Econometrics

Course Calendar	
Week	Main Content
Week 1	Introduction to Simple Regression
Week 2	Simple Regression
Week 3	Simple Regression: r^2 & Hands-on-Exercise
Week 4	Central Limit Theorem, Probability and Probability Density Function (PDF)
Week 5	Hypothesis Testing: Basics
Week 6	Simple Regression: Testing of Hypothesis

Econometrics
Lecture 2. Simple Regression

Geetha Rani Prakasam, Ph.D.

Professor

Recap: PRF and SRF

- Population Regression Function
- Linearity in the parameters
- Stochastic PRF
- Stochastic Disturbance Term u_i plays a critical role in estimating the PRF
- Sample of observations from population
- Stochastic Sample Regression Function SRF is used to estimate the PRF

Recap

- For empirical purposes, it is the stochastic PRF that matters. The stochastic disturbance term u_i plays a critical role in estimating the PRF.
- The PRF is an idealized concept, since in practice one rarely has access to the entire population of interest.
- Generally, one has a sample of observations from population and use the stochastic sample regression (SRF) to estimate the PRF.

Econometrics

Lecture 2:

TWO-VARIABLE REGRESSION MODEL: The problem of Estimation

Outline of Lecture

- The problem of estimation
- Method of ordinary least squares
- Least Square Criterion
- Normal Equations
- Numerical and statistical properties of OLS
- Why assumptions underlying the method of LS?
- Assumptions underlying the method of LS
- Precision or SE of least-squares estimates

TWO-VARIABLE REGRESSION MODEL:
The problem of Estimation

- First task is to estimate the PRF on the basis of the SRF as accurately as possible.
- There are two generally used methods of estimation: OLS and maximum likelihood (ML)
- OLS is used extensively in regression analysis primarily because it is intuitively appealing and mathematically much simpler than the method of ML.
- Besides, in the linear regression context the two methods generally give similar results.

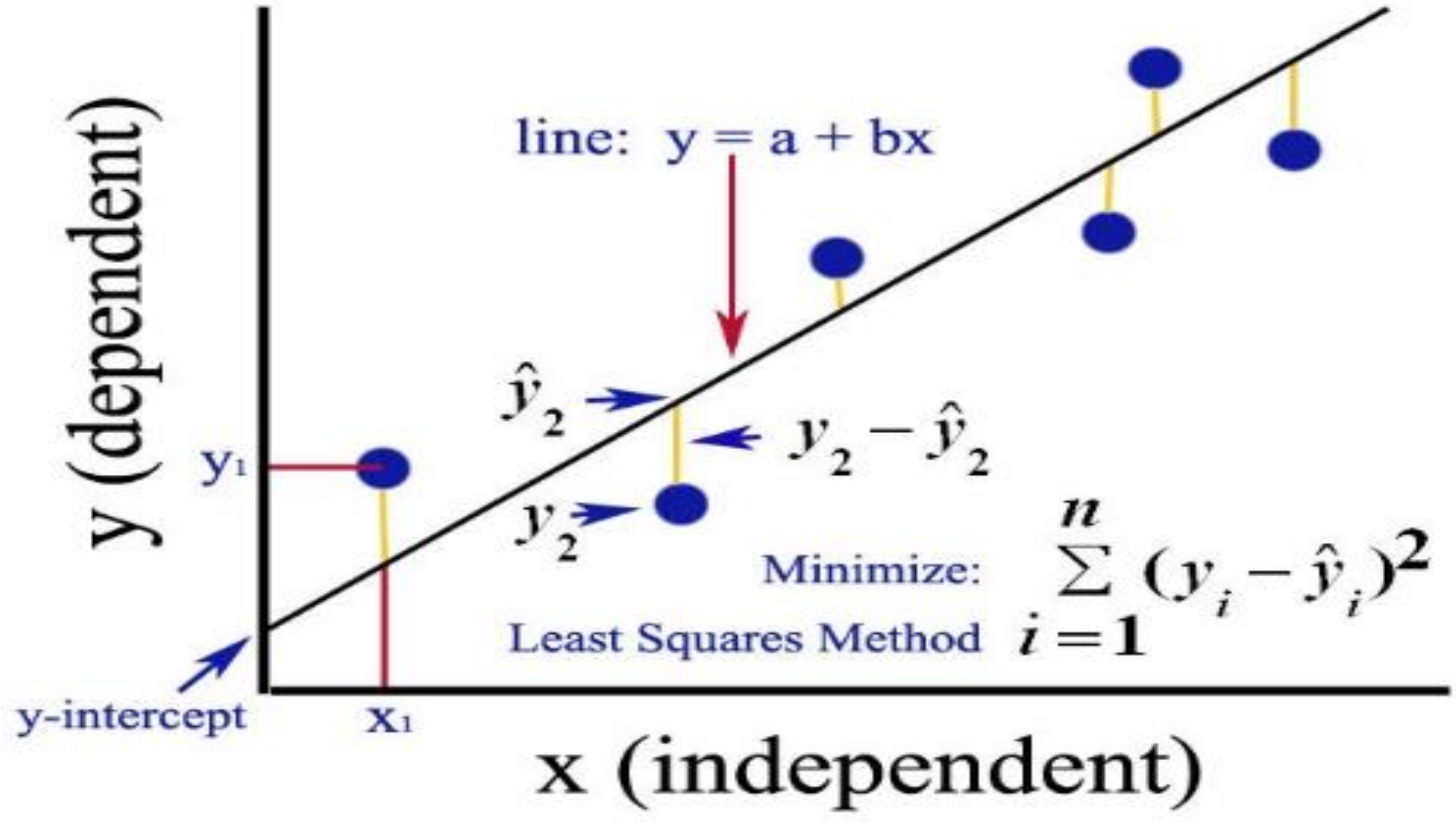
Method of ordinary least squares

- OLS – attributed to Carl Friedrich Gauss
- Attractive statistical properties
- Recall PRF 2.4.2; SRF 2.6.2 and 2.6.3.
- $Yi = ß_1 + ß_2 X_i + U_i$ ---- (2.4.2)
- $Y_i = \beta^\wedge_1 + \beta^\wedge_2 X_i + u^\wedge_I$ ----(2.6.2)

  or $Y_i = Y^\wedge_i + u^\wedge_i$ -------(2.6.3)
- Express equation 2.6.3 as
- $u^\wedge_i = Y_i - Y^\wedge_i$ -------(3.1.1)

