Assumptions of Classical Linear Regression Model

The Classical Linear Regression Model (CLRM) or the Method of Least Squares is based on seven important assumptions. These assumptions are stated below keeping in mind a two variable regression model of the type $Y_i = \beta_1 + \beta_2 X_i + U_i$, where i= 1, 2, 3,...,n, X_i is the explanatory variable assumed to be non-stochastic or fixed regressor, Y_i is the dependent variable or the regressand, β_1 & β_2 are the regression coefficients or parameters and U_i, the error term or stochastic (random) variable.

***Assumption 1: Regression Model is linear in parameters**

The regression model is *linear in parameters*, though it may or may not be linear in the variables. In a linear regression model all the term in the model are either constant or a parameter multiplied by an independent variable. This means that in a two variable regression model of the type $Y_i = \beta_1 + \beta_2 X_i + U_i$, the parameters in β s are raised to the first power only irrespective of the power of X_i and Y_i in the equation.

***Assumption 2: The Mean value of the error term is zero**

Given the value of X_i, the mean or expected value of the error tem or disturbance term U_i is zero.
Symbolically, $1/n \, \Sigma U_i = 0$, i.e. $E(U_i) =0$

Suppose the mean of error is +2. This would imply that our regression model under predicts the observed values. This type of systematic error is called bias. Thus for the regression model to be unbiased the mean of error should always be zero.

***Assumption 3: The independent variable and the error term are uncorrelated**

If by any chance the independent variable is correlated to the error term, then the OLS incorrectly computes some of the variance in Y explained by the error term to the independent variable itself. If the $\mathbf{Cov\ (X_i, U_i) =0}$, then it is proved that the independent variable and the error term are uncorrelated.

***Assumption 4: Homoscedasticity or the variance of the error term should be constant.**

Irrespective of the value of X_i, the variance of the error term should be constant. Symbolically, we express this as $Var\ (U_i) = E[U_i – E(U_i)]^2 = E\ (U_i^2) = \sigma^2$. Homoscedasticity means equal (homo) spread (scedasticity) or equal variance. If the variance of the error term is not constant or is unequal, then the problem of heteroscedasticity arises and reduces the precision of estimation.

***Assumption 5: The disturbance terms or error terms are uncorrelated**

We should not be able to predict the probability of the next error term from observing of one error term. This is possible when given two values X say X_i and X_j ($i \neq j$), the correlation between U_i and U_j ($i \neq j$) must be zero. Expressed symbolically, we have, $Cov(U_i, U_j) = 0$. The violation of this assumption creates the problem of autocorrelation or serial correlation which reduces the exactness of estimation.

***Assumption 6: The number of observations i.e 'n' must be greater than the number of parameters to be estimated**

Suppose we are estimating two parameters. In that case we need more than one pair of data in (X_i, Y_i) to determine the parameters or otherwise the calculation becomes impossible. Hence the number of observations i.e 'n' must be greater than the number of parameters to be estimated or the number of observations must be greater than the number of explanatory variables.

***Assumption 7: The Nature of X variables**

The X values in a given sample must not all be the same. Technically, Var (X) must be a positive number. Moreover, there can be no outliers in the values of X meaning no value of X can be unnaturally large in comparison to the rest of the observations. This is done to ensure that that the regression results are not biased by outliers.

*N.B. Explanations of all the theories have been taken from the following references:

References:

Gujarati, D. N. Porter, D.C., Gunasekar, S. (2009), *Basic econometrics*. (Fifth ed.) McGraw-Hill Education (India).

7 classical assumptions of ordinary least squares (OLS) linear regression. (2019, June 13). Statistics By Jim. https://statisticsbyjim.com/regression/ols-linear-regression-assumptions/