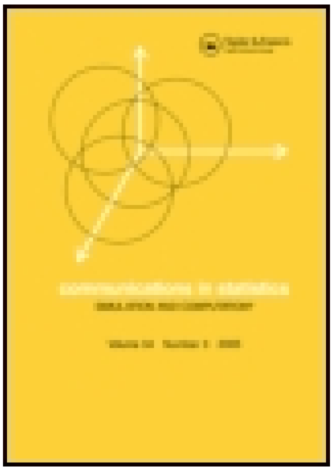

Communications in Statistics - Simulation and Computation
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/lssp20

Updates of statistics in a general linear model: a statistical interpretation and applications

P. Bhimasankaram [a] & S Rao Jammalamadaka [b]

[a] Computer Science Unit , Indian Statistical Institute , 203, Barrackpore Trunk Road, Calcutta, 700 035, India

[b] Department of Statistics & Applied Probability , University of California , Santa Barbara, CA 93106

Published online: 27 Jun 2007.

COMMUN. STATIST.—SIMULA., 23(3), 789-801 (1994)

UPDATES OF STATISTICS IN A GENERAL LINEAR MODEL: A STATISTICAL INTERPRETATION AND APPLICATIONS

P. Bhimasankaram and S. Rao Jammalamadaka [1]

Computer Science Unit
Indian Statistical Institute
203, Barrackpore Trunk Road
Calcutta 700 035, India

Department of Statistics
& Applied Probability
University of California
Santa Barbara, CA 93106

Keywords and phrases: linear model, linear zero function, regression diagnostics, residual, updating formulae, correlated errors.

ABSTRACT

We consider a general linear model $(\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma}$ is a general positive definite matrix and $\boldsymbol{X}$ is possibly rank-deficient. We give updated formulae for various statistical quantities of interest (BLUEs, residual sum of squares, etc.) in the following situations: introduction of an additional observation, deletion of an observation, inclusion of a new regressor and deletion of a regressor. We give the formulae in statistical terminology so that their significance is better understood. We then give an application of these results to Regression Diagnostics in the linear model with correlated errors.

[1] Research supported by Office of Naval Research Grant N00014-93-1-0174.

1. INTRODUCTION AND NOTATION

Updates of Best Linear Unbiased Estimators (BLUEs) and residual sum of squares in linear models have been dealt with, among others, by Plackett (1950), Mitra and Bhimasankaram (1971), McGillchrist and Sandland (1979), Haslett (1985) and Chib, Jammalamadaka and Tiwari (1987), when one or more additional observations become available. While the first two papers considered a model with uncorrelated errors, the later three papers dealt with dependent errors. All the papers except that of Mitra and Bhimasankaram considered full rank design matrix. Golub and Styan (1973), Paige (1978) and Kourouklis and Paige (1981) gave numerically stable recursive computations for a general linear model.

Recently, Bhimasankaram and Jammalamadaka (1993), henceforth referred to as BJ, obtained algebraic expressions for the updates in a general linear model (with a general positive definite matrix and a possibly rank deficient design matrix) for data or model changes. (By a *data change* we mean introducing an additional observation or deletion of an observation and by *model change* we mean introducing an additional regressor or deletion of a regressor.) In addition to the correction terms for the BLUEs and the residual sum of squares, BJ also obtained the correction terms to LRT statistics for tests concerning several estimable linear parametric functions.

One important advantage of the exact algebraic expressions is the statistical interpretability of the correction terms. That is what we intend to do in this paper. In the process, we are able to simplify some formulae and some proofs further and are also able to provide better insight. It is demonstrated that the linear zero functions play an important role in the updates. Sections 2 and 3 deal with the statistical interpretations of the correction terms regarding data and model changes respectively. Section 4 deals mainly with the deletion diagnostics in a general linear model.

We use the following notation. For a matrix $\boldsymbol{A}$, $\boldsymbol{A}'$, $\boldsymbol{A}^-$, $\mathcal{C}(\boldsymbol{A})$ and $\mathcal{R}(\boldsymbol{A})$ denote its transpose, generalised inverse (g-inverse), column space and row space, respectively. The ordered triplet $(\boldsymbol{y}_h, \boldsymbol{X}_h, \boldsymbol{\Sigma}_h)$ denotes the linear model

$$\boldsymbol{y}_h = \boldsymbol{X}_h\boldsymbol{\beta} + \boldsymbol{\epsilon}_h,$$

where h is the number of observations, $E(\boldsymbol{\epsilon}_h) = \mathbf{o}$ and $D(\boldsymbol{\epsilon}_h) = \sigma^2 \boldsymbol{\Sigma}_h$. $R^2_{0_h}$ denotes the residual sum of squares, namely, $(\boldsymbol{y}_h - \boldsymbol{X}_h\widehat{\boldsymbol{\beta}_h})'\boldsymbol{\Sigma}_h{}^{-1}(\boldsymbol{y}_h - \boldsymbol{X}_h\widehat{\boldsymbol{\beta}}_h)$ where $\widehat{\boldsymbol{\beta}}_h$ is any solution to the normal equations $\boldsymbol{X}_h{}'\boldsymbol{\Sigma}_h{}^{-1}\boldsymbol{X}_h\boldsymbol{\beta} = \boldsymbol{X}_h{}'\boldsymbol{\Sigma}_h{}^{-1}\boldsymbol{y}_h$.

