Best Linear Unbiased Estimation in Linear Models

Simo Puntanen*
University of Tampere, Finland

George P. H. Styan†
McGill University, Montréal, Canada

November 4, 2015

Keywords and Phrases: Best linear unbiased, BLUE, BLUP, Gauss–Markov Theorem, Generalized inverse, Ordinary least squares, OLSE.

1 Introduction

In this article we consider the general linear model (Gauss–Markov model)

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \text{or in short } \mathscr{M} = \{\mathbf{y},\, \mathbf{X}\boldsymbol{\beta},\, \sigma^2\mathbf{V}\},$$

where $\mathbf{X}$ is a known $n \times p$ model matrix, the vector $\mathbf{y}$ is an observable n-dimensional random vector, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters, and $\boldsymbol{\varepsilon}$ is an unobservable vector of random errors with expectation $\mathrm{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$, and covariance matrix $\mathrm{cov}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{V}$, where $\sigma^2 > 0$ is an unknown constant. The nonnegative definite (possibly singular) matrix $\mathbf{V}$ is known. In our considerations σ^2 has no role and hence we may put $\sigma^2 = 1$.

As regards the notation, we will use the symbols $\mathbf{A}'$, $\mathbf{A}^-$, $\mathbf{A}^+$, $\mathscr{C}(\mathbf{A})$, $\mathscr{C}(\mathbf{A})^\perp$, and $\mathscr{N}(\mathbf{A})$ to denote, respectively, the transpose, a generalized inverse, the Moore–Penrose inverse, the column space, the orthogonal complement of the column space, and the null space, of the matrix $\mathbf{A}$. By $(\mathbf{A} : \mathbf{B})$ we denote the partitioned matrix with $\mathbf{A}$ and $\mathbf{B}$ as submatrices. By $\mathbf{A}^\perp$ we denote any matrix satisfying $\mathscr{C}(\mathbf{A}^\perp) = \mathscr{N}(\mathbf{A}') = \mathscr{C}(\mathbf{A})^\perp$. Furthermore, we will write $\mathbf{P}_\mathbf{A} = \mathbf{A}\mathbf{A}^+ = \mathbf{A}(\mathbf{A}'\mathbf{A})^-\mathbf{A}'$ to denote the orthogonal projector (with respect to the standard inner product) onto $\mathscr{C}(\mathbf{A})$. In particular, we denote $\mathbf{H} = \mathbf{P}_\mathbf{X}$ and $\mathbf{M} = \mathbf{I}_n - \mathbf{H}$. One choice for $\mathbf{X}^\perp$ is of course the projector $\mathbf{M}$.

*Department of Mathematics and Statistics, FI-33014 University of Tampere, Tampere, Finland. Email: `simo.puntanen@uta.fi`

†Department of Mathematics and Statistics, McGill University, 805 ouest rue Sherbrooke Street West, Montréal (Québec), Canada H3A 2K6. Email: `styan@math.mcgill.ca`

Let $\mathbf{K}'\boldsymbol{\beta}$ be a given vector of parametric functions specified by $\mathbf{K}' \in \mathbb{R}^{q\times p}$. Our object is to find a (homogeneous) linear estimator $\mathbf{Ay}$ which would provide an unbiased and in some sense "best" estimator for $\mathbf{K}'\boldsymbol{\beta}$ under the model $\mathscr{M}$. However, not all parametric functions have linear unbiased estimators; those which have are called *estimable* parametric functions, and then there exists a matrix $\mathbf{A}$ such that

$$\mathrm{E}(\mathbf{Ay}) = \mathbf{AX}\boldsymbol{\beta} = \mathbf{K}'\boldsymbol{\beta} \quad \text{for all } \boldsymbol{\beta} \in \mathbb{R}^p.$$

Hence $\mathbf{K}'\boldsymbol{\beta}$ is estimable if and only if there exists a matrix $\mathbf{A}$ such that $\mathbf{K}' = \mathbf{AX}$, i.e., $\mathscr{C}(\mathbf{K}) \subset \mathscr{C}(\mathbf{X}')$.

The ordinary least squares estimator of $\mathbf{K}'\boldsymbol{\beta}$ is defined as $\mathrm{OLSE}(\mathbf{K}'\boldsymbol{\beta}) = \mathbf{K}'\hat{\boldsymbol{\beta}}$, where $\hat{\boldsymbol{\beta}}$ is any solution to the normal equation $\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$; hence $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$ minimizes $(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$ and it can be expressed as $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y}$, while $\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{Hy}$. Now the condition $\mathscr{C}(\mathbf{K}) \subset \mathscr{C}(\mathbf{X}')$ guarantees that $\mathbf{K}'\hat{\boldsymbol{\beta}}$ is unique, even though $\hat{\boldsymbol{\beta}}$ may not be unique.

