anne C., og Jan Van Ours. 2010. «Layoffs, quits and wage negotiations.» *Economics letters*, 109(2):108-111.
- Gould, Eric D., og Omer Moav. 2016. «Does high inequality attract high skilled immigrants?» *The Economic Journal*, 126(593):1055-1091.
- Hanushek, Eric A., Guido Schwerdt, Simon Wiedershold, og Ludger Woessmann. 2015. «Returns to skills around the world: Evidence from PIAAC.» *European Economic Review*, 73():103-130.

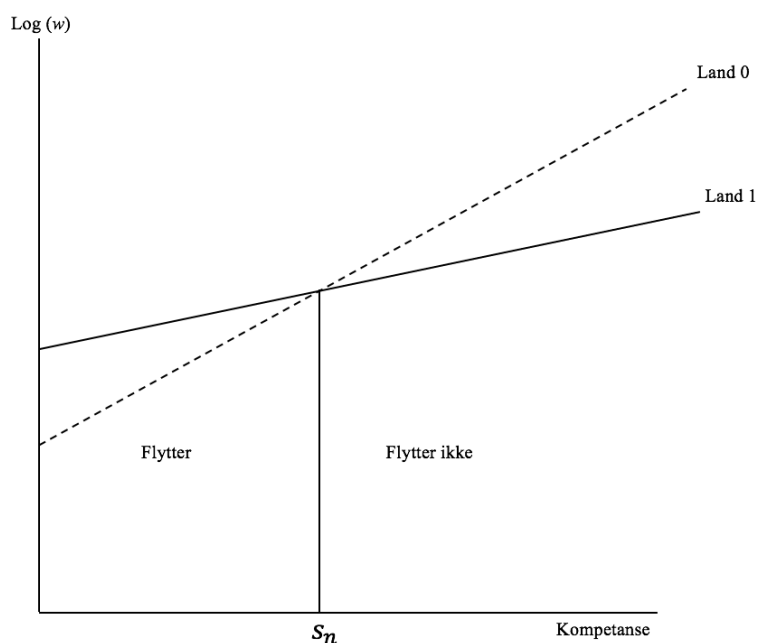
- Hoen, Marie. 2020. "Immigrant and the Tower of Babel: Using Language Barriers to Identify Individual Labor Market Effects of Immigration." *Labour Economics*, 65:101834.
- International organization for migration (IOM). 2020. *World migration report 2020*. Geneva: IOM.
- Joonas, Pernilla Andersson, og Eskil Wadensjö. 2013. «The best and the brightest or the least successful? Self-employment entry among male wage-earners in Sweden.» *Small Business Economics*, 40(1):155-172.
- Kaestner, Robert, og Ofer Malamud. 2014. «Self-selection and international migration: New evidence from Mexico.» *Review of Economics and Statistics*, 96(1):78-91.
- Lemieux, Thomas. 2006. "Increasing residual wage inequality: Composition effects, noisy data, or rising demand for skill?" *American Economic Review*, 96(3): 461-498.
- Moraga, Jesus Fernández-Huertas. 2011. "New evidence on emigrant selection." *Review of Economics and Statistics*, 93(1):72-96.
- NOU 2008:6. 2008. *Kjønn og lønn -Fakta, analyser og virkemidler for likelønn*. Oslo: Barne- og likestillingsdepartementet.
- NOU 2011:7. 2011. *Velferd og migrasjon -den norske modellens framtid*. Oslo: Barne-, likestillings- og inkluderingsdepartementet.
- NOU 2017:2. 2017. *Intergrasjon og tillit- Langsiktig konsekvenser av høy innvandring*. Oslo: Justis og beredsskapsdepartementet.
- OECD. 2021. *Employment rate by age group*. Funnet 30. april, 2021. <https://data.oecd.org/emp/employment-rate-by-age-group.htm>.
- Parey, Mathias, Jens Ruhose, Fabian Waldinger, og Nicolai Netz. 2017. «The selection of high skilled emigrants.» *Review of Economics and Statistics*, 99(5):776-792.
- Patt, Alexander, Jens Ruhose, Simon Wiederhold, og Miguel Flores. 2021. «International emigrant selection on occupational skills.» *Journal of the European Economic Association*, 19(2):1249-1298.
- Pedersen, Peder J., Marianne Røed, og Eskil Wadensjö. 2008. *The common Nordic labour market at 50*. København: Nordic council of ministers.
- Pedersen, Peder J., Marianne Røed, og Lena Schröder. 2003. «Emigration from the Scandinavian welfare states.» I *Alternatives for welfare policy*, redigert av Torben M. Andersen og Per Molander, 76-103. Cambridge: Cambridge University Press.
- Pirttilä, Jukka. 2014. "Is international labour mobility a threat to the welfare state? Evidence from Finland in the 1990s." *Finnish Economic Papers*, 17(1): 18-34.
- Raaum, Oddbjørn, Knut Røed, og Bernt Bratsberg. 2012. «Gjør registerdata AKU overflødig?» I *Økonomiske analyser 5/2012*, redigert av SSB, 46-52. Oslo: SSB.
- Regjeringen. 2021. *Hva EØS-avtalen omfatter*. Funnet 30. mai, 2021. <https://www.regjeringen.no/no/tema/europapolitikk/eos1/hva-avtalen-omfatter/id685024/>.
- Rosso, Anna. 2019. "Emigrant selection and wages: The case of Poland." *Labour Economics*, 60: 148-175.
- Roy, Andrew Donald. 1951. "Some thoughts on the distribution of earnings." *Oxford economic papers*, 3(2):135-146.
- Schröder, Lena. 1996. "Scandinavian skill migration in the 1980s." I *The nordic labour markets in the 1990s part 2*, redigert av Eskil Wadensjö, 71-103. Amsterdam: North-Holland.
- Sheskin, David J. 2003. *Handbook of parametric and nonparametric statistical procedures*. Boca Raton, Florida: CRC Press.
- SSB. 2016. *Norsk standard for utdanningsgruppering 2016 -revidert 2000 dokumentasjon*. Oslo/Kongsvinger: SSB.

