

12.1: Theoretical Specification

As with simple regression, the theoretical multiple regression model contains a **systematic** component — $Y = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik}$ and a **stochastic** component — ϵ_i . The overall theoretical model is expressed as:

$$Y = \alpha + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \epsilon_i$$

where α is the constant term - β_1 through β_k are the parameters of IVs 1 through k - k is the number of IVs - ϵ is the error term

In matrix form the theoretical model can be much more simply expressed as: $y = X\beta + \epsilon$.

The empirical model that will be estimated can be expressed as: $Y_i = A + B_1 X_{i1} + B_2 X_{i2} + \dots + B_k X_{ik} + \epsilon_i$. Therefore, the residual sum of squares (RSS) for the model is expressed as: $RSS = \sum \epsilon_i^2 = \sum (Y_i - \hat{Y}_i)^2 = \sum (Y_i - (A + B_1 X_{i1} + B_2 X_{i2} + \dots + B_k X_{ik}))^2$

12.1.1 Assumptions of OLS Regression

There are several important assumptions necessary for multiple regression. These assumptions include linearity, fixed XX 's, and errors that are normally distributed.

OLS Assumptions

Systematic Component

- *Linearity*
- *Fixed XX*

Stochastic Component

- *Errors have identical distributions*
- *Errors are independent of XX and other ϵ_i*
- *Errors are normally distributed*

Linearity

When OLS is used, it is assumed that a linear functional form is the correct specification for the model being estimated. Note that linearity is assumed in the *parameters* (that is, for the B 's), therefore the expected value of the dependent variable is a linear function of the parameters, not necessarily of the variables themselves. So, as we will discuss in later chapters, it is possible to transform the variables (the X 's) to introduce non-linearity into the model while retaining linear estimated coefficients. For example, a model with a squared XX term can be estimated with OLS:

$$Y = A + B_1 X_{i1} + B_2 X_{i2}^2 + \epsilon_i$$

However, a model with a squared BB term cannot.

Fixed XX

The assumption of fixed values of XX means that the value of XX in our observations is not systematically related to the value of the other XX 's. We can see this most clearly in an experimental setting where the researcher can manipulate the experimental variable while controlling for all other possible X 's through random assignment to a treatment and control group. In that case, the value of the experimental treatment is completely unrelated to the value of the other X 's – or, put differently, the treatment variable is orthogonal to the other X 's. This assumption is carried through to observational studies as well. Note that if XX is assumed to be fixed, then changes in YY are assumed to be a result of the independent variations in the XX 's and error (and nothing else).

This page titled [12.1: Theoretical Specification](#) is shared under a [CC BY 4.0](#) license and was authored, remixed, and/or curated by [Jenkins-Smith et al. \(University of Oklahoma Libraries\)](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.