International Journal of Engineering & Technology, 7 (4.10) (2018) 518-522

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

A Treatise on Ordinary Least Squares Estimation of Parameters of Linear Model

B. Mahaboob [1], B.Venkateswarlu[2]*, C. Narayana [3] , J. Ravi sankar [4] and P. Balasiddamuni [5]

[1]*Department of Mathematics, K.L.E.F (Deemed to be University), Vaddeshwaram, Vijayawada, Andhra Pradesh.*
[2,4]*Department of Mathematics, VIT University, Vellore, Tamilnadu.*
[3]*Department of Mathematics, Sri Harsha institute of P.G Studies, Nellore.*
[5]*Rtd. Professor, Department of Statistics, S.V. University, Tirupati, Andhra Pradesh.*
**Corresponding author E-mail: venkatesh.reddy@vit.ac.in*

Abstract

This research article primarily focuses on the estimation of parameters of a linear regression model by the method of ordinary least squares and depicts Gauss-Mark off theorem for linear estimation which is useful to find the BLUE of a linear parametric function of the classical linear regression model. A proof of generalized Gauss-Mark off theorem for linear estimation has been presented in this memoir. Ordinary Least Squares (OLS) regression is one of the major techniques applied to analyse data and forms the basics of many other techniques, e.g. ANOVA and generalized linear models [1]. The use of this method can be extended with the use of dummy variable coding to include grouped explanatory variables [2] and data transformation models [3]. OLS regression is particularly powerful as it relatively easy to check the model assumption such as linearity, constant, variance and the effect of outliers using simple graphical methods [4]. J.T. Kilmer et.al [5] applied OLS method to evolutionary and studies of algometry.

***Keywords**: BLUE, OLS estimation, mean vector, Covariance matrix, linear regression model.*

1. Introduction

This Regression analysis is a statistical method to establish the relationship between variables. Regression analysis has a wide number of applications in almost all fields of science, including Engineering, Physical and Chemical Sciences; Economics, Management, Social, Life and Biological Sciences. In fact, regression analysis may be the most frequently used statistical technique in practice. Suppose that there exists a linear relationship between a dependent variable Y and an independent variable X. In the scatter diagram, if the points cluster around a straight line then the mathematical form of the linear model may be specified as

$$Y_i = \beta_0 + \beta_1 X_i, \quad i = 1, 2, \ldots n. \tag{1.1}$$

Where β_0 is the intercept and β_1 is the slope.
Generally the data points in the scatter diagram do not fall exactly on a straight line, so equation (2.1.1) should be modified to account for this. Let the difference between the observed value of Y and the straight line $(\beta_0 + \beta_1 X)$ be an error ε. It is convenient to think of ε as a statistical error; that is, it is a random variable that accounts for the failure of the model to fit the data exactly. The error may be made up of the effects of other variables, measurement errors and so forth. Thus, a more plausible model may be specified as

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, i = 1, 2, \ldots, n. \tag{1.2}$$

Equation (1.2) is called a Linear Regression Model or Linear Statistical Model. Customarily X is called the independent variable and Y is called the dependent variable. However, this often causes confusion with the concept of statistical independence, so we refer to X as the Predictor or Regressor variable and Y as the Response variable. Since the equation (2.1.2) involves only one Regressor variable, it is called a 'Simple Linear Regression Model' or a 'Two-Variable Linear Regression Model'. A Three – variable Linear Regression Model may be written as

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i, i = 1, 2, \ldots, n \tag{1.3}$$

This linear regression model contains two regressor variables. The term linear is used because eq. (1.3) is a linear function of the unknown parameters β_0, β_1 and β_2.
In general, the response variable Y may be related to k regressor or predictor variables. The model

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki} + \varepsilon_i, i = 1, 2, \ldots n \tag{1.4}$$

is called a 'Multiple Linear Regression Model' with k independent variables. The parameters β_j, j=0, 1, 2.., k are known as regression coefficients. This model describes a hyperplane in the k – dimensional space of the independent variables X_j's. The parameter β_j represents the expected change in the dependent variable Y per unit change in X_j, when all of the remaining predicted variables

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

X_q's (q≠j) are held constant. Thus, the parameters β_j, j = 1, 2… k are often known as 'Partial Regression Coefficients.

