ECON 4160: Econometrics–Modelling and Systems Estimation:

Computer Class # 3

André Anundsen

Department of Economics, University of Oslo

September 23, 2013

Outline

Data sets

Diagnostics

Monte Carlo

General-to-specific (Gets) modelling

Exogeneity concepts

Data sets for today, posted on the web page:

- KonsDataSim2.zip
- KonsData2Nor.in7
- SimDataGets.xls

We will also work with the following batch files:

- GNC (example of test for GNC)
- Super_exo.fl (example of test for super exogeneity)

Confer also *Note 2 to Computer Class Standard mis-specification tests*

Test battery in PcGive I: Non-normality

Normality, Jarque and Bera (1980): (Note: Small sample correction in Give)
Test the joint hypothesis of no skewness and no excess kurtosis (3^{rd} and 4^{th} moment of the normal distr.):

$$JB = \frac{T}{6}\left(\text{Skewness}^2 + \frac{1}{4}\textit{Excess kurtosis}^2\right) \tag{1}$$

The sample skewness and excess kurtosis (skewness $= 0$ and kurtosis $= 3$ for normal distr.) are defined as follows

$$\text{Skewness} = \frac{\frac{1}{T}\sum_{t=1}^{T}\left(\hat{\varepsilon}_t - \bar{\hat{\varepsilon}}\right)^3}{\left(\frac{1}{T}\sum_{t=1}^{T}\left(\hat{\varepsilon}_t - \bar{\hat{\varepsilon}}\right)^2\right)^{\frac{3}{2}}} = \frac{\frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_t^3}{\left(\frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_t^2\right)^{\frac{3}{2}}} \tag{2}$$

$$\text{Excess kurtosis} = \frac{\frac{1}{T}\sum_{t=1}^{T}\left(\hat{\varepsilon}_t - \bar{\hat{\varepsilon}}\right)^4}{\left(\frac{1}{T}\sum_{t=1}^{T}\left(\hat{\varepsilon}_t - \bar{\hat{\varepsilon}}\right)^2\right)^{\frac{4}{2}}} - 3 = \frac{\frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_t^4}{\left(\frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_t^2\right)^{\frac{4}{2}}} - 3 \tag{3}$$

Null is normality.

Test battery in PcGive II: Heteroskedasticity

Heteroskedasticity, White (1980):
Auxiliary regression:

$$\hat{\varepsilon}_t^2 = \beta_0 + \beta_1 X_t + \beta_2 X_t^2 + u_t \tag{4}$$

Test that $\beta_1 = \beta_2 = 0$ against non-zero using an F-test.
Hetero X (two or more regressors), White (1980)
Auxiliary regression:

$$\hat{\varepsilon}_t^2 = \beta_0 + \beta_1 X_t + \beta_2 Z_t + \beta_3 X_t Z_t + \beta_4 X_t^2 + \beta_5 Z_t^2 + u_t \tag{5}$$

Test that $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$ against non-zero using an F-test.
In both cases the null is homoskedasticity.

Test battery in PcGive III: Autocorrelation

Autocorrelation, Godfrey (1978):
Auxiliary regression:

$$\hat{\varepsilon}_t = \beta_0 + \sum_{i=1}^{p} \beta_i \hat{\varepsilon}_{t-i} + \beta_p X_t + u_t \tag{6}$$

A test for p^{th} order autocorrelation is then to test that $\beta_1 = \beta_2 = \cdots = \beta_p = 0$ against non-zero using an F-test.
Null is no autocorrelation.

Test battery in PcGive IV: Autoregressive conditional heteroskedasticity

ARCH, Engle (1982):
Auxiliary regression:

$$\hat{\varepsilon}_t = \beta_0 + \sum_{i=1}^{p} \beta_i \hat{\varepsilon}_{t-i}^2 + u_t \tag{7}$$

A test for p^{th} order ARCH is then to test that
$\beta_1 = \beta_2 = \cdots = \beta_p = 0$ against non-zero using an F-test.
Null is no ARCH.

