Lecture 20

Serial Correlation

The following are some sketchy notes about serial correlation.

The Durbin-Watson statistic

The Durbin-Watson statistic is the standard method for detecting serial correlation in regression data. Positive serial correlation means that positive residuals tend to lead to positive adjacent residuals and negative residuals tend to lead to negative adjacent residuals. Adjacent here means in the order in which the data was collected, usually a time order. If the data is sorted by predictor variable, positive serial correlation could be a result of curvature in the model. It is not, however, the appropriate test for curvature. You should only worry about positive serial correlation when your data was collected in order and you know that order.

The statistic is:

$$d = \frac{\sum_{t=2}^{n} \left(\hat{\epsilon}_t - \hat{\epsilon}_{t-1}\right)^2}{\sum_{t=1}^{n} \hat{\epsilon}_t^2}$$

The expected value of d when there is no correlation amongst the residuals is 2. If d is substantially lower than 2, that is a sign of positive serial correlation - the most likely possibility and the one with the most serious consequences. If you do not correct for positive serial correlation, you are stating more confidence in your results than you really have. This is called anti-conservatism in statistics and statisticians regard it as a serious error. If d is substantially higher than 2, that is a sign of negative serial correlation, but this is a rarer and less of a problem.

The tables for determining whether a particular value of the Durbin-Watson statistic is statistically significant or not is located at: http://www.jstor.org/stable/2332325

If $d < d_L$, then you have significant positive serial correlation. If $d > d_U$, there is no evidence of positive serial correlation. If $d_L < d < d_U$ you are in a grey area.

References and Readings

Exercises for Lecture 20

1. –

2. –