

Motivere på gruppe- eller individnivå?

- En empirisk studie av incentiver

av

Adrian Sivertsen

Masteroppgave

Masteroppgaven er levert for å fullføre graden

Master i samfunnsøkonomi

Universitetet i Bergen, Institutt for økonomi

Juni 2012

UNIVERSITETET I BERGEN



Forord

Denne oppgaven markerer slutten på min tid som student ved Universitetet i Bergen. Etter en lærerik periode – om enn hektisk og lett krydret med frustrasjon mot slutten – kan jeg snart se tilbake på en fire fine år som samfunnsøkonomistudent.

Først av alt en stor takk til min veileder, Gaute Torsvik, for innspill, råd, veiledning og ikke minst motivasjon gjennom hele arbeidsprosessen. At du hele tiden har vært tilgjengelig for å svare på spørsmål, både saklige og mindre saklige grunnet tilløp til panikk på oppløpssiden, gjør at denne oppgaven endelig ble ferdig.

En stor takk også til de glade sørlendingene, Frank og Signe, for korrekturlesing, og for å være to av dem som har vært med på å gjøre studietiden min til en fantastisk opplevelse.

Til slutt en hilsen til resten av gjengen som har vært med på å skape hverdagen min, stort sett fra morgen til kveld, både på og utenfor institutt for økonomi de siste årene. Takk for nå!

Adrian Sivertsen

Adrian Sivertsen, Bergen 13. juni 2012

Sammendrag

Motivere på gruppe eller individnivå?

- En empirisk studie av incentiver

av

Adrian Sivertsen, Master i samfunnsøkonomi

Universitetet i Bergen, 2012

Veileder: Gaute Torsvik

Problemstillingen i oppgaven er som følger: Vil en bonusordning som vurderer og belønner på individnivå ha en sterkere incentiveeffekt enn en ordning med ren gruppebonus?

Problemstillingen besvares ved å gjennomføre en økonometrisk analyse på data fra kundesenteret i et norsk forsikringselskap. I 2004 ble det gjennomført endringer i de ansattes bonusordning, og man erstattet gradvis den eksisterende ordningen med gruppebonus med en ordning hvor individuelle prestasjoner i større grad ble vurdert og belønnet. Siden ikke alle ansatte var omfattet av denne ordningen gir dette tilgang på en kontrollgruppe, og *difference-in-difference* (DiD) - metoden brukes for å identifisere effekten av reformen. Regresjonene er gjennomført i programvaren Stata. Resultatene tilsier at endringen i bonussystem hadde en signifikant positiv effekt på de ansattes ukentlige salg av forsikringer, og DiD-estimatoren indikerer en økning i gjennomsnittlig salg på mellom 1.506 og 1.583 enheter per uke. Regresjonsmodellene er spesifisert slik at dette resultatet tolkes som endring i salg av forsikringer som følger av at de ansatte yter høyere innsats.

Resultatene er dermed i tråd med teoretiske modeller som predikerer at ordninger med gruppebonus ikke gir arbeidstakere tilstrekkelige incentiver til å yte innsats.

På bakgrunn av regresjonsanalysen kan det ikke påvises forskjeller i hvordan menn og kvinner responderte på reformen. Det kan heller ikke konkluderes med at det var forskjell i hvordan individer eller team med relativ høyt eller lavt salg før reformen ble påvirket.

Innhold

Forord	ii
Sammendrag	iii
Kapittel 1 - Innledning	1
1.1 Bakgrunn for oppgaven.....	1
1.2 Oppgavens struktur	2
Kapittel 2 – Teoretisk rammeverk og tidligere funn	3
2.1 Modell med full informasjon.....	3
2.2 Modell med asymmetrisk informasjon.....	5
2.3 Modell med gruppebonus.....	5
2.4 Modell med individuell bonus.....	6
2.5 Andre forhold	7
2.6 Tidligere funn	8
Kapittel 3 - Bakgrunnsinformasjon	11
3.1 Generelt om bedriften	11
3.2 Reformen.....	11
Kapittel 4 - Metode	13
4.1 Difference-in-Difference estimatoren	13
4.2 Estimeringsmodeller	16

Kapittel 5 - Data	18
5.1 Generelt.....	18
5.2 Variabler.....	18
Kapittel 6 - Regresjonsanalyse	24
6.1 Resultater fra standardutvalget	24
6.2 Valg av estimeringsmodell og spesifisering av regresjonsmodell	26
6.3 Restriksjoner på utvalget.....	28
6.4 Log-level modell	29
6.5 Seleksjonseffekt.....	30
6.6 Inferensgrunnlag.....	31
6.7 Kjønnforskjeller.....	33
6.8 Teamforskjeller.....	34
6.9 Individforskjeller	36
6.10 Utvalgsstørrelse.....	37
Kapittel 7 – Oppsummering og konklusjon.....	39
Kapittel 8 - Avslutning	42
Appendiks A: Regresjonsresultater	43
Appendix B: Tester	50
Appendix C: Figurer	51
Litteraturliste.....	53

Tabeller

Tabell 1: Deskriptiv statistikk for de kontinuerlige variablene, hele utvalget.....	21
Tabell 2: Deskriptiv statistikk for de kontinuerlige variablene for kontroll- og behandlingsgruppen separert.....	22
Tabell 3: Hvor mange henvendelser som kreves for å gjennomføre et salg.....	23
Tabell 4: DiD-estimatoren fra forskjellige estimeringsmodeller og spesifikasjoner av regresjonsmodellen på standardutvalget.....	25
Tabell 5: DiD-estimatoren når det legges restriksjoner på utvalget.....	29
Tabell 6: DiD-estimatoren når avhengig variabel er $\log(\text{sumsalg})$	30
Tabell 7: DiD-estimatoren når individer som ikke opptrer i begge perioder fjernes fra utvalget.....	31
Tabell 8: Cluster robuste, bootstrap, og OLS-standardfeil.....	32
Tabell 9: DiD-estimatoren fra standardutvalget for menn og kvinner separert.....	33
Tabell 10: Koeffisientestimer for interaksjonsledd mellom produktivitet og reform.....	35

Figurer

Figur 1: Difference-in-difference estimatoren.....	15
Figur 2: Utviklingen i gjennomsnittlig salg gjennom observasjonsperioden.....	19
Figur 3: Gjennomsnittlig salg per team før og etter reform.....	34

Kapittel 1

Innledning

1.1 Bakgrunn for oppgaven

«A cornerstone of the theory in personnel economics is that workers respond to incentives».

[Edward P. Lazear (2000, s. 1346)]

I prinsipal-agent teori oppstår behovet for incentivsystemer fordi det finnes en interessekonflikt mellom arbeidstaker (agent) og arbeidsgiver (prinsipal), og det antas at arbeidstaker vil velge en atferd som ikke er forenlig med arbeidsgivers – eller bedriftens – målsettinger.¹ I økonomiske modeller fremstilles dette ved at agentens (nivå av) *innsats* ikke er optimal i forhold til prinsipalens ønsker. Gjennom sin innsats skal agenten bidra til at prinsipalen når sine mål – eksempelvis i form av inntekter eller omsetning – men fordi agenten er tilbøyelig til å «skulke», vil ikke disse målene oppnås.² Man antar også at prinsipalen har liten eller ingen anledning til å observere agentens innsats direkte, og at agenten dermed besitter privat informasjon om egen innsats. På bakgrunn av disse to forholdene – interessekonflikt og asymmetrisk informasjon – er det vanskelig å oppnå effisiente, eller *first-best*, løsninger. Utfordringen for prinsipal-agent teorien er å løse dette problemet gjennom passende incentivsystemer.

Et naturlig oppfølgingsspørsmål til sitatet av Lazear over vil være *hvordan* bedrifter bør utforme bonussystemer for å oppnå ønsket oppførsel hos arbeidstakerne. Det finnes forskjellige syn på hvilke bonusdesign som gir høyest grad av incentiveffekt, og det kan også stilles spørsmål ved om forskjellige bonussystemer egentlig har ønskelig effekt eller påvirker ønskelige forhold. Forskjellige teorier predikerer forskjellige utfall, og man har prøvd å besvare dette både gjennom empiriske analyser og kontrollerte eksperimenter. I praksis opererer også forskjellige bedrifter med ulike former for bonussystem, selv om man intuitivt kan tenke seg at de overordnede formålene med bonussystemer i stor grad vil være

¹ Situasjoner skjult handling kalles ofte *moral hazard*, eller moralsk risiko.

² Det virker kanskje kynisk å anta at de ansatte vil «skulke» fra sine arbeidsoppgaver dersom de får anledning. Andre forklaringer kan eksempelvis være at de ansatte ønsker å utføre arbeidsoppgaver de selv anser som givende (f.eks private prosjekter) men som ikke nødvendigvis er i samsvar med bedriftens ønsker, eller at det er en uoverensstemmelse mellom arbeidsgiver og arbeidstager om hva som bør være bedriftens overordnede mål.

samsvarende.³ Det finnes derfor ikke et entydig svar på spørsmålet om hvordan bonussystemer kan utformes mest hensiktsmessig.

Siden gruppebonus har egenskaper som kan sammenlignes med frivillig bidrag til et kollektivt gode, vil mikroøkonomisk teori predikere at effisiente nivå av innsats ikke oppnås på grunn av gratispassasjeroppførsel. Kort sagt betyr dette at gruppebonus ikke gir arbeidstakere tilstrekkelige incentiver til å yte ønskelig nivå av innsats så lenge innsats har en negativ effekt på individers nytte. Det finnes imidlertid andre syn på dette, og faktorer som eksempelvis altruisme, sosiale normer eller tilhørighet til arbeidsplassen brukes for å argumentere for hvorfor gruppebonuser kan være mer effektivt enn standard økonomisk teori tilsier.

Det faktum at teorien for incentiver og motivasjon gir prediksjoner som peker i ulike retninger, gjør det ekstra viktig å foreta en empirisk studie av om en bonusordning som belønner på individnivå vil kunne ha en sterkere incentiveffekt enn en ordning med gruppebonus. Det resultatet man finner fra en slik empirisk analyse kan være til hjelp for å vurdere de ulike teoretiske modellene for incentiver på arbeidsplassen.

1.2 Oppgavens struktur

Resten av oppgaven er bygd opp som følger: Kapittel 2 viser hvordan økonomisk teori gir forskjellige prediksjoner i forhold til incentiver og motivasjon, og presenterer også noen tidligere funn på området. Kapittel 3 gir en kort presentasjon av bedriften som datasettet er hentet fra, og reformen som ble gjennomført. Kapittel 4 forklarer *difference-in-difference* metoden, som er brukt for å identifisere effekten av endringen i bonussystem. I kapittel 5 presenteres datasettet, mens den økonometriske analysen og resultater gjengis i kapittel 6. Kapittel 7 gir en oppsummering av hovedresultater og konklusjon, mens kapittel 8 avslutter oppgaven.

³ Bragelien (2005) nevner fem positive effekter som intuitivt fremstår som generelle formål dersom man velger å innføre en bonusordning: De ansatte jobber hardere, de ansatte jobber smartere, de ansatte tar bedre beslutninger, bedriften tiltrekker seg bedre kandidater, og bedriften beholder de beste ansatte.

Kapittel 2

Teoretisk rammeverk og tidligere funn

I dette kapitlet settes det opp en enkel modell som beskriver hvordan effisiente (*first-best*) løsninger kan oppnås i tilfeller med perfekt informasjon. Modellen er en forenkling av interaksjonen mellom arbeidsgiver og arbeidstaker, men tjener som en illustrasjon av de sentrale mekanismene i standard prinsipal-agent teori. Deretter brukes modellen for å beskrive hvorfor behovet for incentivsystemer oppstår i tilfeller med skjult handling, hvorfor gruppebonus gir ineffisiente innsatsnivåer så lenge det finnes gratispassasjertilbøyelighet blant de ansatte, og hvordan dette kan løses ved å innføre en individuell bonusordning. Avsnitt 2.5 gir en gjennomgang av andre forhold som kan forklare arbeidstakeres oppførsel, og hvorfor ordninger med gruppebonus ikke nødvendigvis trenger å ha svakere incentiveffekt enn individbonuser. Til slutt gjengir avsnitt 2.6 resultater fra tidligere arbeid på området.

2.1 Full informasjon

Vi tar utgangspunkt i en profittmaksimerende bedrift, som sysselsetter en gruppe $N \geq 2$ homogene arbeidstakere.⁴ Gjennom de ansattes individuelle produksjon blir bedriftens totale produksjon, Y , bestemt ut i fra en prosess gitt ved følgende lineære sammenheng:

$$(2.1) \quad Y = \sum_i^N y_i$$

De individuelle bidragene, y_i , er igjen bestemt ut i fra følgende stokastiske prosess:

$$(2.2) \quad y_i = e_i + \varepsilon_i$$

Her er e_i individ i 's (nivå av) innsats, mens ε_i er en sjokkparameter som fanger opp uobserverbare forhold som kan påvirke et individs produktivitet.⁵ Det antas at $E[\varepsilon_i] = 0$. Variabelen e_i kan forstås som innsats *over* en terskelverdi, slik at $e_i = 0$ ikke betyr at den ansatte ikke yter noen grad av innsats. En positiv verdi av e_i kan heller forstås som at arbeidstakeren yter innsats utover det han eller hun «egentlig har lyst til», og uten kompensasjon ytes ingen innsats utover et minstenivå som er normalisert til 0.

⁴ Modellen som settes opp her tar utgangspunkt i Nalbantian og Schotter (1997), og er en enkel fremstilling av hvordan gruppebonus kan gi lavere innsatsnivå enn individbonus. For eksempler på mer fyldige modeller, se for eksempel Nalbantian og Schotter (1997) eller Holmstrom (1982).

⁵ Parameteren ε_i kan også brukes for å modellere målefeil, dvs. at prinsipalen ikke klarer å identifisere det reelle bidraget fra individ i .

Bedriftens totale inntekter er gitt som en konstant andel, $\gamma \in [0,1]$, av produksjon, og bestemmes dermed ut i fra følgende sammenheng:⁶

$$(2.3) \quad TR = \gamma Y = \gamma \sum_{i=1}^N y_i$$

Å yte innsats medfører kostnader for den enkelte ansatte, og dette må bedriften kompensere for gjennom lønn. Den negative nytten eller kostnaden ved å yte innsats er gitt som et sett av N strengt konvekse kostnadsfunksjoner:

$$(2.4) \quad c(e_i) = e_i^2 / \bar{c},$$

hvor \bar{c} er en positiv konstant. Videre er lønnskostnader de eneste utgiftene til bedriften, slik at bedriftens totale kostnader er gitt ved:

$$(2.5) \quad TC = Nw + \sum_{i=1}^N e_i^2 / \bar{c},$$

hvor w er fast lønn (eller reservasjonslønn) som er lik for alle ansatte. Bedriftens profittfunksjon er da gitt som $TR - TC$:

$$(2.6) \quad \pi = \gamma(\sum_{i=1}^N y_i) - (Nw + \sum_{i=1}^N e_i^2 / \bar{c})$$

Dersom den enkelte arbeidstakers innsatsnivå er observerbart for arbeidsgiver, kan *first-best* løsningen finnes ved å løse bedriftens maksimeringsproblem:

$$(2.7) \quad \max_{e_i} \{ \pi = \gamma(\sum_{i=1}^N (e_i + \varepsilon_i)) - (Nw + \sum_{i=1}^N e_i^2 / \bar{c}) \}$$

Førsteordensbetingelsen for dette maksimeringsproblemet er:

$$(2.8) \quad \frac{\partial \pi}{\partial e_i} = \gamma - 2e_i / \bar{c} = 0$$

Gitt at arbeidstakerne oppnår nytte i henhold til et sett av N nyttefunksjoner:

$$(2.9) \quad u_i = w - e_i^2 / \bar{c},$$

vil førsteordensbetingelsen over gi en Pareto-optimal løsning: Siden hver arbeider kompenserer for hvert nivå av innsats, vil både bedriftens profitt og arbeidernes nytte være

⁶ Det antas at verdien på γ bestemmes eksogent. Eksempelvis kan vi anta at bedriften produserer et gode som omsettes i et frikonkurransemarked, og at γ er markedsprisen på godet.

maksimert så lenge (2.9) holder.⁷ Den enkelte arbeidstakers nytte vil være konstant, og lik fast lønn.⁸ Det individuelle innsatsnivået som løser (2.6) er:

$$(2.10) \quad e_i = \gamma \bar{c} / 2$$

Siden γ blir eksogen bestemt representerer (2.10) en Nash-likevekt.

