The Annals of Mathematical Statistics
1970, Vol. 41, No. 2, 688–691

RECURSIVE ESTIMATION OF A SUBSET OF REGRESSION COEFFICIENTS[1]

By Richard H. Jones

University of Hawaii

1. Summary. Equations are derived for updating estimates of a subset of the regression coefficients in a linear model when more data become available. It is assumed that estimates of the remaining coefficients in the model are of no interest. A simple, but non-trivial special case is the centering of the regression equation by removing an unknown constant.

2. Introduction. In regression analysis, often estimates of only a few of the regression coefficients in the model are of interest. In this case the regression equation can be written

$$y = X\beta + Z\gamma + \varepsilon \tag{1}$$

where y is an $(n \times 1)$ vector of observations on the dependent variable, X is an $(n \times p)$ matrix of known constants, β is a $(p \times 1)$ vector of regression coefficients, Z is an $(n \times q)$ matrix of known constants, γ is a $(q \times 1)$ vector of regression coefficients for which estimates are not needed, and ε is an $(n \times 1)$ vector of uncorrelated errors with mean zero and constant variance. It is assumed that the $[n \times (p+q)]$ matrix $[X\colon Z]$ is of rank $p+q$. It is well known that the estimate of β can be written

$$b = [(X-\hat{X})'(X-\hat{X})]^{-1}(X-\hat{X})'(y-\hat{y}) \tag{2}$$

where

$$\hat{X} = Z(Z'Z)^{-1}Z'X \tag{3}$$

$$\hat{y} = Z(Z'Z)^{-1}Z'y.$$

It is also well known (see, for instance, Bartlett (1951)) that in the standard regression model

$$y = X\beta + \varepsilon \tag{4}$$

where y is $(n \times 1)$, the estimate

$$b_n = \Sigma_n X'y \tag{5}$$

where $\Sigma_n = (X'X)^{-1}$ can be updated when a new observation becomes available. Let y_{n+1} be the new observation corresponding to a new row of the X matrix, x', then

$$b_{n+1} = b_n + \Sigma_n x(y_{n+1} - x'b_n)/d \tag{6}$$

where $d = 1 + x'\Sigma_n x$.

Received May 9 1969.

[1] Research sponsored by the Air Force Office of Scientific Research, Office of Aerospace Research, United States Air Force, under AFOSR Contract No. F44620-69-C-0035.

The updated $(X'X)^{-1}$ matrix is

(7) $$\Sigma_{n+1} = \Sigma_n - \Sigma_n xx'\Sigma_n/d$$

and the residual sum of squares can also be updated (Duncan and Jones, (1966)).

(8) $$(e'e)_{n+1} = (e'e)_n + (y_{n+1} - x'b_n)^2/d.$$

3. Recursive subset estimation. It is not difficult to combine these two estimation techniques. (2) can be written

(9) $$b_n = \Sigma_n G_n$$ where

(10) $$\Sigma_n^{-1} = X'[I - Z(Z'Z)^{-1}Z']X$$
$$G_n = X'[I - Z(Z'Z)^{-1}Z']y.$$

Let x' be the new row of the X matrix, z' be the new row of the Z matrix, and let

(11) $$M_n = X'Z(Z'Z)^{-1}Z'X.$$ Then

(12) $$M_{n+1} = (X'Z + xz')(Z'Z + zz')^{-1}(Z'X + zx').$$ Using

(13) $$(Z'Z + zz')^{-1} = (Z'Z)^{-1} - \frac{(Z'Z)^{-1}zz'(Z'Z)^{-1}}{1 + z'(Z'Z)^{-1}z},$$ gives

(14) $$M_{n+1} = M_n + xx' - \frac{(x - \hat{x})(x - \hat{x})'}{1 + z'(Z'Z)^{-1}z}$$ where

(15) $$\hat{x} = X'Z(Z'Z)^{-1}z.$$ Since

(16) $$\Sigma_{n+1}^{-1} + M_{n+1} = \Sigma_n^{-1} + M_n + xx',$$

(17) $$\Sigma_{n+1}^{-1} = \Sigma_n^{-1} + \frac{(x - \hat{x})(x - \hat{x})'}{1 + z'(Z'Z)^{-1}z}.$$ Similarly

(18) $$G_{n+1} = G_n + \frac{(x - \hat{x})(y_{n+1} - \hat{y}_{n+1})}{1 + z'(Z'Z)^{-1}z}$$ where

