GEOMETRICAL INTERPRETATION OF STEP-WISE ESTIMATION OF PARAMETERS IN LINEAR REGRESSION ANALYSIS

BU-244-M

A. Hedayat

June, 1967

ABSTRACT

This note shows the geometrical approach of step-wise estimation of parameters in linear regression analysis given by Freund, Vail and Clunies-Ross in March, 1961 and by Goldberger and Jochems in March, 1961 and by Goldberger in December, 1961 in Journal of American Statistical Association.

Biometrics Unit, Cornell University, Ithaca, N.Y.

NOTE ON THE GEOMETRICAL INTERPRETATION OF STEP-WISE ESTIMATION OF PARAMETERS IN LINEAR REGRESSION ANALYSIS

Introduction

Least squares estimates of β_1 and β_2 of the partitioned linear model

$$\underset{N\times 1}{Y} = \underset{N\times P_1}{X_1}\ \underset{P_1\times 1}{\beta_1} + \underset{N\times P_2}{X_2}\ \underset{P_2\times 1}{\beta_2} + \underset{N\times 1}{\epsilon_1} \qquad - - - - (0)$$

will be

$$\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix} \begin{bmatrix} X_1'Y \\ X_2'Y \end{bmatrix}$$

$$= \begin{bmatrix} (X_1'X_1)^{-1}[I + X_1'X_2MX_2'X_1(X_1'X_1)^{-1}] & -(X_1'X_1)^{-1}X_1'X_2M^{-1} \\ -M^{-1}X_2'X_1(X_1'X_2)^{-1} & M^{-1} \end{bmatrix} \begin{bmatrix} X_1'Y \\ X_2'Y \end{bmatrix}$$

where $M = X_2'X_2 - X_2'X_1(X_1'X_2)^{-1}X_1'X_2$

$$= \begin{bmatrix} (X_1'X_1)^{-1}X_1'Y - (X_1'X_1)^{-1}X_1'X_2M^{-1}X_2'WY \\ M^{-1}X_2'WY \end{bmatrix} \qquad - - - - (1)$$

where $W = [I - X_1(X_1'X_1)^{-1}X_1]$.

Inserting $M^{-1}X_2'WY = \hat{\beta}_2$ in equation for $\hat{\beta}_1$ we obtain

$$\hat{\beta}_1 = (X_1'X_1)^{-1}X_1'Y - (X_1'X_1)^{-1}X_1'X_2\hat{\beta}_2 \qquad - - - - (2)$$

If one attempts* to estimate β_1 and β_2 in two stages, then as Freund, Vail and Clunies-Ross [1], and Goldberger and Jochems [2] have shown, estimated β_2.

---

* For application of this method see Goldberger and Jochems [2].

and related test will be biased. Later Goldberger [3] has unified the discussion for the estimated β_1 and β_2. Before giving the simple geometrical interpretation of their result, we will summarize their estimation results as follow:

One Stage Estimation of β_1 and Two-Stage Estimation of β_2.

First From

$$Y = X_1\beta_1 + \epsilon_2 \qquad \text{- - - - (3)}$$

estimate β_1, call it b_1

$$b_1 = (X_1'X_1)^{-1}X_1'Y \qquad \text{- - - - (4)}$$

Comparison of (4) and (2) results

$$b_1 = \hat{\beta}_1 + (X_1'X_1)^{-1}X_1'X_2\hat{\beta}_2 \qquad \text{- - - - (5)}$$

Compute the estimated residuals $\hat{\epsilon}_2$

$$\begin{aligned} \hat{\epsilon}_2 &= Y - X_1b_1 \\ &= [I - X_1(X_1'X_1)^{-1}X_1']Y \\ &= WY \quad . \qquad \text{- - - - (6)} \end{aligned}$$

Second From

$$\hat{\epsilon}_2 = X_2\beta_2 + \epsilon_3 \qquad \text{- - - - (7)}$$

estimate β_2, call it b_2

$$\begin{aligned} b_2 &= (X_2'X_2)^{-1}X_2'\hat{\epsilon}_2 \\ &= (X_2'X_2)^{-1}X_2'WY \quad . \qquad \text{- - - - (8)} \end{aligned}$$

Inserting $I = MM^{-1}$ in relation (8) we obtain

$$\begin{aligned} b_2 &= (X_2'X_2)^{-1}X_2'MM^{-1}WY \\ &= (X_2'X_2)^{-1}M(M^{-1}X_2'WY) \\ &= (X_2'X_2)^{-1}M\hat{\beta}_2 \qquad \text{by relation (1)} \end{aligned}$$

$$= [I - (X_2'X_2)^{-1}X_2'X_1(X_1'X_1)^{-1}X_1'X_2]\hat{\beta}_2 \quad \text{by definition of } M$$

$$= \hat{\beta}_2 - [(X_2'X_2)^{-1}X_2'X_1(X_1'X_1)^{-1}X_1'X_2]\hat{\beta}_2. \quad ---- (9)$$

