Erasmus University Rotterdam
ECONOMETRIC INSTITUTE

Report 7406/E

POST-EXPERIMENT INTRODUCTION OF CONSTRAINTS ON PARAMETERS

by T. Kloek and F.B. Lempers

May 1974

POST-EXPERIMENT INTRODUCTION OF CONSTRAINTS ON PARAMETERS

by T. Kloek* and F.B. Lempers**

Contents

1. INTRODUCTION

The highest squared multiple correlation coefficient R^2 often plays an important role when choosing from a set of alternative specifications of the linear model. It is common practice, however, to be wary of choosing functions which have a high R^2 but many "nonsignificant" coefficients, or coefficients whose signs or magnitudes have no theoretical support. (See, e.g., Goldberger (1968), p.130.)

In practice, one often encounters multicollinearity among the explanatory variables. Then, the method of least squares frequently gives unreasonable estimates of the parameters, while another specification with sensible estimates would fit the data almost as well. Forcing one key parameter to a reasonable value frequently makes the others reasonable, but lowers R^2 only slightly (see, e.g., Almon (1969), pp.39-40).

In this note we suggest an extension (or modification) of the output of least-squares computer subroutines which enables the researcher, who is confronted with an unreasonable parameter estimate, to investigate by simple hand computation the consequences of changing this value for the other parameter estimates and the estimated residual variance. (In addition we present the consequences for the variance-covariance matrix.)

* Econometric Institute, Netherlands School of Economics, Erasmus University, Rotterdam.

** Central Planning Bureau, The Hague.

Such a procedure can help the researcher in choosing the parameter to be fixed, and can, therefore, be less time consuming and less expensive than arbitrary use of the computer.

2. DERIVATION OF THE COMPUTING RULES

We consider the linear model

$$y = X\beta + \varepsilon \tag{2.1}$$

where y is an n-dimensional vector of values taken by the dependent variable, X is an n×k matrix (of rank k) of values taken by the k non-stochastic explanatory variables, β is a vector of unknown parameters and ε is a vector of unknown independent and identically distributed random variables with zero mean and unknown variance σ^2, the disturbances.

The least-squares estimator of β is $b = (X'X)^{-1}X'y$. Well-known estimators of σ^2 are $s^2 = (y - Xb)'(y - Xb)/n$ and $\bar{s}^2 = ns^2/(n - k)$. The variance-covariance matrix of b equals $\sigma^2(X'X)^{-1}$. The (unbiased) estimator of the variance-covariance matrix is $\bar{s}^2(X'X)^{-1}$.

Without loss of generality we assume that the first element of the vector b (b_1) is not in accordance with the <u>a priori</u> knowledge available. A possible solution to this problem is to specify an acceptable value, e.g., b_1^*. Then, application of least squares to the transformed model

$$w \equiv y - b_1^* x_1 = X_r\beta_r + \varepsilon \tag{2.2}$$

where the subscript r refers to remainder, yields a new estimator b_r^* of β_r and its variance-covariance matrix.

In order to investigate the differences between the original and the new estimator we partition (2.1) as follows

$$y = [x_1 \quad X_r] \begin{bmatrix} \beta_1 \\ \beta_r \end{bmatrix} + \varepsilon \tag{2.3}$$

so that

$$\begin{bmatrix} b_1 \\ b_r \end{bmatrix} = \begin{bmatrix} x_1'x_1 & x_1'X_r \\ X_r'x_1 & X_r'X_r \end{bmatrix}^{-1} \begin{bmatrix} x_1'y \\ X_r'y \end{bmatrix} \tag{2.4}$$

