



Underrapportering av inntekt blant selvstendig næringsdrivende

Et utgiftsbasert estimat

Stine Opshaug Skjeggestad og Sara Wæhle

Veileder: Jarle Møen

Masterutredning i Økonomisk Styring

NORGES HANDELSHØYSKOLE

Dette selvstendige arbeidet er gjennomført som ledd i masterstudiet i økonomi- og administrasjon ved Norges Handelshøyskole og godkjent som sådan. Godkjenningen innebærer ikke at Høyskolen eller sensorer inntestår for de metoder som er anvendt, resultater som er fremkommet eller konklusjoner som er trukket i arbeidet.

Forord

Denne utredningen er gjennomført som en avsluttende del av masterstudiet i økonomi og administrasjon ved Norges Handelshøyskole, og utgjør 30 studiepoeng av vår hovedprofil, Økonomisk Styring. Utredningen er skrevet med stipend fra et samarbeid mellom Skatteetaten og Norges Handelshøyskole.

Utredningen er en empirisk studie som bruker en utgiftsbasert metode til å estimere en del av Norges svarte økonomi. Denne delen kommer fra underrapportering av inntekt blant selvstendig næringsdrivende. Temaet for oppgaven var et forslag fra vår veileder Jarle Møen, og appellerte spesielt til oss da vi fikk vite at det per dags dato ikke finnes pålitelige estimater på underrapportering av inntekt i Norge. Arbeidet med utredningen har vært meget interessant og vi har opparbeidet oss mye ny kunnskap underveis.

Vi ønsker å takke vår veileder, Jarle Møen, for stort engasjement og interesse. I tillegg vil vi takke Gry Henriksen fra Norsk samfunnsvitenskapelige datatjeneste for hjelp i forbindelse med uthenting av data. Vi vil også takke alle andre som i løpet av prosessen har gitt oss gode råd og innspill.

Bergen, mai 2015

Stine Opshaug Skjeggestad

Sara Wæhle

Sammendrag

I denne masterutredningen undersøker vi graden av underrapportering av inntekt i Norge fra 1975 til 2012 blant selvstendig næringsdrivende. Analysene bygger på en utgiftsbasert estimeringsmetode, utviklet av Pissarides og Weber (1989). Vi tar også for oss ulike faktorer som kan påvirke underrapporteringen; ulike bransjer, skattereformer, tid og antall selvstendig næringsdrivende i husholdningen.

Resultatene våre indikerer at selvstendig næringsdrivende underrapporterer om lag 14 % av sin reelle inntekt. Mye tyder også på at selvstendig næringsdrivende i bygg- og anleggsbransjen underrapporterer en større andel av sin inntekt enn andre bransjer. Innenfor denne bransjen får vi indikasjoner på at også *arbeidstakere* underrapporterer sin inntekt.

For de som arbeider innen helse/undervisning får vi resultater som tyder på at den reelle inntekten blir rapportert. Vi finner ingen statistiske bevis på underrapportering i denne gruppen.

På grunn av få årlige observasjoner og usikre estimater, kan vi ikke konkludere hvorvidt graden av underrapportering har utviklet seg over tid. Vi observerer heller ingen direkte effekt på underrapporteringen som følge av endringer i skattesystemet.

Videre finner vi ingen statistisk signifikante indikasjoner på at flere selvstendig næringsdrivende i husholdningen gir økt grad av underrapportering.

Innholdsfortegnelse

1. INNLEDNING	6
1.1 MOTIVASJON	6
1.2 FORMÅL	7
1.3 PROBLEMSTILLING	7
1.4 STRUKTUR OG OPPBYGGING	8
2. TEORETISK RAMMEVERK	9
2.1 DEFINISJON PÅ SVART ØKONOMI	9
2.2 OMFANG OG KONSEKVENSER AV SVART ØKONOMI	9
2.2.1 Norges svarte økonomi	10
2.3 UTVIKLINGEN AV DET NORSKE SKATTESYSTEMET	11
2.3.1 Skattesystemet før 1992	11
2.3.2 Skattereformen fra 1992	11
2.3.3 Skattereformen fra 2006	12
3. TIDLIGERE STUDIER.....	13
3.1 STORBRIANNIA	15
3.1.1 Pissarides og Weber (1989).....	15
3.1.2 Lyssioutou et al. (2004)	15
3.2 CANADA	16
3.2.1 Schuetze (2002).....	16
3.2.2 Fortin et al. (2010).....	16
3.3 USA	16
3.4 UGANDA	16
3.5 AUSTRALIA	17
3.6 FINLAND	17
3.7 SVERIGE.....	17
3.8 JAMAICA	18
3.9 ESTLAND.....	18
3.10 TYRKIA	19
3.11 SPANIA.....	19
4. DEN UTGIFTSBASERTE ESTIMERINGSMETODEN.....	20
4.1 ESTIMERING VED OLS.....	20
4.2 ESTIMERING VED IV	23
4.2.1 Endogenitetsproblem i inntektsvariabelen.....	23
4.2.2 Permanent inntekt	23

4.2.3 Variablene p_i og k_i	25
4.2.4 Konstantleddet i utgiftsfunksjonen.....	26
4.2.5 Variansen til inntektsresidualet	28
4.2.6 Område for gjennomsnittlig underrapportering.....	30
4.2.7 Estimat på k dersom $\rho \neq 0$	33
4.3 SVAKHETER VED METODEN.....	35
4.3.1 Homogene preferanser	35
4.3.2 Egenproduksjon	36
4.3.3 Andre svakheter	37
5. DATA.....	38
5.1 BESKRIVELSE AV DATASETET	38
5.2 RENSING AV DATASETET	39
5.3 DATABEHANDLING	40
5.3.1 Selvstendig næringsdrivende	40
5.3.2 Ulike bransjer	41
5.4 VALG AV VARIABLER	41
5.4.1 Forklaringsvariabler	42
5.4.2 Instrumentvariabler	43
5.5 BESKRIVENDE STATISTIKK	44
6. RESULTATER OG ANALYSE.....	46
6.1 ESTIMAT PÅ UNDERRAPPORTERING I ULIKE NÆRINGER.....	46
6.1.1 Alle næringer	48
6.1.2 Bygg og anlegg	51
6.1.3 Helse og undervisning	53
6.2 UTVIKLING OVER TID	54
6.3 PÅVIRKNING FRA ULIKE SKATTEREFORMER	57
6.4 ANTALL SELVSTENDIG NÆRINGSDRIVENDE I HUSHOLDNINGEN.....	60
7. AVSLUTNING	63
7.1 KONKLUSJON.....	63
7.2 SVAKHETER VED ANALYSEN	64
8. LITTERATURLISTE.....	65
9. APPENDIKS.....	72

1. Innledning

1.1 Motivasjon

En velferdsstat som Norge er avhengig av skatteinnbetalinger. Dersom det eksisterer stor grad av skatteunndragelse i samfunnet, reduserer dette spillerommet til staten og samfunnet påvirkes i negativ grad. For å kunne iverksette tiltak og begrense negative konsekvenser, er det viktig for myndighetene med mer kunnskap på området som omhandler svart økonomi. Det er samtidig utfordrende å skaffe seg en fullstendig oversikt over størrelsen på den svarte økonomien, til tross for at det finnes forskning på emnet. Dette danner noe av grunnlaget for at vi ønsker å finne estimater på en del av den svarte økonomien. Denne delen omhandler skatteunndragelse i form av underrapportering av inntekt, hovedsakelig blant selvstendig næringsdrivende.

En av årsakene til at det er spesielt interessant å ta for seg selvstendig næringsdrivende, er deres muligheter for underrapportering. Selvstendig næringsdrivende har selv ansvar for inntekts- og kostnadsføring, og må ikke gå via en tredjepart. Denne gruppen har også muligheter for å føre fradragsberettigede kostnader mot inntekten i sin bedrift.

Det er også interessant å undersøke hvorvidt ulike bransjer har ulik grad av underrapportering. Vi vil her ta for oss bygg/anlegg og helse/undervisning. For disse bransjene vil vi åpne opp for at også *arbeidstakere* (som mottar fast lønn) kan underrapportere sin inntekt. Bygg- og anleggsbransjen omfatter praktiske yrker og arbeid som enkelt kan utføres utenfor normal arbeidstid. Denne gruppen har følgelig større muligheter for å arbeide svart. Ved å ikke skattlegge arbeidet, kan kundene bli tilbudt lavere pris enn i markedet. Denne typen svart arbeid er vanskelig for myndighetene å dokumentere. Ideen om å se nærmere på nettopp denne gruppen kommer fra Thor O. Thoresen og Odd Erik Nygård (2014).

Blant de som arbeider innen helse/undervisning, antar vi at selvstendig næringsdrivende i stor grad består av leger som har sin egen praksis, og at arbeidstakere i hovedsak er offentlig ansatte. Datasettet vi benytter tillater oss ikke å få en oversikt over hvem som er leger/offentlig ansatte for alle år. Likevel antar vi at denne gruppen har mindre mulighet til å underrapportere sin inntekt, sammenlignet med andre grupper. Vår veileder, Jarle Møen, har vært pådriveren for å undersøke denne effekten.

1.2 Formål

I denne utredningen vil vi bruke en utgiftsbasert estimeringsmetode utviklet av Christopher A. Pissarides og Guglielmo Weber (1989), som diskuteres nærmere i kapittel 4. Metoden tar utgangspunkt i at selvstendig næringsdrivende underrapporterer sin inntekt. Hovedpoenget er å finne et estimat, k , som skal multipliseres med selvstendig næringsdrivendes rapporterte inntekt for å komme frem til deres ”sanne” inntekt. Vi vil benytte oss av data om matutgifter fra Forbruksundersøkelsen for årene 1975-2012 for å estimere denne k 'en. På denne måten kan vi estimere den delen av svart økonomi i Norge som skyldes underrapportering av inntekt blant selvstendig næringsdrivende.

I hovedsak ønsker vi å finne graden av underrapportering blant selvstendig næringsdrivende i alle næringer, men vil også undersøke hvorvidt denne varierer mellom ulike bransjer, og om den utvikler seg over tid. Vi vil i tillegg åpne opp for at også *arbeidstakere* kan underrapportere sin inntekt i noen deler av analysen.

1.3 Problemstilling

Den overordnede problemstillingen for denne masterutredningen er:

I hvilken grad eksisterer det en svart økonomi i Norge som skyldes underrapportering av inntekt blant selvstendig næringsdrivende?

Videre vil vi også drøfte følgende underproblemstillinger:

1. Er det større grad av underrapportering av inntekt blant de som arbeider med bygg/anlegg, og mindre grad blant de som arbeider med helse/undervisning?
2. Har underrapporteringen utviklet seg signifikant over tid fra 1975 til 2012?
3. Finnes det indikasjoner på at ulike skattereformer har hatt en effekt på utviklingen av underrapporteringen?
4. Har *antall* selvstendig næringsdrivende i en husholdning innvirkning på graden av underrapportering?

1.4 Struktur og oppbygging

Kapittel 1 består av denne innledningen. I kapittel 2 tar vi for oss omfang og konsekvenser av en svart økonomi, og gjennomgår det norske skattesystemet. I kapittel 3 presenteres tidligere studier som har benyttet samme utgiftsbaserte estimeringsmetode som denne utredningen.

Kapittel 4 gir en detaljert gjennomgang av metoden vi benytter. I kapittel 5 presenterer vi datasettet og forklarer hvordan vi har preparert dette. Videre presenteres resultater og analyse i kapittel 6. Kapittel 7 inneholder konklusjonen vår, og her blir det også pekt på svakheter ved analysen.

2. Teoretisk rammeverk

I dette kapitlet vil vi først gi en innføring i begrepet ”svart økonomi”, dets omfang og konsekvenser. Deretter presenterer vi kort noen tidligere studier som har tatt for seg Norges svarte økonomi. Til slutt følger en oversikt over viktige poeng ved skattesystemet i Norge for perioden 1975-2012.

2.1 Definisjon på svart økonomi

Det finnes mange ulike definisjoner på svart økonomi. Eide (2000, s.14) definerer svart økonomi som: ”...all i og for seg lovlig aktivitet som resulterer i transaksjoner som er skjult for myndighetene, først og fremst for skattemyndighetene. Aktivitetene i den svarte økonomi utføres først og fremst for å unngå skatter og avgifter...”. I denne utredningen kan svart økonomi både bestå av lovlige og ulovlige transaksjoner som generer inntekt.

Vi vil kun ta for oss den delen av svart økonomi som skyldes underreportering av inntekt. Dette betyr at vi ikke finner et estimat på den totale svarte økonomien i Norge, men en avgrenset del av den.

2.2 Omfang og konsekvenser av svart økonomi

Her vil vi ta for oss omfang og konsekvenser av den svarte økonomien i Norge. For å gi en god oversikt, presenterer vi til slutt noen utvalgte studier som også har hatt fokus på dette temaet.

Det er vanskelig å få oversikt over den svarte økonomien i Norge ettersom skatteunndragelse er en aktivitet aktørene ønsker å holde skjult. Vi antar at det eksisterer en svart økonomi, men vet ikke i hvilken grad den påvirker samfunnet.

I et høyskatteland som Norge går store deler av befolkningens inntekt til skatter. Det er dermed mulig å ”tjene” mye ved å ikke betale skatt. Skatteunndragelse kan gi store konsekvenser for en velferdsstat som Norge ved at betydelige beløp ikke kommer fellesskapet til gode (ØKOKRIM, 2012). I tillegg kan konkurransen i markedet komme i ubalanse. Dersom man ikke skatter av sin inntekt, kan man sette en lavere pris på arbeidet som

genererer inntekten. Dette fører til bedre konkurransevilkår for de som arbeider svart, og et skjevt marked.

Til tross for store tap for staten, vil skatteunndragelse også kunne ha en positiv effekt for et land. Dette kommer av at det oppstår større verdiskaping i samfunnet når enkelte unngår skattlegging på sitt arbeid, da kjøp av ”svarte” tjenester blir billigere enn i det lovlige markedet.

I følge ØKOKRIM (2012) er underrapportering av inntekt den vanligste formen for skatteunndragelse i Norge. Følgende bransjer hevdes å være særlig utsatte for denne typen skatteunndragelse: tjenesteytende bransjer, bransjer som har stor grad av ufaglært arbeidskraft eller kontantbruk, og bransjer med store og udefinerbare varepartier.

2.2.1 Norges svarte økonomi

For å få en bedre forståelse for Norges svarte økonomi, har vi her tatt for oss noen tidligere studier om dette emnet.

Ved bruk av spørreundersøkelser anslo Isachsen og Strøm (1985) at Norges svarte økonomi utgjorde 4-6 % av BNP på starten av 80-tallet. Schneider (2002) kom frem til at den svarte økonomien hadde økt fra om lag 14,8 % i 1989/1990 til 19 % av BNP i 2001/2002.

Hovedårsaken til veksten i svart økonomi var økt skattetrykk og trygdeavgifter, samt økte statlige reguleringsaktiviteter.

I sin studie fra 2002 kom Goldstein et al. frem til at samlede svarte inntekter i Norge omtrent var uendret fra 1980 til 2001, men at andelen som jobbet svart hadde minket. Noe av grunnlaget for dette var lavere skatt på arbeids- og kapitalinntekt og større grad av automatisk rapportering. Økt yrkesdeltakelse blant kvinner førte til at færre var interessert i strøjobber i det svarte markedet. I tillegg til dette førte flere høyt utdannede til flere jobber med liten mulighet for svart arbeid. Eldre arbeidsstyrke, større tro på kontrollorganene, og en viss bedring i skattemoralen var også viktige årsaker. En svakhet med studien var at resultatene kom fra spørreundersøkelser, noe som kan ha ført til at respondentene utelot viktig informasjon i frykt for at hensynet til anonymitet ikke skulle bli godt nok ivaretatt.

2.3 Utviklingen av det norske skattesystemet

Her vil vi ta for oss hovedmomentene ved skattesystemet i perioden 1975-2012. Ut fra dette kan vi senere drøfte om endringer i skattesystemet har hatt en direkte effekt på utviklingen av underrapportering av inntekt.

2.3.1 Skattesystemet før 1992

Før 1992 besto skattesystemet i Norge generelt av høye og formelle skattesatser, kombinert med skatteordninger som gav skatteutsettelse. Den reelle beskatningen var ikke forenlig med de formelle skattesatsene. Det var mange muligheter for å unngå (deler av) skattebetalingen. Man kan si at skattesystemet før 1992 oppmuntret til skatteplanlegging og disposisjoner som ikke var samfunnsøkonomisk lønnsomme, samt at den omfordelende virkningen ble svekket av de mange fradragsordningene. (Gabrielsen, 1992).

På 1970-tallet var det en skatteomlegging som reduserte de tidligere progressive satsene på inntekts- og formuesskatter i Norge. Frem mot 1988 var det vekselvis heving og senkning av satsene, før man fikk en markert svakere progresjon i satsene for inntektsskatten. (Gabrielsen, 1992).

2.3.2 Skattereformen fra 1992

I følge Gabrielsen (1992) var hovedformålet med den nye skattereformen i 1992 å forenkle skattesystemet, samt å sikre en mer rettferdig fordeling av skattebyrden. Omleggingene gikk i retning av mer generelle og brede skattegrunnlag, samtidig som man dempet progresjonen i skattesatsene (NOU 1999:7, 1999). De marginale skattesatsene ble senket relativt mye, noe som ble oppveid med en skatt på 28 % på all kapitalinntekt, både til bedrift og privatperson. Dette ledet til uttrykket ”den duale inntektsskatten”, der skattesystemet ble delt opp i to; skattlegging på 28 % av all kapitalinntekt og progressiv skattlegging av inntektsskatt (Schjelderup, 2014).

Hovedtanken bak den lineære skatten på 28 % var i følge Christiansen (2004) å unngå skattefavorisering av noen typer investeringer og organisasjonsformer. Skattesatsene ble altså lavere for alminnelig inntekt, men skulle dekke en større bredde av andre former for inntekt.

Delingsmodellen var et annet sentralt aspekt ved denne reformen. Den gikk ut på at arbeidsinntekten til aktive bedriftseiere ble beregnet av deres arbeid i bedriften, basert på kapitalens avkastning. Definisjonen på aktive bedriftseiere var de som eide 2/3 eller mer av bedriften. Hovedårsaken til dette var å sikre lik fordeling, det vil si at bedriftseiere ikke kunne ende opp med å kun betale 28 % skatt av sin inntekt. Problemer oppstod likevel da aktive bedriftseiere solgte seg ned til en eierandel under 2/3 og dermed oppnådde en skatteprosent på 28 % av hele sin inntekt. Dette medførte at ulikhetene i samfunnet økte. (Gabrielsen, 1992).

2.3.3 Skattereformen fra 2006

I følge Finansdepartementet (2011) var formålet med skattereformen fra 2006 å fjerne muligheten til å få skattlagt arbeidsinntekten som utbytte og generelt fjerne skjevheter. Fra 1992 var den maksimale inntektsskatten på 64,7 %, mens utbytte ble skattlagt flatt med 28 %. Dette ble nå jevnet ut slik at høyeste marginalsatt på lønnsinntekt ble satt til 54,3 %. Aksjonærmodellen, som ble innført, førte til at marginalsatt på utbytte kunne bli skattlagt med maksimalt 48,16 % ($1 - 0,72^2$). Det ble også innført et skjermingsfradrag i utbytteskatten. Her ble aksjeinntekter, aksjeutbytte og aksjegevinster, utover en relativt begrenset skattefri avkastning, skattepliktig for personlige skattytere. Skjermingsfradraget tilsvarte risikofri avkastning. Poenget med dette var å forhindre at utbytteskatten svekket bedrifters tilgang på ny, norsk egenkapital. I tillegg til dette ønsket man å tilrettelegge for investeringer og oppstart av ny næringsvirksomhet i Norge. (Meld. St. 11, 2011).

Foretaksmodellen og delingsmodellen sto også sentralt i reformen. Foretaksmodellen gjaldt for enkeltpersonforetak. Den innebar et tillegg til virksomhetens alminnelige inntekt som ble skattlagt løpende på eierens hånd, herav kapitalinntekter og –utgifter fratrukket skjermings- og lønnsfradrag. Deltakermodellen ble innført for beskatning av deltakerlignende selskaper, samtidig som delingsmodellen ble opphevet for aktive eiere. Disse modellene, sammen med aksjonærmodellen, skulle sikre en høy marginalsatt på eierinntekt uavhengig av om inntekten ble opptjent i et aksjeselskap, deltakerlignende selskap eller et enkeltpersonforetak. Skatten ble mindre avhengig av om en var ansatt eller selvstendig næringsdrivende, og det ble en mer jevn beskatning av alle inntektshavere. Skattleggingen av selvstendig næringsdrivende ble lagt så tett som mulig opp til skatteleggingen av aksjonærer. (Meld. St. 11, 2011).

3. Tidligere studier

I dette kapitlet vil vi presentere noen resultater fra forskning i ulike land som tar i bruk samme utgiftsbaserte metode som denne utredningen. En av grunnene for dette er at det kan ligge gode drøftelser og metodiske tilleggsmomenter til den opprinnelige studien som kan være nyttig å bemerke, og som kan være nyttig for fremtidige oppfølgingsstudier.

Forskningsfeltet på svart økonomi er aktivt og levende, og det finnes en relativt omfattende litteratur om emnet. Tabellen nedenfor gir en oversikt over utvalgt litteratur fra ulike land, og under tabellen vil det komme noe mer informasjon om hver av studiene. I tabellen presenteres blant annet estimatet på k (tallet vi multipliserer selvstendig næringsdrivendes rapporterte inntekt med for å få deres ”sanne” inntekt) fra hver studie. Vi vil ikke sammenligne estimatene med hverandre eller med resultatene fra denne utredningen. Dette kommer blant annet av at tidsperioder og valg av variabler vil variere på tvers av land og studier. I tillegg til dette finnes det ingen fellesdefinisjon på begrepet ”svart økonomi” (Pedersen, 1998). Noen betegner for eksempel ”underrapportering av inntekt” som hele landets svarte økonomi, og beregner følgelig den svarte økonomien i prosentandel av BNP direkte fra k . I denne utredningen representerer ”underrapportering av inntekt” kun en avgrenset del av Norges svarte økonomi.

Vi gjengir ikke koeffisienter fra regresjonsresultatene i de ulike studiene. Dette kommer av det ofte blir presentert flere ulike resultater, og det ikke alltid er like klart hvilke tall forfatterne benytter for å kalkulere k .

Tabell 1: *Oversikt over utvalgt litteratur fra andre land*

Studie	Land	Periode	Ant. obs.	Ant. selvst. næringsdr.	k	Instrumentvariabler	Merknad
Pissarides og Weber (1989)	Storbrit.	1982	2 208	196	1,55	Noe uklart. Diskuteres i kap. 5.4.2.	
Schuetze (2002)	Canada	1969-1992	8 463	682	1,12-1,3	Utdanningsnivå, konens arbeidsintensitet, og interaksjoner med selvst. næringsdr. og ant. barn, konens arbeidsintensitet og ant. rom i bolig.	
Hurst et al. (2010)	USA	1980-2003	27 219	2 508	1,43	Dummy-variabler for utdannelse	
Obwona (1999)	Uganda	1992-1993	9 273	6 543	1,002*	-	*dvs. ingen underrapportering
Lyssioutou et al. (2004)	Storbrit.	1993	1 750	-	1,28*	Ant. år med utdannelse for begge hovedpersonene, og dette kvadrert, dummy for hver av ektefellene i ekstrajobb som selvst. næringsdr., og som arbeidstaker, interaksjonsvariabler med hovedperson som selvst. næringsdr. og alle andre variabler benyttet.	*Utregningen utdypes i kap. 3.1.2.
Bradbury (1997)	Australia	1993-1994	-	-	1,34	-	
Johansson (2005)	Finland	1994-1996	2 053	285*	1,165-1,42	Huseierskap, ant. rom i bolig, ant. mnd. ektefelle har arbeidet, ant. mnd. ektefelle har arbeidet i selvst. næringsvirksomhet, ant. mnd. ektefelle har vært arbeidsløs, ant. mnd. ektefelle har hatt deltidsjobb, og produktet av dummy for selvst. næringsdr. med alder, alder kvadrert, barn, barn kvadrert, ant. barn under 7 år, ant. mnd. ektefelle har arbeidet i eget hus og dummy for region.	*Kun hovedinntektstaker er selvst. næringsdr.
Fortin et al. (2010)	Quebec, Canada	1997-2002	6 451	631	-	Alder hovedinntektstaker, alder kvadrert, ekteskapeleg status, kjønn, ant. barn mellom 0 og 15, bosted, interaksjoner; alder og bosted, ekteskapeleg status og ant. barn mellom 0 og 15, alder og ekteskapeleg status.	Studien presenterer kun andel av BNP, se kap. 3.2.2
Engström og Holmlund (2009)	Sverige	1999-2004	6 004*	804*	1,3	Kapitalinntekt og eiendomsskatt.	*Minst én er selvst. næringsdr. eller arbeidst.
Torero et al. (2006)	Jamaica	2001	1 025	-	1,32	-	
Kukk og Staehr (2014)	Estland	2002-2007	6 016	528*	1,39	Utdanningsnivå, kjønn og nasjonalitet, dummy for region	*Basert på identifisering ved rapportert ansettelsesforhold
Davutyan (2008)	Tyrkia	2004-2005	17 103	-	1,25	Alder hovedinntektstaker, alder kvadrert, kjønn, betalt leie, dummy-var. for huseierskap, dummy-var. for urbant bosted	
Martinez-Lopez (2013)	Spania	2006-2009	16 451	-	1,25*	Dummy for hovedinntektstakers nasjonalitet og for utdanningsnivå (to ulike nivåer), interaksjon mellom selvst. næringsdr. og alder, alder kvadrert, og utdanningsvariablene.	*Arbeidstaker kan også underrapportere

Der instrumentvariabler ikke er inkludert, antar vi at studien presenterer OLS-estimer. For studiene Schuetze (2002), Hurst et al. (2010) og Obwona (1999) har vi selv estimert k ut fra oppgitt prosentvis underrapportering, gitt ved formelen: $[k = 1/(1 - P)]$, hvor $P =$ prosentvis underrapportering.

3.1 Storbritannia

3.1.1 Pissarides og Weber (1989)

Pissarides og Weber (1989) estimerte k til å bli 1,55, noe som tilsvarte 5,5 % av Storbritannias BNP i 1982. I denne studien skilte de mellom selvstendig næringsdrivende i ”blue-collar”, som typisk utførte manuelt eller fysisk arbeid, og i ”white-collar”, som typisk hadde et profesjonelt og administrativt arbeid. De fant beviser på at selvstendig næringsdrivende i ”blue-collar” underrapporterte sin inntekt i større grad enn de i ”white-collar”, hvor anslaget var på henholdsvis 60 % og 50 %.

3.1.2 Lyssiotou et al. (2004)

I sin studie modifiserte Lyssiotou et al. (2004) den opprinnelige metoden til Pissarides og Weber (1989) til å bli en ikke-parametrisk metode. Dette gjorde de ved å ta hensyn til at inntekt fra ulike inntektskilder kunne brukes til ulike formål, det vil si at husholdningene hadde heterogene preferanser. De skilte her mellom (uforutsigbar) inntekt fra selvstendig næringsvirksomhet og (forutsigbar) inntekt som kom fra fast lønn til arbeidstakere. På denne måten unngikk de potensielle skjevheter som oppsto ved at heterogene preferanser ble tolket som en inntektseffekt, når det egentlig var en substitusjonseffekt.

