# ESTIMATING THE MEAN VECTOR IN LINEAR MODELS

Shayle R. Searle* and Friedrich Pukelsheim

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

BU-912-M January, 1987

**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 the latter, conditions for the equivalence of BLUE to OLSE and to GLSE are established with new and shorter proofs.

## 1. INTRODUCTION

### 1.1 The model

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

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

with

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

and

$$\mathbf{X}, \text{ of order } N \times p, \text{ has rank } r_{\mathbf{X}} \le p < N. \tag{3}$$

*Most of this paper was prepared while the first author was on leave from Cornell University, supported by a Senior U.S. Scientist Award from the Alexander von Humboldt-Stiftung, Bonn, Federal Republic of Germany.

Frequent use is made of

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

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

The paper deals with estimation of the vector $\mathbf{X}\boldsymbol{\beta}$, all elements of which are estimable, as is every linear combination of them, $\boldsymbol{\lambda}'\mathbf{X}\boldsymbol{\beta}$ for any $\boldsymbol{\lambda}'$. We begin with ordinary least squares estimation (OLSE), followed by weighted least squares estimation (WLSE) using an arbitrary n.n.d. weight matrix $\mathbf{W}$. Neither of these estimation methods involves the dispersion matrix $\text{var}(\mathbf{y}) = \mathbf{V}$. 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 $\mathbf{V}$ non-singular and in Section 3 for $\mathbf{V}$ singular. Section 4 is a summary and Section 5 deals with relationships among the methods. Section 6 is an appendix containing pertinent matrix results (many of them well known). They have equation and lemma numbers prefixed with A.

### 1.2 OLSE: ordinary least squares estimation

Ordinary least squares estimation is based on minimizing

$$S = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \tag{5}$$

with respect to $\boldsymbol{\beta}$. This leads, as is well-known, to equations

$$\mathbf{X}'\mathbf{X}\boldsymbol{\beta}^0 = \mathbf{X}'\mathbf{y} \tag{6}$$

whereupon $\mathbf{X}\boldsymbol{\beta}^0$ is the corresponding estimator of $\mathbf{X}\boldsymbol{\beta}$, to be denoted OLSE:

$$\text{OLSE} = \mathbf{X}\boldsymbol{\beta}^0 = \mathbf{X}(\mathbf{X}'\mathbf{X})^-\mathbf{X}'\mathbf{y} = (\mathbf{I} - \mathbf{M})\mathbf{y}. \tag{7}$$

Invariance of OLSE to $(\mathbf{X}'\mathbf{X})^-$ comes from (A4); and from (A5), OLSE is unbiased for $\mathbf{X\beta}$. Also, using (A3) and (A4)

$$\text{var(OLSE)} = \mathbf{X(X'X)^{-}X'VX(X'X)^{-}X'}. \tag{8}$$

**1.3 WLSE: weighted least squares estimation**

Weighted least squares using an arbitrary non-negative definite (n.n.d.) weight matrix $\mathbf{W}$ (where, through being n.n.d. it can be factored as $\mathbf{W} = \mathbf{H'H}$ for $\mathbf{H}$ of full row rank $r_{\mathbf{W}}$) involves minimizing

$$S_{\mathbf{W}} = (\mathbf{y} - \mathbf{X\beta})'\mathbf{W}(\mathbf{y} - \mathbf{X\beta}) = (\mathbf{Hy} - \mathbf{HX\beta})'(\mathbf{Hy} - \mathbf{HX\beta}). \tag{9}$$

Based on (6) this gives

$$\mathbf{X'H'HX\beta^0} = \mathbf{X'H'Hy}, \text{ i.e., } \mathbf{X'WX\beta^0} = \mathbf{X'Wy}. \tag{10}$$

Then $\mathbf{X\beta^0}$ derived from this is the weighted least squares estimator, to be denoted $\text{WLSE}_{\mathbf{W}}$, and is

$$\text{WLSE}_{\mathbf{W}} = \mathbf{X(X'WX)^{-}X'Wy}. \tag{11}$$

The subscript $\mathbf{W}$ is for $\mathbf{W}$ being the weight matrix; other matrices are used in place of $\mathbf{W}$ in subsequent sections. The dispersion matrix of (11) is

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

The occurrence of a generalized inverse of $\mathbf{X'WX}$ in (11) in the form $\mathbf{X(X'WX)^{-}X'W}$ means that $\text{WLSE}_{\mathbf{W}}$ of (11) is not necessarily invariant to what is used for $\mathbf{(X'WX)^{-}}$. Nor is $\text{WLSE}_{\mathbf{W}}$ necessarily unbiased for $\mathbf{X\beta}$. The following theorem provides conditions under which invariance and unbiasedness arise.

**Theorem 1.** Each of the following conditions implies the other three.

(i) $\text{WLSE}_{\mathbf{W}}$ is invariant to $\mathbf{(X'WX)^{-}}$.

(ii) $\text{WLSE}_{\mathbf{W}}$ is unbiased for $\mathbf{X\beta}$.

(iii) $\mathbf{X} = \mathbf{DHX}$ for some $\mathbf{D}$.

(iv) $\mathbf{X} = \mathbf{CWX}$ for some $\mathbf{C}$.

**Proof.** (iii) and (iv) are equivalent because with $\mathbf{W} = \mathbf{H'H}$ for $\mathbf{H}$ of full row rank the existence of either $\mathbf{D}$ or $\mathbf{C}$ implies existence of the other through the relationships $\mathbf{D(HH')^{-1}H} = \mathbf{C}$ and $\mathbf{CH'} = \mathbf{D}$. Proof that (iii) is a necessary and sufficient condition for each of (i) and (ii) is as follows.

First, if $\mathbf{X} = \mathbf{DHX}$ then from (11)

$$\text{WLSE}_{\mathbf{W}} = \mathbf{DHX(X'H'HX)^{-}X'H'Hy}$$

and

$$E(\text{WLSE}_{\mathbf{W}}) = \mathbf{DHX(X'H'HX)^{-}X'H'HX\beta}. \tag{13}$$

Using $\mathbf{HX}$ for $\mathbf{X}$ in (A4) and (A5) shows, respectively, that $\text{WLSE}_{\mathbf{W}}$ is invariant to $\mathbf{(X'WX)^{-}}$, and that $E(\text{WLSE}_{\mathbf{W}}) = \mathbf{X\beta}$. Thus sufficiency of (iii) is established.

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

$$\begin{aligned}\mathbf{(N'N)^{\sim}N'} &= \mathbf{(N'N)^{-}N'N(N'N)^{-}N'} + \mathbf{[I - (N'N)^{-}N'N]PN'} + \mathbf{Q[I - N'N(N'N)^{-}]N'} \\ &= \mathbf{(N'N)^{-}N'} + \mathbf{[I - (N'N)^{-}N'N]PN'},\end{aligned} \tag{14}$$

on using $\mathbf{N}$ for $\mathbf{X}$ in the transpose of (A5). With $\mathbf{W} = \mathbf{H'H}$, now use $\mathbf{HX}$ for $\mathbf{N}$ in (14) and simultaneously pre-multiply (14) by $\mathbf{X}$ and post-multiply it by $\mathbf{Hy}$. This yields

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

But if $\text{WLSE}_{\mathbf{W}} = \mathbf{X(X'WX)^{-}X'Wy}$ is invariant to $\mathbf{(X'WX)^{-}}$, then (15) reduces to

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

Therefore, by Lemma A2, with $\mathbf{Hy} \neq \mathbf{0}$,

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

Also, if $\text{WLSE}_{\mathbf{W}}$ is unbiased, (11) gives $\mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-}\mathbf{X}'\mathbf{W}\mathbf{X}\boldsymbol{\beta} = \mathbf{X}\boldsymbol{\beta}\ \forall\ \boldsymbol{\beta}$, and this too, implies (17). Hence if $\text{WLSE}_{\mathbf{W}}$ is invariant to $(\mathbf{X}'\mathbf{W}\mathbf{X})^{-}$ and/or is unbiased for $\mathbf{X}\boldsymbol{\beta}$, (17) holds and so $\mathbf{X} = \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-}\mathbf{X}'\mathbf{W}\mathbf{X} = \mathbf{D}\mathbf{H}\mathbf{X}$ for $\mathbf{D} = \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{-}\mathbf{X}'\mathbf{H}'$.

