Projektbereich B
Discussion Paper No. B–326

A BLUE Decomposition in the General Linear Regression Model

by

Hans Joachim WERNER* Bonn, Germany
and
Cemil YAPAR, Trabzon, Turkey

1995

This paper appears in LINEAR ALGEBRA AND ITS APPLICATIONS (5th Special Issue on Statistics, in press for publication in 1995).

*Financial support by Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 303 at the University of Bonn, is gratefully acknowledged.

A BLUE Decomposition in the General Linear Regression Model

Hans Joachim Werner[†]

Institute for Econometrics and Operations Research
Econometrics Unit
University of Bonn
D-53113 Bonn, Germany

and

Cemil Yapar

Karadeniz Technical University
Faculty of Arts and Sciences
Department of Statistics & Computer Science
61080 Trabzon, Turkey

ABSTRACT

In this note we consider the *general* linear regression model $(y,\ X\beta,\ V \mid R_2\beta_2 = r)$ where the block partitioned regressor matrix $X = (\,X_1 \quad X_2\,)$ may be deficient in column rank, the dispersion matrix V is possibly singular, $\beta^t = (\,\beta_1^t \quad \beta_2^t\,)$ - being partitioned according to X - is the vector of unknown regression coefficients, and β_2 is possibly subject to consistent linear constraints $R_2\beta_2 = r$. Of much interest to us is the *traditional* BLUE *(best linear unbiased estimator)* of $X\beta$. We show how this traditional BLUE can obtained from the traditional BLUE of $X_1\beta_1$ in the related model $(y,\ X_1\beta_1,\ V)$. Some properties of the dispersion matrix of the traditional BLUE of $X\beta$ are also given.

Keywords: Linear regression model, rank deficiency, linear constraints, traditional BLUE, dispersion, g-inversion.

JEL Classification: C20

[†]Financial support by Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 303 at the University of Bonn, is gratefully acknowledged.

This paper appears in LINEAR ALGEBRA AND ITS APPLICATIONS (5th Special Issue on Statistics), 1995.

A BLUE Decomposition in the General Linear Regression Model

Hans Joachim Werner[‡]

Institute for Econometrics and Operations Research
Econometrics Unit
University of Bonn
D-53113 Bonn, Germany

and

Cemil Yapar

Karadeniz Technical University
Faculty of Arts and Sciences
Department of Statistics & Computer Science
61080 Trabzon, Turkey

ABSTRACT

In this note we consider the *general* linear regression model $(y,\ X\beta,\ V \mid R_2\beta_2 = r)$ where the block partitioned regressor matrix $X = (X_1 \quad X_2)$ may be deficient in column rank, the dispersion matrix V is possibly singular, $\beta^t = (\beta_1^t \quad \beta_2^t)$ - being partitioned according to X - is the vector of unknown regression coefficients, and β_2 is possibly subject to consistent linear constraints $R_2\beta_2 = r$. Of much interest to us is the *traditional* BLUE *(best linear unbiased estimator)* of $X\beta$. We show how this traditional BLUE can obtained from the traditional BLUE of $X_1\beta_1$ in the related model $(y,\ X_1\beta_1,\ V)$. Some properties of the dispersion matrix of the traditional BLUE of $X\beta$ are also given.

---

1. INTRODUCTION AND PRELIMINARIES

Let $\mathbb{R}^n$, $\mathbb{R}^{n,m}$, and $\mathcal{P}^{n,n}$ denote the set of n-dimensional real column vectors, the set of $n \times m$ real matrices, and the set of real nonnegative definite (nnd) and symmetric $n \times n$ matrices, respectively. Given $A \in \mathbb{R}^{n,m}$,

[‡]Financial support by Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 303 at the University of Bonn, is gratefully acknowledged.

let A^t, $A^\dagger$, $\mathcal{R}(A)$, $\mathcal{N}(A)$, and $\operatorname{rank}(A)$ denote the transpose, the Moore-Penrose inverse, the range space (or column space), the null space, and the rank, respectively, of A. In addition, let A^- denote an arbitrary *g-inverse* of A satisfying $AA^-A = A$; denote the set of all g-inverses of A by $\{A^-\}$. Further, let I and 0, respectively, stand for the identity and zero matrices of whatever size is appropriate to the context.