I denne studien ble både Pissarides og Weber (1989) sin opprinnelige metode og den ikke-parametriske metoden benyttet. Når de benyttet førstnevnte fikk de estimater på den nedre grensen av k på 1,37 for ”blue-collar” funksjonærer og 1,09 for ”white-collar” funksjonærer, og øvre grense på henholdsvis 1,41 og 1,26. Estimater på 1,28, fremstilt i tabell 1, er snittene av øvre og nedre grense for ”blue-” og ”white-collar” funksjonærer ($\frac{1,37+1,41+1,09+1,26}{4}$).

Den ikke-parametriske metoden ga vesentlig høyere estimater; 1,75 for de som var i ”blue-collar” og 1,48 for ”white-collar”. Gjennomsnittsestimatet ble 1,6 ($\frac{1,75+1,48}{2}$), 10,6 % av BNP i Storbritannia i 1993. Den relativt store forskjellen mellom estimatene kan tyde på at det er sentralt å ta hensyn til heterogene preferanser.

3.2 Canada

3.2.1 Schuetze (2002)

I denne studien er husholdningene definert som selvstendig næringsdrivende når minst 30 % av inntekten kommer derfra. Schuetze (2002) fant at disse husholdningene utelater 11-23 % når de rapporterer sin inntekt. Resultatene indikerte at estimatet varierte mellom yrke, alder og antall selvstendig næringsdrivende i husholdningen.

3.2.2 Fortin et al. (2010)

Denne studien gjaldt kun for Quebec i Canada. I tillegg til å estimere størrelsen og veksten på den svarte økonomien, fant Fortin et al. (2010) også det korresponderende tapet av skatteinntekter for myndighetene. Her ble det antatt at den marginale tilbøyeligheten til å konsumere kunne variere mellom ulike inntektskilder, slik som i Storbritannia (Lyssiotou et al., 2004). I følge Fortin et al. (2010) steg den svarte økonomien i Quebec fra 4,6 % av BNP i 1997, til 5,7 % i 2002. Dette tilsvarte et tap på henholdsvis 2 og 3 milliarder for myndighetene.

3.3 USA

Hurst et al. (2010) fant indikasjoner på at selvstendig næringsdrivende underrapporterte sin inntekt med om lag 30 %, relativt til arbeidstakere i spørreundersøkelser som ble tildelt husholdningene. Underrapporteringen var høyere der de marginale skattesatsene var høyere, og lavere blant de som hadde en "avansert" utdannelse.

3.4 Uganda

I Uganda var omfanget av den svarte økonomien på 0,16 % av BNP på starten av 90-tallet, en ubetydelig andel i følge Obwona (1999). Denne studien åpnet opp for at skatteunndragelsen kunne bestå av urapportert inntekt fra egenproduksjon, i tillegg til underrapportering. Egenproduksjon blir diskutert nærmere i kapittel 4.3.2. Obwona (1999) hevdet at det ikke fantes nok inntekt å skattlegge i Uganda, og at estimatet derfor ble så lavt. For å øke landets inntekter, ble det derfor foreslått innføring av "forbruksskatt" og eiendomsskatt i stedet for inntektsskatt.

3.5 Australia

I følge Bradbury (1997) var den ”sanne” konsumkapasiteten på starten av 90-tallet om lag 34 % høyere enn hva den rapporterte inntekten indikerte. Det ble hevdet at dette skyldtes ulik sparing blant selvstendig næringsdrivende og arbeidstakere, i tillegg til underrapportering av inntekt.

Bradbury (1997) fant også at levestandarden blant selvstendig næringsdrivende med lav snittinntekt var høyere enn levestandarden blant arbeidstakere med lav snittinntekt. Samtidig fant studien indikasjoner på at lavt konsum var mer vanlig blant selvstendig næringsdrivende enn blant arbeidstakere. Dette var motstridende resultater.

3.6 Finland

Utvalget i studien til Johansson (2005) ble delt opp i to grupper; én hvor kun hovedinntektstaker var selvstendig næringsdrivende i husholdningen, og én hvor begge hovedpersonene var det. Gjennomsnittlig underrapportering av inntekt ble her estimert til henholdsvis 16,5 % og 42 % i de to gruppene. Estimaten var basert på snittet mellom øvre og nedre grense i hver av de to gruppene. Resultatene stemte godt overens med hypotesen om at større andel av inntekt fra selvstendig næringsvirksomhet fører til større grad av skatteunndragelse.

I Finland var om lag 8 % av den totale inntekten fra selvstendig næringsvirksomhet. Ut fra dette konkluderte Johansson (2005) med at størrelsen på den svarte økonomien var 1,3-3,2 % av BNP.

3.7 Sverige

Engström og Holmlund (2009) undersøkte om selvstendig næringsdrivende i ”unincorporated businesses” underrapporterte større grad av sin inntekt enn de i ”incorporated businesses”. Her ble ”incorporated businesses” definert som bedrifter hvor selvstendig næringsdrivende var ansatt i sitt eget selskap, og hvor de ble lønnet av selskapet. Dette gjaldt AS og tilsvarende selskapsformer. Det var flere regler og reguleringer knyttet til disse bedriftene enn til

”unincorporated businesses”. Blant annet var det krav om at finansielle transaksjoner og bevegelser var offentlig tilgjengelig.

Resultatene indikerte at selvstendig næringsdrivende i ”unincorporated businesses” underrapporterte sin inntekt med 40-50 %, mens de i ”incorporated businesses” underrapporterte med 15-20 %. En av grunnene for høyere underrapportering i ”unincorporated businesses” kan ha vært kombinasjonen av mindre krav til åpenhet og færre regler.

I Engström og Holmlund (2009) kom det også frem at husholdninger med minst én selvstendig næringsdrivende underrapporterte med om lag 30 %, som presentert i tabell 1. Dette kom de frem til ved bruk av både OLS- og IV-estimering, hvor de hentet ut estimatet ved å bruke gjennomsnittet mellom metodene.

3.8 Jamaica

Torero et al. (2006) klassifiserte individene i ”formell” eller ”uformell” sektor. Innenfor disse sektorene så de om individene var arbeidstakere eller selvstendig næringsdrivende.

Av totalt 1578 observasjoner (1025 husholdninger), var 715 ”formelle” arbeidere, mens 863 var ”uformelle”. Resultatene indikerte at ”formelle” arbeidere konsumerte mer enn ”uformelle” arbeidere. De fant også at ”formelle” arbeidere i snitt måtte multiplisere sin rapporterte inntekt med 1,32 for å komme frem til deres reelle inntekt. Dette tilsvarte 9 % av BNP på Jamaica i 2001.

3.9 Estland

Kukk og Staehr (2014) kom frem til at selvstendig næringsdrivende underrapporterte 62 % av sin reelle inntekt. I denne studien fokuserte de på hvordan selvstendig næringsdrivende var identifisert; ved bruk av andel rapportert bedriftsinntekt over en gitt grense, eller ved bruk av rapportert ansettelsesforhold (selvstendig næringsdrivende eller arbeidstaker).

Estimatet på underrapportert inntekt ble om lag dobbelt så stor når husholdningene ble identifisert ved bruk av andel bedriftsinntekt (mellom 2,28 og 2,63), sammenlignet med

ansettelsesforhold (1,39). Det viste seg at relativt mange hadde rapportert seg selv som selvstendig næringsdrivende uten å ha bedriftsinntekt. I tillegg hadde flere rapportert seg selv som arbeidstakere samtidig som de hadde en vesentlig andel bedriftsinntekt. Analysene i denne studiene indikerte at andelen bedriftsinntekt i husholdningene var en bedre identifikasjonsmetode.

3.10 Tyrkia

Davutyan (2008) hevdet at antall kontrolleringer av skattebetalingene hadde betydning for størrelsen på den svarte økonomien. Tyrkia hadde liten grad av skattekontrollører sammenlignet med andre land. Det ble påpekt at disse dårlige kontrollorganene, kombinert med høye inntektsskatter og lave straffer ved å bli oppdaget, kan ha vært årsaker til underrapporteringen.

Videre hevdet Davutyan (2008) at en annen årsak til underrapportering var de høye kostnadene ved å avskjedige noen fra en jobb. Dette førte til stor grad av overtidsjobbing, noe som videre førte til større arbeidsledighet og flere muligheter for underrapportering blant de som var i arbeid.

3.11 Spania

Martinez-Lopez (2013) utvidet tolkningen til Pissarides og Weber (1989) ved å åpne for at også *arbeidstakere* kunne underrapportere en andel av sin inntekt. På denne måten fikk man en bredere tolkning av resultatene. Estimater på underrapportering blant selvstendig næringsdrivende ble dermed ikke målt relativt til selvstendig næringsdrivendes sanne inntekt, men til den inntekten fra selvstendig næringsdrivende som var lik graden av underrapportering fra arbeidstakere.

Resultatene fra denne studien indikerte at rapportert inntekt blant selvstendig næringsdrivende måtte økes med om lag 25 % for å oppnå reell inntekt, noe som tilsvarte 2,5 % av BNP. Martinez-Lopez (2013) hevdet at en svak "skattemoral" i Spania og en stor andel selvstendig næringsdrivende kan ha vært med på å gi et relativt høyt estimat på underrapporteringen.

4. Den utgiftsbaserte estimeringsmetoden

I dette kapitlet gir vi en detaljert beskrivelse av metoden vi benytter for å gjennomføre våre analyser, og svakheter ved den. I all hovedsak har vi brukt statistikkprogrammet STATA til å utføre beregninger og regresjoner.

Den utgiftsbaserte metoden fra Pissarides og Weber (1989) tar utgangspunkt i at det kun er selvstendig næringsdrivende som underrapporterer sin inntekt, og at det i hovedsak gjelder små familiebedrifter.

Hovedideen med metoden er å sammenligne forholdet mellom rapportert inntekt og matutgifter for selvstendig næringsdrivende og arbeidstakere, for et gitt inntektsnivå. Det er dermed sentralt at dette inntektsmålet er riktig.

Pissarides og Weber (1989) diskuterer seg frem til at matkostnaden er den utgiftsposten som gir den mest nøyaktige rapporteringen, da dette ikke faller under kategorien for firmaets utgifter for selvstendig næringsdrivende. I tillegg er det lite sannsynlig at de som fyller ut matutgiftene i forbruksundersøkelser rapporterer for lave matutgifter.

Metoden er en indirekte metode. Dette betyr at respondentene ikke har avgitt sine svar til dette formålet, og dermed ikke har noen grunn for å tilbakeholde informasjon. Ut fra dette er det grunn til å tro at metoden kan gi oss pålitelige estimater.

Videre vil vi ta for oss den utgiftsbaserte estimeringsmetoden mer detaljert, og vil først gå inn på estimering av underrapporteringen ved bruk av minste kvadraters metode (OLS), før vi tar for oss instrumentvariabelmetoden (IV).

4.1 Estimering ved OLS

I denne utredningen tar vi i bruk data fra Forbruksundersøkelsen i Norge. Fra disse dataene aggregerer vi en matvarevariabel, C_i , som presenterer totale matutgifter per husholdning i , hvor i illustrerer om husholdningen er definert som selvstendig næringsdrivende (SE) eller arbeidstaker (EE). Fra undersøkelsen hentes også et sett med husholdningskarakteristika (appendiks A.4) og informasjon om rapportert inntekt etter skatt, som henholdsvis betegnes

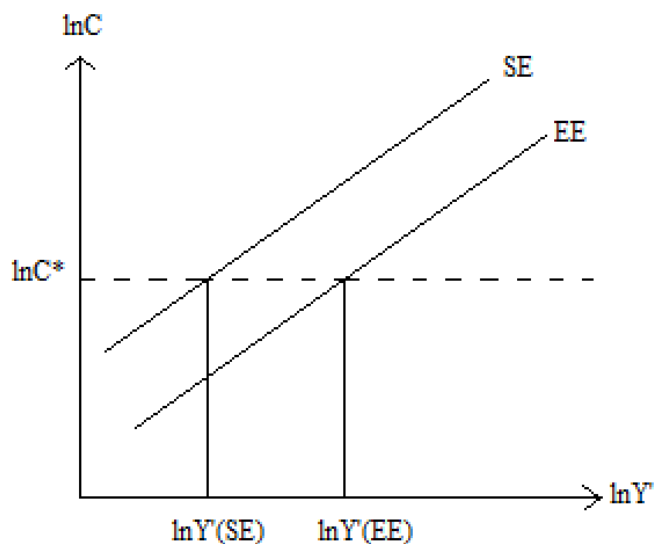
av Z_i og Y_i' . Modellen antar at matutgiftene og husholdningskarakteristikaene er korrekt rapportert av alle husstander. I tillegg antas det at inntekt etter skatt er korrekt rapportert av alle arbeidstakere, men underrapportert av selvstendig næringsdrivende.

Den "sanne" inntekten for husholdning i representeres av Y_i , og den rapporterte inntekten av Y_i' . For alle arbeidstakere har vi at $Y_i = Y_i'$, slik at den rapporterte inntekten er deres "sanne" inntekt. For selvstendig næringsdrivende finnes det en tilfeldig variabel, k_i , som fanger opp underrapporteringen. Den "sanne" inntekten blir dermed gitt ved;

$$Y_i = k_i Y_i', \quad k_i \geq 1, \quad (1)$$

der vi forventer at $k_i = 1$ for arbeidstakere, og $k_i > 1$ for selvstendig næringsdrivende. En høy k_i illustrerer stor grad av underrapportering.

For å komme frem til dette estimatet på k_i , ønsker vi først å illustrere situasjonen ved hjelp av en Engel-kurve, etter idé fra Engström og Holmlund (2009).



Figur 1: Engel-kurver for selvstendig næringsdrivende (SE) og arbeidstakere (EE)

Figuren illustrer to log-lineære inntekts-forbruksprofiler (Engel-kurver), én for selvstendig næringsdrivende og én for arbeidstakere. Engel-kurvene viser konsumentenes endring i matforbruk når inntekten endres. Vi antar at elastisiteten av konsumet med hensyn på

inntekten er lik for begge gruppene, slik at vi får lik helning på de to kurvene. For å fange opp eventuell underrapportering blant selvstendig næringsdrivende, tillater vi imidlertid at skjæringspunktet, altså konstantleddet, er forskjellig.

Fra figur 1 kan vi se at selvstendig næringsdrivende og arbeidstakere rapporterer samme matkonsum, $\ln C_i^*$. Selvstendig næringsdrivende rapporterer imidlertid lavere inntektsnivå enn arbeidstakere, slik at $\ln Y'_{SE} < \ln Y'_{EE}$. Vi antar her, som tidligere, at matkonsum er riktig rapportert for begge husholdningene, at arbeidstakere rapporterer riktig inntekt, men at selvstendig næringsdrivende underrapporterer sin inntekt. Når vi videre antar at disse to husholdningene har samme preferanser, burde inntektsnivået være likt, slik at $\ln Y'_{SE} = \ln Y'_{EE}$. Dette er ikke tilfellet, som vi ser av figur 1. Graden av underrapportering, $\ln k_i$, kan da uttrykkes som $\ln Y'_{EE} - \ln Y'_{SE}$.

For å illustrere dette ved hjelp av en utgiftsfunksjon kan vi, i følge Pissarides og Weber (1989), slå sammen informasjonen for selvstendig næringsdrivende og for arbeidstakere, og estimere følgende:

$$\ln C_i = Z_i \alpha + \beta \ln Y'_i + \gamma SE_i + \varepsilon_i, \quad (2)$$

der α er en vektor, β er en skalar med tolkning; "den marginale tilbøyeligheten" til å konsumere mat, γ er en parameter som fanger opp forskjellen mellom selvstendig næringsdrivende og arbeidstakere, og ε_i er feilleddet. Videre er Z_i en vektor som illustrerer variablene for husholdningskarakteristika som påvirker konsum, og Y'_i den rapporterte inntekten, som vi foreløpig antar er det relevante inntektsmålet. SE_i er en dummy-variabel som tar verdien 1 for selvstendig næringsdrivende og 0 for arbeidstakere.

I ligning (2) fanger parameteren γ opp den vertikale avstanden mellom de to Engel-kurvene i figur 1. Når γ er større enn null får vi en indikasjon på underrapportering blant selvstendig næringsdrivende.

I følge Engström og Holmlund (2009), kan vi nå finne et grovt estimat på graden av underrapportering ved bruk av koeffisientene i utgiftsfunksjonen:

$$\ln k_i = \ln Y'_{EE} - \ln Y'_{SE} = \frac{\gamma}{\beta}. \quad (3)$$

Ut fra dette kan vi finne tallet k_i , som multipliseres med rapportert inntekt for å komme frem til selvstendig næringsdrivendes ”sanne” inntekt;

$$k_i = e^{\frac{\gamma}{\beta}}. \quad (4)$$

4.2 Estimering ved IV

4.2.1 Endogenitetsproblem i inntektsvariabelen

Pissarides og Weber (1989) diskuterer en rekke problemer knyttet til estimering av utgiftsfunksjonen. Dette dreier seg blant annet om hvordan inntektsvariabelen skal behandles. Denne variabelen kan være endogen, noe som kan gi forventningsskjevne estimater ved bruk av OLS-estimering (Woolridge, 2012).

Det ville vært mest hensiktsmessig å benytte permanent inntekt, da dette er det inntektsmålet som påvirker beslutninger angående konsum. Vi har imidlertid ikke informasjon om permanent inntekt, og kan dermed få et endogenitetsproblem. Pissarides og Weber (1989) forsøkte å løse dette problemet ved å behandle rapportert inntekt som en endogen variabel og ta i bruk instrumentvariabelmetoden. De antok at inntektsvariabelen korrelerte med feilledet, noe som kan skyldes målefeil. Vi vil her ta for oss instrumentvariabelmetoden. Et potensielt problem med denne metoden er at man må finne instrumentvariabler som kun påvirker rapportert inntekt uten å påvirke matkonsum direkte i tillegg. Dette er svært krevende.

4.2.2 Permanent inntekt

Ligning (1), som representerer sammenhengen mellom rapportert og ”sann” inntekt, vil også gjelde her: $Y_i = k_i Y'_i$.

Først tar vi for oss hvordan utgiftsfunksjonen ville sett ut ved bruk av permanent inntekt (Pissarides og Weber, 1989);

$$\ln C_i = Z_i \alpha + \beta \ln Y_i^p + \varepsilon_i, \quad (5)$$

der Y_i^p er permanent inntekt. Ligning (5) representerer den log-lineære Engel-kurven beskrevet tidligere. Vi tar utgangspunkt i at den "sanne" inntekten varierer mer for selvstendig næringsdrivende enn for arbeidstakere, da den blir mer preget av svingninger i markedet. Pissarides og Weber (1989) hevdet at graden av underrapportering blant selvstendig næringsdrivende måtte justeres tilsvarende. Sammenhengen mellom "sann" og permanent inntekt blir dermed:

$$Y_i = p_i Y_i^p, \quad (6)$$

der p_i er en tilfeldig variabel, hvor dens forventede verdi avhenger av tilfeldige hendelser. Et "godt" år vil gjennomsnittet til p_i , \bar{p} , være større enn én, og lavere enn én et "dårlig" år. Gjennomsnittet, som vi antar å være lik forventningen til p_i , antas å være likt for selvstendig næringsdrivende og arbeidstakere. Vi antar imidlertid at variansen til p_i er større for selvstendig næringsdrivende enn for arbeidstakere, da denne gruppen ofte står overfor høyere risiko og mer volatil inntekt.

Ved å snu om på ligning (6), slik at $Y_i^p = \frac{Y_i}{p_i}$, og ved å sette inn for Y_i fra ligning (1), slik at

$Y_i^p = \frac{k_i Y_i'}{p_i}$, kan vi finne den naturlige logaritmen for permanent inntekt i ligning (5) slik:

$$\ln Y_i^p = \ln\left(\frac{k_i Y_i'}{p_i}\right) = \ln(k_i Y_i') - \ln p_i$$

$$\ln Y_i^p = \ln Y_i' - \ln p_i + \ln k_i. \quad (7)$$

Utgiftsfunksjonen kan da videre skrives som:

$$\begin{aligned} \ln C_i &= Z_i \alpha + \beta (\ln Y_i' - \ln p_i + \ln k_i) + \varepsilon_i \\ &= \ln Z_i \alpha + \beta \ln Y_i' - \beta \ln p_i + \beta \ln k_i + \varepsilon_i. \end{aligned} \quad (8)$$

4.2.3 Variablene p_i og k_i

På nåværende tidspunkt har vi ikke informasjon om p_i og k_i , og kan dermed ikke estimere utgiftsfunksjonen og Engel-kurven. Pissarides og Weber (1989) antar videre at p_i og k_i er lognormalfordelte og at:

$$\ln p_i = \mu_p + u_i, \quad (9)$$

$$\ln k_i = \mu_k + v_i. \quad (10)$$

Her har de tilfeldige variablene u_i og v_i forventet verdi lik null og konstant varians, σ_u^2 og σ_v^2 , innenfor yrkesgruppene selvstendig næringsdrivende og arbeidstakere. Ligning (9) og (10) vil imidlertid være ulik for disse to gruppene. Siden $k_i = 1$ for arbeidstakere, vil ligning (10) bli lik null for denne gruppen, og dermed kun gjelde for selvstendig næringsdrivende. Ligning (9) vil vi komme nærmere inn på i kapittel 4.2.4. Først vil vi utdype utgiftsfunksjonen fra ligning (8). Når vi setter inn uttrykkene for $\ln p_i$ og $\ln k_i$ fra ligning (9) og (10) får vi:

$$\begin{aligned} \ln C_i &= Z_i \alpha + \beta \left(\ln Y_i' - (\mu_p + u_i) + (\mu_k + v_i) \right) + \varepsilon_i \\ &= Z_i \alpha + \beta \ln Y_i' - \beta \mu_p + \beta \mu_k - \beta u_i + \beta v_i + \varepsilon_i. \end{aligned}$$

Følgende ligning gir da, i følge Pissarides og Weber (1989), et pålitelig estimat på matutgiftsvariabelen:

$$\ln C_i = Z_i \alpha + \beta \ln Y_i' - \beta(\mu_p - \mu_k) - \beta(u_i - v_i) + \varepsilon_i. \quad (11)$$

Her vil uttrykket $-\beta(\mu_p - \mu_k)$ bli en del av konstantleddet, mens $-\beta(u_i - v_i)$ inngår i feilleddet. Først vil vi gå nærmere inn på utvidelsen av feilleddet, før vi ser mer på konstantleddet i kapittel 4.2.4.

Den ekstra komponenten til feilleddet, $-\beta(u_i - v_i)$, inngår også i uttrykket for rapportert inntekt, Y_i' . Av denne grunn vil inntektsvariabelen korrelere med feilleddet og vi har dermed et endogenitetsproblem. Vi kan sette opp et uttrykk for feilleddet i ligning (11) slik:

$$\eta_i = \varepsilon_i - \beta(u_i - v_i),$$

hvor vi antar at ε_i ikke korrelerer med (u_i, v_i) , slik at $cov(\varepsilon_i, u_i) = 0$ og $cov(\varepsilon_i, v_i) = 0$.

Dette kan enklere uttrykkes ved;

$$cov(\varepsilon_i, u_i - v_i) = 0,$$

siden $cov(\varepsilon_i, u_i - v_i) = cov(\varepsilon_i, u_i) - cov(\varepsilon_i, v_i) = 0$. Variansen til konsumresidualet er da gitt ved:

$$\begin{aligned} Var(\eta_i) &= \sigma_\eta^2 = var(\varepsilon_i - \beta(u_i - v_i)) \\ &= var(\varepsilon_i) + var(\beta(u_i - v_i)) - 2cov(\varepsilon_i, \beta(u_i - v_i)) \\ &= var(\varepsilon_i) + var(\beta(u_i - v_i)) - 2\beta cov(\varepsilon_i, u_i - v_i) \\ &= var(\varepsilon_i) + \beta^2 var(u_i - v_i) - 2\beta \cdot 0 \\ &= var(\varepsilon_i) + \beta^2 (var(u_i) + var(v_i) - 2cov(u_i, v_i)) \\ &= \sigma_{\varepsilon_i}^2 + \beta^2 (\sigma_{u_i}^2 + \sigma_{v_i}^2 - 2\rho\sigma_{u_i}\sigma_{v_i}), \end{aligned}$$

og vil være ulik for yrkesgruppene, da selvstendig næringsdrivende generelt antas å ha større varians. Korrelasjonskoeffisienten mellom u_i og v_i er illustrert ved ρ .

4.2.4 Konstantleddet i utgiftsfunksjonen

Her vil vi gå nærmere inn på den ekstra komponenten til konstantleddet fra ligning (11), $-\beta(\mu_p - \mu_k)$. Uttrykket i parentes, $\mu_p - \mu_k$, har ulik verdi for selvstendig næringsdrivende og for arbeidstakere. Ved å sette opp ligning (11) separat for de to gruppene kan vi observere disse ulikhetene:

Selvstendig næringsdrivende:

$$\ln C_{SE} = Z_{SE}\alpha + \beta \ln Y'_{SE} - \beta(\mu_{pSE} - \mu_{kSE}) - \beta(u_{SE} - v_{SE}) + \varepsilon_{SE}$$

Arbeidstakere:

$$\ln C_{EE} = Z_{EE}\alpha + \beta \ln Y'_{EE} - \beta(\mu_{pEE} - \mu_{kEE}) - \beta(u_{EE} - v_{EE}) + \varepsilon_{EE}$$

Vi innfører videre en dummy-variabel for selvstendig næringsdrivende. Denne vil kunne fange opp de ulike konstantleddene mellom gruppene. På denne måten kan vi estimere en felles utgiftsfunksjon for de to gruppene og til slutt estimere k .

I ligning (2) fra OLS presenterte vi kort dummy-variabelen for selvstendig næringsdrivende, SE_i , og dens koeffisient, γ . Videre vil vi gi en mer detaljert beskrivelse av hva komponenten γ inneholder.

