Gujarat Journal of Statistics and Data Science
GUJARAT STATISTICAL ASSOCIATION

Properties of the Matrix $V + XTX'$ in Linear Statistical Models

Stephen J. Haslett[1], Augustyn Markiewicz[2] and Simo Puntanen[3]

1. *School of Mathematical and Computational Sciences & Research Centre for Hauora and Health, Massey University, New Zealand, email: s.j.haslett@massey.ac.nz; Research School of Finance, Actuarial Studies and Statistics, The Australian National University, Australia; NIASRA, Faculty of Engineering and Information Sciences, University of Wollongong, Australia*
2. *Department of Mathematical and Statistical Methods, Poznań University of Life Sciences, Wojska Polskiego 28, PL-60637 Poznań, Poland, email: amark@up.poznan.pl*
3. *Faculty of Information Technology and Communication Sciences, FI-33014 Tampere University, Finland, email: simo.puntanen@tuni.fi*

Received: 27 Nov 2021 / Revised: 02 Feb 2022 / Accepted: 04 Apr 2022

Abstract

It is well known, due originally to C.R. Rao in early 1970s, that the best linear unbiased estimator, BLUE, of $\mathbf{X}\beta$ in the linear model $\mathscr{M} = \{\mathbf{y},\, \mathbf{X}\beta,\, \mathbf{V}\}$ can be expressed in the form $\mathbf{X}(\mathbf{X}'\mathbf{W}^-\mathbf{X})^-\mathbf{X}'\mathbf{W}^-\mathbf{y}$, where $\mathbf{W}$ is a specific matrix of the form $\mathbf{W} = \mathbf{V}+\mathbf{XTX}'$ with $\mathbf{T}$ satisfying the column space condition $\mathscr{C}(\mathbf{W}) = \mathscr{C}(\mathbf{X} : \mathbf{V})$. We denote this class of matrices as $\mathcal{W}$. Choice of $\mathbf{T}$ as an identity matrix gives an obvious member $\mathbf{W} = \mathbf{V} + \mathbf{XX}' \in \mathcal{W}$. The matrices belonging to the class $\mathcal{W}$ have several interesting mathematical properties. In particular, the use of matrix $\mathbf{W} \in \mathcal{W}$ appears to be surprisingly handy and helpful tool when dealing with the linear statistical models. Our aim is to review and collect together some essential features of $\mathbf{W}$ and its use in linear statistical models. While doing this, we go through some related basic properties of the best linear unbiased estimation.

Keywords: Best linear unbiased estimator, BLUE, Column space, Generalized inverse, Löwner ordering,Linear sufficiency, Partitioned linear model.
MSC: 62J05, 62J10

1 Introduction: Basic Tools

We begin this article by introducing the notation and the basic mathematical tools that we are going to use; these matters will occupy the first two sections. In a nutshell, we slowly approach the problems what we meet if we want to use a particular kind of estimator in the linear model to catch the best linear unbiased estimator, BLUE, for the unknown parametric function. In our considerations the matrix of the type $\mathbf{W} = \mathbf{V} + \mathbf{XTX}'$, where $\mathbf{X}'$ is the transpose of $\mathbf{X}$, will have the main role. But before the main goal, we need some basic tools and definitions.

In this article we consider the linear model $\mathbf{y} = \mathbf{X}\beta + \varepsilon$ or shortly

$$\mathscr{M} = \{\mathbf{y}, \mathbf{X}\beta, \mathbf{V}\}\,.$$

Here $\mathbf{y}$ is an n-dimensional observable response variable, and ε is an unobservable random error with a known covariance matrix $\operatorname{cov}(\varepsilon) = \operatorname{cov}(\mathbf{y}) = \mathbf{V}$ (can be singular) and expectation $\operatorname{E}(\varepsilon) = \mathbf{0} \in \mathbb{R}^n$. The matrix $\mathbf{X}$ is a known $n \times p$ matrix, i.e., $\mathbf{X} \in \mathbb{R}^{n\times p}$, and $\boldsymbol{\beta} \in \mathbb{R}^p$ is a vector of fixed (but unknown) parameters. We will denote $\boldsymbol{\mu} = \mathbf{X}\beta$ so that $\operatorname{E}(\mathbf{y}) = \boldsymbol{\mu} = \mathbf{X}\beta$. Sometimes the covariance matrix is of the type $\sigma^2\mathbf{V}$, where σ^2 is an unknown positive constant.

By the partitioned linear model we mean that $\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}$, or shortly denoted

$$\mathscr{M}_{12} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\} = \{\mathbf{y},\, \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2,\, \mathbf{V}\}\,.$$

In addition to the *full* model $\mathscr{M}_{12}$, we will consider the *small* models $\mathscr{M}_i = \{\mathbf{y},\, \mathbf{X}_i\boldsymbol{\beta}_i,\, \mathbf{V}\}$, $i = 1, 2$, and the *reduced* model

$$\mathscr{M}_{12\cdot 2} = \{\mathbf{M}_2\mathbf{y},\, \mathbf{M}_2\mathbf{X}_1\boldsymbol{\beta}_1,\, \mathbf{M}_2\mathbf{V}\mathbf{M}_2\}\,,$$

where $\mathbf{M}_2 = \mathbf{I}_n - \mathbf{P}_{\mathbf{X}_2}$, with $\mathbf{P}_{\mathbf{X}_2}$ being the orthogonal projector onto the column space of $\mathbf{X}_2$ and $\mathbf{I}_n$ is the $n \times n$ identity matrix. Premultiplying the model $\mathscr{M}$ by an $f \times n$ matrix $\mathbf{F}$ yields the *transformed* model

$$\mathbf{F}\mathbf{y} = \mathbf{F}\mathbf{X}\boldsymbol{\beta} + \mathbf{F}\boldsymbol{\varepsilon}\,, \quad \text{or shortly } \mathscr{T} = \{\mathbf{F}\mathbf{y}, \mathbf{F}\mathbf{X}\boldsymbol{\beta}, \mathbf{F}\mathbf{V}\mathbf{F}'\}\,.$$

The reduced model $\mathscr{M}_{12\cdot 2}$ is of course one example of the transformed models. We will also shortly consider the linear model with *new* (unobserved, to be predicted) observations. This means that in addition to $\mathscr{M}$, we are dealing with a $q \times 1$ unobservable random vector $\mathbf{y}_*$ containing new observations. These new observations are assumed to come from $\mathbf{y}_* = \mathbf{X}_*\boldsymbol{\beta} + \boldsymbol{\varepsilon}_*$, where $\mathbf{X}_*$ is a known $q \times p$ matrix, and $\boldsymbol{\varepsilon}_*$ is a q-dimensional random error vector whose (cross-)covariance matrix with $\mathbf{y}$ is known.

As for the notation: $\mathrm{r}(\mathbf{A})$, $\mathbf{A}^-$, $\mathbf{A}^+$, $\mathscr{C}(\mathbf{A})$, $\mathscr{N}(\mathbf{A})$, and $\mathscr{C}(\mathbf{A})^\perp$, denote, respectively, the rank, a generalized inverse, the (unique) Moore–Penrose inverse, the column space, the null space, and the orthogonal complement of the column space of the matrix $\mathbf{A}$. By $\mathbf{A}^\perp$ we denote any matrix satisfying $\mathscr{C}(\mathbf{A}^\perp) = \mathscr{C}(\mathbf{A})^\perp$. Furthermore, we will write $\mathbf{P}_\mathbf{A} = \mathbf{P}_{\mathscr{C}(\mathbf{A})} = \mathbf{A}\mathbf{A}^+ = \mathbf{A}(\mathbf{A}'\mathbf{A})^-\mathbf{A}'$ to denote the orthogonal projector onto $\mathscr{C}(\mathbf{A})$. The orthogonal projector onto $\mathscr{C}(\mathbf{A})^\perp$ is denoted as $\mathbf{Q}_\mathbf{A} = \mathbf{I}_a - \mathbf{P}_\mathbf{A}$. We will shorten our notation as

$$\mathbf{H} = \mathbf{P}_\mathbf{X}\,, \quad \mathbf{M} = \mathbf{I}_n - \mathbf{P}_\mathbf{X}\,, \quad \mathbf{M}_i = \mathbf{I}_n - \mathbf{P}_{\mathbf{X}_i}\,,\ i = 1, 2\,.$$

One obvious choice for $\mathbf{X}^\perp$ is $\mathbf{M}$.

Next we recall some basic concepts when dealing with the best linear unbiased estimation. In particular we explore the problems when figuring out for which choice of matrix $\mathbf{N} \in \mathbb{R}^{n\times n}$ an estimator of the type

$$\mathbf{X}\mathbf{b} = \mathbf{X}(\mathbf{X}'\mathbf{N}\mathbf{X})^-\mathbf{X}'\mathbf{N}\mathbf{y}$$

provides a representation for the best linear estimator, BLUE, of $\mathbf{X}\boldsymbol{\beta}$. Notice that the above representation can be interpreted to arise from solving $\mathbf{b}$ from

$$\mathbf{X}'\mathbf{N}\mathbf{X}\mathbf{b} = \mathbf{X}'\mathbf{N}\mathbf{y}, \tag{1}$$

supposing that (1) is solvable for $\mathbf{b}$; this happens if and only if $\mathbf{X}'\mathbf{N}\mathbf{y} \in \mathscr{C}(\mathbf{X}'\mathbf{N}\mathbf{X})$. We will go through various particular choices of $\mathbf{N}$:

- the first and simplest case is $\mathbf{N} = \mathbf{I}_n$,
- then we take $\mathbf{N} = \mathbf{V}^{-1}$, and $\mathbf{N} = \mathbf{V}^+$, or $\mathbf{N} = \mathbf{V}^-$, i.e., $\mathbf{N} \in \{\mathbf{V}^-\}$,
- and so we slowly approach the most general case which is $\mathbf{N} = \mathbf{W}^- = (\mathbf{V} + \mathbf{X}\mathbf{T}\mathbf{X}')^-$, where $\mathbf{W}$ belongs to a specific class $\mathcal{W}$, say.

A linear statistic $\mathbf{By}$ is said to be a linear unbiased estimator, LUE, for $\mathbf{K}\boldsymbol{\beta}$, where $\mathbf{K} \in \mathbb{R}^{k\times p}$, if its expectation is equal to $\mathbf{K}\boldsymbol{\beta}$, i.e.,

$$\mathrm{E}(\mathbf{By}) = \mathbf{BX}\boldsymbol{\beta} = \mathbf{K}\boldsymbol{\beta} \quad \text{for all } \boldsymbol{\beta} \in \mathbb{R}^p, \quad \text{i.e.,} \quad \mathbf{BX} = \mathbf{K}\,.$$

When $\mathscr{C}(\mathbf{K}') \subseteq \mathscr{C}(\mathbf{X}')$ holds, $\mathbf{K}\boldsymbol{\beta}$ is said to be estimable. The LUE $\mathbf{By}$ is the best LUE, BLUE, of estimable $\mathbf{K}\boldsymbol{\beta}$ if $\mathbf{By}$ has the smallest covariance matrix in the Löwner sense among all linear unbiased estimators of $\mathbf{K}\boldsymbol{\beta}$:

$$\mathrm{cov}(\mathbf{By}) \leq_{\mathsf{L}} \mathrm{cov}(\mathbf{B}_{\#}\mathbf{y}) \quad \text{for all } \mathbf{B}_{\#} : \mathbf{B}_{\#}\mathbf{X} = \mathbf{K}\,,$$

that is, $\mathrm{cov}(\mathbf{B}_{\#}\mathbf{y}) - \mathrm{cov}(\mathbf{By})$ is nonnegative definite for all $\mathbf{B}_{\#} : \mathbf{B}_{\#}\mathbf{X} = \mathbf{K}$.

Under the model $\mathscr{M}$, the ordinary least squares estimator, OLSE, for $\boldsymbol{\beta}$ is the solution minimizing the quantity $\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2$ with respect to $\boldsymbol{\beta}$ yielding to the normal equation $\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{y}$. Thus, if $\mathbf{X}$ has full column rank, the OLSE of $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{X}^{+}\mathbf{y}$. Moreover, the OLSE of $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ is

$$\mathrm{OLSE}(\mathbf{X}\boldsymbol{\beta}) = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y} = \mathbf{XX}^{+}\mathbf{y} = \mathbf{P}_{\mathbf{X}}\mathbf{y} = \mathbf{Hy} = \hat{\boldsymbol{\mu}}\,.$$

Obviously $\hat{\boldsymbol{\mu}} = \mathbf{Hy}$ is a LUE for $\mathbf{X}\boldsymbol{\beta}$; however, $\hat{\boldsymbol{\mu}}$ is the BLUE for $\mathbf{X}\boldsymbol{\beta}$ only under specific conditions. Now the well-known simple version of the Gauss–Markov theorem says that under the model $\mathscr{M}_{\mathbf{I}} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{I}_n\}$, the OLSE of $\mathbf{X}\boldsymbol{\beta}$ is the BLUE of $\mathbf{X}\boldsymbol{\beta}$, or shortly

$$\hat{\boldsymbol{\mu}}(\mathscr{M}_{\mathbf{I}}) = \mathrm{OLSE}(\mathbf{X}\boldsymbol{\beta} \mid \mathscr{M}_{\mathbf{I}}) = \mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta} \mid \mathscr{M}_{\mathbf{I}}) = \tilde{\boldsymbol{\mu}}(\mathscr{M}_{\mathbf{I}})\,. \tag{2}$$

Consider now the model $\mathscr{M}$ where $\mathbf{V}$ is positive definite, and suppose that $\mathbf{V}^{1/2}$ is the positive definite square root of $\mathbf{V}$. Premultiplying $\mathscr{M}$ by $\mathbf{V}^{-1/2}$ yields $\mathscr{M}_{\#} = \{\mathbf{V}^{-1/2}\mathbf{y},\, \mathbf{V}^{-1/2}\mathbf{X}\boldsymbol{\beta},\, \mathbf{I}_n\}$. In light of (2), the BLUE of $\mathbf{X}\boldsymbol{\beta}$ under $\mathscr{M}_{\#}$ equals the OLSE under $\mathscr{M}_{\#}$ and thus

$$\mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta} \mid \mathscr{M}_{\#}) = \tilde{\boldsymbol{\mu}}(\mathscr{M}_{\#}) = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} =: \mathbf{P}_{\mathbf{X};\mathbf{V}^{-1}}\mathbf{y},$$

where $\mathbf{P}_{\mathbf{X};\mathbf{V}^{-1}}$ is the orthogonal projector onto $\mathscr{C}(\mathbf{X})$ when the inner product matrix is $\mathbf{V}^{-1}$. It appears that

$$\mathbf{P}_{\mathbf{X};\mathbf{V}^{-1}}\mathbf{y} = \mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta} \mid \mathscr{M}) = \mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta} \mid \mathscr{M}_{\#})\,. \tag{3}$$

The result (3), sometimes referred to as the Aitken-approach, is well known in statistical textbooks; see Aitken (1935) [1].

It is clear that

$$\begin{aligned}\min_{\boldsymbol{\beta}}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) &= (\mathbf{y} - \mathbf{P}_{\mathbf{X};\mathbf{V}^{-1}}\mathbf{y})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{P}_{\mathbf{X};\mathbf{V}^{-1}}\mathbf{y}) \\ &= (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}_0)'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}_0)\,,\end{aligned}$$

where $\boldsymbol{\beta}_0$ is any solution to the generalized normal equation

$$\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{V}^{-1}\mathbf{y}. \tag{4}$$

Equation (4) is always, i.e., for any $\mathbf{y} \in \mathbb{R}^n$, solvable for $\boldsymbol{\beta}$ and the general solution can be expressed, e.g., as

$$\begin{aligned}\boldsymbol{\beta}_0 &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} + [\mathbf{I}_p - \mathbf{X}'\mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{+}]\mathbf{t} \\ &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} + (\mathbf{I}_p - \mathbf{P}_{\mathbf{X}'})\mathbf{t}\,,\end{aligned}$$

where $\mathbf{t} \in \mathbb{R}^p$ is free to vary. Thus

$$\mathbf{X}\boldsymbol{\beta}_0 = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} = \mathbf{P}_{\mathbf{X};\mathbf{V}^{-1}}\mathbf{y}.$$

What about if $\mathbf{V}$ is singular? Can we straight away replace $\mathbf{V}^{-1}$ with $\mathbf{V}^{+}$ or even with an arbitrary generalized inverse $\mathbf{V}^{-}$? No, we better be careful: we need further information before such replacement can be done.

What happens if we try to use

$$\mathbf{P}_{\mathbf{X};\mathbf{V}^{+}}\mathbf{y} := \mathbf{X}(\mathbf{X}'\mathbf{V}^{+}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{+}\mathbf{y}$$

as a BLUE? First we observe that $\mathbf{P}_{\mathbf{X};\mathbf{V}^{+}}\mathbf{y}$ is a LUE for $\mathbf{X}\boldsymbol{\beta}$ if and only if

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

which, by Proposition 1.1 below, holds if and only if $\mathscr{C}(\mathbf{X}') \subseteq \mathscr{C}(\mathbf{X}'\mathbf{V}^{+}) = \mathscr{C}(\mathbf{X}'\mathbf{V})$, which further is equivalent to each of the following conditions:

$$\mathscr{C}(\mathbf{X}') = \mathscr{C}(\mathbf{X}'\mathbf{V})\,, \quad \mathrm{r}(\mathbf{X}) = \mathrm{r}(\mathbf{X}'\mathbf{V})\,, \quad \mathscr{C}(\mathbf{X}) \cap \mathscr{C}(\mathbf{V})^{\perp} = \{\mathbf{0}\}\,. \tag{5}$$

Above we have used the rank rule of the matrix product

$$\mathrm{r}(\mathbf{AB}) = \mathrm{r}(\mathbf{A}) - \dim \mathscr{C}(\mathbf{A}') \cap \mathscr{C}(\mathbf{B})^{\perp}. \tag{6}$$

From (5) we observe that if $\mathscr{C}(\mathbf{X}) \subseteq \mathscr{C}(\mathbf{V})$, then $\mathbf{P}_{\mathbf{X};\mathbf{V}^{+}}\mathbf{y}$ is unbiased for $\mathbf{X}\boldsymbol{\beta}$. The model $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$, where $\mathscr{C}(\mathbf{X}) \subseteq \mathscr{C}(\mathbf{V})$, is often called a *weakly singular* linear model. We observe that under a weakly singular linear model the product $\mathbf{X}(\mathbf{X}'\mathbf{V}^{+}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{+}$ is invariant for any choice of $(\mathbf{X}'\mathbf{V}^{+}\mathbf{X})^{-}$ in view of the following Proposition, cf. [41, Lemma 2.2.4].

