

Ellen Reitan

Mind the gap: Effekten av yrkesheterogenitet og kunstig intelligens på lønnsulikhet

En empirisk analyse av norske lønnsdata i perioden 1999-2014

Masteroppgave i samfunnsøkonomi
Veileder: Ragnar Torvik (NTNU) og Andreas Kostøl (Norges Bank)
Mai 2019

Norges teknisk-naturvitenskapelige universitet
Fakultet for økonomi
Institutt for samfunnsøkonomi

Forord

Denne masteroppgaven markerer avslutningen på mastergraden min ved Institutt for samfunnsøkonomi på NTNU. Jeg fikk tilbringe siste semester av graden i forskningsavdelingen til Statistisk sentralbyrå, hvor masteroppgaven ble skrevet. Den viktigste takken går til Andreas Kostøl og Ragnar Torvik for ekstraordinær veilederinnsats. Både for verdifulle innspill og avgjørende støtte underveis, men også for tålmodighet med dumme og mindre dumme spørsmål til alle døgnets tider. Dette hadde ikke gått uten dere.

Videre må jeg takke Marek Jasinski for alle innspill og gjennomlesinger med (irriterende) detaljerte tilbakemeldinger. Mastergraden hadde ikke vært den samme uten min trofaste sparringspartner Karine Handegaard Bakken, takk både for godt samarbeid, for trøst i eksamensknekker og for at du alltid, alltid passet pausetida. Takk også til gjengen i SSB som tok imot en halvforvirret trønder og gjorde masterhalvåret veldig morsomt og lærerikt.

Takk til hele instituttet for en fantastisk studietid, og en ekstra takk til Ragnar Torvik for å ha orket alt maset mitt over flere år.

Og til slutt må jeg takke alle som har holdt ut med meg og mitt lidenskapelige forhold til lesesalen i fem år. Lover å komme mer ut i dagslys nå.

Masteroppgaven er i sin helhet mitt selvstendige arbeid, synspunkter og eventuelle feil er derfor mine egne.

Ellen Reitan

Oslo, 31.mai 2019

Sammendrag

Formålet med denne oppgaven er tredelt. Først studerer jeg norske lønnsdata for perioden 1999-2014 for å undersøke om Norge følger den globale trenden med økende inntektsulikhet. Deretter foretar jeg en AKM-dekomponering av lønnsvariabelen til individ- og yrkesfaste effekter. Dette er en metode som ble utviklet av Abowd, Kramarz og Margolis i 1999 for å kunne studere lønnsvariasjon dekomponert til effekter som er spesifikke for individer og firma. Formålet med metoden var altså originalt å undersøke hvor mye av variasjonen i lønn som kan forklares av heterogenitet knyttet til firma sammenlignet med individer. I denne oppgaven utvider jeg denne metodikken til å studere yrkesheterogenitet, i tillegg til firmaspesifikke effekter. Denne dekomponeringen gjør at jeg kan undersøke om yrkesheterogenitet er en sentral driver for økende lønnsforskjeller, eller om utviklingen utelukkende kan tilskrives individeffekter. Til slutt undersøker jeg om automatisering og kunstig intelligens kan forklare de yrkesspesifikke effektene.

Det har lenge vært et politisk mål å redusere økonomisk ulikhet i Norge, senest eksemplifisert med stortingsmeldingen om ulikhet fra 2019. Likevel finner jeg at lønnsulikheten har økt gjennom den aktuelle tidsperioden, i form av at vi observerer en større spredning i lønnsvariabelen. Videre tyder resultatene på at denne økningen ikke kan forklares av yrkesspesifikke effekter, finner også at dekomponering til bedriftseffekter viser at inntektsulikheten drives av de individspesifikke effektene. Analysene av yrkesheterogenitet viser liten effekt av variabler som fanger opp teknologisk utvikling, og tegner et bilde av at automatisering og kunstig intelligens ikke forklarer noe av endringen i de yrkesfaste effektene. Disse funnene er et bidrag til den hurtig voksende litteraturen knyttet til effektene av automatisering på arbeidsmarkedet, men merk at tidsperioden jeg studerer ikke strekker seg lengre enn til 2014. Man kan argumentere for at sannsynligheten vil øke for signifikante effekter fra automatisering på nyere data.

Abstract

The purpose of this thesis is threefold. First, I study Norwegian wage data from the period 1999-2014 to investigate whether Norway follows the global trend with increasing wage inequality. Then I do an AKM decomposition of the wage variable to study the roles of firm- and individual-specific effects. This method was developed by Abowd, Kramarz and Margolis in 1999 to be able to study the wage variation decomposed to effects which is specific to firms and individuals. The original purpose of the method was, in other words, to investigate how much of the variation in wage that can be explained by heterogeneity related to firms compared to individuals. In this thesis I expand this methodology to study the heterogeneity related to occupations, in addition to firm-specific effects. This decomposition makes me able to study whether occupational heterogeneity plays a central role in the rising inequality in wages, or if the development in wages can only be attributed to individual effects. Finally, I investigate if automatization and artificial intelligence can explain the occupation-specific effects.

It has been a political goal to reduce the economic inequality in Norway, most recently exemplified with the white paper about inequality from 2019. Nevertheless, I find that the wage inequality has been rising throughout the period studied in this thesis, measured by a wider wage variable spread. Further, the results suggest that this increase cannot be explained by the occupation-specific effects, and I also find that a decomposition to firm-specific effects shows that the wage inequality is driven by the individual effects. The analyses of the occupational heterogeneity shows little effect of the variables that measures technological progress, and by that indicate that automatization and artificial intelligence does not contribute to explain the changes in the occupation-specific effects. These results are a contribution to the growing literature on the effects of automatization on the labour market, but note that the period I study does not extend beyond 2014. One can argue that the probability will increase for significant effects from automatization with more recent data.

Innholdsfortegnelse

1. Introduksjon og motivasjon	1
2. Teoretisk bakgrunn	5
2.1 Lønnsulikhet	5
2.2 Institusjonelle forhold	9
2.3 Oppsummering	10
3 Datamaterialet	13
3.1 Datautvalg	13
3.2 Avhengig variabel	14
3.3 Tidsvarierende, uavhengige variabler	16
4 Metodisk rammeverk	19
4.1 Modellspesifikasjon	20
4.2 Økonometriske utfordringer	24
5 Analyse	31
5.1 Resultater	31
5.2 Betydning av yrker	38
6 Konklusjon	45
Referanser	I

1 Introduksjon og motivasjon

Økende ulikhet har vært et populært diskusjonstema de siste årene, både i faglitteraturen og i samfunnsdebatten generelt. Samtidig som økonomien globalt vokser, og verden som helhet slik blir rikere, fordeles denne rikdommen stadig mer ulikt mellom oss.¹ En økende grad av litteraturen på feltet setter inntektsutviklingen i nyere tid i kontekst med raskt økende globalisering og teknologisk utvikling.² I praksis handler dette om å forklare økende lønnsulikhet som en konsekvens av at tilbuds- og etterspørselsfaktorer har påvirket produktivitetsgapet mellom høy- og lavkompetansearbeidere, som Card, Heining og Kline (2013) påpeker. Mye tyder på at en økonomi i rask endring også endrer avkastningen på produktivitet, hvor de mest produktive opplever høy inntektsvekst og de mindre produktive møter økende risiko for å bli utkonkurrert av maskiner. Det er ikke nytt at markedet skaper vinnere og tapere, det interessante spørsmålet vil nå være om utviklingen medfører at vinnerne vinner mye mer enn før, slik at distansen ned til de andre øker mer, og raskere enn det vi tidligere har observert. Satt på spissen vil 1. plassen i en “winner takes it all”-økonomi kunne bli milliardær, mens 2. plassen blir arbeidsledig.

Et av økonomifagets klassiske læresetninger går ut på at gitt perfekt konkurranse så skal arbeidernes lønn være lik verdien av marginalproduktiviteten til den enkelte. Sagt på en annen måte: Verdien av det man skaper skal være ekvivalent med verdien av det man tjener. Dette impliserer at identiske arbeidere skal motta identisk lønnskompensasjon. Likevel gjør de i mange tilfeller ikke det. For de fleste av oss utgjør lønn majoriteten av inntekten. Raskt økende lønnsulikhet vil derfor endre de relative økonomiske rammene i samfunnet, og slik fordele makt og muligheter mer skjevt. Å studere hva som driver denne utviklingen, og om disse bakenforliggende faktorene har endret seg over tid, er derfor viktige brikker for å forstå det komplekse fenomenet ulikhet.

Et voksende gap i lønnskompensasjon mellom høy- og lavproduktive arbeidere kan forklare mye av økningen i inntektsulikheten, men det forklarer ikke hvorfor like produktive arbeidere lønnes ulikt. Dette er et fenomen økonomer har forsket på lenge, og det er også opphavet til metoden Abowd, Kramarz og Margolis (AKM) utviklet i 1999 i sin studie av lønnsulikhet.

¹ Ulikheten på tvers av land er fallende, men den øker internt i land (Roser og Ortiz-Ospina, 2016).

² Se for eksempel Goldin og Katz, (2008), og Autor og Acemoglu (2011).

AKM-metoden er benyttet i en rekke studier, og utgjør også fundamentet for metoden jeg bruker i denne oppgaven. Formålet med metoden er å dekomponere variasjon i lønn til å kunne studere hvor mye av variasjonen som kan forklares av firma- og individspesifikke effekter. Jeg foretar en utvidelse av dette metodiske rammeverket ved først å se på firmafaste effekter, og deretter aggregerer til å studere yrkesheterogenitet.

Mitt bidrag til denne litteraturen er en studie av lønnsulikhet i Norge i perioden 1999 til 2014. Jeg skal først undersøke om ulikheten har økt i løpet av den aktuelle tidsperioden, og deretter skal jeg ved hjelp av en AKM-dekomponering undersøke om veksten i ulikhet skyldes heterogenitet knyttet til yrker. For å undersøke inntektsulikheten tar jeg i bruk den samme lønnsmodellen som Card m.fl. benyttet i sin artikkel fra 2013, og ved hjelp av en Two way fixed effects-modell skal jeg forsøke å si noe om hvor mye av lønnsulikheten som kan forklares av individfaste effekter, og hvor mye som kan forklares av yrkesfaste effekter. Nullhypotesen min er at ingenting av økningen i lønnsulikhet kan forklares av yrkesheterogenitet, og at veksten dermed i sin helhet kan tilskrives individfaste effekter. Den empiriske analysen vil i tillegg ta for seg en undersøkelse av firmaspesifikke effekter, etter modell fra Card m.fl. (2013). Jeg tester til slutt om potensielt økende betydning av yrker drives av teknologisk utvikling og automatisering.

AKM-metoden er benyttet for å undersøke bidrag fra firmafaste effekter på data fra en rekke land med ulike problemstillinger, men så vidt meg bekjent har ingen brukt metoden til å dekomponere lønnsulikhet til yrker og individer.

Funnene i denne masteroppgaven kan i all hovedsak oppsummeres i tre konklusjoner. Lønnsulikheten har økt i Norge i løpet av den aktuelle tidsperioden, i tråd med globale trender. Denne økende lønnspredningen kan ikke tilskrives yrkesspesifikke effekter, i stedet drives den økte ulikheten av individeffekter. Gitt konklusjon nummer to blir det dermed ikke så overraskende at jeg ikke finner empirisk dekning for at teknologisk utvikling forklarer endring i yrkesheterogenitet. Dersom yrker har liten betydning for lønnsulikhet følger det naturlig at teknologisk utvikling trolig heller ikke er så viktig for lønn. Det er altså lite som tyder på at automatisering og kunstig intelligens kan forklare den økende inntektsulikheten.

Resten av oppgaven er strukturert i fem seksjoner. Jeg begynner med en teoretisk gjennomgang av lønnsulikhet, med drøfting av noen sentrale institusjonelle forhold samt

gjennomgang av relevant litteratur på feltet. I seksjon 3 gjør jeg rede for datamaterialet som benyttes i oppgaven, og i seksjon 4 går jeg igjennom det metodiske rammeverket. Seksjon 5 gir resultatene fra den empiriske analysen, samt forsøk på å forklare drivere bak yrkesheterogenitet. Oppgaven avsluttes med konklusjon i seksjon 6.

2 Teoretisk bakgrunn

I dette kapitlet skal jeg gjøre rede for fenomenet lønnsulikhet, samt drøfte relevante bidrag til litteraturen på feltet. Jeg begynner med det historiske bakteppet, før jeg videre ser på mulige drivere for trenden med økende ulikhet. Jeg går så igjennom noen sentrale institusjonelle forhold knyttet til lønnsdannelse og arbeidsmarkedet, før jeg til slutt oppsummerer.

2.1 Lønnsulikhet

Målet med denne oppgaven er å analysere effekten av ulike variabler på lønnen til fulltidsansatte arbeidere. Det er derfor viktig å skille lønnsulikhet fra formuesulikhet, fordi formue er skjeverte fordelt enn inntekt (Finansdepartementet, 2019). Likevel tyder mye på at disse henger tett sammen. IMF (2017) finner at inntektsulikhet samvarierer med fordelingen mellom kapital og lønn sin andel av total verdiskaping. Ved redusert lønnsandel stiger inntektsulikheten både før og etter skatt, videre finner de at denne sammenhengen gjelder både i klassisk industrialiserte land og i land med vesentlig lavere BNP. IMF hevder med andre ord at den funksjonelle inntektsfordelingen samvarierer med den personlige inntektsfordelingen, altså den som fordeles på hushold og personer (Aaberge, 2016). OECD (2018) finner at i gjennomsnitt har lønnsandelen falt i sine medlemsland med drøye 3 % i perioden 1995-2013, men merk at den for Norge sin del har økt noe. Hvis man legger IMF sine beregninger til grunn tyder dette på at den økte i lønnsandelen har bremsset økningen i inntektsulikheten i Norge.

2.1.1 Nasjonale og globale trender

Utviklingen i norsk økonomi generelt, og i økonomisk ulikhet, var preget av både høy vekst og store tilbakeslag gjennom 1900-tallet. Børskrakket i New York i 1929 ga en økonomisk tøff start på 30-tallet for mange land, deriblant Norge. Likevel var dette et tiår preget av vekst i både sysselsetting og lønninger, og Grytten (1995) anslår at sysselsettingen gjennomsnittlig økte med 2 % hvert år i perioden 1932-1939. Årene med krig og okkupasjon representerte et stort tilbakeslag, blant annet i form av 50 000 færre sysselsatte i 1945 sammenlignet med

1939 (Hansen og Skoglund, 2008). Store kriger har opp igjennom historien redusert ulikhetene i samfunnene som rammes, blant annet fordi store mengder kapital ødelegges, men for Norges del stoppet ikke trenden med reduserte forskjeller etter krigens slutt i 1945.