2.2 Asymmetrisk informasjon

Problemet er at e_i kan antas å ikke være observerbar for arbeidsgiver (det er asymmetrisk informasjon), og dermed kan ikke lønn avhenge direkte av innsats. Gitt at de ansatte kun utbetales en fast lønn, w , uten å kompenseres for innsats vil den enkelte yte den graden av nytte som maksimerer egen private nytte, gitt ved de individuelle nyttefunksjonene fra (2.9).

Førsteordensbetingelsen som løser individ i 's nyttemaksimeringsproblem er gitt som:

$$(2.11) \quad \frac{\partial u_i}{\partial e_i} = -2e_i / \bar{c} = 0$$

For at (2.10) skal holde må vi ha at $e_i = 0$, og dermed vil et lønnsystem uten noen incentiveeffekt resultere i at de ansatte ikke yter noen innsats utover et minstenivå.

Utfordringen blir derfor å finne mekanismer som gir innsatsnivåene fra (2.10). Forskjellige former for monetære bonuser som gir de ansatte incentiver til å yte høyere innsats på grunnlag av høyere lønnsutbetaling – og dermed også høyere nytte – er mulige løsninger. Spørsmålet blir da hvordan man designer slike bonusordninger for å oppnå ønskelig grad av incentiveeffekt.

2.3 Gruppebonus

Vi antar nå at bedriften velger å innføre en gruppebonus. Fordi bedriften ikke kan observere (eller måle) innsats direkte, blir løsningen å la bonus avhenge av et observerbart mål som det antas at gir en indikasjon på innsats. Et mulig mål vil for bedriften her være total inntekt, γY , siden denne via samlet produksjon bestemmes av den enkeltes innsats. Mer spesifikt antas det

⁷ Vi ser at å maksimere (2.9) er det samme som å minimere (2.4). Siden arbeidstakerne er homogene betyr det at samlet nytte, U , maksimeres gjennom (2.7).

⁸ Vi antar at $u_i = w$ er strengt større enn nytten fra enhver arbeidets beste alternativ. Denne deltakelsesbetingelsen sikrer at ikke noen individer kan oppnå høyere nytte (lønn) ved å ta et alternativt arbeid, og dermed at (2.8) og (2.10) representerer en Nash-likevekt.

at en fast andel, $\alpha \in [0,1]$, deles likt blant de ansatte. Nyttefunksjonen til den enkelte ansatte vil da være gitt ved:

$$(2.12) \quad u_i = w + \alpha\gamma(\sum_{i=1}^N y_i)/N - e_i^2/\bar{c}$$

Førsteordensbetingelsen som løser enhver ansatts maksimeringsproblem er:

$$(2.13) \quad \frac{\partial u_i}{\partial e_i} = \alpha\gamma/N - 2e_i/\bar{c} = 0,$$

og innsatsnivået for individ i som løser førsteordensbetingelsen er gitt ved:

$$(2.14) \quad e_i = \gamma\alpha\bar{c}/2N$$

I en hjørneløsning hvor $\alpha = 1$ vil all inntekt generert av produksjon deles mellom de ansatte. Dette vil likevel ikke gi det Pareto-optimale innsatsnivået fra (2.8), siden antallet gruppe-medlemmer nå også påvirker innsats negativt ved at marginalgevinsten fra bonus blir mindre dersom den totale summen som settes av til bonuspott må deles på et større antall individer. En økning i gruppestørrelsen gir dermed en «fortrengning» (*crowd-out*) av incentiveeffekten. Siden den marginale avkastningen fra bonusen heller ikke avhenger av egen innsats vil også den enkelte ansatte ha incentiver til gratispassasjeroppførsel, og dermed *free-ride* på de andre arbeidstakernes bidrag til samlet produksjon. Dette er også grunnen til at ordninger med gruppebonus kan relateres til situasjoner med frivillig bidrag til et kollektivt gode.

2.4 Individbonus

Alternativt til gruppebonus kan bedriften velge å innføre en ordning med individuell vurdering og utbetaling av bonus. Mer spesifikt kan vi anta at enhver ansatt får utbetalt en andel, $\rho \in [0,1]$, av sin egen produksjon i tillegg til fastlønn. De individuelle nyttefunksjonene vil da ta formen:

$$(2.15) \quad u_i = w + \rho y_i - e_i^2/\bar{c},$$

og førsteordensbetingelsen som løser nyttemaksimeringsproblem til individ i vil være gitt ved:

$$(2.16) \quad \frac{\partial u_i}{\partial e_i} = \rho - 2e_i/\bar{c} = 0$$

For å løse førsteordensbetingelsen må individ i velge et innsatsnivå gitt som:

$$(2.17) \quad e_i = \rho\bar{c}/2$$

Så lenge ρ settes lik γ , ser vi at dette gir samme innsatsnivå som ligning (2.10). Med andre ord vil en ordning hvor arbeidstaker mottar hele verdien av sin produksjon gi incentiver til å yte innsats tilsvarende den Pareto-optimale løsningen med full informasjon.⁹

2.5 Andre forhold

Modellene over bygger på antakelser som fremstår som restriktive i forhold til å beskrive virkeligheten. For det første er antakelsen om homogene arbeideredstakere et forhold som kan tenkes å begrense forklaringskraften til modellen, og det kan antas at individuelle faktorer også påvirker atferd. Dette gjør også at forskjellige incentivsystemer kan antas å ha forskjellig påvirkning på den enkelte medarbeider. Videre kan det også virke kynisk å anta at det kun er inntekt og innsats på arbeidsplassen som bestemmer et individs nytte. Det finnes derfor også teori som vurderer andre faktorer som tenkes å påvirke et individs valg av innsats, og som kan forklare hvorfor gruppebonuser ikke nødvendigvis gir lavere innsatsnivå eller produktivitet relativt til en ordning med individuell bonus.

At individer velger et positivt nivå av privat frivillig bidrag i *public-good* eksperimenter hvor den dominante strategien (per design) er perfekt gratispassasjeroppførsel, tilskrives ofte at de som opptrer i en gruppe har altruistisk motivasjon (Anderson et. al, 1998). Dette betyr at et individ kan ha en nytte som avhenger av *andres* nytte (i tillegg til andre faktorer). I så tilfelle vil ordninger med gruppebonus ikke nødvendigvis gi sub-optimale utfall, siden den enkelte kan velge å yte innsats høyere enn de monetære incentivene tilsier. En annen forklarende faktor kan være at individer velger å besvare andres rettferdige eller urettferdige handlinger med en tilsvarende handling. Slik resiprositet kan også forklare hvorfor gruppebonuser kan være gode instrumenter for å gi incentiver til ønsket oppførsel, så lenge belønningen anses for å være rettferdig av arbeidstakerne og de responderer ved å yte en høyere innsats på bedriftens vegne. I motsetning til altruisme som er ubetinget «godhet», vil ikke resiprositet i seg selv gi ønsket oppførsel siden det også kreves at arbeidstakere anser belønningen fra bedriften som rettferdig. Resiprositet kan også forklare arbeidstakeres oppførsel som respons på kollegaenes oppførsel. Dersom resten av teamet har et høyt bidrag vil agenter velge å gjengjelde dette gjennom eget bidrag og vice versa.

Kuvaas (2008) påpeker at ordninger med overskuddsdeling vil kunne bidra til psykologisk eierskap og fellesskapsfølelse. Også dette betyr at det kan oppstå høyere nivå av innsats enn

⁹ Dette forutsetter at produksjon er et fullverdig mål på innsats. I denne modellen tilsvarer produksjonen bidrag utover minstenivå slik at dette er oppfylt.

de monetære incentivene tilsier. Det kan også være at det generelt eksisterer en grad av ansvarsfølelse overfor bedriften, og det kan tenkes at ansatte ønsker å opptre i forhold til bedriftens beste uavhengig av variabel lønn. Dersom ansvarsfølelse gir tilstrekkelige motivasjon for å yte innsats kan det argumenteres for at ytterligere incentivsystemer er overflødig.

Et annet forhold som kan tenkes å gi incentiver til innsats er overvåkning fra kolleger (*peer monitoring*). Gjensidig overvåkning kan tenkes å oppstå i flere situasjoner, men det kan antas at det er enklere å gjennomføre dersom de ansatte er delt inn i mindre team. Dersom teamene også belønnes for samlet innsats gir dette også ytterligere grunnlag for den enkelte til å overvåke de andre medlemmene av teamet. Loughry og Tosi (2008) skiller mellom direkte og indirekte overvåkning. Direkte overvåkning oppstår når arbeidstakere observerer kollegaenes oppførsel eller resultater og responderer på en direkte måte, for eksempel gjennom tilbakemeldinger og instruksjoner. Indirekte overvåkning oppstår når arbeidstakere «straffer» kolleger som presterer relativt dårlig, eksempelvis gjennom å baksnakke eller unngå dem. Direkte overvåkning kan dermed tenkes å øke ansattes produktivitet gjennom å øke kunnskap eller forbedre rutiner, mens indirekte overvåkning gir den enkelte ansatte incentiver til å bidra for å unngå sosialt stigma. Kandel og Lazear (1992) gir en teoretisk gjennomgang av hvordan press fra kolleger kan gi incentiver til å bidra fremfor å være gratispassasjer, og hvordan straff kan følge av avvik fra en etablert norm innad i en gruppe. Videre kan også utveksling av kunnskap og samarbeid innad i gruppen bidra til at individuelle prestasjoner forbedres. Felles for begge disse faktorene – overvåkning og samarbeid – er at de kan tenkes å oppstå i situasjoner med team, mens innføring av individbonus vil kunne fortrenge dem.

På grunnlag av disse forholdene gir ikke økonomisk teori en entydig prediksjon av effekten av å erstatte gruppebonus med individuell bonus.

2.6 Tidligere funn

Det er gjennomført flere studier som ser på forskjellige incentivsystemer og effekten av disse.¹⁰ Et klassisk eksempel er Lazear (2000), som gjennomfører en empirisk studie for å se på effekten av å erstatte timelønn med et akkordlønnssystem med utgangspunkt i observasjoner fra *Safelite Glass Corporation*. I *Safelite* innførte man individuelle incentiver, og Lazear finner at økningen i produktivitet (antall installerte bilruter) blant de ansatte som

¹⁰ Prendergast (1999), eller Jenkins et. al (1998) gir god oversikt over eksisterende arbeid om incentiver.

følger av dette er på rundt 44 %, og at rundt 22 % kan tilskrives en incentiveeffekt. Også Paarsch og Shearer (1996) påviser en signifikant positiv effekt av incentiver på produktivitet i sin studie av akkordlønn mot timelønn i et canadisk treplantingsfirma. Her er det likevel en annen faktor som også spiller inn: selv om akkordlønn øker *kvantitet* (antall plantede trær) har den en negativ effekt på *kvalitet* (innsatsen de ansatte legger ned i arbeidet for å sørge for at et tre blir riktig plantet). Dette betyr at å gi incentiver til å øke produksjonsvolum ikke nødvendigvis er til bedriftens beste dersom dette går på bekostning av kvalitet.

Nalbantian og Schotter (1997) gjennomfører et eksperiment for å se på produktivitet under forskjellige gruppeincentivsystemer. Resultatene deres tilsier at «skulking» (*shirking*) forekommer dersom deltakerne plasseres i en setting med sterke incentiver til dette. De observerer også at overvåkning fra arbeidsgiver er en effektiv men kostbar måte å oppnå høyere innsats på. Hansen (1997) finner at innføring av gruppebonus blant finansielle rådgivere i *American Express* medførte en konvergenseffekt: De ansatte med lavest produktivitet før gruppebonus ble innført økte sine prestasjoner, mens den estimerte effekten for de mest produktive viste en relativt lavere – og i noen tilfeller også negativ – respons på bonusordningen. Samlet sett kan likevel resultatet av å innføre gruppebonus her anses for å være vellykket, ettersom gjennomsnittlig salg økte etter at ordningen ble innført.

I forhold til artikler som ser på en endring i eksisterende bonusordninger finner Hamilton et. al (2003) en økning i ansattes produktivitet på gjennomsnittlig 14 % som følger av å innføre team i en amerikansk klesprodusent. Samtidig gikk også bedriften fra en ordning med akkordlønn til en ordning med hvor hvert team ble belønnet i henhold til samlet produksjon. Siden lønnen sank for de mest produktive arbeiderne og det ikke var gratispassasjeroppførsel kan dette tolkes som at det er andre faktorer som påvirker produktivitet enn monetære incentiver. Forfatterne påpeker at gjensidig overvåkning kan forklare hvorfor det ikke oppsto gratispassasjertendenser, men at dette ikke er tilstrekkelig for å forklare den signifikante økningen i produktivitet. Denne effekten kan forklares av at innføringen av team økte de ansattes mulighet til å utnytte samarbeidsevner, og derigjennom også produksjonsmuligheter. Dermed var det en positiv effekt av team på produktivitet som ikke kan forklares av monetære incentiver. Helper et. al (2010) ser på forskjellige kompensasjonsformer med utgangspunkt i observasjoner fra to amerikanske bedrifter. Blant resultatene deres er at overskuddsdeling og timelønn gir reduserte prestasjoner for *easy-to-observe* oppgaver (eksempelvis produksjon), men økte prestasjoner for *hard-to-observe* oppgaver (eksempelvis forbedring av produksjonsprosesser og samarbeid). Forfatterne konkluderer på bakgrunn av dette med at

arbeidsomgivelser er av betydning for hvilken bonusordning som er mest passende. I tilfeller hvor arbeiderne produserer et produkt og arbeidsgiver overvåker, er det optimale incentivsystemet individuell akkordlønn. Dersom arbeidsoppgaver som er vanskelig å observere, eksempelvis samarbeid, planlegging, beslutningsprosesser og problemløsning, kreves for å utføre et tilfredsstillende arbeid kan andre typer kompensasjon (potensielt) være optimalt.

Kapittel 3

Bakgrunnsinformasjon

I dette kapitlet gis en kort presentasjon av den aktuelle bedriften datasettet stammer fra, og informasjon om de faktiske endringene som ble gjennomført i bonusordningen.

3.1 Generelt om bedriften

Som tidligere nevnt er datasettet som studeres i den økonometriske analysen hentet fra et kundesenter ved et norsk forsikringsselskap. Av hensyn til selskapets ønske om anonymitet blir ikke ytterligere konsernfakta presentert.

De ansatte ved kundesenteret behandler utelukkende henvendelser per telefon. Dersom en kunde ringer inn til kundesenterets fellesnummer kommer samtalen først inn i et køsystem, før den deretter blir rutet ut til en tilgjengelig agent. De ansatte er i observasjonsperioden organisert i til sammen 19 forskjellige team. To av disse består utelukkende av deltidsansatte og tilkallingsvikarer som ikke blir omfattet av bedriftens bonussystem, til sammen rundt 30 individer. I den økonometriske analysen vil disse fungere som kontrollgruppe. Hvert av de resterende 17 teamene er tilfeldig inndelt og består av rundt ti agenter. Hovedoppgaven til de ansatte er å yte service for selskapets kunder, men også salg av forsikringer og fornyelser av eksisterende kundeforhold er også arbeidsoppgaver som vektlegges. Dette kommer også eksplisitt frem i bonusordningene, som i tillegg til å fremme andre forhold er designet for å direkte belønne – og dermed gi incentiver til – salg av forsikringer. Agentene driver ikke oppsøkende salg, og samtaler som loggføres *ut* fra en agent vil derfor i stor grad kunne antas å ikke generere salg. Videre blir både salg, pålogget tid og antall telefoner målt ukentlig for hver ansatt.

Fra og med første kvartal 2003 hadde bedriften en bonusordning som belønnet de ansatte på bakgrunn av teamprestasjoner. Mer spesifikt ble en sum på inntil 11.000,- utbetalt til det enkelte teammedlem dersom teamet som helhet oppfylte et forhåndsbestemt krav. I denne perioden – frem til den reformen som blir omtalt under – var det dermed kun gruppebonus ved kundeservicesenteret.

3.2 Reformen

I 2004 begynte kundesenteret en gradvis innfasing av et nytt incentivsystem, som i større grad vektla individuelle prestasjoner. Fra og med andre kvartal 2004 ble det stilt krav til den

enkeltes prestasjoner både for å oppnå en individuell utbetaling og for å få tilgang på gruppebonus. Reformen bestod av to kampanjer: Q204 og Q304.

