

**Globalisering og arbeidsmarkedet:  
En empirisk analyse av offshoring og  
etterspørsel etter ulike typer arbeidskraft**

av

Sigurd Elvbakken Birkeland

**Masteroppgave**

Masteroppgaven er levert for å fullføre graden

**Master i samfunnsøkonomi**

Universitetet i Bergen, Institutt for økonomi

Juni 2008

UNIVERSITETET I BERGEN



# Forord

Veiledere: Arild Aakvik og Frode Meland.

Først og fremst må jeg takke mine to fantastiske veiledere! Deres alltid åpne dører har sørget for at jeg leverer i tide.

I tillegg er det verdt å nevne at tiden på instituttet har vært en fryd. Heldigvis har jeg vært omgitt av dyktige, hjelpsomme og trivelige medstudenter, som, blant andre, Erlend Dysvik og Pål Asle Reiersgaard. De har gitt meg uvurderlig hjelp og oppmuntring hele veien.

Arbeidet med denne oppgaven har lært meg mye. Blant annet at datainnsamling kan være en lang og tidkrevende prosess, med et datasettet som først var komplett 11. april 2008, og at dess mer en graver seg ned i problemer, jo lenger unna løsningen kommer en.

Takk til SSB for å gjøre gode data tilgjengelig for alle, og imøtekommende ansatte. Det gleder meg at de tok i mot forespørslene mine, og etter hvert har publisert et datasett jeg etterspurte.

---

Sigurd Elvbakken Birkeland, Bergen 2. juni 2008

# Sammendrag

---

## **Globalisering og arbeidsmarkedet: En empirisk analyse av offshoring og etterspørselen etter ulike typer arbeidskraft**

av

**Sigurd Elvbakken Birkeland, Master i samfunnsøkonomi**

Universitetet i Bergen, 2008

Veiledere: Arild Aakvik og Frode Meland

---

Offshoring er for høyinntektsland antatt å føre til relativt høyere etterspørsel etter kompetent arbeidskraft. Dette forholdet er blitt undersøkt for en rekke vestlige land, og denne oppgaven forsøker å analysere sammenhengen i Norge i perioden 1996 - 2005. Datasettet er samlet inn fra kryssløpstabeller og lønns- og sysselsettingstall publisert av Statistisk Sentralbyrå. Som mål på kompetanse brukes utdanning. Bakgrunnen for analysen er en handelsteoretisk modell utviklet av Feenstra og Hanson. Utledet fra standard mikroøkonomisk teori, er det blitt forsøkt å lage en strukturell modell for estimering av etterspørselsetastisiteter. Dette er blitt gjort ved å ta utgangspunkt i en translog kostnadsfunksjon. Med observasjoner for 51 næringer over 10 år, har det vært naturlig å ta i bruk forskjellige paneldatateknikker. Den økonometriske analysen er utført i STATA. Resultatene som kommer frem i analysen antyder at offshoring fører til reduserte lønnsforskjeller mellom høyt og lavt utdannet arbeidskraft, noe som er overraskende.

1	Innledning.....	1
2	Forskningsstatus.....	5
2.1	Teorier om sammenhengen mellom lønnsgap og handel.....	5
2.2	Empiriske studier.....	9
3	Presentasjon av mikroøkonomisk teori og økonometrisk metode.....	13
3.1	Strukturell modellering.....	14
3.2	Estimering med minste kvadrats metode.....	19
3.3	Estimering med paneldatateknikker.....	20
3.4	Problemer med modellene, og valg av riktig modell.....	24
3.5	Tilsynelatende urelaterte regresjoner (SUR).....	25
3.6	Estimering med dynamiske modeller.....	29
4	Presentasjon av datasettet.....	32
4.1	Datakilder.....	32
4.2	Variabeldefinisjoner.....	34
4.3	Mer om variablene: Deskriptiv statistikk.....	37
5	Resultater fra den økonometriske analysen.....	49
5.1	Sammenslåtte tverrsnittsdata.....	49
5.2	Paneldatateknikker.....	55
5.3	Problemer med modellene.....	59
5.4	Dynamiske modeller.....	64
5.5	Kvantifisering av resultatene.....	68
5.6	Kontroll og alternative spesifiseringer.....	70
6	Oppsummering.....	72
	Referanser.....	75
	Appendiks A.....	77
	Appendiks B.....	81
	Appendiks C.....	82
	Appendiks D.....	84

Tabell 1: Inndeling av arbeidskraft etter høyeste gjennomførte utdanning. ....	34
Tabell 2: Oversikt over variablene .....	37
Tabell 3: Gjennomsnittlige verdier i utvalgte år .....	38
Tabell 4: Årlige, prosentvise endringer i kostnadsandelene, 1996 - 2005 .....	40
Tabell 5: Årlige, prosentvise endringer i inntekt, 1996 - 2005 .....	41
Tabell 6: Årlige, prosentvise endringer i sysselsetting, 1996 - 2005 .....	42
Tabell 7: Utvikling i smal- og bredimport, 1992 - 2005 .....	45
Tabell 8: Utvikling i fast realkapital, produksjon og verdiskapning; 1992 -2005 .....	47
Tabell 9: Utvikling i kjøp av tjenester fra FoU-næringen; 1992 - 2005 .....	48
Tabell 10: Kostnadsandelene regressert med minste kvadrats metode .....	51
Tabell 11: Etterspørselastisiteter basert på regresjon med minste kvadrats metode.....	53
Tabell 12: Kostnadsandelene regressert med fasteffektmodellen .....	55
Tabell 13: Etterspørselastisiteter basert på regresjon med fasteffektmodellen.....	57
Tabell 14: Korrelasjon mellom forklaringsvariablene .....	60
Tabell 15: Autokorrelasjon i fasteffektmodellen .....	62
Tabell 16: Kostnadsandelene regressert med fasteffektmodellen med AR(1)-feilledd .....	65
Tabell 17: Sammenligning av etterspørselastisiteter uten restriksjoner, FE og AR(1) .....	66
Tabell 18: Kostnadsandelene regressert med Arellano-Bond-modellen.....	67
Tabell 19: Effekt av en prosents økning i forkl.var., ettersp.el. basert på FE-modellen.....	69

Figur 1: Uvikling i kostnadsandelene; årlige endringer 1997 -2005.....	41
Figur 2: Inntektsutvikling; årlige endringer 1997 -2005.....	42
Figur 3: Sysselsettingsutvikling; årlige endringer 1997 -2005. ....	43
Figur 4: Utvikling i smalimport; 1992 – 2005. ....	45
Figur 5: Utvikling i bredimport; 1992 - 2005. ....	46
Figur 6: Utvikling i fast realkapital, produksjon og verdiskapning; 1992 - 2005.....	47
Figur 7: Utvikling i kjøp av tjenester fra FoU-næringen; 1992 - 2005 .....	48

# 1 Innledning

Den økte handelen med utlandet har vært gjenstand for stor interesse over lengre tid. Temaet har mange dimensjoner, og økt handel med utlandet kan reise en rekke problemstillinger av så vel politisk som akademisk interesse. Mye av oppmerksomheten har vært rettet mot hvordan globalisering påvirker de nasjonale arbeidsmarkedene rundt om i verden. Det er diskutert om den økte integreringen av verdensmarkedene endrer spillereglene i nasjonale arbeidsmarkeder. Også i Norge har det vært usikkerhet om hvordan etterspørselen etter ulike typer arbeidskraft, og spesielt etter mindre utdannet arbeidskraft, vil endre seg som en følge av utviklingen med økt handel mellom landene som generelt trekk i verdensøkonomien.

Når bedrifter flytter produksjonen til land med billig arbeidskraft, eller kjøper mer varer fra slike land, blir det ofte store medieoppslag om følgene dette kan ha på arbeidsmarkedet og arbeidsplassene i ulike lokalsamfunn i Norge, og ikke minst om hvordan dette påvirker lønn og sysselsetting for norsk arbeidskraft. Selv om lønnsveksten i Norge har vært høy de siste årene, har arbeidstakere med høyest utdanning også vært den gruppen som har opplevd den største relative lønnsøkningen.

Med økt globalisering av økonomien har importen av innsatsvarer økt. Import av innsatsvarer er forbundet med uttrykk som offshoring og outsourcing. Begge begrepene omhandler flytting av produksjon, henholdsvis ut av landet og ut av bedriften. Disse fenomenene kan selvsagt opptre samtidig, for eksempel flytting av produksjon til en annen bedrift i utlandet, *offshore outsourcing*, eller ved flytting til en annen bedrift i hjemlandet *domestic outsourcing*, eller til en utenlandsk del av bedriften, *offshore in-house sourcing*.<sup>1</sup>

I økonomisk sammenheng har debatten om koblingen mellom handel og velferd eksistert lenge, fra Adam Smiths argument om absolutte fortrinn, og Ricardos videreføring av konseptet med komparative fortrinn, til debatten om dagens handelsavtaler. Det blir ofte argumentert for at økt handel totalt sett er bra for alle involverte, men økonomisk handelsteori retter i stor grad oppmerksomheten mot hvordan handelsmønstre påvirker forskjellige deler av samfunnet. Det blir gjort mye forskning på hvem som er skadelidende av nye handelsavtaler,

---

<sup>1</sup> Kilde: <http://www.oecd.org/dataoecd/30/39/38819650.pdf>

og hvem som tjener på dem, og ofte er hensikten med forskningen å finne frem til en forståelse som gjør det mulig å kompensere taperne.

Tidligere var det mest forskningsinteresse knyttet til handel av ferdigvarer, men nyere handelsteori har fokusert på andre aspekter ved økt handel, slik som direkte investeringer i utlandet (FDI), monopolistisk konkurranse, avstand, språk, avgifter og kvoter. Feenstra og Hanson (1996b) påpeker at teorier som inkluderer handel av halvfabrikater, eller innsatsvarer, kan føre til andre konklusjoner om vinnere og tapere, enn teorier som utelukkende fokuserer på handel av ferdigvarer. De argumenterer for at sammenhengen mellom offshoring og etterspørsel etter heterogen arbeidskraft, i betydningene av med ulike kompetansenivå, innebærer at det i begge landene som handler med hverandre vil oppstå økt etterspørsel etter kompetent arbeidskraft. Dette strider mot slutningene fra det såkalte Stolper-Samuelson-teoremet, som lenge var det rådende utgangspunktet for undersøkelser av sammenheng mellom handel og lønns- og sysselsettingsendringer.<sup>2</sup>

Siden midten av nittitallet er det gjennomført mange studier av hvordan sammenhengen mellom offshoring og økende lønns-gap kan beskrives. Det er, så vidt jeg vet, ikke gjort slike studier for Norge. Med forholdsvis ferske studier av fenomenet i blant annet Sverige og England, som har opplevd samme utvikling i importen av innsatsvarer og lønns-gap som Norge, vil en undersøkelse av norske forhold være av interesse. Denne oppgaven følger arbeidene til Ekholm og Hakkala (2005) og Hijzen *et al.* (2005), som omhandler henholdsvis Sverige i perioden 1995 til 2000 og Storbritannia i perioden 1982 til 1996. Andre som har bidratt på dette området er, først og fremst, Feenstra og Hanson (1996a, 1996b, 1999, 2003), som har utviklet modeller og fremgangsmåter for å undersøke dette fenomenet. De har også undersøkt hvordan fenomenet har påvirket lønns-gapet i USA. I tillegg til disse har blant andre Falk og Koebel (2001) undersøkt tilsvarende forhold i Tyskland.

Denne oppgaven tar sikte på å undersøke sammenhengen mellom offshoring og etterspørsel etter forskjellige typer arbeidskraft, og om denne sammenhengen kan forklare det økende lønns-gapet. Ifølge tall publisert av Statistisk Sentralbyrå (SSB) har lønns-gapet mellom

---

<sup>2</sup> Se for eksempel Leamers artikkel: ” *In Search of Stolper-Samuelson Linkages between International Trade and Lower Wages*” (1997).



utdanningsnivåene økt med 2,5 prosentpoeng i perioden 1997 - 2004.<sup>3</sup> Dette er forskjellen mellom sysselsatte med opptil ni års grunnskole, og sysselsatte med mer enn fire års universitets- eller høyskoleutdanning. I denne perioden har også handel med utlandet økt, importen steg med 34 % i samme tidsrom.<sup>4</sup>

Jeg har samlet data for 51 norske næringer. Data inkluderer mål for graden av offshoring, gjennom tall på verdien av importerte innsatsvarer, bruk av fast realkapital, størrelsen på produksjon og verdiskapning i de forskjellige næringene, samlet fra kryssløpstabeller for årene fra 1992 til 2005, og lønns- og sysselsettingstall, for perioden 1996 til 2005, fra SSB. Noe av det som er spesielt med mitt datasett er at det inneholder informasjon om alle størrelsene over hele perioden, mens Ekholm og Hakkala (2005) kun hadde tilgang til kryssløpstabeller for årene 1995 og 2000.

For å estimere sammenhengene økonometrisk har jeg i hovedanalysen utført regresjoner for hvordan kostnadsandelene til fire utdanningsgrupper avhenger av lønnsnivå, fast realkapital, verdiskapning, forskning og utvikling, og offshoring. Kostnadsandelene regnes som hver gruppes andel av variable kostnader, og endringer i disse størrelsene blir tolket som endringer i etterspørselen etter de respektive gruppene arbeidskraft. Hovedsakelig er resultatene funnet ved bruk av fasteffektsmetoden for paneldatasett, med bruk av Zellners metode (SUR) for å øke sikkerheten ved sammenligning av effektene mellom gruppene.<sup>5</sup> Dette gjøres fordi estimeringen av de fire regresjonene er relaterte. Jeg estimerer også en dynamisk modell, noe jeg ikke har sett at er gjort i andre studier på dette området.

Det er grunn til å vurdere resultatene i hovedanalysen som interessante. Hovedsakelig kommer det frem at økt offshoring *ikke* fører til større lønns-gap. Etterspørselen etter den største sysselsettingsgruppen, arbeidstakere med videregående skole, eller tilsvarende, som høyeste gjennomførte utdanning, øker ved økt offshoring. Samtidig finner jeg at gruppen med høyest utdanning opplever avtagende etterspørsel ved økt offshoring. Altså innbærer resultatene at i det i Norge kan se ut til at økt offshoring medfører et negativt press i utviklingen mot økende lønnsforskjeller. Disse resultatene er robuste mot alternative

---

<sup>3</sup>[http://statbank.ssb.no/statistikkbanken/Default\\_FR.asp?PXSid=0&nvl=true&PLanguage=0&tilside=selecttable/hovedtabellHjem.asp&KortnavnWeb=lonnansatt](http://statbank.ssb.no/statistikkbanken/Default_FR.asp?PXSid=0&nvl=true&PLanguage=0&tilside=selecttable/hovedtabellHjem.asp&KortnavnWeb=lonnansatt)

<sup>4</sup>[http://statbank.ssb.no/statistikkbanken/Default\\_FR.asp?PXSid=0&nvl=true&PLanguage=0&tilside=selectvarval/define.asp&Tabellid=05197](http://statbank.ssb.no/statistikkbanken/Default_FR.asp?PXSid=0&nvl=true&PLanguage=0&tilside=selectvarval/define.asp&Tabellid=05197)

<sup>5</sup> SUR står for Seemingly Unrelated Regressions (tilsynelatende uavhengige regresjoner).

spesifiseringer. Dette innebærer at resultatene i denne undersøkelsen skiller seg klart fra resultatene til Ekholm og Hakkala (2005) og Hijzen *et al.* (2005).

Oppgaven er strukturert slik at forskning innenfor området handel og arbeidskraftsetterspørsel presenteres først, i kapittel to. Kapittel tre omhandler metodisk teori, med en presentasjon av hvilke antagelser som ligger til grunn for denne analysen, deretter følger en presentasjon av de forskjellige teknikkene som er brukt i regresjonsanalysen. Kapittel fire inneholder en beskrivelse av datasettet. Videre presenteres resultatene, og sensitivitetsanalysen av disse, i kapittel fem. Til slutt oppsummeres resultatene, og jeg konkluderer prosjektet med forslag til problemstillinger det ville vært interessant og viktig å undersøke nærmere.

## 2 Forskningsstatus

Dette kapittelet omhandler teoretiske og empiriske studier av handel og etterspørsel etter arbeidskraft. Den første delen tar for seg hvordan fokus på problemstillingen har utviklet seg. Del to presenterer en del empiriske arbeider gjort på området.

### **2.1 Teorier om sammenhengen mellom lønns-gap og handel**

Den økende forskjellen i lønn mellom arbeidere med høy og lav utdanning har lenge vært et hett studieobjekt i samfunnsøkonomien.<sup>6</sup> Eksempelvis avtok lønnsforskjellene i USA på 1970-tallet, men økte voldsomt på 1980-tallet. Andelen av arbeidsstyrken med formell kompetanse har økt i den vestlige verdenen, og det kan nok forklare en del av utviklingen på 1970-tallet, men ifølge Bound og Johnson (1992) har ikke tilbudssiden i arbeidsmarkedet påvirket lønnsforskjellene nevneverdig i senere tid, utviklingen i mengden kompetent arbeidskraft har snarere holdt igjen utviklingen. Det er på etterspørselssiden man må lete. To kilder har i litteraturen utkrystallisert seg som de mest sannsynlige grunnene til denne utviklingen; teknologisk utvikling og økende internasjonal handel. Debatten har siden 1980-tallet rast om hvilken av de to som har ansvaret. Mens noen økonomer har holdt teknologisk endring som den kilden med størst forklaringskraft (blant andre Berman *et al.*, 1994, Berman *et al.*, 1998, Lawrence og Slaughter, 1993, Bound og Johnson, 1992 med flere), har andre påstått at også økende integrering av lavinntektsland i verdensøkonomien har gitt et støt i retning økende lønnsforskjeller (Sachs og Shatz, 1994, Feenstra og Hanson, 1996). Blant dem som har diskutert effekten av handel, har det vært debatt om det er pris- eller kvantitetseffekter som er dominerende (to eksempler, henholdsvis Leamer, 1997, Katz og Murphy, 1992). Når det gjelder debatten om teknologisk utvikling, er det skilt mellom *sektorspesifikke* og *faktorspesifikke* teknologiske forbedringer.<sup>7</sup> Senere har også handel av innsatsvarer blitt introdusert som en mulig viktig forklaring på fenomenet. Viktige bidrag er gitt av blant andre Leamer, Deardorff, Krugman, Lawrence, Slaughter, Bhagwati, Feenstra og Hanson.

---

<sup>6</sup> I engelskspråklig litteratur er ofte skillet mellom "skilled" og "unskilled" brukt. I den sammenheng er det ofte såkalte "blue-" og "white collar workers", eller eventuelt produksjons- og ikke-produksjonsarbeidere, det skilles mellom. Skillet mellom "skilled" og "unskilled" er videre i oppgaven ofte referert til som *kompetent* og *ukompetent* arbeidskraft. Selv om dette kan oppleves som diskriminerende, er det her tenkt på som en referanse til *formell* kompetanse, da spesielt (høyere) utdanning.

<sup>7</sup> Henholdsvis *sector biased technological change* og *factor biased technological change*.

Flere studier er gjort hvor det blir undersøkt om faktorinnholdet i handelen kunne ha en sammenheng med økende lønnsforskjeller. Blant andre har, som nevnt, Katz og Murphy (1992) gjort dette. Teorien er at ettersom varer blir produsert med ulike kompetanseintensiteter, vil skiftende handelsmønstre føre til endringer i faktoreterspørselen, og dermed, potensielt, også økende lønnsforskjeller. Basert på mengden arbeidskraft som skal til for å produsere en vare i hjemlig industri, kan import av tilsvarende vare ses på som ekvivalent med import av arbeidskraft. På denne måten kan den effektive endringen i tilbudt arbeidskraft som følge av økt handel kalkuleres. Eventuelt kan importen erstatte hjemlig arbeidskraft slik at etterspørsel etter arbeidskraft går ned. Katz og Murphy følger denne tradisjonen når de prøver å estimere hvordan etterspørselen etter arbeidskraft har endret seg med endrede handelsmønstre (Katz og Murphy, 1992). Selv om de finner en sammenheng mellom handel og økte lønnsforskjeller blant arbeidere med ulike utdanningsnivå, er ”*de forholdsvis små i forhold til den relative økningen i tilbudet av høyere utdannede arbeidere i samme tidsrommet*” (side 65). Også Krugman (2000) argumenterer for at denne fremgangsmåten er berettiget når effekten av handel på lønnsforskjeller skal undersøkes.

I USA økte lønnen til *collage graduates* relativt til *high school graduates* med 11 prosentpoeng i perioden fra 1980 til 1985. Borjas, Freeman og Katz (1992, side 25) har estimert at mellom 15 og 25 % av denne endringen skyldes handel. Handel fremkalte en mindre andel av endringen mellom 1980 og 1988, på grunn av avtagende handelsunderskudd. Dette henger sammen med det ”effektive” innsatsfaktortilbudet handelen representerer. Når det gjelder arbeidstakere som klassifiseres som *high school dropouts* er ikke koblingen mellom endringer i lønn og handel like klar. Dette skyldes at økt immigrasjon presset lønningene ned. De kvantifiserer andelen av den ti prosentpoeng store relative lønnsreduksjonen til denne gruppen, fremkalt av handel og økt immigrasjon, til å være fra tretti til femti prosent. Lawrence og Slaughter (1993) er klar i kritikken av fremgangsmåten til Borjas *et al.* og Katz og Murphy, de mener den er ”*rather weakly grounded in standard trade theory*” (Lawrence og Slaughter, 1993, side 190). De hevder at det ikke er noen klar sammenheng mellom handelsunderskudd og faktorpriser. Bound og Johnson (1992) uttaler muligheten for at handelsunderskuddet som bygget seg opp i USA på 1980-tallet kan ha vært med på å påvirke lønnsforskjellene, men gjennom nedbygging av produksjonssektoren, men finner ikke spesielt klare resultater.

Sachs og Shatz (1994) tar for seg situasjonen som ble observert fra slutten av 1970-tallet, nemlig at det var store sysselsettingsreduksjoner i produksjonen, hvor ikke-kompetent arbeidskraft tok størstedelen av tilpasningen. De bruker kompetanseintensiteten til industriene/bransjene, og undersøker spesielt hvordan handel med utviklings- eller nylig utviklede land påvirker lønns- og sysselsettingsendringene, en tilnærming som inngår i faktorinnholdet-i-handelen-tradisjonen. Undersøkelser av priser på goder importert fra disse landene, viser at det faktisk var en relativ prisreduksjon på ukompetent arbeidskraftsintensive varer. Resultatene av det økonometriske arbeidet deres er at det ser ut til å ha vært et press i retning av større lønnsforskjeller på grunn av handelen med lavinntektsland, selv om de ikke er spesielt signifikante.

Det som hovedsaklig har blitt oppgitt som grunnen til at handel ikke kan være den utslagsgivende kilden til økende lønnsforskjeller, er at handel med lavinntektsland, sett i forhold til BNP, er et relativt lite fenomen. Kritikere av handelsteoretikernes syn hevder at et såpass marginalt fenomen ikke kan ha hatt så stor innvirkning på lønninger som de observerte endringene i lønnsforskjellene tilsier. Som et forsvar mot slike uttalelser er det argumentert at det, i alle fall i et Heckscher-Ohlin-rammeverk (heretter: *HO-rammeverk*), vil være slik at priser og produksjon blir bestemt på marginen. Det vil si at også marginale prisendringer på varer som blir importert, vil kunne ha forholdsvis stor arbeidskraftsetterspørselen (Leamer, 1997).

Når effekten av økt handel med lavinntektsland har blitt undersøkt, har det dermed stort sett blitt gjort med utgangspunkt i et HO-rammeverk. Som en følge av dette rammeverket er Stolper-Samuelson-teoremet etablert. Dette teoremet viser at dersom prisen på et gode øker, vil prisen på den faktoren som brukes relativt mest intensivt i produksjonen av dette godet øke. Med bakgrunn i dette teoremet har flere fremsatt muligheten for at den økte handelen med nylig industrialiserte land, som bør føre til en relativ *prisreduksjon* på det ukompetent arbeidskraftintensive godet for hjemlandet, har gitt nevnte følger i det hjemlige arbeidsmarkedet, *redusert* lønn for lavt utdannet arbeidskraft, og motsatte effekter i de nylig industrialiserte landene. Det er ulikheter i faktorsammensetningen i de forskjellige økonomiene som fører til at handel påvirker lønnsforholdet mellom de ulike typene arbeidskraft.<sup>8</sup>

---

<sup>8</sup> Stolper-Samuelson-teoremet omhandler *reelle* endringer i faktoravlønning.

Edward E. Leamer skriver i artikkelen ” In Search of Stolper-Samuelson Linkages between International Trade and Lower Wages” (1997) om hvilke effekter relativt lavere priser på arbeidsintensive goder hadde på lønnsforskjellene i USA fra 1961 til 1991. Som tittelen antyder er dette i tråd med Stolper-Samuelson-teoremet som artikulerer sammenhengen mellom endringer i relative priser og endringen i relative faktorpriser. Leamer deler inn observasjonene i tre tiår, og han oppdager at historien om globalisering arter seg ganske ulikt i de tre periodene; på 1970-tallet jobbet globaliseringseffekten sterkt mot lavlønnarbeidere, på grunn av fallende priser på ikke-kompetent arbeidskraftintensive goder. På 1980-tallet gikk prispresset fra globalisering i den andre retningen. Leamer hevder at globalisering likevel kan ha påvirket lønnsforskjellene, men gjennom andre kanaler, da spesielt på grunn av økende handelsunderskudd og mindre markedsrett for visse amerikanske sektorer. Han konkluderer med å kalle 1970-tallet for ”*the Stolper-Samuelson decade*” (Leamer, 1997), selv om dette hviler noe på i hvilken grad teknologiske endringer påvirker prisen på goder (*the pass-through rate*).

Lawrence og Slaughter (1993) skriver også om hvordan handel ikke en gang i et HO-rammeverk, kan ha ført til de endringene i arbeidsmarkedet som ble observert i USA på 1980-tallet, ettersom prisforholdet mellom goder som bruker kompetent arbeidskraft intensivt i produksjonen og goder som bruker ukompetent arbeidskraft intensivt ikke endret seg nevneverdig. De observerte at dette prisforholdet endret seg slik at effekten faktisk burde vært at lønnsforskjellene i USA minket (i motsetning til funnene på prisutvikling i Sachs og Shatz, 1994 og Leamer, 1997).

Som en reaksjon på at empirisk forskning innenfor HO-rammeverket ikke fører til klare slutninger om sammenhengen mellom handel og lønnsforskjeller argumenterer Feenstra og Hanson (1996a, 1996b, 1999, 2001) for at man bør skille mellom handel av ferdigvarer og innsatsvarer, varer som er en del av en produksjonsprosess. Slik produksjonsdeling skriver de, påvirker etterspørselen etter arbeidskraft ikke bare i de importkonkurrerende sektorene, men også i de sektorene som bruker de importerte innsatsvarene. De formulerer en modell som tar for seg denne handelen av innsatsvarer.<sup>9</sup> Modellen deres viser at det er en klar teoretisk

---

<sup>9</sup> De lager en enkel modell med tre goder (en ferdigvare, to innsatsvarer), tre faktorer i produksjonen, kapital, kompetent og ikke-kompetent arbeidskraft, og to økonomier, som også blir utvidet til å gjelde for et kontinuum av (uendelig mange) innsatsvarer.

mulighet at handel av innsatsvarer kan øke lønnsforskjellene både i hjemlandet og utlandet. Noe av det som er spennende med denne modellen er at den viser at begge handelspartnerne vil kunne oppleve økende lønnsforskjeller, noe som står i motsetning til resultatene fra HO-rammeverket. De viktigste argumentene mot internasjonal handel som viktig forklaringsfaktor for utviklingen i relative lønninger til ukompetent arbeidskraft er, ifølge Feenstra og Hanson (2003); 1) at handel er et stabilt, lite fenomen i forhold til BNP, 2) usikkerheten tidligere nevnt om retningen på produktprisendringene og 3) at endringene skjer innen industrier ikke mellom industrier slik standard handelsteori (HO-rammeverket) tilsier. Påstanden fra Feenstra og Hanson er at dette ikke er gyldige argumenter dersom det er snakk om handel av innsatsvarer. Når det gjelder 1) er forholdet mellom handel og total BNP ikke så viktig eller nødvendigvis riktig, i og med at det er en stadig voksende servicesektor som det ikke handles med, og samtidig vokser innsatsvareimporten som andel av total import raskt. Det andre argumentet avvises med resonnementet; det er ikke de samme varene som importeres til og som selges i hjemlandet, og at det derfor ikke bør ses på priser mellom industrier, men innen industrier. Dersom man importerer flere innsatsvarer er det sannsynligvis fordi de er tilgjengelig til lavere priser i utlandet enn på hjemmemarkedet. Feenstra og Hanson hevder videre at det ifølge vanlig handelsteori er slik at handel fører til at arbeidskraft blir skiftet mellom industrier, snarere enn innen industriene, og at fokus på handel av innsatsvarer impliserer fokus på skift av arbeidskraft innen industriene. Denne innfalsvinkelen er utgangspunktet for oppgaven, og videre presenteres noen forskningsarbeider med som er sammenlignbare med mitt arbeid.

