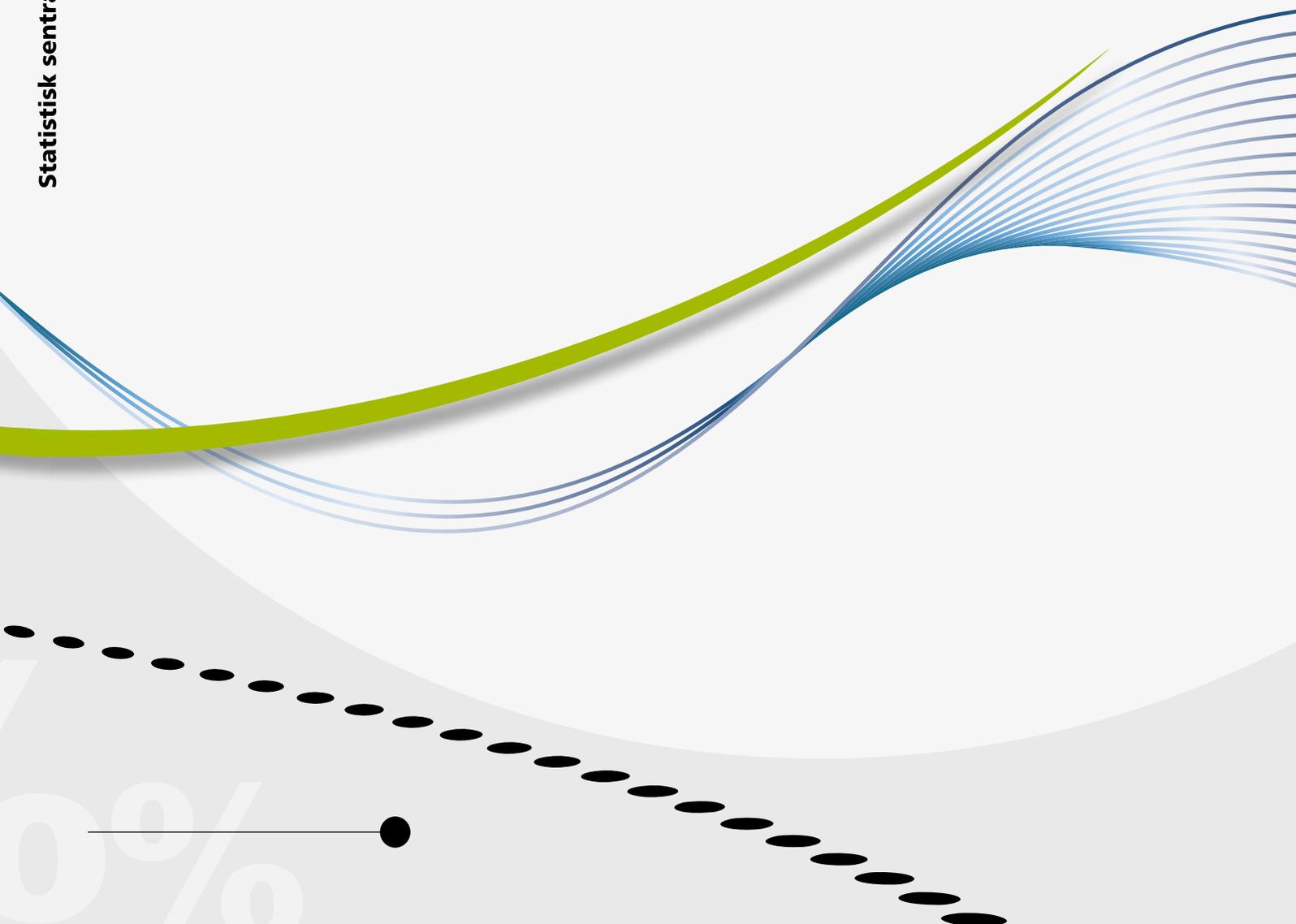
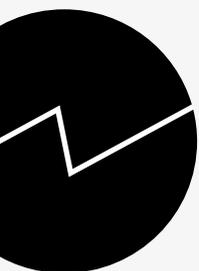


Susie Jentoft og Anna-Karin Mevik

Arbeidskraftundersøkelsen

Imputering av overtid ved indirekte intervju



Susie Jentoft og Anna-Karin Mevik

Arbeidskraftundersøkelsen

Imputering av overtid ved indirekte intervju

	Standardtegn i tabeller	Symbol
© Statistisk sentralbyrå	Tall kan ikke forekomme	.
Ved bruk av materiale fra denne publikasjonen skal Statistisk sentralbyrå oppgis som kilde.	Oppgave mangler	..
Publisert mars 2014	Oppgave mangler foreløpig	...
	Tall kan ikke offentliggjøres	:
	Null	-
ISBN 978-82-537-8886-9 (elektronisk)	Mindre enn 0,5 av den brukte enheten	0
Emne: Arbeid og lønn	Mindre enn 0,05 av den brukte enheten	0,0
	Foreløpig tall	*
	Brudd i den loddrette serien	—
	Brudd i den vannrette serien	
	Desimaltegn	,

Forord

Opprinnelig ble spørsmål om overtid i Arbeidskraftundersøkelsen bare stilt ved direkte intervju. Men fra og med 2006 ble det også stilt ved indirekte intervju. Ved sammenligning av svarene til de med direkte og indirekte intervju, virker det som om disse gruppene svarer forskjellig når det gjelder overtid. Dette er bakgrunnen for analysene og imputeringene som er gjort i dette notatet. Prosjektet er bestilt av seksjon for arbeidsmarkedstatistikk i 2011.

Statistisk sentralbyrå, 27. januar 2014

Hans Henrik Scheel

Sammendrag

I Arbeidskraftundersøkelsen spørres det blant annet om overtid. I dette notatet ser vi på de to overtidsvariablene *overt_tot* og *overt_tot_tim*. Den førstnevnte variabelen angir om intervjuobjektet har jobbet overtid eller ikke i referanseuken, mens den andre variabelen angir hvor mange timer vedkommende har jobbet overtid.

Sammenligning viser at de direkte og indirekte intervjuede svarer forskjellig på disse variablene, selv når vi tar hensyn til variable som påvirker overtidsbruken. Dette kan tyde på at kvaliteten til de indirekte intervjuene ikke er så bra (når det gjelder overtid). Vi anbefaler derfor å ikke bruke rapportert overtid for de indirekte intervjuene, men i stedet imputere overtidsbruken når det skal estimeres tall for publisering.

Vi har testet ulike imputeringsmetoder. For variabelen *overt_tot* foreslår vi å bruke en ikke-stokastisk imputering som baserer seg på en modellert sannsynlighet for å jobbe overtid. For variabelen *overt_tot_tim* anbefaler vi en nærmeste nabo-imputering med predictive mean matching.

Innhold

Forord	3
Sammendrag	4
Innhold	5
1 Innledning	6
2 Arbeidskraftundersøkelsen	6
2.1 Indirekte intervju	6
2.2 Overtid	7
3 Analysedataene	7
4 Analyse av variabelen som angir om det er jobbet overtid eller ikke (overt_tot)	8
4.1 Sammenligning av variabelen for de med direkte og indirekte intervju	8
4.2 Imputering av overtid ved indirekte intervju og partielt frafall	9
4.2.1 Variabelseleksjon	9
4.2.2 Imputeringsmetode	12
4.3 Imputeringsresultat	13
5 Analyse av overtidstimer	16
5.1 Overtidstimer i AKU	16
5.2 Sammenligning av overtidstimer for de med direkte og indirekte intervju	16
5.3 Imputeringsmetode	17
5.3.1 Variabelseleksjon	17
5.3.2 Residual regresjons-imputering	19
5.3.3 Nærmeste nabo-imputering	19
5.4 Imputeringsresultater	20
6 Konklusjon	23
Vedlegg A. Andel intervjuobjekt som har rapportert at de har jobbet overtid (%), per uke i 2012, fordelt etter direkte intervju og indirekte intervju	24
Vedlegg B. Fordeling av antall overtidstimer i 2012 med og uten nærmeste nabo-imputering	25
Referanser	26
Figurregister	27
Tabellregister	27

1 Innledning

I Arbeidskraftundersøkelsen, som er en intervjuundersøkelse, er en del av intervjuene indirekte. På spørsmål om overtid er andelen som rapporterer overtid lavere blant de med indirekte intervju enn de med direkte intervju. I 2012 for eksempel ble det rapportert at 7,4 % av de indirekte intervjuede hadde jobbet overtid, mot 12,3 % av de direkte intervjuede.¹ Antall overtidstimer per person derimot, ble i snitt rapportert litt høyere for de indirekte intervjuede enn for de direkte intervjuede.

I dette notatet skal vi se om det fortsatt er forskjell når vi tar hensyn til sammensettingen av personer som har direkte og indirekte intervju. Vi skal også vurdere ulike måter å imputere overtidvariablene når det er indirekte intervju (eller partielt frafall).

Vi starter i kapittel 2 med en rask presentasjon av Arbeidskraftundersøkelsen. En mer utførlig dokumentasjon er gitt i Arbeidskraftundersøkelsen 2001 (NOS C748). I kapittel 3 gir vi en oversikt over dataene som brukes i analysene. Vi analyserer først variabelen som angir om en person har jobbet overtid eller ikke, så ser vi på variabelen for antall overtidstimer. Resultatene er gitt i henholdsvis kapittel 4 og 5. I kapittel 6 er det en kort konklusjon basert på analysene.

2 Arbeidskraftundersøkelsen

Arbeidskraftundersøkelsen (AKU) er en intervjuundersøkelse som gjennomføres hvert kvartal. Formålet med undersøkelsen er å gi informasjon om utviklingen i sysselsetting og arbeidsledighet, og om ulike befolkningsgruppers tilknytning til arbeidsmarkedet.

