KAMPALA INTERNATIONAL UNIVERSITY

COLLEGE OF ECONOMICS AND MANAGEMENT

DEPARTMENT OF ECONOMICS AND STATISTICS

COURSE OUTLINE FOR ECONOMETRIC METHODS (STA 3101)

- RECAP OF INTRODUCTORY ECONOMETRICS
    - Definition and objectives of econometrics
    - Steps of econometric model analysis
    - Simple econometric model
    - Model assumptions
    - Distribution of the independent variable
- Estimation of parameters using Least squares estimation method (LSE), Best linear unbiased method (BLUE)
    - Estimation of parameters
    - Properties of LS estimators
    - BLUE method
- Estimation using the maximum likelihood estimation method (MLE)
    - Estimation of parameters
    - Properties of MLE
- Multiple regression
    - Estimation of parameters
    - Interval estimation of the estimators
    - Goodness of fit
    - Prediction
    - Hypotheses testing
- VIOLATION OF BASIC ASSUMPTIONS
    - Violation of normality
    - Violation of zero mean
    - Violation of homoscedasticity
    - Violation of non-auto regression
    - Violation of non-multicollinearity

REFERENCES

CHAPTER ONE

1.0 **WHAT IS ECONOMETRICS?**

Is concerned with the testing the theoretical propositions embodied in relations and with estimating the parameters involved. Econometrics is the science that combines economic theory with economic statistics and tries by mathematical and statistical methods to investigate the empirical support of the general law established by economic theory.

It is a composition of economics, mathematics and statistics. Where economics is for developing a hypothesis, mathematics is for model building in a mathematical form and statistics deals with using statistical techniques to analyse the economic model, to estimate the unknown parameters of the model and using the estimates for statistical inference.

Econometrics may be defined as the quantitative analysis of actual economic phenomena based on the concurrent development of theory and observation, related by appropriate methods of inference.

1.1 OBJECTIVES/ GOALS OF ECONOMETRICS

a) To judge the validity of economic theory.
b) To supply the numerical estimates of the coefficients of the economic relationships that may be used for sound economic policies.
c) To forecast the future values of the economic magnitude with a certain degree of probability.

1.4 CATEGORIES OF ECONOMETRICS

It is distinguished into two categories;

i. Theoretical econometrics: deals with the development of the appropriate methods for measuring economic relationships described by econometric models. These methods may be classified into two groups;
   - Single equation techniques (simple regression analysis) which are applied to one relation at a time.
   - Simultaneous equation techniques (multiple regression) which are applied to all relationships of the model simultaneously.

Theoretical econometrics is concerned with spelling out the assumptions of the above methods, their properties and what happens when one or more of the assumptions of the methods are not fulfilled.

ii. Applied econometrics: describes the practical value of econometric research. It deals with the application of econometric techniques developed in theoretical econometrics to different fields of economic theory for its verification and forecasting. Applied econometrics makes it possible to obtain numerical results from studies that are of great importance to planners.

I.3 METHODOLOGY OF ECONOMETRICS

The **traditional** or **classical** methodology, which still dominates empirical research in economics and other social and behavioral sciences involves the following steps:

1. Statement of theory or hypothesis.
2. Specification of the mathematical model of the theory
3. Specification of the statistical, or econometric, model
4. Obtaining the data
5. Estimation of the parameters of the econometric model
6. Hypothesis testing
7. Forecasting or prediction
8. Using the model for control or policy purposes.

CHAPTER TWO

2.0 ESTIMATION OF PARAMETERS USING LSE METHOD

Using the simple regression model of the form $y_i = \beta_o + \beta_1 x_i + \varepsilon_i$ where β_o and β_1 are the parameters to be estimated and ε is the error term.

2.1 **BASIC ASSUMPTIONS OF THE MODEL**

These are referred to as the basic classical assumptions and include;

- Normality of the error term
- Error term has zero mean, i.e $E(\varepsilon_i)=0$
- Constant variance (homoscedasticity) i.e $E(\varepsilon^2)=\delta^2$
- Non-auto-regression i.e $E(\varepsilon_i \varepsilon_j) = 0 \; for \; i \neq j$.
- The regression model is linear in parameters
- Zero covariance between the error term and the explanatory variable i,e $E(E_i x_i)=0$
- Non-stochastic explanatory variable. The values of x are fixed in repeated samples.

- The number of observations n must be greater than the number of parameters to be estimated. Alternatively, the number of observations n must be greater than the number of explanatory variables.
- Variability in x values. The x values in a given sample must not all be the same.
- The regression model is correctly specified. Alternatively, there is no specification bias or error in the model used in empirical analysis.
- There is no perfect multicollinearity, that is there is no perfect linear relationship among the explanatory variables.

2.2 **THE SIGNIFICANCE OF THE STOCHASTIC DISTURBANCE TERM**

The disturbance term is a surrogate for all those variables that are omitted from the model but that collectively affect *Y*. The reasons are many.

1. *Vagueness of theory:* The theory, if any, determining the behavior of *Y* may be, and often is, incomplete.

2. *Unavailability of data:* Even if we know what some of the excluded variables are and therefore consider a multiple regression rather than a simple regression, we may not have quantitative information about these.

3. *Core variables versus peripheral variables:* Assume in a consumption income example that besides income *X*1, the number of children per family *X*2, sex *X*3, religion *X*4, education *X*5, and geographical region *X*6 also affect consumption expenditure. But it is quite possible that the joint influence of all or some of these variables may be so small and at best nonsystematic or random that as a practical matter and for cost considerations it does not pay to introduce them into the model explicitly. One hopes that their combined effect can be treated as a random variable.