## **2.2 Empiriske studier**

Inspirasjonen til denne oppgaven er hentet fra Karolina Ekholm og Katariina Hakkalas artikkel fra 2005: "*The Effect of Offshoring on Labor Demand: Evidence from Sweden*". Som tittelen antyder, handler artikkelen om hvordan offshoring har påvirket etterspørselen etter arbeidskraft i Sverige. Resultatene er basert på data for perioden 1995 til 2000. Ekholm og Hakkala deler arbeidsstokken opp i tre deler etter utdanningsnivå, primær-, sekundær- og tertiærutdannelse og finner tall på import av innsatsvarer i kryssløpstabeller fra 1995 og 2000, for årene i mellom brukes import i forhold til produksjon.<sup>10</sup> I tillegg har de lagt stor vekt på

---

<sup>10</sup> De bruker inndelingen *primary*, *secondary* og *tertiary education*, hvor *primary* tilsvarer norsk grunnskoleutdanning, *secondary* videregående skole, og *tertiary* høyskole- og universitetsutdanning.

hvor importen stammer fra, og kjører flere alternative regresjoner hvor de skiller opprinnelseslandene etter verdensdel og inntektsnivå. Den økonometriske analysen deres er basert på en antagelse om at *“technological change as well as offshoring will affect productivity, but not necessarily in a uniform way across industries”* (Ekholm og Hakkala, 2005, s.16). De baserer estimatene på en translog kostnadsfunksjon, for å finne etterspørselastisiteter for de ulike typene arbeidskraft, hvor offshoring blir behandlet som en faktor som endrer teknologien i den hjemlige industrien. Dette er på linje med det som blir gjort i denne oppgaven. Ettersom de mener at effekten av offshoring har likhetstrekk med faktorfremmende teknologiske endringer, bruker de også forskning(s-) og utviklingsandeler (FoU) i næringene som kontroll.

Resultatene som rapporteres i artikkelen er til en viss grad i tråd med hva de forventet å finne, økninger i importen av innsatsvarer, også fra lavinntektsland, reduserte etterspørselen etter arbeidskraft med sekundærutdanning. I tillegg til dette, fant de ut at effekten av økninger i FoU-intensiteten i næringene, er skift i etterspørselen fra arbeidskraft med kun primærutdanning til arbeidskraft med tertiærutdanning. Offshoring til andre høyinntektsland fører ikke til signifikante endringer i arbeidskraftsetterspørselen. Estimaterne på sammenhengen mellom offshoring til lavinntektsland og økende lønnsforskjeller var til dels store, men i forhold til faktiske handelsvolum viser det seg at de ikke kan oversettes til store tap av arbeidsplasser, ganske enkelt på grunn av at handelen med disse landene er så liten at *“an increase by one percentage point would, in fact, imply a doubling from the present level”* (Ekholm og Hakkala, 2005, s.23).

Ifølge Ekholm og Hakkala ligger deres arbeid særlig tett opptil artikkelen til Hijzen et. al (2005) om offshoring og arbeidskraftsetterspørsel i Storbritannia i perioden 1982 til 1996. I likhet med Ekholm og Hakkalas artikkel, tar artikkelen til Hijzen *et al.* for seg hvordan internasjonal outsourcing, målt som import av innsatsvarer, basert på tall fra kryssløpstabeller, har påvirket lønnsforskjellene mellom arbeidere med ulike kompetansenivå. I denne studien blir det, for å finne etterspørselastisiteter for de ulike typene arbeidskraft, også tatt i bruk en translog kostnadsfunksjon. Resultatene de rapporterer viser at økt import av innsatsvarer har ført til at den relative etterspørselen etter ikke-kompetent arbeidskraft avtok. Økt forbruk på forskning og utvikling i næringene førte til en økt etterspørsel etter kompetent arbeidskraft



En studie fra Tyskland, gjort av Martin Falk og Bertrand M. Koebel (Falk og Koebel, 2002), undersøker effekten av innkjøpte tjenester og importerte innsatsvarer på etterspørselen etter det de kaller heterogen arbeidskraft, i de tyske produksjonssektorene. Med heterogen arbeidskraft menes arbeidskraft på tre ulike kompetansenivå. De tester hypotesen om at det er lettere å erstatte ikke-kompetent arbeidskraft med innkjøpte tjenester og importerte materialer, enn å erstatte kompetent arbeidskraft med innkjøpte tjenester og importerte materialer (Falk og Koebel, 2002, s. 573).<sup>11</sup> Også Falk og Koebel tar i bruk kostnadsfunksjoner for å estimere substitusjons- og etterspørselastisiteter. Hovedresultatene deres kommer fra regresjoner basert på en egendefinert Box-Cox-kostnadsfunksjon, og de behandler pris- og kvantitetseffekter hver for seg. De fremhever at det er viktig å skille mellom nettopp disse effektene, ettersom det i deres analyse kun er kvantitetseffekten som har hatt en påvirkning på etterspørselen etter ”heterogen” arbeidskraft, om enn insignifikant. Det viser seg at følgene av økt import av innsatsvarer og økte kjøp av tjenester, *ikke* er signifikante påvirkninger på lønnsforskjellene i Tyskland i perioden de undersøkte. Altså virker det som om det ikke er lettere å substituere den ene typen arbeidskraft enn den andre med importert av innsatsvarer eller innkjøp av tjenester.

Også Feenstra og Hanson har gjort studier på temaet. I Feenstra og Hanson (1999) tester de effekten av outsourcing og høyteknologisk kapital på lønninger. Fremgangsmåten her er ulik fra de tidligere nevnte studiene, ved at regresjonene baseres på en mer langsiktig null-profitantagelse. Resultatene de kommer frem til er at både internasjonal outsourcing og økt bruk av høyteknologisk kapital har vært med å påvirke de økte lønnsforskjellene i USA i årene 1979 til 1990, kvantifisert til henholdsvis 15 og 35 prosent av endringene. Resultatene deres er svært sensitive til hvordan de definerer kapitalen, noe som jeg kommer tilbake til senere i oppgaven. Feenstra og Hanson har også gjennomført en tilsvarende studie hvor det blir brukt en translog kostnadsfunksjon, og tilhørende faktorandeler av kostnadene (Feenstra og Hanson, 1996, 2003),<sup>12</sup> med tilsvarende resultater.

Det er også gjennomført studier hvor sammenhengen mellom lønninger og outsourcing er blitt undersøkt på individnivå (Staffolani *et al.*, 2007 og Geishecker og Goerg, 2008). Studiene, henholdsvis for Italia i perioden 1985-1997 og Tyskland fra 1991 til 2000, konkluderer med at outsourcing har øvd et negativt press på lønningene til mindre kvalifisert arbeidskraft, i det

---

<sup>11</sup> Forfatterne skiller mellom *skilled* og *unskilled* arbeidskraft.

<sup>12</sup> Også i boken ”*Advanced International Trade: theory and evidence*”, Feenstra, 2004

tyske tilfellet i kontrast med funnene til Falk og Koebel (2002). Studiene registrerer også at lønningene til kompetent arbeidskraft har steget som et resultat av outsourcing.<sup>13</sup>

---

<sup>13</sup> Staffolani *et al.* (2007) skriver ”*leaving unchanged or raising the remunerations of White Collars*”.

### 3 Presentasjon av mikroøkonomisk teori og økonometrisk metode

Økonometrisk metode er statistisk undersøkelse av teoretiske sammenhenger, effekter av politiske beslutninger og prediksjon av fremtidige økonomiske størrelser generelt, med mål om å kvantifisere sammenhenger, og å kartlegge kausalitetsspørsmål. Typiske undersøkelser går ut på testing av økonomiske teorier, om politiske beslutninger har hatt ønsket effekt, og om hvordan økonomiske størrelser, som arbeidsledighet, rente, inflasjon og BNP, utvikler seg, og kommer til å utvikle seg. Tolkning av funnene er *ceteris paribus*, det vil si gitt at alle andre aktuelle og relevante faktorer ikke endrer seg. Det ses på isolerte endringer i en variabel av gangen, noe som er viktig i analysen av resultatene.

Økonometriske undersøkelser bygger vanligvis på teoretiske sammenhenger, det vil si at det bør være en antagelse om hvordan ting henger sammen som ligger til grunn for analysen. Testing av resultatene er dermed ofte en undersøkelse om sammenhengene er slik som predikert av teorien. Dersom modellen som estimeres oppfyller kravene og forutsetningene som er definert i økonomisk teori, er det snakk om en *strukturell* modell. Hvis på den annen side kravene ikke er tilfredsstillt, er det snakk om en modell på *reduisert form*. I spørsmålet om funnene impliserer kausalitet, er bakgrunnen for estimering sentral. Dersom modellen som estimeres er en strukturell modell, kan det, med forankring i teorien, med større grad av sikkerhet kunne antas at sammenhengene som er funnet er kausale. En redusert modell er ikke like solid som bakgrunn for studier, og det følger at resultatene i større grad bør ses på som estimerte sammenhenger mellom variablene.

Kapittel fire er organisert slik at den økonometriske spesifiseringen først presenteres. Så utledes den enkleste statistiske modellen som tas i bruk i denne oppgaven, minste kvadrats metode. Deretter utledes paneldatateknikkene. Videre diskuteres eventuelle problemer med modellene og hva som avgjør hvilken modell som bør velges. Til slutt utledes løsninger på to viktige problemer, autokorrelasjon og avhengighet mellom ulike regresjoner. Den økonometriske delen av kapittelet er primært basert på økonometrilærebøkene til Griffiths *et al.* (1993), Wooldridge (2006) og Verbeek (2004).

### 3.1 Strukturell modellering

Strukturell modellering handler i denne sammenheng om å teste empiriske sammenhenger på en måte konsistent med økonomiske teorier. Ifølge Chintagunta *et al.* (2006) gir en strukturell tilnærming muligheter til å teste teorien modellen er utledet fra, i tillegg til å oppnå prediksjoner som er invariante overfor politiske endringer. I det følgende vil jeg vise hvordan denne undersøkelsen kan sies å være utledet fra mikroøkonomisk teori, og hvilke begrensninger på estimeringen som må være oppfylt for at modellen kan oppfattes som strukturell.

#### 3.1.1 Kostnadsfunksjonen

I tråd med blant andre Feenstra og Hanson (2003), Hakkala og Ekholm (2005) og Hijzen *et al.* (2005) baserer jeg undersøkelsen på en translog kostnadsfunksjon. Dette er en funksjon som gir de *variable* kostnadene i de respektive næringene. Noe av det som er gunstig med denne funksjonstypen, er dens egenskaper med tanke på empiriske undersøkelser. Gitt at kostnadsfunksjonen faktisk oppfyller kravene mikroøkonomisk teori stiller til en slik funksjon, kan det vises at andelen av totale variable kostnader som går til hver innsatsfaktor kan uttrykkes som en lineær funksjon av priser og andre inkluderte størrelser. Dette er en fordel i forhold til estimering av faktoreterspørselsfunksjoner og tilhørende etterspørselastisiteter.

Rasjonell og pristagende adferd i næringene innebærer kostnadsminimering gitt priser på variable innsatsfaktorer. Ifølge Varian (1992) vil ”*enhver monoton, homogen og konkav funksjon av priser være en kostnadsfunksjon for en (eller annen) velfungerende teknologi*” (Varian, 1992, side 209).<sup>14</sup> Kostnadsfunksjonen som brukes som grunnlag for denne studien er gitt ved:

$$\begin{aligned} \ln C_i(w, x, z) = & \alpha_0 + \sum_{j=1}^J \beta_j \ln w_{ij} + \sum_{k=1}^K \beta_k \ln x_{ik} + \sum_{r=1}^R \beta_r z_{ir} \\ & + \frac{1}{2} \sum_{j=1}^J \sum_{s=1}^J \beta_{js} \ln w_{ij} \ln w_{is} + \frac{1}{2} \sum_{k=1}^K \sum_{l=1}^K \beta_{kl} \ln x_{ik} \ln x_{il} + \frac{1}{2} \sum_{r=1}^R \sum_{p=1}^R \beta_{rp} z_{ir} z_{ip} \\ & + \sum_{j=1}^J \sum_{k=1}^K \beta_{jk} \ln w_{ij} \ln x_{ik} + \sum_{j=1}^J \sum_{r=1}^R \beta_{jr} \ln w_{ij} z_{ir} + \sum_{k=1}^K \sum_{r=1}^R \beta_{kr} \ln x_{ik} z_{ir}, \end{aligned} \quad (1)$$

<sup>14</sup> Varian bruker frasen *some well-behaved technology*, som jeg oversetter som over.

hvor  $w_{ij}$  = faktorpriser, for faktorene  $j = 1, \dots, h, \dots, s, \dots, J$ .

$x_{ik}$  = faste innsatsfaktorer og verdiskapning, for  $k = 1, \dots, l, \dots, K$ .

$z_{ir}$  = strukturvariabler,  $r = 1, \dots, p, \dots, R$ .

Alt for industri  $i = 1, \dots, N$ .

Ifølge Feenstra og Hanson (2003) skal alle strukturelle variabler som gir skift i produksjonsfunksjonen inkluderes som argumenter i kostnadsfunksjonen. Som påpekt i kapittel 0 er offshoring en faktor som endrer produktiviteten til de forskjellige gruppene arbeidskraft, og det blir argumentert for at dette skjer på samme måte som økt bruk av forsknings- og utviklingstjenester. Dette innebærer at offshoring og forskning og utvikling inkluderes som vektoren  $z$ . Ekholm og Hakkala (2005) og Hijzen *et al.* (2005) inkluderer disse variablene som rater i regresjonene deres, og de hevder at variablene dermed ikke skal uttrykkes i logaritmisk form i (1).

Når funksjonen over defineres til å gjelde på kort sikt, innebærer det at kapital, verdiskapning, offshoring og forskning og utvikling behandles som delvis faste størrelser. Dette betyr at det i regresjonsanalysen vil være slik at verdiene for disse variablene er eksogene, men at de kan variere mellom periodene, og inkluderes i analysen som mengder og andeler i stedet for priser. De variable innsatsfaktorene, i denne sammenheng arbeidskraft fra ulike utdanningsgrupper, er dermed næringenes eneste beslutningsvariabler.

For at regresjonsanalysen kan baseres på denne kostnadsfunksjonen må funksjonen oppfylle noen generelle krav. Tradisjonelt er en translog kostnadsfunksjon oppfattet som *fleksibel* (Diewert og Wales, 1987), noe som innebærer at den er lineært homogen og konkav i innsatsfaktorpriser. Varian (1992, side 208) argumenterer for at ”*kostnadsfunksjonen er en redusert form av et system av faktoreterspørselsfunksjoner*”. Han presiserer at dette kun er tilfellet under kostnadsminimeringsantagelsen, og viser at det er fire krav som må oppfylles for at en kostnadsfunksjon *de facto* er kostnadsminimerende. En kostnadsfunksjon skal være homogen av grad en i faktorpriser, ikke-negativ for ikke-negative priser og kvanta, og økende og konkav i faktorpriser. Dermed legges det vekt på de parametriske betingelsene som må ligge til grunn for at en translog kostnadsfunksjon skal kunne brukes som utgangspunkt for en

strukturell modell. Diewert og Wales (1987) viser at for at funksjonen skal være lineær homogen i priser, er følgende fem antagelser nødvendige, men også tilstrekkelige:<sup>15</sup>

$$\underbrace{\sum_{j=1}^J \beta_j}_{a} = 1, \quad \underbrace{\sum_{j=1}^J \beta_{js}}_{b} = \underbrace{\sum_{s=1}^J \beta_{js}}_{c} = 0, \quad \underbrace{\sum_{j=1}^J \beta_{jk}}_{d} = 0 \quad \text{og} \quad \underbrace{\sum_{j=1}^J \beta_{jr}}_{e} = 0. \quad (2)$$

Restriksjonene a, b, d og e impliserer at effekter av endringer i en variabel ikke kan føre til at endringene i summen av kostnadsandeler kan endre seg. Restriksjon c gir at en identisk skalering av prisene på innsatsfaktorene ikke fører til omalokering av ressursene, og gjelder for alle kostnadsandeler. I tillegg pålegges analysen betingelser om symmetri. Dette innebærer at krysspriseffektene skal være symmetriske mellom de ulike utdanningsnivåene,  $\beta_{js} = \beta_{sj}$ . Dette er en følge av at kostnadsfunksjonen skal være dobbelt deriverbar.<sup>16</sup> Det kan vises at restriksjonene a, b, d og e automatisk er oppfylles i analysen så lenge kostnadsandelene summerer seg til en.<sup>17</sup> Den siste restriksjonen, c for hver kostnadsandel, og symmetrirestriksjonene må derimot pålegges regresjonene for at kravene til kostnadsfunksjonen skal være oppfylt i regresjonsuttrykkene. Dette er gjort i Hijzen *et al.* (2005), Ekholm og Hakkala (2005) og Feenstra og Hanson (2003) sine studier, og er også det i denne oppgaven.

Fra mikroøkonomisk teori kan betinget faktoretter spørsmål finnes ved Shephards lemma, som gir at kostnadsandelene finnes ved derivasjon av (1) med hensyn på  $\ln w_h$ :

$$\frac{\partial \ln C_i}{\partial \ln w_{ih}} = \frac{\partial C_i}{\partial w_{ih}} \times \frac{w_{ih}}{C_i} = \frac{L_{ih} w_{ih}}{C_i} = S_{ih}, \quad (3)$$

Der  $\frac{\partial C_i}{\partial w_{ih}} = L_{ih}$  kommer fra Shephards lemma.

Når restriksjonene nevnt over er pålagt analysen, kan det vises at derivasjon av kostnadsfunksjonen med hensyn på logaritmen av prisen på en innsatsfaktor ( $\delta \ln C_i / \delta \ln w_{ij}$ ) vil

<sup>15</sup> Se appendiks A1.

<sup>16</sup> Per definisjon er de kryssderiverte av C med hensyn på lønn for gruppe s og j like,  $\delta^2 C / \delta w_j \delta w_s = \delta^2 C / \delta w_s \delta w_j$ , dersom funksjonen er to ganger deriverbar, og det kan vises at  $\beta_{js} = \beta_{sj}$  oppfyller dette.

<sup>17</sup> Se appendiks A2.

gi denne faktorens andel av variable kostnader, kostnadsandelen  $S_{ij}$ , vil bruk av Shephards lemma gi kostnadsminimerende mengde av gitt innsatsfaktor. Dette gir uttrykket for kostnadsandelene;<sup>18</sup>

$$\frac{\partial \ln C_i}{\partial \ln w_{ih}} = S_{ih} = \beta_h + \sum_{j=1}^J \beta_{hj} \ln w_{is} + \sum_{k=1}^K \beta_{hk} \ln x_{ik} + \sum_{r=1}^R \beta_{hr} z_{ir} . \quad (4)$$

Dersom antagelsene om parametrene holder, slik at (1) er en gyldig kostnadsfunksjon, vil tall for kostnadsandeler kunne brukes til å estimere effekten av endringer i aktuelle variabler på etterspørselen etter arbeidskraft. Dette innebærer at (4) er en sentral funksjonen i denne studien. Som det kommer frem av (4), er kostnadsandelene lineært avhengige av argumentene i kostnadsfunksjonen. Dette gjør at funksjonen kan estimeres ved hjelp av lineære økonometriske modeller, som minste kvadrats metode. Videre er en styrke med denne funksjonsformen at kvantifisering av resultatene er relativt ukomplisert.

### 3.1.2 Kvantifisering av resultatene; elastisiteter

Etter å ha estimert sammenhengene mellom de ulike variablene og kostnadsandelene, ved ligning (4), vil det være gunstig å omformulere uttrykkene slik at sammenhengene kan uttrykkes som elastisiteter, og som semielastisiteter for strukturvariablene som inngår i regresjonen som rater.

(4) kan ved bruk av (3), skrives om slik at:

$$L_{ij} = \frac{C_i}{w_{ij}} \left[ \beta_j + \sum_{s=1}^J \beta_{js} \ln w_{ij} + \sum_{k=1}^K \beta_{jk} \ln x_{ik} + \sum_{r=1}^R \beta_{jr} z_{ir} \right] . \quad (5)$$

Differensiering av (5) gir:

$$\dot{L}_{ij} = \dot{C}_i - \dot{w}_{ij} + \frac{1}{S_{ij}} \left[ \sum_{s=1}^J \beta_{js} \dot{w}_{ij} + \sum_{k=1}^K \beta_{jk} \dot{x}_{ik} + \sum_{r=1}^R \beta_{jr} z_{ir} \dot{z}_{ir} \right] , \quad (6)$$

<sup>18</sup> Se appendiks A3 for utledning av den deriverte.

hvor prikk over variablene indikerer relativ endring (for eksempel  $\dot{x} = dx/x$ ) og (4) gir uttrykket for  $S_{ij}$ . Ved å erstatte endringer i totale kostnader i (6) med  $\sum_{s=1}^J S_{is} \dot{w}_{is} = \dot{C}_i$ , og omrokere resultatene, kan endringer i sysselsetting uttrykkes slik:

$$\dot{L}_{ij} = \left( \frac{\beta_{jj}}{S_{ij}} + S_{ij} - 1 \right) \dot{w}_{ij} + \sum_{s=1, s \neq j}^S \left( \frac{\beta_{js}}{S_{ij}} + S_{is} \right) \dot{w}_{is} + \frac{1}{S_{ij}} \left( \sum_{k=1}^K \beta_{jk} \dot{x}_{ik} + \sum_{r=1}^R \beta_{jr} z_{ir} \dot{z}_{ir} \right), \quad (7)$$

hvor  $j \neq s$ . Videre gir (7) etterspørselastisitetene, som gjelder i det alle andre faktorer holdes konstant.

$$\frac{\dot{L}_{ij}}{\dot{w}_{ij}} = \frac{\beta_{jj}}{S_{ij}} + S_{ij} - 1. \quad (8)$$

Egenpriselasiteten for hver variabel innsatsfaktor er gitt av (8). (9) gir de respektive etterspørselastisitetene som funksjoner av prisen på de andre variable innsatsfaktorene.

$$\frac{\dot{L}_{ij}}{\dot{w}_{is}} = \frac{\beta_{js}}{S_{ij}} + S_{is}, \text{ for } s \neq j. \quad (9)$$

Etterspørselen etter innsatsfaktorene som funksjoner av de strukturelle variablene er gitt av (10) og (11). Legg merke til at endringer i  $z$  er i prosentpoeng, mens andre endringer er i prosent.

$$\frac{\dot{L}_{ij}}{\dot{x}_{is}} = \frac{\beta_{jk}}{S_{ij}} \text{ og } \frac{\dot{L}_{ij}}{dz_{ir}} = \frac{\beta_{jr}}{S_{ij}}. \quad (10) \text{ og } (11)$$

I det jeg rapporterer resultatene fra regresjonene som gjennomføres i denne oppgaven, vil jeg også presentere elastisitetene utledet over. Dette hjelper på forståelsen av hvordan endringer i priser på innsatsfaktorer og endringer i strukturvariabler påvirker etterspørselen etter de ulike typene arbeidskraft, i det det er mulig å regne ut hvordan etterspørsel vil endre seg i *antall sysselsatte*.



### 3.2 Estimering med minste kvadrats metode

Minste kvadrats metode er et godt utgangspunkt for å analysere problemstillingen i denne oppgaven. Med formuleringen av kostnadsfunksjonen, som gir kostnadsandelene (4) som lineære funksjoner av strukturvariabler og lønnskostnader, kan regresjonsuttrykket for hver av kostnadsandelsfunksjonene generaliseres til  $j$  lineære funksjoner:

$$S_{itj} = \beta_j + \sum_{s=1}^4 \beta_{js} \ln w_{itj} + \sum_{k=1}^2 \beta_{jk} \ln x_{itk} + \sum_{r=1}^2 \beta_{jr} z_{itr} + e_{itj}. \quad (12)$$

Kostnadsandelene, betegnet med  $S$ , er det 4 av for hver næring  $i$ , i hver periode  $t$ . Det er i tillegg fire sett lønninger  $w$ , verdier for fast realkapital og verdiskapning,  $x$ , og kostnader på offshoring og forskning og utvikling som andeler av verdi på total produksjon,  $z$ , for hver næring i hver periode. De forskjellige  $\beta$ -ene er felles for alle næringene og konstante over tid. Etersom sammenhengen i (12) ikke holder for alle kombinasjoner av forklaringsvariablene, inkluderes et stokastisk feilledd, som inkluderer denne *støyen* som gjør at sammenhengene ikke holder. Det antas at alle utelatte, relevante variabler kan oppsummeres som tilfeldige, og inkluderes i dette feilleddet. Feilleddet har forventingsverdi lik null for alle verdier av forklaringsvariabler. I det videre blir forklaringsvariablene omtalt samlet i vektoren  $x_{it}$ , med de estimerte koeffisientene samlet i vektoren  $\beta$  og de avhengige variablene som  $y$ . Regresjonsuttrykket blir dermed, mer generelt, formulert som følger, for *hver* av de fire regresjonene, med  $\beta_j$  som konstantledd i de fire regresjonene omdøpt til  $\beta_0$ ;

$$y_{it} = \beta_0 + x_{it}\beta + e_{it}. \quad (13)$$

Det er i utgangspunktet fem antagelser som ligger til grunn for estimeringen av sammenhengen mellom den avhengige og de uavhengige variablene. Disse kalles Gauss-Markow-antagelsene, eller -vilkårene. Først og fremst må det kunne antas at den virkelige sammenhengen er *lineær* i parametrene. For at estimatene skal kunne ha gyldighet for, eller kunne beskrive, den sanne populasjonsregresjonen, må i tillegg utvalget estimatene er basert på være representativt for populasjonen. Dette innebærer at det må være et tilfeldig utvalg, og dermed ingen sammenheng mellom observasjonene. Videre må det være variasjon i

variablene, og ingen perfekt korrelasjon mellom dem. Det kan heller ikke være noen systematisk sammenheng mellom forklaringsvariablene og feilleddet, det vil si at for en hver  $x$  skal den forventede verdien av  $e$  være lik null. Variansen i feilleddet skal i tillegg være konstant for alle  $x$ , variansen skal ikke avhenge av størrelsen på forklaringsvariablene.

Antagelsene om feilleddet er sentrale i utledningen av egenskapene til estimatene fra minste kvadrats metode. Dersom Gauss-Markov-vilkårene er oppfylt, kan det vises at estimatorene oppnådd ved minste kvadrats metode er de beste som kan oppnås.<sup>19</sup>

Koeffisientene som estimeres i minste kvadrats metode kan uttrykkes slik:

$$\hat{\beta}_{\text{MKM}} = \frac{\sum_{i=1}^N \sum_{t=1}^T (x_{it} - \bar{x})(y_{it} - \bar{y})}{\sum_{i=1}^N \sum_{t=1}^T (x_{it} - \bar{x})^2}. \quad (14)$$

Hatten indikerer at verdien er *estimert* på basis av utvalgsstørrelser. Koeffisienten estimeres som funksjon av summen av avvik fra totalt gjennomsnitt, det vil si gjennomsnitt over tid og individ, for avhengig variabel  $y$  og uavhengig variabel  $x$ .<sup>20</sup>

### 3.3 Estimering med paneldatateknikker

Dersom det eksisterer informasjon om forhold for en rekke individer over en periode, kan datasettet kalles et paneldatasett. For å dra nytte av denne spesielle informasjonen, og for å få bukt med problemet med brudd på antagelsen om uavhengighet mellom observasjonene, som er viktig i vanlig minste kvadrats metode, er paneldatamodeller gode verktøy. De to paneldatamodellene som presenteres videre, er begge (noe mer) avanserte versjoner av minste kvadrats metode. Det som skiller disse to modellene fra minste kvadrats metode, er inkludering av individspesifikk heterogenitet, og det som skiller paneldatamodellene fra hverandre er hvilke antagelser som blir gjort om hvordan denne heterogeniteten arter seg.

<sup>19</sup> Best linear unbiased estimator; BLUE.

<sup>20</sup> Gjennomsnitt over  $T$  perioder og  $N$  individer:  $\bar{x} = \frac{\sum_{t=1}^T \sum_{i=1}^N x_{it}}{TN}$ , tilsvarende for  $y$

Notasjonsmessig skiller paneldatamodellene seg noe fra minste kvadrats metode. Som nevnt er det individspesifikk heterogenitet som er den sentrale forskjellen, og denne inkluderes som et ekstra feilledd i disse modellene,  $u_i$ :

$$y_{it} = \beta_0 + x_{it}\beta + u_i + e_{it}. \quad (15)$$

### 3.3.1 Fasteffektmodellen

I fasteffektmodellen tas det ikke spesielle antagelser om den individspesifikke heterogeniteten. Det som antas, er at denne komponenten er *ulik* for hvert individ og *konstant over tid*. Som i vanlig minste kvadrats metode antas  $\beta$  å være lik for alle individer, og konstant over tid. Det kan vises at fasteffektsregresjonen er lik minste kvadrats metode med *dummys* for alle individ i datasettet (unntatt ett), som gir individuelle konstantledd. Til sammenligning med (14) er fasteffektestimatorene ( $\beta_{FE}$ ) gitt ved:<sup>21</sup>

$$\hat{\beta}_{FE} = \frac{\sum_{i=1}^N \sum_{t=1}^T (x_{it} - \bar{x}_i)(y_{it} - \bar{y}_i)}{\sum_{i=1}^N \sum_{t=1}^T (x_{it} - \bar{x}_i)^2}. \quad (16)$$

Som det kommer frem av (16) er estimatoren basert kun på individenes variasjon over tid, eller variasjon *innen* individene, i motsetning til estimatoren fra minste kvadrats metode, som utnytter variasjon både over individer og over tid. I tilfellet med minste kvadrats metode med dummys, vil de derav følgende konstantleddene fange opp forskjeller i gjennomsnittsverdier, og  $\beta$  blir lik ved de to estimeringsmetodene. Etersom estimatoren er basert på avvik fra individuelle gjennomsnitt over tid, må det i alle fall for noen individer være variasjon i forklaringsvariablene over tid for at det skal være mulig å estimere effekten av forskjeller i variabelen.

