Analysis of transformations of linear random-effects models

Yongge Tian

China Economics and Management Academy, Central University of Finance and Economics, Beijing 100081, China

Abstract. Assume that a linear random-effects model (LRM) $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ with $\boldsymbol{\beta} = \mathbf{A}\boldsymbol{\alpha} + \boldsymbol{\gamma}$ is transformed as $\mathbf{Ty} = \mathbf{TX}\boldsymbol{\beta} + \mathbf{T}\boldsymbol{\varepsilon} = \mathbf{TXA}\boldsymbol{\alpha} + \mathbf{TX}\boldsymbol{\gamma} + \mathbf{T}\boldsymbol{\varepsilon}$ by pre-multiplying a given matrix $\mathbf{T}$. Estimations/predictions of the unknown parameters under the two models are not necessarily the same because the transformation matrix $\mathbf{T}$ occurs in the statistical inference of the transformed model. This paper presents a general algebraic approach to the problem of best linear unbiased prediction (BLUP) of a joint vector all unknown parameters in the LRM and its linear transformations, and provides a group of fundamental and comprehensive results on mathematical and statistical properties of the BLUP under the LRM.

1 Introduction

Linear models that include random effects, or namely, linear random-effects models (LRMs for short) are statistical models of parameters that vary at more than one level, which have different names in data analysis according to their origination, such as, multilevel models, hierarchical models, nested models, split-plot designs, etc. These models are commonly used to analyze longitudinal and correlated data, which are available to account for the variability of model parameters due to different factors that influence response variables. Statistical inference concerning LRMs are now an important part in data analysis and has attracted much attention since 1970s. Let us consider a general LRM defined by

$$\mathscr{M}: \quad \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\beta} = \mathbf{A}\boldsymbol{\alpha} + \boldsymbol{\gamma}, \tag{1.1}$$

where in the first-stage model,
  $\mathbf{y} \in \mathbb{R}^{n \times 1}$ is a vector of observable response variables,
  $\mathbf{X} \in \mathbb{R}^{n \times p}$ is a known matrix of arbitrary rank,
  $\boldsymbol{\beta} \in \mathbb{R}^{p \times 1}$ is a vector of unobservable random variables,
  $\boldsymbol{\varepsilon} \in \mathbb{R}^{n \times 1}$ is a vector of unobservable random errors,
and in the second-stage model,
  $\mathbf{A} \in \mathbb{R}^{p \times k}$ is known matrix of arbitrary rank,
  $\boldsymbol{\alpha} \in \mathbb{R}^{k \times 1}$ is a vector of fixed but unknown parameters (fixed effects),
  $\boldsymbol{\gamma} \in \mathbb{R}^{p \times 1}$ is a vector of unobservable random variables (random effects).
The LRM in (1.1) is also called a two-level hierarchical linear model in the statistical literature, and are the simplest form of multilevel hierarchical linear models; see, e.g., [3, 5]. Substituting the second equation (1.1) into the first equation yields

$$\mathscr{M}: \quad \mathbf{y} = \mathbf{XA}\boldsymbol{\alpha} + \mathbf{X}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \tag{1.2}$$

which is a special form of the well-known linear mixed model with fixed effects $\boldsymbol{\alpha}$ and fixed effects $\boldsymbol{\gamma}$; see [6, 17, 27, 42].

Pre-multiplying a given matrix $\mathbf{T} \in \mathbb{R}^{m \times n}$ of arbitrary rank to both sides of (1.1) yields a transformed model as follows

$$\mathscr{N}: \quad \mathbf{Ty} = \mathbf{TX}\boldsymbol{\beta} + \mathbf{T}\boldsymbol{\varepsilon} = \mathbf{TXA}\boldsymbol{\alpha} + \mathbf{TX}\boldsymbol{\gamma} + \mathbf{T}\boldsymbol{\varepsilon}. \tag{1.3}$$

The purpose of this paper is to give a comprehensive analysis to the connections of the original model in (1.1) and its transformation in (1.3) via some novel algebraic tools in matrix theory. In order to obtain general conclusions in this approach, we assume that the expectation and dispersion matrix of the combined random vectors of $\boldsymbol{\gamma}$ and $\boldsymbol{\varepsilon}$ in (1.1) are given by

$$E\begin{bmatrix} \boldsymbol{\gamma} \\ \boldsymbol{\varepsilon} \end{bmatrix} = \mathbf{0}, \quad D\begin{bmatrix} \boldsymbol{\gamma} \\ \boldsymbol{\varepsilon} \end{bmatrix} = Cov\left\{ \begin{bmatrix} \boldsymbol{\gamma} \\ \boldsymbol{\varepsilon} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\gamma} \\ \boldsymbol{\varepsilon} \end{bmatrix} \right\} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix} := \boldsymbol{\Sigma}, \tag{1.4}$$

where $\boldsymbol{\Sigma}_{11} \in \mathbb{R}^{p \times p}$, $\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}'_{21} \in \mathbb{R}^{p \times n}$, and $\boldsymbol{\Sigma}_{22} \in \mathbb{R}^{n \times n}$ are known, $\boldsymbol{\Sigma} \in \mathbb{R}^{(p+n) \times (p+n)}$ is non-negative definite matrix of arbitrary rank. We don't assume any further restrictions to the patterns of the submatrices $\boldsymbol{\Sigma}_{ij}$ in (1.4) although they are usually taken as certain prescribed forms for a specified LRM in the statistical literature. In other words, if we meet with the situations where $\boldsymbol{\Sigma}$ is unknown or is given with special patterns, such as,

---

$\boldsymbol{\Sigma} = \mathrm{diag}\{\tau^2\mathbf{I}_p,\ \sigma^2\mathbf{I}_n\}$ or $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}_{p+n}$, where τ^2 and σ^2 are unknown positive numbers, the estimations of $\boldsymbol{\Sigma}$, or τ^2 and σ_2^2 from the observed data in (1.1) are other types of inference work, which we don't consider in this paper.

Before proceeding, we introduce the notation. $\mathbb{R}^{m\times n}$ stands for the collection of all $m \times n$ real matrices. The symbols $\mathbf{A}'$, $r(\mathbf{A})$, and $\mathscr{R}(\mathbf{A})$ stand for the transpose, the rank and the range (column space) of a matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$, respectively. $\mathbf{I}_m$ denotes the identity matrix of order m. The Moore–Penrose generalized inverse of $\mathbf{A}$, denoted by $\mathbf{A}^+$, is defined to be the unique solution $\mathbf{X}$ satisfying the four matrix equations $\mathbf{AGA} = \mathbf{A}$, $\mathbf{GAG} = \mathbf{G}$, $(\mathbf{AG})' = \mathbf{AG}$, and $(\mathbf{GA})' = \mathbf{GA}$. The symbols $\mathbf{P_A}$, $\mathbf{E_A}$, and $\mathbf{F_A}$ stand for the three orthogonal projectors (symmetric idempotent matrices) $\mathbf{P_A} = \mathbf{AA}^+$, $\mathbf{A}^\perp = \mathbf{E_A} = \mathbf{I}_m - \mathbf{AA}^+$, and $\mathbf{F_A} = \mathbf{I}_n - \mathbf{A}^+\mathbf{A}$, where both $\mathbf{E_A}$ and $\mathbf{F_A}$ satisfy $\mathbf{E_A} = \mathbf{F_{A'}}$ and $\mathbf{F_A} = \mathbf{E_{A'}}$, and the ranks of $\mathbf{E_A}$ and $\mathbf{F_A}$ are $r(\mathbf{E_A}) = m - r(\mathbf{A})$ and $r(\mathbf{F_A}) = n - r(\mathbf{A})$. Two symmetric matrices $\mathbf{A}$ and $\mathbf{B}$ of the same size are said to satisfy the inequality $\mathbf{A} \succcurlyeq \mathbf{B}$ in the Löwner partial ordering if $\mathbf{A} - \mathbf{B}$ is nonnegative definite.

Let

$$\widetilde{\mathbf{X}} = [\,\mathbf{X},\ \mathbf{I}_n\,], \quad \mathbf{R} = [\,\mathbf{I}_p,\ \mathbf{0}\,], \quad \mathbf{S} = [\,\mathbf{0},\ \mathbf{I}_n\,], \quad \widehat{\mathbf{X}} = \mathbf{XA}. \tag{1.5}$$

Under the general assumption of the dispersion matrix of unobservable random vectors in (1.1), the dispersion matrix of the observable random vector $\mathbf{y}$, as well as the covariance matrices among $\boldsymbol{\gamma}$, $\boldsymbol{\varepsilon}$, and $\mathbf{X}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}$ are given by

$$D(\mathbf{y}) = D(\mathbf{X}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}) = \widetilde{\mathbf{X}}\boldsymbol{\Sigma}\widetilde{\mathbf{X}}' = \mathbf{X}\boldsymbol{\Sigma}_{11}\mathbf{X}' + \mathbf{X}\boldsymbol{\Sigma}_{12} + \boldsymbol{\Sigma}_{21}\mathbf{X}' + \boldsymbol{\Sigma}_{22} \triangleq \mathbf{V}, \tag{1.6}$$

$$D(\mathbf{Ty}) = D(\mathbf{TX}\boldsymbol{\gamma} + \mathbf{T}\boldsymbol{\varepsilon}) = \mathbf{TVT}', \tag{1.7}$$

$$Cov\{\boldsymbol{\gamma},\ \mathbf{y}\} = Cov\{\boldsymbol{\gamma},\ \mathbf{X}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}\} = \mathbf{R}\boldsymbol{\Sigma}\widetilde{\mathbf{X}}' = \boldsymbol{\Sigma}_{11}\mathbf{X}' + \boldsymbol{\Sigma}_{12} \triangleq \mathbf{V}_1, \tag{1.8}$$

$$Cov\{\mathbf{X}\boldsymbol{\gamma},\ \mathbf{y}\} = Cov\{\mathbf{X}\boldsymbol{\gamma},\ \mathbf{X}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}\} = \mathbf{XR}\boldsymbol{\Sigma}\widetilde{\mathbf{X}}' = \mathbf{X}\boldsymbol{\Sigma}_{11}\mathbf{X}' + \mathbf{X}\boldsymbol{\Sigma}_{12} = \mathbf{XV}_1, \tag{1.9}$$

$$Cov\{\boldsymbol{\varepsilon},\ \mathbf{y}\} = Cov\{\boldsymbol{\varepsilon},\ \mathbf{X}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}\} = \mathbf{S}\boldsymbol{\Sigma}\widetilde{\mathbf{X}}' = \boldsymbol{\Sigma}_{21}\mathbf{X}' + \boldsymbol{\Sigma}_{22} \triangleq \mathbf{V}_2, \tag{1.10}$$

$$\mathscr{R}(\mathbf{V}_1') \subseteq \mathscr{R}(\mathbf{V}), \quad \mathscr{R}(\mathbf{V}_2') \subseteq \mathscr{R}(\mathbf{V}), \quad \mathbf{V} = \mathbf{XV}_1 + \mathbf{V}_2. \tag{1.11}$$

They are all known matrices under the assumptions in (1.1) and (1.4), and occur naturally in the statistical inference of (1.1); see, e.g., [37, 38].

The linear transformation of regression model in (1.3) occurs widely in linear statistical analysis. In order to appreciate the importance of this research, it is helpful to consider special transformations of LRMs (1.3):

(a) If $\mathbf{V}$ in (1.6) is positive definite, and $\mathbf{T} = \mathbf{V}^{-1/2}$, then (1.3) becomes

$$\mathscr{N}: \ \mathbf{V}^{-1/2}\mathbf{y} = \mathbf{V}^{-1/2}\mathbf{X}\boldsymbol{\beta} + \mathbf{V}^{-1/2}\boldsymbol{\varepsilon} = \mathbf{V}^{-1/2}\widehat{\mathbf{X}}\boldsymbol{\alpha} + \mathbf{V}^{-1/2}(\mathbf{X}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}), \tag{1.12}$$

the standard form of linear regression model, where the term $\mathbf{V}^{-1/2}(\mathbf{X}\boldsymbol{\gamma} + \boldsymbol{\varepsilon})$ satisfies the standard assumption $E(\mathbf{V}^{-1/2}(\mathbf{X}\boldsymbol{\gamma} + \boldsymbol{\varepsilon})] = \mathbf{0}$ and $D[(\mathbf{V}^{-1/2}(\mathbf{X}\boldsymbol{\gamma} + \boldsymbol{\varepsilon})] = \mathbf{I}_n$.

(b) If $\mathbf{T} = \widehat{\mathbf{X}}'\mathbf{W}$, where $\mathbf{W}$ is a nonnegative definite weight matrix, then (1.3) becomes

$$\mathscr{N}: \ \widehat{\mathbf{X}}'\mathbf{Wy} = \widehat{\mathbf{X}}'\mathbf{W}\widehat{\mathbf{X}}\boldsymbol{\alpha} + \widehat{\mathbf{X}}'\mathbf{WX}\boldsymbol{\gamma} + \widehat{\mathbf{X}}'\mathbf{W}\boldsymbol{\varepsilon}, \tag{1.13}$$

where the matrix equation $\widehat{\mathbf{X}}'\mathbf{Wy} = \widehat{\mathbf{X}}'\mathbf{W}\widehat{\mathbf{X}}\boldsymbol{\alpha}$ is always consistent, and is the normal matrix equation corresponding to the weighted least-squares problem $(\mathbf{y} - \widehat{\mathbf{X}}\boldsymbol{\alpha})'\mathbf{W}(\mathbf{y} - \widehat{\mathbf{X}}\boldsymbol{\alpha}) = \min.$

(c) If $\mathbf{T} = \mathbf{X}^\perp$, or $\mathbf{T} = \widehat{\mathbf{X}}^\perp$, then (1.3) becomes

$$\mathscr{N}_1: \ \mathbf{X}^\perp\mathbf{y} = \mathbf{X}^\perp\boldsymbol{\varepsilon}, \tag{1.14}$$

$$\mathscr{N}_2: \ \widehat{\mathbf{X}}^\perp\mathbf{y} = \widehat{\mathbf{X}}^\perp\mathbf{X}\boldsymbol{\gamma} + \widehat{\mathbf{X}}^\perp\boldsymbol{\varepsilon}, \tag{1.15}$$

respectively, which can be used to conduct the statistical inference of $\boldsymbol{\gamma}$ and $\boldsymbol{\varepsilon}$ in (1.1) if they are given in variance component forms.

(d) Partition (1.1) as

$$\mathscr{M}: \ \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}\boldsymbol{\beta} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix}(\mathbf{A}\boldsymbol{\alpha} + \boldsymbol{\gamma}) + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \end{bmatrix}, \tag{1.16}$$

and set $\mathbf{T} = [\,\mathbf{I}_{n_1},\ \mathbf{0}\,]$ and $\mathbf{T} = [\,\mathbf{0},\ \mathbf{I}_{n_2}\,]$ in (1.3), respectively, then we obtain two sub-sample models of (1.16) (or two simultaneous LRMs) as follows

$$\mathscr{N}_1: \ \mathbf{y}_1 = \mathbf{X}_1\boldsymbol{\beta} + \boldsymbol{\varepsilon}_1 = \mathbf{X}_1(\mathbf{A}\boldsymbol{\alpha} + \boldsymbol{\gamma}) + \boldsymbol{\varepsilon}_1, \tag{1.17}$$

$$\mathscr{N}_2: \ \mathbf{y}_2 = \mathbf{X}_2\boldsymbol{\beta} + \boldsymbol{\varepsilon}_2 = \mathbf{X}_2(\mathbf{A}\boldsymbol{\alpha} + \boldsymbol{\gamma}) + \boldsymbol{\varepsilon}_2, \tag{1.18}$$

where $\mathbf{y}_1 \in \mathbb{R}^{n_1 \times 1}$, $\mathbf{y}_2 \in \mathbb{R}^{n_2 \times 1}$, $\mathbf{X}_1 \in \mathbb{R}^{n_1 \times p}$, $\mathbf{X}_2 \in \mathbb{R}^{n_2 \times p}$, $\mathbf{A} \in \mathbb{R}^{p \times k}$, $\boldsymbol{\beta} \in \mathbb{R}^{p \times 1}$, $\boldsymbol{\alpha} \in \mathbb{R}^{k \times 1}$, $\boldsymbol{\gamma} \in \mathbb{R}^{p \times 1}$, $\boldsymbol{\varepsilon}_1 \in \mathbb{R}^{n_1 \times 1}$, $\boldsymbol{\varepsilon}_2 \in \mathbb{R}^{n_2 \times 1}$. Under this partition, we can predict/estimate $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, and $\boldsymbol{\alpha}$ simultaneously or separately. Some recent work on the connections of predictors/estimators under the full model in (1.3) and the two sub-sample models in (1.17) and (1.18) can be found in [39]. In particular, if $\mathbf{y}_2$ is unobservable (missing or future observations), we are able to predict $\mathbf{y}_2$ via $\mathbf{y}_1$ in (1.16) under the assumptions in (1.4) and (1.6)–(1.11).