Now let $W \in \mathcal{P}^{n,n}$ and $A \in \mathbb{R}^{n,m}$ be any given matrices. Then we have the following direct sum decomposition:

$$\mathcal{R}(A,\ W) = \mathcal{R}(A) \oplus W\mathcal{N}(A^t) \tag{1.1}$$

for the range space of the block partitioned matrix $(A,\ W)$ (cf. Lemma 3.2 in [5]). It is further convenient to denote by $\mathcal{P}(A \mid W)$ the set of all those projectors $P \in \mathbb{R}^{n,n}$ which satisfy

$$\mathcal{R}(P) = \mathcal{R}(A) \quad \text{and} \quad W\mathcal{N}(A^t) \subseteq \mathcal{N}(P). \tag{1.2}$$

That such a projector does always exist is evident from (1.1). General efficient representations for $\mathcal{P}(A \mid W)$ are given in [6]. Recall that any projector P is an *idempotent* matrix, i.e. $P^2 = P$, and that conversely every idempotent matrix P is a projector onto $\mathcal{R}(P)$ along $\mathcal{N}(P)$, also denoted by $P = P_{\mathcal{R}(P),\mathcal{N}(P)}$. If $P^2 = P$, then $(I-P)^2 = (I-P)$, $\mathcal{R}(I-P) = \mathcal{N}(P)$, and $\mathcal{N}(I-P) = \mathcal{R}(P)$.

The Gauß-Markov model

$$\mathcal{L} := (y,\ X\beta,\ V \mid R\beta = r)$$

characterizes an observable random vector y by requiring $\mathcal{E}(y) = X\beta$, $\mathcal{D}(y) = V$ ($\mathcal{E}$ and $\mathcal{D}$ denote expectation and dispersion, respectively), where the vector β of regression coefficients satisfies linear constraints $R\beta = r$, and the model matrices $X \in \mathbb{R}^{n,m}$, $V \in \mathcal{P}^{n,n}$, $R \in \mathbb{R}^{p,m}$, $r \in \mathbb{R}^p$ are fixed and known. There are no rank assumptions on X, R, V, but the constraints should be feasible, that is, $r \in \mathcal{R}(R)$. For convenience, we put $\bar{A} := A(I - R^-R)$ for any matrix A, where R^- is an arbitrary but fixed g-inverse of R. The linear estimator $Cy + c$ is defined as the *traditional* BLUE (*best linear unbiased estimator*) of $X\beta$ under the model $\mathcal{L}$ whenever

$$Cy + c \in \operatorname{argmin}_{Dy+d\in\mathcal{C}_{\mathcal{L}}}\ \mathcal{D}(Dy + d),$$

where $\mathcal{C}_{\mathcal{L}} := \{Dy + d \mid D\bar{X} = \bar{X},\ d = (I - D)XR^-r\}$ and where the minimization is with respect to the Löwner ordering. Observe that all the statistics contained in $\mathcal{C}_{\mathcal{L}}$ are linear unbiased estimators for $X\beta$ under the model $\mathcal{L}$. In passing, we should also remark that there do exist several extensions of the traditional BLUE approach. We mention here only the Rao-BLUE approach (see Rao [1]) and the more comprehensive W-BLUE approach (see Schönfeld and Werner [4]). For the sequel, it is pertinent to quote the following characterization (see, e.g., (18) in [4] or Section 3 in [3]; compare also Theorem 3.1 in [1]).

THEOREM 1.1. *Let Q be any matrix such that $\mathcal{R}(Q) = \mathcal{N}(\bar{X}^t)$. Then $\hat{\eta} := Cy + c$ is a representation for the traditional BLUE of $\eta := X\beta$ in the model $\mathcal{L}$ iff*

$$C(\bar{X},\ VQ) = (\bar{X},\ 0), \quad c = (I - C)XR^- r. \tag{1.3a, b}$$

In view of (1.3) [observe (1.1)], it is clear that the matrix C such that $Cy+(I-C)XR^- r$ is the traditional BLUE of $X\beta$ is unique iff $\operatorname{rank}(\bar{X}, V) = n$. We see immediately that this happens, in particular, when V is positive definite. But when $\operatorname{rank}(\bar{X}, V) < n$ this is not so. That the numerical value of the traditional BLUE of $X\beta$ is, however, even then unique with probability 1 is the statement of our next corollary; only for the sake of clarity, we add a simple proof of that known result (cf. [3]).

COROLLARY 1.2. *Let $\hat{\eta} = Cy + c$ and $\tilde{\eta} = Dy + d$ be any two traditional BLUEs of $X\beta$ in the model $\mathcal{L}$. Then, with probability 1, $\hat{\eta} = \tilde{\eta}$.*

