Eksamen på Økonomistudiet vinter 2018-19

Økonometri I

Tag-hjem eksamen: 19. december, 2018, kl.10.00-22.00

Praktiske anvisninger til eksamen i Økonometri I:

- Eksamen kan besvares individuelt eller i grupper af max 3 studerende.
- Læs hele eksamensopgaven igennem før der udarbejdes svar.
- Besvar alle spørgsmål og delspørgsmål i Opgave 1 til 5.
- Besvarelsen skal bestå af en samlet rapport med relevante tabeller og figurer.
- Alle sider i rapporten skal forsynes med sidetal og eksamensnumre.
- Rapporten må højst bestå af 8 normalesider (eksklusiv forside og ansvarsfordeling). Til denne eksamen er en normalside defineret som en A4 side med fontstørrelse sat til 12, linjeafstand sat til 1.5 (halvanden), og sidemarginer sat til mindst 2.5cm.
- Rapporten skal forsynes med en forside og der skal udarbejdes en ansvarsfordeling således de enkelte gruppemedlemmers bidrag til besvarelsen fremgår. Dette gøres ved at anvende skabelonen 'forside.doc' som er tilgængelig på Digital Eksamen.
- Eksamen besvares ved at uploade rapporten i PDF-format til Digital Eksamen senest klokken 22 den 19. december. Bemærk at der kun skal uploades en rapport for hver gruppe. Dette gøres ved at følge instruktionerne beskrevet i følgende video: `http://video.ku.dk/digital-eksamen-gruppeaflevering-1`.
- Udover selve rapporten skal der også uploade et STATA program i TXT-format til Digital Eksamen. STATA-programmet skal kunne eksekveres uden fejl og generere resultater i samme rækkefølge som de fremgår af rapporten. Bemærk at STATA-programmet ikke indgår i bedømmelsen.
- Mette Ejrnæs kan kontaktes på telefon 3532 3062 klokken 10-12 den 19. december såfremt der er problemer med at downloade data eller eksamensopgaven er fejlbehæftet. Eventueller beskeder under eksamen annonceres på kursushjemmesiden og allersenest kl. 12.30. Herefter vil enhver form for kommunikation ophøre og eksamen besvares på baggrund af de tilgængelige informationer.
- Hvis Digital Eksamen er ramt af nedbrud, kan besvarelsen i nødstilfælde indsendes til samf-fak@samf.ku.dk. Dette kræver dog en udførlig dokumentation af problemet.
- Det er ikke tilladt at kommunikere med andre grupper under eksamen. Enhver form for kommunikation betragtes som eksamenssnyd og vil bliver behandlet herefter. Reglerne for eksamenssnyd er beskrevet på Din Uddannelsesside og i Rammestudieordningens afsnit 4.12.
- Dette eksamenssæt består af 8 sider i alt incl. denne forside.

Pas på at du ikke begår eksamenssnyd!

Pas på at du ikke begår eksamenssnyd!

Det er fx eksamenssnyd, hvis du ...

- Kopierer andres tekster uden at sætte citationstegn eller kildehenvise, så det ser ud som om det er din egen tekst
- Bruger andres idéer eller tanker uden at kildehenvise, så det ser ud om det er din egen idé eller dine tanker
- Genbruger dele af en opgave, som du tidligere har indleveret og fået en bestå karakter for uden at sætte citationsregn eller kildehenvise (selvplagiering)
- Modtager hjælp fra andre i strid med de regler, som er fastsat i rammestudieordningens afs. 4.12 om samarbejde/sparring

Du kan læse mere om reglerne for eksamenssnyd på Din Uddannelsesside og i Rammestudieordningens afs. 4.12.

Eksamenssnyd sanktioneres altid med en advarsel og bortvisning fra prøven. I de fleste tilfælde bliver den studerende også bortvist fra universitetet i et semester.

Introduktion til opgaven:

"Beregning af fattigdomsgrænser i Vietnam"

Når man beregner den absolutte fattigdomsgrænse, tager man udgangspunkt i de basale behov. Man fastlægger fattigdomsgrænsen således, at en person vil kunne få dækket de absolut basale behov som f.eks. mad, tøj, sundhedsudgifter etc. Der findes flere metoder til at fastlægge fattigdomsgrænser. I denne opgave tager vi udgangspunkt i en metode, der bygger på Engelkurven. Gennemgangen her i opgaven er baseret på rapporten "Well begun but not yet done: Progress and emerging challenges for poverty reduction in Vietnam" 2014 af Valerie Kozel og artiklen "How robust is a poverty profile" af Martin Ravallion og Benu Bidane, *World Bank Review* no 8 pp. 75-102.[1]

