Econometrics 2, Fall 2005

Department of Economics, University of Copenhagen

Heino Bohn Nielsen

PROBLEM SET #3

The first exercise illustrates that the OLS estimate of the coefficient corresponding to a particular explanatory variable in a multiple regression model can also be obtained in a simple regression model where the effects of all other variables have been partialled out. The result is known as the *Frisch-Waugh-Lovell Theorem*.

Exercise 2 introduces a data set for Danish aggregate private consumption and estimates simple time series regressions for the private consumption.

#3.1   THE FRISCH-WAUGH-LOVELL THEOREM

Consider the regression model

$$y = X_1\beta_1 + X_2\beta_2 + \epsilon, \tag{3.1}$$

where X_1 and X_2 are matrices of dimension $N \times K_1$ and $N \times K_2$ respectively, and β_1 and β_2 have conformable dimensions. This can also be written as

$$y = X\beta + \epsilon, \tag{3.2}$$

where $X = (X_1, X_2)$ and $\beta = (\beta_1', \beta_2')'$.

(1) Show that the OLS estimator $\widehat{\beta} = (\widehat{\beta}_1', \widehat{\beta}_2')'$ of β in (3.2) corresponds to the solution to the equations

$$X_1'X_1\widehat{\beta}_1 + X_1'X_2\widehat{\beta}_2 = X_1'y \tag{3.3}$$
$$X_2'X_1\widehat{\beta}_1 + X_2'X_2\widehat{\beta}_2 = X_2'y. \tag{3.4}$$

(2) Solve equation (3.4) for $\widehat{\beta}_2$.

(3) Show that by substituting the obtained $\widehat{\beta}_2$ into (3.3) we obtain the estimator

$$\widehat{\beta}_1 = (X_1'M_2X_1)^{-1}X_1'M_2y, \tag{3.5}$$

where $M_2 = I_N - X_2(X_2'X_2)^{-1}X_2'$.

(4) Show that (3.5) is the OLS estimator in the model

$$M_2 y = (M_2 X_1)\beta_1 + \text{residual}.$$

This is an example of the so-called *Frisch-Waugh-Lovell Theorem*. It implies that the OLS estimate $\widehat{\beta}_1$ of β_1 in (3.1) can be obtained in a sequential way:

(a) First regress y on X_2 and obtain the residual $M_2 y$. [Note: This is a vector of dimension N].

(b) Then regress X_1 on X_2 and obtain the residuals $M_2 X_1$ [Note: This is a matrix of dimension $N \times K_1$].

(c) Finally regress the residual $M_2 y$ on the residual $M_2 X_1$ to get the estimate $\widehat{\beta}_1$.

This result is very useful in theoretical and empirical work, as it allows the correlation of interest, e.g. between y and X_1 to be analyzed conditional on the effect of X_2. A simple application of this idea is the case where X_2 is a constant term and the concentrated regression involves deviations from sample means.

#3.2   TIME SERIES REGRESSIONS FOR PRIVATE CONSUMPTION

In this exercise we consider a data set for aggregate private consumption in Denmark for the period $1971 : 1 - 2003 : 2$. The data file `ConsumptionData.In7` contains observations for the variables

FCP	Private sector aggregate consumption, constant prices.
PCP	Deflator for private consumption, 1995=100.
FYDP	Private disposable income, constant prices.
REALFOR	Private wealth including owner occupied housing, constant prices.
ARBLOS	Expected income loss from changes in unemployment.
IBZ	Average bond rate, fractions, p.a.

All the variables, except the interest rate, are seasonally adjusted and are taken from the data base from the model MONA of the Danish Central Bank.

(1) Read the data into GiveWin and construct the derived variables

$$
\begin{aligned}
c_t &= \log(\mathsf{FCP}_t) \\
y_t &= \log(\mathsf{FYDP}_t) \\
w_t &= \log(\mathsf{REALFOR}_t) \\
p_t &= \log(\mathsf{PCP}_t) \\
\pi_t &= \Delta p_t = p_t - p_{t-1} \\
r_t &= \frac{1}{4} \cdot \mathsf{IBZ}_t
\end{aligned}
$$

Draw time series graphs of the variables: c_t, y_t, w_t, π_t, and ARBLOS_t; and discuss the economic development in Denmark over the period.