Proof. Since $\hat{\eta}$ and $\tilde{\eta}$ are linear unbiased estimators for $X\beta$, clearly $\mathcal{E}(\hat{\eta} - \tilde{\eta}) = 0$. In view of (1.3a), $(C - D)(\bar{X},\ VQ) = 0$. Hence $(C - D)V = 0$ or, equivalently, $\mathcal{D}(\hat{\eta} - \tilde{\eta}) = 0$ as, in virtue of (1.1),

$$\mathcal{R}(\bar{X},\ VQ) = \mathcal{R}(\bar{X},\ V). \tag{1.4}$$

With these observations in mind, the desired result is plain. ∎

It therefore seems reasonable to write $Cy + c = \mathrm{BLUE}_{\mathcal{L}}(X\beta)$ whenever $Cy + c$ coincides almost surely with the traditional BLUE for $X\beta$ in the model $\mathcal{L}$. In what follows we always assume that $\mathcal{L}$ is of the following partitioned form:

$$\mathcal{L} := (y,\ X_1\beta_1 + X_2\beta_2,\ V \mid R_2\beta_2 = r),$$

that is, $R_1 = 0$ is assumed when $R = (R_1,\ R_2)$ is partitioned according to $X = (X_1,\ X_2)$. In addition, we assume throughout that

$$\mathcal{R}(X_2) \subseteq \mathcal{R}(X_1,\ V) \tag{1.5}$$

is satisfied. Our interest focuses in the connection between the traditional BLUE of $X\beta$ under the model $\mathcal{L}$ and the traditional BLUE of $X_1\beta_1$ under the unconstrained model

$$\mathcal{L}_1 := (y,\ X_1\beta_1,\ V).$$

To that end we introduce two further models that are strongly related to $\mathcal{L}$, namely

$$\mathcal{M} := (M_1 y,\ M_1 X_2\beta_2,\ M_1 V M_1 \mid R_2\beta_2 = r)$$

and

$$\mathcal{N} := (M_1 y,\ M_1 X_2\beta_2,\ V \mid R_2\beta_2 = r),$$

where $M_1 := I - X_1 X_1^{\dagger}$. Observe that M_1 coincides with the *orthogonal projector* (with respect to the standard inner product) onto $\mathcal{N}(X_1^t)$ along $\mathcal{R}(X_1)$; cf. Section 2 in [6]. Hence, in particular, $M_1 X_1 = 0$, and so it is clear that the model $\mathcal{M}$ can be obtained from the model $\mathcal{L}$ by premultiplying y by M_1. Whereas $\mathcal{M}$ thus turns out to be a *correctly transformed* model, the dispersion in the so-called *reduced* model $\mathcal{N}$ is defined to be V (cf. [7]). It is pertinent to mention here that as a consequence of (1.5) we have

$$y \in \mathcal{R}(X_1,\ V) \quad \text{with probability } 1, \tag{1.6}$$

irrespective of under which of the different models y is observed.

2. BLUE DECOMPOSITION

We begin this section with deriving representations for the traditional BLUEs of the mean vectors in the respective linear models introduced in Section 1. Notice that for each $R_2^- \in \{R_2^-\}$ we have

$$\begin{pmatrix} 0 \\ R_2^- \end{pmatrix} \in \{R^-\}$$

whenever $R = (\,0 \quad R_2\,)$. Below, the g-inverse R^- is always tacitly chosen in this manner. Therefore

$$I - R^- R = \begin{pmatrix} I & 0 \\ 0 & I - R_2^- R_2 \end{pmatrix},$$

and so

$$\bar{X} = (\,X_1 \quad \tilde{X}_2\,)$$

with $\tilde{X}_2 := X_2(I - R_2^- R_2)$.

THEOREM 2.1. *Consider $\mathcal{L}$, $\mathcal{L}_1$, $\mathcal{M}$, and $\mathcal{N}$, where, as before, $M_1 = I - X_1X_1^\dagger$. Let $P_\mathcal{L}$, $P_{\mathcal{L}_1}$, $P_\mathcal{M}$, and $P_\mathcal{N}$ denote arbitrary but fixed projectors from $\mathcal{P}(\bar{X} \mid V)$, $\mathcal{P}(X_1 \mid V)$, $\mathcal{P}(M_1\tilde{X}_2 \mid M_1VM_1)$, and $\mathcal{P}(M_1\tilde{X}_2 \mid V)$, respectively. With probability 1 we then have:*

(i) $P_\mathcal{L}y + (I - P_\mathcal{L})X_2R_2^- r = \mathrm{BLUE}_\mathcal{L}(X\beta)$,
(ii) $P_{\mathcal{L}_1}y = \mathrm{BLUE}_{\mathcal{L}_1}(X_1\beta_1)$,
(iii) $P_\mathcal{M}M_1y + (I - P_\mathcal{M})M_1X_2R_2^- r = \mathrm{BLUE}_\mathcal{M}(M_1X_2\beta_2)$,
(iv) $P_\mathcal{N}M_1y + (I - P_\mathcal{N})M_1X_2R_2^- r = \mathrm{BLUE}_\mathcal{N}(M_1X_2\beta_2)$.