Vi tar først for oss forskjellen mellom selvstendig næringsdrivende og arbeidstakere i konstantleddet fra ligning (11);

$$\begin{aligned} -\beta(\mu_{pSE} - \mu_{kSE}) - (-\beta(\mu_{pEE} - \mu_{kEE})) &= -\beta\mu_{pSE} + \beta\mu_{kSE} + \beta\mu_{pEE} - \beta\mu_{kEE} \\ &= -\beta[\mu_{pSE} - \mu_{kSE} - \mu_{pEE} + \mu_{kEE}]. \end{aligned}$$

Forventningen til k_i for arbeidstakere er lik 1, så μ_{kEE} forsvinner (fordi $\ln(1) = 0$). Når vi antar at γ er forskjellen på uttrykket $-\beta(\mu_p - \mu_k)$ mellom selvstendig næringsdrivende og arbeidstakere, får vi at:

$$\gamma = -\beta[(\mu_{pSE} - \mu_{pEE}) - \mu_{kSE}] = \beta[\mu_{kSE} - (\mu_{pSE} - \mu_{pEE})].$$

For å komme frem til et mer presist uttrykk for γ , må vi gå nærmere inn på ligning (9), $\ln p_i = \mu_p + u_i$. Fra regneregler om logaritmisk normalfordeling har vi at:

$$E[p_i] = e^{\mu_p + \frac{1}{2}\sigma_u^2}.$$

Når vi antar at gjennomsnittet av p_i er en variant av dens forventning, kan vi si at $\bar{p} = E[p_i]$, og får dermed at:

$$\ln \bar{p} = \ln(E[p_i]) = \ln(e^{\mu_p + \frac{1}{2}\sigma_u^2}) = \mu_p + \frac{1}{2}\sigma_u^2. \quad (12)$$

Dersom vi ser på forskjellen mellom selvstendig næringsdrivende og arbeidstakere i ligning (12), får vi følgende:

$$\ln \bar{p}_{SE} - \ln \bar{p}_{EE} = \left(\mu_{pSE} + \frac{1}{2} \sigma_{uSE}^2 \right) - \left(\mu_{pEE} + \frac{1}{2} \sigma_{uEE}^2 \right),$$

hvor vi har at $\sigma_{uSE}^2 > \sigma_{uEE}^2$ fra antakelsen om at variansen til p_i er størst blant selvstendig næringsdrivende. Videre har vi at forventningen til p_i er lik for begge gruppene, noe som gir oss en venstreside lik null. Vi får da uttrykket:

$$\mu_{pSE} - \mu_{pEE} = -\frac{1}{2} (\sigma_{uSE}^2 - \sigma_{uEE}^2) \leq 0. \quad (13)$$

Ligning (13) kan vi sette inn i det foreløpige estimatet på γ , $\beta[\mu_{kSE} - (\mu_{pSE} - \mu_{pEE})]$, slik at vi får:

$$\gamma = \beta \left[\mu_{kSE} - \left(-\frac{1}{2} (\sigma_{uSE}^2 - \sigma_{uEE}^2) \right) \right] = \beta \left[\mu_{kSE} + \frac{1}{2} (\sigma_{uSE}^2 - \sigma_{uEE}^2) \right]. \quad (14)$$

Nå kan vi estimere utgiftsfunksjonen på nytt. Vi bruker ligning (11) og setter inn γSE_i for $-\beta(\mu_p - \mu_k)$, ettersom vi vet at dette er forskjellen mellom selvstendig næringsdrivende og arbeidstakere. Vi setter også inn η_i for feilleddet $\varepsilon_i - \beta(u_i - v_i)$;

$$\ln C_i = Z_i \alpha + \beta \ln Y_i' + \gamma SE_i + \eta_i. \quad (15)$$

Her representerer γ forskjellen mellom selvstendig næringsdrivende og arbeidstakere, og er en viktig parameter for å kunne utlede et estimat for underrapporteringen, k_i . Vi tar med oss at k_i kun er en stokastisk variabel for selvstendig næringsdrivende, hvor vi antar at $k_{SE} > 1$. Av denne grunn er det kun interessant å finne k_i for selvstendig næringsdrivende, og vi kan videre betrakte k_{SE} som k .

4.2.5 Variansen til inntektsresidualet

Når vi behandler den rapporterte inntekten som en endogen variabel er det nødvendig med et uavhengig estimat på variansen til inntektens feilledd. Dette vil være en vesentlig faktor i utregningen av k . Pissarides og Weber (1989) foreslo å sette opp inntektsfunksjonen i redusert form på denne måten:

$$\ln Y_i' = Z_i \delta_1 + X_i \delta_2 + \zeta_i, \quad (16)$$

hvor X_i er et sett med identifiserende instrumenter, δ_1 og δ_2 er vektorer med koeffisientene, og ζ_i er feilleddet. Feilleddet er gitt ved:

$$\zeta_i = \varepsilon_i - (u_i - v_i),$$

med varians σ_Y^2 . Vi antar også her at $cov(\varepsilon_i, u_i - v_i) = 0$, og at variansen til inntektsresidualet er konstant innenfor gruppene. Denne variansen er da gitt ved:

$$\begin{aligned} var(\zeta_i) &= \sigma_Y^2 = var(\varepsilon_i - (u_i - v_i)) = var(\varepsilon_i) + var(u_i - v_i) - 2cov(\varepsilon_i, u_i - v_i) \\ &= var(\varepsilon_i) + (var(u_i) + var(v_i) - 2cov(u_i, v_i)) - 0 \\ &= \sigma_{\varepsilon_i}^2 + \sigma_{u_i}^2 + \sigma_{v_i}^2 - 2cov(u_i, v_i) = \sigma_{\varepsilon_i}^2 + \sigma_{u_i}^2 + \sigma_{v_i}^2 - 2\rho\sigma_{u_i}\sigma_{v_i}. \end{aligned}$$

Vi antar at også denne variansen er større for selvstendig næringsdrivende enn for arbeidstakere.

Videre antar vi at den delen av σ_Y^2 som er uforklart variasjon i permanent inntekt, ε_i , i ligning (16), har den samme variansen for både selvstendig næringsdrivende og arbeidstakere. Da får vi at $\sigma_{\varepsilon_{SE2}}^2 = \sigma_{\varepsilon_{EE2}}^2$. Vi kan ikke teste denne forutsetningen, men antar at det er realistisk ut fra Pissarides og Webers (1989) begrunnelse om at variasjonen kommer fra utelatte variabler og dummy-variabelen for selvstendig næringsdrivende. I tillegg til dette har vi at $\sigma_{vEE}^2 = 0$ fra antakelsen om at arbeidstakere rapporterer sin reelle inntekt. Når vi skal finne forskjellen på variansen til inntektsresidualet mellom selvstendig næringsdrivende og arbeidstakere, kan vi sette opp:

$$\begin{aligned} var(\zeta_{SE}) - var(\zeta_{EE}) &= \sigma_{YSE}^2 - \sigma_{YEE}^2 \\ &= \sigma_{\varepsilon_{SE}}^2 + \sigma_{u_{SE}}^2 + \sigma_{v_{SE}}^2 - 2\rho\sigma_{u_{SE}}\sigma_{v_{SE}} - (\sigma_{\varepsilon_{EE}}^2 + \sigma_{u_{EE}}^2 + \sigma_{v_{EE}}^2 - 2\rho\sigma_{u_{EE}}\sigma_{v_{EE}}). \end{aligned}$$

Vi setter inn for $\sigma_{vEE} = 0$ og $\sigma_{\varepsilon_{SE}}^2 = \sigma_{\varepsilon_{EE}}^2$, og får videre:

$$\begin{aligned}\sigma_{YSE}^2 - \sigma_{YEE}^2 &= \sigma_{\varepsilon EE}^2 + \sigma_{uSE}^2 + \sigma_{vSE}^2 - 2\rho\sigma_{uSE}\sigma_{vSE} - (\sigma_{\varepsilon EE}^2 + \sigma_{uEE}^2 + 0^2 - 2\rho\sigma_{uEE} \cdot 0) \\ &= \sigma_{uSE}^2 + \sigma_{vSE}^2 - 2\rho\sigma_{uSE}\sigma_{vSE} - \sigma_{uEE}^2.\end{aligned}\quad (17)$$

Ligning (17) er vesentlig når vi skal finne en øvre og nedre grense for den gjennomsnittlige underrapporteringen, \bar{k} .

4.2.6 Område for gjennomsnittlig underrapportering

Fra ligning (1) og (10) har vi henholdsvis at $Y_i = k_i Y_i'$ og $\ln k_i = \mu_k + v_i$. Ved bruk av samme regneregler og antakelser som for $\ln \bar{p}$, får vi at:

$$\ln \bar{k} = \mu_k + \frac{1}{2}\sigma_v^2,$$

hvor σ_v^2 er variansen til k . Når vi ser på forskjellen mellom selvstendig næringsdrivende og arbeidstakere, og husker at vi har $k = 1$ for arbeidstakere, får vi at:

$$\ln \bar{k}_{SE} - \ln \bar{k}_{EE} = \ln \bar{k}_{SE} - 0 = \mu_{kSE} + \frac{1}{2}\sigma_{vSE}^2.$$

Vi får dermed:

$$\ln \bar{k} = \mu_k + \frac{1}{2}\sigma_{vSE}^2. \quad (18)$$

For å kunne isolere effekten av μ_{kSE} fra ligning (14), kan vi omskrive ligning (18). Uttrykket μ_{kSE} betegnes her som μ_k .

$$\mu_k = \ln \bar{k} - \frac{1}{2}\sigma_{vSE}^2,$$

og substituere dette inn i ligning (14) slik at vi får:

$$\gamma = \beta \left[\mu_k + \frac{1}{2}(\sigma_{uSE}^2 - \sigma_{uEE}^2) \right] = \beta \left[(\ln \bar{k} - \frac{1}{2}\sigma_{vSE}^2) + \frac{1}{2}(\sigma_{uSE}^2 - \sigma_{uEE}^2) \right].$$

Ved å omskrive videre får vi:

$$\frac{\gamma}{\beta} = \ln \bar{k} - \frac{1}{2} \sigma_{vSE}^2 + \frac{1}{2} (\sigma_{uSE}^2 - \sigma_{uEE}^2).$$

Dette gir oss følgende uttrykk for $\ln \bar{k}$:

$$\ln \bar{k} = \frac{\gamma}{\beta} + \frac{1}{2} (\sigma_{vSE}^2 - \sigma_{uSE}^2 + \sigma_{uEE}^2). \quad (19)$$

På bakgrunn av at p_i og k_i behandles som tilfeldige variabler, har vi foreløpig ikke mulighet til å finne et konkret estimat på $(\sigma_{vSE}^2 - \sigma_{uSE}^2 + \sigma_{uEE}^2)$, slik at vi kan finne gjennomsnittlig underrapportering. Likevel kan vi, i følge Pissarides og Weber (1989), estimere et *område* for gjennomsnittlig underrapportering. De hevder at dersom $\rho = 0$, vil dette området være lite og veldefinert. Dette kommer av at når kovariansen er lik null, har vi en indikasjon på at uansett hvilken ”sann” inntekt selvstendig næringsdrivende ender opp med å ha et gitt år, vil husholdningen mest sannsynlig rapportere samme andel av den til skattemyndighetene. Dersom $\rho \neq 0$, derimot, kan vi ikke kalkulere disse grensene for k uten mer informasjon om u og v , da andelen av den ”sanne” inntekten som rapporteres vil variere fra år til år. Mer diskusjon vedrørende disse problemstillingene fremkommer i kapittel 4.2.7.

Vi antar videre at $\rho = 0$. Da kan ligning (17) uttrykkes som $\sigma_{YSE}^2 - \sigma_{YEE}^2 = \sigma_{uSE}^2 + \sigma_{vSE}^2 - \sigma_{uEE}^2$. I følge Pissarides og Weber (1989) vil da σ_{uSE}^2 og σ_{vSE}^2 være negativt relaterte. Dette betyr at når variansen til den tilfeldige variabelen p_i for selvstendig næringsdrivende, σ_{uSE}^2 , øker, minker variansen til underrapporteringen k_i for selvstendig næringsdrivende, σ_{vSE}^2 , og omvendt. Ut fra dette hevder Pissarides og Weber (1989) at ligning (19) gir et nedre estimat på $\ln \bar{k}$ når σ_{vSE}^2 tar sin laveste verdi, og et øvre estimat når σ_{uSE}^2 tar sin laveste verdi.

4.2.6.1 Nedre estimat på \bar{k} når $\rho = 0$

Den laveste verdien σ_{vSE}^2 kan ta er 0, og dette forekommer når selvstendig næringsdrivende ikke underrapporterer sin inntekt. Ved å sette inn $\sigma_{vSE}^2 = 0$ og $\rho = 0$ i ligning (17), får vi:

$$\sigma_{YSE}^2 - \sigma_{YEE}^2 = \sigma_{uSE}^2 - \sigma_{uEE}^2.$$

Her observerer vi at:

$$\sigma_{uSE}^2 = \sigma_{YSE}^2 - \sigma_{YEE}^2 + \sigma_{uEE}^2,$$

som vi kan sette inn i ligning (19):

$$\ln \bar{k} = \frac{\gamma}{\beta} + \frac{1}{2} (\sigma_{vSE}^2 - (\sigma_{YSE}^2 - \sigma_{YEE}^2 + \sigma_{uEE}^2) + \sigma_{uEE}^2).$$

Ved å videre sette σ_{vSE}^2 lik 0 her også, indikerer ligning (17) og (19) at den nedre grensen for $\ln \bar{k}$ tilfredsstiller:

$$\ln \bar{k} = \frac{\gamma}{\beta} + \frac{1}{2} (0 - (\sigma_{YSE}^2 - \sigma_{YEE}^2 + \sigma_{uEE}^2) + \sigma_{uEE}^2) = \frac{\gamma}{\beta} - \frac{1}{2} (\sigma_{YSE}^2 - \sigma_{YEE}^2 + \sigma_{uEE}^2 - \sigma_{uEE}^2).$$

Dette gir oss uttrykket:

$$\ln \bar{k}_l = \frac{\gamma}{\beta} - \frac{1}{2} (\sigma_{YSE}^2 - \sigma_{YEE}^2). \quad (20)$$

Her indikerer indeksen l den nedre grensen (lower) for gjennomsnittsestimatet.

4.2.6.2 Øvre estimat på \bar{k} når $\rho = 0$

For å finne det øvre estimatet på k , må vi finne den laveste verdien til σ_{uSE}^2 . Fra tidligere argumenter om at inntekten til selvstendig næringsdrivende har minst like stor varians som inntekten til arbeidstakere, blir den laveste verdien for σ_{uSE}^2 lik σ_{uEE}^2 . Ved å sette inn σ_{uEE}^2 for σ_{uSE}^2 og $\rho = 0$ i ligning (17), får vi:

$$\sigma_{YSE}^2 - \sigma_{YEE}^2 = \sigma_{uEE}^2 + \sigma_{vSE}^2 - 0 - \sigma_{uEE}^2 = \sigma_{vSE}^2,$$

slik at:

$$\sigma_{vSE}^2 = \sigma_{YSE}^2 - \sigma_{YEE}^2.$$

Videre benytter vi dette uttrykket i ligning (19):

$$\ln \bar{k} = \frac{\gamma}{\beta} + \frac{1}{2}((\sigma_{YSE}^2 - \sigma_{YEE}^2) - \sigma_{uSE}^2 + \sigma_{uEE}^2),$$

og setter inn σ_{uEE}^2 for σ_{uSE}^2 , slik at:

$$\ln \bar{k} = \frac{\gamma}{\beta} + \frac{1}{2}((\sigma_{YSE}^2 - \sigma_{YEE}^2) - \sigma_{uEE}^2 + \sigma_{uEE}^2) = \frac{\gamma}{\beta} + \frac{1}{2}((\sigma_{YSE}^2 - \sigma_{YEE}^2) - 0).$$

Det øvre estimatet for underrapportering er da gitt ved:

$$\ln \bar{k}_u = \frac{\gamma}{\beta} + \frac{1}{2}(\sigma_{YSE}^2 - \sigma_{YEE}^2), \tag{21}$$

hvor indeksen u indikerer den øvre grensen (upper) for gjennomsnittsestimatet. Vi observerer fra ligning (20) og (21) at området for den gjennomsnittlige underrapporteringen er symmetrisk, og at det kun er fortegnet som skiller dem.

Vi finner et område for gjennomsnittlig underrapportering blant selvstendig næringsdrivende ved å ta den eksponentielle på begge sider av ligning (20) og (21).

4.2.7 Estimat på k dersom $\rho \neq 0$

Her vil vi drøfte kovariansen og om det er rimelig å anta at denne er lik null.

Pissarides og Weber (1989) hevdet at det ikke vil være mulig å regne ut et område for k uten mer informasjon om u og v når $\rho \neq 0$. Dette kommer av at kovariansen mellom u_{SE} og v_{SE} ikke vil være lik null i dette tilfellet.

For å ikke vekke mistanke hos skattemyndighetene, mente Pissarides og Weber (1989) at selvstendig næringsdrivende ville holde sin rapporterte inntekt tilnærmet konstant over årene. De ville dermed ha større insentiv til å skjule deler av sin inntekt et år med særlig høy inntekt, i motsetning til et år med ”normal” eller lavere inntekt. Dette kommer av at underrapportering i et år med spesielt høy inntekt ikke nødvendigvis vil vekke mistanke hos skattemyndighetene

fordi inntekten likevel vil ligge på samme nivå som tidligere. I tillegg vil det være mer å ”tjene” på skatteunndragelse, noe som styrker insentivene for underrapportering.

Det motsatte gjelder dersom inntekten er særlig lav. Da vil denne gruppen være mer tilbøyelig til å rapportere hele inntekten. Dette kommer av at underrapportering vil medføre at inntekten kommer langt under normalen i dette tilfellet.

Ut fra dette kan vi anta at selvstendig næringsdrivende har en frykt for å bli oppdaget av skattemyndighetene, og dermed ikke tør å underrapportere et år med lav inntekt. Derimot vil de underrapportere dersom inntekten et år er høy, fordi de vet det er mye penger å ”tjene” på dette. Denne atferden indikerer en positiv kovarians mellom u_{SE} og v_{SE} . Dersom dette er tilfellet, argumenterer Pissarides og Weber (1989) for at vi får et underestimat på den øvre grensen av gjennomsnittlig underrapportering, \bar{k}_u , og at den må justeres oppover.

Ved bruk av ligning (17) og (19), kan vi se hva som skjer med det øvre estimatet av underrapporteringen rent teknisk dersom korrelasjonskoeffisienten, ρ , er positiv:

Først flytter vi om på leddene i ligning (17), slik at;

$$\sigma_{vSE} = \sigma_{YSE}^2 - \sigma_{YEE}^2 - \sigma_{uSE}^2 + 2\rho\sigma_{uSE}\sigma_{vSE} + \sigma_{uEE}^2.$$

Dette bruker vi til å substituere inn i ligning (19);

$$\begin{aligned} \ln \bar{k} &= \frac{\gamma}{\beta} + \frac{1}{2}((\sigma_{YSE}^2 - \sigma_{YEE}^2 - \sigma_{uSE}^2 + 2\rho\sigma_{uSE}\sigma_{vSE} + \sigma_{uEE}^2) - \sigma_{uSE}^2 + \sigma_{uEE}^2) \\ &= \frac{\gamma}{\beta} + \frac{1}{2}(\sigma_{YSE}^2 - \sigma_{YEE}^2 - 2\sigma_{uSE}^2 + 2cov(u_{SE}, v_{SE}) + 2\sigma_{uEE}^2). \end{aligned}$$

I og med at vi undersøker den øvre grensen av underrapporteringen, har vi at $\sigma_{uSE}^2 = \sigma_{uEE}^2$, som diskutert i kapittel 4.2.6.2. Når vi setter inn for dette, får vi:

$$\begin{aligned} \ln \bar{k} &= \frac{\gamma}{\beta} + \frac{1}{2}(\sigma_{YSE}^2 - \sigma_{YEE}^2 - 2\sigma_{uEE}^2 + 2cov(u_{SE}, v_{SE}) + 2\sigma_{uEE}^2) \\ &= \frac{\gamma}{\beta} + \frac{1}{2}(\sigma_{YSE}^2 - \sigma_{YEE}^2 + 2cov(u_{SE}, v_{SE})). \end{aligned}$$

Til slutt ender vi opp med:

$$\ln \bar{k} = \frac{\gamma}{\beta} + \frac{1}{2}(\sigma_{YSE}^2 - \sigma_{YEE}^2) + \text{COV}(u_{SE}, v_{SE}). \quad (22)$$

For å kunne regne ut kovariansen i ligning (22) har vi behov for opplysninger om den partielle korrelasjonskoeffisienten, ρ , og standardavvikene $\sigma_{u_{SE}}$ og $\sigma_{v_{SE}}$. I følge Pissarides og Weber (1989) har ikke en positiv korrelasjonskoeffisient mye å si for estimatene av \bar{k} man får ut, da de kommer frem til at den øvre grensen ikke øker vesentlig. Ut fra dette hevder de at man til en viss grad kan se bort fra dette og fokusere på de tilfellene hvor $\rho = 0$. I denne utredningen vil vi dermed kun ta for oss estimering av underrapporteringen under forutsetningen om at $\rho = 0$. I appendiks A.3 finnes nærmere forklaring på utregning for tilfeller hvor $\rho > 0$.

4.3 Svakheter ved metoden

I dette delkapitlet vil vi diskutere antakelser ved metoden som kan føre til svakheter ved estimeringen. Dette kan være nyttig informasjon for analysen og for fremtidige studier.

4.3.1 Homogene preferanser

I følge Tedds (2005) antok Pissarides og Weber (1989) homogene preferanser blant husholdningene. Dette betyr at de ikke tok hensyn til hvilken kilde inntekten kom fra, om det var fra selvstendig næringsvirksomhet eller fra fast lønn til arbeidstakere, som beskrevet i kapittel 3.1.2. I Murashov (2014) ble etterspørselen delt opp i varige og dyre goder, og goder med kort holdbarhet. Resultatene indikerte her at inntektsandelen som ble brukt på varige goder ikke bare var avhengig av inntektsnivået, men også av inntektskilden.

Resultatene fra Lyssiotou et al. (2004) indikerte også at inntektskilden har betydning. Her kom det frem at forutsigbar inntekt i stor grad ble brukt på nødvendigheter, mens uforutsigbar inntekt gikk til andre formål, som luksusgoder. Uforutsigbar inntekt stammet typisk fra selvstendig næringsvirksomhet og var ofte tilknyttet (luksus)goder som kunne føre til høyere fradrag for firmaavgifter, som eksempelvis biler og datamaskiner. En annen grunn for at uforutsigbar inntekt ble brukt på luksusgoder kan ha vært andre preferanser blant selvstendig næringsdrivende.

Uansett hvordan man ser på det, kom det frem i Lyssiotou et al. (2004) at antakelsen om homogene preferanser kunne gi forventningsskjevne estimater. De potensielle skjevhetene kommer av at man ikke klarer å fange opp skillet mellom heterogene preferanser (substitusjonseffekt) og underrapportering (inntektseffekt) tilknyttet ulike inntektskilder, men at alt tolkes som en inntektseffekt.

Til tross for at Lyssiotou et al. (2004) peker på svakheter ved bruk av homogene preferanser, tar vi ikke hensyn til dette for enkelthets skyld. Dette kommer også av at vi ønsker å følge Pissarides og Weber (1989) sin metode i denne utredningen.

4.3.2 Egenproduksjon

Et problem som kan oppstå når vi skal tolke resultatene fra den konsumbaserte tilnærmingen til Pissarides og Weber (1989), er muligheten for egenproduksjon i hjemmet. Håndverkere, altså de som er i bygg- og anleggsbransjen, er for eksempel i bedre posisjon enn andre til å drive egenproduksjon. Dette kommer av at de kan utføre verdifullt arbeid på egen bolig, for eksempel i løpet av et oppussingsprosjekt. Så lenge omfanget ikke blir for stort representerer dette en skattefri inntekt. Denne ”inntekten” kan gi lavere insentiver for underrapportering, og estimatet på k kan dermed bli påvirket.

Sammenlignet med husholdninger som må kjøpe håndverkertjenester i markedet, kan husholdningene med høy grad av egenproduksjon konsumere mer mat for en gitt inntekt. Inntektseffekten fra egenproduksjon kan dermed være en feilkilde i Pissarides og Weber (1989) sitt rammeverk. Davis og Henrekson (2004) presiserte imidlertid at dette er lovlig unngåelse av skatt, men at det blir et problem når for få oppgaver blir tilbudt i markedet på grunn av skattlegging.

Videre påpekte Davis og Henrekson (2004) at skattlegging kan påvirke tidsallokering og valg av produksjonssektor når man skal utføre arbeid på egen bolig. Her har man tre valg; kjøpe tjenesten i markedet, utføre arbeidet selv (egenproduksjon) eller betale noen for å gjøre det svart. I studien kom det frem at få tjenester handles i markedet på grunn av skattleggingen, og at det er for mye tid som brukes i svart sektor. Når skattleggingen av arbeidsinntekt øker, vil denne effekten forsterkes, og den svarte økonomien vil vokse.

I denne utredningen tar vi hensyn til egenproduksjon til en viss grad ved å ekskludere de som driver med jordbruk fra datasettet, da disse forventes å produsere mye av maten selv. Dette blir nærmere beskrevet i kapittel 5.2. Vi tar imidlertid ikke hensyn til effekten av egenproduksjon i form av eget arbeid på hjemmet, da vi ikke har informasjon om hvor mye arbeid husholdningene utfører selv.

4.3.3 Andre svakheter

Her vil vi kort nevne andre svakheter ved metoden til Pissarides og Weber (1989), som har blitt nevnt i kapittel 3 om tidligere studier.

Antakelsen om at det *kun* er selvstendig næringsdrivende som underrapporterer, kan føre til at vi ikke får fanget opp potensielle effekter ved at også *arbeidstakere* kan ha muligheter og insentiver for underrapportering. I denne utredningen undersøker vi dette for bransjene bygg/anlegg og helse/undervisning.

En annen faktor denne metoden ikke fanger opp er forskjellen på sparing mellom selvstendig næringsdrivende og arbeidstakere. Dette har vi ikke muligheter for å undersøke med våre data.

Metoden tar heller ikke hensyn til hvordan selvstendig næringsdrivende blir definert; enten om det er fra selvrapportert status i datasettet eller om det er på grunnlag av andel bedriftsinntekt. I vårt datasett har vi selvrapportert status tilgjengelig og benytter dette for enkelthets skyld.

5. Data

I dette kapitlet vil vi først gi en beskrivelse av Forbruksundersøkelsen. Videre vil vi ta for oss rensing av datasettet, før vi går inn på behandling av data og valg av variabler. Til slutt presenterer vi noe beskrivende statistikk.

5.1 Beskrivelse av datasettet

Analysene i denne utredningen baserer seg på data fra Forbruksundersøkelsen (1975-2012), en undersøkelse utført av Statistisk Sentralbyrå (SSB) i Norge. Undersøkelsen blir blant annet benyttet av SSB for å utarbeide og ajourføre vektgrunnlaget for konsumprisindeksen (Norsk samfunnsvitenskapelig datatjeneste, 2009).

Forbruksundersøkelsen tar for seg private husholdninger i Norge, og baserer seg på besøksintervjuer og utgiftsføring i et representativt utvalg. Husholdningene som deltar i undersøkelsen blir trukket tilfeldig fra folkeregisteret og skal representere alle husholdninger i Norge. Fra 1996 har Forbruksundersøkelsen blant annet blitt tilkoblet opplysninger fra inntektsregisteret, før dette var inntekten selvrapportert. Alle dataene blir anonymisert. (Statistisk Sentralbyrå, 2013a).

Før 2009 ble 2 200 husholdninger tilfeldig utvalgt årlig, mens det fra og med 2012 blir trukket ut hele 7 000 husholdninger. Dette kommer av at periodiske undersøkelser med ulikt intervall skulle utføres fra 2009. Hver husholdning deltar i en periode på to uker som er tilfeldig fordelt utover året. Hovedinntektstaker defineres som den person som bidrar mest til husholdningens underhold. Vi definerer i denne utredningen person 2 som ektefelle, samboer eller registrert partner til hovedinntektstaker. Personer som er bosatt i felleshusholdninger som pensjonat, sykehus og lignende blir ikke inkludert i undersøkelsen. (Statistisk Sentralbyrå, 2013a).

Selvstendig næringsdrivende blir av Skatteetaten (2015) definert som ”enhver som for egen regning og risiko driver en vedvarende virksomhet som er egnet til å gi nettoinntekt”. I Forbruksundersøkelsen rapporterer individene selv om de er selvstendig næringsdrivende eller ikke. En svakhet er at dette ikke blir kontrollert, og vi må dermed ha tillit til at individene rapporterer riktig. På en annen side kan selvrappotering være en styrke. Dette kommer av at

mange er lønsmottakere i egen bedrift, og at det kan være vanskelig å definere hvem som er selvstendig næringsdrivende ut fra registerbasert informasjon.

5.2 Rensing av datasettet

Vi har fått tilsendt Forbruksundersøkelsen for perioden 1975-2012 fra Norsk samfunnsvitenskapelig datatjeneste (NSD). For alle årgangene fikk vi rådata, og det har dermed vært nødvendig med mye forarbeid og rensing. Med over 2500 variabler i hvert datasett, har vi luket ut mange variabler vi ikke har hatt behov for i våre analyser. Dette har vært meget tidkrevende, da variablene sjelden har hatt samme navn fra år til år. En fullstendig oversikt over variablene vi har benyttet i analysen finnes i appendiks A.4.

