ON LEAST SQUARES AND BEST LINEAR UNBIASED ESTIMATION

Shayle R. Searle* and Friedrich Pukelsheim

Biometrics Unit, Cornell University, Ithaca, N.Y., and Institut für Mathematik, Universität Augsburg, Augsburg, Federal Republic of Germany

ABSTRACT

Least squares estimation (in its several forms) and best linear unbiased estimation are reviewed; briefly, for a linear model having non-singular dispersion matrix, and in some detail for the singular case, for which certain equivalences among the estimation methods are established with some new and shorter proofs.

1. INTRODUCTION

1.1 The model

We deal with the linear model $y = X\beta + e$ for y of order $N \times 1$, with expected value $E(y) = X\beta$ for β being a $p \times 1$ vector of parameters and X an $N \times p$ matrix of known values. e is a vector of random errors with $E(e) = 0$ and dispersion matrix V that is symmetric and non-negative definite (n.n.d.). Thus the model is

$$y = X\beta + e, \quad E(y) = X\beta \quad \text{and} \quad \text{var}(y) = V, \tag{1}$$

with

$$V = V' \text{ of order } N, \text{ n.n.d., of rank } r_V; \tag{2}$$

and

$$X, \text{ of order } N \times p, \text{ has rank } r_X \leq p < N. \tag{3}$$

Frequent use is made of the so-called "hat" matrix

$$H = X(X'X)^-X' = XX^+ = H' = H^2 \quad \text{with} \quad HX = X$$

and of

$$M = M' = M^2 = I - XX^+ = I - X(X'X)^-X' \quad \text{with} \quad MX = 0 \qquad (4)$$

where X^+ is the Moore-Penrose inverse of X, and $(X'X)^-$ is a generalized inverse of $X'X$ (see Appendix, Section 6.1).

The paper deals with estimation of the mean vector $E(y) = X\beta$, all elements of which are estimable, as is every linear combination of them, $\lambda'X\beta$ for any vector λ'. We begin with ordinary least squares estimation (OLSE), followed by weighted least squares estimation (WLSE) using an arbitrary n.n.d. weight matrix W. Neither of these estimation methods involves the dispersion matrix $\text{var}(y) = V$. The latter is involved in generalized least squares estimation (GLSE), in maximum likelihood estimation (MLE), and in best linear unbiased estimation (BLUE). These are dealt with in Section 2 for V non-singular and in Section 3 for V singular. Section 4 deals with relationships among the methods, and Section 5 discusses what is called a pseudo BLUE (P-BLUE). Section 6 is an appendix containing pertinent matrix results (most of them well known). They have equation and lemma numbers prefixed with A.

1.2 OLSE: ordinary least squares estimation

Ordinary least squares estimation of $X\beta$ is based on minimizing

$$S = (y - X\beta)'(y - X\beta) \qquad (5)$$

with respect to β. This leads to the well-known normal equations

$$X'X\beta^0 = X'y \qquad (6)$$

where $X\beta^0$ is the corresponding estimator of $X\beta$. This will be denoted OLSE:

$$\text{OLSE} = X\beta^0 = X(X'X)^-X'y = XX^+y = (I - M)y \ . \qquad (7)$$

Through having $E(y) = X\beta$ and $MX = 0$, the unbiasedness of (7) for $X\beta$ is self-evident. The dispersion matrix of (7) is

$$\text{var(OLSE)} = XX^+VXX^+ . \tag{8}$$

1.3 WLSE: weighted least squares estimation

Weighted least squares using an arbitrary non-negative definite (n.n.d.) weight matrix W (where, through being n.n.d. it can be factored as $W = T'T$ for T real and of full row rank r_W) involves minimizing

$$S_W = (y - X\beta)'W(y - X\beta) = (Ty - TX\beta)'(Ty - TX\beta). \tag{9}$$

Based on (6) this gives

$$X'T'TX\beta^0 = X'T'Ty, \text{ i.e., } X'WX\beta^0 = X'Wy . \tag{10}$$

Then $X\beta^0$ derived from this is the weighted least squares estimator, to be denoted $m(W)$, and is

$$m(W) = X(X'WX)^-X'Wy . \tag{11}$$

We use the symbol $m(W)$ because of being interested in special values of W; e.g., OLSE of (7) is $m(I)$. The dispersion matrix of (11) is

$$\text{var}[m(W)] = X(X'WX)^-X'WVWX[(X'WX)^-]'X'. \tag{12}$$

The occurrence of $(X'WX)^-$ in (11) means that $m(W)$ of (11) is not necessarily invariant to what generalized inverse is used for $(X'WX)^-$. Nor is $m(W)$ necessarily unbiased for $X\beta$. The following theorem provides conditions under which invariance and unbiasedness are assured.

Theorem 1. A necessary and sufficient condition for $m(W)$ to be both invariant to $(X'WX)^-$ and unbiased for $X\beta$ is that $X = CWX$ for some C.

Proof. First, sufficiency: if $X = CWX = CT'TX$ then from (11)

$$m(W) = CT'TX(X'T'TX)^-X'T'Ty = C'T'TX(TX)^+Ty , \tag{13}$$

the second equality coming from using TX for X in (A6); and it demonstrates the invariance of $m(W)$ to $(X'T'TX)^-$. Similarly, also using TX for X in (A4),

$$E[m(W)] = CT'TX(X'T'TX)^-X'T'TX\beta = CWX\beta = X\beta .$$

This is sufficiency established.

Second, to show necessity, for any matrix N using $N'N$ as A in (A7) and then post-multiplying the result by N' gives, for arbitrary P and Q,

$$(N'N)^{\sim}N' = (N'N)^-N'N(N'N)^-N' + [I - (N'N)^-N'N]PN' + Q[I - N'N(N'N)^-]N'$$

$$= (N'N)^-N' + [I - (N'N)^-N'N]PN' , \tag{14}$$

on using N for X in the transpose of (A4). Now use TX for N in (14) and simultaneously pre-multiply (14) by X and post-multiply it by Ty. This yields

$$X(X'WX)^{\sim}X'Wy = X(X'WX)^-X'Wy + X[I - (X'WX)^-X'WX]PWy \ \forall \ P . \tag{15}$$

But taking $m(W) = X(X'WX)^-X'Wy$ as invariant to $(X'WX)^-$, reduces (15) to

$$[X - X(X'WX)^-X'WX]PWy = 0 \ \forall \ P . \tag{16}$$

Therefore, by Lemma A2, with $Wy \neq 0$,

$$X - X(X'WX)^-X'WX = 0 . \tag{17}$$

Also, if $m(W)$ is unbiased, (11) gives $X(X'WX)^-X'WX\beta = X\beta \ \forall \ \beta$, and this too, implies (17). Hence if $m(W)$ is invariant to $(X'WX)^-$ and/or is unbiased for $X\beta$, (17) holds and so $X = X(X'WX)^-X'WX = CWX$ for $C = X(X'WX)^-X'$. Q.E.D.

