

Yongge Tian*

Matrix rank and inertia formulas in the analysis of general linear models

DOI 10.1515/math-2017-0013

Received March 15, 2016; accepted January 3, 2017.

Abstract: Matrix mathematics provides a powerful tool set for addressing statistical problems, in particular, the theory of matrix ranks and inertias has been developed as effective methodology of simplifying various complicated matrix expressions, and establishing equalities and inequalities occurred in statistical analysis. This paper describes how to establish exact formulas for calculating ranks and inertias of covariances of predictors and estimators of parameter spaces in general linear models (GLMs), and how to use the formulas in statistical analysis of GLMs. We first derive analytical expressions of best linear unbiased predictors/best linear unbiased estimators (BLUPs/BLUEs) of all unknown parameters in the model by solving a constrained quadratic matrix-valued function optimization problem, and present some well-known results on ordinary least-squares predictors/ordinary least-squares estimators (OLSPs/OLSEs). We then establish some fundamental rank and inertia formulas for covariance matrices related to BLUPs/BLUEs and OLSPs/OLSEs, and use the formulas to characterize a variety of equalities and inequalities for covariance matrices of BLUPs/BLUEs and OLSPs/OLSEs. As applications, we use these equalities and inequalities in the comparison of the covariance matrices of BLUPs/BLUEs and OLSPs/OLSEs. The work on the formulations of BLUPs/BLUEs and OLSPs/OLSEs, and their covariance matrices under GLMs provides direct access, as a standard example, to a very simple algebraic treatment of predictors and estimators in linear regression analysis, which leads a deep insight into the linear nature of GLMs and gives an efficient way of summarizing the results.

Keywords: General linear model, Predictor, Estimator, Covariance matrix, Rank, Inertia, Equality, Inequality

MSC: 15A03, 15A09, 62H12, 62J05

1 Introduction

Throughout this paper, the symbol $\mathbb{R}^{m \times n}$ stands for the collection of all $m \times n$ real matrices. The symbols \mathbf{A}' , $r(\mathbf{A})$, and $\mathcal{R}(\mathbf{A})$ 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

$$\mathbf{A}\mathbf{G}\mathbf{A} = \mathbf{A}, \quad \mathbf{G}\mathbf{A}\mathbf{G} = \mathbf{G}, \quad (\mathbf{A}\mathbf{G})' = \mathbf{A}\mathbf{G}, \quad (\mathbf{G}\mathbf{A})' = \mathbf{G}\mathbf{A}.$$

Further, let $\mathbf{P}_\mathbf{A}$, $\mathbf{E}_\mathbf{A}$ and $\mathbf{F}_\mathbf{A}$ stand for the three orthogonal projectors (symmetric idempotent matrices) $\mathbf{P}_\mathbf{A} = \mathbf{A}\mathbf{A}^+$, $\mathbf{E}_\mathbf{A} = \mathbf{A}^\perp = \mathbf{I}_m - \mathbf{A}\mathbf{A}^+$, and $\mathbf{F}_\mathbf{A} = \mathbf{I}_n - \mathbf{A}^+\mathbf{A}$. All about the orthogonal projectors $\mathbf{P}_\mathbf{A}$, $\mathbf{E}_\mathbf{A}$, and $\mathbf{F}_\mathbf{A}$ with their applications in the linear statistical models can be found in [1–3]. The symbols $i_+(\mathbf{A})$ and $i_-(\mathbf{A})$ for $\mathbf{A} = \mathbf{A}' \in \mathbb{R}^{m \times m}$, called the positive inertia and negative inertia of \mathbf{A} , denote the number of the positive and negative eigenvalues of \mathbf{A} counted with multiplicities, respectively. For brief, we use $i_\pm(\mathbf{A})$ to denote the both numbers. $\mathbf{A} > \mathbf{0}$, $\mathbf{A} \geq \mathbf{0}$, $\mathbf{A} < \mathbf{0}$, and $\mathbf{A} \leq \mathbf{0}$ mean that \mathbf{A} is a symmetric positive definite, positive semi-definite, negative

*Corresponding Author: Yongge Tian: China Economics and Management Academy, Central University of Finance and Economics, Beijing 100081, China, E-mail: yongge.tian@gmail.com

definite, negative semi-definite matrix, respectively. Two symmetric matrices \mathbf{A} and \mathbf{B} of the same size are said to satisfy the inequalities $\mathbf{A} \succ \mathbf{B}$, $\mathbf{A} \succeq \mathbf{B}$, $\mathbf{A} \prec \mathbf{B}$, and $\mathbf{A} \preceq \mathbf{B}$ in the Löwner partial ordering if $\mathbf{A} - \mathbf{B}$ is positive definite, positive semi-definite, negative definite, and negative semi-definite, respectively. Also, it is well known that the Löwner partial ordering is a surprisingly strong and useful property between two symmetric matrices. For more results on connections between inertias and the Löwner partial ordering of real symmetric (complex Hermitian) matrices, as well as applications of inertias and the Löwner partial ordering in statistics, see, e.g., [2, 4–9].

Recall that linear models are the first type of regression models to be studied extensively in statistical inference, which have had a profound impact in the field of statistics and applications and have been regarded without doubt as a kernel part in current statistical theory. A typical form of linear models is defined by

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad E(\boldsymbol{\varepsilon}) = \mathbf{0}, \quad D(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{\Sigma}, \quad (1)$$

where $\mathbf{y} \in \mathbb{R}^{n \times 1}$ is 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 fixed but unknown parameters, $E(\boldsymbol{\varepsilon})$ and $D(\boldsymbol{\varepsilon})$ denote the expectation vector and the dispersion matrix of the random error vector $\boldsymbol{\varepsilon} \in \mathbb{R}^{n \times 1}$, $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$ is a known positive semi-definite matrix of arbitrary rank, and σ^2 is unknown positive number.

Once a general linear model (GLM) is formulated, the first and most important task is to estimate or predict unknown parameters in the model by using various mathematical and statistical tools. As a foundation of current regression theory, there are a sufficient bunch of results on estimators and predictors of parameter spaces in GLMs. Even so, people are still able to obtain more and more new and valuable results on statistical inference of GLMs. Estimation of $\boldsymbol{\beta}$ as well as prediction of $\boldsymbol{\varepsilon}$ in (1) are major concerns in the statistical inference of (1), and it is always desirable, as claimed in [10, 11], to simultaneously identify estimators and predictors of all unknown parameters in GLMs. As formulated in [10, 11], a general vector of linear parametric functions involving the two unknown parameter vectors $\boldsymbol{\beta}$ and $\boldsymbol{\varepsilon}$ in (1) is given by

$$\boldsymbol{\phi} = \mathbf{K}\boldsymbol{\beta} + \mathbf{J}\boldsymbol{\varepsilon}, \quad (2)$$

where $\mathbf{K} \in \mathbb{R}^{k \times p}$ and $\mathbf{J} \in \mathbb{R}^{k \times n}$ are given matrices of arbitrary ranks. Eq. (2) includes all vector and matrix operations in (1) as its special cases. For instance,

- (i) if $\mathbf{K} = \mathbf{X}$ and $\mathbf{J} = \mathbf{I}_n$, (2) becomes $\boldsymbol{\phi} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \mathbf{y}$, the observed response vector;
- (ii) if $\mathbf{J} = \mathbf{0}$, (2) becomes $\boldsymbol{\phi} = \mathbf{K}\boldsymbol{\beta}$, a general vector of linear parametric functions;
- (iii) if $\mathbf{K} = \mathbf{X}$ and $\mathbf{J} = \mathbf{0}$, (2) becomes $\boldsymbol{\phi} = \mathbf{X}\boldsymbol{\beta}$, the mean vector;
- (iv) if $\mathbf{K} = \mathbf{0}$ and $\mathbf{J} = \mathbf{I}_n$, (2) becomes $\boldsymbol{\phi} = \boldsymbol{\varepsilon}$, the random error vector.

Theoretical and applied researches seek to develop various possible predictors/estimators of (2) and its specific cases. During these approaches, the unbiasedness of $\boldsymbol{\phi}$ with respect to the parameter spaces in (1) is an important property. In statistical inference of a GLM, usually there are many unbiased predictors/estimators for the same parameter space. Under the situation that there exist unbiased predictors/estimators for the same parameter space, it is natural to find such an unbiased predictor/estimator that has the smallest dispersion matrix among all the unbiased predictors/estimators. Thus, the unbiasedness and the smallest dispersion matrices of predictors/estimators are most intrinsic requirements in the statistical analysis of GLMs. Based on these requirements, we introduce the following classic concepts of the predictability, estimability, and the BLUP/BLUEs of $\boldsymbol{\phi}$ in (2) and its special cases originated from [12].

Definition 1.1. The vector $\boldsymbol{\phi}$ in (2) is said to be predictable under (1) if there exists an $\mathbf{L} \in \mathbb{R}^{k \times n}$ such that $E(\mathbf{L}\mathbf{y} - \boldsymbol{\phi}) = \mathbf{0}$. In particular, $\mathbf{K}\boldsymbol{\beta}$ is said to be estimable under (1) if there exists an $\mathbf{L} \in \mathbb{R}^{k \times n}$ such that $E(\mathbf{L}\mathbf{y} - \mathbf{K}\boldsymbol{\beta}) = \mathbf{0}$.

Definition 1.2. Let $\boldsymbol{\phi}$ be as given in (2) and assume that it is predictable under (1). If there exists a matrix \mathbf{L} such that

$$E(\mathbf{L}\mathbf{y} - \boldsymbol{\phi}) = \mathbf{0} \quad \text{and} \quad D(\mathbf{L}\mathbf{y} - \boldsymbol{\phi}) = \min \quad (3)$$

holds in the Löwner partial ordering, the linear statistic \mathbf{Ly} is defined to be the best linear unbiased predictor (BLUP) of $\boldsymbol{\phi}$ under (1), and is denoted by

$$\mathbf{Ly} = \text{BLUP}(\boldsymbol{\phi}) = \text{BLUP}(\mathbf{K}\boldsymbol{\beta} + \mathbf{J}\boldsymbol{\varepsilon}). \quad (4)$$

If $\mathbf{J} = \mathbf{0}$ in (2), or $\mathbf{K} = \mathbf{0}$ in (2), then the \mathbf{Ly} satisfying (3) is called the best linear unbiased estimator (BLUE) and the BLUP of $\mathbf{K}\boldsymbol{\beta}$ and $\mathbf{J}\boldsymbol{\varepsilon}$ under (1), respectively, and are denoted by

$$\mathbf{Ly} = \text{BLUE}(\mathbf{K}\boldsymbol{\beta}), \quad \mathbf{Ly} = \text{BLUP}(\mathbf{J}\boldsymbol{\varepsilon}), \quad (5)$$

respectively.

Definition 1.3. Let $\boldsymbol{\phi}$ be as given in (2). The ordinary least-squares estimator (OLSE) of the unknown parameter vector $\boldsymbol{\beta}$ in (1) is defined to be

$$\text{OLSE}(\boldsymbol{\beta}) = \underset{\boldsymbol{\beta} \in \mathbb{R}^{p \times 1}}{\text{argmin}} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}), \quad (6)$$

while the OLSE of $\mathbf{K}\boldsymbol{\beta}$ under (1) is defined to be $\text{OLSE}(\mathbf{K}\boldsymbol{\beta}) = \mathbf{K}\text{OLSE}(\boldsymbol{\beta})$; the ordinary least-squares predictor (OLSP) of the random error vector $\boldsymbol{\varepsilon}$ in (1) is defined to be

$$\text{OLSP}(\boldsymbol{\varepsilon}) = \mathbf{y} - \text{OLSE}(\mathbf{X}\boldsymbol{\beta}), \quad (7)$$

while the OLSP of $\mathbf{J}\boldsymbol{\varepsilon}$ under (1) is defined to be $\text{OLSP}(\mathbf{J}\boldsymbol{\varepsilon}) = \mathbf{J}\text{OLSP}(\boldsymbol{\varepsilon})$. The OLSP of $\boldsymbol{\phi}$ under (1) is defined to be

$$\text{OLSP}(\boldsymbol{\phi}) = \text{OLSE}(\mathbf{K}\boldsymbol{\beta}) + \text{OLSP}(\mathbf{J}\boldsymbol{\varepsilon}). \quad (8)$$

The above definitions enable us to deal with various prediction and estimation problems under most general assumptions of GLMs.

The purpose of this paper is to investigate the performances of $\text{BLUP}(\boldsymbol{\phi})$, $\text{BLUP}(\boldsymbol{\phi}) - \boldsymbol{\phi}$, $\text{OLSP}(\boldsymbol{\phi})$, and $\text{OLSP}(\boldsymbol{\phi}) - \boldsymbol{\phi}$ under the assumption that $\boldsymbol{\phi}$ is predictable under (1) by using the matrix rank/inertia methodology. Our work in this paper includes:

- (I) establishing linear matrix equations and exact analytical expressions of the BLUPs and the OLSPs of (2);
- (II) characterizing algebraic and statistical properties of the BLUPs and the OLSPs of (2);
- (III) deriving formulas for calculating

$$i_{\pm}(\mathbf{A} - \mathbf{D}[\boldsymbol{\phi} - \text{BLUP}(\boldsymbol{\phi})]), \quad r(\mathbf{A} - \mathbf{D}[\boldsymbol{\phi} - \text{BLUP}(\boldsymbol{\phi})]), \quad (9)$$

$$i_{\pm}(\mathbf{A} - \mathbf{D}[\boldsymbol{\phi} - \text{OLSP}(\boldsymbol{\phi})]), \quad r(\mathbf{A} - \mathbf{D}[\boldsymbol{\phi} - \text{OLSP}(\boldsymbol{\phi})]), \quad (10)$$

where \mathbf{A} is a symmetric matrix, such as, dispersion matrices of other predictors and estimators;

- (IV) establishing necessary and sufficient conditions for the following matrix equalities and inequalities

$$\mathbf{D}[\boldsymbol{\phi} - \text{BLUP}(\boldsymbol{\phi})] = \mathbf{A}, \quad \mathbf{D}[\text{BLUP}(\boldsymbol{\phi}) - \boldsymbol{\phi}] > \mathbf{A} (\geq \mathbf{A}, < \mathbf{A}, \leq \mathbf{A}), \quad (11)$$

$$\mathbf{D}[\boldsymbol{\phi} - \text{OLSP}(\boldsymbol{\phi})] = \mathbf{A}, \quad \mathbf{D}[\text{OLSP}(\boldsymbol{\phi}) - \boldsymbol{\phi}] > \mathbf{A} (\geq \mathbf{A}, < \mathbf{A}, \leq \mathbf{A}) \quad (12)$$

to hold, respectively, in the Löwner partial ordering;

- (V) establishing equalities and inequalities between the BLUPs and OLSPs of $\boldsymbol{\phi}$ and their dispersion matrices.

As defined in (1), we do not attach any restrictions to the ranks of the given matrices in the model in order to obtain general results on this group of problems. Regression analysis is a very important statistical method that investigates the relationship between a response variable and a set of other variables named as independent variables. Linear regression models were the first type of models to be studied rigorously in regression analysis, which were regarded without doubt as a noble and magnificent part in current statistical theory. As demonstrated in most statistical textbooks, some common estimators of unknown parameters under a linear regression model, such as, the well-known OLSEs and BLUEs are usually formulated from certain algebraic operations of the observed response vector,

the given model matrix, and the covariance matrix of the error term in the model. Hence, the standard inference theory of linear regression models can be established without tedious and ambiguous assumptions from the exact algebraic expressions of estimators, which is easily acceptable from both mathematical and statistical points of view. In fact, linear regression models are only type of statistical models that have complete and solid supports from linear algebra and matrix theory, while almost all results on linear regression models can be formulated in matrix forms and calculations. It is just based this fact that linear regression models attract a few of linear algebraists to consider their matrix contributions in statistical analysis.

The paper is organized as follows. In Section 2, we introduce a variety of mathematical tools that can be used to simplify matrix expressions and to characterize matrix equalities, and also present a known result on analytical solutions of a quadratic matrix-valued function optimization problems. In Section 3, we present a group of known results on the predictability of (2), the exact expression of the BLUP of (2) and its special cases, as well as various statistical properties and features of the BLUP. In Section 4, we establish a group of formulas for calculating the rank and inertia in (9) and use the formulas to characterize the equality and inequalities in (11). In Section 5, we first give a group of results on the OLSP of (2) and its statistical properties. We then establish a group of formulas for calculating the rank and inertia in (10) and use the formulas to characterize the equality and inequalities in (12). The connections between the OLSP and BLUP of (2), as well as the equality and inequalities between the dispersion matrices of the OLSP and BLUP of (2) are investigated in Section 6. Conclusions and discussions on algebraic tools in matrix theory, as well as the applications of the rank/inertia formulas in statistical analysis are presented in Section 7.

2 Preliminaries in linear algebra

This section begins with introducing various formulas for ranks and inertias of matrices and claims their usefulness in matrix analysis and statistical theory. Recall that the rank of a matrix and the inertia of a real symmetric matrix are two basic concepts in matrix theory, which are the most significant finite nonnegative integers in reflecting intrinsic properties of matrices, and thus are cornerstones of matrix mathematics. The mathematical prerequisites for understanding ranks and inertias of matrices are minimal and do not go beyond elementary linear algebra, while many simple and classic formulas for calculating ranks and inertias of matrices can be found in most textbooks of linear algebra. The intriguing connections between generalized inverses and ranks of matrices were recognized in 1970s. A variety of fundamental formulas for calculating the ranks of matrices and their generalized inverses were established, and many applications of the rank formulas in matrix theory and statistics were presented in [13]. Since then, the matrix rank theory has been greatly developed and has become an influential and effective tool in simplifying complicated matrix expressions and establishing various matrix equalities.

In order to establish and characterize various possible equalities for predictors and estimators under GLMs, and to simplify various matrix equalities composed by the Moore–Penrose inverses of matrices, we need the following well-known results on ranks and inertias of matrices to make the paper self-contained.

Lemma 2.1. *Let $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m \times n}$, or $\mathbf{A} = \mathbf{A}', \mathbf{B} = \mathbf{B}' \in \mathbb{R}^{m \times m}$. Then the following results hold:*

- (a) $\mathbf{A} = \mathbf{B}$ if and only if $r(\mathbf{A} - \mathbf{B}) = 0$,
- (b) $\mathbf{A} - \mathbf{B}$ is nonsingular if and only if $r(\mathbf{A} - \mathbf{B}) = m = n$,
- (c) $\mathbf{A} > \mathbf{B}$ ($\mathbf{A} < \mathbf{B}$) if and only if $i_+(\mathbf{A} - \mathbf{B}) = m$ ($i_-(\mathbf{A} - \mathbf{B}) = m$),
- (d) $\mathbf{A} \geq \mathbf{B}$ ($\mathbf{A} \leq \mathbf{B}$) if and only if $i_-(\mathbf{A} - \mathbf{B}) = 0$ ($i_+(\mathbf{A} - \mathbf{B}) = 0$).

The assertions in Lemma 2.1 directly follow from the definitions of ranks, inertias, definiteness, and semi-definiteness of (symmetric) matrices, and were first summarized and effectively utilized in [4]. This lemma manifests that if certain explicit formulas for calculating ranks/inertias of differences of (symmetric) matrices are established, we can use the formulas to characterize the corresponding matrix equalities and inequalities. Thus there are important and peculiar consequences of establishing formulas for calculating ranks/inertias of matrices. This fact reflects without doubt the most exciting roles of matrix ranks/inertias in matrix analysis and applications, and thus it is

technically necessary to establish exact matrix rank/inertia formulas as many as possible from the theoretical and applied points of view.

