EXTENDING SOME RESULTS AND PROOFS FOR THE SINGULAR LINEAR MODEL

S.R. Searle

Biometrics Unit
N.Y. State College of Agriculture and Life Sciences
Cornell University, Ithaca, NY 14853

BU-1167-MA November 1993

ABSTRACT

Much has been written about the linear model with singular dispersion matrix. Puntanen and Styan (1989) give an excellent review and reference list. This note gives some new forms and shortened proofs for well-established results.

1. DERIVATION AND ALTERNATIVE FORMS OF BLUE($\mathbf{X}\boldsymbol{\beta}$)

a. Ordinary least squares estimation

$\mathrm{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ is always estimable in the linear model $\mathbf{y} \sim (\mathbf{X}\boldsymbol{\beta}, \mathbf{V})$, whether $\mathbf{V} = \mathrm{var}(\mathbf{y})$ is singular or not. We therefore confine attention to estimating $\mathbf{X}\boldsymbol{\beta}$, and start with its ordinary least squares estimator, to be denoted OLSE($\mathbf{X}\boldsymbol{\beta}$). This, as is well-known, is

$$\mathrm{OLSE}(\mathbf{X}\boldsymbol{\beta}) = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y} = \mathbf{X}\mathbf{X}^{+}\mathbf{y} = (\mathbf{I} - \mathbf{M})\mathbf{y} \tag{1}$$

for

$$\mathbf{M} = \mathbf{I} - \mathbf{X}\mathbf{X}^{+} = \mathbf{M}^2 = \mathbf{M}' \text{ with } \mathbf{M}\mathbf{X} = \mathbf{0}\ . \tag{2}$$

$\mathbf{X}^{-}$ represents any generalized inverse satisfying $\mathbf{X}\mathbf{X}^{-}\mathbf{X} = \mathbf{X}$ and $\mathbf{X}^{+}$ is the Moore-Penrose inverse of $\mathbf{X}$. Then, from (1), the OLSE of $\boldsymbol{\lambda}'\mathbf{X}\boldsymbol{\beta}$ for any vector $\boldsymbol{\lambda}'$ is $\mathrm{OLSE}(\boldsymbol{\lambda}'\mathbf{X}\boldsymbol{\beta}) = \boldsymbol{\lambda}'(\mathbf{I} - \mathbf{M})\mathbf{y}$.

b. Best linear unbiased estimation

There are many equivalent forms of the best, linear, unbiased estimator of $\mathbf{X}\boldsymbol{\beta}$, to be denoted BLUE($\mathbf{X}\boldsymbol{\beta}$). Pukelsheim (1974) develops it in terms of the unique Moore-Penrose inverse $(\mathbf{MVM})^{+}$ and so has, equivalent to an expression in Albert (1967),

$$\mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta}) = (\mathbf{I} - \mathbf{M})[\mathbf{I} - \mathbf{VM}(\mathbf{MVM})^{+}\mathbf{M}]\mathbf{y}\ . \tag{3a}$$

Pukelsheim then notes that for any S with MS existing, $(\mathbf{MS})^{+}\mathbf{M} = (\mathbf{MS})^{+}$ and so he has

$$\mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta}) = (\mathbf{I}-\mathbf{M})[\mathbf{I}-\mathbf{VM}(\mathbf{MVM})^{+}]\mathbf{y}\,. \tag{3b}$$

He could just as well have noted that $\mathbf{M}(\mathbf{MVM})^{+}\mathbf{M} = (\mathbf{MVM})^{+}$, and written

$$\mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta}) = (\mathbf{I}-\mathbf{M})[\mathbf{I}-\mathbf{V}(\mathbf{MVM})^{+}]\mathbf{y}\,.$$

Puntanen and Styan (1989) use $\mathbf{H} = \mathbf{I}-\mathbf{M} = \mathbf{XX}^{+}$ and so have

$$\mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta}) = \mathbf{Hy}-\mathbf{HVM}(\mathbf{MVM})^{+}\mathbf{y} = \mathrm{OLSE}(\mathbf{X}\boldsymbol{\beta})-\mathbf{HVM}(\mathbf{MVM})^{+}\mathbf{y}\,. \tag{4}$$

Developing the best linear unbiased estimator (BLUE) of $\boldsymbol{\lambda}'\mathbf{X}\boldsymbol{\beta}$ starts with $\mathbf{y} = (\mathbf{I}-\mathbf{M})\mathbf{y} + \mathbf{My}$. Because $\mathrm{E}[(\mathbf{I}-\mathbf{M})\mathbf{y}] = \mathbf{X}\boldsymbol{\beta}$ and $\mathrm{E}(\mathbf{My}) = \mathbf{0}$, any linear combination of $(\mathbf{I}-\mathbf{M})\mathbf{y}$ and $\mathbf{My}$ can be unbiased for $\boldsymbol{\lambda}'\mathbf{X}\boldsymbol{\beta}$ only if the term in $(\mathbf{I}-\mathbf{M})\mathbf{y}$ is $\boldsymbol{\lambda}'(\mathbf{I}-\mathbf{M})\mathbf{y}$. We therefore ask "for what vector $\boldsymbol{\tau}'$ does adding $\boldsymbol{\tau}'\mathbf{My}$ to $\boldsymbol{\lambda}'(\mathbf{I}-\mathbf{M})\mathbf{y}$ yield the BLUE of $\boldsymbol{\lambda}'\mathbf{X}\boldsymbol{\beta}$?" To answer this we seek $\boldsymbol{\tau}'$ to minimize the variance

$$\mathrm{var}[\boldsymbol{\lambda}'(\mathbf{I}-\mathbf{M})\mathbf{y} + \boldsymbol{\tau}'\mathbf{My}] = \boldsymbol{\lambda}'(\mathbf{I}-\mathbf{M})\mathbf{V}(\mathbf{I}-\mathbf{M})\boldsymbol{\lambda} + 2\boldsymbol{\lambda}'(\mathbf{I}-\mathbf{M})\mathbf{VM}\boldsymbol{\tau} + \boldsymbol{\tau}'\mathbf{MVM}\boldsymbol{\tau}\,.$$