(e) Under (1.16), let $\mathbf{T} = [\mathbf{I}_{n_1}, \mathbf{0}]$ and $\mathbf{T} = [-\mathbf{X}_2\mathbf{X}_1^+, \mathbf{I}_{n_2}]$, respectively. Then we obtain two sub-sample models of (1.1) (or two simultaneous LRMs) as follows

$$\mathscr{N}_1 : \ \mathbf{y}_1 = \mathbf{X}_1(\mathbf{A}\boldsymbol{\alpha} + \boldsymbol{\gamma}) + \boldsymbol{\varepsilon}_1, \tag{1.19}$$

$$\mathscr{N}_2 : \ \mathbf{y}_2 - \mathbf{X}_2\mathbf{X}_1^+\mathbf{y}_1 = (\mathbf{X}_2 - \mathbf{X}_2\mathbf{X}_1^+\mathbf{X}_1)(\mathbf{A}\boldsymbol{\alpha} + \boldsymbol{\gamma}) + \boldsymbol{\varepsilon}_2 - \mathbf{X}_2\mathbf{X}_1^+\boldsymbol{\varepsilon}_1, \tag{1.20}$$

where $\mathbf{y}_1 \in \mathbb{R}^{n_1 \times 1}$, $\mathbf{y}_2 \in \mathbb{R}^{n_2 \times 1}$, $\mathbf{X}_1 \in \mathbb{R}^{n_1 \times p}$, $\mathbf{X}_2 \in \mathbb{R}^{n_2 \times p}$, $\mathbf{A} \in \mathbb{R}^{p \times k}$, $\boldsymbol{\beta} \in \mathbb{R}^{p \times 1}$, $\boldsymbol{\alpha} \in \mathbb{R}^{k \times 1}$, $\boldsymbol{\gamma} \in \mathbb{R}^{p \times 1}$, $\boldsymbol{\varepsilon}_1 \in \mathbb{R}^{n_1 \times 1}$, $\boldsymbol{\varepsilon}_2 \in \mathbb{R}^{n_2 \times 1}$. Both (1.19) and (1.20) are equivalent to (1.16). Under the condition that $\mathscr{R}(\mathbf{X}_2') \subseteq \mathscr{R}(\mathbf{X}_1')$, (1.20) reduces to $\mathscr{N}_1 : \mathbf{y}_1 = \mathbf{X}_1(\mathbf{A}\boldsymbol{\alpha} + \boldsymbol{\gamma}) + \boldsymbol{\varepsilon}_1$ and $\mathscr{N}_2 : \mathbf{y}_2 - \mathbf{X}_2\mathbf{X}_1^+\mathbf{y}_1 = \boldsymbol{\varepsilon}_2 - \mathbf{X}_2\mathbf{X}_1^+\boldsymbol{\varepsilon}_1$. It is easier to make statistical inference under the two sub-sample models, because $\mathscr{N}_2$ involves no $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$, and $\boldsymbol{\gamma}$.

(f) Assume that (1.1) is given by

$$\mathscr{M} : \ \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \end{bmatrix}, \quad \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{bmatrix} \boldsymbol{\alpha} + \begin{bmatrix} \boldsymbol{\gamma}_1 \\ \boldsymbol{\gamma}_2 \end{bmatrix}, \tag{1.21}$$

where $\mathbf{y}_1 \in \mathbb{R}^{n_1 \times 1}$, $\mathbf{y}_2 \in \mathbb{R}^{n_2 \times 1}$, $\mathbf{X}_1 \in \mathbb{R}^{n_1 \times p_1}$, $\mathbf{X}_2 \in \mathbb{R}^{n_2 \times p_2}$, $\mathbf{A}_1 \in \mathbb{R}^{p_1 \times k}$, $\mathbf{A}_2 \in \mathbb{R}^{p_2 \times k}$, $\boldsymbol{\beta}_1 \in \mathbb{R}^{p_1 \times 1}$, $\boldsymbol{\beta}_2 \in \mathbb{R}^{p_2 \times 1}$, $\boldsymbol{\alpha} \in \mathbb{R}^{k \times 1}$, $\boldsymbol{\gamma}_1 \in \mathbb{R}^{p_1 \times 1}$, $\boldsymbol{\gamma}_2 \in \mathbb{R}^{p_2 \times 1}$, $\boldsymbol{\varepsilon}_1 \in \mathbb{R}^{n_1 \times 1}$, $\boldsymbol{\varepsilon}_2 \in \mathbb{R}^{n_2 \times 1}$, then two sub-sample models of (1.21) (or two simultaneous LRMs) are

$$\mathscr{N}_1 : \ \mathbf{y}_1 = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}_1 = \mathbf{X}_1(\mathbf{A}_1\boldsymbol{\alpha} + \boldsymbol{\gamma}_1) + \boldsymbol{\varepsilon}_1, \tag{1.22}$$

$$\mathscr{N}_2 : \ \mathbf{y}_2 = \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}_2 = \mathbf{X}_2(\mathbf{A}_2\boldsymbol{\alpha} + \boldsymbol{\gamma}_2) + \boldsymbol{\varepsilon}_2. \tag{1.23}$$

Under this partition, we can predict/estimate $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, $\boldsymbol{\gamma}_1$, $\boldsymbol{\gamma}_2$, and $\boldsymbol{\alpha}$ simultaneously or separately. In particular, if $\mathbf{y}_2$ is unobservable, we are able to predict $\mathbf{y}_2$, $\boldsymbol{\beta}_2$, and $\boldsymbol{\gamma}_2$ via $\mathbf{y}_1$ in (1.21) under the assumptions in (1.4) and (1.6)–(1.11).

(g) Assume that (1.1) is given by

$$\mathscr{M} : \ \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \end{bmatrix}, \quad \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \end{bmatrix} + \begin{bmatrix} \boldsymbol{\gamma}_1 \\ \boldsymbol{\gamma}_2 \end{bmatrix}, \tag{1.24}$$

where $\mathbf{y}_1 \in \mathbb{R}^{n_1 \times 1}$, $\mathbf{y}_2 \in \mathbb{R}^{n_2 \times 1}$, $\mathbf{X}_1 \in \mathbb{R}^{n_1 \times p_1}$, $\mathbf{X}_2 \in \mathbb{R}^{n_2 \times p_2}$, $\mathbf{A}_1 \in \mathbb{R}^{p_1 \times k_1}$, $\mathbf{A}_2 \in \mathbb{R}^{p_2 \times k_2}$, $\boldsymbol{\beta}_1 \in \mathbb{R}^{p_1 \times 1}$, $\boldsymbol{\beta}_2 \in \mathbb{R}^{p_2 \times 1}$, $\boldsymbol{\alpha}_1 \in \mathbb{R}^{k_1 \times 1}$, $\boldsymbol{\alpha}_2 \in \mathbb{R}^{k_2 \times 1}$, $\boldsymbol{\gamma}_1 \in \mathbb{R}^{p_1 \times 1}$, $\boldsymbol{\gamma}_2 \in \mathbb{R}^{p_2 \times 1}$, $\boldsymbol{\varepsilon}_1 \in \mathbb{R}^{n_1 \times 1}$, $\boldsymbol{\varepsilon}_2 \in \mathbb{R}^{n_2 \times 1}$, then two sub-sample models of (1.24) (or two simultaneous LRMs) are

$$\mathscr{N}_1 : \ \mathbf{y}_1 = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}_1 = \mathbf{X}_1(\mathbf{A}_1\boldsymbol{\alpha}_1 + \boldsymbol{\gamma}_1) + \boldsymbol{\varepsilon}_1, \tag{1.25}$$

$$\mathscr{N}_2 : \ \mathbf{y}_2 = \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}_2 = \mathbf{X}_2(\mathbf{A}_2\boldsymbol{\alpha}_2 + \boldsymbol{\gamma}_2) + \boldsymbol{\varepsilon}_2. \tag{1.26}$$

These two sub-sample models are special cases of so-called seemingly unrelated regression (SUR) models, because all the vectors and matrices in the two models are different. Under this partition, we can predict/estimate $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, $\boldsymbol{\gamma}_1$, $\boldsymbol{\gamma}_2$, $\boldsymbol{\alpha}_1$, and $\boldsymbol{\alpha}_2$ simultaneously or separately. In particular, if $\mathbf{y}_2$ is unobservable, we are able to predict/estimate $\mathbf{y}_2$, $\boldsymbol{\beta}_2$, $\boldsymbol{\gamma}_2$, and $\boldsymbol{\alpha}_2$ via $\mathbf{y}_1$ in (1.24) under the assumptions in (1.4) and (1.6)–(1.11). In this case, the covariance matrix of the four random vectors $\boldsymbol{\gamma}_1$, $\boldsymbol{\varepsilon}_1$, $\boldsymbol{\gamma}_2$, and $\boldsymbol{\varepsilon}_2$ is a 4×4 block matrix. Thus, it is tedious to make statistical inference under a general covariance matrix assumption for the four random vectors.

(h) Assume that (1.1) is given by the partitioned form

$$\mathscr{M} : \quad \mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}, \quad \boldsymbol{\beta}_1 = \mathbf{A}_1\boldsymbol{\alpha}_1 + \boldsymbol{\gamma}_1, \quad \boldsymbol{\beta}_2 = \mathbf{A}_2\boldsymbol{\alpha}_2 + \boldsymbol{\gamma}_2, \tag{1.27}$$

where $\mathbf{X}_1 \in \mathbb{R}^{n \times p_1}$, $\mathbf{X}_2 \in \mathbb{R}^{n \times p_2}$, $\mathbf{A}_1 \in \mathbb{R}^{p_1 \times k_1}$, $\mathbf{A}_2 \in \mathbb{R}^{p_2 \times k_2}$, $\boldsymbol{\beta}_1 \in \mathbb{R}^{p_1 \times 1}$, $\boldsymbol{\beta}_2 \in \mathbb{R}^{p_2 \times 1}$, $\boldsymbol{\alpha}_1 \in \mathbb{R}^{k_1 \times 1}$, $\boldsymbol{\alpha}_2 \in \mathbb{R}^{k_2 \times 1}$, $\boldsymbol{\gamma}_1 \in \mathbb{R}^{p_1 \times 1}$, $\boldsymbol{\gamma}_2 \in \mathbb{R}^{p_2 \times 1}$, $\boldsymbol{\varepsilon} \in \mathbb{R}^{n \times 1}$. Pre-multiplying $\mathbf{X}_2^\perp$ and $\mathbf{X}_1^\perp$ to the both sides of (1.27), we obtain two reduced models as follows

$$\mathscr{N}_1 : \quad \mathbf{X}_2^\perp\mathbf{y} = \mathbf{X}_2^\perp\mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2^\perp\boldsymbol{\varepsilon} = \mathbf{X}_2^\perp\mathbf{X}_1(\mathbf{A}_1\boldsymbol{\alpha}_1 + \boldsymbol{\gamma}_1) + \mathbf{X}_2^\perp\boldsymbol{\varepsilon}, \tag{1.28}$$

$$\mathscr{N}_2 : \quad \mathbf{X}_1^\perp\mathbf{y} = \mathbf{X}_1^\perp\mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{X}_1^\perp\boldsymbol{\varepsilon} = \mathbf{X}_1^\perp\mathbf{X}_2(\mathbf{A}_2\boldsymbol{\alpha}_2 + \boldsymbol{\gamma}_2) + \mathbf{X}_1^\perp\boldsymbol{\varepsilon}. \tag{1.29}$$

These two reduced models can be used to predict/estimate $\boldsymbol{\beta}_i$, $\boldsymbol{\gamma}_i$, and $\boldsymbol{\alpha}_i$, separately.

The above examples and discussions show that (1.3) covers many patterns of LRMs as its special cases, and therefore, they prompt us to characterize relationship between statistical inference of an original model and its transformed models. Note that the linear transformation $\mathbf{Ty}$ of the observable random vector $\mathbf{y}$ in (1.1) may preserve enough information to release predictions/estimations of unknown parameters. Thus, one of classic research problems about an original model and its transformed models is: when a linear transformation $\mathbf{Ty}$ contains all information of (1.1). This kind of problems were considered by many authors; see, e.g., [1, 7, 8, 10, 16, 18, 19, 20, 21, 22, 26, 28, 40]. This work can equivalently be converted to establishing possible equalities for predictors/estimators of unknown parameters in an original model and its transformed models. In the literature about regression models, the core issue is the estimation and hypothesis testing about the unknown parameters in the models. However, in various applications, it is utmost important for practitioners to predict future values of response variables in regression models. Prediction is also an important aspect of decision-making process through statistical methodology, while linear regression models play important roles in predicting unknown values of response variables corresponding to known values of explanatory variables. In some cases, it is of great practical interest to simultaneously identify the important predictors that correspond to both the fixed- and random-effects components in LRMs. Rao [32] showed a lemma on matrix function optimization problem and established a unified theory of linear estimation/prediction of all unknown parameters in a general linear model with fixed or mixed effects; see also Lemma 4.7 in [33]. Some previous work on simultaneous estimations/predictions of combined unknown parameters under linear regression models can be found in [4, 12, 21, 35, 41]. As a subsequent work on statistical inference of general LRMs, we consider in this paper BLUPs/BLUEs of all unknown parameters in (1.3), as well as relations between the BLUPs/BLUEs of all unknown parameters under (1.1) and (1.3). In order to obtain general results on simultaneous predictions/estimations of all unknown parameters in (1.2), we construct a joint vector of parametric functions containing the fixed effects, random effects, and error terms in (1.2) as follows

$$\boldsymbol{\phi} = [\,\mathbf{F}, \, \mathbf{G}, \, \mathbf{H}\,] \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\gamma} \\ \boldsymbol{\varepsilon} \end{bmatrix} = \mathbf{F}\boldsymbol{\alpha} + \mathbf{G}\boldsymbol{\gamma} + \mathbf{H}\boldsymbol{\varepsilon}, \qquad (1.30)$$

where $\mathbf{F} \in \mathbb{R}^{s \times k}$, $\mathbf{G} \in \mathbb{R}^{s \times p}$, and $\mathbf{H} \in \mathbb{R}^{s \times n}$ are known matrices of arbitrary ranks. In this case, we obtain the following formulas for the expectation and covariance matrices

$$E(\boldsymbol{\phi}) = \mathbf{F}\boldsymbol{\alpha}, \qquad (1.31)$$

$$D(\boldsymbol{\phi}) = \mathbf{J}\boldsymbol{\Sigma}\mathbf{J}' = \mathbf{G}\boldsymbol{\Sigma}_{11}\mathbf{G}' + \mathbf{G}\boldsymbol{\Sigma}_{12}\mathbf{H}' + \mathbf{H}\boldsymbol{\Sigma}_{21}\mathbf{G}' + \mathbf{H}\boldsymbol{\Sigma}_{22}\mathbf{H}', \qquad (1.32)$$

$$Cov\{\boldsymbol{\phi}, \, \mathbf{y}\} = \mathbf{J}\boldsymbol{\Sigma}\widetilde{\mathbf{X}}' = \mathbf{G}\boldsymbol{\Sigma}_{11}\mathbf{X}' + \mathbf{G}\boldsymbol{\Sigma}_{12} + \mathbf{H}\boldsymbol{\Sigma}_{21}\mathbf{X}' + \mathbf{H}\boldsymbol{\Sigma}_{22} = \mathbf{G}\mathbf{V}_1 + \mathbf{H}\mathbf{V}_2, \qquad (1.33)$$

$$Cov\{\boldsymbol{\phi}, \, \mathbf{Ty}\} = \mathbf{J}\boldsymbol{\Sigma}(\mathbf{T}\widetilde{\mathbf{X}})' = \mathbf{G}\boldsymbol{\Sigma}_{11}\mathbf{X}'\mathbf{T}' + \mathbf{G}\boldsymbol{\Sigma}_{12}\mathbf{T}' + \mathbf{H}\boldsymbol{\Sigma}_{21}\mathbf{X}'\mathbf{T}' + \mathbf{H}\boldsymbol{\Sigma}_{22}\mathbf{T}' = \mathbf{G}\mathbf{V}_1\mathbf{T}' + \mathbf{H}\mathbf{V}_2\mathbf{T}', \qquad (1.34)$$

where $\mathbf{J} = [\mathbf{G}, \, \mathbf{H}]$, Eq. (1.30) includes all vector operations in (1.1) and (1.3) as its special cases. For instance, if $\mathbf{F} = \mathbf{T}\widehat{\mathbf{X}}$, $\mathbf{G} = \mathbf{TX}$, and $\mathbf{H} = \mathbf{T}$, then (1.30) becomes (1.3):

$$\boldsymbol{\phi} = \mathbf{T}\widehat{\mathbf{X}}\boldsymbol{\alpha} + \mathbf{TX}\boldsymbol{\gamma} + \mathbf{T}\boldsymbol{\varepsilon} = \mathbf{Ty}.$$

Thus, the statistical inference of $\boldsymbol{\phi}$ is a comprehensive work, and will play prescriptive role for various special statistical inference problems under (1.1) from both theoretical and applied points of view. In order to establish a standard theory of the Best Linear Unbiased Predictor (BLUP) of $\boldsymbol{\phi}$ be as given in (1.30), we need the following classic statistical concepts and definitions originated from [9].