Utviklingen med redusert ulikhet vedvarte i Norge, og i mange andre vestlige land, frem til midten av 1980-tallet (Aaberge og Modalsli, 2014).

For Norges del kan noe av trendsiftet på 80-tallet tilskrives deregulering av kapitalmarkedet. Fra en situasjon i etterkrigsårene hvor bankene var svært strengt regulert ble det etter dereguleringen betydelig lettere å låne penger (Hansen og Skoglund, 2008). I kombinasjon med lavere beskatning av kapitalinntekter bidro dette trolig til at ulikhetene begynte å øke. Det er likevel svært sannsynlig at en liten, åpen økonomi i stor grad preges av internasjonale trender.

Stortingsmeldingen om ulikhet trekker frem at økende lønnsulikhet globalt skyldes en kombinasjon av oppgang i inntekten til de rikeste og lav lønnsvekst for de på bunn, men at lite har skjedd i midtsjiktet (Finansdepartementet, 2019, side 121). Aaberge og Modalsli (2014) deler denne oppfatningen, og viser i sin forskning til at inntektsandelen til de rikeste faktisk økte mer i Norge på 1990-tallet enn i de fleste andre europeiske land. Driverne bak økende ulikhet kan forklares av faktorer som endrede skattesystemer,³ deregulering av markeder og økt innvandring. Men årsakene kan også forklares av langt mer komplekse drivere knyttet til globalisering og teknologisk utvikling. Jeg kommer nærmere tilbake til dette nedenfor. Selv om trenden med økende ulikhet fra 80-tallet har vedvart, er inntektsforskjellene lavere i Norge enn i de fleste andre land. Både målt ved OECDs tall for gini-koeffisient og for P90/P10-fordelingen rangeres Norge som det landet med 5. lavest inntektsulikhet (OECD, 2019a).^{4 5}

Utviklingen jeg nå svært overfladisk har skissert kan på mange måter sies å være bakteppet for den voksende litteraturen på ulikhet. Mange av de refererte studiene har tatt for seg et bredere ulikhetsbegrep, også de som har sett på inntektsulikhet, enn det jeg skal analysere i denne oppgaven. Jeg vil likevel argumentere for at den historiske utviklingen og den

³ Omfattende skattereform ble vedtatt i Norge i 1992. Fulgte internasjonal trend med bredere skattegrunnlag og lavere sats.

⁴ Gini-koeffisienten er et populært mål på inntektsulikhet. Den tar verdi mellom 0 og 1, og måler fordelingen av inntekt i befolkningen. Der 0 indikerer at alle innbyggere mottar akkurat like stor inntekt, og 1 indikerer at én person i økonomien mottar all inntekt. P90/P10 er en rate som viser forholdet mellom inntekten til individet med 90 % høyest inntekt og individet med 10 % lavest inntekt i økonomien (OECD, 2016).

⁵ Basert på tall fra 2017.

omfattende empirien på feltet er et relevant utgangspunkt for min analyse av økende lønnsulikhet.

2.1.2 Globalisering og teknologi

En illustrerende, om enn banal, beskrivelse av fenomenet globalisering er at verden blir mindre. I dette legger mange økonomer at økonomiske aktører oppfører seg mer og mer som om alle er en del av et og samme marked. Det som tidligere var mange separate økonomier adskilt av landegrensler, handelsbarrierer og andre reguleringer har de siste tiårene blitt mer og mer integrert i et stort verdensmarked. Utviklingen er drevet frem av et samspill mellom internasjonalt samarbeid mellom land og teknologisk fremgang som har gjort flyt av varer, kapital og tjenester både enklere og billigere (Kolb, 2018). Det er konsensus blant økonomer om at det globale resultatet har vært økonomisk vekst av en slik karakter som neppe hadde vært mulig uten den økende integrasjonen mellom de ulike landene. En viktig del av dette bildet er hvordan denne veksten har fordelt seg mellom land, men av særlig interesse for denne oppgaven, hvordan den har fordelt seg mellom individer. Tabloid formulert viser empirien at det ikke eksisterer en automatikk i at større kake å fordele gir mer kake til alle. Et av de fremste eksemplene på dette er India, som er en av de hurtigst voksende økonomiene i verden, men hvor de 1 % rikeste sitter på 73 % av all rikdom (Oxfam, 2019). Det eksisterer derfor en omfattende litteratur rundt implikasjonene for økonomisk ulikhet av globalisering og økt handel.

Teknologisk utvikling har ikke bare fungert som rammeverk og døråpner for økt internasjonal handel. Det har også fundamentalt endret måten økonomien er skrudd sammen på. Den hurtige utviklingen med blant annet robotisering og digitalisering omtales av mange som den fjerde industrielle revolusjonen, hvor automatiseringen som tidligere erstattet manuelle og repetitive oppgaver nå i større og større grad erstatter kognitive arbeidsprosesser som mennesker til nå har hatt et konkurransefortrinn i (Acemoglu og Restrepo, 2018). Sagt enklere, det som tidligere erstattet muskelkraft erstatter nå i større og større grad hjernekraft. I sin artikkel fra 2003 argumenterer Autor, Levy og Murnane for at f.eks. bilkjøring ikke ville bli gjenstand for automatisering, fordi oppgavens innhold er av en så kompleks karakter at den ikke kan løses av maskiner. Dette utsagnet, som støtter seg på blant annet Pinkers artikkel fra 1997, illustrerer godt hvor raskt teknologien utvikler seg. Ideen om at komplekse oppgaver

best lar seg løse av det menneskelige intellektet er, drøyt 15 år etter Autors påstand, allerede avleggs.

Teknologisk utvikling er ikke et nytt fenomen, det er heller ikke debatten rundt effekten av automatisering på etterspørselen etter arbeidskraft. Den britiske økonomen John Maynard Keynes hevdet i 1930 at automatisering vil skape «teknologisk arbeidsledighet», fordi maskinenes inntog i økonomien skulle gjøre arbeidskraft overflødig (Keynes, 1930). Motstykket til Keynes kan på mange måter sies å være Peter Drucker. Den østerrikske samfunnsviteren hevdet i 1954 at teknologisk endring ikke ville gjøre arbeidskraft overflødig, det ville snarere tvert imot skape etterspørsel etter høykompetent arbeidskraft for å utvikle og jobbe med de nye verktøyene (Autor m.fl., 2003). Felles for begge prediksjonene er store skift i etterspørselen etter arbeidskraft, som igjen gir endringer i *prisen* på arbeidskraft.

Selv om Keynes fortsatt kan få rett, må de fleste være enige om at hittil er det Drucker som har hatt rett. Spådommen fra Drucker gjenspeiles i at høykompetent arbeidskraft over tid har opplevd høy lønnsvekst, relativt til resten av arbeidsstokken (Donovan og Bradley, 2018).⁶ Dette adresseres i produktivitetskommisjonens 2. rapport, hvor det blant annet vises til at Kvaløy (2014) hevder at teknologisk utvikling favoriserer høykompetent arbeidskraft fordi disse arbeiderne har lettere for å ta i bruk den nye teknologien. Teknologi gjør slik de allerede produktive arbeiderne enda mer produktive.

Hvem som henter ut gevinsten av ny teknologi er bare den ene siden av problemstillingen. For å få et mer helhetlig bilde av effekten av teknologisk utvikling på lønnsulikhet er det avgjørende å identifisere hvem som bærer kostnadene. I praksis betyr dette å studere hvilke jobber som forsvinner, og dermed hvilke arbeidere som opplever redusert etterspørsel etter arbeidskraften de tilbyr. Autor (2015) mener å finne at det er arbeidskraften på midten av lønnskalaen som automatiseres bort. Dette forklarer han delvis med at arbeidskraften på bunnen er så billig at den er for dyr å automatisere vekk. På denne måten tvinges de som opprinnelig befant seg på midten av lønnsfordelingen til å ta jobber som lønnes relativt lavt, hevder Autor. Dette omtaler han som lønnpolarisering, og er slik et bidrag til litteraturen som søker å forklare den økende lønnsulikheten.

⁶ Studie basert på amerikanske data.

2.2 Institusjonelle forhold

Her skal jeg kort sammenligne noen sentrale forhold ved det norske og det tyske arbeidsmarkedet. Dette er et nødvendig teoretisk bakteppe for den empiriske analysen, fordi metodologien jeg anvender og utvider i stor grad bygger på arbeidet til Card m.fl. på vest-tyske data, og det er nyttig å kjenne til en del institusjonelle strukturer når resultatene skal tolkes og forstås.

2.2.1 Lønnsdannelse

Et av hovedfunnene til Card m.fl. er at økende bedriftsheterogenitet forklarer en stor del av den økende lønnsulikheten.⁷ Videre finner de empirisk belegg for å hevde at heterogeniteten øker blant bedrifter som har droppet ut av ordningen med kollektive lønnsforhandlinger, mens heterogeniteten blant bedrifter som praktiserer kollektive lønnsforhandlinger er relativt konstant. Artikkelforfatterne argumenterer med bakgrunn i blant annet dette for at den økende lønnsulikheten i Vest-Tyskland kan relateres til institusjonelle endringer i lønnsdannelsen (Card et al., 2013). Kollektive lønnsforhandlinger innebærer at lønn forhandles frem for en større gruppe arbeidere. Denne formen for lønnsdannelse er normen i Norge, noe som representerer et særtrekk ved det nordiske arbeidsmarkedet sammenlignet med andre land (Bjørnstad og Nymo, 2015). Card m.fl. viser til at dekningsgraden av kollektive lønnsforhandlinger i Vest-Tyskland var 63 % i 2007.⁸ Ved å benytte OECD sine tall ser en at den norske dekningsgraden har ligget opp mot 70 % i dataperioden for denne oppgaven (OECD, 2019b). Fra 2010 til 2016 ligger Tyskland under, og Norge over, gjennomsnittlig dekningsgrad i EU (ETUI, 2015).^{9 10 11} Det er verdt å merke seg at Norges dekningsgrad ligger 8-10 %-poeng høyere enn Tyskland.¹² I tillegg er det et sentralt poeng at hovedregelen i Norge er at også ikke-organiserte arbeidere omfattes av resultatet fra kollektive forhandlinger i en bedrift (Bjørnstad og Nymo, 2015).¹³

⁷ I artikkelen defineres bedriftsheterogenitet som bedriftsspesifikke lønnsgevinster som alle ansatte mottar.

⁸ Basert på OECD sin definisjon "Collective bargaining coverage rate corresponds to the ratio of employees covered by collective agreements, divided by all wage earners with right to bargaining" (OECD, 2019).

⁹ Data strekker seg kun til og med 2016.

¹⁰ Gjennomsnittstall fra 2015.

¹¹ Merk: Ved sammenligning av dekningsgrad mellom land er det viktig å huske at det er usikkerhet knyttet til tallene, samt at det er varierende hva som ligger i forhandlingene ulike steder. (ETUI, 2015).

¹² OECD skiller ikke på Vest- og Øst-Tyskland i sine data.

¹³ Individuelle lønnsforhandlinger er vanligere på toppledernivå.

I sin kartlegging av sammenhengen mellom lønnsspredning og typer lønnsdannelse har OECD pekt på at den norske lønnsdannelsen genererer små lønnsforskjeller (OECD 2018). Dette understøttes av at et bredt spekter av studier finner at kollektive lønnsforhandlinger reduserer lønnsspredning.¹⁴ Isolert sett impliserer dette høyere lønnsulikhet i Tyskland sammenlignet med Norge.

2.2.2 Arbeidsmarkedsreformer

Det er ikke bare lønnsdanningsstrukturen som skiller det norske fra det vest-tyske arbeidsmarkedet. Fra midten av 90-tallet og til et stykke ut på 2000-tallet gjennomførte tyske myndigheter en rekke reformer som liberaliserte det tyske arbeidsmarkedet. Disse reformene dreide seg blant annet om åpning for mer utstrakt bruk av midlertidige ansettelser, reduksjon i sykelønnsordningen og en generell senking av ytelsene til arbeidsledige (Card et al., 2013). Dette er tiltak som bidrar til å skille det tyske arbeidsmarkedet fra det norske i tidsperioden som studeres. For eksempel har norske arbeidstakere siden 1978 hatt rett på full lønn ved sykdom i opp til 52 uker (NAV, 2019), for tyske arbeidstakere begrenser ordningen seg til seks uker med full lønnskompensasjon (OECD, 2010). Riktignok ble adgangen til midlertidige ansettelser moderat utvidet i Norge i juli 2015,¹⁵ noe som ga en liten økning i andelen som var midlertidig ansatt i 2016, men denne utviklingen vedvarte ikke i 2017 (SSB, 2017). Det er med andre ord differanser i stillingsvern og sosialt sikkerhetsnett som kan spille en rolle for effekten av bedriftsheterogenitet på arbeidernes lønn.¹⁶

2.3 Oppsummering

Skissert litteratur og teori styrker hypotesen om økende lønnsulikhet også i Norge. Det vil særlig være grunn til å tro at digitalisering og automatisering øker forskjellene, noe som impliserer raskere vekst i ulikhet utover i tidsperioden vi studerer. Adapsjon av ny teknologi varierer mellom ulike sektorer, men i hvilken grad man tar i bruk teknologi vil med stor

¹⁴ Se f.eks. Hibbs og Locking (1996), Kahn (1998), Falch og Strøm (2006).

¹⁵ Fra og med 1. juli 2015 ble det åpnet for generell adgang til midlertidig ansettelse i inntil tolv måneder. Arbeidstaker har krav på fast stilling etter tre år som midlertidig ansatt.

¹⁶ Merk at Card m.fl. ser på bedriftseffekter, ikke på yrkeseffekter.

sannsynlighet også variere mellom bedrifter innen samme sektor. Dette er et argument for at bedriftsspesifikke effekter får økende betydning i lønnssettingen. Dessuten peker litteraturen på feltet i retning av at den teknologiske utviklingen også øker betydningen av yrker, fordi automatisering erstatter spesifikke arbeidsoppgaver som gjør arbeidskraft overflødig i konkrete yrker. Lønnsdannelse og andre forhold ved arbeidsmarkedet peker i retning av lavere vekst i lønnsforskjellene i Norge relativt til både Tyskland og OECD som helhet.

3 Datamaterialet

Jeg skal i denne seksjonen presentere datamaterialet som benyttes i den empiriske analysen i oppgaven. Jeg begynner med å forklare hvilke avgrensninger i data som er foretatt, beskriver deretter variablene som benyttes i den empiriske analysen og presenterer tilhørende deskriptiv statistikk.

