ESTIMATION OF PARAMETERS USING BEST LINEAR UNBIASED AND LIKELIHOOD METHOD

a) BEST LINEAR UNBIASED ESTIMATION METHOD

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 $y_i = \alpha + \beta x_i + \varepsilon_i$
- $substituting \; for \; y$
- $E(\sum c_i y_i) = \sum c_i E(\alpha + \beta x_i + \varepsilon_i) = \sum c_i(\alpha + \beta x_i) = \alpha \sum c_i + \beta \sum c_i x_i$
- $for \; \tilde{\beta}$ 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 $\tilde{\beta}$.

b) 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 σ^2 is $\Sigma\hat{}u^2_i /n$. This estimator is biased, whereas the OLS estimator of $\sigma 2 = \Sigma\hat{}u_i^2/n-2$, is unbiased. But comparing these two estimators of σ^2, we see that as the sample size n gets larger the two estimators of σ^2 tend to be equal. Thus, asymptotically (i.e., as n increases indefinitely), the ML estimator of σ^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 σ^2 do not differ vastly.

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 = σ^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
$f(Y1, Y2, \ldots, Yn \mid \beta 1 + \beta 2Xi, \sigma 2) = f(Y1 \mid \beta 1 + \beta 2Xi, \sigma 2)\, f(Y2 \mid \beta 1 + \beta 2Xi, \sigma 2) \cdots f(Yn \mid \beta 1 + \beta 2Xi, \sigma 2)$ **(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
$\ln LF = -n \ln \sigma - n2\ln(2\pi) - \Sigma(Yi - \beta 1 - \beta 2Xi)^2/2\sigma 2 = -n2\ln \sigma 2 - n2\ln(2\pi) - \Sigma(Yi - \beta 1 - \beta 2Xi)2/2\sigma 2$ **(5)**

If $\beta 1$, $\beta 2$, and $\sigma 2$ 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 $\beta 1$, $\beta 2$, and $\sigma 2$, we obtain
$\partial \ln LF/\partial\beta 1 = -1/\sigma 2(Yi - \beta 1 - \beta 2Xi)(-1)$ **(6)**
$\partial \ln LF/\partial\beta 2 = -1/\sigma 2(Yi - \beta 1 - \beta 2Xi)(-Xi)$ **(7)**

∂ ln LF/$\partial\sigma 2 = -\,n2\sigma 2 + 1/2\sigma 4$^$(Yi - \beta 1 - \beta 2Xi\,)2$ **(8)**
Setting these equations equal to zero (the first-order condition for optimization) and letting ˜ $\beta 1$, ˜ $\beta 2$, and ˜σ 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 ˜ β's, are the same as the OLS estimators, the ˆ β'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 ˜σ 2 as;$\tilde{}\sigma 2 = 1/n\Sigma(Yi - \tilde{}\,\beta 1 - \tilde{}\,\beta 2Xi\,)2 = 1/\mathrm{n}\Sigma\hat{}u_{\mathrm{i}}^2$ (14).
From Equation (14) it is obvious that the ML estimator ˜σ 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 $\sigma 2$. Thus, the ML estimator of $\sigma 2$ 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
$E(\tilde{}\sigma\,2) = 1/nE\Sigma\tilde{}u_{\mathrm{I}}\,2 = \Sigma n - 2/n\Sigma\sigma 2 = \sigma 2 - 2n\sigma 2$ (15) which shows that ˜σ 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), ˜σ 2 is *unbiased* too, that is, lim $E(\tilde{}\sigma\,2) = \sigma 2$ as $n\to\infty$. It can further be proved that ˜σ 2 is also a **consistent** estimator4; that is, as n increases indefinitely, ˜σ 2 converges to its true value $\sigma 2$.

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.

Example: using the data below for minimum bank deposits in thousands of shillings and number of new accounts opened., estimate the regression equation assuming regression through the origin, establish the variance S^2_{yx} and the raw coefficient of determination.

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