**Q.E.D.**

The name WLSE warrants comment. There is considerable agreement that $\text{WLSE}_{\mathbf{W}}$ is indeed the weighted least squares estimator of $\mathbf{X}\boldsymbol{\beta}$ based on the (n.n.d.) weight matrix $\mathbf{W}$. It is a generalization of the Aitken (1935) estimator with $\mathbf{X}$ of full column rank and $\mathbf{W}$ non-singular: $\mathbf{X}(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}^{-1}\mathbf{y}$. In this case there is no problem about invariance; nor is there with

$$\text{WLSE}_{\mathbf{W}^{-1}} = \mathbf{X}(\mathbf{X}'\mathbf{W}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{W}^{-1}\mathbf{y} \tag{18}$$

for then $\mathbf{X} = \mathbf{C}\mathbf{W}^{-1}\mathbf{X}$ of Theorem 1 is satisfied for $\mathbf{C} = \mathbf{W}$. Difficulty arises only when $\mathbf{X}$ has less than full column rank and some kind of generalized inverse has to be used, so giving rise to invariance requirements as in Theorem 1. Of course, these could be avoided by always using $(\mathbf{X}'\mathbf{W}\mathbf{X})^{+}$ and defining $\text{WLSE}_{\mathbf{W}}$ accordingly, calling it $\text{WLSE}_{\mathbf{W}(MP)}$:

$$\text{WLSE}_{\mathbf{W}(MP)} = \mathbf{X}(\mathbf{X}'\mathbf{W}\mathbf{X})^{+}\mathbf{X}'\mathbf{W}\mathbf{y}.$$

This seems restrictive, and we prefer to direct attention to WLSE of (11).

## 2. NON-SINGULAR V

### 2.1 GLSE: generalized least squares estimation

Although the Aitken estimator is sometimes called the generalized least squares estimator, we reserve this name for the weighted least squares estimator when $\mathbf{V}^{-1}$ (or, later, $\mathbf{V}^{-}$) is used in place of $\mathbf{W}$:

$$\text{GLSE} = \text{WLSE}_{\mathbf{V}^{-1}} = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}. \tag{19}$$

with dispersion matrix, from (12),

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

Clearly, the condition $\mathbf{X} = \mathbf{CWX}$ of Theorem 1 with $\mathbf{W} = \mathbf{V}^{-1}$ is satisfied for $\mathbf{C} = \mathbf{V}$, and so (19) is invariant to its generalized inverse and is unbiased for $\mathbf{X}\boldsymbol{\beta}$

**2.2 MLE: maximum likelihood estimation**

On assuming that elements of $\mathbf{y}$ have a multivariate normal distribution, $\mathbf{y} \sim \mathcal{N}(\mathbf{X}\boldsymbol{\beta},\mathbf{V})$, maximum likelihood estimation is based on maximizing

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

This leads to minimizing $S_{\mathbf{M}} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$ which is simply $\mathbf{S}_{\mathbf{W}}$ of (9) with $\mathbf{V}^{-1}$ in place of $\mathbf{W}$. Hence, by comparison with (10) the estimation equations are

$$\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\boldsymbol{\beta}^{0} = \mathbf{X}'\mathbf{V}^{-1}\mathbf{y}. \tag{21}$$

so that

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

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

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

so that 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 $\mathbf{t}'$ such that (i) $\mathbf{t}'\mathbf{y}$ is unbiased for $\boldsymbol{\lambda}'\mathbf{X}\boldsymbol{\beta}$ and (ii) from all unbiased estimators $\mathbf{t}'\mathbf{y}$ of $\boldsymbol{\lambda}'\mathbf{X}\boldsymbol{\beta}$ the one to be described as "best" is that which has minimum variance.

Requirement (i) demands having

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

Requirement (ii) demands choosing $\mathbf{t}$ to simultaneously satisfy (23) and minimize $v(\mathbf{t}'\mathbf{y}) = \mathbf{t}'\mathbf{V}\mathbf{t}$. Using Lagrange multipliers $2\boldsymbol{\theta}'$ this means minimizing

$$\mathcal{L} = \mathbf{t}'\mathbf{V}\mathbf{t} + 2\boldsymbol{\theta}'(\mathbf{X}'\boldsymbol{\lambda} - \mathbf{X}'\mathbf{t})$$

with respect to $\mathbf{t}$ and $\boldsymbol{\theta}$, which leads to equations

$$\mathbf{V}\mathbf{t} = \mathbf{X}\boldsymbol{\theta} \tag{24}$$

and

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

With $\mathbf{V}$ non-singular, equations (24) are consistent for solving for $\mathbf{t}$, the solution being

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

Substituting (26) into (25) gives

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

Then using $\boldsymbol{\theta}$ from (27) in (26) gives

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

Hence the BLUE of $\boldsymbol{\lambda}'\mathbf{X}\boldsymbol{\beta}$, being $\mathbf{t}'\mathbf{y}$ is

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

Therefore, on letting $\boldsymbol{\lambda}'$ be successive rows of $\mathbf{I}_N$, the BLUE of $\mathbf{X}\boldsymbol{\beta}$ is

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

Immediately it is seen that

$$\text{BLUE} = \text{WLSE}_{V^{-1}}$$

## 3. SINGULAR V

Estimators of $\mathbf{X}\boldsymbol{\beta}$ in the case of singular $\mathbf{V}$ are denoted by the symbols GLSE$^*$, MLE$^*$, and BLUE$^*$ when $\mathbf{V}^-$ is involved; special cases are GLSE$^+$ and MLE$^+$ when $\mathbf{V}^+$ is involved. BLUE is reserved for the most general form of best linear unbiased estimator, which requires explicit use of neither $\mathbf{V}^-$ nor $\mathbf{V}^+$, as in Section 3.4.

### 3.1 GLSE$^+$ and MLE$^+$, based on $\mathbf{V}^+$.

With $\mathbf{V}$ being singular but n.n.d., factor $\mathbf{V}$ as

$$\mathbf{V} = \mathbf{L'L} \text{ for } \mathbf{L} \text{ of full row rank, the rank of } \mathbf{V}.$$

Then defining

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

and

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

Then carrying out OLSE on $\mathbf{w}$ of (33) gives, similar to (10),

$$\mathbf{X'T'TX}\boldsymbol{\beta}^0 = \mathbf{X'T'w} = \mathbf{X'T'Ty}. \tag{34}$$

From (32) this is

$$\mathbf{X'V^+X}\boldsymbol{\beta}^0 = \mathbf{X'V^+y}$$

and so

$$\mathrm{GLSE}^+ = \mathbf{X}\boldsymbol{\beta}^0 = \mathbf{X}(\mathbf{X'V^+X})^-\mathbf{X'V^+y} = \mathrm{WLSE}_{\mathbf{V}^+}, \tag{35}$$

with variance

$$\mathrm{var}(\mathrm{GLSE}^+) = \mathbf{X}(\mathbf{X'V^+X})^-\mathbf{X'}. \tag{36}$$

The last equality in (35) is by comparison with (11); and finally, from Theorem 1, GLSE$^+$ is invariant to $(\mathbf{X'V^+X})^-$ if and only if $\mathbf{X} = \mathbf{CV^+X}$ for some $\mathbf{C}$.

For $\mathbf{y} \sim \mathcal{N}(\mathbf{X\beta}, \mathbf{V})$ as in Section 2.2, $\mathbf{w} = \mathbf{Ty} \sim \mathcal{N}(\mathbf{TX\beta}, \mathbf{I})$. Therefore, similar to (21), the estimation equations are $(\mathbf{TX})'\mathbf{I}^{-1}\mathbf{TX\beta}^0 = (\mathbf{TX})'\mathbf{I}^{-1}\mathbf{Ty}$, which are the same as (34) and so

$$\text{MLE}^{+} = \text{GLSE}^{+} = \text{WLSE}_{\mathbf{V}^{+}} = \mathbf{X}(\mathbf{X}'\mathbf{V}^{+}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{+}\mathbf{y}. \tag{37}$$

**3.2 Using Lagrange multipliers**

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 $\boldsymbol{\theta}$. 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 form of solution depends on whether or not $\mathbf{VV}^{-}\mathbf{X} = \mathbf{X}$, we consider two cases: when $\mathbf{VV}^{-}\mathbf{X}$ equals $\mathbf{X}$ and when it does not.

<u>3.2a BLUE$^*$, based on $\mathbf{VV}^{-}\mathbf{X} = \mathbf{X}$</u>

First, as in Lemma A4, when $\mathbf{VV}^{-}\mathbf{X} = \mathbf{X}$ is true for *some* $\mathbf{V}^{-}$, it is true for *all* $\mathbf{V}^{-}$. Second, when $\mathbf{VV}^{-}\mathbf{X} = \mathbf{X}$, equations $\mathbf{Vt} = \mathbf{X\theta}$ have solution $\mathbf{t} = \mathbf{V}^{-}\mathbf{X\theta}$ because then $\mathbf{Vt} = \mathbf{VV}^{-}\mathbf{X\theta} = \mathbf{X\theta}$. Therefore equations $\mathbf{Vt} = \mathbf{X\theta}$ are consistent for a solution for $\mathbf{t}$, its form being $\mathbf{t} = \mathbf{V}^{-}\mathbf{X\theta}$. This leads to what shall be called BLUE$^*$, in exactly the same way as (30) is derived:

$$\text{BLUE}^{*} = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-}\mathbf{y} = \text{WLSE}_{\mathbf{V}^{-}} \text{ when } \mathbf{VV}^{-}\mathbf{X} = \mathbf{X}. \tag{38}$$

Identifying BLUE$^*$ as $\text{WLSE}_{\mathbf{V}^{-}}$ comes from comparison with (11). Variance is

$$\text{var}(\text{BLUE}^{*}) = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-}\mathbf{X})^{-1}\mathbf{X}. \tag{39}$$

BLUE of (3) utilizes $\mathbf{V}^{-1}$ when $\mathbf{V}$ is non-singular. Comparison with (38) shows that BLUE$^*$, providing $\mathbf{VV}^{-}\mathbf{X} = \mathbf{X}$, is just BLUE with $\mathbf{V}^{-}$ in place of $\mathbf{V}^{-1}$. By analogy with (31) we could also label BLUE$^*$ as GLSE$^*$ and MLE$^*$.