(19) $$\hat{y}_{n+1} = y'Z(Z'Z)^{-1}z.$$ Using (13),

(20) $$\Sigma_{n+1} = \Sigma_n - \Sigma_n(x - \hat{x})(x - \hat{x})'\Sigma_n/d$$ where

(21) $$d = 1 + z'(Z'Z)^{-1}z + (x - \hat{x})'\Sigma_n(x - \hat{x}).$$

The updated regression coefficients are

(22) $$b_{n+1} = b_n + \Sigma_n(x - \hat{x})[y_{n+1} - \hat{y}_{n+1} - (x - \hat{x})'b_n]/d,$$

and the new sum of squares of residuals is

$$(e'e)_{n+1} = (e'e)_n + (y_{n+1} - \hat{y}_{n+1} - (x - \hat{x})'b_n)^2/d. \tag{23}$$

(23) can be derived directly from (8) by replacing $x'b_n$ by $x'b_n + z'\gamma_n$ where $\gamma_n = (Z'Z)^{-1}Z'(y - Xb_n)$, and by replacing $1 + x'\Sigma_n x$ by

$$1 + [x' \vdots z'] \begin{bmatrix} X'X & \vdots & X'Z \\ \cdots & & \cdots \\ Z'X & \vdots & Z'Z \end{bmatrix} \begin{bmatrix} x \\ \cdots \\ y \end{bmatrix}$$

using the well-known formula

$$\begin{bmatrix} X'X & \vdots & X'Z \\ \cdots & & \cdots \\ Z'X & \vdots & Z'Z \end{bmatrix}^{-1} = \begin{bmatrix} \Sigma_n^{-1} & \vdots & -\Sigma_n^{-1}X'Z(Z'Z)^{-1} \\ \cdots & & \cdots \\ -(Z'Z)^{-1}Z'X\Sigma_n^{-1} & \vdots & (Z'Z)^{-1} + (Z'Z)^{-1}Z'X\Sigma_n^{-1}X'Z(Z'Z)^{-1} \end{bmatrix}, \tag{24}$$

where Σ_n^{-1} is defined in (10).

4. Outline of steps. The recursive procedure can now be summarized. At the beginning of each step the following matrices will be available (the numbers in parentheses below each matrix show the dimensions)

$$\underset{(p \times 1)}{b_n}, \quad \underset{(p \times p)}{\Sigma_n}, \quad \underset{(1 \times 1)}{(e'e)_n}, \quad \underset{(q \times q)}{(Z'Z)_n^{-1}}, \quad \underset{(p \times q)}{(X'Z)_n}, \quad \underset{(1 \times q)}{(y'Z)_n}.$$

Now

$$\underset{(p \times 1)}{x}, \quad \underset{(q \times 1)}{z}, \quad \text{and} \quad \underset{(1 \times 1)}{y_{n+1}}$$

are observed. $\hat{x}$ is calculated from (15), and $\hat{y}_{n+1}$, d, b_{n+1}, $(e'e)_{n+1}$ from (19), (21), (22) and (23). Σ_{n+1} is calculated from (20). The rows and columns of Σ_{n+1} can be divided by the square roots of the diagonal elements giving the correlation matrix of the estimated regression coefficients, b_{n+1}. The recursion is completed by calculating $(Z'Z)_{n+1}^{-1}$ from (13) and

$$\begin{aligned} (X'Z)_{n+1} &= (X'Z)_n + xz' \\ (y'Z)_{n+1} &= (y'Z)_n + y_{n+1}z'. \end{aligned} \tag{25}$$

5. Example. The simplest example is the centering of a regression equation by removing a constant. In this case Z becomes a column of ones, $(Z'Z)_n^{-1} = 1/n$, $(X'Z)_n$ is the sums of the column of X arranged in a column vector,$(y'Z)_n$ is the sum of the observations on the dependent variable. $\hat{x} = n^{-1}(X'Z)$ are the means of the columns of X and $\hat{y}_{n+1}$ is the mean of y. Then

$$d = 1 + n^{-1} + (x - \hat{x})'\Sigma_n(x - \hat{x}).$$

Therefore, when centering a regression equation, the means of all previous observations on the dependent and independent variables are subtracted from the new observations. The recursions are then carried out as in ordinary recursive estimation (6) through (8) except that the 1 in d is replaced by $1+n^{-1}$.

REFERENCES

[1] BARTLETT, M. S. (1951). An inverse matrix adjustment arising in discriminant analysis. *Ann. Math. Statist.* **22** 107–11.

[2] DUNCAN, D. B. and JONES, R. H. (1966). Multiple regression with stationary errors. *J. Amer. Statist. Assoc.* **61** 917–928.