$m(W)$ is the weighted least squares estimator of $X\beta$ based on the (n.n.d.) weight matrix W. It is a generalization of the Aitken (1935) estimator that has X of full column rank and W non-singular: $X(X'W^{-1}X)^{-1}X'W^{-1}y$. In this case there is no problem about invariance; nor is there with

$$m(W^{-1}) = X(X'W^{-1}X)^-X'W^{-1}y ,$$

for then $X = CW^{-1}X$ of Theorem 1 is satisfied for $C = W$. Difficulty arises only when X has less than full column rank and W is singular, for then the desirable properties of invariance and unbiasedness are met only when the condition of Theorem 1 is satisfied, namely when there exists a C such that $X = CWX$. Nevertheless, in view of $m(W)$ being a generalization of the Aitken estimator, which has often been called a weighted least squares estimator (WLSE), and because this name and the name generalized least squares estimator (GLSE) have each been used for a variety of cases, and sometimes interchangeably (see, e.g., Puntanen and Styan, 1989) — for these reasons, and because of the generality of $m(W)$, there would be merit in giving $m(W)$ a name. In keeping with Plackett (1960), who describes $(X'WX)^{-1}X'Wy$ as coming from an "extended" principle of least squares, we therefore might call $m(W)$ the EWLSE, "estimated weighted least squares estimator."

Of course if one wishes to use $m(W)$ without the condition of the theorem being satisfied, the invariance property could be defined away by always using $X(X'WX)^{+}X'Wy$. In doing so, an extension of a result in Baksalary and Kala (1983) can be used.

Lemma For

$$\overset{o}{b} = (X'WX)^{+}X'Wy \quad \text{and} \quad \overset{\cdot}{b} = (X'X)^{+}X'y \tag{18}$$

a necessary and sufficient conditon for $X(X'WX)^{+}X'Wy = X\overset{o}{b}$ to equal OLSE $= X\overset{\cdot}{b}$ is $\overset{o}{b} = \overset{\cdot}{b}$.

Proof. Sufficiency is obvious. Proving necessity starts with $X\overset{o}{b} = X\overset{\cdot}{b}$ so that

$$X^{+}X\overset{o}{b} = X^{+}X\overset{\cdot}{b} . \tag{19}$$

But $X^{+} = X'(XX')^{-}X(X'X)^{-}X'$ of (A5), which is $X^{+} = X'(XX')^{-}XX^{+}$, has two consequences:

for any R, $(RX)^+ = X'U$ for some U; and $X^+XX' = X'$.

Hence each b in (18) has the form

$$b = X'Uy = X'u \text{ for some } U \text{ and } u .$$

Therefore each side of (19) has the form

$$X^+Xb = X^+XX'u = X'u = b .$$

Thus $\overset{o}{b} = \overset{.}{b}$.

Q.E.D.

Despite this lemma, confining attention to $X(X'WX)^+XWy$ seems restrictive

and we prefer to direct attention to $m(W)$ and some special cases thereof.

In doing so we deal first with the easy and commonly discussed case of non-

singular V.

2. NON-SINGULAR V

2.1 GLSE: generalized least squares estimation

Although the Aitken estimator is sometimes called the generalized

least squares estimator (GLSE), we reserve this name for the special case

of $m(W)$ when V^{-1} is used in place of W:

$$\text{GLSE} = m(V^{-1}) = X(X'V^{-1}X)^-X'V^{-1}y . \tag{20}$$

with dispersion matrix, from (12),

$$\text{var}[m(V^{-1})] = X(X'V^{-1}X)^-X' . \tag{21}$$

Clearly, the condition $X = CWX$ of Theorem 1 with $W = V^{-1}$ is satisfied for

$C = V$, and so (20) is invariant to its generalized inverse and is unbiased

for $X\beta$.

2.2 MLE: maximum likelihood estimation

When elements of y have a multivariate normal distribution, i.e., $y \sim$

$\mathcal{N}(X\beta, V)$, maximum likelihood estimation is based on maximizing

$$(2\pi)^{\frac{1}{2}N}|V|^{\frac{1}{2}} \exp -\tfrac{1}{2}(y - X\beta)'V^{-1}(y - X\beta) .$$

This leads to minimizing S_W of (9) with $\mathbf{W}$ replaced by $\mathbf{V}^{-1}$. Hence the estimation equations are (11) with the same replacement, so that the maximum likelihood estimator (MLE) is

$$\mathbf{m}(\mathbf{V}^{-1}) = \text{MLE} = \text{GLSE} = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \ . \tag{22}$$

Hence the MLE has the properties of the GLSE (particularly invariance and unbiasedness). Moreover,

$$\text{MLE} \sim \mathcal{N}[\mathbf{X\beta}, \ \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'] \ .$$

Thus MLE also has all the properties that normality implies, additional to those of GLSE, which demands no assumption of normality on $\mathbf{y}$.

2.3 BLUE: best linear unbiased estimation

This method of estimation involves deriving a vector $\mathbf{t}'$ for given $\boldsymbol{\lambda}'$ such that (i) $\mathbf{t}'\mathbf{y}$ is unbiased for $\boldsymbol{\lambda}'\mathbf{X\beta}$ and (ii) from all unbiased estimators $\mathbf{t}'\mathbf{y}$ of $\boldsymbol{\lambda}'\mathbf{X\beta}$ the one to be described as "best" is that which has minimum variance. Requirement (i) demands having

$$\mathbf{t}'\mathbf{X\beta} = \boldsymbol{\lambda}'\mathbf{X\beta} \ \forall \ \boldsymbol{\beta}; \ \text{i.e.,} \ \mathbf{t}'\mathbf{X} = \boldsymbol{\lambda}'\mathbf{X}. \tag{23}$$

Requirement (ii) demands choosing $\mathbf{t}$ to simultaneously satisfy (23) and minimize $v(\mathbf{t}'\mathbf{y}) = \mathbf{t}'\mathbf{Vt}$. Using Lagrange multipliers $2\boldsymbol{\theta}'$ this means minimizing $\mathbf{t}'\mathbf{Vt} + 2\boldsymbol{\theta}'(\mathbf{X}'\boldsymbol{\lambda} - \mathbf{X}'\mathbf{t})$ with respect to $\mathbf{t}$ and $\boldsymbol{\theta}$, which leads to equations

$$\mathbf{Vt} = \mathbf{X\theta} \tag{24}$$

and

$$\mathbf{t}'\mathbf{X} = \boldsymbol{\lambda}'\mathbf{X} \ . \tag{25}$$

With $\mathbf{V}$ non-singular, (24) is consistent for a solution for $\mathbf{t}$:

$$\mathbf{t} = \mathbf{V}^{-1}\mathbf{X\theta} \ . \tag{26}$$

Substituting (26) into (25) gives

$$\mathbf{X}'\mathbf{V}^{-1}\mathbf{X\theta} = \mathbf{X}'\boldsymbol{\lambda}, \ \text{i.e.,} \ \boldsymbol{\theta} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\boldsymbol{\lambda} \ . \tag{27}$$

