Title
Inclusion and exclusion of data or parameters in the general linear model

Journal
Statistics & Probability Letters, 77(12)

ISSN
0167-7152

Authors
Jammalamadaka, S Rao
Sengupta, D

Publication Date
2007-07-01

DOI
10.1016/j.spl.2007.03.008

Peer reviewed

Available online at www.sciencedirect.com
ScienceDirect
Statistics & Probability Letters 77 (2007) 1235–1247
STATISTICS & PROBABILITY LETTERS
www.elsevier.com/locate/stapro

Inclusion and exclusion of data or parameters in the general linear model

S. Rao Jammalamadaka[a,*], D. Sengupta[b]

[a]Department of Statistics and Applied Probability, University of California, Santa Barbara, CA 93106, USA
[b]Applied Statistics Unit, Indian Statistical Institute, Kolkatta 700108, India

Available online 16 March 2007

Abstract

This paper revisits the topic of how linear functions of observations having zero expectation, play an important role in our statistical understanding of the effect of addition or deletion of a set of observations in the general linear model. The effect of adding or dropping a group of parameters is also explained well in this manner. Several sets of update equations were derived by previous researchers in various special cases of the general set-up that we consider here. The results derived here bring out the common underlying principles of these update equations and help integrate these ideas. These results also provide further insights into recursive residuals, design of experiments, deletion diagnostics and selection of subset models.

Keywords: Linear zero functions; Singular linear model; Updating equations; Diagnostics; Subset selection; Recursive residuals

1. Introduction and preliminaries

Consider the general linear model

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

where the error vector ε has expectation $E(\varepsilon) = \mathbf{0}$ and dispersion matrix $D(\varepsilon) = \sigma^2 V$, V being a known matrix. This article reviews the changes that result in the computation of the best linear unbiased estimators (BLUEs) of estimable parametric functions, the variance–covariance matrices of such estimators, the error sum of squares, and likelihood ratio tests for testable linear hypotheses (under normal theory), when some observations are added or deleted, as well as when some explanatory variables are added or deleted. It attempts to summarize and update some of the material contained in Sengupta and Jammalamadaka (2003).

There has been extensive work done in this area by various authors. Placket (1950) gave update expressions for the least squares estimates in a linear model with $V = I$ and full rank X, when a single observation is added. Subsequent researchers (see Mitra and Bhimasankaram, 1971; McGilchrist and Sandland, 1979; Haslett, 1985; Bhimasankaram et al., 1995; Bhimasankaram and Jammalamadaka, 1994a, b; Jammalamadaka

*Corresponding author. Tel.: +1 805 8933119; fax: +1 805 8932334.
E-mail address: rao@pstat.ucsb.edu (S.R. Jammalamadaka).

0167-7152/$ - see front matter
doi:10.1016/j.spl.2007.03.008

and Sengupta, 1999) sought to extend this work to data deletion, variable inclusion/exclusion, heteroscedastic and correlated model errors, rank-deficient $\boldsymbol{V}$, rank-deficient $\boldsymbol{X}$, inclusion/exclusion of multiple observations or variables, and so on. Another stream of research focussed on numerically stable methods of recursive estimation in the linear model (see, e.g., Chambers, 1975; Gragg et al., 1979; Kourouklis and Paige, 1981; Farebrother, 1988).

Our goal here is to discuss, in the most general linear model set-up, expressions which provide a good understanding of the problem(s) and allow statistical interpretations, rather than focussing on numerically stable computations. For instance, a good understanding of the update mechanism in the case of additional observations can provide insights into strategies for sequential design, whereas updating for exclusion of observations has implications in deletion diagnostics. The inclusion and exclusion of explanatory variables are useful for comparison of various subset models. These and other applications of the update relations are mentioned in the last section.

Following the work of Sengupta and Jammalamadaka (2003) (hereafter referred to as SJ), we will use linear zero functions (LZF)—linear functions of $\boldsymbol{y}$ having expectation zero—as the main tool in the derivation of the updates. See Section 4.1 of SJ for some basic results pertaining to the LZFs. The LZFs characterize the BLUEs in the linear model: a linear function is the BLUE of its expectation if and only if it is uncorrelated with every LZF. Every LZF is a linear function of $(\boldsymbol{I}-\boldsymbol{P_X})\boldsymbol{y}$, where $\boldsymbol{P_X}=X(X'X)^-X'$ is the orthogonal projection matrix for the column space of $\boldsymbol{X}$ (for any matrix A, we use A^- to denote a generalized inverse of it). Thus, an ordinary linear unbiased estimator can be turned into a BLUE by the removal of its correlation with $(\boldsymbol{I}-\boldsymbol{P_X})\boldsymbol{y}$, by means of the following lemma (Proposition 3.1.2 of SJ).

Lemma 1.1. *Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be random vectors having first and second order moments with $E(\boldsymbol{v})=\boldsymbol{0}$. Then the linear compound $\boldsymbol{u}-\boldsymbol{B}\boldsymbol{v}$ is uncorrelated with $\boldsymbol{v}$ if and only if*

$$\boldsymbol{B}\boldsymbol{v}=Cov(\boldsymbol{u},\boldsymbol{v})[D(\boldsymbol{v})]^-\boldsymbol{v}\quad \textit{with probability } 1.$$

By putting $\boldsymbol{u}=\boldsymbol{y}$ and $\boldsymbol{v}=(\boldsymbol{I}-\boldsymbol{P_X})\boldsymbol{y}$ in the above lemma, one gets the following expression for the BLUE of $\boldsymbol{X\beta}$:

$$\widehat{\boldsymbol{X\beta}}=[\boldsymbol{I}-\boldsymbol{V}(\boldsymbol{I}-\boldsymbol{P_X})\{(\boldsymbol{I}-\boldsymbol{P_X})\boldsymbol{V}(\boldsymbol{I}-\boldsymbol{P_X})\}^-(\boldsymbol{I}-\boldsymbol{P_X})]\boldsymbol{y}.$$

The residual vector is $\boldsymbol{e}=\boldsymbol{y}-\widehat{\boldsymbol{X\beta}}$. It can be shown that

$$D(\widehat{\boldsymbol{X\beta}})=\sigma^2[\boldsymbol{V}-\boldsymbol{V}(\boldsymbol{I}-\boldsymbol{P_X})\{(\boldsymbol{I}-\boldsymbol{P_X})\boldsymbol{V}(\boldsymbol{I}-\boldsymbol{P_X})\}^-(\boldsymbol{I}-\boldsymbol{P_X})\boldsymbol{V}],$$
$$D(\boldsymbol{e})=\sigma^2\boldsymbol{V}(\boldsymbol{I}-\boldsymbol{P_X})\{(\boldsymbol{I}-\boldsymbol{P_X})\boldsymbol{V}(\boldsymbol{I}-\boldsymbol{P_X})\}^-(\boldsymbol{I}-\boldsymbol{P_X})\boldsymbol{V}.$$

The column spaces of these matrices are $\mathscr{C}(D(\widehat{\boldsymbol{X\beta}}))=\mathscr{C}(\boldsymbol{X})\cap\mathscr{C}(\boldsymbol{V})$ and $\mathscr{C}(D(\boldsymbol{e}))=\mathscr{C}(\boldsymbol{V}(\boldsymbol{I}-\boldsymbol{P_X}))$.

The set of elements of the vector $(\boldsymbol{I}-\boldsymbol{P_X})\boldsymbol{y}$ may be called a *generating set* of all LZFs in the linear model $(\boldsymbol{y},\boldsymbol{X\beta},\sigma^2\boldsymbol{V})$ in the sense that all LZFs of this model are linear functions of these LZFs. A *standardized basis set* of LZFs is a generating set of LZFs having uncorrelated elements with variance σ^2. If $\boldsymbol{z}$ is any vector whose elements constitute a standardized basis set of LZFs of the model $(\boldsymbol{y},\boldsymbol{X\beta},\sigma^2\boldsymbol{V})$, then it can be shown that $\boldsymbol{z}$ has $\rho(\boldsymbol{V}:\boldsymbol{X})-\rho(\boldsymbol{X})$ elements (where $\rho(\cdot)$ indicates the rank of the matrix concerned), and that the value of $\boldsymbol{z}'\boldsymbol{z}$ does not depend on the choice of the standardized basis set. Further, $\boldsymbol{z}'\boldsymbol{z}$ happens to be the residual sum of squares denoted by R_0^2, i.e., the minimized value of $(\boldsymbol{y}-\boldsymbol{X\beta})'\boldsymbol{V}^-(\boldsymbol{y}-\boldsymbol{X\beta})$ subject to the restriction $\boldsymbol{y}-\boldsymbol{X\beta}\in\mathscr{C}(\boldsymbol{V})$. On the other hand, if $\boldsymbol{z}$ is a vector whose elements constitute merely a generating set of LZFs of the model $(\boldsymbol{y},\boldsymbol{X\beta},\sigma^2\boldsymbol{V})$, then $R_0^2=\boldsymbol{z}'[D(\boldsymbol{z})]^-\boldsymbol{z}$, and the rank of $D(\boldsymbol{z})$ is always $\rho(\boldsymbol{V}:\boldsymbol{X})-\rho(\boldsymbol{X})$, which is the residual degrees of freedom. In particular, one can use the generating set $\boldsymbol{z}=(\boldsymbol{I}-\boldsymbol{P_X})\boldsymbol{y}$.

