1   Introduction

This part of the module has been about the building of economic, financial and investment models. In most cases the aim is to get a greater understanding of the relationship between variables are to use that to improve decision-making.

The variable that is the focus of attention, the entity that we would like to understand, is called the *dependent* variable. The variables that will be used to explain the behaviour of the dependent variable are called the explanatory variables. It is possible to summarise this relationship mathematically.

$$y_t = \beta_{1,t} + \beta_{2,t} X_t + \varepsilon_t \tag{1}$$

Where y_t is the dependent variable at time t; β_1 and β_2 are coefficients to be estimated; X_t is the explanatory variable; and ε is an error term that will represent all the other factors that affect the behaviour of the dependent variable that is the object of attention.

2   Ordinary least squares

In this example the return on Bank of America is the phenomenon that we aim to know more about and the return on the S&P 500 is the variable that will be used to I understand the return from this bank investment. To understand more about the way that the return on the market effects the return on Bank of America, we would like to estimate the parameters of Equation 1. That means finding the values of β_1 and β_2 that are most likely.

There are a number of different techniques that can be used to make this estimation. The easiest, and in many cases the best, is *ordinary least squares*. If Bank of America returns are represented by y and the return from an investment in the S&P 500 is denoted x, the relationship can be viewed graphically.

Figure 1 goes here

The aim is to draw a line that will express this relationship between x and y. If we were to draw a line arbitrarily to represent the relationship, it could be denoted

$$y_t = a + bX_t + u_t \tag{2}$$

The difference between Equation 1 and Equation 2 is that the former is a theoretical model and the latter is an actual equation. The variable ε in Equation 1 is the *error term* that represents all the factors that have not been included in the model; u_t is the *residual* that is the difference between the estimated value of the dependent variable and the actual vale of the dependent variable. As such, a good model would be one that makes the residuals as small as possible.

However, if the residuals are just added together the positive and negative numbers will just cancel each other out. Therefore it is necessary to square the residuals before the addition to create the *sum of squared residuals* (SSR).

There are two main ways that the model with the smallest squared residuals can be identified: the first is to choose an equation, sum the residuals and repeat this process until a minimum has been found; the second , which can be used in this case but not all of them, is to use analytical methods to calculate the coefficients that make the minimum.

The first method is easily performed by a computer and will not be discussed in detail here. The second uses calculus. From Equation 2,

$$u_t = y_t - a - bx_t \tag{3}$$

and

$$u^2 = (y_t - a - bx_t)(y_t - a - bx_t) \tag{4}$$

so,

$$SSR = (y_t a - bx_t)^2 \tag{5}$$

Therefore, take the derivatives to get

$$\frac{\delta u}{\delta a} = \sum_{t=1}^{t=T} 2(y_t - a - bX_t) = \qquad 0$$

$$\frac{\delta u}{\delta b} = \sum_{t=1}^{t=T} 2x_t(y_t - a - bx_t) = \qquad 0$$

Re-arrange to get

$$\hat{b} = \frac{\sum x \sum y - \frac{\sum x \sum y}{T}}{\sum x^2 - \frac{(\sum x)^2}{T}}$$

$$\hat{a} = \bar{Y} - b\bar{X}$$

In matrix form

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{u}$$
$$\mathbf{u} = \mathbf{y} - \mathbf{X}\beta$$
$$\mathbf{u`u} = (\mathbf{X}\beta + \mathbf{u})`(\mathbf{X}\beta + \mathbf{u})$$

Taking derivative and re-arranging (see textbook for proof)

$$\beta = (\mathbf{X`X})^{-1}\mathbf{X`y}$$

3   OLS results

The Adjusted R^2 is 0.29, therefore nearly 30% of the BAC returns are explained by the returns of the market. 95% confidence intervals for the β are 1.39 to 2.27.

Ordinary Least Squares is one way of estimating the model. Given a number of assumptions, the OLS coefficients are **BLUE**. That is the **B**est **L**inear **U**nbiased **E**stimator. The assumptions are

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.0130	0.0105	1.23	0.2203
SPY.R	1.8303	0.2240	8.17	0.0000

1. The errors have a zero mean

2. The errors are *independent and identically distributed* (iid)

   - No serial correlation (errors related to each other)
   - Hetroskedasticity (some errors are systematically larger than others)

3. Explanatory variables are not related to the error

4. Additionally, assume *normal errors* if we want to use normal assumption to compute *t-tests* of coefficients

4 Tests of OLS Assumptions

More can be found in the Eviews Help File *Why Test Residuals? Part 2 p. 159* Assumptions for OLS to be BLUE

1. Residuals should be iid

   - No auto or serial correlation: residuals are *independent*
   - No Hetroskedasticity: residuals are *identically distributed*

2. We would also like the residuals to be *normally distributed*: so that we can use the known quantiles from the normal distribution to test assertions about the model