

Program for øvelserne:

Gruppearbejde og plenumdiskussion
SAS øvelser

Øvelsesopgave: Vækstregressioner (fortsat)

Ugeseddel 13 fortsætter den empiriske analyse af vækstregressionen fra ugeseddel 8. Opgaven tager udgangspunkt i, at to af de forklarende variable i regressionen kan være endogene. Vi søger at løse problemet ved at anvende Instrumentvariabel (IV) estimation. Det involverer bl.a. test af, om instrument variable er brugbare, test for eksogenitet, brug af 2SLS estimator og test af overidentificerende restriktioner.

Vores økonometriske ligning for væksten er givet ved (se også ugeseddel 8):

$$g_i = \beta_0 + \beta_1 \log y_{i0} + \beta_2 (\log s_{K,i} - \log(n_i + 0.075)) + \beta_3 (\log s_{H,i} - \log(n_i + 0.075)) + \beta_4 (\text{Safrika} * \log y_{i0}) + \beta_5 \text{Latinam}_i + u_i \quad (1.1)$$

hvor den afhængige variabel, g_i , repræsenterer den gennemsnitlige vækstrate i realt BNP per beskæftiget for perioden 1960-2000. De forklarende variable dækker over den initiale indkomst, investeringskvoter for fysisk og human kapital samt dummyvariabler, der korrigerer for specielle forhold for lande i Afrika syd for Sahara og i Latinamerika. u_i repræsenterer et fejleddet for hvert af de 80 lande, $i = 1, \dots, 80$.

Modellen er vel-specificeret set i forhold til heteroskedasticitet i fejleddet ifølge vores resultater fra Ugeseddel 9. MLR.5 kan altså antages at holde. Givet de øvrige Gauss-Markov antagelser MLR.1-4 også holder, er OLS er den bedste lineære og middeltrette estimator for parametrene i modellen.

Litteraturen kritiserer dog ofte brugen af OLS estimatoren på konvergensligninger som (1.1).¹ Argumentet er, at en eller flere af de forklarende variable er korreleret med fejleddet. Modellen opfylder i så fald ikke "zero correlation" antagelsen på side 178 i Wooldridge, (og dermed heller ikke den strengere antagelse MLR.4,) så den forklarende variabel er endogen. Konkret ser vi her på initialindkomstniveauet, $\log y_{i0}$.

¹ Fx i Robert Hall og Charles Jones: "Why do some countries produce so much more output per worker than others?" *Quarterly Journal of Economics*, februar 1999.

Gruppearbejde og plenumdiskussion:

Diskuter følgende spørgsmål:

- a. Hvad betyder det for OLS-estimatorens egenskaber, hvis "zero correlation" antagelsen *ikke* holder?
- b. Barro (1991) peger på, at en midlertidig målefejl i initialindkomsten $\log y_{i0}$ kan give inkonsistente OLS estimater for ligning (1.1). Hvordan kan en midlertidig målefejl i initialindkomsten medføre korrelation mellem fejleddet og den forklarende variabel $\log y_{i0}$?
- c. Antag, at midlertidige målefejl eksisterer, og at vi vil løse problemet ved at estimere model (1.1) via IV estimatoren. Kan følgende variable bruges som instrument for $\log y_{i0}$?
 - i) Logaritmen til det samlede forbrug per beskæftiget i land i , $\log c_{i0}$.
 - ii) Den absolutte værdi af afstanden mellem land i og ækvator, $alatitude_i$ (målt ved breddegraden som en andel af 90 grader).

SAS-øvelser:

Du kan med fordel tage udgangspunkt i SAS-programmer fra tidligere ugesedler. Programmerne ligger på hjemmesiden for øvelserne.

Én gruppe bedes skrive en kort opsamling (1/2-1 side) af spørgsmål 7 og sende den til alice.klynge@econ.ku.dk (og din øvelseslærer) senest mandag kl. 12.00.

