Exercises to PcGive lectures
2. September 2009

We outline 10 exercises for this course in econometrics. The exercises have been designed such that students should be able to carry out their own empirical research using the datasets which belong to the different exercises. Data are organized in the form of Lotus 123 spreadsheets (*.wk1), Excel spreadsheets (*.xls) or PcGive data files (*.in7 *.bn7). Sample batch files (*.fl) are available to get a head start to analyze the data. Results will be discussed during the lectures. See also the associated handout file with results.

Contents

A *Exercise 1A* Inflation and excess demand

Let p_t be (the log) of headline CPI, y_t denotes (the log of) output in the Norwegian mainland economy (fixed prices), and y_t^* is the log of a measure of potential output (equilibrium growth path) for the mainland economy.

Assume that the following simple relationship holds between the annual rate of CPI growth, $\Delta_4 p_t$, and "output gap" measured by the (log) difference $(y_t - y_t^*)$.
What is the interpretation of the "output gap" in words?

(1) $\Delta_4 p_t = \alpha + \beta \Delta_4 p_{t-1} + \gamma (y - y^*)_t + \varepsilon_t$

assuming $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$.
Disposition of PcGive batch file E01-INFL.FL

- Read and calculate necessary variables
- Calculate "potential output" y_t^*
- Calculate the "output gap" $(y_t - y_t^*)$
- Formulate and estimate the model using OLS
- Evaluate the model
- Consider alternative model specifications

Batch-commands in the example-program E01-INFL.FL

module	module selection, e.g. PcGive (single equation modeling)
loaddata	load data from disk (in7,bn7),wk1,xls,...
algebra	data transformations
system	model specification
println	print line to output file (Results)
estsystem	estimation e.g. using OLS
store	stores "fitted" values of regression to the database

B *Exercise 1B* Investment relationships for US (Greene)

Read the batch-file `E01-GINV.FL` into a suitable editor (e.g., OxEdit), or load the batch file directly into GiveWin. The program reads the dataset `GREENINV` (see Greene) which contains annual observations of the following six variables: Annual time trend (years), $YEAR$, nominal GNP, $NGNP$, nominal gross investment, $NINV$, headline CPI, CPI, interest rates (discount rate) $RDIS$, annual rate of inflation, $INFL$, for the period from 1968 to 1982. Write a batch file which carries out the following steps:

a) Read data

b) Make new variables $y = NGNP/(10 * CPI), g = NINV/(10 * CPI), r = RDIS, p = INFL$ as described in Greene and give a motivation for the two first transforms.

c) Plot data for y and g

e) Estimate the coefficients in the investment equation

(2) $\quad y_t = \alpha + \beta_g g_t + \beta_r r_t + \beta_p p_t + \beta_t t_t + u_t$

under the assumption that $u_t \sim N(0, \sigma_u^2)$. Give a motivation for the relationship, and interpret the estimated coefficients in light of economic theory.

f) Calculate the model residuals $\hat{u}_t = y_t - \hat{y}_t$ and interpret the misspecification diagnostics which are calculated by the command `testsummary;` in the batch-file.

C *Exercise 1C* The US macro consumption function (Maddala)

Read the batch-file `E01-USCP.FL`. The program reads the dataset `USCP` which contain observations of an annual time trend, $YEAR$, private consumption, CP, and household real disposable income YP. There are missing observations for some years, and we must take this into consideration in the empirical analysis.

a) Estimate the coefficients in the consumption function

(3) $C_t = \alpha + \beta Y_t + u_t$

under the assumption that $u_t \sim N(0, \sigma_u^2)$.

b) Calculate the model residuals $\hat{u}_t = C_t - \hat{C}_t$, and interpret the misspecification diagnostics which are calculated by the command `testsummary;` in the batch-file.

D *Exercise 2* Using PcGive to create and analyze artificial data

1. Use the ranseed(arg) and rann() functions to generate the stochastic variables $eps1$ and $eps2$
2. Then generate $z1$ and $z2$ according to the formulae in the batch-file. What are their properties?
3. Generate y_t and x_t according to the formulae in the batch-file. What are their properties?
4. Regress y_t on x_t and a constant. What is a sensible estimate of the regression coefficient and how should this result be interpreted?
5. Regress y_t on $z1_t$ and $z2_t$. What are the expected values of the OLS-estimates?
6. What will happen if x_t is heteroskedastic?
7. What happens if $z2_t$ is heteroskedastic?

