Chapter 11 in *Multivariate, Multilinear and Mixed Linear Models*,
K. Filipiak, A. Markiewicz and D. von Rosen, eds.
Springer International Publishing, Cham, 2021. Pages 265–318.
DOI, Book's DOI.

A review of the linear sufficiency and linear prediction sufficiency in the linear model with new observations

Stephen J. Haslett
Research School of Finance, Actuarial Studies and Statistics, The Australian National University, Canberra, Australia;
Centre for Public Health Research, Massey University, Wellington, New Zealand;
School of Fundamental Sciences, Massey University, Palmerston North, New Zealand
email: stephen.haslett@anu.edu.au, s.j.haslett@massey.ac.nz

Jarkko Isotalo
Faculty of Information Technology and Communication Sciences, FI-33014 Tampere University, Finland
email: jarkko.isotalo@tuni.fi

Radosław Kala
Department of Mathematical and Statistical Methods, Poznań University of Life Sciences, Wojska Polskiego 28, PL-60637 Poznań, Poland
email: kalar@up.poznan.pl

Augustyn Markiewicz
Department of Mathematical and Statistical Methods, Poznań University of Life Sciences, Wojska Polskiego 28, PL-60637 Poznań, Poland
email: amark@up.poznan.pl

Simo Puntanen[1]
Faculty of Information Technology and Communication Sciences, FI-33014 Tampere University, Finland
email: simo.puntanen@tuni.fi

- **Submitted: 21 December 2018**
- **Revised: 7 February 2020. Accepted: 8 February 2020.**

[1]Author responsible for correspondence

Abstract

We consider the general linear model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, denoted as $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$, supplemented with the new unobservable random vector $\mathbf{y}_*$, coming from $\mathbf{y}_* = \mathbf{X}_*\boldsymbol{\beta} + \boldsymbol{\varepsilon}_*$, where the covariance matrix of $\mathbf{y}_*$ is known as well as the cross-covariance matrix between $\mathbf{y}_*$ and $\mathbf{y}$. A linear statistic $\mathbf{Fy}$ is called linearly sufficient for $\mathbf{X}_*\boldsymbol{\beta}$ if there exists a matrix $\mathbf{A}$ such that $\mathbf{AFy}$ is the best linear unbiased estimator, BLUE, for $\mathbf{X}_*\boldsymbol{\beta}$. The concept of linear sufficiency with respect to a predictable random vector is defined in the corresponding way but considering the best linear unbiased predictor, BLUP, instead of BLUE. In this paper, we consider the linear sufficiency of $\mathbf{Fy}$ with respect to $\mathbf{y}_*$, $\mathbf{X}_*\boldsymbol{\beta}$, and $\boldsymbol{\varepsilon}_*$. We also apply our results into the linear mixed model.

The concept of linear sufficiency was essentially introduced in early 1980s by Baksalary, Kala and Drygas. Recently several papers providing further properties of the linear sufficiency have been published by the present authors. Our aim is to provide an easy-to-read review of recent results and while doing that, we go through some basic concepts related to linear sufficiency. As a review paper, we do not provide many proofs, instead, our goal is to explain and clarify the central results.

Key words and phrases: Best linear unbiased estimator, BLUE, best linear unbiased predictor, BLUP, linear sufficiency, linear model with new observations, Löwner ordering, linear mixed model, orthogonal projector, transformed linear model.

MSC: 62J05, 62J10.

1 Preliminaries and introduction to the models

In this section we introduce the notation to be used and briefly go through the various versions of linear models that we are interested in. Also, we present some handy matrix-algebraic tools that will be needed later on.

The symbol $\mathbb{R}^{m\times n}$ denotes the set of $m \times n$ real matrices, while $\mathbf{A}'$, $\mathbf{A}^-$, $\mathbf{A}^+$, $\mathscr{C}(\mathbf{A})$, $\mathscr{C}(\mathbf{A})^\perp$, $\mathrm{r}(\mathbf{A})$, and $\mathscr{N}(\mathbf{A})$, denote, respectively, the transpose, a generalized inverse, the Moore–Penrose inverse, the column space, the orthogonal complement of the column space, rank, and the null space of the matrix $\mathbf{A}$. If $\mathbf{G}$ satisfies $\mathbf{AGA} = \mathbf{A}$, then we denote $\mathbf{G} = \mathbf{A}^-$. The Moore–Penrose inverse $\mathbf{A}^+$ is defined as a unique matrix satisfying the following four conditions:

$$\mathbf{AA}^+\mathbf{A} = \mathbf{A}\,, \quad \mathbf{A}^+\mathbf{AA}^+ = \mathbf{A}^+, \quad (\mathbf{AA}^+)' = \mathbf{AA}^+, \quad (\mathbf{A}^+\mathbf{A})' = \mathbf{A}^+\mathbf{A}\,. \tag{1.1}$$

By $(\mathbf{A} : \mathbf{B})$ we denote the columnwise partitioned matrix with $\mathbf{A}_{a\times b}$ and $\mathbf{B}_{a\times c}$ as submatrices. By $\mathbf{A}^\perp$ we denote any matrix satisfying $\mathscr{C}(\mathbf{A}^\perp) = \mathscr{C}(\mathbf{A})^\perp = \mathscr{N}(\mathbf{A}')$. Notice that if $\mathbf{A} \in \mathbb{R}^{a\times b}$, then $\mathbf{A}^\perp \in \mathbb{R}^{a\times d}$, where $d \geq a - \mathrm{r}(\mathbf{A})$. Notation $\mathbf{P_A} = \mathbf{AA}^+ = \mathbf{A}(\mathbf{A}'\mathbf{A})^-\mathbf{A}'$ stands for the orthogonal projector (with respect to the standard inner product) onto the column space $\mathscr{C}(\mathbf{A})$, and so $\mathbf{P}_{\mathbf{A}'} = \mathbf{A}^+\mathbf{A}$. The orthogonal projector onto $\mathscr{C}(\mathbf{A})^\perp$ is denoted as $\mathbf{Q_A} = \mathbf{I}_a - \mathbf{P_A}$, where $\mathbf{I}_a$ refers to the $a \times a$ identity matrix and a is the number of rows of $\mathbf{A}$. It appears convenient to use the short notation

$$\mathbf{M} = \mathbf{I}_n - \mathbf{P_X}\,, \tag{1.2}$$

where $\mathbf{X}_{n\times p}$ refers the matrix determining the expectation subspace in the linear model. One handy choice (for its symmetry and idempotence) for $\mathbf{X}^\perp$ is $\mathbf{M}$.

The concept of nonnegative definite matrices plays an important role in statistics. Formally, a symmetric $n \times n$ matrix $\mathbf{A}$ is said to be nonnegative definite (or positive semidefinite), denoted as $\mathbf{A} \in \mathrm{NND}_n$, if

$$\mathbf{x}'\mathbf{Ax} \geq 0 \quad \text{for all } \mathbf{x} \in \mathbb{R}^n, \quad \text{or equivalently,} \quad \mathbf{A} = \mathbf{C}'\mathbf{C} \quad \text{for some } \mathbf{C}\,. \tag{1.3}$$

Such matrices can be partially ordered. If $\mathbf{A}, \mathbf{B} \in \mathrm{NND}_n$ and simultaneously $\mathbf{B} - \mathbf{A} \in \mathrm{NND}_n$, then $\mathbf{A}$ is said to be below $\mathbf{B}$ in the Löwner partial order. This fact will be denoted as $\mathbf{A} \leq_{\mathrm{L}} \mathbf{B}$.

Next we shortly describe the various statistical models that we are interested in.

(a) *General linear model, $\mathscr{M}$.* Our main interest lies in the general linear model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \text{or shortly } \mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}, \tag{1.4}$$

where $\mathbf{y}$ is an observable n-dimensional random vector, $\mathbf{X}_{n\times p}$ is a known model matrix, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters, and $\boldsymbol{\varepsilon}$ is an unobservable vector of random errors with expectation $\mathrm{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$, and covariance matrix $\mathrm{cov}(\boldsymbol{\varepsilon}) = \mathbf{V}$. The nonnegative definite covariance matrix $\mathbf{V}$ is known and can be singular.

(b) *Partitioned linear model, $\mathscr{M}_{12}$.* One special case of $\mathscr{M}$ is the partitioned linear model $\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}$, or shortly denoted

$$\mathscr{M}_{12} = \{\mathbf{y},\, \mathbf{X}\boldsymbol{\beta},\, \mathbf{V}\} = \{\mathbf{y},\, \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2,\, \mathbf{V}\}. \tag{1.5}$$

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

$$\mathscr{M}_{12\cdot 2} = \{\mathbf{M}_2\mathbf{y},\, \mathbf{M}_2\mathbf{X}_1\boldsymbol{\beta}_1,\, \mathbf{M}_2\mathbf{V}\mathbf{M}_2\}, \tag{1.6}$$

which is obtained by premultiplying the model $\mathscr{M}_{12}$ by $\mathbf{M}_2 = \mathbf{I}_n - \mathbf{P}_{\mathbf{X}_2}$. Transformation by $\mathbf{M}_2$ eliminates the "nuisance parameter" $\boldsymbol{\beta}_2$, as some authors say.

(c) *Transformed model, $\mathscr{T}$.* Premultiplying the model $\mathscr{M}$ by an $f \times n$ matrix $\mathbf{F}$ yields the transformed model

$$\mathbf{F}\mathbf{y} = \mathbf{F}\mathbf{X}\boldsymbol{\beta} + \mathbf{F}\boldsymbol{\varepsilon}\,, \quad \text{or shortly } \mathscr{T} = \{\mathbf{F}\mathbf{y}, \mathbf{F}\mathbf{X}\boldsymbol{\beta}, \mathbf{F}\mathbf{V}\mathbf{F}'\}\,. \tag{1.7}$$

The reduced model $\mathscr{M}_{12\cdot 2}$ is of course one example of the transformed models. The transformed model $\mathscr{T}$ will play a crucial role in our considerations. Loosely described, one of the main questions in this article will be the following: does the multiplication of the response $\mathbf{y}$ by $\mathbf{F}$ keep the estimation/prediction "undisturbed", that is, do we lose anything essential as a consequence? Notice that in $\mathscr{M}$ the response $\mathbf{y}$ is n-dimensional while in $\mathscr{T}$ the response $\mathbf{F}\mathbf{y}$ is f-dimensional. In principle the number of rows in $\mathbf{F}$, f, can be greater than, less than or equal to n but intuitively it seems clear that the rows of $\mathbf{F}$ could be chosen linearly independent, see (3.17). In the partitioned model (1.5) the reduction into $\mathscr{M}_{12\cdot 2}$ has been done by $\mathbf{M}_2$ which has n rows but another essentially identical reduction could be carried out by choosing $\mathbf{F}' = \mathbf{X}_2^{\perp}$, where the columns of $n \times r_2$ matrix $\mathbf{X}_2^{\perp}$ would span $\mathscr{C}(\mathbf{M}_2)$, $\mathrm{r}(\mathbf{M}_2) = r_2$.

(d) *Linear model with new observations, $\mathscr{M}_*$.* Let $\mathbf{y}_*$ denote a $q \times 1$ unobservable random vector containing new observations. The new observations are assumed to be generated from

$$\mathbf{y}_* = \mathbf{X}_*\boldsymbol{\beta} + \boldsymbol{\varepsilon}_*\,, \tag{1.8}$$

where $\mathbf{X}_*$ is a known $q \times p$ matrix, $\boldsymbol{\beta}$ is the same vector of fixed but unknown parameters as in $\mathscr{M}$, and $\boldsymbol{\varepsilon}_*$ is a q-dimensional random error vector. We further assume that

$$\mathrm{E}\begin{pmatrix}\mathbf{y}\\ \mathbf{y}_*\end{pmatrix} = \begin{pmatrix}\mathbf{X}\boldsymbol{\beta}\\ \mathbf{X}_*\boldsymbol{\beta}\end{pmatrix} = \begin{pmatrix}\mathbf{X}\\ \mathbf{X}_*\end{pmatrix}\boldsymbol{\beta}\,, \quad \mathrm{cov}\begin{pmatrix}\mathbf{y}\\ \mathbf{y}_*\end{pmatrix} = \begin{pmatrix}\mathbf{V} & \mathbf{V}_{12}\\ \mathbf{V}_{21} & \mathbf{V}_{22}\end{pmatrix} = \boldsymbol{\Gamma}, \tag{1.9}$$

where $\boldsymbol{\Gamma} \in \mathrm{NND}_{n+q}$ is known. We denote this setup shortly as

$$\mathscr{M}_* = \left\{ \begin{pmatrix}\mathbf{y}\\ \mathbf{y}_*\end{pmatrix}, \begin{pmatrix}\mathbf{X}\\ \mathbf{X}_*\end{pmatrix}\boldsymbol{\beta}, \begin{pmatrix}\mathbf{V} & \mathbf{V}_{12}\\ \mathbf{V}_{21} & \mathbf{V}_{22}\end{pmatrix} \right\}. \tag{1.10}$$

We call $\mathscr{M}_*$ "the linear model with new observations". Of course, the word "new" need not be taken here literally. Our main interest in $\mathscr{M}_*$ lies in predicting $\mathbf{y}_*$ on the basis of observable $\mathbf{y}$, but we will also be interested in predicting $\boldsymbol{\varepsilon}_*$. Notice the key role of the (cross-)covariance matrix $\operatorname{cov}(\mathbf{y}, \mathbf{y}_*) = \mathbf{V}_{12} \in \mathbb{R}^{n \times q}$.

(e) *Transformed linear model with new observations, $\mathscr{T}_*$.* Suppose we transform $\mathscr{M}$ into $\mathscr{T}$ and do the prediction of the new observations with the "help" of $\mathbf{Fy}$. Corresponding to $\mathscr{M}_*$ we have now the following setup:

$$\mathscr{T}_* = \left\{ \begin{pmatrix} \mathbf{Fy} \\ \mathbf{y}_* \end{pmatrix}, \begin{pmatrix} \mathbf{FX} \\ \mathbf{X}_* \end{pmatrix} \boldsymbol{\beta}, \begin{pmatrix} \mathbf{FVF}' & \mathbf{FV}_{12} \\ \mathbf{V}_{21}\mathbf{F}' & \mathbf{V}_{22} \end{pmatrix} \right\}. \tag{1.11}$$

(f) *Mixed linear model, $\mathscr{L}$.* One application of the model $\mathscr{M}_*$ is the linear mixed model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Zu} + \mathbf{e}, \quad \text{or shortly,} \quad \mathscr{L} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta} + \mathbf{Zu}, \mathbf{D}, \mathbf{R}, \mathbf{S}\}, \tag{1.12}$$

where $\mathbf{X}_{n \times p}$ and $\mathbf{Z}_{n \times q}$ are known matrices, $\boldsymbol{\beta} \in \mathbb{R}^p$ is a vector of unknown fixed effects, $\mathbf{u}$ is an unobservable vector (q elements) of random effects with $\mathrm{E}(\mathbf{u}) = \mathbf{0}$, $\operatorname{cov}(\mathbf{u}) = \mathbf{D}$, $\operatorname{cov}(\mathbf{e}, \mathbf{u}) = \mathbf{S}$, and $\mathrm{E}(\mathbf{e}) = \mathbf{0}$, $\operatorname{cov}(\mathbf{e}) = \mathbf{R}$. (Often in applications $\mathbf{S} = \mathbf{0}$.) In this situation we have

$$\operatorname{cov}\begin{pmatrix} \mathbf{e} \\ \mathbf{u} \end{pmatrix} = \begin{pmatrix} \mathbf{R} & \mathbf{S} \\ \mathbf{S}' & \mathbf{D} \end{pmatrix}, \quad \operatorname{cov}\begin{pmatrix} \mathbf{y} \\ \mathbf{u} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Sigma} & \mathbf{ZD} + \mathbf{S} \\ (\mathbf{ZD} + \mathbf{S})' & \mathbf{D} \end{pmatrix}, \tag{1.13}$$

$$\boldsymbol{\Sigma} = \operatorname{cov}(\mathbf{y}) = \operatorname{cov}(\mathbf{Zu} + \mathbf{e}) = \mathbf{ZDZ}' + \mathbf{R} + \mathbf{ZS}' + \mathbf{SZ}'. \tag{1.14}$$

In Section 9 it is shown how the mixed model can be expressed as a version of the model with new observations.

Our matrix expressions will use generalized inverses heavily and in this context it is essential to know whether the expressions are independent of the choice of the generalized inverses involved. Lemma 2.2.4 of Rao and Mitra (1971) gives the condition under which the matrix product $\mathbf{AB}^-\mathbf{C}$ is invariant with respect to the choice of $\mathbf{B}^-$.

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

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

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

Suppose that the matrix equation

$$\mathbf{YB} = \mathbf{A} \tag{1.15}$$

is solvable for $\mathbf{Y}$, i.e., $\mathscr{C}(\mathbf{A}') \subset \mathscr{C}(\mathbf{B}')$. Then it is well known, see, e.g., Rao and Mitra (1971, p. 24) and Ben-Israel and Greville (2003, p. 52), that the general solution $\mathbf{Y}_0$ to (1.15) can be written as

$$\mathbf{Y}_0 = \{\text{one solution to (1.15)}\} + \mathbf{E}_0\mathbf{Q}_{\mathbf{B}}, \quad \text{where } \mathbf{E}_0 \text{ is free to vary.} \tag{1.16}$$

There is one special class of matrices worth particular attention and that is the set $\mathcal{W}$ of nonnegative definite matrices defined as

$$\mathcal{W} = \big\{\mathbf{W} \in \mathbb{R}^{n\times n} : \mathbf{W} = \mathbf{V} + \mathbf{X}\mathbf{U}\mathbf{U}'\mathbf{X}',\ \mathscr{C}(\mathbf{W}) = \mathscr{C}(\mathbf{X} : \mathbf{V})\big\}\,. \tag{1.17}$$

In (1.17) $\mathbf{U}$ can be any matrix comprising p rows as long as $\mathscr{C}(\mathbf{W}) = \mathscr{C}(\mathbf{X} : \mathbf{V})$ is satisfied. One obvious choice is $\mathbf{U} = \mathbf{I}_p$. In particular, if $\mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{V})$, we can choose $\mathbf{U} = \mathbf{0}$. Occasionally, we may use the notation $\mathcal{W}_{\mathscr{A}}$ to indicate that the model $\mathscr{A}$, say, is under consideration. We will also use the phrase "$\mathbf{W}_{\mathscr{A}}$ is a $\mathbf{W}$-matrix under the model $\mathscr{A}$". The set $\mathcal{W}$ in (1.17) is of course $\mathcal{W}_{\mathscr{M}}$. The following result is easy to confirm:

$$\mathbf{W} \in \mathcal{W}_{\mathscr{M}} \implies \mathbf{F}\mathbf{W}\mathbf{F}' \in \mathcal{W}_{\mathscr{T}}\,. \tag{1.18}$$

We will later utilize some particular properties of $\mathcal{W}$ and of the corresponding extended set

$$\mathcal{W}^{\#} = \big\{\mathbf{W} \in \mathbb{R}^{n\times n} : \mathbf{W} = \mathbf{V} + \mathbf{X}\mathbf{U}\mathbf{X}',\ \mathscr{C}(\mathbf{W}) = \mathscr{C}(\mathbf{X} : \mathbf{V})\big\}\,. \tag{1.19}$$

Above $\mathbf{U} \in \mathbb{R}^{p\times p}$ is free to vary subject to condition $\mathscr{C}(\mathbf{W}) = \mathscr{C}(\mathbf{X} : \mathbf{V})$. Notice that $\mathbf{W}$ belonging to $\mathcal{W}^{\#}$ is not necessarily nonnegative definite and it can be nonsymmetric. The following Lemma 1.2 collects together some important properties of the class $\mathcal{W}^{\#}$.

Lemma 1.2. *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 $\mathbf{W} = \mathbf{V} + \mathbf{X}\mathbf{U}\mathbf{X}'$, where $\mathbf{U}$ is a $p \times p$ matrix. Then the following statements are equivalent:*

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

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

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

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

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

Moreover, each of these statements is equivalent also to $\mathscr{C}(\mathbf{X} : \mathbf{V}) = \mathscr{C}(\mathbf{W}')$, and hence to the statements (b)–(e) by replacing $\mathbf{W}$ with $\mathbf{W}'$. Observe that obviously $\mathscr{C}(\mathbf{W}) = \mathscr{C}(\mathbf{W}')$ and that the invariance properties in (d) and (e) concern also the choice of $\mathbf{W} \in \mathcal{W}^{\#}$. For further properties of $\mathcal{W}^{\#}$, see, e.g., Baksalary et al. (1990, Th. 2), Baksalary and Mathew (1990, Th. 2), Harville (1997, p. 468), and Puntanen et al. (2011, Sec. 12.3).

For the following lemma, see, e.g., Isotalo et al. (2008a), Puntanen et al. (2011, Prop. 15.2) and Markiewicz and Puntanen (2019a, Sec. 4).

Lemma 1.3. *Consider the partitioned linear model $\mathscr{M} = \{\mathbf{y}, (\mathbf{X}_1 : \mathbf{X}_2)\boldsymbol{\beta}, \mathbf{V}\}$, let $\mathbf{W} = \mathbf{V} + \mathbf{X}\mathbf{U}\mathbf{U}'\mathbf{X}' \in \mathcal{W}$ and denote $\mathbf{M}_2 = \mathbf{I}_n - \mathbf{P}_{\mathbf{X}_2}$ and*

$$\dot{\mathbf{M}} = \mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^-\mathbf{M}\,, \quad \dot{\mathbf{M}}_{2W} = \mathbf{M}_2(\mathbf{M}_2\mathbf{W}\mathbf{M}_2)^-\mathbf{M}_2\,. \tag{1.20}$$

Then the following equalities hold:

(a) $\mathbf{X}(\mathbf{X}'\mathbf{W}^-\mathbf{X})^-\mathbf{X}'\mathbf{W}^+ = \mathbf{P_W} - \mathbf{VM}(\mathbf{MVM})^-\mathbf{MP_W} = \mathbf{P_W} - \mathbf{V}\dot{\mathbf{M}}\mathbf{P_W}\,,$

(b) $\mathbf{X}(\mathbf{X}'\mathbf{W}^-\mathbf{X})^-\mathbf{X}' = \mathbf{W} - \mathbf{WM}(\mathbf{MVM})^-\mathbf{MW} = \mathbf{V} - \mathbf{V}\dot{\mathbf{M}}\mathbf{V} + \mathbf{XUU}'\mathbf{X}',$

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

(d) $\mathbf{X}_2(\mathbf{X}_2'\mathbf{W}^-\mathbf{X}_2)^-\mathbf{X}_2' = \mathbf{W} - \mathbf{WM}_2(\mathbf{M}_2\mathbf{WM}_2)^-\mathbf{M}_2\mathbf{W},$

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

As noted by Isotalo et al. (2008a, p. 1439), the matrix $\dot{\mathbf{M}} = \mathbf{M}(\mathbf{MVM})^-\mathbf{M}$ is not necessarily unique with respect to the choice of the generalized inverse $(\mathbf{MVM})^-$. It is unique if and only if $\mathbb{R}^n = \mathscr{C}(\mathbf{X} : \mathbf{V})$. However, for example, $\mathbf{V}\dot{\mathbf{M}}\mathbf{P_W}$ is unique. It is noteworthy that using the Moore–Penrose inverse the following holds:

$$\mathbf{M}(\mathbf{MVM})^+\mathbf{M} = (\mathbf{MVM})^+\mathbf{M} = \mathbf{M}(\mathbf{MVM})^+ = (\mathbf{MVM})^+. \tag{1.21}$$

If $\mathbf{V}$ is positive definite and $\mathbf{Z} \in \{\mathbf{X}^\perp\}$, then

$$\begin{aligned}\dot{\mathbf{M}} &= \mathbf{M}(\mathbf{MVM})^-\mathbf{M} = \mathbf{Z}(\mathbf{Z}'\mathbf{VZ})^-\mathbf{Z}' \\ &= \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^-\mathbf{X}'\mathbf{V}^{-1}.\end{aligned} \tag{1.22}$$

The following result concerning the matrix $\mathbf{F}'\mathbf{Q_{FX}}$, where $\mathbf{Q_{FX}} = \mathbf{I}_f - \mathbf{P_{FX}}$, is useful for our considerations. For the proof, see Rao and Mitra (1971, Complement 7, p. 118). For related results, see also Markiewicz and Puntanen (2019a, 2015, Sec. 2, Lemma 3).

Lemma 1.4. *Suppose that* $\mathbf{F}$ *is an* $f \times n$ *matrix. Then*

$$\mathscr{C}(\mathbf{F}'\mathbf{Q_{FX}}) = \mathscr{C}(\mathbf{F}') \cap \mathscr{C}(\mathbf{M})\,, \tag{1.23}$$

and denoting $\mathbf{N} = \mathbf{P}_{\mathbf{F}'\mathbf{Q_{FX}}} = \mathbf{P}_{\mathscr{C}(\mathbf{F}')\cap\mathscr{C}(\mathbf{M})}$, *we have*

$$\mathbf{MF}'\mathbf{Q_{FX}} = \mathbf{F}'\mathbf{Q_{FX}}\,, \qquad \mathbf{N} = \mathbf{MN} = \mathbf{NM}\,. \tag{1.24}$$

We assume the model $\mathscr{M}$ to be consistent in the sense that $\mathbf{y}$ lies in $\mathscr{C}(\mathbf{X} : \mathbf{V})$ with probability 1; see, e.g., Baksalary et al. (1992) and Groß (2004). Hence we assume that under the model $\mathscr{M}$ the observed numerical value of $\mathbf{y}$ satisfies

$$\mathbf{y} \in \mathscr{C}(\mathbf{X} : \mathbf{V}) = \mathscr{C}(\mathbf{X} : \mathbf{VX}^\perp) = \mathscr{C}(\mathbf{X} : \mathbf{VM}) = \mathscr{C}(\mathbf{X}) \oplus \mathscr{C}(\mathbf{VM})\,, \tag{1.25}$$

where "$\oplus$" refers to the direct sum, implying that

$$\mathscr{C}(\mathbf{X}) \cap \mathscr{C}(\mathbf{VX}^\perp) = \{\mathbf{0}\}. \tag{1.26}$$

For the equality $\mathscr{C}(\mathbf{X} : \mathbf{V}) = \mathscr{C}(\mathbf{X} : \mathbf{VM})$, we refer to Rao (1974, Lemma 2.1). There is a related decomposition, see, e.g., Puntanen et al. (2011, Th. 8), that is worth mentioning in this context: for any conformable matrices $\mathbf{A}$ and $\mathbf{B}$ we have

$$\mathscr{C}(\mathbf{A} : \mathbf{B}) = \mathscr{C}(\mathbf{A} : \mathbf{Q}_{\mathbf{A}}\mathbf{B})\,, \text{ and thereby } \mathbf{P}_{(\mathbf{A}:\mathbf{B})} = \mathbf{P}_{\mathbf{A}} + \mathbf{P}_{\mathbf{Q}_{\mathbf{A}}\mathbf{B}}\,. \tag{1.27}$$

Thus we can obtain the following useful results for the partitioned linear model; for the part (b) in Lemma 1.5, see the rank rule of the matrix product of Marsaglia and Styan (1974, Cor. 6.2).

Lemma 1.5. *Consider* $\mathbf{X} = (\mathbf{X}_1 : \mathbf{X}_2)$ *and let* $\mathbf{M}_2 = \mathbf{I}_n - \mathbf{P}_{\mathbf{X}_2}$. *Then*

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

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

Let $\mathbf{A}$ and $\mathbf{B}$ be arbitrary $m \times n$ matrices. Then, in the consistent linear model $\mathscr{M}$, the estimators $\mathbf{Ay}$ and $\mathbf{By}$ are said to be equal with probability 1 if

$$\mathbf{Ay} = \mathbf{By} \quad \text{for all } \mathbf{y} \in \mathscr{C}(\mathbf{X} : \mathbf{V}) = \mathscr{C}(\mathbf{W})\,, \tag{1.28}$$

where $\mathbf{W} \in \mathcal{W}$. Thus, if $\mathbf{A}$ and $\mathbf{B}$ satisfy (1.28), then $\mathbf{A} - \mathbf{B} = \mathbf{C}\mathbf{Q}_{\mathbf{W}}$ for some matrix $\mathbf{C}$. When talking about the equality of estimators like $\mathbf{Ay} = \mathbf{By}$, we often drop off the phrase "with probability 1".

The following lemma collects together some equivalent expressions for (1.28). For part (d) of Lemma 1.6, see Groß and Trenkler (1998, Th. 1).

Lemma 1.6. *Let* $\mathbf{A}$ *and* $\mathbf{B}$ *be* $m \times n$ *matrices. Then under the model* $\mathscr{M}$ *the identity* $\mathbf{Ay} = \mathbf{By}$ *holds with probability* 1 *if and only if any of the following equivalent conditions holds:*

(a) $\mathbf{AX} = \mathbf{BX}$ *and* $\mathbf{AV} = \mathbf{BV}$,

(b) $\mathbf{AX} = \mathbf{BX}$ *and* $\mathbf{AVM} = \mathbf{BVM}\,,$

(c) $\mathbf{AX} = \mathbf{BX}$ *and* $\mathrm{cov}(\mathbf{Ay} - \mathbf{By}) = \mathbf{0}\,,$

(d) $\mathbf{AX} = \mathbf{BX}$, $\mathrm{cov}(\mathbf{Ay}) = \mathrm{cov}(\mathbf{By})$, *and* $2\,\mathrm{cov}(\mathbf{Ay}) = \mathrm{cov}(\mathbf{Ay}, \mathbf{By}) + \mathrm{cov}(\mathbf{By}, \mathbf{Ay})$.

One more notational matter is worth mentioning. Let $\mathbf{A}$ be a nonnegative definite $n \times n$ matrix with $\mathrm{r}(\mathbf{A}) = r$ and let $\mathbf{A} = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}'$ be its eigenvalue decomposition in terms of nonzero eigenvalues; here $\boldsymbol{\Lambda}$ is an $r \times r$ diagonal matrix of the nonzero eigenvalues of $\mathbf{A}$. Then $\mathbf{A}^{1/2}$ refers to the nonnegative definite square root of $\mathbf{A}$, i.e., $\mathbf{A}^{1/2} = \mathbf{Q}\boldsymbol{\Lambda}^{1/2}\mathbf{Q}'$. Notation $\mathbf{A}^{+1/2} = \mathbf{Q}\boldsymbol{\Lambda}^{-1/2}\mathbf{Q}'$ refers to $(\mathbf{A}^+)^{1/2}$ or equivalently $(\mathbf{A}^{1/2})^+$. Notice that $\mathbf{A}^{1/2}\mathbf{A}^{+1/2} = \mathbf{Q}\mathbf{Q}' = \mathbf{A}\mathbf{A}^+ = \mathbf{P}_{\mathbf{A}}$.

As regards the structure of this paper, below is an outline. As a review paper, we do not provide many proofs, instead, our goal is to explain and clarify the relevant central results.

- Section 2: We go through some basic properties of the best linear unbiased estimators and predictors, BLUPs and BLUEs, and introduce the well-known fundamental BLUP and BLUE equations. When dealing with linear predictors, our aim is to predict $\mathbf{y}_*$ and $\boldsymbol{\varepsilon}_*$, and when considering linear estimators, we are interested in estimating $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ or some estimable parametric function $\boldsymbol{\mu}_* = \mathbf{X}_*\boldsymbol{\beta}$.

- Section 3: We recall some known conditions for the linear sufficiency. When dealing with the predictors, we may sometimes use the term "linearly prediction sufficient" instead of the phrase "linearly sufficient".

- Section 4: We consider the transformed model $\mathscr{T} = \{\mathbf{Fy}, \mathbf{FX}\boldsymbol{\beta}, \mathbf{FVF}'\}$ and introduce various properties of the linear sufficiency using this transformed model. Explicit expressions for the BLUEs and BLUPs under the original and under the transformed model are given.

- Section 6: We explore the coincidence of the multipliers of $\mathbf{y}$ providing the BLUEs for $\mathbf{X}_*\boldsymbol{\beta}$ under $\mathscr{M}$ and under $\mathscr{T}$. If every representation of the BLUE of $\mathbf{X}_*\boldsymbol{\beta}$ under $\mathscr{T}$ is BLUE also under $\mathscr{M}$, we can denote

  $$\{\mathrm{BLUE}(\mathbf{X}_*\boldsymbol{\beta}) \mid \mathscr{T}\} \subset \{\mathrm{BLUE}(\mathbf{X}_*\boldsymbol{\beta}) \mid \mathscr{M}\}\,. \tag{1.29}$$

  Proposition 6.2 provides a new interesting characterization for the equality in (1.29).

- Section 5: We shortly discuss the concept of relative linear sufficiency, introduced by Kala et al. (2017b, Sec. 5).

- Section 7: We consider, in the spirit of Markiewicz and Puntanen (2019a), a partitioned linear model $\mathscr{M}_{12} = \{\mathbf{y}, \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2, \mathbf{V}\}$. Particular attention is paid on the situation when the transformation matrix is $\mathbf{M}_2 = \mathbf{I}_n - \mathbf{P}_{\mathbf{X}_2}$, so that the transformed model is $\mathscr{M}_{12\cdot 2} = \{\mathbf{M}_2\mathbf{y}, \mathbf{M}_2\mathbf{X}_1\boldsymbol{\beta}_1, \mathbf{M}_2\mathbf{V}\mathbf{M}_2\}$.

- Section 8: We explore the mutual relations of linear sufficiencies. In addition, we go through some interesting connections between the covariance matrices of the BLUPs and the linear sufficiencies. We also comment on the upper bounds of the Euclidean distance between the BLUPs when the prediction is based on the original model $\mathscr{M}$ and when it is based on the transformed model $\mathscr{T}$.

- Section 9: In this section we consider the linear mixed model in the spirit of Isotalo et al. (2018) and Haslett et al. (2020). The key feature here is the fact that linear mixed model can be interpreted as a special case of the model with new observations.