Proposition 1.1. *For nonnull matrices $\mathbf{A}$ and $\mathbf{C}$ the following holds:*

(a) $\mathbf{AB}^{-}\mathbf{C} = \mathbf{AB}^{+}\mathbf{C}$ *for all* $\mathbf{B}^{-}$ $\iff$ $\mathscr{C}(\mathbf{C}) \subseteq \mathscr{C}(\mathbf{B})\ \&\ \mathscr{C}(\mathbf{A}') \subseteq \mathscr{C}(\mathbf{B}')$.

(b) $\mathbf{AA}^{-}\mathbf{C} = \mathbf{C}$ *for some (and hence for all)* $\mathbf{A}^{-}$ $\iff$ $\mathscr{C}(\mathbf{C}) \subseteq \mathscr{C}(\mathbf{A})$.

(c) $\mathbf{C}'\mathbf{A}^{-}\mathbf{A} = \mathbf{C}'$ *for some (and hence for all)* $\mathbf{A}^{-}$ $\iff$ $\mathscr{C}(\mathbf{C}) \subseteq \mathscr{C}(\mathbf{A}')$.

Things become a bit trickier when we consider an estimator like

$$\mathbf{P}_{\mathbf{X};\mathbf{V}^{-}}\mathbf{y} := \mathbf{X}(\mathbf{X}'\mathbf{V}^{-}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-}\mathbf{y},$$

where $\mathbf{V}^{-}$ is a given generalized inverse of $\mathbf{V}$. Supposing that $\mathscr{C}(\mathbf{X}) \subseteq \mathscr{C}(\mathbf{V})$ we see that the observed value of $\mathbf{P}_{\mathbf{X};\mathbf{V}^{-}}\mathbf{y}$ is invariant for $\mathbf{V}^{-}$ if and only if $\mathbf{y} \in \mathscr{C}(\mathbf{V})$. Do we know that this holds? The answer is *yes* in the case of a *consistent* linear model by which we mean such a model where the observed value of $\mathbf{y}$ belongs to $\mathscr{C}(\mathbf{X} : \mathbf{V})$:

$$\mathbf{y} \in \mathscr{C}(\mathbf{X} : \mathbf{V}) = \mathscr{C}(\mathbf{X}) \oplus \mathscr{C}(\mathbf{VM}) = \mathscr{C}(\mathbf{X}) \boxplus \mathscr{C}(\mathbf{MV})\,, \tag{7}$$

where "$\oplus$" refers to the direct sum and "$\boxplus$" to the direct sum of orthogonal subspaces. For decompositions in (7), see [37, Lemma 2.1]. The models we consider are assumed to be consistent in the sense of (7) and sometimes we use phrase "$\mathbf{y}$ belongs to $\mathscr{C}(\mathbf{X} : \mathbf{V})$ with probability 1, or shortly w.p. 1". For consistency, see, e.g., [8].

There is a related decomposition, see, e.g., [35, Th. 8]: for any conformable matrices $\mathbf{A}$ and $\mathbf{B}$ we have

$$\mathscr{C}(\mathbf{A} : \mathbf{B}) = \mathscr{C}(\mathbf{A} : \mathbf{Q_A B}) \,, \text{ and thereby } \mathbf{P}_{(\mathbf{A}:\mathbf{B})} = \mathbf{P_A} + \mathbf{P}_{\mathbf{Q_A B}} \,. \tag{8}$$

Thus if $\mathbf{X} = (\mathbf{X}_1 : \mathbf{X}_2)$ and

$$\mathbf{M} = \mathbf{I}_n - \mathbf{P}_{(\mathbf{X}_1:\mathbf{X}_2)} = \mathbf{I}_n - (\mathbf{P}_{\mathbf{X}_2} + \mathbf{P}_{\mathbf{M}_2\mathbf{X}_1}) = \mathbf{M}_2\mathbf{Q}_{\mathbf{M}_2\mathbf{X}_1} \,,$$

and by (6),

$$\mathrm{r}(\mathbf{M}_2\mathbf{X}_1) = \mathrm{r}(\mathbf{X}_1) - \dim \mathscr{C}(\mathbf{X}_1) \cap \mathscr{C}(\mathbf{X}_2) \,.$$

As for the structure of the paper, in the next section we recall the fundamental BLUE equation which literally has a fundamental role for our considerations. In Section 3 we go through some mathematical properties of the so-called $\mathcal{W}$-matrices, i.e., the class $\mathcal{W}$, and in Section 4 we introduce some representations of the BLUEs. The use of the class $\mathcal{W}$ in the partitioned model is explored in Section 5. Sections 6 and 7 are devoted to particular properties of the perp-operator $\perp$ and for the linear sufficiency, respectively. In Section 8 we deal with the equality of the BLUEs under two models and in the last section we briefly discuss the model with new future observations. This paper is a review paper containing no essentially new results. However, we believe that our review provides a useful summary of the its area and thereby increases the insights and appreciation to the presented approach to best linear unbiased estimation.

2 The Fundamental BLUE Equation

In what follows, we frequently refer to the following Proposition, sometimes called the fundamental BLUE equation, see, e.g., Drygas [11, p. 55], Rao [38, p. 282], and Baksalary [2].

Proposition 2.1. *Consider the linear model* $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$. *Then* $\mathbf{Gy}$ *is the* BLUE *for* $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ *if and only if* $\mathbf{G}$ *satisfies the equation*

$$\mathbf{G}(\mathbf{X} : \mathbf{V}\mathbf{X}^{\perp}) = (\mathbf{X} : \mathbf{0}) \,. \tag{9}$$

The corresponding condition for $\mathbf{By}$ *to be the* BLUE *of an estimable* $\mathbf{K}\boldsymbol{\beta}$ *is*

$$\mathbf{B}(\mathbf{X} : \mathbf{V}\mathbf{X}^{\perp}) = (\mathbf{K} : \mathbf{0}) \,. \tag{10}$$

Proposition 2.1 offers an extremely handy tool to check whether a given estimator is a BLUE. Moreover, it provides a convenient way to introduce various representations for the BLUE. Equation (9) is always solvable for $\mathbf{G}$ while (10) is solvable for $\mathbf{B}$ if and only if $\mathbf{K}\boldsymbol{\beta}$ is estimable. The solutions are unique if and only if $\mathscr{C}(\mathbf{X} : \mathbf{V}\mathbf{X}^{\perp}) = \mathbb{R}^n$. As said, one choice for $\mathbf{X}^{\perp}$ is $\mathbf{M} = \mathbf{I}_n - \mathbf{P_X}$. We can define the set $\{\mathbf{P}_{\boldsymbol{\mu}|\mathscr{M}}\}$ as follows:

$$\mathbf{G} \in \{\mathbf{P}_{\boldsymbol{\mu}|\mathscr{M}}\} \iff \mathbf{G}(\mathbf{X} : \mathbf{VM}) = (\mathbf{X} : \mathbf{0}) \,.$$

If $\mathbf{G}_0$ is one particular solution for (9) then the general solution can be expressed as

$$\mathbf{G}_0 + \mathbf{E}(\mathbf{I}_n - \mathbf{P}_{(\mathbf{X}:\mathbf{V})}) \,,$$

where $\mathbf{E} \in \mathbb{R}^{n\times n}$ is free to vary.

We see at once that under a weakly singular linear model we have

$$\mathbf{P}_{\mathbf{X};\mathbf{V}^+}(\mathbf{X} : \mathbf{VM}) = \mathbf{X}(\mathbf{X}'\mathbf{V}^+\mathbf{X})^-\mathbf{X}'\mathbf{V}^+(\mathbf{X} : \mathbf{VM}) = (\mathbf{X} : \mathbf{0}) \,,$$

and it actually appears that

$$\mathbf{X}(\mathbf{X}'\mathbf{V}^{+}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{+}\mathbf{y} = \mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta}) \iff \mathscr{C}(\mathbf{X}) \subseteq \mathscr{C}(\mathbf{V});$$

see [44, Cor. 1.1] and [34, p. 286]. Moreover, if $\mathscr{C}(\mathbf{X}) \subseteq \mathscr{C}(\mathbf{V})$, then for any $\mathbf{V}^{-}$,

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

and $\mathbf{P}_{\mathbf{X};\mathbf{V}^{-}}\mathbf{y}$ is invariant for all generalized inverses involved assuming that the model is consistent.

Following Rao (1971) [40, Sec. 4] we can consider a matrix $\mathbf{W}$ defined as

$$\mathbf{W} = \mathbf{V} + \mathbf{X}\mathbf{U}\mathbf{U}'\mathbf{X}' \in \mathsf{NND}_n,$$

where NND_n stands for the set of nonnegative definite (symmetric) $n \times n$ matrices and $\mathbf{U} \in \mathbb{R}^{p\times s}$ (for some s) is such that $\mathscr{C}(\mathbf{W}) = \mathscr{C}(\mathbf{X} : \mathbf{V})$; then we may denote $\mathbf{W} \in \mathcal{W}_{\geq}$. A more general class is such where

$$\mathbf{W} = \mathbf{V} + \mathbf{X}\mathbf{T}\mathbf{X}' \in \mathbb{R}^{n\times n},$$

with $\mathbf{T}$ being any $p \times p$ matrix such that $\mathscr{C}(\mathbf{W}) = \mathscr{C}(\mathbf{X} : \mathbf{V})$. The set of such matrices $\mathbf{W}$ will be denoted as $\mathcal{W}$. Consider then the estimator

$$\mathbf{P}_{\mathbf{X};\mathbf{W}^{-}}\mathbf{y} = \mathbf{X}(\mathbf{X}'\mathbf{W}^{-}\mathbf{X})^{-}\mathbf{X}'\mathbf{W}^{-}\mathbf{y}, \quad \text{where } \mathbf{W} \in \mathcal{W}.$$

Now by Proposition 1.1, $\mathbf{X}'\mathbf{W}^{-}\mathbf{X}$ is invariant for the choice of the generalized inverse of $\mathbf{W}$ if and only if

$$\mathscr{C}(\mathbf{X}) \subseteq \mathscr{C}(\mathbf{W}') \quad \text{and} \quad \mathscr{C}(\mathbf{X}) \subseteq \mathscr{C}(\mathbf{W}). \tag{11}$$

It can be shown that

$$\mathscr{C}(\mathbf{W}') = \mathscr{C}(\mathbf{W}), \tag{12}$$

and thereby (11) holds, because assumption $\mathscr{C}(\mathbf{W}) = \mathscr{C}(\mathbf{X} : \mathbf{V})$ obviously implies $\mathscr{C}(\mathbf{X}) \subseteq \mathscr{C}(\mathbf{W})$. Using Proposition 1.1 we can conclude that $\mathbf{P}_{\mathbf{X};\mathbf{W}^{-}}\mathbf{y}$ is invariant for any choice of generalized inverses involved supposing that the model is consistent.

We further observe that $\mathbf{P}_{\mathbf{X};\mathbf{W}^{-}}(\mathbf{X} : \mathbf{V}\mathbf{M}) = (\mathbf{X} : \mathbf{0})$ can be written as

$$\mathbf{X}(\mathbf{X}'\mathbf{W}^{-}\mathbf{X})^{-}\mathbf{X}'\mathbf{W}^{-}(\mathbf{X} : \mathbf{W}\mathbf{M}) = (\mathbf{X} : \mathbf{0}), \tag{13}$$

where the second part $\mathbf{P}_{\mathbf{X};\mathbf{W}^{-}}\mathbf{W}\mathbf{M} = \mathbf{0}$ holds in light of (11). By Proposition 1.1, the first part of (13),

$$\mathbf{X}(\mathbf{X}'\mathbf{W}^{-}\mathbf{X})^{-}\mathbf{X}'\mathbf{W}^{-}\mathbf{X} = \mathbf{X} \tag{14}$$

holds if and only if $\mathscr{C}(\mathbf{X}') \subseteq \mathscr{C}[\mathbf{X}'(\mathbf{W}^{-})'\mathbf{X}]$, i.e.,

$$\mathrm{r}(\mathbf{X}) = \mathrm{r}[\mathbf{X}'(\mathbf{W}^{-})'\mathbf{X}] = \mathrm{r}(\mathbf{X}'\mathbf{W}^{-}\mathbf{X}). \tag{15}$$

The above equality holds in view of

$$\mathrm{r}(\mathbf{X}'\mathbf{W}^{-}\mathbf{X}) = \mathrm{r}[\mathbf{X}'\mathbf{W}^{-}(\mathbf{X} : \mathbf{W}\mathbf{M})] = \mathrm{r}(\mathbf{X}'\mathbf{W}^{-}\mathbf{W}) = \mathrm{r}(\mathbf{X}),$$

where we have used the assumption $\mathscr{C}(\mathbf{X}) \subseteq \mathscr{C}(\mathbf{W}')$. Thereby, under the model $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$,

$$\mathbf{X}(\mathbf{X}'\mathbf{W}^{-}\mathbf{X})^{-}\mathbf{X}'\mathbf{W}^{-}\mathbf{y} = \mathrm{BLUE}(\boldsymbol{\mu} \mid \mathscr{M}), \text{ i.e., } \mathbf{P}_{\mathbf{X};\mathbf{W}^{-}} \in \{\mathbf{P}_{\boldsymbol{\mu}|\mathscr{M}}\}. \tag{16}$$

There is one interesting approach to demonstrate the usefulness of matrix $\mathbf{W} \in \mathcal{W}_{\geq}$. Namely it is clear that

$$\mathbf{G}(\mathbf{X} : \mathbf{VM}) = (\mathbf{X} : \mathbf{0}) \iff \mathbf{G}(\mathbf{X} : \mathbf{WM}) = (\mathbf{X} : \mathbf{0}),$$

where $\mathbf{W} = \mathbf{V} + \mathbf{XUU'X'} \in \mathcal{W}_{\geq}$. Observing that $\mathscr{M}_{\mathbf{W}} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{W}\}$ is a weakly singular linear model we can conclude that

$$\mathbf{X}(\mathbf{X}'\mathbf{W}^{-}\mathbf{X})^{-}\mathbf{X}'\mathbf{W}^{-}\mathbf{y} = \mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta} \mid \mathscr{M}_{\mathbf{W}}) = \mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta} \mid \mathscr{M}). \tag{17}$$

For (17) see also [9, Th. 10.1.3].

After this longish Introduction to Basic Tools and the Fundamental BLUE Equation, we will focus in more details on the properties of matrix which is of type $\mathbf{W}$ and its usage in BLUE-related matters. It appears to be surprisingly useful and powerful tool when dealing with linear statistical models.

3 Properties of the Class $\mathcal{W}$

For a given linear model $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$, let the set $\mathcal{W}$ of $n \times n$ matrices be defined as

$$\mathcal{W} = \left\{\mathbf{W} \in \mathbb{R}^{n\times n} : \mathbf{W} = \mathbf{V} + \mathbf{XTX}',\ \mathscr{C}(\mathbf{W}) = \mathscr{C}(\mathbf{X} : \mathbf{V})\right\}. \tag{18}$$

In (18), $\mathbf{T}$ can be any $p \times p$ matrix as long as $\mathscr{C}(\mathbf{W}) = \mathscr{C}(\mathbf{X} : \mathbf{V})$ is satisfied. It is clear that we can always choose $\mathbf{T} = \alpha^2\mathbf{I}_n$, where α is an arbitrary nonzero scalar. Moreover, $\mathbf{V} \in \mathcal{W}$ if and only if $\mathscr{C}(\mathbf{X}) \subseteq \mathscr{C}(\mathbf{V})$. If there is a need to emphasize that there is a particular model $\mathscr{M}$, say, under consideration we will use notation $\mathcal{W}(\mathscr{M})$. Sometimes we use the phrases like "$\mathbf{A}$ is a $\mathcal{W}$-matrix" indicating that $\mathbf{A} \in \mathcal{W}$.

Choosing $\mathbf{T}$ in (18) nonnegative definite, i.e., putting $\mathbf{T} = \mathbf{UU}'$ (for some $\mathbf{U}$), we get the set $\mathcal{W}_{\geq}$ of nonnegative definite matrices defined as

$$\mathcal{W}_{\geq} = \left\{\mathbf{W} \in \mathrm{NND}_n : \mathbf{W} = \mathbf{V} + \mathbf{XUU'X}',\ \mathscr{C}(\mathbf{W}) = \mathscr{C}(\mathbf{X} : \mathbf{V})\right\}. \tag{19}$$

In (19), $\mathbf{U}$ can be any matrix comprising p rows so that $\mathscr{C}(\mathbf{W}) = \mathscr{C}(\mathbf{X} : \mathbf{V})$ is satisfied. Using $\mathcal{W}_{\geq}$ instead of $\mathcal{W}$ some considerations can become simpler, as can be expected.

Proposition 3.1 collects together some important properties of the class $\mathcal{W}$; see, e.g., [35, Prop. 12.1].