This minimization leads, after some straightforward vector calculus (e.g., Searle, 1982, Section 12.8b and c), to $\boldsymbol{\tau} = -(\mathbf{MVM})^{-}\mathbf{MV}(\mathbf{I}-\mathbf{M})\boldsymbol{\lambda}$. Hence

$$\mathrm{BLUE}(\boldsymbol{\lambda}'\mathbf{X}\boldsymbol{\beta}) = \boldsymbol{\lambda}'(\mathbf{I}-\mathbf{M})\mathbf{y} + \boldsymbol{\tau}'\mathbf{My}$$

becomes, on letting $\boldsymbol{\lambda}'$ be successive rows of $\mathbf{I}$,

$$\mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta}) = (\mathbf{I}-\mathbf{M})[\mathbf{I}-\mathbf{VM}(\mathbf{MVM})^{-}\mathbf{M}]\mathbf{y}\,. \tag{5}$$

This is identical to (3) save for (3) having the unique Moore-Penrose inverse of MVM, whereas (5) has any generalized inverse. Thus (3) is unique, whereas (5) appears not to be: but we proceed to show that it is, and is thus equal to (3).

c. Invariance to $(\mathbf{MVM})^{-}$

Since $\mathbf{V}$ is a variance-covariance matrix, it is symmetric and non-negative definite. It can therefore be written as $\mathbf{V} = \mathbf{LL}'$ with $\mathbf{L}$ a real, full column rank matrix of rank t, the rank of $\mathbf{V}$; and so $\mathbf{L}'\mathbf{L}$ is non-singular. Then with $\mathbf{y} \sim (\mathbf{X}\boldsymbol{\beta}, \mathbf{V})$ it is always possible to write

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Lw} \tag{6}$$

where $\mathrm{var}(\mathbf{w}) = \mathbf{I}$ and

$$\mathbf{V} = \mathrm{var}(\mathbf{y}) = \mathbf{LIL}' = \mathbf{LL}'\,.$$

Now consider the identity

$$\mathbf{MVM(MVM)^{-}MVM = MVM}\ .$$

This, using $\mathbf{V = LL'}$ and the symmetry of $\mathbf{M}$, is

$$\mathbf{MVM(MVM)^{-}MLL'M' = MLL'M'}\ .$$

Then, on using the result (e.g. Searle, 1987, p. 63) for any real matrices $\mathbf{P}$, $\mathbf{Q}$ and $\mathbf{T}$ that $\mathbf{PTT' = QTT'}$ implies $\mathbf{PT = QT}$, we get

$$\mathbf{MVM(MVM)^{-}ML = ML}\ .$$

Therefore for $\mathbf{w}$ of (6)

$$\mathbf{MVM(MVM)^{-}MLw = MLw}\ . \tag{7}$$

But since $\mathbf{MX = 0}$, pre-multiplying of (6) by $\mathbf{M}$ gives

$$\mathbf{My = MLw}$$

and so in (7)

$$\mathbf{MVM(MVM)^{-}My = My}\ . \tag{8}$$

Suppose $\mathbf{(MVM)^{\sim}}$ is a generalized inverse of $\mathbf{MVM}$ different from $\mathbf{(MVM)^{-}}$. Replacing $\mathbf{My}$ of (5) by (8) with $\mathbf{(MVM)^{-}}$ replaced by $\mathbf{(MVM)^{\sim}}$ then gives the term $\mathbf{VM(MVM)^{-}My}$ of (5) as

$$\begin{aligned}\mathbf{VM(MVM)^{-}My} &= \mathbf{VM(MVM)^{-}MVM(MVM)^{\sim}My}\\ &= \mathbf{VM(MVM)^{\sim}My}\ ,\end{aligned}$$

because $\mathbf{VM(MVM)^{-}MVM = LL'M'(L'M')^{+}L'M = VM}$. Thus BLUE $(\mathbf{X}\beta)$ of (5) is invariant to the choice of $\mathbf{(MVM)^{-}}$, and in particular $\mathbf{(MVM)^{+}}$ could be used, in which case (5) would be identical to (3).

d. A simplification

Multiplying out the right-hand side of (5) gives

$$\mathrm{BLUE}(\mathbf{X}\beta) = \mathbf{y - My - VM(MVM)^{-}My + MVM(MVM)^{-}My}$$

and using (8) reduces this to the new and simpler

$$\mathrm{BLUE}(\mathbf{X}\beta) = \mathbf{y - VM(MVM)^{-}My}\ . \tag{9}$$

We can note in (9) that $\mathbf{M(MVM)^{-}M}$ 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 MVM can be used in place of $\mathbf{M}(\mathbf{MVM})^{-}\mathbf{M}$ in (9) because not every generalized inverse of MVM has M as a left and a right factor and that is an essential feature of (9). We therefore leave (9) as it is.

e. Mean and variance of BLUE($\mathbf{X}\beta$)

With MX being null and $E(\mathbf{y}) = \mathbf{X}\beta$ it is clear from (9) that

$$E[\text{BLUE}(\mathbf{X}\beta)] = \mathbf{X}\beta - \mathbf{VM}(\mathbf{MVM})^{-}\mathbf{MX}\beta = \mathbf{X}\beta ,$$

and (10)

$$\text{var}[\text{BLUE}(\mathbf{X}\beta)] = \mathbf{V} - \mathbf{VM}(\mathbf{MVM})^{-}\mathbf{MV} .$$

And, of course, this variance is invariant to $(\mathbf{MVM})^{-}$ because $\mathbf{VM}(\mathbf{MVM})^{-}\mathbf{MV} = \mathbf{L}[\mathbf{L}'\mathbf{M}'(\mathbf{MLL}'\mathbf{M}')^{-}\mathbf{ML}]\mathbf{L}'$ wherein the term within the square brackets is invariant to the generalized inverse.