Unbiasedness of $\text{BLUE}^*$ and invariance to the choice of $(\mathbf{X}'\mathbf{V}^-\mathbf{X})^-$ is assured by Theorem 1 because the condition $\mathbf{X} = \mathbf{CWX}$ of that theorem is satisfied with $\mathbf{W} = \mathbf{V}^-$ and $\mathbf{C} = \mathbf{V}$ when $\mathbf{VV}^-\mathbf{X} = \mathbf{X}$, which is when $\text{BLUE}^*$ exists. Furthermore $\text{BLUE}^*$ is invariant to the choice of $\mathbf{V}^-$ by Lemma A4, and so is equal to $\text{GLSE}^+$ of (35). Thus if $\mathbf{VV}^-\mathbf{X} = \mathbf{X}$, then

$$\begin{aligned}\text{BLUE}^* &= \mathbf{X}(\mathbf{X}'\mathbf{V}^-\mathbf{X})^-\mathbf{X}'\mathbf{V}^-\mathbf{y}\\ &= \text{WLSE}_{\mathbf{V}^-} = \text{GLSE}^* = \text{MLE}^*\\ &= \mathbf{X}(\mathbf{X}'\mathbf{V}^+\mathbf{X})^-\mathbf{X}'\mathbf{V}^+\mathbf{y}\\ &= \text{WLSE}_{\mathbf{V}^+} = \text{GLSE}^+ = \text{MLE}^+ .\end{aligned}$$

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

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

after using $\mathbf{LL}^- = \mathbf{I}$ that comes from $\mathbf{L}$ having full row rank (see Lemma A3). Thus, $\mathbf{VV}^-\mathbf{X} = \mathbf{X}$ is a necessary and sufficient condition for equations $\mathbf{Vt} = \mathbf{X\theta}$ to be consistent for a solution for $\mathbf{t}$, for any $\boldsymbol{\theta}$; and $\text{BLUE}^*$ follows. But there can be cases where, for *some* $\boldsymbol{\theta}$ (or $\boldsymbol{\theta}$s), the equations $\mathbf{Vt} = \mathbf{X\theta}$ have a solution for $\mathbf{t}$ without being consistent for *every* $\boldsymbol{\theta}$; i.e., there can be a solution $\mathbf{t} = \mathbf{V}^-\mathbf{X\theta}$ for that $\boldsymbol{\theta}$ (or $\boldsymbol{\theta}$s) without $\mathbf{VV}^-\mathbf{X} = \mathbf{X}$ being true. We illustrate this possibility in an example.

3.2b An example when $\mathbf{VV^-X} \neq \mathbf{X}$

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

$$\mathbf{V} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{X} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \quad . \tag{40}$$

Then $\mathbf{Vt} = \mathbf{X\theta}$ is

$$\begin{aligned} t_1 &= \theta \\ t_2 &= \theta \\ 0 &= \theta \end{aligned} \tag{41}$$

with solution $t_1 = t_2 = \theta = 0$: and $t_3$ remains unspecified. When $\mathbf{X't} = \mathbf{X'\lambda}$ is taken into account, it being $\Sigma\mathbf{t} = \Sigma\boldsymbol{\lambda}$ 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 (41), the solution to $\mathbf{Vt} = \mathbf{X\theta}$ and $\mathbf{X't} = \mathbf{X'\lambda}$ as

$$\mathbf{t}' = [0 \quad 0 \quad \Sigma\lambda_i] \quad \text{and} \quad \theta = 0. \tag{42}$$

Thus $\mathbf{t'y} = (\Sigma\lambda_i)y_3$ is the BLUE of $\boldsymbol{\lambda}'\mathbf{X}\beta = (\Sigma\lambda_i)\beta$, i.e. $\hat{\beta} = y_3$. Clearly, this is correct, since $y_3$ has variance zero as seen from var($\mathbf{y}$) = $\mathbf{V}$ of (40).

This solution has been derived without using any $\mathbf{V}^-$. It is therefore not $\mathbf{V^-X\theta}$ analogous to $\mathbf{V^{-1}X\theta}$ of (26) in the non-singular $\mathbf{V}$ case. Nevertheless, it *is* possible in this example (and in others) to find certain forms of generalized inverse of $\mathbf{V}$, to be denoted $\mathbf{V}^{\sim}$ such that $\mathbf{t}'$ is $\boldsymbol{\lambda}'\mathbf{X}(\mathbf{X'V^{\sim}X})^-\mathbf{X'V}^-$, analogous to (28). One possibility in the example is

$$\mathbf{V}^{\sim} = \begin{bmatrix} 1 & 0 & a \\ 0 & 1 & b \\ -1 & -1 & c \end{bmatrix} \tag{43}$$

for any $\alpha$, $\beta$ and $\gamma$ so long as $\alpha + \beta + \gamma \neq 1$. For then, for $a + b + c \neq 0$

$$\boldsymbol{\lambda}'\mathbf{X}(\mathbf{X}'\mathbf{V}^{\sim}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{\sim} = [\lambda_1\ \lambda_2\ \lambda_3] \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} (\alpha + \beta + \alpha)^{-1}\ [0\quad 0\quad \alpha + \beta + \gamma]$$

$$= [0\quad 0\quad \Sigma\lambda] = \mathbf{t}' \text{ of } (42).$$

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

$$\mathbf{V}\mathbf{V}^{\sim}\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & \alpha \\ 0 & 1 & \beta \\ -1 & -1 & \gamma \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 + \alpha \\ 1 + \beta \\ 0 \end{bmatrix} \neq \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \mathbf{X},$$

and hence $\mathbf{V}\mathbf{V}^{-}\mathbf{X} \neq \mathbf{X}$ for every $\mathbf{V}^{-}$. Thus although for this $\mathbf{V}^{\sim}$ we have $\mathbf{t}'$ of (42) equalling $\boldsymbol{\lambda}'\mathbf{X}(\mathbf{X}'\mathbf{V}^{\sim}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{\sim}$, the latter is not invariant to $\mathbf{V}^{-}$ in general, nor to $(\mathbf{X}'\mathbf{V}^{-}\mathbf{X})^{-}$.

### 3.2c P-BLUE* when $\mathbf{V}\mathbf{V}^{-}\mathbf{X} \neq \mathbf{X}$

A general solution for $\mathbf{t}$ and $\boldsymbol{\theta}$ to $\mathbf{V}\mathbf{t} = \mathbf{X}\boldsymbol{\theta}$ and $\mathbf{X}'\mathbf{t} = \mathbf{X}'\boldsymbol{\lambda}$ when $\mathbf{V}\mathbf{V}^{-}\mathbf{X} \neq \mathbf{X}$ can be derived as follows. Since $\mathbf{V}$ is singular, there is no loss of generality (save for the use of permutation matrices) in partitioning $\mathbf{V}$ as

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_1 \\ \mathbf{V}_2 \end{bmatrix} \text{ and similarly } \mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}\ , \tag{44}$$

where $\mathbf{V}_1$ has full row rank, the same rank as $\mathbf{V}$. Then

$$\mathbf{V}_2 = \mathbf{K}\mathbf{V}_1 \text{ for some } \mathbf{K}. \tag{45}$$

Now equations $\mathbf{Vt} = \mathbf{X}\boldsymbol{\theta}$ are, by (44), equivalent to the two equations $\mathbf{V}_1\mathbf{t} = \mathbf{X}_1\boldsymbol{\theta}$ and $\mathbf{V}_2\mathbf{t} = \mathbf{X}_2\boldsymbol{\theta}$. Multiplying the first of these by $\mathbf{K}$ and subtracting the second yields, by (45),

$$(\mathbf{K}\mathbf{V}_1 - \mathbf{V}_2)\mathbf{t} = \mathbf{0} = (\mathbf{K}\mathbf{X}_1 - \mathbf{X}_2)\boldsymbol{\theta}. \tag{46}$$

