Autocorrelation

This implies that taking the time series regression

$$Y_t = \alpha + \beta X_t + u_t$$

but in this case there is some relation between the error terms across observations.

$$E(u_t) = 0$$
$$E(u_t^2) = \sigma^2$$
$$E(u_s u_p) \neq 0$$

- Thus the error covariances are not zero.
- This mean that one of the assumption that makes OLS BLU does not hold.

Likely causes:

1. Omit variable that ought to be included.
   The errors will then contain the trended variable
   So say the true model is
   $$Y_t = \alpha_0 + \alpha_1 X_t + \alpha_2 Z_t + u_t$$
   $$Y_t = \alpha_0 + \alpha_1 X_t + \alpha_2 Z_t + u_t$$
   and you estimate
   $$Y_t = \beta_0 + \beta_1 X_t + w_t$$
   then
   $$w_t = \alpha_2 Z_t + u_t$$
   If the omitted variable is uncorrleated with the included independent variable then can use the methods to account for serial correlation that we discuss below but if this is not the case, say $Z_t$was X_t^2 then the estimates are both inconsistent and inefficient.
2. Mispecification of the functional form. This is most obvious where a straight line is put through a curve of dots. This would clearly show up in plots of residuals.
3. Errors of measurement in the dependent variable. If the errors are not random then the error term will pick up any sytematic mistakes.

The Problem

OLS is not the best estimation method.

- It will underestimate the true variance.
- the t values will look too good
- will reject H_0 when it is true

So estimates will be unbiased but inefficient, but there can be considerable understatement of the sampling variances and the R^2 as well as the t and F tests will tend to be exaggerated

Focus on simplest form of relation over time: first order autocorrelation which can be written as

$$u_t = \rho u_{t-1} + \varepsilon_t$$

Obviously there could be more complicated forms.

Tests

1. Plot the residuals over time or against a particular variable and see if there is a pattern.

little change of sign $\Rightarrow$ positive autocorrelation

$$u_t = ue_{t-1} + \varepsilon_t$$

expect if high value then next one will be high as well with extreme where $\rho = 1$

2. Durbin Watson Statistic: commonly used

$$DW = \frac{\sum(\hat{u}_t - \hat{u}_{t-1})^2}{\sum \hat{u}_t^2}$$

$$= \frac{\sum \hat{u}_t^2 + \sum \hat{u}_{t-1}^2 - 2\sum \hat{u}_t \hat{u}_{t-1}}{\sum \hat{u}_t^2}$$

Now when the number of observations is very large $\sum \hat{u}_t^2$ and $\sum u_{t-1}^2$ will be almost the same, so

$$= \frac{2\sum \hat{u}_t^2}{\sum \hat{u}_t^2} - \frac{\sum \hat{u}_t \hat{u}_{t-1}}{\sum \hat{u}_t^2}$$

$$= 2\left(1 - \frac{\sum \hat{u}_t \hat{u}_{t-1}}{\sum \hat{u}_t^2}\right)$$

$$= 2(1 - \hat{\rho})$$

so we have

$$DW \approx 2(1 - \hat{\rho})$$

- if strong positive autocorrelation then $\hat{\rho} = 1$ and $DW = 0$
- if strong negative autocorrelation then $\hat{\rho} = -1$ and $DW = 4$
- if no autocorrelation then $\hat{\rho} = 0$ and $DW = 2$

  So the best can hope for is a DW of 2

  But sampling distribution of the DW depends on the values of the explanatory variables and so can only derive upper and lower limits

-
  - $DW < DW_L$ reject hypothesis no autocorrelation
  - $DW > DW_U$ don't reject
  - $DW_L < DW < DW_U$ inconclusive

- increasing the observations shrinks the indeterminacy region
- increasing variables increases the indeterminacy region
- Rule of thumb: lower limit for positive autocorrelation = 1.6

Note

1. Only works for first order
   - Higher order tests are available
   - Can't use with LDV -can use Durbin's h -biased towards 2, but extent of bias reduced as add more variables.
   - If have LDV and DW not close to 2 have problem -likely higher order autocorrelation
   - LM test in Microfit works for higher order
2. DW test can also be considered a general mispecification test if there is no autocorrelation
   - if correctly specified but autocorrelated DW will pick up the autocorrelation
   - if incorrectly specified and no autocorrelation DW will pick up the mispecification