2. A GENERALIZATION: ARBITRARY WEIGHTS

It is well known that the estimation equations coming from weighted least squares using an arbitrary non-null non-negative definite (n.n.d.) weighted matrix W (where, through being n.n.d. it can be factored as $\mathbf{W} = \mathbf{T}'\mathbf{T}$ for T real and of full column rank $r_{\mathbf{W}}$) are

$$\mathbf{X}'\mathbf{W}\mathbf{X}\beta^{0} = \mathbf{X}'\mathbf{W}\mathbf{y} \quad . \tag{11}$$

We might wish to denote the estimator of $\mathbf{X}\beta$ coming from (11) as WLSE($\mathbf{X}\beta$): with W as the weight matrix, $\text{WLSE}(\mathbf{X}\beta) = \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-}\mathbf{X}'\mathbf{W}\mathbf{y}$. But for easier notation, and to emphasize the dependence on W, we represent $\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-}\mathbf{X}'\mathbf{W}\mathbf{y}$ by $\hat{\mu}(\mathbf{W})$:

$$\hat{\mu}(\mathbf{W}) = \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-}\mathbf{X}'\mathbf{W}\mathbf{y} \quad . \tag{12}$$

The utility of this is that it is a generalized form of several familiar estimators; e.g. $\hat{\mu}(\mathbf{I})$ is OLSE($\mathbf{X}\beta$) of (1), and when V is non-singular $\hat{\mu}(\mathbf{V}^{-1})$ is the familiar $\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$ – a form which we later show (after Theorem 2) is a special case of BLUE($\mathbf{X}\beta$) of (9).

A problem with (12) is that the occurrence therein of $(\mathbf{X}'\mathbf{W}\mathbf{X})^{-}$ means that $\hat{\mu}(\mathbf{W})$ is not necessarily invariant to $(\mathbf{X}'\mathbf{W}\mathbf{X})^{-}$. Nor is $\hat{\mu}(\mathbf{W})$ necessarily unbiased for $\mathbf{X}\beta$. The desired invariance and unbiasedness are provided by the necessary and sufficient condition $\mathbf{X} = \mathbf{CWX}$ (with $\mathbf{X}'\mathbf{W} \neq \mathbf{0}$) for some $\mathbf{C}$ as in Theorem 1. But first, for the necessity proof, we have the following lemma.

Lemma 1: $\mathbf{FRB} = \mathbf{0}\ \forall\ \mathbf{R} \Rightarrow \mathbf{F} = \mathbf{0}$ for $\mathbf{B} \neq \mathbf{0}$.

Proof: Since $\mathbf{B} \neq \mathbf{0}$, there exists a vector $\boldsymbol{\tau} = \mathbf{Bu} \neq \mathbf{0}$. Then $\mathbf{FRB} = \mathbf{0} \Rightarrow \mathbf{FR}\boldsymbol{\tau} = \mathbf{0}$ for all $\mathbf{R}$. One possible $\mathbf{R}$ is $\mathbf{R} = \boldsymbol{\nu}\boldsymbol{\tau}'/\boldsymbol{\tau}'\boldsymbol{\tau}$ for any $\boldsymbol{\nu} \neq \mathbf{0}$. Then $\mathbf{FR}\boldsymbol{\tau} = \mathbf{0} \Rightarrow \mathbf{F}\boldsymbol{\nu} = \mathbf{0}\ \forall\ \boldsymbol{\nu} \neq \mathbf{0}$ and so $\mathbf{F} = \mathbf{0}$. **Q.E.D.**

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

Proof of sufficiency: That $\mathbf{X} = \mathbf{CWX}$ (with $\mathbf{X'W} \neq \mathbf{0}$) implies invariance and unbiasedness.

Using $\mathbf{X} = \mathbf{CWX}$ and $\mathbf{W} = \mathbf{T'T}$ gives (12) as

$$\hat{\mu}(\mathbf{W}) = \mathbf{CT'TX(X'T'TX)^-X'T'Ty} = \mathbf{CT'TX(TX)^+Ty}\ . \tag{13}$$

The second equality in (13) comes from using $\mathbf{TX}$ in place of $\mathbf{X}$ in the standard result $\mathbf{X(X'X)^-X'} = \mathbf{XX^+}$ implicit in (1); and occurrence of the unique $\mathbf{(TX)^+}$ in (13) ensures invariance of $\hat{\mu}(\mathbf{W})$ to $(\mathbf{X'WX})^-$. Similarly, because $\mathbf{TX(X'T'TX)^-X'T'TX} = \mathbf{TX}$,

$$\mathrm{E}[\hat{\mu}(\mathbf{W})] = \mathbf{CT'TX(X'T'TX)^-X'T'TX}\boldsymbol{\beta} = \mathbf{CWX}\boldsymbol{\beta} = \mathbf{X}\boldsymbol{\beta}\ .$$

Proof of necessity: That invariance and unbiasedness each imply $\mathbf{X} = \mathbf{CWX}$ for some $\mathbf{C}$.

We begin with the standard result that given any $(\mathbf{A'A})^-$, another generalized inverse of $\mathbf{A'A}$ is

$$(\mathbf{A'A})^{\sim} = (\mathbf{A'A})^-\mathbf{A'A}(\mathbf{A'A})^- + [\mathbf{I} - (\mathbf{A'A})^-\mathbf{A'A}]\mathbf{R} + \mathbf{S}[\mathbf{I} - \mathbf{A'A}(\mathbf{A'A})^-\ ] \tag{14}$$

for any $\mathbf{R}$ and $\mathbf{S}$ (e.g., Searle, 1982, p. 220). Post-multiplying (14) by $\mathbf{A}'$ reduces its first term to $(\mathbf{A'A})^-\mathbf{A}'$ and the term in $\mathbf{S}$ becomes null; then replacing $\mathbf{A}$ by $\mathbf{TX}$, pre-multiplying by $\mathbf{X}$ and post-multiplying by $\mathbf{Ty}$ and using $\mathbf{T'T} = \mathbf{W}$ gives

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

Given that $\hat{\mu}(\mathbf{W}) = \mathbf{X(X'WX)^-X'Wy}$ is invariant to $(\mathbf{X'WX})^-$, the left-hand side of (15) then equals the first term of its right side and so (15) becomes

$$\mathbf{X}[\mathbf{I} - (\mathbf{X'WX})^-\mathbf{X'WX}]\mathbf{RX'Wy} = \mathbf{0}\ \forall\ \mathbf{R} \text{ and } \mathbf{y} \neq \mathbf{0}\ .$$