- Section 10: We give necessary and sufficient conditions that $\mathbf{Fy}$ continues to be linearly sufficient for $\mathbf{X}_*\boldsymbol{\beta}$ (or $\boldsymbol{\varepsilon}_*$) under the misspecified model $\underline{\mathscr{M}}$. The misspecification concerns the covariance part of the setup.

2 BLUEs and BLUPs

In this section we go through some basic properties of the best linear unbiased estimators and predictors in the frames of the general linear model. We will talk about estimation (and estimators) when our main focus is in the fixed parametric function, $\mathbf{X}_*\boldsymbol{\beta}$, say, and about prediction (and predictors), when the main interest lies in predicting a random vector like $\mathbf{y}_*$ which is believed to be generated by $\mathbf{y}_* = \mathbf{X}_*\boldsymbol{\beta} + \boldsymbol{\varepsilon}_*$ in the frames of the model $\mathscr{M}_*$. Not everybody agrees with this division between estimators and predictors; see Robinson (1991, Sec. 1).

A linear statistic $\mathbf{By}$ is said to be linear unbiased estimator (LUE) for $\boldsymbol{\mu}_* = \mathbf{X}_*\boldsymbol{\beta}$ in $\mathscr{M}$ if its expectation is equal to $\boldsymbol{\mu}_*$, i.e.,

$$\mathrm{E}(\mathbf{By} - \boldsymbol{\mu}_*) = (\mathbf{BX} - \mathbf{X}_*)\boldsymbol{\beta} = \mathbf{0} \quad \text{for all } \boldsymbol{\beta} \in \mathbb{R}^p, \tag{2.1}$$

which happens if and only if $\mathbf{X}_*' = \mathbf{X}'\mathbf{B}'$. When $\mathscr{C}(\mathbf{X}_*') \subset \mathscr{C}(\mathbf{X}')$ holds, the linear parametric function $\boldsymbol{\mu}_* = \mathbf{X}_*\boldsymbol{\beta}$ is said to be estimable in $\mathscr{M}$. The LUE $\mathbf{By}$ is the best linear unbiased estimator, BLUE, of estimable $\mathbf{X}_*\boldsymbol{\beta}$ if $\mathbf{By}$ has the smallest covariance matrix in the Löwner sense among all LUEs of $\mathbf{X}_*\boldsymbol{\beta}$:

$$\mathrm{cov}(\mathbf{By}) \leq_{\mathrm{L}} \mathrm{cov}(\mathbf{B}_{\#}\mathbf{y}) \quad \text{for all } \mathbf{B}_{\#} : \mathbf{B}_{\#}\mathbf{X} = \mathbf{X}_* \,. \tag{2.2}$$

Correspondingly, the linear predictor $\mathbf{Ay}$ is said to be unbiased for $\mathbf{y}_*$ if the expected prediction error is zero, i.e., $\mathrm{E}(\mathbf{y}_* - \mathbf{Ay}) = \mathbf{0}$ for all $\boldsymbol{\beta} \in \mathbb{R}^p$, which happens if and only if $\mathbf{X}_*' = \mathbf{X}'\mathbf{A}'$. When $\mathscr{C}(\mathbf{X}_*') \subset \mathscr{C}(\mathbf{X}')$ holds, we will say that $\mathbf{y}_*$ is predictable under $\mathscr{M}_*$, that is, $\mathbf{y}_*$ is predictable whenever $\mathbf{X}_*\boldsymbol{\beta}$ is estimable. Now a linear unbiased predictor $\mathbf{Ay}$ is the best linear unbiased predictor, BLUP, for $\mathbf{y}_*$, if we have the Löwner ordering

$$\mathrm{cov}(\mathbf{y}_* - \mathbf{Ay}) \leq_{\mathrm{L}} \mathrm{cov}(\mathbf{y}_* - \mathbf{A}_{\#}\mathbf{y}) \quad \text{for all } \mathbf{A}_{\#} : \mathbf{A}_{\#}\mathbf{X} = \mathbf{X}_* \,. \tag{2.3}$$

Notice that

- in (2.3) we are minimizing the covariance matrix of the prediction error subject to the unbiasedness of the prediction, while
- in (2.2) we minimize the covariance matrix of the estimator subject to the unbiasedness of the estimation.

Consider then the BLUP of $\boldsymbol{\varepsilon}_*$. Obviously $\mathbf{Dy}$ is an unbiased predictor for $\boldsymbol{\varepsilon}_*$ if and only if $\mathbf{DX} = \mathbf{0}$, i.e., $\mathbf{D} = \mathbf{LM}$ for some $\mathbf{L}$. Thus the unbiased $\mathbf{Dy}$ is the BLUP for $\boldsymbol{\varepsilon}_*$ if and only if

$$\mathrm{cov}(\boldsymbol{\varepsilon}_* - \mathbf{Dy}) \leq_{\mathrm{L}} \mathrm{cov}(\boldsymbol{\varepsilon}_* - \mathbf{D}_{\#}\mathbf{y}) \quad \text{for all } \mathbf{D}_{\#} : \mathbf{D}_{\#}\mathbf{X} = \mathbf{0} \,, \tag{2.4}$$

or equivalently,

$$\mathrm{cov}(\boldsymbol{\varepsilon}_* - \mathbf{Dy}) \leq_{\mathrm{L}} \mathrm{cov}(\boldsymbol{\varepsilon}_* - \mathbf{LMy}) \quad \text{for all } \mathbf{L} \in \mathbb{R}^{q \times n}. \tag{2.5}$$

For Proposition 2.1, characterizing the BLUE, see, e.g., Rao (1973, p. 282), and Drygas (1970, p. 55), Kala (1981, Th. 3.1), and the BLUP, see, e.g., Christensen (2011, p. 294), and Isotalo and Puntanen (2006b, p. 1015). For part (c), see Isotalo et al. (2018, Th. 3.1). For the general reviews of the BLUP-properties, see, e.g., Robinson (1991), Searle (1997), Tian (2015a,b) and Haslett and Puntanen (2017).

Proposition 2.1. *Consider the linear model with new observations defined as* $\mathscr{M}_*$ *in* (1.10), *where* $\mathscr{C}(\mathbf{X}'_*) \subset \mathscr{C}(\mathbf{X}')$, *i.e.,* $\mathbf{y}_*$ *is predictable.*

(a) *The linear predictor* $\mathbf{Ay}$ *is the* BLUP *for* $\mathbf{y}_*$ *if and only if*

$$\mathbf{A}(\mathbf{X} : \mathbf{V}\mathbf{X}^{\perp}) = (\mathbf{X}_* : \mathbf{V}_{21}\mathbf{X}^{\perp})\,. \tag{2.6}$$

(b) *The linear estimator* $\mathbf{By}$ *is the* BLUE *of* $\boldsymbol{\mu}_* = \mathbf{X}_*\boldsymbol{\beta}$ *if and only if*

$$\mathbf{B}(\mathbf{X} : \mathbf{V}\mathbf{X}^{\perp}) = (\mathbf{X}_* : \mathbf{0})\,. \tag{2.7}$$

In particular, $\mathbf{Cy}$ *is the* BLUE *for* $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ *if and only if*

$$\mathbf{C}(\mathbf{X} : \mathbf{V}\mathbf{X}^{\perp}) = (\mathbf{X} : \mathbf{0})\,. \tag{2.8}$$

(c) *The linear predictor* $\mathbf{Dy}$ *is the* BLUP *for* $\boldsymbol{\varepsilon}_*$ *if and only if*

$$\mathbf{D}(\mathbf{X} : \mathbf{V}\mathbf{X}^{\perp}) = (\mathbf{0} : \mathbf{V}_{21}\mathbf{X}^{\perp})\,. \tag{2.9}$$

Equations (2.6) and (2.7) are sometimes called the fundamental BLUP and BLUE equations, respectively. It is noteworthy that equations (2.6) and (2.7) are solvable for $\mathbf{A}$ and $\mathbf{B}$, respectively, if and only if $\mathbf{X}_*\boldsymbol{\beta}$ is estimable while (2.8) and (2.9) are always solvable for $\mathbf{C}$ and $\mathbf{D}$, respectively. If $\mathbf{V}_{12} = \mathbf{0}$, then (2.6) and (2.7) become the same and the $\mathrm{BLUP}(\mathbf{y}_*)$ is the $\mathrm{BLUE}(\mathbf{X}_*\boldsymbol{\beta})$ and the $\mathrm{BLUP}(\boldsymbol{\varepsilon}_*) = \mathbf{0}$.

Putting (2.7) and (2.9) together yields

$$\begin{pmatrix} \mathbf{B} \\ \mathbf{D} \end{pmatrix} (\mathbf{X} : \mathbf{V}\mathbf{X}^{\perp}) = \begin{pmatrix} \mathbf{X}_* & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_{21}\mathbf{X}^{\perp} \end{pmatrix}, \tag{2.10}$$

which implies that

$$(\mathbf{B} + \mathbf{D})(\mathbf{X} : \mathbf{V}\mathbf{X}^{\perp}) = (\mathbf{X}_* : \mathbf{V}_{21}\mathbf{X}^{\perp})\,, \tag{2.11}$$

and thereby $(\mathbf{B} + \mathbf{D})\mathbf{y}$ is a BLUP for $\mathbf{y}_*$ and we have the following result, see, e.g., Isotalo et al. (2018, Sec. 3):

Proposition 2.2. *Under the linear model* $\mathscr{M}_*$, *where* $\mathbf{y}_*$ *is predictable, the following decomposition holds* (*with probability* 1)*:*

$$\mathrm{BLUP}(\mathbf{y}_*) = \mathrm{BLUE}(\mathbf{X}_*\boldsymbol{\beta}) + \mathrm{BLUP}(\boldsymbol{\varepsilon}_*)\,, \quad \textit{or shortly,} \;\; \tilde{\mathbf{y}}_* = \tilde{\boldsymbol{\mu}}_* + \tilde{\boldsymbol{\varepsilon}}_*\,. \tag{2.12}$$

Let us define the sets $\{\mathbf{P}_{\mathbf{y}_*|\mathscr{M}}\}$, $\{\mathbf{P}_{\mathbf{X}_*|\mathscr{M}}\}$, and $\{\mathbf{P}_{\boldsymbol{\varepsilon}_*|\mathscr{M}}\}$ as follows:

$$\mathbf{A} \in \{\mathbf{P}_{\mathbf{y}_*|\mathscr{M}_*}\} \iff \mathbf{A}(\mathbf{X} : \mathbf{VM}) = (\mathbf{X}_* : \mathbf{V}_{21}\mathbf{X}^{\perp}) , \tag{2.13a}$$

$$\mathbf{B} \in \{\mathbf{P}_{\mathbf{X}_*|\mathscr{M}}\} \iff \mathbf{B}(\mathbf{X} : \mathbf{VM}) = (\mathbf{X}_* : \mathbf{0}) , \tag{2.13b}$$

$$\mathbf{D} \in \{\mathbf{P}_{\boldsymbol{\varepsilon}_*|\mathscr{M}_*}\} \iff \mathbf{D}(\mathbf{X} : \mathbf{VM}) = (\mathbf{0} : \mathbf{V}_{21}\mathbf{X}^{\perp}) . \tag{2.13c}$$

Using Lemma 1.2 we can obtain, for example, the following well-known solutions to (2.7) and(2.8):

$$\mathbf{G}_* := \mathbf{X}_*(\mathbf{X}'\mathbf{W}^-\mathbf{X})^-\mathbf{X}'\mathbf{W}^- \in \{\mathbf{P}_{\mathbf{X}_*|\mathscr{M}}\} , \tag{2.14a}$$

$$\mathbf{G} := \mathbf{X}(\mathbf{X}'\mathbf{W}^-\mathbf{X})^-\mathbf{X}'\mathbf{W}^- \in \{\mathbf{P}_{\mathbf{X}|\mathscr{M}}\} , \tag{2.14b}$$

where $\mathbf{W} \in \mathcal{W}$ and we can freely choose the generalized inverses involved. Alternative solutions to (2.13b) are, for example,

$$\mathbf{B}_1 := \mathbf{X}_*(\mathbf{X}'\mathbf{W}^-\mathbf{X})^-\mathbf{X}'\mathbf{W}^+, \quad \mathbf{B}_2 := (\mathbf{X}_* : \mathbf{0})(\mathbf{X} : \mathbf{VM})^+. \tag{2.15}$$

Notice that $\mathbf{B}_1$ is unique with respect to choice of $\mathbf{W}^-$ and $(\mathbf{X}'\mathbf{W}^-\mathbf{X})^-$. In (2.14a)– (2.14b) it is actually enough if $\mathbf{W}$ belongs to the extended set $\mathcal{W}^{\#}$ but for simplicity we deal here with $\mathcal{W}$. The *general* solutions for $\mathbf{B}$ and $\mathbf{C}$ in (2.8) and (2.7), respectively, can be expressed, for example, as

$$\mathbf{P}_{\mathbf{X}_*|\mathscr{M}} = \mathbf{G}_* + \mathbf{E}\mathbf{Q}_{\mathbf{W}} , \quad \mathbf{P}_{\mathbf{X}|\mathscr{M}} = \mathbf{G} + \mathbf{E}_1\mathbf{Q}_{\mathbf{W}} , \tag{2.16}$$

where $\mathbf{E} \in \mathbb{R}^{q\times n}$ and $\mathbf{E}_1 \in \mathbb{R}^{n\times n}$ are free to vary and $\mathbf{Q}_{\mathbf{W}} = \mathbf{I}_n - \mathbf{P}_{\mathbf{W}}$. It is worth emphasizing that the matrix $\mathbf{G}$ in (2.14b) is a kind of extended version of the projector: it is a projector onto $\mathscr{C}(\mathbf{X})$ along $\mathscr{C}(\mathbf{VM})$. The same concerns any member of the class $\{\mathbf{P}_{\mathbf{X}|\mathscr{M}}\}$, i.e., any matrix of the form $\mathbf{G} + \mathbf{E}_1\mathbf{Q}_{\mathbf{W}}$. For generalized projectors, see, e.g., Rao (1974), Kala (1981), and Puntanen et al. (2011, Sec. 2.5).

Under the consistency of $\mathscr{M}$, for a given $\mathbf{y} \in \mathscr{C}(\mathbf{X} : \mathbf{V})$, we can write $\mathbf{y}$ as

$$\mathbf{y} = \mathbf{Xa} + \mathbf{VMb} \quad \text{for some } \mathbf{a} \in \mathbb{R}^p \text{ and } \mathbf{b} \in \mathbb{R}^n, \tag{2.17}$$

where $\mathbf{Xa}$ and $\mathbf{VMb}$ are unique. Thus, in view of Lemma 1.2, the observed numerical value of the BLUE is uniquely $\mathbf{By} = \mathbf{X}_*\mathbf{a}$ even though $\mathbf{B} \in \{\mathbf{P}_{\mathbf{X}_*|\mathscr{M}}\}$ may not be unique; $\mathbf{B}$ is unique if and only if $\mathscr{C}(\mathbf{W}) = \mathbb{R}^n$. The properties of the BLUE deserve particular attention when $\mathscr{C}(\mathbf{W}) = \mathbb{R}^n$ does not hold: then there is an infinite number of multipliers $\mathbf{B}$ such that $\mathbf{By}$ is BLUE.

By Lemma 1.3 and taking into account the consistency of the model $\mathscr{M}$, we obtain, for $\mathbf{W} = \mathbf{V} + \mathbf{XUU}'\mathbf{X}' \in \mathcal{W}$,

$$\tilde{\boldsymbol{\mu}} = \mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta}) = \mathbf{X}(\mathbf{X}'\mathbf{W}^-\mathbf{X})^-\mathbf{X}'\mathbf{W}^-\mathbf{y} = [\mathbf{I}_n - \mathbf{VM}(\mathbf{MVM})^-\mathbf{M}]\mathbf{y}, \tag{2.18}$$

which further gives the BLUE's residual

$$\mathbf{y} - \mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta}) = \mathbf{VM}(\mathbf{MVM})^-\mathbf{My}, \tag{2.19}$$

and the covariance matrix of $\tilde{\boldsymbol{\mu}}$, cf. part (b), Lemma 1.3,

$$\operatorname{cov}(\tilde{\boldsymbol{\mu}}) = \mathbf{X}(\mathbf{X}'\mathbf{W}^{-}\mathbf{X})^{-}\mathbf{X}' - \mathbf{X}\mathbf{U}\mathbf{U}'\mathbf{X} = \mathbf{V} - \mathbf{V}\mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}\mathbf{V}. \tag{2.20}$$

From (2.9) we observe that $\mathbf{D}\mathbf{y}$ is the BLUP for $\boldsymbol{\varepsilon}_*$ if $\mathbf{D} = \mathbf{L}\mathbf{M}$ for some matrix $\mathbf{L} \in \mathbb{R}^{q\times n}$ such that $\mathbf{L}\mathbf{M}\mathbf{V}\mathbf{M} = \mathbf{V}_{21}\mathbf{M}$, from which one solution to $\mathbf{L}$ is $\mathbf{L} = \mathbf{V}_{21}\mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}$ yielding the following expression:

$$\operatorname{BLUP}(\boldsymbol{\varepsilon}_*) = \mathbf{D}\mathbf{y} = \mathbf{V}_{21}\mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}\mathbf{y} = \mathbf{V}_{21}\dot{\mathbf{M}}\mathbf{y}, \tag{2.21}$$

where $\dot{\mathbf{M}} = \mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}$. By (2.18) and (2.19), we have, for example, the following further representations, see Haslett et al. (2014, Th. 2):

$$\begin{aligned}\operatorname{BLUP}(\boldsymbol{\varepsilon}_*) &= \mathbf{V}_{21}\mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}\mathbf{y} = \mathbf{V}_{21}\mathbf{V}^{-}\mathbf{V}\mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}\mathbf{y}\\ &= \mathbf{V}_{21}\mathbf{W}^{-}\mathbf{W}\mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}\mathbf{y} = \mathbf{V}_{21}\mathbf{V}^{-}[\mathbf{y} - \operatorname{BLUE}(\mathbf{X}\boldsymbol{\beta})]\\ &= \mathbf{V}_{21}\mathbf{V}^{-}(\mathbf{I}_n - \mathbf{G})\mathbf{y} = \mathbf{V}_{21}\mathbf{W}^{-}(\mathbf{I}_n - \mathbf{G})\mathbf{y},\end{aligned} \tag{2.22}$$

where $\mathbf{G} = \mathbf{X}(\mathbf{X}'\mathbf{W}^{-}\mathbf{X})^{-}\mathbf{X}'\mathbf{W}^{-}$ and $\mathbf{V}_{21}\mathbf{V}^{-}\mathbf{V} = \mathbf{V}_{21}\mathbf{W}^{-}\mathbf{W} = \mathbf{V}_{21}$.

If $\mathbf{V}$ is positive definite and $\operatorname{r}(\mathbf{X}) = p$, as in Goldberger (1962), we obtain

$$\begin{aligned}\operatorname{BLUP}(\mathbf{y}_*) &= \operatorname{BLUE}(\mathbf{X}_*\boldsymbol{\beta}) + \operatorname{BLUP}(\boldsymbol{\varepsilon}_*)\\ &= \mathbf{X}_*\tilde{\boldsymbol{\beta}} + \mathbf{V}_{21}\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}})\\ &= \mathbf{X}_*\tilde{\boldsymbol{\beta}} + \mathbf{V}_{21}\mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}\mathbf{y},\end{aligned} \tag{2.23}$$

where $\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$. Apparently Goldberger (1962) was the first to use the term best linear unbiased predictor.

3 Conditions for linear sufficiency

Let us formally define the concept of linear sufficiency as done by Baksalary and Kala (1981, 1986) and Drygas (1983).

Definition 3.1. *Suppose that $\boldsymbol{\mu}_* = \mathbf{X}_*\boldsymbol{\beta}$, where $\mathbf{X}_* \in \mathbb{R}^{q\times p}$, is estimable under the model $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$. Then a linear statistic $\mathbf{F}\mathbf{y}$, where $\mathbf{F} \in \mathbb{R}^{f\times n}$, is called linearly sufficient for $\boldsymbol{\mu}_*$ if there exists a matrix $\mathbf{A} \in \mathbb{R}^{q\times f}$ such that $\mathbf{A}\mathbf{F}\mathbf{y}$ is the* BLUE *for $\boldsymbol{\mu}_*$; that is, there exists a matrix $\mathbf{A}$ such that*

$$\mathbf{A}\mathbf{F}(\mathbf{X} : \mathbf{V}\mathbf{M}) = (\mathbf{X}_* : \mathbf{0})\,. \tag{3.1}$$

Of course, $\mathbf{F}\mathbf{y}$ is linearly sufficient for $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$, if there exists a matrix $\mathbf{A} \in \mathbb{R}^{n\times f}$ such that $\mathbf{A}\mathbf{F}(\mathbf{X} : \mathbf{V}\mathbf{M}) = (\mathbf{X} : \mathbf{0})$.

The concept of linear prediction sufficiency is defined in the corresponding way:

Definition 3.2. *Let $\mathbf{y}_* = \mathbf{X}_*\boldsymbol{\beta} + \boldsymbol{\varepsilon}_*$ be predictable under the model $\mathscr{M}_*$, i.e., $\mathscr{C}(\mathbf{X}_*') \subset \mathscr{C}(\mathbf{X}')$. Then $\mathbf{F}\mathbf{y}$ is called linearly prediction sufficient for $\mathbf{y}_*$ if there* *exists a matrix* $\mathbf{A}$ *such that* $\mathbf{AFy}$ *is the* BLUP *for* $\mathbf{y}_*$*; that is, there exists a matrix* $\mathbf{A}$ *such that*

$$\mathbf{AF}(\mathbf{X} : \mathbf{VM}) = (\mathbf{X}_* : \mathbf{V}_{21}\mathbf{M})\,. \tag{3.2}$$

Moreover, $\mathbf{Fy}$ *is linearly prediction sufficient for* $\boldsymbol{\varepsilon}_*$ *if there exists a matrix* $\mathbf{A}$ *such that* $\mathbf{AFy}$ *is the* BLUP *for* $\boldsymbol{\varepsilon}_*$*; that is, there exists a matrix* $\mathbf{A}$ *such that*

$$\mathbf{AF}(\mathbf{X} : \mathbf{VM}) = (\mathbf{0} : \mathbf{V}_{21}\mathbf{M})\,. \tag{3.3}$$

Sometimes we will use the phrases "BLUE-sufficient" and "BLUP-sufficient" when dealing with estimation and with prediction, respectively, and the short notations like $\mathbf{Fy} \in \mathcal{S}(\mathbf{y}_*)$, $\mathbf{Fy} \in \mathcal{S}(\mathbf{X}_*\boldsymbol{\beta})$, or $\mathbf{Fy} \in \mathcal{S}(\boldsymbol{\varepsilon}_*)$. However, the division into BLUE-sufficiency vs. BLUP-sufficiency is not necessary and we can simply refer to linear sufficiency of $\mathbf{Fy}$ with respect to $\mathbf{y}_*$, $\mathbf{X}_*\boldsymbol{\beta}$ or $\boldsymbol{\varepsilon}_*$. Thus we have,

$$\mathcal{S}(\mathbf{y}_*) = \{\mathbf{Fy} : \mathbf{AF}(\mathbf{X} : \mathbf{VM}) = (\mathbf{X}_* : \mathbf{V}_{21}\mathbf{M}) \text{ for some } \mathbf{A} \in \mathbb{R}^{q\times f}\}\,, \tag{3.4a}$$

$$\mathcal{S}(\mathbf{X}_*\boldsymbol{\beta}) = \{\mathbf{Fy} : \mathbf{AF}(\mathbf{X} : \mathbf{VM}) = (\mathbf{X}_* : \mathbf{0}) \text{ for some } \mathbf{A} \in \mathbb{R}^{q\times f}\}\,, \tag{3.4b}$$

$$\mathcal{S}(\boldsymbol{\varepsilon}_*) = \{\mathbf{Fy} : \mathbf{AF}(\mathbf{X} : \mathbf{VM}) = (\mathbf{0} : \mathbf{V}_{21}\mathbf{M}) \text{ for some } \mathbf{A} \in \mathbb{R}^{q\times f}\}\,. \tag{3.4c}$$

As Kala et al. (2017a, Remark 2) point out, the notation of the above type is merely symbolic and it is not meant to refer to a set containing only one element which is a single fixed vector resulting from transformation of an observed vector $\mathbf{y}$, or is a single random vector variable being a specific linear transformation of the random vector $\mathbf{y}$. We are, of course, actually interested in the matrices $\mathbf{F}$ satisfying a certain property.

Gourieroux and Monfort (1980, Sec. 3F) and Baksalary and Kala (1981) considered corresponding problems without using the term "linear sufficiency" which is due to Drygas (1983, Sec. 3). Baksalary and Kala (1981) used the phrase "linear transformations preserving best linear unbiased estimators". For further related references and concepts like minimal sufficiency and linear completeness, see Müller et al. (1984), Baksalary and Mathew (1986), Müller (1987), Baksalary and Drygas (1992), Groß (1998), Isotalo and Puntanen (2006a), and Sengupta and Jammalamadaka (2003, Sec. 11.1). In this paper we shall very briefly handle minimal sufficiency but skip over the linear completeness concept. The concept of linear minimal sufficiency, introduced by Drygas (1983), is defined as follows:

Definition 3.3. *A linear statistic* $\mathbf{Fy}$ *is called linearly minimal sufficient if for any other linearly sufficient statistics* $\mathbf{Sy}$*, there exists a matrix* $\mathbf{A}$ *such that* $\mathbf{Fy} = \mathbf{ASy}$ *almost surely. Notation* $\mathbf{Fy} \in \mathcal{S}_0(\mathbf{X}\boldsymbol{\beta})$ *indicates that* $\mathbf{Fy}$ *is linearly minimal sufficient for* $\mathbf{X}\boldsymbol{\beta}$*.*

The minimal *prediction* sufficiency can be defined in an analogous way; see Isotalo and Puntanen (2006b, Def. 3.2).

Suppose that $\mathbf{F}_0\mathbf{y} \in \mathcal{S}_0(\mathbf{X}\boldsymbol{\beta})$ and that $\mathbf{Sy}$ is an arbitrary member of $\mathcal{S}(\mathbf{X}\boldsymbol{\beta})$. Then

$$\mathscr{C}(\mathbf{X}) = \mathscr{C}(\mathbf{WF}_0')\,, \text{ and } \mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{WS}')\,, \text{ where } \mathbf{W} \in \mathcal{W}. \tag{3.5}$$

By Definition 3.3, there exists a matrix $\mathbf{A}$ such that $\mathbf{F}_0\mathbf{y} = \mathbf{A}\mathbf{S}\mathbf{y}$ almost surely, i.e., $\mathbf{F}_0\mathbf{W} = \mathbf{A}\mathbf{S}\mathbf{W}$, which further means that

$$\mathscr{C}(\mathbf{X}) = \mathscr{C}(\mathbf{W}\mathbf{F}_0') \subset \mathscr{C}(\mathbf{W}\mathbf{S}')\,. \tag{3.6}$$

The column space of $\mathbf{W}$ can be termed as the general linear model subspace while $\mathscr{C}(\mathbf{X})$ determines the expectation subspace of $\mathscr{M}$. Postmultiplying $\mathbf{W}$ by any matrix does not increase its rank. Therefore, the relation (3.6) means that the transformation $\mathbf{F}_0$ leads to the maximum possible reduction of $\mathscr{C}(\mathbf{W}) = \mathscr{C}(\mathbf{X} : \mathbf{V})$ to the expectation subspace $\mathscr{C}(\mathbf{X})$ in which the $\mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta})$ is contained. Simultaneously, $\mathbf{F}_0\mathbf{y}$ is the smallest amount of information necessary to reconstruct the $\mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta})$.

Remark 3.1. [Linear error-sufficiency.] Groß (1998) introduced the notion of linear error-sufficiency while considering linear sufficient statistics for the prediction of the random error term $\boldsymbol{\varepsilon}$ in the general linear model. As pointed out by Isotalo et al. (2018, Sec. 3), this is nothing but the BLUP-sufficiency of $\boldsymbol{\varepsilon}$. Namely, if we, instead of $\boldsymbol{\varepsilon}_*$, wish to find the BLUP for $\boldsymbol{\varepsilon}$, we have to minimize (in Löwner sense) $\operatorname{cov}(\boldsymbol{\varepsilon} - \mathbf{A}\mathbf{y})$ subject to $\mathbf{A}\mathbf{y}$ being unbiased for $\boldsymbol{\varepsilon}$, that is, parallel to (2.5), the unbiased $\mathbf{A}\mathbf{y}$ must satisfy

$$\operatorname{cov}(\boldsymbol{\varepsilon} - \mathbf{A}\mathbf{y}) \leq_{\mathrm{L}} \operatorname{cov}(\boldsymbol{\varepsilon} - \mathbf{K}\mathbf{M}\mathbf{y}) \quad \text{for all } \mathbf{K} \in \mathbb{R}^{n\times n}. \tag{3.7}$$

Corresponding to (2.9), the matrix $\mathbf{A}$ is a solution to $\mathbf{A}(\mathbf{X} : \mathbf{V}\mathbf{M}) = (\mathbf{0} : \mathbf{V}\mathbf{M})$, and thus

$$\mathrm{BLUP}(\boldsymbol{\varepsilon}) = \mathbf{V}\mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}\mathbf{y}, \tag{3.8}$$

and on account of (2.19) we have the decomposition

$$\mathbf{y} = \mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta}) + \mathrm{BLUP}(\boldsymbol{\varepsilon})\,. \tag{3.9}$$

For the BLUP of $\boldsymbol{\varepsilon}$, see also Arendacká and Puntanen (2015, Lemma 1). □

In Proposition 3.1 we collect some well-known equivalent conditions for $\mathbf{F}\mathbf{y}$ being linearly sufficient for $\mathbf{X}\boldsymbol{\beta}$. For the proofs of parts (c) and (d), see Baksalary and Kala (1981). [Actually, according to Drygas (1983, p. 92), (c) was originally proved by Baksalary and Kala (1978b).] For part (e), see Baksalary and Kala (1986, Cor. 2); and part (f), Müller (1987, Prop. 3.1a). For further related references, see Drygas (1983), Baksalary and Mathew (1986), Baksalary and Drygas (1992), Groß (1998), Isotalo and Puntanen (2006a,b), Kornacki (2007), and Kala and Pordzik (2009).

Proposition 3.1. *The statistic* $\mathbf{F}\mathbf{y}$ *is linearly sufficient for* $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ *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) $\mathscr{C}\begin{pmatrix}\mathbf{X}'\\ \mathbf{0}\end{pmatrix} \subset \mathscr{C}\begin{pmatrix}\mathbf{X}'\mathbf{F}'\\ \mathbf{M}\mathbf{V}\mathbf{F}'\end{pmatrix}$,

(b) $\mathscr{N}(\mathbf{F}\mathbf{X} : \mathbf{F}\mathbf{V}\mathbf{M}) \subset \mathscr{N}(\mathbf{X} : \mathbf{0})$,

(c) $\mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{WF}')$, *where* $\mathbf{W} \in \mathcal{W}$,

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

(e) $\mathscr{C}(\mathbf{X}'\mathbf{F}') = \mathscr{C}(\mathbf{X}')$ *and* $\mathscr{C}(\mathbf{FX}) \cap \mathscr{C}(\mathbf{FVM}) = \{\mathbf{0}\}$,

(f) $\mathscr{N}(\mathbf{F}) \cap \mathscr{C}(\mathbf{X} : \mathbf{V}) \subset \mathscr{C}(\mathbf{VM})$.

Moreover, $\mathbf{Fy}$ *is linearly minimal sufficient for* $\mathbf{X}\boldsymbol{\beta}$ *if and only if* $\mathscr{C}(\mathbf{X}) = \mathscr{C}(\mathbf{WF}')$, *or equivalently, the equality holds in* (a), (b) *or* (f).

Example 3.1. [When is $\mathbf{X}'\mathbf{y}$ linearly sufficient?] Following Kala et al. (2017b, Sec. 4), let us pose a question under which condition the statistic $\mathbf{X}'\mathbf{y}$ is linearly sufficient for $\mathbf{X}\boldsymbol{\beta}$ under $\mathscr{M}$. Now $\mathbf{X}'\mathbf{y} \in \mathcal{S}(\mathbf{X}\boldsymbol{\beta})$ if and only if $\mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{WX})$, where $\mathbf{W} = \mathbf{V} + \mathbf{XX}'$, i.e., $\mathscr{C}(\mathbf{X}) = \mathscr{C}(\mathbf{WX})$, which, noting that we always have $\mathrm{r}(\mathbf{WX}) = \mathrm{r}(\mathbf{X})$, further holds if and only if

$$\mathbf{P_X WX} = \mathbf{WX}\,. \tag{3.10}$$

Rewriting (3.10) yields $\mathbf{P_X}(\mathbf{V} + \mathbf{XX}')\mathbf{X} = (\mathbf{V} + \mathbf{XX}')\mathbf{X}$, i.e., $\mathbf{P_X VX} = \mathbf{VX}$, or in other words,

$$\mathscr{C}(\mathbf{VX}) \subset \mathscr{C}(\mathbf{X})\,. \tag{3.11}$$

The column space inclusion (3.11) is the well-known necessary and sufficient condition for the equality between the BLUE of $\mathbf{X}\boldsymbol{\beta}$ and the ordinary least-squares estimator, OLSE, of $\mathbf{X}\boldsymbol{\beta}$ under the model $\mathscr{M}$; see, e.g., Rao (1967), Zyskind (1967), and Puntanen and Styan (1989). We can express our conclusion as follows:

Proposition 3.2. *The statistic* $\mathbf{X}'\mathbf{y}$ *is linearly sufficient for* $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ *under the model* $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$ *if and only if* $\mathrm{OLSE}(\mathbf{X}\boldsymbol{\beta}) = \mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta})$. *In this situation* $\mathbf{X}'\mathbf{y}$ *is linearly minimal sufficient.*

The corresponding result as in Proposition 3.2, for a positive definite $\mathbf{V}$, appears also in Baksalary and Kala (1981, p. 913). We may further mention that requesting $\mathbf{P_X y} = \mathrm{OLSE}(\mathbf{X}\boldsymbol{\beta})$ to be linearly sufficient for $\mathbf{X}\boldsymbol{\beta}$ leads to the same condition as in Proposition 3.2, i.e.,

$$\mathbf{P_X y} \in \mathcal{S}(\mathbf{X}\boldsymbol{\beta}) \iff \mathbf{P_X y} = \mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta})\,. \tag{3.12}$$

As a simple special case of the linear sufficiency of $\mathbf{X}'\mathbf{y}$, let us consider the model $\mathscr{A} = \{\mathbf{y}, \mathbf{1}\alpha, \mathbf{V}\}$, where $\mathbf{V}$ is positive definite and $\mathbf{1}$ is a vector of ones. We know that the BLUE of α is $\tilde{\alpha} = (\mathbf{1}'\mathbf{V}^{-1}\mathbf{1})^{-1}\mathbf{1}'\mathbf{V}^{-1}\mathbf{y}$. Now $\mathbf{1}'\mathbf{y} \in \mathcal{S}(\alpha)$ if and only if there exists a scalar a such that $\tilde{\alpha} = a\,\mathbf{1}'\mathbf{y}$ for all $\mathbf{y} \in \mathbb{R}^n$, i.e.,

$$(\mathbf{1}'\mathbf{V}^{-1}\mathbf{1})^{-1}\mathbf{1}'\mathbf{V}^{-1} = a\,\mathbf{1}'. \tag{3.13}$$

The equation (3.13) means that $\mathbf{1}$ is an eigenvector of $\mathbf{V}$, i.e.,

$$\mathbf{V1} = \lambda\mathbf{1} \quad \text{for some } \lambda \in \mathbb{R}\,, \tag{3.14}$$

which corresponding to (3.11) can be expressed as that $\mathscr{C}(\mathbf{V1}) = \mathscr{C}(\mathbf{1})$. □

Baksalary and Kala (1986) proved parts (a)–(e) of the following theorem. For other claims, see Kala et al. (2017a) and Kala et al. (2017b).