Test battery in PcGive V: Regression Specification Test

RESET, Ramsey (1969):
Auxiliary regression (Note 2,3 in PcGive means squares and cubes!):

$$\hat{\varepsilon}_t = \beta_0 + \beta_1 X_t + \beta_2 \hat{Y}_t^2 + \beta_3 \hat{Y}_t^3 + u_t \tag{8}$$

A test for specification error is then to test that $\beta_2 = \beta_3 = 0$ against non-zero using an F-test.
Null is no specification error.

What is Monte Carlo simulation? I

Say that the process that has generated the data (the data generating process, or the DGP) takes the following form:

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t \tag{9}$$

where ε_t is WN. Now, assume that we had some data $t = 1, \ldots, T$ on y_t and x_t, and that we want to pin down β_0 and β_1 (our parameters of interest)

As long as $Cov(x_t, \varepsilon_t) = 0$, you know that the OLS estimator is BLUE! How could we confirm this by simulation?

What is Monte Carlo simulation? II

1. Fix β_0 and β_1 in (9) at some values, e.g. $\beta_0 = 2$ and $\beta_1 = 1.5$
2. Generate some numbers for the time series x_t on a sample $t = 1 \ldots, T$
3. Assume e.g. that $\varepsilon_t \sim N(0, 1)$, and draw T numbers from the standard normal distribution
4. Then, the time series y_t for the sample $t = 1, \ldots T$ will follow by definition from the DGP!
5. Estimate an equation of the form (9) by OLS and collect your β_0 and β_1 estimates; call them $\hat{\beta}_0^1$ and $\hat{\beta}_1^1$
6. Now, repeat the steps 1–5 M times, and calculate $\beta_0^{MC} = \frac{\sum_{m=1}^{M} \hat{\beta}_0^m}{M}$ and $\beta_1^{MC} = \frac{\sum_{m=1}^{M} \hat{\beta}_1^m}{M}$!

What is Monte Carlo simulation? III

As we have seen, β_0^{MC} and β_1^{MC} are nothing but the mean estimates obtained by repeating 1–5 M times. From the law of large numbers, we know that β_0^{MC} and β_1^{MC} converge to $E(\hat{\beta}_0)$ and $E(\hat{\beta}_1)$ as $M \to \infty$.

But then, we know that a measure of the biases of these estimates would be: $\beta_0^{MC} - \beta_0$ and $\beta_1^{MC} - \beta_1$. What we can also do is to consider a recursive Monte Carlo, where we e.g. vary the sample size from say $T = 20$ to $T = 500$ in increments of 20 to check the bias of the OLS estimator for different samples!

Experiment 1: The properties of the OLS estimator when we estimate the true DGP I

Suppose the DGP is given by:

$$y_t = 2 + 1.5x_t + \varepsilon_t$$

We want to explore the properties of the OLS estimator for different sample sizes when the classical assumptions are satisfied (we keep the number of MC replications at $M = 1000$ for all samples). Do we expect the OLS estimator to be unbiased? Let's first derive this theoretically! In matrix notation, the equation of interest takes the following form:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

where in our example $\mathbf{y}$ is $T \times 1$, $\mathbf{X}$ is $T \times 2$ (with ones in the first column), $\boldsymbol{\beta}$ is 2×1 and $\boldsymbol{\varepsilon}$ is $T \times 1$. The OLS estimator is given as:

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}$$

Experiment 1: The properties of the OLS estimator when we estimate the true DGP II

By substituting in for $\mathbf{y}$, we get:

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\left(\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}\right)\\ &= \left(\mathbf{X}'\mathbf{X}\right)^{-1}\left(\mathbf{X}'\mathbf{X}\right)\boldsymbol{\beta}+\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\\ &= \boldsymbol{\beta}+\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\end{aligned}$$

Taking expectations, we get:

$$\begin{aligned}E\left(\hat{\boldsymbol{\beta}}\right) &= \boldsymbol{\beta}+E\left(\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\right)\\ &= \boldsymbol{\beta}+\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'E\left(\boldsymbol{\varepsilon}\right)=\boldsymbol{\beta}\end{aligned}$$

i.e. we expect the OLS estimator to be unbiased! Let's check this out by MC!