Applying Lemma 1 to this, with $\mathbf{B} = \mathbf{X'W} \neq \mathbf{0}$ gives

$$\mathbf{X} = \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^-\mathbf{X}'\mathbf{W}\mathbf{X} = \mathbf{C}\mathbf{W}\mathbf{X} \text{ for } \mathbf{C} = \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^-\mathbf{X}' . \tag{16}$$

Thus, with $\mathbf{X}'\mathbf{W} \neq \mathbf{0}$, invariance implies $\mathbf{X} = \mathbf{CWX}$. Similarly, $\hat{\mu}(\mathbf{W})$ being unbiased for $\mathbf{X}\beta$ implies $\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^-\mathbf{X}'\mathbf{W}\mathbf{X}\beta = \mathbf{X}\beta \ \forall \ \beta$. This equality, ignoring the somewhat obscure occasions on which $\mathbf{AX}\beta = \mathbf{X}\beta$ does not imply $\mathbf{AX} = \mathbf{X}$ [see Christensen (1990) and Harville (1990), for example] also implies (16). Hence for the characteristics of invariance of $\hat{\mu}(\mathbf{W})$ to $(\mathbf{X}'\mathbf{W}\mathbf{X})^-$ and unbiasedness for $\mathbf{X}\beta$, each implies the other and (16). **Q.E.D.**

$\hat{\mu}(\mathbf{W})$ is the weighted least squares estimator of $\mathbf{X}\beta$ based on the (n.n.d.) weight matrix $\mathbf{W}$. It is a generalization of the Aitken (1935) estimator $\hat{\mu}(\mathbf{W}^{-1}) = \mathbf{X}(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}^{-1}\mathbf{y}$, which has $\mathbf{X}$ of full column rank and $\mathbf{W}$ non-singular. In that case there is no problem about invariance, for then $\mathbf{X} = \mathbf{C}\mathbf{W}^{-1}\mathbf{X}$ of Theorem 1 is satisfied for $\mathbf{C} = \mathbf{W}$. Difficulty arises with $\hat{\mu}(\mathbf{W})$ of (12) only when $\mathbf{X}$ has less than full column rank and $\mathbf{W}$ is singular, for then invariance and unbiasedness are met only when there exists a $\mathbf{C}$ such that $\mathbf{X} = \mathbf{CWX}$. Nevertheless, in view of $\hat{\mu}(\mathbf{W})$ 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 $\hat{\mu}(\mathbf{W})$, there would be merit in giving $\hat{\mu}(\mathbf{W})$ a name. In keeping with Plackett (1960), who describes $(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{y}$ as coming from an "extended" principle of least squares, we therefore might call $\hat{\mu}(\mathbf{W})$ the EWLSE, "extended weighted least squares estimator".

3. USING A $\mathbf{V}^-$ AS A WEIGHT MATRIX

If one wanted to use $\hat{\mu}(\mathbf{W})$ without the condition of Theorem 1, the invariance property could be defined away by using $\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^+\mathbf{X}'\mathbf{W}\mathbf{y}$. However, confining ourselves to $\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^+\mathbf{X}\mathbf{W}\mathbf{y}$ seems restrictive, and so we direct attention to $\hat{\mu}(\mathbf{W})$ and special cases thereof. In particular, we consider

$$\hat{\mu}(\mathbf{V}^-) = \mathbf{X}(\mathbf{X}'\mathbf{V}^-\mathbf{X})^-\mathbf{X}'\mathbf{V}^-\mathbf{y} \tag{17}$$

which is the analog (for singular $\mathbf{V}$) of the familiar (for non-singular $\mathbf{V}$) estimator $\hat{\mu}(\mathbf{V}^{-1}) = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$ that is both the BLUE of $\mathbf{X}\beta$ and, under normality with non-singular $\mathbf{V}$, the maximum likelihood estimator (MLE) of $\mathbf{X}\beta$.

An invariance property

Although $\hat{\mu}(\mathbf{V}^-)$ of (17) is not invariant to the choice of generalized inverses $\mathbf{V}^-$ and $(\mathbf{X}'\mathbf{V}^-\mathbf{X})^-$, we use the following lemma to show that it is when $\mathbf{V}\mathbf{V}^-\mathbf{X} = \mathbf{X}$, a condition that arises in Theorem 2 for the equality of $\hat{\mu}(\mathbf{V}^-)$ and $\text{BLUE}(\mathbf{X}\boldsymbol{\beta})$.

Lemma 2: $\mathbf{V}\mathbf{V}^-\mathbf{X} = \mathbf{X}$ implies (i) $\mathbf{V}\mathbf{V}^-\mathbf{X} = \mathbf{X}\ \forall\ \mathbf{V}^-$ and (ii) $\mathbf{X}'\mathbf{V}^-\mathbf{X}$ is invariant to $\mathbf{V}^-$.

Also, for almost all $\mathbf{y} \sim (\mathbf{X}\boldsymbol{\beta}, \mathbf{V})$,

$\mathbf{V}\mathbf{V}^-\mathbf{X} = \mathbf{X}$ implies (iii) $\mathbf{V}\mathbf{V}^-\mathbf{y} = \mathbf{y}\ \forall\ \mathbf{V}^-$ and (iv) $\mathbf{X}'\mathbf{V}^-\mathbf{y}$ is invariant to $\mathbf{V}^-$.

Proof. Suppose $\mathbf{X} = \mathbf{V}\mathbf{V}^-\mathbf{X}$ for some particular $\mathbf{V}^-$. Let $\mathbf{V}^\sim$ be a generalized inverse of $\mathbf{V}$ different from $\mathbf{V}^-$. Then we have the following.