Det vanlige feilleddet i fasteffektmodellen ( $e_{it}$ ) antas å oppfylle samme vilkår som i minste kvadrats metode, og estimatoren er forventingsrett dersom feilleddet ikke er korrelert med

---

<sup>21</sup> Gjennomsnitt over T perioder:  $\bar{x}_i = \frac{\sum_{t=1}^T x_{it}}{T}$ , tilsvarende for y.

forklaringsvariabler i samme periode. Fasteffektestimatorene er konsistent (både for høy N og høy T) dersom kravet om at feilleddet har forventet verdi lik null gitt  $x$  i *alle* perioder:

$$E\{x_{it}e_{is}\} = 0 \text{ for alle } s, t.$$

Dette innebærer at forklaringsvariablene må være eksogene, noe som utelukker bruk av *laggede* avhengige variabler som forklaringsvariabel. I tillegg er det verdt å nevne at modellen ikke skiller mellom endringer i forklaringsvariabler på grunn av forskjell mellom periodene, eller mellom individer.

Når fasteffektmodellen estimeres i STATA, oppgir programmet et konstantledd. Verdien på konstantleddet er lik gjennomsnittet til de individspesifikke fasteffektene.<sup>22</sup>

### 3.3.2 Tilfeldigeffektmodellen

Tilfeldigeffektmodellen baseres på en annen antagelse om den individspesifikke heterogeniteten. I denne modellen antas komponenten å være en del av feilleddet. Det vil si at, som i vanlig minste kvadrats metode, ikke-inkluderte, relevante variabler antas å kunne oppsummeres som et tilfeldig feilledd. I tilfeldigeffektmodellen deles feilleddet dermed opp i to deler. Individspesifikk heterogenitet som er konstant over tid, antas å i gjennomsnitt være lik null, med konstant varians, tilsvarende den generelle delen av feilleddet, som er felles i de tre modellene nevnt så langt.

Tilfeldigeffektmodellen må dermed inkludere variasjon mellom individene i uttrykket for estimatoren ( $\beta_{RE}$ ). Det kan vises at  $\beta_{RE}$  er gitt ved:

$$\hat{\beta}_{RE} = \frac{\sum_{i=1}^N \sum_{t=1}^T (x_{it} - \bar{x}_i)(y_{it} - \bar{y}_i) + \psi T \sum_{i=1}^N (\bar{x}_i - \bar{\bar{x}})(\bar{y}_i - \bar{\bar{y}})}{\sum_{i=1}^N \sum_{t=1}^T (x_{it} - \bar{x}_i)^2 + \psi T \sum_{i=1}^N (\bar{x}_i - \bar{\bar{x}})^2}. \quad (17)$$

Estimatoren fra tilfeldigeffektmodellen er basert på et *effisient* vektet gjennomsnitt av variasjon innen individene og mellom individene (Verbeek, 2004).<sup>23</sup>  $\psi$  er funksjon av

<sup>22</sup> <http://www.stata.com/support/faqs/stat/xtreg2.html>

variansene i det individspesifikke, tidskonstante feilleddet og det vanlige feilleddet.<sup>24</sup> Dersom  $T$  går mot uendelig, det vil si at det er veldig mange tidsperioder i datasettet, vil  $\psi$  gå mot null, og  $\beta_{RE} = \beta_{FE}$ . Dersom  $\psi$  er lik en, vil  $\beta_{RE}$  være lik estimatoren fra minste kvadrats metode.  $\psi$  kan innta verdien en dersom variansen i individspesifikke fasteffekter er lik null. Dersom dette er tilfellet, vil det ikke være noen variasjon mellom individene, og ikke nødvendig å estimere en mer avansert modell enn minste kvadrats metode.

Forutsetningen om at forventingsverdien til individuelle gjennomsnitt i regressor gitt størrelse på individspesifikk fasteffekt, alltid skal være lik null, er sentral for at resultatene i tilfeldigeffektmodellen skal gi konsistente estimater. Med sammenheng mellom regressor og individspesifikk heterogenitet vil tilfeldigeffektmodellen gi inkonsistente estimater. Modellen vil være mer effisient enn minste kvadrats metode og fasteffektmodellen dersom det ikke er slik sammenheng. Dette kommer av at tilfeldigeffektmodellen inkluderer både variasjon mellom og innen individene, ved et vektet gjennomsnitt av de to, denne vektingen er basert på variansen til hver del av feilleddene.

### 3.3.3 Hausmantesten

Avveilingen om det er fasteffektmodellen eller tilfeldigeffektmodellen som skal brukes er ikke alltid lett. Fasteffektmodellen kan brukes, den gir forventingsrette resultater, både dersom individspesifikk heterogenitet ikke henger sammen med forklaringsvariablene og dersom den gjøre det. Tilfeldigeffektmodellen gir i det første tilfellet mer effisiente resultater enn fasteffektmodellen, men gir i det andre tilfellet inkonsistente resultater. Dette gjør at fasteffektestimatorene vil være den mer robuste av dem. I tillegg til dette vil det være slik at om fasteffektmodellen skal brukes, må det være (i alle fall noe) individuell variasjon i forklaringsvariablene for at sammenhenger skal kunne estimeres. Dersom dummyer er viktige forklaringsvariabler, vil altså tilfeldigeffektmodellen være å foretrekke. Videre vil fasteffektmodellen være et godt valg dersom kvantifisering av individspesifikk heterogenitet er interessant, eller dersom utvalget ikke kan ses på som tilfeldig trukket, dersom individene for eksempel er land, næringer, eller tilsvarende enheter som kan ses på som *one of a kind*

---

<sup>23</sup> Med fasteffektmodellen for variasjon *innen individene* og den såkalte betweenmodellen (fra modellen  $\bar{y}_i = \mu + \bar{x}_i'\beta + u_i + \bar{\epsilon}_i$ , regresjon for *gjennomsnittene* til variablene) for variasjonen *mellom individene*.

<sup>24</sup> 
$$\psi = \frac{\sigma_e^2}{\sigma_e^2 + T\sigma_u^2}$$

(Verbeek, 2004, side 351). Fasteffektmodellen estimerer da sammenhenger mellom regressor og regressant *gitt* de individspesifikke forskjellene. Hvis ikke noen av disse er tilfellet, er konklusjonen intuitivt uklar. Hausmantesten kan brukes for å finne ut hvilken modell som bør foretrekkes. Nullhypotesen i denne testen er at den individspesifikke komponenten ikke er korrelert med regressor. Estimatorene fra de to modellene sammenlignes, og store forskjeller indikerer at nullhypotesen ikke holder, ettersom de to modellene da gir like koeffisienter. Testobservatoren fra denne testen kan formuleres som en funksjon av estimerte koeffisienter og variansen i disse:

$$\tau_H = (\hat{\beta}_{FE} - \hat{\beta}_{RE})' [\hat{V}(\hat{\beta}_{FE}) - \hat{V}(\hat{\beta}_{RE})]^{-1} (\hat{\beta}_{FE} - \hat{\beta}_{RE}). \quad (18)$$

Hatt over  $\beta$  og  $V$ , variansene til  $\beta$ -koeffisientene, indikerer estimerte størrelser. Dersom forskjellene mellom estimatorene er tilstrekkelig store i forhold til forskjellene i varianser, vil testobservatoren, som har en asymptotisk chikvadratfordeling med  $K$  (antall forklaringsvariabler i modellen) frihetsgrader, være høyere enn kritisk verdi, og nullhypotesen kan forkastes, og fasteffektmodellen bør foretrekkes.

En forutsetning for at testobservatoren skal kunne bli regnet ut, er at kovariansmatrisene i (18) er positiv definitte, dette er fordi det må til for å kunne regne ut de inverse av kovariansmatrisene. Det kan være andre grunner til at koeffisientene skiller seg fra hverandre, og slike problemer er ikke uvanlige, noe som gjør at validiteten av hausmantesten skal derfor ikke alltid tas for gitt.

### **3.4 Problemer med modellene, og valg av riktig modell**

Som nevnt er de ulike regresjonsteknikkene basert på antagelser om hvordan data er og hvordan modeller oppfører seg. Dette innebærer at noen av modellene er bedre egnet enn andre gitt tilgjengelige data, og hvilke antagelser som eventuelt er brutt. Som tidligere nevnt er implikasjonene av brutte antagelser forskjellige ettersom hvilken antagelse som er brutt. Det viktigste med modellen er å få forventingsrette estimatorene.

Endogenitet er en kilde til problemer med forventningskjevne resultater. Det er primært tre måter dette kan oppstå på. Dersom det er *målefeil* i data, som gjør at det er systematiske avvik

i verdiene på variablene, vil estimerte sammenhenger ikke være de sammenhengene som er ønsket. Endogenitet kan oppstå som følge av *simultanitet*, i det variabler i modellen er bestemt samtidig og gjensidig med avhengig variabel. En annen måte endogenitet kan oppstå på, er dersom det er utelatte variabler som er korrelert med regressor. Med utelatt menes en variabel som er med å forklare avhengig variabel, men ikke inkludert i modellen. Slike problemer er det viktig å gjøre noe med. Det kan vises at fasteffektmodellen er mer robust mot slike problemer enn tilfeldigeffektmodellen og minste kvadrats metode, ettersom utelatte variabler vil inkluderes i det individspesifikke feilleddet, som det igjen er tillatt å korrelere med regressor.

Andre økonometriske problemer fører ofte til *inferensproblemer*, som innebærer at testing av koeffisienter ikke er til å stole på. Multikolaritet, som innebærer at det er sterk korrelasjon mellom noen av forklaringsvariablene, fører til inferensproblemer. Dette fordi det er vanskelig å si nøyaktig hvilken av variablene som påvirker den avhengige variabelen. Dette fører til større standardfeil, og i noen tilfeller uventede størrelser og fortegn på estimerte koeffisienter. Antagelsen om konstant varians i feilleddene, og uavhengighet mellom feilledd og forklaringsvariabler er sentral når det gjelder inferensproblemer. Heteroskedastisitet og autokorrelasjon er problemer som bryter med denne antagelsen, og som også gir inferensproblemer. Under presenterer jeg to modeller som bøter på disse problemene, og når resultatene presenteres kommer jeg igjen nærmere inn på disse problemene. Det er også ønskelig med normalfordelte feilledd, slik at også estimatorene blir normalfordelte, og testobservatorer blir fordelt med funksjoner av normalfordelingen.

### **3.5 Tilsynelatende urelaterte regresjoner (SUR)**

Ettersom det er fire kostnadsandeler som skal regresseres, og disse fire andelene henger sammen, ved at de i *sum er 1*, kan det vises at feilleddene i regresjonene er korrelerte. I litteraturen henvises det til Zellners metode for *seemingly unrelated regressions*, eller tilsynelatende urelaterte regresjoner. Denne metoden for estimering av flere regresjoner samtidig krever her at en ligning utelates, fordi kovariansmatrisen vil bli singular dersom alle fire regresjonene inkluderes samtidig. Dette følger av at når ligningene summeres til 1, vil en ligning være avhengig av de andre, og det er dermed en lineær avhengighet.

I matrisenotasjon kan de tre gjenværende regresjonene presenteres slik:

$$\begin{bmatrix} y_{1it} \\ y_{2it} \\ y_{3it} \end{bmatrix} = \begin{bmatrix} x_{1it} & 0 & 0 \\ 0 & x_{2it} & 0 \\ 0 & 0 & x_{3it} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} e_{1it} \\ e_{2it} \\ e_{3it} \end{bmatrix}, \quad (19)$$

hvor argumentene i (19) er vektorer og matriser som i (13), bortsett fra at  $\beta_0$  er inkludert i vektoren  $\beta$ , og at  $y_{it}$ ,  $e_{it}$  og  $x_{it}$  er samlet i vektorene  $y$  og  $e$  og matrisen  $X$ :

$$y = X\beta + e. \quad (20)$$

Når feilleddene i et sett regresjoner er korrelerte mellom regresjonene, bør modellen estimeres i SUR-rammeverket med en, eller annen, form for minste kvadrats metode.<sup>25</sup> Fra (14) har vi, i matrisenotasjon, et uttrykk for den estimerte koeffisienten, for vanlig minste kvadrats metode:

$$\hat{\beta}_{MKM} = (X'X)^{-1} X'y.$$

Til sammenligning er tilsvarende Zellners SUR-modell-estimator slik:

$$\hat{\beta}_{SUR} = (X'W^{-1}X)^{-1} X'W^{-1}y. \quad (21)$$

Feilleddenes varians-kovarians-matrise som er inkludert i (21) er gitt ved;

$$W = E[ee'] = \begin{bmatrix} \sigma_1^2 I_T & \sigma_{12} I_T & \sigma_{13} I_T \\ \sigma_{21} I_T & \sigma_2^2 I_T & \sigma_{23} I_T \\ \sigma_{31} I_T & \sigma_{32} I_T & \sigma_3^2 I_T \end{bmatrix}, \quad (22)$$

med  $\sigma_{ij} = \sigma_{ji}$ , for  $i \neq j$ . (22) viser antagelsen om at feilleddene er korrelert mellom de tre forskjellige regresjonene. Identitetsmatrisen  $I_T$  er en diagonalmatrise som viser at

---

<sup>25</sup> For eksempel fasteffektmodellen og tilfeldigeffektmodellen. I STATA er det greit å bruke SUR på fasteffektmodellen ettersom det er lett å uttrykke denne modellen i "minste kvadrats metode-terminologi" (med dummyer). Det er nok også mulig å implementere tilfeldigeffektmodellen, eller andre generaliserte minste kvadrats metoder, ved å uttrykke dem som minste kvadrats metode med manuelle transformasjoner.



korrelasjonen er antatt å bare eksistere innen samme tidsperiode. SUR-modellen, (21), inkluderer altså varians-kovariansmatrisen i argumentene, på en lignende måte som vanlig generalisert minste kvadrats metode, og tilfeldigeffektmodellen, inkluderer uttrykk for varians. Griffiths *et al.* (1993) skriver at estimatoren fra SUR-modellen er lik estimatoren fra minste kvadrats metode dersom  $x_{1it} = x_{2it} = x_{3it}$ , noe som også er tilfellet i denne oppgaven. Ettersom de tre kostnadsfunksjonene som estimeres inneholder nøyaktig samme forklaringsvariabler. Dette innebærer at estimering av regresjonene i SUR-modellen fører til samme koeffisienter som estimering i minste kvadrats metode. Det som derimot er fordelen med SUR-modellen, er at testing av hypoteser om sammenhengene mellom koeffisientene i de ulike delene av modellen blir mer effisiente.

Avgjørelsen om bruk av denne metoden er nødvendig tas på bakgrunn av en test av estimerte kovarianser i samme perioder, med nullhypotesen at det ikke er kovarians mellom de ulike delene av modellen;  $H_0: \sigma_{12} = \sigma_{13} = \sigma_{23} = 0$ , og alternativhypotesen at det er kovarians mellom i alle fall et av parene;  $H_1: \text{minst en av } \sigma_{ij} \neq 0, i \neq j$ . Ifølge Griffiths *et al.* (1993) er testobservatoren asymptotisk chikvadratfordelt, med  $M(M-1)/2$  frihetsgrader ( $M$  er antall ligninger). Noe som i dette tilfellet gir tre frihetsgrader. Gitt at testobservatoren er høyere enn kritisk verdi, kan nullhypotesen om ingen samtidig korrelasjon forkastes, og SUR-modellen bør benyttes. STATA rapporterer en slik test etter estimering av SUR-modellen. Testen blir da referert til som *Breusch-Pagans* uavhengighetstest.<sup>26</sup>

For å teste hypotesen om at koeffisientene for forskjellige gruppene er statistisk signifikant forskjellige fra null samtidig, som i tilfellet kostnadsandeler impliserer ulik effekt av for eksempel offshoring på de forskjellige arbeidstakergruppene, brukes en Wald-test. Nullhypotesen er dermed at en eller flere av settene koeffisienter er like for alle delene i modellen.<sup>27</sup> Med utgangspunkt i denne nullhypotesen kan en testobservator kalkuleres. Høye verdier for testobservatoren ( $F$ ) kalkulert under antagelsen om at nullhypotesen holder, regnes som bevis *mot* nullhypotesen. Fra generell testing av lineære hypoteser i regresjonsanalyse basert på minste kvadrats metode, kommer formuleringen av null- og alternativhypotesene:

$$H_0 : R\beta = r \quad \text{og} \quad H_1 : R\beta \neq r, \quad (23)$$

<sup>26</sup> *Breusch-Pagan test of independence.*

<sup>27</sup> For eksempel  $\beta_{1\text{offshoring}} = \beta_{2\text{offshoring}} = \beta_{3\text{offshoring}}$ , tilsvarende  $\beta_{jr}$  fra (4).

med  $J$  antall lineære hypoteser.  $R$  er en  $(J \times K)$  matrise, hvor radene definerer lineære kombinasjoner av elementene i  $\beta$ , og  $r$  er en  $(J \times 1)$  vektor som presiserer antagelsene i hypotesene. For å understreke hvordan dette ser ut dersom effekten av endringer i en variabel er forskjellig i tre regresjoner, vil  $H_0$  i (23) se slik ut:

$$\beta_{12} - \beta_{22} = 0$$

$$\beta_{12} - \beta_{32} = 0.$$

Når  $H_0$  er sann, vil  $R\beta$  være normalfordelt med gjennomsnitt  $r$  og varians  $R\text{cov}(\beta)R'$ :

$$R\hat{\beta} \sim N(r, RCR').$$

Med  $\beta$ s varians-kovariansmatrise gitt ved  $C = \text{cov}(\hat{\beta}) = (X'WX)^{-1}$ . Den "sanne" testobservatoren kan dermed formuleres som:

$$g = (R\hat{\beta} - r)'(RCR')^{-1}(R\hat{\beta} - r) \sim \chi_{(J)}^2. \quad (24)$$

Ettersom kovariansmatrisen  $C$ , og varians-kovariansmatrisen  $W$  som inngår i  $C$ , ikke er kjent, må den erstattes av estimert  $W$ , som baseres på kovarians og varians fra estimatorene fra minste kvadrats metode. Dette gir en ny versjon av (24), hvor hatter indikerer at verdiene er basert på estimerte størrelser:

$$\hat{g} = (R\hat{\beta} - r)'(R\hat{C}R')^{-1}(R\hat{\beta} - r) \xrightarrow{d} \chi_{(J)}^2. \quad (25)$$

Denne estimerte testobservatoren er chikvadratfordelt med  $J$  frihetsgrader, og er et asymptotisk resultat (Griffiths *et al.*, 1993). Denne testobservatoren kan regnes ut i STATA som postkommando etter estimering av en SUR-modell, eller ved bruk av SUEST kommandoen, som samler estimerte regresjoner, og gjør det mulig å utføre tester som nevnt over.

Selv om denne presentasjonen er basert på sammenligning mellom vanlig minste kvadrats metode, er det som tidligere nevnt fullt mulig å kjøre mer avanserte modeller i dette rammeverket, som for eksempel fasteffektmodellen eller versjoner av generalisert minste kvadrats metode.

### **3.6 Estimering med dynamiske modeller**

Det er også aktuelt at spesifiseringen basert på kostnadsfunksjonen ikke tar høyde for alle de prosessene som er med å bestemme verdiene av kostnadsandelene. Det er klart at endringen i kostnadsandelene mellom to perioder ikke er tilfeldig. Næringene bruker tid på omstilling, og dersom noe endrer seg, vil ikke effekten slå fullt ut med en gang. Slike rigiditeter kan gi seg utslag i at kostnadsandelene er sterkt korrelert over tid. Det er mulig å estimere dynamiske paneldatamodeller i STATA, med modeller med en autoregressiv prosess som et godt alternativ. Ifølge Veerbek (2004, side 360) er muligheten til å modellere individuell dynamikk en *unik styrke ved paneldata*.

#### **3.6.1 Autoregressiv prosess**

Generelt kan det sies at dersom sammenhengen mellom verdiene for en variabel i to påfølgende perioder er sterkt korrelert vil positive avvik i en periode, det vil si positivt feilledd, føre til positive avvik i neste. En måte å undersøke tilstedeværelsen av dynamikk er å fokusere på feilleddene i den statiske modellen. Etter å ha estimert en statisk modell slik som de over, i et datasett hvor det er observasjoner over flere perioder, er det aktuelt å undersøke om det er sammenhenger i data mellom periodene. En måte å gjøre dette på er å sjekke om det er korrelasjon mellom feilleddene mellom periodene. Dersom dette er tilfellet kan det være snakk om at *autokorrelasjon* er et problem.

$$y_{it} = \beta_0 + \beta x_{it} + u_i + e_{it}. \quad (26)$$

Ved estimering av modellen over, er det mulig å undersøke om det eksisterer en sammenheng mellom feilleddene i forskjellige perioder. Dersom positive avvik i en periode følges av positive avvik i neste periode, betegnes det som positiv autokorrelasjon, og i motsatt tilfelle, positive avvik følges av negative avvik, betegnes det som negativ autokorrelasjon.

$$e_{it} = \rho e_{it-1} + v_{it}. \quad (27)$$

For å teste om autokorrelasjon er et problem i modellen kan det tas utgangspunkt i antagelsen om at det ikke eksisterer autokorrelasjon, for så å utlede en testobservator som gir høye verdier dersom antagelsen (nullhypotesen) ikke holder. Testen utføres på estimerte feilledd fra den aktuelle modellen. Dette indikeres med hatt over feilleddene,  $\hat{e}$ . Dessverre er ikke mulig å definere en bestemt kritisk verdi for det Durbin-Watson-testen, men det eksisterer veiledende kritiske verdier som gir intervaller for kritisk verdi, hvor konklusjonen er usikker, men med testobservator under 1,8 vil generelt gi konklusjon om forekomst av autokorrelasjon.

$$dw_{\rho} = \frac{\sum_{i=1}^N \sum_{t=2}^T (\hat{e}_{it} - \hat{e}_{i,t-1})^2}{\sum_{i=1}^N \sum_{t=1}^T \hat{e}_{it}^2}. \quad (28)$$

Dersom testing gir grunn til å forkaste nullhypotesen om fravær av autokorrelasjon, vil det være naturlig å omformulere det opprinnelige regresjonsuttrykket slik at dette problemet blir tatt høyde for. I denne sammenhengen finnes det metoder som estimerer modeller med såkalte autoregressive prosesser. I utgangspunktet kan slike prosesser være av forskjellig orden, i den forstand at sammenhengen mellom feilleddene kan eksistere over flere perioder. I STATA er flere tester tilgjengelige, som tester for autokorrelasjon av høyere orden, og også modifiserte Durbin-Watson-tester, blant andre Baltagi-Wu-LBI test, og tester i analysen nevnt over.<sup>28</sup> Jeg tar for meg situasjonen hvor det bare er sammenheng mellom to *påfølgende* perioder. Dette er en autoregressiv prosess av første orden, med sammenheng som i (27).

STATA har inkorporert en metode som estimerer paneldatamodellene, også når det er problemer med autokorrelasjon, ved antagelsen om at modellen er gitt ved (26), med en prosess i feilleddet som beskrevet av (27).

### 3.6.2 Arellano-Bond-estimatoren

En annen innfallsvinkel er å se på verdien av forklaringsvariabelen i en periode som  $\Delta y_{it}$ , i alle fall delvis, bli forklart av verdien i den foregående. Blant annet fordi omstillingen næringene må gjennom som følge av endringer i noen av forklaringsvariablene tar tid, er verdien i en

<sup>28</sup> Baltagi-Wu locally best invariant test for om  $\rho = 0$ . I tillegg rapporteres Arellano-Bond test for autokorrelasjon ved estimering av dynamisk modell med kommandoen *xtabond* i STATA.

periode en god indikator på hva den vil være i neste, og kan inkluderes som en forklaringsvariabel i regresjonsmodellen. Dette kan mer spesifikt formuleres slik:

$$y_{it} = \beta_0 + \gamma y_{it-1} + x_{it}\beta + e_{it}. \quad (29)$$

Dette uttrykket indikerer at det er de allerede kjente forklaringsvariablene sammen med størrelsen på samme avhengige variabel i perioden før som bestemmer størrelsen på avhengig variabel i denne periode  $t$ .

Verbeek (2004) viser at estimering av (29) vil gi forventningskjevne estimater for  $\gamma$  ved fast  $T$ , dersom vanlig fasteffektsmodell brukes. Han foreslår videre en metode for å estimere modeller som i (29), med en autoregressiv prosess av første orden. Denne metoden tar utgangspunkt i differensiering av uttrykket over, for så estimere sammenhenger ved bruk av en instrumentvariabelmetode, basert på arbeider av Arellano og Bond, for å oppnå et estimat for  $\gamma$ . Instrumentene blir oppnådd ved antagelser om fravær av autokorrelasjon av høyere orden (enn 1), som gir momentrestriksjoner som kan brukes til å danne instrumentvariabler til bruk i estimeringen.

## 4 Presentasjon av datasettet

Presentasjonen av datasettet er delt inn i tre; hvor data er samlet inn og hvordan de er organisert, så defineres variablene. I den tredje delen presenteres deskriptiv statistikk.

### 4.1 Datakilder

Datasettet består av data for Norge i perioden fra 1992 til 2005. Alle data stammer fra SSB, og er primært delt opp i to grupper; inntekts- og sysselsettingsdata, og andre næringsdata, alt på næringsnivå etter standardgrupperingen NACE rev. 1 (Nomenclature générale des Activités économiques dans les Communautés Européenes). Totalt er det i datasettet tall for 57 næringer, opplistet i appendikset. De andre næringsdataene omfatter hovedforklaringsvariabelen offshoring, kapitalstokk, verdiskapning og produksjon, og er hentet fra *kryssløpstabeller* publisert av SSB. De avhengige variablene i den økonometriske analysen er dannet på bakgrunn av inntekts- og sysselsettingsdata, og de uavhengige variablene er inntekt og andre data på næringsnivå.

#### 4.1.1 Næringsinndeling

Datasettet er en samling data for norske næringer i tidsrommet 1992 til 2005. Næringene er som nevnt delt inn etter NACE-kode-systemet, som, ifølge SSB, danner ”*grunnlag for koding av enheter etter viktigste aktivitet i Statistisk sentralbyrås bedrifts- og foretakregister og for enheter i Enhetsregisteret*”.<sup>29</sup> Dette innebærer at bedrifter i Norge blir gruppert i næringer etter hovedaktivitet. Det er totalt 60 næringer på dette nivået, men i denne oppgaven er det kun 51. Grunnen til at noen næringer er utelatt, er hovedsakelig manglende data.

I tillegg kan det nevnes at næringer med få ansatte i de forskjellige utdanningsgruppene vil oppleve store relative endringer i størrelsene nevnt selv med en ny ansettelse. Næringene med NACE nummer 10 og 95 (henholdsvis næringen bryting av steinkull og brunkull og næringen private husholdninger med ansatte personer) er spesielle i denne sammenhengen, med spesielt få ansatte med høyere universitetsutdanning. Med ingen import av innsatsvarer fra tilsvarende næringer i utlandet, og noen ekstremobservasjoner i datasettet konkluderer jeg med at det er

---

<sup>29</sup> <http://www3.ssb.no/stabas/ItemsFrames.asp?ID=3152101&Language=nb&VersionLevel=ClassVersion>

hensiktsmessig å ekskludere disse næringene fra den videre analysen. Samme skjebne lider NACE-næringene 1, 2 og 5,<sup>30</sup> på grunn av manglende inntekts- og sysselsettingstall. Næring 12 (utvinning av uran etc.) finnes ikke i Norge og næring 99 (internasjonale organisasjoner) er ikke inkludert i de norske kryssløpstabellene, og er dermed heller ikke inkludert i analysen. I tillegg bør det nevnes at næringene 15 og 16 (nærings- og nytelsesmiddelindustri og tobakksindustri) og 23 og 24 (petroleumsindustrien og kjemisk industri) er slått sammen ettersom tobakks- og petroleumsindustriene fra 2001 ble konfidensielle, og slått sammen med henholdsvis industri 15 og 24 i kryssløpstabellene fra SSB. En fullstendig liste over næringer som er med i datasettet finnes i appendikset.

I noen av regresjonene kan det være aktuelt å slå sammen næringene i grupper alt ettersom hvordan jeg bruker forklaringsvariablene. Typisk vil dette være aktuelt i det jeg undersøker hvordan alternative mål på forskning og utvikling påvirker etterspørselen etter arbeidskraft. I de tilfellene dette skjer vil det uttrykkelig på bli gjort merke på det i teksten.

