Afrika Statistika
Vol. 13 (1), 2018, pages 1511–1530.
DOI: http://dx.doi.org/10.16929/as/1511.117

Afrika Statistika

ISSN 2316-090X

Further properties of linear prediction sufficiency and the BLUPs in the linear model with new observations

Augustyn Markiewicz[1] **and Simo Puntanen**[2,*]

[1]Department of Mathematical and Statistical Methods, Poznań University of Life Sciences, Wojska Polskiego 28, PL-60637 Poznań, Poland
[2]Faculty of Natural Sciences, FI-33014 University of Tampere, Tampere, Finland

Selected Papers presented at the East African Conference of Statistical Mathematics and Applications, EACSMA 2017

Received August 7, 2017; Accepted February 05, 2018

Abstract. A linear statistic $\mathbf{Fy}$ is called linearly sufficient for the estimable parametric function of $\mathbf{X}_*\boldsymbol{\beta}$ under the linear model $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$ 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 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 $\boldsymbol{\varepsilon}_*$, when the random vector $\mathbf{y}_*$ comes from $\mathbf{y}_* = \mathbf{X}_*\boldsymbol{\beta} + \boldsymbol{\varepsilon}_*$, and the prediction is based on the linear model $\mathscr{M}$. Our main results concern the mutual relations of these sufficiencies. In addition, we give an extensive review of some interesting properties of the covariance matrices of the BLUPs of $\boldsymbol{\varepsilon}_*$. We also apply our results into the linear mixed model.

Key words: Best linear unbiased estimator; Best linear unbiased predictor; Linear sufficiency; Linear mixed model; Transformed linear model.
AMS 2010 Mathematics Subject Classification : 62J05; 62J10.

---

*Corresponding author Simo Puntanen: simo.puntanen@uta.fi
Augustyn Markiewicz: amark@up.poznan.pl

Résumé. Une statistique linéaire $\mathbf{Fy}$ est dite linéairement suffisante pour la fonction paramétrique estimable de $\mathbf{X}_*\boldsymbol{\beta}$ sous le modèle linéaire $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$ s'il existe une matrice $\mathbf{A}$ telle que $\mathbf{AFy}$ soit le meilleur estimateur linéaire sans biais, BLUE, pour$\mathbf{X}_*\boldsymbol{\beta}$. Le concept de suffisance linéaire par rapport à un vecteur aléatoire prévisible est défini de manière similaire mais en considérant le meilleur prédicteur linéaire sans biais, BLUP, au lieu du BLUE. Dans cet article, nous considérons la suffisance linéaire de $\mathbf{Fy}$ par rapport à $\mathbf{y}_*$, $\mathbf{X}_*\boldsymbol{\beta}$ et $\boldsymbol{\varepsilon}_*$, lorsque le vecteur aléatoire $\mathbf{y}_*$ provient de $\mathbf{y}_* = \mathbf{X}_*\boldsymbol{\beta}+\boldsymbol{\varepsilon}_*$, et la prédiction est basée sur le modèle linéaire $\mathscr{M}$. Nos principaux résultats concernent les relations mutuelles de ces suffisances. En outre, nous donnons un examen approfondi de certaines propriétés intéressantes des matrices de covariance des BLUP des $\boldsymbol{\varepsilon}_*$. Nous appliquons également nos résultats dans le modèle mixte linéaire.

1. Introduction

In this section we introduce some preliminary concepts and results that are needed in our main considerations. First some words about the notation. The symbol $\mathbb{R}^{m\times n}$ denotes the set of $m \times n$ real matrices, while $\mathbf{A}'$, $\mathbf{A}^-$, $\mathbf{A}^+$, $\mathscr{C}(\mathbf{A})$, and $\mathscr{C}(\mathbf{A})^\perp$, denote, respectively, the transpose, a generalized inverse, the Moore–Penrose inverse, the column space, and the orthogonal complement of the column space of the matrix $\mathbf{A}$. 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 partitioned matrix with $\mathbf{A}_{a\times b}$ and $\mathbf{B}_{c\times d}$ as submatrices, where $a = c$. By $\mathbf{A}^\perp$ we denote any matrix satisfying $\mathscr{C}(\mathbf{A}^\perp) = \mathscr{C}(\mathbf{A})^\perp$. Furthermore, we will write $\mathbf{P}_\mathbf{A} = \mathbf{AA}^+ = \mathbf{A}(\mathbf{A}'\mathbf{A})^-\mathbf{A}'$ to denote the orthogonal projector (with respect to the standard inner product) onto $\mathscr{C}(\mathbf{A})$. The orthogonal projector onto $\mathscr{C}(\mathbf{A})^\perp$ is denoted as $\mathbf{Q}_\mathbf{A} = \mathbf{I}_a - \mathbf{P}_\mathbf{A}$, where $\mathbf{I}_a$ refers to the $a \times a$ identity matrix and a is the number of rows of $\mathbf{A}$. In particular, we use notation $\mathbf{M} = \mathbf{I}_n - \mathbf{P}_\mathbf{X}$, where $\mathbf{X}_{n\times p}$ refers to the model matrix; see (1.2) below. One choice for $\mathbf{X}^\perp$ is of course $\mathbf{M}$.

Our focus lies in the general linear model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\,, \quad \text{or shortly } \mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}\,, \tag{1.2}$$

where $\mathbf{X}_{n\times p}$ is a known fixed model matrix, the vector $\mathbf{y}$ is an observable n-dimensional random vector, $\boldsymbol{\beta}$ is a $p\times 1$ vector of unknown (but fixed) 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 matrix $\mathbf{V}$ is known and can be singular. Premultiplying the model $\mathscr{M}$ by an $f \times n$ matrix $\mathbf{F}$ yields the transformed model

$$\mathbf{Fy} = \mathbf{FX}\boldsymbol{\beta} + \mathbf{F}\boldsymbol{\varepsilon}\,, \quad \text{or shortly } \mathscr{M}_t = \{\mathbf{Fy}, \mathbf{FX}\boldsymbol{\beta}, \mathbf{FVF}'\}\,. \tag{1.3}$$

The transformed model $\mathscr{M}_t$ will play a crucial role in our considerations.

Let $\mathbf{y}_*$ denote a $q \times 1$ unobservable random vector containing new future observations. The new observations are assumed to be generated from

$$\mathbf{y}_* = \mathbf{X}_*\boldsymbol{\beta} + \boldsymbol{\varepsilon}_*\,, \tag{1.4}$$

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 have

$$\mathrm{E}\begin{pmatrix}\mathbf{y}\\ \mathbf{y}_*\end{pmatrix} = \begin{pmatrix}\mathbf{X}\boldsymbol{\beta}\\ \mathbf{X}_*\boldsymbol{\beta}\end{pmatrix} = \begin{pmatrix}\mathbf{X}\\ \mathbf{X}_*\end{pmatrix}\boldsymbol{\beta}, \quad \mathrm{cov}\begin{pmatrix}\mathbf{y}\\ \mathbf{y}_*\end{pmatrix} = \begin{pmatrix}\mathbf{V} & \mathbf{V}_{12}\\ \mathbf{V}_{21} & \mathbf{V}_{22}\end{pmatrix} = \boldsymbol{\Gamma}, \tag{1.5}$$

where the covariance matrix $\boldsymbol{\Gamma}$ is assumed to be known. We denote this setup shortly as

$$\mathscr{M}_* = \left\{ \begin{pmatrix}\mathbf{y}\\ \mathbf{y}_*\end{pmatrix}, \begin{pmatrix}\mathbf{X}\\ \mathbf{X}_*\end{pmatrix}\boldsymbol{\beta}, \begin{pmatrix}\mathbf{V} & \mathbf{V}_{12}\\ \mathbf{V}_{21} & \mathbf{V}_{22}\end{pmatrix} \right\}. \tag{1.6}$$

We call $\mathscr{M}_*$ "the linear model with new future observations". Of course, the phrase "new future" 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 $\boldsymbol{\varepsilon}_*$.

Suppose we transform $\mathscr{M}$ into $\mathscr{M}_t$ and do the prediction using the resulting transformed model. Corresponding to $\mathscr{M}_*$, we now have the following transformed setup:

$$\mathscr{M}_{t*} = \left\{ \begin{pmatrix}\mathbf{Fy}\\ \mathbf{y}_*\end{pmatrix}, \begin{pmatrix}\mathbf{FX}\\ \mathbf{X}_*\end{pmatrix}\boldsymbol{\beta}, \begin{pmatrix}\mathbf{FVF}' & \mathbf{FV}_{12}\\ \mathbf{V}_{21}\mathbf{F}' & \mathbf{V}_{22}\end{pmatrix} \right\}. \tag{1.7}$$

There is one further model that we will pay attention to, it is the linear mixed model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Zu} + \mathbf{e}, \quad \text{or shortly } \mathscr{L} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta} + \mathbf{Zu}, \mathbf{D}, \mathbf{R}, \mathbf{S}\}, \tag{1.8}$$

where $\mathbf{Z}_{n\times q}$ is a known matrix, $\mathbf{y}$, $\mathbf{X}$, and $\boldsymbol{\beta}$ are as in $\mathscr{M}$ but $\mathbf{u}$ is an unobservable q-dimensional random effect with $\mathrm{E}(\mathbf{u}) = \mathbf{0}$, $\mathrm{cov}(\mathbf{u}) = \mathbf{D}$, and $\mathbf{e}$ is a random error vector with $\mathrm{E}(\mathbf{e}) = \mathbf{0}$, $\mathrm{cov}(\mathbf{e}) = \mathbf{R}$, $\mathrm{cov}(\mathbf{e}, \mathbf{u}) = \mathbf{S}$.

A parametric function $\mathbf{X}_*\boldsymbol{\beta}$ is said to be estimable if it has a linear unbiased estimator $\mathbf{Cy}$. Such a matrix $\mathbf{C} \in \mathbb{R}^{q\times n}$ exists only when $\mathscr{C}(\mathbf{X}_*') \subset \mathscr{C}(\mathbf{X}')$. The linear unbiased estimator $\mathbf{Cy}$ is the best linear unbiased estimator, BLUE, of estimable $\mathbf{X}_*\boldsymbol{\beta}$ if $\mathbf{Cy}$ has the smallest covariance matrix in the Löwner sense among all linear unbiased estimators of $\mathbf{X}_*\boldsymbol{\beta}$:

$$\mathrm{cov}(\mathbf{Cy}) \leq_{\mathrm{L}} \mathrm{cov}(\mathbf{C}_{\#}\mathbf{y}) \quad \text{for all } \mathbf{C}_{\#} : \mathbf{C}_{\#}\mathbf{X} = \mathbf{X}_*, \tag{1.9}$$

i.e.,

$$\mathrm{cov}(\mathbf{C}_{\#}\mathbf{y}) - \mathrm{cov}(\mathbf{Cy}) \quad \text{is nonnegative definite for all } \mathbf{C}_{\#} : \mathbf{C}_{\#}\mathbf{X} = \mathbf{X}_*. \tag{1.10}$$

Some clarifying words about the Löwner ordering may be in place. First, 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

$$\mathbf{x}'\mathbf{Ax} \geq 0 \ \text{ for all } \mathbf{x} \in \mathbb{R}^n, \quad \text{or equivalently,} \quad \mathbf{A} = \mathbf{KK}' \ \text{ for some } \mathbf{K}. \tag{1.11}$$

Let $\mathbf{A}, \mathbf{B} \in \mathrm{NND}_n$. Then $\mathbf{A}$ is said to be below $\mathbf{B}$ in the Löwner sense and denoted as

$$\mathbf{A} \leq_{\mathrm{L}} \mathbf{B}, \quad \text{or} \quad \mathbf{B} - \mathbf{A} \geq_{\mathrm{L}} \mathbf{0}, \tag{1.12}$$

if $\mathbf{B} - \mathbf{A} \in \mathrm{NND}_n$, i.e., $\mathbf{x}'(\mathbf{B} - \mathbf{A})\mathbf{x} \geq 0$ so that $\mathbf{x}'\mathbf{Ax} \leq \mathbf{x}'\mathbf{Bx}$ for all $\mathbf{x} \in \mathbb{R}^n$. Löwner ordering is a very powerful and useful ordering in statistics. For example, (1.12) implies the following inequalities:

$$\det(\mathbf{A}) \leq \det(\mathbf{B}), \quad \mathrm{tr}(\mathbf{A}) \leq \mathrm{tr}(\mathbf{B}), \quad \mathrm{ch}_i(\mathbf{A}) \leq \mathrm{ch}_i(\mathbf{B}), \quad a_{ii} \leq b_{ii}, \tag{1.13}$$

where $i = 1, \ldots, n$ and $\det(\cdot), \mathrm{tr}(\cdot)$ and $\mathrm{ch}_i(\cdot)$ refer to the determinant, trace and the ith largest eigenvalue of the matrix argument, respectively.