4. *Intrinsic randomness in human behavior:* Even if we succeed in introducing all the relevant variables into the model, there is bound to be some“intrinsic” randomness in individual *Y*’s that cannot be explained no matter how hard we try. The disturbances, may very well reflect this intrinsic randomness.

5. *Poor proxy variables:* Although the classical regression model assumes that the variables *Y* and *X* are measured accurately, in practice the data may be plagued by errors of measurement.

6. *Principle of parsimony:* we would like to keep our regression model as simple as possible. If we can explain the behavior of *Y* "substantially" with two or three explanatory variables and if our theory is not strong enough to suggest what other variables might be included, why introduce more variables? Let *the error term* represent all other variables. Of course, we should not exclude relevant and important variables just to keep the regression model simple.

7. *Wrong functional form:* Even if we have theoretically correct variables explaining a phenomenon and even if we can obtain data on these variables, very often we do not know the form of the functional relationship between the regressand and the regressors. Is consumption expenditure a linear (invariable) function of income or a nonlinear (invariable) function? If it is the former, $Yi = \beta 1 + B2Xi + Ei$ is the proper functional relationship between *Y* and *X*, but if it is the latter, $Yi = \beta 1 + \beta 2Xi + \beta 3X2i + \mathrm{E}i$ may be the correct functional form. In two-variable models the functional form of the relationship can often be judged from the scatter gram. But in a multiple regression model, it is not easy to determine the appropriate functional form, for graphically we cannot visualize scattergrams in multiple dimensions.

Using LSE method;

- From the model , make the error term the subject
- Form sums of squares
- Minimize the sums of squares (differentiate with respect to the parameters) and equate the results to zero.
- Form normal regression equations and solve for the parameters using any methods e.g matrix method.

2.3 Properties of Least-Squares Estimators: The Gauss–Markov Theorem

The least squares estimates possess some ideal or optimum properties. These properties are contained in the well-known **Gauss–Markov**

theorem. To understand this theorem, we need to consider the **best linear unbiasedness property** of an estimator.

An estimator, say the OLS estimator $\hat{\beta}2$, is said to be a best linear unbiased estimator (BLUE) of $\beta 2$ if the following hold:
1. It is **linear,** that is, a linear function of a random variable, such as the dependent variable Y in the regression model.

2. It is **unbiased,** that is, its average or expected value, $E(\hat{\beta}2)$, is equal to the true value, $\beta 2$.
3. It has minimum variance in the class of all such linear unbiased estimators; an unbiased estimator with the least variance is known as an **efficient estimator.**
In the regression context it can be proved that the OLS estimators are BLUE. This is the gist of the famous Gauss–Markov theorem, which can be stated as follows:

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

2.4 Properties of OLS Estimators under the Normality Assumption

With the assumption that *the* disturbance terms follow the normal distribution ,the OLS estimators have the following properties(desirable statistical properties of estimators):
1. They are unbiased.
2. They have minimum variance. Combined with 1, this means that they are **minimum variance unbiased,** or **efficient estimators.**
3. They have **consistency;** that is, as the sample size increases indefinitely, the estimators converge to their true population values.
4. $\hat{\beta}1$ (being a linear function of the error term) is *normally distributed* with Mean: $E(\hat{\beta}1) = \beta 1$
5. $\hat{\beta}2$ (being a linear function of the error term) is *normally* distributed
$\hat{\beta}2 \sim N\_(\beta 2, \sigma 2)$, Mean: $E(\hat{\beta}2) = \beta 2$ and var $(\hat{\beta}2)$: $\sigma 2/\sum x_i^2$.
6. $(n - 2)/(\hat{\sigma} 2/\sigma 2)$ is distributed as the $\chi 2$ (chi-square) distribution with $(n - 2)$df. This knowledge will help us to draw inferences about the true $\sigma 2$ from the estimated $\sigma 2$.
7. $(\hat{\beta}1, \hat{\beta}2)$ are distributed independently of $\hat{\sigma} 2$.

8. ˆ β1 *and* ˆ β2 *have minimum variance in the entire class of unbiased estimators, whether linear or not.* This result, due to Rao, is very powerful because, unlike the Gauss–Markov theorem, it is not restricted to the class of linear estimators only. Therefore, we can say that the least-squares estimators are **best unbiased estimators (BUE);** that is, they have minimum variance in the entire class of unbiased estimators.

To sum up: The important point to note is that the normality assumption enables us to derive the probability, or sampling, distributions of ˆ β1 and ˆ β2 (both normal) and ˆσ 2 (related to the chi square). This simplifies the task of establishing confidence intervals and testing (statistical) hypotheses.

EXAMPLE I: show that LSE are unbiased.

SOLUTION

Required to show that the estimators $\widehat{\beta}\ \ and\ \hat{\alpha}$ are unbiased for the parameters $\beta\ and\ \alpha$ in the model $y = \alpha + \beta x_i + \varepsilon_{i.}$ that is ; $E\left(\hat{\beta}\right) = \beta\ and\ E(\hat{\alpha}) = \alpha.$

2.4 **MEAN AND VARIANCE OF THE ESTIMATORS**

The mean for $\hat{\alpha} = E(\hat{\alpha}) = \alpha$ and the variance $v(\hat{\alpha}) = v\left(\bar{y} - \hat{\beta}\bar{x}\right) this\ is\ from\ \hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$. Therefore; $v(\hat{\alpha}) = v(\bar{y}) + \bar{x}^2 v\left(\hat{\beta}\right)$.

The mean for $\hat{\beta} = E\left(\hat{\beta}\right) = \beta$ and the variance v$(\hat{\beta}) = \frac{\delta^2}{\sum(x_i - \bar{x})^2}$ where $\delta^2 = \frac{\sum e_i^2}{n-p} = \sum \frac{(y_i - \hat{y}_i)^2}{n-p}$.