Definition 1.1. Let $\boldsymbol{\phi}$ be as given in (1.30).

(a) The $\boldsymbol{\phi}$ is said to be *predictable* under (1.1) if $E(\mathbf{L}_1\mathbf{y} - \boldsymbol{\phi}) = \mathbf{0}$ holds.

(b) The $\boldsymbol{\phi}$ is said to be *predictable* under (1.3) if $E(\mathbf{L}_2\mathbf{Ty} - \boldsymbol{\phi}) = \mathbf{0}$ holds.

Definition 1.2. Let $\boldsymbol{\phi}$ be as given in (1.30).

(a) If there exists a matrix $\mathbf{L}_1$ such that

$$E(\,\mathbf{L}_1\mathbf{y} - \boldsymbol{\phi}\,) = \mathbf{0} \text{ and } D(\,\mathbf{L}_1\mathbf{y} - \boldsymbol{\phi}\,) = \min \qquad (1.35)$$

hold in the Löwner partial ordering, the linear statistic $\mathbf{L}_1\mathbf{y}$ is defined to be the BLUP of $\boldsymbol{\phi}$ in (1.30), and is denoted by

$$\mathrm{BLUP}_{\mathscr{M}}(\boldsymbol{\phi}) = \mathbf{L}_1\mathbf{y} = \mathrm{BLUP}_{\mathscr{M}}(\,\mathbf{F}\boldsymbol{\alpha} + \mathbf{G}\boldsymbol{\gamma} + \mathbf{H}\boldsymbol{\varepsilon}\,). \qquad (1.36)$$

If $\mathbf{G} = \mathbf{0}$ and $\mathbf{H} = \mathbf{0}$, the linear statistic $\mathbf{L}_1\mathbf{y}$ in (1.36) is the well-known BLUE of $\mathbf{F}\boldsymbol{\alpha}$ under (1.2), and is denoted by

$$\mathbf{L}_1\mathbf{y} = \mathrm{BLUE}_{\mathscr{M}}(\mathbf{F}\boldsymbol{\alpha}). \qquad (1.37)$$

4

(b) If there exists a matrix $\mathbf{L}_2$ such that

$$E(\mathbf{L}_2\mathbf{Ty} - \boldsymbol{\phi}) = \mathbf{0} \text{ and } D(\mathbf{L}_2\mathbf{Ty} - \boldsymbol{\phi}) = \min \tag{1.38}$$

hold in the Löwner partial ordering, the linear statistic $\mathbf{L}_2\mathbf{Ty}$ is defined to be the BLUP of $\boldsymbol{\phi}$ in (1.30), and is denoted by

$$\text{BLUP}_{\mathcal{N}}(\boldsymbol{\phi}) = \mathbf{L}_2\mathbf{Ty} = \text{BLUP}_{\mathcal{N}}(\mathbf{F}\boldsymbol{\alpha} + \mathbf{G}\boldsymbol{\gamma} + \mathbf{H}\boldsymbol{\varepsilon}). \tag{1.39}$$

If $\mathbf{G} = \mathbf{0}$ and $\mathbf{H} = \mathbf{0}$, the linear statistic $\mathbf{L}_2\mathbf{Ty}$ in (1.39) are the well-known BLUE of $\mathbf{F}\boldsymbol{\alpha}$ under (1.3), and is denoted by

$$\mathbf{L}_2\mathbf{Ty} = \text{BLUE}_{\mathcal{N}}(\mathbf{F}\boldsymbol{\alpha}). \tag{1.40}$$

The theory of predictors/estimators under linear regression models belongs to the classical issues of mathematical statistics. The BLUEs of unknown parameters, such as $\boldsymbol{\alpha}$ in (1.2), as well as the BLUPs of unknown random variables such as $\boldsymbol{\beta}$ in (1.1), generally depend on the dispersion matrix of the observed random vector, and the covariances between the observed vector and the predicted random variable, as demonstrated in Theorem 3.1 below. Although the concept of BLUP is easy to understand from statistical and mathematical points of view, derivations of analytical results on BLUPs of unknown parameters under general regression models are challenging tasks. Tian [37] recently solved a constrained quadratic matrix-valued function optimization problem in the Löwner partial ordering, and used the solutions to obtain a group of analytical formulas for calculating the BLUPs of $\boldsymbol{\phi}$ in (1.30) under (1.1); see also the subsequent work in [38, 39]. The results in these papers provide solid mathematical foundation for conducting statistical inference of (1.1), and we can use them to derive many new mathematical and statistical properties of the BLUPs.

The structure of the paper is as follows. Section 2 presents a direct method of solving the matrix optimization problem in (1.38), and gives a variety of analytical formulas for calculating the BLUP/BLUE and their properties. Section 3 derives necessary and sufficient conditions for $\text{BLUP}_{\mathcal{M}}(\boldsymbol{\phi}) = \text{BLUP}_{\mathcal{N}}(\boldsymbol{\phi})$ to hold, and present various consequences for different choices of the unknown parameter vector $\boldsymbol{\phi}$. The proofs of the main results are given in Appendix.

2 Equations and formulas for BLUPs under transformed LRMs

In this section, we first introduce the concept of consistency of LRM in (1.1), and characterize the predictability of $\boldsymbol{\phi}$ in (1.30). We then derive a fundamental linear matrix equation associated with the BLUP of $\boldsymbol{\phi}$, and present a variety of fundamental properties of the BLUP.

Under the assumptions in (1.1), it is easy to see that

$$E\left[(\mathbf{I}_n - [\widehat{\mathbf{X}}, \mathbf{V}][\widehat{\mathbf{X}}, \mathbf{V}]^+)\mathbf{y}\right] = (\mathbf{I}_n - [\widehat{\mathbf{X}}, \mathbf{V}][\widehat{\mathbf{X}}, \mathbf{V}]^+)\widehat{\mathbf{X}}\boldsymbol{\alpha} = \mathbf{0},$$

$$D\left[(\mathbf{I}_n - [\widehat{\mathbf{X}}, \mathbf{V}][\widehat{\mathbf{X}}, \mathbf{V}]^+)\mathbf{y}\right] = (\mathbf{I}_n - [\widehat{\mathbf{X}}, \mathbf{V}][\widehat{\mathbf{X}}, \mathbf{V}]^+)\mathbf{V}(\mathbf{I}_n - [\widehat{\mathbf{X}}, \mathbf{V}][\widehat{\mathbf{X}}, \mathbf{V}]^+)' = \mathbf{0}.$$

These two equalities imply that $\mathbf{y} \in \mathscr{R}[\widehat{\mathbf{X}}, \mathbf{V}]$ holds with probability 1. In this case, (1.1) is said to be consistent; see [30, 31]. In this case, (1.3) is consistent as well, that is, $\mathbf{Ty} \in \mathscr{R}[\mathbf{T}\widehat{\mathbf{X}}, \mathbf{TVT}']$ holds with probability 1.

Substituting (1.2) and (1.30) into $\mathbf{L}_2\mathbf{Ty} - \boldsymbol{\phi}$ gives

$$\mathbf{L}_2\mathbf{Ty} - \boldsymbol{\phi} = \mathbf{L}_2\mathbf{T}\widehat{\mathbf{X}}\boldsymbol{\alpha} + \mathbf{L}_2\mathbf{TX}\boldsymbol{\gamma} + \mathbf{L}_2\mathbf{T}\boldsymbol{\varepsilon} - \mathbf{F}\boldsymbol{\alpha} - \mathbf{G}\boldsymbol{\gamma} - \mathbf{H}\boldsymbol{\varepsilon}$$

$$= (\mathbf{L}_2\mathbf{T}\widehat{\mathbf{X}} - \mathbf{F})\boldsymbol{\alpha} + (\mathbf{L}_2\mathbf{TX} - \mathbf{G})\boldsymbol{\gamma} + (\mathbf{L}_2\mathbf{T} - \mathbf{H})\boldsymbol{\varepsilon}.$$

Then, the expectation of $\mathbf{L}_2\mathbf{Ty} - \boldsymbol{\phi}$ is

$$E(\mathbf{L}_2\mathbf{Ty} - \boldsymbol{\phi}) = (\mathbf{L}_2\mathbf{T}\widehat{\mathbf{X}} - \mathbf{F})\boldsymbol{\alpha}; \tag{2.1}$$

the MMSE of $\mathbf{L}_2\mathbf{Ty} - \boldsymbol{\phi}$ is

$$E[(\mathbf{L}_2\mathbf{Ty} - \boldsymbol{\phi})(\mathbf{L}_2\mathbf{Ty} - \boldsymbol{\phi})']$$

$$= (\mathbf{L}_2\mathbf{T}\widehat{\mathbf{X}} - \mathbf{F})\boldsymbol{\alpha}\boldsymbol{\alpha}'(\mathbf{L}_2\mathbf{T}\widehat{\mathbf{X}} - \mathbf{F})' + [\mathbf{L}_2\mathbf{TX} - \mathbf{G}, \mathbf{L}_2\mathbf{T} - \mathbf{H}]\boldsymbol{\Sigma}[\mathbf{L}_2\mathbf{TX} - \mathbf{G}, \mathbf{L}_2\mathbf{T} - \mathbf{H}]'$$

$$= (\mathbf{L}_2\mathbf{T}\widehat{\mathbf{X}} - \mathbf{F})\boldsymbol{\alpha}\boldsymbol{\alpha}'(\mathbf{L}_2\mathbf{T}\widehat{\mathbf{X}} - \mathbf{F})' + (\mathbf{L}_2\mathbf{T}\widetilde{\mathbf{X}} - \mathbf{J})\boldsymbol{\Sigma}(\mathbf{L}_2\mathbf{T}\widetilde{\mathbf{X}} - \mathbf{J})'; \tag{2.2}$$

the dispersion matrix of $\mathbf{L}_2\mathbf{Ty} - \boldsymbol{\phi}$ is

$$D(\mathbf{L}_2\mathbf{Ty} - \boldsymbol{\phi}) = (\mathbf{L}_2\mathbf{T}\widetilde{\mathbf{X}} - \mathbf{J})\boldsymbol{\Sigma}(\mathbf{L}_2\mathbf{T}\widetilde{\mathbf{X}} - \mathbf{J})' \overset{\triangle}{=} f(\mathbf{L}_2). \tag{2.3}$$

Hence, the constrained covariance matrix minimization problem in (1.38) converts to a mathematical problem of minimizing the quadratic matrix-valued function $f(\mathbf{L}_2)$ subject to $\mathbf{L}_2\mathbf{T}\widehat{\mathbf{X}} = \mathbf{F}$.

Concerning the predictability of $\boldsymbol{\phi}$ in (1.30), we have the following result.

Lemma 2.1. *Let $\mathscr{N}$ be as given in (1.3). Then, the parameter vector $\boldsymbol{\phi}$ in (1.30) is predictable by $\mathbf{Ty}$ in (1.3) if and only if*

$$\mathscr{R}[(\mathbf{T}\widehat{\mathbf{X}})'] \supseteq \mathscr{R}(\mathbf{F}'). \tag{2.4}$$

We first give the matrix equation associated with the BLUP of $\boldsymbol{\phi}$ in (1.30), and present a variety of fundamental properties of the BLUP.

Theorem 2.2. (*Fundamental BLUP equation*) *Let $\mathscr{N}$ be as given in (1.3), and assume that the parameter vector $\boldsymbol{\phi}$ in (1.30) is predictable by $\mathbf{Ty}$ in (1.3). Then*

$$E(\mathbf{L}_2\mathbf{Ty} - \boldsymbol{\phi}) = \mathbf{0} \text{ and } D(\mathbf{L}_2\mathbf{Ty} - \boldsymbol{\phi}) = \min \Leftrightarrow \mathbf{L}_2[\,\mathbf{T}\widehat{\mathbf{X}},\, D(\mathbf{Ty})(\mathbf{T}\widehat{\mathbf{X}})^{\perp}\,] = [\,\mathbf{F},\, Cov\{\boldsymbol{\phi}, \mathbf{Ty}\}(\mathbf{T}\widehat{\mathbf{X}})^{\perp}\,]. \tag{2.5}$$

The matrix equation in (2.5), called the fundamental BLUP equation, is consistent, i.e,,

$$[\,\mathbf{F},\, Cov\{\boldsymbol{\phi}, \mathbf{Ty}\}(\mathbf{T}\widehat{\mathbf{X}})^{\perp}\,][\,\mathbf{T}\widehat{\mathbf{X}},\, D(\mathbf{Ty})(\mathbf{T}\widehat{\mathbf{X}})^{\perp}\,]^{+}[\,\mathbf{T}\widehat{\mathbf{X}},\, D(\mathbf{Ty})(\mathbf{T}\widehat{\mathbf{X}})^{\perp}\,] = [\,\mathbf{F},\, Cov\{\boldsymbol{\phi}, \mathbf{Ty}\}(\mathbf{T}\widehat{\mathbf{X}})^{\perp}\,] \tag{2.6}$$

holds under (2.4), while the general solution of $\mathbf{L}_2$ and the corresponding BLUP of $\boldsymbol{\phi}$ are given by

$$\begin{aligned}
\mathrm{BLUP}_{\mathscr{N}}(\boldsymbol{\phi}) &= \mathbf{L}_2\mathbf{Ty} \\
&= \Big([\,\mathbf{F},\, (\mathbf{G}\mathbf{V}_1 + \mathbf{H}\mathbf{V}_2)\mathbf{T}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}\,][\,\mathbf{T}\widehat{\mathbf{X}},\, \mathbf{T}\mathbf{V}\mathbf{T}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}\,]^{+}\mathbf{T} + \mathbf{U}[\,\mathbf{T}\widehat{\mathbf{X}},\, \mathbf{T}\mathbf{V}\mathbf{T}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}\,]^{\perp}\mathbf{T}\Big)\mathbf{y}, \tag{2.7}
\end{aligned}$$

where $\mathbf{U} \in \mathbb{R}^{s\times m}$ is arbitrary. Further, the following results hold.

(a) $r[\,\mathbf{T}\widehat{\mathbf{X}},\, \mathbf{T}\mathbf{V}\mathbf{T}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}\,] = r[\,\mathbf{T}\widehat{\mathbf{X}},\, \mathbf{T}\widetilde{\mathbf{X}}\boldsymbol{\Sigma}]$, $\mathscr{R}[\,\mathbf{T}\widehat{\mathbf{X}},\, \mathbf{T}\mathbf{V}\mathbf{T}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}\,] = \mathscr{R}[\,\mathbf{T}\widehat{\mathbf{X}},\, \mathbf{T}\widetilde{\mathbf{X}}\boldsymbol{\Sigma}]$, *and* $\mathscr{R}(\mathbf{T}\widehat{\mathbf{X}}) \cap \mathscr{R}[\mathbf{T}\mathbf{V}\mathbf{T}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}] = \{\mathbf{0}\}$.

(b) $\mathbf{L}_2$ *is unique if and only if* $r[\,\mathbf{T}\widehat{\mathbf{X}},\, \mathbf{T}\widetilde{\mathbf{X}}\boldsymbol{\Sigma}] = m$.

(c) $\mathbf{L}_2\mathbf{T}$ *is unique if and only if* $\mathscr{R}[\,\mathbf{T}\widehat{\mathbf{X}},\, \mathbf{T}\widetilde{\mathbf{X}}\boldsymbol{\Sigma}] = \mathscr{R}(\mathbf{T})$.