The linear predictor $\mathbf{By}$ is said to be unbiased for $\mathbf{y}_*$ if the expected prediction error is zero, i.e., $\mathrm{E}(\mathbf{y}_* - \mathbf{By}) = \mathbf{0}$ for all $\boldsymbol{\beta} \in \mathbb{R}^p$, which happens if and only if $\mathbf{X}_*' = \mathbf{X}'\mathbf{B}'$. 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 $\mathbf{By}$ is the best linear unbiased predictor, BLUP, for $\mathbf{y}_*$, if we have the Löwner ordering

$$\mathrm{cov}(\mathbf{y}_* - \mathbf{By}) \leq_{\mathrm{L}} \mathrm{cov}(\mathbf{y}_* - \mathbf{B}_{\#}\mathbf{y}) \quad \text{for all } \mathbf{B}_{\#} : \mathbf{B}_{\#}\mathbf{X} = \mathbf{X}_* \,. \tag{1.14}$$

In other words, the BLUP provides the minimal covariance matrix for the prediction error. Notice that in (1.9) we minimize the covariance matrix of the linear unbiased estimator while in (1.14) we minimize the covariance matrix of the linear unbiased prediction error.

Lemma 1 characterizes the BLUE; see, e.g., Rao (1973, p. 282), and Lemma 2 characterizes the BLUP; see, e.g., Christensen (2011, p. 294), and Isotalo & Puntanen (2006, p. 1015).

Lemma 1. *Under the linear model $\mathscr{M}$, the estimator $\mathbf{Ay}$ is the* BLUE *for $\mathbf{X}\boldsymbol{\beta}$ if and only if*

$$\mathbf{A}(\mathbf{X} : \mathbf{V}\mathbf{X}^{\perp}) = (\mathbf{X} : \mathbf{0}) \,. \tag{1.15}$$

Correspondingly, $\mathbf{Cy}$ is the BLUE *of an estimable parametric function $\mathbf{X}_*\boldsymbol{\beta}$ if and only if*

$$\mathbf{C}(\mathbf{X} : \mathbf{V}\mathbf{X}^{\perp}) = (\mathbf{X}_* : \mathbf{0}) \,. \tag{1.16}$$

Lemma 2. *Consider the linear model $\mathscr{M}_*$, where $\mathscr{C}(\mathbf{X}_*') \subset \mathscr{C}(\mathbf{X}')$, i.e., $\mathbf{y}_*$ is predictable. The linear predictor $\mathbf{By}$ is the* BLUP *for $\mathbf{y}_*$ if and only if*

$$\mathbf{B}(\mathbf{X} : \mathbf{V}\mathbf{X}^{\perp}) = (\mathbf{X}_* : \mathbf{V}_{21}\mathbf{X}^{\perp}) = [\mathbf{X}_* : \mathrm{cov}(\mathbf{y}_*, \mathbf{y})\mathbf{X}^{\perp}] \,. \tag{1.17}$$

For the reviews of the BLUP-properties, see, e.g., Robinson (1991) and Haslett & Puntanen (2017).

We will frequently utilise Lemma 2.2.4 of Rao & Mitra (1971), which says that for nonnull matrices $\mathbf{A}$ and $\mathbf{C}$ the following holds:

$$\mathbf{AB}^{-}\mathbf{C} = \mathbf{AB}^{+}\mathbf{C} \iff \mathscr{C}(\mathbf{C}) \subset \mathscr{C}(\mathbf{B}) \;\&\; \mathscr{C}(\mathbf{A}') \subset \mathscr{C}(\mathbf{B}') \,. \tag{1.18}$$

We will have several matrix expressions involving generalized inverses and of course it is crucial to know whether they are dependent on the choice of the generalized inverses.

One well-known solution for $\mathbf{A}$ in (1.15) (which is always solvable) is

$$\mathbf{P}_{\mathbf{X};\mathbf{W}^-} := \mathbf{X}(\mathbf{X}'\mathbf{W}^-\mathbf{X})^-\mathbf{X}'\mathbf{W}^-, \tag{1.19}$$

where $\mathbf{W}$ is a matrix belonging to the set of nonnegative definite matrices defined as

$$\mathcal{W} = \left\{\mathbf{W} \in \mathbb{R}^{n\times n} : \mathbf{W} = \mathbf{V} + \mathbf{XUU}'\mathbf{X}',\ \mathscr{C}(\mathbf{W}) = \mathscr{C}(\mathbf{X} : \mathbf{V})\right\}. \tag{1.20}$$

For a review of the properties of $\mathcal{W}$, see, e.g., Puntanen et al. (2011, Sec. 12.3).

We assume the model $\mathscr{M}$ to be consistent in the sense that the observed value of $\mathbf{y}$ lies in $\mathscr{C}(\mathbf{X} : \mathbf{V})$ with probability 1. Hence we assume that under the model $\mathscr{M}$,

$$\mathbf{y} \in \mathscr{C}(\mathbf{X} : \mathbf{V}) = \mathscr{C}(\mathbf{X} : \mathbf{V}\mathbf{X}^{\perp}) = \mathscr{C}(\mathbf{X} : \mathbf{V}\mathbf{M}) = \mathscr{C}(\mathbf{X}) \oplus \mathscr{C}(\mathbf{V}\mathbf{M}) . \tag{1.21}$$

For the equality $\mathscr{C}(\mathbf{X} : \mathbf{V}) = \mathscr{C}(\mathbf{X} : \mathbf{V}\mathbf{M})$, we refer to Rao (1974, Lemma 2.1). The corresponding consistency as in (1.21) is assumed in all models that we will deal with. There is a related decomposition, see, e.g., Puntanen et al. (2011, Th. 8), that is worth mentioning in this context: for conformable matrices $\mathbf{A}$ and $\mathbf{B}$ we have

$$\mathscr{C}(\mathbf{A} : \mathbf{B}) = \mathscr{C}(\mathbf{A} : \mathbf{Q}_{\mathbf{A}}\mathbf{B}) . \tag{1.22}$$

Let $\mathbf{A}$ and $\mathbf{B}$ be $m \times n$ 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

$$\mathbf{Ay} = \mathbf{By} \quad \text{for all } \mathbf{y} \in \mathscr{C}(\mathbf{X} : \mathbf{V}) , \tag{1.23}$$

which will be a crucial property in our considerations. Hence we state the following lemma collecting together some equivalent expressions for (1.23). For part (d) of Lemma 3, see Groß & Trenkler (1998, Th. 1).

Lemma 3. *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* $\operatorname{cov}(\mathbf{Ay} - \mathbf{By}) = \mathbf{0}$,
(d) $\mathbf{AX} = \mathbf{BX}$, $\operatorname{cov}(\mathbf{Ay}) = \operatorname{cov}(\mathbf{By})$, *and* $2\operatorname{cov}(\mathbf{Ay}) = \operatorname{cov}(\mathbf{Ay}, \mathbf{By}) + \operatorname{cov}(\mathbf{By}, \mathbf{Ay})$.

The structure of the paper is as follows. In Section 2 we present some well-known conditions for the linear sufficiency. In Section 3 we provide some useful comments on the BLUPs for $\mathbf{y}_*$ and in particular for the error term $\boldsymbol{\varepsilon}_*$. According to our experience, the BLUP of $\boldsymbol{\varepsilon}_*$ has not received much attention in statistical literature. In Section 4 we study the equality of the BLUPs of $\boldsymbol{\varepsilon}_*$ under the original and the transformed model. Section 5 provides linear sufficiency charaterizations via certain covariance matrices and shows how the linear sufficiencies of $\mathbf{Fy}$ with respect to $\mathbf{y}_*$, $\mathbf{X}_*\boldsymbol{\beta}$, and $\boldsymbol{\varepsilon}_*$ are mutually related. In Section 6 we apply our results into the mixed linear model. In our paper, our attempt has been to call well-known (or pretty well-known) results Lemmas, while Theorems refer to our own contributions or clarifications. To increase the readability of our paper, the proofs of Theorems 2 and 5 are put into a separate Section 8. Our approach is a theoretical one and we focus on mathematical properties of the models.

2. Conditions for linear sufficiency

A linear statistic $\mathbf{Fy}$, where $\mathbf{F} \in \mathbb{R}^{f \times n}$, is called linearly sufficient for $\mathbf{X}\boldsymbol{\beta}$ under the model $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$, if there exists a matrix $\mathbf{A} \in \mathbb{R}^{n \times f}$ such that $\mathbf{AFy}$ is the BLUE for $\mathbf{X}\boldsymbol{\beta}$. Correspondingly, $\mathbf{Fy}$ is linearly sufficient for estimable $\mathbf{X}_*\boldsymbol{\beta}$, where $\mathbf{X}_* \in \mathbb{R}^{k \times p}$, if there exists a matrix $\mathbf{A} \in \mathbb{R}^{k \times f}$ such that $\mathbf{AFy}$ is the BLUE for $\mathbf{X}_*\boldsymbol{\beta}$. Sometimes we will use the phrase "BLUE-sufficient" and the notation $\mathbf{Fy} \in \mathcal{S}(\mathbf{X}_*\boldsymbol{\beta})$.

For the following Lemma 4, see, e.g., Baksalary & Kala (1981, 1986), Drygas (1983), Tian & Puntanen (2009, Th. 2.8), and Kala, Puntanen & Tian (2017, Th. 2). We will use the notation

$$\boldsymbol{\mu}_* = \mathbf{X}_*\boldsymbol{\beta}, \quad \tilde{\boldsymbol{\mu}}_* = \mathrm{BLUE}(\boldsymbol{\mu}_* \mid \mathscr{M}_*), \quad \tilde{\boldsymbol{\mu}}_{t*} = \mathrm{BLUE}(\boldsymbol{\mu}_* \mid \mathscr{M}_{t*}). \tag{2.1}$$

Lemma 4. *The statistic* $\mathbf{Fy}$ *is* BLUE-*sufficient for* $\mathbf{X}\boldsymbol{\beta}$ *under the model* $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$ *if and only if any of the following equivalent statements holds:*

(a) $\mathscr{C}\begin{pmatrix}\mathbf{X}' \\ \mathbf{0}\end{pmatrix} \subset \mathscr{C}\begin{pmatrix}\mathbf{X}'\mathbf{F}' \\ \mathbf{MVF}'\end{pmatrix}$, (b) $\mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{WF}')$, *where* $\mathbf{W} \in \mathcal{W}$.

The statistics $\mathbf{Fy}$ *is* BLUE-*sufficient for estimable* $\mathbf{X}_*\boldsymbol{\beta}$ *if and only if*

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

Moreover, let $\boldsymbol{\mu}_* = \mathbf{X}_*\boldsymbol{\beta}$ *be estimable under* $\mathscr{M}$ *and* $\mathscr{M}_t$. *Then the following statements are equivalent:*

(d) $\mathbf{Fy}$ *is* BLUE-*sufficient for* $\mathbf{X}_*\boldsymbol{\beta}$, *i.e.,* $\mathbf{Fy} \in \mathcal{S}(\mathbf{X}_*\boldsymbol{\beta})$,
(e) $\tilde{\boldsymbol{\mu}}_* = \tilde{\boldsymbol{\mu}}_{t*}$ *with probability* 1,
(f) $\mathrm{cov}(\tilde{\boldsymbol{\mu}}_*) = \mathrm{cov}(\tilde{\boldsymbol{\mu}}_{t*})$.

The concept of linear prediction sufficiency is defined analogically as follows: Let $\mathbf{y}_*$ be predictable under the model $\mathscr{M}_*$, i.e., $\mathscr{C}(\mathbf{X}'_*) \subset \mathscr{C}(\mathbf{X}')$. Then $\mathbf{Fy}$ 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

$$\mathbf{AF}(\mathbf{X} : \mathbf{VM}) = (\mathbf{X}_* : \mathbf{V}_{21}\mathbf{M}). \tag{2.2}$$

Corresponding to the phrase "BLUE-sufficient", we may use the term "BLUP-sufficient" and the notation $\mathbf{Fy} \in \mathcal{S}(\mathbf{y}_*)$.

The following lemma collects together some important properties of the linear prediction sufficiency. For the proof, see Isotalo & Puntanen (2006), and Isotalo et al. (2017).

Lemma 5. *Suppose that* $\mathbf{y}_*$ *is predictable under* $\mathscr{M}_*$ *and* $\mathscr{M}_{t*}$. *Then the following statements are equivalent:*

(b) $\mathbf{Fy}$ *is* BLUP-*sufficient for* $\mathbf{y}_*$, *or shortly* $\mathbf{Fy} \in \mathcal{S}(\mathbf{y}_*)$.
(c) $\mathscr{C}\begin{pmatrix}\mathbf{X}'_* \\ \mathbf{MV}_{12}\end{pmatrix} \subset \mathscr{C}\begin{pmatrix}\mathbf{X}'\mathbf{F}' \\ \mathbf{MVF}'\end{pmatrix}$.
(d) $\mathrm{BLUP}(\mathbf{y}_* \mid \mathscr{M}_*) = \mathrm{BLUP}(\mathbf{y}_* \mid \mathscr{M}_{t*})$, *or shortly,* $\tilde{\mathbf{y}}_* = \tilde{\mathbf{y}}_{t*}$ *with probability* 1.

