

Eksamen på Økonomistudiet sommer 2019

Økonometri I

Tag-hjem eksamen

13. juni 2019

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 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 2.5cm.
- 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 13. maj. 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 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.
- Rasmus Jørgensen kan kontaktes på telefon 3532 3075 klokken 10-12 den 13. juni såfremt der er problemer med at downloade data eller eksamensopgaven er fejlbehæftet. Eventuelle 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 blive behandlet herefter.
- Dette eksamenssæt består af 7 sider i alt.

Pas på at du ikke begår eksamenssnyd!

Det er eksamenssnyd, hvis du:

- Kopierer andres tekster uden at sætte citationstegn eller kildehenviser, så det ser ud som om det er din egen tekst.
- Bruger andres idéer eller tanker uden at kildehenviser, 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 citationstegn eller kildehenviser (selvplagiering).
- Modtager hjælp fra andre i strid med de regler, som er fastsat i rammestudieordningens afsnit 4.12 om samarbejde/sparring.

Du kan læse mere om reglerne for eksamenssnyd på Din Uddannelsesside og i Rammestudieordningens afsnit 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:

”Øger international handel landes levestandard?”

Siden Adam Smith og David Ricardo har økonomer argumenteret for, at international handel øger landes levestandard. Denne sammenhæng er som regel baseret på argumenter om specialisering og stordriftsfordele. International handel gør det således muligt for lande kun at producere de varer og services som de er relativt bedst til, for derefter at bytte dem med andre landes varer og services igennem import og eksport. International handel øger dermed velstanden på tværs af lande, fordi landenes ressourcer udnyttes bedre igennem specialisering og stordriftsfordele. Det er også velkendt, at disse velfærdsgevinster ikke deles ligeligt blandt befolkningen i et land. Således vil international handel øge velstanden for nogle individer, mens andre vil opleve en velstandsnedgang.

Der er adskillige økonometriske udfordringer forbundet med en empirisk analyse af sammenhængen mellem international handel og landes levestandard. For det første findes der kun få historiske eksempler på lande, som har oplevet en pludselig transition fra lukket til åben økonomi —og det er næppe tilfældigt hvilke lande der har gennemgået en sådan transition. En simpel sammenligning af landes levestandard før og efter overgangen til åben økonomi er derfor ikke mulig.

Empiriske studier af sammenhængen mellem international handel og landes levestandard tager derfor ofte udgangspunkt i sammenligninger på tværs af lande. I denne type analyser undersøges det, om BNP per capita i gennemsnit er højere for lande med et højt forhold af import og eksport til BNP. Den primære økonometriske udfordring ved denne metode er, at sammenhængen mellem landes levestandard og international handel kan skyldes andre uobserverbare, lande-specifikke faktorer.

Frankel og Romer (1999) argumenterer for, at et lands geografiske karakteristika er afgørende for hvor meget landet handler med resten af verden.¹ Derfor forventes et relativt isoleret land som New Zealand at have mindre international handel end et land som Belgien, der er tæt på mange mulige handelspartnere. Denne eksamensopgave har til formål at undersøge sammenhængen mellem levestandard, international handel og geografiske forhold.

Adgang til data

Til eksamen i Økonometri I er der adgang til ti datasæt på Digital Eksamen. Følg disse instruktioner til at vælge 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 svarer til gruppenummet bestemt i trin 1.

Eksempel: En gruppe bestående af eksamensnumre 75, 82, 174 har gruppenummer 5 og downloader derfor ‘groupdata5.dta’ fra Digital Eksamen.

3. Åbn datafilen i STATA og verificer at data kan indlæses uden fejl.

¹J.A. Frankel and D. Romer (1999): “Does Trade Cause Growth?”, *American Economic Review*, vol. 89, no. 3, pp. 379-399.

Dokumentation af data

Datasættet indeholder observationer for 125 lande i 2005. Følgende variable er tilgængelige:

Tabel 1: Variable i ‘groupdataX.dta’

Navn	Beskrivelse
land	Landenavn
iso	Landekode
Y	BNP per capita i 2005-USD
T	Handelsindeks = (import + eksport)/BNP
P	Befolkning (i mio)
A	Areal (i km ²)
pTi	Prædiktioner af T (defineret i ligning (3))
pTa	Prædiktioner af T (defineret i ligning (4))
N	Antal af handelspartnere

Note: ‘groupdataX.dta’ kan ikke bruges til andre formål end at besvare denne eksamensopgave.