The estimated variance-covariance matrix can be written as

$$(2.5)\qquad \bar{s}^2 \begin{bmatrix} x_1'x_1 & x_1'X_r \\ X_r'x_1 & X_r'X_r \end{bmatrix}^{-1} = \bar{s}^2 \begin{bmatrix} \alpha & a' \\ a & A \end{bmatrix} = \begin{bmatrix} \bar{\alpha} & \bar{a}' \\ \bar{a} & \bar{A} \end{bmatrix}$$

where

$$(2.6)\qquad \begin{cases} \bar{\alpha} = \bar{s}^2/[x_1'\{I - X_r'(X_r'X_r)^{-1}X_r'\}x_1] \\ \bar{a} = \bar{\alpha}p_1 \\ \bar{A} = \bar{s}^2(X_r'X_r)^{-1} + \bar{\alpha}p_1p_1' \end{cases}$$

and

$$(2.7)\qquad p_1 = -(X_r'X_r)^{-1}X_r'x_1$$

If the computer output of the least-squares subroutine contains the estimated variance-covariance matrix of b, viz. $\bar{s}^2(X'X)^{-1}$, one has $\bar{\alpha}$, $\bar{a}$, and $\bar{A}$ at his disposal.

We now turn to the minimization of the sum of the squared estimated disturbances under the linear constraint $b_1 = b_1^*$. For this we refer to Theil (1971), pp. 43-45, who derives the least-squares adjustment under the linear constraint

$$(2.8)\qquad r = Rz$$

where z stands for the parameter vector and r and R are given matrices of order q×1 and q×k, respectively. Under the condition that R has rank q, he obtains for the constrained least-squares solution, to be denoted by b* (compare Theil (1971) p.44, eq. (8.9)):

$$(2.9)\qquad b^* = b + (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}(r - Rb)$$

For the case we consider we can write

$$(2.10)\qquad r = b_1^* \qquad R = [1 \quad 0'] \qquad z' = [b_1 \quad b_r']$$

where 0' is a (k-1) dimensional zero row vector.

On combining (2.6), (2.9) and (2.10) one easily derives

$$(2.11) \qquad b_r^* = b_r + (b_1^* - b_1)\,\bar{a}/\bar{\alpha}$$
$$= b_r + (b_1^* - b_1)p_1$$

In words, the difference between the original and the new estimator of a certain element of β_r is equal to the product of the difference between the original estimate and the new value of β_1 and the corresponding element of the first column vector of $\bar{s}^2(X'X)^{-1}$ divided by the first diagonal element.[1] Generalization for the case of changing an arbitrary element of β is straightforward. For further comments we refer to Section 3.

The next problem is the derivation of the estimated residual variance. From Theil (1971), p.45, eq.(8.10), it follows that

$$(2.12) \qquad (y - Xb^*)'(y - Xb^*) = y'My + (r - Rb)'[R(X'X)^{-1}R']^{-1}(r - Rb)$$

where $M = I - X(X'X)^{-1}X'$. Making use of $y'My = ns^2$, and denoting $ns^{*2} = (y - Xb^*)'(y - Xb^*)$, we obtain from (2.6), (2.10), and (2.12)

$$(2.13) \qquad ns^{*2} = ns^2 + (b_1 - b_1^*)^2\, \bar{s}^2/\bar{\alpha}$$

Since $\bar{s}^{*2} = ns^{*2}/(n - k + 1)$, we can derive from (2.13)

$$(2.14) \qquad \bar{s}^{*2} = \bar{s}^2[\{(n - k) + (b_1 - b_1^*)^2/\bar{\alpha}\}/(n - k + 1)]$$

This expression can be written as

$$(2.15) \qquad \bar{s}^{*2} = \bar{s}^2\, [1 + \frac{t_1^{*2} - 1}{n - k - 1}]$$

where $t_1^* = (b_1 - b_1^*)\bar{\alpha}^{-\frac{1}{2}}$ is the corresponding t-value of the difference $b_1 - b_1^*$. Note that $\bar{s}_1^{*2} < \bar{s}^2$ if $|t^*| < 1$. The relative change in the residual variance $(\bar{s}^{*2} - \bar{s}^2)/\bar{s}^2 = (t_1^{*2} - 1)/(n - k + 1)$ can simply be computed by hand. The same holds for (2.14), provided that $\bar{s}^2$ is available.

---

[1] For an alternative interpretation, see the Appendix.