2. Ordinary Least Squares Estimation of Parameters of Linear Model

An Consider the Classical Linear Regression model

$$Y_{nx1} = X_{nxk}\beta_{kx1} + \varepsilon_{nx1} \quad (2.1)$$

with usual assumptions such as

$$E(\varepsilon) = 0,\ E(\varepsilon\varepsilon') = \sigma^2 I_n \quad (2.2)$$

Write the residual sum of squares as

$$e'e = \left(Y - X\hat{\beta}\right)'\left(Y - X\hat{\beta}\right) \quad (2.3)$$

$$= Y'Y - \hat{\beta}'X'Y - Y'X\hat{\beta} + \hat{\beta}'X'X\hat{\beta}$$

$$\Rightarrow e'e = Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta} \qquad \left[\because Y'X\hat{\beta} = \hat{\beta}'X'Y\right]$$

Where $\hat{\beta}$ is the least squares estimator of β.

By the least squares estimation method, $\hat{\beta}$ minimizes the residual sum of squares $e'e$.

First order condition:

$$\frac{\partial}{\partial\hat{\beta}}(e'e) = O \Rightarrow -2X'Y + 2X'X\hat{\beta} = O$$

$$\Rightarrow X'X\hat{\beta} = X'Y \quad (2.4)$$

The system (2.7.4) contains 'n' simultaneous linear equations, which is called the 'System of Normal Equations'. Since, the system of normal equations is always consistent, these exists at least a non-zero solution of $\hat{\beta}$, which gives the ordinary least squares (OLS) estimator of β.

i.e. $\hat{\beta} = \left(X'X\right)^{-1}X'Y$ (2.5)

Further, consider the OLS residual vector

$$e = Y - X\hat{\beta} \quad (2.6)$$

$$= X\beta + \varepsilon - X\left(X'X\right)^{-1}X'(X\beta + \varepsilon)$$

$$= \varepsilon - X\left(X'X\right)^{-1}X'\varepsilon$$

$$= \left(I_n - X\left(X'X\right)^{-1}X'\right)\varepsilon$$

[I_n is a unit matrix of order n]

$$\Rightarrow e = M\varepsilon \quad (2.7)$$

where

$M = \left(I_n - X\left(X'X\right)^{-1}X'\right)$ is a symmetric idempotent matrix such that $M'M=M$,

$M'=M$ and $MX=O$.

Now, consider the OLS residual sum of squares

$$e'e = \left(M\varepsilon\right)'\left(M\varepsilon\right) = \varepsilon'M\varepsilon$$

$$\Rightarrow E\left(e'e\right) = E\left(\varepsilon'M\varepsilon\right) = E\left(trace\,\varepsilon'M\varepsilon\right)\ \left[\because \varepsilon'M\varepsilon \text{ is a scalar}\right]$$

$$= E\left(trace\,M\varepsilon\varepsilon'\right)$$

$$= \left(trace\,M\right)E\left(\varepsilon\varepsilon'\right) = \sigma^2 trace\,M \qquad \left[\because E\left(\varepsilon\varepsilon'\right) = \sigma^2 I_n\right]$$

$$= \sigma^2 \text{trace}\left(I_n - X\left(X'X\right)^{-1}X'\right)$$

$$= \sigma^2\left[\text{trace}\,I_n - \text{trace}\left(X'X\right)^{-1}\left(X'X\right)\right]$$

$$= \sigma^2\left[n - trace\,I_k\right]$$

$$\Rightarrow E\left(e'e\right) = \sigma^2\left(n-k\right)$$

$$or\ E\left(\frac{e'e}{n-k}\right) = \sigma^2 \quad (2.8)$$

Or

$E\left(S^2\right) = \sigma^2$, where $S^2 = \dfrac{e'e}{n-k}$ is an unbiased estimator of σ^2.

3. Gauss-Mark-Off Theorem for Linear Estimation

This theorem is useful to find the Best Linear Unbiased Estimator (BLUE) of a linear parametric function of the classical linear regression model.