Proposition 3.3. *Let $\boldsymbol{\mu}_* = \mathbf{X}_*\boldsymbol{\beta}$ be an estimable parametric function under $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$, i.e., $\mathscr{C}(\mathbf{X}_*') \subset \mathscr{C}(\mathbf{X}')$. Then the following statements hold.*

(a) $\mathbf{Fy}$ *is linearly sufficient for* $\mathbf{X}_*\boldsymbol{\beta}$ *under* $\mathscr{M}$ *if and only if any of the following equivalent statements holds:*

(a$_1$) $\mathscr{C}\begin{pmatrix}\mathbf{X}_*' \\ \mathbf{0}\end{pmatrix} \subset \mathscr{C}\begin{pmatrix}\mathbf{X}'\mathbf{F}' \\ \mathbf{MVF}'\end{pmatrix}$,

(a$_2$) $\mathscr{N}(\mathbf{FX} : \mathbf{FVX}^{\perp}) \subset \mathscr{N}(\mathbf{X}_* : \mathbf{0})$,

(a$_3$) $\mathscr{C}[\mathbf{X}(\mathbf{X}'\mathbf{W}^-\mathbf{X})^-\mathbf{X}_*'] \subset \mathscr{C}(\mathbf{WF}')$, *where* $\mathbf{W} \in \mathcal{W}$.

(b) $\mathbf{Fy}$ *is linearly minimal sufficient for* $\mathbf{X}_*\boldsymbol{\beta}$ *if and only if equality holds in* (a$_1$), (a$_2$) *or equivalently in* (a$_3$).

(c) *If* $\mathscr{C}(\mathbf{X}'\mathbf{F}') = \mathscr{C}(\mathbf{X}_*')$, *then* $\mathbf{Fy} \in \mathcal{S}(\mathbf{X}_*\boldsymbol{\beta}) \iff \mathscr{C}(\mathbf{FX}) \cap \mathscr{C}(\mathbf{FVM}) = \{\mathbf{0}\}$.

(d) $\mathbf{Fy} \in \mathcal{S}(\mathbf{FX}\boldsymbol{\beta}) \iff \mathscr{C}(\mathbf{FX}) \cap \mathscr{C}(\mathbf{FVM}) = \{\mathbf{0}\}$.

(e) $\mathbf{Fy} \in \mathcal{S}(\mathbf{X}_*\boldsymbol{\beta})$ *for every estimable* $\mathbf{X}_*\boldsymbol{\beta}$ *if and only if* $\mathbf{Fy} \in \mathcal{S}(\mathbf{X}\boldsymbol{\beta})$.

(f) $\mathscr{C}(\mathbf{W}) \cap \mathscr{C}(\mathbf{F}')^{\perp} = \{\mathbf{0}\} \implies \mathbf{Fy} \in \mathcal{S}(\mathbf{X}_*\boldsymbol{\beta})$; *holds e.g. if* $\mathbf{F}$ *is invertible.*

(g) $\mathrm{r}(\mathbf{LF}) = \mathrm{r}(\mathbf{F}) \implies \mathbf{LFy} \in \mathcal{S}(\mathbf{X}_*\boldsymbol{\beta})$.

(h) $\mathscr{C}(\mathbf{F}') \subset \mathscr{C}(\mathbf{F}_1') \implies \mathbf{F}_1\mathbf{y} \in \mathcal{S}(\mathbf{X}_*\boldsymbol{\beta})$.

(i) $(\mathbf{F}' : \mathbf{F}_2')'\mathbf{y}$ *is linearly sufficient for* $\mathbf{X}_*\boldsymbol{\beta}$ *for any conformable* $\mathbf{F}_2$.

As a curiosity, we may mention that Tian (2013, Th. 3.1) and Tian and Puntanen (2009, Th. 2.8) express (a$_1$) of Proposition 3.3 in the form

$$\mathscr{C}\begin{pmatrix}\mathbf{X}_*' \\ \mathbf{0}\end{pmatrix} \subset \mathscr{C}\begin{pmatrix}\mathbf{X}'\mathbf{F}' & \mathbf{0} \\ \mathbf{VF}' & \mathbf{X}\end{pmatrix}. \tag{3.15}$$

This follows at once from

$$\begin{pmatrix}\mathbf{X}'\mathbf{F}' \\ \mathbf{MVF}'\end{pmatrix} = \begin{pmatrix}\mathbf{X}'\mathbf{F}' & \mathbf{0} \\ \mathbf{VF}' & \mathbf{X}\end{pmatrix}\begin{pmatrix}\mathbf{I}_f \\ -\mathbf{X}^+\mathbf{VF}'\end{pmatrix}. \tag{3.16}$$

It is also noteworthy that in view of part (h) of Proposition 3.3, it is the basis of $\mathscr{C}(\mathbf{F}') \subset \mathbb{R}^n$ that matters, that is, we can choose the columns of $\mathbf{F}'$ linearly independent but spanning $\mathscr{C}(\mathbf{F}')$. In this context, we might ask: how many rows there are in $\mathbf{F}$, i.e., what is f? Suppose that $\mathbf{Fy} \in \mathcal{S}(\mathbf{X}_*\boldsymbol{\beta})$ and $\mathrm{r}(\mathbf{F}) = f$. Then by part (a$_3$) of Proposition 3.3,

$$\mathrm{r}[\mathbf{X}(\mathbf{X}'\mathbf{W}^-\mathbf{X})^-\mathbf{X}_*'] = \mathrm{r}(\mathbf{X}_*) \leq \mathrm{r}(\mathbf{WF}') \leq f \leq n. \tag{3.17}$$

For parts (a) and (b) of the following Proposition, see Isotalo and Puntanen (2006b), and Isotalo et al. (2018), respectively.

Proposition 3.4. *Consider the linear model with new observations defined as* $\mathscr{M}_*$ *in* (1.10)*, where* $\mathscr{C}(\mathbf{X}_*') \subset \mathscr{C}(\mathbf{X}')$*, i.e.,* $\mathbf{y}_*$ *is predictable.*

(a) $\mathbf{Fy} \in \mathcal{S}(\mathbf{y}_*)$ *if and only if any of the following equivalent conditions holds:*

(i) $\mathscr{C}\begin{pmatrix} \mathbf{X}_*' \\ \mathbf{MV}_{12} \end{pmatrix} \subset \mathscr{C}\begin{pmatrix} \mathbf{X'F'} \\ \mathbf{MVF'} \end{pmatrix}$,

(ii) $\mathscr{N}(\mathbf{FX} : \mathbf{FVM}) \subset \mathscr{N}(\mathbf{X}_* : \mathbf{V}_{21}\mathbf{M})$,

(iii) $\mathscr{N}(\mathbf{F}) \cap \mathscr{C}(\mathbf{W}) \subset \mathscr{C}\Big((\mathbf{X} : \mathbf{VM})\big(\begin{smallmatrix} \mathbf{X}_*' \\ \mathbf{MV}_{12} \end{smallmatrix}\big)^{\perp}\Big)$.

(b) $\mathbf{Fy} \in \mathcal{S}(\boldsymbol{\varepsilon}_*)$ *if and only if any of the following equivalent conditions holds:*

(iv) $\mathscr{C}\begin{pmatrix} \mathbf{0} \\ \mathbf{MV}_{12} \end{pmatrix} \subset \mathscr{C}\begin{pmatrix} \mathbf{X'F'} \\ \mathbf{MVF'} \end{pmatrix}$,

(v) $\mathscr{C}(\mathbf{MV}_{12}) \subset \mathscr{C}(\mathbf{MVF'Q_{FX}})$.

In particular, if $\mathbf{Fy} \in \mathcal{S}(\mathbf{X}\boldsymbol{\beta})$*, then* (iv) *becomes*

(vi) $\mathscr{C}(\mathbf{MV}_{12}) \subset \mathscr{C}(\mathbf{MVF'})$.

Moreover, the minimal prediction sufficiency above is obtained if and only if the corresponding inclusion is equality.

Remark 3.2. [Properties of $\mathscr{C}(\mathbf{WF'})$.] Consider the following three questions (a), (b) and (c) related to the linear sufficiency condition (c) of Proposition 3.1:

$$\mathbf{Fy} \in \mathcal{S}(\mathbf{X}\boldsymbol{\beta}) \iff \mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{WF'}), \quad \text{where } \mathbf{W} \in \mathcal{W}. \tag{3.18}$$

(a) The matrix $\mathbf{W}$ in (3.18) belongs to the set $\mathcal{W}$ of (symmetric) nonnegative definite matrices. One question: is the column space $\mathscr{C}(\mathbf{WF'})$ unique, i.e., does it remain invariant for any choice of $\mathbf{W} \in \mathcal{W}$? It might be somewhat tempting to conjecture that for a given $\mathbf{F}$, the column space $\mathscr{C}(\mathbf{WF'})$ would be invariant. However, Kala et al. (2017b, Ex. 1) provide a counterexample showing that this is not the case.

(b) Kala et al. (2017b, Sec. 4) also studied whether the column space $\mathscr{C}(\mathbf{WF'})$ is invariant for any choice of $\mathbf{W}$ if $\mathbf{Fy} \in \mathcal{S}(\mathbf{X}\boldsymbol{\beta})$. The Proposition 3.5 is a reply to this question. We formulate it in a more general setup using the set $\mathcal{W}^{\#}$ instead of $\mathcal{W}$.

Proposition 3.5. *Consider the linear model* $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$*, let* $\mathbf{W} \in \mathcal{W}^{\#}$ *and suppose that* $\mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{WF'})$*. Then the column space* $\mathscr{C}(\mathbf{WF'})$ *is invariant for any choice of* $\mathbf{W} \in \mathcal{W}^{\#}$ *and*

$$\mathscr{C}(\mathbf{WF'}) = \mathscr{C}(\mathbf{X}) \oplus \mathscr{C}(\mathbf{MVF'}) = \mathscr{C}(\mathbf{W'F'}). \tag{3.19}$$

(c) Kala et al. (2017b, Sec. 4) were wondering whether in (3.18) the set $\mathcal{W}$ can be replaced with the more general set $\mathcal{W}^{\#}$. The answer is interestingly (but not trivially) positive and can be expressed as follows.

Proposition 3.6. *Let* $\mathbf{W} \in \mathcal{W}^{\#}$*. Then 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*

$$\mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{WF}')\,, \quad \textit{or, equivalently,} \quad \mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{W}'\mathbf{F}')\,. \tag{3.20}$$

According to our knowledge, 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. Of course, sometimes it can be simpler to have $\mathbf{W}$ from set $\mathcal{W}$. For detailed proofs of Propositions 3.5 and 3.6, see Kala et al. (2017b, Sec. 4). $\square$

Example 3.2. [Random walk and linear sufficiency.] One interesting example of the BLUP is described by Isotalo and Puntanen (2006b, p. 1021) and Haslett et al. (2014, Sec. 1). They consider the model

$$y_t = \beta t + \varepsilon_t\,, \tag{3.21}$$

where t denotes discrete time variable, β is the unknown parameter and ε_t is a random walk process, see, e.g., Davidson and MacKinnon (2004, p. 606), with a form $\varepsilon_t = \varepsilon_{t-1} + u_t$, $\varepsilon_0 = 0$, $u_t \sim \mathrm{IID}(0,1)$. Suppose that the time series y_t given in (3.21) is observable at times of $t = 1, 2, 3, 4$. Then putting $\mathbf{x} = (1,2,3,4)'$, we have in matrix terms

$$\mathbf{y} = \mathbf{x}\beta + \boldsymbol{\varepsilon}, \text{ where } \boldsymbol{\varepsilon} = \mathbf{Du}\,,\ \mathbf{D} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix},\ \mathrm{cov}(\mathbf{u}) = \mathbf{I}_4\,. \tag{3.22}$$

Being interested in predicting the outcome of y_t at time of $t = 5$ based on observable random variables $y_1, \dots, y_4$, we can write the model as

$$\mathscr{M}_* = \left\{ \begin{pmatrix} \mathbf{y} \\ y_* \end{pmatrix}, \begin{pmatrix} \mathbf{x}\beta \\ x_*\beta \end{pmatrix}, \begin{pmatrix} \mathbf{V} & \mathbf{v}_{12} \\ \mathbf{v}_{12}' & v_{22} \end{pmatrix} \right\}, \tag{3.23}$$

where $y_* = y_5$, $x_* = 5$, the variance of y_5 is $v_{22} = 5$, and

$$\mathrm{cov}(\mathbf{y}) = \mathbf{V} = \mathbf{DD}' = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 \\ 1 & 2 & 3 & 3 \\ 1 & 2 & 3 & 4 \end{pmatrix},\ \mathrm{cov}(\mathbf{y}, y_5) = \mathbf{v}_{12} = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}. \tag{3.24}$$

Now $\mathbf{x} = \mathbf{v}_{12}$, and thereby $\mathscr{C}(\mathbf{x}^{\perp}) = \mathscr{C}(\mathbf{v}_{12}^{\perp})$ and

$$\mathbf{Q}_{\mathbf{x}} = \mathbf{I}_4 - \mathbf{v}_{12}\mathbf{v}_{12}'/\mathbf{v}_{12}'\mathbf{v}_{12} = \mathbf{I}_4 - \mathbf{Vf}\,\mathbf{f}'\mathbf{V}/\mathbf{f}'\mathbf{V}^2\mathbf{f}\,, \tag{3.25}$$

where $\mathbf{f} = (0,0,0,1)'$ and $\mathbf{v}_{12} = \mathbf{Vf}$. We further have

$$\mathbf{f}'\mathbf{x} = 4\,, \quad \mathbf{f}'\mathbf{V}\mathbf{Q}_{\mathbf{x}} = (0,0,0,0)\,, \quad x_* = 5, \quad \mathbf{v}_{12}'\mathbf{Q}_{\mathbf{x}} = (0,0,0,0)\,. \tag{3.26}$$

Thus

$$\mathscr{N}(\mathbf{f}'\mathbf{x} : \mathbf{f}'\mathbf{V}\mathbf{Q}_{\mathbf{x}}) = \mathscr{N}(x_* : \mathbf{v}_{12}'\mathbf{Q}_{\mathbf{x}})\,, \tag{3.27}$$

which by Proposition 3.4 means that $\mathbf{f}'\mathbf{y} = y_4$ is linearly minimal sufficient predictor for y_5. The BLUP for y_5 in this situation is simply $\frac{5}{4}\,y_4$. $\square$

4 The transformed model

Consider the model $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$ and let $\mathbf{F} \in \mathbb{R}^{f\times n}$ be such a matrix that $\mathbf{Fy}$ is linearly sufficient for $\mathbf{X}\boldsymbol{\beta}$. Then the transformation $\mathbf{F}$ applied to $\mathbf{y}$ induces the transformed model

$$\mathscr{T} = \{\mathbf{Fy}, \mathbf{FX}\boldsymbol{\beta}, \mathbf{FVF}'\}\,. \tag{4.1}$$

As the statistic $\mathbf{Fy}$ is linearly sufficient for $\mathbf{X}\boldsymbol{\beta}$, it sounds intuitively natural that both models provide the same starting point for obtaining the BLUE of $\mathbf{X}\boldsymbol{\beta}$. Indeed this is true as proved by Baksalary and Kala (1981, 1986). We can also do the prediction of $\mathbf{y}_*$ (or $\boldsymbol{\varepsilon}_*$) based on the transformed model using the following setup:

$$\mathscr{T}_* = \left\{ \begin{pmatrix} \mathbf{Fy} \\ \mathbf{y}_* \end{pmatrix}, \begin{pmatrix} \mathbf{FX} \\ \mathbf{X}_* \end{pmatrix} \boldsymbol{\beta}, \begin{pmatrix} \mathbf{FVF}' & \mathbf{FV}_{12} \\ \mathbf{V}_{21}\mathbf{F}' & \mathbf{V}_{22} \end{pmatrix} \right\}. \tag{4.2}$$

Recall that trivially the BLUE-considerations are identical under $\mathscr{M}$ and $\mathscr{M}_*$, i.e., $\mathrm{BLUE}(\mathbf{X}_*\boldsymbol{\beta} \mid \mathscr{M}) = \mathrm{BLUE}(\mathbf{X}_*\boldsymbol{\beta} \mid \mathscr{M}_*)$, etc.

In order to find BLUEs and BLUPs we need to have some estimability conditions.

Proposition 4.1. *Consider the models $\mathscr{M}$ and $\mathscr{T}$. Then the following estimability conditions hold:*

(a) $\mathbf{X}_*\boldsymbol{\beta}$ *is estimable under* $\mathscr{M} \iff \mathscr{C}(\mathbf{X}_*') \subset \mathscr{C}(\mathbf{X}')$,

(b) $\mathbf{X}_*\boldsymbol{\beta}$ *is estimable under* $\mathscr{T} \iff \mathscr{C}(\mathbf{X}_*') \subset \mathscr{C}(\mathbf{X}'\mathbf{F}')$,

(c) $\mathbf{X}\boldsymbol{\beta}$ *is estimable under* $\mathscr{T} \iff \mathscr{C}(\mathbf{X}') = \mathscr{C}(\mathbf{X}'\mathbf{F}')$, *i.e.,* $\mathrm{r}(\mathbf{X}) = \mathrm{r}(\mathbf{FX})$.

Moreover,

(d) *the column space condition in* (b) *implies that one in* (a),

(e) *the column space conditions in* (a) *and* (c) *imply that one in* (b).

Suppose that $\mathbf{X}_*\boldsymbol{\beta}$ is estimable under the transformed model $\mathscr{T}$. Then $\mathbf{CFy}$ is the BLUE for $\mathbf{X}_*\boldsymbol{\beta}$ under $\mathscr{T}$ if and only if $\mathbf{C}$ satisfies the condition

$$\mathbf{C}(\mathbf{FX} : \mathbf{FVF}'\mathbf{Q}_{\mathbf{FX}}) = (\mathbf{X}_* : \mathbf{0})\,, \tag{4.3}$$

or equivalently

$$\mathbf{C}(\mathbf{FX} : \mathbf{FVMF}'\mathbf{Q}_{\mathbf{FX}}) = (\mathbf{X}_* : \mathbf{0})\,, \quad \text{i.e.,} \quad \mathbf{C}(\mathbf{FX} : \mathbf{FVMN}) = (\mathbf{X}_* : \mathbf{0})\,, \tag{4.4}$$

where $\mathbf{Q}_{\mathbf{FX}} = \mathbf{I}_f - \mathbf{P}_{\mathbf{FX}}$, and $\mathbf{N} = \mathbf{P}_{\mathscr{C}(\mathbf{M})\cap\mathscr{C}(\mathbf{F}')}$; see Lemma 1.4. We use the notation

$$\mathbf{C} \in \{\mathbf{P}_{\mathbf{X}_*|\mathscr{T}}\} \iff \mathbf{C}(\mathbf{FX} : \mathbf{FVF}'\mathbf{Q}_{\mathbf{FX}}) = (\mathbf{X}_* : \mathbf{0})\,. \tag{4.5}$$

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

$$\mathbf{D} \in \{\mathbf{P}_{\mathbf{X}_*|\mathscr{T}}\mathbf{F}\} \iff \mathbf{D} = \mathbf{CF}, \text{ where } \mathbf{C} \in \{\mathbf{P}_{\mathbf{X}_*|\mathscr{T}}\}. \tag{4.6}$$

Recall, by (2.13b), that the set $\{\mathbf{P}_{\mathbf{X}_*|\mathscr{M}}\}$ providing the BLUEs for $\mathbf{X}_*\boldsymbol{\beta}$ under the original model $\mathscr{M}$ is defined as

$$\mathbf{B} \in \{\mathbf{P}_{\mathbf{X}_*|\mathscr{M}}\} \iff \mathbf{B}(\mathbf{X} : \mathbf{VM}) = (\mathbf{X}_* : \mathbf{0})\,. \tag{4.7}$$

Baksalary and Kala (1986, Th. 1) consider $\mathbf{X}_*\boldsymbol{\beta}$, which is estimable under $\mathscr{M}$. They assume that $\mathbf{Fy}$ is linearly sufficient for $\mathbf{X}_*\boldsymbol{\beta}$ so that there exists some matrix $\mathbf{A}$ such that $\mathbf{AFy}$ is the BLUE of $\mathbf{X}_*\boldsymbol{\beta}$ in $\mathscr{M}$, i.e., $\mathbf{AF} \in \{\mathbf{P}_{\mathbf{X}_*|\mathscr{M}}\}$ so that

$$\mathbf{AF}(\mathbf{X} : \mathbf{VM}) = (\mathbf{X}_* : \mathbf{0})\,. \tag{4.8}$$

They show that for each $\mathbf{C} \in \{\mathbf{P}_{\mathbf{X}_*|\mathscr{T}}\}$ satisfying the equation

$$\mathbf{C}(\mathbf{FX} : \mathbf{FVF}'\mathbf{Q}_{\mathbf{FX}}) = (\mathbf{X}_* : \mathbf{0})\,, \tag{4.9}$$

the equality

$$\mathbf{AF}(\mathbf{X} : \mathbf{V}) = \mathbf{CF}(\mathbf{X} : \mathbf{V}) \tag{4.10}$$

holds and thereby $\mathbf{AFy}$ equals $\mathbf{CFy}$ with probability 1. In other words,

$$\mathrm{BLUE}(\mathbf{X}_*\boldsymbol{\beta} \mid \mathscr{M}) = \mathrm{BLUE}(\mathbf{X}_*\boldsymbol{\beta} \mid \mathscr{T}) \quad \text{with probability 1.} \tag{4.11}$$

It is worth emphasizing that (4.10) holds for *every* $\mathbf{C} \in \{\mathbf{P}_{\mathbf{X}_*|\mathscr{T}}\}$ and *at least for one matrix* $\mathbf{A}$.

Equality (4.10) also implies that

$$\mathbf{AF}(\mathbf{X} : \mathbf{VM}) = \mathbf{CF}(\mathbf{X} : \mathbf{VM}) = (\mathbf{X}_* : \mathbf{0})\,. \tag{4.12}$$

Now (4.12) means that every multiplier $\mathbf{CF}$ of $\mathbf{y}$ (where $\mathbf{C} \in \{\mathbf{P}_{\mathbf{X}_*|\mathscr{T}}\}$) provides the BLUE under the original model $\mathscr{M}$. In other notation,

$$\mathbf{Fy} \in \mathcal{S}(\mathbf{X}_*\boldsymbol{\beta}) \implies \{\mathbf{P}_{\mathbf{X}_*|\mathscr{T}}\mathbf{F}\} \subset \{\mathbf{P}_{\mathbf{X}_*|\mathscr{M}}\}\,. \tag{4.13}$$

It is obvious that the implication in (4.13) holds also in the reverse direction.

Moreover, we can conclude the following: there exists at least one representation of BLUE under $\mathscr{M}$ [this is $\mathbf{AFy}$, where $\mathbf{A}$ satisfies (4.12)] which is BLUE also under the transformed model $\mathscr{T}$. To confirm this we have to show that

$$\mathbf{AF}(\mathbf{X} : \mathbf{VM}) = \mathbf{A}(\mathbf{FX} : \mathbf{FVM}) = (\mathbf{X}_* : \mathbf{0}) \tag{4.14}$$

implies

$$\mathbf{A}(\mathbf{FX} : \mathbf{FVF}'\mathbf{Q}_{\mathbf{FX}}) = \mathbf{A}(\mathbf{FX} : \mathbf{FVMF}'\mathbf{Q}_{\mathbf{FX}}) = (\mathbf{X}_* : \mathbf{0})\,. \tag{4.15}$$

where we have used Lemma 1.4. Now it is obvious that (4.14) implies (4.15). The feature that there exists at least one representation of BLUE under $\mathscr{M}$ which is BLUE also under the transformed model $\mathscr{T}$ can be denoted as

$$\{\mathbf{P}_{\mathbf{X}_*|\mathscr{T}}\mathbf{F}\} \cap \{\mathbf{P}_{\mathbf{X}_*|\mathscr{M}}\} \neq \{\emptyset\}. \tag{4.16}$$

As a matter of fact, the above proof provides the following result, where the concept of linear sufficiency is not explicitly present:

Proposition 4.2. *Suppose that $\mathbf{X}_*\boldsymbol{\beta}$ is estimable under $\mathscr{M}$. Then*

(a) $\mathbf{AF} \in \{\mathbf{P}_{\mathbf{X}_*|\mathscr{M}}\} \implies \mathbf{A} \in \{\mathbf{P}_{\mathbf{X}_*|\mathscr{T}}\}$,

(b) $\mathbf{AF} \in \{\mathbf{P}_{\boldsymbol{\varepsilon}_*|\mathscr{M}_*}\} \implies \mathbf{A} \in \{\mathbf{P}_{\boldsymbol{\varepsilon}_*|\mathscr{T}_*}\}$,

(c) $\mathbf{AF} \in \{\mathbf{P}_{\mathbf{y}_*|\mathscr{M}_*}\} \implies \mathbf{A} \in \{\mathbf{P}_{\mathbf{y}_*|\mathscr{T}_*}\}$.

Notice that if (4.8) is solvable for $\mathbf{A}$ then (4.9) is solvable for $\mathbf{C}$ and $\mathbf{X}_*\boldsymbol{\beta}$ is estimable under the transformed model $\mathscr{T}$. Another noteworthy fact is that (4.16) implies that $\mathbf{Fy} \in \mathcal{S}(\mathbf{X}_*\boldsymbol{\beta})$.

The matrix $\mathbf{C} \in \{\mathbf{P}_{\mathbf{X}_*|\mathscr{T}}\}$ satisfying the equation (4.9) is unique if and only if

$$\mathrm{r}(\mathbf{FX} : \mathbf{FVF}'\mathbf{Q}_{\mathbf{FX}}) = f\,, \quad \text{i.e.,} \quad \mathrm{r}(\mathbf{FW}) = f. \tag{4.17}$$

The rank rule of the matrix product of Marsaglia and Styan (1974, Cor. 6.2) gives

$$\mathrm{r}(\mathbf{FW}) = \mathrm{r}(\mathbf{F}) - \dim \mathscr{C}(\mathbf{F}') \cap \mathscr{C}(\mathbf{W})^{\perp}, \tag{4.18}$$

and thus the matrix $\mathbf{C} \in \{\mathbf{P}_{\mathbf{X}_*|\mathscr{T}}\}$ is unique if and only if

$$\mathrm{r}(\mathbf{F}) = f \quad \text{and} \quad \mathscr{C}(\mathbf{Q}_{\mathbf{F}'} : \mathbf{W}) = \mathbb{R}^n. \tag{4.19}$$

Kala et al. (2017a, Th. 1) represented the parts (a)–(f) of the following proposition dealing with statistical implications of the linear sufficiency.

Proposition 4.3. *Suppose that $\mathbf{Fy}$ is linearly sufficient for estimable $\mathbf{X}_*\boldsymbol{\beta}$ under the model $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$ and let $\mathscr{T} = \{\mathbf{Fy}, \mathbf{FX}\boldsymbol{\beta}, \mathbf{FVF}'\}$ denote the transformed model and $\mathbf{W} \in \mathcal{W}$. Then the following statements hold.*

(a) *$\mathbf{X}_*\boldsymbol{\beta}$ is estimable also under the transformed model $\mathscr{T}$, i.e., $\mathscr{C}(\mathbf{X}_*') \subset \mathscr{C}(\mathbf{X}'\mathbf{F}')$.*

(b) *$\mathrm{BLUE}(\mathbf{X}_*\boldsymbol{\beta} \mid \mathscr{M}) = \mathrm{BLUE}(\mathbf{X}_*\boldsymbol{\beta} \mid \mathscr{T})$ with probability 1, in other notation, $\mathbf{BW} = \mathbf{CFW}$, where $\mathbf{B} \in \{\mathbf{P}_{\mathbf{X}_*|\mathscr{M}}\}$, and $\mathbf{C} \in \{\mathbf{P}_{\mathbf{X}_*|\mathscr{T}}\}$.*

(c) *Every representation of the BLUE of $\mathbf{X}_*\boldsymbol{\beta}$ under $\mathscr{T}$ is BLUE also under $\mathscr{M}$, i.e., $\{\mathbf{P}_{\mathbf{X}_*|\mathscr{T}}\mathbf{F}\} \subset \{\mathbf{P}_{\mathbf{X}_*|\mathscr{M}}\}$.*

(d) *There exists at least one representation of BLUE of $\mathbf{X}_*\boldsymbol{\beta}$ under $\mathscr{M}$ which is BLUE also under the transformed model $\mathscr{T}$, i.e., there exists at least one matrix in the set $\{\mathbf{P}_{\mathbf{X}_*|\mathscr{M}}\}$ which belongs also to $\{\mathbf{P}_{\mathbf{X}_*|\mathscr{T}}\mathbf{F}\}$.*

(e) *If* $\mathbf{By}$ *is an arbitrary* BLUE *for* $\mathbf{X}_*\boldsymbol{\beta}$ *under* $\mathscr{M}$*, then there exists* $\mathbf{C}$ *such that* $\mathbf{By} = \mathbf{CFy}$ *for all* $\mathbf{y} \in \mathscr{C}(\mathbf{X} : \mathbf{V})$*, and* $\mathbf{CFy}$ *is the* BLUE *for* $\mathbf{X}_*\boldsymbol{\beta}$ *under* $\mathscr{T}$*.*

(f) *If* $\mathscr{C}(\mathbf{X}_*') = \mathscr{C}(\mathbf{X}'\mathbf{F}')$*, then* $\mathbf{Fy} \in \mathcal{S}(\mathbf{X}_*\boldsymbol{\beta}) \iff \mathbf{Fy} \in \mathcal{S}(\mathbf{FX}\boldsymbol{\beta})$.

(g) *There is only one matrix in the class* $\{\mathbf{P}_{\mathbf{X}_*|\mathscr{T}}\mathbf{F}\}$ *if and only if* $\mathrm{r}(\mathbf{F}) = f$ *and* $\mathscr{C}(\mathbf{Q}_{\mathbf{F}'} : \mathbf{W}) = \mathbb{R}^n$.

Moreover, if $\mathbf{Fy}$ *is linearly sufficient for* $\mathbf{X}\boldsymbol{\beta}$ *then the statements* (a)–(e) *hold also when* $\mathbf{X}_*$ *is replaced with* $\mathbf{X}$.