Metoden går ud på først at fastlægge det basale behov for mad. Forskellige studier har vist, at man regner med, at en person i Vietnam skal have 2230 kilokalorier per dag for at få dækket sit behov.[2] Dernæst bestemmes, hvor store udgifterne til mad vil være for at dække 2230 kilokalorier. For at gøre dette sammensættes så en diæt, som præcist vil dække kaloriebehovet. I diæten brugt i dette studie kommer 2/3 af kalorierne fra ris og den sidste 1/3 fra grønsager og kød (kød udgør dog mindre end 6 procent af kalorierne). Udfra forbrugerpriserne på madvarer i Vietnam udregnes prisen på denne diæt. Prisen på en diæt, som dækker minimumsbehovet for kalorier i 2010, er opgjort til 4.116.000 vietnamesiske Dong årligt per person.[3] Denne størrelse kalder vi $x^{\min food}$.

Udfordringen er nu at bestemme, hvor store de basale udgifter til andre ting end mad er (f.eks. tøj og sundhedsudgifter). Til at gøre det anvendes Engelkurven. Ud fra Engelkurven kan det beregnes, hvor stor en andel en husholdning vil bruge på mad, når vi antager, at de samlede udgifter er 4.116.000 Dong. Denne størrelse kalder vi $s^{\min}$. Det viser sig, at disse husholdninger har en andel, der er mindre end 1 ($s^{\min} < 1$). Det betyder, at disse husholdninger bruger mindre på mad, end der skal til for at opretholde de daglige minimumsindtag af kalorier. Heraf kan sluttes, at der er andre udgifter, som husholdningerne anser for så vigtige (basale), at de er villige til have et lavere indtag af kalorier end minimumsindtaget. Når fattigdomsgrænsen skal bestemmes, bør man derfor tage disse udgifter med. Størrelsen af disse øvige basale udgifter er $x^{\min nonfood} = x^{\min food}(1-s^{\min})$. Vi kan nu udregne de mindste samlede udgifter, som sikrer, husholdningerne vil have $x^{\min food}$ til mad og $x^{\min nonfood}$ til andre basale udgifter. Udgifterne til

$$\begin{aligned}\text{samlet minimum udgifter} &= x^{\min food} + x^{\min nonfood} \\ &= x^{\min food} + x^{\min food}(1 - s^{\min}).\end{aligned}$$

Fattigdomsgrænsen defineres som de samlede minimumsudgifter til mad og til at dække andre basale behov. Altså er fattigdomsgrænsen

$$\text{fattigdomsgrænse}{=}x^{\min food} + x^{\min food}(1 - s^{\min}).$$

[1] Alt den information, der skal bruges i opgaven står i opgaveteksten. Resultaterne i eksamenopgaven vil ikke være de samme som i rapporten, da der kun bruges et udsnit af data til eksamen.

[2] Der findes lidt forskel på hvor mange kalorier man regner for at være nødvendige. Tallet ligger mellem 1800 til 3000 kilokalorier om dagen.

[3] Dette svarer ca. til 5800 kr.

Adgang til data

Til eksamen i Økonometri I er der adgang til ti datasæt på Digital Eksamen. Følg disse instruktioner til at udvælge det korrekte datasæt for jeres gruppe:

1. Bestem jeres gruppenummer som det sidste ciffer i det mindste eksamensnummer blandt gruppemedlemmerne.
2. Download filen 'groupdataX.dta' fra Digital Eksamen, hvor X er lig gruppenummeret bestemt i trin 1.
3. *Eksempel:* En gruppe bestående af eksamensnumre 75, 82, 174 har gruppenummer 5 og downloader derfor 'groupdata5.dta' fra Digital Eksamen.
4. Åbn datafilen i STATA og verificer at data kan indlæses uden fejl.

Dokumentation af data

Data stammmer fra den vietnamesiske forbrugsundersøgelse foretaget i 2010. Data består af 7554 husholdninger, som har angivet deres udgifter til mad, samlede udgifter, samlet indkomst, husholdningens størrelse samt en dummy for om husholdningen bor på landet. Alle variablene refererer til hele hustandens forbrug/indkomst.

