Statistics Norway
Statistisk sentralbyrå

Svein Erik Førre

Er store foretak mer forskningsintensive?

En anvendelse av diagnostiske metoder

Svein Erik Førre

Er store foretak mer forskningsintensive?
En anvendelse av diagnostiske metoder

Statistisk sentralbyrå • Statistics Norway
Oslo–Kongsvinger 1997

Standardtegn i tabeller	Symbols in tables	Symbol
Tall kan ikke forekomme	Category not applicable	.
Oppgave mangler	Data not available	..
Oppgave mangler foreløpig	Data not yet available	...
Tall kan ikke offentliggjøres	Not for publication	:
Null	Nil	-
Mindre enn 0,5 av den brukte enheten	Less than 0.5 of unit employed	0
Mindre enn 0,05 av den brukte enheten	Less than 0.05 of unit employed	0,0
Foreløpige tall	Provisional or preliminary figure	*
Brudd i den loddrette serien	Break in the homogeneity of a vertical series	—
Brudd i den vannrette serien	Break in the homogeneity of a horizontal series	I
Rettet siden forrige utgave	Revised since the previous issue	r

ISBN 82-537-4413-7
ISSN 0806-2056

Emnegruppe
10.90 Metoder, modeller, dokumentasjon

Emneord
Diagnostiske metoder
Forskning og utvikling
FoU
Industriforetak
Innovasjon
Schumpeter

Design: Enzo Finger Design
Trykk: Statistisk sentralbyrå

Sammendrag

Svein Erik Førre

Er store foretak mer forskningsintensive?
En anvendelse av diagnostiske metoder

Rapporter 97/11 • Statistisk sentralbyrå 1997

Denne analysen ser på sammenhengen mellom forskningsintensitet og foretaksstørrelse for norske industriforetak. En presisering av Schumpeters hypotese om at store foretak er mer innovative enn små danner utgangspunktet for de empiriske analysene. Ulike modelleringer av sammenhengen mellom FoU og størrelse blir diskutert, og diagnostiske metoder blir nyttet som et veiledende redskap ved valg av modell og drøftingen av estimatene som følger av modellvalget. Resultatene viser at det hersker en nær positiv sammenheng mellom FoU og størrelse, men at FoU-intensiteten generelt avtar.

Emneord: Diagnostiske metoder, forskning og utvikling, FoU, industriforetak, innovasjon, Schumpeter

Prosjektstøtte: Dette prosjektet er støttet med midler fra Norges Forskningsråd (Næringslos, 102795/531)

Innhold

1. Innledning 7

2. En formell presisering av den Schumpeterianske hypotese 9

3. Beskrivelse av datamaterialet 11

4. Empiriske sammenhenger mellom FoU og størrelse 12
4.1. En ikke-parametrisk tilnærming 12
4.2. En analyse basert på lineær regresjon 14
4.2.1. En diagnose av robustheten i estimatene på de næringsavhengige elastisitetene β_{1k} 14
4.2.2. En anvendelse av «Added variable plot» 16
4.3. Sensurering - en ikke lineær analyse 19
4.3.1. Utledning av modellen 19
4.3.2. Estimeringsresultater 21
4.4. Størrelsesavhengig 23
4.4.1. Utledning av modellen 23
4.4.2. Estimeringsresultater 23

5. Konklusjon 27

Referanser 29

Vedlegg A: Diagnostiske metoder 30

De sist utgitte publikasjonene i serien Rapporter 33

1. Innledning*

hans liv var viet mikroøkonometri
økonomenes håpefulle alkymi
i en verden som hastet kaotisk forbi
var han selv fylt av underlig harmoni

hans tanker kretset om det generelle
høyt hevet over alt det trivielle
matematikken ble hans trylleformular
kun den var egnet til fullverdige svar

i flere artikler som han tålmodig produserte
forsøkte han å forklare det ytterst kompliserte
overbevist om det underliggende strukturelle
rettferdiggjorde han det snevre og spesielle

hans håp om å bli en av de store
gav han styrke i det han gjorde
så konsentert var han om det sære
at han ble uberørt av det nære

i formlenes abstraherte estetikk
slukket han sitt behov for erotikk
sikker på at hans eneste misjon
var den vitenskaplige evolusjon

I økonomisk-politiske diskusjoner synes det å herske enighet om at graden av teknologisk utvikling er betinget av et foretaks størrelse. Særlig blir dette momentet fremhevet når man diskuterer norske foretaks konkurranseulemper i møte med store utenlandske konsern. Diskusjonen om sammenhengen mellom størrelse og teknologisk utvikling og innovasjon har sine røtter tilbake til Schumpeter (1942) som argumenterte for et skarpt skille mellom den optimale statiske organisering og den optimale dynamiske organisering av bedrifter og markeder. Tanken var at den optimale statiske ressursallokering vil bli best ivaretatt av små bedriftsenheter i et frikonkurransemarked, mens forutsetningen for optimal teknologisk utvikling er store bedriftsenheter og høy markedskonsentrasjon. Schumpeter mente at en monopolist ville generere en større produksjon av innovasjoner siden «there are advantages which, though not strictly attainable on the competetive level of enterprice, are as a matter of fact secured only on the monopoly level»[1]

Selv om Schumpeter vektla kombinasjonen av størrelse og markedsmakt, har empiriske studier hovedsakelig konsentrert seg om størrelse alene (se Cohen og Levin, 1989). I diskusjonen av de empiriske resultatene har to hovedargumenter gjort seg gjeldende. Det første argumentet postulerer at det er *økende utbytte på skalaen ved utgifter til forskning og utvikling (FoU)*. Argumentet består av to deler: (a) En stor FoU-stab er mer effektiv enn en liten og (b) en FoU-stab av en gitt størrelse er mer effektiv i et stort foretak enn i et lite. Argumentasjonen for (a) har vært at et stort foretak gir mer rom for spesialisert personell, både rene forskere og rådgivere samtidig som sannsynligheten for at enkelte blir i foretaket over lengre tid og dermed opparbeider seg lang erfaring er større. Argumentasjonen for (b)

* Denne rapporten er en revidert hovedfagsoppgave i sosialøkonomi ved UiO under veiledning av Forsker Tor Jakob Klette i SSB.

[1] Schumpeter (1942), s. 101.

knytter seg til risikoen som er forbundet med forskning: Et stort foretak har gjerne mer diversifiserte aktiviteter enn et lite foretak og kan derfor lettere dra nytte av et FoU prosjekt der utfallet er usikkert[2].

Det andre hovedargumentet som er blitt reist er at det er *økende utbytte på skalaen i finansmarkedet*. Store bedrifter har tilgang til flere finansmarkeder enn små, samtidig som de lettere kan oppnå gunstige lån[3]. Dette resulterer i at store foretak kan finansiere en *relativt* større FoU stab enn et lite foretak. Store foretak vil også være mindre risikoaverse siden gode lånemuligheter kan kompensere for risikoen knyttet til et usikkert FoU prosjekt.

I de senere år har flere argumentert mot postulatet om at store foretak entydig skal ha et konkuransemessige fortrinn ved FoU (se f.eks Scherer og Ross 1990, s. 652-3). Det har særlig blitt pekt på at en byråkratisk organisasjonsstruktur i større foretak vil kunne trenere nye prosjekter og vanskeliggjøre risikofylt innovasjon og dermed forhindre nye FoU prosjekter. Samtidig som byråkratiske restriksjoner vil kunne bidra til at vitenskapelig personell forlater store foretak til fordel for små. Det har også blitt påpekt at selv om store foretak har større tilgang på egenkapital vil de være varsomme med å låne penger til risikofylt investering i kapitalmarkedet dersom de er avhengige av lån for ekspansjon på andre fronter.

Siden det er problematisk å måle avkastningen på FoU har empiriske tester konsentrert seg om FoU *innsatsen* og testingen av det som Fisher og Temin (1973) kaller «den Schumpeterianske hypotese». *«Den Schumpeterianske hypotese» postulerer at stordriftsfordeler ved FoU vil kunne testes ved en null hypotese om at FoU-innsatsen vokser proporsjonalt med foretaksstørrelsen.* Denne hypotesen er derfor vesensforskjellig fra det som Fisher og Temin (1973) kaller «Schumpeters egentlige hypotese»[4] som postulerer at *FoU-produksjonen*, eksempelvis antall patenter eller innovasjoner skal vokse mer enn proporsjonalt med foretaksstørrelsen. Det eksisterer en meget omfattende empirisk litteratur som har tatt «den Schumpeterianske hypotesen» som sitt utgangspunkt og studert sammenhengen mellom FoU-intensitet og foretakstørrelse innenfor vanlige lineære modeller. Denne litteraturen er diskutert i flere oversiktsartikler se f.eks Cohen og Levin (1989), Baldwin og Scott (1987). Cohen og Klepper (1996) oppsummerer resultatene og konkluderer med at det eksisterer overveldende bevis på tvers av utvalg, spesifikasjoner og estimeringsmetoder for at FoU innsatsen vokser monotont med foretaksstørrelsen og proporsjonalt for foretak over en viss størrelse. Dette stiliserte faktum har i neste omgang, i lys av «den Scumpeterianske hypotese», ført til en konklusjon om at ikke eksisterer stordriftsfordeler ved FoU.

Fisher og Temin (1973) var de første som prøvde å formalisere Scumpeters opprinnelige argument og sammenhengen mellom dette argumentet og «den Schumpeterianske hypotese». De konkluderte med at «den Schumpeterianske hypotesen» var uegnet som indikator på eventuelle stordriftsfordeler ved FoU og advarte sterkt mot å trekke konklusjoner om eventuelle stordriftsfordeler på basis av FoU-intensiteten. Nylig har Cohen og Klepper (1996) vist at en konstant (avtakende) FoU-intensitet kan være fullt ut forenlig med stordriftsfordeler ved FoU. I del 2 vil vi komme tilbake til disse bidragene.

Selv om empiriske studier av sammenhengen mellom FoU og størrelse ikke er egnet til å belyse eventuelle stordriftsfordeler ved FoU, betyr ikke dette at denne sammenhengen er uinteressant.

Dokumentasjon og verifisering av gyldigheten i empiriske «lovmessigheter» er viktige for utviklingen og vurderingen av økonomiske modeller. Stiliserete fakta kan inngå som grunnleggende hypoteser i en teori eller tjene som indikatorer på validiteten av teorien (Dasgupta 1986). Dersom en kan dokumentere en proporsjonal sammenheng mellom FoU og størrelse vil dette kunne tjene som en svært forenklende antakelse i en større modell (se Klette og Grilliches 1996). Tidligere er sammenhengen mellom FoU og størrelse ikke undersøkt for norske industriforetak. En empirisk dokumentasjon av denne sammenhengen er viktig for å kunne vurdere hvorvidt internasjonale (amerikanske) «stiliserte fakta» også er gyldige for den norske økonomiske virkelighet. Hovedmotivasjon for denne studien er derfor først og fremst deskriptiv: Hvilken sammenhengen hersker mellom FoU og størrelse i norske industriforetak ? Analysen har en klar eksplorativ karakter og eksplorative metoder eller det som går under navnet *diagnostiske metoder* inngår derfor som en vesentlig del av analysen.

I del 2 vil de ulike forsøkene på å formalisere «den Schumpeterianske hypotese» bli diskutert. I del 3 vil datamaterialet bli introdusert. I del 4 følger analysen av dataene.

---

[2] Her er det implisitt forutsatt at det ikke eksisterer et marked for ny kunnskap. En teoretisk begrunnelse for dette er gitt i K. Arrow (1962).

[3] Se F. Johansen (1995) for en analyse av sammenhengen mellom investering og finansielle forhold i norske industribedrifter.

[4] Cohen & Klepper (1996) påpeker at Schumpeter strengt tatt ikke har fremsatt en hypotese som postulerer en kontinuerlig sammenheng mellom størrelse og FoU, men at han kun viste til at store byråkratiske foretak stod for hovedvekten av de innovasjoner som ble foretatt i midten av vårt århundre.

2. En formell presisering av den Schumpeterianske hypotese

Det stiliserte faktum at FoU-intensiteten ikke vokser med foretaksstørrelsen og mer direkte tester basert på innovasjonstall og patenter som har vist at det relative antallet innovasjoner og patenter avtar med foretaksstørrelsen (Bound et.al.(1984); Acs og Audretsch (1988)), har resultert i en utbredt enighet om at store foretak ikke har noen konkurransemessige fortrinn ved FoU. Fisher og Temin (1973, 1979); heretter F&T; retter imidlertid en klar advarsel mot å trekke slutninger om stordriftsfordeler på basis av sammenhengen mellom FoU og størrelse. Deres artikkel fra 1973 er det første forsøk på å gi en presis formulering av problemstillingen. Schumpeters opprinnelige argument ble av F&T antatt å være basert på et postulat om økende utbytte på skalaen både mhp. størrelsen på FoU-innsatsen og den totale foretaksstørrelsen. De definerte en produktfunksjon F(R,N) som uttrykte «the dollar value of the *per worker* output of the R&D staff, or the average labour productivity of research and development» og var en funksjon av R, antall sysselsatte i FoU virksomhet, og N, antall sysselsatte i foretaket som var knyttet produksjon. Foretaket ble så antatt å maksimere profitten gitt ved:

(1) $N) - wR + \Pi = R\,F(R,H(N)$

wR er lønnkostnader ved FoU mens H (N) uttrykker *netto*-inntekter uavhengig av om foretaket utfører FoU. RF(R,N) blir av F&T forklart som «the difference in the profits (exclusive direct R&D costs) obtained by the firm of operating staff size N which engages in R&D to the extent measured by R and those which would be obtained by the same firm if it did no R&D.» Det totale antall sysselsatte i foretaket er da gitt ved summen av R og N.

Schumpeters postulat er da, hevder F&T, at den gjennomsnittlige FoU produktivitet skal vokse med størrelsen på FoU-staben, R slik at :

(2) $F_1 > 0$

og at den skal vokse med størrelsen på foretaket, N slik at:

(3) $F_2 > 0$

«Den Schumpeterianske hypotesen» kan da videre uttrykkes som:

(4) $$El_S RF \equiv \frac{S}{RF}\frac{d(RF)}{dS} = El_S R + \frac{S}{F}\frac{dF}{dS} > 1$$

mens Schumpeters «egentlige» argument ville kunne uttrykkes som

(5) $$El_S RF \equiv \frac{S}{RF}\frac{d(RF)}{dS} = El_S R + \frac{S}{F}\frac{dF}{dS} > 1$$

F&T hevder at Schumpeters eget argument synes å være at (2) og (3) impliserer (5) mens den empiriske litteraturen synes å hevde at (2) og (3) impliserer (4) som igjen impliserer (5). De viser at ingen av disse sammenhengene generelt eksisterer og konkluderer med at estimater av El_SFoU hverken vil bekrefte eller avkrefte den sentrale hypotesen gitt ved (2) og (3).

En logisk glipp i argumentasjonen til F&T, poengtert av Rodriguez (1979), resulterte i at disse konklusjonene ble trukket i tvil. Rodriguez viser at en optimal tilpasning av foretaket forutsetter at $F_1 < 0$. Dette bryter imidlertid med den grunnleggende antakelsen til F&T i (2). I et svar til Rodriguez foreslår F&T (1979) at man erstatter (2) med:

(2') $RF_1 + NF_2 > 0$

her er F_1 antatt å være negativ i optimum. Den nye formuleringen antar at tilpasningen er slik at dersom man øker R og N prosentvis like mye vil dette gi et samlet økende utbytte på skalaen ved FoU. Denne nye formuleringen innebærer imidlertid at F&T ikke tar hensyn til Schumpeters første hypotese. De opprinnelige konklusjonene fremstår imidlertid som styrkede etter reformuleringen.

Kohn og Scott (1982) videreutvikler ideene til F&T innenfor det samme rammeverket. De introduserer en egen produktfunksjon for FoU-produksjon, Q=G(R)

der R er utgifter til FoU. Ved å la Q inngå som handlingsparameter i optimeringsbetingelsen i stedet for R (jf. (1)) er deres alternative formulering både konsistent med en profittmaksimerende tilpasning for foretaket og postulat (2). K&S konstaterer, som F&T, at en test av «den Scumpeterianske hypotesen» ikke vil kunne kaste lys over de grunnleggende postulatene om stordriftsfordeler gitt ved (2) og (3).

