A general prediction analysis to linear random-effects models with restrictions and new observations

Yongge Tian

CEMA, Central University of Finance and Economics, Beijing 100081, China

Abstract. This paper presents a unified approach to the problem of best linear unbiased prediction (BLUP) of a joint vector of all unknown parameters in a general linear random-effects model (LRM) with restrictions and new observations via some state-of-the-art tools in matrix mathematics. We first establish the fundamental matrix equation and the exact algebraic expression for calculating the Best Linear Unbiased Predictor (BLUP) of a joint vector of all unknown parameters in the LRM by directly solving a constrained quadratic matrix-valued function optimization problem. We then present a variety of special cases and uniform decompositions of the BLUP, as well as many algebraic and statistic properties of the BLUP. In particular, we show how to use various matrix rank/inertia formulas and matrix tricks in establishing and simplifying various complicated matrix expressions related to the covariances matrices of BLUPs, and to derive necessary and sufficient conditions for equalities and equalities of covariance matrices of BLUPs to hold. The whole work in this paper provides a precise algebraic study to LRM, and can be utilized as standard tools in statistical analysis and inference of various types of LRM.

Keywords: linear random-effects model, parameter restriction, new observation, BLUP, BLUE, rank, inertia

1 Introduction

Throughout this paper, $\mathbb{R}^{m \times n}$ stands for the collection of all $m \times n$ real matrices. The symbols $A'$, $r(A)$, and $\mathcal{R}(A)$ stand for the transpose, the rank, and the range (column space) of a matrix $A \in \mathbb{R}^{m \times n}$, respectively; $I_m$ denotes the identity matrix of order $m$. The Moore–Penrose generalized inverse of $A$, denoted by $A^+$, is defined to be the unique solution $X$ satisfying the four matrix equations $AGA$, $GAG = G$, $(AG)' = AG$, and $(GA)' = GA$. Further, let $P_A$, $E_A$, and $F_A$ stand for the three orthogonal projectors (symmetric idempotent matrices) $P_A = AA'$, $E_A = A\perp = I_m - AA'$, and $F_A = I_n - A^+A$. Both $i_+(A)$ and $i_-(A)$, called the partial inertia of $A = A' \in \mathbb{R}^{m \times m}$, are defined to be the number of the positive and negative eigenvalues of $A$ counted with multiplicities, respectively, both of which satisfy $r(A) = i_+(A) + i_-(A)$. For brief, we use $(A)$ to denote the both numbers. $A \succeq 0$, $A \succ 0$, $A \prec 0$, and $A \preceq 0$ mean that $A$ is a symmetric positive definite, positive semi-definite, negative definite, negative semi-definite matrix, respectively. Two symmetric matrices $A$ and $B$ of the same size are said to satisfy the inequalities $A \succeq B$, $A \succ B$, $A \prec B$, and $A \preceq B$ in the Löwner partial ordering if $A - B$ is positive definite, positive semi-definite, negative definite, and negative semi-definite, respectively. It is well known that the Löwner partial ordering is a surprisingly strong and useful property between two Hermitian matrices. For more issues about connections between the inertias and the Löwner partial ordering of Hermitian matrices, as well as applications of the matrix inertia and Löwner partial ordering in statistics, see, e.g., [36, 45, 48].

Consider a general Linear Random-effects Model (LRM for short) with parameter restrictions defined by

$$y = X\alpha + \varepsilon, \quad \beta = A\gamma + \varepsilon, \quad B\beta = b,$$

where in the first-stage model,

- $y \in \mathbb{R}^{n \times 1}$ is a vector of observable response variables,
- $X \in \mathbb{R}^{n \times p}$ is a known matrix of arbitrary rank,
- $\varepsilon \in \mathbb{R}^{n \times 1}$ is a vector of unobservable random errors, while in the second-stage model,
- $\beta \in \mathbb{R}^{p \times 1}$ is a vector of unobservable random variables,
- $A \in \mathbb{R}^{p \times q}$ is known matrix of arbitrary rank,
- $\alpha \in \mathbb{R}^{q \times 1}$ is a vector of fixed but unknown parameters (fixed effects),
- $\gamma \in \mathbb{R}^{q \times 1}$ is a vector of unobservable random variables (random effects),
- in the third matrix equation, $B \in \mathbb{R}^{r \times p}$ is known matrix of arbitrary rank,
- $b \in \mathbb{R}^{r \times 1}$ is a known vector with $b \in \text{range}(BA)$.

Random-effects models are statistical models of parameters that vary at more than one level, which have different names in data analysis according to their originations, such as, multilevel models, hierarchical models, nested models, random-coefficient models, split-plot designs, etc. Model (1) is also called a nested linear model or two-stage hierarchical linear model in the statistical literature, where the two equations are called the first-stage model and the second-stage model, respectively. Linear models that involve both fixed effects and random effects are now widely used to account for the variability due to different factors that influence response variables, while statistical inferences concerning LRM are now an important part in data analysis and have attracted much attention since 1970s.

E-mail address: yongge.tian@gmail.com
Assume now that new observations of response variables in the future follow the model

\[ \mathcal{M}_f: \ y_f = X_f \beta_f + \epsilon_f, \ \beta = A \alpha + \gamma, \ B \beta = b, \]  

(2)

where

- \( y_f \in \mathbb{R}^{n_f \times 1} \) is a vector of future observations (unknown),
- \( X_f \in \mathbb{R}^{n \times p} \) is a known model matrix associated with the future observations,
- \( \epsilon_f \in \mathbb{R}^{n_f \times 1} \) is a vector of measurement errors associated with future observations.

The parameter restriction \( B \beta = b \) in (1) and (2) is often available as an extraneous information for the unknown parameter vector \( \beta \) to satisfy. This restriction often occurs, for example, in the linear hypothesis testing on the parameter vector in (1). Linear models with parameter restrictions belong to the classic studies in regression analysis; see, e.g., [2, 10, 16]. Because of the occurrence of the restricted parameters in (1), derivations of estimators of unknown parameter vector \( \beta \) in (1) usually become complicated.

