Some identities useful in the analysis of residuals from linear regression

by

. Christopher Bingham
University of Minnesota

Technical Report No. 300

August, 1977

Abstract:

Relationships between regression statistics computed for the complete data model $Y = X\beta + \epsilon$ and two submodels are presented. The models considered omit either the last n_2 rows (observations) of X and Y, or the last p_2 columns (independent variables) of X. The identities are useful in the analysis of residuals. In particular it is shown that the last n_2 residuals from the complete data model are expressible in terms of the deviations of the last n_2 Y values from predictions based on a model fit to only the first $n_1 = n - n_2$ rows, and vice versa. Also the difference between the least squares coefficients based on the complete data model and the submodel omitting n_2 observations is expressible in terms of these same residuals. Cook's distance for detecting influential observations is generalized to measure the influence of sets of n_2 observations and expressed in terms of these same residuals.

Key words: Cook's distance, matrix identities, testing for outliers, regression updating, residual analysis.

Some identities useful in the analysis of residuals from linear regression

by Christopher Bingham
University of Minnesota, St. Paul

1. Introduction

Consider the standard linear model situation

$$Y = X\beta + \epsilon \ , \tag{1.1}$$

where Y and ϵ are n by 1, X is n by p, β is p by 1 and $\mathrm{Cov}[\epsilon] = \sigma^2 I_n$. This will be referred to as the complete data model. Presented below are a number of identities relating matrices and vectors defined in terms of submodels of (1.1) with corresponding quantities for the complete data model. By a submodel is meant a model omitting from (1.1) rows (observations) or columns (variables), or both. Without loss of generality these can always be taken to be the last n_2 rows or p_2 columns. Many of the results are not new. Some are, indeed, well known. However the author and others have found it very useful to have them gathered together in a uniform notation. Where appropriate, some details of the statistical applications of these identities will be given.

The only case considered in detail is the full rank case. Most of the identities can be generalized to the non-full rank case by appropriate uses of generalized inverses.

Although the identities herein are mathematically correct, this does not necessarily imply that they should be used as building blocks in computational algorithms. See Chambers (1971) for a discussion of some of the numerical problems associated with expanding or contracting models.

Throughout the paper, the matrix of sums of squares and cross-products for the complete data model will be denoted by

$$S = X'X \ , \tag{1.2}$$

and the complete data vector of least squares estimates as

$$\hat{\beta} = S^{-1}X'Y \ . \tag{1.3}$$

2. Identities related to adding or deleting observations

In this section we assume that X and Y are partitioned matrices of the form

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \ , \qquad Y = \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} \ , \tag{2.1}$$

where X_1 is n_1 by p, X_2 is n_2 by p, Y_1 is n_1 by 1, and Y_2 is n_2 by 1, where $n_1 + n_2 = n$. In some applications, the first n_1 rows will be considered n_1 "good" observations, and the remaining rows as n_2 "suspect" observations. In other situations, the final n_2 rows just represent n_2 additional observations for which it is desired to update statistics computed for the first n_1. We define the matrices of sums of squares and cross-products as

$$S_j = X_j'X_j \ , \quad j = 1,2 \ ; \quad S_1 + S_2 = S \ . \tag{2.2}$$

Throughout this section we assume that $\text{rank}(X_1) = \text{rank}(S_1) = \text{rank}(S) = p$.

The starting place is the following well known result (see, e.g., Dempster (1969), eq. 7.5.19, 7.5.25):

Proposition 1:

$$S^{-1} = S_1^{-1} - S_1^{-1}X_2'(I_{n_2} + U)^{-1}X_2S_1^{-1} \quad \text{(adding rows)} \tag{2.3}$$

$$S_1^{-1} = S^{-1} + S^{-1}X_2'(I_{n_2} - D)^{-1}X_2S^{-1} \quad \text{(deleting rows)} \tag{2.4}$$

where

$$U = X_2 S_1^{-1} X_2' \ , \quad D = X_2 S^{-1} X_2' \qquad \text{(U for Update, D for Delete)} \ . \tag{2.5}$$