Statement:

In the Gauss-Mark off linear model, $Y = X\beta + \varepsilon$ with usual assumptions; the BLUE of a linear parametric function $C'\beta$ is given by $C'\hat{\beta}$, where $\hat{\beta}$ is the ordinary least squares estimator of β. Here, C is a (kx1) vector of known coefficients.

Proof:

Consider the Gauss-Mark-off linear model

$$Y_{nx1} = X_{nxk}\beta_{kx1} + \varepsilon_{nx1} \quad (3.1)$$

Such that

$E\left(\varepsilon\right) = O, E\left(\varepsilon\varepsilon'\right) = \sigma^2 I_n$ and $\varepsilon \sim N\left(O, \sigma^2 I_n\right)$. Suppose that the linear parametric function $C'\beta$ is estimable. Then, there exists a linear function of observations vector $P'Y$ such that $E\left(P'Y\right) = C'\beta$, P is a vector of unknown coefficients.

$$\Rightarrow P'X\beta = C'\beta\ \left[\because E\left(Y\right) = X\beta\right]\ or\ X'P = C \quad (3.2)$$

One may have

$$Var\left(P'Y\right) = \left(P'P\right)Var\left(Y\right) = \sigma^2\left(P'P\right)$$

The BLUE of $C'\beta$ can be obtained by minimizing the $Var(P'Y)$ with respect to P subject to the restriction $X'P = C$. Write the constrained minimization function as

$$\phi = P'P - 2\lambda'(X'P - C) \tag{3.3}$$

Where λ is a (kx1) vector of unknown as Lagrangian multipliers First order condition:

$$\frac{\partial\phi}{\partial P} = 0 \Rightarrow 2P - 2X\lambda = 0$$ or

$$P = X\lambda \tag{3.4}$$

From (3.2) and (3.4), one may obtain

$$X'X\lambda = C \tag{3.5}$$

The BLUE of $C'\beta$ is given by

$$P'Y = \lambda'X'Y \tag{3.6}$$

It can be shown that $\lambda'X'Y = C'\hat{\beta}$. Here, $\hat{\beta} = (X'X)^{-1}X'Y$ is the OLS estimator of β.
From the ordinary least squares estimation method, one may write the system of normal equations as $X'X\hat{\beta} = X'Y$. One may obtain,

$$\lambda'X'Y = \lambda'X'X\hat{\beta} = C'\hat{\beta}\left[\because X'XC = C\right]$$

Hence, the BLUE of linear parametric function $C'\beta$ is given by $C'\hat{\beta}$, where $\hat{\beta} = (X'X)^{-1}X'Y$.

4. Mean Vector and Covariance Matrix of Blue

Consider the Gauss-Mark-off linear model

$$Y_{nx1} = X_{nxk}\beta_{kx1} + \varepsilon_{nx1} \tag{4.1}$$

such that

$E(Y) = X\beta$ and $Var(Y) = \sigma^2 I_n$. Suppose that a linear parametric function $C'\beta$ is estimable. Then, there exists a linear function of observation vector $\lambda'Y$ such that

$$E(\lambda'Y) = C'\beta \quad \text{or} \quad \lambda'X\beta = C'\beta$$
$$\text{or} \quad X'\lambda = C \tag{4.2}$$

By the condition for the existence of BLUE, one may have

$$X'XP = C \tag{4.3}$$

or

$$\left.\begin{aligned} (X'X)^{-1}C = P, &\quad \text{if} \quad \rho(X) = k \\ (X'X)^{g}C = P, &\quad \text{if} \quad \rho(X) < k \end{aligned}\right\} \qquad 4.4)$$

Further, one may write the BLUE for $C'\beta$ as

$$C'\hat{\beta} = \lambda'Y = P'X'Y \tag{4.5}$$

Consider, $Var(\lambda'Y) = (\lambda'\lambda)Var(Y) = \sigma^2(\lambda'\lambda) \quad [\because \lambda = XP]$

$$= \sigma^2(P'X'XP) \tag{4.6}$$

(i) $Var(\lambda'Y) = \sigma^2\left[C'(X'X)^{-1}(X'X)(X'X)^{-1}C\right] \quad \text{if} \ \rho(X) = k$ (4.7)

$$\Rightarrow Var(\lambda'Y) = \sigma^2\left[C'(X'X)^{-1}C\right] \quad \text{if} \ \rho(X) = k$$