#### 4.1.2 Utdanningsnivå

Oppgavens hovedproblemstilling er å finne ut om outsourcing, eller import av innsatsvarer, har påvirket etterspørselen etter heterogen arbeidskraft. Med heterogen arbeidskraft forstås i denne sammenheng arbeidskraft med ulik grad av kompetanse. Arbeidskraften er delt inn i *fire* utdanningsnivå, noe som skiller seg noe fra eksisterende litteratur. Som nevnt før blir det vanligvis skilt mellom to grupper, *skilled* og *unskilled* (eventuelt *white* og *blue collar*) *workers*. Dette er spesielt vanlig i undersøkelser av det amerikanske arbeidsmarkedet, se for eksempel Feenstra og Hansson (1996), mens europeiske studier ofte deler inn arbeidskraften i opptil tre-fire nivå. I de artiklene som denne oppgaven er basert på kan det nevnes at Ekholm og Hakkala (2005) bruker tre nivåer basert på utdanning, Hijzen *et al.* (2005) bruker tre nivåer basert på stillingstype.

Utdanningsnivåene er utledet fra *norsk standard for utdanningsgruppering*, som inneholder svært detaljert informasjon om utdanning. Ettersom jeg ønsker å skille mellom bare fire utdanningsgrupper, holder det med informasjon på såkalt 1-siffer nivå. Disse fire utdanningsgruppene jeg har definert omfatter henholdsvis grunnskoleutdanning, videregående

---

<sup>30</sup> Henholdsvis jordbruks-, skogsbruks- og fiskerinæringene.

utdanning, laveregrads universitets- og høyskoleutdanning og høyeregrads universitets- og høyskoleutdanning, se Tabell 1 for nærmere presisering.

**Tabell 1: Inndeling av arbeidskraft etter høyeste gjennomførte utdanning.**

<i>Firedeling</i>	<i>Utd.nivå</i>	<i>Beskrivelse</i>	<i>Klassetrinn</i>
Gruppe 1	0	Ingen utdanning eller førskoleutdanning	Ingen
	1	Barneskoleutdanning	1-7
	2	Ungdomsskoleutdanning	8-10
Gruppe 2	3	Videregående, grunnutdanning	11-12
	4	Videregående, avsluttende utdanning	13+
	5	Påbygging til videregående utdanning	14+
Gruppe 3	6	Lavere universitetsutdanning	14-17
Gruppe 4	7	Høyere universitetsutdanning	18-19
	8	Forskerutdanning	20+

Kilde: SSB (norsk standard for utdanningsgruppering), med egen firedeling.

## **4.2 Variabeldefinisjoner**

Med fire kostnadsandeler som skal regresseres blir det fire avhengige variabler. Disse regresseres på seks uavhengige variabler.

### **4.2.1 De avhengige variablene**

#### Kostnadsandelene

På grunnlag av egenskapene presentert i forbindelse med bruken av translog-kostnadsfunksjoner, er en innsatsfaktors andel av totale variable kostnader godt egnet som avhengig variabel når en skal undersøke hvordan etterspørselen etter en innsatsfaktor er påvirket av aktuelle variabler. I denne forbindelse er det etterspørselen etter forskjellige typer arbeidskraft som er aktuelt, og de forskjellige typene arbeidskrafts kostnadsandeler som er de avhengige variablene.

Med tilgjengelige data på inntekt og sysselsetting etter utdanningsnivå for næringene i perioden 1996 til 2005, har jeg hatt muligheten til å generere de respektive kostnadsandelene etter formelen;



$$S_{ijt} = \frac{w_{ijt} L_{ijt}}{\sum_{i=1}^4 w_{ijt} L_{ijt}},$$

med  $j=1,2,3,4$  som de fire utdanningsnivåene,  $i=1,\dots,51$  næringer og  $t = 1996,\dots,2005$  som tidsindikatorer.

Sysselsettingstallene ( $L_{ijt}$ ) er antall menn med positiv pensjonsgivende inntekt i de respektive næringene, med inntekt ( $w_{ijt}$ ) som den gjennomsnittlige årlige pensjonsgivende inntekten disse mennene fikk utbetalt. Tallene er oppgitt i løpende norske kroner.

## 4.2.2 De uavhengige variablene

### Inntekt

Inntektsdata opptrer også på høyresiden av regresjonsligningene. Som nevnt over er dette gjennomsnittlig årlig pensjonsgivende inntekt for menn i de respektive næringene for årene fra 1996 til 2005 oppgitt i løpende norske kroner.

### Fast realkapital

Fast realkapital kommer fra SSBs kryssløpstabeller, hvor det heter *fixed capital stock*, og er oppgitt for årene 1992 til 2000 og 2002. For årene 2001 og 2003 til 2005 er det, grunnet ny standard for kryssløpstabeller, ikke oppgitt. Begrepet omfatter alle typer varig kapital som er produsert og kan benyttes gjentatte ganger.

Tallene er oppgitt i løpende millioner kroner

### Produksjon

Mål på næringens totale produksjon, verdien på varer og tjenester fra innenlands produksjonsaktivitet. Dette er inkludert varer og tjenester som blir forbrukt innen bedriftene, og er ikke det samme som verdien på salg av varer og tjenester. Tallene er oppgitt i løpende millioner norske kroner, og er inkludert eventuelle produktsubsidier men ikke merverdiavgift og andre skatter.

### Verdiskapning/bruttoprodukt

Denne variabelen omfatter forskjellen mellom produksjon og produktinnsats. På samme måte som med produksjon, er eventuelle subsidier inkludert, men skatter og avgifter ekskludert. Oppgitt i løpende millioner norske kroner.

### Smalimport

Smalimport er det som i den engelskspråklige litteraturen kalles *narrow offshoring*, introdusert av Feenstre og Hansson (1996). Det er forholdet mellom importerte innsatsvarer fra samme næring i utlandet og total produksjon. Informasjonen kommer fra SSBs kryssløpstabeller, som ikke inneholder informasjon om hvilken næring den eksporterende bedriften i utlandet hører til, men informasjonen er knyttet til varen. Det er ligger dermed en antagelse til grunn i utarbeidelsen av kryssløpstabellene om at kombinasjonen vare - næring er den samme i utlandet som i Norge. Dette stemmer nok ikke alltid, men spørsmålet er om det er grunn til å anta at det er systematiske avvik fra denne antagelsen, og ettersom det mest sannsynlig ikke er det skal det ikke være noe problem å bruke disse dataene.

I litteraturen på dette området er det en del ulike definisjoner på denne variabelen, men denne er i tråd med Ekholm og Hakkala (2005). Andre velger å måle importen relativt til total innsatsvarebruk eller verdiskapning. Poenget er uansett å ta høyde for endringer i import i forhold til produksjon, verdiskapning eller innsatsvarebruk generelt, med de tre siste som samvarierende mål på næringens størrelse. Tallene rapporteres som andeler, *relsmal*, og ligger dermed mellom 0 og 1.

Det er også tall i datasettet på såkalt bredimport (*broad offshoring*). Som inkluderer alle importerte innsatsvarer, vektet mot produksjon på samme måte som for smalimport.

### Total73

Som mål på bruk av forsknings- og utviklingstjenester og -varer bruker jeg næringenes bruk av innsatsvarer fra næring NACE 73, Forsknings- og utviklingstjenester, hentet fra SSBs kryssløpstabeller. Tallene er opprinnelig oppgitt i millioner løpende kroner, men som med de foregående variablene bruker jeg tallene i forhold til produksjon, altså blir variabelen *reltot73* et tall mellom 0 og 1.

### Forskjellige mål på FoU

SSB publiserer også tall på forskning og utvikling (fou) utført i næringene. Tallene her er noe grovere aggregert enn tallene fra kryssløpstabellene, og heller ikke for like mange år. Tallene er for annethvert år fra 1995 til 2001, og så årlig frem til 2005. Tallene som er oppgitt er kostnader forbundet med egenutført fou, driftskostnader av fou, investeringskostnader for fou og innkjøpte fou-tjenester, alle oppgitt i millioner løpende kroner. Ettersom det er et annet aggregeringsnivå på disse tallene vil regresjoner med disse opplysningene føre til mindre utvalg, men til gjengjeld mer presise tall. Denne problemstillingen blir nærmere drøftet senere i oppgaven (se avsnitt 5.6, side 70).

**Tabell 2: Oversikt over variablene**

Variabelnavn	Definisjon av variabel	Måleenhet	Minimums- og maksimumsverdier
S1	Kostnadsandel utd.gruppe 1	Andel	0,021 – 0,347
S2	Kostnadsandel utd.gruppe 2	Andel	0,124 – 0,813
S3	Kostnadsandel utd.gruppe 3	Andel	0,053 – 0,559
S4	Kostnadsandel utd.gruppe 4	Andel	0,003 – 0,660
W1	Inntekt utd.gruppe 1	Løp. NOK	156 674 - 665 538
W2	Inntekt utd.gruppe 2	Løp. NOK	177 136 - 649 486
W3	Inntekt utd.gruppe 3	Løp. NOK	190 328 - 820 614
W4	Inntekt utd.gruppe 4	Løp. NOK	249 125 - 901 591
L1	Sysselsatte utd.gruppe 1	Menn	6 – 17 604
L2	Sysselsatte utd.gruppe 2	Menn	60 – 95 245
L3	Sysselsatte utd.gruppe 3	Menn	12 – 33 547
L4	Sysselsatte utd.gruppe 4	Menn	1 – 15 879
K	Fast realkapital/kapitalstokk	Mill. Løp. NOK	2 - 1 321 414
V	Verdiskapning	Mill. Løp. NOK	6 – 447 660
P	Produksjon	Mill. Løp. NOK	6 – 405 955
Smalimpo	Import fra samme utenlandske næring	Mill. Løp. NOK	0 - 18 109
Tot73	Innkjøp fra NACE-næring 73 (forsknings- og utviklingstjenester)	Mill. Løp. NOK	0 - 1 445

### **4.3 Mer om variablene: Deskriptiv statistikk**

Oppgaven følger en følger en rekke forskningsprosjekter som har søkt å kartlegge sammenhengen mellom økt handel av innsatsvarer og økende inntektsgap mellom arbeidskraft med ulike kompetansenivå. Med dette som bakgrunn er det viktig å kartlegge i hvilken grad den norske økonomien virkelig har opplevd slike endringer.

### 4.3.1 Kostnadsandeler, inntekts- og sysselsettingsdata

Gjennomsnittsinntekten til en hver tid i de forskjellige næringene varierer mye. Dette får man en indikasjon på ved å kaste et blikk på minimums og maksimumsverdiene i Tabell 2. Dette kan delvis forklares av konstruksjonen av variabelen inntekt. Den er definert som den *gjennomsnittlige pensjonsgivende inntekten til menn med positiv pensjonsgivende inntekt*. Dette innebærer at personer med deltidsstillinger påvirker den gjennomsnittlige inntekten for næringene forholdsvis sterkt, og at det ideelt sett burde vært regnet med heltidsekvivalenter, med antall årsverk som sysselsettingsvariabel. Dette er dessverre ikke mulig med det aktuelle datagrunnlaget. Tabell 3 viser gjennomsnittlige nivå for sysselsettingsandelene, inntekt og sysselsetting i 51 næringer for første, midterste og siste år i datasettet.

**Tabell 3: Gjennomsnittlige verdier i utvalgte år**

	1996	2000	2005	1996 - 2005	Totalt
S1	0,159 (0,081)	0,125 (0,066)	0,107 (0,056)	0,129 (0,069)	
S2	0,543 (0,1469)	0,580 (0,159)	0,572 (0,174)	0,567 (0,160)	
S3	0,203 (0,123)	0,195 (0,121)	0,207 (0,122)	0,202 (0,122)	
S4	0,095 (0,114)	0,100 (0,118)	0,114 (0,125)	0,102 (0,117)	
W1	238130 (51560)	277765 (57684)	344771 (85033)	291118 (73451)	
W2	251388 (50105)	306270 (61063)	374316 (73695)	314214 (73937)	
W3	326619 (65990)	407157 (83043)	508261 (101542)	419835 (102393)	
W4	419225 (71815)	503367 (98285)	607899 (128642)	519880 (121672)	
L1	2770 (3319)	2401 (3053)	2223 (2990)	2444 (3090)	124668 (11447)
L2	10120 (12520)	11618 (15198)	11568 (16139)	11243 (14812)	573372 (25677)
L3	3469 (5945)	3647 (6240)	3821 (6526)	3689 (6262)	188145 (5465)
L4	1352 (2858)	1562 (3228)	1802 (3607)	1572 (3223)	79548 (7381)

Standardavvik oppgitt i parenteser, første, midterste og siste inntekts-, sysselsettings- og kostnadsobservasjoner, Kolonnen ”Totalt” inneholder gjennomsnittlig antall sysselsatte til sammen i de 51 næringene.

Det kommer frem fra tabellen over at flere av variablene har hatt tilsynelatende jevn utvikling i analyseperioden. Kostnadsandelene til utdanningsgruppe 1 og gruppen med høyere universitetsutdanning, utdanningsgruppe 4, har henholdsvis sunket fra 15,9 % via 12,5 % til 10,7 % og økt fra 9,5 %, via 10 %, til 11,4 %. På den annen side har kostnadsandelene til de med videregående utdanning hatt en økning fra 1996 til 2000, for så å avta noe frem til 2005. Det motsatte er tilfellet for utdanningsgruppe 3, sysselsatte med lavere universitetsutdanning, som fra et 10 % lavere nivå (1 prosentpoeng) i 2000, kommer tilbake på omtrent nøyaktig samme nivå i 2005 som i 1995.

Når det gjelder utviklingen i inntekt for de fire utdanningsgruppene, ser det ut til at det har vært en jevn økning i inntektsnivået til alle utdanningsgruppene. En undersøkelse av prosentvise endringer kommer under. Ved å undersøke nivåene kommer det frem at inntektsveksten har vært klart høyest for gruppene med høyest utdanning.

Sysselsettingen av de forskjellige typene arbeidskraft reflekterer til en viss grad trenden i kostnadsandelene, med en jevn reduksjon for utdanningsgruppe 1 og en jevn økning for utdanningsgruppe 4. Sysselsetting av utdanningsgruppe 2 øker en del fra 1996 til 2000, for så å stabilisere seg frem til 2005. Utdanningsgruppe 3 opplever som utdanningsgruppe 4 en jevn økning i sysselsettingen fra 1996 til 2005.

Standardavvikene, oppgitt i parentes, viser at variasjonen mellom næringene ikke er ubetydelig. Det er særlig stor varians i hvor mye ressurser hver næring bruker på gruppene med høyest utdanning. Tabellene under gir en klarer fremstilling av hvordan næringene er ulike i sammensetningen av arbeidskraft.

### Kostnadsandelene

Tabell 4 viser hvordan utviklingen i kostnadsandelene har vært i perioden fra 1996 til 2005. Som Tabell 3 viste hadde utdanningsgruppe 1 en mindre del av totale inntektskostnader i 2005 enn i 1996, og motsatt for utdanningsgruppe 4. Det kommer frem av tabellen under at de hadde en gjennomsnittlig utvikling på rundt 4 % per år i hver sin retning. Kostnadsandelene forbundet med sysselsatte fra utdanningsgruppene 2 og 3 har totalt sett hatt en svak økning. Oppdelingen av standardavvikene gjør at det kommer klarere frem hva som skyldes variansen i variablene, om næringene har blitt påvirket på forskjellige måter, eller om det er variasjon innen næringene som er utslagsgivende. Det som omtales som *mellom*, er variansen til de

forskjellige næringenes gjennomsnittsverdier. Høye standardavvik betyr at næringene er svært ulike. Med *innen* forstås næringenes variasjon fra eget gjennomsnitt. Globalt gjennomsnitt er addert inn slik at maksimums- og minimumsverdiene skal være sammenlignbare med tilsvarende verdier for mellom-varians.

**Tabell 4: Årlige, prosentvise endringer i kostnadsandelene, 1996 - 2005**

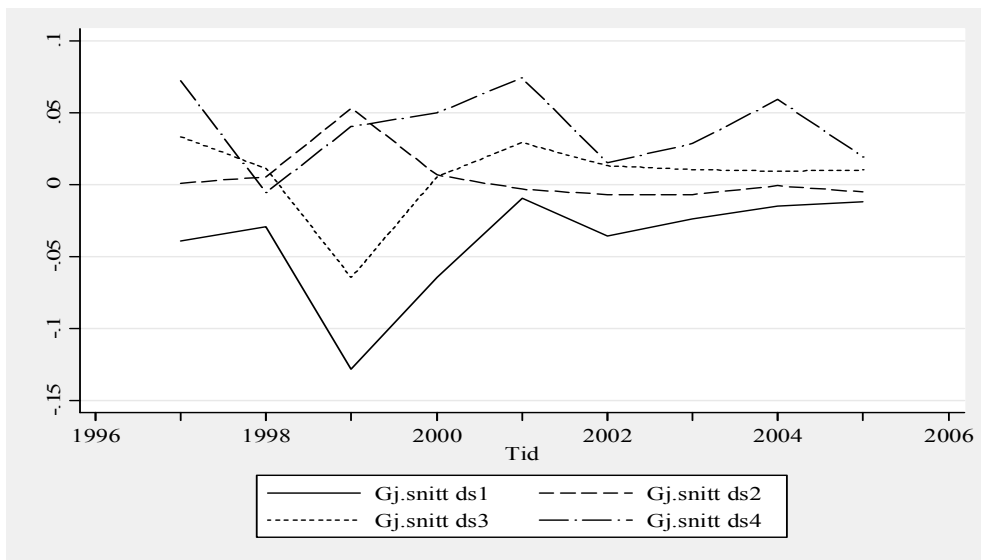
Variabel	Oppdeling	Gj.snitt	Std.avv.	Min	Maks	Observasjoner
S1	Totalt	-0,0393	0,0789	-0,4018	0,3695	Obs = 455
	Mellom		0,0221	-0,0887	0,0411	N= 51
	Innen		0,0758	-0,4822	0,2890	Tid = 8,92
S2	Totalt	0,0049	0,0467	-0,2019	0,6826	Obs = 455
	Mellom		0,0153	-0,0246	0,0735	N= 51
	Innen		0,0442	-0,1893	0,6140	Tid = 8,92
S3	Totalt	0,0067	0,0778	-0,6728	0,6263	Obs = 455
	Mellom		0,0185	-0,0525	0,0507	N= 51
	Innen		0,0756	-0,6136	0,6059	Tid = 8,92
S4	Totalt	0,0395	0,1978	-0,6298	2,1161	Obs = 455
	Mellom		0,0587	-0,0802	0,3224	N= 51
	Innen		0,1908	-0,7489	1,9533	Tid = 8,92

Tabellen viser prosentvis endring fra foregående år for de respektive gruppene.

Det er klart at selv om gjennomsnittlig ressurser brukt på utdanningsgruppene 1 og 4 har endret seg i samme størrelsesorden, er det langt større varians i endringen for utdanningsgruppe 4. For alle utdanningsgruppene er det variasjon *innen* næringene som har vært mest utslagsgivende i variansen til variablene. Dette betyr at...

Figur 1 viser hvordan utviklingen i de fire variablene har vært gjennom analyseperioden. De totale endringene fra 1996 til 2005 er ikke spesielt store, men det kommer klart frem av figuren er at det har vært store variasjoner i løpet av perioden. Særlig fremtredende er de reduserte kostnadsandelene til utdanningsgruppe 1 og 3 i 1999. Årsaken til denne utviklingen er ikke klar. Arbeidsledigheten var i 1999 spesielt lav, så jeg ser på dette med en viss skepsis, og kontrollerer i analysen om utelukkning av dette året påvirker resultatene. I tillegg kommer det frem at også kostnadsandelene til utdanningsgruppe 4 har forholdsvis stor variasjon, men uten de helt store toppene og dalene.

**Figur 1: Uvikling i kostnadsandelene; årlige endringer 1997 -2005.**



Verdiene er gjennomsnittlig endring fra foregående år for de 51 næringene.

### Inntekt

Tabell 5 viser hvordan inntektsutviklingen for de fire utdanningsgruppene har vært i perioden 1996 – 2005. Som tidligere nevnt har inntektsutviklingen for de fire utdanningsgruppene vært forholdsvis lik i de aktuelle årene, med unntak av gruppe 3 som har hatt en noe høyere vekst. Det som er spesielt her er at *mellom*-standardavviket er langt lavere enn det var for kostnadsandelene. Dette speiler det norske lønnsforhandlingssystemet, med sentraliserte forhandlinger som favner om store deler av arbeidslivet. Dette innebærer at det er forholdsvis liten variasjon i inntektsveksten mellom næringene. Også variasjonen innen hver næring er forholdsvis lav, i og med at det stort sett har vært gode tider i norsk økonomi i denne perioden.

**Tabell 5: Årlige, prosentvise endringer i inntekt, 1996 - 2005**

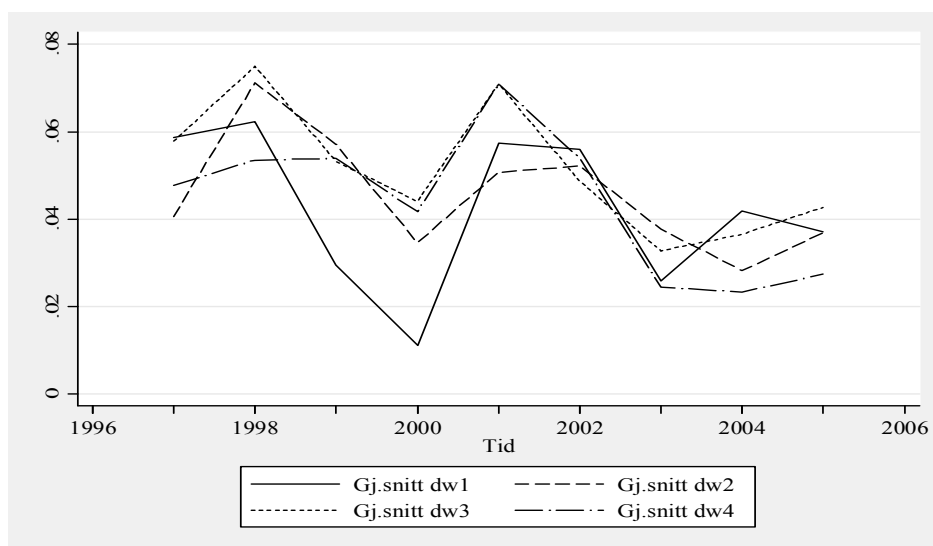
Variabel	Oppdeling	Gj.Snitt	Std.avv.	Min	Maks	Observasjoner
W1	Totalt	0,0422	0,0442	-0,1872	0,3973	obs = 459
	Mellom		0,0076	0,0258	0,0612	N= 51
	Innen		0,0436	-0,1989	0,3856	Tid = 9
W2	Totalt	0,0455	0,0238	-0,0805	0,1715	Obs = 459
	Mellom		0,0054	0,0312	0,0566	N= 51
	Innen		0,0232	-0,0792	0,1646	Tid = 9
W3	Totalt	0,0513	0,0446	-0,1269	0,3646	Obs = 459
	Mellom		0,0081	0,0370	0,0752	N= 51
	Innen		0,0438	-0,1404	0,3512	Tid = 9

W4	Totalt	0,0441	0,0640	-0,2421	0,3489	Obs = 455
	Mellom		0,0183	-0,0023	0,1398	N= 51
	Innen		0,0620	-0,1960	0,2874	Tid = 8,92

Tabellen viser prosentvis endring fra foregående år, for de respektive gruppene.

Figur 2 viser hvordan inntektsveksten har variert i perioden fra 1997 – 2005. Særlig inntektsutviklingen til utdanningsgruppene 2 og 3 har fulgt lignende bane, på et generelt høyere nivå enn spesielt utdanningsgruppe 1. Det kan se ut som særlig gruppe 1 blir påvirket av sentrale lønnsforhandlinger, i og med at veksten i lønn endrer seg særlig mye annethvert år, som for eksempel fra 2000 til 2002.

**Figur 2: Inntektsutvikling; årlige endringer 1997 -2005.**



Verdiene er gjennomsnittlig endring fra foregående år for de 51 næringene.

### Sysselsetting

Ettersom utdanningsgruppe 4 stod for en større andel av inntektsutgiftene i 2005 enn i 1996, mens det samtidig har vært en forholdsvis jevn inntektsutvikling, innebærer det at det er sysselsettingsnivåene som har variert, og endret seg, mest. Tabell 6 viser hvordan sysselsettingsutviklingen har vært i analyseperioden.

**Tabell 6: Årlige, prosentvise endringer i sysselsetting, 1996 - 2005**

Variabel	Oppdeling	Gj.snitt	Std.avv.	Min	Maks	Observasjoner
L1	Totalt	-0,0306	0,1182	-0,6364	0,7451	Obs = 459
	Mellom		0,0462	-0,1315	0,0866	N= 51
	Innen		0,1089	-0,5416	0,6730	Tid = 9

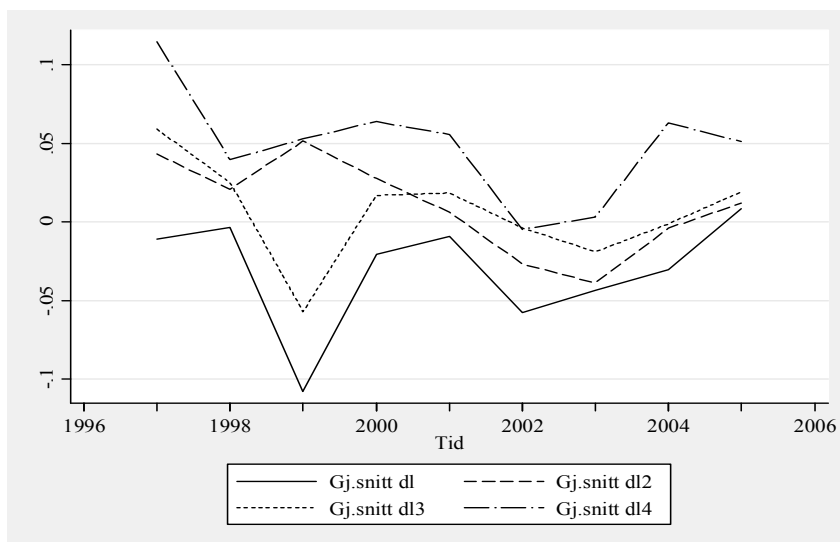


L2	Totalt	0,0102	0,1031	-0,5745	0,6199	Obs = 459
	Mellom		0,0411	-0,0811	0,1433	N= 51
	Innen		0,0947	-0,4832	0,4869	Tid = 9
L3	Totalt	0,0064	0,1154	-0,7120	0,6696	Obs = 459
	Mellom		0,0393	-0,0887	0,1220	N= 51
	Innen		0,1086	-0,6444	0,7324	Tid = 9
L4	Totalt	0,0486	0,2161	-0,6947	2,3333	Obs = 455
	Mellom		0,0581	-0,0932	0,2672	N= 51
	Innen		0,2083	-0,9133	2,1148	Tid = 8,92

Tabellen viser prosentvis endring fra foregående år, for de respektive gruppene.

Som det kommer frem i tabellen over, er det primært gjennom endringer i sysselsetting at kostnadsandelene til utdanningsgruppe 1 og 4 endret seg. Årlig har det blitt 3 % færre sysselsatte med grunnskoleutdanning, mens det har vært nesten 5 % økning årlig i sysselsetting av individer fra utdanningsgruppe 4, noe som korresponderer bra med utviklingen til kostnadsandelene. Utdanningsgruppe 2 og 3 har hatt forholdsvis stabile sysselsettingstall. Det kommer frem ved nærmere undersøkelse at 1999 var et spesielt år med hensyn på sysselsettingsendringer. Som det kommer frem i Figur 3, opplevde utdanningsgruppene 1 og 3 langt lavere sysselsetting dette året enn årene før og etter. Også årene 2001, 2002 og 2003 hadde lavere sysselsettingsvekst enn årene før og etter.

**Figur 3: Sysselsettingsutvikling; årlige endringer 1997 -2005.**



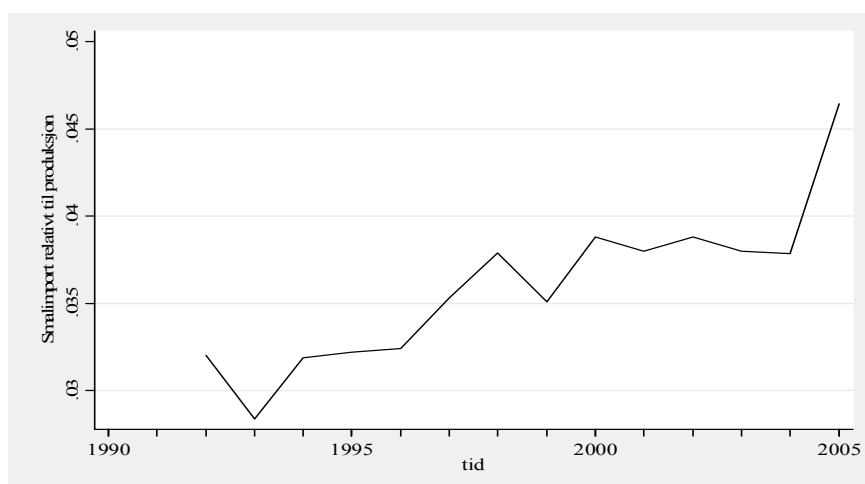
Verdiene er gjennomsnittlig endring fra foregående år for de 51 næringene.