Proof. First, let $P_\mathcal{L} \in \mathcal{P}(\bar{X} \mid V)$, and let Q be as in Theorem 1.1. Then, by definition of $\mathcal{P}(\bar{X} \mid V)$,

$$P_\mathcal{L}\begin{pmatrix} \bar{X} & VQ \end{pmatrix} = \begin{pmatrix} \bar{X} & 0 \end{pmatrix},$$

and it now follows from Theorem 1.1 that $P_\mathcal{L}y+(I-P_\mathcal{L})X_2R_2^- r$ represents, with probability 1, the BLUE for $X\beta$ in model $\mathcal{L}$. With obvious modifications, (ii), (iii), and (iv) can be obtained on similar lines by applying Theorem 1.1 to $\mathcal{L}_1$, $\mathcal{M}$, and $\mathcal{N}$, respectively, in lieu of $\mathcal{L}$. ∎

We proceed with proving an interesting relationship between the *generally oblique* projectors considered in the previous theorem.

THEOREM 2.2. *Let M_1, $\bar{X}$, $\tilde{X}_2$, $\mathcal{P}(\bar{X} \mid V)$, $\mathcal{P}(X_1 \mid V)$, and $\mathcal{P}(M_1\tilde{X}_2 \mid M_1VM_1)$ be all defined as before. For each $P_{\mathcal{L}_1} \in \mathcal{P}(X_1 \mid V)$ and for each $P_\mathcal{M} \in \mathcal{P}(M_1\tilde{X}_2 \mid M_1VM_1)$ we then have*

$$P_{\mathcal{L}_1} + (I - P_{\mathcal{L}_1})P_\mathcal{M}M_1 \in \mathcal{P}(\bar{X} \mid V). \tag{2.1}$$

Proof. Let $P_{\mathcal{L}_1} \in \mathcal{P}(X_1 \mid V)$ and $P_\mathcal{M} \in \mathcal{P}(M_1\tilde{X}_2 \mid M_1VM_1)$ be arbitrary but fixed. Then, by definition of $\mathcal{P}(X_1 \mid V)$, we have $\mathcal{R}(P_{\mathcal{L}_1}) = \mathcal{R}(X_1)$, $V\mathcal{N}(X_1^t) \subseteq \mathcal{N}(P_{\mathcal{L}_1})$ and $P_1^2 = P_1$. Since $P_\mathcal{M} \in \mathcal{P}(M_1\tilde{X}_2 \mid M_1VM_1)$ and since $M_1VM_1\mathcal{N}(\tilde{X}_2^tM_1) = M_1VM_1\left[\mathcal{R}(X_1) \oplus \mathcal{N}(\bar{X}^t)\right] = M_1V\mathcal{N}(\bar{X}^t)$, we further have $\mathcal{R}(P_\mathcal{M}) = \mathcal{R}(M_1\tilde{X}_2)$, $M_1V\mathcal{N}(\bar{X}^t) \subseteq \mathcal{N}(P_\mathcal{M})$, and $P_\mathcal{M}^2 = P_\mathcal{M}$. Now put

$$P := P_{\mathcal{L}_1} + (I - P_{\mathcal{L}_1})P_\mathcal{M}M_1. \tag{2.2}$$

For establishing $P \in \mathcal{P}(\bar{X} \mid V)$ it suffices to show that the conditions $\mathcal{R}(P) = \mathcal{R}(\bar{X})$, $V\mathcal{N}(\bar{X}^t) \subseteq \mathcal{N}(P)$, and $P^2 = P$ all hold true. Notice that, according to our convention, $\bar{X} = \begin{pmatrix} X_1 & \tilde{X}_2 \end{pmatrix}$. Since $P_{\mathcal{L}_1}X_1 = X_1$ and $M_1X_1 = 0$, clearly $PX_1 = X_1$. As $P_\mathcal{M}M_1\tilde{X}_2 = M_1\tilde{X}_2$ and $(I - P_{\mathcal{L}_1})M_1 = I - P_{\mathcal{L}_1}$, also $P\tilde{X}_2 = P_{\mathcal{L}_1}\tilde{X}_2 + (I - P_{\mathcal{L}_1})P_\mathcal{M}M_1\tilde{X}_2 = P_{\mathcal{L}_1}\tilde{X}_2 + (I - P_{\mathcal{L}_1})\tilde{X}_2 = \tilde{X}_2$. Hence $P\bar{X} = \bar{X}$, and so $\mathcal{R}(\bar{X}) \subseteq \mathcal{R}(P)$. The converse inclusion is seen as follows. Of course, $\mathcal{R}(P) \subseteq \mathcal{R}(P_{\mathcal{L}_1}) + (I - P_{\mathcal{L}_1})\mathcal{R}(P_\mathcal{M}M_1) \subseteq$

