Journal of Econometrics 7 (1978) 115–117. © North-Holland Publishing Company

ON OBTAINING THE RIGHT SIGN OF A COEFFICIENT ESTIMATE BY OMITTING A VARIABLE FROM THE REGRESSION

Ignazio VISCO

Banca d'Italia, Servizio Studi, Rome, Italy

Received December 1976, final version received May 1977

It has been recently proved by Leamer (1975) that when a variable is dropped from a least-squares linear regression equation there can be no change in the sign of any coefficient estimate that has a larger t-ratio than the estimate of the coefficient of the dropped variable. This means that if the estimate of a coefficient comes out with the 'wrong' sign (relative to what the researcher expects, for example on the basis of economic theory), there is no way to obtain an estimate having the 'right' sign if we simply rerun the regression deleting an explanatory variable which had a coefficient estimate statistically less significant than that which came out originally with the 'wrong' sign. We will show in this note that the converse is not generally true, that is, deleting a variable which had a coefficient estimate more significant than that with the 'wrong' sign is not sufficient (even if necessary) to have this sign reversed. We will examine below what the necessary and sufficient conditions are, and we will consider the relevance and meaning of these conditions in two simple situations.

Consider the linear model

$$Y = \alpha+\beta_1 X_1+\dots+\beta_i X_i+\beta_j X_j+\dots+\beta_k X_k+u. \tag{1}$$

The least-squares estimate of the coefficient vector $\beta = (\beta_1, \dots, \beta_k)'$ can be written as $\hat{\beta} = VX'y$, where $X = (x_1, \dots, x_k)$, y and x_i $(i = 1, \dots, k)$ being the dependent and explanatory variables expressed as deviations from means, and $V = (X'X)^{-1}$ is the variance–covariance matrix of $\hat{\beta}$ (with ijth element V_{ij}) when the variance of the random process u is set equal to one (this assumption has no relevance for the results below).

Constraining β_i to zero, it can be proved that the least-squares estimate of the vector β will be[1]

$$\hat{\beta}^* = \hat{\beta}-V_i\hat{\beta}_i/V_{ii}, \tag{2}$$

where V_i is the ith column of V. Obviously (2) verifies $\hat{\beta}_i^* = 0$.

[1]See, for example, Goldberger (1964, p. 257) and Leamer (1975, p. 388).

Our objective is to obtain, dropping X_i from the regression, a reversal in the sign of the estimate of β_j, that is,

$$\hat{\beta}_j \hat{\beta}_j^* < 0. \tag{3}$$

But, using (2),

$$\hat{\beta}_j \hat{\beta}_j^* = \hat{\beta}_j(\hat{\beta}_j - V_{ij}\hat{\beta}_i/V_{ii}) = V_{jj}(t_j^2 - \rho_{ij} t_i t_j),$$

where t_i and t_j are the t-ratios for the coefficients $\hat{\beta}_i$ and $\hat{\beta}_j$ (i.e., $t_i = \hat{\beta}_i/V_{ii}^{\frac{1}{2}}$, $t_j = \hat{\beta}_j/V_{jj}^{\frac{1}{2}}$), and $\rho_{ij} = \mathrm{cov}(\hat{\beta}_i, \hat{\beta}_j) = V_{ij}/(V_{ii}V_{jj})^{\frac{1}{2}}$.

Since $V_{jj} > 0$, necessary and sufficient condition for (3) to hold is

$$t_j^2 < \rho_{ij} t_i t_j, \tag{4}$$

which can be divided into two parts:[2]

$$|t_i| > |t_j|/|\rho_{ij}|, \tag{4a}$$

$$\rho_{ij}\hat{\beta}_i \gtrless 0 \quad \text{as} \quad \hat{\beta}_j \gtrless 0. \tag{4b}$$

This proves the following:

Theorem. Necessary and sufficient conditions for the constrained least-squares estimate of β_j, with β_i set equal to zero, to have opposite sign than the least-squares estimate of β_j obtained in the absence of any restriction on β_i, are the following: (a) *the t-ratio for the* (*unconstrained*) *estimate of β_j must be in absolute value less than the t-ratio for the estimate of β_i by a factor smaller than the absolute value of the correlation coefficient between the two estimates* [*condition* (4a)]*; the signs of the estimate of β_i and of the correlation coefficient between the estimates of β_i and β_j must be equal or opposite, depending on the estimate of β_j being positive or negative, respectively* [*condition* (4b)].

It is interesting to examine what these conditions can be reduced to in two simple cases. For a regression on two explanatory variables, X_1 and X_2,[3] the estimate of β_1 would be

$$\hat{\beta}_1 = A(r_{1y} - r_{12}r_{2y})/(1 - r_{12}^2), \tag{5}$$

[2]Condition (4a) obviously includes the (necessary) condition $|t_i| > |t_j|$ since $|\rho_{ij}| \leqq 1$, V being positive definite. Condition (4b) is derived from (4) as follows: if $t_j > 0$ (i.e., $\hat{\beta}_j > 0$) (4) becomes $\rho_{ij}t_i > t_j > 0$, so that sign (ρ_{ij}) = sign (t_i) = sign $(\hat{\beta}_i)$; conversely, if $t_j < 0$ (i.e., $\hat{\beta}_j < 0$), (4) becomes $\rho_{ij}t_i < t_j < 0$ and ρ_{ij} must have a sign different than that of t_i (and $\hat{\beta}_i$).

[3]Also other explanatory variables, X_i ($i > 2$), could be present in the regression, provided that they were orthogonal to X_1 and X_2.

where $A = [(y'y)/(x_1'x_1)]^{\frac{1}{2}} > 0$, r_{1y} is the sample correlation coefficient between X_1 and X_2 (in this case, $r_{12} = -\rho_{12}$), while r_{1y} and r_{2y} are the sample correlation coefficients between X_1 and Y and X_2 and Y, respectively. Constraining β_2 to zero, we will have

$$\hat{\beta}_1^* = Ar_{1y}. \tag{6}$$

It is immediately evident from (6) that the sign of the estimate of β_1 will be reversed, omitting X_2 from the regression, if and only if

$$r_{1y} \lesseqgtr 0 \quad \text{as} \quad \hat{\beta}_1 \gtreqless 0. \tag{7}$$

a result which could have been easily obtained starting from conditions (4a) and (4b) and going through some algebra.

Similarly, it is easy to show that in the presence of a third explanatory variable, X_3, a reversal in the sign of the estimate of β_1 will occur, dropping X_2 from the regression, if and only if

$$r_{1y} \lesseqgtr r_{13}r_{3y} \quad \text{as} \quad \hat{\beta}_1 \gtreqless 0. \tag{8}$$

In the two simple situations considered, then, conditions (4a) and (4b), which involve correlation coefficients between the estimates of β_i and β_j, reduce to simple conditions involving only sample correlation coefficients of the variables present in the regression.

References

Goldberger, A.S., 1964, Econometric theory (Wiley, New York).

Leamer, E.E., 1975, A result on the sign of restricted least-squares estimates, Journal of Econometrics 3.