(d) $\mathrm{BLUP}_{\mathscr{N}}(\boldsymbol{\phi})$ *is unique if and only if* $\mathbf{Ty} \in \mathscr{R}[\,\mathbf{T}\widehat{\mathbf{X}},\, \mathbf{T}\mathbf{V}\mathbf{T}'\,]$ *with probability 1, namely, (1.3) is consistent.*

(e) *The dispersion matrix of* $\mathrm{BLUP}_{\mathscr{N}}(\boldsymbol{\phi})$ *is*

$$\begin{aligned}
D[\mathrm{BLUP}_{\mathscr{N}}(\boldsymbol{\phi})] &= [\,\mathbf{F},\, (\mathbf{G}\mathbf{V}_1 + \mathbf{H}\mathbf{V}_2)\mathbf{T}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}\,][\,\mathbf{T}\widehat{\mathbf{X}},\, \mathbf{T}\mathbf{V}\mathbf{T}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}\,]^{+}\mathbf{T}\mathbf{V}\mathbf{T}' \\
&\quad \times \Big([\,\mathbf{F},\, (\mathbf{G}\mathbf{V}_1 + \mathbf{H}\mathbf{V}_2)\mathbf{T}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}\,][\,\mathbf{T}\widehat{\mathbf{X}},\, \mathbf{T}\mathbf{V}\mathbf{T}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}\,]^{+}\Big)'; \tag{2.8}
\end{aligned}$$

the covariance matrix between $\mathrm{BLUP}_{\mathscr{N}}(\boldsymbol{\phi})$ *and* $\boldsymbol{\phi}$ *is*

$$Cov\{\mathrm{BLUP}_{\mathscr{N}}(\boldsymbol{\phi}),\, \boldsymbol{\phi}\} = [\,\mathbf{F},\, (\mathbf{G}\mathbf{V}_1 + \mathbf{H}\mathbf{V}_2)\mathbf{T}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}\,][\,\mathbf{T}\widehat{\mathbf{X}},\, \mathbf{T}\mathbf{V}\mathbf{T}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}\,]^{+}\mathbf{T}(\mathbf{G}\mathbf{V}_1 + \mathbf{H}\mathbf{V}_2)'; \tag{2.9}$$

the difference of the dispersion matrices of $\boldsymbol{\phi}$ *and* $\mathrm{BLUP}_{\mathscr{N}}(\boldsymbol{\phi})$ *is*

$$\begin{aligned}
D(\boldsymbol{\phi}) - D[\mathrm{BLUP}_{\mathscr{N}}(\boldsymbol{\phi})] &= \mathbf{J}\boldsymbol{\Sigma}\mathbf{J}' - [\,\mathbf{F},\, (\mathbf{G}\mathbf{V}_1 + \mathbf{H}\mathbf{V}_2)\mathbf{T}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}\,][\,\mathbf{T}\widehat{\mathbf{X}},\, \mathbf{T}\mathbf{V}\mathbf{T}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}\,]^{+}\mathbf{T}\mathbf{V}\mathbf{T}' \\
&\quad \times \Big([\,\mathbf{F},\, (\mathbf{G}\mathbf{V}_1 + \mathbf{H}\mathbf{V}_2)\mathbf{T}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}\,][\,\mathbf{T}\widehat{\mathbf{X}},\, \mathbf{T}\mathbf{V}\mathbf{T}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}\,]^{+}\Big)' \tag{2.10}
\end{aligned}$$

the dispersion matrix of the difference of $\boldsymbol{\phi}$ *and* $\mathrm{BLUP}_{\mathscr{N}}(\boldsymbol{\phi})$ *is*

$$\begin{aligned}
D[\boldsymbol{\phi} - \mathrm{BLUP}_{\mathscr{N}}(\boldsymbol{\phi})] &= \Big([\,\mathbf{F},\, (\mathbf{G}\mathbf{V}_1 + \mathbf{H}\mathbf{V}_2)\mathbf{T}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}\,][\,\mathbf{T}\widehat{\mathbf{X}},\, \mathbf{T}\mathbf{V}\mathbf{T}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}\,]^{+}\mathbf{T}\widetilde{\mathbf{X}} - \mathbf{J}\Big)\boldsymbol{\Sigma} \\
&\quad \times \Big([\,\mathbf{F},\, (\mathbf{G}\mathbf{V}_1 + \mathbf{H}\mathbf{V}_2)\mathbf{T}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}\,][\,\mathbf{T}\widehat{\mathbf{X}},\, \mathbf{T}\mathbf{V}\mathbf{T}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}\,]^{+}\mathbf{T}\widetilde{\mathbf{X}} - \mathbf{J}\Big)'. \tag{2.11}
\end{aligned}$$

(f) *The BLUP of* $\boldsymbol{\phi}$ *can be decomposed as the sum*

$$\mathrm{BLUP}_{\mathscr{N}}(\boldsymbol{\phi}) = \mathrm{BLUE}_{\mathscr{N}}(\mathbf{F}\boldsymbol{\alpha}) + \mathrm{BLUP}_{\mathscr{N}}(\mathbf{G}\boldsymbol{\gamma}) + \mathrm{BLUP}_{\mathscr{N}}(\mathbf{H}\boldsymbol{\varepsilon}), \tag{2.12}$$

and the following formulas for covariance matrices hold

$$Cov\{\mathrm{BLUE}_{\mathscr{N}}(\mathbf{F}\boldsymbol{\alpha}),\, \mathrm{BLUP}_{\mathscr{N}}(\mathbf{G}\boldsymbol{\gamma} + \mathbf{H}\boldsymbol{\varepsilon})\} = \mathbf{0}, \tag{2.13}$$

$$D[\mathrm{BLUP}_{\mathscr{N}}(\boldsymbol{\phi})] = D[\mathrm{BLUE}_{\mathscr{N}}(\mathbf{F}\boldsymbol{\alpha})] + D[\mathrm{BLUP}_{\mathscr{N}}(\mathbf{G}\boldsymbol{\gamma} + \mathbf{H}\boldsymbol{\varepsilon})]. \tag{2.14}$$

(g) $\mathrm{BLUP}_{\mathscr{N}}(\mathbf{P}\boldsymbol{\phi}) = \mathbf{P}\mathrm{BLUP}_{\mathscr{N}}(\boldsymbol{\phi})$ *holds for any matrix* $\mathbf{P} \in \mathbb{R}^{t\times s}$.

By setting $\mathbf{T} = \mathbf{I}_n$ in Theorem 2.2, we obtain the following known result in [37].

Corollary 2.3. *Let $\mathscr{M}$ be as given in* (1.1), $\mathbf{V}$, $\mathbf{V}_1$, *and* $\mathbf{V}_2$ *be as given in* (1.6), (1.8), *and* (1.10), *and denote The vector* $\boldsymbol{\phi}$ *in* (1.30) *is predictable by* $\mathbf{y}$ *in* (1.1) *if and only if*

$$\mathscr{R}(\widehat{\mathbf{X}}') \supseteq \mathscr{R}(\mathbf{F}'). \tag{2.15}$$

In this case, the following matrix equation

$$\mathbf{L}_1[\widehat{\mathbf{X}}, \mathbf{V}\widehat{\mathbf{X}}^{\perp}] = [\mathbf{F}, (\mathbf{G}\mathbf{V}_1 + \mathbf{H}\mathbf{V}_2)\widehat{\mathbf{X}}^{\perp}] \tag{2.16}$$

is consistent, and the general solution of $\mathbf{L}_1$ *and the corresponding BLUP of* $\boldsymbol{\phi}$ *are given by*

$$\mathrm{BLUP}_{\mathscr{M}}(\boldsymbol{\phi}) = \mathbf{L}_1\mathbf{y} = \left([\mathbf{F}, (\mathbf{G}\mathbf{V}_1 + \mathbf{H}\mathbf{V}_2)\widehat{\mathbf{X}}^{\perp}][\widehat{\mathbf{X}}, \mathbf{V}\widehat{\mathbf{X}}^{\perp}]^{+} + \mathbf{U}[\widehat{\mathbf{X}}, \mathbf{V}\widehat{\mathbf{X}}^{\perp}]^{\perp}\right)\mathbf{y}, \tag{2.17}$$

where $\mathbf{U} \in \mathbb{R}^{s\times n}$ *is arbitrary. Further, the following results hold.*

(a) $r[\widehat{\mathbf{X}}, \mathbf{V}\widehat{\mathbf{X}}^{\perp}] = r[\widehat{\mathbf{X}}, \mathbf{V}]$, $\mathscr{R}[\widehat{\mathbf{X}}, \mathbf{V}\widehat{\mathbf{X}}^{\perp}] = \mathscr{R}[\widehat{\mathbf{X}}, \mathbf{V}]$, *and* $\mathscr{R}(\widehat{\mathbf{X}}) \cap \mathscr{R}(\mathbf{V}\widehat{\mathbf{X}}^{\perp}) = \{\mathbf{0}\}$.

(b) $\mathbf{L}_1$ *is unique if and only if* $r[\widehat{\mathbf{X}}, \mathbf{V}] = n$.

(c) $\mathrm{BLUP}_{\mathscr{M}}(\boldsymbol{\phi})$ *is unique if and only if* $\mathbf{y} \in \mathscr{R}[\widehat{\mathbf{X}}, \mathbf{V}]$ *with probability 1, namely,* (1.1) *is consistent.*

(d) *The following covariance matrix equalities hold*

$$D[\mathrm{BLUP}_{\mathscr{M}}(\boldsymbol{\phi})] = [\mathbf{F}, (\mathbf{G}\mathbf{V}_1 + \mathbf{H}\mathbf{V}_2)\widehat{\mathbf{X}}^{\perp}][\widehat{\mathbf{X}}, \mathbf{V}\widehat{\mathbf{X}}^{\perp}]^{+}\mathbf{V}([\mathbf{F}, (\mathbf{G}\mathbf{V}_1 + \mathbf{H}\mathbf{V}_2)\widehat{\mathbf{X}}^{\perp}][\widehat{\mathbf{X}}, \mathbf{V}\widehat{\mathbf{X}}^{\perp}]^{+})', \tag{2.18}$$

$$Cov\{\mathrm{BLUP}_{\mathscr{M}}(\boldsymbol{\phi}), \boldsymbol{\phi}\} = [\mathbf{F}, (\mathbf{G}\mathbf{V}_1 + \mathbf{H}\mathbf{V}_2)\widehat{\mathbf{X}}^{\perp}][\widehat{\mathbf{X}}, \mathbf{V}\widehat{\mathbf{X}}^{\perp}]^{+}(\mathbf{G}\mathbf{V}_1 + \mathbf{H}\mathbf{V}_2)', \tag{2.19}$$

and

$$\begin{aligned}&D(\boldsymbol{\phi}) - D[\mathrm{BLUP}_{\mathscr{M}}(\boldsymbol{\phi})]\\&= \mathbf{J}\boldsymbol{\Sigma}\mathbf{J}' - [\mathbf{F}, (\mathbf{G}\mathbf{V}_1 + \mathbf{H}\mathbf{V}_2)\widehat{\mathbf{X}}^{\perp}][\widehat{\mathbf{X}}, \mathbf{V}\widehat{\mathbf{X}}^{\perp}]^{+}\mathbf{V}([\mathbf{F}, (\mathbf{G}\mathbf{V}_1 + \mathbf{H}\mathbf{V}_2)\widehat{\mathbf{X}}^{\perp}][\widehat{\mathbf{X}}, \mathbf{V}\widehat{\mathbf{X}}^{\perp}]^{+})',\end{aligned} \tag{2.20}$$

$$\begin{aligned}&D[\boldsymbol{\phi} - \mathrm{BLUP}_{\mathscr{M}}(\boldsymbol{\phi})]\\&= \left([\mathbf{F}, (\mathbf{G}\mathbf{V}_1 + \mathbf{H}\mathbf{V}_2)\widehat{\mathbf{X}}^{\perp}][\widehat{\mathbf{X}}, \mathbf{V}\widehat{\mathbf{X}}^{\perp}]^{+}\widetilde{\mathbf{X}} - \mathbf{J}\right)\boldsymbol{\Sigma}\left([\mathbf{F}, (\mathbf{G}\mathbf{V}_1 + \mathbf{H}\mathbf{V}_2)\widehat{\mathbf{X}}^{\perp}][\widehat{\mathbf{X}}, \mathbf{V}\widehat{\mathbf{X}}^{\perp}]^{+}\widetilde{\mathbf{X}} - \mathbf{J}\right)'.\end{aligned} \tag{2.21}$$

(e) *The BLUP of* $\boldsymbol{\phi}$ *can be decomposed as the sum*

$$\mathrm{BLUP}_{\mathscr{M}}(\boldsymbol{\phi}) = \mathrm{BLUE}_{\mathscr{M}}(\mathbf{F}\boldsymbol{\alpha}) + \mathrm{BLUP}_{\mathscr{M}}(\mathbf{G}\boldsymbol{\gamma}) + \mathrm{BLUP}_{\mathscr{M}}(\mathbf{H}\boldsymbol{\varepsilon}), \tag{2.22}$$

and the following formulas for covariance matrices hold

$$Cov\{\mathrm{BLUE}_{\mathscr{M}}(\mathbf{F}\boldsymbol{\alpha}), \mathrm{BLUP}_{\mathscr{M}}(\mathbf{G}\boldsymbol{\gamma} + \mathbf{H}\boldsymbol{\varepsilon})\} = \mathbf{0}, \tag{2.23}$$

$$D[\mathrm{BLUP}_{\mathscr{M}}(\boldsymbol{\phi})] = D[\mathrm{BLUE}_{\mathscr{M}}(\mathbf{F}\boldsymbol{\alpha})] + D[\mathrm{BLUP}_{\mathscr{M}}(\mathbf{G}\boldsymbol{\gamma} + \mathbf{H}\boldsymbol{\varepsilon})]. \tag{2.24}$$

(f) $\mathrm{BLUP}_{\mathscr{M}}(\mathbf{P}\boldsymbol{\phi}) = \mathbf{P}\mathrm{BLUP}_{\mathscr{M}}(\boldsymbol{\phi})$ *holds for any matrix* $\mathbf{P} \in \mathbb{R}^{t\times s}$.

Theorem 2.2 and Corollary 2.3 show that the BLUPs of all unknown parameters in LRMs can jointly be determined by certain linear matrix equations composed by all given matrices in the models. So that it is easy to present a simple yet general algebraic treatment of the BLUPs of all unknown parameters under general LRMs via basic linear matrix equations, while the BLUPs have a large number of properties that are technically convenient from the analytical solutions of the matrix equations. The whole results in Theorem 2.2 and Corollary 2.3 in fact provide a unified theory about BLUPs of all unknown parameters and their fundamental properties under general LRMs.

3   Equalities for BLUPs under original and transformed models

In order to characterize equalities between two random vectors, we need the following definition.

Definition 3.1. *Let* $\mathbf{y}$ *be a random vector.*

(a) *The equality* $\mathbf{G}_1\mathbf{y} = \mathbf{G}_2\mathbf{y}$ *is said to hold definitely if* $\mathbf{G}_1 = \mathbf{G}_2$.

(b) *The equality* $\mathbf{G}_1\mathbf{y} = \mathbf{G}_2\mathbf{y}$ *is said to hold with probability 1 if both* $E(\mathbf{G}_1\mathbf{y} - \mathbf{G}_2\mathbf{y}) = \mathbf{0}$ *and* $D(\mathbf{G}_1\mathbf{y} - \mathbf{G}_2\mathbf{y}) = \mathbf{0}$ *hold.*

(c) $\mathbf{G}_1\mathbf{y}$ and $\mathbf{G}_2\mathbf{y}$ are said to have the same expectation and dispersion matrix if both $E(\mathbf{G}_1\mathbf{y}) = E(\mathbf{G}_2\mathbf{y})$ and $D(\mathbf{G}_1\mathbf{y}) = D(\mathbf{G}_2\mathbf{y})$ hold.

The three types of equality are not necessarily equivalent since they are defined from different criteria. Because of the non-commutativity of matrix algebra, it is usually a challenging task to characterize the equalities of estimators under general linear models. Due to recent some developments of tools in characterizing matrix equalities, we now be able to derive satisfactory conclusions on this kind of issues.

The establishments of the decomposition equalities of BLUPs in the previous sections enable us to compare predictors under various conditions. Through comparing the formulas of the BLUPs of the same unknown parameters under different models, it is easy to propose some possible equalities, and then seek identifying conditions for them to hold. Note that the pair of models in (1.1) and (1.3) are not the same, the two BLUPs' equations in (2.5) and (2.16) of the same set of predictable functions have different structures, and their solutions are not necessarily the same. In this case, possible equalities for the BLUPs under two different LRMs were recently considered in [13, 14, 15]. This work in fact can convert to compare coefficient matrices of the BLUPs (i.e., solutions of matrix equations) under two LRMs.

Note that (2.4) implies (2.15). Hence,

$$\boldsymbol{\phi} \text{ in } (1.30) \text{ is predictable under } (1.3) \Longrightarrow \boldsymbol{\phi} \text{ in } (1.30) \text{ is predictable under } (1.1).$$

In this case, the BLUPs of $\boldsymbol{\phi}$ under (1.1) and (1.3) can be written as (2.7) and (2.17), respectively. However, these two types of predictor are not necessarily the same. In this section, we study relations between the BLUPs of $\boldsymbol{\phi}$ under the two LRMs, and present necessary and sufficient conditions for the BLUPs to be equal.