Hovedkilden for AKU er kvartalsvise utvalgsundersøkelser. Et kvartalsutvalg består av ca 24 000 personer i alderen 15-74 år. Utvalget trekkes ved at det først trekkes familieenheter, så tas alle familiemedlemmene til de uttrukne familieenheterne med i utvalget, hvis de er i den aktuelle aldersgruppen. Hver person som trekkes ut til å være med i AKU, er med i 8 påfølgende kvartalsutvalg.

Hver person i et kvartalsutvalg intervjues om sin tilknytning til arbeidsmarkedet i en nærmere spesifisert referanseuke. For å sikre at alle ukene blir kartlagt, blir utvalget spredt utover hele kvartalet.

2.1 Indirekte intervju

I utgangspunktet skal det gjøres direkte intervju. Direkte intervju betyr at en person gir opplysninger om seg selv. Men i noen tilfeller der vi ikke får tak i den vi skal samle inn opplysninger om, gjennomføres det indirekte intervju. Indirekte intervju betyr at opplysningene om en person gis av et annet familiemedlem.

Regelen for indirekte intervju er at foreldre kan svare for sine barn, og ektefeller/samboere kan svare for hverandre. Barn kan ikke svare for andre enn seg selv. Ikke alle spørsmålene i spørreskjemaet blir stilt ved indirekte intervju.

I perioden 2009-2012 er det ca 14 % indirekte intervju. Det er litt færre indirekte intervju om vi begrenser oss til å kun se på de som er ansatte i inntektsgivende arbeid (det er denne gruppen vi ser på når vi analyserer overtidvariablene). Da er det ca 11 % indirekte intervju.

¹ Dette gjelder overtid i hovedarbeidsforholdet, målt i % av de som er ansatte i inntektsgivende arbeid.

2.2 Overtid

I AKU spørres det om overtid, både i hovedarbeidsforholdet og i eventuelt biarbeidsforhold. I dette notatet ser vi på overtid i hovedarbeidsforholdet.

Med overtid menes arbeidstid utover avtalt arbeidstid, utført i løpet av referanseuken. Så vel betalt som ubetalt overtid skal inkluderes. For nøyere beskrivelse av overtid, se Bø og Håland, 2002 (kapittel 4.3).

Opprinnelig ble spørsmål om overtid bare stilt ved direkte intervju. Men fra og med 2006 er det også stilt ved indirekte intervju. Som nevnt er det færre som rapporterer overtid ved indirekte intervju enn ved direkte. Samtidig er timetallet som rapporteres i snitt litt større for de indirekte intervjuene enn de direkte.

3 Analysedataene

Vi har sett på data fra 2009, 2010, 2011 og 2012.

Utvalgene og de publiserte tallene er kvartalsvise. Men pga. sesongjusteringen estimeres det per måned, og så beregnes kvartalstallene ut fra månedstallene. Derfor analyserer vi på månedsdataene.

Vi skal se på overtidsbruken i hovedarbeidsforholdet. Ikke alle i utvalget får spørsmål om dette. Det er kun de som er ansatt i inntektsgivende arbeid, og som ikke er midlertidig fraværende fra hovedarbeidsforholdet, som blir spurt om det. Med variablene som fins i datasettene, tilsvarende dette personer som har $Hstat = '111'$ og $SYS26 = '1', '2'$ eller $'3'$. Det er denne gruppen vi analyserer på.

Når vi tar bort enhetsfrafallet får vi månedlige analysedata som består av omtrent 3 600 intervjuobjekt. Unntaket er juli måned, som har ca 2 000 intervjuobjekt.

Andelen indirekte intervju varierer fra måned til måned. De fleste månedene har en andel på ca 11 %, mens de ekstremeste månedene har 6,6 og 15,0 % indirekte intervju.

Vi baserer våre analyser på de to overtidsvariablene $overt_tot$ og $overt_tot_tim$ som fins i analysedataene. Variablen $overt_tot$ registrerer hvorvidt intervjuobjektet har jobbet overtid eller ikke i referanseuken, mens $overt_tot_tim$ angir antall timer intervjuobjektet har jobbet overtid i løpet av referanseuken.² Begge variablene gjelder hovedarbeidsforholdet.

Variablene $overt_tot$ er kodet på følgende måte

$$overt_tot = \begin{cases} 1 & \text{hvis intervjuobjektet har jobbet overtid i referanseuken} \\ & \text{hvis intervjuobjektet ikke har jobbet overtid i referanseuken} \\ 9 & \text{hvis partielt frafall (dvs. ubesvart)} \end{cases}$$

Det er veldig få som har partielt frafall, færre enn 0,5 % av analysedataene.

Timeantallet i $overt_tot_tim$ er avrundet til hele timer. Hvis intervjuobjektet ikke har jobbet overtid, mangler $overt_tot_tim$ verdi. Hvis $overt_tot$ har partielt frafall, har $overt_tot_tim$ også partielt frafall, og mangler i så fall verdi. $overt_tot_tim$ kan også ha partielt frafall uten at $overt_tot$ har det (men dette forekommer svært sjelden). Variablen er i så fall kodet med 998.

² Overtidsarbeid som varer mindre enn en halv time totalt i løpet av referanseuken, blir pga avrunding ikke registrert som overtid i disse variablene.

4 Analyse av variabelen som angir om det er jobbet overtid eller ikke (*overt_tot*)

4.1 Sammenligning av variabelen for de med direkte og indirekte intervju

Andelen som rapporterer at de har jobbet overtid er lavere for de med indirekte intervju enn de med direkte intervju. I utvalget for januar 2012 er det for eksempel 7,4 % som har jobbet overtid blant de med indirekte intervju, mot 11,9 % blant de med direkte intervju (se tabell 1).³

Tabell 1. Andel intervjuobjekt som har jobbet overtid (%), per måned 2012, fordelt etter type intervju

	Direkte intervju	Indirekte intervju
Januar	11,92	7,41
Februar	12,07	7,72
Mars	11,65	6,13
April	9,95	4,10
Mai	11,22	7,49
Juni	14,99	10,61
Juli	12,49	6,62
August	12,29	7,49
September	14,49	6,09
Oktober	13,77	6,84
November	12,76	8,56
Desember	11,38	8,99

I vedlegg A er det en figur som viser andel intervjuobjekt som har jobbet overtid per uke.

Deler vi opp utvalget etter kjønn og alder, er overtidsbruken fortsatt lavere for de med indirekte intervju, med noen få unntak. Tall for januar 2012 er vist i tabell 2.

Tabell 2. Andel intervjuobjekt som har jobbet overtid (%), januar 2012, fordelt etter kjønn, alder og type intervju

	Direkte intervju	Indirekte intervju
Menn		
15-24 år	13,61	6,98
25-34 år	16,62	7,14
35-54 år	17,07	8
55-74 år	9,32	10
Kvinner		
15-24 år	7,02	5,26
25-34 år	8,53	10,53
35-54 år	9,92	6,82
55-74 år	6,49	6,25

Spørsmålet er om det fortsatt er forskjell mellom direkte og indirekte intervju når vi tar hensyn til flere variable som har betydning for overtidsbruken.

For å sjekke dette har vi modellert sannsynligheten for at en person jobber overtid ved hjelp av logistisk regresjon. Vi har brukt to sett med forklaringsvariable. Det ene består av variable for alder, kjønn, uketype, heltids-/deltidsarbeid, yrke og lederansvar/ikke lederansvar. Det andre består i tillegg av variable som angir ekteskaplig status, antall arbeidsforhold og om intervjuobjektet har hatt noe fravær i referanseuken. (Variablene blir nøyere beskrevet i neste avsnitt). For å teste om intervjutype har betydning for sannsynligheten for å jobbe overtid, bruker vi i tillegg en forklaringsvariabel som angir om det er et direkte intervju eller ikke.

³ Forskjellen kan ikke forklares med det partielle frafallet. Det partielle frafallet ligger mellom 0,00-0,43 % og 0,00-1,73 % for hhv direkte og indirekte intervju, så selv om alle med partielt frafall for indirekte intervju jobber overtid, mens alle med partielt frafall for direkte intervju ikke jobber overtid, er det fortsatt forskjell i overtidsbruken.

Resultatet av modelltilpassingene viser tydelig at de med indirekte intervju har lavere sannsynlighet for å jobbe overtid enn de med direkte intervju. Dvs. selv når vi korrigerer for de nevnte forklaringsvariablene, er overtidsbruken mindre for de med indirekte intervju.

Dette tyder på at overtidsvariabelen *overt_tot* blir underrapportert ved indirekte intervju. Tas de indirekte intervjuene med i estimeringen, medfører det i så fall til underestimering av populasjonstotaler og -andeler. Derfor anbefaler vi å imputere overtidsbruken til de med indirekte intervju, i stedet for å bruke de rapporterte verdiene, når AKU-tall skal estimeres og publiseres.

4.2 Imputering av overtid ved indirekte intervju og partielt frafall

Dette kapitlet foreslår hvordan *overt_tot* kan imputeres når det er indirekte intervju eller partielt frafall.

Imputeringsmetoden bygger på følgende forutsetning: når vi tar hensyn til en del variable som påvirker overtidsbruken, skiller ikke overtidsbruken til de med indirekte intervju eller partielt frafall seg fra overtidsbruken til de med direkte intervju. Dette tilsvarer å anta at gitt en del variable som påvirker overtidsbruken, så er det tilfeldig hvem som har indirekte intervju eller partielt frafall.

