Journal of Statistical and Econometric Methods, Vol 13, No 1, 2024, 1-40
ISSN: 2241-0384 (print), 2241-0376 (online)
https://doi.org/10.47260/jsem/1311
Scientific Press International Limited

Regression Analysis: A Theoretical Approach

Teshome Hailemeskel Abebe[1]

Abstract

The main objective of this document is to provide a comprehensive understanding in the area of simple regression, especially for undergraduate students majoring in economics, finance and statistics.

JEL classification numbers: C01, C013, C51, C52.
Keywords: Econometrics, single equation model, Estimation.

1. Introduction to Econometrics

From various economics courses, we have learned simply the relationships among variables. For instance, in microeconomics we learn the demand and supply models in which the quantities demanded and supplied of a good depend on its own price. In macroeconomics, we study investment function to explain the amount of aggregate investment in the economy as the rate of interest changes and consumption function that relates aggregate consumption to the level of aggregate disposable income. However, as economist we need to ask questions like if one variable changes in a certain magnitude, by how much will another variable change? In addition, given the value of one variable; can we forecast or predict the corresponding value of another? The purpose of studying the relationships among economic variables and attempting to answer such types of questions leads to the foundation of econometrics. Thus, if empirical data verify the relationship proposed by economic theory, we accept the theory as valid. If the theory is incompatible with the observed behavior, we either reject the theory or in the light of the empirical evidence of the data, modify the theory. To provide a better understanding of economic relationships and a better guidance for economic policymaking we also need to know the quantitative relationships between the different economic variables.

---

[1] Department of Economics, Ambo University, Ambo, Ethiopia.

Article Info: *Received:* November 11, 2023. *Revised:* November 28, 2023.
Published online: January 24, 2024.

Thus, we need a way to find quantitative answers to quantitative questions using econometrics. Therefore, the main objective of econometrics is to identify a causal effect of one or more variables (independent) on another variable (dependent). In addition, econometrics can test and refine economic theory. For example, economic theory may be ambiguous about the impact of a policy change; however, econometrics can evaluate the policy program. It is also true that econometric analysis is useful to appropriate decision making.

1.1 Definition of Econometrics

While econometrics may go back as far as the work of Davenant and King in the 17 century, it did not come into self-consciousness as a separate field until the foundation of the Econometric Society in 1933. The society defined econometrics as “economic theory in its relation to statistics and mathematics” and its objective as the “unification of the theoretical-quantitative and the empirical-quantitative approach to economic problems” (cited by Ragnar Frisch, 1933, p. 1).
Different economist defined econometrics from their point of view, but all of them are arriving at the same conclusions and we can boil down the whole definition in to the following.
Econometrics is a social science, which applies economics, mathematics and statistical inference to the analysis of economic phenomena i.e., combination of statistical methods, mathematics, economics, and data to answer empirical questions in economics (Arthur S. Goldberger, 1965).
Econometrics is the application of statistical and mathematical methods to the analysis of economic data with a purpose of giving empirical content to economic theories and verifying them or refuting them (Maddala, 1992).
Econometrics is also defined as statistical methods for estimating economic relationships, testing economic theories, and evaluating and implementing government and business policy (Wooldridge, 2002).
Generally, econometrics is the branch of economics that use mathematical methods and statistical tools to analyze economic relationship or economic data with a purpose of verifying or refuting economic theories through test of economic theories, estimate the coefficient of the variables and predict or forecast the future the value.

1.2 Scope of Econometrics

Why econometrics is a separate discipline?
Econometrics: is an amalgamation/integration of economic theory, economic statistics, mathematical and statistics.
Thus, in a simple circular flow we can depict the relationship as below:

- **Economic theory** it is an observation from the real-world phenomena. However, the complexity of the real-world economy makes us impossible to understand all interrelationships at once. Thus, why we build our economic theory based on ceteris paribus assumptions. Therefore, economic theory simply explains economic relationship using ceteris paribus assumptions.

Example: consumption depend on current income (Y_t) and previous income (Y_{t-1}) of an individual other thing being equal. This theory does not give any insight how current income and previous income will affect consumption by giving numerical values.

- **Economic statistics** is concerned with descriptive statistics (is mainly concerned with collecting, processing, and presenting economic data in the form of charts and tables). It attempts to describe the pattern in their development over time and perhaps detect some relationship between various economic magnitudes. It does not provide explanations of the development of the various variables and it does not provide measurements the coefficients of economic relationships.

Example: it describe economic relationship using numerical data like mean, median, standard division etc., but it does not make reliable the relationship between variables.

- **Mathematical economics** is to express economic theory in equation/mathematical form without empirical verification of the theory. Hence, mathematical economics states economic theory in terms of mathematical symbols, i.e., there is no essential difference between mathematical economics and economic theory since both states the same relationships. However, economic theory uses verbal exposition. Both express economic relationships in an exact or deterministic form, i.e., neither mathematical economics nor economic theory allows for random elements, which might affect the relationship and make it stochastic. Furthermore, they do not provide numerical values for the coefficients of economic relationships.

Example: $\mathrm{C_t = \beta_0 + \beta_1 Y_t + \beta_2 Y_{t-1}}$ (1.1)

where C_t; consumption, Y_t current income, Y_{t-1} previous income

Thus, mathematical equation explain the relationship between dependent variable (C_t) and independent variables (Y_t & Y_{t-1}) by ignoring other variables that affect consumption.

- **Mathematical statistics** it deals with statistics from mathematical point of view using probability theory. Mathematical (or inferential) statistics deals with the method of measurement that is developed based on controlled experiments. But statistical methods of measurement are not appropriate for a number of economic relationships because for most economic relationships controlled (carefully planned) experiments cannot be designed due to the fact that the nature of relationships among economic variables are stochastic or random. Yet the fundamental ideas of inferential statistics are applicable in econometrics, but they must be adapted to the problem of economic life. Econometric methods are adjusted so that they may become appropriate for the measurement of economic relationships, which are stochastic. The adjustment consists primarily in specifying the stochastic (random) elements that are supposed to operate in the real world and enter into the determination of the observed data.

In all of the above methods, they completely ignore the other factors that will affect the economic relationships, but econometrics captures all other factor that affect economic relationship through random term given as:

$$C_t = \underbrace{\beta_0 + \beta_1 Y_t + \beta_2 Y_{t-1}}_{Deterministic\ component} + \underbrace{\varepsilon_t}_{Stochastic\ component} \tag{1.2}$$

where β_0, β_1 & β_2 are unknown but fixed parameters and are called the regression coefficients. The parameter β_0 is the intercept coefficient and β_1 & β_2, are the slope coefficient. ε_t is random term which represents all other factor that will affect consumption. These factor may be many such as wealth, tradition, invention of new product etc. thus, econometrics by considering other factors (denoted by ε_t) will find numerical value for the coefficients of the variables that will explain the relationship to verify economic theories.

What is a 'Model'? Difference between Economic and Econometrics model

Model is a simplified representation of real world process. However, the choice of a simple model to explain a complex real-world phenomenon leads to two criticisms: such as the model is over simplified and the assumptions are unrealistic (Maddala, 1992).

Model is neither a hypothesis nor a theory (Levins R., 1966). Unlike scientific hypotheses, a model is not verifiable directly by an experiment. For all models of true or false, the validation of a model is not that it is true" but that it generates good testable hypotheses relevant to important problems."

The meaning of a *model* in Oxford Advanced Learner's dictionary is a simple description of a system, which used for explaining how something works or calculating what might happen (Hornby, 2000).

Economic model Vs. Econometrics

An economic model is a set of assumptions that approximately describes the behavior of an economy (or a sector of an economy).

***Specifically, an economic model* focuses on**:

- "how" and "why"
- Indicates only the relationship between variables
- The sign or partial derivatives

***Econometrics* focuses on:**

- "How much" and "by how much"
- Specification of the functional form of the relationship
- The specification of the necessary time lags
- Specification of the stochastic characteristics of the system (the probability distribution of the disturbances

Therefore, stochastic properties are the basic determinants of behavior of economic variables. Since economic variables are not experimental, rather than observational (behavioral), they are random or stochastic. If the variables are stochastic, economic model cannot handle, rather econometrics model which capture random effect is an appropriate model.

Thus, the existence of stochastic relationship helps us to:

- Testing economic theory, estimate the parameters and test hypotheses about them
- Use the relationship for forecasting and policy controls

These stochastic properties of economic variables are also important for construction of econometric models, interpretation of the findings and forecasting. Econometricians use the scientific way of thinking to develop a new econometric model or a theory to explain the economic system.

Example:

Economist: "If the government increases alcohol excise tax, consumers will cut down on their alcohol consumption."

Econometrician: "If the government increases alcohol excise tax by 20%, consumers will reduce their alcohol consumption by 1%. Thus, econometrics is vital in applying economic theories in practice.

Thus, econometrics differs from mathematical economics in that econometrics assumes random relationships among economic variables. Econometric methods are designed to take into account random disturbances, which relate deviations from exact behavioral patterns suggested by economic theory and mathematical economics. Furthermore, econometric methods provide numerical values of the coefficients of economic relationships.

Therefore, we can distinguish two types of relationships:

- Deterministic relationship which is expressed using mathematical model
- Statistical relationship which does not give unique values for Y for a given values of X, but can be described exactly in probabilistic terms.

2. The Goals of Econometrics

The end goals of econometrics are:

- Estimating the coefficient of economic relationship (estimation of parameter values)
- Hypothesis testing/Testing of economic theory
- Forecasting the future values of economic magnitude

2.1 Estimating

This mean by applying different methods of econometric techniques we can obtain individual numerical values for the coefficients of economic relationship. Using these numerical values, a decision can be undertaken by different economic agents. Econometrics can supply MPC, elasticity's, MC, MR etc. using these magnitudes (numerical values) decision will be undertaken.

Example:

$$D = \beta_0 + \beta_1 I + \beta_2 Ex + \beta_3 PI + \beta_4 PEx + \varepsilon_t \tag{1.3}$$

where D= devaluation, I is volume of import, Ex is volume of export, PI is price of import, PEx is price of export, then devaluation will depend on all these explanatory

variable coefficients. From these coefficient we can have
β_1 denotes marginal propensity to import
β_2= marginal propensity to export
β_3 & β_4 are marginal propensity to price of import and export, respectively
Then based on these coefficients of numerical values, the government will decide whether devaluation will illuminate the countries deficit or not. Generally, it is used for ***evaluating*** government and business policy.

2.2 Testing of economic theories/analysis

This concern with testing hypothesis, that is testing the significance of the coefficient and verification of economic theories and thereby knows and decides how well they explain the observed behavior of the economic units.

2.3 Forecasting

It means using the numerical values of the coefficients of economic relationships, we can judge whether to take any policy measure in order to influence the future value of economic variables or not. In other words, it refers to explaining and predicting the future changes of economic phenomena based on historical data.
Assuming that the estimated results from the Ethiopian economy for the year 1985-1995

$$\hat{Y} = -261.09 + 0.2453X_i \tag{1.4}$$

where Y consumption expenditure, X_i personal disposable income, then on the bases of the above results the government can able to know his expenditure in any year after 1995 using the above equation. If disposable income X_i will be one million in 1999, then expenditure on imported goods will be $\hat{Y} = -261.09 + 0.2453(1{,}000{,}000) = 245038.91$ by the year 1999.
Then since the government knows the future values of expenditure on imported goods and services, it can take any measure to increase or cut down imports using these numerical values. Forecasting is used for both developed and developing countries in different ways, i.e., developed countries used if it for regulation of their economies whereas developing countries used it for planning purpose.

2.4 Branch of Econometrics

Just like any subject econometrics also decomposed in two branches: theoretical and applied econometrics.
Theoretical econometrics: it is the development of appropriate econometric methods for measuring economic relationship between variables in theoretical econometrics.
The data used for measurement purpose are observations from the real world, but are not derived from control experiment. Moreover, econometric relationships are not exact
The econometric method that will be used in the theoretical econometrics may be

classified in to two:

- Single equation techniques, i.e., one side relationship between variable at a time.

Example:

$$Q_d = \beta_0 + \beta_1 P_i + \varepsilon_t \tag{1.5}$$

Means quantity demand depends up on the price of the commodity but not price depends up on quantity. Then we have on side causations. Then we can apply econometrics technique only for this equation.

- Simultaneous equation model: when there is two side causation. Example equation (1.5) explains that quantity demand depends on the price of the commodity but if the price of the commodity is in turn depends on the quantity of commodity supplied then we will have two side equation.

$$P_i = \beta_0 + \beta_1 Q_s + \varepsilon_t \tag{1.6}$$

Econometric techniques will applied for three equations:

$$Q_d = \alpha_0 + \alpha_1 P_i + \varepsilon_t \quad demand\ equation \tag{1.7}$$

$$Q_s = \beta_0 + \beta_1 P_i + \varepsilon_t \quad supply\ \ equation \tag{1.8}$$

$$Q_d = Q_s \quad identity \tag{1.9}$$

Then in this case, we applied econometrics techniques simultaneously for all equations at a time.

Applied econometrics: refers to the application of theoretical econometrics method to specific branch of economic theory i.e. application of theoretical econometrics for verification and forecasting of demand, cost, supply, production, investment, consumption and other related field of economic theory.

3. Methodology of Econometrics

Econometric research is concerned with the measurement of the parameters of economic relationships and with the predication of the true values of economic variables. Broadly speaking, starting with the postulated theoretical relationships among economic variables, econometric research or inquiry generally proceeds along the following stages:

1. Statement of economic theory or hypothesis
2. Specification of the mathematical model of the theory
3. Specification of the econometric model
4. Collecting the data
5. Estimation of the parameters of the econometric model
6. Evaluation of estimates (Model diagnostic and Hypotheses testing)
7. Forecasting or prediction
8. Evaluation of the forecasting accuracy of the model
9. Using the model for control or policy purposes

To illustrate the preceding steps, let us consider the well-known Keynesian theory of consumption function.

3.1 Statement of theory or hypothesis

The first step in econometrics methodology is identification of the economic relationship, i.e., economic theory tells us about the relationship between two or more variables.

Example 1: Keynes stated consumption increases as income increases, but not as much as the increase in income". It means that "the marginal propensity to consume (MPC) for a unit change in income is greater than zero but less than unit".

3.2 Specification of the mathematical model of the theory

This refers to the transformation of economic theory into mathematical model that explain the relationship between economic variables. Under this stage, we will have the following:

- **Selection of variables**: it involves determining the dependent (endogenous or explained) variable and independent (exogenous or explanatory) variables of the theory. In short, it is expected to choose the dependent variable and independent variables and how they should be measured.

Example: consumption of an individual at time t is dependent variable and disposable income are independent variables.