Moreover,

$$U = D(I - D)^{-1} \ , \quad D = U(I + U)^{-1} \tag{2.6}$$

and

$$(I + U)^{-1} = I - D \ , \quad (I - D)^{-1} = I + U \ . \tag{2.7}$$

Proof: Multiplying the r.h.s. of (2.3) by $S = S_1 + X_2'X_2$ yields

$$I_p - X_2'(I + U)^{-1}X_2S_1^{-1} + X_2'X_2S_1^{-1} - X_2'X_2S_1^{-1}X_2'(I + U)^{-1}X_2S_1^{-1}$$

$$= I_p - X_2'[(I + U)^{-1} - I + U(I + U)^{-1}]X_2S_1^{-1} = I_p \ ,$$

since $I - U(I + U)^{-1} = (I + U)^{-1}$. Hence (2.3) is valid. Equation (2.4) is verified the same way, multiplying the r.h.s. by $S_1 = S - X_2'X_2$. Equations (2.6) and (2.7) are trivial consequences of (2.3) and (2.4). □

Proposition 1 is valid in the non-full rank case with the substitution of generalized inverses S^- and S_1^- in place of S^{-1} and S_1^{-1}, provided $\text{rank}(X_1) = \text{rank}(X)$ (Rao and Mitra 1971, problem 20, p. 70).

The least squares estimate of β based on the first n_1 rows is

$$\hat{\beta}^{(1)} = S_1^{-1}X_1'Y_1 \ . \tag{2.8}$$

Define the residuals associated with the submodel consisting of the first n_1 rows as

$$r^{(1)} = \begin{bmatrix} r_1^{(1)} \\ r_2^{(1)} \end{bmatrix} = \begin{bmatrix} Y_1 - X_1\hat{\beta}^{(1)} \\ Y_2 - X_2\hat{\beta}^{(1)} \end{bmatrix} = Y - X\hat{\beta}^{(1)} \ , \tag{2.9}$$

and the residuals associated with the complete data model as

$$r = \begin{bmatrix} r_1 \\ r_2 \end{bmatrix} = \begin{bmatrix} Y_1 - X_1\hat{\beta} \\ Y_2 - X_2\hat{\beta} \end{bmatrix} = Y - X\hat{\beta} \ . \tag{2.10}$$

Then r and $r^{(1)}$ can be expressed as

$$r = MY \quad \text{and} \quad r^{(1)} = M^{(1)}Y \ ,$$

where the residual projection operators M and $M^{(1)}$ are

$$M = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} = I_n - XS^{-1}X' = \begin{bmatrix} I_{n_1} - X_1S^{-1}X_1' & -X_1S^{-1}X_2' \\ -X_2S^{-1}X_1' & I_{n_2} - X_2S^{-1}X_2' \end{bmatrix}, \quad (2.11)$$

and

$$M^{(1)} = I_{n_1} - X_1S_1^{-1}X_1' . \quad (2.12)$$

Note that

$$M_{22} = I - D = (I + U)^{-1} . \quad (2.13)$$

Also, using (2.3),

$$M_{12} = M_{21}' = -X_1S_1^{-1}X_2'(I - (I + U)^{-1}U) = -X_1S_1^{-1}X_2'(I + U)^{-1} , \quad (2.14)$$

and

$$\begin{aligned} M_{11} &= I - X_1S_1^{-1}X_1' + X_1S_1^{-1}X_2'(I + U)^{-1}X_2S_1^{-1}X_1' \\ &= M^{(1)} + M_{12}M_{22}^{-1}M_{21} . \end{aligned} \quad (2.15)$$

From (2.13) and (2.14) follow

$$M_{12}M_{22}^{-1} = -X_1S_1^{-1}X_2' . \quad (2.16)$$

We can use the above results to express the residuals r from the complete data model in terms of the residuals $r^{(1)}$ from the submodel and vice versa:

<u>Proposition 2</u>:

$$r_2 = (I + U)^{-1}r_2^{(1)} = (I - D)r_2^{(1)} \quad (2.17)$$

$$r_2^{(1)} = (I - D)^{-1}r_2 = (I + U)r_2 \quad (2.18)$$

$$r_1 = r_1^{(1)} - X_1S_1^{-1}X_2'(I + U)^{-1}r_2^{(1)} = r_1^{(1)} - X_1S_1^{-1}X_2'r_2 \quad (2.19)$$

$$r_1^{(1)} = r_1 + X_1S^{-1}X_2'(I - D)^{-1}r_2 = r_1 + X_1S^{-1}X_2'r_2^{(1)} . \quad (2.20)$$

Proof:

Using (2.13) and (2.16)

$$r_2 = M_{21}Y_1 + M_{22}Y_2 = M_{22}(Y_2 - X_2S_1^{-1}X_1'Y_1) = (I + U)^{-1}r_2^{(1)} .$$

Equations (2.19) and (2.20) are similarly proved. □

We see from (2.17) and (2.18) that the complete data model residuals associated with the last n_2 observations are a linear transformation of the residuals from the predicted means of these rows as computed from the submodel omitting these rows, and vice versa. We can use this correspondence to relate $\hat{\beta}$ and $\hat{\beta}^{(1)}$.

Proposition 3:

$$\hat{\beta} = \hat{\beta}^{(1)} + S_1^{-1}X_2'(I + U)^{-1}r_2^{(1)} = \hat{\beta}^{(1)} + S_1^{-1}X_2'r_2 \tag{2.21}$$

$$\hat{\beta}^{(1)} = \hat{\beta} - S^{-1}X_2'(I - D)^{-1}r_2 = \hat{\beta} - S^{-1}X_2'r_2^{(1)} \tag{2.22}$$

Proof: We have, using (2.3),

$$\begin{aligned}\hat{\beta} &= S^{-1}(X_1'Y_1 + X_2'Y_2) = S_1^{-1}X_1'Y_1 - S_1^{-1}X_2'(I + U)^{-1}X_2S_1^{-1}X_1'Y_1 \\ &\qquad + S_1^{-1}X_2'Y_2 - S_1^{-1}X_2'(I + U)^{-1}X_2S_1^{-1}X_2'Y_2 \\ &= \hat{\beta}^{(1)} - S_1^{-1}X_2'[(I + U)^{-1}X_2\hat{\beta}^{(1)} - Y_2 + (I + U)^{-1}UY_2] \\ &= \hat{\beta}^{(1)} + S_1^{-1}X_2'(I + U)^{-1}r_2^{(1)} = \hat{\beta}^{(1)} + S_1^{-1}X_2'r_2 ,\end{aligned}$$

thus establishing (2.21). Equation (2.22) is proved similarly, or one can apply to (2.21) the easily established identity

$$S^{-1}X_2' = S_1^{-1}X_2'(I + U)^{-1} . \quad \square \tag{2.23}$$

A particular case of (2.22) ($n_2 = 1$) is given by Miller (1974).

Cook (1977a, 1977b) has considered the amount of change in the least squares estimates when one or more observations are deleted from the complete data model as a measure of how influential those observations are. To express this change as a single quantity, it is natural to consider statistics of the form

$$D_Q^2 = (\hat{\beta} - \hat{\beta}^{(1)})'Q(\hat{\beta} - \hat{\beta}^{(1)}) \ , \tag{2.24}$$

for a suitably chosen positive semi-definite matrix Q. Using (2.21) or (2.22) we have

$$D_Q^2 = r_2'X_2'S_1^{-1}QS_1^{-1}X_2r_2 = r_2^{(1)\prime}X_2'S^{-1}QS^{-1}X_2r_2^{(1)} \ . \tag{2.25}$$