Theorem 3.2. *Let $\mathscr{M}$ and $\mathscr{N}$ be as given in (1.1) and (1.3), respectively, and assume that $\boldsymbol{\phi}$ in (1.30) is predictable under (1.3). Then, the following statements are equivalent:*

(a) $\mathrm{BLUP}_{\mathscr{M}}(\boldsymbol{\phi}) = \mathrm{BLUP}_{\mathscr{N}}(\boldsymbol{\phi})$ *holds definitely .*

(b) $\mathrm{BLUP}_{\mathscr{M}}(\boldsymbol{\phi}) = \mathrm{BLUP}_{\mathscr{N}}(\boldsymbol{\phi})$ *holds with probability 1.*

(c) $r\begin{bmatrix} \mathbf{T}\widehat{\mathbf{X}} & Cov\{\mathbf{Ty}, \mathbf{y}\} \\ \mathbf{0} & \widehat{\mathbf{X}}' \\ \mathbf{F} & Cov\{\boldsymbol{\phi}, \mathbf{y}\} \end{bmatrix} = r\begin{bmatrix} \mathbf{T}\widehat{\mathbf{X}} & Cov\{\mathbf{Ty}, \mathbf{y}\} \\ \mathbf{0} & \widehat{\mathbf{X}}' \end{bmatrix}.$

(d) $r\begin{bmatrix} \mathbf{T}\widehat{\mathbf{X}} & Cov\{\mathbf{Ty}, \widehat{\mathbf{X}}^{\perp}\mathbf{y}\} \\ \mathbf{F} & Cov\{\boldsymbol{\phi}, \widehat{\mathbf{X}}^{\perp}\mathbf{y}\} \end{bmatrix} = r[\,\mathbf{T}\widehat{\mathbf{X}}, Cov\{\mathbf{Ty}, \widehat{\mathbf{X}}^{\perp}\mathbf{y}\}\,].$

(e) $\mathscr{R}([\,\mathbf{T}\widehat{\mathbf{X}}, Cov\{\mathbf{Ty}, \widehat{\mathbf{X}}^{\perp}\mathbf{y}\}\,]') \supseteq \mathscr{R}([\,\mathbf{F}, Cov\{\boldsymbol{\phi}, \widehat{\mathbf{X}}^{\perp}\mathbf{y}\}\,]').$

Corollary 3.3. *Let $\mathscr{M}$ and $\mathscr{N}$ be as given in (1.1) and (1.3), respectively. Then, the following results hold.*

(a) $\mathbf{X}\boldsymbol{\beta}$ *is predictable under (1.3) if and only if $r(\mathbf{T}\widehat{\mathbf{X}}) = r\widehat{\mathbf{X}}$. In this case, the following statements are equivalent:*

    (i) $\mathrm{BLUP}_{\mathscr{M}}(\mathbf{X}\boldsymbol{\beta}) = \mathrm{BLUP}_{\mathscr{N}}(\mathbf{X}\boldsymbol{\beta})$ *holds definitely (with probability 1).*

    (ii) $r\begin{bmatrix} \mathbf{T}\widehat{\mathbf{X}} & Cov\{\mathbf{Ty}, \mathbf{y}\} \\ \mathbf{0} & \widehat{\mathbf{X}}' \\ \widehat{\mathbf{X}} & Cov\{\mathbf{X}\boldsymbol{\beta}, \mathbf{y}\} \end{bmatrix} = r\begin{bmatrix} \mathbf{T}\widehat{\mathbf{X}} & Cov\{\mathbf{Ty}, \mathbf{y}\} \\ \mathbf{0} & \widehat{\mathbf{X}}' \end{bmatrix}.$

    (iii) $r\begin{bmatrix} \mathbf{T}\widehat{\mathbf{X}} & Cov\{\mathbf{Ty}, \widehat{\mathbf{X}}^{\perp}\mathbf{y}\} \\ \widehat{\mathbf{X}} & Cov\{\mathbf{X}\boldsymbol{\beta}, \widehat{\mathbf{X}}^{\perp}\mathbf{y}\} \end{bmatrix} = r[\,\mathbf{T}\widehat{\mathbf{X}}, Cov\{\mathbf{Ty}, \widehat{\mathbf{X}}^{\perp}\mathbf{y}\}\,].$

    (v) $\mathscr{R}([\,\mathbf{T}\widehat{\mathbf{X}}, Cov\{\mathbf{Ty}, \widehat{\mathbf{X}}^{\perp}\mathbf{y}\}\,]') \supseteq \mathscr{R}([\,\widehat{\mathbf{X}}, Cov\{\mathbf{X}\boldsymbol{\beta}, \widehat{\mathbf{X}}^{\perp}\mathbf{y}\}]').$

(b) $\widehat{\mathbf{X}}\boldsymbol{\alpha}$ *is estimable under (1.3) if and only if $r(\mathbf{T}\widehat{\mathbf{X}}) = r\widehat{\mathbf{X}}$. In this case, the following statements are equivalent:*

    (i) $\mathrm{BLUE}_{\mathscr{M}}(\widehat{\mathbf{X}}\boldsymbol{\alpha}) = \mathrm{BLUE}_{\mathscr{N}}(\widehat{\mathbf{X}}\boldsymbol{\alpha})$ *holds definitely (with probability 1).*

    (ii) $r\begin{bmatrix} Cov\{\mathbf{Ty}, \mathbf{y}\} \\ \widehat{\mathbf{X}}' \end{bmatrix} = r[\,\mathbf{T}\widehat{\mathbf{X}}, Cov\{\mathbf{Ty}, \widehat{\mathbf{X}}^{\perp}\mathbf{y}\}\,].$

    (iii) $r\begin{bmatrix} \mathbf{T}\widehat{\mathbf{X}} & Cov\{\mathbf{Ty}, \widehat{\mathbf{X}}^{\perp}\mathbf{y}\} \\ \widehat{\mathbf{X}} & \mathbf{0} \end{bmatrix} = r[\,\mathbf{T}\widehat{\mathbf{X}}, Cov\{\mathbf{Ty}, \widehat{\mathbf{X}}^{\perp}\mathbf{y}\}\,].$

    (iv) $\mathscr{R}([\,\mathbf{T}\widehat{\mathbf{X}}, Cov\{\mathbf{Ty}, \widehat{\mathbf{X}}^{\perp}\mathbf{y}\}\,]') \supseteq \mathscr{R}([\,\widehat{\mathbf{X}}, \mathbf{0}]').$

(c) $\mathbf{X\gamma}$ is always predictable under (1.1) and (1.3), and the following statements are equivalent:

    (i) $\mathrm{BLUP}_{\mathscr{M}}(\mathbf{X\gamma}) = \mathrm{BLUP}_{\mathscr{N}}(\mathbf{X\gamma})$ holds definitely (with probability 1).

    (ii) $r\begin{bmatrix} \mathbf{T\widehat{X}} & Cov\{\mathbf{Ty, y}\} \\ \mathbf{0} & \mathbf{\widehat{X}'} \\ \mathbf{0} & Cov\{\mathbf{X\gamma, y}\} \end{bmatrix} = r\begin{bmatrix} \mathbf{T\widehat{X}} & Cov\{\mathbf{Ty, y}\} \\ \mathbf{0} & \mathbf{\widehat{X}'} \end{bmatrix}.$

    (iii) $r\begin{bmatrix} \mathbf{T\widehat{X}} & Cov\{\mathbf{Ty, \widehat{X}^{\perp}y}\} \\ \mathbf{0} & Cov\{\mathbf{X\gamma, \widehat{X}^{\perp}y}\} \end{bmatrix} = r[\,\mathbf{T\widehat{X}},\, Cov\{\mathbf{Ty, \widehat{X}^{\perp}y}\}\,].$

    (iv) $\mathscr{R}([\,\mathbf{T\widehat{X}},\, Cov\{\mathbf{Ty, \widehat{X}^{\perp}y}\}\,]') \supseteq \mathscr{R}([\,\mathbf{0},\, Cov\{\mathbf{X\gamma, \widehat{X}^{\perp}y}\}\,]').$

    (v) $\mathscr{R}([Cov\{(\mathbf{T\widehat{X}})^{\perp}\mathbf{Ty}, \mathbf{\widehat{X}^{\perp}y}\}]') \supseteq \mathscr{R}([Cov\{\mathbf{X\gamma}, \mathbf{\widehat{X}^{\perp}y}\}]').$

(d) $\boldsymbol{\varepsilon}$ is always predictable under (1.1) and (1.3), and the following statements are equivalent:

    (i) $\mathrm{BLUP}_{\mathscr{M}}(\boldsymbol{\varepsilon}) = \mathrm{BLUP}_{\mathscr{N}}(\boldsymbol{\varepsilon})$ holds definitely (with probability 1).

    (ii) $r\begin{bmatrix} \mathbf{T\widehat{X}} & Cov\{\mathbf{Ty, y}\} \\ \mathbf{0} & \mathbf{\widehat{X}'} \\ \mathbf{0} & Cov\{\boldsymbol{\varepsilon}, \mathbf{y}\} \end{bmatrix} = r\begin{bmatrix} \mathbf{T\widehat{X}} & Cov\{\mathbf{Ty, y}\} \\ \mathbf{0} & \mathbf{\widehat{X}'} \end{bmatrix}.$

    (iii) $r\begin{bmatrix} \mathbf{T\widehat{X}} & Cov\{\mathbf{Ty, \widehat{X}^{\perp}y}\} \\ \mathbf{0} & Cov\{\boldsymbol{\varepsilon}, \mathbf{\widehat{X}^{\perp}y}\} \end{bmatrix} = r[\,\mathbf{T\widehat{X}},\, Cov\{\mathbf{Ty, \widehat{X}^{\perp}y}\}\,].$

    (iv) $\mathscr{R}([\,\mathbf{T\widehat{X}},\, Cov\{\mathbf{Ty, \widehat{X}^{\perp}y}\}\,]') \supseteq \mathscr{R}([\,\mathbf{0},\, Cov\{\boldsymbol{\varepsilon}, \mathbf{\widehat{X}^{\perp}y}\}\,]').$

    (v) $\mathscr{R}([Cov\{(\mathbf{T\widehat{X}})^{\perp}\mathbf{Ty}, \mathbf{\widehat{X}^{\perp}y}\}]') \supseteq \mathscr{R}([Cov\{\boldsymbol{\varepsilon}, \mathbf{\widehat{X}^{\perp}y}\}]').$

One of the main contributions of this paper is establishing the following decompositions of the transformed data vector $\mathbf{Ty}$ with respect to the BLUPs under (1.1) and (1.3).

Theorem 3.4. *Let $\mathscr{M}$ and $\mathscr{N}$ be as given in (1.1) and (1.3), respectively. Then, $\mathbf{TX\beta}$, $\mathbf{T\widehat{X}\alpha}$, $\mathbf{TX\gamma}$, and $\mathbf{T\varepsilon}$ are always predictable and estimable under (1.1) and (1.3), respectively, and they satisfy the following decomposition equalities*

$$\mathbf{Ty} = \mathrm{BLUE}_{\mathscr{M}}(\mathbf{T\widehat{X}\alpha}) + \mathrm{BLUP}_{\mathscr{M}}(\mathbf{TX\gamma}) + \mathrm{BLUP}_{\mathscr{M}}(\mathbf{T\varepsilon}) = \mathbf{P_1 y} + \mathbf{P_2 y} + \mathbf{P_3 y}, \tag{3.1}$$

$$\mathbf{Ty} = \mathrm{BLUE}_{\mathscr{N}}(\mathbf{T\widehat{X}\alpha}) + \mathrm{BLUP}_{\mathscr{N}}(\mathbf{TX\gamma}) + \mathrm{BLUP}_{\mathscr{N}}(\mathbf{T\varepsilon}) = \mathbf{Q_1 y} + \mathbf{Q_2 y} + \mathbf{Q_3 y}, \tag{3.2}$$

where

$$\mathbf{P_1} = [\,\mathbf{T\widehat{X}},\, \mathbf{0}\,][\,\mathbf{\widehat{X}},\, \mathbf{V\widehat{X}^{\perp}}\,]^{+} + \mathbf{TU}[\,\mathbf{\widehat{X}},\, \mathbf{V\widehat{X}^{\perp}}\,]^{\perp},$$

$$\mathbf{P_2} = [\,\mathbf{0},\, \mathbf{TXV_1\widehat{X}^{\perp}}\,][\,\mathbf{\widehat{X}},\, \mathbf{V\widehat{X}^{\perp}}\,]^{+} + \mathbf{TU}[\,\mathbf{\widehat{X}},\, \mathbf{V\widehat{X}^{\perp}}\,]^{\perp},$$

$$\mathbf{P_3} = [\,\mathbf{0},\, \mathbf{TV_2\widehat{X}^{\perp}}\,][\,\mathbf{\widehat{X}},\, \mathbf{V\widehat{X}^{\perp}}\,]^{+} + \mathbf{TU}[\,\mathbf{\widehat{X}},\, \mathbf{V\widehat{X}^{\perp}}\,]^{\perp},$$

$$\mathbf{Q_1} = [\,\mathbf{T\widehat{X}},\, \mathbf{0}\,][\,\mathbf{T\widehat{X}},\, \mathbf{TVT'(T\widehat{X})^{\perp}}\,]^{+}\mathbf{T} + \mathbf{U}[\,\mathbf{T\widehat{X}},\, \mathbf{TVT'(T\widehat{X})^{\perp}}\,]^{\perp}\mathbf{T},$$

$$\mathbf{Q_2} = [\,\mathbf{0},\, \mathbf{TXV_1T'(T\widehat{X})^{\perp}}\,][\,\mathbf{T\widehat{X}},\, \mathbf{TVT'(T\widehat{X})^{\perp}}\,]^{+}\mathbf{T} + \mathbf{U}[\,\mathbf{T\widehat{X}},\, \mathbf{TVT'(T\widehat{X})^{\perp}}\,]^{\perp}\mathbf{T},$$

$$\mathbf{Q_3} = [\,\mathbf{0},\, \mathbf{TV_2T'(T\widehat{X})^{\perp}}\,][\,\mathbf{T\widehat{X}},\, \mathbf{TVT'(T\widehat{X})^{\perp}}\,]^{+}\mathbf{T} + \mathbf{U}[\,\mathbf{T\widehat{X}},\, \mathbf{TVT'(T\widehat{X})^{\perp}}\,]^{\perp}\mathbf{T},$$

$(\mathbf{Ty})'(\mathbf{Ty})$ *can be decomposed as*

$$(\mathbf{Ty})'(\mathbf{Ty}) = \mathbf{y'T'P_1 y} + \mathbf{y'T'P_2 y} + \mathbf{y'T'P_3 y}, \tag{3.3}$$

$$(\mathbf{Ty})'(\mathbf{Ty}) = \mathbf{y'T'Q_1 y} + \mathbf{y'T'Q_2 y} + \mathbf{y'T'Q_3 y}, \tag{3.4}$$

the dispersion matrix of $\mathbf{Ty}$ satisfies

$$D(\mathbf{Ty}) = D[\mathrm{BLUE}_{\mathscr{M}}(\mathbf{T\widehat{X}\alpha})] + D[\,\mathrm{BLUP}_{\mathscr{M}}(\mathbf{TX\gamma}) + \mathrm{BLUP}_{\mathscr{M}}(\mathbf{T\varepsilon})\,], \tag{3.5}$$

$$D(\mathbf{Ty}) = D[\mathrm{BLUE}_{\mathscr{N}}(\mathbf{T\widehat{X}\alpha})] + D[\,\mathrm{BLUP}_{\mathscr{N}}(\mathbf{TX\gamma}) + \mathrm{BLUP}_{\mathscr{N}}(\mathbf{T\varepsilon})\,]. \tag{3.6}$$

Eqs. (3.1)–(3.6) hold unconditionally, and can be regarded as certain types of built-in restriction to BLUPs under LRMs. These equalities show that BLUPs/BLUEs of all unknown parameters under linear models are closely linked one another, and can also be called the expansion formulas of the transformed response vector $\mathbf{Ty}$ with respect to the BLUPs/BLUEs of the components in the two LRMs $\mathscr{M}$ and $\mathscr{N}$. The establishments of these expansion formulas sufficiently demonstrates the exclusive uses of mathematical tools in statistics, in particular, the orthodox role of BLUPs/BLUEs in statistical inference of LRMs, as claimed in [23]. It seems

that only BLUPs have such nice decompositions from both mathematical and statistical points of view. Some previous discussions on built-in restrictions to BLUPs can be found, e.g., in [2, 25, 34].

Note that the terms on the right-hand sides of (3.1) and (3.2) are not necessarily the same. By comparing these terms, we obtain the following result.