Q204 inntraff i uke 13 i 2004. Med utgangspunkt i fire måleparametere – individuelt salg, samlet salg for teamet, kvalitet og fornyelsesprosent av privatkunder – kunne enhver ansatt nå, avhengig av egne og resten av teamets prestasjoner, oppnå en ekstrautbetaling på inntil 12.000,- for kvartalet. Den neste kampanjen ble implementert etter at Q204 var avsluttet, og gjaldt for 3. kvartal 2004. Igjen var det samme fire måleparametrene som dannet bonusgrunnlag, og den maksimale ekstrautbetalingen var fremdeles 12.000,- for hele kampanjeperioden. En forskjell fra Q204 var at det relative forholdet mellom belønning på bakgrunn av individuelle prestasjoner og teamets samlede prestasjoner endret seg. Det ble nå gitt høyere belønning for oppfyllelse av individuelt budsjett, mens belønningen knyttet til teamets oppfyllelse av budsjett gikk tilsvarende ned. For begge kampanjene ble individuelle salgsbudsjett fastsatt før kampanjestart, og teambudsjett ble bestemt som summen av individbudsjettene. Individuelt budsjett var likt for alle ansatte, og var på 1200 solgte enheter i året. Salg ble maskinelt rapportert via et internt system, og det var dermed ingen anledning for de ansatte til å manipulere sine resultater. Kvalitet i utført arbeid ble målt som prosentvis oppfyllelse av et internt måleinstrument, mens kriteriet i forhold til fornyelsesgrad var at denne skulle være prosentvis høyere ved utgangen av kampanjeperioden enn ved inngangen til kampanjeperioden.

Siden både kvalitet og fornyelsesgrad ble målt på team eller senternivå er det salg av forsikringer som (delvis) blir individualisert gjennom reformen. Det er derfor ukentlig salg som er utfallsvariabelen i den økonometriske analysen, og som reformen kan tenkes å gi utslag i fordi de ansatte gis sterkere incentiver til å gjennomføre salg.

Kapittel 4

Metode

Å estimere effekten av forskjellige *behandlinger* på en undergruppe av populasjonen er ofte av interesse for økonomer, men også innenfor andre disipliner som for eksempel psykologi og medisin. Dette kalles i litteraturen ofte for *effektevaluering*, og man ønsker å finne den kausale effekten av behandlingen på et observerbart mål, eller *utfallsvariabel*. Terminologien stammer i stor grad fra medisinsk forskning, men innenfor økonomi kan behandling for et individ være å delta i et arbeidsmarkedsprogram, inngå medlemskap i en fagforening, eller – som blir studert her – få en endring i incentivsystem. Dette kapitlet gir en gjennomgang av metoden som er brukt i denne oppgaven.

4.1 Difference-in-Difference estimatoren

Det finnes flere forskjellige tilnæringsmåter til effektevaluering, og den mest passende metoden bestemmes ofte ut i fra tilgjengelige data.¹¹ Datasettet som danner grunnlaget for den økonometriske analysen kommer fra et såkalt naturlig, eller *quasi*-eksperiment.

Utgangspunktet er at en eksogen hendelse inntreffer, som forandrer det økonomiske miljøet agentene opererer i. Denne hendelsen kalles ofte behandling, tiltak eller reform, og medfører at agentene deles inn i to grupper: *Behandlingsgruppen* som direkte blir utsatt for tiltaket, og *kontrollgruppen* som ikke berøres direkte. At hendelsen som er eksogen tilsier at den er utenfor eksperimentators kontroll, og at individer tilfeldig velges til de to gruppene.¹²

En mye brukt metode i effektevaluering er *difference-in-difference* (DiD)-estimatoren. I tillegg til observasjoner av både kontroll- og behandlingsgruppe forutsetter denne fremgangsmåten også at hele utvalget observeres i minst to tidsperioder, det vil si både før og etter behandlingen inntreffer. Dette er for å kunne kontrollere for systematiske forskjeller mellom kontroll- og behandlingsgruppen. Videre må vi også ha at *trenden* i utfallsvariabelen er den samme for de to gruppene, dvs. at utviklingen i utfallsvariabelen for behandlingsgruppen må være den samme som for kontrollgruppen.¹³ Grunnen til dette er at kontrollgruppen skal representere en kontrafaktisk situasjon: Utviklingen i utfallsvariabelen

¹¹ Se for eksempel Heckman, LaLonde & Smith (1999) for eksempler.

¹² Dette blir i motsetning til kontrollerte eksperimenter som utføres og kontrolleres av en eksperimentator, som også deler inn eksperimentdeltakerne i kontroll- og behandlingsgrupper.

¹³ Alternativt kan man kontrollere for trend, noe som er gjort i denne oppgaven.

for behandlingsgruppen i fravær av behandling. I det følgende presenteres intuisjonen bak DiD-estimatoren.¹⁴

En utfallsvariabel, y_{it} , observeres for hele utvalget og det antas at behandlingen skal ha en effekt på denne. Utvalget består av $i = 1, 2, \dots, N$ individer som observeres for perioder $t = 1, 2, \dots, T$. Vi tar utgangspunkt i følgende fast-effekt modell:

$$(4.1) \quad y_{it} = \alpha_i + \phi_t + \delta G_{it} + \varepsilon_{it}$$

Variabelen ϕ_t er en tidsspesifikk fast effekt, mens α_i er en individspesifikk fast effekt. Av notasjonen ser vi at ϕ_t kun varierer over tid, mens α_i kun varierer over individ. Regressoren G_{it} er en binær variabel som kan ta verdien 1 eller 0 avhengig av om et individ mottar behandling i periode t :

$$G_{it} = \begin{cases} 1 & \text{dersom individ } i \text{ mottar behandling i periode } t \\ 0 & \text{dersom individ } i \text{ ikke mottar behandling i periode } t \end{cases}$$

G_{it} kan kalles en statusvariabel, siden den angir individ i 's status på tidspunkt t . Til slutt er ε_{it} det idiosynkratiske feilleddet i modellen, og vi antar at dette er identisk og uavhengig fordelt med forventning lik 0 og konstant varians lik σ_ε^2 . For enkelhets inneholder ikke (4.1) flere forklaringsvariabler, men modellen kan utvides til å kontrollere for flere forhold.

Ved å ta førstedifferansen (over tid) av ligning (4.1) fjernes den individspesifikke faste effekten:

$$(4.2) \quad y_{it} - y_{i,t-1} = (\alpha_i - \alpha_i) + (\phi_t - \phi_{t-1}) + \delta(G_{it} - G_{i,t-1}) + (\varepsilon_{it} - \varepsilon_{i,t-1}),$$

eller på endringsform:¹⁵

$$(4.3) \quad \Delta y_{it} = \Delta \phi_t + \delta \Delta G_{it} + \Delta \varepsilon_{it}$$

Vi tar utgangspunkt i to perioder, slik at $t = 1, 2$. Behandling inntreffer først i periode 2, slik at for $t = 1$ har vi $G_{i1} = 0 \forall i$, mens for $t = 2$ vil G_{i2} være 1 for en undergruppe av utvalget (behandlingsgruppen). Videre defineres $\Delta \bar{y}^T$ som utvalgsgjennomsnittet til Δy_i for behandlingsgruppen ($G_i = 1$), og $\Delta \bar{y}^{NT}$ som utvalgsgjennomsnittet til Δy_i for kontrollgruppen ($G_i = 0$). OLS-estimatet av δ fra (4.2) eller (4.3) er da gitt som:

¹⁴ Utledningen tar utgangspunkt i Verbeek (2006, s. 363).

¹⁵ Endringsoperatoren Δ indikerer her endring over tid, fra periode $t-1$ til periode t .

$$(4.4) \quad \hat{\delta} = \Delta \bar{y}^T - \Delta \bar{y}^{NT} = (\bar{y}_2^T - \bar{y}_2^{NT}) - (\bar{y}_1^T - \bar{y}_1^{NT})$$

Dette er den såkalte difference-in-difference estimatoren. Navnet kommer av at man først tar tidsdifferansen mellom de to gruppene (behandling og kontroll), og deretter tar differansen mellom de «behandlede» og «ikke-behandlede».¹⁶ Under visse forutsetninger er $\hat{\delta}$ et konsistent estimat for δ .¹⁷

Med regresjon kan DiD-estimatoren finnes med utgangspunkt i følgende lineære sammenheng:¹⁸

$$(4.5) \quad y_{it} = \alpha_i + \beta_1 G_i + \beta_2 T_t + \delta G_i * T_t + \varepsilon_{it}$$

Her er α_i fremdeles den tidsinvariante individsspesifikke faste effekten, G_i er en binær variabel som angir gruppetilhørighet, mens T_t er en tidsdummy som endrer verdi fra 0 til 1 på tidspunktet hvor behandlingen inntreffer. ε_{it} er det idiosynkratiske feilleddet, og vi antar fremdeles at $\varepsilon_{it} \sim i. i. d(0, \sigma_\varepsilon^2)$. Fra (4.5) kan DiD-estimatoren, $\hat{\delta}$, finnes ved å estimere koeffisienten til interaksjonsleddet $G_i * T_t$. Av egenskapene til G_i og T_t følger det at produktet $G_i * T_t$ kun vil anta verdien 1 for individer i behandlingsgruppen etter at behandlingen har inntruffet. Det kan vises at $plim \hat{\delta} = (\bar{y}_2^T - \bar{y}_2^{NT}) - (\bar{y}_1^T - \bar{y}_1^{NT})$, og vi ser at høyresiden av dette uttrykket er det samme som i ligning (4.4). figur 1 illustrerer intuisjonen bak DiD-estimatoren fra (4.5).

	Før Behandling	Etter Behandling	Differanse
Kontrollgruppe	α_i	$\alpha_i + \beta_2$	β_2
Behandlingsgruppe	$\alpha_i + \beta_1$	$\alpha_i + \beta_1 + \beta_2 + \delta$	$\beta_2 + \delta$
Differanse	β_1	$\beta_1 + \delta$	δ

Figur 1: Difference-in-difference estimatoren

¹⁶ Estimatoren $\hat{\delta}$ kalles også den gjennomsnittlige behandlingseffekten (*average treatment effect*), siden den estimerer effekten av behandling på gjennomsnittsutfallet for utfallsvariabelen.

¹⁷ Så lenge $E\{\Delta G_{it} \Delta \varepsilon_{it}\} = 0$, dvs. at endringen i status er ukorreletert med endringen i de uobserverbare faktorene i feilleddet.

¹⁸ Også (3.5) kan på samme måte som (3.1) utvides til å inneholde flere regressorer, noe som er gjort i den faktiske modellen som brukes for å estimere tiltakseffekten.

Av figur 1 ser vi at «differansen i differansen», eller «differansen mellom tidsdifferansene for de to undergruppene av utvalget» er gitt ved parameteren δ . Diagrammet kan også brukes for å vise hvorfor observasjoner av både kontroll- og behandlingsgruppe pre og post behandling er viktig for å identifisere effekten av behandlingen. Vi antar at $E[\hat{\delta}] = \delta$. Dersom man ikke har tilgang på en kontrollgruppe og kun ser på forskjellen før og etter behandling for behandlingsgruppen, vil (den estimerte) differansen i utfallsvariabelen være gitt ved $\hat{\beta}_2 + \hat{\delta}$. Dersom dette skal være den sanne effekten av behandling må vi ha at $E[\hat{\beta}_2] = \beta_2 = 0$, noe som betyr at det ikke eksisterer noen tidstrend i utfallsvariabelen. Dersom man på en annen side har en kontrollgruppe men mangler observasjoner av gruppene før behandling, vil differansen mellom de to gruppene post-behandling være gitt ved $\hat{\beta}_1 + \hat{\delta}$. Dette betyr at $E[\hat{\beta}_1] = \beta_1 = 0$ for at denne differansen skal gi den sanne effekten av behandlingen. At $\beta_1 = 0$ betyr at de to gruppene må være identiske, slik at det ikke finnes noen permanente forskjeller i utfallsvariabelen for de kontroll- og behandlingsgruppen.¹⁹

4.2 Estimeringsmodeller

For å finne et estimat for δ fra (4.5) finnes det flere forskjellige fremgangsmåter.²⁰ En mulighet er å gjennomføre en standard OLS-regresjon. Dersom alle forklaringsvariabler er ukorrelert med heterogeniteten og «samtidig» (*contemporaneously*) ukorrelert med det idiosynkratiske feilleddet er OLS-estimatoren forventningsrett, og konsistent for $N \rightarrow \infty$ og $T \rightarrow \infty$. OLS-estimatoren utnytter begge dimensjonene av datasettet (*within* og *between*), men ikke effisient så lenge de individuelle feilleddene er korrelert over tid. En estimator som i så tilfelle utnytter begge dimensjoner av datasettet på en effisient måte er *tilfeldig-effekt* (RE) estimatoren. Dette er en såkalt *feasible general least squares* estimator, og kan bestemmes som et OLS-estimat fra en transformert modell. Dersom alle forklaringsvariabler er strengt eksogene og ukorrelert med den individspesifikke heterogeniteten er RE-estimatoren konsistent for $N \rightarrow \infty$ og $T \rightarrow \infty$.

Et forhold som ofte kan tenkes å skape problemer i mikroøkonometrisk analyse er *individuell heterogenitet*, det vil si at de økonomiske agentene har individuelle egenskaper som ikke lar seg observere, eller i det minste ikke tallfeste. Denne heterogeniteten skaper problemer

¹⁹ I følge Albouy (2004) er dette er dette ikke nødvendigvis en feil antakelse i et kontrollert eksperiment: Dersom kontroll- og behandlingsgruppe tilfeldig velges kan det være grunn til å tro at de er tilnærmet identiske.

²⁰ *Estimeringsmodeller* blir gjennomgående brukt for å beskrive de forskjellige modellene som brukes for å identifisere effekter (f. eks fast-effekt modellen), mens *regresjonsmodell* brukes om regresjonslikningen

dersom disse uobserverbare egenskapene påvirker utfallsvariabelen, og i tillegg er korrelert med én eller flere av de observerte regressorene. Hverken OLS- eller RE-estimatoren håndterer dette problemet direkte, og forutsetter at heterogeniteten er ukorrelert med alle forklaringsvariabler for forventningsrette og konsistente estimater. Et paneldatasett gir imidlertid anledning til å håndtere problemet med individuell heterogenitet med mindre restriktive antakelser. I ligning (4.5) inngår heterogeniteten, α_i , som en individspesifikk, tidskonstant faktor, og (4.5) er dermed et eksempel på en såkalt *fast-effekt* (FE) modell. I FE-rammeverket løser man heterogenitetsproblemet ved å foreta en transformasjon av (4.5) som eliminerer alle tidskonstante elementer, inkludert heterogeniteten, før koeffisientene estimeres med OLS. FE-estimatoren, også kalt *within*-estimatoren, utnytter informasjon om forskjeller innen grupper, dvs. individuelle forskjeller over tid, men ikke mellom individ. Selv om den kun utnytter en dimensjon av datasettet kan FE-estimatoren være å foretrekke fordi den er mer robust i tilfeller med tidskonstant heterogenitet.²¹

²¹ Se for eksempel Verbeek (2008, kap. 10) for en grundigere gjennomgang av de forskjellige estimeringsmodellene.

Kapittel 5

Data

5.1 Generelt

Datasettet som danner grunnlag for den økonometriske analysen er et panel som inneholder ukentlige observasjoner på individnivå. Individene er ansatte ved kundesenteret til et norsk forsikringsselskap, og i utgangspunktet er det 261 individer som opptrer innenfor observasjonsperioden. Datasettet er ikke balansert, noe som betyr at det er manglende observasjoner for et eller flere individer i én eller flere tidsperioder (her: uker). Dette er imidlertid ikke et problem så lenge det kan antas at de manglende observasjonene skyldes tilfeldige fremfor systematiske årsaker.²²

Selv om datasettet inneholder observasjoner fra og med uke 1, 2003 og helt frem til uke 52, 2010, er tidsrommet som danner estimeringsgrunnlag begrenset til 40 uker før og etter endringen innføres. Dette er fordi det i tidsrommet fra 2003 til 2010 inntreffer flere forhold som kan forstyrre resultatene. Avgjørelsen tas på bakgrunn av en avveining mellom hvor forstyrrende disse forholdene er for resultatene våre, og hvilke konsekvenser en reduisering av observasjoner har for utvalgsstørrelsen. Siden det vanskelig lar seg gjøre å isolere effekten av reformen dersom det forekommer et høyt antall eksogene sjokk, blir valget å begrense lengden på panelet.

