Problems with OLS

Considering :

$$Y_i = \alpha + \beta X_i + u_i$$

we assume

$$E(u_i) = 0$$
$$E(u_i^2) = \sigma^2 \text{ or } var(u_i) = \sigma^2$$
$$E(u_i u_j) = 0 \text{ or } cov(u_i, u_j) = 0$$

We have seen that we have to make very specific assumptions about u_i in order to get OLS estimates with the desirable properties.

If these assumptions don't hold than the OLS estimators are not necessarily BLU.

We can respond to such problems by changing specification and/or changing the method of estimation.

First we consider the problems that might occur and what they imply. In all of these we are basically looking at the residuals to see if they are random.

- 1. The errors are serially dependent ⇒ autocorrelation/serial correlation.
  2. The error variances are not constant ⇒ heteroscedasticity
  3. In multivariate analysis two or more of the independent variables are closely correlated ⇒ multicollinearity
  4. The function is non-linear
  5. There are problems of outliers or extreme values -but what are outliers?
  6. There are problems of missing variables ⇒can lead to missing variable bias

Of course these problems do not have to come separately, nor are they likely to

- Note that in terms of significance things may look OK and even the $R^2$the regression may not look that bad.
- Really want to be able to identify a misleading regression that you may take seriously when you should not.
- The tests in Microfit cover many of the above concerns, but you should always plot the residuals and look at them.

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.
- Means that one of the assumption that makes OLS BLU does not hold.

Likely causes:

1. Omit variable that ought to be included.
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 (not least variance)

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
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

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

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
- Durbin's h used if LDV
- LM test in Microfit works for higher order
- DW test can also be considered a general mispecification test if there is no autocorrelation

Solutions

- Find cause
- increase number of observations
- find missing values
- specify correctly
- Microfit provides number of procedures eg Cochrane Orcutt -last resort
- Most important: It is easy to confuse mispecified dynamics with serial correlation in the errors.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,

Heteroscedasticity

In this case

$$Y_i = \alpha + \beta X_i + u_i$$

we assume

$$E(u_i) = 0$$
$$E(u_i^2) = \sigma_i^2$$
$$E(u_i u_j) = 0$$

So in this case the errors do not have a common variance.

The effect of this will be to leave the OLS estimator of β unbiased, but the estimated standard error will be biased.

- Number of tests available including that reported in Microfit
- Solution: Correct the standard errors, using the White Heteroscedastic robust errors available in Microfit.

Multicollinearity

- When we have more than one explanatory variable there is a possible problem.
- There may be two or more variables that explain the dependent variable well, but they may be closely correlated.
- This could mean that is difficult to distinguish the individual effects of both variables.

Problem and identification

Have considered perfect multicollinearity in exercise

In practise unlikely to find this extreme , but may get close to it.

There has to be some multicollinearity, so the question is identifying when it is important and so when it is a problem.

- What will tend to get is high $R^2$and F test statistics, but low individual significance of the individual coefficients.
- Parameter estimates become very sensitive to the addition or deletion of observations. So can test by dropping observations and seeing what happens
- Predictions will be worse than those of a model with only a subset of variables
- Standard errors of the regression coefficients will be high. Though in fact this is not necessarily the result of multicollinearity alone.

Cures

- Get more data. But what is important is not the number of observations but the informational content.
- Drop variables:may work in some cases where not interested in individual parameter values, but there is a problem of omitted variable bias.
- Simply present the OLS estimates and the variance-covariance matrix to allow an assessment of multicollinearity.
- Rather than the zero restriction try some others across the variables.
- Use extraneous estimates: e.g. from cross section estimates
- Transform data: Using ratios or first differences will get rid of any multicollinearity caused by common trend. But see discussion of dynamic models.
- Principal components and ridge regression

So no easy solution to multicollinearity, but have to be aware of the problem when dropping seemingly insignificant variables from the regression. Consider:

	Xs important	Xs not important
Include Xs	✓	inefficiency
Exclude Xs	specification bias	✓

Functional form

When estimate models often don’t have information from theory on functional form.

- Use of linear is for simplicity of estimation
- Note that it is linear in the parameters that allows estimation by OLS not linearity in the

variables

- There are methods of estimating non linear relations, but they are not straightforward so we try to find a way of transforming relation to make it linear in the parameters
- There are a wide variety of functional forms that can be approximated using the powers of the independent variables and applying OLS -though problems of multicollinearity possible

$$Y_i = \alpha_0 + \alpha_1 X_i + \alpha_2 X_i^2$$

- Can also use logs or reciprocal transformations

$$Y_i = \alpha_0 + \alpha_1 X_i$$

$$\ln Y_i = \alpha_0 + \alpha_1 \ln X_i$$

$$\ln Y_i = \alpha_0 + \alpha_1 X_i$$

$$Y_i = \alpha_0 + \alpha_1 \ln X_i$$

$$Y_i = \alpha_0 + \alpha_1 (1/X_i)$$

$$\ln Y_i = \alpha_0 + \alpha_1 (1/X_i)$$

- A common example is the Cobb Douglas production function which is multiplicative:

$$Q = AK^{\beta_1} L^{\beta_2}$$

  - to estimate we take logs and add an error

$$\log Q = \log A + \beta_1 \log K + \beta_2 \log L + u$$

  - and then rewrite

$$\log Q = \beta_0 + \beta_1 \log K + \beta_2 \log L + u$$

  - Note the error could have been assumed as $\exp(u)$ in the initial form .
- Using logs has another advantage in that it also give us the elasticities. Consider a demand function:

$$Q_t = \beta_0 + \beta_1 P_t + u_t$$

  - Now β_1 is the slope coefficient

$$\frac{\delta Q}{\delta P} = \beta_1$$

  but it is not the elasticity
  - Instead if we had taken logs

$$\log Q_t = \alpha_o + \alpha_1 \log P_t + e_t$$

  - then α_1 is the elasticity as

$$\frac{\delta \log Q_t}{\delta \log P_t} = \frac{\delta Q_t}{\delta P_t} \bullet \frac{P_t}{Q_t}$$

- Also the difference in logs approximates the growth rate

$$\Delta \log Y_t = \log Y_t - \log Y_{t-1} \approx \frac{Y_t - Y_{t-1}}{Y_{t-1}}$$

- Sometimes theory gives information on possible functional form but it is unusual for it to be specific enough

- May be able to compare using R^2as a guide or SER

  remember cant use R^2 to compares different dependent variables, such as linear vs log linear
- Microfit gives the Reset test which is designed to check the choice of an inappropriate functional form.

Outliers

Problem

Regression parameters can be influenced by a few extreme values or outliers

- Should be able to spot from a careful analysis of the residuals $\hat{u}_t$
- In the case of a simple bivariate regression you can simply plot the data.
- Outlier is an observation that is very different: usually generated by some unusual factor
- Least squares estimates are very sensitive to outliers, particularly in small samples
- Maddala P89-90 gives examples of data sets that when plotted look very different, but give the same regression results. In two cases this is caused by a single extreme value.

Actions

- Drop the observations with large residuals and reestimate the equation. This should really be a last resort
- The outliers may provide important information. They may not be outliers at all. An example of this is the relation between infant mortality and GDP per capita in Asian countries.
- For cross section should maybe try to get more data rather than drop observations. Also for time series.
- Problem of what is an outlier also relates to leverage: need variation in the data or cant estimate any relationship. Its not always obvious when information on a system becomes an outlier.
- Can treat extreme observations with dummy variables

Omitted variable bias

If we miss out an important variable it not only means our model is poorly specified it also means that any estimated parameters are likely to be biased.

- Incorrect omission of variables leads to biased estimates of the parameters that are included
- Incorrect inclusion only produces inefficient estimates, so don’t have minimum variance
- So better to include the wrong variables rather than exclude the right ones.

Dynamic Models

1. The problem with specifying the dynamic form of a regression model is that normally the theory provides little information on lag lengths, nature of adjustments etc. So seems better to use the theory to specify the variables to be included, but to allow the data to determine what the dynamic model should look like
   a. Consider the consumption function being used in the exercises:

$$C_t = \beta_0 + \beta_1 Y_t + \beta_2 Y_{t-1} + \beta_3 C_{t-1} + u_t$$

   This encompasses a number of different models all with different dynamic structures they are 'nested' in this model meaning the restrictions on the parameters can be written down and they can be tested.
2. Static model: absolute income hypothesis $\beta_2 = \beta_3 = 0$

$$C_t = \alpha_0 + \alpha_1 Y_t + v_{1t}$$

3. AR(1) model: first order autoregression $\beta_1 = \beta_2 = 0$

$$C_t = \alpha_0 + \alpha_1 C_{t-1} + v_{2t}$$

4. Partial adjustment/habit persistence model $\beta_2 = 0$

$$C_t = \alpha_0 + \alpha_1 Y_t + \alpha_2 C_{t-1} + v_{3t}$$

   This is a habit model if $\alpha_2 > 0$ and a partial adjustment model if $|\alpha_2| < 1$
5. Distributed lag model $\beta_3 = 0$

$$C_t = \alpha_0 + \alpha_1 Y_t + \alpha_2 Y_{t-1} + v_{4t}$$

6. First difference $\beta_3 = 1$ and $\beta_2 = -\beta_1$

$$\Delta C_t = \alpha_0 + \alpha_1 \Delta Y_t + v_{5t}$$

   where $\Delta C_t = C_t - C_{t-1}$and $\Delta Y_t = Y_t - Y_{t-1}$
7. Error correction model

$$\Delta C_t = \alpha_0 + \alpha_1 \Delta Y_t + \alpha_2 (C_{t-1} - Y_{t-1}) + v_{6t}$$

   where $\beta_3 = 1 + \alpha_2$ and $\beta_2 = -(\alpha_1 + \alpha_2)$

This is now a very commonly used model, because of its use in cointegration analysis

- All of these restrictions can be tested and you will do this in the exercise
- Look for most parsimonious
- Another issue is the long run solutions of these models
- Set $Y_t = Y_{t-1} = Y$ and do the same for X gives the long run solution
- Note that the first difference equation does not have one so if you were to find this was the best model you would only have short run dynamics.
- It is possible that this problem can be dealt with but need to discuss the concept of cointegration