Thus D_Q^2 is also expressible as a quadratic form in the residuals associated with the deleted observations. Two attractive choices are $Q = S$ and $Q = S_1$. For these we have

$$D_S^2 = (\hat{\beta} - \hat{\beta}^{(1)})'S(\hat{\beta} - \hat{\beta}^{(1)}) = r_2^{(1)\prime}Dr_2^{(1)} = r_2'(I - D)^{-1}D(I - D)^{-1}r_2 \tag{2.26}$$

and

$$D_{S_1}^2 = (\hat{\beta} - \hat{\beta}^{(1)})'S_1(\hat{\beta} - \hat{\beta}^{(1)}) = r_2'Ur_2 = r_2^{(1)\prime}(I + U)^{-1}U(I + U)^{-1}r_2^{(1)} . \tag{2.27}$$

When divided by ps^2, where s^2 is the residual mean square from the full data model, D_S^2 represents a generalization to a set of n_2 observations of Cook's (1977a) distance measure for detecting influential observations. Another expression for D_S^2 is, since $S = X'X$,

$$D_S^2 = (\hat{Y} - \hat{Y}^{(1)})'(\hat{Y} - \hat{Y}^{(1)}) \ , \quad \hat{Y} = X\hat{\beta} \ , \quad \hat{Y}^{(1)} = X\hat{\beta}^{(1)} \ . \tag{2.28}$$

We can use (2.19) to relate the sum of squared residuals for the submodel with the sum of squared residuals for the complete data model.

<u>Proposition 4</u>:

$$\begin{aligned} r'r &= r_1^{(1)\prime}r_1^{(1)} + r_2^{(1)\prime}(I + U)^{-1}r_2^{(1)} \\ &= r_1^{(1)\prime}r_1^{(1)} + r_2'(I + U)r_2 \ . \end{aligned} \tag{2.29}$$

Proof:

Using the orthogonality relationship $X_1'r_1^{(1)} = 0$, from (2.19) we have

$$r_1'r_1 = r_1^{(1)\prime}r_1^{(1)} + r_2'X_2S_1^{-1}X_1'X_1S_1^{-1}X_2'r_2 = r_1^{(1)\prime}r_1^{(1)} + r_2'Ur_2 \ .$$

Since $r'r = r_1'r_1 + r_2'r_2$ and $r_2^{(1)} = (I + U)^{-1}r_2$, the result is proved. □

Beckman and Trussell (1974) give a particular case ($n_2 = 1$) of (2.29).

Corollary:

Let $s_1^2 = r_1^{(1)\prime}r_1^{(1)}/(n_1 - p)$ and $s^2 = r'r/(n - p)$ be the residual mean squares for the deleted observation submodel and the complete data model, respectively. Then

$$s^2 = [(n_1 - p)s_1^2 + r_2^{(1)\prime}(I + U)^{-1}r_2^{(1)}]/(n - p) \qquad (2.30)$$

$$s_1^2 = [(n - p)s^2 - r_2'(I - D)^{-1}r_2]/(n_1 - p) \ . \qquad (2.31)$$

Consider now what may be termed the multiple outlier problem -- testing whether suspected observations (X_2, Y_2) fit the same linear model as describes (X_1, Y_1). Using the residuals $r_2^{(1)} = Y_2 - X_2\hat{\beta}^{(1)}$ based on the "good" model, a natural statistic to use is (since $\text{Cov}[r_2^{(1)}] = (I + U)\sigma^2$)

$$T_1^2 = [n_2^{-1}r_2^{(1)\prime}(I + U)^{-1}r_2^{(1)}]/s_1^2 \ . \qquad (2.32)$$

Under the null hypothesis and assuming normality, the numerator of T_1^2 is distributed as $\sigma^2\chi^2(n_2)/n_2$, independently of the denominator which is $\sigma^2\chi^2(n_1 - p)/(n_1 - p)$. Hence T_1^2 is distributed as $F(n_2, n_1 - p)$. Of course, if the rows in (X_2, Y_2) were chosen after inspection of the data, the use of the F-distribution would not be correct.

