Økonometri 1
Efterår 2006

Ugeseddel 9

Program for øvelserne:
Opsamling på Ugeseddel 8
Gruppearbejde
SAS øvelser

Ugeseddel 9 består i at undersøge, om der er heteroskedasticitet i vores model for væksten og i så fald, hvordan vi korrigerer for det. Ved heteroskedasticitet er variansen på fejlleddet forskellig for forskellige observationer.

Til analysen benytter vi grafiske og formelle test og viser, hvordan en vægtet regressionsmetode estimerer modellens parametre efficient. Vi introducerer såkaldte robuste standardfejl, som giver konsistente estimater af OLS-estimatorens standardfejl uanset eksistensen af heteroskedasticitet.

Øvelsesopgave: Vækstregressioner (fortsat)

Ugeseddel 8 gav os følgende model for udviklingen i vækstraterne i 80 lande:

$$\begin{aligned} g_i = \beta_0 + \beta_1 \log y_{i0} + \beta_2(\log s_{K,i} - \log(n_i + 0.075)) + \beta_3(\log s_{H,i} - \log(n_i + 0.075)) \\ + \beta_4(Safrica_i \cdot \log y_{i0}) + \beta_5 Latinam + u_i \end{aligned} \tag{1.1}$$

hvor $Safrica_i$ og $Latinam_i$ er dummyvariable for lande i Afrika syd for Sahara og lande i Latinamerika.

De forklarende variabler (undtagen dummyerne) er transformeret via logaritmen. Log-transformationen bruges ofte mere *ad hoc* med det formål at fjerne heteroskedasticitet i fejlleddet.

Vi vil undersøge, om transformationen også har den effekt for vækstregressionen. Konkret vil vi gøre som Barro (1991), hvor initialindkomsten optræder uden log-transformation. Barro undersøger tilstedeværelsen af heteroskedasticitet og finder, at resultaterne af hans analyse ikke er følsomme. Vi vil undersøge, om vi får tilsvarende resultater.

Vi analyserer derfor en udgave af modellen, hvor initialindkomsten ikke transformeres:

$$\begin{aligned} g_i = \delta_0 + \delta_1 y_{i0} + \delta_2(\log s_{K,i} - \log(n_i + 0.075)) + \delta_3(\log s_{H,i} - \log(n_i + 0.075)) \\ + \delta_4(Safrica_i \cdot y_{i0}) + \delta_5 Latinam + v_i \end{aligned} \tag{1.2}$$

Bemærk at y_{i0} optræder to steder i modellen.

Gruppearbejde:

Diskuter:

1. Antag, at antagelse MLR.5 i Wooldridge *ikke* holder i model (1.2). Hvad betyder det for OLS-estimatorens egenskaber?

2. Antag, at initialindkomsten styrer variansen på fejlleddet og dermed er en relevant ”skalavariabel” for heteroskedasticiteten. Hvad kan begrunde det valg?

3. Skitser sammenhængen mellem fejlleddet v_i og initialindkomsten y_{i0}, hvis fejlleddets varians er ligefrem proportional med y_{i0}. Ligeledes hvis fejlleddets varians er omvendt proportional med y_{i0}. Hvordan ser en WLS estimation (Wooldridge p. 284-85) ud for hvert af de to tilfælde?

SAS-øvelser:

Én gruppe bedes skrive en kort opsamling (½-1 side) af spørgsmål h) og sende den til alice.klynge@econ.ku.dk senest mandag kl. 12.00.

Heteroskedasticitet i model (1.2)

a) Estimer model (1.2) og gem residualerne

Vores data er fortsat det udvidede PWT datasæt. Du kan med fordel tage udgangspunkt i SAS-programmerne fra tidligere ugesedler.

Udfør de nødvendige transformationer af data til at opstille model (1.2). Estimer modellens parametre ved OLS i Proc Reg. Gem residualerne, $\hat{v}_i$, sammen med de oprindelige variabler i et datasæt.

b) Grafisk test af heteroskedasticitet

Beregn de kvadrerede residualer, $\hat{v}_i^2$, fra (1.2). Lav plots af $\hat{v}_i$ og $\hat{v}_i^2$ over for de forudsagte værdier fra (1.2), $\hat{g}_i = g_i - \hat{v}_i$, og over for y_{i0}; i alt fire plots. Er der grafisk set tegn på heteroskedasticitet, og i så fald af hvilken form?

c) Breusch-Pagan test af heteroskedasticitet

Udfør Breusch-Pagan testet for heteroskedasticitet ved at regressere $\hat{v}_i^2$ mod variablene på højre side i model (1.2). Angiv nul- og alternativhypotesen og resultatet af testet i ord og matematisk. Beregn både *F*- og LM udgaven af testet. Er der tegn på heteroskedasticitet?