Kort oppsummert går imputeringsmetoden ut på først å beskrive sannsynligheten for å jobbe overtid. Så brukes denne sannsynligheten til å finne det forventede antall personer som jobber overtid i ulike grupper. Dette forventede antall plukkes så ut fra de personene det skal imputeres for, og de får imputert at de har jobbet overtid (*overt_tot* = 1). Resten får imputert at de ikke har jobbet overtid (*overt_tot* = 0).

For å beskrive overtidssannsynligheten bruker vi en logistisk regresjonsmodell. Dvs. vi modellerer sannsynligheten for at en person jobber overtid med

$$p(x_1, x_2, \dots) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots}}$$

der x_1, x_2, \dots er variable som har betydning for overtidsbruken.

4.2.1 Variabelseleksjon

For å plukke ut variable som har betydning for overtidsbruken, har vi brukt en trinnvis seleksjonsprosedyre i SAS. Prosedyren er en blanding av forlengs og baklengs seleksjonsmetode. Den starter som en forlengs metode, men kan også kaste ut variable under veis.⁴

Vi har kjørt seleksjonsprosedyren på alle de månedlige analysedataene (vi bruker bare de som har direkte intervju). Variablene vi har testet er en blanding av registervariable og variable som innhentes i intervjuet. Det varierer litt fra måned til måned hvilke variable som tas med.

Vi har valgt å bare bruke variable som kommer med minst ¾ av de i alt 48 månedene vi tester på. Det er følgende seks variable (uten kryssledd):

- kjønn
- alder
- uketype

⁴ Seleksjonsprosedyren heter "stepwise", og vi har brukt "slentry"=0,15 og "slstay"=0,20 ("slentry" er grensen for å ta inn ny variabel, mens "slstay" er grensen for å kaste ut en variabel). For mer om denne seleksjonsprosedyren i SAS, se Heldal, 2006, kapittel 6.3.

- heltids-/deltidsarbeid
- yrke
- lederansvar/ikke lederansvar

Alder er delt inn intervaller. Vi har prøvd med litt ulike intervallinndelinger, og valgt den som forklarte overtidsbruken best:

1. 15-24 år
2. 25-34 år
3. 35-54 år
4. 55-74 år

Uketype er en klassifisering av referanseuke, og deler inn referanseukene i vanlige og spesielle uker. Spesiell uke betyr at det generelt er færre som jobber overtid. Etter en del testing er følgende uker klassifisert som spesiell:

- uke 1, hvis 1. nyttårsdag faller på en ukedag i denne uken
- påskeukene (de to ukene som har Skjærtorsdag, Langfredag, 1. og 2. påskedag)
- uke 51, hvis julaften faller på en ukedag i denne uken
- uke 52 (og uke 53 hvis gjeldende år har 53 uker)
- hvis det i mai/juni fins en uke som har to røde⁵ ukedager, klassifiseres denne uken som spesiell

Med ukedag mener vi mandag til fredag.

Yrke er delt opp i tre grupper:

1. ledere, akademiske yrker, høyskoleyrker og kontoryrker
2. håndverkere, prosess- og maskinoperatører, transportarbeidere mv.
3. andre yrker

Heltids-/deltidsarbeid skiller mellom heltidsarbeid og deltidsarbeid, mens lederansvar/ikke lederansvar angir om det inngår lederansvar i arbeidsoppgavene.

Tabell 3. Estimerte modellparametre (parametrisert med referansegruppe), januar 2012

Parameter	Estimat	Standard feil	p-verdi
Konstantledd	-1,72	0,21	<,0001
Kjønn			
Kvinne	-0,24	0,13	0,0549
Alder			
25-34 år	-0,13	0,22	0,561
35-54 år	-0,08	0,20	0,6996
55-74 år	-0,60	0,24	0,0114
Heltids-/deltidsarbeid			
Deltid	-0,64	0,18	0,0004
Yrke			
2	0,50	0,15	0,0005
3	-0,47	0,17	0,0073
Lederansvar/ikke lederansvar			
Lederansvar	0,28	0,12	0,0177

I denne tabellen er ikke uketype tatt med, fordi januar 2012 bare har vanlige uker.

⁵ Røde dager i mai/juni er 1. mai, 17. mai, Kristi Himmelfartsdag, 1. og 2. pinsedag.

Tabell 4. Estimerte modellparametre (parametrisert med referansegruppe), april 2012

Parameter	Estimat	Standard feil	p-verdi
Konstantledd	-1,75	0,23	<,0001
Kjønn			
Kvinne	-0,31	0,14	0,029
Alder			
25-34 år	-0,05	0,24	0,8175
35-54 år	-0,13	0,22	0,5536
55-74 år	-0,76	0,27	0,0043
Uketype			
Spesiell	-0,48	0,13	0,0001
Heltids-/deltidsarbeid			
Deltid	-0,49	0,20	0,0127
Yrke			
2	0,53	0,16	0,0007
3	-0,47	0,19	0,0138
Lederansvar/ikke lederansvar			
Lederansvar	0,36	0,13	0,0047

I tabell 3 og 4 ser vi eksempel på de estimerte β_j 'ene i modellen, med standardfeil og p-verdi for nullhypotesen $\beta_j = 0$. Alle forklaringsvariablene er kategoriske, og vi har brukt parametrisering med referansegruppe. Som referansegruppe har vi brukt mann, 15-24 år, vanlig uke, heltidsarbeid, yrkegruppe 1 og ikke lederansvar.

Vi ser at den estimerte parameteren til "kjønn kvinner" er negativ. Det betyr at kvinner har lavere sannsynlighet for å jobbe overtid enn menn. Tilsvarende er det mindre sannsynlig å jobbe overtid i en spesiell uke, sammenlignet med en vanlig uke, og for de som jobber deltid, sammenlignet med de som jobber heltid. For de som har lederansvar er parameteren positiv, som betyr at de har større sannsynlighet for å jobbe overtid enn de uten lederansvar.

For yrke ser vi at det er yrkesgruppe 2 som har størst sannsynlighet for å jobbe overtid, mens yrkesgruppe 3 har minst. Det fins noen få tilfeller hvor rangeringen avviker fra dette.

For alder er det 55-74-åringene som jobber minst overtid. Forskjellen mellom de andre aldersgruppene er mindre tydelig, og rangeringen av dem kan variere fra måned til måned. I ca. halvparten av tilfellene er det 25-34 åringene som jobber mest overtid, tett fulgt av 15-24-åringene, og så 35-54-åringene. I ca 1/3 av tilfellene har vi tilsvarende rangering som i tabell 4, der det er 15-24-åringene som jobber mest overtid, tett fulgt av 25-34-åringene, og så 35-54-åringene. I de resterende tilfellene varierer rangeringen (men 55-74-åringene har alltid minst overtid).

De andre variablene vi har vurdert, men valgt å ikke ta med i modellen, er utdanning, kommuneklassifisering, kommunesentralitet, antall arbeidsforhold, ekteskaplig status og fravær i løpet av referanseuken. Utdanning, kommuneklassifisering og kommunesentralitet hadde så å si ingen betydning for overtidsbruken (i alle fall ikke når de valgte forklaringsvariablene er med). Antall arbeidsforhold og ekteskaplig status ser ut til å ha betydning i noen av de 48 månedene vi ser på, men ikke mange nok til at vi har valgt å ta dem med i modellen.

Fravær i løpet av referanseuken ser ut til å ha stor betydning for overtidsbruken. (Det er snakk om fravær i deler av uken, hvis det er fravær i hele uken blir ikke spørsmål om overtid stilt). Grunnen til at vi ikke tar med denne variabelen er at den har mye partielle frafallet. Det er omtrent 10 % partielt frafall for de indirekte intervjuene, og 1,5 % for de direkte intervju.

4.2.2 Imputeringsmetode

Imputere med $p(\underline{x})$

Alle intervjuobjektene, også de med indirekte intervju, har nå en individuell sannsynlighet $p(\underline{x})$ for å jobbe overtid. Denne sannsynligheten (etter at den er estimert) kan brukes som imputert verdi for overtidsvariabelen. Dette er en imputering som vi mener gir bra resultat med tanke på estimerte populasjonstotaler og -andeler (forutsatt at overtidsmodellen også passer for de med indirekte intervju). Men med denne imputeringen blir ikke den imputerte overtidsvariabelen en 0/1-variabel, slik det er ønsket. Derfor er denne imputeringsmetoden ikke aktuell.

Imputere med trekking fra Bernoullifordeling

En måte å imputere 0/1-verdier på, er å trekke fra en Bernoullifordeling med "suksesssannsynlighet" $p(\underline{x})$.⁶ Dette er en stokastisk imputering. Testing viser at resultatet kan bli nokså forskjellig ved gjentatte imputeringer. For januar 2012 fikk vi at antall intervjuobjekter som ble imputert at de jobbet overtid, varierte fra 21 til 39 når vi gjentok imputeringen 12 ganger. De estimerte overtidsandelene ble ikke så veldig forskjellige når vi beregnet dem for hele analysepopulasjonen, de varierte fra 12,0 til 12,5 %. Men for 15-24 åringene ble estimatene veldig forskjellige, de varierte fra 10,3 til 12,4 %.

På grunn av dette mener vi at hvis denne imputeringsmetoden skal benyttes, bør den kombineres med multippel imputering. Multippel imputering betyr at man gjentar imputeringen flere ganger, og bruker gjennomsnittet av estimatene fra hver imputering som endelig estimat (for mer om multippel imputering, se Bjørnstad, 2007). Dermed får vi ikke bare ett imputert datasett, slik det er ønsket. Derfor er heller ikke denne imputeringsmetoden aktuell.