2.5 INTERVAL ESTIMATION AND HYPOTHESIS TESTING

2.51 Interval Estimation

Instead of relying on the point estimate alone, we may construct an interval around the point estimator, say within two or three standard errors on either side of the point estimator, such that this interval has, say, 95 percent probability of including the true parameter value. This is roughly the idea behind **interval estimation.**

Given a model of the form $y = \beta_1 + \beta_2 x_i + \varepsilon_{i,}a$ssume that we want to find out how "close," say, ˆ β2 is to β2. For this purpose we try to find out two positive numbers δ and α, the latter lying between 0 and 1, such that the probability that the **random interval** ($\hat{\beta}_2 - \delta$, $\hat{\beta}_2 + \delta$) contains the true β_2 is $1 - \alpha$. Symbolically, $\Pr(\hat{\beta}_2 - \delta \le \beta_2 \le \hat{\beta}_2 + \delta) = 1 - \alpha$. Such an interval, if it exists, is known as a **confidence interval;** $1 - \alpha$ is known as the **confidence coefficient;** and α ($0 < \alpha < 1$) is known as the **level of significance.** The endpoints of the confidence interval are known as the **confidence limits** (also known as *critical* values), $\hat{\beta}_2 - \delta$ being the **lower confidence** *limit* and $\hat{\beta}_2 + \delta$ the **upper confidence** *limit*.

An **interval estimator,** in contrast to a point estimator, is an interval constructed in such a manner that it has a specified probability $1 - \alpha$ of including within its limits the true value of the parameter.

It is very important to know the following aspects of interval estimation:
1. Since β_2, although an unknown, is assumed to be some fixed number, either it lies in the interval or it does not. For the method described, the probability of constructing an interval that contains β_2 is $1 - \alpha$.
2. The interval is a **random interval;** that is, it will vary from one sample to the next because it is based on $\hat{\beta}_2$, which is random.
3. Since the confidence interval is random, the probability statements attached to it should be understood in the long-run sense, that is, repeated sampling.

4. The interval is random so long as $\hat{\beta}_2$ is not known. But once we have a specific sample and once we obtain a specific numerical value of $\hat{\beta}_2$, the interval is no longer random; it is fixed. In this case, we cannot say that the probability is $1 - \alpha$ that a given *fixed* interval includes the true β_2. In this situation, β_2 is either in the fixed interval or outside it. Therefore, the probability is either 1 or 0.

b) Confidence Intervals for Regression Coefficients β_1 and β_2

Confidence Interval for β_2

We can use the normal distribution to make probabilistic statements about β_2 provided the true population variance σ^2 is known. If σ^2 is known, an important property of a normally distributed variable with mean μ and variance σ^2 is that the area under the normal curve between $\mu \pm \sigma$ is about 68 percent, that between the limits $\mu \pm 2\sigma$ is about 95 percent, and that between $\mu \pm 3\sigma$ is about 99.7 percent. But σ^2 is rarely known, and in practice it is determined by the unbiased estimator $\hat{\sigma}^2$. If we replace σ by $\hat{\sigma}$, the t statistic is $t = \frac{\hat{\beta}_2 - \beta_2}{se(\hat{\beta}_2)}$. where se($\hat{\beta}_2$) now refers to the estimated standard error. It can be shown that the t variable thus defined follows the t distribution with $n - 2$ df.

We can use the t distribution to establish a confidence interval for $\beta 2$ as follows: Pr ($-t\alpha/2 \leq t \leq t\alpha/2$) = 1 – α. where the t value in the middle of this double inequality is the *t statistic value computed* and where $t\alpha/2$ is the value of the t variable obtained from the t distribution for $\alpha/2$ level of significance and $n - 2$ df; it is often called the **critical** t value at $\alpha/2$ level of significance.

The width of the confidence interval is proportional to the standard error of the estimator. That is, the larger the standard error, the larger is the width of the confidence interval. Put differently, the larger the standard error of the estimator, the greater is the uncertainty of estimating the true value of the unknown parameter. Thus, the standard error of an estimator is often described as a measure of the **precision** of the estimator (i.e., how precisely the estimator measures the true population value).

$\hat{\beta}1 \pm t\alpha/2$ se ($\hat{\beta}1$)

Pr [$\hat{\beta}1 - t\alpha/2$ se ($\hat{\beta}1$) ≤ $\beta 1$ ≤ $\hat{\beta}1 + t\alpha/2$ se ($\hat{\beta}1$)] = $1 - \alpha$

$\hat{\beta}2 \pm t\alpha/2$ se ($\hat{\beta}2$)

Pr [$\beta 2 - t\alpha/2$ se ($\hat{\beta}2$) ≤ $\beta 2$ ≤ $\hat{\beta}2 + t\alpha/2$ se ($\hat{\beta}2$)] = $1 - \alpha$ with n-2 df.

EXAMPLE: CONSTRUCT A 95% CONFIDENCE INTERVAL FOR THE PARAMETER ESTIMATES USING THE DATA BELOW.

X	77	50	71	72	81	94	96	99	67
y	82	66	78	34	47	85	99	99	98

c) Confidence Interval for $\sigma 2$

Under the normality assumption, the variable $\chi 2 = (n-2)\frac{\hat{\delta}^2}{\delta^2}$ follows the $\chi 2$ distribution with $n - 2$ df. Therefore, we can use the $\chi 2$ distribution to establish a confidence interval for $\sigma 2$. Pr ($\chi 2\ 1{-}\alpha/2 \leq \chi 2 \leq \chi 2\alpha/2$)= $1 - \gamma$ where $\chi 21{-}\alpha/2$ and $\chi 2\ \alpha/2$ are two values of $\chi 2$ (the **critical** $\chi 2$ values) obtained from the chisquare table for $n - 2$ df in such a manner that they cut off $100(\gamma/2)$ percent tail areas of the $\chi 2$ distribution which gives the $100(1 - \gamma)$% confidence interval for $\sigma 2$.

