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

ABSTRACT


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 $\beta_1$ and $\beta_2$ of the partitioned linear model

$$Y = X_1 \beta_1 + X_2 \beta_2 + \epsilon_1$$

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_2X_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'Y \\
M^{-1}X_2'Y
\end{bmatrix}

$$

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

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

If one attempts to estimate $\beta_1$ and $\beta_2$ in two stages, then as Freund, Vail and Clunies-Ross [1], and Goldberger and Jochems [2] have shown, estimated $\beta_2$.

* For application of this method see Goldberger and Jochems [2].
and related test will be biased. Later Goldberger [3] has unified the dis-
cussion for the estimated $\beta_1$ and $\beta_2$. Before giving the simple geometrical interpretation of their result, we will summarize their estimation results as follow:

One Stage Estimation of $\beta_1$ and Two-Stage Estimation of $\beta_2$.

First From

$$Y = X_1 \beta_1 + \epsilon_2$$  \hspace{1cm} (3)

estimate $\beta_1$, call it $b_1$

$$b_1 = (X_1'X_1)^{-1}X_1'Y$$  \hspace{1cm} (4)

Comparison of (4) and (2) results

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

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

$$\hat{\epsilon}_2 = Y - X_1 b_1$$

$$= [I - X_1(X_1'X_1)^{-1}X_1']Y$$

$$= WY$$  \hspace{1cm} (6)

Second From

$$\hat{\epsilon}_2 = X_2 \beta_2 + \epsilon_3$$  \hspace{1cm} (7)

estimate $\beta_2$, call it $b_2$

$$b_2 = (X_2'X_2)^{-1}X_2'\hat{\epsilon}_2$$

$$= (X_2'X_2)^{-1}X_2'WY$$  \hspace{1cm} (8)

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

$$b_2 = (X_2'X_2)^{-1}X_2'MM^{-1}WY$$

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

$$= (X_2'X_2)^{-1}MM^{-1}\hat{\beta}_2$$  \hspace{1cm} by relation (1)
\[
-3-
= [I - (x_2'x_2)^{-1}x_2'x_1(x_1'x_1)^{-1}x_1'x_2] \hat{\beta}_2 \quad \text{by definition of } \hat{\beta}_2
\]
\[
= \hat{\beta}_2 - [(x_2'x_2)^{-1}x_2'x_1(x_1'x_1)^{-1}x_1'x_2] \beta_2.
\]

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

\[
E_1 = E_{\hat{\beta}_1} = \beta_1
\]
\[
E_2 = E_{\hat{\beta}_2} = \beta_2
\]

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 \( \hat{\beta}_1 = \frac{\| OK \|}{\| x_1 \|} , \hat{\beta}_2 = \frac{\| OH \|}{\| x_2 \|} \) \hspace{1cm} - - - - (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 \( \tilde{y} \) 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 \( \hat{\varepsilon}_1 \) of regression of \( y \) on \( x_1 \).

Now the regression of \( \hat{\varepsilon}_1 \) 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

\[ b_1 = \frac{\| OP_1 \|}{\| x_1 \|} , \quad b_2 = \frac{\| OM \|}{\| x_2 \|} \] \hspace{1cm} - - - - (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**