Reformen inntreffer i uke 13 i 2003, og var som nevnt i kapittel 3 egentlig to endringer i bedriftens bonussystem. Fordi begge endringene inntreffer innenfor et relativt kort tidsrom, blir estimeringsgrunnlaget for lavt til å isolere effekten av Q204 gjennom ytterligere interaksjonsledd. Å snevre inn tidsrommet før og etter Q204 gir ikke signifikante resultater, mest sannsynlig fordi utvalget blir for lavt til å identifisere en effekt.²³ For Q304 er lengden på perioden etter at kampanjen er gjennomført begrenset av en ytterligere endring i 2005.

5.2 Variabler

Utfallsvariabelen i analysen er *sumsalg*, som angir det ukentlige salget av forsikringer for enhver ansatt. Dette er fordi det som nevnt i avsnitt 3.2 er salg av forsikringer som

²² Cameron og Trivedi (2009, s. 230).

²³ Både ytterligere interaksjonsledd og kortere tidsrom ble forsøkt i Stata.

individualiseres gjennom reformen. Figur 2 viser gjennomsnittlig salg gjennom observasjonsperioden.

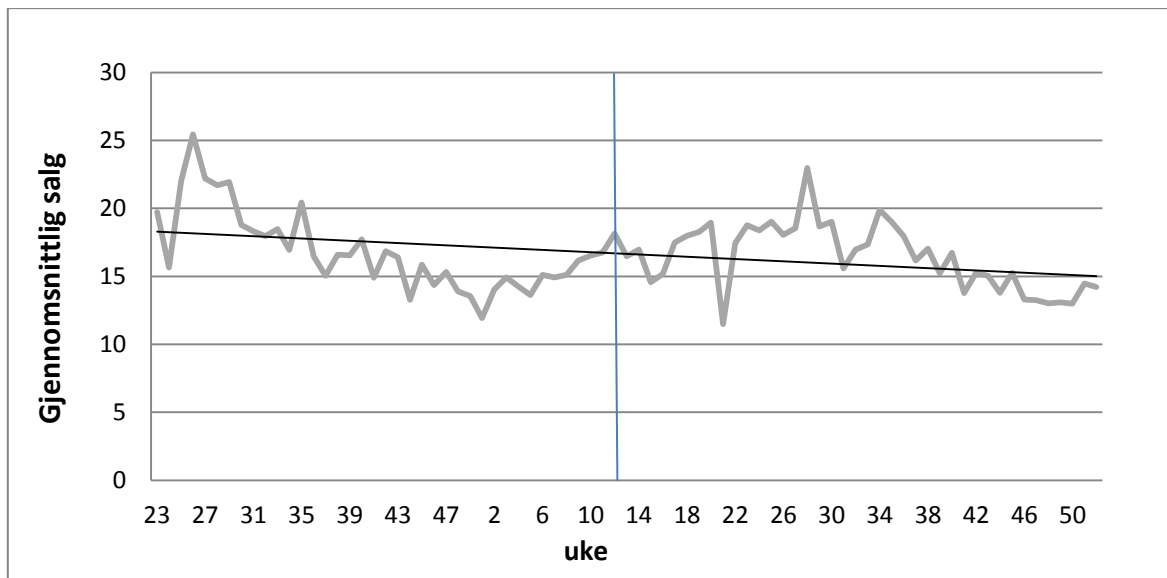


Fig. 2: Utviklingen i gjennomsnittlig salg gjennom observasjonsperioden

Trendlinjen i figur 2 indikerer at det generelt var en nedadgående trend i salget gjennom hele observasjonsperioden. Grafen er imidlertid ikke korrigert for mulige sesongvariasjoner i trafikken inn til kundesenteret eller den generelle etterspørselen etter forsikringer, og heller ikke for antallet telefoner som besvares i det samme tidsrommet. Derfor sier denne figuren i seg selv lite om effekten av reformen. Disse forholdene kontrolleres det for i regresjonsanalysen. At trendlinjen i figuren indikerer en vedvarende nedgang i salg også etter at Q204 er implementert betyr dermed ikke nødvendigvis at reformen ikke hadde en effekt eller at denne var negativ.

Variabelen *treatment* er en binær indikatorvariabel som tar verdien 1 dersom individet er i behandlingsgruppen, og 0 dersom individet er i kontrollgruppen. Videre er også variabelen *reform* binær. Denne tar verdien 0 i alle uker frem til reformen inntreffer, hvorpå den tar verdien 1 for alle etterfølgende uker. Interaksjonsleddet mellom *treatment* og *reform*, *treat_ref*, er fokusvariabel i analysen. Dette er fordi koeffisientestimatet til *treat_ref* gir DiD-estimatoren [se ligning (4.5)]. Matematisk er denne variabelen gitt ved $treatment * reform$, og vil dermed per konstruksjon kun ta verdien 1 for individer i behandlingsgruppen etter at Q204 er innført.

Et forhold det må kontrolleres for er den faktiske arbeidstiden til de ansatte. Siden henvendelser fra kunder er en forutsetning for å gjennomføre et salg, kan forskjeller i salg både mellom individer, grupper og over tid skyldes at det er variasjon i antallet henvendelser den enkelte besvarer. Reformen kan også tenkes å gi incentiver til å besvare flere henvendelser. Derfor inkluderes variabler for å kontrollere for antall henvendelser en ansatt behandler, eller antall minutter han eller hun er pålogget kundesenterets interne system. Dette gjør at positive verdier av DiD-estimatoren tolkes som endringen i salg som følger av at de ansatte responderer på reformen ved å yte en høyere innsats, gitt at de besvarer en ytterligere henvendelse eller gitt et ekstra minutt pålogget tid. Variabler som kan brukes for å kontrollere for henvendelser eller pålogget tid er *telefoner*, *totalt1* og *loggtid*. *Telefoner* angir totalt antall samtaler inn eller ut fra en ansatts apparat²⁴ per uke, mens *totalt1* kun inkluderer henvendelser *inn* til agenten (inkludert *callback*, som behandles som innkommende samtaler fordi det forutsetter at kunden som ringes opp først har henvendt seg til kundesenteret). *Loggtid* indikerer antall påloggede minutter per uke. Disse variablene blir samlet referert til som FT-variabler, og vil på grunn av at de er høyt korrelerte ikke inkluderes i modellen samtidig på grunn av multikollinearitet.²⁵ Variablene kan dermed anses for å være gjensidig utelukkende.

Det kan også tenkes at det er andre eksogene forhold som gjør at antallet henvendelser inn til kundesenteret vil variere fra uke til uke innenfor en gitt periode. Kontrollgruppen består av deltidsansatte og tilkallingsvikarer, og det kan tenkes at disse oftere er tilstede på arbeidsplassen i uker med stor pågang. I tillegg vil det kunne antas at potensialet for å få gjennomført et salg også er høyere enn normalt i uker med et relativt høyt antall henvendelser. Til sammen vil dette kunne medføre at salgstrenden innenfor en gitt periode ikke er lik for de to gruppene. Regresjonsmodellene inneholder derfor også et sett med ukedummier, samlet referert til som *\$wdummy*, for å kontrollere for dette. Dette er for å sikre at den indentifiserende antakelsen for DiD-estimatoren om at (utviklingen i) tidstrend for kontroll- og behandlingsgruppen ikke brytes på grunn av ukentlige variasjoner. Tabell 1 og 2 viser deskriptiv statistikk for de kontinuerlige variablene.

²⁴ Med apparat menes den enkelte ansatte fysiske telefon, og samtalene er dermed ikke nødvendigvis via kundesentersentralen.

²⁵ En enkel OLS-regresjon av *telefoner* på *totalt1* gir $R^2 = 0.9311$. Generelt vil $\text{Var}(\hat{\beta}_j) \rightarrow \infty$ når $R^2 \rightarrow 1$. Høy, men ikke perfekt, korrelasjon mellom to eller flere bakgrunnsvariabler kalles *multikollinearitet*, og vil på samme måte som lav utvalgsstørrelse kunne gi høy varians i koeffisientestimatene (Wooldridge 2006, s. 102).

Tabell 1: Deskriptiv statistikk for de kontinuerlige variablene, hele utvalget.

<u>Hele Observasjonsperioden</u>			
Variabel	Beskrivelse	Gj.snitt	Std.avvik
<i>Sumsalg</i>	Angir ukentlig salg for den enkelte agent, målt i enheter	16.4939	11.1547
<i>Loggtid</i>	Angir ukentlig pålogget tid for den enkelte agent, målt i minutter	1408.635	592.6829
<i>Telefoner</i>	Angir totalt antall telefoner, innkommende og utgående, til den enkelte agents apparat	138.2194	69.195
<i>Totaltl</i>	Angir alle innkommende samtaler for den enkelte agent	109.2459	57.9712
<u>Før Reform</u>			
<i>Sumsalg</i>		16.7009	11.0988
<i>Loggtid</i>		1430.382	586.7595
<i>Telefoner</i>		148.1292	72.1870
<i>Totaltl</i>		117.7908	61.1957
<u>Etter Reform</u>			
<i>Sumsalg</i>		16.29667	11.2052
<i>Loggtid</i>		1388.687	597.4225
<i>Telefoner</i>		129.1293	65.0235
<i>Totaltl</i>		101.4077	53.6690

Merk: uke 1, 51 og 52 slettes fra observasjonsperioden på grunn av at ukene rundt jul er lite representative for resten av året.

Tabell 2: Deskriptiv statistikk for de kontinuerlige variablene for kontroll- og behandlingsgruppen separat.

<u>Behandlingsgruppen</u>			
	Variabel	Gj.snitt	Std. avvik
Før reform	<i>Sumsalg</i>	17.42533	11.1101
	<i>Loggtid</i>	1460.834	560.0305
	<i>Telefoner</i>	151.8217	70.2832
	<i>Totalt1</i>	121.177	60.2302
Etter reform	<i>Sumsalg</i>	17.255	11.344
	<i>Loggtid</i>	1417.014	551.3969
	<i>Telefoner</i>	131.8863	61.9355
	<i>Totalt1</i>	104.2225	52.0903
<u>Kontrollgruppen</u>			
	Variabel	Gj.snitt	Std. avvik
Før reform	<i>Sumsalg</i>	10.0327	8.4696
	<i>Loggtid</i>	1163.757	731.0478
	<i>Telefoner</i>	115.8	80.2067
	<i>Totalt1</i>	88.1437	61.6379
Etter reform	<i>Sumsalg</i>	10.3945	8.1112
	<i>Loggtid</i>	1208.769	810.0013
	<i>Telefoner</i>	111.6174	79.8210
	<i>Totalt1</i>	83.5289	59.7952

Tabellene viser at både gjennomsnittlig salg og gjennomsnittlig antall besvarte henvendelser gikk ned i reformperioden, og at dette gjelder for begge grupper. Det er også relativt stor spredning i observasjonene rundt utvalgsgjennomsnittene for begge gruppene. En måte å studere endringen i produktivitet på ut i fra disse tallene, kan være å se på det relative forholdet mellom antall besvarte henvendelser og salg. Gjennomsnittlig antall telefoner delt på gjennomsnittlig antall salg kan tolkes som hvor mange henvendelser som kreves for at et salg skal gjennomføres, og gir dermed en indikasjon på hvor effektive de ansatte var før og etter reformen ble implementert.

Tabell 3: *Hvor mange henvendelser som kreves for å gjennomføre et salg*

<u>Behandlingsgruppen</u>			
	Før reform	Etter reform	Differanse
<i>(Telefoner/Sumalg)</i>	8.713	7.643	1.07
<u>Kontrollgruppen</u>			
	Før reform	Etter reform	Differanse
<i>(Telefoner/Sumalg)</i>	11.542	10.738	0.804

Tabell 3 viser endringen for både kontroll- og behandlingsgruppen. Begge grupper har en positiv endring i produktivitet etter reformen, men økningen er større for behandlingsgruppen enn for kontrollgruppen. I prosent var økningen for behandlingsgruppen 12.28 %, mens den var 6.97 % for kontrollgruppen. Dette gir dermed en indikasjon på at reformen stimulerte til høyere innsats i behandlingsgruppen. Det er likevel for tidlig å konkludere med noe på bakgrunn av dette, siden det ikke kontrolleres for tidstrend her.

Kapittel 6

Regresjonsanalyse

I dette kapitlet gjennomgås resultatene fra den økonometriske analysen, og det blir også gitt begrunnelse for de forskjellige valgene som er tatt for spesifisering av modellene og endring av datasettet underveis. I avsnitt 6.1 presenteres resultater fra forskjellige estimeringsmodeller på utvalget før det gjennomføres noen restriksjoner eller endringer. Utvalget som brukes her refereres til som *standardutvalget*. I avsnitt 6.2 vurderes det hvilken estimeringsmodell som anses for å være mest passende for dette panelet, og hvilken FT-variabel som er mest egnet for å kontrollere for innsats. Deretter følger forskjellige endringer av datasettet, for å se om resultatene fra standardutvalget endrer seg. I avsnitt 6.6 blir inferensgrunnlaget vurdert, mens avsnitt 6.7 og 6.8 ser på om det er forskjeller i hvordan menn og kvinner eller høy- og lavproduktive team og individer responderte på reformen.

6.1 Resultater fra standardutvalget

Tabell 4 gjengir resultater fra FE- og RE- modellene samt OLS-estimerer fra sammenslått panel (*pooled OLS regression*), fra en regresjon av *sumsalg* på forklaringsvariablene presentert i forrige kapittel. Resten av koeffisientestimatene fra disse regresjonene er gjengitt i tabell A.1 (se app. A).²⁶ Resultatene gir et utgangspunkt for å se på forskjeller som følger av hvilken FT-variabel som inkluderes, og også for å vurdere hvilken estimeringsmodell som er best egnet.

Resultatene tilsier en positiv effekt av reformen siden DiD-estimatoren har et positivt fortegn i alle tilfeller, med unntak av RE-modellen med *loggtid* inkludert som FT-variabel. Det er imidlertid variasjon i både verdien til koeffisientestimatene og den statistiske forklaringskraften til forklaringsvariablene. På bakgrunn av den kalkulerede p-verdien (for en nullhypotese om at koeffisientestimatet er lik 0) gir FE-modellen med *telefoner* eller *totalt1* som FT-variabel resultater som er statistisk signifikante på et akseptabelt nivå.²⁷ I forhold til *goodnes-of-fit* målet, R^2 , tilsier verdien av disse at modellen har høyest forklaringskraft dersom *totalt1* inkluderes som FT-variabel. Selv om R^2 er uegnet for å vurdere forskjellige

²⁶ Koeffisientestimatene til *\$wdummy* utelates av praktiske årsaker fra tabell A1, men er gjengitt for FE-regresjon av standardutvalget med *totalt1* som FT-variabel i tabell A.7 (se app. A).

²⁷ I tolkningen av resultatene vil en p-verdi lavere enn 0.1 generelt anses for å være et tilfredsstillende.

estimeringsmodeller mot hverandre, kan det brukes for å vurdere forskjellige spesifiseringer av samme modell.

Tabell 4: DiD-estimatoren fra forskjellige estimeringsmodeller og spesifikasjoner av regresjonsmodellen på standardutvalget.

Modell	DiD-estimator	FT-variabel	R^2		overall
			<i>within</i>	<i>between</i>	
FE	1.089 (0.977)	<i>Loggtid</i>	.2112	.4108	.2588
FE	1.686** (0.834)	<i>Telefoner</i>	.2412	.6058	.3530
FE	1.583* (0.857)	<i>Totalt1</i>	.2516	.6439	.3762
RE	-0.755 (0.624)	<i>Loggtid</i>	.2227	.3651	.2727
RE	0.0563 (0.491)	<i>Telefoner</i>	.2497	.5511	.3637
RE	0.132 (0.491)	<i>Totalt1</i>	.2585	.5949	.3857
OLS	0.687 (0.492)	<i>Loggtid</i>			.2811
OLS	1.252*** (0.460)	<i>Telefoner</i>			.3708
OLS	1.195*** (0.443)	<i>Totalt1</i>			.3912

Avhengig variabel: *sumsalg*

N:	261
T:	80
Totalt antall observasjoner:	9602

Merk: Cluster-robuste standardfeil i parenteser under koeffisientestimerer,

asterisker angir signifikans: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Siden panelet er ubalansert tilsvarer ikke det totale antallet observasjoner $N \cdot T$

Regresjonsmodellene inneholder ukedummier for å kontrollere for tidstrend.