- . 2018. *Brattere trapp til lønnstoppen*. Funnet 30. mai, 2021. <https://www.ssb.no/arbeid-og-lonn/artikler-og-publikasjoner/brattere-trapp-til-lonnstoppen>.
- . 2020. *Befolkningens utdanningsnivå*. Funnet 24. mars, 2021. <https://www.ssb.no/utniv/>.
- . 2021a. *Tabell 06574: Gjennomsnittlig månedsførtjeneste for heltidsekvivalenter, etter kjønn, sektor og desiler (avslutta serie) 1997 - 2008*. Funnet 1. mars, 2021. <https://www.ssb.no/statbank/table/06574/>.
- . 2021b. *Yrkesinntekter*. Funnet 21. april, 2021. <https://www.ssb.no/a/metadata/conceptvariable/vardok/49/nb>.
- Statistikkloven. 2019. *Lov om offisiell statistikk og Statistisk sentralbyrå av 21. juni 2019 nr. 32*. <https://lovdata.no/lov/2019-06-21-32>.
- Taeger, Dirk, og Sonja Kuhnt. 2014. *Statistical Hypothesis Testing with SAS and R*. Chichester, UK: John Wiley & Sons.
- Tønnesen, Marianne. 2021. *Migrant*. Funnet 13. mai, 2021. <https://snl.no/migrant>.

Appendiks

A.1 Negativ seleksjon

Figur A.1



A.2 Inndeling av utdanningskategorier

Inndelingen i utdanningskategoriene tar utgangspunkt i «Norsk standard for utdanningsgruppering» som er utarbeidet av SSB. Standarden har blitt revidert flere ganger, men den siste oppdateringen blir omtalt som NUS2000. Standarden er et 6-sifret kodesystem som klassifiserer utdanningsaktiviteter etter nivå og fag (SSB 2016). Registerdataene er anonymisert av Institutt for økonomi ved UiB. Det medfører at jeg kun har tilgang til de tre første sifrene i kodesystemet. De tre første sifrene innebærer at jeg kan dele inn etter utdanningsnivå og utdanningsgruppe på et overordnet nivå. Det er ikke mulig å identifisere den spesifikke utdanningen individet har.

Nivåinndelingen i NUS2000 er bygd opp på følgende måte:

Tabell A.1
Nivåinndeling i NUS2000

Tredeling av nivå	Nivå	Nivånavn
Obligatorisk utdanning	1	Barneskoleutdanning
	2	Ungdomsskoleutdanning
Mellomutdanning	3	Videregående, grunnutdanning
	4	Videregående, avsluttende utdanning
	5	Påbygg til videregående skole
Universitets- og høyskoleutdanning	6	Universitets- og høyskoleutdanning, lavere nivå
	7	Universitets- og høyskoleutdanning, høyere nivå
	8	Forskerutdanning

Merknad: Påbygg til videregående skole omfatter utdanninger som er over videregående skole, men som ikke regnes som høyere utdanning. I denne oppgaven omtaler jeg det som fagskoleutdanning. For en nærmere beskrivelse av denne kategorien og de andre kategoriene se SSB (2016).

Jeg har følgende inndeling av utdanningskategoriene:

Tabell A.2
Nivåinndeling av utdanning i denne oppgaven

Tredeling av nivå	Nivånavn	Hva inngår?
Obligatorisk utdanning	Grunnskoleutdanning	NUS2000-nivå 1 og 2
Mellomutdanning	Videregående, Studieforberevende	NUS200-nivå 3 og 4. Kode 301 og 401.
	Videregående, yrkesfag	NUS2000-nivå 3 og 4. Inkluderer alle koder på nivå 3 og 4 utenom 301 og 401.
	Fagskoleutdanning	NUS2000-nivå 5
Universitets- og høyskoleutdanning	Bachelor eller tilsvarende	NUS2000-nivå 6
	Master eller tilsvarende	NUS2000-nivå 7
	Forskerutdanning	NUS2000-nivå 8

Kategorien grunnskoleutdanning inneholder alle individer med enten barneskole- eller ungdomsskoleutdanning. Det er få individer med kun barneskoleutdanning, og jeg har derfor

valgt å plassere dem sammen med individene som har ungdomsskoleutdanning som høyeste utdanningsnivå.

Jeg skiller mellom individer som har yrkesfaglig og studieforbereende utdanning fra videregående skole. Studieforbereende utdanning er en samlebetegnelse for videregående utdanninger som gir studiekompetanse. Etter Reform 94 ble det opprettet to nye studieretninger som gir studiekompetanse. Det gjelder studieretningene Musikk, dans og drama og Idrettsfag (SSB 2016). Disse to studieretningene faller ikke under kategorien «allmenne fag» i NUS2000-klassifiseringen. Dersom det eksisterer individer i mitt utvalg med disse videregående utdannelsene vil de i mitt utvalg være plassert i kategorien «videregående, yrkesfag». Det er fordi jeg ikke kan identifisere den spesifikke utdanningen til individene i utvalget. Det yngste årskullet i mitt utvalg, 1977-kullet, fylte 17 år da Reform 94 ble innført. Det innebærer at det mest sannsynlig er få individer som har fullført en videregående utdanning som ble opprettet etter Reform 94.