(ii) $Var(\lambda'Y) = \sigma^2\left[C'(X'X)^{g}(X'X)(X'X)^{g}C\right] \quad \text{if} \ \rho(X) < k$

$$\Rightarrow Var(\lambda'Y) = \sigma^2\left[C'(X'X)^{g}C\right] \quad \text{if} \ \rho(X) < k \tag{4.8}$$

Thus, the mean vector and covariance matrix of BLUE $C'\hat{\beta}$ are given by

(i) $E(C'\hat{\beta}) = C'\beta$ (4.9)

and

$$\left.\begin{aligned} &\text{(ii)} \ Var(C'\hat{\beta}) = \sigma^2\left[C'(X'X)^{-1}C\right] \quad \text{if} \ \rho(X) = k \\ &\text{and} \ Var(C'\hat{\beta}) = \sigma^2\left[C'(X'X)^{g}C\right] \quad \text{if} \ \rho(X) < k \end{aligned}\right\} \tag{4.10}$$

Remarks:
By taking C as a (kx1) vector of one's one may obtain,

(i) $E(\hat{\beta}) = \beta$ (4.11)

And

$$\left.\begin{aligned} &\text{(ii)} \ Var(\hat{\beta}) = \sigma^2(X'X)^{-1} \quad \text{if} \ \rho(X) = k \\ &\text{and} \ Var(\hat{\beta}) = \sigma^2(X'X)^{g} \quad \text{if} \ \rho(X) < k \end{aligned}\right\} \tag{4.12}$$

5. Generalized Gauss-Mark-Off Theorem for Linear Estimation

One may obtain the Generalized Gauss-Mark-off linear model by violating the assumption $E(\varepsilon\varepsilon') = \sigma^2 I_n$, in the Gauss- Mark-off Linear model.
Consider the linear regression model,

$$Y_{n\times1} = X_{n\times k}\,\beta_{k\times1} + \varepsilon_{n\times1} \tag{5.1}$$

such that

$$\left.\begin{aligned} &E(\varepsilon) = O \quad \text{or} \quad E(Y) = X\beta \quad \text{and} \quad E(\varepsilon\varepsilon') = \sigma^2\Omega \\ &\text{or} \quad Var(Y) = \sigma^2\Omega \end{aligned}\right\} \tag{5.2}$$

Where σ^2 is known and Ω is a known positive definite symmetric matrix. The linear regression model (5.1) along with assumptions (5.2) is known as the Generalized Gauss- Mark-off linear model, which was first given by Aitken in year 1932.

Statement:
In the Generalized Gauss-Mark-off linear model $Y_{nx1} = X_{nxk}\beta_{kx1} + \varepsilon_{nx1}$ such that $E(\varepsilon) = O, \quad E(\varepsilon\varepsilon') = \sigma^2\Omega$; the BLUE for a linear parametric function $C'\beta$ is given by $C'\tilde{\beta}$, where $\tilde{\beta}$ is the unique Generalized Least Squares (GLS) estimator for β, which can be obtained by solving a system of Generalized Normal Equations $(X'\Omega^{-1}X)\tilde{\beta} = (X'\Omega^{-1}Y)$. Also an unbiased estimator of σ^2 is given by

$$\tilde{\sigma}^2 = \frac{e'\Omega^{-1}e}{n-r},$$

Where $e = Y - \hat{Y}$ is OLS residual vector and $\rho(X) = r$.
Here Ω is a known positive definite symmetric matrix.

