

## SØK 3001 V 17. Sensorveiledning

### Oppgave 1.

a) Eksogenitet/endogenitet i økonometrisk forstand er diskutert i Woolridge kap. 3.3.

La den økonometriske modellen være

$$(1) y = \alpha + \beta x + u$$

Kort fortalt innebærer eksogenitet en forutsetning om at  $E(u|x) = 0$ , som impliserer at  $\text{cov}(x, u) = 0$  og  $E(u) = 0$ . Endogenitet innebærer at  $E(u|x) \neq 0$  eller alternativt  $\text{cov}(x, u) \neq 0$ .

Årsaker: Utelatte variable, målefeil i  $x$ , eller at  $x$  og  $y$  er simultant bestemt. Litt forklaring av hver av disse årsakene er ønskelig, men det er ikke nødvendig med fullstendige og omstendelige utledninger.

b) Bør starte med å presisere konsistensbegrepet, dvs forklare hva som menes med en konsistent estimator for  $\beta$  i ligninga (se kap. 5 i Woolridge). Konsistens innebærer at

$\text{plim} \hat{\beta} = \beta$  der  $\hat{\beta}$  er en estimator for den ukjente parameteren  $\beta$ . Må få fram at OLS-estimatoren i kolonne (1) er konsistent estimator dersom  $x$  er eksogen, dvs.  $\text{plim} \hat{\beta}^{\text{OLS}} = \beta$  dersom  $\text{cov}(x, u) = 0$ .

IV-estimatoren i kolonne (2) er konsistent dvs  $\text{plim} \hat{\beta}^{\text{IV/2SLS}} = \beta$  dersom  $z_1$  og  $z_2$  er valide instrumentvariable for  $x$ , (se kap. 15 i Woolridge), formelt

$$\text{cov}(z_1, u) = 0 \text{ og } \text{cov}(z_2, u) = 0 \text{ (eksklusjonskriteriet)}$$

$$\text{cov}(x, z_1) \neq 0, \text{ cov}(x, z_2) \neq 0 \text{ (Relevanskriteriet)}$$

c) Estimatorene i kolonne (2) er framkommet ved 2SLS, der 1. steget er en OLS på regresjonen

$$(2) x = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + v$$

Estimerte parametre i denne brukes til å lage predikert  $x$ ,  $\hat{x}$ .  $\hat{x}$  kan da benyttes som instrumentvariabel for  $x$  i strukturligninga. Metoden kan implementeres med å erstatte  $x$  med  $\hat{x}$ , i strukturligninga (1) og estimere den da framkomne ligning med OLS, i steg 2, se kap. 15 i Woolridge.

Relevansen av  $z_1$  og  $z_2$  som instrumenter innebærer å teste om  $z_1$  og  $z_2$  bidrar signifikant til å forklare variasjonen i  $x$  representert ved 1.steags-ligninga (2). Test

$$H_0: \pi_1 = \pi_2 = 0,$$

Kan bruke R-squared varianten av F-test av om alle forklaringsvariablene i en OLS-regresjon bidrar til å forklare  $x$  (se Woolridge kap. 4-5).

$$F = \frac{R^2/k}{(1-R^2)/(n-k-1)}$$

Her er  $k=2$ ,  $n=100$  og  $R^2=0.333$  i kolonne (2) i tabell 1, hvilket gir  $F=24.2$  og nullhypotesen forkastes klart, hvilket indikerer at instrumentene er relevante. F-verdien er også klart over tommelfingerregelen på  $F=10$ .

d) Endogenitetstest er forklart i Woolridge kap. 15 s. 481, og implementeres ved å estimere OLS-regresjonen som er rapportert i kolonne (4) der strukturrelasjonen er utvidet med residualen,  $e_3$  fra 1.steget i 2SLS regresjonen (kolonne (3)). Intuisjonen er at dersom  $x$  er eksogen og  $z_1$  og  $z_2$  er valide instrumenter, skal residualen  $e_3$  som representerer den variasjonen i  $x$  som ikke skyldes  $z_1$  og  $z_2$  ikke bidra signifikant i forklaringen av  $y$ . Eksogen  $x$  innebærer dermed en hypotese om at koeffisienten foran  $e_3$  i kolonne (4) er lik null.  $t$ -observator på  $0.999/0.1 = 6.1$  innebærer at nullhypotesen om at  $x$  er eksogen forkastes.