Then using θ from (27) in (26) gives

$$t' = \lambda'X(X'V^{-1}X)^-X'V^{-1} \ . \tag{28}$$

Hence the BLUE of $\lambda'X\beta$, being $t'y$, is

$$(\text{BLUE of } \lambda'X\beta) = \lambda'X(X'V^{-1}X)^-X'V^{-1}y \ . \tag{29}$$

Therefore, on letting λ' be successive rows of I_N, the BLUE of $X\beta$ is

$$m(V^{-1}) = \text{BLUE} = X(X'V^{-1}X)^-X'V^{-1}y \ . \tag{30}$$

Immediately it is seen that

$$m(V^{-1}) = \text{BLUE} = \text{MLE} = \text{GLSE} \ . \tag{31}$$

3. SINGULAR V

Some estimators of $X\beta$ when V is singular have the form of $m(W)$ using V^+ or V^-. But in some cases the BLUE requires explicit use of neither V^- nor V^+, as in Section 3.4.

3.1 GLSE and MLE using the Moore-Penrose inverse, V^+.

With V being singular but n.n.d., factor V as $V = L'L$ for L of full row rank, the rank of V. Then defining T (different from the T in Section 1.3) as

$$T = (LL')^{-1}L, \text{ gives } TVT' = I \text{ and } \overset{+}{V} = T'T \ , \tag{32}$$

and the vector of transformed data

$$w = Ty = TX\beta + Te \text{ has } \text{var}(w) = TVT' = I \ . \tag{33}$$

Hence OLSE of $X\beta$ from w of (33) gives, similar to (10), $X'T'TX\beta^0 = X'T'w = X'T'Ty$ which, from (32), is $X'V^+X\beta^0 = X'V^+y$. Hence, from (11) the estimator is

$$m(V^+) = X(X'V^+X)^-X'V^+y \ , \tag{34}$$

with variance $\text{var}[m(V^+)] = X(X'V^+X)^-X'$.

From Theorem 1, $m(V^+)$ is invariant to $(X'V^+X)^-$ if and only if $X = CV^+X$ for some C.

Clearly, $\mathbf{m}(\mathbf{V}^+)$ is the GLSE of $\mathbf{X\beta}$ that has been obtained from $\mathbf{Ty}$ of (33); and under normality it is also the MLE, as can be obtained from the likelihood of $\mathbf{Ty} \sim \mathcal{N}(\mathbf{X\beta}, \mathbf{I})$. Thus

$$\mathbf{m}(\mathbf{V}^+) = \text{MLE} = \text{GLSE} = \mathbf{X}(\mathbf{X}'\mathbf{V}^+\mathbf{X})^-\mathbf{X}'\mathbf{V}^+\mathbf{y} \ . \tag{35}$$

3.2 BLUE when $\mathbf{VV}^-\mathbf{X} = \mathbf{X}$ is $\mathbf{m}(\mathbf{V}^-)$.

Specification of best linear unbiased estimation is exactly the same for singular $\mathbf{V}$ as for non-singular $\mathbf{V}$ in (23), and through using Lagrange multipliers leads to equations (24) and (25) that have to be solved for $\mathbf{t}$ and $\mathbf{\theta}$. For non-singular $\mathbf{V}$ this is done very easily, starting with $\mathbf{t} = \mathbf{V}^{-1}\mathbf{X\theta}$ of (26), obtained from (24). But for singular $\mathbf{V}$ the solving of (24) and (25) can be somewhat different because it does not always involve just the simple replacement of $\mathbf{V}^{-1}$ by $\mathbf{V}^-$. Indeed, since the solution is a form of $\mathbf{m}(\mathbf{W})$ when $\mathbf{VV}^-\mathbf{X} = \mathbf{X}$, and is quite different in appearance when that equality is not assumed, we consider two cases: first, when $\mathbf{VV}^-\mathbf{X} = \mathbf{X}$ and second, a more general result (of which the first is a special case).

As a preliminary it is worth emphasizing, as in Lemma A4, that when $\mathbf{X} = \mathbf{VV}^-\mathbf{X}$ is true for *some* $\mathbf{V}^-$ it is true for all $\mathbf{V}^-$ (because $\mathbf{X} = \mathbf{VV}^-\mathbf{X} = \mathbf{VV}^\sim\mathbf{VV}^-\mathbf{X} = \mathbf{VV}^\sim\mathbf{X}$). Therefore when $\mathbf{X} = \mathbf{VV}^-\mathbf{X}$ we think of it being true not just for one $\mathbf{V}^-$ but for every $\mathbf{V}^-$.

Using the Lagrange multiplier approach when $\mathbf{V}$ is non-singular involves solving (24) and (25). The equations $\mathbf{Vt} = \mathbf{X\theta}$ of (24) when $\mathbf{VV}^-\mathbf{X} = \mathbf{X}$ become $\mathbf{Vt} = \mathbf{VV}^-\mathbf{X\theta}$ and are therefore consistent for a solution for $\mathbf{t}$, its form being $\mathbf{t} = \mathbf{V}^-\mathbf{X\theta}$. This leads to BLUE in exactly the same way as (30) is derived:

$$\mathbf{m}(\mathbf{V}^-) = \text{BLUE} = \mathbf{X}(\mathbf{X}'\mathbf{V}^-\mathbf{X})^-\mathbf{X}'\mathbf{V}^-\mathbf{y} = \text{when } \mathbf{VV}^-\mathbf{X} = \mathbf{X} \tag{36}$$

Its variance matrix is

$$\text{var}[\mathbf{m}(\mathbf{V}^-)] = \mathbf{X}(\mathbf{X}'\mathbf{V}^-\mathbf{X})^-\mathbf{X}'\mathbf{V}^-\mathbf{X}(\mathbf{X}'\mathbf{V}^-\mathbf{X})^-\mathbf{X} \tag{37}$$

Unbiasedness of BLUE in (38) and invariance to the choice of $(X'V^-X)^-$ is assured by Theorem 1 because the condition $X = CWX$ of that theorem is satisfied with $W = V^-$ and $C = V$ when $VV^-X = X$. Furthermore BLUE is invariant to the choice of V^- by Lemma A4, and so is equal to GLSE of (35). Thus

$$m(V^-) = m(V^+) = \text{BLUE} = \text{MLE} = \text{GLSE when } VV^-X = X \ . \tag{38}$$

The equality of BLUE, MLE and GLSE when $VV^-X = X$ is thus the same as in (31) for non-singular V.

