SOME CONSEQUENCES OF A SINGULAR (CO)VARIANCE MATRIX IN THE LINEAR MODEL

Shayle R. Searle

Biometrics Unit and Statistics Center
Cornell University, Ithaca, New York 14853

BU-1254-M July 1994

ABSTRACT

Some of the effects of $\mathbf{V}$ being singular in the usual linear model $\mathbf{y} \sim (\mathbf{X}\boldsymbol{\beta}, \mathbf{V})$ are summarized.

1. INTRODUCTION

We deal with the linear model $\mathbf{y} \sim (\mathbf{X}\boldsymbol{\beta}, \mathbf{V})$, meaning that $\mathbf{y}$ has expected value $\mathrm{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$, and variance-covariance matrix $\mathrm{var}(\mathbf{y}) = \mathbf{V}$. The vector $\boldsymbol{\beta}$ is considered to have elements that are unknown constants, $\mathbf{X}$ is a known matrix of order $N \times p$ and rank r, with $r \leq p < N$, and $\mathbf{V}$ is symmetric, non-negative definite.

Every linear combination of elements of $\mathbf{X}\boldsymbol{\beta}$ is estimable and so rather than consider estimation of $\boldsymbol{\beta}$ or of $\boldsymbol{\lambda}'\mathbf{X}\boldsymbol{\beta}$ for some vector $\boldsymbol{\lambda}$ we consider estimation of just $\mathbf{X}\boldsymbol{\beta}$. The ordinary least squares estimator of $\mathbf{X}\boldsymbol{\beta}$ is

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

where $(\mathbf{X}'\mathbf{X})^{-}$ is any generalized inverse of $\mathbf{X}'\mathbf{X}$ and so satisfies $\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X} = \mathbf{X}'\mathbf{X}$. When $\mathbf{V}$ is non-singular, the best linear unbiased estimator of $\mathbf{X}\boldsymbol{\beta}$ is

$$\mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta}) = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} . \tag{2}$$

This note summarizes some results for the preceding model when $\mathbf{V}$ is singular.

First, we deal with conditions on $\mathbf{A}$ such that $\mathbf{Ay}$ is unbiased for $\mathbf{X}\boldsymbol{\beta}$; i.e.,

$$\mathrm{E}(\mathbf{Ay}) = \mathbf{AX}\boldsymbol{\beta} = \mathbf{X}\boldsymbol{\beta} . \tag{3}$$

In doing so we show there are matrices $\mathbf{A}$ satisfying (3) without $\mathbf{AX}$ equaling $\mathbf{X}$; i.e., with $\mathbf{AX} \neq \mathbf{X}$. And then we consider the effect on this situation when $\mathbf{V}$ is singular. The quoted results are from McCulloch and Searle (1995), wherein details of the derivations will be found.

Second, we consider the effect on (2) of $\mathbf{V}$ being singular, quoting from Searle (1994) which draws on the references therein, particularly Pukelsheim (1974).

2. EXAMPLES OF $\mathbf{AX}\beta = \mathbf{X}\beta$ WITH $\mathbf{AX} \neq \mathbf{X}$

2.1 A being a function of y

Suppose

$$\mathrm{E}(\mathbf{y}) = \mathrm{E}\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_1 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \mathbf{X}\beta \quad \text{for} \quad \mathbf{X} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}. \tag{4}$$

Then, so long as $y_2 \neq 0$ with probability 1, it will be found that defining

$$\mathbf{A}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & y_1/y_2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{gives} \quad \mathrm{E}(\mathbf{A}_1\mathbf{y}) = \mathbf{X}\beta \quad \text{but} \quad \mathbf{A}_1\mathbf{X} \neq \mathbf{X}\,.$$

Suppose further that

$$\mathbf{V} = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix}.$$

Then for

$$\mathbf{A}_2 = \begin{bmatrix} 1-y_3-y_4 & 0 & y_3 & y_3 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{we have} \quad \mathrm{E}(\mathbf{A}_2\mathbf{y}) = \mathbf{X}\beta \quad \text{but} \quad \mathbf{A}_2\mathbf{X} \neq \mathbf{X}\,.$$

