

# Eksamen på økonomi-studiet vinter 2020-2021

## Økonometri I

Tag-hjem eksamen: 17. december, 2020, kl. 10.00-22.00

### Praktiske anvisninger til eksamen i Økonometri I:

- Eksamenen kan besvares individuelt eller i grupper af maks. 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 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 (husk at trykke "Aflever") senest klokken 22 den 17. december 2020. 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 en pdf-fil (inklusiv bilag) navngivet udelukkende med eksamensnummeret, f.eks. 1234.pdf eller 55.pdf.
- Ud over 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.
- Daniel le Maire kan kontaktes på telefon 3532 3063 klokken 10-12 den 17. december, 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](mailto: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 7 sider i alt inklusiv forside.

## Pas på du ikke begår eksamenssnyd!

Det er for eksempel 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:

John Maynard Keynes forudså i 1929, at den hurtige spredning af teknologi til at automatisere produktionsprocesser ville lede til en massiv teknologisk arbejdsløshed. Selv om denne forudsigelse indtil videre ikke har holdt stik, er der en fornyet interesse i at forstå om robotteknologi og kunstig intelligens vil betyde en nedlæggelse af store dele af de eksisterende jobs. Der findes en hel del empirisk evidens for, at automatiseringen af jobs med færdigheder på lavt og mellem niveau omkring slutningen af det 20. århundrede har betydet mere ulighed. Der er kun i de senere år kommet direkte evidens for effekten af robotisering på arbejdsmarkedet. I denne eksamensopgave undersøger effekten af industrirobotter på beskæftigelsen.<sup>1</sup>

## 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 jeres **eksamensnummer** ikke jeres KUident.
2. Download filen 'groupdataX.dta' fra Digital Eksamen, hvor X er lig gruppenummet 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 data-filen i STATA og verificer, at data kan indlæses uden fejl.

## Dokumentation af data

Datasættet indeholder information for 300 regioner i 1995, 2005 og 2015. Alle regioner er del af det samme land. Variablene, der er indeholdt i datasættet, er beskrevet i tabel 1. Data i denne opgave er konstruerede og kan kun bruges til at besvare denne opgave.

## Opgave 1 (10%)

- a Lav en tabel med beskrivende statistik separat for hvert år.
- b Lav en figur, der viser den samlede nationale udvikling over tid i beskæftigelsen for personer med kompetencegivende uddannelse og beskæftigelsen for personer uden kompetencegivende uddannelse.

---

<sup>1</sup>Eksamensopgaven bygger løst på Dauth, Findeisen, Suedekum og Woessner (2020), "The Adjustment of Labor Markets to Robots", working paper, og Acemoglu og Restrepo (2020), "Robots and Jobs: Evidence from US Labor Markets", *Journal of Political Economy*, Vol. 128, No. 6. Data er simuleret og adskiller sig på adskillige punkter fra de nævnte papirer. Bl.a. antages det i denne opgave, at data indeholder information om industrirobotter på regionsniveau. I Dauth et al. (2020) og Acemoglu og Restrepo (2020) har de imidlertid kun information om industrirobotter på *nationalt* branche niveau.

Variabelnavn(e)	Variabelbeskrivelse
region	Geografisk regionsnr.
aar	År
besk_ejkomp	Antal beskæftigede med ikke-kompetencegivende uddannelse
besk_komp	Antal beskæftigede med kompetencegivende uddannelse
robot	Antal industrirobotter per 1000 arbejder i region
nord	Dummy for om region ligger i den nordlige del af landet
alder	Gennemsnitsalder i arbejdsdygtig alder i region
kvinde	Andel af kvinder i arbejdsdygtig alder i region
branche1-branch8	Andel af branche1-branch8 i region i 1995
udlandrobot1-udlandrobot8	Ændringen i antallet industrirobotter i branche1-branch8 i andre højindkomst lande i de seneste 10 år

Tabel 1: Oversigt over variable.

