SOME FOLLOW-UP ON AITKEN'S LEAST SQUARES EQUATIONS[1]

Shayle R. Searle

Biometrics Unit, Cornell University, Ithaca, N. Y., 14853, U.S.A.

BU-1282-M April 1995

ABSTRACT

Extensions of Aitken's (1934) weighted least squares equations are reviewed. Where Aitken used a model matrix of full column rank, and a variance-covariance matrix of full rank as the weight matrix, recent extensions relax these rank conditions, even to including an arbitrary non-negative definite weight matrix. A special extension is Henderson's mixed model equations, which simultaneously provide BLUE of the fixed effects and BLUP of the random effects in mixed models.

Key words: OLSE, WLSE, GLSE, BLUE, singular variance matrix, MME, BLUP, mixed model.

[1] Invited paper for the A.C. Aitken Centenary Conference, Dunedin, New Zealand, August 28 – September 1, 1995.

1. INTRODUCTION

Origins of the method of least squares are entangled in the controversy between Legendre and Gauss (with Laplace being implicated too), as ably described by Plackett (1949, 1972), especially in the second of these. Plackett, in quoting Merriman (1877), asserts that the method was first named by Legendre in an 1805 publication, *Nouvelles méthodes pour la determination des orbites des comètes.* And Harter (1983) quotes an early use as Adrain (1808). However, Galle (1924) indicates that there is evidence that Gauss had applied the method as early as 1794 or 1795; and therein lies the controversy as to priority (see Placket, 1972). However, although Gauss had in 1806 a German language version of his book, which contained his least squares work, it was not published until 1809 – and then only in Latin, with an English translation not coming until 1857 (*loc. cit.*). So no wonder priority is somewhat confusing.

The object of the wrangling that went on between Legendre, Gauss, and Laplace was the establishment of, in today's notation, equations of the form

$$\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\hat{\beta} = \mathbf{X}'\mathbf{V}^{-1}\mathbf{y}\ . \tag{1}$$

In the initial formulation of (1), where $\mathbf{y}$ is a vector of data with expected value of the underlying random variables being $\mathbf{X}\beta$, the $\mathbf{V}^{-1}$ in (1) was taken to be the diagonal matrix of the reciprocals of the variances of those same random variables (of which $\mathbf{y}$ is a realized value).

Our starting point is (1) as derived by Aitken from minimizing the sum of squares $(\mathbf{y} - \mathbf{X}\beta)'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\beta)$ with respect to elements of β. In his treatment he describes $\mathbf{V}$ as being the "variances and product variances [covariances, no doubt] of error arrayed by the elements of a symmetric matrix." Curiously, in that 1934 paper Aitken deems it necessary to define a transposed matrix as being one obtained by "substituting rows for columns, columns for rows"; but he does not define a symmetric matrix, simply stating that $\mathbf{V}$ is symmetric. It is also interesting that Aitken describes equations (1) as normal equations – with no indication of where that use of the word 'normal' has come from.

Today's interest in equation (1) is for obtaining a solution $\hat{\beta}$, so it is somewhat strange to see that Aitken's (1934) interest was to derive what he calls an "approximate representation" of $\mathbf{y}$. For this he obtains from (1) what we now call "predicted $\mathbf{y}$" or $\hat{\mathbf{y}} = \mathbf{X}\hat{\beta}$. To do this Aitken takes $\mathbf{X}$ (for which he used the letter $\mathbf{P}$) as having full column rank; and his $\hat{\mathbf{y}}$ is then

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\beta} \qquad \text{for} \qquad \hat{\beta} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \;, \tag{2}$$

with $\hat{\beta}$ being the familiar solution of (1) for $\mathbf{X}$ of full-column rank and $\mathbf{V}$ non-singular. Under these conditions on $\mathbf{X}$ and $\mathbf{V}$, again for to-day's world, we can see that $\hat{\mathbf{y}}$ is

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\beta} = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \,. \tag{3}$$

2. TWO GENERALIZATIONS

For the symbol $\mathbf{y}$ doing double duty to represent, as appropriate, both a realized value (data) of a vector of random variables, and that vector itself, we deal with a model that is based on

$$\mathrm{E}(\mathbf{y}) = \mathbf{X}\beta \;, \tag{4}$$

where E represents expectation, β is a vector of parameters and $\mathbf{X}$ a known matrix. Our two generalizations are those of having $\mathbf{X}$ of less than full column rank and of having any symmetric, real, non-negative definite matrix in place of $\mathbf{V}^{-1}$.

2.1 Having X of less than full column rank

The first generalization is to accommodate the many applications wherein $\mathbf{X}$ has less than full column rank, in which case $\mathbf{X}'\mathbf{X}$ is singular. Then, in place of the regular inverse $(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}$ use is made of a generalized inverse $(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}$ defined by

$$\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{X} = \mathbf{X}'\mathbf{V}^{-1}\mathbf{X} \,. \tag{5}$$

Then a solution to $\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\hat{\beta}^{o} = \mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$ of (1) is

$$\beta^{o} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \;, \tag{6}$$

where the notation β^{o} in place of $\hat{\beta}$ is used for emphasizing that β^{o} of (6) is only a solution of (1). It is not an estimator of β; it depends on what is used for $(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}$.

In contrast

$$\mathbf{X}\beta^{o} = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \tag{7}$$

is an estimator of $\mathbf{X}\beta$ and is invariant to the choice of $(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}$. Confining attention to $\mathbf{X}\beta$ rather than β avoids the need for considering estimability: for any non-null row vector λ' the function $\lambda'\mathbf{X}\beta$ is estimable.