In the investigation of restricted linear models, (1) is usually handled by transforming into an implicitly restricted model. The most popular transformation is combining the three parts in (1) as new vector form

\[ \tilde{\mathcal{M}}: \ \tilde{y} = \tilde{X} \beta + \tilde{\epsilon} = \tilde{X} \alpha + \tilde{X} \gamma + \tilde{\epsilon}, \]  

(4)

while combining (2) and (4) gives

\[ \tilde{\mathcal{M}}: \ \tilde{y} = \tilde{X} \beta + \tilde{\epsilon} = \tilde{X} \alpha + \tilde{X} \gamma + \tilde{\epsilon}, \]  

(6)

In order to establish a unified theory on statistical inferences of (6), we assume that the expectation vector and dispersion matrix of the combined random vector in (6) are given by

\[ E \begin{bmatrix} \gamma \\ \epsilon_f \end{bmatrix} = 0, \ D \begin{bmatrix} \gamma \\ \epsilon_f \end{bmatrix} = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{23} \\ \Sigma_{31} & \Sigma_{32} & \Sigma_{33} \end{bmatrix} := \Sigma, \]  

(7)

where we don’t attach any further restrictions to the patterns of the submatrices \( \Sigma_{ij} \) in (7), although they are usually taken as certain prescribed forms for a specified random-effects model in statistical literatures. Correspondingly, the dispersion matrix of the combined random vector \( \tilde{y} \) in (4) is given by

\[ D(\tilde{y}) = D(\tilde{X} \gamma + \tilde{\epsilon}) = Z \Sigma Z', \ Z = [\tilde{X}, \tilde{I}_n, 0], \ \tilde{I}_n = \begin{bmatrix} I_n \\ 0 \end{bmatrix}. \]  

(8)

This \( D(\tilde{y}) \) is a known matrix under the assumptions in (1)–(7), and will occur in the statistical inferences of (4) and (6). Note from (6) and (7) that under the general assumptions in (1)–(7), \( \tilde{y} \) and \( y_f \) are correlated. Hence, it is possible to give predictions of \( y_f, X_f \beta \), and \( \epsilon_f \) in (2) from the original observation vector \( \tilde{y} \) in (4) under the assumptions in (1)–(7).

In the statistical analysis and inference of a regression model, a first step is to establish estimators of the unknown parameters in the model. However, in various applications, it is utmost important for practitioners to predict future values of response variables in regression models. Prediction is also an important aspect of decision-making process through statistical methodology, while linear regression models play important roles in predicting unknown values of response variables corresponding to known values of explanatory variables. Traditional method in statistical inferences is to establish estimators/predictors of unknown parameters separately, so that formulas for each unknown parameter vector are complicated, and their statistical performances are confused. In this case, it is most desirable to simultaneously identify the estimators and predictors that correspond to both the fixed- and random-effects components in linear
models. There seems to be very little general literature explicitly discussing the estimations/predictions of all unknown parameters in linear models with original and future observations. [37] showed a lemma on optimization of a matrix function in the Löwner partial ordering and established a unified theory of linear estimations/predictions of all unknown parameters in general linear models with original and future observations by directly solving certain constrained quadratic matrix-valued function optimization problems in the Löwner partial ordering. Some other work on simultaneous estimations/predictions of joint vectors of unknown parameters under regression models can be found, e.g., in [8, 43, 51]. As formulated in [37], a general vector of parametric functions involving the fixed effects, random effects, and error terms in (6) is given by

$$\phi = F\alpha + G\gamma + He + H_f\epsilon_f,$$

(9)

where $F \in \mathbb{R}^{s \times k}$, $G \in \mathbb{R}^{s \times p}$, $H \in \mathbb{R}^{s \times n}$, and $H_f \in \mathbb{R}^{s \times m'}$ are given matrices of arbitrary ranks. In this case,

$$E(\phi) = F\alpha, \quad D(\tilde{\phi}) = \begin{bmatrix} D(\tilde{\gamma}) & Cov(\tilde{\gamma}, \phi) \\ Cov(\phi, \tilde{\gamma}) & D(\phi) \end{bmatrix} = \begin{bmatrix} Z\Sigma\Sigma' & Z\Sigma J' \\ J\Sigma\Sigma' & J\Sigma J' \end{bmatrix},$$

(10)

where $J = [G, H, H_f]$. Eq. (9) includes all vector operations in (1)–(6) as its special cases. For instance, if $F = TXA$, $G = TX$, $H = T \begin{bmatrix} I_n \\ 0 \\ 0 \end{bmatrix}$, and $H_f = T \begin{bmatrix} 0 \\ 0 \\ 1_{m'} \end{bmatrix}$, then (9) becomes

$$\phi = TXA\alpha + TX\gamma + T\bar{\epsilon} = T\bar{\gamma},$$

which includes $\gamma$, $\gamma_f$, and $\bar{\gamma}$ as its special cases. Another well-known form of $\phi$ is the following target function discussed by [8, 9, 43, 51], which allows the prediction of both $y_f$ and $E(y_f)$,

$$\tau = \lambda y_f + (1 - \lambda)E(y_f) = X_fA\alpha + \lambda X_f\gamma + \lambda \epsilon_f,$$

where $\lambda$ (0 ≤ $\lambda$ ≤ 1) is a non-stochastic scalar assigning weights to actual and expected values of $y_f$. Other special forms of $\phi$ in (9) and their predictions/estimations can be found, e.g., in [21, 27, 35]. Thus, the statistical inference of $\phi$ is a comprehensive work, and will play prescriptive role for various special statistical inference problems under (6) from both theoretical and applied points of view. Note that there are 14 given matrices in (1)–(7) and (9). Hence, the statistical inference of $\phi$ is not easy task, and we will encounter many tedious matrix operations for the given 14 matrices, as demonstrated in Section 3 below.

In the inferences theory of linear models, there has been considerable interest in establishing estimators of the unknown parameters by certain linear functions of the observed response vectors in the models. In order to establish standard theory of the Best Linear Unbiased Predictor (BLUP) of the general unknown parameter vector $\phi$ in (9) under (4), we need the following classic statistical concepts and definitions originated from [17]. The linear statistic $L\hat{\gamma}$, where $L \in \mathbb{R}^{s \times (n+q)}$, is said to have the same expectation with $\phi$ in (9), or $\phi$ in (9) is said to be predictable by $\hat{\gamma}$ in (4), if and only if $E(L\hat{\gamma} - \phi) = 0$ holds. The matrix $L$ satisfying this equality is not necessarily unique. In this case, if there exists a matrix $L$ such that

$$E(L\hat{\gamma} - \phi) = 0 \quad \text{and} \quad D(L\hat{\gamma} - \phi) = \min$$

(11)

hold in the the Löwner partial ordering, then the linear statistic $L\hat{\gamma}$ is defined to be the BLUP of $\phi$ in (9), and is denoted by

$$L\hat{\gamma} = \text{BLUP}(\phi) = \text{BLUP}(F\alpha + G\gamma + He + H_f\epsilon_f).$$

(12)

In particular, if $\phi = F\alpha$, then the linear statistic $L\hat{\gamma}$ satisfying (11) is the well-known BLUE of $F\alpha$ under (4), and is denoted by

$$L\hat{\gamma} = \text{BLUE}(F\alpha).$$

(13)

The theory of linear predictors/estimators under linear regression models belongs to the classical methods of mathematical statistics. The Best Linear Unbiased Estimators (BLUEs) of unknown parameters, such as $\alpha$ in (1), as well as the BLUP of unknown random variables such as $\beta$ in (1), generally depend on the dispersion matrix of the observed random vector, and on the vector of covariances between the observed vector and the predicted random variable, as demonstrated in Theorem 2 below. To account for general prediction/estimation problems of unknown parameters in linear models, it is common
Hence, the expectation and the covariance matrix of predictors/estimators can easily be established.

