# The legend of the equality of OLSE and BLUE: highlighted by C.R. Rao in 1967

Augustyn Markiewicz, Simo Puntanen and George P.H. Styan **3 May 2021**

Appeared in *Methodology and Applications of Statistics – A Volume in Honor of C.R. Rao on the Occasion of his 100th Birthday*.
B.C. Arnold, N. Balakrishnan and C.A. Coelho, Eds., Springer, Cham,
pages 51–76, DOI, Book's DOI.

**Abstract** In this article we go through some crucial developments regarding the equality of the ordinary least squares estimator and the best linear unbiased estimator in the general linear model. C.R. Rao (Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability. University of California Press, Berkeley, pp. 355–372, 1967) appears to be the first to provide necessary and sufficient conditions for the general case when both the model matrix and the random error term's covariance matrix are possibly deficient in rank. We describe the background of the problem area and provide some examples. We also consider some personal CRR-related glimpses of our research careers and provide a rather generous list of references.

*Key words and phrases*: Best linear unbiased estimator, BLUE, efficiency of ordinary least squares, estimability, generalized inverse, ordinary least squares estimator, OLSE, linear model, Löwner ordering.

## 1 Introduction and background

Let us begin by quoting the beginning of Chapter 10, entitled "BLUE", of the *Matrix Tricks Book* by Puntanen, Styan & Isotalo (2011):

Augustyn Markiewicz
Department of Mathematical and Statistical Methods, Poznań University of Life Sciences, Wojska Polskiego 28, PL-60637 Poznań, Poland, e-mail: amark@up.poznan.pl

Simo Puntanen
Faculty of Information Technology and Communication Sciences, FI-33014 Tampere University, Finland, e-mail: simo.puntanen@tuni.fi

George P.H. Styan
Department of Mathematics and Statistics, McGill University, 805 Sherbrooke Street West, Montréal, Québec, Canada H3A 0B9, e-mail: geostyan@gmail.com

Over the years, one of our favourite research topics in linear models has been the equality between OLSE and BLUE of $\mathbf{X}\boldsymbol{\beta}$. In Proposition 10.1 [in the present article Theorem 4] we collect together some necessary and sufficient conditions for their equality. We find this collection very useful and we believe it includes several interesting linear algebraic problems.

While preparing this book, for a long time the title of this chapter was “OLSE vs BLUE”, but the simpler version “BLUE” describes better the matrix tricks under consideration. In the sections of this chapter we consider, for example, the best unbiased linear predictors, BLUPs, and mixed models.

We will closely follow the article “The equality of the ordinary least squares estimator and the best linear unbiased estimator” in *The American Statistician* by Puntanen & Styan (1989), the most cited joint paper by these authors, as well as the articles by Baksalary, Puntanen & Styan (1990b) and by Puntanen & Styan (1996): “A brief biography and appreciation of Calyampudi Radhakrishna Rao, with a bibliography of books and papers”. Our aim is to give an easy-to-read review for a non-expert of the area and illustrate the role of C.R. Rao in its development. This article contains no new technical results, makes no claim at completeness; this is a brief survey—but we believe that the years after 1989 have matured our insight into this area. Browsing again through the old material was very interesting. Hopefully we can express this in what follows.

To give some perspective, we start by going through some background in the spirit of Puntanen & Styan (1996).

In 1954 C.R. Rao received some data collected in Japan in order to study the long-term effects of radiation on atom bomb casualties in Hiroshima and Nagasaki. The statistical analysis involved finding a matrix to replace the inverse of $\mathbf{X}'\mathbf{X}$, where $\mathbf{X}$ is the model matrix in the linear model and $\mathbf{X}'$ stands for its transpose; here the matrix $\mathbf{X}'\mathbf{X}$ was singular. This led to a *pseudoinverse* which was introduced by Rao (1955) in *Sankhyā*. This was the same year that Penrose (1955) published his paper on generalized inverses. Rao then discovered that the key condition for a generalized inverse $\mathbf{G}$ of a matrix $\mathbf{A}$ was the equation $\mathbf{AGA} = \mathbf{A}$, introducing the notation $\mathbf{G} = \mathbf{A}^-$. The calculus of generalized inverses and the unified theory of linear estimation were then presented by Rao (1962), in the *Journal of the Royal Statistical Society, Ser. B.* The subject of generalized inverses was further developed leading to the monograph (1971a) with Sujit Kumar Mitra entitled *Generalized Inverse of Matrices and its Applications*.

As a sidetrack, below is an excerpt from S.K. Mitra’s interview, carried out in February 1993 in the Indian Statistical Institute, New Delhi, see Puntanen & Styan (2012). Professor Mitra was replying to the following question: When was the decision made that you will start writing that book with Professor Rao?

In 1967 we had a summer school at the ISI, with a lot of students participating. Often new areas of statistics and mathematics were exposed to the students during these six weeks of summer. I was in fact once the programme director of such a summer school.