2.2 A general weight matrix

Gauss and Laplace used as a weight matrix a diagonal matrix of reciprocals of the variances of the elements of $\mathbf{y}$. Aitken extended this to the inverse of the (non-singular) variance-covariance matrix of $\mathbf{y}$, thus permitting correlation among elements of $\mathbf{y}$ to be taken into account. As our second generalization, suppose we use a weight matrix $\mathbf{W}$, restricted only to being real, symmetric and non-negative definite. Then we have equations

$$\mathbf{X'WX}\beta^o = \mathbf{X'Wy} \ , \tag{8}$$

and on solving them for β^o and defining $\mathbf{X}\beta^o$ as $\hat{\mu}(\mathbf{W})$ gives

$$\hat{\mu}(\mathbf{W}) = \mathbf{X}(\mathbf{X'WX})^{-}\mathbf{X'Wy} \ . \tag{9}$$

Clearly, this is quite a broad generalization of (7). Nevertheless, it can be unbiased for $\mathbf{X}\beta^o$, as indicated by the following theorem from Searle (1994).

Theorem 1 A necessary and sufficient condition for $\hat{\mu}(\mathbf{W})$ to be either invariant to $(\mathbf{X'WX})^{-}$ or unbiased for $\mathbf{X}\beta$ is that $\mathbf{X} = \mathbf{CWX}$ (with $\mathbf{WX} \neq \mathbf{0}$) for some $\mathbf{C}$; and then both invariance and unbiasedness are assured.

3. SEVERAL FORMS OF LEAST SQUARES ESTIMATION

To this point no names have been mentioned for solutions of normal equations involving weight matrices. This is because the literature is not unequivocal in this matter. A brief review is therefore in order.

3.1 Ordinary least squares estimation (OLSE)

The simplest weight matrix is $\mathbf{W} = \mathbf{I}$, or equivalently, $\mathbf{W} = \mathbf{V}^{-1}$ with $\mathbf{V} = \sigma^2\mathbf{I}$. Then, as is well known, the least squares equations (1) are

$$\mathbf{X'X}\beta^o = \mathbf{X'y} \qquad \text{with solution} \qquad \beta^o = (\mathbf{X'X})^{-}\mathbf{X'y} \tag{10}$$

for any generalized inverse $(\mathbf{X'X})^{-}$. Then $\mathbf{X}\beta^o$ for β^o of (10) is $\hat{\mu}(\mathbf{I})$, and is the ordinary least squares estimator (OLSE) of $\mathbf{X}\beta$:

$$\text{OLSE}(\mathbf{X}\beta) = \mathbf{X}(\mathbf{X'X})^{-}\mathbf{X'y} \ . \tag{11}$$

It is invariant to the choice of $(\mathbf{X'X})^{-}$.

3.2 Generalized least squares estimation (GLSE)

Although (11) has been developed from the viewpoint of using $\mathbf{W} = \sigma^2\mathbf{I}$ in $\hat{\mu}(\mathbf{W})$, a more basic development starting from

$$\mathbf{y} \sim (\mathbf{X}\boldsymbol{\beta}, \quad \sigma^2\mathbf{I}) \tag{12}$$

is to simply minimize the sum of squares of the elements of $\mathbf{y} - \mathrm{E}(\mathbf{y})$, i.e., minimize $(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$. This yields (1) and (2).

For $\mathbf{V}$ being other than $\sigma^2\mathbf{I}$, Aitken (1934) offers not a word on why one would want to minimize $(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$ and so get (1). But for

$$\mathbf{y} \sim (\mathbf{X}\boldsymbol{\beta}, \quad \mathbf{V}) \tag{13}$$

one can get (1) by recognizing that the nature of $\mathbf{V}$ permits writing $\mathbf{V}^{-1} = \mathbf{L}\mathbf{L}'$ for non-singular $\mathbf{L}$. Using $\mathbf{L}$ to consider $\mathbf{L}\mathbf{y}$ in (13) instead of $\mathbf{y}$ then gives

$$\mathbf{L}\mathbf{y} \sim [\mathbf{L}\mathbf{X}\boldsymbol{\beta}, \quad \mathbf{L}(\mathbf{L}'\mathbf{L})^{-1}\mathbf{L}'] \sim (\mathbf{L}\mathbf{X}\boldsymbol{\beta}, \quad \mathbf{I})$$

and applying (10) to this (i.e., $\mathbf{L}\mathbf{X}$ in place of $\mathbf{X}$ and $\mathbf{L}\mathbf{y}$ in place of $\mathbf{y}$) gives

$$\mathbf{X}'\mathbf{L}'\mathbf{L}\mathbf{X}\boldsymbol{\beta}^o = \mathbf{X}'\mathbf{L}'\mathbf{L}\mathbf{y} \quad \Rightarrow \quad \mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\boldsymbol{\beta}^o = \mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \tag{14}$$

which is (1). Then $\mathbf{X}\boldsymbol{\beta}^o$ for $\boldsymbol{\beta}^o$ of (14) is known as the generalized least squares estimator (GLSE) of $\mathbf{X}\boldsymbol{\beta}$:

$$\mathrm{GLSE}(\mathbf{X}\boldsymbol{\beta}) = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}\,. \tag{15}$$

3.3 Best linear unbiased estimation (BLUE)

An oft-used estimator is that which is a linear combination of the data which is unbiased for $\mathbf{t}'\mathbf{X}\boldsymbol{\beta}$ for any given $\mathbf{t}'$ and which, at the same time, has minimum variance. These requirements lead to

$$\mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta}) = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}\,, \tag{16}$$

which is exactly the same as $\mathrm{GLSE}(\mathbf{X}\boldsymbol{\beta})$ of (15).