Imputere per gruppe

Det vi i stedet foreslår er følgende. Del opp intervjuobjektene i grupper ved hjelp av forklaringsvariablene. Innen hver gruppe har alle lik sannsynlighet for å jobbe overtid, og det forventede antallet som jobber overtid, blant dem det skal imputeres for, er gitt ved

$$n_g p_g$$

der n_g er antall intervjuobjekt det skal imputeres for (i gruppe g), og p_g er sannsynligheten for å jobbe overtid (i gruppe g).⁷ Dette forventede antallet plukkes ut fra dem det skal imputeres for, og de får imputert at de har jobbet overtid ($overt_tot = 1$). Resten får imputert at de ikke har jobbet overtid ($overt_tot = 0$).

For å plukke ut de som skal få imputert $overt_tot = 1$, sorteres intervjuobjektene først etter estimeringsvekten. Dette gjøres per gruppe. Så gjøres det en systematisk utplukking. For eksempel, hvis det skal plukkes ut tre stykker i en gruppe, er det 1. kvartil, median og 3. kvartil som plukkes ut. På den måten blir estimeringsvekten omtrent likt fordelt for de som får imputert $overt_tot = 1$ og for de som får imputert $overt_tot = 0$.

I stedet for å gjøre en systematisk utplukking av de som skal få imputert $overt_tot = 1$, kunne vi gjort en enkel tilfeldig trekking (per gruppe). Vi har vurdert dette, men kommet fram til at også denne stokastiske imputeringen bør kombineres med multippel imputering. Derfor er tilfeldig trekking ikke aktuelt.

⁶ En Bernoullifordeling er fordelingen til en stokastisk variabel som bare kan ha verdiene 0 og 1.

⁷ p_g er ukjent og må først estimeres. Vi bruker bare de med direkte intervju til estimeringen.

Det er viktig å merke seg at hvis man skal beregne variansen til de estimerte populasjonstotalene og -andelene, underslår man usikkerheten hvis man ikke tar hensyn til at det er imputert. (Med multipl imputering kunne vi tatt høyde for dette).

4.3 Imputeringsresultat

For å se effekten av imputeringen har vi estimert andelen personer som jobber overtid i analysepopulasjonen. Vi har estimert både med og uten imputering. Når vi har estimert uten imputering, har vi estimert en gang med kun de direkte intervjuene, og en gang med alle intervjuene (med rapporterte verdier for de indirekte).

I tabell 5 ser vi estimatene for 2012, når vi har aggregert over hele analysepopulasjonen. Estimaten med imputerte verdier avviker ikke mye fra estimatene med kun direkte intervju. Estimaten med imputerte verdier ligger hele tiden mellom estimatene med kun direkte intervju og estimatene med alle intervjuene, og det er estimatene med kun direkte intervju som er størst. Men på lavere aggregeringsnivå gjelder dette ikke alltid. For eksempel er rekkefølgen på estimatene snudd om for kvinner som er 25-34 år. (Se tabell 6).

Tabell 5. Estimert andel som har jobbet overtid (%), per måned 2012

	Bruker kun direkte intervju	Bruker alle intervjuene, med imputerte verdier for de indirekte og det partielle frafallet	Bruker alle intervjuene, med rapporterte verdier for de indirekte
Januar	12,34	12,19	11,94
Februar	12,13	12,09	11,76
Mars	11,78	11,69	11,34
April	10,19	10,09	9,70
Mai	11,38	11,30	10,98
Juni	14,98	14,92	14,51
Juli	12,52	12,48	11,77
August	12,06	12,00	11,46
September	14,63	14,49	13,59
Oktober	13,91	13,74	13,15
November	12,82	12,72	12,36
Desember	11,58	11,43	11,21

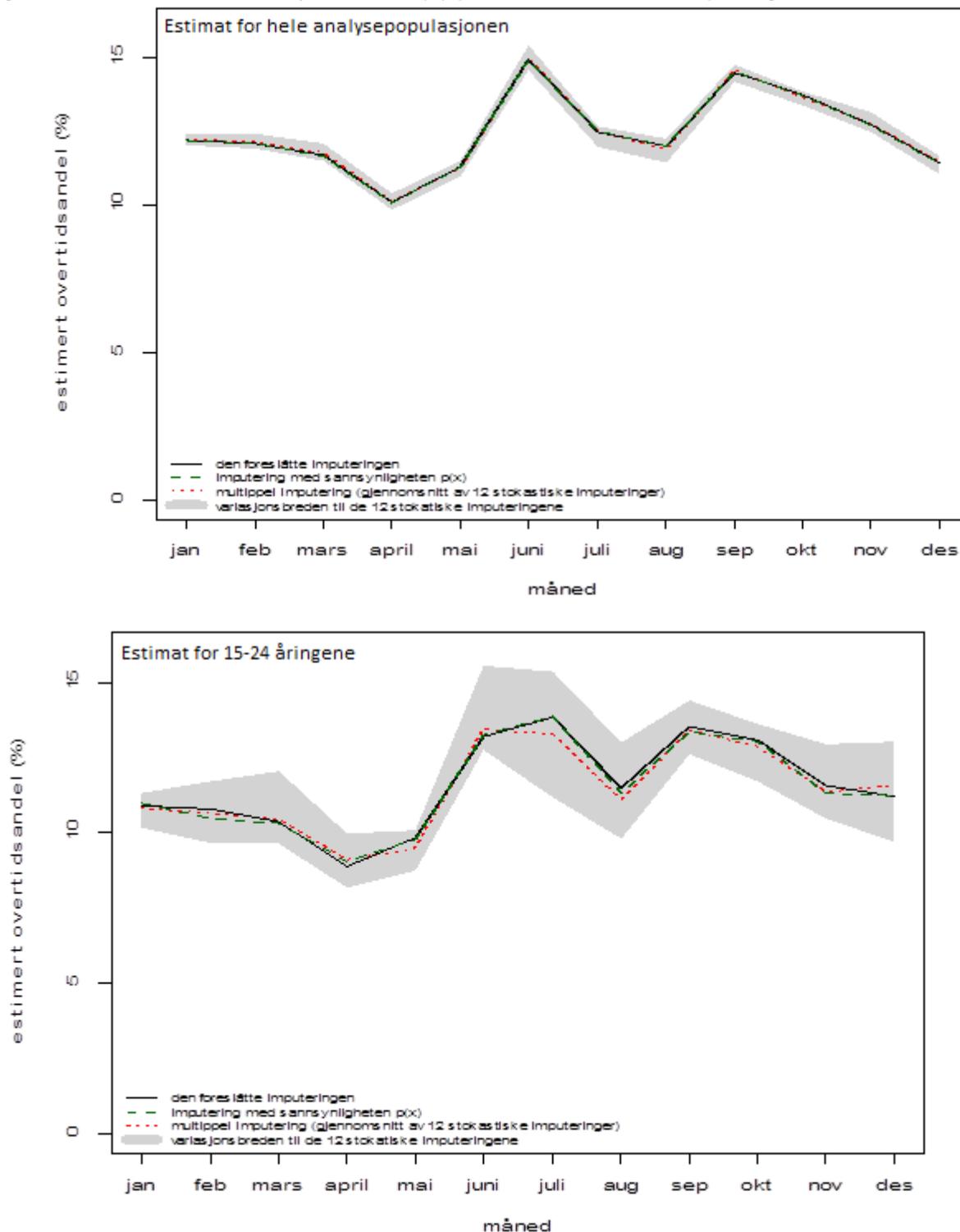
Tabell 6. Estimert andel som har jobbet overtid (%), januar 2012, fordelt etter kjønn og alder

	Bruker kun direkte intervju	Bruker alle intervjuene, med imputerte verdier for de indirekte og det partielle frafallet	Bruker alle intervjuene, med rapporterte verdier for de indirekte
Menn			
15-24 år	15,04	14,52	13,35
25-34 år	17,11	17,04	16,71
35-54 år	16,81	16,77	16,32
55-74 år	9,70	9,58	9,63
Kvinner			
15-24 år	7,62	7,19	7,25
25-34 år	8,46	8,60	8,69
35-54 år	10,04	9,98	9,85
55-74 år	7,56	7,43	7,31

I figur 1 kan vi sammenligne den foreslåtte imputeringsmetoden med metodene der vi imputerer med sannsynligheten $p(\underline{x})$, og der vi bruker stokastisk imputering (trekking fra en Bernoullifordeling) kombinert med multipl imputering. Den heltrukne svarte linjen er estimatene vi får med den foreslåtte metoden. Den grønne stiplede linjen er estimatene vi får med $p(\underline{x})$ -imputering. Den prikkede røde linjen er estimatene vi får med multipl imputering med 12 stokastiske imputeringer. Det går feltet markerer variasjonsområdet til estimatene for de 12 imputeringene.

Av den øverste figuren ser vi at på populasjonsnivå gir de tre metodene veldig like tall, og variasjonsbredden til de 12 stokastiske imputeringene er ikke så stor (med flere imputeringer ville variasjonsbredden økt). Ser vi på den nederste figuren, som gjelder 15-24 åringene, er variasjonsbredden veldig stor. Av imputeringsmetodene er det hovedsaklig den multiple imputeringen som skiller seg litt fra de andre. (Med flere imputeringer ville multippel imputering gitt samme resultat som $p(\underline{x})$ -imputeringen).