  $= Y_i - \beta^\wedge_1 - \beta^\wedge_2 X_i$

Fig 3.1 Least-squares criterion



[Ordinary Least Square (OLS) Method for Linear Regression | by Aishwarya Gulve | Analytics Vidhya | Medium]

3.1. Least-square criterion

- If we adopt the *least-squares criterion, which states that the SRF* can be fixed in such a way that Minimizing $\sum U\hat{}^2_i = \sum(Y_i - Y\hat{}_i)^2$
- $= \sum(Y_i - \beta\hat{}_1 - \beta\hat{}_2 X_i)^2$ –(3.1.2)
- *is as small as possible, where* $u\hat{}^2_i$ *are the squared residuals.*
- The principle or the **method of least squares** chooses $\beta\hat{}_1$ and $\beta\hat{}_2$ in such a manner that, for a given sample or set of data, $\sum u\hat{}_I{}^2$ is as small as possible.
- In other words, for a given sample, the method of LS provides us with unique estimates of *β1 and β2 that give the smallest possible value of* $\sum u\hat{}_I{}^2$.

3-1. The method of OLS

- How is this accomplished? This is a straightforward exercise in differential calculus.
- Minimizing $\sum U\hat{}^2_i = \sum(Y_i - Y\hat{}_i)^2$
  $= \sum(Y_i - \beta\hat{}_1 - \beta\hat{}_2 X)^2$ (3.1.2)
- the process of differentiation yields the following equations for estimating *β1 and β2:*
- Normal Equation and solving it for $\beta\hat{}_1$ and $\beta\hat{}_2$ = Least-square estimators [See (3.1.6)(3.1.7)]

Normal Equations

- The process of differentiation yields the following equations for estimating β_1 and β_2
- $\sum Y_i = n\hat{\beta}_1 + \hat{\beta}_2 \sum X_i$ (3.1.4)
- $\sum Y_i X_i = \hat{\beta}_1 \sum X_i + \hat{\beta}_2 \sum X_i^2$ (3.1.5)
- where *n is the sample size.*
- *These simultaneous equations are known as the* ***normal*** **equations.**
- Solving the normal equations simultaneously, we obtain

Least Square estimators

- $\beta\hat{}_2 = n \sum X_i\ Y_i - \sum X_i \sum Y_i / n \sum X_i^{2} - (\sum X_i)^2$
-

  $= \sum (X_i - \bar{x})\ (Y_i - \bar{y}) / (\sum X_i - \bar{x})^2$

  $= \sum x_i\ y_i / \sum x_i^2$ --------------(3.1.6)

  Small / lower case letter denote the deviation from mean values.

- $\beta\hat{}_1 = \sum X_i^{2} \sum Y_i - \sum X_i \sum X_i\ Y_i / n \sum X_i^{2} - (\sum X_i)^2$
- $= \bar{y} - \beta\hat{}_2\ \bar{x}$ ---------------(3.1.7)

3-1. Numerical and statistical properties of OLS:

- Hold only under certain assumptions about the data were generated (Assumptions of OLS).
- OLS estimators are expressed solely in terms of observable quantities.
- They are point estimators: provide only a single value of the relevant population parameter.
- The sample regression line passes through sample means of X and Y.
- The mean value of the estimated Y = $\hat{Y}_i$ is equal to the mean value of the actual Y, where use is made of the fact that $\sum(X_i - \bar{x}) = 0$

3-1. Numerical and statistical properties of OLS:

- The mean value of the estimated Y^ is equal to the mean value of the actual Y: $E(Y) = E(Y\hat{})$
- The mean value of the residuals $U\hat{}_i$ is zero: $E(u\hat{}_i)=0$
- The residuals $u\hat{}_i$ are uncorrelated with the predicted $Y\hat{}_i$ and with X_i :
- That are $\sum u\hat{}_i Y\hat{}_i = 0$; $\sum u\hat{}_i X_i = 0$

3-2. Why assumptions underlying the method of LS?

- What is our objective?
- our objective is not only to obtain $\hat{\beta}_1$ *and* $\hat{\beta}_2$ *but also to draw inferences about the true* β_1 *and* β_2 *.*
- would like to know how close $\hat{Y}_i$ *is to the true* $E(Y \mid X_i)$*.*
- To that end, we must not only specify the functional form of the model, as in Eq. (2.4.2), but also make certain assumptions about the manner in which *Yi are generated.*
- *To see why this requirement is needed, look at the PRF:*