$\mathcal{R}(P_{\mathcal{L}_1}) + (I - P_{\mathcal{L}_1})\mathcal{R}(P_{\mathcal{M}})$. But $\mathcal{R}(P_{\mathcal{M}}) = \mathcal{R}(M_1\tilde{X}_2) = M_1\mathcal{R}(\tilde{X}_2)$ and $(I - P_{\mathcal{L}_1})M_1 = I - P_{\mathcal{L}_1}$. Therefore $\mathcal{R}(P) \subseteq \mathcal{R}(P_{\mathcal{L}_1}) + \mathcal{R}(\tilde{X}_2) = \mathcal{R}(X_1) + \mathcal{R}(\tilde{X}_2) = \mathcal{R}(X_1, \tilde{X}_2) = \mathcal{R}(\bar{X})$, that is, $\mathcal{R}(P) \subseteq \mathcal{R}(\bar{X})$. As desired, we therefore have $\mathcal{R}(P) = \mathcal{R}(\bar{X})$. By exploiting $P_{\mathcal{M}}M_1V\mathcal{N}(\bar{X}^t) = \{0\}$, $V\mathcal{N}(\bar{X}^t) \subseteq V\mathcal{N}(X_1^t)$, and $P_{\mathcal{L}_1}V\mathcal{N}(X_1^t) = \{0\}$, we further get $PV\mathcal{N}(\bar{X}^t) = \{0\}$, that is, $V\mathcal{N}(\bar{X}^t) \subseteq \mathcal{N}(P)$. Because of $P_{\mathcal{L}_1}(I - P_{\mathcal{L}_1}) = 0$, $M_1P_{\mathcal{L}_1} = 0$, $M_1(I - P_{\mathcal{L}_1}) = M_1$, $M_1P_{\mathcal{M}} = P_{\mathcal{M}}$, and $P_{\mathcal{M}}^2 = P_{\mathcal{M}}$, we also obtain $P^2 = [P_{\mathcal{L}_1} + (I - P_{\mathcal{L}_1})P_{\mathcal{M}}M_1]^2 = P_{\mathcal{L}_1}^2 + (I - P_{\mathcal{L}_1})P_{\mathcal{M}}M_1(I - P_{\mathcal{L}_1})P_{\mathcal{M}}M_1 = P_{\mathcal{L}_1} + (I - P_{\mathcal{L}_1})P_{\mathcal{M}}M_1P_{\mathcal{M}}M_1 = P_{\mathcal{L}_1} + (I - P_{\mathcal{L}_1})(P_{\mathcal{M}})^2 M_1 = P_{\mathcal{L}_1} + (I - P_{\mathcal{L}_1})P_{\mathcal{M}}M_1 = P$, and the proof is complete. ■

THEOREM 2.3. *Let $\mathcal{L}$, $\mathcal{L}_1$, $\mathcal{M}$, and $\mathcal{N}$ be defined as before. In addition, let $P_{\mathcal{L}_1}$ be an arbitrary but fixed projector from $\mathcal{P}(X_1 \mid V)$. The traditional BLUEs of $X\beta$, $X_1\beta_1$, and $M_1X_2\beta_2$ in $\mathcal{L}$, $\mathcal{L}_1$, and $\mathcal{M}$, respectively, are then, with probability 1, related as follows:*

$$\mathrm{BLUE}_{\mathcal{L}_1}(X_1\beta_1) + (I - P_{\mathcal{L}_1}) \cdot \mathrm{BLUE}_{\mathcal{M}}(M_1X_2\beta_2) = \mathrm{BLUE}_{\mathcal{L}}(X\beta). \quad (2.3)$$

Moreover, if $V\mathcal{N}(X_1^t) \subseteq \mathcal{N}(X_1^t)$, then

$$\mathrm{BLUE}_{\mathcal{L}_1}(X_1\beta_1) + (I - P_{\mathcal{L}_1}) \cdot \mathrm{BLUE}_{\mathcal{N}}(M_1X_2\beta_2) = \mathrm{BLUE}_{\mathcal{L}}(X\beta) \quad (2.4)$$

with probability 1.

Proof. Let $P_{\mathcal{L}_1}$ be an arbitrary but fixed projector from $\mathcal{P}(X_1 \mid V)$. Then, by Theorem 2.1(ii), $P_{\mathcal{L}_1}y = \mathrm{BLUE}_{\mathcal{L}_1}(X_1\beta_1)$ with probability 1. By Theorem 2.1(ii), for each $P_{\mathcal{M}} \in \mathcal{P}(M_1\tilde{X}_2 \mid M_1VM_1)$, likewise $P_{\mathcal{M}}M_1y + (I - P_{\mathcal{M}})M_1X_2R_2^-r = \mathrm{BLUE}_{\mathcal{M}}(M_1X_2\beta_2)$ with probability 1. Now choose any $P_{\mathcal{M}} \in \mathcal{P}(M_1\tilde{X}_2 \mid M_1VM_1)$ and put $P_{\mathcal{L}} := P_{\mathcal{L}_1} + (I - P_{\mathcal{L}_1})P_{\mathcal{M}}M_1$. Then, in view of (2.1), $P_{\mathcal{L}} \in \mathcal{P}(\bar{X} \mid V)$. According to Theorem 2.1(i), therefore $P_{\mathcal{L}}y + (I - P_{\mathcal{L}})X_2R_2^-r = \mathrm{BLUE}_{\mathcal{L}}(X\beta)$. Next observe that $(I - P_{\mathcal{L}_1})M_1 = I - P_{\mathcal{L}_1}$. Combining all these observations now results in