Note that $\mathbf{A}_1$ and $\mathbf{A}_2$ are both functions of $\mathbf{y}$, but whereas elements of $\mathbf{A}_1\mathbf{y}$ are linear in elements of $\mathbf{y}$, those of $\mathbf{A}_2\mathbf{y}$ are not. Hence, we cannot write $E(\mathbf{A}_2\mathbf{y})$ as equaling $\mathbf{A}\,E(\mathbf{y})$ which is, of course, $\mathbf{AX}\beta$ and yet $E(\mathbf{A}_2\mathbf{y})$ does equal $\mathbf{X}\beta$.

2.2. Estimable constraints on β

Suppose in (4) we know *a priori*, or are prepared to assume, $\beta_1 = \beta_2$. Then for

$$\mathbf{A}_3 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{we have} \quad E(\mathbf{A}_3\mathbf{y}) = \mathbf{X}\beta \quad \text{but} \quad \mathbf{A}_3\mathbf{X} \neq \mathbf{X}\,.$$

The result $E(\mathbf{A}_3\mathbf{y}) = \mathbf{X}\beta$ occurs through using $\beta_1 = \beta_2$ which, in the words of Christensen (1990) "is more information than is given by the model", in this case, (4). Knowing $\beta_1 = \beta_2$ *a priori* we would usually reformulate the model as

$$E(\mathbf{y}) = \begin{bmatrix} \beta_1 \\ \beta_1 \\ \beta_1 \\ \beta_1 \end{bmatrix} = \mathbf{X}^*\beta^* \quad \text{for} \quad \mathbf{X}^* = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \text{and } \beta^* = \beta_1\,, \tag{5}$$

for which $\mathbf{A}_3\mathbf{X}^* = \mathbf{X}^*$.

Knowing *à priori*, the constraint $\beta_1 - \beta_2 = 0$ in the preceding example can be generalized to knowing

$$\mathbf{TX}\beta = \mathbf{d} \tag{6}$$

for given $\mathbf{T}$ and $\mathbf{d}$ such that $\mathbf{TX}$ has full row rank and $\mathbf{d}$ is in the column space of $\mathbf{TX}$. At the same time we generalize $E(\mathbf{Ay}) = \mathbf{X}\beta$ to wanting $E(\mathbf{Ay} + \mathbf{c}) = \mathbf{X}\beta$. In this situation McCulloch and Searle (1995) give a detailed example of always being able to find an $\mathbf{A}$ such that $E(\mathbf{Ay} + \mathbf{c}) = \mathbf{X}\beta$ but with $\mathbf{AX} \neq \mathbf{X}$. They also prove a theorem that this can always be done in general. Their development depends on partitioning $\mathbf{X}$ and β to rewrite (6) as

$$\mathbf{TX}_1\beta_1 + \mathbf{TX}_2\beta_2 = \mathbf{d} \tag{7}$$

with $\mathbf{TX}_1$ being non-singular. Then, on defining

$$\mathbf{W} = [\mathbf{I} - \mathbf{X}_1(\mathbf{TX}_1)^{-1}\mathbf{T}]\mathbf{X}_2\,,$$

their result is that for

$$\mathbf{A} = \mathbf{W}\mathbf{W}^{-} \quad \text{and} \quad \mathbf{c} = -(\mathbf{A} - \mathbf{I})\mathbf{X}_1(\mathbf{T}\mathbf{X}_1^{-1})\mathbf{d}$$

one will have

$$\mathrm{E}(\mathbf{Ay} + \mathbf{c}) = \mathbf{X}\boldsymbol{\beta} \quad \text{but} \quad \mathbf{AX} \neq \mathbf{X} \, .$$

The alternative to all this is, of course, to incorporate (6) into the model (4) so as to reformulate the model in the manner illustrated by (5).