3.1 Datautvalg

Lønnsdataene er hentet fra SSB sine registerdata, og er satt sammen til et paneldatasett som rommer tidsperioden 1999-2014. Deler av analysen i oppgaven er en replikering av arbeidet til Card m.fl. fra 2013 på vest-tyske data, hvor jeg anvender deres metodikk på mine norske data. Jeg har valgt å foreta den samme avgrensningen av data med tanke på alder og kjønn som Card m.fl., og studerer derfor heltidsansatte menn i aldersgruppen 20 til 61 år.¹⁷ I 2015 byttet SSB system for registrering av lønn og arbeid, dette representerer et brudd i data som forstyrrer analysen i så stor grad at jeg velger å utelate data fra den nye ordningen.¹⁸

Alle individer er sysselsatt i ulike yrker som i datasettet er listet opp som yrkeskoder. Utgangspunktet er sjusifrede styrk-koder, etter mal fra SSB. Dette er en svært detaljert yrkesinndeling, som jeg senere aggregerer til en firesifret styrk-kode etter modell fra Kostøl (2017). Jeg foretar denne aggregeringen fordi jeg i siste del av analysen ønsker å koble yrkeskodene sammen med ulike mål på teknologisk utvikling, kommer nærmere tilbake til dette i seksjon 5. Yrkeskoder ble innlemmet i SSBs registerdata først i 2003, og analysen av yrkesspesifikke effekter vil derfor kun strekke seg fra 2003-2014.¹⁹ Jeg har fulgt Card m.fl. sin metode for å sortere individer til firma ved å summere lønn tjent av et gitt individ fra hver jobb i ett år, og plasserer individet i det firmaet som har generert høyest inntekt. Det fremstår plausibelt å definere dette som “hovedarbeidssted”, særlig siden de aller fleste kun har en jobb

¹⁷ Card studerte aldersspennet 20-60 år. Jeg har valgt 20-61 fordi man i Norge kan gå av med tidligpensjon fra 62 år, og det er derfor naturlig å avgrense utvalget her.

¹⁸ SSB gikk i 2015 over til å bruke data fra a-ordningen, som er et rapporteringssystem levert av Skatteetaten (SSB, 2014).

¹⁹ Bedriftsidentitet er tilgjengelig tilbake til 1999, og analyse av bedriftsfaste effekter vil derfor ta i bruk data fra hele tidsperioden datasettet rommer.

pr år. Dataperioden deles i to, der første intervall inkluderer årene 1999-2008 og andre intervall tar for seg 2009-2014.

3.2 Avhengig variabel

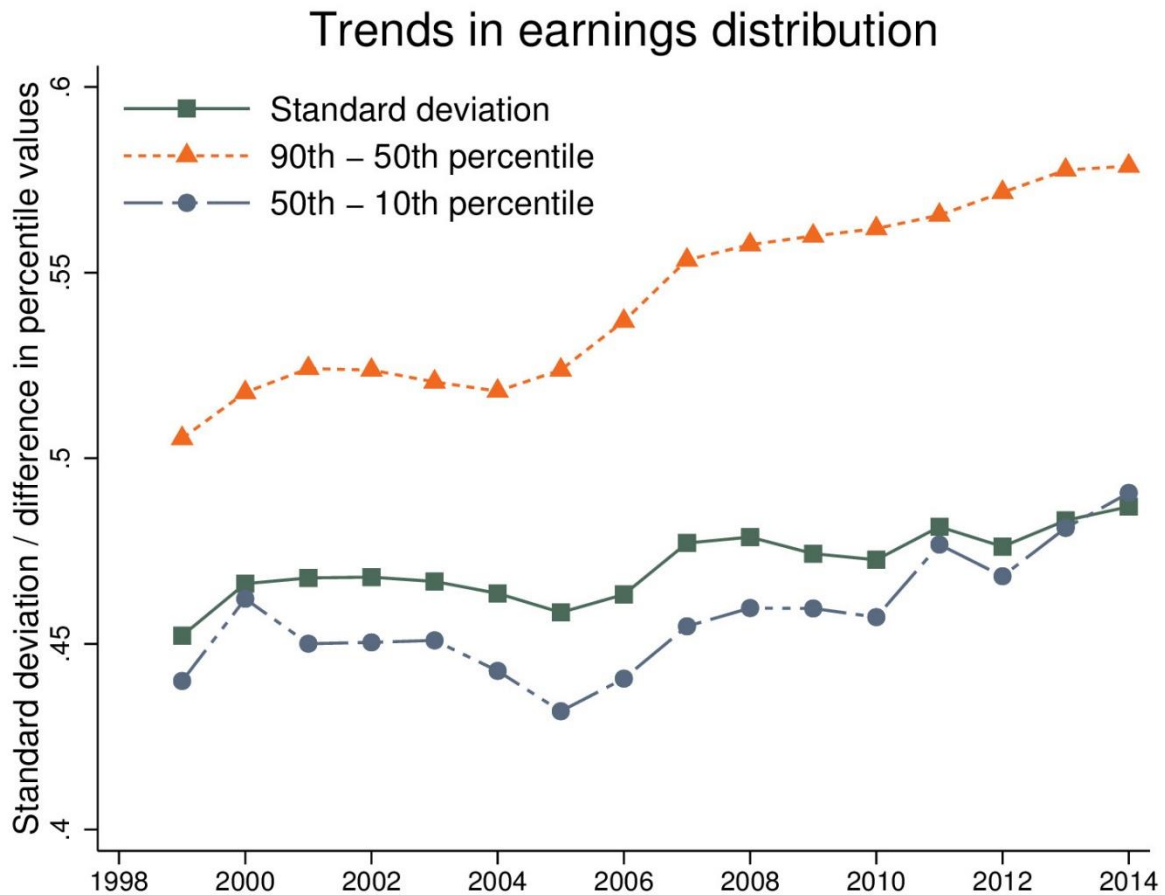
Gjennom hele oppgaven er avhengig variabel inflasjonsjustert kontantlønn. I begrepet kontantlønn inkluderes overtidsgodtgjørelse, feriepenger og utbetalt lønn ved sykdom innenfor arbeidsgiverperioden (16 dager i henhold til dagens sykelønnsordning). Dette skiller seg med andre ord fra den bredere “lønnsinntekt” som i tillegg til kontantlønn inkluderer naturalytelser, sykepenger og dagpenger ved arbeidsledighet (SSB, 2012). Data på kontantlønn er gjennom hele tidsperioden hentet fra SSBs statistikkbank. Lønnsdata rapporteres på månedsnivå, siden tidsenheten i analysen er år er kontantlønsvariabelen et gjennomsnitt for månedene i det gjeldende kalenderåret. I den empiriske analysen benyttes den naturlige logaritmen til lønnsvariabelen. Dette er en vanlig transformasjon når en betrakter lønn, der en av fordelene er at variabelen slik får mindre spennvidde. Dette gjør at ekstremverdiene tillegges mindre vekt i estimeringen (Woolridge, 2015, 172).

Tabell 1: Lønnsstatistikk for fulltidsansatte menn fra hele populasjonen

	Obs (1)	Mean (2)	Std.Dev (3)
<i>Fulltidsansatte menn</i>			
1999	669075	10.75	0.447
2002	671664	10.82	0.463
2006	680482	10.81	0.458
2010	715802	10.82	0.465
2014	756813	10.79	0.478

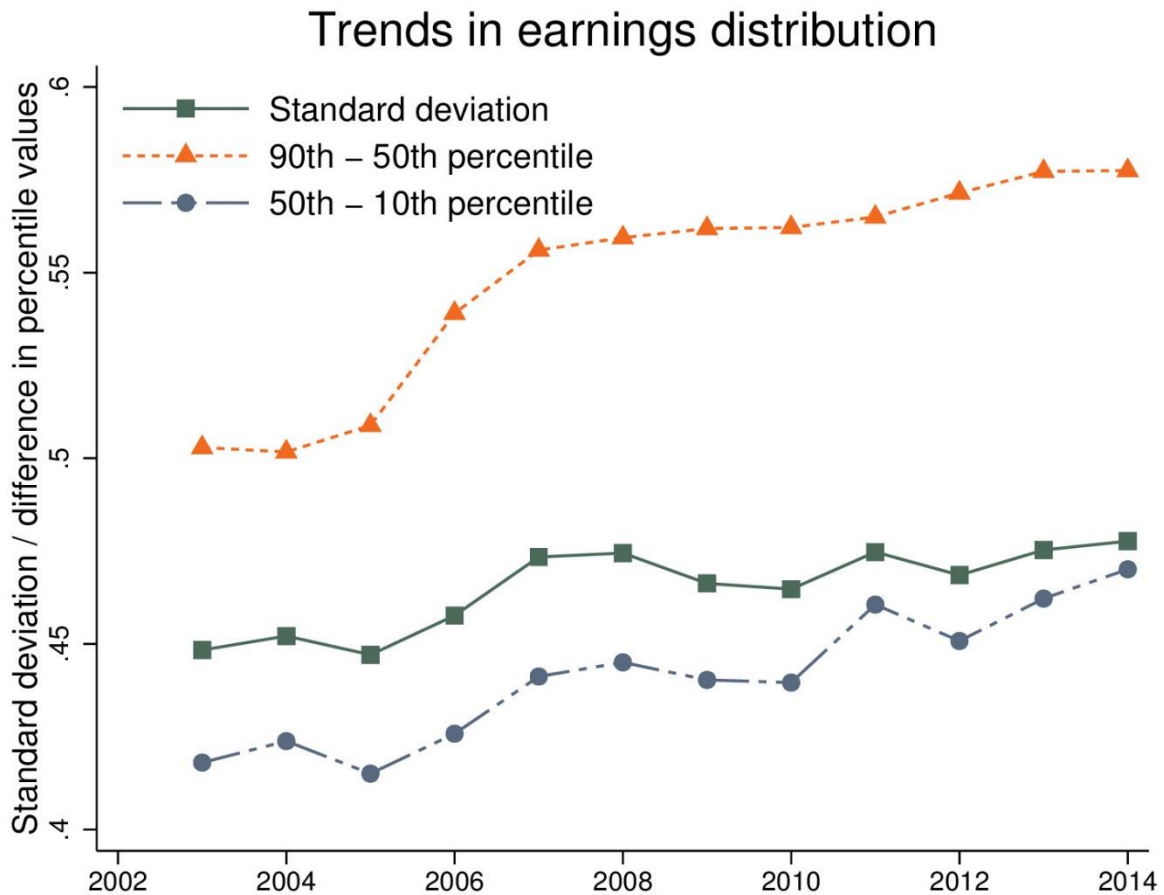
Tabell 1 viser lønnsstatistikk fra fem ulike år for fulltidsansatte menn i utvalget vi studerer. Vi ser i kolonne 1 at antall fulltidsansatte arbeidere øker fra nesten 670 000 til drøyt 750 000 observasjoner. Vi ser videre at logaritmen til den inflasjonsjusterte lønna er relativt stabil, mens standardavviket øker gjennom tidsperioden vi studerer. Dette innebærer at lønnsvariasjonen ser ut til å øke over tid.

Figur 1: Utvikling i lønnsbredning for mannlig populasjon



Figur 1 viser lønnsbredningen for menn i tidsperioden vi studerer. Vi betrakter utviklingen til standardavviket, samt differansen mellom 90-50-persentilene og 50-10-persentilene. Vi ser at lønnsbredningen vokser raskere for de øverste persentilene i lønnsfordelingen, målt ved 90-50-gapet, sammenlignet med de nederste. Veksten er dessuten mer stabil for 90-50 sammenlignet med 50-10. Sistnevnte har flere ganger i løpet av tidsperioden negativ vekst i lønnsbredningen. Merk at alle tre grafer får et veksthopp fra rundt 2005 frem til rundt 2008-2009. Sluttet på denne vekstperioden sammenfaller med finanskrisen.

Figur 2: Utvikling i lønnsbredning i utvalget vi betrakter.



Figur 2 viser lønnsbredningen for utvalget vi skal studere. Jeg har her fjernet foretak med færre enn 30 ansatte og alle individ uten en yrkeskode. Vi betrakter de samme målene som i figur 1, og ser at trendene i de to figurene er nokså like. Vi har her en litt større differanse mellom 50-10-persentilen og standardavvikene, men kurvene beveger seg nokså likt. Videre ser vi at veksten for 90-50-persentilen er enda sterkere i figur 2 sammenlignet med figur 1, noe som impliserer større lønnsbredning.

3.3 Tidsvarierende, uavhengige variabler

Formålet med analysen er å se på hvor mye av lønnsulikhetene som kan forklares av heterogenitet knyttet til individ og yrke. Det er derfor nødvendig å isolere de uobserverbare faktorene fra elementer som er mer åpenbare lønnsdrivere. Av observerbare faktorer velger jeg i analysen å kontrollere for alder, hvor mange barn under 18 år individene har, og hvor mye utdanning hver enkelt har fullført. Jeg bruker her alder som mål på potensiell erfaring, og

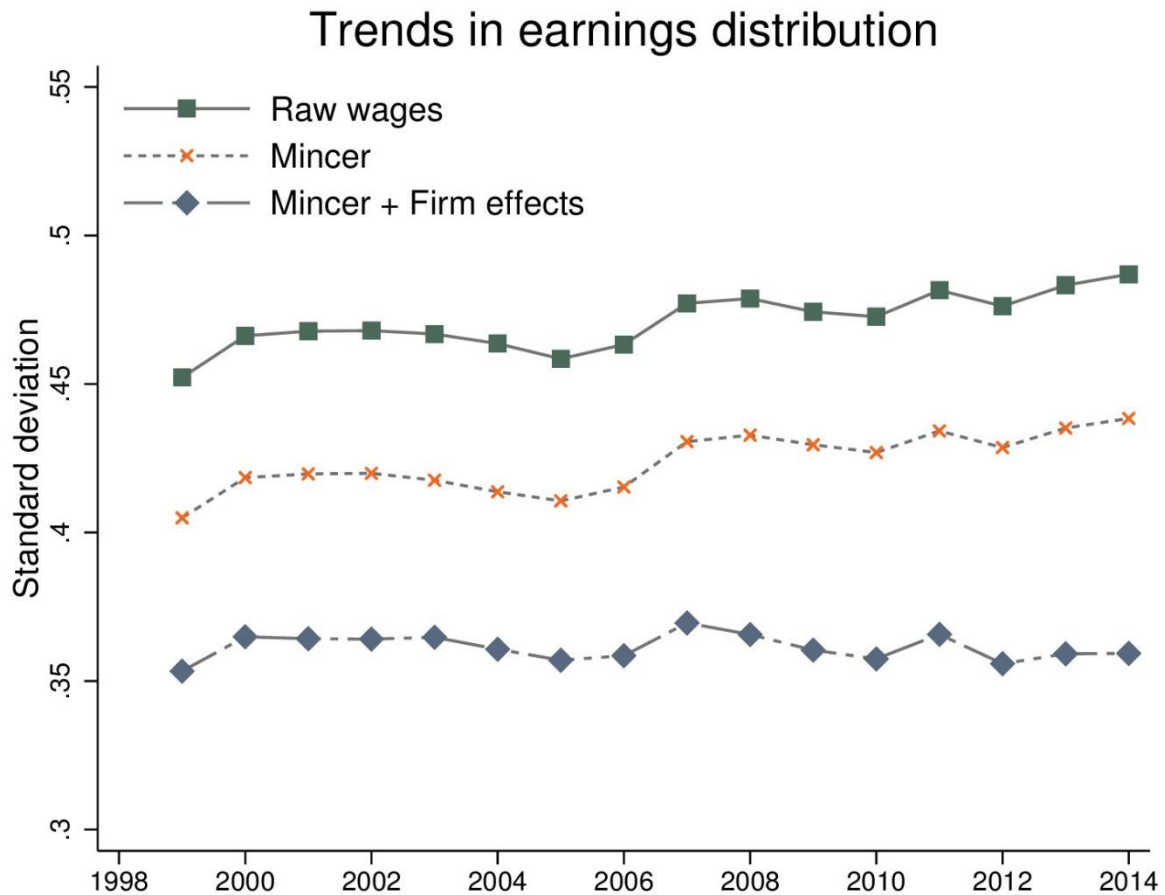
utdanning sier trolig noe om produktiviteten til arbeiderne. Alle disse variablene varierer i både tverrsnitts- og i tidsdimensjonen, altså både mellom individer og over år. Selv om en kan argumentere for at for “godt voksne” vil utdanningsvariabelen være noenlunde konstant.