e) Overidentifikasjonstest er behandlet i Woolridge kap. 15 s. 482. Identifikasjon i en ligning med en endogen høyresidevariabel innebærer at vi har minst én eksogen variabel (ett instrument) som er korrelert med  $x$ , men ukorrelert med restleddet i strukturligningen (1). Overidentifikasjon innebærer at vi har mer enn ett valid instrument. Intuisjonen er at dersom  $z_1$  og  $z_2$  begge er valide instrumenter vil IV-estimatorer basert på h.h.vis  $z_1$  og  $z_2$  som instrument, begge være konsistente, og dermed ikke være systematisk forskjellige. Som forklart i Woolridge s. 483-484 kan overidentifikasjonstesten implementeres ved å estimere en hjelperegresjon med residualen fra 2SLS regresjonen,  $e_2$ , som avhengig variabel med  $z_1$  og  $z_2$  som regressorer og danne testobservatoren  $n \cdot R^2$ , der  $R^2$  er multiple determinasjonskoeffisienten fra hjelperegresjonen i kolonne (5). Under  $H_0$  om overidentifikasjon er observatoren kjikvadrat-fordelt med antall frihetsgrader lik antall overidentifiserende restriksjoner som her er 1. Her er testobservatoren lik  $0.363 \cdot 100 = 3.63$  som er marginalt lavere enn kritisk verdi ved 5% signifikansnivå (3.84). Vi kan derfor ikke forkaste hypotesen om overidentifikasjon.

## Oppgave 2.

Her kan det være til nyttig å formulere de tre modellvariantene eksplisitt først

$$\text{Kol (1) } la = \alpha + \beta_1 lp + u$$

$$\text{Kol (3) } la = \alpha + \beta_0 lp + \beta_1 lp_{-1} + \gamma la_{-1} + u$$

$$\text{Kol (5) } la - la_{-1} = \alpha + (\beta_0 + \beta_1)lp + (\gamma - 1)la_{-1} + u$$

a) (1) er en statisk modell og kan tolkes som at prisen på ris slår momentant ut i arealet tilplantet med ris. Her kan det ikke skilles mellom kort og langtidseffekt. Priselasitet er her 0.776. (3) er en dynamisk modellvariant der langtidslivekten defineres ved en situasjon der alle variable er i ro. Langtids priselasitet blir her  $(\beta_0 + \beta_1)/(1 - \gamma)$  og estimert elasitet blir følgelig  $(0.7766 - 0.6109)/(1 - 0.4043) = 0.275$ . Korttids priselasitet er  $\beta_0$ , estimert til 0.77. Studentene bør være i stand til å se at modellen estimert i kolonne (5) er en ren transformasjon av modellen i kolonne (3) og at de underliggende koeffisientene her er de samme og langtids og korttidselasitetene dermed også er de samme. Modellformuleringen i (5) er nyttig når det skal testes hypoteser i neste delspørsmål

b) Her er det enkelt å benytte modellen estimert i kolonne (5). Siden langtids priselasitet er  $(\beta_0 + \beta_1)/(1 - \gamma)$ , vil en enkel test av hypotesen om at den er null være en test av om koeffisienten foran  $lp$  i kolonne (5) er lik 0.  $t$ -observatoren her er  $0.1658/0.3844 = 0.43$ . Siden den er langt under kritisk verdi i  $t$ -fordelinga kan vi klart ikke forkaste hypotesen om langtids priselasitet lik 0.

c) Seriekorrelasjon er behandlet i læreboka kap. 12.

Forventer at studentene definerer presist hva som menes med seriekorrelasjon i restleddet i en økonometrisk modell som i kap. 12.1.