3.3 BLUE – the general case

Derivation of BLUE $= m(V^-)$ in (36) is based on the same methodology as used for deriving BLUE in the case of non-singular V, namely, using Lagrange multipliers. But, in general, the technique of Lagrange multipliers does not always give the same results for minimization subject to side conditions as do other techniques. We therefore consider an alternative derivation.

Starting with $(I - M)y$ as the OLSE of $X\beta$ from (7) gives $\lambda'(I - M)y$ as the OLSE of $\lambda'X\beta$. [Equation (4) defines M.] We now ascertain what linear function of the OLSE residuals My should be added to $\lambda'(I - M)y$ to yield the BLUE of $\lambda'X\beta$ in the form

$$\text{BLUE}(\lambda'X\beta) = \lambda'(I - M)y + \tau'My \tag{39}$$

for some vector τ. Since $E[(I - M)y] = X\beta$ and $E(My) = 0$, no linear combination of $(I - M)y$ and My can be unbiased for $\lambda'X\beta$ unless the term in $(I - M)y$ is $\lambda'(I - M)y$, as in (39). We therefore obtain the BLUE by choosing τ to minimize the variance of (39), which is

$$v = \lambda'(I - M)V(I - M)\lambda + 2\lambda'(I - M)VM\tau + \tau'MVM\tau \ . \tag{40}$$

This leads to $\tau = -(MVM)^-MV(I - M)\lambda$ and so letting λ' be successive rows of I gives, from (39)

$$BLUE = (I - M)[I - VM(MVM)^-M]y. \tag{41}$$

There are many equivalent forms of this result. Pukelsheim (1974) develops (41) wholly in terms of $(MVM)^+$ rather than $(MVM)^-$ and so has, equivalent to an expression in Albert (1967)

$$BLUE = (I - M)[I - VM(MVM)^+M]y . \tag{42}$$

Pukelsheim then observes that for any matrix S

$$(MS)^+M = (MS)^+ \tag{43}$$

and so he has

$$BLUE = (I - M)[I - VM(MVM)^+]y . \tag{44}$$

Puntanen and Styan (1989) use $H = I - M = XX^+$ and so have (42) as

$$BLUE = Hy - HVM(MVM)^+y$$

$$= OLSE - HVM(MVM)^+y . \tag{45}$$

One can also observe that since V is non-negative definite, $V = LL'$, and so on applying (A4),

$$MVM(MVM)^-M = M . \tag{46}$$

Hence from (41)

$$BLUE = [I - VM(MVM)^-M]y . \tag{47}$$

Moreover, on substituting (46), used with $(MVM)^\sim$ different from $(MVM)^-$, into the last M of $VM(MVM)^-M$ occurring in (47) gives

$$VM(MVM)^-M = VM(MVM)^-MVM(MVM)^\sim M$$

$$= VM(MVM)^\sim M$$

and so BLUE is invariant to whatever value is used for $(MVM)^-$. Similar arguments yield

$$var(BLUE) = V - VM(MVM)^-MV$$

and, with $MX = 0$, using (47)

$$E(BLUE) = X\beta - VM(MVM)^-MX\beta = X\beta .$$

4. EQUALITY OF BLUE TO $m(V^-)$ AND TO OLSE

We have three theorems connecting BLUE for singular V to $m(V^-)$ and to OLSE.

Theorem 1. BLUE = $m(V^-)$ if and only if $VV^-X = X$.

The sufficiency part of this theorem shows that the general form of BLUE in Section 3.3 reduces to the BLUE in Section 3.2 when $VV^-X = X$. An early proof of sufficiency is due to Rao and Mitra (1971), and of necessity to Pukelsheim (1974). We offer new proofs that are shorter than theirs.

Proof.

First, define

$$Q = V^- - V^-X(X'V^-X)^-X'V^- = Q' \text{ with } QX = 0$$

and so observe that

$$QM = Q(I - XX^+) = Q = MQ, \text{ using the symmetry of } Q \text{ and } M,$$

$$= MQM \text{ because } MQM = M^2Q = MQ.$$

Hence

$$MVMQMVM = MVQVM = MVM - MVV^-X(X'V^-X)^-X'V^-VM. \tag{48}$$

Proof of sufficiency starts with $VV^-X = X$. Then in (48)

$$MVMQMVM = MVM - MX(X'V^-X)^-X'M = MVM;$$

i.e., Q is a generalized inverse of MVM. Hence

$$BLUE = y - VM(MVM)^-My$$

$$= y - VMQMy, \text{ from } (MVM)^- = Q$$

$$= y - VQy, \text{ because } Q = MQM$$

$$= y - [VV^-y - VV^-X(X'V^-X)^-X'V^-y]$$

$$= y - y + X(X'V^-X)^-X'V^-y, \text{ from Lemma A4,}$$

$$= X(X'V^-X)^-X'V^-y$$

$$= m(V^-) \text{ of } (36) \ .$$

Thus is sufficiency established.

For necessity, we show that $m(V^+)$ = BLUE implies $VV^+X = X$, whereupon $VV^-X = X$ and $m(V^+) = m(V^-)$ and so, in general, $m(V^-)$ = BLUE implies $VV^-X = X$. Therefore we start with

$$X(X'V^+X)^-X'V^+y = y - VM(MVM)^-My \ \forall \ y \ . \tag{49}$$

For convenience define T (different from the Ts of Sections 1.3 and 3.1) as

$$T = T' = M(MVM)^-M$$

and then, on applying Lemma A5 to (49), get

$$X(X'V^+X)^-X'V^+V = V - VTV \ . \tag{50}$$

Pre-multiplying (50) by $X'V^+$ gives

$$X'V^+X(X'V^+X)^-X'V^+V = X'V^+V - X'V^+VTV$$

which is

$$X'V^+V = X'V^+V - X'V^+VTV \ .$$

Therefore

$$X'V^+VTV = 0 \ \text{and so} \ VTVV^+X = 0 \ . \tag{51}$$

Post-multiplying (50) by V^+X,

$$X(X'V^+X)^-X'V^+VV^+X = VV^+X - VTVV^+X \ ,$$

and using (51) gives

$$X(X'V^+X)^-X'V^+X = VV^+X \ . \tag{52}$$

But a standard result based on (A4) is that the left-hand side of (52) is X; and so $X = VV^+X$, and hence, as in Lemma A4, $X = VV^-X$.

Q.E.D.

Corollary

$$\text{BLUE} = m(V_p^-) \quad \text{for} \quad V_p = V + XX' \ .$$

This is the form of BLUE given by Rao and Mitra (1971).

