

Eksamen på Økonomistudiet sommeren 2021

Økonometri I

Tag-hjem eksamen: 15. juni, 2021, 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 normale sider (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. 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.
- Eksamen besvares ved at uploade rapporten i PDF-format til Digital Eksamen senest klokken 22 den 15. juni. 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-gruppe>
- 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 15. 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 eksamen.
- 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 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 inkl. 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 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:

"Syg af arbejde eller for syg til at arbejde"

Opgaven går ud på at undersøge, hvilken betydning pensionering har på helbredet. I litteraturen er forskellige hypoteser om effekten af pensionering på helbredet blevet fremsat. En hypotese er, at arbejdet er nedslidende, og det derfor har en gavnlig effekt på helbredet at gå på pension. En anden hypotese går på, at arbejdet er med til at holde os i gang, og derfor har det en gavnlig effekt på helbredet at fortsætte med at arbejde.

De fleste deskriptive analyser viser, at personer, som er pensioneret, har dårligere helbred end personer, som fortsat arbejder. Problemet med disse deskriptive studier er, at de ikke tager højde for, om dem, som går tidligt på pension, måske gør det på grund af helbredsproblemer. Det betyder, at disse deskriptive studier ikke nødvendigvis beskriver den kausale effekt af pensionering på helbred. Der er lavet en del empiriske studier af effekten af pensionering på helbredet, konklusionerne har ikke være entydige.

I denne opgave undersøges den kausale effekt af pensionering på helbred. Vi benytter her to mål for helbred. Det første mål er tidlig død: død mellem 66-70 år.¹ Det andet helbredsmaal, vi benytter, er antallet af lægebesøg som 70 årige. Vi benytter data fra en tilfældig stikprøve af personer født i 1938-1940.²

I opgaven benyttes også en pensionsreform i 1999, hvor pensionsalderen for folkepensionen blev sænket fra 67 år til 65 år. Før 1999 kunne personer i Danmark gå på folkepension, når de fyldte 67 år. I 1999 ændrer man så folkepensionsalderen, så man kunne gå på folkepension, fra man fyldte 65 år. Den nye regel blev introduceret således, at den gjaldt for alle, som var født efter d. 1. juli 1939. Det betød, at den officielle folkepensionsalder for personer født i perioden 1938-1940 var følgende:

Født	Folkepensionsalder
1.1 1938 - 30.6 1939	67 år
1.7 1939 - 31.12 1940	65 år

Implementeringen af denne reform benyttes i opgaven.

¹Middellevetiden i Danmark er i dag over 80 år.

²Opgaven er inspireret af artiklen Nick Fabrin Nielsen (2019): "Sick of Retirement", Journal of Health Economics vol 65, pp. 133-152.

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 (husk at anvende jeres eksamensnummer og ikke jeres KU-ident).
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 består af et tilfældig udtræk af personer født i 1938-1940, som stadig er på arbejdsmarkedet som 65 årig og bosiddende i Danmark. Datasættet indeholder 1.530 observationer og otte variable. Variablene er beskrevet nedenfor.

Tabel 1: Variable i 'groupdataX.dta'

Variabelnavn	Indhold
fodaar	Fødselsår
fodmaaned	Fødselsmåned
kvinde	Dummy (1: kvinde, 0: mand)
indkomst	Årlig indkomst ved 65 år
udd_kat	Uddannelseskategorier (1: grundskole, 2: faglært, 3: kort udd, 4: mellem og lang udd)
pension66	Dummy (1: pensionering før 66 år, 0 ellers)
dod70	Dummy (1: død inden 70 år, 0: i live som 70 årig)
laege70	antal af lægebesøg som 70 årige

Noter: 'groupdataX.dta' indeholder konstruerede data og kan derfor ikke .
bruges til andre formål end at besvare denne eksamensopgave

Effekten af pensionering på helbredet

Eksamensopgaven tager udgangspunkt i en estimation af en lineær sandsynlighedsmodel, som beskriver sandsynligheden for at dø, inden man er fyldt 70 år. Regressionsmodellen er givet ved:

$$\begin{aligned} \text{dod70} = & \beta_0 + \beta_1 \text{pension66} + \beta_2 \log(\text{indkomst}) + \beta_3 \text{kvinde} + \\ & \beta_4 \text{faglaert} + \beta_5 \text{kortudd} + \beta_6 \text{mludd} + u \end{aligned} \quad (1)$$

hvor *faglaert* er en dummyvariabel for faglært uddannelse, *kortudd* er dummyvariabel for kort uddannelse og *mludd* er en dummyvariabel for mellem- og lang uddannelse, *log* er den naturlige logaritme og *u* er fejleddet.