3. ESTABLISHING THE NECESSITY OF $\mathbf{AX} = \mathbf{X}$ FOR $\mathrm{E}(\mathbf{Ay} + \mathbf{c}) = \mathbf{X}\boldsymbol{\beta}$

The sufficiency of $\mathbf{AX} = \mathbf{X}$ and $\mathbf{c} = \mathbf{0}$ for $\mathrm{E}(\mathbf{Ay} + \mathbf{c}) = \mathbf{X}\boldsymbol{\beta}$ to be satisfied is obvious. But the necessity of $\mathbf{AX} = \mathbf{X}$ seems to be negated by the previous section. However, the situation there consists of two separate cases: (i) $\mathbf{A}$ being a function of $\mathbf{y}$, and (ii) having $\mathbf{TX}\boldsymbol{\beta} = \mathbf{d}$ known, *à priori*. Both of these warrant exclusion: (i) because it allows non-linear estimators, and (ii) because it can be avoided by model reformulation. On excluding (i) and (ii) it is then easy for non-singular $\mathbf{V}$ to show that $\mathbf{AX} = \mathbf{X}$ and $\mathbf{c} = \mathbf{0}$ are necessary conditions for having

$$\mathrm{E}(\mathbf{Ay} + \mathbf{c}) = \mathbf{X}\boldsymbol{\beta} \, . \tag{8}$$

It is the singular $\mathbf{V}$ case that has generated discussion, (e.g., Christensen, 1990; Harville, 1990, and Puntanen and Styan, 1989). With $\mathbf{V}$ being singular there is always a matrix $\mathbf{T}$, of maximal full row rank such that $\mathbf{TV} = \mathbf{0}$ and so $\mathrm{var}(\mathbf{Ty}) = \mathbf{0}$. This means

$$\mathbf{Ty} = \text{some constant, } \mathbf{b} \text{ say, with probability } 1 \, . \tag{9}$$

Hence

$$\mathrm{E}(\mathbf{Ty}) = \mathbf{TX}\boldsymbol{\beta} \quad \Rightarrow \quad \mathrm{E}(\mathbf{b}) = \mathbf{TX}\boldsymbol{\beta} \, ,$$

i.e.,

$$\mathbf{TX}\boldsymbol{\beta} = \mathbf{b}, \text{ with probability } 1 \, . \tag{10}$$

Initially one thinks of $\mathbf{TX}\boldsymbol{\beta} = \mathbf{b}$ of (10) as the same kind of equation as $\mathbf{TX}\boldsymbol{\beta} = \mathbf{d}$ of (6). However, there is a big difference: the $\mathbf{d}$ of (6) is a given constant whereas the $\mathbf{b}$ of (10) is not. This is because we do not know the value of $\mathbf{b}$ and are unwilling to exclude any possible values for it. Based on this distinction McCulloch and Searle (1995) show that so long as $\mathbf{A}$ does not depend on $\mathbf{y}$, singular $\mathbf{V}$ implies that $\mathbf{AX} = \mathbf{X}$ and $\mathbf{c} = \mathbf{0}$ are necessary for having $\mathrm{E}(\mathbf{Ay} + \mathbf{c}) = \mathbf{X}\boldsymbol{\beta}$. Thus, on excluding both the dependence of **A** on **y**, and the use of estimable constraints on β, the joint conditions $\mathbf{AX} = \mathbf{X}$ and $\mathbf{c} = \mathbf{0}$ are both necessary and sufficient for $E(\mathbf{Ay} + \mathbf{c}) = \mathbf{X}\beta$.