The algebraic expressions of predictors/estimators of unknown parameter vectors, as well as properties of the quadratic matrix-valued function minimization problem in the Löwner partial ordering, while exact algebraic expressions of predictors/estimators of the unknown parameter vector in the model. It has the advantage that no further distributional assumptions of error terms need to be made in the model, other then about the first- and second-order moments. In particular, parameter vector in the model. In order to finish this task, we need to use many rank formulas for matrices and their generalized inverses, while heavy matrix operations are involved.

2 Equation and formulas for BLUPs of all unknown parameters

In the statistical analysis and inference of a given linear regression model, many types of optimality criterion in mathematics were used to define predictors/estimators of unknown parameters in the model. In comparison, the minimization of matrix mean squared error (MMSE) of unknown parameter vector with respect to linear statistic is a primary criterion for deriving best predictors/estimators of the unknown parameter vector in the model. It has the advantage that no further distributional assumptions of error terms need to be made in the model, other then about the first- and second-order moments. In particular, the minimization the covariance matrix of a linear statistic with respect to predictable/estimable vector terms need to be made in the model, other then about the first- and second-order moments. In particular, the minimization the covariance matrix of a linear statistic with respect to predictable/estimable vector terms need to be made in the model, other then about the first- and second-order moments.

Concerning the predictability of \( \phi \) in (9), we have the following result.

**Lemma 1.** The vector \( \phi \) in (9) is predictable by \( \hat{y} \) in (4), i.e., \( E(L\hat{y} - \phi) = 0 \) holds for some \( L \), if and only if

\[
\mathcal{R}[\hat{X}A'] \supseteq \mathcal{R}(F').
\]

Concerning the matrix equation and the exact algebraic expression of the BLUP of \( \phi \) in (9), as well as the properties of the BLUPs, we have the following comprehensive results.

**Theorem 2.** (Fundamental BLUP equation) Assume that \( \phi \) in (9) is predictable by \( \hat{y} \) in (4), namely, (17) holds, and let \( \hat{X}_n \), \( Z \), and \( J \) be as given in (8) and (10), respectively. Then,

\[
E(L_0\hat{y} - \phi) = 0 \quad \text{and} \quad D(L_0\hat{y} - \phi) = \min \Leftrightarrow L_0[\hat{X}A, D(\hat{y})(\hat{X}A)'^+] = [F, Cov(\phi, \hat{y})(\hat{X}A)^+] .
\]

The matrix equation in (18), called the fundamental BLUP equation, is consistent as well under (17). In this case, BLUP(\( \phi \)) can be written as

\[
\text{BLUP}(\phi) = L_0\hat{y},
\]
Further, the following results hold.

(a) \( r[\hat{X}A, \Sigma\Sigma'(\hat{X}A)^+] = r[\hat{X}A, \Sigma], \Sigma[\hat{X}A, \Sigma\Sigma'(\hat{X}A)^+] = \Sigma[\hat{X}A, \Sigma], \) and \( \Sigma(\hat{X}A) \cap \Sigma[\Sigma\Sigma'(\hat{X}A)^+] = \{0\}. \)

(b) \( L \) is unique if and only if \( r[\hat{X}A, \Sigma] = n + q. \)

(c) BLUP(\( \phi \)) is unique with probability 1 if and only if \( \hat{y} \in \Sigma[\hat{X}A, \Sigma], \) i.e., (4) is consistent.

(d) The dispersion matrix of BLUP(\( \phi \)) is
\[
D[BLUP(\phi)] = \left( [F, J\Sigma Z'(\hat{X}A)^+] \parallel [\hat{X}A, D(y)(\hat{X}A)^+] \right) + U[\hat{X}A, D(y)(\hat{X}A)^+] \right),
\]
where \( U \in \mathbb{R}^{s \times (n+q)} \) is arbitrary. The corresponding \( f(L_0) \) under (16), namely, the dispersion matrix of the difference of BLUP(\( \phi \)) - \( \phi \) is
\[
f(L_0) = D[BLUP(\phi) - \phi] = \left( [F, J\Sigma Z'(\hat{X}A)^+] \parallel [\hat{X}A, D(y)(\hat{X}A)^+] \right) \Sigma
\]
\[
\times \left( [F, J\Sigma Z'(\hat{X}A)^+] \parallel [\hat{X}A, D(y)(\hat{X}A)^+] \right).
\]

Further, the following results hold.

(a) \( r[\hat{X}A, \Sigma\Sigma'(\hat{X}A)^+] = r[\hat{X}A, \Sigma], \Sigma[\hat{X}A, \Sigma\Sigma'(\hat{X}A)^+] = \Sigma[\hat{X}A, \Sigma], \) and \( \Sigma(\hat{X}A) \cap \Sigma[\Sigma\Sigma'(\hat{X}A)^+] = \{0\}. \)

(b) \( L \) is unique if and only if \( r[\hat{X}A, \Sigma] = n + q. \)

(c) BLUP(\( \phi \)) is unique with probability 1 if and only if \( \hat{y} \in \Sigma[\hat{X}A, \Sigma], \) i.e., (4) is consistent.

(d) The dispersion matrix of BLUP(\( \phi \)) is
\[
D[BLUP(\phi)] = \left( [F, J\Sigma Z'(\hat{X}A)^+] \parallel [\hat{X}A, D(y)(\hat{X}A)^+] \right) + U[\hat{X}A, D(y)(\hat{X}A)^+] \right),
\]
where \( U \in \mathbb{R}^{s \times (n+q)} \) is arbitrary. The corresponding \( f(L_0) \) under (16), namely, the dispersion matrix of the difference of BLUP(\( \phi \)) - \( \phi \) is
\[
f(L_0) = D[BLUP(\phi) - \phi] = \left( [F, J\Sigma Z'(\hat{X}A)^+] \parallel [\hat{X}A, D(y)(\hat{X}A)^+] \right) \Sigma
\]
\[
\times \left( [F, J\Sigma Z'(\hat{X}A)^+] \parallel [\hat{X}A, D(y)(\hat{X}A)^+] \right).
\]
Corollary 3. Let $\phi$ be as given in (9). Then, the following results hold.