(i) $\quad \mathbf{V}\mathbf{V}^\sim\mathbf{X} = \mathbf{V}\mathbf{V}^\sim\mathbf{V}\mathbf{V}^-\mathbf{X} = \mathbf{V}\mathbf{V}^-\mathbf{X} = \mathbf{X}.$

(ii) $\quad \mathbf{X}'\mathbf{V}^\sim\mathbf{X} = \mathbf{X}'\mathbf{V}^{-'}\mathbf{V}'\mathbf{V}^\sim\mathbf{V}\mathbf{V}^-\mathbf{X} = \mathbf{X}'\mathbf{V}^{-'}\mathbf{V}\mathbf{V}^-\mathbf{X} = \mathbf{X}'\mathbf{V}^-\mathbf{X}\,.$

[Recall that $\mathbf{V}' = \mathbf{V}$, $\mathbf{V}\mathbf{V}^\sim\mathbf{V} = \mathbf{V}$ and $\mathbf{V}^{-'}$ is a generalized inverse of $\mathbf{V}$.]

(iii)
$$\begin{aligned}\mathbf{0} &= (\mathbf{I}-\mathbf{V}\mathbf{V}^-)\mathbf{V}(\mathbf{I}-\mathbf{V}\mathbf{V}^-)' \\ &= (\mathbf{I}-\mathbf{V}\mathbf{V}^-)\mathrm{E}[(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'](\mathbf{I}-\mathbf{V}\mathbf{V}^-)' \\ &= \mathrm{E}(\mathbf{z}\mathbf{z}') \text{ for } \mathbf{z} = (\mathbf{I}-\mathbf{V}\mathbf{V}^-)(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})\,.\end{aligned}$$

But, with probability one, $\mathrm{E}(\mathbf{z}\mathbf{z}') = \mathbf{0}$ implies $\mathbf{z} = \mathbf{0}$. Hence $(\mathbf{I}-\mathbf{V}\mathbf{V}^-)(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}) = \mathbf{0}$. With $\mathbf{X} = \mathbf{V}\mathbf{V}^-\mathbf{X} = \mathbf{X}$ this gives $\mathbf{y} = \mathbf{V}\mathbf{V}^-\mathbf{y}$ and then

$$\mathbf{V}\mathbf{V}^\sim\mathbf{y} = \mathbf{V}\mathbf{V}^\sim\mathbf{V}\mathbf{V}^-\mathbf{y} = \mathbf{V}\mathbf{V}^-\mathbf{y} = \mathbf{y}\,.$$

(iv) $\quad \mathbf{X}'\mathbf{V}^-\mathbf{y} = \mathbf{X}'\mathbf{V}^-\mathbf{V}\mathbf{V}^\sim\mathbf{y} = \mathbf{X}'\mathbf{V}^\sim\mathbf{y}\,.$ Q.E.D.

Now, with $\mathbf{X} = \mathbf{V}\mathbf{V}^-\mathbf{X}$ consider

$$\hat{\mu}(\mathbf{V}^-) = \mathbf{X}'(\mathbf{X}'\mathbf{V}^-\mathbf{X})^-\mathbf{X}'\mathbf{V}^-\mathbf{y} = \mathbf{V}\mathbf{V}^-\mathbf{X}(\mathbf{X}'\mathbf{V}^-\mathbf{X})^-\mathbf{X}'\mathbf{V}^-\mathbf{y}\,. \tag{18}$$

In the right-most term of (18), $\mathbf{V}^-$ is n.n.d. (because $\mathbf{V}$ is) and so $\mathbf{V}^- = \mathbf{K}'\mathbf{K}$ for some $\mathbf{K}$, and then letting $\mathbf{N} = \mathbf{K}\mathbf{X}$,

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

where $\mathbf{N}(\mathbf{N}'\mathbf{N})^-\mathbf{N}'$ is invariant to $(\mathbf{N}'\mathbf{N})^-$, this being a standard result pertaining to $(\mathbf{N}'\mathbf{N})^-$ for any real $\mathbf{N}$ (e.g., Searle, 1982, Section 8.6c). And since by parts (ii) and (iv) of Lemma 2 the terms $\mathbf{X}'\mathbf{V}^-\mathbf{X}$ and

$\mathbf{X}'\mathbf{V}^-\mathbf{y}$ in $\hat{\mu}(\mathbf{V}^-)$ are invariant to $\mathbf{V}^-$ when $\mathbf{X} = \mathbf{V}\mathbf{V}^-\mathbf{X}$, we then have $\hat{\mu}(\mathbf{V}^-)$ being unique, for given $\mathbf{V}$, when $\mathbf{X}$ and $\mathbf{V}^-$ are such that $\mathbf{X} = \mathbf{V}\mathbf{V}^-\mathbf{X}$.

The algebraic similarity of $\hat{\mu}(\mathbf{V}^-)$ of (17) to $\hat{\mu}(\mathbf{V}^{-1}) = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^-\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$, the well-known BLUE of $\mathbf{X}\boldsymbol{\beta}$ when $\mathbf{V}$ is non-singular, is in sharp contrast to the dissimilarity of both $\hat{\mu}(\mathbf{V}^-)$ and $\hat{\mu}(\mathbf{V}^{-1})$ to $\text{BLUE}(\mathbf{X}\boldsymbol{\beta}) = \mathbf{y} - \mathbf{VM}(\mathbf{MVM})^-\mathbf{My}$ of (9). This begs the question "When does $\hat{\mu}(\mathbf{V}^-)$ equal $\text{BLUE}(\mathbf{X}\boldsymbol{\beta})$?" Theorem 2 provides the answer: when $\mathbf{V}\mathbf{V}^-\mathbf{X} = \mathbf{X}$. And paralleling that is Theorem 4 which answers the question "When does $\hat{\mu}(\mathbf{V}^-) = \text{OLSE}(\mathbf{X}\boldsymbol{\beta})$?" with "When both $\hat{\mu}(\mathbf{V}^-)$ and $\text{OLSE}(\mathbf{X}\boldsymbol{\beta})$ equal $\text{BLUE}(\mathbf{X}\boldsymbol{\beta})$." And in between is Theorem 3 that $\text{OLSE}(\mathbf{X}\boldsymbol{\beta})$ equals $\text{BLUE}(\mathbf{X}\boldsymbol{\beta})$ when $\mathbf{VX} = \mathbf{XB}$ for some $\mathbf{B}$. In all of Theorems 2, 3 and 4, the conditions for equality are necessary and sufficient. We now state and prove those theorems.