## Opgave 2 (25%)

Vi ønsker at estimere effekten af robotter på beskæftigelsen af henholdsvis arbejdere med og uden kompetencegivende uddannelse. Lad  $i$  angive region og  $t$  angive år. Outcome variablene, logaritmen til  $besk_{it}^{komp}$  og logaritmen til  $besk_{it}^{ejkomp}$ , angiver logaritmen til beskæftigelsen af henholdsvis arbejdere med kompetencegivende uddannelse og arbejdere uden kompetencegivende uddannelse. Vi vil benytte den følgende ligning

$$\log(besk_{it}^{type}) = \beta_0 + \beta_1 robot_{it} + \beta_2 nord_i + \beta_3 alder_{it} + \beta_4 kvinde_{it} + \beta_5 d_{it}^{1995} + \beta_6 d_{it}^{2005} + a_i + u_{it} \quad (1)$$

hvor  $type = \{komp, ejkomp\}$ ,  $d_{it}^{1995}$ ,  $d_{it}^{2005}$  er tidsdummies,  $a_i$  er en uobserverbar tidskonstant heterogenitet og  $u_{it}$  er et tidsvarierende fejllid.

- Estimer ligning (1) med pooled OLS for hver type af beskæftigede. Rapporter parameter estimater og standardfejl. Foretag et formelt test for om  $\beta_1 = 0$ . Fortolk på resultaterne. Kan vi give resultaterne en kausal fortolkning?
- Estimer ligning (1) med pooled OLS for hver af de to typer af beskæftigede, hvor kun konstantleddet og robotvariablen er inkluderet. Test restriktionerne, der er lagt på den restriktede model, i forhold til urestriktede model estimeret med pooled OLS i det foregående delspørgsmål.
- Undersøg om der er heteroskedasticitet i den urestriktede model i ligning (1) ved at benytte Breusch-Pagan testet. I givet fald det konkluderes, at der er heteroskedasticitet, undersøg grafisk hvilke(n) variable, der styrer variansen.

## Opgave 3 (25%)

- a Tag førstedifferenser af ligning (1) og estimer den resulterende ligning, hvor den afhængige variabel er henholdsvis logaritmen til  $besk_{it}^{komp}$  og logaritmen til  $besk_{it}^{ejkomp}$ . Rapporter parameterestimer og standardfejl. Fortolk på resultaterne.
- b Diskuter hvordan  $\Delta robot$  kunne tænkes at være endogen i ligningen, der er estimeret i det foregående delspørgsmål og instrumenter  $\Delta robot$  ved at konstruere et såkaldt *shift-share* instrument. Her er *share*-delen andelen af arbejdspladser, der hører til branche  $j = 1, 2, \dots, 8$  i 1995, mens *shift*-delen er ændringen i antallet af industrirobotter per 100.000 arbejder i de sidste 10 år for andre højindkomst lande opgjort separat for hver branche  $j = 1, 2, \dots, 8$ . Dvs. *shift-share* instrumentet er givet ved

$$z_{it} = \sum_{j=1}^8 \text{branche}_{ij,1995} \text{udlandrobot}_{jt} \quad (2)$$

hvor *udlandrobot* er *shift*-delen af instrumentet og er opgjort for tidsperioderne 1995-2005 og 2005-2015. Estimer ligningen med førstedifferenser fra det foregående delspørgsmål med 2SLS, hvor  $\Delta robot_{it}$  instrumenteres med  $z_{it}$ . Rapporter estimationsresultaterne for både *first stage* og *second stage* af 2SLS.

- c Undersøg om  $\Delta robot$  er endogen, og diskuter kort om instrumentet  $z_{it}$  er et validt instrument. [Til diskussionen om validitet af instrumentet kan løsning af opgave 4 og 5 muligvis hjælpe.]

## Opgave 4 (20%)

I denne delopgave vil vi teoretisk undersøge *shift-share* instrumentet for en simpel lineær regressionsmodel. Vi har, at

$$y_i = \beta_0 + \beta_1 x_i + u_i \quad (3)$$

hvor  $i = 1, 2, \dots, n$ , og hvor  $x_i$  er endogen. Vi har imidlertid et instrument  $z_i$ , så vi kan estimere 2SLS med den følgende første trins ligning

$$x_i = \pi_0 + \pi_1 z_i + \epsilon_i \quad (4)$$

- a Opskriv estimatorerne,  $\hat{\pi}_0$  og  $\hat{\pi}_1$  i det første trin, hvor vi regresserer den endogene variabel  $x$  mod  $z$  ved at estimere den simple lineære regressionsmodel.
- b Antag nu, at vores instrument  $z$  er et *shift-share* instrument. Vi antager, at der er to brancher, som har haft forskellig vækst, og at vi kan bruge det til at konstruere *shift-share* instrumentet. Lad andelen være givet ved  $s_1$  for branche 1 og  $s_2$  for branche 2. Vi har, at de to andele summerer til 1, så det er en fordel at bruge, at  $s_1 = s$  og  $s_2 = 1 - s$ . Vi kan dermed skrive vores *shift-share* instrument som  $z_i = s_i g_1 + (1 - s_i) g_2 = g_2 + s_i (g_1 - g_2)$ , hvor  $g_1$  er væksten for branche 1 og  $g_2$  er væksten for branche 2. Udregn estimatorerne,  $\hat{\pi}_0$  og  $\hat{\pi}_1$  ved at benytte definitionen af  $z$ .