2 The Best Linear Unbiased Estimator (BLUE)

The expectation $\mathbf{X}\boldsymbol{\beta}$ is trivially estimable and $\mathbf{Gy}$ is unbiased for $\mathbf{X}\boldsymbol{\beta}$ whenever $\mathbf{GX} = \mathbf{X}$. An unbiased linear estimator $\mathbf{Gy}$ for $\mathbf{X}\boldsymbol{\beta}$ is defined to be the *best* linear unbiased estimator, BLUE, for $\mathbf{X}\boldsymbol{\beta}$ under $\mathscr{M}$ if

$$\mathrm{cov}(\mathbf{Gy}) \leq_{\mathrm{L}} \mathrm{cov}(\mathbf{Ly}) \quad \text{for all } \mathbf{L}\colon \mathbf{LX} = \mathbf{X},$$

where "$\leq_{\mathrm{L}}$" refers to the Löwner partial ordering. In other words, $\mathbf{Gy}$ has the smallest covariance matrix (in the Löwner sense) among all linear unbiased estimators. We denote the BLUE of $\mathbf{X}\boldsymbol{\beta}$ as $\mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta}) = \mathbf{X}\tilde{\boldsymbol{\beta}}$. If $\mathbf{X}$ has full column rank, then $\boldsymbol{\beta}$ is estimable and an unbiased estimator $\mathbf{Ay}$ is the BLUE for $\boldsymbol{\beta}$ if $\mathbf{AVA}' \leq_{\mathrm{L}} \mathbf{BVB}'$ for all $\mathbf{B}$ such that $\mathbf{BX} = \mathbf{I}_p$. The Löwner ordering is a very strong ordering implying for example

$$\mathrm{var}(\tilde{\beta}_i) \leq \mathrm{var}(\beta_i^*)\,, \quad i = 1, \dots, p,$$

$$\mathrm{trace}[\mathrm{cov}(\tilde{\boldsymbol{\beta}})] \leq \mathrm{trace}[\mathrm{cov}(\boldsymbol{\beta}^*)], \qquad \det[\mathrm{cov}(\tilde{\boldsymbol{\beta}})] \leq \det[\mathrm{cov}(\boldsymbol{\beta}^*)],$$

for any linear unbiased estimator $\boldsymbol{\beta}^*$ of $\boldsymbol{\beta}$; here var refers to the variance and "det" denotes the determinant.

The following theorem gives the "Fundamental BLUE equation"; see, e.g., Rao (1967), Zyskind (1967) and Puntanen, Styan and Werner (2000).

Theorem 1. *Consider the general linear model* $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$. *Then the estimator* $\mathbf{Gy}$ *is the* BLUE *for* $\mathbf{X}\boldsymbol{\beta}$ *if and only if* $\mathbf{G}$ *satisfies the equation*

$$\mathbf{G}(\mathbf{X} : \mathbf{VX}^{\perp}) = (\mathbf{X} : \mathbf{0}). \tag{1}$$

The corresponding condition for $\mathbf{Ay}$ *to be the* BLUE *of an estimable parametric function* $\mathbf{K}'\boldsymbol{\beta}$ *is* $\mathbf{A}(\mathbf{X} : \mathbf{VX}^{\perp}) = (\mathbf{K}' : \mathbf{0})$.

It is sometimes convenient to express (1) in the following form, see Rao (1971).

Theorem 2 (Pandora's Box)**.** *Consider the general linear model* $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$. *Then the estimator* $\mathbf{Gy}$ *is the* BLUE *for* $\mathbf{X}\boldsymbol{\beta}$ *if and only if there exists a matrix* $\mathbf{L} \in \mathbb{R}^{p\times n}$ *so that* $\mathbf{G}$ *is a solution to*

$$\begin{pmatrix} \mathbf{V} & \mathbf{X} \\ \mathbf{X}' & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{G}' \\ \mathbf{L} \end{pmatrix} = \begin{pmatrix} \mathbf{0} \\ \mathbf{X}' \end{pmatrix}.$$