### 4.3.2 Offshoring

Etter en grundig gjennomgang av indikatorer på etterspørselen etter arbeidskraft, er det den viktigste forklaringsvariabelen som nå blir presentert. Offshoring er et begrep som favner om situasjoner hvor bedrifter skaffer varer fra utlandet snarere enn å skaffe dem fra hjemlandet. Dette være seg tilfeller hvor bedriftene, i stedet for å produsere varen i en hjemlig del av bedriften, flytter produksjonen til en del av bedriften som holder til i utlandet, eller tilfeller hvor bedriften kjøper varene inn fra andre bedrifter i utlandet. Som tidligere nevnt er definisjonen på offshoring ikke alltid like klar, men i denne oppgaven baserer jeg meg på det som Feenstra og Hanson (1996b) kaller *narrow offshoring*. *Smalimport* (som er det jeg kaller *narrow offshoring* i datasettet) fanger situasjoner hvor bedrifter kjøper inn varer fra samme næringer i utlandet som de selv holder til i, og er dermed et mål på substituering av norsk og utenlandsk arbeidskraft.

De aller fleste næringene importerer innsatsvarer fra utlandet, og veldig mange importerer varer fra samme næring i utlandet, med utviklingen for årene 1992 – 2005 i en klar positiv trend. Som en del av produksjonsinnsatsen er denne importen kun et substitutt for norsk arbeidskraft dersom det blir sett i forhold til nivået på produksjonen. Det at importen har økt i den nevnte perioden, trenger ikke å bety at norsk arbeidskraft har blitt substituert bort i stor stil. De forskjellige målene på denne variabelen som kommer frem i litteratur på dette området, har en ting til felles idet man alltid snakker om offshoring relativt til et mål på størrelse på næringen, for eksempel produksjon eller verdiskapning. Som tidligere nevnt er *smalimport* i denne oppgaven vektet mot produksjon, og nevnes som *relsmal*.

**Figur 4: Utvikling i smalimport; 1992 – 2005.**



Som det kommer frem av Figur 4 har det vært en klar økning i bruken av importerte varer fra samme utenlandske næring, i forhold til total produksjon. Selv om det ses bort fra den spesielt store økningen fra 2004 til 2005, har bruken offshoring økt med over 25 %, fra 4,2 % til 5,4 %, fra 1992 til 2004. Sammenlignet med Ekholm og Hakkalas (2005) tall for Sverige, som er på 4 % i 1995 og 4,2 % i 2000, tilsvarende 10 % økning på fem år, er dette store tall. Hijzen *et al.* (2005) finner for Storbritannia at smalimport relativt til *verdiskapning* økte fra 11 % til 16 % fra 1984 til 1995. Til sammenligning er tilsvarende tall i mitt datasett 13,4 % i 1992 og 16,8 % i 2004 (19,5 % i 2005). Det kan se ut som om Norge ligger på nivå med Sverige, og noe under Storbritannia i bruk av importerte innsatsvarer.

**Tabell 7: Utvikling i smal- og bredimport, 1992 - 2005**

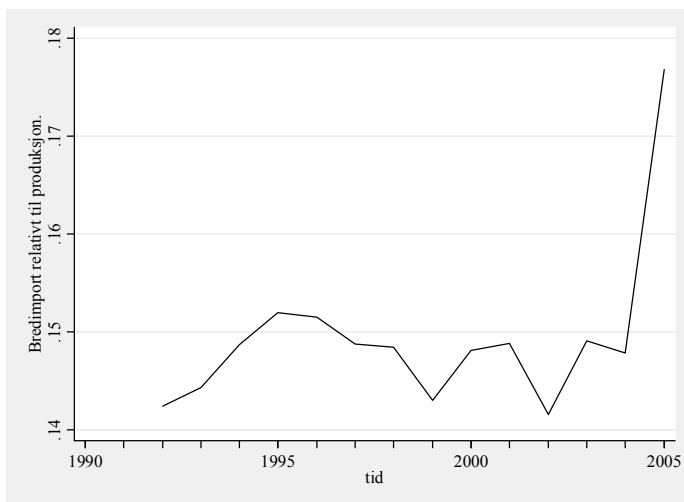
Variabel		Gj.snitt	Std.avv.	Min	Maks	Observasjoner
Relsmal	Totalt	0,036	0,069	0,000	0,440	Obs = 714
	Mellom		0,067	0,000	0,310	N = 51
	Innen		0,016	-0,114	0,178	Tid = 14
Relbred	Totalt	0,149	0,111	0,000	0,557	Obs = 714
	Mellom		0,109	0,026	0,486	N = 51
	Innen		0,025	-0,114	0,361	Tid = 14

Verdiene er oppgitt som import relativt til produksjon.

Det er stor variasjon i intensiteten i innsatsvarebruken. Tabell 7 viser nettopp dette. Med et stort mellom-næring-standardavvik kommer det frem at næringene bruker muligheten for offshoring ulikt. Både variasjonen i bruk import av innsatsvarer generelt og import av innsatsvarer fra samme utenlandske næring er preget av større variasjon mellom næringene

enn innen næringene. Mengden import av innsatsvarer totalt (bredimport), har ikke så stor innen-næring-varians, da veksten i denne variabelen ikke har vært så sterk som for smalimport, utenom for det siste året, hvor det er en økning på omtrent 3 prosentpoeng, eller om lag 20 %. Hva den siste endringen skyldes, kan til en viss grad forklares av fall i produksjonen fra 2004 til 2005 (se Figur 6, side 47), som faller med omtrent 5 %, samtidig som importen også økte betydelig.

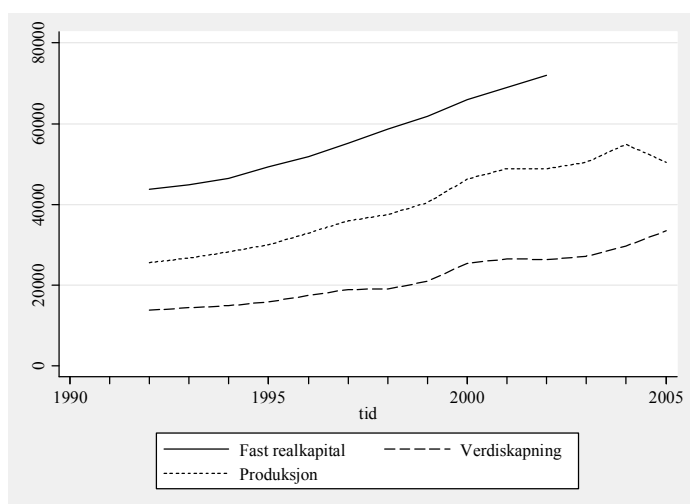
**Figur 5: Utvikling i bredimport; 1992 - 2005.**



### 4.3.3 Verdiskapning, fast realkapital og produksjon

Utviklingen i disse variablene har vært forholdsvis stabilt positiv i undersøkelsesperioden. Spesielt er det akkumulasjonen av fast realkapital som har vokst raskt, men også verdiskapning og produksjon vokste nærmest hele perioden. I 2005 sank produksjonen med i gjennomsnitt 5 % i forhold til i 2004.

**Figur 6: Utvikling i fast realkapital, produksjon og verdiskapning; 1992 - 2005**



Tabell 8 viser at det er liten variasjon innen næringene da de alle har en sterk tidstrend i disse variablene. Det er også klart at det er store variasjoner mellom næringen, noe som er naturlig ettersom det er stor forskjell dem i mellom.

**Tabell 8: Utvikling i fast realkapital, produksjon og verdiskapning; 1992 -2005**

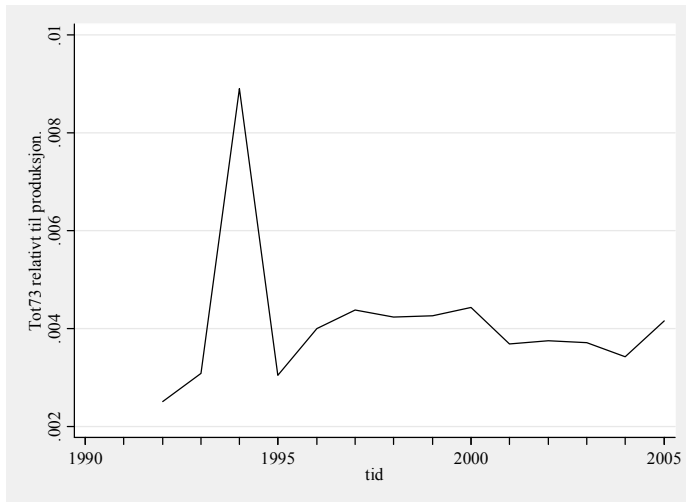
Variabel		Gj.snitt	Std.avv.	Min	Maks	Observasjoner
Fast realkapital	Totalt	54 978	148 058	2	1 321 414	Obs = 510
	Mellom		146 543	115	949 094	N = 51
	Innen		28 742	-161 711	427 299	Tid = 10
Verdiskapning	Totalt	21 698	38 781	6	447 660	Obs = 714
	Mellom		35 186	165	212 939	N = 51
	Innen		16 983	-97 963	256 420	Tid = 14
Produksjon	Totalt	39 804	50 132	6	405 955	Obs = 714
	Mellom		46 820	493	241 028	N = 51
	Innen		19 001	-82 640	204 731	Tid = 14

Mangler observasjoner på fast realkapital for årene 2001 og 2003-2005

#### 4.3.4 Forskning og utvikling

Bruken av innkjøpte varer fra næring 73, som er forsknings- og utviklingstjenester, har økt noe i analyseperioden, se Figur 7, men har med unntak av i 1994 vært forholdsvis stabil rundt 0,4 prosentpoeng. Selv om toppen i 1994 kan se ekstrem ut, noe den i og for seg er, omtrent 200 % prosent høyere enn årene før og etter, er det kun en endring på omtrent 4 promillepoeng. Tabell 9 viser at variasjonen i ressurser brukt på dette er store både mellom næringene og over tid.

**Figur 7: Utvikling i kjøp av tjenester fra FoU-næringen; 1992 - 2005**



**Tabell 9: Utvikling i kjøp av tjenester fra FoU-næringen; 1992 - 2005**

<i>Variabel</i>		<i>Gj.snitt</i>	<i>Std.avv.</i>	<i>Min</i>	<i>Maks</i>	<i>Observasjoner</i>
Reltot73	Totalt	0,004	0,004	0,000	0,061	Obs = 714
	Mellom		0,002	0,001	0,014	N = 51
	Innen		0,003	-0,005	0,052	Tid = 10

## 5 Resultater fra den økonometriske analysen

I dette kapitlet presenteres funnene fra regresjonsanalysen på datasettet som ble presentert i forrige kapittel. Noe av det som er spesielt med datasettet og problemstillingen, er at det blir flere sett regresjoner. Disse settene består av fire regresjoner, en for hver kostnadsandel. Dette innebærer at det blir mange resultat å forholde seg til. Som tidligere vist, er det noen restriksjoner som må oppfylles i analysen (se 3.1.1). Forutsetningene om symmetri og at lønneffektene i hver kostnadsandelsfunksjon skal summe seg til 1 må pålegges regresjonsuttrykkene eksplisitt.<sup>31</sup> De andre er som vist automatisk oppfylte. I tillegg estimeres ligningene i SUR-rammeverket, slik at restriksjonene kan implementeres og slik at relevante tester kan gjennomføres. Ettersom ligningene estimeres simultant, rapporterer jeg resultatene for alle fire regresjonene sammen. Det vil også gi en mer oversiktlig presentasjon, og det blir lettere å sammenligne resultatene.

Kapitlet er strukturert som følger; først presenteres estimatene for det sammenslåtte paneldatasettet. Deretter presenterer jeg paneldatamodellene uten dynamikk, for så å avslutte hovedanalysen med en undersøkelse av hvordan inklusjon av dynamikk i modellene vil påvirke resultatene. Dette blir etterfulgt av en utdyping og diskusjon av resultatene. Til slutt diskuteres alternative spesifiseringer, og bruk av andre forklaringsvariabler.

### 5.1 Sammenslåtte tverrsnittsdata

Jeg begynner med en undersøkelse av såkalte sammenslåtte tverrsnittsdata. Dette innebærer å bruke minste kvadrats metode på hele datasettet *som om* det var et stort sett uavhengige tverrsnittsobservasjoner. Som tidligere nevnt, er en av de viktige forutsetningene for minste kvadrats metode at det ikke skal være en sammenheng mellom forskjellige observasjoner, noe som dermed åpenbart ikke er tilfellet her, da samme næring har opptil ni observasjoner (en for hver tidsperiode). Kostnadsandelene for de forskjellige utdanningsgruppene i de ulike næringene er så absolutt forbundet med tilsvarende verdier i tidsperiodene før og etter. Dermed er ikke Gauss-Markow-vilkåret om uavhengighet mellom observasjonene oppfylt.

---

<sup>31</sup>  $\beta_{js} = \beta_{sj}$  og  $\sum_{s=1}^J \beta_{js} = 0$

Dette medfører at svekket inferens, i og med at estimerte standardfeil, og dermed også t-verdier, ikke er riktige. Likevel skal estimatorene være forventingsrette gitt at det ikke er avhengighet mellom regressor og individspesifikke feilledd.

Det er likevel et rimelig utgangspunkt for undersøkelse, og som Hijzen *et al.* (2005) og Ekholm og Hakkala (2005) gjør, undersøker jeg først datasettet med bruk av minste kvadrats metode. Regresjonene som utføres kommer fra uttrykk (4), men er her omrokkert noe, og noen relevante kontrollvariabler er inkludert;

$$S_{ijt} = \alpha_{ij} + \beta_{j\text{smal}} \text{Relsmal}_{it} + \beta_{j\text{fou}} \text{Reltot73}_{it} + \beta_{jv} \ln V_{it} + \beta_{jk} \ln K_{it} + \beta_{j1} \ln W_{i1t} + \beta_{j2} \ln W_{i2t} + \beta_{j3} \ln W_{i3t} + \beta_{j3} \ln W_{i4t} + \text{tidsdummyer} + e_{it}$$

Med  $j=1,2,3,4$  utdanningsgrupper,  $i=1, \dots, 51$  næringer og tidsindikatorer;  $t=1996, \dots, 2005$ .

S er de respektive kostnadsandelene i hver næring til ethvert tidspunkt, Relsmal er offshoring relativt til produksjon, Reltot73 er forbruk av forsknings- og utviklingstjenester relativt til produksjon,  $\ln V$  er verdiskapning på logaritmeform,  $\ln K$  fast realkapital på logaritmeform og  $\ln W$  er de forskjellige utdanningsnivåenes inntekter på logaritmisk form. Alt er oppgitt for næringer og tidsperioder.

Tidsdummyer er inkludert i regresjonen for å fange opp konjunkturvirkninger som er felles for alle næringer. Blant annet vil disse fange opp utviklingen i verdier på forklaringsvariablene som øker på grunn av at dataene er oppgitt i løpende priser.

I analysen er det et problem med simultanitet. Det er en vanlig betraktning at inntekt, som i utgangspunktet befinner seg på begge sider av regresjonslikningene, og sysselsettingsnivå blir fremforhandlet simultant. Derfor har jeg brukt inntekt fra perioden før som proxy for inntekt i hver periode, inspirert av Verbeek (2004), som gjør dette i et eksempel på hvordan man kan estimere etterspørselen etter arbeidskraft. For å understreke at det brukes laggede inntektsmål i analysen, omtales inntekten til de forskjellige gruppene som  $\ln W_{i\_lag}$  i tabellene, med  $\_lag$  som indikator på at det er verdien fra foregående periode. Det er en sterk korrelasjon mellom inntekt i påfølgende perioder, og ettersom det er et år mellom inntekts- og sysselsettingsobservasjonene, brukes dette som en strategi for å omgå simultanitetsproblemet.



Alle tabellene i dette kapittelet inneholder estimerte størrelser på  $\beta$ -ene,<sup>32</sup> med tilhørende standardfeil oppgitt i parentes under og signifikansnivå markert med stjerner, \*\*\* for 1 % nivå, \*\* for 5 % og \* for 10 %.

**Tabell 10: Kostnadsandelene regressert med minste kvadrats metode**

	S1	S2	S3	S4
Relsmal	-0,2182 (0,0412)***	-0,4959 (0,1057)***	0,2576 (0,0814)***	0,4565 (0,0751)***
Reltot73	-5,7981 (1,2228)***	-23,4944 (3,1237)***	13,1672 (2,4110)***	16,1254 (2,2214)***
lnV	-0,0241 (0,0025)***	-0,0365 (0,0064)***	0,0369 (0,0049)***	0,0238 (0,0045)***
lnK	0,0029 (0,0018)	0,0098 (0,0046)**	-0,0055 (0,0036)	-0,0072 (0,0033)**
lnW1_lag	0,0228 (0,0246)	-0,0854 (0,0289)***	-0,0507 (0,0252)**	0,1133 (0,0159)***
lnW2_lag	-0,0854 (0,0289)***	-0,0600 (0,0586)	0,0560 (0,0458)	0,0894 (0,0350)**
lnW3_lag	-0,0507 (0,0252)**	0,0560 (0,0458)	0,2249 (0,0541)***	-0,2302 (0,0297)***
lnW4_lag	0,1133 (0,0159)***	0,0894 (0,0350)**	-0,2302 (0,0297)***	0,0275 (0,0285)
Konstantledd	0,3298 (0,0258)***	0,8371 (0,0642)***	-0,0828 (0,0507)	-0,0840 (0,0452)*
Tidsdummyer	Ja	Ja	Ja	Ja
Næringsdummyer	Nei	Nei	Nei	Nei
Observasjoner	455	455	455	455
Parametre	15	15	15	15
RMSE	0,0566	0,1459	0,1120	0,1041
R <sup>2</sup>	0,287	0,209	0,184	0,250
F-stat	14,81	7,65	10,51	12,94
P	0,00	0,00	0,00	0,00

STATA rapporterer modellenes forklaringskraft, betegnet som  $R^2$ . Denne størrelsen viser andelen av variasjonen i avhengig variabel som kan forklares av variasjon i uavhengige variabler. I tilfellene over ligger forklaringskraften mellom 18 % og 29 %, hvor særlig de to regresjonene med lavest forklaringskraft har lav  $R^2$  til å være estimert med minste kvadrats metode.

F-stat er en testobservator som reflekterer forklaringskraften i modellen, og er den tilstrekkelig høy, betyr det at modellen kan stoles på, noe som vil si at ikke alle koeffisientene

<sup>32</sup> Med mindre noe annet er spesifisert.

er insignifikant forskjellige fra null samtidig. P viser på hvilket signifikansnivå man minst kan godta modellen. I tabellen over godtas alle fire regresjonene på alle signifikansnivå.

Tabell 10 viser stort sett signifikante sammenhenger mellom de strukturelle forklaringsvariablene og kostnadsandelene, hovedsaklig på 1 % nivå, utenom for fast realkapital. Koeffisientene for sammenheng mellom inntekt og kostnadsandeler er i noe mindre grad signifikante enn koeffisientene for de andre variablene.

Effektene av økt offshoring, når produksjonen holdes fast, er i tråd med det som predikeres i teorien, og er for alle utdanningsnivåene signifikant på 1 % nivå. Utdanningsgruppe 2 blir negativt påvirket av offshoring, noe som er på linje med funnene i Ekholm og Hakkala (2005). Utdanningsgruppe 1, sysselsatte med grunnskoleutdanning, er også under negativ påvirkning av økt offshoring. Den gruppen som får den største økningen i kostnadsandelen er de med høyere universitetsutdanning. Gruppen med lavere universitetsutdanning er utsatt for noe mindre, men positiv, påvirkning.

I tråd med teorien ser det ut til at næringenes bruk av forsknings- og utviklingstjenester øker produktiviteten til arbeidstakere med universitetsutdanning. Dermed går deres del av inntektskostnadene opp, og andelene til dem med mindre utdanning går ned.

Koeffisientene for sammenhengene mellom verdiskapning og kostnadsandelene er signifikante på 1 % nivå, og viser at økt verdiskapning fører til en reduksjon i andelen av kostnadene til gruppene med lavere utdanning, og økning i kostnadsandelene til arbeidstakere med universitetsutdanning.

Ellers kommer det frem at fast realkapital kan ses på som komplementær med arbeidskraft med lavere utdanning, ettersom kostnadsandelene til disse gruppene øker når den faste realkapitalen øker. Motsatt er det for de to gruppene med universitetsutdanning, som dermed kan ses på som substitutter for kapital. Ettersom bare to av koeffisientene er signifikante på 5 % nivå, kan dette bare karakteriseres som en trend, i det estimatene for de to andre gruppene ikke er signifikant forskjellige fra null.

**Tabell 11: Etterspørselastisiteter basert på regresjon med minste kvadrats metode**

	<i>Relsmal</i>	<i>Reltot73</i>	<i>V</i>	<i>K</i>	<i>W1</i>	<i>W2</i>	<i>W3</i>	<i>W4</i>
L1	-1,741 (0,329)***	-46,255 (5,484)***	-0,193 (0,020)***	0,023 (0,014)*	-0,692 (0,196)***	-0,112 (0,231)	-0,202 (0,201)	1,006 (0,127)***
L2	-0,871 (0,186)***	-41,246 (5,484)***	-0,064 (0,011)***	0,017 (0,008)**	-0,025 (0,051)	-0,580 (0,103)***	<b>0,301</b> (0,080)***	0,260 (0,061)***
L3	1,273 (0,402)***	65,057 (11,912)***	0,182 (0,024)***	-0,027 (0,018)	-0,125 (0,124)	0,846 (0,226)***	0,314 (0,267)	-1,035 (0,147)***
L4	4,448 (0,732)***	157,115 (21,644)***	0,232 (0,044)***	-0,070 (0,032)**	1,229 (0,155)***	1,441 (0,341)***	-2,040 (0,289)***	-0,629 (0,278)**

Størrelsene betegner for *relsmal* og *reltot73* prosentvis endring i sysselsetting som svar på ett prosentpoengs endring i de foran nevnte variablene. For de andre variablene er det prosentvis endring i sysselsetting etter en prosent høyere verdiskaping og fast realkapital.

Tabell 11 viser hvordan estimatene fra regresjonstabellene kan kvantifiseres som prosentvise endringer i sysselsetting, alt annet likt. Det vil si at elastisitetene viser hvordan sysselsettingen endres dersom det kommer en økning i offshoring med ett prosentpoeng, eller dersom bruken av forsknings- og utviklingstjenester øker med ett prosentpoeng. For alle andre variabler er responsen i sysselsetting en følge av en prosentvis endring i den aktuelle variabelen. Disse elastisitetene er ikke konstante, de er vektet av størrelsen på kostnadsandelene, og i presentasjonen er de evaluert på gjennomsnittlige verdier for kostnadsandelene.

Som det kommer frem, er det relativt høye elastisiteter for de to første variablene, offshoring og forskning og utvikling. Dette kommer av at en økning på ett *prosentpoeng*, innebærer henholdsvis en økning på henholdsvis 28 % og 250 % på om lag for disse variablene. Dette kommer av at størrelsen offshoring i gjennomsnitt er 3,6 % av produksjonen i de 51 næringene, og forbruk på forsknings- og utviklingstjenester er på vel 0,4 % i gjennomsnitt. Tabellen gir likevel en indikasjon på størrelsesorden på sammenhengene når elastisitetene sammenlignes mellom utdanningsgruppene.

Inntektseffektene som er estimert, er i varierende grad statistisk signifikante. Ettersom en endring i en gruppes inntekt bør påvirke gruppens andel av kostnadene, både gjennom endret sysselsetting og selvsagt gjennom endret kostnad, er ikke koeffisientene lette å sammenligne, og det er mest naturlig å se på de respektive etterspørselastisitetene.

Tabell 11 viser også etterspørselastisitetene for arbeidskraft fra de forskjellige utdanningsgruppene. Selv om det i regresjonen er pålagt restriksjoner om symmetri i

parametrene, er ikke elastisitetene nødvendigvis symmetriske. Dette er fordi det, som vist i 3.1.2, vektet med hensyn på størrelsen på kostnadsandelene. De uttrykker prosentvis endring i sysselsettingen av de forskjellige typene arbeidskraft,  $L_j$ , ved en prosents økning i inntekten,  $W_j$ , til de forskjellige gruppene. Egenpriselasititeten til en gruppe forventes å være negativ. Dette er en nødvendig, men ikke tilstrekkelig, betingelse for konkavitet i kostnadsfunksjonen (Diewert og Wales, 1987). Selv om egenpriselasititeten til utdanningsgruppe 3 ikke er negativ, problematiserer jeg ikke dette videre, ettersom den ikke er statistisk signifikant.

Tolkning av elastisitetene går hovedsakelig på hvilket fortegn de har, og om de er større eller mindre enn 1. Som nevnt bør egenpriselasititetene være negative, som impliserer at en prisøkning på en innsatsfaktor fører til mindre bruk av denne. Dersom krysspriselasititetene er negative betyr det at de to faktorene er komplementære i produksjonene. Dette innebærer at det er vanskelig å erstatte den ene typen arbeidskraft med den andre. Positive krysspriselasititeter impliserer det motsatte. Symmetrirestriksjonene fører til at identiske fortegn på de symmetriske krysspriselasititetene.

Resultatene tyder på at gruppene 1 og 4, 2 og 4, og 2 og 3 er substitutter for hverandre. Dette innebærer at det ved en prisøkning på arbeidstakere med høyere universitetsutdanning fører til økt etterspørsel etter arbeidskraft med videregående utdanning eller mindre, og tilsvarende mellom arbeidstakere med lavere universitetsutdanning og arbeidstakere med videregående utdanning. I motsatt retning er forholdet mellom arbeidstakere med grunnskoleutdanning, og gruppene med videregående utdanning og lavere universitetsutdanning. I dette tilfellet er gruppene komplementære, og dersom lønningene for de med grunnskoleutdanning går opp, substitueres disse arbeidstakerne med arbeidstakere med høyere utdanning. Også de gruppene med universitetsutdanning er komplementære.

Inntektsmålene er sterkt korrelert med hverandre, og som nevnt vil multikolinearitetsproblemer kunne føre til usikre estimater, og i noen tilfeller koeffisienter med overraskende størrelser og fortegn. Dette gjør at resultatene presentert over skal tolkes med forsiktighet.

## 5.2 Paneldatateknikker

Ettersom datasettet inneholder observasjoner på 51 næringer over perioden 1996 til 2005, er det ønskelig å utnytte den ekstra kunnskapen dette kan gi. Da er det spesielt muligheten for undersøke, og spesifisere, variasjonen innenfor hver næring over tid som kan være fruktbar å utforske. Som nevnt tidligere kan man bruke både fasteffektmodellen og tilfeldigeffektmodellen til å undersøke data. Jeg begynner med fasteffektmodellen estimert med Zellners metode, diskuterer deretter om tilfeldigeffektmodellen bør brukes, og undersøker til slutt om korreksjon for autoregressiv prosess av første orden forandrer resultatene.

### 5.2.1 Fasteffektmodellen

Denne modellen baseres på variasjon innen næringene og omtales ifølge Verbeek (2004) ofte for *within*-estimatoren (innen-estimatoren). I analysen har jeg estimert modellen i SUR-rammeverket i STATA. Dette innebærer at jeg har estimert modellen med minste kvadrats metode med næringsdummyer, som gir de samme resultatene som ved vanlig estimering av fasteffektmodellen.

**Tabell 12: Kostnadsandelene regressert med fasteffektmodellen**

	S1	S2	S3	S4
Relsmal	-0,0314 (0,0411)	0,2421 (0,1000)**	-0,1276 (0,0790)	-0,0831 (0,0439)*
Reltot73	3,1934 (0,5331)***	-3,6407 (1,3094)***	-1,1680 (1,0345)	1,6153 (0,5737)***
InV	0,0076 (0,0030)	-0,0114 (0,0074)	-0,0043 (0,0058)	0,0081 (0,0032)**
InK	0,0007 (0,0005)	0,0023 (0,0012)*	-0,0022 (0,0009)**	-0,0007 (0,0005)
InW1	0,0403 (0,0155)***	-0,0547 (0,0186)***	-0,0028 (0,0144)	0,0172 (0,0078)**
InW2	-0,0547 (0,0186)***	0,0866 (0,0432)**	0,0027 (0,0316)	-0,0346 (0,0152)**
InW3	-0,0028 (0,0144)	0,0027 (0,0316)	0,0002 (0,0307)	-0,0001 (0,0122)
InW4	0,0172 (0,0078)**	-0,0346 (0,0152)**	-0,0001 (0,0122)	0,0175 (0,0092)*
Konstantledd	-0,0114 (0,0371)	0,6254 (0,0902)***	0,3033 (0,0711)***	0,0827 (0,0395)**
Tidsdummyer	Ja	Ja	Ja	Ja
Næringsdummyer	Ja	Ja	Ja	Ja

Observasjoner	455	455	455	455
Parametre	65	65	65	65
RMSE	0,0108	0,0265	0,0210	0,0117
R <sup>2</sup>	0,977	0,9768	0,9746	0,9916
F-stat	253,4	251,8	228,8	707,6
P	0,00	0,00	0,00	0,00

Som det kommer frem i Tabell 12 har koeffisientene for relsmal endret fra det som var tilfellet i modellen estimert med vanlig minste kvadrats metode. For det første har fortegnene for gruppene 2, 3 og 4 snudd, noe som indikerer at det er arbeidstakere med videregående utdanning som kommer best ut av det ved økt offshoring. Dette strider ikke bare mot resultatene fra forrige modell, men også mot resultatene til Ekholm og Hakkala (2005) og Hijzen *et al.* (2005). For det andre ga minste kvadrats metode signifikante resultater for alle utdanningsnivåene, mens det i fasteffektmodellen bare er virkningen av offshoring på utdanningsgruppe 2 og 4 som kan sies å være signifikant ulike null på henholdsvis 5 % og 10 % signifikansnivå. Flere andre forklaringsvariabler har skiftet fortegn. En del av disse koeffisientene er ikke lenger statistisk signifikant forskjellige fra null, noe som gjør at fortegnsendringen ikke trenger å være så viktig. De er nå ikke med sikkerhet ulike null, og det er dermed ikke sikkert at de virkelig har skiftet fortegn. Som det kom frem i Figur 3 (side 43) var det et spesielt fall i sysselsetting for utdanningsgruppene 1 og 3 i 1999. Resultatene fra regresjonene utført uten observasjonene fra 1999, for å kontrollere om sysselsettingsendringene det året påvirker resultatene, viser at retningene på koeffisientene er de samme, men at effekten av relsmal på gruppe 3 i denne estimeringen er signifikant på 10 % nivå.