Consider a hypothesis $\mathcal{H} : \boldsymbol{A\beta} = \boldsymbol{\xi}$ where $\mathcal{R}(\boldsymbol{A}) \subseteq \mathcal{R}(\boldsymbol{X}_h)$. Let $T_h = \min_{\boldsymbol{\beta}:\boldsymbol{A\beta}=\boldsymbol{\xi}} (\boldsymbol{y}_h - \boldsymbol{X}_h\boldsymbol{\beta})'\boldsymbol{\Sigma}_h{}^{-1}(\boldsymbol{y}_h - \boldsymbol{X}_h\boldsymbol{\beta})$. We use $R^2_{H_h}$ to denote $T_h - R^2_{0_h}$. Observe that $D(\boldsymbol{A}\widehat{\boldsymbol{\beta}}_h) = \sigma^2\boldsymbol{A}(\boldsymbol{X}_h{}'\boldsymbol{\Sigma}_h{}^{-1}\boldsymbol{X}_h)^-\boldsymbol{A}'$. Using Wald's representation, we write

$$R^2_{H_h} = \frac{1}{\sigma^2}(\boldsymbol{A}\widehat{\boldsymbol{\beta}}_h - \boldsymbol{\xi})'[D(\boldsymbol{A}\widehat{\boldsymbol{\beta}}_h)]^-(\boldsymbol{A}\widehat{\boldsymbol{\beta}}_h - \boldsymbol{\xi}).$$

In a model with $h + 1$ observations, the $(h + 1)$th observation is denoted by $y(h + 1)$ and $\boldsymbol{y}_{h+1}$ denotes the vector $(\boldsymbol{y}_h{}' : y(h + 1))'$. The row vector in the design matrix $\boldsymbol{X}_{h+1}$ corresponding to the observation $y(h + 1)$ is denoted by $\boldsymbol{x}'(h + 1)$ and the variance of $y(h + 1)$ by $\sigma^2 c$. The covariance of $\boldsymbol{y}_h$ with $y(h + 1)$ is denoted by $\sigma^2\boldsymbol{c}$. Thus the dispersion matrix of $\boldsymbol{y}_{h+1}$ is

$$\boldsymbol{\Sigma}_{h+1} = \begin{pmatrix} \boldsymbol{\Sigma}_h & \boldsymbol{c} \\ \boldsymbol{c}' & c \end{pmatrix}.$$

We assume that $\boldsymbol{\Sigma}_{h+1}$ is also positive definite.

In connection with the results on tests of hypotheses we assume the multivariate normality for $\boldsymbol{y}$ without explicitly mentioning it each time.

2. UPDATING FORMULAE FOR A DATA-CHANGE

In this section we give the updates of BLUEs of estimable linear functions of $\boldsymbol{\beta}$ and the conventional unbiased estimator of σ^2 concerning a general linear model when an additional (possibly correlated) observation is introduced into the model or an observation is deleted from the model. We also give the updates of the likelihood ratio test (LRT) statistics for testing $\boldsymbol{A\beta} = \boldsymbol{\xi}$ (when $\boldsymbol{A\beta}$ is estimable) in the above situations. We also provide the statistical interpretations of the correction terms.

First, we dispose off a simple case in the following

Theorem 2.1. Consider the models $(\boldsymbol{y}_h, \boldsymbol{X}_h, \boldsymbol{\Sigma}_h)$ and $(\boldsymbol{y}_{h+1}, \boldsymbol{X}_{h+1}, \boldsymbol{\Sigma}_{h+1})$. If $\boldsymbol{x}_{h+1} \notin \mathcal{C}(\boldsymbol{X}_h{}')$, then the following hold.

(a) The classes of linear zero functions under both the models coincide. (Hence, in $(\boldsymbol{y}_{h+1}, \boldsymbol{X}_{h+1}, \boldsymbol{\Sigma}_{h+1})$, no linear zero function involves $y(h+1)$.)

(b) BLUEs of $\boldsymbol{X}_h\boldsymbol{\beta}$ under both the models coincide. (Denote the same by $\widehat{\boldsymbol{X}_h\boldsymbol{\beta}}$.) $D(\widehat{\boldsymbol{X}_h\boldsymbol{\beta}})$ is also the same in both the models.

(c) BLUE of $\boldsymbol{x}'(h+1)\boldsymbol{\beta}$ under $(\boldsymbol{y}_{h+1}, \boldsymbol{X}_{h+1}, \boldsymbol{\Sigma}_{h+1})$ is $y(h+1) - \boldsymbol{c}'\boldsymbol{\Sigma}_h{}^{-1}(\boldsymbol{y}_h - \widehat{\boldsymbol{X}_h\boldsymbol{\beta}})$.

(d) The residual sum of squares and the conventional unbiased estimators of σ^2 under the two models coincide.