Theorem 2: (a) **Sufficiency** if $\mathbf{V}\mathbf{V}^-\mathbf{X} = \mathbf{X}$, then $\hat{\mu}(\mathbf{V}^-) = \text{BLUE}(\mathbf{X}\boldsymbol{\beta})$.

(a) **Necessity** if $\hat{\mu}(\mathbf{V}^-) = \text{BLUE}(\mathbf{X}\boldsymbol{\beta})$ for some symmetric reflexive generalized inverse $(\mathbf{V}^-\mathbf{V}\mathbf{V}^- = \mathbf{V}^- = \mathbf{V}^{-\prime})$ then $\mathbf{V}\mathbf{V}^{\sim}\mathbf{X} = \mathbf{X}$ for every generalized inverse $\mathbf{V}^{\sim}$.

When using only unique Moore-Penrose inverses, an early proof of the sufficiency part of this theorem is due to Rao and Mitra (1971), and of the necessity part to Pukelsheim (1974). We offer new proofs which are somewhat shorter than theirs, and which do not involve Moore-Penrose inverses.

Proof of sufficiency: First note that because $\mathbf{V}$ is non-negative definite, so is $\mathbf{V}^-$ and therefore $\mathbf{V}^- = \mathbf{LL}'$ for some $\mathbf{L}$, and hence $\mathbf{V}^-\mathbf{X} = \mathbf{V}^-\mathbf{X}(\mathbf{X}'\mathbf{V}^-\mathbf{X})^-\mathbf{X}'\mathbf{V}^{-\prime}\mathbf{X}$. Then define

$$\mathbf{Q} = \mathbf{V}^- - \mathbf{V}^-\mathbf{X}(\mathbf{X}'\mathbf{V}^-\mathbf{X})^-\mathbf{X}'\mathbf{V}^{-\prime} = \mathbf{Q}' \text{ with } \mathbf{QX} = \mathbf{0} \tag{19}$$

and so observe that, with $\mathbf{Q}$ and $\mathbf{M}$ being symmetric,

$$\mathbf{QM} = \mathbf{Q}(\mathbf{I} - \mathbf{XX}^+) = \mathbf{Q} = \mathbf{MQ} = \mathbf{MQM}\,. \tag{20}$$

Hence

$$\mathbf{MVMQMVM} = \mathbf{MVQVM} = \mathbf{MVM} - \mathbf{MVV}^-\mathbf{X}(\mathbf{X}'\mathbf{V}^-\mathbf{X})^-\mathbf{X}'\mathbf{V}^{-\prime}\mathbf{VM}\,. \tag{21}$$

Then if $\mathbf{V}\mathbf{V}^-\mathbf{X} = \mathbf{X}$, the second term of (21) contains $\mathbf{MX} = \mathbf{0}$, and so then $\mathbf{MVMQMVM} = \mathbf{MVM}$; i.e., $\mathbf{Q}$ is a generalized inverse of $\mathbf{MVM}$, or $(\mathbf{MVM})^- = \mathbf{Q}$. Hence from (9)

$$\begin{aligned}\mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta}) &= \mathbf{y} - \mathbf{VM}(\mathbf{MVM})^{-}\mathbf{My} \\ &= \mathbf{y} - \mathbf{VMQMy} \\ &= \mathbf{y} - \mathbf{VQy}, \text{ from (20)} \\ &= \mathbf{y} - [\mathbf{VV}^{-}\mathbf{y} - \mathbf{VV}^{-}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-}\mathbf{y}] \; . \end{aligned} \tag{22}$$

But with $\mathbf{X} = \mathbf{VV}^{-}\mathbf{X}$ we have from Lemma 2(iii) $\mathbf{y} = \mathbf{VV}^{-}\mathbf{y}$ (almost surely) and this, together with $\mathbf{X} = \mathbf{VV}^{-}\mathbf{X}$ itself, reduces (22) to

$$\mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta}) = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-}\mathbf{y} \; , \tag{23}$$

$$= \hat{\mu}(\mathbf{V}^{-}) \; . \qquad \textbf{Q.E.D.}$$

And, when $\mathbf{X} = \mathbf{VV}^{-}\mathbf{X}$ we have $\hat{\mu}(\mathbf{V}^{-})$ invariant to the generalized inverses therein, as explained following (18).

Proof of necessity Using $\mathbf{V}^{-}$ as any generalized inverse of $\mathbf{V}$, we start with $\hat{\mu}(\mathbf{V}^{-}) = \mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta})$ which, from (17) and (9), is

$$\mathbf{X}(\mathbf{X}'\mathbf{V}^{-}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-}\mathbf{y} = \mathbf{y} - \mathbf{VM}(\mathbf{MVM})^{-}\mathbf{My} \; . \tag{24}$$

We want this to hold for all $\mathbf{y}$. Hence we want

$$\mathbf{X}(\mathbf{X}'\mathbf{V}^{-}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-} = \mathbf{I} - \mathbf{VM}(\mathbf{MVM})^{-}\mathbf{M} \; . \tag{25}$$

Post-multiplying (25) by $\mathbf{X}$ and using $\mathbf{MX} = \mathbf{0}$ gives

$$\mathbf{X}(\mathbf{X}'\mathbf{V}^{-}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-}\mathbf{X} = \mathbf{X} \; . \tag{26}$$

Pre-multiplying (25) by $\mathbf{X}'\mathbf{V}^{-}$ gives

$$\mathbf{X}'\mathbf{V}^{-}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-} = \mathbf{X}'\mathbf{V}^{-} - \mathbf{X}'\mathbf{V}^{-}\mathbf{VM}(\mathbf{MVM})^{-}\mathbf{M} \; . \tag{27}$$

But, on using (26), the left-hand side of (27) reduces to $\mathbf{X}'\mathbf{V}^{-}$ and so (27) becomes

$$\mathbf{X}'\mathbf{V}^{-}\mathbf{VM}(\mathbf{MVM})^{-}\mathbf{M} = \mathbf{0} \; . \tag{28}$$

