Assumption 5: No autocorrelation between the disturbances. Given any two X values, $X_i$ and $X_j$ ($i \neq j$), the correlation between any two $U_i$ and $U_j$ ($i \neq j$) is zero.

$$\text{Cov}(U_i, U_j / X_i, X_j) = E(\{U_i - E(U_i)\}/X_i) \cdot (\{U_j - E(U_j - E(U_j))\}/X_j) = E(U_i / X_i) \cdot (U_j / X_j) = 0$$

Assumption 5 implies absence of any systematic pattern of U distribution.

Suppose in PRF ($Y_t = \beta_1 + \beta_2 X_t + U_t$) that $U_t$ and $U_{t-1}$ are positively correlated, then $Y_t$ depends not only on $X_t$ but also on $U_{t-1}$ to some extent determines $U_t$. We will only consider the systematic effect, if any of $X_t$ on $Y_t$ and not worry about the other influences that might act on $Y$ as a result of the possible inter-correlation among the U’s.

Assumption 6: Zero covariance between $U_i$ and $X_i$

$$E(U_i, X_i) = 0$$

the disturbance U and explanatory variable X are uncorrelated when we express PRF we assumed that X and U (which may represent the influence of all omitted variables) have separate (and additive) influence on Y. but if X and U are correlated, it is not possible to assess their individual effects on Y. Thus, if X and U are positively correlated, X increases when U increases and it decreases when U decreases, Assumption 6 is automatically fulfilled if X variable is non-random or nonstochastic.

Assumption 7: The number of observations $n$ must be greater than the number of parameters to be estimated (we need some variation).

Assumption 8: The X values in a given sample must not all be the same. Variability in X values.

$$\text{Var} (X) = \text{a finite positive number}$$

If all values are identical for X, then $X_i = \bar{X}$ and the denominator of that equation will be zero making it impossible to estimate $\beta_2$ and therefore $\beta_1$. The variation in both Y and X is essential to use regression analysis as a research tool.

Assumption 9: The regression model is correctly specified. Alternatively, there is no specification bias or error in the model used in empirical analysis.

The classical econometric methodology assumes implicitly, if not explicitly that the model used to test an economic theory is “correctly specified.”
Assumption 10: There is no perfect multicollinearity. That is, there are no perfect relationships among the explanatory variables.

**Precision of Standard Errors of Least-Squares Estimates.**

\[
\text{Var}(\hat{\beta}_2) = \frac{G^2}{\sum x_i^2} \quad \hat{G}^2 = \frac{\sum \hat{u}_i^2}{n-2}
\]

\[
SE(\hat{\beta}_2) = \frac{G}{\sqrt{\sum x_i^2}}
\]

where \( \hat{G}^2 \) is the OLS estimator of the true but unknown \( G^2 \) and \( n-2 \) is degrees of freedom (df); \( \sum \hat{U}_i^2 \) is the sum of the residuals squared or the residual sum of squares (RSS).

Once \( \sum u_i^2 \) is known, \( \hat{G}^2 \) can be easily computed.

\[
\sum u_i^2 = \sum y_i^2 - \beta_2^2 \sum x_i^2
\]

The positive square root of \( \hat{G}^2 \) is \( \hat{G} = \sqrt{\frac{\sum u_i^2}{n-2}} \) is known as the standard error of estimate or the standard error of the regression (SE). It is simply the standard deviation of the Y values about the estimated regression line and is often used as a summery measure of the “goodness of fit” of the estimated regression line.

Given \( X_i, G^2 \) represents the (conditional) variance of both \( U_i \) and \( Y_i \). Therefore, the standard error of the estimate can also be called the (conditional) standard deviation of \( U_i \) and \( Y_i \).

\( G^2 \) and \( G_Y \) represent, respectively, the unconditional variance and unconditional standard deviation of Y. Features of the variance (and therefore the standard errors) of \( \hat{\beta}_1 \) and \( \hat{\beta}_2 \):

1) The variance of \( \hat{\beta}_2 \) is directly proportional to \( G^2 \) but inversely proportional to \( \sum x_i^2 \).

