# The Unique, Minimum Variance, Linear, Unbiased Stepwise Estimator

*Sung-Yeung Kwack**

Arthur Goldberger and others have discussed the degree of the bias of the stepwise estimators of the parameters in a multiple linear regression model $\mathbf{Y}=\mathbf{X}_1\beta_1+\mathbf{X}_2\beta_2+\varepsilon$. The present paper attempts to develop a stepwise method for estimating the parameters, by which the unique, minimum variance, linear, unbiased estimators can be obtained, even if the orthogonality of $\mathbf{X}_1$ to $\mathbf{X}_2$ is not assumed.

1. Consider the multiple linear regression model: $\mathbf{Y}=\mathbf{X}\beta+\varepsilon$, as partitioned as follow:

$$\mathbf{Y}=\mathbf{X}_1\beta_1+\mathbf{X}_2\beta_2+\varepsilon, \tag{1}$$

where $\mathbf{Y}$ is the $N\times 1$ vector of the dependent variables,

$\mathbf{X}_1$ is the $N\times K_1$ matrix of observations of the first $K_1$ independent variables,

$\mathbf{X}_2$ is the $N\times K_2$ matrix of observations of the second $K_2$ independent variables,

$\beta_1$ is the $K_1\times 1$ vector of regression coefficients of the $\mathbf{X}_1$,

$\beta_2$ is the $K_2\times 1$ vector of regression coefficients of the $\mathbf{X}_2$,

$\varepsilon$ is the $N\times 1$ vector of random disturbances.

It is assumed that:

(i) the vector of random disturbances satisfies:

$$E(\varepsilon)=\mathbf{0} \tag{2}$$

$$V(\varepsilon)=\sigma^2\mathbf{I} \tag{3}$$

where $\mathbf{I}$ is the identity matrix of order $N$, and $\sigma^2$ is a constant parameter representing the variance of random disturbances.

(ii) $\mathbf{X}_1$ and $\mathbf{X}_2$ are of rank $K_1$ and $K_2$, respectively.

2. Let $\mathbf{M}$ be a matrix such that $\mathbf{M}$ is orthogonal to the sub-space $V_{K_1}$ of Euclidean $N$-space spanned by the $K_1$ vectors of $\mathbf{X}_1$: for instance, $\mathbf{M}$ may be defined as

---

* The author is a graduate student in Department of Economics at the University of California, Berkeley. He is indebted to Professor Hyung-Yoon Byun for encouragement on the research. Views and remaining errors are solely the author's.

$$\mathbf{M}=[\mathbf{I}-\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1']. \tag{4}$$

By pre-multiplying(1) by **M**,

$$\mathbf{MY}=\mathbf{MX}_2\beta_2+\mathbf{M}\varepsilon. \tag{5}$$

(5) could satisfy the hypotheses of the Gauss-Markov Theorem on estimation as generalized by Aitkin[1], for the rank of $[\mathbf{MX}_2]$ is $K_2$, and the variance-covariance matrix of $[\mathbf{M}\varepsilon]$ is semi-positive definite(1):

$$V(\mathbf{M}\varepsilon)=\sigma_2\ \mathbf{M}. \tag{6}$$

Thus, the estimator $\beta_2^S$ of the vector of regression coefficients of $\mathbf{X}_1$, $\beta_2$, is obtained by projecting $[\mathbf{MY}]$ on the sub-space spanned by $[\mathbf{M}\ \mathbf{X}_2]$:

$$\beta_2^S=(\mathbf{X}_1'\mathbf{MX}_2)^{-1}\mathbf{X}_2'\mathbf{MY}, \tag{7}$$

which has variance-covarince matrix:

$$V(\beta_2^S)=\sigma^2(\mathbf{X}_2'\mathbf{MX}_2)^{-1}. \tag{8}$$

Again, by projecting the sum of the residuals $[\mathbf{Y}-\mathbf{X}_2\beta_2^S]$ on the sub-space $V_{K1}$, the estimator $\beta_1^S$ of $\beta_1$ is:

$$\beta_1^S=(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'[\mathbf{I}-\mathbf{X}_2(\mathbf{X}_2'\mathbf{MX}_2)^{-1}\mathbf{X}_2'\mathbf{M}]\mathbf{Y}, \tag{9}$$

which has variance-covariance matrix:

$$V(\beta_1^S)=\sigma^2(\mathbf{X}_1'\mathbf{X}_1)^{-1}[\mathbf{I}+\mathbf{X}_1'\mathbf{X}_2(\mathbf{X}_2'\mathbf{MX}_2)^{-1}\mathbf{X}_2'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}]. \tag{10}$$

The estimator$(\beta_1^S\ \ \beta_2^S)$ is identical to the ordinary least squares estimator.(2) Thus, it is sufficient to demonstrate that the estimator derived by the method described above is the unique minimum variance linear unbiased estimator of the parameters $(\beta_1\ \beta_2)$, even in the case in which $\mathbf{X}_1'\mathbf{X}_2\neq 0$. In order to differentiate the estimator from the ordinary stepwise estimator(3), the estimator may be called as the unique, minimum variance, linear, unbiased stepwise estimator.