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

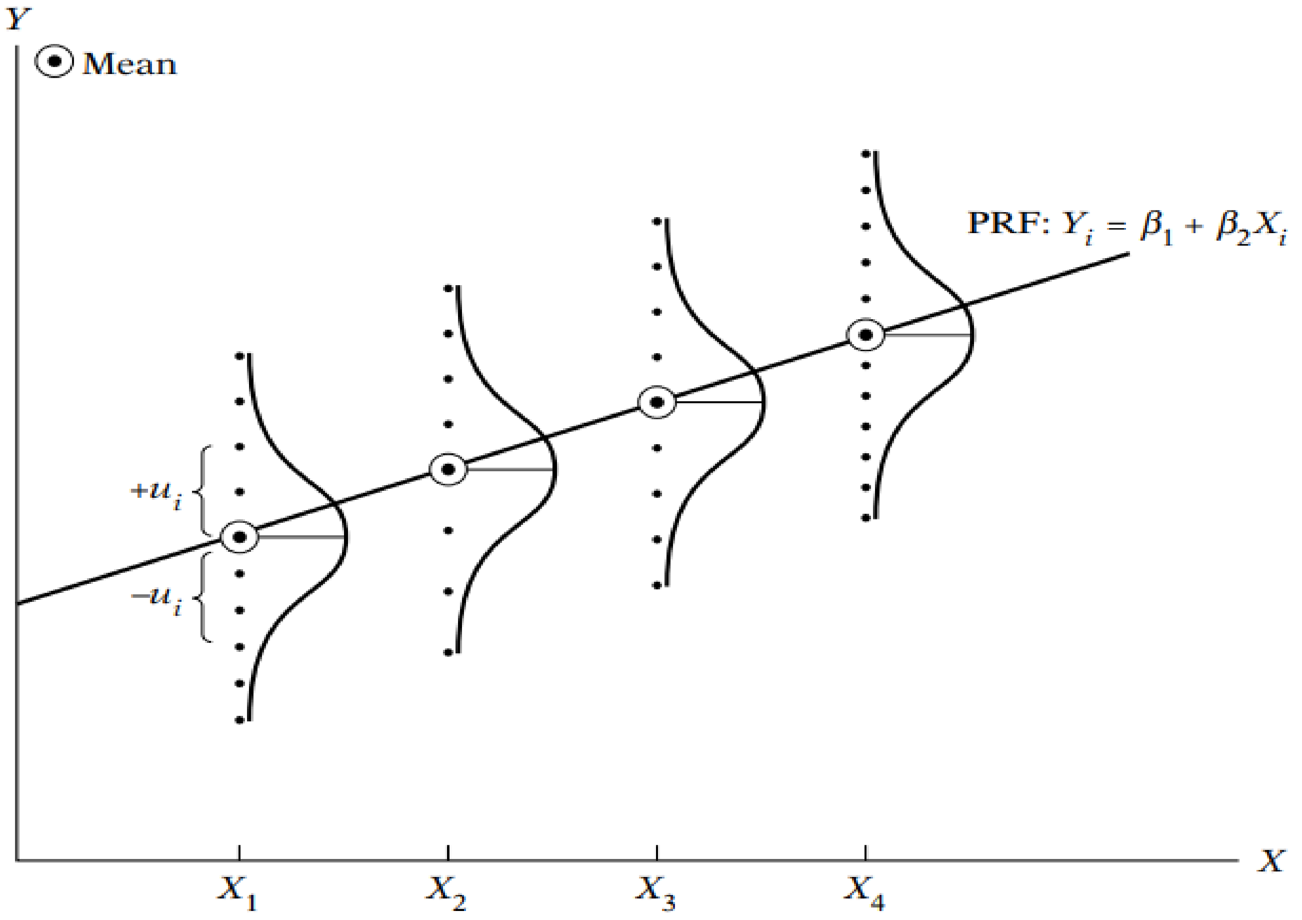
3-2. Why assumptions underlying the method of LS?

- It shows that Y_i depends on both X_i and u_i.
- Therefore, unless we are specific about how X_i and u_i are created or generated, there is no way we can make any statistical inference about the Y_i and also, as we shall see, about β_1 and β_2.
- *Thus, **the*** **assumptions made about the *Xi variable(s) and the error term are extremely critical to the*** **valid interpretation of the regression estimates**

3-2. Assumptions underlying the method of LS

- **The Gaussian, standard, or classical linear regression model (CLRM)**
- Ass 1: Linear regression model (linear in parameters)
- Ass 2: X values are fixed in repeated sampling; X is assumed to be nonstochastic. This means our regression analysis is **conditional regression analysis**, i.e. conditional on the given values of the regressor(s) X.
- Ass 3: Zero mean value of u_i : $E(u_i \mid X_i)=0$; this assumption pictured in Fig 3.3.

