

Eksamen på Økonomistudiet sommer 2020

Økonometri I

Tag-hjem eksamen: 19. juni, 2020, kl.10.00-22.00

Praktiske anvisninger til eksamen i Økonometri I:

- Eksamenen kan besvares individuelt eller i grupper af max 3 studerende.
- Læs hele eksamensopgaven igennem før, der udarbejdes svar.
- Besvar alle spørgsmål og delspørsmål 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 normalsider (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 2.5 cm.
- Rapporten skal forsynes med en forside og der skal udarbejdes en ansvarsfordeling, således de enkelte gruppemedlemmers bidrag til besvarelsen fremgår. Alle studerende i gruppen skal være ansvarlig for mindst et **a-, b- og c- spørgsmål**. Dette gøres ved at anvende skabelonen 'forside.doc', som er tilgængelig på Digital Eksamen.
- Eksamenen besvares ved at uploade rapporten i pdf-format til Digital Eksamen senest klokken 22 den 19. juni. Bemærk at der kun skal uploades én rapport for hver gruppe. Dette gøres ved at følge instruktionerne beskrevet i følgende video: <http://video.ku.dk/digital-eksamen-gruppeaflevering-1>.
- Besvarelsen uploades på Digital Eksamen som én pdf.fil (inkl. bilag) navngivet udelukkende med eksamensnummeret, f.eks. 1234.pdf eller 55.pdf
- Udover selve rapporten skal der også uploades 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. juni, såfremt der er problemer med at downloade data. Hvis der mod forventning skulle være fejl eller uklarheder i eksamensopgaven, vil I blive kontakte af eksamensadministrationen via email. I skal derfor sørge for at tjekke jeres email under eksamenen.
- 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 eksamenen. Enhver form for kommunikation betragtes som eksamenssnyd og vil blive behandlet herefter. Reglerne for eksamenssnyd er beskrevet på Din Uddannelsesside og i Rammestudieordningens afsnit 4.12.
- Dette eksamenssæt består af 8 sider i alt inkl.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, som om det er din egen idé eller dine tanker
- Genbruger dele af en opgave, som du tidligere har indleveret og fået en bestået 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:

"Hvilken betydning har arbejdsløshed på kriminalitet"¹

Beckers teori fra 1968 om kriminalitet bygger på, at det enkelte individ laver en afvejning af fordele og ulemper ved legale aktiviteter i forhold til illegale aktiviteter. Hvis den samlede nytte ved illegale aktiviteter er større end ved legale aktiviteter, så vil individet vælge at engagere sig i illegale aktiviteter. Forskere har tidligere set på forhold vedr. ulemper ved kriminel aktivitet så som risikoen for at blive straffet og straffens størrelse. Andre forskere har set på betydningen af arbejdsmarkedsforhold på kriminalitet, hvor arbejdsmarkedsforhold påvirker fordele og ulemper ved legal aktivitet og herved valget om at indgå i illegal aktivitet. Empiriske analyser viser, at et højt lønniveau er forbundet med lavere kriminalitet. Dette kan forklares med, at når lønnen er højere ved legal aktivitet (at have et arbejde), så vil fordelene ved legal aktivitet være højere og overstige fordelene ved illegal aktivitet. De empiriske resultater for, hvordan arbejdsløshed påvirker kriminalitet, er derimod mindre entydige. Det har været nævnt, at de forskellige resultater kan skyldes den måde, arbejdsløshed er blevet opgjort på, og at der er blevet anvendt forskellige økonometriske metoder.

I denne opgave er fokus på unge mænd i alderen 15-29, hvor I skal undersøge, hvordan arbejdsløshed påvirker kriminalitetsraten for unge mænd.

¹Denne eksamensopgave er inspireret af Fougère, Denis, et al. "Youth Unemployment and Crime in France." *Journal of the European Economic Association*, vol. 7, no. 5, 2009, pp. 909–938. I denne opgave anvendes andre data, så resultaterne kan afvige fra artiklen.

Adgang til data

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

1. Bestem jeres gruppenummer som det sidste ciffer i det mindste eksamensnummer blandt gruppemedlemmerne. Bemærk at I skal bruge **eksamensnummer** ikke jeres KUident.
2. Download filen 'groupdataX.dta' fra Digital Eksamen, hvor X er lig gruppenummeret bestemt i trin 1.
3. *Eksempel:* En gruppe bestående af eksamensnumrene 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 stammer fra Statistikbanken, Danmarks Statistik. Her er indhentet oplysninger om alle 98 kommuner i Danmark (Christiansø er ikke medtaget). Hver kommune er observeret to gange: i året 2012 og året 2017. I analysen fokuseres på mænd i alderen 15-29 år. Datasættet består af 196 observationer og 11 variable.