Experiment 2: The properties of the OLS estimator when we have an omitted variable I

Suppose the DGP is given by:

$$y_t = 2 + 1.5x_t - 2z_t + u_t$$

This means that, if we were to estimate the true DGP, we would consider a model of the following form:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\delta + \mathbf{u}$$

where the notation is as previously, $\mathbf{Z}$ is a $T \times 1$ matrix and δ is a scalar. Suppose instead that the model we consider is the same model as considered previously:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

with the OLS estimator given by:

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}$$

Experiment 2: The properties of the OLS estimator when we have an omitted variable II

Would this be an unbiased estimator? Let's do exactly the same as previously (substitute in for **y** from the DGP). We get:

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\boldsymbol{\beta}+\mathbf{Z}\delta+\mathbf{u}) \\ &= (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{X})\boldsymbol{\beta}+(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}\delta+(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} \\ &= \boldsymbol{\beta}+(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}\delta+(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\end{aligned}$$

Taking expectations, we get:

$$\begin{aligned}E\left(\hat{\boldsymbol{\beta}}\right) &= \boldsymbol{\beta}+E\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}\delta\right)+E\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\right) \\ &= \boldsymbol{\beta}+(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}\delta+(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\boldsymbol{\varepsilon}) \\ &= \boldsymbol{\beta}+(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}\delta \neq \boldsymbol{\beta}\end{aligned}$$

i.e. we expect the OLS estimator to be **biased**! Again, let's check this out by some simulations!

Example of omitted variables bias using simulated consumption data

- Consider a three-variable equation system for y, x and z (a VAR)
- The conditional expectation function for y is then a function of x and z (we will simplify and abstract from lags now), i.e. the conditional model for y depends on x and z (confer Lecture # 3, CC #2 and Lecture note 3)
- Analyze this in OxMetrics/PcGive with the use of the artificial data set in *KonsDataSim2.zip* on the web page
- We will only use the variables named $y = C$ (consumption), $x = I$ (income) and $z = F$ (households' wealth)
- The Data Generating Process (DGP) is

$$C = 45 + 0.8I + 0.05F$$

- Choose the same sample size as in the first computer class: 1960-2007

Heuristics of Gets modelling I

Usually, we are not in the fortunate situation of knowing the process that has generated the data, i.e. we cannot simply estimate the DGP. In empirical research, it is therefore imperative to combine:

1. Economic theory
2. Previous empirical results
3. Institutional knowledge

The Gets approach utilizes information on these things to try to model the DGP

Heuristics of Gets modelling II

Suppose that the DGP (that we don't know) is given by:

$$y_t = 2 + 1.5x_t + 0.2x_{t-1} - 2z_t - 0.3z_{t-3} + u_t$$

Even though we don't know the DGP, we might have some theory stating that y_t is linked to x_t and z_t. However, there might be uncertainty e.g. about the dynamics (the lags), the theory might be incomplete, we might have different theories etc..

Heuristics of Gets modelling III

The idea behind the Gets approach is to allow for a general specification at the outset (the *General Unrestricted Model*, or the *GUM*) that nests the DGP (to the extent it is possible) and then use statistical testing to pin down the DGP. As an example, suppose we started out with a model (nesting our DGP) of the following form:

$$y_t = \beta_0 + \sum_{i=1}^{5} \beta_i x_{t+1-i} + \sum_{i=0}^{4} \gamma_i z_{t-i} + e_t \tag{10}$$

The goal is then to use statistical testing to select the relevant regressors (x_t,x_{t-1}, z_t and z_{t-3}).

Note: It is key that your GUM is well specified, since the statistical testing we shall utilize relies on no residual mis-specification!

Heuristics of Gets modelling IV

There will be two main challenges:

1. *Retain* the variables that are *relevant*
2. *Eliminate* the variables that are *irrelevant*

In principle, if we include N variables in our GUM, then there will exist 2^N possible models! How do we proceed to search for the DGP?

A common approach is to do sequential t-testing testing, i.e. drop insignificant variables successively, starting with the least significant variable and stopping when only significant variables are retained.