Therefore confidence interval is $P\left[\frac{(n-2)\hat{\delta}^2}{x^2_{\frac{\gamma}{2}}} \leq \delta^2 \leq \frac{(n-2)\hat{\delta}^2}{x^2_{1-\frac{\gamma}{2}}}\right] = 1-\gamma.$

Hypothesis Testing: General Comments

The theory of hypothesis testing is concerned with developing rules or procedures for deciding whether to reject or not reject the null hypothesis. There are two *mutually complementary* approaches for devising such rules, namely, **confidence interval** and **test of significance.** Both these approaches predicate that the variable (statistic or estimator) under consideration has some probability distribution and that hypothesis testing involves making statements or assertions about the value(s) of the parameter(s) of such distribution.

For example, we know that with the normality assumption ˆ $\beta 2$ is normally distributed with mean equal to $\beta 2$ and variance. If we hypothesize that $\beta 2 = 1$, we are making an assertion about one of the parameters of the normal distribution, namely, the mean.

2.52 HYPOTHESIS TESTING

a) the confidence-interval approach

Two-Sided or Two-Tail Test

To illustrate the confidence interval approach, from regression results given as follows; the slope coefficient is 0.7240. Suppose we postulate that

$H0$: $\beta 2 = 0.5$

$H1$: $\beta 2 < 0.5$. That is, the true slope coefficient is 0.5 under the null hypothesis but less than 0.5 under the alternative hypothesis. The null hypothesis is a simple hypothesis, statistical hypothesis is called a **simple hypothesis** if it specifies the precise value(s) of the parameter(s) of a probability density function; otherwise, it is called a **composite hypothesis** because it does not have a specific value.

Two-sided hypothesis e.g H_1: $\theta \neq \theta_0$, such a two-sided alternative hypothesis reflects the fact that we do not have a strong a priori or theoretical expectation about the direction in which the alternative hypothesis should move fr
om the null hypothesis.

Thus, the confidence interval provides a set of plausible null hypotheses. Therefore, if $\beta 2$ under $H0$ falls within the $100(1 - \alpha)\%$ confidence interval, we do not reject the null hypothesis; if it lies outside the interval, we may reject it.

In statistics, when we reject the null hypothesis, we say that our finding is **statistically significant.** On the other hand, when we do not reject the null hypothesis, we say that our finding is **not statistically significant.**

b) The Test-of-Significance Approach

Testing the Significance of Regression Coefficients: The *t* Test

An *alternative but complementary approach* to the confidence-interval method of testing statistical hypotheses is the **test-of-significance approach.**

A test of significance is a procedure by which sample results are used to verify the truth or falsity of a null hypothesis. The key idea behind tests of significance is that of a **test statistic** (estimator) and the sampling distribution of such a statistic under the null hypothesis. The decision to accept or reject *H*0 is made on the basis of the value of the test statistic obtained from the data at hand.

If the value of true $\beta 2$ is specified under the null hypothesis, the *t* value can readily be computed from the available sample, and therefore it can serve as a test statistic. And since this test statistic follows the *t* distribution, confidence-interval statements such as the following can be made:
Pr($-t\alpha/2 \leq$ ˆ ($\beta 2 - \beta{*}2$)/se (ˆ $\beta 2$) $\leq t\alpha/2$) = 1 − α where $\beta{*}2$ is the value of $\beta 2$ under *H*0 and where $-t\alpha/2$ and $t\alpha/2$ are the values of *t* (the **critical** *t* values) obtained from the *t* table for ($\alpha/2$) level of significance and $n - 2$ df
In the confidence-interval procedure we try to establish a range or an interval that has a certain probability of including the true but unknown $\beta 2$, whereas in the test-of-significance approach we hypothesize some value for $\beta 2$ and try to see whether the computed ˆ $\beta 2$ lies within reasonable (confidence) limits around the hypothesized value.

To test the hypothesis that there is no relationship between the variables x and Y using the model $y = \alpha + \beta x$, the null hypothesis is stated as ; H_0: β=0 [no relationship between x and y]. if no prior information about the values of the regression parameters is available , the alternative hypothesis is stated as; H_A: β≠0.

The test statistic is given as $t = \frac{\beta}{s_\beta}$ at n-2 degrees of freedom. For a two tailed test (β≠0), the acceptance region is $-t_{\frac{\propto}{2},n-2} \leq \frac{\beta}{s_\beta} \leq t_{\frac{\propto}{2},n-2}$.
The best test is achieved if we take the alternative hypothesis as β<0, where the rejection region is $-t_{\propto,n-2} \leq \frac{\beta}{s_\beta}$.
If there is prior knowledge about the values of the parameters,
$H_0: \widehat{\beta} = \beta_0$ *and the test statistic is* $t = \frac{\widehat{\beta}-\beta_0}{s_\beta} \sim t_{n-2}$.