If one prefers to work with the residuals $r_2 = Y_2 - X_2\hat{\beta}$ from the complete data model, the natural "studentized" statistic would be (since $\text{Cov}[r_2] = (I - D)\sigma^2$)

$$T^2 = [n_2^{-1} r_2'(I - D)^{-1} r_2]/s^2 \ . \tag{2.33}$$

But by (2.17) or (2.18), $r_2'(I - D)^{-1} r_2 = r_2^{(1)\prime}(I + U)^{-1} r_2^{(1)}$. Hence, by (2.31),

$$\begin{aligned} T_1^2 &= n_2^{-1} r_2'(I - D)^{-1} r_2/[(n_1 - p)^{-1}((n - p)s^2 - r_2'(I - D)^{-1} r_2)] \\ &= (n_1 - p)T^2/(n - p - n_2 T^2) \ . \end{aligned} \tag{2.34}$$

Thus T^2 and T_1^2 are one-to-one monotonic functions of each other and are thus equivalent test statistics. Moreover (2.34) shows how to transform T^2 to an F-statistic. Equation (2.34) directly generalizes a result of Beckman and Trussell (1974) for the single outlier problem.

3. Identities related to adding and deleting variables

Although the identities involved in adding or removing variables (columns) from the linear model (1.1) are better known than those associated with adding or deleting observations, it seems useful to include some of them here. The independent variable matrix is now considered to be partitioned as

$$X = [\widetilde{X}_1, \widetilde{X}_2], \ \widetilde{X}_j \text{ n by } p_j, \ j = 1,2 \ ; \quad p_1 + p_2 = p \ . \tag{3.1}$$

Correspondingly the matrix of sums of squares and cross-products is partitioned as

$$S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix}, \quad S_{ij} = \widetilde{X}_i'\widetilde{X}_j \ , \quad i,j, = 1,2, \tag{3.2}$$

and its inverse is similarly partitioned as

$$S^{-1} = \begin{bmatrix} S^{11} & S^{12} \\ S^{21} & S^{22} \end{bmatrix}. \tag{3.3}$$

Throughout this section we assume that $\text{rank}(\widetilde{X}_1) = p_1$, $\text{rank}(X) = p$.

A fundamental identity is the well known form for the inverse of a partitioned matrix (see, e.g., Graybill (1969), Theorem 8.2.5, p. 164).

Proposition 5:

$$S^{-1} = \begin{bmatrix} (S_{11} - S_{12}S_{22}^{-1}S_{21})^{-1} & -S^{11}S_{12}S_{22}^{-1} \\ -S_{22}^{-1}S_{21}S^{11} & S_{22}^{-1} + S_{22}^{-1}S_{21}S^{11}S_{12}S_{22}^{-1} \end{bmatrix} \tag{3.4}$$

and

$$S^{-1} = \begin{bmatrix} S_{11}^{-1} + S_{11}^{-1}S_{12}S^{22}S_{21}S_{11}^{-1} & -S_{11}^{-1}S_{12}S^{22} \\ -S^{22}S_{21}S_{11}^{-1} & (S_{22} - S_{21}S_{11}^{-1}S_{12})^{-1} \end{bmatrix} \tag{3.5}$$

Proof: Multiplying the r.h.s. of (3.4) and (3.5) by S yields I_p. □

Proposition 5 is valid for arbitrary non-singular partitioned matrices when the required inverses exist. Moreover in the non-full rank case one can substitute generalized inverses for inverses throughout, provided $\mathcal{M}(S_{12}) \subseteq \mathcal{M}(S_{11})$ and $\mathcal{M}(S_{21}') \subseteq \mathcal{M}(S_{11}')$ (for 3.4)), or $\mathcal{M}(S_{12}') \subseteq \mathcal{M}(S_{22}')$ and $\mathcal{M}(S_{21}) \subseteq \mathcal{M}(S_{22})$ (for (3.5)), where $\mathcal{M}(A)$ is the column space of a matrix A (see Rohde 1965 for the case when S is a cross-product matrix, as is the case here).