$$\begin{aligned}
& \mathrm{BLUE}_{\mathcal{L}_1}(X_1\beta_1) + (I - P_{\mathcal{L}_1}) \cdot \mathrm{BLUE}_{\mathcal{M}}(M_1X_2\beta_2) \\
= \; & P_{\mathcal{L}_1}y + (I - P_{\mathcal{L}_1})\left[P_{\mathcal{M}}M_1y + (I - P_{\mathcal{M}})M_1X_2R_2^-r\right] \\
= \; & [P_{\mathcal{L}_1} + (I - P_{\mathcal{L}_1})P_{\mathcal{M}}M_1]\, y + (I - P_{\mathcal{L}_1})(I - P_{\mathcal{M}})M_1X_2R_2^-r \\
= \; & [P_{\mathcal{L}_1} + (I - P_{\mathcal{L}_1})P_{\mathcal{M}}M_1]\, y + (I - P_{\mathcal{L}_1})(I - P_{\mathcal{M}}M_1)X_2R_2^-r \\
= \; & P_{\mathcal{L}}y + (I - P_{\mathcal{L}})X_2R_2^-r \\
= \; & \mathrm{BLUE}_{\mathcal{L}}(X\beta) \quad \text{with probability 1,}
\end{aligned}$$

and the proof of (2.3) is complete. Next let $V\mathcal{N}(X_1^t) \subseteq \mathcal{N}(X_1^t)$. Then $M_1VM_1 = VM_1$, since $\mathcal{R}(M_1) = \mathcal{N}(X_1^t)$. Moreover, as in the proof of

Theorem 2.2, $VM_1\mathcal{N}(\tilde{X}_2^t M_1) = V\mathcal{N}(\bar{X}^t)$. Therefore $M_1 V M_1 \mathcal{N}(\tilde{X}_2^t M_1) = V\mathcal{N}(\bar{X}^t)$. For each $P_{\mathcal{N}} \in \mathcal{P}(M_1\tilde{X}_2 \mid V)$ we have, by definition of $\mathcal{P}(M_1\tilde{X}_2 \mid V)$, that $V\mathcal{N}(\tilde{X}_2^t M_1) = V\left[\mathcal{R}(X_1) \oplus \mathcal{N}(\bar{X}^t)\right] = V\mathcal{R}(X_1) + V\mathcal{N}(\bar{X}^t) \subseteq \mathcal{N}(P_{\mathcal{N}})$. So, in particular, $V\mathcal{N}(\bar{X}^t) \subseteq \mathcal{N}(P_{\mathcal{N}})$. As $M_1 V M_1 \mathcal{N}(\tilde{X}_2^t M_1) = V\mathcal{N}(\bar{X}^t)$, therefore $\mathcal{P}(M_1\tilde{X}_2 \mid V) \subseteq \mathcal{P}(M_1\tilde{X}_2 \mid M_1 V M_1)$, and (2.4) follows from (2.3), Theorem 2.1(iii), and Theorem 2.1(iv). ∎

Theorem 2.3 needs some further discussion. First observe that (2.3) along with Theorem 2.1 tells us how the traditional BLUE of $X\beta$ in the model $\mathcal{L}$ can be obtained from the traditional BLUE of $X_1\beta_1$ in the model $\mathcal{L}_1$. Relation (2.3) can thus serve as an *updating formula* when additional regressors are to be taken into account. Of course, from the practical point of view and in the light of Theorem 2.1, a natural question then is: How can we explicitly compute a member of a set like $\mathcal{P}(A \mid W)$ in terms of A and W? It is therefore pertinent to mention here that a satisfactory answer to this question is given in ([6], Section 2). Next recall that the inclusion $V\mathcal{N}(X_1^t) \subseteq \mathcal{N}(X_1^t)$ holds iff $\mathrm{OLSE}_{\mathcal{L}_1}(X_1\beta_1) = \mathrm{BLUE}_{\mathcal{L}_1}(X_1\beta_1)$ (with probability 1), that is, iff the OLS method (*ordinary least squares method*) gives the traditional BLUE of $X_1\beta_1$ in the model $\mathcal{L}_1$; cf. [2].

The BLUE decomposition (2.3) also enables us to derive the following interesting properties of the dispersion matrix of the traditional BLUE of $X\beta$ under the model $\mathcal{L}$.