Tabel 1: Variable i 'groupdataX.dta'

Variabelnavn	Indhold
$Krimrate_{it}$	Andelen af 15-29 årige mænd som er dømt i år t i kommune i
$Ledighed_1524_{it}$	Andelen af 15-24 årige mænd som er arbejdsløse i år t i kommune i
$Ledighed_2529_{it}$	Andelen af 25-29 årige mænd som er arbejdsløse i år t i kommune i
$Studentrate_{it}$	Andelen af 15-29 årige mænd som er studerende i år t i kommune i
$Koen_balance_{it}$	Andelen af 15-29 årige mænd i forhold til alle 15-29 årige i år t i kommune i
$Storby_{it}$	Dummy for KBH, FRB, Aarhus, Odense, Aalborg og Esbjerg
$pop_mand1529_{it}$	Antallet af 15-29 årige i år t i kommune i
Dz_{it}	Forventet vækstrate i antal job i perioden 2012-2017 for kommune i
aar_{it}	År
kom_nr_{it}	Kommunenr. i kommune i
$kommune_{it}$	Kommunenavn i kommune i

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.

Variablen for den forventede vækst i antal job i kommunen er konstrueret ud fra de nationale jobvækstrater i perioden 2012 til 2017 inden for hver sektor j : g_j . Vi benytter nu andelen af jobs i kommune i i 2012 i sektor j : τ_{ij} . Den forventede vækstrate i job i kommune i er:

$$Dz_{it} = \sum_{j=1}^J g_j \tau_{ij}.$$

Analyse af betydningen af ungdomsarbejdsløshed på kriminalitet

Opgave 1 (20%)

- a Udfør en deskriptiv analyse af datasættet `groupdataX.dta`. Beregn gennemsnittet for følgende variable: *krimrate*, *ledighed_1524*, *ledighed_2529*, *studentrate* og *koen_balance*. Gennemsnittet beregnes separat for de to tidsperioder.
- b Beskriv regressionsmodel (1):

$$\begin{aligned} \text{krimrate}_{it} = & \beta_0 + \beta_1 \text{ledighed_1524}_{it} + \beta_2 \text{ledighed_2529}_{it} + \beta_3 \text{studentrate}_{it} \\ & + \beta_4 \text{koen_balance}_{it} + \beta_5 \text{Storby}_{it} + \beta_6 D2017_{it} + v_{it} \quad t = 2012, 2017 \end{aligned} \quad (1)$$

hvor $D2017$ er en dummy for 2017. Angiv fortolkningen af parametrene β_1 og β_2 og det forventede fortegn for β_1 og β_2 . Estimer modellen med OLS og rapporter parameterestimer og relevante standardfejl i en tabel. Kan OLS estimatet for β_1 fortolkes som den estimerede kausale effekt af ungdomsledighed på kriminalitetsraten?

- c Vi antager nu, at variansen på fejleddet er givet ved

$$\text{Var}(v_{it}|X_{it}) = \frac{\sigma^2}{\text{pop_mand1529}_{it}},$$

hvor X_{it} indeholder alle variablene bortset fra *krimrate*. Diskuter hvorfor det er en rimelig antagelse og lav dernæst en vægtet estimation, hvor model (1) estimeres med WLS. Angiv vægtene som benyttes. Angiv hvilke antagelser der skal være opfyldt for, at WLS er konsistent.

Opgave 2 (20%)

- a Vi antager nu, at fejleddet v_{it} kan skrives som $v_{it} = a_i + u_{it}$. Desuden antages, at *ledigheden_1524* og *ledigheden_2529* er korreleret med a_i men ukorreleret med u_{it} . Diskuter, om denne antagelse er realistisk. De øvrige variable antages at være ukorrelerede med v . Angiv en konsistent estimator for modellens parametre i dette scenarie og estimer modellen med denne estimator og angiv de relevante standardfejl.
- b Vi antager nu, at variansen af u_{it} er givet ved

$$\text{Var}(u_{it}|X_{it}) = \frac{\sigma_u^2}{\text{pop_mand1529}_{i2017}} \quad t = 2012, 2017. \quad (2)$$

Estimer modellen med en vægtet estimator, som er konsistent under antagelserne angivet i spørgsmål 2.a. Angiv hvilke vægte, som anvendes.

- c Angiv fordele og ulemper ved estimatoren i spørgsmål 2.a og 2.b. Test hypotesen, at det kun er ungdomsledigheden (*ledighed_1524*) og tiden ($D2017$), som påvirker ungdomskriminaliteten i den foretrukne model. Opstil nulhypotesen og angiv teststørrelsen. Forklar hvorledes testet udføres og skriv konklusionen på testet.

- d Lav en graf, der viser sammenhængen mellem ændringen i kriminalitetsraten fra 2012 til 2017 og ændringen i ungdomsledigheden ($ledighed_1524$) fra 2012 til 2017. Grafen skal vise både data og model-prædiktioner fra den fortrukne model.

Opgave 3 (20%)

Vi tager i denne opgave udgangspunkt i den simple model

$$krimrate_{it} = \beta_0 + \beta_1 ledighed_1524_{it} + \beta_6 D2017 + a_i + u_{it}, \quad (3)$$

Vi antager nu, at $ledighed_1524_{it}$ potentielt er korreleret med både a_i og u_{it} .