- **Determining the theoretical values**: refers to a prior expectation of the sign and magnitude of the parameters. This needs only a theoretical background to determine the relationship between the dependent and independent variables, i.e., negative or positive relationship between variables. From our example, we can have the following sign or direction of relationship between variables.

Example: the relationship between consumption at time t and income at time t has positive relationship.

Demand equation for a final consumption good can be defined as:

$$Q_d \quad = f\left(\overset{-}{P}, \overset{+}{Y}, \overset{-}{P_c}, \overset{+}{P_s}\right)$$

where p is own price, Y is income, P_c & P_s are price of complementary and substitute goods.

The signs in the above variables indicate the hypothesized sign of the respective regression coefficient in a linear model.

- **Specification of the model:** In this stage, we specify the relationship between the dependent and independent variables based on economic theories. We also determine the number of equations (single equation or simultaneous equation model) & the type of equation i.e. whether the relationship between economic variables explained using linear or non-linear equations. Let us specify our previous theoretical relationship.

Example 1: $$C_t = \beta_0 + \beta_1 Y_t \quad (1.10)$$

where C_t aggregate consumption at time t, Y_t aggregate income at time t

Example 2: $$C_t = \beta_0 Y_t^{\ \beta_1} r_t^{\ \beta_2} \quad (1.11)$$

where C_t aggregate consumption at time t, Y_t aggregate income at time t, r_t is

future rate of return.
All the above equation is single equation model but equation 1.10 is linear equations & equation 1.11 is non- linear equation. Magnitude of the coefficient of the variables ($\beta_0,\ \beta_1,\ \beta_2$), what will be the likely magnitude of these coefficients? The magnitude or size of the numerical values of the coefficient of the variable ($\beta_0,\ \beta_1$, β_2) are determined by the economic theory & empirical observation of the real world. In equation 1.10 & 1.11 the coefficient of β_1 refers to marginal propensity to consume and determined by economic theory. The explanation of equation 1.10 is differ from equation 1.11 since the former is linear while the latter is non-linear. Example in equation 1.10 if income increases by 1 birr on the average consumption will increase by β_1 amount. But in equation 1.11, β_1 & β_2 explains elasticities, i.e., if income increases by 1% consumption will increase on average by $\beta_1\%$ and for β_2 if rate of interest is increasing by 1% consumption will be cut down on the average by $\beta_2\%$.
Therefore, under mathematical model we cannot easily estimate the value of the parameter. Since the data never perfectly fit the mathematical model. There is uncertainty.
Thus, statistics is used to determine the value of β given the model, the data and the uncertainty. Hence, we need an econometrics model which incorporate uncertainty using statistical concept.

3.3 Specification of the econometric model of the theory

The relationships between economic variables that we express above are generally exact. In addition to income, other variables may affect consumption expenditure. For example, sizes of family, ages of family members, family religion, etc., are likely to exert some influence on consumption.
Specification of the econometric model will be based on knowledge of economic theory and on any available information related to the phenomena under investigation.
To allow for the random relationships between economic variables, equation (1.10) is written as:

$$\mathrm{Y}_I = \overbrace{\beta_0 + \beta_1 \mathrm{X}_I}^{\text{Deterministic component}} + \overbrace{\varepsilon_t}^{\text{stochastic component}} \tag{1.12}$$

where Y = consumption expenditure (as dependent variable), X = income (independent or explanatory variable), β_0 = the intercept coefficient, β_1 = the slope coefficient which indicates the MPC in our above example, and ε_t is disturbance term or error term. It is a random or stochastic variable, which is unobservable and has well-defined probabilistic properties. The disturbance term ε_t may well represent all those factors that affect consumption, but are not taken into account explicitly.
Thus, the inclusion of ε_t in mathematical economics model (in exact relationship between variables) will transform the model into econometric model (inexact relationship between variables since ε capture unexplained variables). The most

common errors of specification are:

- Omissions of some important variables from the function
- Inclusion of irrelevant variable
- Wrong specification of the model (for example, instead of simultaneous equations model, apply single equation model, linear model if the relationship is non-linear and viscera).

4. Obtaining data

4.1 Types of Data

- **Cross-Sectional Data**

A cross-sectional data: data collected on one or more variables collected at particular period of time. Example: a sample of individuals, households, firms, cities, states, countries, or a variety of other units, taken at a given point in time. We often assume that these data have been obtained by random sampling.

- **Time Series Data**

A time series dataset consists of observations on a variable or several variables over several periods of time (days, weeks, months, years).
A key feature of time series data is that, typically, observations are correlated across time which results the absence of random sample. This time correlation introduces very important issues in the estimation and testing of time series econometric models.

4.2 Pooled cross sections

Pooled data occur when we have a time series of cross sections, but the observations in each cross section do not necessarily refer to the same unit.

- **Panel or Longitudinal Data**

In panel data, we have a group of individuals (or households, firms, countries, etc.) who are observed at several points in time.
That is, we have time series data for each individual in the sample. That is it consists of a time series for each cross-sectional member in the data set. The key feature of panel data that distinguishes them from pooled cross sections is that the same individuals are followed over a given period of time.
Thus, panel data refers to samples of the same cross-sectional units observed at multiple points in time. In pane data, we cannot assume that the observations are independently distinguished across time and serial correlation of regression residuals becomes an issue. A panel-data observation has two dimensions: $x_{i,t}$, where i runs from 1 to N and denotes the cross-sectional unit and t runs from 1 to T and denotes the time of the observation.

4.3 Problems in accuracy of data

However, plenty of data are available for research purpose but the quality of data matter in arriving at a good result. The quality of data may not be good for different reason:

- Since most social science data are not experimental in nature, so omission error may occur
- An approximate & round off the numbers will have error of measurement
- In questioner, non-response error may be occur
- Economic data are available at aggregate level & errors may be committed in aggregation

Because of the above reasons, one can deduce that the results obtained by any researchers are highly depending up on the quality of the data. Then if you get unsatisfactory results the reason may be the quality of the data if you correctly specifying the model.

5. Estimating the parameters of econometric model

Regression analysis is the main tool used to obtain the estimates. However, we need to choice appropriate economic techniques for estimation, i.e. to decide a specific econometric method to be applied in the estimation. Using this technique and the data given in Table 1.2, we obtain the following estimates of the parameters denoted by Greek letters, β_0 and β_1, namely, 231.8 and 0.7194, respectively. Parameters are unknown quantities that characterize a model. Thus, the estimated consumption functions were obtained using OLS methods of estimation as given below:

$$\hat{Y} = 231.8 + 0.7194\mathrm{X} \tag{1.13}$$

Where: MPC is about 0.72 and it means that for the sample period when real income increases by 1 USD results on average increases in real consumption expenditure about 72 cents.

<u>Note</u>: A hat symbol (^) on the above equation signify an estimator of the relevant population value.

However, the selection of the methods of estimation depends upon many factors.

The nature of relationship between economic variables and their identification:

If we studied the econometric relationship using a **single equation**, the estimation methods are ordinary least square (OLS), Maximum likelihood (ML), methods of moment, Mixed estimation Technique, etc.

Example of single equation model:

$$Q_{di} = \beta_0 + \beta_1 P_i + \varepsilon_t \tag{1.14}$$

Where: Q_d is quantity demand, P_i is price. In this case, OLS is the best method. However, if the relationships between economic variables are in a function of **simultaneous equation**: indirect leas square (ILS or reduced form techniques), two stage least square (2SLS), three stage least square (3SLS) & the Full information maximum likelihood (FIML) methods are used.

- On the properties of estimated coefficient obtained from each method is that a good estimate should give the properties of unbiasedness, consistency, efficiency, & sufficiency or a desirable characteristics than any other estimates from other methods, then that techniques, which possess more of the desirable characteristics, will be selected.

- On the purpose of econometric research: if the purpose of the model is forecasting the property of minimum variance is very important, i.e. the techniques, which will give the minimum variance of the coefficients of the variables, will be selected. However, if the purpose of the research is policymaking (analysis) that techniques which gives unbiasedness of the variable will be selected.
- On the simplicity of technique: if our interest is simply computation, we can select that technique which involves simple computation & less data requirement.
- Time and cost required for computation of the coefficients of the variables may determine the selection of estimation methods.
- The linearity and non-linearity in variables as well as in parameters.

6. Evaluation of estimates (Model diagnostic and hypothesis testing)

After we have estimated the values of the parameters, we need to evaluate the accuracy of the model. At this stage, we are evaluating the reliability of the results whether they are theoretically meaningful & statistically satisfactory results. Thus, confirmation or refutation of economic theories based on sample evidence is object of statistical analysis.

To evaluate the reliability of the estimates we follow the following steps:

- **Economic prior criterion**: economic interpretation of the results

In this stage, we should confirm that whether the estimated values explain the economic theory or not i.e. it refers to the sign & magnitude of the estimated coefficients of the variables.

Example: if we have the following consumption function:

$$C_i = \beta_0 + \beta_1 Y_i + \varepsilon_i \tag{1.15}$$

where C_i is consumption expenditure, Y_i is income

From economic theory (economic relationship between consumption & income) it is known that β_1 denotes marginal propensity to consume (MPC). Then on the base of a prior economic criteria, it is determined that the sign of β has to be positive & the magnitude (size) β again is in between zero & one $(0 < \beta < 1)$. If the estimated results of the above consumption function gives

$$\hat{C}_i = -3.32 + 0.2033 Y_i \tag{1.16}$$

From economic relationship explained by economic theory states that if your income increases by 1 birr your consumption will increase on average by less than one birr i.e. 0.203 cents. Then the value of β_1 is less than one & greater than zero in its magnitude (size) again the sign of β_1 is positive. Therefore, the estimated models explain the economic theory (economic relationship between consumption & income) or satisfies the prior-economic criteria. If another estimation of the model using other data gives the following estimated results.

$$\hat{C}_i = -24.45 + -5.091Y_i \tag{1.17}$$

where C_i is consumption expenditure, Y_i is income. From economic theory it is known that β_1 has to be positive & its magnitude is greater than zero & less than one. However, the estimated model results that the sign of β_1 is negative & its magnitude is greater than one in absolute value then we reject the model because the results are contradictory or do not confirm the economic theory.

In the evaluation of estimates of the model, we should take into consideration the sign & magnitudes of the estimated coefficients. If the sign and magnitude of the parameter do not confirm the economic relationship between variables explained by the economic theory then the model will be rejected. However, if there is a good reason to accept the model then the reason should be clearly stated. In general, if the prior theoretical criteria's are not satisfied, then the estimates should be considered as unsatisfactory. In most of the cases, the deficiencies of empirical data utilized for the estimation of the model are responsible for the occurrence of wrong sign or size of the estimated parameters. The deficiency of the empirical data means either the sample observation may not represents the population (due to sampling procedure problem or collecting inadequate data or some assumption of the method employed are violated). In general, if a priority criterion is not satisfied, the estimates should be considered as unsatisfactory.

- **Statistical criterion**/ first order criteria: statistical interpretation of the results

If the model passes prior-economic criteria, the reliability of the estimate of the parameters will be evaluated using statistical criteria. The most widely statistical criteria are:

1. The correlation coefficient –R^2/r^2
2. The standard error/deviation / S.E of the estimate
3. t-ratio or t-test and F-test of the estimates

Since the estimated value is obtained from a sample of observations taken from the population, the statistical test of the estimated values will help to find out how accurate these estimates are (how they accurately explain the population?).

- ✓ R^2 explain that the percentage of the total variation of the dependent variable explained by the change of the explanatory variables(how much % of the dependent variable is explained by the explanatory variables). So to say the result is statistically significant its value should be larger (more than 60%).
- ✓ $S.E$ (Standard error or deviation): measures the dispersion of the sample estimates around the true population parameters. The lower the standard error, the higher the reliability (the sample estimates are closer to the population parameter) of the estimates & vice-versa.
- ✓ t-ratio or t-test of the estimates shows the significance of the individual variable coefficients. The higher the t- test, the higher the significance level.

- **Econometric criterion**/second order condition:

After conducting a prior test & statistical test, the investigator should check the reliability of the estimates whether the econometric assumptions (example, non-

normality, heteroscedasticity, autocorrelation, etc.) are holds true or not.
If any one of the assumption of econometrics are violated.

- The estimates of the parameters cease to possess some of the desirable properties (unbiasedness, consistency, sufficiency etc.)
- The statistical criteria losses their validity & become unreliable

If the assumptions of econometric techniques are violated then the researcher has to re-specifying the already utilized model. To do so the researcher introduce additional variable in to the model or omit some variables from the model or transform the original variables etc. by re-specify the model the investigator proceeds with re-estimation & re-application of all the tests (a prior, statistical & econometric) until the estimates satisfies all the tests.
Thus, we are expected to conduct residual analysis (the stochastic assumptions) by looking at a plot of the residuals in graphical form as well as formal test like using a normality test, autocorrelation test, heteroscedasticity tests, etc. Moreover, if we have more than one explanatory variable we should conduct multicollinearity test.

7. Forecasting or prediction

Forecasting is one of the primary aims of econometric research. Especially, in time series data, the objective is to forecast the future value of the series based on the observed value of historical data given the model that we have used in the estimation process. The estimated model may economically meaningful, statistically & econometrically correct for the sample period. However, given all these, it may not have a good power of forecasting due to the inaccuracy of the explanatory variables & deficiency of the data used in obtaining the estimated values.
If this happens, the estimated value (i.e. forecasted) should be compared with the actual realized value magnitude of the relevant dependent variable. The difference between the actual & forecasted value is tested statistically. If the difference is significant, we conclude that the forecasting power of the model is poor. If it is statistically insignificant, the forecasting power of the model is good.
To illustrate, suppose we want to predict the mean consumption expenditure for 1994, income value for 1994 was 6000 billion dollars. What is the forecasted consumption expenditure?
The predicted value of Y is

$$\hat{Y} = 231.8 + 0.7194(6000) = 4084.6 \qquad (1.18)$$

The actual value of consumption expenditure reported in 1994 was 4000 billion dollars (see Table 1.1). The estimated model (1.17) over-predicted the actual consumption expenditure by 84.6 billion dollars. We could say the forecast error is about 84.6 billion dollars. By fiscal and monetary policy, government can manipulate the control variable X (in our case income) to get the desired level of target variable Y(in our case consumption).