But the absence of consistency means that $\mathbf{K}\mathbf{X}_1 \neq \mathbf{X}_2$ and so (46) has a solution

$$\boldsymbol{\theta} = \mathbf{0}. \tag{47}$$

In many cases, a solution can be found for $\mathbf{t}$. Part of $\mathbf{Vt} = \mathbf{Xt}$ is $\mathbf{V}_1\mathbf{t} = \mathbf{X}_1\boldsymbol{\theta}$ which, with $\underset{\sim}{\theta} = 0$, is $\mathbf{V}_1\mathbf{t} = \mathbf{0}$ and this, since $\mathbf{V}_1$ has full row rank, has solution

$$\mathbf{t} = (\mathbf{I} - \mathbf{V}_1^{-}\mathbf{V}_1)\mathbf{w}$$

for any $\mathbf{w}$. But to satisfy $\mathbf{X}'\mathbf{t} = \mathbf{X}'\boldsymbol{\lambda}$ we must chose $\mathbf{w}$ to satisfy (25),i.e.,

$$\mathbf{X}'\mathbf{t} = \mathbf{X}'(\mathbf{I} - \mathbf{V}_1^{-}\mathbf{V}_1)\mathbf{w} = \mathbf{X}'\boldsymbol{\lambda}.$$

Hence, if these latter equations are consistent for a solution to $\mathbf{w}$,

$$\mathbf{t} = (\mathbf{I} - \mathbf{V}_1^{-}\mathbf{V}_1)[\mathbf{X}'(\mathbf{I} - \mathbf{V}_1^{-}\mathbf{V}_1)]^{-}\mathbf{X}'\boldsymbol{\lambda}. \tag{48}$$

For $\mathbf{V}_{11}$ non-singular, of rank the same as $\mathbf{V}$, consider partitioning $\mathbf{V}$ as

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_1 \\ \mathbf{V}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{V}_1 \\ \mathbf{K}\mathbf{V}_1 \end{bmatrix} = \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{11}\mathbf{K}' \\ \mathbf{K}\mathbf{V}_{11} & \mathbf{K}\mathbf{V}_{11}\mathbf{K}' \end{bmatrix} \tag{49}$$

with

$$\mathbf{V}^- = \begin{bmatrix} \mathbf{V}_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \quad \text{and} \quad \mathbf{V}_1^- = \begin{bmatrix} \mathbf{V}_{11}^{-1} \\ \mathbf{0} \end{bmatrix} \quad .$$

Hence

$$\mathbf{V}_1^-\mathbf{V}_1 = \begin{bmatrix} \mathbf{V}_{11}^{-1} \\ \mathbf{0} \end{bmatrix} [\mathbf{V}_{11} \quad \mathbf{V}_{11}\mathbf{K}'] = \begin{bmatrix} \mathbf{I} & \mathbf{K}' \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \quad .$$

Therefore (48) is

$$\mathbf{t} = \begin{bmatrix} \mathbf{0} & -\mathbf{K}' \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \left[ [\mathbf{X}_1' \quad \mathbf{X}_2'] \begin{pmatrix} \mathbf{0} & -\mathbf{K}' \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \right]^- \mathbf{X}'\boldsymbol{\lambda}$$

which reduces to

$$\mathbf{t} = \mathbf{B}'(\mathbf{X}'\mathbf{B}')^-\mathbf{X}'\boldsymbol{\lambda} \text{ for } \mathbf{B} = [-\mathbf{K} \quad \mathbf{I}]. \tag{50}$$

This and $\boldsymbol{\theta} = \mathbf{0}$ of (47) are therefore a solution to (24) and (25) when $\mathbf{Vt} = \mathbf{X}\boldsymbol{\theta}$ are not consistent for $\mathbf{t}$, provided $\mathbf{X}'(\mathbf{I} - \mathbf{V}_1^-\mathbf{V}_1)\mathbf{w} = \mathbf{X}'\boldsymbol{\lambda}$ is consistent for a solution in w. Then (50) gives what shall be called a pseudo-BLUE$^*$, namely

$$\text{P-BLUE}^* = \mathbf{X}(\mathbf{BX})^-\mathbf{By} \tag{51}$$

for

$$\mathbf{B} = [-\mathbf{K} \quad \mathbf{I}] \text{ with } \mathbf{K} = \mathbf{V}_2\mathbf{V}_1'(\mathbf{V}_1\mathbf{V}_1')^{-1}, \tag{52}$$

the latter coming from (45).

**Example.**

Applying (44) and (45) to (40) gives

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

Therefore (50) 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 (42).

**Verification of solution**

From (49),

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_{11} \\ \mathbf{KV}_{11} \end{bmatrix}[\mathbf{I} \quad \mathbf{K}']. \tag{53}$$

Therefore from (50) and (53), $\mathbf{Vt}$ involves the product

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

and so $\mathbf{Vt} = \mathbf{0} = \mathbf{X\theta}$ for $\boldsymbol{\theta} = \mathbf{0}$ of (47). Thus (24) is satisfied. For (25) recall that we are considering only the case when equations $\mathbf{X}'(\mathbf{I} - \mathbf{V}_1^-\mathbf{V}_1)\mathbf{w} = \mathbf{X}'\boldsymbol{\lambda}$ are consistent. And, as in the development of (50), $\mathbf{X}'\mathbf{B}' = \mathbf{X}'(\mathbf{I} - \mathbf{V}_1^-\mathbf{V}_1)$, so that $\mathbf{X}'\mathbf{Bw} = \mathbf{X}'\boldsymbol{\lambda}$ are being taken as consistent. But it is a standard result for consistent equations $\mathbf{Ax} = \mathbf{y}$ having solution $\mathbf{x} = \mathbf{A}^-\mathbf{y}$ that $\mathbf{AA}^-\mathbf{y} = \mathbf{y}$. Therefore for $\mathbf{t}$ of (50), $\mathbf{X}'\mathbf{t} = \mathbf{X}'\mathbf{B}'(\mathbf{X}'\mathbf{B}')^-\mathbf{X}'\boldsymbol{\lambda}' = \mathbf{X}'\boldsymbol{\lambda}'$. Thus (25) is satisfied. Hence (47) and (50) satisfy (24) and (25) when equations $\mathbf{Vt} = \mathbf{X\theta}$ are not consistent for $\mathbf{t}$, provided $\mathbf{X}'\mathbf{Bw} = \mathbf{X}'\boldsymbol{\lambda}$ is consistent for a solution in $\mathbf{w}$.

3.2d An apparent BLUE$^*$

The end of section 3.2b illustrates for the example the existence of a $\mathbf{V}^-$ such that $\mathbf{t}' = \boldsymbol{\lambda}'\mathbf{X}(\mathbf{BX})^-\mathbf{B}$ of (50) equals $\boldsymbol{\lambda}'\mathbf{X}(\mathbf{X}'\mathbf{V}^\sim\mathbf{X})^-\mathbf{X}'\mathbf{V}^\sim$, the form of the BLUE$^*$ in (38) using $\mathbf{V}^\sim$. But in this case $\mathbf{V}\mathbf{V}^\sim\mathbf{X} \neq \mathbf{X}$ and so the invariance properties of BLUE$^*$ do not apply. Nevertheless it is of interest to investigate when such a $\mathbf{V}^\sim$ exists and how it might be derived. Its existence is what prompts the name P-BLUE$^*$ in (51).

Begin with the equality

$$\mathbf{X}(\mathbf{BX})^-\mathbf{B} = \mathbf{X}(\mathbf{X}'\mathbf{V}^\sim\mathbf{X})^-\mathbf{X}'\mathbf{V}^\sim . \tag{54}$$

Post-multiplication by $\mathbf{X}$ yields

$$\mathbf{X}(\mathbf{BX})^-\mathbf{BX} = \mathbf{X}.$$

But this means that $\mathbf{BX}$ and $\mathbf{X}$ have the same rank. Therefore $r_{\mathbf{BX}} = r_{\mathbf{X}}$ is a necessary condition for (54) to be true.

Assuming that rank condition, pre-multiplying (54) by $\mathbf{X}'\mathbf{V}^\sim$ gives

$$\mathbf{X}'\mathbf{V}^\sim\mathbf{X}(\mathbf{BX})^-\mathbf{B} = \mathbf{X}'\mathbf{V}^\sim \tag{55}$$

Now, with $\mathbf{V}^-$ that follows (49),

$$\mathbf{V}^-\mathbf{V} = \begin{bmatrix} \mathbf{I} & \mathbf{K}' \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \quad \text{and} \quad \mathbf{V}\mathbf{V}^- = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{K} & \mathbf{0} \end{bmatrix} . \tag{56}$$

Therefore, since from (A7) the general form of ~~a~~ a generalized inverse of $\mathbf{V}$, $\mathbf{V}^g$ say, is

$$\mathbf{V}^g = \mathbf{V}^-\mathbf{V}\mathbf{V}^- + (\mathbf{I} - \mathbf{V}^-\mathbf{V})\mathbf{P} + \mathbf{Q}(\mathbf{I} - \mathbf{V}\mathbf{V}^-) \tag{57}$$

for any square $\mathbf{P}$ and $\mathbf{Q}$ of order N, using (56) in (57), together with $\mathbf{B}$ of (50) gives

$$\mathbf{V}^g = \mathbf{V}^- + [\mathbf{0} \quad \mathbf{B}']\mathbf{P} + \mathbf{Q}\begin{bmatrix} \mathbf{0} \\ \mathbf{B} \end{bmatrix} . \tag{58}$$