In the simplest case, suppose all the potential regressors are mutually orthogonal, then choose some significance level α (with an associated critical value from the t-distribution given by t_α). Then estimate the GUM and compute the square of the t-value for all N regressors ($\hat{t}_i^2 \; \forall \; i = 1, \ldots N$) to avoid considering the signs. Rank these squared t-statistics from smallest to largest and eliminate all variables where $\hat{t}_i^2 < t_\alpha^2$!

Heuristics of Gets modelling V

You now see that by selecting the significance level, we can control how many irrelevant variables we will retain on average! When the significance level is α, and we consider N independent potential regressors, we will retain only αN irrelevant variables on average when all of them are irrelevant!

In a same way, we can calculate the probability that none of the t-tests will reject the null when all N variables are irrelevant:

$$(1-\alpha)^N$$

meaning that in $\left(1-(1-\alpha)^N\right) \times 100\%$ of the cases, we will select the wrong model.

Simulation studies show that the problem with Gets modelling is not eliminating irrelevant variables, but retaining the relevant variables (the power).

Heuristics of Gets modelling VI

Simulation studies have, however, shown that the Gets approach is successful also in retaining the relevant variables (you'll get an example shortly).

Recommended literature on Gets modelling:

- Hoover, K. D. and S. J. Perez (1999). Data Mining Reconsidered: Encompassing and the General-to-Specific Approach to Specification Search. Econometrics Journal 2, 167–191.
- Hendry, D. F. and H.-M. Krolzig (2005). The Properties of Automatic Gets Modelling. Economic Journal 115, 32–61.

Example of manual Gets

I have generated some data in accordance with the following DGP:

$$y_t = 2 + 1.5x_t + 0.2x_{t-1} - 2z_t + -0.3z_{t-3} + u_t$$

Start with a specification of the following form:

$$y_t = \beta_0 + \sum_{i=1}^{5} \beta_i x_{t+1-i} + \sum_{i=0}^{4} \gamma_i z_{t-i} + e_t \tag{11}$$

Download the data set contained in *SimDataGets.zip* and use manual Gets to see what you find!

Automated Gets modelling (*Autometrics*)

In *PcGive* there is a program called *Autometrics* that automatizes a more complicated version of the Gets approach (the computer does the job for you)! The algorithm does a multi-path search, and even though we won't have time for the details now, I recommend that you read the following:

- Doornik, J. (2009). Autometrics. In J. L. Castle and N. Shephard (Eds.), The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry. Oxford: Oxford University Press.

Use the same data set to see what *Autometrics* returns!

Monte Carlo to see how *Autometrics* is performing I

Suppose now that we have the following DGP:

$$y_t = 1 + 0.8x_{1,t} + u_t \; t = 1, \ldots T$$

But that we started our with a GUM of the following form:

$$y_t = \beta_0 + \sum_{i=1}^{20} \beta_i x_{i,t} + e_t$$

i.e. we include 19 irrelevant regressors! Let's use a significance level of $\alpha = 0.05$

We then expect to retain an irrelevant variable about 5% of the times, and we expect to find the true model a total of $(1 - 0.05)^{20} \times 100 = 36\%$ of the times!

Let's do this for $T = 200$ observations using $M = 1000$ MC

Monte Carlo to see how *Autometrics* is performing II

Some interesting questions:

- When we apply Autometrics, how many times do we falsely retain an irrelevant variable?
- How many times do we retain the relevant variables?
- How many times do we find the true DGP?
- Even if we retain some irrelevant variables, how does that bias our estimates?

Four concepts of exogeneity

1. Weak exogeneity
2. Strong exogeneity
3. Super exogeneity
4. Strict exogeneity and pre-determinedness

VAR

At CC #2, we considered a bi-variate VAR model of first order:

$$\begin{pmatrix} y_{1,t} \\ y_{2,t} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} \tag{12}$$

A more compact notation gives:

$$\mathbf{y}_t = \boldsymbol{\mu} + \mathbf{A}\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t \tag{13}$$

Let us now assume that $\boldsymbol{\varepsilon}_t \sim MVN(\mathbf{0}_{2\times 1}, \boldsymbol{\Sigma})$, i.e.:

$$\begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} \sim MVN\left(\mathbf{0}_{2\times 1}, \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix}\right) \tag{14}$$