Proof [of Part (e)]]. Part (e) of the above Proposition 4.3 was given, without a proof, in Kala et al. (2017a, Th. 1). For completeness, we briefly go through the proof. Suppose that $\mathbf{Fy} \in \mathcal{S}(\mathbf{X}_*\boldsymbol{\beta})$ and $\mathbf{B} \in \{\mathbf{P}_{\mathbf{X}_*|\mathscr{M}}\}$. Then the general representation of $\mathbf{B}$ is

$$\mathbf{B}_0 = \mathbf{X}_*(\mathbf{X}'\mathbf{W}^-\mathbf{X})^-\mathbf{X}'\mathbf{W}^- + \mathbf{EQ_W}\,, \tag{4.20}$$

where $\mathbf{E} \in \mathbb{R}^{q\times n}$ is arbitrary. Postmultiplying (4.20) by $\mathbf{W}$ yields

$$\mathbf{B}_0\mathbf{W} = \mathbf{X}_*(\mathbf{X}'\mathbf{W}^-\mathbf{X})^-\mathbf{X}' = \mathbf{CFW}, \tag{4.21}$$

for some $\mathbf{C} \in \mathbb{R}^{q\times f}$; here we have used part (a_3) of Proposition 3.3, which says that under linear sufficiency, $\mathbf{X}(\mathbf{X}'\mathbf{W}^-\mathbf{X})^-\mathbf{X}_*' = \mathbf{WF}'\mathbf{C}'$ for some $\mathbf{C}$. Equality (4.21) means that $\mathbf{B}_0\mathbf{y} = \mathbf{CFy}$ for all $\mathbf{y} \in \mathscr{C}(\mathbf{W})$. It remains to show that $\mathbf{C} \in \{\mathbf{P}_{\mathbf{X}_*|\mathscr{T}}\}$, which immediately follows from Proposition 4.2. □

Next we consider the explicit expression for the BLUE of $\mathbf{X}\boldsymbol{\beta}$ under the transformed model $\mathscr{T}$. To do this we have to assume that $\mathbf{X}\boldsymbol{\beta}$ is estimable under $\mathscr{T}$. Then the statistic $\mathbf{CFy}$ is the BLUE for $\mathbf{X}\boldsymbol{\beta}$ under $\mathscr{T}$ if and only if $\mathbf{C}$ satisfies

$$\mathbf{C}(\mathbf{FX} : \mathbf{FVF}'\mathbf{Q}_{\mathbf{FX}}) = (\mathbf{X} : \mathbf{0})\,, \quad \text{or shortly,} \quad \mathbf{C} \in \{\mathbf{P}_{\mathbf{X}|\mathscr{T}}\}. \tag{4.22}$$

One solution for $\mathbf{C}$ in (4.22) is

$$\mathbf{C}_1 := \mathbf{X}[\mathbf{X}'\mathbf{F}'(\mathbf{FWF}')^-\mathbf{FX}]^-\mathbf{X}'\mathbf{F}'(\mathbf{FWF}')^- \in \{\mathbf{P}_{\mathbf{X}|\mathscr{T}}\}\,. \tag{4.23}$$

Notice that by (1.18), $\mathbf{FWF}'$ is a $\mathbf{W}$-matrix under the model $\mathscr{T}$. Thus, see Kala et al. (2017b, Sec. 6) and Markiewicz and Puntanen (2019a, Sec. 3), the BLUE of $\mathbf{X}\boldsymbol{\beta}$ under $\mathscr{T}$ has, for example, the representation

$$\mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta} \mid \mathscr{T}) = \mathbf{C}_1\mathbf{Fy} = \mathbf{G}_t\mathbf{y}, \tag{4.24}$$

where

$$\mathbf{G}_t = \mathbf{C}_1\mathbf{F} = \mathbf{X}[\mathbf{X}'\mathbf{F}'(\mathbf{FWF}')^-\mathbf{FX}]^-\mathbf{X}'\mathbf{F}'(\mathbf{FWF}')^-\mathbf{F} \in \{\mathbf{P}_{\mathbf{X}|\mathscr{T}}\mathbf{F}\}\,. \tag{4.25}$$

Correspondingly, if $\mathbf{X}_*\boldsymbol{\beta}$ is estimable under $\mathscr{T}$, then $\mathbf{X}_* = \mathbf{LFX}$ for some $\mathbf{L} \in \mathbb{R}^{q\times f}$ and $\mathbf{CFy}$ is the BLUE for $\mathbf{X}_*\boldsymbol{\beta}$ if $\mathbf{C} \in \{\mathbf{P}_{\mathbf{X}_*|\mathscr{T}}\}$. One such a $\mathbf{C}$ is

$$\mathbf{C}_2 := \mathbf{X}_*[\mathbf{X}'\mathbf{F}'(\mathbf{FWF}')^-\mathbf{FX}]^-\mathbf{X}'\mathbf{F}'(\mathbf{FWF}')^- \in \{\mathbf{P}_{\mathbf{X}_*|\mathscr{T}}\}\,, \tag{4.26}$$

and the BLUE of $\mathbf{X}_*\boldsymbol{\beta}$ under $\mathscr{T}$ can be expressed as

$$\mathrm{BLUE}(\mathbf{X}_*\boldsymbol{\beta} \mid \mathscr{T}) = \mathbf{C}_2\mathbf{Fy} = \mathbf{LF}\,\mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta} \mid \mathscr{T}) = \mathbf{LFG}_t\mathbf{y}. \tag{4.27}$$

The general representations of matrices in $\{\mathbf{P}_{\mathbf{X}|\mathscr{T}}\mathbf{F}\}$ and $\{\mathbf{P}_{\mathbf{X}_*|\mathscr{T}}\mathbf{F}\}$ providing the BLUE for $\mathbf{X}\boldsymbol{\beta}$ and $\mathbf{X}_*\boldsymbol{\beta}$, under $\mathscr{T}$, can be expressed as

$$\mathbf{P}_{\mathbf{X}|\mathscr{T}}\mathbf{F} = (\mathbf{C}_1 + \mathbf{E}_2\mathbf{Q}_{\mathbf{FW}})\mathbf{F}, \quad \mathbf{P}_{\mathbf{X}_*|\mathscr{T}}\mathbf{F} = (\mathbf{C}_2 + \mathbf{E}_3\mathbf{Q}_{\mathbf{FW}})\mathbf{F}, \tag{4.28}$$

respectively; here $\mathbf{E}_2 \in \mathbb{R}^{n\times f}$ and $\mathbf{E}_3 \in \mathbb{R}^{q\times f}$ are free to vary.

The statistic $\mathbf{Dy}$ is the BLUP for $\boldsymbol{\varepsilon}_*$ under the transformed model $\mathscr{T}_*$ if and only if $\mathbf{D} = \mathbf{CF}$, where $\mathbf{C}$ satisfies

$$\mathbf{C}(\mathbf{FX} : \mathbf{FVF}'\mathbf{Q}_{\mathbf{FX}}) = (\mathbf{0} : \mathbf{V}_{21}\mathbf{F}'\mathbf{Q}_{\mathbf{FX}})\,, \quad \text{or shortly,} \quad \mathbf{C} \in \{\mathbf{P}_{\boldsymbol{\varepsilon}_*|\mathscr{T}_*}\}. \tag{4.29}$$

Thus the BLUP of $\boldsymbol{\varepsilon}_*$ under $\mathscr{T}_*$ can be expressed as

$$\mathrm{BLUP}(\boldsymbol{\varepsilon}_* \mid \mathscr{T}_*) = \mathbf{V}_{21}\mathbf{F}'\mathbf{Q}_{\mathbf{FX}}(\mathbf{Q}_{\mathbf{FX}}\mathbf{FVF}'\mathbf{Q}_{\mathbf{FX}})^-\mathbf{Q}_{\mathbf{FX}}\mathbf{Fy}. \tag{4.30}$$

Recall that in (4.30), $\mathbf{F}'\mathbf{Q}_{\mathbf{FX}} = \mathbf{MF}'\mathbf{Q}_{\mathbf{FX}}$ and hence

$$\mathbf{Q}_{\mathbf{FX}}\mathbf{FVF}'\mathbf{Q}_{\mathbf{FX}} = \mathbf{Q}_{\mathbf{FX}}\mathbf{FWF}'\mathbf{Q}_{\mathbf{FX}}\,. \tag{4.31}$$

It can further be shown that (4.30) can be expressed also as

$$\mathrm{BLUP}(\boldsymbol{\varepsilon}_* \mid \mathscr{T}_*) = \mathbf{V}_{21}\mathbf{N}(\mathbf{NVN})^-\mathbf{Ny} = \tilde{\boldsymbol{\varepsilon}}_{t*}\,, \tag{4.32}$$

where $\mathbf{N} = \mathbf{P}_{\mathbf{F}'\mathbf{Q}_{\mathbf{FX}}}$. This is based on the fact that if $\mathscr{C}(\mathbf{A}) = \mathscr{C}(\mathbf{B})$, then

$$\mathbf{P}_{\mathbf{W}}\mathbf{A}(\mathbf{A}'\mathbf{WA})^-\mathbf{A}'\mathbf{P}_{\mathbf{W}} = \mathbf{P}_{\mathbf{W}}\mathbf{B}(\mathbf{B}'\mathbf{WB})^-\mathbf{B}'\mathbf{P}_{\mathbf{W}}\,. \tag{4.33}$$

Thus, see also Isotalo et al. (2018, Sec. 4), we have the following proposition.

Proposition 4.4. *Let $\mathbf{y}_*$ be predictable under $\mathscr{M}_*$, so that $\mathbf{X}_* = \mathbf{KX}$ for some $\mathbf{K} \in \mathbb{R}^{q\times n}$, and $\mathbf{G} = \mathbf{X}(\mathbf{X}'\mathbf{W}^-\mathbf{X}')^-\mathbf{X}'\mathbf{W}^-$. Then the* BLUP$(\mathbf{y}_*)$ *under $\mathscr{M}_*$ can be written as*

$$\begin{aligned}
\mathrm{BLUP}(\mathbf{y}_* \mid \mathscr{M}_*) &= \mathrm{BLUE}(\boldsymbol{\mu}_* \mid \mathscr{M}) + \mathbf{V}_{21}\mathbf{V}^-[\mathbf{y} - \mathrm{BLUE}(\boldsymbol{\mu} \mid \mathscr{M})] \\
&= \mathbf{KGy} + \mathbf{V}_{21}\mathbf{V}^-(\mathbf{I}_n - \mathbf{G})\mathbf{y} \\
&= \mathbf{KGy} + \mathbf{V}_{21}\mathbf{W}^-(\mathbf{I}_n - \mathbf{G})\mathbf{y} \\
&= \mathbf{KGy} + \mathbf{V}_{21}\mathbf{M}(\mathbf{MVM})^-\mathbf{My} \\
&= \mathrm{BLUE}(\boldsymbol{\mu}_* \mid \mathscr{M}) + \mathrm{BLUP}(\boldsymbol{\varepsilon}_* \mid \mathscr{M}_*)\,,
\end{aligned} \tag{4.34}$$

or shortly,

$$\tilde{\mathbf{y}}_* = \tilde{\boldsymbol{\mu}}_* + \tilde{\boldsymbol{\varepsilon}}_* \,. \tag{4.35}$$

Let $\mathbf{y}_*$ *be predictable under* $\mathscr{T}_*$, *so that* $\mathbf{X}_* = \mathbf{LFX}$ *for some* $\mathbf{L} \in \mathbb{R}^{q\times f}$, *and* $\mathbf{G}_t$ *is defined as in* (4.25). *Then the* $\mathrm{BLUP}(\mathbf{y}_*)$ *under* $\mathscr{T}_*$ *can be written as*

$$\begin{aligned}\mathrm{BLUP}(\mathbf{y}_* \mid \mathscr{T}_*) &= \mathrm{BLUE}(\boldsymbol{\mu}_* \mid \mathscr{T}) + \mathbf{V}_{21}\mathbf{F}'(\mathbf{FVF}')^{-}\mathbf{F}[\mathbf{y} - \mathrm{BLUE}(\boldsymbol{\mu} \mid \mathscr{T})] \\ &= \mathbf{LFG}_t\mathbf{y} + \mathbf{V}_{21}\mathbf{F}'(\mathbf{FVF}')^{-}\mathbf{F}(\mathbf{I}_n - \mathbf{G}_t)\mathbf{y} \\ &= \mathbf{LFG}_t\mathbf{y} + \mathbf{V}_{21}\mathbf{N}(\mathbf{NVN})^{-}\mathbf{N}\mathbf{y} \\ &= \mathrm{BLUE}(\boldsymbol{\mu}_* \mid \mathscr{T}) + \mathrm{BLUP}(\boldsymbol{\varepsilon}_* \mid \mathscr{T}_*)\,, \end{aligned} \tag{4.36}$$

or shortly,

$$\tilde{\mathbf{y}}_{t*} = \tilde{\boldsymbol{\mu}}_{t*} + \tilde{\boldsymbol{\varepsilon}}_{t*} \,, \tag{4.37}$$

where $\mathbf{N} = \mathbf{P}_{\mathbf{F}'\mathbf{Q}_{\mathbf{FX}}} = \mathbf{P}_{\mathscr{C}(\mathbf{F}')\cap\mathscr{C}(\mathbf{M})}$.

For the following Proposition 4.5, representing four equivalent statements, see, e.g., Baksalary and Kala (1981, 1986), Drygas (1983), Tian and Puntanen (2009, Th. 2.8), and Kala et al. (2017a, Th. 2). To the properties of the covariance matrices of the BLUEs we return in Propositions 4.6 and 7.4.

Proposition 4.5. *Consider the model* $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$ *and its transformed version* $\mathscr{T} = \{\mathbf{Fy}, \mathbf{FX}\boldsymbol{\beta}, \mathbf{FVF}'\}$, *and let* $\boldsymbol{\mu}_* = \mathbf{X}_*\boldsymbol{\beta}$ *be estimable under* $\mathscr{T}$ *(and thereby under* $\mathscr{M}$*). Then the following five statements are equivalent:*

(a) $\mathbf{Fy}$ *is* BLUE-*sufficient for* $\boldsymbol{\mu}_* = \mathbf{X}_*\boldsymbol{\beta}$.

(b) $\tilde{\boldsymbol{\mu}}_* = \tilde{\boldsymbol{\mu}}_{t*}$ *with probability* 1.

(c) $\mathrm{cov}(\tilde{\boldsymbol{\mu}}_*) = \mathrm{cov}(\tilde{\boldsymbol{\mu}}_{t*})$.

(d) $\{\mathbf{P}_{\mathbf{X}_*|\mathscr{T}}\mathbf{F}\} \cap \{\mathbf{P}_{\mathbf{X}_*|\mathscr{M}}\} \neq \{\emptyset\}$.

(e) $\{\mathbf{P}_{\mathbf{X}_*|\mathscr{T}}\mathbf{F}\} \subset \{\mathbf{P}_{\mathbf{X}_*|\mathscr{M}}\}$.

Example 4.1. [Centering the model.] Consider the partitioned linear model

$$\mathscr{M}_{12} = \{\mathbf{y},\, \mathbf{1}\alpha + \mathbf{X}_0\boldsymbol{\beta}_{\mathbf{x}}\,,\, \mathbf{I}_n\} = \{\mathbf{y},\, (\mathbf{1} : \mathbf{X}_0)\begin{pmatrix}\alpha \\ \boldsymbol{\beta}_{\mathbf{x}}\end{pmatrix},\, \mathbf{I}_n\}\,, \tag{4.38}$$

where $\mathbf{1} \in \mathbb{R}^n$ is a vector of ones. Assume that $\mathbf{X} = (\mathbf{1} : \mathbf{X}_0)$ has full column rank. Premultiplying the model $\mathscr{M}_{12}$ by the centering matrix $\mathbf{Q}_{\mathbf{1}} = \mathbf{I}_n - \mathbf{P}_{\mathbf{1}}$ yields the centered model

$$\mathscr{M}_{12\cdot 1} = \{\mathbf{Q}_{\mathbf{1}}\mathbf{y},\ \mathbf{Q}_{\mathbf{1}}\mathbf{X}_0\boldsymbol{\beta}_{\mathbf{x}}\,,\ \mathbf{Q}_{\mathbf{1}}\}\,. \tag{4.39}$$

In this centered model we have a singular covariance matrix and hence it may seem that finding a BLUE would be problematic. However, between the covariance matrix $\mathbf{Q}_{\mathbf{1}}$ and the model matrix $\mathbf{Q}_{\mathbf{1}}\mathbf{X}_0$ we have the relation

$$\mathscr{C}(\mathbf{Q}_{\mathbf{1}} \cdot \mathbf{Q}_{\mathbf{1}}\mathbf{X}_0) = \mathscr{C}(\mathbf{Q}_{\mathbf{1}}\mathbf{X}_0)\,. \tag{4.40}$$

Thus, corresponding to (3.11), we have the equality between OLSE($\mathbf{Q_1}\mathbf{X}_0\boldsymbol{\beta}_\mathbf{x}$) and BLUE($\mathbf{Q_1}\mathbf{X}_0\boldsymbol{\beta}_\mathbf{x}$), and thus

$$\mathrm{BLUE}(\boldsymbol{\beta}_\mathbf{x} \mid \mathscr{M}_{12\cdot 1}) = \mathrm{OLSE}(\boldsymbol{\beta}_\mathbf{x} \mid \mathscr{M}_{12\cdot 1}) = (\mathbf{X}_0'\mathbf{Q_1}\mathbf{X}_0)^{-1}\mathbf{X}_0'\mathbf{Q_1}\mathbf{y}. \tag{4.41}$$

On the other hand, it is well known that

$$\mathrm{BLUE}(\boldsymbol{\beta}_\mathbf{x} \mid \mathscr{M}_{12}) = (\mathbf{X}_0'\mathbf{Q_1}\mathbf{X}_0)^{-1}\mathbf{X}_0'\mathbf{Q_1}\mathbf{y}. \tag{4.42}$$

Now the equality of the BLUEs in (4.41) and (4.42) means that $\mathbf{Q_1}\mathbf{y}$ is linearly sufficient for $\boldsymbol{\beta}_\mathbf{x}$. This is a simple example of the Frisch–Waugh–Lovell theorem, into which we return in Section 7. □

Example 4.2. [Linear sufficiency of $\mathbf{V}^{-1/2}\mathbf{y}$ and $\mathbf{V}^{+1/2}\mathbf{y}$.] For a positive definite $\mathbf{V}$ the linear sufficiency condition (c) of Proposition 3.1 becomes simply $\mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{V}\mathbf{F}')$. Thus $\mathbf{V}^{-1/2}\mathbf{y}$ is linearly sufficient for $\mathbf{X}\boldsymbol{\beta}$ and the BLUE of $\mathbf{X}\boldsymbol{\beta}$ under the transformed model

$$\mathscr{T} = \{\mathbf{V}^{-1/2}\mathbf{y},\, \mathbf{V}^{-1/2}\mathbf{X}\boldsymbol{\beta},\, \mathbf{I}_n\} \tag{4.43}$$

is the same as in the original model $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$, i.e., the BLUE($\mathbf{X}\boldsymbol{\beta}$) under $\mathscr{M}$ equals the BLUE under $\mathscr{T}$:

$$\mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta} \mid \mathscr{M}) = \mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta} \mid \mathscr{T}) = \mathrm{OLSE}(\mathbf{X}\boldsymbol{\beta} \mid \mathscr{T}). \tag{4.44}$$

This technique, sometimes referred to as the Aitken-approach, see Aitken (1935), is well known in statistical textbooks. However, as remarked by Kala et al. (2017b, Sec. 3), usually these textbooks do not mention anything about linear sufficiency feature of this transformation.

Consider then so-called weakly singular model $\mathscr{M}_{ws}$, say, which means that $\mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{V})$. Transforming $\mathscr{M}$ by $\mathbf{V}^{+1/2}$ leads to the transformed model

$$\mathscr{T}_{ws} = \{\mathbf{V}^{+1/2}\mathbf{y},\, \mathbf{V}^{+1/2}\mathbf{X}\boldsymbol{\beta},\, \mathbf{P_V}\}, \tag{4.45}$$

because $\mathbf{V}^{+1/2}\mathbf{V}\mathbf{V}^{+1/2} = \mathbf{P_V}$. Under the model $\mathscr{M}_{ws}$, the statistic $\mathbf{V}^{+1/2}\mathbf{y}$ is linearly sufficient for $\mathbf{X}\boldsymbol{\beta}$; this is due to the fact that $\mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{V}\mathbf{V}^{+1/2}) = \mathscr{C}(\mathbf{V})$. The $\mathbf{W}$-matrix in $\mathscr{T}_{ws}$ is $\mathbf{P_V}$ whose Moore–Penrose inverse is $\mathbf{P_V}$ and hence under $\mathscr{T}_{ws}$ we have

$$\begin{aligned}\mathrm{BLUE}(\mathbf{V}^{+1/2}\mathbf{X}\boldsymbol{\beta}) &= \mathbf{V}^{+1/2}\,\mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta}) \\ &= \mathbf{V}^{+1/2}\mathbf{X}(\mathbf{X}'\mathbf{V}^{+1/2}\mathbf{P_V}\mathbf{V}^{+1/2}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{+1/2}\mathbf{P_V}\mathbf{V}^{+1/2}\mathbf{y} \\ &= \mathbf{V}^{+1/2}\mathbf{X}(\mathbf{X}'\mathbf{V}^{+}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{+}\mathbf{y}. \end{aligned} \tag{4.46}$$

Premultiplying (4.46) by $\mathbf{V}^{1/2}$ gives the following presentation for the BLUE of $\mathbf{X}\boldsymbol{\beta}$ under a weakly singular linear model:

$$\mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta} \mid \mathscr{M}_{ws}) = \mathbf{X}(\mathbf{X}'\mathbf{V}^{+}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{+}\mathbf{y}, \tag{4.47}$$

where $\mathbf{V}^+$ can be replaced with any $\mathbf{V}^-$. For the weakly singular models, see Zyskind and Martin (1969). □

Choosing $\mathbf{W} = \mathbf{V} + \mathbf{XUU'X'} \in \mathcal{W}$, we have, in light of (2.20), the following representations for the covariance matrix of the BLUE for $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$:

$$\begin{aligned}\operatorname{cov}(\tilde{\boldsymbol{\mu}} \mid \mathscr{M}) &= \mathbf{V} - \mathbf{VM}(\mathbf{MVM})^-\mathbf{MV} = \mathbf{X}(\mathbf{X}'\mathbf{W}^+\mathbf{X})^-\mathbf{X}' - \mathbf{T} \\ &= \mathbf{X}(\mathbf{X}'\mathbf{W}^{+1/2}\mathbf{W}^{+1/2}\mathbf{X})^-\mathbf{X}' - \mathbf{T}, \end{aligned} \tag{4.48}$$

where $\mathbf{T} = \mathbf{XUU'X'}$; for further representations, see, e.g., Isotalo et al. (2008b). Assuming that $\mathbf{X}\boldsymbol{\beta}$ is estimable under the transformed model $\mathscr{T}$, we have

$$\begin{aligned}\operatorname{cov}(\tilde{\boldsymbol{\mu}} \mid \mathscr{T}) &= \mathbf{X}[\mathbf{X}'\mathbf{F}'(\mathbf{FWF}')^-\mathbf{FX}]^-\mathbf{X}' - \mathbf{T} \\ &= \mathbf{X}(\mathbf{X}'\mathbf{W}^{+1/2}\mathbf{P}_{\mathbf{W}^{1/2}\mathbf{F}'}\mathbf{W}^{+1/2}\mathbf{X})^-\mathbf{X}' - \mathbf{T}. \end{aligned} \tag{4.49}$$

Using (4.48) and (4.49), Markiewicz and Puntanen (2019a, Th. 1) gave some characterizations of the linear sufficiency in terms of covariance matrices:

Proposition 4.6. *Let $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ be estimable under $\mathscr{T}$ and let $\mathbf{W} \in \mathcal{W}$. Then*

$$\operatorname{cov}(\tilde{\boldsymbol{\mu}} \mid \mathscr{M}) \leq_{\mathrm{L}} \operatorname{cov}(\tilde{\boldsymbol{\mu}} \mid \mathscr{T})\,. \tag{4.50}$$

Moreover, the following statements are equivalent:

(a) $\mathbf{Fy}$ *is linearly sufficient for* $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ *under* $\mathscr{M}$,

(b) $\operatorname{cov}(\tilde{\boldsymbol{\mu}} \mid \mathscr{M}) = \operatorname{cov}(\tilde{\boldsymbol{\mu}} \mid \mathscr{T})$,

(c) $\mathbf{X}(\mathbf{X}'\mathbf{W}^+\mathbf{X})^-\mathbf{X}' = \mathbf{X}(\mathbf{X}'\mathbf{W}^{+1/2}\mathbf{P}_{\mathbf{W}^{1/2}\mathbf{F}'}\mathbf{W}^{+1/2}\mathbf{X})^-\mathbf{X}$,

(d) $\mathbf{X}'\mathbf{W}^+\mathbf{X} = \mathbf{X}'\mathbf{W}^{+1/2}\mathbf{P}_{\mathbf{W}^{1/2}\mathbf{F}'}\mathbf{W}^{+1/2}\mathbf{X}$,

(e) $\mathscr{C}(\mathbf{W}^{+1/2}\mathbf{X}) \subset \mathscr{C}(\mathbf{W}^{1/2}\mathbf{F}')$.

The following proposition collects together some important properties of the linear prediction sufficiency. For further details, see Isotalo and Puntanen (2006b), and Isotalo et al. (2018). We will denote

$$\{\mathbf{P}_{\mathbf{y}_*|\mathscr{M}_*}\} = \{\mathbf{A} : \mathbf{Ay} = \mathrm{BLUP}(\mathbf{y}_* \mid \mathscr{M}_*)\}\,, \tag{4.51a}$$

$$\{\mathbf{P}_{\boldsymbol{\varepsilon}_*|\mathscr{M}_*}\} = \{\mathbf{B} : \mathbf{By} = \mathrm{BLUP}(\boldsymbol{\varepsilon}_* \mid \mathscr{M}_*)\}\,. \tag{4.51b}$$

The classes $\{\mathbf{P}_{\mathbf{y}_*|\mathscr{T}_*}\}$ and $\{\mathbf{P}_{\boldsymbol{\varepsilon}_*|\mathscr{T}_*}\}$ are defined in the corresponding way.

Proposition 4.7. *Suppose that $\mathbf{y}_*$ is predictable under $\mathscr{T}_*$. Then the following properties hold.*

(a) *The following seven statements are equivalent:*

(i) $\mathbf{Fy} \in \mathcal{S}(\mathbf{y}_*)$.	(ii) $\tilde{\mathbf{y}}_* = \tilde{\mathbf{y}}_{t*}$ *with probability* 1.
(iii) $\operatorname{cov}(\tilde{\mathbf{y}}_* - \tilde{\mathbf{y}}_{t*}) = \mathbf{0}$.	(iv) $\operatorname{cov}(\mathbf{y}_* - \tilde{\mathbf{y}}_*) = \operatorname{cov}(\mathbf{y}_* - \tilde{\mathbf{y}}_{t*})$.

(v) $\{\mathbf{P}_{\mathbf{y}_*|\mathscr{T}_*}\mathbf{F}\} \cap \{\mathbf{P}_{\mathbf{y}_*|\mathscr{M}_*}\} \neq \{\emptyset\}$. (vi) $\{\mathbf{P}_{\mathbf{y}_*|\mathscr{T}_*}\mathbf{F}\} \subset \{\mathbf{P}_{\mathbf{y}_*|\mathscr{M}_*}\}$.

(vii) $\operatorname{cov}(\tilde{\mathbf{y}}_*) = \operatorname{cov}(\tilde{\mathbf{y}}_{t*}) = \operatorname{cov}(\tilde{\mathbf{y}}_*, \tilde{\mathbf{y}}_{t*})$.

(b) *The following six statements are equivalent:*

(viii) $\mathbf{Fy} \in \mathcal{S}(\boldsymbol{\varepsilon}_*)$. (ix) $\tilde{\boldsymbol{\varepsilon}}_* = \tilde{\boldsymbol{\varepsilon}}_{t*}$ *with probability* 1.

(x) $\operatorname{cov}(\tilde{\boldsymbol{\varepsilon}}_*) = \operatorname{cov}(\tilde{\boldsymbol{\varepsilon}}_{t*})$. (xi) $\operatorname{cov}(\boldsymbol{\varepsilon}_* - \tilde{\boldsymbol{\varepsilon}}_*) = \operatorname{cov}(\boldsymbol{\varepsilon}_* - \tilde{\boldsymbol{\varepsilon}}_{t*})$.

(xii) $\{\mathbf{P}_{\boldsymbol{\varepsilon}_*|\mathscr{T}_*}\mathbf{F}\} \cap \{\mathbf{P}_{\boldsymbol{\varepsilon}_*|\mathscr{M}_*}\} \neq \{\emptyset\}$. (xiii) $\{\mathbf{P}_{\boldsymbol{\varepsilon}_*|\mathscr{T}_*}\mathbf{F}\} \subset \{\mathbf{P}_{\boldsymbol{\varepsilon}_*|\mathscr{M}_*}\}$.

Let us prove the equivalence of (iii) and (vii). Denoting $\tilde{\mathbf{y}}_* = \mathbf{Ay}$, $\tilde{\mathbf{y}}_{t*} = \mathbf{By}$, we observe that

$$\begin{aligned}
\operatorname{cov}(\tilde{\mathbf{y}}_* - \tilde{\mathbf{y}}_{t*}) &= (\mathbf{A} - \mathbf{B})\mathbf{V}(\mathbf{A} - \mathbf{B})' \\
&= \mathbf{AVA}' + \mathbf{BVB}' - \mathbf{AVB}' - \mathbf{BVA}' \\
&= \operatorname{cov}(\tilde{\mathbf{y}}_*) + \operatorname{cov}(\tilde{\mathbf{y}}_{t*}) - \operatorname{cov}(\tilde{\mathbf{y}}_*, \tilde{\mathbf{y}}_{*t}) - \operatorname{cov}(\tilde{\mathbf{y}}_{t*}, \tilde{\mathbf{y}}_*) \\
&= \mathbf{0}
\end{aligned} \tag{4.52}$$

holds if and only if $\mathbf{AV} = \mathbf{BV}$. Now $\mathbf{AV} = \mathbf{BV}$ immediately implies (vii). On the other hand, (vii) obviously implies (4.52).

Notice that in part (a) of Proposition 4.7 we do *not* have the covariance equality condition $\operatorname{cov}(\tilde{\mathbf{y}}_*) = \operatorname{cov}(\tilde{\mathbf{y}}_{t*})$; instead we have $\operatorname{cov}(\tilde{\mathbf{y}}_* - \tilde{\mathbf{y}}_{t*}) = \mathbf{0}$. In view of Lemma 1.6 the equality $\tilde{\mathbf{y}}_* = \tilde{\mathbf{y}}_{t*}$ holds with probability 1 if and only if $\operatorname{cov}(\tilde{\mathbf{y}}_* - \tilde{\mathbf{y}}_{t*}) = \mathbf{0}$. We return into this feature in Section 8.

5 Relative linear sufficiency

When studying the relative efficiency of OLSE vs BLUE of $\boldsymbol{\beta}$ we are dealing with two linear models $\mathscr{M}_{\mathbf{I}} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{I}_n\}$, and $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$, where $\mathbf{X}$ has full column rank and $\mathbf{V}$ is positive definite. The corresponding BLUEs are $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ and $\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$. Assuming that the model $\mathscr{M}$ is the correct one, the relative goodness of $\hat{\boldsymbol{\beta}}$ with respect to $\tilde{\boldsymbol{\beta}}$ can be measured by various means. The most common measure is the Watson efficiency, see Watson (1955) and Bloomfield and Watson (1975),

$$\phi = \frac{|\operatorname{cov}(\tilde{\boldsymbol{\beta}})|}{|\operatorname{cov}(\hat{\boldsymbol{\beta}})|} = \frac{|(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}|}{|(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}|} = \frac{|\mathbf{X}'\mathbf{X}|^2}{|\mathbf{X}'\mathbf{V}\mathbf{X}| \cdot |\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}|}\,, \tag{5.1}$$

where $|\cdot|$ refers to the determinant. Obviously $0 < \phi \leq 1$ and the upper bound is attained when $\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}$.

Kala et al. (2017b, Sec. 5) consider the models $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$ and $\mathscr{T} = \{\mathbf{Fy}, \mathbf{FX}\boldsymbol{\beta}, \mathbf{FVF}'\}$, and aim to do something similar with

$$\operatorname{BLUE}(\boldsymbol{\beta} \mid \mathscr{T}) = \tilde{\boldsymbol{\beta}}_t = [\mathbf{X}'\mathbf{F}'(\mathbf{FVF}')^{-}\mathbf{FX}]^{-1}\mathbf{X}'\mathbf{F}'(\mathbf{FVF}')^{-}\mathbf{Fy}, \tag{5.2a}$$

$$\operatorname{cov}(\tilde{\boldsymbol{\beta}}_t) = [\mathbf{X}'\mathbf{F}'(\mathbf{FVF}')^{-}\mathbf{FX}]^{-1}, \tag{5.2b}$$

where it is assumed that $\mathbf{FX}$ is estimable under $\mathscr{T}$. Corresponding to the Watson efficiency, we could consider the ratio

$$\gamma = \frac{|\mathrm{cov}(\tilde{\boldsymbol\beta})|}{|\mathrm{cov}(\tilde{\boldsymbol\beta}_t)|} = \frac{|(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}|}{|[\mathbf{X}'\mathbf{F}'(\mathbf{FVF}')^{-}\mathbf{FX}]^{-1}|} = \frac{|\mathbf{X}'\mathbf{V}^{-1/2}\mathbf{P}_{\mathbf{V}^{1/2}\mathbf{F}'}\mathbf{V}^{-1/2}\mathbf{X}|}{|\mathbf{X}'\mathbf{V}^{-1/2}\mathbf{V}^{-1/2}\mathbf{X}|}\,. \tag{5.3}$$

Clearly $0 < \gamma \leq 1$, where the upper bound is attained if and only if $\mathbf{Fy}$ is linearly sufficient for $\boldsymbol\beta$. We could keep $\mathbf{X}$ and $\mathbf{V}$ given and try to figure out which $\mathbf{F}$ yields the minimum of γ subject to the condition $\mathrm{r}(\mathbf{X}) = \mathrm{r}(\mathbf{FX})$. The lower bound for the Watson efficiency was found by Bloomfield and Watson (1975). However, it seems to be nontrivial to find the lower bound for γ. The (attainable) lower bound zero does not make sense, of course.

Kala et al. (2017b, Sec. 5) consider also another measure of the relative linear sufficiency based on the linear sufficiency condition $\mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{WF}')$ which is equivalent to

$$\mathbf{P}_{\mathbf{WF}'}\mathbf{X} = \mathbf{X}\,. \tag{5.4}$$

Hence one can wonder how "badly" (5.4) is satisfied by considering the difference $\mathbf{D} = \mathbf{X} - \mathbf{P}_{\mathbf{WF}'}\mathbf{X}$. The "size" of $\mathbf{D}$ could be measured by the Frobenius norm as

$$\|\mathbf{D}\|_F^2 = \mathrm{tr}(\mathbf{D}'\mathbf{D}) = \mathrm{tr}(\mathbf{X}'\mathbf{X}) - \mathrm{tr}(\mathbf{X}'\mathbf{P}_{\mathbf{WF}'}\mathbf{X})\,. \tag{5.5}$$

Hence the relative linear sufficiency of $\mathbf{Fy}$ could be defined as

$$\psi = \frac{\mathrm{tr}(\mathbf{X}'\mathbf{P}_{\mathbf{WF}'}\mathbf{X})}{\mathrm{tr}(\mathbf{X}'\mathbf{X})}\,, \tag{5.6}$$

where $\mathrm{tr}(\cdot)$ refers to the trace. Now $0 \leq \psi \leq 1$, where the lower bound is attained when $\mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{WF}')^{\perp}$ and the upper bound when $\mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{WF}')$, i.e., when $\mathbf{Fy}$ is linearly sufficient for $\mathbf{X}\boldsymbol\beta$.