Uavhengig av FT-variabel tilsier også resultatene fra FE-modellen at det generelt var en nedgang i salget etter at reformen ble innført (se tabell A.1). Dette betyr at det var en nedadgående trend i salget etter at reformen ble innført, men at effekten av at reformen ble innført dempet for noe av denne nedgangen. DiD-estimatoren tolkes som effekten av reformen uavhengig av koeffisientestimatet til *reform*, mens selve endringen i salg etter at reformen er inntruffet er gitt ved summen av de to koeffisientestimatene.

Resultatene fra RE-modellen tilsier en lavere effekt av reformen enn resultatene fra FE-modellen, og dersom *loggtid* benyttes som FT-variabel har faktisk DiD-estimatoren annet fortegn enn de andre DiD-estimatorene. Ingen av disse resultatene er imidlertid signifikante, og i så måte indikerer resultatene fra RE-modellen at reformen ikke hadde noen effekt på salget. Et forhold som kan forstyrre resultatene fra RE-modellen er at den ikke er robust for uobserverbar heterogenitet. Det er heller ikke OLS-estimatoren, og som nevnt i avsnitt 4.2 er OLS-estimatoren i utgangspunktet heller ikke effisient i tilfeller med panel. Neste avsnitt ser derfor på noen tester for å vurdere hvilken av estimeringsmodellene som er mest egnet for dette datasettet.

6.2 Valg av estimeringsmodell og spesifisering av regresjonsmodell

Et første spørsmål er om RE- og FE-modellene alltid vil være mer egnet enn standard OLS i møte med paneldata, og dermed om sistnevnte kan utelukkes *a priori*. Et forhold som gjør at OLS-estimatoren kan være å foretrekke er dersom det *ikke* er noen heterogenitet (eller *paneleffekt*) i datasettet. Hvis regresjonsmodellen er så velspesifisert at de inkluderte regressorene fanger opp alle uobserverbare faktorer kan man velge å gjennomføre OLS-regresjon på et sammenslått panel. Dougherty (2007) nevner to fordeler med dette: For det første er det en gevinst i forhold til effisiens, fordi man ikke trenger å tillate (ikke-eksisterende) seriekorrelasjon innenfor grupper. Videre vil man også kunne utnytte «endelig-utvalg» (*finite-sample*) egenskapene til OLS-estimatoren fremfor å stole på de asymptotiske egenskapene til RE-estimatoren.

Breusch & Pagan (1980) har foreslått en test for tilfeldige effekter, og under nullhypotesen er det ingen paneleffekt (det er med andre ord ingen signifikante forskjeller mellom individer). Dersom nullhypotesen ikke kan forkastes kan OLS på sammenslått panel velges på bakgrunn av argumentene over. Resultatene fra denne testen er gjengitt i app. B. Nullhypotesen *kan* forkastes, og på grunnlag av dette fremstår RE-estimatoren som et bedre valg enn standard OLS.

Siden det også er forskjeller i estimatene fra FE- og RE-modellene blir neste spørsmål hvilken av disse modellene som er best egnet for dette datasettet. Hausman (1978) har foreslått en test for en nullhypotese om at heterogeniteten er ukorrelert med forklaringsvariablene. Tanken bak den såkalte Hausman-testen er å sammenligne to estimatorene, hvorav den ene er konsistent under både nullhypotesen og alternativhypotesen, mens den andre er konsistent og typisk også effisient) kun under nullhypotesen. Essensielt tester Hausman-testen for om det er en signifikant forskjell mellom to estimatorene, og avgjør på bakgrunn av dette om nullhypotesen kan forkastes. Et potensielt problem med denne testen er at den ikke er kompatibel med cluster-robuste standardfeil, som til nå har blitt brukt som inferensgrunnlag på grunnlag av mistanke om seriekorrelasjon i de individuelle feilleddene.²⁸ Dette blir sett nærmere på i avsnitt 5.6. Et robust alternativ til Hausman-testen er å teste for overidentifiserende restriksjoner i den ene av modellene, her RE-modellen, som bygger på strengere antagelser for konsistente estimater enn FE-modellen.²⁹ Resultatene fra begge testene er gjengitt i app. B, og begge taler i favør av FE-modellen. Videre er heller ikke fokusvariabelen, *treat_ref*, tidskonstant. På bakgrunn av dette vil FE-modellen brukes som estimeringsmodell i resten av analysen.

I forhold til spesifisering av modellen fremstår *totalt1* som det beste valget for FT-variabel. For det første tilsier R^2 tilsier at modellen har høyest forklaringskraft dersom *totalt1* inkluderes i regresjonsmodellen. I tillegg har denne variabelen relativt lik tolkning som *telefoner*, fordi den største andelen av telefoner til eller fra en agent fanges opp av *totalt1*. Det er likevel en fordel at *totalt1* kan korrigere for samtaler som ikke er arbeidsrelaterte, og som dermed i stor grad ikke kan generere salg. I forhold til *loggtid* er heller ikke denne intuitivt like egnet som *totalt1*, siden antall telefoner uansett indikerer at agenten er pålogget systemet. Det er gjennom samtaler med kunder at salg kan gjennomføres, og ikke på bakgrunn av å være pålogget systemet i seg selv. Derfor velges også *loggtid* bort til fordel for *totalt1*, og resten av analysen vil se på resultater når regresjonsmodellen estimeres med FE-modellen og *totalt1* inkluderes som FT-variabel. App. A gjengir likevel resultatene når *loggtid* og *telefoner* inkluderes i regresjonsmodellen for sammenligning.

²⁸ Både den rapporterte rho-verdien (.305) fra FE-regresjon av standardutvalget og et scatterplot av residualene mot residualene med et lag tilsier seriekorrelasjon (se fig. C.3 i app. C).

²⁹ I Stata implementeres testen ved hjelp av kommandoen *xtoverid*. Denne er et bidrag fra Scaffer og Stillman (2010), og bygger på resultater fra blant annet Arellano (1993).

Med dette som utgangspunkt tilsier resultatene så langt at reformen hadde en positiv effekt på innsatsen de ansatte la ned i enhver samtale for å gjennomføre et salg, selv om salget generelt fikk en nedgang i perioden etter reformen ble innført. Den estimerte effekten er på 1.583 salg per uke, og estimatet er statistisk signifikant. Dette kan igjen tolkes som at innføring av individuell bonus medførte en sterkere incentiveeffekt enn ordningen med gruppebonus.

I de følgende avsnittene gjennomføres forskjellige restriksjoner og endringer av datasettet for å sjekke robustheten til resultatet så langt, og for å studere andre forhold som kan være av interesse.

6.3 Restriksjoner på utvalget

På bakgrunn av at forstyrrende observasjoner kan tenkes å påvirke resultatene, kan det være hensiktsmessig å foreta visse begrensninger av utvalget.³⁰ En løsning er å fjerne observasjoner av individer i behandlingsgruppen som er pålogget kundesenterets internsystemer mindre enn 5000 minutter, eller har gjennomsnittlig ukentlig salg lavere enn 4, i en av periodene.

Modellen kan også testes ytterligere ved å stille enda strengere krav til behandlingsgruppen, men dette blir en avveining mellom hvilke observasjoner som faktisk er forstyrrende i forhold til å finne den reelle effekten av behandlingen, og hva som kan bli ansett for å være manipulering av datasettet. Et annet forhold og ta hensyn til er at utvalget åpenbart blir mindre dersom observasjoner slettes.

Resultatene etter at utvalget er begrenset som beskrevet over er gjengitt i tabell 5.

Begrensningen av utvalget medfører kun kvantitative endringer for koeffisientestimatene, og resultatene kvalitativt er som før. Ytterligere restriksjoner på utvalget gir tilsvarende resultater. DiD-estimatoren tilsier nå en ceteris-paribus effekt av reformen på 1.506, og dette er signifikant på et 10 % nivå. Til sammenligning hadde DiD-estimatoren en verdi på 1.583 når den ble estimert på standardutvalget. At det ikke er særlig stor differanse i absoluttverdien til estimatene kan forklares ved at ikke mange observasjoner ble fjernet, slik at estimeringsgrunnlaget ikke forandret seg mye. Gitt at observasjonene som ble droppet av begrensningen kun var forstyrrende forhold, kan koeffisientestimatet 1.506 anses for å være et mer nøyaktig mål på effekten av reformen. Problemet kan være at observasjoner som skulle vært med også fjernes. En løsning kunne vært å slette de aktuelle individobservasjonene fra

³⁰ Blant annet har det vært situasjoner for overordnede ved kundesenteret har besvart henvendelser i spesielt hektiske perioder, som har kunnet medføre salg. Siden disse personene ikke passer inn hverken i kontrollgruppen eller behandlingsgruppen, kan de anses for å være forstyrrende observasjoner.

utvalget, men dette er ikke en mulighet.³¹ Det kan derfor ikke konkluderes med at 1.506 er et mer korrekt mål på effekten en 1.583.

Tabell 5: DiD-estimatoren når det legges restriksjoner på utvalget

Modell	DiD-estimator	FT-variabel	R^2			Avhengig Variabel
			<i>within</i>	<i>between</i>	<i>overall</i>	
FE	1.506* (0.879)	<i>TotaltI</i>	.2520	.6671	.3744	<i>Sumsalg</i>
N:		221				
T:		80				
Totalt antall observasjoner:		9446				

*Merk: Cluster-robuste standardfeil i parenteser under koeffisientestimer, asterisker angir signifikans: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$*

Siden panelet er ubalansert tilsvarer ikke det totale antallet observasjoner $N \cdot T$

Regresjonsmodellene inneholder ukedummier for å kontrollere for tidstrend.

6.4 Log-level modell

En mye brukt metode er å la den avhengige variabelen være den naturlige logaritmen til utfallsvariabelen. Dersom de forklaringsvariablene i modellen ikke endres på gir dette en såkalt log-level modell. Det finnes ulike grunner til å bruke slike modeller, og Wooldridge (2006) nevner flere fordeler: For det første kan man dempe – om ikke fjerne – problemer med strengt positive variabler som har betinget fordeling som er heteroskedastisk eller skjev dersom man tar logaritmen av variabelen. Videre vil man også vanligvis kunne innsnevre «området» (*range*) til en variabel slik at estimatene blir mindre sensitive for observasjoner med ekstremverdier. Den estimerte tetthetsfunksjonen til *sumsalg* og logaritmen til *sumsalg* er gjengitt i app. C.

Resultatene når avhengig variabel er $\log(\text{sumsalg}) := \log(\text{sumsalg})$ er gjengitt i tabell 6. Heller ikke denne transformasjonen av modellen endrer kvalitativt på den estimerte effekten av reformen for behandlingsgruppen. I en log-level modell får koeffisientestimatene en annen tolkning enn i en level-level modell: koeffisientestimatene fra en log-level modell skalert med 100 kalles ofte *semi-elastisiteten* til den avhengige variabelen med hensyn på den respektive forklaringsvariabelen, og det kan vises at $\% \Delta y \approx (100\beta_k)x_k$ for hver av de k koeffisientestimatene i modellen. Gitt denne forståelsen av koeffisientestimatene tilsier

³¹ Fordi datasettet er anonymisert kan ikke identitetsindikatorerne knyttes til enkelte ansatte.

resultatene at *ceteris paribus* effekten av reformen på ukentlig salg var en omtrentlig økning på 17 %, og koeffisientestimatet er signifikant forskjellig fra 0.³²

Tabell 6: *DiD-estimatoren når avhengig variabel er Logsumsalg*

Modell	DiD-estimator	FT-variabel	<i>within</i>	R^2		Avhengig Variabel
				<i>between</i>	<i>overall</i>	
FE	0.171** (0.0783)	<i>Totalt1</i>	.2843	.6155	.4061	<i>Logsumsalg</i>
N:		261				
T:		80				
Totalt antall observasjoner:		9602				

Merk: Cluster-robuste standardfeil i parenteser under koeffisientestimater,

*asterisker angir signifikans: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$*

Siden panelet er ubalansert tilsvarer ikke det totale antallet observasjoner $N \cdot T$

Regresjonsmodellene inneholder ukesdummier for å kontrollere for tidstrend.

6.5 Seleksjonseffekt

Innledningsvis i kapittel 5 ble det nevnt at datasettet ikke er balansert. En grunn til dette er at det er individer som slutter og begynner ved kundesenteret i løpet av observasjonsperioden. En effekt av bonusordninger som ikke har blitt omtalt tidligere i oppgaven er såkalt *seleksjonseffekt*. Dette betyr at bedriften tiltrekker seg mer produktive kandidater gjennom å tilby gunstige kompensasjonsformer, mens ansatte som er mindre produktive slutter fordi lønnen blir for lav. Dersom dette er tilfellet her vil den *motivasjonseffekten* som ønskes identifisert kunne forveksles med en slik seleksjonseffekt.

En måte å sjekke dette på er å droppe observasjoner for alle individer i behandlingsgruppen som ikke opptrer i begge perioder. Dersom DiD-estimatoren endrer seg mye i verdi kan det være en indikasjon på at deler av den estimerte effekten av reformen skyldes en seleksjonseffekt. Tabell 7 gjengir resultater både fra standardutvalget og utvalget med restriksjoner etter at individer i behandlingsgruppen som ikke observeres både før og etter reformen er implementert slettes.

³² Gibbons (2009) påpeker et viktig forhold: Dersom den forventede trenden i *logsumsalg* er konstant, kan ikke den forventede trenden i *sumsalg* være konstant. Dette betyr at en av modellene bryter med den identifiserende antakelsen om at trend er konstant over tid og individer. Siden begge modellene kontrollerer for ukesvariasjoner kan dette tenkes å være korrigert for, men det må likevel tas hensyn til når resultatene tolkes.

DiD-estimatoren er i begge tilfeller tilnærmet identisk med resultatene gjengitt i tabell 2 og III, og forskjellen i absoluttverdi er henholdsvis $|0.006|$ og $|0,005|$. Dette gir en god indikasjon på at den estimerte effekten ikke kan tilskrives en seleksjonsprosess.

Tabell 7: DiD-estimatoren når individer som ikke opptrer i begge perioder fjernes fra utvalget

<u>Standardutvalget etter at observasjoner er droppet</u>						
Modell	DiD-estimator	FT-variabel	<i>within</i>	R^2 <i>between</i>	<i>overall</i>	Avhengig Variabel
FE	1.589* (0.853)	<i>Totalt1</i>	.2532	.6844	.3770	<i>Sumsalg</i>
N:		191				
T:		80				
Totalt antall observasjoner:		8868				

<u>Utvalg med restriksjoner etter at observasjoner er droppet</u>						
Modell	DiD-estimator	FT-variabel	<i>within</i>	R^2 <i>between</i>	<i>overall</i>	Avhengig Variabel
FE	1.511* (0.875)	<i>Totalt1</i>	.2535	.6676	.3758	<i>Sumsalg</i>
N:		179				
T:		80				
Totalt antall observasjoner:		8763				

Merk: Cluster-robuste standardfeil i parenteser under koeffisientestimer,

*asterisker angir signifikans: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$*

*Siden panelet er ubalansert tilsvarer ikke det totale antallet observasjoner $N*T$*

Regresjonsmodellene inneholder ukesdummier for å kontrollere for tidstrend.

6.6 Inferensgrunnlag

I avsnitt 5.2 ble det nevnt en mistanke om *cluster* i observasjonene, og generelt betyr dette at selv om det er uavhengighet på tvers av grupper vil det være korrelasjon mellom observasjoner innenfor grupper. Med et paneldatasett på individnivå vil hvert individ være et cluster så lenge det er seriekorrelasjon innenfor de individuelle feilleddene. Dette vil imidlertid ikke bryte med forutsetningene for konsistente estimater, så absoluttverdi og fortegn for disse vil ikke påvirkes. Problemet ligger i at cluster forstyrrer inferensgrunnlaget, noe som kan medføre at variabler tilegnes forklaringskraft selv om de respektive koeffisientestimater egentlig ikke er signifikant forskjellige fra null. Så langt har såkalte

cluster-robuste standardfeil blitt brukt for å danne inferensgrunnlag, en metode som i stor grad kan tilskrives flere bidragsytere.³³ Bertrand et. al (2004) ser på forskjellige måter å håndtere problemet med cluster på, og foreslår å bruke en *bootstrap* metode. En bootstrap estimering av standardfeilene kan også gjennomføres her for å se om dette gir store forskjeller relativt til de rapporterte cluster-robuste standardfeilene. Det antas ikke her at bootstrap metoden er å foretrekke fremfor cluster-robuste standardfeil, slik at denne ytterligere estimeringen av standardfeil kun gjennomføres for å se om det gir store forskjeller. Bootstrap standardfeil med 200 og 400 repetisjoner for DiD-estimatoren fra tabell 4, 5, 6 og 7 er gjengitt i tabell 8.³⁴ Tabellen inneholder også cluster-robuste og standard OLS estimater for standardfeilene for sammenligning.