In the presence of the vector restriction $\boldsymbol{A\beta}=\boldsymbol{\xi}$ where $\boldsymbol{A\beta}$ is an estimable function, it can be shown that the vector $\widehat{\boldsymbol{A\beta}}-\boldsymbol{\xi}$ (where $\widehat{\boldsymbol{A\beta}}$ is the BLUE of $\boldsymbol{A\beta}$) is a vector LZF in the restricted model which is uncorrelated with $(\boldsymbol{I}-\boldsymbol{P_X})\boldsymbol{y}$ and which, together with the latter set of LZFs, constitute a generating set of LZFs in the restricted model. It follows that the residual sum of squares under the restriction or hypothesis $\boldsymbol{A\beta}=\boldsymbol{\xi}$, is given by

$$R_H^2 = R_0^2 + (\boldsymbol{A}\widehat{\boldsymbol{\beta}} - \boldsymbol{\xi})'[\sigma^{-2}D(\boldsymbol{A}\widehat{\boldsymbol{\beta}})]^{-}(\boldsymbol{A}\widehat{\boldsymbol{\beta}} - \boldsymbol{\xi}).$$

The corresponding number of degrees of freedom is $\rho(\boldsymbol{V} : \boldsymbol{X}) - \rho(\boldsymbol{X}) + \rho(D(\boldsymbol{A}\widehat{\boldsymbol{\beta}}))$.

In the four sections that follow, we deal with the inclusion/exclusion of data and of parameters along with the updating equations for the quantities $\widehat{\boldsymbol{X\beta}}$, $D(\widehat{\boldsymbol{X\beta}})$, R_0^2, R_H^2 and the degrees of freedom associated with these two sums of squares, and conclude with some remarks in the final section.

2. Inclusion of observations

Let us denote the linear model with n observations by

$$\mathscr{M}_n = (\boldsymbol{y}_n, \boldsymbol{X}_n\boldsymbol{\beta}, \sigma^2\boldsymbol{V}_n).$$

In this section we track the transition from $\mathscr{M}_m = (\boldsymbol{y}_m, \boldsymbol{X}_m\boldsymbol{\beta}, \sigma^2\boldsymbol{V}_m)$ to $\mathscr{M}_n$ for $m < n$. We refer to $\mathscr{M}_m$ as the 'initial' model and $\mathscr{M}_n$ as the 'augmented' model. Note that each LZF in the initial model $\mathscr{M}_m$ is also an LZF in the augmented model $\mathscr{M}_n$. From the discussion of the foregoing section, the number of nontrivial and uncorrelated LZFs exclusive to the augmented model, which are uncorrelated with the LZFs common to both the models, turns out to be $[\rho(\boldsymbol{X}_n : \boldsymbol{V}_n) - \rho(\boldsymbol{X}_n)] - [\rho(\boldsymbol{X}_m : \boldsymbol{V}_m) - \rho(\boldsymbol{X}_m)]$. The clue to the update relationships lies in the identification of these LZFs.

2.1. Linear zero functions gained

Let $\boldsymbol{y}_n$, $\boldsymbol{X}_n$ and $\boldsymbol{V}_n$ be partitioned as shown below

$$\boldsymbol{y}_n = \begin{pmatrix} \boldsymbol{y}_m \\ \boldsymbol{y}_l \end{pmatrix}, \quad \boldsymbol{X}_n = \begin{pmatrix} \boldsymbol{X}_m \\ \boldsymbol{X}_l \end{pmatrix}, \quad \boldsymbol{V}_n = \begin{pmatrix} \boldsymbol{V}_m & \boldsymbol{V}_{ml} \\ \boldsymbol{V}_{lm} & \boldsymbol{V}_l \end{pmatrix}, \tag{2}$$

where $l = n - m$. Let $l_* = \rho(\boldsymbol{X}_n : \boldsymbol{V}_n) - \rho(\boldsymbol{X}_m : \boldsymbol{V}_m)$. Note that

$$0 \leqslant \rho(\boldsymbol{X}_n) - \rho(\boldsymbol{X}_m) \leqslant l_* \leqslant l.$$

The integer l_* coincides with l when $\boldsymbol{V}_n$ is nonsingular. Four cases can arise

(a) $0 < \rho(\boldsymbol{X}_n) - \rho(\boldsymbol{X}_m) = l_*$,
(b) $0 = \rho(\boldsymbol{X}_n) - \rho(\boldsymbol{X}_m) < l_*$,
(c) $0 = \rho(\boldsymbol{X}_n) - \rho(\boldsymbol{X}_m) = l_*$,
(d) $0 < \rho(\boldsymbol{X}_n) - \rho(\boldsymbol{X}_m) < l_*$.

Case (a) implies that there are some additional estimable linear parametric functions (LPFs) in the augmented model, but no new LZF. In case (b), $\mathscr{C}(\boldsymbol{X}_l') \subseteq \mathscr{C}(\boldsymbol{X}_m')$, which means that there are some additional LZFs in the augmented model but no new estimable LPF. Case (c) corresponds to no new LZF or estimable LPF. This case can only arise when $\boldsymbol{V}_n$ is singular. Case (d) indicates that there are some additional LZFs as well as additional estimable LPFs in the augmented model. This case can arise only if $l > 1$.

There is no new LZF to be identified in cases (a) and (c). In case (d), we can permute the rows of $\boldsymbol{X}_l$ in such a way that each of the top few rows, when appended successively to $\boldsymbol{X}_m$, increase the rank of the matrix by 1, and the remaining rows belong to the row space of the concatenated matrix. This permuted version of $\boldsymbol{X}_l$ can be partitioned as $(\boldsymbol{X}_{l_1}' : \boldsymbol{X}_{l_2}')'$, where $\boldsymbol{X}_{l_1}$ has full row rank and

$$\rho\begin{pmatrix} \boldsymbol{X}_m \\ \boldsymbol{X}_{l_1} \end{pmatrix} - \rho(\boldsymbol{X}_m) = \rho(\boldsymbol{X}_n) - \rho(\boldsymbol{X}_m).$$

The elements of $\boldsymbol{y}_l$, $\boldsymbol{V}_{ml}$ and $\boldsymbol{V}_l$ can also be permuted accordingly. Thus, the inclusion of the l observations can be viewed as a two-step process: the inclusion of the first set of observations entails additional estimable LPFs but no new LZF, as in case (a), while the inclusion of the remaining observations result in additional LZFs but no new estimable LPF, as in case (b). Thus, it is enough to identify the set of new LZFs in the augmented model in case (b), which we do through the next theorem.

Theorem 2.1. *In the above set-up, let $l_*>0$ and $\mathscr{C}(X_l') \subseteq \mathscr{C}(X_m')$. Then a vector of LZFs of the model $\mathscr{M}_n$ that is uncorrelated with all the LZFs of $\mathscr{M}_m$ is given by*

$$w_l = y_l - X_l\widehat{\beta}_m - V_{ml}'V_m^-(y_m - X_m\widehat{\beta}_m). \tag{3}$$

Further, all LZFs of the augmented model are linear combinations of w_l and the LZFs of the initial model.

Proof. It is easy to see that $y_l - X_l\widehat{\beta}_m$ is indeed an LZF in the augmented model. The expression for w_l is obtained by making it uncorrelated with $(I_m - P_{X_m})y_m$ as per Theorem 1.1, and simplifying it.

We shall prove the second part of the theorem by showing that there is no LZF of the augmented model which is uncorrelated with w_l and the LZFs of the initial model. Suppose, for contradiction, that $u'(I - P_{X_n})y_n$ is such an LZF. Consequently, it is uncorrelated with $(I - P_{X_m})y_m$ and $(y_l - X_l\widehat{\beta}_m)$. Therefore,

$$(I - P_{X_m})(V_m : V_{ml})(I - P_{X_n})u = 0,$$

$$(V_{lm} : V_l)(I - P_{X_n})u - X_lX_m^-(V_m : V_{ml})(I - P_{X_n})u = 0.$$

The first condition is equivalent to $(V_m : V_{ml})(I - P_{X_n})u \in \mathscr{C}(X_m)$. It follows from this and the second condition that

$$\begin{pmatrix} X_m \\ X_l \end{pmatrix} X_m^-(V_m : V_{ml})(I - P_{X_n})u = \begin{pmatrix} V_m & V_{ml} \\ V_{lm} & V_l \end{pmatrix}(I - P_{X_n})u,$$

that is, $V(I - P_{X_n})u \in \mathscr{C}(X_n)$. This implies that $u'(I - P_{X_n})y_n$ is a trivial LZF with zero variance. □

There is no unique choice of the LZF with the properties stated in Theorem 2.1. Any linear function of w_l having the same rank of the dispersion matrix would suffice. However, the expression in (3) is invariant under the choice of the g-inverse of V_m.

A standardized basis set of LZFs in the augmented model has l_* extra elements, in comparison with a corresponding set for the initial model. Since all the LZFs of the augmented model that are uncorrelated with those of the initial model, are linear functions of w_l, the rank of $D(w_l)$ must be l_*.

The LZF w_l can be written as the prediction error $y_l - \widehat{y}_l$, where $\widehat{y}_l$ is the BLUP of y_l on the basis of the model $\mathscr{M}_m$. In the special case $l = 1$ and $V_n = I$, Brown et al. (1975) calls this quantity the *recursive residual* of the additional observation. Recursive residuals have the attractive property that these are uncorrelated. These are used as diagnostic tools, particularly when there is a natural order among the observations (see Kianifard and Swallow, 1996). McGilchrist and Sandland (1979) extends the recursive residual to the case of any positive definite V_n, while Haslett (1985) extends it to the case of multiple observations ($l \geqslant 1$). In the case of possibly singular V_n, the vector w_l is a recursive residual for y_l. Jammalamadaka and Sengupta (1999) termed a scaled version of w_l as the recursive group residual of y_l.

2.2. Update equations

Consider case (b)—the main case of interest for data augmentation mentioned in Section 2.1. We now use the result of Theorem 2.1 to update various statistics.