Kala et al. (2017b, Remark 3) also consider two transformation matrices $\mathbf{F}_1$ and $\mathbf{F}_2$ and the corresponding transformed models

$$\mathscr{T}_i = \{\mathbf{F}_i\mathbf{y},\, \mathbf{F}_i\mathbf{X}\boldsymbol\beta,\, \mathbf{F}_i\mathbf{V}\mathbf{F}_i'\}\,,\ i \in \{1,2\}, \tag{5.7}$$

and assume that $\mathrm{r}(\mathbf{F}_1\mathbf{X}) = \mathrm{r}(\mathbf{F}_2\mathbf{X}) = \mathrm{r}(\mathbf{X}) = p$, so that $\boldsymbol\beta$ is estimable in both models. Then the Löwner ordering $\mathrm{cov}(\tilde{\boldsymbol\beta}_{t1}) \leq_{\mathrm{L}} \mathrm{cov}(\tilde{\boldsymbol\beta}_{t2})$ holds if and only if

$$\mathbf{X}'\mathbf{V}^{-1/2}\mathbf{P}_{\mathbf{V}^{1/2}\mathbf{F}_2'}\mathbf{V}^{-1/2}\mathbf{X} \leq_{\mathrm{L}} \mathbf{X}'\mathbf{V}^{-1/2}\mathbf{P}_{\mathbf{V}^{1/2}\mathbf{F}_1'}\mathbf{V}^{-1/2}\mathbf{X}\,, \tag{5.8}$$

i.e.,

$$\mathbf{X}'\mathbf{V}^{-1/2}(\mathbf{P}_{\mathbf{V}^{1/2}\mathbf{F}_1'} - \mathbf{P}_{\mathbf{V}^{1/2}\mathbf{F}_2'})\mathbf{V}^{-1/2}\mathbf{X} \geq_{\mathrm{L}} \mathbf{0}\,. \tag{5.9}$$

The matrix $\mathbf{P}_{\mathbf{V}^{1/2}\mathbf{F}_1'} - \mathbf{P}_{\mathbf{V}^{1/2}\mathbf{F}_2'}$ is nonnegative definite if and only if [see, e.g., Puntanen et al. (2011, p. 152)]

$$\mathscr{C}(\mathbf{F}_2') \subset \mathscr{C}(\mathbf{F}_1')\,. \tag{5.10}$$

Hence (5.8) holds if $\mathscr{C}(\mathbf{F}_2') \subset \mathscr{C}(\mathbf{F}_1')$. In this case we can say that in a sense $\mathbf{F}_1\mathbf{y}$ is "more than or equally linearly sufficient" than $\mathbf{F}_2\mathbf{y}$ even though neither of them need to be "fully linearly sufficient". However, it may well be that there is no Löwner ordering between the covariance matrices $\mathrm{cov}(\tilde{\boldsymbol\beta}_{t1})$ and $\mathrm{cov}(\tilde{\boldsymbol\beta}_{t2})$. Then some other criteria should be used to compare the "linear sufficiency" of $\mathbf{F}_1\mathbf{y}$ and $\mathbf{F}_2\mathbf{y}$. Using the matrix rank method, Dong et al. (2014) provide an extensive study of the relations between the covariance matrices of the BLUEs under $\mathscr{M}_{t1}$ and $\mathscr{M}_{t2}$.

We conclude this section with a curious result related to the Watson efficiency and linear sufficiency. Chu et al. (2004, 2005) showed that in the partitioned (weakly singular) linear model the Watson efficiency can be decomposed into the product

$$\mathrm{eff}(\hat{\boldsymbol\beta} \mid \mathscr{M}_{12}) = \mathrm{eff}(\hat{\boldsymbol\beta}_1 \mid \mathscr{M}_1) \cdot \mathrm{eff}(\hat{\boldsymbol\beta}_2 \mid \mathscr{M}_{12}) \cdot \frac{1}{\mathrm{eff}(\hat{\boldsymbol\beta}_1 \mid \mathscr{M}_{1H})}\,, \tag{5.11}$$

where $\mathscr{M}_1 = \{\mathbf{y}, \mathbf{X}_1\boldsymbol\beta_1, \mathbf{V}\}$, and $\mathscr{M}_{1H} = \{\mathbf{Hy}, \mathbf{X}_1\boldsymbol\beta_1, \mathbf{HVH}\}$, and $\mathbf{H} = \mathbf{P}_{\mathbf{X}}$. Thus $\mathscr{M}_{1H}$ is a transformed version of $\mathscr{M}_1$, transformation matrix being $\mathbf{F} = \mathbf{H}$. Chu et al. (2005, Th. 2.2) proved that a necessary and sufficient condition for

$$\mathrm{eff}(\hat{\boldsymbol\beta} \mid \mathscr{M}_{12}) = \mathrm{eff}(\hat{\boldsymbol\beta}_2 \mid \mathscr{M}_{12}) \tag{5.12}$$

is that $\mathbf{Hy}$ is linearly sufficient for $\mathbf{X}_1\boldsymbol\beta_1$ under $\mathscr{M}_1$, i.e., $\mathscr{C}(\mathbf{X}_1) \subset \mathscr{C}(\mathbf{VH}) = \mathscr{C}(\mathbf{VX})$. Given that (5.11) holds, the claim (5.12) is easy to prove using Proposition 4.6, which says that under linear sufficiency,

$$\mathrm{cov}(\tilde{\boldsymbol\beta}_1 \mid \mathscr{M}_1) = \mathrm{cov}(\tilde{\boldsymbol\beta}_1 \mid \mathscr{M}_{1H})\,. \tag{5.13}$$

Of course, $\mathrm{cov}(\hat{\boldsymbol\beta}_1 \mid \mathscr{M}_1) = \mathrm{cov}(\hat{\boldsymbol\beta}_1 \mid \mathscr{M}_{1H}) = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{V}\mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}$. We find the reduction of (5.11) into (5.12) somewhat unexpected and it is not obvious to find a "natural explanation" for it.

6 The "vice versa" problem

Let us take a close look at the Theorem of Baksalary and Kala (1981, p. 914). In part (ii) they write the following (in our notation):

(a) "If the condition $\mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{WF}')$ is satisfied, then each BLUE of $\mathbf{X}\boldsymbol\beta$ in the transformed model $\mathscr{T} = \{\mathbf{Fy}, \mathbf{FX}\boldsymbol\beta, \mathbf{FVF}'\}$ is also a BLUE of $\mathbf{X}\boldsymbol\beta$ in the original model $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol\beta, \mathbf{V}\}$, and vice versa."

On the other hand, Baksalary and Kala (1986, Th. 1) state the following:

(b) "If $\mathbf{Fy}$ is linearly sufficient for estimable $\mathbf{X}_*\boldsymbol\beta$, then every representation of the BLUE of $\mathbf{X}_*\boldsymbol\beta$ in the induced model $\mathscr{T}$ is also the BLUE of $\mathbf{X}_*\boldsymbol\beta$ in the original model $\mathscr{M}$."

It is the phrase *vice versa* that may cause some confusion as stated by Kala et al. (2017a, Sec. 4). Let us discuss the meaning of the *vice versa* part.

Consider the multipliers of the response vector $\mathbf{y}$ when playing with the BLUEs under $\mathscr{M}$ and under $\mathscr{T}$. Let $\mathbf{X}_*\boldsymbol{\beta}$ be an estimable parametric function under the model $\mathscr{T}$ (and thereby also under $\mathscr{M}$) and denote

$$\begin{aligned} \mathcal{A} &= \{\mathbf{A} : \mathbf{AFy} = \mathrm{BLUE}(\mathbf{X}_*\boldsymbol{\beta} \mid \mathscr{M})\} \\ &= \{\mathbf{A} : \mathbf{AF}(\mathbf{X} : \mathbf{VM}) = (\mathbf{X}_* : \mathbf{0})\}\,, \end{aligned} \tag{6.1a}$$

$$\begin{aligned} \mathcal{C} &= \{\mathbf{C} : \mathbf{CFy} = \mathrm{BLUE}(\mathbf{X}_*\boldsymbol{\beta} \mid \mathscr{T})\} \\ &= \{\mathbf{C} : \mathbf{C}(\mathbf{FX} : \mathbf{FVF}'\mathbf{Q}_{\mathbf{FX}}) = (\mathbf{X}_* : \mathbf{0})\} \\ &= \{\mathbf{C} : \mathbf{CF}(\mathbf{X} : \mathbf{VMF}'\mathbf{Q}_{\mathbf{FX}}) = (\mathbf{X}_* : \mathbf{0})\} \\ &= \{\mathbf{P}_{\mathbf{X}_*|\mathscr{T}}\}\,, \end{aligned} \tag{6.1b}$$

where we have used Lemma 1.4. Assume further that $\mathbf{Fy}$ is linearly sufficient for $\mathbf{X}_*\boldsymbol{\beta}$, which actually guarantees that $\mathcal{A}$ and $\mathcal{C}$ are not empty.

By Proposition 4.2. we observe immediately that $\mathcal{A} \subset \mathcal{C}$. To go the other way, let us pick $\mathbf{C} \in \mathcal{C} = \{\mathbf{P}_{\mathbf{X}_*|\mathscr{T}}\}$ and utilize part (c) of Proposition 4.3 which says that (under linear sufficiency of $\mathbf{Fy}$) $\mathbf{CF} \in \{\mathbf{P}_{\mathbf{X}_*|\mathscr{M}}\}$, i.e., $\mathbf{CF}$ satisfies $\mathbf{CF}(\mathbf{X} : \mathbf{VM}) = (\mathbf{X}_* : \mathbf{0})$. Thus we have confirmed the following; see Kala et al. (2017a, Th. 3).

Proposition 6.1. *Suppose that* $\mathbf{Fy}$ *is linearly sufficient for the estimable parametric function* $\mathbf{X}_*\boldsymbol{\beta}$ *under the linear model* $\mathscr{M}$*, and let the sets of matrices* $\mathcal{A}$ *and* $\mathcal{C}$ *be defined as above. Then* $\mathcal{A} = \mathcal{C} = \{\mathbf{P}_{\mathbf{X}_*|\mathscr{T}}\}$ *and*

$$\mathcal{B} = \big\{\mathbf{B} : \mathbf{B} = \mathbf{AF} \in \{\mathbf{P}_{\mathbf{X}_*|\mathscr{M}}\}\big\} = \{\mathbf{P}_{\mathbf{X}_*|\mathscr{T}}\mathbf{F}\}\,. \tag{6.2}$$

Assume that $\mathbf{Fy}$ is linearly sufficient for estimable $\boldsymbol{\mu}_*$ under $\mathscr{M}$. Then the Proposition 6.1 means that $\mathbf{CFy}$ is the BLUE for $\boldsymbol{\mu}_*$ under $\mathscr{T}$ if and only if $\mathbf{CFy}$ is the BLUE for $\boldsymbol{\mu}_*$ under $\mathscr{M}$. In other words, for each matrix $\mathbf{C}$ such that $\mathbf{CFy}$ is the BLUE of $\boldsymbol{\mu}_*$ in the transformed model $\mathscr{T}$, the statistic $\mathbf{CFy}$ is also the BLUE of $\boldsymbol{\mu}_*$ in the original model $\mathscr{M}$, and vice versa. Notice that in this statement the "vice versa" means that we consider such $\mathbf{C}$ for which $\mathbf{CFy}$ is BLUE under $\mathscr{M}$, not the set of matrices $\mathbf{B} \in \{\mathbf{P}_{\mathbf{X}_*|\mathscr{M}}\}$ such that $\mathbf{By}$ is BLUE under $\mathscr{M}$.

It is noteworthy that we have the inclusion

$$\mathcal{B} = \big\{\mathbf{B} : \mathbf{B} = \mathbf{AF} \in \{\mathbf{P}_{\mathbf{X}_*|\mathscr{M}}\}\big\} \subset \{\mathbf{P}_{\mathbf{X}_*|\mathscr{M}}\}\,, \tag{6.3}$$

which immediately implies part (c) of Proposition 4.3:

$$\{\mathbf{P}_{\mathbf{X}_*|\mathscr{T}}\mathbf{F}\} \subset \{\mathbf{P}_{\mathbf{X}_*|\mathscr{M}}\}\,. \tag{6.4}$$

It is clear that it is not necessary that every matrix from the set $\{\mathbf{P}_{\mathbf{X}_*|\mathscr{M}}\}$ belongs to set $\mathcal{B}$.

The total sets of multipliers of $\mathbf{y}$ for the BLUEs of $\mathbf{X}_*\boldsymbol{\beta}$ under $\mathscr{M}$ and $\mathscr{T}$ are $\{\mathbf{P}_{\mathbf{X}_*|\mathscr{T}}\mathbf{F}\}$ and $\{\mathbf{P}_{\mathbf{X}_*|\mathscr{M}}\}$, respectively. Kala et al. (2017a, Sec. 4) show that the equality in (6.4) holds if $\mathbf{F}$ is a nonsingular square matrix. Using a different approach, Tian (2013, Th. 3.1) has shown that the nonsingularity of $\mathbf{F}$ is also necessary for the equality in (6.4). However, Tian's assumptions are slightly different from those of ours and thus in our considerations the equality in (6.4) can appear without $\mathbf{F}$ being nonsingular. A simple example is obtained by considering model $\mathscr{M}$, where $\mathbf{X}$ has full column rank and $\mathbf{V}$ is positive definite. Then choosing $\mathbf{F} = \mathbf{X}'\mathbf{V}^{-1}$, we have $\mathbf{Fy} \in \mathcal{S}(\boldsymbol{\beta})$ and

$$\mathscr{T} = \{\mathbf{X}'\mathbf{V}^{-1}\mathbf{y},\, \mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\boldsymbol{\beta},\, \mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\}\,, \tag{6.5}$$

yielding

$$\mathrm{BLUE}(\boldsymbol{\beta} \mid \mathscr{M}) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} = \mathrm{BLUE}(\boldsymbol{\beta} \mid \mathscr{T})\,. \tag{6.6}$$

Thus there is only one member, $(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}$, in sets $\{\mathbf{P}_{\mathbf{X}_*|\mathscr{T}}\mathbf{F}\}$ and $\{\mathbf{P}_{\mathbf{X}_*|\mathscr{M}}\}$; yet $\mathbf{F} = \mathbf{X}'\mathbf{V}^{-1} \in \mathbb{R}^{p\times n}$ is not a square matrix and of course not invertible.

A deeper look at the possible equality in (6.4) yields the Proposition 6.2 (which according to our knowledge is new). For this purpose we denote

$$\begin{aligned}\{\mathbf{P}_{\boldsymbol{\varepsilon}_*|\mathscr{T}_*}\} &= \{\mathbf{D} : \mathbf{DFy} = \mathrm{BLUP}(\boldsymbol{\varepsilon}_* \mid \mathscr{T}_*)\} \\ &= \{\mathbf{D} : \mathbf{D}(\mathbf{FX} : \mathbf{FVF}'\mathbf{Q}_{\mathbf{FX}}) = (\mathbf{0} : \mathbf{V}_{21}\mathbf{M})\}\,.\end{aligned} \tag{6.7}$$

Proposition 6.2. *Let $\mathbf{X}_*\boldsymbol{\beta}$ be estimable under $\mathscr{T}$ and consider condition*

$$\mathbf{Q}_{\mathbf{W}} = \mathbf{Q}_{\mathbf{W}}\mathbf{P}_{\mathbf{F}'}\,, \quad \textit{i.e.,} \quad \mathscr{C}(\mathbf{W})^{\perp} \subset \mathscr{C}(\mathbf{F}')\,. \tag{6.8}$$

(a) *If $\mathbf{Fy} \in \mathcal{S}(\mathbf{X}_*\boldsymbol{\beta})$, then $\{\mathbf{P}_{\mathbf{X}_*|\mathscr{M}}\} = \{\mathbf{P}_{\mathbf{X}_*|\mathscr{T}}\mathbf{F}\} \iff$ (6.8) holds.*

(b) *If $\mathbf{Fy} \in \mathcal{S}(\boldsymbol{\varepsilon}_*)$, then $\{\mathbf{P}_{\boldsymbol{\varepsilon}_*|\mathscr{M}_*}\} = \{\mathbf{P}_{\boldsymbol{\varepsilon}_*|\mathscr{T}_*}\mathbf{F}\} \iff$ (6.8) holds.*

(c) *If $\mathbf{Fy} \in \mathcal{S}(\mathbf{y}_*)$, then $\{\mathbf{P}_{\mathbf{y}_*|\mathscr{M}_*}\} = \{\mathbf{P}_{\mathbf{y}_*|\mathscr{T}_*}\mathbf{F}\} \iff$ (6.8) holds.*

In other words, under the linear sufficiency and condition (6.8)*, each representation of the* BLUE *of $\mathbf{X}_*\boldsymbol{\beta}$ under $\mathscr{M}$ is a representation of the* BLUE *under $\mathscr{T}$ and vice versa, and the same concerns the* BLUP *for $\boldsymbol{\varepsilon}_*$ and $\mathbf{y}_*$.*

Proof. We know, by (6.4) and part (c) of Proposition 4.3, that linear sufficiency of $\mathbf{Fy}$ with respect to $\mathbf{X}_*\boldsymbol{\beta}$ implies the inclusion $\{\mathbf{P}_{\mathbf{X}_*|\mathscr{T}}\mathbf{F}\} \subset \{\mathbf{P}_{\mathbf{X}_*|\mathscr{M}}\}$. Thus the claim (a) means that if $\mathbf{Fy} \in \mathcal{S}(\mathbf{X}_*\boldsymbol{\beta})$, then

$$(6.8) \iff \{\mathbf{P}_{\mathbf{X}_*|\mathscr{M}}\} \subset \{\mathbf{P}_{\mathbf{X}_*|\mathscr{T}}\mathbf{F}\}\,. \tag{6.9}$$

The general expression for $\mathbf{B}_0 \in \{\mathbf{P}_{\mathbf{X}_*|\mathscr{M}}\}$ is

$$\mathbf{B}_0 = \mathbf{B}_1 + \mathbf{E}\mathbf{Q}_{\mathbf{W}}\,, \tag{6.10}$$

where $\mathbf{B}_1$ is one matrix from the set $\{\mathbf{P}_{\mathbf{X}_*|\mathscr{M}}\}$ and $\mathbf{E} \in \mathbb{R}^{q\times n}$ is arbitrary. Because $\mathbf{Fy}$ is linearly sufficient for $\mathbf{X}_*\boldsymbol{\beta}$, we know that there exists a matrix $\mathbf{A} \in \mathbb{R}^{q\times f}$ such that $\mathbf{AF} \in \{\mathbf{P}_{\mathbf{X}_*|\mathscr{M}}\}$. Choosing $\mathbf{B}_1 = \mathbf{AF}$ gives

$$\mathbf{B}_0 = \mathbf{AF} + \mathbf{E}\mathbf{Q}_{\mathbf{W}}\,, \tag{6.11}$$

Assume now that

$$\{\mathbf{P}_{\mathbf{X}_*|\mathscr{M}}\} \subset \{\mathbf{P}_{\mathbf{X}_*|\mathscr{T}}\mathbf{F}\}\,. \tag{6.12}$$

We observe that (6.12) implies that $\mathbf{B}_0$ is of the form $\mathbf{B}_0 = \mathbf{LF}$ for some $\mathbf{L} \in \mathbb{R}^{q\times f}$, which further means that

$$\mathbf{B}_0\mathbf{P}_{\mathbf{F}'} = \mathbf{B}_0\,. \tag{6.13}$$

Substituting (6.13) into (6.11) gives

$$\mathbf{AF} + \mathbf{EQ}_\mathbf{W}\mathbf{P}_{\mathbf{F}'} = \mathbf{AF} + \mathbf{EQ}_\mathbf{W}\,, \tag{6.14}$$

i.e.,

$$\mathbf{E}(\mathbf{Q}_\mathbf{W} - \mathbf{Q}_\mathbf{W}\mathbf{P}_{\mathbf{F}'}) = \mathbf{0}\,. \tag{6.15}$$

Equation (6.15) has to hold for any $\mathbf{E} \in \mathbb{R}^{q\times n}$ and hence we necessarily must have

$$\mathbf{P}_{\mathbf{F}'}\mathbf{Q}_\mathbf{W} = \mathbf{Q}_\mathbf{W}\,, \quad \text{i.e.,} \quad \mathbf{P}_{\mathbf{F}'}(\mathbf{I}_n - \mathbf{P}_\mathbf{W}) = (\mathbf{I}_n - \mathbf{P}_\mathbf{W})\,, \tag{6.16}$$

which is precisely the condition (6.8). Thus we have shown that if $\mathbf{Fy} \in \mathcal{S}(\mathbf{X}_*\boldsymbol{\beta})$, then (6.12) implies (6.8), i.e.,

$$\{\mathbf{P}_{\mathbf{X}_*|\mathscr{M}}\} = \{\mathbf{P}_{\mathbf{X}_*|\mathscr{T}}\mathbf{F}\} \implies (6.8). \tag{6.17}$$

It remains to show that if $\mathbf{Fy} \in \mathcal{S}(\mathbf{X}_*\boldsymbol{\beta})$, then

$$(6.8) \implies \{\mathbf{P}_{\mathbf{X}_*|\mathscr{M}}\} \subset \{\mathbf{P}_{\mathbf{X}_*|\mathscr{T}}\mathbf{F}\}\,. \tag{6.18}$$

In view of (6.11) we know that there exists a matrix $\mathbf{A}$ such that the general expression for $\mathbf{B}_0 \in \{\mathbf{P}_{\mathbf{X}_*|\mathscr{M}}\}$ is

$$\mathbf{B}_0 = \mathbf{AF} + \mathbf{EQ}_\mathbf{W}\,, \tag{6.19}$$

where $\mathbf{E}$ is free to vary. Postmultiplying (6.19) by $\mathbf{P}_{\mathbf{F}'} = \mathbf{F}^+\mathbf{F}$ yields

$$\mathbf{B}_0\mathbf{P}_{\mathbf{F}'} = \mathbf{AF} + \mathbf{EQ}_\mathbf{W}\mathbf{P}_{\mathbf{F}'}\,. \tag{6.20}$$

Assuming that (6.8) holds, we have $\mathbf{Q}_\mathbf{W}\mathbf{P}_{\mathbf{F}'} = \mathbf{Q}_\mathbf{W}$ and thereby in light of (6.19) and (6.20) we have

$$\mathbf{B}_0 = \mathbf{B}_0\mathbf{P}_{\mathbf{F}'} = \mathbf{B}_0\mathbf{F}^+ \cdot \mathbf{F} := \mathbf{B}_2\mathbf{F}, \tag{6.21}$$

where $\mathbf{B}_2 = \mathbf{B}_0\mathbf{F}^+$. It remains to show that $\mathbf{B}_2 = \mathbf{B}_0\mathbf{F}^+ \in \{\mathbf{P}_{\mathbf{X}_*|\mathscr{T}}\}$. This follows at once from Proposition 4.2 which says that

$$\mathbf{B}_0\mathbf{y} = \mathbf{B}_2\mathbf{Fy} = \mathrm{BLUE}(\mathbf{X}_*\boldsymbol{\beta} \mid \mathscr{M}) \implies \mathbf{B}_2 \in \{\mathbf{P}_{\mathbf{X}_*|\mathscr{T}}\}\,. \tag{6.22}$$

Thus the proof of claim (a) is completed.

The proofs of parts (b) and (c) of Proposition 6.2 are parallel to that of (a). □

As an aside we may mention that if $\mathscr{C}(\mathbf{W}) = \mathbb{R}^n$, then $\{\mathbf{P}_{\mathbf{X}_*|\mathscr{M}}\} = \{\mathbf{P}_{\mathbf{X}_*|\mathscr{T}}\mathbf{F}\}$ holds for any $\mathbf{Fy} \in \mathcal{S}(\mathbf{X}_*\boldsymbol{\beta})$. Notice that in that situation the BLUE's representation is unique.

7 Partitioned linear model

Consider a partitioned linear model $\mathscr{M}_{12} = \{\mathbf{y}, \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2, \mathbf{V}\}$, where $\mathbf{X}_i \in \mathbb{R}^{n\times p_i},\ i = 1, 2,\ p_1 + p_2 = p$. We denote the transformed model as

$$\mathscr{T}_{12} = \{\mathbf{F}\mathbf{y},\, \mathbf{F}\mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{F}\mathbf{X}_2\boldsymbol{\beta}_2,\, \mathbf{F}\mathbf{V}\mathbf{F}'\}\,. \tag{7.1}$$

Let the transformation matrix be $\mathbf{M}_2 = \mathbf{I}_n - \mathbf{P}_{\mathbf{X}_2}$, so that the transformed model is

$$\mathscr{M}_{12\cdot 2} = \{\mathbf{M}_2\mathbf{y},\, \mathbf{M}_2\mathbf{X}_1\boldsymbol{\beta}_1,\, \mathbf{M}_2\mathbf{V}\mathbf{M}_2\}\,. \tag{7.2}$$

Then, in view of part (d) of Proposition 3.3, the statistic $\mathbf{M}_2\mathbf{y}$ is linearly sufficient for its expectation $\mathbf{M}_2\mathbf{X}_1\boldsymbol{\beta}_1$ under $\mathscr{M}_{12}$ if and only if

$$\mathscr{C}(\mathbf{M}_2\mathbf{X}_1) \cap \mathscr{C}(\mathbf{M}_2\mathbf{V}\mathbf{M}) = \{\mathbf{0}\}\,. \tag{7.3}$$

Using the decomposition $\mathbf{M} = \mathbf{M}_2\mathbf{Q}_{\mathbf{M}_2\mathbf{X}_1}$, (7.3) becomes

$$\mathscr{C}(\mathbf{M}_2\mathbf{X}_1) \cap \mathscr{C}(\mathbf{M}_2\mathbf{V}\mathbf{M}_2\mathbf{Q}_{\mathbf{M}_2\mathbf{X}_1}) = \{\mathbf{0}\}\,, \tag{7.4}$$

which obviously, see (1.26), holds and thus $\mathbf{M}_2\mathbf{y}$ is linearly sufficient for $\mathbf{M}_2\mathbf{X}_1\boldsymbol{\beta}_1$; see also Kala and Pordzik (2009, Th. 1).

Is $\mathbf{M}_2\mathbf{y}$ linearly minimal sufficient? By part (b) of Proposition 3.3, the answer is positive if

$$\mathrm{r}[\mathbf{X}(\mathbf{X}'\mathbf{W}^-\mathbf{X})^-\mathbf{X}_*'] = \mathrm{r}(\mathbf{W}\mathbf{M}_2)\,, \tag{7.5}$$

where $\mathbf{X}_* = \mathbf{M}_2\mathbf{X} = (\mathbf{M}_2\mathbf{X}_1 : \mathbf{0})$. Now

$$\mathrm{r}[\mathbf{X}(\mathbf{X}'\mathbf{W}^-\mathbf{X})^-\mathbf{X}_*'] = \mathrm{r}(\mathbf{X}_*) = \mathrm{r}(\mathbf{M}_2\mathbf{X})\,. \tag{7.6}$$

Using the rank rule of Marsaglia and Styan (1974, Cor. 6.2) it can be shown that $\mathrm{r}(\mathbf{M}_2\mathbf{W}) = \mathrm{r}(\mathbf{M}_2\mathbf{X})$ if and only if $\mathrm{r}(\mathbf{W}) = \mathrm{r}(\mathbf{X})$, i.e.,

$$\mathbf{M}_2\mathbf{y} \in \mathcal{S}_0(\mathbf{M}_2\mathbf{X}_1\boldsymbol{\beta}_1) \iff \mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{V})\,. \tag{7.7}$$

The following proposition characterizes the estimability of $\boldsymbol{\mu}_1 = \mathbf{X}_1\boldsymbol{\beta}_1$ under $\mathscr{M}_{12}$, $\mathscr{T}_{12}$ and under $\mathscr{M}_{12\cdot 2} = \{\mathbf{M}_2\mathbf{y}, \mathbf{M}_2\mathbf{X}_1\boldsymbol{\beta}_1, \mathbf{M}_2\mathbf{V}\mathbf{M}_2\}$.

Proposition 7.1. *Consider the models $\mathscr{M}_{12}$ and its transformed versions $\mathscr{T}_{12}$ and $\mathscr{M}_{12\cdot 2}$. Then the followings statements hold:*

(a) $\mathbf{X}_1\boldsymbol{\beta}_1$ *is estimable under $\mathscr{M}_{12}$ if and only if $\mathscr{C}(\mathbf{X}_1') = \mathscr{C}(\mathbf{X}_1'\mathbf{M}_2)$.*

(b) $\mathbf{X}_1\boldsymbol{\beta}_1$ *is estimable under $\mathscr{T}_{12}$ if and only if $\mathscr{C}(\mathbf{X}_1') = \mathscr{C}(\mathbf{X}_1'\mathbf{F}'\mathbf{Q}_{\mathbf{F}\mathbf{X}_2})$, or, equivalently, if and only if $\mathscr{C}(\mathbf{X}_1') = \mathscr{C}(\mathbf{X}_1'\mathbf{F}')$ and $\mathscr{C}(\mathbf{F}\mathbf{X}_1) \cap \mathscr{C}(\mathbf{F}\mathbf{X}_2) = \{\mathbf{0}\}$.*

(c) $\mathbf{X}_1\boldsymbol{\beta}_1$ *is estimable under $\mathscr{M}_{12\cdot 2}$ if and only if $\mathscr{C}(\mathbf{X}_1') = \mathscr{C}(\mathbf{X}_1'\mathbf{M}_2)$.*

It is noteworthy that if $\mathscr{C}(\mathbf{X}_1') = \mathscr{C}(\mathbf{X}_1'\mathbf{M}_2)$, i.e., $\mathscr{C}(\mathbf{X}_1) \cap \mathscr{C}(\mathbf{X}_2) = \{\mathbf{0}\}$, which is a condition for the estimability of $\mathbf{X}_1\boldsymbol{\beta}_1$, then by part (f) of Proposition 4.3,

$$\mathbf{M}_2\mathbf{y} \in \mathcal{S}(\mathbf{M}_2\mathbf{X}\boldsymbol{\beta}) = \mathcal{S}(\mathbf{M}_2\mathbf{X}_1\boldsymbol{\beta}_1) \iff \mathbf{M}_2\mathbf{y} \in \mathcal{S}[(\mathbf{X}_1 : \mathbf{0})\boldsymbol{\beta}] = \mathcal{S}(\mathbf{X}_1\boldsymbol{\beta}_1)\,, \tag{7.8}$$

i.e., assuming the estimability of $\mathbf{X}_1\boldsymbol{\beta}_1$, $\mathbf{M}_2\mathbf{y}$ is linearly sufficient for $\mathbf{M}_2\mathbf{X}_1\boldsymbol{\beta}_1$ if and only if $\mathbf{M}_2\mathbf{y}$ is linearly sufficient for $\mathbf{X}_1\boldsymbol{\beta}_1$.

The next result is an interesting corollary to Proposition 6.2.

Proposition 7.2. *Consider the partitioned linear model $\mathscr{M}_{12}$ and the reduced model $\mathscr{M}_{12\cdot 2} = \{\mathbf{M}_2\mathbf{y}, \mathbf{M}_2\mathbf{X}_1\boldsymbol{\beta}_1, \mathbf{M}_2\mathbf{V}\mathbf{M}_2\}$. Then*

$$\{\mathbf{P}_{\mathbf{M}_2\mathbf{X}|\mathscr{M}_{12}}\} = \{\mathbf{P}_{\mathbf{M}_2\mathbf{X}|\mathscr{M}_{12\cdot 2}}\mathbf{M}_2\}\,, \tag{7.9}$$

that is, each representation of the BLUE *of $\mathbf{M}_2\mathbf{X}\boldsymbol{\beta} = \mathbf{M}_2\mathbf{X}_1\boldsymbol{\beta}_1$ under $\mathscr{M}_{12}$ is a representation of the* BLUE *under $\mathscr{M}_{12\cdot 2}$ and vice versa. Moreover, if $\mathbf{X}_1\boldsymbol{\beta}_1$ is estimable under $\mathscr{M}$, i.e., $\mathscr{C}(\mathbf{X}_1) \cap \mathscr{C}(\mathbf{X}_2) = \{\mathbf{0}\}$, then*

$$\{\mathbf{P}_{(\mathbf{X}_1:\mathbf{0})|\mathscr{M}_{12}}\} = \{\mathbf{P}_{(\mathbf{X}_1:\mathbf{0})|\mathscr{M}_{12\cdot 2}}\mathbf{M}_2\}\,. \tag{7.10}$$

Proof. By Proposition 6.2, (7.9) holds if

$$\mathscr{C}(\mathbf{F}')^{\perp} \subset \mathscr{C}(\mathbf{W})\,, \tag{7.11}$$

where now $\mathbf{F}' = \mathbf{M}_2$. Trivially (7.11) in this case holds. The second part of Proposition 7.2 is obvious. □

An alternative proof of Proposition 7.2 appears in Puntanen et al. (2011, Sec. 15.4). Proposition 7.2 is the general formulation of the well-known Frisch–Waugh–Lovell theorem, see, for example, Frisch and Waugh (1933), Lovell (1963, 2008), Groß and Puntanen (2000, 2005), Bhimasankaram and Sengupta (1996, Th. 6.1), and Arendacká and Puntanen (2015, Th. 1). Actually Groß and Puntanen (2000, Th. 4) show that $\{\mathbf{P}_{\mathbf{M}_2\mathbf{X}|\mathscr{M}_{12\cdot 2}}\mathbf{M}_2\} \subset \{\mathbf{P}_{\mathbf{M}_2\mathbf{X}|\mathscr{M}_{12}}\}$ but state “boldly” that “It is obvious that also the reverse relation holds”.