Lemma 2.2 ([13]). 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

$$r[\mathbf{A}, \mathbf{B}] = r(\mathbf{A}) + r(\mathbf{E}_\mathbf{A}\mathbf{B}) = r(\mathbf{B}) + r(\mathbf{E}_\mathbf{B}\mathbf{A}), \quad (13)$$

$$r \begin{bmatrix} \mathbf{A} \\ \mathbf{C} \end{bmatrix} = r(\mathbf{A}) + r(\mathbf{C}\mathbf{F}_\mathbf{A}) = r(\mathbf{C}) + r(\mathbf{A}\mathbf{F}_\mathbf{C}), \quad (14)$$

$$r \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{0} \end{bmatrix} = r(\mathbf{B}) + r(\mathbf{C}) + r(\mathbf{E}_\mathbf{B}\mathbf{A}\mathbf{F}_\mathbf{C}). \quad (15)$$

In particular, the following results hold.

- (a) $r[\mathbf{A}, \mathbf{B}] = r(\mathbf{A}) \Leftrightarrow \mathcal{R}(\mathbf{B}) \subseteq \mathcal{R}(\mathbf{A}) \Leftrightarrow \mathbf{A}\mathbf{A}^+\mathbf{B} = \mathbf{B} \Leftrightarrow \mathbf{E}_\mathbf{A}\mathbf{B} = \mathbf{0}$.
- (b) $r \begin{bmatrix} \mathbf{A} \\ \mathbf{C} \end{bmatrix} = r(\mathbf{A}) \Leftrightarrow \mathcal{R}(\mathbf{C}') \subseteq \mathcal{R}(\mathbf{A}') \Leftrightarrow \mathbf{C}\mathbf{A}^+\mathbf{A} = \mathbf{C} \Leftrightarrow \mathbf{C}\mathbf{F}_\mathbf{A} = \mathbf{0}$.
- (c) $r[\mathbf{A}, \mathbf{B}] = r(\mathbf{A}) + r(\mathbf{B}) \Leftrightarrow \mathcal{R}(\mathbf{A}) \cap \mathcal{R}(\mathbf{B}) = \{\mathbf{0}\} \Leftrightarrow \mathcal{R}[(\mathbf{E}_\mathbf{A}\mathbf{B})'] = \mathcal{R}(\mathbf{B}') \Leftrightarrow \mathcal{R}[(\mathbf{E}_\mathbf{B}\mathbf{A})'] = \mathcal{R}(\mathbf{A}')$.
- (d) $r \begin{bmatrix} \mathbf{A} \\ \mathbf{C} \end{bmatrix} = r(\mathbf{A}) + r(\mathbf{C}) \Leftrightarrow \mathcal{R}(\mathbf{A}') \cap \mathcal{R}(\mathbf{C}') = \{\mathbf{0}\} \Leftrightarrow \mathcal{R}(\mathbf{C}\mathbf{F}_\mathbf{A}) = \mathcal{R}(\mathbf{C}) \Leftrightarrow \mathcal{R}(\mathbf{A}\mathbf{F}_\mathbf{C}) = \mathcal{R}(\mathbf{A})$.
- (e) $r \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{0} \end{bmatrix} = r(\mathbf{B}) + r(\mathbf{C}) \Leftrightarrow \mathbf{E}_\mathbf{B}\mathbf{A}\mathbf{F}_\mathbf{C} = \mathbf{0}$.
- (f) $r(\mathbf{A} + \mathbf{B}) = r(\mathbf{A}) + r(\mathbf{B}) \Leftrightarrow \mathcal{R}(\mathbf{A}) \cap \mathcal{R}(\mathbf{B}) = \{\mathbf{0}\}$ and $\mathcal{R}(\mathbf{A}') \cap \mathcal{R}(\mathbf{B}') = \{\mathbf{0}\}$.

Lemma 2.3 ([14]). If $\mathcal{R}(\mathbf{A}'_1) \subseteq \mathcal{R}(\mathbf{B}'_1)$, $\mathcal{R}(\mathbf{A}_2) \subseteq \mathcal{R}(\mathbf{B}_1)$, $\mathcal{R}(\mathbf{A}'_2) \subseteq \mathcal{R}(\mathbf{B}'_2)$, and $\mathcal{R}(\mathbf{A}_3) \subseteq \mathcal{R}(\mathbf{B}_2)$, then

$$r(\mathbf{A}_1\mathbf{B}_1^+\mathbf{A}_2) = r \begin{bmatrix} \mathbf{B}_1 & \mathbf{A}_2 \\ \mathbf{A}_1 & \mathbf{0} \end{bmatrix} - r(\mathbf{B}_1), \quad (16)$$

$$r(\mathbf{A}_1\mathbf{B}_1^+\mathbf{A}_2\mathbf{B}_2^+\mathbf{A}_3) = r \begin{bmatrix} \mathbf{0} & \mathbf{B}_2 & \mathbf{A}_3 \\ \mathbf{B}_1 & \mathbf{A}_2 & \mathbf{0} \\ \mathbf{A}_1 & \mathbf{0} & \mathbf{0} \end{bmatrix} - r(\mathbf{B}_1) - r(\mathbf{B}_2). \quad (17)$$

The results collected in the following lemma are obvious or well known.

Lemma 2.4. Let $\mathbf{A} = \mathbf{A}' \in \mathbb{R}^{m \times m}$, $\mathbf{B} = \mathbf{B}' \in \mathbb{R}^{n \times n}$, $\mathbf{Q} \in \mathbb{R}^{m \times n}$, and assume that $\mathbf{P} \in \mathbb{R}^{m \times m}$ is nonsingular. Then

$$r(\mathbf{A}) = i_+(\mathbf{A}) + i_-(\mathbf{A}), \quad (18)$$

$$i_\pm(\mathbf{P}\mathbf{A}\mathbf{P}') = i_\pm(\mathbf{A}) \quad (\text{Sylvester's law of inertia}), \quad (19)$$

$$i_\pm(\mathbf{A}^+) = i_\pm(\mathbf{A}), \quad i_\pm(-\mathbf{A}) = i_\mp(\mathbf{A}), \quad (20)$$

$$i_\pm \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{bmatrix} = i_\pm(\mathbf{A}) + i_\pm(\mathbf{B}), \quad (21)$$

$$i_+ \begin{bmatrix} \mathbf{0} & \mathbf{Q} \\ \mathbf{Q}' & \mathbf{0} \end{bmatrix} = i_- \begin{bmatrix} \mathbf{0} & \mathbf{Q} \\ \mathbf{Q}' & \mathbf{0} \end{bmatrix} = r(\mathbf{Q}). \quad (22)$$

Lemma 2.5 ([4]). Let $\mathbf{A} = \mathbf{A}' \in \mathbb{R}^{m \times m}$, $\mathbf{B} \in \mathbb{R}^{m \times n}$, and $\mathbf{D} = \mathbf{D}' \in \mathbb{R}^{n \times n}$. Then

$$i_\pm \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}' & \mathbf{0} \end{bmatrix} = r(\mathbf{B}) + i_\pm(\mathbf{E}_\mathbf{B}\mathbf{A}\mathbf{E}_\mathbf{B}), \quad r \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}' & \mathbf{0} \end{bmatrix} = 2r(\mathbf{B}) + r(\mathbf{E}_\mathbf{B}\mathbf{A}\mathbf{E}_\mathbf{B}), \quad (23)$$

$$i_\pm \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}' & \mathbf{D} \end{bmatrix} = i_\pm(\mathbf{A}) + i_\pm \begin{bmatrix} \mathbf{0} & \mathbf{E}_\mathbf{A}\mathbf{B} \\ \mathbf{B}'\mathbf{E}_\mathbf{A} & \mathbf{D} - \mathbf{B}'\mathbf{A} + \mathbf{B} \end{bmatrix}. \quad (24)$$

In particular,

$$i_+ \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}' & \mathbf{0} \end{bmatrix} = r[\mathbf{A}, \mathbf{B}], \quad i_- \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}' & \mathbf{0} \end{bmatrix} = r(\mathbf{B}), \quad r \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}' & \mathbf{0} \end{bmatrix} = r[\mathbf{A}, \mathbf{B}] + r(\mathbf{B}), \quad \text{if } \mathbf{A} \succeq \mathbf{0}, \quad (25)$$

$$i_{\pm} \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}' & \mathbf{D} \end{bmatrix} = i_{\pm}(\mathbf{A}) + i_{\pm}(\mathbf{D} - \mathbf{B}'\mathbf{A}^+\mathbf{B}) \quad \text{if } \mathcal{R}(\mathbf{B}) \subseteq \mathcal{R}(\mathbf{A}). \quad (26)$$

Lemma 2.6 ([4]). Let $\mathbf{A} = \mathbf{A}' \in \mathbb{R}^{m \times m}$, $\mathbf{B} \in \mathbb{R}^{q \times n}$, $\mathbf{D} = \mathbf{D}' \in \mathbb{R}^{n \times n}$, $\mathbf{P} \in \mathbb{R}^{q \times m}$ with $\mathcal{R}(\mathbf{A}) \subseteq \mathcal{R}(\mathbf{P}')$ and $\mathcal{R}(\mathbf{B}) \subseteq \mathcal{R}(\mathbf{P})$. Also let $\mathbf{M} = \begin{bmatrix} -\mathbf{A} & \mathbf{P}' & \mathbf{0} \\ \mathbf{P} & \mathbf{0} & \mathbf{B} \\ \mathbf{0} & \mathbf{B}' & \mathbf{D} \end{bmatrix}$. Then

$$i_{\pm}[\mathbf{D} - \mathbf{B}'(\mathbf{P}')^+\mathbf{A}\mathbf{P}^+\mathbf{B}] = i_{\pm}(\mathbf{M}) - r(\mathbf{P}), \quad (27)$$

$$r[\mathbf{D} - \mathbf{B}'(\mathbf{P}')^+\mathbf{A}\mathbf{P}^+\mathbf{B}] = r(\mathbf{M}) - 2r(\mathbf{P}). \quad (28)$$

Hence,

$$\mathbf{B}'(\mathbf{P}')^+\mathbf{A}\mathbf{P}^+\mathbf{B} = \mathbf{D} \Leftrightarrow r(\mathbf{M}) = 2r(\mathbf{P}),$$

$$\mathbf{B}'(\mathbf{P}')^+\mathbf{A}\mathbf{P}^+\mathbf{B} \succeq \mathbf{D} \Leftrightarrow i_+(\mathbf{M}) = r(\mathbf{P}),$$

$$\mathbf{B}'(\mathbf{P}')^+\mathbf{A}\mathbf{P}^+\mathbf{B} \preceq \mathbf{D} \Leftrightarrow i_-(\mathbf{M}) = r(\mathbf{P}).$$

Lemma 2.7 ([15]). The linear matrix equation $\mathbf{A}\mathbf{X} = \mathbf{B}$ is consistent if and only if $r[\mathbf{A}, \mathbf{B}] = r(\mathbf{A})$, or equivalently, $\mathbf{A}\mathbf{A}^+\mathbf{B} = \mathbf{B}$. In this case, the general solution of the equation can be written in the parametric form $\mathbf{X} = \mathbf{A}^+\mathbf{B} + \mathbf{F}_\mathbf{A}\mathbf{U}$, where \mathbf{U} is an arbitrary matrix.

In statistical inference of parametric regression models, the unknown parameters in them are usually predicted/estimated by various optimization methods or algorithms. A brief survey on modern optimization methods in statistical analysis can be found in [16]. In any case, we expect that the optimization problems occurred in parameter predictions/estimations under a GLM have analytical solutions, so that we can use the analytical solutions to establish a perfect theory on the statistical inference of the GLM. The notion of BLUP was well established in the literature, but it is not easy to believe that there exist analytical results on BLUPs of a very general nature due to the lack of explicit solutions of BLUPs' optimization problems. The present author recently developed an algebraic method of solving quadratic matrix-valued function optimization problems in [17], and used the method to derive many new analytical matrix equations and formulas of BLUPs of all unknown parameters in GLMs with random effects. By using this optimization method, we are now able to directly obtain exact formulas for calculating the BLUPs of $\boldsymbol{\phi}$ in (2) and their dispersion matrices, and to use the formulas in dealing with various statistical inference problems under general assumptions. In order to directly solve the matrix minimization problem associated with the BLUPs under a GLM, we need the following known result on analytical solutions of a constrained quadratic matrix-valued function minimization problem.

Lemma 2.8 ([17]). Let

$$f(\mathbf{L}) = (\mathbf{L} + \mathbf{C})\mathbf{M}(\mathbf{L} + \mathbf{C})' \quad \text{s.t. } \mathbf{L}\mathbf{A} = \mathbf{B},$$

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

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

holds for all solutions of $\mathbf{L}\mathbf{A} = \mathbf{B}$, while the matrix \mathbf{L}_0 satisfying the inequality is determined by the following consistent matrix equation

$$\mathbf{L}_0[\mathbf{A}, \mathbf{M}\mathbf{A}^\perp] = [\mathbf{B}, -\mathbf{C}\mathbf{M}\mathbf{A}^\perp].$$

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

$$\mathbf{L}_0 = \underset{\mathbf{L}\mathbf{A}=\mathbf{B}}{\operatorname{argmin}} f(\mathbf{L}) = [\mathbf{B}, -\mathbf{C}\mathbf{M}\mathbf{A}^\perp][\mathbf{A}, \mathbf{M}\mathbf{A}^\perp]^+ + \mathbf{U}[\mathbf{A}, \mathbf{M}]^\perp,$$

$$f(\mathbf{L}_0) = \min_{\mathbf{L}\mathbf{A}=\mathbf{B}} f(\mathbf{L}) = \mathbf{K}\mathbf{M}\mathbf{K}' - \mathbf{K}\mathbf{M}(\mathbf{A}^\perp\mathbf{M}\mathbf{A}^\perp)^+ \mathbf{M}\mathbf{K}',$$

$$f(\mathbf{L}) - f(\mathbf{L}_0) = (\mathbf{L}\mathbf{M}\mathbf{A}^\perp + \mathbf{C}\mathbf{M}\mathbf{A}^\perp)(\mathbf{A}^\perp\mathbf{M}\mathbf{A}^\perp)^+ (\mathbf{L}\mathbf{M}\mathbf{A}^\perp + \mathbf{C}\mathbf{M}\mathbf{A}^\perp)',$$

where $\mathbf{K} = \mathbf{B}\mathbf{A}^+ + \mathbf{C}$ and $\mathbf{U} \in \mathbb{R}^{n \times p}$ is arbitrary.

The assertions in this lemma show that a clear way of obtaining analytical solutions of a typical constrained quadratic matrix-valued function minimization problem is established, and all the properties and features of the minimization problem can be discovered from the analytical solutions. With the supports of the assertions in Lemma 2.1 and the rank/inertia formulas in Lemmas 2.2–2.6, we are able to convert many inference problems in statistics into algebraic problems of characterizing matrix equalities and inequalities composed by matrices and their generalized inverses, and to derive, as demonstrated in Sections 4–6 below, analytical solutions of the problems by using the methods of matrix equations, matrix rank formulas, and various tricky partitioned matrix calculations.

3 Formulas and properties of BLUPs

In what follows, we assume that (1) is consistent, that is, $\mathbf{y} \in \mathcal{R}[\mathbf{X}, \boldsymbol{\Sigma}]$ holds with probability 1; see [18, 19].

Notice from (3) that the BLUP of $\boldsymbol{\phi}$ in (2) is defined from the minimization of the dispersion matrix of $\mathbf{L}\mathbf{y} - \boldsymbol{\phi}$ subject to $\mathbf{E}(\mathbf{L}\mathbf{y} - \boldsymbol{\phi}) = \mathbf{0}$. The advantage of the minimization property of the BLUP has created many problems in statistical inference of GLMs. In particular, the minimum dispersion matrix of $\mathbf{L}\mathbf{y} - \boldsymbol{\phi}$ can be utilized to compare the optimality and efficiency of other types of predictor of $\boldsymbol{\phi}$ under (1). In fact, BLUPs are primary choices in all possible predictors due to their simple and optimality properties, and have wide applications in both pure and applied disciplines of statistical inference. The theory of BLUPs under GLMs belongs to the classical methods of mathematical statistics, and has been core issues of researches in the field of statistics and applications. It should be pointed out that (3) can equivalently be assembled as certain constrained matrix-valued function optimization problem in the Löwner partial ordering. This kind of equivalences between dispersion matrix minimization problems and matrix-valued function minimization problems were firstly characterized in [11]; see also, [20, 21]. Along with some new development of optimization methods in matrix theory, it is now easy to deal with various complicated matrix operations occurring in the statistical inference of (2).

We next show how to translate the above statistical problems under GLMs into mathematical problems on matrix analysis, and solve the problems by using various results and methods in matrix algebra. Under (1) and (2),

$$\mathbf{E}(\boldsymbol{\phi}) = \mathbf{K}\boldsymbol{\beta}, \quad \mathbf{D}(\boldsymbol{\phi}) = \sigma^2 \mathbf{J}\boldsymbol{\Sigma}\mathbf{J}', \quad \text{Cov}\{\boldsymbol{\phi}, \mathbf{y}\} = \sigma^2 \mathbf{J}\boldsymbol{\Sigma}, \quad (29)$$

while $\mathbf{L}\mathbf{y} - \boldsymbol{\phi}$ can be rewritten as

$$\mathbf{L}\mathbf{y} - \boldsymbol{\phi} = \mathbf{L}\mathbf{X}\boldsymbol{\beta} + \mathbf{L}\boldsymbol{\varepsilon} - \mathbf{K}\boldsymbol{\beta} - \mathbf{J}\boldsymbol{\varepsilon} = (\mathbf{L}\mathbf{X} - \mathbf{K})\boldsymbol{\beta} + (\mathbf{L} - \mathbf{J})\boldsymbol{\varepsilon}. \quad (30)$$

Then, the expectation of $\mathbf{L}\mathbf{y} - \boldsymbol{\phi}$ can be expressed as

$$\mathbf{E}(\mathbf{L}\mathbf{y} - \boldsymbol{\phi}) = \mathbf{E}[(\mathbf{L}\mathbf{X} - \mathbf{K})\boldsymbol{\beta}] = (\mathbf{L}\mathbf{X} - \mathbf{K})\boldsymbol{\beta}; \quad (31)$$

the matrix mean square error of $\mathbf{L}\mathbf{y} - \boldsymbol{\phi}$ is

$$\mathbf{E}[(\mathbf{L}\mathbf{y} - \boldsymbol{\phi})(\mathbf{L}\mathbf{y} - \boldsymbol{\phi})'] = (\mathbf{L}\mathbf{X} - \mathbf{K})\boldsymbol{\beta}\boldsymbol{\beta}'(\mathbf{L}\mathbf{X} - \mathbf{K})' + \sigma^2(\mathbf{L} - \mathbf{J})\boldsymbol{\Sigma}(\mathbf{L} - \mathbf{J})'; \quad (32)$$

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

$$\mathbf{D}(\mathbf{L}\mathbf{y} - \boldsymbol{\phi}) = \sigma^2(\mathbf{L} - \mathbf{J})\boldsymbol{\Sigma}(\mathbf{L} - \mathbf{J})' \triangleq \sigma^2 f(\mathbf{L}). \quad (33)$$

Hence, the constrained covariance matrix minimization problem in (3) converts to a mathematical problem of minimizing the quadratic matrix-valued function $f(\mathbf{L})$ subject to $(\mathbf{L}\mathbf{X} - \mathbf{K})\boldsymbol{\beta} = \mathbf{0}$. A general method for solving this kind of matrix optimization problems in the Löwner partial ordering was formulated in Lemma 2.8. In particular, the following comprehensive results were established in [22]; see also [23].