4. BLUE($\mathbf{X}\beta$) FOR SINGULAR V

4.1 Early results

Discussion of BLUE($\mathbf{X}\beta$) makes great use of

$$\mathbf{M} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}' = \mathbf{I} - \mathbf{X}\mathbf{X}^{+} = \mathbf{M}' = \mathbf{M}^2, \text{ with } \mathbf{MX} = \mathbf{0}\,,$$

where $\mathbf{X}^{+}$ is the Moore-Penrose inverse of **X**. Pukelsheim's (1974) results are the first two of

$$\text{BLUE}(\mathbf{X}\beta) = (\mathbf{I} - \mathbf{M})[\mathbf{I} - \mathbf{VM}(\mathbf{MVM})^{+}\mathbf{M}]\mathbf{y} \tag{11a}$$

$$= (\mathbf{I} - \mathbf{M})[\mathbf{I} - \mathbf{VM}(\mathbf{MVM})^{+}]\mathbf{y} \tag{11b}$$

$$= (\mathbf{I} - \mathbf{M})[\mathbf{I} - \mathbf{V}(\mathbf{MVM})^{+}]\mathbf{y} \tag{11c}$$

the second, as he notes, because $(\mathbf{MS})^{+}\mathbf{M} = (\mathbf{MS})^{+}$, and the third, as noted in Searle (1994), because $\mathbf{M}(\mathbf{MVM})^{+}\mathbf{M} = (\mathbf{MVM})^{+}$. And (11a) is similar to a result in Albert (1967).

4.2 Reduction for non-singular V

Despite the obvious dissimilarity of (11) with BLUE($\mathbf{X}\beta$) $= \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$ for the non-singular **V**, any of equations (11) do indeed reduce to this when **V** is non-singular. To show this define

$$\mathbf{Q} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1} = \mathbf{Q}' \quad \text{with} \quad \mathbf{QX} = \mathbf{0}\,.$$

Then observe that

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

Moreover

$$\mathbf{VQ} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1} \quad \text{and so} \quad \mathbf{MVQ} = \mathbf{M}\,.$$

It can then be shown that $\mathbf{Q} = \mathbf{Q}' = (\mathbf{MVM})^{+}$ and so (11a) is

$$\begin{aligned}
\text{BLUE}(\mathbf{X}\beta) &= (\mathbf{I} - \mathbf{M})(\mathbf{I} - \mathbf{VMQM})\mathbf{y} \\
&= (\mathbf{I} - \mathbf{M} - \mathbf{VMQM} + \mathbf{MVMQM})\mathbf{y} \\
&= (\mathbf{I} - \mathbf{M} - \mathbf{VQ} + \mathbf{MVQ})\mathbf{y} = (\mathbf{I} - \mathbf{M} - \mathbf{VQ} + \mathbf{M})\mathbf{y} \\
&= \mathbf{V}(\mathbf{V}^{-1} - \mathbf{Q})\mathbf{y} \\
&= \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}\,.
\end{aligned}$$

4.3 Using any generalized inverse of MVM

Expressions (11) for BLUE($X\beta$) somewhat lack generality through their dependence on the unique, Moore-Penrose inverse $(\mathbf{MVM})^+$ rather than on any generalized inverse $(\mathbf{MVM})^-$. However, by starting with $\mathbf{y} = (\mathbf{I}-\mathbf{M})\mathbf{y} + \mathbf{My}$, where $E[(\mathbf{I}-\mathbf{M})\mathbf{y}] = X\beta$ and $E(\mathbf{My}) = \mathbf{0}$, we see that any linear combination of $(\mathbf{I}-\mathbf{M})\mathbf{y}$ and $\mathbf{My}$ can be unbiased for $\lambda' X\beta$ only if the term in $(\mathbf{I}-\mathbf{M})\mathbf{y}$ in that combination is $\lambda'(\mathbf{I}-\mathbf{M})\mathbf{y}$. Therefore, in asking "For what vector τ' does adding $\tau'\mathbf{My}$ to $\lambda'(\mathbf{I}-\mathbf{M})\mathbf{y}$ yield BLUE($\mathbf{X}\beta$)?" we seek τ' to minimize the variance of $\lambda'(\mathbf{I}-\mathbf{M})\mathbf{y} + \tau'\mathbf{My}$. This leads, as is shown in Searle (1994) to

$$\mathrm{BLUE}(\mathbf{X}\beta) = (\mathbf{I}-\mathbf{M})[\mathbf{I}-\mathbf{VM}(\mathbf{MVM})^-\mathbf{M}]\mathbf{y}\,. \tag{12}$$