Tabell 2: Kovariater for hele utvalget

	Mean (1)	Count (2)	Std.Dev (3)
<i>a) 2003</i>			
Alder	40.90	669279	11.09
Barn under 18	0.860	669279	1.079
Antall års utdanning	13.01	669279	2.367
<i>b) 2008</i>			
Alder	40.93	711689	11.22
x_child_bel18	0.852	711689	1.064
Antall års utdanning	13.13	711689	2.410
<i>c) 2014</i>			
Alder	41.00	756813	11.28
Barn under 18	0.821	756813	1.043
Antall års utdanning	13.38	756813	2.433

Tabell 2 gir en beskrivelse av utvalget vi studerer basert på inkluderte kovariater. Jeg viser her sammensetningen for heltidsansatte menn i 2003, 2008 og 2014. Ser av kolonne (1) at gjennomsnittsalderen nesten ikke endres fra 2003 til 2014, og at antall barn under 18 år og utdanningsnivå også endres lite i tidsperioden vi betrakter. Av kolonne (3) kan vi se at alder og utdanningsnivå i utvalget varierer noe mer med tiden, ved at standardavviket for begge variabler øker svakt.

Figur 3: Utvikling i lønnsspredning i utvalget vi betrakter, med kontroll for utdanning og firmafaste effekter.



Figur 3 gir utviklingen i lønnsspredning i utvalget i tidsperioden 1998-2014. Den øverste grafen viser trenden i standardavviket til lønn, uten noen pålagte korrigeringer. Grafen i midten er en Mincer-funksjon, denne kontrollerer for utdanningsnivået til utvalget ved hjelp av dummyvariabler. Ser at dette endrer i svært liten grad utviklingen i lønnsspredning sammenlignet med lønnsmodellen over. Dette tyder på at ulikt utdanningsnivå i liten grad driver økende lønnsulikhet. I den nederste grafen har jeg kontrollert for både utdanning og firmaspesifikke effekter, ser at dette i stor grad fjerner veksten i lønnsulikhet. Dette tyder på at firmaeffektene spiller en viktig rolle for å forklare utviklingen i lønnsspredning i perioden vi studerer. Dette resultatet sammenfaller med funnene til Card m.fl. (2013), og i likhet med dem vil jeg poengtere at dette funnet må behandles med en viss varsomhet. Arbeidere allokeres ikke tilfeldig til ulike bedrifter, og slike firmaspesifikke effekter kan slik drives av såkalt «sortering» av arbeidere. Dette kan eksemplifiseres ved at vi har en økende allokering av høyinntektspersoner til høyinntektsfirma, og slik vil effekten av bedriftsheterogenitet kunne forstyrres av uobserverbare faktorer knyttet til arbeiderne.

4 Metodisk rammeverk: Two-way fixed effects

Seksjonen som helhet tar for seg den økonometriske strategien jeg har valgt, og som slik legger rammene for den empiriske analysen i seksjon 5. Jeg begynner med å presentere den økonometriske modellspesifikasjonen, og tar deretter for meg hvilke forutsetninger som ligger til grunn og potensielle utfordringer knyttet til disse.

Paneldata er en datastruktur som tillater oss å observere den samme enheten over tid. Det vil si at vi har observasjoner som varierer i flere dimensjoner, hvor det klassiske eksempelet er variasjon i både tverrsnitts- og i tidsdimensjon (Woolridge, 2015, 403). I dette tilfellet varierer data over år, samt mellom yrker og individer, med henholdsvis fotskrift t, j og i . Vi står derfor ovenfor en situasjon med variasjon i tre dimensjoner, ofte omtalt som multilevel eller hierarkisk datastruktur (Andrews, Schank og Upward, 2006). I vårt tilfelle observeres individene én gang pr tidsenhet t i yrke j . Individene kan med andre ord bytte yrke over tid (og vi er avhengige av at et tilstrekkelig antall gjør nettopp det for å identifisere effekter). Dette innebærer at enhetene på laveste nivå (individer) ikke er nøstet på det høyere nivået (yrker) (Verbeek, 2017, 71-72).

Sammenlignet med rene tverrsnitts- eller tidsseriedata er det flere fordeler med paneldatastruktur når en skal utføre empiriske analyser. For det første inneholder dataene mer informasjon, og dermed fanger de i større grad opp kompleksiteten i sammenhengene vi ønsker å studere. Dette vil gi analysene et høyere presisjonsnivå, og øker slik sannsynligheten for å trekke korrekte slutninger. Vårt datasett inneholder tre dimensjoner, som synliggjør enda mer informasjon enn et standard paneldatasett med én tverrsnittsdimensjon og tidsdimensjon. I vårt tilfelle kan vi skille effektene av beslutninger tatt på individ- og på yrkesnivå fra hverandre, som er en stor fordel når en studerer arbeidsmarkedet. For det andre kan man med paneldata kontrollere for uobserverbare faktorer som kun varierer i tverrsnittet, altså faktorer som er konstante over tid, men som varierer mellom enhetene. Man kan også kontrollere for faktorer som er felles for enhetene og som kun har variasjon i tidsdimensjonen, i litteraturen ofte eksemplifisert med klassiske makrovariabler (Woolridge, 2015, 435). Dette er kvaliteter ved paneldata som potensielt kan løse noen økonometriske utfordringer, som for eksempel et datasett med rene tverrsnittssdata ikke er i stand til å adressere. Jeg kommer tilbake til dette når jeg skisserer de økonometriske utfordringene.

Datasettet som benyttes gjør det mulig å undersøke variasjonen i lønn mellom individer og yrker over tid. Formålet med analysen er å undersøke årsaker til økende lønnsulikhet i Norge, og mer konkret å studere hvor mye av variansen i lønn som skyldes individeffekter og hvor mye som skyldes yrkeseffekter.

4.1 Modellspesifikasjon

Modellen jeg skal estimere er den samme som Card m.fl. benytter i sin artikkel fra 2013, med den ene forskjellen at jeg studerer yrker i stedet for bedrifter. Denne tilsvarer den originale lønnslikningen benyttet til å utvikle AKM-dekomponeringen.²⁰

Modellen som estimeres gis av følgende likning:

$$(1) y_{it} = \alpha_i + \psi_{j(i,t)} + X'_{it}\beta + r_{it}$$

Den avhengige variabelen y_{it} representerer logaritmen til kontantlønnen til individ i i år t . α_i og $\psi_{j(i,t)}$ er tidskonstante og uobserverbare faktorer som påvirker lønn, og som henholdsvis er spesifikk for individer og for yrker. Å kjøre en enkel MKM-regresjon på (1) vil føre til at α_i og $\psi_{j(i,t)}$ behandles som en del av restleddet, fordi faktorene er uobserverbare. Jeg følger her Card m.fl. og tolker yrkeskomponenten som en proporsjonal lønnsgevinst (kan være negativ) som betales av yrke j til alle ansatte. $X'_{it}\beta$ er en indeks av observerbare karakteristikk som varierer i tid, med tilhørende koeffisientvektor β . Kovariatene i X ble skissert i seksjon 3. Indeksen inneholder også årsummier.²¹ r_{it} er et stokastisk restledd.

Jeg antar at restleddet er sammensatt av tre komponenter:

$$(2) r_{it} = \eta_{ij(i,t)} + \xi_{it} + \varepsilon_{it}$$

²⁰ Basert på «High Wage Workers and High Wage Firm» av John M. Abowd, Francis Kramarz and David N. Margolis fra 1999.

²¹ Dette innebærer at restleddet ikke inneholder en uobserverbar tidskomponent. Antar at denne estimeres direkte ved hjelp av årsummiene. Uten denne forenklingen hadde vi stått ovenfor en three-way fixed effects-modell. Antakelsen er praktisk så lenge tidsdimensjonen i panelet er relativt kort (Andrews, Schank og Upward, 2006).

$\eta_{ij(i,t)}$ er en idiosynkratisk komponent som fanger opp lønnsgevinsten (kan også være negativ) som individ i sysselsatt i yrke j tjener på toppen av det basismodellen gitt av (1) indikerer. Card m.fl. forklarer opphavet til denne restleddskomponenten med at det eksisterer en idiosynkratisk produktivitetskomponent knyttet til enhver potensiell “jobbmatch”, og at den aktuelle arbeideren mottar en andel av denne gevinsten som en del av sin lønn. Jinkins og Morin (2018) på sin side forklarer denne komponenten med at den fanger opp bidraget til lønnen av produktivitetskarakteristika fra «matchen» mellom arbeider i og firma (i mitt tilfelle yrke) $J_{(i,t)}$ som ansetter arbeider i på tidspunkt t .²² ξ_{it} er en enhetsrotkomponent som fanger opp trend i den enkeltes opptjeningsevne, alternativt den enkeltes evne til å generere lønn. Dette kan dreie seg om ulike sjokk som påvirker arbeiderens produktivitet, for eksempel helse relaterte sjokk som gir en individuell lønnstrend. ε_{it} er en komponent som fanger opp forbigående sykliske faktorer som er utelatt fra modellspesifikasjonen. Vi antar at gjennomsnittsverdien for alle komponenter er lik null for alle individer. Det sammensatte restleddet fanger i sum opp alt som ikke modelleres eksplisitt i likning (1). Restleddet inneholder med andre ord både observerbare og uobserverbare variabler, og det er særlig sistnevnte som kan by på utfordringer. Jeg kommer tilbake til dette senere.

I utgangspunktet innebærer gjeldende datastruktur at den foretrukne estimeringsmetoden er Random effects (RE), da dette er den effisiente estimatoren (Woolridge, 2015, 444).²³ Forutsetningene denne estimatoren bygger på er svært restriktive, og å anta at disse holder vil i mange tilfeller være urealistisk. Den sentrale forutsetningen som må holde for å velge RE som estimeringsstrategi er at den enhetsspesifikke komponenten i restleddet må være ukorrelert med alle forklaringsvariable i modellen.²⁴ Dette er en forutsetning man i mange tilfeller ikke ønsker å pålegge. I følge Andrews, Gill, Schank og Upward (2008) er det rimelig å anta at heterogenitetene er korrelert med de observerbare faktorene som varierer i sitt tverrsnitt, altså at α_i og ψ_j er korrelert med variabler som varierer i henholdsvis individ- og yrkesdimensjonen. Jeg vil derfor argumentere for at Fixed effects (FE) er et bedre valg. Denne estimatoren tillater korrelasjon mellom enhetsspesifikke restleddskomponenter og X-ene, men har den svakheten at den ikke kan identifisere effekten av tverrsnittspesifikke variabler (Woolridge, 2015, 436). Dette innebærer at vi ikke kan inkludere tidskonstante

²² Jinkins og Morin baserer seg også på AKM, og undersøker lønnsbidraget fra firmaeffekter på danske data.

²³ Effisient estimator betyr at av alle estimatoren man kan velge mellom har denne lavest varians (Woolridge, 2015, 162).

²⁴ Som nevnt vil standard OLS på (1) føre til at α_i og $\psi_{j(i,t)}$ inngår i restleddet. RE forutsetter at disse komponentene ikke korrelerer med observerbare, eksplisitt modellerte faktorer.

variabler i modellspesifikasjonen. Dette er åpenbart problematisk dersom interessevariabelen i analysen ikke varierer i tid, men en generell svakhet med FE er dessuten at den utnytter mindre av den tilgjengelige informasjonen i data, ved å kun utnytte variasjon i tidsdimensjonen.

Jeg følger AKM sin konklusjon knyttet til at Least square dummy variable-estimatoren (LSDV) med dette datasettet blir for omfattende å benytte, fordi antall individer og yrker gir matriser av en størrelsesorden det i praksis blir umulig å jobbe med. Vi vil få identiske resultater ved å benytte en standard two way fixed effects-modell dersom vi inkluderer yrkesspesifikke dummyvariabler som fanger opp de yrkesspesifikke effektene. I den følgende analysen benyttes altså FE ved å inkludere N-1 yrkesspesifikke dummyvariabler. Disse fanger opp effekten av alle faktorer som varierer mellom yrkene, men som er konstante over tid. Modellen estimeres med minste kvadraters metode (MKM),²⁵ som gitt at underliggende forutsetninger holder er den beste lineære forventningsrette estimatoren (BLUE).^{26 27}

Metoden bygger på en AKM-dekomponering av lønnsvariabelen. Formålet med dette er å identifisere hvor mye av endringen i avhengig variabel som kan forklares av individ- og yrkeseffekter. I praksis blir dette som å ta ut effektene av variablene i indeksen med observerbare karakteristikker som varierer i tid. Slik sitter vi igjen med en residualisert lønnsvariabel som forklares av heterogenitet mellom individer og mellom yrker.

Jeg skriver (1) om til matriseform:²⁸

$$(3) y = D\alpha + F\psi + X\beta + r$$

y uttrykker kolonnevektoren $N^* \times 1$ bestående av lønnsverdier sortert etter individ og år, hvor N^* er antall person-år-observasjoner i datasettet. D defineres som $[d^1, \dots, d^N]$, og har form som en $N \times N^*$ -designmatrise bestående av arbeiderindikatorer, der N uttrykker totalt antall arbeidere. Definerer F som $[f^1, \dots, f^J]$, og har form som en $N^* \times J$ -designmatrise bestående av

²⁵ Omtales ofte som «pooled» OLS

²⁶ Se Woolridge (2015) s. 89 for forutsetninger som ligger til grunn for BLUE.

²⁷ Ved tilstrekkelig store utvalg (vanlig tommelfingerregel sier minimum 30) kan man lene seg på sentralgrenseteoremet. Dette er tilfellet i denne analysen, og dermed vil restleddet være tilnærmet normalfordelt uavhengig av hvilken fordeling lønnsvariabelen føyer best (Woolridge, 2015, side 684).