Theorem 3.1. The vector ϕ in (2) is predictable by y in (1) if and only if

$$\mathcal{R}(\mathbf{X}') \supseteq \mathcal{R}(\mathbf{K}'). \quad (34)$$

Proof. It follows from (31) that $E(\mathbf{L}y - \phi) = \mathbf{0} \Leftrightarrow (\mathbf{L}\mathbf{X} - \mathbf{K})\beta = \mathbf{0}$ for all $\beta \Leftrightarrow \mathbf{L}\mathbf{X} = \mathbf{K}$. From Lemma 2.7, the matrix equation is consistent if and only if (34) holds. \square

Theorem 3.2 (Fundamental BLUP equation). Assume that ϕ in (2) is predictable under (1). Then

$$E(\mathbf{L}y - \phi) = \mathbf{0} \text{ and } D(\mathbf{L}y - \phi) = \min \Leftrightarrow \mathbf{L}[\mathbf{X}, \Sigma\mathbf{X}^\perp] = [\mathbf{K}, \mathbf{J}\Sigma\mathbf{X}^\perp]. \quad (35)$$

The equation in (35), called the fundamental BLUP equation, is consistent, i.e.,

$$[\mathbf{K}, \mathbf{J}\Sigma\mathbf{X}^\perp][\mathbf{X}, \Sigma\mathbf{X}^\perp]^+[\mathbf{X}, \Sigma\mathbf{X}^\perp] = [\mathbf{K}, \mathbf{J}\Sigma\mathbf{X}^\perp] \quad (36)$$

holds under (34), while the general expression of \mathbf{L} and the corresponding $\text{BLUP}(\phi)$ are given by

$$\text{BLUP}(\phi) = \mathbf{L}y = \left([\mathbf{K}, \mathbf{J}\Sigma\mathbf{X}^\perp][\mathbf{X}, \Sigma\mathbf{X}^\perp]^+ + \mathbf{U}[\mathbf{X}, \Sigma\mathbf{X}^\perp]^\perp \right) y, \quad (37)$$

where $\mathbf{U} \in \mathbb{R}^{k \times n}$ is arbitrary. In particular,

$$\text{BLUE}(\mathbf{K}\beta) = \left([\mathbf{K}, \mathbf{0}][\mathbf{X}, \Sigma\mathbf{X}^\perp]^+ + \mathbf{U}_1[\mathbf{X}, \Sigma\mathbf{X}^\perp]^\perp \right) y, \quad (38)$$

$$\text{BLUP}(\mathbf{J}\epsilon) = \left([\mathbf{0}, \mathbf{J}\Sigma\mathbf{X}^\perp][\mathbf{X}, \Sigma\mathbf{X}^\perp]^+ + \mathbf{U}_2[\mathbf{X}, \Sigma\mathbf{X}^\perp]^\perp \right) y, \quad (39)$$

where $\mathbf{U}_1, \mathbf{U}_2 \in \mathbb{R}^{k \times n}$ are arbitrary. Furthermore, the following results hold.

- (a) $r[\mathbf{X}, \Sigma\mathbf{X}^\perp] = r[\mathbf{X}, \Sigma]$, $\mathcal{R}[\mathbf{X}, \Sigma\mathbf{X}^\perp] = \mathcal{R}[\mathbf{X}, \Sigma]$, and $\mathcal{R}(\mathbf{X}) \cap \mathcal{R}(\Sigma\mathbf{X}^\perp) = \{\mathbf{0}\}$.
- (b) \mathbf{L} is unique if and only if $r[\mathbf{X}, \Sigma] = n$.
- (c) $\text{BLUP}(\phi)$ is unique with probability 1 if and only if (1) is consistent.
- (d) The dispersion matrices of $\text{BLUP}(\phi)$ and $\phi - \text{BLUP}(\phi)$, as well as the covariance matrix between $\text{BLUP}(\phi)$ and ϕ are unique, and satisfy the following equalities

$$D[\text{BLUP}(\phi)] = \sigma^2[\mathbf{K}, \mathbf{J}\Sigma\mathbf{X}^\perp][\mathbf{X}, \Sigma\mathbf{X}^\perp]^+ \Sigma ([\mathbf{K}, \mathbf{J}\Sigma\mathbf{X}^\perp][\mathbf{X}, \Sigma\mathbf{X}^\perp]^+)', \quad (40)$$

$$\text{Cov}\{\text{BLUP}(\phi), \phi\} = \sigma^2[\mathbf{K}, \mathbf{J}\Sigma\mathbf{X}^\perp][\mathbf{X}, \Sigma\mathbf{X}^\perp]^+ \Sigma \mathbf{J}', \quad (41)$$

$$D(\phi) - D[\text{BLUP}(\phi)] = \sigma^2 \mathbf{J} \Sigma \mathbf{J}' - \sigma^2[\mathbf{K}, \mathbf{J}\Sigma\mathbf{X}^\perp][\mathbf{X}, \Sigma\mathbf{X}^\perp]^+ \Sigma ([\mathbf{K}, \mathbf{J}\Sigma\mathbf{X}^\perp][\mathbf{X}, \Sigma\mathbf{X}^\perp]^+)', \quad (42)$$

$$D[\phi - \text{BLUP}(\phi)] = \sigma^2 \left(\mathbf{J} - [\mathbf{K}, \mathbf{J}\Sigma\mathbf{X}^\perp][\mathbf{X}, \Sigma\mathbf{X}^\perp]^+ \right) \Sigma \left(\mathbf{J} - [\mathbf{K}, \mathbf{J}\Sigma\mathbf{X}^\perp][\mathbf{X}, \Sigma\mathbf{X}^\perp]^+ \right)'. \quad (43)$$

- (e) $\text{BLUP}(\phi)$, $\text{BLUE}(\mathbf{K}\beta)$, and $\text{BLUP}(\mathbf{J}\epsilon)$ satisfy

$$\text{BLUP}(\phi) = \text{BLUE}(\mathbf{K}\beta) + \text{BLUP}(\mathbf{J}\epsilon), \quad (44)$$

$$\text{Cov}\{\text{BLUE}(\mathbf{K}\beta), \text{BLUP}(\mathbf{J}\epsilon)\} = \mathbf{0}, \quad (45)$$

$$D[\text{BLUP}(\phi)] = D[\text{BLUE}(\mathbf{K}\beta)] + D[\text{BLUP}(\mathbf{J}\epsilon)]. \quad (46)$$

- (f) $\mathbf{T}\phi$ is predictable for any matrix $\mathbf{T} \in \mathbb{R}^{l \times k}$, and $\text{BLUP}(\mathbf{T}\phi) = \mathbf{T}\text{BLUP}(\phi)$ holds.

- (g) In particular,

$$\text{BLUE}(\mathbf{X}\beta) = \left([\mathbf{X}, \mathbf{0}][\mathbf{X}, \Sigma\mathbf{X}^\perp]^+ + \mathbf{U}[\mathbf{X}, \Sigma\mathbf{X}^\perp]^\perp \right) y, \quad (47)$$

$$D[\text{BLUE}(\mathbf{X}\beta)] = \sigma^2[\mathbf{X}, \mathbf{0}][\mathbf{X}, \Sigma\mathbf{X}^\perp]^+ \Sigma ([\mathbf{X}, \mathbf{0}][\mathbf{X}, \Sigma\mathbf{X}^\perp]^+)', \quad (48)$$

where $\mathbf{U} \in \mathbb{R}^{n \times n}$ is arbitrary, and

$$\text{BLUP}(\epsilon) = \left([\mathbf{0}, \Sigma\mathbf{X}^\perp][\mathbf{X}, \Sigma\mathbf{X}^\perp]^+ + \mathbf{U}[\mathbf{X}, \Sigma\mathbf{X}^\perp]^\perp \right) y = \left(\Sigma(\mathbf{X}^\perp \Sigma \mathbf{X}^\perp)^+ + \mathbf{U}[\mathbf{X}, \Sigma\mathbf{X}^\perp]^\perp \right) y, \quad (49)$$

$$\text{Cov}\{\text{BLUP}(\epsilon), \epsilon\} = D[\text{BLUP}(\epsilon)] = \sigma^2 \Sigma(\mathbf{X}^\perp \Sigma \mathbf{X}^\perp)^+ \Sigma, \quad (50)$$

$$D[\epsilon - \text{BLUP}(\epsilon)] = D(\epsilon) - D[\text{BLUP}(\epsilon)] = \sigma^2 \Sigma - \sigma^2 \Sigma(\mathbf{X}^\perp \Sigma \mathbf{X}^\perp)^+ \Sigma, \quad (51)$$

where $\mathbf{U} \in \mathbb{R}^{n \times n}$ is arbitrary.

(h) \mathbf{y} , $\text{BLUE}(\mathbf{X}\boldsymbol{\beta})$, and $\text{BLUP}(\boldsymbol{\varepsilon})$ satisfy

$$\mathbf{y} = \text{BLUE}(\mathbf{X}\boldsymbol{\beta}) + \text{BLUP}(\boldsymbol{\varepsilon}), \quad (52)$$

$$\text{Cov}\{\text{BLUE}(\mathbf{X}\boldsymbol{\beta}), \text{BLUP}(\boldsymbol{\varepsilon})\} = \mathbf{0}, \quad (53)$$

$$D(\mathbf{y}) = D[\text{BLUE}(\mathbf{X}\boldsymbol{\beta})] + D[\text{BLUP}(\boldsymbol{\varepsilon})]. \quad (54)$$

Proof. Eq. (3) is equivalent to finding a solution \mathbf{L}_0 of $\mathbf{L}_0\mathbf{X} = \mathbf{K}$ such that

$$f(\mathbf{L}) \succcurlyeq f(\mathbf{L}_0) \text{ s.t. } \mathbf{LX} = \mathbf{K} \quad (55)$$

holds in the Löwner partial ordering. From Lemma 2.8, there always exists a solution \mathbf{L}_0 of $\mathbf{L}_0\mathbf{X} = \mathbf{K}$ such that $f(\mathbf{L}) \succcurlyeq f(\mathbf{L}_0)$ holds for all solutions of $\mathbf{LX} = \mathbf{K}$, and the \mathbf{L}_0 is determined by the matrix equation $\mathbf{L}_0[\mathbf{X}, \boldsymbol{\Sigma}\mathbf{X}^\perp] = [\mathbf{K}, \mathbf{J}\boldsymbol{\Sigma}\mathbf{X}^\perp]$, establishing the matrix equation in (35). Solving the matrix equation in (35) yields (37). Eqs. (38) and (39) follow directly from (37).

The three formulas in (a) are well known for the coefficient matrix in (37); see e.g., [1] and [2, p. 123]. Note that $[\mathbf{X}, \boldsymbol{\Sigma}\mathbf{X}^\perp]^\perp = \mathbf{0} \Leftrightarrow [\mathbf{X}, \boldsymbol{\Sigma}\mathbf{X}^\perp][\mathbf{X}, \boldsymbol{\Sigma}\mathbf{X}^\perp]^\perp = \mathbf{I}_n \Leftrightarrow r[\mathbf{X}, \boldsymbol{\Sigma}\mathbf{X}^\perp] = r[\mathbf{X}, \boldsymbol{\Sigma}] = n$. Combining this fact with (37) leads to (b). Setting the term $[\mathbf{X}, \boldsymbol{\Sigma}\mathbf{X}^\perp]^\perp \mathbf{y} = \mathbf{0}$ in (37) leads to (c). Eqs. (40)–(43) follow from (1), (29), (33), and (37).

Rewrite (37) as

$$\begin{aligned} \text{BLUP}(\boldsymbol{\phi}) &= ([\mathbf{K}, \mathbf{J}\boldsymbol{\Sigma}\mathbf{X}^\perp][\mathbf{X}, \boldsymbol{\Sigma}\mathbf{X}^\perp]^\perp + \mathbf{U}[\mathbf{X}, \boldsymbol{\Sigma}\mathbf{X}^\perp]^\perp) \mathbf{y} \\ &= ([\mathbf{K}, \mathbf{0}][\mathbf{X}, \boldsymbol{\Sigma}\mathbf{X}^\perp]^\perp + \mathbf{U}_1[\mathbf{X}, \boldsymbol{\Sigma}\mathbf{X}^\perp]^\perp) \mathbf{y} + ([\mathbf{0}, \mathbf{J}\boldsymbol{\Sigma}\mathbf{X}^\perp][\mathbf{X}, \boldsymbol{\Sigma}\mathbf{X}^\perp]^\perp + \mathbf{U}_2[\mathbf{X}, \boldsymbol{\Sigma}\mathbf{X}^\perp]^\perp) \mathbf{y} \\ &= \text{BLUE}(\mathbf{K}\boldsymbol{\beta}) + \text{BLUP}(\mathbf{J}\boldsymbol{\varepsilon}), \end{aligned}$$

establishing (44).

From (38) and (39), the covariance matrix between $\text{BLUE}(\mathbf{K}\boldsymbol{\beta})$ and $\text{BLUP}(\mathbf{J}\boldsymbol{\varepsilon})$ is

$$\begin{aligned} \text{Cov}\{\text{BLUE}(\mathbf{K}\boldsymbol{\beta}), \text{BLUP}(\mathbf{J}\boldsymbol{\varepsilon})\} \\ &= \sigma^2 ([\mathbf{K}, \mathbf{0}][\mathbf{X}, \boldsymbol{\Sigma}\mathbf{X}^\perp]^\perp + \mathbf{U}_1[\mathbf{X}, \boldsymbol{\Sigma}\mathbf{X}^\perp]^\perp) \boldsymbol{\Sigma} ([\mathbf{0}, \mathbf{J}\boldsymbol{\Sigma}\mathbf{X}^\perp][\mathbf{X}, \boldsymbol{\Sigma}\mathbf{X}^\perp]^\perp + \mathbf{U}_2[\mathbf{X}, \boldsymbol{\Sigma}\mathbf{X}^\perp]^\perp)' \\ &= \sigma^2 [\mathbf{K}, \mathbf{0}][\mathbf{X}, \boldsymbol{\Sigma}\mathbf{X}^\perp]^\perp \boldsymbol{\Sigma} ([\mathbf{0}, \mathbf{J}\boldsymbol{\Sigma}\mathbf{X}^\perp][\mathbf{X}, \boldsymbol{\Sigma}\mathbf{X}^\perp]^\perp)' . \end{aligned}$$

Applying (17) to this matrix and simplifying, we obtain

$$\begin{aligned} &r(\text{Cov}\{\text{BLUE}(\mathbf{K}\boldsymbol{\beta}), \text{BLUP}(\mathbf{J}\boldsymbol{\varepsilon})\}) \\ &= r([[\mathbf{K}, \mathbf{0}][\mathbf{X}, \boldsymbol{\Sigma}\mathbf{X}^\perp]^\perp \boldsymbol{\Sigma} ([\mathbf{0}, \mathbf{J}\boldsymbol{\Sigma}\mathbf{X}^\perp][\mathbf{X}, \boldsymbol{\Sigma}\mathbf{X}^\perp]^\perp)') \\ &= r \begin{bmatrix} \mathbf{0} & \begin{bmatrix} \mathbf{X}' \\ \mathbf{X}^\perp \boldsymbol{\Sigma} \end{bmatrix} \begin{bmatrix} \mathbf{0} \\ \mathbf{X}^\perp \boldsymbol{\Sigma} \mathbf{J}' \end{bmatrix} \\ [\mathbf{X}, \boldsymbol{\Sigma}\mathbf{X}^\perp] & \boldsymbol{\Sigma} & \mathbf{0} \\ [\mathbf{K}, \mathbf{0}] & \mathbf{0} & \mathbf{0} \end{bmatrix} - 2r[\mathbf{X}, \boldsymbol{\Sigma}\mathbf{X}^\perp] \\ &= r \begin{bmatrix} \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{X}^\perp \boldsymbol{\Sigma} \mathbf{X}^\perp \end{bmatrix} \begin{bmatrix} \mathbf{X}' \\ \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{0} \\ \mathbf{X}^\perp \boldsymbol{\Sigma} \mathbf{J}' \end{bmatrix} \\ [\mathbf{X}, \mathbf{0}] & \boldsymbol{\Sigma} & \mathbf{0} \\ [\mathbf{K}, \mathbf{0}] & \mathbf{0} & \mathbf{0} \end{bmatrix} - 2r[\mathbf{X}, \boldsymbol{\Sigma}] \\ &= r \begin{bmatrix} \mathbf{0} & \mathbf{X}' \\ \mathbf{X} & \boldsymbol{\Sigma} \\ \mathbf{K} & \mathbf{0} \end{bmatrix} + r[\mathbf{X}^\perp \boldsymbol{\Sigma} \mathbf{X}^\perp, \mathbf{X}^\perp \boldsymbol{\Sigma} \mathbf{J}'] - 2r[\mathbf{X}, \boldsymbol{\Sigma}] \\ &= r \begin{bmatrix} \mathbf{X} \\ \mathbf{K} \end{bmatrix} + r \begin{bmatrix} \mathbf{X}' \\ \boldsymbol{\Sigma} \end{bmatrix} + r[\mathbf{X}, \boldsymbol{\Sigma}\mathbf{X}^\perp, \boldsymbol{\Sigma}\mathbf{J}'] - r(\mathbf{X}) - 2r[\mathbf{X}, \boldsymbol{\Sigma}] \quad (\text{by (13) and (25)}) \end{aligned}$$

$$\begin{aligned}
&= r(\mathbf{X}) + r[\mathbf{X}, \boldsymbol{\Sigma}] + r[\mathbf{X}, \boldsymbol{\Sigma}] - r(\mathbf{X}) - 2r[\mathbf{X}, \boldsymbol{\Sigma}] \quad (\text{by (34)}) \\
&= 0,
\end{aligned}$$

establishing (45). Eq. (46) follows from (44) and (45). Result (f) follows directly from (37). Results (g) and (h) follow from (38), (39), (d) and (e). \square

Eq. (35) shows that the BLUPs of all unknown parameters in (2) can jointly be determined by a linear matrix equation composed by the two given coefficient matrices \mathbf{K} and \mathbf{J} , the given model matrix \mathbf{X} , the dispersion matrix of the observed random vector \mathbf{y} , and the covariance matrix between $\boldsymbol{\phi}$ and \mathbf{y} . So that we find it convenient to present a simple yet general algebraic treatment of the BLUPs of all unknown parameters in a GLM via a basic linear matrix equation, while the BLUPs have a large number of algebraic and statistical properties and features that are technically convenient from analytical solutions of the matrix equation. Matrix equations and formulas for BLUPs like those in (35) and (37) under GLMs were established in the statistical literature by using various direct and indirect methods, for instance, the BLUE of $\mathbf{K}\boldsymbol{\beta}$ and the BLUP of $\mathbf{J}\boldsymbol{\varepsilon}$, as well as (52) were established separately in [11]. In comparison, the whole results collected in Theorem 3.2 provide an exclusive and unified theory about BLUPs and BLUEs of parameter spaces and their magnificent properties and features under GLMs. As demonstrated in [22, 23], the results in Theorem 3.2 can serve as basic and useful references being applied in the statistical inference of GLMs.

