Univariate Ordinary Least Squares Estimator
Regression Constant Is Zero

Gary Schurman MBE, CFA

May, 2011

In this white paper we will estimate the unknown value of a dependent variable Y based upon the known value of an independent variable X. We can define the actual value of the dependent variable Y to be a function of the estimated value of Y plus an error term. We will define $\hat{Y}$ (Y hat) to be the estimated value of the dependent variable Y. The equation for the estimated value of Y is...

$$\hat{Y} = \beta\, X \tag{1}$$

The actual value of Y is therefore the estimated value of Y plus an error term. Using Equation (1) above the equation for the actual value of the dependent variable Y where ϵ is the error term is...

$$\begin{aligned} Y &= \hat{Y} + \epsilon \\ &= \beta\, X + \epsilon \end{aligned} \tag{2}$$

How do we go about estimating the value β in equation (1) above? We will set that variable value such that the sum of squared estimation errors is minimized. To demonstrate these techniques we will work through the following problem...

The Equation for Total Sum of Squared Errors and it's Derivative

The estimation error squared is the squared difference between the actual value of Y and the estimated value of Y. The equation for the total sum of squared errors (SSE) is therefore...

$$\begin{aligned} SSE &= \sum_{i=1}^{N} \Bigg(Y_i - \beta X_i \Bigg)^2 \\ &= \sum_{i=1}^{N} \Bigg[Y_i^2 - 2\beta Y_i X_i + \beta^2 X_i^2 \Bigg] \\ &= \sum_{i=1}^{N} \Bigg[Y_i^2 \Bigg] - \sum_{i=1}^{N} \Bigg[2\beta Y_i X_i \Bigg] + \sum_{i=1}^{N} \Bigg[\beta^2 X_i^2 \Bigg] \\ &= \sum_{i=1}^{N} \Bigg[Y_i^2 \Bigg] - 2\beta \sum_{i=1}^{N} \Bigg[Y_i X_i \Bigg] + \beta^2 \sum_{i=1}^{N} \Bigg[X_i^2 \Bigg] \end{aligned} \tag{3}$$

Note that $\bar{X}$ (X bar) is the mean of the observed values of the independent variable X and $\bar{Y}$ (Y bar) is the mean of the observed values of the dependent variable Y.

To calculate values for the parameters in equation (1) we will need the derivative of equation (3) with respect to β. The derivative of equation (3) with respect to β is...

$$\frac{\delta SSE}{\delta \beta} = 2\beta \sum_{i=1}^{N} \Bigg[X_i^2 \Bigg] - 2 \sum_{i=1}^{N} \Bigg[Y_i X_i \Bigg] \tag{4}$$

To solve for β in equation (1) above we will set the derivative in equation (4) equal to zero and jointly solve for these two parameters. The value of β is...

$$\frac{\delta SSE}{\delta\beta} = 0$$

$$2\beta\sum_{i=1}^{N}\Bigg[X_i^2\Bigg] - 2\sum_{i=1}^{N}\Bigg[Y_iX_i\Bigg] = 0$$

$$\beta\sum_{i=1}^{N}\Bigg[X_i^2\Bigg] = \sum_{i=1}^{N}\Bigg[Y_iX_i\Bigg]$$

$$\beta = \sum_{i=1}^{N}\Bigg[Y_iX_i\Bigg]\Bigg/\sum_{i=1}^{N}\Bigg[X_i^2\Bigg] \tag{5}$$

Goodness of Fit and Standard Error of Estimate

A well-fitting regression model results in predicted values close to the observed (i.e. actual) values. The mean model, which uses the mean of the data series for every predicted value, would be used if there were no informative predictor variables. The fit of a proposed regression model should therefore be better than the fit of the mean model.

We defined SSE to be the sum of squared errors using our proposed regression model (Equation (3) above). We will define SST to be the sum of squared errors using the mean model. The equations for SSE and SST are...

$$SSE = \sum_{i=1}^{N}\Bigg(Y_i - \beta X_i\Bigg)^2 \text{ ...and... } SST = \sum_{i=1}^{N} Y_i^2 \tag{6}$$

The difference between SST and SSE is the improvement in predictive power from using the regression model as compared to the mean model. The equation for R-squared is..

$$\text{R-squared} = \frac{SST - SSE}{SST} \tag{7}$$

The value of R-squared from zero to one. Zero means that the proposed model does not improve prediction over the mean model. One indicates perfect prediction. The closer to one that the value of R-squared gets the better the predictive power of the regression model.

Using Equation (6) above, the equation for the regression standard error of estimate is...

$$SEE = \sqrt{\frac{SSE}{N-1}} \tag{8}$$