Kandidatene bør få fram at foruten i en standard-situasjon med strikt eksogene forklaringsvariable der strukturligninga er korrekt spesifisert, men restleddet er korrelert mellom observasjonsperioder, kan restleddet også framstå som seriekorrelert i en situasjon der dynamikken er feilspesifisert, f.eks dersom det estimeres en statisk modell, mens den «sanne» modellen er dynamisk.

**Konsekvenser**, nyttig å skille mellom to situasjoner:

A. Strukturlikninga korrekt spesifisert og strikt eksogene forklaringsvariable:

i) OLS-estimatoren forventningsrett og konsistent, men ikke effisient (eksisterer lineær forventningsrett estimator med lavere varians, GLS-estimator)

ii) Standard variansformel under Gauss Markov forutsetningene gjelder ikke, dvs. standard estimert standardavvik feilestimerer sanne standardavvik, som gjør at t og F tester ikke lenger blir valide

B. Seriekorrelasjon pga utelatt variabel, f.eks feilspesifisert dynamikk:

Kan enkelt illustrere det med modell med og uten lagga endogen variabel.

La «sann» modell være:  $la = \alpha + \beta_0 lp + \gamma la_{-1} + u$  der u er hvit støy, mens vi estimerer den statiske modellen  $la = \alpha + \beta_0 lp + v$

Her vil  $v = \alpha + \beta_0 lp + \gamma la_{-1} + u$

Dermed vil  $v_{-1} = \alpha + \beta_0 lp_{-1} + \gamma la_{-2} + u_{-1}$  og restleddet på tidspunkt t vil åpenbart være korrelert med restleddet på tidspunkt t-1 i den feilspesifisert modellen vi estimerer i tillegg til at vi normalt vil ha utelatt variabel-skjevhet i dette tilfellet. Her vil altså seriekorrelasjonen i restleddet være en indikasjon på at modellen er feilspesifisert.

d) Test av seriekorrelasjon.

Behandlet i kap 12-2 i Woolridge.

Baseres på hjelperegresjonene i kolonnene (2) og (4).

Test av nullhypotesen om fravær av 1. ordens seriekorrelasjon i modellene i kolonne (1) og (3) gjennomføres ved å teste med t-test om koeffisienten foran lagga residualer,  $\hat{e}$  og  $\hat{e}D$  er lik 0 i henholdsvis kolonne (2) og (4). t-verdier er henholdsvis  $0.3986/0.1673=2.38$  og  $0.0656/0.1868=0.35$ . Konklusjon: Basert på t-testen forkastes hypotesen om fravær av 1.ordens seriekorrelasjon i statisk modell i kolonne (1). Tilsvarende hypetese forkastes ikke i dynamisk modell i kolonne (3). Rimelig konklusjon er at statiske modellen er feilspesifisert, jfr svaret på c) ovenfor. Enkle t-tester på koeffisientene foran  $la_{-1}$  og  $lp_{-1}$  viser også at vi må forkaste hypoteser om at disse koeffisientene er null.

### Oppgave 3.

a) loglinearisert stokastisk versjon av Cobb Douglas blir  $ly = \beta_0 + \beta_1 \ln + \beta_2 lk + u$

der  $\beta_0$  representerer teknikkfaktoren,  $\beta_1$  representerer grense-elasticiteten av arbeidskraft og  $\beta_2$  representer grense-elasticiteten av kapital og  $u$  er et stokastisk restledd.

i) Kolonne (3)-(6) som alle inkluderer land-dummier vil ivareta permanente forskjeller i produksjonsnivå mellom land,

ii) kolonne (5) og (6) som har *både* land-dummier og årsdummier ivaretar både permanente forskjeller i produksjonsnivå mellom land og virkningen av felles teknologisk endring

b) Restriksjoner:

**Kolonne 1 og 2:** Alle koeffisientene er like på tvers av land (dvs. at konstantledd og effektene av realkapital og sysselsetting ikke varierer mellom land)

koeffisientene er like i ulike tidsperioder (dvs. at konstantledd og effektene av realkapital og sysselsetting ikke varierer mellom tidsperioder)