- a Vi ser nu på følgende regressionsmodel

$$\Delta krimrate_{i2017} = \gamma_0 + \gamma_1 \Delta ledighed_1524_{i2017} + \omega_{i2017}. \quad (4)$$

Vis hvordan model (4) er relateret til model (3). $\Delta ledighed_1524_{it}$ er potentielt en endogen variabel. Vi ønsker at anvende den forventede vækstrate i antallet af job (Dz_{it}) som instrument for $\Delta ledighed_1524_{it}$. Angiv hvilke antagelser, som skal være opfyldt for at have et validt instrument. Diskuter om de er opfyldt. Lav IV estimationen og rapporter estimationsresultater og relevante standardfejl.

- b Undersøg om $\Delta ledighed_1524_{it}$ er en eksogen variabel.
- c Vi antag nu, som i spørgsmål 2.b, at variansen af u_{it} er givet ved (2). Udfør en vægtet IV estimation af model (3). Sammenlign estimationsresultaterne fra opgave 1, 2 og 3 og kommenter på evt. forskelle. Angiv den kausale effekt af ungdomsarbejdsløshed på kriminalitetsraten baseret på den fortrukne estimation.

Opgave 4 (20%)

Betragt følgende statistiske model:

$$y_{it} = \beta_0 + \beta_1 x_{it} + a_i + u_{it} \quad (5)$$

hvor $i = 1, \dots, n$ er antallet af kommuner. Vi observerer hver kommune til tidspunkt 1 og 2. For modellen antages følgende:

$$Cov(x_{it}, a_i) = \delta, Cov(x_{it}, u_{it}) = \rho, Var(x_{it}) = \sigma_x^2 > 0. \quad (6)$$

Desuden antages uafhængighed mellem kommunerne.

- a Udregn den asymptotiske bias for OLS estimatoren til model (5): $p\lim(\hat{\beta}_1^{OLS}) - \beta_1$. Angiv betingelserne på parameterne δ og ρ for, at OLS estimatoren er konsistent.

b Vi antager nu, at vi har en tredje variabel z . Om z gælder der:

$$\begin{aligned} Cov(x_{it}, z_{it}) &= \sigma_{zx} \neq 0 \\ Cov(z_{it}, u_{it}) &= 0, Cov(z_{it}, a_i) = \theta. \end{aligned} \quad (7)$$

Hvis vi anvender z_{it} som instrument for x_{it} i model (5), er IV estimatoren givet ved

$$\hat{\beta}_1^{IV} = \frac{\sum_{t=1}^2 \sum_{i=1}^n (y_{it} - \bar{y})(z_{it} - \bar{z})}{\sum_{t=1}^2 \sum_{i=1}^n (x_{it} - \bar{x})(z_{it} - \bar{z})}.$$

Redegør for, hvornår IV estimatoren er konsistent.

c Forstå en estimator som er konsistent for alle værdier af δ, ρ og θ under antagelserne (6) og (7). Redegør for, at den foreslåede estimator er konsistent. Angiv hvis der er behov for yderligere antagelser.

Opgave 5 (20%)

Denne opgave går ud på at sammenligne forskellige estimators for model (5) ved et simulationssstudie. Vi tager udgangspunkt i modellen fra opgave 4. Betragt følgende datagenererende proces (DGP):

$$\begin{aligned} y_{it} &= \beta_0 + \beta_1 x_{it} + a_i + u_{it} \\ a_i &\sim N(0, 1), u_{it} \sim N(0, 1), x_{it}^* \sim N(4, 4) \\ x_{it} &= x_{it}^* + \delta a_i + \rho u_{it} \\ z_{it} &= -0.5 x_{it}^* + \theta a_i \\ \beta_0 &= 3, \beta_1 = 1, \delta = -0.7, \rho = -0.2, \theta = 0.5 \end{aligned}$$

Det antages, at x_{it}^*, a_i og u_{it} er indbyrdes uafhængige.

a Lav eksperimentet med en stikprøve på 100 kommuner og to tidsperioder og brug 1000 replikationer. I simulationsstudiet skal OLS estimatoren $\hat{\beta}_1^{OLS}$ sammenlignes med First Difference estimatoren $\hat{\beta}_1^{FD}$ og IV estimatoren $\hat{\beta}_1^{IV}$ (som defineret i opgave 4.b). Rapporter deskriptiv statistik til at dokumentere simulationseksperimentet. Skriv hvad eksperimentet viser.

[Hint: Følgende STATA kode vil generere et panel datasæt med to perioder og følgende variable: a_i, t , og en variabel som angiver kommune nr (id):

```
*SET NUMBER OF INDIVIDUALS
set obs 100
*DATA GENERATING PROCESS
```

```

generate a = rnormal()
gen id=_n
expand 2
bysort id: gen t=_n
]

```

- b Sammenlign bias i simulationseksperimentet for OLS estimatoren og IV estimatoren med de teoretiske bias udregnet i spørgsmål 4.a og 4.b.
- c Implementer følgende estimator i simulationsprogrammet:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{100} \Delta y_{i2}(z_{i2} - \bar{z}_2)}{\sum_{i=1}^{100} \Delta x_{i2}(z_{i2} - \bar{z}_2)}.$$

Udfør simulationsstudiet og lav en tabel og en graf med histogrammer, hvor de tre estimators sammenlignes. Kommenter på simulationsresultaterne.