Theorem 3.5. *Let $\mathscr{M}$ and $\mathscr{N}$ be as given in* (1.1) *and* (1.3), *respectively. Then, the following results hold.*

(a) *The following statements are equivalent*:

   (i) $\mathrm{BLUP}_{\mathscr{M}}(\mathbf{TX}\boldsymbol{\beta}) = \mathrm{BLUP}_{\mathscr{N}}(\mathbf{TX}\boldsymbol{\beta})$ *holds definitely* (*with probability* 1).

   (ii) $\mathrm{BLUP}_{\mathscr{M}}(\mathbf{T}\boldsymbol{\varepsilon}) = \mathrm{BLUP}_{\mathscr{N}}(\mathbf{T}\boldsymbol{\varepsilon})$ *holds definitely* (*with probability* 1).

   (iii) $r\begin{bmatrix} \mathbf{T}\widehat{\mathbf{X}} & Cov\{\mathbf{Ty}, \mathbf{y}\} \\ \mathbf{0} & \widehat{\mathbf{X}}' \\ \mathbf{T}\widehat{\mathbf{X}} & Cov\{\mathbf{TX}\boldsymbol{\beta}, \mathbf{y}\} \end{bmatrix} = r\begin{bmatrix} \mathbf{T}\widehat{\mathbf{X}} & Cov\{\mathbf{Ty}, \mathbf{y}\} \\ \mathbf{0} & \widehat{\mathbf{X}}' \end{bmatrix}.$

   (iv) $r\begin{bmatrix} \mathbf{T}\widehat{\mathbf{X}} & Cov\{\mathbf{Ty}, \widehat{\mathbf{X}}^{\perp}\mathbf{y}\} \\ \mathbf{T}\widehat{\mathbf{X}} & Cov\{\mathbf{TX}\boldsymbol{\beta}, \widehat{\mathbf{X}}^{\perp}\mathbf{y}\} \end{bmatrix} = r[\,\mathbf{T}\widehat{\mathbf{X}},\, Cov\{\mathbf{Ty}, \widehat{\mathbf{X}}^{\perp}\mathbf{y}\}\,].$

   (v) $r\begin{bmatrix} \mathbf{T}\widehat{\mathbf{X}} & Cov\{\mathbf{Ty}, \widehat{\mathbf{X}}^{\perp}\mathbf{y}\} \\ \mathbf{0} & Cov\{\mathbf{T}\boldsymbol{\varepsilon}, \widehat{\mathbf{X}}^{\perp}\mathbf{y}\} \end{bmatrix} = r[\,\mathbf{T}\widehat{\mathbf{X}},\, Cov\{\mathbf{Ty}, \widehat{\mathbf{X}}^{\perp}\mathbf{y}\}\,].$

   (vi) $\mathscr{R}([\,\mathbf{T}\widehat{\mathbf{X}},\, Cov\{\mathbf{Ty}, \widehat{\mathbf{X}}^{\perp}\mathbf{y}\}\,]') \supseteq \mathscr{R}([\,\mathbf{T}\widehat{\mathbf{X}},\, Cov\{\mathbf{TX}\boldsymbol{\beta}, \widehat{\mathbf{X}}^{\perp}\mathbf{y}\}\,]').$

   (vii) $\mathscr{R}([\,\mathbf{T}\widehat{\mathbf{X}},\, Cov\{\mathbf{Ty}, \widehat{\mathbf{X}}^{\perp}\mathbf{y}\}\,]') \supseteq \mathscr{R}([\,\mathbf{0},\, Cov\{\mathbf{T}\boldsymbol{\varepsilon}, \widehat{\mathbf{X}}^{\perp}\mathbf{y}\}\,]').$

   (viii) $\mathscr{R}([\,Cov\{(\mathbf{T}\widehat{\mathbf{X}})^{\perp}\mathbf{Ty}, \widehat{\mathbf{X}}^{\perp}\mathbf{y}\}\,]') \supseteq \mathscr{R}([\,Cov\{\mathbf{T}\boldsymbol{\varepsilon}, \widehat{\mathbf{X}}^{\perp}\mathbf{y}\}\,]').$

(b) *The following statements are equivalent*:

   (i) $\mathrm{BLUE}_{\mathscr{M}}(\mathbf{T}\widehat{\mathbf{X}}\boldsymbol{\alpha}) = \mathrm{BLUE}_{\mathscr{N}}(\mathbf{T}\widehat{\mathbf{X}}\boldsymbol{\alpha})$ *holds definitely* (*with probability* 1).

   (ii) $r\begin{bmatrix} Cov\{\mathbf{Ty}, \mathbf{y}\} \\ \widehat{\mathbf{X}}' \end{bmatrix} + r(\mathbf{T}\widehat{\mathbf{X}}) = r[\,\mathbf{T}\widehat{\mathbf{X}},\, Cov\{\mathbf{Ty}, \widehat{\mathbf{X}}^{\perp}\mathbf{y}\}\,] + r(\widehat{\mathbf{X}}).$

   (iii) $r(Cov\{\mathbf{Ty}, \widehat{\mathbf{X}}^{\perp}\mathbf{y}\}) = r(Cov\{(\mathbf{T}\widehat{\mathbf{X}})^{\perp}\mathbf{Ty}, \widehat{\mathbf{X}}^{\perp}\mathbf{y}\}).$

(c) *The following statements are equivalent*:

   (i) $\mathrm{BLUP}_{\mathscr{M}}(\mathbf{TX}\boldsymbol{\gamma}) = \mathrm{BLUP}_{\mathscr{N}}(\mathbf{TX}\boldsymbol{\gamma})$ *holds definitely* (*with probability* 1).

   (ii) $r\begin{bmatrix} \mathbf{T}\widehat{\mathbf{X}} & Cov\{\mathbf{Ty}, \mathbf{y}\} \\ \mathbf{0} & \widehat{\mathbf{X}}' \\ \mathbf{0} & Cov\{\mathbf{TX}\boldsymbol{\gamma}, \mathbf{y}\} \end{bmatrix} = r\begin{bmatrix} \mathbf{T}\widehat{\mathbf{X}} & Cov\{\mathbf{Ty}, \mathbf{y}\} \\ \mathbf{0} & \widehat{\mathbf{X}}' \end{bmatrix}.$

   (iii) $r\begin{bmatrix} \mathbf{T}\widehat{\mathbf{X}} & Cov\{\mathbf{Ty}, \widehat{\mathbf{X}}^{\perp}\mathbf{y}\} \\ \mathbf{0} & Cov\{\mathbf{TX}\boldsymbol{\gamma}, \widehat{\mathbf{X}}^{\perp}\mathbf{y}\} \end{bmatrix} = r[\,\mathbf{T}\widehat{\mathbf{X}},\, Cov\{\mathbf{Ty}, \widehat{\mathbf{X}}^{\perp}\mathbf{y}\}\,].$

   (iv) $\mathscr{R}([\,\mathbf{T}\widehat{\mathbf{X}},\, Cov\{\mathbf{Ty}, \widehat{\mathbf{X}}^{\perp}\mathbf{y}\}\,]') \supseteq \mathscr{R}([\,\mathbf{0},\, Cov\{\mathbf{TX}\boldsymbol{\gamma}, \widehat{\mathbf{X}}^{\perp}\mathbf{y}\}\,]').$

   (v) $\mathscr{R}([\,Cov\{(\mathbf{T}\widehat{\mathbf{X}})^{\perp}\mathbf{Ty}, \widehat{\mathbf{X}}^{\perp}\mathbf{y}\}\,]') \supseteq \mathscr{R}([\,Cov\{\mathbf{TX}\boldsymbol{\gamma}, \widehat{\mathbf{X}}^{\perp}\mathbf{y}\}\,]').$

4   Conclusions

We presented a new investigation to the specified LRM in (1.1) and its transformed models via some novel algebraic tools in matrix theory, and obtained a variety of analytical conclusions on statistical inference of LRMs under most general assumptions. In particular, we established exact linear matrix equations associated with the BLUPs of the general vector of all unknown parameters in (1.1) and (1.3), and obtained many explicit formulas associated with the BLUPs by solving the linear matrix equations. These formulas and corresponding are easy to understand and use in dealing with various prediction/estimation inference problems on LRMs. Hence, it is expected more valuable results on statistical inference of LRMs can derived from these analytical formulas and conclusions. In particular, applying Sections 2 and 3 to the specified transformed models in (d)–(h) in Section 1 will produce a variety of concrete results on the statistical inference of the models.

Appendix

Statistical methods in many areas of application often involve mathematical computations with vectors and matrices. In particular, formulas and algebraic tricks for handling matrices in linear algebra and matrix theory play important roles in the derivation of predictors/estimators and characterizations of their properties under linear regression models. Recall that the rank of matrix is a conceptual foundation in linear algebra and matrix theory, which is the most significant finite nonnegative integer in reflecting intrinsic properties of matrices. The mathematical prerequisites for understanding the rank of matrix are minimal and do not go beyond elementary linear algebra. It has long history to establish rank formulas for block matrices and use the formulas in statistical inference, and a pioneer work in this aspect can be found in [11]. The intriguing connections between generalized inverses of matrices and rank formulas of matrices were recognized in 1970s, and a seminal work on rank formulas for matrices and their generalized inverses was presented in [24]. In order to establish and characterize various possible equalities for predictors/estimators under LRMs, and to simplify various matrix equalities composed by the Moore–Penrose inverses of matrices, we need some fundamental matrix rank formulas for matrices and their Moore–Penrose generalized inverses to make the paper self-contained.

Lemma A.1 ([24, 36])**.** *Let* $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{B} \in \mathbb{R}^{m \times k}$, *and* $\mathbf{C} \in \mathbb{R}^{l \times n}$. *Then*

$$r[\mathbf{A}, \mathbf{B}] = r(\mathbf{A}) + r(\mathbf{E_A B}) = r(\mathbf{B}) + r(\mathbf{E_B A}), \tag{A.1}$$

$$r\begin{bmatrix} \mathbf{A} \\ \mathbf{C} \end{bmatrix} = r(\mathbf{A}) + r(\mathbf{CF_A}) = r(\mathbf{C}) + r(\mathbf{AF_C}), \tag{A.2}$$

$$r\begin{bmatrix} \mathbf{AA}' & \mathbf{B} \\ \mathbf{B}' & \mathbf{0} \end{bmatrix} = r[\mathbf{A}, \mathbf{B}] + r(\mathbf{B}). \tag{A.3}$$

If $\mathscr{R}(\mathbf{A}'_1) \subseteq \mathscr{R}(\mathbf{B}'_1)$, $\mathscr{R}(\mathbf{A}_2) \subseteq \mathscr{R}(\mathbf{B}_1)$, $\mathscr{R}(\mathbf{A}'_2) \subseteq \mathscr{R}(\mathbf{B}'_2)$ *and* $\mathscr{R}(\mathbf{A}_3) \subseteq \mathscr{R}(\mathbf{B}_2)$, *then*

$$r(\mathbf{A}_1 \mathbf{B}_1^+ \mathbf{A}_2) = r\begin{bmatrix} \mathbf{B}_1 & \mathbf{A}_2 \\ \mathbf{A}_1 & \mathbf{0} \end{bmatrix} - r(\mathbf{B}_1), \tag{A.4}$$

$$r(\mathbf{A}_1 \mathbf{B}_1^+ \mathbf{A}_2 \mathbf{B}_2^+ \mathbf{A}_3) = r\begin{bmatrix} \mathbf{0} & \mathbf{B}_2 & \mathbf{A}_3 \\ \mathbf{B}_1 & \mathbf{A}_2 & \mathbf{0} \\ \mathbf{A}_1 & \mathbf{0} & \mathbf{0} \end{bmatrix} - r(\mathbf{B}_1) - r(\mathbf{B}_2). \tag{A.5}$$

In addition, the following results hold.

(a) $r[\mathbf{A}, \mathbf{B}] = r(\mathbf{A}) \Leftrightarrow \mathscr{R}(\mathbf{B}) \subseteq \mathscr{R}(\mathbf{A}) \Leftrightarrow \mathbf{AA}^+ \mathbf{B} = \mathbf{B} \Leftrightarrow \mathbf{E_A B} = \mathbf{0}$.

(b) $r\begin{bmatrix} \mathbf{A} \\ \mathbf{C} \end{bmatrix} = r(\mathbf{A}) \Leftrightarrow \mathscr{R}(\mathbf{C}') \subseteq \mathscr{R}(\mathbf{A}') \Leftrightarrow \mathbf{CA}^+ \mathbf{A} = \mathbf{C} \Leftrightarrow \mathbf{CF_A} = \mathbf{0}$.

(c) $r[\mathbf{A}, \mathbf{B}] = r(\mathbf{A}) + r(\mathbf{B}) \Leftrightarrow \mathscr{R}(\mathbf{A}) \cap \mathscr{R}(\mathbf{B}) = \{\mathbf{0}\} \Leftrightarrow \mathscr{R}[(\mathbf{E_A B})'] = \mathscr{R}(\mathbf{B}') \Leftrightarrow \mathscr{R}[(\mathbf{E_B A})'] = \mathscr{R}(\mathbf{A}')$.

(d) $r\begin{bmatrix} \mathbf{A} \\ \mathbf{C} \end{bmatrix} = r(\mathbf{A}) + r(\mathbf{C}) \Leftrightarrow \mathscr{R}(\mathbf{A}') \cap \mathscr{R}(\mathbf{C}') = \{\mathbf{0}\} \Leftrightarrow \mathscr{R}(\mathbf{CF_A}) = \mathscr{R}(\mathbf{C}) \Leftrightarrow \mathscr{R}(\mathbf{AF_C}) = \mathscr{R}(\mathbf{A})$.

Lemma A.2 ([29])**.** *The linear matrix equation* $\mathbf{AX} = \mathbf{B}$ *is consistent if and only if* $r[\mathbf{A}, \mathbf{B}] = r(\mathbf{A})$, *or equivalently,* $\mathbf{AA}^+ \mathbf{B} = \mathbf{B}$. *In this case, the general solution of the equation can be written in the following parametric form* $\mathbf{X} = \mathbf{A}^+ \mathbf{B} + (\mathbf{I} - \mathbf{A}^+ \mathbf{A})\mathbf{U}$, *where* $\mathbf{U}$ *is an arbitrary matrix.*

In order to directly solve the matrix minimization problem in (1.38), we need the following result on a constrained quadratic matrix-valued function minimization problem, which was proved, utilized and extended in [37, 38, 39].

Lemma A.3. *Let*

$$f(\mathbf{L}) = (\mathbf{LC} + \mathbf{D})\mathbf{M}(\mathbf{LC} + \mathbf{D})' \quad s.t. \quad \mathbf{LA} = \mathbf{B},$$

where $\mathbf{A} \in \mathbb{R}^{p \times q}$, $\mathbf{B} \in \mathbb{R}^{n \times q}$, $\mathbf{C} \in \mathbb{R}^{p \times m}$ *and* $\mathbf{D} \in \mathbb{R}^{n \times m}$ *are given,* $\mathbf{M} \in \mathbb{R}^{m \times m}$ *is nnd, and the matrix equation* $\mathbf{LA} = \mathbf{B}$ *is consistent. Then, there always exists a solution* $\mathbf{L}_0$ *of* $\mathbf{L}_0 \mathbf{A} = \mathbf{B}$ *such that*

$$f(\mathbf{L}) \succcurlyeq f(\mathbf{L}_0)$$

holds for all solutions of $\mathbf{LA} = \mathbf{B}$. *In this case, the matrix* $\mathbf{L}_0$ *satisfying the above inequality is determined by the following consistent matrix equation*

$$\mathbf{L}_0 \left[\mathbf{A}, \mathbf{CMC}' \mathbf{A}^\perp \right] = \left[\mathbf{B}, -\mathbf{DMC}' \mathbf{A}^\perp \right].$$

In this case, the general expression of $\mathbf{L}_0$ and the corresponding $f(\mathbf{L}_0)$ and $f(\mathbf{L})$ are given by

$$\mathbf{L}_0 = \underset{\mathbf{LA}=\mathbf{B}}{\arg\min}\, f(\mathbf{L}) = [\mathbf{B},\, -\mathbf{DMC}'\mathbf{A}^{\perp}]\,[\mathbf{A},\, \mathbf{CMC}'\mathbf{A}^{\perp}]^{+} + \mathbf{U}\,[\mathbf{A},\, \mathbf{CMC}']^{\perp},$$

$$f(\mathbf{L}_0) = \underset{\mathbf{LA}=\mathbf{B}}{\min}\, f(\mathbf{L}) = \mathbf{KMK}' - \mathbf{KMC}'\mathbf{TCMK}',$$

$$f(\mathbf{L}) = f(\mathbf{L}_0) + \left(\mathbf{LCMC}'\mathbf{A}^{\perp} + \mathbf{DMC}'\mathbf{A}^{\perp}\right)\mathbf{T}\left(\mathbf{LCMC}'\mathbf{A}^{\perp} + \mathbf{DMC}'\mathbf{A}^{\perp}\right)',$$

where $\mathbf{K} = \mathbf{BA}^{+}\mathbf{C} + \mathbf{D}$, $\mathbf{T} = \left(\mathbf{A}^{\perp}\mathbf{CMC}'\mathbf{A}^{\perp}\right)^{+}$, and $\mathbf{U} \in \mathbb{R}^{n \times p}$ is arbitrary.