Proposition 3.1. *Let $\mathbf{V}$ be an $n \times n$ nonnegative definite matrix, let $\mathbf{X}$ be an $n \times p$ matrix, and define $\mathbf{W}$ as $\mathbf{W} = \mathbf{V} + \mathbf{XTX}'$, where $\mathbf{T}$ is a $p \times p$ matrix. Then the following statements are equivalent:*

(a) $\mathscr{C}(\mathbf{X} : \mathbf{V}) = \mathscr{C}(\mathbf{W})$,

(b) $\mathscr{C}(\mathbf{X}) \subseteq \mathscr{C}(\mathbf{W})$,

(c) $\mathrm{r}(\mathbf{X} : \mathbf{V}) = \mathrm{r}(\mathbf{W})$,

(d) $\mathbf{X}'\mathbf{W}^{-}\mathbf{X}$ *is invariant for any choice of* $\mathbf{W}^{-}$,

(e) $\mathscr{C}(\mathbf{X}'\mathbf{W}^{-}\mathbf{X})$ *is invariant for any choice of* $\mathbf{W}^{-}$,

(f) $\mathscr{C}(\mathbf{X}'\mathbf{W}^{-}\mathbf{X}) = \mathscr{C}(\mathbf{X}')$ *for any choice of* $\mathbf{W}^{-}$,

(g) $\mathrm{r}(\mathbf{X}'\mathbf{W}^{-}\mathbf{X}) = \mathrm{r}(\mathbf{X})$ *irrespective of the choice of* $\mathbf{W}^{-}$,

(h) $\mathrm{r}(\mathbf{X}'\mathbf{W}^{-}\mathbf{X})$ *is invariant with respect to the choice of* $\mathbf{W}^{-}$,

(i) $\mathbf{X}(\mathbf{X}'\mathbf{W}^-\mathbf{X})^-\mathbf{X}'\mathbf{W}^-\mathbf{X} = \mathbf{X}$ *for any choices of* $\mathbf{W}^-$ *and* $(\mathbf{X}'\mathbf{W}^-\mathbf{X})^-$.

Moreover, each of these statements is equivalent to

(a') $\mathscr{C}(\mathbf{X} : \mathbf{V}) = \mathscr{C}(\mathbf{W}')$,

and hence to the statements (b')–(i') *obtained from* (b)–(i)*, by setting* $\mathbf{W}'$ *in place of* $\mathbf{W}$.

Shortly said, given the model $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$, $\mathbf{W} \in \mathcal{W}(\mathscr{M})$ if and only if any of the conditions in Proposition 3.1 holds. Observe that the invariance properties in (f)–(i) concern also the choice of $\mathbf{W} \in \mathcal{W}$; not only its generalized inverse. Actually, we will return to this property in due course.

As references to Proposition 3.1, in addition to [40, Sec. 4], we may mention, e.g., [6, Th. 1], [7, Th. 2], [5, Th. 2], [14, p. 468], and [35, Sec. 12.3].

Notice that the equivalence of (g) and (h) of Proposition 3.1, is the same as that between (14) and (15). Moreover, we can conclude that the statement

$$\mathbf{P}_{\mathbf{X};\mathbf{W}^-} \in \{\mathbf{P}_{\mu|\mathscr{M}}\} \quad \text{for any choices of } \mathbf{W}^- \text{ and } (\mathbf{X}'\mathbf{W}^-\mathbf{X})^- \tag{20}$$

is equivalent to the conditions in Proposition 3.1.

Let's take a quick look at some some developments of the equivalence of the statements of Proposition 3.1; see, in particular, Baksalary & Mathew (1990) [5, Sec. 3]. Consider the model $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$, and the following generalized normal equation:

$$\mathbf{X}'\mathbf{A}\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{A}\mathbf{y}, \tag{21}$$

where $\mathbf{A}$ is a given $n \times n$ matrix. If $\mathbf{A}$ is nonnegative definite (and symmetric) then (21) has a solution for $\boldsymbol{\beta}$ for every $\mathbf{y}$ and the solution minimizes

$$(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{A}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2_{\mathbf{A}}\,.$$

Rao (1971) [40, p. 372] pointed out that we can consider a more general class of matrices $\mathbf{A}$ by allowing $\mathbf{A}$ to be any matrix for which $\boldsymbol{\beta}$ is solvable from (21). Assuming that the model $\mathscr{M}$ is consistent, i.e., $\mathbf{y} \in \mathscr{C}(\mathbf{X} : \mathbf{V})$, we observe that (21) is solvable for any $\mathbf{y} \in \mathscr{C}(\mathbf{X} : \mathbf{V})$ if and only if $\mathbf{X}'\mathbf{A}(\mathbf{X} : \mathbf{V})\mathbf{t} \in \mathscr{C}(\mathbf{X}'\mathbf{A}\mathbf{X})$ for all $\mathbf{t} \in \mathbb{R}^{n+p}$, i.e.,

$$\mathscr{C}(\mathbf{X}'\mathbf{A}\mathbf{V}) \subseteq \mathscr{C}(\mathbf{X}'\mathbf{A}\mathbf{X})\,. \tag{22}$$

Rao [40, Th. 4.2] showed that if (22) holds, then for any solution $\boldsymbol{\beta}_0$ of (21) the estimator $\mathbf{X}\boldsymbol{\beta}_0$ is the BLUE of $\mathbf{X}\boldsymbol{\beta}$ if and only if $\mathbf{A}$ is of the form

$$\mathbf{A} = (\mathbf{V} + \mathbf{X}\mathbf{T}\mathbf{X}')^- + \mathbf{J}\,, \tag{23}$$

and satisfies the equality $\mathrm{r}(\mathbf{X}'\mathbf{A}\mathbf{X}) = \mathrm{r}(\mathbf{X})$, with $\mathbf{T}$ and $\mathbf{J}$ being arbitrary matrices such that

$$\mathscr{C}(\mathbf{X} : \mathbf{V}) = \mathscr{C}(\mathbf{V} + \mathbf{X}\mathbf{T}\mathbf{X}') = \mathscr{C}(\mathbf{V} + \mathbf{X}\mathbf{T}'\mathbf{X}')\,, \tag{24}$$

and $\mathbf{X}'\mathbf{J}(\mathbf{X} : \mathbf{V}) = (\mathbf{0} : \mathbf{0})$. Baksalary & Puntanen (1989) [6, Th. 1] proved that the condition (24) may be simplified because

$$\mathscr{C}(\mathbf{X} : \mathbf{V}) = \mathscr{C}(\mathbf{V} + \mathbf{X}\mathbf{T}\mathbf{X}') \iff \mathscr{C}(\mathbf{X} : \mathbf{V}) = \mathscr{C}(\mathbf{V} + \mathbf{X}\mathbf{T}'\mathbf{X}')\,,$$

c.f. (12), and that, under (24) the condition $\mathrm{r}(\mathbf{X}'\mathbf{A}\mathbf{X}) = \mathrm{r}(\mathbf{X})$ is redundant since for every $\mathbf{A}$ of the form (23):

$$\mathscr{C}(\mathbf{X} : \mathbf{V}) = \mathscr{C}(\mathbf{V} + \mathbf{X}\mathbf{T}\mathbf{X}') \implies \mathrm{r}(\mathbf{X}'\mathbf{A}\mathbf{X}) = \mathrm{r}(\mathbf{X})\,. \tag{25}$$

Baksalary et al. [7, Th. 2], showed that the implication (25) may be reversed, in the sense that if

$$\mathrm{r}[\mathbf{X}'(\mathbf{V} + \mathbf{X}\mathbf{T}\mathbf{X}')^{-}\mathbf{X}] = \mathrm{r}(\mathbf{X}) \quad \text{for every } (\mathbf{V} + \mathbf{X}\mathbf{T}\mathbf{X}')^{-} \tag{26}$$

then $\mathscr{C}[\mathbf{X}'(\mathbf{V}+\mathbf{X}\mathbf{T}\mathbf{X}')^{-}\mathbf{X}] = \mathscr{C}(\mathbf{X} : \mathbf{V})$; this confirms the equivalence of (f) and (g) in Proposition 3.1. Moreover, they raised the question whether it is possible to relax the condition (26) by requiring only that

$$\mathrm{r}[\mathbf{X}'(\mathbf{V} + \mathbf{X}\mathbf{T}\mathbf{X}')^{-}\mathbf{X}] \quad \text{is invariant for every } (\mathbf{V} + \mathbf{X}\mathbf{T}\mathbf{X}')^{-}. \tag{27}$$

Baksalary & Mathew [5, Th. 2] showed that the answer is positive to this question; thus (g) and (d) in Proposition 3.1 are equivalent.

4 Representations of the BLUE

In this section we present an important matrix decomposition in Proposition 4.1 and some its consequences. Before it, however, a few words about the matrix $\mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}$ which we denote as

$$\dot{\mathbf{M}} = \mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}\,.$$

The matrix $\dot{\mathbf{M}}$ is not necessarily unique for any $(\mathbf{M}\mathbf{V}\mathbf{M})^{-}$; it is unique if and only if $\mathrm{r}(\mathbf{X} : \mathbf{V}) = n$. However, we always have

$$\mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{+}\mathbf{M} = (\mathbf{M}\mathbf{V}\mathbf{M})^{+}\mathbf{M} = \mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{+} = (\mathbf{M}\mathbf{V}\mathbf{M})^{+}.$$

In particular, for a positive definite $\mathbf{V}$ we have, for any $(\mathbf{M}\mathbf{V}\mathbf{M})^{-}$,

$$\begin{aligned}\mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M} &= \mathbf{V}^{-1/2}\mathbf{P}_{\mathbf{V}^{1/2}\mathbf{M}}\mathbf{V}^{-1/2} \\ &= \mathbf{V}^{-1/2}(\mathbf{I}_n - \mathbf{P}_{(\mathbf{V}^{1/2}\mathbf{M})^{\perp}})\mathbf{V}^{-1/2} \\ &= \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1},\end{aligned}$$

where we have used the obvious fact $\mathscr{C}(\mathbf{V}^{1/2}\mathbf{M})^{\perp} = \mathscr{C}(\mathbf{V}^{-1/2}\mathbf{X})$.

Proposition 4.1. *Consider the linear model* $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$. *Let* $\mathbf{T}$ *be any* $p \times p$ *matrix such that the matrix* $\mathbf{W} = \mathbf{V} + \mathbf{X}\mathbf{T}\mathbf{X}'$ *satisfies the condition* $\mathscr{C}(\mathbf{W}) = \mathscr{C}(\mathbf{X} : \mathbf{V})$, *i.e.,* $\mathbf{W} \in \mathcal{W}(\mathscr{M})$, *and denote* $\dot{\mathbf{M}} = \mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}$. *Then*

(a) $\mathbf{P}_{\mathbf{W}}\mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}\mathbf{P}_{\mathbf{W}} = \mathbf{W}^{+} - \mathbf{W}^{+}\mathbf{X}(\mathbf{X}'\mathbf{W}^{-}\mathbf{X})^{-}\mathbf{X}'\mathbf{W}^{+}$,

(b) $\mathbf{P}_{\mathbf{W}}\mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}\mathbf{P}_{\mathbf{W}} = (\mathbf{M}\mathbf{V}\mathbf{M})^{+} = \mathbf{P}_{\mathbf{W}}\dot{\mathbf{M}}\mathbf{P}_{\mathbf{W}}\,$,

(c) $\mathbf{P}_{\mathbf{X};\mathbf{W}^{+}} = \mathbf{X}(\mathbf{X}'\mathbf{W}^{-}\mathbf{X})^{-}\mathbf{X}'\mathbf{W}^{+} = \mathbf{P}_{\mathbf{W}} - \mathbf{V}\mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}\mathbf{P}_{\mathbf{W}}\,$,

(d) $\mathbf{P}_{\mathbf{X};\mathbf{W}^{+}} \in \{\mathbf{P}_{\mu|\mathscr{M}}\}\,$.

For the proof of (a), see [35, Prop. 15.2] and [22, Cor. 2.2]. Some related considerations (in full rank case) appear also in [28, pp. 415–416] and [26, pp. 323–324].

We observe that in light of (8) we have

$$\mathbf{P}_{\mathbf{W}} = \mathbf{P}_{\mathbf{X}} + \mathbf{P}_{\mathbf{M}\mathbf{V}} = \mathbf{H} + \mathbf{P}_{\mathbf{M}\mathbf{V}\mathbf{M}}\,,$$

which implies (b) of Proposition 4.1. Premultiplying (a) by $\mathbf{W}$ and using $\mathscr{C}(\mathbf{X}) \subseteq \mathscr{C}(\mathbf{W}') = \mathscr{C}(\mathbf{W})$ gives

$$\begin{aligned} \mathbf{X}(\mathbf{X}'\mathbf{W}^{-}\mathbf{X})^{-}\mathbf{X}'\mathbf{W}^{+} &= \mathbf{P_W} - \mathbf{VM}(\mathbf{MVM})^{-}\mathbf{MP_W} \\ &= (\mathbf{I}_n - \mathbf{V}\dot{\mathbf{M}})\mathbf{P_W}\,. \end{aligned} \tag{28}$$

From (28) we immediately confirm that $\mathbf{X}(\mathbf{X}'\mathbf{W}^{-}\mathbf{X})^{-}\mathbf{X}'\mathbf{W}^{+}$ is invariant with respect to the choice of $\mathbf{W} \in \mathcal{W}$ as was pointed out in the context of Proposition 3.1.

Premultiplying (28) by $\mathbf{H} = \mathbf{P_X}$ gives further expressions:

$$\begin{aligned} \mathbf{P}_{\mathbf{X};\mathbf{W}^+} &= \mathbf{X}(\mathbf{X}'\mathbf{W}^{-}\mathbf{X})^{-}\mathbf{X}'\mathbf{W}^{+} \\ &= \mathbf{P_W} - \mathbf{VM}(\mathbf{MVM})^{-}\mathbf{MP_W} \\ &= \mathbf{P_W} - \mathbf{V}(\mathbf{MVM})^{+} \\ &= \mathbf{H} - \mathbf{HVM}(\mathbf{MVM})^{-}\mathbf{MP_W} \\ &= \mathbf{H} - \mathbf{HVM}(\mathbf{MVM})^{+}\mathbf{M}\,. \end{aligned} \tag{29}$$

It is worth emphasizing that in (28) and (29) we use the Moore–Penrose inverse $\mathbf{A}^{+}$, wherever it is marked while the notation $\mathbf{A}^{-}$ means that we can use any generalized inverse.

As we have already in (16) observed we have under $\mathscr{M}$

$$\mathbf{P}_{\mathbf{X};\mathbf{W}^-}\mathbf{y} = \mathbf{X}(\mathbf{X}'\mathbf{W}^{-}\mathbf{X})^{-}\mathbf{X}'\mathbf{W}^{-}\mathbf{y} = \mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta}) = \tilde{\boldsymbol{\mu}}\,.$$

Notice that $\mathbf{X}(\mathbf{X}'\mathbf{W}^{-}\mathbf{X})^{-}\mathbf{X}'$ is invariant with respect to the choice of generalized inverses involved and

$$\mathbf{P}_{\mathbf{X};\mathbf{W}^+} = \mathbf{X}(\mathbf{X}'\mathbf{W}^{+}\mathbf{X})^{+}\mathbf{X}'\mathbf{W}^{+} = \mathbf{X}(\mathbf{X}'\mathbf{W}^{-}\mathbf{X})^{-}\mathbf{X}'\mathbf{W}^{+}$$

for any choice of $\mathbf{W}^{-}$ and $(\mathbf{X}'\mathbf{W}^{-}\mathbf{X})^{-}$.

From (29) we can conclude that under the consistent model $\mathscr{M}$, i.e., assuming that $\mathbf{y} \in \mathscr{C}(\mathbf{W}) = \mathscr{C}(\mathbf{X} : \mathbf{V})$,

$$\tilde{\boldsymbol{\mu}} = \mathbf{X}(\mathbf{X}'\mathbf{W}^{-}\mathbf{X})^{-}\mathbf{X}'\mathbf{W}^{-}\mathbf{y} = \mathbf{y} - \mathbf{VM}(\mathbf{MVM})^{-}\mathbf{My} = \mathbf{Hy} - \mathbf{HVM}(\mathbf{MVM})^{-}\mathbf{My},$$

and

$$\mathbf{y} - \tilde{\boldsymbol{\mu}} = \mathbf{VM}(\mathbf{MVM})^{-}\mathbf{My} = \tilde{\boldsymbol{\varepsilon}} = \mathrm{BLUE}\text{'s residual.} \tag{30}$$

The covariance matrix of the $\tilde{\boldsymbol{\mu}} = \mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta})$ can be expressed as

$$\begin{aligned} \mathrm{cov}(\tilde{\boldsymbol{\mu}}) &= \mathbf{HVH} - \mathbf{HVM}(\mathbf{MVM})^{-}\mathbf{MVH} \\ &= \mathrm{cov}(\mathbf{Hy}) - \mathbf{HVM}(\mathbf{MVM})^{-}\mathbf{MVH}\,, \end{aligned}$$

as well as $\mathrm{cov}(\tilde{\boldsymbol{\mu}}) = \mathbf{V} - \mathbf{VM}(\mathbf{MVM})^{-}\mathbf{MV}$. Notice that

$$\mathrm{cov}(\hat{\boldsymbol{\mu}} - \tilde{\boldsymbol{\mu}}) = \mathrm{cov}(\hat{\boldsymbol{\mu}}) - \mathrm{cov}(\tilde{\boldsymbol{\mu}}) = \mathbf{HVM}(\mathbf{MVM})^{-}\mathbf{MVH}\,.$$

Postmultiplying (28) by $\mathbf{W}$ yields

$$\mathbf{X}(\mathbf{X}'\mathbf{W}^{-}\mathbf{X})^{-}\mathbf{X}' = \mathbf{W} - \mathbf{VM}(\mathbf{MVM})^{-}\mathbf{MV}$$

and thereby the BLUE's covariance matrix has a representation

$$\mathrm{cov}(\tilde{\boldsymbol{\mu}}) = \mathbf{X}(\mathbf{X}'\mathbf{W}^{-}\mathbf{X})^{-}\mathbf{X}' - \mathbf{XTX}'. \tag{31}$$

The form (31) was first expressed, using $\mathbf{T} = \alpha^2\mathbf{I}_n$, by Rao (1971) [40, p. 382] and Rao & Mitra (1971) [41, p. 289]. Rao [40, p. 384–385] pointed out the use of $\mathbf{W} \in \mathcal{W}$ with condition $\mathscr{C}(\mathbf{W}') = \mathscr{C}(\mathbf{X} : \mathbf{V})$, which, as stated earlier, is actually not needed. For further references regarding the $\mathrm{cov}(\tilde{\boldsymbol{\mu}})$, see, e.g., [7] and [22, 23].