THEOREM 2.4. *Let $\mathcal{L}$, $\mathcal{L}_1$, $\mathcal{M}$, M_1, and $\tilde{X}_2$ be defined as before. Moreover, let $P_{\mathcal{L}_1}$ be an arbitrary but fixed projector from $\mathcal{P}(X_1 \mid V)$. In the model $\mathcal{L}$, then*

$$\begin{aligned}\mathcal{D}(\mathrm{BLUE}_{\mathcal{L}}(X\beta)) = \mathcal{D}(\mathrm{BLUE}_{\mathcal{L}_1}(X_1\beta_1)) + (I - P_{\mathcal{L}_1})\cdot \\ \cdot\mathcal{D}(\mathrm{BLUE}_{\mathcal{M}}(M_1X_2\beta_2)) \cdot (I - P_{\mathcal{L}_1})^t.\end{aligned} \tag{2.5}$$

Hence,

$$\mathcal{D}(\mathrm{BLUE}_{\mathcal{L}}(X\beta)) = \mathcal{D}(\mathrm{BLUE}_{\mathcal{L}_1}(X_1\beta_1)) \quad \text{iff} \quad \mathcal{R}(\tilde{X}_2) \subseteq \mathcal{R}(X_1) \tag{2.6}$$

or, equivalently, iff $M_1\tilde{X}_2 = 0$.

Proof. First recall that under the model $\mathcal{L}$ we have $\mathcal{D}(Ay + a) = AVA^t$ for each statistic $Ay + a$ with $A \in \mathbb{R}^{k,n}$ and $a \in \mathbb{R}^k$ both fixed and known. Next observe that $P_{\mathcal{L}_1} V M_1 = 0$ holds for any choice of $P_{\mathcal{L}_1} \in \mathcal{P}(X_1 \mid V)$. With these observations as well as Theorem 2.1 in mind, it is now not difficult to obtain (2.5) from (2.3); details are left to the reader. In view of

(2.5) and Theorem 2.1(iii), clearly $\mathcal{D}(\text{BLUE}_{\mathcal{L}}(X\beta)) = \mathcal{D}(\text{BLUE}_{\mathcal{L}_1}(X_1\beta_1))$ iff

$$(I - P_{\mathcal{L}_1})P_{\mathcal{M}}M_1V = 0. \tag{2.7}$$

So it suffices to show that (2.7) is equivalent to $\mathcal{R}(\tilde{X}_2) \subseteq \mathcal{R}(X_1)$. As a consequence of (1.5), $\mathcal{R}(\tilde{X}_2) \subseteq \mathcal{R}(X_1,\ V)$. Since $\mathcal{R}(X_1,\ V) = \mathcal{R}(X_1) \oplus \mathcal{R}(VM_1)$ [see (1.1)], therefore $\tilde{X}_2 = X_1U + VM_1T$ for some matrices U and T; notice that X_1U and VM_1T are uniquely determined. Then $M_1\tilde{X}_2 = M_1VM_1T$, which in turn implies $\mathcal{R}(M_1\tilde{X}_2) \subseteq \mathcal{R}(M_1V)$ on account of $\mathcal{R}(M_1VM_1) = \mathcal{R}(M_1V)$. Consequently $\mathcal{R}(P_{\mathcal{M}}M_1V) = \mathcal{R}(M_1\tilde{X}_2)$, as $\mathcal{R}(P_{\mathcal{M}}) = \mathcal{R}(M_1\tilde{X}_2)$. But now it is clear that (2.7) is equivalent to

$$(I - P_{\mathcal{L}_1})M_1\tilde{X}_2 = 0. \tag{2.8}$$

Since $(I - P_{\mathcal{L}_1})M_1 = I - P_{\mathcal{L}_1}$ and $\mathcal{N}(I - P_{\mathcal{L}_1}) = \mathcal{R}(X_1)$, Equation (2.8) holds iff $\mathcal{R}(\tilde{X}_2) \subseteq \mathcal{R}(X_1)$ or, equivalently, iff $M_1\tilde{X}_2 = 0$. This completes the proof. ■

For a moment, let us now consider a situation where there is uncertainty about the inclusion of some regressors in the model and where the *true* model may be assumed to be given by $\mathcal{L}_1$ or $\mathcal{L}$. The relation (2.5) then represents the (intuitively clear) property that under (1.5) the dispersion of the BLUE of $\mathcal{E}(y)$ is inflated when $\mathcal{L}$ is chosen, although $\mathcal{L}_1$ is the *true* model. In other words, the dispersion of the BLUE of the expectation vector always either remains unchanged or increases when additional regressors are taken into consideration.