Finally, post-multiplying (25) again, this time by $\mathbf{VV}^{-}\mathbf{X}$ (which here, in this proof of necessity is not given as equaling $\mathbf{X}$ — this is what we are try to prove), gives

$$\mathbf{X}(\mathbf{X}'\mathbf{V}^{-}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-}\mathbf{VV}^{-}\mathbf{X} = \mathbf{VV}^{-}\mathbf{X} - \mathbf{VM}(\mathbf{MVM})^{-}\mathbf{MVV}^{-}\mathbf{X} \; . \tag{29}$$

At this point we use the assumed symmetry of V^-; the right-most term of (29) is th n $\mathbf{0}$ – from (28). And we also now use the assumed reflexiveness of V, namely $V^-VV^- = V^-$. Thus (29) becomes

$$X(X'V^-X)^-X'V^-X = VV^-X\,. \tag{30}$$

Using the same symmetric reflexive V^- in (26) in combination with (30) therefore yields

$$VV^-X = X, \qquad \text{for that } V^-\,.$$

But by Lemma 2 we then have

$$VV^{\sim}X = X, \qquad \text{for any } V^{\sim}\,. \tag{31}$$

Q.E.D.

An interesting feature of this necessity condition is that it must start with $\hat{\mu}(V^-) = \text{BLUE}(X\beta)$ with V^- being a symmetric reflexive inverse but it finishes up with $VV^{\sim}X = X$ for any generalized inverse $V^{\sim}$. The only uses of the symmetry and reflexive properties are, respectively, for the right-most term of (29) to be null, from (28), and for the left-hand side of (29) to be the left-hand side of (30). Note, too, that if V^- is not symmetric and reflexive it can be replaced by $V^{-\prime}VV^-$, which is, and in making that replacement VV^-X becomes $VV^{-\prime}VV^-X = VV^-X$, i.e., VV^-X is unchanged by the replacement.

And now that $VV^-X = X$ has been established as the necessary condition, we know from the argument following (18) that $\hat{\mu}(V^-)$ is invariant V^-.

a. Summary

The preceding theorem gives the useful result that if $VV^-X = X$ then BLUE($X\beta$) of (3), (4), (5) or (9), which include M, can in fact be calculated as $\hat{\mu}(V^-)$; i.e., $\text{BLUE}(X\beta) = y - VM(MVM)^-My = \hat{\mu}(V^-) = X(X'V^-X)^-X'V^-y$.

b. Non-singular V

In the special case of non-singular V (when $V^- = V^{-1}$) the condition $VV^-X = X$ is certainly satisfied and we have the existence of V^{-1} implying $\text{BLUE}(X\beta) = y - VM(MVM)^-My = X(X'V^{-1}X)^-X'V^{-1}y$. And, of course, when X has full column rank this reduces further, to the familiar $X\hat{\beta} = X(X'V^{-1}X)^{-1}X'V^{-1}y$.

c. $\mathbf{V} + \mathbf{XX}'$ in place of $\mathbf{V}$

Rao and Mitra (1971) suggest that BLUE($\mathbf{X}\beta$) can be calculated using $\mathbf{U} = \mathbf{V} + \mathbf{XX}'$ in place of $\mathbf{V}$. This is so because in any of the equivalent forms (3), (4), (5) or (9) for BLUE($\mathbf{X}\beta$), the matrix $\mathbf{V}$ occurs only in the form $\mathbf{VM}$. Therefore, in replacing $\mathbf{V}$ by $\mathbf{U}$, the product $\mathbf{VM}$ would become $\mathbf{UM} = \mathbf{VM} + \mathbf{XX}'\mathbf{M} = \mathbf{VM}$ because $\mathbf{XX}'\mathbf{M} = \mathbf{X}(\mathbf{MX})' = \mathbf{0}$ since $\mathbf{MX} = \mathbf{0}$. Hence BLUE($\mathbf{X}\beta$) is not affected by using $\mathbf{U}$ in place of $\mathbf{V}$. Therefore, we will have $\hat{\mu}(\mathbf{U}^-) =$ BLUE($\mathbf{X}\beta$) provided $\mathbf{UU}^-\mathbf{X} = \mathbf{X}$. This is so because we get from

$$(\mathbf{I} - \mathbf{UU}^-)\mathbf{U}(\mathbf{I} - \mathbf{UU}^-)' = \mathbf{0}\ ,$$

$$(\mathbf{I} - \mathbf{UU}^-)\mathbf{V}(\mathbf{I} - \mathbf{UU}^-)' + (\mathbf{I} - \mathbf{UU}^-)\mathbf{XX}'(\mathbf{I} - \mathbf{UU}^-)' = \mathbf{0}\ .$$

Each of the two terms in this sum is n.n.d.; therefore the sum is null only if each term is null. Hence $(\mathbf{I} - \mathbf{UU}^-)\mathbf{XX}'(\mathbf{I} - \mathbf{UU}^-)' = \mathbf{0}$ and so, because $\mathbf{U}$ and $\mathbf{X}$ are real, $(\mathbf{I} - \mathbf{UU}^-)\mathbf{X} = \mathbf{0}$, i.e., $\mathbf{UU}^-\mathbf{X} = \mathbf{X}$.

5. WHEN DOES BLUE($\mathbf{X}\beta$) = OLSE($\mathbf{X}\beta$)?

Theorem 3. BLUE($\mathbf{X}\beta$) = OLSE($\mathbf{X}\beta$) if and only if $\mathbf{VX} = \mathbf{XB}$ for some $\mathbf{B}$.

This result is due to Zyskind (1967). It is part of a whole series of equivalent results; see, e.g., Puntanen and Styan, 1989.

Proof. Given $\mathbf{VX} = \mathbf{XB}$, we prove sufficiency by noting that

$$\begin{aligned}\mathbf{MVM} &= \mathbf{MV}(\mathbf{I} - \mathbf{XX}^+) = \mathbf{MV} - \mathbf{MVXX}^+ = \mathbf{MV} - \mathbf{MXBX}^+ \\ &= \mathbf{MV} \text{ because } \mathbf{MX} = \mathbf{0} \\ &= \mathbf{VM} \text{ because } \mathbf{MVM} \text{ is symmetric, and hence so is } \mathbf{MV}\ .\end{aligned}$$

Therefore in (4), $\mathbf{HVM} = (\mathbf{I} - \mathbf{M})\mathbf{VM} = \mathbf{0}$ and so (4) reduces to BLUE($\mathbf{X}\beta$) = OLSE($\mathbf{X}\beta$).