Consider then the estimation of $\boldsymbol{\mu}_1 = \mathbf{X}_1\boldsymbol{\beta}_1 = (\mathbf{X}_1 : \mathbf{0})\boldsymbol{\beta}$ under $\mathscr{M}_{12}$, where $\mathscr{C}(\mathbf{X}_1) \cap \mathscr{C}(\mathbf{X}_2) = \{\mathbf{0}\}$, so that $\boldsymbol{\mu}_1$ is estimable. Now $\mathbf{X}_* = (\mathbf{X}_1 : \mathbf{0})$ and on account of (2.14a), one expression for the BLUE of $\boldsymbol{\mu}_1$ is

$$\mathbf{A}\mathbf{y} = \mathbf{X}_*(\mathbf{X}'\mathbf{W}^-\mathbf{X})^-\mathbf{X}'\mathbf{W}^-\mathbf{y}. \tag{7.12}$$

An alternative expression for the BLUE of $\boldsymbol{\mu}_1$, obtainable from $\mathscr{M}_{12\cdot 2}$, is

$$\begin{aligned}\mathbf{B}\mathbf{y} &= \mathbf{X}_1[\mathbf{X}_1'\mathbf{M}_2(\mathbf{M}_2\mathbf{W}\mathbf{M}_2)^-\mathbf{M}_2\mathbf{X}_1]^-\mathbf{X}_1'\mathbf{M}_2(\mathbf{M}_2\mathbf{W}\mathbf{M}_2)^-\mathbf{M}_2\mathbf{y} \\ &= \mathbf{X}_1(\mathbf{X}_1'\dot{\mathbf{M}}_{2W}\mathbf{X}_1)^-\mathbf{X}_1'\dot{\mathbf{M}}_{2W}\mathbf{y},\end{aligned} \tag{7.13}$$

where, on account of (1.18), $\mathbf{M}_2\mathbf{W}\mathbf{M}_2 \in \mathcal{W}_{\mathscr{M}_{12\cdot 2}}$ and we have denoted

$$\dot{\mathbf{M}}_{2W} = \mathbf{M}_2(\mathbf{M}_2\mathbf{W}\mathbf{M}_2)^-\mathbf{M}_2\,. \tag{7.14}$$

It is clear that in (7.14) $\mathbf{W}$ can be replaced with any matrix of the form $\mathbf{W}_1 = \mathbf{V} + \mathbf{X}_1\mathbf{U}_1\mathbf{U}_1'\mathbf{X}_1'$ such that $\mathscr{C}(\mathbf{W}_1) = \mathscr{C}(\mathbf{V} : \mathbf{X}_1)$.

Now we have $\mathbf{Ay} = \mathbf{By}$ for all $\mathbf{y} \in \mathscr{C}(\mathbf{W})$, i.e., $\mathbf{AW} = \mathbf{BW}$ which implies

$$\mathbf{W}\dot{\mathbf{M}}_{2W}\mathbf{X}_1(\mathbf{X}_1'\dot{\mathbf{M}}_{2W}\mathbf{X}_1)^-\mathbf{X}_1' = \mathbf{X}(\mathbf{X}'\mathbf{W}^-\mathbf{X})^-\mathbf{X}_*' \,. \tag{7.15}$$

It is easy to confirm that $\mathscr{C}[\mathbf{W}\dot{\mathbf{M}}_{2W}\mathbf{X}_1(\mathbf{X}_1'\dot{\mathbf{M}}_{2W}\mathbf{X}_1)^-\mathbf{X}_1'] = \mathscr{C}(\mathbf{W}\dot{\mathbf{M}}_{2W}\mathbf{X}_1)$. Thus, in view of (a$_3$) of Proposition 3.3,

$$\mathbf{Fy} \in \mathcal{S}(\mathbf{X}_1\boldsymbol{\beta}_1) \iff \mathscr{C}(\mathbf{W}\dot{\mathbf{M}}_{2W}\mathbf{X}_1) \subset \mathscr{C}(\mathbf{WF}') \,. \tag{7.16}$$

From (7.16) we immediately see that $\mathbf{X}_1'\dot{\mathbf{M}}_{2W}\mathbf{y}$ is linearly minimal sufficient for $\mathbf{X}_1\boldsymbol{\beta}_1$, as observed by Isotalo and Puntanen (2006a, Th. 2).

In Proposition 7.3 we collect some interesting properties of linearly sufficient statistics of $\boldsymbol{\mu}_1$. See Markiewicz and Puntanen (2019a, Th. 2).

Proposition 7.3. *Let $\boldsymbol{\mu}_1 = \mathbf{X}_1\boldsymbol{\beta}_1$ be estimable under $\mathscr{M}_{12}$ and let $\mathbf{W} \in \mathcal{W}$. Then the statistic $\mathbf{Fy}$ is linearly sufficient for $\boldsymbol{\mu}_1$ under $\mathscr{M}_{12}$ if and only if*

$$\mathscr{C}(\mathbf{W}\dot{\mathbf{M}}_{2W}\mathbf{X}_1) \subset \mathscr{C}(\mathbf{WF}') \,, \tag{7.17}$$

where $\dot{\mathbf{M}}_{2W} = \mathbf{M}_2(\mathbf{M}_2\mathbf{W}\mathbf{M}_2)^-\mathbf{M}_2$, or, equivalently, by Lemma 1.3,

$$\mathscr{C}\{[\mathbf{I}_n - \mathbf{X}_2(\mathbf{X}_2'\mathbf{W}^+\mathbf{X}_2)^-\mathbf{X}_2'\mathbf{W}^+]\mathbf{X}_1\} \subset \mathscr{C}(\mathbf{WF}') \,. \tag{7.18}$$

Moreover, the following statements hold.

(a) $\mathbf{M}_2\mathbf{y}$ *is linearly sufficient for* $\boldsymbol{\mu}_1$.

(b) *If* $\mathscr{C}(\mathbf{A}') = \mathscr{C}(\mathbf{M}_2)$ *then* $\mathbf{Ay}$ *is linearly sufficient for* $\boldsymbol{\mu}_1$.

(c) $\dot{\mathbf{M}}_{2W}\mathbf{y}$ *is linearly sufficient for* $\boldsymbol{\mu}_1$.

(d) $\mathbf{M}_2\mathbf{y}$ *is linearly minimal sufficient for* $\boldsymbol{\mu}_1$ *if and only if* $\mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{V})$.

(e) $\mathbf{X}_1'\dot{\mathbf{M}}_{2W}\mathbf{y}$ *is linearly minimal sufficient for* $\boldsymbol{\mu}_1$.

(f) $\mathbf{Q}_{\mathbf{M}_2\mathbf{V}\mathbf{M}}\mathbf{M}_2\mathbf{y}$ *is linearly sufficient for* $\boldsymbol{\mu}_1$.

(g) *If* $\mathscr{C}(\mathbf{X}_1) \subset \mathscr{C}(\mathbf{X}_2 : \mathbf{V})$, (7.17) *becomes*

$$\mathscr{C}(\mathbf{W}\dot{\mathbf{M}}_2\mathbf{X}_1) \subset \mathscr{C}(\mathbf{WF}') \,, \quad \textit{where} \;\; \dot{\mathbf{M}}_2 = \mathbf{M}_2(\mathbf{M}_2\mathbf{V}\mathbf{M}_2)^-\mathbf{M}_2 \,, \tag{7.19}$$

and $\mathbf{V}\mathbf{M}_2(\mathbf{M}_2\mathbf{V}\mathbf{M}_2)^-\mathbf{M}_2\mathbf{y} \in \mathcal{S}(\boldsymbol{\mu}_1)$.

(h) *If* $\mathbf{V}$ *is positive definite,* (7.17) *becomes* $\mathscr{C}(\dot{\mathbf{M}}_2\mathbf{X}_1) \subset \mathscr{C}(\mathbf{F}')$.

(i) *If* $\boldsymbol{\beta}_1$ *is estimable under* $\mathscr{M}_{12}$, *then*

$$\mathbf{Fy} \in \mathcal{S}(\mathbf{X}_1\boldsymbol{\beta}_1 \mid \mathscr{M}_{12}) \iff \mathbf{Fy} \in \mathcal{S}(\boldsymbol{\beta}_1 \mid \mathscr{M}_{12}) \,. \tag{7.20}$$

Part (f) of Proposition 7.3 is due to Groß and Puntanen (2000, Sec. 3). They consider the following choice of $\mathbf{F}$:

$$\begin{aligned}\mathbf{F} &= (\mathbf{I}_n - \mathbf{P}_{\mathbf{M}_2\mathbf{V}\mathbf{M}})\mathbf{M}_2 = \mathbf{M}_2 - \mathbf{P}_{\mathbf{M}_2\mathbf{V}\mathbf{M}} \\ &= \mathbf{I}_n - \mathbf{P}_{\mathbf{X}_2} - \mathbf{P}_{\mathbf{M}_2\mathbf{V}\mathbf{M}} = \mathbf{I}_n - \mathbf{P}_{(\mathbf{X}_2 : \mathbf{V}\mathbf{M})} \,,\end{aligned} \tag{7.21}$$

where we have used part (a) of Lemma 1.5. For this $\mathbf{F}$ we have

$$\mathscr{C}(\mathbf{F}') = \mathscr{C}(\mathbf{F}) = \mathscr{C}(\mathbf{M}_2) \cap \mathscr{C}(\mathbf{V}\mathbf{M})^{\perp} = \mathscr{C}(\mathbf{M}_2) \cap \mathscr{N}(\mathbf{M}\mathbf{V}) \,. \tag{7.22}$$

It is easy to observe that $\mathscr{C}(\dot{\mathbf{M}}_{2W}\mathbf{X}_1) \subset \mathscr{C}(\mathbf{F}')$, and thereby (7.17) holds.

The next proposition gives some further characterizations for $\mathbf{F}\mathbf{y}$ being linearly sufficient for $\boldsymbol{\mu}_1$. For more details, see Markiewicz and Puntanen (2019a), who show, for example, that the following properties hold:

$$\begin{aligned}\operatorname{cov}(\tilde{\boldsymbol{\mu}}_1 \mid \mathscr{M}_{12}) &= \mathbf{X}_1(\mathbf{X}_1'\dot{\mathbf{M}}_{2W}\mathbf{X}_1)^{-}\mathbf{X}_1' - \mathbf{T}_1 \\ &= \mathbf{X}_1[\mathbf{X}_1'\mathbf{M}_2(\mathbf{M}_2\mathbf{W}\mathbf{M}_2)^{-}\mathbf{M}_2\mathbf{X}_1]^{-}\mathbf{X}_1' - \mathbf{T}_1 \\ &= \mathbf{X}_1[\mathbf{X}_1'\mathbf{W}^{+1/2}\mathbf{P}_{\mathbf{W}^{1/2}\mathbf{M}_2}\mathbf{W}^{+1/2}\mathbf{X}_1]^{-}\mathbf{X}_1' - \mathbf{T}_1 \,,\end{aligned} \tag{7.23a}$$

$$\tilde{\boldsymbol{\mu}}_1(\mathscr{T}_{12}) = \mathbf{X}_1[\mathbf{X}_1'\mathbf{N}_2(\mathbf{N}_2\mathbf{W}\mathbf{N}_2)^{-}\mathbf{N}_2\mathbf{X}_1]^{-}\mathbf{X}_1'\mathbf{N}_2(\mathbf{N}_2\mathbf{W}\mathbf{N}_2)^{-}\mathbf{N}_2\mathbf{y}, \tag{7.23b}$$

$$\begin{aligned}\operatorname{cov}(\tilde{\boldsymbol{\mu}}_1 \mid \mathscr{T}_{12}) &= \mathbf{X}_1[\mathbf{X}_1'\mathbf{N}_2(\mathbf{N}_2\mathbf{W}\mathbf{N}_2)^{-}\mathbf{N}_2\mathbf{X}_1]^{-}\mathbf{X}_1' - \mathbf{T}_1 \\ &= \mathbf{X}_1(\mathbf{X}_1'\mathbf{W}^{+1/2}\mathbf{P}_{\mathbf{W}^{1/2}\mathbf{N}_2}\mathbf{W}^{+1/2}\mathbf{X}_1)^{-}\mathbf{X}_1' - \mathbf{T}_1 \,,\end{aligned} \tag{7.23c}$$

where $\mathbf{W} = \mathbf{V} + \mathbf{X}\mathbf{U}\mathbf{U}'\mathbf{X}' \in \mathcal{W}$, $\mathbf{U}' = (\mathbf{U}_1' : \mathbf{U}_2')$ and

$$\mathbf{T}_1 = \mathbf{X}_1\mathbf{U}_1\mathbf{U}_1'\mathbf{X}_1', \quad \mathbf{N}_2 = \mathbf{P}_{\mathbf{F}'\mathbf{Q}_{\mathbf{F}\mathbf{X}_2}} = \mathbf{P}_{\mathscr{C}(\mathbf{F}')\cap\mathscr{C}(\mathbf{M}_2)} \,. \tag{7.24}$$

Proposition 7.4. *Let $\boldsymbol{\mu}_1 = \mathbf{X}_1\boldsymbol{\beta}_1$ be estimable under $\mathscr{T}_{12}$ (and hence under $\mathscr{M}_{12}$), let $\mathbf{W} \in \mathcal{W}$ and denote $\mathbf{N}_2 = \mathbf{P}_{\mathbf{F}'\mathbf{Q}_{\mathbf{F}\mathbf{X}_2}} = \mathbf{P}_{\mathscr{C}(\mathbf{F}')\cap\mathscr{C}(\mathbf{M}_2)}$. Then*

$$\operatorname{cov}(\tilde{\boldsymbol{\mu}}_1 \mid \mathscr{M}_{12}) \leq_{\mathrm{L}} \operatorname{cov}(\tilde{\boldsymbol{\mu}}_1 \mid \mathscr{T}_{12}) \,. \tag{7.25}$$

Moreover, the following statements are equivalent:

(a) $\operatorname{cov}(\tilde{\boldsymbol{\mu}}_1 \mid \mathscr{M}_{12}) = \operatorname{cov}(\tilde{\boldsymbol{\mu}}_1 \mid \mathscr{T}_{12})$.

(b) $\mathscr{C}(\mathbf{X}_1) \subset \mathscr{C}(\mathbf{X}_2 : \mathbf{W}\mathbf{N}_2) = \mathscr{C}(\mathbf{X}_2 : \mathbf{M}_2\mathbf{W}\mathbf{N}_2)$.

(c) $\mathscr{C}(\mathbf{X}_1) \subset \mathscr{C}(\mathbf{X}_2) \oplus [\mathscr{C}(\mathbf{W}\mathbf{F}') \cap \mathscr{C}(\mathbf{W}\mathbf{M}_2)]$.

(d) $\mathbf{W}\mathbf{M}_2(\mathbf{M}_2\mathbf{W}\mathbf{M}_2)^{-}\mathbf{M}_2\mathbf{X}_1 = \mathbf{W}\mathbf{N}_2(\mathbf{N}_2\mathbf{W}\mathbf{N}_2)^{-}\mathbf{N}_2\mathbf{X}_1$.

(e) *The statistic $\mathbf{F}\mathbf{y}$ is linearly sufficient for $\mathbf{X}_1\boldsymbol{\beta}_1$ under $\mathscr{M}_{12}$.*

Notice the correspondence between the notation $\mathbf{N}$ and $\mathbf{N}_2$:

$$\mathbf{N} = \mathbf{P}_{\mathbf{F}'\mathbf{Q}_{\mathbf{F}\mathbf{X}}} = \mathbf{P}_{\mathscr{C}(\mathbf{F}')\cap\mathscr{C}(\mathbf{M})} \,, \quad \mathbf{N}_2 = \mathbf{P}_{\mathbf{F}'\mathbf{Q}_{\mathbf{F}\mathbf{X}_2}} = \mathbf{P}_{\mathscr{C}(\mathbf{F}')\cap\mathscr{C}(\mathbf{M}_2)} \,. \tag{7.26}$$

Consider the small model $\mathscr{M}_1 = \{\mathbf{y}, \mathbf{X}_1\boldsymbol{\beta}_1, \mathbf{V}\}$ and full model $\mathscr{M}_{12}$. We may ask, for example, what is the condition that the following implication holds:

$$\mathbf{F}\mathbf{y} \in \mathcal{S}(\boldsymbol{\mu}_1 \mid \mathscr{M}_1) \implies \mathbf{F}\mathbf{y} \in \mathcal{S}(\boldsymbol{\mu}_1 \mid \mathscr{M}_{12}) . \tag{7.27}$$

Markiewicz and Puntanen (2019a, Th. 4) provided the following solution to (7.27).

Proposition 7.5. *Consider the models $\mathscr{M}_{12}$ and $\mathscr{M}_1$ and suppose that $\boldsymbol{\mu}_1 = \mathbf{X}_1\boldsymbol{\beta}_1$ is estimable under $\mathscr{M}_{12}$ and $\mathscr{C}(\mathbf{X}_2) \subset \mathscr{C}(\mathbf{X}_1 : \mathbf{V})$. Then the following statements are equivalent:*

(a) $\mathbf{X}_1'\mathbf{W}_1^{+}\mathbf{X}_2 = \mathbf{0}$,

(b) $\mathrm{BLUE}(\boldsymbol{\mu}_1 \mid \mathscr{M}_1) = \mathrm{BLUE}(\boldsymbol{\mu}_1 \mid \mathscr{M}_{12})$ *with probability* 1,

(c) $\mathbf{F}\mathbf{y} \in \mathcal{S}(\boldsymbol{\mu}_1 \mid \mathscr{M}_1) \iff \mathbf{F}\mathbf{y} \in \mathcal{S}(\boldsymbol{\mu}_1 \mid \mathscr{M}_{12})$.

From a different angle, the linear sufficiency in a partitioned linear model has been considered, e.g., in Isotalo and Puntanen (2006a, 2009), Markiewicz and Puntanen (2009), and Kala and Pordzik (2009). Baksalary (1984, 1987, Sec. 3.3, Sec. 5) considered linear sufficiency under $\mathscr{M}_{12}$ and $\mathscr{M}_1$ assuming that $\mathbf{V} = \mathbf{I}_n$.

8 Mutual relations of linear sufficiencies

In this section we explore the mutual relations of linear sufficiencies. In addition, we go through some interesting connections between the covariance matrices of the BLUPs and the linear sufficiencies. We also comment on the upper bounds of the Euclidean distance between the BLUPs when the prediction is based on the original model $\mathscr{M}$ and when it is based on the transformed model $\mathscr{T}$.

Following Markiewicz and Puntanen (2018a), let us take a closer look at $\tilde{\mathbf{y}}_* = \tilde{\boldsymbol{\mu}}_* + \tilde{\boldsymbol{\varepsilon}}_*$ and $\tilde{\mathbf{y}}_{t*} = \tilde{\boldsymbol{\mu}}_{t*} + \tilde{\boldsymbol{\varepsilon}}_{t*}$. It can be seen that $\tilde{\boldsymbol{\mu}}_*$ and $\tilde{\boldsymbol{\varepsilon}}_*$ are uncorrelated and the corresponding property holds also for $\tilde{\boldsymbol{\mu}}_{t*}$ and $\tilde{\boldsymbol{\varepsilon}}_{t*}$. Hence

$$\mathrm{cov}(\tilde{\mathbf{y}}_*) = \mathrm{cov}(\tilde{\boldsymbol{\mu}}_*) + \mathrm{cov}(\tilde{\boldsymbol{\varepsilon}}_*) , \quad \mathrm{cov}(\tilde{\mathbf{y}}_{t*}) = \mathrm{cov}(\tilde{\boldsymbol{\mu}}_{t*}) + \mathrm{cov}(\tilde{\boldsymbol{\varepsilon}}_{t*}) . \tag{8.1}$$

Now we have

$$\tilde{\boldsymbol{\varepsilon}}_* = \mathbf{V}_{21}\mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}\mathbf{y}, \quad \tilde{\boldsymbol{\varepsilon}}_{t*} = \mathbf{V}_{21}\mathbf{N}(\mathbf{N}\mathbf{V}\mathbf{N})^{-}\mathbf{N}\mathbf{y}, \tag{8.2}$$

with covariance matrices

$$\mathrm{cov}(\tilde{\boldsymbol{\varepsilon}}_*) = \mathbf{V}_{21}\mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}\mathbf{V}_{12} , \quad \mathrm{cov}(\tilde{\boldsymbol{\varepsilon}}_{t*}) = \mathbf{V}_{21}\mathbf{N}(\mathbf{N}\mathbf{V}\mathbf{N})^{-}\mathbf{N}\mathbf{V}_{12} . \tag{8.3}$$

Notice that in view of (1.21), the matrix products $\mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}$ and $\mathbf{N}(\mathbf{N}\mathbf{V}\mathbf{N})^{-}\mathbf{N}$ in (8.2) and (8.3) could be replaced with $(\mathbf{M}\mathbf{V}\mathbf{M})^{+}$ and $(\mathbf{N}\mathbf{V}\mathbf{N})^{+}$, respectively.

Straightforward calculation shows that $\mathrm{cov}(\tilde{\boldsymbol{\varepsilon}}_*, \tilde{\boldsymbol{\varepsilon}}_{t*}) = \mathrm{cov}(\tilde{\boldsymbol{\varepsilon}}_{t*})$, and

$$\mathrm{cov}(\tilde{\boldsymbol{\varepsilon}}_* - \tilde{\boldsymbol{\varepsilon}}_{t*}) = \mathrm{cov}(\tilde{\boldsymbol{\varepsilon}}_*) - \mathrm{cov}(\tilde{\boldsymbol{\varepsilon}}_{t*}) , \tag{8.4}$$

and thereby we have the Löwner ordering $\mathrm{cov}(\tilde{\boldsymbol{\varepsilon}}_*) \geq_{\mathrm{L}} \mathrm{cov}(\tilde{\boldsymbol{\varepsilon}}_{t*})$. Moreover, in view of Lemma 1.6 and (8.4), the equality $\tilde{\boldsymbol{\varepsilon}}_* = \tilde{\boldsymbol{\varepsilon}}_{t*}$ holds with probability 1 if and only if $\mathrm{cov}(\tilde{\boldsymbol{\varepsilon}}_*) = \mathrm{cov}(\tilde{\boldsymbol{\varepsilon}}_{t*})$. This confirms that (b) of Proposition 4.7 indeed holds and so:

$$\mathbf{Fy} \in \mathcal{S}(\boldsymbol{\varepsilon}_*) \iff \mathrm{cov}(\tilde{\boldsymbol{\varepsilon}}_*) = \mathrm{cov}(\tilde{\boldsymbol{\varepsilon}}_{t*})\,. \tag{8.5}$$

Proposition 8.1 provides some further characterizations for $\mathbf{Fy}$ being linearly sufficient foe $\boldsymbol{\varepsilon}_*$; see Markiewicz and Puntanen (2018a, Th. 2).

Proposition 8.1. *Denoting* $\mathbf{N} = \mathbf{P}_{\mathbf{F}'\mathbf{Q}_{\mathbf{FX}}}$ *, the following statements are equivalent:*

(a) $\mathbf{V}_{21}\mathbf{M} = \mathbf{V}_{21}\mathbf{N}(\mathbf{NVN})^-\mathbf{NVM}$,

(b) $\mathscr{C}(\mathbf{MV}_{12}) \subset \mathscr{C}(\mathbf{MVN}) = \mathscr{C}(\mathbf{MVF}'\mathbf{Q}_{\mathbf{FX}})$,

(c) $\mathscr{C}(\mathbf{V}_{12}) \subset \mathscr{C}(\mathbf{VN} : \mathbf{X}) = \mathscr{C}(\mathbf{VF}'\mathbf{Q}_{\mathbf{FX}} : \mathbf{X})$,

(d) $\mathbf{V}_{21}\mathbf{M}(\mathbf{MVM})^-\mathbf{MV}_{12} = \mathbf{V}_{21}\mathbf{N}(\mathbf{NVN})^-\mathbf{NV}_{12}$.

Moreover, each of the above conditions is a necessary and sufficient condition for the statistic $\mathbf{Fy}$ *to be linearly sufficient for* $\boldsymbol{\varepsilon}_*$ *under* $\mathscr{M}_*$.

Recall that

$$\text{(i) } \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon\varepsilon}} := \mathrm{cov}(\tilde{\boldsymbol{\varepsilon}}_* - \tilde{\boldsymbol{\varepsilon}}_{t*}) = \mathrm{cov}(\tilde{\boldsymbol{\varepsilon}}_*) - \mathrm{cov}(\tilde{\boldsymbol{\varepsilon}}_{t*})\,, \quad \text{(ii) } \mathrm{cov}(\tilde{\boldsymbol{\varepsilon}}_{t*}) \leq_{\mathrm{L}} \mathrm{cov}(\tilde{\boldsymbol{\varepsilon}}_*)\,. \tag{8.6}$$

Moreover, Markiewicz and Puntanen (2018a, Sec. 5) showed that

$$\text{(i) } \boldsymbol{\Sigma}_{\boldsymbol{\mu\mu}} := \mathrm{cov}(\tilde{\boldsymbol{\mu}}_* - \tilde{\boldsymbol{\mu}}_{t*}) = \mathrm{cov}(\tilde{\boldsymbol{\mu}}_{t*}) - \mathrm{cov}(\tilde{\boldsymbol{\mu}}_*)\,, \quad \text{(ii) } \mathrm{cov}(\tilde{\boldsymbol{\mu}}_{t*}) \geq \mathrm{cov}(\tilde{\boldsymbol{\mu}}_*)\,. \tag{8.7}$$

In other words,

$$\boldsymbol{\Sigma}_{\boldsymbol{\mu\mu}} = \mathrm{cov}(\tilde{\boldsymbol{\mu}}_* - \tilde{\boldsymbol{\mu}}_{t*}) = \mathrm{cov}(\tilde{\boldsymbol{\mu}}_{t*}) - \mathrm{cov}(\tilde{\boldsymbol{\mu}}_*)\,, \tag{8.8a}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon\varepsilon}} = \mathrm{cov}(\tilde{\boldsymbol{\varepsilon}}_* - \tilde{\boldsymbol{\varepsilon}}_{t*}) = \mathrm{cov}(\tilde{\boldsymbol{\varepsilon}}_*) - \mathrm{cov}(\tilde{\boldsymbol{\varepsilon}}_{t*})\,. \tag{8.8b}$$

However, the following does *not* necessarily hold:

$$\boldsymbol{\Sigma}_{\mathbf{yy}} := \mathrm{cov}(\tilde{\mathbf{y}}_* - \tilde{\mathbf{y}}_{t*}) = \mathrm{cov}(\tilde{\mathbf{y}}_*) - \mathrm{cov}(\tilde{\mathbf{y}}_{t*})\,. \tag{8.8c}$$

In view of Lemma 1.6, the equality $\tilde{\mathbf{y}}_* = \tilde{\mathbf{y}}_{t*}$ holds with probability 1 if and only if $\mathrm{cov}(\tilde{\mathbf{y}}_* - \tilde{\mathbf{y}}_{t*}) = \mathbf{0}$, which thereby is a condition for $\mathbf{Fy}$ being linearly sufficient for $\mathbf{y}_*$. Thus in terms of covariance matrices, we have

$$\mathbf{Fy} \in \mathcal{S}(\boldsymbol{\mu}_*) \iff \mathrm{cov}(\tilde{\boldsymbol{\mu}}_*) = \mathrm{cov}(\tilde{\boldsymbol{\mu}}_{t*})\,, \tag{8.9a}$$

$$\mathbf{Fy} \in \mathcal{S}(\boldsymbol{\varepsilon}_*) \iff \mathrm{cov}(\tilde{\boldsymbol{\varepsilon}}_*) = \mathrm{cov}(\tilde{\boldsymbol{\varepsilon}}_{t*})\,, \tag{8.9b}$$

$$\mathbf{Fy} \in \mathcal{S}(\mathbf{y}_*) \iff \mathrm{cov}(\tilde{\mathbf{y}}_* - \tilde{\mathbf{y}}_{t*}) = \mathbf{0}\,. \tag{8.9c}$$

The above statements (8.9a)–(8.9c) are all appearing in Propositions 4.5 and 4.7.

Denoting

$$\boldsymbol{\Sigma}_{\boldsymbol{\mu}\boldsymbol{\varepsilon}} = \operatorname{cov}(\tilde{\boldsymbol{\mu}}_* - \tilde{\boldsymbol{\mu}}_{t*},\, \tilde{\boldsymbol{\varepsilon}}_* - \tilde{\boldsymbol{\varepsilon}}_{t*})\,, \tag{8.10}$$

it can be shown that $\boldsymbol{\Sigma}_{\boldsymbol{\mu}\boldsymbol{\varepsilon}} = -\operatorname{cov}(\tilde{\boldsymbol{\mu}}_{t*},\, \tilde{\boldsymbol{\varepsilon}}_*)$ and

$$\boldsymbol{\Sigma}_{\mathbf{yy}} = \boldsymbol{\Sigma}_{\boldsymbol{\mu}\boldsymbol{\mu}} + \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}} + (\boldsymbol{\Sigma}_{\boldsymbol{\mu}\boldsymbol{\varepsilon}} + \boldsymbol{\Sigma}'_{\boldsymbol{\mu}\boldsymbol{\varepsilon}})\,. \tag{8.11}$$

Markiewicz and Puntanen (2018a, Th. 3) expressed the linear sufficiency of $\mathbf{Fy}$ for $\mathbf{y}_*$ in terms of covariance matrices as follows.

Proposition 8.2. *Denoting* $\boldsymbol{\Sigma}_{\boldsymbol{\mu}\boldsymbol{\varepsilon}} = \operatorname{cov}(\tilde{\boldsymbol{\mu}}_* - \tilde{\boldsymbol{\mu}}_{t*},\, \tilde{\boldsymbol{\varepsilon}}_* - \tilde{\boldsymbol{\varepsilon}}_{t*})$, *the following statements are equivalent:*

(a) $\mathbf{Fy}$ *is* BLUP-*sufficient for* $\mathbf{y}_*$,

(b) $\operatorname{cov}(\tilde{\mathbf{y}}_*) = \operatorname{cov}(\tilde{\mathbf{y}}_{t*})$ *and* $\operatorname{cov}(\tilde{\boldsymbol{\varepsilon}}_{t*}) - \operatorname{cov}(\tilde{\boldsymbol{\varepsilon}}_*) = \frac{1}{2}(\boldsymbol{\Sigma}_{\boldsymbol{\mu}\boldsymbol{\varepsilon}} + \boldsymbol{\Sigma}'_{\boldsymbol{\mu}\boldsymbol{\varepsilon}})$.

The mutual relations of the linear sufficiency of $\mathbf{Fy}$ for $\mathbf{X}_*\boldsymbol{\beta}$, $\boldsymbol{\varepsilon}_*$, and $\mathbf{y}_*$, can be characterized as follows; see Markiewicz and Puntanen (2018a, Th. 5).

Proposition 8.3. *Consider the following three statements:*

(a) $\mathbf{Fy} \in \mathcal{S}(\mathbf{X}_*\boldsymbol{\beta})$, (b) $\mathbf{Fy} \in \mathcal{S}(\boldsymbol{\varepsilon}_*)$, (c) $\mathbf{Fy} \in \mathcal{S}(\mathbf{y}_*)$.

Then above, any two conditions together imply the third one. Moreover, the equality

$$\boldsymbol{\Sigma}_{\boldsymbol{\mu}\boldsymbol{\varepsilon}} = -\boldsymbol{\Sigma}'_{\boldsymbol{\mu}\boldsymbol{\varepsilon}}\,, \quad \textit{where} \;\; \boldsymbol{\Sigma}_{\boldsymbol{\mu}\boldsymbol{\varepsilon}} = -\operatorname{cov}(\tilde{\boldsymbol{\mu}}_{t*},\, \tilde{\boldsymbol{\varepsilon}}_*)\,, \tag{8.12}$$

is a necessary and sufficient condition for the implication (c) $\implies$ (a) *and* (b).

We end this section by some remarks concerning the upper bounds for the Euclidean distance between the BLUPs when the prediction is based on the original model $\mathscr{M}$ and when it is based on the transformed model $\mathscr{T}$.

Markiewicz and Puntanen (2018b) considered the Euclidean norms of differences

$$\tilde{\boldsymbol{\varepsilon}}_* - \tilde{\boldsymbol{\varepsilon}}_{t*}\,, \quad \tilde{\boldsymbol{\mu}}_* - \tilde{\boldsymbol{\mu}}_{t*} \quad \text{and} \quad \tilde{\mathbf{y}}_* - \tilde{\mathbf{y}}_{t*}\,. \tag{8.13}$$

For this purpose, let $\operatorname{ch}_1(\cdot)$ denote the largest eigenvalue of the matrix argument and let the matrix norm be defined as $\|\mathbf{A}\|_2 = \sqrt{\operatorname{ch}_1(\mathbf{AA}')}$. In particular, $\|\mathbf{a}\|_2^2 = \mathbf{a}'\mathbf{a}$, where $\mathbf{a} \in \mathbb{R}^n$. Without going into any further details we cite below the result of Markiewicz and Puntanen (2018b, Th. 4.1).

Proposition 8.4. *Consider the model* $\mathscr{M}_*$. *Then for all* $\mathbf{y} = \mathbf{Xa} + \mathbf{VMb}$,

$$\begin{aligned} \|\tilde{\boldsymbol{\varepsilon}}_* - \tilde{\boldsymbol{\varepsilon}}_{t*}\|_2^2 &= \|\mathrm{BLUP}(\boldsymbol{\varepsilon}_* \mid \mathscr{M}_*) - \mathrm{BLUP}(\boldsymbol{\varepsilon}_* \mid \mathscr{T}_*)\|_2^2 \\ &\le \operatorname{ch}_1(\mathbf{AA}')\, \mathbf{b}'\mathbf{Mb} := \alpha_1\,, \end{aligned} \tag{8.14}$$

where

$$\mathbf{A} = \mathbf{V}_{21}\mathbf{M}(\mathbf{I}_n - \mathbf{B}) \in \mathbb{R}^{q\times n}, \quad \mathbf{B} = \mathbf{N}(\mathbf{NVN})^-\mathbf{NVM} \in \mathbb{R}^{n\times n}, \tag{8.15}$$

and $\mathbf{N} = \mathbf{P}_{\mathbf{F}'\mathbf{Q}_{\mathbf{FX}}}$. *If* $\mathbf{b} \notin \mathscr{C}(\mathbf{X})$, *then the upper bound* α_1 *in* (8.14) *is equal to zero if and only if* $\mathbf{A} = \mathbf{0}$, *i.e.,* $\mathbf{Fy}$ *is linearly sufficient for* $\boldsymbol{\varepsilon}_*$.