Proof. One only needs to show that $V_p V_p^- X = X$. Since

$$(I - V_p V_p^-)V_p(I - V_p V_p^-)' = 0 \ ,$$

$$(I - V_p V_p^-)V(I - V_p V_p^-)' + (I - V_p V_p^-)XX'(I - V_p V_p^-)' = 0 . \tag{53}$$

Since each of the two terms in (53) is n.n.d., their sum is null only if each term is null. Therefore $(I - V_p V_p^-)XX'(I - V_p V_p^-)' = 0$ and so, because all matrices are real in this context, $(I - V_p V_p^-)X = 0$, i.e., $V_p V_p^- X = X$.

<div align="right">Q.E.D.</div>

Theorem 2. BLUE = OLSE = $XX^+ y$ if and only if $VX = XQ$ for some Q.

This result is due to Zyskind (1967) as part of a series of equivalent results of which this is but one form.

Proof. If $VX = XQ$

$$MVM = MV(I - XX^+) = MV - MVXX^+ = MV - MXQX^+$$

$$= MV \text{ because } MX = 0$$

$$= VM \text{ because } MVM \text{ is symmetric, and hence so is } MV .$$

Therefore $HVM = (I - M)VM = 0$ and so in (45) BLUE = OLSE. Thus is sufficiency established.

Proving necessity begins with BLUE = OLSE:

$$y - VM(MVM)^- My = y - My \ \forall \ y .$$

Therefore $[VM(MVM)^- M - M]y = 0 \ \forall \ y$. Hence, by Lemma A5, $[VM(MVM)^- M - M]V = 0$. Therefore $VM(MVM)^- MV = MV = VM$, the last equality arising from the symmetry of $VM(MVM)^- MV$. Hence, with $MV = VM$ and M being $I - XX^+$,

$$(I - XX^+)V = V(I - XX^+) ,$$

which yields $XX^+ VX = VXX^+ X = VX$; i.e., $VX = XX^+ VX = XQ$ for $Q = X^+ VX$.

<div align="right">Q.E.D.</div>

Theorem 3. $m(V^+)$ = OLSE if and only if $m(V^+)$ = BLUE and OLSE = BLUE; in which case $VV^-X = X$ and $VX = XQ$ for some Q.

Proof . Sufficiency is obvious: if $m(V^+)$ and OLSE each equal BLUE then they equal each other. Proving necessity starts with $m(V^+)$ = OLSE, i.e.,

$$X(X'V^+X)^-X'V^+y = X(X'X)^-X'y \ . \tag{54}$$

Applying Lemma A5 to (54) gives

$$X(X'V^+X)^-X'V^+V = X(X'X)^-X'V$$

and pre-multiplying this by X' and post-multiplying it by V^+X gives

$$X'X(X'V^+X)^-X'V^+VV^+X = X'X(X'X)^-X'VV^+X \ . \tag{55}$$

Then because $X(X'V^+X)^-X'V^+X = X$, (55) reduces to $X'X = X'VV^+X$. Therefore

$$X'X - X'VV^+X = X'(I - VV^+)X = [(I - VV^+)X]'[(I - VV^+)X] = 0 \ ,$$

and so $(I - XX^+)V = 0$, thus giving $X = VV^+X$ and hence $X = VV^-X$. Thus $m(V^+)$ = OLSE implies $X = VV^-X$ which, as we have seen, implies $m(V^+)$ = $m(V^-)$ = BLUE.

<div align="right">Q.E.D.</div>

5. m(V⁻) when VV⁻X ≠ X

It was shown in Section 3.2 that $VV^-X = X$ is sufficient for equations $Vt = X\theta$ to be consistent for a solution for t, whatever the choice of θ. It is also a necessary condition. This is so because, since V is n.n.d. and so can be factored as $V = L'L$ as used in (32), the consistency of $Vt = X\theta$ means that $X = L'Z$ for some Z. Then, because $V^- = (L'L)^- = L^-(LL')^{-1}L$ we have

$$VV^-X = L'LL^-(LL')^{-1}LL'Z = L'LL^-Z = L'Z = X ,$$

after using $LL^- = I$ that comes from L having full row rank (see Lemma A3). Thus, $VV^-X = X$ is a necessary and sufficient condition for equations $Vt = X\theta$ to be consistent for a solution for t, for any θ; and $m(V^-) = $ BLUE follows. But there can be cases where there are some forms of V^- (without VV^-X equalling X) that lead to $m(V^-) = X(X'V^-X)^-X'V^-y$ being the same as BLUE $= [I - VM(MVM)^-M]y$ of (47). This means for *some* θ (or θs), the equations $Vt = X\theta$ have a solution for t without being consistent for *every* θ; i.e., there can be a solution $t = V^-X\theta$ for that θ (or θs) without VV^-X being X. We illustrate this possibility in an example, and then investigate what form V^- must have.

5.1 An example of VV⁻X ≠ X

Consider as an example (suggested by Baksalary, 1986) the case of

$$V = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad X = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \tag{56}$$

where rank of $\bf{V}^-$ is r = 2. We use this to illustrate the possibility of being able, in some cases, to solve the Lagrange multiplier equations, (24) and (25), in the form $\bf{t} = \bf{V}^-\bf{X}\theta$ without having $\bf{VV}^-\bf{X}$. Then $\bf{m}(\bf{V}^-)$ can be calculated, and in this example it is BLUE, even though $\bf{VV}^-\bf{X}$ is not $\bf{X}$.

Using (56), the equation $\bf{Vt} = \bf{X}\theta$ is

$$t_1 = \theta \qquad t_2 = \theta \qquad \text{and} \qquad 0 = \theta \ , \tag{57}$$

i.e., $t_1 = t_2 = 0$; and t_3 remains unspecified. On taking $\bf{X}'\bf{t} = \bf{X}'\lambda$ of (25) into account, it being $\Sigma t = \Sigma\lambda$ for this example, i.e., $t_1 + t_2 + t_3 = \lambda_1 + \lambda_2 + \lambda_3$, we have in conjunction with $t_1 = t_2 = \theta = 0$ of (57), the solution to $\bf{Vt} = \bf{X}\theta$ and $\bf{X}'\bf{t} = \bf{X}'\lambda$ as

$$\bf{t}' = [0 \quad 0 \quad \Sigma\lambda_i] \quad \text{and} \quad \theta = 0 \ . \tag{58}$$

Thus $\bf{T}'\bf{y} = (\Sigma\lambda_i)y_3$ is the BLUE of $\lambda'\bf{X}\beta = (\Sigma\lambda_i)\beta$, i.e., $\hat{\beta} = y_3$. Clearly, this is correct, since y_3 has variance zero as seen from var($\bf{y}$) = $\bf{V}$ of (56).