Proving necessity begins with BLUE($\mathbf{X}\beta$) = OLSE($\mathbf{X}\beta$), which from (9) and (1) gives $\mathbf{VM}(\mathbf{MVM})^-\mathbf{My} = \mathbf{My}\ \forall\ \mathbf{y}$. Hence, $\mathbf{VM}(\mathbf{MVM})^-\mathbf{M} = \mathbf{M}$, post-multiplication of which by $\mathbf{VM}$ gives $\mathbf{VM} = \mathbf{MVM}$, and so because $\mathbf{MVM}$ is symmetric $\mathbf{MV} = \mathbf{VM}$. This, with $\mathbf{M} = \mathbf{I} - \mathbf{XX}^+$, is

$$(\mathbf{I} - \mathbf{XX}^+)\mathbf{V} = \mathbf{V}(\mathbf{I} - \mathbf{XX}^+)\ ,$$

which yields $\mathbf{XX}^+\mathbf{VX} = \mathbf{VXX}^+\mathbf{X} = \mathbf{VX}$; i.e., $\mathbf{VX} = \mathbf{XX}^+\mathbf{VX} = \mathbf{XB}$ for $\mathbf{B} = \mathbf{X}^+\mathbf{VX}$. **Q.E.D.**

6. WHEN DOES $\hat{\mu}(\mathbf{V}^-) = \mathrm{OLSE}(\mathbf{X}\boldsymbol{\beta})$?

Theorem 4. The two estimators $\hat{\mu}(\mathbf{V}^-)$ and OLSE($\mathbf{X}\boldsymbol{\beta}$) are equal for any $\mathbf{V}^-$ if and only if they each equal BLUE($\mathbf{X}\boldsymbol{\beta}$); in which case $\mathbf{VV}^-\mathbf{X} = \mathbf{X}$ and $\mathbf{VX} = \mathbf{XB}$ for some $\mathbf{B}$.

Proof. Sufficiency is obvious: if $\hat{\mu}(\mathbf{V}^-)$ and OLSE($\mathbf{X}\boldsymbol{\beta}$) each equal BLUE($\mathbf{X}\boldsymbol{\beta}$) then they equal each other, and by Theorems 2 and 3, $\mathbf{VV}^-\mathbf{X} = \mathbf{X}$, and $\mathbf{VX} = \mathbf{XB}$ for some $\mathbf{B}$.

Proving necessity starts with $\hat{\mu}(\mathbf{V}^-) = \mathrm{OLSE}(\mathbf{X}\boldsymbol{\beta})$ which, from (17) and (1) is

$$\mathbf{X}(\mathbf{X}'\mathbf{V}^-\mathbf{X})^-\mathbf{X}'\mathbf{V}^-\mathbf{y} = \mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{y} \;\; \forall \;\; \mathbf{y} .$$

Therefore we want

$$\mathbf{X}(\mathbf{X}'\mathbf{V}^-\mathbf{X})^-\mathbf{X}'\mathbf{V}^- = \mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}' . \tag{32}$$

Post-multiplying (32) by $\mathbf{X}$ gives

$$\mathbf{X}(\mathbf{X}'\mathbf{V}^-\mathbf{X})^-\mathbf{X}'\mathbf{V}^-\mathbf{X} = \mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{X} = \mathbf{X} . \tag{33}$$

And pre-multiplying (32) by $\mathbf{X}^-$ and post-multiplying it by $\mathbf{VV}^-\mathbf{X}$ gives

$$\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{V}^-\mathbf{X})^-\mathbf{X}'\mathbf{V}^-\mathbf{VV}^-\mathbf{X} = \mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{VV}^-\mathbf{X} = \mathbf{X}'\mathbf{VV}^-\mathbf{X} .$$

Now, as in Theorem 2, treat $\mathbf{V}^-$ as reflexive and get

$$\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{V}^-\mathbf{X})^-\mathbf{X}'\mathbf{V}^-\mathbf{X} = \mathbf{X}'\mathbf{VV}^-\mathbf{X} ,$$

which, on using (33) is

$$\mathbf{X}'\mathbf{X} = \mathbf{X}'\mathbf{VV}^-\mathbf{X} . \tag{34}$$

Since $\mathbf{V}^+$ is a permissible (symmetric, reflexive) form for $\mathbf{V}^-$, with $\mathbf{VV}^+$ being symmetric, we take (34) as

$$\mathbf{0} = \mathbf{X}'\mathbf{X} - \mathbf{X}'\mathbf{VV}^+\mathbf{X} = \mathbf{X}'(\mathbf{I} - \mathbf{VV}^+)\mathbf{X} = [(\mathbf{I} - \mathbf{VV}^+)\mathbf{X}]'[(\mathbf{I} - \mathbf{VV}^+)\mathbf{X}] .$$

Therefore, because $(\mathbf{I} - \mathbf{VV}^+)\mathbf{X}$ is real, it is null, which gives $\mathbf{X} = \mathbf{VV}^+\mathbf{X} = \mathbf{VV}^-\mathbf{X}$. Thus $\hat{\mu}(\mathbf{V}^-) = \mathrm{OLSE}(\mathbf{X}\boldsymbol{\beta})$ implies $\mathbf{X} = \mathbf{VV}^-\mathbf{X}$ which, as in Theorem 2, implies $\hat{\mu}(\mathbf{V}^-) = \mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta})$. **Q.E.D.**

Acknowledgments

Grateful thanks go to the referee for many helpful and inciteful comments, and to Friedrich Pukelsheim for numerous discussions over the years on this and allied topics.

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.

Christensen, R. (1990). Comment on Puntanen and Styan (1989). *The Amer. Statist.* 44:191-192.

Harville, D.A. (1990). Comment on Puntanen and Styan (1989). *The Amer. Statist.* 44:192.

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. *The Amer. Statist.* 43:153-161.

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: 289-230.

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

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.