

Program for øvelserne:

Opsamling fra ugeseddel 11
Gruppearbejde
Plenumdiskussion
SAS-øvelser

Øvelsesopgave: Målefejl og instrument variabel estimation

Formålet med ugeseddel 12 er at se nærmere på et tilfælde, hvor den forklarende variabel er endogen som følge af målefejl. Generelt vil en forklarende variabel være endogen, hvis den er korreleret med fejleddet. Årsagen kan være målefejl, en udeladt forklarende variabel, forkert funktionel form, o. lign. Korrelationen betyder, at OLS-estimatoren ikke er middelret og konsistent.

Opgaven består dels i at udregne effekten på OLS estimatoren analytisk og via simulation, dels i at korrigere for målefejlen via en instrumentvariabel (IV) estimator.

Udgangspunktet er vores økonometriske model fra de første ugesedler. Modellen beskrev sammenhængen mellem de samlede udgifter, $xtot_i$, og budgetandelen til mad, $\frac{xfath_i}{xtot_i}$, som:

$$\frac{xfath_i}{xtot_i} = \delta_0 + \delta_1 \ln\left(\frac{xtot_i}{price_i}\right) + u_i$$

I modellen kan det samlede forbrug, $xtot_i$, betragtes som en endogen variabel af flere grunde. Dels kan data indeholde målefejl. Dels indgår forbruget af mad som en del af det totale forbrug.

Problemet kan korrigeres via en IV-estimator baseret på et velegnet instrument. I opgaven benytter vi en variabel for personens indkomst efter skat, $nety$, som instrument for det samlede forbrug.

Gruppearbejde

Diskuter:

1. Hvordan påvirker målefejl i udgifterne OLS-estimerne?
2. Hvilke betingelser skal en variabel z opfylde for at være et gyldigt instrument for x ?
3. Er indkomsten efter skat, $nety$, et gyldigt instrument for det samlede forbrug, $xtot_i$?
4. Opskriv IV-estimatoren på matrixform.
5. Opskriv Z-matricen af instrumenter.

SAS-øvelser

SAS-øvelsen går ud på at estimere Engelkurven for andelen af mad for enlige mænd via IV estimatoren. Som instrument for den endogene variabel, $xtot_i$, bruges variabelen $nety_i$. IV estimatoren udregnes ved brug af PROC IML. Man kan med fordel bruge programmet fra ugeseddel 3.

Én gruppe bedes skrive en kort opsamling (1/2-1 side) af spørgsmål 5 og 6 og sende den til alice.klynge@econ.ku.dk senest mandag kl. 12.00.

1. Skab variabelen for instrumentet, $\log\left(\frac{nety_i}{price_i}\right)$.
2. Check, at instrumentet er korreleret med den endogene variabel, $\log\left(\frac{xtot_i}{price_i}\right)$.
3. Indlæs data for instrumentet i en vektor i PROC IML (ligesom data for y og x er indlæst i programmet på ugeseddel 3).
4. Dan Z-matricen (husk, at matricen skal indeholde et konstantled).
5. Udregn estimator for IV estimatoren
6. Sammenlign resultaterne fra IV estimatoren med dem fra OLS estimatoren.

Simulationsopgave:

Simulationsopgaven består i at sammenligne egenskaberne for OLS estimatoren og IV estimatoren, når den forklarende variabel indeholder målefejl.

Lad følgende datagenererende proces danne baggrund for eksperimentet:

$$y_i = \beta_0 + \beta_1 x_i^* + u_i, \quad u_i \sim N(0, \sigma^2) \quad (1)$$

hvor $\beta_0 = 1$, $\beta_1 = 4$ og $\sigma^2 = 1$. Antag, at den forklarende variabel, x_i^* , og fejleddet, u_i , er indbyrdes uafhængige og at x_i^* generes som uafhængige trækninger fra $N(2,4)$. Antallet af observationer er repræsenteret ved n og kan variere.

Antag, at den forklarende variabel, x_i^* , er uobserverbar, og at vi i stedet observerer $x_i = x_i^* + e_i$. Lad e_i være en klassisk målefejl: $E(e_i) = 0$, $V(e_i) = \sigma_e^2$ og $Cov(e_i, x_i^*) = 0$, hvor $\sigma_e^2 = 1$. Generer e_i som uafhængige trækninger fra $N(0,1)$.

2. Beregn OLS estimator for 10.000 replikationer af modellen

Brug SAS-koden fra Ugeseddel 5 som udgangspunkt for at konstruere en algoritme i IML, der kan udføre 10.000 replikationer af modellen og for hver replikation gemme OLS estimatorerne, $\hat{\beta}_0$ og $\hat{\beta}_1$. Husk at generere den afhængige variabel, y , på basis af den korrekte forklarende variabel, x^* , og at estimere modellen på basis af x (x^* plus målefejl). Udfør eksperimentet for:

- A. $n = 100$ observationer.

B. $n = 10$ observationer.

Tegn et histogram for Monte Carlo fordelingen af hvert sæt af estimater og beregn fordelingernes middelværdi og varians via Proc Univariate. Stemmer OLS estimaterne overens med de sande parametre?

2. Beregn IV estimater for 10.000 replikationer af modellen

Antag, vi har en variabel z givet ved $z = \gamma_0 + \gamma_1 x^* + w$. Lad fejleddet være fordelt ved: $w \sim N(0,1)$, og lad parametrene være givet ved $\gamma_0 = 3$ og $\gamma_1 = 1$. Under hvilke betingelser er variabelen z et gyldigt instrument for x ?

Beregn IV estimater for 10.000 replikationer af modellen for:

- A. $n = 100$ observationer.
- B. $n = 10$ observationer.

Tegn et histogram for Monte Carlo fordelingen af hvert sæt af estimater og beregn fordelingernes middelværdi og varians. Stemmer IV estimaterne overens med de sande parametre?

3. Sammenlign OLS og IV estimater

Hvilke ligheder og forskelle er der i fordelingerne for OLS og IV estimaterne for $n = 100$ og $n = 10$? Hvilke(n) estimator ser ud til at være konsistent?

4. Konklusion på simulationsstudiet

Hvad er dine konklusioner af simulationsstudiet for en model med klassisk målefejl?

Hjemmearbejde

Teoriopgaver:

Alternativ estimator i den simple regressionsmodel

Vi ønsker at estimere parametrene for en simpel regressionsmodel:

$$y = \beta_0 + \beta_1 x^* + u \quad (2)$$

hvor modellen opfylder antagelserne SLR 1-SLR 4. Lad den forklarende variabel, x^* , være uobserverbar, mens vi observerer x givet ved:

$$x = x^* + e$$

hvor e er en klassisk målefejl.

1. Hvad er betingelserne for, at e er en klassisk målefejl?
2. Bevis analytisk, at OLS estimatoren er inkonsistent.
3. Udregn den analytiske bias og sammenlign med resultaterne fra simulationsstudiet.