²⁸ Følger samme oppsett som Card m.fl. 2013.

yrkesindikatorer, der J uttrykker totalt antall yrker. Definerer X som $[x^1, \dots, x^K]$, og har form som en $N^* \times K$ -matrise bestående av tidsvarierende kovariater. Til sist er r en $N^* \times 1$ -vektor av sammensatte restledd.

Likning (3) kan skrives om til:

$$(4) y = Z'\xi + r$$

Definerer Z som $[D, F, X]$ og ξ som $[\alpha', \psi', \beta']$. Estimeringen av likning (3) ved hjelp av MKM gir estimater som løser standard normallikningene:²⁹

$$(5) Z'Z\xi = Z'y$$

En unik løsning krever at alle rader i $Z'Z$ -matrisen er lineært uavhengige slik at matrisen lar seg invertere. I likhet med Card m.fl. (2013) forenkler jeg estimeringen til det største sammenkoblede datasettet av yrker som er knyttet til arbeidsmobilitet. Denne forenklingen er nødvendig for å kunne sammenligne yrkeseffekter på tvers av ulike yrker. Grovt forenklet kan vi se for oss at arbeidsplasser er ulike av to årsaker; en skyldes at noen firma benytter bedre teknologi enn de andre og slik er mer produktive, og den andre forskjellen er knyttet til kvalitet på arbeiderne som jobber der. For å utføre analysen av yrkesheterogenitet må vi derfor kontrollere for kvaliteten på arbeiderne. Vi gjør dette ved å gruppere arbeiderne slik at alle som har jobbet i samme yrke i løpet av tidsperioden settes sammen, og vi lager slik et nettverk bestående av alle arbeiderne som har jobbet i et yrke og alle yrkene de aktuelle arbeiderne har jobbet i i løpet av tidsperioden. To separate nettverk innebærer dermed at ingen individer i det ene nettverket har jobbet i et yrke i det andre nettverket, og motsatt. Siden jeg har delt tidsperioden opp i to intervaller er denne prosessen kjørt i to ulike omganger, en for hvert intervall. Jeg velger så det nettverket som inkluderer flest arbeidere og yrker for å utføre analysen. Dette er årsaken til at vi er avhengige av at et tilstrekkelig antall arbeidere bytter jobb i løpet av tidsperioden vi studerer, og vi skal i neste seksjon se at denne strategien innebærer å utnytte en veldig høy andel av observasjonene i det originale datasettet. Jeg normaliserer her yrkeseffektene ved å utelate den siste yrkesdummien.

²⁹ Se Verbeek (2017) side 8.

4.2 Økonometriske utfordringer

For å oppnå troverdige resultat med den valgte estimeringsstrategien må vi knytte en del forutsetninger til det sammensatte restleddet.³⁰ Formålet med disse forutsetningene er å sørge for at vi unngår endogenitet, samt sikre at det vil være mulig å trekke inferens. For at estimeringen skal gi oss de sanne effektene er vi avhengig av at de inkluderte forklaringsvariablene ikke er korrelert med elementer i restleddet. Dersom dette ikke er tilfellet vil vi få et endogenitetsproblem. Endogenitetsproblemer gir opphav til forventningsskjevne og inkonsistente estimater for effekten av forklaringsvariablene på avhengig variabel, lønn (Wooldridge, 2015, 76-77). Resultatet vil enten bli systematisk over- eller underestimert av de ulike effektene på lønn, avhengig av om korrelasjonen med restleddet er positiv eller negativ. Fortegnet på skjevheten kan være vanskelig å avgjøre da flere enn én forklaringsvariabel kan ha korrelasjon med restleddet, og fortegnet på disse korrelasjonene trenger ikke være like. Å kunne trekke inferens fra de estimerte koeffisientene er avgjørende for å kunne si noe om populasjonen som helhet (Verbeek, 2017, 98-99). Målet med en empirisk analyse er å bruke egenskaper ved utvalget studien baserer seg på til å si noe om egenskaper ved hele populasjonen. I denne delseksjonen skisserer jeg kort noen scenarioer som kan gjøre statistisk inferens ugyldig, og slik risikere å gi oss villedende resultater. Jeg skal i det følgende diskutere mulige brudd på forutsetningene, og hvilke konsekvenser slike brudd vil få for analysen.

4.2.1 Mobilitet

I den etter hvert relativt omfattende litteraturen som bygger på AKM omtales ofte mobilitet som den største endogenitetsrisikoen. Jinkins og Morin (2018) formulerer dette som at betinget på de faste effektene kan ikke arbeiderne sorteres til bedrifter hvor de mottar ekstraordinært høy lønn. I praksis innebærer dette at forutsetningen for eksogenitet hviler på om arbeidernes allokering til ulike yrker avhenger av uobserverbare elementer i restleddet. Risikoen for dette er lettere å analysere hvis vi konsentrerer oss om de tre restleddskomponentene separat. Vi kan ha en underliggende sortering av arbeidere til yrker basert på den enkelte arbeiders lønnsgevinst, η_{ij} . Dette innebærer at vår tolkning av

³⁰ Se forutsetninger nærmere forklart i Woolridge (2015, side 74-82 og 106).

estimatorene knyttet til de yrkesspesifikke elementene ikke blir korrekte, fordi at i ethvert yrke vil ulike arbeidere ha ulik lønnsgevinst. Hvis det er tilfellet vil endogenitet oppstå og estimatorene våre blir forventningsskjevne og inkonsistente. Dette er en kilde til skjevhet som kan la seg undersøke. Hvis to arbeidere med lik formell kompetanse og arbeidserfaring bytter fra hvert sitt yrke til den andres yrke, alt annet konstant, skal differansen mellom arbeidernes lønnsgevinst og -tap assosiert med jobbskifte si oss noe om den potensielle skjevheten. Dersom en sortering basert på individuell lønnsgevinst knyttet til “jobbmatch” driver arbeidernes mobilitet mellom yrker skulle vi i det skisserte eksempelet se at begge individer opplever lønnsgevinst. Dersom forutsetningene våre holder, og det ikke eksisterer en slik sortering, skal den enes lønnsgevinst med et slikt jobbskifte nulles ut av den andres lønnstap. Card m.fl. undersøkte denne mekanismen i sin studie av mobilitet mellom bedrifter, og finner at det ikke eksisterer noen slik “mobilitetsgevinst” for arbeidere som flytter mellom bedrifter. Jeg vil i det videre gå ut i fra at det samme vil være tilfellet mellom yrker.

Dersom trend i forventet lønn en arbeider vil tjene i ethvert yrke styrer jobbskifter, altså korrelasjon mellom yrkesindikatorene og ξ_{it} , vil vi også få endogenitet. Hvis en arbeider får varig nedsatt helsetilstand, og slik blir mindre produktiv, vil vedkommende trolig oppleve redusert lønn (eventuelt redusert lønnsvekst). Dette vil igjen øke sannsynligheten for at arbeideren ved et jobbskifte bytter til et yrke med relativt sett lavere lønnskompensasjon. Dette er et eksempel på at mobiliteten mellom yrker påvirkes av faktorer i restleddet, fordi denne lønnstrenden ikke er eksplisitt modellert. Den siste restleddskomponenten kan by på endogenitetsproblemer dersom vi har korrelasjon mellom arbeidernes jobbskifter og fluktusjon i elementer som påvirker ulike yrkers lønnsnivå. For eksempel at konjunkturer i enkelte bransjer påvirker lønnsnivået til spesifikke yrker, som igjen spiller inn på sannsynligheten for bytte av jobb. Card m.fl. viser til at de ikke finner systematiske lønnstrender knyttet til jobbskifter, og jeg vurderer det som lite sannsynlig at dette ikke også skal gjelde for mine data. Analysen i seksjon 5 vil derfor være basert på antakelsen om at jobbmobilitet ikke forstyrrer de estimerte effektene.

4.2.2 Utelatt variabel

En sentral kilde til endogenitet er utelatt variabel-skjevhet, også omtalt som underspesifisering av modellen. Dette problemet oppstår dersom variabler som påvirker avhengig variabel *og* som er korrelerte med én eller flere forklaringsvariabler er utelatt fra modellspesifikasjonen. Dersom vi har utelatt en variabel som påvirker lønn, men som ikke har korrelasjon med noen av forklaringsvariablene får vi ikke et endogenitetsproblem. Resultatet vil bare bli mindre presise estimat. Årsakene til at relevante variabler kan bli utelatt er ofte enten at man mangler data (data er ikke tilgjengelig for de som utfører analysen, eller det dreier seg om faktorer som er vanskelig å kvantifisere), eller det dreier seg om uobserverbare faktorer. Vi kan også se for oss at vi i modellen inkluderer irrelevante variabler, altså tar med forklaringsvariabler som ikke har noen betydning for den avhengige variabelen. Dette vil ikke representere brudd på noen av forutsetningene vi opererer med, og vil dermed ikke skade analysen. Kostnaden ved å inkludere forklaringsvariabler som ikke påvirker lønn, gjerne omtalt som overspesifisering av modellen, vil kun være høyere varians (Wooldridge, 2015, 77-78).

Det klassiske eksempelet på utelatt variabel-skjevhet i lønnsestimering er evner, som er en uobserverbar faktor. Det er naturlig å anta at hvilken utdanning du velger vil ha en viss sammenheng med hvilke evner du har, evner vil være noe som mer eller mindre er konstant over tid, men som varierer på tvers av individer. Videre kan vi se for oss at disse evnene gjør at visse individer er predisponert for høy inntekt, i den forstand at de relativt sett er mer produktive og slik vil tjene bedre enn de fleste andre. Høyinntektsindivider vil trolig være overrepresentert på studier som i snitt genererer høy inntekt og som har høye karakterkrav for inntak, for eksempel medisin. Slik vil evner være med å påvirke hvilken utdanning individer velger, og hva de vil komme til å tjene. Dette er et eksempel på at vi har elementer i restleddet som både er korrelert med lønn og med den eksplisitt modellerte forklaringsvariabelen utdanning. Dette vil overvurdere effekten av yrke på lønn, og dermed gi skjeve estimat.

I vårt tilfelle vil det være svært sannsynlig at restleddet inneholder uobserverbare elementer som påvirker arbeidernes lønn. Dette innebærer at en standard MKM-regresjon på modellen ville gitt forventningsskjeve og inkonsistente estimat. Fordelen med paneldata er, som nevnt tidligere, at datastrukturen gjør det mulig å potensielt løse denne utfordringen. Vi kan

kontrollere for den uobserverbare heterogeniteten ved å inkludere enhetsspesifikke dummyvariabler. Disse vil fange opp alle tidskonstante faktorer i restleddet som varierer mellom enhetene. Dette er ekvivalent med å foreta en “within group”-transformasjon av modellen, som innebærer å transformere vekk alle tidskonstante variabler (Woolridge, 2015, 438). Dummyvariablene vil altså kontrollere for alle faktorer som varierer mellom de ulike yrkene og individene, men som ikke varierer over tid. Som forklart i 4.1 er dette vår foretrukne estimeringsstrategi i denne oppgaven. Dermed blir utelatte variabler kun et problem dersom vi har korrelasjon mellom forklaringsvariablene og uobserverbare elementer som varierer mellom enhetene og over tid.

4.2.3 Seleksjon

Empiriske analyser bygger på å teste visse egenskaper ved utvalg av varierende størrelse, fordi det i de fleste sammenhenger ikke vil være mulig å benytte seg av data fra hele populasjonen. Det kreves at utvalget man studerer er tilfeldig trukket og sammensatt for at slutningene man treffer skal ha noen overføringsverdi til populasjonen som helhet. Dersom dette ikke er tilfellet vil vi få seleksjonsskjevhet (Verbeek, 2017, 265). En kilde til seleksjonsskjevhet er at utvalget man har data for består av en viss type individer, altså er de en selektert gruppe. Hvis man for eksempel driver analyse av partivalg og har et utvalg kun bestående av menn i en viss aldersgruppe med høyere utdanning vil konklusjonen ha liten ekstensiv validitet, da man ikke har et utvalg som er representativ for befolkningen som helhet. På mer generelt grunnlag kan man si at dersom sannsynligheten for at et gitt individ er inkludert i utvalget er avhengig av fenomenet som studeres vil man få seleksjonsskjevhet (Verbeek, 2017, 266). I mitt tilfelle har jeg begrenset utvalget til å bestå av menn i fulltidsstilling. I utgangspunktet gir dette en selektert gruppe fordi det for eksempel vil ekskludere de individene i samme aldersgruppe som av ulike årsaker ikke jobber 100 %. Jeg vil likevel argumentere for at når utfallsvariabelen min er lønn for fulltidsansatte vil ikke denne utvalgsavgrensningen gi seleksjonsskjevhet som forstyrrer resultatene.

4.2.4 Simultanitet

Dersom en eller flere uavhengige variabler bestemmes samtidig som den avhengige variabelen, også omtalt som toveis kausalitet, gir dette opphav til simultanitetsskjevheter som resulterer i endogene forklaringsvariabler (Woolridge, 2015, 503-504). Endogeniteten oppstår fordi den eller de forklaringsvariablene som bestemmes simultant med avhengig variabel vil være korrelert med restleddet, og vi har dermed brudd på forutsetningene som sikrer eksogene forklaringsvariable. I denne analysen vil jeg se det som relativt usannsynlig at noen forklaringsvariable bestemmes simultant med lønn, og vil derfor se vekk i fra risikoen for simultanitetsskjevheter.

4.2.5 Målefeil

Hvis vi har avvik mellom observert og faktisk verdi på en variabel har vi det som omtales som målefeil (Wooldridge, 2015, 287-292). Konsekvensene av målefeil kan være ulike, og av ulik alvorlighetsgrad, avhengig av hvilke variabler det er snakk om, og omfanget på målefeilen. Datagrunnlaget i denne analysen kan potensielt ha målefeil i flere variabler.

Dersom vi har målefeil i avhengig variabel innebærer det altså at verdien på lønnsvariabelen ikke er korrekt. Dette kan forekomme dersom arbeidsgiver rapporterer inn galt, eller dersom SSB gjør feil i sin bearbeidelse av data. For at dette skal by på alvorlige problemer for analysen må de være systematiske, dermed vil en systematisk gal innrapportering eller bearbeidelse av data øke risikoen for skjeve estimat. Dersom det er store avvik mellom observert og faktisk verdi på lønnsvariabelen, uten at det foreligger noe systematisk i feilen, vil konsekvensen bli høy restleddsvarians og slik mindre presis estimering.

De observerbare forklaringsvariablene er i sin helhet hentet fra SSB sin statistikkbank, og jeg anser det som relativt lite sannsynlig at det eksisterer noen systematisk eller særlig omfattende feil i registerdataene knyttet til inkluderte kovariater. Generelt kan utfordringer knyttet til målefeil øke dersom vi har målefeil i en forklaringsvariabel som varierer lite over tid, for eksempel utdanningsnivå i vår modell, fordi vi fjerner mye av variasjonen med å benytte FE og målefeilen blir slik mer dominerende.