Det er to NUS2000-nivåer for videregående utdanning. Det er «videregående, grunnutdanning» og «videregående, avsluttende utdanning». Til og med skoleåret 1994/1995 angir både NUS2000-nivå 3 og 4 individer som har fullført en videregående utdanning. Fra og med skoleåret 1995/1996 er det kun NUS2000-nivå 4 som angir om individet har fullført en videregående utdanning. Dersom et individ har blitt tildelt NUS2000-nivå 3 før eller i skoleåret 1994/1995 har individet, ut fra SSBs definisjon, fullført en videregående utdanning (SSB 2020). Utvalget mitt inkluderer individer i alderen 25-54 år i årene 1993-2003. Det er derfor ikke mulig for meg å identifisere alle individer som har blitt tildelt NUS2000-nivå 3 etter skoleåret 1994/1995. Individer som har fullført deler av en videregående utdanning kan derfor være plassert i kategoriene «videregående, studieforbereende» eller «videregående, yrkesfag». De inkluderte årskullene i utvalget mitt medfører at de fleste individene som tilhører NUS2000-nivå 3 har havnet i den grupperingen før skoleåret 1994/1995. Jeg legger ikke skjul på at det kan være feil i grupperingen av individer med videregående utdanning, men det er trolig få individer det gjelder.

Mine resterende utdanningskategorier har identisk inndeling som NUS2000. Fra og med skoleåret 1998/1999 inkluderer NUS2000-nivå 6 også individer som har avlagt 120 studiepoeng eller mer i universitets- og høyskolesystemet, uavhengig om de har fullført en grad på høyere nivå. NUS2000-nivå 7 inkluderer alle som har fullført høyere utdanning på mer enn 4 år. NUS2000-nivå 8 inkluderer alle som har fullført en forskerutdanning uavhengig av varighet.

A.3 Metode for å undersøke om fordelingene er ulike

Roy-modellen predikerer at den kumulative fordelingsfunksjonen til emigrantene skal førsteordens stokastisk dominere fordelingen til ikke-emigrantene dersom emigrantene er positivt selektert fra populasjonen. Dersom jeg får resultater som tyder på at emigrantene er positivt selektert, vil formålet med denne metoden være å identifisere intervallet hvor verdiene til den kumulative fordelingsfunksjonen til ikke-emigrantene er statistisk signifikant større enn verdiene til fordelingen til emigrantene³⁶. Formelt prøver metoden å finne det lengste intervallet $[\hat{w}^-, \hat{w}^+]$ som nullhypotesen

$$H_0: \max_{w \in [\hat{w}^-, \hat{w}^+]} [F_M(w) - F_N(w)] \geq 0 \quad (13)^{37}$$

kan bli forkastet i favør av alternativhypotesen av sterk stokastisk dominans.³⁸

$$H_1: \max_{w \in [\hat{w}^-, \hat{w}^+]} [F_M(w) - F_N(w)] < 0 \quad (14)$$

Jeg estimerer differansen mellom de kumulative fordelingsfunksjonene $\Delta[F(w)] = F_N(w) - F_M(w)$ med utvalgsverdiene $\hat{\Delta}[F(w)]$ og regner ut konfidensintervall. Davidson and Duclos (2000) viser at estimatoren $\hat{\Delta}[F(w)]$ er konsistent og asymptotisk normalfordelt. La $\hat{\sigma}(w)$ være standardfeilen til estimatoren $\hat{\Delta}[F(w)]$. Videre, θ angir signifikansnivået og la $z(\theta)$ være 100(1 - θ) persentilpoeng av en standard normalfordeling. Standardfeilene blir estimert ved å bruke Taylor linearisering og blir gruppert på individ (clustering). Ligning (15) viser hvordan den nedre grensen til konfidensintervallet blir regnet ut.

$$\widehat{LB}_{\Delta(F(w))} = \hat{\Delta}[F(w)] - \hat{\sigma}(w)z(\theta) \quad (15)$$

Ligning (16) viser hvordan den øvre grensen til konfidensintervallet blir regnet ut.

$$\widehat{UB}_{\Delta(F(w))} = \hat{\Delta}[F(w)] + \hat{\sigma}(w)z(\theta) \quad (16)$$

³⁶ Dersom en får resultater som indikerer at emigrantene er negativt selektert vil formålet med metoden være å identifisere intervallet hvor verdiene til den kumulative fordelingsfunksjonen til emigrantene er statistisk signifikant større enn verdiene til fordelingen til ikke-emigrantene

³⁷ $F_M(z)$ ($F_N(z)$) representerer den kumulative fordelingsfunksjonen til emigrantene (ikke-emigrantene).

³⁸ Alternativet er sterk dominans, ettersom svak dominans ikke kan bli separert statistisk fra nullhypotesen av ikke-dominans.

For å utføre disse utregningene benytter jeg statapakken til Araar og Duclos (2013). Statapakken regner ut differansen mellom kumulative fordelingsfunksjoner, og viser om differansen er statistisk signifikant ved å rapportere et 95 % konfidensintervall.

A.4 Kolmogorov-Smirnov-test

Følgende framstilling er basert på Sheskin (2003) og Taeger og Kuhnt (2014). En Kolmogorov-Smirnov toutvalgs-test tester om to uavhengige utvalg stammer fra den samme fordelingen. Kolmogorov-Smirnov-testen er en ikke-parametrisk test. Testen krever ikke at en spesifiserer fordelingene utenom at fordelingene er kontinuerlige.