Tabell 8: Cluster-robuste, bootstrap, og OLS-standardfeil

DiD-estimator	<u>Robust</u>	<u>Bootstrap</u>		<u>OLS</u>	Avhengig variabel
	$\hat{\sigma}_\delta$	$\hat{\sigma}_\delta(200)$	$\hat{\sigma}_\delta(400)$	$\hat{\sigma}_\delta$	
1.583	(0.857) [0.066]	(0.898) [0.078]	(0.879) [0.072]	(0.601) [0.008]	<i>Sumsalg</i>
1.506	(0.879) [0.088]	(0.965) [0.119]	(0.891) [0.091]	(0.604) [0.013]	<i>Sumsalg</i>
1.589	(0.853) [0.064]	(0.952) [0.095]	(0.900) [0.077]	(0.606) [0.009]	<i>Sumsalg</i>
0.171	(0.0783) [0.030]	(0.082) [0.037]	(0.082) [0.037]	(0.044) [0.000]	<i>Logsumsalg</i>

Merk: P-verdier i klammeparanteser under std.feil

Koeffisientestimerer er estimert med fast-effekt modellen.

Antall individer og tidsperioder er lik som i tabellene hvor de respektive koeffisientestimatene opptrer.

Selv om det er forskjell i de estimerte standardfeilene er det kun et tilfelle hvor endringen er så stor at et estimat ikke lenger kan godtas innenfor samme nivå som med cluster-robuste standardfeil (DiD-estimatoren fra tabell 3 med 200 repetisjoner). OLS standardfeilene er mindre enn de cluster-robuste, og dette viser også hvorfor det er viktig å ta høyde for cluster. Feilestimering av standardfeil øker forkastingsområdet til en nullhypotese om at koeffisientestimatet ikke er signifikant forskjellig fra 0. Dette øker igjen sannsynligheten for å

³³ Se for eksempel Nichols & Schaffer (2007) for en oversikt. Cluster robuste standardfeil refereres også til som blant annet Huber-white- eller sandwich-standardfeil.

³⁴ Grunnlaget for velge disse antallene repetisjoner er Cameron & Trivedi (2009, s.419).

påvise og akseptere en effekt, selv om den respektive forklaringsvariabelen egentlig ikke har forklaringskraft i regresjonsmodellen. Mindre standardfeil (og varians) er å foretrekke, men ikke dersom det er på grunn av feilestimering.

6.7 Kjønnforskjeller

Et interessant spørsmål kan være om det er forskjell i hvordan menn og kvinner responderer på incentiver. Siden FE-modellen ikke klarer å fange opp effekter av tidskonstante faktorer blir løsningen å estimere to modeller: én for menn og én for kvinner. Dersom det er signifikante forskjeller i DiD-estimatoren kan disse resultatene brukes for å si noe om hvorvidt det er slik at individincentiver har en sterkere effekt på kvinner enn på menn, eller vice versa. Tabell 7 gjengir resultatene for både menn og kvinner.

Tabell 9: *DiD-estimatoren fra standardutvalget for menn og kvinner separert*

Modell	DiD-estimator	FT-variabel	Menn		<i>overall</i>	P > t
			<i>within</i>	R^2 <i>between</i>		
FE	1.545 (1.341)	<i>Totalt1</i>	.2220	.4268	.2870	.251
N:		128				
T:		80				
Totalt antall observasjoner:		4662				
Modell	DiD-estimator	FT-variabel	Kvinner		<i>overall</i>	P > t
			<i>within</i>	R^2 <i>between</i>		
FE	1.487 (1.163)	<i>Totalt1</i>	.2635	.6149	.3783	.204
N:		147				
T:		80				
Totalt antall observasjoner:		4940				

*Merk: Cluster-robuste standardfeil i parenteser under koeffisientestimer, Siden panelet er ubalansert tilsvarer ikke det totale antallet observasjoner $N*T$ Regresjonsmodellene inneholder ukedummier for å kontrollere for tidstrend. Siden resultatene ikke er signifikante innenfor 10 % nivå inkluderes eksakt p-verdi.*

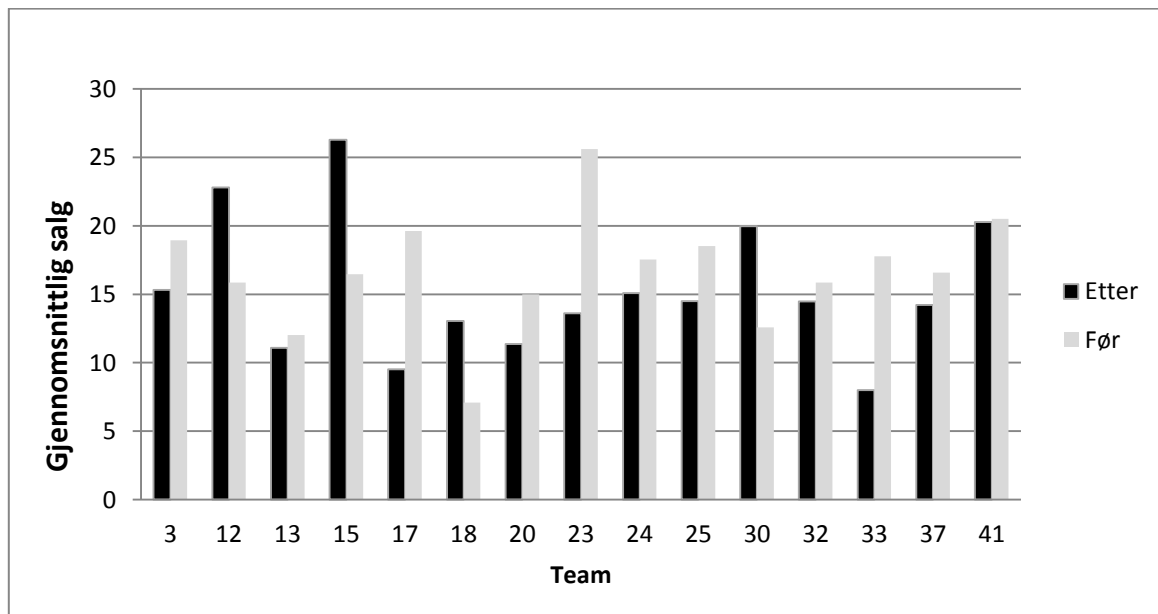
Ingen av resultatene er signifikante, og dette skyldes med stor sannsynlighet at utvalget blir for lite når det skilles mellom kjønn. Dermed blir estimeringsgrunnlaget for DiD-estimatoren ikke tilstrekkelig for å påvise en effekt av reformen for hverken menn eller kvinner isolert. Dersom resultatene likevel godtas som meningsfulle tilsier de at effekten av reformen var noe større for menn. Differansen er likevel på kun 0,058, en tilnærmet neglisjerbar forskjell.

Alternativt kan det sies at resultatene tilsier at det ikke er noen forskjell i hvordan menn og kvinner responderer på incentiver. I verdi er ikke DiD-estimatoren for noen av kjønnene veldig forskjellig fra den estimerte effekten på 1.583 i den opprinnelige modellen, og dette kan tolkes som en indikasjon på at det ikke er noen kjønnsforskjeller.

På grunn av usikkerhet rundt resultatene på grunn av det lave estimeringsgrunnlaget vil det ikke likevel ikke konkluderes med noe i forhold til dette.

6.8 Teamforskjeller

Et annet forhold av interesse kan være å se på om det er forskjeller i hvordan team med relativt høyt salg responderte på effekten i forhold til team med relativt lavt salg. Figur 2 viser gjennomsnittlig salg per uke for teamene i kontrollgruppen, og indikerer at det er til dels store forskjeller i hvordan teamene presterte før og etter reformen ble innført.



Figur 3: Gjennomsnittlig salg per team før og etter reform

En fremgangsmåte kan være å dele opp teamene i høyt, «normalt» og lavt salg ut i fra gjennomsnittlig salg i perioden før reform, og estimere en regresjonsmodell som inkluderer interaksjonsledd mellom en slik «produktivitetsindikator» og variabelen *reform*. Intuitivt kan det da tenkes at hvert av koeffisientestimatene til interaksjonsleddene vil gi effekten av reformen for den respektive gruppen, mens de samlet vil gi den totale effekten som tidligere er estimert til å ligge på 1.583 (fra standardutvalget). Resultatene er gjengitt i tabell 8, og det kan se ut som om det var teamene med «normalt» salg før reform som gav opphav til den estimerte endringen i salg som følger av reformen. Hverken teamene med gjennomgående høyt eller gjennomgående lavt salg før reformen har en estimert effekt signifikant forskjellig fra 0, mens det for teamene mellom var en signifikant økning utover det tidligere resultater tilsier. I tolkningen av disse resultatene er det imidlertid flere forbehold som må tas: For det første er det mulig at regresjonsmodellen ikke er korrekt spesifisert, og at resultatene dermed ikke kan tilegnes den tiltenkte tolkningen. Videre kan også grunnen til at det ikke kan identifiseres en signifikant effekt for høy- og lavproduktivitetsteamene, være at estimeringsgrunnlaget igjen er for lavt. Det kan derfor ikke konkluderes med at det var teamene i mellomstjiktet som responderte sterkest på effekten.³⁵

Tabell 10: Koeffisientestimater for interaksjonsledd mellom produktivitet og reform

	<u>Interaksjonsledd:</u>		
	<i>Høy_ref</i>	<i>Normalt_ref</i>	<i>Lavt_ref</i>
Koeffisient- estimat:	-0.471 (0.847)	2.530*** (0.867)	0.245 (0.837)
R ² : within:	.2561		
between:	.6287		
overall:	.3780		
N:	261		
T:	80		
Totalt antall observasjoner:	9602		

*Merk: Cluster-robuste standardfeil i parenteser under koeffisientestimater, asterisker angir signifikans: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$*

Siden panelet er ubalansert tilsvarende ikke det totale antallet observasjoner $N \cdot T$. Regresjonsmodellene inneholder ukesdummier for å kontrollere for tidstrend.

³⁵ En annen mulighet er å kun se på de to teamene som skiller seg ut med høyest og lavest salg før reformen. Dette ble forsøkt, men gav heller ikke signifikante resultater.

6.9 Individforskjeller

Et annet forhold som kan tenkes å skape komplikasjoner når teamprestasjoner vurderes, er at teamene ikke nødvendigvis er de samme gjennom hele observasjonsperioden. Dersom en ansatt bytter team eller det forekommer nyansettelser eller oppsigelser, er det ikke sikkert at teamene som måles før og etter kan sammenlignes. Dette gir grunnlag for å se på de enkelte ansatte fremfor team, og heller spørre om det er forskjeller i hvordan individer med relativt høyt eller relativt lavt salg før reformen responderte på endringen i bonusordning.

Fremgangsmåten her er som følger: Først identifiseres en gruppe ansatte som kan karakteriseres som «høyproduktive», og deretter en gruppe som kan karakteriseres som «lavproduktive» i perioden før reform (se fig. C.4 og C.5 i app. C). Elleve individer skiller seg ut som høyproduktive (gjennomsnittlig ukentlig salg høyere enn 25), og av disse er det to som slutter i løpet av observasjonsperioden. De resterende ni danner dermed høyproduksjonsgruppen. Deretter plukkes det ut ni individer som danner lavproduktivitetsgruppen (gjennomsnittlig ukentlig salg lavere enn 10). Alle individer i begge gruppene er i behandlingsgruppen. Ved å inkludere interaksjonsledd mellom hver av de to gruppene og *reform* i regresjonsmodellen, kan det tenkes at det er mulig å estimere en forskjell mellom høyproduktive og lavproduktive.

Resultatene er gjengitt i tabell A.5 (se app. A), og det er ingen signifikant forskjell i effekten av reformen for noen av de to gruppene. Dette kan indikere at det ikke var noen effekt av reformen for disse individene, men igjen gjør det lave estimeringsgrunnlaget at dette ikke kan konkluderes med.

6.10 Utvalgsstørrelse

Det er noen forhold som må tas hensyn til i tolkningen av resultatene. Et eksempel kan være dersom det ikke var endringen i bonusordning i seg selv som hadde en effekt, men at de ansatte økte innsatsen som svar på den økte overvåkingen fra bedriften. Dersom individene i kontrollgruppen korrekt forventer å bli mer grundig overvåket så lenge individuelle salgsresultater rapporteres, kan det være at det er dette som gir den sterkeste incentiveeffekten.³⁶ Et annet forhold er den relativt lave utvalgsstørrelsen, og dette blir nærmere gjennomgått her.

Datasettet denne regresjonsanalysen bygger på består av et panel med 261 individuelle observasjoner over en periode på 80 uker.³⁷ Blom et al. (2011) diskuterer et lignende problem i sin studie av hvordan forskjellig ledelse kan påvirke bedriftens prestasjoner. Utvalget deres består av 20 produksjonsanlegg som observeres i 100 uker. Dette er riktignok et lavere tverrsnittsutvalg enn det som er brukt i denne oppgaven, men noen av de samme resonnementene kan brukes her. For det første er det et spørsmål om utvalget er for lite til å identifisere signifikante endringer. Gjennom regresjonsanalysen er det påvist endringer som er signifikante innenfor et akseptabelt nivå. At resultatene er signifikante kan skyldes feilestimering av standardfeilene, men det er forsøkt flere fremgangsmåter for å ta høyde for dette. Forklaringen på at datasettet er rikt nok til å identifisere en effekt kan da gis på samme grunnlag som fra overnevnte studie. For det første er datamaterialet i stor grad er loggført og rapportert maskinelt, slik at det kan antas at det er liten tilstedeværelse av målefeil.³⁸ Videre vil den høye frekvensen av observasjoner over behandlingsperioden (ukentlig) kunne redusere utvalgsstørrelsen som trengs for å identifisere effekten av behandlingen.

Gitt at det blir påvist en effekt av reformen blir det neste spørsmålet om resultatene kan la seg utvide til å gjelde mer generelt. Dette kommer i stor grad an på hvordan man velger å tolke resultatene. På grunn av at både bonusordningene før og etter reform og arbeidsoppgaver er

³⁶ Det kan også spørres om det er andre faktorer som gjør at behandlingsgruppen alltid selger mer enn kontrollgruppen f.o.m uke 13. En måte å sjekke dette på er å lage en kontrafaktisk reform i uke 13, 2003 for å se om det finnes en effekt her, selv om tidsrommet før denne «placeboreformen» blir kort (et kvartal). DiD-estimatoren er gjengitt i tabell A.6 (se app. A), og dersom resultatet godtas på tross av det lave estimeringsgrunnlaget indikerer det at det ikke generelt er høyere salg blant behandlingsgruppen i tidsrommet etter uke 13.

³⁷ Merk at siden datasettet er ubalansert vil antallet observasjoner være lavere enn et tverrsnitt for 261 individer over 80 perioder tilsier.

³⁸ Målefeil kan skape både endogenitetsproblemer og større varians i koeffisientestimatene. Se Wooldridge (2006, kap. 9) for en nærmere gjennomgang av problemer som følger av målefeil.

relativt spesifikke for kundesenteret vi ser på, kan ikke datasettet sies å være representativt for populasjonen dersom dette forutsetter likhet på disse punktene. Forskjellige bedrifter operer med forskjellige bonusordninger, og ansattes produktivitet måles ut fra forskjellige størrelser. Dermed kan man argumentere for at disse resultatene i stor grad blir for snevre – de påviser en positiv effekt av å endre bonussystem i en spesifikk bedrift – og i beste fall kan dette tenkes å la seg utvide til lignende bedrifter som for eksempel kundesenter i andre forsikringsselskaper eller banker.