Cohen og Klepper (1996); heretter C&K; tar et annet utgangspunkt enn F&T idet de *ikke* antar at stordriftsfordelene ved FoU oppstår som et resultat av økt produktivitet, men at gevinsten et foretak kan hente inn ved innovasjon avhenger av det produksjonsapparatet foretaket har til rådighet ved innovasjonstidspunktet. Modellen de presenter hviler på to grunnleggende forutsetninger: (i) Det eksisterer ikke noe marked for innovasjoner, (ii) gevinsten ved en innovasjon blir ikke hentet ut ved å ekspandere produksjonen men ved en økt pris-kostnadsmargin. Store foretak har derfor en stordriftsfordel idet de ved en innovasjon har større produksjonskapasitet. Med dette utgangspunktet har de utviklet en teori som er i overensstemmelse med det de regner som stiliserte fakta i den empiriske litteraturen om FoU og størrelse: (a) sannsynligheten for å utføre FoU øker med foretaksstørrelsen, (b) det eksisterer en nær og positiv sammenheng mellom FoU og størrelse, (c) FoU vokser proporsjonalt med størrelse innenfor de fleste næringer og (d) antall patenter og innovasjoner generert per FoU-krone avtar med størrelsen. *C&K viser at disse stiliserte fakta er fullt ut forenlig med en modell der det er stordriftfordeler ved FoU.* Innenfor den «Schumpeterianske» tradisjon er (c) og (d) blitt tolket som stordriftsulemper ved FoU. C&K viser ved en enkel modell at (c) og (d) er den naturlige følge av at store foretak med et stort produksjonsapparat har et konkurransemessig fortrinn. En avtakende FoU-intensitet kan også vises å være forenlig med modellen. Resonnementet er som følger: Pris-kostnadsmarginen antas å være positivt voksende med graden FoU. Prosessinnovasjoner reduserer produksjonskostnaden, mens produktinnovasjoner øker kvaliteten på produktet som selges og gir dermed grunnlag for en høyere pris. Produktiviteten til et FoU-prosjekt er uavhengig av foretaksstørrelsen. Forskjeller i utgifter til FoU er et resultat av at gevinsten ved en innovasjon avhenger av den initiale produksjonskapasiteten. Siden lønnsomheten i et FoU prosjekt er direkte proporsjonal med størrelsen på foretakets produksjonskapasitet, vil utgiftene til FoU også være proporsjonale med produksjonskapasiteten. Videre følger det at store foretak utfører mer marginale (dvs. mindre produktive) FoU prosjekter enn små foretak fordi de i større skala vil kunne utnytte gevinsten av den økte pris-kostnadsmarginen ved en innovasjon. Disse marginale FoU-prosjektene resulterer i at det gjennomsnittlige FoU prosjektets produktivitet er lavere for store foretak. I motsetning til den allmenne oppfatning er denne lavere produktiviteten nettopp et uttrykk for stordriftsfordeler ved FoU. Desto større initial produksjonskapasitet, desto større profitt ved en innovasjon.

I lys av dette bør en avstå fra å trekke slutninger om stordriftsfordeler med utgangspunkt i om FoU-intensiteten vokser med størrelse. Mens F&T retter et kritisk blikk på spriket mellom de argumentene som var blitt reist for eventuelle stordriftfordeler ved FoU og de empiriske slutningene som kan trekkes ved å studere sammenhengen mellom FoU og størrelse, bidrar C&K med en teoretisk modell som viser at en konstant/avtakende FoU intensitet er fullt ut forenlig med en hypotese om stordriftsfordeler ved FoU.

3. Beskrivelse av datamaterialet

Utgangspunktet for denne analysen er data for årene 1982, 1983, 1984, 1985, 1987 og 1989. To ulike datasett, SSBs Industristatistikk og NTNFs data for FoU, er blitt koblet sammen. I koblingen av disse datasettene gjorde to ulike problemer seg gjeldende. Et problem var at NTNF benyttet en annen koding av foretakene enn SSB gjør i Industristatistikken. Hvert NTNFs-foretak fikk derfor tilordnet et foretaksnummer samsvarende med Industristatistikken. Som tabell 3.1 viser fikk ikke alle bransjeenhetene tilordnet et slikt foretaksnummer. Disse ble utelatt. Utgifter til FoU er oppgitt for bransjeenheter på 4-siffernivå (Standard for næringsgruppering)[5]. En bransjeenhet er alle bedriftene i et flerbedriftsforetak som produserer innenfor samme næring. I analysen ble bransjeenheter på 3-siffernivå studert. Hver bransjeenhet i NTNFs datasett ble koblet til en tilsvarende bransjeenhet i Industristatistikken. En første kobling viste at opp mot en mot en tredjedel av bransjeenhetene i NTNF-databasen ikke fant noen makker i Industristatistikken. To hovedårsaker gjør seg her gjeldende: Omlag halvparten av disse bransjeenhetene fantes ikke i Industristatistikken for den gjeldende årgangen. Den andre halvparten fantes, men oppga utgifter til FoU i en annen 3-siffer næring enn den produserte i. Siden foretakene i NTNFs kartlegging ble bedt om å oppgi produktrelevant FoU og mange av enbedriftsforetakene oppga FoU i en annen næring en de i henhold til industristatistikken produserte i, er det rimelig å anta at "feil" produktgruppe for FoU er oppgitt. Produktgruppen til disse foretakene ble endret slik at den samsvarte med foretakets produktgruppe i Industristatistikken. En slik systematisk endring var uproblematisk for enbedriftsforetak og for de av flerbedriftsforetakene der alle bedriftsenhetene produserte innenfor samme 3-siffer næring. For de flerbedriftsforetakene med produksjon i ulike 3-siffer næringer var situasjonen mer problematisk. FoU-virksomheten kunne her være rettet mot bransjeenheter i ulike næringer. Da det ikke dreide seg om mer enn 1-3 foretak pr. år, ble disse i mangel av noe bedre alternativ utelatt fra de videre analyser. Tabell 3.1 viser tall fra koblingen av datasettene. Det nye datasettet inneholder data for finansiering av og kostnader ved FoU i tillegg til de variable som allerede er inneholdt i Industristatistikken.

Tabell 3.1. Koblingen av NTNFs FoU-data og SSBs Industristatistikk

NTNF-enheter	304	337	400	445	484	443
Enheter med tilordnet foretaksnunner	273	312	377	432	469	440
Bransjeenheter på 3-siffernivå	243	286	342	400	449	418
Kobling til Industristat. før omkodning	187	223	279	325	278	295
Omkoding av produktgruppe	38	41	39	57	63	68
Kobling til Industristat. etter omkoding[1]	225	264	316	381	341	358

[1] Antall observasjoner for hver årgang etter at bransjeenhetene har fått produktgruppen for FoU endret på 4-siffer nivå og deretter aggregert til 3-siffernivå og koblet til Industristatistikken.

[5] Bedriftene i Industristatistikken er klassifisert etter de varene og tjenestene de produserer. Denne klassifiseringen er basert på FNs internasjonale grupperingsstandard ISIC- International Standard Industrial Classification of all Economic Activities. Standarden er utformet som et 5-sifret pyramidisk grupperingssystem der det 5-sifferede nivået vil være det mest detaljerte nivået.

4. Empiriske sammenhenger mellom FoU og størrelse

4.1. En ikke-parametrisk tilnærming

Tabell 4.1 gir en oversikt over hvordan bransjeenhetene fordeler seg mellom næringer, antallet bransjeenheter som utfører FoU innen hver næring og hvor mange bransjeenheter som er del av et større foretak. I tillegg er gjennomsnittsmål for bruttoproduksjon, FoU og FoU-intensitet angitt. Som det fremgår av tabellen oppgir kun en liten andel av bransjeenhetene i industrien at de utfører FoU.

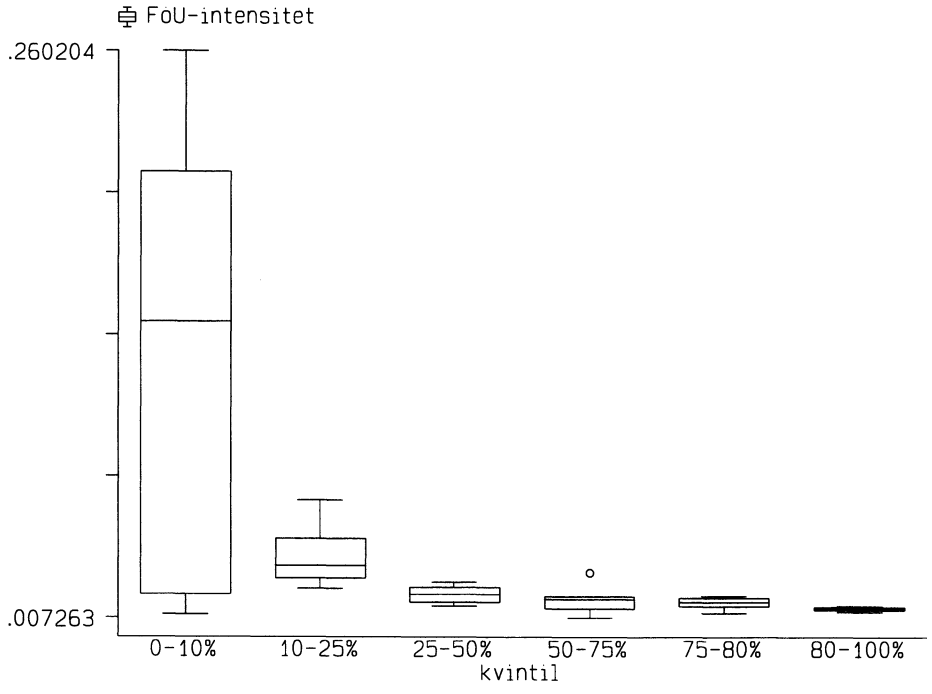
Det er gode grunner til å anta at ikke all FoU som utføres i industrien er beskrevet ved dette utvalget. Det at en bransjeenhet *ikke rapporterer* FoU innebærer ikke nødvendigvis at den *ikke utfører* FoU. Det kan være rimelig å anta at mange *små* bransjeenheter *ikke* opererer med noe klart skille mellom FoU og annen produksjon. Disse bransjeenhetene oppgir ikke utgifter til FoU siden de ikke eksplisitt budsjetterer for FoU. Videre vil de av de minste bedriftene som *oppgir* FoU, oppgi FoU fordi en vesentlig del av virksomheten er knyttet til FoU. Dette er et klassisk seleksjonsproblem. En økonometrisk analyse som ikke tar hensyn til dette problemet vil resultere i forventningsskjeve estimater siden kun de mest forskningsintensive av de små bransjeenhetene er representert blant de bransjeenhetene som oppgir FoU.

I figur 4.1.1 og 4.1.2 er bransjeenhetene gruppert i ulike kvintiler etter størrelsen på bruttoproduksjonsverdien. Figurene viser spredningen i FoU-intensiteten mellom de ulike årene i periden 1982-89 innenfor hver kvintil i størrelsesfordelingen ved et *boxplot*. Et boxplot viser medianen og interkvartil avstanden dvs. avstanden mellom den 1. og 3. kvartil i fordelingen (avstanden som spenner over de midtre 50 prosent av observasjonsårene). Det fremgår av figurene at sammenhengen mellom FoU-intensiteten og størrelsen på bransjeenhetene vil avhenge om vi baserer oss på alle bransjeenhetene i industrien eller kun de som oppgir FoU. Av figur 4.1.1 ser vi at ved å beregne FoU-intensiteten med utgangspunkt i alle bransjeenhetene (og sette utgifter til FoU lik null dersom bransjeenheten ikke oppgir FoU), vil medianen og interkvartilavstanden til FoU-intensiteten ikke ligge systematisk lavere utover i størrelsesfordelingen. Av figur 4.1.2 ser vi imidlertid at dersom vi i stedet kun tar utgangspunkt i de bransje

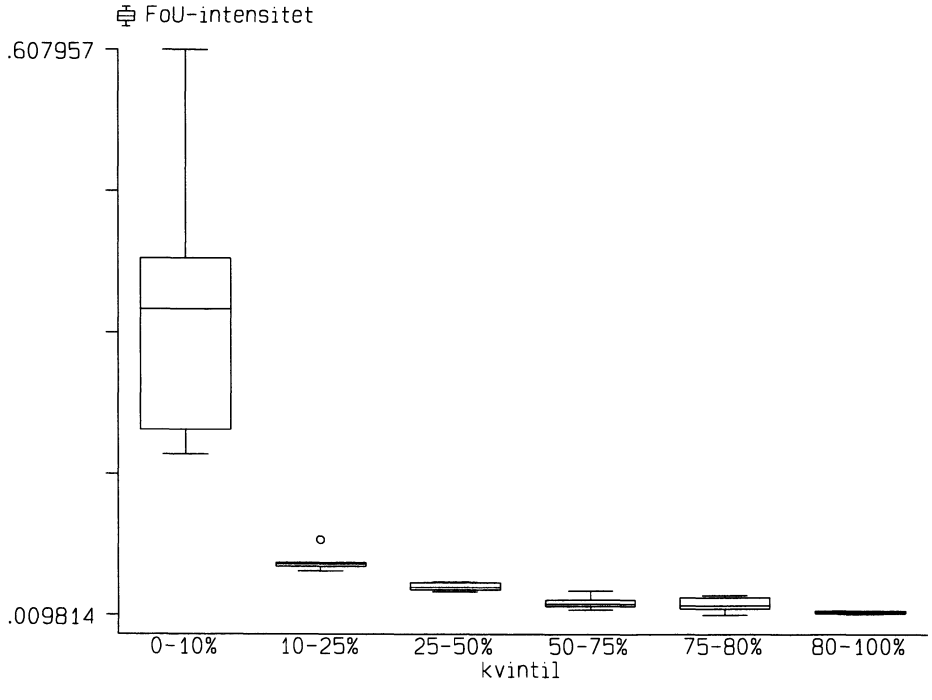
Tabell 4.1. Oversiktstabell for noen sentrale størrelser i analysen

Næring	Næringskode	Bransjeenheter	Gjennomsnittlig bruttoproduksjon	Antall bransjeenheter som er del av et større foretak	Antall bransjeenheter som utfører FoU	Gjennomsnittlig FoU	Gjennomsnittlig FoU-intensitet
Næringsmidler	1	2332	36829,5	7	33	2912,2	0,017
Tekstilvarer og lærvarer	2	718	10476,7	4	15	1311,3	0,076
Trevarer	3	1730	14944,8	4	30	958,4	0,010
Treforedling	4	1437	30265,8	9	24	2339,9	0,007
Kjemiske produkter	5	185	165964,6	13	33	15512,3	0,062
Raff. og prod. av jordolje	6	41	535733,4	2	4	3395,4	0,018
Gummi og plastvarer	7	390	20232,0	4	16	1198,0	0,031
Mineralske produkter	8	558	20717,5	3	16	2321,2	0,018
Metaller	9	125	311055,4	12	30	16064,2	0,056
Metallvarer	10	1433	12497,4	15	40	2025,2	0,048
Maskiner	11	967	51815,8	10	68	12834,0	0,077
Elektroprodukter	12	425	46579,6	9	63	14121,0	0,208
Transportmidler	13	884	32803,8	5	35	5710,5	0,333
Teknisk vitenskl. prod.	14	69	20021,29	0	8	5876,2	0,294
Industriprod. ellers	15	162	9681,7	1	4	1366,4	0,028

Figur 4.1.1. FoU-intensiteten i ulike kvintiler etter bransjeenhetsstørrelse. Alle bransjeenheter i industrien



Figur 4.1.2. FoU-intensiteten i ulike kvintiler etter bransjeenhetsstørrelse. Kun FoU-rapporterende bransjeenheter



enhetene som utfører FoU, vil medianen og interkvartil avstanden til FoU-intensiteten ligge systematisk lavere utover i størrelsesfordelingen.

Et sentralt poeng i den videre analysen er å forsøke å klarlegge hvorvidt den generelle trenden er en avtakende eller konstant FoU-intensitet. I denne sammenheng vil spørsmålet om hvordan en skal forholde seg til de bransjeenhetene som ikke oppgir FoU være viktig. Dersom det skjer en seleksjon der kun de mest FoU intensive av de minste bransjeenhetene oppgir at de utfører FoU, vil en analyse som ikke tar hensyn til dette problemet underpredikere elastisiteten som beskriver sammenhengen mellom størrelse og FoU. Men det kan også tenkes at det skjer en annen form for seleksjon. Dersom det eksisterer høye faste kostnader ved FoU, og man antar som Cohen og Keller (1996) at lønnsomheten ved FoU avhenger av den initiale produksjonskapasiteten, vil FoU kun være lønnsom i de mest produktive av de små bransjeenhetene. Den høye FoU-intensiteten i de minste bransjeenhetene vil da være et naturlig resultat av dette. Det mest nærliggende er imidlertid å anta at begge disse seleksjonsmekanismene gjør seg gjeldene. I kapittel 4.3 og 4.4 vil seleksjonsproblemene bli drøftet nærmere.