Testen har følgende antakelser:

- Dataene er i det minste målt på en ordinal skala.
- Observasjonene må være identisk og uavhengig fordelt (i.i.d.). Ettersom jeg har paneldata er observasjonene for et individ trolig korrelert over tid. Det innebærer at antakelsen om uavhengighet på tvers av observasjonene ikke er oppfylt. Det er likevel rimelig å tro at denne antakelsen er oppfylt på tvers av individene.
- De to utvalgene er uavhengige med kontinuerlige fordelingsfunksjoner $F_M(w)$ og $F_N(w)$.

Jeg utfører hovedsakelig en tohaletest. Av den grunn formulerer jeg følgende null- og alternativhypotese.

$$H_0 = F_M(w) = F_N(w)$$

$$H_1 = F_M(w) \neq F_N(w) \text{ for minst en } w$$

Teststatistikken til Kolmogorov-Smirnov-testen er forskjellig om en utfører en tohale- eller enhaletest. Teststatistikken til en tohaletest er den største absolutte avstanden mellom de to kumulative fordelingsfunksjonene:

$$D = \max_w |(F_M(w) - F_N(w))|$$

En forkaster nullhypotese dersom teststatistikken er større eller lik den kritiske verdien. Formelt kan det vises som $D \geq d_{\theta, n_M, n_N}$. θ angir signifikansnivået, og n_M og n_N angir utvalgsstørrelsen til de to utvalgene som benyttes i testen. Den kritiske verdien for det aktuelle signifikansnivået og utvalgsstørrelsene, d_{θ, n_M, n_N} , kan bli funnet i tabell A.23 i Sheskin (2003).

A.5 Lineær sannsynlighetsmodell

Tabell A.3
Lineær sannsynlighetsmodell

	(1) Menn	(2) Kvinner	(3) Menn	(4) Kvinner
Avhengig variabel: Emigrasjonsstatus				
Barn			-0,0009*** (0,0000)	-0,0011*** (0,0000)
Gift			-0,0002*** (0,0000)	-0,0003*** (0,0001)
VGS, studiefor.	0,0011*** (0,0001)	0,0005*** (0,0001)	0,0011*** (0,0001)	0,0005*** (0,0001)
VGS, yrkesfag	0,0000 (0,0000)	-0,0000 (0,0000)	0,0001** (0,0001)	0,0000 (0,0001)
Fagskole	0,0010*** (0,0001)	0,0008*** (0,0001)	0,0011*** (0,0001)	0,0009*** (0,0001)
Bachelor	0,0021*** (0,0001)	0,0013*** (0,0001)	0,0022*** (0,0001)	0,0014*** (0,0001)
Master	0,0047*** (0,0001)	0,0036*** (0,0001)	0,0048*** (0,0001)	0,0036*** (0,0001)
PhD	0,0087*** (0,0005)	0,0059*** (0,0008)	0,0089*** (0,0005)	0,0060*** (0,0008)
Konstantledd	0,0016*** (0,0001)	0,0028*** (0,0002)	0,0019*** (0,0001)	0,0030*** (0,0002)
N	6 678 597	4 141 205	6 678 597	4 141 205
Alders. effekter	Ja	Ja	Ja	Ja
Tids. effekter	Ja	Ja	Ja	Ja
R ²	0,0015	0,0010	0,0016	0,0012

Merknad: Den avhengige variabelen er emigrasjonsstatus. Det er en dummyvariabel som tar verdien 1 dersom individet emigrerer ut av Norge påfølgende år, og 0 ellers. Inkluderer både korttids- og langtidsemigranter. Klyngerobuste standardfeil i parentes som er gruppert på individ. Referansekategori er grunnskoleutdanning.

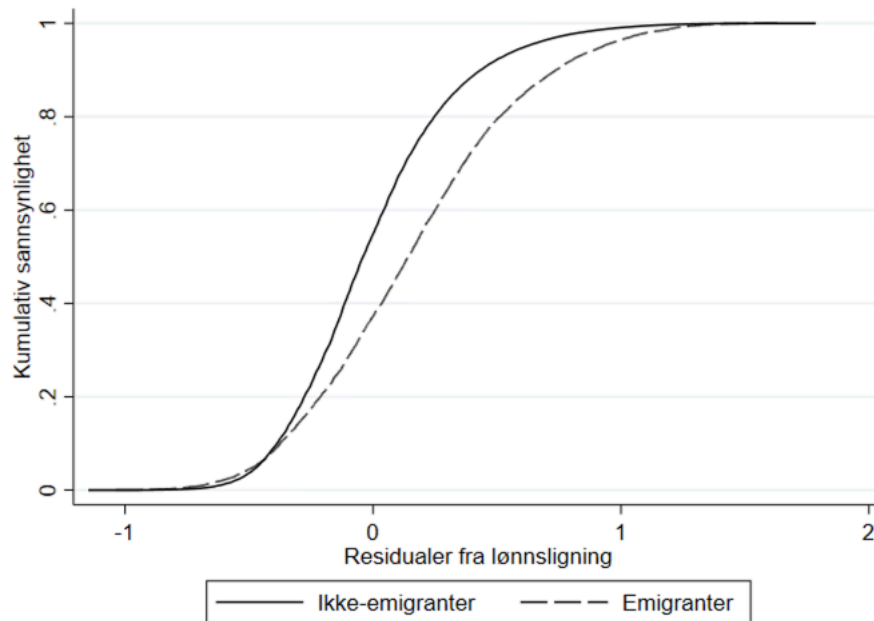
* $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