Tabel 1: Variable i 'groupdataX.dta'

Variabelnavn	Indhold
$foodexp_i$	Udgifter (årlige) til mad i 1000 vietnamesiske Dong
$totalexp_i$	Samlede (årlige) udgifter i 1000 vietnamesiske Dong
$income_i$	Årlig indkomst i 1000 vietnamesiske Dong
$hhsize_i$	Antal medlemmer i husholdningen
$rural_i$	Dummy for om husholdningen bor på landet

Noter: 'groupdataX.dta' indeholder konstruerede data og kan derfor ikke bruges til andre formål end at besvare denne eksamensopgave. Beregninger i denne opgave kan derfor afvige fra beregninger i originalmaterialet.

Beregning af fattigdomgrænsen i Vietnam

Eksamensopgaven tager udgangpunkt i en estimation af Engelkurven, altså følgende lineære regressionsmodel:

$$\frac{foodexp}{totalexp} = \beta_0 + \beta_1 \log(\frac{totalexp}{hhsize}) + \beta_2 \log(hhsize) + u \tag{1}$$

hvor log er den naturlige logaritme og u er fejlleddet.

Opgave 1 (20%)

1. Beskriv regressionsmodel (1).
2. Fortolk parametrene β_1 og β_2 og angiv deres forventede fortegn.
3. Udfør en deskriptiv analyse af data. Beregn den gennemsnitlige andel til mad for husholdninger på landet ($Rural = 1$) og i byerne ($Rural = 0$). Udregn også den gennemsnitlige andel til mad for husholdninger, som har et samlet forbrug per person, som er mindre end 4.116.000 Dong.

Opgave 2 (20%)

1. Udfør estimation af (1) ved OLS. Rapporter parameterestimater og relevante standardfejl i en tabel. Fortolk estimationsresultaterne. Redegør for, om OLS-estimatoren er konsistent.
2. Hvis der er målefejl i variablen *totalexp*, kan det påvirke OLS estimationen (se evt. spørgsmål 4 og 5). Derfor benyttes en IV estimation, hvor $\log(\frac{income}{hhsize})$ anvendes som instrument for $\log(\frac{totalexp}{hhsize})$. Diskuter og undersøg om $\log(\frac{income}{hhsize})$ kan anvendes som instrument. Udfør estimation af (1) ved IV estimation. Rapporter og fortolk estimationsresultaterne. Redegør for, under hvilke antagelser at IV-estimatoren er konsistent.
3. Udfør et test for om $\log(\frac{totalexp}{hhsize})$ er en endogen variabel. Angiv konklusionen og skriv hvad det betyder.

Opgave 3 (20%)

Når man estimerer Engelkurver, anvender man ofte en mere fleksibel funktionel form, hvor også variablen $\left[\log(\frac{totalexp}{hhsize})\right]^2$ indgår:

$$\frac{foodexp}{totalexp} = \beta_0 + \beta_1 \log(\frac{totalexp}{hhsize}) + \beta_2 \log(hhsize) + \beta_3 \left[\log(\frac{totalexp}{hhsize})\right]^2 + u. \tag{2}$$

1. Undersøg om model (1) bør udvides. Udfør et statistisk test for om model (2) kan reduceres til model (1). Angiv jeres foretrukne model og parameterestimater. Undersøg også om der er behov for at korrigere for husholdningens størrelse ($\beta_2 = 0$). Giv en fortolkning af parametrene i jeres fortrukne model.

2. Det oplyses, at i Vietnam vil en person skulle bruge 4.116.000 Dong for at kunne købe mad, der svarer til 2230 kilokalorier. Udregn $\hat{s}^{\min}$, dvs. den forventede andel af budgettet, som vil anvendes til mad for en husholdning på 4 personer, som har et total forbrug på 4.116.000 Dong per person. Angiv andelen som bruges til mad, $\hat{s}^{\min}$ og konfindensintervallet for $\hat{s}^{\min}$. Brug teksten i starten af opgaven og forklar, hvordan $\hat{s}^{\min}$ skal fortolkes.

3. Fattigdomsgrænsen kan nu beregnes som

$$\text{fattigdomsgrænse} = x^{\min food} + (1 - \hat{s}^{\min}) \cdot x^{\min food}.$$

   Beregn fattigdomsgrænsen. Undersøg dernæst, hvor mange procent af husholdningerne, som har et total forbrug per person, som ligger under fattigdomsgrænsen. Undersøg om der er forskel på andelen af fattige på landet og i byerne.

