X-547-65-497

THE EFFECT OF ADDITIONAL OBSERVATIONS ON A PREVIOUS LEAST SQUARES ESTIMATE

by
Raymond V. Borchers

December 1965

GODDARD SPACE FLIGHT CENTER
Greenbelt, Maryland

CONTENTS

LIST OF SYMBOLS

X true value of $k \times 1$ vector of regression parameters

$\hat{X}_n$ original least squares estimate of X depending upon n observations

$\bar{X}$ $k \times 1$ vector of deviations in the regression parameters from reference; $\bar{X} = X - \hat{X}$

y $n \times 1$ vector of observations or measurements

$\hat{y}$ computed value of y vector

$\bar{y}$ vector of deviations in the observations; $\bar{y} = y - \hat{y}$

Y $(n + p) \times 1$ vector of deviations in the weighted observations; $Y = \sqrt{W}\,\bar{y}$

e $(n + p) \times 1$ noise vector associated with the observations

Q covariance matrix of the observational error e.

σ_i probable error in the ith observation

W $\text{diag}\left(\frac{1}{\sigma_1^2}, \frac{1}{\sigma_2^2}, \ldots, \frac{1}{\sigma_n^2}\right)$ "weighting" matrix

C_n covariance matrix of $\bar{X}$ depending upon n observations

I identity matrix

$E[\ \ ]$ expected value of $[\ \ ]$

$(\ )^T$ transpose of matrix $(\ )$

$(\ )^{-1}$ inverse of matrix $(\ )$

THE EFFECT OF ADDITIONAL OBSERVATIONS ON A PREVIOUS LEAST SQUARES ESTIMATE

INTRODUCTION

In a number of applications of the method of least squares it may be desirable to compute the influence of additional observations on a previously determined least squares estimate of the regression parameters and their associated covariance matrix. A particular application is presented in which the parameters are assumed to be constants, although the method can be extended to dynamic systems as shown by Gainer in reference [1]. Matrix equations which provide for additional observations are derived for the method of weighted least squares with uncorrelated errors. These recursion formulas reduce to the ones presented by Gainer in reference [1] in the case where only a single observation is added.

DERIVATION OF METHOD

The nonlinear estimation problem is usually solved by assuming a linear approximation of the true regression parameters in a neighborhood of a nominal or reference set of parameters. Let us assume that we have obtained a least squares estimate and also the covariance matrix of the regression parameters. We are concerned here with the problem of determining a revised least squares estimate and its associated covariance matrix when additional observations become available.

Consider the following system of n + p observation equations relating deviations in "weighted" observations to deviations in the k $(n > k)$ unknowns

$$\sqrt{W_1}\,\left(a_{11}\bar{X}_1 + a_{12}\bar{X}_2 + \ldots + a_{1k}\bar{X}_k\right) = \sqrt{W_1}\,\left(\bar{y}_1 + e_1\right)$$

$$\sqrt{W_2}\,\left(a_{21}\bar{X}_1 + a_{22}\bar{X}_2 + \ldots + a_{2k}\bar{X}_k\right) = \sqrt{W_2}\,\left(\bar{y}_2 + e_2\right)$$

$$\text{- -}$$

$$\sqrt{W_n}\,\left(a_{n1}\bar{X}_1 + a_{n2}\bar{X}_2 + \ldots + a_{nk}\bar{X}_k\right) = \sqrt{W_n}\,\left(\bar{y}_n + e_n\right) \qquad (1)$$

$$\text{- -}$$

$$\sqrt{W_{n+p}}\,\left(a_{n+p,1}\bar{X}_1 + a_{n+p,2}\bar{X}_2 + \ldots + a_{n+p,k}\bar{X}_k\right) = \sqrt{W_{n+p}}\,\left(\bar{y}_{n+p} + e_{n+p}\right).$$