4.2. En analyse basert på lineær regresjon

Sammenhengen mellom FoU og foretaksstørrelse danner utgangspunktet for denne analysen. Datamaterialet er aggeregert til 3-siffernivå slik at vi kan separere bransjeenheten fra resten av foretaket og dermed analysere både sammenhengen mellom FoU og størrelsen på bransjeenheten men også betydningen av resten av foretaket. Det empiriske utgangspunktet er den enkle log-lineære modellen analysert i Bound et.al.(1984) :

$$(4.2.1)\quad FoU = \alpha^{*}S^{\beta}$$

der FoU er de totale utgiftene til FoU, S er foretakets bruttoproduksjonsverdi, α^{*} er en konstant og β er elastisiteten av FoU mhp. foretaksstørrelsen. Vi utvider modellen (4.1.1) ved å definere S som bransjeenhetens bruttoproduksjonsverdi og Se som restforetakets bruttoproduksjonsverdi:

$$(4.2.2)\quad FoU = \alpha^{*}\, S^{\beta_1}Se^{\beta_2}$$

De videre analysene tar utgangspunkt i den log-lineære sammenhengen:

$$(4.2.3)\quad \ln FoU = \alpha + \beta_1 \ln S + \beta_2 \ln Se + \eta D(Se=0) + \varepsilon$$

der lnFOU, lnS og lnSe er henholdsvis de naturlige logaritmene til FoU, S og Se, α og β er ukjente parametre og ε antas å være normalfordelt med forventning null og varians σ^2. Vi har satt Se vilkårlig lik én i tilfellet Se=0, og for å korrigere for dette inkluderer vi en dummy D(Se=0). Et opplagt problem ved å nytte FoU i det enkelte år som et mål på innovasjon er forskjellen i tid mellom de løpende utgiftene til FoU og innovasjonentidspunktet. For å korrigere for denne tidsforskyvningen og eventuelle konjunkturelle variasjoner mellom næringer, baserer vi oss på gjennomsnittsverdier for perioden 1982-89 (bedrifter som bare inngår i en enkeltårgang er utelatt). Utenlandsk kapital vil kunne påvirke den FoU-aktivitet som foregår innen et foretak. Ved å inkludere en dummy i modellen blir det korrigert for utenlandsk eierskap. Det er videre naturlig å anta store næringsmessige forskjeller i nivået på FoU. For å ta hensyn til dette kategoriserer vi observasjonene i 15 produktgrupper og tillater variasjoner i parametre og i nivået på FoU mellom næringene.

Etter at vi har utvidet modellen slik at vi tar hensyn til næringsmessige forskjeller og effekten av utenlandsk eierskap har vi følgende log-lineære sammenheng:

$$(4.2.4)\quad \ln FOU_i = \beta_{1k}\ln S_i + \beta_2 \ln Se_i + \eta D(Se_i=0) + \gamma D(N_{ik}) + \theta D(UK_i=0) + \varepsilon_i,$$

$$\text{der } i=1,\ldots,425 \text{ og } k=1,\ldots,15$$

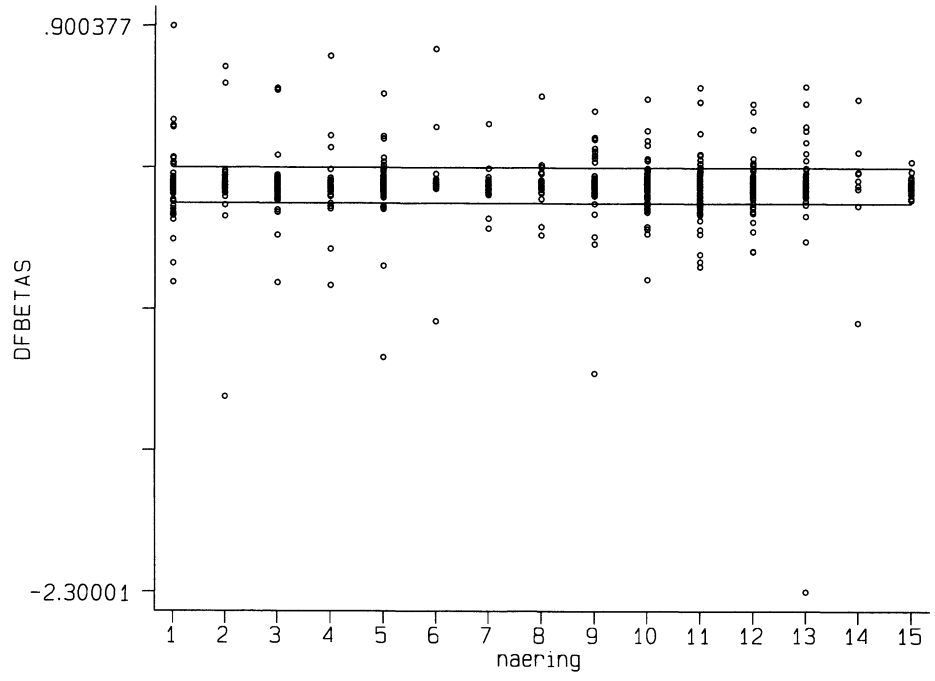
Vi har inkludert en dummy variabel, $D(N_{ik})$, som tar hensyn til nivåforskjeller mellom næringene. Videre har vi innført en dummy, $D(UK_i=0)$, for en utenlandsk eierandel større enn 20 prosent. I relasjon (4.2.4) tillater vi elastisiten, β_{1k}, å variere mellom de 15 næringene. Effekten av restforetakets størrelse, β_2, antas imidlertid lik for alle næringer. I det neste avsnittet vil estimatene for de næringsavhengige elastisitetene (β_{1k} , k=1,..15) bli presentert, og i denne sammenheng vil et diagnostisk mål bli introdusert og tjene som et middel til å vurdere robustheten i disse estimatene.

4.2.1. En diagnose av robustheten i estimatene på de næringsavhengige elastisitetene β_{1k}

I en regresjonsanalyse vil det ofte være et utilfredsstillende stort avvik mellom den idealiserte teoretiske basis og den praktiske anvendelsen. Inferens basert på minste kvadraters metode kan være sterkt påvirket av noen få enkeltobservasjoner og modellens prediksjoner vil kunne reflektere disse avvikende observasjonene i større grad enn det generelle mønstret i datasettet. To hovedretninger er blitt fulgt for å redusere dette gapet mellom teori og praksis. Den ene hovedretningen baserer seg på *robuste metoder* i estimering og testing som krever færre strukturelle forutsetninger. Den andre hovedretningen, *diagnostiske metoder*, tar utgangspunkt i den idealiserte postulerte modellen og søker å identifisere de faktorene som avviker fra denne spesifikasjonen.

Diagnostiske metoder nyttes for å identifisere observasjoner som ikke er i overensstemmelse med den modellsammenhengen som er postulert. Identifikasjon av

Figur 4.2.1. DFBETAS$_i$ for de næringsavhengige elastisitene



disse observasjonene vil være viktig for å kunne vurdere styrken i en estimert sammenheng, validiteten i modellen som helhet og mulige modifikasjoner av den gjeldende modellen. En nyttig distinksjon kan gjøres mellom avvikende observasjoner og innflytelsesrike observasjoner. *Avvikende observasjoner* er observasjoner som avviker fra det generelle mønsteret i datamaterialet mens *innflytelsesrike observasjoner* er observasjoner som påvirker de estimerte parametrene på en avgjørende måte. En observasjon kan selvfølgelig inneha begge disse egenskapene. Innflytelsen kan være på prediksjonen av høyresidevariabelen, på de estimerte parameterene under ett eller på en enkelt parameter.

En enkeltobservasjons innflytelse kan måles ved endringen på estimatet(ene) eller prediksjonen etterat den gjeldende observasjonen er utelatt. «R*ow deletion methods*» er en fellesbetegnelse på metoder som har et slikt utgangspunkt. I det følgende vil vi konsentrere oss om et diagnostisk mål som måler innflytelse på *enkeltkoeffisientene* i en lineær regresjonsmodell.

De næringsavhengige elastisiteter i modell 4.2.4 viser seg å resultere i lite robuste estimater på de individuelle elastisitene $\beta_{1,k}$ (k=1,..15). En måte å se dette på er å nytte det diagnostiske målet *DFBETAS$_i$* . Betydningen av enkeltobservasjoner på de estimerte koeffisientene blir ved DFBETAS$_i$ målt ved det skalerte avviket i estimatet $b_{1,k}$ for $\beta_{1,k}$, i forhold til det tilsvarende estimatet etter at den i-te observasjonen er utelatt, $b_{1,k}(i)$ (Belsley et.al.1980)[6]:

$$DFBETAS_i = \frac{b_{1,k}(i) - b_{1,k}}{s(i)\sqrt{(X'X)^{-1}_{ii}}}$$

der $(X'X)_{ii}$ er det i'te diagonale element av (X'X) og X er matrisen av eksogene variable i en regresjonsmodell. Differansen er skalert, ikke med sitt faktiske standardavvik men med standardavviket til $b_{1,k}$ og et estimat på variansen til restleddet (σ) etter at den i-te observasjonen er utelatt, s(i). En tolkning av DFBETAS$_i$ er at den gir et anslag på hvor mange standardavvik estimatet på parameteren endres når den gjeldende observasjonen utelates. Fordelingen til DFBETAS$_i$ er ikke kjent, men en *veiledende* kritisk grense er foreslått lik $2/\sqrt{n}$, der n er antall observasjoner. Dersom en observasjon resulterer i en DFBETAS$_i$ som i absoluttverdi er større enn denne grensen blir den identifisert som *potensielt* innflytelsesrik. Disse veiledende grensene vil i praksis ofte være overflødige, viktigere vil det være å studere spredningen og ekstreme verdier av DFBETAS$_i$.
Kolonne 3 i Tabell 4.2.1 viser estimatene på de individuelle elastisitene $\beta_{1,j}$ gitt ved relasjon 4.2.4. Kolonne 4 viser p-verdiene som følger av en F-test av hypotesen om at elastisitene, $\beta_{1,j}$, er signifikant forskjellige fra én.

[6] Se A3 i vedlegget.

Tabell 4.2.1. Estimater på de næringsavhengige elastisitene $\beta_{1,k}$ i relasjon 4.1.4. Øvre og nedre grenser for estimatene er angitt

Produktkategorier	Parametre	Relasjon 4.2.4	$H_0: \beta_{1,i}=1$ p-verdi	Øvre grense	$H_0: \beta_{1,i}=1$ p-verdi	Nedre grense	$H_0: \beta_{1,i}=1$ p-verdi
Næringsmidler	$\beta_{1,1}$	0,611**	0,01	0,688**	0,03	0,476**	0,00
		(0,142)		(0,141)		(0,142)	
Tekstilvarer og lærvarer	$\beta_{1,2}$	0,173	0,01	0,571	0,26	-0,004	0,00
		(0,316)		(0,380)		(0,342)	
Trevarer	$\beta_{1,3}$	0,690**	0,06	0,785**	0,23	0,567*	0,01
		(0,164)		(0,179)		(0,167)	
Treforedling	$\beta_{1,4}$	0,632**	0,08	0,754**	0,25	0,451**	0,01
		(0,209)		(0,209)		(0,213)	
Kjemiske produkter	$\beta_{1,5}$	0,691**	0,03	0,833**	0,24	0,624*	0,01
		(0,140)		(0,143)		(0,140)	
Raff. og prod. av jordolje	$\beta_{1,6}$	0,03	0,01	0,42	0,17	-2,372	0,04
		(0,376)		(0,424)		(1,629)	
Gummi og plastvarer	$\beta_{1,7}$	0,392	0,01	0,439	0,03	0,303	0,00
		(0,220)		(0,263)		(0,226)	
Mineralske produkter	$\beta_{1,8}$	0,755**	0,38	0,839**	0,56	0,592	0,20
		(0,274)		(0,274)		(0,274)	
Metaller	$\beta_{1,9}$	0,640**	0,01	0,845**	0,31	0,580**	0,00
		(0,133)		(0,151)		(0,138)	
Metallvarer	$\beta_{1,10}$	0,293	0,00	0,380**	0,00	0,177	0,00
		(0,153)		(0,151)		(0,277)	
Maskiner	$\beta_{1,11}$	0,692**	0,00	0,700**	0,00	0,631**	0,00
		(0,090)		(0,090)		(0,090)	
Elektroprodukter	$\beta_{1,12}$	0,662**	0,00	0,687**	0,00	0,584**	0,00
		(0,100)		(0,098)		(0,100)	
Transportmidler	$\beta_{1,13}$	0,255	0,00	0,620**	0,03	0,150	0,00
		(0,153)		(0,173)		(0,157)	
Teknisk vitenskl. prod.	$\beta_{1,14}$	0,319	0,04	0,621	0,03	0,129	0,02
		(0,339)		(0,381)		(0,338)	
Industriprod. ellers	$\beta_{1,15}$	0,970	0,96	0,970	0,96	0,832	0,78
		(0,635)		(0,635)		(0,612)	

Vi ser at samtlige estimater er lavere enn én. I de 8 tilfellene vi finner signifikante estimater vil imidlertid tre av estimatene ikke være signifikant forskjellig fra én. Dette gjelder næringene «trevarer», «treforedling» og «mineralske produkter». For å vurdere robustheten i disse estimatene har vi i tabell 4.2.1 plottet de estimerte $DFBETAS_{ij}$ for estimatene for de ulike 15 næringene. De horisontale linjene angir de kritiske grensene. Verdt å merke seg i denne figuren er de observasjonene med særlig sterk innflytelse på estimatene. Eksempelvis ser vi av figuren at én observasjon i næring 13 bidrar til å redusere estimatet med anslagsvis to standardavvik. I tabellen under har vi reestimert relasjon 4.2.4 etter at vi har utelatt den observasjonen i hver næring som har størst innflytelse i retning av å redusere estimatet. Eliminasjon av disse observasjonene definerer en «øvre grense» for estimatene. Tilsvarende kan vi ved å utelate de observasjonene som i sterkest grad bidrar til å øke estimatet på elastisiteten få definert en «nedre grense». For næring 13 ser vi at utelatelse av den ovenfor nevnte observasjonen bidrar til å øke estimatet fra et insignifikant estimat på 0,255 til et signifikant estimat på 0,620. Den «øvre» og «nedre» grensen definerer et intervall som estimatoren vil befinne seg innenfor avhengig av hvilke enkeltobervasjoner som utelates. I dette konkrete tilfellet vil intervallet være gitt ved [0.155, 0.620]. Generelt ser vi at de fleste estimatene viser en stor relativ variasjonsbredde. Hvorvidt vi kan forkaste hypotesen om at elastisiteten er signifikant lik én er i flere tilfeller kritisk avhengig av enkeltobservasjoners innflytelse. Estimatene på de næringsavhengige elastisitetene er derfor lite robuste.

4.2.2. En anvendelse av «Added variable plot»

Den foregående modellen med næringsavhengige elastistister viste seg å gi lite robuste estimater. Vi vil nå innføre en mer generell modell:

$$(4.2.5)\quad \ln FOU_i = \beta_1 \ln S_i + \beta_2 \ln Se_i + \eta D(Se_i=0) + \gamma D(N_{ik}) + \theta D\ k=1,...,15(UK_i=0) + \varepsilon_i,\ \text{der } i=1,...,425 \text{ og}$$

Her der elastisten av FoU mhp. bransjeenheten ikke lenger næringsspesifikk. I tabell 4.2.2 kollonne (I) har vi gjengitt estimatene på koeffisintene som følger av minste kvadraters metode. Elastisitenen av FoU mhp. størrelse blir her estimert til å være lik 0,58. En F-test av hypotesen H_0 om at $\beta_1 = 1$ gir en testobservator $F(1,405)=93,93$ og forkaster derfor H_0. Som et neste steg vil vi nå igjen drøfte robustheten i estimatene. Dette vil bli gjort ved å kombinere DFBETAS med et *«added variable plot»* som nå vil bli introdusert.

