Eryk Wdowiak
economics@doviak.net

Introductory Statistics

Homework #7

This homework introduces you to time-series analysis and to show you that – when working with time series – you must make sure that serial correlation does not affect your estimates.

♦ ♦ ♦

First, we need to understand the problems associated with serial correlation, so load the dataset of GDP per worker from the Penn World Table (i.e. "pwt-4countries.csv") into Gretl.

1. Compute the log of GDP per worker in France ("FRA"), Germany ("GER"), Morocco ("MAR") and Nepal ("NPL") by going to Gretl's "Add" menu and selecting "Logs of selected variables." The new series will appear at the bottom of the list of variables.

2. Obtain the correlation matrix for log of GDP per worker in France ("l_FRA"), Germany ("l_GER"), Morocco ("l_MAR") and Nepal ("l_NPL"). Why are the correlation coefficients so high?

3. Create time-series plots of log of GDP per worker in France, Germany, Morocco and Nepal by go to Gretl's "View" menu, selecting "Multiple graphs" and then selecting "Time series." Do these plots help explain why the correlation coefficients are so high?

When the difference between two values is small, the *log difference* between those two values is approximately equal to the percentage change. For example, German GDP per worker was $67,075 in 2008 and $63,904 in 2009, so the percentage change from 2008 to 2009 was –4.7 percent. Taking the natural log of both values and computing the difference between the two yields –0.048, or –4.8 percent, which is approximately equal to the percentage change.

4. Compute the log difference of GDP per worker in France, Germany, Morocco and Nepal by going to the "Add" menu and selecting "Log differences of selected variables." The new series will appear at the bottom of the list of variables.

5. Obtain the correlation matrix for log differences of GDP per worker in France ("ld_FRA"), Germany ("ld_GER"), Morocco ("ld_MAR") and Nepal ("ld_NPL"). Why are the correlation coefficients of the log differenced variables so much lower than the correlation coefficients of the log variables?

6. Create time-series plots of the log difference in GDP per worker in France, Germany, Morocco and Nepal by going to the "View" menu, selecting "Graph specified vars" and then selecting "Time series plot." Plot Germany and France on one graph and plot Morocco and Nepal on another. Do these plots help explain why the correlation coefficients of the log differenced variables are so much lower?

continued on next page

Now, let's look at the US labor market data. So load the “usa_labor-mkt.csv” dataset into Gretl.

7. Make sure Gretl gives the data a time-series interpretation. To do that, go to the “Data” menu and select “Dataset structure.” Select “time-series,” then select “monthly” and select Jan. 1960 as the first month (i.e. “1960:01”).

8. Next, we must prevent serial correlation from affecting our estimates, so:
   a. compute the log difference of: nonfarm employment (“NonfarmEmp”), average hours worked in manufacturing (“HoursManuf”), unfilled orders for durable goods (“DurGoodsUnfillO”) and the consumer price index (“CPIu”).
   b. compute the first difference of the unemployment rate (“UnempRate”) and the coincident index (“Coincident”).

9. Obtain the correlation matrix for the differenced and log differenced variables (i.e.: “d_Coincident,” “d_UnempRate,” “ld_NonfarmEmp,” “ld_HoursManuf,” “ld_DurGoodsUn” and “ld_CPIu”).

10. Which of those variables are highly correlated with each other? which are weakly correlated?