Proof:
Consider the Generalized Gauss-Mark-off linear regression model,

$$Y_{nx1} = X_{nxk}\beta_{kx1} + \varepsilon_{nx1} \tag{5.3}$$

Such that $E(\varepsilon) = O \text{ and } E(\varepsilon\varepsilon') = \sigma^2\Omega$. Since, Ω is known positive definite symmetric matrix, these exists a non-singular matrix M such that

$$MM' = \Omega \text{ or } (M'\Omega^{-1}M) = I \tag{5.4}$$

Pre-multiplying on both of sides (2.10.3) by M^{-1} gives

$$M^{-1}Y = M^{-1}X\beta + M^{-1}\varepsilon \text{ or } Y^* = X^*\beta + \varepsilon^* \tag{5.5}$$

Where $Y^* = M^{-1}Y, \quad X^* = M^{-1}X \text{ and } \varepsilon^* = M^{-1}\varepsilon$

Consider (i) $E(\varepsilon^*) = E(M^{-1}\varepsilon) = M^{-1}E(\varepsilon) = O$ (5.6)

(ii) $E(\varepsilon^*\varepsilon^{*'}) = E[(M^{-1}\varepsilon)(M^{-1}\varepsilon)'] = \sigma^2(M^{-1}\Omega M'^{-1})$

$$= \sigma^2(M'\Omega^{-1}M)^{-1}$$

$$\Rightarrow E(\varepsilon^*\varepsilon^{*'}) = \sigma^2 I \tag{5.7}$$

Thus, the Generalized Gauss-Mark -off linear regression model reduces to an ordinary Gauss-Mark-off linear model given in (5.5), (5.6) and (5.7). Now, the unique GLS estimator of β can be obtained by solving the system of normal equations,

$$X^{*'}X^*\tilde{\beta} = X^{*'}Y^*$$

$$\Rightarrow (M^{-1}X)'(M^{-1}X)\tilde{\beta} = (M^{-1}X)'(M^{-1}Y)$$

$$\Rightarrow \left[X'(MM')^{-1}X\right]\tilde{\beta} = X'(MM')^{-1}Y$$

$$\Rightarrow (X'\Omega^{-1}X)\tilde{\beta} = X'\Omega^{-1}Y \tag{5.8}$$

One BLUE of $C'\beta$ is given by $C'\tilde{\beta}$.
Where $\tilde{\beta} = (X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}Y), \quad \text{if } \rho(X) = k$

or $\tilde{\beta} = (X'\Omega^{-1}X)^{g}(X'\Omega^{-1}Y), \quad \text{if } \rho(X) < k$

Also an unbiased estimator of error variance σ^2 is given by

$$\tilde{\sigma}^2 = \frac{e^{*'}e^*}{n-r}, \text{ where } e^* = Y^* - X^*\hat{\beta}$$

$$\text{or } \tilde{\sigma}^2 = \frac{(Y - X\hat{\beta})'(MM')^{-1}(Y - X\hat{\beta})}{n-r} = \frac{e'\Omega^{-1}e}{n-r}, \tag{5.9}$$

where $r = \rho(X)$

6. Properties of OLS Estimators

Following are some important properties of the OLS estimators $\hat{\beta}$ and $\hat{\sigma}^2$:

The OLS estimator $\hat{\beta}$ is the BLUE for β. The mean vector and the covariance matrix are respectively given by β and $\sigma^2(X'X)^{-1}$.

The OLS estimator $\hat{\beta}$ is the maximum likelihood estimator for β and hence, it is consistent.

The OLS estimator $\hat{\beta}$ follows multivariate normal distribution with mean vector β and the covariance matrix $\sigma^2(X'X)^{-1}$.

One OLS estimator $\hat{\sigma}^2 = \frac{e'e}{n-k}$ is an unbiased and consistent estimator of σ^2.

$\left[\frac{e'e}{\sigma^2}\right]$ or $\left[\frac{(n-k)\hat{\sigma}^2}{\sigma^2}\right]$ follows χ^2-distribution with (n-k) degrees of freedom.

The variance of $\hat{\sigma}^2$ is given by

$$Var(\hat{\sigma}^2) = \frac{2\sigma^4}{n-k}. \tag{6.1}$$

The OLS estimators $\hat{\beta}$ and $\hat{\sigma}^2$ are the efficient estimators of β and σ^2 respectively.