Theorem 2.2. *Under the set-up of Section 2.1, let $\mathscr{C}(X_l') \subseteq \mathscr{C}(X_m')$ and let $l_* = \rho(X_n : V_n) - \rho(X_m : V_m) > 0$. Suppose further that $A\beta$ is estimable with $D(A\widehat{\beta}_m)$ not identically zero, and w_l is the recursive residual given in (3). Then*

(a) $X_m\widehat{\beta}_n = X_m\widehat{\beta}_m - Cov(X_m\widehat{\beta}_m, w_l)[D(w_l)]^- w_l$.
(b) $D(X_m\widehat{\beta}_n) = D(X_m\widehat{\beta}_m) - Cov(X_m\widehat{\beta}_m, w_l)[D(w_l)]^- Cov(w_l), (X_m\widehat{\beta}_m)$.
(c) $R^2_{0_n} = R^2_{0_m} + \sigma^2 w_l'[D(w_l)]^- w_l$.

(d) *The change in* R^2_H *corresponding to the hypothesis* $\boldsymbol{A\beta}=\boldsymbol{\xi}$ *is* $R^2_{H_n}=R^2_{H_m}+\sigma^2(\boldsymbol{w}_l-\widehat{\boldsymbol{w}}_l)'[D(\boldsymbol{w}_l)-D(\widehat{\boldsymbol{w}}_l)]^-(\boldsymbol{w}_l-\widehat{\boldsymbol{w}}_l)$, *where*
$\widehat{\boldsymbol{w}}_l=Cov(\boldsymbol{w}_l,\boldsymbol{A}\widehat{\boldsymbol{\beta}}_m)[D(\boldsymbol{A}\widehat{\boldsymbol{\beta}}_m)]^-(\boldsymbol{A}\widehat{\boldsymbol{\beta}}_m-\boldsymbol{\xi})$.

(e) *Inclusion of the l additional observations increases the degrees of freedom of* R^2_0 *and* R^2_H *by* l_* *and* $\rho(D(\boldsymbol{w}_{l*}))$, *respectively*.

Proof. Note that $\boldsymbol{X}_m\widehat{\boldsymbol{\beta}}_m$ is an unbiased estimator of $\boldsymbol{X}_m\boldsymbol{\beta}$ that is already uncorrelated with the LZFs of $\mathscr{M}_m$. By making it uncorrelated with the new LZFs $\boldsymbol{w}_l$ through Theorem 1.1, we have the expression of part (a). Since $\boldsymbol{X}_m\widehat{\boldsymbol{\beta}}_n$ must be uncorrelated with the increment term in part (a), we have

$$D(\boldsymbol{X}_m\widehat{\boldsymbol{\beta}}_m)=D(\boldsymbol{X}_m\widehat{\boldsymbol{\beta}}_n)+Cov(\boldsymbol{X}_m\widehat{\boldsymbol{\beta}}_m,\boldsymbol{w}_l)[D(\boldsymbol{w}_l)]^-Cov(\boldsymbol{w}_l,\boldsymbol{X}_m\widehat{\boldsymbol{\beta}}_m),$$

which is equivalent to the result of part (b). Part (c) follows from the characterization of R^2_0 through a standardized basis set of LZFs. By a similar argument—after $\boldsymbol{w}_l$ is corrected for correlation with the additional LZF $\boldsymbol{A}\widehat{\boldsymbol{\beta}}_m$ as per Lemma 1.1—we get $R^2_{H_n}=R^2_{H_m}+\sigma^2\boldsymbol{w}'_{l*}[D(\boldsymbol{w}_{l*})]^-\boldsymbol{w}_{l*}$, where $\boldsymbol{w}_{l*}=\boldsymbol{w}_l-Cov(\boldsymbol{w}_l,\boldsymbol{A}\widehat{\boldsymbol{\beta}}_m)[D(\boldsymbol{A}\widehat{\boldsymbol{\beta}}_m)]^-(\boldsymbol{A}\widehat{\boldsymbol{\beta}}_m-\boldsymbol{\xi})=\boldsymbol{w}_l-\widehat{\boldsymbol{w}}_l$. It follows that $\boldsymbol{w}_{l*}$ and $\widehat{\boldsymbol{w}}_l$ are uncorrelated, and hence, $D(\boldsymbol{w}_{l*})=D(\boldsymbol{w}_l)-D(\widehat{\boldsymbol{w}}_l)$. Part (d) follows immediately. Part (e) is a consequence of the fact that the additional error degrees of freedom coincide with the number of nontrivial LZFs of the augmented model that are uncorrelated with the old ones as well as among themselves. □

The variances and covariances involved in the update formulae can be computed from the expressions given in Section 1. The explicit algebraic expressions in the general case, given by Pordzik (1992a) and Bhimasankaram et al. (1995) are somewhat complicated.

When a single observation is included ($l=1$), $D(\boldsymbol{w}_l)$ reduces to a scalar. Here, the assumptions of Theorem 2.2 imply that $\rho(D(\boldsymbol{w}_l))$ is equal to 1. The rank of $D(\boldsymbol{w}_{l*})$ must also be equal to 1 (it is zero if and only if $\boldsymbol{w}_l$ is a linear function of the BLUEs of $\mathscr{M}_m$, which is impossible).

If $\boldsymbol{V}_m$ is nonsingular, the unscaled recursive group residual defined in (3) can be written as

$$\boldsymbol{w}_l=\boldsymbol{s}_l-\widehat{\boldsymbol{s}}_l,$$

where

$$\boldsymbol{s}_l=\boldsymbol{y}_l-\boldsymbol{V}'_{ml}\boldsymbol{V}_m^{-1}\boldsymbol{y}_m,$$

$$\widehat{\boldsymbol{s}}_l=\boldsymbol{Z}_l\widehat{\boldsymbol{\beta}}_m\quad\text{and}\quad\boldsymbol{Z}_l=(\boldsymbol{X}_l-\boldsymbol{V}'_{ml}\boldsymbol{V}_m^{-1}\boldsymbol{X}_m).$$

(A similar decomposition is possible even if $\boldsymbol{V}_m$ is singular, but the quantities $\boldsymbol{s}_l$ and $\widehat{\boldsymbol{s}}_l$ are not uniquely defined in such a case.) The quantity $\boldsymbol{s}_l$ is a part of $\boldsymbol{y}_l$ which is uncorrelated with $\boldsymbol{y}_m$. On the other hand, $\widehat{\boldsymbol{s}}_l$ can be interpreted as the BLUP of $\boldsymbol{s}_l$ under the model $(\boldsymbol{y}_m,\boldsymbol{X\beta},\sigma^2\boldsymbol{V}_m)$. Clearly, $Cov(\boldsymbol{s}_l,\widehat{\boldsymbol{s}}_l)=\boldsymbol{0}$. It follows that

$$D(\boldsymbol{w}_l)=D(\boldsymbol{s}_l)+D(\widehat{\boldsymbol{s}}_l),$$

$$Cov(\boldsymbol{X}_m\widehat{\boldsymbol{\beta}}_m,\boldsymbol{w}_l)=-Cov(\boldsymbol{X}_m\widehat{\boldsymbol{\beta}}_m,\widehat{\boldsymbol{s}}_l).$$

If, in addition, $\boldsymbol{X}_m$ has full column rank, then we can work directly with $\widehat{\boldsymbol{\beta}}_m$ (instead of $\boldsymbol{X}_m\widehat{\boldsymbol{\beta}}_m$). Thus, we have the following simplifications:

$$D(\boldsymbol{s}_l)=\sigma^2(\boldsymbol{V}_l-\boldsymbol{V}'_{ml}\boldsymbol{V}_m^-\boldsymbol{V}_{ml}),$$

$$D(\widehat{\boldsymbol{s}}_l)=\sigma^2\boldsymbol{Z}_l(\boldsymbol{X}'_m\boldsymbol{V}_m^{-1}\boldsymbol{X}_m)^{-1}\boldsymbol{Z}'_l,$$

$$Cov(\widehat{\boldsymbol{\beta}}_m,\widehat{\boldsymbol{s}}_l)=\sigma^2(\boldsymbol{X}'_m\boldsymbol{V}_m^{-1}\boldsymbol{X}_m)^{-1}\boldsymbol{Z}'_l=-Cov(\widehat{\boldsymbol{\beta}}_m,\boldsymbol{w}_l),$$

$$\widehat{\boldsymbol{\beta}}_n=\widehat{\boldsymbol{\beta}}_m+Cov(\widehat{\boldsymbol{\beta}}_m,\widehat{\boldsymbol{s}}_l)[D(\boldsymbol{s}_l)+D(\widehat{\boldsymbol{s}}_l)]^-(\boldsymbol{s}_l-\widehat{\boldsymbol{s}}_l),$$

$$D(\widehat{\boldsymbol{\beta}}_n)=D(\widehat{\boldsymbol{\beta}}_m)-Cov(\widehat{\boldsymbol{\beta}}_m,\widehat{\boldsymbol{s}}_l)[D(\boldsymbol{s}_l)+D(\widehat{\boldsymbol{s}}_l)]^-Cov(\widehat{\boldsymbol{\beta}}_m,\widehat{\boldsymbol{s}}_l)',$$

$$R^2_{0_n}=R^2_{0_m}+\sigma^2(\boldsymbol{s}_l-\widehat{\boldsymbol{s}}_l)'[D(\boldsymbol{s}_l)+D(\widehat{\boldsymbol{s}}_l)]^-(\boldsymbol{s}_l-\widehat{\boldsymbol{s}}_l).$$

The above formulae for $\widehat{\boldsymbol{\beta}}_n$, $D(\widehat{\boldsymbol{\beta}}_n)$ and $R^2_{0_n}$ are essentially the same as those given by McGilchrist and Sandland (1979) (for $l=1$) and Haslett (1985) (for $l\geqslant 1$).