A.6 Kumulative fordelingsfunksjoner

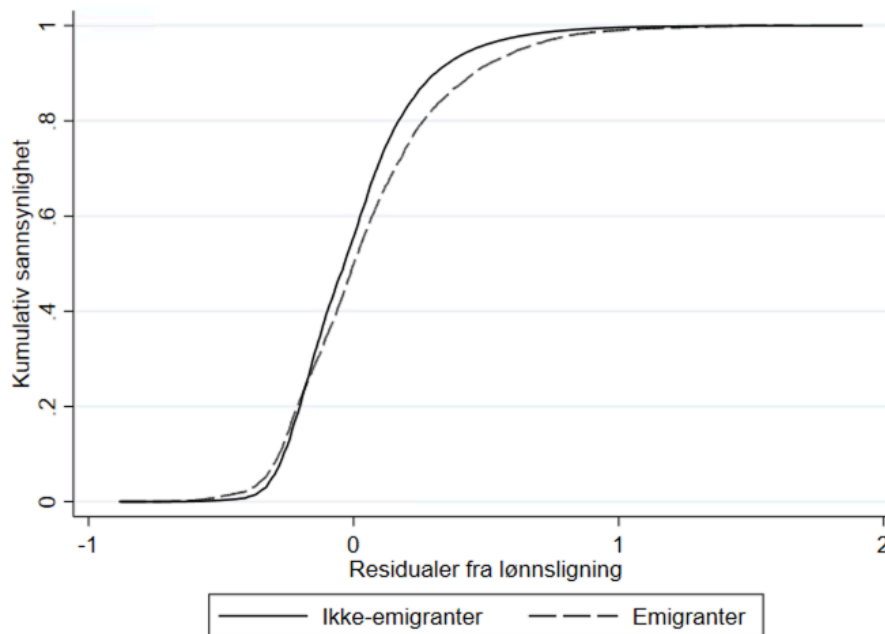
Figur A.2

Kumulative fordelingsfunksjoner av residualer fra lønnslikning med ikke-standardisert inntekt som avhengig variabel. Forskjell mellom ikke-emigranter og emigranter

a) Menn



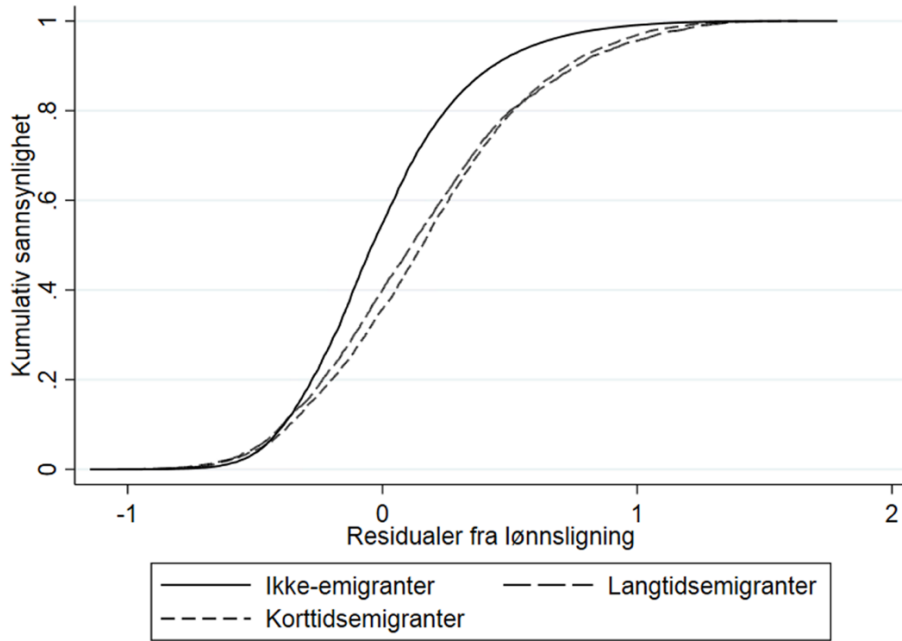
b) kvinner



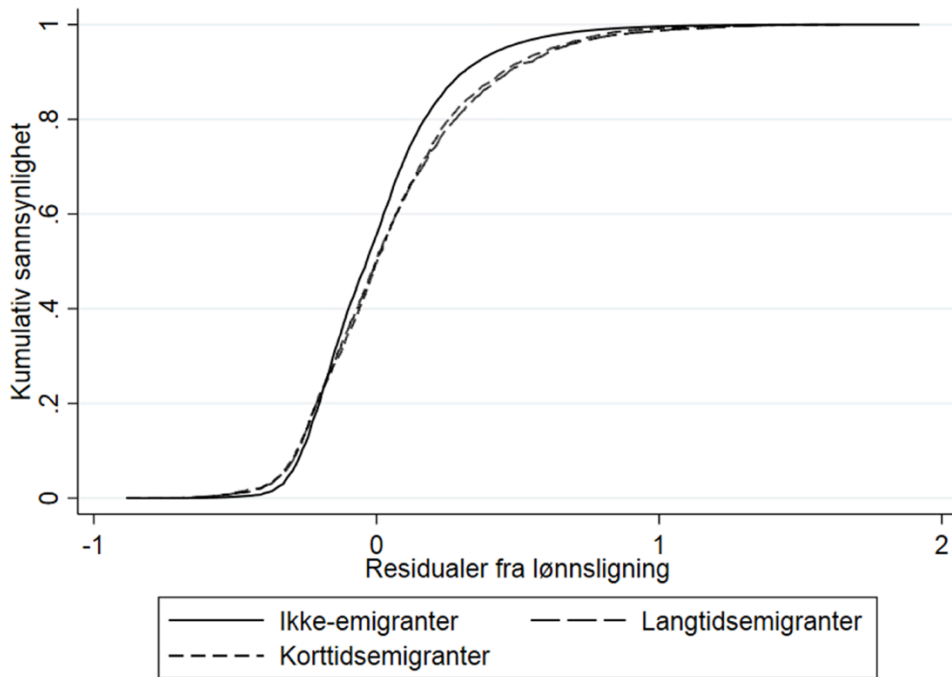
Figur A.3

Kumulative fordelingsfunksjoner av residualer fra lønnslikning for ikke-emigranter, langtidsemigranter og korttidsemigranter. Ikke-standardisert inntekt som avhengig variabel