Opgave 1 (20%)

- Beskriv regressionsmodel (1) og fortolk parametrene β_1 og β_3 .
- Konstruer en dummyvariable (*FP*), som er 1 for de personer, hvor den officielle folkepensionsalder er 65 år, og 0 for de personer hvor den officielle folkepensionsalder er 67 år.³ Konstruer de relevante dummyvariable for uddannelse. Udfør en deskriptiv analyse af data, som indgår i model (1). Opdel personerne efter deres officielle folkepensionsalder (*FP*) og rapporter gennemsnit og standardafvigelse for de relevante variable for de to grupper.
- Udfør estimation af (1) ved OLS. Rapporter parameterestimer og relevante standardfejl i en tabel. Fortolk estimationsresultaterne. Redegør for, om OLS-estimatoren er konsistent.

Opgave 2 (20%)

- Modellen estimeres nu ved en IV estimation, hvor dummien for den officielle folkepensionsalder *FP* anvendes som instrument for *pension66*. Diskuter og undersøg om *FP* kan anvendes som instrument. Udfør estimation af (1) ved IV estimation. Rapporter estimationresultaterne. Redegør for, under hvilke antagelser, at IV-estimatoren er konsistent.
- Udfør et test for om *pension66* er en eksogen variabel. Angiv konklusionen og skriv hvad det betyder. Angiv estimatet for β_1 og β_3 fra den foretrukne estimation, og skriv hvordan resultaterne skal fortolkes.
- Lav et histogram over prædikterede sandsynligheder for tidlig død for hhv. mænd og kvinder. Vurder på baggrund af histogrammet om modellen er vel-specificeret.

Opgave 3 (20%)

I denne opgave ser vi på et andet mål for helbred. Her benytter vi antallet af lægebesøg som 70 årig. Vi benytter følgende model:

$$\begin{aligned} \text{laege70} = & \gamma_0 + \gamma_1 \text{pension66} + \gamma_2 \log(\text{indkomst}) + \gamma_3 \text{kvinde} + \\ & \gamma_4 \text{faglaert} + \gamma_5 \text{kortudd} + \gamma_6 \text{mludd} + v, \end{aligned} \quad (2)$$

hvor *v* er et fejleddet.

³Hint: I STATA kan dette gøres ved kommandoen: `gen FP=((fodaar==1940)|(fodaar==1939 & fodmaaned>6))`

- Estimer model (2) ved en IV estimation hvor dummi'en for den officielle folkepensionsalder FP anvendes som instrument for $pension66$. Rapporter og fortolk estimationsresultaterne. Diskuter hvorfor der indgår færre observationer i denne regression end i regressionen for model (1) og overvej, om det kan være et problem.
- Udfør et test for om uddannelse har betydning for antallet af lægebesøg. Opskriv hypotesen, teststørrelse og skriv konklusionen på testet.
- Vi ønsker nu at undersøge, om pensionering har en forskellig effekt på helbredet for mænd og kvinder. Derfor udvides modellen således:

$$laege70 = \gamma_0 + \gamma_1 pension66 + \gamma_2 \log(indkomst) + \gamma_3 kvinde + \gamma_4 faglært + \gamma_5 kortudd + \gamma_6 mludd + \gamma_7 pension66 \cdot kvinde + v. \quad (3)$$

Estimer model (3) konsistent og test om pensionering har samme effekt på mænds og kvinders helbred. Angiv hypoteser, teststørrelse og skriv hvad konklusionen på testet er. Angiv på baggrund af estimaterne, hvad man kan konkludere vedr. effekten af pensionering på antallet af lægebesøg.

Opgave 4 (20%)

Betragt følgende regressionsmodel:

$$y_i = \beta_0 + \beta_1 x_i + u_i,$$

hvor $i = 1, \dots, n$. Vi antager, at MLR.1-MLR.3 er opfyldt. Desuden antager vi, at $E(u) = 0$ og $E(xu) = 0.10$.