Remark 1. The referee of our paper interestingly commented as follows:

For singular models where we use the matrix $\mathbf{W}$ to get representations for the BLUE of $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$, the covariance matrix is a somewhat lengthy matrix expression, not as elegant as the expression given above for a nonsingular $\mathbf{V}$. For singular models, are there some choices of $\mathbf{W}$ that gives some elegant simplifications for the covariance matrix of the BLUE of $\boldsymbol{\mu}$?

Indeed, for a positive definite $\mathbf{V}$ we can choose $\mathbf{W} = \mathbf{V}$ and thus (31) gives the well-known formula $\operatorname{cov}(\tilde{\boldsymbol{\mu}}) = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'$. Similarly, for a weakly singular linear model, i.e., when $\mathscr{C}(\mathbf{X}) \subseteq \mathscr{C}(\mathbf{V})$, we can again choose $\mathbf{W} = \mathbf{V}$ and obtain

$$\operatorname{cov}(\tilde{\boldsymbol{\mu}}) = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-}\mathbf{X})^{-}\mathbf{X}'.$$

But for the question of further choices of $\mathbf{W}$ yielding elegant simplifications we are afraid that we must raise our hands and postpone it for further research. However, as pointed out by Rao (1978) [36], missing the role of the matrix $\mathbf{T}$ in (31) can yield wrong results. Rao points out that the choice of $\mathbf{T} = \alpha^2\mathbf{I}_p$ has some advantages, like even if $\mathbf{V}$ is singular, the matrix $\mathbf{W} = \mathbf{V} + \alpha^2\mathbf{X}\mathbf{X}'$ may be positive definite.

There is a related curious problem: suppose $\mathbf{W}$ is defined as

$$\mathbf{W} = \mathbf{V} + \mathbf{X}\mathbf{U}\mathbf{U}'\mathbf{X}' = (\mathbf{V}^{1/2} : \mathbf{X}\mathbf{U})(\mathbf{V}^{1/2} : \mathbf{X}\mathbf{U})',$$

so that $\mathscr{C}(\mathbf{W}) = \mathscr{C}(\mathbf{X} : \mathbf{V})$. What is the choice for $\mathbf{U}$ making $\mathbf{W}^{-}$ to be also a generalized inverse of $\mathbf{V}$, i.e.,

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

Groß [12] showed that one such choice is $\mathbf{U} = \mathbf{X}^{+}(\mathbf{I}_n - \mathbf{P}_{\mathbf{V}})$. For related discussion, see also [32].

The ordinary, unweighted sum of squares of errors SSE is defined as

$$\operatorname{SSE}(\mathbf{I}) = \min_{\boldsymbol{\beta}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2 = \mathbf{y}'\mathbf{M}\mathbf{y}\,,$$

while the weighted SSE, when $\mathbf{V}$ is positive definite, is

$$\begin{aligned}\operatorname{SSE}(\mathbf{V}) &= \min_{\boldsymbol{\beta}} \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2_{\mathbf{V}^{-1}} = \|\mathbf{y} - \mathbf{P}_{\mathbf{X};\mathbf{V}^{-1}}\mathbf{y}\|^2_{\mathbf{V}^{-1}} \\ &= \mathbf{y}'[\mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}]\mathbf{y} \\ &= \mathbf{y}'\mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}\mathbf{y} = \mathbf{y}'\dot{\mathbf{M}}\mathbf{y}\,.\end{aligned}$$

In the general case, the weighted SSE can be defined as

$$\operatorname{SSE}(\mathbf{W}) = (\mathbf{y} - \tilde{\boldsymbol{\mu}})'\mathbf{W}^{-}(\mathbf{y} - \tilde{\boldsymbol{\mu}})\,,$$

where $\mathbf{W} \in \mathcal{W}$. Then, recalling that by (30), the BLUE's residual is $\tilde{\boldsymbol{\varepsilon}} = \mathbf{y} - \tilde{\boldsymbol{\mu}} = \mathbf{V}\dot{\mathbf{M}}\mathbf{y}$, we observe the following:

$$\begin{aligned}\operatorname{SSE}(\mathbf{W}) &= \tilde{\boldsymbol{\varepsilon}}'\mathbf{W}^{-}\tilde{\boldsymbol{\varepsilon}} = \tilde{\boldsymbol{\varepsilon}}'\mathbf{V}^{-}\tilde{\boldsymbol{\varepsilon}} = \mathbf{y}'\dot{\mathbf{M}}\mathbf{y} \\ &= \mathbf{y}'[\mathbf{W}^{-} - \mathbf{W}^{-}\mathbf{X}(\mathbf{X}'\mathbf{W}^{-}\mathbf{X})^{-}\mathbf{X}'\mathbf{W}^{-}]\mathbf{y}.\end{aligned}$$

It can be further shown that under $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{V}\}$, $\operatorname{SSE}(\mathbf{W})$ provides an unbiased estimator of σ^2:

$$\operatorname{E}(\mathbf{y}'\dot{\mathbf{M}}\mathbf{y}/f) = \sigma^2, \quad \text{where } f = \operatorname{r}(\mathbf{V}\mathbf{M})\,.$$

Remark 2. Baksalary et al. (1990) [7, Th. 3] considered the model $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$ and proved that if $\mathbf{W} = \mathbf{V} + \mathbf{XTX}' \in \mathcal{W}(\mathscr{M})$ then the equality

$$\mathbf{W} = \mathbf{VB}(\mathbf{B}'\mathbf{VB})^{-}\mathbf{B}'\mathbf{V} + \mathbf{X}(\mathbf{X}'\mathbf{W}^{-}\mathbf{X})^{-}\mathbf{X}' \tag{32}$$

holds for a matrix $\mathbf{B}$ if and only if

$$\mathscr{C}(\mathbf{VW}^{-}\mathbf{X}) \subseteq \mathscr{C}(\mathbf{B})^{\perp} \text{ and } \mathscr{C}(\mathbf{VM}) \subseteq \mathscr{C}(\mathbf{VB})\,. \tag{33}$$

It is clear that the choice of $\mathbf{B} = \mathbf{M}$ satisfies (33) and thereby also (32) holds for $\mathbf{B} = \mathbf{M}$. Postmultiplying (32) by $\mathbf{W}^{+}$ in that situation gives (c) of Proposition 4.1.

Remark 3. Wang & Liski (1998) [43, p. 45] introduce an interesting matrix inequality by considering estimator $\mathbf{Ay}$ which is unbiased for $\mathbf{BX}\boldsymbol{\beta}$ under $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$, i.e., $\mathbf{AX} = \mathbf{BX}$. Let $\mathbf{W} = \mathbf{V} + \mathbf{XUU}'\mathbf{X}' \in \mathcal{W}_{\geq}(\mathscr{M})$. Then $\mathbf{BP}_{\mathbf{X};\mathbf{W}^{+}}\mathbf{y} = \mathrm{BLUE}(\mathbf{BX}\boldsymbol{\beta})$ and for any $\mathbf{A}$ and $\mathbf{B}$ satisfying $\mathbf{AX} = \mathbf{BX}$,

$$\mathrm{cov}(\mathbf{BP}_{\mathbf{X};\mathbf{W}^{+}}\mathbf{y}) \leq_{\mathsf{L}} \mathrm{cov}(\mathbf{Ay})\,,$$

i.e.,

$$\mathbf{B}\big[\mathbf{X}(\mathbf{X}'\mathbf{W}^{-}\mathbf{X})^{-}\mathbf{X}' - \mathbf{XUU}'\mathbf{X}'\big]\mathbf{B}' \leq_{\mathsf{L}} \mathbf{AVA}',$$

which is further equivalent to

$$\mathbf{B}\big[\mathbf{X}(\mathbf{X}'\mathbf{W}^{-}\mathbf{X})^{-}\mathbf{X}'\big]\mathbf{B}' \leq_{\mathsf{L}} \mathbf{AWA}'. \tag{34}$$

Equality appears in (34) if and only if $\mathbf{Ay}$ is the $\mathrm{BLUE}(\mathbf{BX}\boldsymbol{\beta})$ which by Proposition 2.1 happens if and only if $\mathscr{C}(\mathbf{VA}') \subseteq \mathscr{C}(\mathbf{X})$.

Another interesting application of the $\mathcal{W}_{\geq}$-matrix is given by [27] who considered the upper bound for

$$\delta = \mathrm{trace}[\mathrm{cov}(\hat{\boldsymbol{\mu}} \mid \mathscr{M}) - \mathrm{cov}(\tilde{\boldsymbol{\mu}} \mid \mathscr{M})] = \mathrm{trace}[\mathbf{HVM}(\mathbf{MVM})^{-}\mathbf{MVH}]\,,$$

where $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$. Without going into more details we may only mention that they based their proof by noting that

$$\delta = \mathrm{trace}[\mathrm{cov}(\hat{\boldsymbol{\mu}} \mid \mathscr{M}_{\mathbf{W}}) - \mathrm{cov}(\tilde{\boldsymbol{\mu}} \mid \mathscr{M}_{\mathbf{W}})]\,,$$

where $\mathscr{M}_{\mathbf{W}} = \{\mathbf{y},\ \mathbf{X}\boldsymbol{\beta},\ \mathbf{W}\}$, with $\mathbf{W} = \mathbf{V} + \alpha^2\mathbf{XX}' \in \mathcal{W}_{\geq}(\mathscr{M})$ and could generalize the result of Rao (1985) [39] given for a positive definite $\mathbf{V}$.

5 Partitioned Linear Model

Consider then the estimation of $\boldsymbol{\mu}_1 = \mathbf{X}_1\boldsymbol{\beta}_1$ under the partitioned model $\mathscr{M}_{12} = \{\mathbf{y}, \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2, \mathbf{V}\}$ assuming that $\boldsymbol{\mu}_1$ is estimable which is well known to hold if and only if

$$\mathscr{C}(\mathbf{X}_1) \cap \mathscr{C}(\mathbf{X}_2) = \{\mathbf{0}\}\,, \quad \text{i.e.,} \quad \mathrm{r}(\mathbf{M}_2\mathbf{X}_1) = \mathrm{r}(\mathbf{X}_1)\,.$$

Let us denote the small models as $\mathscr{M}_i = \{\mathbf{y}, \mathbf{X}_i\boldsymbol{\beta}_i, \mathbf{V}\}, i = 1, 2$. Corresponding to (19), $\mathbf{W}_i \in \mathcal{W}_{\geq}(\mathscr{M}_i)$ if there exists a matrix $\mathbf{L}_i$ such that

$$\mathbf{W}_i = \mathbf{V} + \mathbf{X}_i\mathbf{L}_i\mathbf{L}_i'\mathbf{X}_i'\,, \quad \mathscr{C}(\mathbf{W}_i) = \mathscr{C}(\mathbf{X}_i : \mathbf{V})\,, \quad i = 1, 2\,. \tag{35}$$

Premultiplying $\mathscr{M}_{12}$ by $\mathbf{M}_2 = \mathbf{I}_n - \mathbf{P}_{\mathbf{X}_2}$ yields the reduced model

$$\mathscr{M}_{12\cdot 2} = \{\mathbf{M}_2\mathbf{y},\ \mathbf{M}_2\mathbf{X}_1\boldsymbol{\beta}_1,\ \mathbf{M}_2\mathbf{VM}_2\}\,,$$

which is a special case of the transformed model $\mathscr{T} = \{\mathbf{Fy}, \mathbf{FX}\beta, \mathbf{FVF}'\}$, where $\mathbf{F} \in \mathbb{R}^{f \times n}$. In view of the Frisch–Waugh–Lovell theorem, see, e.g., [13, Sec. 6], the BLUEs of $\boldsymbol{\theta}_1 = \mathbf{M}_2\mathbf{X}_1\boldsymbol{\beta}_1$ under $\mathscr{M}_{12}$ and $\mathscr{M}_{12\cdot 2}$ coincide. It is noteworthy that $\boldsymbol{\theta}_1$ is estimable under $\mathscr{M}_{12}$ as well as under $\mathscr{M}_{12\cdot 2}$. An explicit expression for the BLUE of $\boldsymbol{\theta}_1$ under $\mathscr{M}_{12\cdot 2}$ can be obtained from

$$\mathrm{BLUE}(\boldsymbol{\theta}_1 \mid \mathscr{M}_{12\cdot 2}) = \mathrm{BLUE}(\boldsymbol{\theta}_1 \mid \mathscr{M}_{12}) = \mathbf{P}_{\mathbf{M}_2\mathbf{X}_1;\mathbf{W}_{rm}^-}\mathbf{M}_2\mathbf{y},$$

where

$$\mathbf{P}_{\mathbf{M}_2\mathbf{X}_1;\mathbf{W}_{rm}^-} = \mathbf{M}_2\mathbf{X}_1(\mathbf{X}_1'\mathbf{M}_2\mathbf{W}_{rm}^-\mathbf{M}_2\mathbf{X}_1)^-\mathbf{X}_1'\mathbf{M}_2\mathbf{W}_{rm}^- \in \{\mathbf{P}_{\boldsymbol{\theta}_1|\mathscr{M}_{12\cdot 2}}\},$$

and $\mathbf{W}_{rm}$ is an arbitrary $\mathcal{W}$-matrix in $\mathscr{M}_{12\cdot 2}$, i.e., $\mathbf{W}_{rm} \in \mathcal{W}(\mathscr{M}_{12\cdot 2})$. Notice that

$$\mathbf{P}_{\mathbf{M}_2\mathbf{X}_1;\mathbf{W}_{rm}^-}\mathbf{M}_2 \in \{\mathbf{P}_{\boldsymbol{\theta}_1|\mathscr{M}_{12}}\}.$$

Clearly any matrix of the form $\mathbf{M}_2(\mathbf{V} + \mathbf{X}_1\mathbf{K}_1\mathbf{K}_1'\mathbf{X}_1')\mathbf{M}_2$ satisfying

$$\mathscr{C}[\mathbf{M}_2(\mathbf{V} : \mathbf{X}_1\mathbf{K}_1)] = \mathscr{C}[\mathbf{M}_2(\mathbf{V} : \mathbf{X}_1)] = \mathscr{C}(\mathbf{M}_2\mathbf{W}_1), \tag{36}$$

is a $\mathcal{W}_\geq$-matrix in $\mathscr{M}_{12\cdot 2}$. Putting $\mathbf{K}_1 = \mathbf{L}_1$ as in (35) we can choose

$$\mathbf{W}_{rm} = \mathbf{M}_2\mathbf{W}_1\mathbf{M}_2 \in \mathcal{W}_\geq(\mathscr{M}_{12\cdot 2}).$$

Thus the BLUE of $\boldsymbol{\theta}_1 = \mathbf{M}_2\mathbf{X}_1\boldsymbol{\beta}_1$ under $\mathscr{M}_{12\cdot 2}$ can be expressed as

$$\mathrm{BLUE}(\boldsymbol{\theta}_1 \mid \mathscr{M}_{12\cdot 2}) = \mathbf{M}_2\mathbf{X}_1(\mathbf{X}_1'\dot{\mathbf{M}}_2\mathbf{X}_1)^-\mathbf{X}_1'\dot{\mathbf{M}}_2\mathbf{y}, \tag{37}$$

where

$$\dot{\mathbf{M}}_2 = \mathbf{M}_2(\mathbf{M}_2\mathbf{W}_1\mathbf{M}_2)^-\mathbf{M}_2.$$

Notice that by Proposition 4.1 the matrix

$$\mathbf{P}_{\mathbf{M}_2\mathbf{X}_1;\mathbf{W}_{rm}^+} = \mathbf{M}_2\mathbf{X}_1(\mathbf{X}_1'\mathbf{M}_2\mathbf{W}_{rm}^-\mathbf{M}_2\mathbf{X}_1)^-\mathbf{X}_1'\mathbf{M}_2\mathbf{W}_{rm}^+$$

belonging to $\{\mathbf{P}_{\boldsymbol{\theta}_1|\mathscr{M}_{12\cdot 2}}\}$ is unique with respect to the choice of generalized inverses indicated by "$^-$" as well as with the choice of $\mathbf{W}_{rm} \in \mathcal{W}_\geq(\mathscr{M}_{12\cdot 2})$. It is easy to confirm that $\mathscr{C}(\mathbf{X}_1'\mathbf{M}_2) = \mathscr{C}(\mathbf{X}_1'\dot{\mathbf{M}}_2\mathbf{X}_1) = \mathscr{C}(\mathbf{X}_1')$ where the last iquality holds if $\boldsymbol{\mu}_1$ is estimable in $\mathscr{M}_{12}$.