It is easy to see from the relation (5) and (9) that

$$Eb_1 \neq E\hat{\beta}_1 = \beta_1$$

$$Eb_2 \neq E\hat{\beta}_2 = \beta_2$$

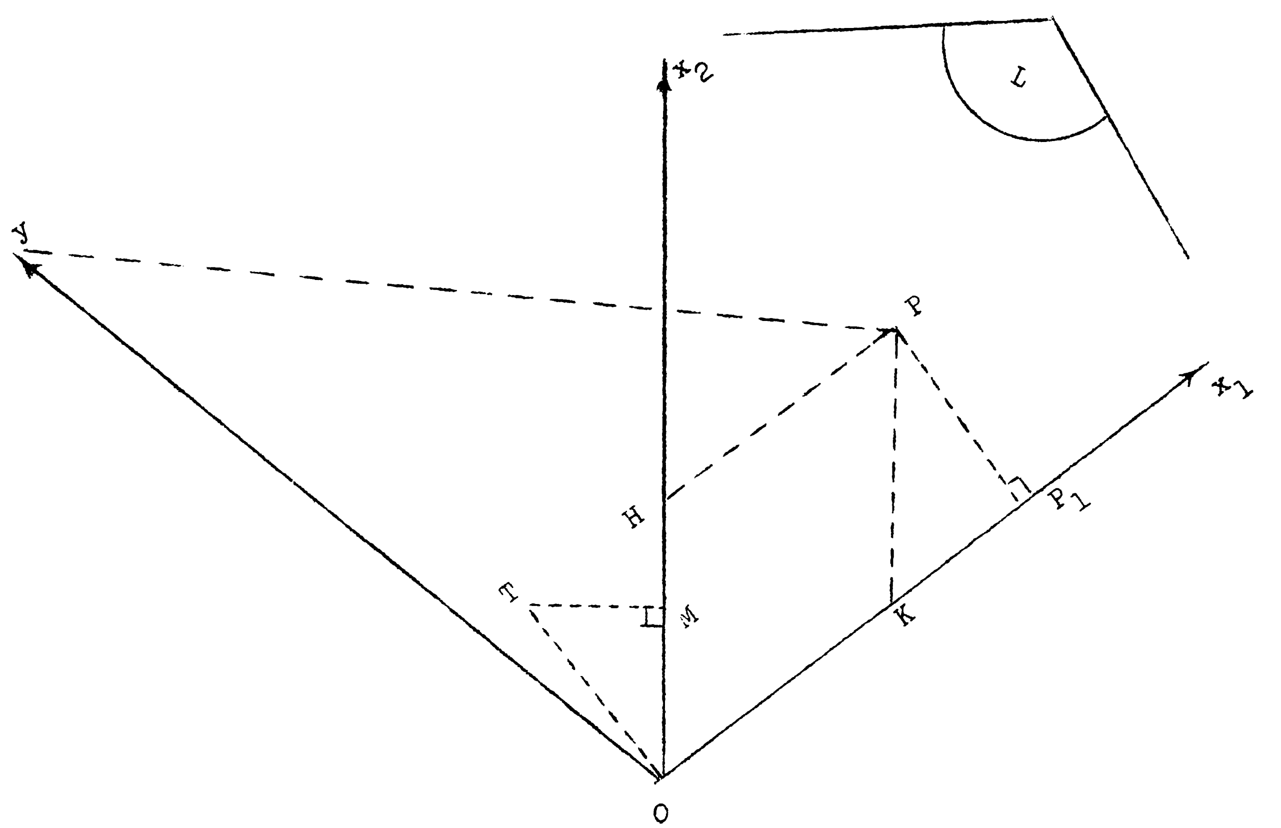
unless $(X_1'X_2) = 0$ or $\beta_2 = 0$.

Geometrical Approach

Consider the case where X_1 and X_2 consist of a single column (the only reason for this choice is so we can provide a diagram). And for convenience we measure Y, X_1, X_2 as deviation from their means. Therefore our model will be

$$y = x_1\beta_1 + x_2\beta_2 + \epsilon.$$

Let L be the linear subspace generated by the column vectors x_1 and x_2. Regression of y on x_1 and x_2 is the orthogonal projection of y on L denoted by P. Coordinates of P gives $\hat{y}$.



Then $\hat{\beta}_1 = \frac{\| OK \|}{\| x_1 \|}$, $\hat{\beta}_2 = \frac{\| OH \|}{\| x_2 \|}$ - - - - (10)

where OK and OH are the components of OP on x_1 and x_2 respectively.

Regression of y on x_1 is the orthogonal projection of y on x_1 or equivalently is the orthogonal projection of $\hat{y}$ on x_1 denoted by P_1. If we complete the parallelogram which has OP as diagonal and OP_1 and PP_1 as sides, we obtain the parallelogram $OTPP_1$. Coordinates of T gives the estimated residuals $\hat{\epsilon}_1$ of regression of y on x_1.

Now the regression of $\hat{\epsilon}_1$ on x_2 is represented by the projection OM of OT on x_2. b_1 and b_2 in relations (5) and (9) for our simple model will be

$$b_1 = \frac{\| OP_1 \|}{\| x_1 \|} , \quad b_2 = \frac{\| OM \|}{\| x_2 \|} \qquad \text{- - - - (11)}$$

Comparing (10) and (11) and by referring to the picture we see that

$$b_1 \neq \beta_1 , \quad b_2 \neq \beta_2 \quad \text{unless} \quad x_1 \perp x_2$$

Remark: There is an unbiased two-stage least squares estimation of parameters. For algebraic derivation see Freund [1] and for the geometric interpretation see pp 112-115 of Applied Regression Analysis by Draper and Smith.

Literature Cited

[1] Freund, R. J., Vail, R. W. and Clunies-Ross, C. W. Residual Analysis. Journal of American Stat. Association 56, 1961.

[2] Goldberger, A. S. and Jochems, D. B. Note on Stepwise Least Squares. Journal of American Stat. Association 56, 1961.

[3] Goldberger, A. S. Stepwise Least Squares: Residual Analysis and Specification Error. Journal of American Stat. Association 56, 1961.

[4] Draper, N. R. and Smith, H. Applied Regression Analysis. John Wiley and Son, Inc., New York, 1966.