3.4 Weighted least squares estimation (WLSE)

In the most general sense, weighted least squares would seem to be the appropriate name for $\hat{\mu}(\mathbf{W})$ of (9). Yet that name is, in fact, often used for GLSE. That being so we have

$$\mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta}) = \left\{ \begin{matrix} \mathrm{GLSE}(\mathbf{X}\boldsymbol{\beta}) \\ \mathrm{WLSE}(\mathbf{X}\boldsymbol{\beta}) \end{matrix} \right\} = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}\,, \tag{17}$$

A further name when $\mathbf{y} \sim \mathrm{N}(\mathbf{X}\boldsymbol{\beta},\ \mathbf{V})$ is $\mathrm{MLE}(\mathbf{X}\boldsymbol{\beta})$: maximum likelihood estimator of $\mathbf{X}\boldsymbol{\beta}$.

3.5 Extended weighted least squares estimation (EWLSE)

Whichever name is preferred in (17) one might also want a name for $\hat{\boldsymbol{\mu}}(\mathbf{W}) = \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-}\mathbf{X}'\mathbf{W}\mathbf{y}$ of (9) when $\mathbf{X} = \mathbf{C}\mathbf{W}\mathbf{X}$ for some $\mathbf{C}$. For then, as in Theorem 1, $\hat{\boldsymbol{\mu}}(\mathbf{W})$ is invariant to $(\mathbf{X}'\mathbf{W}\mathbf{X})^{-}$ and is unbiased for $\mathbf{X}\boldsymbol{\beta}$. As noted in Searle (1994), Plackett (1960) describes $(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{y}$ as coming from an "extended" principle of least squares, and so "extended weighted least squares estimator" (EWLSE) seems suitable for $\hat{\boldsymbol{\mu}}(\mathbf{W})$:

$$\mathrm{EWLSE}(\mathbf{X}\boldsymbol{\beta}) = \hat{\boldsymbol{\mu}}(\mathbf{W}) = \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-}\mathbf{X}'\mathbf{W}\mathbf{y}$$

where $\mathbf{X} = \mathbf{C}\mathbf{W}\mathbf{X}$ for some $\mathbf{C}$. In passing, we see that

$$\hat{\boldsymbol{\mu}}(\sigma^2\mathbf{I}) = \mathrm{OLSE}(\mathbf{X}\boldsymbol{\beta}) \qquad \text{and} \qquad \hat{\boldsymbol{\mu}}(\mathbf{V}^{-1}) = \mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta})\ ;$$

and in Section 5 we find for singular $\mathbf{V}$ that $\hat{\boldsymbol{\mu}}(\mathbf{V}^{-})$ when $\mathbf{V}\mathbf{V}^{-}\mathbf{X} = \mathbf{X}$ (so satisfying Theorem 1) is useful too.

It is also interesting to note that neither "generalized" nor "weighted" least squares are mentioned in some places where one might expect to find them: not in Aitken (1934), where general weights appear to have been first suggested, not in Plackett (1949, 1972), and not in Harter's (1983) article on least squares in the *Encyclopedia of Statistical Sciences* (Wiley, 1982-9). In that 10-volume encyclopedia there are, of course, articles on weighted least squares (Read, 1983) and on iteratively reweighted least squares (Rubin, 1988). Merriman (1877; p. 69 of 8'th edition, post 1911) does discuss weights, in terms of the square of either measure of precision or of probable error.

An interesting variation of $\mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta})$ is when $\mathbf{V}$ is functionally related to $\boldsymbol{\beta}$. Then $\mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta})$ gets used iteratively; Dear (1994) is an example.

4. ESTIMATION AND SAMPLING VARIANCES

In using a general non-negative $\mathbf{W}$ it is assumed that $\mathbf{W}$ is known, numerically. Thus $\hat{\boldsymbol{\mu}}(\mathbf{W})$ can be calculated as it stands: and for $\mathrm{var}(\mathbf{y}) = \mathbf{V}$,

$$\mathrm{var}[\hat{\boldsymbol{\mu}}(\mathbf{W})] = \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-}\mathbf{X}'\mathbf{W}\mathbf{V}\mathbf{W}\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-}\ .$$

If $\mathbf{W} = \sigma^2\mathbf{I}$ or $(1/\sigma^2)\mathbf{I}$

$$\mathrm{OLSE}(\mathbf{X}\beta) = \mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{y}$$

which can always be calculated. Its variance is

$$\mathrm{var}[\mathrm{OLSE}(\mathbf{X}\beta)] = \mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}'$$

and this is

$$\mathrm{var}[\mathrm{OLSE}(\mathbf{X}\beta)] = \mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}'\sigma^2 \quad \text{for} \quad \mathbf{V} = \sigma^2\mathbf{I}\ , \tag{18}$$

which is a common assumption. Thus to calculate (18) all that is needed is an estimate of σ^2 usually taken as

$$\hat{\sigma}^2 = \mathbf{y}'\mathbf{M}\mathbf{y}\,/\,(\mathrm{N} - \mathrm{r}_{\mathbf{X}}) \quad \text{for} \quad \mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}'\ ,$$

where $\mathbf{y}$ contains N data values and $\mathrm{r}_{\mathbf{X}}$ is the rank of $\mathbf{X}$.