a) Menn



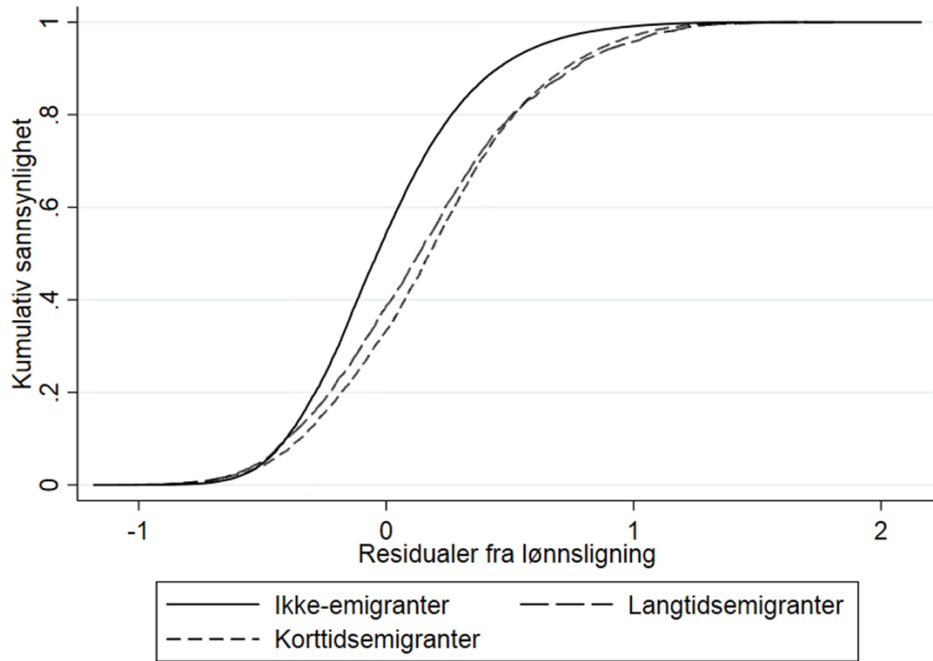
b) Kvinner



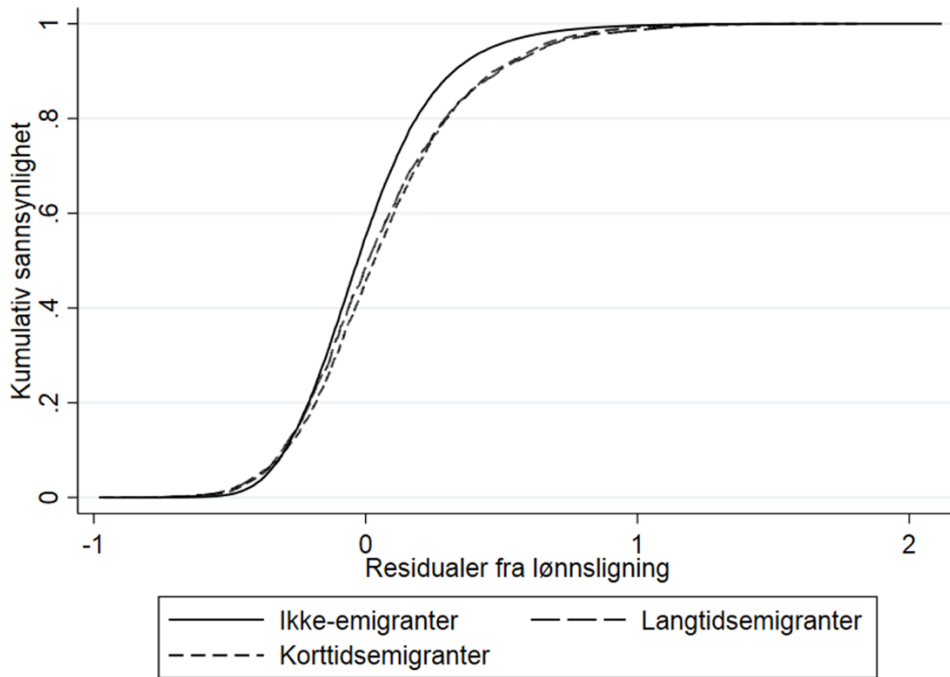
Figur A.4

Kumulative fordelingsfunksjoner av residualer fra lønnslikning for ikke-emigranter, langtidsemigranter og korttidsemigranter. Standardisert inntekt som avhengig variabel