Hint:
- F-testet kommer ud automatisk som det samlede test for signifikans af hjælperegressionen.
- Til LM testet skal du bruge antallet af observationer og R^2 fra hjælperegressionen. De kritiske værdier for LM testet finder du i χ^2 fordelingen i Tabel G.4 i bogen.

Se også Wooldridge p. 280.

d) Breusch-Pagan test af heteroskedasticitet for en specifik variansstyrende variabel

Udfør et Breusch-Pagan test med udgangspunkt i, at initialindkomsten y_{i0} er den mest sandsynlige årsag til heteroskedasticitet. Udfør testet ved at regressere $\hat{v}_i^2$ på y_{i0} og en konstant (Woolridge p. 282). Angiv hypotesen og resultatet af testet i ord og matematisk. Hvad konkluderer du?

e) Sammenhæng mellem variansen på fejlleddet og initialindkomsten

Antag, at fejlledsvariansen i (1.2) opfylder følgende sammenhæng:

$$Var(v_i \mid x_i) = \sigma^2 / y_{i0} \qquad (1.3)$$

hvor x_i betegner sættet af variable på højresiden af (1.2) for den i'te observation. Svarer det til den type af tegn på heteroskedasticitet, du fandt under b) og d)?

f) Vægtet mindste kvadraters estimation (WLS) via Weight optionen

Udfør en vægtet mindste kvadraters estimation (WLS) af model (1.2) (Woolridge p. 285-86). Antag herunder, at variansen på fejlleddet opfylder (1.3).

Sammenlign estimaterne fra WLS estimationen med OLS estimaterne fra spørgsmål a). Hvilke(t) sæt af estimater er konsistente for δ_j parametrene? Hvilket sæt af estimater er det bedste, forstået som estimaterne med mindst varians?

Til WLS kan du anvende følgende SAS-stump:

```
proc reg data = netdata ;
model g60_00 = yr_60 strucK strucH intafyrniv latinam;
weight yr_60;
run;
```

Optionen vægter alle variable med kvadratroden af værdierne af den variabel, der er angivet i `weight`. Se også ”Proc Reg: Weight statement” i SAS' hjælpefunktion.

g) Vægtet mindste kvadraters estimation (WLS) via egen udregning

Gennemfør den vægtede estimation ved at bruge Proc Reg, men *uden* at anvende WEIGHT optionen. Lav i stedet et DATA step, hvor alle variable i regressionen multipliceres med værdierne af $\sqrt{y_{i0}}$. Sammenlign parameterestimater, standardfejl, R^2, F-test og lignende med dem fra spørgsmål f)?

Hint: Du får brug for Proc reg optionen /NOINT, som sikrer, at et konstantled ikke indgår i modellen.

h) Robuste standardfejl

Wooldridge p. 272-75 beskriver en estimator for robuste standardfejl. Det vil sige, en estimator for standardfejlen på OLS estimatoren, som er konsistent uanset tilstedeværelsen af heteroskedasticitet. Formlen (8.4) på p. 274 for en multipel lineær regressionsmodel er ækvivalent med følgende udtryk på matrixform (se også forelæsningsnoten om robust kovariansestimation):

$$\widehat{Var}(\hat{\delta}) = n(X'X)^{-1}S(X'X)^{-1}$$
$$S = \frac{1}{n}\sum_{i=1}^{n}\hat{v}_i^2 x_i x_i' \qquad (1.4)$$

hvor X betegner $n \times (k+1)$ matricen af *k* forklarende variabler plus konstantleddet, vektoren x_i består af den *i*'te række af *X*, og $\hat{v}_i^2$ er de kvadrerede residualer fra model (1.2).

Robuste standardfejl kan beregnes i SAS ved at tilføje optionen / ACOV efter Proc Reg kommandoen. SAS udskriver den robuste kovariansmatrix for parametrene og bruger den som grundlag for efterfølgende TEST kommandoer.

Regresser model (1.2) med ACOV optionen. Brug den robuste kovariansmatrix til at opstille en tabel, hvor du rapporterer estimaterne $\hat{\delta}$ med deres robuste standardfejl og robuste *t*-test for hypotesen: $H_0: \ \delta_j = 0$. Indsæt også de sædvanlige OLS standardfejl i tabellen og de tilsvarende *t*-test, som du beregnede i spørgsmål a). Kommenter på tabellen.

Heteroskedasticitet i model (1.1)

i) Test for heteroskedasticitet i model (1.1)

Gennemfør en analyse svarende til pkt. a) til d) for model (1.1). Er der heteroskedasticitet i modellen? Er vores resultater af analyserne på Ugeseddel 6-8 gyldige eller ej?

Hjemmearbejde

Lav resten af SAS-øvelserne, hvis det ikke er nået til øvelserne.