Notice, somewhat interestingly, that we cannot add the condition $\mathrm{cov}(\tilde{\mathbf{y}}_*) = \mathrm{cov}(\tilde{\mathbf{y}}_{t*})$ into Lemma 5. We return into this feature in Section 5.

According to Isotalo et al. (2017), the statistic $\mathbf{Cy}$ is the BLUP for $\boldsymbol{\varepsilon}_*$ if and only if

$$\mathbf{C}(\mathbf{X} : \mathbf{VM}) = (\mathbf{0} : \mathbf{V}_{21}\mathbf{M}), \tag{2.3}$$

or, equivalently, $\mathbf{C} = \mathbf{AM}$ for some matrix $\mathbf{A}$ such that $\mathbf{AMVM} = \mathbf{V}_{21}\mathbf{M}$, which yields the following representation for $\mathrm{BLUP}(\boldsymbol{\varepsilon}_*)$:

$$\mathrm{BLUP}(\boldsymbol{\varepsilon}_*) = \mathbf{V}_{21}\mathbf{M}(\mathbf{MVM})^-\mathbf{My}. \tag{2.4}$$

Moreover, with probability 1,

$$\mathrm{BLUP}(\mathbf{y}_*) = \mathrm{BLUE}(\mathbf{X}_*\boldsymbol{\beta}) + \mathrm{BLUP}(\boldsymbol{\varepsilon}_*)\,. \tag{2.5}$$

The following lemma gives some BLUP-sufficiency properties of $\mathbf{Fy}$ for $\boldsymbol{\varepsilon}_*$; see Isotalo et al. (2017).

Lemma 6. *The statistic* $\mathbf{Fy}$ *is* BLUP*-sufficient for* $\boldsymbol{\varepsilon}_*$ *under* $\mathscr{M}_*$ *if and only if any of the following equivalent conditions holds:*

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

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

(c) $\mathrm{BLUP}(\boldsymbol{\varepsilon}_* \mid \mathscr{M}_*) = \mathrm{BLUP}(\boldsymbol{\varepsilon}_* \mid \mathscr{M}_{t*})$, *or shortly,* $\tilde{\boldsymbol{\varepsilon}}_* = \tilde{\boldsymbol{\varepsilon}}_*$ *with probability* 1.

In this case, we can add the condition $\mathrm{cov}(\tilde{\boldsymbol{\varepsilon}}_*) = \mathrm{cov}(\tilde{\boldsymbol{\varepsilon}}_*)$ into Lemma 6; we will deal with this property in Sections 3 and 4.

3. Some comments on the BLUPs under the original and the transformed model

Assume that the parametric function $\mathbf{X}_*\boldsymbol{\beta}$ is estimable under $\mathscr{M}$ as well as under $\mathscr{M}_t$, which happens if and only if $\mathscr{C}(\mathbf{X}_*') \subset \mathscr{C}(\mathbf{X}') \cap \mathscr{C}(\mathbf{X}'\mathbf{F}') = \mathscr{C}(\mathbf{X}'\mathbf{F}')$, so that

$$\mathbf{X}_* = \mathbf{LFX} \quad \text{for some matrix } \mathbf{L} \in \mathbb{R}^{q\times f}. \tag{3.1}$$

Throughout the paper we will assume that (3.1) holds. Recall that this also means that $\mathbf{y}_*$ is predictable under $\mathscr{M}_*$ and $\mathscr{M}_{t*}$. Consulting (1.18), we observe that $\mathbf{X}_* = \mathbf{X}_*\mathbf{X}^-\mathbf{X}$ for any choice of $\mathbf{X}^-$ and hence we can express $\mathbf{X}_*$, for example, as

$$\mathbf{X}_* = \mathbf{X}_*\mathbf{X}^+\mathbf{X} = \mathbf{X}_*\mathbf{P}_{\mathbf{X}'}\,. \tag{3.2}$$

We will use notation $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ and

$$\boldsymbol{\mu}_* = \mathbf{X}_*\boldsymbol{\beta} = \mathbf{LFX}\boldsymbol{\beta} = \mathbf{LF}\boldsymbol{\mu} = \mathbf{X}_*\mathbf{X}^+\mathbf{X}\boldsymbol{\beta} = \mathbf{X}_*\mathbf{X}^+\boldsymbol{\mu}\,. \tag{3.3}$$

The parametric function $\mathbf{X}\boldsymbol{\beta}$ is of course always estimable under $\mathscr{M}$ while under $\mathscr{M}_t$ it is estimable whenever

$$\mathscr{C}(\mathbf{X}') = \mathscr{C}(\mathbf{X}'\mathbf{F}')\,, \quad \text{i.e.,} \quad \mathrm{rank}(\mathbf{X}) = \mathrm{rank}(\mathbf{FX})\,. \tag{3.4}$$

There is no need to assume (3.4) throughout all our considerations; we need it only when dealing with the BLUE of $\mathbf{X}\boldsymbol{\beta}$ under $\mathscr{M}_t$.

Denoting

$$\mathbf{G} = \mathbf{X}(\mathbf{X}'\mathbf{W}^{-}\mathbf{X})^{-}\mathbf{X}'\mathbf{W}^{-} = \mathbf{P}_{\mathbf{X};\mathbf{W}^{-}}\,, \tag{3.5}$$

where $\mathbf{W} \in \mathcal{W}$, we have

$$\mathbf{G}\mathbf{y} = \mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta} \mid \mathscr{M}) = \mathrm{BLUE}(\boldsymbol{\mu} \mid \mathscr{M}) = \tilde{\boldsymbol{\mu}}\,. \tag{3.6}$$

If $\mathbf{X}\boldsymbol{\beta}$ is estimable under $\mathscr{M}_t$, then, in light of Lemma 1, $\mathbf{BFy}$ is the BLUE for $\mathbf{X}\boldsymbol{\beta}$ under $\mathscr{M}_t$ if and only if $\mathbf{B}$ satisfies

$$\mathbf{B}(\mathbf{FX} : \mathbf{FVF}'\mathbf{Q}_{\mathbf{FX}}) = (\mathbf{X} : \mathbf{0})\,. \tag{3.7}$$

Thus, see Kala, Markiewicz & Puntanen (2017, Sec. 6) and Markiewicz & Puntanen (2017, Sec. 3), the BLUE of $\mathbf{X}\boldsymbol{\beta}$ under $\mathscr{M}_t$ has, for example, the representation

$$\mathbf{G}_t\mathbf{y} = \mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta} \mid \mathscr{M}_t) = \mathrm{BLUE}(\boldsymbol{\mu} \mid \mathscr{M}_t) = \tilde{\boldsymbol{\mu}}_t\,, \tag{3.8}$$

where

$$\mathbf{G}_t = \mathbf{X}[\mathbf{X}'\mathbf{F}'(\mathbf{FWF}')^{-}\mathbf{FX}]^{-}\mathbf{X}'\mathbf{F}'(\mathbf{FWF}')^{-}\mathbf{F}. \tag{3.9}$$

Notice that $\mathbf{G}_t$ satisfies the equations

$$\mathbf{G}_t(\mathbf{X} : \mathbf{VF}'\mathbf{Q}_{\mathbf{FX}}) = (\mathbf{X} : \mathbf{0})\,, \tag{3.10a}$$
$$\mathbf{LFG}_t(\mathbf{X} : \mathbf{VF}'\mathbf{Q}_{\mathbf{FX}}) = (\mathbf{LFX} : \mathbf{0}) = (\mathbf{X}_* : \mathbf{0})\,, \tag{3.10b}$$

which means that $\mathbf{LFG}_t\mathbf{y} = \mathrm{BLUE}(\boldsymbol{\mu}_* \mid \mathscr{M}_t)$.

According to Isotalo et al. (2017, Sec. 4), the $\mathrm{BLUP}(\mathbf{y}_*)$ under $\mathscr{M}_*$ can be written as

$$\begin{aligned}\mathrm{BLUP}(\mathbf{y}_* \mid \mathscr{M}_*) &= \mathrm{BLUE}(\boldsymbol{\mu}_* \mid \mathscr{M}) + \mathbf{V}_{21}\mathbf{V}^{-}[\mathbf{y} - \mathrm{BLUE}(\boldsymbol{\mu} \mid \mathscr{M})]\\ &= \mathbf{LFGy} + \mathbf{V}_{21}\mathbf{V}^{-}(\mathbf{I}_n - \mathbf{G})\mathbf{y}\\ &= \mathbf{LFGy} + \mathbf{V}_{21}\mathbf{M}(\mathbf{MVM})^{-}\mathbf{My}\\ &= \mathrm{BLUE}(\boldsymbol{\mu}_* \mid \mathscr{M}) + \mathrm{BLUP}(\boldsymbol{\varepsilon}_* \mid \mathscr{M}_*)\,,\end{aligned} \tag{3.11}$$

or shortly,

$$\tilde{\mathbf{y}}_* = \tilde{\boldsymbol{\mu}}_* + \tilde{\boldsymbol{\varepsilon}}_*\,, \tag{3.12}$$

and

$$\begin{aligned}\mathrm{BLUP}(\mathbf{y}_* \mid \mathscr{M}_{t*}) &= \mathrm{BLUE}(\boldsymbol{\mu}_* \mid \mathscr{M}_t) + \mathbf{V}_{21}\mathbf{F}'(\mathbf{FVF}')^{-}\mathbf{F}[\mathbf{y} - \mathrm{BLUE}(\boldsymbol{\mu} \mid \mathscr{M}_t)]\\ &= \mathbf{LFG}_t\mathbf{y} + \mathbf{V}_{21}\mathbf{F}'(\mathbf{FVF}')^{-}\mathbf{F}(\mathbf{I}_n - \mathbf{G}_t)\mathbf{y}\\ &= \mathbf{LFG}_t\mathbf{y} + \mathbf{V}_{21}\mathbf{N}(\mathbf{NVN})^{-}\mathbf{Ny}\\ &= \mathrm{BLUE}(\boldsymbol{\mu}_* \mid \mathscr{M}_t) + \mathrm{BLUP}(\boldsymbol{\varepsilon}_* \mid \mathscr{M}_{t*})\,,\end{aligned} \tag{3.13}$$

or shortly,

$$\tilde{\mathbf{y}}_{t*} = \tilde{\boldsymbol{\mu}}_{t*} + \tilde{\boldsymbol{\varepsilon}}_{t*}\,, \tag{3.14}$$

where

$$\mathbf{N} = \mathbf{P}_{\mathbf{F}'\mathbf{Q}_{\mathbf{FX}}} = \mathbf{P}_{\mathscr{C}(\mathbf{F}')\cap\mathscr{C}(\mathbf{M})}\,. \tag{3.15}$$

The orthogonal projector $\mathbf{N}$ will have an important role in our considerations. We see that

$$\mathscr{C}(\mathbf{N}) = \mathscr{C}(\mathbf{F}'\mathbf{Q}_{\mathbf{FX}}) = \mathscr{C}(\mathbf{F}') \cap \mathscr{C}(\mathbf{M}) \,, \quad \mathbf{MF}'\mathbf{Q}_{\mathbf{FX}} = \mathbf{F}'\mathbf{Q}_{\mathbf{FX}} \,, \quad \mathbf{N} = \mathbf{MN} \,. \tag{3.16}$$

For properties like (3.16), see Markiewicz & Puntanen (2017, Sec. 2).

A couple of short remarks are worth mentioning. First, notice that in (3.11) and (3.13) the matrix $\mathbf{V}$ can be replaced with $\mathbf{W} \in \mathcal{W}$. Secondly, recall that in (3.11) and (3.13) we could replace $\mathbf{LF}$ with $\mathbf{X}_*\mathbf{X}^+$ Moreover, notice that the use of term $\mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta} \mid \mathscr{M}_t)$, as in the first expression in (3.13), requires, of course, that $\mathbf{X}\boldsymbol{\beta}$ is estimable under the transformed model $\mathscr{M}_t$. The use of other expressions in (3.13) does not require this assumption; the estimability of $\mathbf{X}_*\boldsymbol{\beta}$ under $\mathscr{M}_t$ is only needed.