We also that showed that the bi-variate VAR(1) in 12 could expressed in **model form** by:

$$y_{1,t} = \phi_0 + \beta_0 y_{2,t} + \beta_1 y_{2,t-1} + \phi_1 y_{1,t-1} + \epsilon_t \quad (15)$$

$$y_{2,t} = \mu_2 + a_{21} y_{1,t-1} + a_{22} y_{2,t-1} + \varepsilon_{2,t} \quad (16)$$

$$\begin{aligned} E(\epsilon_t) &= 0 \; \forall \; t \\ Var(\epsilon_t) &= \sigma_1^2 \left(1 - \rho_{12}\right) \; \forall \; t \\ E(y_{2,t}\epsilon_t) &= 0 \; \forall \; t \\ E(\varepsilon_{2,t}\epsilon_t) &= 0 \; \forall \; t \end{aligned}$$

Weak exogeneity

If there are no efficiency gains from considering the system (15 and 16) rather than only considering equation 15, we say that $y_{2,t}$ is weakly exogenous (WE) with respect to the parameters of interest, i.e. we don't "loose" anything by abstracting from the marginal model for $y_{2,t}$!

Granger non-causality (GNC)

- We see that in one sense (15) and (16) define a recursive model, since given the history (represented by $y_{1,t-1}$ and $y_{2,t-1}$), then $y_{2,t}$ is determined first, and given this, $y_{1,t}$ is determined. Moreover, $Cov(\varepsilon_{2,t}, \epsilon_t) = 0$
- However, the two variables are clearly jointly determined over time, since in general $y_{2,t-1} \longrightarrow y_{1,t}$ and $y_{1,t-1} \longrightarrow y_{2,t}$. In econometrics we call this joint Granger causality. Only if $a_{21} = 0$ can we say that we have a recursive causal chain
- With $a_{21} = 0$ imposed, we say that $y_{1,t-1}$ is not Granger-causing $y_{2,t}$ while $y_{2,t-1}$ is Granger-causing $y_{1,t}$.

Strong exogeneity and super exogeneity

Strong exogeneity:

If $y_{2,t}$ is WE wrt. the parameters of interest and $y_{1,t}$ is not Granger causing $y_{2,t}$, we say that the variable $y_{2,t}$ is strongly exogenous (SE)!

Super exogeneity:

If the coefficient on $y_{2,t}$ in the conditional model for $y_{1,t}$ (β_0 in equation (15)) is invariant to structural breaks in the marginal model for $y_{2,t}$, then we say that $y_{2,t}$ is super exogenous (SpE) in the conditional model for $y_{1,t}$!

Invariance, super exogeneity and autonomy

- Super-exogeneity is the property that we have constant parameters in the conditional model for $y_{1,t}$ even in periods where there is a structural break in the marginal equation for $y_{2,t}$
- Super exogeneity is defined for conditional models, but the concept is related to the more general idea of autonomy
- Econometric models with parameters that are invariant in the face of wide range of structural breaks have a high degree of autonomy

Testing invariance and super exogeneity and autonomy

To test a hypothesis of lack of invariance we need to investigate two issues:

1. Test the null hypothesis of no-structural breaks in the marginal model
2. Test the null hypothesis of stability in the conditional model

We have several tools available:

- Recursive estimation and recursive graphs
- Recursive Chow tests (more on this next time)
- Find significant dummy variables that represent structural breaks in a marginal model, and test the significance of those dummies in the conditional model (here we can use the Autometrics algorithm we've learnt about today)

Strict exogeneity and pre-determinedness

Strict exogeneity:

With reference to (15), if $E(\epsilon_t y_{2,t}) = 0$, $y_{2,t}$ is a strictly exogenous variable

Pre-determinedness:

Again, if in equation (15), we have that $E(\epsilon_{t+j} y_{1,t-1}) = 0 \; \forall \; j \geq 0$, we say that $y_{1,t-1}$ is a pre-determined variable

Examples using

- KonsData2Nor.in7
- GNC
- Super_exo.fl