For completeness we also consider the variance-covariance matrix of b_r^*. We then need an expression for $(X_r'X_r)^{-1}$. From (2.6) we find

$$(2.16) \qquad (X_r'X_r)^{-1} = \{1/\bar{s}^2\}[\bar{A} - \bar{\alpha}p_1p_1'] = \{1/\bar{s}^2\}[\bar{A} - \bar{a}\bar{a}'/\bar{\alpha}]$$

Upon combining (2.14) and (2.16) we conclude that the estimated variance-covariance matrix of b_r^* equals

$$(2.17) \qquad \bar{s}^{*2}(X_r'X_r)^{-1} = [\{(n-k) + (b_1-b_1^*)^2/\bar{\alpha}\}/(n-k+1)][\bar{A} - \bar{a}\bar{a}'/\bar{\alpha}]$$

which in principle can also be computed by hand since all necessary figures are available, although this formula is somewhat less simple than (2.11) and (2.14).

3. POSSIBILITIES OF APPLICATION

In this section we discuss how the results of the preceding section could be applied.

Let c_{ij} be the typical element of the variance-covariance matrix $\bar{s}^2(X'X)^{-1}$ and let p_{ij} be defined by

$$(3.1) \qquad p_{ij} = c_{ij}/c_{jj}$$

Note that p_1 defined in (2.7) and used in (2.11) equals

$$\begin{bmatrix} p_{21} \\ p_{31} \\ \vdots \\ p_{k1} \end{bmatrix}$$

Now we may consider the four following possibilities for the output of a routine computer program:[2]

[2] It is understood that point estimates, residual variance, R^2 and Durbin-Watson statistic are printed anyway.

(i) Print standard errors (or t-values);

(ii) Print standard errors and variance-covariance matrix;

(iii) Print standard errors and $[p_{ij}]$;

(iv) Print standard errors, variance-covariance matrix, and $[p_{ij}]$.

In case (i) valuable information is suppressed anyhow. Case (ii) is rather common nowadays. In our view it could be improved upon by either (iii) or (iv). Given the p_{ij} the result (2.11) can simply be computed by hand. An additional advantage of printing p_{ij} is that the squared correlations are equal to $p_{ij}p_{ji}$ which is simpler to compute than $c_{ij}^2/c_{ii}c_{jj}$. The formula for the revised residual variance (2.15) does not require special print-out.

In the context of maximum-likelihood or (non-informative) Bayesian methods the same or a slightly modified procedure can be followed.

REFERENCES

Almon, C.A. (1967), Matrix Methods in Economics, Addison-Wesley, Reading, Mass.

Goldberger, A.S. (1968), Topics in Regression Analysis, Macmillan, Collier-Macmillan, London.

Theil, H. (1971), Principles of Econometrics, North-Holland Publishing Company, Amsterdam.

APPENDIX

In this appendix we give a different approach to derive our results. In our view this is helpful for understanding their meaning.

On combining (2.6) and (2.11) we find

(A.1) $$b_r^* = b_r + (b_1^* - b_1)(X_r'X_r)^{-1}X_r'x_1$$

The expresion $(X_r'X_r)^{-1}X_r'x_1$ at the rhs of (A.1) can be interpreted as the least-squares estimator d of the auxiliary specification

(A.2) $$x_1 = X_r\delta + \eta$$

so that we obtain (compare (2.6))

(A.3) $$x_1 = X_r d + f \qquad \text{and} \qquad d = -\bar{a}/\bar{\alpha} = -p_1$$

If we denote the estimated disturbances of (2.1) by e, we obtain

(A.4) $$y = x_1 b_1 + X_r b_r + e$$

On combining (A.4), (2.2), and (A.3) we can write

(A.5) $$y - b_1^* x_1 = X_r[b_r + (b_1 - b_1^*)d] + e + (b_1 - b_1^*)f$$

From (2.13), (A.4),(2.6), and (A.3) it follows that

(A.6) $$ns^{*2} = e'e + (b_1 - b_1^*)^2 f'f$$

where $e'e = ns^2$, which is equal to the sum of squares of the residuals of (A.5) since $e'f = 0$.

Now we can give the following interpretation. If the explanatory variables in X_r hardly "explain" x_1, d will be relatively small and f relatively large. We then expect a small change in b_r (compare (A.5)) and a large difference between ns^{*2} and ns^2; see (A.6). If, on the other had, X_r "explains" x_1 very well (the case of multicollinearity) we can expect a large difference between b_r and b_r^*, and a relatively small increase (if not decrease) in the sum of the squared estimated disturbances.