Let us have a closer look at $\tilde{\mathbf{y}}_* = \tilde{\boldsymbol{\mu}}_* + \tilde{\boldsymbol{\varepsilon}}_*$ and $\tilde{\mathbf{y}}_{t*} = \tilde{\boldsymbol{\mu}}_{t*} + \tilde{\boldsymbol{\varepsilon}}_{t*}$. We observe that the random vectors $\tilde{\boldsymbol{\mu}}_*$ and $\tilde{\boldsymbol{\varepsilon}}_*$ are uncorrelated and the corresponding property holds also for $\tilde{\boldsymbol{\mu}}_{t*}$ and $\tilde{\boldsymbol{\varepsilon}}_{t*}$. Hence we have

$$\mathrm{cov}(\tilde{\mathbf{y}}_*) = \mathrm{cov}(\tilde{\boldsymbol{\mu}}_*) + \mathrm{cov}(\tilde{\boldsymbol{\varepsilon}}_*) \,, \quad \mathrm{cov}(\tilde{\mathbf{y}}_{t*}) = \mathrm{cov}(\tilde{\boldsymbol{\mu}}_{t*}) + \mathrm{cov}(\tilde{\boldsymbol{\varepsilon}}_{t*}) \,. \tag{3.17}$$

Now we have

$$\tilde{\boldsymbol{\varepsilon}}_* = \mathbf{V}_{21}\mathbf{M}(\mathbf{MVM})^-\mathbf{My}, \quad \tilde{\boldsymbol{\varepsilon}}_{t*} = \mathbf{V}_{21}\mathbf{N}(\mathbf{NVN})^-\mathbf{Ny}, \tag{3.18}$$

with covariance matrices

$$\mathrm{cov}(\tilde{\boldsymbol{\varepsilon}}_*) = \mathbf{V}_{21}\mathbf{M}(\mathbf{MVM})^-\mathbf{MV}_{12} \,, \quad \mathrm{cov}(\tilde{\boldsymbol{\varepsilon}}_{t*}) = \mathbf{V}_{21}\mathbf{N}(\mathbf{NVN})^-\mathbf{NV}_{12} \,. \tag{3.19}$$

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

$$\mathrm{cov}(\tilde{\boldsymbol{\varepsilon}}_* - \tilde{\boldsymbol{\varepsilon}}_{t*}) = \mathrm{cov}(\tilde{\boldsymbol{\varepsilon}}_*) - \mathrm{cov}(\tilde{\boldsymbol{\varepsilon}}_{t*}) \,, \tag{3.20}$$

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 3 and (3.20), equality $\tilde{\boldsymbol{\varepsilon}}_* = \tilde{\boldsymbol{\varepsilon}}_{t*}$ holds with probability 1 if and only if $\mathrm{cov}(\tilde{\boldsymbol{\varepsilon}}_*) = \mathrm{cov}(\tilde{\boldsymbol{\varepsilon}}_{t*})$. Thus, in light of Lemma 6,

$$\mathbf{Fy} \in \mathcal{S}(\boldsymbol{\varepsilon}_*) \iff \mathrm{cov}(\tilde{\boldsymbol{\varepsilon}}_*) = \mathrm{cov}(\tilde{\boldsymbol{\varepsilon}}_{t*}) \,. \tag{3.21}$$

We return to (3.21) in Theorem 2 and Remark 1 in Section 4.

The covariance matrix of the prediction error of $\boldsymbol{\varepsilon}_* - \tilde{\boldsymbol{\varepsilon}}_*$ is

$$\begin{aligned} \mathrm{cov}(\boldsymbol{\varepsilon}_* - \tilde{\boldsymbol{\varepsilon}}_*) &= \mathrm{cov}[\boldsymbol{\varepsilon}_* - \mathbf{V}_{21}\mathbf{M}(\mathbf{MVM})^-\mathbf{My}] \\ &= \mathbf{V}_{22} - \mathbf{V}_{21}\mathbf{M}(\mathbf{MVM})^-\mathbf{MV}_{12} \\ &= \mathrm{cov}(\boldsymbol{\varepsilon}_*) - \mathrm{cov}(\tilde{\boldsymbol{\varepsilon}}_*) \,, \end{aligned} \tag{3.22}$$

while the covariance matrix of the prediction error of $\boldsymbol{\varepsilon}_* - \tilde{\boldsymbol{\varepsilon}}_{t*}$ is

$$\begin{aligned} \mathrm{cov}(\boldsymbol{\varepsilon}_* - \tilde{\boldsymbol{\varepsilon}}_{t*}) &= \mathrm{cov}[\boldsymbol{\varepsilon}_* - \mathbf{V}_{21}\mathbf{N}(\mathbf{NVN})^-\mathbf{Ny}] \\ &= \mathbf{V}_{22} - \mathbf{V}_{21}\mathbf{N}(\mathbf{NVN})^-\mathbf{NV}_{12} \\ &= \mathrm{cov}(\boldsymbol{\varepsilon}_*) - \mathrm{cov}(\tilde{\boldsymbol{\varepsilon}}_{t*}) \,. \end{aligned} \tag{3.23}$$

Thus the covariance matrices of the prediction errors are equal if and only if $\mathrm{cov}(\tilde{\boldsymbol{\varepsilon}}_*) = \mathrm{cov}(\tilde{\boldsymbol{\varepsilon}}_{t*})$.

For clarity, let us collect some of our observations together.

Theorem 1. *Consider the* BLUP*s* $\tilde{\varepsilon}_*$ *and* $\tilde{\varepsilon}_{t*}$ *for* ε_* *under* $\mathscr{M}_*$ *and* $\mathscr{M}_{t*}$*, respectively. Denote* $\mathbf{N} = \mathbf{P}_{\mathscr{C}(\mathbf{F}')\cap\mathscr{C}(\mathbf{M})}$ *and let* $\mathbf{W} \in \mathcal{W}$*. Then the following statements hold:*

(a) $\tilde{\varepsilon}_* = \mathbf{V}_{21}\mathbf{M}(\mathbf{MVM})^-\mathbf{My}, \quad \tilde{\varepsilon}_{t*} = \mathbf{V}_{21}\mathbf{N}(\mathbf{NVN})^-\mathbf{Ny},$
(b) $\operatorname{cov}(\tilde{\varepsilon}_*) = \mathbf{V}_{21}\mathbf{M}(\mathbf{MVM})^-\mathbf{MV}_{12} = \mathbf{V}_{21}\mathbf{W}^{+1/2}\mathbf{P}_{\mathbf{W}^{1/2}\mathbf{M}}\mathbf{W}^{+1/2}\mathbf{V}_{12}\,,$
(c) $\operatorname{cov}(\tilde{\varepsilon}_{t*}) = \mathbf{V}_{21}\mathbf{N}(\mathbf{NVN})^-\mathbf{NV}_{12} = \mathbf{V}_{21}\mathbf{W}^{+1/2}\mathbf{P}_{\mathbf{W}^{1/2}\mathbf{N}}\mathbf{W}^{+1/2}\mathbf{V}_{12}\,,$
(d) $\operatorname{cov}(\tilde{\varepsilon}_* - \tilde{\varepsilon}_{t*}) = \operatorname{cov}(\tilde{\varepsilon}_*) - \operatorname{cov}(\tilde{\varepsilon}_{t*})\,,$
(e) $\operatorname{cov}(\tilde{\varepsilon}_{t*}) \leq_{\mathrm{L}} \operatorname{cov}(\tilde{\varepsilon}_*)\,,$
(f) $\operatorname{cov}(\varepsilon_* - \tilde{\varepsilon}_*) = \operatorname{cov}(\varepsilon_*) - \operatorname{cov}(\tilde{\varepsilon}_*)\,,$
(g) $\operatorname{cov}(\varepsilon_* - \tilde{\varepsilon}_{t*}) = \operatorname{cov}(\varepsilon_*) - \operatorname{cov}(\tilde{\varepsilon}_{t*})\,,$
(h) $\operatorname{cov}(\varepsilon_* - \tilde{\varepsilon}_*) \leq_{\mathrm{L}} \operatorname{cov}(\varepsilon_* - \tilde{\varepsilon}_{t*})\,,$
(i) $\operatorname{cov}(\varepsilon_* - \tilde{\varepsilon}_*) = \operatorname{cov}(\varepsilon_* - \tilde{\varepsilon}_{t*}) \iff \operatorname{cov}(\tilde{\varepsilon}_*) = \operatorname{cov}(\tilde{\varepsilon}_{t*})\,.$

4. Equality of the BLUPs of error term under the original and the transformed model

Let us study when the following holds:

$$\mathrm{BLUP}(\varepsilon_* \mid \mathscr{M}_*) = \mathrm{BLUP}(\varepsilon_* \mid \mathscr{M}_{t*}) \quad \text{with probability } 1, \tag{4.1}$$

i.e., for all $\mathbf{y} \in \mathscr{C}(\mathbf{X} : \mathbf{VM})$:

$$\mathbf{V}_{21}\mathbf{M}(\mathbf{MVM})^-\mathbf{My} = \mathbf{V}_{21}\mathbf{N}(\mathbf{NVN})^-\mathbf{Ny}, \tag{4.2}$$

where $\mathbf{N} = \mathbf{P}_{\mathbf{F}'\mathbf{Q}_{\mathbf{FX}}}$ and we know that $\mathbf{N}$ has properties like in (3.16). For $\mathbf{y} \in \mathscr{C}(\mathbf{X})$ we get zeros on both sides of (4.2). For $\mathbf{y} \in \mathscr{C}(\mathbf{VM})$ we get

$$\begin{aligned}\mathbf{V}_{21}\mathbf{M} &= \mathbf{V}_{21}\mathbf{N}(\mathbf{NVN})^-\mathbf{NVM}\\ &= \mathbf{V}_{21}\mathbf{MN}(\mathbf{NVN})^-\mathbf{NVM}\\ &:= \mathbf{V}_{21}\mathbf{ME}\,,\end{aligned} \tag{4.3}$$

where $\mathbf{E} = \mathbf{N}(\mathbf{NVN})^-\mathbf{NVM} \in \mathbb{R}^{n\times n}$. It is interesting to confirm (algebraically) that (4.3) is equivalent to the inclusion

$$\mathscr{C}(\mathbf{MV}_{12}) \subset \mathscr{C}(\mathbf{MVN}) = \mathscr{C}(\mathbf{MVF}'\mathbf{Q}_{\mathbf{FX}})\,, \tag{4.4}$$

which is a necessary and sufficient condition for $\mathbf{Fy}$ being linearly sufficient for ε_*. We can formulate this and some related result as a theorem.

Theorem 2. *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* ε_* *under* $\mathscr{M}_*$*.*

Remark 1. Some of the equivalences in Theorem 2 could be proved using appropriate characterizations of linear sufficiency of $\mathbf{Fy}$ for ε_*; see (3.21). However, it is of great interest to prove the equivalences of Theorem 2 using linear algebraic tools as done in Section 8. □

5. Linear sufficiency characterizations via covariance matrices

Recall that

$$\text{(i)}\ \operatorname{cov}(\tilde{\boldsymbol{\varepsilon}}_* - \tilde{\boldsymbol{\varepsilon}}_{t*}) = \operatorname{cov}(\tilde{\boldsymbol{\varepsilon}}_*) - \operatorname{cov}(\tilde{\boldsymbol{\varepsilon}}_{t*})\,, \quad \text{(ii)}\ \operatorname{cov}(\tilde{\boldsymbol{\varepsilon}}_{t*}) \leq_{\mathrm{L}} \operatorname{cov}(\tilde{\boldsymbol{\varepsilon}}_*)\,. \tag{5.1}$$

Moreover, we can show the following:

$$\text{(i)}\ \operatorname{cov}(\tilde{\boldsymbol{\mu}}_* - \tilde{\boldsymbol{\mu}}_{t*}) = \operatorname{cov}(\tilde{\boldsymbol{\mu}}_{t*}) - \operatorname{cov}(\tilde{\boldsymbol{\mu}}_*)\,, \quad \text{(ii)}\ \operatorname{cov}(\tilde{\boldsymbol{\mu}}_{t*}) \geq_{\mathrm{L}} \operatorname{cov}(\tilde{\boldsymbol{\mu}}_*)\,. \tag{5.2}$$

To confirm (i) of (5.2) [which further implies (ii) of (5.2)] we notice that choosing $\mathbf{W} = \mathbf{V} + \mathbf{XUU'X'} \in \mathcal{W}$, we have

$$\begin{aligned}
\operatorname{cov}(\tilde{\boldsymbol{\mu}}_*) &= \operatorname{cov}(\mathbf{LFGy}) = \mathbf{LFGVG'F'L'} \\
&= \mathbf{LF}[\mathbf{X}(\mathbf{X'W^-X'})^-\mathbf{X'} - \mathbf{XUU'X'}]\mathbf{F'L'}, \qquad (5.3)\\
\operatorname{cov}(\tilde{\boldsymbol{\mu}}_*, \tilde{\boldsymbol{\mu}}_{t*}) &= \operatorname{cov}(\mathbf{LFGy}, \mathbf{LFG}_t\mathbf{y}) = \mathbf{LFGVG}_t'\mathbf{F'L'} \\
&= \operatorname{cov}(\tilde{\boldsymbol{\mu}}_*)\,, \qquad (5.4)
\end{aligned}$$

because

$$\begin{aligned}
\mathbf{GVG}_t' &= \mathbf{G}(\mathbf{W} - \mathbf{XUU'X'})\mathbf{G}_t' \\
&= \mathbf{GWG}_t' - \mathbf{XUU'X'} \\
&= \mathbf{X}(\mathbf{X'W^-X})^-\mathbf{X'W^-WG}_t' - \mathbf{XUU'X'} \\
&= \mathbf{X}(\mathbf{X'W^-X})^-\mathbf{X'G}_t' - \mathbf{XUU'X'} \\
&= \mathbf{X}(\mathbf{X'W^-X})^-\mathbf{X'} - \mathbf{XUU'X'} \\
&= \mathbf{GVG'}.
\end{aligned} \tag{5.5}$$

Now on account of (5.4), $\operatorname{cov}(\tilde{\boldsymbol{\mu}}_* - \tilde{\boldsymbol{\mu}}_{t*}) = \operatorname{cov}(\tilde{\boldsymbol{\mu}}_*) + \operatorname{cov}(\tilde{\boldsymbol{\mu}}_{t*}) - 2\operatorname{cov}(\tilde{\boldsymbol{\mu}}_*)$ and hence (i) of (5.2) indeed holds.