(a) If $\phi$ is predictable by $\tilde{y}$ in (4), then $T\phi$ is predictable by $\tilde{y}$ in (4) as well for any matrix $T \in \mathbb{R}^{t \times s}$, and $\text{BLUE}(T\phi) = T\text{BLUE}(\phi)$ holds.

(b) If $\phi$ in (9) is predictable by $\tilde{y}$ in (4), then $F\tilde{a}$ is estimable by $\tilde{y}$ in (4) as well, and the $\text{BLUP}$ of $\phi$ can be decomposed as the sum

$$\text{BLUP}(\phi) = \text{BLUE}(F\tilde{a}) + \text{BLUE}(G\gamma) + \text{BLUE}(He) + \text{BLUE}(Hf_1\varepsilon_1),$$

and the following formulas for covariance matrices hold

$$\text{Cov}\{\text{BLUE}(F\tilde{a}), \text{BLUE}(G\gamma + He + Hf_1\varepsilon_1)\} = 0,$$

$$D[\text{BLUP}(\phi)] = D[\text{BLUE}(F\tilde{a})] + D[\text{BLUE}(G\gamma + He + Hf_1\varepsilon_1)].$$

(c) If $\alpha$ in (9) is estimable under (4), then the $\phi$ in (9) is predictable by $\tilde{y}$ in (4). In this case,

$$\text{BLUP} \begin{bmatrix} \alpha \\ \gamma \\ \varepsilon \\ \varepsilon_f \end{bmatrix} = \begin{bmatrix} \text{BLUE}(\alpha) \\ \text{BLUP}(\gamma) \\ \text{BLUP}(\varepsilon) \\ \text{BLUP}(\varepsilon_f) \end{bmatrix},$$

$$\text{BLUP}(\phi) = \text{FBLUE}(\alpha) + \text{GBLUP}(\gamma) + \text{HBLUP}(\varepsilon) + Hf\text{BLUP}(\varepsilon_f).$$

Corollary 4. Let $\Sigma$ and $Z$ be as given in (7) and (8), and assume that

$$\phi_1 = F_1\alpha + G_1\gamma + H_1\varepsilon + H_1\varepsilon_f, \quad \phi_2 = F_2\alpha + G_2\gamma + H_2\varepsilon + H_2\varepsilon_f$$

are predictable under (6), where $F_1, F_2 \in \mathbb{R}^{s \times k}, G_1, G_2 \in \mathbb{R}^{s \times p}, H_1, H_2 \in \mathbb{R}^{r \times n},$ and $H_1, H_2 \in \mathbb{R}^{s \times n_y}$ are known matrices, and denote $J_i = [G_i, H_i, H_{1i}], \quad i = 1, 2.$ Then, the following results hold.

(a) The sum $\phi_1 + \phi_2$ is predictable under (6), and its $\text{BLUP}$ satisfies

$$\text{BLUP}(\phi_1 + \phi_2) = \text{BLUP}(\phi_1) + \text{BLUP}(\phi_2).$$

(b) $\text{BLUP}(\phi_1) = \text{BLUP}(\phi_2) \iff F_1 = F_2$ and $\mathcal{R}(Z\Sigma J_1') = \mathcal{R}(Z\Sigma J_2').$

Finally, we give some equalities of the $\text{BLUP}$ of $\tilde{y}$ in (6).

Corollary 5. (Fundamental $\text{BLUP}$ decompositions) The combined vector $\tilde{y}$ in (6) is predictable by $\tilde{y}$ in (4) if and only if $\mathcal{R}(X\tilde{A})' \supseteq \mathcal{R}(X_fy_f')$. In this case, $B\beta$ is predictable by $\tilde{y}$ in (4), and the following decomposition equalities hold

$$y = \text{BLUP}(X\beta) + \text{BLUP}(\varepsilon) = \text{BLUE}(X\alpha) + \text{BLUE}(X\gamma) + \text{BLUE}(\varepsilon),$$

$$b = \text{BLUP}(B\beta) = \text{BLUE}(BA\alpha) + \text{BLUE}(B\gamma),$$

$$\text{BLUP}(y_f) = \text{BLUP}(X_f\beta) + \text{BLUP}(\varepsilon_f) = \text{BLUE}(X_fA\alpha) + \text{BLUE}(X_f\gamma) + \text{BLUE}(\varepsilon_f),$$

$$\text{BLUP}(\tilde{y}) = \begin{bmatrix} \gamma \\ b \\ \text{BLUP}(y_f) \end{bmatrix}.$$
Theorem 6. (Fundamental rank/inertia formulas) Let $\phi$ be as given in (9), and assume that $\phi$ is predictable by $\hat{y}$ in (4). Also, let $S \in \mathbb{R}^{xx}$ be a symmetric matrix, and denote

$$
M = \begin{bmatrix}
D(\hat{y}) & Cov\{\hat{y}, \phi\} & \hat{X}A \\
Cov\{\phi, \hat{y}\} & D(\phi) - S & F \\
(\hat{X}A)' & F' & 0
\end{bmatrix} = \begin{bmatrix}
Z\Sigma Z' & Z\Sigma J' & \hat{X}A \\
J\Sigma Z' & J\Sigma J' - S & F \\
(\hat{X}A)' & F' & 0
\end{bmatrix}.
$$

Then, the following inertia and rank formulas hold. In consequence, the following results hold.

(a) $D[BLUP(\phi) - \phi] \succ 0 \Leftrightarrow i_+(M) = r[\hat{X}A, Z\Sigma] + s.$

(b) $D[BLUP(\phi) - \phi] \prec 0 \Leftrightarrow i_-(M) = r(\hat{X}A) + s.$

(c) $D[BLUP(\phi) - \phi] \succ 0 \Leftrightarrow i_+(M) = r(\hat{X}A).$

(d) $D[BLUP(\phi) - \phi] \preceq 0 \Leftrightarrow i_+(M) = r[\hat{X}A, Z\Sigma].$

(e) $D[BLUP(\phi) - \phi] = 0 \Leftrightarrow r(M) = r[\hat{X}A, Z\Sigma] + r(\hat{X}A).$

Many consequences can be derived from Theorem 6 for different choices of the joint vector $\phi$ and the symmetric matrix $S$. We present two of them in the following two corollaries.