The equation (1) has a unique solution for $\mathbf{G}$ if and only if $\mathscr{C}(\mathbf{X} : \mathbf{V}) = \mathbb{R}^n$. Notice that under $\mathscr{M}$ we assume that the observed value of $\mathbf{y}$ belongs to the subspace $\mathscr{C}(\mathbf{X} : \mathbf{V})$ with probability 1; this is the consistency condition of the linear model, see, e.g., Baksalary, Rao and Markiewicz (1992). The consistency condition means, for example, that whenever we have some statements which involve the random vector $\mathbf{y}$, these statements need hold only for those values of $\mathbf{y}$ that belong to $\mathscr{C}(\mathbf{X} : \mathbf{V})$. The general solution for $\mathbf{G}$ can be expressed, for example, in the following ways:

$$\mathbf{G}_1 = \mathbf{X}(\mathbf{X}'\mathbf{W}^-\mathbf{X})^-\mathbf{X}'\mathbf{W}^- + \mathbf{F}_1(\mathbf{I}_n - \mathbf{W}\mathbf{W}^-),$$

$$\mathbf{G}_2 = \mathbf{H} - \mathbf{HVM}(\mathbf{MVM})^-\mathbf{M} + \mathbf{F}_2[\mathbf{I}_n - \mathbf{MVM}(\mathbf{MVM})^-]\mathbf{M},$$

where $\mathbf{F}_1$ and $\mathbf{F}_2$ are arbitrary matrices, $\mathbf{W} = \mathbf{V} + \mathbf{XUX}'$ and $\mathbf{U}$ is any arbitrary conformable matrix such that $\mathscr{C}(\mathbf{W}) = \mathscr{C}(\mathbf{X} : \mathbf{V})$. Notice that even though $\mathbf{G}$ may not be unique, the numerical value of $\mathbf{Gy}$ is unique because $\mathbf{y} \in \mathscr{C}(\mathbf{X} : \mathbf{V})$. If $\mathbf{V}$ is positive definite, then $\mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta}) = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^-\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$. Clearly $\mathrm{OLSE}(\mathbf{X}\boldsymbol{\beta}) = \mathbf{Hy}$ is the BLUE under $\{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}\}$. It is also worth noting that the matrix $\mathbf{G}$ satisfying (1) can be interpreted as a projector: it is a projector onto $\mathscr{C}(\mathbf{X})$ along $\mathscr{C}(\mathbf{VX}^\perp)$, see Rao (1974).

3 OLSE vs. BLUE

Characterizing the equality of the Ordinary Least Squares Estimator (OLSE) and the BLUE has received a lot of attention in the literature, since Anderson (1948), but the major breakthroughs were made by Rao (1967) and Zyskind (1967); for a detailed review, see Puntanen and Styan (1989). For some further references from those years we may mention Kruskal (1968), Watson (1967), and Zyskind and Martin (1969).

We present below six characterizations for the OLSE and the BLUE to be equal (with probability 1).

Theorem 3 (OLSE vs. BLUE)**.** *Consider the general linear model* $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$. *Then* $\mathrm{OLSE}(\mathbf{X}\boldsymbol{\beta}) = \mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta})$ *if and only if any one of the following six equivalent conditions holds.* (*Note:* $\mathbf{V}$ *may be replaced by its Moore–*

Penrose inverse $\mathbf{V}^+$ *and* $\mathbf{H}$ *and* $\mathbf{M} = \mathbf{I}_n - \mathbf{H}$ *may be interchanged.*)

(1) $\mathbf{HV} = \mathbf{VH}$,

(2) $\mathbf{HVM} = \mathbf{0}$,

(3) $\mathscr{C}(\mathbf{VH}) \subset \mathscr{C}(\mathbf{H})$,

(4) $\mathscr{C}(\mathbf{X})$ *has a basis comprising* $r = \operatorname{rank}(\mathbf{X})$ *orthonormal eigenvectors of* $\mathbf{V}$,

(5) $\mathbf{V} = \mathbf{HAH} + \mathbf{MBM}$ *for some* $\mathbf{A}$ *and* $\mathbf{B}$,

(6) $\mathbf{V} = \alpha\mathbf{I}_n + \mathbf{HKH} + \mathbf{MLM}$ *for some* $\alpha \in \mathbb{R}$, *and* $\mathbf{K}$ *and* $\mathbf{L}$.

Theorem 3 shows at once that under $\{\mathbf{y},\, \mathbf{X}\boldsymbol{\beta},\, \mathbf{I}_n\}$ the OLSE of $\mathbf{X}\boldsymbol{\beta}$ is trivially the BLUE; this result is often called the Gauss–Markov Theorem.