- Antag at x er en dummyvariabel, hvor der gælder, at $E(x) = P(x = 1) = 0.15$. Angiv variansen af x og udregn den asymptotiske bias af OLS estimatoren for β_1 .
- Vis, at OLS estimatoren $\hat{\beta}_1$ for β_1 , når x er en dummyvariabel kan skrives som

$$\hat{\beta}_1 = \bar{y}^{x=1} - \bar{y}^{x=0},$$

hvor $\bar{y}^{x=1}$ er gennemsnittet af y når $x = 1$: $\bar{y}^{x=1} = \frac{\sum_{i=1}^n y_i x_i}{\sum_{i=1}^n x_i}$, $\bar{y}^{x=0} = \frac{\sum_{i=1}^n y_i (1 - x_i)}{\sum_{i=1}^n (1 - x_i)}$.

Hint:

$$\sum_{i=1}^n y_i x_i = n \cdot \bar{x} \cdot \bar{y}^{x=1}$$

$$\bar{y} = \bar{x} \cdot \bar{y}^{x=1} + (1 - \bar{x}) \cdot \bar{y}^{x=0}.$$

Vi antager, at vi har en dummyvariabel d_i . Om denne variabel gælder der, at

$$\begin{aligned} E(d \cdot u) &= 0 \\ \bar{x}^{d=1} &\neq \bar{x}^{d=0}, \end{aligned}$$

$$\text{hvor } \bar{x}^{d=1} = \frac{\sum_{i=1}^n x_i d_i}{\sum_{i=1}^n d_i}, \bar{x}^{d=0} = \frac{\sum_{i=1}^n x_i (1-d_i)}{\sum_{i=1}^n (1-d_i)}.$$

c. Vis, at estimator $\tilde{\beta}_1$ er en konsistent estimator for β_1

$$\tilde{\beta}_1 = \frac{\bar{y}^{d=1} - \bar{y}^{d=0}}{\bar{x}^{d=1} - \bar{x}^{d=0}}$$

hvor

$$\bar{y}^{d=1} = \frac{\sum_{i=1}^n y_i d_i}{\sum_{i=1}^n d_i}, \bar{y}^{d=0} = \frac{\sum_{i=1}^n y_i (1-d_i)}{\sum_{i=1}^n (1-d_i)}$$

Opgave 5 (20%)

Denne opgave går ud på at illustrere egenskaberne i en lineær sandsynlighedsmodel. Vi generer data fra følgende model:

$$y_i^* = \gamma_0 + \gamma_1 x_i + u_i \quad (4)$$

$$y_i = \begin{cases} 1 & \text{hvis } y_i^* > 0 \\ 0 & \text{ellers} \end{cases}$$

$$\gamma_0 = 3, \gamma_1 = -2, x \sim iiN(2, 1), u \sim iiN(0, 1)$$

Vi antager nu, at y^* er uobserveret, men y og x er observerede. Vi estimerer følgende regressionsmodel

$$y_i = \beta_0 + \beta_1 x_i + v_i$$

- Undersøg ved et simulationseksperiment egenskaberne for OLS estimatoren $\hat{\beta}_1$. Vælg et seed nummer efter eget valg og angiv det i besvarelsen. Lav eksperimentet med en stikprøve på 500 observationer og brug 1000 replikationer. Rapportér deskriptiv statistik og et histogram til at dokumentere simulationseksperimentet. Skriv hvad eksperimentet viser.
- Beregn i simulationsstudiet både de robuste og ikke robuste standardfejlen for OLS estimatoren af β_1 . Rapportér de gennemsnitlige standardfejl. Sammenlign standardfejlene med standardafvigelsen af den simulerede OLS estimatorer $\hat{\beta}_1$. Beskriv konsekvenserne i denne model, hvis man anvender de ikke robuste standardfejl.
- Den teoretiske sandsynlighed for at $y = 1$, når $x = k$ kan beregnes som

$$P(y = 1|x = k) = P(\gamma_0 + \gamma_1 k + u > 0) = 1 - \Phi(-\gamma_0 - \gamma_1 k),$$

hvor Φ er den kummulative fordelingsfunktion for $N(0, 1)$. Beregn ændringen i den teoretiske sandsynlighed når x går fra 0 til 1: $P(y = 1|x = 1) - P(y = 1|x = 0)$ og sammenlign med den gennemsnitlige værdi for OLS estimatoren for β_1 , som er fundet i simulationseksperimentet.