Eric Doviak

Introduction to Economic and Business Statistics

Homework #6

This homework introduces you to time-series analysis and VAR modelling. It isn't much different from the regression analysis that we studied before, but – when working with time series – you must make sure that serial correlation does not affect your estimates.

The purpose of this exercise is to introduce you to time-series analysis and to show you how the assumptions of your model can affect your results.

This homework makes use of Mark Thoma's description of Christopher Sims' work (on VAR modelling) and the paper that Sean MacDonald and I wrote on "Terraced VARs." (You should recall these papers from homework #4).

♦ ♦ ♦

Before jumping into VARs, we first 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?

♦ ♦ ♦

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. Given your reading of our "Terraced VARs" paper, why do you think the changes in the coincident index are so highly correlated with the change in the unemployment rate and with the log difference in nonfarm employment?

♦ ♦ ♦

Suppose that we're interested in the relationship between the coincident index and nonfarm employment. Given the fact that nonfarm employment is used to construct the coincident index, changes in nonfarm employment should affect the coincident index. And since the coincident index is designed to capture the underlying state of the economy, the changes in the coincident index should affect nonfarm employment.

Suppose further that we have the following model:

$$NF_t = -a_0 \cdot CI_t + a_1 \cdot NF_{t-1} + a_2 \cdot CI_{t-1} + u_t$$
$$CI_t = -b_0 \cdot NF_t + b_1 \cdot NF_{t-1} + b_2 \cdot CI_{t-1} + v_t$$

Note that if we assume that $a_0 \neq 0$ and $b_0 = 0$, then we are assuming that changes in the coincident index immediately affect nonfarm employment, but changes in nonfarm employment do not immediately affect the coincident index.

Conversely, if we assume that $a_0 = 0$ and $b_0 \neq 0$, then we are assuming that changes in nonfarm employment immediately affect the coincident index, but changes in the coincident index do not immediately affect nonfarm employment.

11. Which assumption do you think is better? To answer the question, create a time-series plot of the changes in the coincident index and the log difference in nonfarm employment. Which variable appears to predict movements in the other?

12. Note however that we cannot estimate both equations in their current form because one is simply a linear transformation of the other. So solve the model for NF_t and CI_t.

13. The solution that you just obtained provides a reduced form model that we can estimate. Estimate the model in Gretl by going to the “Model” menu, selecting “Time series” and then selecting “Vector Autoregression.”

    Notice that there is more than one endogenous variable in a VAR. Specifically, there are two in this case – the log difference in nonfarm employment and the change in the coincident index – so place them both in the endogenous variable field.

14. Notice also that Gretl provides several options. It asks for the lag order (i.e. the number of past values to include in the model). It asks if you want to include a constant term in the model. It asks if you want to include a trend, etc.

    What is the effect of increasing the lag order? What is the effect of adding a constant term to the model? What is the effect of adding a trend?

15. For simplicity, estimate a VAR of “ld_Nonfarm” and “d_Coincident” with two lags, no constant and no trend. Notice that the VAR you just ran consists of two regression models.

16. Why does the absence of a constant and the absence of a trend imply that the changes in the coincident index and nonfarm employment will go to zero in the long run?

17. Suppose that we're in the long run. There is no change in either the coincident index or nonfarm employment. Now suppose that a shock suddenly reduces the value of the coincident index. If we assume that $a_0 = 0$, then the shock to “d_Coincident” has no immediate effect on “ld_Nonfarm.” Conversely, if we assume that $a_0 \neq 0$, then “d_Coincident” will immediately affect “ld_Nonfarm.”

    Using the model above, trace out this effect in the immediate period and in the subsequent period. This is called an “impulse response.”

♦ ♦ ♦

It is very important whether we assumed that $a_0 = 0$ or that $b_0 = 0$. Unfortunately, Gretl makes this distinction in a vague way. When you go to the “Graphs” menu (in the window that contains the VAR regression results) and select “Impulse responses (combined),” Gretl asks for the “Cholesky ordering.”

If “ld_Nonfarm” is listed first in the ordering, then Gretl assumes that $a_0 = 0$ and $b_0 \neq 0$, so that “d_Coincident” has no immediate effect on “ld_Nonfarm.” Conversely, if “d_Coincident” is listed first, then Gretl assumes that $a_0 \neq 0$ and $b_0 = 0$, so “ld_Nonfarm” has no immediate effect on “d_Coincident.”

18. So run the create the impulse responses under both assumptions. Notice how the impulse responses depend on whether it is assumed that $a_0 = 0$ or it is assumed that $b_0 = 0$.

19. What have you learned about how your assumptions affect your results?