4 Two Linear Models

Consider now two linear models $\mathscr{M}_1 = \{\mathbf{y},\, \mathbf{X}\boldsymbol{\beta},\, \mathbf{V}_1\}$ and $\mathscr{M}_2 = \{\mathbf{y},\, \mathbf{X}\boldsymbol{\beta},\, \mathbf{V}_2\}$, which differ only in their covariance matrices. For the proof of the following proposition and related discussion, see, e.g., Rao (1971, Th. 5.2, Th. 5.5), and Mitra and Moore (1973, Th. 3.3, Th. 4.1–4.2).

Theorem 4. *Consider the linear models* $\mathscr{M}_1 = \{\mathbf{y},\, \mathbf{X}\boldsymbol{\beta},\, \mathbf{V}_1\}$ *and* $\mathscr{M}_2 = \{\mathbf{y},\, \mathbf{X}\boldsymbol{\beta},\, \mathbf{V}_2\}$, *and let the notation* $\{\mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta} \mid \mathscr{M}_1)\} \subset \{\mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta} \mid \mathscr{M}_2)\}$ *mean that every representation of the* BLUE *for* $\mathbf{X}\boldsymbol{\beta}$ *under* $\mathscr{M}_1$ *remains the* BLUE *for* $\mathbf{X}\boldsymbol{\beta}$ *under* $\mathscr{M}_2$. *Then the following statements are equivalent:*

(1) $\{\mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta} \mid \mathscr{M}_1)\} \subset \{\mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta} \mid \mathscr{M}_2)\}$,

(2) $\mathscr{C}(\mathbf{V}_2\mathbf{X}^{\perp}) \subset \mathscr{C}(\mathbf{V}_1\mathbf{X}^{\perp})$,

(3) $\mathbf{V}_2 = \mathbf{V}_1 + \mathbf{X}\mathbf{N}_1\mathbf{X}' + \mathbf{V}_1\mathbf{M}\mathbf{N}_2\mathbf{M}\mathbf{V}_1$, *for some* $\mathbf{N}_1$ *and* $\mathbf{N}_2$,

(4) $\mathbf{V}_2 = \mathbf{X}\mathbf{N}_3\mathbf{X}' + \mathbf{V}_1\mathbf{M}\mathbf{N}_4\mathbf{M}\mathbf{V}_1$, *for some* $\mathbf{N}_3$ *and* $\mathbf{N}_4$.

Notice that obviously

$$\{\mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta} \mid \mathscr{M}_1)\} = \{\mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta} \mid \mathscr{M}_2)\} \iff \mathscr{C}(\mathbf{V}_2\mathbf{X}^{\perp}) = \mathscr{C}(\mathbf{V}_1\mathbf{X}^{\perp}).$$

For the equality between the BLUEs of $\mathbf{X}_1\boldsymbol{\beta}_1$ under two partitioned models, see Haslett and Puntanen (2010a).

5 Model with New Observations: Best Linear Unbiased Predictor (BLUP)

Consider the model $\mathscr{M} = \{\mathbf{y},\, \mathbf{X}\boldsymbol{\beta},\, \mathbf{V}\}$, and let $\mathbf{y}_f$ denote an $m \times 1$ unobservable random vector containing *new observations.* The new observations are assumed to follow the linear model $\mathbf{y}_f = \mathbf{X}_f\boldsymbol{\beta} + \boldsymbol{\varepsilon}_f$, where $\mathbf{X}_f$ is a known $m \times p$ model matrix associated with new observations, $\boldsymbol{\beta}$ is the same vector of unknown parameters as in $\mathscr{M}$, and $\boldsymbol{\varepsilon}_f$ is an $m \times 1$ random error vector associated with new observations. Our goal is to predict the random vector $\mathbf{y}_f$ on the basis of $\mathbf{y}$. The expectation and the covariance matrix are

$$\mathrm{E}\begin{pmatrix}\mathbf{y}\\ \mathbf{y}_f\end{pmatrix} = \begin{pmatrix}\mathbf{X}\boldsymbol{\beta}\\ \mathbf{X}_f\boldsymbol{\beta}\end{pmatrix}, \quad \operatorname{cov}\begin{pmatrix}\mathbf{y}\\ \mathbf{y}_f\end{pmatrix} = \begin{pmatrix}\mathbf{V} = \mathbf{V}_{11} & \mathbf{V}_{12}\\ \mathbf{V}_{21} & \mathbf{V}_{22}\end{pmatrix},$$