When $\boldsymbol{V}_n=\boldsymbol{I}$, we have $\boldsymbol{s}_l=\boldsymbol{y}_l$, $\widehat{\boldsymbol{s}}_l=\boldsymbol{X}_l\widehat{\boldsymbol{\beta}}_m$, $D(\boldsymbol{s}_l)=\sigma^2\boldsymbol{I}$, $D(\widehat{\boldsymbol{s}}_l)=\sigma^2\boldsymbol{X}_l(\boldsymbol{X}'_m\boldsymbol{X}_m)^{-1}\boldsymbol{X}'_l$ and $Cov(\widehat{\boldsymbol{\beta}}_m,\widehat{\boldsymbol{s}}_l)=\sigma^2(\boldsymbol{X}'_m\boldsymbol{X}_m)^{-1}\boldsymbol{X}'_l$. Thus, we have the following simplification of Theorem 2.2.

Corollary 2.1. *Under the above set-up, let* $\boldsymbol{X}'_m$ *have full rank, and* $\boldsymbol{V}_n=\boldsymbol{I}$. *Further, let* $\boldsymbol{w}_l=\boldsymbol{y}_l-\boldsymbol{X}_l\widehat{\boldsymbol{\beta}}_m$, $\boldsymbol{H}=\boldsymbol{X}_l(\boldsymbol{X}'_m\boldsymbol{X}_m)^{-}\boldsymbol{X}'_l$ *and* $\boldsymbol{C}=(\boldsymbol{X}'_m\boldsymbol{X}_m)^{-}\boldsymbol{X}'_l$. *Then*

(a) $\widehat{\boldsymbol{\beta}}_n=\widehat{\boldsymbol{\beta}}_m+\boldsymbol{C}[\boldsymbol{I}+\boldsymbol{H}]^{-1}\boldsymbol{w}_l$.
(b) $D(\widehat{\boldsymbol{\beta}}_n)=D(\widehat{\boldsymbol{\beta}}_m)-\sigma^2\boldsymbol{C}[\boldsymbol{I}+\boldsymbol{H}]^{-1}\boldsymbol{C}'$.
(c) $R^2_{0_n}=R^2_{0_m}+\boldsymbol{w}'_l[\boldsymbol{I}+\boldsymbol{H}]^{-1}\boldsymbol{w}_l$.
(d) *The change in* R^2_H *corresponding to the hypothesis* $\boldsymbol{A\beta}=\boldsymbol{\xi}$ *is* $R^2_{H_n}=R^2_{H_m}+(\boldsymbol{w}_l-\widehat{\boldsymbol{w}}_l)'[\boldsymbol{I}+\boldsymbol{H}-\boldsymbol{C}'\boldsymbol{A}'\boldsymbol{D}_A^{-}\boldsymbol{A}\boldsymbol{C}]^{-1}(\boldsymbol{w}_l-\widehat{\boldsymbol{w}}_l)$, *where*
$\widehat{\boldsymbol{w}}_l=-\boldsymbol{C}'\boldsymbol{A}'\boldsymbol{D}_A^{-}(\boldsymbol{A}\widehat{\boldsymbol{\beta}}_m-\boldsymbol{\xi})$ *and* $\boldsymbol{D}_A=\boldsymbol{A}(\boldsymbol{X}'_m\boldsymbol{X}_m)^{-1}\boldsymbol{A}'$.
(e) *The number of degrees of freedom of both* R^2_0 *and* R^2_H *increases by* l *as a result of the inclusion of the additional observation.*

Update equations like those given in Corollary 2.1 are obtained by Placket (1950) and Mitra and Bhimasankaram (1971).

We now turn to cases (a), (c) and (d) of Section 2.1. In case (c), the additional observations of the augmented model are essentially linear functions of the initial model. Therefore, the LZFs and the BLUEs remain the same in the appended model. There is no change whatsoever in any statistic of interest. It has already been explained that data augmentation in case (d) essentially consists of two steps of augmentation classifiable as cases (a) and (b), respectively. In case (a), there is no additional LZF in the augmented model. Hence, the BLUEs of the LPFs which are estimable in $\mathscr{M}_m$, their dispersions, the error sum of squares and the corresponding degrees of freedom are the same under the two models. The error sum of squares under the restriction $\boldsymbol{A\beta}=\boldsymbol{\xi}$ and the corresponding degrees of freedom also remains the same after data augmentation. However, the additional observations contribute to the estimation of the LPFs that are estimable only under the augmented model, as shown in the next theorem.

Theorem 2.3. *Under the set-up used in this section, let* $\rho(\boldsymbol{X}'_n)-\rho(\boldsymbol{X}'_m)=l_*$. Then

(a) $\boldsymbol{X}_l\widehat{\boldsymbol{\beta}}_n=\boldsymbol{y}_l-\boldsymbol{V}_{lm}\boldsymbol{V}_m^{-}(\boldsymbol{y}_m-\boldsymbol{X}_m\widehat{\boldsymbol{\beta}}_m)$.
(b) $D(\boldsymbol{X}_l\widehat{\boldsymbol{\beta}}_n)=\sigma^2\boldsymbol{V}_l-\boldsymbol{V}_{lm}\boldsymbol{V}_m^{-}D(\boldsymbol{y}_m-\boldsymbol{X}_m\widehat{\boldsymbol{\beta}}_m)\boldsymbol{V}_m^{-}\boldsymbol{V}_{ml}$.

Proof. The LZFs of the augmented and original models coincide. Therefore, the BLUE of $\boldsymbol{X}_l\boldsymbol{\beta}$ is obtained by adjusting $\boldsymbol{y}_l$ for its covariance with the LZFs of the original model. We choose $\boldsymbol{y}_m-\boldsymbol{X}_m\widehat{\boldsymbol{\beta}}_m$ as a representative vector of LZFs. If we write this vector as

$$\begin{aligned}\boldsymbol{y}_m-\boldsymbol{X}_m\widehat{\boldsymbol{\beta}}_m&=\boldsymbol{V}_m(\boldsymbol{I}-\boldsymbol{P}_{X_m})\{(\boldsymbol{I}-\boldsymbol{P}_{X_m})\boldsymbol{V}_m^{-}(\boldsymbol{I}-\boldsymbol{P}_{X_m})\}^{-}(\boldsymbol{I}-\boldsymbol{P}_{X_m})\boldsymbol{y}_m)\\&=\boldsymbol{V}_m\boldsymbol{R}_m\boldsymbol{y}_m,\end{aligned}$$

then the required BLUE is

$$\begin{aligned}\boldsymbol{X}_l\widehat{\boldsymbol{\beta}}_n&=\boldsymbol{y}_l-Cov(\boldsymbol{y}_l,\boldsymbol{y}_m-\boldsymbol{X}_m\widehat{\boldsymbol{\beta}}_m)[D(\boldsymbol{y}_m-\boldsymbol{X}_m\widehat{\boldsymbol{\beta}}_m)]^{-}(\boldsymbol{y}_m-\boldsymbol{X}_m\widehat{\boldsymbol{\beta}}_m)\\&=\boldsymbol{y}_l-Cov(\boldsymbol{y}_l,\boldsymbol{V}_m\boldsymbol{R}_m\boldsymbol{y}_m)[D(\boldsymbol{V}_m\boldsymbol{R}_m\boldsymbol{y}_m)]^{-}(\boldsymbol{V}_m\boldsymbol{R}_m\boldsymbol{y}_m)\\&=\boldsymbol{y}_l-\boldsymbol{V}_{lm}\boldsymbol{V}_m^{-}D(\boldsymbol{V}_m\boldsymbol{R}_m\boldsymbol{y}_m)[D(\boldsymbol{V}_m\boldsymbol{R}_m\boldsymbol{y}_m)]^{-}(\boldsymbol{V}_m\boldsymbol{R}_m\boldsymbol{y}_m)\\&=\boldsymbol{y}_l-\boldsymbol{V}_{lm}\boldsymbol{V}_m^{-}(\boldsymbol{y}_m-\boldsymbol{X}_m\widehat{\boldsymbol{\beta}}_m).\end{aligned}$$

The expression of part (b) follows immediately. □

When $V_{lm} = \mathbf{0}$, it is clear that the fitted value of y_l is equal to its observed value. This may not hold when $V_{lm} \neq \mathbf{0}$. However, the new LZF is the BLUP of ε_l from the original model, which is a function of the LZFs of the initial model and it does not alter the error sum of squares or the degrees of freedom.

3. Exclusion of observations

In this section we track the transition from the model $\mathscr{M}_n = (y_n, X_n\beta, \sigma^2 V_n)$ to the model $\mathscr{M}_m = (y_m, X_m\beta, \sigma^2 V_m)$, where $l = n - m > 0$. We refer to $\mathscr{M}_n$ as the 'initial' model and $\mathscr{M}_m$ as the 'deleted' model.

3.1. Linear zero functions lost

Let us consider once again the four cases described in Section 2.1. No LZF is lost in cases (a) and (c). It also follows from the discussion of that section that we only have to identify the LZFs lost in case (b). Case (d) can be thought of as a two-step exclusion where the steps correspond to cases (b) and (a), respectively. Therefore, we deal mainly with case (b).

In the simple case of data exclusion, the unscaled recursive group residual (w_l) of Section 2.1 cannot be used as a pivot for computations, as it is expressed in terms of the residuals of $\mathscr{M}_m$, which is not available before the data exclusion takes place. We need a modification of w_l which can be used in the present context. Note that w_l can be written as $d_l(\widehat{\beta}_m)$, where

$$d_l(\beta) = y_l - X_l\beta - V_{lm}V_m^-(y_m - X_m\beta),$$

which is the part of the model error of y_l that is uncorrelated with the model error of y_m. SJ show that a vector of LZFs of the model $\mathscr{M}_n$ that is uncorrelated with all the LZFs of $\mathscr{M}_m$ is given by

$$r_l = d_l(\widehat{\beta}_n) = y_l - X_l\widehat{\beta}_n - V_{lm}V_m^-(y_m - X_m\widehat{\beta}_n). \tag{4}$$

Further, if $\rho(V_n) = n$, then there is no nontrivial LZF of $\mathscr{M}_n$ which is uncorrelated with r_l and the LZFs of $\mathscr{M}_m$.