Alternatively, the F- test can be used to test for a relationship between the variables. The acceptance region for the hypothesis is $\frac{SSR}{(\frac{SSE}{N-2})} \leq F_{(1,N-2)}$.

c)The *t* Test of Significance: Decision Rules

Type of *H*0: The *H*1: The Alternative Decision Rule:
Reject *H*0 If
Two-tail: $\beta 2 = \beta 2^*$ Vs $\beta 2 \neq \beta 2^*$ reject H_0 if $|t| > t\alpha/2$,df
Right-tail : $\beta 2 \leq \beta 2^*$ Vs $\beta 2 > \beta 2^*$ reject if $t > t\alpha$,df
Left-tail: $\beta 2 \geq \beta 2^*$ Vs $\beta 2 < \beta 2^*$ *t reject if* $t < -t\alpha$,df

Notes: β^*2 is the hypothesized numerical value of $\beta 2$,$|t|$ means the absolute value of *t*, $t\alpha$ or $t\alpha/2$ means the critical *t* value at the α or $\alpha/2$ level of significance.
df: degrees of freedom, $(n - 2)$ for the two-variable model, $(n - 3)$ for the three-variable model, and so on.
The same procedure holds to test hypotheses about $\beta 1$.

d) **TESTING FOR SIGNIFICANCE OF REGRESSION USING THE F-TEST.**

The hypothesis is stated as H_o: model not significance versus H_A: model is significant at a given level of significance. The critical region is given as $F_c \geq F_{\propto,[k-1,N-k]}$ (reject H_o). where k are the parameters estimated and n is the sample size.

Using Analysis of variance (ANOVA), total variation is split into the explained variation and the unexplained variation; SST=SSR+SSE.

Using the ANOVA table for regression, the significance of regression can be determined using the f-test.

Source of variation	Sum of squares	Degrees of freedom	Mean sum of squares	f-computed
Regression	SSR	K-1	SSR/K-1=MSR	
Error	SSE	N-K	SSE/N-K=MSE	F_C=MSR/MSE
Total	SST	N-1		

Compare the computed f-statistic with the tabulated statistic at a level of significance.

Alternatively, for model significance the hypothesis can be stated as follows; H_o: $\beta_1 = \beta_2 = \cdots = \beta_k = 0\ versus\ H_A: not\ all\ \beta' s equal\ to\ zero.$ the f-statistic is obtained using $F_c = \frac{\frac{R^2}{(k-1)}}{\frac{(1-R^2)}{(N-K)}}$, where R^2 is the coefficient of determination.

The critical region is $f_c \geq f_{\propto,[k-1,N-K]}$

EXAMPLE II: refer to example I above. Test for significance of regression using the f-test at 5% level of significance.

- H_0: not significant vs H_A: significant
- C.r: $f \geq f_{\propto,[1,n-k]}$, $f_{0.05,[1.5]=6.61}$

Anova table

S.o.v	Degrees of freedom	Sum of squares	Mean sum of squares	f-computed
Regression	1	1383.929	1383.929	1383.929/87.5=15.816
Error	5	437.50	87.5	

Total	6	1821.429		

- Decision: reject H_0. It is significant.

EXAMPLE

Given the data below for minimum bank deposits in thousands of shillings and number of new accounts opened.

Branch	Minimum deposit (x)	New accounts (y)
A	125	160
B	100	112
C	200	124
D	75	28
E	150	152
F	175	156
G	75	42
H	175	124
I	125	150
J	200	104
K	100	136

i. Estimate regression model of the form $\hat{y}=\beta_0 + \beta_1 x_i$.

$$\beta_1 = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} = \frac{(11 * 186{,}200) - (1500 * 1288)}{(11 * 226250) - 1500^2} = \frac{116200}{238750} = 0.487.$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x} = 117.091 - (0.487 * 136.364) = 50.682$$

Equation: $y = 50.682 + 0.487x_i$.

ii. Variance for the estimator β_1.

$$v(\beta_1) = \frac{\delta^2}{\sum(x-\bar{x})^2}$$

$$where\ \delta^2\ can\ be\ estimated\ using\ s_{yx}^2 = \frac{\sum y^2 - \beta_0 \sum y - \beta_1 \sum xy}{n-2}$$

$$= \frac{170{,}696 - (50.682 * 1288) - (0.487 * 186200)}{11-2}$$

$$= 1637.576$$

$$therefore; v(\beta_1) = \frac{1637.576}{21704.546} = 0.0755.$$

iii. Test the hypothesis that $\beta_1 = 0\ against\ \beta_1 > 0$ at 0.05 level of significance.

- H_o: $\beta_1 = 0\ vs\ H_A: \beta_1 > 0$
- L.o.s α=0.05
- C.r: $t_c > t_{\propto,n-2},\ t_{0.05,9} = 1.833$
- $t_c = \frac{\beta_1}{s.e(\beta_1)} = \frac{0.487}{\sqrt{0.0755}} = 1.7724$
- Decision: fail to reject H_0.

e) Testing the Significance of *σ*2: The *χ*2 Test

As another illustration of the test-of-significance methodology, consider the following variable:

*χ*2 = (*n* − 2)ˆ*σ*2/*σ*2 which follows the *χ*2 distribution with *n* − 2 df. For example, ˆ*σ*2 = 0.8937 and df = 11. If we postulate that *H*0: *σ*2 = 0.6 versus *H*1: *σ*2 _= 0.6, it can be found that under *H*0, *χ*2 = 16.3845. If we assume *α* = 5%, the critical *χ*2 values are 3.81575 and 21.9200. Since the computed *χ*2 lies between these limits, the data support the null hypothesis and we do not reject it. This test procedure is called the **chi-square test of significance.**

2.6 BEST LINEAR UNBIASED ESTIMATION METHOD (BLUE)

To derive the BLUE of the parameters $\alpha\ and\ \beta$ requires the estimators to be;

i. A linear combination of sample observations. That is , $\tilde{\beta} = \sum c_i y_i$ where c =(c_1,c_2,…,c_n) are constants to be determined such that ,

ii. $\tilde{\beta}$is unbiased

iii. $\tilde{\beta}$ has minimum variance.

For unbiasedness $E(\tilde{\beta}) = E(\sum c_i y_i) = \beta$.