Fig 3.3.Conditional Distribution of Ui



Basic Econometrics by Damodar Gujrati, Page.67

3-2. Assumptions underlying the method of LS

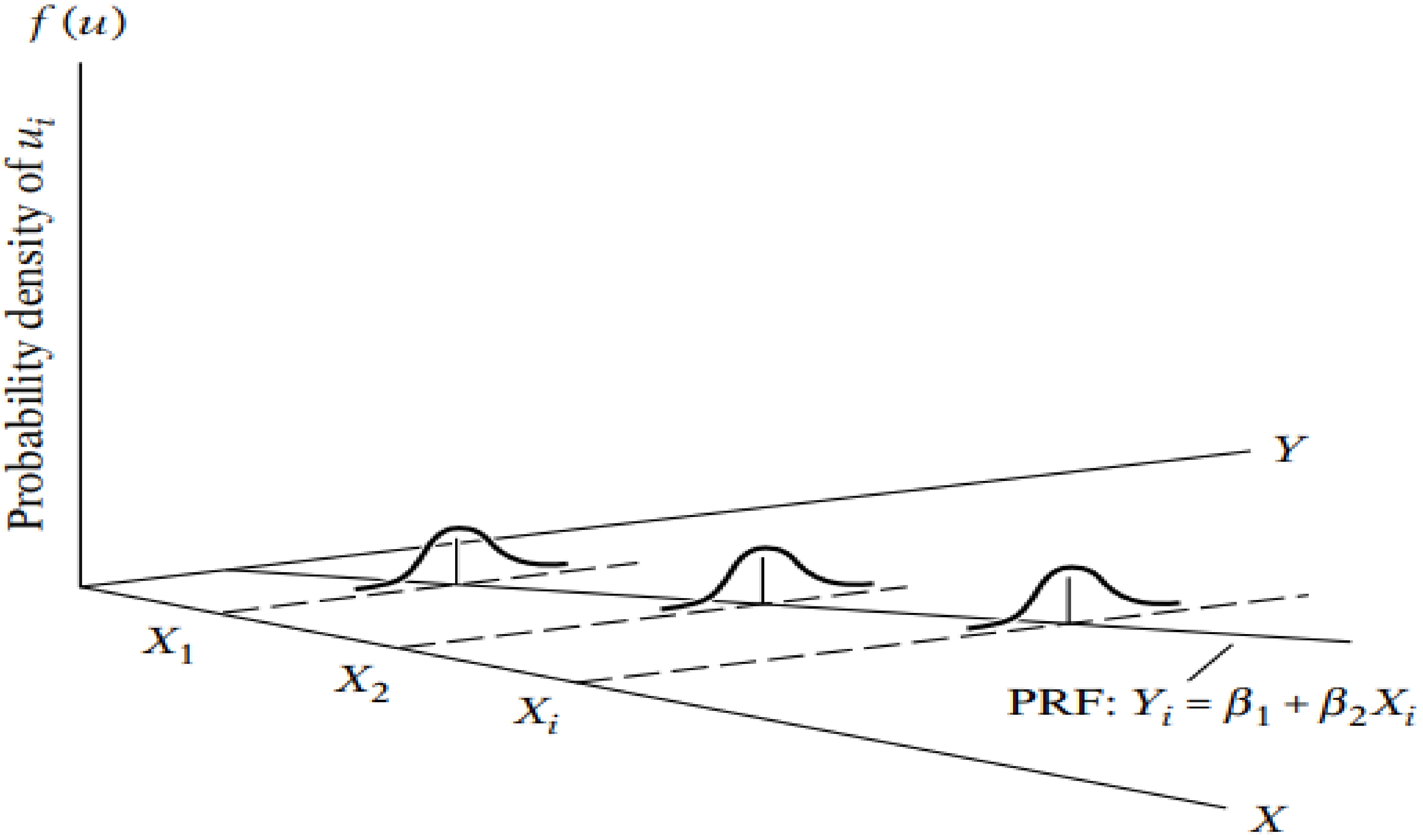
- This assumption says that the factors not explicitly included in the model, therefore subsumed in u_i do not systematically affect the mean value of Y;
- the positive and negative ui values cancel out, so that their average mean effect on Y is zero.
- $E(u_i | X_i)=0$ implies that $E(Y_i | X_i)=$ $\beta_1 + \beta_2 Xi$
- Ass 4: Homoscedasticity or equal or constant variance of u_i regardless of the values of X:

$$Var\ (u_i | X_i) = \sigma^2 \quad \text{------------}(3.2.2)$$

Homoscedasticity Assumption

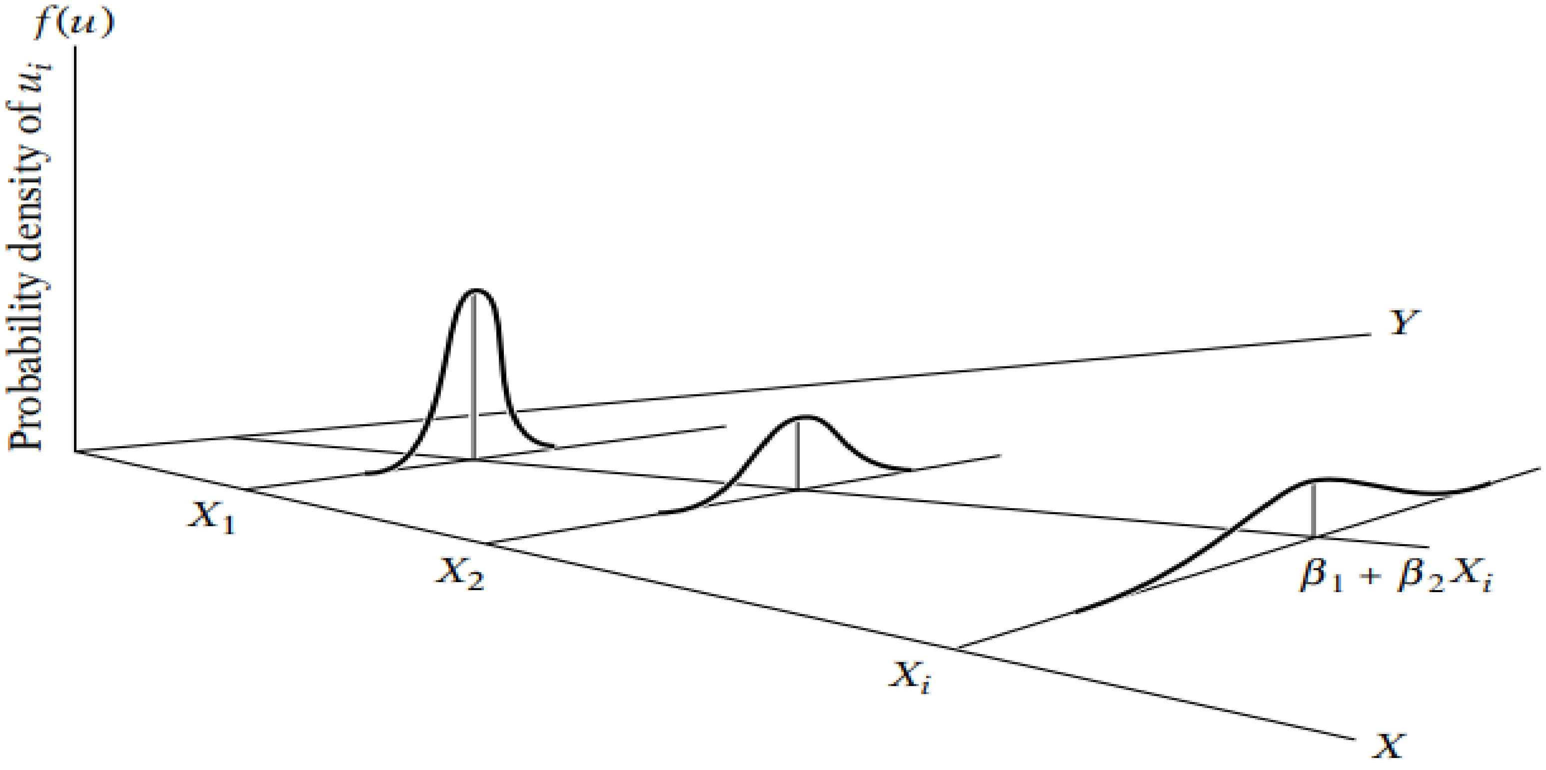
- Equation 3.2.2 states that the variance of ui for each Xi (i.e., the conditional variance of ui) is some positive constant number equal to σ^2.
- Greek verb skedanime means to disperse or scatter.
- This assumption in Eq. (3.2.2) means that the Y populations corresponding to various X values have the same variance.
- Put simply, the variation around the regression line (which is the line of average relationship between Y and X) is the same across the X values; it neither increases nor decreases as X varies.
- Diagrammatically, the situation is as depicted in Figure 3.4.