Note that the condition $\mathscr{C}(X_l') \subseteq \mathscr{C}(X_m')$ was *not* needed in the proof of the above result. Thus, it covers case (d) of Section 2.1, i.e., the case where some LZFs and estimable LPFs are lost due to data exclusion.

It can be shown that whenever $\mathscr{C}(X_l') \subseteq \mathscr{C}(X_m')$ and V_n is nonsingular, r_l and w_l are linearly transformed versions of one another. When V_n is singular, w_l may not be a function of r_l. In particular, r_l may even have zero dispersion whereas the dispersion matrix of w_l must have rank l_*. Evidently, r_l can serve as a pivot for updates in the general linear model if and only if $\rho(D(r_l)) = l_* - [\rho(X_n) - \rho(X_m)]$. The latter condition is satisfied when V_n is nonsingular.

3.2. Update equations

Let us assume that $0 < l_* - [\rho(X_n) - \rho(X_m)] = \rho(D(r_l))$, i.e., some LZFs (represented adequately by r_l) are lost because of data exclusion. We have

$$X_m\widehat{\beta}_m = X_m\widehat{\beta}_n + Cov(X_m\widehat{\beta}_m, r_l)[D(r_l)]^- r_l. \tag{5}$$

The covariance on the right-hand side have to be expressed in terms of the known quantities in the current model. From (5) it follows that

$$\begin{aligned} &Cov(X_m\widehat{\beta}_m, d_l(\beta)) \\ &\quad = Cov(X_m\widehat{\beta}_n, d_l(\beta)) + Cov(X_m\widehat{\beta}_m, r_l)[D(r_l)]^- Cov(r_l, d_l(\beta)). \end{aligned}$$

Since $d_l(\beta)$ is uncorrelated with y_m while $X_m\widehat{\beta}_m$ is a linear function of it, the left-hand side is zero. On the other hand, $Cov(r_l, d_l(\beta)) - D(r_l)$ is the covariance of r_l with a BLUE in $\mathscr{M}_n$ which must be zero. Therefore, the second term in the right-hand side reduces to $Cov(X_m\widehat{\beta}_m, r_l)$, which can be replaced by $-Cov(X_m\widehat{\beta}_n, d_l(\beta))$ in (5). This simplification, together with the results of Theorem 2.2, leads to the update relationships given below.

Theorem 3.1. *Let* $0 < l_* - [\rho(\boldsymbol{X}_n) - \rho(\boldsymbol{X}_m)] = \rho(D(\boldsymbol{r}_l))$ *and* $\boldsymbol{A\beta}$ *be estimable in either model with* $D(\boldsymbol{A}\widehat{\boldsymbol{\beta}}_n)$ *not identically zero. Then the updated statistics for the deleted model are as follows*:

(a) $\boldsymbol{X}_m\widehat{\boldsymbol{\beta}}_m = \boldsymbol{X}_m\widehat{\boldsymbol{\beta}}_n - Cov(\boldsymbol{X}_m\widehat{\boldsymbol{\beta}}_n, \boldsymbol{d}_l(\boldsymbol{\beta}))[D(\boldsymbol{r}_l)]^-\boldsymbol{r}_l.$

(b) $D(\boldsymbol{X}_m\widehat{\boldsymbol{\beta}}_m) = D(\boldsymbol{X}_m\widehat{\boldsymbol{\beta}}_n) + Cov(\boldsymbol{X}_m\widehat{\boldsymbol{\beta}}_n, \boldsymbol{d}_l(\boldsymbol{\beta}))[D(\boldsymbol{r}_l)]^- Cov(\boldsymbol{d}_l(\boldsymbol{\beta}), \boldsymbol{X}_m\widehat{\boldsymbol{\beta}}_n).$

(c) $R^2_{0_m} = R^2_{0_n} - \sigma^2\boldsymbol{r}_l'[D(\boldsymbol{r}_l)]^-\boldsymbol{r}_l.$

(d) *The reduction in the error sum of squares under the restriction* $\boldsymbol{A\beta} = \boldsymbol{\xi}$ *is given by* $R^2_{H_m} = R^2_{H_n} - \sigma^2\boldsymbol{r}_{l*}'[D(\boldsymbol{r}_{l*})]^-\boldsymbol{r}_{l*}$, *where*
$\boldsymbol{r}_{l*} = \boldsymbol{r}_l + Cov(\boldsymbol{d}_l(\boldsymbol{\beta}), \boldsymbol{A}\widehat{\boldsymbol{\beta}}_n)[\boldsymbol{D}(\boldsymbol{A}\widehat{\boldsymbol{\beta}}_n)]^-(\boldsymbol{A}\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\xi}).$

(e) *The degrees of freedom of* R^2_0 *and* R^2_H *reduce by* l_* *and* $\rho(D(\boldsymbol{r}_{l*}))$, *respectively, as a result of data exclusion.*

The difficulty of finding an update equation in the case $\rho(D(\boldsymbol{r}_l)) < l_* - [\rho(\boldsymbol{X}_n) - \rho(\boldsymbol{X}_m)]$ can be appreciated by considering the model with

$$\boldsymbol{y}_n = \begin{pmatrix} \boldsymbol{y}_m \\ y_{m+1} \\ y_{m+2} \end{pmatrix}, \quad \boldsymbol{X}_n = \begin{pmatrix} & \boldsymbol{X}_m & \\ 1 & & 0 \\ 1 & & 1 \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \quad \boldsymbol{V}_n = \begin{pmatrix} \boldsymbol{V}_m \\ \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix}.$$

If $\rho(\boldsymbol{X}_m) = 2$ and $\boldsymbol{V}_m = \boldsymbol{I}_{m\times m}$, then $l_* - [\rho(\boldsymbol{X}_n) - \rho(\boldsymbol{X}_m)] = 2$, while $\rho(D(\boldsymbol{r}_l)) = \boldsymbol{0}$. It is clear that $\widehat{\boldsymbol{\beta}}_n = (y_{m+1} : y_{m+2} - y_{m+1})'$ and $D(\widehat{\boldsymbol{\beta}}_n) = \boldsymbol{0}$, but $\widehat{\boldsymbol{\beta}}_m$ has to be calculated afresh. There is no way of 'utilizing' the computations of the model with n observations.

Bhimasankaram and Jammalamadaka (1994a) give algebraic expressions for the updates given in Theorem 3.1 in the special case when $l = 1$ and $\boldsymbol{V}_n$ is nonsingular. Bhimasankaram and Jammalamadaka (1994b) give statistical interpretations of these results along the lines of Theorem 3.1. Another interesting interpretation is given by Chib et al. (1987) in the multivariate normal case. Bhimasankaram et al. (1995) give update equations for data exclusion in all possible cases, using the inverse partition matrix approach. Generalizing a result due to Haslett (1999), SJ show that

$$\boldsymbol{X}_n\widehat{\boldsymbol{\beta}}_m = \boldsymbol{X}_n\widehat{\boldsymbol{\beta}}_n - \boldsymbol{B}_n\begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{w}_l \end{pmatrix},$$

where $\boldsymbol{B}_n$ is such that $\boldsymbol{X}_n\widehat{\boldsymbol{\beta}}_n = \boldsymbol{B}_n\boldsymbol{y}_n$. This elegant result is not amenable to recursive computation, as the right-hand side combines a summary statistic of $\mathscr{M}_n$ with an LZF of $\mathscr{M}_m$. One can simplify the results in the special case of $\boldsymbol{V}_n = \boldsymbol{I}$ and $\boldsymbol{X}_m$ full column rank.

4. Exclusion of explanatory variables

In the present section and the next, we examine the connection between the models $\mathscr{M}_{(k)} = (\boldsymbol{y}, \boldsymbol{X}_{(k)}\boldsymbol{\beta}_{(k)}, \sigma^2\boldsymbol{V})$ and $\mathscr{M}_{(h)} = (\boldsymbol{y}, \boldsymbol{X}_{(h)}\boldsymbol{\beta}_{(h)}, \sigma^2\boldsymbol{V})$ $(k > h)$, where the subscript within parentheses represents the number of *explanatory variables* in the model, $\boldsymbol{X}_{(k)} = (\boldsymbol{X}_{(h)} : \boldsymbol{X}_{(j)})$, and $\boldsymbol{\beta}_{(k)} = \binom{\boldsymbol{\beta}_{(h)}}{\boldsymbol{\beta}_{(j)}}$. We shall refer to $\mathscr{M}_{(k)}$ and $\mathscr{M}_{(h)}$ as the *larger* and *smaller* model, respectively.

The model $\mathscr{M}_{(h)}$ can be viewed as a restricted version of the model $\mathscr{M}_{(k)}$, where the restriction is $\boldsymbol{\beta}_{(j)} = \boldsymbol{0}$. For the consistency of the smaller model with the data, we assume that $(\boldsymbol{I} - \boldsymbol{P}_V)\boldsymbol{y} \in \mathscr{C}((\boldsymbol{I} - \boldsymbol{P}_V)\boldsymbol{X}_{(h)})$. It follows that the data is consistent with the larger model as well.

We consider the transition from the larger to the smaller model ($\mathscr{M}_{(k)}$ to $\mathscr{M}_{(h)}$) in this section, and the reverse transition in Section 5.

4.1. Linear zero functions gained

It is easy to see that every LZF in the larger model is an LZF in the smaller model. A standardized basis of LZFs for the model with k parameters contains $\rho(\boldsymbol{X}_{(k)} : \boldsymbol{V}) - \rho(\boldsymbol{X}_{(k)})$ LZFs. If this set is extended to a standardized basis for the smaller model, then the number of uncorrelated LZFs exclusive to the smaller model is $j_* = \rho(\boldsymbol{X}_{(h)} : \boldsymbol{V}) - \rho(\boldsymbol{X}_{(k)} : \boldsymbol{V}) - \rho(\boldsymbol{X}_{(h)}) + \rho(\boldsymbol{X}_{(k)})$. It is clear that $0 \leqslant j_* \leqslant \rho(\boldsymbol{X}_{(j)})$.