Før vi aggregerte matvariabelen, som var nødvendig for våre analyser, gjennomgikk vi hver enkelt matvarevariabel og erstattet manglende og negative verdier med null. Dersom den aggregerte variabelen likevel ble lik null etter dette, luket vi ut husholdningene dette gjaldt for. Dette var nødvendig fordi vi bruker logaritmen til denne variabelen i analysedelen.

Den aggregerte matvarevariabelen inneholder alle former for mat fra dagligvarehandel, inkludert mineralvann. Alkoholholdig drikke og tobakk er ekskludert. I tillegg har vi utelatt restaurant- og cafébesøk, da husholdninger vil ha ulikt forbruksmønster her, avhengig av preferanser og prioriteringer.

Husholdninger som mangler verdi for inntektsvariabelen er fjernet fra datasettet. Dette kommer av at denne variabelen er vesentlig for våre analyser. I tillegg til dette har vi kontrollert at det ikke finnes husholdninger med negative verdier for inntektsvariabelen. Vi observerte tre husholdninger fra datasettet med ekstremt høye inntekter (over kr 10 000 000). Disse har vi luket ut, da vi mener de kan gjøre mer skade enn nytte i våre analyser.

I løpet av prosessen har vi sett det nødvendig å ekskludere noen husholdninger fra datasettet som ikke er relevante for vår analyse. Dette gjelder blant annet husholdninger hvor hovedinntektstaker er over 65 år, da vi antar at slike husholdninger nærmer seg pensjonisttilværelsen. Husholdninger hvor hovedinntektstaker ikke er gift, har samboer eller registrert partner ekskluderer vi også. Dette kommer av at vi ønsker å analysere husholdninger med to voksne, i tråd med Pissarides og Weber (1989). Vi ønsker husholdninger hvor

hovedinntektstaker er arbeidstaker eller selvstendig næringsdrivende, og ekskluderer dermed husholdninger hvor vi ikke har informasjon om dette. Husholdninger hvor minst én av hovedpersonene driver med jordbruk blir også fjernet. Dette kommer av at vi antar at denne gruppen har en annen utgiftsfunksjon på mat ettersom de har større tilgang på egenprodusert mat.

Etter rensingen har vi 29 097 husholdninger tilgjengelig. På grunn av manglende informasjon om husholdningsinntekten for 1979, faller denne årgangen bort i analysen. I tillegg til dette forsvinner all informasjon fra 1986 når vi bruker IV-estimering, da opplysninger om en av instrumentvariablene (*boligens samlede nettoflate*) mangler dette året.

5.3 Databehandling

I dette delkapitlet vil vi gi en beskrivelse av hvordan vi har behandlet dataene for å forberede til analysen.

5.3.1 Selvstendig næringsdrivende

Vi har opprettet en dummy-variabel for selvstendig næringsdrivende (*SE*) som vi benytter i regresjonsanalysene. Her har vi valgt å bruke definisjonen ”minst én i husholdningen er selvstendig næringsdrivende”. Pissarides og Weber (1989) tok kun hensyn til hovedinntektstaker når husholdningene ble definert som selvstendig næringsdrivende. Vi har grunn til å tro at begge ”hovedpersonene” i husholdningen har et mer likestilt ansvar til å bidra med inntekt i dag enn hva som var tilfellet på 80-tallet, og tar dermed hensyn til begge i vår definisjon.

I tillegg har vi opprettet tre dummy-variabler for ulike kategorier av selvstendig næringsdrivende. Dette har vi gjort for å kunne undersøke om antall selvstendig næringsdrivende i husholdningen påvirker graden av underrapportering. Her har vi opprettet én for husholdninger hvor kun hovedinntektstaker er selvstendig næringsdrivende, én hvor kun person 2 er selvstendig næringsdrivende, og en der begge er selvstendig næringsdrivende.

5.3.2 Ulike bransjer

Når vi tar for oss ulike bransjer antar vi at også *arbeidstakere* har muligheter til å underrapportere. Av denne grunn oppretter vi en dummy-variabel for arbeidstakere (*EE*), som er lik 1 dersom det ikke er selvstendig næringsdrivende i husholdningen, 0 hvis det er. Denne bruker vi i interaksjonsvariablene beskrevet nedenfor.

For å undersøke hvorvidt underrapporteringen varierer mellom ulike bransjer, opprettet vi først dummy-variabler som er lik 1 dersom hovedinntektstaker er i en gitt bransje (bygg/anlegg (*BA*) eller helse/undervisning (*HU*)), 0 hvis ikke. Deretter opprettet vi interaksjonsvariabler mellom disse variablene og dummy-variablene for selvstendig næringsdrivende (*SE*) og for arbeidstakere (*EE*).

Disse interaksjonsvariablene medfører imidlertid potensielt noen problemer. Når vi kun tar hensyn til hovedinntektstakers bransje, vil vi ikke fange opp potensielle effekter fra person 2 sin bransje. Dette er en svakhet ved denne utredningen.

Et annet mulig problem er at vi vil kunne fange uønskede effekter fra en viss type husholdninger. Dette gjelder husholdninger hvor hovedinntektstaker for eksempel er bygg- og anleggsarbeider, mens person 2 er den som er selvstendig næringsdrivende. Her vil husholdningen defineres som selvstendig næringsdrivende i bygg- og anleggsbransjen, til tross for at den i husholdningen som faktisk er selvstendig næringsdrivende arbeider i en annen bransje. Dette kommer av at vi definerer husholdningene som selvstendig næringsdrivende når minst én av hovedpersonene er det. Det er imidlertid kun 0,19 % av det totale utvalget som faller i denne kategorien, og vi ser dermed bort fra potensielle innvirkninger dette kan ha på resultatene våre.

5.4 Valg av variabler

Vi har valgt å inkludere langt færre forklarings- og instrumentvariabler enn Pissarides og Weber (1989). Dette kommer blant annet av at vi mener deres variabler er noe utdaterte, da studien kun tar utgangspunkt i data fra 1982. I tillegg til dette har vi ikke alle variablene de benytter tilgjengelig i vårt datasett. Pissarides og Weber (1989) skiller ikke tydelig mellom

hva de bruker som forklarings- eller instrumentvariabler. Dermed tar vi videre noen antakelser om hvilke variabler de har benyttet til ulike formål.

5.4.1 Forklaringsvariabler

Vi har valgt å inkludere følgende husholdningskarakteristika i tråd med Pissarides og Weber (1989); *alder, alder kvadrert, antall barn, antall barn kvadrert*. Vi har også inkludert *antall barn under 7 år*, etter idé fra Johansson (2005). Pissarides og Weber (1989) inkluderer i stedet *antall barn mellom 0 og 4 år* og *antall barn mellom 5 og 16 år*. Vi mener *antall barn under 7 år* skiller ”småbarnsfamilier” fra andre husholdninger på en god måte ved at barn i denne alderen stort sett har andre preferanser og behov enn et eldre barn. I tillegg var dette definert som en egen variabel i datasettet vårt.

Pissarides og Weber (1989) har benyttet en del forklaringsvariabler vi ikke har sett nødvendig å inkludere i våre analyser. Vi antar at dette blant annet gjelder dummy-variabler for *geografiske områder, TV, sentralvarme, vaskemaskin* og hvilket *kvartal* intervjuet fant sted. Vi har ikke inkludert *geografiske områder*, da vi ikke har hatt tilstrekkelig informasjon om dette i alle årganger. *TV, sentralvarme* og *vaskemaskin* er variabler vi antar forklarte mer på 80-tallet, da de fleste husholdninger i dag har TV og vaskemaskin, og sentralvarme ikke påvirker husholdningsutgiftene i like stor grad som tidligere.

Vi velger å ikke inkludere dummy-variabler for kvartal, da vi ikke har informasjon om dette for år 2003. I tillegg ser vi i tabellen under at føringsperiodene er relativt tilfeldig spredt utover året.

Tabell 2: *Oversikt over andel selvstendig næringsdrivende og arbeidstakere i hvert kvartal*

	Selvstendig næringsdrivende	Arbeidstakere
Kvartal 1	23,04 %	23,33 %
Kvartal 2	27,65 %	26,32 %
Kvartal 3	21,32 %	21,43 %
Kvartal 4	27,99 %	28,92 %

Fordelingen av kvartaler er basert på når føringsperioden starter.

Problemer ved å ikke inkludere kvartalsdummyer oppstår for eksempel dersom en høy andel selvstendig næringsdrivende svarer på Forbruksundersøkelsen i kvartaler hvor en sesongmessig konsumerer mye mat (jul etc.), og samtidig en høy andel arbeidstakere i

kvartaler hvor en sesongmessig konsumerer lite mat. Ettersom tabell 2 viser at andelen selvstendig næringsdrivende og arbeidstakere er tilnærmet lik for hvert kvartal, ser ikke dette ut til å være tilfellet. Vi antar dermed at manglende kvartalsdummyer ikke vil skape forventningsskjeve estimater.

5.4.2 Instrumentvariabler

Valg av instrumentvariabler har vært utfordrende, da det ikke finnes noen som er ”perfekte” for inntektsvariabelen.

Vi har i hovedsak tatt utgangspunkt i instrumentvariablene som er benyttet av Pissarides og Weber (1989), siden vi følger deres metode, og Johansson (2005), fordi vi mener at gode og forklarende instrumenter er inkludert i denne studien. Vi antar at Pissarides og Weber (1989) blant annet har benyttet følgende instrumentvariabler: *antall biler i husholdningen*, *antall rom i boligen*, dummy-variabler for om nåværende og ”normal” inntekt varierer, om *konen jobber fulltid* og om *konen jobber deltid*, samt flere ulike interaksjonsvariabler med dummy-variabelen for selvstendig næringsdrivende. Johansson (2005) har benyttet følgende instrumentvariabler: *huseierskap*, *antall rom i boligen*, *antall måneder ektefelle har vært i arbeid*, *antall måneder ektefelle har vært i selvstendig næringsvirksomhet*, *antall måneder ektefelle har vært arbeidsløs*, *antall måneder ektefelle har vært i deltidsjobb*, og produktet av dummy-variabelen for selvstendig næringsdrivende med *alder*, *alder kvadrert*, *barn*, *barn kvadrert*, *antall barn under 7 år*, *antall måneder ektefelle har arbeidet i eget hus* og *regioner*.

Etter å ha vurdert nevnte instrumentvariabler, falt valget på å inkludere følgende i vår analyse:

- *Antall arbeidstimer per uke for person 2*
- *Antall arbeidstimer per uke for hovedinntektstaker*
- *Boligens samlede nettoflate*

Vi tar utgangspunkt i at disse variablene korrelerer med husholdningsinntekten, men ikke med matkonsumet.

Antall arbeidstimer per uke for person 2 er noe tilsvarende dummy-variablene Pissarides og Weber (1989) har inkludert. Johansson (2005) har også inkludert en lignende variabel. Vi

mener flere arbeidstimer øker inntektsnivået, og at en slik variabel vil gi en god indikasjon på husholdningens inntekt. Det samme gjelder for variabelen *antall arbeidstimer per uke for hovedinntektstaker*.

Det siste instrumentet vi har valgt å inkludere er *boligens samlede nettoflate*. Vi mener denne variabelen er mer hensiktsmessig enn *antall rom i boligen*, da man kan ha få rom i boligen, men likevel store arealer.

Vi velger å ikke inkludere interaksjonsvariabler med dummy-variabelen for selvstendig næringsdrivende, da vi ikke tror selvstendig næringsdrivende har en egen delvis effekt på faktorer som eksempelvis *alder*. Når vi har undersøkt ulike interaksjonsvariabler, har vi i tillegg ikke fått andre resultater enn de presenterte.

Flere tidligere studier har inkludert utdanningsnivå som en instrumentvariabel. Dette er en variabel som også kunne vært relevant i denne analysen, men som vi ikke har informasjon om for årene 1975-2001. Av denne grunn har vi ikke inkludert denne variabelen.

5.5 Beskrivende statistikk

I dette delkapitlet vil vi presentere og diskutere noe beskrivende statistikk. I tabellen under har vi delt opp i arbeidstakere og ulike definisjoner av selvstendig næringsdrivende.

Tabell 3: *Beskrivende statistikk, 1975-2012*

	Minst én er selvst. næringsdr.	Kun hovedinntektst. er selvst. næringsdr.	Kun person 2 er selvst. næringsdr.	Begge er selvst. næringsdr.	Arbeidstakere
Ant. obs.	3 534	2 401	756	377	25 563
Alder	44,08 (0,18)	44,7 (0,22)	41,48 (0,37)	45,3 (0,54)	42 (0,07)
Antall barn	1,5 (0,02)	1,49 (0,03)	1,53 (0,04)	1,55 (0,07)	1,3 (0,01)
ln(inntekt)	12,32 (0,01)	12,24 (0,02)	12,56 (0,02)	12,35 (0,04)	12,48 (0,01)
ln(matutgift)	10,38 (0,01)	10,32 (0,02)	10,57 (0,02)	10,4 (0,04)	10,36 (0,01)

Standardavvik i parenteser. Alle tall i tabellen, foruten antall observasjoner, er gjennomsnittstall for husholdningen. Alder gjelder kun for hovedinntektstaker.

Tabell 3 viser at 12 % av utvalget vårt er husholdninger hvor minst én er selvstendig næringsdrivende. Av disse er det 68 % hvor kun hovedinntektstaker er selvstendig næringsdrivende, 21 % hvor kun person 2 er det, og 11 % hvor begge er det.

Gjennomsnittstallene rundt inntektsvariabelen fra tabellen gir indikasjoner på at arbeidstakere rapporterer høyere inntekt enn selvstendig næringsdrivende. Dette kan være et tegn på underrapportering av inntekt blant selvstendig næringsdrivende. Det kan også tyde på at arbeidstakere generelt tjener mer. For husholdninger hvor kun person 2 er selvstendig næringsdrivende er snittet for inntektsvariabelen på 12,56, som er det høyeste snittet. Her er hovedinntektstaker arbeidstaker. Den største delen av husholdningsinntekten kommer følgelig fra arbeidstakeren i husholdningen. Det kan derfor være naturlig med høyere inntekt her enn for øvrige kategorier av selvstendig næringsdrivende, til tross for at husholdninger som defineres som selvstendig næringsdrivende kan ha større muligheter for underrapportering.

Vi observerer at gjennomsnittlig matkonsum er på høyde med, eller større, for flere av selvstendig næringsdrivende-kategoriene enn for arbeidstakere. Dette er interessant, da denne gruppen ofte rapporterer lavere inntekt enn arbeidstakerne. Ved å konstruere et 95 %-konfidensintervall (appendiks A.8) for gjennomsnittlig matkonsum, kan vi ikke påvise en signifikant forskjell mellom gruppene, bortsett fra for husholdninger hvor kun person 2 er selvstendig næringsdrivende. Det høye matkonsumet blant selvstendig næringsdrivende kan tyde på et generelt høyere konsum av mat i forhold til arbeidstakere. I tillegg kan det tyde på at denne gruppen bruker ”svarte penger” på mat.

I tabellen nedenfor gir vi en oversikt over antall selvstendig næringsdrivende og arbeidstakere i de ulike bransjene.

Tabell 4: *Oversikt over antall observasjoner i ulike næringsgrupper*

	Totalt	SE	Bygg*SE	Helse*SE	EE	Bygg*EE	Helse*EE
Ant. obs.	29 097	3 534	644	238	25 563	2 459	3 025

Tabell 4 viser at om lag 11 % av det totale utvalget arbeider med bygg/anlegg, hvorav 21 % er selvstendig næringsdrivende og 79 % er arbeidstakere. Helse/undervisning står også for om lag 11 % av det totale utvalget. Her er 7 % selvstendig næringsdrivende og 93 % arbeidstakere.

6. Resultater og analyse

Formålet med dette kapitlet er å drøfte ulike estimater på underrapportering av inntekt i Norge. Vi tar utgangspunkt i datasettet presentert i kapittel 5 og metodedelen fra kapittel 4. Først tar vi for oss underrapporteringen i alle næringer, før vi systematisk diskuterer underproblemstillingene fra kapittel 1.

6.1 Estimat på underrapportering i ulike næringer

For å finne estimater på underrapporteringen, har vi kjørt regresjonsanalyser av utgiftsfunksjonen fra kapittel 4. Resultatene fra estimering ved OLS og to-steps-minste-kvadraters-metode (2SLS) fra IV presenteres i tabell 5. Kolonne 1-2 representerer alle næringer, og er dermed vårt hovedresultat. I kolonne 3-4 er bygg/anlegg (*BA*) inkludert, og i 5-6 er helse/undervisning (*HU*) inkludert. Under tabellen følger diskusjon av resultatene.

Tabell 5: Regresjonsresultater for $\ln(\text{matutgifter})$, 1975-2012

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	2SLS	OLS	2SLS	OLS	2SLS
$\ln(\text{inntekt}) (\beta)$	0,293 (0,009)***	0,600 (0,040)***	0,293 (0,010)***	0,600 (0,040)***	0,293 (0,0095)***	0,607 (0,040)***
SE (γ)	0,042 (0,009)***	0,062 (0,010)***	0,036 (0,010)***	0,047 (0,010)***	0,031 (0,0105)***	0,045 (0,010)***
Bygg*			0,058 (0,021)***	0,106 (0,020)***	0,056 (0,0207)***	0,102 (0,020)***
SE (γ_1)						
Helse/und.*					-0,025 (0,040)	-0,056 (0,040)
SE (γ_2)						
Bygg*			0,039 (0,010)***	0,042 (0,010)***	0,032 (0,010)***	0,035 (0,010)***
EE (δ_1)						
Helse/und.*					-0,045 (0,010)***	-0,046 (0,010)***
EE (δ_2)						
Alder	0,026 (0,003)***	0,003 (0,004)	0,026 (0,003)***	0,003 (0,004)	0,026 (0,003)***	0,003 (0,004)
Alder kvadrert	0,000 (0,000)***	0,000 (0,000)	0,000 (0,000)***	0,000 (0,000)	0,000 (0,000)***	0,000 (0,000)
Antall barn	0,328 (0,009)***	0,278 (0,010)***	0,328 (0,009)***	0,278 (0,010)***	0,327 (0,009)***	0,276 (0,010)***
Ant. barn kvdr.	-0,034 (0,002)***	-0,026 (0,020)***	-0,034 (0,002)***	-0,026 (0,020)***	-0,034 (0,002)***	-0,026 (0,002)***
Ant. barn u 7 år	-0,025 (0,005)***	-0,017 (0,005)***	-0,025 (0,005)***	-0,017 (0,005)***	-0,025 (0,005)***	-0,017 (0,005)***
Konstantledd	5,119 (0,108)***	2,260 (0,377)***	5,102 (0,108)***	2,210 (0,380)***	5,100 (0,108)***	2,180 (0,380)***
Tidstrend (t)	0,050 (0,001)***	0,021 (0,004)***	0,050 (0,001)***	0,020 (0,004)***	0,050 (0,001)***	0,020 (0,004)***
Tid kvdr. (t^2)	-0,001 (0,000)***	0,000 (0,000)***	-0,001 (0,000)***	0,000 (0,000)***	-0,001 (0,000)***	0,000 (0,000)***
Ant. obs.	29 097	28 034	29 097	28 034	29 097	28 034
Forkl.kraft (R^2)	0,5646	0,5394	0,5088	0,5390	0,5652	0,5388
F-test for felles signifikans, p-verdi i parentes		111,50 (0,000)		110,53 (0,000)		110,56 (0,000)
Wu-Hausman-test for eksogenitet av $\ln Y$, p-verdi		0,000		0,000		0,000
Sargan-test for valide instrumenter, p-verdi		0,9267		0,9338		0,9207

Robuste standardavvik i parentesene. * $p < 0,1$; ** $p < 0,05$; *** $p < 0,01$.

Instrumentvariabler for 2SLS; boligens samlede nettoflate, antall arbeidstimer per uke for hovedinntektstaker og antall arbeidstimer per uke for person 2.

Tabell 5 viser resultatene for hele tidsperioden vi har tilgjengelig, 1975 til 2012. Vi observerer at den lineære tidstrenden er positiv og signifikant, noe som tyder på at matkonsumet øker med om lag 5 % hvert år fra 1975 til 2012 (ved OLS-estimering). Valg av tidstrend diskuteres i appendiks A.5.

6.1.1 Alle næringer

Her vil vi ta for oss hovedresultatene fra tabell 5, først ved OLS-estimering (kolonne (1)) og deretter ved IV-estimering (kolonne (2)).

6.1.1.1 OLS-resultater

I kolonne (1) observerer vi at den marginale tilbøyeligheten til å konsumere mat, β , er 0,293 og signifikant på 1 %-nivå. Dette betyr at matkonsumet øker med 0,293 % for hver prosent inntekten øker.

Koeffisienten for selvstendig næringsdrivende, γ , er også signifikant på 1 %-nivå og lik 0,042. Dette indikerer et ”overforbruk” av mat blant selvstendig næringsdrivende på 4,2 %, sammenlignet med arbeidstakere med samme inntektsnivå. Den positive koeffisienten gir også en indikasjon på underrapportering.

Vi vil ikke diskutere de øvrige forklaringsvariablene, da disse ikke er relevante for den videre analysen.

Vi kalkulerer estimatet på k for alle næringer, for å finne det tallet vi må multiplisere selvstendig næringsdrivendes rapporterte inntekt med for å finne ”sann” inntekt:

$$k_{SE} = e^{\frac{\gamma}{\beta}} = e^{\frac{0,042}{0,293}} \approx 1,16.$$

En k lik 1,16 kan tolkes som at husholdninger med minst én selvstendig næringsdrivende underrapporterer sin reelle inntekt med nærmere 14 % ($0,16/1,16$). Ved å se på 95 %-konfidensintervallet for β og γ , finner vi at k kan variere mellom 1,08 og 1,25 (appendiks A.7). Dette indikerer en relativt stor variasjon i estimatet. På grunnlag av dette er det

vanskelig å si noe konkret om størrelsen på k , men resultatene tyder på at den er større enn én, og at denne gruppen dermed underrapporterer sin inntekt.

Vi finner at en underrapportering på 14 % tilsvarer 0,91 % av Norges BNP for 2012, nærmere 27 milliarder NOK. Disse tallene indikerer at hver selvstendig næringsdrivende i snitt underrapporterer mer enn 80 000 kroner av sin reelle inntekt årlig, noe som gir et inntektstap på om lag 7,3 milliarder NOK for staten (appendiks A.6). Estimatenes er dog noe usikre, og vi vil dermed ikke diskutere dette nærmere. (Statistisk Sentralbyrå, 2013b; Statistisk Sentralbyrå, 2015).

Forklaringskraften ligger mellom 50 % og 60 %, noe som tyder på at også andre forklaringsvariabler kan påvirke resultatene våre. Likevel er det ikke forventet å få en mye høyere forklaringskraft i denne typen studier. Tidligere studier har fått en noe lavere forklaringskraft enn hva vi får. Pissarides og Weber (1989), Johansson (2005), Engström og Holmlund (2009) og Kukk og Staehr (2014) fikk en forklaringskraft på henholdsvis 30 %, 33 %, 20 % og 45 %. Dette kan gi en indikasjon på at enkelte valg i denne utredningen har gitt positivt utslag sammenlignet med andre studier. Forklaringskraften er riktignok ikke helt sammenlignbar på tvers av studier. Dette kommer blant annet av at studiene er fra ulike land, og at det benyttes ulikt antall observasjoner, ulike forklaringsvariabler og andre tidsperioder.

6.1.1.2 IV-resultater

Resultatene fra IV-estimeringen finnes i kolonne (2), (4) og (6) i tabell 5.

Testene som presenteres nederst i tabellen blir nærmere forklart i appendiks A.5. Kort fortalt bekrefter Wu-Hausman-testen at det er et endogenitetsproblem i inntektsvariabelen, og følgelig at IV er å foretrekke fremfor OLS. Resultatene fra F-testen forteller oss at de valgte instrumentvariablene er signifikante sammen, og p-verdien fra Sargan-testen indikerer at det ikke er problemer med disse instrumentvariablene.

Resultatene fra 2SLS viser at koeffisientene for inntektsvariabelen, β , og selvstendig næringsdrivende, γ , blir vesentlig høyere ved IV (0,60) enn OLS (0,29) i alle regresjonene. Koeffisientene er signifikante på 1 %-nivå.

Mye tyder på at koeffisienten til inntektsvariabelen er spesielt sensitiv til om vi bruker OLS eller IV, hvor den kan være underestimert ved bruk av OLS. Dette kan føre til en overestimering av k . I følge Hurst et al. (2010) kan dette skyldes forbigående variasjon rundt permanent inntekt, samt målefeil i den rapporterte inntekten. I denne studien kom det også frem at elastisiteten til matkonsumet på inntekten dempes ved OLS-estimering på grunn av at rapportert inntekt kun er et grovt estimat på permanent inntekt.

Fra gjennomgangen av metoden (kapittel 4.2.1) vet vi at det er nødvendig med mer informasjon enn hva vi finner i tabell 5, for å kunne estimere k ved IV. Her har vi behov for variansen til inntektsresidualet for å kunne kalkulere en øvre og en nedre grense.

Tabell 6: *Estimater fra instrumentvariabelmetoden, 1975-2012*

	β	γ	σ_{YSE}^2	σ_{YEE}^2
Alle næringer	0,600 (0,040)***	0,062 (0,010)***	0,284	0,173

Standardavvik i parenteser. γ er koeffisienten foran dummy-variabelen for selvstendig næringsdrivende (kolonne 2, tabell 5), β er marginal tilbøyelighet til å konsumere (kolonne 2, tabell 5), og σ_Y^2 er variansen til inntektsresidualet (som vi har funnet ved å kjøre en regresjon av ligning (16) med samme forklarings- og instrumentvariabler som i tabell 5).

Ved bruk av informasjonen fra tabell 6 og ligning (20) og (21) fra kapittel 4, finner vi nedre og øvre grense slik:

$$\bar{k}_l = e^{\frac{\gamma}{\beta} - \frac{1}{2}(\sigma_{YSE}^2 - \sigma_{YEE}^2)} = e^{\frac{0,062}{0,6} - \frac{1}{2}(0,284 - 0,173)} = 1,05$$

$$\bar{k}_u = e^{\frac{\gamma}{\beta} + \frac{1}{2}(\sigma_{YSE}^2 - \sigma_{YEE}^2)} = e^{\frac{0,062}{0,6} + \frac{1}{2}(0,284 - 0,173)} = 1,17$$

Den øvre grensen på 1,17 er nærmest estimatet fra OLS-estimeringen på 1,16. Dette kan igjen tyde på et overestimert av k ved bruk av OLS. Likevel får vi fremdeles tegn på signifikant underrapportering blant selvstendig næringsdrivende i alle næringer ved bruk av IV.

Til tross for at vi har påvist endogenitet i inntektsvariabelen, er det ikke gitt at inntekten svinger vesentlig mer for selvstendig næringsdrivende enn for arbeidstakere. I så tilfelle får vi at $\sigma_{YSE}^2 = \sigma_{YEE}^2$, og kan regne ut k på samme måte som for OLS:

$$k_{SE} = e^{\frac{0,062}{0,6}} = 1,11.$$

Vi vil ikke kommentere dette estimatet nærmere, da resultatene fra tabell 6 tyder på at inntekten svinger i dette tilfellet. Større forskjell mellom variansen til inntektsresidualene fører imidlertid til større forskjell mellom den øvre og nedre grensen til k .