Proof of Theorem 2.1 It is obvious that $E\left(\mathbf{L}\widehat{\mathbf{y}} - \boldsymbol{\phi}\right) = \mathbf{0} \Leftrightarrow \mathbf{L}\widehat{\mathbf{X}}\mathbf{A}\boldsymbol{\alpha} - \mathbf{F}\boldsymbol{\alpha} = \mathbf{0}$ for all $\boldsymbol{\alpha} \Leftrightarrow \mathbf{L}\widehat{\mathbf{X}}\mathbf{A} = \mathbf{F}$. From Lemma A.2, the matrix equation is consistent if and only if (2.4) holds. $\quad\square$

Proof of Theorem 2.2 Under (2.3), (1.38) is equivalent to finding a solution $\mathbf{L}_0$ of $\mathbf{L}_0\mathbf{T}\widehat{\mathbf{X}} = \mathbf{F}$ such that

$$f(\mathbf{L}) \succcurlyeq f(\mathbf{L}_0)\ \ \text{s.t.}\ \ \mathbf{LT}\widehat{\mathbf{X}} = \mathbf{F} \tag{A.6}$$

holds in the Löwner partial ordering. From Lemma A.3, there always exists a solution $\mathbf{L}_0$ of $\mathbf{L}_0\mathbf{T}\widehat{\mathbf{X}} = \mathbf{F}$ such that $f(\mathbf{L}) \succcurlyeq f(\mathbf{L}_0)$ holds for all solutions of $\mathbf{LT}\widehat{\mathbf{X}} = \mathbf{F}$, and the $\mathbf{L}_0$ is determined by the matrix equation $\mathbf{L}_0[\mathbf{T}\widehat{\mathbf{X}},\, \mathbf{TVT}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}] = [\mathbf{F},\, (\mathbf{GV}_1 + \mathbf{HV}_2)\mathbf{T}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}]$, establishing the matrix equation in (2.5). The general solution of the equation is given in (2.7) by Lemma A.2.

Results (a)–(d) are routine consequences of (2.7).

Taking dispersion operation of (2.7) yields (2.8) and (2.10). Also from (1.33) and (2.7), the covariance matrix between $\text{BLUP}_{\mathcal{N}}(\boldsymbol{\phi})$ and $\boldsymbol{\phi}$ is

$$Cov\{\,\text{BLUP}_{\mathcal{N}}(\boldsymbol{\phi}),\, \boldsymbol{\phi}\,\} = \mathbf{LT}Cov\{\,\mathbf{y},\, \boldsymbol{\phi}\,\}$$

$$= [\mathbf{F},\, (\mathbf{GV}_1 + \mathbf{HV}_2)\mathbf{T}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}][\mathbf{T}\widehat{\mathbf{X}},\, \mathbf{TVT}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}]^{+}\mathbf{T}(\mathbf{GV}_1 + \mathbf{HV}_2)',$$

thus establishing (2.9). Further by (2.7),

$$\mathbf{LT}\widetilde{\mathbf{X}} = \left([\mathbf{F},\, (\mathbf{GV}_1 + \mathbf{HV}_2)\mathbf{T}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}][\mathbf{T}\widehat{\mathbf{X}},\, \mathbf{TVT}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}]^{+} + \mathbf{U}[\mathbf{T}\widehat{\mathbf{X}},\, \mathbf{TVT}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}]^{\perp}\right)\mathbf{T}\widetilde{\mathbf{X}}.$$

Substituting it into (2.3) yields

$$Cov(\boldsymbol{\phi} - \mathbf{LTy}) = \left([\mathbf{F},\, (\mathbf{GV}_1 + \mathbf{HV}_2)\mathbf{T}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}][\mathbf{T}\widehat{\mathbf{X}},\, \mathbf{TVT}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}]^{+}\mathbf{T}\widetilde{\mathbf{X}} - \mathbf{J}\right)\boldsymbol{\Sigma}$$

$$\times \left([\mathbf{F},\, (\mathbf{GV}_1 + \mathbf{HV}_2)\mathbf{T}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}][\mathbf{T}\widehat{\mathbf{X}},\, \mathbf{TVT}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}]^{+}\mathbf{T}\widetilde{\mathbf{X}} - \mathbf{J}\right)',$$

thus establishing (2.11).

Note that $[\mathbf{F},\, (\mathbf{GV}_1 + \mathbf{HV}_2)\mathbf{T}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}][\mathbf{T}\widehat{\mathbf{X}},\, \mathbf{TVT}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}]^{+}\mathbf{T}$ in (2.7) can be decomposed as the sum of three terms

$$[\mathbf{F},\, (\mathbf{GV}_1 + \mathbf{HV}_2)\mathbf{T}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}][\mathbf{T}\widehat{\mathbf{X}},\, \mathbf{TVT}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}]^{+}\mathbf{T}$$

$$= [\mathbf{F},\, \mathbf{0}][\mathbf{T}\widehat{\mathbf{X}},\, \mathbf{TVT}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}]^{+}\mathbf{T}$$

$$+ [\mathbf{0},\, \mathbf{G}[\mathbf{I}_p,\, \mathbf{0}]\boldsymbol{\Sigma}(\mathbf{T}\widetilde{\mathbf{X}})'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}][\mathbf{T}\widehat{\mathbf{X}},\, \mathbf{TVT}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}]^{+}\mathbf{T}$$

$$+ [\mathbf{0},\, \mathbf{H}[\mathbf{0},\, \mathbf{I}_n]\boldsymbol{\Sigma}(\mathbf{T}\widetilde{\mathbf{X}})'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}][\mathbf{T}\widehat{\mathbf{X}},\, \mathbf{TVT}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}]^{+}\mathbf{T}$$

$$:= \mathbf{Q}_1 + \mathbf{Q}_2 + \mathbf{Q}_3.$$

Hence

$$\text{BLUP}_{\mathcal{N}}(\boldsymbol{\phi}) = \left(\mathbf{Q}_1 + \mathbf{U}_1[\mathbf{T}\widehat{\mathbf{X}},\, \mathbf{TVT}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}]^{\perp}\mathbf{T}\right)\mathbf{y}$$

$$+ \left(\mathbf{Q}_2 + \mathbf{U}_2[\mathbf{T}\widehat{\mathbf{X}},\, \mathbf{TVT}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}]^{\perp}\mathbf{T}\right)\mathbf{y}$$

$$+ \left(\mathbf{Q}_3 + \mathbf{U}_3[\mathbf{T}\widehat{\mathbf{X}},\, \mathbf{TVT}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}]^{\perp}\mathbf{T}\right)\mathbf{y}$$

$$= \text{BLUE}_{\mathcal{N}}(\mathbf{F}\boldsymbol{\alpha}) + \text{BLUP}_{\mathcal{N}}(\mathbf{G}\boldsymbol{\gamma}) + \text{BLUP}_{\mathcal{N}}(\mathbf{H}\boldsymbol{\varepsilon}),$$

thus establishing (2.12).

From (2.7), the covariance matrix between $\text{BLUE}_{\mathcal{N}}(\mathbf{F}\boldsymbol{\alpha})$ and $\text{BLUP}_{\mathcal{N}}(\mathbf{G}\boldsymbol{\gamma} + \mathbf{H}\boldsymbol{\varepsilon})$ is

$$Cov\{\,\text{BLUE}_{\mathcal{N}}(\mathbf{F}\boldsymbol{\alpha}),\, \text{BLUP}_{\mathcal{N}}(\mathbf{G}\boldsymbol{\gamma} + \mathbf{H}\boldsymbol{\varepsilon})\,\}$$

$$= [\mathbf{F},\, \mathbf{0}][\mathbf{T}\widehat{\mathbf{X}},\, \mathbf{TVT}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}]^{+}\mathbf{TVT}'\left([\mathbf{0},\, (\mathbf{GV}_1 + \mathbf{HV}_2)\mathbf{T}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}][\mathbf{T}\widehat{\mathbf{X}},\, \mathbf{TVT}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}]^{+}\right)'. \tag{A.7}$$

Applying (A.5) to (A.7) and simplifying, we obtain

$$r(Cov\{\operatorname{BLUE}_{\mathcal{N}}(\mathbf{F}\boldsymbol{\alpha}),\ \operatorname{BLUP}_{\mathcal{N}}(\mathbf{G}\boldsymbol{\gamma}+\mathbf{H}\boldsymbol{\varepsilon})\})$$

$$= r\left([\mathbf{F},\ \mathbf{0}][\mathbf{T}\widehat{\mathbf{X}},\ \mathbf{T}\mathbf{V}\mathbf{T}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}]^{+}\mathbf{T}\mathbf{V}\mathbf{T}'\left([\mathbf{0},\ (\mathbf{G}\mathbf{V}_1+\mathbf{H}\mathbf{V}_2)\mathbf{T}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}][\mathbf{T}\widehat{\mathbf{X}},\ \mathbf{T}\mathbf{V}\mathbf{T}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}]^{+}\right)'\right)$$

$$= r\begin{bmatrix} \mathbf{0} & \begin{bmatrix} (\mathbf{T}\widehat{\mathbf{X}})' \\ (\mathbf{T}\widehat{\mathbf{X}})^{\perp}\mathbf{T}\mathbf{V}\mathbf{T}' \end{bmatrix} & \begin{bmatrix} \mathbf{0} \\ (\mathbf{T}\widehat{\mathbf{X}})^{\perp}\mathbf{T}\widetilde{\mathbf{X}}\boldsymbol{\Sigma}\mathbf{J}' \end{bmatrix} \\ [\mathbf{T}\widehat{\mathbf{X}},\ \mathbf{T}\mathbf{V}\mathbf{T}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}] & \mathbf{T}\mathbf{V}\mathbf{T}' & \mathbf{0} \\ [\mathbf{F},\ \mathbf{0}] & \mathbf{0} & \mathbf{0} \end{bmatrix}$$

$$\quad - 2r[\mathbf{T}\widehat{\mathbf{X}},\ \mathbf{T}\mathbf{V}\mathbf{T}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}]$$

$$= r\begin{bmatrix} \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -(\mathbf{T}\widehat{\mathbf{X}})^{\perp}\mathbf{T}\mathbf{V}\mathbf{T}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp} \end{bmatrix} & \begin{bmatrix} (\mathbf{T}\widehat{\mathbf{X}})' \\ \mathbf{0} \end{bmatrix} & \begin{bmatrix} \mathbf{0} \\ (\mathbf{T}\widehat{\mathbf{X}})^{\perp}\mathbf{T}(\mathbf{G}\mathbf{V}_1+\mathbf{H}\mathbf{V}_2)' \end{bmatrix} \\ [\mathbf{T}\widehat{\mathbf{X}},\ \mathbf{0}] & \mathbf{T}\mathbf{V}\mathbf{T}' & \mathbf{0} \\ [\mathbf{F},\ \mathbf{0}] & \mathbf{0} & \mathbf{0} \end{bmatrix} - 2r[\mathbf{T}\widehat{\mathbf{X}},\ \mathbf{T}\widetilde{\mathbf{X}}\boldsymbol{\Sigma}]$$

$$= r\begin{bmatrix} \mathbf{0} & (\mathbf{T}\widehat{\mathbf{X}})' \\ \mathbf{T}\widehat{\mathbf{X}} & \mathbf{T}\mathbf{V}\mathbf{T}' \\ \mathbf{F} & \mathbf{0} \end{bmatrix} + r[(\mathbf{T}\widehat{\mathbf{X}})^{\perp}\mathbf{T}\mathbf{V}\mathbf{T}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp},\ (\mathbf{T}\widehat{\mathbf{X}})^{\perp}\mathbf{T}(\mathbf{G}\mathbf{V}_1+\mathbf{H}\mathbf{V}_2)'] - 2r[\mathbf{T}\widehat{\mathbf{X}},\ \mathbf{T}\widetilde{\mathbf{X}}\boldsymbol{\Sigma}]$$

$$= r\begin{bmatrix} \mathbf{T}\widehat{\mathbf{X}} \\ \mathbf{F} \end{bmatrix} + r\begin{bmatrix} (\mathbf{T}\widehat{\mathbf{X}})' \\ \mathbf{T}\mathbf{V}\mathbf{T}' \end{bmatrix} + r[\mathbf{T}\widehat{\mathbf{X}},\ \mathbf{T}\widetilde{\mathbf{X}}\boldsymbol{\Sigma},\ \mathbf{T}(\mathbf{G}\mathbf{V}_1+\mathbf{H}\mathbf{V}_2)'] - r(\mathbf{T}\widehat{\mathbf{X}}) - 2r[\mathbf{T}\widehat{\mathbf{X}},\ \mathbf{T}\widetilde{\mathbf{X}}\boldsymbol{\Sigma}] \quad \text{(by (A.3))}$$

$$= r(\mathbf{T}\widehat{\mathbf{X}}) + r\begin{bmatrix} (\mathbf{T}\widehat{\mathbf{X}})' \\ \boldsymbol{\Sigma}(\mathbf{T}\widetilde{\mathbf{X}})' \end{bmatrix} + r[\mathbf{T}\widehat{\mathbf{X}},\ \mathbf{T}\widetilde{\mathbf{X}}\boldsymbol{\Sigma}] - r(\mathbf{T}\widehat{\mathbf{X}}) - 2r[\mathbf{T}\widehat{\mathbf{X}},\ \mathbf{T}\widetilde{\mathbf{X}}\boldsymbol{\Sigma}]$$

$$= 0,$$

which implies that $Cov\{\operatorname{BLUE}_{\mathcal{N}}(\mathbf{F}\boldsymbol{\alpha}),\ \operatorname{BLUP}_{\mathcal{N}}(\mathbf{G}\boldsymbol{\gamma}+\mathbf{H}\boldsymbol{\varepsilon})\}$ is a zero matrix, establishing (2.13). Equation (2.14) follows from (2.12) and (2.13). Result (g) follows from (2.7). $\quad\square$

Proof of Theorem 3.2 From Definition 3.1(a), $\operatorname{BLUP}_{\mathcal{M}}(\boldsymbol{\phi}) = \operatorname{BLUP}_{\mathcal{N}}(\boldsymbol{\phi})$ holds definitely if and only if the coefficient matrix of $\mathbf{y}$ in (2.7) satisfies the matrix equation in (2.16), i.e.,

$$\left([\mathbf{F},\ (\mathbf{G}\mathbf{V}_1+\mathbf{H}\mathbf{V}_2)\mathbf{T}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}][\mathbf{T}\widehat{\mathbf{X}},\ \mathbf{T}\mathbf{V}\mathbf{T}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}]^{+}\mathbf{T} + \mathbf{U}[\mathbf{T}\widehat{\mathbf{X}},\ \mathbf{T}\mathbf{V}\mathbf{T}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}]^{\perp}\mathbf{T}\right)[\widehat{\mathbf{X}},\ \mathbf{V}\widehat{\mathbf{X}}^{\perp}]$$

$$= [\mathbf{F},\ (\mathbf{G}\mathbf{V}_1+\mathbf{H}\mathbf{V}_2)\widehat{\mathbf{X}}^{\perp}]. \tag{A.8}$$

From Lemma A.2, this equation is solvable for $\mathbf{U}$ if and only if

$$r\begin{bmatrix} [\mathbf{F},\ (\mathbf{G}\mathbf{V}_1+\mathbf{H}\mathbf{V}_2)\widehat{\mathbf{X}}^{\perp}] - [\mathbf{F},\ (\mathbf{G}\mathbf{V}_1+\mathbf{H}\mathbf{V}_2)\mathbf{T}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}][\mathbf{T}\widehat{\mathbf{X}},\ \mathbf{T}\mathbf{V}\mathbf{T}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}]^{+}[\mathbf{T}\widehat{\mathbf{X}},\ \mathbf{T}\mathbf{V}\widehat{\mathbf{X}}^{\perp}] \\ [\mathbf{T}\widehat{\mathbf{X}},\ \mathbf{T}\mathbf{V}\mathbf{T}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}]^{\perp}[\mathbf{T}\widehat{\mathbf{X}},\ \mathbf{T}\mathbf{V}\widehat{\mathbf{X}}^{\perp}] \end{bmatrix}$$

$$= r([\mathbf{T}\widehat{\mathbf{X}},\ \mathbf{T}\mathbf{V}\mathbf{T}'(\mathbf{T}\widehat{\mathbf{X}})^{\perp}]^{\perp}[\mathbf{T}\widehat{\mathbf{X}},\ \mathbf{T}\mathbf{V}\widehat{\mathbf{X}}^{\perp}]). \tag{A.9}$$