The covariance matrix between $\tilde{\boldsymbol{\mu}}_{t*}$ and $\tilde{\boldsymbol{\varepsilon}}_*$ is

$$\begin{aligned}
\operatorname{cov}(\tilde{\boldsymbol{\mu}}_{t*}, \tilde{\boldsymbol{\varepsilon}}_*) &= \operatorname{cov}[\mathbf{LFG}_t\mathbf{y}, \mathbf{V}_{21}\mathbf{M}(\mathbf{MVM})^-\mathbf{My}] \\
&= \mathbf{LFG}_t\mathbf{VM}(\mathbf{MVM})^-\mathbf{MV}_{12}\,,
\end{aligned} \tag{5.6}$$

while $\tilde{\boldsymbol{\mu}}_*$ and $\tilde{\boldsymbol{\varepsilon}}_{t*}$ are uncorrelated:

$$\begin{aligned}
\operatorname{cov}(\tilde{\boldsymbol{\mu}}_*, \tilde{\boldsymbol{\varepsilon}}_{t*}) &= \operatorname{cov}[\mathbf{LFGy}, \mathbf{V}_{21}\mathbf{N}(\mathbf{NVN})^-\mathbf{Ny}] \\
&= \mathbf{LFGVN}(\mathbf{NVN})^-\mathbf{NV}_{12} \\
&= \mathbf{0}\,,
\end{aligned} \tag{5.7}$$

where we have used $\mathbf{GVN} = \mathbf{GVMN} = \mathbf{0}$. Now

$$\begin{aligned}
\operatorname{cov}(\tilde{\mathbf{y}}_* - \tilde{\mathbf{y}}_{t*}) &= \operatorname{cov}[(\tilde{\boldsymbol{\mu}}_* - \tilde{\boldsymbol{\mu}}_{t*}) + (\tilde{\boldsymbol{\varepsilon}}_* - \tilde{\boldsymbol{\varepsilon}}_{t*})] \\
&= \operatorname{cov}(\tilde{\boldsymbol{\mu}}_* - \tilde{\boldsymbol{\mu}}_{t*}) + \operatorname{cov}(\tilde{\boldsymbol{\varepsilon}}_* - \tilde{\boldsymbol{\varepsilon}}_{t*}) + \boldsymbol{\Sigma}_{\boldsymbol{\mu\varepsilon}} + \boldsymbol{\Sigma}_{\boldsymbol{\mu\varepsilon}}'\,,
\end{aligned} \tag{5.8}$$

where, recalling that $\operatorname{cov}(\tilde{\boldsymbol{\mu}}_*, \tilde{\boldsymbol{\varepsilon}}_*) = \operatorname{cov}(\tilde{\boldsymbol{\mu}}_{t*}, \tilde{\boldsymbol{\varepsilon}}_{t*}) = \operatorname{cov}(\tilde{\boldsymbol{\mu}}_*, \tilde{\boldsymbol{\varepsilon}}_{t*}) = \mathbf{0}$,

$$\boldsymbol{\Sigma}_{\boldsymbol{\mu\varepsilon}} = \operatorname{cov}(\tilde{\boldsymbol{\mu}}_* - \tilde{\boldsymbol{\mu}}_{t*}, \tilde{\boldsymbol{\varepsilon}}_* - \tilde{\boldsymbol{\varepsilon}}_{t*}) = -\operatorname{cov}(\tilde{\boldsymbol{\mu}}_{t*}, \tilde{\boldsymbol{\varepsilon}}_*)\,. \tag{5.9}$$

Thus,

$$\begin{aligned}\operatorname{cov}(\tilde{\mathbf{y}}_* - \tilde{\mathbf{y}}_{t*}) &= \operatorname{cov}[(\tilde{\boldsymbol{\mu}}_* - \tilde{\boldsymbol{\mu}}_{t*}) + (\tilde{\boldsymbol{\varepsilon}}_* - \tilde{\boldsymbol{\varepsilon}}_{t*})] \\ &= \operatorname{cov}(\tilde{\boldsymbol{\mu}}_* - \tilde{\boldsymbol{\mu}}_{t*}) + \operatorname{cov}(\tilde{\boldsymbol{\varepsilon}}_* - \tilde{\boldsymbol{\varepsilon}}_{t*}) + (\boldsymbol{\Sigma}_{\boldsymbol{\mu\varepsilon}} + \boldsymbol{\Sigma}'_{\boldsymbol{\mu\varepsilon}}) \\ &= [\operatorname{cov}(\tilde{\boldsymbol{\mu}}_{t*}) - \operatorname{cov}(\tilde{\boldsymbol{\mu}}_*)] + [\operatorname{cov}(\tilde{\boldsymbol{\varepsilon}}_*) - \operatorname{cov}(\tilde{\boldsymbol{\varepsilon}}_{t*})] + (\boldsymbol{\Sigma}_{\boldsymbol{\mu\varepsilon}} + \boldsymbol{\Sigma}'_{\boldsymbol{\mu\varepsilon}}) , \end{aligned} \tag{5.10}$$

or, using a shorter notation,

$$\boldsymbol{\Sigma}_{\mathbf{yy}} = \boldsymbol{\Sigma}_{\boldsymbol{\mu\mu}} + \boldsymbol{\Sigma}_{\boldsymbol{\varepsilon\varepsilon}} + (\boldsymbol{\Sigma}_{\boldsymbol{\mu\varepsilon}} + \boldsymbol{\Sigma}'_{\boldsymbol{\mu\varepsilon}}) . \tag{5.11}$$

We now have

$$\boldsymbol{\Sigma}_{\boldsymbol{\mu\mu}} = \operatorname{cov}(\tilde{\boldsymbol{\mu}}_* - \tilde{\boldsymbol{\mu}}_{t*}) = \operatorname{cov}(\tilde{\boldsymbol{\mu}}_{t*}) - \operatorname{cov}(\tilde{\boldsymbol{\mu}}_*) , \tag{5.12a}$$

$$\boldsymbol{\Sigma}_{\boldsymbol{\varepsilon\varepsilon}} = \operatorname{cov}(\tilde{\boldsymbol{\varepsilon}}_* - \tilde{\boldsymbol{\varepsilon}}_{t*}) = \operatorname{cov}(\tilde{\boldsymbol{\varepsilon}}_*) - \operatorname{cov}(\tilde{\boldsymbol{\varepsilon}}_{t*}) , \tag{5.12b}$$

but the following does *not* necessarily hold:

$$\boldsymbol{\Sigma}_{\mathbf{yy}} = \operatorname{cov}(\tilde{\mathbf{y}}_* - \tilde{\mathbf{y}}_{t*}) = \operatorname{cov}(\tilde{\mathbf{y}}_*) - \operatorname{cov}(\tilde{\mathbf{y}}_{t*}) . \tag{5.12c}$$

In terms of linear sufficiency, we have

$$\mathbf{Fy} \in \mathcal{S}(\boldsymbol{\mu}_*) \iff \operatorname{cov}(\tilde{\boldsymbol{\mu}}_*) = \operatorname{cov}(\tilde{\boldsymbol{\mu}}_{t*}) , \tag{5.13a}$$

$$\mathbf{Fy} \in \mathcal{S}(\boldsymbol{\varepsilon}_*) \iff \operatorname{cov}(\tilde{\boldsymbol{\varepsilon}}_*) = \operatorname{cov}(\tilde{\boldsymbol{\varepsilon}}_{t*}) , \tag{5.13b}$$

$$\mathbf{Fy} \in \mathcal{S}(\mathbf{y}_*) \iff \operatorname{cov}(\tilde{\mathbf{y}}_* - \tilde{\mathbf{y}}_{t*}) = \mathbf{0} . \tag{5.13c}$$

Here again the last statement “differs” from the others. Actually, it is of interest to prove the following:

Theorem 3. *Let $\tilde{\mathbf{y}}_*$, $\tilde{\boldsymbol{\varepsilon}}_*$ and $\tilde{\boldsymbol{\mu}}_*$ denote the* BLUP*s and* BLUE *under $\mathscr{M}_*$ and $\tilde{\mathbf{y}}_{t*}$, $\tilde{\boldsymbol{\varepsilon}}_{t*}$ and $\tilde{\boldsymbol{\mu}}_{t*}$ the corresponding* BLUP*s and* BLUE *under $\mathscr{M}_{t*}$ and*

$$\boldsymbol{\Sigma}_{\boldsymbol{\mu\varepsilon}} = \operatorname{cov}(\tilde{\boldsymbol{\mu}}_* - \tilde{\boldsymbol{\mu}}_{t*}, \tilde{\boldsymbol{\varepsilon}}_* - \tilde{\boldsymbol{\varepsilon}}_{t*}) = -\operatorname{cov}(\tilde{\boldsymbol{\mu}}_{t*}, \tilde{\boldsymbol{\varepsilon}}_*) . \tag{5.14}$$

Then the following statements are equivalent:

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

Proof. Combining

$$\begin{aligned}\operatorname{cov}(\tilde{\mathbf{y}}_*) - \operatorname{cov}(\tilde{\mathbf{y}}_{t*}) &= \operatorname{cov}(\tilde{\boldsymbol{\mu}}_*) + \operatorname{cov}(\tilde{\boldsymbol{\varepsilon}}_*) - \operatorname{cov}(\tilde{\boldsymbol{\mu}}_{t*}) - \operatorname{cov}(\tilde{\boldsymbol{\varepsilon}}_{t*}) \\ &= [\operatorname{cov}(\tilde{\boldsymbol{\mu}}_*) - \operatorname{cov}(\tilde{\boldsymbol{\mu}}_{t*})] + [\operatorname{cov}(\tilde{\boldsymbol{\varepsilon}}_*) - \operatorname{cov}(\tilde{\boldsymbol{\varepsilon}}_{t*})] \\ &= -[\operatorname{cov}(\tilde{\boldsymbol{\mu}}_{t*}) - \operatorname{cov}(\tilde{\boldsymbol{\mu}}_*)] + [\operatorname{cov}(\tilde{\boldsymbol{\varepsilon}}_*) - \operatorname{cov}(\tilde{\boldsymbol{\varepsilon}}_{t*})] \\ &= -\operatorname{cov}(\tilde{\boldsymbol{\mu}}_* - \tilde{\boldsymbol{\mu}}_{t*}) + \operatorname{cov}(\tilde{\boldsymbol{\varepsilon}}_* - \tilde{\boldsymbol{\varepsilon}}_{t*}) , \end{aligned} \tag{5.15}$$

and

$$\begin{aligned}\operatorname{cov}(\tilde{\mathbf{y}}_* - \tilde{\mathbf{y}}_{t*}) &= \operatorname{cov}[(\tilde{\boldsymbol{\mu}}_* - \tilde{\boldsymbol{\mu}}_{t*}) + (\tilde{\boldsymbol{\varepsilon}}_* - \tilde{\boldsymbol{\varepsilon}}_{t*})] \\ &= \operatorname{cov}(\tilde{\boldsymbol{\mu}}_* - \tilde{\boldsymbol{\mu}}_{t*}) + \operatorname{cov}(\tilde{\boldsymbol{\varepsilon}}_* - \tilde{\boldsymbol{\varepsilon}}_{t*}) + (\boldsymbol{\Sigma}_{\boldsymbol{\mu\varepsilon}} + \boldsymbol{\Sigma}'_{\boldsymbol{\mu\varepsilon}}) , \end{aligned} \tag{5.16}$$

and using $\operatorname{cov}(\tilde{\boldsymbol{\mu}}_* - \tilde{\boldsymbol{\mu}}_{t*}) = \operatorname{cov}(\tilde{\boldsymbol{\mu}}_{t*}) - \operatorname{cov}(\tilde{\boldsymbol{\mu}}_*)$ proves the claim. □

Let us return to the linear sufficiency of $\mathbf{Fy}$ for $\mathbf{y}_*$, which can be expressed interestingly in terms of covariance matrices as follows.