which we may write as

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

A linear predictor $\mathbf{Ay}$ is said to be unbiased for $\mathbf{y}_f$ if $\mathrm{E}(\mathbf{Ay}) = \mathrm{E}(\mathbf{y}_f) = \mathbf{X}_f\boldsymbol{\beta}$ for all $\boldsymbol{\beta} \in \mathbb{R}^p$. Then the random vector $\mathbf{y}_f$ is said to be unbiasedly predictable. Now an unbiased linear predictor $\mathbf{Ay}$ is the best linear unbiased predictor, BLUP, for $\mathbf{y}_f$ if the Löwner ordering

$$\operatorname{cov}(\mathbf{Ay} - \mathbf{y}_f) \leq_{\mathrm{L}} \operatorname{cov}(\mathbf{By} - \mathbf{y}_f)$$

holds for all $\mathbf{B}$ such that $\mathbf{By}$ is an unbiased linear predictor for $\mathbf{y}_f$.

The following theorem characterizes the BLUP; see, e.g., Christensen (2002, p. 283), and Isotalo and Puntanen (2006, p. 1015).

Theorem 5 (Fundamental BLUP equation)**.** *Consider the linear model $\mathscr{M}_f$, where $\mathbf{X}_f\boldsymbol{\beta}$ is a given estimable parametric function. Then the linear estimator $\mathbf{Ay}$ is the best linear unbiased predictor* (BLUP) *for $\mathbf{y}_f$ if and only if $\mathbf{A}$ satisfies the equation*

$$\mathbf{A}(\mathbf{X} : \mathbf{VX}^{\perp}) = (\mathbf{X}_f : \mathbf{V}_{21}\mathbf{X}^{\perp}).$$

In terms of Pandora's Box (Theorem 2), $\mathbf{Ay}$ is the BLUP *for $\mathbf{y}_f$ if and only if there exists a matrix $\mathbf{L}$ such that $\mathbf{A}$ satisfies the equation*

$$\begin{pmatrix}\mathbf{V} & \mathbf{X}\\ \mathbf{X}' & \mathbf{0}\end{pmatrix}\begin{pmatrix}\mathbf{A}'\\ \mathbf{L}\end{pmatrix} = \begin{pmatrix}\mathbf{V}_{12}\\ \mathbf{X}'_f\end{pmatrix}.$$

6 The Mixed Model

A mixed linear model can be presented as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \quad \text{or shortly} \quad \mathscr{M}_{\mathrm{mix}} = \{\mathbf{y},\, \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma},\, \mathbf{D},\, \mathbf{R}\},$$

where $\mathbf{X} \in \mathbb{R}^{n\times p}$ and $\mathbf{Z} \in \mathbb{R}^{n\times q}$ are known matrices, $\boldsymbol{\beta} \in \mathbb{R}^p$ is a vector of unknown fixed effects, $\boldsymbol{\gamma}$ is an unobservable vector (q elements) of *random effects* with $\operatorname{cov}(\boldsymbol{\gamma}, \boldsymbol{\varepsilon}) = \mathbf{0}_{q\times p}$ and

$$\mathrm{E}(\boldsymbol{\gamma}) = \mathbf{0}_q, \quad \operatorname{cov}(\boldsymbol{\gamma}) = \mathbf{D}_{q\times q}, \quad \mathrm{E}(\boldsymbol{\varepsilon}) = \mathbf{0}_n\,, \quad \operatorname{cov}(\boldsymbol{\varepsilon}) = \mathbf{R}_{n\times n}.$$

This leads directly to:

Theorem 6. *Consider the mixed model* $\mathscr{M}_{\text{mix}} = \{\mathbf{y},\, \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma},\, \mathbf{D},\, \mathbf{R}\}$. *Then the linear estimator* $\mathbf{By}$ *is the* BLUE *for* $\mathbf{X}\boldsymbol{\beta}$ *if and only if*

$$\mathbf{B}(\mathbf{X} : \boldsymbol{\Sigma}\mathbf{X}^{\perp}) = (\mathbf{X} : \mathbf{0}),$$

where $\boldsymbol{\Sigma} = \mathbf{ZDZ}' + \mathbf{R}$. *Moreover,* $\mathbf{Ay}$ *is the* BLUP *for* $\boldsymbol{\gamma}$ *if and only if*

$$\mathbf{A}(\mathbf{X} : \boldsymbol{\Sigma}\mathbf{X}^{\perp}) = (\mathbf{0} : \mathbf{DZ}'\mathbf{X}^{\perp}).$$