From a graphical inspection, do any of the variables appear stationary?

(2) Consider the following transformed variables

$$
\begin{aligned}
\Delta c_t &= c_t - c_{t-1} \\
\Delta y_t &= y_t - y_{t-1} \\
\Delta w_t &= w_t - w_{t-1} \\
\Delta \pi_t &= \pi_t - \pi_{t-1} \\
\mathsf{ECM}_t &= c_t - 0.536566 - 0.263579 \cdot y_t - 0.482082 \cdot w_t + 2.12642 \cdot \Delta p_t.
\end{aligned}
$$

Draw time series graphs of the variables Δc_t, Δy_t, Δw_t, $\Delta \pi_t$, and ECM_t.

From a graphical inspection, do the transformed variables appear stationary?

[Hint: The variable ECM_t can be thought of as the deviation of consumption, c_t, from the equilibrium value in period t and it is formally constructed using cointegration. We will return to this issue later in the course, for now just think of ECM_t as a particular explanatory variable.]

(3) In the construction of time series regression models, why is it important that the included variables are stationary?

(4) When you construct an empirical model for the consumption data, do you think it is preferable to start with a simple model and then successively include more explanatory variables; or do think it is preferable to begin with a general model and delete insignificant variables? Motivate your answer.

(5) Start PcGive. Choose [Econometric Modelling] in the [Packages] menu and [Single Equation Dynamic Modelling] in the [Model] menu.

Consider the linear regression model

$$
\begin{aligned}
\Delta c_t &= \beta_1 + \beta_2 \cdot \Delta c_{t-1} + \beta_3 \cdot \Delta y_t + \beta_4 \cdot \Delta y_{t-1} + \beta_5 \cdot \Delta w_t + \beta_6 \cdot \Delta w_{t-1} + \beta_7 \cdot \Delta \pi_t \\
&\quad + \beta_8 \cdot \Delta \pi_{t-1} + \beta_9 \cdot \mathsf{ARBLOS}_t + \beta_{10} \cdot \mathsf{ARBLOS}_{t-1} + \beta_{11} \cdot \mathsf{ECM}_{t-1} + \epsilon_t,
\end{aligned}
$$

for $t = 1973 : 3 - 2003 : 2$. Interpret the signs and magnitudes of the coefficients.

(6) Choose [Graphic Analysis] in the [Test] menu, and try the possibilities. Do the residuals look well behaved?

(7) Use PcGive to calculate the LM test for no first order autocorrelation in the residuals. This is done by selecting [Test...] in the [Test] menu and specifying the Error Autocorrelation test from lag 1 to lag 1. What is the distribution of the test statistic, and what do you conclude?

Save the estimated residuals $\widehat{\epsilon}_t$ from the original regression and construct manually the LM test. [Hint: Remember to insert zero values for the residuals in the beginning of the sample and use the same estimation sample for the auxiliary regression that you used for the original regression. You should get exactly the same test statistic as the one reported by PcGive.]

(8) Calculate more misspecification tests using [Test Summary] in the [Test] menu.
Discuss how the tests are constructed and how they should be interpreted.
What do you conclude regarding the specification of the model.

(9) Now simplify the model by removing regressors with insignificant coefficients. Begin
by deleting the variable with smallest $t-$ratio and continue until all coefficients are
significant. *Always retain the constant term.*
Recalculate the misspecification tests.

(10) How do you think that the interest rate should affect consumption?
Analyze whether there is a role for the bond rate, r_t, or the real bond rate, $r_t - \pi_t$,
in your empirical model.

(11) Now reestimate your preferred model for $t = 1973 : 1 - 1998 : 4$ and retain the last
18 observations for post-sample predictions.
To evaluate the post-sample prediction performance of the regression model, consider
the one-step ahead prediction errors. They are defined as the out-of-sample predicted
values from the estimated equations, when the actual values from the database are
used for Δc_{t-1} in the calculation of the forecasts.
What do you conclude regarding the ability of the model to forecast Danish aggregate
consumption for the recent years?
[Hint: In PcGive, estimate the model for the full sample, but save 18 observations
for forecasts. Then select [Forecast] in the [Test] menu.]