Det som skiller minste kvadrats metode og fasteffektmodellen fra hverandre er at fasteffektmodellen inkluderer konstantledd for alle næringene, noe som innebærer at forskjellen mellom næringene blir fanget opp der, og de estimerte  $\beta$  fanger opp variasjonen innen næringene over tid. Den forskjellen mellom næringene som er korrelert med forklaringsvariablene blir i fasteffektmodellen inkludert i de næringsspesifikke konstantleddene. I minste kvadrats metode havner all slik heterogenitet i feilleddet. Dermed kan det tyde på at det er de næringsspesifikke konstantleddene som fanger opp så mye av variasjonen at estimatene fra de to modellene blir så ulike som de er.

Ved å uttrykke fasteffektmodellen eksplisitt i STATA, det vil si som minste kvadrats metode med konstantledd for alle næringene, rapporteres alle disse konstantleddene. Dette gir muligheten til å inspisere hvordan heterogeniteten arter seg. Veldig mange av konstantleddene er signifikante på 1 % nivå, noe som gjør det tydelig at det er stor variasjon mellom næringene. Med olje- og gassnæringen som referansenæring, er det slik at det som betegnes som industrinæringene stort sett har større utgifter knyttet til utdanningsgruppe 1 enn referansenæringen, med tekstil-, gjenvinningsnæringene og lær og lærprodukter som næringer med spesielt høye innslag av arbeidskraft med kun grunnskoleutdanning. På den annen side bruker offentlig administrasjon og utdanningssektoren lite ressurser på samme arbeidskraftsgruppe. På samme måte er det stor variasjon i hvor mye ressurser som blir brukt på arbeidskraft fra de forskjellige utdanningsgruppene mellom næringene. Ettersom det å inkludere de næringsspesifikke fastleddene innebærer å inkludere flere forklaringsvariabler i modellen er det ikke overraskende at en del koeffisienter endrer seg. Dette tyder på at det er en sammenheng mellom fastleddene og forklaringsvariablene for øvrig.

For å teste minste kvadrats metode mot fasteffektmodellen, rapporteres det i STATA en F-test dersom fasteffektmodellen estimeres eksplisitt som en fasteffektmodell. Ettersom jeg har estimert modellene som tilsynelatende uavhengige regresjoner, rapporterte STATA ikke disse F-testene. Disse F-testene tester om alle feilleddene/ konstantleddene er lik null samtidig. Ved å estimere regresjonene som vanlige fasteffektmodeller i STATA, konkluderer programmet med at fasteffektmodellen bør foretrekkes foran minste kvadrats metode i alle tilfellene på alle signifikansnivå.<sup>33</sup>

**Tabell 13: Etterspørselastisiteter basert på regresjon med fasteffektmodellen**

	<i>Relsmal</i>	<i>Reltot73</i>	<i>V</i>	<i>K</i>	<i>W1</i>	<i>W2</i>	<i>W3</i>	<i>W4</i>
L1	-0,251 (0,328)	25,476 (2,299)***	0,061 (0,024)**	0,005 (0,004)	-0,553 (0,124)***	0,133 (0,149)	0,180 (0,115)	0,240 (0,062)***
L2	0,425 (0,176)**	-6,391 (2,299)***	-0,020 (0,013)	0,004 (0,002)*	0,029 (0,033)	-0,526 (0,076)***	0,207 (0,055)***	0,042 (0,027)
L3	-0,631 (0,390)	-5,771 (5,112)	-0,021 (0,029)	-0,011 (0,005)**	0,111 (0,071)	0,583 (0,156)***	-0,796 (0,152)***	0,102 (0,060)*
L4	-0,809 (0,428)*	15,738 (5,590)***	0,079 (0,032)**	-0,007 (0,005)	0,293 (0,076)***	0,233 (0,148)	0,201 (0,119)*	-0,727 (0,090)***

Størrelsene betegner for *relsmal* og *reltot73* prosentvis endring i sysselsetting som svar på ett prosentpoengs endring i de foran nevnte variablene. For de andre variablene er det prosentvis endring i sysselsetting etter en prosent høyere verdiskaping og fast realkapital.

<sup>33</sup> Med testobservatorer fra 197 til 635, som gir forkastning av nullhypotese på alle signifikansnivå.

Tabell 13 rapporterer elastisitetene kalkulert på resultatene fra fasteffektmodellen. Egenpriselastisitetene er negative for alle utdanningsgruppene. Dette er et godt tegn med tanke på konkavitetskravet til kostnadsfunksjonen. Samtidig kommer det frem at alle typene arbeidskraft er substitutter for hverandre (alle krysspriselastisitetene er positive).

Styrken ved å bruke SUR-rammeverket er muligheten til å teste hypoteser som går på kryss av regresjonene. De mest sentrale testene er hovedsakelig om offshoring påvirker utdanningsgruppene på forskjellig vis, og tilsvarende for forskning og utvikling. Disse testene gir gjennomgående klare resultater. I denne analysen er effekten av offshoring på etterspørselen etter arbeidskraft ulik for de fire utdanningsgruppene. Det kommer også frem at effekten av offshoring i forhold til den av endringer i forsknings- og utviklingsvariabelen  $\ln Y_{it}$ , er Det er også aktuelt å teste om effekten av offshoring og forskning og utvikling er ulik. I tillegg er testing av de pålagte restriksjonene viktig, det bør undersøkes om implementeringen av disse restriksjonene

Testing av restriksjonene innebærer å teste om de holder selv i modeller som ikke er pålagt dem. Restriksjonene om at priseffektene skal summes til null i hver regresjon kunne ikke forkastes i modellene med utdanningsgruppene 3 og 4, mens for gruppene 1 og 2, ble de forkastet på alle signifikansnivå. Symmetrirestriksjonen forkastes på alle signifikansnivå. Dette innebærer at restriksjonene må pålegges analysen for at modellen skal kunne tolkes som strukturell.

Videre er det aktuelt å teste om offshoring påvirker de forskjellige typene arbeidskraft på ulike måter. Offshoring har ulik påvirkning på de fire utdanningsgruppene, signifikant på 10 % nivå. Det samme gjelder forskning og utvikling, men testen forkaster på alle signifikansnivå nullhypotesen om likhet. For utdanningsgruppe 3 er påvirkningene av offshoring og forskning og utvikling ikke signifikant forskjellig fra hverandre, men for de andre gruppene er det signifikante (på alle signifikansnivå) forskjeller mellom disse to effektene.



## **5.2.2 Tilfeldigeffektmodellen**

Vanligvis brukes Hausman-testen til å undersøke hvilken av paneldatamodellene som bør brukes (se 3.3.3). Jeg har utført denne testen, men resultatene som STATA rapporterer er ikke entydige, og det er generelt problemer med resultatene. Det første poenget som kommer frem er at, ettersom det er fire regresjoner som utføres, bør det være samme konklusjonen i de fire Hausman-testene som utføres. Dette er ikke tilfellet. Ifølge Hausman-testen, bør fasteffektmodellen brukes for utdanningsnivå 2 i regresjonsanalysen, mens tilfeldigeffektmodellen bør foretrekkes for gruppene 1 og 3. For utdanningsgruppe 4 får jeg negativ chi-kvadratverdi. Jeg har ikke fått til å estimere paneldatateknikkene med restriksjoner slik at det var mulig å bruke Hausman-testen på de regresjonene jeg faktisk bruker. Men konklusjonene fra Hausman-testene gjort på paneldatamodellene uten restriksjoner, ga så uklare resultater, at jeg velger å bruke fasteffektmodellen. Avgjørelsen er i tillegg basert på det faktum at fasteffektmodellen er den mest robuste av de to modellene, ettersom den gir forventingsrette estimater uansett sammenhengen mellom individspesifikk heterogenitet og regressor, men er også motivert av praktiske aspekter som at det er mulig å estimere modellen med restriksjoner, samt å teste hypoteser i SUR-rammeverket i STATA.

## **5.3 Problemer med modellene**

Økonometrisk arbeid dreier seg ofte om å identifisere problemer som kan forstyrre analysen. I dette delkapittelet vurderes noen av de formene for problemer som ble introdusert i kapittel 3.4, side 24.

### **5.3.1 Multikolinearitet**

Multikolinearitet er et problem som kan oppstå dersom det er sterk korrelasjon mellom forklaringsvariablene. Hva som menes med sterk korrelasjon er ikke helt entydig definert, men korrelasjonen mellom de forskjellige inntektene kan være et problem.

**Tabell 14: Korrelasjon mellom forklaringsvariablene**

	<i>lnW1</i>	<i>lnW2</i>	<i>lnW3</i>	<i>lnW4</i>	<i>Relsmal</i>	<i>Reltot73</i>	<i>lnV</i>	<i>lnK</i>
<i>lnW1</i>	1							
<i>lnW2</i>	0,9389	1						
<i>lnW3</i>	0,8575	0,9068	1					
<i>lnW4</i>	0,7443	0,7341	0,8327	1				
<i>Relsmal</i>	0,0752	0,0485	0,1946	0,1501	1			
<i>Reltot73</i>	0,1004	0,0702	-0,0496	-0,1105	-0,169	1		
<i>lnV</i>	0,0782	0,0544	-0,0671	0,0643	-0,3586	-0,1502	1	
<i>lnK</i>	0,0727	0,0605	-0,0445	0,0919	-0,257	-0,1516	0,7017	1

Det fremkommer i Tabell 14 at det er høy korrelasjon mellom de ulike inntektene. Også korrelasjonen mellom fast realkapital og verdiskapning er høy. Korrelasjonen mellom fast realkapital og verdiskapning kommer an på hvordan verdiene for fast realkapital i årene 2001 og 2003 – 2005 blir definert,<sup>34</sup> men det er uansett høy korrelasjon mellom de to. Den høye korrelasjonen mellom inntektene kan føre til multikolaritetsproblemer, som innebærer at sammenhengen mellom disse variablene og kostnadsandelene kan bli unøyaktig estimert, fordi det er vanskelig å si hvilken lønn det er som påvirker kostnadsandelene. Dette betyr ikke at estimatene for variabler som ikke er korrelerte med ”problemvariablene” er forventningsskjeve. Det kunne vært mulig å overkomme disse problemene ved å finne andre variabler som fanger opp inntekt til de forskjellige gruppene men som samtidig ikke er like sterkt korrelert mellom gruppene, men dette er en tidkrevende prosess, som ikke er får plass i denne oppgaven. Jeg velger å inkludere inntekt i regresjonene, men tar forbehold for multikolaritetsproblemer når jeg tolker resultatene for disse variablene.

### 5.3.2 Endogenitet

Som nevnt i kapittel 3.4 er problemer med endogenitet alvorlige i økonometrisk analyse. En form for dette, simultanitet, oppstår i det en eller flere av de uavhengige variablene er felles bestemt med den avhengige variabelen, i dette tilfellet gjelder dette inntektsvariablene og sysselsettingsnivåene. Dette gjør at jeg bruker de såkalt *laggede* verdiene for inntekt som proxy for inntekt.<sup>35</sup> Korrelasjonen mellom inntekt i to påfølgende perioder ligger på henholdsvis 0,98, 0,995, 0,984 og 0,963, men den laggede inntekten er ikke den *samme* som er inkludert i kostnadsandelene.

<sup>34</sup> Som nevnt i datakapittelet, mangler observasjoner på fast realkapital for de nevnte årene.

<sup>35</sup> Jeg estimerer også en regresjon med sysselsettingsandelene som avhengige variabler (se kapittel 5.6, side 70).

Utelatte variabler er et problem som ofte oppstår i økonometriske undersøkelser. Det kan være mye som forklarer størrelsen på den avhengige variabelen, og det er stort sett ikke mulig å inkludere alle relevante faktorer. Derfor er det viktig å være bevisst hvilke problemer som oppstår dersom viktige faktorer er utlatt, og hva som gjør at det oppstår problemer. Resultatene som blir rapportert er forventningsskjevne dersom det er et problem med utelatte variabler. Per definisjon oppstår problemet i det en relevant variabel, det vil si en variabel som er med å forklare avhengig variabel, som i tillegg er korrelert med andre forklaringsvariabler, er utelatt.

Effekten av utelatte variabler er altså at koeffisientene blir forventningsskjevne. Dersom det er klart hvilke variabler som er utelatt, og det er kjent hvordan sammenhengen er mellom både utelatt variabel og avhengig variabel, og mellom utelatt variabel og forklaringsvariablene, er det mulig å si noe om retningen av skjevheten. I denne oppgaven er det ikke så klart hva som er utelatt, men det kom frem i overgangen mellom minste kvadrats metode og fasteffektmodellen at heterogenitet var utelatt ettersom koeffisientene endret seg såpass mye som de gjorde. Det er fortsatt muligheter for at det er utelatte variabler som kan skape problemer, men ved bruk av fasteffektmodellen er analysen i mindre grad utsatt for dette problemet, ettersom det individspesifikke feilleddet tillates å korrelere med regressor, i motsetning til hva som er tilfellet i tilfeldigeffektmodellen.

Som en videreføring av argumentasjonen om utelatte variabler heterogenitet, kan feilspesifisert funksjonsform nevnes. Det er implisitt i modellen antatt at alle næringer i utvalget har samme  $\beta$ . Ved estimering i fasteffektmodellen er det tatt høyde for noe heterogenitet. Det er likevel mulig å se for seg at fenomen som offshoring og forskning og utvikling kan påvirke næringene på ulike måter. Det også aktuelt at sammenhengene mellom forklaringsvariablene og avhengig variabel ikke er slik som modellert i denne oppgaven. Dette er nok et aspekt som det kunne vært aktuelt å undersøke nærmere, men det får bli ved en senere anledning. Selv om problemene knyttet til endogenitet og spesifisering er alvorlige, ser jeg meg nødt til å stole på robustheten til fasteffektestimatorene i denne oppgaven.

### 5.3.3 Autokorrelasjon

Det er ofte klare sammenhenger i både uavhengige og avhengige variabler over tid, og det er ikke uvanlig å overse visse typer dynamikk ved bruk av vanlige paneldatamodeller som fasteffektmodellen og tilfeldigeffektmodellen. Dermed er det viktig å kontrollere om autokorrelasjon er et problem i de aktuelle regresjonene. Det er estimert fire regresjoner, og jeg presenterer autokorrelasjonsmatrisene for dem i Tabell 15.

**Tabell 15: Autokorrelasjon i fasteffektmodellen**

S1				S2					
$\varepsilon_{\hat{t}}$	$\varepsilon_{\hat{t-1}}$	$\varepsilon_{\hat{t-2}}$	$\varepsilon_{\hat{t-3}}$	$\varepsilon_{\hat{t}}$	$\varepsilon_{\hat{t-1}}$	$\varepsilon_{\hat{t-2}}$	$\varepsilon_{\hat{t-3}}$		
$\varepsilon_{\hat{t}}$	1			$\varepsilon_{\hat{t}}$	1				
$\varepsilon_{\hat{t-1}}$	0,5210	1		$\varepsilon_{\hat{t-1}}$	0,6951	1			
$\varepsilon_{\hat{t-2}}$	0,1070	0,4467	1	$\varepsilon_{\hat{t-2}}$	0,1913	0,4044	1		
$\varepsilon_{\hat{t-3}}$	-0,0618	0,1712	0,5175	1	$\varepsilon_{\hat{t-3}}$	-0,1400	0,0709	0,5854	1

S3				S4					
$\varepsilon_{\hat{t}}$	$\varepsilon_{\hat{t-1}}$	$\varepsilon_{\hat{t-2}}$	$\varepsilon_{\hat{t-3}}$	$\varepsilon_{\hat{t}}$	$\varepsilon_{\hat{t-1}}$	$\varepsilon_{\hat{t-2}}$	$\varepsilon_{\hat{t-3}}$		
$\varepsilon_{\hat{t}}$	1			$\varepsilon_{\hat{t}}$	1				
$\varepsilon_{\hat{t-1}}$	0,7452	1		$\varepsilon_{\hat{t-1}}$	0,5628	1			
$\varepsilon_{\hat{t-2}}$	0,1242	0,2739	1	$\varepsilon_{\hat{t-2}}$	0,2178	0,5127	1		
$\varepsilon_{\hat{t-3}}$	-0,1957	-0,0871	0,5367	1	$\varepsilon_{\hat{t-3}}$	-0,0813	0,1521	0,5762	1

Ettersom sammenhengene mellom 4 påfølgende perioder måles, forsvinner en del observasjoner. Antall observasjoner i denne tabellen er 352.

Med verdier på mellom 0,5 og 0,75, er det såpass sterk korrelasjon mellom feilleddene i påfølgende perioder at det sannsynligvis er problemer med autokorrelasjon. Ettersom korrelasjonen mellom feilleddene avtar med antall lag, kan det se ut som det her er snakk om autokorrelasjon av første orden. Det er også slik at korrelasjonen mellom to påfølgende perioders feilledd ikke er lik for alle periodene, disse er i tabellen markert med kursiv. Neste trinn er å gjennomføre en test for autokorrelasjon, og Verbeek (2004) foreslår flere, blant annet en versjon av Durbin-Watson-testen. Denne testen er basert på en nullhypotese om at  $\rho$  i en regresjon hvor feilleddene fra fasteffektmodellen regresseres på tilsvarende feilledd i perioden før, vil være lik null;

$$e_{it} = \rho e_{it-1} + v_{it}$$

Med  $e$  som feilledd i de estimerte fasteffektmodellene, og  $v$  som vanlig feilledd i denne regresjonen.  $\rho$  betegner sammenhengen mellom feilledd i påfølgende perioder. Denne estimeringen innebærer en lineær sammenheng mellom feilleddene i påfølgende perioder, som er lik mellom alle perioder. Dette er, som det kommer frem av Tabell 15, ikke en nøyaktig antagelse. Testobservatoren dannes på følgende måte;<sup>36</sup>

$$dw_{\rho} = \frac{\sum_{i=1}^N \sum_{t=2}^T (\hat{e}_{it} - \hat{e}_{i,t-1})^2}{\sum_{i=1}^N \sum_{t=1}^T \hat{e}_{it}^2} .$$

Med nullhypotesen:  $\rho = 0$ , at det ikke er noen autokorrelasjon, og alternativhypotese, ensidig sådan, at  $\rho > 0$  eller  $\rho < 0$ . Størrelsen har kritiske verdier tilpasset antallet observasjoner per periode, antall perioder og antall parametre i den opprinnelige regresjonen. De kritiske verdiene danner et intervall hvor konklusjonen er uklar dersom testobservatoren inntar en verdi i dette intervallet, men med lavere testobservator enn intervallet vil nullhypotesen beholdes. Ved innsetting av verdiene for de respektive feilleddene, predikert fra fasteffektmodellene, får jeg testobservatorer med verdiene 0,664, 0,593, 0,694 og 0,630, henholdsvis for gruppene 1, 2, 3 og 4. Alle er under den nedre delen av kritisk intervall i Durbin-Watson-testen, som i dette tilfellet går til rundt 1,88.<sup>37</sup> Dermed kan jeg forkaste nullhypotesen, og med rimelig sikkerhet anta at det er problemer med autokorrelasjon. Disse resultatene stemmer overens med konklusjonene fra Baltagi-Wu-testen og Arellano-Bond-test for autokorrelasjon rapportert i STATA. Arellano-Bond-testen tester også for autokorrelasjon av andre orden, og beholder nullhypotesen om fravær av dette, i alle fire modellene (se Tabell 18, side 67). Jeg konkluderer med at denne saken er verdt å forfølge (se kapittel 5.4), gitt mulighetene STATA gir til alternative modeller som tester for denne typen dynamikk.

### 5.3.4 Andre økonometriske problemer

Når det gjelder situasjoner som kan føre til inferensproblemer er heteroskedastisitet et typisk eksempel. Som nevnt tidligere (kapittel 3.4) er dette situasjoner der variansen i feilleddet

<sup>36</sup> Med hatt som betegnelse på estimerte verdier.

<sup>37</sup> For  $T=10$ ,  $N=100$  og ni estimerte koeffisienter er den nedre grensen 1,878, ettersom det er enda færre observasjoner i dette datasettet er den virkelige kritiske verdien noe lavere, men fortsatt over alle testobservatorene jeg har kalkulert.

avhenger av størrelsen på forklaringsvariablene. Ved å estimere feilleddene i modellen, det mulig, som en første undersøkelse, å måle korrelasjonen mellom kvadrerte feilledd og forklaringsvariablene.<sup>38</sup> Korrelasjon her tyder på at det kan være problemer med heteroskedastisitet. Vanlig omgås dette problemet ved å spesifisere en robust-kommando i STATA. Dette innebærer å implementere en metode for robuste standardavvik som gjør bot på problemet ved ukjent form på heteroskedastisiteten. Dessverre har jeg ikke hatt mulighet til å bruke denne kommandoen i forbindelse med estimering i SUR-rammeverket.

Normalfordeling av feilleddene er nødvendig for at t- og F-tester skal være til å stole på. Med bakgrunn i sentralgrenseteoremet, vil tilstrekkelig store utvalg vanligvis sørge for at dette kravet blir oppfylt. Ekholm og Hakkala (2005) bruker *bootstrapede* feilledd i sin analyse. De skriver at dette gjøres på grunn av at antagelsen om normalfordelte feilledd er brutt i denne typen data. Etter å ha utført regresjonene i fasteffektmodellen har jeg estimert feilledd, som jeg undersøkte fordelingen til. Etersom normalfordelte feilledd ikke er et krav for store utvalg, og det automatisk er oppfylt som følge av sentralgrenseteoremet, er ikke dette et problem i denne oppgaven. Definisjonen på hva som er et stort utvalg er ikke klar, men med 51 næringer over ti år, og dermed med totalt 510 observasjoner, er datasettet i denne undersøkelsen alle fall i nærheten av noe som kan kalles et stort utvalg.<sup>39</sup>

## **5.4 Dynamiske modeller**

Med en sterk sammenheng i de respektive kostnadsandelene mellom hver periode, og tegn som peker mot problemer med autokorrelasjon, vil jeg videre undersøke om det hjelper på inferensen å ta høyde for at en form for dynamikk påvirker resultatene. I STATA er det en enkel prosedyre å estimere regresjonene under antagelsen om at det er en underliggende prosess som styrer størrelsene på kostnadsandelene, og at denne prosessen er autoregressiv av første orden. Dette passer bra med funnene i Tabell 15, som viser at prosessen ser ut til å *være* av nettopp første orden. Dessverre tillater ikke kommandoen *xtreg* i STATA implementering av restriksjonene som er implementert i modellene over. Dette gjør at det ved sammenligning av modellene blir usikkert om endringene kommer av fraværet av restriksjoner eller på grunn av inkludert dynamikk. Derfor velger jeg å sammenligne resultatene fra den dynamiske modellen med resultatene jeg får ved å estimere fasteffektmodellen uten restriksjoner.

---

<sup>38</sup> Se appendiks B, Appendikstabell 1.

<sup>39</sup> I appendiks C vises fire tabeller som indikerer at feilleddene er tilnærmet normalfordelt.

Forskjellen mellom de statiske modellene er ikke så store når det gjelder de strukturelle variablene, men effekten av endringer i inntekt er noe større. Dette gjør at jeg hovedsakelig fokuserer på forskjeller på de strukturelle elastisitetene.

**Tabell 16: Kostnadsandelene regressert med fasteffektmodellen med AR(1)-feilledd**

	<i>S1</i>	<i>S2</i>	<i>S3</i>	<i>S4</i>
Relsmal	-0,0002 (0,0303)	0,2156 (0,0708)***	-0,0652 (0,0607)	-0,1579 (0,0315)***
Reltot73	1,0577 (0,4370)**	-1,1522 (1,018)	0,4107 (0,8786)	-0,4666 (0,4524)
lnV	-0,0033 (0,0027)	0,0000 (0,0064)	0,0080 (0,0055)	-0,0046 (0,0028)
lnK	0,0001 (0,0005)	0,0015 (0,0012)	-0,0011 (0,001)	-0,0005 (0,0005)
lnW1_lag	-0,0123 (0,0113)	0,0075 (0,0262)	-0,0666 (0,0228)***	0,0713 (0,0116)***
lnW2_lag	-0,0179 (0,0223)	-0,0483 (0,0519)	0,0681 (0,0447)	-0,0014 (0,0231)
lnW3_lag	0,0182 (0,0115)	0,0496 (0,0268)*	-0,0431 (0,0233)*	-0,0214 (0,0119)*
lnW4_lag	0,0146 (0,0073)**	0,0003 (0,0171)	0,0011 (0,0148)	-0,0179 (0,0076)**
Konstantledd	0,1118 (0,0817)	0,4106 (0,1807)**	0,6756 (0,1882)***	-0,2243 (0,0758)***
Tidsdummyer	Ja	Ja	Ja	Ja
Rho-AR	0,71	0,73	0,67	0,74
R <sup>2</sup> : Innen	0,224	0,206	0,175	0,234
Mellom	0,096	0,0008	0,001	0,003
I alt	0,112	0,002	0,002	0,006
Observasjoner	404	404	404	404
Grupper	51	51	51	51
Obs. per gr.	7,9	7,9	7,9	7,9
F(15,338)	6,5	5,86	4,79	6,88
Prob>F	0	0	0	0

Tabell 16 viser koeffisientene i fasteffektmodellen med autoregressiv prosess av første orden. Det første som kan legges merke til er at *rho-AR* ligger på rundt 0,7 i alle regresjonene. Dette betyr at det er en relativt sterk tidstrend i de avhengige variablene. Disse estimatene er det mulig å teste i STATA. STATA kan rapportere Baltagi-Wu-lbi-testobservatoren,<sup>40</sup> og for

<sup>40</sup> *lbi* står for *locally best invariant*. Denne testen er en modifisert dw-test, med tilsvarende kritiske verdier.

regresjonene over, forkastes nullhypotesene om at  $\rho = 0$ , med verdi på testobservatoren rundt 1 for alle fire regresjonene.

Sammenhengen mellom endring i offshoring og kostnadsandelene går i samme retning i de to modellene, spesielt er effekten av relsmal på utdanningsgruppe 2 og 4 nå med større sikkerhet ulike null, med en endring fra statistisk signifikans på 5 % nivå til signifikans på 1 % nivå. Koeffisientene blir begge noe mindre, men forskjellen mellom dem er blitt noe større. Effekten av endringer i utgifter på forsknings- og utviklingstjenester på kostnadsandelene for utdanningsnivåene 1, 2 og 4 er i fasteffektmodellen signifikante på 1 % nivå. I modellen hvor det er tatt høyde for den autoregressive prosessen er det kun kostnadsandelen til utdanningsgruppe 1 som på 5 % signifikansnivå kan sies å bli påvirket av dette.

Sammenhengen mellom utgifter til forsknings- og utviklingstjenester og størrelsen på kostnadsandelene til utdanningsnivå 2 og 4 er i denne modellen ikke signifikant ulike null.

**Tabell 17: Sammenligning av etterspørselastisiteter uten restriksjoner, FE og AR(1)**

	Fasteffektmodellen				Fasteffektmodell med AR(1) prosess			
	Relsmal	Reltot73	lnV	lnK	Relsmal	Reltot73	lnV	lnK
L1	-0,2019 (0,2963)	26,17 (3,85)***	0,0595 (0,0220)***	0,0058 (0,0035)*	-0,0017 (0,2420)	8,44 (3,49)**	-0,0262 (0,0218)	0,0008 (0,0041)
L2	0,5696 (0,1622)***	-6,72 (2,11)***	-0,0209 (0,0120)*	0,0044 (0,0019)**	0,3785 (0,1242)***	0,57 (1,79)	0,0000 (0,0112)	0,0027 (0,0021)
L3	-0,5918 (0,3628)	-5,55 (4,72)	-0,0208 (0,0269)	-0,0118 (0,0042)***	-0,3221 (0,3001)	2,03 (4,34)	0,0395 (0,0271)	-0,0056 (0,0048)
L4	-0,8268 (0,3982)**	16,27 (5,18)***	0,0841 (0,0295)***	-0,0082 (0,0046)*	-1,5385 (0,3070)***	-4,55 (4,41)	-0,0445 (0,0276)	-0,0053 (0,0053)

Størrelsene betegner for relsmal og reltot73 prosentvis endring i sysselsetting som svar på ett prosentpoengs endring i de foran nevnte variablene. For de andre variablene er det prosentvis endring i sysselsetting etter en prosent høyere verdiskapning og fast realkapital.