From the fundamental matrix equation and formulas in Theorem 3.2, we are now able to derive many new and valuable consequences on properties of BLUPs of parameter spaces in GLMs under various assumptions. For instance, one of the well-known special cases of (1) is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad E(\boldsymbol{\varepsilon}) = \mathbf{0}, \quad D(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}_n, \quad (56)$$

where σ^2 is unknown positive scalar. In this setting, the BLUP and the OLSP of $\boldsymbol{\phi}$ in (2) coincide under (56), while Theorem 3.2 reduces to the following results.

Corollary 3.3. Assume that $\boldsymbol{\phi}$ is predictable under (56), i.e., $\mathcal{R}(\mathbf{X}') \supseteq \mathcal{R}(\mathbf{K}')$ holds. Then

$$E(\mathbf{L}\mathbf{y} - \boldsymbol{\phi}) = \mathbf{0} \text{ and } D(\mathbf{L}\mathbf{y} - \boldsymbol{\phi}) = \min \Leftrightarrow \mathbf{L}[\mathbf{X}, \mathbf{X}^\perp] = [\mathbf{K}, \mathbf{J}\mathbf{X}^\perp].$$

The matrix equation on the right-hand side is consistent as well, and the unique solution of \mathbf{L} and the corresponding $\mathbf{L}\mathbf{y}$ are given by

$$\text{BLUP}(\boldsymbol{\phi}) = \mathbf{L}\mathbf{y} = (\mathbf{K}\mathbf{X}^+ + \mathbf{J}\mathbf{X}^\perp) \mathbf{y}.$$

In particular,

$$\begin{aligned}
\text{BLUE}(\mathbf{K}\boldsymbol{\beta}) &= \mathbf{K}\mathbf{X}^+ \mathbf{y}, \quad \text{BLUP}(\mathbf{J}\boldsymbol{\varepsilon}) = \mathbf{J}\mathbf{X}^\perp \mathbf{y}, \\
\text{BLUE}(\mathbf{X}\boldsymbol{\beta}) &= \mathbf{X}\mathbf{X}^+ \mathbf{y}, \quad \text{BLUP}(\boldsymbol{\varepsilon}) = \mathbf{X}^\perp \mathbf{y}.
\end{aligned}$$

Furthermore, the following results hold.

(a) $\text{BLUP}(\boldsymbol{\phi})$ satisfies the following covariance matrix equalities

$$\begin{aligned}
D[\text{BLUP}(\boldsymbol{\phi})] &= \sigma^2 (\mathbf{K}\mathbf{X}^+ + \mathbf{J}\mathbf{X}^\perp) (\mathbf{K}\mathbf{X}^+ + \mathbf{J}\mathbf{X}^\perp)', \\
\text{Cov}\{\text{BLUP}(\boldsymbol{\phi}), \boldsymbol{\phi}\} &= \sigma^2 (\mathbf{K}\mathbf{X}^+ + \mathbf{J}\mathbf{X}^\perp) \mathbf{J}', \\
D(\boldsymbol{\phi}) - D[\text{BLUP}(\boldsymbol{\phi})] &= \sigma^2 \mathbf{J}\mathbf{J}' - \sigma^2 (\mathbf{K}\mathbf{X}^+ + \mathbf{J}\mathbf{X}^\perp) (\mathbf{K}\mathbf{X}^+ + \mathbf{J}\mathbf{X}^\perp)', \\
D[\boldsymbol{\phi} - \text{BLUP}(\boldsymbol{\phi})] &= \sigma^2 (\mathbf{K}\mathbf{X}^+ - \mathbf{J}\mathbf{P}_\mathbf{X}) (\mathbf{K}\mathbf{X}^+ - \mathbf{J}\mathbf{P}_\mathbf{X})'.
\end{aligned}$$

(b) $\text{BLUP}(\boldsymbol{\phi})$, $\text{BLUE}(\mathbf{K}\boldsymbol{\beta})$, and $\text{BLUP}(\mathbf{J}\boldsymbol{\varepsilon})$ satisfy

$$\begin{aligned}
\text{BLUP}(\boldsymbol{\phi}) &= \text{BLUE}(\mathbf{K}\boldsymbol{\beta}) + \text{BLUP}(\mathbf{J}\boldsymbol{\varepsilon}), \\
\text{Cov}\{\text{BLUE}(\mathbf{K}\boldsymbol{\beta}), \text{BLUP}(\mathbf{J}\boldsymbol{\varepsilon})\} &= \mathbf{0}, \\
D[\text{BLUP}(\boldsymbol{\phi})] &= D[\text{BLUE}(\mathbf{K}\boldsymbol{\beta})] + D[\text{BLUP}(\mathbf{J}\boldsymbol{\varepsilon})].
\end{aligned}$$

- (c) $\text{BLUP}(\mathbf{T}\boldsymbol{\phi}) = \mathbf{T}\text{BLUP}(\boldsymbol{\phi})$ holds for any matrix $\mathbf{T} \in \mathbb{R}^{t \times k}$.
 (d) $\text{BLUE}(\mathbf{X}\boldsymbol{\beta})$ and $\text{BLUP}(\boldsymbol{\varepsilon})$ satisfy

$$\begin{aligned}\text{BLUE}(\mathbf{X}\boldsymbol{\beta}) &= \mathbf{P}_X \mathbf{y}, & \text{BLUP}(\boldsymbol{\varepsilon}) &= (\mathbf{I}_n - \mathbf{P}_X) \mathbf{y}, \\ \text{D}[\text{BLUE}(\mathbf{X}\boldsymbol{\beta})] &= \sigma^2 \mathbf{P}_X, & \text{Cov}\{\text{BLUP}(\boldsymbol{\varepsilon}), \boldsymbol{\varepsilon}\} &= \text{D}[\text{BLUP}(\boldsymbol{\varepsilon})] = \sigma^2 (\mathbf{I}_n - \mathbf{P}_X), \\ \text{D}[\boldsymbol{\varepsilon} - \text{BLUP}(\boldsymbol{\varepsilon})] &= \text{D}(\boldsymbol{\varepsilon}) - \text{D}[\text{BLUP}(\boldsymbol{\varepsilon})] = \sigma^2 \mathbf{P}_X.\end{aligned}$$

- (e) \mathbf{y} , $\text{BLUE}(\mathbf{X}\boldsymbol{\beta})$, and $\text{BLUP}(\boldsymbol{\varepsilon})$ satisfy

$$\begin{aligned}\mathbf{y} &= \text{BLUE}(\mathbf{X}\boldsymbol{\beta}) + \text{BLUP}(\boldsymbol{\varepsilon}), \\ \text{Cov}\{\text{BLUE}(\mathbf{X}\boldsymbol{\beta}), \text{BLUP}(\boldsymbol{\varepsilon})\} &= \mathbf{0}, \quad \text{D}(\mathbf{y}) = \text{D}[\text{BLUE}(\mathbf{X}\boldsymbol{\beta})] + \text{D}[\text{BLUP}(\boldsymbol{\varepsilon})].\end{aligned}$$

4 Rank/inertia formulas for dispersion matrices of BLUPs

Once predictors/estimators of parameter spaces in GLMs are established, more attention is paid to investigating algebraical and statistical properties and features of the predictors/estimators. Since BLUPs/BLEs are fundamental statistical methodology to predict and estimate unknown parameters under GLMs, they play an important role in statistical inference theory, and are often taken as cornerstones for comparing the efficiency of different predictors/estimators due to the minimization property of the BLUPs'/BLEs' dispersion matrices. As demonstrated in Section 3, we are now able to give exact expressions of BLUPs/BLEs under GLMs, so that we can derive various algebraical and statistical properties of BLUPs/BLEs and utilize the properties in the inference of GLMs. Since dispersion matrix of random vector is a conceptual foundation in statistical analysis and inference, statisticians are interested in studying dispersion matrices of predictors/estimators and their algebraic properties. Some previous and present work on the dispersion matrices of BLUPs/BLEs and their properties under GLMs can be found, e.g., in [24–27]. As is known to all, equalities and inequalities of dispersion matrices of predictors/estimators under GLMs play an essential role in characterizing behaviors of the predictors/estimators. Once certain equalities and inequalities are established for dispersion matrices of predictors/estimators under various assumptions, we can use them to describe performances of the predictors/estimators. This is, however, not an easy task from both mathematical and statistical points of view, because dispersion matrices of predictors/estimators often include various complicated matrix operations of given matrices and their generalized inverses in GLMs, as formulated in (40)–(43). In recent years, the theory of matrix ranks and inertias have been introduced to the statistical analysis of GLMs. We are able to establish various equalities and inequalities for dispersion matrices of predictors/estimators under GLMs by using the matrix rank/inertia methodology; see [5, 6, 9]. Note from (3) that $\text{D}[\boldsymbol{\phi} - \text{BLUP}(\boldsymbol{\phi})]$ has the smallest dispersion matrix among all $\mathbf{L}\mathbf{y} - \boldsymbol{\phi}$ with $\text{E}(\mathbf{L}\mathbf{y} - \boldsymbol{\phi}) = \mathbf{0}$. So that the dispersion matrix plays a key role in characterizing performances of the BLUP of $\boldsymbol{\phi}$ in (2). In order to establish possible equalities and inequalities for dispersion matrices of BLUPs, we first establish three basic formulas for calculating the rank/inertia described in (9).

Theorem 4.1. Assume that $\boldsymbol{\phi}$ in (2) is predictable under (1), and let $\text{BLUP}(\boldsymbol{\phi})$ be as given in (37). Also let $\mathbf{A} = \mathbf{A}' \in \mathbb{R}^{k \times k}$, and denote

$$\mathbf{M} = \begin{bmatrix} \sigma^2 \boldsymbol{\Sigma} & \mathbf{0} & \mathbf{X} \\ \mathbf{0} & -\mathbf{A} & \mathbf{K} - \mathbf{J}\mathbf{X} \\ \mathbf{X}' & \mathbf{K}' - \mathbf{X}'\mathbf{J}' & \mathbf{0} \end{bmatrix}.$$

Then

$$i_+(\mathbf{A} - \text{D}[\boldsymbol{\phi} - \text{BLUP}(\boldsymbol{\phi})]) = i_-(\mathbf{M}) - r(\mathbf{X}), \quad (57)$$

$$i_-(\mathbf{A} - \text{D}[\boldsymbol{\phi} - \text{BLUP}(\boldsymbol{\phi})]) = i_+(\mathbf{M}) - r[\mathbf{X}, \boldsymbol{\Sigma}], \quad (58)$$

$$r(\mathbf{A} - \text{D}[\boldsymbol{\phi} - \text{BLUP}(\boldsymbol{\phi})]) = r(\mathbf{M}) - r[\mathbf{X}, \boldsymbol{\Sigma}] - r(\mathbf{X}). \quad (59)$$

In consequence, the following results hold.

- (a) $\text{D}[\boldsymbol{\phi} - \text{BLUP}(\boldsymbol{\phi})] \succ \mathbf{A} \Leftrightarrow i_+(\mathbf{M}) = r[\mathbf{X}, \boldsymbol{\Sigma}] + k$.

- (b) $D[\phi - \text{BLUP}(\phi)] < \mathbf{A} \Leftrightarrow i_-(\mathbf{M}) = r(\mathbf{X}) + k$.
 (c) $D[\phi - \text{BLUP}(\phi)] \geq \mathbf{A} \Leftrightarrow i_-(\mathbf{M}) = r(\mathbf{X})$.
 (d) $D[\phi - \text{BLUP}(\phi)] \leq \mathbf{A} \Leftrightarrow i_+(\mathbf{M}) = r[\mathbf{X}, \Sigma]$.
 (e) $D[\phi - \text{BLUP}(\phi)] = \mathbf{A} \Leftrightarrow r(\mathbf{M}) = r[\mathbf{X}, \Sigma] + r(\mathbf{X})$.

Proof. Note from (43) that

$$\begin{aligned} \mathbf{A} - D[\phi - \text{BLUP}(\phi)] &= \mathbf{A} - \sigma^2 \left([\mathbf{K}, \mathbf{J}\Sigma\mathbf{X}^\perp][\mathbf{X}, \Sigma\mathbf{X}^\perp]^+ - \mathbf{J} \right) \Sigma \left([\mathbf{K}, \mathbf{J}\Sigma\mathbf{X}^\perp][\mathbf{X}, \Sigma\mathbf{X}^\perp]^+ - \mathbf{J} \right)' \\ &= \sigma^2 \left(\sigma^{-2}\mathbf{A} - \left([\mathbf{K}, \mathbf{J}\Sigma\mathbf{X}^\perp][\mathbf{X}, \Sigma\mathbf{X}^\perp]^+ - \mathbf{J} \right) \Sigma \left([\mathbf{K}, \mathbf{J}\Sigma\mathbf{X}^\perp][\mathbf{X}, \Sigma\mathbf{X}^\perp]^+ - \mathbf{J} \right)' \right). \end{aligned} \quad (60)$$

Also note that $\mathcal{R}([\mathbf{K}, \mathbf{J}\Sigma\mathbf{X}^\perp]') \subseteq \mathcal{R}([\mathbf{X}, \Sigma\mathbf{X}^\perp]')$ and $\mathcal{R}(\Sigma) \subseteq \mathcal{R}[\mathbf{X}, \Sigma\mathbf{X}^\perp]$. Then applying (26) to (60) and simplifying by Lemmas 2.4 and 2.5, and congruence operations, we obtain

$$\begin{aligned} i_\pm(\mathbf{A} - D[\phi - \text{BLUP}(\phi)]) &= i_\pm \left(\sigma^{-2}\mathbf{A} - \left([\mathbf{K}, \mathbf{J}\Sigma\mathbf{X}^\perp][\mathbf{X}, \Sigma\mathbf{X}^\perp]^+ - \mathbf{J} \right) \Sigma \left([\mathbf{K}, \mathbf{J}\Sigma\mathbf{X}^\perp][\mathbf{X}, \Sigma\mathbf{X}^\perp]^+ - \mathbf{J} \right)' \right) \\ &= i_\pm \left[\begin{array}{cc} \Sigma & \Sigma([\mathbf{K}, \mathbf{J}\Sigma\mathbf{X}^\perp][\mathbf{X}, \Sigma\mathbf{X}^\perp]^+ - \mathbf{J})' \\ ([\mathbf{K}, \mathbf{J}\Sigma\mathbf{X}^\perp][\mathbf{X}, \Sigma\mathbf{X}^\perp]^+ - \mathbf{J})\Sigma & \sigma^{-2}\mathbf{A} \end{array} \right] - i_\pm(\Sigma) \\ &= i_\pm \left(\left[\begin{array}{cc} \Sigma & -\Sigma\mathbf{J}' \\ -\mathbf{J}\Sigma & \sigma^{-2}\mathbf{A} \end{array} \right] + \left[\begin{array}{cc} \Sigma & 0 \\ 0 & [\mathbf{K}, \mathbf{J}\Sigma\mathbf{X}^\perp] \end{array} \right] \left[\begin{array}{cc} 0 & [\mathbf{X}, \Sigma\mathbf{X}^\perp]' \\ [\mathbf{X}, \Sigma\mathbf{X}^\perp]' & 0 \end{array} \right]^+ \left[\begin{array}{cc} \Sigma & 0 \\ 0 & [\mathbf{K}, \mathbf{J}\Sigma\mathbf{X}^\perp]' \end{array} \right] \right) - i_\pm(\Sigma) \\ &= i_\pm \left[\begin{array}{ccccc} 0 & -\mathbf{X} & -\Sigma\mathbf{X}^\perp & \Sigma & 0 \\ -\mathbf{X}' & 0 & 0 & 0 & \mathbf{K}' \\ -\mathbf{X}^\perp\Sigma & 0 & 0 & 0 & \mathbf{X}^\perp\Sigma\mathbf{J}' \\ \Sigma & 0 & 0 & \Sigma & -\Sigma\mathbf{J}' \\ 0 & \mathbf{K} & \mathbf{J}\Sigma\mathbf{X}^\perp & -\mathbf{J}\Sigma & \sigma^{-2}\mathbf{A} \end{array} \right] - i_\mp \left[\begin{array}{cc} 0 & [\mathbf{X}, \Sigma\mathbf{X}^\perp]' \\ [\mathbf{X}, \Sigma\mathbf{X}^\perp]' & 0 \end{array} \right] - i_\pm(\Sigma) \\ &= i_\pm \left[\begin{array}{ccccc} -\Sigma & -\mathbf{X} & -\Sigma\mathbf{X}^\perp & 0 & \Sigma\mathbf{J}' \\ -\mathbf{X}' & 0 & 0 & 0 & \mathbf{K}' \\ -\mathbf{X}^\perp\Sigma & 0 & 0 & 0 & \mathbf{X}^\perp\Sigma\mathbf{J}' \\ 0 & 0 & 0 & \Sigma & 0 \\ \mathbf{J}\Sigma & \mathbf{K} & \mathbf{J}\Sigma\mathbf{X}^\perp & 0 & \sigma^{-2}\mathbf{A} - \mathbf{J}\Sigma\mathbf{J}' \end{array} \right] - r[\mathbf{X}, \Sigma\mathbf{X}^\perp] - i_\pm(\Sigma) \quad (\text{by (22)}) \\ &= i_\pm \left[\begin{array}{ccccc} -\Sigma & -\mathbf{X} & -\Sigma\mathbf{X}^\perp & \Sigma\mathbf{J}' & \\ -\mathbf{X}' & 0 & 0 & \mathbf{K}' & \\ -\mathbf{X}^\perp\Sigma & 0 & 0 & \mathbf{X}^\perp\Sigma\mathbf{J}' & \\ \mathbf{J}\Sigma & \mathbf{K} & \mathbf{J}\Sigma\mathbf{X}^\perp & \sigma^{-2}\mathbf{A} - \mathbf{J}\Sigma\mathbf{J}' & \end{array} \right] - r[\mathbf{X}, \Sigma] \quad (\text{by (21)}) \\ &= i_\pm \left[\begin{array}{cccc} -\Sigma & -\mathbf{X} & 0 & 0 \\ -\mathbf{X}' & 0 & 0 & \mathbf{K}' - \mathbf{X}'\mathbf{J}' \\ 0 & 0 & \mathbf{X}^\perp\Sigma\mathbf{X}^\perp & 0 \\ 0 & \mathbf{K} - \mathbf{J}\mathbf{X} & 0 & \sigma^{-2}\mathbf{A} \end{array} \right] - r[\mathbf{X}, \Sigma] \\ &= i_\pm \left[\begin{array}{ccc} -\Sigma & 0 & \mathbf{X} \\ 0 & \sigma^{-2}\mathbf{A} & \mathbf{J}\mathbf{X} - \mathbf{K} \\ \mathbf{X}' & \mathbf{X}'\mathbf{J}' - \mathbf{K}' & 0 \end{array} \right] + i_\pm(\mathbf{X}^\perp\Sigma\mathbf{X}^\perp) - r[\mathbf{X}, \Sigma] \quad (\text{by (21)}) \\ &= i_\mp \left[\begin{array}{ccc} \sigma^2\Sigma & 0 & \mathbf{X} \\ 0 & -\mathbf{A} & \mathbf{K} - \mathbf{J}\mathbf{X} \\ \mathbf{X}' & \mathbf{K}' - \mathbf{X}'\mathbf{J}' & 0 \end{array} \right] + i_\pm(\mathbf{X}^\perp\Sigma\mathbf{X}^\perp) - r[\mathbf{X}, \Sigma] \quad (\text{by (20)}), \end{aligned} \quad (61)$$

that is,

$$\begin{aligned}
i_+(\mathbf{A} - \mathbf{D}[\boldsymbol{\phi} - \text{BLUP}(\boldsymbol{\phi})]) &= i_-(\mathbf{M}) + i_+(\mathbf{X}^\perp \boldsymbol{\Sigma} \mathbf{X}^\perp) - r[\mathbf{X}, \boldsymbol{\Sigma}] = i_-(\mathbf{M}) + r(\mathbf{X}^\perp \boldsymbol{\Sigma}) - r[\mathbf{X}, \boldsymbol{\Sigma}] \\
&= i_-(\mathbf{M}) - r(\mathbf{X}), \\
i_-(\mathbf{A} - \mathbf{D}[\boldsymbol{\phi} - \text{BLUP}(\boldsymbol{\phi})]) &= i_+(\mathbf{M}) + i_-(\mathbf{X}^\perp \boldsymbol{\Sigma} \mathbf{X}^\perp) - r[\mathbf{X}, \boldsymbol{\Sigma}] = i_+(\mathbf{M}) - r[\mathbf{X}, \boldsymbol{\Sigma}]
\end{aligned}$$

by (13), establishing (57) and (58). Adding the two equalities in (57) and (58) yields (59). Applying Lemma 2.1 to (57)–(59) yields (a)–(e). \square

Eqs. (57)–(59) establish links between the dispersion matrices of the BLUPs of $\boldsymbol{\phi}$ and any symmetric matrix. Hence, they can be applied to characterize behaviors of the BLUPs, especially, they can be used to establish many equalities and inequalities for BLUPs' dispersion matrices under various assumptions. Because the five matrices \mathbf{A} , \mathbf{K} , \mathbf{J} , \mathbf{X} , and $\boldsymbol{\Sigma}$ occur separately in the symmetric block matrix \mathbf{M} in Theorem 4.1, it is easy to further simplify (57)–(59) for different choices of the five matrices in the formulas. We next present several special cases of (57)–(59) and give lots of interesting consequences on dispersion matrices of BLUPs/BLUEs and their operations.