Now let us apply the method of partitioning to the matrix equation

$$A\bar{X} = Y + e, \qquad (2)$$

and define

$$
A = \begin{pmatrix} \sqrt{W_n} \; A_n \\ \hline \sqrt{W_p} \; B_p \end{pmatrix} = \begin{pmatrix} E_n \\ \hline F_p \end{pmatrix}, \quad Y = \begin{pmatrix} \sqrt{W_1} \; \bar{y}_1 \\ \vdots \\ \sqrt{W_n} \; \bar{y}_n \\ \hline \sqrt{W_{n+1}} \; \bar{y}_{n+1} \\ \vdots \\ \sqrt{W_{n+p}} \; \bar{y}_{n+p} \end{pmatrix} = \begin{pmatrix} u_n \\ \hline v_p \end{pmatrix},
$$

$$
e = \begin{pmatrix} \sqrt{W_1} \; e_1 \\ \vdots \\ \sqrt{W_n} \; e_n \\ \hline \sqrt{W_{n+1}} \; e_{n+1} \\ \vdots \\ \sqrt{W_{n+p}} \; e_{n+p} \end{pmatrix} = \begin{pmatrix} \mathcal{E}_1 \\ \hline \mathcal{E}_2 \end{pmatrix} \tag{3}
$$

where

A_n is a $n \times k$ matrix of coefficients a_{ij}, $(i = 1, \ldots, n;\ j = 1, \ldots, k)$,

B_p is a $p \times k$ matrix of coefficients a_{ij}, $(i = n + 1, \ldots, n + p;\ j = 1, \ldots, k)$,

E_n is a $n \times k$ matrix of coefficients $\sqrt{W_i}\, a_{ij}$, $(i = 1, \ldots, n;\ j = 1, \ldots, k)$,

F_p is a $p \times k$ matrix of coefficients $\sqrt{W_i}\, a_{ij}$, $(i = n + 1, \ldots, n + p;\ j = 1, \ldots, k)$,

u_n is a $n \times 1$ vector of weighted observations,

v_p is a $p \times 1$ vector of additional weighted observations,

$\mathcal{E}_1$ is a $n \times 1$ vector of errors in weighted observations,

$\mathcal{E}_2$ is a $p \times 1$ vector of errors in additional weighted observations,

W_n is a $n \times 1$ diagonal "weighting" matrix.

Let us assume that the errors in the observations are random variables having a Gaussian distribution with zero mean and $E\left[ee^T\right] = Q$, where Q is a diagonal covariance matrix with elements σ_i^2. It can be shown that the least squares estimate of $\overline{X}$ which depends upon n observations is given by

$$\hat{\overline{X}}_n = \left(E_n^T E_n\right)^{-1} E_n^T, \ u_n = \left(A_n^T W_n A_n\right)^{-1} A_n^T W_n u_n \tag{4}$$

where the diagonal "weighting" matrix is given by

$$W_n = \begin{pmatrix} \frac{1}{\sigma_i^2} & & 0 \\ & \frac{1}{\sigma_2^2} & \\ & & \ddots & \\ 0 & & & \frac{1}{\sigma_n^2} \end{pmatrix}.$$

Now $C_n = E\left[\left(\hat{\overline{X}}_n - \overline{X}\right)\left(\hat{\overline{X}}_n - \overline{X}\right)^T\right] = \left(A_n^T W_n A_n\right)^{-1} = \left(E_n^T E_n\right)^{-1}$ is the covariance matrix of $\hat{\overline{X}}_n$, and it is important for its statistical information. Let $\overline{X}_i$, $i = 1, 2, \ldots, k$, be the parameters to be estimated. Then an unbiased estimate of the variance of $\overline{X}_i$ is given by

$$\hat{\sigma}_{\overline{X}_i}^2 = \frac{e^T W e}{N - k} b_{ii} \tag{5}$$

where

b_{ii} is the ith diagonal element of C_n,

N is the number of observation equations,

k is the number of parameters to be estimated.

It would be of interest to obtain a revised estimate of $\overline{X}_n$ and C_n by considering p additional observations. The estimate of $\overline{X}$ which depends upon $n + p$ observations is

$$\hat{\bar{X}}_{n+p} = \left(\begin{pmatrix} E_n \\ \hline F_p \end{pmatrix}^T \begin{pmatrix} E_n \\ \hline F_p \end{pmatrix}\right)^{-1} \begin{pmatrix} E_n \\ \hline F_p \end{pmatrix}^T \begin{pmatrix} u_n \\ \hline v_p \end{pmatrix}, \quad p \geq 1 \tag{6}$$

$$= \left(\begin{pmatrix} \underset{k \times n}{E_n^T} & \vdots & \underset{k \times p}{F_p^T} \end{pmatrix} \begin{pmatrix} E_n \\ \hline F_p \end{pmatrix}_{\substack{n \times k \\ p \times k}}\right)^{-1} \begin{pmatrix} \underset{k \times n}{E_n^T} & \vdots & \underset{k \times p}{F_p^T} \end{pmatrix} \begin{pmatrix} u_n \\ \hline v_p \end{pmatrix}_{\substack{n \times 1 \\ p \times 1}}$$

$$\hat{\bar{X}}_{n+p} = \left(E_n^T E_n + F_p^T F_p\right)^{-1}_{k \times k} \left(E_n^T u_n + F_p^T v_p\right)_{k \times 1} \tag{7}$$