(e) Let $\boldsymbol{A\beta}$ be estimable under $(\boldsymbol{y}_h, \boldsymbol{X}_h, \boldsymbol{\Sigma}_h)$. Then the LRT statistics for testing $\boldsymbol{A\beta} = \boldsymbol{\xi}$ and their null distributions under the two models coincide.

(a) is easy to prove. (b) and (c) are simple consequences of (a). (d) follows from (b) and (c). (e) follows from (b) and (d). ∎

Thus, no correction term is needed when an observation is introduced into the model or deleted from the model if the corresponding row of the dispersion matrix does not belong to the space spanned by the rest of the rows. We shall comment later on the importance of the above result in regression diagnostics and in missing plot techniques.

In what follows, we shall consider only the case when $\boldsymbol{x}(h+1) \in \mathcal{C}(\boldsymbol{X}_h{}')$.

2.1. Introducing an additional observation

We shall give the correction terms to $\widehat{\boldsymbol{\beta}}_h$, the unbiased estimator of σ^2 and the LRT statistic for testing $\boldsymbol{A\beta} = \boldsymbol{\xi}$ (where $\mathcal{R}(\boldsymbol{A}) \subseteq \mathcal{R}(\boldsymbol{X}_h)$) when an additional observation is introduced into $(\boldsymbol{y}_h, \boldsymbol{X}_h, \boldsymbol{\Sigma}_h)$ to get $(\boldsymbol{y}_{h+1}, \boldsymbol{X}_{h+1}, \boldsymbol{\Sigma}_{h+1})$. As already remarked, we assume that $\boldsymbol{x}(h+1) \in \mathcal{C}(\boldsymbol{X}_h{}')$.

First, we introduce a few quantities in terms of which we give the correction terms:

(i) $\boldsymbol{v}' = \boldsymbol{x}'(h+1) - \boldsymbol{c}'\boldsymbol{\Sigma}_h{}^{-1}\boldsymbol{X}_h$.

(ii) $d = y(h+1) - \boldsymbol{c}'\boldsymbol{\Sigma}_h^{-1}\boldsymbol{y}_h$. Clearly $E(d) = \boldsymbol{v}'\boldsymbol{\beta}$. Notice that $\boldsymbol{c}'\boldsymbol{\Sigma}_h^{-1}\boldsymbol{y}_h + \boldsymbol{v}'\boldsymbol{\beta}$ is the linear regression of $\boldsymbol{y}_{h+1}$ on $\boldsymbol{y}_h$. Hence d is the residual part of $y(h+1)$ not explainable by a linear function of $\boldsymbol{y}_h$. We call d the *residue* (as opposed to $y(h+1) - \boldsymbol{x}'(h+1)\widehat{\boldsymbol{\beta}}_h$ called the *residual* in the regression diagnostics terminology).

(iii) $\alpha = V(d)/\sigma^2$.

(iv) $\theta = V(\widehat{d}_h)/\sigma^2$ where $\widehat{d}_h = \boldsymbol{v}'\widehat{\boldsymbol{\beta}}_h$ is the predictor of d under $(\boldsymbol{y}_h, \boldsymbol{X}_h, \boldsymbol{\Sigma}_h)$.

(v) $r_h = d - \widehat{d}_h$. r_h is the residual of d under $(\boldsymbol{y}_h, \boldsymbol{X}_h, \boldsymbol{\Sigma}_h)$. Clearly $V(r_h) = (\alpha+\theta)\sigma^2$ as $Cov(d, \widehat{d}_h) = 0$.

(vi) $\mu =$ Regression of r_h on $\boldsymbol{A}\widehat{\boldsymbol{\beta}}_h - \boldsymbol{\xi}$.

(vii) $\delta = (1/\sigma^2)(V(r_h) - V(r_h|\boldsymbol{A}\widehat{\boldsymbol{\beta}}_h))$. Notice that $\delta = V(r_h)$ would imply that r_h is a linear function of $\boldsymbol{A}\widehat{\boldsymbol{\beta}}_h$ which is not possible since $\boldsymbol{\Sigma}_{h+1}$ is p.d.

The following theorem gives the correction terms to various quantities of interest.

Theorem 2.2. Consider the models $(\boldsymbol{y}_h, \boldsymbol{X}_h, \boldsymbol{\Sigma}_h)$ and $(\boldsymbol{y}_{h+1}, \boldsymbol{X}_{h+1}, \boldsymbol{\Sigma}_{h+1})$ where $\boldsymbol{x}(h+1) \in \mathcal{C}(\boldsymbol{X}_h')$. Let $d, v, r_h, \alpha, \theta, \mu$ and δ be as defined above. Then the following hold:

(a) $\widehat{\boldsymbol{\beta}}_{h+1} = \widehat{\boldsymbol{\beta}}_h - r_h \dfrac{Cov(\widehat{\boldsymbol{\beta}}_h, r_h)}{V(r_h)}$.

(b) $R^2_{0_{h+1}} = R^2_{0_h} + \dfrac{r_h^2}{\alpha+\theta} = R^2_{0_h} + \dfrac{r_h^2\sigma^2}{V(r_h)}$.