Concerning the ranks and inertias of $\mathbf{D}[\boldsymbol{\phi} - \text{BLUP}(\boldsymbol{\phi})]$ and $\mathbf{D}(\boldsymbol{\phi}) - \mathbf{D}[\text{BLUP}(\boldsymbol{\phi})]$ in (42) and (43), we have the following result.

Theorem 4.2. Assume that $\boldsymbol{\phi}$ in (2) is predictable under (1), and let

$$\mathbf{M} = \begin{bmatrix} -\boldsymbol{\Sigma} & \mathbf{0} & \mathbf{X} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma} & \mathbf{0} & \mathbf{X} & \mathbf{0} \\ \mathbf{X}' & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{K}' \\ \mathbf{0} & \mathbf{X}' & \mathbf{0} & \mathbf{0} & \mathbf{X}'\mathbf{J}' \\ \mathbf{0} & \mathbf{0} & \mathbf{K} & \mathbf{J}\mathbf{X} & \mathbf{0} \end{bmatrix}.$$

Then

$$r(\mathbf{D}[\boldsymbol{\phi} - \text{BLUP}(\boldsymbol{\phi})]) = r \begin{bmatrix} \boldsymbol{\Sigma} & \mathbf{X} \\ \mathbf{J}\boldsymbol{\Sigma} & \mathbf{K} \end{bmatrix} - r[\mathbf{X}, \boldsymbol{\Sigma}], \quad (62)$$

and

$$i_{\pm}(\mathbf{D}(\boldsymbol{\phi}) - \mathbf{D}[\text{BLUP}(\boldsymbol{\phi})]) = i_{\pm}(\mathbf{M}) - r(\mathbf{X}) - r[\mathbf{X}, \boldsymbol{\Sigma}], \quad (63)$$

$$r(\mathbf{D}(\boldsymbol{\phi}) - \mathbf{D}[\text{BLUP}(\boldsymbol{\phi})]) = r(\mathbf{M}) - 2r(\mathbf{X}) - 2r[\mathbf{X}, \boldsymbol{\Sigma}]. \quad (64)$$

Proof. Eq. (62) follows from (59) by setting $\mathbf{A} = \mathbf{0}$. Note from (42) that $\mathbf{D}(\boldsymbol{\phi}) - \mathbf{D}[\text{BLUP}(\boldsymbol{\phi})] = \sigma^2 \mathbf{J}\boldsymbol{\Sigma}\mathbf{J}' - \sigma^2 [\mathbf{K}, \mathbf{J}\boldsymbol{\Sigma}\mathbf{X}^\perp][\mathbf{X}, \boldsymbol{\Sigma}\mathbf{X}^\perp]^+ \boldsymbol{\Sigma} ([\mathbf{K}, \mathbf{J}\boldsymbol{\Sigma}\mathbf{X}^\perp][\mathbf{X}, \boldsymbol{\Sigma}\mathbf{X}^\perp]^+)'$. Applying (27) and simplifying by congruence transformations gives

$$\begin{aligned}
i_{\pm}(\mathbf{D}(\boldsymbol{\phi}) - \mathbf{D}[\text{BLUP}(\boldsymbol{\phi})]) &= i_{\pm} \left(\mathbf{J}\boldsymbol{\Sigma}\mathbf{J}' - [\mathbf{K}, \mathbf{J}\boldsymbol{\Sigma}\mathbf{X}^\perp][\mathbf{X}, \boldsymbol{\Sigma}\mathbf{X}^\perp]^+ \boldsymbol{\Sigma} ([\mathbf{K}, \mathbf{J}\boldsymbol{\Sigma}\mathbf{X}^\perp][\mathbf{X}, \boldsymbol{\Sigma}\mathbf{X}^\perp]^+)' \right) \\
&= i_{\pm} \begin{bmatrix} -\boldsymbol{\Sigma} & \mathbf{X} & \boldsymbol{\Sigma}\mathbf{X}^\perp & \mathbf{0} \\ \mathbf{X}' & \mathbf{0} & \mathbf{0} & \mathbf{K}' \\ \mathbf{X}^\perp \boldsymbol{\Sigma} & \mathbf{0} & \mathbf{0} & \mathbf{X}^\perp \boldsymbol{\Sigma}\mathbf{J}' \\ \mathbf{0} & \mathbf{K} & \mathbf{J}\boldsymbol{\Sigma}\mathbf{X}^\perp & \mathbf{J}\boldsymbol{\Sigma}\mathbf{J}' \end{bmatrix} - r[\mathbf{X}, \boldsymbol{\Sigma}\mathbf{X}^\perp] \\
&= i_{\pm} \begin{bmatrix} -\boldsymbol{\Sigma} & \mathbf{X} & \mathbf{0} & \mathbf{0} \\ \mathbf{X}' & \mathbf{0} & \mathbf{0} & \mathbf{K}' \\ \mathbf{0} & \mathbf{0} & \mathbf{X}^\perp \boldsymbol{\Sigma}\mathbf{X}^\perp & \mathbf{X}^\perp \boldsymbol{\Sigma}\mathbf{J}' \\ \mathbf{0} & \mathbf{K} & \mathbf{J}\boldsymbol{\Sigma}\mathbf{X}^\perp & \mathbf{J}\boldsymbol{\Sigma}\mathbf{J}' \end{bmatrix} - r[\mathbf{X}, \boldsymbol{\Sigma}] \\
&= i_{\pm} \begin{bmatrix} -\boldsymbol{\Sigma} & \mathbf{X} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{X}' & \mathbf{0} & \mathbf{0} & \mathbf{K}' & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Sigma} & \boldsymbol{\Sigma}\mathbf{J}' & \mathbf{X} \\ \mathbf{0} & \mathbf{K} & \mathbf{J}\boldsymbol{\Sigma} & \mathbf{J}\boldsymbol{\Sigma}\mathbf{J}' & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{X}' & \mathbf{0} & \mathbf{0} \end{bmatrix} - r(\mathbf{X}) - r[\mathbf{X}, \boldsymbol{\Sigma}] \quad (\text{by (23)})
\end{aligned}$$

$$\begin{aligned}
&= i_{\pm} \begin{bmatrix} -\Sigma & \mathbf{X} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{X}' & \mathbf{0} & \mathbf{0} & \mathbf{K}' & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \Sigma & \mathbf{0} & \mathbf{X} \\ \mathbf{0} & \mathbf{K} & \mathbf{0} & \mathbf{0} & -\mathbf{JX} \\ \mathbf{0} & \mathbf{0} & \mathbf{X}' & -\mathbf{X}'\mathbf{J}' & \mathbf{0} \end{bmatrix} - r(\mathbf{X}) - r[\mathbf{X}, \Sigma] \\
&= i_{\pm} \begin{bmatrix} -\Sigma & \mathbf{0} & \mathbf{X} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Sigma & \mathbf{0} & \mathbf{X} & \mathbf{0} \\ \mathbf{X}' & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{K}' \\ \mathbf{0} & \mathbf{X}' & \mathbf{0} & \mathbf{0} & -\mathbf{X}'\mathbf{J}' \\ \mathbf{0} & \mathbf{0} & \mathbf{K} & -\mathbf{JX} & \mathbf{0} \end{bmatrix} - r(\mathbf{X}) - r[\mathbf{X}, \Sigma] \\
&= i_{\pm} \begin{bmatrix} -\Sigma & \mathbf{0} & \mathbf{X} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Sigma & \mathbf{0} & \mathbf{X} & \mathbf{0} \\ \mathbf{X}' & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{K}' \\ \mathbf{0} & \mathbf{X}' & \mathbf{0} & \mathbf{0} & \mathbf{X}'\mathbf{J}' \\ \mathbf{0} & \mathbf{0} & \mathbf{K} & \mathbf{JX} & \mathbf{0} \end{bmatrix} - r(\mathbf{X}) - r[\mathbf{X}, \Sigma] \\
&= i_{\pm}(\mathbf{M}) - r(\mathbf{X}) - r[\mathbf{X}, \Sigma],
\end{aligned}$$

establishing (63) and (64). \square

Many consequences can be derived from the previous two theorems for different choices of \mathbf{K} , \mathbf{J} , and \mathbf{A} in them. Here, we only give the rank/inertia formulas for the difference $D(\mathbf{A}\mathbf{y}) - D[\boldsymbol{\phi} - \text{BLUP}(\boldsymbol{\phi})]$.

Corollary 4.3. Assume that $\boldsymbol{\phi}$ in (2) is predictable under (1), and let $\text{BLUP}(\boldsymbol{\phi})$ be as given in (37). Also let $\mathbf{A} \in \mathbb{R}^{k \times n}$, and denote

$$\mathbf{M} = \begin{bmatrix} \Sigma & \mathbf{0} & \mathbf{X} \\ \mathbf{0} & -\mathbf{A}\Sigma\mathbf{A}' & \mathbf{K} - \mathbf{JX} \\ \mathbf{X}' & \mathbf{K}' - \mathbf{X}'\mathbf{J}' & \mathbf{0} \end{bmatrix}.$$

Then

$$\begin{aligned}
i_+ (D(\mathbf{A}\mathbf{y}) - D[\boldsymbol{\phi} - \text{BLUP}(\boldsymbol{\phi})]) &= i_-(\mathbf{M}) - r(\mathbf{X}), \\
i_- (D(\mathbf{A}\mathbf{y}) - D[\boldsymbol{\phi} - \text{BLUP}(\boldsymbol{\phi})]) &= i_+(\mathbf{M}) - r[\mathbf{X}, \Sigma], \\
r (D(\mathbf{A}\mathbf{y}) - D[\boldsymbol{\phi} - \text{BLUP}(\boldsymbol{\phi})]) &= r(\mathbf{M}) - r[\mathbf{X}, \Sigma] - r(\mathbf{X}).
\end{aligned}$$

5 Formulas and properties of OLSPs

The method of least squares in statistics is a standard approach for estimating unknown parameters in linear statistical models, which was first proposed as an algebraic procedure for solving overdetermined systems of equations by Gauss (in unpublished work) in 1795 and independently by Legendre in 1805, as remarked in [28–31]. The notion of least-squares estimation is well established in the literature, and we briefly review the derivations of OLSEs and OLSPs. It is easy to verify that the norm $(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$ in (6) can be decomposed as the sum

$$(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{y}'\mathbf{E}_{\mathbf{X}}\mathbf{y} + (\mathbf{P}_{\mathbf{X}}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{P}_{\mathbf{X}}\mathbf{y} - \mathbf{X}\boldsymbol{\beta}),$$

where $\mathbf{y}'\mathbf{E}_{\mathbf{X}}\mathbf{y} \geq 0$ and $(\mathbf{P}_{\mathbf{X}}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{P}_{\mathbf{X}}\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \geq 0$. Hence,

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^{p \times 1}} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{y}'\mathbf{E}_{\mathbf{X}}\mathbf{y} + \min_{\boldsymbol{\beta} \in \mathbb{R}^{p \times 1}} (\mathbf{P}_{\mathbf{X}}\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{P}_{\mathbf{X}}\mathbf{y} - \mathbf{X}\boldsymbol{\beta});$$

see also [7, 8]. The matrix equation $\mathbf{X}\boldsymbol{\beta} = \mathbf{P}_{\mathbf{X}}\mathbf{y}$, which is equivalent to the so-called normal equation $\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{y}$ by pre-multiplying \mathbf{X}' , is always consistent; see, e.g., [32, p. 114] and [33, pp. 164–165]. Solving this linear matrix equation by Lemma 2.7 yields the following general results.

Theorem 5.1. Assume that ϕ in (2) is predictable by \mathbf{y} in (1), and $\mathbf{K}\beta$ is estimable under (1). Then the OLSPs and OLSEs of ϕ and its special cases are given by

$$\text{OLSE}(\mathbf{K}\beta) = \mathbf{KX}^+ \mathbf{y}, \quad (65)$$

$$\text{OLSE}(\mathbf{X}\beta) = \mathbf{XX}^+ \mathbf{y}, \quad (66)$$

$$\text{OLSP}(\epsilon) = (\mathbf{I}_n - \mathbf{XX}^+) \mathbf{y} = \mathbf{X}^\perp \mathbf{y}, \quad (67)$$

$$\text{OLSP}(\mathbf{J}\epsilon) = \mathbf{JX}^\perp \mathbf{y}, \quad (68)$$

$$\text{OLSP}(\phi) = (\mathbf{KX}^+ + \mathbf{JX}^\perp) \mathbf{y}. \quad (69)$$

Furthermore, the following results hold.

(a) $\text{OLSP}(\phi)$ satisfies the following equalities

$$D[\text{OLSP}(\phi)] = \sigma^2 (\mathbf{KX}^+ + \mathbf{JX}^\perp) \Sigma (\mathbf{KX}^+ + \mathbf{JX}^\perp)', \quad (70)$$

$$\text{Cov}\{\text{OLSP}(\phi), \phi\} = \sigma^2 (\mathbf{KX}^+ + \mathbf{JX}^\perp) \Sigma \mathbf{J}', \quad (71)$$

$$D(\phi) - D[\text{OLSP}(\phi)] = \sigma^2 \mathbf{J} \Sigma \mathbf{J}' - \sigma^2 (\mathbf{KX}^+ + \mathbf{JX}^\perp) \Sigma (\mathbf{KX}^+ + \mathbf{JX}^\perp)', \quad (72)$$

$$D[\phi - \text{OLSP}(\phi)] = \sigma^2 (\mathbf{KX}^+ - \mathbf{JP}_X) \Sigma (\mathbf{KX}^+ - \mathbf{JP}_X)'. \quad (73)$$

(b) $\text{OLSP}(\phi)$, $\text{OLSE}(\mathbf{K}\beta)$, and $\text{OLSP}(\mathbf{J}\epsilon)$ satisfy

$$\text{OLSP}(\phi) = \text{OLSE}(\mathbf{K}\beta) + \text{OLSP}(\mathbf{J}\epsilon), \quad (74)$$

$$\text{Cov}\{\text{OLSE}(\mathbf{K}\beta), \text{OLSP}(\mathbf{J}\epsilon)\} = \sigma^2 \mathbf{KX}^+ \Sigma \mathbf{X}^\perp \mathbf{J}'. \quad (75)$$

(c) $\text{OLSP}(\mathbf{T}\phi) = \mathbf{T} \text{OLSP}(\phi)$ holds for any matrix $\mathbf{T} \in \mathbb{R}^{t \times k}$.

(d) \mathbf{y} , $\text{OLSE}(\mathbf{X}\beta)$, and $\text{OLSP}(\epsilon)$ satisfy

$$\mathbf{y} = \text{OLSE}(\mathbf{X}\beta) + \text{OLSP}(\epsilon), \quad (76)$$

$$\text{Cov}\{\text{OLSE}(\mathbf{X}\beta), \text{OLSP}(\epsilon)\} = \sigma^2 \mathbf{P}_X \Sigma \mathbf{X}^\perp, \quad (77)$$

$$D[\text{OLSE}(\mathbf{X}\beta)] = D[\epsilon - \text{OLSP}(\epsilon)] = \sigma^2 \mathbf{P}_X \Sigma \mathbf{P}_X, \quad (78)$$

$$D[\text{OLSP}(\epsilon)] = \sigma^2 \mathbf{X}^\perp \Sigma \mathbf{X}^\perp, \quad (79)$$

$$D(\epsilon) - D[\text{OLSP}(\epsilon)] = \sigma^2 \Sigma - \sigma^2 \mathbf{X}^\perp \Sigma \mathbf{X}^\perp. \quad (80)$$

We next give a group of rank/inertia formulas related to OLSPs and their consequences.

Theorem 5.2. Assume that ϕ in (2) is predictable under (1), and let $\text{OLSP}(\phi)$ be as given in (69). Also let $\mathbf{A} = \mathbf{A}' \in \mathbb{R}^{k \times k}$, and denote

$$\mathbf{M} = \begin{bmatrix} \sigma^2 \mathbf{X}' \Sigma \mathbf{X} & \mathbf{0} & \mathbf{X}' \mathbf{X} \\ \mathbf{0} & -\mathbf{A} & \mathbf{K} - \mathbf{JX} \\ \mathbf{X}' \mathbf{X} & \mathbf{K}' - \mathbf{X}' \mathbf{J}' & \mathbf{0} \end{bmatrix}.$$

Then

$$i_{\pm} \{\mathbf{A} - D[\phi - \text{OLSP}(\phi)]\} = i_{\mp}(\mathbf{M}) - r(\mathbf{X}), \quad (81)$$

$$r \{\mathbf{A} - D[\phi - \text{OLSP}(\phi)]\} = r(\mathbf{M}) - 2r(\mathbf{X}). \quad (82)$$

In consequence, the following results hold.

(a) $D[\phi - \text{OLSP}(\phi)] \succ \mathbf{A} \Leftrightarrow i_+(\mathbf{M}) = r(\mathbf{X}) + k$.

(b) $D[\phi - \text{OLSP}(\phi)] \prec \mathbf{A} \Leftrightarrow i_-(\mathbf{M}) = r(\mathbf{X}) + k$.

(c) $D[\phi - \text{OLSP}(\phi)] \succeq \mathbf{A} \Leftrightarrow i_-(\mathbf{M}) = r(\mathbf{X})$.