Since $C_n^{-1} = \left(E_n^T E_n\right)_{k \times k}$ we write

$$\left(E_n^T E_n + F_p^T F_p\right)^{-1} = \left(C_n^{-1} + F_p^T F_p\right)^{-1} = (I - H)^{-1} C_n \tag{8}$$

where $H = - C_n F_p^T F_p$.

Substitution of (8) into (7) gives

$$\hat{\bar{X}}_{n+p} = (I - H)^{-1} C_n \left(E_n^T u_n + F_p^T v_p\right) \tag{9}$$

$$= (I - H)^{-1} \left(C_n E_n^T u_n + C_n F_p^T v_p\right)$$

$$\hat{\bar{X}}_{n+p} = (I - H)^{-1}_{k \times k} \left(\hat{\bar{X}}_n + C_n F_p^T v_p\right)_{k \times 1} . \tag{10}$$

Evaluation of $(I - H)^{-1}$

For the inversion of the matrix $(I - H)$ we find application of the following formula in reference [2]

$$\left(A + U S^{-1} V^T\right)^{-1} = A^{-1} - A^{-1} U \left(S + V^T A^{-1} U\right)^{-1} V^T A^{-1} \tag{11}$$

where A and S are nonsingular matrices, U and V are rectangular.

Putting

$$A = I_{k \times k}, \quad S = I_{p \times p}, \quad \underset{k\times p}{U} = \underset{k\times k}{C_n} \; \underset{k\times p}{F_p^T}, \quad V^T = \underset{p\times k}{F_p}$$

the form of (11) becomes

$$(I - H)^{-1}_{k\times k} = \underset{k\times k}{I} - \underset{k\times k}{C_n} \; \underset{k\times p}{F_p^T} \left(\underset{p\times p}{I} + \underset{p\times k}{F_p} \; \underset{k\times k}{C_n} \; \underset{k\times p}{F_p^T}\right)^{-1}_{p\times p} \underset{p\times k}{F_p} \tag{12}$$

Substitution of (12) into (10) gives

$$\hat{\bar{X}}_{n+p} = \left[I - C_n F_p^T \left(I + F_p C_n F_p^T\right)^{-1} F_p\right] \hat{\bar{X}}_n + C_n F_p^T v_p \tag{13}$$

We may select either of the equations (10) or (13) for the evaluation of $\hat{\bar{X}}_{n+p}$. It is to be observed that in equation (10) one always inverts a matrix of order k, where k is the number of unknowns; whereas in equation (13) it is necessary to invert a matrix of order p, where p is the number of additional observations. Consequently, if the situation arises where k is much greater than p, it would probably be desirable to use equation (13).

Special Case When $p = 1$

If we select $p = 1$ we find that (13) reduces to

$$\hat{\bar{X}}_{n+1} = \hat{\bar{X}}_n + \frac{C_n F_1^T}{1 + F_1 C_n F_1^T}\left(v_1 - F_1 \hat{\bar{X}}_n\right) \tag{14}$$

where $\left(F_1 C_n F_1^T\right)$ is now a scalar quantity for the case of a single additional observation. We see that (14) is equivalent to equation (36) in reference [1]. We can also observe that the estimate $\hat{\bar{X}}_{n+1}$ does not require any matrix inversion. Hence it is clear that the recursive equation (14) may be used as a least square sequential estimation procedure to obtain a revised estimate for the parameter vector.

Evaluation of C_{n+p}

The revised covariance matrix C_{n+p} is given by

$$C_{n+p} = \left(E_n^T E_n + F_p^T F_p\right)^{-1} = (I - H)^{-1} C_n \tag{15}$$

$$= \left[I - C_n F_p^T \left(I + F_p C_n F_p^T\right)^{-1} F_p\right] C_n .$$

Special Case When $p = 1$

Let $p = 1$. Then we find that (15) reduces to

$$C_{n+1} = C_n - \frac{C_n F_1^T F_1 C_n}{1 + F_1 C_n F_1^T} . \tag{16}$$

Let $\underset{k \times 1}{J} = \underset{k \times k}{C_n} \; \underset{k \times 1}{F_1^T}$. Then (16) becomes

$$C_{n+1} = C_n - \frac{J J^T}{1 + F_1 J} . \tag{17}$$

It is to be observed that (16) is equivalent to equation (34) in reference [1] and that it does not require matrix inversion.