Eksamensopgave

Opgave 1 (10%)

1. Udfør en deskriptiv analyse af data.
2. Rapportér to figurer som grafisk undersøger sammenhængen mellem: *i)* N og T , og *ii)* N og $\log Y$. Kommentér kort på graferne.

Opgave 2 (30%)

Eksamensopgaven tager udgangspunkt i følgende regressionsmodel:

$$\log Y = \beta_0 + \beta_1 T + \beta_2 \log P + \beta_3 \log A + u \quad (1)$$

hvor \log er den naturlige logaritme og u er et fejledd.

1. Udfør estimation af (1) ved OLS. Rapportér parameterestimer og relevante standardfejl i en tabel. Fortolk estimationsresultaterne. Redegør for om OLS-estimatoren er konsistent.
2. Undersøg om fejleddet i (1) har konstant varians. Beregn Breusch-Pagan testet for heteroskedasticitet. Redegør for konklusionen.
3. Undersøg om den funktionelle form i (1) er korrekt specificeret. Beregn RESET. Redegør for konklusionen.
4. Opstil en nulhypotese om, at handelsindekset *ikke* har nogen betydning for \log BNP per capita over for en alternativhypotese, at handelsindekset har en *positiv* effekt på \log BNP per capita. Beregn den relevante teststørrelse og p-værdi. Redegør for konklusionen på testet.

Opgave 3 (20%)

Ifølge Frankel og Romer (1999) har et lands geografiske karakteristika stor betydning for hvor meget international handel landet har med resten af verden. De opstiller derfor følgende lineære regressionsmodel til at undersøge sammenhængen mellem *bilateral* handel og geografiske forhold for afsender- og modtagerlandet:

$$\log\left(\frac{M_{ij} + X_{ij}}{BNP_i}\right) = f(\text{afstand}_{ij}, \text{grænse}_{ij}, \text{hav}_i, \text{hav}_j, P_i, P_j, A_i, A_j) \quad (2)$$

hvor M_{ij} er land i 's import fra land j , mens X_{ij} er land i 's eksport til land j . Den afhængige variabel i model (2) angiver land i 's (log-transformerede) bilaterale handel med land j relativt til land i 's BNP. f er en lineær funktion, som ikke er beskrevet nærmere i denne eksamensopgave, udover at f er en funktion af:

- i. Den fysiske *afstand* mellem land i og j
- ii. En dummy for om i og j deler en fysisk landegrænse
- iii. To dummies for om i og j har havadgang til søtransport
- iv. Landenes befolkningsstørrelse (P_i, P_j)
- v. Landenes samlede areal (A_i, A_j)

OLS-estimation af model (2) beregner en forklaringsgrad på 32.3 pct ($R^2 = 0.323$) med data fra 2005. På baggrund af OLS-estimerne prædikteres den afhængige variabel som vi betegner som $\widehat{\log T_{ij}}$. Variablene pTi og pTa konstrueres derefter som:

$$pTi_i = \sum_{j=1}^{N_i} \exp(\widehat{\log T_{ij}}) \quad (3)$$

$$pTa_i = \sum_{j=1}^{202} \exp(\widehat{\log T_{ij}}) \quad (4)$$

hvor $\widehat{\log T_{ij}}$ er land i 's prædikterede (log-transformerede) bilaterale handel med land j relativt til land i 's BNP fra en OLS-estimation af model (2). En vigtig forskel på ligning (3) og (4) er antallet af lande der indgår i summen. I ligning (3) summeres der over N_i , som angiver hvor mange handelspartnere land i har. Variablen pTi er derfor en sum af (exp-transformerede) prædiktioner, for de landepar som har en observeret samhandel. Sagt med andre ord er pTi konstrueret på baggrund af *in-sample* prædiktioner fra model (2). OLS-estimationen af model (2) tager udgangspunkt i data for 203 lande. Hvis alle lande handler med hinanden, bør data indeholde $203 \times 202 = 41,006$ observationer af den afhængige variabel i model (2). Der indgår dog kun 25,432 observationer af den afhængige variabel i datasættet. Ergo er der 15,574 observationer, som ikke optræder i datasættet, fordi de pågældende lande ikke handler med hinanden. Model (2) kan dog bruges til beregne *out-of-sample* prædiktioner for de lande, der ikke handler med hinanden, ud fra deres geografiske karakteristika. Variablen pTa er konstrueret på baggrund af *in-sample* og *out-of-sample* prædiktioner fra model (2) for alle tænkelige handelspartnere.