(d) $D[\phi - \text{OLSP}(\phi)] \preceq \mathbf{A} \Leftrightarrow i_+(\mathbf{M}) = r(\mathbf{X})$.

$$(e) \ D[\boldsymbol{\phi} - \text{OLSP}(\boldsymbol{\phi})] = \mathbf{A} \Leftrightarrow r(\mathbf{M}) = 2r(\mathbf{X}).$$

Proof. Note from (73) that

$$\begin{aligned} \mathbf{A} - D[\boldsymbol{\phi} - \text{OLSP}(\boldsymbol{\phi})] &= \mathbf{A} - \sigma^2 (\mathbf{K}\mathbf{X}^+ - \mathbf{J}\mathbf{P}_\mathbf{X}) \boldsymbol{\Sigma} (\mathbf{K}\mathbf{X}^+ - \mathbf{J}\mathbf{P}_\mathbf{X})' \\ &= \mathbf{A} - \sigma^2 (\mathbf{K} - \mathbf{J}\mathbf{X}) \mathbf{X}^+ \boldsymbol{\Sigma} (\mathbf{X}')^+ (\mathbf{K}' - \mathbf{X}'\mathbf{J}') \\ &= \mathbf{A} - \sigma^2 (\mathbf{K} - \mathbf{J}\mathbf{X}) (\mathbf{X}'\mathbf{X})^+ \mathbf{X}' \boldsymbol{\Sigma} \mathbf{X} (\mathbf{X}'\mathbf{X})^+ (\mathbf{K}' - \mathbf{X}'\mathbf{J}'). \end{aligned} \quad (83)$$

Applying (27) to (83) gives

$$\begin{aligned} i_{\pm} \{ \mathbf{A} - D[\boldsymbol{\phi} - \text{OLSP}(\boldsymbol{\phi})] \} &= i_{\pm} \left[\mathbf{A} - \sigma^2 (\mathbf{K} - \mathbf{J}\mathbf{X}) (\mathbf{X}'\mathbf{X})^+ \mathbf{X}' \boldsymbol{\Sigma} \mathbf{X} (\mathbf{X}'\mathbf{X})^+ (\mathbf{K}' - \mathbf{X}'\mathbf{J}') \right] \\ &= i_{\pm} \begin{bmatrix} -\sigma^2 \mathbf{X}' \boldsymbol{\Sigma} \mathbf{X} & \mathbf{X}'\mathbf{X} & \mathbf{0} \\ \mathbf{X}'\mathbf{X} & \mathbf{0} & \mathbf{K}' - \mathbf{X}'\mathbf{J}' \\ \mathbf{0} & \mathbf{K} - \mathbf{J}\mathbf{X} & \mathbf{A} \end{bmatrix} - r(\mathbf{X}) \\ &= i_{\mp} \begin{bmatrix} \sigma^2 \mathbf{X}' \boldsymbol{\Sigma} \mathbf{X} & \mathbf{0} & \mathbf{X}'\mathbf{X} \\ \mathbf{0} & -\mathbf{A} & \mathbf{K} - \mathbf{J}\mathbf{X} \\ \mathbf{X}'\mathbf{X} & \mathbf{K}' - \mathbf{X}'\mathbf{J}' & \mathbf{0} \end{bmatrix} - r(\mathbf{X}), \end{aligned}$$

establishing (81) and (82). Applying Lemma 2.1 to (81) and (82) yields the results in (a)–(e). \square

Theorem 5.3. Assume that $\boldsymbol{\phi}$ in (2) is predictable under (1), and let $\text{OLSP}(\boldsymbol{\phi})$ be as given in (69). Then \mathbf{y} , $\text{OLSE}(\mathbf{X}\boldsymbol{\beta})$, and $\text{OLSP}(\boldsymbol{\varepsilon})$ satisfy

$$\begin{aligned} r\{D(\mathbf{y}) - D[\text{OLSE}(\mathbf{X}\boldsymbol{\beta})] - D[\text{OLSP}(\boldsymbol{\varepsilon})]\} &= r\{D[\boldsymbol{\varepsilon} - \text{OLSP}(\boldsymbol{\varepsilon})] - D(\boldsymbol{\varepsilon}) + D[\text{OLSP}(\boldsymbol{\varepsilon})]\} \\ &= 2r[\mathbf{X}, \boldsymbol{\Sigma}\mathbf{X}] - 2r(\mathbf{X}), \end{aligned} \quad (84)$$

$$r(\text{Cov}\{\text{OLSE}(\mathbf{X}\boldsymbol{\beta}), \text{OLSP}(\boldsymbol{\varepsilon})\}) = r[\mathbf{X}, \boldsymbol{\Sigma}\mathbf{X}] - r(\mathbf{X}). \quad (85)$$

Proof. It comes from (78)–(80) that

$$\begin{aligned} D(\mathbf{y}) - D[\text{OLSE}(\mathbf{X}\boldsymbol{\beta})] - D[\text{OLSP}(\boldsymbol{\varepsilon})] &= D(\boldsymbol{\varepsilon}) - D[\boldsymbol{\varepsilon} - \text{OLSP}(\boldsymbol{\varepsilon})] - D[\text{OLSP}(\boldsymbol{\varepsilon})] \\ &= \text{Cov}\{\text{OLSE}(\mathbf{X}\boldsymbol{\beta}), \text{OLSP}(\boldsymbol{\varepsilon})\} + \text{Cov}\{\text{OLSP}(\boldsymbol{\varepsilon}), \text{OLSE}(\mathbf{X}\boldsymbol{\beta})\} \\ &= \sigma^2 \mathbf{P}_\mathbf{X} \boldsymbol{\Sigma} \mathbf{X}^\perp + \sigma^2 \mathbf{X}^\perp \boldsymbol{\Sigma} \mathbf{P}_\mathbf{X}. \end{aligned}$$

Then by (13) and Lemma 2.2(f),

$$\begin{aligned} r\{D(\mathbf{y}) - D[\text{OLSE}(\mathbf{X}\boldsymbol{\beta})] - D[\text{OLSP}(\boldsymbol{\varepsilon})]\} &= r\{D[\boldsymbol{\varepsilon} - \text{OLSP}(\boldsymbol{\varepsilon})] - D(\boldsymbol{\varepsilon}) + D[\text{OLSP}(\boldsymbol{\varepsilon})]\} \\ &= r(\mathbf{P}_\mathbf{X} \boldsymbol{\Sigma} \mathbf{X}^\perp + \mathbf{X}^\perp \boldsymbol{\Sigma} \mathbf{P}_\mathbf{X}) = r(\mathbf{P}_\mathbf{X} \boldsymbol{\Sigma} \mathbf{X}^\perp) + r(\mathbf{X}^\perp \boldsymbol{\Sigma} \mathbf{P}_\mathbf{X}) \\ &= 2r(\mathbf{X}^\perp \boldsymbol{\Sigma} \mathbf{P}_\mathbf{X}) = 2r[\mathbf{X}, \boldsymbol{\Sigma} \mathbf{P}_\mathbf{X}] - 2r(\mathbf{X}) \\ &= 2r[\mathbf{X}, \boldsymbol{\Sigma}\mathbf{X}] - 2r(\mathbf{X}), \end{aligned}$$

as required for (84). By (77) and (13),

$$r(\text{Cov}\{\text{OLSE}(\mathbf{X}\boldsymbol{\beta}), \text{OLSP}(\boldsymbol{\varepsilon})\}) = r(\mathbf{X}^\perp \boldsymbol{\Sigma} \mathbf{P}_\mathbf{X}) = r[\mathbf{X}, \boldsymbol{\Sigma}\mathbf{X}] - r(\mathbf{X}),$$

as required for (85). \square

6 Twists of BLUPs and OLSPs under GLMs

After establishing BLUPs and OLSPs of all the unknown parameter vectors in (1) and their algebraic and statistical properties in the previous three sections, it may be of interest to consider the relations between the BLUPs and

OLSPs and to establish possible equalities of these predictors. In fact, the equivalence problems of the OLSEs and BLUEs in linear regression theory were initialized and approached in late 1940s from theoretical and applied points of view, which was previously reviewed in [34], and considerable literature exists with many results and discussions on this kind of classic problems; see, e.g., [2, 27, 35–58]. After the sufficient preparations in the previous sections, we are now able to establish possible equalities between BLUPs and OLSPs under various assumptions.

Theorem 6.1. Assume that ϕ in (2) is predictable under (1), and let $\text{BLUP}(\phi)$ and $\text{OLSP}(\phi)$ be as given in (37) and (69), respectively. Then the following results hold.

(a) The following statements are equivalent:

- (i) $\text{OLSP}(\phi) = \text{BLUP}(\phi)$.
- (ii) $(\mathbf{K} - \mathbf{JX})\mathbf{X}^+ \Sigma \mathbf{X}^\perp = \mathbf{0}$.
- (iii) $r \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\Sigma\mathbf{X}^\perp \\ \mathbf{K} - \mathbf{JX} & \mathbf{0} \end{bmatrix} = r(\mathbf{X})$.
- (iv) $r \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\Sigma \\ \mathbf{0} & \mathbf{X}' \\ \mathbf{K} - \mathbf{JX} & \mathbf{0} \end{bmatrix} = 2r(\mathbf{X})$.
- (v) $\mathcal{R}\{[(\mathbf{K} - \mathbf{JX})\mathbf{X}^+ \Sigma]'\} \subseteq \mathcal{R}(\mathbf{X})$.
- (vi) $\mathcal{R}([\mathbf{K} - \mathbf{JX}, \mathbf{0}]') \subseteq \mathcal{R}([\mathbf{X}'\mathbf{X}, \mathbf{X}'\Sigma\mathbf{X}^\perp]')$.
- (vii) $\mathcal{R}([\mathbf{K} - \mathbf{JX}, \mathbf{0}]') \subseteq \mathcal{R}\left(\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\Sigma \\ \mathbf{0} & \mathbf{X}' \end{bmatrix}'\right)$.

(b) [47, 55] The following statements are equivalent:

- (i) $\text{OLSE}(\mathbf{K}\beta) = \text{BLUE}(\mathbf{K}\beta)$.
- (ii) $\mathbf{KX}^+ \Sigma \mathbf{X}^\perp = \mathbf{0}$.
- (iii) $r \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\Sigma\mathbf{X}^\perp \\ \mathbf{K} & \mathbf{0} \end{bmatrix} = r(\mathbf{X})$.
- (iv) $r \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\Sigma \\ \mathbf{0} & \mathbf{X}' \\ \mathbf{K} & \mathbf{0} \end{bmatrix} = 2r(\mathbf{X})$.
- (v) $\mathcal{R}[(\mathbf{KX}^+ \Sigma)'] \subseteq \mathcal{R}(\mathbf{X})$.
- (vi) $\mathcal{R}([\mathbf{K}, \mathbf{0}]') \subseteq \mathcal{R}([\mathbf{X}'\mathbf{X}, \mathbf{X}'\Sigma\mathbf{X}^\perp]')$.
- (vii) $\mathcal{R}([\mathbf{K}, \mathbf{0}]') \subseteq \mathcal{R}\left(\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\Sigma \\ \mathbf{0} & \mathbf{X}' \end{bmatrix}'\right)$.

(c) The following statements are equivalent:

- (i) $\text{OLSP}(\mathbf{J}\epsilon) = \text{BLUP}(\mathbf{J}\epsilon)$.
- (ii) $\mathbf{JXX}^+ \Sigma \mathbf{X}^\perp = \mathbf{0}$.
- (iii) $r \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\Sigma\mathbf{X}^\perp \\ \mathbf{JX} & \mathbf{0} \end{bmatrix} = r(\mathbf{X})$.
- (iv) $r \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\Sigma \\ \mathbf{0} & \mathbf{X}' \\ \mathbf{JX} & \mathbf{0} \end{bmatrix} = 2r(\mathbf{X})$.
- (v) $\mathcal{R}[(\mathbf{JXX}^+ \Sigma)'] \subseteq \mathcal{R}(\mathbf{X})$.
- (vi) $\mathcal{R}([\mathbf{JX}, \mathbf{0}]') \subseteq \mathcal{R}([\mathbf{X}'\mathbf{X}, \mathbf{X}'\Sigma\mathbf{X}^\perp]')$.
- (vii) $\mathcal{R}([\mathbf{JX}, \mathbf{0}]') \subseteq \mathcal{R}\left(\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\Sigma \\ \mathbf{0} & \mathbf{X}' \end{bmatrix}'\right)$.

(d) [34] The following statements are equivalent:

- (i) $\text{OLSE}(\mathbf{X}\beta) = \text{BLUE}(\mathbf{X}\beta)$.
- (ii) $\text{OLSP}(\epsilon) = \text{BLUP}(\epsilon)$.
- (iii) $\mathbf{P}_X \Sigma = \Sigma \mathbf{P}_X$.
- (iv) $\mathbf{X}^\perp \Sigma = \Sigma \mathbf{X}^\perp$.

- (v) $r[\mathbf{X}, \Sigma\mathbf{X}] = r(\mathbf{X})$.
- (vi) $r[\mathbf{X}^\perp, \Sigma\mathbf{X}^\perp] = r(\mathbf{X}^\perp)$.
- (vii) $\mathcal{R}(\Sigma\mathbf{X}) \subseteq \mathcal{R}(\mathbf{X})$.
- (viii) $\mathcal{R}(\Sigma\mathbf{X}^\perp) \subseteq \mathcal{R}(\mathbf{X}^\perp)$.
- (ix) $\mathcal{R}(\Sigma\mathbf{X}) = \mathcal{R}(\Sigma) \cap \mathcal{R}(\mathbf{X})$.
- (x) $\mathcal{R}(\Sigma\mathbf{X}^\perp) = \mathcal{R}(\Sigma) \cap \mathcal{R}(\mathbf{X}^\perp)$.

Proof. From (37) and (69), the difference of OLSP(ϕ) and BLUP(ϕ) is given by

$$\text{OLSP}(\phi) - \text{BLUP}(\phi) = (\mathbf{K}\mathbf{X}^+ + \mathbf{J}\mathbf{X}^\perp - [\mathbf{K}, \mathbf{J}\Sigma\mathbf{X}^\perp][\mathbf{X}, \Sigma\mathbf{X}^\perp]^+ - \mathbf{U}[\mathbf{X}, \Sigma\mathbf{X}^\perp]^\perp)\mathbf{y}.$$

Hence OLSP(ϕ) = BLUP(ϕ) holds if and only if the coefficient matrix of \mathbf{y} is null, i.e.,

$$\mathbf{K}\mathbf{X}^+ + \mathbf{J}\mathbf{X}^\perp - [\mathbf{K}, \mathbf{J}\Sigma\mathbf{X}^\perp][\mathbf{X}, \Sigma\mathbf{X}^\perp]^+ - \mathbf{U}[\mathbf{X}, \Sigma\mathbf{X}^\perp]^\perp = \mathbf{0}.$$

From Lemma 2.7, the matrix equation is solvable for \mathbf{U} if and only if

$$r \begin{bmatrix} \mathbf{K}\mathbf{X}^+ + \mathbf{J}\mathbf{X}^\perp - [\mathbf{K}, \mathbf{J}\Sigma\mathbf{X}^\perp][\mathbf{X}, \Sigma\mathbf{X}^\perp]^+ \\ [\mathbf{X}, \Sigma\mathbf{X}^\perp]^\perp \end{bmatrix} = r([\mathbf{X}, \Sigma\mathbf{X}^\perp]^\perp), \quad (86)$$

where by (13),

$$\begin{aligned} & r \begin{bmatrix} \mathbf{K}\mathbf{X}^+ + \mathbf{J}\mathbf{X}^\perp - [\mathbf{K}, \mathbf{J}\Sigma\mathbf{X}^\perp][\mathbf{X}, \Sigma\mathbf{X}^\perp]^+ \\ [\mathbf{X}, \Sigma\mathbf{X}^\perp]^\perp \end{bmatrix} \\ &= r \begin{bmatrix} \mathbf{K}\mathbf{X}^+ + \mathbf{J}\mathbf{X}^\perp - [\mathbf{K}, \mathbf{J}\Sigma\mathbf{X}^\perp][\mathbf{X}, \Sigma\mathbf{X}^\perp]^+ & \mathbf{0} \\ \mathbf{I}_n & [\mathbf{X}, \Sigma\mathbf{X}^\perp] \end{bmatrix} - r[\mathbf{X}, \Sigma\mathbf{X}^\perp] \\ &= r \begin{bmatrix} \mathbf{K}\mathbf{X}^+ + \mathbf{J}\mathbf{X}^\perp & \mathbf{K} \mathbf{J}\Sigma\mathbf{X}^\perp \\ \mathbf{I}_n & \mathbf{X} \Sigma\mathbf{X}^\perp \end{bmatrix} - r[\mathbf{X}, \Sigma] \\ &= r \begin{bmatrix} \mathbf{0} & \mathbf{0} \mathbf{J}\Sigma\mathbf{X}^\perp - (\mathbf{K}\mathbf{X}^+ + \mathbf{J}\mathbf{X}^\perp)\Sigma\mathbf{X}^\perp \\ \mathbf{I}_n & \mathbf{0} \end{bmatrix} - r[\mathbf{X}, \Sigma] \\ &= n + r[(\mathbf{K} - \mathbf{J}\mathbf{X})\mathbf{X}^+\Sigma\mathbf{X}^\perp] - r[\mathbf{X}, \Sigma] \\ &= n + r[(\mathbf{K} - \mathbf{J}\mathbf{X})(\mathbf{X}'\mathbf{X})^+\mathbf{X}'\Sigma\mathbf{X}^\perp] - r[\mathbf{X}, \Sigma] \\ &= r \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\Sigma\mathbf{X}^\perp \\ \mathbf{K} - \mathbf{J}\mathbf{X} & \mathbf{0} \end{bmatrix} + n - r(\mathbf{X}) - r[\mathbf{X}, \Sigma] \quad (\text{by (16)}) \\ &= r \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\Sigma \\ \mathbf{0} & \mathbf{X}' \end{bmatrix} + n - 2r(\mathbf{X}) - r[\mathbf{X}, \Sigma] \quad (\text{by (14)}), \end{aligned} \quad (87)$$

$$r([\mathbf{X}, \Sigma\mathbf{X}^\perp]^\perp) = n - r[\mathbf{X}, \Sigma]. \quad (88)$$

Substituting (87) and (88) into (86) yields the following equalities

$$(\mathbf{K} - \mathbf{J}\mathbf{X})\mathbf{X}^+\Sigma\mathbf{X}^\perp = \mathbf{0}, \quad r \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\Sigma\mathbf{X}^\perp \\ \mathbf{K} - \mathbf{J}\mathbf{X} & \mathbf{0} \end{bmatrix} = r(\mathbf{X}), \quad r \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\Sigma \\ \mathbf{0} & \mathbf{X}' \\ \mathbf{K} - \mathbf{J}\mathbf{X} & \mathbf{0} \end{bmatrix} = 2r(\mathbf{X}),$$

establishing the equivalences of (i), (ii), (iii), and (iv) in (a). The equivalences of (ii) and (v), (iii) and (vi), (iv) and (vii) in (a) follow from Lemma 2.2(a) and (b). Results (b)–(d) are special cases of (a) for different choices of \mathbf{K} and \mathbf{J} . \square

Note that $(\mathbf{K} - \mathbf{J}\mathbf{X})\mathbf{X}^+\Sigma\mathbf{X}^\perp = \mathbf{0}$ in (ii) of Theorem 6.1(a) is a linear matrix equation for \mathbf{K} and \mathbf{J} to satisfy. Solving this equation will produce all ϕ such that OLSP(ϕ) = BLUP(ϕ) hold. Concerning the relationships between $D[\phi - \text{BLUP}(\phi)]$ and $D[\phi - \text{OLSP}(\phi)]$, we have the following results.