The real difficulty of calculating an estimate of $\mathbf{X}\beta$ is when using (17):

$$\mathrm{BLUE}(\mathbf{X}\beta) = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^-\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}\ .$$

So long as $\mathbf{V}$ is known, calculation of this, and of its variance,

$$\mathrm{var}[\mathrm{BLUE}(\mathbf{X}\beta)] = \mathbf{X}'(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^-\mathbf{X}'$$

clearly presents no difficulty.

However, $\mathbf{V} = \mathrm{var}(\mathbf{y})$ is often not known: for example, in mixed models where elements of $\mathbf{V}$ are usually zero or various sums of variance components. One often estimates these from the same data vector $\mathbf{y}$ as is available for estimating $\mathbf{X}\beta$. At first thought an "obvious" way of calculating an estimate of $\mathrm{BLUE}(\mathbf{X}\beta) = \hat{\mu}(\mathbf{V}^{-1})$ of (17) is to use some estimate $\hat{\mathbf{V}}$ of $\mathbf{V}$. This yields

$$\hat{\mu}(\hat{\mathbf{V}}^{-1}) = \mathbf{X}(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^-\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{y}\ ,$$

which is, of course, readily calculable. But it must be emphasized that $\hat{\mu}(\hat{\mathbf{V}}^{-1})$ is not $\mathrm{BLUE}(\mathbf{X}\beta)$. And, in general, it has few, if any, attractive properties. It might be tempting to take the expected value of $\hat{\mu}(\hat{\mathbf{V}}^{-1})$ as $\mathbf{X}(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^-\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X}\beta$, but doing so would be to ignore that $\hat{\mathbf{V}}^{-1}$ obtained from $\mathbf{y}$ involves functions of $\mathbf{y}$, and this has to be taken into account in deriving $\mathrm{E}[\hat{\mu}(\hat{\mathbf{V}}^{-1})]$. However, Kackar and Harville (1981) have shown for variance components models that if the variance components are estimated as translation-invariant even-valued functions of $\mathbf{y}$ [e.g., $\mathrm{s}(\mathbf{y}) = \mathrm{s}(-\mathbf{y})$] for all $\mathbf{y}$ and β, then $\hat{\mu}(\hat{\mathbf{V}}^{-1})$ is indeed unbiased for $\mathbf{X}\beta$.

Similar difficulties exist for the sampling variance of $\hat{\mu}(\hat{\mathbf{V}}^{-1})$. That variance is neither

$$\mathbf{X}(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-}\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{V}\hat{\mathbf{V}}^{-1}\mathbf{X}(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-}\mathbf{X}' \quad \text{nor} \quad \mathbf{X}(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-}\mathbf{X}' ,$$

where the latter comes from the former by replacing $\mathbf{V}$ with $\hat{\mathbf{V}}$. These difficulties of $\hat{\mu}(\hat{\mathbf{V}}^{-1})$ are dealt with in Kackar and Harville (1984), and are summarized in Searle *et al.* (1992, p. 320).

5. SINGULAR V

5.1 Least squares estimation

For singular $\mathbf{V}$, take $\mathbf{V}^{-}$ as a generalized inverse of $\mathbf{V}$ to be symmetric and reflexive, i.e.,

$$\mathbf{V}^{-} = \mathbf{V}^{-'} \quad \text{and} \quad \mathbf{V}^{-}\mathbf{V}\mathbf{V}^{-} = \mathbf{V}^{-} . \tag{19}$$

If this is not the case we can, without loss of generality, in place of $\mathbf{V}^{-}$ use $\mathbf{V}^{\sim} = \mathbf{V}^{-}\mathbf{V}\mathbf{V}^{-'}$, which *is* symmetric and reflexive.

Now, despite there being an infinity of matrices $\mathbf{V}^{-}$ satisfying (19) for singular $\mathbf{V}$, using one of them for $\mathbf{W}$ in $\hat{\mu}(\mathbf{W})$ gives

$$\hat{\mu}(\mathbf{V}^{-}) = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-}\mathbf{y} . \tag{20}$$

This is not invariant to $\mathbf{V}^{-}$, a deficiency that could be defined away by using the unique Moore-Penrose inverse $\mathbf{V}^{+}$ in place of $\mathbf{V}^{-}$. However, the algebraic similarity of $\hat{\mu}(\mathbf{V}^{-})$ to the invariant $\hat{\mu}(\mathbf{V}^{-1}) =$ BLUE$(\mathbf{X}\beta)$ of (16) begs several questions, the first of which is answered by the following lemma from Searle (1994, Lemma 2).

Lemma If $\mathbf{V}^{-}$ is such that $\mathbf{V}\mathbf{V}^{-}\mathbf{X} = \mathbf{X}$ then $\mathbf{V}\mathbf{V}^{-}\mathbf{X} = \mathbf{X} \ \forall \ \mathbf{V}^{-}$, and $\mathbf{X}'\mathbf{V}^{-}\mathbf{X}$ is invariant to $\mathbf{V}^{-}$; and, for almost all $\mathbf{y} \sim (\mathbf{X}\beta, \ \mathbf{V})$, we also have $\mathbf{V}\mathbf{V}^{-}\mathbf{y} = \mathbf{y} \ \forall \ \mathbf{V}^{-}$ and $\mathbf{V}\mathbf{V}^{-}\mathbf{y}$ invariant to $\mathbf{V}^{-}$.

Immediately we have Theorem 2.

Theorem 2

$$\hat{\mu}(\mathbf{V}^{-}) = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-}\mathbf{y}$$

is invariant to $\mathbf{V}^{-}$ provided $\mathbf{V}\mathbf{V}^{-}\mathbf{X} = \mathbf{X}$.