Målefeil i de estimerte yrkes- og individeffektene er en problemstilling som adresseres i Card m.fl. sin artikkel fra 2013, og i flere andre studier i AKM-litteraturen. Slike målefeil vil ikke bare kunne gi forventingsskjevne variansestimater, vi risikerer også systematisk underestimering av kovariansen mellom yrkes- og individeffektene. Andrews m.fl. (2008) viser at denne skjevheten øker i lavere jobbmobilitet, dvs. problemet blir større hvis få individer bytter jobb i løpet av tidsperioden vi studerer. De viser også at skjevheten reduseres i økt T, altså når tidsperioden man studerer er relativt lang. I likhet med Card m.fl. vil jeg ikke forsøke å konstruere korrigerede estimater for variansen og kovariansen til effektene, men jeg baserer tolkningen av estimatene i seksjon 5 på at skjevhetene er like i tidlig og sent tidsintervall.

Generelt vil målefeil alltid være en risiko, og det kan være vanskelig å korrigere for. Det er derfor viktig å være oppmerksom på muligheten for at det kan forstyrre resultatene.

4.2.6 Seriekorrelasjon og heteroskedastisitet

Konstant restleddsvarians, eller homoskedastisitet, er en av forutsetningene som må på plass for at MKM skal være BLUE. Fravær av homoskedastisitet kalles heteroskedastisitet, og innebærer at variansen til restleddet ikke er konstant (Wooldridge, 2015, 395). Vi vil for eksempel ha heteroskedastisitet dersom variansen til utelatte faktorer som påvirker lønn varierer mellom de ulike yrkeskodene. Dette fører ikke til forventingsskjevne estimater, men gjør at statistisk inferens blir ugyldig.

Seriekorrelasjon oppstår dersom vi har tidskonstante forskjeller mellom individer og yrker, ψ_j . Dette skyldes at komponenter som ikke varierer over tid fører til at restleddet blir korrelert med seg selv, og slik ikke får en standard struktur. Dette problemet forsvinner ved å benytte FE, fordi tidskonstante effekter transformeres bort. Men vi kan også ha seriekorrelasjon i idiosynkratiske restleddskomponenter. Både seriekorrelasjon og heteroskedastisitet kan løses med å benytte cluster-robuste standardfeil, dette sikrer at vi kan trekke inferens.

5 Analyse

Jeg skal i denne seksjonen presentere resultatene i den empiriske analysen. Jeg har brukt Stata og Matlab til å gjennomføre analysene, og det er i alle regresjoner benyttet cluster-robuste standardfeil på individnivå. Jeg begynner med å presentere resultatene fra en estimering med bedriftsfaste effekter, altså en replikering av det Card m.fl. gjorde på vest-tyske data. Deretter går jeg igjennom resultatet av hovedanalysen min, som dreier seg om effekten av yrkesfaste effekter. Jeg har estimert modell (1) på to delutvalg av datasettet, hvor jeg først har sett på perioden 1999-2008 og deretter 2009-2014. Som forklart tidligere inngår yrkeskoder i SSB sitt datamateriale fra 2003, og analysen av yrkesfaste effekter for det første tidsintervallet begrenser seg derfor til 2003-2008. Jeg har benyttet lineære modeller med additive individ- og yrkeseffekter som skissert i seksjon 4. Vi står overfor et ubalansert panel i alle regresjoner, fordi både antall individer og antall firma varierer gjennom tidsintervallene.

5.1 Resultater

5.1.1 Bedriftsfaste effekter

Tabell 3: Deskriptiv statistikk, bedriftsfaste effekter

	(1) 1998-2008	(2) 2009-2014
<i>a) Originalt datasett</i>		
P-y-obs	6758546	4421456
Workers	1118049	1036334
Firms	12683	11209
Mean wage	10.701	10.699
Var wage	0.173	0.180
<i>b) Største sammenkoblede datasett</i>		
P-y-obs	6757817	4419773
Movers	546337	348099
Firms	12614	11088
Mean wage	10.7009	10.6994
Var wage	0.173	0.180

Jeg begynner med å betrakte bedriftsfaste effekter. Panel a i tabell 3 viser statistikk for det originale datasettet for begge tidsintervaller. Tabellen gir observasjoner for antall individ-år-observasjoner for fulltidsansatte mannlige arbeidere, antall unike individer, samt gjennomsnittsverdi og varians for log-lønn. Kolonne (1) viser tidlig tidsperiode og kolonne (2) viser sen tidsperiode. Vi ser at antall individ-år-observasjoner reduseres fra drøyt 6,7 millioner i det tidlige intervallet til nesten 4,5 millioner i det andre. Antall unike arbeidere faller med cirka 80 000 fra tidlig til sen periode, og bedriftsobservasjonene faller fra drøyt 12 500 i første tidsintervall til drøyt 11 000 i andre tidsintervall. Vi ser at gjennomsnittlig log-lønn er stabil gjennom tidsperiodene, men merk at variansen øker. Varians er et spredningsmål som sier noe om variasjonen innad i en fordeling. Varians er det kvadrerte standardavviket som gir gjennomsnittlig avstand fra gjennomsnittsverdien, og vil i denne sammenhengen slik være et mål på lønnsspredning (Woolridge, 2015, 45).

Som forklart i seksjon 4 baserer analysen seg på det største sammenkoblede datasettet. Verdiene for dette datasettet er gitt i panel b i tabell 3. Vi ser at i begge tidsintervall inneholder datasettet vi baserer analysen på rundt 99 % av firmaene fra det originale datasettet, og for individ-år-observasjoner er tallet i begge tidsintervall 99,9 % av originalt datasett. Vi ser at gjennomsnittlig log-lønn og varians i det største sammenkoblede datasettet er nesten identisk som i det originale datasettet. Dette styrker antakelsen om at en analyse basert på denne sammenkoblede gruppen av observasjoner vil gi ekstern validitet.

Fundert i distinksjoner knyttet til institusjonelle forhold pekte teorien i seksjon 2 i retning av lavere lønnsulikhet i Norge, sammenlignet med Tyskland. Verdien på lønnsvariansen i tabell 3 gir støtte til denne prediksjonen, for selv om tallene viser økende lønnsulikhet også i Norge så er både nivået og veksten på variasjonen lavere enn det Card m.fl. fant.

Tabell 4: Estimeringsresultat for AKM-estimering basert på bedriftsfaste effekter

	(1) 1998-2008	(2) 2009-2014
<i>a) Hovedresultat</i>		
Var firma	0.021	0.023
Var indv	0.088	0.112
Cov	0.005	0.004
Corr	0.120	0.084
R ²	0.69	0.79
Adj R ²	0.63	0.73
<i>b) Lønn mot individ</i>		
R ²	0.58	0.67

Tabell 4 gir resultatene fra estimeringen for begge tidsintervall, henholdsvis tidlig i kolonne (1) og sent i kolonne (2). I panel a har jeg valgt å rapportere variansen til de estimerte bedrifts- og individeffektene, samt kovariansen og korrelasjonen mellom disse effektene. Jeg rapporterer i tillegg R² og justert R².³¹ Vi ser at variansen til bedriftseffektene faller fra tidlig til sen periode, men individeffektene øker. Dermed har vi at variasjonen til individeffektene øker over tid, mens variasjonen til bedriftseffektene reduseres. Korrelasjonen mellom effektene faller fra 0,120 til 0,084. Vi ser at modellens forklaringskraft øker fra 0,63 til 0,73 fra tidlig til sent tidsintervall, noe som tyder på at bedrifts- og individeffektene over tid har blitt viktigere forklaringsfaktorer bak lønnsutviklingen. Disse resultatene harmonerer med det Card m.fl. fant for vest-tyske data. Selv om verdiene på de ulike estimatene varierer er trenden den samme, med økende lønnsvarians og forklaringskraft. Men merk at korrelasjonen mellom effektene øker hos Card, i motsetning til i vårt tilfelle der den faller.

Panel b gir resultatet for R² fra en regresjon av lønnsvariabelen utelukkende mot individeffektene.³² Vi ser slik at forklaringskraften i all hovedsak kommer fra disse, og at også økningen mellom tidsintervallene drives av individer, ikke av firma. Dette resultatet strider med funnene til Card m.fl., men kan ha sammenheng med at variasjonen til firmaeffektene faller i mine data. Det strider også med den illustrerte lønnstrenden i figur 3 hvor bedriftseffekter så ut til å spille en rolle for lønnsulikheten.

³¹ R² gir modellens forklaringskraft justert for antall frihetsgrader. Standard R² vil «alltid» øke når modellen utvides, selv om de nye forklaringsvariablene ikke egentlig har noen forklaringskraft for avhengig variabel (Verbeek, 2017, 22).

³² Verdien på R² i panel b er beregnet både manuelt og i Stata med identisk resultat.

Det er verdt å knytte noen ytterligere kommentarer til disse funnene. Som nevnt i forbindelse med figur 3 i seksjon 3 er det en viss risiko for at effekten av firmafaste effekter drives av sortering av arbeidere. Implikasjonen av dette er at den firmaspesifikke lønnsgevinsten i liten grad forklarer trenden i lønnsulikhet, og at bytte av arbeidssted dermed ikke vil føre til systematiske lønnsendringer for arbeiderne. Card m.fl. finner imidlertid empirisk belegg for at dette jevnt over ikke er tilfellet, da en oppdeling i lønnspercentiler faktisk avdekker en systematikk i lønnsendringene for en arbeider som bytter fra en bedrift med høyt lønnsnivå til en bedrift med lavere, og motsatt. Jeg har ikke undersøkt denne sammenhengen på mine data, men vil argumentere for at Card m.fl. sine funn på generell basis gir kredibilitet til betydningen av firmafaste effekter. I tillegg vil det i dette tilfellet være mindre relevant all den tid de skisserte estimeringsresultatene tyder på at firmaspesifikke effekter ikke er av stor betydning for den økende lønnsulikheten.

5.1.2 Yrkesfaste effekter

Tabell 5: Deskriptiv statistikk, yrkesfaste effekter. 4-sifret styrk.

	(1) 2003-2008	(2) 2009-2014
<i>a) Originalt datasett</i>		
P-y-obs	3045489	3860237
Workers	819012	938577
Jobs	258	258
Mean wage	10.7132	10.6978
Var wage	0.185	0.189
<i>b) Største sammenkoblede datasett</i>		
P-y-obs	3045476	3860209
Movers	228466	255916
Jobs	256	253
Mean wage	10.7132	10.6978
Var wage	0.185	0.189

Jeg går nå over til å studere effektene av yrkesfaste effekter. Panel a og b i tabell 5 viser deskriptiv statistikk for henholdsvis det originale datasettet og for det største sammenkoblede datasettet. Vi ser data for tidlig tidsintervall i kolonne (1), og tilsvarende for sent intervall i

kolonne (2). Merk at vi her betrakter 2003-2008 i det første intervallet. I likhet med i analysen av bedriftseffekter gir tabell 5 observasjoner for antall individ-år-observasjoner for fulltidsansatte arbeidere, antall unike individer, samt gjennomsnittsverdi og varians for log-lønn. Vi ser at vi her får en økning i antall individ-år-observasjoner fra tidlig til sen periode. Dette er et naturlig resultat, med tanke på at vi nå betrakter to tidsintervaller med likt antall år, i tillegg til at vi vet fra dataseksjonen at antall individer i utvalget har økt over tid. Vi ser videre at antall unike yrkeskoder er 258 i begge tidsintervall. Gjennomsnittsverdien på log-lønn er redusert marginalt, mens variansen har økt med cirka 4 % fra tidlig til sen periode. Vi ser av panel b at i begge tidsintervall utgjør antall individ-år-observasjoner 99,9 % av tilsvarende observasjoner i det originale datasettet. For antall yrkeskoder er 99,2 % av originalt datasett med i tidlig tidsintervall, mens 98,1 % er med i sent tidsintervall. Konklusjonen blir dermed også her en styrket antakelse om ekstern validitet ved å estimere basert på det største sammenkoblede datasettet.

Tabell 6: Estimeringsresultat for AKM-estimering basert på yrkesfaste effekter. 4-sifret styrk.

	(1) 2003-2008	(2) 2009-2014
<i>a) Hovedresultat</i>		
Var job	0.0034	0.0031
Var indv	0.1292	0.1340
Cov	0.0046	0.0058
Corr	0.2189	0.2802
R ²	0.77	0.79
Adj R ²	0.68	0.72
<i>b) Lønn mot individ</i>		
R ²	0.75	0.77

I tabell 6 finner vi resultatene for begge tidsintervall fra estimeringen på det største sammenkoblede datasettet. I panel a rapporterer jeg de samme estimatene som for analysen av bedriftseffekter, ved å vise estimert verdi på variansen til begge effekter, kovariansen og korrelasjonen mellom dem, i tillegg til R² og justert R². Kolonne (1) gir verdiene for tidlig tidsintervall og kolonne (2) gir tilsvarende verdier for sent tidsintervall. Vi ser at variansen til yrkeseffektene faller fra tidlig til sen periode, men merk at reduksjonen er lav og at den estimerte variansverdien i begge tidsperioder er liten. Estimert variasjon i individeffekter øker

mellom tidsintervallene, noe også korrelasjonen mellom effektene gjør. Vi ser at både R^2 og justert R^2 øker fra tidlig til sen tidsperiode, slik at estimeringen tilsier at forklaringskraften øker mellom periodene, noe vi også så for bedriftseffekter. Men merk at økningen fra tidlig til sen periode er mindre i modellen for yrkeseffekter sammenlignet med modellen for bedriftseffekter. Vi ser at forskjellen primært drives av at forklaringskraften i tidlig tidsintervall er vesentlig større for estimeringen med yrkeseffekter. Merk likevel at vi bør være forsiktige med å sammenligne resultatene fra tidlig tidsintervall mellom disse to modellene, da de ikke måler helt samme tidsperiode.

I likhet med i tabell 4 så gir panel b resultatet for R^2 fra en regresjon av lønnsvariabelen utelukkende mot individeffektene.³³ Vi ser slik at forklaringskraften også her i all hovedsak kommer fra de individfaste effektene, og at dette også gjelder for økningen i R^2 mellom tidsintervallene. Dette tyder på at yrkesheterogenitet ikke kan forklare den økende lønnsulikheten vi observerer.

Tabell 7: Deskriptiv statistikk, yrkesfaste effekter. 7-sifret styrk.

	(1) 2003-2008	(2) 2009-2014
<i>a) Originalt datasett</i>		
P-y-obs	4080777	4421431
Workers	970486	1036331
Jobs	8923	6982
Mean wage	10.699	10.6993
Var wage	0.174	0.180
<i>b) Største sammenkoblede datasett</i>		
P-y-obs	4079385	4418880
Movers	483929	397915
Jobs	8565	6561
Mean wage	10.699	10.6993
Var wage	0.174	0.180

³³ Verdien på R^2 i panel b er beregnet både manuelt og i Stata med identisk resultat.