Tidligere har testresultatene fra tabell 5 indikert at IV er en bedre estimeringsmetode enn OLS, og vi observerer at de to metodene gir ulike resultater. Likevel ønsker vi å fokusere på OLS videre i denne utredningen. Dette kommer blant annet av at det er enklere å estimere k ved bruk av OLS. I tillegg til dette forventer vi ikke at resultatene fra IV og OLS vil variere vesentlig. En annen faktor er at vi ikke kan være sikker på at de valgte instrumentvariablene er ”perfekte” for å representere den permanente inntekten.

6.1.2 Bygg og anlegg

Her vil vi drøfte resultatene for bygg- og anleggsbransjen. Dette er en tjenesteytende bransje med trolig høy kontantbruk, og er dermed spesielt utsatt for underrapportering i følge ØKOKRIM (2012). Resultatene er presentert i tabell 5, kolonne (3). Vi viser først noe beskrivende statistikk fra kapittel 5.5 i tabellen nedenfor for å gi en oversikt over hvor mange husholdninger vi har i denne bransjen. Her, og i kapittel 6.1.3, antar vi at også arbeidstakere kan underrapportere sin inntekt, og vil dermed estimere k for denne gruppen i tillegg til selvstendig næringsdrivende.

Tabell 7: *Oversikt over antall observasjoner i bygg- og anleggsbransjen, 1975-2012*

	Totalt	SE	Bygg*SE	EE	Bygg*EE
Antall observasjoner	29 097	3 534	644	25 563	2 459

Fra tabell 5 observerer vi at begge interaksjonsvariablene for bygg- og anleggsbransjen er positive og signifikante på 1 %-nivå. Variablene indikerer at selvstendig næringsdrivende og arbeidstakere i denne bransjen bruker henholdsvis 5,8 % og 3,9 % mer på mat enn tilsvarende grupper i andre bransjer. Dette kan tyde på underrapportering i begge grupper. Koeffisienten for arbeidstakere på 0,039 er spesielt interessant, da den utgiftsbaserte estimeringsmetoden kun antar underrapportering blant selvstendig næringsdrivende.

Estimat på k for selvstendig næringsdrivende, ekskludert bygg- og anleggsbransjen:

$$k_{SE \text{ (ekskl. BA)}} = e^{\frac{\gamma}{\beta}} = e^{\frac{0,036}{0,293}} \approx 1,13$$

Når vi ekskluderer bygg- og anleggsbransjen fra gruppen med selvstendig næringsdrivende, får vi et lavere estimat enn når alle bransjer er inkludert. Dette kan tyde på at denne bransjen øker snittet på underrapporteringen.

Estimat på k for selvstendig næringsdrivende i bygg- og anleggsbransjen:

$$k_{SE/BA} = e^{\frac{\gamma+\gamma_1}{\beta}} = e^{\frac{0,036+0,058}{0,293}} \approx 1,38$$

Estimatet på 1,38 for selvstendig næringsdrivende i bygg- og anleggsbransjen indikerer at denne gruppen underrapporterer sin reelle inntekt med om lag 28 %.

Estimat på k for arbeidstakere i bygg- og anleggsbransjen:

$$k_{EE/BA} = e^{\frac{\delta_1}{\beta}} = e^{\frac{0,039}{0,293}} \approx 1,14$$

I denne bransjen finner vi at også *arbeidstakere* underrapporterer sin inntekt, med om lag 12 %. Dette estimatet er nesten like høyt som for selvstendig næringsdrivende i alle næringer.

Metoden vi benytter fører til at vi ikke får fanget opp potensielle effekter fra eksempelvis egenproduksjon, en effekt som kan indikere at estimatene vi har fått for denne bransjen er underestimer. Effekten fra potensielle heterogene preferanser kan også gi andre estimater. Ut fra dette vet vi ikke sikkert om estimatene våre er over- eller underestimer.

Kort oppsummert tyder resultatene fra denne delen av analysen på at både arbeidstakere og selvstendig næringsdrivende i bygg- og anleggsbransjen underrapporterer sin inntekt. Her får vi statistisk signifikante resultater for begge gruppene.

6.1.3 Helse og undervisning

Her vil vi ta for oss resultatene for de som arbeider innen helse eller undervisning. Dette er presentert i kolonne (5) i tabell 5. Vi henter også her noe beskrivende statistikk fra kapittel 5.5 for å gi en oversikt over hvor mange husholdninger vi har i denne bransjen.

Tabell 8: *Oversikt over antall observasjoner innenfor helse og undervisning, 1975-2012*

	Totalt	SE	Helse*SE	EE	Helse*EE
Antall observasjoner	29 097	3 534	238	25 563	3 025

Fra tabell 5 observerer vi at koeffisienten for selvstendig næringsdrivende innen helse og undervisning er lik -0,025. En negativ koeffisient kan indikere at denne gruppen bruker mindre andel av sin rapporterte inntekt på mat enn andre selvstendig næringsdrivende (ekskludert bygg- og anleggsbransjen). Koeffisienten er imidlertid ikke signifikant, og vi kan dermed ikke statistisk konkludere med ulikt matforbruk mellom de to gruppene.

For selvstendig næringsdrivende innen helse/undervisning, finner vi k til å bli:

$$k_{SE/HU} = e^{\frac{\gamma + \gamma_2}{\beta}} = e^{\frac{0,031 + (-0,025)}{0,293}} \approx 1,021$$

Til tross for at estimatet er usikkert, kan det likevel gi svake indikasjoner på at denne gruppen rapporterer tilnærmet riktig inntekt. Selvstendig næringsdrivende i helse/undervisning står kun for 0,82 % av det totale utvalget, noe som kan være grunnlaget for det usikre estimatet. Den ikke-signifikante koeffisienten fører til at vi ikke kan konkludere noe annet for denne gruppen enn at det ikke er statistiske tegn på underrapportering.

Koeffisienten til arbeidstakere innen helse og undervisning er også negativ (-0,045). Denne er imidlertid signifikant på 1 %-nivå. Dette indikerer et lavere matforbruk og mindre underrapportering relativt til gjennomsnittet for andre bransjer. Estimatet på underrapportering i denne gruppen blir:

$$k_{EE/HU} = e^{\frac{\delta_2}{\beta}} = e^{\frac{-0,045}{0,293}} \approx 0,86$$

Estimatet indikerer en stor ”negativ” effekt, og at denne gruppen *overrapporterer* sin inntekt. Vi velger å tolke dette som indikasjoner på at arbeidstakere *hvert fall* ikke underrapporterer sin inntekt. Dette kan for eksempel komme av at denne gruppen er spesielt ”ærlig”. På en annen side kan dette komme av at denne gruppen er bedre på å utnytte sitt matbudsjett eller at de har andre preferanser, for eksempel at de ikke prioriterer matkonsum like høyt for en gitt inntekt. Dette er effekter metoden ikke klarer å skille mellom.

Resultatene her gir ingen indikasjoner på at de som arbeider innen helse/undervisning underrapporterer sin inntekt. Dette kan tyde på at denne gruppen faktisk rapporterer sin reelle inntekt.

6.2 Utvikling over tid

Her vil vi undersøke utviklingen av underrapporteringen over tid på to måter. Først vil vi ta for oss estimater for hvert år fra 1975 til 2012. Deretter inkluderer vi en interaksjonsvariabel mellom dummy-variabelen til selvstendig næringsdrivende (SE) og tidsvariabelen (t) i hovedregresjonen fra tabell 5 (kolonne (1)), for å se om denne variabelen er signifikant. Vi går her tilbake til å kun se på underrapportering blant selvstendig næringsdrivende, og inkluderer dermed ikke arbeidstakere i dette delkapitlet.

En av grunnene for at vi ønsker å undersøke tidsutviklingen er innføringen av forhåndsutfylte selvangivelser for arbeidstakere på slutten av 90-tallet. Arbeidstakere fikk da begrensede muligheter for underrapportering. Dette kan ha gitt en økning i estimatet på k for selvstendig næringsdrivende i perioden etter innføringen, ettersom arbeidstakere er vår referansegruppe.

Økt grad av digitalisering og automatisering de senere årene har gitt enklere kontrollarbeid for skattemyndighetene. Det er interessant å undersøke om vi finner indikasjoner på at dette har ført til en nedgang i k . I tillegg er det interessant å se om selvrapportert inntekt før 1996 kan ha hatt en innvirkning.

Tabell 9: Regresjonsresultater for hvert enkelt år, 1975-2012

	ln(inntekt) (β)	SE (γ)	Underrapp. (k)	Antall observasjoner	Antall selvstendig næringsdrivende	Forkl.kraft (R^2)
1975	0,36 (0,06)***	0,12 (0,07)*	1,38	641	73	0,3286
1976	0,25 (0,05)***	-0,04 (0,08)	0,86	677	75	0,3368
1977	0,17 (0,05)***	-0,01 (0,06)	0,96	574	65	0,2017
1978	0,26 (0,07)***	0,02 (0,05)	1,09	725	101	0,2671
1980	0,42 (0,06)***	0,05 (0,07)	1,14	656	81	0,294
1981	0,31 (0,05)***	0,09 (0,06)	1,33	905	96	0,2588
1982	0,30 (0,05)***	0,11 (0,06)*	1,44	882	107	0,2559
1983	0,22 (0,05)***	0,01 (0,06)	1,05	870	93	0,2331
1984	0,22 (0,04)***	0,12 (0,05)***	1,7	856	103	0,2716
1985	0,33 (0,05)***	0,10 (0,05)**	1,35	876	113	0,3006
1986	0,10 (0,04)**	-0,02 (0,04)	0,80	827	176	0,2626
1987	0,08 (0,04)**	0,01 (0,04)	1,11	796	165	0,2284
1988	0,09 (0,04)**	-0,06 (0,04)	0,50	795	171	0,2164
1989	0,11 (0,07)	-0,05 (0,06)	0,66	702	130	0,2267
1990	0,22 (0,05)***	0,10 (0,05)**	1,60	700	111	0,2453
1991	0,22 (0,36)***	0,02 (0,04)	1,11	737	123	0,2922
1992	0,17 (0,03)***	0,01 (0,03)	1,09	985	174	0,2899
1993	0,15 (0,03)***	-0,05 (0,04)	0,73	950	170	0,2343
1994	0,09 (0,03)***	-0,03 (0,04)	0,72	983	179	0,2329
1995	0,08 (0,04)*	-0,08 (0,04)*	0,37	955	187	0,2352
1996	0,40 (0,05)***	-0,01 (0,05)	0,98	944	129	0,4034
1997	0,27 (0,06)***	-0,001 (0,04)	1	890	112	0,4135
1998	0,40 (0,07)***	0,13 (0,06)**	1,38	811	60	0,4246
1999	0,32 (0,05)***	0,06 (0,07)	1,22	807	69	0,399
2000	0,40 (0,07)***	0,16 (0,09)*	1,48	554	40	0,3822
2001	0,38 (0,06)***	-0,06 (0,09)	0,85	724	44	0,4102
2002	0,40 (0,06)***	0,02 (0,06)	1,05	735	51	0,4373
2003	0,42 (0,06)***	-0,05 (0,10)	0,89	799	58	0,3783
2004	0,40 (0,06)***	0,06 (0,05)	1,15	782	62	0,4028
2005	0,38 (0,05)***	0,08 (0,07)	1,22	775	58	0,4604
2006	0,37 (0,06)***	0,10 (0,07)	1,31	700	53	0,3932
2007	0,32 (0,08)***	-0,004 (0,08)	0,99	779	56	0,3343
2008	0,46 (0,07)***	-0,04 (0,10)	0,92	831	57	0,3719
2009	0,47 (0,05)***	0,15 (0,06)**	1,37	792	57	0,4039
2012	0,46 (0,05)***	0,05 (0,04)	1,11	2082	135	0,3219

Robuste standardavvik i parentesene. * $p < 0,1$; ** $p < 0,05$; *** $p < 0,01$. Samme forklaringsvariabler som i tabell 5. Tidstrend er ikke inkludert, da dette er estimater fra år til år.

Tabell 9 viser at det er stor variasjon på k opp gjennom årene, fra 0,37 i 1995 til 1,7 i 1984. Koeffisienten for selvstendig næringsdrivende er sjelden signifikant, noe som blant annet kan skyldes få observasjoner hvert år. Dette danner noe av grunnlaget for at vi ikke kan konkludere noe sikkert om tidsutviklingen fra tabell 9. Resultatene kan tyde på at det er nødvendig med flere observasjoner for å få signifikante koeffisienter.

Til tross for usikre estimater, er det en faktor som kan ha innvirkning på resultatene. Dette er at inntektsvariabelen er ulikt definert for periodene 1975-1985, 1986-1995 og 1996-2012, noe som blir forklart nærmere i kapittel 7.2. For å undersøke effekten av dette kan vi sammenligne koeffisienten til inntektsvariabelen, β , og forklaringskraften, R^2 , for periodene der definisjonen på inntektsbegrepet er ulik. Her observerer vi et relativt klart mønster. For perioden før 1985, da inntektsvariabelen representerte ”inntekt ved statsskatteligningen”, observerer vi at både β og R^2 er relativt høye. I perioden 1986-1995, da inntekten ble definert som ”netto inntekt, sum for alle medlemmer”, synker både koeffisienten og forklaringskraften. Fra 1996-2012 defineres variabelen som ”inntekt etter skatt”. I denne perioden stiger disse to og blir vesentlig høyere enn for de andre periodene. Dette kan tyde på at inntektsvariabelen ikke representerer samme ”type” inntekt over alle årene, noe som er en svakhet. Vi observerer at forklaringskraften er høyest fra 1996 og utover, noe som kan indikere at ”inntekt etter skatt” er det mest korrekte inntektsmålet.

Tabell 10: Regresjonsresultater med tidsutvikling (interaksjonsvariabelen ($SE*t$)), 1975-2012

	OLS
ln(inntekt) (β)	0,293 (0,010)***
Selvstendig næringsdrivende (γ)	0,070 (0,020)***
Selvstendig næringsdrivende* tid	-0,002 (0,001)
Tid (t)	0,050 (0,001)***
Tid kvadrert (t^2)	-0,001 (0,000)***
Antall observasjoner	29 097

Robuste standardavvik i parentesene. * $p < 0,1$; ** $p < 0,05$; *** $p < 0,01$. Samme forklaringsvariabler som i tabell 5.

Fra tabell 10 observerer vi at koeffisienten til interaksjonsvariabelen er negativ, noe som kan tyde på en nedgang i underrapporteringen i løpet av perioden. Koeffisienten er riktignok ikke signifikant, og vi kan dermed ikke konkludere noe om utviklingen av underrapporteringen over tid fra disse resultatene heller.

6.3 Påvirkning fra ulike skattereformer

For å undersøke om endringer i skattesystemet har hatt innvirkning på underrapporteringen blant selvstendig næringsdrivende, har vi valgt å se på ulike perioder. Først vil vi ta for oss perioden før og etter innføring av 1992-reformen. Deretter vil vi dele opp i periodene 1975-1992, 1993-2006 og 2007-2012, for å se om 2006-reformen har påvirket estimatet. Vi inkluderer her interaksjonsvariablene for bygg- og anleggsbransjen, da vi tidligere har funnet sterke indikasjoner på at denne gruppen har en høyere k enn andre næringer.

Tabell 11: Regresjonsresultater og estimat på underrapp. før og etter skattereformen i 1992

	1975-1992	1993-2012
ln(inntekt) (β)	0,232 (0,010)***	0,340 (0,010)***
Selvstendig næringsdrivende (γ)	0,038 (0,014)***	-0,005 (0,015)
Bygg*	0,036	0,090
SE (γ_1)	(0,026)	(0,030)***
Bygg*	0,052	0,030
EE (δ_1)	(0,013)***	(0,016)*
Antall observasjoner	13 204	15 893
Antall selvstendig næringsdrivende	1 957	1 577
Forklaringskraft (R^2)	0,5938	0,3938
$k_{SE} = e^{\frac{\gamma}{\beta}}$	1,18	0,99
$k_{BA/SE} = e^{\frac{\gamma_1 + \gamma}{\beta}}$	1,38	1,28
$k_{BA/EE} = e^{\frac{\delta_1}{\beta}}$	1,25	1,09

Robuste standardavvik i parentesene. * $p < 0,1$; ** $p < 0,05$; *** $p < 0,01$. Samme forklaringsvariabler som i tabell 5. Tidstrend (lineær og kvadratisk) er også inkludert.

Fra tabell 11 observerer vi at det er relativt stor forskjell på estimatene for β og γ før og etter skattereformen i 1992. Til tross for den positive tidstrenden i tabell 5, er dette noe overraskende. Spesielt gjelder dette for koeffisienten til inntektsvariabelen. Det er ingen grunn til å tro at innføring av skattereformen har ført til vesentlig høyere inntekt. En av

forklaringene på denne økningen kan være de ulike inntektsmålene som ble beskrevet kort i kapittel 6.2. Endrede preferanser i løpet av perioden kan også være med på å bidra til dette.

For første periode får vi sterke indikasjoner på underrapportering blant selvstendig næringsdrivende som ikke arbeider med bygg/anlegg. Estimaten på $-0,005$ for γ den siste perioden er ikke signifikant, noe som blant annet kan skyldes tilfeldigheter. Dette gjør at vi ikke kan si noe sikkert om underrapporteringen i perioden etter 1992.

I bygg- og anleggsbransjen er kun koeffisienten for arbeidstakere signifikant før 1992, mens begge blir signifikante på minst 10 %-nivå etter 1992. Resultatene kan indikere at denne gruppen underrapporterte mindre etter at skattereformen ble innført.

Til tross for usikre estimater, kan et lavere estimat på k etter 1992 gi en svak indikasjon på at innføring av denne reformen reduserte underrapporteringen, spesielt for bygg- og anleggsarbeidere. Dette kan blant annet skyldes reduserte insentiver ved at lavere skatter og bedre fordeling av skattebetalingene ble innført. Delingsmodellen kan også ha bidratt til reduserte insentiver på grunn av dens gunstige beskatning for de som nå var blitt ikke-aktive bedriftseiere (på marginen).

Likevel kan vi ikke konkludere noe om påvirkningen av denne reformen på grunn av ikke-signifikante koeffisienter. Forklaringskraften for perioden 1993-2012 er også vesentlig lavere enn perioden før, noe som gir enda en indikasjon på usikre estimater.

Vi vil nå undersøke videre om vi kan finne signifikante forskjeller i estimatene ved å også inkludere skattereformen fra 2006.

Tabell 12: Regresjonsresultater og estimat på underrapportering for tre perioder

	1975-1992	1993-2006	2007-2012
ln(inntekt) (β)	0,232 (0,010)***	0,309 (0,016)***	0,430 (0,030)***
Selvstendig næringsdrivende (γ)	0,038 (0,014)***	-0,013 (0,017)	0,028 (0,030)
Bygg*	0,036 (0,026)	0,085 (0,034)**	0,140 (0,097)
SE (γ_1)			
Bygg*	0,052 (0,013)***	0,033 (0,017)*	0,020 (0,040)
EE (δ_1)			
Antall observasjoner	13 204	11 409	4 484
Antall selvstendig næringsdrivende	1 957	1 272	305
Forklaringskraft (R^2)	0,5093	0,3737	0,3491
$k_{SE} = e^{\frac{\gamma}{\beta}}$	1,18	0,96	1,07
$k_{BA/SE} = e^{\frac{\gamma_1 + \gamma}{\beta}}$	1,38	1,26	1,48
$k_{BA/EE} = e^{\frac{\delta_1}{\beta}}$	1,25	1,11	1,05

Robuste standardavvik i parentesene. * $p < 0,1$; ** $p < 0,05$; *** $p < 0,01$. Samme forklaringsvariabler som i tabell 5. Tidstrend (lineær og kvadratisk) er også inkludert.

For de to periodene etter 1992 er γ ikke-signifikant. Det er også usikre estimater for bygg- og anleggsbransjen. Dette kan skyldes få observasjoner eller ren tilfeldighet. Vi observerer at forklaringskraften er vesentlig lavere for de to siste periodene.

Estimatet på k faller relativt mye for alle næringer etter 1992, fra 1,18 til 0,96. Fra 2007 øker k til å bli 1,07. Til tross for ikke-signifikante koeffisienter, diskuterer vi noe av grunnlaget for hvorfor det virker rimelig med nedgang i k etter 1992-reformen og oppgang etter 2006-reformen.

Før 1992 førte høye og formelle skattesatser til at store deler av inntekten gikk ”tapt” til skatter, noe som kan ha ført til høyere insentiver for underrapportering. I tillegg til dette var det enklere for selvstendig næringsdrivende å rapportere sin inntekt som annen ikke-skattepliktig inntekt eller benytte seg av skatteutsettelse over en lengre periode.

Insentivene for underrapportering kan ha økt noe etter innføring av ny reform i 2006. Dette kan blant annet skyldes redusert skattetilpasning, som var et nøkkelord i denne reformen. I tillegg til dette kan en årsak være at det ble vanskeligere for bedriftseiere (og andre

selvstendig næringsdrivende) å unngå de høye marginals kattene, da delingsmodellen ble opphevet og erstattet av blant annet deltakermodellen.

På bakgrunn av ikke-signifikante resultater, kan vi ikke konkludere om endring i skattesystemet har hatt en direkte effekt på underrapporteringen.

6.4 Antall selvstendig næringsdrivende i husholdningen

Resultatene vi har presentert hittil gjelder for vår hoveddefinisjon av selvstendig næringsdrivende; at minst én i husholdningen er det. Videre er det også interessant å undersøke om *antall* selvstendig næringsdrivende i husholdningen har betydning for graden av underrapportering.

I tabell 13, kolonne (2), presenterer vi regresjonsresultater for ulike definisjoner av selvstendig næringsdrivende; kun hovedinntektstaker er selvstendig næringsdrivende (*SE1*), kun person 2 er selvstendig næringsdrivende (*SE2*) og begge personene er selvstendig næringsdrivende (*SE3*). I kolonne (1) er hovedresultatene fra tabell 5 presentert, med definisjonen ”minst én selvstendig næringsdrivende”.

Tabell 13: Regresjonsresultater for ulike definisjoner av SE, 1975-2012

	(1) Hovedreg. (tabell 5)	(2) Ulike definisjoner av SE
ln(inntekt) (β)	0,293 (0,010)***	0,321 (0,011)***
Minst én (γ)	0,042 (0,009)***	
Kun hovedinnt. (γ_{SE1})		0,033 (0,012)***
Kun person 2 (γ_{SE2})		0,092 (0,017)***
Begge (γ_{SE3})		0,035 (0,030)
Antall observasjoner	29 097	25 973
Forklaringskraft (R^2)	0,5646	0,5350

Robuste standardavvik i parentesene. * $p < 0,1$; ** $p < 0,05$; *** $p < 0,01$. Samme forklaringsvariabler som i tabell 5. Tidstrend (lineær og kvadratisk) er også inkludert.

Her har vi hentet en oversikt over antall observasjoner fra kapittel 5.5:

Tabell 14: *Oversikt over antall selvstendig næringsdrivende i ulike kategorier, 1975-2012*

Ant. obs.	Kun			Begge	Totalt EE og SE
	Minst én	hovedinntektstaker	Kun person 2		
	3 534	2 401	756	377	29 097

Fra kolonne (2) i tabell 13 observerer vi at koeffisienten til inntektsvariabelen på 0,321 er noe høyere enn ved den opprinnelige definisjonen på 0,293.

Vi observerer imidlertid at det er færre observasjoner i kolonne (2), noe som kan ha innvirkning på resultatene. Grunnen for dette er 3 124 manglende verdier for ansettelsesforholdet til person 2. I disse husholdningene har vi ikke tilstrekkelig informasjon om person 2 er selvstendig næringsdrivende, arbeidstaker eller familiemedlem. Dette kan for eksempel skyldes at person 2 er hjemmeværende kone og dermed ikke oppfyller noen av kategoriene, eller at husholdningen har glemt å fylle inn denne informasjonen. Vi får da like mange manglende verdier for variabelen *SE2* (kun person 2 er selvstendig næringsdrivende). Vi har valgt å utelate disse husholdningene fra analysen, da vi ikke ønsker å anta noe om hvilken kategori person 2 faller inn under. Dette kan ha betydning for at vi får ulik β og forklaringskraft i de to kolonnene.

Koeffisienten for de ulike definisjonene av selvstendig næringsdrivende er likevel høyest for ”kun person 2 er selvstendig næringsdrivende”, og estimatet på 0,092 er signifikant. Dette tyder på at det er høyest grad av underrapportering i disse husholdningene, sammenlignet med andre kategorier av selvstendig næringsdrivende. Dette er noe overraskende, da person 2 har lavest inntjening til husholdningsinntekten. Vi observerer for øvrig to avvikende husholdninger i datasettet som kan ha innvirkning på disse resultatene; i 1992 finner vi en husholdning bestående av tre personer med samlet husholdningsinntekt på kr 273 609 og et samlet matforbruk på kr 247 612. I 1999 finner vi en husholdning med fire personer som har inntekt på kr 222 167 og matforbruk på kr 204 704. Disse husholdningene bruker over 90 % av sin samlede inntekt på mat, noe som kan være med på å bidra til forventningsskjevne estimater.

Vi observerer relativt liten forskjell mellom koeffisientene for ”kun hovedinntektstaker er SE” (0,033) og ”begge er SE” (0,035), der sistnevnte ikke er signifikant. Dette kan gi svake

indikasjoner på at flere selvstendig næringsdrivende i husholdningen ikke nødvendigvis øker graden av underrapportering.

Ut fra resultatene i kolonne (2) finner vi følgende estimater på k :

Kun hovedinntektstaker er selvstendig næringsdrivende:

$$k_{SE1} = e^{\frac{\gamma_{SE1}}{\beta}} = e^{\frac{0,033}{0,321}} \approx 1,11$$

Kun person 2 er selvstendig næringsdrivende:

$$k_{SE2} = e^{\frac{\gamma_{SE2}}{\beta}} = e^{\frac{0,0917}{0,321}} \approx 1,33$$

Begge hovedpersonene er selvstendig næringsdrivende:

$$k_{SE3} = e^{\frac{\gamma_{SE3}}{\beta}} = e^{\frac{0,035}{0,321}} \approx 1,12$$

Estimatene tyder på at husholdninger hvor kun hovedinntektstaker er selvstendig næringsdrivende underrapporterer sin inntekt med om lag 10 %. Dette indikerer at også person 2 er med på å påvirke graden av underrapportering, da vi finner et estimat på 14 % når vi bruker definisjonen ”minst én selvstendig næringsdrivende”.

Resultatene indikerer også at husholdninger hvor hovedinntektstaker er arbeidstaker og person 2 er selvstendig næringsdrivende underrapporterer sin inntekt med 25 %. Dette kan komme av tilfeldige årsaker. En annen faktor som kan ha innvirkning er person 2 sitt yrke, noe vi ikke har tatt hensyn til i denne utredningen på grunn av fokuset på hovedinntektstaker.

For husholdninger hvor begge hovedpersonene er selvstendig næringsdrivende får vi usikre estimater, noe som kan skyldes få observasjoner. Ut fra diskusjonen kan vi dermed ikke konkludere noe om hvorvidt antall selvstendig næringsdrivende i husholdningen har innvirkning på graden av underrapportering.

7. Avslutning

Her presenterer vi vår konklusjon, samt svakheter ved analysen.

7.1 Konklusjon

I denne utredningen har vi funnet et estimat på k , som representerer det tallet selvstendig næringsdrivende må multiplisere sin rapporterte inntekt med for å komme frem til den ”sanne” inntekten. Dette har vi gjort basert på en utgiftsbasert estimeringsmetode.