Påvirkningen på kostnadsandelene av verdien på fast realkapital, som var signifikant på minst 10 % nivå for alle utdanningsgruppene i fasteffektmodellen, er nå insignifikant, selv på 10 % nivå, i alle fire regresjonene. Også effekten av økt verdiskapning på kostnadsandelene er statistisk insignifikant forskjellig fra null i alle regresjonene.

Det er noen forskjeller mellom estimatene for lønn i de to modellene. Det er til en viss grad en tendens til at endringen i estimatene er at størrelsen på elastisitetene øker. Og for etterspørselen etter sysselsatte fra utdanningsnivå 2 med hensyn på lønn til utdanningsgruppe 1 forandrer koeffisienten fortegn



Det er også en mulighet til å se på etterspørselselastisitetene på lang sikt, ved å vekte elastisitetene som kommer fra den autoregressive modellen med  $\rho$  fra samme modellen,  $\beta/(1-\rho)$ . Dette er imidlertid en noe tvilsom fremgangsmåte, ettersom antagelsen om at kostnadsfunksjonen regresjonene er basert på bare er gyldig på kort sikt.

Selv om det ser ut til at inklusjon av dynamikk øker presisjonen i resultatene, er det ikke dermed sagt at dette er den riktige formen for dynamikk. Det kan være at prosessen er av en høyere orden, eller at den er slik at kostnadsandelene fra foregående perioder bør inkluderes som forklaringsvariabel. Jeg har estimert denne typen dynamikk i STATA, modellert med Arellano-Bond-estimatoren.

**Tabell 18: Kostnadsandelene regressert med Arellano-Bond-modellen**

	S1	S2	S3	S4
Si	0,6480 (0,0428)***	0,4320 (0,1065)***	0,5104 (0,1199)***	0,5709 (0,0604)***
Relsmal	-0,0017 (0,0385)	0,2475 (0,0758)***	-0,0628 (0,0753)	-0,2330 (0,0379)***
Reltot73	1,4956 (0,5593)***	-3,1108 (1,1297)***	-0,0202 (1,0872)	0,4775 (0,5682)
InV	0,0016 (0,0031)	-0,0111 (0,0058)*	0,0047 (0,0064)	-0,0036 (0,0031)
InK	0,0003 (0,0004)	0,0013 (0,0010)	-0,0009 (0,0008)	-0,0009 (0,0005)*
InW1_lag	-0,0638 (0,0149)***	0,0401 (0,0320)	-0,0751 (0,0281)***	0,0621 (0,0143)***
InW2_lag	0,0434 (0,0264)*	-0,1070 (0,0657)	0,1301 (0,0536)*	0,0146 (0,0267)
InW3_lag	0,0176 (0,0149)	0,1027 (0,0306)***	-0,0804 (0,0292)*	-0,0313 (0,0156)**
InW4_lag	0,0077 (0,0081)	0,0152 (0,0177)	0,0038 (0,0162)	-0,0305 (0,0093)***
Konstantledd	0,0001 (0,0019)	-0,0014 (0,0039)***	0,0041 (0,0035)	-0,0025 (0,0020)
Tidsdummyer	Ja	Ja	Ja	Ja
Observasjoner	404	404	404	404
Parametre	18	18	18	18
Wald chi	505,95	246,43	105,57	165,66
Prob. over ID	0,000	0,667	0,006	0,000
Prob A-B autokov 1	0,000	0,000	0,000	0,000
Prob A-B autokov 2	0,101	0,989	0,915	0,143

Det som kommer frem ved å estimere dynamikken på denne måten, er at effekten av økninger i relsmal går i samme retning som i de foregående modellene, men at forskjellen mellom utdanningsgruppe 2 og 4 er noe større enn i AR(1)-modellen, med estimerte koeffisienter på henholdsvis 0,25 og -0,23, begge statistisk signifikante på 1 % nivå. Effekten av økninger i reltot<sup>73</sup> er signifikant for utdanningsgruppe 1 og 2, med negativ effekt for gruppe 1 og positiv for gruppe 2. Endringer i fast realkapital og verdiskapning er stort sett insignifikant forskjellig fra null for alle gruppene. Dette med unntak av negative effekter av økninger i verdiskapning på gruppe 2 og i fast realkapital på gruppe 4, begge signifikant ulike null på 10 % nivå. Lønnseffektene er igjen ikke sammenlignbare med resultatene i de statiske modellene på grunn av fravær av restriksjoner. Estimerte *kovarianser* mellom forklaringsvariabler i påfølgende perioder ligger mellom 0,43 (for gruppe 2) og 0,65 (for gruppe 1), alle signifikante på 1 % nivå.

STATA rapporterer resultat fra Sargans test for overidentifiserende restriksjoner hvor nullhypotesen forkastes for regresjonene 1, 3 og 4 (i denne sammenheng testes restriksjoner på kovarians mellom feilledd over tid, som legges for å skape nok instrumentvariabler til å estimere sammenhengene i regresjonen, med p-verdier rapport i Tabell 18 som prob over ID). Det at hypotesen ikke forkastes for regresjonen for gruppe 2, fører til at resultatene fra denne regresjonen må tolkes med forsiktighet.

## **5.5 Kvantifisering av resultatene**

De estimerte elastisitetene viser endringer i etterspørsel etter arbeidskraft fra de forskjellige utdanningsgruppene ved endringer i aktuelle variabler. Selv om det er interessant å observere ulikhetene i størrelse mellom dem, er de ikke direkte sammenlignbare når det gjelder etterspørselsendringer i *antall* sysselsatte. Dette kommer av at det ved en proSENTS endring i sysselsettingen av utdanningsgruppe 2 er snakk om rundt 5700 stykker (menn). Til sammenligning er det i gjennomsnitt rundt 190 000 menn fra utdanningsgruppe 3 sysselsatt i hvert år, som gir at en reduksjon på en prosent tilsvarer 1900 personer.

På samme måte er en økning på ett prosentpoeng, en veldig forskjellig mengde for forskning og utvikling og for offshoring. Offshoring ligger i gjennomsnitt på rundt 3,6 % i undersøkelsesperioden, og reltot<sup>73</sup>, forsknings- og utviklingsstørrelsen, på rundt 0,4 %. Dette innebærer at en økning på et prosentpoeng er store relative endringer for begge, men spesielt

stort for reltot73. Endringer i lønn, verdiskapning og fast realkapital regnes i prosent i forhold til utgangspunktet, så disse kan sammenlignes.

Elastisitetene rapportert i Tabell 13 er evaluert i gjennomsnittlige sysselsettingsandeler. Datasettet inneholder data på alle menn med pensjonsgivende inntekt i Norge i den aktuelle perioden. For å kvantifisere størrelsen på elastisitetene kan gjennomsnittlig sysselsetting tas til følge. Det kan gjøres ved å undersøke hvordan etterspørselen etter arbeidskraft endrer seg i antall sysselsatte ved en prosents endring i forklaringsvariablene. Tabell 19 viser slike resultater.

**Tabell 19: Effekt av en prosents økning i forkl.var., ettersp.el. basert på FE-modellen**

	<i>Relsmal</i>	<i>Reltot73</i>	<i>V</i>	<i>K</i>	<i>W1</i>	<i>W2</i>	<i>W3</i>	<i>W4</i>
L1	-11	127***	76***	6	-689***	166	224	299***
L2	87**	-147***	-115	23*	166	-3016***	1187***	241
L3	-42	-43	-40**	-21	209	1097***	-1498***	192*
L4	-23*	50***	63	-6***	233***	185	160*	-578***

Elastisiteter vurdert for gjennomsnittlige størrelser, og gir endring i antall etterspurte sysselsatte i de fire utdanningsgruppene. Statistisk signifikansnivå indikert med stjerner, \*\*\* for 1 % nivå, \*\* 5 % nivå og \* 10 % nivå.

Tallene i Tabell 19 er som nevnt basert på gjennomsnittlige størrelser, men gir en indikasjon på hvor stor effekt endringer i de forskjellige forklaringsvariablene har på etterspørselen etter de ulike gruppene arbeidskraft. Det kommer frem at det etterspørres 87 færre menn fra utdanningsgruppe 2 nasjonalt, dersom relsmal øker med 1 prosent. Det kommer frem av tabellen at noen av effektene er på grensen av det som kan kalles økonomisk signifikante, særlig gjelder dette effekten av endringer i fast realkapital.

Verdien av import av innsatsvarer som andel av produksjon har over perioden fra 1996 til 2005 økt med 43,5 %, eller 1,41 prosentpoeng. Dette innebærer at, alt annet likt, har offshoring ført til økt etterspørsel etter arbeidskraft fra utdanningsgruppe 2 med 3786 sysselsatte i dette panelet. Tilsvarende tall for utdanningsgruppe 4 er en reduksjon på 1000. Ettersom datasettet inkluderer alle menn med pensjonsgivende inntekt i denne perioden, er disse tallene til en viss grad generaliserbare til å gjelde for populasjonen som en helhet, det vil si at endringene i antall sysselsatte nok er større når kvinner tas med. Men når det er sagt vil inklusjon av kvinner i analysen føre til at kvinnedominerte næringer vil tillegges forskjellig vekt, og at estimerte koeffisienter kan endres.

Det er i tillegg et poeng å understreke koblingen mellom endret etterspørsel etter arbeidskraft og endringer i lønninger. Målet med oppgaven er som kjent å undersøke om økt offshoring fører til økte lønnsforskjeller i Norge. Dette er blitt gjort ved å estimere etterspørselen etter arbeidskraft. Endringer i etterspørselen etter arbeidskraft gir seg utslag antall sysselsatte og i inntekten til de som er sysselsatte. Slik som undersøkelsen er lagt opp, blir det naturlig å se på etterspørselastisitetene som indikatorer på hvordan lønnsutviklingen er. Teorien utviklet av Feenstra og Hanson (1996a, 1996b, osv.) tar utgangspunkt i antagelsen om full sysselsetting, som innebærer at alle relative etterspørselsendringer gir seg utslag i endringer i relative lønninger. Det kan diskuteres om situasjonen i Norge oppfyller antagelsen om full sysselsetting. Det er til dels tilfellet i 2008, men har ikke vært tilfellet i hele undersøkelsesperioden.

## **5.6 Kontroll og alternative spesifiseringer**

Det er vanlig prosedyre i andre arbeider å kontrollere resultatene fra kostnadsandelsregresjonene ved å kjøre tilsvarende regresjoner på sysselsettingsandelene. Tilsvarende tabell presenteres i appendikset. Tabellen viser at estimering av sysselsettingsandeler gir tilnærmet like svar som ved estimering av kostnadsandelene. Dette er et tegn på at estimeringen av de foregående regresjonene er meningsfulle.

Hvilke næringer som undersøkes kan være viktig for hvilke resultater som avdekkes. Det kom frem at det er store variasjoner mellom næringene da jeg undersøkte konstantleddene i fasteffektmodellen. Dette betyr at det kan være at resultatene for visse grupper av næringer er ulike de for andre. Blant annet undersøker Falk og Koebel (2002) de tyske industrinæringene. Ekholm og Hakkala (2005) undersøker 20 produksjonsindustrier. For å sjekke om det at jeg har inkludert de fleste næringene i mine regresjoner har gjort at resultatene mine avviker fra deres, har jeg utført et sett regresjoner hvor jeg inkluderte 21 næringer, som kan betegnes som industrier, eller produserende næringer.<sup>41</sup> Resultatene fra disse regresjonene er også presentert i appendikset. Det kommer frem at resultatene går i samme retning som i regresjonene utført med hele utvalget.

---

<sup>41</sup> Næringer med NACE-kode 17 til 37. Se appendiks B.

Dersom det brede målet på offshoring inkluderes, er retningen på sammenhengene de samme som i den opprinnelige modellen. Appendikstabell 6 viser regresjonsresultatene.

Ettersom det i teorien og tidligere forskningsarbeider er lagt så stor vekt på forskning og utvikling som en forklaring på økte lønnsforskjeller, er det en mulighet for at variabelen tot73, det vil si innkjøpte tjenester fra forsknings- og utviklingsnæringen, ikke fanger opp nok av den utviklingen som foregår i næringene. Av den grunn er det interessant å bruke bedre mål på forskning og utvikling. SSB publiserer tall som viser totale kostnader på forskning og utvikling i næringene, men som nevnt i datakapittelet, er ikke disse tallene publisert på samme aggregeringsnivå som tallene i kryssløpstabellene. Av den grunn valgte jeg i hovedanalysen å bruke tot73 som proxy for forskning og utvikling. For likevel å kunne kontrollere om valget av forsknings- og utviklingsvariabel har påvirket resultatene, har jeg omgruppert datasettet mitt slik at det passer til forsknings- og utviklingstallene fra SSB. Dette innebærer at datasettet ikke lenger er gruppert med tall for 51 næringer. Regresjonene utført på disse forsknings- og utviklingstallene inneholder derimot tall for 31 grupper og enkelt næringer. Resultatene fra dette settet regresjoner er mindre sikre.<sup>42</sup> For det første er det nå kun sammenhengen mellom forskning og utvikling og kostnadsandelen til utdanningsgruppe 1 som er statistisk signifikant. For det andre har ikke retningen, det vil si fortegnene, på de fire koeffisientene endret seg mellom de to modellene. Det er heller ikke slik at estimatene for de andre koeffisientene endrer seg når jeg endrer regresjonene på denne måten.

---

<sup>42</sup> Det er verdt å nevne at det mangler observasjoner på denne FoU-variabelen for årene 1996, 1998 og 2000. Jeg har antatt at variabelen tar gjennomsnittsverdiene av årene før og etter disse tre årene. Dette gjør resultatene mer usikre, da estimatene ikke er gjort kun på de sanne verdiene, men også på estimerte.

## 6 Oppsummering

I denne masteroppgaven har jeg undersøkt hvordan de ulike utdanningsgruppenes kostnadsandeler blir påvirket av endringer i offshoring, innkjøp av forsknings- og utviklingstjenester, verdiskapning og fast realkapital og de fire utdanningsgruppenes lønnsnivå. Jeg har brukt ulike tilnæringsmåter til å studere problemstillingen. Første tilnærming var estimering av regresjonene med minste kvadrats metode. Resultatene ga en klar indikasjon på at sammenhengen mellom offshoring og etterspørsel etter heterogen arbeidskraft var av en slik karakter at økt offshoring førte til økte ulikheter i arbeidsmarkedet. Dette kom frem i de positive koeffisientene for offshoring (relsmal) for utdanningsgruppene 3 og 4, og de negative koeffisientene for utdanningsgruppene 1 og 2. Dette vil si at resultatene antyder at det ved økt offshoring vil være slik at etterspørselen etter arbeidskraft med lavere utdanning vil reduseres relativt til etterspørselen etter arbeidskraft med høyere utdanning, og at lønnsforskjellene vil øke.

Ettersom datasettet er organisert som et paneldatasett, med opplysninger om forhold i 51 næringer over 13 år, vil ikke estimatoren fra vanlig minste kvadrats metode fange opp de næringsspesifikke kjennetegnene i datasettet. Derfor var det nærliggende å bruke paneldatateknikker for å få sikrere estimat for sammenhengene mellom offshoring og lønnsgapet. Fra analysen av datasettet med fasteffektmodellen, kom det frem at sammenhengene ikke er så tydelige som når det blir estimert med minste kvadrats metode, de endres betydelig. Både basert på estimatene fra fasteffektmodellen uten dynamikk, men estimert med Zellners metode, og fra fasteffektmodellen med autoregressiv prosess av første orden, er resultatet at etterspørselen etter arbeidstakere fra utdanningsgruppe 2 *øker* med økt offshoring.

Sammenhengen er signifikant i begge modellene på henholdsvis 5 % nivå og 1 % nivå. Sammen med funnet av at etterspørselen etter arbeidskraft fra utdanningsnivå 4 faller med økt offshoring, som også er signifikant på henholdsvis 5 % og 1 % nivå, overrasker dette. Som tidligere påpekt går disse funnene på tvers av det Ekholm og Hakkala (2005) fant for Sverige, og det Hijzen *et al.* (2005) fant for Storbritannia.

Grunnene til at resultatene skiller seg slik fra Sverige og Storbritannia kan være mange. Det kan være at det rett og slett er slik at den norske økonomien er ulik den svenske og den britiske. Samtidig er det heller ikke usannsynlig at det er noe ved datasettet i denne oppgaven som skiller seg fra Ekholm og Hakkala og Hijzen *et al.* sine. For eksempel hadde de førstnevnte gode opplysninger om opprinnelsesland for importen av innsatsvarer. Funnene deres tyder på at offshoring samlet sett øker forskjellen mellom deres gruppe som korresponderer med gruppe 2 i denne studien, og gruppen som korresponderer med grupper 3 og 4 i mitt datasett. Funnene deres, i det de skiller mellom høyinntekts- og lavinntektsland, er at offshoring til høyinntektsland ikke fører til statistisk signifikante forskjeller, mens offshoring til lavinntektsland fører til større forskjeller. Selv om jeg hadde tilgang til handelstall etter opprinnelsesland, var de ikke av en slik kvalitet at jeg kunne gjennomføre regresjoner med *nok* observasjoner til at sammenhengene kunne estimeres på en robust måte. Dersom temaet skal undersøkes videre, er dette noe som det godt kan legges mer vekt på.

Resultatene Hijzen *et al.* (2005) rapporterer er langt mer entydige, med gjennomgående resultater som peker på økende forskjeller ved økt offshoring. De har heller ikke undersøkt hvor offshoringen blir kanalisert, men har, som jeg har gjort i denne oppgaven, brukt kryssløpstabellene med informasjon om samlet import fra alle land. Det som skiller deres fremgangsmåte fra min, er at de inkluderer *materialer* i kostnadsfunksjonen på samme måte som arbeidskraft, men dette burde ikke påvirke effektene av offshoring på de forskjellige typene arbeidskraft, i alle fall ikke forholdet dem i mellom. Hvilke variabler som inkluderes i regresjonene er viktige. Det er selvsagt flere ting som påvirker kostnadene i en næring enn de som er inkludert i denne undersøkelsen. Videre forskning på området kan innebære å estimere en mer fullstendig kostnadsfunksjon.

Resultatene i denne oppgaven hviler til en viss grad på antagelsen om funksjonsform på kostnadsfunksjonen, ettersom denne bestemmer sammenhengen mellom de avhengige og de uavhengige variablene, som skal estimeres. Valget av translog kostnadsfunksjon ble blant annet motivert av to praktiske aspekt, den gir kostnadsandeler som *lineære* funksjoner av priser og strukturvariabler. Det gjør det mulig å estimere sammenhengene ved hjelp av minste kvadrats metode. Et annet praktisk argument som taler for translog kostnadsfunksjon, er at det er en enkel prosedyre å kvantifisere resultatene fra regresjonsanalysen. Et tredje argument er at fremgangsmåten er brukt av Feenstra og Hanson (2003), Ekholm og Hakkala (2005) og Hijzen *et al.* (2005), som er artikler jeg følger tett. Samtidig viser Falk og Koebel (2001) at

denne funksjonsformen ikke nødvendigvis er den beste, og foreslår videre bruk av en Box-Cox-kostnadsfunksjon. Valg av funksjonstype kan påvirke resultatene, blant annet fordi det kan lede til feilspesifisering av modellen.

Det er vanlig å undersøke produksjonsnæringene spesielt. Men, etter også å ha kontrollert for hvilke næringer som er inkludert i regresjonene og for valg av forsknings- og utviklingstall, mener jeg likevel det er riktig å konkludere med at funnene i denne studien er robuste, for den sammenhengen som er undersøkt, nemlig offshoring totalt, det vil si til alle deler av verden.

Før det undersøkes hvorfor offshoring slår ut slik som det er funnet i denne oppgaven, ser jeg for meg at videre arbeid på dette felt vil innebære å teste resultatene fra denne oppgaven nærmere. Undersøkelsene kan gjøres bedre ved å søke opp disaggregerte data på offshoringaktiviteter, for å oppnå bedre presisjon i forhold til hvilke oppgaver som offshores, men også for å oppnå en mindre grov inndeling av næringene som offshorer. Det vil være ønskelig å kombinere dette med hvilke land og regioner det offshores til, slik at det er mulig å kontrollere for kompetansenivå hos handelspartner.

Resultatene fra denne undersøkelsen synes forholdsvis klare. Med utgangspunkt i de sammenhengene som her er analysert, blir konklusjonen at økt offshoring totalt sett, ser ut til å gi mindre lønnsforskjeller mellom utdanningsgruppene i Norge.



## Referanser

Berndt, E. R. og D. O. Wood (1975): "Technology, Prices, and the Derived Demand for Energy," *The Review of Economics and Statistics*, Vol. 57, No. 3, 259 - 268

Berman, E., J. Bound og Z. Griliches (1994), "Changes in the Demand for Skilled Labor within U.S. Manufacturing: Evidence from the Annual Survey of Manufacturers", *The Quarterly Journal of Economics*, Vol. 109, No. 2. 367-397

Berman, E., J. Bound og S. Manchin (1998): "Implications of Skill-Biased Technological Change: International Evidence", *The Quarterly Journal of Economics*, Vol. 113, No. 4, 1245-1279

Borjas, G. J., R. B. Freeman og L. F. Katz (1992): "On the Labor Market Effects of Immigration and Trade," i G. Borjas og R. Freeman, red., *Immigration and the work force*. Chicago: University of Chicago Press, 213 – 244

Bound, J. og G. Johnson (1992): "Changes in the Structure of Wages in the 1980's: An Evaluation of Alternative Explanations", *The American Economic Review*, Vol. 82, No. 3, 371-392

Chintagunta, P., T. Erdem, P. E. Rossi og M. Wedel (2006): "Structural Modeling in Marketing: Review and Assessment", *Marketing Science*, Vol. 25, No. 6, 604-616

Diewert, W. E. og T. J. Wales (1987): "Flexible Functional Forms and Global Curvature Conditions", *Econometrica*, Vol. 55, No. 1, 43-68

Ekholm, K. og K. Hakkala (2005), "The Effect of Offshoring on Labor Demand: Evidence from Sweden", *Working Paper No 654 IUI* (Industriens Utredningsinstitut, nå Institutet för näringslivsforskning, INF) Stockholm

Falk, M. og B.M. Koebel (2002), "Outsourcing, Imports and Labour Demand", *Scandinavian Journal of Economics*, 104, 567-586

Feenstra, R. C. (2004): *Advanced International Trade: theory and evidence*. Princeton: Princeton University Press

Feenstra, R. C. og G. H. Hanson (1996a): "Foreign Investment, Outsourcing and Relative Wages," in Robert C. Feenstra, Gene M. Grossman og Douglas A. Irwin, red., *The Political Economy of Trade Policy: Papers in Honor of Jagdish Bhagwati*, MIT Press, 1996, 89-127

Feenstra, R.C. og G. H. Hanson (1996b), "Globalization, Outsourcing, and Wage Inequality", *The American Economic Review*, 86, 240-244

Feenstra, R.C. og G. H. Hanson (1999): "The Impact of Outsourcing and High-Technology Capital on Wages: Estimates for the United States, 1979-1990". *The Quarterly Journal of Economics*, 114, 907-940

Feenstra, R.C. og G. H. Hanson (2003): "Global Production Sharing and Rising Inequality: a Survey of Trade and Wages", i E. Kwan Choi og J. Harrigan, red., *Handbook of International Trade*. Oxford: Blackwell

Geishecker, I. og H. Goerg (2008), "Winners and losers: a micro-level analysis of international outsourcing and wages", *Canadian Journal of Economics*, Vol. 41, No. 1, 234 - 270

Griffiths, W. E., R. C. Hill, G. G. Judge, (1993): *Learning and Practicing Econometrics* John Wiley & Sons, Inc.

Hijzen, A., H. Goerg, og R. Hine (2005), "International outsourcing and the skill structure of labour demand in the united kingdom", *The Economic Journal*, 115, 860-878

Katz, L. F. og K. M. Murphy (1992), "Changes in relative wages, 1963-1987: supply and demand factors", *The Quarterly Journal of Economics*, Vol. 107, No. 1. pp. 35-78

Krugman, P. R. (1999), "Technology, trade and factor prices", *Journal of International Economics*, 50 (2000) 51-71

Lawrence, R. Z. og M. J. Slaughter (1993), "International Trade and American Wages in the 1980s: Giant Sucking Sound or Small Hiccup?", *Brookings Papers on Economic Activity*, Vol. 1993, No. 2, 161-226 (med kommentarer fra Hall og Topel)

Leamer E. E. (1997): "In Search of Stolper-Samuelson Linkages between International Trade and Lower Wages", i Susan Collins, red., *Imports, Exports and the American Worker* Brookings, 141 - 214, publisert på nytt i E. E. Leamer, *International Economics*, Worth Publishers, 2001

Sachs, J. D. og H. J. Shatz (1994): "Trade and Jobs in U.S", Manufacturing, *Brookings Papers on Economic Activity*, Vol. 1994, No. 1., 1-84

Staffolani, S., A. Lo Turco, A. F. Presbitero og C. Broccolini, (2007): "International Outsourcing, vs. ICT in Explaining the Wagegap in Italian Manufacturing" *Quaderni di Ricerca*, n. 299, Dipartimento di Economia, Università Politecnica delle Marche

Varian, H. R.(1992): *Microeconomic Analysis 3rd ed.* W. W. Norton & Company, Inc., New York

Verbeek, M., (2004): *A guide to modern econometrics, 2nd ed.* John Wiley & Sons

Wooldridge, J. M., (2006): *Introductory econometrics: A Modern Approach 3rd ed.* Thompson Learning, South-Western College Publishing, Stamford

## Appendiks A

### Appendiks A.1: Utledning av egenskapene til kostnadsfunksjonen

Inspirert av Hijzen *et al.* (2005), Ekholm og Hakkala (2005), Berndt og Wood (1975) osv., er den transcendentalt logaritmiske (translog) kostnadsfunksjonen i denne oppgaven gitt som følger:

$$\begin{aligned}
 \ln C_i(w, x, z) &= \alpha_0 + \sum_{j=1}^J \beta_j \ln w_{ij} + \sum_{k=1}^K \beta_k \ln x_{ik} + \sum_{r=1}^R \beta_r z_{ir} \\
 &+ \frac{1}{2} \sum_{j=1}^J \sum_{s=1}^J \beta_{js} \ln w_{ij} \ln w_{is} + \frac{1}{2} \sum_{k=1}^K \sum_{l=1}^K \beta_{kl} \ln x_{ik} \ln x_{il} + \frac{1}{2} \sum_{r=1}^R \sum_{p=1}^R \beta_{rp} z_{ir} z_{ip} \\
 &+ \sum_{j=1}^J \sum_{k=1}^K \beta_{jk} \ln w_{ij} \ln x_{ik} + \sum_{j=1}^J \sum_{r=1}^R \beta_{jr} \ln w_{ij} z_{ir} + \sum_{k=1}^K \sum_{r=1}^R \beta_{kr} \ln x_{ik} z_{ir}
 \end{aligned} \tag{30}$$

For å vise at denne funksjonstypen under gitte forutsetninger er lineært homogen i faktorpriser ( $w_{ij}$ ) viser jeg omregningen når alle faktorprisene øker med en positiv faktor  $t$ .

$$\begin{aligned}
 \ln C_i(tw, x, z) &= \alpha_0 + \sum_{j=1}^J \beta_j \ln(tw_{ij}) + \sum_{k=1}^K \beta_k \ln x_{ik} + \sum_{r=1}^R \beta_r z_{ir} \\
 &+ \frac{1}{2} \sum_{j=1}^J \sum_{s=1}^J \beta_{js} (\ln tw_{ij})(\ln tw_{is}) + \frac{1}{2} \sum_{k=1}^K \sum_{l=1}^K \beta_{kl} \ln x_{ik} \ln x_{il} + \frac{1}{2} \sum_{r=1}^R \sum_{p=1}^R \beta_{rp} z_{ir} z_{ip} \\
 &+ \sum_{j=1}^J \sum_{k=1}^K \beta_{jk} (\ln tw_{ij}) \ln x_{ik} + \sum_{j=1}^J \sum_{r=1}^R \beta_{jr} (\ln tw_{ij}) z_{ir} + \sum_{k=1}^K \sum_{r=1}^R \beta_{kr} \ln x_{ik} z_{ir}
 \end{aligned} \tag{31}$$

Ved bruk av regelen at logaritmen av et produkt er lik summen av logaritmene av hver faktor deler jeg opp logaritmeuttrykkene.

$$\begin{aligned}
 \ln C_i(tw, x, z) &= \alpha_0 + \underbrace{\sum_{j=1}^J \beta_j (\ln w_{ij} + \ln t)}_a + \sum_{k=1}^K \beta_k \ln x_{ik} + \sum_{r=1}^R \beta_r z_{ir} \\
 &+ \frac{1}{2} \underbrace{\sum_{j=1}^J \sum_{s=1}^J \beta_{js} (\ln w_{ij} + \ln t)(\ln w_{is} + \ln t)}_b + \frac{1}{2} \sum_{k=1}^K \sum_{l=1}^K \beta_{kl} \ln x_{ik} \ln x_{il} + \frac{1}{2} \sum_{r=1}^R \sum_{p=1}^R \beta_{rp} z_{ir} z_{ip} \\
 &+ \underbrace{\sum_{j=1}^J \sum_{k=1}^K \beta_{jk} (\ln w_{ij} + \ln t) \ln x_{ik}}_c + \underbrace{\sum_{j=1}^J \sum_{r=1}^R \beta_{jr} (\ln w_{ij} + \ln t) z_{ir}}_d + \sum_{k=1}^K \sum_{r=1}^R \beta_{kr} \ln x_{ik} z_{ir}
 \end{aligned} \tag{32}$$