Solutions

- Find cause
- increase number of observations
- find missing values
- specify correctly

If know how autocorrelated then can use quasi differencing to solve

$$Y_t = \beta_0 + \beta_1 X_t + u_t$$

and

$$u_t = \rho u_{t-1} + \varepsilon_t$$

then

$$Y_{t-1} = \beta_0 + \beta_1 X_{t-1} + u_{t-1}$$
$$\rho Y_{t-1} = \beta_0 \rho + \beta_1 \rho X_{t-1} + \rho u_{t-1}$$

and taking the equation for ρY_{t-1} from that for Y_t giving

$$Y_t - \rho Y_{t-1} = \beta_0(1-\rho) + \beta_1(X_t - \rho X_{t-1}) + (u_t - \rho u_{t-1})$$

So could estimate a transformed regression

$$Y_t^* = \beta_0 + \beta_1 X_t^* + w_t$$

This would give an unbiased estimator of β which would be better than the OLS.

- Note that the procedure will not use all of the data because of the lag.
- To use all observations there are techniques that replace the first observation: Generalised Least Squares

In practise don't know what the autocorrelation is as have to estimate it

Microfit provides a number of ways of doing this, the best known is the Cochrane Orcutt iterative method

1. Estimate by OLS:

$$Y_t = \beta_0 + \beta_1 X_t + u_t$$

assuming

$$u_t = \rho u_{t-1} + \varepsilon_t$$

2. Take the estimated residuals $\hat{u}_t$ and compute an estimate of ρ

$$\hat{\rho} = \frac{\sum \hat{u}_t \hat{u}_{t-1}}{\sum \hat{u}_t^2}$$

(Can use $\hat{\rho} = 1 - \frac{DW}{2}$ in some cases)

3. Continue to do this until it converges on a value
4. Can generalise to more than a first order process
   - Problems of multiple solutions led to development of other -grid search- methods: won't consider here
   - Durbin's preferred method, especially when there are a number of explanatory variables was to estimate

$$Y_t = \gamma_0 + \gamma_1 Y_{t-1} + \gamma_2 X_t + \gamma_3 X_{t-1} + \varepsilon_t$$

This is because if we take the static model with autoregressive errors:

$$Y_t = \beta_0 + \beta_1 X_t + u_t$$
$$u_t = \rho u_{t-1} + \varepsilon_t$$

substituting for u_t in the first part gives

$$Y_t = \beta_0 + \beta_1 X_t + \rho u_{t-1} + \varepsilon_t$$
$$as\ :\ u_t = Y_t - \beta_0 - \beta_1 X_t \Rightarrow u_{t-1} = Y_{t-1} - \beta_0 - \beta_1 X_{t-1}$$
$$Y_t = \beta_0 + \beta_1 X_t + \rho(Y_{t-1} - \beta_0 - \beta_1 X_{t-1}) + \varepsilon_t$$

multiplying out and rearranging gives

1.

$$Y_t = \beta_0(1-\rho) + \rho Y_{t-1} + \beta_1 X_t - \beta_1 \rho X_{t-1} + e_t$$

which is the estimating equation above where so the coefficient on Y_{t-1} is then an estimate of ρ

- This estimate $\hat{\rho}$ can then be used in the quasi-first difference regression.
- The standard errors of the transformed equation are only correct asymptotically.
- Note that with a LDV in the original equation this is not true.

Should consider

1. In small samples the fact that you have to estimate $\hat{\rho}$ may mean that you don't really gain anything in terms of the mean square error (MSE). Have to consider bias and efficiency
2. Have assumed first order but may be more complex and first order may not be the best way

to start off

3. Need to take care using quarterly data
4. Most important: It is easy to confuse mispecified dynamics with serial correlation in the errors.
5. In fact it is best to always start from a general dynamic models and test the restrictions before applying the tests for serial correlation.
   - The AR(1) is only one possible dynamic model, so to start from static and then use C-O if there are problems with the DW statistic is not a good idea. It is the way a lot of applied work was done years ago.
   - The general to specific methodology became the way to do applied work and it is the way you will be doing the applied exercises