(c) $R^2_{H_{h+1}} = R^2_{H_h} + \dfrac{r_h\sigma^2}{V(r_h)} \cdot \left(2\mu + \dfrac{\delta r_h}{V(r_h)}\right) + \dfrac{\sigma^2}{V(r_h|\boldsymbol{A}\widehat{\boldsymbol{\beta}}_h)} \cdot \left(\mu + \dfrac{\delta r_h \sigma^2}{V(r_h)}\right)^2$.

The proof follows once the algebraic quantities in Theorem 3.1 of BJ are identified in statistical terms. □

Remark 1: Let $\boldsymbol{p}'\boldsymbol{\beta}$ be estimable. (Notice that the classes of all estimable functions of $\boldsymbol{\beta}$ coincide for the two models under consideration.) Then from (a) of the preceding theorem we have the following. The difference between the BLUEs of $\boldsymbol{p}'\boldsymbol{\beta}$, namely $\boldsymbol{p}'\widehat{\boldsymbol{\beta}}_{h+1} - \boldsymbol{p}'\widehat{\boldsymbol{\beta}}_h$, is $-\dfrac{Cov(\boldsymbol{p}'\widehat{\boldsymbol{\beta}}_h, r_h)}{V(r_h)} r_h$, which is the negative of the linear regression of $\boldsymbol{p}'\widehat{\boldsymbol{\beta}}_h - \boldsymbol{p}'\boldsymbol{\beta}$ on r_h (the residual of the residue).

Remark 2: The ranks of $\boldsymbol{X}_{h+1}$ and $\boldsymbol{X}_h$ are the same. So an additional linear zero function, which is uncorrelated with every linear zero function under $(\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\Sigma})$, has come in due to the introduction of $y(h+1)$. Clearly, this is r_h. This is the reason why $R^2_{0_{h+1}}$ is larger than $R^2_{0_h}$ by $\sigma^2 r_h^2 / V(r_h)$. (See (b) of the preceding theorem.)

2.2. Deleting an observation

Here we consider the model $(\boldsymbol{y}_{h+1}, \boldsymbol{X}_{h+1}, \boldsymbol{\Sigma}_{h+1})$ where $\boldsymbol{x}(h+1) \in \mathcal{C}(\boldsymbol{X}_h{}')$. We shall give the correction terms to $\widehat{\boldsymbol{\beta}}_{h+1}$, the conventional unbiased estimator of σ^2 and the LRT statistic for testing $\boldsymbol{A\beta} = \boldsymbol{\xi}$ (where $\mathcal{R}(\boldsymbol{A}) \subseteq \mathcal{R}(\boldsymbol{X}_h)$) when the observation $y(h+1)$ is dropped to get the model $(\boldsymbol{y}_h, \boldsymbol{X}_h, \boldsymbol{\Sigma}_h)$.

Let d and $\boldsymbol{v}$ be as defined in Section 2.1. Let $\widehat{d}_{h+1} = \boldsymbol{v}'\widehat{\boldsymbol{\beta}}_{h+1}$. Let $r_{h+1} = d - \widehat{d}_{h+1}$. Thus r_{h+1} is the residual of the residue under $(\boldsymbol{y}_{h+1}, \boldsymbol{X}_{h+1}, \boldsymbol{\Sigma}_{h+1})$. It is easy to show that $\boldsymbol{x}(h+1) \in \mathcal{C}(\boldsymbol{X}_h{}')$ if and only if r_{h+1} is not a constant (in fact 0, with probability 1). We have

Theorem 2.3. Consider $(\boldsymbol{y}_{h+1}, \boldsymbol{X}_{h+1}, \boldsymbol{\Sigma}_{h+1})$ and let $\boldsymbol{x}(h+1) \in \mathcal{C}(\boldsymbol{X}_h{}')$. Then, corresponding to the deletion of the observation $y(h+1)$, the following hold:

(a) $\widehat{\boldsymbol{\beta}}_h = \widehat{\boldsymbol{\beta}}_{h+1} + \dfrac{r_{h+1} Cov(\widehat{\boldsymbol{\beta}}_{h+1}, d)}{V(r_{h+1})}$.

(b) $R^2_{0_h} = R^2_{0_{h+1}} - \dfrac{r^2_{h+1}\sigma^2}{V(r_{h+1})}$.

(b) $R^2_{H_h} = R^2_{H_{h+1}} - \dfrac{r^2_{h+1}\sigma^2}{V(r_{h+1})} - \dfrac{(r_{h+1}+\gamma)^2\sigma^2}{V(d) - V(\widehat{d}_{h+1} | \boldsymbol{A}\widehat{\boldsymbol{\beta}}_{h+1})}$,

where $\sigma^2\gamma$ is the linear regression of $\widehat{d}_{h+1} - \boldsymbol{v}'\boldsymbol{\beta}$ on $\boldsymbol{A}\widehat{\boldsymbol{\beta}}_{h+1} - \boldsymbol{\xi}$.