Fig 3.4: Homoscedasticity



Basic Econometrics by Damodar Gujrati, Page.68

Fig 3.5: Heteroscedasticity



Basic Econometrics by Damodar Gujrati, Page.69

Homosedasticity VS. Heteroscedasticity

- When conditional variance of the Y population varies with X, such situation is known as **heteroscedasticity, or unequal** spread, or variance.
- Symbolically, in this situation, Eq. (3.2.2) can be written as var $(u_i \mid X_i) = \sigma^2{}_i$ --------(3.2.3)
- Not all Y values corresponding to the various X's will be equally reliable, reliability being judged by how closely or distantly the Y values are distributed around their means, that is, the points on the PRF.
- By invoking Assn. 4, all Y values corresponding to the various X's are equally important.
- Assn. 4 implies that the conditional variances of Yi are also homoscedastic. That is,
- var (Yi |Xi) = σ2 ------------(3.2.4)

No Serial Correlation or No Autocorrelation

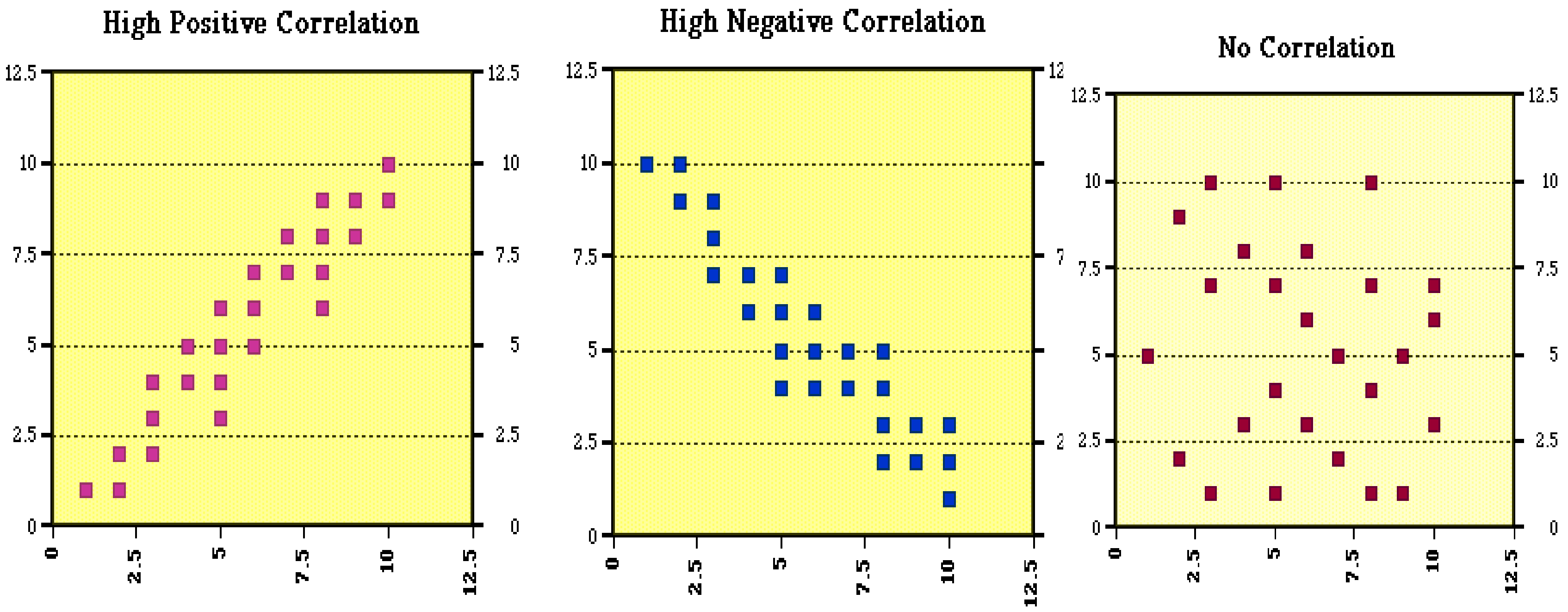
- Ass 5: No autocorrelation between the disturbances:

  $Cov(u_i, u_j | X_i, X_j) = 0$ ----(3.2.5) with i # j [VS. Correlation, + or -]

- This assn. postulates that the disturbances ui and uj are uncorrelated.
- This means that, given Xi, the deviations of any two Y values from their mean value do not exhibit patterns such as those shown in Figures 3.6(a) (b) and ©

Autocorrelation: fig 3.6

(a) (b) ©



Assn 5: No Serial Correlation or No Autocorrelation

- In Figure 3.6(a), we see that the u’s are positively correlated, a positive u followed by a positive u or a negative u followed by a negative u.
- In Figure 3.6(b), the u’s are **negatively correlated, a positive u** followed by a negative u and vice versa.
- If the disturbances (deviations) follow systematic patterns, such as those shown in Figures 3.6(a) and (b), there is auto- or serial correlation, and what Assn. 5 requires is that such correlations be absent.
- Figure 3.6(c) shows that there is no systematic pattern to the u’s, thus indicating zero correlation.
- Type of data to be noted.

Assumptions underlying the method of LS

- **Ass 6:** The Number of Observations n Must Be Greater than the Number of Parameters to Be Estimated.
- Alternatively, the number of observations must be greater than the number of explanatory variables.
- **Ass 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. Furthermore, there can be no **outliers in** the values of the X variable, that is, values that are very large in relation to the rest of the observations.
- **Ass 8: Zero covariance between u_i and X_i**
- $\mathrm{Cov}(u_i, X_i) = E(u_i, X_i) = 0$

3-2. The assumptions underlying the method of LS

- **Ass 9:** The regression model is correctly specified
- **Ass 10:** There is no perfect multicollinearity between Xs
- It is important to note that all of these assumptions pertain to the PRF only and not the SRF.
- How realistic are all these assumptions?
- The "reality of assumptions" is an age-old question in the philosophy of science.
- Some argue that it does not matter whether the assumptions are realistic.
- What matters are the predictions based on those assumptions.

3-3. Precision or std. errors of LS estimates

- From Eqs. (3.1.6) and (3.1.7),
- = $\sum x_i y_i / \sum x_i^2$ -------------(3.1.6)
- = $\bar{y}$ - $\hat{\beta}_2 \bar{X}$ --------------(3.1.7)
- it is evident that least-squares estimates are a function of the sample data.
- But since the data are likely to change from sample to sample, the estimates will change ipso facto.
- Therefore, what is needed is some measure of "reliability" or **precision of the estimators** $\hat{\beta}_1$ ***and*** $\hat{\beta}_2$**.**

3-3. Precision or SE of least-squares estimates

- In statistics the precision of an estimate is measured by its standard error (SE).
- The **standard error is nothing but the standard deviation of the sampling distribution of the estimator,** and the sampling distribution of an estimator is simply a probability or frequency distribution of the estimator, that is, a distribution of the set of values of the estimator obtained from all possible samples of the same size from a given population.
- Sampling distributions are used to draw inferences about the values of the population parameters on the basis of the values of the estimators calculated from one or more samples.

3-3. Precision or SE of least-squares estimates

- $\text{var}(\hat{\beta}_2) = \sigma^2 / \sum x_i^2$ (3.3.1)
- $\text{se}(\hat{\beta}_2) = \sqrt{\text{Var}(\hat{\beta}_2)}$ (3.3.2)
- $\text{var}(\hat{\beta}_1) = \sigma^2 \sum X_i^2 / n \sum x_i^2$ (3.3.3)
- $\text{se}(\hat{\beta}_1) = \sqrt{\text{Var}(\hat{\beta}_1)}$ (3.3.4)
- where var = variance and se = standard error and where *σ2 is the constant or* homoscedastic variance of *ui of Assn. 4.* is standard error of the estimate.

3-3. Precision or SE of least-squares estimates

- All the quantities entering into the preceding equations except *$\sigma 2$ can be estimated from* the data.
- $\hat{\sigma}^2 = \sum \hat{u}^2_i / (n - 2)$ (3.3.5)
- where $\hat{\sigma}^2$ *is the OLS estimator of the true but unknown* σ^2 *and where the expression* $n - 2$ is known as the **number of degrees of freedom (df),** $\sum \hat{u}^2_i$ being the sum of the residuals squared or the **residual sum of squares (RSS).**
- It is simply the standard deviation of the *Y values about the estimated regression line and is* often used as a summary measure of the "goodness of fit" of the estimated regression line.

3-3. **Features of the variance (**and therefore the standard errors) of $\beta^{\wedge}_1$ and $\beta^{\wedge}_2$:

- (i) **var($\beta^{\wedge}_2$) is proportional to σ^2 and inversely proportional to $\sum x^2_i$.**
- That is, given σ2, the larger the variation in the X values, the smaller the variance of **$\beta^{\wedge}_2$** and hence the greater the precision with which **$\beta^{\wedge}_2$** can be estimated.
- In short, given σ2, if there is substantial variation in the X values, **β_2** can be measured more accurately than when the Xi do not vary substantially.