It is clear that for estimable $\boldsymbol{\mu}_1$ we have

$$\mathrm{BLUE}(\boldsymbol{\mu}_1 \mid \mathscr{M}_{12\cdot 2}) = \mathrm{BLUE}(\boldsymbol{\mu}_1 \mid \mathscr{M}_{12}) = \mathbf{X}_1(\mathbf{X}_1'\dot{\mathbf{M}}_2\mathbf{X}_1)^-\mathbf{X}_1'\dot{\mathbf{M}}_2\mathbf{y}. \tag{38}$$

Notice that by (36) $\mathbf{M}_2\mathbf{V}\mathbf{M}_2 \in \mathcal{W}_\geq(\mathscr{M}_{12\cdot 2})$ if and only if $\mathscr{C}(\mathbf{M}_2\mathbf{X}_1) \subseteq \mathscr{C}(\mathbf{M}_2\mathbf{V})$, i.e.,

$$\mathscr{C}(\mathbf{X}_1) \subseteq \mathscr{C}(\mathbf{X}_2 : \mathbf{V}).$$

Thus, for example for a positive definite $\mathbf{V}$ and full-rank $\mathbf{X}$ we have

$$\tilde{\boldsymbol{\beta}}_1(\mathscr{M}_{12}) = (\mathbf{X}_1'\dot{\mathbf{M}}_{2V}\mathbf{X}_1)^{-1}\mathbf{X}_1'\dot{\mathbf{M}}_{2V}\mathbf{y}, \quad \mathrm{cov}[\tilde{\boldsymbol{\beta}}_1(\mathscr{M}_{12})] = (\mathbf{X}_1'\dot{\mathbf{M}}_{2V}\mathbf{X}_1)^{-1},$$

where

$$\dot{\mathbf{M}}_{2V} = \mathbf{M}_2(\mathbf{M}_2\mathbf{V}\mathbf{M}_2)^-\mathbf{M}_2 = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}_2(\mathbf{X}_2'\mathbf{V}^{-1}\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{V}^{-1}.$$

Consider now the following choice for $\mathbf{W}_\ell \in \mathcal{W}_\geq(\mathscr{M}_{12})$ (with obvious partitioning)

$$\mathbf{W}_\ell = \mathbf{V} + \mathbf{XUU}'\mathbf{X}' = \mathbf{V} + \mathbf{X}\begin{pmatrix}\mathbf{U}_1\\ \mathbf{U}_2\end{pmatrix}(\mathbf{U}_1' : \mathbf{U}_2')\mathbf{X}',$$

where $\mathbf{U} = (\mathbf{U}_1' : \mathbf{U}_2')'$ has the property

$$\mathscr{C}(\mathbf{W}_\ell) = \mathscr{C}(\mathbf{V} : \mathbf{X}_1\mathbf{U}_1 + \mathbf{X}_2\mathbf{U}_2) = \mathscr{C}(\mathbf{V} : \mathbf{X}_1 : \mathbf{X}_2)\,. \tag{39}$$

Premultiplying (39) with $\mathbf{M}_2$ gives

$$\mathscr{C}(\mathbf{M}_2\mathbf{W}_\ell) = \mathscr{C}[\mathbf{M}_2(\mathbf{V} : \mathbf{X}_1\mathbf{U}_1)] = \mathscr{C}(\mathbf{M}_2\mathbf{W}_1)\,.$$

The nonnegative definiteness of $\mathbf{W}_\ell$ means that $\mathscr{C}(\mathbf{M}_2\mathbf{W}_\ell) = \mathscr{C}(\mathbf{M}_2\mathbf{W}_\ell\mathbf{M}_2)$ and so we have proved the following:

$$\mathbf{W} \in \mathcal{W}_{\geq}(\mathscr{M}_{12}) \implies \mathbf{M}_2\mathbf{W}\mathbf{M}_2 \in \mathcal{W}_{\geq}(\mathscr{M}_{12\cdot2})\,. \tag{40}$$

Thus the $\mathrm{BLUE}(\boldsymbol{\theta}_1 \mid \mathscr{M}_{12})$ has a representation $\mathbf{P}_{\mathbf{M}_2\mathbf{X}_1;\mathbf{W}_\ell^-}\mathbf{M}_2\mathbf{y}$, where

$$\mathbf{P}_{\mathbf{M}_2\mathbf{X}_1;\mathbf{W}_\ell^-} = \mathbf{M}_2\mathbf{X}_1(\mathbf{X}_1'\mathbf{M}_2\mathbf{W}_\ell^-\mathbf{M}_2\mathbf{X}_1)^-\mathbf{X}_1'\mathbf{M}_2\mathbf{W}_\ell^- \in \{\mathbf{P}_{\boldsymbol{\theta}_1|\mathscr{M}_{12\cdot2}}\}$$

so that

$$\mathrm{BLUE}(\boldsymbol{\theta}_1 \mid \mathscr{M}_{12}) = \mathbf{M}_2\mathbf{X}_1(\mathbf{X}_1'\dot{\mathbf{M}}_{2W}\mathbf{X}_1)^-\mathbf{X}_1'\dot{\mathbf{M}}_{2W}\mathbf{y},$$
$$\mathrm{BLUE}(\boldsymbol{\mu}_1 \mid \mathscr{M}_{12}) = \mathbf{X}_1(\mathbf{X}_1'\dot{\mathbf{M}}_{2W}\mathbf{X}_1)^-\mathbf{X}_1'\dot{\mathbf{M}}_{2W}\mathbf{y},$$

where

$$\dot{\mathbf{M}}_{2W} = \mathbf{M}_2(\mathbf{M}_2\mathbf{W}_\ell\mathbf{M}_2)^-\mathbf{M}_2 \quad\text{and}\quad \mathbf{W}_\ell \in \mathcal{W}_{\geq}(\mathscr{M}_{12})\,.$$

Denoting

$$\mathbf{P}_{\mathbf{M}_2\mathbf{X}_1;\mathbf{W}_\ell^+} = \mathbf{M}_2\mathbf{X}_1(\mathbf{X}_1'\mathbf{M}_2\mathbf{W}_\ell^-\mathbf{M}_2\mathbf{X}_1)^-\mathbf{X}_1'\mathbf{M}_2\mathbf{W}_\ell^+ \in \{\mathbf{P}_{\boldsymbol{\theta}_1|\mathscr{M}_{12\cdot2}}\}\,,$$

we observe by Proposition 4.1 that

$$\mathbf{P}_{\mathbf{M}_2\mathbf{X}_1;\mathbf{W}_\ell^+} = \mathbf{P}_{\mathbf{M}_2\mathbf{X}_1;\mathbf{W}_{rm}^+}\,,$$
$$\mathrm{BLUE}(\boldsymbol{\theta}_1 \mid \mathscr{M}_{12}) = \mathbf{P}_{\mathbf{M}_2\mathbf{X}_1;\mathbf{W}_\ell^-}\mathbf{M}_2\mathbf{y} = \mathbf{P}_{\mathbf{M}_2\mathbf{X}_1;\mathbf{W}_{rm}^-}\mathbf{M}_2\mathbf{y}\,.$$

Now we can wonder whether (40) holds for *any* $\mathbf{W} \in \mathcal{W}(\mathscr{M}_{12})$ so that $\mathbf{W}$ is not necessarily symmetric nor nonnegative definite. So, let $\mathbf{W}_t$ be of the form $\mathbf{W}_t = \mathbf{V} + \mathbf{X}\mathbf{T}\mathbf{X}'$, where

$$\mathscr{C}(\mathbf{W}_t) = \mathscr{C}(\mathbf{V} + \mathbf{X}\mathbf{T}\mathbf{X}') = \mathscr{C}(\mathbf{X} : \mathbf{V})\,. \tag{41}$$

To have (40) holding for any $\mathbf{W}_t \in \mathcal{W}(\mathscr{M}_{12})$ we should have

$$\mathscr{C}(\mathbf{M}_2\mathbf{W}_t\mathbf{M}_2) = \mathscr{C}[\mathbf{M}_2(\mathbf{X} : \mathbf{V})]\,. \tag{42}$$

By (41) we observe that

$$\mathscr{C}(\mathbf{M}_2\mathbf{W}_t\mathbf{M}_2) \subseteq \mathscr{C}(\mathbf{M}_2\mathbf{W}_t) = \mathscr{C}[\mathbf{M}_2(\mathbf{X} : \mathbf{V})]\,, \tag{43}$$

so that (42) holds if and only if

$$\mathscr{C}(\mathbf{M}_2\mathbf{W}_t\mathbf{M}_2) = \mathscr{C}(\mathbf{M}_2\mathbf{W}_t)\,. \tag{44}$$

In other words, the implication

$$\mathbf{W} \in \mathcal{W}(\mathscr{M}_{12}) \implies \mathbf{M}_2\mathbf{W}\mathbf{M}_2 \in \mathcal{W}(\mathscr{M}_{12\cdot2}) \tag{45}$$

holds if and only if (44) holds; this happens, e.g., if $\mathbf{W}$ is nonnegative definite.

One alternative expression for the BLUE of $\boldsymbol{\mu}_1$ can be obtained by premultiplying the fundamental BLUE-equation

$$\mathbf{X}(\mathbf{X}'\mathbf{W}^-\mathbf{X})^-\mathbf{X}'\mathbf{W}^-(\mathbf{X}_1 : \mathbf{X}_2 : \mathbf{VM}) = (\mathbf{X}_1 : \mathbf{X}_2 : \mathbf{0})\,, \text{ where } \mathbf{W} \in \mathcal{W},$$

by $\mathbf{M}_2$ yielding

$$(\mathbf{M}_2\mathbf{X}_1 : \mathbf{0})(\mathbf{X}'\mathbf{W}^-\mathbf{X})^-\mathbf{X}'\mathbf{W}^-(\mathbf{X}_1 : \mathbf{X}_2 : \mathbf{VM}) = (\mathbf{M}_2\mathbf{X}_1 : \mathbf{0} : \mathbf{0})\,. \tag{46}$$

Because $\mathrm{r}(\mathbf{M}_2\mathbf{X}_1) = \mathrm{r}(\mathbf{X}_1)$, we can, by the rank cancellation rule of Marsaglia & Styan (1974) [31], cancel $\mathbf{M}_2$ in (46) and thus an alternative expression for (38) is

$$\tilde{\boldsymbol{\mu}}_1(\mathscr{M}_{12}) = (\mathbf{X}_1 : \mathbf{0})(\mathbf{X}'\mathbf{W}^-\mathbf{X})^-\mathbf{X}'\mathbf{W}^-\mathbf{y}.$$

Let us figure out what is the covariance matrix of the BLUE of estimable $\boldsymbol{\mu}_1 = \mathbf{X}_1\boldsymbol{\beta}_1$ under $\mathscr{M}_{12}$ when

$$\tilde{\boldsymbol{\mu}}_1(\mathscr{M}_{12}) = \mathbf{X}_1(\mathbf{X}_1'\dot{\mathbf{M}}_{2W}\mathbf{X}_1)^-\mathbf{X}_1'\dot{\mathbf{M}}_{2W}\mathbf{y} =: \mathbf{Ay}.$$

Notice that $\mathrm{cov}(\tilde{\boldsymbol{\mu}}_1 \mid \mathscr{M}_{12})$ is obviously unique and hence invariant for the choice of representation of the BLUE of $\boldsymbol{\mu}_1$. Choosing $\mathbf{W} = \mathbf{V} + \mathbf{XUU}'\mathbf{X}' \in \mathcal{W}_{\geq}(\mathscr{M}_{12})$, where $\mathbf{U} = (\mathbf{U}_1' : \mathbf{U}_2')'$, we get (after straightforward calculation)

$$\begin{aligned}\mathrm{cov}(\tilde{\boldsymbol{\mu}}_1 \mid \mathscr{M}_{12}) &= \mathbf{AVA}' = \mathbf{A}(\mathbf{W} - \mathbf{XUU}'\mathbf{X}')\mathbf{A}' \\ &= \mathbf{X}_1(\mathbf{X}_1'\dot{\mathbf{M}}_{2W}\mathbf{X}_1)^-\mathbf{X}_1' - \mathbf{X}_1\mathbf{U}_1\mathbf{U}_1'\mathbf{X}_1'\,,\end{aligned}$$

where, in light of part (e) of Proposition 6.1 in Section 6, $\mathbf{X}_1'\dot{\mathbf{M}}_{2W}\mathbf{X}_1$ can be written as

$$\mathbf{X}_1'\dot{\mathbf{M}}_{2W}\mathbf{X}_1 = \mathbf{X}_1'\Big[\mathbf{W}^+ - \mathbf{W}^+\mathbf{X}_2(\mathbf{X}_2'\mathbf{W}^+\mathbf{X}_2)^-\mathbf{X}_2'\mathbf{W}^+\Big]\mathbf{X}_1\,.$$

Remark 4. We can generalise the considerations in (41)–(45) for the transformed model $\mathscr{T} = \{\mathbf{Fy}, \mathbf{FX}\boldsymbol{\beta}, \mathbf{FVF}'\}$, where $\mathbf{F} \in \mathbb{R}^{f\times n}$. Then the set of $\mathcal{W}$-matrices is defined as

$$\mathcal{W}(\mathscr{T}) = \Big\{\mathbf{W} : \mathbf{W} = \mathbf{F}(\mathbf{V} + \mathbf{XNX}')\mathbf{F}',\ \mathscr{C}(\mathbf{W}) = \mathscr{C}[\mathbf{F}(\mathbf{X} : \mathbf{V})]\Big\}\,.$$

Choosing $\mathbf{W}_t = \mathbf{V} + \mathbf{XTX}' \in \mathcal{W}(\mathscr{M})$ we have

$$\mathscr{C}(\mathbf{FW}_t\mathbf{F}') \subseteq \mathscr{C}(\mathbf{FW}_t) = \mathscr{C}[\mathbf{F}(\mathbf{X} : \mathbf{V})]\,. \tag{47}$$

If we want that $\mathbf{FW}_t\mathbf{F}' \in \mathcal{W}(\mathscr{T})$, we need to have the equality in (47), which happens if and only if $\mathrm{r}(\mathbf{FW}_t\mathbf{F}') = \mathrm{r}(\mathbf{FW}_t)$. Thus one representation for the BLUE of $\mathbf{FX}\boldsymbol{\beta}$ under $\mathscr{T}$ is

$$\mathbf{FX}[\mathbf{X}'\mathbf{F}'(\mathbf{FW}_t\mathbf{F}')^-\mathbf{FX}]^-\mathbf{X}'\mathbf{F}'(\mathbf{FW}_t\mathbf{F}')^-\mathbf{Fy}\,,$$

where $\mathbf{W}_t \in \mathcal{W}(\mathscr{M})$ and $\mathrm{r}(\mathbf{FW}_t\mathbf{F}') = \mathrm{r}(\mathbf{FW}_t)$ as pointed out by [24, p. 287].

Remark 5. To simplify the considerations in the partitioned model $\mathscr{M}_{12}$ we could consider the subclass $\mathcal{W}_{\#}(\mathscr{M}_{12})$ of $\mathcal{W}_{\geq}(\mathscr{M}_{12})$ defined so that $\mathbf{W} \in \mathcal{W}_{\#}(\mathscr{M}_{12})$ if $\mathbf{W}_i = \mathbf{V} + \mathbf{X}_i\mathbf{L}_i\mathbf{L}_i'\mathbf{X}_i'$, $\mathscr{C}(\mathbf{W}_i) = \mathscr{C}(\mathbf{X}_i : \mathbf{V})$, $i = 1, 2$, and

$$\mathbf{W} = \mathbf{V} + \mathbf{X}_1\mathbf{L}_1\mathbf{L}_1'\mathbf{X}_1' + \mathbf{X}_2\mathbf{L}_2\mathbf{L}_2'\mathbf{X}_2'\,.$$

The benefit in using $\mathcal{W}_{\#}(\mathscr{M}_{12})$ instead of $\mathcal{W}_{\geq}(\mathscr{M}_{12})$ is that some calculations become simpler. For example, if $\mathbf{W} \in \mathcal{W}_{\#}(\mathscr{M}_{12})$, then

$$\mathbf{M}_2\mathbf{W}\mathbf{M}_2 = \mathbf{M}_2\mathbf{W}_1\mathbf{M}_2 \in \mathcal{W}(\mathscr{M}_{12\cdot 2}) ,$$
$$\dot{\mathbf{M}}_2 = \mathbf{M}_2(\mathbf{M}_2\mathbf{W}_1\mathbf{M}_2)^{\cdot}\mathbf{M}_2 = \mathbf{M}_2(\mathbf{M}_2\mathbf{W}\mathbf{M}_2)^{\cdot}\mathbf{M}_2 = \dot{\mathbf{M}}_{2W} ,$$

while $\mathbf{W} \in \mathcal{W}_{\geq}(\mathscr{M}_{12})$ implies

$$\mathbf{M}_2\mathbf{W}\mathbf{M}_2 \in \mathcal{W}(\mathscr{M}_{12\cdot 2}) \ \text{ and } \ \mathbf{M}_2\mathbf{W}_1\mathbf{M}_2 \in \mathcal{W}(\mathscr{M}_{12\cdot 2})$$

but the equality $\mathbf{M}_2\mathbf{W}\mathbf{M}_2 = \mathbf{M}_2\mathbf{W}_1\mathbf{M}_2$ does not necessarily hold.

6 Some Properties of the $\perp$

It is interesting to take a further look at the $\perp$-operation and its usefulness in linear models. Let's begin by citing [35, Sec. 5.13].