Theorem 4. *Using the notation of Theorem 3, 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* $\operatorname{cov}(\tilde{\varepsilon}_{t*}) - \operatorname{cov}(\tilde{\varepsilon}_*) = \frac{1}{2}(\mathbf{\Sigma}_{\boldsymbol{\mu\varepsilon}} + \mathbf{\Sigma}'_{\boldsymbol{\mu\varepsilon}})$.

Proof. On account of Lemma 3, the equality $\tilde{\mathbf{y}}_* = \tilde{\mathbf{y}}_{t*}$ holds with probability 1 [and thereby $\mathbf{Fy} \in \mathcal{S}(\mathbf{y}_*)$] if and only if $\operatorname{cov}(\tilde{\mathbf{y}}_*) = \operatorname{cov}(\tilde{\mathbf{y}}_{t*})$ holds along with

$$2\operatorname{cov}(\tilde{\mathbf{y}}_*) = \operatorname{cov}(\tilde{\mathbf{y}}_*, \tilde{\mathbf{y}}_{t*}) + \operatorname{cov}(\tilde{\mathbf{y}}_{t*}, \tilde{\mathbf{y}}_*) \,. \tag{5.17}$$

Straightforward calculation yields

$$\operatorname{cov}(\tilde{\mathbf{y}}_*, \tilde{\mathbf{y}}_{t*}) = \operatorname{cov}(\tilde{\boldsymbol{\mu}}_*) + \operatorname{cov}(\tilde{\varepsilon}_{t*}) + \operatorname{cov}(\tilde{\varepsilon}_*, \tilde{\boldsymbol{\mu}}_{t*}) = \operatorname{cov}(\tilde{\boldsymbol{\mu}}_*) + \operatorname{cov}(\tilde{\varepsilon}_{t*}) - \mathbf{\Sigma}'_{\boldsymbol{\mu\varepsilon}} \,. \tag{5.18}$$

Substituting (5.18) into (5.17) gives

$$2[\operatorname{cov}(\tilde{\boldsymbol{\mu}}_*) + \operatorname{cov}(\tilde{\varepsilon}_*)] = 2[\operatorname{cov}(\tilde{\boldsymbol{\mu}}_*) + \operatorname{cov}(\tilde{\varepsilon}_{t*})] - (\mathbf{\Sigma}_{\boldsymbol{\mu\varepsilon}} + \mathbf{\Sigma}'_{\boldsymbol{\mu\varepsilon}}) \,, \tag{5.19}$$

i.e., $2[\operatorname{cov}(\tilde{\varepsilon}_{t*}) - \operatorname{cov}(\tilde{\varepsilon}_*)] = \mathbf{\Sigma}_{\boldsymbol{\mu\varepsilon}} + \mathbf{\Sigma}'_{\boldsymbol{\mu\varepsilon}}$, and so the proof is completed. □

Next we characterize the mutual relations of the linear sufficiency of $\mathbf{Fy}$ for $\mathbf{X}_*\boldsymbol{\beta}$, $\boldsymbol{\varepsilon}_*$, and $\mathbf{y}_*$.

Theorem 5. *Consider the following three statements:*

(a) $\mathbf{Fy}$ *is* BLUE-*sufficient for* $\mathbf{X}_*\boldsymbol{\beta}$.
(b) $\mathbf{Fy}$ *is* BLUP-*sufficient for* $\boldsymbol{\varepsilon}_*$.
(c) $\mathbf{Fy}$ *is* BLUP-*sufficient for* $\mathbf{y}_*$.

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

$$\mathbf{\Sigma}_{\boldsymbol{\mu\varepsilon}} = -\mathbf{\Sigma}'_{\boldsymbol{\mu\varepsilon}} \,, \tag{5.20}$$

where $\mathbf{\Sigma}_{\boldsymbol{\mu\varepsilon}} = -\operatorname{cov}(\tilde{\boldsymbol{\mu}}_{t*}, \tilde{\varepsilon}_*)$, *is a necessary and sufficient condition for the implication*

$$\text{(c)} \implies \text{(a) } \textit{and} \text{ (b)} \,. \tag{5.21}$$

For the proof of Theorem 5, see Section 8.

According to Isotalo et al. (2017, Th. 3.4), the condition

$$\mathscr{C}(\mathbf{X}_*) \cap \mathscr{C}(\mathbf{V}_{21}\mathbf{M}) = \{\mathbf{0}\} \tag{5.22}$$

is a sufficient condition for (5.21). Next we show that this can be concluded from Theorem 5. To do this, notice first that the condition (5.20) is

$$\mathbf{LFG}_t\mathbf{VM}(\mathbf{MVM})^-\mathbf{MV}_{12} = -\mathbf{V}_{21}\mathbf{M}(\mathbf{MVM})^-\mathbf{MVG}_t'\mathbf{F}'\mathbf{L}', \tag{5.23}$$

or shortly

$$\mathbf{A}_1\mathbf{VA}_2' = -\mathbf{A}_2\mathbf{VA}_1' \,, \tag{5.24}$$

where

$$\begin{aligned}\mathbf{A}_1 &= \mathbf{LFG}_t = \mathbf{LFX}[\mathbf{X}'\mathbf{F}'(\mathbf{FWF}')^-\mathbf{FX}]^-\mathbf{X}'\mathbf{F}'(\mathbf{FWF}')^-\mathbf{F} \\ &= \mathbf{X}_*[\mathbf{X}'\mathbf{F}'(\mathbf{FWF}')^-\mathbf{FX}]^-\mathbf{X}'\mathbf{F}'(\mathbf{FWF}')^-\mathbf{F}, \end{aligned} \tag{5.25}$$

$$\mathbf{A}_2 = \mathbf{V}_{21}\mathbf{M}(\mathbf{MVM})^-\mathbf{M}\,. \tag{5.26}$$

Suppose that (c) holds which means that

$$(\mathbf{A}_1 - \mathbf{A}_2)\mathbf{V} = \mathbf{0}\,, \quad \text{i.e.,} \quad \mathbf{A}_1\mathbf{V} = \mathbf{A}_2\mathbf{V}. \tag{5.27}$$

Now $\mathscr{C}(\mathbf{A}_1\mathbf{V}) \subset \mathscr{C}(\mathbf{X}_*)$ and $\mathscr{C}(\mathbf{A}_2\mathbf{V}) \subset \mathscr{C}(\mathbf{V}_{21}\mathbf{M})$ and hence the disjointness (5.22) implies that $\mathbf{A}_1\mathbf{V} = \mathbf{A}_2\mathbf{V}$ holds if only if $\mathbf{A}_1\mathbf{V} = \mathbf{A}_2\mathbf{V} = \mathbf{0}$ which further implies (5.24). Thus we have confirmed, using Theorem 5, that (5.22) is a sufficient condition for (5.21).

6. Applications to linear mixed model

Consider the linear mixed model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Zu} + \mathbf{e}\,, \quad \text{denoted as} \quad \mathscr{L} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta} + \mathbf{Zu}, \mathbf{D}, \mathbf{R}, \mathbf{S}\}\,, \tag{6.1}$$

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 $\mathrm{E}(\mathbf{u}) = \mathbf{0}$, $\mathrm{cov}(\mathbf{u}) = \mathbf{D}_{q\times q}$, $\mathrm{cov}(\mathbf{e}, \mathbf{u}) = \mathbf{S}_{n\times q}$, and $\mathrm{E}(\mathbf{e}) = \mathbf{0}$, $\mathrm{cov}(\mathbf{e}) = \mathbf{R}_{n\times n}$. In this situation

$$\mathrm{cov}\begin{pmatrix}\mathbf{e}\\ \mathbf{u}\end{pmatrix} = \begin{pmatrix}\mathbf{R} & \mathbf{S}\\ \mathbf{S}' & \mathbf{D}\end{pmatrix}, \quad \mathrm{cov}\begin{pmatrix}\mathbf{y}\\ \mathbf{u}\end{pmatrix} = \begin{pmatrix}\boldsymbol{\Sigma} & \mathbf{ZD}+\mathbf{S}\\ (\mathbf{ZD}+\mathbf{S})' & \mathbf{D}\end{pmatrix}, \tag{6.2}$$

and $\mathrm{cov}(\mathbf{y}) = \mathbf{ZDZ}' + \mathbf{R} + \mathbf{ZS}' + \mathbf{SZ}' = \boldsymbol{\Sigma}$.

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

$$\mathscr{L}_* := \left\{ \begin{pmatrix}\mathbf{y}\\ \mathbf{g}\end{pmatrix}, \begin{pmatrix}\mathbf{X}\\ \mathbf{X}\end{pmatrix}\boldsymbol{\beta}, \begin{pmatrix}\boldsymbol{\Sigma} & (\mathbf{ZD}+\mathbf{S})\mathbf{Z}'\\ \mathbf{Z}(\mathbf{DZ}'+\mathbf{S}') & \mathbf{ZDZ}'\end{pmatrix} \right\}. \tag{6.3}$$

Corresponding to (1.2) and (1.4), we have

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\,, \quad \boldsymbol{\varepsilon} = \mathbf{Zu} + \mathbf{e}\,, \quad \mathrm{cov}(\boldsymbol{\varepsilon}) = \boldsymbol{\Sigma}\,, \tag{6.4a}$$

$$\mathbf{Zu} = \boldsymbol{\varepsilon}_*\,, \quad \mathrm{cov}(\boldsymbol{\varepsilon}_*) = \mathbf{ZDZ}', \quad \mathrm{cov}(\boldsymbol{\varepsilon}, \boldsymbol{\varepsilon}_*) = (\mathbf{ZD}+\mathbf{S})\mathbf{Z}'. \tag{6.4b}$$

Suppose that we transform the mixed model $\mathscr{L}$ by premultiplying it by $\mathbf{F}$ and do the prediction of $\mathbf{g} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Zu}$ using this transformed model. Then the resulting transformed setup is

$$\mathscr{L}_{t*} := \left\{ \begin{pmatrix}\mathbf{Fy}\\ \mathbf{g}\end{pmatrix}, \begin{pmatrix}\mathbf{FX}\\ \mathbf{X}\end{pmatrix}\boldsymbol{\beta}, \begin{pmatrix}\mathbf{F}\boldsymbol{\Sigma}\mathbf{F}' & \mathbf{F}(\mathbf{ZD}+\mathbf{S})\mathbf{Z}'\\ \mathbf{Z}(\mathbf{DZ}'+\mathbf{S}')\mathbf{F}' & \mathbf{ZDZ}'\end{pmatrix} \right\}. \tag{6.5}$$

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

$$\begin{pmatrix}\mathbf{B}_1\\ \mathbf{B}_2\end{pmatrix}(\mathbf{X} : \boldsymbol{\Sigma}\mathbf{M}) = \begin{pmatrix}\mathbf{X} & \mathbf{0}\\ \mathbf{0} & \mathbf{Z}(\mathbf{DZ}'+\mathbf{S}')\mathbf{M}\end{pmatrix} = \begin{pmatrix}\mathbf{X} & \mathbf{0}\\ \mathbf{0} & \mathrm{cov}(\mathbf{Zu}, \mathbf{y})\mathbf{M}\end{pmatrix}. \tag{6.6}$$

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

$$\mathbf{B}_3(\mathbf{X} : \boldsymbol{\Sigma}\mathbf{M}) = \big[\mathbf{X} : \mathbf{Z}(\mathbf{D}\mathbf{Z}' + \mathbf{S}')\mathbf{M}\big] = \big[\mathbf{X} : \operatorname{cov}(\mathbf{g}, \mathbf{y})\mathbf{M}\big] \,. \tag{6.7}$$

Thus the following holds:

$$\begin{aligned} \mathrm{BLUP}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} \mid \mathscr{L}) &= \mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta} \mid \mathscr{L}) + \mathrm{BLUP}(\mathbf{Z}\mathbf{u} \mid \mathscr{L}) \\ &= \mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta} \mid \mathscr{L}) + \mathbf{Z}\,\mathrm{BLUP}(\mathbf{u} \mid \mathscr{L})\,, \end{aligned} \tag{6.8}$$