3-3. **Features of the variance** (and therefore the standard errors) of $\beta\hat{}_1$ and $\beta\hat{}_2$:

- Also, given $\sum \mathbf{x}^2_i$, the larger the variance of σ^2, the larger the variance of $\boldsymbol{\beta}_2$.
- Note that as the sample size n increases, the number of terms in the sum, $\sum \mathbf{x}^2_i$ will increase.
- As n increases, the precision with which $\boldsymbol{\beta}_2$ can be estimated also increases.

3-3. Features of the variance of $\beta\hat{}_1$ and $\beta\hat{}_2$:

- (ii). var($\beta\hat{}_1$) is proportional to σ^2 and $\sum X^2_i$ but inversely proportional to $\sum x^2_i$ &the sample size n.
- (iii). Since $\beta\hat{}_1$ *and* $\beta\hat{}_2$ *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, this dependence being measured by the covariance between them.
- cov ($\beta\hat{}_1$, $\beta\hat{}_2$) = -X bar var($\beta\hat{}_2$)
- = -Xbar(σ^2 / $\sum x^2_i$) --------- (3.3.9)
- shows the independence between $\beta\hat{}_1$ and $\beta\hat{}_{2.}$

3-3. Features of the variance of $\beta\text{^}_1$ and $\beta\text{^}_2$:

- Since var($\beta\text{^}_2$) is always positive, as is the variance of any variable, the nature of the covariance between $\beta\text{^}_1$ and $\beta\text{^}_2$ depends on the sign of Xbar. If Xbar is positive, then as the formula shows, the covariance will be negative.
- Thus, if the slope coefficient $\beta\text{^}_2$ is overestimated (i.e., the slope is too steep), the intercept coefficient $\beta\text{^}_1$ will be underestimated (i.e., the intercept will be too small).

3-3. Features of the variance of $\hat{\beta}_1$ and $\hat{\beta}_2$:

- The utility of studying the covariances between the estimated regression coefficients – helpful in addressing Multicollinearity
- How do the variances and standard errors of the estimated regression coefficients enable one to judge the reliability of these estimates?
- This is a problem in statistical inference.

3-4. Properties of least-squares estimators: The Gauss-Markov Theorem

- As noted earlier, given the assumptions of the classical linear regression model, the least squares estimates possess some ideal or optimum properties.
- These properties are contained in the well-known **Gauss–Markov theorem.**
- **To understand this theorem, we need** to consider the **best linear unbiasedness property of an estimator.**
- An estimator, say the OLS estimator $\beta\hat{}_2$, is said to be a best linear unbiased estimator (BLUE) of β_2 if the following hold:

3-4. Properties of least-squares estimators: The Gauss-Markov Theorem

- An OLS estimator is said to be BLUE if :
- It is *linear*, that is, a linear function of a random variable, such as the dependent variable Y in the regression model.
- It is *unbiased* , that is, its average or expected value, $E(\hat{\beta}_2)$, is equal to the true value β_2
- 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*

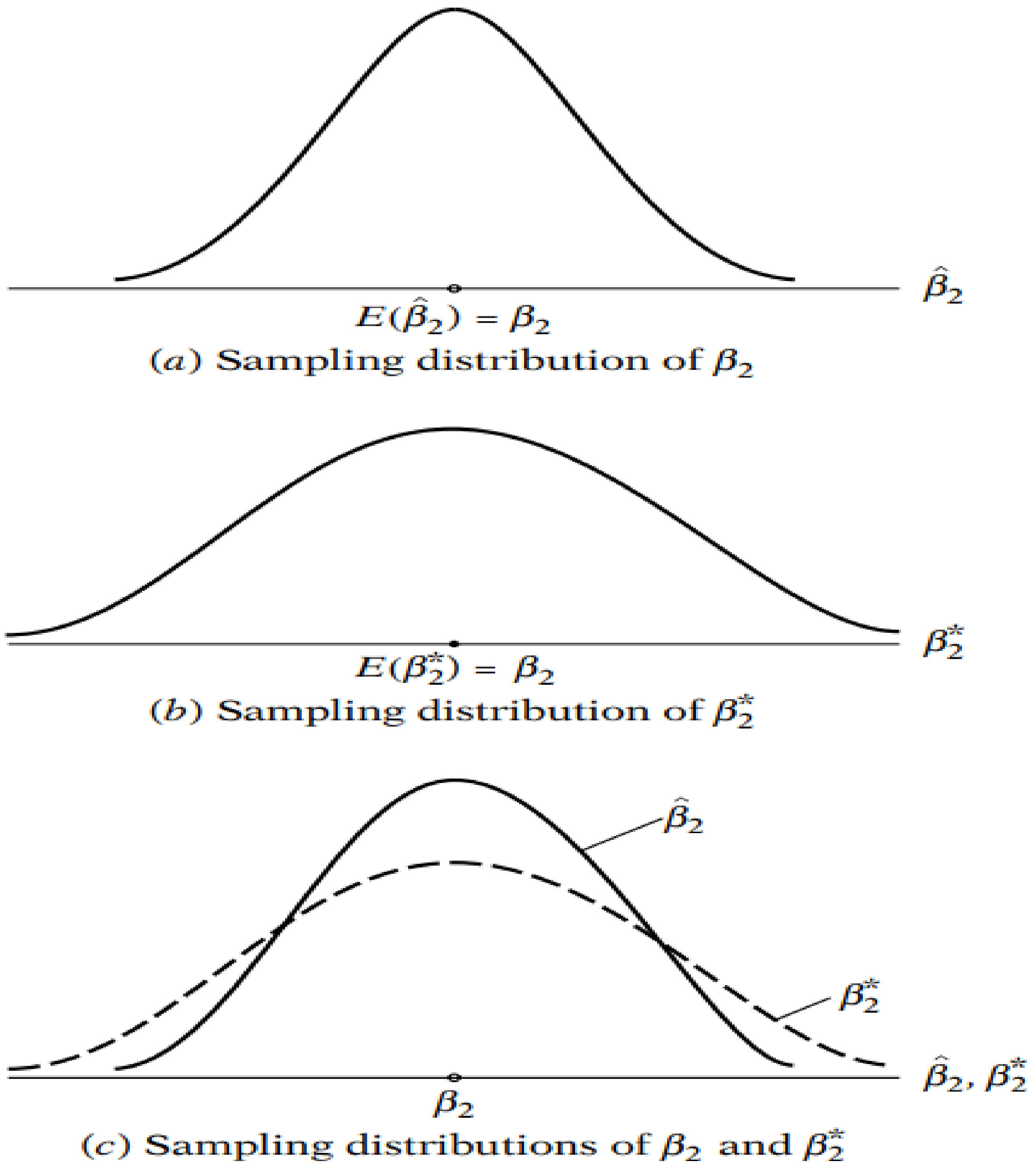
3-4. Properties of least-squares estimators: **The Gauss-Markov Theorem**