Tabell 8: Estimeringsresultat for AKM-estimering basert på yrkesfaste effekter. 7-sifret styrk.

	(1) 2003-2008	(2) 2009-2014
Var job	0.0097	0.0105
Var indv	0.1110	0.1139
Cov	0.0048	0.0082
Corr	0.1458	0.2357
R ²	0.748	0.781
Adj R ²	0.669	0.713

Tabell 7 og 8 viser deskriptiv statistikk og resultat fra estimering av yrkesfaste effekter på en modell med sju-sifrede styrk-koder. Dette innebærer en langt mer detaljert inndeling i yrkeskategorier enn de firesifrede styrk-kodene benyttet i tabell 5 og 6. Vi ser dette i tabell 7 i kolonne (1) og (2) hvor vi har henholdsvis nesten 9000 og 7000 unike yrker i det originale datasettet. Vi ser at også denne modelleringen gir nokså likt resultat som de skissert ovenfor. Variasjonen til lønnsvariabelen er økende, det samme er modellens forklaringskraft og variansen til individeffektene. Vi ser dessuten at korrelasjonen mellom yrkes- og individeffekter øker, samt en økning i variansen til yrkeseffektene.

5.1.3 Oppsummering

Vi observerer altså en gjennomgående økning i variansen til lønnsvariabelen. Dette funnet er i tråd med hypotesen om økende lønnsulikhet, og stemmer overens med empirien skissert i seksjon 2. Merk at estimeringen er foretatt på et utvalg som kun består av fulltidsansatte menn. Dette resultatet innebærer med andre ord at økende inntektsforskjeller i samfunnet ikke bare drives av lønnsforskjeller mellom kjønn, eller mellom arbeidere i hel- og deltidsstilling, og heller ikke bare av differansen mellom individer utenfor og innenfor arbeidsmarkedet.

Vi observerer i tillegg at en modell basert på individ- og yrkesfaste effekter forklarer mer av variasjonen i lønn over tid, med andre ord spiller individ- og yrkesfaste effekter en større og større rolle for lønnen arbeiderne mottar. Kontrollert for kun individeffekter viser det seg

imidlertid at yrkesheterogenitet ikke spiller noen særlig rolle for lønnsforskjellene. Resultatene mine tyder altså på at det er permanente forskjeller mellom arbeiderne som driver den økte inntektsulikheten.

5.2 Betydning av yrker

Skissert litteratur i seksjon 2 peker i retning av at ny teknologi og økt automatisering vil føre til økt betydning av yrkeseffekter, fordi, som Autor (2015) argumenterer for, vil teknologiutviklingen gjøre noen av oss mer produktive mens andre blir overflødige. Resultatene i seksjon 5.1 tyder på at yrker ikke er viktige bidragsytere for trenden med økende lønnsvariasjon, og for konsistens bør derfor en analyse av virkningen av automatisering på yrkesheterogenitet gi lite effekt. Vi husker fra modellspesifikasjonen i seksjon 4.1 at komponenten som fanger opp yrkesheterogeniteten kan forklares som en proporsjonal lønnsgevinst alle arbeidere som er sysselsatt i det spesifikke yrket mottar. Dersom det stemmer at yrker forklarer lite av økende inntektsulikhet bør ikke teknologisk utvikling være en sentral driver bak endringene i den yrkesspesifikke lønnsgevinsten over tid.

Jeg skal derfor nå forsøke å teste betydningen av teknologisk utvikling på de yrkesfaste effektene. Med bakgrunn i resultatene for yrkeseffekter er hypotesen min at den akselererende utviklingen i, og implementeringen av, ny teknologi ikke er viktige drivere bak heterogeniteten mellom yrker.

For å måle effekten av ny teknologi og automatisering har jeg benyttet fire ulike variabler. *AI-exp* i tabell 9 er et mål på hvor utsatt et yrke er for kunstig intelligens, i form av en effektscore utviklet av Felten, Raj og Seamans (2018). Kort forklart er scoren konstruert ved å benytte mål på historisk utvikling i kunstig intelligens på oppgavenivå fra The Electronic Frontier Foundation (EFF),³⁴ og videre koble dette sammen med yrkesdatabasen til O*NET som består av et stort sett yrkestitler og drøyt 50 ulike egenskaper som i varierende

³⁴ EFF måler utviklingen i teknologiens evne til å utføre definerte oppgaver, f.eks. ansiktsgjenkjenning, sortert i 16 ulike kategorier. Historisk utvikling avgjør hvilken økningsrate hver enkelt kategori tildeles (Felten et al. 2018).

grad er viktig i de ulike yrkene.³⁵ En sammenkobling av disse dataene gir en score som tildeles hver enkelt yrkeskode, og som slik gjør det mulig å sammenligne hvor utsatt ulike yrker er for kunstig intelligens.^{36 37} De tre neste variablene, henholdsvis *Routine*, *Manual* og *Abstracts*, er hentet fra Autor og Dorn (2013), som bygger sitt empiriske arbeid på Autor m.fl. (2003). Disse tre variablene er konstruert ved å identifisere hvilke arbeidsoppgaver som løses i ulike yrker, og deretter utforme en indeks over hvor intensiv de ulike yrkene er i de tre variablene som studeres. Vi får slik en oversikt over hvilke yrker som i stor grad er preget av rutinearbeid, manuelt arbeid og hvilke som krever mye av elementer som kreativitet og problemløsning.³⁸ Dette er ment å måle hvor lett de ulike arbeidsoppgavene, og dermed de ulike yrkene, kan utføres av maskiner.

³⁵ Databasen inneholder en oversikt over hvor utbredt og hvor viktig disse egenskapene er på yrkesnivå.

³⁶ Verdien på scoren er arbitrær, men den gjør det mulig å sammenligne på tvers av ulike yrker

³⁷ Benyttet denne effektscoren ved å kjøre crosswalk fra amerikanske til norske yrkeskoder

³⁸ Se Autor og Dorn (2013) for nærmere beskrivelse av variablene. Har også her crosswalket fra amerikanske yrkeskoder.

Tabell 9: Estimeringsresultat for yrkeseffekter mot teknologivariabler

	(1) PSI-AI	(2) PSI-Rout	(3) PSI-Manu	(4) PSI-Abstr	(5) PSI-All
<i>a) Yrkeseffekter tidlig periode</i>					
AI-exp	0.066*** (0.011)				0.046*** (0.013)
Routine		-0.000 (0.005)			-0.002 (0.006)
Manual			0.007 (0.006)		0.011 (0.008)
Abstracts				0.017*** (0.003)	0.010*** (0.004)
<i>N</i>	241	256	256	256	241
<i>R</i> ²	0.110	0.000	0.004	0.071	0.130

b) Yrkeseffekter sen periode

AI-exp	0.056*** (0.009)				0.049*** (0.014)
Routine		0.004 (0.003)			-0.000 (0.003)
Manual			0.002 (0.006)		0.002 (0.007)
Abstracts				0.012*** (0.002)	0.004 (0.004)
<i>N</i>	239	253	253	253	239
<i>R</i> ²	0.178	0.006	0.001	0.084	0.183

c) Yrkeseffekter endring

AI-exp	0.009 (0.013)				-0.004 (0.017)
Routine		-0.003 (0.006)			-0.002 (0.007)
Manual			0.006 (0.006)		0.010 (0.008)
Abstracts				0.004 (0.003)	0.006 (0.004)
<i>N</i>	238	252	252	252	238
<i>R</i> ²	0.002	0.002	0.004	0.004	0.013

Robuste standardfeile i parantes

* p<0.05, ** p<0.01, ***p<0.001

Tabell 9 består av tre ulike panel, og gir resultatene fra regresjoner av de yrkesspesifikke effektene mot fire ulike variabler som potensielt kan vise effekten av teknologisk utvikling på yrkesnivå. Den valgte estimeringsstrategien benyttet i seksjon 5.1 gjør det mulig å hente opp igjen de estimerte verdiene på individ- og yrkeseffektene. Dette utnytter jeg i denne delen av analysen for å kunne studere faktorer som kan påvirke yrkesheterogeniteten. Panel a gir estimeringsresultatene for yrkeseffektene i tidlig tidsintervall og panel b gir tilsvarende resultat for sent tidsintervall. I panel c har jeg sett på endringen i yrkeseffektene fra tidlig til sen periode.

Kolonne (1)-(4) i tabellen gir estimeringsresultater fra en enkel regresjon med yrkeseffekter mot henholdsvis AI-exp, Routine, Manual og Abstracts. Kolonne (5) gir resultatene fra en regresjon med yrkeseffekter mot alle fire forklaringsvariabler. Vi ser av kolonne (1) at estimatet for eksponering for kunstig intelligens er statistisk signifikant forskjellig fra null, og at dette resultatet både gjelder for tidlig og sent tidsintervall. Det fremgår av kolonne (5) at variabelen er signifikant også når det kontrolleres for de andre forklaringsvariablene, og at også dette resultatet gjelder i begge tidsperioder. Merk at verdien på koeffisienten reduseres over tid i kolonne (1), men at den øker svakt fra tidlig til sent tidsintervall i kolonne (5). Vi ser i kolonne (4) at estimert verdi på Abstracts også er statistisk signifikant forskjellig fra null i begge tidsperioder, men det er kun i tidlig intervall at estimatet også er signifikant når det kontrolleres for andre variabler. Estimert verdi faller fra tidlig til sen tidsperiode i både kolonne (4) og (5) for Abstracts. Hverken Routine eller Manual har signifikante estimater i noen tidsintervall, og er heller ikke signifikante i kolonne (5) hvor alle forklaringsvariabler er inkludert. De robuste standardfeilene er jevnt over ganske små, og endres lite mellom tidsintervallene for alle koeffisienter. Vi ser i panel c at ingen estimat er statistisk signifikant forskjellig fra null når vi estimerer med endring i yrkeseffekter som avhengig variabel. Vi ser videre at de estimerte verdiene jevnt over er lave, og at for de fleste forklaringsvariablene er de estimerte standardfeilene høyere enn de estimerte koeffisientverdiene.

Forklaringskraften i panel a og b er høyere enn i panel c, men den varierer mye mellom de ulike kolonnene. Konstruksjonen til R^2 gjør det naturlig at forklaringskraften er høyest i kolonne (5), fordi R^2 "alltid" øker når man legger til flere variabler i estimeringen. Det er derfor verdt å merke seg at forklaringskraften øker relativt lite fra kolonne (1) til kolonne (5), særlig i panel b. Økningen i R^2 fra tidlig til sen periode for kolonne (1) og (5) reflekterer at

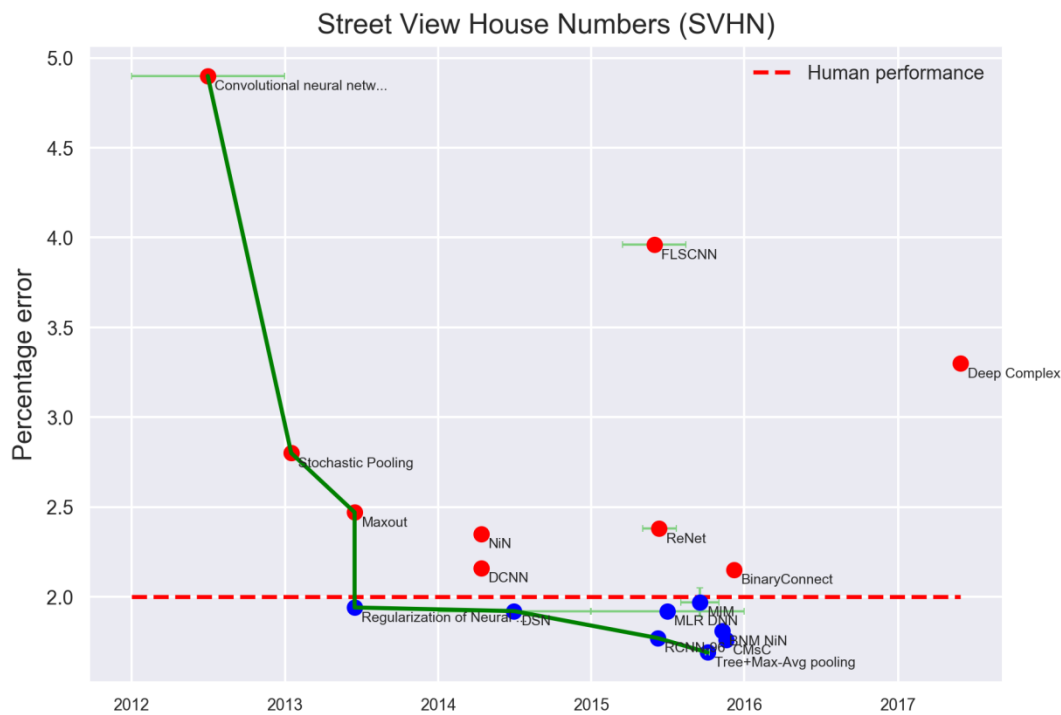
forklaringsvariablene forklarer mer av «mellom yrke»-variansen i tverrsnittet over tid, selv om ingen av variablene ser ut til å ha noen betydning for endringen i yrkeseffekter.

Resultatene i tabell 9 er interessante av flere årsaker. Først og fremst er mangelen på signifikante effekter i panel c konsistent med funnene i seksjon 5.1, knyttet til at økende lønnsulikhet ikke kan forklares av yrkesspesifikke effekter. Med bakgrunn i dette er det derfor ikke overraskende at endringen i yrkesheterogenitet fra tidlig til sen periode ikke kan forklares av de inkluderte målene på automatisering og eksponering for kunstig intelligens. Ser dette av at ingen av de estimerte verdiene er statistisk signifikante.

Basert på panel a og b kan resultatene tyde på at yrker som scorer høyt på elementene som ligger i “Abstracts” påvirker den yrkesspesifikke lønnsgevinsten, men altså ikke i endringene i yrkeseffektene gitt i panel c. Dersom yrkeseffekter hadde forklart mye av de økte lønnsulikhetene kunne de signifikante verdiene i panel a og b muligens blitt tolket som støtte til Autor (2015), som hevder at teknologiutvikling øker produktiviteten til de på toppen av lønnsfordelingen. Han predikerte videre at arbeidskraften “på midten” møter størst risiko for automatisering, og Autor og Dorn (2013) argumenterer for at dette er klassisk rutinepreget arbeid, eksemplifisert ved regnskapsføring og kontorarbeid. Dette gir resultatene i tabell 9 lite støtte til, ved at ingen av verdiene er signifikante.