The condition in this theorem, that $\mathbf{V}\mathbf{V}^{-}\mathbf{X} = \mathbf{X}$, satisfies the condition of Theorem 1, for the invariance of $\hat{\mu}(\mathbf{V}^{-})$ to $(\mathbf{X}'\mathbf{V}^{-}\mathbf{X})^{-}$ and the unbiasedness of $\hat{\mu}(\mathbf{V}^{-})$ for $\mathbf{X}\beta$. Thus providing $\mathbf{V}\mathbf{V}^{-}\mathbf{X} = \mathbf{X}$

we have $\hat{\mu}(\mathbf{V}^-)$ as an unbiased estimator of $\mathbf{X}\beta$ and as invariant to the choice of both generalized inverses $\mathbf{V}^-$ and $(\mathbf{X}'\mathbf{V}^-\mathbf{X})^-$. And once BLUE($\mathbf{X}\beta$) has been established for singular $\mathbf{V}$, we have in Theorem 3 that $\hat{\mu}(\mathbf{V}^-)$ equals that BLUE.

5.2 BLUE($\mathbf{X}\beta$) for singular V

We have seen for non-singular $\mathbf{V}$ that $\hat{\mu}(\mathbf{V}^{-1}) = \text{BLUE}(\mathbf{X}\beta)$. But that is no assurance for singular $\mathbf{V}$ that $\hat{\mu}(\mathbf{V}^-)$ is BLUE($\mathbf{X}\beta$). Indeed, for singular $\mathbf{V}$, we have yet to explicitly consider what BLUE($\mathbf{X}\beta$) is. Various expressions given by Albert (1961), Pukelsheim (1974) and Puntanen and Styan (1989) are discussed by Searle (1994) whose new development is as follows.

Write $\mathbf{y} = (\mathbf{I}-\mathbf{M})\mathbf{y} + \mathbf{My}$, and note that $\text{E}[(\mathbf{I}-\mathbf{M})\mathbf{y}] = \mathbf{X}\beta$ and $\text{E}(\mathbf{My}) = \mathbf{0}$, because $\text{E}(\mathbf{y}) = \mathbf{X}\beta$ and $\mathbf{MX} = \mathbf{0}$. Therefore any linear combination of $(\mathbf{I}-\mathbf{M})\mathbf{y}$ and $\mathbf{My}$ (being a linear function of elements of $\mathbf{y}$) can be unbiased for $\lambda'\mathbf{X}\beta$ only if the term in $(\mathbf{I}-\mathbf{M})\mathbf{y}$ is $\lambda'(\mathbf{I}-\mathbf{M})\mathbf{y}$. For finding the BLUE of $\lambda'\mathbf{X}\beta$ this leads to asking "for what vector τ' does adding $\tau'\mathbf{My}$ to $\lambda'(\mathbf{I}-\mathbf{M})\mathbf{y}$ yield BLUE($\lambda'\mathbf{X}\beta$)?" This is achieved by choosing τ' to minimize $\text{var}[\lambda'(\mathbf{I}-\mathbf{M})\mathbf{y} + \tau'\mathbf{My}]$, and yields $\tau = -(\mathbf{MVM})^-\mathbf{MV}(\mathbf{I}-\mathbf{M})\lambda$ which in turn gives (see Searle, 1994)

$$\text{BLUE}(\mathbf{X}\beta) = (\mathbf{I}-\mathbf{M})[\mathbf{I}-\mathbf{VM}(\mathbf{MVM})^-\mathbf{M}]\mathbf{y}\ . \tag{21}$$

Pukelsheim (1974) has this result except for having $(\mathbf{MVM})^+$ where (21) has $(\mathbf{MVM})^-$. The latter suggests that (21) is not invariant to $(\mathbf{MVM})^-$; but, in fact, it is, as in Searle (*loc. cit.*). Moreover, that same paper also shows that (21) can be simplified to

$$\text{BLUE}(\mathbf{X}\beta) = \mathbf{y} - \mathbf{VM}(\mathbf{MVM})^-\mathbf{My}\ . \tag{22}$$

This is certainly a simpler expression than (21). And note that in (22) the term $\mathbf{M}(\mathbf{MVM})^-\mathbf{M}$ is a generalized inverse of $\mathbf{MVM}$, say $(\mathbf{MVM})^*$, but it is not unique, as is $(\mathbf{MVM})^+$ in the equality $\mathbf{M}(\mathbf{MVM})^+\mathbf{M} = (\mathbf{MVM})^+$. Hence not any generalized inverse of $\mathbf{MVM}$ can be used in place of $\mathbf{M}(\mathbf{MVM})^+\mathbf{M}$ in (22), because not every generalized inverse of $\mathbf{MVM}$ has $\mathbf{M}$ as a left and a right factor, and that is an essential feature in (22).

5.3 Three estimators for $\mathbf{X}\beta$

For singular $\mathbf{V}$ we now have three estimators for $\mathbf{X}\beta$: (i) $\hat{\mu}(\mathbf{V}^-)$ of (22) when $\mathbf{VV}^-\mathbf{X} = \mathbf{X}$,

(ii) BLUE($\mathbf{X}\beta$) of (21), and (iii) OLSE($\mathbf{X}\beta$) = $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y}$ of (11), which ignores $\mathbf{V}$, be it singular or not. Having three estimators prompts the question "when are they equal?" It is answered in the next three theorems. Theorem 3 indicates when $\hat{\mu}(\mathbf{V}^{-})$ equals BLUE($\mathbf{X}\beta$), Theorem 4 shows when BLUE($\mathbf{X}\beta$) equals OLSE($\mathbf{X}\beta$), and Theorem 5 combines Theorems 3 and 4 in specifying when all three are equal.