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 $\mathbf{g} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}$:

$$\begin{aligned} \mathrm{BLUP}(\mathbf{g} \mid \mathscr{L}) = \tilde{\mathbf{g}} &= \mathbf{T}\mathbf{y} + \mathbf{Z}(\mathbf{D}\mathbf{Z}' + \mathbf{S}')\boldsymbol{\Sigma}^-(\mathbf{I}_n - \mathbf{T})\mathbf{y} \\ &= \mathbf{T}\mathbf{y} + \mathbf{Z}(\mathbf{D}\mathbf{Z}' + \mathbf{S}')\mathbf{M}(\mathbf{M}\boldsymbol{\Sigma}\mathbf{M})^-\mathbf{M}\mathbf{y} \\ &= \tilde{\boldsymbol{\mu}} + \mathbf{Z}\tilde{\mathbf{u}}\,, \end{aligned} \tag{6.9}$$

where $\mathbf{T} = \mathbf{X}(\mathbf{X}'\mathbf{W}_{\boldsymbol{\Sigma}}^-\mathbf{X})^-\mathbf{X}'\mathbf{W}_{\boldsymbol{\Sigma}}^-$ and $\mathbf{W}_{\boldsymbol{\Sigma}} \in \mathcal{W}_{\boldsymbol{\Sigma}}$,

$$\mathcal{W}_{\boldsymbol{\Sigma}} = \big\{\mathbf{W}_{\boldsymbol{\Sigma}} \in \mathbb{R}^{n\times n} : \mathbf{W}_{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma} + \mathbf{X}\mathbf{U}\mathbf{U}'\mathbf{X}',\ \mathscr{C}(\mathbf{W}_{\boldsymbol{\Sigma}}) = \mathscr{C}(\mathbf{X} : \boldsymbol{\Sigma})\big\}\,. \tag{6.10}$$

The BLUP of $\mathbf{g} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}$ under the transformed model $\mathscr{L}_{t*}$ can be expressed as

$$\begin{aligned} \mathrm{BLUP}(\mathbf{g} \mid \mathscr{L}_{t*}) = \tilde{\mathbf{g}}_t &= \mathbf{T}_t\mathbf{y} + \mathbf{Z}(\mathbf{D}\mathbf{Z}' + \mathbf{S}')\mathbf{F}'\boldsymbol{\Sigma}^-(\mathbf{I}_n - \mathbf{T}_t)\mathbf{y} \\ &= \mathbf{T}_t\mathbf{y} + \mathbf{Z}(\mathbf{D}\mathbf{Z}' + \mathbf{S}')\mathbf{N}(\mathbf{N}\boldsymbol{\Sigma}\mathbf{N})^-\mathbf{N}\mathbf{y} \\ &:= \tilde{\boldsymbol{\mu}}_t + \mathbf{Z}\tilde{\mathbf{u}}_t\,, \end{aligned} \tag{6.11}$$

where $\mathbf{N} = \mathbf{P}_{\mathbf{F}'\mathbf{Q}_{\mathbf{F}\mathbf{X}}}$ and

$$\mathbf{T}_t = \mathbf{X}[\mathbf{X}'\mathbf{F}'(\mathbf{F}\mathbf{W}_{\boldsymbol{\Sigma}}\mathbf{F}')^-\mathbf{F}\mathbf{X}]^-\mathbf{X}'\mathbf{F}'(\mathbf{F}\mathbf{W}_{\boldsymbol{\Sigma}}\mathbf{F}')^-\mathbf{F}. \tag{6.12}$$

The following result now follows from Theorem 5.

Theorem 6. *Consider the following three statements:*

(a) $\mathbf{F}\mathbf{y}$ *is* BLUE-*sufficient for* $\mathbf{X}\boldsymbol{\beta}$.
(b) $\mathbf{F}\mathbf{y}$ *is* BLUP-*sufficient for* $\mathbf{Z}\mathbf{u}$.
(c) $\mathbf{F}\mathbf{y}$ *is* BLUP-*sufficient for* $\mathbf{g} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}$.

Then above, any two conditions together imply the third one. Moreover, using the notation as in (6.9) *and* (6.11)*, the equality*

$$\operatorname{cov}(\tilde{\boldsymbol{\mu}}_t,\, \mathbf{Z}\tilde{\mathbf{u}}) = -\operatorname{cov}(\mathbf{Z}\tilde{\mathbf{u}},\, \tilde{\boldsymbol{\mu}}_t) \tag{6.13}$$

is a necessary and sufficient condition for the implication

$$\text{(c)} \quad\Longrightarrow\quad \text{(a) } \textit{and} \text{ (b)}\,. \tag{6.14}$$

7. Final remarks

This paper is dealing with the linear model $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$ and its transformed version $\mathscr{M}_t = \{\mathbf{Fy}, \mathbf{FX}\boldsymbol{\beta}, \mathbf{FVF}'\}$. Our observed response $\mathbf{y}$ is coming out of $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ and we are interested in predicting an unobservable random vector $\mathbf{y}_*$, which is believed to come from $\mathbf{y}_* = \mathbf{X}_*\boldsymbol{\beta} + \boldsymbol{\varepsilon}_*$. Here $\mathbf{X}_*\boldsymbol{\beta}$ is an estimable parametric function and $\boldsymbol{\varepsilon}_*$ is a random error whose covariance matrix with $\boldsymbol{\varepsilon}$ is known. The original setup for a linear model with new observations can be described as

$$\mathscr{M}_* = \left\{ \begin{pmatrix} \mathbf{y} \\ \mathbf{y}_* \end{pmatrix}, \begin{pmatrix} \mathbf{X} \\ \mathbf{X}_* \end{pmatrix} \boldsymbol{\beta}, \begin{pmatrix} \mathbf{V} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{pmatrix} \right\}, \tag{7.1}$$

while the corresponding transformed model is

$$\mathscr{M}_{t*} = \left\{ \begin{pmatrix} \mathbf{Fy} \\ \mathbf{y}_* \end{pmatrix}, \begin{pmatrix} \mathbf{FX} \\ \mathbf{X}_* \end{pmatrix} \boldsymbol{\beta}, \begin{pmatrix} \mathbf{FVF}' & \mathbf{FV}_{12} \\ \mathbf{V}_{21}\mathbf{F}' & \mathbf{V}_{22} \end{pmatrix} \right\}. \tag{7.2}$$

A linear statistic $\mathbf{Fy}$ is called linearly sufficient for the estimable parametric function of $\mathbf{X}_*\boldsymbol{\beta}$ under $\mathscr{M}$ if there exists a matrix $\mathbf{A}$ such that $\mathbf{AFy}$ is the best linear unbiased estimator, BLUE, for $\mathbf{X}_*\boldsymbol{\beta}$. Similarly, $\mathbf{Fy}$ is called linearly (prediction) sufficient for $\mathbf{y}_*$, if there exists a matrix $\mathbf{A}$ such that $\mathbf{AFy}$ is the best linear unbiased predictor, BLUP, for $\mathbf{y}_*$. What this means is that nothing essential has been lost if we base our prediction on the response $\mathbf{Fy}$ instead of $\mathbf{y}$. The concept of linear sufficiency is strongly connected to the transformed model $\mathscr{M}_t$, because for example, if $\mathbf{Fy}$ is linearly sufficient for $\mathbf{y}_*$, then every representation of the BLUP for $\mathbf{y}_*$ under the transformed model $\mathscr{M}_{t*}$ is BLUP also under the original model $\mathscr{M}_*$.

In this paper, we pay particular attention to the linear sufficiency of $\mathbf{Fy}$ with respect to $\mathbf{y}_*$, $\mathbf{X}_*\boldsymbol{\beta}$, and $\boldsymbol{\varepsilon}_*$ and the mutual relations between these sufficiencies. According to our experience, the BLUP of $\boldsymbol{\varepsilon}_*$ has not received much attention in statistical literature. In particular, we give an extensive review of some interesting properties of the covariance matrices of the BLUPs of $\boldsymbol{\varepsilon}_*$.

In general, as mentioned by Haslett & Puntanen (2017), best linear unbiased prediction has a wide range of applications, for example, plant variety trials, animal breeding, selection indices in quantitative genetics, quality estimation, time series, Kalman filtering and small area estimation. In practical applications, the covariance matrices involved may be unknown and that complicates the considerations substantially. The review article Haslett & Puntanen (2017) provides a short discussion on this matter. See also Robinson (1991). However, it seems to be quite a big challenge to apply the linear sufficiency concepts in such situations.

8. Some proofs

Proof of Theorem 2.
Consider first the equivalence of (b) and (c), which follows from the following:

$$\mathscr{C}(\mathbf{C}) \subset \mathscr{C}(\mathbf{A} : \mathbf{B}) \iff \mathscr{C}(\mathbf{Q_A C}) \subset \mathscr{C}(\mathbf{Q_A B}), \tag{8.1}$$

where $\mathbf{A}, \mathbf{B}$, and $\mathbf{C}$ are conformable matrices. To prove property (8.1), we note that pre-multiplying

$$\mathscr{C}(\mathbf{C}) \subset \mathscr{C}(\mathbf{A} : \mathbf{B}) = \mathscr{C}(\mathbf{A} : \mathbf{Q_A B}) \tag{8.2}$$

by $\mathbf{Q_A}$ yields $\mathscr{C}(\mathbf{Q_A C}) \subset \mathscr{C}(\mathbf{Q_A B})$. For the equality $\mathscr{C}(\mathbf{A} : \mathbf{B}) = \mathscr{C}(\mathbf{A} : \mathbf{Q_A B})$, see (1.22). On the other hand, we obviously have

$$\mathbf{C} = \mathbf{Q_A C} + \mathbf{P_A C}. \tag{8.3}$$

Using $\mathscr{C}(\mathbf{Q_A C}) \subset \mathscr{C}(\mathbf{Q_A B})$, (8.3) implies that for some matrices $\mathbf{D}_1, \dots, \mathbf{D}_4$ we have

$$\mathbf{C} = \mathbf{Q_A B D}_1 + \mathbf{A D}_2 = (\mathbf{Q_A B} : \mathbf{A})\mathbf{D}_3 = (\mathbf{A} : \mathbf{B})\mathbf{D}_4\,, \tag{8.4}$$

which means that (8.1) indeed holds. The equivalence of (b) and (c) follows from (8.1).

Consider then the claim (a). We observe that (a) holds if and only if $(\mathbf{I}_n - \mathbf{E}')\mathbf{M}\mathbf{V}_{12} = \mathbf{0}$, i.e.,

$$\mathscr{C}(\mathbf{M}\mathbf{V}_{12}) \subset \mathscr{N}(\mathbf{I}_n - \mathbf{E}') = \mathscr{C}(\mathbf{I}_n - \mathbf{E})^{\perp}. \tag{8.5}$$

Our task is now to show that

$$\mathscr{N}(\mathbf{I}_n - \mathbf{E}') = \mathscr{C}(\mathbf{MVN}) = \mathscr{C}(\mathbf{MVF}'\mathbf{Q_{FX}})\,. \tag{8.6}$$

We first observe that $(\mathbf{I}_n - \mathbf{E}')\mathbf{MVN} = \mathbf{0}$, i.e., $\mathscr{C}(\mathbf{MVN}) \subset \mathscr{N}(\mathbf{I}_n - \mathbf{E}')$. For the dimension of $\mathscr{N}(\mathbf{I}_n - \mathbf{E}')$ we get

$$\dim \mathscr{N}(\mathbf{I}_n - \mathbf{E}') = n - \operatorname{rank}(\mathbf{I}_n - \mathbf{E}) = \operatorname{rank}(\mathbf{E})\,, \tag{8.7}$$

where we have used the fact that $\mathbf{E}$ is idempotent, and that for an idempotent matrix $\mathbf{E} \in \mathbb{R}^{n\times n}$, $\operatorname{rank}(\mathbf{E}) = n - \operatorname{rank}(\mathbf{I}_n - \mathbf{E})$. The proof of (8.6) is completed by noting that

$$\begin{aligned}
\operatorname{rank}(\mathbf{NVM}) &\geq \operatorname{rank}[\mathbf{N}(\mathbf{NVN})^{-}\mathbf{NVM}] = \operatorname{rank}(\mathbf{E}) \\
&\geq \operatorname{rank}[\mathbf{NVN}(\mathbf{NVN})^{-}\mathbf{NVM}] \\
&= \operatorname{rank}(\mathbf{NVM})\,.
\end{aligned} \tag{8.8}$$

Consider then the claim (d). Notice that in the expressions of the covariance matrices of $\tilde{\varepsilon}_*$ and $\tilde{\varepsilon}_{t*}$ the matrix $\mathbf{V}$ can be replaced with $\mathbf{W} \in \mathcal{W}$ and thereby we have

$$\operatorname{cov}(\tilde{\varepsilon}_*) = \mathbf{V}_{21}\mathbf{W}^{+1/2}\mathbf{P}_{\mathbf{W}^{1/2}\mathbf{M}}\mathbf{W}^{+1/2}\mathbf{V}_{12}\,, \tag{8.9a}$$

$$\operatorname{cov}(\tilde{\varepsilon}_{t*}) = \mathbf{V}_{21}\mathbf{W}^{+1/2}\mathbf{P}_{\mathbf{W}^{1/2}\mathbf{N}}\mathbf{W}^{+1/2}\mathbf{V}_{12}\,. \tag{8.9b}$$