Proposition 6.1. *Consider the model $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\beta, \mathbf{V}\}$ and let $\mathbf{W} \in \mathcal{W}(\mathscr{M})$. Then*

$$\mathscr{C}(\mathbf{V}\mathbf{X}^{\perp}) = \mathscr{C}(\mathbf{W}^{-}\mathbf{X} : \mathbf{I}_n - \mathbf{W}^{-}\mathbf{W})^{\perp}, \tag{48}$$

where $\mathbf{W}^{-}$ is an arbitrary (but fixed) generalized inverse of $\mathbf{W}$. The column space $\mathscr{C}(\mathbf{V}\mathbf{X}^{\perp})$ can be expressed also as

$$\mathscr{C}(\mathbf{V}\mathbf{X}^{\perp}) = \mathscr{C}\big[(\mathbf{W}^{-})'\mathbf{X} : \mathbf{I}_n - (\mathbf{W}^{-})'\mathbf{W}'\big]^{\perp}.$$

Moreover, let $\mathbf{V}$ be possibly singular and assume that $\mathscr{C}(\mathbf{X}) \subseteq \mathscr{C}(\mathbf{V})$. Then

$$\mathscr{C}(\mathbf{V}\mathbf{X}^{\perp}) = \mathscr{C}(\mathbf{V}^{-}\mathbf{X} : \mathbf{I}_n - \mathbf{V}^{-}\mathbf{V})^{\perp} \subseteq \mathscr{C}(\mathbf{V}^{-}\mathbf{X})^{\perp},$$

where the inclusion becomes equality if and only if $\mathbf{V}$ is positive definite.

It is of particular interest to note that the perp symbol $\perp$ falls down, so to say, very "nicely" when $\mathbf{V}$ is positive definite:

$$\mathscr{C}(\mathbf{V}\mathbf{X}^{\perp})^{\perp} = \mathscr{C}(\mathbf{V}^{-1}\mathbf{X}) ,$$

but when $\mathbf{V}$ is singular we have to use a much more complicated rule (48).

Markiewicz & Puntanen [29] reviewed various features of the perp-operation, and proved, e.g., the following: If $\mathbf{W} \in \mathcal{W}$, then

$$\mathscr{C}(\mathbf{V}\mathbf{X}^{\perp}) = \mathscr{C}(\mathbf{W}^{-}\mathbf{X})^{\perp} \iff \mathscr{C}(\mathbf{X} : \mathbf{V}) = \mathbb{R}^n.$$

For the following Proposition 6.2, see [35, Sec. 5.13] and [30, Lemma 4]. In this lemma the notation $\mathbf{A}^{1/2}$ stands for the nonnegative definite square root of a nonnegative definite matrix $\mathbf{A}$. Similarly $\mathbf{A}^{+1/2}$ denotes the Moore–Penrose inverse of $\mathbf{A}^{1/2}$ so that $\mathbf{P}_{\mathbf{A}} = \mathbf{A}^{1/2}\mathbf{A}^{+1/2} = \mathbf{A}^{+1/2}\mathbf{A}^{1/2}$.

Proposition 6.2. *Let $\mathbf{W} \in \mathcal{W}_{\geq}(\mathscr{M}_{12})$ and $\dot{\mathbf{M}}_{2W} = \mathbf{M}_2(\mathbf{M}_2\mathbf{W}\mathbf{M}_2)^{-}\mathbf{M}_2$. Then:*

(a) $\mathscr{C}(\mathbf{V}\mathbf{M})^{\perp} = \mathscr{C}(\mathbf{W}\mathbf{M})^{\perp} = \mathscr{C}(\mathbf{W}^{+}\mathbf{X} : \mathbf{Q}_{\mathbf{W}})$, *where $\mathbf{Q}_{\mathbf{W}} = \mathbf{I}_n - \mathbf{P}_{\mathbf{W}}$,*

(b) $\mathscr{C}(\mathbf{W}^{1/2}\mathbf{M}_2)^{\perp} = \mathscr{C}(\mathbf{W}^{+1/2}\mathbf{X}_2 : \mathbf{Q}_{\mathbf{W}})$,

(c) $\mathscr{C}(\mathbf{W}^{1/2}\mathbf{M}_2) = \mathscr{C}(\mathbf{W}^{+1/2}\mathbf{X}_2 : \mathbf{Q}_{\mathbf{W}})^{\perp} = \mathscr{C}(\mathbf{W}^{+1/2}\mathbf{X}_2)^{\perp} \cap \mathscr{C}(\mathbf{W})$,

(d) $\mathbf{P}_{\mathbf{W}^{1/2}\mathbf{M}_2} = \mathbf{P}_\mathbf{W} - \mathbf{P}_{\mathbf{W}^{+1/2}\mathbf{X}_2} = \mathbf{P}_{\mathscr{C}(\mathbf{W}^{+1/2}\mathbf{X}_2)^\perp \cap \mathscr{C}(\mathbf{W})}$,

(e) $\mathbf{P}_\mathbf{W}\dot{\mathbf{M}}_{2W}\mathbf{P}_\mathbf{W} = \mathbf{W}^+ - \mathbf{W}^+\mathbf{X}_2(\mathbf{X}_2'\mathbf{W}^+\mathbf{X}_2)^-\mathbf{X}_2'\mathbf{W}^+$,

(f) $\mathbf{W}\dot{\mathbf{M}}_{2W}\mathbf{X}_1 = [\mathbf{I}_n - \mathbf{X}_2(\mathbf{X}_2'\mathbf{W}^+\mathbf{X}_2)^-\mathbf{X}_2'\mathbf{W}^+]\mathbf{X}_1$.

Proof. Claim (a) follows from Proposition 6.1. Let us take a look, in more details as [30, Sec. 2], at the other statements of Proposition 6.2. We observe that $(\mathbf{W}^{1/2}\mathbf{M}_2)'(\mathbf{W}^{+1/2}\mathbf{X}_2 : \mathbf{Q}_\mathbf{W}) = \mathbf{0}$ so that

$$\mathscr{C}(\mathbf{W}^{+1/2}\mathbf{X}_2 : \mathbf{Q}_\mathbf{W}) \subseteq \mathscr{C}(\mathbf{W}^{1/2}\mathbf{M}_2)^\perp. \tag{49}$$

We further have

$$\begin{aligned} \mathrm{r}(\mathbf{W}^{+1/2}\mathbf{X}_2 : \mathbf{Q}_\mathbf{W}) &= \mathrm{r}(\mathbf{W}^{+1/2}\mathbf{X}_2) + \mathrm{r}(\mathbf{Q}_\mathbf{W}) \\ &= \mathrm{r}(\mathbf{X}_2) + n - \mathrm{r}(\mathbf{W}), \\ \mathrm{r}(\mathbf{W}^{1/2}\mathbf{M}_2)^\perp &= n - \mathrm{r}(\mathbf{W}^{1/2}\mathbf{M}_2) \\ &= n - [\mathrm{r}(\mathbf{W}^{1/2}) - \dim \mathscr{C}(\mathbf{W}^{1/2}) \cap \mathscr{C}(\mathbf{X}_2)] \\ &= n - \mathrm{r}(\mathbf{W}) + \mathrm{r}(\mathbf{X}_2), \end{aligned}$$

which confirms the equality in (49), i.e., claim (b) which is obviously equivalent to (c). Part (d) follows from (c):

$$\begin{aligned} \mathbf{P}_{\mathbf{W}^{1/2}\mathbf{M}_2} &= \mathbf{I}_n - \mathbf{P}_{(\mathbf{W}^{+1/2}\mathbf{X}_2 : \mathbf{Q}_\mathbf{W})} = \mathbf{I}_n - (\mathbf{Q}_\mathbf{W} + \mathbf{P}_{\mathbf{W}^{+1/2}\mathbf{X}_2}) \\ &= \mathbf{P}_\mathbf{W} - \mathbf{P}_{\mathbf{W}^{+1/2}\mathbf{X}_2} = \mathbf{P}_{\mathscr{C}(\mathbf{W}^{+1/2}\mathbf{X}_2)^\perp \cap \mathscr{C}(\mathbf{W})}. \end{aligned}$$

In view of (d) we have

$$\begin{aligned} \mathbf{P}_\mathbf{W}\dot{\mathbf{M}}_{2W}\mathbf{P}_\mathbf{W} &= \mathbf{P}_\mathbf{W}\mathbf{M}_2(\mathbf{M}_2\mathbf{W}\mathbf{M}_2)^-\mathbf{M}_2\mathbf{P}_\mathbf{W} \\ &= \mathbf{W}^{+1/2}\mathbf{P}_{\mathbf{W}^{1/2}\mathbf{M}_2}\mathbf{W}^{+1/2} \\ &= \mathbf{W}^{+1/2}(\mathbf{P}_\mathbf{W} - \mathbf{P}_{\mathbf{W}^{+1/2}\mathbf{X}_2})\mathbf{W}^{+1/2} \\ &= \mathbf{W}^+ - \mathbf{W}^+\mathbf{X}_2(\mathbf{X}_2'\mathbf{W}^+\mathbf{X}_2)^-\mathbf{X}_2'\mathbf{W}^+, \end{aligned}$$

and hence

$$\begin{aligned} \mathbf{W}\dot{\mathbf{M}}_{2W}\mathbf{X}_1 &= \mathbf{W}[\mathbf{W}^+ - \mathbf{W}^+\mathbf{X}_2(\mathbf{X}_2'\mathbf{W}^+\mathbf{X}_2)^-\mathbf{X}_2'\mathbf{W}^+]\mathbf{X}_1 \\ &= [\mathbf{I}_n - \mathbf{X}_2(\mathbf{X}_2'\mathbf{W}^+\mathbf{X}_2)^-\mathbf{X}_2'\mathbf{W}^+]\mathbf{X}_1. \end{aligned}$$

which completes the proof.

Remark 6. Markiewicz & Puntanen [30, p. 11] mention that in claim (f) of Proposition 6.2 the matrix $\mathbf{W}$ can be replaced with $\mathbf{W}_1$ to obtain

$$\mathbf{W}_1\dot{\mathbf{M}}_2\mathbf{X}_1 = [\mathbf{I}_n - \mathbf{X}_2(\mathbf{X}_2'\mathbf{W}_1^+\mathbf{X}_2)^-\mathbf{X}_2'\mathbf{W}_1^+]\mathbf{X}_1, \tag{51}$$

where $\dot{\mathbf{M}}_2 = \mathbf{M}_2(\mathbf{M}_2\mathbf{W}_1\mathbf{M}_2)^-\mathbf{M}_2$. However, (51) does not hold in general; it holds, for example, if $\mathscr{C}(\mathbf{X}_2) \subseteq \mathscr{C}(\mathbf{W}_1)$.

For completeness we state the following related result, due to [41, p. 140].

Proposition 6.3. *Consider the linear model $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$ and denote $\mathbf{W} = \mathbf{V} + \mathbf{X}\mathbf{T}\mathbf{X}'$, where $\mathscr{C}(\mathbf{W}) = \mathscr{C}(\mathbf{X} : \mathbf{V})$, and let $\mathbf{W}^-$ be an arbitrary generalized inverse of $\mathbf{W}$. Then*

$$\begin{aligned} &\mathscr{C}(\mathbf{W}^-\mathbf{X}) \oplus \mathscr{C}(\mathbf{X})^\perp = \mathbb{R}^n, &&\mathscr{C}(\mathbf{W}^-\mathbf{X})^\perp \oplus \mathscr{C}(\mathbf{X}) = \mathbb{R}^n, \\ &\mathscr{C}[(\mathbf{W}^-)'\mathbf{X}] \oplus \mathscr{C}(\mathbf{X})^\perp = \mathbb{R}^n, &&\mathscr{C}[(\mathbf{W}^-)'\mathbf{X}]^\perp \oplus \mathscr{C}(\mathbf{X}) = \mathbb{R}^n. \end{aligned}$$

7 Linear Sufficiency

A linear statistic $\mathbf{Fy}$, where $\mathbf{F} \in \mathbb{R}^{f\times n}$, is called linearly sufficient for $\mathbf{X}\boldsymbol{\beta}$ under the model $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$, if there exists a matrix $\mathbf{A} \in \mathbb{R}^{n\times f}$ such that $\mathbf{AFy}$ is the BLUE for $\mathbf{X}\boldsymbol{\beta}$. Correspondingly, $\mathbf{Fy}$ is linearly sufficient for estimable $\mathbf{K}\boldsymbol{\beta}$, where $\mathbf{K} \in \mathbb{R}^{k\times p}$, if there exists a matrix $\mathbf{A} \in \mathbb{R}^{k\times f}$ such that $\mathbf{AFy}$ is the BLUE for $\mathbf{K}\boldsymbol{\beta}$.

The concept of linear sufficiency was essentially introduced in early 1980s by Baksalary & Kala [4, 3] and by Drygas [10]. [4] talked about "linear transformations preserving best linear unbiased estimators" and Drygas [10] introduced the term "linear sufficiency".

By definition, $\mathbf{Fy}$ is linearly sufficient for estimable $\mathbf{K}\boldsymbol{\beta}$ if and only if the equation

$$\mathbf{AF}(\mathbf{X} : \mathbf{VM}) = (\mathbf{K} : \mathbf{0})$$

has a solution for $\mathbf{A}$, which happens if and only if

$$\mathscr{C}\begin{pmatrix}\mathbf{K}'\\ \mathbf{0}\end{pmatrix} \subseteq \mathscr{C}\begin{pmatrix}\mathbf{X}'\mathbf{F}'\\ \mathbf{MVF}'\end{pmatrix}.$$

Sometimes we may use the notation $\mathbf{Fy} \in \mathcal{S}(\mathbf{K}\boldsymbol{\beta})$ to indicate that $\mathbf{Fy}$ is linearly sufficient for $\mathbf{K}\boldsymbol{\beta}$. Moreover, we can denote, symbolically,

$$\mathcal{S}(\mathbf{K}\boldsymbol{\beta}) = \{\mathbf{Fy} : \mathbf{AF}(\mathbf{X} : \mathbf{VM}) = (\mathbf{K} : \mathbf{0}) \text{ for some } \mathbf{A} \in \mathbb{R}^{k\times f}\}.$$

For the proofs of parts (a) and (b) of Proposition 7.1, see [4], and for (c), [3].

Proposition 7.1. *The statistic $\mathbf{Fy}$ is linearly sufficient for $\mathbf{X}\boldsymbol{\beta}$ under the linear model $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$ if and only if any of the following equivalent statements holds:*

(a) $\mathscr{C}(\mathbf{X}) \subseteq \mathscr{C}(\mathbf{WF}')$, *where* $\mathbf{W} \in \mathcal{W}_{\geq}$,

(b) $\mathrm{r}(\mathbf{X} : \mathbf{VF}') = \mathrm{r}(\mathbf{WF}')$, *where* $\mathbf{W} \in \mathcal{W}_{\geq}$.

Moreover, $\mathbf{Fy}$ is linearly sufficient for estimable $\mathbf{K}\boldsymbol{\beta}$ under $\mathscr{M}$ if and only if

(c) $\mathscr{C}[\mathbf{X}(\mathbf{X}'\mathbf{W}^-\mathbf{X})^-\mathbf{K}'] \subseteq \mathscr{C}(\mathbf{WF}')$, *where* $\mathbf{W} \in \mathcal{W}_{\geq}$.

The crucial connection between the concept of linear sufficiency and the transformed model $\mathscr{T} = \{\mathbf{Fy}, \mathbf{FX}\boldsymbol{\beta}, \mathbf{FVF}'\}$ was was proved by Baksalary & Kala [4, 3]: if $\mathbf{K}\boldsymbol{\beta}$ is estimable under $\mathscr{M}$ and $\mathscr{T}$, then

$$\mathbf{Fy} \in \mathcal{S}(\mathbf{K}\boldsymbol{\beta}) \iff \mathrm{BLUE}(\mathbf{K}\boldsymbol{\beta} \mid \mathscr{M}) = \mathrm{BLUE}(\mathbf{K}\boldsymbol{\beta} \mid \mathscr{T}) \text{ w.p. } 1.$$

Thus we do not lose anything essential if we estimate $\mathbf{K}\boldsymbol{\beta}$ under the transformed model $\mathscr{T}$ instead of $\mathscr{M}$.

The next proposition characterizes when $\mathbf{Fy}$ is linearly sufficient for $\boldsymbol{\mu}_1$.

Proposition 7.2. *Let $\boldsymbol{\mu}_1 = \mathbf{X}_1\boldsymbol{\beta}_1$ be estimable under $\mathscr{M}_{12}$ and let $\mathbf{W} \in \mathcal{W}_{\geq}$. Then $\mathbf{Fy}$ is linearly sufficient for $\boldsymbol{\mu}_1$ under $\mathscr{M}_{12}$ if and only if any of the following equivalent conditions holds:*

(a) $\mathscr{C}(\mathbf{W}\dot{\mathbf{M}}_{2W}\mathbf{X}_1) \subseteq \mathscr{C}(\mathbf{WF}')$, *where* $\dot{\mathbf{M}}_{2W} = \mathbf{M}_2(\mathbf{M}_2\mathbf{WM}_2)^-\mathbf{M}_2$.

(b) $\mathscr{C}\big\{[\mathbf{I}_n - \mathbf{X}_2(\mathbf{X}_2'\mathbf{W}^+\mathbf{X}_2)^-\mathbf{X}_2'\mathbf{W}^+]\mathbf{X}_1\big\} \subseteq \mathscr{C}(\mathbf{WF}')$.