It is enlightening to note that (3.4) provides an alternate path to (2.3) and all the identities in Section 2. Deletion of the last n_2 rows of (1.1) can be shown to be equivalent by adding $p_2 = n_2$ columns of the form $\tilde{X}_2 = [0, I_{n_2}]'$ to (1.1). The least squares coefficient vector for the new model is $[\hat{\beta}^{(1)\prime}, r^{(1)\prime}]'$, and the test statistic T_1^2 is simply the usual partial F for testing whether a set of regression coefficients is zero.

As is well known, $(S^{22})^{-1}$ is the cross-product matrix of the residuals of $\tilde{X}_2$ regressed on $\tilde{X}_1$. This is more formally expressed as follows.

Proposition 6:

$$(S^{22})^{-1} = (\tilde{X}_2 - \tilde{X}_1S_{11}^{-1}S_{12})'(\tilde{X}_2 - \tilde{X}_1S_{11}^{-1}S_{12}) \; . \tag{3.6}$$

Proof: From (3.5), $(S^{22})^{-1} = S_{22} - S_{21}S_{11}^{-1}S_{12}$. It is easily checked that the r.h.s. of (3.6) reduces to the same expression. □

Proposition 5 leads to the following interesting identity:

Proposition 7:

$$\begin{bmatrix} S_{11}^{-1} & 0 \\ 0 & 0 \end{bmatrix} S + S^{-1} \begin{bmatrix} 0 & 0 \\ 0 & (S^{22})^{-1} \end{bmatrix} = I_p . \tag{3.7}$$

Proof: The l.h.s. of (3.7) is

$$\begin{bmatrix} I_{p_1} & S_{11}^{-1}S_{12} \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & S^{12}(S^{22})^{-1} \\ 0 & I_{p_2} \end{bmatrix} = I_p ,$$

since $S^{12} = -S_{11}^{-1}S_{12}S^{22}$ by (3.5). □

Equation (3.7) provides the basis for the following representations of S and S^{-1}:

Proposition 8:

$$S^{-1} = \begin{bmatrix} -S_{11}^{-1}S_{12} \\ I_{p_2} \end{bmatrix} S^{22} [-S_{21}S_{11}^{-1} , I_{p_2}] + \begin{bmatrix} S_{11}^{-1} & 0 \\ 0 & 0 \end{bmatrix} \tag{3.8}$$

$$S = \begin{bmatrix} I_{p_2} \\ S_{21}S_{11}^{-1} \end{bmatrix} S_{11} [I_{p_1} , S_{11}^{-1}S_{12}] + \begin{bmatrix} 0 & 0 \\ 0 & (S^{22})^{-1} \end{bmatrix} \tag{3.9}$$

Proof: Multiply (3.7) on the left by S or on the right by S^{-1} and simplify. □

From Proposition 8 we easily obtain two important and familiar partitions.

Proposition 9:

$$XS^{-1}X' = \tilde{X}_1S_{11}^{-1}\tilde{X}_1' + (\tilde{X}_2 - \tilde{X}_1S_{11}^{-1}S_{12})S^{22}(\tilde{X}_2 - \tilde{X}_1S_{11}^{-1}S_{12})' \tag{3.10}$$

and, if $\alpha = [\alpha_1', \alpha_2']'$ is an arbitrary p by 1 vector,

$$\alpha' S\alpha = \alpha_1^{*\prime} S_{11}\alpha_1^* + \alpha_2'(S^{22})^{-1}\alpha_2 \ , \tag{3.11}$$

where

$$\alpha_1^* = [I_{p_1}\ , \ S_{11}^{-1}S_{12}]\alpha = \alpha_1 + S_{11}^{-1}S_{12}\alpha_2 \ . \tag{3.12}$$

<u>Proof</u>: These follow directly by substitution of the expressions for S^{-1} and S given by (3.8) and (3.9). □