The proof follows once the algebraic quantities in Theorem 3.6 of BJ are identified in statistical terms. □

Remark: The difference in the BLUEs of an estimable function $\boldsymbol{p}'\boldsymbol{\beta}$, namely $\boldsymbol{p}'\widehat{\boldsymbol{\beta}}_h - \boldsymbol{p}'\widehat{\boldsymbol{\beta}}_{h+1}$, is $\dfrac{Cov(\boldsymbol{p}'\widehat{\boldsymbol{\beta}}_{h+1}, d)}{V(r_{h+1}}r_{h+1}$. Compare it with the expression in Remark 1 after Theorem 3.2. There we had $Cov(\boldsymbol{p}'\widehat{\boldsymbol{\beta}}_h, r_h)$. Here, $Cov(\boldsymbol{p}'\widehat{\boldsymbol{\beta}}_{h+1}, r_{h+1})$ is in fact zero, as r_{h+1} is a linear zero function. When we drop the $y(h+1)$, r_{h+1} is the linear zero function that is lost. This is also evident from (b) of the preceding theorem. Compare this with Remark 2 after Theorem 2.2.

3. UPDATING FORMULAE FOR A MODEL CHANGE

In this section, we consider the model $(\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\Sigma})$ and give the correction terms for the estimators of $\boldsymbol{\beta}$ and σ^2 when an additional regressor is introduced into the model or a regressor is dropped from the model. We also give the correction terms for the LRT statistic for testing $\boldsymbol{A\beta} = \boldsymbol{\xi}$ in the above situations.

If the i-th column of $\boldsymbol{X}$ is a linear combination of the remaining columns (this is the exact collinearity situation) then β_i, the corresponding β − parameter is not estimable. Let $\boldsymbol{X}_{(i)}$ be the matrix obtained from $\boldsymbol{X}$ after deleting the i-th column of $\boldsymbol{X}$ without altering the order of the other columns. $\boldsymbol{\beta}_{(i)}$ is defined similarly. Then the residual sums of squares and the LRT statistics for testing $\boldsymbol{A\beta} = \boldsymbol{\xi}$ are identical under $(\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\Sigma})$ and $(\boldsymbol{y}, \boldsymbol{X}_{(i)}, \boldsymbol{\Sigma})$. If $\widehat{\boldsymbol{\beta}}_{(i)}$ is a generalized least squares estimator of $\boldsymbol{\beta}_{(i)}$ then a choice of $\widehat{\boldsymbol{\beta}}$ is $(\widehat{\boldsymbol{\beta}}'_{(i)} : 0)'$.

Thus, if we include an additional regressor which is a linear combination of the existing regressors, or if we delete a regressor which is a linear combination of the rest of the regressors, the least squares analysis does not change in any appreciable way.

Henceforth we shall consider only regressors which are not linearly dependent on the other regressors for inclusion or exclusion from the model.

3.1. Inclusion of an additional regressor

Consider the model $(\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\Sigma})$. Let $\boldsymbol{x}$ be the data on a new regressor leading to the model

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{x}\nu + \boldsymbol{\epsilon}, \qquad E(\boldsymbol{\epsilon}) = \boldsymbol{o}, \qquad D(\boldsymbol{\epsilon}) = \sigma^2\boldsymbol{\Sigma}.$$

Write $\boldsymbol{W} = (\boldsymbol{X} : \boldsymbol{x})$ and $\boldsymbol{\theta} = (\boldsymbol{\beta}' : \nu)'$. Thus the new model after introducing the new regressor can be denoted by $(\boldsymbol{y}, \boldsymbol{W}, \boldsymbol{\Sigma})$. As already remarked, we assume $\boldsymbol{x} \notin \mathcal{C}(\boldsymbol{X})$.

Let $\boldsymbol{u} = (\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{x}$. Thus $\boldsymbol{u}$ is the vector of regression coefficients when the new regressor is regressed on the existing regressors. Let

$$\alpha = \boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{x} - \boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\boldsymbol{u}$$

which is the residual sum of squares corresponding to the above regression. In fact, $\alpha = 0$ if and only if $\boldsymbol{x} \in \mathcal{C}(\boldsymbol{X})$. Let $\tilde{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\beta}}$ denote generalized least squares estimators of $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$ in the models $(\boldsymbol{y}, \boldsymbol{W}, \boldsymbol{\Sigma})$ and $(\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\Sigma})$, respectively. Let $R^2_{0_{new}}$ and $R^2_{0_{old}}$ denote the residual sum of squares under $(\boldsymbol{y}, \boldsymbol{W}, \boldsymbol{\Sigma})$ and $(\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\Sigma})$, respectively. $R^2_{H_{new}}$ and $R^2_{H_{old}}$ are defined similarly in connection with hypothesis testing. We have the following

Theorem 3.1. Consider the models $(\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\Sigma})$ and $(\boldsymbol{y}, \boldsymbol{W}, \boldsymbol{\Sigma})$, where $\boldsymbol{x} \notin \mathcal{C}(\boldsymbol{X})$. Then the following hold.

(a) $\tilde{\boldsymbol{\theta}} = \begin{pmatrix} \tilde{\boldsymbol{\beta}} \\ \tilde{\nu} \end{pmatrix}$ where $\tilde{\nu} = (1/\alpha)\boldsymbol{x}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})$ and $\tilde{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}} - \tilde{\nu}\boldsymbol{u}$.