We first show that the above expression for j_* can be simplified to $\rho(\boldsymbol{X}_{(k)}) - \rho(\boldsymbol{X}_{(h)})$, if we dispose of a pathological special case. Suppose that $\boldsymbol{x}$ is an explanatory variable exclusive to the larger model which is not in $\mathscr{C}(\boldsymbol{X}_{(h)} : \boldsymbol{V})$. Then $\boldsymbol{l} = (\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}_{(h)}:\boldsymbol{V}})\boldsymbol{x}$ must be a nontrivial vector. Consistency of the smaller model dictates that $\boldsymbol{l}'\boldsymbol{y} = 0$ with probability 1, while that of the larger model requires $\boldsymbol{l}'\boldsymbol{y} = (\boldsymbol{l}'\boldsymbol{x})\beta = \|\boldsymbol{l}\|^2\beta$, where β is the coefficient of $\boldsymbol{x}$ in the larger model. These two conditions hold simultaneously only if β is identically zero, that is, when $\boldsymbol{x}$ is *useless* as an explanatory variable. *We now assume that there is no useless explanatory variable in the larger model, that is,* $\rho(\boldsymbol{X}_{(k)} : \boldsymbol{V}) = \rho(\boldsymbol{X}_{(h)} : \boldsymbol{V})$. Consequently $j_* = \rho(\boldsymbol{X}_{(k)}) - \rho(\boldsymbol{X}_{(h)})$.

Another trivial case occurs when $j_* = 0$. Under this condition, the explanatory variables exclusive to the larger model are *redundant* in the presence of the other explanatory variables, so that each model is a reparametrization of the other. The various statistics of interest under the two models are essentially the same. The case of main interest is $0 < j_* \leqslant \rho(\boldsymbol{X}_{(j)})$.

Recall that j_* is the maximum number of uncorrelated LZFs in the smaller model that are uncorrelated with all the LZFs in the larger model. A vector of LZFs having this property must be a BLUE in the larger model. The following result provides a set of such linear functions.

Theorem 4.1. *The linear function* $\boldsymbol{v} = (\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}_{(h)}})\boldsymbol{X}_{(j)}\widehat{\boldsymbol{\beta}}_{(j)}$, *is a vector of BLUEs in the model* $\mathscr{M}_{(k)}$ *and a vector of LZFs in the model* $\mathscr{M}_{(h)}$. *Further,* $\rho(D(\boldsymbol{v})) = j_*$.

Proof. The parametric function $(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}_{(h)}})\boldsymbol{X}_{(k)}\boldsymbol{\beta}_{(k)}$ is estimable in the larger model. The BLUE of this function is $\boldsymbol{v}$. Under the smaller model, $E(\boldsymbol{v}) = 0$. Since the column space of $D(\boldsymbol{X}_{(k)}\widehat{\boldsymbol{\beta}}_{(k)})$ is $\mathscr{C}(\boldsymbol{X}_{(k)}) \cap \mathscr{C}(\boldsymbol{V})$ (see Section 1), that of $D(\boldsymbol{v})$ must be $\mathscr{C}((\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}_{(h)}})\boldsymbol{X}_{(k)}) \cap \mathscr{C}((\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}_{(h)}})\boldsymbol{V})$. Note that

$$\begin{aligned}\mathscr{C}((\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}_{(h)}})\boldsymbol{X}_{(k)}) &\subseteq \mathscr{C}((\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}_{(h)}})(\boldsymbol{X}_{(k)} : \boldsymbol{V})) \\ &= \mathscr{C}((\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}_{(h)}})(\boldsymbol{X}_{(h)} : \boldsymbol{V})) = \mathscr{C}((\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}_{(h)}})\boldsymbol{V}).\end{aligned}$$

Hence, $\mathscr{C}(D(\boldsymbol{v})) = \mathscr{C}((\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}_{(h)}})\boldsymbol{X}_{(k)})$. Consequently, $\rho(D(\boldsymbol{v})) = \rho((\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}_{(h)}})\boldsymbol{X}_{(k)}) = j_*$. □

The quantity $\boldsymbol{v}$ introduced in the above theorem is a special case of $\boldsymbol{A}\widehat{\boldsymbol{\beta}} - \boldsymbol{\xi}$ with $\boldsymbol{A} = (\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}_{(h)}})\boldsymbol{X}_{(k)}$ and $\boldsymbol{\xi} = \boldsymbol{0}$. The equivalent linear restriction is essentially the 'testable part' of the generally untestable restriction $\boldsymbol{\beta}_{(j)} = \boldsymbol{0}$.

4.2. Update equations

The only functions of $\boldsymbol{\beta}_{(h)}$ that are estimable under the larger model are linear combinations of $(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}_{(j)}})\boldsymbol{X}_{(h)}\boldsymbol{\beta}_{(h)}$. This is only a subset of the linear combinations of $\boldsymbol{X}_{(h)}\boldsymbol{\beta}_{(h)}$, all of which are estimable in the smaller model. The rank of $(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}_{(j)}})\boldsymbol{X}_{(h)}$ is j_*, which is the maximum number of uncorrelated LZFs which are exclusive to the smaller model. Hence, a necessary and sufficient condition for all the estimable functions in the *smaller* model to be estimable under the larger model is that $j_* = \rho(\boldsymbol{X}_{(j)})$.

Even if $0 < j_* < \rho(\boldsymbol{X}_{(j)})$, there may be *some* functions of $\boldsymbol{\beta}_{(h)}$ that are estimable under both the models. We now proceed to obtain the update of the BLUE of such a function when the last j explanatory variables are dropped from the larger model. Once again, we use a 'tilde' for the estimators under the smaller model and a 'hat' for those under the larger model.

The results given below follow along the lines of Theorem 2.2.

Theorem 4.2. *Under the above set-up, let* $\rho(\boldsymbol{X}_{(k)} : \boldsymbol{V}) = \rho(\boldsymbol{X}_{(h)} : \boldsymbol{V})$, *and* $j_* = \rho(\boldsymbol{X}_{(k)}) - \rho(\boldsymbol{X}_{(h)})$. *Let* $\boldsymbol{A}\boldsymbol{\beta}_{(h)}$ *be estimable under the larger model. Then*

(a) $\boldsymbol{A}\widetilde{\boldsymbol{\beta}}_{(h)} = \boldsymbol{A}\widehat{\boldsymbol{\beta}}_{(h)} - Cov(\boldsymbol{A}\widehat{\boldsymbol{\beta}}_{(h)}, \boldsymbol{v})[D(\boldsymbol{v})]^-\boldsymbol{v}$, *where* $\boldsymbol{v} = (\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}_{(h)}})\boldsymbol{X}_{(j)}\widehat{\boldsymbol{\beta}}_{(j)}$.

(b) $D(\boldsymbol{A}\widetilde{\boldsymbol{\beta}}_{(h)}) = D(\boldsymbol{A}\widehat{\boldsymbol{\beta}}_{(h)}) - Cov(\boldsymbol{A}\widehat{\boldsymbol{\beta}}_{(h)}, \boldsymbol{v})[D(\boldsymbol{v})]^- Cov(\boldsymbol{v}, \boldsymbol{A}\widehat{\boldsymbol{\beta}}_{(h)})$.

(c) $R^2_{0_{(h)}} = R^2_{0_{(k)}} + \boldsymbol{v}'[\sigma^{-2}D(\boldsymbol{v})]^-\boldsymbol{v}$.

(d) *The change in* R_H^2 *corresponding to the hypothesis* $\boldsymbol{A\beta}_{(h)} = \boldsymbol{\xi}$ *is*
$R^2_{H_{(h)}} = R^2_{H_{(k)}} + \boldsymbol{v}'_*[\sigma^{-2}D(\boldsymbol{v}_*)]^-\boldsymbol{v}_*$, *where*
$\boldsymbol{v}_* = \boldsymbol{v} - Cov(\boldsymbol{v}, \boldsymbol{A}\widehat{\boldsymbol{\beta}}_{(h)})[D(\boldsymbol{A}\widehat{\boldsymbol{\beta}}_{(h)})]^-(\boldsymbol{A}\widehat{\boldsymbol{\beta}}_{(h)} - \boldsymbol{\xi})$.

(e) *As a result of exclusion of the explanatory variables, the degrees of freedom of* R_0^2 *and* R_H^2 *increase by* j_* *and* $\rho(D(\boldsymbol{v}_*))$, *respectively.* □

Depending on the special case at hand, one may use a different form of $\boldsymbol{v}$ that would have the requisite properties. For instance, if $j_* = \rho(\boldsymbol{X}_{(j)})$, it can be chosen as $\boldsymbol{X}_{(j)}\widehat{\boldsymbol{\beta}}_{(j)}$. If $j_* = j$, $\boldsymbol{v}$ can be chosen as $\widehat{\boldsymbol{\beta}}_{(j)}$.