Now partition $\mathbf{P}$ and $\mathbf{Q}$ as $\mathbf{P} = [\mathbf{U}' \quad \mathbf{F}']'$ and $\mathbf{Q} = [\mathbf{R} \quad \mathbf{S}]$ with $\mathbf{F}'$ and $\mathbf{S}$ each of order $N \times (N-r)$. Then (58) is

$$\mathbf{V}^g = \mathbf{V}^- + \mathbf{B}'\mathbf{F} + \mathbf{S}\mathbf{B}. \tag{59}$$

Using this for $\mathbf{V}^{\sim}$ in (55) gives

$$\mathbf{X}'(\mathbf{V}^- + \mathbf{B}'\mathbf{F} + \mathbf{S}\mathbf{B})\mathbf{X}(\mathbf{B}\mathbf{X})^-\mathbf{B} = \mathbf{X}'(\mathbf{V}^- + \mathbf{B}'\mathbf{F} + \mathbf{S}\mathbf{B}). \tag{60}$$

Therefore when any $\mathbf{F}$ and $\mathbf{S}$ that satisfy (60) are used in $\mathbf{V}^g$ of (59), that $\mathbf{V}^g$ used as $\mathbf{V}^{\sim}$ in $\boldsymbol{\lambda}'\mathbf{X}(\mathbf{X}'\mathbf{V}^{\sim}\mathbf{X})^-\mathbf{X}'\mathbf{V}^{\sim}$ yields $\mathbf{t}' = \boldsymbol{\lambda}'\mathbf{X}(\mathbf{B}\mathbf{X})^-\mathbf{B}$ of (50). Then

$$\text{P-BLUE}^* = \mathbf{X}(\mathbf{B}\mathbf{X})^-\mathbf{B}\mathbf{y} = \mathbf{X}(\mathbf{X}'\mathbf{V}^g\mathbf{X})^-\mathbf{X}'\mathbf{V}^g\mathbf{y} \tag{61}$$

for $\mathbf{V}^g$ of (59). But (61) is <u>not</u> a BLUE$^*$ because $\mathbf{V}\mathbf{V}^g\mathbf{X} \neq \mathbf{X}$ for $\mathbf{V}^g$ of (59).

**Example** (continued)

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

Since $N = 3$ and $r = 2$, $\mathbf{F}'$ and $\mathbf{S}$ have the form $\mathbf{F} = [f_1 \quad f_2 \; f_3]$ and $\mathbf{S}' = [s_1 \quad s_2 \quad s_3]$. Using these in (60) gives

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

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

This is

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

which is

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

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 (57)

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

$$= \begin{bmatrix} 1 & 0 & s_1 \\ 0 & 1 & s_2 \\ -1 & -1 & 2+s_3 \end{bmatrix},$$

and for $s_1 = a$, $s_2 = b$ and $s_3 = \gamma - 2$ this is $\mathbf{V}^-$ of (43).

Special Case. In the example, $\mathbf{BX}$ is scalar; whenever this is so, or when $\mathbf{BX}$ is non-singular, the terms in $\mathbf{S}$ in (60) cancel - as they do when, even more generally, $\mathbf{BX}(\mathbf{BX})^-\mathbf{B} = \mathbf{B}$. And then (60) reduces to

$$\mathbf{X}'(\mathbf{V}^- + \mathbf{B}'\mathbf{F})\mathbf{X}(\mathbf{BX})^-\mathbf{B} = \mathbf{X}'(\mathbf{V}^- + \mathbf{B}'\mathbf{F}) \qquad (62)$$

and if any $\mathbf{F}$ that satisfies (62) is used in (59) with $\mathbf{S} = \mathbf{0}$, i.e., in

$$\mathbf{V}^g = \mathbf{V}^- + \mathbf{B}'\mathbf{F}, \tag{63}$$

then that $\mathbf{V}^g$ used in (61) yields $\mathbf{t}' = \boldsymbol{\lambda}'\mathbf{X}(\mathbf{BX})^-\mathbf{B}$ of (50), but $\mathbf{VV}^g\mathbf{X} \neq \mathbf{X}$.

### 3.3 BLUE - the general case

Derivation of BLUE* and P-BLUE* in (38) and (51) respectively is based on the same methodology as used for deriving BLUE in the case of non-singular $\mathbf{V}$, 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 $(\mathbf{I} - \mathbf{M})\mathbf{y}$ as the OLSE of $\mathbf{X}\boldsymbol{\beta}$ from (7) gives $\boldsymbol{\lambda}'(\mathbf{I} - \mathbf{M})\mathbf{y}$ as the OLSE of $\boldsymbol{\lambda}'\mathbf{X}\boldsymbol{\beta}$. We now ascertain what linear function of the OLSE residuals $\mathbf{My}$ should be added to $\boldsymbol{\lambda}'(\mathbf{I} - \mathbf{M})\mathbf{y}$ to yield the BLUE of $\boldsymbol{\lambda}'\mathbf{X}\boldsymbol{\beta}$ in the form

$$\boldsymbol{\lambda}'\mathbf{X}\boldsymbol{\beta}^0 = \boldsymbol{\lambda}'(\mathbf{I} - \mathbf{M})\mathbf{y} + \boldsymbol{\mu}'\mathbf{My} \tag{64}$$

for some vector $\boldsymbol{\mu}$. Since $E[(\mathbf{I} - \mathbf{M})\mathbf{y}] = \mathbf{X}\boldsymbol{\beta}$ and $E(\mathbf{My}) = \mathbf{0}$, no linear combination of $(\mathbf{I} - \mathbf{M})\mathbf{y}$ and $\mathbf{My}$ can be unbiased for $\boldsymbol{\lambda}'\mathbf{X}\boldsymbol{\beta}$ unless the term in $(\mathbf{I} - \mathbf{M})\mathbf{y}$ is $\boldsymbol{\lambda}'(\mathbf{I} - \mathbf{M})\mathbf{y}$, as in (64). We therefore obtain the BLUE by choosing $\boldsymbol{\mu}$ to minimize the variance of (64), which is

$$\text{var}(\boldsymbol{\lambda}'\mathbf{X}\boldsymbol{\beta}^\circ) = \boldsymbol{\lambda}'(\mathbf{I} - \mathbf{M})\mathbf{V}(\mathbf{I} - \mathbf{M})\boldsymbol{\lambda} + 2\boldsymbol{\lambda}'(\mathbf{I} - \mathbf{M})\mathbf{VM}\boldsymbol{\mu} + \boldsymbol{\mu}'\mathbf{MVM}\boldsymbol{\mu}. \tag{65}$$

Differentiating v with respect to elements of $\boldsymbol{\mu}$ and equating the result to $\mathbf{0}$ gives

$$\mathbf{MV}(\mathbf{I} - \mathbf{M})\boldsymbol{\lambda} + \mathbf{MVM}\boldsymbol{\mu} = \mathbf{0}.$$

Hence $\boldsymbol{\mu} = -(\mathbf{MVM})^{-}\mathbf{MV}(\mathbf{I} - \mathbf{M})\boldsymbol{\lambda}$ and so letting $\boldsymbol{\lambda}'$ be successive rows of $\mathbf{I}$ gives, from (64)

$$\text{BLUE} = (\mathbf{I} - \mathbf{M})[\mathbf{I} - \mathbf{VM}(\mathbf{MVM})^{-}\mathbf{M}]\mathbf{y}. \tag{66}$$

But, from Lemma A5, $\mathbf{My} = \mathbf{MVM}(\mathbf{MVM})^{-}\mathbf{My}$ and so

$$\text{BLUE} = \mathbf{y} - \mathbf{VM}(\mathbf{MVM})^{-}\mathbf{My}. \tag{67}$$

Pukelsheim (1974), using $(\mathbf{MVM})^{+}$ in place of $(\mathbf{MVM})^{-}$, gives (66), which is equivalent to a form given in Albert (1973); (67) is new.