- **Gauss- Markov Theorem:**

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

It is sufficient to note here that the theorem has theoretical as well as practical importance.

Fig 3.7: BLUE



Basic Econometrics by
Damodar Gujrati, Page.80

What does **the Gauss-Markov Theorem** mean?

- **FIGURE 3.7:** Sampling distribution of OLS estimator $\hat{\beta}_2$ and alternative estimator β_2*.
- In Fig 3.7(a)we have shown the **sampling distribution of the OLS estimator** $\hat{\beta}_2$**, i.e,** the distribution of the values taken by $\hat{\beta}_2$ in repeated sampling experiments (recall Tab 3.1).
- For convenience we have assumed $\hat{\beta}_2$ to be distributed symmetrically. As the figure shows, the mean of the $\hat{\beta}_2$ values, $E(\hat{\beta}_2)$, is equal to the true β_2 . In this situation we say that $\hat{\beta}_2$ is an unbiased estimator of β_2 .

What does **the Gauss-Markov Theorem** mean?

- In Fig 3.7(b) we have shown the sampling distribution of β∗2 , an alternative estimator of β2 obtained by using another (i.e., other than OLS) method.
- For convenience, assume that $\beta_2 *$, like $\beta\hat{}_2$, is unbiased, that is, its average or expected value is equal to β2.Assume further that both ˆ β2 and β∗2 are linear estimators, that is, they are linear functions of Y. Which estimator, $\beta\hat{}_2$ or $\beta\hat{}_2 *$, would you choose?

What does **the Gauss-Markov Theorem** mean?

- To answer this question, superimpose the two figures, as in Figure 3.7(c). It is obvious that although both $\beta\text{^}_2$ and $\beta\text{^}_2$ * are unbiased the distribution of $\beta_2 *$ is more diffused or widespread around the mean value than the distribution of $\beta\text{^}_2$.
- In other words, the variance of $\beta\text{^}_2 *$ is larger than the variance of $\beta\text{^}_2$. Now given two estimators that are both linear and unbiased, one would choose the estimator with the smaller variance because it is more likely to be close to β_2 than the alternative estimator.
- In short, one would choose the BLUE estimator.
- The Gauss–Markov theorem is remarkable in that it makes no assumptions about the probability distribution of the random variable *ui , and therefore of Yi (in the next chapter we* will take this up).

What does **the Gauss-Markov Theorem** mean?

- As long as the assumptions of CLRM are satisfied, the theorem holds. As a result, we need not look for another linear unbiased estimator, for we will not find such an estimator whose variance is smaller than the OLS estimator.
- Of course, if one or more of these assumptions do not hold, the theorem is invalid. For example, if we consider nonlinear in- the-parameter regression models, we may be able to obtain estimators that may perform better than the OLS estimators.
- Also, as we will show in the chapter on heteroscedasticity, if the assumption of homoscedastic variance is not fulfilled, the OLS estimators, although unbiased and consistent, are no longer minimum variance estimators even in the class of linear estimators.

The Gauss-Markov Theorem

- The statistical properties that we have just discussed are known as **finite sample properties:**
- **These properties hold regardless of the sample size on which the estimators are** based.
- Later we will have occasions to consider the **asymptotic properties, that is, properties** that hold only if the sample size is very large (technically, infinite).

Summary

- The problem of estimation
- The meaning of the Method of ordinary least squares, minimizing the residual sum of squares
- Least Square Criterion
- Two Normal Equations after the partial derivates with respect to $\beta\text{^}_2$ and $\beta\text{^}_1$
- Deriving the formula for $\beta\text{^}_2$ and $\beta\text{^}_1$
- Numerical and statistical properties of OLS

Summary

- Why assumptions underlying the method of LS?
- Ten Assumptions underlying the method of LS
- Precision or SE of least-squares estimates of $\hat{\beta}_2$ and $\hat{\beta}_1$
- Their formula and their importance
- Gauss Markov Theorem
- BLUE Properties

References

- Chapter 3: TWO-VARIABLE REGRESSION MODEL: **The problem of Estimation, in Basic Econometrics Book by Domodar Gujarati**
- **Images taken from the worldwide web and reference given in each of these figures.**

What Next?

- Working on the Problem of Estimation in Excel
- Solutions can be checked after completing the work for verification.