a) Menn



b) Kvinner



A.7 Sensitivitetsanalyse for kvinner

Tabell A.4

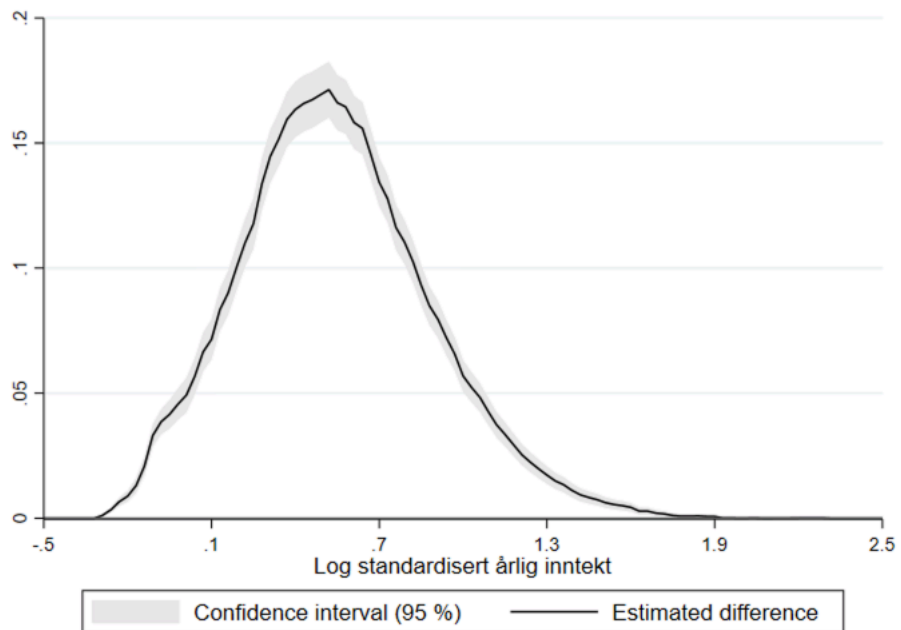
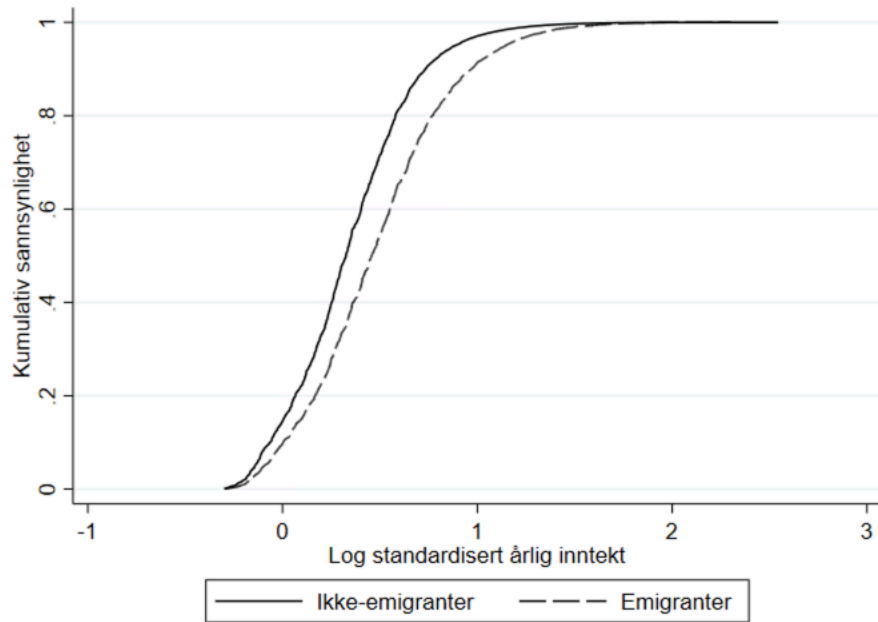
Deskriptiv statistikk for kvinner, inntettsgrense på 150 000 (2015-kroner)

	Alle	Ikke- emigranter	Emigranter	Korttids- emigranter	Langtids- emigranter
Observasjoner	4 778 146	4 770 488	7 658	5 127	2 531
Alder					
Gjennomsnitt	39,64	39,65	35,52	35,56	35,43
Median	40	40	34	34	33
Årlig inntekt					
Gjennomsnitt	5,81	5,81	6,49	6,50	6,49
Median	5,46	5,46	5,86	5,86	5,72
Standardisert årlig inntekt					
Gjennomsnitt	1,49	1,48	1,74	1,76	1,71
Median	1,38	1,38	1,58	1,61	1,53
Utdanning (år)					
Gjennomsnitt	13,23	13,23	14,77	15,08	14,13
Median	13	13	15	16	14
Utdanningsgrupper (%)					
Grunnskole	17,47	17,48	8,81	6,67	13,16
VGS, studieforbredende	13,92	13,93	10,76	9,44	13,49
VGS, yrkesfag	29,06	29,09	15,57	13,58	19,06
Fagskole	2,94	2,94	3,47	3,34	3,75
Bachelor og tilsvarende	31,65	31,62	46,67	50,26	39,39
Master og tilsvarende	4,76	4,74	13,85	15,62	10,27
Forskerutdanning (PhD)	0,2	0,2	0,86	1,09	0,4

Merknad: Årlig inntekt måles i enheter, hvor en enhet er 50 000 kr. Årlig inntekt er deflatert til 2015-kroner. De to første kolonnene angir observasjoner, og de tre siste angir individer.