Theorem 3

$$\hat{\mu}(\mathbf{V}^{-}) = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-}\mathbf{y} \quad \text{equals} \quad \text{BLUE}(\mathbf{X}\beta) = \mathbf{y} - \mathbf{VM}(\mathbf{MVM})^{-}\mathbf{My}$$

if and only if $\mathbf{VV}^{-}\mathbf{X} = \mathbf{X}$.

When using only Moore-Penrose inverses, early proof of the sufficiency part of this theorem is due to Rao and Mitra (1971) and of the necessity part to Pukelsheim (1974). New proofs, which are somewhat shorter than theirs, and which do not rely on Moore-Penrose inverses and their uniqueness property, are given in Searle (1994).

Theorem 4

$$\text{BLUE}(\mathbf{X}\beta) = \mathbf{y} - \mathbf{VM}(\mathbf{MVM})^{-}\mathbf{My} \quad \text{equals} \quad \text{OLSE}(\mathbf{X}\beta) = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y}$$

if and only if $\mathbf{VX} = \mathbf{XB}$ for some $\mathbf{B}$.

This result, due to Zyskind (1967), is part of numerous equivalent results; see, for example, Puntanen and Styan (1989).

Theorem 5

$$\hat{\mu}(\mathbf{V}^{-}) = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-}\mathbf{y} \quad \text{equals} \quad \text{OLSE}(\mathbf{X}\beta) = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y}$$

if and only if $\mathbf{VV}^{-}\mathbf{X} = \mathbf{X}$ and $\mathbf{VX} = \mathbf{XB}$ for some $\mathbf{B}$.

5.4 Non-singular V

BLUE($\mathbf{X}\beta$) = $\mathbf{y} - \mathbf{VM}(\mathbf{MVM})^{-}\mathbf{My}$ looks to be so different from BLUE($\mathbf{X}\beta$) = $\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$ when $\mathbf{V}$ is non-singular that it is interesting to see the reduction of the former to the latter. To do this, define

$$\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1} = \mathbf{P}' \quad \text{with} \quad \mathbf{PX} = \mathbf{0}\,. \tag{23}$$

Then, with $\mathbf{P}$ and $\mathbf{M}$ being symmetric

$$\mathbf{PM} = \mathbf{P}(\mathbf{I} - \mathbf{XX}^{+}) = \mathbf{P} = \mathbf{MP} = \mathbf{PMP} \,. \tag{24}$$

Therefore, using $\mathbf{MX} = \mathbf{0}$

$$\mathbf{MVMPMVM} = \mathbf{MVPVM} = \mathbf{MVM} \,,$$

which allows us to write

$$\mathbf{P} = (\mathbf{MVM})^{-} \,. \tag{25}$$

Hence the result for $\mathbf{V}$ singular,

$$\text{BLUE}(\mathbf{X}\boldsymbol{\beta}) = \mathbf{y} - \mathbf{VM}(\mathbf{MVM})^{-}\mathbf{My} \text{ of } (21) \,,$$

becomes for $\mathbf{V}$ non-singular

$$\begin{aligned}
\text{BLUE}(\mathbf{X}\boldsymbol{\beta}) &= \mathbf{y} - \mathbf{VMPMy} \,, \quad \text{from} \;\; (25) \\
&= \mathbf{y} - \mathbf{VPy} \,, \quad \text{from} \;\; (24) \\
&= \mathbf{y} - [\mathbf{y} - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}] \,, \;\; \text{from} \;\; (23) \\
&= \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \\
&= \text{BLUE}(\mathbf{X}\boldsymbol{\beta}) \text{ of } (17) \,.
\end{aligned}$$

6. MIXED MODEL EQUATIONS

In (4) we have $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$. Defining $\boldsymbol{\epsilon}$ as $\boldsymbol{\epsilon} = \mathbf{y} - E(\mathbf{y})$ then gives the familiar model equation

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

Clearly, $E(\boldsymbol{\epsilon}) = \mathbf{0}$; and defining $\text{var}(\boldsymbol{\epsilon}) = \mathbf{V}$ gives $\text{var}(\mathbf{y}) = \mathbf{V}$.

Suppose $\mathbf{V}$ could be explained by modeling $\boldsymbol{\epsilon}$ as

$$\boldsymbol{\epsilon} = \mathbf{Zu} + \mathbf{e} \,,$$

where $\mathbf{Z}$ is known, and $\mathbf{u}$ and $\mathbf{e}$ are vectors of random variables with the following properties:

$$\begin{gathered}
E(\mathbf{u}) = \mathbf{0} \,, \qquad E(\mathbf{e}) = \mathbf{0} \,, \\
\text{var}(\mathbf{u}) = \mathbf{D} \,, \qquad \text{var}(\mathbf{e}) = \mathbf{R} \,, \qquad \text{and} \qquad \text{cov}(\mathbf{u},\; \mathbf{e}') = \mathbf{0} \,.
\end{gathered} \tag{26}$$

Hence

$$\mathbf{V} = \text{var}(\mathbf{y}) = \text{var}(\boldsymbol{\epsilon}) = \text{var}(\mathbf{ZU} + \mathbf{e}) = \mathbf{ZDZ}' + \mathbf{R} \tag{27}$$

with

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Zu} + \mathbf{e} \,. \tag{28}$$