That is, given \( G^2 \), the larger the variation in the X values, the smaller the variance of \( \hat{\beta}_2 \) and hence the greater the precision with which \( \beta_2 \) can be estimated. That means the larger sample gives more information to estimate slope better. More observations always better than less. And the larger the \( G^2 \) (total) the larger the variance for \( \hat{\beta}_2 \), which means less precision of estimating \( \hat{\beta}_2 \).
2) The variance of \( \hat{\beta}_1 \) is directly proportional to \( G^2 \) and \( \sum x_i^2 \), but inversely proportional to \( \sum x_i^2 \) and the sample size.

3) Since \( \hat{\beta}_1 \) and \( \hat{\beta}_2 \) are estimators, they will not only vary from sample to sample, but in a given sample they are likely to be dependent on each other, which is measure by the covariance between them.

\[
COV(\hat{\beta}_1, \hat{\beta}_2) = -\bar{x} \text{var}(\hat{\beta}_2) = -\bar{x}(\frac{G^2}{\sum x_i^2})
\]

Since \text{var}(\hat{\beta}_2) is always positive, as is the variance of any variable, the nature of the covariance between \( \hat{\beta}_1 \) and \( \hat{\beta}_2 \) depends on the sign of \( \bar{x} \). If \( \bar{x} \) is positive, then the covariance will be negative. Thus, if the slope coefficient \( \beta_2 \) is overestimated, the intercept coefficient \( \hat{\beta}_1 \) will be underestimated.

**Properties of Least-Squares Estimators: The Gauss-Markov Theorem**

An estimator is said to be a best linear unbiased estimator (BLUE) if the following hold:

1) It is linear, that is, a linear function of a random variable, such as the dependent variable \( Y \) in the regression model.

2) It is unbiased, that is, its average or expected value, \( E(\hat{\beta}_2) \), is equal to the true value, \( \beta_2 \).

3) It has minimum variance in the class of all such linear unbiased estimators. An unbiased estimator with the least variance is known as an efficient estimator.

In the regression context it can be proved that the OLS estimators are BLUE.

**Gauss-Markov Theorem**: Given the assumptions of the classical linear regression model, the least-squares estimators, in the class of unbiased linear estimators, have minimum variance, that is, they are BLUE.

\[
\sum y_i^2 = \sum (Y_i - \bar{Y})^2 = \text{total variation of the actual } Y \text{ values about their sample mean} \quad \text{(total sum of squares, TSS)}
\]

\[
\sum \hat{y}_i^2 = \sum (\hat{Y}_i - \bar{Y})^2 = \sum (\hat{Y}_i - \bar{Y})^2 = \text{variation of the estimated } Y \text{ values about their mean} \quad \text{(\( \hat{Y} = Y \) which is the sum of squares due to regression line (due to the explanatory variables = ESS is explained sum of squares)}
\]
\[ \sum \hat{U}_i^2 = \text{residual or unexplained variation of the Y values about the regression line} = \text{RSS} \]

is residual sum of squares. TSS = ESS + RSS

TSS = ESS + RSS

\[ 1 = \frac{ESS}{TSS} + \frac{RSS}{TSS} \]

\[ r^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS} \] (sample coefficient of determination)

\( r^2 \) measures the proportion or percentage of the total variation in Y explained by the regression model.

If \( r^2 = 1 \) then perfect fit \( \hat{Y}_i = Y_i \) for each \( i \).

If \( r^2 = 0 \), the \( \hat{\beta}_2 = 0 \), in this case \( \hat{Y}_i = \hat{\beta}_1 = \bar{Y} \) that is the best prediction of any Y value is simply its mean value. (regression line is horizontal to the X axis).

The coefficient of correlation is the measure of the degree of association between two variables: \( r = \pm \sqrt{r^2} \) (sample correlation coefficient)

\[ r^2 = \frac{(\sum y_i \hat{y}_i)^2}{(\sum y_i^2)(\sum \hat{y}_i^2)} \]

where \( y_i = \text{actual Y} \), \( \hat{y}_i = \text{estimated} \) and \( \bar{y} = \bar{y} = \text{the mean of Y} \).