ECON 4160: Econometrics–Modelling and Systems Estimation:

Computer Class 1

Ragnar Nymoen

Department of Economics, University of Oslo

Last revision: August 29, 2014

Aims and purpose I

- Use computer program to learn about:
  1. Econometric theory
     - Complementary and supplementary to the lectures
  2. Applied econometrics
     - Integrated with the seminar series

Aims and purpose II

- In the CC class, we will demonstrate the implications of the theoretical results established in the lectures through Monte Carlo simulations, and we will apply the methods you learn to real-world data!
- To each seminar, you will be given a set of exercises. The exercises are of two types:
  - Applied modeling tasks, where you use your (theoretical) econometric skills in combination with the skills you acquire in the computer class to analyze real-world data. You will then hold a live computer presentation of your solution proposal at the seminar
  - The other type of seminar exercises will be more theoretical and algebraic, establishing results that are extremely important to have in mind when analyzing a data set

How do I choose which software package to use?II

- Menus, batch language and programming capability
  - Menus are good for getting started, and for demonstrations
  - Collecting many commands in batch-files is important for:
    1. Efficient work (once you become an "expert")
    2. Documentation (colleagues, yourself and journals!)
    3. Communication (e.g., between supervisor and yourself)
- Batch files in OxMetrics serve the same function as do-files in Stata.
- Programming capability can be important to increase flexibility (can't always do "everything "using menus and batch language)

How do I choose which one to use? I

- The purpose of your project!
  - Econometric programs are like (specialized) tools: They are designed to do specific tasks efficiently
  - Each program has its strengths and weaknesses
  - Unless you are very specialized yourself, you will probably end up using more than one program

Why do we use OxMetrics/PcGive in this course?

- OxMetrics is a powerful package including an option to estimate all models we consider in this course (particularly good at simultaneous equation systems! Give = Generalized Instrumental Variable Estimation)
  - and... the names of the models in the program are close to the econometric models you will become familiar with throughout the course
- Fairly easy to do Monte Carlo simulations demonstrating how e.g. heteroskedasticity, autocorrelation and endogneity affects estimated coefficients when we use OLS
- Finally, it is relatively easy to use and provides a lot of output that are essential to any econometric analysis!

PcGive

Three books are on the web page of this course.
Or use the help menu in the program (we'll see later)
Note also:

http://pcgive.com/pcgive/index.html

Topics for first computer class

- Loading the data into the program and get to "know" the data (always start by looking at graphs in case there are data contamination, unusual shifts in the time series etc.!)
- Variable transformations
- Simple regression and mis-specification testing
- Regression and mis-specification testing with the use of the batch language
- Stability of regression models

Reference notes for today

On the course web page:

www.uio.no/studier/emner/sv/oekonomi/ECON4160/h13/index.html,

you will find the following notes:

- *A first regression in OxMetrics/PcGive/Introduction to OxMetrics*
- *Note 1 to Computer class: Use of the natural logarithm*
- *Note 2 to Computer class: Standard mis-specification tests*

You may keep them open as a reference during today's class.

Reference note: A first regression in OxMetrics/PcGive

- Download the zip-file **KonsDataSim.zip** from the course web page . Right click and choose *extract all*
- Load the data **konsum_sim.in7** into the program! Do the same using **konsum_sim.xls**, just to see how easy it is to switch between the different formats!
- Now, we shall follow the step-by-step instruction in *A first regression in OxMetrics/PcGive* to do exactly that!

Regression with experimental and non-experimental data

Consider the modelling task with experimental/lab data:

$$\underset{\text{result}}{Y_i} = \underset{\text{input}}{g(X_i)} + \underset{\text{shock}}{v_i} \tag{1}$$

and compare with the situation with non-experimental, real-world data:

$$\underset{\text{observed}}{Y_i} = \underset{\text{explained}}{f(X_i)} + \underset{\text{remainder}}{\varepsilon_i} \tag{2}$$

Clearly, we know much less about the match between $f(X_i)$ and Y_i in the non-experimental case: We simply can't control the input and then study the output. Often, we may not even know all variables that are included in the vector X_i! In Lecture 1 we developed this by factorization of the joint pdf of Y and X. All our choices of functional form, $f(\cdot)$, and explanatory variables, X_i, will be reflected in the remainder ε_i:

$$\varepsilon_i = Y_i - f(X_i) \tag{3}$$

However, we will follow custom and refer to ε_i as the *disturbance* and the estimated counterpart $\hat{\varepsilon}_i$ as the *residual*.