Let us prove Proposition 7.2 along the lines of [24, Sec. 3]. For a different proof, see [21, Th. 2]. One expression for the BLUE of $\boldsymbol{\mu}_1 = \mathbf{X}_1\boldsymbol{\beta}_1$, obtainable from $\mathscr{M}_{12\cdot 2}$, is

$$\mathbf{Ay} := \mathbf{X}_1(\mathbf{X}_1'\dot{\mathbf{M}}_{2W}\mathbf{X}_1)^-\mathbf{X}_1'\dot{\mathbf{M}}_{2W}\mathbf{y},$$

where $\dot{\mathbf{M}}_{2W} = \mathbf{M}_2(\mathbf{M}_2\mathbf{W}\mathbf{M}_2)^-\mathbf{M}_2$ and $\mathbf{W} \in \mathcal{W}_{\geq}(\mathscr{M}_{12})$. On the other hand, the BLUE of $\boldsymbol{\mu}_1$ can be written also as

$$\mathbf{By} := (\mathbf{X}_1 : \mathbf{0})(\mathbf{X}'\mathbf{W}^-\mathbf{X})^-\mathbf{X}'\mathbf{W}^-\mathbf{y} = \mathbf{K}(\mathbf{X}'\mathbf{W}^-\mathbf{X})^-\mathbf{X}'\mathbf{W}^-\mathbf{y},$$

where $\mathbf{K} = (\mathbf{X}_1 : \mathbf{0}) \in \mathbb{R}^{n\times p}$ and $\mathbf{W} \in \mathcal{W}_{\geq}(\mathscr{M}_{12})$. By the consistency of the model $\mathscr{M}_{12}$ we have $\mathbf{Ay} = \mathbf{By}$ with probability 1, i.e., $\mathbf{AW} = \mathbf{BW}$, which can be transposed to give

$$\mathbf{W}\dot{\mathbf{M}}_{2W}\mathbf{X}_1(\mathbf{X}_1'\dot{\mathbf{M}}_{2W}\mathbf{X}_1)^-\mathbf{X}_1' = \mathbf{X}(\mathbf{X}'\mathbf{W}^-\mathbf{X})^-\mathbf{K}'.$$

In light of part (c) of Proposition 7.1, the claim (a) is confirmed by showing that

$$\mathscr{C}[\mathbf{W}\dot{\mathbf{M}}_2\mathbf{X}_1(\mathbf{X}_1'\dot{\mathbf{M}}_{2W}\mathbf{X}_1)^-\mathbf{X}_1'] = \mathscr{C}(\mathbf{W}\dot{\mathbf{M}}_{2W}\mathbf{X}_1)\,,$$

i.e., $\mathrm{r}\Big[\mathbf{W}\dot{\mathbf{M}}_{2W}\mathbf{X}_1(\mathbf{X}_1'\dot{\mathbf{M}}_{2W}\mathbf{X}_1)^-\mathbf{X}_1'\Big] = \mathrm{r}(\mathbf{W}\dot{\mathbf{M}}_{2W}\mathbf{X}_1)$, which follows from

$$\begin{aligned}\mathrm{r}(\mathbf{W}\dot{\mathbf{M}}_{2W}\mathbf{X}_1) &\geq \mathrm{r}\Big[\mathbf{W}\dot{\mathbf{M}}_{2W}\mathbf{X}_1(\mathbf{X}_1'\dot{\mathbf{M}}_{2W}\mathbf{X}_1)^-\mathbf{X}_1'\Big]\\ &\geq \mathrm{r}\Big[\mathbf{W}\dot{\mathbf{M}}_{2W}\mathbf{X}_1(\mathbf{X}_1'\dot{\mathbf{M}}_2\mathbf{X}_1)^-\mathbf{X}_1'\dot{\mathbf{M}}_{2W}\mathbf{X}_1\Big]\\ &= \mathrm{r}(\mathbf{W}\dot{\mathbf{M}}_{2W}\mathbf{X}_1)\,.\end{aligned}$$

The claim (b) follows from part (f) of Proposition 6.2.

Remark 7. It is of interest to consider some particular properties related to the linear sufficiency condition (c) of Proposition 7.1:

$$\mathbf{Fy} \in \mathcal{S}(\mathbf{X}\boldsymbol{\beta}) \iff \mathscr{C}(\mathbf{X}) \subseteq \mathscr{C}(\mathbf{WF}')\,, \quad \text{where } \mathbf{W} \in \mathcal{W}_{\geq}(\mathscr{M}). \tag{52}$$

The matrix $\mathbf{W}$ in (52) belongs to the set $\mathcal{W}$ of (symmetric) nonnegative definite matrices. One could wonder whether the column space $\mathscr{C}(\mathbf{WF}')$ is unique once $\mathbf{F}$ is given, i.e., does it remain invariant for any choice of $\mathbf{W} \in \mathcal{W}_{\geq}$? Kala et al. [24, Ex. 1] provide a counterexample showing that this is not the case. Kala et al. [24, Sec. 4] also studied whether the column space $\mathscr{C}(\mathbf{WF}')$ is invariant for any choice of $\mathbf{W} \in \mathcal{W}(\mathscr{M})$ if $\mathbf{Fy} \in \mathcal{S}(\mathbf{X}\boldsymbol{\beta})$. The answer is positive, and moreover,

$$\mathscr{C}(\mathbf{WF}') = \mathscr{C}(\mathbf{X}) \oplus \mathscr{C}(\mathbf{MVF}') = \mathscr{C}(\mathbf{W}'\mathbf{F}')\,.$$

Kala et al. [24, Th. 4] were also wondering whether in (52) the set $\mathcal{W}_{\geq}$ can be replaced with the more general set $\mathcal{W}$. Interestingly, the answer is positive. As far as we know, in all linear sufficiency considerations appearing in literature, it is assumed that $\mathbf{W}$ is nonnegative definite. However, this is not necessary, and $\mathbf{W}$ can also be nonsymmetric. It may be mentioned, in passing, that the proof is parallel to that of [4, p. 914] who utilize the fact that $\mathbf{By}$ is a BLUE of $\mathbf{X}\boldsymbol{\beta}$ if and only if

$$\mathbf{BW} = \mathbf{X}(\mathbf{X}'\mathbf{W}^-\mathbf{X})^-\mathbf{X}', \quad \text{where } \mathbf{W} \in \mathcal{W}_{\geq}\,. \tag{53}$$

It is easy to confirm that in (53) the set $\mathcal{W}_{\geq}$ can be replaced with $\mathcal{W}$. Namely we know that $\mathbf{B} \in \{\mathbf{P}_{\mu|\mathscr{M}}\}$ if and only if

$$\mathbf{B} = \mathbf{X}(\mathbf{X}'\mathbf{W}^-\mathbf{X})^-\mathbf{X}'\mathbf{W}^- + \mathbf{E}\mathbf{Q}_{\mathbf{W}} \tag{54}$$

for some $\mathbf{E} \in \mathbb{R}^{n\times n}$. Postmultiplying (54) by $\mathbf{W}$ and using $\mathbf{X}'\mathbf{W}^-\mathbf{W} = \mathbf{X}'$ gives (53). On the other hand, if $\mathbf{B}$ satisfies (53) then $\mathbf{B}$ is necessarily of the form (54) for some $\mathbf{E}$ and thereby $\mathbf{B} \in \{\mathbf{P}_{\mu|\mathscr{M}}\}$.

Remark 8. Baksalary & Kala (1981) [4, p. 914] write the following (in our notation):

(a) "If the condition $\mathscr{C}(\mathbf{X}) \subseteq \mathscr{C}(\mathbf{W}\mathbf{F}')$, where $\mathbf{W} \in \mathcal{W}_{\geq}(\mathscr{M})$, is satisfied, then each BLUE of $\mathbf{X}\beta$ in the transformed model $\mathscr{T}$ is also a BLUE of $\mathbf{X}\beta$ in the original model $\mathscr{M}$, and vice versa."

It is the phrase *vice versa* that may cause some confusion as stated by [25, Sec. 4]. Let us discuss the meaning of the *vice versa* part along the lines of [16, Sec. 11.6].

Suppose that $\boldsymbol{\eta} = \mathbf{K}\beta$ is estimable under the transformed model $\mathscr{T}$ (and thereby also under $\mathscr{M}$). Then $\mathbf{CFy}$ is the BLUE for $\mathbf{K}\beta$ under $\mathscr{T}$ if and only if $\mathbf{C}$ belongs to the set $\{\mathbf{P}_{\eta|\mathscr{T}}\}$ which is defined as

$$\mathbf{C} \in \{\mathbf{P}_{\eta|\mathscr{T}}\} \iff \mathbf{C}(\mathbf{FX} : \mathbf{FVF}'\mathbf{Q}_{\mathbf{FX}}) = (\mathbf{K} : \mathbf{0})\,.$$

where $\mathbf{Q}_{\mathbf{FX}} = \mathbf{I}_f - \mathbf{P}_{\mathbf{FX}}$. The set of products $\mathbf{CF}$, where $\mathbf{C} \in \{\mathbf{P}_{\eta|\mathscr{T}}\}$, will be denoted as $\{\mathbf{P}_{\eta|\mathscr{T}}\mathbf{F}\}$. It means that each matrix $\mathbf{D} \in \{\mathbf{P}_{\eta|\mathscr{T}}\mathbf{F}\}$ applied to $\mathbf{y}$ provides the BLUE for $\mathbf{K}\beta$ under the transformed model $\mathscr{T}$, i.e.,

$$\mathbf{D} \in \{\mathbf{P}_{\eta|\mathscr{T}}\mathbf{F}\} \iff \mathbf{D} = \mathbf{CF}, \text{ where } \mathbf{C} \in \{\mathbf{P}_{\eta|\mathscr{T}}\}.$$

Consider the multipliers of the response vector $\mathbf{y}$ when playing with the BLUEs under $\mathscr{M}$ and under $\mathscr{T}$; these sets are $\{\mathbf{P}_{\eta|\mathscr{M}}\}$ and $\{\mathbf{P}_{\eta|\mathscr{T}}\mathbf{F}\}$, respectively. Assume further that $\mathbf{Fy}$ is linearly sufficient for $\boldsymbol{\eta}$. Then the inclusion $\{\mathbf{P}_{\eta|\mathscr{T}}\mathbf{F}\} \subseteq \{\mathbf{P}_{\eta|\mathscr{M}}\}$ is straightforward but corresponding equality is more problematic. The following solution was given by [16, Prop. 11.17].

Proposition 7.3. *Let $\boldsymbol{\eta} = \mathbf{K}\beta$ be estimable under $\mathscr{T}$, $\mathbf{W} \in \mathcal{W}_{\geq}(\mathscr{M})$ and assume that $\mathbf{Fy} \in \mathcal{S}(\boldsymbol{\eta})$. Then $\{\mathbf{P}_{\eta|\mathscr{M}}\} = \{\mathbf{P}_{\eta|\mathscr{T}}\mathbf{F}\}$ holds if and only if*

$$\mathbf{Q}_{\mathbf{W}} = \mathbf{Q}_{\mathbf{W}}\mathbf{P}_{\mathbf{F}'}\,, \quad \textit{i.e.,} \quad \mathscr{C}(\mathbf{W})^{\perp} \subseteq \mathscr{C}(\mathbf{F}')\,. \tag{55}$$

In other words, under the linear sufficiency and condition (55)*, each representation of the* BLUE *of* $\mathbf{K}\beta$ *under* $\mathscr{M}$ *is a representation of the* BLUE *under* $\mathscr{T}$ *and vice versa.*

8 Equality of the BLUEs Under Two Models

Let us consider two linear models, $\mathscr{A} = \{\mathbf{y}, \mathbf{X}\beta, \mathbf{V}_a\}$ and $\mathscr{B} = \{\mathbf{y}, \mathbf{X}\beta, \mathbf{V}_b\}$, having different covariance matrices. Let $\mathbf{W}_a \in \mathcal{W}_{\geq}(\mathscr{A})$ so that for some $\mathbf{U}$

$$\mathbf{W}_a = \mathbf{V}_a + \mathbf{XUU}'\mathbf{X}', \text{ where } \mathscr{C}(\mathbf{W}_a) = \mathscr{C}(\mathbf{X} : \mathbf{V}_a)\,.$$

Then one representation for the BLUE of $\mathbf{X}\beta$ under $\mathscr{A}$ is

$$\mathbf{P}_{\mathbf{X};\mathbf{W}_a^+}\mathbf{y} = \mathbf{X}(\mathbf{X}'\mathbf{W}_a^+\mathbf{X})^-\mathbf{X}'\mathbf{W}_a^+\mathbf{y}.$$

We can now ask whether $\mathbf{P}_{\mathbf{X};\mathbf{W}_a^+}\mathbf{y}$ continues to be BLUE under $\mathscr{B}$. This happens if and only if

$$\mathbf{X}(\mathbf{X}'\mathbf{W}_a^+\mathbf{X})^-\mathbf{X}'\mathbf{W}_a^+(\mathbf{X} : \mathbf{V}_b\mathbf{M}) = (\mathbf{X} : \mathbf{0})\,,$$

which is obviously equivalent to $\mathbf{X}'\mathbf{W}_a^+\mathbf{V}_b\mathbf{M} = \mathbf{0}$. Further equivalent conditions are given in Proposition 8.1 below which appears in Mitra & Moore (1973) [33, Th. 2.1, Th. 2.2, Note 1]. Some parts they did not prove in details, giving only hints. For a complete proof, see [18, Th. 1].

Proposition 8.1. *Using the earlier notation, $\mathbf{P}_{\mathbf{X};\mathbf{W}_a^+}\mathbf{y}$ is the* BLUE *for $\mathbf{X}\beta$ also under $\mathscr{B}$ if and only if any of the following equivalent conditions holds:*

(a) $\mathbf{X}'\mathbf{W}_a^+\mathbf{V}_b\mathbf{M} = \mathbf{0}\,,$ (b) $\mathscr{C}(\mathbf{V}_b\mathbf{M}) \subseteq \mathscr{C}(\mathbf{W}_a^+\mathbf{X})^{\perp} =: \mathscr{C}(\mathbf{Z}),$

(c) $\mathbf{V}_b = \mathbf{XRX}' + \mathbf{ZSZ}'$ *for some* $\mathbf{R}$ *and* $\mathbf{S}$, *and* $\mathbf{Z} \in \{(\mathbf{W}_a^+\mathbf{X})^\perp\}$,

(d) $\mathbf{P}_{\mathbf{X};\mathbf{W}_a^+}\mathbf{V}_b$ *is symmetric,*

(e) $\mathscr{C}(\mathbf{W}_a^+\mathbf{X})$ *is spanned by a set of* r *proper eigenvectors of* $\mathbf{V}_b$ *with respect to* $\mathbf{W}_a$; $r = \mathrm{r}(\mathbf{X})$,

(f) $\mathscr{C}(\mathbf{X})$ *is spanned by a set of* r *eigenvectors of* $\mathbf{V}_b\mathbf{W}_a^+$.

In Proposition 8.1 we utilize the concept of *proper eigenvectors* following Rao & Mitra (1971) [41, Sec. 6.3]; see also [34], and [42]. To have a brief look at these concepts, let $\mathbf{A}$ and $\mathbf{B}$ be two symmetric $n \times n$ matrices of which $\mathbf{B}$ is nonnegative definite. Let $\lambda \in \mathbb{R}$ be a scalar and $\mathbf{u}$ a vector such that

$$\mathbf{Au} = \lambda\mathbf{Bu}, \quad \mathbf{Bu} \neq \mathbf{0}\,.$$

Rao & Mitra [41, Sec. 6.3] call λ a proper eigenvalue and $\mathbf{u}$ a proper eigenvector of $\mathbf{A}$ with respect to $\mathbf{B}$, or shortly, $(\lambda, \mathbf{u})$ is a proper eigenpair for $(\mathbf{A}, \mathbf{B})$. If $\mathbf{B}$ is singular, there may exist a vector $\mathbf{u} \neq \mathbf{0}$ such that $\mathbf{Au} = \mathbf{Bu} = \mathbf{0}$, in which case

$$(\mathbf{A} - \lambda\mathbf{B})\mathbf{u} = \mathbf{0}$$

is satisfied with arbitrary λ. Such a vector $\mathbf{u} \in \mathbb{R}^n$ is called an improper eigenvector of $\mathbf{A}$ with respect to $\mathbf{B}$. The space of improper eigenvectors is precisely $\mathscr{N}(\mathbf{A}) \cap \mathscr{N}(\mathbf{B}) = \mathscr{C}(\mathbf{A} : \mathbf{B})^\perp$.

What about if we request that *every* representation of BLUE of $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ under $\mathscr{A}$ continues to be BLUE under $\mathscr{B}$, or shortly

$$\{\mathrm{BLUE}(\boldsymbol{\mu} \mid \mathscr{A})\} \subseteq \{\mathrm{BLUE}(\boldsymbol{\mu} \mid \mathscr{B})\}, \quad \text{i.e.,} \quad \{\mathbf{P}_{\mu|\mathscr{A}}\} \subseteq \{\mathbf{P}_{\mu|\mathscr{B}}\}\,. \tag{56}$$

As an arbitrary member of $\{\mathbf{P}_{\mu|\mathscr{A}}\}$ can be expressed as

$$\mathbf{X}(\mathbf{X}'\mathbf{W}_a^+\mathbf{X})^-\mathbf{X}'\mathbf{W}_a^+ + \mathbf{EQ}_{\mathbf{W}_a}\,, \quad \text{where } \mathbf{E} \in \mathbb{R}^{n\times n} \text{ is free to vary,}$$

we conclude that (56) holds if and only if

$$[\mathbf{X}(\mathbf{X}'\mathbf{W}_a^+\mathbf{X})^-\mathbf{X}'\mathbf{W}_a^+ + \mathbf{EQ}_{\mathbf{W}_a}](\mathbf{X} : \mathbf{V}_b\mathbf{M}) = (\mathbf{X} : \mathbf{0})\,. \tag{57}$$

It is straightforward to conclude that (57) holds for any $\mathbf{E}$ if and only if

$$\mathscr{C}(\mathbf{V}_b\mathbf{M}) \subseteq \mathscr{C}(\mathbf{V}_a\mathbf{M})\,.$$

For the conditions like (56) see, e.g., [33] and [15].