Proof:

- $E(\tilde{\beta}) = E(\sum c_i y_i) = \sum c_i E(y_i)$ but $y_i = \alpha + \beta x_i + \varepsilon_i$
- $substituing\ for\ y$
- $E(\sum c_i y_i) = \sum c_i E(\alpha + \beta x_i + \varepsilon_i) = \sum c_i(\alpha + \beta x_i) = \alpha \sum c_i + \beta \sum c_i x_i$
- $for\ \ \tilde{\beta}$ to be unbiased; $\sum c_i = 0\ and\ \sum c_i x_i = 1$.
- $when\ the$ above conditions are satisfied then; $E(\tilde{\beta}) = E(\sum c_i y_i) = \beta\ and\ therefore\ unbiased$.

The variance for $\tilde{\beta} = v(\sum c_i y_i) = \frac{\delta^2}{\sum(x_i - \bar{x})^2}$.

EXERCISE: DETERMINE THE FORMULA FOR THE ESTIMATOR $\tilde{\beta}$.

CHAPTER THREE

3.0 ESTIMATION USING THE MAXIMUM LIKELIHOOD ESTIMATION METHOD (MLE)

A method of point estimation with some stronger theoretical properties than the method of OLS is the method of **maximum likelihood (ML). since from** the regression model, the error terms *ui* are assumed to be normally distributed, the ML and OLS estimators of the regression coefficients, the β's, are identical, and this is true of simple as well as multiple regressions. The ML estimator of $\sigma 2$ is $\Sigma \hat{} u^2_i /n$. This estimator is biased, whereas the OLS estimator of $\sigma 2 = \Sigma \hat{} u_i^2/n\text{-}2$, is unbiased. But comparing these two estimators of $\sigma 2$, we see that as the sample size *n* gets larger the two estimators of $\sigma 2$ tend to be equal. Thus, asymptotically (i.e., as *n* increases indefinitely), the ML estimator of $\sigma 2$ is also unbiased.

(*a*) Compared to ML, the OLS is easy to apply; (*b*) the ML and OLS estimators of $\beta 1$ and $\beta 2$ are identical (which is true of multiple regressions too); and (*c*) even in moderately large samples the OLS and ML estimators of $\sigma 2$ do not differ vastly.

3.1 Maximum Likelihood Estimation of Two-Variable Regression Model

Assume that in the two-variable model $Yi = \beta 1 + \beta 2Xi + ui$,the *Yi* are normally and independently distributed with mean = $\beta 1 + \beta 2Xi$ and variance = $\sigma 2$. As a result, the joint probability density function of $Y1, Y2, \ldots, Yn$, given the preceding mean and variance, can be written as $f(Y1, Y2, \ldots, Yn \mid \beta 1 + \beta 2Xi, \sigma 2)$.

But in view of the independence of the *Y*'s, this joint probability density function can be written as a product of *n* individual density functions as
$f(Y1, Y2, \ldots, Yn \mid \beta 1 + \beta 2Xi, \sigma 2) = f(Y1 \mid \beta 1 + \beta 2Xi, \sigma 2)\, f(Y2 \mid \beta 1 + \beta 2Xi, \sigma 2) \cdot \cdot \cdot f(Yn \mid \beta 1 + \beta 2Xi, \sigma 2)$ **(1)**
where ; $f(Yi) = 1/\sigma\sqrt{2}\pi \exp^{-(Yi - \beta 1 - \beta 2Xi)2/2\sigma 2}$ **(2)** which is the density function of a normally distributed variable with the given mean and variance.

Substituting Equation (2) for each *Yi* into Equation (1) gives
$f(Yi, Y2, \ldots, Yn \mid \beta 1 + \beta 2Xi, \sigma 2) = 1/\sigma^n(\sqrt{2}\pi)^n \exp^{-\Sigma}(Yi - \beta 1 - \beta 2Xi)^2/2\sigma^2$ **(3)**
If $Y1, Y2, \ldots, Yn$ are known or given, but $\beta 1$, $\beta 2$, and $\sigma 2$ are not known, the function in Equation (3) is called a **likelihood function,** denoted by $LF(\beta 1, \beta 2, \sigma 2)$, and written as $LF(\beta 1, \beta 2, \sigma 2) = 1/\sigma^n(\sqrt{2}\pi)^n \exp^{-\Sigma(Yi - \beta 1 - \beta 2Xi)2/2\sigma 2}$ **(4).**

The **method of maximum likelihood,** as the name indicates, consists of estimating the unknown parameters in such a manner that the probability of observing the given *Y*'s is as high (or maximum) as possible. Therefore, we have to find the maximum of the function in Equation (4). For differentiation it is easier to express Equation (4) in the log term as follows.2

$\ln LF = -n \ln \sigma - n2\ln(2\pi) - \Sigma(Yi - \beta1 - \beta2Xi)^2 /2\sigma2 = -n2\ln \sigma2 - n2\ln(2\pi) - \Sigma(Yi - \beta1 - \beta2Xi)2/2\sigma2$ **(5)**

If $\beta1$, $\beta2$, and $\sigma2$ are known but the *Yi* are not known, Eq. (4) represents the joint probability density function—the probability of jointly observing the *Yi*.