(b) $R^2_{0_{new}} = R^2_{0_{old}} - \tilde{\nu}\boldsymbol{y}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \widehat{\boldsymbol{x}})$, where $\widehat{\boldsymbol{x}} = \boldsymbol{X}\boldsymbol{u}$, the predictor of $\boldsymbol{x}$ based on $\boldsymbol{X}$.

(c) Let $\boldsymbol{A\beta}$ be estimable under $(\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\Sigma})$. Then it is estimable under $(\boldsymbol{y}, \boldsymbol{W}, \boldsymbol{\Sigma})$ also and

$$R^2_{H_{new}} = R^2_{H_{old}} + \tilde{\nu}^2\boldsymbol{v}'\boldsymbol{A}\boldsymbol{u} - 2\tilde{\nu}\boldsymbol{v}'(\boldsymbol{A}\widehat{\boldsymbol{\beta}} - \boldsymbol{\xi}) - \frac{[\boldsymbol{v}'(\boldsymbol{A}\widehat{\boldsymbol{\beta}} - \boldsymbol{\xi})]^2}{\alpha + \boldsymbol{v}'\boldsymbol{A}\boldsymbol{u}},$$

where $\boldsymbol{v} = (1/\sigma^2)[D(\boldsymbol{A}\widehat{\boldsymbol{\beta}})]^{-}A\boldsymbol{u}$.

(d) ν is estimable under $(\boldsymbol{y}, \boldsymbol{W}, \boldsymbol{\Sigma})$ and the usual statistic to test $\mathcal{H} : \nu = 0$ is

$$\frac{\boldsymbol{x}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})}{\sqrt{(\boldsymbol{x} - \widehat{\boldsymbol{x}})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \widehat{\boldsymbol{x}})\dfrac{R^2_{0_{new}}}{n - rank(\boldsymbol{X}) - 1}}},$$

which has Student's t distribution with $n - rank(\boldsymbol{X}) - 1$ degrees of freedom under $\mathcal{H}$. (n is the number of observations, that is, $\boldsymbol{y}$ is of the order $n \times 1$.)

The proof follows once the algebraic quantities in Theorem 4.3 of BJ are identified in statistical terms. □

$\tilde{\nu}$ as expected is the regression coefficient in the linear regression of the residual of $\boldsymbol{y}$, $\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}$, on the new regressor. Assuming that we are performing regression with intercept, $\boldsymbol{x}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}-\boldsymbol{X}\widehat{\boldsymbol{\beta}})$ is the corrected sum of products corresponding to the new variable and the residual of the criterion variable when regressed on the original regressors. $\boldsymbol{y}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\widehat{\boldsymbol{x}})$ has a similar interpretation.

3.2. Deletion of a regressor

The statistical interpretation of the differences in the quantities of interest in the context of the models $(\boldsymbol{y},\boldsymbol{X},\boldsymbol{\Sigma})$ and $(\boldsymbol{y},\boldsymbol{W},\boldsymbol{\Sigma})$ is clear from Theorem 3.1 and the following discussion. In this section, we reproduce from BJ the updates when the last regressor is deleted from $(\boldsymbol{y},\boldsymbol{W},\boldsymbol{\Sigma})$ in terms of the computations available from the analysis of the model $(\boldsymbol{y},\boldsymbol{W},\boldsymbol{\Sigma})$. As before we assume that $\boldsymbol{x} \notin \mathcal{C}(\boldsymbol{X})$.

We give the correction terms in terms of the following quantities and α and $\boldsymbol{u}$ as defined in Section 3.1.

$$\begin{aligned}\zeta_1 &= \boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}\\ \zeta_2 &= \boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{y}\end{aligned}$$

We assume without loss of generality that a nonnegative definite (nnd) g-inverse of $\boldsymbol{W}'\boldsymbol{\Sigma}^{-1}\boldsymbol{W}$ is available. (If $\boldsymbol{G}$ is a g-inverse of $\boldsymbol{W}'\boldsymbol{\Sigma}^{-1}\boldsymbol{W}$, then $\boldsymbol{G}(\boldsymbol{W}'\boldsymbol{\Sigma}^{-1}\boldsymbol{W})\boldsymbol{G}'$ is an nnd g-inverse of $\boldsymbol{W}'\boldsymbol{\Sigma}^{-1}\boldsymbol{W}$.)