GEOMETRICAL INTERPRETATION OF THE LEAST SQUARES PROBLEM

The presentation which follows utilizes certain basic concepts set forth by Householder in references [2] and [3] in providing a geometric interpretation of the method of weighted least squares.

Consider the linear least squares problem where one has given a $n \times k$ matrix A with $n > k$ rows and a n-vector y. One desires a k-vector X such that

$$A X = y + e, \quad P^T e = 0$$

where

e is the vector of residuals,

$$P = W A = \operatorname{diag}\left(\frac{1}{\sigma_i^2}, \frac{1}{\sigma_2^2}, \ldots \frac{1}{\sigma_n^2}\right) A .$$

Let us now consider the "weighting" matrix W as a metric for the space and hence define the orthogonality and lengths of the vectors with respect to this metric. One can give a geometrical interpretation to the method of least squares, i.e., the projection of an arbitrary given vector upon a certain subspace. Let the columns of the rectangular matrix A be linearly independent. Then the vector AX is a vector in the same space spanned by the columns of A and AX is said to be the orthogonal projection with respect to W of the vector y whenever $y - AX$ is orthogonal to A with respect to the positive definite and symmetric matrix W, i.e.,

$$A^T W (y - AX) = 0$$

or

$$X = \left(A^T W A\right)^{-1} A^T W y,$$

which is exactly the least squares solution of the usual normal equations. The assumption that the columns of A are linearly independent together with the fact that W is a positive definite symmetric matrix leads to the following definition of a generalized projector reference [3].

<u>Definition:</u> A matrix T_A is the generalized projector with respect to W for the space of A if $T_{A;W} = A \left(A^T W A\right)^{-1} A^T W$.

The generalized projector has the following properties:

(a) $T_{A;W}$ is a unique projector since if A is replaced by AK, where K is any nonsingular matrix, one obtains the same projector. This can be verified directly, i.e.,

$$T_{AK;W} = AK\left[(AK)^T W (AK)\right]^{-1} (AK)^T W$$

$$= AK\left[K^T \left(A^T W A\right) K\right]^{-1} K^T A^T W$$

$$= A\left(KK^{-1}\right)\left(A^T W A\right)^{-1}\left(K^T\right)^{-1} K^T A^T W$$

$$T_{AK;W} = A\left(A^T WA\right)^{-1} A^T W = T_{A;W}$$

(b) $WT_{A;W}$ is symmetric, i.e., $\left(WT_{A;W}\right)^T = WT_{A;W}$

Now $\left(WT_{A;W}\right)^T = \left[WA\left(A^T WA\right)^{-1} A^T W\right]^T = WA\left(A^T WA\right)^{-1} A^T W$

$= WT_{A;W}$

(c) $T_{A;W}$ is idempotent, i.e., $T^2_{A;W} = T_{A;W}$.

$T^2_{A;W} = A\left(A^T WA\right)^{-1} A^T WA\left(A^T WA\right)^{-1} A^T W = A\left(A^T WA\right)^{-1} A^T W$

$= T_{A;W}$.

(d) $T_{A;W}\,y$ is the orthogonal projection of the vector y. This is shown by taking the generalized scalar product of $T_{A;W}\,y$ and the residual vector $\left(y - T_{A;W}\,y\right)$, i.e.,

$\left(T_{A;W}\,y\right)^T W\left(I - T_{A;W}\right) y = y^T\left(WT_{A;W}\right)^T\left(I - T_{A;W}\right) y$

$= y^T WT_{A;W}\left(I - T_{A;W}\right) y = y^T W\left(T_{A;W} - T^2_{A;W}\right) y = 0$

since by (b) $WT_{A;W}$ is symmetric and by (c) $T_{A;W}$ is idempotent.

REFERENCES

1. Patrick A. Gainer, "A Method for Computing the Effect of An Additional Observation On A Previous Least-Squares Estimate", NASA TN D-1599, January 1963.

2. Alston S. Householder, "Principles of Numerical Analysis", McGraw-Hill Book Company, Inc., 1953.

3. Alston S. Householder, "The Theory of Matrices in Numerical Analysis", Blaisdell Publishing Co., 1964.