Invariance of the BLUE to $(\mathbf{MVM})^{-}$ is easily established from (67). For $(\mathbf{MVM})^{\sim}$ being a generalized inverse of $\mathbf{MVM}$ different from $(\mathbf{MVM})^{-}$,

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

and because, with $\mathbf{V}$ being n.n.d. and $\mathbf{M}$ being symmetric, $\mathbf{VM}(\mathbf{MVM})^{-}\mathbf{VM} = \mathbf{VM}$, and so,

$$\text{BLUE} = \mathbf{y} - \mathbf{VM}(\mathbf{MVM})^{\sim}\mathbf{My};$$

i.e., BLUE is invariant to $(\mathbf{MVM})^{-}$. Similar arguments yield

$$\text{var(BLUE)} = \mathbf{V} - \mathbf{VM}(\mathbf{MVM})^{-}\mathbf{MV}$$

and, with $\mathbf{MX} = \mathbf{0}$,

$$\text{E(BLUE)} = \mathbf{X}\boldsymbol{\beta} - \mathbf{VM}(\mathbf{MVM})^{-}\mathbf{MX}\boldsymbol{\beta} = \mathbf{X}\boldsymbol{\beta}.$$

## 4. SUMMARY

### 4.1 For $\mathbf{V} = \sigma^2\mathbf{I}$

From (7), (11), (19), (22) and (30), when $\mathbf{V} = \sigma^2\mathbf{I}$ we have

$$\text{OLSE} = (\mathbf{I} - \mathbf{M})\mathbf{y} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y} = \mathbf{XX}^{+}\mathbf{y} = \text{WLSE}_{\sigma^2\mathbf{I}} = \text{GLSE} = \text{MLE} = \text{BLUE} \tag{68}$$

**4.2 For V non-singular**

From (11), (19), (22) and (30), when $\mathbf{V}^{-1}$ exists,

$$\mathrm{WLSE}_{\mathbf{V}^{-1}} = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} = \mathrm{GLSE} = \mathrm{MLE} = \mathrm{BLUE}. \tag{69}$$

**4.3 For V singular**

From (11), (35) and (37), when V is singular,

$$\mathrm{WLSE}_{\mathbf{V}^{+}} = \mathbf{X}(\mathbf{X}'\mathbf{V}^{+}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{+}\mathbf{y} = \mathrm{GLSE}^{+} = \mathrm{MLE}^{+}; \tag{70}$$

$\mathrm{BLUE}^{*}$ exists if and only if $\mathbf{VV}^{-}\mathbf{X} = \mathbf{X}$ and from (38)

$$\mathrm{BLUE}^{*} = \mathrm{WLSE}_{\mathbf{V}^{-}} = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-}\mathbf{y}. \tag{71}$$

P-BLUE* exists when $\mathbf{VV}^{-}\mathbf{X} \neq \mathbf{X}$ and is, from (51) and (52), providing $\mathbf{X}'(\mathbf{I} - \mathbf{V}_1^{-}\mathbf{V}_1)$ has the same rank as $\mathbf{X}'$,

$$\text{P-BLUE*} = \mathbf{X}(\mathbf{BX})^{-}\mathbf{By} \text{ for } \mathbf{B} = [-\mathbf{V}_2\mathbf{V}_1'(\mathbf{V}_1\mathbf{V}_1')^{-1} \quad \mathbf{I}]$$

where $\mathbf{V}_1$ is of full row rank, the rank of $\mathbf{V}$. Further, providing that $\mathbf{BX}$ and $\mathbf{X}$ have equal rank, and providing also, that $\mathbf{F}$ and $\mathbf{S}$ can be found that satisfy (60), then using $\mathbf{F}$ and $\mathbf{S}$ in $\mathbf{V}^{g} = \mathbf{V}^{-} + \mathbf{B}'\mathbf{F} + \mathbf{SB}$ of (59) permits one to write $\text{P-BLUE}^{*}$ as $\mathbf{X}(\mathbf{X}'\mathbf{V}^{-}\mathbf{X})^{g}\mathbf{X}'\mathbf{V}^{g}\mathbf{y}$ as in (61), giving the appearance of a $\mathrm{BLUE}^{*}$; but $\text{P-BLUE}^{*}$ is not $\mathrm{BLUE}^{*}$ because $\mathbf{VV}^{g}\mathbf{X} \neq \mathbf{X}$.

The BLUE for singular $\mathbf{V}$ is, from (67),

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

## 5. EQUALITY OF BLUE (FOR SINGULAR V) TO OTHER ESTIMATORS

We have four theorems connecting BLUE for singular $\mathbf{V}$ to OLSE, to BLUE$^*$, to GLSE$^+$ and to a calculation using $\mathbf{V}_* = \mathbf{V} + \mathbf{XX}'$.

**Theorem 2.** BLUE = OLSE = $\mathbf{XX}^+\mathbf{y}$ if and only if $\mathbf{VX} = \mathbf{XQ}$ for some $\mathbf{Q}$.

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

**Proof.** If $\mathbf{VX} = \mathbf{XQ}$

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

Therefore

$$\begin{aligned}\text{BLUE} &= \mathbf{y} - \mathbf{VM}(\mathbf{MVM})^-\mathbf{My} = \mathbf{y} - \mathbf{MVM}(\mathbf{MVM})^-\mathbf{My} \text{ because } \mathbf{VM} = \mathbf{MVM},\\ &= \mathbf{y} - \mathbf{My} \text{ from Lemma A5}\\ &= \mathbf{XX}^+\mathbf{y} = \text{OLSE}.\end{aligned}$$

Thus is sufficiency established. Proving necessity begins with BLUE = OLSE:

$$\mathbf{y} - \mathbf{VM}(\mathbf{MVM})^-\mathbf{My} = \mathbf{y} - \mathbf{My} \;\forall\; \mathbf{y}.$$

Therefore $[\mathbf{VM}(\mathbf{MVM})^-\mathbf{M} - \mathbf{M}]\mathbf{y} = \mathbf{0} \;\forall\; \mathbf{y}$. Hence, by Lemma A6, $[\mathbf{VM}(\mathbf{MVM})^-\mathbf{M} - \mathbf{M}]\mathbf{V} = \mathbf{0}$. Therefore $\mathbf{VM}(\mathbf{MVM})^-\mathbf{MV} = \mathbf{MV} = \mathbf{VM}$, the last equality arising from symmetry of $\mathbf{VM}(\mathbf{MVM})^-\mathbf{MV}$, based on (A4). Hence, with $\mathbf{MV} = \mathbf{VM}$, and $\mathbf{M}$ being $\mathbf{I} - \mathbf{XX}^+$,

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

so that

$$\mathbf{XX}^+\mathbf{VX} = \mathbf{VXX}^+\mathbf{X} = \mathbf{VX};$$

i.e.,

$$\mathbf{VX} = \mathbf{XX}^{+}\mathbf{VX} = \mathbf{XQ} \text{ for } \mathbf{Q} = \mathbf{X}^{+}\mathbf{VX}.$$

Q. E. D.

**Theorem 3.** BLUE = GLSE$^{+}$ = BLUE$^{*}$ = $\mathbf{X}(\mathbf{X}'\mathbf{V}^{-}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-}\mathbf{y}$ if and only if $\mathbf{VV}^{-}\mathbf{X} = \mathbf{X}$.

The sufficiency part of this theorem is due to Rao and Mitra (1971) and the necessity part to Pukelsheim (1974). We offer new proofs that are shorter than theirs.

**Proof.** As motivation, we first show that when $\mathbf{V}^{-1}$ exists (for which $\mathbf{VV}^{-}\mathbf{X} = \mathbf{VV}^{-1}\mathbf{X} = \mathbf{X}$) then BLUE = GLSE. This is so because by invariance to $(\mathbf{MVM})^{-}$,

$$\text{BLUE} = \mathbf{y} - \mathbf{VM}(\mathbf{MVM})^{-}\mathbf{My} = \mathbf{y} - \mathbf{VM}(\mathbf{MVM})^{+}\mathbf{My} = \mathbf{y} - \mathbf{VMPMy}$$

for $\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}$ of Lemma A7. Therefore from the proof of that lemma yields $\mathbf{P} = \mathbf{PM} = \mathbf{MP} = \mathbf{MPM}$ and so

$$\begin{aligned} \text{BLUE} &= \mathbf{y} - \mathbf{VPy} = \mathbf{y} - [\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}]\mathbf{y} \\ &= \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} = \text{GLSE of (19).} \end{aligned}$$

Now the general theorem: first, to prove sufficiency of $\mathbf{VV}^{-}\mathbf{X} = \mathbf{X}$, define

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

and observe that $\mathbf{QX} = \mathbf{0}$ and hence

$$\begin{aligned} \mathbf{QM} &= \mathbf{Q}(\mathbf{I} - \mathbf{XX}^{+}) = \mathbf{Q} = \mathbf{MQ}, \text{ using the symmetry of } \mathbf{Q} \text{ and } \mathbf{M}, \\ &= \mathbf{MQM} \text{ because } \mathbf{MQM} = \mathbf{M}^{2}\mathbf{Q} = \mathbf{MQ}. \end{aligned}$$

Hence

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

Therefore, if $\mathbf{VV}^{-}\mathbf{X} = \mathbf{X}$

$$\mathbf{MVMQMVM} = \mathbf{MVM} - \mathbf{MX}(\mathbf{X}'\mathbf{V}^-\mathbf{X})^-\mathbf{X}'\mathbf{M} = \mathbf{MVM};$$

i.e., $\mathbf{Q}$ is a generalized inverse of $\mathbf{MVM}$. Hence

$$\begin{aligned}
\text{BLUE} &= \mathbf{y} - \mathbf{VM}(\mathbf{MVM})^-\mathbf{My} \\
&= \mathbf{y} - \mathbf{VMQMy}, \text{ from } (\mathbf{MVM})^- = \mathbf{Q} \\
&= \mathbf{y} - \mathbf{VQy}, \text{ because } \mathbf{Q} = \mathbf{MQM} \\
&= \mathbf{y} - [\mathbf{VV}^-\mathbf{y} - \mathbf{VV}^-\mathbf{X}(\mathbf{X}'\mathbf{V}^-\mathbf{X})^-\mathbf{X}'\mathbf{V}^-\mathbf{y}] \\
&= \mathbf{y} - \mathbf{y} + \mathbf{X}(\mathbf{X}'\mathbf{V}^-\mathbf{X})^-\mathbf{X}'\mathbf{V}^-\mathbf{y}, \text{ from Lemma A5,} \\
&= \mathbf{X}(\mathbf{X}'\mathbf{V}^-\mathbf{X})^-\mathbf{X}'\mathbf{V}^-\mathbf{y} \\
&= \text{BLUE}^* \text{ of (38).}
\end{aligned}$$

Thus is sufficiency established, since $\mathbf{VV}^-\mathbf{X} = \mathbf{X}$ also implies $\text{BLUE}^* = \text{GLSE}^* = \text{GLSE}^+$.