The l.h.s. of (3.10) is the matrix of the orthogonal projection on the column space $\mathcal{M}(X)$ of X, while the r.h.s. is the direct sum of the matrices of the orthogonal projections on $\mathcal{M}(\widetilde{X}_1)$ and on $\mathcal{M}(\widetilde{X}_2 - \widetilde{X}_1 S_{11}^{-1}S_{12})$, the orthogonal complement of $\mathcal{M}(\widetilde{X}_1)$ in $\mathcal{M}(X)$. Note also that if $\alpha = \hat{\beta} = S^{-1}X'Y$ is the vector of complete data least squares estimates, then

$$\hat{\beta}_1^* = [I\ , \ S_{11}^{-1}S_{12}]\hat{\beta} = S_{11}^{-1}\widetilde{X}_1'Y \tag{3.13}$$

is the vector of least squares estimates in the submodel $(\widetilde{X}_1, Y)$. Thus we recognize (3.11) as the usual analysis of variance partition of the regression sum of squares into the regression sum of squares for $(\widetilde{X}_1, Y)$ and the "partial" regression sum of squares due to including $\widetilde{X}_2$ after $\widetilde{X}_1$. Conversely to (3.13) we have

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} \hat{\beta}_1^* - S_{11}^{-1}S_{12}\hat{\beta}_2 \\ \hat{\beta}_2 \end{bmatrix} , \quad \hat{\beta}_2 = S^{22}(\widetilde{X}_2 - \widetilde{X}_1 S_{11}^{-1}S_{12})'Y \ , \tag{3.14}$$

thus providing an updating formula for the estimates when variables are added to a submodel. The residual vector $\tilde{r} = Y - \widetilde{X}_1\hat{\beta}_1^*$ for the submodel can be updated to the residual vector $r = Y - X\hat{\beta}$ for the complete data model as

$$r = \tilde{r} - (\widetilde{X}_2 - \widetilde{X}_1 S_{11}^{-1}S_{12})\hat{\beta}_2 \ , \tag{3.15}$$

and the residual sum of squares as

$$r'r = r'r - \hat{\beta}_2'(S^{22})^{-1}\hat{\beta}_2 = \tilde{r}'\tilde{r} - Y'(\widetilde{X}_2 - \widetilde{X}_1 S_{11}^{-1}S_{12})S^{22}(\widetilde{X}_2 - \widetilde{X}_1 S_{11}^{-1}S_{12})'Y \ . \tag{3.16}$$

4. Identities related to adding or deleting both rows and columns

Because adding or deleting both rows and columns of the complete data model (1.1) can be accomplished in stages, first dealing with rows and then with columns or vice versa, there is little that can usefully be included here.

One remark somewhat expands the scope for application of the identities in Section 2. In several of the results in Section 3, the expression $\tilde{X}_2-\tilde{X}_1S_{11}^{-1}S_{12}$ appears. This represents the least squares residual vector in the regression of $\tilde{X}_2$ on $\tilde{X}_1$. Hence, relationships between residuals including and excluding the least n_2 rows are applicable. For instance, let X now be partitioned as

$$X = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix}, \quad X_{ij}\ n_i \text{ by } p_j,\ i,j = 1,2 \ , \tag{4.1}$$

and define

$$S_{ijk} = X'_{ki}X_{kj} \ , \quad i,j,k = 1,2 \ . \tag{4.2}$$

Then, by (2.17),

$$X_{22} - X_{21}S_{11}^{-1}S_{12} = (I - U_1)^{-1}(X_{22} - X_{21}S_{111}^{-1}S_{121}) \ , \tag{4.3}$$

where

$$U_1 = X_{21}S_{111}^{-1}X'_{21} \ . \tag{4.4}$$

Similarly, by (2.19)

$$X_{12}-X_{11}S_{11}^{-1}S_{12} = X_{12}-X_{11}S_{111}^{-1}S_{12}-X_{11}S_{111}X'_{21}(I+U_1)^{-1}(X_{22}-X_{21}S_{111}^{-1}S_{121}) \ . \tag{4.5}$$