This solution has been derived without using any $\bf{V}^-$. It is therefore not $\bf{V}^-\bf{X}\theta$ analogous to $\bf{V}^{-1}\bf{X}\theta$ of (26) in the non-singular $\bf{V}$ case. Nevertheless, it *is* possible in this example (and in others) to find certain forms of generalized inverse of $\bf{V}$, to be denoted $\bf{\tilde{V}}$ such that $\bf{t}'\bf{y}$ is $\lambda'\bf{m}(\bf{\tilde{V}})$, analogous to (29) but with $\bf{VV}^{\tilde{}}\bf{X} \neq \bf{X}$. One possibility in the example is

$$\bf{\tilde{V}} = \begin{bmatrix} 1 & 0 & a \\ 0 & 1 & b \\ -1 & -1 & c \end{bmatrix} \tag{59}$$

for any a, b and c so long as a + b + c $\neq$ 0. For then,

$$\lambda'\bf{X}(\bf{X}'\bf{\tilde{V}}\bf{X})^-\bf{X}'\bf{\tilde{V}} = [\lambda_1 \ \lambda_2 \ \lambda_3] \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} (a + b + c)^{-1} [0 \quad 0 \quad a + b + c]$$

$$= [0 \quad 0 \quad \Sigma\lambda] = \bf{t}' \text{ of (58)} \ .$$

Note, too, that $VV^{\sim}X \neq X$:

$$VV^{\sim}X = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & a \\ 0 & 1 & b \\ -1 & -1 & c \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1+a \\ 1+b \\ 0 \end{bmatrix} \neq \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = X \ ,$$

and hence $VV^{-}X \neq X$ for every V^{-}. Thus, although for the particular $V^{\sim}$ of (59), we have $t'y$ equalling $m(V^{\sim}) = \lambda'X(X'V^{\sim}X)^{-}X'V^{\sim}y$, the latter is not invariant to any choice of generalized inverse of V used in place of $V^{\sim}$. Nevertheless, since in the particular case of $V^{\sim}$ it is BLUE, we give the name pseudo-BLUE (P-BLUE) to $m(V^{-})$ when $VV^{-}X \neq X$, and proceed to find a general form for P-BLUE and then provide an answer to the question "When is P-BLUE equal to BLUE?". This demands finding what form V^{-} must have, without $VV^{-}X = X$, such that $m(V^{-})$ is BLUE.

5.2 Pseudo-BLUE: $m(V^{-})$ when $VV^{-}X \neq X$

Whenever it is possible to solve $Vt = X\theta$ and $X't = X'\lambda$ for t and some particular θ or θs (as in the preceding example), and when a V^{-} can be derived without $VV^{-}X$ equalling X, we call the resulting $m(V^{-})$ a pseudo-BLUE (P-BLUE). We here derive a general form for P-BLUE.

Since V is singular, there is no loss of generality (save for the use of permutation matrices) in partitioning V as

$$V = \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} \quad \text{and similarly} \quad X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \ , \tag{60}$$

where V_1 has full row rank, the same rank as V. Hence

$$V_2 = KV_1 \text{ for some } K \ . \tag{61}$$

Then equations $Vt = X\theta$ are, by (60), equivalent to the two equations $V_1 t = X_1\theta$ and $X_2 t = X_2\theta$. Multiplying the first of these by K and subtracting the second yields, by (61),

$$(KV_1 - V_2)t = 0 = (KX_1 - X_2)\theta \ . \tag{62}$$

But the absence of consistency means that $KX_1 \neq X_2$ and so (62) has a solution $\theta = 0$.

In many cases, a solution can be found for t. Part of $Vt = V\theta$ is $V_1 t = X_1 \theta$ which, with $\theta = 0$, is $V_1 t = 0$ and this, since V_1 has full row rank, has solution $t = (I - V_1^- V_1)w$ for any w. Therefore, to satisfy $X't = X'\lambda$ of (25), we must choose w to satisfy

$$X'(I - V_1^- V_1)w = X'\lambda \ . \tag{63}$$

Solving (63) for w and substituting into t gives

$$t = (I - V_1^- V_1)[X'(I - V_1^- V_1)]^- X'\lambda \ . \tag{64}$$

Now, for V_{11} non-singular, of rank the same as V, consider partitioning V as

$$V = \begin{bmatrix} V_1 \\ V_2 \end{bmatrix} = \begin{bmatrix} V_1 \\ KV_1 \end{bmatrix} = \begin{bmatrix} V_{11} & V_{11}K' \\ KV_{11} & KV_{11}K' \end{bmatrix} \tag{65}$$

with

$$V^- = \begin{bmatrix} V_{11}^{-1} & 0 \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad V_1^- = \begin{bmatrix} V_{11}^{-1} \\ 0 \end{bmatrix} \ .$$

Hence

$$V_1^- V_1 = \begin{bmatrix} V_{11}^{-1} \\ 0 \end{bmatrix} [V_{11} \quad V_{11}K'] = \begin{bmatrix} I & K' \\ 0 & 0 \end{bmatrix} \ . \tag{66}$$

As a result of (66), it is readily shown that (64) reduces to

$$t = B'(X'B')^- X'\lambda \quad \text{for} \quad B = [-K \quad I], \quad \text{with} \quad K = V_2 V_1'(V_1 V_1')^{-1} \ , \tag{67}$$

the latter coming from (61). Thus (67) and $\theta = 0$ are a solution to (24) and (25) when $VT = X\theta$ is not consistent for a solution in t *for all* θ, and so $VV^{-1}X \neq X$, as is easily confirmed. Then (67) gives what we call P-BLUE, namely

$$\text{P-BLUE} = X(BX)^- By \quad \text{for} \quad B = [-K \quad I] \ .$$

Example.

Applying (60) and (61) to (56) gives

$$\mathbf{K} = [0 \quad 0], \quad \mathbf{B} = [0 \quad 0 \quad 1] \quad \text{and} \quad \mathbf{X'B'} = 1 .$$

Therefore (67) is

$$\mathbf{t} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} (1)^{-} [1 \quad 1 \quad 1] \begin{bmatrix} \lambda_1 \\ \lambda_1 \\ \lambda_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \Sigma\lambda \end{bmatrix} ,$$

the same result as in (58).