Figur 1. Estimert andel som har jobbet overtid (%), per måned 2012, med ulike imputeringsmetoder

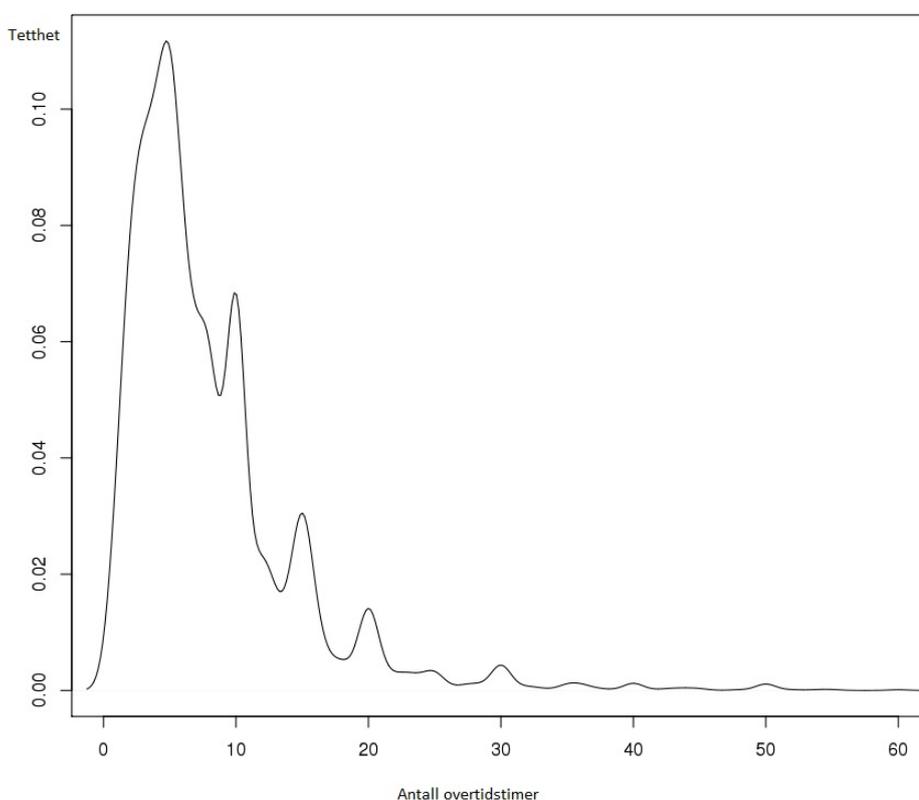


5 Analyse av overtidstimer

5.1 Overtidstimer i AKU

Overtidstimer er registrert for intervjuobjekt (IO) som jobber overtid i referanseuken i AKU. Variabelen *overt_tot_tim* er brukt for antall overtidstimer i hovedarbeidet i referanseuken. Variabelen *overt_tot_tim* strekker seg fra 1 til 86 timer. Figuren nedenfor viser at IO svarer oftere med 5-ganger tall (5, 10, 15 ... timer) enn andre. Dette indikerer at IO avrunder antall overtidstimer når de svare i AKU, eller at de avrunder overtidstimer når de registrer dem på arbeidsplassen.

Figur 2. Tetthetsdiagram av antall overtidstimer per uke for de som jobbet overtid. 2012



5.2 Sammenligning av overtidstimer for de med direkte og indirekte intervju

Overtidstimer er gjennomsnittlig høyere for de som svarer indirekte enn de svarene som innhentes med direkte intervju. Tabellen nedenfor viser vektet gjennomsnitt for hvert kvartal i 2009-2012. Forskjellen mellom direkte og indirekte intervju kan være på grunn av målefeil som henger med indirekte intervju, eller skjevhet i de som svare indirekte. For de som ble imputert at de jobbet overtid (*overtid_imp=1*), skal vi også imputere antall overtidstimer.

Tabell 7. Vektet ukegjennomsnitt av antall overtidstimer, fordelt etter type intervju og kvartal

	Direkte	(n)	Indirekte	(n)	Direkte + Indirekte
2009					
1. kvartal	8,05	1 285	9,56	92	8,15
2. kvartal	8,07	1 205	9,67	90	8,18
3. kvartal	8,55	959	9,22	76	8,60
4. kvartal	8,42	1 091	8,65	70	8,43
2010					
1. kvartal	7,93	1 050	9,12	80	8,01
2. kvartal	8,07	1 047	8,70	64	8,10
3. kvartal	8,41	871	11,20	67	8,61
4. kvartal	7,81	1 076	8,61	88	7,87
2011					
1. kvartal	7,73	1 136	7,56	92	7,71
2. kvartal	8,18	991	8,15	59	8,18
3. kvartal	7,79	910	11,17	64	8,00
4. kvartal	8,36	1 242	8,01	57	8,35
2012					
1. kvartal	7,58	1 211	8,88	61	7,64
2. kvartal	7,62	1 173	7,93	81	7,64
3. kvartal	7,94	1 043	9,55	82	8,05
4. kvartal	7,86	1 169	8,61	101	7,92

5.3 Imputeringsmetode

Imputering av overtidstimer ble gjort for de med imputert verdi for overtid, og for de som svarte direkte men mangler eller har feilkodet antall overtidstimer. Feilkoding for antall overtidstimer er registrert som "998" i dataene. To imputeringsmetoder er forsøkt her: residual regresjons-imputering og en nærmeste nabo imputering ved *predictive mean matching*. Først beskrives variabelseleksjonen.

5.3.1. Variabelseleksjon

Begge imputeringsmetodene bruker forklaringsvariabler i en modell. Derfor er begge metodene avhengig av å finne de beste variablene som forklarer dataene. Variablene som blir vurdert til å inkluderes i en modell er: kjønn, alder (numerisk), aldersgrupper (4-nivå), yrke (3- og 10-nivå), utdanning (3-nivå), heltid/deltidsarbeid, lederstilling (ja/nei), sivilstatus (gift/ikke gift) og landsdel. Regresjonsmodeller ble kjørt med alle hovedeffekter (byttet mellom numerisk og gruppert alder og 3 og 10-nivå for yrke). For å se på variabelseleksjonen ble dataene modellert årlig, men for imputering ble det månedsvis. Vi brukte *backward AIC selection* for å finne en modell som passer data. Modellene ble kjørt i R 3.0.1. Variablene som var konsekvent i de beste modellene i årene 2009-2012 var:

- kjønn
- heltid/deltid
- yrke
- lederstilling

Tabell 8. Vektet ukegjennomsnitt av antall overtidstimer, fordelt etter kjønn, type intervju og måned. Antall observasjoner (n). 2012

	Menn				Kvinner			
	Direkte	(n)	Indirekte	(n)	Direkte	(n)	Indirekte	(n)
Januar	8,75	246	10,56	11	6,78	139	6,99	9
Februar	7,54	267	11,03	17	6,51	129	7,20	5
Mars	7,82	271	8,76	12	6,74	159	4,61	7
April	7,94	205	7,30	9	8,35	104	3,98	3
Mai	8,38	310	7,75	26	6,42	153	11,95	5
Juni	7,81	251	7,52	25	6,08	150	9,23	13
Juli	8,47	144	18,29	7	8,59	80	7,12	12
August	7,94	263	9,28	26	6,63	136	6,19	11
September	8,95	265	9,13	17	6,18	155	11,32	9
Oktober	8,59	266	7,32	16	6,23	149	8,02	11
November	8,16	299	8,16	29	6,83	179	7,19	12
Desember	8,73	183	11,40	23	7,35	92	7,39	10

Heltid/deltidsarbeid (heldelt2)

Overtidstimer varierer med hel/deltidsarbeid. De som jobber deltid, jobber gjennomsnittlig flere timer overtid i 9 av månedene i 2012 (direkte intervju). Tabell 9 viser forskjell etter måned i 2012.

Tabell 9- Vektet ukegjennomsnitt av overtidstimer, fordelt etter heltid/deltidsarbeid, type intervju og måned. Antall observasjoner (n). 2012

	Heltid				Deltid			
	Direkte	(n)	Indirekte	(n)	Direkte	(n)	Indirekte	(n)
Januar	8,14	340	9,16	16	7,47	45	7,80	4
Februar	7,20	347	11,49	18	7,36	49	4,09	4
Mars	7,43	374	6,80	15	7,50	56	9,03	4
April	7,54	270	6,69	11	11,57	39	4,00	1
Mai	7,95	419	8,88	26	5,99	44	5,44	5
Juni	7,41	349	8,52	35	5,81	52	4,64	3
Juli	7,79	186	12,60	14	12,30	38	6,59	5
August	7,29	351	7,86	33	9,00	48	11,14	4
September	7,90	365	9,65	20	8,48	55	10,47	6
Oktober	7,77	370	7,39	26	7,79	45	12,00	1
November	7,50	420	7,75	37	8,87	58	8,96	4
Desember	7,93	242	10,63	24	10,48	33	9,15	9

Lederstilling (sys22)

Variabelen *sys22* indikerer om IO jobber i en lederstilling (1=ja, 2=nei).

Overtidstimer er gjennomsnittlig lavere for de som ikke har lederstilling i 9 av månedene i 2012 (tabell 10).