Theorem 6.2. Assume that ϕ in (2) is predictable under (1), and let $\text{BLUP}(\phi)$ and $\text{OLSP}(\phi)$ be as given in (37) and (69), respectively. Then

$$\begin{aligned} i_{-}\{D[\phi - \text{BLUP}(\phi)] - D[\phi - \text{OLSP}(\phi)]\} &= r\{D[\phi - \text{BLUP}(\phi)] - D[\phi - \text{OLSP}(\phi)]\} \\ &= r \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\boldsymbol{\Sigma} \\ \mathbf{0} & \mathbf{X}' \\ \mathbf{K} - \mathbf{J}\mathbf{X} & \mathbf{0} \end{bmatrix} - 2r(\mathbf{X}). \end{aligned} \quad (89)$$

In particular,

$$\begin{aligned} i_{-}\{D[\text{BLUE}(\mathbf{K}\boldsymbol{\beta})] - D[\text{OLSE}(\mathbf{K}\boldsymbol{\beta})]\} &= r\{D[\text{BLUE}(\mathbf{K}\boldsymbol{\beta})] - D[\text{OLSE}(\mathbf{K}\boldsymbol{\beta})]\} \\ &= r \begin{bmatrix} \boldsymbol{\Sigma}\mathbf{X} & \mathbf{X} & \mathbf{0} \\ \mathbf{X}'\mathbf{X} & \mathbf{0} & \mathbf{K}' \end{bmatrix} - 2r(\mathbf{X}) \text{ if } \mathbf{K}\boldsymbol{\beta} \text{ is estimable,} \end{aligned} \quad (90)$$

$$\begin{aligned} i_{-}\{D[\text{BLUE}(\mathbf{X}\boldsymbol{\beta})] - D[\text{OLSE}(\mathbf{X}\boldsymbol{\beta})]\} &= r\{D[\text{BLUE}(\mathbf{X}\boldsymbol{\beta})] - D[\text{OLSE}(\mathbf{X}\boldsymbol{\beta})]\} \\ &= r[\mathbf{X}, \boldsymbol{\Sigma}\mathbf{X}] - r(\mathbf{X}) < n, \end{aligned} \quad (91)$$

$$\begin{aligned} i_{-}\{D[\text{BLUE}(\boldsymbol{\beta})] - D[\text{OLSE}(\boldsymbol{\beta})]\} &= r\{D[\text{OLSE}(\boldsymbol{\beta})] - D[\text{BLUE}(\boldsymbol{\beta})]\} \\ &= r[\boldsymbol{\Sigma}\mathbf{X}, \mathbf{X}] - p < n \text{ if } \boldsymbol{\beta} \text{ is estimable,} \end{aligned} \quad (92)$$

$$\begin{aligned} i_{-}\{D[\boldsymbol{\varepsilon} - \text{BLUP}(\boldsymbol{\varepsilon})] - D[\boldsymbol{\varepsilon} - \text{OLSP}(\boldsymbol{\varepsilon})]\} &= r\{D[\boldsymbol{\varepsilon} - \text{BLUP}(\boldsymbol{\varepsilon})] - D[\boldsymbol{\varepsilon} - \text{OLSP}(\boldsymbol{\varepsilon})]\} \\ &= r[\mathbf{X}, \boldsymbol{\Sigma}\mathbf{X}] - r(\mathbf{X}) < n, \end{aligned} \quad (93)$$

$$\begin{aligned} i_{+}\{D[\text{BLUP}(\boldsymbol{\varepsilon})] - D[\text{OLSP}(\boldsymbol{\varepsilon})]\} &= i_{-}\{D[\text{BLUP}(\boldsymbol{\varepsilon})] - D[\text{OLSP}(\boldsymbol{\varepsilon})]\} \\ &= r[\mathbf{X}, \boldsymbol{\Sigma}\mathbf{X}] - r(\mathbf{X}) < n, \end{aligned} \quad (94)$$

$$r\{D[\text{BLUP}(\boldsymbol{\varepsilon})] - D[\text{OLSP}(\boldsymbol{\varepsilon})]\} = 2r[\mathbf{X}, \boldsymbol{\Sigma}\mathbf{X}] - 2r(\mathbf{X}) < 2n. \quad (95)$$

In consequence, the following results hold.

$$(a) \ D[\phi - \text{BLUP}(\phi)] < D[\phi - \text{OLSP}(\phi)] \Leftrightarrow r \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\boldsymbol{\Sigma} \\ \mathbf{0} & \mathbf{X}' \\ \mathbf{K} - \mathbf{J}\mathbf{X} & \mathbf{0} \end{bmatrix} = 2r(\mathbf{X}) + k.$$

$$(b) \ D[\text{BLUE}(\mathbf{K}\boldsymbol{\beta})] < D[\text{OLSE}(\mathbf{K}\boldsymbol{\beta})] \Leftrightarrow r \begin{bmatrix} \boldsymbol{\Sigma}\mathbf{X} & \mathbf{X} & \mathbf{0} \\ \mathbf{X}'\mathbf{X} & \mathbf{0} & \mathbf{K}' \end{bmatrix} = 2r(\mathbf{X}) + k.$$

Proof. By (37), (43), (69) and (73)

$$\begin{aligned} &D[\phi - \text{BLUP}(\phi)] - D[\phi - \text{OLSP}(\phi)] \\ &= \sigma^2 \left([\mathbf{K}, \mathbf{J}\boldsymbol{\Sigma}\mathbf{X}^{\perp}] [\mathbf{X}, \boldsymbol{\Sigma}\mathbf{X}^{\perp}]^{+} - \mathbf{J} \right) \boldsymbol{\Sigma} \left([\mathbf{K}, \mathbf{J}\boldsymbol{\Sigma}\mathbf{X}^{\perp}] [\mathbf{X}, \boldsymbol{\Sigma}\mathbf{X}^{\perp}]^{+} - \mathbf{J} \right)' \\ &\quad - \sigma^2 \left(\mathbf{K}\mathbf{X}^{+} - \mathbf{J}\mathbf{P}_{\mathbf{X}} \right) \boldsymbol{\Sigma} \left(\mathbf{K}\mathbf{X}^{+} - \mathbf{J}\mathbf{P}_{\mathbf{X}} \right)', \end{aligned} \quad (96)$$

$$\begin{aligned} &D[\text{OLSP}(\phi) - \text{BLUP}(\phi)] \\ &= \sigma^2 \left(\mathbf{K}\mathbf{X}^{+} + \mathbf{J}\mathbf{X}^{\perp} - [\mathbf{K}, \mathbf{J}\boldsymbol{\Sigma}\mathbf{X}^{\perp}] [\mathbf{X}, \boldsymbol{\Sigma}\mathbf{X}^{\perp}]^{+} \right) \boldsymbol{\Sigma} \left(\mathbf{K}\mathbf{X}^{+} + \mathbf{J}\mathbf{X}^{\perp} - [\mathbf{K}, \mathbf{J}\boldsymbol{\Sigma}\mathbf{X}^{\perp}] [\mathbf{X}, \boldsymbol{\Sigma}\mathbf{X}^{\perp}]^{+} \right)'. \end{aligned} \quad (97)$$

Applying (81) to (96), we first obtain

$$i_{\pm} (D[\phi - \text{BLUP}(\phi)] - D[\phi - \text{OLSP}(\phi)]) = i_{\mp} \begin{bmatrix} \sigma^2 \mathbf{X}'\boldsymbol{\Sigma}\mathbf{X} & \mathbf{0} & \mathbf{X}'\mathbf{X} \\ \mathbf{0} & -D[\phi - \text{BLUP}(\phi)] & \mathbf{K} - \mathbf{J}\mathbf{X} \\ \mathbf{X}'\mathbf{X} & \mathbf{K}' - \mathbf{X}'\mathbf{J}' & \mathbf{0} \end{bmatrix} - r(\mathbf{X}). \quad (98)$$

Furthermore, substituting $D[\phi - \text{BLUP}(\phi)]$ in (43) into the block matrix in (98) and applying (61), we obtain

$$i_{\mp} \begin{bmatrix} \sigma^2 \mathbf{X}'\boldsymbol{\Sigma}\mathbf{X} & \mathbf{0} & \mathbf{X}'\mathbf{X} \\ \mathbf{0} & -D[\phi - \text{BLUP}(\phi)] & \mathbf{K} - \mathbf{J}\mathbf{X} \\ \mathbf{X}'\mathbf{X} & \mathbf{K}' - \mathbf{X}'\mathbf{J}' & \mathbf{0} \end{bmatrix}$$

$$\begin{aligned}
&= i_{\mp} \begin{bmatrix} \mathbf{X}'\Sigma\mathbf{X} & 0 & \mathbf{X}'\mathbf{X} \\ 0 & -([\mathbf{K}, \mathbf{J}\Sigma\mathbf{X}^{\perp}][\mathbf{X}, \Sigma\mathbf{X}^{\perp}]^{+} - \mathbf{J})\Sigma([\mathbf{K}, \mathbf{J}\Sigma\mathbf{X}^{\perp}][\mathbf{X}, \Sigma\mathbf{X}^{\perp}]^{+} - \mathbf{J})' & \mathbf{K} - \mathbf{JX} \\ \mathbf{X}'\mathbf{X} & \mathbf{K}' - \mathbf{X}'\mathbf{J}' & 0 \end{bmatrix} \\
&= i_{\mp} \left(\begin{bmatrix} \mathbf{X}'\Sigma\mathbf{X} & 0 & \mathbf{X}'\mathbf{X} \\ 0 & 0 & \mathbf{K} - \mathbf{JX} \\ \mathbf{X}'\mathbf{X} & \mathbf{K}' - \mathbf{X}'\mathbf{J}' & 0 \end{bmatrix} \right. \\
&\quad \left. - \begin{bmatrix} 0 \\ \mathbf{I}_k \\ 0 \end{bmatrix} \left(([\mathbf{K}, \mathbf{J}\Sigma\mathbf{X}^{\perp}][\mathbf{X}, \Sigma\mathbf{X}^{\perp}]^{+} - \mathbf{J})\Sigma([\mathbf{K}, \mathbf{J}\Sigma\mathbf{X}^{\perp}][\mathbf{X}, \Sigma\mathbf{X}^{\perp}]^{+} - \mathbf{J})' \begin{bmatrix} 0 & \mathbf{I}_k & 0 \end{bmatrix} \right) \right) \\
&= i_{\pm} \left[\begin{bmatrix} \Sigma & 0 & 0 & 0 & \mathbf{X} \\ 0 & -\mathbf{X}'\Sigma\mathbf{X} & 0 & -\mathbf{X}'\mathbf{X} & 0 \\ 0 & 0 & 0 & -\mathbf{K} + \mathbf{JX} & \mathbf{K} - \mathbf{JX} \\ 0 & -\mathbf{X}'\mathbf{X} & -\mathbf{K}' + \mathbf{X}'\mathbf{J}' & 0 & 0 \\ \mathbf{X}' & 0 & \mathbf{K}' - \mathbf{X}'\mathbf{J}' & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ 0 \\ \mathbf{K} - \mathbf{JX} \\ 0 \\ 0 \end{bmatrix} \right] + i_{\mp}(\mathbf{X}^{\perp}\Sigma\mathbf{X}^{\perp}) - r[\mathbf{X}, \Sigma] \\
&= i_{\pm} \left[\begin{bmatrix} \Sigma & 0 & 0 & 0 & \mathbf{X} \\ 0 & -\mathbf{X}'\Sigma\mathbf{X} & 0 & -\mathbf{X}'\mathbf{X} & 0 \\ 0 & 0 & 0 & -\mathbf{K} + \mathbf{JX} & \mathbf{K} - \mathbf{JX} \\ 0 & -\mathbf{X}'\mathbf{X} & -\mathbf{K}' + \mathbf{X}'\mathbf{J}' & 0 & 0 \\ \mathbf{X}' & 0 & \mathbf{K}' - \mathbf{X}'\mathbf{J}' & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ 0 \\ \mathbf{K} - \mathbf{JX} \\ 0 \\ 0 \end{bmatrix} \right] + i_{\mp}(\mathbf{X}^{\perp}\Sigma\mathbf{X}^{\perp}) - r[\mathbf{X}, \Sigma] \\
&= i_{\pm} \left[\begin{bmatrix} \Sigma & 0 & 0 & 0 & \mathbf{X} \\ 0 & -\mathbf{X}'\Sigma\mathbf{X} & 0 & -\mathbf{X}'\mathbf{X} & -\mathbf{X}'\mathbf{X} \\ 0 & 0 & 0 & -\mathbf{K} + \mathbf{JX} & 0 \\ 0 & -\mathbf{X}'\mathbf{X} & -\mathbf{K}' + \mathbf{X}'\mathbf{J}' & 0 & 0 \\ \mathbf{X}' & -\mathbf{X}'\mathbf{X} & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ 0 \\ \mathbf{K} - \mathbf{JX} \\ 0 \\ 0 \end{bmatrix} \right] + i_{\mp}(\mathbf{X}^{\perp}\Sigma\mathbf{X}^{\perp}) - r[\mathbf{X}, \Sigma] \\
&= i_{\pm} \left[\begin{bmatrix} \Sigma & \Sigma\mathbf{X} & 0 & 0 & \mathbf{X} \\ \mathbf{X}'\Sigma & 0 & 0 & -\mathbf{X}'\mathbf{X} & 0 \\ 0 & 0 & 0 & -\mathbf{K} + \mathbf{JX} & 0 \\ 0 & -\mathbf{X}'\mathbf{X} & -\mathbf{K}' + \mathbf{X}'\mathbf{J}' & 0 & 0 \\ \mathbf{X}' & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ 0 \\ \mathbf{K} - \mathbf{JX} \\ 0 \\ 0 \end{bmatrix} \right] + i_{\mp}(\mathbf{X}^{\perp}\Sigma\mathbf{X}^{\perp}) - r[\mathbf{X}, \Sigma] \\
&= i_{\pm} \left[\begin{bmatrix} \Sigma & 0 & \Sigma\mathbf{X} & \mathbf{X} & 0 \\ 0 & 0 & \mathbf{X}'\mathbf{X} & 0 & \mathbf{K}' - \mathbf{X}'\mathbf{J}' \\ \mathbf{X}'\Sigma & \mathbf{X}'\mathbf{X} & 0 & 0 & 0 \\ \mathbf{X}' & 0 & 0 & 0 & 0 \\ 0 & \mathbf{K} - \mathbf{JX} & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ 0 \\ \mathbf{K} - \mathbf{JX} \\ 0 \\ 0 \end{bmatrix} \right] + i_{\mp}(\mathbf{X}^{\perp}\Sigma\mathbf{X}^{\perp}) - r[\mathbf{X}, \Sigma]. \tag{99}
\end{aligned}$$

Substituting (99) into (98) gives

$$\begin{aligned}
&i_{\pm}(\mathbf{D}[\boldsymbol{\phi} - \text{BLUP}(\boldsymbol{\phi})] - \mathbf{D}[\text{OLSP}(\boldsymbol{\phi}) - \boldsymbol{\phi}]) \\
&= i_{\pm} \left[\begin{bmatrix} \Sigma & 0 & \Sigma\mathbf{X} & \mathbf{X} & 0 \\ 0 & 0 & \mathbf{X}'\mathbf{X} & 0 & \mathbf{K}' - \mathbf{X}'\mathbf{J}' \\ \mathbf{X}'\Sigma & \mathbf{X}'\mathbf{X} & 0 & 0 & 0 \\ \mathbf{X}' & 0 & 0 & 0 & 0 \\ 0 & \mathbf{K} - \mathbf{JX} & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{X} \\ 0 \\ \mathbf{K} - \mathbf{JX} \\ 0 \\ 0 \end{bmatrix} \right] + i_{\mp}(\mathbf{X}^{\perp}\Sigma\mathbf{X}^{\perp}) - r(\mathbf{X}) - r[\mathbf{X}, \Sigma].
\end{aligned}$$

Hence,

$$i_{+}(\mathbf{D}[\boldsymbol{\phi} - \text{BLUP}(\boldsymbol{\phi})] - \mathbf{D}[\boldsymbol{\phi} - \text{OLSP}(\boldsymbol{\phi})])$$

$$\begin{aligned}
&= i_+ \begin{bmatrix} \Sigma & 0 & \Sigma X & X & 0 \\ 0 & 0 & X'X & 0 & K' - X'J' \\ X'\Sigma & X'X & 0 & 0 & 0 \\ X' & 0 & 0 & 0 & 0 \\ 0 & K - JX & 0 & 0 & 0 \end{bmatrix} + i_-(X^\perp \Sigma X^\perp) - r(X) - r[X, \Sigma] \\
&= r \begin{bmatrix} \Sigma & 0 & \Sigma X & X & 0 \\ 0 & 0 & X'X & 0 & K' - X'J' \end{bmatrix} - r(X) - r[X, \Sigma] \\
&= r \begin{bmatrix} \Sigma & 0 & 0 & X & 0 \\ 0 & 0 & X'X & 0 & 0 \end{bmatrix} - r(X) - r[X, \Sigma] = 0, \\
&i_-(D[\phi - \text{BLUP}(\phi)] - D[\phi - \text{OLSP}(\phi)]) \\
&= i_- \begin{bmatrix} \Sigma & 0 & \Sigma X & X & 0 \\ 0 & 0 & X'X & 0 & K' - X'J' \\ X'\Sigma & X'X & 0 & 0 & 0 \\ X' & 0 & 0 & 0 & 0 \\ 0 & K - JX & 0 & 0 & 0 \end{bmatrix} + r(X^\perp \Sigma) - r(X) - r[X, \Sigma] \\
&= r \begin{bmatrix} X'\Sigma & X'X \\ X' & 0 \\ 0 & K - JX \end{bmatrix} - 2r(X),
\end{aligned}$$

establishing (89). Eqs. (90)–(93) follow directly from (89). Eqs. (94) and (95) were proved in [5]. Applying Lemma 2.1 to (89) and (90) yields (a) and (b). \square

The following results follow from Theorems 5.3, 6.1, and 6.2.

Corollary 6.3. Assume that ϕ in (2) is predictable under (1), and let $\text{BLUP}(\phi)$ and $\text{OLSP}(\phi)$ be as given in (37) and (69), respectively. Then the following statements are equivalent:

- (a) $\text{BLUP}(\phi) = \text{OLSP}(\phi)$, i.e., $\phi - \text{BLUP}(\phi) = \phi - \text{OLSP}(\phi)$.
- (b) $D[\phi - \text{BLUP}(\phi)] = D[\phi - \text{OLSP}(\phi)]$.
- (c) $r \begin{bmatrix} X'X & X'\Sigma \\ 0 & X' \\ K - JX & 0 \end{bmatrix} = 2r(X)$, i.e., $\mathcal{R} \begin{bmatrix} (K - JX)' \\ 0 \end{bmatrix} \subseteq \mathcal{R} \begin{bmatrix} X'X & 0 \\ \Sigma X & X \end{bmatrix}$.