Våre hovedfunn indikerer at det eksisterer en svart økonomi i Norge som skyldes underrapportering av inntekt blant selvstendig næringsdrivende. For alle næringer sett under ett, finner vi et estimat på 1,16. Dette tilsvarer en underrapportering på nærmere 14 % av reell inntekt.

Videre finner vi sterke indikasjoner på at de som arbeider i bygg- og anleggsbransjen underrapporterer større grad av sin inntekt enn andre bransjer. Dette gjelder både for selvstendig næringsdrivende og arbeidstakere. De to gruppene underrapporterer henholdsvis 28 % og 12 % av sin reelle inntekt.

Resultatene tyder på at de som arbeider innen helse og undervisning faktisk rapporterer sin reelle inntekt. Vi finner ingen statistiske bevis på underrapportering i denne gruppen.

Vi kan ikke konkludere noe sikkert om den generelle tidsutviklingen av underrapporteringen, og har heller ikke fanget opp en direkte effekt som følge av endringer i skattereformene.

Videre tyder ikke resultatene på at antall selvstendig næringsdrivende i en husholdning påvirker graden av underrapportering.

7.2 Svakheter ved analysen

Med 29 097 husholdninger tilgjengelig, har vi en relativt stor datamengde. I snitt er det om lag 765 observasjoner årlig. Likevel indikerer noen av resultatene at større datamengde er nødvendig for å få signifikante resultater.

Det er noen svakheter tilknyttet Forbruksundersøkelsen. Den kan blant annet inneholde frafallsfeil, utvalgsfeil og føringsfeil. Frafallsfeil kommer av at nettoutvalget som svarer på undersøkelsen ikke representerer hele befolkningen, da det kun er bruttoutvalget som er trukket ut helt tilfeldig. Utvalgsfeil kommer fra variansen som oppstår fra blant annet lengden av registreringsperioden og størrelsen på utvalget. Føringsfeil er feilene som forekommer i forbindelse med innsamling eller utarbeidelsen av dataene. (Statistisk Sentralbyrå, 2013a).

Inntektsvariabelen kan også føre med seg noen svakheter, da den er ulikt definert over perioden 1975-2012. Før 1985 var det stort sett skattevariabler fra ligningsregisteret som var tilgjengelig, og vi benyttet variabelen ”inntekt ved statsskatteligningen”. Dette er et nettoinntektsbegrep (skattepliktig inntekt minus inntektsfradrag) som, i følge Statistisk Sentralbyrå (e-post, 4. mars 2015), ikke nødvendigvis sier så mye om den disponible inntekten til hver husholdning. For årene 1986-1995 har vi tatt i bruk variabelen ”netto inntekt, sum for alle medlemmer”, og fra og med 1996 har vi benyttet variabelen ”inntekt etter skatt”. Vi har ingen informasjon om hvordan disse variablene skiller seg fra hverandre.

Det var to tilgjengelige inntektsvariabler fra samleundersøkelsen i 2012; en for 2011 og en for 2012 (2010 var ikke inkludert). Vi har benyttet snittet av disse to. Matkonsumet som er tilgjengelig i dette datasettet gjelder imidlertid kun for 2012. Vi får dermed ikke målt inntekten direkte mot matkonsumet hvert år. Ettersom datasettet fra 2012 er såpass stort, har vi likevel inkludert dette på lik linje med andre årganger for å få mest mulig data.

8. Litteraturliste

Allingham, M. G & Sandmo, A. (1972) *Income Tax Evasion: A Theoretical Analysis* [Internett]. Side 323-338 i *Journal of Public Economics 1* (1972). Tilgjengelig fra: <<http://www3.nccu.edu.tw/~klueng/tax%20paper/1.pdf>> [Lest: 28.01.15]

Benedek, D. & Lelkes, O. (2011) *The Distributional Implications of Income Under-Reporting in Hungary* [Internett]. Side 539-560 i *Fiscal Studies* (2011), Vol. 32, No. 4. Tilgjengelig fra: <<http://onlinelibrary.wiley.com/doi/10.1111/j.1475-5890.2011.00150.x/abstract;jsessionid=DC8CFC99DE4A50A086E7EDE48D8A2F2A.f04t03>> [Lest: 16.05.15]

Bradbury, B. (1997) *The Living Standards of the Low Income Self-Employed* [Internett]. Side 374-389 i *The Australian Economic Review* (1997), vol. 30, No. 4. Tilgjengelig fra: <<http://onlinelibrary.wiley.com/doi/10.1111/1467-8462.304036/abstract>> [Lest: 05.05.15]

Christiansen, V. (2004) *Norwegian Income Tax Reforms* [Internett]. Side 9-14 i *CESifo DICE Report* (2004), No. 3. Tilgjengelig fra: <<http://core.ac.uk/download/pdf/6631096.pdf>> [Lest: 16.02.15]

Davis, S. J. & Henrekson, M. (2004) *Tax Effect on Work Activity, Industry Mix and Shadow Economy Size: Evidence from Rich-Country Comparisons* [Internett]. *NBER Working Paper Series* (2004), No. 10509. Tilgjengelig fra: <<http://www.nber.org/papers/w10509>> [Lest: 29.04.15]

Davutyan, N. (2008) *Estimating the Size of Turkey's Informal Sector: an Expenditure-Based Approach* [Internett]. Side 261-271 i *Journal of Economic Policy Reform* (2008), 4. utg., volum 11. Tilgjengelig fra: <http://www.erf.org.eg/CMS/uploads/pdf/1211433730_403.pdf> [Lest: 03.03.15]

Eide, E. (2000) *Oversikt over litteratur om svart arbeid og skatteunndragelser* [Internett]. Fra prosjektet *Svart økonomi i Norge* (2000), No. 6, Stiftelsen Frischsenteret for samfunnsøkonomisk forskning. Tilgjengelig fra:

<http://www.frisch.uio.no/publikasjoner/pdf/rapp00_06.pdf> [Lest: 13.02.15]

Engström, P. & Holmlund, B. (2009) *Tax evasion and self-employment in a high-tax country: evidence from Sweden* [Internett]. Side 2419-2430 i *Applied Economics*, vol. 41. Tilgjengelig fra: <<http://www.tandfonline.com/doi/full/10.1080/00036840701765452>> [Lest: 03.10.14]

Finansdepartementet. (2011) *Evaluering av skattereformen 2006* [Internett], Regjeringen.no. Tilgjengelig fra: <<https://www.regjeringen.no/nb/aktuelt/evaluering-av-skattereformen-2006/id637651/>> [Lest: 12.02.15]

Fortin, B., Lacroix, G & Pinard, D. (2010) *Evaluation of the Underground Economy in Quebec: A Microeconomic Approach* [Internett]. Side 463-479 i *International Economic Journal* (2010), Vol. 24, No. 4. Tilgjengelig fra: <<http://www.tandfonline.com/doi/abs/10.1080/10168737.2010.525983>> [Lest: 05.05.15]

Gabrielsen, I. (1992) *Det norske skattesystemet 1992* [Internett], *Sosiale og økonomiske studier* (1992), Statistisk Sentralbyrå. Tilgjengelig fra: <https://www.ssb.no/a/histstat/sos/sos_079.pdf> [Lest: 16.02.15]

Goldstein, H., Hansen, W. G., Ognedal, T. & Strøm, S. (2002) *Svart arbeid fra 1980 til 2001* [Internett]. Fra prosjektet *Svart økonomi i Norge* (2002), No. 3, Stiftelsen Frischsenteret for samfunnsøkonomisk forskning. Tilgjengelig fra: <https://www.frisch.uio.no/publikasjoner/pdf/rapp02_03.pdf> [Lest: 11.02.15]

Hurst, E., Li, G. & Pugsley, B. (2010) *Are Household Surveys Like Tax-Forms: Evidence from Income Underreporting of the Self Employed* [Internett]. *NBER Working Paper Series* (2010), No. 16527. Tilgjengelig fra: <<http://core.ac.uk/download/pdf/6477024.pdf>> [Lest: 06.05.15]

Isachsen, A. J. & Strøm, S. (1985) *The Size and Growth of the Hidden Economy in Norway* [Internett]. Side 21-38 i *Review of Income and Wealth* (1985), Vol. 31. Tilgjengelig fra: <<http://onlinelibrary.wiley.com/doi/10.1111/j.1475-4991.1985.tb00496.x/abstract>> [Lest: 09.01.15]

Johansson, E. (2005) *An estimate of Self-employment income underreporting in Finland* [Internett]. Side 99-109 i *Nordic Journal of Political Economy* (2005), Vol. 31. Tilgjengelig fra: <http://www.nopecjournal.org/NOPEC_2005_a05.pdf> [Lest: 06.10.14]

Kleven, H. J., Knudsen, M. B., Kreiner, C. T., Pedersen, S & Saez, E. (2011) *Unwilling or Unable to Cheat? Evidence From a Tax Audit Experiment in Denmark* [Internett], side 651-692 i *Econometrica* (2011), Vol. 79, No. 3. Tilgjengelig fra: <[http://darp.lse.ac.uk/papersdb/Kleven_etal_\(Ecm11\).pdf](http://darp.lse.ac.uk/papersdb/Kleven_etal_(Ecm11).pdf)> [Lest: 29.01.15]

Kukk, M & Staehr, K. (2014) *Identification of Income Underreporting by the Self-Employed: Employment Status og Reported Business Income?* [Internett]. *TUTECON Working Paper* (2014), No. 1. Tilgjengelig fra: <<http://tutecon.eu/index.php/TUTECON/article/view/8>> [Lest: 07.05.15]

Lyssiotou, P., Pashardes, P. & Stengos T. (2004) *Estimates of the Black Economy Based on Consumer Demand Approaches* [Internett]. Side 622-640 i *The Economic Journal* (2004), Vol. 114. Tilgjengelig fra: <<http://onlinelibrary.wiley.com/doi/10.1111/j.1468-0297.2004.00234.x/abstract?deniedAccessCustomisedMessage=&userIsAuthenticated=false>> [Lest: 06.05.15]

Martinez-Lopez, D. (2013) *The underreporting of income by self-employed workers in Spain* [Internett]. Side 353-371 i *SERIEs* (2013), Vol. 4, No. 4. Tilgjengelig fra: <<http://link.springer.com/article/10.1007/s13209-012-0093-8>> [Lest: 06.05.15]

Meld. St. 11 (2010-2011). (2011) *Evaluering av skattereformen 2006* [Internett], Regjeringen.no. Tilgjengelig fra: <<https://www.regjeringen.no/contentassets/4402291a902b4d129854f4fe447b56ce/no/pdfs/stm201020110011000dddpdfs.pdf>> [Lest: 13.02.15]

Mirus, R. & Smith, R. S. (1989) *Canada's Underground Economy* [Internett]. Side 267-280 i *The Underground Economies* (1989). Tilgjengelig fra:

<<https://books.google.dk/books?hl=no&lr=&id=CmbKloygryEC&oi=fnd&pg=PA267&dq=Canada's+Underground+Economy&ots=aqpVBNbdQm&sig=8WzCZDRm7N-0dNfmpCOJ7gxuwV0#v=onepage&q=Canada's%20Underground%20Economy&f=false>> [Lest: 26.01.15]

Murashov, Y. (2014) *Estimation of the size of informal economy in Russian Federation based on household budget survey data* [Internett], HAL archives-ouvertes. Tilgjengelig fra:

<<http://dumas.ccsd.cnrs.fr/dumas-01095923/document>> [Lest: 08.05.15]

Norsk samfunnsvitenskapelig datatjeneste (NSD). (2009) *Datasett: Forbruksundersøkelsen, 2009* [Internett], NSD. Tilgjengelig fra:

<<http://nsddata.nsd.uib.no/webview/?submode=abstract&mode=documentation&study=http://nsddata.nsd.uib.no/obj/fStudy/NSD1906&mode=documentation>> [Lest: 20.01.15]

Norsk samfunnsvitenskapelig datatjeneste (NSD). (2015) *Forbruksundersøkelsen (1975-2012)*. Utlevert til NSD fra Statistisk sentralbyrå, gjort tilgjengelig av Gry Henriksen, NSD, Bergen

NOU 1999:7. (1999) *Flatere skatt* [Internett], Regjeringen.no. Tilgjengelig fra:

<http://omega.regjeringen.no/Rpub/NOU/19991999/007/PDFA/NOU199919990007000DDD_PDFA.pdf> [Lest: 12.02.15]

Nygård, O. E. & Thoresen, T. O. (2014) *The expenditure approach and hidden economy analysis*. Upublisert presentasjon i Stockholm (november, 2014) i forbindelse med forskning for Statistisk Sentralbyrå.

Obwona, M. (1999) *Estimating unreported income of the self-employed and tax-evasion in Uganda: An expenditure-based approach* [Internett]. Research Series No. 9, Economic Policy Research Center. Tilgjengelig fra:

<http://www.eprc.or.ug/pdf_files/researchseries/series9.pdf> [Lest: 05.05.15]

Pedersen, S. (1998) *The Shadow Economy in Western Europe, Measurement and Results for Selected Countries* [Internett]. The Rockwool Foundation Research Unit, No. 5. Tilgjengelig fra: <http://www.rff.dk/files/RFF-site/Publikations%20upload/Study%20Papers%20-%20uk%20site/The%20Shadow%20Economy%20in%20Western%20Europe.%20Measurement%20and%20Results%20for%20Selected%20Countries.study_5.pdf> [Lest: 15.05.15]

Pissarides, C. & Weber, G. (1989) *An expenditure – based estimate of Britain's black economy* [Internett]. Side 17-32 i *Journal of Public Economics* (1989), Vol. 39. Tilgjengelig fra: <<http://www.sciencedirect.com/science/article/pii/0047272789900522>> [Lest: 10.10.14]

Rasmussen, S. & Rogde, H. (2013) *Omfanget av skyggeøkonomien: en oversikt over litteratur og metoder*, Masterutredning i Økonomisk Styring ved Norges Handelshøyskole, Bergen

Schjelderup, G. (2014) Forelesning 1 FIE432 (20.08.14): *Det norske skattesystemet*, Bergen, Norges Handelshøyskole

Schneider, F. (2002) *The Size and Development of the Shadow Economies of 22 Transition and 21 OECD Countries* [Internett]. *IZA Discussion paper series* (2002), No. 154. Tilgjengelig fra: <<http://www.econstor.eu/bitstream/10419/21427/1/dp514.pdf>> [Lest: 23.01.15]

Schneider, F. (2007) *Shadow Economies and Corruption All Over the World: New Estimates for 145 Countries* [Internett]. *Economics* (2007), No. 9. Tilgjengelig fra: <http://www.economics-ejournal.org/economics/journalarticles/2007-9/version_1> [Lest: 22.01.15]

Schneider, F. (2013) *The Shadow Economy in Europe, 2013* [Internett]. Et samarbeid med Visa Europe og A.T.Kearny. Tilgjengelig fra: <<http://www.atkearney.com/documents/10192/1743816/The+Shadow+Economy+in+Europe+2013.pdf>> [Lest: 23.01.15]

Schuetze, H. J. (2002) *Profiles of Tax Non-Compliance Among the Self-Employed in Canada: 1969 to 1992* [Internett]. Side 219-238 i *Canadian Public Policy* (2002), Vol. 28, No. 2.

Tilgjengelig fra:

<<http://www.jstor.org/discover/10.2307/3552326?uid=2&uid=4&sid=21106785557593>>

[Lest: 08.05.15]

Skatteetaten. (2015) *Grensedragningen mellom arbeidstakere og selvstendig næringsdrivende* [Internett], Skatteetaten. Tilgjengelig fra:

<<http://www.skatteetaten.no/nn/Radgivar/Rettskjelder/Frasegner/Aktuelle-skattmessige-tilpassingar/Grensedragningen-mellom-arbeidstakere-og-selvstendig-naringsdrivende-/>>

[Lest: 16.04.15]

Statistisk Sentralbyrå (SSB). (2013a) *Forbruksundersøkelsen, 2012* [Internett], SSB.no.

Tilgjengelig fra: <[https://ssb.no/inntekt-og-forbruk/statistikker/fbu/aar/2013-12-](https://ssb.no/inntekt-og-forbruk/statistikker/fbu/aar/2013-12-17?fane=om#content)

[17?fane=om#content](https://ssb.no/inntekt-og-forbruk/statistikker/fbu/aar/2013-12-17?fane=om#content)> [Lest: 20.01.15]

Statistisk Sentralbyrå (SSB). (2013b) *Inntekter, personlig næringsdrivende, 2012* [Internett],

SSB.no. Tilgjengelig fra: <[https://www.ssb.no/inntekt-og-forbruk/statistikker/ifpn/aar/2013-](https://www.ssb.no/inntekt-og-forbruk/statistikker/ifpn/aar/2013-12-17)

[12-17](https://www.ssb.no/inntekt-og-forbruk/statistikker/ifpn/aar/2013-12-17)> [Lest: 19.05.15]

Statistisk Sentralbyrå (SSB). (2015) *Årlig nasjonalregnskap, 2014* [Internett], SSB.no.

Tilgjengelig fra: <[http://ssb.no/nasjonalregnskap-og-konjunkturer/statistikker/nr/aar/2015-02-](http://ssb.no/nasjonalregnskap-og-konjunkturer/statistikker/nr/aar/2015-02-11?fane=tabell&sort=nummer&tabell=217894)

[11?fane=tabell&sort=nummer&tabell=217894](http://ssb.no/nasjonalregnskap-og-konjunkturer/statistikker/nr/aar/2015-02-11?fane=tabell&sort=nummer&tabell=217894)> [Lest: 19.05.15]

Steineke, J. M. (2000) *Regional development and political response: local, regional og sectoral differentiation in Norway's shadow economy* [Internett]. Artikkel i forbindelse med presentasjon under "European Urban and Regional Studies Conference" 14.-17. september 2000. Tilgjengelig fra:

<[http://gammelweb.iris.no/internet/student.nsf/199f312efd2a0cacc125680e00635b85/0df9ee4e90043564c1257b2b002c3a4a/\\$FILE/_final_.PDF](http://gammelweb.iris.no/internet/student.nsf/199f312efd2a0cacc125680e00635b85/0df9ee4e90043564c1257b2b002c3a4a/$FILE/_final_.PDF)> [Lest: 05.05.15]

Tedds, L. M. (2005) *Nonparametric Expenditure-Based Estimation of Income Under-Reporting and the Underground Economy* [Internett]. Department of Economics, McMaster University. Tilgjengelig fra:

<http://www.researchgate.net/profile/Lindsay_Tedds/publication/5169249_Nonparametric_expenditure-based_estimation_of_income_under-reporting_and_the_underground_economy/links/00b49524d81b4cfda0000000.pdf> [Lest: 06.05.15]

Torero, M., De La Roca, J., Hernandez, M., Robles, M. & Webber, M. (2006) *The Informal Sector in Jamaica* [Internett]. Rapport i samarbeid med Inter-American Development Bank. Tilgjengelig fra:

<<http://publications.iadb.org/bitstream/handle/11319/4326/The%20Informal%20Sector%20in%20Jamaica.pdf?sequence=1>> [Lest: 07.05.2015]

Wangen, K. R. (2005) *An Expenditure Based Estimate of Britian's Black Economy Revisited* [Internett. Discussion Paper No. 414, Statistisk Sentralbyrå. Tilgjengelig fra:

<<http://brage.bibsys.no/xmlui/bitstream/handle/11250/180919/dp414.pdf?sequence=1>> [Lest: 25.02.15]

Woolridge, J. (2012) *Introductory Econometrics A Modern Approach*. 5 utg. Mason, South-Western, Cengage Learning

ØKOKRIM (2012) *Trusselvurdering – økonomisk kriminalitet og miljøkriminalitet 2013-2014* [Internett], Økokrim. Tilgjengelig fra:

<[http://www.okokrim.no/www/okokrim/resource.nsf/files/www93sfde-trusselvurdering_2013-20141/\\$FILE/trusselvurdering_2013-20141.pdf](http://www.okokrim.no/www/okokrim/resource.nsf/files/www93sfde-trusselvurdering_2013-20141/$FILE/trusselvurdering_2013-20141.pdf)> [Lest: 09.02.15]

9. Appendiks

A.1 Annen forskning på svart økonomi

Opp gjennom årene har flere ulike metoder blitt tatt i bruk for å finne estimater på svart økonomi. Vi vil nå kort presentere noen av studiene som har blitt utført i Norge og andre land.

A.1.1 Norge

For å få oversikt over tidligere studier i Norge, har vi blant annet tatt utgangspunkt i mastergradutredningen ”Omfanget av skyggeøkonomien: En oversikt over litteratur og metoder”, av Rogde og Rasmussen (2013).

A.1.1.1 Spørreundersøkelser

I 1980 og 1983 utførte Isachsen og Strøm (1985) spørreundersøkelser i Norge. Resultatene fra denne studien tydet på at Norges svarte økonomi utgjorde 4-6 % av BNP for disse årene. Spørreundersøkelsene indikerte at rundt halvparten av dette estimatet kom fra skjult inntekt, og at store deler av transaksjonene i den svarte økonomien ble utført ved bruk av kontanter.

A.1.1.2 Valutaetterspørsel

I sin studie fra 2002 tok Schneider for seg aktivitetsendringer i svart økonomi som forklaring på bevegelser i kontantbeholdninger. Han fant ut at Norge var et av OECD-landene med størst svart økonomi. Det ble videre hevdet at høy sosial sikkerhet og høyere skatter var med på å bidra til skatteunndragelse. Resultatene tydet på at Norges svarte økonomi hadde vokst fra om lag 14,8 % av BNP i 1989/1990 til 19 % i 2001/2002. Hovedårsaken til denne veksten og størrelsen var økende skattetrykk og trygdeavgifter, samt en økning i statlige reguleringsaktiviteter.

Schneider (2007) fant tegn på en nedgang i den norske skyggeøkonomien med 18,4 % i 2002/2003 og 16,8 % i 2004/2005. Likevel var Norges estimat stadig over OECD-snittet.

A.1.1.3 Elektroniske betalinger

Flere studier bygger på antakelsen om at transaksjoner innenfor svart økonomi gjennomføres ved bruk av kontanter. Schneider (2013), i samarbeid med Visa Europe og A. T. Kearney, publiserte en studie som så på sammenhengen mellom svart økonomi og bruken av elektroniske betalinger fremfor kontanter. Schneider tok i bruk den modellbaserte metoden, MIMIC (multiple indicators, multiple causes). MIMIC er en statistisk teknikk som oppretter en strukturell modell for den svarte økonomien og undersøker forholdene mellom denne økonomien og andre inndata, som for eksempel sosial sikkerhet og direkte skatt.

Denne studien anslo at den svarte økonomien var om lag 14 % av BNP i 2013, og viste en nedgang fra Schneiders tidligere studier. Videre ble det også konkludert at det var en negativ sammenheng mellom svart økonomi i et land og dets elektroniske betalinger.

A.1.2 Allingham-Sandmo modellen

Her vil vi presentere en modell utviklet av Michael G. Allingham og Agnar Sandmo (1972), som er en teoretisk analyse om sammenhengen mellom skattlegging og risikotaking.

Allingham og Sandmo (1972) analyserte skattebetalerens beslutninger vedrørende skatteunndragelse ved underrapportering av inntekt. Modellen foreslo at endring i skatterater kunne få to effekter som trakk i motsatt retning. Den første effekten gikk ut på at skatteunndragelsen ville synke når skattesatsen ble satt ned, som en substitusjonseffekt. Dette ble begrunnet med at mindre profitt på unndragelse skapte et insentiv for skattebetaleren til å redusere unndragelsen, ved å substituere rapportert inntekt med unndratt inntekt. Den andre effekten gikk ut på at lavere skattesatser kunne ha en motsatt effekt, en inntektseffekt, som førte til økt skatteunndragelse dersom skattebetalerens risikoaversjon avtok når inntekten økte.

Allingham og Sandmo konkluderte med at de viktigste verktøyene myndighetene hadde for å motvirke skatteunndragelser var skattesatsene i seg selv, straffen for å bli tatt og hvor mye ressurser de avsatte til etterforskning på området.

A.1.3 Danmark

Kleven et al. (2011) undersøkte graden og utbredelsen av skatteunndragelse i Danmark. Den teoretiske litteraturen bygget på Allingham-Sandmo modellen, som presentert ovenfor.

Deres forskning ledet til tre empiriske funn. Først fant de et anslag på skatteunndragelse lik null for inntekt der en tredjepart var ansvarlig for rapporteringen. Der inntekten ble selvrapportert var anslaget av betydelig størrelse. Videre fant de, ved bruk av kvasiexperimentell variasjon, at marginale skattesatser hadde en positiv effekt på skatteunndragelse for selvrapportert inntekt. Det ble også konkludert at tidligere revideringer, samt ”trusler” om revidering, hadde en signifikant effekt på selvrapportert inntekt, men ingen effekt på inntekt rapportert av en tredjepart.

A.1.4 Canada

Mirus og Smith (1989) kom frem til at den svarte økonomien i Canada var om lag 12-15 % av BNP. Her hadde de inkludert ulovlige aktiviteter, men ekskludert ”byttehandler”. Dette estimatet ble beregnet med ulike metoder, som intervjuer, ulike inntekts- og utgiftsbaserte metoder (ikke lik den vi benytter i denne utredningen), samt indirekte metoder. De påpekte svakheter ved alle metoder i sin studie, og benyttet av den grunn alle de nevnte metodene for å komme frem til sitt estimat.

Studien fokuserte mer på hvordan man kan forebygge en svart økonomi, fremfor hvordan å beregne et estimat. For å redusere omfanget foreslo Mirus og Smith (1989) blant annet at man kunne redusere skattesatsen og samtidig utvide skattegrunnlaget, satse på større oppfølging og kontroll, i tillegg til å øke internasjonalt samarbeid for deling av skatteinformasjon.

A.1.5 Ungarn

Benedek og Lelkes (2011) satte fokus på de store variasjonene i den faktiske skattebyrden til individer med likt inntektsnivå, som oppsto som følge av skatteunndragelse. Det ble her benyttet en noe annerledes metode, der de sammenlignet brutto rapporterte inntekter i selvangivelsen til nesten 230 000 tilfeldig valgte personer, hvor deres inntekt var oppgitt i en husholdningsundersøkelse.

Resultatene fra denne studien viste at den gjennomsnittlige underrapporteringen (for selvstendig næringsdrivende og arbeidstakere) lå på 9-13 %, men at snittet ble vesentlig dratt opp av selvstendig næringsdrivende. Videre fant de at denne formen for skatteunndragelse reduserte kostnadene for personlig skattebetaling med 16-20 %.

Benedek og Lelkes (2011) fant også at husholdninger med høy inntekt unndro forholdsmessig mer skatt, ettersom andelen ”fattige” steg minimalt, mens inntektsulikheten økte signifikant. Progressiviteten i skattesystemet ble også redusert, og mistet mye av sin effekt på grunn av skatteunndragelse.

A.2 Revidert versjon av Pissarides og Weber (1989)

Wangen (2005) kom med en revidert utgave av Pissarides og Weber (1989) sin metode, hvor han tok for seg det han mener var feil og mangler ved den opprinnelige metoden. I denne utredningen har vi ikke tatt hensyn til dette.

Wangen (2005) foreslo tre nye estimeringsintervaller. Det første forslaget kom på bakgrunn av at Pissarides og Weber (1989) sitt resonnement for utledningen var for uformell. Wangen (2005) mente at man måtte behandle estimatet på k som en målfunksjon, avhengig av de tre argumentene; σ_{vSE}^2 , σ_{uSE}^2 og σ_{YSE}^2 . I følge Wangens beregninger var det korrekte intervallet for alle verdier av $\rho \in [0,1]$:

$$\left(\frac{\gamma}{\beta} + \frac{\sigma_{YEE}^2}{2} \right) \pm \frac{\sigma_{YSE}^2}{2\sqrt{1-\rho^2}}$$

Ved å sammenligne dette intervallet med estimatet på Pissarides og Webers (1989) intervall, fant han den største forskjellen når $\rho = 1$.