Added variable plots (AV-plott)[7] er et nyttig diagnostisk redskap som illustrerer grafisk hvordan eventuelle

[7] Se A5 i vedlegget.

innflytelsesrike observasjoner påvirker en enkeltparameter i en multivariat modell. AV-plottet viser sammenhengen som gjelder mellom venstresidevariabelen og en gitt høyresidevariabel *etter at en har kontrollert for de andre høyresidevariablene* og kan utledes på følgende måte. Anta at den opprinnelige modellen er gitt ved:

$$\boldsymbol{Y} = \boldsymbol{X}\beta + \varepsilon$$

Betrakt så følgende alternativ til den opprinnelige modellen:

$$(*) \quad \boldsymbol{Y} = \boldsymbol{X}\beta + \theta\boldsymbol{z} + \varepsilon$$

der vi har inkludert en ny forklaringsvariabel **z**. Vi kan nå teste hvorvidt denne variabelen er signifikant på vanlig måte ved en t-test for $\theta=0$. Men vi kan også utlede et plot som illustrer grafisk hvor sterk denne sammenhengen er. Dersom vi multipliserer begge sider med $\mathbf{M_X} \equiv \mathbf{I_n} - \mathbf{X(X'X)^{-1}X'}$ får vi at:

$$M_X Y = M_X X\beta + \theta M_X z + M_X\varepsilon$$

Venstre side vil da være vektoren med OLS residualer. $\mathbf{M_X X}$ vil per definisjon være lik null siden $\mathbf{M_X}$ projiserer til planet ortogonalt på **X**. Dersom vi nå tar forventningen på begge sider av likhetstegnet har vi at:

$$(**) \quad \boldsymbol{E(e)} = \theta\boldsymbol{M_X z}$$

Dette gir at et plot av **e** mot $\theta\mathbf{M_X z}$ som tilnærmet liniært gjennom origo. Det kan videre vises ved Frisch-Waugh-Lovell Teoremet (se R. Davidson og J. MacKinnon, s.19-24) at estimatet for θ ved en regresjon av **e** mot $\theta\mathbf{M_X z}$ vil være identisk med estimatet for θ i (*). Dermed vil den estimerte helningskoeffisienten være lik den partielle koeffisienten i en multivariat modell. Ved å studere den bivariate sammenhengen gitt ved (**) kan en danne oss et bilde av robustheten i estimat av en partiell koeffisient i en multivariat analyse. Dette plottet har fått navnet «added variable plot» AV-plot (eller «partial regression leverage plot»).

Et «Added Variable Plot» for lnS er gitt i figur 4.2.5 der vi har antydet størrelsen på $DFBETAS_i$ ved størrelsen på sirklene. Observasjoner som er «signifikant» innflytelsesrike er angitt ved sitt observasjonsnummer. Av figuren ser vi at et fremtredende trekk er at de mest innflytelsesrike observasjonene opptrer i to hovedklynger: De befinner seg enten i gruppen av *store bransjeenheter med høy FoU-innsats* eller i gruppen av *små bransjeenheter med høy FoU-innsats.* En konveks sammenheng mellom lnFoU og lnS ville føyd disse observasjonene bedre. En antakelse om en konstant skalaelastisitet resulterer i at elastisiteten for de minste bransjeenhetene blir overpredikert mens den blir underpredikert for de mellomstore og store bransjeenhetene.

Vi ser at den mest innflytelsesrike bransjeenheten er 255. Dette er den samme bransjeenheten som øvet avgjørende innflytelse på den næringsavhengige FoU elastisiteten i næring 13. I tabell 4.2.2, kolonne (II) er parameterne reestimert etter at denne observasjonen er utelatt. Vi ser at estimatet på β_1 har økt til 0,61 mens de andre parametrene ikke påvirkes i nevneverdig grad.

Cohen og Klepper (1996) gir en forklaring på hva denne konveksiteten kan skyldes. Deres modell som ble referert i del 2 bygget på forutsetningen om at stordriftsfordelen ved FoU oppsto idet gevinsten ved en innovasjon ble antatt å avhenge av produksjonskapasiteten på innovasjonstidspunktet og den økte priskostnadsmargin den gav opphav til. I en situasjon der det ikke er faste kostnader ved FoU og ingen produktivitetsforskjeller i produksjonen av FoU, viser deres modell at det vil være en proporsjonal sammenheng mellom FoU og størrelse. Dersom en imidlertid slakker på disse forutsetningene og antar at ulik produktivitet i bransjeenhetene vil den observerte konveksiteten oppstå som et resultat. Deres argument er som følger: Selv om produktiviteten i bransjeenhetene ikke er korrelert med størrelsen, vil den være det for de bedriftene som utfører FoU. Disse utgjør et selektert utvalg siden små bedrifter med et lite produksjonsapparat må ha en høyere produktivitet ved FoU enn store for at de faste kostnadene ved FoU skal dekkes. De små bransjeenhetene som utfører FoU vil derfor måtte ha en gjennomsnittlig høyere produktivitet ved FoU enn de store. De vil derfor ha en høyere FoU intensitet og resultatet er den observerte konveksiteten.

I avsnitt 4.1 ble også en annen mulig årsak til denne konveksiteten diskutert. Den vil kunne oppstå som et resultat av at kun de mest FoU-intensive av bransjeenhetene som utfører FoU, rapporterer denne virksomheten. Små bransjeenheter som reelt sett utfører FoU, vil ikke budsjettere for FoU siden denne virksomheten utgjør en liten og integrert del av produksjonen. I

Tabell 4.2.2. log FoU regresjon (med og uten observasjon 255)

Variable	Relasjon 4.2.5	
	I	II
lnS	0,579**	0,606**
	(0,042)	(0,043)
lnSe	0,204**	0,201**
	(0,059)	(0,693)
D(Se=1)	2,100**	2,100**
	(0,698)	(0,693)
D(UK=1)	0,371*	0,308*
	(0,176)	(0,174)
R^2	0,49	0,51
H_0: $\beta_1 = 1$	F(1,405)=93,93 p-verdi=0,000	F(1,404)=84,02 p-verdi=0,000

Regresjonene inneholder 15 dummy-variable for næring. Relasjon 4.2.5 (I) og relasjon 4.2.5 (II) refererer til estimater henholdsvis før og etter at observasjon 255 er utelatt.

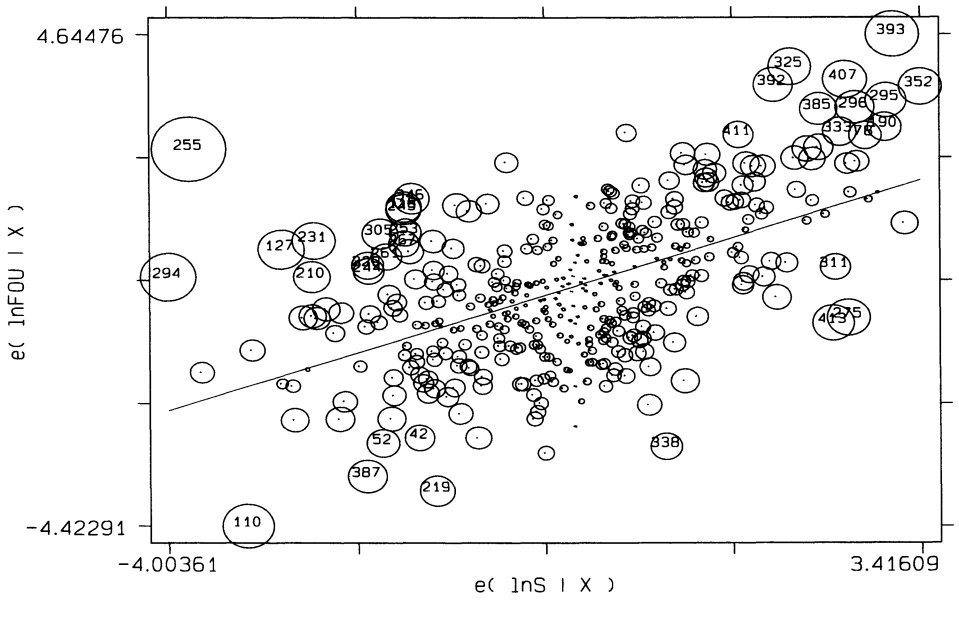
del 4.2 og 4.3 vil det bli drøftet nærmere hvordan en kan ta hensyn til disse seleksjonsproblemene ved estimering.

En tredje årsak til at flere av de minste bransjeenhetene er svært FoU-intensive kan være at de er del av et større foretak og utfører forskning også for andre deler av dette foretaket. En undersøkelse av de 15 «signifikant innflytelsesrike» observasjonene som opptrer i øvre venstre halvdel av figuren viser at 5 av disse er del av et større foretak. Dette gjelder forøvrig ikke for den mest innflytelsesrike observasjonen, 255.

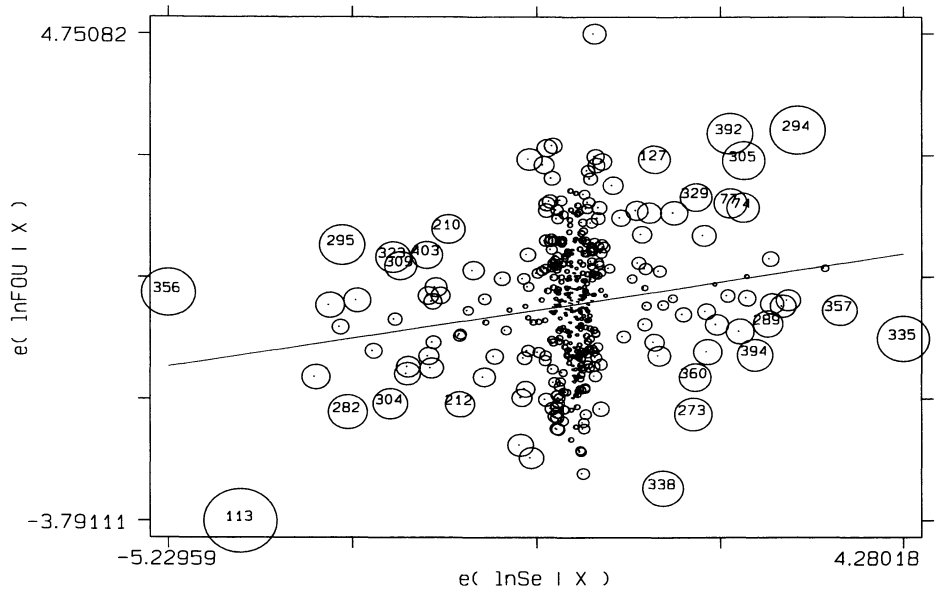
Det ble i innledningen pekt på at store foretak med diversifiserte aktiviteter lettere kunne dra nytte av et FoU prosjekt der utfallet er usikkert. Bedre muligheter til å anvende resultatene av et FoU-prosjekt kan ses på som en *«economies of scope»* effekt ved FoU og man skulle derfor forvente at FoU-virksomheten i en bransjeenhet også påvirkes av størrelsen på resten av foretaket. Resultatene av estimeringene viser både en signifikant nivåforskjell og en tiltakende tendens: Bransjeenheter som er del av et større foretak utfører både gjennomsnittlig mer FoU, samtidig øker FoU innsatsen med størrelsen på flerbedriftsforetaket. Dette kan være en indikasjon på at det er positive eksternaliteter ved FoU forbundet med det å være en del av et større foretak. Av tabellen ser vi at elastisiteten av FoU mhp. bransjeenhetsstørrelsen ble estimert til å være 0,2. I figur 4.2.6 er det vist et «added variable plot» for lnSe. Plottet viser at observasjon 113 er innflytelsesrik og har en avgjørende betyding for størrelsen på «scope-elastisiteten». Dersom vi eliminerer denne observasjonen avtar estimatet fra 0,20 til 0,17 men forblir signifikant på 1 prosentnivå. Eliminasjon av observasjon 356 og 335 bidrar på den annen side til å øke estimatet fra 0,20 til henholdsvis 0,23 og 0,22. Sammenhengen mellom størrelsen på restforetaket og nivået på FoU-innsatsen er derfor positiv og klart signifikant, men størrelsen på estimatet er likevel noe følsom overfor enkeltobservasjoners innflytelse.

Vi har inkludert en dummy-variabel for de bransjeenhetene som har en utenlandsk eierandel på mer enn 20 prosent. Av estimeringene i tabell 4.2.3 ser vi at disse bransjeenhetene har en signifikant høyere FoU aktivitet. Utenlandsk (del)eide foretak synes derfor å utføre mer FoU enn andre. En forklaring på dette kan være den internasjonale karakter vitenskap og teknologi antar. Utenlandsk (del)eierskap vil være en måte å skaffe seg lettere tilgang til den internasjonale kunnskapsbasen ved at det f.eks letter tilgangen på kvalifisert personell, sikrer bedre tilgang til vitenskapelig og teknologisk infrastruktur, gir økte muligheter til finansiell støtte, sikrer bredere og bedre markedsføring av produktene etc.

Figur 4.2.5. Added Variable Plot for lnS



Figur 4.2.6. Added Variable Plot for lnSe



4.3. Sensurering - en ikke lineær analyse

4.3.1. Utledning av modellen

Utgangspunktet for de foregående analysene, var de bransjeenhetene som oppga utgifter til FoU i NTNFs spørreundersøkelse. Det er imidlertid lite sannsynlig at den totale FoU virksomheten i industrien er fullt ut beskrevet ved dette utvalget: Det at en bransjeenhet *ikke rapporterer* FoU innebærer ikke nødvendigvis at den *ikke utfører* FoU. Mange små bransjeenheter rapporterer ikke FoU fordi de ikke opererer med noe klart skille mellom FoU og andre produksjonskostnader. Dersom den beslutningsprosessen som er bestemmende for om FoU blir rapportert, er avhengig av nivået på FoU eller de faktorene som er bestemmende for nivået på FoU, vil minste kvadraters metode kunne resultere i forventningsskjeve estimater. En vanlig Tobit modell vil også resultere i forventningsskjevhet siden dataene vil være generert ved en beslutningsprosess og derfor ikke vil være gitt ved en vanlig trunkering av sannsynlighetsfordelingen.

Heckman (1976) utviklet en modell som korrigerer for den type seleksjonsskjevhet vi her står overfor. Vi vil i det følgende ta utgangspunkt i en versjon av denne modellen behandlet i Bound et.al.(1980) og senere i Crepon et.al.(1994), der beslutningen om å rapportere FoU er modellert eksplisitt. Vi antar to latente variable gitt ved relasjonene (4.3.1) og (4.3.2). FoU_i^* beskriver bransjeenhetens faktiske utgifter til FoU og vil kun være observerbar dersom FoU blir rapportert. Videre antar vi at FoU blir rapportert dersom ln FoU_i^* er større enn en stokastisk nedre grense gitt ved ln C_i^*:

(4.3.1) $\ln FOU_i^* = X_{1i}\beta + \varepsilon_{1i}$

(4.3.2) $\ln C_i^* = X_{2i}\xi + \varepsilon_{2i}$

og

(4.3.3) $\ln FOU_i = \ln FOU_i^*$ hvis $|\ln FOU_i^* \geq \ln C_i^*$

(4.3.4) $\ln FOU_i = 0$ hvis $\ln C_i^* > \ln FOU_i^*$

der x_{ij} (j=1,2 ; i=1,...,n) er vektorer av eksogene variable for relasjon j, β og ξ er ukjente parametre og ε_{ij} er stokastiske restledd som antas å ha en binormal fordeling:

$$\begin{pmatrix}\varepsilon_1\\ \varepsilon_2\end{pmatrix} \sim N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\sigma_1^2 & \rho^*\sigma_1\sigma_2\\ \rho^*\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix}\right)$$

La D=1 hvis vi observerer FoU og la D=0 ellers. Vi har da at:

(4.3.6) $D = 1$ hvis $\ln FOU_i^* > \ln C_i^*$
dvs. $X_{1i}\beta - X_{2i}\xi + (\varepsilon_{1i} - \varepsilon_{2i}) > 0$

$$D = 0 \quad \text{hvis} \quad \ln FOU_i^* < \ln C_i^*$$
$$\text{dvs.} \quad X_{1i}\beta - X_{2i}\xi + (\varepsilon_{1i} - \varepsilon_{2i}) < 0$$

alternativt:

(4.3.7) $D=1$ hvis $X_{1i}\beta - X_{2i}\xi + u_{1i} > 0$

$D=0$ hvis $X_{1i}\beta - X_{2i}\xi + u_{1i} < 0$

der $u_{1i} = (\varepsilon_{1i} - \varepsilon_{2i})$. Vi har da at beslutningsprosessen vil være gitt ved en standard probit modell:

(4.3.8) $\Pr\{ \text{FoU observert} \} = 1 - F[(X_{1i}\beta - X_{2i}\xi)/\sigma]$,

der σ er variansen til restleddet u_{1i} og $F(\cdot)$ er den kumulative normalfordelings funksjonen. Vi ønsker å estimere relasjonen (4.3.1) men observerer bare utgifter til FoU dersom $D=1$.