Lecture 1: When we choose regression, the $f(X_i)$ function is the conditional expectation:

$$f(X_i) = E(Y_i|X_i) \tag{4}$$

and it follows that

$$E(\varepsilon_i|X_i) = E((Y_i - f(X_i))|X_i) = 0 \tag{5}$$

- But if $E(Y_i|X_i)$ is incomplete or wrong relative to the DGP, ε_i will in general be correlated with omitted variables (even non-linear functions of X!)
- The assumptions that we make about the disturbances , e.g., the "classical assumptions" in applied modelling, are only tentative, and that we need to validate them after estimation
- This is called residual mis-specification testing

Residual mis-specification overview

	Disturbances ε_i are:			
X_i	heteroscedastic		autocorrelated	
	$\hat{\beta}_1$	$\widehat{Var}(\hat{\beta}_1)$	$\hat{\beta}_1$	$\widehat{Var}(\hat{\beta}_1)$
exogenous	unbiased consistent	wrong	unbiased consistent	wrong
predetermined	biased consistent	wrong	biased inconsistent	wrong

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.):

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

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

$$\text{Skewness} = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{\varepsilon}_i - \bar{\hat{\varepsilon}}\right)^3}{\left(\frac{1}{n}\sum_{i=1}^{n}\left(\hat{\varepsilon}_i - \bar{\hat{\varepsilon}}\right)^2\right)^{\frac{3}{2}}} = \frac{\frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^3}{\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^2\right)^{\frac{3}{2}}} \tag{7}$$

$$\text{Excess kurtosis} = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{\varepsilon}_i - \bar{\hat{\varepsilon}}\right)^4}{\left(\frac{1}{n}\sum_{i=1}^{n}\left(\hat{\varepsilon}_i - \bar{\hat{\varepsilon}}\right)^2\right)^{\frac{4}{2}}} = \frac{\frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^4}{\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^2\right)^{2}} \tag{8}$$

Null is normality.

Test battery in PcGive II: Heteroskedasticity

Heteroskedasticity, White (1980):
Auxiliary regression:

$$\hat{\varepsilon}_i^2 = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + u_i \tag{9}$$

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

$$\hat{\varepsilon}_i^2 = \beta_0 + \beta_1 X_i + \beta_2 Z_i + \beta_3 X_i Z_i + \beta_4 X_i^2 + \beta_5 Z_i^2 + u_i \tag{10}$$

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 (note, *i* is now time unit! Call it *t*):

$$\hat{\varepsilon}_t = \beta_0 + \sum_{i=1}^{p} \beta_i \hat{\varepsilon}_{t-i} + \beta_p X_t + u_t \tag{11}$$

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:

$$\hat{\varepsilon}_t = \beta_0 + \sum_{i=1}^{p} \beta_i \hat{\varepsilon}^2_{t-i} + u_t \tag{12}$$

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!):

$$\hat{\varepsilon}_t = \beta_0 + \beta_1 X_t + \beta_2 \hat{Y}_t^2 + \beta_3 \hat{Y}_t^3 + u_t \tag{13}$$

A test for p^{th} order ARCH 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 date (the data generating process, the DGP) takes the following form:

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t \tag{14}$$

where ε_t is normally distributed. 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_1 (our parameter of interest)

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

What is Monte Carlo simulation? II

So, what do we do?

1. Fix β_0 and β_1 in (14) 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. Say that $\varepsilon_t \sim N(0,1)$, and draw T numbers from the standard normal distribution
4. Then, y_t will follow by definition from the DGP!
5. Estimate an equation of the form (14) by OLS and collect your β_1 estimate; call it $\hat{\beta}_1^1$
6. Now, repeat the steps 1–5 M times, and calculate $\beta_1^{MC} = \frac{\sum_{m=1}^{M} \hat{\beta}_1^m}{M}$! This is just the mean estimator, which by the law of large numbers converges to $E(\hat{\beta}_1)$ as $M \to \infty$.

What is Monte Carlo simulation? III

But then, we know that the estimator is unbiased if $\beta_1^{MC} - \beta_1 = 0$. Let us vary the sample size from $T = 20$ to $T = 500$ in increments of 20 and do this experiment with $M = 1000$ to check the unbiasedness of the OLS estimator!