1. Check udgangspunkt

Datasættet i MERGEDATA bør give de samme 80 brugbare observationer som hidtil. Estimer model (1.1) med OLS estimatoren og check, at resultaterne stemmer overens med dem på Ugeseddel 8.

2. 6-trinsprocedure for IV-estimation

Vi ønsker at anvende 6-trinsproceduren for IV-estimation beskrevet i IV-noten, afsnit 8, for at løse problemet med, at $\log y_{i0}$ potentielt er en endogen forklarende variabel. Bemærk her følgende:

Ad Trin 1: Opstil den økonometriske model og udpeg de endogene regressorer:

Model (1.1) specificerer den strukturelle ligning.

Variablen $\log y_{i0}$ kan være endogen som følge af målefejl. Da variabelen $\log y_{i0}$ også indgår i interaktionsleddet med Afrika-dummen, er der potentielt to endogene regressorer i ligningen, $\log y_{i0}$ og $Safrica \cdot \log y_{i0}$. Resten af variablerne antages at være eksogene.

Ad Trin 2: Find instrumentvariable:

Betragt $alatitude_i$ som et muligt instrument for $\log y_{i0}$. Da variabelen $\log y_{i0}$ også indgår i interaktionsleddet med Afrika-dummen betragter vi også interaktionen mellem Afrika-dummen og $alatitude_i$ som et muligt instrument.

Antag, at $alatitude_i$ er ukorreleret med u_i og således kan indgå i det samlede sæt af eksogene variable, Z (også som interaktion med Afrika-dummen).

Ad Trin 3: Test, om instrumenterne er korreleret med de potentielt endogene regressorer:

Opstil en hjælpe ligning (reduceret form) for hver potentiel endogen variabel, $\log y_{i0}$ og $Safrica \cdot \log y_{i0}$. Regresser hver endogen variabel på alle de eksogene variable, Z_i , ved OLS. Test i hver ligning, om koefficienterne til $alatitude_i$ og $Safrica \cdot alatitude_i$ er signifikant forskellige fra nul.

Ad Trin 4: Udregn residualer fra hjælpe ligningerne:

Gem de to sæt af residualer \hat{E}_{i1} og \hat{E}_{i2} fra de to hjælpe ligninger. Gem også de predikterede værdier \hat{X}_{i1} og \hat{X}_{i2} .

Ad Trin 5: Test af om regressorerne er exogene:

Tilføj de to sæt af residualer fra hjælpe ligningerne som ekstra regressorer i den strukturelle ligning (1.1). Estimer modellen ved hjælp af OLS. Udfør et test for, om $\log y_{i0}$ og $Safrica \cdot \log y_{i0}$ er endogene på grundlag af signifikansen af \hat{E}_{i1} og \hat{E}_{i2} .

Ad Trin 6: Estimer den strukturelle ligning:

Ser det ud til, at midlertidige målefejl i $\log y_{i0}$ faktisk udgør et problem i forhold til konsistent estimation af ligning (1.1)? Er det nødvendigt at bruge IV estimatoren i det tilfælde? Kan man bruge OLS estimatoren? Estimér ligningen givet ved (1.1) via din foretrukne estimator.

3. 6-trinsprocedure for IV-estimation, når antal instrumenter er større end antal endogene regressorer

I spørgsmål 2 var antallet af instrumenter to og dermed lig antal endogene regressorer. Vi vil nu genoverveje spørgsmålet om endogenitet og se på mulighederne ved at bruge et ekstra instrument i form af variabelen $yrsoopen_i$. Hall og Jones beskriver variabelen $yrsoopen_i$ således: "[$yrsoopen_i$] measures the fraction of years during the period 1950 to 1994 that the economy has been open and is measured on a {0,1} scale."

Overvej, om $yrsoopen_i$ kan være et gyldigt instrument for initialindkomsten. Gennemfør 6-trinsproceduren igen, men denne gang med tre instrumenter.