Partition the nnd g-inverse of $\boldsymbol{W}'\boldsymbol{\Sigma}^{-1}\boldsymbol{W}$ as

$$(\boldsymbol{W}\boldsymbol{\Sigma}^{-1}\boldsymbol{W})^{-} = \begin{pmatrix} \boldsymbol{B} & \boldsymbol{\eta} \\ \boldsymbol{\eta}' & \tau \end{pmatrix}$$

where $(\boldsymbol{\eta}' : \tau)$ is the last row of $(\boldsymbol{W}'\boldsymbol{\Sigma}^{-1}\boldsymbol{W})^{-}$ and τ is a scalar. It is easy to identify that $\tau = (1/\alpha)$ and $\boldsymbol{\eta} = -(1/\alpha)\boldsymbol{u}$. Let $\tilde{\boldsymbol{\beta}}$, $R^2_{0_{old}}$ and $R^2_{H_{old}}$ with their usual meanings correspond to the model $(\boldsymbol{y},\boldsymbol{W},\boldsymbol{\Sigma})$ and let $\widehat{\boldsymbol{\beta}}$, $R^2_{0_{new}}$ and $R^2_{H_{new}}$ correspond to the model $(\boldsymbol{y},\boldsymbol{X},\boldsymbol{\Sigma})$. Then we have

Theorem 3.2 (BJ). Consider the models $(\boldsymbol{y}, \boldsymbol{W}, \boldsymbol{\Sigma})$ and $(\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\Sigma})$ where $\boldsymbol{x} \notin \mathcal{C}(\boldsymbol{X})$. Then the following hold:

(a) $\widehat{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}} + (1/\alpha)(\zeta_2 - \boldsymbol{u}'\boldsymbol{\zeta}_1)\boldsymbol{u}$.

(b) $R^2_{0_{new}} = R^2_{0_{old}} + \dfrac{1}{\alpha}(\zeta_2 - \boldsymbol{u}'\boldsymbol{\zeta}_1)^2 + \dfrac{1-\alpha}{\alpha}\zeta_2\boldsymbol{u}'\boldsymbol{\zeta}_1$.

(c) Let $\boldsymbol{A\beta}$ be estimable under both the models. Consider $\mathcal{H} : \boldsymbol{A\beta} = \boldsymbol{\xi}$. Then

$$R^2_{H_{new}} = R^2_{H_{old}} + \boldsymbol{\eta}'\boldsymbol{A}'\boldsymbol{R}^-\boldsymbol{A\eta}(\zeta_2 - \boldsymbol{u}'\boldsymbol{\zeta}_1)^2 - 2q(\zeta_2 - \boldsymbol{u}'\boldsymbol{\zeta}_1) + \frac{q^2}{\tau - \boldsymbol{\eta}'\boldsymbol{A}'\boldsymbol{R}^-\boldsymbol{A\eta}}$$

where $\boldsymbol{R} = (1/\sigma^2)D(\boldsymbol{A}\tilde{\boldsymbol{\beta}})$ and $q = \boldsymbol{\eta}'\boldsymbol{A}'\boldsymbol{R}^-(\boldsymbol{A}\tilde{\boldsymbol{\beta}} - \boldsymbol{\xi})$.

The proof follows once the algebraic quantities in Theorem 4.4 of BJ are identified in statistical terms. □

It is easy to see that $\zeta_2 - \boldsymbol{u}'\boldsymbol{\zeta}_1$ is equal to $\boldsymbol{x}'\boldsymbol{\Sigma}^{-1}(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})$, which is the corrected sum of products of the deleted variable with the residual from the regression of the criterion variable on the other predictor variables.

4. APPLICATIONS

The most important application of the updates when an observation is deleted is to the deletion diagnostics in Regression Analysis. The deletion diagnostics are well-developed and studied for the case of uncorrelated errors (see for example Belsley, Kuh and Welsch (1980), Chatterjee and Hadi (1986). We demonstrate here that, even though the algebraic expressions for the diagnostics are different when the observations are correlated, the expressions in terms of certain statistical quantities remain unchanged. We use the same notation as in Belsley, Kuh and Welsch (1980) for the diagnostics, namely $DFBETA$, $DFFIT$ etc. We consider the model $(\boldsymbol{y}_{h+1}, \boldsymbol{X}_{h+1}, \boldsymbol{\Sigma}_{h+1})$ and study the effect of deleting the $(h+1)$th observation using the results of Section 2.2. The effect of deleting any other observation can be studied in the same way by using a suitable change in the notation. We consider only the case $\boldsymbol{x}(h+1) \in \mathcal{C}(\boldsymbol{X}_h{}')$ as there is no change in the other case. By

$DFBETA_{h+1,p}$ we mean the difference $\boldsymbol{p}'\widehat{\boldsymbol{\beta}}_{h+1} - \boldsymbol{p}'\widehat{\boldsymbol{\beta}}_h$ where $\boldsymbol{p}'\boldsymbol{\beta}$ is estimable. We have

$$DFBETA_{h+1,p} = \frac{r_{h+1}}{V(r_{h+1})} Cov(\boldsymbol{p}'\widehat{\boldsymbol{\beta}}_{h+1}, d).$$