8. Evaluation of the forecasting accuracy of the model

Forecasting is one of the aims of econometric research. However, before applying the model that used in forecasting the future values of the series for other purpose like policy control, it should be economically meaningful and statistically and econometrically valid or correct for the sample period for which the model that has been used in forecasting purpose. Thus, we use statistical loss functions, which are based on the mean forecast error. The forecast error for l lead times is defined as the difference between the actual forecast and its conditional expectation and can be expressed as:

$$\varepsilon_i(l) = \sigma_{i+l}^2 - \hat{\sigma}_i{}^2(l) \qquad (1.19)$$

The most common loss functions: the Mean Square Error (MSE) and the Mean Absolute Error (MAE), are defined as:

$$\text{MSE} = \frac{1}{\text{n}}\sum_{\text{i}=1}^{\text{n}}\left(\sigma_{i+l}^2 - \hat{\sigma}_i{}^2(l)\right)^2 \qquad (1.20)$$

$$\text{MAE} = \frac{1}{\text{n}}\sum_{\text{i}=1}^{\text{n}}\left|\sigma_{i+l}^2 - \hat{\sigma}_i{}^2(l)\right| \qquad (1.21)$$

where σ_{i+l}^2 refers to the $l - lead$ actual level of conditional variance and $\hat{\sigma}_i{}^2(l)$ refers to the $l - lead$ forecasted level of the conditional variance.

8.1 Using model for control or policy purposes

After the validity of the model is checked, it can be used to evaluate or control the government policy through manipulation of the objective variables given reliable model and forecasts of the economic phenomena.

For example, the government believes that 5000 level of consumption level keeps inflation rate at 10%. Thus, $Y = 5000 = -231.8 + 0.7194\,X$, when X=7266. Given the estimate of MPC = 0.72, an income of $7266 Bill will produce an expenditure of $5000 Bill.

By fiscal and monetary policy, government can manipulate the control variable X to get the desired level of target variable Y.

Simple Linear Regression

Regression analysis consists of techniques for modeling the relationship between a dependent variable (also called response variable) and one or more independent variables (also known as explanatory variables or predictors). In regression, the dependent variable is modeled as a function of independent variables, corresponding regression parameters (coefficients), and a random error term, which represents variation in the dependent variable unexplained, by the function of the dependent variables and coefficients. In linear regression, the dependent variable is modeled as a linear function of a set of regression parameters and a random error. The parameters need to be estimated so that the model gives the best fit to the data. The parameters are estimated based on predefined criterion.

The most commonly used criterion is the least squares method, but other criteria have also been used that will result in different estimators of the regression parameters. The statistical properties of the estimator derived using different criteria will be different from the estimator using the least squares principle. In this article the least squares principle will be utilized to derive estimates of the regression parameters. If a regression model adequately reflects the true relationship between the response variable and independent variables, this model can be used for predicting dependent variable, identifying important independent variables, and establishing desired causal relationship between the response variable and independent variables.

Therefore, in this study, we cover both simple linear regression and multiple linear regression.

Simple linear regression analysis explains the relationship between a single dependent (Endogenous) variable, Y, as a function of a single independent (exogenous) variable, X. Under the macroeconomic theory of income hypothesis, consumption is a function of current income other things remain constant (ceteris paribus) given a linear relationship. However, consumption expenditure is not determined only by income, since it can be affected by previous income, tradition, wealth etc. Then this inexact relationship in a simple linear regression model is captured by random term (ε_t) given as follows:

$$C_i = \beta_0 + \beta_1 Y_i + \varepsilon_t$$

where C_i is consumption (as dependent variable) for the ith observations, Y_i is disposable income (as explanatory variable), β_0 & β_1 are coefficients or regression parameters, and ε_i stochastic disturbance or error term.

We introduce the stochastic error term (ε_i) in the regression model because we cannot capture every influence on a dependent variable in the model. Specifically, the stochastic error term can captures the effect of omitted variables, erratic nature of human being, misspecification of mathematical model, measurement error in the response and independent variables, errors in aggregation and sampling error, etc.

The population regression function (PRF) relates the conditional mean with the independent variable is given by:

$$\mathrm{Y_i} = \beta_0 + \beta_1 X_i + \varepsilon_i \quad , i = 1, 2, \ldots, n \tag{2.1}$$

where β_0 and β_1 are unknown but fixed parameters known as Y intercept and slope coefficients, respectively. Y is the dependent variable, X is the independent variable, and ε_i is an independently identically distributed (i.i.d.) random error term that surrogate for all variables that are omitted from the model but they collectively affect Y and it is non-systematic component, *i* =1, …,*n* denote a random sample of size *n* from the population. It is usually assumed that error ε is normally distributed with $E(\varepsilon) = 0$ and a constant variance $Var(\varepsilon) = \sigma^2$ in the simple linear regression.

In practice, however, we do not have directly data on the population, so we rely on the sample. Thus, the sample counterpart of the population regression function is the *sample regression function* (*SRF*). Therefore, when we substitute the estimates of the parameters $\hat{\beta}_0$ and $\hat{\beta}_1$ in population regression function, we obtain the sample regression function as given by:

$$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\varepsilon}_i \tag{2.2}$$

where $\hat{\beta}_0$ and $\hat{\beta}_1$ are estimators of the unknown but fixed parameters, $\hat{\varepsilon}_i$ denotes the (sample) residual term which is an estimator of population disturbance term and $\hat{Y}_i$ is estimator of $\mathrm{E}(\mathrm{Y_i}/X_i)$.

The primary objective of regression analysis is to estimate the unknown population regression coefficients (β_0 and β_1) on the basis of the SRF coefficients ($\hat{\beta}_0$ and $\hat{\beta}_1$) given n observations on Y and X. To do so, we need to specify basic assumptions of the model.

8.2 Gauss-Markov Assumptions

In general, we categorized the assumption of simple linear regression model into:

- Assumption of the parameters of the model
- Assumption about the distribution of error term (ε_t)
- Assumption of the relationship between error term (ε_t) and explanatory variable (X's)
- Assumptions of the dependent variable (Y)

1) The regression model is linear in the parameters

The classical assumed that the model should be linear in the parameters regardless of whether the relationship between explanatory variable and the dependent variable is linear or not. Linear in the parameters means that the parameters are not multiplied together, divided, squared or cubed.

2) The Explanatory variable (X) are fixed in repeated sampling

The value taken by independent variable is considered to be fixed in repeated samples. That is, the regressor is assumed to be non-stochastic ((i.e., when repeating the experiment, choose exactly the same set of X values on each occasion so that they remain unchanged). However, having a larger spread of values (i.e. a larger variance) of the explanatory variable (X) in the sample improves the accuracy of estimation.

3) The observed data represent a random sample from the population described by the model.
4) The number of observations is greater than the number of parameters to be estimated, usually written n > k.
5) The error term (ε) is a random variable and its mean value in any particular period is zero

$$E(\varepsilon_i|X_i) = E(\varepsilon_i) = 0$$

This means, for each value of X, the random variables **(**ε_i) may assume various values, some greater than zero and some smaller than zero, but if we considered all the possible values of the random variables **(**ε_i) for any given value of X, they should have on average value equal to zero. In other words, the positive and negative values of the random variables **(**ε_i**)** cancel each other. Indirectly, the zero mean of the disturbances implies that no relevant regressors have been omitted from the model.

6) Constant variance (Homoscedasticity)

Under the classical regression assumptions, the conditional variance of the error term is constant (homoscedastic) and does not vary as a function of the explanatory variable. The error variance is a measure of model uncertainty, while homoscedasticity implies the model uncertainty is identical across observations.

$$\mathrm{Var}(\varepsilon_i|X_i) = \mathrm{E}[\varepsilon_i - \mathrm{E}(\varepsilon_i)/X_i]^2$$

By assumption 5, $\mathrm{E}(\varepsilon_i|X_i) = 0$, then

$\mathrm{Var}(\varepsilon_i) = \mathrm{E}({\varepsilon_i}^2|X_i) = \sigma^2, \ \forall_i = 1, 2, \dots, \mathrm{n}$

7) No Autocorrelation

This means the value of the random terms assumed in one observation does not depend on the value, which it assumed in any other observation. That is, the random terms of different observations (ε_i and ε_j) are, independent.

$$Cov(\varepsilon_i, \varepsilon_j) = 0, \qquad \forall, i \neq j$$

Given any two X values, X_i and X_j ($i \neq$ j), the correlation between any ε_i and ε_j ($i \neq$ j) is zero.

$$Cov(\varepsilon_i, \varepsilon_j|X_iX_j) = E\{[\varepsilon_i - E(\varepsilon_i|X_i)][\varepsilon_j - E(\varepsilon_j|X_j)]\}$$

By assumption 5, $E(\varepsilon_i|X_i)\ and\ E(\varepsilon_j|X_j) = 0$, then

$$Cov(\varepsilon_i, \varepsilon_j|X_iX_j) = E\{[\varepsilon_i - 0][\varepsilon_j - 0]\}$$

Given an independently, identically distributed (i.i.d.) error term,

$$Cov(\varepsilon_i, \varepsilon_j|X_iX_j) = E[\varepsilon_i]E[\varepsilon_j]$$

$$Cov(\varepsilon_i, \varepsilon_j|X_iX_j) = Cov(\varepsilon_i, \varepsilon_j) = E[\varepsilon_i]E[\varepsilon_j] = 0$$

8) The random variables (ε_i) is independent of the explanatory variables.

This means ε_i and X's are not moving together or zero covariance between X_i & ε_i.

$Cov(\varepsilon_i, X_i) = E\{[\varepsilon_i - E(\varepsilon_i)][X_i - E(X_i]\}, by\ assumption\ \ E(\varepsilon_i) = 0$

$$= E[\varepsilon_i(X_i - E(X_i)]$$

$= E(\varepsilon_i X_i) - E(X_i E(\varepsilon_i)),\ \ E(X_i) = X_i\ is\ nonstochastic$ and $E(\varepsilon_i) = 0$

$$= E(\varepsilon_i X_i) = 0$$

$= X_i E(\varepsilon_i) = 0$, since X_i is non-stochastic

9) Assumption about the dependent variable Y_i

The response variable is normally distributed, i.e.

$$Y_i \sim N(\beta_0 + \beta_1 X_i, \qquad \sigma^2)$$

Proof:

Mean: $E(Y_i) = E(\beta_0 + \beta_1 X_i + \varepsilon_i)$

$$= \beta_0 + \beta_1 X_i \text{ since } E(\varepsilon_i) = 0$$

Variance: $\mathrm{var}(Y_i) = E(Y_i - E(Y_i))^2$

$$= E(\beta_0 + \beta_1 X_i + \varepsilon_i - (\beta_0 + \beta_1 X_i))^2$$

$$= E(\varepsilon_i)^2$$

$$= \sigma^2 \text{ since } E(\varepsilon_i)^2 = \sigma^2$$

The shape of the distribution of Y_i is determined by the shape of the distribution of ε_i. Since $\beta_0\ and\ \beta_1$ are being constant, they don't affect the distribution of Y_i. Furthermore, the values of the explanatory variable X are a set of fixed value by

assumption 2 and therefore, it doesn't affect the shape of the distribution of Y_i.
The other assumption on the dependent variable is, the successive values of the dependent variable are independent, i.e. $cov(Y_i, Y_j) = 0$

Proof:

$$cov(Y_i, Y_j) = E\{[Y_i - E(Y_i)][Y_j - E(Y_j)]\}$$

$$= E\{[\beta_0 + \beta_1 X_i + \varepsilon_i - E(\beta_0 + \beta_1 X_i + \varepsilon_i)][\beta_0 + \beta_1 X_j + \varepsilon_j - E(\beta_0 + \beta_1 X_j + \varepsilon_j)]\}$$

(since $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ and $Y_j = \beta_0 + \beta_1 X_j + \varepsilon_j$)

$$= E\{[\beta_0 + \beta_1 X_i + \varepsilon_i - \beta_0 - \beta_1 X_i][\beta_0 + \beta_1 X_j + \varepsilon_j - \beta_0 - \beta_1 X_j]\}$$

$= E[\varepsilon_i \varepsilon_j] = 0, i \neq j$, since $E(\varepsilon_i) \& E(\varepsilon_j) = 0$ by assumption 8

Therefore, $cov(Y_i, Y_j) = 0$

8.3 Method of Estimations

After specifying the model and stated their assumption, the next step is estimation of the numerical values of the parameters of economic relationships. Depending on the nature of the data and functional form, there are different methods of estimating the parameters. The most widely used methods of estimation in the parameters of the simple linear regression model is the ordinarily Least Squares (OLS) method.

8.3.1 The Ordinary Least Square (OLS) Methods

In regression analysis, the researcher is interested in analyzing the behavior of a dependent variable "Y_i" given the information contained in a set of explanatory variables "X_i". Ordinary Least Squares is a standard approach to specify a linear regression model and estimate its unknown parameters by minimizing the sum of squared errors. This leads to an approximation of the mean function of the conditional distribution of the dependent variable. Thus, the principle of least squares is based on the residuals. For any line, the residuals are the deviations of the dependent variables Y_i away from the line. If the line is the true regression line of the model, then the residuals are exactly random errors.
Note that, the better the line fits the data, the smaller the residuals will be. Thus, we can use the 'sizes' of the residuals as a measure of how well a proposed line fits the data.
Consider the model, $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ is called the true relationship between Y and X because Y and X represent their respective population value, and β_0 and β_1 are called the true parameters since they can be estimated from the population value of Y and X. However, it is difficult to obtain the population value of Y and X because of technical or economic reasons. So we are forced to take the sample value of Y and X.
The parameters estimated from the sample value of Y and X are called the estimators of the true parameters β_0 and $\beta_1$and are symbolized as $\hat{\beta}_0$ and $\hat{\beta}_1$.
Recall from equation (2), the population regression function is given: $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$. However, the population regression function is not directly observable

thus; we estimate it from the sample regression function as stated in equation (3):$Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\varepsilon}_i = \hat{Y}_i + \hat{\varepsilon}_i$

where $\hat{Y}_i$ is read as "*Y*-hat' 'which refers to the estimated (conditional mean) value of Y_i.
Now we can derive the sample residual term $(\hat{\varepsilon}_i)$ which is the difference between the actual and the fitted/estimated/predicted values of Y as follow:

$$\hat{\varepsilon}_i = Y_i - \hat{Y}_i = Y_i - \hat{\beta}_0 + \hat{\beta}_1 X_i \qquad (2.3)$$

Now, given data on both X and Y, our objective is to find the sample regression function that best "fits" the population regression function or the values of estimators which are as close as possible to population parameters. Therefore, the goal is to find the best-fit line that minimizes the sum of the error terms.

- **Criterion I: Minimize the sum of residuals**

That is, minimize $\sum_{i=1}^{n} \hat{\varepsilon}_i = 0$
However, this criterion is not good as it gives equal weight to all kinds of residuals (large, medium and small) and residuals of different signs can be compensated (cancel out each other).

- **Criterion II: Minimize Absolute Values of Residuals Criterion**

In order to avoid the compensation of positive residuals with negative ones, the absolute values from the residuals are taken. That is minimize $\sum_{i=1}^{n} |\hat{\varepsilon}_i| = 0$.
Unfortunately, although the estimators thus obtained have some interesting properties, their calculation is complicated and requires resolving the problem of linear programming or applying a procedure of iterative calculation.