The OLS estimators $\hat{\beta}$ and $(n-k)\hat{\sigma}^2$ are two joint sufficient statistics of β and σ^2.

$\left(\frac{n-k}{n}\right)\hat{\sigma}^2$ is the maximum likelihood estimator for σ^2.

The Rao – Cramer lower bounds for the variances of $\hat{\beta}$ and $\hat{\sigma}^2$ are respectively given by $\sigma^2(X'X)^{-1}$ and $\frac{2\sigma^4}{n}$.

The OLS estimators $\hat{\beta}$ and $\hat{\sigma}^2$ are asymptotically efficient estimators of β and σ^2 respectively.

$$\sqrt{n}\left(\hat{\beta}-\beta\right)\overset{asy}{\sim} N\left[O,\sigma^2\left(\underset{n\to\infty}{Lim}\left(\frac{X'X}{n}\right)^{-1}\right)\right]$$

and

$$\sqrt{n}\left(\hat{\sigma}^2-\sigma^2\right)\overset{asy}{\sim} N\left[O,2\sigma^4\right].$$

7. Problems of Linear Model by Violating the Assumptions

Several problems can be arised by violating the crucial assumptions about the linear regression model such as:

1. Problems of biased and Inconsistent estimators for β will be arised if $E(\varepsilon)\neq 0$;
2. Problems of heteroscedasticity and autocorrelation will be arised if $E(\varepsilon\varepsilon')\neq\sigma^2 I_n$;
3. Problem of multicollinearity will be arised if $\rho(X)\neq k$;
4. Problem of stochastic regressors will be arised if the data matrix X is a stochastic matrix;
5. Problem of errors in variables will be arised if there are errors in the independent variables;
6. Problems of non-normal errors and non-parametric linear regression analysis will be arised if ε does not follow multivariate normal distribution; and Problem of Random Coefficient Regression (RCR) models will arise if the regression coefficients governed by some probability distribution.

8. Conclusion

In The above discourse presented OLS estimation method of a classical linear regression model and the BLUE of a linear parametric function of the v model. In addition to these mean vector and covariance matrix of BLUE have been derived in this article. Properties of OLS estimators have been proposed and the problems which can be arised by violating the crucial assumptions about the linear regression model have been exhaustively discussed.

References

[1] Rutherford A, "Introducing ANOVA and ANCOVA: a GLM approach,London: Sage publications (2001).

[2] Hutcheson, G.D. Ordinary Least-Squares Regression. In: Moutinho, L. and Hutcheson, G.D., the SAGE Dictionary of Quantitative Management Research, SAGE Publications, Thousand Oaks, (2011), 224-228.

[3] Fox. J, An Rand S- Plus companion to Applied Regression London: Sage publications, (2002).

[4] Hutcheson G.D and Sofroniou .N, "The multivariate social Scientist London, Sage publications, (1999).

[5] J.T. Kilmer, R.L. Rodriguez, "Ordinary least squares regression is indicated for studies of allesnetry", Journal of Evolutionary, Biology, Vol. (31), issue 1, (2016), Pp: 4-12

[6] Agrestic A, "An introduction to categorical Data Analysis", John Wiley and Sons. Inc, (1996).

[7] Hutcheson G.D and Moutinho .L,"Statistical modelling for management", Sage publications, (2008).

[8] Koteswara .R.K, "Testing for the Independence of regression Disturbances, Econometric a, 38, (1970), and Pp: 97-117

[9] Ryan .T. P. "Modern Regression methods Chichester: John Wiley and sons, (1997).

[10] Marno Verbeek, "using linear regression to establish empirical relationships", IZA, World of Labour, (2017).

[11] Gordon K. Smyth. "Nonlinear regression, Encyclopaedia of Environ metrics Vol. (3), (2002) Pp: 1405-1461.

[12] Gurleen K. Popli. "A note on the instrumental variable estimators in the nonlinear models", Journal of Quantitative Economics Vol.16. no.2, (2002), Pp: 31-36.

[13] E. Grafarent and J. Awange. "Application of linear and nonlinear models", Springer Geophysics, (2012).