This is the same as (11a) but with the unique Moore-Penrose inverse $(\mathbf{MVM})^+$ of (11a), replaced by a non-unique generalized inverse $(\mathbf{MVM})^-$ in (12). However, Searle (1994) shows that $\mathbf{VM}(\mathbf{MVM})^-\mathbf{My}$ is invariant to the choice of $(\mathbf{MVM})^-$. But although $\mathbf{M}(\mathbf{MVM})^-\mathbf{M}$ is a generalized inverse of $\mathbf{MVM}$, we cannot replace $\mathbf{M}(\mathbf{MVM})^-\mathbf{M}$ in (12) by any generalized inverse of $\mathbf{MVM}$ because not every generalized inverse of $\mathbf{MVM}$ has $\mathbf{M}$ as a left and right factor as does $\mathbf{M}(\mathbf{MVM})^-\mathbf{M}$, and that is an essential feature of (12). Thus although (12) equals (11a) we cannot extend (12) to have (11b) and (11c) with $(\mathbf{MVM})^+$ replaced by $(\mathbf{MVM})^-$. It is the uniqueness of a Moore-Penrose inverse that permits (11a) to yield (11b) and (11c); and that uniqueness does not exist for $(\mathbf{MVM})^-$.

4.4 A simple form of BLUE($X\beta$)

Nevertheless, one can further simplify (12) to yield

$$\mathrm{BLUE}(\mathbf{X}\beta) = \mathbf{y} - \mathbf{VM}(\mathbf{MVM})^-\mathbf{My} \tag{13}$$

through using the equality $\mathbf{MVM}(\mathbf{MVM})^-\mathbf{My} = \mathbf{My}$ established in Searle (1994). This also leads to showing that (13) is invariant to the choice of generalized inverse for $(\mathbf{MVM})^-$.

4.5 An "obvious" generalization of BLUE($X\beta$) for non-singular V

Suppose in (2), the $\mathrm{BLUE}(X\beta) = X(X'V^{-1}X)^-X'V^{-1}y$ of the non-singular $\mathbf{V}$ case, we simply replace $\mathbf{V}^{-1}$ by any generalized inverse $\mathbf{V}^-$ for the singular $\mathbf{V}$ case. Denote such an expression by $\hat{\mu}(\mathbf{V}^-)$; i.e.,

$$\hat{\mu}(\mathbf{V}^-) = \mathbf{X}(\mathbf{X}'\mathbf{V}^-\mathbf{X})^-\mathbf{X}'\mathbf{V}^-\mathbf{y}\ . \tag{14}$$

A first concern about $\hat{\mu}(\mathbf{V}^-)$ is that it involves two generalized inverses, $\mathbf{V}^-$ and $(\mathbf{X}'\mathbf{V}^-\mathbf{X})^-$. However, Searle (1994) shows that if $\mathbf{V}\mathbf{V}^- = \mathbf{X}$, then $\hat{\mu}(\mathbf{V}^-)$ is invariant to those inverses. This is so because $\mathbf{V}\mathbf{V}^-\mathbf{X} = \mathbf{X}$ for any particular $\mathbf{V}^-$ implies that $\mathbf{V}\mathbf{V}^-\mathbf{X} = \mathbf{X}$ is true for every $\mathbf{V}^-$; and that $\mathbf{X}'\mathbf{V}^-\mathbf{X}$ and $\mathbf{X}'\mathbf{V}^-\mathbf{y}$ (for almost all $\mathbf{y}$) are invariant to $\mathbf{V}^-$.

The widespread familiarity of (2), which motivates considering (14), together with the dissimilarity of (14) from (12), assuredly prompts the question "Of what use is (14)?" And, more particularly, "Can (14) ever be the same as (12)?" The answer to this second question is yes, as given by the following theorem.