- **Criterion III: Sum of Squared Residuals Criterion**

According to this criterion, find the SRF, which minimize the sum of the squared residuals, $\sum_{i=1}^{n} \hat{\varepsilon}_i^2$. Alternatively, we choose $\hat{\beta}_0$ and $\hat{\beta}_1$ such that the sum of squared residuals is minimized. This criterion is very important because, it gives more weight to larger residuals and less weight to smaller residuals as define above on OLS properties. For example, consider the following three residuals, 2 4, and 8. Each residual is twice as large as the preceding residual. That is, 4 is twice as large as 2 and 8 is twice as large as 4. When the residual is squared, the squared values are 4, 16, and 64 in the OLS objective function. This trait places a larger weight on the objective function when the estimated Y-value is far away from the actual value than when the estimated Y-value is close to the actual value. Moreover, it insures the residuals that are equal in magnitude are given equal weight. Consider the two residuals –6 and 6. In both of these observations, the estimated y-value is equal distance from the observed y-value, 6 units. It just happens you overestimated y in the first case and underestimated y in the second case. By squaring the residuals, both values are 16 in the objective function.

Deriving the OLS Estimators

From the sample regression function of $Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{\varepsilon}_i = \hat{Y}_i + \hat{\varepsilon}_i$, the fitted value is $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$. Then the residual is given by $\hat{\varepsilon}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$.
The sum of the squared residuals as defined by S is given by:

$$S = \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2 = \sum_{i=1}^{n} \hat{\varepsilon}_i^2 \tag{2.4}$$

Now we determine (estimate) $\hat{\beta}_0$ and $\hat{\beta}_1$ in such a way that $\sum_{i=1}^{n} \hat{\varepsilon}_i^2$ is minimum. Hence, we minimize $\sum_{i=1}^{n} \hat{\varepsilon}_i^2$ subject to $\hat{\beta}_0$ and $\hat{\beta}_1$.

Note that an estimator, also known as a (sample) statistic, is simply a rule, formula, or method that tells how to estimate the population parameter from the information provided by the sample at hand. A particular numerical value obtained by the estimator in an application is known as an estimate.

To minimize sum of the squared residuals ($\sum_{i=1}^{n} \hat{\varepsilon}_i^2$), we apply first order (necessary) condition with respect to $\hat{\beta}_0$ and $\hat{\beta}_1$.

The necessary condition for intercept term ($\hat{\beta}_0$):

$$\frac{\partial \sum_{i=1}^{n} \hat{\varepsilon}_i^2}{\partial \hat{\beta}_0} = 0$$

$$\frac{\partial \sum_{i=1}^{n} \hat{\varepsilon}_i^2}{\partial \hat{\beta}_0} = \sum_{i=1}^{n} 2\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right)(-1) = 0$$

Apply sum to the whole values and divided both side by 2 results:

$$-\sum_{i=1}^{n} Y_i + n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} X_i = 0$$

Then the normal equation is given by:

$$\sum_{i=1}^{n} Y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} X_i$$

Divide both sides by n results:

$$\bar{Y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{X}$$

Therefore, the intercept coefficient

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} \tag{2.5}$$

The necessary condition for slope coefficient $\hat{\beta}_1$ (continued):

$$\frac{\partial \sum_{i=1}^{n} \hat{\varepsilon}_i^2}{\partial \hat{\beta}_1} = 0$$

$$\frac{\partial \sum_{i=1}^{n} \hat{\varepsilon}_i^2}{\partial \hat{\beta}_1} = \sum_{i=1}^{n} 2\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right)(-X_i) = 0$$

$$= -\sum_{i=1}^{n} Y_i X_i + \hat{\beta}_0 \sum_{i=1}^{n} X_i + \hat{\beta}_1 \sum_{i=1}^{n} X_i^2 = 0$$

$$\sum_{i=1}^{n} Y_i X_i = \hat{\beta}_0 \sum_{i=1}^{n} X_i + \hat{\beta}_1 \sum_{i=1}^{n} X_i^2 \tag{2.6}$$

Then, substitute equations (2.5) into equation (2.6) and obtained:

$$\sum_{i=1}^{n} Y_i X_i = (\bar{Y} - \hat{\beta}_1 \bar{X}) \sum_{i=1}^{n} X_i + \hat{\beta}_1 \sum_{i=1}^{n} X_i^2$$

$$\sum_{i=1}^{n} Y_i X_i = \bar{Y} \sum_{i=1}^{n} X_i - \hat{\beta}_1 \bar{X} \sum_{i=1}^{n} X_i + \hat{\beta}_1 \sum_{i=1}^{n} X_i^2$$

$$\sum_{i=1}^{n} Y_i X_i = \bar{Y} \sum_{i=1}^{n} X_i + \hat{\beta}_1 \sum_{i=1}^{n} X_i^2 - \hat{\beta}_1 \bar{X} \sum_{i=1}^{n} X_i$$

$$\sum_{i=1}^{n} Y_i X_i - \bar{Y} \sum_{i=1}^{n} X_i = \hat{\beta}_1 \left[\sum_{i=1}^{n} X_i^2 - \bar{X} \sum_{i=1}^{n} X_i \right]$$

Divided both side by $\left[\sum_{i=1}^{n} X_i^2 - \bar{X} \sum_{i=1}^{n} X_i\right]$ results

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} Y_i X_i - \bar{Y} \sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} X_i^2 - \bar{X} \sum_{i=1}^{n} X_i} = \frac{\sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y}}{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2}$$

Now let us rewrite the $\hat{\beta}_1$ function in other way or deviation form as follows.

$\sum_{i=1}^{n} Y_i X_i - n\bar{Y}\bar{X} = \sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y} + n\bar{X}\bar{Y} - n\bar{Y}\bar{X}$

$$= \sum_{i=1}^{n} Y_i X_i - \bar{X} \sum_{i=1}^{n} Y_i - \bar{Y} \sum_{i=1}^{n} X_i + n\bar{X}\bar{Y}$$

$$\sum_{i=1}^{n} Y_i X_i - n\bar{Y}\bar{X} = \sum_{i=1}^{n} (Y_i - \bar{Y})(X_i - \bar{X})$$

and

$$\sum_{i=1}^{n} X_i^2 - n\bar{X}^2 = \sum_{i=1}^{n} X_i^2 - n\bar{X}^2 + 2n\bar{X}.\bar{X}$$

$$= \sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{n} \bar{X}^2 - 2\bar{X} \sum_{i=1}^{n} X_i$$

$$\sum_{i=1}^{n} X_i^2 - n\bar{X}^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2$$

Then we can simplify the formulas for beta hat as:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (Y_i - \bar{Y})(X_i - \bar{X})}{\sum_{i=1}^{n} (X_i - \bar{X})^2} = \frac{\sum_{i=1}^{n} y_i x_i}{\sum_{i=1}^{n} x_i^2} \tag{2.7}$$

where $x_i = X_i - \bar{X}$, that is x is in deviation form of the observed data, X and $\bar{X} = \sum_{i=1}^{n} X_i / n$

$y_i = Y_i - \bar{Y}$, that is y is in deviation form of the observed data, Y and $\bar{Y} = \sum_{i=1}^{n} Y_i / n$.

Alternatively, $\hat{\beta}_1$ can be expressed in probability point of view given both Y and X as a random variable by divided both the numerator and denominator by $n - 1$.

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(Y_i - \bar{Y})(X_i - \bar{X})/n-1}{\sum_{i=1}^{n}(X_i - \bar{X})^2/n-1} = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)}$$

Therefore, the sign of $\hat{\beta}_1$ is the same as the sign of the covariance.

A $\hat{\beta}_1$ coefficient measures the partial effect of the regressor X on Y holding the other regressors fixed. However, if the variable is in logarithm form we interpreted as percentage or elasticity's.

Now let us see whether the intercept and slope coefficients satisfy the second order (sufficient) condition. i.e. proof whether the OLS estimates have a global minimum which is expected to have a positive value.

The sufficient condition for intercept coefficient ($\hat{\beta}_0$)

$$\frac{\partial^2(\sum_{i=1}^{n}\hat{\varepsilon}_i^2)}{\partial^2\hat{\beta}_0^2} = \frac{\partial^2\left(\sum_{i=1}^{n}\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right)^2\right)}{\partial^2\hat{\beta}_0^2}$$

$$= \frac{\partial\left(\sum_{i=1}^{n} 2\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right)(-1)\right)}{\partial\hat{\beta}_0}$$

$$\frac{\partial^2(\sum_{i=1}^{n}\hat{\varepsilon}_i^2)}{\partial^2\hat{\beta}_0^2} = 2\sum_{i=1}^{n} 1 = 2n \qquad (2.8)$$

The sufficient condition for slope coefficient ($\hat{\beta}_1$)

$$\frac{\partial^2(\sum_{i=1}^{n}\hat{\varepsilon}_i^2)}{\partial^2\hat{\beta}_1^2} = \frac{\partial^2\left(\sum_{i=1}^{n}\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right)^2\right)}{\partial^2\hat{\beta}_1^2} =$$

$$= \frac{\partial\left(\sum_{i=1}^{n} 2\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right)(-X_i)\right)}{\partial\hat{\beta}_1}$$

$$\frac{\partial^2(\sum_{i=1}^{n}\hat{\varepsilon}_i^2)}{\partial^2\hat{\beta}_1^2} = 2\sum_{i=1}^{n} X_i^2 \qquad (2.9)$$

The sufficient conditions for covariance of intercept and slope coefficient, $\mathrm{cov}(\hat{\beta}_0, \hat{\beta}_1)$ is given as

$$\frac{\partial^2(\sum_{i=1}^{n}\hat{\varepsilon}_i^2)}{\partial\hat{\beta}_0\hat{\beta}_1} = \frac{\partial^2\left(\sum_{i=1}^{n}\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right)^2\right)}{\partial\hat{\beta}_0\hat{\beta}_1} =$$

$$= \frac{\partial\left(\sum_{i=1}^{n} 2\left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i\right)(-1)\right)}{\partial\hat{\beta}_1}$$

$$\frac{\partial^2(\sum_{i=1}^{n}\hat{\varepsilon}_i^2)}{\partial\hat{\beta}_0\hat{\beta}_1} = 2\sum_{i=1}^{n} X_i = 2n\bar{X} \qquad (2.10)$$

The Hessian matrix (H^*) which is the matrix of second order partial derivatives in this case is given as:

$$H^* = \begin{pmatrix} \frac{\partial^2(\sum_{i=1}^{n}\hat{\varepsilon}_i^2)}{\partial^2\hat{\beta}_0^2} & \frac{\partial^2(\sum_{i=1}^{n}\hat{\varepsilon}_i^2)}{\partial\hat{\beta}_0\hat{\beta}_1} \\ \frac{\partial^2(\sum_{i=1}^{n}\hat{\varepsilon}_i^2)}{\partial\hat{\beta}_0\hat{\beta}_1} & \frac{\partial^2(\sum_{i=1}^{n}\hat{\varepsilon}_i^2)}{\partial^2\hat{\beta}_1^2} \end{pmatrix} = 2\begin{pmatrix} n & n\bar{X} \\ n\bar{X} & \sum_{i=1}^{n} X_i^2 \end{pmatrix}$$

The matrix H^* is a positive definite if its determinate and the element in the first row and column of H^* are positive. The determinant of H^* is given by:

$$|H^*| = 2\left(n\sum_{i=1}^{n} X_i^2 - n^2\bar{X}^2\right) = 2n\sum_{i=1}^{n}(X_i - \bar{X})^2 \geq 0 \qquad (2.11)$$

Since $\sum_{i=1}^{n}(X_i - \bar{X})^2$ is positive definite because it is a quadratic (square) function The case when $\sum_{i=1}^{n}(X_i - \bar{X})^2 = 0$ is not interesting because all the observations in this case are identical, i.e. $X_i = c$ (some constant). In such case, there is no relationship between X and Y in the context of regression analysis. That is why we say sample variability in X value is necessary. Since $\sum_{i=1}^{n}(X_i - \bar{X})^2 > 0$, therefore, $|H^*| > 0$. So H^* is positive definite for any $(\hat{\beta}_0, \hat{\beta}_1)$ and hence the estimates $(\hat{\beta}_0 \text{ and } \hat{\beta}_1)$ has a value that gives a global minimum error.

8.3.2 The Statistical Properties of Ordinary Least Square Estimators

In this section we discuss the statistical properties of the least squares estimates for the simple linear regression. The optimum properties that the *Ordinary Least Square (OLS)* estimates possess may be summarized by well-known theorem known as the ***Gauss-Markov Theorem.*** According to this theorem, under the basic assumptions of the classical linear regression model, the least squares estimators are linear, unbiased and have minimum variance (i.e. are best of all linear unbiased estimators). Sometimes the theorem referred as the BLUE theorem i.e. Best, Linear, and Unbiased Estimator. The estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ determined by OLS are known as Best Linear Unbiased Estimators (BLUE).

9. Linearity

The first property of OLS estimates is the linearity property. That is the estimates $\hat{\beta}_0$ & $\hat{\beta}_1$ are linear in Y

Theorem 1. The least squares estimator $\hat{\beta}_1$ is a linear estimate of β_1, i.e. the slope estimator $\hat{\beta}_1$ is a linear function of the dependent variable Y.

Proof (1):

Recall from equation (2.7):

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i}{\sum x_i^2}$$

Now our objective is to show whether the slope estimator $\hat{\beta}_1$ is linear with the observed dependent variable Y_i. So express in terms of observed Y_i, we should express the deviated form of above equation in observed form as follows.

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i(Y_i - \bar{Y})}{\sum x_i^2} = \frac{\sum_{i=1}^{n} x_i Y_i - \bar{Y}\sum_{i=1}^{n} x_i}{\sum x_i^2}$$

However, $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n}(X_i - \bar{X}) = \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \bar{X} = \sum_{i=1}^{n} X_i - n\bar{X}$

$$= \sum_{i=1}^{n} X_i - n\sum_{i=1}^{n} X_i/n = \sum_{i=1}^{n} X_i - \sum_{i=1}^{n} X_i = 0$$

$$\therefore \qquad \sum_{i=1}^{n} x_i = 0 \qquad (2.12)$$

Therefore,

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i Y_i}{\sum x_i^2}$$

Now, let us define the observation weights $k_i = \frac{x_i}{\sum x_i^2}, \rightarrow (i = 1, 2, \ldots, n)$ which is a function of fixed or non-stochastic explanatory variable X_i.

$$\hat{\beta}_1 = \sum_{i=1}^{n} k_i Y_i = k_1 Y_1 + k_2 Y_2 + k_3 Y_3 + \cdots + k_n Y_n \qquad (2.13)$$

Therefore, $\hat{\beta}_1$ is a linear weighted sum of Y_i

In order to establish the remaining properties of $\hat{\beta}_1$, it is necessary to know the arithmetic properties of the weights k_i.

