Introduction to Eviews and Ordinary Least Squares

Rob Hayward

today

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 and 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.

$$y_t = \beta_0 + \beta_1 X_t + \varepsilon_t \tag{1}$$

Where y_t is the dependent variable at time t; β_0 and β_1 are coefficients to be estimated; X_t is a matrix representing the explanatory variables at time t; and ε is an error term that will represent all the other factors that affect the behaviour of the dependent variable.

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 explain the Bank of America 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_0 and β_1 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.

BAC return vs SPY return
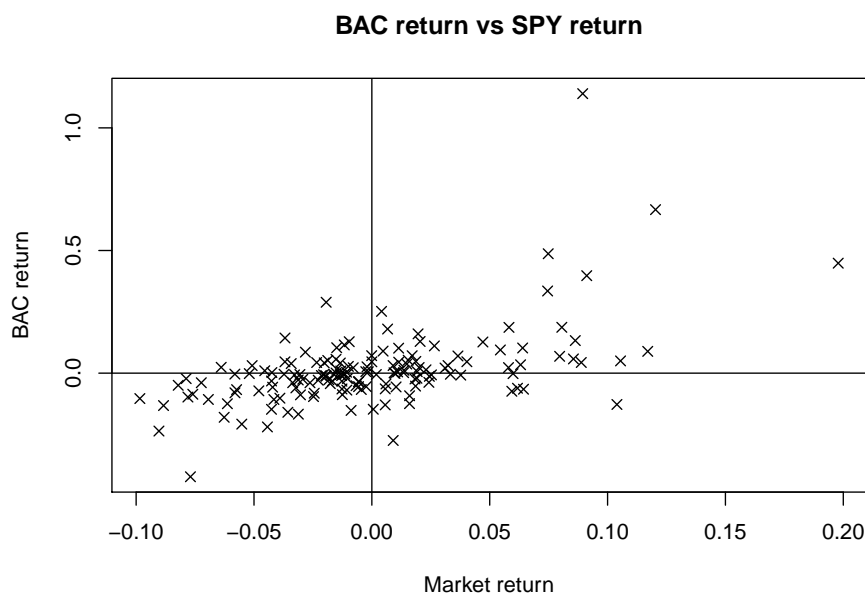
It seems that there is a positive relationship between the two variables. The aim is to draw a line that will express this relationship between x and y. If we were to draw a line arbitrarily to represent the relationship, it could be denoted

$$y_t = a + bX_t + u_t \qquad (2)$$

Where, as usual, a would be the intecept on the y axis and b would be the gradient. The gradient b is the relationship between the return on the S&P 500 and the return on Bank of America. This is what we want to find out. 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.

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 *Residual Sum of Squares* (RSS). There are two main ways that the model with the smallest RSS 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 here but not in all cases, is to use analytical methods to calculate the coefficients that make the minimum.

The first method of trial and error will not be discussed in detail here. The second method uses calculus to find the minimum value of the RSS. Using linear algebra for a multiple regression (there is more than one explanatory variable) the solution that will be used by excel, Eviews or any other package is

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{u}$$

$$\mathbf{u} = \mathbf{y} - \mathbf{X}\beta$$

$$\mathbf{u}'\mathbf{u} = (\mathbf{X}\beta + \mathbf{u})'(\mathbf{X}\beta + \mathbf{u})$$

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

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

Knowing this is not important for now. It will be calculated by the software. It can be useful to understand this analytical solution if you want to know more about the effect of relaxing the OLS assumptions. You will cover this next year in the Econometrics option.

1 OLS results

Using that formula, the following solution is found.

The regresion results are shown in Table 1. The column *Estimate* shows the estimates of the β_0 and β_1 from Equation 1. The estimate for β_0 is 1.83. This indicates that, on average, the effect of a 1 percent S&P 500 return is a 1.8% return for Bank of America. This is the main information that we require.

	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

Scatter plot of BAC and Market Returns
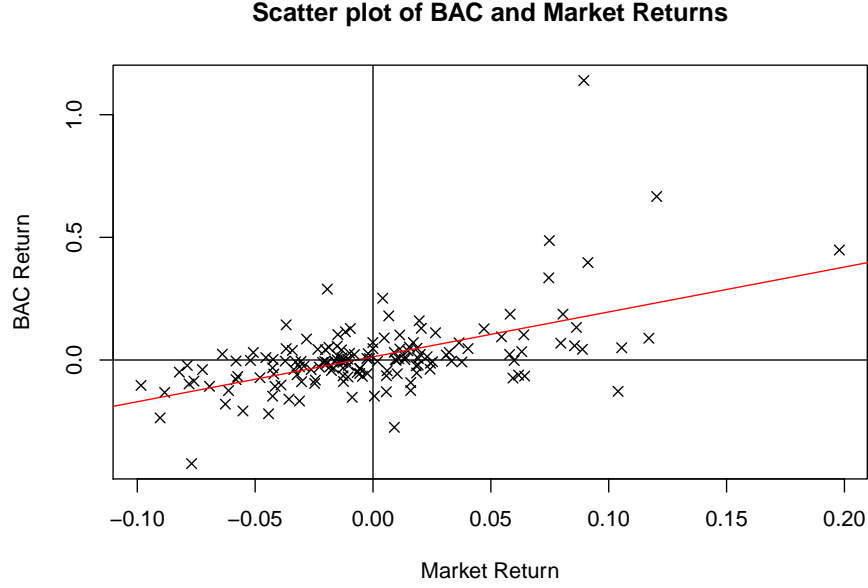
The Adjusted R^2 is 0.29, therefore nearly 30% of the BAC returns are explained by the returns of the market. The *Standard Error* is the estimate of the standard deviation of the estimated coefficient value. This means that, assuming a normal distribtion, 95% of the estimates would stand between 2 standard errors of the mean. Therefore, the 95% confidence intervals for the β_1 are 1.39 to 2.27. This means that under repeated sampling 95% ofthe estimates of β_1 should be between these values. As, if value of β_1 were really zero, it would be extremly unlikely to get a figure of 1.8% for this estimate. The probability of getting this value when it was really zero is in the final column. This probablity is based on the *t value* which is calculated as

$$t = \frac{\hat{\beta}_1 - \beta_n}{SE(\hat{\beta}_1)} \tag{3}$$

Where $\hat{\beta}_1$ is the estimate of the coefficient (1.83 in our case), β_n is the hypothesised value of the coefficient (zero in our case) and $SE(\hat{\beta}_1)$ is the estimate of

5

the standard deviation of the estimates of the coeficient (0.224). Therefore,

$$t = \frac{1.8 - 0}{0.224} = 8.17 \tag{4}$$

This is the number of standard deviations that the coefficient is away from the hypothesised value of zero. If the absolute value of the *t-statistic* is more than 2, there is less than 5 chances in 100 that the coefficient would be see when the actual value was zero. In that situation we say that the coefficient is *statistically different from zero*. 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

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