Differentiating Equation (5) partially with respect to $\beta1$, $\beta2$, and $\sigma2$, we obtain

$\partial \ln LF/\partial\beta1 = -1/\sigma2(Yi - \beta1 - \beta2Xi)(-1)$ **(6)**

$\partial \ln LF/\partial\beta2 = -1/\sigma2(Yi - \beta1 - \beta2Xi)(-Xi)$ **(7)**

$\partial \ln LF/\partial\sigma2 = -n2\sigma2 + 1/2\sigma4\hat{}(Yi - \beta1 - \beta2Xi)2$ **(8)**

Setting these equations equal to zero (the first-order condition for optimization) and letting $\tilde{\beta}1$, $\tilde{\beta}2$, and $\tilde{\sigma}2$ denote the ML estimators, we obtain

$1/\tilde{\sigma}2(Yi - \tilde{\beta}1 - \tilde{\beta}2Xi) = 0$ **(9)**

$1/\tilde{\sigma}2(Yi - \tilde{\beta}1 - \tilde{\beta}2Xi)Xi = 0$ **(10)**

$-n2\tilde{\sigma}2 + 1/2\tilde{\sigma}4\_(Yi - \tilde{\beta}1 - \tilde{\beta}2Xi)2 = 0$ **(11)**

After simplifying, Eqs. (9) and (10) yield

$\Sigma Yi = n\tilde{\beta}1 + \tilde{\beta}2\Sigma Xi$ **(12)** $\Sigma Yi\, Xi = \tilde{\beta}1\Sigma Xi + \tilde{\beta}2\Sigma X2i$ **(13)**

which are precisely the *normal equations* of the least-squares theory . Therefore, the ML estimators, the $\tilde{\beta}$'s, are the same as the OLS estimators, the $\hat{\beta}$'s.This equality is not accidental. Examining the likelihood (5), we see that the last term enters with a negative sign. Therefore, maximizing Equation (5) amounts to minimizing this term, which is precisely the least-squares approach.

Substituting the ML (= OLS) estimators into Equation (11) and simplifying, we obtain the ML estimator of $\tilde{\sigma}2$ as; $\tilde{\sigma}2 = 1/n\Sigma(Yi - \tilde{\beta}1 - \tilde{\beta}2Xi)2 = 1/n\Sigma\hat{u}_i^2$ (14). From Equation (14) it is obvious that the ML estimator $\tilde{\sigma}2$ differs from the OLS estim ator $\hat{\sigma}^2 = [1/(n-2)]\Sigma\hat{u}_i^2$, which was shown to be an unbiased estimator of $\sigma2$. Thus, the ML estimator of $\sigma2$ is biased. The magnitude of this bias can be easily determined as follows.

We use ˜ (tilde) for ML estimators and ˆ (cap or hat) for OLS estimators.

Taking the mathematical expectation of Equation (14) on both sides, we obtain

$E(\tilde{\sigma}2) = 1/nE\Sigma\hat{u}_I 2 = \Sigma n - 2/n\Sigma\sigma 2 = \sigma 2 - 2n\sigma 2$ (15) which shows that $\tilde{\sigma}2$ is biased downward (i.e., it underestimates the true $\sigma 2$) in small samples. But notice that as *n*, the sample size, increases indefinitely, the second term in Equation (15), the bias factor, tends to be zero. Therefore, *asymptotically* (i.e., in a very large sample), $\tilde{\sigma}2$ is *unbiased* too, that is, lim $E(\tilde{\sigma}2) = \sigma 2$ as $n\rightarrow\infty$. It can further be proved that $\tilde{\sigma}2$ is also a **consistent** estimator4; that is, as *n* increases indefinitely, $\tilde{\sigma}2$ converges to its true value $\sigma 2$.

CHAPTER FOUR

3.0 REGRESSION THROUGH THE ORIGIN

There are occasions when the two-variable Population Regression Function (PRF) assumes the following form:

$Y_i = \beta_2 X_i + u$

In this model the intercept term is absent or zero, hence the name regression through the origin.

Given a Sample Regression Function (SRF) namely,

$Yi = \hat{\beta}_2 X_i + \hat{u}_i$ …………………………………….*(a)*

Now applying the OLS method we obtain the following formulas for $\hat{\beta}2$ and its variance .

$\hat{\beta}_2 = \Sigma XiYi/\Sigma X_i^2$, $v(\widehat{\beta_2}) = \frac{\delta^2}{\sum x_i^2}$ and $\widehat{\delta^2} = \frac{\sum \hat{u}_i^2}{n-1}$

when the intercept term is included in the model: $Yi = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i$

The formulas for the estimators are;

- $\beta_2 = \Sigma xi\ yi / \Sigma x_i^2$
- var ($\hat{\beta}_2) = \sigma^2/\Sigma x_i^2$
- $\hat{\sigma}^2 = \Sigma\hat{u}_i^2 / n - 2$, where u_i^2 is the error term determined using $\sum Y - \hat{Y})^2$.

Where x and y are deviations from the mean.

Although the interceptless or zero intercept model may be appropriate on occasions, there are some features of this model that need to be noted. First, $\Sigma\hat{u}i$, which is always zero for the model with the intercept term (the conventional model), need not be zero when that term is absent. In short, $\Sigma\hat{u}i$ need not be zero for the regression through the origin.

Second, r^2, the coefficient of determination, which is always nonnegative for the conventional model, can on occasions turn out to be *negative* for the interceptless model! This anomalous result arises because r^2 explicitly assumes that the intercept is included in the model. Therefore, the conventionally computed

r^2 may not be appropriate for regression-through-the-origin models. But one can compute what is known as the **raw r^2** for such models, which is defined as raw $r^2 = (\Sigma XiYi)^2 / \Sigma X_i^2 \, \Sigma Y_i^2$
Note: These are raw (i.e., not mean-corrected) sums of squares and cross products.
Although this raw r^2 satisfies the relation $0 < r^2 < 1$, it is not directly comparable to the conventional r^2 value. For this reason some people do not report the r^2 value for zero intercept regression models.

Because of these special features of this model, one needs to exercise great caution in using the zero intercept regression model. *Unless there is very strong a priori expectation*, one would be well advised to stick to the conventional, intercept-present model. This has a dual advantage. First, if the intercept term is included in the model but it turns out to be statistically insignificant (i.e., statistically equal to zero), for all practical purposes we have a regression through the origin. Second, and more important, if in fact there is an intercept in the model but we insist on fitting a regression through the origin, we would be committing a **specification error,** thus violating Assumption 9 of the classical linear regression model.