Lemma 1: $\sum_i k_i = 0$, i.e. the weights k_i sum to zero, how?

Proof:

$$\sum_{i=1}^{n} k_i = \sum \frac{x_i}{\sum_{i=1}^{n} x_i^2} = \frac{1}{\sum_{i=1}^{n} x_i^2}\sum_{i=1}^{n} x_i$$

By equation (2.12), $\sum_{i=1}^{n} x_i = 0$, thus the sum of the weights k_i is given by

$$\sum_{i=1}^{n} k_i = \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} x_i^2} = \frac{0}{\sum_{i=1}^{n} x_i^2} = 0$$

Lemma 2: $\sum_i k_i^2 = \frac{1}{\sum_i x_i^2}$, how?

Proof:

$$\sum_i k_i^2 = \sum_i \left(\frac{x_i}{\sum_i x_i^2}\right)^2 = \sum_i \frac{x_i^2}{(\sum_i x_i^2)^2} = \frac{\sum_i x_i^2}{(\sum_i x_i^2)^2} = \frac{1}{\sum_i x_i^2}$$

$$\therefore \quad \sum_i k_i^2 = \frac{1}{\sum_i x_i^2}$$

Lemma 3: $\sum_i k_i x_i = \sum_i k_i X_i$, how?

Proof:

$$\sum_i k_i x_i = \sum_i k_i (X_i - \bar{X}) = \sum_i k_i X_i - \bar{X} \sum_i k_i$$

By Lemma 1, $\sum_i k_i = 0$

$\therefore \quad \sum_i k_i x_i = \sum_i k_i X_i$ since $\sum_i k_i = 0$ by assumptions (**Lemma 1**)

Lemma 4: $\sum_{i=1}^{n} k_i X_i = 1$, how?

Proof:

$$\sum_{i=1}^{n} k_i X_i = \sum_{i=1}^{n} \left(\frac{x_i}{\sum_{i=1}^{n} x_i^2} \right) X_i = \frac{\sum_{i=1}^{n} x_i X_i}{\sum_{i=1}^{n} x_i^2} = \frac{\sum_{i=1}^{n} (X_i - \bar{X}) X_i}{\sum_{i=1}^{n} x_i^2}$$

$$= \frac{\sum_{i=1}^{n} X_i^{\,2} - \bar{X} \sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} X_i^{\,2} - n\,\bar{X}^2} = \frac{\sum_{i=1}^{n} X_i^{\,2} - n\bar{X}^2}{\sum_{i=1}^{n} X_i^{\,2} - n\,\bar{X}^2} = 1$$

$$\therefore \quad \sum_{i=1}^{n} k_i X_i = 1$$

Theorem 2. The least squares estimator $\hat{\beta}_0$ is a linear estimate of β_0, i.e. the intercept estimator $\hat{\beta}_0$ is a linear function of the dependent variable Y.

Recall equation (2.5)

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

Now we should again express $\hat{\beta}_0$ as a function of observed value of Y_i. To do so, we should substitute equation (2.13) in equation (2.5) value of $\hat{\beta}_1$.

Proof (2):

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X} = \bar{Y} - \left(\sum K_i Y_i \right) \bar{X}$$

$$= \sum \frac{1}{n} Y_i - \left(\sum K_i Y_i \right) \bar{X}$$

$$= \sum \left(\frac{1}{n} - \bar{X} K_i \right) Y_i$$

$$\hat{\beta}_0 = \sum \left(\frac{1}{n} - \bar{X} K_i \right) Y_i = \sum Z_i \, Y_i \qquad (2.14)$$

where $Z_i = \left(\frac{1}{n} - \bar{X} K_i \right)$ which is function of fixed values of n &X.

Therefore, $\hat{\beta}_0$ is linear weighted sum of Y_i

9.1 Unbiasedness

The basic idea of unbiasedness lie down the sample is drawn randomly and independently from the population. In this case, the OLS estimators $\hat{\beta}_0$ & $\hat{\beta}_1$ become an unbiased estimates of the true parameters β_0 & β_1. However, the property of unbiasedness does not mean that $\hat{\beta}_1 = \beta_1$; it says only that, if we could undertake repeated sampling an infinite number of times, we would get the correct estimate "on the average." To show whether those sample statistic (estimates) are unbiased estimators of the population parameters.

Now let us see the proof as follows:

Theorem 3: The OLS slope coefficient is an unbiased estimate of the population slope parameter (i.e., $E(\hat{\beta}_1) = \beta_1$).

Proof (1):

Prove that $\hat{\beta}_1$ is unbiased estimate of β_1, i.e. $E(\hat{\beta}_1) = \beta_1$

We know from equation (2.13), $\hat{\beta}_1 = \sum_{i=1}^{n} k_i Y_i$

Now substitute the population regression function in Y_i given by equation (2.1) since our objective is to show whether the sample coefficient is an unbiased estimator of the population parameter.

$$\hat{\beta}_1 = \sum_{i=1}^{n} k_i Y_i = \sum_{i=1}^{n} k_i\,(\beta_0 + \beta_1 X_i + \varepsilon_i)$$

$$= \beta_0 \sum_{i=1}^{n} k_i + \beta_1 \sum_{i=1}^{n} k_i\, X_i + \sum_{i=1}^{n} k_i\, \varepsilon_i$$

However, from Lemma 1 and lemma 4, $\sum_{i=1}^{n} k_i = 0$ and $\sum_i k_i X_i = 1$, respectively.

$$\hat{\beta}_1 = \beta_1 + \sum_{i=1}^{n} k_i \varepsilon_i$$

$$\hat{\beta}_1 - \beta_1 = \sum_{i=1}^{n} k_i \varepsilon_i \qquad (2.15)$$

Taking expectation:

$E(\hat{\beta}_1) = E(\beta_1) + \sum_{i=1}^{n} k_i E(\varepsilon_i)$ since k_i are fixed

$E(\hat{\beta}_1) = \beta_1$, since $E(\varepsilon_i) = 0$

$$E(\hat{\beta}_1) = \beta_1 \qquad (2.16)$$

Therefore, $\hat{\beta}_1$ is unbiased estimator of β_1.

Theorem 4: The least squares estimator $\hat{\beta}_0$ is an unbiased estimate of β_0 (i.e., $E(\hat{\beta}_0) = \beta_0$).

Proof (2): P*rove that* $\hat{\beta}_0$ *is unbiased i.e:* $E(\hat{\beta}_0) = \beta_0$

From the proof of linearity property of $\hat{\beta}_0$ in equation (2.14), we have:

$$\hat{\beta}_0 = \sum \left(\frac{1}{n} - \bar{X} k_i\right) Y_i$$

From equation (2.1) $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$, then

$$\hat{\beta}_0 = \sum \left[\left(\frac{1}{n} - \bar{X} k_i\right)(\beta_0 + \beta_1 X_i + \varepsilon_i)\right]$$

$$= \beta_0 + \beta_1 \frac{1}{n} \sum X_i + \frac{1}{n} \sum \varepsilon_i - \beta_0 \bar{X} \sum k_i - \beta_1 \bar{X} \sum k_i X_i - \bar{X} \sum k_i\, \varepsilon_i$$

$$= \beta_0 + \beta_1 \bar{X} + \frac{1}{n} \sum \varepsilon_i - \beta_1 \bar{X} - \bar{X} \sum k_i\, \varepsilon_i$$

By Lemma 1, $\sum \mathrm{k_i} = 0$ and Lemma 4, $\sum \mathrm{k_i X_i} = 1$, then

$$\hat{\beta}_0 = \beta_0 + \frac{1}{n}\sum \varepsilon_i - \bar{X}\sum \mathrm{k_i}\,\varepsilon_i$$

$$\hat{\beta}_0 - \beta_0 = \frac{1}{n}\sum \varepsilon_i - \bar{X}\sum \mathrm{k_i}\,\varepsilon_i = \sum\left(\frac{1}{n} - \bar{X}\mathrm{k_i}\right)\varepsilon_i \tag{2.17}$$

To find the unbiased estimator, taking expectation:

$$E(\hat{\beta}_0) = \beta_0 + \frac{1}{n}\sum E(\varepsilon_i) - \bar{X}\sum \mathrm{k_i}\,E(\varepsilon_i)$$

By assumption 5, $E(\varepsilon_i) = 0$, then we have

$$E(\hat{\beta}_0) = \beta_0 \tag{2.18}$$

Therefore, $\hat{\beta}_0$ is an unbiased estimator of β_0

9.2 Minimum variance

In order to test whether the linear and unbiased estimators of $\hat{\beta}_0$ and $\hat{\beta}_1$ possess the smallest sampling variances (efficiency), we shall first obtain the variance of $\hat{\beta}_0$ and $\hat{\beta}_1$ and then establish that each estimates has the minimum variance in comparison with the variances of other linear and unbiased estimators obtained by any other econometric methods of estimations than OLS.

a. Variance of $\hat{\beta}_1$

$$\mathrm{var}(\hat{\beta}_1) = \mathrm{E}\big(\hat{\beta}_1 - \mathrm{E}(\hat{\beta}_1)\big)^2 = E(\hat{\beta}_1 - \beta_1)^2 \tag{2.19}$$

Substitute equation (2.15) in (2.19) and we get

$$\mathrm{var}(\hat{\beta}_1) = \mathrm{E}\Big(\sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{k_i}\varepsilon_i\Big)^2$$

$$= \mathrm{E}(\mathrm{k_1\varepsilon_1 + k_2\varepsilon_2 + \cdots + k_{n-1}\varepsilon_{n-1} + k_n\varepsilon_n})^2$$

$$= E[k_1^2\varepsilon_1^2 + k_2^2\varepsilon_2^2 + \cdots + k_n^2\varepsilon_n^2 + 2\mathrm{k_1k_2\varepsilon_1\varepsilon_2} + \cdots + 2\mathrm{k_{n-1}k_n\varepsilon_{n-1}\varepsilon_n}]$$

$$= E[k_1^2\varepsilon_1^2 + k_2^2\varepsilon_2^2 + \cdots + k_n^2\varepsilon_n^2] + \mathrm{E}[2\mathrm{k_1k_2\varepsilon_1\varepsilon_2} + \cdots + 2\mathrm{k_{n-1}}k_n\mathrm{\varepsilon_{n-1}\varepsilon_n}]$$

$$= E\left[\sum k_i^2\varepsilon_i^2\right] + 2\mathrm{E}\left[\sum_{i=1}^{n-1}\sum_{j=2}^{n}\mathrm{k_i}k_j\mathrm{\varepsilon_i\varepsilon_j}\right] \quad , \rightarrow i \neq j$$

$$= \sum k_i^2\, E(\varepsilon_i^2) + 2\sum_{i=1}^{n-1}\sum_{j=2}^{n}\mathrm{k_i}k_j\mathrm{E}(\varepsilon_i\varepsilon_j) \quad , \qquad since\ \mathrm{k_i}\ and\ \mathrm{k}_j\ are\ constant$$

By classical linear regression assumption (6), $E(\varepsilon_i^2) = \sigma^2$ and assumption (7), $\mathrm{E}(\varepsilon_i\varepsilon_j) = 0)$, for $i \neq j$.

$$\mathrm{var}(\hat{\beta}_1) = \sum k_i^2\,\sigma^2$$

$$\therefore\ \mathrm{var}(\hat{\beta}_1) = \sigma^2\sum k_i^2 = \frac{\sigma^2}{\sum x_i^2} \tag{2.20}$$

b. Variance of $\hat{\beta}_0$

$$\mathrm{var}(\hat{\beta}_0) = E\left(\hat{\beta}_0 - E(\hat{\beta}_0)\right)^2 = E(\hat{\beta}_0 - \beta_0)^2 \tag{2.21}$$

Substituting equation (2.16) in (2.21) and we get,

$$\mathrm{var}(\hat{\beta}_0) = E\left(\sum\left(\frac{1}{n} - \bar{X}K_i\right)\varepsilon_i\right)^2$$

$$= E\left[\left(\frac{1}{n} - \bar{X}K_1\right)\varepsilon_1 + \left(\frac{1}{n} - \bar{X}K_2\right)\varepsilon_2 + \cdots + \left(\frac{1}{n} - \bar{X}K_{n-1}\right)\varepsilon_{n-1} + \left(\frac{1}{n} - \bar{X}K_n\right)\varepsilon_n\right]^2$$

$$= E\left[\left(\frac{1}{n} - \bar{X}K_1\right)^2 \varepsilon_1^2 + \left(\frac{1}{n} - \bar{X}K_2\right)^2 \varepsilon_2^2 + \cdots + \left(\frac{1}{n} - \bar{X}K_n\right)^2 \varepsilon_n^2\right.$$

$$+ 2\left(\frac{1}{n} - \bar{X}K_1\right)\left(\frac{1}{n} - \bar{X}K_2\right)\varepsilon_1\varepsilon_2 + \cdots$$

$$\left.+ 2\left(\frac{1}{n} - \bar{X}K_{n-1}\right)\left(\frac{1}{n} - \bar{X}K_n\right)\varepsilon_{n-1}\varepsilon_n\right]$$

$$= E\left[\left(\frac{1}{n} - \bar{X}K_1\right)^2 \varepsilon_1^2 + \left(\frac{1}{n} - \bar{X}K_2\right)^2 \varepsilon_2^2 + \cdots + \left(\frac{1}{n} - \bar{X}K_n\right)^2 \varepsilon_n^2\right]$$

$$+ E\left[2\left(\frac{1}{n} - \bar{X}K_1\right)\left(\frac{1}{n} - \bar{X}K_2\right)\varepsilon_1\varepsilon_2 + \cdots\right.$$

$$\left.+ 2\left(\frac{1}{n} - \bar{X}K_{n-1}\right)\left(\frac{1}{n} - \bar{X}K_n\right)\varepsilon_{n-1}\varepsilon_n\right]$$

$$= E\sum_{i=1}^{n}\left(\frac{1}{n} - \bar{X}K_i\right)^2\varepsilon_i{}^2 + 2\mathrm{E}\left[\sum_{i=1}^{n-1}\sum_{j=2}^{n}\left(\frac{1}{n} - \bar{X}k_i\right)\left(\frac{1}{n} - \bar{X}k_j\right)\varepsilon_{\mathrm{i}}\varepsilon_{\mathrm{j}}\right], \quad i \neq j$$