Tabell 10. Vektet ukegjennomsnitt av overtidstimer, fordelt etter leder/ikke lederstilling, intervju type og måned. Antall observasjoner (n). 2012

	Leder				Ikke leder			
	Direkte	(n)	Indirekte	(n)	Direkte	(n)	Indirekte	(n)
Januar	9,13	169	10,34	8	7,26	216	7,82	12
Februar	7,15	188	14,85	7	7,29	208	8,24	15
Mars	7,78	195	9,24	7	7,15	235	6,04	12
April	8,22	142	3,00	1	7,95	167	6,76	11
Mai	8,36	204	6,86	12	7,27	259	9,38	19
Juni	7,77	187	10,22	14	6,71	214	6,92	24
Juli	7,80	73	17,35	6	8,87	151	8,42	13
August	7,86	193	9,29	11	7,18	206	7,87	26
September	8,46	181	11,64	10	7,58	239	8,77	16
Oktober	8,90	201	7,44	12	6,72	214	7,71	15
November	7,57	209	7,71	23	7,74	269	8,08	18
Desember	8,68	118	8,24	15	7,93	157	12,05	18

Yrke

Yrke er en nyttig forklaringsvariabel for overtidstimer også. Med yrkesfelt⁸ (10 grupper) som inndeling er det for få IO i gruppene til å være gjennomførbart. En tredeling (lav, middels og høy, med tanke på antall overtidstimer) kan være fornuftig. Optimal gruppering viste seg å være litt annerledes enn for variabelen *overt_tot*:

⁸ Standard for yrkesklassifisering

- Lav, ble kategorisert som yrkesfelt 2, 3 og 4
- Middels, ble kategorisert som yrkesfelt 5, 7, 8, 9
- Høy, ble kategorisert som yrkesfelt 0, 1, 6

Modellene ble testet med yrke som en ordinal (3 nivå) og en kategorisk (3 grupper) variabel. *Akaike Information Criterion* (AIC) er en måte å sammenligne forskjellige modeller for å finne den beste tilpassning av dataene. Vi brukte AIC for å se hvilken spesifisering av yrke som var beste i modellen. Ordinal spesifisering av yrkesvariabelen hadde beste tilpassning i 9 av månedene i 2012. Dette er selv om tabellen nedenfor viser at yrkesvariabelen ikke er helt ordinal i forhold til overtidstimer i 6 av månedene i 2012. Dette kan delvis forklares av at tabellen nedenfor ikke tar hensyn til andre variable eller viser hvor stor variasjonen er innen gruppene. Få observasjoner i høy-overtidsgruppen enkelte måneder kan bidra til ustabilitet i estimatene med kategorisk spesifisering. Det er ønskelig at modellen ikke forandres fra måned til måned så vi valgt å bruke ordinal yrkesgruppering for alle månedene.

Tabell 11. Vektet ukegjennomsnitt av overtidstimer, fordelt etter yrkesgrupper, intervju type og måned. Antall observasjoner (n). 2012

	Lav		Middels				Høy					
	Direkte	(n)	Indirekte	(n)	Direkte	(n)	Indirekte	(n)	Direkte	(n)	Indirekte	(n)
Januar	7,81	210	7,17	7	7,56	145	7,31	9	12,11	30	14,93	4
Februar	6,66	215	9,17	9	7,78	141	7,57	10	8,34	40	21,60	3
Mars	6,83	235	8,64	8	8,26	152	4,56	9	7,84	43	14,29	2
April	7,32	159	6,62	6	8,85	124	6,26	6	9,36	26	NA	0
Mai	7,15	246	7,83	11	7,89	164	8,39	17	9,90	53	10,44	3
Juni	6,49	207	8,02	19	8,23	157	6,59	16	6,95	37	17,12	3
Juli	8,33	100	6,08	9	8,86	112	12,61	8	6,82	12	31,68	2
August	7,40	208	5,55	15	7,70	147	9,00	18	7,23	44	14,81	4
September	7,37	235	6,54	8	8,50	153	10,39	15	9,87	32	15,19	3
Oktober	6,84	221	6,92	8	8,56	160	7,99	15	10,25	34	7,43	4
November	7,49	258	7,44	18	7,81	174	8,29	18	8,04	46	7,87	5
Desember	6,92	139	10,40	10	9,71	116	9,41	20	8,57	20	13,93	3

5.3.2. Residual regresjons-imputering

I residual regresjons-imputering, blir manglende enheter erstatt med den predikerte verdien pluss en residual som er trukket tilfeldig. Dette er en stokastisk imputeringsmetode som kan beholde fordelingen til variabelen og redusere selektiv skjevhet. Metoden har en fordel i at variansberegning er mulig ved å benytte multipl imputering [1]. En ulempe er at umulige tall kan imputeres. Det kan også være en ulempe at ved enkel-imputering, vil utfallet variere fra gang til gang.

Imputeringen gjøres som beskrevet nedenfor:

- Regresjon er tilpasset data der det er gjort direkte intervju, med oppgitt Y_i . Log-transformering brukes for å sørge for positive tall:

$$\log(y_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \varepsilon_i$$

Hvor x_{1i} er en dummyvariabel for kjønn til individ i , x_{2i} er en dummyvariabel som indikerer om individ i jobber heltid (35+ avtalt timer) eller ikke, x_{3i} er numeriske yrke til individ i (3 –nivå) og x_{4i} er en dummyvariabel som indikerer om individ i jobber i en lederstilling.

- Modellen brukes for å prediktere verdier av $\log(y_k)$ for alle k enheter som skal imputeres.
- Residualer fra modellen er trukket tilfeldig med tilbakelegging for hver enhet k .
- Hver enhet k imputeres ved predikert verdi pluss den tilfeldige residualen.

Metoden er gjennomført månedsvis.

5.3.3. Nærmeste nabo-imputering

Donorimputering betyr at de manglende enhetene blir erstatt med observerte tall fra liknende enheter (giver). Metoden har en fordel ved at kun observerte, og derfor mulige tall, blir imputert. Donorimputering er generelt mindre sårbar for feil

spesifisering av modell enn parametrisk imputering [2]. Ulempen er at beregning av varians er komplekst.

Nærmeste nabo-imputering er en spesiell type av donorimputering som bruker forklaringsvariabler for å finne giveren som ligner mest på resipienten (mottakeren). Avstanden mellom giver og resipient kan måles på forskjellige måter. Vi bruker *predictive mean matching* (ppm). Denne metoden beskrives i Little [3] og er godt dokumentert. En fordel med metoden er at variablene som forklarer målvariabelen best, har mest påvirkning på avstandsmålet. I Andridge og Littles [4] anmeldelse av donorimputering, konkluderer de med at ppm er en god måte å bruke forklaringsvariabler, som kan bidra til plausibel imputering med gyldige fordelinger og kvantil-estimer.

Her beskriver vi nærmeste nabo-imputering (NNI), hvor y_i er observert antall overtidstimer for individ i som har oppgitt tall og svart ved direkte intervju.

- Regresjonen er tilpasset dataene med direkte intervju som har oppgitt y_i verdi:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + \varepsilon_i \quad (1)$$

der x_{1i} er en dummyvariabel for kjønn til individ i , x_{2i} er en dummyvariabel som indikerer om individ i jobber heltid (35+avtalt timer) eller ikke, x_{3i} er numeriske yrke for individ i (3-nivå) og x_{4i} er en dummyvariabel som indikerer om individ i jobber i en lederstilling. Siden disse variablene stort sett er kategoriske, ligner metoden på hot-deck imputering innen grupper. Siden vi ikke ønsker at imputeringen forandres så mye hver gang man kjører imputerings-programmet, brukes en ekstra variabel, x_{5i} som er månedsvekt, for å skille mellom observasjoner og stabilisere estimatene.

- β -koeffisientene fra modellen brukes for å predikere y , både for observert (\hat{y}_i) og indirekte (\hat{y}_k), der k indikerer enheten som skal få imputert y
- Donor for enhet k er enhet d som har predikert verdi \hat{y}_d nærmest den predikerte verdien \hat{y}_k .
- Observert verdi til donoren (y_d) er imputert for y_k

Dermed minimerer imputeringsmetoden avstandsfunksjonen:

$$D(k, i) = |\hat{y}(X_k) - \hat{y}(X_i)|$$

der X_k er vektoren av forklaringsvariable for enhet k , og X_i er tilsvarende vektor for observert enhet i .

Imputeringen er gjennomført månedvis.

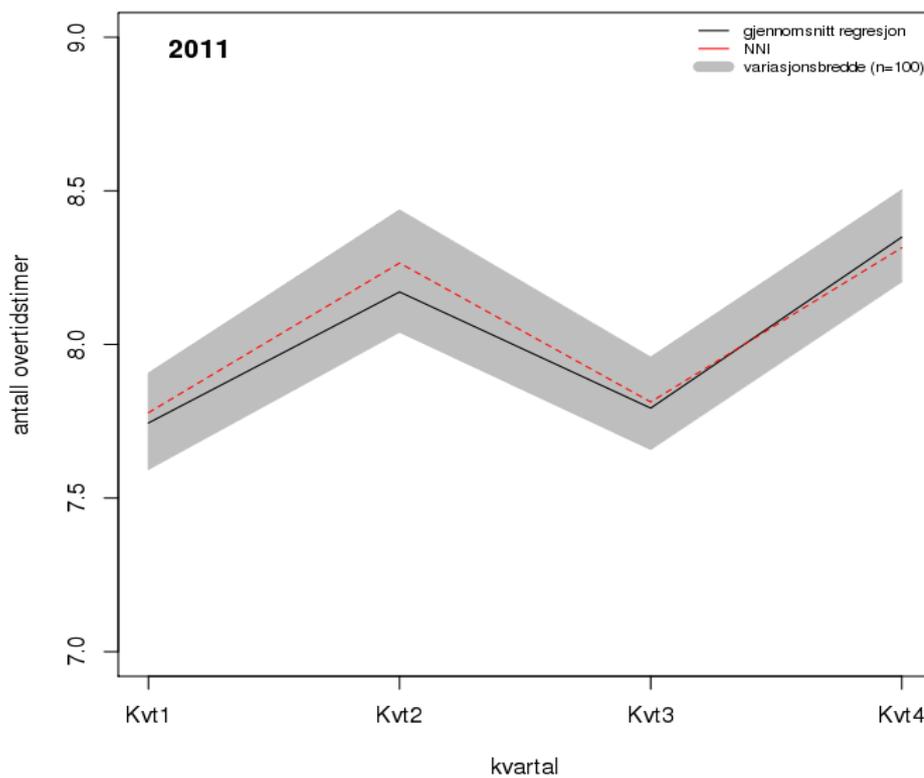
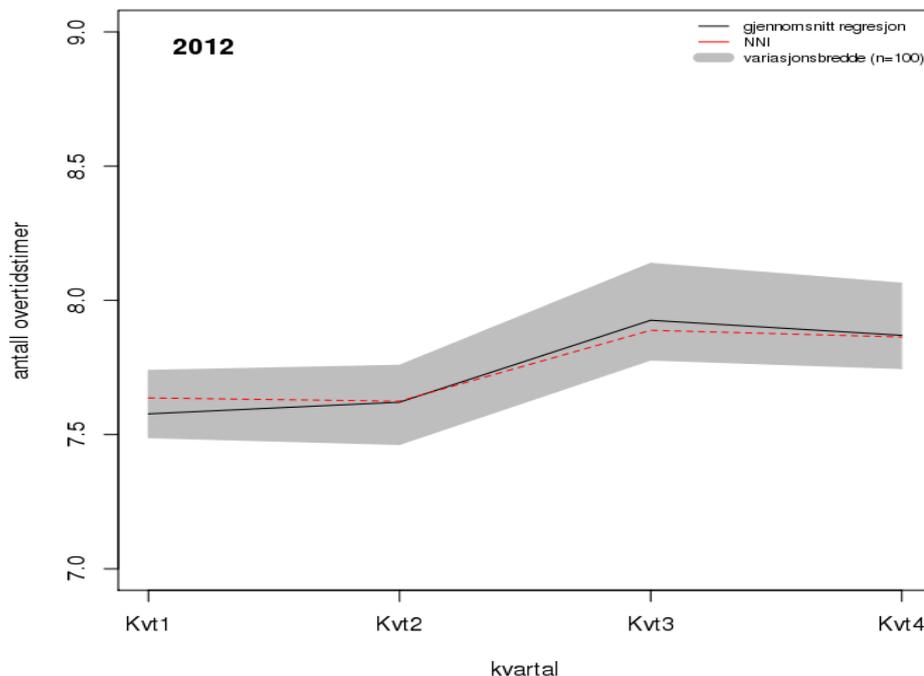
5.4 Imputeringsresultater