The vector $\boldsymbol{v}_*$ of part (d) is the BLUE of $(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}_{(h)}})\boldsymbol{X}_{(j)}\boldsymbol{\beta}_{(j)}$ in the larger model under the restriction $\boldsymbol{A\beta}_{(h)} = \boldsymbol{\xi}$. If $\boldsymbol{\beta}_{(k)}$ is entirely estimable under the original model, then we have

$$\widetilde{\boldsymbol{\beta}}_{(h)} = \widehat{\boldsymbol{\beta}}_{(h)} - Cov(\widehat{\boldsymbol{\beta}}_{(h)}, \widehat{\boldsymbol{\beta}}_{(j)})[D(\widehat{\boldsymbol{\beta}}_{(j)})]^-\widehat{\boldsymbol{\beta}}_{(j)},$$

$$D(\widetilde{\boldsymbol{\beta}}_{(h)}) = D(\widehat{\boldsymbol{\beta}}_{(h)}) - Cov(\widehat{\boldsymbol{\beta}}_{(h)}, \widehat{\boldsymbol{\beta}}_{(j)})[D(\widehat{\boldsymbol{\beta}}_{(j)})]^- Cov(\widehat{\boldsymbol{\beta}}_{(h)}, \widehat{\boldsymbol{\beta}}_{(j)})',$$

$$\widetilde{\boldsymbol{\beta}}_{(j)} = \boldsymbol{0}, \quad D(\widetilde{\boldsymbol{\beta}}_{(j)}) = \boldsymbol{0}.$$

These updates only involve $\widehat{\boldsymbol{\beta}}_{(k)}$ and its dispersion.

Bhimasankaram and Jammalamadaka (1994b) give the update formulae for the exclusion of a single explanatory variable when $\boldsymbol{V}$ is nonsingular and these can be obtained as a special case of Theorem 4.2.

5. Inclusion of explanatory variables

We now consider the transition from the model $\mathscr{M}_{(h)} = (\boldsymbol{y}, \boldsymbol{X}_{(h)}\boldsymbol{\beta}_{(h)}, \sigma^2\boldsymbol{V})$ to the model $\mathscr{M}_{(k)} = (\boldsymbol{y}, \boldsymbol{X}_{(k)}\boldsymbol{\beta}_{(k)}, \sigma^2\boldsymbol{V})$ $(k > h)$, where $\boldsymbol{X}_{(k)} = (\boldsymbol{X}_{(h)} : \boldsymbol{X}_{(j)})$, and $\boldsymbol{\beta}_{(k)} = (\boldsymbol{\beta}'_{(h)} : \boldsymbol{\beta}_{(j)})'$. As in Section 4, we refer to $\mathscr{M}_{(k)}$ as the *larger model*, and to $\mathscr{M}_{(h)}$ as the *smaller model*.

5.1. Linear zero functions lost

We need a pivot which can be computed in terms of the statistics of the smaller model. Such a vector is presented below.

Theorem 5.1. *A vector of LZFs in the smaller model that is also a BLUE in the larger model is*

$$\boldsymbol{t} = \boldsymbol{X}'_{(j)}(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}_{(h)}})\{(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}_{(h)}})\boldsymbol{V}(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}_{(h)}})\}^-(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}_{(h)}})\boldsymbol{y}. \tag{6}$$

Further, $\rho(D(\boldsymbol{t})) = j_*$.

Proof. It is clear that $\boldsymbol{t}$ is an LZF in the smaller model. Let $\boldsymbol{l}'\boldsymbol{y}$ be an LZF in the augmented model. We can conclude without loss of generality that $\boldsymbol{X}'_{(j)}\boldsymbol{l} = 0$ and $\boldsymbol{X}'_{(h)}\boldsymbol{l} = 0$. Writing $\boldsymbol{l}$ as $(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}_{(h)}})\boldsymbol{s}$, we have,

$$\begin{aligned} &Cov(\boldsymbol{t}, \boldsymbol{l}'\boldsymbol{y}) \\ &\quad = \sigma^2\boldsymbol{X}'_{(j)}(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}_{(h)}})\{(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}_{(h)}})\boldsymbol{V}(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}_{(h)}})\}^-(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}_{(h)}})\boldsymbol{V}(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}_{(h)}})\boldsymbol{s} \\ &\quad = \sigma^2\boldsymbol{X}'_{(j)}(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}_{(h)}})\boldsymbol{s} = \sigma^2\boldsymbol{X}'_{(j)}\boldsymbol{l} = 0. \end{aligned}$$

Here, we have used the fact that $\mathscr{C}(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}_{(h)}})\boldsymbol{X}_{(j)})$ is a subset of $\mathscr{C}((\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}_{(h)}})\boldsymbol{V})$ or $\mathscr{C}((\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}_{(h)}})\boldsymbol{V}(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}_{(h)}}))$, which follows from the assumption $\boldsymbol{X}_{(j)} \in \mathscr{C}(\boldsymbol{X}_{(h)} : \boldsymbol{V})$. Being uncorrelated with all LZFs in the larger model, $\boldsymbol{t}$ must be a BLUE there. The rank condition follows from the fact that $\mathscr{C}(D(\boldsymbol{y} - \boldsymbol{X}_{(h)}\widetilde{\boldsymbol{\beta}}_{(h)})) = \mathscr{C}(\boldsymbol{V}(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}_{(h)}}))$, i.e., $\mathscr{C}(D(\boldsymbol{t})) = \mathscr{C}(\boldsymbol{X}'_{(j)}(\boldsymbol{I} - \boldsymbol{P}_{\boldsymbol{X}_{(h)}}))$. □

Recall that $\mathscr{C}(\boldsymbol{X}_{(j)})$ is assumed to be a subset of $\mathscr{C}(\boldsymbol{X}_{(h)} : \boldsymbol{V})$. If $\boldsymbol{X}_{(j)} = \boldsymbol{X}_{(h)}\boldsymbol{B} + \boldsymbol{VC}$, then $\boldsymbol{t}$ is the same as $\boldsymbol{C}'\boldsymbol{y}_{\rm res}$, where $\boldsymbol{y}_{\rm res}$ is the residual of $\boldsymbol{y}$ from the smaller model. The vector $\boldsymbol{t}$ can also be interpreted as $\boldsymbol{X}'_{(j)_{\rm res}}\boldsymbol{V}^-\boldsymbol{y}_{\rm res}$ where $\boldsymbol{X}_{(j)_{\rm res}} = \boldsymbol{RX}_{(j)}$, the 'residual' of $\boldsymbol{X}_{(j)}$ when regressed (one column at a time) on $\boldsymbol{X}_{(h)}$. Similarly, $D(\boldsymbol{t})$ is the same as $\sigma^2\boldsymbol{X}'_{(j)_{\rm res}}\boldsymbol{V}^-\boldsymbol{X}_{(j)_{\rm res}}$.

The expectations of $\boldsymbol{v}$ and $\boldsymbol{t}$, defined in Theorems 4.1 and 5.1, respectively, are linear functions of $\boldsymbol{\beta}_{(j)}$. These linear parametric functions are estimable in the model $(\boldsymbol{y}_{\text{res}}, \boldsymbol{X}_{(j)_{\text{res}}}\boldsymbol{\beta}_{(j)}, \sigma^2\boldsymbol{W})$, where $\boldsymbol{W} = \boldsymbol{RV}$. Moreover, $\boldsymbol{v}$ and $\boldsymbol{t}$ are BLUEs of the corresponding parametric functions in this 'residual' model, which is obtained from the original (larger) model by pre-multiplying both the systematic and error parts by $\boldsymbol{R}$. When $\boldsymbol{V}$ is positive definite and a *single* explanatory variable is included, the BLUE of the coefficient of the new variable in the augmented model can be interpreted as the estimated (simple) regression coefficient in the 'residual' model.

5.2. Update equations

We now provide the update relations for the larger model where the BLUE is denoted with a 'hat', in terms of the statistics of the smaller model, where the BLUE is denoted with a 'tilde'.

Theorem 5.2. *If $\boldsymbol{A\beta}_{(h)}$ is estimable under the larger model, then*

(a) $\widehat{\boldsymbol{A\beta}}_{(h)} = \widetilde{\boldsymbol{A\beta}}_{(h)} + Cov(\boldsymbol{BX}^-_{(k)}\boldsymbol{y}, \boldsymbol{t})[D(\boldsymbol{t})]^-\boldsymbol{t}$, *where $\boldsymbol{t}$ is as in* (6) *and $\boldsymbol{B}$ is a matrix of the form $(\boldsymbol{A} : \boldsymbol{0})$ having k columns.*

(b) $D(\widehat{\boldsymbol{A\beta}}_{(h)}) = D(\widetilde{\boldsymbol{A\beta}}_{(h)}) + Cov(\boldsymbol{BX}^-_{(k)}\boldsymbol{y}, \boldsymbol{t})[D(\boldsymbol{t})]^- Cov(\boldsymbol{t}, \boldsymbol{BX}^-_{(k)}\boldsymbol{y})$.

(c) $R^2_{0_{(k)}} = R^2_{0_{(h)}} - \sigma^2\boldsymbol{t}'[D(\boldsymbol{t})]^-\boldsymbol{t}$.