Vi har at:

(4.3.9) $E[FoU^* \mid X_{1i}, X_{2i} \text{ og } D=1]$
$= X_{1i}\beta + E[\varepsilon_{1i} \mid X_{1i}, X_{2i} \text{ og } D=1]$
$= X_{1i}\beta + (\rho\sigma_1 / \sigma)\, E[u_{1i} \mid X_{2i} \text{ og } D=1]$
$\neq X_{1i}\beta$,

der ρ er korrelasjonen mellom u_{1i} og ε_{1i}. Vi ser her at korrelasjonen mellom det stokastiske restleddet i relasjonen som beskriver beslutningsprosessen og relasjonen som beskriver sammenhengen mellom størrelse og FoU gjør at vi ved kun å ta utgangspunkt i de FoU rapporterende bedriftene og estimere ved minste kvadraters metode vil kunne få forventningsskjeve estimater. For å få et inntrykk av graden av forventningsskjevhet kan vi utlede et uttrykk for $E[u_{i1} \mid X_{2i} \text{ og } D=1]$. Vi har at:

$E[u_{1i} \mid X_{2i} \text{ og } D=1] = E[u_{1i} \mid u_{1i} > -(X_{1i}\beta - X_{2i}\xi)]$
$= \sigma E[u_{1i}/\sigma \mid u_{1i}/\sigma > -(X_{1i}\beta - X_{2i}\xi)/\sigma]$
$= \sigma\, E[Z_i \mid Z_i > -k_i]$,

der $Z_i = u_{1i}/\sigma \sim N(0,1)$ og $k_i = (X_{1i}\beta - X_{2i}\xi)/\sigma$. Vi har videre at:

$$E[Z_i \mid Z_i > -k_i] = \int_{-k_i}^{\infty} z_i f(z_i \mid z_i > -k_i)\, dz_i$$

$$= \int_{-k_i}^{\infty} z_i \frac{f(z_i)}{1 - F(-k_i)}\, dz_i$$

Siden den deriverte av tetthetsfunksjonen til normalfordelingen er $-uf(u)$ og $(1 - F(-k)) = F(k)$ kan vi forenkle uttrykket over:

$$E[Z_i \mid Z_i > -k_i] = \frac{1}{F(k_i)}\left(-f(z_i)\Big|_{-k_i}^{\infty}\right) = \frac{f(k_i)}{F(k_i)}$$

siden $f(\infty)=0$ og $f(-k)=f(k)$. Denne raten er kjent som "Mills raten". Ved å sette $f(k)/F(k) = M(k)$ kan vi nå uttrykke (4.3.9) som:

(4.3.10) $E\ (FoU_i^* \mid X_{1i}, X_{2i} \text{ og } D=1)$
$= X_{1i}\beta + \rho\sigma_1 M(k_i)$

Så lenge korrelasjonen ρ er forskjellig fra null, vil minste kvadraters metode estimater som ignorerer $M(k)$ gi forventningsskjeve estimater. Det er to måter å løse dette problemet på. Vi kan estimere probit relasjonen som beskriver beslutningsprosessen (4.3.7) simultant med (4.3.1) ved maksimum likelihood. Vi kan også ta utgangspunkt i den prosedyren som ble foreslått av Heckman (1976). Metoden, en to-trinns-prosedyre, består i at vi først finner et estimat på $(X_{1i}\beta - X_{2i}\xi)/\sigma$ i (4.3.8) ved en probit analyse. Ved å nytte dette estimatet kan vi kalkulere $M(\hat{k}_i)$ eksplisitt og separat for hver enkelt observasjon. Vi kan da estimere relasjonen (4.3.10) ved å erstatte $M(k_i)$ med $M(\hat{k}_i)$. Heckman viser at dette vil gi konsistente estimater for ρ og β.

Vi vil følge begge disse prosedyrene. Vi estimerte først relasjonene ved Heckmans totrinnsprosedyre og deretter simultant ved maksimum likelihood (ML). Maksimum likelihood modellen vil ikke beskrives nærmere her; i stedet henvises det til Grilliches et.al (1978) og Nelson (1977). ML-metoden vil gi effisiente estimater og vil være å foretrekke ved hypotesetesting, mens Heckmans to-trinns-metode vil muliggjøre bruk av diagnostiske metoder. I modelleringen av beslutningsprosessen antok vi at bransjeenhetens størrelse (bruttoproduksjonsverdien, S) og bransjeenhetens kapitalintensitet (gitt ved K/S der K er definert ved brannsforsikringsverdien av maskiner og inventar) er avgjørende for den stokastiske nedre grensen C* som var bestemmende for hvorvidt bransjeenheten rapporterer FoU. Både S og K/S ble funnet i gjennomsnitt å være høyere for de FoU rapporterende bedriftene[8]. Gitt våre modellantagelser får vi i henhold til relasjon (4.3.7) følgende probit modell:

$D=1$ hvis $[\xi_1 \ln S_i + \xi_2 \ln(K_i/S_i) + \eta_{2i} + \gamma_{2k} + \theta_{2i}]$
$- [\beta_1 \ln S_i + \beta_2 \ln Se_i + \eta_{1i} + \gamma_{1k} + \theta_{1i}]$
$> \varepsilon_{2i} - \varepsilon_{1i}$

$D=0$ ellers

Alternativt kan vi skrive:

(4.3.11) $D = 1$ hvis $(\xi_1 - \xi_2 - \beta_1)\ln S_i + \xi_2 \ln K_i - \beta_2 \ln Se_i$
$+ (\eta_{2i} - \eta_{1i}) + (\gamma_{2k} - \gamma_{1k}) - \theta_2 > \varepsilon_{2i} - \varepsilon_{1i}$
$D = 0$ ellers

[8] Gjennomsnittlig bruttoproduksjonsverdi for rapporterende bransjeenheter var 29 mill. kroner og for ikke rapporterende 25 mill. kroner, mens gjennomsnittlig kapitalintensitet var henholdsvis 0,74 og 0,60.

(4.3.12) $D = 1$ hvis $\delta_1 \ln S_i + \delta_2 \ln K_i - \delta_3 \ln Se_i + \eta D(Se_i=0) + \gamma D_k - \theta D(UK_i=0) > u_{1i}$

$D = 0$ ellers

Ved å estimere probitmodellen gitt ved (4.3.10) kan vi kalkulere $\hat{M}_i = M(\hat{k}_i)$ der $\hat{k}_i = \hat{\delta}_1 \ln S_i + \hat{\delta}_2 \ln K_i - \hat{\delta}_3 \ln Se_i + \hat{\eta} D(Se_i=0) + \hat{\gamma} D_k - \hat{\theta} D(UK_i=0)] / \hat{\sigma}$. Vi kan nå i tråd med (4.3.10) inkludere $\hat{M}$ i FoU-relasjonen og estimere trinn to av Heckman prosedyren:

(4.3.13) $\ln FoU_i = \beta_1 \ln S_i + \beta_2 \ln Se_i + \beta_3 \hat{M}_i + \eta D(Se_i=0) + \gamma D(N_{ki}) + \theta D(UK_i=0) + \varepsilon_i$,

der i=1,...,418 og k=1,...,15

4.3.2. Estimeringsresultater

Etter at vi har korrigert for seleksjonsskjevhet ser vi av tabell 4.3.1 at estimatet for elastisiteten av FoU mhp. bransjeenhetstørrelsen er blitt omlag 30 prosent høyere samtidig som den ikke lenger er signifikant forskjellig fra én. Dette tyder på at modellen har klart å korrigere for at mange små bedrifter som i praksis er innovative ikke oppgir FoU samtidig som de små bedriftene som oppgir FoU er svært FoU-intensive. Strørrelsen på restforetaket er fortsatt signifikant men har avtatt noe. Bedrifter med utenlandsk kapitalinnskudd har imidlertid ikke lenger noe signinifikant høyere FoU-aktivitet. Av estimatene fra probit ligningene ser vi både størrelsen på bransjeenheten og kapitalinnsatsen inngår som signifikante forklaringsvariable og bidrar positivt til sannsynligheten for å oppgi FoU.

Etter å ha tatt hensyn til seleksjonsskjevhet kan vi ikke lenger forkaste hypotesen om proporsjonalitet mellom nivået på FoU og foretakstørrelsen. Heckmanns totrinnsprosedyre avslører imidlertid et alvorlig problem ved seleksjonsmodellen. I kolonne 4 ser vi at selv om inkludering av $M(\hat{k}_i)$ endrer estimatet på β_1 drastisk, vil den ikke selv være signifikant forskjellig fra null. Dette skyldes at $M(\hat{k}_i)$ som er en ikke-lineær funksjon av variablene gitt i tabellen, er sterkt korrelert med variablene som inngår i relasjonen for lnFoU. Vi står derfor overfor et kollinearitetsproblem og den observerte seleksjonseffekten vil derfor kunne avhenge av noen få avvikende observasjoner. For å få undersøkt dette nærmere kan det diagnostiske verktøyet igjen nyttes. Disse diagnostiske metodene er utledet innenfor rammen av en lineær modell, men Heckmans totrinnsprosedyre muliggjør likevel en anvendelse.

Dersom vi betrakter figur (4.3.1) og (4.3.2) ser vi at observasjonene 333, 104 og 313 til høyre i figurene og observasjon 110 nederst til venstre alle øver en signifikant innflytelse på den estimerte elastisiteten av FoU mhp. bransjeenhetsstørrelsen og koeffisienten til Millsraten. I figur 4.3.4 har vi plottet den bivariate sammenhengen mellom Millsraten og lnS. Dette plottet avslører en sterk korrelasjon (ρ=0,95). Samtidig er de observasjonene som var blant de mest innflytelsesrike på parameterestimatene til elastisiteten og Millsraten, alle i i ytterkanten av observajonssvermen. Disse observasjonene reduserer alle kollinerariteten og er derfor avgjørende for den seleksjonseffekten vi klarer å identifisere. Dersom vi utelukker disse observasjonene og estimerer på nytt, finner vi at maksimum likelihood estimatet på β_1 reduseres fra 0,82 til 0,73 samtidig som en χ^2 –test vil forkaste hypotesen om at β_1 lik én (p-verdi=0,031). Verifiseringen av hypotesen om at elastisiteten av FoU mhp.bransjeenhetsstørrelsen er lik én, er derfor avhengig av tre innflytelsesrike observasjoner.

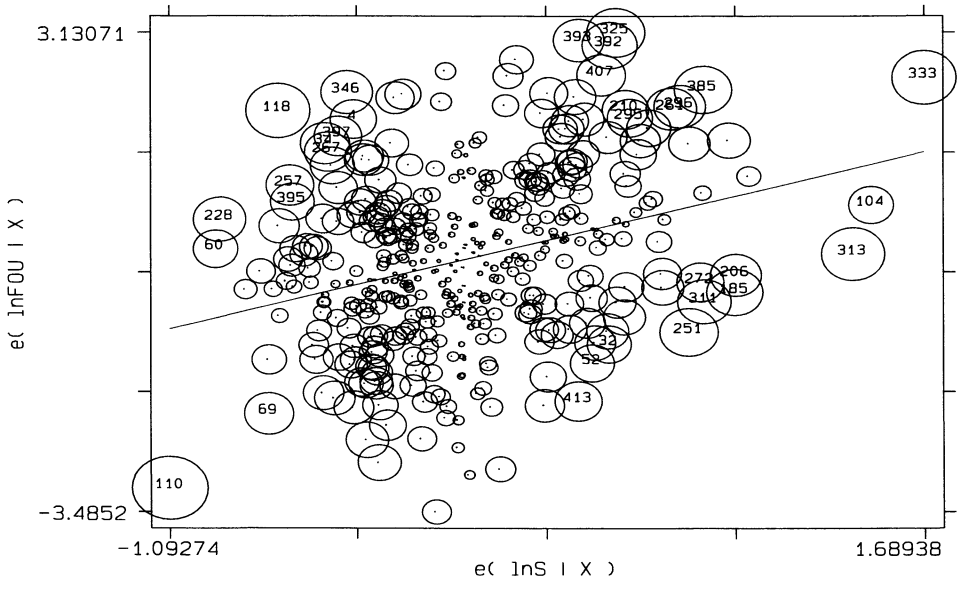
Tabell 4.3.1. log FoU regresjon korrigert for seleksjonskjevhet

Variable	MKM	Heckmanns 2-trinnsprosedyre		Maksimum likelihood (ML)	
	lnFoU relasjon	Probit regresjon	lnFoU	Probit regresjon	lnFoU
ln(S)	0,606** [1)]	0,259**	0,854**	0,274**	0,828**
	(0,043)	(0,042)	(0,142)	(0,426)	(0,107)
ln(Se)	0,201**	-0,013	0,158**	-0,013	0,161**
	(0,059)	(0,035)	(0,064)	(0,035)	(0,059)
ln(K)	-	0,300**	-	0,272**	-
		(0,037)		(0,035)	
$M(\hat{k}_i)$	-	-	0,589	-	-
			(0,340)		
D(Se=1)	2,100**	-0,188	1,599**	-0,197	1,632*
	(0,693)	(0,402)	(0,747)	(0,624)	(0,698)
D(UK=1)	0,308	-0,781**	-0,074	-0,777**	-0,048
	(0,174)	(0,090)	(0,259)	(0,088)	(0,232)
R^2	0,497	-	0,521	-	-
H_0: $\beta_1 = 1$	F(1,404)=84,02	F(1,402)=1,05		$\chi(1)$=2,58	
	p-verdi=0,000	p-verdi=0,305		p-verdi=0,108	

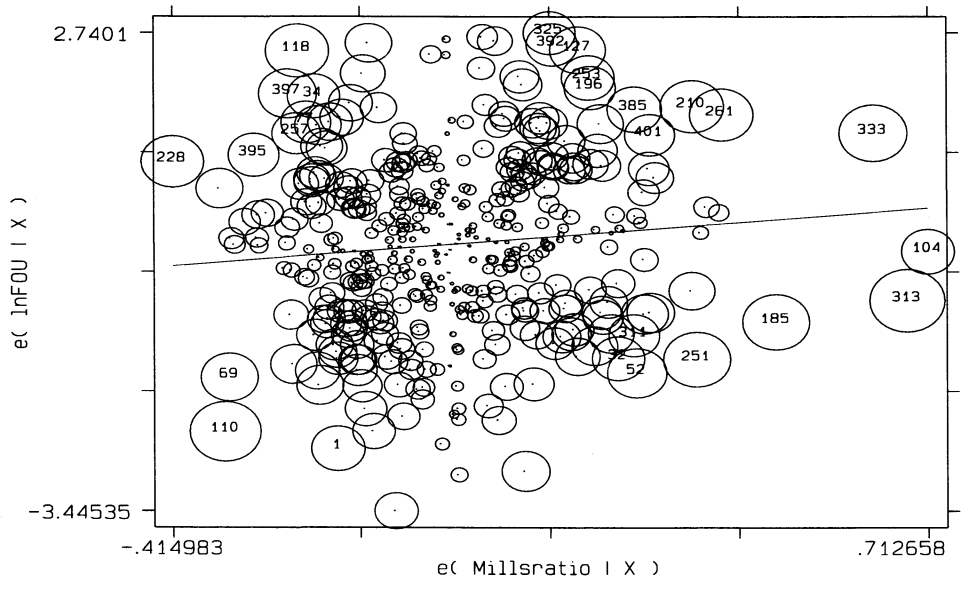
Alle regresjonene inneholder 15 næringsdummier. Siden vi vil ha hetroskedastiske restledd i trinn to av heckmanns to-trinnsprosedyre, har vi nyttet en robust metode i kalkuleringen av standardavvikene til estimatene.

[1] Observasjon 255 er nå utelatt fra estimeringene.