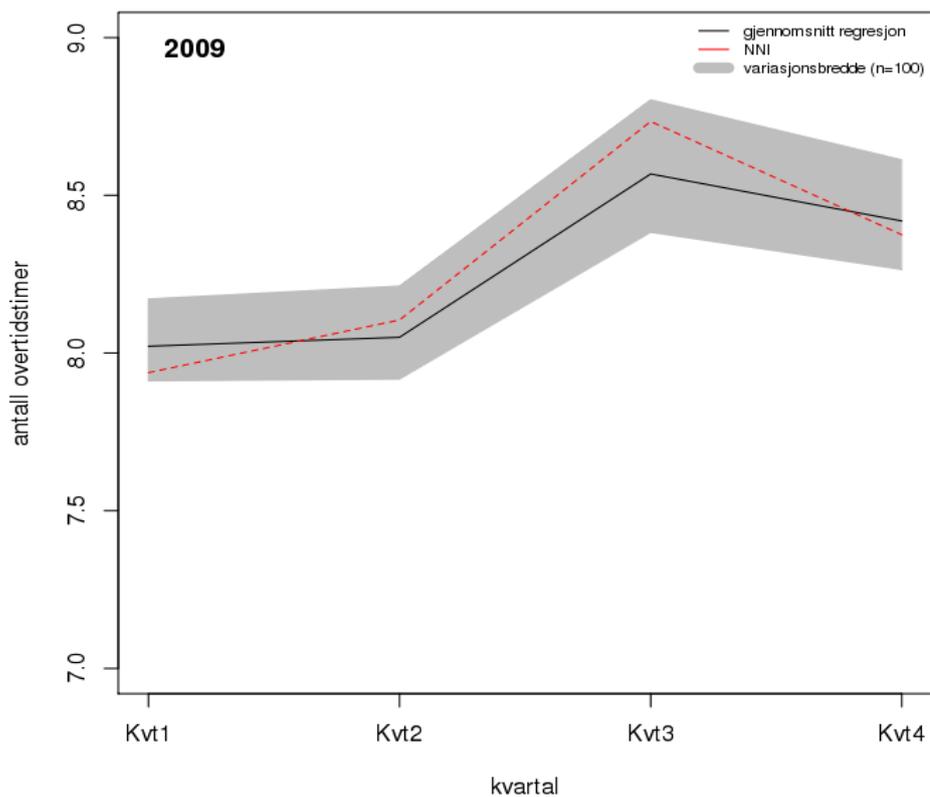
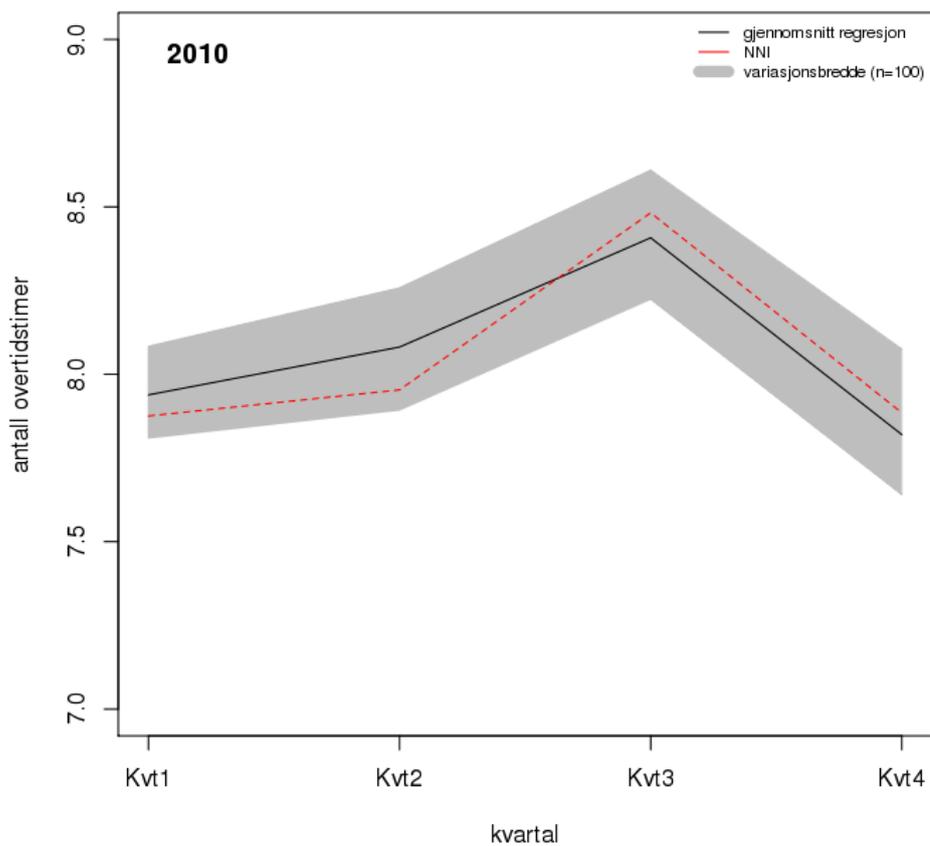
En sammenligning av de to imputeringsmetodene vises i figur 3. Det er lite forskjell mellom nærmeste nabo-imputering og gjennomsnittstall fra regresjons-imputering. Regresjonsimputeringen ble kjørt 100 ganger for å lage en variasjonsbredde. Nærmeste nabo-imputering gir tall som ligger innenfor variasjonen til regresjons-imputeringen. Figurene er skalert for å vise forskjell i imputeringsmetodene, med det resultat at overtidstallene ser ut til å variere mer enn de faktisk gjør.

Siden metoderesultatene stort sett er like, anbefaler vi nærmeste nabo-imputering som er en enklere metode å implementere. Den har en fordel fremfor residual regressions-imputering ved at den gir samme resultat hver gang imputeringen kjøres. Residual regressions-imputering produserer tall som er litt forskjellige hver gang den kjøres og kan produsere tall langt fra gjennomsnittet i visse tilfeller.

Multiple imputering er en løsning til dette, men er ikke ønskelig i produksjonsprosessen.

Figur 3. Gjennomsnittlig antall overtidstimer per uke, etter kvartal, for direkte og imputerte tall etter imputeringsmetodene. Dette er kun for de som jobbet (eller ble imputert at de jobbet) overtid. 2009-2012



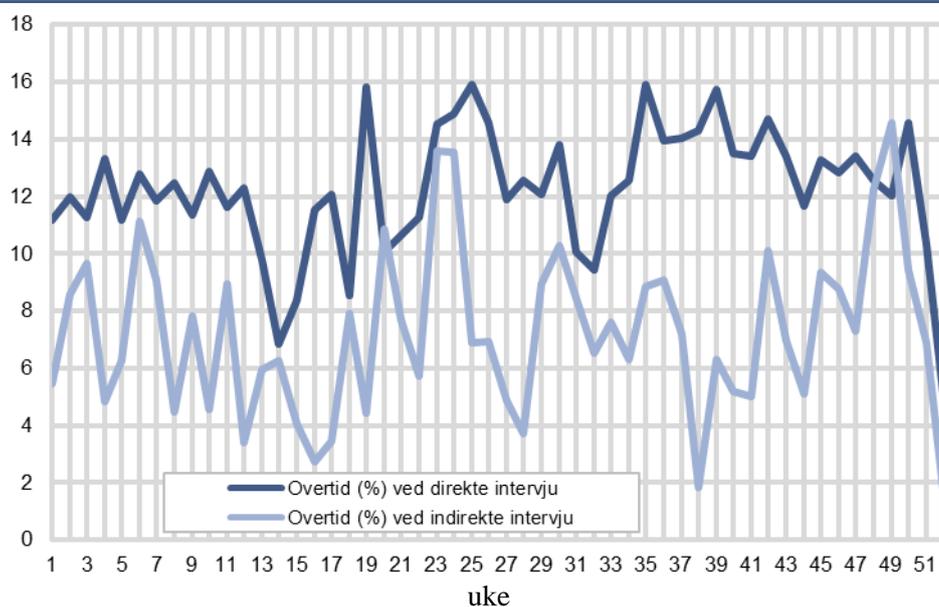


6 Konklusjon

Analysene vi har gjort i dette notatet tyder på at kvaliteten til de indirekte intervjuene ikke er så bra (når det gjelder overtid). Vi anbefaler derfor å ikke bruke rapportert overtid for de indirekte intervjuene, men i stedet imputere overtidsbruken.