Example E02-RANN (DGP formulae):

First, construct two normally distributed random variables $\varepsilon_{1t}, \varepsilon_{2t}$ with mean zero but with different variance:

$$\varepsilon_{1t} \sim \text{Niid}(0, (0.01)^2), \varepsilon_{2t} \sim \text{Niid}(0, (0.02)^2)$$

Then we define the variables z_{1t}, z_{2t} as follows:

$$z_{1t} = \varepsilon_{1t} + 0.5\varepsilon_{2t} + 0.005t + 0.01\text{rann}()$$

$$z_{2t} = \varepsilon_{1t} - 0.5\varepsilon_{2t} + 0.0025t + 0.01\text{rann}()$$

We note that z_{1t}, z_{2t} are linear functions of $\varepsilon_{1t}, \varepsilon_{2t}$, a linear trend t and a normally distributed random term generated by the function `rann()`.

Now, construct the variables y_t, x_t as

$$y_t = 0.4z_{1t} + 0.2z_{2t} + 0.01\text{rann}()$$

$$x_t = 0.6z_{1t} - 0.2z_{2t} + 0.01\text{rann}()$$

1. Regress y_t on x_t and a constant. What is a sensible estimate of the regression coefficient and how should this result be interpreted?

2. Regress y_t on $z1_t$ and $z2_t$. What are the expected values of the OLS-estimates?

3. What will happen if x_t is heteroskedastic?

4. What will happen if $z2_t$ is heteroskedastic?

E *Exercise 3* Bivariate regression analysis and model evaluation

Read the dataset `SP8101`, which contains 81 observations of the variables X and Y. Analyze the properties of the following simple regression model for Y_t given X_t

(4) $\quad Y_t = \alpha + \beta X_t + u_t$

a) What is it reasonable to assume about the errors u_t?.
b) Interpret the results, is it reasonable to conclude that $\beta \neq 1$ on the basis of the estimated results. What is the interpretation of R^2?
c) Is this model an adequate conditional model of Y_t given X_t?
d) What are the criteria we use to evaluate this model?
e) Compare the results with those of the `inverted` model, i.e. X_t given Y_t.

f) Are the results reasonable in light of the evaluation criteria?
g) Estimate the parameters in a more general dynamic model where also Y_{t-1} and X_{t-1} are included as explanatory variables in a conditional model of Y_t. Evaluate the results.

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \gamma_1 Y_{t-1} + u_t \tag{5}$$

h) Consider the three models given by

$$\begin{aligned}
&(6)\quad Model1)\quad Y_t &=& \alpha + \beta_0 X_t + u_t \\
&(7)\quad Model2)\quad Y_t &=& \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \gamma_1 Y_{t-1} + u_t \\
&(8)\quad Model3)\quad Y_t &=& \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \gamma_1 Y_{t-1} + \gamma_2 Y_{t-2} + u_t
\end{aligned}$$

i) Formulate a strategy for evaluating the three models against each other. Explain the difference between "simple-to-general" and "general-to-simple" modeling strategies.

F *Exercise 4* Bivariate regressions with auto-correlated errors

Repeat the analysis in Exercise 3 on the data in `SP8102` and `SP8103` (in addition to `SP8101`. All three datasets contain 81 observations of X and Y. consider the following models:

Model 1: $Y_t = \alpha + \beta X_t + \varepsilon_t$

Model 2: $Y_t = \alpha' + \beta' X_t + \gamma' X_{t-1} + \delta' Y_{t-1} + \varepsilon_t$

Model 4: $Y_t = \alpha + \beta X_t + u_t, \quad u_t = \rho u_{t-1} + \varepsilon_t$

where $\varepsilon_t \sim Niid(0, \sigma^2)$.

We can rewrite model 3 as

$$\begin{aligned}
(Y_t - \alpha - \beta X_t) &= \rho(Y_{t-1} - \alpha - \beta X_{t-1}) + \varepsilon_t \\
Y_t &= \underbrace{(1-\rho)\alpha}_{\alpha'} + \underbrace{\beta}_{\beta'} X_t - \underbrace{\rho\beta}_{-\gamma'} X_{t-1} + \underbrace{\rho}_{\delta'} Y_{t-1} + \varepsilon_t
\end{aligned}$$

Estimate Model 1 and Model 2 using the OLS method and Model 4 using the RALS method. Explain and interpret the results.

