A Note
On Derivation of the Least Squares Estimator

Eric Iksoon Im*

Working Paper Series No. 96-11
June 1, 1996

Abstract

Derivation of Least Squares (LS) estimators of intercept and slope in bivariate regression model has been solely calculus-based. Herein, for the first time in the chronicle of regression, we provide a derivation of the LS estimators in very basic algebra within the grasp of the intended readers of many introductory books in statistics and econometrics. Also, we provide a similar derivation of the LS estimator of a parameter vector for the multiple regression model which takes only a few steps of basic matrix operation.

Key words: Derivation of LS Estimator; Bivariate Regression; Calculus-based; Self-contained; Multiple Regression; Matrix Calculus; Orthogonality.

---

* mailing address: Department of Economics, University of Hawai'i at Manoa
Room 542, Porteus Hall, 2424 Maile Way, Honolulu, Hawai'i 96822
phone: (808)956-2323; fax: (808)956-4347
e-mail address: eim@uhunix.uhcc.hawaii.edu

1. Introduction

Virtually with no exception, introductory textbooks in statistics and econometrics treat the LS estimators of the intercept and slope for bivariate linear regression model as an integral part of their contents. However, though it is stated in their prefaces either explicitly or implicitly that calculus is not required for the intended readers, the derivation of the LS estimators, if provided, is usually on the basis of calculus, or at least so referred to elsewhere. In the case in which the derivation is based alternatively on two normal equations (which are in essence the very first order conditions for minimization in calculus), as is often the case, no explanation is given on how the normal equations relate to the LS estimation.

Such treatments of the derivation substantially reduces the level of self-containment of the books, leaving many intended readers to a varying degree puzzled or even intimidated. To help remove this long standing snag from the introductory books, we offer a derivation solely in rudimentary algebra well within the grasp of intended readers of the introductory texts.

2. Derivation

For the standard bivariate linear regression model $Y_i = \alpha + \beta X_i + \varepsilon_i,\ i = 1,\ldots,n,$ the estimated error term can be written as $e_i = Y_i - a - bX_i$ where a and b can be any estimators of α and β, respectively. Define for the economies of notation $q = \overline{Y} - a - b\overline{X}$ where $\overline{Y}$ and $\overline{X}$ denote the respective sample means of Y_i and X_i. Further defining y_i and x_i as deviations of Y_i and X_i from their respective means, we can work out the lower bound for $ESS(a,b)$, the error sum of squares as a function of a and b, with several steps of basic algebraic manipulation:

$$ESS(a,b) = \Sigma \left[Y_i - a - bX_i\right]^2$$

$$= \Sigma \left[(\overline{Y}+y_i) - a - b(\overline{X}+x_i)\right]^2$$

$$= \Sigma \left[q - (bx_i - y_i)\right]^2$$

$$= nq^2 + \Sigma(bx_i - y_i)^2$$

$$= nq^2 + \left(b^2 - 2b\Sigma x_i y_i/\Sigma x_i^2 + \Sigma y_i^2/\Sigma x_i^2\right)\Sigma x_i^2$$

$$= nq^2 + \left(b^2 - 2b\Sigma x_i y_i/\Sigma x_i^2 + \left[\Sigma x_i y_i/\Sigma x_i^2\right]^2 + \Sigma y_i^2/\Sigma x_i^2 - \left[\Sigma x_i y_i/\Sigma x_i^2\right]^2\right)\Sigma x_i^2$$

$$= nq^2 + \left(b - \Sigma x_i y_i/\Sigma x_i^2\right)^2 \Sigma x_i^2 + \Sigma y_i^2\left\{1 - \left[\Sigma x_i y_i/\sqrt{\Sigma x_i^2 \Sigma y_i^2}\right]^2\right\}$$

$$\geq \Sigma y_i^2\left\{1 - \left[\Sigma x_i y_i/\sqrt{\Sigma x_i^2 \Sigma y_i^2}\right]^2\right\}, \quad (1)$$

noting that the first two terms in the last equality are both non-negative definite for all a and b. The last expression in (1) defines the lower bound for $ESS(a,b)$ which is reached at $a = a^*$ and $b = b^*$ such that $q = \overline{Y} - a^* - b^*\overline{X} = 0$ and $b^* - \Sigma x_i y_i/\Sigma x_i^2 = 0$. Solving for a^* and b^* these two equations, we come up with the LS estimators of intercept and slope: $a^* = \overline{Y} - b^*\overline{X}$ and $b^* = \Sigma x_i y_i/\Sigma x_i^2$.

3. Analogy for Multiple Regression

There are two existing approaches to deriving the LS estimator of parameter vector β in the case of the standard multiple regression model $y = X\beta + u$ where matrix dimensions are $y(n\times 1)$, $X(n\times k)$, $\beta(k\times 1)$, and $u(n\times 1)$. One is based on the first order condition for minimizing $ESS(b) = [y - Xb]'[y - Xb]$, b denoting any estimator of β, and the other based on the orthogonality condition in analytical geometry (for an example of both, see Greene (1993)).

Analogous to (1), however, we can establish the lower bound for $ESS(b)$ for the multiple regression model without going beyond only a few elementary matrix operations. Defining $P = X(X'X)^{-1}X'$ and $M = I_n - P$ to facilitate a short analogy, we arrive rather quickly at the lower bound for $ESS(b)$:

$$ESS(b) = [y - Xb]'[y - Xb]$$

$$= [(Py - Xb) + My]'[(Py - Xb) + My]$$

$$= (Py - Xb)'(Py - Xb) + y'My$$

$$\geq y'My \tag{2}$$

noting that $M' = M = M^2$, $X'M = 0$, and $P'M = 0$, and also that the first term in the last equality is nonnegative definite for all b. (Note that 0 in this section denotes a null matrix of an appropriate order.)

From (2) it is immediately clear that $ESS(b)$ reaches its lower bound at $b = b^*$ such that $Py - Xb^* = 0$. Solving the matrix equation for b^*, $b^* = (X'X)^{-1}X'Py = (X'X)^{-1}X'y$ reflecting that $X'P = X'$.

4. Final Remarks

In this note we have derived the LS estimators of intercept and slope for the bivariate regression model, keeping the level of algebra within the grasp of intended readers of the introductory textbooks in statistics and econometrics. The derivation herein provided is self-contained and expository: self-contained in the sense that we arrive at the final results without going beyond the perimeters of basic algebra appropriate for the intended readers of the elementary books in statistics and econometrics; expository in the sense that the lower bound for the error sum of squares explicitly plays a pivotal role in deriving the LS estimators.

A similar derivation of the LS estimator for multiple regression herein provided requires only the basic knowledge of matrix operation. Given the basic knowledge, one may well find this approach more comfortable relative to existing ones which involve matrix calculus and the orthogonality condition in analytic geometry.

5. References

Greene, W. H. (1993), *Econometric Analysis*, 2nd ed., N.Y.: MacMillan