In terms of Pandora's Box (Theorem 2), $\mathbf{Ay} = \text{BLUP}(\boldsymbol{\gamma})$ *if and only if there exists a matrix* $\mathbf{L}$ *such that* $\mathbf{A}$ *satisfies the equation*

$$\begin{pmatrix} \boldsymbol{\Sigma} & \mathbf{X} \\ \mathbf{X}' & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{A}' \\ \mathbf{L} \end{pmatrix} = \begin{pmatrix} \mathbf{ZD} \\ \mathbf{0} \end{pmatrix}.$$

For the equality between the BLUPs under two mixed models, see Haslett and Puntanen (2010b, 2010c).

6.1 Note

Reprinted with permission from Lovric, Miodrag (2011), *International Encyclopedia of Statistical Science.* Heidelberg: Springer Science+Business Media, LLC.

7 References

1. Anderson, T. W. (1948). On the theory of testing serial correlation. *Skandinavisk Aktuarietidskrift*, **31**, 88–116.

2. Baksalary, Jerzy K.; Rao, C. Radhakrishna and Markiewicz, Augustyn (1992). A study of the influence of the 'natural restrictions' on estimation problems in the singular Gauss–Markov model, *Journal of Statistical Planning and Inference*, **31**, 335–351.

3. Christensen, Ronald (2002). *Plane Answers to Complex Questions: The Theory of Linear Models,* 3rd Edition. Springer, New York.

4. Haslett, Stephen J. and Puntanen, Simo (2010a). Effect of adding regressors on the equality of the BLUEs under two linear models. *Journal of Statistical Planning and Inference*, **140**, 104–110,

5. Haslett, Stephen J. and Puntanen, Simo (2010b). Equality of BLUEs or BLUPs under two linear models using stochastic restrictions. *Statistical Papers*, **51**, 465–475.

6. Haslett, Stephen J. and Puntanen, Simo (2010c). On the equality of the BLUPs under two linear mixed models. *Metrika*, available online, DOI 10.1007/s00184-010-0308-6.

7. Isotalo, Jarkko and Puntanen, Simo (2006). Linear prediction sufficiency for new observations in the general Gauss–Markov model. *Communications in Statistics: Theory and Methods*, **35**, 1011–1023.

8. Kruskal, William (1967). When are Gauss–Markov and least squares estimators identical? A coordinate-free approach. *The Annals of Mathematical Statistics*, **39**, 70–75.

9. Mitra, Sujit Kumar and Moore, Betty Jeanne (1973). Gauss–Markov estimation with an incorrect dispersion matrix. *Sankhyā, Series A*, **35**, 139–152.

10. Puntanen, Simo and Styan, George P. H. (1989). The equality of the ordinary least squares estimator and the best linear unbiased estimator [with comments by Oscar Kempthorne and by Shayle R. Searle and with "Reply" by the authors]. *The American Statistician*, **43**, 153–164.

11. Puntanen, Simo; Styan, George P. H. and Werner, Hans Joachim (2000). Two matrix-based proofs that the linear estimator Gy is the best linear unbiased estimator. *Journal of Statistical Planning and Inference*, **88**, 173–179.

12. Rao, C. Radhakrishna (1967). Least squares theory using an estimated dispersion matrix and its application to measurement of signals. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability: Berkeley, California, 1965/1966*, vol. 1 (Eds. Lucien M. Le Cam & Jerzy Neyman), University of California Press, Berkeley, pp. 355–372.

13. Rao, C. Radhakrishna (1971). Unified theory of linear estimation. *Sankhyā, Series A*, **33**, 371–394. [Corrigenda (1972), **34**, p. 194 and p. 477.]

14. Rao, C. Radhakrishna (1974). Projectors, generalized inverses and the BLUE's. *Journal of the Royal Statistical Society, Series B*, **36**, 442–448.

15. Watson, Geoffrey S. (1967). Linear least squares regression. *The Annals of Mathematical Statistics*, **38**, 1679–1699.

16. Zyskind, George (1967). On canonical forms, non-negative covariance matrices and best and simple least squares linear estimators in linear models. *The Annals of Mathematical Statistics*, **38**, 1092–1109.

17. Zyskind, George and Martin, Frank B. (1969). On best linear estimation and general Gauss–Markov theorem in linear models with arbitrary non-negative covariance structure. *SIAM Journal on Applied Mathematics*, **17**, 1190–1202.