Since n, $\bar{X}$ and k_i are constant

$$\mathrm{var}(\hat{\beta}_0) = \sum\left(\frac{1}{n} - \bar{X}k_i\right)^2 E(\varepsilon_i{}^2)$$

$$+ 2\left[\sum_{i=1}^{n-1}\sum_{j=2}^{n}\left(\frac{1}{n} - \bar{X}k_i\right)\left(\frac{1}{n} - \bar{X}k_j\right)E(\varepsilon_{\mathrm{i}}\varepsilon_{\mathrm{j}})\right], \text{for } i \neq j$$

Again by assumption (6) & (7), $E(\varepsilon_i{}^2) = \sigma^2$ and $E(\varepsilon_{\mathrm{i}}\varepsilon_{\mathrm{j}}) = 0, for\ i \neq j$

$$\mathrm{var}(\hat{\beta}_0) = \sigma^2 \sum\left(\tfrac{1}{n} - \bar{X}\mathrm{k_i}\right)^2$$

$$= \sigma^2 \sum\left(\frac{1}{n^2} - {}^2/_n\,\bar{X}\mathrm{k_i} + \bar{X}^2 k_i^2\right)$$

$$= \sigma^2\left(\frac{1}{n} - {}^2/_n\,\bar{X}\sum \mathrm{k_i} + \bar{X}^2 \sum k_i^2\right)$$

By Lemma (1), $\sum \mathrm{k_i} = 0$

$$\mathrm{var}(\hat{\beta}_0) = \sigma^2\left(\frac{1}{n} + \bar{X}^2 \sum k_i^2\right)$$

$$= \sigma^2\left(\frac{1}{n} + \bar{X}^2 \sum k_i^2\right)$$

By Lemma (2), $\sum k_i^2 = \frac{1}{\sum x_i^2}$

$$\mathrm{var}(\hat{\beta}_0) = \sigma^2\left(\frac{1}{n} + \frac{\bar{X}^2}{\sum x_i^2}\right)$$

where $\frac{1}{n} + \frac{\bar{X}^2}{\sum x_i^2} = \frac{\sum x_i^2 + n\bar{X}^2}{n\sum x_i^2} = \frac{\sum (X_i - \bar{X})^2 + n\bar{X}^2}{n\sum x_i^2} = \frac{\sum\left(X_i^2 - 2X_i\bar{X} + \bar{X}^2\right) + n\bar{X}^2}{n\sum x_i^2}$

$$= \frac{\sum X_i^2 - 2\bar{X}\sum X_i + \sum \bar{X}^2 + n\bar{X}^2}{n\sum x_i^2} = \frac{\sum X_i^2 - 2\bar{X}\sum X_i + n\bar{X}^2 + n\bar{X}^2}{n\sum x_i^2}$$

$$\frac{1}{n} + \frac{\bar{X}^2}{\sum x_i^2} = \frac{\sum X_i^2 - 2n\bar{X}^2 + 2n\bar{X}^2}{n\sum x_i^2} = \frac{\sum X_i^2}{n\sum x_i^2} \tag{2.22}$$

$$\therefore \operatorname{var}(\hat{\beta}_0) = \sigma^2\left(\frac{1}{n} + \frac{\bar{X}^2}{\sum x_i^2}\right) = \sigma^2\left(\frac{\sum X_i^2}{n\sum x_i^2}\right) \tag{2.23}$$

We have computed the variances of OLS estimates. Now, it is time to check whether the variances of OLS estimates ($\hat{\beta}_0$ & $\hat{\beta}_1$) do possess minimum variance property compared to the variances of other estimators of the true β_0 and β_1.

To establish that $\hat{\beta}_0$ and $\hat{\beta}_1$ possess minimum variance property, we compare their variances with that of the variances of some other alternative linear and unbiased estimators of β_0 and β_1, say $\hat{\beta}_0^*$ *and* $\hat{\beta}_1^*$. Now, we want to prove that any other linear and unbiased estimator of the true population parameter obtained from any other econometric method has larger variance than that of OLS estimators. Let us first show the minimum variance of $\hat{\beta}_1$ and then that of $\hat{\beta}_0$.

9.3 Minimum variance of $\hat{\beta}_1$

Suppose: $\hat{\beta}_1^*$ is an alternative linear and unbiased estimator of β_1.

Linearity property of $\hat{\beta}_1^*$ is given by:

$$\text{Let } \hat{\beta}_1^* = \sum w_i Y_i \tag{2.24}$$

where w_i is the weighted value in other linear and unbiased estimator, which is differ from the collection of constant term, k_i (i.e. $w_i \neq k_i$), rather $w_i = k_i + c_i$. c_i is another constant term whose value is derived from an assumption.

Unbiasedness of $\hat{\beta}_1^*$ is given by:

$\hat{\beta}_1^* = \sum w_i(\beta_0 + \beta_1 X_i + \varepsilon_i)$, since $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$

$= \beta_0 \sum w_i + \beta_1 \sum w_i X_i + \sum w_i \varepsilon_i$

Since $\hat{\beta}_1^*$ is assumed to be an unbiased estimator. Then $\hat{\beta}_1^*$ is to be an unbiased estimator of β_1, there must be true that $\sum w_i = 0$ and $\sum w_i X_i = 1$ in the above equation.

However, $w_i = k_i + c_i$, thus $\sum w_i = \sum(k_i + c_i) = \sum k_i + \sum c_i$

From Lemma (1), $\sum k_i = 0$,

Thus, $\sum w_i$ to be zero, then $\sum c_i$ should be zero: i.e.

$$\sum c_i = 0 \tag{2.25}$$

Therefore, based on lemma (1) equation (2.25):

$$\sum w_i = 0 \tag{2.26}$$

Moreover, $\sum w_i X_i = \sum(k_i + c_i) X_i = \sum k_i X_i + \sum c_i X_i$

Again from lemma (4), $\sum k_i X_i = 1$.

Thus, $\sum w_i X_i$ to be one, then $\sum c_i X_i$ should be zero, How?

Let derived it from $\sum c_i x_i$, , where $x_i = X_i - \bar{X}$.

Thus, $\sum c_i x_i = \sum c_i\,(X_i - \bar{X})$

$$= \sum c_i\,X_i - \bar{X}\sum c_i$$

From equation (2.25) $\sum c_i = 0$, then

$$\sum c_i x_i - \sum c_i\,X_i = -\bar{X}\sum c_i = 0$$

Hence, $\sum c_i x_i$ and $\sum c_i X_i = 0$.

$$\sum c_i X_i = 0 \tag{2.27}$$

Therefore, from lemma (4) and equation (2.27):

$$\sum w_i X_i = 1 \tag{2.28}$$

From equation (2.27 & 2.28), $\sum c_i\,X_i = 0$ and $\sum c_i = 0$, respectively, thus:

$$\sum c_i x_i = 0 \tag{2.29}$$

Thus, we have to put $\hat{\beta}_1^*$ as in the form

$$\hat{\beta}_1^* = \beta_1 + \sum w_i \varepsilon_i$$

$$\hat{\beta}_1^* - \beta_1 = \sum w_i \varepsilon_i \tag{2.30}$$

Taking expectation

$$E\left(\hat{\beta}_1^*\right) = \beta_1 + \sum w_i\,E(\varepsilon_i)$$

Since w_i is fixed and the only random variable is ε_i, we apply expectation only for random variables, ε_i and its expected value is zero, ($E(\varepsilon_i) = 0$).

Finally, we have unbiased estimate given as below:

$$E\left(\hat{\beta}_1^*\right) = \beta_1 \tag{2.31}$$

Variance of $\hat{\beta}_1^*$ is given by:

To prove whether $\hat{\beta}_1^*$ has minimum variance, let's compute $var(\hat{\beta}_1^*)$ with $var(\hat{\beta}_1^*)$.

$$\mathrm{var}\left(\hat{\beta}_1^*\right) = \mathrm{E}\left(\hat{\beta}_1^* - \mathrm{E}(\hat{\beta}_1^*)\right)^2 = \mathrm{E}\left(\hat{\beta}_1^* - \beta_1\right)^2$$

From equation (2.30), $\hat{\beta}_1^* - \beta_1 = \sum w_i \varepsilon_i$, then

$$var\left(\hat{\beta}_1^*\right) = E\left(\sum w_i \varepsilon_i\right)^2$$

$$= \mathrm{E}(w_1\varepsilon_1 + \mathrm{w}_2\varepsilon_2 + \cdots + w_{\mathrm{n}-1}\varepsilon_{\mathrm{n}-1} + \mathrm{w_n}\varepsilon_{\mathrm{n}})^2$$

$$= E[w_1^2\varepsilon_1^2 + w_2^2\varepsilon_2^2 + \cdots + w_n^2\varepsilon_n^2 + 2\mathrm{w}_1\mathrm{w}_2\varepsilon_1\varepsilon_2 + \cdots + 2w_{\mathrm{n}-1}\mathrm{w}_n\varepsilon_{\mathrm{n}-1}\varepsilon_{\mathrm{n}}]$$

$$= E[w_1^2\varepsilon_1^2 + w_2^2\varepsilon_2^2 + \cdots + w_n^2\varepsilon_n^2] + \mathrm{E}[2w_1w_2\varepsilon_1\varepsilon_2 + \cdots + 2\mathrm{w}_{\mathrm{n}-1}w_n\varepsilon_{\mathrm{n}-1}\varepsilon_{\mathrm{n}}]$$

$$= E\left[\sum w_i^2\varepsilon_i^2\right] + 2\mathrm{E}\left[\sum_{i=1}^{n-1}\sum_{j=2}^{n} \mathrm{w_i}w_j\varepsilon_{\mathrm{i}}\varepsilon_{\mathrm{j}}\right], \quad \text{for } i \neq j$$

$$= \sum w_i^2\,E(\varepsilon_i^2) + 2\sum_{i=1}^{n-1}\sum_{j=2}^{n} \mathrm{w_i}w_j\mathrm{E}\left(\varepsilon_{\mathrm{i}}\varepsilon_{\mathrm{j}}\right)\ , since\ \mathrm{w_i}\ and\ \mathrm{w}_j\ are\ constant$$

By assumption (6) & (7), $E(\varepsilon_i^2) = \sigma^2$ and $\mathrm{E}(\varepsilon_i\varepsilon_j) = 0, for \;\; i \neq j$

$$var(\hat{\beta}_1^*) = \sigma^2 \sum w_i^2$$

However, $\sum w_i^2 = \sum(k_i + c_i)^2 = \sum k_i^2 + 2\sum k_i c_i + \sum c_i^2$

Now let us check whether $\sum k_i c_i = 0$.

We know the weight $k_i = \frac{x_i}{\sum x_i^2}$ and from equation (2.29), $\sum c_i x_i = 0$.

Then

$$\sum k_i c_i = \frac{\sum x_i c_i}{\sum x_i^2} = \frac{0}{\sum x_i^2} = 0 \tag{2.32}$$

Thus, $\sum w_i^2 = \sum k_i^2 + \sum c_i^2$ (2.33)

Therefore, $var(\hat{\beta}_1^*) = \sigma^2(\sum k_i^2 + \sum c_i^2) \Longrightarrow \sigma^2 \sum k_i^2 + \sigma^2 \sum c_i^2$

From equation (2.20), $var(\hat{\beta}_1) = \sigma^2 \sum k_i^2$, then

$$var(\hat{\beta}_1^*) = var(\hat{\beta}_1) + \sigma^2 \sum c_i^2 \tag{2.34}$$

Given that c_i is an arbitrary constant, $\sigma^2 \sum c_i^2$ is a positive i.e., it is greater than zero. Thus, $var(\hat{\beta}_1^*) > var(\hat{\beta}_1)$. This proves that $\hat{\beta}_1$ possess minimum variance property. In the similar way we can prove that the least square estimate of the constant intercept ($\hat{\beta}_0$) possess minimum variance.

9.4 Minimum Variance of $\hat{\beta}_0$

We take a new estimator $\hat{\beta}_0^*$ which we assume to be a linear and unbiased estimator of the function of β_0. The least square estimator $\hat{\beta}_0$ is given by:

$$\hat{\beta}_0 = \sum \left(\frac{1}{n} - \bar{X}\mathrm{k_i}\right) Y_\mathrm{i} \tag{2.35}$$

By analogy with that the proof of the minimum variance property of $\hat{\beta}_1$, let's use the weights $w_i = \mathrm{c_i} + \mathrm{k_i}$.

Linearity of $\tilde{\beta}_0$ is given by:

$$\hat{\beta}_0^* = \sum \left(\frac{1}{n} - \bar{X}\mathrm{w_i}\right) Y_\mathrm{i} \tag{2.36}$$

Unbisedness of $\hat{\beta}_0^*$ is given as:

Since we want $\hat{\beta}_0^*$ to be an unbiased estimator of the true β_0, that is, $E(\hat{\beta}_0^*) = \beta_0$, we substitute for $Y_\mathrm{i} = \beta_0 + \beta_1 X_i + \varepsilon_i$ in $\hat{\beta}_0^*$ and find the expected value of $\hat{\beta}_0^*$.

$$\hat{\beta}_0^* = \sum \left(\frac{1}{n} - \bar{X}\mathrm{w_i}\right) (\beta_0 + \beta_1 X_i + \varepsilon_i)$$

$$= \sum \left(\frac{\beta_0}{n} + \frac{\beta_1 X_i}{n} + \frac{\varepsilon_i}{n} - \bar{X}\mathrm{w_i}\beta_0 - \beta_1 \bar{X} X_i \mathrm{w_i} - \bar{X}\mathrm{w_i}\varepsilon_i\right)$$

$$= \beta_0 + \beta_1\bar{X} + \sum \varepsilon_i/n - \beta_0\bar{X}\sum \mathrm{w_i} - \beta_1\bar{X}\sum \mathrm{w_i}X_i - \bar{X}\sum \mathrm{w_i}\varepsilon_i$$

For $\hat{\beta}_0^*$ to be an unbiased estimator of the true β_0, the following must hold. That is given equation (2.26 & 2.28), $\sum \mathrm{w_i} = 0$ and $\sum \mathrm{w_i} X_i = 1$, we have

$$\hat{\beta}_0^* = \beta_0 + \beta_1\bar{X} + \frac{1}{n}\sum \varepsilon_i - \beta_1\bar{X} - \bar{X}\sum w_i\varepsilon_i$$
$$\hat{\beta}_0^* = \beta_0 + \frac{1}{n}\sum \varepsilon_i - \bar{X}\sum w_i\varepsilon_i$$
$$\hat{\beta}_0^* - \beta_0 = \sum\left(\frac{1}{n} - \bar{X}w_i\right)\varepsilon_i \tag{2.37}$$

Taking expectation

$$E(\hat{\beta}_0^*) = \beta_0 \ + \sum (\frac{1}{n} - \bar{X}w_i)\, E(\varepsilon_i)$$