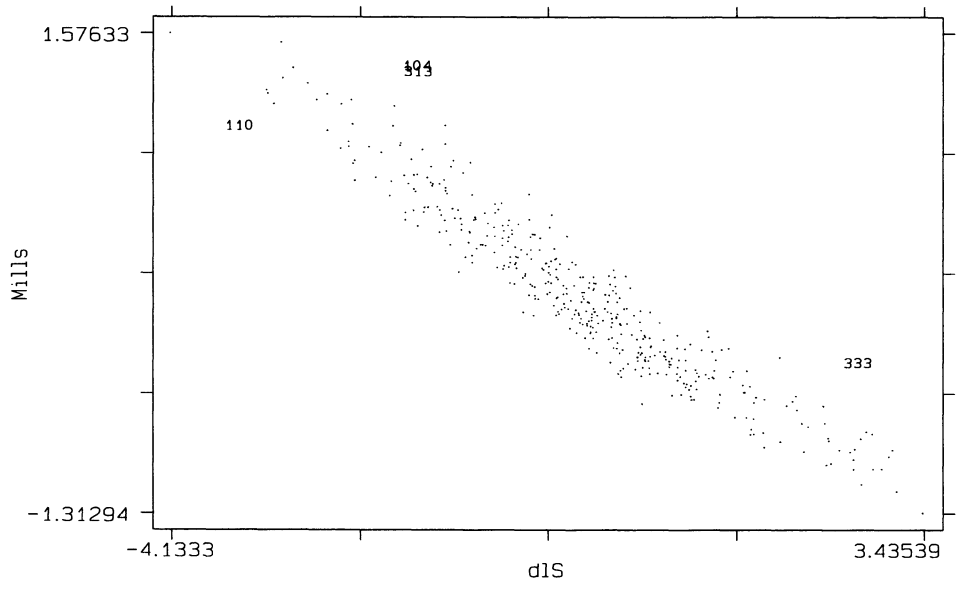
Figur 4.3.1. Added Variable Plot for lnS



Figur 4.3.2. Added Variable Plot for Millsraten (M)



Figur 4.3.3. Plot av Millsraten mot lnS



4.4. Størrelsesavhengig

4.4.1. Utledning av modellen

Etter å ha studert et "Added Variable Plot" for lnS i den lineære modellen (4.2.5) konkluderte vi med at antagelsen om konstant FoU elastisitet mhp. bransjeenhetsstørrelsen var for restriktiv siden mange av de minste bransjeenhetene ble funnet å være svært innflytelsesrike. To mulige og ikke gjensidig utelukkende forklaringer ble presentert. I avsnitt (4.3) antok vi at den høye gjennomsnittlige FoU-intensiteten ikke var reell, men var et resultat av at mange små innoverende bransjeenheter ikke oppgav FoU. Det resulterende seleksjonsproblemet ville dersom det ikke ble tatt hensyn til, føre til at den faktiske elastisiteten ble underpredikert. Beslutningsprosessen som ble antatt å være avgjørende for om en bransjeenhet oppgav FoU ble modellert eksplisitt og det resulterte i at estimatet av elastisiteten til FoU mhp. bransjeenhetsstørrelsen økte slik at den ikke lenger var signifikant forskjellig fra én. Den andre forklaringen er presentert i Cohen&Keller (1996): Det høye FoU-intensiteten blant de minste FoU-rapporterende bransjeenhetene kan skyldes at lønnsomheten knyttet til et FoU-prosjekt er avhengig utnyttelsen av innovasjonen ved produksjon, bestemt ved den initiale produksjonskapasiteten. Dersom det er forbundet store faste kostnader ved FoU og bransjeenhetene har en ulik FoU-produktivitet, vil kun de mest FoU-produktive av de minste bransjeenhetene finne det lønnsomt å innovere. Den høyere gjennomsnittlige FoU-intensiteten vil derfor beskrive et faktisk forhold og det vil være ønskelig å modellere dette eksplisitt. Ved å innføre lnS kvadrert kan vi i tråd med avsnitt (4.3) utlede relasjonene:

$$(4.4.1)\quad [\ln FOU_i \mid D=1] = \beta_1 \ln S_i + \beta_2 (\ln S_i)^2 + \beta_3 \ln Se_i + \eta D(Se_i=0) + \gamma D(N_{ik}) + \theta D(UK_i=0) + \varepsilon_{1i},$$

der $i=1,\ldots,418$ og $k=1,\ldots,15$.

Beslutningsprosessen vil i da være gitt ved probit modellen:

$$(4.4.2)\quad D = 1 \text{ hvis } \delta_1 \ln S_i + \delta_2 (\ln S_i)^2 + \delta_3 \ln K_i - \delta_4 \ln Se_i + \eta D(Se_i=0) + \gamma D_k - \theta D(UK_i=0) > u_{1i}$$
$$D = 0 \text{ ellers}$$

I denne modellen vil elastisiteten av FoU mhp. bransjeenhetsstørrelsen være gitt ved:

$$(4.4.2)\quad El_S FoU = \beta_1 + 2\beta_2 \ln S$$

og avhenger derfor av størrelsen på bransjeenheten.

4.4.2. Estimeringsresultater

I tabell (4.4.1) har vi gjengitt resultatene av estimeringene. Innenfor den opprinnelige lineære modellen finner vi en signifikant konveks sammenheng som gir

at elastisiteten til FoU mhp. bransjeenhetstørrelsen tiltar med bransjeenhetsstørrelsen men med en avtakende rate. Det følger av dette at FoU- intensiteten vil avta fra små til mellomstore/store bransjeenheter mens den for de største bransjeenhetene igjen vil tilta. Dersom vi inkluderer et kvadratisk ledd i seleksjonsmodellen vil seleksjonseffekten bli svakere samtidig som den konvekse sammenhengen ikke lenger er signifikant. Et forsøk på å ta hensyn til begge seleksjonsmekanismene i modelleringen er derfor problematisk siden Heckman-prosedyren og det kvadratiske leddet begge korrigerer for de samme bedriftene: Små bedrifter med høy FoU-intensitet. Det vil ikke være mulig å skille de to seleksjonsmekanismene fra hverandre.

For å få et inntrykk av sammenhengen mellom de estimerte skalaelastisitetene og bransjeenhetsstørrelsen ved de ulike estimeringsprosedyrene har vi plottet elastisiteten mot bransjeenhetsstørrelsen i figur (4.4.1). Spredningen av observasjonene er forsøkt vist ved kvadratene nederst i figuren. Det fremkommer da at gruppen av observasjoner med estimert skalaelastisitet større enn én omfatter relativt få bransjeenheter. For majoriteten av de norske bransjeenhetene vil derfor FoU-intensiteten avta og først tilta for de aller største bransjeenhetene. Det følger av estimatene at elastisiteten av FoU mhp. bransjeenhetsstørrelse vil være større enn én kun for bransjeenheter med bruttoproduksjon som er større enn 1,2 mrd.kr. dersom vi ikke korrigerer for seleksjon og 730 mill.kr. dersom vi korrigerer for seleksjon. Dette utgjør henholdsvis 4 prosent og 9 prosent av bransjeenhetene i vårt utvalg.

Et problem med en konveks spesifikasjon er at den i sterk grad kan drives av enkelte innflytelsesrike observasjoner. Figurene (4.4.2) og (4.4.3) viser «added variable plot» for lnS og $(\ln S)^2$. Plottene viser at ingen enkeltobservasjoner alene er avgjørende for den konvekse sammenhengen. Det motsatte synes derimot å være tilfelle: De to mest innflytelsesrike observasjonene, 110 og 206, reduserer konveksiteten. Observasjon 110 er en liten bransjeenhet og observasjon 206 en stor bransjeenhet som begge, målt ut fra næringsgjennomsnittet, har en lav FoU-intensitet. Den konvekse spesifikasjonen føyer seg mer etter gruppene av veldig små og veldig store *FoU-intensive* bransjeenheter og resulterer i at observasjoner som 110 og 206 blir mer innflytelsesrike. Ved eliminasjon av disse finner vi at minste kvadraters metode estimatet for β_1 reduseres fra -1,19 til -1,50 og ML-estimatet for β_1 fra -1,08 til -1,42, mens etimatet for β_2 øker fra 0,08 til 0,09 i begge modellene. Dette innebærer i praksis at resultatene ved minste kvadraters metode og innenfor seleksjonsmodellen i høy grad sammenfaller. Av figur 4.4.4 ser vi at 110 er den mest innflytelsesrike observasjonen på parameteren til Millsraten. De estimerte parametrene i seleksjonsmodellen avhenger igjen på en avgjørende måte av noen få innflytelsesrike observasjoner. Millsraten inngår ikke i modellen med en signifikant parameter siden den er sterkt korrelert med de andre parametrene.

Tabell 4.4.2. Minimum bransjeenhetsstørrelse som resulterer i tiltakende FoU-intensitet

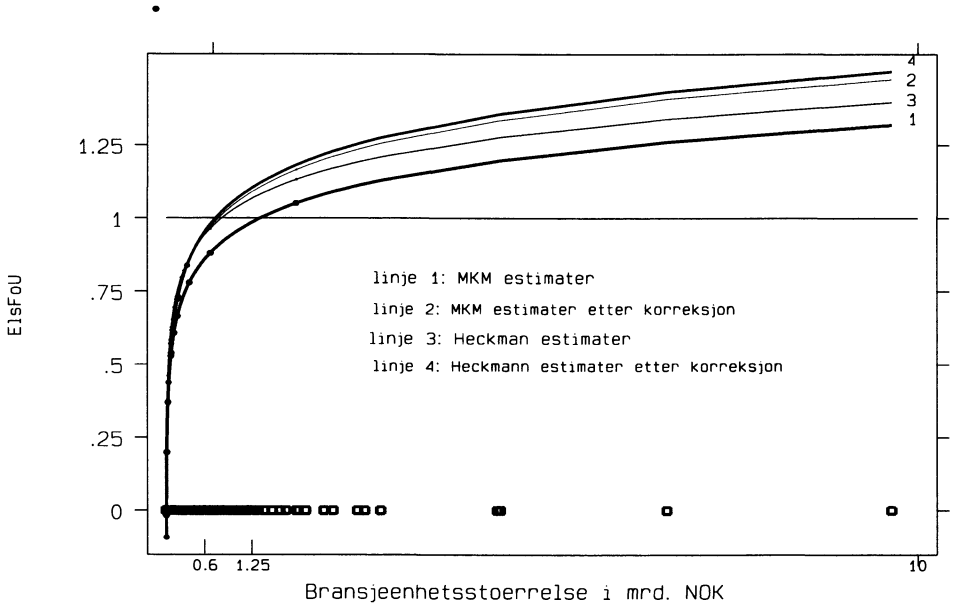
Estimering	Bransjeenhets-størrelse (S) som gir $El_sFoU > 1$	Utvalgsandel med estimert $El_sFoU > 1$
Minste kvadraters metode (MKM)	1,25 mrd. kr	4 pst.
Heckmann estimater	730 mill. kr	9 pst.
Minste kvadraters metode (uten observasjon 110 og 204)	672 mill. kr	10 pst.
Heckman estimater (uten observasjon 110 og 204)	612 mill. kr	11 pst.

Tabell 4.4.1. log FoU regresjon med størrelsesavhengig elastisitet og korreksjon for seleksjonsskjevhet

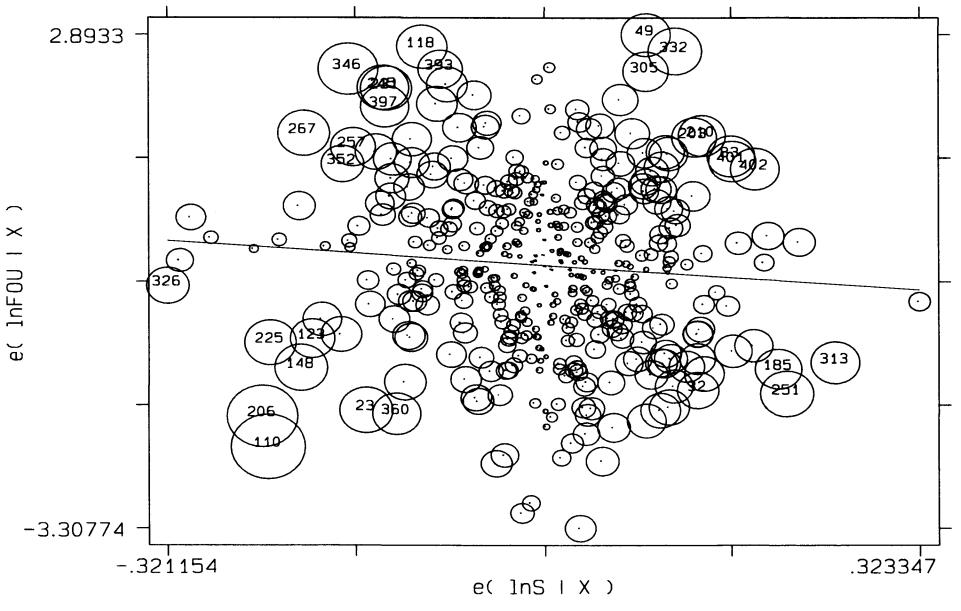
Variable	MKM	Heckmanns 2-trinnsprosedyre		Maksimum likelihood	
	lnFoU regresjon	Probit regresjon	lnFoU	Probit regresjon	lnFoU
ln(S)	-1,192**	1,127**	-1,078	0,127**	-1,078
	(0,409)	(0,223)	(0,600)	(0,222)	(0,582)
$(\ln S)^2$	0,078**	-0,041**	0,077**	-0,041**	0,077**
	(0,017)	(0,010)	(0,022)	(0,010)	(0,228)
ln(Se)	0,177**	-0,005	0,137*	-0,049	0,137*
	(0,057)	(0,035)	(0,058)	(0,035)	(0,056)
ln(K)	-	0,294**	-	0,294**	-
		(0,037)		(0,038)	
M(R)	-	-	0,147	-	-
			(0,355)		
D(Se=1)	1,834**	-0,121	1,443*	-0,121	1,443*
	(0,679)	(0,402)	(0,681)	(0,624)	(0,030)
D(UK*)=1)	0,325	-0,789**	-0,186	-0,788**	0,186
	(0,169)	(0,087)	(0,265)	(0,087)	(0,257)
R^2	0,540	-	0,543	-	-

Alle regresjonene inneholder 15 næringsdummier. Siden vi vil ha hetroskedastiske restledd i trinn to av heckmanns to-trinnsprosedyre, har vi nyttet en robust metode i kalkuleringen av standardavvikene til estimatene.