9 Further Remarks

In this paper we have reviewed the properties of matrix $\mathbf{W}$ belonging to the class

$$\mathcal{W}(\mathscr{M}) = \Big\{\mathbf{W} \in \mathbb{R}^{n\times n} : \mathbf{W} = \mathbf{V} + \mathbf{XTX}',\ \mathscr{C}(\mathbf{W}) = \mathscr{C}(\mathbf{X} : \mathbf{V})\Big\},$$

where $\mathbf{T}$ can be any $p \times p$ matrix as long as $\mathscr{C}(\mathbf{W}) = \mathscr{C}(\mathbf{X} : \mathbf{V})$ is satisfied and $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$. Corresponding considerations can be done in other models, like linear model with *new observations*, which we will denote as $\mathscr{M}_*$. The *mixed* linear model is a special case of of the model with new observations. In this article we skip the mixed model but will briefly go through the linear model with new observations.

We can extend the model $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$ by considering a $q \times 1$ random vector $\mathbf{y}_*$, which is an unobservable random vector containing new future observations. These new observations are assumed to be generated from

$$\mathbf{y}_* = \mathbf{X}_*\boldsymbol{\beta} + \boldsymbol{\varepsilon}_* = \boldsymbol{\mu}_* + \boldsymbol{\varepsilon}_*\,,$$

where $\mathbf{X}_*$ is a known $q \times p$ matrix, $\boldsymbol{\beta} \in \mathbb{R}^p$ is the same vector of fixed but unknown parameters as in $\mathscr{M}$, and $\boldsymbol{\varepsilon}_*$ is a q-dimensional random error vector with $\mathrm{E}(\boldsymbol{\varepsilon}_*) = \mathbf{0}$. The covariance matrix of $\mathbf{y}_*$ and the cross-covariance matrix between $\mathbf{y}$ and $\mathbf{y}_*$ are assumed to be known and thus we have

$$\mathrm{E}\begin{pmatrix}\mathbf{y}\\ \mathbf{y}_*\end{pmatrix} = \begin{pmatrix}\boldsymbol{\mu}\\ \boldsymbol{\mu}_*\end{pmatrix} = \begin{pmatrix}\mathbf{X}\\ \mathbf{X}_*\end{pmatrix}\boldsymbol{\beta}\,, \quad \mathrm{cov}\begin{pmatrix}\mathbf{y}\\ \mathbf{y}_*\end{pmatrix} = \begin{pmatrix}\mathbf{V} & \mathbf{V}_{12}\\ \mathbf{V}_{21} & \mathbf{V}_{22}\end{pmatrix}.$$

This setup can be denoted shortly as

$$\mathscr{M}_* = \left\{\begin{pmatrix}\mathbf{y}\\ \mathbf{y}_*\end{pmatrix}, \begin{pmatrix}\mathbf{X}\\ \mathbf{X}_*\end{pmatrix}\boldsymbol{\beta}, \begin{pmatrix}\mathbf{V} & \mathbf{V}_{12}\\ \mathbf{V}_{21} & \mathbf{V}_{22}\end{pmatrix}\right\}.$$

Our aim is to predict the unobservable $\mathbf{y}_*$ on the basis of the observable $\mathbf{y}$.

The random vector $\mathbf{A}\mathbf{y}$ is a linear unbiased predictor (LUP) of $\mathbf{y}_*$ if $\mathrm{E}(\mathbf{y}_* - \mathbf{A}\mathbf{y}) = \mathbf{0}$ for all $\boldsymbol{\beta} \in \mathbb{R}^p$. Such a matrix $\mathbf{A} \in \mathbb{R}^{q\times n}$ exists if and only if $\mathscr{C}(\mathbf{X}_*') \subseteq \mathscr{C}(\mathbf{X}')$, i.e., $\mathbf{X}_*\boldsymbol{\beta}$ is estimable under $\mathscr{M}$ and then we say that $\mathbf{y}_*$ is predictable under $\mathscr{M}_*$. Now a LUP $\mathbf{A}\mathbf{y}$ is the best linear unbiased predictor, BLUP, for $\mathbf{y}_*$, if the covariance matrix of the prediction error, subject to the unbiasedness of the prediction, is minimized:

$$\mathrm{cov}(\mathbf{y}_* - \mathbf{A}\mathbf{y}) \leq_{\mathsf{L}} \mathrm{cov}(\mathbf{y}_* - \mathbf{A}_{\#}\mathbf{y}) \quad \text{for all } \mathbf{A}_{\#} : \mathbf{A}_{\#}\mathbf{X} = \mathbf{X}_*\,.$$

It appears that the linear predictor $\mathbf{A}\mathbf{y}$ is the BLUP for $\mathbf{y}_*$ if and only if $\mathbf{A} \in \mathbb{R}^{q\times n}$ satisfies the the so-called fundamental BLUP equation

$$\mathbf{A}(\mathbf{X} : \mathbf{V}\mathbf{X}^{\perp}) = (\mathbf{X}_* : \mathbf{V}_{21}\mathbf{X}^{\perp})\,. \tag{58}$$

For (58), see, e.g., [9, p. 294], and [20, p. 1015]. Corresponding to (58), $\mathbf{B}\mathbf{y}$ is the $\mathrm{BLUP}(\boldsymbol{\varepsilon}_*)$ whenever

$$\mathbf{B}(\mathbf{X} : \mathbf{V}\mathbf{X}^{\perp}) = (\mathbf{0} : \mathbf{V}_{21}\mathbf{X}^{\perp})\,.$$

Now the $\mathrm{BLUP}(\mathbf{y}_*)$ under $\mathscr{M}_*$, see, e.g., [17, Sec. 2] and [19, Sec. 4], can be written for example as

$$\begin{aligned}\mathrm{BLUP}(\mathbf{y}_*) &= \mathrm{BLUE}(\boldsymbol{\mu}_*) + \mathrm{BLUP}(\boldsymbol{\varepsilon}_*)\\ &= \mathbf{X}_*\mathbf{B}\mathbf{y} + \mathbf{V}_{21}\mathbf{V}^{-}(\mathbf{I}_n - \mathbf{G})\mathbf{y}\\ &= \mathbf{X}_*\mathbf{B}\mathbf{y} + \mathbf{V}_{21}\mathbf{W}^{-}(\mathbf{I}_n - \mathbf{G})\mathbf{y}\\ &= \mathbf{X}_*\mathbf{B}\mathbf{y} + \mathbf{V}_{21}\mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}\mathbf{y},\end{aligned}$$

where $\mathbf{B} = (\mathbf{X}'\mathbf{W}^{-}\mathbf{X})^{-}\mathbf{X}'\mathbf{W}^{-}$ and $\mathbf{G} = \mathbf{X}\mathbf{B} = \mathbf{P}_{\mathbf{X};\mathbf{W}^{-}}$ and $\mathbf{W} \in \mathcal{W}(\mathscr{M})$. In particular, if $\mathbf{V}$ is positive definite and $\mathrm{r}(\mathbf{X}) = p$, we obtain

$$\begin{aligned}\mathrm{BLUP}(\mathbf{y}_*) &= \mathbf{X}_*\tilde{\boldsymbol{\beta}} + \mathbf{V}_{21}\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})\\ &= \mathbf{X}_*\tilde{\boldsymbol{\beta}} + \mathbf{V}_{21}\mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}\mathbf{y},\end{aligned}$$

where $\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$.

Acknowledgements

The authors are grateful to Professor Bikas Kumar Sinha for the invitation to prepare this article for possible publication in the *Gujarat Journal of Statistical and Data Sciences*. Thanks go also to the anonymous referee for constructive comments.

References

[1] A C Aitken. "On least squares and linear combination of observations." In: *Proc. Roy. Soc. Edinburgh Sect. A* 55 (1935), pp. 42–49. DOI: 10.1017/S0370164600014346.

[2] J K Baksalary. "An elementary development of the equation characterizing best linear unbiased estimators." In: *Linear Algebra Appl.* 388 (2004), pp. 3–6. DOI: 10.1016/S0024-3795(03)00396-3.

[3] J K Baksalary and R Kala. "Linear sufficiency with respect to a given vector of parametric functions." In: *J. Stat. Plann. Inference* 14 (1986), pp. 331–338. DOI: 10.1016/0378-3758(86)90171-0.

[4] J K Baksalary and R Kala. "Linear transformations preserving best linear unbiased estimators in a general Gauss–Markoff model." In: *Ann. Stat.* 9 (1981), pp. 913–916. DOI: 10.1214/aos/1176345533.

[5] J K Baksalary and T Mathew. "Rank invariance criterion and its application to the unified theory of least squares." In: *Linear Algebra Appl.* 127 (1990), pp. 393–401. DOI: 10.1016/0024-3795(90)90352-D.

[6] J K Baksalary and S Puntanen. "Weighted-least-squares estimation in the general Gauss–Markov model." In: *Statistical Data Analysis and Inference*. Ed. by Y Dodge. Amsterdam: North-Holland, 1989, pp. 355–368. DOI: 10.1016/B978-0-444-88029-1.50038-1.

[7] J K Baksalary, S Puntanen, and G P H Styan. "A property of the dispersion matrix of the best linear unbiased estimator in the general Gauss–Markov model." In: *Sankhyā Ser. A* 52 (1990), pp. 279–296. URL: http://www.jstor.org/stable/25050797.

[8] J K Baksalary, C R Rao, and A Markiewicz. "A study of the influence of the 'natural restrictions' on estimation problems in the singular Gauss–Markov model." In: *J. Stat. Plann. Inference* 31 (1992), pp. 335–351. DOI: 10.1016/0378-3758(92)90141-E.

[9] R Christensen. *Plane Answers to Complex Questions: the Theory of Linear Models*. 4th ed. New York: Springer, 2011. DOI: 10.1007/978-1-4419-9816-3.

[10] H Drygas. "Sufficiency and completeness in the general Gauss–Markov model." In: *Sankhyā Ser. A* 45 (1983), pp. 88–98. URL: http://www.jstor.org/stable/25050416.

[11] H Drygas. *The Coordinate-Free Approach to Gauss–Markov Estimation*. Berlin: Springer, 1970. DOI: 10.1007/978-3-642-65148-9.

[12] J Groß. "Special generalized inverse matrices connected with the theory of unified least squares." In: *Linear Algebra Appl.* 264 (1997), pp. 325–327. DOI: 10.1016/S0024-3795(95)07778-2.

[13] J. Groß and S Puntanen. "Estimation under a general partitioned linear model." In: *Linear Algebra Appl.* 321 (2000), pp. 131–144. DOI: 10.1016/S0024-3795(00)00028-8.

[14] D A Harville. *Matrix Algebra From a Statistician's Perspective*. New York: Springer, 1997. DOI: 10.1007/b98818.

[15] S J Haslett and S Puntanen. "A review of conditions under which BLUEs and/or BLUPs in one linear mixed model are also BLUEs and/or BLUPs in another." In: *Calcutta Stat. Assoc. Bull.* 65 (2013), pp. 25–41. DOI: 10.1177/0008068320130103.

[16] S J Haslett et al. "A review of the linear sufficiency and linear prediction sufficiency in the linear model with new observations." In: *Multivariate, Multilinear and Mixed Linear Models*. Ed. by D von Rosen K Filipiak A Markiewicz. Cham: Springer, 2021, pp. 265–318. DOI: 10.1007/978-3-030-75494-5_11.

[17] S J Haslett et al. "Equalities between OLSE, BLUE and BLUP in the linear model." In: *Stat. Pap.* 55 (2014), pp. 543–56. DOI: 10.1007/s00362-013-0500-7.

[18] J Hauke, A Markiewicz, and S Puntanen. "Revisiting the BLUE in a linear model via proper eigenvectors." In: *Combinatorial Matrix Theory and Generalized Inverses of Matrices*. Ed. by R B Bapat et al. India: Springer, 2013, pp. 73–83. DOI: 10.1007/978-81-322-1053-5_7.

[19] J Isotalo, A Markiewicz, and S Puntanen. "Some properties of linear prediction sufficiency in the linear model." In: *Trends and Perspectives in Linear Statistical Inference: LinStat, Istanbul, 2016*. Ed. by D von Rosen M Tez. Cham: Springer, 2018, pp. 111–129. DOI: 10.1007/978-3-319-73241-1_8.

[20] J Isotalo and S Puntanen. "Linear prediction sufficiency for new observations in the general Gauss–Markov model." In: *Commun. Stat. Theory Methods* 35 (2006), pp. 1011–1023. DOI: 10.1080/03610920600672146.

[21] J Isotalo and S Puntanen. "Linear sufficiency and completeness in the partitioned linear model." In: *Acta Comment Univ. Tartu Math.* 10 (2006), pp. 53–67. DOI: 10.12697/ACUTM.2006.10.05.

[22] J Isotalo, S Puntanen, and G P H Styan. "A useful matrix decomposition and its statistical applications in linear regression." In: *Commun. Stat. Theory Methods* 37 (2008), pp. 1436–1457. DOI: 10.1080/03610920701666328.

[23] J Isotalo, S Puntanen, and G P H Styan. "The BLUE's covariance matrix revisited: a review." In: *J. Stat. Plann. Inference* 138 (2008), pp. 2722–2737. DOI: 10.1016/j.jspi.2008.03.029.

[24] R Kala, A Markiewicz, and S Puntanen. "Some further remarks on the linear sufficiency in the linear model." In: *Applied and Computational Matrix Analysis: MatTriad, Coimbra, Portugal, 2015*. Ed. by N Bebiano. Cham: Springer, 2017, pp. 275–294. DOI: 10.1007/978-3-319-49984-0_19.

[25] R Kala, S Puntanen, and Y Tian. "Some notes on linear sufficiency." In: *Stat. Pap* 58 (2017), pp. 1–17. DOI: 10.1007/s00362-015-0682-2.

[26] L R LaMotte. "A direct derivation of the REML likelihood function." In: *Stat. Pap.* 48 (2007), pp. 321–327. DOI: 10.1007/s00362-006-0335-6.

[27] E P Liski, S Puntanen, and S G Wang. "Bounds for the trace of the difference of the covariance matrices of the OLSE and BLUE." In: *Linear Algebra Appl.* 176 (1992), pp. 121–130. DOI: 10.1016/0024-3795(92)90214-U.

[28] A Markiewicz. "On dependence structures preserving optimality." In: *Statist. Probab. Lett.* 53 (2001), pp. 415–419. DOI: 10.1016/S0167-7152(01)00064-5.

[29] A Markiewicz and S Puntanen. "All about the $\perp$ with its applications in the linear statistical models." In: *Open Mathematics* 13 (2015), pp. 33–50. DOI: 10.1515/math-2015-0005.

[30] A Markiewicz and S Puntanen. "Further properties of the linear sufficiency in the partitioned linear model." In: *Matrices, Statistics and Big Data: Selected Contributions from IWMS 2016*. Ed. by S Puntanen S E Ahmed F Carvalho. Cham: Springer, 2019, pp. 1–22. DOI: 10.1007/978-3-030-17519-1_1.

[31] G Marsaglia and G P H Styan. "Equalities and inequalities for ranks of matrices." In: *Linear Multilinear Algebra* 2 (1974), pp. 269–292. DOI: 10.1080/03081087408817070.

[32] S K Mitra. "Unified least squares approach to linear estimation in a general Gauss–Markov model." In: *SIAM J. Appl. Math.* 25 (1973), pp. 671–680. DOI: 10.1137/0125065.

[33] S K Mitra and B J Moore. "Gauss–Markov estimation with an incorrect dispersion matrix." In: *Sankhyā Ser. A* 35 (1973), pp. 139–152. URL: https://www.jstor.org/stable/25049864.

[34] S K Mitra and C R Rao. "Some results in estimation and tests of linear hypotheses under the Gauss–Markoff model." In: *Sankhyā Ser. A* 30 (1968), pp. 281–290. URL: https://www.jstor.org/stable/25041359.

[35] S Puntanen, G P H Styan, and J Isotalo. *Matrix Tricks for Linear Statistical Models: Our Personal Top Twenty*. Heidelberg: Springer, 2011. DOI: 10.1007/978-3-642-10473-2.

[36] C R Rao. “Least squares theory for possibly singular models.” In: *Canad. J. Statist.* 6 (1978), pp. 19–23. DOI: 10.2307/3314821.

[37] C R Rao. “Projectors, generalized inverses and the BLUE's.” In: *J. Roy. Statist. Soc. Ser. B* 36 (1974), pp. 442–448. DOI: 10.1111/j.2517-6161.1974.tb01019.x.

[38] C R Rao. “Representations of best linear estimators in the Gauss–Markoff model with a singular dispersion matrix.” In: *J. Multivariate Anal.* 3 (1973), pp. 276–292. DOI: 10.1016/0047-259X(73)90042-0.

[39] C R Rao. “The inefficiency of least squares: extensions of the Kantorovich inequality.” In: *Linear Algebra Appl.* 70 (1985), pp. 249–255. DOI: 10.1016/0024-3795(85)90056-4.

[40] C R Rao. “Unified theory of linear estimation.” In: *Sankhyā Ser. A* 33 (1971), pp. 371–394. URL: https://www.jstor.org/stable/25049750.

[41] C R Rao and S K Mitra. *Generalized Inverse of Matrices and Its Applications*. New York: Wiley, 1971.

[42] A J Scott and G P H Styan. “On a separation theorem for generalized eigenvalues and a problem in the analysis of sample surveys.” In: *Linear Algebra Appl.* 70 (1985), pp. 209–224. DOI: 10.1016/0024-3795(85)90053-9.

[43] S G Wang and E P Liski. “Some matrix versions of Cauchy–Schwartz inequality: a statistical approach.” In: *Mathematical Statistics and Applied Probability* 13 (1998), pp. 43–50.

[44] G Zyskind and F B Martin. “On best linear estimation and general Gauss–Markov theorem in linear models with arbitrary nonnegative covariance structure.” In: *SIAM J. Appl. Math.* 17 (1969), pp. 1190–1202. DOI: 10.1137/0117110.