Men, formålet med regresjonsanalysen var å se om individbonus alt annet like påvirket ansattes innsats relativt til en ordning med gruppebonus. Den økonomiske tolkningen av resultatene er da at ansatte ceteris paribus får høyere motivasjon til å yte innsats dersom de belønnes individuelt i forhold til dersom belønningen skjer på gruppenivå. Dersom resultatene gir denne tolkningen er det mulig at resultatene kan la seg generalisere dersom de ansatte ved kundesenteret kan anses for å være et representativt utvalg for arbeidstakere. Spørsmålet blir imidlertid stående ubesvart, og den eneste løsningen som virker tilstrekkelig for å konkludere med om resultatet kan utvides er å studere dette empirisk i flere bedrifter.

Kapittel 7

Oppsummering og konklusjon

Hovedresultatene fra den økonometriske analysen kan oppsummeres som følger:

1. Resultatene tilsier at reformen alt annet like hadde en positiv effekt på salg. Den estimerte effekten er på 1.506 og 1.583 gjennomsnittlig solgte enheter, avhengig av om observasjoner som kan tenkes å være forstyrrende fjernes. Disse estimatene er statistisk signifikante på et 10 % nivå. Resultatene fra en log-level modell tilsier at økningen i salg som følger av reformen var på rundt 17 %.
2. Det er kun mindre endringer i verdien til DiD-estimatoren når individer som slutter eller kommer til i utvalget underveis i observasjonsperioden fjernes fra utvalget, og dette gir en god indikasjon på at den estimerte effekten ikke skyldes en seleksjonsprosess.
3. Det kan ikke konkluderes med at det var forskjeller i hvordan menn og kvinner responderte på endringen i bonussystem på bakgrunn av denne analysen, siden modellen ikke påviser signifikante forskjeller mellom kjønnene.
4. Det kan heller ikke konkluderes med team eller individer relativt høyt eller lavt salg responderte forskjellig på reformen. Resultatene indikerer, om noe, at det var team og ansatte som lå i mellomsjiktet med hensyn på salg før reformen som gir grunnlag for den positive effekten, men dette konkluderes det ikke med.

Hovedkonklusjonen blir dermed at å innføre individbonus økte innsatsen de ansatte la ned i hver telefonsamtale for å få gjennomført et salg. Dette forstås som at systemet med gruppebonus hadde en sub-optimal incentiveeffekt. Den videre tolkningen av dette kan være at det fantes tilfeller av gratispassasjeroppførsel som ble dempet som følger av at individbonus ble innført. I så måte er resultatene i tråd med klassisk økonomisk teori som predikerer at gruppeincentiver gir lavere innsats enn individuelle incentivmekanismer. Det tyder også da på at gratispassasjeroppførsel (helt eller delvis) fortrenget de positive effektene ved gruppebonus, og det kan finnes flere forklaringer på dette. For det første kan det være at gjensidig

overvåkning er vanskelig så lenge det ikke er anledning for den enkelte ansatte å observere kollegenes salgsresultater. Gitt at disse ikke offentliggjøres kan det tenkes å være at den enkelte ikke har anledning til å observere sidemannens salg (eller forsøk på salg), samtidig som man selv skal utføre et tilfredsstillende arbeid. Å gjennomføre et salg kan også anses for å være noe man ikke nødvendigvis «lærer seg». Det kan derfor tenkes at etter at en ansatt har opparbeidet seg grunnleggende ferdigheter i interne systemer og kunnskap nødvendig for å drive kundebehandling, er det lite som kan læres av de andre team-medlemmene for å øke salg. Selv om det finnes gode råd for hvordan man kan vinkle en kundesamtale mot et salg, kan det argumenteres for at å utvikle seg til en god selger er en prosess som hver enkelt må ta ansvar for selv. I så tilfelle vil faktorer som samarbeid og utveksling av kunnskap kun i begrenset grad kunne øke salg. Det er mulig at en eller flere av de positive sidene ved gruppebonus var tilstedeværende før reformen, men det var likevel forbedringspotensialer i kundesenterets incentivsystem siden de ansattes innsats økte etter at reformen var innført.

Resultatene tilsier også at den estimerte effekten av reformen ikke kan tilskrives en seleksjonsprosess, da det kun er en neglisjerbar endring i DiD-estimatoren dersom individer som slutter eller begynner ved kundesenteret i løpet av observasjonsperioden fjernes fra utvalget.

Det kan ikke påvises forskjeller i hvordan menn og kvinner responderte på reformen. Resultatene tilsier at det ikke var noen forskjell for noen av kjønnene isolert sett, og dette kommer mest sannsynlig av at estimeringsgrunnlaget er for lavt til å identifisere en effekt. Videre kan resultatene gi en indikasjon på at reformen hadde høyest effekt på motivasjonen til ansatte med et «normalt» salg før reformen, men heller ikke her er resultatene tilstrekkelige til at dette konkluderes med.

Hvorvidt det nye systemet gav en optimal tilpasning i forhold til innsatsnivå sett fra selskapets ståsted kan det ikke konkluderes med på bakgrunn av denne oppgaven, siden dette forutsetter en teori for hva som er en optimal tilpasning. Et videre spørsmål kan derfor være om en reform med enda sterkere monetære incentiver kunne hatt en ytterligere positiv effekt, gitt at potensiell lønnsutbetaling fremdeles var akseptabel for bedriften. Analysen sier heller ingenting om hvorvidt den nye bonusordningen medførte en forbedring for bedriften dersom andre forhold tas i betraktning. Et eksempel kan være kundenes opplevelser av kontakt med agenter som har høyt fokus på salg. Dersom det kan påvises at kundene ble mindre fornøyd med kundesenteret etter at reformen ble innført, kan det tenkes at det økte fokuset på salg har

gått på bekostning av serviceytelsen. Dette blir imidlertid kun spekulering. Formålet med analysen var da heller ikke å vurdere om ordninger med individbonus nødvendigvis er *bedre* enn ordninger med gruppebonus langs alle dimensjoner, men å studere endringen i innsats som følger av en endring i incentivsystem.

Problemstillingen det ble tatt sikte på å besvare var:

Vil en bonusordning som vurderer og belønner på individnivå ha en sterkere incentiveffekt enn en ordning med ren gruppebonus?

Gitt resultatene er svaret på dette ja. For den aktuelle bedriften kan det sies at den nye bonusordningen gav de ansatte motivasjon til å yte høyere innsats for å gjennomføre salg. Hvorvidt dette er et resultat som kan la seg generalisere kan det argumenteres for og i mot, men uten ytterligere empiriske studier lar det seg vanskelig gjøre å konkludere med noe.

Kapittel 8

Avslutning

Hovedformålet med denne oppgaven var å estimerte effekten en endring i bonussystem hadde på ansattes produktivitet, for å kunne si noe om forskjellige incentivsystemer. Gjennom en økonometrisk analyse har jeg fått resultater som tilsier at å gå fra en bonusordning som vurderte og belønnet ansatte på team-nivå til en ordning større individuelt fokus, alt annet like gav sterkere incentiver til å yte innsats for å gjennomføre salg. På bakgrunn av dette kan det argumenteres for at ordningen med gruppebonus hadde en sub-optimal incentiveffekt. Jeg har også prøvd å ta høyde for andre forhold og eventuelle kilder til problemer både i estimeringen og tolkningen, slik at jeg med større sikkerhet kan gå god for resultatene mine.

Å vurdere hvorvidt ordninger med individuell bonus er *bedre* enn ordninger med gruppebonuser dersom alle relevante forhold tas i betraktning, var ikke hensikten med denne oppgaven. Dermed blir heller ikke dette konkludert med. En oppgave av denne størrelsen vil uansett kun berøre temaet incentiveffekter, og det finnes generelt mange problemstillinger som står ubesvart.

Appendiks A: Regresjonsresultater

Tabell A.1: Resultater fra standardutvalget

Estimeringsmodell:	FE	FE	FE	RE	RE	RE	OLS	OLS	OLS
Bakgrunnsvariabel:									
<i>Reform</i>	-1.331 (0.839)	-0.457 (0.680)	-0.179 (0.705)	-0.0532 (0.544)	0.733* (0.375)	0.894** (0.369)	-1.115** (0.452)	0.325 (0.421)	0.570 (0.404)
<i>Treatment</i>				7.328*** (0.967)	6.146*** (0.850)	5.469*** (0.775)	4.998*** (0.378)	4.407*** (0.356)	3.979*** (0.346)
<i>Treat_ref</i>	1.089 (0.977)	1.686** (0.834)	1.583* (0.857)	-0.755 (0.624)	0.0563 (0.491)	0.132 (0.491)	0.687 (0.492)	1.252*** (0.460)	1.195*** (0.443)
<i>Loggtid</i>	0.00665*** (0.000319)			0.00664*** (0.000293)			0.00863*** (0.000161)		
<i>Telefoner</i>		0.0689*** (0.00284)			0.0702*** (0.00277)			0.0905*** (0.00157)	
<i>Total1</i>			0.0865*** (0.00344)			0.0886*** (0.00344)			0.112*** (0.00195)
<i>_cons</i>	4.584*** (0.703)	4.315*** (0.675)	4.472*** (0.644)	-2.859*** (0.929)	-2.099** (0.821)	-1.314* (0.748)	-3.150*** (0.647)	-3.062*** (0.607)	-2.215*** (0.599)
R ² : within:	0.2112	0.2412	0.2516	0.2227	0.2497	0.2584			
between:	0.4108	0.6058	0.6439	0.3651	0.5511	0.5949			
overall:	0.2588	0.3530	0.3762	0.2727	0.3637	0.3857	0.2811	0.3708	0.3912
N: 261	T: 80	Totalt antall observasjoner: 9602			Avhengig variabel: <i>sumsalg</i>				

Merk: Cluster-robuste standardfeil i parenteser, asterisker angir signifikans: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Siden panelet er ubalansert tilsvarende ikke det totale antallet observasjoner $N \cdot T$

Regresjonsmodellene inneholder ukesdummier for å kontrollere for tidstrend.

Tabell A.2: Resultater med restriksjoner på utvalget

Bakgrunnsvariabel:	Koeffisientestimat:		
<i>Reform</i>	-1.238 (0.860)	-0.363 (0.688)	-0.0880 (0.717)
<i>Treat_ref</i>	0.990 (1.008)	1.606* (0.854)	1.506* (0.879)
<i>Loggtid</i>	0.00665*** (0.000323)		
<i>Telefoner</i>		0.0689*** (0.00287)	
<i>Totalt1</i>			0.0866*** (0.00348)
<i>_cons</i>	4.621*** (0.705)	4.302*** (0.676)	4.466*** (0.648)
R ² : within:	0.2113	0.2416	0.2520
between:	0.4193	0.6239	0.6671
overall:	0.2553	0.3504	0.3744

Estimeringsmodell: Fast effekt

Avhengig variabel: *Sumsalg*

N:	221
T:	80
Totalt antall observasjoner:	9446

*Merk: Cluster-robuste standardfeil i parenteser, asterisker angir signifikans: * p < 0.1, ** p < 0.05, *** p < 0.01*
*Siden panelet er ubalansert tilsvarer ikke det totale antallet observasjoner N*T*
Regresjonsmodellene inneholder ukesdummier for å kontrollere for tidstrend.

Tabell A.3: Resultater fra log-level modell (standardutvalget)

Bakgrunnsvariabel:	Koeffisientestimat:		
<i>Reform</i>	-0.122* (0.0724)	-0.0490 (0.0653)	-0.0284 (0.0727)
<i>Treat_ref</i>	0.129 (0.0787)	0.181** (0.0717)	0.171** (0.0783)
<i>Loggtid</i>	0.000612*** (0.0000227)		
<i>Telefoner</i>		0.00589*** (0.000207)	
<i>Totalt1</i>			0.00720*** (0.000257)
<i>_cons</i>	1.531*** (0.0476)	1.574*** (0.0443)	1.612*** (0.0424)
R ² : within:	0.2692	0.2825	0.2843
between:	0.5118	0.6131	0.6155
overall:	0.3493	0.4008	0.4061

Estimeringsmodell: Fast effekt

Avhengig variabel: *Logsumsalg*

N:	261
T:	80
Totalt antall observasjoner:	9602

*Merk: Cluster-robuste standardfeil i parenteser, asterisker angir signifikans: * p < 0.1, ** p < 0.05, *** p < 0.01*
*Siden panelet er ubalansert tilsvarende ikke det totale antallet observasjoner N*T*
Regresjonsmodellene inneholder ukesdummier for å kontrollere for tidstrend.

Tabell A.4: Resultater når observasjoner av individer som ikke opptrer i begge perioder er droppet fra standardutvalget

Bakgrunnsvariabel:	Koeffisientestimat:		
<i>Reform</i>	-1.333 (0.827)	-0.449 (0.674)	-0.176 (0.700)
<i>Treat_ref</i>	1.086 (0.964)	1.691** (0.828)	1.589* (0.853)
<i>Loggtid</i>	0.00678*** (0.000338)		
<i>Telefoner</i>		0.0694*** (0.00300)	
<i>Totalt1</i>			0.0871*** (0.00363)
<i>_cons</i>	4.328*** (0.744)	4.141*** (0.716)	4.326*** (0.683)
R ² : within:	0.2154	0.2427	0.2532
between:	0.4743	0.6569	0.6844
overall:	0.2639	0.3554	0.3770

Estimeringsmodell: Fast effekt

Avhengig variabel: *Sumsalg*

N:	191
T:	80
Totalt antall observasjoner:	8868

Merk: Cluster-robuste standardfeil i parenteser,

*asterisker angir signifikans: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$*

Siden panelet er ubalansert tilsvarende ikke det totale antallet observasjoner $N \cdot T$

Regresjonsmodellene inneholder ukedummier for å kontrollere for tidstrend.

Tabell A.5: Individforskjeller

	<u>Interaksjonsledd:</u>		
	<i>Høy_ref</i>	<i>Lav_ref</i>	<i>Treat_ref</i>
Koeffisient- estimat:	0.528 (1.399)	0.320 (0.836)	1.529* (0.875)
R ² : within:	.2517		
between:	.6485		
overall:	.3790		

Estimeringsmodell: Fast effekt

Avhengig variabel: *Sumsalg*

N:	261
T:	80
Totalt antall observasjoner:	9602

*Merk: Cluster-robuste standardfeil i parenteser under koeffisientestimer, Siden panelet er ubalansert tilsvarer ikke det totale antallet observasjoner N*T Regresjonsmodellene inneholder ukesdummier for å kontrollere for tidstrend*

Tabell A.6: DiD-estimatoren for en «placeboreform» i uke 13, 2003

Modell	DiD-estimator	FT-variabel	R ²			P > t
			<i>within</i>	<i>between</i>	<i>overall</i>	
FE	0.737 (1.066)	<i>Totalt1</i>	.2282	.6816	.3678	.490

Estimeringsmodell: Fast effekt

Avhengig variabel: *sumsalg*

N:	195
T:	48
Totalt antall observasjoner:	5969

*Merk: Cluster-robuste standardfeil i parenteser under koeffisientestimer, asterisker angir signifikans: * p < 0.1, ** p < 0.05, *** p < 0.01 Siden panelet er ubalansert tilsvarer ikke det totale antallet observasjoner N*T Regresjonsmodellene inneholder ukesdummier for å kontrollere for tidstrend Siden resultatene ikke er signifikante innenfor 10 % nivå inkluderes eksakt p-verdi.*