Corollary 7. Let $\phi$ be as given in (9), and assume that $\phi$ is predictable by $\hat{y}$ in (4). Then,

$$
r(D[BLUP(\phi) - \phi]) = r \left[ D(\hat{y}), \hat{X}A \right].
$$

In consequence, the following results hold.

(a) $D[BLUP(\phi) - \phi] \succ 0 \Leftrightarrow r \left[ D(\hat{y}), \hat{X}A \right] = r[D(\hat{y}), \hat{X}A] + s.$

(b) $D[BLUP(\phi) - \phi] = 0 \Leftrightarrow BLUP(\phi) = \phi$ holds with probability 1 $\Leftrightarrow$

$$
r \left[ D(\hat{y}), \hat{X}A \right] = r[D(\hat{y}), \hat{X}A].
$$

Corollary 8. Assume that $\phi$ in (9) is predictable by $\hat{y}$ in (4), and let $L$ satisfy $E(L\hat{y} - \phi) = 0$. Then,

$$
r(D(L\hat{y} - \phi) - D[BLUP(\phi) - \phi]) = r \left[ L\Sigma Z' - J\Sigma J' \right] \hat{X}A - r(\hat{X}A).
$$

We established a group of standard results on the matrix equation and the exact algebraic expression of the BLUP of a joint vector of all unknown parameters under the general assumptions in (1)–(10). We also obtained a set of analytical formulas for calculating ranks/inertias of covariance matrices of the BLUPs/BLUEs, and used the formulas to characterize a variety of basic equalities and inequalities for the covariance matrices of the BLUPs/BLUEs. Linear models, estimators, predictors, dispersion matrices, etc were conceptual foundations in regression theory, and are prominent objects of study in statistical analysis and inference. In particular, the variety of features and performances of BLUPs/BLUEs were widely discussed in the statistical literature. However, there are still many new and interesting problems on these classic issues that can be proposed from theoretical and applied points of view, and can be solved by various known and novel tools in mathematics, as demonstrated in the previous sections. It is believed that more insightful and sophisticated formulas, equalities, and inequalities associated with predictors/estimators can be derived under linear models with various general assumptions, which, we believe, will help deeply rebuilding mathematical foundation of regression analysis from theoretical point of view.
We also need the following known formulas on ranks of matrices.

The assertions in Lemma 9 arise directly from the definitions of rank and inertia of matrix, which
established in the past decades. We now are able to use matrix inertia formulas in establishing and
simplifying various complicated matrix expressions, and in the intuitive and rigorous derivations of matrix
equalities and inequalities that involve generalized inverses of matrices. In statistic analysis and inference of linear regression models, we can use matrix rank/inertia formulas to make many original, innovative
equalities and inequalities that involve generalized inverses of matrices. In statistic analysis and inference
of linear regression models, we can use matrix rank/inertia formulas to make many original, innovative
applications in some lemmas below to make the paper self-contained.

**Lemma 9.** Let $A, B \in \mathbb{R}^{m \times n}$, or $A = A', B = B' \in \mathbb{R}^{n \times m}$. Then, the following results hold.

(a) $A = B$ if and only if $r(A - B) = 0$.

(b) $A - B$ is nonsingular if and only if $r(A - B) = m$.

(c) $A \succ B$ ($A \prec B$) if and only if $i_+(A - B) = m$ ($i_-(A - B) = m$).

(d) $A \succ B$ ($A \prec B$) if and only if $i_-(A - B) = 0$ ($i_+(A - B) = 0$).

The assertions in Lemma 9 arise directly from the definitions of rank and inertia of matrix, which
show the simple links between ranks and inertias of matrices with equalities and inequalities of matrices.
In order to technically use Lemma 14 in establishing inequalities for the dispersion matrices in Section 3,
we need a variety of classic and new matrix rank/inertia formulas. The following are some known results
due to [32 ] on ranks of matrices.

We also need the following known formulas on ranks of matrices.
Lemma 10 ([32]). Let $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{m \times k}$, and $C \in \mathbb{R}^{l \times n}$. Then,

$$r[A, B] = r(A) + r(E_AB) = r(B) + r(E_BA),$$  \hspace{1cm} (41)

$$r\begin{bmatrix} A \\ C \end{bmatrix} = r(A) + r(CF_A) = r(C) + r(AF_C),$$  \hspace{1cm} (42)

$$r\begin{bmatrix} AA' \\ B' \end{bmatrix} = r[A, B] + r(B).$$  \hspace{1cm} (43)

Lemma 11 ([44]). If $\mathcal{R}(A_1') \subseteq \mathcal{R}(B_1')$, $\mathcal{R}(A_2') \subseteq \mathcal{R}(B_1')$, $\mathcal{R}(A_2') \subseteq \mathcal{R}(B_2')$ and $\mathcal{R}(A_3) \subseteq \mathcal{R}(B_2)$, then

$$r(A_1B_1'A_2) = r\begin{bmatrix} B_1 & A_2 \\ A_1 & 0 \end{bmatrix} - r(B_1),$$  \hspace{1cm} (44)

$$r(A_1B_1'A_2B_2'A_3) = r\begin{bmatrix} 0 & B_2 & A_3 \\ B_1 & A_2 & 0 \\ A_1 & 0 & 0 \end{bmatrix} - r(B_1) - r(B_2).$$  \hspace{1cm} (45)

The fundamental inertia formulas in the following lemma are well known or follow directly from the definition of inertia of symmetric matrix.

Lemma 12. Let $A = A' \in \mathbb{R}^{m \times m}$, $B = B' \in \mathbb{R}^{n \times n}$, $Q \in \mathbb{R}^{m \times n}$, and assume that $P \in \mathbb{R}^{m \times m}$ is nonsingular. Then,

$$i_{\pm}(PAP') = i_{\pm}(A) \quad \text{(Sylvester's law of inertia)},$$  \hspace{1cm} (46)

$$i_{\pm}(A^+) = i_{\pm}(A), \quad i_{\pm}(-A) = i_{\mp}(A),$$  \hspace{1cm} (47)

$$i_{\pm}\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} = i_{\pm}(A) + i_{\pm}(B),$$  \hspace{1cm} (48)

$$i_{\pm}\begin{bmatrix} 0 & Q \\ Q' & 0 \end{bmatrix} = i_{\pm}\begin{bmatrix} 0 & Q \\ Q' & 0 \end{bmatrix} = r(Q).$$  \hspace{1cm} (49)

In order to characterize equalities and inequalities of matrices from the assertions in Lemma 14, we also need the following matrix inertia formulas.