The relationships between the unique, minimum variance, linear, unbiased and the ordinary stepwise estimators are expressed as follows(4):

$$\mathbf{b}_2=[\mathbf{I}-(\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2]\beta_2^S \tag{11}$$

$$\mathbf{b}_1=\beta_1^S+(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2\beta_2^S, \tag{12}$$

where **b** is the ordinary stepwise estimator or $\beta_j$ $(j=1,\ 2)$.

---

(1) $\mathbf{M}=\mathbf{M}'=\mathbf{MM}'>0$ for any matrix **A** such that $\mathbf{A}\neq\mathbf{X}_1$. Gauss-Markov Theorem can be applied to this specification so long as $(\mathbf{X}_2'\mathbf{MX}_2)^{-1}$ exists, as it is the case, even if **M** is semi-positive definite.

(2) (7) and (9) are identical to (2.6) in Arthur Goldberger, "Stepwise Least Squares: Residual Analysis and Specification Error," *Journal of the American Stastical Association*, 56(December,1961), p. 999. Also, the same formulars are to be found from(17) in Dale W. Jorgenson, "Minimum Variance, Linear, Unbiased Seasonal Adjustment of Economic Time Series," *Journal of the American Statistical Association*, 59(September, 1964), p. 697.

(3) A. Goldberger, *op. cit.*, p. 998, i.e., $\bar{\mathbf{b}}_1=(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{Y}$ and $\bar{\mathbf{b}}_2=(\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{Y}$, where $\bar{\mathbf{b}}_j$ is the ordinary stepwise estimator of $\beta_j(j=1,\ 2)$.

(4) (11) and (12) are identical to (2. 7) of Goldberger's, *op. cit.*, p. 999.

(11) implies that[5]:

$$\beta_2^S \neq \mathbf{b}_2 \tag{13}$$

$$\beta_1^S \neq \mathbf{b}_2 \tag{14}$$

(12) implies that the direction of the bias of the coefficients $\mathbf{b}_1$ may upward or downward; it depends on not only $(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2$, but also the sign of $\beta_1^S$, unless $\mathbf{X}_2'\mathbf{X}_1=0$ or $E(\beta_1^S)=0$[6].

3. An unbiased estimator of the parameter $\sigma^2$, say $\sigma^{*2}$, is given by:

$$\sigma^{*2}=\frac{\mathbf{e}'\mathbf{e}}{N-K_1-K_2} \tag{15}$$

where $\mathbf{e}=\mathbf{Y}-\mathbf{X}_1\beta_1^S-\mathbf{X}_2\beta_2^S$.

Under the assumption that the random disturbances are independently and normally distributed with expection of zero and a constant varaince $\sigma^2$:

$$\varepsilon \text{ is } N(0,\ \sigma^2), \tag{16}$$

normal sampling theory can be applied for statical inference on problems such as testing a linear hypothesis on regression coefficients.

## REFERENCES:


[1] Aitkin, A.C., "On Least Squares and Linear Combination of Observations," *Proceedings of the Royal Society of Edinburgh*, 55(1934—5), 42—8.

[2] Goldberger, S.S., "Stepwise Least Squares: Residual Analysis and Specification Error," *Journal of the American Statical Association*, 56(December, 1961), 998—1000.

[3] Goldberger, A.S. and Jochems, "Note on Stepwise Least Squares," *Journal of the American Statistical Association*, 56(March, 1961), 105—110.

[4] Jorgenson, D.W., "Minimum Variance, Linear, Unbiased Seasonal Adjustment of Economic Time Series," *Journal of the American Statistical Association*, 56 (September, 1964), 681—724.

[5] Kendall, M. G., *A Course in the Geometry of N Dimensions*, (London: Charels Griffin, 1961).

[6] Schaffé, H., *The Analysis of Variance*(New York: John Wiley & Sons, 1959).


(5) Goldberger and Jochems derived the same results; see, "Note on Stepwise Least Squares," *Journal of the American Stistical Association*, 56(September, 1961), 105—6.

(6) In the special case where $K_2=1$, the quantity in the brackets on the right of (11) is, in fact, $[1-r_{21}^2]$ where $r_{21}^2$ is the coefficient of determination of $\mathbf{X}_2$ on the set $\mathbf{X}_1$. In that case, $1>1-r_{21}^2>0$ and (11) may indicate the direction as follows:

$$\beta_2^s \geq \mathbf{b}_2 \geq 0 \text{ for } \beta_2^s \geq 0$$

$$\beta_2^s < \mathbf{b}_2 < 0, \text{ otherwise.}$$