Since w_i and $\bar{X}$ are fixed and the only random variable is ε_i, we apply expectation only for random variables and its expected value is zero, ($E(\varepsilon_i) = 0$).
Finally, we have unbiased estimate given as below:
As we know, $E(\varepsilon_i) = 0$ and other variable are fixed, we obtained an unbiased estimate of $\hat{\beta}_0^*$ as:

$$E(\hat{\beta}_0^*) = \beta_0 \tag{2.38}$$

Variance of $\hat{\beta}_0^*$ is given by:
To prove whether $\hat{\beta}_0$ has minimum variance, let's compute $var(\hat{\beta}_0^*)$ with $var(\hat{\beta}_0)$.

$$var(\hat{\beta}_0^*) = E\left(\hat{\beta}_0^* - E(\hat{\beta}_0^*)\right)^2 = E\left(\hat{\beta}_0^* - \beta_0\right)^2$$

From equation (2.37), $\hat{\beta}_0^* - \beta_0 = \sum\left(\frac{1}{n} - \bar{X}w_i\right)\varepsilon_i$, then $var(\hat{\beta}_0^*)$ is given by:

$$var(\hat{\beta}_0^*) = E\left(\sum\left(\frac{1}{n} - \bar{X}w_i\right)\varepsilon_i\right)^2$$
$$= E\left[\left(\frac{1}{n} - \bar{X}w_1\right)\varepsilon_1 + \left(\frac{1}{n} - \bar{X}w_2\right)\varepsilon_2 + \cdots + \left(\frac{1}{n} - \bar{X}w_{n-1}\right)\varepsilon_{n-1} + \left(\frac{1}{n} - \bar{X}w_n\right)\varepsilon_n\right]^2$$
$$= E\Bigg[\left(\frac{1}{n} - \bar{X}w_1\right)^2\varepsilon_1^2 + \left(\frac{1}{n} - \bar{X}w_2\right)^2\varepsilon_2^2 + \cdots + \left(\frac{1}{n} - \bar{X}w_n\right)^2\varepsilon_n^2 + 2\left(\frac{1}{n} - \bar{X}w_1\right)\left(\frac{1}{n} - \bar{X}w_2\right)\varepsilon_1\varepsilon_2 + \cdots + 2\left(\frac{1}{n} - \bar{X}w_{n-1}\right)\left(\frac{1}{n} - \bar{X}w_n\right)\varepsilon_{n-1}\varepsilon_n\Bigg]$$
$$= E\left[\left(\frac{1}{n} - \bar{X}w_1\right)^2\varepsilon_1^2 + \left(\frac{1}{n} - \bar{X}w_2\right)^2\varepsilon_2^2 + \cdots + \left(\frac{1}{n} - \bar{X}w_n\right)^2\varepsilon_n^2\right] + E\left[2\left(\frac{1}{n} - \bar{X}w_1\right)\left(\frac{1}{n} - \bar{X}w_2\right)\varepsilon_1\varepsilon_2 + \cdots + 2\left(\frac{1}{n} - \bar{X}w_{n-1}\right)\left(\frac{1}{n} - \bar{X}w_n\right)\varepsilon_{n-1}\varepsilon_n\right]$$

$$= E\sum(\frac{1}{n} - \bar{X}w_i)^2\,\varepsilon_i{}^2 + 2\mathrm{E}\left[\sum_{i=1}^{n-1}\sum_{j=2}^{n}\left(\frac{1}{n} - \bar{X}w_i\right)\left(\frac{1}{n} - \bar{X}w_j\right)\varepsilon_\mathrm{i}\varepsilon_\mathrm{j}\right], \rightarrow i \neq j$$

Since n, $\bar{X}$ and w_i are constant, apply expectation only for random variable, ε_i

$$= \sum\left(\frac{1}{n} - \bar{X}w_i\right)^2 E(\varepsilon_i{}^2) + 2\left[\sum_{i=1}^{n-1}\sum_{j=2}^{n}\left(\frac{1}{n} - \bar{X}w_i\right)\left(\frac{1}{n} - \bar{X}w_j\right)E(\varepsilon_\mathrm{i}\varepsilon_\mathrm{j})\right], \rightarrow i \neq j$$

By assumption (6) & (7), $E(\varepsilon_i{}^2) = \sigma^2$ and $cov(\varepsilon_i\varepsilon_j) = 0,\ i \neq j$, respectively, then

$$var(\hat{\beta}_0^*) = \left(\sum(\frac{1}{n} - \bar{X}w_i)^2\,\sigma^2\right)$$

$$= \sigma^2\sum(\frac{1}{n} - \bar{X}w_i)^2$$

$$= \sigma^2\sum\left(\frac{1}{n^2} + \bar{X}^2w_i^2 - 2\frac{1}{n}\bar{X}w_i\right)$$

$$= \sigma^2\left(\frac{1}{n} + \bar{X}^2\sum w_i^2 - 2\frac{1}{n}\bar{X}\sum w_i\right)$$

From equation (2.26) $\sum w_i = 0$, then

$$var(\hat{\beta}_0^*) = \sigma^2\left(\frac{1}{n} + \bar{X}^2\sum w_i^2\right)$$

Moreover, from equation (2.33), $\sum w_i^2 = \sum k_i^2 + \sum c_i^2$, thus

$$var(\hat{\beta}_0^*) = \sigma^2\left(\frac{1}{n} + \bar{X}^2\left(\sum k_i^2 + \sum c_i^2\right)\right) = \sigma^2\frac{1}{n} + \sigma^2\bar{X}^2\sum k_i^2 + \sigma^2\bar{X}^2\sum c_i^2$$

$$= \sigma^2\left(\frac{1}{n} + \bar{X}^2\sum k_i^2\right) + \sigma^2\bar{X}^2\sum c_i^2$$

$$var(\hat{\beta}_0^*) = \sigma^2\left(\frac{1}{n} + \frac{\bar{X}^2}{\sum x_i^2}\right) + \sigma^2\bar{X}^2\sum c_i^2 \tag{2.39}$$

From equation (2.22), $\frac{1}{n} + \frac{\bar{X}^2}{\sum x_i^2} = \frac{\sum X_i^2}{n\sum x_i^2}$

$$var(\hat{\beta}_0^*) = \sigma^2\left(\frac{\sum X_i^2}{n\sum x_i^2}\right) + \sigma^2\bar{X}^2\sum c_i^2 \tag{2.40}$$

From equation (2.23), $var(\hat{\beta}_0) = \sigma^2\left(\frac{\sum X_i^2}{n\sum x_i^2}\right)$

Thus, $var(\hat{\beta}_0^*) = var(\hat{\beta}_0) + \sigma^2\bar{X}^2\sum c_i^2$

Hence, $var(\hat{\beta}_0^*) > var(\hat{\beta}_0)$, since $\sigma^2\bar{X}^2\sum c_i^2$ is quadratic (all are squared) its value is positive. i.e. $\sigma^2\bar{X}^2\sum c_i^2 > 0$.

Therefore, we have proved that the least square estimators of linear regression model are best, linear, and unbiased estimators. The sampling variance of the OLS estimators ($var(\hat{\beta}_0)$ and $var(\hat{\beta}_1)$) measure the statistical precision of the coefficient of OLS estimators of $\hat{\beta}_0$ and $\hat{\beta}_1$.

The $var(\hat{\beta}_0)$ and $var(\hat{\beta}_1)$ measure the statistical **precision** of the coefficient of

OLS estimators of $\hat{\beta}_0$ and $\hat{\beta}_1$. Therefore, the estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ to be precise, the variance of $\hat{\beta}_0$ and $\hat{\beta}_1$ ($var(\hat{\beta}_0)$ and $var(\hat{\beta}_1)$) should be minimum.
Thus, the $var(\hat{\beta}_0)$ and $var(\hat{\beta}_1)$ become minimum if:

- The smaller the error variance σ^2 i.e., the smaller the variance of the unobserved and unknown random influences on Y_i.
- The larger is the sample variation of X_i about their sample mean, i.e., the larger the values of $x_i^2 = (X_i - \bar{X})^2$, i=1,2….,n.
- The larger is the size of the sample, i.e., the larger is n.

9.5 Covariance $\hat{\boldsymbol{\beta}}_0$ and $\hat{\boldsymbol{\beta}}_1$

The covariance of the OLS coefficient estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ is defined as:

$$cov(\hat{\beta}_0, \hat{\beta}_1) = E\{[\hat{\beta}_0 - \mathrm{E}(\hat{\beta}_0)][\hat{\beta}_1 - \mathrm{E}(\hat{\beta}_1)]\}$$

Derivation of expression for $cov(\hat{\beta}_0, \hat{\beta}_1)$:

$$cov(\hat{\beta}_0, \hat{\beta}_1) = E\{[\hat{\beta}_0 - \mathrm{E}(\hat{\beta}_0)][\hat{\beta}_1 - \mathrm{E}(\hat{\beta}_1)]\}$$

From equation (2.15 & 2.17), $(\hat{\beta}_1 - \beta_1) = \sum \mathrm{k_i}\, \varepsilon_i$ and $(\hat{\beta}_0 - \beta_0) = \sum \left(\frac{1}{n} - \bar{X}\mathrm{k_i}\right) \varepsilon_i$, respectively, then

$$cov(\hat{\beta}_0, \hat{\beta}_1) = E\left\{\left[\sum \mathrm{k_i}\, \varepsilon_i\right]\left[\sum \left(\frac{1}{n} - \bar{X}\mathrm{k_i}\right) \varepsilon_i\right]\right\}$$

$$= E\left\{[k_1\varepsilon_1 + k_2\varepsilon_2 + \cdots + k_n\varepsilon_n]\left[\left(\frac{1}{n} - \bar{X}\mathrm{k_1}\right)\varepsilon_1 + \left(\frac{1}{n} - \bar{X}\mathrm{k_2}\right)\varepsilon_2 + \cdots + \left(\frac{1}{n} - \bar{X}\mathrm{k_n}\right)\varepsilon_n\right]\right\}$$

$$= E\left\{\left[k_1\left(\frac{1}{n} - \bar{X}\mathrm{k_1}\right)\varepsilon_1\varepsilon_1 + k_2\left(\frac{1}{n} - \bar{X}\mathrm{k_2}\right)\varepsilon_2\varepsilon_2 + \cdots + k_n\left(\frac{1}{n} - \bar{X}\mathrm{k_n}\right)\varepsilon_n\varepsilon_n\right] + 2k_1\left(\frac{1}{n} - \bar{X}\mathrm{k_2}\right)\varepsilon_1\varepsilon_2 + \cdots + 2k_{n-1}\left(\frac{1}{n} - \bar{X}k_n\right)\varepsilon_{n-1}\varepsilon_n\right\}$$

$$= E\sum k_i(\frac{1}{n} - \bar{X}k_i)\,{\varepsilon_i}^2 + 2\mathrm{E}\left[\sum_{i=1}^{n-1}\sum_{j=2}^{n} k_i\left(\frac{1}{n} - \bar{X}k_j\right)\varepsilon_\mathrm{i}\varepsilon_\mathrm{j}\right], \rightarrow i \neq j$$

$$= \sum k_i\left(\frac{1}{n} - \bar{X}k_i\right)E({\varepsilon_i}^2) + 2\mathrm{E}\left[\sum_{i=1}^{n-1}\sum_{j=2}^{n} k_i\left(\frac{1}{n} - \bar{X}k_j\right)E(\varepsilon_\mathrm{i}\varepsilon_\mathrm{j})\right], \rightarrow i \neq$$

$$= \sum k_i\left(\frac{1}{n} - \bar{X}k_i\right)\sigma^2, \qquad since\ E({\varepsilon_i}^2)\ and\ \mathrm{E}(\varepsilon_\mathrm{i}\varepsilon_\mathrm{j}) = 0, i \neq$$

$$= \sigma^2\left(\frac{1}{n}\sum k_i - \bar{X}\sum k_i^2\right)$$

$cov(\hat{\beta}_0, \hat{\beta}_1) = \sigma^2(-\bar{X}\sum k_i^2)$, since $\sum k_i = 0$

$= \sigma^2\left(-\frac{\bar{X}}{\sum x_i^2}\right) = -\bar{X}\left(\frac{\sigma^2}{\sum x_i^2}\right)$, since $\sum k_i^2 = \frac{1}{\sum x_i^2}$

$= -\bar{X}\left(Var(\hat{\beta}_1)\right)$ since $Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum x_i^2}$

$$\therefore Cov(\hat{\beta}_0, \hat{\beta}_1) = -\bar{X}\left(\frac{\sigma^2}{\sum x_i^2}\right) = -\bar{X}\left(Var(\hat{\beta}_1)\right) \tag{2.41}$$

Since both σ^2 and $\sum x_i^2$ are positive, the sign of $Cov(\hat{\beta}_0, \hat{\beta}_1)$ dependes on the sign of $-\bar{X}$.

If $\bar{X} > 0$, $Cov(\hat{\beta}_0, \hat{\beta}_1) < 0$: the sampling errors $(\hat{\beta}_0 - \beta_0)$ and $(\hat{\beta}_1 - \beta_1)$ are of opposite sign.

If $\bar{X} < 0$, $Cov(\hat{\beta}_0, \hat{\beta}_1) > 0$: the sampling errors $(\hat{\beta}_0 - \beta_0)$ and $(\hat{\beta}_1 - \beta_1)$ are of same sign.

9.6 Estimation of the Population Variance

As we know that the variances of the OLS estimates incorporate σ^2, which is the population variance of the random or disturbance term. However, it is difficult to obtain for the population value of the disturbance term because of technical and economic reasons. Hence, it is difficult to compute σ^2, and estimating the variances of OLS estimates are also difficult. However, we can compute these variances if we take the unbiased estimate of σ^2 which is $\hat{\sigma}^2$ computed from the sample value of the disturbance term $\hat{\varepsilon}_i$ in the expression:

$$\hat{\sigma}_\varepsilon^2 = \frac{\sum\left(Y_i - \hat{Y}\right)^2}{n-2} = \frac{\sum \hat{\varepsilon}_i^2}{n-2} \tag{2.42}$$

To use $\hat{\sigma}^2$ in the expressions for the variances of $\hat{\beta}_0$ *and* $\hat{\beta}_1$ we have to prove whether $\hat{\sigma}^2$ is an unbiased estimator of σ^2, i.e., $E(\hat{\sigma}^2) = E\left(\frac{\sum \hat{\varepsilon}_i^2}{n-2}\right) = \sigma^2$.