The necessity of $\mathbf{VV}^-\mathbf{X} = \mathbf{X}$ depends only on the equality of $\text{GLSE}^+$ and BLUE because, once necessity is proven the equality of $\text{GLSE}^+ = \mathbf{X}(\mathbf{X}'\mathbf{V}^+\mathbf{X})^-\mathbf{X}'\mathbf{V}^+\mathbf{y}$ to $\text{GLSE}^* = \mathbf{X}(\mathbf{X}'\mathbf{V}^-\mathbf{X})^-\mathbf{X}'\mathbf{V}^-\mathbf{y}$ is immediate. Therefore we start with

$$\mathbf{X}(\mathbf{X}'\mathbf{V}^+\mathbf{X})^-\mathbf{X}'\mathbf{V}^+\mathbf{y} = \mathbf{y} - \mathbf{VM}(\mathbf{MVM})^-\mathbf{My} \;\forall\; \mathbf{y}. \tag{72}$$

For convenience define

$$\mathbf{T} = \mathbf{T}' = \mathbf{M}(\mathbf{MVM})^-\mathbf{M}$$

and then, on applying Lemma A6 to (72), get

$$\mathbf{X}(\mathbf{X}'\mathbf{V}^+\mathbf{X})^-\mathbf{X}'\mathbf{V}^+\mathbf{V} = \mathbf{V} - \mathbf{VTV}. \tag{73}$$

Pre-multiplying (73) by $\mathbf{X}'\mathbf{V}^+$ gives

$$\mathbf{X}'\mathbf{V}^+\mathbf{X}(\mathbf{X}'\mathbf{V}^+\mathbf{X})^-\mathbf{X}'\mathbf{V}^+\mathbf{V} = \mathbf{X}'\mathbf{V}^+\mathbf{V} - \mathbf{X}'\mathbf{V}^+\mathbf{VTV}$$

which is

$$\mathbf{X}'\mathbf{V}^+\mathbf{V} = \mathbf{X}'\mathbf{V}^+\mathbf{V} - \mathbf{X}'\mathbf{V}^+\mathbf{VTV}.$$

Therefore

$$\mathbf{X'V^+VTV = 0} \text{ and so } \mathbf{VTVV^+X = 0}. \tag{74}$$

Post-multiplying (73) by $\mathbf{V^+X}$,

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

and using (74) gives

$$\mathbf{X(X'V^+X)^-X'V^+X = VV^+X}. \tag{75}$$

But

$$\mathbf{X(X'V^+X)^-X'V^+X = X}. \tag{76}$$

Thus (75) and (76) give $\mathbf{VV^+X = X}$, and so from Lemma A4, $\mathbf{VV^-X = X}$.

**Q.E.D.**

**Theorem 4.** $\text{GLSE}^+ = \text{OLSE}$ if and only if $\text{GLSE}^+ = \text{BLUE}$ and OLSE = BLUE; in which case $\mathbf{VV^-X = X}$ and $\mathbf{VX = XQ}$ for some $\mathbf{Q}$.

**Proof .** Sufficiency is obvious: if $\text{GLSE}^+$ and OLSE each equal BLUE then they equal each other. Proving necessity starts with $\text{GLSE}^+ = \text{OLSE}$, i.e.,

$$\mathbf{X(X'V^+X)^-X'V^+y = X(X'X)^-X'y}. \tag{77}$$

Equating the expected values of both sides gives $\mathbf{X(X'V^+X)^-X'V^+X\beta = X\beta}\ \forall\ \boldsymbol{\beta}$, and so

$$\mathbf{X(X'V^+X)^-X'V^+X = X}. \tag{78}$$

Applying Lemma A6 to (77) gives

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

and pre-multiplying this by $\mathbf{X}'$ and post-multiplying it by $\mathbf{V}^{+}\mathbf{X}$ gives

$$\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{V}^{+}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{+}\mathbf{V}\mathbf{V}^{+}\mathbf{X} = \mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{V}\mathbf{V}^{+}\mathbf{X}, \tag{79}$$

Then using (78) gives (79) as $\mathbf{X}'\mathbf{X} = \mathbf{X}'\mathbf{V}\mathbf{V}^{+}\mathbf{X}$. Therefore

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

and so using (A9) gives $\mathbf{X} = \mathbf{V}\mathbf{V}^{+}\mathbf{X}$, and thus $\mathbf{X} = \mathbf{V}\mathbf{V}^{-}\mathbf{X}$. Then OLSE = GLSE$^{+}$ = BLUE$^{*}$ = BLUE.

**Q.E.D.**

**Theorem 5. BLUE = $\mathbf{X}(\mathbf{X}'\mathbf{V}_{*}^{-}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}_{*}^{-}\mathbf{y}$ for $\mathbf{V}_{*} = \mathbf{V} + \mathbf{X}\mathbf{X}'$.**

Using $\mathbf{V}_{*}$ is the form of the BLUE given by Rao and Mitra (1971).

**Proof.** Theorem 3 shows that BLUE = WLSE$_{\mathbf{V}^{-}}$ if and only if $\mathbf{V}\mathbf{V}^{-}\mathbf{X} = \mathbf{X}$; i.e.

$$\mathbf{y} - \mathbf{V}\mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}\mathbf{y} = \text{WLSE}_{\mathbf{V}^{-}} \text{ if } \mathbf{V}\mathbf{V}^{-}\mathbf{X} = \mathbf{X}.$$

But for $\mathbf{V}_{*} = \mathbf{V} + \mathbf{X}\mathbf{X}'$, Lemma A8 gives $\mathbf{V}_{*}\mathbf{V}_{*}^{-}\mathbf{X} = \mathbf{X}$ and so

$$\mathbf{y} - \mathbf{V}_{*}\mathbf{M}(\mathbf{M}\mathbf{V}_{*}\mathbf{M})^{-}\mathbf{M}\mathbf{y} = \mathbf{X}(\mathbf{X}'\mathbf{V}_{*}^{-}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}_{*}^{-}\mathbf{y}. \tag{80}$$

But $\mathbf{V}_{*}\mathbf{M} = (\mathbf{V} + \mathbf{X}\mathbf{X}')\mathbf{M} = \mathbf{V}\mathbf{M}$; therefore (80) is

$$\mathbf{y} - \mathbf{V}\mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}\mathbf{y} = \mathbf{X}(\mathbf{X}'\mathbf{V}_{*}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}_{*}^{-}\mathbf{y}, \tag{80}$$

which is BLUE = $\mathbf{X}(\mathbf{X}'\mathbf{V}_{*}^{-}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}_{*}^{-}\mathbf{y}$.

**Q.E.D.**

Proof of this theorem is also given by Baksalary and Kala (1978) using projection operators.

# 6. APPENDIX

## 6.1 Generalized inverses

The Moore-Penrose inverse of any non-null matrix $\mathbf{A}$ is the unique matrix $\mathbf{A}^+$satisfying

$$\mathbf{AA}^+\mathbf{A} = \mathbf{A},\ \mathbf{A}^+\mathbf{AA}^+ = \mathbf{A} \text{ and having both } \mathbf{AA}^+ \text{ and } \mathbf{A}^+\mathbf{A} \text{ symmetric.} \tag{A1}$$

A generalized inverse, $\mathbf{A}^-$, satisfies just the first condition in (A1):

$$\mathbf{AA}^-\mathbf{A} = \mathbf{A}. \tag{A2}$$

The following properties (e.g., Searle, 1982, p. 216) of a generalized inverse $(\mathbf{X'X})^-$ of $\mathbf{X'X}$ a (A2)

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

$$(\mathbf{X'X})^{-\prime} \text{ is a generalized inverse of } \mathbf{X'X}; \tag{A3}$$

$$\mathbf{XX}^+ = \mathbf{X}(\mathbf{X'X})^-\mathbf{X'} \text{ is symmetric and invariant to } (\mathbf{X'X})^- \tag{A4}$$

$$\mathbf{X}(\mathbf{X'X})^-\mathbf{X'X} = \mathbf{X} \text{ is invariant to } (\mathbf{X'X})^-; \tag{A5}$$

$$\mathbf{M} = \mathbf{I} - \mathbf{XX}^+ = \mathbf{M'} = \mathbf{M}^2 \text{ with } \mathbf{MX} = \mathbf{0}. \tag{A6}$$

**Lemma A1.** For $\mathbf{A}^-$ being a generalized inverse of $\mathbf{A}$, so is

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

for any $\mathbf{P}$ and $\mathbf{Q}$. Furthermore, for $\mathbf{A}^-$ being some particular generalized inverse of $\mathbf{A}$, say $\mathbf{A}^*$, a suitable $\mathbf{P}$ and $\mathbf{Q}$ can be found for (A7); e.g., $\mathbf{P} = \mathbf{A}^*$ and $\mathbf{Q} = \mathbf{A}^-\mathbf{AA}^*$.