Simplifying both sides by elementary block matrix operations, we obtain

$$
r\left[\begin{array}{c} [\mathbf{F},\,(\mathbf{GV}_1+\mathbf{HV}_2)\widehat{\mathbf{X}}^\perp\,]-[\mathbf{F},\,(\mathbf{GV}_1+\mathbf{HV}_2)\mathbf{T}'(\mathbf{T}\widehat{\mathbf{X}})^\perp\,][\mathbf{T}\widehat{\mathbf{X}},\,\mathbf{TVT}'(\mathbf{T}\widehat{\mathbf{X}})^\perp\,]^+[\mathbf{T}\widehat{\mathbf{X}},\,\mathbf{TV}\widehat{\mathbf{X}}^\perp\,] \\ [\mathbf{T}\widehat{\mathbf{X}},\,\mathbf{TVT}'(\mathbf{T}\widehat{\mathbf{X}})^\perp\,]^+[\mathbf{T}\widehat{\mathbf{X}},\,\mathbf{TV}\widehat{\mathbf{X}}^\perp\,] \end{array}\right]
$$

$$
=r\left[\begin{array}{cc} [\mathbf{F},\,(\mathbf{GV}_1+\mathbf{HV}_2)\widehat{\mathbf{X}}^\perp\,]-[\mathbf{F},\,(\mathbf{GV}_1+\mathbf{HV}_2)\mathbf{T}'(\mathbf{T}\widehat{\mathbf{X}})^\perp\,][\mathbf{T}\widehat{\mathbf{X}},\,\mathbf{TVT}'(\mathbf{T}\widehat{\mathbf{X}})^\perp\,]^+[\mathbf{T}\widehat{\mathbf{X}},\,\mathbf{TV}\widehat{\mathbf{X}}^\perp\,] & \mathbf{0} \\ [\mathbf{T}\widehat{\mathbf{X}},\,\mathbf{TV}\widehat{\mathbf{X}}^\perp\,] & [\mathbf{T}\widehat{\mathbf{X}},\,\mathbf{TVT}'(\mathbf{T}\widehat{\mathbf{X}})^\perp\,] \end{array}\right]
$$

$$
-\,r[\mathbf{T}\widehat{\mathbf{X}},\,\mathbf{TVT}'(\mathbf{T}\widehat{\mathbf{X}})^\perp\,]
$$

$$
=r\left[\begin{array}{cc} [\mathbf{F},\,(\mathbf{GV}_1+\mathbf{HV}_2)\widehat{\mathbf{X}}^\perp\,] & [\mathbf{F},\,(\mathbf{GV}_1+\mathbf{HV}_2)\mathbf{T}'(\mathbf{T}\widehat{\mathbf{X}})^\perp\,] \\ [\mathbf{T}\widehat{\mathbf{X}},\,\mathbf{TV}\widehat{\mathbf{X}}^\perp\,] & [\mathbf{T}\widehat{\mathbf{X}},\,\mathbf{TVT}'(\mathbf{T}\widehat{\mathbf{X}})^\perp\,] \end{array}\right]-r[\mathbf{T}\widehat{\mathbf{X}},\,\mathbf{T}\widetilde{\mathbf{X}}\boldsymbol{\Sigma}\,]
$$

$$
=r\left[\begin{array}{cc} \mathbf{F} & (\mathbf{GV}_1+\mathbf{HV}_2)\widehat{\mathbf{X}}^\perp \\ \mathbf{T}\widehat{\mathbf{X}} & \mathbf{TV}\widehat{\mathbf{X}}^\perp \end{array}\right]-r[\mathbf{T}\widehat{\mathbf{X}},\,\mathbf{T}\widetilde{\mathbf{X}}\boldsymbol{\Sigma}\,]
$$

$$
=r\left[\begin{array}{cc} \mathbf{T}\widehat{\mathbf{X}} & \mathbf{TV} \\ \mathbf{0} & \widehat{\mathbf{X}}' \\ \mathbf{F} & (\mathbf{GV}_1+\mathbf{HV}_2) \end{array}\right]-r(\widehat{\mathbf{X}})-r[\mathbf{T}\widehat{\mathbf{X}},\,\mathbf{T}\widetilde{\mathbf{X}}\boldsymbol{\Sigma}\,]\quad\text{(by (A.2))},
$$

$$
=r\left[\begin{array}{cc} \mathbf{T}\widehat{\mathbf{X}} & Cov\{\mathbf{Ty},\,\mathbf{y}\} \\ \mathbf{0} & \widehat{\mathbf{X}}' \\ \mathbf{F} & Cov\{\boldsymbol{\phi},\,\mathbf{y}\} \end{array}\right]-r(\widehat{\mathbf{X}})-r[\mathbf{T}\widehat{\mathbf{X}},\,\mathbf{T}\widetilde{\mathbf{X}}\boldsymbol{\Sigma}\,]
$$

$$
=r\left[\begin{array}{cc} \mathbf{T}\widehat{\mathbf{X}} & Cov\{\mathbf{Ty},\,\widehat{\mathbf{X}}^\perp\mathbf{y}\} \\ \mathbf{F} & Cov\{\boldsymbol{\phi},\,\widehat{\mathbf{X}}^\perp\mathbf{y}\} \end{array}\right]-r[\mathbf{T}\widehat{\mathbf{X}},\,\mathbf{T}\widetilde{\mathbf{X}}\boldsymbol{\Sigma}\,],
$$

and

$$
r([\mathbf{T}\widehat{\mathbf{X}},\,\mathbf{TVT}'(\mathbf{T}\widehat{\mathbf{X}})^\perp\,]^+[\mathbf{T}\widehat{\mathbf{X}},\,\mathbf{TV}\widehat{\mathbf{X}}^\perp\,])=r\left[\begin{array}{cc} \mathbf{T}\widehat{\mathbf{X}} & Cov\{\mathbf{Ty},\,\mathbf{y}\} \\ \mathbf{0} & \widehat{\mathbf{X}}' \end{array}\right]-r(\widehat{\mathbf{X}})-r[\mathbf{T}\widehat{\mathbf{X}},\,\mathbf{T}\widetilde{\mathbf{X}}\boldsymbol{\Sigma}\,]
$$

$$
=r[\mathbf{T}\widehat{\mathbf{X}},\,Cov\{\mathbf{Ty},\,\widehat{\mathbf{X}}^\perp\mathbf{y}\}\,]-r[\mathbf{T}\widehat{\mathbf{X}},\,\mathbf{T}\widetilde{\mathbf{X}}\boldsymbol{\Sigma}\,].
$$

Substituting the two equalities into (A.9) leads to the equivalence of (a), (c), (d) and (e).

From Definition 3.1(b), $\mathrm{BLUP}_{\mathscr{M}}(\boldsymbol{\phi})=\mathrm{BLUP}_{\mathscr{N}}(\boldsymbol{\phi})$ holds with probability 1 if and only if the coefficient matrices in (2.7) and (2.17) satisfy

$$
\left([\mathbf{F},\,(\mathbf{GV}_1+\mathbf{HV}_2)\widehat{\mathbf{X}}^\perp\,][\widehat{\mathbf{X}},\,\mathbf{V}\widehat{\mathbf{X}}^\perp\,]^++\mathbf{U}_1[\widehat{\mathbf{X}},\,\mathbf{V}\widehat{\mathbf{X}}^\perp\,]^\perp\right)[\widehat{\mathbf{X}},\,\mathbf{V}\widehat{\mathbf{X}}^\perp\,]
$$

$$
=\left([\mathbf{F},\,(\mathbf{GV}_1+\mathbf{HV}_2)\mathbf{T}'(\mathbf{T}\widehat{\mathbf{X}})^\perp\,][\mathbf{T}\widehat{\mathbf{X}},\,\mathbf{TVT}'(\mathbf{T}\widehat{\mathbf{X}})^\perp\,]^+\mathbf{T}+\mathbf{U}[\mathbf{T}\widehat{\mathbf{X}},\,\mathbf{TVT}'(\mathbf{T}\widehat{\mathbf{X}})^\perp\,]^\perp\mathbf{T}\right)[\widehat{\mathbf{X}},\,\mathbf{V}\widehat{\mathbf{X}}^\perp\,],
$$

which can further reduce to (A.8) from

$$
[\mathbf{F},\,(\mathbf{GV}_1+\mathbf{HV}_2)\widehat{\mathbf{X}}^\perp\,][\widehat{\mathbf{X}},\,\mathbf{V}\widehat{\mathbf{X}}^\perp\,]^+[\widehat{\mathbf{X}},\,\mathbf{V}\widehat{\mathbf{X}}^\perp\,]=[\mathbf{F},\,(\mathbf{GV}_1+\mathbf{HV}_2)\widehat{\mathbf{X}}^\perp\,]\ \text{ and }\ [\widehat{\mathbf{X}},\,\mathbf{V}\widehat{\mathbf{X}}^\perp\,]^\perp[\widehat{\mathbf{X}},\,\mathbf{V}\widehat{\mathbf{X}}^\perp\,]=\mathbf{0}.
$$

Hence, (a) and (b) are equivalent. $\qquad\square$

Proof of Theorem 3.4 Eqs. (3.1) and (3.2) follow from (2.12) and (2.22) by setting $\boldsymbol{\phi}=\mathbf{Ty}=\mathbf{T}\widehat{\mathbf{X}}\boldsymbol{\alpha}+\mathbf{TX}\boldsymbol{\gamma}+\mathbf{T}\boldsymbol{\varepsilon}$. $\qquad\square$

Proof of Theorem 3.5 Results (a)–(c) follow from Theorem 3.2 by setting $\boldsymbol{\phi}=\mathbf{TX}\boldsymbol{\beta}$, $\mathbf{T}\widehat{\mathbf{X}}\boldsymbol{\alpha}$, $\mathbf{TX}\boldsymbol{\gamma}$, $\mathbf{T}\boldsymbol{\varepsilon}$, respectively.

References

[1] J.K. Baksalary, R. Kala. Linear transformations preserving best linear unbiased estimators in a general Gauss–Markoff model. Ann. Statist. 9 (1981) 913–916.

[2] B.A. Brumback. On the built-in restrictions in linear mixed models, with Application to smoothing spline analysis of variance. Commun. Stat. Theory Methods 39 (2010) 579–591.

[3] A.S. Bryk, S.W. Raudenbush. Hierarchical Linear Models: Applications and Data Analysis Methods. 2nd Edn, Newbury Park, CA: Sage, 2002.

[4] A. Chaturvedi., S. Kesarwani, R. Chandra. Simultaneous prediction based on shrinkage estimator. In: Recent Advances in Linear Models and Related Areas, Essays in Honour of Helge Toutenburg. Springer, Berlin, 2008, pp. 181–204.

[5] J. DeLeeuw, I. Kreft. Random efficient models for multilevel analysis. J. Educ. Statist. 11 (1986) 57–85.

[6] E. Demidenko. Mixed Models: Theory and Applications. Wiley, New York, 2004.

[7] B. Dong, W. Guo, Y. Tian. On relations between BLUEs under two transformed linear models. J. Multivariate Anal. 131 (2014) 279–292.

[8] H. Drygas. Sufficiency and completeness in the general Gauss–Markov model. Sankhyā Ser. A 45 (1983) 88–98.

[9] A.S. Goldberger. Best linear unbiased prediction in the generalized linear regression models. J. Amer. Stat. Assoc. 57 (1962) 369–375.

[10] J. Groß. A note on the concepts of linear and quadratic sufficiency. J. Statist. Plann. Inference 70 (1998) 69–76.

[11] L. Guttman. General theory and methods for matric factoring. Psychometrika 9 (1944) 1–16.

[12] D. Harville. Extension of the Gauss–Markov theorem to include the estimation of random effects. Ann. Statist. 4 (1976) 384–395.

[13] S.J. Haslett, S. Puntanen. Equality of BLUEs or BLUPs under two linear models using stochastic restrictions. Stat. Papers 51 (2010) 465–475.

[14] S.J. Haslett, S. Puntanen. Note on the equality of the BLUPs for new observations under two linear models. Acta Comm. Univ. Tartu. Math. 14 (2010) 27–33.

[15] S.J. Haslett, S. Puntanen. On the equality of the BLUPs under two linear mixed models. Metrika 74 (2011) 381–395.

[16] J. Isotalo, S. Puntanen. Linear prediction sufficiency for new observations in the general Gauss–Markov model. Commun. Stat. Theory Methods 35 (2006) 1011–1023.

[17] J. Jiang. Linear and Generalized Linear Mixed Models and Their Applications. Springer, New York, 2007.

[18] R. Kala, P.R. Pordzik. Estimation in singular partitioned, reduced or transformed linear models. Stat. Papers 50 (2009) 633–638.

[19] R. Kala, S. Puntanen, Y. Tian. Some notes on linear sufficiency. Stat. Papers 2016, DOI:10.1007/s00362-015-0682-2.

[20] A. Kornacki. The linear sufficiency and invariant linear sufficiency in a multivariate Gauss–Markov model. Metrika 42 (1995) 240–241.

[21] A. Kornacki. Different kinds of sufficiency in the general Gauss–Markov model. Math. Slovaca 57 (2007) 389–392.

[21] N. Laird, J. Ware. Random-effects models for longitudinal data. Biometrics 38 (1982) 963–974.

[22] E.P. Liski, G. Trenkler, J. Groß. Estimation from transformed data under the linear regression model. Statistics 29 (1997), 205–219.

[23] R. Ma, B. Jørgensen. Nested generalized linear mixed models: an orthodox best Linear unbiased predictor approach. J. Roy. Statist. Soc. Ser. B. 69 (2007) 625–641.

[24] G. Marsaglia, G.P.H. Styan. Equalities and inequalities for ranks of matrices. Linear Multilinear Algebra 2 (1974) 269–292.

[25] R.A. McLean, W.L. Sanders, W.W. Stroup. A unified approach to mixed linear models. Amer. Statist. 45 (1991) 54–64.

[26] J. Mueller. Sufficency and completeness in the linear model. J. Multivariate Anal. 21 (1987) 312–323.

[27] K.E. Muller, P.W. Stewart. Linear Model Theory Univariate, Multivariate, and Mixed Models. Wiley, New York, 2006.

[28] W. Oktaba, A. Kornacki, J. Wawrzosek. Invariant linearly sufficient trasformations of the general Gauss–Markov model. Estimation and testing. Scand. J. Statist. 15 (1988) 117–124.

[29] R. Penrose. A generalized inverse for matrices. Proc. Cambridge Phil. Soc. 51 (1955) 406–413

[30] C.R. Rao. Unified theory of linear estimation. Sankhyā, Ser. A 33 (1971) 371–394.

[31] C.R. Rao. Representations of best linear unbiased estimators in the Gauss–Markoff model with a singular dispersion matrix. J. Multivariate Anal. 3 (1973) 276–292.

[32] C.R. Rao. A lemma on optimization of matrix function and a review of the unified theory of linear estimation. In: Statistical Data Analysis and Inference, Y. Dodge (ed.), Elsevier, North-Holland, 1989, pp. 397–417.

[33] C.R. Rao. H. Toutenburg, Shalabh, C. Heumann. Linear Models and Generalizations Least Squares and Alternatives. 3rd Edn., Springer, Berlin, 2008.

[34] S.R. Searle. Built-in restrictions on best linear unbiased predictors (BLUP) of random effects in mixed Models. Amer. Statist. 51 (1997) 19–21.

[35] Shalabh. Performance of Stein-rule procedure for simultaneous prediction of actual and average values of study variable in linear regression models. Bull. Internat. Stat. Instit. 56 (1995) 1375–1390.

[36] Y. Tian. More on maximal and minimal ranks of Schur complements with applications. Appl. Math. Comput. 152 (2004) 675–692

[37] Y. Tian. A new derivation of BLUPs under random-effects model. Metrika 78 (2015) 905–918.

[38] Y. Tian. A matrix handling of predictions under a general linear random-effects model with new observations. Electron. J. Linear Algebra 29 (2015) 30–45.

[39] Y. Tian, B. Jiang. An algebraic study of BLUPs under two linear random-effects models with correlated covariance matrices. Linear Multilinear Algebra, 2016, DOI:10.1080/03081087.2016.1155533.

[40] Y. Tian, S. Puntanen. On the equivalence of estimations under a general linear model and its transformed models. Linear Algebra Appl. 430 (2009) 2622–2641.

[41] H. Toutenburg. Prior Information in Linear Models. Wiley, New York, 1982.

[42] G. Verbeke, G. Molenberghs. Linear Mixed Models for Longitudinal Data. Springer, New York, 2000.