Proof:

To proof this we have to compute $\sum \hat{\varepsilon}_i^2$ from the expressions of Y, $\hat{Y}$, y, $\hat{y}$ and $\hat{\varepsilon}_i$.

$$Y = \hat{\beta}_0 + \hat{\beta}_1 X + \hat{\varepsilon}$$
$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$$

$$Y = \hat{Y} + \hat{\varepsilon} \tag{2.43}$$

$$\hat{\varepsilon} = Y - \hat{Y} \tag{2.44}$$

Apply summing for equation (2.43) will result the following expression

$$\sum Y_i = \sum \hat{Y}_i + \sum \hat{\varepsilon}_i$$
$$\sum Y_i = \sum \hat{Y}_i$$

Given that $\sum \hat{\varepsilon}_i = 0$, now let us check whether the sum of residual is equal to zero.

$$\sum_{i=1}^{n} \hat{\varepsilon}_i = \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i) = \sum_{i=1}^{n} Y_i - n\,\hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^{n} X_i$$
$$= n\bar{Y} - n\,\hat{\beta}_0 - n\,\hat{\beta}_1 \bar{X} = n(\bar{Y} - \hat{\beta}_0 - \hat{\beta}_1 \bar{X})$$

$$\sum_{i=1}^{n} \hat{\varepsilon}_i = n\left(\bar{Y} - \left(\bar{Y} - \hat{\beta}_1 \bar{X}\right) - \hat{\beta}_1 \bar{X}\right) = 0$$

Dividing both sides the above by the number of observation 'n' will give us

$$\frac{\sum Y_i}{n} = \frac{\sum \hat{Y}_i}{n} \rightarrow \bar{Y} = \bar{\hat{Y}} \tag{2.45}$$

Putting equation (2.43) and (2.44) together and subtract

$$Y = \hat{Y} + \hat{\varepsilon}$$
$$\bar{Y} = \bar{\hat{Y}}$$
$$\Rightarrow (Y - \bar{Y}) = \left(\hat{Y} - \bar{\hat{Y}}\right) + \hat{\varepsilon}$$
$$\Rightarrow y_i = \hat{y} + \hat{\varepsilon} \tag{2.46}$$

From equation (2.46):

$$\hat{\varepsilon} = y - \hat{y} \tag{2.47}$$

where the y's are in deviation form.
Now, we should to find $y \; and \; \hat{y}$ in functional expression as derived below.
The regression equation is given by:

$$Y = \beta_0 + \beta_1 X_i + \varepsilon$$

Then the average value of the regression is given by:

$$\bar{Y} = \beta_0 + \beta_1 \bar{X} + \bar{\varepsilon}$$

Note: In this case, we assumed earlier that, $E(\varepsilon_i) = 0$, i.e. in taking a very large number of samples we expect ε_i to have a mean value of zero, but in any particular single sample $\bar{\varepsilon}$ is not necessarily zero.
Now by subtraction we obtain

$$y = (Y - \bar{Y}) = \beta_1 (X - \bar{X}) + (\varepsilon - \bar{\varepsilon}) = \beta_1 x + (\varepsilon - \bar{\varepsilon})$$
$$y = \beta_1 x_i + (\varepsilon - \bar{\varepsilon}) \tag{2.48}$$

Similarly from the fitted value:

$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$

Then the average value of the fitted line is given by:

$\bar{\hat{Y}} = \hat{\beta}_0 + \hat{\beta}_1 \bar{X}$

By subtraction of the mean value of the fitted value from the total fitted value we get

$$\hat{y} = \hat{Y} - \bar{\hat{Y}} = \hat{\beta}_0 + \hat{\beta}_1 X - \left(\hat{\beta}_0 + \hat{\beta}_1 \bar{X}\right) = \hat{\beta}_1 (X - \bar{X})$$
$$\hat{y} = \hat{\beta}_1 x \tag{2.49}$$

Substituting (2.48) and (2.49) in (2.47) we get

$$\hat{\varepsilon} = y - \hat{y} = \beta_1 x + (\varepsilon - \bar{\varepsilon}) - \hat{\beta}_1 x$$
$$= (\varepsilon - \bar{\varepsilon}) - \left(\hat{\beta}_1 - \beta_1\right) x$$

The summation of the squared residuals over the 'n' samples on both side yields:

$$\sum \hat{\varepsilon}_i^2 = \sum \left[(\varepsilon_i - \bar{\varepsilon}) - \left(\hat{\beta}_1 - \beta_1\right) x_i\right]^2$$

$$=\sum\left[(\varepsilon_i-\bar{\varepsilon})^2+\left(\hat{\beta}_1-\beta_1\right)^2x_i^2-2\left(\hat{\beta}_1-\beta_1\right)x_i(\varepsilon_i-\bar{\varepsilon})\right]$$
$$=\sum(\varepsilon_i-\bar{\varepsilon})^2+\left(\hat{\beta}_1-\beta_1\right)^2\sum{x_i}^2-2\left[\left(\hat{\beta}_1-\beta_1\right)\sum x_i(\varepsilon_i-\bar{\varepsilon})\right]$$

To find the unbiasedness, we apply expectation for the whole function

$$E\left(\sum\hat{\varepsilon}_i^2\right)=E\left(\sum(\varepsilon_i-\bar{\varepsilon})^2\right)+E\left[\left(\hat{\beta}_1-\beta_1\right)^2\sum{x_i}^2\right]$$
$$-2E\left[\left(\hat{\beta}_1-\beta_1\right)\sum x_i(\varepsilon_i-\bar{\varepsilon})\right]\qquad(2.50)$$

The right hand side terms of equation (2.50) may be rearranged as follows

a. $E\left(\sum(\varepsilon_i-\bar{\varepsilon})^2\right)=E\sum({\varepsilon_i}^2-2\bar{\varepsilon}\varepsilon_i+\bar{\varepsilon}^2)=E\left(\sum\varepsilon_i^2-2\bar{\varepsilon}\sum\varepsilon_i+\bar{\varepsilon}\sum\varepsilon_i\right)$

$$=E\left(\sum\varepsilon_i^2-\bar{\varepsilon}\sum\varepsilon_i\right)$$
$$=E\left(\sum\varepsilon_i^2-\frac{(\sum\varepsilon_i)^2}{n}\right)=\sum E(\varepsilon_i^2)-\frac{E(\sum\varepsilon_i)^2}{n}$$

$=\sum{\sigma_\varepsilon}^2-\frac{1}{n}E(\varepsilon_1+\varepsilon_2+\cdots+\varepsilon_n)^2=n{\sigma_\varepsilon}^2-\frac{1}{n}E(\varepsilon_1+\varepsilon_2+\cdots+\varepsilon_n)^2$ since $E(\varepsilon_i^2)=\sigma_\varepsilon^2$

$$=n{\sigma_\varepsilon}^2-\frac{1}{n}E\left(\sum{\varepsilon_i}^2+2\sum_{i=1}^{n-1}\sum_{j=2}^{n}\varepsilon_i\varepsilon_j\right)$$
$$=n{\sigma_\varepsilon}^2-\frac{1}{n}\sum E({\varepsilon_i}^2)-\frac{2}{n}\sum E\left(\varepsilon_i,\varepsilon_j\right),i\neq j$$
$$=n{\sigma_\varepsilon}^2-{\sigma_\varepsilon}^2,\left(\,given\;E\left(\varepsilon_i,\varepsilon_j\right)=0\right),for\;\;i\neq j$$
$$={\sigma_\varepsilon}^2(n-1)\qquad(2.51)$$

b. $E\left[\left(\hat{\beta}_1-\beta_1\right)^2\sum{x_i}^2\right]=\sum{x_i}^2\,E\left(\hat{\beta}_1-\beta_1\right)^2$

Given that the values of x_i is derived from X's which are fixed in all samples and we know that

$$E\left(\hat{\beta}_1-\beta_1\right)^2=var\left(\hat{\beta}_1\right)={\sigma_\varepsilon}^2\frac{1}{\sum{x_i}^2}$$
$$=\sum{x_i}^2\,{\sigma_\varepsilon}^2\frac{1}{\sum{x_i}^2}$$

Hence,

$$\sum{x_i}^2E\left(\hat{\beta}_1-\beta_1\right)^2={\sigma_\varepsilon}^2\qquad(2.52)$$

c. $-2E\left[\left(\hat{\beta}_1-\beta_1\right)\sum x_i(\varepsilon_i-\bar{\varepsilon})\right]=-2E\left[\left(\hat{\beta}_1-\beta_1\right)\left(\sum x_i\varepsilon_i-\bar{\varepsilon}\sum x_i\right)\right]$

$=-2E\left[\left(\hat{\beta}_1-\beta_1\right)\left(\sum x_i\varepsilon_i\right)\right]$ since $\sum x_i=0$

However, from (2.8), $\left(\hat{\beta}_1-\beta_1\right)=\sum k\varepsilon_i$ and substitute it in the above expression, we will get:

$$=-2E\left[\left(\sum k_i\varepsilon_i\right)\left(\sum x_i\varepsilon_i\right)\right]$$

where $k_i=\frac{x_i}{\sum{x_i}^2}=$

$$-2E\left[\sum\left(\left(\frac{x_i}{\sum x_i{}^2}\right)\varepsilon_i\right)\left(\sum x_i\varepsilon_i\right)\right] = -2E\left[\frac{(\sum x_i\varepsilon_i)}{\sum x_i{}^2}\left(\sum x_i\varepsilon_i\right)\right]$$
$$= -2E\left[\frac{(\sum x_i\varepsilon_i)^2}{\sum x_i{}^2}\right]$$
$$= -2E\left[\frac{(x_1\varepsilon_1 + x_2\varepsilon_2 \ldots + x_{n-1}\varepsilon_{n-1} + x_n\varepsilon_n)^2}{\sum x_i{}^2}\right]$$
$$= -2E\left[\frac{(\sum x_i{}^2\varepsilon_i{}^2) + 2(\sum_{i=1}^{n-1}\sum_{j=2}^{n} x_ix_j\varepsilon_i\varepsilon_j}{\sum x_i{}^2}\right]$$
$$= -2\left[\frac{\sum x_i{}^2\, E(\varepsilon_i{}^2)}{\sum x_i{}^2} + \frac{2(\sum\sum x_ix_jE(\varepsilon_i\varepsilon_j)}{\sum x_i{}^2}\right], \qquad i \neq j$$
$$= -2\frac{\sum x_i{}^2\, E(\varepsilon_i{}^2)}{\sum x_i{}^2}\,(given\, E(\varepsilon_i\varepsilon_j) = 0,\;\; i \neq j)$$
$$-2E(\varepsilon_i{}^2) = -2\sigma^2 \tag{2.53}$$

Consequently, equation (51) can be written in terms of (2.51), (2.52), and (2.53) as follows:

$$E\left(\sum \hat{\varepsilon}_i^2\right) = \sigma_\varepsilon{}^2(n-1) + \sigma_\varepsilon{}^2 - 2\sigma_\varepsilon{}^2 \tag{2.54}$$

From which we get

$$E\left(\sum \hat{\varepsilon}_i^2\right) = \sigma_\varepsilon{}^2 n - 2\sigma_\varepsilon{}^2 = \sigma_\varepsilon{}^2(n-2)$$
$$E\left(\sum \hat{\varepsilon}_i^2\right) = \sigma_\varepsilon{}^2(n-2)$$

Divide both sides by $n-2$ results

$$E\left(\frac{\sum \hat{\varepsilon}_i^2}{n-2}\right) = \sigma_\varepsilon{}^2$$

By equation (2.42), $\hat{\sigma}_\varepsilon^2 = \frac{\sum(Y_i - \hat{Y})^2}{n-2} = \frac{\sum \hat{\varepsilon}_i^2}{n-2}$. where $n-2$ is the df.

Thus,

$$E\left(\frac{\sum \hat{\varepsilon}_i^2}{n-2}\right) = E(\hat{\sigma}_\varepsilon^2) = \sigma_\varepsilon{}^2 \tag{2.53}$$

Therefore, the sample variance of an error term, $\hat{\sigma}_\varepsilon^2 = \frac{\sum \hat{\varepsilon}_i^2}{n-2}$ is unbiased estimate of the true or population variance of the error term ($\sigma_\varepsilon{}^2$).

Now, we can substitute $\hat{\sigma}_\varepsilon^2 = \frac{\sum \hat{\varepsilon}_i^2}{n-2}$ for $\sigma_\varepsilon{}^2$ in the variance expression of $\hat{\beta}_0$ *and* $\hat{\beta}_1$, since $E(\hat{\sigma}_\varepsilon^2) = \sigma_\varepsilon{}^2$.

Hence the formula of variance of $\hat{\beta}_0$ *and* $\hat{\beta}_1$, becomes;

$$var(\hat{\beta}_1) = \frac{\hat{\sigma}_\varepsilon^2}{\sum x_i^2} = \frac{\sum \hat{\varepsilon}_i^2}{(n-2)\sum x_i^2} \tag{2.54}$$
$$var(\hat{\beta}_0) = \hat{\sigma}_\varepsilon^2\left(\frac{\sum X_i^2}{n\sum x_i^2}\right) = \frac{\sum \hat{\varepsilon}_i^2 \sum X_i^2}{n(n-2)\sum x_i^2} \tag{2.55}$$

where $var(\hat{\beta}_1)$ and $var(\hat{\beta}_0)$ measure the statistical precision of the OLS coefficient estimators $(\hat{\beta}_1)$ and $\hat{\beta}_0$, respectively.

10. Conclusion

The classical regression model is still the fundamental even for the development of new approaches, which is based on extensions of the classical assumptions. Following that, in this study, we try to cover the basic concept of regression analysis especially simple linear regression, their assumptions known as Gauss-Markov assumptions, estimation of the parameters using ordinary least square (OLS), and proofing the best linear unbiased estimator properties of OLS estimates. Thus, in the estimation of classical linear regression model using an ordinary least square should satisfy the Gauss Markov assumptions. If the Gauss–Markov assumptions hold true, the OLS procedure creates the best possible estimates.

Reference

[1] Arthur, S. Goldberger. (1965). Econometric Theory. The Swedish Journal of Economics, vol. 67, No. 3, pp. 256-258, https://doi.org/10.2307/3439032
[2] Hornby, A. S. (2000). Oxford Advanced Learner's Dictionary, 6th ed., Oxford University Press.
[3] Levins, R. (1966). The Strategy of Model Building in Population Biology. American Scientist, 54(4), 421–431. http://www.jstor.org/stable/27836590.
[4] Maddala, G. S. (1992). Introduction to Econometrics, 2nd Edition. Macmillan Publishing Company, 866 Third Avenue, New York, New York 10022.
[5] Ragnar Frisch. (1933). Econometrica. Journal of the Econometric Society, Vol. 1, No. 1.
[6] Wooldridge, J. M. (2002). Econometric Analysis of Cross Section and Panel Data. Cambridge, MA: MIT Press.