Verification of solution

From (65),

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_{11} \\ \mathbf{KV}_{11} \end{bmatrix} [\mathbf{I} \quad \mathbf{K'}] . \tag{69}$$

Therefore from (67) and (69), **Vt** involves the product

$$[\mathbf{I} \quad \mathbf{K'}]\mathbf{B'} = [\mathbf{I} \quad \mathbf{K'}] \begin{bmatrix} -\mathbf{K'} \\ \mathbf{I} \end{bmatrix} = 0$$

and so $\mathbf{Vt} = 0 = \mathbf{X\theta}$ for $\mathbf{\theta} = 0$ of the solution. Thus (24) is satisfied. For (25) it will be found in using (66) to derive (67) from (64) that

$$\mathbf{X'}(\mathbf{I} - \mathbf{V}_1^{-}\mathbf{V}_1) = [0 \quad \mathbf{X'B'}] .$$

Hence equations (63) are

$$[0 \quad \mathbf{X'B'}]\mathbf{w} = \mathbf{X'\lambda} ,$$

i.e., for $\mathbf{w}_2$ the appropriate sub-vector of **w**,

$$\mathbf{X'B'w}_2 = \mathbf{X'\lambda} .$$

Clearly, these are consistent. Therefore a solution is $\mathbf{w}_2 = (\mathbf{X'B'})^{-}\mathbf{X'\lambda}$, and so $\mathbf{X'B'}(\mathbf{X'B'})^{-}\mathbf{X'\lambda} = \mathbf{X'\lambda}$. But from (67) $\mathbf{X't} = \mathbf{X'B'}(\mathbf{X'B'})^{-}\mathbf{X'\lambda}$ and so $\mathbf{X't} = \mathbf{X'\lambda}$, which is (25). Thus $\mathbf{\theta} = 0$ and **t** of (67) satisfy (24) and (25).

5.3 When does P-BLUE have the form $m(\tilde{V})$?

Section 5.1 illustrates the existence of a $\tilde{V}$ such that P-BLUE of (68) equals $m(\tilde{V})$, but with $X\tilde{X}X \neq X$; and the invariance properties of BLUE do not apply. Nevertheless it is of interest to investigate when such a $\tilde{V}$ exists and how it might be derived.

Begin with the equality

$$X(BX)^-B = X(X'\tilde{V}X)^-X'\tilde{V} \ . \tag{70}$$

Post-multiplication by X yields

$$X(BX)^-BX = X \ .$$

But this holds if, and only if, BX and X have the same rank. Therefore $r_{BX} = r_X$ is a necessary condition for (70) to be true.

Assuming that rank condition, pre-multiplying (70) by $X'\tilde{V}$ gives

$$X'\tilde{V}X(BX)^-B = X'\tilde{V} \ . \tag{71}$$

Now, with V^- that follows (65),

$$V^-V = \begin{bmatrix} I & K' \\ 0 & 0 \end{bmatrix} \quad \text{and} \quad VV^- = \begin{bmatrix} I & 0 \\ K & 0 \end{bmatrix} \ . \tag{72}$$

Therefore, since from (A7) the general form of generalized inverse of V, V^g say, is

$$V^g = V^-VV^- + (I - V^-V)P + Q(I - VV^-) \tag{73}$$

for any square P and Q of order N, using (72) in (73), together with B of (67) gives

$$V^g = V^- + [0 \quad B']P + Q\begin{bmatrix} 0 \\ B \end{bmatrix} \ . \tag{74}$$

Now partition P as $P = [U' \quad F']'$ and Q as $Q = [R \quad S]$ with F' and S each of order $N \times (N-r)$. Then (74) is

$$V^g = V^- + B'F + SB \ . \tag{75}$$

Using this for $\tilde{V}$ in (71) gives

$$X'(V^- + B'F + SB)X(BX)^-B = X'(V^- + B'F + SB) . \qquad (76)$$

Therefore when any F and S that satisfy (76) are used in V^g of (75), that V^g used as $\tilde{V}$ in $\lambda'X(X'\tilde{V}X)^-X'\tilde{V}$ yields $t' = \lambda'X(BX)^-B$ of (67). Then

$$P\text{-BLUE} = X(BX)^-By = X(X'V^gX)^-X'V^gy \qquad (77)$$

for V^g of (75). But (77) is <u>not</u> BLUE because $VV^gX \neq X$ for V^g of (75).

Example (continued)

$$B = [0 \quad 0 \quad 1], \quad X' = [1 \quad 1 \quad 1] \text{ and } BX = 1 .$$

Hence $r(BX) = r(X) = 1$. Since $N = 3$ and $r_V = 2$, F' and S have the form $F = [f_1 \quad f_2 \quad f_3]$ and $S' = [s_1 \quad s_2 \quad s_3]$. Using these in (76) gives

$$[1 \ 1 \ 1] \left\{ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} [f_1 \ f_2 \ f_3] + \begin{bmatrix} s_1 \\ s_2 \\ s_3 \end{bmatrix} [0 \ 0 \ 1] \right\} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} (1)^- [0 \ 0 \ 1]$$

$$= [1 \ 1 \ 1] \left\{ \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} [f_1 \ f_2 \ f_3] + \begin{bmatrix} s_1 \\ s_2 \\ s_3 \end{bmatrix} [0 \ 0 \ 1] \right\} .$$

This is

$$(2 + \Sigma f_i + \Sigma s_i)[0 \ 0 \ 1] = [1 \ 1 \ 0] + [f_1 \quad f_2 \quad f_3] + [0 \ 0 \ \Sigma s_i] .$$

which is

$$[0 \ 0 \ (2 + \Sigma f_i + \Sigma s_i)] = [1 + f_1 \quad 1 + f_2 \quad f_3 + \Sigma s_i] .$$

Therefore a solution is $f_1 = -1$, $f_2 = -1$ and $f_3 = 2$, with s_1, s_2, s_3 being anything. Thus in (73)

$$V^g = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} [-1 \quad -1 \quad 2] + \begin{bmatrix} s_1 \\ s_2 \\ s_3 \end{bmatrix} [0 \ 0 \ 1] = \begin{bmatrix} 1 & 0 & s_1 \\ 0 & 1 & s_2 \\ -1 & -1 & 2+s_3 \end{bmatrix} ,$$

and for $s_1 = a$, $s_2 = b$ and $s_3 = c - 2$ this is V^- of (59).

Special Case 1: when BX is scalar.

It is easily shown that **BX** has order $(N - r) \times p$. Therefore when, as in the preceding example, **BX** is scalar, $N - r = 1 = p$. Moreover, **F** and **S** then have order $1 \times N$ and $N \times 1$, respectively. Therefore, on writing **B**, **F**, **S** and **X** as **b'**, **f'**, **s** and **x**, respectively, (76) becomes

$$\frac{x'V^-xb'}{b'x} + \frac{x'bf'xb'}{b'x} + \frac{x'sb'xb}{b'x} = x'V^- + x'bf' + x'sb' \ .$$

This simplifies to

$$\frac{x'V^-xb'}{b'x} + f'xb' = x'V^- + x'bf' \qquad (78)$$

which contains no **s**. Then, when any **f'** satisfying (78) is used in (75) with **s** $= 0$ (when $N - r = 1 = p$), i.e., in

$$V^g = V^- + bf' \ ,$$

then that V^g used in (77) yields $t' = \lambda'xb'/b'x$, but $VV^g \neq X$. This is exactly what happened in the example

$$t' = [\lambda_1 \ \ \lambda_2 \ \ \lambda_3] \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} [0 \ \ 0 \ \ 1]/1 = [0 \ \ 0 \ \ \Sigma\lambda] \ .$$

Special case 2: when $BX(BX)^-B = B$, i.e., when **BX** and **B** have the same rank. Then (76) becomes

$$X'(V^- + B'F)X(BX)^-B = X'(V^- + B'F) \qquad (79)$$

and if any **F** that satisfies (78) is used in (75) with **S** $= 0$, i.e., in

$$V^g = V^- + B'F \ , \qquad (80)$$

then that V^g used in (77) yields $t' = \lambda'X(BX)^-B$ of (67), but $VV^gX \neq X$.