Videre blir a, b, c og d delt opp som vist under, i tillegg til at rekkefølgen på noen summer blir endret og at konstante ledd blir flyttet ut av summasjonene:

$$\begin{aligned}
\ln C_i(tw, x, z) = & \alpha_0 + \underbrace{\sum_{j=1}^J \beta_j \ln w_{ij} + \ln t \sum_{j=1}^J \beta_j}_{a} + \sum_{k=1}^K \beta_k \ln x_{ik} + \sum_{r=1}^R \beta_r z_{ir} \\
& + \frac{1}{2} \left( \underbrace{\ln t \sum_{j=1}^J (\ln w_{ij}) \sum_{s=1}^J \beta_{js} + \ln t \sum_{s=1}^J (\ln w_{is}) \sum_{j=1}^J \beta_{js} + (\ln t)^2 \sum_{j=1}^J \sum_{s=1}^J \beta_{js} + \sum_{j=1}^J \sum_{s=1}^J \beta_{js} \ln w_{ij} \ln w_{is}}_b \right) \\
& + \frac{1}{2} \sum_{k=1}^K \sum_{l=1}^K \beta_{kl} \ln x_{ik} \ln x_{il} + \frac{1}{2} \sum_{r=1}^R \sum_{p=1}^R \beta_{rp} z_{ir} z_{ip} \\
& + \underbrace{\sum_{j=1}^J \sum_{k=1}^K \beta_{jk} \ln w_{ij} \ln x_{ik} + \ln t \sum_{k=1}^K \ln x_{ik} \sum_{j=1}^J \beta_{jk}}_c \\
& + \underbrace{\sum_{j=1}^J \sum_{r=1}^R \beta_{jr} \ln w_{ij} z_{ir} + \ln t \sum_{r=1}^R z_{ir} \sum_{j=1}^J \beta_{jr} + \sum_{k=1}^K \sum_{r=1}^R \beta_{kr} \ln x_{ik} z_{ir}}_d
\end{aligned} \tag{33}$$

I alle disse fire komponentene er det en del som er lik som i den opprinnelige translog(e) kostnadsfunksjonen. Disse samles med de andre uforandrede leddene og omformuleres til  $\ln C_i(w, x, z)$ . Videre inkluderes de tretten (evt. fem, ettersom det summeres over J) antagelsene som skal til for at funksjonen er lineær homogen i priser;

$$\sum_{j=1}^J \beta_j = 1, \quad \sum_{j=1}^J \beta_{js} = \underbrace{\sum_{s=1}^J \beta_{js}}_r = 0, \quad \sum_{j=1}^J \beta_{jk} = 0 \quad \text{og} \quad \sum_{j=1}^J \beta_{jr} = 0 \tag{34}$$

Med disse antagelsene blir den translog(e) kostnadsfunksjonen slik:

$$\begin{aligned}
\ln C_i(tw, x, z) = & \underbrace{\ln t \sum_{j=1}^J \beta_j}_{=1} + \frac{1}{2} \ln t \sum_{j=1}^J (\ln w_{ij}) \underbrace{\sum_{s=1}^J \beta_{js}}_{=0(\text{problem})} + \frac{1}{2} \ln t \sum_{s=1}^J (\ln w_{is}) \underbrace{\sum_{j=1}^J \beta_{js}}_{=0} \\
& + \frac{1}{2} (\ln t)^2 \sum_{s=1}^J \sum_{j=1}^J \underbrace{\beta_{js}}_{=0} + \ln t \sum_{k=1}^K (\ln x_{ik}) \underbrace{\sum_{j=1}^J \beta_{jk}}_{=0} + \ln t \sum_{r=1}^R (z_{ir}) \underbrace{\sum_{j=1}^J \beta_{jr}}_{=0} + \ln C_i(w, x, z)
\end{aligned} \tag{35}$$

Ved å følge antagelsene er det klart at uttrykket for den translog(e)kostnadsfunksjonen blir som følger;

$$\ln C_i(tw, x, z) = \ln t + \ln C_i(w, x, z) = \tag{36}$$

Dermed vil en eksponentialisering av uttrykkene gi et uttrykk for den ”underliggende” kostnadsfunksjonen, som viser at den under antagelsene er homogen av grad 1 i faktorpriser.

$$C_i(tw, x, z) = e^{(\ln t + \ln C_i(w, x, z))} = t C_i(w, x, z) \tag{37}$$

## Appendiks A.2: Automatisk oppfylte betingelser

Restriksjonene utenom restriksjon c i teksten og symmetrirestriksjonene er nødvendigvis oppfylt ved estimering av kostnadsandelsfunksjonene i minste kvadrats metode. Dette kommer av:

$$\bar{S}_j = \beta_{0j} + \beta_{js} \bar{W} + \beta_{jk} \bar{X} + \beta_{jr} \bar{Z} + \bar{e}_{ji} \text{ med } \bar{e}_j = 0 \forall j, i \quad (38)$$

Summering over de fire kostnadsandelene gir dette:

$$1 = \underbrace{\sum_{j=1}^4 \bar{S}_j}_{(i)} = \sum_{j=1}^4 \beta_{0j} + \sum_{j=1}^4 \beta_{js} \bar{W} + \sum_{j=1}^4 \beta_{jk} \bar{X} + \sum_{j=1}^4 \beta_{jr} \bar{Z}$$

Per definisjon er (i) alltid lik 1, og  $\beta$ -ene gjelder for alle X, med forventingsverdien av feilleddene  $e_j$  er lik 0. Dette gir at sammenhengen mellom summen av kostnadsandeler kan uttrykkes slik:

$$\sum_{j=1}^4 S_{ji} = 1 = \sum_{j=1}^4 \beta_{0j} + \sum_{j=1}^4 \beta_{js} W_i + \sum_{j=1}^4 \beta_{jk} X_i + \sum_{j=1}^4 \beta_{jr} Z_i + \sum_{j=1}^4 e_{ji}$$

Ettersom denne sammenhengen må holde for alle *individer / næringer* i datasettet, er det to måter likheten holder, gitt betingelsen om null i betinget gjennomsnitt for feilleddet holder:<sup>43</sup>

1)  $\sum_{j=1}^4 \beta_{0j} \neq 1$  og  $\sum_{j=1}^4 \beta_{jw} \neq 0$ ,  $\sum_{j=1}^4 \beta_{jx} \neq 0$  og/eller  $\sum_{j=1}^4 \beta_{jz} \neq 0$ , W, X, og/eller Z må være konstant

2)  $\sum_{j=1}^4 \beta_{0j} = 1$ ,  $\sum_{j=1}^4 \beta_{jw} = 0$ ,  $\sum_{j=1}^4 \beta_{jx} = 0$  og  $\sum_{j=1}^4 \beta_{jz} = 0$ , W, X og Z kan variere

Dersom 1) skal holde, er det **én** kombinasjon av W, X og Z som minimerer de kvadrerte avvikene, og som oppfyller kravet om likhet med null. Hvis en av forklaringsvariablene endrer seg, vil ikke likheten holde. Ettersom koeffisientene gjelder for alle individ, og individene har ulike verdier på forklaringsvariablene, er det kun 2) som kan holde for denne typen regresjon, med lineære koeffisienter. Dette innebærer at estimering av regresjonene hver for seg vil gi koeffisienter som stemmer med likheten, og at alle restriksjonene utenom ”problemrestriksjonen” er nødvendigvis oppfylt.

---

<sup>43</sup>  $E\left(\sum_{j=1}^4 e_j\right) = \sum_{j=1}^4 E(e_j) = 0$

## Appendiks A.3 Derivasjon av kostnadsfunksjonen.

Kostnadsfunksjonen er gitt som translog kostnadsfunksjon fra teksten:

$$\begin{aligned} \ln C_i(w, x, z) &= \alpha_0 + \underbrace{\sum_{j=1}^J \beta_j \ln w_{ij}}_a + \sum_{k=1}^K \beta_k \ln x_{ik} + \sum_{r=1}^R \beta_r z_{ir} \\ &+ \underbrace{\frac{1}{2} \sum_{j=1}^J \sum_{s=1}^J \beta_{js} \ln w_{ij} \ln w_{is}}_b + \frac{1}{2} \sum_{k=1}^K \sum_{l=1}^K \beta_{kl} \ln x_{ik} \ln x_{il} + \frac{1}{2} \sum_{r=1}^R \sum_{p=1}^R \beta_{rp} z_{ir} z_{ip} \\ &+ \underbrace{\sum_{j=1}^J \sum_{k=1}^K \beta_{jk} \ln w_{ij} \ln x_{ik}}_c + \underbrace{\sum_{j=1}^J \sum_{r=1}^R \beta_{jr} \ln w_{ij} z_{ir}}_d + \sum_{k=1}^K \sum_{r=1}^R \beta_{kr} \ln x_{ik} z_{ir}. \end{aligned}$$

Leddene a, b, c og d inneholder uttrykk for lønn, og er dermed som er interessante. Ved derivasjon med hensyn på logaritmen til lønn til gruppe  $h$ , får vi;

$$\frac{\partial \ln C_i(w, x, z)}{\partial \ln w_{ih}} = \beta_h + \underbrace{\sum_{s \neq h} \frac{1}{2} \beta_{hs} \ln w_{is} + \sum_{j \neq h} \frac{1}{2} \beta_{jh} \ln w_{ij}}_b + \beta_{hh} \ln w_{ih} + \sum_{k=1}^K \beta_{hk} \ln x_{ik} + \sum_{r=1}^R \beta_{hr} z_{ir}$$

Leddene som ikke inneholdt lønn i kostnadsfunksjonen faller bort, mens a, c og d er enkle å ha med å gjøre. Ledd b er derimot noe spesielt. Ettersom lønn inngår i begge deler av dobbelsummen, må det deriverte uttrykket kunne settes opp som over. Del b inneholder dermed to summer over  $s$  og  $j$  for alle  $J$  utenom  $h$ , og ett ledd med egenprisindeffekten for  $h$ . Dette kan slås sammen, slik at begge disse summene også inkluderer  $h$ ;

$$\frac{\partial \ln C_i(w, x, z)}{\partial \ln w_{ih}} = \beta_h + \underbrace{\sum_{s=1}^J \frac{1}{2} (\beta_{hs} + \beta_{sh}) \ln w_{is}}_q + \sum_{k=1}^K \beta_{hk} \ln x_{ik} + \sum_{r=1}^R \beta_{hr} z_{ir}. \quad (39)$$

Estimering av dette uttrykket gir ikke identifiserbare  $\beta$ -verdier, kun gjennomsnitt av parene. Dersom antagelsen om symmetri inkluderes, det vil si at vi antar at  $\beta_{js} = \beta_{sj}$ , kan kostnadsandelsfunksjonen til faktor  $h$  uttrykkes som følger:

$$\frac{\partial \ln C_i(w, x, z)}{\partial \ln w_{ih}} = S_h = \beta_h + \underbrace{\sum_{s=1}^J \beta_{hs} \ln w_{is}}_b + \sum_{k=1}^K \beta_{hk} \ln x_{ik} + \sum_{r=1}^R \beta_{hr} z_{ir}, \quad \beta_{js} = \beta_{sj}.$$

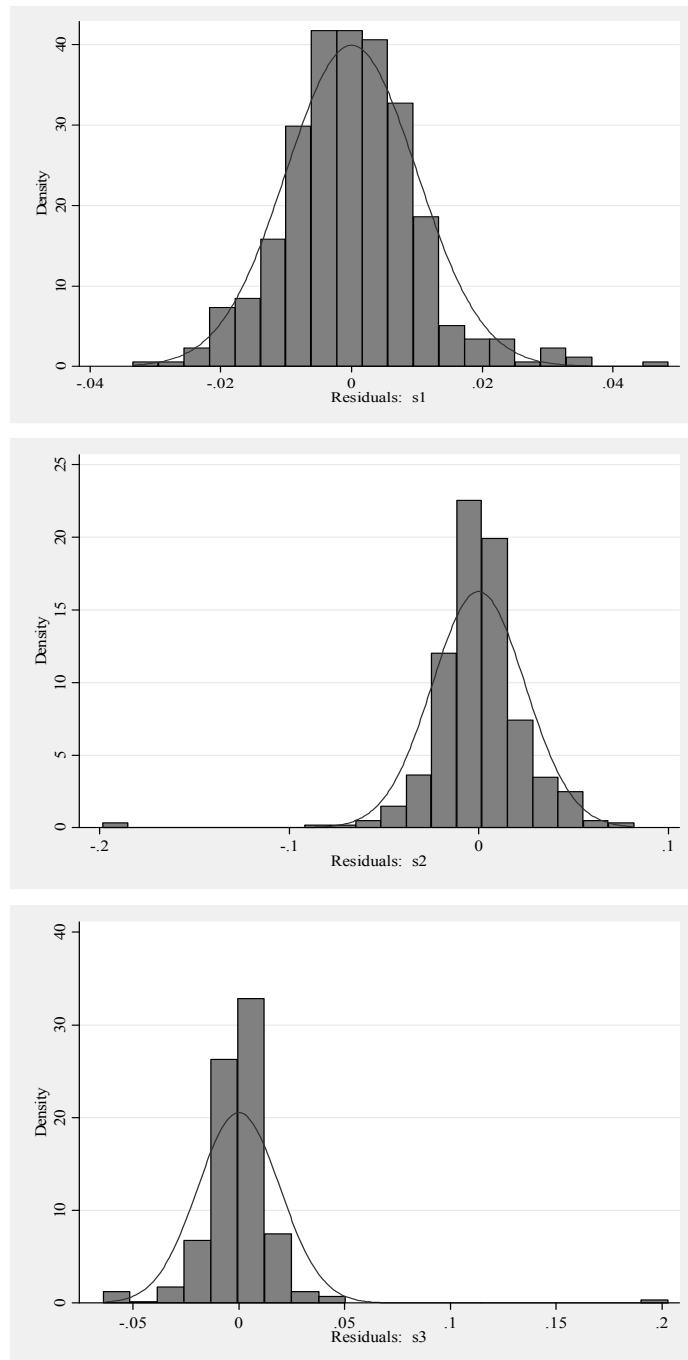
Problemet er bare at den erklærte restriksjonen implementert i STATA er at  $q$  i ligning (39) skal være lik mellom regresjonene, det vil si at gjennomsnittet av  $\beta_{hs}$  og  $\beta_{sh}$  skal være det samme over forskjellige regresjoner. Dessverre identifiseres ikke eksplisitte verdier for  $\beta_{hs}$  og  $\beta_{sh}$ , det er uendelig mange måter som oppfyller at gjennomsnittet av de to er likt i over regresjoner, ikke bare at de to underliggende koeffisientene er like hverandre

## Appendiks B

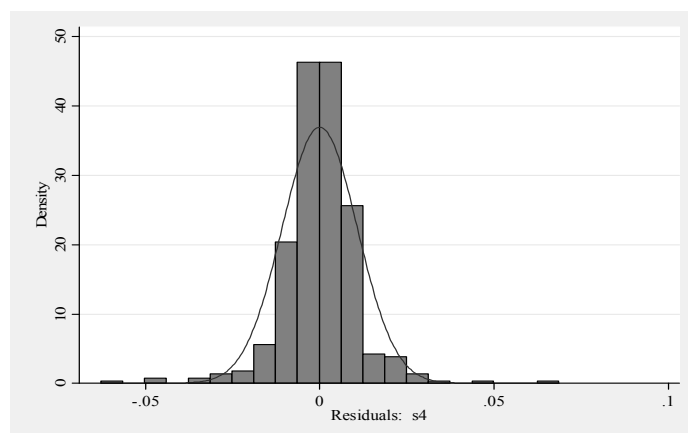
Nr.	Nace	Navn på næring
1	01	Products of agriculture, hunting and related services
2	02	Products of forestry, logging and related services
3	05	Fish and other fishing products; services incidental of fishing
4	10	Coal and lignite; peat
5	11	Crude petroleum and natural gas; services incidental to oil and gas extraction excluding surveying
6	12	Uranium and thorium ores
7	13	Metal ores
8	14	Other mining and quarrying products
9	15	Food products and beverages
10	16	Tobacco products
11	17	Textiles
12	18	Wearing apparel; furs
13	19	Leather and leather products
14	20	Wood and products of wood and cork (except furniture); articles of straw and plaiting materials
15	21	Pulp, paper and paper products
16	22	Printed matter and recorded media
17	23	Coke, refined petroleum products and nuclear fuels
18	24	Chemicals, chemical products and man-made fibres
19	25	Rubber and plastic products
20	26	Other non-metallic mineral products
21	27	Basic metals
22	28	Fabricated metal products, except machinery and equipment
23	29	Machinery and equipment n.e.c.
24	30	Office machinery and computers
25	31	Electrical machinery and apparatus n.e.c.
26	32	Radio, television and communication equipment and apparatus
27	33	Medical, precision and optical instruments, watches and clocks
28	34	Motor vehicles, trailers and semi-trailers
29	35	Other transport equipment
30	36	Furniture; other manufactured goods n.e.c.
31	37	Secondary raw materials
32	40	Electrical energy, gas, steam and hot water
33	41	Collected and purified water, distribution services of water
34	45	Construction work
35	50	Trade, maintenance and repair services of motor vehicles and motorcycles; rt. sale of automotive fuel
36	51	Wholesale trade and commission trade services, except of motor vehicles and motorcycles
37	52	Retail tr. services, except of motor vehic. and motorcycl; repair services of persn. and househ. goods
38	55	Hotel and restaurant services
39	60	Land transport; transport via pipeline services
40	61	Water transport services
41	62	Air transport services
42	63	Supporting and auxiliary transport services; travel agency services
43	64	Post and telecommunication services
44	65	Financial intermediation services, except insurance and pension funding services
45	66	Insurance and pension funding services, except compulsory social security services
46	67	Services auxiliary to financial intermediation
47	70	Real estate services
48	71	Renting services of machinery and equipment without operator and of personal and household goods
49	72	Computer and related services
50	73	Research and development services
51	74	Other business services
52	75	Public administration and defence services; compulsory social security services
53	80	Education services
54	85	Health and social work services
55	90	Sewage and refuse disposal services, sanitation and similar services
56	91	Membership organisation services n.e.c.
57	92	Recreational, cultural and sporting services
58	93	Other services
59	95	Private households with employed persons

# Appendiks C

## Appendiks C.1: FORDELING AV PREIDKERTE FEILLEDD







## Appendiks C.2: Heteroskedastisitet: Breusch-Pagan....

**Appendikstabell 1: KORRELASJON MELLOM ESTIMERT KVADRERT FEILLEDD OG X**

	$\hat{e}_{s1}^2$	$\hat{e}_{s2}^2$	$\hat{e}_{s3}^2$	$\hat{e}_{s4}^2$
$\hat{e}_{s1}^2$	1			
$\hat{e}_{s2}^2$	0,1045	1		
$\hat{e}_{s3}^2$	-0,023	0,9639	1	
$\hat{e}_{s4}^2$	0,0604	0,109	0,0043	1
Relsmal	-0,0304	-0,0208	-0,0418	0,3281
Reltot73	0,0977	-0,0953	-0,1009	0,0016
lnV	-0,2814	-0,0217	0,0408	-0,2127
lnK	-0,1525	0,0281	0,0686	-0,1569
lnW1_lag	-0,1408	-0,0018	-0,0094	0,0603
lnW2_lag	-0,1363	-0,0209	-0,0194	-0,0128
lnw3_lag	-0,1112	-0,005	-0,0118	0,0342
lnW4_lag	-0,1197	0,1028	0,098	0,0042
Tid1998	0,0559	0,1149	0,0957	0,0796
Tid1999	0,0051	-0,0242	-0,026	-0,0216
Tid2000	-0,0778	-0,0508	-0,028	-0,0761
Tid2001	-0,0719	-0,0461	-0,0311	-0,0409
Tid2002	-0,1432	-0,0518	-0,0341	-0,0481
Tid2003	-0,0432	-0,0418	-0,0296	-0,0462
Tid2004	0,0166	-0,0221	-0,0219	-0,0375
Tid2005	0,0476	-0,0127	-0,02	0,0392

## Appendiks D

Appendikstabell 2: VANLIG FE UTEN RESTRIKSJONER

	S1	S2	S3	S4
Relsmal	-0,0253 (0,0371)	0,2299 (0,0924)**	-0,1198 (0,0734)	-0,0849 (0,0409)**
Reltot73	3,2802 (0,4828)***	-3,8268 (1,2011)***	-1,1228 (0,9546)	1,6694 (0,5312)***
lnV	0,0075 (0,0028)***	-0,0119 (0,0069)*	-0,0042 (0,0054)	0,0086 (0,0030)***
lnK	0,0007 (0,0004)*	0,0025 (0,0011)**	-0,0024 (0,0009)***	-0,0008 (0,0005)*
lnW1_lag	0,0270 (0,0144)*	-0,0225 (0,0359)	-0,0331 (0,0286)	0,0286 (0,0159)*
lnW2_lag	-0,1264 (0,0234)***	0,1517 (0,0582)***	-0,0222 (0,0462)	-0,0031 (0,0257)
lnW3_lag	-0,0085 (0,0151)	0,0576 (0,0377)	-0,0235 (0,0299)	-0,0255 (0,0167)
lnW4_lag	0,0085 (0,0082)	-0,0584 (0,0204)***	0,0256 (0,0162)	0,0242 (0,0090)***
Konstantledd	1,2844 (0,2875)***	-1,0442 (0,7151)	0,9933 (0,5684)*	-0,2336 (0,3163)
Tidsdummyer	Ja	Ja	Ja	Ja
Industridummyer	Ja	Ja	Ja	Ja
Observasjoner	455	455	455	455
Parametre	66	66	66	66
RMSE	0,0098	0,0243	0,0193	0,0107
R <sup>2</sup>	0,978	0,977	0,975	0,992
Chikvadrat	20220	19547	17637	54413
Prob>chi	0	0	0,59	0

**Appendikstabell 3: FE sysselsettingsandeler**

	<i>SL1</i>	<i>SL2</i>	<i>SL3</i>	<i>SL4</i>
Relsmal	-0,0354 (0,0425)	0,2360 (0,0898)***	-0,0857 (0,0694)	-0,1149 (0,0343)***
Reltot73	3,3132 (0,5515)***	-3,1800 (1,1745)***	-1,2305 (0,9074)	1,0974 (0,4477)**
LnV	0,0081 (0,0031)***	-0,0055 (0,0067)	-0,0039 (0,0051)	0,0013 (0,0025)
LnK	0,0005 (0,0005)	0,0014 (0,0010)	-0,0015 (0,0008)*	-0,0005 (0,0004)
InW1_lag	0,0172 (0,0160)	-0,0459 (0,0191)**	0,0122 (0,0142)	0,0165 (0,0074)**
InW2_lag	-0,0459 (0,0191)**	0,0219 (0,0401)	0,0256 (0,0282)	-0,0017 (0,0127)
InW3_lag	0,0122 (0,0142)	0,0256 (0,0282)	-0,0268 (0,0266)	-0,0111 (0,0100)
InW4_lag	0,0165 (0,0074)**	-0,0017 (0,0127)	-0,0111 (0,0100)	-0,0037 (0,0071)
Konstantledd	-0,0144 (0,0384)	0,5974 (0,0811)***	0,2763 (0,0624)	0,1407 (0,0309)***
Tidsdummyer	Ja	Ja	Ja	Ja
Næringsdummyer	Ja	Ja	Ja	Ja
Observasjoner	455	455	455	455
Parametre	65	65	65	65
	0,0112	0,0238	0,0184	0,0091
RMSE				
R <sup>2</sup>	0,979	0,977	0,980	0,993
Chi-kvadrat	271,1	249,9	286,9	807,6
P	0	0	0	0

**Appendikstabell 4: FE industri; 21 næringer**

	S1	S2	S3	S4
Relsmal	0,0339 (0,0463)	0,2024 (0,0905)**	-0,1465 (0,0412)***	-0,0897 (0,0541)*
Reltot73	3,4707 (1,4697)**	-10,1941 (2,9053)***	3,4197 (1,3163)***	3,3037 (1,7446)*
InV	0,0049 (0,0045)	-0,0108 (0,0088)	-0,0040 (0,0040)	0,0099 (0,0052)*
InK	0,0009 (0,0005)	0,0018 (0,0011)*	-0,0016 (0,0005)***	-0,0010 (0,0006)
InW1_lag	0,0226 (0,0245)	0,0037 (0,0249)	-0,0495 (0,0168)***	0,0232 (0,0109)**
InW2_lag	0,0037 (0,0249)	-0,0449 (0,0436)	0,0699 (0,0215)***	-0,0287 (0,0198)
InW3_lag	-0,0495 (0,0168)***	0,0699 (0,0215)***	-0,0111 (0,0207)	-0,0093 (0,0100)
InW4_lag	0,0232 (0,0109)**	-0,0287 (0,0198)	-0,0093 (0,0100)	0,0148 (0,0135)
Konstantledd	0,1555 (0,0283)***	0,7736 (0,0520)***	0,0970 (0,0247)***	-0,0261 (0,0311)
Tidsdummyer	Ja	Ja	Ja	Ja
Næringsdummyer	Ja	Ja	Ja	Ja
Observasjoner	203	203	203	455
Parametre	37	37	37	15
RMSE	0,012	0,023	0,010	0,104
R <sup>2</sup>	0,973	0,962	0,987	0,250
Chi-kvadrat	161,8	113,9	344,0	12,9
P	0	0	0	0

**Appendikstabell 5: FE med gode FoU-tall; 31 næringer**

	S1	S2	S3	S4
Relsmal	-0,0401 (0,0477)	0,1847 (0,1491)	-0,0663 (0,1100)	-0,0782 (0,0556)
RelFoU	0,2898 (0,1440)**	-0,1960 (0,4506)	-0,1802 (0,3333)	0,0863 (0,1680)
InV	0,0012 (0,0053)	0,0108 (0,0163)	-0,0189 (0,0120)	0,0070 (0,0061)
InK	0,0331 (0,0158)**	-0,1596 (0,0494)***	0,0900 (0,0365)**	0,0365 (0,0184)*
InW1_lag	0,0906 (0,0215)***	-0,1145 (0,0248)***	-0,0104 (0,0169)	0,0342 (0,0122)***
InW2_lag	-0,1145 (0,0248)***	0,2414 (0,0655)***	-0,0831 (0,0448)*	-0,0439 (0,0231)*
InW3_lag	-0,0104 (0,0169)	-0,0831 (0,0448)*	0,0804 (0,0378)**	0,0131 (0,0176)
InW4_lag	0,0342 (0,0122)***	-0,0439* (0,0231)	0,0131 (0,0176)	-0,0034 (0,0166)
Konstantledd	-0,3504 (0,1836)*	2,5943 (0,5744)***	-0,8609 (0,4240)**	-0,3829 (0,2146)*
Tidsdummyer	Ja	Ja	Ja	Ja
Næringsdummyer	Ja	Ja	Ja	Ja
Observasjoner	202	202	202	202
Parametre	43	43	43	43
	0,0107	0,0336	0,0248	0,0125
RMSE				
R <sup>2</sup>	0,980	0,962	0,959	0,993
Chi-kvadrat	178,2	92,1	84,9	496,7
P	0,00	0,00	0,00	0,00

**Appendikstabell 6: FE bredimport**

	S1	S2	S3	S4
Relbred	-0,0649 (0,0272)**	0,3666 (0,0650)***	-0,1572 (0,0524)***	-0,1445 (0,0287)***
RelFoU	2,9478 (0,5398)***	-2,2885 (1,2916)*	-1,7407 (1,0433)*	1,0813 (0,5688)*
InV	0,0062 (0,0031)***	-0,0042 (0,0073)	-0,0072 (0,0059)	0,0051 (0,0032)
InK	0,0010 (0,0005)**	0,0006 (0,0012)	-0,0015 (0,0009)	0,0000 (0,0005)
InW1_lag	0,0426 (0,0153)***	-0,0622 (0,0186)***	-0,0006 (0,0144)	0,0201 (0,0078)***
InW2_lag	-0,0622 (0,0186)***	0,1212 (0,0426)***	-0,0106 (0,0315)	-0,0485 (0,0149)***
InW3_lag	-0,0006 (0,0144)	-0,0106 (0,0315)	0,0057 (0,0308)	0,0054 (0,0122)
InW4_lag	0,0201 (0,0078)***	-0,0485 (0,0149)***	0,0054 (0,0122)	0,0230 (0,0091)***
Konstantledd	0,0037 (0,0373)*	0,5514 (0,0884)***	0,3316 (0,0712)***	0,1133 (0,0389)
Tidsdummyer	Ja	Ja	Ja	Ja
Næringsdummyer	Ja	Ja	Ja	Ja
Observasjoner	455	455	455	455
Parametre	65	65	65	65
RMSE	0,0107	0,0257	0,0208	0,0114
R <sup>2</sup>	0,977	0,978	0,975	0,992
Chi-kvadrat	256,71	269,33	233,05	745,66
P	0,00	0,00	0,00	0,00