1. Diskutér om pTi og pTa kan bruges som instrument for T i model (1). Rapporter og fortolk IV-resultaterne. Redegør for om IV-estimatoren er konsistent.
2. Undersøg om T er en endogen variabel i model (1).

Opgave 4 (20%)

Betragt følgende regressionsmodel:

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u \quad (5)$$

Antag at model (5) er korrekt specificeret og opfylder MLR.1-MLR.4. Antag derudover at x_1, x_2, x_3 er indbyrdes relateret som beskrevet i følgende kovarians-matrix:

$$Var(X) = \begin{bmatrix} var(x_1) & cov(x_1, x_2) & cov(x_1, x_3) \\ cov(x_2, x_1) & var(x_2) & cov(x_2, x_3) \\ cov(x_3, x_1) & cov(x_3, x_2) & var(x_3) \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{bmatrix} \quad (6)$$

1. Betragt først en regression af y på x_1 . Udled den asymptotiske bias for OLS-estimatoren, $\hat{\beta}_{x_1}^{M1}$, for denne regression som vi kalder model 1 (M1).
2. Betragt nu en regression af y på x_1 og x_2 . Udled den asymptotiske bias for OLS-estimatoren, $\hat{\beta}_{x_1}^{M2}$, for denne regression som vi kalder model 2 (M2).
3. x_3 er en udeladt variabel i model 1 og 2. Dette medfører, at $\hat{\beta}_{x_1}^{M1}$ og $\hat{\beta}_{x_1}^{M2}$ lider af "udeladt variabel bias". Diskutér og begrund hvorvidt følgende udsagn er korrekte:
 - i. Udeladt variabel-bias er mindre, når x_2 medtages i modellen udover x_1 .
 - ii. Udeladt variabel-bias kan fortegnbestemmes ud fra β_3 og σ_{13} .

Opgave 5 (20%)

Betragt følgende datagenererende proces (DGP):

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u \quad (7)$$

$$\beta = (\beta_0, \beta_1, \beta_2, \beta_3)' = (1, 0.5, -0.75, 1.25)' \quad (8)$$

$$u \sim N(0, 1) \quad (9)$$

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \sim N(\mu, \Sigma) \quad (10)$$

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix} \quad (11)$$

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{pmatrix} = \begin{pmatrix} 3 & 1.25 & 0.75 \\ 1.25 & 1 & 0.6 \\ 0.75 & 0.6 & 1 \end{pmatrix} \quad (12)$$

Udfør et simulationseksperiment af OLS-estimatoren på baggrund af ovenstående datagenererende proces. Vælg et *seed* nummer efter eget valg og angiv det i besvarelsen. Træk stikprøver med $N = 200$ observationer og gentag eksperimentet 1000 gange. Beregn følgende størrelser (for hver stikprøve):

- i. OLS-estimatet, $\hat{\beta}_{x_1}^{M1}$, for en regression af y på x_1 (model 1).
- ii. OLS-estimatet, $\hat{\beta}_{x_1}^{M2}$, for en regression af y på x_1 og x_2 (model 2).

Udfør simulationseksperimentet for hver af følgende definitioner af kovarians-matricen, Σ :

1. Σ er lig udtrykket i (12).
2. Σ er lig udtrykket i (12), bortset fra at $\sigma_{13} = 0$.
3. Σ er lig udtrykket i (12), bortset fra at $\sigma_{23} = 0$.
4. Σ er lig udtrykket i (12), bortset fra at $\sigma_{12} = 0$.

Udfør en deskriptiv analyse af de simulerede resultater. Redegør for og kommentér på OLS-estimatorens statistiske egenskaber på baggrund af simulationsstudiet. Hvordan påvirker udeladte variable OLS-estimatorens egenskaber for en multipel lineær regressionsmodel? Hvilke kilder til bias er der? Kan biasen fortegnbestemmes ud fra β_3 og σ_{13} ?

Hint: STATA's *drawnorm*-kommando kan med fordel bruges til at trække tilfældige tal fra en multivariat normalfordeling. Nedenstående program trækker 200 tilfældige tal fra normalfordelingen angivet i (10)-(12):

```
program simdata, rclass
    drop _all
    matrix my=(2,2,2)
    matrix cov=(3,1.25,0.75\1.25,1,0.6\0.75,0.6,1)
    drawnorm x1 x2 x3, n(200) means(my) cov(cov)
end
```