Let us take a look at the Euclidean distance between the BLUEs of $\boldsymbol{\mu}_* = \mathbf{X}_*\boldsymbol{\beta}$ in the original and the transformed model, assuming that $\mathbf{X}_* = \mathbf{LFX}$ for some $\mathbf{L} \in \mathbb{R}^{q\times f}$. Then, for all $\mathbf{y} \in \mathscr{C}(\mathbf{W})$, and $\boldsymbol{\mu}_* = \mathbf{LFX}\boldsymbol{\beta}$, we have, using the consistency and multiplicativity of the matrix norm $\|\cdot\|_2$, see, for example, Ben-Israel and Greville (2003, pp. 19–20),

$$
\begin{aligned}
\|\tilde{\boldsymbol{\mu}}_* - \tilde{\boldsymbol{\mu}}_{t*}\|_2^2 &= \|\mathbf{LF}(\mathbf{G}_t - \mathbf{G})\mathbf{y}\|_2^2 \\
&= \|\mathbf{LFG}_t\mathbf{VM}(\mathbf{MVM})^-\mathbf{My}\|_2^2 \\
&\le \|\mathbf{LFG}_t\mathbf{VM}\|_2^2 \; \|(\mathbf{MVM})^+\|_2^2 \; \|\mathbf{My}\|_2^2 \\
&= \|\mathbf{C}\|_2^2 \; \|(\mathbf{MVM})^+\|_2^2 \; \|\mathbf{My}\|_2^2 \\
&= \frac{a}{b^2}\,\mathbf{y}'\mathbf{My} := \alpha_2\,,
\end{aligned} \tag{8.16}
$$

where $\mathbf{G}$ and $\mathbf{G}_t$ are defined as in Proposition 4.4, $\mathbf{C} = \mathbf{LFG}_t\mathbf{VM}$, the scalar a is the largest eigenvalue of $\mathbf{CC}'$, and b is the smallest nonzero eigenvalue of $\mathbf{MVM}$. A model with property $\mathbf{VM} = \mathbf{0}$ is called a degenerated model, see Groß (2004, p. 317). If $\mathscr{M}$ is not a degenerated model then α_2 is zero if and only if $\mathbf{Fy}$ is linearly sufficient for $\boldsymbol{\mu}_*$. For (8.16), see also Kala et al. (2017b, Th. 5).

For the upper bound of $\|\tilde{\mathbf{y}}_* - \tilde{\mathbf{y}}_{t*}\|_2^2$, we refer to Markiewicz and Puntanen (2018b, Th. 4.2). The properties of the Euclidean norm of the difference $\mathrm{OLSE}(\mathbf{X}\boldsymbol{\beta}) - \mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta})$ have been studied by Baksalary and Kala (1978a, 1980) and for the BLUEs under two models by Hauke et al. (2012); see also Pordzik (2012), Baksalary et al. (2013), and Haslett et al. (2014).

9 Mixed linear model

In this section we consider the linear mixed model in the spirit of Isotalo et al. (2018) and Haslett et al. (2020), defined as

$$
\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Zu} + \mathbf{e}\,, \quad \text{denoted as} \quad \mathscr{L} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta} + \mathbf{Zu}, \mathbf{D}, \mathbf{R}, \mathbf{S}\}\,. \tag{9.1}
$$

Here $\mathbf{X}_{n\times p}$ and $\mathbf{Z}_{n\times q}$ are known matrices, $\boldsymbol{\beta} \in \mathbb{R}^p$ is a vector of unknown fixed effects, $\mathbf{u}$ is an unobservable vector (q elements) of random effects with $\mathrm{E}(\mathbf{u}) = \mathbf{0}$, $\mathrm{cov}(\mathbf{u}) = \mathbf{D}_{q\times q}$, $\mathrm{cov}(\mathbf{e},\mathbf{u}) = \mathbf{S}_{n\times q}$, and $\mathrm{E}(\mathbf{e}) = \mathbf{0}$, $\mathrm{cov}(\mathbf{e}) = \mathbf{R}_{n\times n}$. In this situation

$$
\mathrm{cov}\begin{pmatrix}\mathbf{e}\\ \mathbf{u}\end{pmatrix} = \begin{pmatrix}\mathbf{R} & \mathbf{S}\\ \mathbf{S}' & \mathbf{D}\end{pmatrix} =: \dot{\mathbf{V}}, \quad \mathrm{cov}\begin{pmatrix}\mathbf{y}\\ \mathbf{u}\end{pmatrix} = \begin{pmatrix}\boldsymbol{\Sigma} & \mathbf{ZD} + \mathbf{S}\\ (\mathbf{ZD} + \mathbf{S})' & \mathbf{D}\end{pmatrix}, \tag{9.2}
$$

and denoting $\mathbf{v} = \binom{\mathbf{e}}{\mathbf{u}}$,

$$
\begin{aligned}
\boldsymbol{\Sigma} &= \mathrm{cov}(\mathbf{y}) = \mathrm{cov}(\mathbf{e} + \mathbf{Zu}) = \mathrm{cov}[(\mathbf{I}_n : \mathbf{Z})\mathbf{v}] \\
&= (\mathbf{I}_n : \mathbf{Z})\dot{\mathbf{V}}(\mathbf{I}_n : \mathbf{Z})' = \mathbf{ZDZ}' + \mathbf{R} + \mathbf{ZS}' + \mathbf{SZ}'.
\end{aligned} \tag{9.3}
$$

The mixed model can be expressed as a version of the model with "new observations", the new observations being, for example, in $\mathbf{g} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Zu}$:

$$\mathscr{L}_* := \left\{ \begin{pmatrix} \mathbf{y} \\ \mathbf{g} \end{pmatrix}, \begin{pmatrix} \mathbf{X} \\ \mathbf{X} \end{pmatrix} \boldsymbol{\beta}, \begin{pmatrix} \boldsymbol{\Sigma} & (\mathbf{ZD} + \mathbf{S})\mathbf{Z}' \\ \mathbf{Z}(\mathbf{ZD} + \mathbf{S})' & \mathbf{ZDZ}' \end{pmatrix} \right\}. \tag{9.4}$$

Corresponding to (1.4) and (1.8), we have

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} = \mathbf{Zu} + \mathbf{e}, \quad \operatorname{cov}(\boldsymbol{\varepsilon}) = \operatorname{cov}(\mathbf{y}) = \boldsymbol{\Sigma}, \tag{9.5a}$$

$$\mathbf{y}_* = \mathbf{g} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Zu}, \quad \mathbf{X}_* = \mathbf{X}, \tag{9.5b}$$

$$\boldsymbol{\varepsilon}_* = \mathbf{Zu}, \quad \operatorname{cov}(\boldsymbol{\varepsilon}_*) = \mathbf{ZDZ}', \quad \operatorname{cov}(\boldsymbol{\varepsilon}, \boldsymbol{\varepsilon}_*) = (\mathbf{ZD} + \mathbf{S})\mathbf{Z}'. \tag{9.5c}$$

Choosing the "new observations" being as $\mathbf{u}$, we get

$$\left\{ \begin{pmatrix} \mathbf{y} \\ \mathbf{u} \end{pmatrix}, \begin{pmatrix} \mathbf{X} \\ \mathbf{0} \end{pmatrix} \boldsymbol{\beta}, \begin{pmatrix} \boldsymbol{\Sigma} & \mathbf{ZD} + \mathbf{S} \\ (\mathbf{ZD} + \mathbf{S})' & \mathbf{D} \end{pmatrix} \right\}. \tag{9.6}$$

Now, see, e.g., Haslett et al. (2015, Lemma 2), under the mixed model $\mathscr{L}$, $\mathbf{B}_1\mathbf{y}$ is the BLUE for $\mathbf{X}\boldsymbol{\beta}$ and $\mathbf{B}_2\mathbf{y}$ is the BLUP for $\mathbf{Zu}$ if and only if

$$\begin{pmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \end{pmatrix} (\mathbf{X} : \boldsymbol{\Sigma}\mathbf{M}) = \begin{pmatrix} \mathbf{X} & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}(\mathbf{ZD} + \mathbf{S})'\mathbf{M} \end{pmatrix} = \begin{pmatrix} \mathbf{X} & \mathbf{0} \\ \mathbf{0} & \operatorname{cov}(\mathbf{g}, \mathbf{y})\mathbf{M} \end{pmatrix}. \tag{9.7}$$

Similarly, $\mathbf{B}_3\mathbf{y}$ is the BLUP for $\mathbf{g} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Zu}$ if and only if

$$\mathbf{B}_3(\mathbf{X} : \boldsymbol{\Sigma}\mathbf{M}) = \big[\mathbf{X} : \mathbf{Z}(\mathbf{ZD} + \mathbf{S})'\mathbf{M}\big] = \big[\mathbf{X} : \operatorname{cov}(\mathbf{g}, \mathbf{y})\mathbf{M}\big]. \tag{9.8}$$

Hence $(\mathbf{B}_1 + \mathbf{B}_2)(\mathbf{X} : \boldsymbol{\Sigma}\mathbf{M}) = \mathbf{B}_3(\mathbf{X} : \boldsymbol{\Sigma}\mathbf{M})$ and the following holds:

$$\begin{aligned} \mathrm{BLUP}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Zu}) &= \mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta}) + \mathrm{BLUP}(\mathbf{Zu}) \\ &= \mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta}) + \mathbf{Z}\,\mathrm{BLUP}(\mathbf{u}), \end{aligned} \tag{9.9}$$

which can be denoted as $\tilde{\mathbf{g}} = \tilde{\boldsymbol{\mu}} + \mathbf{Z}\tilde{\mathbf{u}}$, and we have the following representations for the BLUP of $\mathbf{g} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Zu}$:

$$\begin{aligned} \mathrm{BLUP}(\mathbf{g}) = \tilde{\mathbf{g}} &= \mathbf{G}_m\mathbf{y} + \mathbf{Z}(\mathbf{ZD} + \mathbf{S})'\boldsymbol{\Sigma}^-(\mathbf{I}_n - \mathbf{G}_m)\mathbf{y} \\ &= \mathbf{G}_m\mathbf{y} + \mathbf{Z}(\mathbf{ZD} + \mathbf{S})'\mathbf{M}(\mathbf{M}\boldsymbol{\Sigma}\mathbf{M})^-\mathbf{My} \\ &= \tilde{\boldsymbol{\mu}} + \mathbf{Z}\tilde{\mathbf{u}}, \end{aligned} \tag{9.10}$$

where $\mathbf{G}_m = \mathbf{X}(\mathbf{X}'\mathbf{W}_m^-\mathbf{X})^-\mathbf{X}'\mathbf{W}_m^-$ and $\mathbf{W}_m \in \mathcal{W}_{\mathscr{L}}$,

$$\mathcal{W}_{\mathscr{L}} = \big\{\mathbf{W}_m \in \mathbb{R}^{n\times n} : \mathbf{W}_m = \boldsymbol{\Sigma} + \mathbf{XUU}'\mathbf{X}',\ \mathscr{C}(\mathbf{W}_m) = \mathscr{C}(\mathbf{X} : \boldsymbol{\Sigma})\big\}. \tag{9.11}$$

For example, in the simple situation when $\mathbf{X}$ has full column rank, $\mathbf{S} = \mathbf{0}$ and $\boldsymbol{\Sigma}_0 = \mathbf{ZDZ}' + \mathbf{R}$ is positive definite, we have

$$\mathrm{BLUP}(\mathbf{u}) = \mathbf{DZ}'\boldsymbol{\Sigma}_0^{-1}(\mathbf{I}_n - \mathbf{X}\tilde{\boldsymbol{\beta}}_0), \quad \tilde{\boldsymbol{\beta}}_0 = (\mathbf{X}'\boldsymbol{\Sigma}_0^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}_0^{-1}\mathbf{y}. \tag{9.12}$$

We futher note that $\mathbf{B}_4\mathbf{y}$ is the BLUP for $\boldsymbol{\eta} = \mathbf{K}\boldsymbol{\beta} + \mathbf{L}\mathbf{u}$ (where $\mathbf{K}\boldsymbol{\beta}$ is estimable) if and only if

$$\mathbf{B}_4(\mathbf{X} : \boldsymbol{\Sigma}\mathbf{M}) = \big[\mathbf{K} : \mathbf{L}(\mathbf{Z}\mathbf{D} + \mathbf{S})'\mathbf{M}\big] = \big[\mathbf{K} : \operatorname{cov}(\boldsymbol{\eta}, \mathbf{y})\mathbf{M}\big] . \tag{9.13}$$

Now obviously we get the following:

$$\mathbf{F}\mathbf{y} \in \mathcal{S}(\mathbf{g} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}) \iff \mathscr{C}\begin{pmatrix} \mathbf{X}' \\ \mathbf{M}(\mathbf{Z}\mathbf{D} + \mathbf{S})\mathbf{Z}' \end{pmatrix} \subset \mathscr{C}\begin{pmatrix} \mathbf{X}'\mathbf{F}' \\ \mathbf{M}\boldsymbol{\Sigma}\mathbf{F}' \end{pmatrix}, \tag{9.14a}$$

$$\mathbf{F}\mathbf{y} \in \mathcal{S}(\mathbf{X}\boldsymbol{\beta}) \iff \mathscr{C}\begin{pmatrix} \mathbf{X}' \\ \mathbf{0} \end{pmatrix} \subset \mathscr{C}\begin{pmatrix} \mathbf{X}'\mathbf{F}' \\ \mathbf{M}\boldsymbol{\Sigma}\mathbf{F}' \end{pmatrix}, \tag{9.14b}$$

$$\mathbf{F}\mathbf{y} \in \mathcal{S}(\mathbf{Z}\mathbf{u}) \iff \mathscr{C}\begin{pmatrix} \mathbf{0} \\ \mathbf{M}(\mathbf{Z}\mathbf{D} + \mathbf{S})\mathbf{Z}' \end{pmatrix} \subset \mathscr{C}\begin{pmatrix} \mathbf{X}'\mathbf{F}' \\ \mathbf{M}\boldsymbol{\Sigma}\mathbf{F}' \end{pmatrix}, \tag{9.14c}$$

and thereby the following holds.

Proposition 9.1. *Consider the mixed model $\mathscr{L} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}, \mathbf{D}, \mathbf{R}, \mathbf{S}\}$, and the following statements:*

(a) $\mathbf{F}\mathbf{y} \in \mathcal{S}(\mathbf{X}\boldsymbol{\beta})$, (b) $\mathbf{F}\mathbf{y} \in \mathcal{S}(\mathbf{Z}\mathbf{u})$, (c) $\mathbf{F}\mathbf{y} \in \mathcal{S}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u})$.

Then any of the two conditions above imply the third one.

As Haslett et al. (2014, Sec. 4) point out, premultiplying (9.1) by $\mathbf{M}$ produces a reduced model from which $\mathbf{X}\boldsymbol{\beta}$ has been eliminated. Actually, $\mathbf{M}\mathbf{y}$ is linearly sufficient for $\mathbf{Z}\mathbf{u}$ under the mixed model because then the condition (9.14c) becomes

$$\mathscr{C}[\mathbf{M}(\mathbf{Z}\mathbf{D} + \mathbf{S})\mathbf{Z}'] \subset \mathscr{C}(\mathbf{M}\boldsymbol{\Sigma}\mathbf{M}) = \mathscr{C}(\mathbf{M}\boldsymbol{\Sigma}) , \tag{9.15}$$

which obviously holds in view of $\mathscr{C}[(\mathbf{Z}\mathbf{D} + \mathbf{S})\mathbf{Z}'] \subset \mathscr{C}(\boldsymbol{\Sigma})$.

Liu et al. (2008, p. 1511) have a slightly different definition for the linear sufficiency. According to them, the statistic $\mathbf{F}\mathbf{y}$ is BLUP-sufficient if for all predictable parametric functions $\boldsymbol{\eta} = \mathbf{K}\boldsymbol{\beta} + \mathbf{L}\mathbf{u}$ there exists a matrix $\mathbf{A}$ such that $\mathbf{A}\mathbf{F}\mathbf{y}$ is the BLUP for $\boldsymbol{\eta}$ in the original model. Since $\boldsymbol{\eta} = \mathbf{K}\boldsymbol{\beta} + \mathbf{L}\mathbf{u}$ is predictable if and only if $\mathbf{K} = \mathbf{J}\mathbf{X}$ for some matrix $\mathbf{J}$ (while $\mathbf{L}$ can be any conformable matrix) we can re-express this definition as follows.

Definition 9.1. *The statistic* $\mathbf{F}\mathbf{y}$ *is* BLUP-*sufficient under the model* $\mathscr{L}$ *if for all* $\mathbf{J}$ *and* $\mathbf{L}$, *there exists a matrix* $\mathbf{A}$ *such that* $\mathbf{A}\mathbf{F}\mathbf{y}$ *is the* BLUP *for* $\boldsymbol{\eta} = \mathbf{J}\mathbf{X}\boldsymbol{\beta} + \mathbf{L}\mathbf{u}$, *and then we denote* $\mathbf{F}\mathbf{y} \in \mathcal{S}_r(\boldsymbol{\beta}, \mathbf{u})$.

We see that the difference between Definitions 3.2 and 9.1 is that in Definition 3.2 our object of estimation/prediction is a given predictable combination of fixed parameters and random effect like $\mathbf{g} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}$ (where $\mathbf{X}$ and $\mathbf{Z}$ are given and fixed) while in Definition 9.1 we consider *all* predictable combinations of the type $\boldsymbol{\eta} = \mathbf{K}\boldsymbol{\beta} + \mathbf{L}\mathbf{u}$. Actually Kala and Pordzik (2009, p. 635) use the linear sufficiency concept in the spirit of Definition 9.1 when saying that a statistic $\mathbf{F}\mathbf{y}$ is linearly sufficient if it is linearly sufficient for all estimable parametric functions of the model.

Haslett et al. (2014, Th. 2) proved the following result.

Proposition 9.2. *Consider the mixed model* $\mathscr{L} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}, \mathbf{D}, \mathbf{R}, \mathbf{S}\}$, *and denote* $\boldsymbol{\Sigma} = \operatorname{cov}(\mathbf{y}) = \mathbf{ZDZ}' + \mathbf{R} + \mathbf{ZS}' + \mathbf{SZ}'$, *and let* $\mathbf{W} \in \mathcal{W}_{\mathscr{L}}$, *where the class* $\mathcal{W}_{\mathscr{L}}$ *of matrices is defined as in* (9.11). *Then the following statements are equivalent:*

(a) $\mathbf{Fy} \in \mathcal{S}(\mathbf{X}\boldsymbol{\beta}) \cap \mathcal{S}(\mathbf{u})$, *i.e.,* $\mathbf{Fy}$ *is linearly sufficient for* $\mathbf{X}\boldsymbol{\beta}$ *and for* $\mathbf{u}$.

(b) $\mathbf{Fy} \in \mathcal{S}_r(\boldsymbol{\beta}, \mathbf{u})$, *i.e.,* $\mathbf{Fy}$ *is linearly sufficient for every predictable* $\mathbf{K}\boldsymbol{\beta} + \mathbf{L}\mathbf{u}$.

(c) $\mathscr{C}\begin{pmatrix} \mathbf{X}' & \mathbf{0} \\ \mathbf{0} & \mathbf{M}(\mathbf{ZD} + \mathbf{S}) \end{pmatrix} \subset \mathscr{C}\begin{pmatrix} \mathbf{X}'\mathbf{F}' \\ \mathbf{M}\boldsymbol{\Sigma}\mathbf{F}' \end{pmatrix}$.

(d) $\mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{WF}')$ *and* $\mathscr{C}[\mathbf{M}(\mathbf{ZD} + \mathbf{S})] \subset \mathscr{C}(\mathbf{MWF}')$.

(e) $\mathscr{C}(\mathbf{X} : \mathbf{ZD} + \mathbf{S}) \subset \mathscr{C}(\mathbf{WF}')$.

As Isotalo et al. (2018, Sec. 5) point out, here is one further interesting link between the mixed model and the following extended partitioned model:

$$\mathscr{A} = \{\dot{\mathbf{y}}, \dot{\mathbf{X}}\boldsymbol{\pi}, \dot{\mathbf{V}}\} = \left\{ \begin{pmatrix} \mathbf{y} \\ \mathbf{y}_0 \end{pmatrix}, \begin{pmatrix} \mathbf{X} & \mathbf{Z} \\ \mathbf{0} & -\mathbf{I}_q \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{\gamma} \end{pmatrix}, \begin{pmatrix} \mathbf{R} & \mathbf{S} \\ \mathbf{S}' & \mathbf{D} \end{pmatrix} \right\}, \tag{9.16}$$

where both $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are *fixed* effects parameters. Expressed in error terms we have

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}\,, \tag{9.17a}$$

$$\mathbf{y}_0 = \qquad -\boldsymbol{\gamma} + \boldsymbol{\varepsilon}_0\,, \tag{9.17b}$$

where $\operatorname{cov}\binom{\mathbf{y}}{\mathbf{y}_0} = \operatorname{cov}\binom{\boldsymbol{\varepsilon}}{\boldsymbol{\varepsilon}_0} = \dot{\mathbf{V}}$. Premultiplying $\mathscr{A}$ by the matrix $\dot{\mathbf{F}} = (\mathbf{I}_n : \mathbf{Z})$, as in Arendacká and Puntanen (2015, Sec. 2), yields the equation

$$\mathbf{y} + \mathbf{Z}\mathbf{y}_0 = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\varepsilon}_0 + \boldsymbol{\varepsilon}\,, \tag{9.18}$$

and in matrix terms we get the transformed model

$$\mathscr{B} = \{\dot{\mathbf{F}}\dot{\mathbf{y}}, \dot{\mathbf{F}}\dot{\mathbf{X}}\boldsymbol{\pi}, \dot{\mathbf{F}}\dot{\mathbf{V}}\dot{\mathbf{F}}'\} = \{\mathbf{w}, \mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}\}\,, \tag{9.19}$$

where $\mathbf{w} = \mathbf{y} + \mathbf{Z}\mathbf{y}_0$ and

$$\boldsymbol{\Sigma} = \operatorname{cov}(\mathbf{w}) = \mathbf{ZDZ}' + \mathbf{R} + \mathbf{ZS}' + \mathbf{SZ}'\,. \tag{9.20}$$

Now (9.18) can be interpreted as a mixed model where the observable response is $\mathbf{w} = \mathbf{y} + \mathbf{Z}\mathbf{y}_0$, $\boldsymbol{\varepsilon}_0$ is the unobservable random effect, and using the mixed model notation we have

$$\mathscr{B} = \{\mathbf{w},\, \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\varepsilon}_0,\, \mathbf{D},\, \mathbf{R},\, \mathbf{S}\}\,. \tag{9.21}$$

It would be now interesting to know whether the BLUEs of $\mathbf{X}\boldsymbol{\beta}$ under $\mathscr{A}$ and $\mathscr{B}$ are equal. Isotalo et al. (2018) answer to this question using the linear sufficiency concept while Haslett et al. (2015) and Arendacká and Puntanen

(2015) solved this problem using different approach. Isotalo et al. (2018) show that $\dot{\mathbf{F}}\dot{\mathbf{y}} = (\mathbf{I}_n : \mathbf{Z})\dot{\mathbf{y}}$ is linearly sufficient for $\mathbf{X}\boldsymbol{\beta}$ and thereby

$$\mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta} \mid \mathscr{A}) = \mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta} \mid \mathscr{B})\,. \tag{9.22}$$

Corresponding considerations appear also in Baksalary and Kala (1986, Sec. 3) but without referring to the mixed model. The connection between the models $\mathscr{A}$ and $\mathscr{B}$ can be used as a tool to calculate the BLUEs and BLUPs in mixed model and it is often referred to as a Henderson's method; see, e.g., Henderson et al. (1959) and McCulloch et al. (2008, Ch. 8). As a reference to rank and inertia formulas for covariance matrices of BLUPs in linear mixed models we may mention Güler and Büyükkaya (2019).

10 Linear sufficiency in the misspecified linear model

Consider the models $\mathscr{M}_*$ and $\underline{\mathscr{M}}_*$, where $\mathscr{C}(\mathbf{X}_*') \subset \mathscr{C}(\mathbf{X}')$:

$$\mathscr{M}_* = \left\{ \begin{pmatrix} \mathbf{y} \\ \mathbf{y}_* \end{pmatrix}, \begin{pmatrix} \mathbf{X} \\ \mathbf{X}_* \end{pmatrix}\boldsymbol{\beta}, \begin{pmatrix} \mathbf{V} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{pmatrix} \right\}, \tag{10.1a}$$

$$\underline{\mathscr{M}}_* = \left\{ \begin{pmatrix} \mathbf{y} \\ \mathbf{y}_* \end{pmatrix}, \begin{pmatrix} \mathbf{X} \\ \mathbf{X}_* \end{pmatrix}\boldsymbol{\beta}, \begin{pmatrix} \underline{\mathbf{V}} & \underline{\mathbf{V}}_{12} \\ \underline{\mathbf{V}}_{21} & \underline{\mathbf{V}}_{22} \end{pmatrix} \right\}. \tag{10.1b}$$

Thus the difference appears only in the covariance matrices

$$\boldsymbol{\Gamma} = \begin{pmatrix} \mathbf{V} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{pmatrix}, \quad \underline{\boldsymbol{\Gamma}} = \begin{pmatrix} \underline{\mathbf{V}} & \underline{\mathbf{V}}_{12} \\ \underline{\mathbf{V}}_{21} & \underline{\mathbf{V}}_{22} \end{pmatrix}. \tag{10.2}$$

By $\mathscr{M}$ and $\underline{\mathscr{M}}$ we of course mean the models $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$ and $\underline{\mathscr{M}} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \underline{\mathbf{V}}\}$.

Suppose that $\mathbf{Fy}$ is linearly sufficient for $\mathbf{X}_*\boldsymbol{\beta}$ under $\mathscr{M}_*$, i.e., $\mathbf{Fy} \in \mathcal{S}(\mathbf{X}_*\boldsymbol{\beta} \mid \mathscr{M}_*)$. Obviously we have

$$\mathcal{S}(\mathbf{X}_*\boldsymbol{\beta} \mid \mathscr{M}_*) = \mathcal{S}(\mathbf{X}_*\boldsymbol{\beta} \mid \mathscr{M})\,. \tag{10.3}$$

We can now pose the following question: what is the condition that the same $\mathbf{Fy}$ continues to be linearly sufficient under the misspecified model $\underline{\mathscr{M}}_*$, i.e.,

$$\mathcal{S}(\mathbf{X}_*\boldsymbol{\beta} \mid \mathscr{M}) \subset \mathcal{S}(\mathbf{X}_*\boldsymbol{\beta} \mid \underline{\mathscr{M}}) \;? \tag{10.4}$$

Observe that in view of (10.3) we have above dropped off the subscript $*$ from $\mathscr{M}$ and $\underline{\mathscr{M}}$.

We use the following notations:

$$\mathcal{W} = \left\{ \mathbf{W} \in \mathbb{R}^{n\times n} : \mathbf{W} = \mathbf{V} + \mathbf{XUU}'\mathbf{X}',\ \mathscr{C}(\mathbf{W}) = \mathscr{C}(\mathbf{X} : \mathbf{V}) \right\}, \tag{10.5a}$$

$$\underline{\mathcal{W}} = \left\{ \underline{\mathbf{W}} \in \mathbb{R}^{n\times n} : \mathbf{W} = \underline{\mathbf{V}} + \mathbf{X}\underline{\mathbf{U}}\underline{\mathbf{U}}'\mathbf{X}',\ \mathscr{C}(\underline{\mathbf{W}}) = \mathscr{C}(\mathbf{X} : \underline{\mathbf{V}}) \right\}. \tag{10.5b}$$

Thus, in light of part (a_3) of Proposition 3.3, the claim (10.4) can be expressed as

$$\mathscr{C}(\mathbf{Z}) \subset \mathscr{C}(\mathbf{W}\mathbf{F}') \implies \mathscr{C}(\underline{\mathbf{Z}}) \subset \mathscr{C}(\underline{\mathbf{W}}\mathbf{F}')\,, \tag{10.6}$$

where $\mathbf{W} \in \mathcal{W}$, $\underline{\mathbf{W}} \in \underline{\mathcal{W}}$, and

$$\mathbf{Z} = \mathbf{X}(\mathbf{X}'\mathbf{W}^-\mathbf{X})^-\mathbf{X}_*'\,, \quad \underline{\mathbf{Z}} = \mathbf{X}(\mathbf{X}'\underline{\mathbf{W}}^-\mathbf{X})^-\mathbf{X}_*'\,. \tag{10.7}$$

Correspondingly, in view of part (b) of Proposition 3.4, $\mathcal{S}(\boldsymbol{\varepsilon}_* \mid \mathscr{M}) \subset \mathcal{S}(\boldsymbol{\varepsilon}_* \mid \underline{\mathscr{M}})$ can be expressed as

$$\mathscr{C}(\mathbf{M}\mathbf{V}_{12}) \subset \mathscr{C}(\mathbf{M}\mathbf{V}\mathbf{F}'\mathbf{Q}_{\mathbf{FX}}) \implies \mathscr{C}(\mathbf{M}\underline{\mathbf{V}}_{12}) \subset \mathscr{C}(\mathbf{M}\underline{\mathbf{V}}\mathbf{F}'\mathbf{Q}_{\mathbf{FX}})\,. \tag{10.8}$$

In this section we shortly present, following Markiewicz and Puntanen (2019b), solutions to implications (10.6) and (10.8). Baksalary and Mathew (1986) allowed misspecification also in the $\mathbf{X}$-part and considered the following inclusion:

$$\mathcal{S}(\mathbf{X}\boldsymbol{\beta} \mid \mathscr{M}) \subset \mathcal{S}(\underline{\mathbf{X}}\boldsymbol{\beta} \mid \underline{\underline{\mathscr{M}}})\,, \tag{10.9}$$

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

Below is a solution to (10.4), see Markiewicz and Puntanen (2019b, Th. 3.1).

Proposition 10.1. *Consider the linear models $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$ and $\underline{\mathscr{M}} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \underline{\mathbf{V}}\}$, let $\mathscr{C}(\mathbf{X}_*') \subset \mathscr{C}(\mathbf{X}')$, and denote*

$$\mathbf{Z} = \mathbf{X}(\mathbf{X}'\mathbf{W}^-\mathbf{X})^-\mathbf{X}_*'\,, \quad \underline{\mathbf{Z}} = \mathbf{X}(\mathbf{X}'\underline{\mathbf{W}}^-\mathbf{X})^-\mathbf{X}_*'\,. \tag{10.10}$$

Then the inclusion $\mathcal{S}(\mathbf{X}_*\boldsymbol{\beta} \mid \mathscr{M}) \subset \mathcal{S}(\mathbf{X}_*\boldsymbol{\beta} \mid \underline{\mathscr{M}})$ holds if and only if the following two conditions hold:

(a) $\mathscr{C}(\underline{\mathbf{Z}}) \subset \mathscr{C}(\underline{\mathbf{W}}\mathbf{W}^+\mathbf{Z})$, *i.e.,* $(\mathbf{W}^+\mathbf{Z})'\mathbf{y} \in \mathcal{S}(\mathbf{X}_*\boldsymbol{\beta} \mid \underline{\mathscr{M}})$,

(b) $\mathscr{C}(\underline{\mathbf{W}}) \subset \mathscr{C}(\mathbf{W})$.

Let us briefly sketch the main idea of the proof of Proposition 10.1. Assume now that the inclusion $\mathcal{S}(\mathbf{X}_*\boldsymbol{\beta} \mid \mathscr{M}) \subset \mathcal{S}(\mathbf{X}_*\boldsymbol{\beta} \mid \underline{\mathscr{M}})$ holds, i.e.,

$$\mathscr{C}(\mathbf{Z}) \subset \mathscr{C}(\mathbf{W}\mathbf{F}') \tag{10.11}$$

implies

$$\mathscr{C}(\underline{\mathbf{Z}}) \subset \mathscr{C}(\underline{\mathbf{W}}\mathbf{F}')\,. \tag{10.12}$$

Choosing $\mathbf{F}' = \mathbf{W}^-\mathbf{Z}$, the condition (10.11) is satisfied and thereby $(\mathbf{W}^-\mathbf{Z})'\mathbf{y}$ is linearly sufficient for $\mathbf{X}_*\boldsymbol{\beta}$ under $\mathscr{M}$ for any choice of $\mathbf{W}^-$. By assumption this same $\mathbf{F}\mathbf{y}$ (for any $\mathbf{W}^-$) now belongs to $\mathcal{S}(\mathbf{X}_*\boldsymbol{\beta} \mid \underline{\mathscr{M}})$, which means that we must have

$$\mathscr{C}(\underline{\mathbf{Z}}) \subset \mathscr{C}(\underline{\mathbf{W}}\mathbf{W}^-\mathbf{Z})\,. \tag{10.13}$$

Then the proof proceeds by utilizing some conditions under which (10.13) is holding for any choice of $\mathbf{W}^-$; for details, see Markiewicz and Puntanen (2019b, Th. 3.1).

The following proposition gives the condition for the equality in Proposition 10.1.

Proposition 10.2. *The equality* $\mathcal{S}(\mathbf{X}_*\boldsymbol{\beta} \mid \mathscr{M}) = \mathcal{S}(\mathbf{X}_*\boldsymbol{\beta} \mid \underline{\mathscr{M}})$ *holds if and only if the following two conditions hold:*

(a) $\mathscr{C}(\mathbf{W}^+\mathbf{Z}) = \mathscr{C}(\underline{\mathbf{W}}^+\underline{\mathbf{Z}})\,,$

(b) $\mathscr{C}(\underline{\mathbf{W}}) = \mathscr{C}(\mathbf{W})\,.$

Let us see how Proposition 10.1 changes when we put $\mathbf{X}_* = \mathbf{X}$. Using

$$\mathscr{C}[\mathbf{X}(\mathbf{X}'\mathbf{W}^-\mathbf{X})^-\mathbf{X}'] = \mathscr{C}[\mathbf{X}(\mathbf{X}'\underline{\mathbf{W}}^-\mathbf{X})^-\mathbf{X}'] = \mathscr{C}(\mathbf{X})\,, \tag{10.14}$$

the statement (a) in Proposition 10.1 becomes $\mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\underline{\mathbf{W}}\mathbf{W}^+\mathbf{X})$, i.e., $\mathscr{C}(\mathbf{X}) = \mathscr{C}(\underline{\mathbf{W}}\mathbf{W}^+\mathbf{X})$. This yields the following result.