Lemma 13 ([45]). Let $A = A' \in \mathbb{R}^{m \times m}$, $B \in \mathbb{R}^{m \times n}$, and $D = D' \in \mathbb{R}^{n \times n}$. Then,

$$i_{\pm}\begin{bmatrix} A & B \\ B' & D \end{bmatrix} = i_{\pm}(A) + i_{\pm}\begin{bmatrix} 0 & E_AB \\ B'E_A & D - B'A^+B \end{bmatrix}.$$  \hspace{1cm} (50)

In particular,

$$i_{\pm}\begin{bmatrix} A & B \\ B' & D \end{bmatrix} = i_{\pm}(A) + i_{\pm}(D - B'A^+B) \quad \text{if } \mathcal{R}(B) \subseteq \mathcal{R}(A).$$  \hspace{1cm} (51)

The following lemma is well known; see [34].

Lemma 14. The linear matrix equation $AX = B$ is consistent if and only if $r[A, B] = r(A)$, or equivalently, $AA^+B = B$. In this case, the general solution of the equation can be written in the following parametric form $X = A^+B + (I - A^+A)U$, where $U$ is an arbitrary matrix.

The following result on analytical solution of a constrained matrix-valued function optimization problem was given in [46], see also [?], which can directly be utilized to establish a linear matrix equation associated with the BLUP of $\phi$ in (9).

Lemma 15. Let

$$f(L) = (LC + D)(MC + D)' \quad \text{s.t. } LA = B,$$  \hspace{1cm} (52)

where $A \in \mathbb{R}^{p \times q}$, $B \in \mathbb{R}^{n \times q}$, $C \in \mathbb{R}^{m \times n}$, $D \in \mathbb{R}^{n \times m}$ are given, $M \in \mathbb{R}^{m \times m}$ is positive semi-definite, and the matrix equation $LA = B$ is consistent. Then, there always exists a solution $L_0$ of $L_0A = B$ such that

$$f(L) \Rightarrow f(L_0)$$  \hspace{1cm} (53)

holds for all solutions of $LA = B$. In this case, the matrix $L_0$ satisfying (53) is determined by the following consistent matrix equation

$$L_0[A, CMC'A^+] = [B, -DMC'A^+]$$  \hspace{1cm} (54)
while the general expression of $L_0$ and the corresponding $f(L_0)$ are given by

$$L_0 = \arg \min_{L_{a-b}} f(L) = [B, -DMC'A^\perp][A, CMC'A^\perp]^+ + V|A, CMC'|^+, \quad (55)$$

$$f(L_0) = \min_{L_{a-b}} f(L) = KMK' - KMC'(A^\perp CMC'A^\perp)^+ CMK',$n

$$f(L) - f(L_0) = (LCMC'A^\perp + DMC'A^\perp)(A^\perp CMC'A^\perp)^+(LCMC'A^\perp + DMC'A^\perp)^\perp, \quad (57)$$

where $K = BA^+C + D$, and $V \in \mathbb{R}^{n \times p}$ is arbitrary.

**Proof of Lemma 1.** It is obvious that $E(L\hat{y} - \phi) = 0 \iff L\tilde{X}\alpha - F\alpha = 0$ for all $\alpha \iff L\tilde{X} = F$. From Lemma 14, the matrix equation is consistent if and only if (17) holds. \hfill \Box

**Proof of Theorem 2.** Under (16), we see from Lemma 14 that the first part of (18) is equivalent to finding a solution $L_0$ of the consistent matrix equation $L_0\tilde{X} = F$ such that

$$f(L) \supseteq f(L_0) \text{ s.t. } L\tilde{X} = F \quad (58)$$

holds in the Löwner partial ordering. Further from Lemma 15, there always exists a solution $L_0$ of $L_0\tilde{X} = F$ such that (58) holds, and the $L_0$ is determined by the matrix equation $L_0[\tilde{X}, Z\Sigma Z'(\tilde{X})A^\perp] = [F, J\Sigma Z'(\tilde{X})A^\perp]$, establishing the matrix equation in (18). Solve the equation by Lemma 14 to give (19). Also from (56),

$$f(L_0) = D(L_0\tilde{y} - \phi) = D(\phi - L_0\tilde{y}) = (L_0Z - J)\Sigma(L_0Z - J)^\perp,$$

thus establishing (21). Applying (57) to (16) yields (22). Result (a) is well known. Result (b) follows directly from (19). Taking dispersion matrix of (19) yields (23). From (9) and (19),

$$\text{Cov} \{ \text{BLUP}(\phi), \phi \} = \text{Cov}(L\tilde{y}, \phi) = L\Sigma(L\tilde{y}, \Sigma Z'(\tilde{X})A^\perp)^+ \Sigma Z,'$$

thus establishing (24). Eq. (25) follows from (10) and (23). \hfill \Box

**Proof of Corollary 3.** The estimability of $T\phi$ follows from $\mathcal{R}[(\tilde{X})A'] \supseteq \mathcal{R}(F') = \mathcal{R}(F'T')$. Also from (19),

$$\text{BLUP}(T\phi) = \left( [TF, TJ\Sigma Z'(\tilde{X})A^\perp][\tilde{X}, Z\Sigma Z'(\tilde{X})A^\perp] + U[\tilde{X}, Z\Sigma Z'(\tilde{X})A^\perp]^\perp \right) \tilde{y}$$

$$= T \left( [F, J\Sigma Z'(\tilde{X})A^\perp][\tilde{X}, Z\Sigma Z'(\tilde{X})A^\perp] + U[\tilde{X}, Z\Sigma Z'(\tilde{X})A^\perp]^\perp \right) \tilde{y}$$

$$= \text{TBUP}(\phi),$$

where $U = TU_1$, as required for (a).