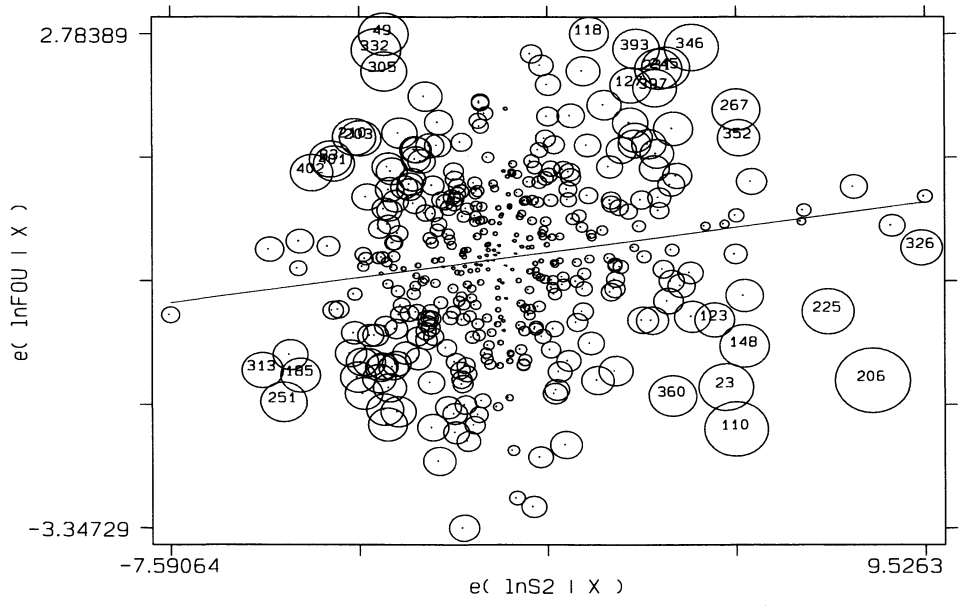
Figur 4.4.1. Elastisiteten av FoU mh. bransjeenhetsstørrelsen som funksjon av bransjeenhetstørrelsen



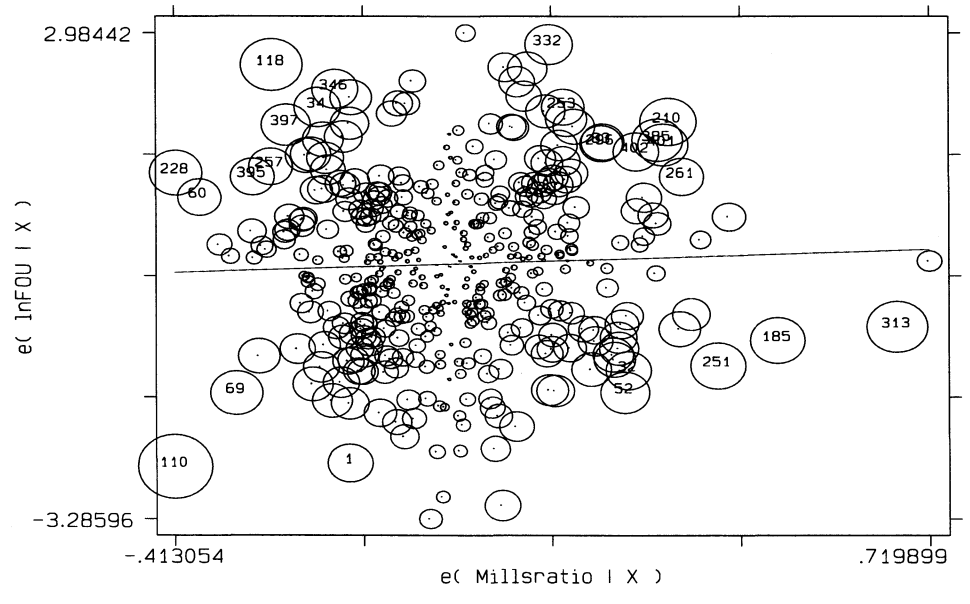
Figur 4.4.2. Added Variable Plot for lnS



Figur 4.4.3. Added Variable Plot for (lnS)²



Figur 4.4.4. Added Variable Plot for Millsraten(M)



5. Konklusjon

Det sentrale temaet i denne rapporten har vært sammenhengen mellom forskningsintensitet og foretaksstørrelse. Det blir regnet som et stilisert faktum at det eksisterer en nær og positiv sammenheng mellom utgifter til forskning og utvikling (FoU) og foretaksstørrelse, og at FoU vokser proporsjonalt med størrelse innenfor de fleste næringer. Et sentralt poeng har vært å forsøke å belyse hvorvidt dette stiliserte faktum er forenlig med den norske økonomiske virkelighet. Resultatet av analysene viser at det også for norske industriforetak eksisterer det en nær positiv sammenheng mellom FoU og størrelse. Det er imidlertid vanskeligere å påvise noen proporsjonal sammenheng mellom FoU-innsatsen og bransjeenhetstørrelsen.

Modellvalget var avgjørende for resultatene, og diagnostiske metoder tjente som et veiledende redskap både i modellvalg og i vurderingen av robustheten i de sammenhengene som ble estimert. Et sentralt problem ved modelleringen var de foretakene som ikke oppgir utgifter til FoU. I avsnitt 4.1 viste vi at sammenhengen mellom FoU og størrelse vil avhenge av om vi baserer oss på alle industribedriftene eller kun de som oppgir utgifter til FoU. Dersom vi tar utgangspunkt i alle industribedriftene og beregner den gjennomsnittlige FoU-intensiteten innenfor ulike kvintiler i størrelsesfordelingen, finner vi at den er konstant og derfor uavhengig av størrelse. Dersom vi imidlertid kun baserer oss på de FoU-rapporterende bransjeenhetene avtar FoU intensiteten monotont med størrelsen.

Ved kun å ta hensyn til de FoU-rapporterende bransjeenhetene og forutsette at elastisiteten til FoU mhp. bransjeenhetsstørrelsen er størrelsesuavhengig, blir elastisiteten (ved minste kvadraters metode) estimert til å være 0,6 og signifikant lavere enn én. Det er imidlertid lite sannsynlig at utvalget av FoU-rapporterende foretak fanger opp all den FoU-aktivitet som utføres i norsk industri. Dersom det er slik at sannsynligheten for at et innoverende foretak rapporterer FoU er avhengig av forhold som er sterkt korrelert med størrelsen på foretaket, vil vi stå overfor et seleksjonsproblem. De minste bransjeenhetene fremstår med en langt høyere gjennomsnittlig FoU-intensitet og det avgjørende spørsmålet blir da om dette skyldes mangelfull rapportering av FoU. Med en utgangshypotese om at mange små innovative foretak ikke skiller mellom FoU og annen produksjon og derfor ikke rapporterer FoU, ble en generalisert Tobit-modell introdusert. Her er beslutningen om å oppgi FoU modellert eksplisitt. Ved å estimere innenfor den generaliserte Tobit-modellen økte estimatet på elastisiteten med omlag 30 prosent til 0,83 og den var ikke lenger signifikant forskjellig fra null. Denne modellspesifikasjonen viste seg imidlertid som følge av et kollinearitetsproblem å være lite robust og diagnostiske metoder avdekket at noen få innflytelsesrike observasjoner var avgjørende for effekten av den foreslåtte seleksjonsmekanismen.

Mangelfull rapportering av FoU kan være en forklaring på den høye gjennomsnittlige FoU-intensiteten blant de minste FoU-rapportende foretakene. Det kan imidlertid også tenkes at det skjer en faktisk seleksjon ved at kun de mest produktive i FoU av de små bransjeenhetene finner det lønnsomt å innovere. Cohen og Klepper (1996) forklarer dette ved at lønnsomheten til FoU-prosjekt avhenger av en bransjeenhets initiale produksjonskapasitet og små bransjeenheter vil ikke, med unntak av de med høyest FoU-produktivitet, ha mulighet til å få dekket de faste kostnadene som er forbundet med FoU. Resultatet er at de minste foretakene som rapporterer FoU er svært innovative og har en gjennomsnittlig svært høy FoU-intensitet. For bransjeenheter med større produksjonskapasitet vil også de med lavere produktivitet i FoU finne det lønnsomt å innovere. Dette resulterer i en *reell* konveks sammenheng der de minste og største bransjeenhetene er mer FoU-intensive enn de mellomstore. Dette kan modelleres ved å anta en størrelsesavhengig FoU-elastisitet og en reformulering av den opprinnelige modellen resulterer i *signifikant konveks* sammenheng mellom størrelse og FoU slik at FoU-intensiteten først er avtakende men siden tiltakende for de aller største bransjeenhetene.

Det er rimelig å anta at begge disse seleksjonsmekanismene gjør seg gjeldende. Et forsøk på å estimere en størrelsesavhengig FoU-elastisitet innenfor den

generaliserte Tobit-modellen er imidlertid problematisk: Tobit-modellen og den størrelses-avhengige FoU-elastisiteten tar begge hensyn til den høye gjennomsnittlige FoU-intensiteten blant minste bransjeenhetene. Siden vi ikke vet hvilken av seleksjonsmekanismene som i hvert enkelt tilfelle gjør seg gjeldende vil et forsøk på å ta hensyn til begge typer av seleksjon innenfor dette rammeverket ikke være mulig.

I dataene var FoU oppgitt bransjeenhetsnivå og ikke foretaksnivå. Dette gjorde det mulig å separere effekten av størrelsen på bransjeenheten fra effekten av størrelsen på resten av foretaket. Betydningen av resten av foretaket på den enkelte bransjeeenhets FoU-innsats («scope»-effekten) kunne dermed estimeres. Resultatene viste at bransjeenheter som var del av et større foretak hadde signifikant høyere FoU-aktivitet enn andre bransjeenheter og at FoU aktiviteten tiltok med størrelsen på resten av foretaket. Elastisiteten av FoU mhp. størrelsen på resten av foretaket ble funnet å være omlag 0,2.

Referanser

Acs, J. S. og D.B. Audreutsch (1988): Innovation in Large and Small Firms: An empirical analysis, *American Economic Review* **78**, 680-690.

Belsley, D.A., K. Edwin og R.E. Welsch (1980): *Regression diagnostics: Identifying influental data and sources of collinearity,* New York: John Wiley & Sons.

Bound, J., C. Cummins, Z. Grilliches, B.H. Hall, og A. Jaffe (1984): *Who does R&D and who patents? R&D, patents and productivity,* Chicago: University of Chicago Press.

Chatterjee, S. og A.S Hadi (1986): Influential Observations, High Leverage Points, and Outliers in Linear Regression, *Statistical Science* **3**, 379-416.

Cohen, W.M og R.C. Levin (1989): «Empirical studies of innovation and market structure» i R. Schmalensee og R. D. Willig (red.): *Handbook of Industrial Organization* **2**, Amsterdam: North-Holland.

Cohen, R.M. og S. Klepper (1996): A reprise of size and R&D, *The Economic Journal* **106**, 925-951.

Cook, R.D. og S. Weisberg (1982): *Residuals and Influence in Regression,* New York: Chapman & Hall.

Crepon, B., E. Duguet og I. Kable (1994): A moderat support to Schumpeterian conjectures from various innovation measures, INSEE/CREST.

Dasgupta, P. (1986): «The theory of technological competition» i Stiglitz, J.E. og G.F. Mathewson, (red.): *New Developments in the analysis of marketstructure. Proceeding of a conferance held by the International Economic Association*, London: Macmillan.

Davidson, R og J.G. MacKinnon (1993): *Estimation and inference in econometrics,* Oxford: Oxford University Press.

Fisher, F.M. og P. Temin (1973): Returns to scale in reasearch and development: What does the Schumpeterian hypotesis imply? *Journal of Political Economy* **81**, 56-70.

Fisher, F.M. og P. Temin (1979): The Schumpeterian hypotesis, a reply? *Journal of Political Economy* **87**, 386-9.

Heckman, J. (1976): The common structure of statistical models of truncation, sample selection, and limited dependent variables and a simple estimator of such models, *The annuals of economic and social measurement* **5**, 475-492.

Krasker, W.S., E. Kuh og R.E. Welsch (1983): «Estimation for dirty data and flawed models» i Intriligator, M.D. og Griliches Z. (red): *Handbook of econometrics,* Amsterdam: North-Holland.

Klette, T.J. og Z. Grilliches (1997): Empirical patterns of firm growth and R&D-investment: A quality ladder model interpretation, Discussion Paper 188, Statistisk sentralbyrå.

Kohn, M.G. og J.T. Scott (1982): Scale economics in research and development: The Schumpeterian Hypothesis, *Journal of Industial Economics* **30**, 239-249.

Nelson, F.D. (1977): Censored regression models with unobserved stochastic censoring thresholds, *Journal of Econometrics* **6**, 309-327.

Rodriguez, C.A (1979): A comment on Fisher and Temin on the Scumpeterian Hypothesis. *Journal of Political Economy* **87**, 383-5.

Vedlegg A

Diagnostiske metoder

A1. Innledning

I en regresjonsanalyse vil det ofte være et utilfredstillende stort avvik mellom den idealiserte teoretiske basis og den praktiske anvendelsen. Inferens basert på minste kvadraters metode kan være sterkt påvirket av noen få enkeltobservasjoner og modellens prediksjonen vil kunne reflektere disse avvikende observasjonene i større grad enn det generelle mønstret i datamaterialet. To hovedretninger er fulgt for å redusere dette gapet mellom teori og praksis. Den ene hovedretningen baserer seg på robuste metoder i estimering og testing som krever færre strukturelle forutsetninger. Den andre hovedretningen tar utgangspunkt i den idealiserte postulerte modellen og søker å identifisere de faktorene som avviker fra denne spesifikasjonen. Utgangspunktet her er ulike diagnostiske mål som skal identifisere faktorer som ikke er i overenstemmelse med den modell-sammenhengen som er postulert. Identifikasjon av disse observasjonene vil kunne være interessant i seg selv samtidig som de vil kunne peke på svakheter og mulige modifikasjoner av den gjeldende modellen. Vi vil her gi en kort avgrenset redegjørelse for deler av den raskst voksende teorien for diagnostisering. Vi tar utgangspunkt i den vanlige lineære regresjons-modellen:

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon,$$

der $\mathbf{y}$ er en vektor av avhengige variable, $\mathbf{X}$ er en n×k matrise av k forklaringsvariable med rang k<n, β er en k-vektor av uavhengige parametre, og ε er en n-vektor av stokastiske restledd med betinget forventning og varians gitt ved $E(\varepsilon|\mathbf{X})=0$ og $Var(\varepsilon|\mathbf{X})=\sigma^2\mathbf{I}_n$, der σ^2 er ukjent parameter og $\mathbf{I}_n$ er identitetsmatrisen av orden n. Gitt denne modellstrukturen vil vi kunne basere oss på den tradisjonelle teorien for inferens som tar normalfordelingen som sitt utgangspunkt. Det er imidlertid avgjørende å undersøke i hvilken grad disse forutsetningene synes oppfylt i en gitt analyse-sammenheng. Vi vil nå utlede en del viktige mål som vil kunne uttrykke i avvik fra de ovenfor gitte antagelsene.

A2. Avvikende observasjoner og innflytelsesrike observasjoner

En nyttig distinksjon kan gjøres mellom avvikende observasjoner og inflytelsesrike observasjoner. *Avvikende observasjoner* er observasjoner som på bestemt måte avviker fra de andre observasjonene. *Inflytelsesrike observasjoner* er observasjoner som påvirker de estimerte parametrene på en avgjørende måte. To grunnleggende statistiske mål, hver for seg eller i kombinasjon, karakteriserer avvikende observasjone og inflytelsesrike observasjoner:

(i) Leverage
(ii) Skalerte residualer

En observasjon vil være gitt ved $\{y_i,\mathbf{x}_i\}$. Den kan være en avvikende observasjon fordi den avviker fra de andre punktene i $\delta(\mathbf{X})$. Dette vil være observasjoner med *høy potensiell innflytelse*. En observasjon kan også ligge fjernt i forhold til $E(y_i|\mathbf{x}_i)$. Dette vil resultere i *store residualer*. Vi kan selvfølgelig også ha observasjoner som innehar begge disse karakteristika.

Årsakene til disse avvikende observasjonene kan være mange og håndteringen av dem vil derfor igjen avhenge av det konkrete tilfellet vi står overfor. Videre analyser vil så kunne avgjøre om vi står overfor feil i datamaterialet eller hvorvidt vi har reelle avvik fra vår modellspesifikasjon som peker i retning av andre og bedre spesifikasjoner.

A3. Potensiell innflytelse

Minste kvadraters estimatene for β er gitt ved $\mathbf{b}=(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$. Vektoren av predikerte verdier er gitt ved

$$\mathbf{y}^P = \mathbf{X}\mathbf{b} = \mathbf{P}_\mathbf{X}\mathbf{y},$$

der

$$\mathbf{P}_\mathbf{X} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

er den *ortogonale projeksjonsmatrisen.* $\mathbf{P}_\mathbf{X}$ er symmetrisk og indempotent. La $\delta(\mathbf{X})$ være underrommet spent av $\mathbf{X}$, $\mathbf{P}_\mathbf{X}$ *projiserer* vektoren $\mathbf{y}$ ortogonalt ned på $\delta(\mathbf{X})$. Gitt ethvert underrom av $\mathbf{R}^n$ vil det alltid være to ortogonale projeksjonsmatriser, en som projiserer ethvert punkt i $\mathbf{R}^n$ ned i underrommet og en som projiserer punktet inn i underrommets ortogonale komplement, $\delta^\perp(\mathbf{X})$. Den matrisen som projiserer til $\delta^\perp(\mathbf{X})$ er $\mathbf{M}_\mathbf{X} \equiv \mathbf{I} - \mathbf{P}_\mathbf{X} \equiv \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. $\mathbf{P}_\mathbf{X}$ og $\mathbf{M}_\mathbf{X}$ vil vise seg å være svært nyttige i de videre utledningene.

Et mål for en observasjons potensielle innflytelse er gitt ved:

$$h_i = x_i(X'X)^{-1}x_i'$$

der h_i er det i-te diagonale element i matrisen $\mathbf{P}_\mathbf{X}$. De predikerte verdiene kan skrives som

$$y^P = \sum_{k=1}^{n} h_{ik}\ y_k = \sum_{k\neq i} h_{ik}\ y_k + h_{ii}\ y_i$$

Vi ser at effekten av $\mathbf{y}_i$ på $\mathbf{y}_i^P$ er kontrollert av det korresponderende diagonale elementet h_{ii} ($\equiv h_i$). En relativ stor h_i vil bety at den i'te observasjonen har sterk innflytelse på prediksjonen av y.