Det signifikante resultatet på eksponering for kunstig intelligens i panel a og b er interessant på flere måter. Resultatene tyder isolert sett på at risiko for automatisering påvirker variasjonen i de yrkesspesifikke effektene. Merk at modellens forklaringskraft ikke er spesielt høy, men relativt lav forklaringskraft er naturlig i en situasjon hvor vi tester en avhengig variabel vi vet består av mange faktorer opp mot ett konkret mål på teknologi. En klar svakhet med denne analysen er at tidsperioden jeg studerer stopper i 2014, og det vil være lett å argumentere for at effekten og betydningen av kunstig intelligens ville vært større med nyere data. Nettopp med hensyn på tidsaspektet i analysen er funnene i panel a og b noe oppsiktsvekkende. Selv om ingen av målene på teknologi ser ut til å ha en effekt på *endring* i yrkeseffekter, så tegner likevel tabell 9 et bilde av at eksponering for kunstig intelligens har en innvirkning på yrkeseffektene i begge perioder. Dette er et resultat som kanskje særlig kan virke urealistisk i tidlig tidsintervall.

Figur 4: Utvikling i kunstig intelligens målt mot det menneskelige intellektet.



Kilde: www.eff.org

Figur 4 viser tidsutviklingen til én bestemt kategori kunstig intelligens målt mot menneskers evne til å utføre den samme oppgaven. Her eksemplifisert ved evnen til å gjenkjenne tall. Vi ser at 2013 markerer bruddpunktet for hvor teknologien ble mer effektiv enn mennesket, slik at automatisering fra og med dette tidspunktet øker produktiviteten i denne konkrete arbeidsoppgaven, alt annet konstant. Basert på dette vil det være mulig å se for seg at kunstig intelligens påvirker yrkesheterogeniteten i sen periode i vårt datasett, men med tanke på at estimert koeffisientverdi er *høyere* i tidlig enn i sen er det grunn til å mistenke at noe forstyrrer resultatene.

Det finnes flere svakheter knyttet til dette målet jeg benytter på eksponering for kunstig intelligens. Felten et al. peker blant annet på at det å benytte historisk utvikling som mål på fremtidig utviklingshastighet kan vise seg å være lite presist. En tilleggsulempe med dette er at for en del kategorier mangler EFF helt eller i stor grad data på historisk utvikling, dette gjør estimering av helning på fremtidig utvikling vanskelig. For å kunne benytte scoren på norske yrkeskoder måtte jeg dessuten kjøre en crosswalk fra de amerikanske kodene, en utfordring med dette er at O*NET opererer med nesten 1000 unike koder mens jeg i denne analysen har benyttet 4-sifrede styrk-koder med 259 unike koder. Jeg har derfor måttet aggregere scoren

slik at hver av de 259 kodene fikk én score hver. Dette reduserer presisjonsnivået sammenlignet med Felten et. al sine data. Dette er en svakhet som også gjelder de tre andre variablene jeg har benyttet i denne delseksjonen, fordi også de er crosswalket fra amerikanske yrkeskoder. Det kan dessuten være fristende å kritisere at jeg her implisitt antar at kunstig intelligens vil påvirke yrker i Norge likt som yrker i USA, samt at de samme yrkene sorteres under henholdsvis routine, manual og abstracts. Likevel ser jeg liten grunn til at arbeidsoppgavene som utføres i ulike yrker er så ulike på tvers av land at det gjør variablene ugyldige på norske data.

5.2.2 Oppsummering

I tråd med hypotesen finner jeg ingen effekt av målene på automatisering og kunstig intelligens på endring i de yrkesspesifikke effektene. To av målene, AI-exp og Abstracts, gir signifikante utslag på yrkeseffektene i tidlig og sen periode, likevel er det en del elementer som tyder på at disse resultatene ikke bør tillegges for mye vekt. Med bakgrunn i tidsaspektet for analysen fremstår det lite sannsynlig at effekten av eksponering for kunstig intelligens har sterkere betydning i tidlig tidsintervall sammenlignet med i sent. Hovedpoenget er likevel resultatene i panel c som viser ingen signifikant effekt på endring i yrkesheterogeniteten, noe som er konsistent med at vi i seksjon 5.1 fant at yrker ikke har stor betydning for lønnsulikheten.

6 Konklusjon

I denne oppgaven har jeg undersøkt lønnsulikhet i Norge i perioden 1999 til 2014. Ved hjelp av en AKM-dekomponering av lønn har jeg studert bidraget til yrkesspesifikke effekter opp mot individspesifikke effekter, samt sett på bedriftsfaste effekter som en replikering av Card m.fl. sitt arbeid på vest-tyske data. Jeg har i tillegg undersøkt om teknologisk utvikling har hatt en effekt på yrkesheterogeniteten over tid. Jeg definerer her yrkesheterogenitet som en proporsjonal lønnsgevinst som alle arbeidere i det spesifikke yrket mottar.

I tråd med relevant empiri viser mine undersøkelser at lønnsulikheten i Norge har økt i løpet av den gjeldende tidsperioden, men at variasjonen i lønn er mindre enn det Card m.fl. fant for Tyskland. Dette er konsistent med både differanser i institusjonelle forhold mellom landene, samt IMF (2017) sine beregninger som knytter fallende lønnsandel til økende inntektsulikhet. Husker fra seksjon 2 at Norge har opplevd økende lønnsandel i motsetning til de fleste OECD-land.

Hovedformålet med oppgaven har vært å benytte den etterhvert så hyppig brukte AKM-metoden på en ny måte, nemlig ved å dekomponere lønnsvariansen til individ- og yrkeseffekter på norske data. Hovedanalysen viser noe oppsiktsvekkende at den økte lønnsulikheten ikke kan tilskrives yrkesspesifikke effekter. Dekomponeringen til bedriftseffekter understøtter også at lønnsforskjellene i Norge drives av individer.

Del to av den empiriske analysen i oppgaven studerer virkningen av automatisering og kunstig intelligens på yrkesheterogenitet. Jeg finner her at ingen av de inkluderte «teknologivariablene» har en signifikant effekt på endringen i de yrkesspesifikke effektene. Betinget på resultatet fra AKM-dekomponeringen er dette et logisk resultat. Hvis yrker ikke er spesielt viktig for lønnsulikhet så vil automatisering trolig heller ikke være så viktig for lønn. Merk likevel at denne analysen omfatter en tidsperiode som slutter i 2014, og en kan argumentere for at mye har skjedd på dette feltet siden da.

Konklusjonen blir dermed tredelt, i den forstand at jeg finner at lønnsulikheten har økt, denne økningen er ikke i særlig grad drevet av yrkesheterogenitet og automatisering er ikke en viktig forklaringskraft for de yrkesspesifikke effektene.

Referanseliste

- Aaberge, R. og Modalsli, H. (2014). Fordelingen av inntekter i Norge fra 1875 til i dag.
Hentet fra:
https://www.sv.uio.no/econ/english/research/publications/files/2014/fordelingen_av_inntekter_i_norge_fra_1875_til_i_dag.pdf
- Aaberge, R. (2016). Inntektsulikhet i Norge i lys av Piketty-debatten. *Samfunnsspeilet*, (1).
Hentet fra: https://www.ssb.no/inntekt-og-forbruk/artikler-og-publikasjoner/_attachment/260238?_ts=1537457cb00
- Abowd, J., Kramarz, F., og Margolis, D. (1999) High Wage Workers and High Wage Firms. *Econometrica*, 67(2), doi: 10.1111/1468-0262.00020
- Acemoglu, D. og Restrepo, P. (2018). Automation and New Tasks: How Technology Displaces and Reinstates Labor. *Journal of Economic Perspectives*, 33(2).
- Andrews, M., Schank, T. og Upward, R. (2006). Practical fixed-effects estimation methods for three-way error-components model. *The Stata Journal*, 6(4).
- Andrews, M.J., Gill, L., Schank, T. og Upward R. (2008). High Wage Workers and Low Wage Firms: Negative Assortative Matching or Limited Mobility Bias. *Journal of the Royal Statistical Society*, 171(3).
- Autor, D., Levy, F. og Murnane, R. (2003). The Skill Content of Recent Technological Change: An Empirical Exploration. *The Quarterly Journal of Economics*, 118(4),
- Autor, D. og Acemoglu, D. (2011) Skills, Tasks, and Technologies: Implications for Employment and Earnings. *Handbook of Labor Economics*, 4(B),
- Autor, D. og Dorn, D. (2013). The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market. *American Economic Review*, 103(5).

- Autor, D. (2015). Why Are There Still So Many Jobs? The History and Future of Workplace Automation. *Journal of Economic perspectives*, 29(3).
- Bjørnstad, R. og Nymoen R. (2015). *Frontfagsmodellen i fortid, nåtid og framtid*. (1). Hentet fra: <https://static1.squarespace.com/static/576280dd6b8f5b9b197512ef/t/577508e09de4bbbd55a58d55/1467287781204/SFL+R1-2015+Den+norske+frontfagsmodellen.pdf>
- Card, D., Heining, J. og Kline, P. (2013) Workplace heterogeneity and the rise of West German wage inequality. *The Quarterly Journal of Economics*, 128(3),
- Donovan, S. og Bradley, D. (2018). Real Wage Trends, 1979 to 2017. Hentet fra: <https://fas.org/sgp/crs/misc/R45090.pdf>
- ETUI. (2015). Collective Bargaining. Hentet fra: <https://www.worker-participation.eu/National-Industrial-Relations/Across-Europe/Collective-Bargaining2>
- Falch, T. og Strøm B. (2006). Local flexibility in wage setting: evidence from the Norwegian local public sector. *Empirical Economics*, 31(1),
- Felten, E., Raj, M. og Seamans, R. (2018). A Method to Link Advances in Artificial Intelligence to Occupational Abilities. *AEA Papers and Proceedings*, 108.
- Finansdepartementet. (2019). *Muligheter for alle*. (Meld. St. 13 2018-2019). Hentet fra <https://www.regjeringen.no/no/dokumenter/meld.-st.-13-20182019/id2630508/sec1>
- Goldin, C. og Lawrence F. K., 2008. *The Race between Education and Technology*. Cambridge: Harvard University Press.
- Grytten, O. (1995): Dagens arbeidsledighet i lys av arbeidsledigheten i 1920- og 1930-årene. *Sosialøkonomen*, 49(6).

- Hansen, S. og Skoglund, T. (2008). Utviklingen i sysselsetting og lønn etter 1930. *Økonomiske analyser* 6/2008. Hentet fra:
https://www.ssb.no/a/publikasjoner/pdf/oa_200806/skoglund.pdf
- Hibbs, D. og Locking, H. (1996). Wage compression, wage drift and wage inflation in Sweden. *Labour Economics*, 3(2),
- IMF. (2017, oktober). World Economic Outlook. Hentet fra:
<https://www.imf.org/en/Publications/WEO/Issues/2017/09/19/world-economic-outlook-october-2017>
- Jenkins, D. og Morin, A. (2018). Job-to-job transitions, sorting, and wage growth. *Labour Economics*, 55.
- Kahn, L. (1998). Collective Bargaining and the Interindustry Wage Structure: International Evidence. *Economica*, 65(260),
- Keynes, J. (1930). Economic Possibilities for our Grandchildren. Hentet fra:
<https://www.marxists.org/reference/subject/economics/keynes/1930/our-grandchildren.htm>
- Kostøl, A.R. (2017): *Mismatch and the Consequences of Job Loss* (Doktorgradsavhandling). Universitetet i Bergen.
- Kolb, M. (2018, 29. oktober). What Is Globalization?. Hentet fra:
<https://piie.com/microsites/globalization/what-is-globalization.html>
- Kvaløy, O. (2014). Offer for egen suksess. *Dagen Næringsliv*, 21. mars 2014.
- NAV. (2018, 9. desember). Sykepenger til arbeidstakere. Hentet fra:
<https://www.nav.no/no/Person/Arbeid/Sykmeldt%2C+arbeidsavklaringspenger+og+yrkesskade/Sykepenger/Sykepenger+til+arbeidstakere>

- NOU 2016: 3. (2016). Ved et vendepunkt: Fra ressursøkonomi til kunnskapsøkonomi. Hentet fra: <https://www.regjeringen.no/no/dokumenter/nou-2016-3/id2474809/>
- OECD. (2010). *Sickness, Disability and Work: Breaking the Barriers*. Hentet fra: <https://www.oecd.org/publications/sickness-disability-and-work-breaking-the-barriers-9789264088856-en.htm>
- OECD. (2016). *OECD Factbook 2015-2016: Economic, Environmental and Social Statistics*. Paris: OECD Publishing.
- OECD. (2018, 4. juli). *Employment Outlook 2018*. Hentet fra: <http://www.oecd.org/social/oecd-employment-outlook-19991266.htm>
- OECD. (2019a). Income inequality (indicator). Hentet 28. mai 2019 fra: <https://data.oecd.org/inequality/income-inequality.htm>
- OECD. (2019b). Collective bargaining coverage. Hentet 28. Mai 2019 fra: <https://stats.oecd.org/Index.aspx?DataSetCode=CBC#>
- Oxfam. (2019, mars). Mind the gap: The state of employment in India. Hentet fra: <https://www.oxfamindia.org/sites/default/files/2019-03/Full%20Report%20-%20Low-Res%20Version%20%28Single%20Pages%29.pdf>
- Pinker, S. (1997). *How the Mind Works*. Hentet fra: <https://nyaspubs.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1749-6632.1999.tb08538.x>
- Roser, M. og Ortiz-Ospina, E. (2016, oktober). *Income Inequality*. Hentet fra <https://ourworldindata.org/income-inequality>
- Regjeringen. (2015, 8. mai). Endringer i reglene om midlertidig ansettelse fra 1. juli 2015. Hentet fra: <https://www.regjeringen.no/no/tema/arbeidsliv/arbeidsmiljo-og-sikkerhet/innsikt/arbeidsmiljolooven/lovendringeri-arbeidsmiljolooven-fra-1.-juli--2015/endringer-i-reglene-om-midlertidig-ansettelse-fra-1.-juli-2015/id2410607/>

SSB. (2012, 16. april). Begreper i nasjonalregnskapet. Hentet fra:

<https://www.ssb.no/nasjonalregnskap-og-konjunkturer/begreper-i-nasjonalregnskapet>

SSB. (2014, 12. desember). A-ordningen – én datakilde for tre etater. Hentet fra:

<https://www.ssb.no/omssb/om-oss/nyheter-om-ssb/a-ordningen-en-datakilde-for-tre-etater>

SSB. (2017, 6. september). Over 200 000 midlertidig ansatte. Hentet fra:

<https://www.ssb.no/arbeid-og-lonn/artikler-og-publikasjoner/over-200-000-midlertidig-ansatte>

Verbeek, M. (2017). A Guide to Modern Econometrics (5 utg.). Chichester: John Wiley & Sons.

Wooldridge, J.M. (2015). Introductory Econometrics: A Modern Approach (6 utg.). Boston: Cengage Learning.