6. APPENDIX

6.1 Generalized inverses

The Moore-Penrose inverse of any non-null matrix A is the unique matrix A^+ satisfying

$$AA^+A = A, \quad A^+AA^+ = A \text{ and having both } AA^+ \text{ and } A^+A \text{ symmetric} . \qquad (A1)$$

A generalized inverse, A^-, satisfies just the first condition in (A1):

$$AA^-A = A . \qquad (A2)$$

The following properties (e.g., Searle, 1982, p. 216) of a generalized inverse $(X'X)^-$ of $X'X$ are especially useful:

$$(X'X)^{-\prime} \text{ is a generalized inverse of } X'X , \qquad (A3)$$

$$X(X'X)^-X'X = X , \qquad (A4)$$

and

$$X^+ = X'(XX')^-X(X'X)^-X' , \qquad (A5)$$

from which is is easily seen that

$$XX^+ = X(X'X)^-X' , \qquad (A6)$$

with XX^+ being symmetric and invariant to the choice of $(X'X)^-$.

Lemma A1. For A^- being a generalized inverse of A, so is

$$A^{\sim} = A^-AA^- + (I - A^-A)P + Q(I - AA^-) \qquad (A7)$$

for any P and Q. Furthermore, for $A^{\sim}$ being some particular generalized inverse of A, say A_g, a suitable P and Q can always be found for (A7); e.g., $P = A_g$ and $Q = A^-AA_g$ (as is easily verified by substitution).

6.2 Non-negative definite (n.n.d.) matrices

If W is any n.n.d. (real) matrix of rank r then there exists (see Searle, 1982, Section 7.7) a matrix H of full row rank r such that

$$W = K'K, \quad (K'K)^{-1} \text{ exists and } W^+ = K'(KK')^{-2}K \qquad (A8)$$

Verification of W^+ is easily established from (A1).

If, for any real matrix B,

$$BB' = 0 \text{ then } B = 0. \qquad (A9)$$

6.3 Other pertinent results

Lemma A2. If $FSG = 0$ for all S then either $F = 0$ or $G = 0$.

Proof. If $G \neq 0$, there exists a vector $\tau = Gu \neq 0$. Then

$$FSG = 0 \Rightarrow FSGu = 0 \Rightarrow FS\tau = 0 \ \forall \ S \ .$$

For any non-null vector v take $S = v\tau'/\tau'\tau$. Then

$FS\tau = 0 \Rightarrow Fv\tau'\tau/\tau'\tau = 0 \Rightarrow Fv = 0 \ \forall \ \tau \neq 0$, and so $F = 0$.

Similar arguments applied to $G'S'F' = 0$ when $F \neq 0$ yield $G = 0$.

<div align="right">Q.E.D.</div>

Lemma 3. For R of full row rank, $RR^- = I$ for all R^-.

Proof. $R^\sim = R'(RR')^{-1}$ is a generalized inverse of R with $RR^\sim = I$. Therefore $I = RR^\sim = RR^- RR^\sim = RR^- I = RR^-$.

<div align="right">Q.E.D.</div>

Lemma A4. If $VV^- X = X$ then

    (i) $VV^- X = X$ for every V^-,

    (ii) $X'V^- X$ is invariant to V^-; and for almost all $y \sim (X\beta, V)$,

    (iii) $VV^- y = y$ for every V^-, and

    (iv) $X'V^- y$ is invariant to V^-.

Proof. Let $V^\sim$ be a generalized inverse of V different from V^-.

    (i) $X = VV^- X = VV^\sim VV^- X = VV^\sim X$.

    (ii) Write $X' = (VV^\sim X)' = X'V^\sim V$ because V is symetric and if $V^\sim$ is not (although it can always be made so) then $V^{\sim'}$ is a generalized inverse of V anyway. Then

$$X'V^- X = X'V^\sim VV^- X = X'V^\sim X \ .$$

(iii) $0 = (I - VV^-)V(I - VV^-)'$

$\qquad = (I - VV^-)E[(y - X\beta)(y - X\beta)'](I - VV^-)'$

$\qquad = E(zz')$ for $z = (I - VV^-)(y - X\beta)$.

Therefore, since $E(zz') = 0$ implies, with probability 1.0, that $z = 0$, we have $(I - VV^-)(y - X\beta) = 0$. But $(I - VV^-)X = 0$. Therefore $(I - VV^-)y = 0$ and so $y = VV^-y$ and then $y = VV^-y = VV^-\tilde{V}Vy = VV^-\tilde{y}$.

(iv) Using (iii) of proof is the same as that of (ii), but with its final X replaced by y.

<div align="right">Q.E.D.</div>

Lemma A5. If $Ky = 0$ with probability unity, then $KV = 0$.

Proof. $Ky = 0$ for almost all y implies $\text{var}(Ky) = 0 \Rightarrow KVK' = 0$. $V = LL'$ as used in (32), and so $KVK' = 0 \Rightarrow KL'LK' = 0 \Rightarrow KL' = 0 \Rightarrow KL'L = 0 \Rightarrow KV = 0$.

<div align="right">Q.E.D.</div>

REFERENCES

Aitken, A. C. (1935). On least squares and linear combinations of observations. *Proc. Roy. Soc. Edinburgh*, 55, 42-48.

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

Baksalary, J. K. (1986). Personal communication.

Baksalary, J. K. and Kala, R. (1978). Relationships between some representations of the best linear unbiased estimator in the general Gauss-Markov model. *SIAM J. Appl. Math.*, 35, 515-520.

Baksalary, J. K. and Kala R. (1983). On equalities between BLUEs, WLSEs and SLSEs. *Canad. J. Statist.*, 11, 119-123.

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

Pukelsheim, F. (1974). Schätzen von Mittelwert und Streuungsmatrix in Gauss - Markoff - Modellen. *Diplomarbeit*, 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. *Amer. Statist.* 43 (in press).

Rao, C. R. and Mitra, S. K. (1971). Further contributions to the theory of generalized inverses of matrices and its applications. *Sankhya, Ser. A*, 33,289-30.

Searle, S. R. (1982). *Matrix Algebra Useful for Statistics*. Wiley and Sons, New York, N.Y.

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