Using the property

$$\mathbf{P_A} - \mathbf{P_B} = \mathbf{P}_{\mathscr{C}(\mathbf{A})\cap\mathscr{C}(\mathbf{B})^{\perp}}\,, \tag{8.10}$$

which holds for conformable matrices $\mathbf{A}$ and $\mathbf{B}$ such that $\mathscr{C}(\mathbf{B}) \subset \mathscr{C}(\mathbf{A})$, see, e.g., Puntanen et al. (2011, p. 152), we observe that

$$\mathbf{P}_{\mathbf{W}^{1/2}\mathbf{M}} - \mathbf{P}_{\mathbf{W}^{1/2}\mathbf{N}} = \mathbf{P}_{\mathscr{C}(\mathbf{W}^{1/2}\mathbf{M})\cap\mathscr{C}(\mathbf{W}^{1/2}\mathbf{N})^{\perp}}\,. \tag{8.11}$$

Hence

$$\begin{aligned}\operatorname{cov}(\tilde{\varepsilon}_*) - \operatorname{cov}(\tilde{\varepsilon}_{t*}) &= \mathbf{V}_{21}\mathbf{W}^{+1/2}(\mathbf{P}_{\mathbf{W}^{1/2}\mathbf{M}} - \mathbf{P}_{\mathbf{W}^{1/2}\mathbf{N}})\mathbf{W}^{+1/2}\mathbf{V}_{12} \\ &= \mathbf{V}_{21}\mathbf{W}^{+1/2}\,\mathbf{P}_{\mathscr{C}(\mathbf{W}^{1/2}\mathbf{M})\cap\mathscr{C}(\mathbf{W}^{1/2}\mathbf{N})^\perp}\,\mathbf{W}^{+1/2}\mathbf{V}_{12} \\ &= \mathbf{0} \end{aligned} \tag{8.12}$$

if and only if

$$\begin{aligned}\mathscr{C}(\mathbf{W}^{+1/2}\mathbf{V}_{12}) &\subset \mathscr{C}[(\mathbf{W}^{1/2}\mathbf{M})^\perp : \mathbf{W}^{1/2}\mathbf{N}] \\ &= \mathscr{C}(\mathbf{W}^{+1/2}\mathbf{X} : \mathbf{Q}_\mathbf{W} : \mathbf{W}^{1/2}\mathbf{N})\,, \end{aligned} \tag{8.13}$$

where we have used $\mathscr{C}(\mathbf{W}^{1/2}\mathbf{M})^\perp = \mathscr{C}(\mathbf{W}^{+1/2}\mathbf{X} : \mathbf{Q}_\mathbf{W})$; see Markiewicz & Puntanen (2017, Lemma 4). Premultiplying (8.13) by $\mathbf{W}^{1/2}$ yields an equivalent inclusion

$$\mathscr{C}(\mathbf{V}_{12}) \subset \mathscr{C}(\mathbf{X} : \mathbf{WN}) = \mathscr{C}(\mathbf{X} : \mathbf{WMN}) = \mathscr{C}(\mathbf{X} : \mathbf{VN})\,, \tag{8.14}$$

where we have used $\mathbf{N} = \mathbf{MN}$. Thus we have shown the equivalence of (d) and (b) and the proof is completed. □

Proof of Theorem 5.
Consider the decomposition

$$\mathbf{\Sigma}_{\mathbf{yy}} = \mathbf{\Sigma}_{\boldsymbol{\mu\mu}} + \mathbf{\Sigma}_{\boldsymbol{\varepsilon\varepsilon}} + (\mathbf{\Sigma}_{\boldsymbol{\mu\varepsilon}} + \mathbf{\Sigma}'_{\boldsymbol{\mu\varepsilon}})\,, \tag{8.15}$$

i.e.,

$$\operatorname{cov}(\tilde{\mathbf{y}}_* - \tilde{\mathbf{y}}_{t*}) = \operatorname{cov}(\tilde{\boldsymbol{\mu}}_* - \tilde{\boldsymbol{\mu}}_{t*}) + \operatorname{cov}(\tilde{\varepsilon}_* - \tilde{\varepsilon}_{t*}) + (\mathbf{\Sigma}_{\boldsymbol{\mu\varepsilon}} + \mathbf{\Sigma}'_{\boldsymbol{\mu\varepsilon}})\,, \tag{8.16}$$

or, using other notation:

$$\begin{aligned}\operatorname{cov}(\tilde{\mathbf{y}}_* - \tilde{\mathbf{y}}_{t*}) &= \operatorname{cov}[(\tilde{\boldsymbol{\mu}}_* - \tilde{\boldsymbol{\mu}}_{t*}) + (\tilde{\varepsilon}_* - \tilde{\varepsilon}_{t*})] \\ &= \operatorname{cov}(\mathbf{A}_1\mathbf{y} + \mathbf{A}_2\mathbf{y}) \\ &= \mathbf{A}_1\mathbf{VA}_1' + \mathbf{A}_2\mathbf{VA}_2' + \mathbf{A}_1\mathbf{VA}_2' + \mathbf{A}_2\mathbf{VA}_1'\,, \end{aligned} \tag{8.17}$$

where

$$\mathbf{\Sigma}_{\boldsymbol{\mu\varepsilon}} = \mathbf{A}_1\mathbf{VA}_2' = \operatorname{cov}(\tilde{\boldsymbol{\mu}}_* - \tilde{\boldsymbol{\mu}}_{t*},\, \tilde{\varepsilon}_* - \tilde{\varepsilon}_{t*}) = -\operatorname{cov}(\tilde{\boldsymbol{\mu}}_{t*},\, \tilde{\varepsilon}_*)\,. \tag{8.18}$$

In terms of (8.15), the three claims of Theorem 5 can be expressed as follows:

$$\text{(a) } \mathbf{\Sigma}_{\boldsymbol{\mu\mu}} = \mathbf{0}, \qquad \text{(b) } \mathbf{\Sigma}_{\boldsymbol{\varepsilon\varepsilon}} = \mathbf{0}, \qquad \text{(c) } \mathbf{\Sigma}_{\mathbf{yy}} = \mathbf{0}. \tag{8.19}$$

Notice that $\mathbf{A}_1\mathbf{VA}_1' = \mathbf{0}$ implies $\mathbf{A}_1\mathbf{VA}_2' = \mathbf{0}$ and similarly $\mathbf{A}_2\mathbf{VA}_2' = \mathbf{0}$ implies $\mathbf{A}_1\mathbf{VA}_2' = \mathbf{0}$. Thus we can conclude that the first part of Theorem 5 indeed holds.

The second claim in Theorem 5 concerns the condition under which (c) $\mathbf{\Sigma}_{\mathbf{yy}} = \mathbf{0}$ would imply (a) $\mathbf{\Sigma}_{\boldsymbol{\mu\mu}} = \mathbf{0}$ and (b) $\mathbf{\Sigma}_{\boldsymbol{\varepsilon\varepsilon}} = \mathbf{0}$. In other words, we want to study when the following implication holds:

$$\mathbf{\Sigma}_{\boldsymbol{\mu\mu}} + \mathbf{\Sigma}_{\boldsymbol{\varepsilon\varepsilon}} = -(\mathbf{\Sigma}_{\boldsymbol{\mu\varepsilon}} + \mathbf{\Sigma}'_{\boldsymbol{\mu\varepsilon}}) \implies \mathbf{\Sigma}_{\boldsymbol{\mu\mu}} = \mathbf{\Sigma}_{\boldsymbol{\varepsilon\varepsilon}} = \mathbf{0}\,, \tag{8.20}$$

i.e.,

$$\mathbf{\Sigma}_{\boldsymbol{\mu\mu}} + \mathbf{\Sigma}_{\boldsymbol{\varepsilon\varepsilon}} = -(\mathbf{\Sigma}_{\boldsymbol{\mu\varepsilon}} + \mathbf{\Sigma}'_{\boldsymbol{\mu\varepsilon}}) \implies \mathbf{\Sigma}_{\boldsymbol{\mu\mu}} + \mathbf{\Sigma}_{\boldsymbol{\varepsilon\varepsilon}} = \mathbf{0}\,. \tag{8.21}$$

This clearly happens if and only if

$$\mathbf{\Sigma}_{\boldsymbol{\mu\varepsilon}} + \mathbf{\Sigma}'_{\boldsymbol{\mu\varepsilon}} = \mathbf{0}, \quad \text{i.e.,} \quad \mathbf{\Sigma}_{\boldsymbol{\mu\varepsilon}} = -\mathbf{\Sigma}'_{\boldsymbol{\mu\varepsilon}}\,, \tag{8.22}$$

which completes the proof. In passing we may recall the matrix $\mathbf{\Sigma}_{\boldsymbol{\mu\varepsilon}}$ satisfying (8.22) is called skew-symmetric (or antisymmetric). □

Acknowledgements

Part of this research was done during the meeting of an International Research Group on Multivariate and Mixed Linear Models in the Mathematical Research and Conference Center, Będlewo, Poland, March 2017, supported by the Stefan Banach International Mathematical Center. A talk based on this article, was given by the second author in The First East African Conference on Mathematical Statistics with Applications, Kigali, Rwanda, 14–16 June 2017. Thanks go to the organizers for great hospitality.

References

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 (http://dx.doi.org/10.1214/aos/1176345533)

Baksalary, J.K. & Kala, R. (1986). Linear sufficiency with respect to a given vector of parametric functions. *J. Stat. Plan. Inf.*, 14, 331–338. DOI (http://dx.doi.org/10.1016/0378-3758(86)90171-0)

Christensen, R. (2011). *Plane Answers to Complex Questions: the Theory of Linear Models*, 4th Edition. Springer, New York. DOI.(http://dx.doi.org/10.1007/978-1-4419-9816-3)

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

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. [Proceedings of the Second International Triennial Calcutta Symposium on Probability and Statistics, Calcutta, India, 30 December 1994 – 2 January 1995.]

Haslett, S.J. & Puntanen, S. (2017). Best linear unbiased prediction (BLUP). *Wiley StatsRef: Statistics Reference Online.* DOI.(http://dx.doi.org/10.1002/9781118445112.stat08120)

Haslett, S.J., Puntanen, S. & Arendacká, B. (2015). The link between the mixed and fixed linear models revisited. *Stat. Pap.*, 56, 849–861. DOI (http://dx.doi.org/10.1007/s00362-014-0611-9)

Isotalo, J., Markiewicz, A. & Puntanen, S. (2017). Some properties of linear prediction sufficiency in the linear model. *Proceedings of the LINSTAT-2016* (International Conference on Trends and Perspectives in Linear Statistical Inference, Istanbul, Turkey, 22–25 August 2016). (M. Tez & D. von Rosen, eds.) Springer, in press. Conference website.(http://www.linstat2016.com/en/default.asp)

Isotalo, J. & Puntanen, S. (2006). Linear prediction sufficiency for new observations in the general Gauss–Markov model. *Commun. Stat. Theory Methods*, 35, 1011–1023. DOI (http://dx.doi.org/10.1080/03610920600672146)

Kala, R., Markiewicz, A. & Puntanen, S. (2017). Some further remarks on the linear sufficiency in the linear model. *Applied and Computational Matrix Analysis: MatTriad, Coimbra, Portugal, September 2015, Selected, Revised Contributions.* (N. Bebiano, editor), Springer Proceedings in Mathematics & Statistics, vol. 192, 275–294. DOI. (http://dx.doi.org/10.1007/978-3-319-49984-0_19)

Kala, R., Puntanen, S. & Tian, Y. (2017). Some notes on linear sufficiency. *Stat. Pap.*, 58, 1–17. DOI (http://dx.doi.org/10.1007/s00362-015-0682-2)

Markiewicz, A. & Puntanen, S. (2015). All about the ⊥ with its applications in the linear statistical models. *Open Mathematics*, 13, 33–50. DOI.(http://dx.doi.org/10.1515/math-2015-0005)

Markiewicz, A. & Puntanen, S. (2017). Further properties of the linear sufficiency in the partitioned linear model. *Matrices, Statistics and Big Data* (Proceedings of the 25th International Workshop on Matrices and Statistics, IWMS-2016, held in Funchal, Madeira. Portugal, 6–9 June 2016). (S.E. Ahmed, F. Carvalho & S. Puntanen, eds.) Springer, in press.

Puntanen, S., Styan, G.P.H. & Isotalo, J. (2011). *Matrix Tricks for Linear Statistical Models: Our Personal Top Twenty.* Springer, Heidelberg. DOI (http://dx.doi.org/10.1007/978-3-642-10473-2)

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.DO DOI (http://dx.doi.org/10.1016/0047-259X(73)90042-0)

Rao, C.R. (1974). Projectors, generalized inverses and the BLUE's. *J. Roy. Statist. Soc. Ser. B*, 36, 442–448. DOI (http://www.jstor.org/stable/2984930)

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 (http://dx.doi.org/10.1214/ss/1177011926)

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 (http://dx.doi.org/10.1016/j.laa.2008.09.016)