Theorem 1 A necessary and sufficient condition (N.S.C.) for $\hat{\mu}(\mathbf{V}^-)$ to equal BLUE($\mathbf{X}\beta$) of (11a), (11b), (11c), (12) or (13) is that $\mathbf{V}\mathbf{V}^-\mathbf{X} = \mathbf{X}$. This, be it noted, is the same condition for $\hat{\mu}(\mathbf{V}^-)$ to be invariant to $\mathbf{V}^-$ and to $(\mathbf{X}'\mathbf{V}^-\mathbf{X})^-$.

Proof is detailed in Searle (1994).

4.4 Equality to OLSE(Xβ)

Expression (1) shows that OLSE($\mathbf{X}\beta$) does not involve $\mathbf{V}$; and so is much easier to calculate than BLUE($\mathbf{X}\beta$). (For example, knowing the variance and covariances – or having to estimate them – that go into $\mathbf{V}$ is avoided.) Consequently, conditions under which BLUE($\mathbf{X}\beta$) equals OLSE($\mathbf{X}\beta$) are of great interest; and they have been established by Zyskind (1967) and recently discussed at length by Puntanen and Styan (1989). Of the many equivalent conditions, the one used here is that BLUE($\mathbf{X}\beta$) = OLSE($\mathbf{X}\beta$) if and only if $\mathbf{V}\mathbf{X} = \mathbf{X}\mathbf{B}$ for some $\mathbf{B}$. Having established when $\hat{\mu}(\mathbf{V}^-)$ equals BLUE($\mathbf{X}\beta$) we now ask "When does $\hat{\mu}(\mathbf{V}^-)$ = OLSE($\mathbf{X}\beta$)?" This is answered by Theorem 2.

Theorem 2 $\hat{\mu}(\mathbf{V}^-)$ = OLSE($\mathbf{X}\beta$) for any $\mathbf{V}^-$ if and only if $\hat{\mu}(\mathbf{V}^-)$ and OLSE($\mathbf{X}\beta$) each equal BLUE($\mathbf{X}\beta$); in which case, by Theorem 1 and Zyskind's results, respectively, $\mathbf{V}\mathbf{V}^-\mathbf{X} = \mathbf{X}$ and $\mathbf{V}\mathbf{X} = \mathbf{X}\mathbf{B}$ for some $\mathbf{B}$.

Proof is given in Searle (1994). It amounts to confining attention to invariant $\hat{\mu}(\mathbf{V}^-)$ which means that $\mathbf{V}\mathbf{V}^-\mathbf{X} = \mathbf{X}$, which in turn gives $\hat{\mu}(\mathbf{V}^-) = \text{BLUE}(\mathbf{X}\beta)$, by Theorem 1. And so Zyskind's result gives $\hat{\mu}(\mathbf{V}^-)$ [already equaling $\text{BLUE}(\mathbf{X}\beta)] = \text{OLSE}(\mathbf{X}\beta)$ if and only if $\mathbf{VX} = \mathbf{XB}$. Additionally, it is shown that $\mathbf{V}\mathbf{V}^-\mathbf{X}$ is a necessary condition for $\hat{\mu}(\mathbf{V}^-) = \text{OLSE}(\mathbf{X}\beta)$.

References

Albert, A. (1967). The Gauss-Markov theorem for regression model with possibly singular covariances. *Siam J. of Applied Mathematics* **24**, 182-7.

Christensen, R. (1990). Comment on Puntanen and Styan (1989). *The American Statistician* **44**, 191-2.

Harville, D.A. (1990). Comment on Puntanen and Styan (1989). *The American Statistician* **44**, 192.

McCulloch, C.E. and Searle, S.R. (1995). On an identity derived from unbiasedness in linear models. *The American Statistician* (in press).

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

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

Puntanen, S. and Styan, G.P.H. (1990). Reply to Christensen (1990) and Harville (1990). *The American Statistician* **44**, 192.

Searle, S.R. (1994). Extending some results and proofs for the singular linear model *Linear Algebra and its Applications*, Third Special Issue on Matrices and Statistics (in press).

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