Equations (26), (27) and (28) represent the general form of the usual analysis of variance mixed model. A simple example is the randomized complete block experiment where the observation on the i'th treatment in the j'th block has model equation

$$y_{ij} = \mu + \alpha_i + \beta_j + e_{ij} \, . \tag{29}$$

In this equation μ and α_i are constants (fixed effects). The β_js are random variables (random block effects) usually with $E(\beta_j) = 0$, $\text{var}(\beta_j) = \sigma^2_\beta \; \forall$ j, $\text{cov}(\beta_j, \beta_{j'}) = 0 \; \forall \; j \neq j'$, and $\text{cov}(\beta_j, \; e_{st}) = 0 \; \forall$ j, s and t. Thus (29) is a particular case of (28): the μ and α_is of (29) constitute $\boldsymbol{\beta}$ of (28) and the β_js of (29) constitute **u** of (28).

The extension of (29) to (28) for several random effects factors is to have $\mathbf{u}'$ of (28) partitioned as

$$\mathbf{u}' = [\mathbf{u}'_1 \;\; \mathbf{u}'_2 \;\; \cdots \;\; \mathbf{u}'_i \;\; \cdots \;\; \mathbf{u}'_r] \qquad \text{and} \qquad \mathbf{Z} = [\mathbf{Z}_1 \;\; \mathbf{Z}_2 \;\; \cdots \;\; \mathbf{Z}_i \;\; \cdots \;\; \mathbf{Z}_r]$$

conformable with **u**. Each $\mathbf{u}_i$ in **u** is a vector of q_i random effects occurring in the data corresponding to a random effects factor with, just like (26),

$$E(\mathbf{u}_i) = \mathbf{0} \, , \quad \text{var}(\mathbf{u}_i) = \sigma^2_i \mathbf{I}_{q_i} \, , \quad \text{cov}(\mathbf{u}_i, \; \mathbf{e}') = \mathbf{0} \quad \text{and} \quad \text{cov}(\mathbf{u}_i, \; \mathbf{u}'_j) = \mathbf{0} \tag{30}$$

$\forall \; i \neq j$. Then **D** of (26) is

$$\text{var}(\mathbf{u}) = \mathbf{D} = \left\{_d \; \sigma^2_i \mathbf{I}_{q_i} \right\}_{i=1}^{r} \, , \tag{31}$$

a block diagonal matrix of matrices

$$\sigma^2_i \mathbf{I}_{q_i} \, , \quad \text{of order } q_i \times q_i \, , \quad \text{for } i = 1, 2, \cdots, r \, .$$

Thus from (27), (30) and (31)

$$\mathbf{V} = \text{var}(\mathbf{y}) = \mathbf{ZDZ}' + \mathbf{R} = \sum_{i=1}^{r} \mathbf{Z}_i \mathbf{Z}'_i \sigma^2_i + \mathbf{R} \, . \tag{32}$$

And usually

$$\mathbf{R} = \sigma^2_e \mathbf{I} \, . \tag{33}$$

With these details for **V** we still have

$$\text{BLUE}(\mathbf{X}\boldsymbol{\beta}) = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \tag{34}$$

of (17) with **V** of (32) taken as non-singular. However, in 1958, C.R. Henderson, professor of animal breeding at Cornell University, showed in Henderson et al. (1959) that an alternative method for calculating $\text{BLUE}(\mathbf{X}\boldsymbol{\beta})$ as $\mathbf{X}\boldsymbol{\beta}^o$ is to calculate $\boldsymbol{\beta}^o$ by solving

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{D}^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}^o \\ \tilde{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix} . \tag{35}$$

Henderson derived these equations by maximizing what he thought was a likelihood function, but which in fact was a distribution function. The motivation was an exercise in Mood (1950) – see Searle *et al.* (1992, Sec. 7.1).

Equations (35) have come to be known as the mixed model equations (MMEs). Whatever their source they have the following useful features.

1. The solution to the MMEs for $\boldsymbol{\beta}^o$ is identical to

$$\boldsymbol{\beta}^o = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}\ , \tag{36}$$

a result that depends on the now well-known identity

$$(\mathbf{ZDZ}' + \mathbf{R})^{-1} = \mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{Z}(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{D}^{-1})^{-1}\mathbf{Z}'\mathbf{R}^{-1}\ . \tag{37}$$

It was not always well known: see Henderson and Searle (1981) who trace an interesting history of this identity. And (37) is $\mathbf{V}^{-1}$, from (27).

2. It is $\mathbf{u}$ representing random effects that introduces $\mathbf{D}$ into $\mathbf{V}$. Suppose $\mathbf{u}$ represented fixed effects. Then $\mathbf{D}$ would not exist, var($\mathbf{y}$) would be $\mathbf{R}$ and the MMEs of (34) would reduce to

$$\begin{bmatrix} \mathbf{X}' \\ \mathbf{Z}' \end{bmatrix} \mathbf{R}^{-1} [\mathbf{X} \quad \mathbf{Z}] \begin{bmatrix} \boldsymbol{\beta}^* \\ \mathbf{u}^* \end{bmatrix} = \begin{bmatrix} \mathbf{X}' \\ \mathbf{Z}' \end{bmatrix} \mathbf{R}^{-1}\mathbf{y}\ . \tag{38}$$

These are precisely the normal equations for weighted least squares estimators $\boldsymbol{\beta}^*$ and $\mathbf{u}^*$ (i.e., Aitken's equations) using $\mathbf{R}^{-1}$ as the weight matrix.