Proposition 10.3. [Baksalary and Mathew (1986, Th. 1)] *The inclusion*

$$\mathcal{S}(\mathbf{X}\boldsymbol{\beta} \mid \mathscr{M}) \subset \mathcal{S}(\mathbf{X}\boldsymbol{\beta} \mid \underline{\mathscr{M}}) \tag{10.15}$$

holds if and only if the following two conditions hold:

(a) $\mathscr{C}(\mathbf{X}) = \mathscr{C}(\underline{\mathbf{W}}\mathbf{W}^+\mathbf{X})$, *i.e.,* $(\mathbf{W}^+\mathbf{X})'\mathbf{y} \in \mathcal{S}(\mathbf{X}\boldsymbol{\beta} \mid \underline{\mathscr{M}})\,,$

(b) $\mathscr{C}(\underline{\mathbf{W}}) \subset \mathscr{C}(\mathbf{W})\,.$

Baksalary and Mathew (1986, Th. 1) give Proposition 10.3 in the situation when also the **X**-parts can be different; see (10.9).

Consider then a partitioned linear model $\mathscr{M}_{12}$. We know that $\mathbf{Fy}$ is linearly sufficient for $\boldsymbol{\mu}_1 = \mathbf{X}_1\boldsymbol{\beta}_1$ if and only if

$$\mathscr{C}[\mathbf{X}(\mathbf{X}'\mathbf{W}^-\mathbf{X})^-\mathbf{X}_*'] \subset \mathscr{C}(\mathbf{W}\mathbf{F}')\,, \tag{10.16}$$

where $\mathbf{X}_* = (\mathbf{X}_1 : \mathbf{0})$. In this situation, in view of (7.15), we can express the matrix $\mathbf{Z}$ as follows:

$$\mathbf{Z} = \mathbf{X}(\mathbf{X}'\mathbf{W}^-\mathbf{X})^-\mathbf{X}_*' = \mathbf{W}\dot{\mathbf{M}}_{2W}\mathbf{X}_1(\mathbf{X}_1'\dot{\mathbf{M}}_{2W}\mathbf{X}_1)^-\mathbf{X}_1'\,, \tag{10.17}$$

where $\dot{\mathbf{M}}_{2W} = \mathbf{M}_2(\mathbf{M}_2\mathbf{W}\mathbf{M}_2)^-\mathbf{M}_2$ and column space of $\mathbf{Z}$ is

$$\mathscr{C}(\mathbf{Z}) = \mathscr{C}(\mathbf{W}\dot{\mathbf{M}}_{2W}\mathbf{X}_1)\,. \tag{10.18}$$

Using (10.18) and Proposition 10.1, Markiewicz and Puntanen (2019b, Sec. 4) showed the following.

Proposition 10.4. *Let* $\boldsymbol{\mu}_1 = \mathbf{X}_1\boldsymbol{\beta}_1$ *be estimable under* $\mathscr{M}_{12}$ *and let* $\mathbf{W} \in \mathcal{W}$ *and* $\underline{\mathbf{W}} \in \underline{\mathcal{W}}$. *Then the inclusion* $\mathcal{S}(\mathbf{X}_1\boldsymbol{\beta}_1 \mid \mathscr{M}_{12}) \subset \mathcal{S}(\mathbf{X}_1\boldsymbol{\beta}_1 \mid \underline{\mathscr{M}}_{12})$ *holds if and only if the following two conditions hold:*

(a) $\mathscr{C}(\mathbf{M}_2\mathbf{X}_1) \subset \mathscr{C}[\mathbf{M}_2\underline{\mathbf{W}}\mathbf{M}_2(\mathbf{M}_2\mathbf{W}\mathbf{M}_2)^-\mathbf{M}_2\mathbf{X}_1]\,,$

(b) $\mathscr{C}(\underline{\mathbf{W}}) \subset \mathscr{C}(\mathbf{W})\,.$

Consider then the misspecification and the linear sufficiency with respect to the error term, that is, when is the following holding: $\mathcal{S}(\boldsymbol{\varepsilon}_* \mid \mathscr{M}_*) \subset \mathcal{S}(\boldsymbol{\varepsilon}_* \mid \underline{\mathscr{M}}_*)$, in other words,

$$\mathscr{C}(\mathbf{MV}_{12}) \subset \mathscr{C}(\mathbf{MVF}'\mathbf{Q}_{\mathbf{FX}}) \implies \mathscr{C}(\mathbf{M}\underline{\mathbf{V}}_{12}) \subset \mathscr{C}(\mathbf{M}\underline{\mathbf{V}}\mathbf{F}'\mathbf{Q}_{\mathbf{FX}}) . \tag{10.19}$$

The following proposition gives the result.

Proposition 10.5. *Consider the linear models (with new observations) $\mathscr{M}_*$ and $\underline{\mathscr{M}}_*$. Then the inclusion $\mathcal{S}(\boldsymbol{\varepsilon}_* \mid \mathscr{M}_*) \subset \mathcal{S}(\boldsymbol{\varepsilon}_* \mid \underline{\mathscr{M}}_*)$ holds if and only if the following two conditions hold:*

(a) $\mathscr{C}(\mathbf{M}\underline{\mathbf{V}}_{12}) \subset \mathscr{C}[\mathbf{M}\underline{\mathbf{V}}\mathbf{M}(\mathbf{MVM})^+\mathbf{MV}_{12}]$,

(b) $\mathscr{C}(\underline{\mathbf{W}}) \subset \mathscr{C}(\mathbf{W})$,

where (a) *can be equivalently expressed in the following two forms:*

(c) $\mathbf{V}_{21}\mathbf{M}(\mathbf{MVM})^+\mathbf{My} \in \mathcal{S}(\boldsymbol{\varepsilon}_* \mid \underline{\mathscr{M}}_*)$, *i.e.,* $\mathrm{BLUP}(\boldsymbol{\varepsilon}_* \mid \mathscr{M}_*) \in \mathcal{S}(\boldsymbol{\varepsilon}_* \mid \underline{\mathscr{M}}_*)$,

(d) $\mathscr{C}(\underline{\mathbf{V}}_{12}) \subset \mathscr{C}[\mathbf{X} : \underline{\mathbf{V}}\mathbf{M}(\mathbf{MVM})^+\mathbf{MV}_{12}]$.

Using Proposition 8.3 we can write the following result.

Proposition 10.6. *Consider the linear models (with new observations) $\mathscr{M}_*$ and $\underline{\mathscr{M}}_*$ and the following statements:*

(a) $\mathcal{S}(\mathbf{X}_*\boldsymbol{\beta} \mid \mathscr{M}_*) \subset \mathcal{S}(\mathbf{X}_*\boldsymbol{\beta} \mid \underline{\mathscr{M}}_*)$,

(b) $\mathcal{S}(\boldsymbol{\varepsilon}_* \mid \mathscr{M}_*) \subset \mathcal{S}(\boldsymbol{\varepsilon}_* \mid \underline{\mathscr{M}}_*)$,

(c) $\mathcal{S}(\mathbf{y}_* \mid \mathscr{M}_*) \subset \mathcal{S}(\mathbf{y}_* \mid \underline{\mathscr{M}}_*)$.

Then each of the two statements above imply the third one. In particular (a) *and* (b) *imply* (c), *i.e.,* (c) *holds if*

(i) $\mathscr{C}(\mathbf{M}\underline{\mathbf{V}}_{12}) \subset \mathscr{C}[\mathbf{M}\underline{\mathbf{V}}\mathbf{M}(\mathbf{MVM})^+\mathbf{MV}_{12}]$,

(ii) $\mathscr{C}(\underline{\mathbf{Z}}) \subset \mathscr{C}(\underline{\mathbf{W}}\mathbf{W}^+\mathbf{Z})$,

(iii) $\mathscr{C}(\underline{\mathbf{W}}) \subset \mathscr{C}(\mathbf{W})$.

It would be interesting to find a necessary and sufficient condition for (c) in Proposition 10.6. This does not seem to be easy to do, so this question remains a topic for future research.

We complete this section by briefly commenting the models

$$\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}, \quad \text{and} \quad \underline{\underline{\mathscr{M}}} = \{\mathbf{y}, \underline{\mathbf{X}}\boldsymbol{\beta}, \underline{\mathbf{V}}\}. \tag{10.20}$$

Thus the difference appears not only in the covariance matrices but also in the $\mathbf{X}$-part. We can now characterize the condition under which

$$\mathcal{S}(\mathbf{X}_*\boldsymbol{\beta} \mid \mathscr{M}) \subset \mathcal{S}(\underline{\mathbf{X}}_*\boldsymbol{\beta} \mid \underline{\underline{\mathscr{M}}}). \tag{10.21}$$

We assume that $\mathbf{X}_*\boldsymbol{\beta}$ is estimable under $\mathscr{M}$ and $\underline{\mathbf{X}}_*\boldsymbol{\beta}$ is estimable under $\underline{\underline{\mathscr{M}}}$, i.e.,

$$\mathscr{C}(\mathbf{X}_*') \subset \mathscr{C}(\mathbf{X}') \quad \text{and} \quad \mathscr{C}(\underline{\mathbf{X}}_*') \subset \mathscr{C}(\underline{\mathbf{X}}') \,. \tag{10.22}$$

Using the notation

$$\underline{\underline{\mathcal{W}}} = \big\{ \underline{\underline{\mathbf{W}}} : \underline{\underline{\mathbf{W}}} = \underline{\mathbf{V}} + \underline{\mathbf{X}}\underline{\mathbf{U}}\underline{\mathbf{U}}'\underline{\mathbf{X}}',\ \mathscr{C}(\underline{\underline{\mathbf{W}}}) = \mathscr{C}(\underline{\mathbf{X}} : \underline{\mathbf{V}}) \big\} \,, \tag{10.23}$$

the claim (10.21) above can be expressed as

$$\mathscr{C}(\mathbf{Z}) \subset \mathscr{C}(\mathbf{W}\mathbf{F}') \implies \mathscr{C}(\underline{\underline{\mathbf{Z}}}) \subset \mathscr{C}(\underline{\underline{\mathbf{W}}}\mathbf{F}') \,, \tag{10.24}$$

where $\mathbf{W} \in \mathcal{W}$, $\underline{\underline{\mathbf{W}}} \in \underline{\underline{\mathcal{W}}}$, and

$$\mathbf{Z} = \mathbf{X}(\mathbf{X}'\mathbf{W}^-\mathbf{X})^-\mathbf{X}_*' \,, \quad \underline{\underline{\mathbf{Z}}} = \underline{\mathbf{X}}(\underline{\mathbf{X}}'\underline{\underline{\mathbf{W}}}^-\underline{\mathbf{X}})^-\underline{\mathbf{X}}_*' \,. \tag{10.25}$$

The proof of the following result is parallel to that of Proposition 10.1.

Proposition 10.7. *Consider the models* $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$ *and* $\underline{\underline{\mathscr{M}}} = \{\mathbf{y}, \underline{\mathbf{X}}\boldsymbol{\beta}, \underline{\mathbf{V}}\}$. *Assume that* (10.22) *holds and denote*

$$\mathbf{Z} = \mathbf{X}(\mathbf{X}'\mathbf{W}^-\mathbf{X})^-\mathbf{X}_*' \,, \quad \underline{\underline{\mathbf{Z}}} = \underline{\mathbf{X}}(\underline{\mathbf{X}}'\underline{\underline{\mathbf{W}}}^-\underline{\mathbf{X}})^-\underline{\mathbf{X}}_*' \,, \tag{10.26}$$

where $\mathbf{W} \in \mathcal{W}$, $\underline{\underline{\mathbf{W}}} \in \underline{\underline{\mathcal{W}}}$. *Then the inclusion*

$$\mathcal{S}(\mathbf{X}_*\boldsymbol{\beta} \mid \mathscr{M}) \subset \mathcal{S}(\underline{\mathbf{X}}_*\boldsymbol{\beta} \mid \underline{\underline{\mathscr{M}}}) \tag{10.27}$$

holds if and only if the following two conditions hold:

(a) $\mathscr{C}(\underline{\underline{\mathbf{Z}}}) \subset \mathscr{C}(\underline{\underline{\mathbf{W}}}\mathbf{W}^+\mathbf{Z})$, *i.e.,* $(\mathbf{W}^+\mathbf{Z})'\mathbf{y} \in \mathcal{S}(\underline{\mathbf{X}}_*\boldsymbol{\beta} \mid \underline{\underline{\mathscr{M}}})$,

(b) $\mathscr{C}(\underline{\underline{\mathbf{W}}}) \subset \mathscr{C}(\mathbf{W})$,

where the part (a) *can be replaced with any of the following equivalent conditions:*

(c) $\mathscr{C}(\underline{\underline{\mathbf{W}}}^+\underline{\underline{\mathbf{Z}}}) \subset \mathscr{C}(\mathbf{P}_{\underline{\underline{\mathbf{W}}}}\mathbf{W}^+\mathbf{Z})$,

(d) $\mathscr{C}(\mathbf{Q}_{\underline{\underline{\mathbf{W}}}\mathbf{Q}_{\underline{\underline{\mathbf{Z}}}}}\underline{\underline{\mathbf{Z}}}) \subset \mathscr{C}(\mathbf{P}_{\underline{\underline{\mathbf{W}}}}\mathbf{Q}_{\mathbf{W}\mathbf{Q}_{\mathbf{Z}}}\mathbf{Z})$.

11 Conclusions

The idea of transforming $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$ by a matrix $\mathbf{F}$ of order $f \times n$ follows from a desire of reduction of the initial information delivered by an observed value of a random vector variable $\mathbf{y}$ in such a way that it is still possible to obtain the BLUE of $\mathbf{X}\boldsymbol{\beta}$ from the transformed model $\mathscr{T} = \{\mathbf{F}\mathbf{y}, \mathbf{F}\mathbf{X}\boldsymbol{\beta}, \mathbf{F}\mathbf{V}\mathbf{F}'\}$. Hence the concept of the linear sufficiency has an essential role when studying the connection between $\mathscr{M}$ and its transformed version $\mathscr{T}$ and thereby in our paper a lot of attention has been paid to the properties of the transformed model. The same concerns the prediction of the new unknown $\mathbf{y}_*$ on the basis of observable $\mathbf{y}$: we say that $\mathbf{Fy}$ is linearly (prediction) sufficient for $\mathbf{y}_*$ if the BLUP of $\mathbf{y}_*$ is obtainable by $\mathbf{AFy}$ for some matrix $\mathbf{A}$.

So the key point is that we do not lose anything essential when doing estimation or prediction if instead of $\mathscr{M}$ we use $\mathscr{T}$. In these considerations we believe that the new unobservable random vector $\mathbf{y}_*$ is coming from $\mathbf{y}_* = \mathbf{X}_*\boldsymbol{\beta} + \boldsymbol{\varepsilon}_*$, where the expectation of $\mathbf{y}_*$ is $\mathbf{X}_*\boldsymbol{\beta}$ and the covariance matrix of $\mathbf{y}_*$ is known as well as the cross-covariance matrix between $\mathbf{y}_*$ and $\mathbf{y}$. We denote the supplemented setup shortly as

$$\mathscr{M}_* = \left\{ \begin{pmatrix} \mathbf{y} \\ \mathbf{y}_* \end{pmatrix}, \begin{pmatrix} \mathbf{X} \\ \mathbf{X}_* \end{pmatrix} \boldsymbol{\beta}, \begin{pmatrix} \mathbf{V} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{pmatrix} \right\}. \tag{11.1}$$

In this paper, we have considered the linear sufficiency of $\mathbf{Fy}$ with respect to $\mathbf{y}_*$, $\mathbf{X}_*\boldsymbol{\beta}$, and $\boldsymbol{\varepsilon}_*$. We have also applied our results into the linear mixed model which can be seen as special case of the model with new observations. We omit some related concepts like linear completeness and discuss only briefly the minimal linear sufficiency. Regarding Section 10, we may mention that the issue of misspecified covariances in linear models also applies when covariances must be estimated (as is nearly always true in practice). This happens very frequently, indeed almost any time a linear mixed model is fitted. A recent review article by Haslett and Welsh (2019) attempts to disentangle the literature on the effect of estimating covariances.

Dong et al. (2014), and Tian (2013, 2017) study interesting connections between the BLUEs under two transformed models. Yongge Tian and his collaborators have presented a comprehensive investigation to the linear transformations using so called matrix-rank method. This approach provides effective matrix algebraic tools and the results often appear in somewhat complicated forms. Of course, the questions that the matrix-rank group presents are often more complicated than those that we have been posing.

The concept of linear sufficiency was essentially introduced in early 1980s by Baksalary, Kala and Drygas. Drygas in (1983, p. 97) pointed out that "The concept of a linearly sufficient statistic is rather unknown in statistical literature. Besides the paper by Baksalary and Kala (1978b), who prove [Drygas' Theorem 3.2] without using the concept of sufficiency, there is only the paper by Barnard (1963, p. 232)." Recently several papers providing further properties of the linear sufficiency have been published and a bunch is by the present authors. Some of these papers have been under the auspices of recent meetings of an International Research Group on Multivariate and Mixed Linear Models in Będlewo, Poland. Our aim in this article is to provide an easy-to-read, i.e., a self-readable review of recent results that "ordinary mortals can appreciate", cf. Lindley (2003, p. 232), and while doing that, to go through some basic concepts related to linear sufficiency.

Acknowledgements

Thanks for helpful discussions go to Barbora Arendacká, Tadeusz Caliński, Xu-Qing Liu, and Yongge Tian. We express special thanks to anonymous referees for constructive comments. Part of this research has been done during the biannual meetings of an International Research Group on Multivariate and Mixed Linear Models in the Mathematical Research and Conference Center, Będlewo, Poland, from November 2013 to November 2018, supported by the Stefan Banach International Mathematical Center.

References

Aitken, A.C. (1935). On least squares and linear combination of observations. *Proc. Roy. Soc. Edinburgh Sect. A*, 55, 42–49. DOI

Arendacká, B., Puntanen, S. (2015). Further remarks on the connection between fixed linear model and mixed linear model. *Stat. Pap.*, 56, 1235–1247. DOI.

Baksalary, J.K. (1984). A study of the equivalence between a Gauss–Markoff model and its augmentation by nuisance parameters. *Math. Operationsforsch. Stat. Ser. Stat.*, 15, 2–35. DOI

Baksalary, J.K. (1987). Algebraic characterizations and statistical implications of the commutativity of orthogonal projectors. *Proceedings of the Second International Tampere Conference in Statistics.* (T. Pukkila, S. Puntanen, eds.) Dept. of Math. Sciences/Statistics, Univ. of Tampere, vol. A 184, pp. 113–142.

Baksalary, J.K., Drygas, H. (1992). A note on the concepts of sufficiency in the general Gauss–Markov model: a coordinate-free approach. Forschungsbericht 92/2, Universität Dortmund, Fachbereich Statistik.

Baksalary, J.K., Kala, R. (1978a). A bound for the Euclidean norm of the difference between the least squares and the best linear unbiased estimators. *Ann. Stat.*, 6, 1390–1393. DOI

Baksalary, J.K., Kala, R. (1978b). Linear transformations preserving the best linear unbiased estimators in a general linear model. Paper presented at the 6th International Conference on Mathematical Statistics, Wisła, Poland, December 1978.

Baksalary, J.K., Kala, R. (1980). A new bound for the Euclidean norm of the difference between the least squares and the best linear unbiased estimators. *Ann. Stat.*, 8, 679–681. DOI

Baksalary, J.K., Kala, R. (1981). Linear transformations preserving best linear unbiased estimators in a general Gauss–Markoff model. *Ann. Stat.*, 9, 913–916. DOI

Baksalary, J.K., Kala, R. (1986). Linear sufficiency with respect to a given vector of parametric functions. *J. Stat. Plan. Inference*, 14, 331–338. DOI

Baksalary, J.K., Mathew, T. (1986). Linear sufficiency and completeness in an incorrectly specified general Gauss–Markov model. *Sankhyā Ser. A*, 48, 169–180. DOI

Baksalary, J.K., Mathew, T. (1990). Rank invariance criterion and its application to the unified theory of least squares. *Linear Algebra Appl.*, 127, 393–401. DOI

Baksalary, J.K., Puntanen, S., Styan, G.P.H. (1990). A property of the dispersion matrix of the best linear unbiased estimator in the general Gauss–Markov model. *Sankhyā Ser. A*, 52, 279–296. DOI

Baksalary, J.K., Rao, C.R., Markiewicz, A. (1992). A study of the influence of the "natural restrictions" on estimation problems in the singular Gauss–Markov model. *J. Stat. Plan. Inference*, 31, 335–351. DOI

Baksalary, O.M., Trenkler, G., Liski, E. (2013). Let us do the twist again. *Stat. Pap.*, 54, 1109–1119. DOI

Ben-Israel, A., Greville, T.N.E. (2003). *Generalized Inverses: Theory and Applications*, 2nd Edition. Springer, New York. DOI

Barnard, G.A. (1963). The logic of least squares. *J. Roy. Statist. Soc. Ser. B*, 25, 12–127. DOI

Bhimasankaram, P., Sengupta, D. (1996). The linear zero functions approach to linear models. *Sankhyā Ser. B*, 58, 338–351. DOI

Bloomfield, P., Watson, G.S. (1975). The inefficiency of least squares. *Biometrika*, 62, 121–128. DOI

Chu, K.L, Isotalo, J., Puntanen, S., Styan, G.P.H. (2004). On decomposing the Watson efficiency of ordinary least squares in a partitioned weakly singular linear model. *Sankhyā*, 66, 634–651. DOI

Chu, K.L, Isotalo, J., Puntanen, S., Styan, G.P.H. (2005). Some further results concerning the decomposition of the Watson efficiency in partitioned linear models. *Sankhyā*, 67, 74–89. DOI

Christensen, R. (2011). *Plane Answers to Complex Questions: the Theory of Linear Models*, 4th Edition. Springer, New York. DOI

Davidson, R., MacKinnon, J.G. (2004). *Econometric Theory and Methods.* Oxford University Press, New York. www

Dong, B., Guo, W., Tian, Y. (2014). On relations between BLUEs under two transformed linear models. *J. Multivariate Anal.*, 13, 279–292. DOI

Drygas, H. (1970). *The Coordinate-Free Approach to Gauss–Markov Estimation.* Springer, Berlin. DOI

Drygas, H. (1983). Sufficiency and completeness in the general Gauss–Markov model. *Sankhyā Ser. A*, 45, 88–98. DOI

Frisch, R., Waugh, F.V. (1933). Partial time regressions as compared with individual trends. *Econometrica*, 1, 387–401. DOI

Goldberger, A.S. (1962). Best linear unbiased prediction in the generalized linear regression model. *J. Am. Stat. Assoc.*, 58, 369–375. DOI

Gourieroux, C., Monfort, A. (1980). Sufficient linear structures: econometric applications. *Econometrica*, 48, 1083–1097. DOI

Groß, J. (1998). A note on the concepts of linear and quadratic sufficiency. *J. Stat. Plan. Inference*, 70, 88–98. DOI

Groß, J. (2004). The general Gauss–Markov model with possibly singular dispersion matrix. *Stat. Pap.*, 45, 311–336. DOI

Groß, J., Puntanen, S. (2000). Estimation under a general partitioned linear model. *Linear Algebra Appl.*, 321, 131–144. DOI

Groß, J., Puntanen, S. (2005). Extensions of the Frisch–Waugh–Lovell Theorem. *Discuss. Math. Probab. Stat.*, 25, 39–49. Online.

Groß, J., Trenkler, G. (1998). On the equality of linear statistics in general Gauss–Markov Model. *Frontiers in Probability and Statistics.* (S.P. Mukherjee, S.K. Basu, B.K. Sinha, eds.) Narosa Publishing House, New Delhi, pp. 189–194.

Güler, N., Büyükkaya, E. (2019). Rank and inertia formulas for covariance matrices of BLUPs in general linear mixed models. *Commun. Stat. Theory Methods*, available online. DOI

Harville, D.A. (1997). *Matrix Algebra From a Statistician's Perspective.* Springer, New York. DOI

Haslett, S.J., Isotalo, J., Liu, Y., Puntanen, S. (2014). Equalities between OLSE, BLUE and BLUP in the linear model. *Stat. Pap.*, 55, 543–561. DOI

Haslett, S.J., Liu, X.-Q., Markiewicz, A. Puntanen, S. (2020). Some properties of linear sufficiency and the BLUPs in the linear mixed model. *Stat. Pap.*, 61, 385–401. DOI

Haslett, S.J., Puntanen, S. (2017). Best linear unbiased prediction (BLUP). *Wiley StatsRef: Statistics Reference Online.* stat08120. (M. Davidian, R. Kenett, N.T. Longford, G. Molenberghs, W.W. Piegorsch, F. Ruggeri, eds.) Wiley, Chichester. 6 pp. DOI

Haslett, S.J., Puntanen, S., Arendacká, B. (2015). The link between the mixed and fixed linear models revisited. *Stat. Pap.*, 56, 849–861. DOI

Haslett, S.J., Welsh, A.H. (2019). Empirical best linear unbiased prediction (EBLUP). *Wiley StatsRef: Statistics Reference Online.* stat08180. (M. Davidian, R. Kenett, N.T. Longford, G. Molenberghs, W.W. Piegorsch, F. Ruggeri, eds.) Wiley, Chichester. 10 pp. DOI

Hauke, J., Markiewicz, A., Puntanen, S. (2012). Comparing the BLUEs under two linear models. *Commun. Stat. Theory Methods*, 41, 2405–2418. DOI

Henderson, C.R., Kempthorne, O., Searle, S.R., von Krosigh, C.N. (1959). The estimation of environmental and genetic trends from records subject to culling. *Biometrics*, 15, 192–218. DOI

Isotalo, J., Markiewicz, A., Puntanen, S. (2018). Some properties of linear prediction sufficiency in the linear model. *Trends and Perspectives in Linear Statistical Inference: LinStat, Istanbul, 2016.* (M. Tez. D. von Rosen, eds.) Springer, pp. 111–129. DOI

Isotalo, J., Puntanen, S. (2006a). Linear sufficiency and completeness in the partitioned linear model. *Acta Comment. Univ. Tartu. Math.*, 10, 53–67. www

Isotalo, J., Puntanen, S. (2006b). Linear prediction sufficiency for new observations in the general Gauss–Markov model. *Commun. Stat. Theory Methods*, 35, 1011–1023. DOI

Isotalo, J., Puntanen, S. (2009). Linear sufficiency and completeness in the context of estimating the parametric function in the general Gauss–Markov model. *J. Stat. Plan. Inference*, 139, 722–733. DOI.

Isotalo, J., Puntanen, S., Styan, G.P.H. (2008a). A useful matrix decomposition and its statistical applications in linear regression. *Commun. Stat. Theory Methods*, 37, 1436–1457. DOI

Isotalo, J., Puntanen, S., Styan, G.P.H. (2008b). The BLUE's covariance matrix revisited: a review. *J. Stat. Plan. Inference*, 138, 2722–2737. DOI

Kala, R. (1981). Projectors and linear estimation in general linear models. *Commun. Stat. Theory Methods*, 10, 849–873. DOI

Kala, R., Puntanen, S., Tian, Y. (2017a). Some notes on linear sufficiency. *Stat. Pap.*, 58, 1–17. DOI

Kala, R., Markiewicz, A., Puntanen, S. (2017b). Some further remarks on the linear sufficiency in the linear model. *Applied and Computational Matrix Analysis: Mat-*

Triad, Coimbra, Portugal, September 2015, Selected, Revised Contributions. (N. Bebiano, ed.) Springer Proceedings in Mathematics and Statistics, vol. 192, pp. 275–294. DOI

Kala, R., Pordzik, P.R. (2009). Estimation in singular partitioned, reduced or transformed linear models. *Stat. Pap.*, 50, 633–638. DOI

Lindley, D.V. (2003). Professor George A. Barnard, 1915–2002, Obituary. *The Statistician*, 52, 231–234. DOI

Lovell, M.C. (1963). Seasonal adjustment of economic time series and multiple regression analysis. *J. Am. Stat. Assoc.*, 58, 993–1010. DOI

Lovell, M.C. (2008). A simple proof of the FWL Theorem. *J. Econ. Educ.*, 39, 88–91. DOI

Kornacki, A. (2007). Different kinds of sufficiency in the general Gauss–Markov model. *Math. Slovaca*, 57, 389–392. DOI

Liu, X.-Q., Rong, J.-Y., Liu, J.-Y. (2008). Best linear unbiased prediction for linear combinations in general mixed linear models. *J. Multivar. Anal.*, 99, 1503–1517. DOI

Markiewicz, A. (1998). Comparison of linear restricted models with respect to the validity of admissible and linearly sufficient estimators. *Statist. Probab. Lett.*, 38, 347–354. DOI

Markiewicz, A., Puntanen, S. (2009). Admissibility and linear sufficiency in linear model with nuisance parameters. *Stat. Pap.*, 50, 847–854. DOI.

Markiewicz, A., Puntanen, S. (2015). All about the $\perp$ with its applications in the linear statistical models. *Open Mathematics*, 13, 33–50. DOI

Markiewicz, A., Puntanen, S. (2018a). Further properties of linear prediction sufficiency and the BLUPs in the linear model with new observations. *Afrika Statistika*, 13, 1511–1530. DOI

Markiewicz, A., Puntanen, S. (2018b). Upper bounds for the Euclidean distances between the BLUPs. *Special Matrices*, 6, 249–261. DOI

Markiewicz, A., Puntanen, S. (2019a). Further properties of the linear sufficiency in the partitioned linear model. *Matrices, Statistics and Big Data: Selected Contributions from IWMS 2016.* (S.E. Ahmed, F. Carvalho and S. Puntanen, eds.) Springer, pp. 1–22. DOI

Markiewicz, A., Puntanen, S. (2019b). Linear prediction sufficiency in the misspecified linear model. *Commun. Stat. Theory Methods*, available online. DOI

Marsaglia, G., Styan, G.P.H. (1974). Equalities and inequalities for ranks of matrices. *Linear Multilinear Algebra*, 2, 269–292. DOI

McCulloch, C.E., Searle, S.R., Neuhaus, J.M. (2008). *Generalized, Linear, and Mixed Models*, 2nd Edition. Wiley, New York.

Müller, J. (1987). Sufficiency and completeness in the linear model. *J. Multivar. Anal.*, 21, 312–323. DOI

Müller, J., Rao, C.R., Sinha, B.K. (1984). Inference on parameters in a linear model: a review of recent results. *Experimental design, statistical models, and genetic statistics* (K. Hinkelmann, ed.). Dekker, New York, pp. 277–295.

Pordzik, P.R. (2012). A bound for the Euclidean distance between restricted and unrestricted estimators of parametric functions in the general linear model. *Stat. Pap.*, 53, 299–304. DOI

Puntanen, S., Styan, G.P.H. (1989). The equality of the ordinary least squares estimator and the best linear unbiased estimator [with comments by O. Kempthorne and by S.R. Searle and with "Reply" by the authors]. *Amer. Statist.*, 43, 153–164. DOI

Puntanen, S., Styan, G.P.H., Isotalo, J. (2011). *Matrix Tricks for Linear Statistical Models: Our Personal Top Twenty.* Springer, Heidelberg. DOI

Rao, C.R. (1967). Least squares theory using an estimated dispersion matrix and its application to measurement of signals. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability: Berkeley, California, 1965/1966*, vol. 1. (L.M. Le Cam, J. Neyman, eds.) Univ. of California Press, Berkeley, pp. 355–372. DOI

Rao, C.R. (1973). Representations of best linear estimators in the Gauss–Markoff model with a singular dispersion matrix. *J. Multivariate Anal.*, 3, 276–292. DOI

Rao, C.R. (1974). Projectors, generalized inverses and the BLUE's, *J. Roy. Statist. Soc. Ser. B*, 36, 442–448. DOI

Rao, C.R., Mitra, S.K. (1971). *Generalized Inverse of Matrices and Its Applications.* Wiley, New York.

Robinson, G.K. (1991). That BLUP is a good thing: the estimation of random effects (discussion: pp. 32–51). *Stat. Sci.*, 6, 15–51. DOI

Rong, J.-Y., Liu, X.-Q. (2010). On misspecification of the dispersion matrix in mixed linear models. *Stat. Pap.*, 51, 445–453. DOI

Searle, S.R. (1997). The matrix handling of BLUE and BLUP in the mixed linear model. *Linear Algebra Appl.*, 264, 291–311. DOI

Sengupta, D., Jammalamadaka, S.R. (2003). *Linear Models: An Integrated Approach.* World Scientific, River Edge.

Tian, Y. (2013). On properties of BLUEs under general linear regression models. *J. Statist. Plann. Inference*, 143, 771–782. DOI

Tian, Y. (2015a). A matrix handling of predictions of new observations under a general random-effects model. *Electron. J. Linear Algebra*, 29, 30–45. DOI

Tian, Y. (2015b). A new derivation of BLUPs under random-effects model. *Metrika*, 78, 905–918. DOI

Tian, Y. (2017). Transformation approaches of linear random-effects models. *Stat. Methods Appl.*, 26, 583–608. DOI

Tian, Y., Puntanen, S. (2009). On the equivalence of estimations under a general linear model and its transformed models. *Linear Algebra Appl.*, 430, 2622–2641. DOI

Watson, G.S. (1955). Serial correlation in regression analysis, I. *Biometrika*, 42, 327–341. DOI

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

Zyskind, G., Martin, F.B. (1969). On best linear estimation and general Gauss–Markov theorem in linear models with arbitrary nonnegative covariance structure. *SIAM J. Appl. Math.*, 17, 1190–1202. DOI