Note that $[F, J\Sigma Z'(\tilde{X})A^\perp][\tilde{X}, Z\Sigma Z'(\tilde{X})A^\perp]^\perp$ in (19) can be decomposed as

$$\left[ F, J\Sigma Z'(\tilde{X})A^\perp][\tilde{X}, Z\Sigma Z'(\tilde{X})A^\perp]^\perp$$

$$= \left[ [F, 0][\tilde{X}, Z\Sigma Z'(\tilde{X})A^\perp]^\perp + [0, G, 0, 0]\Sigma Z'(\tilde{X})A^\perp][\tilde{X}, Z\Sigma Z'(\tilde{X})A^\perp]^\perp + [0, 0, H, 0]\Sigma Z'(\tilde{X})A^\perp][\tilde{X}, Z\Sigma Z'(\tilde{X})A^\perp]^\perp \right]$$

$$= \text{BLUE}(Fa) + \text{BLUP}(G\gamma) + \text{BLUP}(H\varepsilon) + \text{BLUP}(H_i\varepsilon),$$

establishing (26). We also obtain from (19) the covariance matrix between $\text{BLUE}(Fa)$ and $\text{BLUP}(G\gamma + H\varepsilon + H_i\varepsilon)$ as follows

$$\text{Cov} \{ \text{BLUE}(Fa), \text{BLUP}(G\gamma + H\varepsilon + H_i\varepsilon) \}$$

$$= \left[ [F, 0][\tilde{X}, Z\Sigma Z'(\tilde{X})A^\perp]^\perp + Z\Sigma Z' \left( [0, J\Sigma Z'(\tilde{X})A^\perp][\tilde{X}, Z\Sigma Z'(\tilde{X})A^\perp]^\perp \right) \right]. \quad (59)$$
Applying (45) to (59) and simplifying, we obtain

\[ r(Cov\{BLUE(Fa), BLUP(G\gamma + H\epsilon + H_f\epsilon_f)\}) \]

\[ = r\left( [F, 0] [\hat{X}A, Z\Sigma Z'(\hat{X}A)^+]^+ Z\Sigma Z'(0, J\Sigma Z'(\hat{X}A)^+) Z\Sigma Z'(\hat{X}A)^+]^+ \right) \]

\[ = r\left( \begin{bmatrix} 0 & (\hat{X}A)' \\ (\hat{X}A)^+ Z\Sigma Z' \\ Z\Sigma Z' \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \right) \]

\[ - 2r[\hat{X}A, Z\Sigma Z'(\hat{X}A)^+] \]

\[ = r\left( \begin{bmatrix} 0 & (\hat{X}A)' \\ - (\hat{X}A)^+ Z\Sigma Z'(\hat{X}A)^+ \\ F, 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \right) \]

\[ - 2r[\hat{X}A, Z\Sigma Z'] \]

\[ = r[\hat{X}A, Z\Sigma Z', Z\Sigma J'] - r(\hat{X}A) \]

which implies that \( Cov\{BLUE(Fa), BLUP(G\gamma + H\epsilon + H_f\epsilon_f)\} \) is a zero matrix, establishing (27). Eq. (28) follows from (26) and (27). Eqs. (29) and (30) follow from (a) and (26). \( \square \)

**Proof of Theorem 4.** Eq. (31) follows from Corollary 3(a) and (26). From Theorem 2, the two equations for the coefficient matrices of \( L_1[\hat{X}A, Z\Sigma Z'(\hat{X}A)^+] = [F_1, J_1\Sigma Z'(\hat{X}A)^+] \), \( L_2[\hat{X}A, Z\Sigma Z'(\hat{X}A)^+] = [F_2, J_2\Sigma Z'(\hat{X}A)^+] \).

This pair of matrix equations have a common solution if and only if

\[ r\left( \begin{bmatrix} \hat{X}A & Z\Sigma Z'(\hat{X}A)^+ & \hat{X}A & Z\Sigma Z'(\hat{X}A)^+ \\ F_1 & J_1\Sigma Z'(\hat{X}A)^+ & F_2 & J_2\Sigma Z'(\hat{X}A)^+ \end{bmatrix} \right) \]

\[ = r[\hat{X}A, Z\Sigma Z'(\hat{X}A)^+, \hat{X}A, Z\Sigma Z'(\hat{X}A)^+] \]

(60)

where

\[ r\left( \begin{bmatrix} \hat{X}A & Z\Sigma Z'(\hat{X}A)^+ & \hat{X}A & Z\Sigma Z'(\hat{X}A)^+ \\ F_1 & J_1\Sigma Z'(\hat{X}A)^+ & F_2 & J_2\Sigma Z'(\hat{X}A)^+ \end{bmatrix} \right) \]

\[ = r\left( \begin{bmatrix} \hat{X}A & Z\Sigma Z'(\hat{X}A)^+ & 0 & 0 \\ 0 & 0 & F_2 - F_1 & (J_2\Sigma Z' - J_1\Sigma Z')(\hat{X}A)^+ \end{bmatrix} \right) \]

\[ = r[\hat{X}A, Z\Sigma Z'(\hat{X}A)^+] + r[F_2 - F_1, (J_2\Sigma Z' - J_1\Sigma Z')(\hat{X}A)^+] \]

\[ = r[\hat{X}A, Z\Sigma Z'(\hat{X}A)^+] \]

Hence, (60) is equivalent to \([F_2 - F_1, (J_2\Sigma Z' - J_1\Sigma Z')(\hat{X}A)^+] = 0\), which is further equivalent to (b). \( \square \)

**Proof of Theorem 6.** Note from (21) that

\[ S - D[BLUP(\phi) - \phi] \]

\[ = S - \left( [F, J\Sigma Z'(\hat{X}A)^+] [\hat{X}A, Z\Sigma Z'(\hat{X}A)^+]^+ Z - J \right) \Sigma \]

\[ \times \left( [F, J\Sigma Z'(\hat{X}A)^+] [\hat{X}A, Z\Sigma Z'(\hat{X}A)^+]^+ Z - J \right)^t. \]
Applying (51) to this expression and simplifying, we obtain