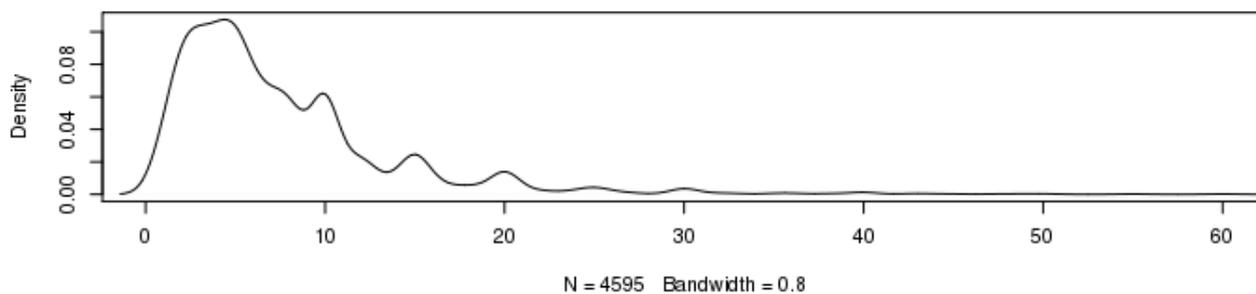
Vi har testet flere imputeringsmetoder. Av metodene vi vurderte for variabelen *overt_tot* foreslår vi å bruke den ikkestokastiske imputeringen som baserer seg på en modellbasert sannsynlighet for å jobbe overtid. Av metodene vi vurderte for variabelen *overt_tot_tim* anbefaler vi nærmeste nabo-imputeringen med predictive mean matching

Vedlegg A. Andel intervjuobjekt som har rapportert at de har jobbet overtid (%), per uke i 2012, fordelt etter direkte intervju og indirekte intervju

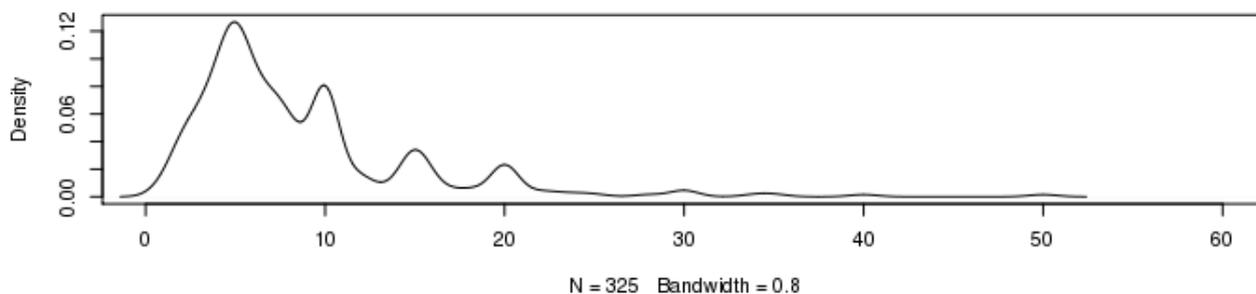


Vedlegg B. Fordeling av antall overtidstimer i 2012 med og uten nærmeste nabo-imputering

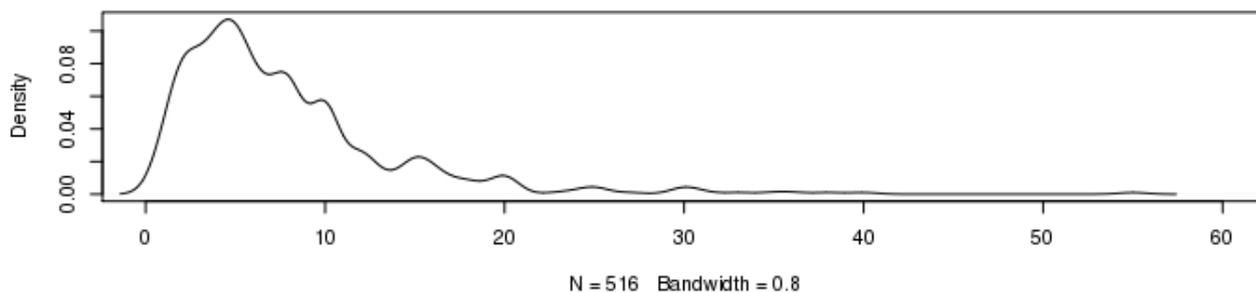
direkt



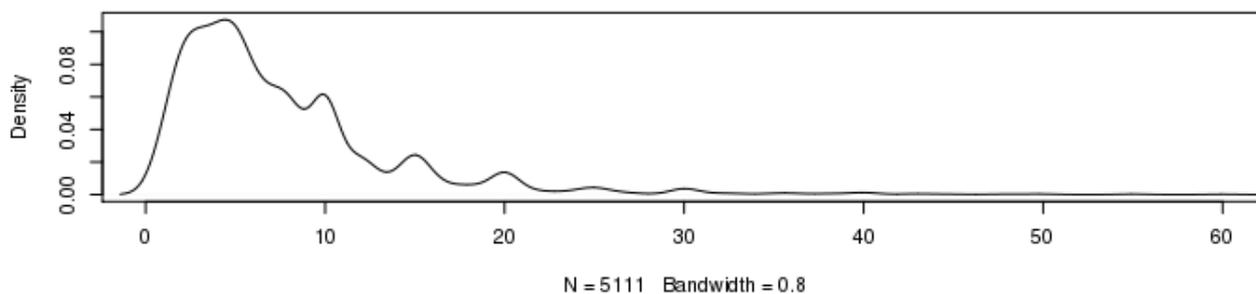
indirekt



imputert



direkt + imputert



Referanser

- Bjørnstad, J.F., *Non-Bayesian Multiple Imputation*. Journal of Official Statistics, 2007. 23(4).
- Durrant, G.B., Imputation Methods for handling Item Nonresponse in Social Sciences: A Methodological Review, in NCRM Methods Review Papers2005, ESRC National Centre for Research Methods and Southampton Statistical Sciences Research Institute (S3RI), University of Southampton.
- Little, R.J.A., Missing-Data Adjustments in Large Surveys. Journal of Business & Economic Statistics, 1988. 6(3): p. 287-296.
- Andridge, R.R. and R.J.A. Little, A review of hot deck imputation for survey non-response. International Statistical Review, 2010. 78(1): p. 40-64.
- Statistisk sentralbyrå (2001): *Arbeidskraftundersøkelsen 2001*, NOS C748.
- Bø, T.P. og Håland, I. (2002): Dokumentasjon av arbeidskraftundersøkelsen (AKU) Korrigert versjon, Notater 2002/24, Statistisk sentralbyrå.
- Heldal, J. (2006): *Logistisk regresjon - kurskompendium i byråskolens kurs SM507*, Notater 2006/54, Statistisk sentralbyrå.

Figurregister

Figur 1.	Estimert andel som har jobbet overtid (%), per måned 2012, med ulike imputeringsmetoder	15
Figur 2.	Tetthetsdiagram av antall overtidstimer per uke for de som jobbet overtid. 2012.....	16
Figur 3.	Gjennomsnittlig antall overtidstimer per uke, etter kvartal, for direkte og imputerte tall etter imputeringsmetodene. Dette er kun for de som jobbet (eller ble imputert at de jobbet) overtid. 2009-2012.....	21

Tabellregister

Tabell 1.	Andel intervjuobjekt som har jobbet overtid (%), per måned 2012, fordelt etter type intervju.....	8
Tabell 2.	Andel intervjuobjekt som har jobbet overtid (%), januar 2012, fordelt etter kjønn, alder og type intervju	8
Tabell 3.	Estimerte modellparametre (parametrisert med referansegruppe), januar 2012	10
Tabell 4.	Estimerte modellparametre (parametrisert med referansegruppe), april 2012	11
Tabell 5.	Estimert andel som har jobbet overtid (%), per måned 2012	13
Tabell 6.	Estimert andel som har jobbet overtid (%), januar 2012, fordelt etter kjønn og alder	13
Tabell 7.	Vektet ukegjennomsnitt av antall overtidstimer, fordelt etter type intervju og kvartal	17
Tabell 8.	Vektet ukegjennomsnitt av antall overtidstimer, fordelt etter kjønn, type intervju og måned. Antall observasjoner (n). 2012	18
Tabell 9-	Vektet ukegjennomsnitt av overtidstimer, fordelt etter heltid/deltidsarbeid, type intervju og måned. Antall observasjoner (n). 2012.....	18
Tabell 10.	Vektet ukegjennomsnitt av overtidstimer, fordelt etter leder/ikke lederstilling, intervju type og måned. Antall observasjoner (n). 2012.....	18
Tabell 11.	Vektet ukegjennomsnitt av overtidstimer, fordelt etter yrkesgrupper, intervju type og måned. Antall observasjoner (n). 2012.....	19

Statistisk sentralbyrå

Postadresse:
Postboks 8131 Dep
NO-0033 Oslo

Besøksadresse:
Akersveien 26, Oslo
Oterveien 23, Kongsvinger

E-post: ssb@ssb.no
Internett: www.ssb.no
Telefon: 62 88 50 00

ISBN 978-82-537-8886-9 (elektronisk)



Statistisk sentralbyrå
Statistics Norway