**Proof.** The product $\mathbf{AA}^{\sim}\mathbf{A}$ easily is always $\mathbf{A}$; and $\mathbf{P} = \mathbf{A}^*$ and $\mathbf{Q} = \mathbf{A}^-\mathbf{AA}^*$ reduce the right-hand side of (A7) to $\mathbf{A}^*$.

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

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

$$\mathbf{W} = \mathbf{H}'\mathbf{H},\ (\mathbf{H}'\mathbf{H})^{-1} \text{ exists and } \mathbf{W}^{+} = \mathbf{H}'(\mathbf{H}\mathbf{H}')^{-2}\mathbf{H}. \tag{A8}$$

Verification of $\mathbf{W}^{+}$ is easily established from (A1).

If, for any real matrix $\mathbf{B}$,

$$\mathbf{B}\mathbf{B}' = \mathbf{0} \text{ then } \mathbf{B} = \mathbf{0}. \tag{A9}$$

### 6.3 Other pertinent results

**Lemma A2.** If $\mathbf{FSG} = \mathbf{0}$ for all $\mathbf{S}$ then either $\mathbf{F} = \mathbf{0}$ or $\mathbf{G} = \mathbf{0}$.

**Proof.** If $\mathbf{G} \neq \mathbf{0}$, there exists a vector $\boldsymbol{\tau} = \mathbf{Gu} \neq \mathbf{0}$. Then

$$\mathbf{FSG} = \mathbf{0} \Rightarrow \mathbf{FSGu} = \mathbf{0} \Rightarrow \mathbf{FS}\boldsymbol{\tau} = \mathbf{0}\ \forall\ \mathbf{S}.$$

For any non-null vector $\boldsymbol{\nu}$ take $\mathbf{S} = \boldsymbol{\nu}\boldsymbol{\tau}'/\boldsymbol{\tau}'\boldsymbol{\tau}$. Then

$$\mathbf{FS}\boldsymbol{\tau} = \mathbf{0} \Rightarrow \mathbf{F}\boldsymbol{\nu}\boldsymbol{\tau}'\boldsymbol{\tau}/\boldsymbol{\tau}'\boldsymbol{\tau} = \mathbf{0} \Rightarrow \mathbf{F}\boldsymbol{\nu} = \mathbf{0}\ \forall\ \boldsymbol{\tau} \neq \mathbf{0}, \text{ and so } \mathbf{F} = \mathbf{0}.$$

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

**Q.E.D.**

**Lemma 3.** For $\mathbf{R}$ of full row rank, $\mathbf{R}\mathbf{R}^{-} = \mathbf{I}$ for all $\mathbf{R}^{-}$.

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

**Q.E.D.**

**Lemma A4.** If $\mathbf{VV^-X = X}$ then for $\mathbf{y} \sim (\mathbf{X\beta}, \mathbf{V})$

(i) $\mathbf{VV^-X = X}$ for every $\mathbf{V^-}$, (ii) $\mathbf{X'V^-X}$ is invariant to $\mathbf{V^-}$;

and for almost all $\mathbf{y}$

(iii) $\mathbf{VV^-y = y}$ for every $\mathbf{V^-}$, (iv) $\mathbf{X'V^-y}$ is invariant to $\mathbf{V^-}$,

and

(v) $\mathbf{X(X'V^-X)^-X'V^-y}$ is invariant to $\mathbf{V^-}$ and to $\mathbf{(X'V^-X)^-}$; and is unbiased.

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

$$\text{(i) } \mathbf{X = VV^-X = VV^\sim VV^-X = VV^\sim X}.$$

$$\text{(ii) } \mathbf{X'V^\sim X = X'V^\sim VV^-X = (V'V^{\sim\prime}X)'V^-X = (VV^-X)'V^-X = X'V^-X}.$$

This derivation turns on the symmetry of $\mathbf{V}$, and thus $\mathbf{V^{\sim\prime}}$ is a generalized inverse of $\mathbf{V}$ and so $\mathbf{X = VV^-X}$, invariant to $\mathbf{V^-}$, as in (i), implies $\mathbf{X = VV^\sim X}$ $= \mathbf{VV^{\sim\prime}X = V'\tilde{V}'X}$.

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

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

(iv) Proof is the same as (ii), with its final $\mathbf{X}$ replaced by $\mathbf{y}$.

(v) Theorem 1 applies, because $\mathbf{X = CWX}$ is $\mathbf{X = VV^-X}$.

**Q.E.D.**

**Lemma A5.** $\mathbf{My = MVM(MVM)^-My}$.

**Proof.** Define $\mathbf{Y} = \text{var}(\mathbf{My}) = \mathbf{MVM = E[My(My)']}$, because $\mathbf{MX = 0}$. Then

$$\mathbf{0 = (I - YY^-)MVM(I - YY^-)' = E(ww')} \text{ for } \mathbf{w = (I - YY^-)My}.$$

Therefore, with probability unity, $\mathbf{E(ww')} = \mathbf{0}$ implies $\mathbf{w} = \mathbf{0}$, i.e., $(\mathbf{I} - \mathbf{YY}^-)\mathbf{My} = \mathbf{0}$ and so $\mathbf{My} = \mathbf{MVM}(\mathbf{MVM})^-\mathbf{My}$.

**Q.E.D.**

**Lemma A6.** If $\mathbf{Ky} = \mathbf{0}$ with probability unity, then $\mathbf{KV} = \mathbf{0}$.

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

**Q.E.D.**

**Lemma A7.** For $\mathbf{V}$ non-singular, $(\mathbf{MVM})^+ = \mathbf{P}$ for $\mathbf{P} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X'V}^{-1}\mathbf{X})^-\mathbf{X'V}^{-1}$.

**Proof.** $\mathbf{PX} = \mathbf{0}$, using (A5); hence $\mathbf{PVM} = \mathbf{M}$ and so the symmetry of $\mathbf{M}$, $\mathbf{V}$ and $\mathbf{P}$ gives $\mathbf{PVM} = \mathbf{M} = \mathbf{MVP}$. And $\mathbf{PM} = \mathbf{P}(\mathbf{I} - \mathbf{XX}^+) = \mathbf{P}$, because $\mathbf{PX} = \mathbf{0}$, with $\mathbf{PM} = \mathbf{MP}$, because $\mathbf{P}$ and $\mathbf{M}$ are symmetric, and $\mathbf{MP} = \mathbf{MPM}$, because $\mathbf{P} = \mathbf{PM}$. With these properties for $\mathbf{P}$ and $\mathbf{M}$ it is easily shown that $\mathbf{P}$ satisfies the four conditions (A1) that define $(\mathbf{MVM})^+$.

**Q.E.D.**

**Lemma A8.** With $\mathbf{V}$ n.n.d., and $\mathbf{V}_* = \mathbf{V} + \mathbf{XX'}$, then $\mathbf{V}_*\mathbf{V}_*^-\mathbf{X} = \mathbf{X}$.

**Proof.** $(\mathbf{I} - \mathbf{V}_*\mathbf{V}_*^-)\mathbf{V}_*(\mathbf{I} - \mathbf{V}_*\mathbf{V}_*^-)' = \mathbf{0}$. Therefore

$$(\mathbf{I} - \mathbf{V}_*\mathbf{V}_*^-)\mathbf{V}(\mathbf{I} - \mathbf{V}_*\mathbf{V}_*^-)' + (\mathbf{I} - \mathbf{V}_*\mathbf{V}_*^-)\mathbf{XX'}(\mathbf{I} - \mathbf{V}_*\mathbf{V}_*^-)' = \mathbf{0}. \tag{A12}$$

Since each of the two terms in (A12) is n.n.d., their sum is null only if each term is null. Therefore $(\mathbf{I} - \mathbf{V}_*\mathbf{V}_*^-)\mathbf{XX'}(\mathbf{I} - \mathbf{V}_*\mathbf{V}_*^-)' = \mathbf{0}$ and so $(\mathbf{I} - \mathbf{V}_*\mathbf{V}_*^-)\mathbf{X} = \mathbf{0}$, i.e., $\mathbf{V}_*\mathbf{V}_*^-\mathbf{X} = \mathbf{X}$.

**Q.E.D.**

## REFERENCES

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

Albert, A. (1937). 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.

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

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.