Potensiell innflytelse er et mål på en observasjons relative avstand i forhold til de andre observasjonene i $\delta(\mathbf{X})$. Vi vil nå se at h_i har en slik avstands-intrepretasjon. La $\mathbf{X}^*$ være $\mathbf{X}$-matrisen sentrert ved kolonnegjennomsnittet $\bar{x}$, slik at $h_i^* = (x_i - \bar{x})(X^{*\prime}X)^{-1}(x_i - \bar{x})'$. Vi ser av dette at observerasjoner med store verdier på h_i^* vil ligge relativt langt fra senteret av observasjonene målt i $(\mathbf{X}^{*\prime}\mathbf{X}^*)$-koordinatsystemet . En høy verdi av h_i^* vil være ekvivalent med en høy verdi av h_i, h_i vil derfor karakterisere en observasjon som ligger langt fra sentert av observasjoner i $\delta(\mathbf{X})$.

For å undersøke hvilke grenser h_i varierer innenfor kan vi omskrive uttrykket ovenfor:

$$h_i = e_i' P_X e_i = |P_X e_i|^2$$

der $\mathbf{e}_i$ er en n-vektor med 1 i den i-te posisjon og null ellers, dette følger av definisjonen av $\mathbf{P}_X$ og av at $\mathbf{e}_i' \mathbf{X} = \mathbf{X}_i$. Uttrykket viser at $\mathbf{h}_i$ er den kvadratiske lengden av en gitt vektor, dette sikrer at $h_t \geq 0$. Videre har vi at siden $|e_i| = 1$ og P_x er en ortogonal projeksjon av e_i vil $P_x e_i$ ikke kan ha lengde større enn e_i slik at $h_t = |P_X e_i|^2 \leq 1$. Vi kan dermed slutte at :

$$0 \leq h_i \leq 1$$

Videre følger det at summen av h_i over alle i er lik k. Dette kan vi se ved å nytte trase-operatoren:

$$\sum_{t=1}^{n} h_i = Tr(P) = Tr(X(X'X)^{-1}X)$$

$$= Tr((X'X)^{-1}X'X) = Tr(I_k) = k$$

h_i vil da i gjennomsnitt vil være lik k/n. Observasjoner som har h_i større enn 2k/n blir av Belsley et.al. (1980) foreslått som signifikante avvikende observasjoner.

A4. Innflytelse

En innflytelsesrik observasjon er en observasjon som har en unormalt sterk effekt på regresjonsresultatene. Denne innflytelsen kan være på:

(i) prediksjonen av **y**,
(ii) **b**,
(iii) b_j,

Hvert element i vektoren av OLS estimater for β er et veid gjennomsnitt av elementene i vektoren **y**. For å se dette kan vi definere $\mathbf{c}_i$ som den i'te rad i matrisen $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ og skrive

$$b = c_i y$$

Siden hvert element i **b** er et veid snitt, vil noen observasjoner ha større inflytelse på **b** enn andre. Hvis noen få observasjoner er sterkt innflytelsesrike, i den forstand at eliminasjon av disse vil forandre estimatene radikalt, vil det være av avgjørende betydning å få kartlagt disse observasjonene. Til dette formålet finnes det en rekke ulike mål som er ment å karakterisere observasjoners inflytelse.

Et utgangspunkt for en analyse av innflytelse er endringer i estimater og prediksjon ved eliminasjon av enkeltobservasjoner *"row deletion methods"*. Effekten av f.eks. en enkeltobservasjon på **b** kan sees ved en sammenligning med **b(i)**, estimatet for β på et utvalg der den i-te observasjonen er utelatt. La e_i være det i'te element av en vektor av OLS residualer: $\mathbf{e} = \mathbf{M}_x\mathbf{y} = (\mathbf{I}_n - \mathbf{P}_X)\mathbf{y} = \mathbf{y} - \mathbf{Xb}$. Et fundmentalt resultat er da at:

$$b(i) = b - \left(\frac{1}{1 - h_i}\right)(X'X)^{-1}X'e_i$$

Av dette utrykket ser vi at når e_i er stor og/eller $1-h_i$ er liten vil effekten av den i'te observasjonen på i hvertfall noen av elementene i **b** være stor. Basert på dette kan vi konstruere et mål for inflytelse kjent i litteraturen som *DFBETAS*$_i$ (Belsley et Al., 1980):

$$DFBETAS_i = \frac{b_i(i) - b_i}{s(i)\sqrt{(X'X)_{ii}^{-1}}} = \frac{e_i}{s(i)(1 - h_i)}$$

der $(X'X)_{ii}$ er det i'te diagonale element av (X'X). Dette uttrykket er utledet ved å skalere $b_i - b_i(i)$, ikke med sitt faktiske standardavvik, men med standardavviket til b_i, samtidig som s(i) erstatter σ.

Vi kan også undersøke hvordan utelatelsen av en observasjon påvirker modellens prediksjon. I Belsley et.al. (1980) er det vist at:

$$X_i\, b(i) = X_i\, b - \left(\frac{h_i}{1 - h_i}\right)e_i$$

Basert på dette målet kan vi konstruere et annet mål, *DFFIT*$_i$:

$$DFFIT_i = X_i\, b(i) - X_i\, b = \left(\frac{h_i}{1 - h_i}\right)e_i$$

Skalering av dette målet gir *DFFITS*:

$$DFFITS_i = \frac{X_i\, b(i) - X_i\, b}{s(i)\sqrt{h_i}} = \frac{\frac{h_i e_i}{(1 - h_i)}}{s(i)\sqrt{h_i}}$$

$$= \sqrt{h_i}\,\frac{e_i}{s(i)(1 - h_i)}$$

Vi ser at både $DFFIT_i$ og $DFBETA_{ij}$ er gitt ved en konstant multiplisert med $\hat{e}_i$. Belsley et.al. argumenterer for bruk av DFFITS og/eller $DFBETAS_{ij}$ og foreslår veiledende kritiske grenser for begge målene. Disse grensene er ikke absolutte, men er

avhengige av antallet observasjoner i den forstand at uansett datasettets størrelse vil alltid tilnærmet den samme andelen potensielt innflytelsesrike observasjoner bli identifisert.

DFFITS: I en perfekt balansert matrise av eksogene variable vil hver observasjon ha den samme potensiell innflytelse lik gjennomsnittet: $h_i = k/n$, $\forall i$. Innsatt for h_i i formelen for DFFITS får vi da at:

$$DFFITS_i = \sqrt{\frac{k}{n-k}}\, e_i^*$$

Vi har at $Pr\{e^* > 2\} \approx 0.05$. En veiledende kritisk grense kan da utledes ved:

$$|DFFITS_i| = |\sqrt{\frac{k}{n-k}}\, e_i^*| > 2\sqrt{\frac{k}{n-k}}$$
$$\approx 2\sqrt{\frac{k}{n}}$$

Den vil avhenge av utvalgstørrelsen og ta hensyn til at DFFITS vokser når k vokser. Det perfekte balanserte tilfellet vil her fungere som standarden som DFFITS blir målt mot. $DFBETA_{ij}$: Belsley et Al. foreslår å nytte $\pm 2/\sqrt{n}$ som kritisk grense.

A5. Added Variable Plots

Added Variable Plots er et nyttig redskap når vi ønsker å studere hvilken rolle variabelen **z** spiller dersom den inkluderes i en modell. I et slikt plot kan vi få et bilde av hvordan eventuelle innflytelsesrike observasjoner påvirker en enkeltparameter partielt innenfor en multivariat modell og er derfor et nyttig hjelpemiddel i en multivariat analyse. For å generere disse plottene velger vi en alternativ formulerig av modellen:

$$Y = X\beta + \theta z + \varepsilon$$

Vi har nå inkludert en ny forklaringsvariabel **z** og kan teste hvorvidt denne variabelen er signifikant på vanlig måte ved en t-test for $\theta=0$. Vi kan imidlertid også utlede et plot som illustrer grafisk hvor sterk denne sammenhengen er. Vi multipliserer begge sider med $\mathbf{M_x}$, matrisen som projiserer ned på $\delta^{\perp}(\mathbf{X})$. Vi får da at:

$$M_x Y = M_x X\beta + \theta M_x z + M_x\varepsilon$$

Venstre side vil da være vektoren med OLS residualer. $\mathbf{M_x}\,\mathbf{X}$ vil per definisjon være lik null siden **X** ligger i planet ortogonalt på $\delta^{\perp}(\mathbf{X})$. Dersom vi nå tar forventningen på begge sider av likhetstegnet har vi at:

$$E(\mathbf{e}) = \theta\mathbf{M_x}\,\mathbf{z}$$

Dette gir at et plot av **e** mot $\theta\mathbf{M_x}\,\mathbf{z}$ som tilnærmet lineært gjennom origo. Dette plottet har fått navnet "added variable plot" siden det er ment å vise effekten av å inkludere en ekstra variabel i modellen. Det kan videre vises ved Frisch-Waugh-Lovell Teoremet (ref.: Davidson et.al.1993, s.19-24) at estimatet for θ ved en regresjon av **e** mot $\theta\mathbf{M_x}\,\mathbf{z}$ vil være identisk med estimatet for θ i den ovenstående utvidede modell.

Variabelen **z** kan også representere en variabel allerede introdusert i modellen. Hvis U_k er matrisen som projiserer ortogonalt på underrommet spent av av alle vektorene i **X** bortsett fra x_k vil vi tilsvarende kunne plotte $(I - U_k)\,Y$ mot $(I - U_k)\,X_k$. Dette plottet har av Belsley et.al.(1980) blitt kalt "partial leverage regression plots" og viser hvordan de ulike observasjonene påvirker estimatet til den k-te parameter i modellen. For å få et inntrykk av innflytelse kan det være relevant å studere dette plottet i sammenheng men med DFBETAS.

A6. Konkluderende betraktninger

Vi har nå utledet en rekke ulike diagnostike mål, men i den praktiske anvendelsen av disse målene i denne konkrete analysen viste det seg snart at de i høy grad var overlappende og at DFBETAS og kombinert med "added variable plot" alene var effektivt og tilstrekkelig verktøy. De diagnostiske målene, ment å beskrive innflytelse på hele modellen under ett (DFFITS, h), gav innenfor den modellrammen vi braktet lite tilleggsinformansjon. Observasjoner som viste seg å være sterkt innflytelsesrike på enkeltparametre i modellen, var også avgjørende for modellens prediksjon eller sagt på en annen måte: En observasjon som øvet sterk innflytelse på modellens prediksjon framkom også som signifikant innflytelsesrik på en av enkeltparametrene. Det må imidlertid presiseres at dette ikke er noen generell konklusjon. Det kan tenkes situasjoner der en observasjon ikke er "signifikant" innflytelsesrik på noen enkeltparameter, men påvirker alle parametrene under ett og dermed likevel øver "signifikant" innflytelse på modellens prediksjon.

De sist utgitte publikasjonene i serien Rapporter
Recent publications in the series Reports

96/8 K.E. Rosendahl: Helseeffekter av luftforurensning og virkninger på økonomisk aktivitet: Generelle relasjoner med anvendelse på Oslo. 1996. 40s. 80 kr. ISBN 82-537-4277-0

96/9 S.-E. Mamelund og J.-K. Borgan: Kohort- og periodedødelighet i Norge 1846-1994. 1996. 236s. 165 kr. ISBN 82-537-4278-9

96/10 A. Schjalm: Kvalitetsundersøkelsen for Folke- og boligtelling 1990. 1996. 36s. 80 kr. ISBN 82-537-4279-7

96/11 K. Skrede og M. Ryen: Levekår i støpeskjeen. Status og utvikling i ungdomsgenerasjonenes materielle levekår 1990-1995. 1996. 80s. 95 kr. ISBN 82-537-4284-3

96/12 K.H. Alfsen, P. Boug and D. Kolsrud: Energy Demand, Carbon Emissions and Acid Rain: Consequences of a Changing Western Europe. 1996. 26s. 80 kr. ISBN 82-537-4285-1

96/13 M.W. Arneberg: Theory and Practice in the World Bank and IMF Economic Policy Models: Case study Mozambique. 1996. 28s. 80 kr. ISBN 82-537-4296-7

96/14 O. Skorge, F. Foyn og G. Frengen: Forsknings- og utviklingsvirksomhet i norsk industri 1993. 1996. 57s. 95 kr. ISBN 82-537-4306-8

96/15 K.O. Oftedal: Framskriving av markedssituasjonen for helse- og sosialpersonell fram mot år 2030. 1996. 66s. 95 kr. ISBN 82-537-4307-6

96/16 M.I. Hansen, T.A. Johnsen og J.Ø. Oftedal: Det norske kraftmarkedet til år 2020: Nasjonale og regionale fremskrivninger. 1996. 39s. 80 kr. ISBN 82-537-4316-5

96/17 K. Flugsrud og K. Rypdal: Utslipp til luft fra innenriks sjøfart, fiske og annen sjøtrafikk mellom norske havner. 1996. 52s. 95 kr. ISBN 82-537-4321-1

96/18 T. Fæhn og T. Hægeland: Effektive satser for næringsstøtte 1994. 1996. 79s. 95 kr. ISBN 82-537-4323-8

96/19 A. Bråten og L. Sandberg: Priser på jordbruksvarer: En analyse av statistiske kilder. 1996. 84s. 95 kr. ISBN 82-537-4325-4

96/20 E. Gulløy, S. Gåsemyr og A. Vedø: Forslag til et nytt system for norsk bistandsstatistikk. 1996. 50s. 95 kr. ISBN 82-537-4338-6

96/21 A. Thomassen og T. Tørstad: Prisstatistikk for næringseiendommer: Prøveundersøkelse for Oslo og Akershus. 1996. 31s. 80 kr. ISBN 82-537-4340-8

96/22 A.K. Essilfie: Investeringer, kostnader og gebyrer i den kommunale avløpssektoren: Resultater fra undersøkelsen i 1995. 1996. 44s. 80 kr. ISBN 82-537-4344-0

96/23 S. Glomsrød, A.C. Hansen og K.E. Rosendahl: Integrering av miljøkostnader i makroøkonomiske modeller. 1996. 46s. 95 kr. ISBN 82-537-4348-3

97/1 R. Jule: Produksjonsindeks for bygg og anlegg. 1997. 38s. 80 kr. ISBN 82-537-4355-6

97/2 T. Eika og K.-G. Lindquist: Konjunktur-impulser fra utlandet. 1997. 28s. ISBN 82-537-4357-2

97/3 T. Skjerpen and A.R. Swensen: Forecasting Manufacturing Investment Using Survey Information. 1997. 23s. ISBN 82-537-4374-2

97/4 E. Midtlyng: Arbeidsmiljø i skolen. 1997. 62s. 95 kr. ISBN 82-537-4390-4

97/5 B. Bjørlo og P. Schøning: Resultatkontroll jordbruk 1997: Gjennomføring av tiltak mot forurensninger. 1997. 85s. 95 kr. ISBN 82-537-4397-1

97/6 R.H. Kitterød: Leid hjelp til husarbeid? Bruk av privat rengjøringshjelp 1980-1995. 1997. 59s. 95 kr. ISBN 82-537-4399-8

97/7 S. Holtskog og K. Rypdal: Energibruk og utslipp til luft fra transport i Norge. 1997. 47s. 80 kr. ISBN 82-537-4400-5

97/8 K.O. Oftedal: Arbeidstilbudet fra sykepleiere og leger ved endret studie- og arbeidsmønster. 1997. 27s. 80 kr. ISBN 82-537-4401-3

97/11 S.E. Førre: Er store foretak mer forskningsintensive? En anvendelse av diagnostiske metoder. 1997. 33s. 100 kr. ISBN 82-537-4413-7

B

Returadresse:
Statistisk sentralbyrå
Postboks 8131 Dep.
N-0033 Oslo

Publikasjonen kan bestilles fra:

Statistisk sentralbyrå
Salg-og abonnementservice
Postboks 8131 Dep.
N-0033 Oslo

Telefon: 22 00 44 80
Telefaks: 22 86 49 76

eller:
Akademika – avdeling for
offentlige publikasjoner
Møllergt. 17
Postboks 8134 Dep.
N-0033 Oslo

Telefon: 22 11 67 70
Telefaks: 22 42 05 51

ISBN 82-537-4413-7
ISSN 0806-2056

Pris kr 100,00 inkl. mva.

Statistisk sentralbyrå
Statistics Norway

Falch Hurtigtrykk, Oslo