- c Udregn den prædikterede værdi,  $\hat{x}_i = \hat{\pi}_0 + \hat{\pi}_1 z_i$ , fra første trin, når vi bruger instrumentet  $z_i$ . Sammenlign med den prædikterede værdi  $\hat{x}_i$ , hvis vi i stedet for havde benyttet de initiale andele,  $s_i$  som instrument. Hvilken betydning har det om vi vælger  $z$  eller  $s$  som instrument?

## Opgave 5 (20%)

I denne opgave skal vi lave en simulation for at se nærmere på IV estimation med et shift-share instrument. Antag at vi har et tværsnitsdatasæt med  $N$  regioner. Vi vil begynde med at konstruere et shift-share instrument for region  $i$ ,  $z_i = \sum_{j=1}^J g_j s_{ij}$ , der består af en *shift*-del,  $g_j$ , og en *share*-del,  $s_{ij}$ . Vi vil sørge for, at *share*-variablene summerer til 1. For at gøre dette defineres en hjælpevariabel  $q_{ij} \sim U(0, 1)$  for de  $j = 1, 2, \dots, J$  forskellige brancher. Dermed kan vi skrive *share*-variablen som  $s_{ij} = \frac{q_{ij}}{\sum_{j=1}^J q_{ij}}$ . Hvis vi lader antallet af brancher være angivet med den globale variabel  $J$ , så kan vi danne *share*-variablen  $s_{ij}$  for  $j = 1, 2, \dots, J$  med den følgende Stata kode:<sup>2</sup>

```
forvalues j=1/$J {
    gen q'j'=runiform()
}
egen naevner=rowtotal(q*)
forvalues j=1/$J {
    gen s'j'=q'j'/naevner
}
```

hvor funktionen `rowtotal(q*)` for hver observation summerer alle variable, der begynder med `q`. Derfor er det vigtigt for, at koden virker efter hensigten, at der ikke er andre variable i datasættet, der begynder med `q`.

For at gøre det simpelt at udregne *shift*-variablen, defineres den som  $g_j = j/J$  for  $j = 1, 2, \dots, J$ .

Betragt derefter den følgende model

$$y_i = \beta_0 + \beta_1 x_i + u_i \quad (5)$$

hvor  $i = 1, 2, \dots, N$ . Lad  $u_i \sim N(0, 1)$ ,  $x_i = 1 + 1.5z_i + 0.2u_i$ . Ydermere antager vi, at  $N = 1000$ , samt  $\beta_0 = 1$  og  $\beta_1 = 2$ . Til Monte-Carlo simulationerne nedenfor benytter vi 500 repetitioner.

- Foretag et simulationsstudie, hvor der antages, at  $J = 2$ . I simulationen estimeres ligning (5) med OLS og med 2SLS, hvor instrumentet  $z_i$  benyttes til at instrumentere  $x_i$ . Lav en deskriptiv tabel for  $\beta_1$  estimerterne og et histogram med fordelingen af de to sæt estimerter.
- Foretag et simulationsstudie, hvor der antages, at  $J = 2$ . I simulationen estimeres ligning (5) med 2SLS, hvor *share*-variablen  $s_{i1}$  benyttes som instrument. Sammenlign  $\beta_1$  estimerterne med 2SLS, hvor *shift-share* instrumenten bliver benyttet.

<sup>2</sup>Til de første to spørgsmål kan man nøjes med at benytte  $J = 2$ , og så kan man klare sig med en simplere kodning.

- c Foretag et simulationsstudie med  $J = 5$  ved at estimere ligning (5) med 2SLS, hvor  $z_i = \sum_{j=1}^5 g_j s_{ij}$  benyttes som instrument, og ved at estimere ligning (5) med 2SLS, hvor share-variablene  $s_{i1}, s_{i2}, \dots, s_{i4}$  benyttes som instrumenter. Lav en deskriptiv tabel for estimerne og et histogram med fordelingen af de to sæt estimer.