(d) $R^2_{H_{(k)}} = R^2_{H_{(h)}} - \sigma^2\boldsymbol{t}'_*[D(\boldsymbol{t}_*)]^-\boldsymbol{t}_*$, *where*

$$\boldsymbol{t}_* = D(\boldsymbol{t})[D(\boldsymbol{t}) + Cov(\boldsymbol{t}, \boldsymbol{BX}^-_{(k)}\boldsymbol{y})[D(\widetilde{\boldsymbol{A\beta}}_{(h)})]^- Cov(\boldsymbol{t}, \boldsymbol{BX}^-_{(k)}\boldsymbol{y})']^- \times [\boldsymbol{t} + Cov(\boldsymbol{t}, \boldsymbol{BX}^-_{(k)}\boldsymbol{y})[D(\widetilde{\boldsymbol{A\beta}}_{(h)})]^-(\widetilde{\boldsymbol{A\beta}}_{(h)} - \boldsymbol{\xi})].$$

(e) *The increase in the degrees of freedom of R^2_0 and R^2_H with the exclusion of the explanatory variables are given by j_* and $\rho(D(\boldsymbol{t}_*))$, respectively.*

Proof. Since $\boldsymbol{t}$ contains j_* uncorrelated LZFs of the smaller model that turn into BLUEs in the larger model, we have

$$\widetilde{\boldsymbol{A\beta}}_{(h)} = \widehat{\boldsymbol{A\beta}}_{(h)} - Cov(\widehat{\boldsymbol{A\beta}}_{(h)}, \boldsymbol{t})[D(\boldsymbol{t})]^-\boldsymbol{t}.$$

Write $\widehat{\boldsymbol{A\beta}}_{(h)}$ as

$$\widehat{\boldsymbol{A\beta}}_{(h)} = \boldsymbol{BX}^-_{(k)}[\boldsymbol{X}_{(k)}\widehat{\boldsymbol{\beta}}_{(k)}] = \boldsymbol{BX}^-_{(k)}\boldsymbol{y} - \boldsymbol{BX}^-_{(k)}[\boldsymbol{y} - \boldsymbol{X}_{(k)}\widehat{\boldsymbol{\beta}}_{(k)}].$$

The second term is an LZF in the larger model and hence is uncorrelated with $\boldsymbol{t}$. Therefore, $Cov(\widehat{\boldsymbol{A\beta}}_{(h)}, \boldsymbol{t}) = Cov(\boldsymbol{BX}^-_{(k)}\boldsymbol{y}, \boldsymbol{t})$, and we have the expression given in part (a). Part (b) follows directly from part (a), after observing that $\widetilde{\boldsymbol{A\beta}}_{(h)}$ is uncorrelated with $\boldsymbol{t}$. Part (c) is a consequence of the characterization of the residual sum of squares in terms of a basis set of LZFs. Part (d) is obtained similarly, after $\boldsymbol{t}$ is adjusted for its correlation with $\widehat{\boldsymbol{A\beta}}_{(h)} - \boldsymbol{\xi}$ which is a BLUE in the larger model under the hypothesis $\boldsymbol{A\beta}_{(h)} = \boldsymbol{\xi}$. The adjusted vector is $\boldsymbol{t}_* = \boldsymbol{t} - Cov(\boldsymbol{t}, \widehat{\boldsymbol{A\beta}}_{(h)})[D(\widehat{\boldsymbol{A\beta}}_{(h)})]^-(\widehat{\boldsymbol{A\beta}}_{(h)} - \boldsymbol{\xi})$, which is further expressed in terms of the statistics of the original model by using parts (a) and (b). Part (e) is easy to prove. □

The vector $\boldsymbol{BX}^-_{(k)}\boldsymbol{y}$ used in parts (a) and (b) depends on the choice of the generalized inverse of $\boldsymbol{X}_{(k)}$, but its covariance with $\boldsymbol{t}$ does not.

6. Concluding remarks

The update equations for data inclusion can be used to derive various diagnostics for model violation, to obtain optimal design of additional observations in a linear model (see Sengupta, 1995; Bhaumik and Mathew, 2001) and even to derive the Kalman filter for recursive prediction in state space models. Nieto and Guerrero (1995) and Sengupta (2004) give different derivations of the Kalman filter in the singular dispersion case. The update equations for excluded data can be used for deriving case deletion diagnostics which were popularized by Belsley et al. (1980), Cook and Weisberg (1994) and others. These can also be used to do missing plot substitution for a designed experiment suffering from limited data loss, under very general circumstances. Update equations for inclusion/exclusion of variables have application in added variable plots and other aspects of regression model building. Sengupta and Jammalamadaka (2003) give a review of some of these applications.

There is some literature on updates for simultaneous change in data and model, see Haslett (1996) for details and an application to Kalman filter.

A dropped variable amounts to setting its coefficient (in the original model) to zero. Inclusion of a linear restriction is an extension of this operation, and can be investigated in terms of the additional LZF. Kala and Klaczyński (1988) and Pordzik (1992b) obtain the explicit formulae for sequential inclusion of linear restrictions in the case when $\boldsymbol{V}$ is nonsingular.

There is an interesting connection between the exclusion of observations from the homoscedastic linear model and inclusion of some special variables to it. If we wish to drop the last l observations, then the corresponding updates are given in Theorem 3.1. This theorem uses a key LZF, $\boldsymbol{r}_l$, which is uncorrelated with all the LZFs of the depleted model. This LZF is lost when the observations are dropped. If $\boldsymbol{V}_{lm} = \boldsymbol{0}$, then the expression for $\boldsymbol{r}_l$ reduces to $\boldsymbol{e}_l$, the residuals of the last l observations. If, instead of dropping l observations, we seek to include l explanatory variables (in the form of an $n \times l$ matrix $\boldsymbol{Z}$ concatenated to the columns of $\boldsymbol{X}$), then the appropriate 'lost' LZF is given by $\boldsymbol{t}_l$ (see (6)). This LZF reduces to $\boldsymbol{Ze}$ in the present case, $\boldsymbol{e}$ being the residual vector in the original model. The key LZFs in the two cases would be identical if $\boldsymbol{Z}$ is chosen to be the last l columns of an $n \times n$ identity matrix. Since the LZFs are identical, all the updates would naturally be identical. Thus, the dropping of the observations is equivalent to the inclusion of the explanatory variables. A special case of this result (when $\boldsymbol{V} = \boldsymbol{I}$ and $l = 1$) has been known to researchers for a long time (see, for instance, Schall and Dunne, 1988), and has been applied to analysis of covariance with unbalanced data (see SJ) and diagnostics in the Cox regression model (see Storer and Crowley, 1985).

References

Belsley, D.A., Kuh, E., Welsch, R.E., 1980. Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. Wiley Series in Probability and Mathematical Statistics. Wiley, New York.

Bhaumik, D., Mathew, T., 2001. Optimal data augmentation for the estimation of a linear parametric function in linear models. Sankhyā Ser. B 63, 10–26.

Bhimasankaram, P., Jammalamadaka, S.R., 1994a. Recursive estimation and testing in general linear models with applications to regression diagnostics. Tamkang J. Math. 25, 353–366.

Bhimasankaram, P., Jammalamadaka, S.R., 1994b. Updates of statistics in a general linear model: a statistical interpretation and applications. Comm. Statist. Simulation Comput. 23, 789–801.

Bhimasankaram, P., Sengupta, D., Ramanathan, S., 1995. Recursive inference in a general linear model. Sankhyā Ser. A 57, 227–255.

Brown, R.L., Durbin, J., Evans, J.M., 1975. Methods of investigating whether a regression relationship is constant over time (with discussion). J. Roy. Statist. Soc. Ser. B 37, 149–192.

Chambers, J.M., 1975. Updating methods for linear models for the addition or deletion of observations. In: Srivastava, J.N. (Ed.), A Survey of Statistical Design and Linear Models. North-Holland, Amsterdam, pp. 53–65.

Chib, S., Jammalamadaka, S.R., Tiwari, R., 1987. Another look at some results on the recursive estimation in the general linear model. Amer. Statist. 41, 56–58.

Cook, R.D., Weisberg, S., 1994. An Introduction to Regression Graphics. Wiley Series in Probability and Mathematical Statistics. Wiley, New York.

Farebrother, R.W., 1988. Linear Least Squares Computations. Mercel Dekker, New York.

Gragg, W.B., LeVeque, R.J., Trangenstein, J.A., 1979. Numerically stable methods for updating regressions. J. Amer. Statist. Assoc. 74, 161–168.

Haslett, S., 1996. Updating linear models with dependent errors to include additional data and/or parameters. Linear Algebra Appl. 237/238, 329–349.

Haslett, S., 1985. Recursive estimation of the general linear model with dependent errors and multiple additional observations. Austral. J. Statist. 27, 183–188.

Haslett, J., 1999. A simple derivation of deletion diagnostic results for the general linear model with correlated errors. J. Roy. Statist. Soc. Ser. B 61, 603–609.

Jammalamadaka, S.R., Sengupta, D., 1999. Changes in the general linear model: a unified approach. Linear Algebra Appl. 289, 225–242.

Kala, R., Klaczyński, K., 1988. Recursive improvement of estimates in a Gauss–Markov model with linear restrictions. Canad. J. Statist. 16, 301–305.

Kianifard, F., Swallow, W., 1996. A review of the development and application of recursive residuals in linear models. J. Amer. Statist. Assoc. 91, 391–400.

Kourouklis, S., Paige, C.C., 1981. A constrained least squares approach to the general Gauss–Markov linear model. J. Amer. Statist. Assoc. 76, 620–625.

McGilchrist, C.A., Sandland, R.L., 1979. Recursive estimation of the general linear model with dependent errors. J. Roy. Statist. Soc. Ser. B 41, 65–68.

Mitra, S.K., Bhimasankaram, P., 1971. Generalized inverses of partitioned matrices and recalculation of least squares estimators for data and model changes. Sankhyā Ser. A 33, 395–410.

Nieto, F.H., Guerrero, V.M., 1995. Kalman filter for singular and conditional state-space models when the system state and the observational error are correlated. Statist. Probab. Lett. 22, 303–310.

Placket, R.L., 1950. Some theorems in least squares. Biometrika 37, 149–157.

Pordzik, P.R., 1992a. A lemma on g-inverse of the bordered matrix and its application to recursive estimation in the restricted model. Comput. Statist. 7, 31–37.

Pordzik, P.R., 1992b. Adjusting of estimates in general linear model with respect to linear restrictions. Statist. Probab. Lett. 15, 125–130.

Schall, R., Dunne, T.T., 1988. A unified approach to outliers in the general linear model. Sankhyā Ser. B 50, 157–167.

Sengupta, D., 1995. Optimal choice of a new observation in a linear model. Sankhyā Ser. A 57, 137–153.

Sengupta, D., Jammalamadaka, S.R., 2003. Linear Models An Integrated Approach. World Scientific, New Jersey.

Storer, B.E., Crowley, J., 1985. A diagnostic for Cox regression for and general conditional likelihoods. J. Amer. Statist. Assoc. 389, 139–147.