**Kolonne 3 og 4** Alle koeffisientene unntatt konstantledd er like på tvers av land.

Konstantledd varierer mellom land (Landspesifikke konstantledd representert ved landdummiene)

**Kolonne 5 og 6** Alle koeffisientene med unntak av konstantledd er like på tvers av land og tidsperiode. Konstantleddet varierer både mellom land og over tid (Landspesifikke og årsspesifikke konstantledd representert ved land-dummiene og årsdummiene).

c) Her er det viktig å innse at (2) er en ren transformasjon av (1), (4) er ren transformasjon av (3), mens (6) er ren transformasjon av (5). Dette er nyttig innsikt for å foreta testing av konstant skala-avkastning i spørsmål d). Koeffisienten foran  $\ln$  i kolonnene (2), (4) og (6) er lik  $1 - (\beta_1 + \beta_2)$ , mens koeffisienten foran  $lk - \ln$  i (2), (4) og (6) er grense-elasticiteten av kapital ( $\beta_2$ ). De estimerte grense-elasticitetene av sysselsetting og kapital er således 0.1373 og 0.8624 i kolonne (1) og (2), 0.2908 og 0.5862 i kolonne (3) og (4) og 0.1602 og 0.5148 i kolonne (5) og (6).

95% konfidensintervall beregnes rett fram som tilnærma lik

koeffisientestimert  $\pm 2se$ ,

der  $se$  = estimert standardavvik til koeffisienten.

Følgende konfidensintervaller finnes (tilnærma)

Grense- elastisitet	Kolonne 1	Kolonne 3	Kolonne (5)
Sysselsetting	(0.1237,0.1509)	(0.2354,0.3462)	(0.0948, 0.2256)
kapital	(0.8524,0.8724)	(0.565,0.6074)	(0.4902,0.5394)

d) Konstant skala-avkastning innebærer hypotesen  $H_0: (\beta_1 + \beta_2) = 1$ , Test av hypotesen gjennomføres ved rett fram tosidig t-test av om koeffisienten  $1 - (\beta_1 + \beta_2)$  foran ln i henholdsvis kolonne (2), (4) og (6) er lik null. Tolking av hypotesen: Konstant skalaavkastning innebærer at p prosent økning i begge faktorene gir p prosent økning i produksjonen. t-verdiene er henholdsvis 0.055, 6.32 og 10.3. Konklusjon er at hypotesen om konstant skalavkastning ikke forkastes i modellen i kolonne (1) og (2), mens den forkastes i de øvrige modellene. Gode kandidater bør knytte dette til hva som skiller modellene, dvs hva slags utelatte variable som er ivaretatt, jfr. svaret på spørsmål a). Ut fra dette synes resultatene fra kolonne (5) og (6) mest troverdige siden de ivaretar både effekten av permanente forskjeller mellom land så vel som innflytelsen fra felles teknologiske fremskritt.

e) Her er svaret rett fram at siden DEM slik den er definert, ikke variere over tid innen land, vil den være perfekt korrelert med land-dummiene og vil derfor ikke være mulig å identifisere effekten av den i kolonne (2) og (3).

f) Rett fram å utvide modellene i (1), (3) og (5) med en interaksjonsvariabel  $lk*DEM$ . Siden dette vil være en variabel som har variasjon både mellom land og over tid, vil det være mulig å identifisere effekten av den også i de versjonene som har landspesifikke effekter. Dersom koeffisienten foran  $lk*DEM$  er positiv (negativ) indikerer det at grenselastisiten av realkapital er høyere (lavere) i et demokratisk land enn i et ikke-demokratisk land. Helt tilsvarende i de transformerte modellene.

g) Det er mulig å inkludere en slik variabel (WP) i kolonne (1)-(4). Gode kandidater vil innse at konsistent anslag av WP i disse variantene krever at variabelen ikke fanger opp effekten av andre utelatte variable som bare varierer over tid. WP vil derimot være perfekt korrelert med årsummiene som er inkludert i kolonne (5) og (6) og kan ikke inkluderes i disse to variantene.