Theory

It is assumed that the user is knowledgeable about the fundamentals of regression analysis. The general bivariate linear regression equation takes the form of

\[ Y = \beta_0 + \beta_1 X + \varepsilon \]

where \( \beta_0 \) is the intercept, \( \beta_1 \) is the slope, and \( \varepsilon \) is the error term. It is bivariate as there are only two variables, a \( Y \), or dependent variable, and an \( X \), or independent variable, where \( X \) is also known as the regressor (sometimes a bivariate regression is also known as a univariate regression as there is only a single independent variable \( X \)). The dependent variable is so named because it depends on the independent variable; for example, sales revenue depends on the amount of marketing costs expended on a product’s advertising and promotion, making the dependent variable “sales” and the independent variable “marketing costs.” An example of a bivariate regression is seen as simply inserting the best-fitting line through a set of data points in a two-dimensional plane, as seen on the left in Figure 1. In other cases, a multivariate regression can be performed, where there are multiple, or \( k \) number of, independent \( X \) variables or regressors, where the general regression equation will now take the form of

\[ Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 \ldots + \beta_k X_k + \varepsilon \]

In this case, the best-fitting line will be within a \( k + 1 \) dimensional plane.

However, fitting a line through a set of data points in a scatter plot as in Figure 1 may result in numerous possible lines. The best-fitting line is defined as the single unique line that minimizes the total vertical errors, that is, the sum of the absolute distances between the actual data points (\( Y_i \)) and the estimated line (Error! Objects cannot be created from editing field codes), as shown on the right of Figure 1. To find the best-fitting unique line that minimizes the errors, a more sophisticated approach is applied, using regression analysis. Regression analysis, therefore, finds the unique best-fitting line by requiring that the total errors be minimized, or by calculating

\[ \text{Min} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \]
where only one unique line minimizes this sum of squared errors. The errors (vertical distances between the actual data and the predicted line) are squared to avoid the negative errors from canceling out the positive errors. Solving this minimization problem with respect to the slope and intercept requires calculating first derivatives and setting them equal to zero:

\[
\frac{d}{d\beta_0} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = 0 \quad \text{and} \quad \frac{d}{d\beta_1} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = 0
\]

which yields the bivariate regression’s least squares equations:

\[
\beta = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n} (X_i - \bar{X})^2} = \frac{n\sum_{i=1}^{n} X_i Y_i - \sum_{i=1}^{n} X_i \sum_{i=1}^{n} Y_i}{n\sum_{i=1}^{n} X_i^2 - \left(\sum_{i=1}^{n} X_i\right)^2}
\]

\[
\beta_0 = \bar{Y} - \beta \bar{X}
\]

For multivariate regression, the analogy is expanded to account for multiple independent variables, where

\[Y_i = \beta_1 + \beta_2 X_{2,i} + \beta_3 X_{3,i} + \varepsilon_i\]

and the estimated slopes can be calculated by:

\[
\hat{\beta}_2 = \frac{n\sum_{i=1}^{n} Y_i X_{2,i} - \sum_{i=1}^{n} Y_i \sum_{i=1}^{n} X_{2,i} X_{3,i}}{n\sum_{i=1}^{n} X_{2,i}^2 - \left(\sum_{i=1}^{n} X_{2,i}\right)^2}
\]

\[
\hat{\beta}_3 = \frac{\sum_{i=1}^{n} Y_i X_{3,i} - \sum_{i=1}^{n} Y_i \sum_{i=1}^{n} X_{3,i}}{\sum_{i=1}^{n} X_{2,i}^2 \sum_{i=1}^{n} X_{3,i} - \left(\sum_{i=1}^{n} X_{2,i} X_{3,i}\right)^2}
\]

In running multivariate regressions, great care must be taken to set up and interpret the results. For instance, a good understanding of econometric modeling is required (e.g., identifying regression pitfalls such as structural breaks, multicollinearity, heteroskedasticity, autocorrelation, specification tests, nonlinearities, and so forth) before a proper model can be constructed.

**Procedure**

- Start Excel and type in or open your existing dataset (the illustration in Figure 2 uses the file *Multiple Regression* in the examples folder).
- Check to make sure that the data are arranged in columns and select the data including the variable headings, and click on *Risk Simulator | Forecasting | Multiple Regression*.
- Select the dependent variable and check the relevant options (lags, stepwise regression, nonlinear regression, and so forth) and click **OK** (Figure 2).
Figure 2. Running a Multivariate Regression

Results Interpretation

Figure 3 (on the next page) illustrates a sample multivariate regression result report generated. The report comes complete with all the regression results, analysis of variance results, fitted chart, and hypothesis test results.

In “Multivariate Regression, Part 2,” you will learn about a powerful automated approach to regression analysis known as “stepwise regression” and about how goodness-of-fit statistics provide a glimpse into the accuracy and reliability of the estimated regression model.
Figure 3. Multivariate Regression Results

TO BE CONCLUDED IN “Multivariate Regression, Part 2”