4. IV-estimator beregnet ved OLS i to trin

Den netop gennemførte 6-trinsprocedure lægger op til, at variabelen $\log y_{i0}$ er endogen, hvorfor vi skal bruge IV estimatoren for at regressere ligning (1.1) konsistent. Når vi har flere instrumenter end endogene regressorer beregnes IV-estimatoren ved hjælp af OLS i to trin, Two Stage Least Squares (2SLS)-estimatoren.

Første trin af 2SLS estimationen er allerede sket i spørgsmål 3, da den består i at regressere hjælpe ligningerne og beregne de forudsagte værdier for de endogene regressorer. Udfør nu andet trin i 2SLS estimationen: Estimer model (1.1) via OLS estimatoren, hvor du erstatter $\log y_{i0}$ og $Safrica \cdot \log y_{i0}$ med de predikterede værdier, \hat{X}_{i1} og \hat{X}_{i2} , fra hjælperegressionerne i trin (4). Sammenlign de estimerede koefficienter med resultaterne fra trin (5) i spørgsmål 3.

5. Beregn korrekte standardfejl for 2SLS estimaterne

Koefficienterne fra spørgsmål 4 giver de korrekte 2SLS estimater, men ikke de korrekte standardfejl, da de rapporterede OLS standardfejl ikke tager højde for, at \hat{X}_{i1} og \hat{X}_{i2} er estimerede regressorer. Andre størrelser fra OLS regressionen i trin 2, fx residualerne, kan heller ikke umiddelbart bruges.

De korrekte 2SLS standardfejl kan beregnes ved hjælp af proceduren Proc Syslin i SAS, som (blandt andet) kan udføre 2SLS estimation. Brug 2SLS standardfejlene til at teste hypotesen, at *strucK* ikke indgår signifikant i konvergensligningen. Formuler hypotesen og udfør det relevante test.

Du kan bruge følgende programstump:

```
proc syslin 2SLS data = mergedata out=out2SLS3 ;
endogenous lyr_60 intAfyr ;
instruments alatitude strucK strucH intAfal latinam yrsopen ;
model g60_00 = lyr_60 strucK strucH intAfyr latinam ;
output residual=res2SLS3 ;
run ;
```

Kommandoen "instruments" skal indeholde alle eksogene variable, dvs. både de eksogene variable fra den strukturelle ligning (her: *strucK*, *strucH* og *latinam*) og de egentlige instrumentvariable (her: *alatitude*, *Safrica* · *alatitude_i* og *yrsopen*).

Output kommandoen sørger for, at residualerne fra 2SLS estimationen bliver gemt i datasættet OUT2SLS3.

Du kan få mere at vide om Proc Syslin ved at søge i SAS' hjælpefunktion.

6. Test af overidentificerende restriktioner

Udfør et test af overidentificerende restriktioner ved at regressere 2SLS residualerne, \hat{u}_i , på hele sættet af eksogene variable, Z_i . Beregn teststatistikken som nR^2 , hvor n er antal observationer og R^2 kommer fra hjælperegressionen. Hvor mange overidentificerende restriktioner er der her? Hvad bliver konklusionen om de tre instrumenter ud fra testet?

Hvad ville blive resultatet af et tilsvarende "test" i spørgsmål 2, hvor der kun var to instrumenter? (prøv at køre en hjælperegression, der svarer til spørgsmål 2, men husk at bruge de rigtige residualer).

7. Tabel med sammenfatning af resultater

Opstil en tabel, der sammenfatter de estimater af parametrene i den strukturelle ligning (1.1), som du har opnået i spørgsmål 1, 2 og 5. Kommenter kort på tabellen, herunder på, hvordan de forskellige modeller forholder sig til hinanden og hvilke(t) sæt af estimater, du vil lægge til grund for din vurdering af parametrene i konvergensligningen i en Solow model.

Hjemmearbejde

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