Serial Correlation

Our original regression model assumed that the error terms were independent of one another.

In most applications involving time series data- e.g. macroeconomic and financial variables, this assumption is no longer valid.

This is because this data is often of a "cyclical" nature.

When errors associated with observations of different time periods are related to each other, we refer to the errors as being **serially correlated**.

Most often it's the case that errors associated with adjacent observations are correlated, and errors for observations which are "far apart" are not. Thus serial correlation is only a concern for data sets exhibiting a natural ordering of some sort.

We will focus on a particular type of serial correlation, referred to as **first-order serial correlation**. This basically means that errors in one time period are directly correlated with errors in the next time period.

In most applications, this type of serial correlation will be positive, and that's the case that we'll be covering.

Our discussion of serial correlation will be completely parallel to our discussion of heteroscedasticity. Specifically, we will aim to answer the following 3 questions?

1. What are the consequences of serial correlation in my data?

2. How can I correct my estimation procedure to allow for serial correlation?

3. How can I test for the presence of serial correlation in my data set?

Effects of Serial Correlation on Least Squares Estimates

As was the case with heteroscedasticity, if serial correlation is present in our data, our least squares estimator will still be unbiased, but no longer B.L.U.E. Moreover, in the case of positive serial correlation, our estimates of the standard errors will be lower than they should be (i.e. they will be biased downward.)

This will mean our confidence intervals are too small, and we will sometimes reject our null hypothesis when we shouldn't.

Finally, our value of R^2 will be higher than it should be, and our estimator of the error variance will be smaller than it should be.

Correction Procedure for Serially Correlated Data

Consider the following multiple regression model which allows for first order serial correlation:

$$y_t = b_1 + b_2 x_{2t} + \ldots b_k x_{kt} + e_t$$

$$e_t = r e_{t-1} + n_t \qquad 0 \leq | r | < 1$$

where we assume:

$$n_t \sim N(0, s^2{}_n)$$

$$e_t \sim N(0, s^2{}_e)$$

and that n_t is independent of e_t but e_t is not independent of other error terms, such as e_{t-1}.

In other words the value of the error term in the current period is determined by reducing the value of the error in the previous period and adding the effect of a different mean 0 random variable.

This type of error behavior is referred to as a **first order autoregressive process.**

How does the value of the error in the current time period affect the error value in other periods?

Note it can be easily shown that:

$$Var(e_t) = s^2{}_e = \frac{s^2{}_n}{1 - r^2}$$

and the correlation between error terms peters out as the time periods get farther apart:

$$Cov(e_t, e_{t-1}) = r s^2_e$$

$$Cov(e_t, e_{t-2}) = r^2 s^2_e$$

$$Cov(e_t, e_{t-3}) = r^3 s^2_e$$

Now suppose r were known to us. There is a simple procedure for transforming the data so serial correlation is no longer present.

Consider an observation of our model in the *previous* time period:

$$y_{t-1} = b_1 + b_2 x_{2t-1} + \ldots b_k x_{kt-1} + e_{t-1}$$

$$y_t = b_1 + b_2 x_{2t} + \ldots b_k x_{kt} + e_t$$

and multiply this equation by r :

$$ry_{t-1} = rb_1 + b_2 rx_{2t-1} + \ldots b_k rx_{kt-1} + re_{t-1}$$

now subtract this equation from our original equation observed in time period *t*. We now have the equation:

$$y_t^* = b_1(1-r) + b_2 x^*_{2t} + \ldots b_k x^*_{kt} + n_t$$

where:

$$y_t^* = y_t - ry_{t-1}$$

$$x_{jt}^* = x_{jt} - rx_{jt-1} \qquad j = 2,3,\ldots k$$

$$n_t = e_t - re_{t-1}$$

Note for our transformed model, the error term n_t is serially uncorrelatedso the assumptions of the Gauss Markov theorem are once again satisfied. Thus if we run least squares on the transformed model, our estimates will be B.L.U.E.

Cochrane-Orcutt Procedure

$$y_t^* = y_t - r y_{t-1}$$

Note the transformation we just talked about usually can't be applied in practice, since we rarely know what the value of r .

However, we can employ the following iterative procedure:

1. Run ordinary least squares, and get residual values $\hat{e}_t$.

2. With these residual values run the regression:

   $$\hat{e}_t = r\hat{e}_{t-1} + n_t$$

   and get the least squares estimate $\hat{r}$.

3. Using the value of $\hat{r}$, transform the data the way we just discussed. Run least squares on the transformed data.

4. Go back to 2.

If our sample size is large, our estimates using this procedure should be close to the estimates we would obtain if r were known.

Test for Serial Correlation

A very popular test for serial correlation is the **Durbin-Watson** test. It's based on the intuitive idea that if there is first order serial correlation, the sample correlation between the residual in period *t* and period *t*-1 should be far away from 0.

The value of the Durbin Watson statistic is:

$$DW = 2(1 - \hat{r})$$

where $\hat{r}$ is define as before- the regression coefficient when we regress $\hat{e}_t$ on $\hat{e}_{t-1}$.

To do the test, however, we have to look up critical values on a new table (page 610).

Value of DW	**Conclusion**
$4 - d_l < DW < 4$	Reject null in favor of neg. ser. corr.
$4 - d_u < DW < 4 - d_l$	Indeterminate
$2 < DW < 4 - d_u$	Accept null
$d_u < DW < 2$	Accept null
$d_l < DW < d_u$	Indeterminate
$0 < DW < d_l$	Reject null in favor of pos. ser. corr.