Tabell A.7: Koeffisientestimat til ukedummiene fra standardutvalget

variabel	koef.est.	variabel	koef.est	variabel	koef.est
<i>wdummy3</i>	0.945* (0.501)	<i>wdummy19</i>	2.526** (1.169)	<i>wdummy35</i>	4.067*** (0.713)
<i>wdummy4</i>	0.0626 (0.685)	<i>wdummy20</i>	3.128*** (0.810)	<i>wdummy36</i>	2.449*** (0.635)
<i>wdummy5</i>	-0.186 (0.722)	<i>wdummy21</i>	-1.475** (0.746)	<i>wdummy37</i>	1.062** (0.507)
<i>wdummy6</i>	1.055 (0.644)	<i>wdummy22</i>	1.021 (1.101)	<i>wdummy38</i>	2.099*** (0.665)
<i>wdummy7</i>	0.975 (0.698)	<i>wdummy23</i>	4.112*** (0.831)	<i>wdummy39</i>	1.542** (0.637)
<i>wdummy8</i>	1.024 (0.781)	<i>wdummy24</i>	2.100*** (0.669)	<i>wdummy40</i>	3.087*** (0.663)
<i>wdummy9</i>	2.332*** (0.746)	<i>wdummy25</i>	4.598*** (0.731)	<i>wdummy41</i>	0.338 (0.594)
<i>wdummy10</i>	2.212*** (0.813)	<i>wdummy26</i>	5.308*** (0.891)	<i>wdummy42</i>	1.665*** (0.619)
<i>wdummy11</i>	2.547*** (0.744)	<i>wdummy27</i>	2.908*** (0.907)	<i>wdummy43</i>	1.882*** (0.571)
<i>wdummy12</i>	3.656*** (0.874)	<i>wdummy28</i>	5.247*** (0.922)	<i>wdummy44</i>	0.0293 (0.514)
<i>wdummy13</i>	1.494** (0.717)	<i>wdummy29</i>	3.887*** (0.726)	<i>wdummy45</i>	1.109** (0.558)
<i>wdummy14</i>	1.927** (0.751)	<i>wdummy30</i>	2.988*** (0.814)	<i>wdummy46</i>	0.286 (0.595)
<i>wdummy15</i>	2.974*** (0.897)	<i>wdummy31</i>	1.477** (0.715)	<i>wdummy47</i>	0.680 (0.579)
<i>wdummy16</i>	1.051 (0.822)	<i>wdummy32</i>	1.776*** (0.631)	<i>wdummy48</i>	-0.289 (0.537)
<i>wdummy17</i>	1.598** (0.788)	<i>wdummy33</i>	2.401*** (0.738)	<i>wdummy49</i>	0.165 (0.533)
<i>wdummy18</i>	1.824* (1.006)	<i>wdummy34</i>	2.184*** (0.604)	<i>wdummy50</i>	-0.846 (0.535)

Estimeringsmodell: Fast effekt

FT-variabel: *Totalt1*

Avhengig variabel: *sumsalg*

N:	261
T:	80
Totalt antall observasjoner:	9602

Merk: Cluster-robuste standardfeil i parenteser, asterisker angir signifikans: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Siden panelet er ubalansert tilsvarende ikke det totale antallet observasjoner $N \cdot T$

Dummyvariabelen for uke 2 er utelatt fra regresjonsmodellene for å fungere som referanse (benchmark)

Tabell A.8: Gjennomsnittsverdier og «within» standardavvik for de kontinuerlige variablene

	Variabel	Gj.snitt	Std.avvik (<i>within</i>)
Hele perioden	<i>Sumsalg</i>	16.4939	8.5596
	<i>Loggtid</i>	1408.635	582.6062
	<i>Telefoner</i>	138.2194	53.1332
	<i>Totaltl</i>	109.2459	43.70
Før reform	<i>Sumsalg</i>	16.7009	8.2123
	<i>Loggtid</i>	1430.382	448.8456
	<i>Telefoner</i>	148.1292	51.2419
	<i>Totaltl</i>	117.7908	42.46211
Etter reform	<i>Sumsalg</i>	16.2967	8.4665
	<i>Loggtid</i>	1388.687	487.3616
	<i>Telefoner</i>	129.1293	50.0903
	<i>Totaltl</i>	101.4077	40.87088

Appendix B: Tester

Breusch-Pagan test LM test for random effects

$$\text{sumsalg}[\text{memo_id},t] = Xb + u[\text{memo_id}] + e[\text{memo_id},t]$$

Estimated results:

	Var	sd = sqrt(Var)
<i>sumsalg</i>	124.4273	11.1547
e	56.00327	7.483533
u	12.01814	3.466718

H₀: Var(u) = 0

chibar2(01)	=	20206.78
Prob > chibar2	=	0.0000

Hausman test

b = consistent under H₀ and H_a; obtained from xtreg

B = inconsistent under H_a, efficient under H₀; obtained from xtreg

H₀: difference in coefficients not systematic

chi2(52)	=	(b-B)'[(V _b -V _B) ^{(-1)](b-B)}
	=	77.97
Prob>chi2	=	0.0114

(V_b-V_B is not positive definite)

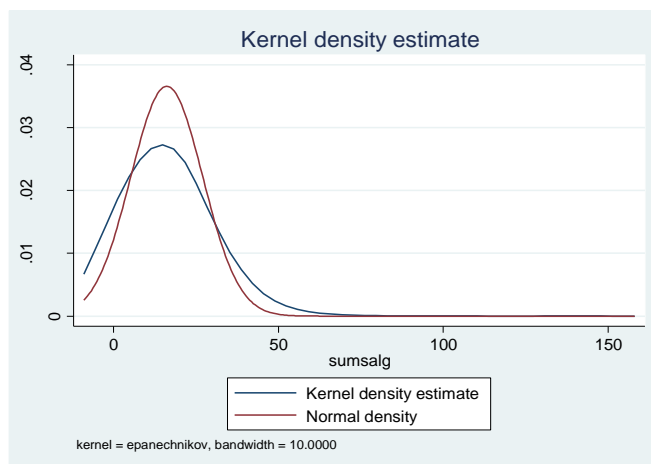
Test of overidentifying restrictions: fixed vs random effects

H₀: extra orthogonality conditions are valid

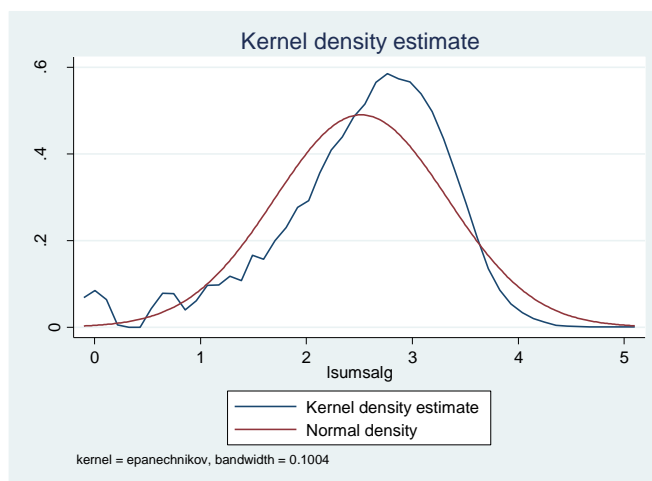
Cross-section time-series model: xtreg re robust cluster(memo_id)

Sargan-Hansen statistic	Chi-sq(52)	=	308.407
	P-value	=	0.0000

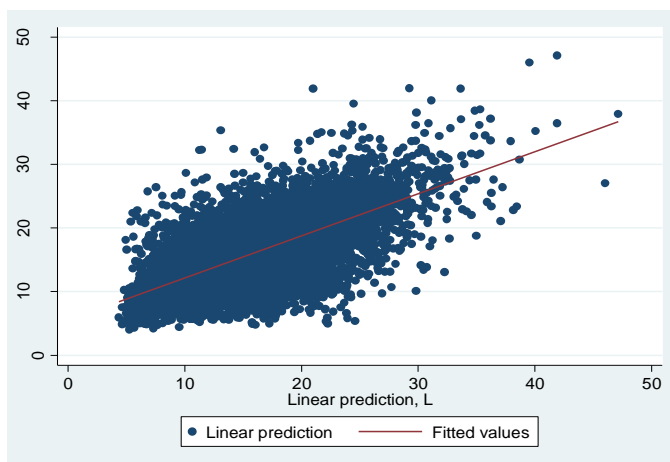
Appendix C: Figurer



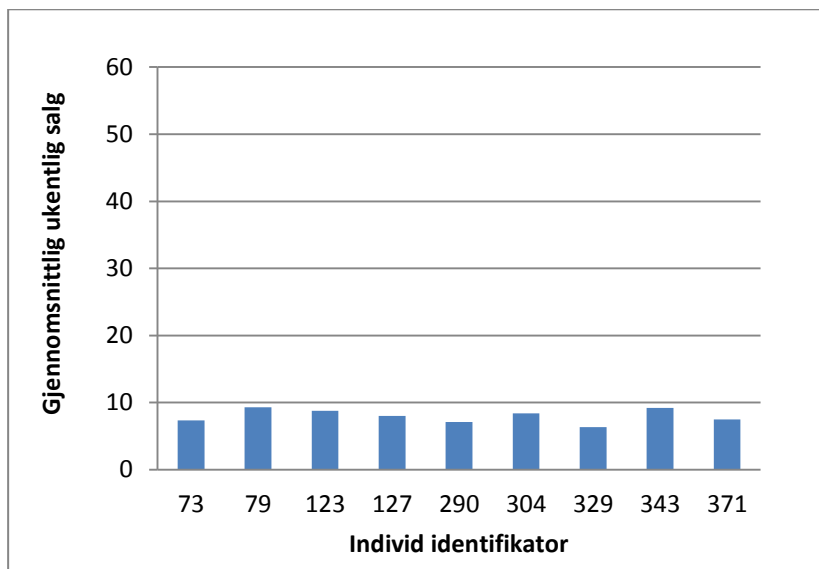
Figur C.1: Estimert tetthetsfunksjon og normalfordeling for *sumsalg* fra standardutvalget



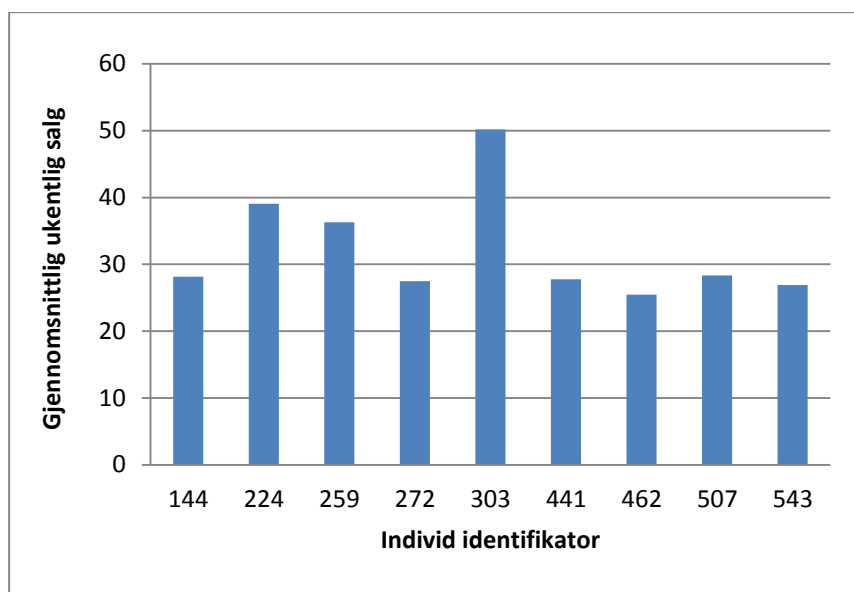
Figur C.1: Estimert tetthetsfunksjon og normalfordeling for *logsumsalg* fra standardutvalget



Figur C.3: Scatterplot av residualer mot residualer med et lag



Figur C.4: Ansatte med relativt lavt salg før reform



Figur C.5: Ansatte med relativt høyt salg før reform

Litteraturliste

Albouy, David (2004) *Program Evaluation and the Difference in Difference Estimator*. Notat i Public Economics, Economics 131, (vår 2004), University of California, Berkely, California. Tilgjengelig fra: <http://emlab.berkeley.edu/users/webfac/saez/e131_s04/diff.pdf> [Nedlastet 27.02.12]

Anderson, Simon P., Goeree, Jacob K., & Holt, Charles A. (1998) A Theoretical Analysis of Altruism and Decision Error in Public Goods Games. *Journal of Public Economics*, Vol. 70 (feb.), s. 279-323

Arellano, Manuel. (1993) On the Testing of Correlated Effects Within Panel Data. *Journal of Econometrics*, 59, s. 87-97.

Bloom, Nicholas, Eifert, Benn, Mahajan, Aprajit, McKenzie, David, Roberts, John. (2011) Does Management Matter? Evidence from India. *National Bureau of Economic Research Working Paper Series*, Working Paper 16658, JEL Nr. L2, M2, O14, O32, O33

Bragelien, Iver. (2005) 10 Bonustabber – Hvordan Lære av Teori og Praksis? *Praktisk Økonomi & Finans*, 21(2), s. 25-35

Breusch, Trevor S, Pagan, Adrian R. (1980) The Lagrange Multiplier Test and its Applications to Model Specification in Econometrics. *The Review of Economic Studies*, 47(1), s.239-253

Cameron, A. Colin, Trivedi, Pravin K. (2009) *Microeconometrics Using Stata*. Texas, Stata Press

Dougherty, Christopher. (2007) *Introduction to Econometrics*, 3. Etd, New York Oxford University Press Inc.

Gibbons, Charlie. (2009) *What's the Difference? Fixed Effects and Difference-in-Difference Approaches*. Handout i kurset Political Science (239): Quantitative Methodology in the Social Sciences Seminar, (25.02.2009), University of California, Berkeley, California. Tilgjengelig fra: <http://cgibbons.berkeley.edu/Courses/PS239_S09/FixedEffects.pdf> [Nedlastet 22.03.12]

- Guan, Weihua. (2003) From the help desk: Bootstrapped Standard Errors. *The Stata Journal*, 3(1), s. 71-80
- Hamilton, Barton H., Nickerson, Jack A., & Owan, Hideo. (2003) Empirical Analysis of the Impact of Teams on Productivity and Participation. *Journal of Political Economy*, 111(3), s. 465-497
- Hansen, Daniel G. (1997) Worker Performance and Group Incentives: A Case Study. *Industrial and Labor Relations Review*, 51(1), s. 37-49
- Hausman, Jerry A: Specification Tests in Econometrics. *Econometrica*, 46(6), s. 1251-1271
- Heckman, James J., Lalonde, Robert J. & Smith, Jeffrey A. (1999) The economics and econometrics of active labor market programs, I: Ashenfelter, Orley & Card, David. red. *Handbook of Labor Economics*, vol.3A, Amsterdam, Elsevier Science B.V., s. 1865-2097
- Helper, Susan, Kleiner, Morris M., & Wang, Yingchun (2010) Analyzing Compensation Methods in Manufacturing: Piece Rates, Time Rates, or Gain-Sharing? *National Bureau of Economic Research Working Paper Series*, Working Paper 16540, JEL Nr. J3,J31,L1,L2,L23,L6
- Hoechle, Daniel. (2007) Robust Standard Errors for Panel Regressions With Cross-Sectional Dependence. *The Stata Journal*,7(3), s. 281-312
- Jenkins, G. Douglas Jr, Mitra, Atul, Gupta, Nina, & Shaw, Jason D. (1998) Are Financial Incentives Related to Performance? A Meta-Analytic Review of Empirical Research. *Journal of Applied Psychology*, 83(5), s. 777-787
- Kandel, Eugene, & Lazear, Edward P. (1992) Peer Pressure and Partnerships. *Journal of Political Economy*, 100(4), s. 801-817
- Kuvaas, Bård. (2008) Hvorfor prestasjonsbasert belønning ofte skaper flere problemer enn det løser. *Praktisk Økonomi & Finans*, 24(2), s. 9-20
- Lazear, Edward P. (2000) Performance Pay and Productivity. *The American Economic Review*, 90(5), s.1346-1361
- Loughry, Misty L., & Tosi, Henry L. (2008) Performance Implications of Peer Monitoring. *Organization Science*, 19(6), s. 876-890.

Nalbantian, Haig R., & Schotter, Andrew. (1997) Productivity Under Group Incentives: An Experimental Study. *The American Economic Review*, 87(3), s. 314-341

Nichols, Austin, Schaffler, Mark. (2007) *Clustered Errors in Stata*. Presentasjon fra Stata 2007 UK Users Group Meeting, Cass Business School. Tilgjengelig fra:
<http://www.stata.com/meeting/13uk/nichols_crse.pdf> [Nedlastet 24.04.12]

Paarsch, Harry J, Shearer, Bruce. (1996) Piece Rates, Fixed Wages, and Incentive Effects: Statistical Evidence from Payroll Records. *Scientific Series*, Cirano, Working Paper 96s-31, JEL Nr. D2, J3, L2

Schaffer, M.E., Stillman, S. (2010) *xtoverid: Stata module to calculate tests of overidentifying restrictions after xtreg, xtivreg, xtivreg2 and xthtaylor*. Tilgjengelig fra:
<<http://ideas.repec.org/c/boc/bocode/s456779.html>> [Nedlastet 29.02.12]

Verbeek, Marno. (2008) *A Guide to Modern Econometrics*, 3. Etd, Chichester, John Wiley & Sons, Ltd.

Wooldridge, Jeffrey M. (2006) *Introductory Econometrics*, 3. Etd, Mason, Thomson South-Western