One can use the partition (3.11) to find an expression for the numerator in Cook's distance measure associated with a subset of the coefficients, say the last p_2 of them. By Cook (1977a), the required quantity is

$$\tilde{D}^2 = (\hat{\beta}_2 - \hat{\beta}_2^{(1)})'(S^{22})^{-1}(\hat{\beta}_2 - \hat{\beta}_2^{(1)}) \ . \qquad (4.6)$$

Note that $\tilde{D}^2$ is of the form D_Q^2 with Q = block diag $[0, (S^{22})^{-1}]$. But by (2.26) and (3.11) we have

$$\tilde{D}^2 = D_S^2 - (\hat{\beta}_2 - \hat{\beta}_2^{(1)})^{*\prime} S_{11}(\hat{\beta}_2 - \hat{\beta}_2^{(1)})^* ,$$

where $(\hat{\beta}_2 - \hat{\beta}_2^{(1)})^*$ is defined by (3.12). Now by (2.22), $\hat{\beta} - \hat{\beta}^{(1)}$ is the coefficient vector in the least squares regression of $[0, r_2^{(1)\prime}]'$ on X. Hence, by (3.13)

$$(\hat{\beta}_2 - \hat{\beta}^{(1)})^* = S_{11}^{-1} X_1' \begin{bmatrix} 0 \\ r_2^{(1)} \end{bmatrix} = S_{11}^{-1} X_{21}' r_2^{(1)} \ .$$

Thus by (2.26)

$$\tilde{D}^2 = D_S^2 - r_2^{(1)\prime} X_{21} S_{11}^{-1} X_{21}' r_2^{(1)} = r_2^{(1)\prime}(D - D_1) r_2^{(1)},\ D_1 = X_{21} S^{-1} X_{21}' \ . \qquad (4.7)$$

Note that $\sigma^2 D = \sigma^2 X_2 S^{-1} X_2' = \mathrm{Cov}[\hat{Y}_2]$ for the complete data model and $\sigma^2 D_1 = \mathrm{Cov}[\hat{Y}_2]$ in the submodel $(\tilde{X}_1, Y)$. This generalizes a result of Cook (1977b) for the case of a single residual.

An alternative form for $\tilde{D}^2$, derivable using Proposition 6, is

$$\tilde{D}^2 = (\hat{Y} - \hat{Y}^{(1)})'(I - \tilde{X}_1 S_{11}^{-1} \tilde{X}_1')(\hat{Y} - \hat{Y}^{(1)}), \qquad (4.8)$$

expressing $\tilde{D}^2$ as the squared length of the projection of $\hat{Y} - \hat{Y}^{(1)}$ on the space orthogonal to $\mathcal{M}(\tilde{X}_1)$.

5. Acknowledgement

The author would like to express his appreciation to Professor R. Dennis Cook for many discussions on the results in this paper and for his encouragement that it should be published.

References

Beckman, R.J. and Trussell, H.J. (1974), "The Distribution of an Arbitrary Studentized Residual and the Effects of Updating in Multiple Regression," Journal of the American Statistical Association, 69, 199-201.

Chambers, John M. (1971), "Regression Updating," Journal of the American Statistical Association, 66, 744-748.

Cook, R. Dennis (1977a), "Detection of Influential Observations in Linear Regression," Technometrics, 19, 15-18.

---- (1977b), "Isolating the Effects of Influential Observations," Technical Report #283, Department of Applied Statistics, University of Minnesota, Saint Paul.

Dempster, A.P. (1969), Elements of Continuous Multivariate Analysis, Reading, Massachusetts: Addison-Wesley Publishing Company.

Graybill, Franklin A. (1969), Introduction to Matrices with Application to Statistics, Belmont, California: Wadsworth Publishing Company.

Miller, Rupert G., Jr. (1974), "An unbalanced jackknife," The Annals of Statistics, 2, 880-891.

Rao, C. Radhakrishna and Mitra, Sujit Kumar (1971), Generalized Inverse of Matrices and its Applications, New York: John Wiley & Sons.

Rohde, Charles A. (1965), "Generalized Inverse of Partitioned Matrices," Journal of the Society for Industrial and Applied Mathematics, 12, 1033-1035.