Videre hevdet Wangen (2005) at Pissarides og Weber (1989) ikke hadde fått frem hvor viktig likheten mellom de sammensatte variansene til utgifts- og inntektsfunksjonen var. Disse ledet til to ulike estimatorene. Det første var det enkleste av alle tre intervallene, og kunne bli utledet ved å kun ta i bruk matutgiftsfunksjonen. Det andre intervallet ble utledet som:

$$\left(\frac{\gamma}{\beta} + \frac{\sigma_{\eta EE}^2}{2\beta^2} \right) \pm \frac{\sigma_{\eta SE}^2}{2\beta\sqrt{1-\rho^2}}$$

Det tredje intervallet fant Wangen (2005) ved å gjøre en liten endring på Pissarides og Weber (1989) sin modell. Det ble hevdet at i stedet for å avvise inntektsfunksjonen, kunne man inkludere de strukturelle likhetene mellom residualene til utgifts- og inntektsfunksjonen i estimeringen. De sammensatte residualene ville dermed oppfylle følgende restriksjon:

$$\frac{\sigma_{\eta SE}^2 - \sigma_{\eta EE}^2}{\beta^2} = \sigma_{YSE}^2 - \sigma_{YEE}^2$$

Ut fra dette hevdet Wangen (2005) at utgifts- og inntektsfunksjonen for selvstendig næringsdrivende og arbeidstakere, både innenfor ”white-collar” og ”blue-collar”, kunne blitt estimert samtidig. Denne estimeringen ville avhenge av forholdet mellom estimatene til σ_{YEE}^2 og $\frac{\sigma_{\eta EE}^2}{\beta^2}$. Dersom $\sigma_{YEE}^2 > \frac{\sigma_{\eta EE}^2}{\beta^2}$ fulgte man det første intervallet, og dersom $\sigma_{YEE}^2 < \frac{\sigma_{\eta EE}^2}{\beta^2}$ fulgte man det andre intervallet. Det er her viktig å notere seg at selv om formlene er de samme, vil estimatene av de grunnleggende parameterne vanligvis være forskjellige mellom den simultane estimatmodellen til Wangen (2005) og ligning-for-ligning metoden utviklet av Pissarides og Weber (1989).

Resultatene Wangen (2005) kom frem til var avhengig av hvilken estimeringsmetode av de foreslåtte han benyttet. Ved bruk av den første metoden, kom han frem til at den øvre grensen for underrapportering var 0,1 % av BNP høyere for begge grupper av arbeidere enn hva Pissarides og Weber (1989) fant.

Den andre metoden han foreslo ledet til store sprik mellom den øvre og den nedre grensen. Her tydet et av resultatene på at selvstendig næringsdrivende *overrapporterte* sin inntekt, og et annet at deres ”sanne” inntekt var 3,5 ganger større enn den rapporterte, altså 12,4 % av BNP. Disse resultatene skilte seg vesentlig fra resultatene til Pissarides og Weber (1989).

Et anslag basert på det tredje intervallet ble ikke presentert, da dette krevde en full reestimering. Wangen (2005) presiserte imidlertid at valg av metode har stor effekt på resultatet.

A.3 Estimering av k når $\rho > 0$

Fra Pissarides og Weber (1989) vet vi at variansen til u_i , σ_u^2 , ikke kan overstige variansen til inntektsresidualet til arbeidstakere, σ_{YEE}^2 , ettersom denne er en sum av variansen til u_i og variansen til residualet til den permanente inntekten. Dersom vi bruker estimatet av σ_{YEE}^2 som en øvre grense for σ_{uEE}^2 , og tar med oss videre at $\sigma_{uSE}^2 = \sigma_{uEE}^2$, får vi fra ligning (17) at;

$$\sigma_{YSE}^2 - \sigma_{YEE}^2 = \sigma_{uEE}^2 + \sigma_{vSE}^2 - 2\rho\sigma_{uEE}\sigma_{vSE} - \sigma_{uEE}^2 = \sigma_{vSE}^2 - 2\rho\sigma_{uEE}\sigma_{vSE},$$

og sitter da igjen med den øvre grensen:

$$\sigma_{YSE}^2 - \sigma_{YEE}^2 = \sigma_{vSE}^2 - 2\rho\sigma_{YEE}\sigma_{vSE}. \quad (23)$$

Videre kan vi løse for σ_{vSE} ved hjelp av en andregradsligning formulert ut i fra ligning (23):

$$\sigma_{vSE}^2 - (2\rho\sigma_{YEE})\sigma_{vSE} - (\sigma_{YSE}^2 - \sigma_{YEE}^2) = 0.$$

På bakgrunn av dette kan vi beregne $cov(u_{SE}, v_{SE}) = 2\rho\sigma_{YEE}\sigma_{vSE}$, og dermed finne $\ln \bar{k}$ ved bruk av ligning (22) for de tilfellene hvor vi har positiv korrelasjonskoeffisient, $\rho > 0$.

A.4 Variabeloversikt

Tabell A.1: *Oversikt over variabler brukt i regresjonsanalysene*

Forklaringsvariabel	Variabelbeskrivelse
lnC	matkonsum, ln-transformert
lnY	inntekt, ln-transformert
SE	dummy lik 1 dersom minst én er selvstendig næringsdrivende, 0 hvis ikke
alder	alderen til hovedinntektstaker
alder_sq	alderen til hovedinntektstaker opphøyd i andre potens
ant_barn	antall barn i husholdningen
ant_barn_sq	antall barn i husholdningen opphøyd i andre potens
antbu7	antall barn i husholdningen som er under 7 år
t	lineær tidstrend
t2	kvadratisk tidstrend
SE1	dummy lik 1 dersom kun hovedinntektstaker er selvstendig næringsdrivende, 0 hvis ikke
SE2	dummy lik 1 dersom kun person 2 er selvstendig næringsdrivende, 0 hvis ikke
SE3	dummy lik 1 dersom begge er selvstendig næringsdrivende, 0 hvis ikke
SE_tid	Interaksjonsvariabel mellom SE og t (SE*t)
BA_SE	Interaksjonsvariabel mellom dummy for bygg- og anleggsbransjen og selvstendig næringsdrivende
BA_EE	Interaksjonsvariabel mellom dummy for bygg- og anleggsbransjen og arbeidstakere
HU_SE	Interaksjonsvariabel mellom dummy for helse- og undervisningssektoren og selvstendig næringsdrivende
HU_EE	Interaksjonsvariabel mellom dummy for helse- og undervisningssektoren og arbeidstakere
Instrumentvariabel	Variabelbeskrivelse
boligflate	boligens samlede nettoflate
arbtimer	antall arbeidstimer per uke for hovedinntektstaker
arbtimer_p2	antall arbeidstimer per uke for person 2

Husholdningskarakteristika er uthevet.

A.5 Statistiske tester

Teorien som er benyttet i dette delkapitlet er hentet fra Woolridge (2012).

A.5.1 Heteroskedastisitet

I tverrsnittsdata er ofte heteroskedastisitet et problem, altså at variablene ikke har konstant varians. Vi utførte en Breusch-Pagan test i STATA, og fikk en høy chi-kvadrert på 999 og en lav p-verdi på 0,000. Ut fra dette forkastet vi nullhypotesen om at det ikke eksisterer heteroskedastisitet, og har dermed kjørt alle regresjoner med robuste standardavvik.

A.5.2 Endogenitetsproblem

Vi utførte Wu-Hausmans test for endogenitet i inntektsvariabelen. Her tar man utgangspunkt i at IV-estimatene alltid er forventningsrette, og at OLS-estimatene er det dersom $cov(x, u) = 0$. Vi forkastet nullhypotesen om at inntektsresidualet var forskjellig fra null. Dette tilsier at det er korrelasjon mellom inntekten og feilledet i regresjonen. Inntektsvariabelen er følgelig endogen, og instrumentvariabelmetoden gir dermed bedre estimater enn OLS.

A.5.3 F-test

For å undersøke hvorvidt de valgte instrumentvariablene var signifikante sammen, utførte vi en F-test. Resultatene er presentert i tabell 5. Vi gjennomførte denne testen fordi vi valgte flere instrumentvariabler for én forklaringsvariabel. Tommelfingerregelen for at instrumentvariablene sammen er signifikant forskjellig fra null, er at F-verdien er større enn 10. Dette kravet oppfyller vi med en F-verdi på over 100 og en p-verdi på 0,000. Dermed kan vi konkludere med at våre instrumenter er signifikante sammen.

A.5.4 Sargan-test

For å avgjøre validiteten av overidentifiserende restriksjoner, utførte vi en Sargan-test. Fra tabell 5 observerer vi en høy p-verdi, som forteller oss at vi ikke kan forkaste nullhypotesen om at alle overidentifiserte restriksjoner er valide. Dette bekrefter at det ikke er problemer med våre instrumenter.

A.5.5 Tidstrend

Vi har inkludert både en lineær og en kvadratisk tidstrend for å justere for uobserverbare trendfaktorer som kun er tidsavhengige, som inflasjon og prisnivå. Grunnlaget for dette er at vi forventer at inntekt og matkonsum øker relativt jevnt fra 1975 til 2012. Ved å inkludere tidstrenden, sørger vi for at regresjonsligningen registrerer at den avhengige variabelen, $\ln C$, kan vokse/minke over tid på grunnlag av årsaker som ikke kan forklares av forklaringsvariablene.

Før vi tok avgjørelsen om å inkludere tidstrend, undersøkte vi andre tidsvarianter og resultatene fra disse. Først undersøkte vi hvordan regresjonsresultatene ble med et fast konstantledd, dette fremheves i tabell A.2.

Tabell A.2: Regresjonsresultater for $\ln(\text{matutgifter})$, med fast konstantledd, 1975-2012

	Alle næringer
$\ln(\text{inntekt}) (\beta)$	0,549 (0,006)***
Selvstendig næringsdrivende (γ)	0,053 (0,01)***
Alder	0,0115 (0,003)***
Alder kvadrert	-0,00008 (0,00003)**
Antall barn	0,2997 (0,008)***
Antall barn kvadrert	-0,029 (0,002)***
Antall barn under 7 år	-0,022 (0,005)***
Konstantleddet	2,88 (0,005)***
Antall observasjoner	29097
Forklaringskraft (R^2)	0,5294

Robuste standardavvik i parentesene. * $p < 0,1$; ** $p < 0,05$; *** $p < 0,01$.

Til tross for signifikante koeffisienter i tabell A.2, antar vi at det må justeres for tidsavhengige faktorer. Av denne grunn undersøkte vi årsummier. Resultatene fra dette fremkommer i tabell A.3.

Tabell A.3: Regresjonsresultater for $\ln(\text{matutgifter})$, med årsummier fra 1976-2012
(1975 er utelatt)

	Alle næringer
$\ln(\text{inntekt}) (\beta)$	0,29 (0,01)***
Selvstendig næringsdrivende (γ)	0,028 (0,01)***
Alder	0,029 (0,002)***
Alder kvadrert	-0,0003 (0,000)***
Antall barn	0,322 (0,009)***
Antall barn kvadrert	-0,032 (0,002)***
Antall barn under 7 år	-0,022 (0,005)***
Årsummier:	
1976	0,051 (0,03)*
1977	0,0998 (0,03)***
1978	0,13 (0,028)***
1980	0,17 (0,03)***
1981	0,37 (0,028)***
1982	0,48 (0,028)***
1983	0,55 (0,029)***
1984	0,57 (0,028)***
1985	0,6 (0,028)***
1986	0,65 (0,027)***
1987	0,68 (0,028)***
1988	0,67 (0,03)***
1989	0,74 (0,03)***
1990	0,77 (0,03)***
1991	0,84 (0,028)***
1992	0,80 (0,028)***
1993	0,79 (0,028)***
1994	0,80 (0,029)***
1995	0,75 (0,029)***
1996	0,77 (0,029)***
1997	0,78 (0,03)***
1998	0,75 (0,03)***
1999	0,78 (0,03)***
2000	0,796 (0,03)***
2001	0,80 (0,03)***
2002	0,81 (0,03)***
2003	0,82 (0,03)***
2004	0,88 (0,03)***
2005	0,86 (0,03)***
2006	0,87 (0,03)***
2007	0,89 (0,036)***
2008	0,92 (0,036)***
2009	0,99 (0,034)***
2012	1,06 (0,03)***
Konstantleddet	4,99 (0,109)***
Antall observasjoner	29097
Forklaringskraft (R^2)	0,5767

Robuste standardavvik i parentesene. * $p < 0,1$; ** $p < 0,05$; *** $p < 0,01$.

Resultatene fra tabell A.3 viser at koeffisientene til γ og β ikke er veldig forskjellig fra regresjonsresultatene vi får når vi inkluderer t og t^2 . Vi observerer også at konstantleddet øker nokså jevnt gjennom 70- og 80-tallet, flater ut på 90-tallet, og så minker litt, før det igjen øker relativt jevnt fra 1998 og frem mot 2012. Ut fra mønsteret på dummykoeffisientene er det ikke helt åpenbart hva som er det riktige valget mellom tidstrend og årsummy. Til tross for at det ikke ser helt stabilt ut fra år til år, velger vi å inkludere lineær og kvadratisk tidstrend, fordi vi i utgangspunktet forventer gradvis endring opp gjennom årene. Når vi inkluderer trenden, definerer vi tidsvariabelen som ($t = \text{år} - 1975$). Vi har imidlertid ikke deflatert konsum- og inntektsvariabelen, da det ikke ser ut til at den øvrige litteraturen har gjort dette.

A.6 Estimat på BNP-andel

Fra Statistisk Sentralbyrå (2013b) har vi at selvstendig næringsdrivendes bruttoinntekt i snitt var kr 572 500 i 2012, og at antall selvstendig næringsdrivende var 336 661 i Norge. Til tross for at vi antar høy inntektsvarians, tar vi utgangspunkt i at disse tallene multiplisert med hverandre vil gi oss total inntekt fra selvstendig næringsvirksomhet i Norge:

$$kr\ 572\ 500 \cdot 336\ 661 = kr\ 192\ 738\ 422\ 500$$

Vi finner at bruttonasjonalprodukt (BNP) var på om lag 2 965 milliarder NOK i Norge i 2012 (Statistisk Sentralbyrå, 2015), og kan dermed finne andelen av BNP som stammer fra selvstendig næringsvirksomhet slik:

$$\frac{192\ 738\ 422\ 500}{2\ 965\ 000\ 000\ 000} = 0,065,$$

altså 6,5 %. Dette benytter vi videre når vi skal finne andelen av underrapportering av BNP. Vi har et estimat på underrapportering av inntekt på 14 % for alle næringer, og kan dermed kalkulere denne andelen slik:

$$0,14 \cdot 0,065 = 0,0091$$

noe som tilsvarer 0,91 %. Vi kan dermed si at den delen av svart økonomi i Norge som skyldes underrapportering av inntekt er på 0,91 % av BNP, noe som tilsvarer nærmere 27 milliarder NOK ($0,0091 \cdot 2\ 965\ 000\ 000\ 000$). Det er imidlertid et noe usikkert estimat, da vi kun har benyttet BNP-tall fra 2012, og estimatet vårt på underrapporteringen gjelder for hele perioden fra 1975 til 2012.

Med 27 milliarder kroner unndratt og 336 661 selvstendig næringsdrivende, kan vi også finne ut hvor mye hver selvstendig næringsdrivende har underrapportert sin inntekt med i snitt;

$$\frac{27\ 000\ 000\ 000}{336\ 661} = 80\ 199.$$

Ved å anta at underrapporteringen i snitt skulle vært beskattet med 27 % (antakeligvis et noe forsiktig estimat), kan vi finne det korresponderende tapet for staten;

$$27\,000\,000\,000 \cdot 0,27 = 7\,290\,000\,000,$$

som altså tilsvarer 7,29 milliarder kroner.

A.7 Konfidensintervall for k

Regresjonsutskriften nedenfor viser 95 %-konfidensintervallet vi har benyttet for å kalkulere minste og største estimat på k (for alle næringer). Vi har brukt konfidensintervallene for β og γ , og formelen $k = e^{\frac{\gamma}{\beta}}$. På denne måten kan vi finne ut om k er signifikant forskjellig fra null. Utrekningene for konfidensintervallet til k finnes under utskriften.

Linear regression

Number of obs = 29097
 F(9, 29087) = 3511.97
 Prob > F = 0.0000
 R-squared = 0.5646
 Root MSE = .50898

lnC	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]	
lnY	.2926086	.0094987	30.81	0.000	.2739907	.3112264
SE	.0422876	.0091577	4.62	0.000	.024338	.0602372
alder	.0258919	.0027739	9.33	0.000	.0204549	.0313289
alder_sq	-.0002171	.0000317	-6.85	0.000	-.0002792	-.0001549
ant_barn	.3276704	.0089896	36.45	0.000	.3100504	.3452904
ant_barn_sq	-.0337266	.0023876	-14.13	0.000	-.0384063	-.0290468
antbu7	-.0248934	.0047165	-5.28	0.000	-.034138	-.0156487
t	.0502585	.0014063	35.74	0.000	.047502	.053015
t2	-.0007426	.0000296	-25.07	0.000	-.0008006	-.0006845
_cons	5.118824	.1078995	47.44	0.000	4.907336	5.330312

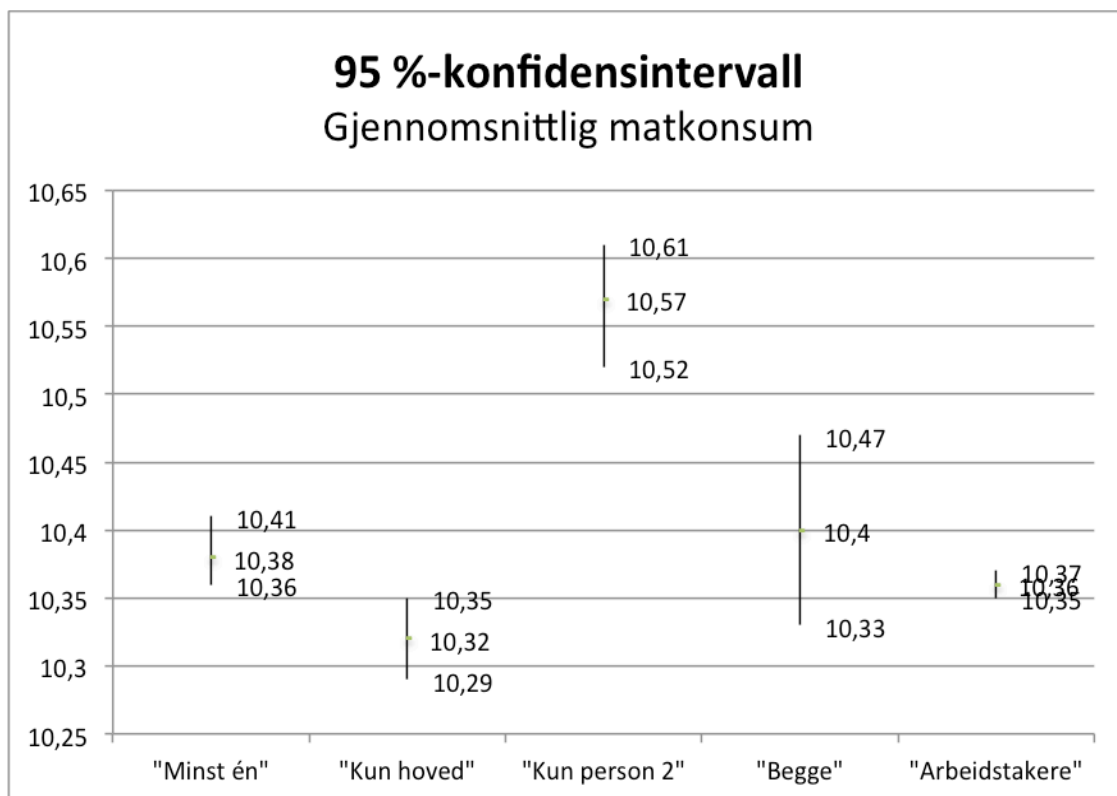
Laveste estimat: $k = e^{\frac{0,024338}{0,3112264}} \approx 1,08$.

Høyeste estimat: $k = e^{\frac{0,0602372}{0,2739907}} \approx 1,25$.

Konfidensintervallet for k blir dermed [1,08, 1,25].

A.8 Konfidensintervall for gjennomsnittlig matkonsum

Figuren nedenfor illustrerer 95 %-konfidensintervall for gjennomsnittlig matkonsum, fordelt på de ulike kategoriene av selvstendig næringsdrivende og arbeidstakere.



A.9 Dofil for behandling av data

Vi fikk tilsendt ett datasett med rådata av Forbruksundersøkelsen for hvert år fra 1975 til 2012. Variabelnavnene var forskjellig i de fleste datasettene, og vi har derfor laget én dofil per datasett for å rense bort de variablene vi ikke trenger, og for å gjøre alle datasettene klare til å slås sammen til ett stort datasett. Her vil vi kun presentere dofilen vi benyttet for datasettet fra 2009 (da dette var det første vi fikk tilsendt). Dofilene for de andre årgangene er tilsvarende, hvor eneste forskjellen er variabelnavnene.

I denne dofilen renser vi datasettene ved å slette variabler vi ikke trenger, endre variabelnavn og luke ut ekstremverdier og husholdninger vi ikke mener er relevante for vår analyse. Vi finner også beskrivende statistikk i denne dofilen, som er presentert i tabell 3.

DOFIL:

*Valg av arbeidsområde:

```
use "M:\Masteroppgave\Datasettene\2000-tallet\Forbruk2009.dta"
```

*Dropper overflødig informasjon og endrer navn på ulike variabler:

```
drop verkn112-verkn825
drop verkn832-verkn851
rename verkn831 restaurant
drop verkn006-verkn010
drop verkn016-verkn020
drop verkn027-verkn030
rename hefte1a fisk_janei
drop verva* mgdva* mgdkn* refnr3 ald3 kjonn3 ektest3 siv3 arbeid3 arbb3 tuke3 knaer3 /*
*/ yrke3 ledende3 ansf3 arb065_3 st_type3 ikkej3 spm61a3 spm61b3 refnr4 /*
*/ ald4 kjonn4 ektest4 siv4 arbeid4 arbb4 tuke4 knaer4 yrke4 /*
*/ ledende4 ansf4 arb065_4 st_type4 ikkej4 spm61a4 spm61b4 refnr5 ald5 /*
*/ kjonn5 ektest5 siv5 arbeid5 arbb5 tuke5 knaer5 yrke5 ledende5 /*
*/ ansf5 arb065_5 st_type5 ikkej5 spm61a5 spm61b5 refnr6 ald6 kjonn6 /*
*/ ektest6 siv6 arbeid6 arbb6 tuke6 knaer6 yrke6 ledende6 ansf6 /*
*/ arb065_6 st_type6 ikkej6 spm61a6 spm61b6 refnr7 ald7 kjonn7 ektest7 /*
*/ siv7 arbeid7 arbb7 tuke7 knaer7 yrke7 ledende7 ansf7 arb065_7 /*
*/ st_type7 ikkej7 spm61a7 spm61b7 refnr8 ald8 kjonn8 ektest8 siv8 /*
*/ arbeid8 arbb8 tuke8 knaer8 yrke8 ledende8 ansf8 arb065_8 st_type8 /*
*/ ikkej8 spm61a8 spm61b8 refnr9 ald9 kjonn9 ektest9 siv9 arbeid9 /*
*/ arbb9 tuke9 knaer9 yrke9 ledende9 ansf9 arb065_9 st_type9 ikkej9 /*
*/ spm61a9 spm61b9 hpalder spm1 spm1a* spm1b* spm1c* spm2 spm2a spm2b spm3a spm3b /*
*/ spm6b spm6c spm7 spm8 spm9b ksum blbelop rentavd0 komavg00 samlavg* spm4 spm6a spm20* /*
*/ spm21 insbelop spm22b spm23* spm24* rentavd2 spm25* spm26* spm27* spm28* spm29* /*
*/ spm30b spm31a filbelop spm32* rentavd1 komavg01 spm33 spm34* spm35 spm36* spm37* /*
*/ spm380 spm381 spm382 spm39002 spm383 spm39* spm386 spm387 spm41* spm42* sp43eg* /*
*/ kodeeg* spm43* spm45kg* kodekg* spm45* spm46* kode46* spmmobel spm47* spm48* spm49* /*
*/ spm57 kode* hm_barn* utgift* offpriv* spm59 campt motors baat komfyr mbovn efrihus innfaar /*
*/ kjfrys oppvask kjskap fryser toerktr videos videok esymask vekt rt nuts2 paryrk fvekt pulje spm61a* /*
*/ spm61b* h_hefte arb065*
rename hefte2a mat_mottatt
drop hefte1b hefte2b ektest* refnr* siv* st_type* antfenh avdbolig gjbolig landsd spm40* arbeid1 /*
*/ arbb1 yrke1 ledende1 ikkej1 ald2 arbeid2 yrke2 ledende2 ikkej2 bstroek* hndfelt/*
rename efbolig eier_leier
drop kommtype htype* hpkjoeenn ybarnald naering hustype aastall spm10 spm11 spm14 spm15a /*
*/ spm16a spm16b spm16d1 spm16d2 spm16e spm16f spm16ib spm17 spm18a spm18b spm19 /*
*/ spm384 spm51b spm53b spm52 spm54a spm70 spm71 arbb2
rename spm5b ant_ror
drop spm9
```

```

rename spm3 hustype
rename spm5a boligflate
rename spm6d sentralvarme
rename spm9a leieform
rename spm16c festeavgift
rename spm16f2 eiendomsskatt
rename spm30a hytteeier
rename spm51a reise_kjøpt
rename spm53a reise_mottatt
drop spm16g spm16h spm16i spm16ia spm50 spm54avr spm54b spm54bvr

```

*Oversikt over alle missing values i datasettet:

```
misstable summarize
```

*Vi antar at missing values for følgende kan erstattes med null (fordi det er beløp):

*Matvarer:

```

foreach var of varlist verkn*{
  replace `var' = 0 if (`var' == .)
}

```

*Dropper husholdningene hvor vi ikke har informasjon om inntekt:

```
drop if g_innt>=.
```

*Erstatter negative verdier for hver matvare med null:

```

foreach var of varlist verkn*{
  summarize `var', detail
  replace `var'=0 if (`var'<0)
}

```

GENERER VARIABLER

/*Vi dropper husholdningene hvor vi ikke har informasjon om ansettelsesforholdet til hovedinntektstaker (for da kan vi ikke vite om han/hun er ansatt eller selvstendig næringsdrivende)*/

```
drop if ansf1>=.
```

*Dummy lik 1 dersom kun hovedinntektstaker er SE:

```
gen SE1=0
```

```
replace SE1=1 if (ansf1==1 & ansf2!=1)
```

```
replace SE1=. if missing(ansf1)
```

```
label variable SE1 "Dummy lik 1 dersom kun hovedinntektstaker er SE, 0 hvis ikke"
```

*Dummy lik 1 dersom person 2 er SE:

```
gen SE2=0
```

```
replace SE2=1 if (ansf2==1 & ansf1!=1)
```

```
replace SE2=. if missing(ansf2)
```

```
label variable SE2 "Dummy lik 1 dersom kun person 2 er SE, 0 hvis ikke"
```

*Dummy lik 1 dersom begge er SE:

```
gen SE3=0
```

```
replace SE3=1 if (ansf1==1 & ansf2==1)
```

```
label variable SE3 "Dummy lik 1 dersom begge er SE, 0 hvis ikke"
```