3. Thus on comparing (34) and (37) we see that the MMEs of (35) are just weighted least squares equations (38) adapted by adding $\mathbf{D}^{-1}$ to the $\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z}$ coefficient of $\mathbf{u}^*$ in the second equation of (38).

4. A computational advantage of the MMEs for deriving $\boldsymbol{\beta}^o$ is that they usually involve much less effort for matrix inversion than does (36). This is because in (36) the inverse is needed of $\mathbf{V}$, a matrix of order N, the number of observations in the data vector. But in the MMEs of (35), the inverses $\mathbf{R}^{-1}$ and $\mathbf{D}^{-1}$ are easy when, as is usually the case, $\mathbf{R}$ and $\mathbf{D}$ are diagonal, as in (32) and (31); and the order of the MMEs is p + q, the total number of levels of all the fixed and random effects in the data – a number which is usually much less than N.

5. Finally, the solution for $\tilde{\mathbf{u}}$ to the MMEs is useful – very useful. It is the best linear unbiased predictor (BLUP) of $\mathbf{u}$, expressible in its simplest form as

$$\tilde{\mathbf{u}} = \text{BLUP}(\mathbf{u}) = \mathbf{DZ'V}^{-1}(\mathbf{y} - \mathbf{X}\beta^{o}) \ .$$

It is also a Bayes estimator of $\mathbf{u}$; and, indeed, there are numerous derivations of $\mathbf{u}$ (Searle *et al.*, 1993, Sec. 7.4 and 7.5). One of the greatest uses of BLUP($\mathbf{u}$), and right here in New Zealand, is for assessing the genetic value of young dairy bulls from the milk production records of a small sample of their daughters. This assessment is used to select bulls having the highest $\mathbf{u}$-values, and those bulls will then be used to sire, via artificial insemination, large numbers of cows throughout the country's dairy farms. This practice, in use for some thirty years, has definitely contributed to impressive gains in per-cow milk production – in numerous countries around the world. **Moral:** statistics *is* useful!

REFERENCES

Adrain, R. (1808). ... *Analyst* **1**, 93-109.

Albert, A. (1967). The Gauss-Markov theorem for regression model with possibly singular covariances. *Siam J. Applied. Mathematics* **24**, 182-187.

Dear, K. B. G. (1994). Iterative generalized least squares for meta-analysis of survival data at multiple times. *Biometrics* **50**, 989-1002.

Galle, A. (1924). *Über die Geodätischer Arbeiten von Gauss.* C.F. Gauss Werke, **11**(2). Berlin: Springer.

Harter, H. L. (1983). Least squares. In *Encyclopedia of Statistical Sciences*, Eds. S. Kotz, N.L. Johnson and C.R. Read, Wiley, NY, 593-598.

Henderson, C. R., Kempthorne, O., Searle, S. R. and von Krosigk, C. N. (1959). Estimation of environmental and genetic trends from records subject to culling. *Biometrics* **13**, 192-218.

Henderson, H. V. and Searle, S. R. (1981). On deriving the inverse of a sum of matrices. *SIAM Review* **23**, 53-60.

Kackar, R. N. and Harville, D. A. (1981). Unbiasedness of two-stage estimation and prediction procedures for mixed linear models. *Communications in Statistics A: Theory and Methods* **10**, 1249-1261.

Kackar, R. N. and Harville, D. A. (1984). Approximation for standard errors of estimators of fixed and random effects in mixed linear models. *J. American Statistical Association* **79**, 853-867.

Merriman, M. (1877). A list of writings relating to the method of least squares, with historical and critical notes. *Transactions Connecticut Academy of Arts and Sciences* **4**, 151-232.

Mood, A. M. (1950). *Introduction to the Theory of Statistics.* McGraw-Hill, New York.

Plackett, R. L. (1949). A historical note on the method of least squares. *Biometrika* **36**, 458-460.

Plackett, R. L. (1960). *Principles of Regression Analysis.* Clarendon Press: Oxford.

Plackett, R. L. (1972). Studies in the history of probability and statistics. XXIX: The discovery of the method of least squares. *Biometrika* **59**, 239-251.

Pukelsheim, F. (1974). Schätzen von Mittelwert und Streuungsmatrix in Gauss-Markoff-Modellen. *Diplomarbet*, Freiburg im Breisgau.

Puntanen, S. and Styan, G. P. H. (1989). On the equating of the ordinary least squares estimator and the best linear unbiased estimator. *The American Statistician* **43**, 153-161.

Rao, C. R. and Mitra, S. K. (1971). Further contributions to the theory of generalized inverses of matrices and its applications. *Sankhyā, Series A* **33**, 289-230.

Read, C. R. (1988). Weighted least squares. In *Encyclopedia of Statistical Sciences*, Eds. S. Kotz, N.L. Johnson and C.R. Read, Wiley & Sons, **9**, 576-578.

Rubin, D. B. (1983). Iteratively reweighted least squares. In *Encyclopedia of Statistical Sciences*, Eds. S. Kotz, N.L. Johnson and C.R. Read, Wiley & Sons, **4**, 272-275.

Searle, S. R. (1994). Extending some results and proofs for the singular linear model. *Linear Algebra and Its Applications*, **210**, 139-151.

Searle, S. R., Casella, G. and McCulloch, C. E. (1992). *Variance Components*, Wiley & Sons, N. Y.

Zyskind, G. (1967). On canonical forms, non-negative covariance matrices and best and simple least squares linear estimators in linear models. *Annals Mathematical Statistics* **38**, 1092-1109.