Most Efficient Least Squares Estimator in
Multivariate Regression

The model in matrix notation is

\[ y = X \beta + u \]

with \( X \) standardized, i.e., \( X'X = T_l \). The regression coefficient estimator is

\[ b = (X'X)^{-1}X'y = \frac{1}{T}X'y = \beta + \frac{1}{T}X'u. \]

Let

\[ b_1 = \frac{1}{T}R'y \]

be any linear estimator of \( \beta \). The object is to show that the minimum value of \( \delta \), where

\[ \delta = \text{E}(y_1 - \eta)'(y_1 - \eta) \]

with

\[ y_1 = Xb_1; \quad \eta = X\eta \]

is attained for \( R = X \).

From (1), (3) and (5),

\[ y_1 = \frac{1}{T}XR'(X\eta + u). \]

Hence

\[ y_1 - \eta = -X\eta + \frac{1}{T}XR'X + \frac{1}{T}XR'u, \]

so that, from (4),

\[ \delta = -\beta'X'R\eta - \beta'R'X\eta + \frac{1}{T}\beta'X'R'R'X\beta + \frac{1}{T}\text{Eu}'R'R'u, \]

where \( \delta \) is the value of the terms in \( R \) in \( \delta (4) \).
We now propose finding the minimum value of \( \varphi \) from

\[
\frac{\partial \varphi}{\partial r_{ti}} = 0, \quad t = 1, 2, \ldots, T; \quad i = 1, 2, \ldots, k.
\]

or, rather, showing that \( \varphi \) is satisfied by \( r_{ti} = x_{ti} \), these being the respective elements of \( R \) and \( X \). Call the terms on the r.s. of (8) \( T_1, T_2, T_3, T_4 \). Clearly, \( \text{w.l.o.g., } T_4 \), which equals \( \varphi^2 TRR' \), can be taken as constant independent of \( R \), e.g. \( \varphi^2 \). Of course, \( T_1 = T_2 \). Set

\[
X_{\cdot j} = Z = \{z_1, z_2, \ldots, z_T\}
\]

it can then easily be shown that

\[
\frac{\partial T_1}{\partial r_{ti}} = -z_{t\cdot i} = \frac{\partial T_2}{\partial r_{ti}}
\]

and not quite so easily that

\[
\frac{\partial T_3}{\partial r_{ti}} = \frac{2}{T} \sum_{t} \frac{T}{s=1} z_{s} r_{si}
\]

with, from (10),

\[
z_{s} = \sum_{j=1}^{k} x_{s} x_{ij}.
\]

On substitution for \( z_{s} \), given by (13) in the brackets \( ( ) \) in (12), setting \( r_{\cdot i} = x_{si} \) and using the orthogonal property of \( X \), we find

\[
\frac{\partial T_3}{\partial r_{ti}} = 2z_{t\cdot i}.
\]

Accordingly, from (11) and (14),

\[
\frac{\partial}{\partial r_{ti}} (T_1 + T_2 + T_3) = 0,
\]

for \( r_{ti} = x_{ti} \) or \( R = X \).
We must bear in mind, however, that \( R \) has been conditioned by \( T_4 = \epsilon^2 \), which should be introduced in Lagrangean form into the expression to be minimized, i.e.

\[
(16) \quad q = T_1 + T_2 + T_3 + T_4 - \lambda (T_4 - \epsilon^2)
\]

so that \( \frac{\partial q}{\partial r_{ti}} = 0 \) with \( T_4 = \epsilon^2 \) are satisfied by

\[ r_{ti} = x_{ti} \text{ and } \lambda = 1. \]

So we have proved the intuitive result that the best linear estimator of the coefficient matrix \( \beta \) is the regression estimator \( b \). No novelty is claimed: it is a Gauss-Markoff property. The matrix treatment and the use of standardization of \( X \) may have some interest.

16 June 1967
Revised 17 October 1967

R. C. Geary