Corollary 6.4. Assume that $K\beta$ is estimable under (1). Then the following statistical facts are equivalent:

- (a) $\text{BLUE}(K\beta) = \text{OLSE}(K\beta)$.
- (b) $D[\text{BLUE}(K\beta)] = D[\text{OLSE}(K\beta)]$.
- (c) $\text{Cov}\{\text{OLSE}(K\beta), y\} = \text{Cov}\{\text{BLUE}(K\beta), y\}$.
- (d) $\text{Cov}\{\text{OLSE}(K\beta), y\} = \text{Cov}\{\text{BLUE}(K\beta), \text{OLSE}(X\beta)\}$.
- (e) $\text{Cov}\{\text{OLSE}(K\beta), \text{OLSP}(\varepsilon)\} = 0$.

Corollary 6.5. Under the assumption in (1), the following statistical facts are equivalent:

- (a) $\text{BLUE}(X\beta) = \text{OLSE}(X\beta)$.
- (b) $\text{BLUP}(\varepsilon) = \text{OLSP}(\varepsilon)$.
- (c) $D[\text{BLUE}(X\beta)] = D[\text{OLSE}(X\beta)]$.
- (d) $D[\text{BLUP}(\varepsilon)] = D[\text{OLSP}(\varepsilon)]$.
- (e) $D[\varepsilon - \text{BLUP}(\varepsilon)] = D[\varepsilon - \text{OLSP}(\varepsilon)]$.
- (f) $D(y) = D[\text{OLSE}(X\beta)] + D[\text{OLSP}(\varepsilon)]$.
- (g) $D[\varepsilon - \text{OLSP}(\varepsilon)] = D(\varepsilon) - D[\text{OLSP}(\varepsilon)]$.
- (h) $\text{Cov}\{\text{OLSE}(X\beta), y\} = \text{Cov}\{\text{BLUE}(X\beta), y\}$.
- (i) $\text{Cov}\{\text{OLSE}(X\beta), y\} = \text{Cov}\{\text{BLUE}(X\beta), \text{OLSE}(X\beta)\}$.
- (j) $\text{Cov}\{\text{OLSE}(X\beta), \text{OLSE}(X\beta)\} = \text{Cov}\{\text{BLUE}(X\beta), y\}$.

- (k) $\text{Cov}\{\text{OLSE}(\mathbf{X}\boldsymbol{\beta}), \text{OLSE}(\mathbf{X}\boldsymbol{\beta})\} = \text{Cov}\{\text{BLUE}(\mathbf{X}\boldsymbol{\beta}), \text{OLSE}(\mathbf{X}\boldsymbol{\beta})\}.$
- (l) $\text{Cov}\{\text{OLSE}(\mathbf{X}\boldsymbol{\beta}), \mathbf{y}\} = \text{Cov}\{\mathbf{y}, \text{OLSE}(\mathbf{X}\boldsymbol{\beta})\}.$
- (m) $\text{Cov}\{\text{OLSP}(\boldsymbol{\varepsilon}), \mathbf{y}\} = \text{Cov}\{\mathbf{y}, \text{OLSP}(\boldsymbol{\varepsilon})\}.$
- (n) $\text{Cov}\{\text{OLSE}(\mathbf{X}\boldsymbol{\beta}), \text{OLSP}(\boldsymbol{\varepsilon})\} = \text{Cov}\{\text{OLSP}(\boldsymbol{\varepsilon}), \text{OLSE}(\mathbf{X}\boldsymbol{\beta})\}.$
- (o) $\text{Cov}\{\text{OLSE}(\mathbf{X}\boldsymbol{\beta}), \text{OLSP}(\boldsymbol{\varepsilon})\} + \text{Cov}\{\text{OLSP}(\boldsymbol{\varepsilon}), \text{OLSE}(\mathbf{X}\boldsymbol{\beta})\} = \mathbf{0}.$
- (p) $\text{Cov}\{\text{OLSE}(\mathbf{X}\boldsymbol{\beta}), \text{OLSP}(\boldsymbol{\varepsilon})\} = \mathbf{0}.$

Some of the equivalent facts in Corollaries 6.4 and 6.5 were proved in [5, 58]. Furthermore, many interesting vector and matrix norm equalities for OLSEs to be BLUEs can be established. We shall present them in another paper.

7 Conclusions

We have offered a predominantly theoretical coverage of the statistical predictions and estimations by establishing two groups of standard result on exact algebraic expressions of the BLUPs and OLSPs of all unknown parameters and their fundamental properties under (1). We have also established a variety of exact algebraic formulas for calculating ranks and inertias of the matrices associated with the BLUPs and OLSPs, and have used the formulas to characterize many interesting and valuable equalities and inequalities for the dispersion matrices of the BLUPs and OLSPs under (1). The whole work contains a massive amount of useful results related to the world of GLMs and can serve as a comprehensive description of rank/inertia theory in the parameter prediction and estimation problems under GLMs.

Statistical theory and methods often require various 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 derivations and characterizations of estimators and predictors and their features and performances under linear regression models. So that matrix theory provides a powerful tool set for addressing statistical problems. There is a long list of handbooks on matrix algebra for statistical analysis published since 1960s; see, e.g., [2, 59–74], while various new algebraic methods were developed regularly in matrix mathematics. But it is rarely the case that the algebraic techniques in matrix theory are ready-made to address statistical challenges. This is why the dialogue between matrix theory and statistics benefits both disciplines.

Although the ranks/inertias of matrices are the conceptual foundation in elementary linear algebra and are the most significant finite integers in reflecting intrinsic properties of matrices, it took a long time in the development of mathematics to establish various analytical and valuable formulas for calculating ranks/inertias of matrices and to use the formulas, as demonstrated in the previous sections, in the intuitive and rigorous derivations of matrix equalities and inequalities in the statistical analysis of GLMs. The present author has been devoting to this subject and has proved thousands of matrix rank/inertia formulas since 1980s by using various skillful calculations for partitioned matrices. This work has provided significant advances to general algebraical and statistical methodology, and a state-of-the-art theory on matrix ranks/inertias with applications has been established. We are now able to use matrix rank/inertia formulas to describe many fundamental properties and features of matrices, such as, establishing and simplifying various complicated matrix expressions, deriving matrix equalities and inequalities that involve generalized inverses of matrices, characterizing definiteness and semi-definiteness of symmetric matrices, and solving matrix optimization problems in the Löwner sense. In the past decade, the present author has been working on applications of the matrix rank/inertia methodology in linear regression analysis, while much experience of using various rank/inertia formulas in statistical inference of GLMs has been achieved, and many fundamental and valuable mathematical and statistical features of predictors/estimators under GLMs have been obtained during this approach. Some recent contributions on this topic by the present author and his collaborators are presented in [5–7, 9, 17, 22, 57, 58, 75–79], which contain a massive amount of useful results related to the world of GLMs. The whole findings in these papers provide significant advances to algebraical methodology in statistical analysis and inference of GLMs, and can merge into the essential part of unified theory of GLMs.

Acknowledgement: The author thanks anonymous referees for their helpful comments and suggestions on an earlier version of this paper.

This work was supported by the National Natural Science Foundation of China (Grant No. 11271384).

References

- [1] Markiewicz A., Puntanen S., All about the \perp with its applications in the linear statistical models, *Open Math.*, 2015, 13, 33–50
- [2] Puntanen S., Styan G.P.H., Isotalo J., *Matrix Tricks for Linear Statistical Models: Our Personal Top Twenty*, Springer, Berlin Heidelberg, 2011
- [3] Rao C.R., Mitra S.K., *Generalized Inverse of Matrices and Its Applications*, Wiley, New York, 1971
- [4] Tian Y., Equalities and inequalities for inertias of Hermitian matrices with applications, *Linear Algebra Appl.*, 2010, 433, 263–296
- [5] Tian Y., Some equalities and inequalities for covariance matrices of estimators under linear model, *Stat. Papers*, 2015, DOI:10.1007/s00362-015-0707-x
- [6] Tian Y., Guo W., On comparison of dispersion matrices of estimators under a constrained linear model, *Stat. Methods Appl.*, 2016, 25, 623–649
- [7] Tian Y., Jiang B., Matrix rank/inertia formulas for least-squares solutions with statistical applications, *Spec. Matrices*, 2016, 4, 130–140
- [8] Tian Y., Jiang B., Quadratic properties of least-squares solutions of linear matrix equations with statistical applications, *Comput. Statist.*, DOI:10.1007/s00180-016-0693-z
- [9] Dong B., Guo W., Tian Y., On relations between BLUEs under two transformed linear models, *J. Multivariate Anal.*, 2014, 131, 279–292
- [10] Lowerre J.M., Some simplifying results on BLUEs, *J. Amer. Stat. Assoc.*, 1977, 72, 433–437
- [11] Rao C.R., A lemma on optimization of matrix function and a review of the unified theory of linear estimation, In: Y. Dodge (ed.), *Statistical Data Analysis and Inference*, North-Holland, Elsevier, 1989, 397–417
- [12] Goldberger A.S., Best linear unbiased prediction in the generalized linear regression models, *J. Amer. Stat. Assoc.*, 1962, 57, 369–375
- [13] Marsaglia G., Styan G.P.H., Equalities and inequalities for ranks of matrices, *Linear Multilinear Algebra*, 1974, 2, 269–292
- [14] Tian Y., More on maximal and minimal ranks of Schur complements with applications, *Appl. Math. Comput.*, 2004, 152, 675–692
- [15] Penrose R., A generalized inverse for matrices, *Proc. Cambridge Phil. Soc.*, 1955, 51, 406–413
- [16] Lange K., Chi E.C., Zhou H., A brief survey of modern optimization for statisticians, *Internat. Stat. Rev.*, 2014, 82, 46–70
- [17] Tian Y., A new derivation of BLUPs under random-effects model, *Metrika*, 2015, 78, 905–918
- [18] Rao C.R., Unified theory of linear estimation, *Sankhyā Ser. A*, 1971, 33, 371–394
- [19] Rao C.R., Representations of best linear unbiased estimators in the Gauss–Markoff model with a singular dispersion matrix, *J. Multivariate Anal.*, 1973, 3, 276–292
- [20] Rao C.R., Toutenburg H., Shalabh, Heumann C., *Linear Models and Generalizations Least Squares and Alternatives*, 3rd ed., Springer, Berlin Heidelberg, 2008
- [21] Searle S.R., The matrix handling of BLUE and BLUP in the mixed linear model, *Linear Algebra Appl.*, 1997, 264, 291–311
- [22] Gan S., Sun Y., Tian Y., Equivalence of predictors under real and over-parameterized linear models, *Comm. Stat. Theory Meth.*, 2016, DOI:10.1080/03610926.2015.1100742
- [23] Tian Y., Jiang B., A new analysis of the relationships between a general linear model and its mis-specified forms, *J. Korean Stat. Soc.*, 2016, DOI:10.1016/j.jkss.2016.08.004
- [24] Baksalary J.K., Puntanen S., Characterizations of the best linear unbiased estimator in the general Gauss–Markov model with the use of matrix partial orderings, *Linear Algebra Appl.*, 1990, 127, 363–370
- [25] Baksalary J.K., Puntanen S., Styan G.P.H., A property of the dispersion matrix of the best linear unbiased estimator in the general Gauss–Markov model, *Sankhyā Ser. A*, 1990, 52, 279–296
- [26] Isotalo J., Puntanen S., Styan G.P.H., The BLUE's covariance matrix revisited: A review, *J. Stat. Plann. Inference*, 2008, 138, 2722–2737
- [27] Puntanen S., Styan G.P.H., Tian Y., Three rank formulas associated with the covariance matrices of the BLUE and the OLSE in the general linear model, *Econometric Theory*, 2005, 21, 659–664
- [28] Abdulle A., Wanner G., 200 years of least squares method, *Elem. Math.*, 2002, 57, 45–60
- [29] Farebrother R.W., Some early statistical contributions to the theory and practice of linear algebra, *Linear Algebra Appl.*, 1996, 237/238, 205–224
- [30] Paris Q., The dual of the least-squares method, *Open J. Statist.*, 2015, 5, 658–664
- [31] Stigler S.M., Gauss and the invention of least squares, *Ann. Stat.*, 1981, 9, 465–474
- [32] Graybill F.A., *An Introduction to Linear Statistical Models*, Vol. I, McGraw–Hill, New York, 1961
- [33] Searle S.R., *Linear Models*, Wiley, New York, 1971

- [34] Puntanen S., Styan G.P.H., The equality of the ordinary least squares estimator and the best linear unbiased estimator, with comments by O. Kempthorne, S.R. Searle, and a reply by the authors, *Amer. Statistician*, 1989, 43, 153–164
- [35] Alalouf I.S., Styan G.P.H., Characterizations of the conditions for the ordinary least squares estimator to be best linear unbiased, in: Y.P. Chaubey, T.D. Dwivedi (Eds.), *Topics in Applied Statistics*, Concordia University, Montréal, 1984, 331–344
- [36] Baksalary J.K., Criteria for the equality between ordinary least squares and best linear unbiased estimators under certain linear models, *Canad. J. Stat.*, 1988, 16, 97–102
- [37] Baksalary J.K., Kala R., An extension of a rank criterion for the least squares estimator to be the best linear unbiased estimator, *J. Stat. Plann. Inference*, 1977, 1, 309–312
- [38] Baksalary J.K., Kala R., Simple least squares estimation versus best linear unbiased prediction, *J. Stat. Plann. Inference*, 1981, 5, 147–151
- [39] Baksalary J.K., van Eijnsbergen A.C., Comparison of two criteria for ordinary-least-squares estimators to be best linear unbiased estimators, *Amer. Statistician*, 1988, 42, 205–208
- [40] Baksalary O.M., Trenkler G., Between OLSE and BLUE, *Aust. N.Z.J. Stat.*, 2011, 53, 289–303
- [41] Baksalary O.M., Trenkler G., Liski E.P., Let us do the twist again, *Stat. Papers*, 2013, 54, 1109–1119
- [42] Haslett S.J., Isotalo J., Liu Y., Puntanen S., Equalities between OLSE, BLUE and BLUP in the linear model, *Stat. Papers*, 2014, 55, 543–561
- [43] Haslett S.J., Puntanen S., A note on the equality of the BLUPs for new observations under two linear models, *Acta Comment. Univ. Tartu. Math.*, 2010, 14 27–33
- [44] Haslett S.J., Puntanen S., Equality of BLUEs or BLUPs under two linear models using stochastic restrictions, *Stat. Papers*, 2010, 51, 465–475
- [45] Haslett S.J., Puntanen S., On the equality of the BLUPs under two linear mixed models, *Metrika*, 2011, 74, 381–395
- [46] Herzberg A.M., Aleong J., Further conditions on the equivalence of ordinary least squares and weighted least squares estimators with examples, in: J. Lanke, G. Lindgren (Eds.), *Contributions to Probability and Statistics in Honour of Gunnar Blom*, University of Lund, 1985, 127–142
- [47] Isotalo J., Puntanen S., A note on the equality of the OLSE and the BLUE of the parametric functions in the general Gauss–Markov model, *Stat. Papers*, 2009, 50, 185–193
- [48] Jiang B., Sun Y., On the equality of estimators under a general partitioned linear model with parameter restrictions, *Stat. Papers*, 2016, DOI:10.1007/s00362-016-0837-9
- [49] Kruskal W., When are Gauss–Markov and least squares estimators identical? A coordinate-free approach, *Ann. Math. Statist.*, 1968, 39, 70–75
- [50] Liski E.P., Puntanen S., Wang S., Bounds for the trace of the difference of the covariance matrices of the OLSE and BLUE, *Linear Algebra Appl.*, 1992, 176, 121–130
- [51] McElroy F.W., A necessary and sufficient condition that ordinary least-squares estimators be best linear unbiased, *J. Amer. Stat. Assoc.*, 1967, 62, 1302–1304
- [52] Milliken G.A., Albohali M., On necessary and sufficient conditions for ordinary least squares estimators to be best linear unbiased estimators, *Amer. Statistician*, 1984, 38, 298–299
- [53] Norlén U., The covariance matrices for which least squares is best linear unbiased, *Scand. J. Statist.*, 1975, 2, 85–90
- [54] Styan G.P.H., When does least squares give the best linear unbiased estimate?, in: D.G. Kabe, R.P. Gupta (Eds.), *Multivariate Statistical Inference*, North-Holland, Amsterdam, 1973, 241–246
- [55] Tian Y., On equalities of estimations of parametric functions under a general linear model and its restricted models, *Metrika*, 2010, 72, 313–330
- [56] Tian Y., On properties of BLUEs under general linear regression models, *J. Stat. Plann. Inference*, 2013, 143, 771–782
- [57] Tian Y., Zhang J., Some equalities for estimations of partial coefficients under a general linear regression model, *Stat. Papers*, 2011, 52, 911–920
- [58] Tian Y., Zhang X., On connections among OLSEs and BLUEs of whole and partial parameters under a general linear model, *Stat. Prob. Lett.*, 2016, 112, 105–112
- [59] Abadir K.M., Magnus J.R., *Matrix Algebra*, Cambridge University Press, 2005
- [60] Banerjee S., Roy A., *Linear Algebra and Matrix Analysis for Statistics*, CRC Press, New York, 2014
- [61] Bapat R.B., *Linear Algebra and Linear Models*, 3rd ed., Springer, Berlin Heidelberg, 2012
- [62] Eldén L., *Matrix Methods in Data Mining and Pattern Recognition*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, 2007
- [63] Fieller N., *Basics of Matrix Algebra for Statistics with R*, Chapman and Hall/CRC, 2015
- [64] Gentle J.E., *Numerical Linear Algebra for Applications in Statistics*, Springer, Berlin Heidelberg, 1998
- [65] Gentle J.E., *Matrix Algebra: Theory, Computations, and Applications in Statistics*, Springer, Berlin Heidelberg, 2007
- [66] Graybill F.A., *Matrices with Applications in Statistics*, 2nd ed., Brooks/Cole, 2002
- [67] Harville D.A., *Matrix Algebra From a Statistician's Perspective*, Springer, New York, 1997
- [68] Harville D.A., *Matrix Algebra: Exercises and Solutions*, Springer, New York, 2001
- [69] Healy M.J.R., *Matrices for Statistics*, 2nd ed., Oxford University Press, 2000
- [70] Magnus J.R., Neudecker H., *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Revised edition of the 1988 original, Wiley, New York, 1999

- [71] Rao C.R., Rao M.B., *Matrix Algebra and Its Applications to Statistics and Econometrics*, World Scientific, Singapore, 1998
- [72] Schott J.R., *Matrix Analysis for Statistics*, 2nd ed., Wiley, Hoboken, NJ, 2005
- [73] Searle S.R., *Matrix Algebra Useful for Statistics*, Wiley, New York, 1982
- [74] Seber G.A.F., *A Matrix Handbook for Statisticians*, Wiley, New York, 2008
- [75] Lu C., Gan S., Tian Y., Some remarks on general linear model with new regressors, *Stat. Prob. Lett.*, 2015, 97, 16–24
- [76] Tian Y., A matrix handling of predictions under a general linear random-effects model with new observations, *Electron. J. Linear Algebra*, 2015, 29, 30–45
- [77] Tian Y., Jiang B., Equalities for estimators of partial parameters under linear model with restrictions, *J. Multivariate Anal.*, 2016, 143, 299–313
- [78] Tian Y., Jiang B., An algebraic study of BLUPs under two linear random-effects models with correlated covariance matrices, *Linear Multilinear Algebra*, 2016, 64, 2351–2367
- [79] Zhang X., Tian Y., On decompositions of BLUEs under a partitioned linear model with restrictions, *Stat. Papers*, 2016, 57, 345–364