\[
\begin{align*}
&= i_\pm (S - D[\text{BLUP}(\phi) - \phi]) \\
&= i_\pm \left(S - \left[ F, J\Sigma Z'(\hat{X}A)^+ \right] [\hat{X}A, Z\Sigma Z'(\hat{X}A)^+]^+ Z - J \right) \Sigma \\
&\quad \times \left[ [F, J\Sigma Z'(\hat{X}A)^+] [\hat{X}A, Z\Sigma Z'(\hat{X}A)^+]^+ Z - J \right] \Sigma \\
&= i_\pm \left[ \Sigma \right] + \left[ \Sigma Z' \right] o \left[ F, J\Sigma Z'(\hat{X}A)^+ \right] \\
&\times \left[ [\hat{X}A, Z\Sigma Z'(\hat{X}A)^+]^+ \right] [\hat{X}A, Z\Sigma Z'(\hat{X}A)^+]^+ Z \Sigma \left[ F, J\Sigma Z'(\hat{X}A)^+ \right] \\
&= i_\pm \left( \left[ \begin{array}{ccc}
0 & -\hat{X}A & -Z\Sigma Z'(\hat{X}A)^+ \\
-\hat{X}A' & 0 & 0 & 0 \\
-\hat{X}A' & 0 & 0 & 0 \\
\end{array} \right] \right) - r[\hat{X}A, Z\Sigma Z'] \\
&= i_\pm \left( \left[ \begin{array}{ccc}
-Z\Sigma Z' & -\hat{X}A & -Z\Sigma Z'(\hat{X}A)^+ \\
-A\hat{X}' & 0 & 0 & 0 \\
-A\hat{X}' & 0 & 0 & 0 \\
\end{array} \right] \right) - r[\hat{X}A, Z\Sigma Z'] \\
&= i_\pm \left( \left[ \begin{array}{ccc}
-Z\Sigma Z' & -\hat{X}A & -Z\Sigma Z'(\hat{X}A)^+ \\
-A\hat{X}' & 0 & 0 & 0 \\
-A\hat{X}' & 0 & 0 & 0 \\
\end{array} \right] \right) - r[\hat{X}A, Z\Sigma Z'] \\
&= i_\pm \left( \left[ \begin{array}{ccc}
-Z\Sigma Z' & -\hat{X}A & Z\Sigma J' \\
-A\hat{X}' & 0 & 0 & 0 \\
-A\hat{X}' & 0 & 0 & 0 \\
\end{array} \right] \right) - r[\hat{X}A, Z\Sigma Z'] \\
&= i_\pm \left( \left[ \begin{array}{ccc}
-Z\Sigma Z' & -\hat{X}A & Z\Sigma J' \\
-A\hat{X}' & 0 & 0 & 0 \\
-A\hat{X}' & 0 & 0 & 0 \\
\end{array} \right] \right) - r[\hat{X}A, Z\Sigma Z'] \\
&= i_\pm \left( \left[ \begin{array}{ccc}
-Z\Sigma Z' & -\hat{X}A & Z\Sigma J' \\
-A\hat{X}' & 0 & 0 & 0 \\
-A\hat{X}' & 0 & 0 & 0 \\
\end{array} \right] \right) - r[\hat{X}A, Z\Sigma Z'] \\
&= i_\pm \left( \left[ \begin{array}{ccc}
-Z\Sigma Z' & -\hat{X}A & Z\Sigma J' \\
-A\hat{X}' & 0 & 0 & 0 \\
-A\hat{X}' & 0 & 0 & 0 \\
\end{array} \right] \right) - r[\hat{X}A, Z\Sigma Z'] \\
&= i_\pm \left( \left[ \begin{array}{ccc}
-Z\Sigma Z' & -\hat{X}A & Z\Sigma J' \\
-A\hat{X}' & 0 & 0 & 0 \\
-A\hat{X}' & 0 & 0 & 0 \\
\end{array} \right] \right) - r[\hat{X}A, Z\Sigma Z'] \\
&= i_\pm (M) + i_\pm ([\hat{X}A]^+ Z\Sigma Z'(\hat{X}A)^+) - r[\hat{X}A, Z\Sigma Z'] \\
&= i_\pm (M) + i_\pm ([\hat{X}A]^+ Z\Sigma Z'(\hat{X}A)^+) - r[\hat{X}A, Z\Sigma Z'].
\end{align*}
\]

Hence, we further obtain from (41) that

\[
\begin{align*}
i_\pm (S - D[\text{BLUP}(\phi) - \phi]) &= i_\pm (M) + r[\hat{X}A, Z\Sigma] - r[\hat{X}A, Z\Sigma] \\
&= i_\pm (M) - r(\hat{X}A), \\
i_\pm (S - D[\text{BLUP}(\phi) - \phi]) &= i_\pm (M) - r[\hat{X}A, Z\Sigma], \\
r(S - D[\text{BLUP}(\phi) - \phi]) &= r(M) - r[\hat{X}A, Z\Sigma] - r(\hat{X}A),
\end{align*}
\]
as required for (37)–(36). Applying Lemma 9 to (37)–(36) yields (a)–(e). □
Proof of Corollary 7. Setting \( S = 0 \) in (36) and simplifying by (43), we obtain

\[
\begin{align*}
    r(D[\text{BLUP}(\phi) - \phi]) &= r \begin{bmatrix}
        D\hat{y} & Cov\{\hat{y}, \phi\} & \hat{X}A \\
        Cov\{\phi, \hat{y}\} & D(\phi) & F \\
        (\hat{X}A)' & F' & 0
    \end{bmatrix} - r[\hat{X}A, Z\Sigma] - r(\hat{X}A) \\
    &= r \begin{bmatrix}
        D\hat{y} & Cov\{\hat{y}, \phi\} & \hat{X}A \\
        Cov\{\phi, \hat{y}\} & D(\phi) & F \\
        (\hat{X}A)' & F' & 0
    \end{bmatrix} - r[\hat{X}A, Z\Sigma] \\
    &= r \begin{bmatrix}
        D\hat{y} & \hat{X}A \\\n        \hat{X}A & F
    \end{bmatrix} - r(D\hat{y}, \hat{X}A),
\end{align*}
\]

thus establishing (39).

Proof of Corollary 8. Note that \( [(\hat{X}A)^{\dagger}Z\Sigma Z'(\hat{X}A)^{\dagger}]^{\dagger} \geq 0 \) and

\[
\mathcal{A}(LZ\Sigma Z'(\hat{X}A)^{\dagger} - J\Sigma Z'(\hat{X}A)^{\dagger}) \subseteq \mathcal{A}(\hat{X}A)^{\dagger}Z\Sigma Z'(\hat{X}A)^{\dagger} = \mathcal{A}(\hat{X}A)^{\dagger}Z\Sigma.
\]

Then, we obtain from (22) and (42) that

\[
\begin{align*}
    r(D(L\hat{y} - \phi) - D[\text{BLUP}(\phi) - \phi]) &= r(LZ\Sigma Z'(\hat{X}A)^{\dagger} - J\Sigma Z'(\hat{X}A)^{\dagger})[(\hat{X}A)^{\dagger}Z\Sigma Z'(\hat{X}A)^{\dagger}]^{\dagger} \\
    &= r(LZ\Sigma Z'(\hat{X}A)^{\dagger} - J\Sigma Z'(\hat{X}A)^{\dagger}) \\
    &= r \begin{bmatrix}
        LZ\Sigma - J\Sigma Z' & \hat{X}A
    \end{bmatrix} - r(\hat{X}A),
\end{align*}
\]

as required for (40).