In view of Theorem 2.4, one could be tempted to believe that in model $\mathcal{L}$ we always have $\text{BLUE}_{\mathcal{L}_1}(X_1\beta_1) = \text{BLUE}_{\mathcal{L}}(X\beta)$ with probability 1 whenever (2.6) is satisfied. This, however, is generally not true, because, even under (2.6), the statistic $\text{BLUE}_{\mathcal{L}_1}(X_1\beta_1)$ can fail to be an unbiased estimator for $X\beta$ in the model $\mathcal{L}$. As can be easily checked, $\text{BLUE}_{\mathcal{L}_1}(X_1\beta_1)$ is an unbiased estimator for $X\beta$ in model $\mathcal{L}$ iff

$$\mathcal{R}(\tilde{X}_2) \subseteq \mathcal{R}(X_1) \quad \text{and} \quad X_2R_2^-r \in \mathcal{R}(X_1) \tag{2.9a, b}$$

where R_2^- denotes, as before, an arbitrary but fixed g-inverse of R_2. Needless to say, the conditions (2.9) are hence also necessary and sufficient for

$$\text{BLUE}_{\mathcal{L}_1}(X_1\beta_1) = \text{BLUE}_{\mathcal{L}}(X\beta)$$

to hold with probability 1 in the model $\mathcal{L}$. The following theorem should now be obvious.

THEOREM 2.5. *Let $\mathcal{L}$, $\mathcal{L}_1$, and M_1 be defined as before. In addition, let R_2^- denote an arbitrary but fixed g-inverse of R_2. The following statements are then equivalent:*

(i) $\mathrm{BLUE}_{\mathcal{L}_1}(X_1\beta_1) = \mathrm{BLUE}_{\mathcal{L}}(X\beta)$ *with probability 1;*
(ii) $\mathrm{BLUE}_{\mathcal{L}_1}(X_1\beta_1)$ *is an unbiased estimator for $X\beta$ in the model $\mathcal{L}$;*
(iii) $\mathcal{D}(\mathrm{BLUE}_{\mathcal{L}_1}(X_1\beta_1)) = \mathcal{D}(\mathrm{BLUE}_{\mathcal{L}}(X\beta))$ *and* $M_1X_2R_2^-r = 0$*;*
(iv) $M_1\tilde{X}_2 = 0$ *and* $M_1X_2R_2^-r = 0$*.*

In connection with (2.9) and Theorem 2.5 the following concluding remark is worth making. Let $\mathrm{BLUE}_{\mathcal{L}_1}(X_1\beta_1)$ be an unbiased estimator for $X\beta$ in the model $\mathcal{L}$, that is, let (2.9) hold. It should be stressed that even then $X_1\beta_1$ need not necessarily be linearly unbiasedly estimable in the model $\mathcal{L}$, although in the model $\mathcal{L}_1$ it trivially is. In order to see this, consider the *unconstrained* model $\mathcal{L}$, that is, let $R_2 = 0$ as well as $r = 0$. It is then not difficult to check that $X_1\beta_1$ is linearly unbiasedly estimable in the *unconstrained* model $\mathcal{L}$ iff $\mathcal{R}(X_1^t) \subseteq X_1^t\mathcal{N}(X_2^t)$ or, equivalently, iff

$$\mathcal{R}(X_1) \cap \mathcal{R}(X_2) = \{0\}. \tag{2.10}$$

But when $\mathcal{L}$ is unconstrained, the condition (2.9b) is redundant, whereas condition (2.9a) reduces to $\mathcal{R}(X_2) \subseteq \mathcal{R}(X_1)$. Equation (2.10) thus holds iff $X_2 = 0$, so that our claim becomes obvious.

REFERENCES

1 C. R. Rao, Representations of the best linear unbiased estimators in the Gauß-Markoff model with a singular dispersion matrix, *J. Multivariate Anal.* 3:276–292 (1973).
2 S. Puntanen and G. P. H. Styan, The equality of the ordinary least squares estimator and the best linear unbiased estimator, *Amer. Statist.* 43:153–161 (1989).
3 P. Schönfeld and H. J. Werner, Beiträge zur Theorie und Anwendung linearer Modelle, in *Ökonomische Prognose-, Entscheidungs- und Gleichgewichtsmodelle* (W. Krelle, Ed.), VCH Verlagsgesellschaft, Weinheim, 1986, pp. 251–262.
4 P. Schönfeld and H. J. Werner, A note on C.R. Rao's wider definition BLUE in the general Gauß-Markov model, *Sankhyā Ser. B* 49:1–8 (1987).
5 H. J. Werner, More on BLU estimation in regression models with possibly singular covariances, *Linear Algebra Appl.* 67:207–214 (1985).
6 H. J. Werner and C. Yapar, On inequality constrained generalized least squares selections in the general possibly singular Gauß-Markov model: A projector theoretical approach, *Linear Algebra Appl.* (1995, to appear; Fifth Special Issue devoted to Linear Algebra and Statistics).

7 H. J. Werner and C. Yapar, More on partitioned possibly restricted linear regression, in *Proceedings of the 5th Tartu Conference on Multivariate Statistics* (Ene-Margit Tiit et al., Eds.), VSP Press, Utrecht, in press for publication in 1995.