Opgave 4 (20%)

Betragt følgende statistiske model:

$$\frac{y_i^*}{x_i^*} = \beta_0 + \beta_1 \log(x_i^*) + u_i$$

hvor $i = 1, ..., n$. Vi antager, at MLR.1-MLR.4 er opfyldt. Vi antager, at y_i^* og x_i^* er uobserverbare. I stedet observeres y_i og x_i med målefejl. Vi antager, at der er multiplikative målefejl a_i ($a_i > 0$), og at målefejlen er den samme for y_i og x_i :

$$\begin{aligned} y_i &= y_i^* a_i \\ x_i &= x_i^* a_i \end{aligned}$$

Det betyder, at hvis $a_i = 1.05$ så overvurderer husholdning i både y_i^* og x_i^* med 5 procent og hvis $a_i = 0.96$ så undervurderer husholdning i både y_i^* og x_i^* med 4 procent. Vi antager også, at målefejlene a er uafhængige af u, x^*og y^*. Vi antager desuden, at

$$\begin{aligned} Var(\log(x^*)) &= \sigma_{x^*}^2 \\ Var(\log(a_i)) &= \sigma_a^2 \end{aligned}$$

1. Diskuter om antagelserne om målefejl er rimelige i forhold til en model, hvor y er udgifter til mad og x er de totale udgifter.

2. Opskriv regressionsmodellen for de observerede variable y_i og x_i.

3. Lad $\hat{\beta}_1$ være OLS estimatoren af β_1 i en model

$$\frac{y_i}{x_i} = \beta_0 + \beta_1 \log(x_i) + v_i.$$

   Udregn, grænsen i sandsynlighed for OLS estimatoren, når antallet af observationer går mod uendeligt: $p\lim \hat{\beta}_1$.

4. Antag at $\beta_1 < 0$. Angiv om den asymptotiske bias er negativ eller positiv.

Opgave 5 (20%)

Denne opgave går ud på at illustrere i et simulationsstudie, hvordan målefejl påvirker OLS estimatoren. Vi tager udgangspunkt i modellen fra opgave 4. Betragt følgende datagenererende proces (DGP) :

$$\frac{y_i^*}{x_i^*} = \beta_0 + \beta_1 \log(x_i^*) + u_i \tag{3}$$

$$\beta_0 = 1.8,\ \beta_1 = -0.2 \tag{4}$$

$$x^* \sim U(500, 1000), u \sim U(-0.125, 0.125) \tag{5}$$

Vi antager nu, at y^*og x^* er de sande værdier, men at de er uobserverede. I stedet observerer vi y og x, som også indeholder en målefejl. Vi antager, at målefejlene er multiplikative og de samme for x og y

$$\begin{aligned} y_i &= y_i^* a_i \\ x_i &= x_i^* a_i. \end{aligned}$$

Om målefejlene gælder

$$a = \exp(e) \text{ hvor } e \sim U(-0.125, 0.125).$$

Desuden antages målefejlene at være uafhængige af y^*og x^*.

1. Undersøg ved et simulationseksperiment, hvilke konsekvenser det har for OLS estimatoren $\hat{\beta}_1$, hvis vi anvender følgende regressionsmodel

   $$\frac{y_i}{x_i} = \beta_0 + \beta_1 \log(x_i) + \varepsilon_i$$

   Vælg et seed nummer efter eget valg og angiv det i besvarelsen. Lav eksperimentet med en stikprøve på 2000 observationer og brug 1000 replikationer. Rapporter deskriptiv statistik og et histogram til at dokumentere simulationseksperimentet. Skriv hvad eksperimentet viser. [*Hint*: Følgende STATA kode vil generere x^*, u, y^* og a :

```
generate xstar = 500+500*runiform()
generate u = 0.25*(runiform()-0.5)
generate ystar = xstar*(1.8-0.2*ln(xstar)+u)
generate a = exp(0.25*(runiform()-0.5))
```

]

2 Udfør endnu et simulationseksperiment, hvor vi nu antager, at målefejlene i x og y er ukorrelerede:

$$
\begin{aligned}
y_i &= y_i^* c_i \\
x_i &= x_i^* a_i,
\end{aligned}
$$

hvor

$$
\begin{aligned}
a_i &= \exp(e) \text{ hvor } e \sim U(-0.125, 0.125) \\
c_i &= \exp(v) \text{ hvor } v \sim U(-0.125, 0.125)
\end{aligned}
$$

og e og v er uafhængige.

Lav eksperimentet med en stikprøve på 2000 observationer og brug 1000 replikationer. Rapporter deskriptiv statistik og et histogram til at dokumentere simulationseksperimentet. Skriv hvad eksperimentet viser. Forklar, hvorledes resultaterne i dette studie afviger fra resultaterne med klassiske målefejl.