CHAPTER FIVE

5.0 Multiple Regression Analysis:

This is the analysis involving more than one explanatory variable. Given a regression model as $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots . + \beta_k x_k + u_i$, where β_0 is the intercept term. As usual, it gives the mean or average effect on Y of all the variables excluded from the model, although its mechanical interpretation is ;the Average of Y when $X1$ and $X2$ etc are set equal to zero. The coefficients $\beta 1$, $\beta 2$ etc are called the **partial regression coefficients.**
Assuming a model with two explanatory variables x_1 and x_2 and the model $= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u_i$, the following classical assumptions are fulfilled if x1 and $X2$ are nonstochastic;
1. Linear regression model, or *linear in the parameters.*
2. Fixed X values or X values independent of the error term. Here, this means we require zero covariance between ui and each X variables.
cov (ui , $X1i$) and cov (ui , $X2i$) = 0
3. Zero mean value of disturbance ui . $E(ui \,|X1i, X2i) = 0$ for each i.
4. Homoscedasticity or constant variance of ui .that is; var (ui) = σ^2

5. No autocorrelation, or serial correlation, between the disturbances. that is; cov (*ui* , *uj*) = 0 for $i \neq j$.
6. The number of observations *n* must be greater than the number of parameters to be estimated, which is 3 in our current case.
7. There must be variation in the values of the *X* variables.
8. No exact collinearity between the *X* variables. No **exact linear relationship** between *X*1 and *X2*.
9. There is no *specification bias*. The model is correctly specified.

4.1 The Meaning of Partial Regression Coefficients

As mentioned earlier, the regression coefficients $\beta 1$ and $\beta 2$ are known as **partial regression** or **partial slope coefficients.** The meaning of partial regression coefficient is as follows: $\beta 1$ measures the *change* in the mean value of *Y*, *E*(*Y*), per unit change in *X*1, holding the value of *X2* constant. Put differently, it gives the "direct" or the "net" effect of a unit change in *X1* on the mean value of *Y*, net of any effect that *X*2 may have on mean *Y*.

Likewise, $\beta 2$ measures the change in the mean value of *Y* per unit change in *X2*, holding the value of *X*1 constant. That is, it gives the "direct" or "net" effect of a unit change in *X*2 on the mean value of *Y*, net of any effect that *X*1 may have on mean *Y*.

4.2 ESTIMATION OF PARAMETERS

When we assume that the basic assumptions hold, the least squares estimators can be determined. For two explanatory variables x_1 and x_2, the least squares estimators can be determined using partial derivatives.

- For the model $y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$, determine sums of squares and differentiate with respect to the parameters.

- The coefficient of determination shows the percentage of Y explained by variations due to changes in x_1 and x_2 and is a measure of the goodness of fit. It is given by $R^2 = \frac{SSR}{SST} = \frac{\sum \hat{y}_i^2}{\sum y_i^2} = \frac{\sum(\hat{Y}_i - \overline{Y})^2}{\sum(Y_i - \bar{Y})^2}$ OR

$R^2 = \frac{\hat{\beta}_1 \sum y_i x_{1i} + \hat{\beta}_2 \sum y_i x_{2i}}{\sum y_i^2}$. This formula does not put into consideration the degrees of freedom left in introducing a new explanatory variable to the model. To overcome this, the adjusted coefficient of multiple determination is used. This is given by; $R^2 = 1 - \frac{(\frac{\sum e_i^2}{n-k})}{(\frac{\sum y_i^2}{n-1})}$.

- The variances of the estimators of the parameters are estimated using;

$v(\hat{\beta}_1) = \frac{\delta^2 \sum x_2^2}{[\Sigma x_1^2 \sum x_2^2 - (\sum x_1 x_2)^2]}$, $v(\hat{\beta}_2) = \frac{\delta^2 \sum x_1^2}{[\Sigma x_1^2 \sum x_2^2 - (\sum x_1 x_2)^2]}$ and

$v(\hat{\alpha}) = \delta^2[\frac{1}{n} + \frac{\bar{X}_1^2 \sum x_2^2 + \bar{X}_2^2 \sum x_1^2 - 2\bar{X}_1 \bar{X}_2 \sum x_1 x_2}{[\Sigma x_1^2 \sum x_2^2 - (\sum x_1 x_2)^2]}]$ Where $\delta^2 = \frac{\sum e_i^2}{n-k} = \frac{\sum(Y_i - \hat{Y}_i)^2}{n-k}$

4.3 TESTING FOR SIGNIFICANCE OF REGRESSION

Given the H_o: β_i =0 [not significant] and the H_A: $\beta_i \neq 0$ [significant] and a level of significance λ, the critical region is given as a two tailed test such that; $-t_{\frac{\lambda}{2},(n-k)} \leq t \leq t_{\frac{\lambda}{2}(n-k)}$ where k are the parameters estimated. For two explanatory variables k=2 and the computed t-statistic is given as $t = \hat{\beta}_i / se(\hat{\beta}_i)$.

EXAMPLE: for the data below estimate the regression equation, variances and test the significance of the parameters at 5%

Y	10	8	7	7	5	6	9	10	11	6
X_1	5	7	6	6	8	7	5	4	3	9
X_2	10	6	12	5	3	4	13	11	13	3

SOLUTION

- **Fit a model of the form $y = \alpha + \beta_1 x_1 + \beta_2 x_2$**
- **Obtain the variances**
- **Determine SST,SSE and SSR**
- **Test for significance of the model using the f-test**