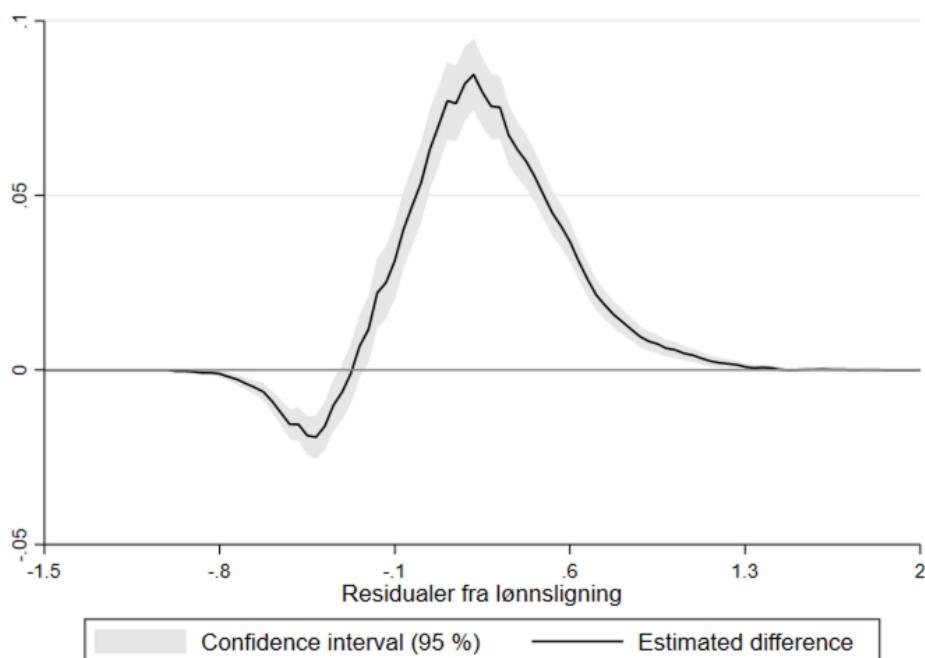
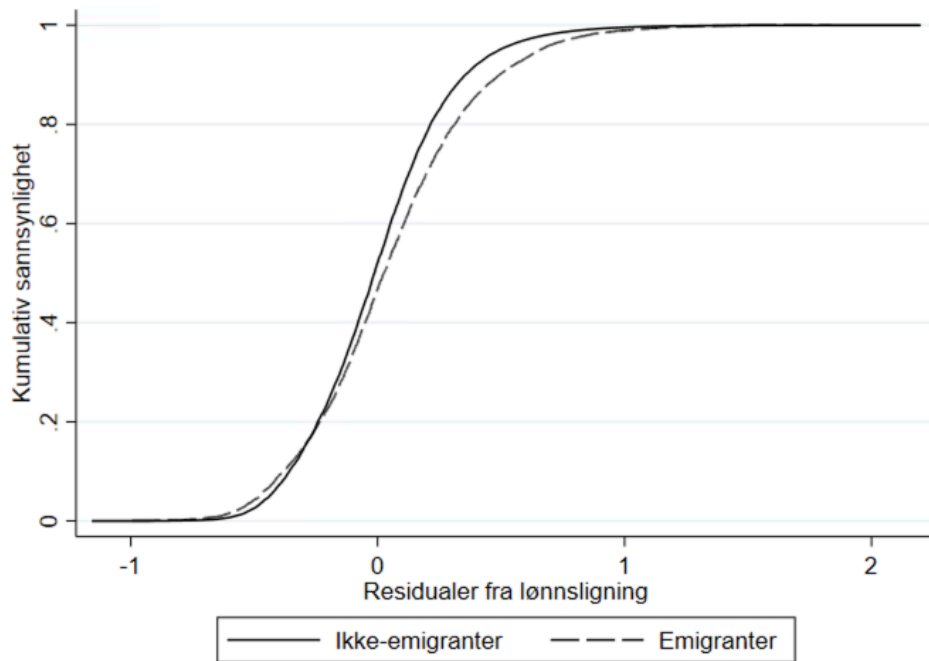
Figur A.5

Kumulative fordelingsfunksjoner av standardisert inntekt for kvinnelige ikke-emigranter og emigranter. Tilhørende figur er estimert differanse med konfidensintervall. Utvalget er observasjoner med inntekt over 150 000 kr (2015-kroner)



Figur A.6

Kumulative fordelingsfunksjoner av residuallønninger for kvinnelige ikke-emigranter og emigranter fra regresjon med standardisert inntekt som avhengig variabel. Tilhørende figur er estimert differanse med konfidensintervall. Utvalget er observasjoner med inntekt over 150 000 kr (2015-kroner)



A.8 Deskriptiv statistikk for emigranter med kjent destinasjonsområde

Tabell A.5

Deskriptiv statistikk for emigranter med kjent destinasjonsområde

	Menn			Kvinner		
	Ikke-emigranter	Emigranter Norden	Emigranter u. Norden	Ikke-emigranter	Emigranter Norden	Emigranter u. Norden
Observasjoner	6 666 868	2 261	6 527	4 134 383	1 325	3 830
Alder						
Gjennomsnitt	39,57	35,15	37,65	39,78	34,39	36,09
Median	40	33	37	40	32	34,5
Årlig inntekt						
Gjennomsnitt	8,30	9,58	11,63	6,16	6,71	6,91
Median	7,33	7,95	10,40	5,72	6,09	6,09
Standardisert årlig inntekt						
Gjennomsnitt	1,20	1,45	1,73	1,57	1,77	1,88
Median	1,07	1,26	1,60	1,45	1,62	1,71
Utdanning (år)						
Gjennomsnitt	13,06	13,99	15,77	13,40	14,35	15,4
Median	13	13	16	13	14	16
Utdanningsgrupper (%)						
Grunnskole	19,71	17,51	4,21	15,85	12,23	4,62
VGS, studieforberevende	8,46	10,08	7,12	13,75	10,57	9,11
VGS, yrkesfag	39,57	24,86	12,66	28,01	19,7	10,65
Fagskole	4,76	4,56	5,03	3,03	3,02	3,39
Bachelor og tilsvarende	19,06	26,36	35,33	33,82	42,34	52,9
Master og tilsvarende	7,90	15,08	31,58	5,30	11,09	18,25
Forskerutdanning (PhD)	0,55	1,55	4,08	0,23	1,06	1,07

Merknad: Årlig inntekt måles i enheter, hvor en enhet er 50 000 kr. Årlig inntekt er deflatert til 2015-kroner.