G *Exercise 5* Variable transformations

The data on $\{X_t, Y_t\}_{t=1}^{81}$ in the dataset `SP8102` are used in Exercise 4 above. Estimate the following regression model

(9) $Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \gamma_1 Y_{t-1} + u_t$

under the assumption that $u_t \sim N(0, \sigma_u^2)$.

Now, subtract Y_{t-1} from both sides of the equation and add and subtract $\beta_0 X_{t-1}$ on the RHS of the equation. The model can now be written:

(10) $\Delta Y_t = \delta + \phi_0 \Delta X_t + \phi_1 X_{t-1} + \psi_1 Y_{t-1} + u_t$

a) What is the difference between the two models? In particular, what is the relationship between the parameters of the two equations?
b) Estimate the parameters and comment on the results.
c) How should we evaluate these models in light of R^2 and the estimated coefficients?
d) Comment on the problem with multicollinearity in these regressions? (Hint: Calculate the correlation matrix of the RHS variables)

H *Exercise 6* Multicollinearity

Load the datasets `SE10001,SE10002 and SE10003`. All datasets contain 100 observations of the three variables Y, $X1$ and $X2$.

a) For each dataset: Estimate the parameters in the model

(11) $Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + u_t$

under the assumption that $u_t \sim N(0, \sigma_u^2)$.

b) What can we conclude about the magnitude of the parameters β_i, $i = 0 \ldots 2$ in each of the three cases?

Estimate the following model:

(12) $Y_t = \beta_0 + \beta_1 X_{1t} + u_t$

c) We want to test whether this is a valid simplification or not. How can we test this in PcGive?
d) Write the batch-file such that the test is automatically calculated.
e) Interpret and comment the results for each of the three datasets.

I *Oppgave 7* Residual regression.

We use the data set `SE10001` introduced in exercise 6.
Consider the following regression model where we have removed the linear effect from X_{2t} on Y_t and X_{1t} respectively.

$$(13)\quad [Y_t - b_{y2}X_{2t}] = \beta_0 + \beta_1[X_{1t} - b_{12}X_{2t}] + (\beta_2 + \beta_1\frac{m_{12}}{m_{22}} - \frac{m_{y2}}{m_{22}}) + u_t$$

m_{ij} denote empirical moments between variable i and variable j.

Note that the final parenthesis on the right hand side is exactly 0, and β_1 is determined in auxilliary regressions where $b_{y2} = m_{y2}/m_{22}$ and $b_{12} = m_{12}/m_{22}$, i.e., regressing Y_t and X_{1t} on X_{2t} respectively. Then we find the *residuals* in the parenthesis in (13) above, and finally we find the OLS estimator of β_1 as

$$(14)\quad \hat{\beta}_1 = \frac{m_{y1.2}}{m_{11.2}}$$

where $m_{ij.2}$ denote the empirical moment between variable i and variable j where we have removed the effect from X_{2t}. Compare with the results from OLS estimation of the model:

$$(15)\quad y_t = \beta_0 + \beta_1 X_{1t} + \beta_{2t} X_{2t} + u_t$$

J *Exercise 8* More about heteroskedasticity.

The data set `TVV200`, contain 200 observations of the variables Y, $X1$ and $X2$, i.e., a sequence of observations $\{Y_t, X1_t, X2_t\}_{t=1}^{200}$.

a) Estimate first the regression model

(16) $Y_t = \beta_0 + \beta_1 X1_t + \beta_2 X2_t + u_t$

using OLS.

b) Suggest suitable tests which can detect whether model residuals are heteroskedastic.

c) Which conclusions can be drawn with respect to the size of the parameters?

K *Exercise 9* Estimation of product functions (Maddala)

The data set `PRFU` contain 39 observations of the variables YEAR, X, L1, L2, K1 and K2 from table 3.11 i the textbook by Maddala(1989).

a) Estimate the parameters in the following production function

(17) $X_t = \beta_0 + \beta_1 L_t + u_t$

Define $x_t = \ln X_t, \ell_t = \ln L_t$ etc.

b) Estimate the parameters in the log-linear model

(18) $x_t = \beta_0 + \beta_1 \ell_t + u_t$

c) Analyze the model residuals in the two regression models. What can they tell about the model specification?

d) Test whether $k_t = \ln K_t$ should be included in the models, i.e., whether the production function can be written as:

(19) $x_t = \beta_0 + \beta_1 \ell_t + \beta_2 k_t + u_t$

Define labour productivity $xl_t = x_t - \ell_t$ and the the capital-labour ratio $kl_t = k_t - \ell_t$.

e) Estimate the following model with OLS:

(20) $xl_t = \beta_0 + \beta_1 kl_t + u_t$

f) Compare with the results above and interpret the estimated parameters.
g) Which are the (linear) restrictions that the new parameters satisfies?
h) Discuss the necessary exogeneity assumptions.
i) Discuss under which assumptions it may be more appropriate to estimate the following model:

(21)
$$\begin{aligned} \ell_t &= \beta_0 + \beta_1 x_t + u_{1t} \\ k_t &= \alpha_0 + \alpha_1 x_t + u_{2t} \end{aligned}$$

j) How should these two equations be estimated ? (discuss) k) What are the reasonable assumptions to make about the error terms u_{1t}, u_{2t}?

k) Comment on the results obtained from using `OLS` separately on each equation with the results if the two equations are estimated simultaneously using the `SURE` method.

L *Exercise 10* A simple market model.

The datasets `BFM101` and `BFM201` each contain 100 observations of the variables $Y1$ (quantity) and $Y2$ (price) for a consumer good, and contain in addition observations of two variables $X1$ (supply shocks) and $X2$ (demand shocks).
For each of the two datasets the following steps to develop a simple simultaneous system should be undertaken. Compare the results for alternative assumptions about the identification of the parameters in the structural equation:

a) Formulate a simple market model where prices and quantities are simultaneously determined. Specify a supply relationship and a demand relationship for this market. Discuss under which conditions the two relationships are identified.
b) Estimate the parameters in the reduced-form equations (RF) for this model. c) Under the assumption that the parameters in the supply- and demand-relationships (SF) can be identified, estimate the parameters in these relationships using

i) indirect least squares `ILS`,
ii) ordinary least squares `OLS`,
iii) instrumental variables `IV` (discuss the choice of instruments),
iv) two stage least squares `2SLS`,
In each case, discuss whether the estimators are unbiased, consistent and/or efficient. Make the necessary additional assumptions, and discuss whether these assumptions have testable implications.

Supply- and demand relationships (SF)

$$\begin{aligned} p_t &= \phi q_t + \rho s_t + \varepsilon_{1t} \\ q_t &= -\psi p_t + \eta d_t + \varepsilon_{2t} \end{aligned}$$

Price- and quantity equations (RF)

$$\begin{aligned} p_t &= \frac{\varphi\eta}{1+\varphi\psi} d_t + \frac{\rho}{1+\varphi\psi} s_t + \frac{1}{1+\varphi\psi}(\varphi\varepsilon_{2t} + \varepsilon_{1t}) \\ q_t &= \frac{\eta}{1+\varphi\psi} d_t - \frac{\psi\rho}{1+\varphi\psi} s_t + \frac{1}{1+\varphi\psi}(\varepsilon_{2t} - \psi\varepsilon_{1t}) \end{aligned}$$

M Data and sample programs (PcGive *.FL batch files)

```
-------------------------------------------------------------------------
Ex:      Data:                         Program:            Output:
        .WK1 .XLS .IN7 (.BN7)
-------------------------------------------------------------------------
01)     V04INFL  .XLS .IN7             E01_INFL.FL         E01_INFL.OUT
        GREENINV .WK1                  E01_GINV.FL         E01_GINV.OUT
        GREENGAS .WK1
        USCP     .WK1                  E01_USCP.FL         E01_USCP.OUT

02)                                    E02_RANN.FL         E02_RANN.OUT

03-05)  SP8101   .WK1 .XLS             E03_SP01.FL         E03_SP01.OUT
        SP8102   .WK1 .XLS             E03-SP02.FL         E03-SP02.OUT
        SP8103   .WK1 .XLS

06-07)  SE10001  .WK1                  E06_SE  .FL
        SE10002  .WK1                  E07_RERE.FL
        SE10003  .WK1

        SE40001  .WK1
        SE40002  .WK1
        SE40003  .WK1

08)     TVV200   .WK1

09)     PRFU     .WK1 .XLS

10)     BFM101   .WK1 .XLS             E10-BFM1.FL         E10-BFM1.OUT
        BFM201   .WK1 .XLS             E10-BFM2.FL         E10-BFM2.OUT
```