*Dummy lik 1 dersom minst en av hovedpersonene er SE:

```
gen SE=0
```

```
replace SE=1 if (SE1==1|SE2==1|SE3==1)
```

```
label variable SE "Dummy lik 1 dersom minst en i husholdningen er SE, 0 hvis ikke"
```

*Lager aggregert matvarevariabel:

```
egen float C=rowtotal(verkn001-verkn111)
```

```
label variable C "Aggregert matvareutgift"
```

*Dropper hver enkelt av matvarevariablene:

```
drop verkn001-verkn111
```

*In-transformerer matvariablene:

```
gen lnC=ln(C)
```

```
label variable lnC "Matvareutgift ln-transformert"
```

*Husholdningen må bestå av to voksne i ekteskap/registrert partnerskap/samboerskap:

*Dummy som er lik 1 hvis parforhold, 0 hvis ikke:

```
gen d_par=0
```

```
replace d_par=1 if (slekt2==2 | slekt2==3)
```

```
replace d_par=. if missing(slekt2)
```

```
label variable d_par "Dummy lik 1 hvis parforhold, 0 hvis ikke"
```

*Dropper hvis ikke parforhold:

```
drop if d_par==0
```

```
drop d_par
```

*Husholdningene kan ikke drive med jordbruk:

*Dummy lik 1 hvis en av hovedpersonene driver med jordbruk, 0 hvis ikke:

```

gen jordbruk=0
replace jordbruk=1 if (knaer1==1 | knaer2==1)
label variable jordbruk "Dummy lik 1 hvis jordbruk, 0 hvis ikke"
*Dropper hvis en av personene driver med jordbruk:
drop if jordbruk==1
drop jordbruk
*Dummy lik 1 dersom hovedinntektstaker er over 65 år, 0 hvis ikke:
gen d_65=0
replace d_65=1 if ald1>65
replace d_65=. if missing(ald1)
*Hovedinntektstaker må være under 65 år:
drop if d_65==1
drop d_65
*Genererer næringsgrupper:
*Dummy for bygg og anlegg:
gen bygg_anlegg=0
replace bygg_anlegg=1 if (knaer1==50 & aar>1974 & aar<1996)
replace bygg_anlegg=1 if (knaer1==45 & aar>1995 & aar<2013)
replace bygg_anlegg=. if missing(knaer1)
label variable bygg_anlegg "Dummy lik 1 hvis hovedinntektstaker jobber i bygg- og anleggsbransjen, 0 hvis ikke"
*Dummy for helse og undervisning:
gen helse_und=0
replace helse_und=1 if (knaer1==93 & aar>1974 & aar<1996)
replace helse_und=1 if ((knaer1==80 | knaer1==50) & aar>1995 & aar<2013)
replace helse_und=. if missing(knaer1)
label variable helse_und "Dummy lik 1 hvis hovedinntektstaker jobber i helsesektor eller med undervisning, 0 hvis ikke"
*Interaksjonsvariabel mellom bygg og anlegg og SE:
gen BA_SE=0
replace BA_SE=1 if (bygg_anlegg==1 & SE==1)
label variable BA_SE "Dummy lik 1 dersom BA og SE, 0 hvis ikke"
*Interaksjonsvariabel mellom bygg og anlegg og EE:
gen BA_EE=0
replace BA_EE=1 if (bygg_anlegg==1 & SE==0)
label variable BA_EE "Dummy lik 1 dersom BA og EE, 0 hvis ikke"
*Interaksjonsvariabel mellom helse og undervisning og SE:
gen HU_SE=0
replace HU_SE=1 if (helse_und==1 & SE==1)
label variable HU_SE "Dummy lik 1 dersom HU og SE, 0 hvis ikke"
*Interaksjonsvariabel mellom helse og undervisning og EE:
gen HU_EE=0
replace HU_EE=1 if (helse_und==1 & SE==0)
label variable HU_EE "Dummy lik 1 dersom HU og EE, 0 hvis ikke"
*Inntektsvariabelen:
rename g_innt Y
*ln-transformerer inntekt (Y), som er den rapporterte inntekten etter skatt:
gen lnY=ln(Y)
label variable lnY "Inntekt ln-transformert"
*Generer en dummy-variabel for antall barn i en husholdning
gen barn3=(slekt3==4 | slekt3==5)
gen barn4=(slekt4==4 | slekt4==5)
gen barn5=(slekt5==4 | slekt5==5)
gen barn6=(slekt6==4 | slekt6==5)
gen barn7=(slekt7==4 | slekt7==5)
gen barn8=(slekt8==4 | slekt8==5)
gen barn9=(slekt9==4 | slekt9==5)
gen ant_barn=( barn3+ barn4+ barn5+ barn6+ barn7+ barn8+ barn9)
label variable ant_barn "Antall barn i husholdningen"
drop barn*
*Opphøyer barnevariabelen i annen:
gen ant_barn_sq=(ant_barn)^2
label variable ant_barn_sq "Antall barn i husholdningen opphøyd i annen"
drop slekt*
*Antar at dersom vi har missing values for antall arbeidstimer for person 2, tilsvarer dette at personen har jobbet null timer:
foreach var of varlist tuke2{
replace `var`=0 if missing(`var`)

```

}

*Lagre:

save "M:\Masteroppgave\Steg 2 og 3\Datasett etter steg 2 og 3\2009_2.dta"

*Kombinerer alle datasettene

append using "M:\Masteroppgave\Steg 2 og 3\Datasett etter steg 2 og 3\1975_2.dta"/*

/ "M:\Masteroppgave\Steg 2 og 3\Datasett etter steg 2 og 3\1976_2.dta"/

/ "M:\Masteroppgave\Steg 2 og 3\Datasett etter steg 2 og 3\1977_2.dta"/

/ "M:\Masteroppgave\Steg 2 og 3\Datasett etter steg 2 og 3\1978_2.dta"/

/ "M:\Masteroppgave\Steg 2 og 3\Datasett etter steg 2 og 3\1979_2.dta"/

/ "M:\Masteroppgave\Steg 2 og 3\Datasett etter steg 2 og 3\1980_2.dta"/

/ "M:\Masteroppgave\Steg 2 og 3\Datasett etter steg 2 og 3\1981_2.dta"/

/ "M:\Masteroppgave\Steg 2 og 3\Datasett etter steg 2 og 3\1982_2.dta"/

/ "M:\Masteroppgave\Steg 2 og 3\Datasett etter steg 2 og 3\1983_2.dta"/

/ "M:\Masteroppgave\Steg 2 og 3\Datasett etter steg 2 og 3\1984_2.dta"/

/ "M:\Masteroppgave\Steg 2 og 3\Datasett etter steg 2 og 3\1985_2.dta"/

/ "M:\Masteroppgave\Steg 2 og 3\Datasett etter steg 2 og 3\1986_2.dta"/

/ "M:\Masteroppgave\Steg 2 og 3\Datasett etter steg 2 og 3\1987_2.dta"/

/ "M:\Masteroppgave\Steg 2 og 3\Datasett etter steg 2 og 3\1988_2.dta"/

/ "M:\Masteroppgave\Steg 2 og 3\Datasett etter steg 2 og 3\1989_2.dta"/

/ "M:\Masteroppgave\Steg 2 og 3\Datasett etter steg 2 og 3\1990_2.dta"/

/ "M:\Masteroppgave\Steg 2 og 3\Datasett etter steg 2 og 3\1991_2.dta"/

/ "M:\Masteroppgave\Steg 2 og 3\Datasett etter steg 2 og 3\1992_2.dta"/

/ "M:\Masteroppgave\Steg 2 og 3\Datasett etter steg 2 og 3\1993_2.dta"/

/ "M:\Masteroppgave\Steg 2 og 3\Datasett etter steg 2 og 3\1994_2.dta"/

/ "M:\Masteroppgave\Steg 2 og 3\Datasett etter steg 2 og 3\1995_2.dta"/

/ "M:\Masteroppgave\Steg 2 og 3\Datasett etter steg 2 og 3\1996_2.dta"/

/ "M:\Masteroppgave\Steg 2 og 3\Datasett etter steg 2 og 3\1997_2.dta"/

/ "M:\Masteroppgave\Steg 2 og 3\Datasett etter steg 2 og 3\1998_2.dta"/

/ "M:\Masteroppgave\Steg 2 og 3\Datasett etter steg 2 og 3\1999_2.dta"/

/ "M:\Masteroppgave\Steg 2 og 3\Datasett etter steg 2 og 3\2000_2.dta"/

/ "M:\Masteroppgave\Steg 2 og 3\Datasett etter steg 2 og 3\2001_2.dta"/

/ "M:\Masteroppgave\Steg 2 og 3\Datasett etter steg 2 og 3\2002_2.dta"/

/ "M:\Masteroppgave\Steg 2 og 3\Datasett etter steg 2 og 3\2003_2.dta"/

/ "M:\Masteroppgave\Steg 2 og 3\Datasett etter steg 2 og 3\2004_2.dta"/

/ "M:\Masteroppgave\Steg 2 og 3\Datasett etter steg 2 og 3\2005_2.dta"/

/ "M:\Masteroppgave\Steg 2 og 3\Datasett etter steg 2 og 3\2006_2.dta"/

/ "M:\Masteroppgave\Steg 2 og 3\Datasett etter steg 2 og 3\2007_2.dta"/

/ "M:\Masteroppgave\Steg 2 og 3\Datasett etter steg 2 og 3\2008_2.dta"/

/ "M:\Masteroppgave\Steg 2 og 3\Datasett etter steg 2 og 3\2012_2.dta"/

*/

Nå har vi ett stort datasett med alle årganger samlet. Det er dette ene datasettet vi bruker videre.

*Sorterer etter årstall:

sort aar

*Lager tidsvariabel:

gen t=aar-1975

gen t2=t^2

*Dropper flere variabler:

drop kjonn1 loepenr bostrok bostrok2 kjonn2 arbtimer foerst foersl

rename tuke1 arbtimer

*Har en missing value for aar, fjerner denne:

drop if missing(aar)

rename ald1 alder

gen alder_sq=(alder)^2

label variable alder_sq "Alder hovedinntektstaker kvadrert"

*Dropper fordi eller så får vi ikke noe estimat når vi ln-transformerer:

drop if Y==0

*Dropper verdier for mat som er lik null (eller av en eller annen grunn fortsatt er negative),

*eller så får vi ikke ln-transformert

foreach var of varlist C{

drop if `var`<0

```
drop if `var'==0
}
*Ekstremverdier for inntekt:
foreach var of varlist Y {
  summarize `var', detail
  drop if `var'<0
  drop if `var'>10000000
}

*BESKRIVENDE STATISTIKK*
*Tabell 3:
mean(alder) if SE==1
mean(alder) if SE1==1
mean(alder) if SE2==1
mean(alder) if SE3==1
mean(alder) if SE==0

mean(ant_barn) if SE==1
mean(ant_barn) if SE1==1
mean(ant_barn) if SE2==1
mean(ant_barn) if SE3==1
mean(ant_barn) if SE==0

mean(lnY) if SE==1
mean(lnY) if SE1==1
mean(lnY) if SE2==1
mean(lnY) if SE3==1
mean(lnY) if SE==0

mean(lnC) if SE==1
mean(lnC) if SE1==1
mean(lnC) if SE2==1
mean(lnC) if SE3==1
mean(lnC) if SE==0

*Tabell 4:
count if SE==1
count if BA_SE==1
count if HU_SE==1
count if SE==0
count if BA_EE==1
count if HU_EE==1

*Lagre:
save "M:\Masteroppgave\Steg 4\Datsett etter steg 4\Datsett.dta"
```


A.10 Dofil for regresjoner ved OLS

I denne dofilen utfører vi analyser ved bruk av OLS, og får ut tall til tabellene presentert i analysedelen.

DOFIL:

*Arbeidsområde:

use "M:\Masteroppgave\Steg 4\Datasett etter steg 4\Datasett.dta"

*Tester først om det er heteroskedastisitet:

quietly reg lnC lnY SE alder alder_sq ant_barn ant_barn_sq antbu7 t t2
estat hettest

HOVEDREGRESJON – tabell 5

*Alle næringer - kolonne 1:

reg lnC lnY SE alder alder_sq ant_barn ant_barn_sq antbu7 t t2, robust

*Finner k for alle næringer:

display exp(_b[SE]/_b[lnY])

*Bygg og anlegg – kolonne 3:

reg lnC lnY SE BA_SE BA_EE alder alder_sq ant_barn ant_barn_sq antbu7 t t2, robust

*Finner k for alle som ikke er i bygg- og anleggsbransjen:

display exp(_b[SE]/_b[lnY])

*Finner k for SE i bygg- og anleggsbransjen:

display exp((_b[SE]+_b[BA_SE])/_b[lnY])

*Finner k for EE i bygg- og anleggsbransjen:

display exp(_b[BA_EE]/_b[lnY])

*Helse og undervisning – kolonne 5:

reg lnC lnY SE BA_SE HU_SE BA_EE HU_EE alder alder_sq ant_barn ant_barn_sq antbu7 t t2, robust

*Finner k for SE i helse og undervisning:

display exp((_b[SE]+_b[HU_SE])/_b[lnY])

*Finner k for EE i helse og undervisning:

display exp(_b[HU_EE]/_b[lnY])

SKATTEREFORMER

*1975-1992:

reg lnC lnY SE BA_SE BA_EE alder alder_sq ant_barn ant_barn_sq antbu7 t t2 if (aar<1993), robust

display exp(_b[SE]/_b[lnY])

display exp((_b[SE]+_b[BA_SE])/_b[lnY])

display exp(_b[BA_EE]/_b[lnY])

count if (SE==1 & aar<1993)

*1993-2006:

reg lnC lnY SE BA_SE BA_EE alder alder_sq ant_barn ant_barn_sq antbu7 t t2 if (aar>1992 & aar<2007), robust

display exp(_b[SE]/_b[lnY])

display exp((_b[SE]+_b[BA_SE])/_b[lnY])

display exp(_b[BA_EE]/_b[lnY])

count if (SE==1 & (aar>1992 & aar<2007))

*2007-2012:

reg lnC lnY SE BA_SE BA_EE alder alder_sq ant_barn ant_barn_sq antbu7 t t2 if (aar>2006), robust

display exp(_b[SE]/_b[lnY])

display exp((_b[SE]+_b[BA_SE])/_b[lnY])

display exp(_b[BA_EE]/_b[lnY])

count if (SE==1 & aar>2006)

*Etter 1992:

reg lnC lnY SE BA_SE BA_EE alder alder_sq ant_barn ant_barn_sq antbu7 t t2 if (aar>1992), robust

display exp(_b[SE]/_b[lnY])

display exp((_b[SE]+_b[BA_SE])/_b[lnY])

display exp(_b[BA_EE]/_b[lnY])

count if (SE==1 & aar>1992)

ULIKE DEFINISJONER PÅ SE

reg lnC lnY SE1 SE2 SE3 alder alder_sq ant_barn ant_barn_sq antbu7 t t2, robust

*Finner k for husholdningene der kun hovedinntektstaker er SE

```

display exp(_b[SE1]/_b[lnY])
*Finner k for husholdningene der kun person 2 er SE
display exp(_b[SE2]/_b[lnY])
*Finner k for husholdningene der begge er SE
display exp(_b[SE3]/_b[lnY])

*UTVIKLING OVER TID*
gen SE_tid=SE*t
reg lnC lnY SE SE_tid alder alder_sq ant_barn ant_barn_sq antbu7 t t2, robust

*år for år
reg lnC lnY SE alder alder_sq ant_barn ant_barn_sq antbu7 if (aar==1975), robust
display exp(_b[SE]/_b[lnY])
count if (SE==1 & aar==1975)
reg lnC lnY SE alder alder_sq ant_barn ant_barn_sq antbu7 if (aar==1976), robust
display exp(_b[SE]/_b[lnY])
count if (SE==1 & aar==1976)
reg lnC lnY SE alder alder_sq ant_barn ant_barn_sq antbu7 if (aar==1977), robust
display exp(_b[SE]/_b[lnY])
count if (SE==1 & aar==1977)
reg lnC lnY SE alder alder_sq ant_barn ant_barn_sq antbu7 if (aar==1978), robust
display exp(_b[SE]/_b[lnY])
count if (SE==1 & aar==1978)
reg lnC lnY SE alder alder_sq ant_barn ant_barn_sq antbu7 if (aar==1979), robust
display exp(_b[SE]/_b[lnY])
count if (SE==1 & aar==1979)
reg lnC lnY SE alder alder_sq ant_barn ant_barn_sq antbu7 if (aar==1980), robust
display exp(_b[SE]/_b[lnY])
count if (SE==1 & aar==1980)
reg lnC lnY SE alder alder_sq ant_barn ant_barn_sq antbu7 if (aar==1981), robust
display exp(_b[SE]/_b[lnY])
count if (SE==1 & aar==1981)
reg lnC lnY SE alder alder_sq ant_barn ant_barn_sq antbu7 if (aar==1982), robust
display exp(_b[SE]/_b[lnY])
count if (SE==1 & aar==1982)
reg lnC lnY SE alder alder_sq ant_barn ant_barn_sq antbu7 if (aar==1983), robust
display exp(_b[SE]/_b[lnY])
count if (SE==1 & aar==1983)
reg lnC lnY SE alder alder_sq ant_barn ant_barn_sq antbu7 if (aar==1984), robust
display exp(_b[SE]/_b[lnY])
count if (SE==1 & aar==1984)
reg lnC lnY SE alder alder_sq ant_barn ant_barn_sq antbu7 if (aar==1985), robust
display exp(_b[SE]/_b[lnY])
count if (SE==1 & aar==1985)
reg lnC lnY SE alder alder_sq ant_barn ant_barn_sq antbu7 if (aar==1986), robust
display exp(_b[SE]/_b[lnY])
count if (SE==1 & aar==1986)
reg lnC lnY SE alder alder_sq ant_barn ant_barn_sq antbu7 if (aar==1987), robust
display exp(_b[SE]/_b[lnY])
count if (SE==1 & aar==1987)
reg lnC lnY SE alder alder_sq ant_barn ant_barn_sq antbu7 if (aar==1988), robust
display exp(_b[SE]/_b[lnY])
count if (SE==1 & aar==1988)
reg lnC lnY SE alder alder_sq ant_barn ant_barn_sq antbu7 if (aar==1989), robust
display exp(_b[SE]/_b[lnY])
count if (SE==1 & aar==1989)
reg lnC lnY SE alder alder_sq ant_barn ant_barn_sq antbu7 if (aar==1990), robust
display exp(_b[SE]/_b[lnY])
count if (SE==1 & aar==1990)
reg lnC lnY SE alder alder_sq ant_barn ant_barn_sq antbu7 if (aar==1991), robust
display exp(_b[SE]/_b[lnY])
count if (SE==1 & aar==1991)
reg lnC lnY SE alder alder_sq ant_barn ant_barn_sq antbu7 if (aar==1992), robust
display exp(_b[SE]/_b[lnY])
count if (SE==1 & aar==1992)
reg lnC lnY SE alder alder_sq ant_barn ant_barn_sq antbu7 if (aar==1993), robust
display exp(_b[SE]/_b[lnY])

```

```

count if (SE==1 & aar==1993)
reg lnC lnY SE alder alder_sq ant_barn ant_barn_sq antbu7 if (aar==1994), robust
display exp(_b[SE]/_b[lnY])
count if (SE==1 & aar==1994)
reg lnC lnY SE alder alder_sq ant_barn ant_barn_sq antbu7 if (aar==1995), robust
display exp(_b[SE]/_b[lnY])
count if (SE==1 & aar==1995)
reg lnC lnY SE alder alder_sq ant_barn ant_barn_sq antbu7 if (aar==1996), robust
display exp(_b[SE]/_b[lnY])
count if (SE==1 & aar==1996)
reg lnC lnY SE alder alder_sq ant_barn ant_barn_sq antbu7 if (aar==1997), robust
display exp(_b[SE]/_b[lnY])
count if (SE==1 & aar==1997)
reg lnC lnY SE alder alder_sq ant_barn ant_barn_sq antbu7 if (aar==1998), robust
display exp(_b[SE]/_b[lnY])
count if (SE==1 & aar==1998)
reg lnC lnY SE alder alder_sq ant_barn ant_barn_sq antbu7 if (aar==1999), robust
display exp(_b[SE]/_b[lnY])
count if (SE==1 & aar==1999)
reg lnC lnY SE alder alder_sq ant_barn ant_barn_sq antbu7 if (aar==2000), robust
display exp(_b[SE]/_b[lnY])
count if (SE==1 & aar==2000)
reg lnC lnY SE alder alder_sq ant_barn ant_barn_sq antbu7 if (aar==2001), robust
display exp(_b[SE]/_b[lnY])
count if (SE==1 & aar==2001)
reg lnC lnY SE alder alder_sq ant_barn ant_barn_sq antbu7 if (aar==2002), robust
display exp(_b[SE]/_b[lnY])
count if (SE==1 & aar==2002)
reg lnC lnY SE alder alder_sq ant_barn ant_barn_sq antbu7 if (aar==2003), robust
display exp(_b[SE]/_b[lnY])
count if (SE==1 & aar==2003)
reg lnC lnY SE alder alder_sq ant_barn ant_barn_sq antbu7 if (aar==2004), robust
display exp(_b[SE]/_b[lnY])
count if (SE==1 & aar==2004)
reg lnC lnY SE alder alder_sq ant_barn ant_barn_sq antbu7 if (aar==2005), robust
display exp(_b[SE]/_b[lnY])
count if (SE==1 & aar==2005)
reg lnC lnY SE alder alder_sq ant_barn ant_barn_sq antbu7 if (aar==2006), robust
display exp(_b[SE]/_b[lnY])
count if (SE==1 & aar==2006)
reg lnC lnY SE alder alder_sq ant_barn ant_barn_sq antbu7 if (aar==2007), robust
display exp(_b[SE]/_b[lnY])
count if (SE==1 & aar==2007)
reg lnC lnY SE alder alder_sq ant_barn ant_barn_sq antbu7 if (aar==2008), robust
display exp(_b[SE]/_b[lnY])
count if (SE==1 & aar==2008)
reg lnC lnY SE alder alder_sq ant_barn ant_barn_sq antbu7 if (aar==2009), robust
display exp(_b[SE]/_b[lnY])
count if (SE==1 & aar==2009)
reg lnC lnY SE alder alder_sq ant_barn ant_barn_sq antbu7 if (aar==2012), robust
display exp(_b[SE]/_b[lnY])
count if (SE==1 & aar==2012)

```

ÅRSDUMMYER – appendiks (tester bare hovedreg.):

```

reg lnC lnY SE alder alder_sq ant_barn ant_barn_sq antbu7 i.aar, robust
display exp(_b[SE]/_b[lnY])

```

FAST KONSTANTLEDD - appendiks (tester bare hovedreg.):

```

reg lnC lnY SE alder alder_sq ant_barn ant_barn_sq antbu7, robust
display exp(_b[SE]/_b[lnY])

```

A.11 Dofil for regresjoner ved IV

I denne dofilen utfører vi analyser ved bruk av IV, og får ut tall til tabellene presentert i analysedelen. Vi utfører også tester som er beskrevet tidligere.

DOFIL:

*Arbeidsområde:

use "M:\Masteroppgave\Steg 4\Datsett etter steg 4\Datsett.dta"

rename tuke2 arbtimer_p2

REGRESJONER TABELL 5

*Kolonne 2:

*Estimat på gamma og beta:

ivregress 2sls lnC (lnY = boligflate arbtimer arbtimer_p2) SE alder alder_sq ant_barn ant_barn_sq antbu7/*
*/ t t2, robust

*For å få varians til inntektsresidualet:

quietly reg lnC lnY SE alder alder_sq ant_barn ant_barn_sq antbu7 t t2, robust

reg lnY alder alder_sq ant_barn ant_barn_sq antbu7 t t2 boligflate arbtimer arbtimer_p2 if SE==1, robust

quietly reg lnC lnY SE alder alder_sq ant_barn ant_barn_sq antbu7 t t2, robust

reg lnY alder alder_sq ant_barn ant_barn_sq antbu7 t t2 boligflate arbtimer arbtimer_p2 if SE==0, robust

*Kolonne 4:

ivregress 2sls lnC (lnY = boligflate arbtimer arbtimer_p2) SE BA_SE BA_EE alder alder_sq ant_barn /*

*/ant_barn_sq antbu7 t t2, robust

*Kolonne 6:

ivregress 2sls lnC (lnY = boligflate arbtimer arbtimer_p2) SE BA_SE HU_SE BA_EE HU_EE alder /*

*/ alder_sq ant_barn ant_barn_sq antbu7 t t2, robust

WU-HAUSMAN-TEST

*Kolonne 2:

ivregress 2sls lnC (lnY = boligflate arbtimer arbtimer_p2) SE alder alder_sq ant_barn ant_barn_sq antbu7/*

*/ t t2, robust

estat endogenous

*Kolonne 4:

ivregress 2sls lnC (lnY = boligflate arbtimer arbtimer_p2) SE BA_SE BA_EE alder alder_sq ant_barn/*

*/ ant_barn_sq antbu7 t t2, robust

estat endogenous

*Kolonne 6:

ivregress 2sls lnC (lnY = boligflate arbtimer arbtimer_p2) SE BA_SE HU_SE BA_EE HU_EE alder/*

*/ alder_sq ant_barn ant_barn_sq antbu7 t t2, robust

estat endogenous

F-TEST

*Kolonne 2:

*Tester om instrumentene er signifikante sammen (jointly significant):

*Gjør dette når vi har flere instrumenter for en forklaringsvariabel

quietly reg lnC lnY SE alder alder_sq ant_barn ant_barn_sq antbu7 t t2, robust

quietly reg lnY SE alder alder_sq ant_barn ant_barn_sq antbu7 t t2 boligflate arbtimer arbtimer_p2, robust

test boligflate arbtimer arbtimer_p2

*instrumentene er jointly signifikante (tommelfingerregelen er at F skal være større enn 10).

*Kolonne 4:

quietly reg lnC lnY SE BA_SE BA_EE alder alder_sq ant_barn ant_barn_sq antbu7 t t2, robust

quietly reg lnY SE BA_SE BA_EE alder alder_sq ant_barn ant_barn_sq antbu7 t t2 boligflate/*

*/ arbtimer arbtimer_p2, robust

test boligflate arbtimer arbtimer_p2

*Kolonne 6:

quietly reg lnC lnY SE BA_SE HU_SE BA_EE HU_EE alder alder_sq ant_barn ant_barn_sq antbu7 t t2, robust

quietly reg lnY SE BA_SE HU_SE BA_EE HU_EE alder alder_sq ant_barn ant_barn_sq antbu7 t t2 /*

*/ boligflate arbtimer arbtimer_p2, robust

test boligflate arbtimer arbtimer_p2

SARGAN-TEST

*Kolonne 2:

```
ivregress 2sls lnC (lnY = boligflate arbtimer arbtimer_p2) SE alder alder_sq ant_barn ant_barn_sq antbu7/*  
*/ t t2, robust  
estat overid
```

*Kolonne 4:

```
ivregress 2sls lnC (lnY = boligflate arbtimer arbtimer_p2) SE BA_SE BA_EE alder alder_sq ant_barn/*  
*/ ant_barn_sq antbu7 t t2, robust  
estat overid
```

*Kolonne 6:

```
ivregress 2sls lnC (lnY = boligflate arbtimer arbtimer_p2) SE BA_SE HU_SE BA_EE HU_EE alder/*  
*/ alder_sq ant_barn ant_barn_sq antbu7 t t2, robust  
estat overid
```