The scaled change in $\boldsymbol{p}'\widehat{\boldsymbol{\beta}}_{h+1}$ when the $(h+1)$th observation is deleted is given by

$$DFBETAS_{h+1,p} = \frac{r_{h+1} Cov(\boldsymbol{p}'\widehat{\boldsymbol{\beta}}_{h+1}, d)}{V(r_{h+1})\sqrt{\widehat{V}(\boldsymbol{p}'\widehat{\boldsymbol{\beta}}_{h+1})}},$$

where $\widehat{V}(\boldsymbol{p}'\widehat{\boldsymbol{\beta}}_{h+1})$ is $\boldsymbol{p}'(\boldsymbol{X}_h'\boldsymbol{\Sigma}^{-1}{}_{h+1}\boldsymbol{X}_{h+1})^{-}\boldsymbol{p}\dfrac{R^2_{0_h}}{n - rank(\boldsymbol{X}_h)}$. [Here $\boldsymbol{y}_h$ is of order $n \times 1$.] As a special case, the scaled change in fit of the jth observation y_j when $(h+1)$th observation is deleted is given by

$$DFFITS_{h+1,j} = \widehat{y}_{j,h} - \widehat{y}_{j,h+1} = \frac{r_{h+1}}{V(r_{h+1})} \cdot \frac{Cov(\widehat{y}_{j,h+1}, d)}{\sqrt{\widehat{V}(\widehat{y}_{j,h+1})}}.$$

As pointed out in Bhimasankaram, Sengupta and Ramanathan (1994), $|DFFITS_{h+1,h+1}|$ need not be the maximum among $|DFFITS_{h+1,j}|$ for $j = 1, 2, \ldots, h+1$, when the observations are correlated.

The change in the residual sum of squares after deleting the $(h+1)$th observation is

$$R^2_{0_h} - R^2_{0_{h+1}} = -\frac{r^2_{h+1}\sigma^2}{V(r_{h+1})}.$$

As noted earlier r_{h+1} is the zero function that is lost due to the deletion of the $(h+1)$th observation.

The Cook's distance (square) is given by

$$\frac{r^2_{h+1}}{V(r_{h+1})^2} \cdot \frac{Cov(\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{h+1}, d)'\boldsymbol{\Sigma}^{-1}_{h+1} Cov(\boldsymbol{X}\widehat{\boldsymbol{\beta}}_{h+1}, d)}{kR^2_{0_{h+1}}} \cdot (h + 1 - rank(\boldsymbol{X}_{h+1})),$$

where k is the number of predictors.

We now turn to Theorem 2.1. The case considered there might look pathological at the first glance, but it is far from it. This has an important application to analysis with deleted (or missing) observations. Suppose we want to delete the i-th observation from $(\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\Sigma})$. Consider a new model

$$\boldsymbol{y} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}_i\gamma + \boldsymbol{\epsilon}, \qquad E(\boldsymbol{\epsilon}) = o, \qquad D(\boldsymbol{\epsilon}) = \sigma^2\boldsymbol{\Sigma},$$

where e_i is the i-th column of the identity matrix. Clearly, in this model, the i-th row of $(X : e_i)$ does not belong to the space spanned by the other rows. Hence by using Theorem 2.1 it can be seen that analysing the model (y, X, Σ) after deleting the i-th observation is equivalent to analysing the model $(y, (X : e_i) : \Sigma)$. This has been fruitfully utilised in the deletion diagnostics. Also this is the principle used in the analysis of covariance method used in the missing plot technique.

ACKNOWLEDGEMENTS

We thank the referee for his comments on an earlier draft of the paper which led to a more focussed presentation. We thank Dr. Debasis Sengupta for several useful discussions.

BIBLIOGRAPHY

Belsley, D.A., Kuh, E. and Welsch, R.E. (1980). *Regression diagnostics*, John Wiley, New York.

Bhimasankaram, P. and Jammalamadaka, S.R. (1993). "Recursive estimation and testing in general linear models", to appear in *Tamkang Journal of Mathematics*.

Bhimasankaram, P., Sengupta, D. and Ramanathan, S. (1994). "Recursive inference in a general linear model", to appear in *Sankhyā Ser. A*.

Chatterjee, S. and Hadi, A.S. (1986). *Sensitivity Analysis in Linear Regression*, John Wiley, New York.

Chib, S., Jammalamadaka, S.R. and Tiwari, R.C. (1987). "Another look at some results on the recursive estimation in the general linear model", *American Statistician*, **41**, 56–58.

Golub, G.H. and Styan, G.P.H. (1973). "Numerical computations for univariate linear models", *Journal of Statistical Computing and Simulation*, **2**, 253–274.

Haslett, S. (1985). "Recursive estimation of the general linear model with dependent errors and multiple additional observations", *Australian Journal of Statistics*, **27**, 183–188.

Kourouklis, S. and Paige, C.C. (1981). "A constrained least squares approach to the general Gauss-Markov linear model", *Journal of the American Statistical Association*, **76**, 620–625.

McGillchrist, C.A. and Sandland, R.L. (1979). "Recursive estimation of the general linear model with dependent errors", *Journal of the Royal Statistical Society, Ser. B*, **41**, 65–68.

Mitra, S.K. and Bhimasankaram, P. (1971). "Generalized inverses of partitioned matrices and recalculation of least squares estimators for data or model changes" *Sankhyā Ser. A*, **33**, 396–410.

Paige, C.C. (1978). "Numerically stable computations for general univariate linear model", *Communications in Statistics Ser. B*, **B7**, 437–453.

Plackett, R.L. (1950). "Some theorems in least squares", *Biometrika*, **37**, 149–157.

Received June, 1993; Revised February, 1994.