As a member of the summer school, I was able to get the best of C.R. Rao’s papers and manuscripts. So I taught a course in the summer school and then by the time I had completed the course, I myself had some new results. In fact my first two papers on generalized inverses, which appeared in 1968, were essentially papers that appeared in their first form in these summer schools.

> Professor C.R. Rao had at that time already decided to write a book on generalized inverses all by himself. It was also announced as a forthcoming publication of the Statistical Publishing Society in Calcutta. He must have seen my new results and in a few days he invited me be a co-author. That is how that book started.

Using the concept of generalized inverse, Rao (1971, 1973c) further developed a unified theory for linear estimation, noting that generalized inverses were particularly helpful with explicit expressions for projectors. We wish to cite a few words from the Appendix of the article by Rao (1971). The title of the Appendix was "The Atom Bomb and Generalized Inverse". Below the first and last paragraph of the Appendix are quoted.

> The author was first led to the definition of a pseudo-inverse (now called generalized inverse or g-inverse) of a singular matrix in 1945–1955 when he undertook to carry out multivariate analysis of anthropometric data obtained on families of Hiroshima and Nagasaki to study the effects of radiation due atom bomb explosions, on request from Dr. W.J. Schull of the University of Michigan. The computation and use of a pseudo-inverse are given in a statistical report prepared by the author, which is incorporated in Publication No. 461 of the National Academy of Sciences, U.S.A., by Neel & Schull (1956). It may be of interest to the audience to know the circumstances under which the pseudoinverse had to be introduced.
>
> It is hard to believe that scientists have found in what has been described as the greatest tragedy a source for providing material and simulation for research in many directions.

## 2 What is OLSE, what is BLUE?

Let us quickly recall the definition of the ordinary least squares estimator, OLSE, and the best linear unbiased estimator, BLUE, and before that, the linear statistical model under discussion. We will consider the general linear model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}\,, \quad \text{or shortly the triplet } \mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}\,, \tag{1}$$

where $\mathbf{X}$ is a known $n \times p$ model matrix, $\mathbf{y}$ is an observable $n$-dimensional random vector, $\boldsymbol{\beta}$ is $p$-dimensional vector of unknown but fixed parameters, and $\boldsymbol{\varepsilon}$ is an unobservable vector of random errors with expectation $\mathrm{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$, and covariance matrix $\mathrm{cov}(\boldsymbol{\varepsilon}) = \mathbf{V}$. We will denote $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ so that $\mathrm{E}(\mathbf{y}) = \boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$. Often the covariance matrix is of the type $\sigma^2\mathbf{V}$, where $\sigma^2$ is an unknown positive constant. However, in our considerations $\sigma^2$ has no role. The nonnegative definite matrix $\mathbf{V}$ is known and can be singular. If $\mathbf{V}$ is not known things get much more complicated; for the so-called empirical best linear unbiased predictors in the linear mixed model, see, for example, Haslett & Welsh (2019).

Then some words about the notation. The symbols $\mathbf{A}^-$, $\mathbf{A}^+$, $\mathbf{A}'$, $\mathscr{C}(\mathbf{A})$, and $\mathscr{C}(\mathbf{A})^\perp$, denote, respectively, a generalized inverse, the (unique) Moore–Penrose inverse, the transpose, the column space, and the orthogonal complement of the column space of the matrix $\mathbf{A}$. Notation $\mathbf{A}^-$ refers to any matrix satisfying $\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A}$ and $\mathbf{A}^+$ satisfies the four Moore–Penrose conditions. By $(\mathbf{A} : \mathbf{B})$ we denote the partitioned matrix with $\mathbf{A}_{a\times b}$ and $\mathbf{B}_{a\times c}$ as submatrices. The symbol $\mathbf{A}^\perp$ stands for any matrix satisfying $\mathscr{C}(\mathbf{A}^{\perp}) = \mathscr{C}(\mathbf{A})^{\perp}$. Furthermore, we will use $\mathbf{P_A} = \mathbf{AA}^{+} = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-}\mathbf{A}'$ to denote the orthogonal projector (with respect to the standard inner product) onto the column space $\mathscr{C}(\mathbf{A})$, and $\mathbf{Q_A} = \mathbf{I} - \mathbf{P_A}$, where $\mathbf{I}$ refers to the identity matrix of appropriate order. In particular, we denote shortly

$$\mathbf{H} = \mathbf{P_X}, \quad \mathbf{M} = \mathbf{I}_n - \mathbf{P_X}. \tag{2}$$

A linear statistic $\mathbf{By}$ is said to be a linear unbiased estimator, LUE, for the parametric function $\mathbf{K}\boldsymbol{\beta}$, where $\mathbf{K} \in \mathbb{R}^{q\times p}$, if its expectation is equal to $\mathbf{K}\boldsymbol{\beta}$, i.e.,

$$\mathrm{E}(\mathbf{By}) = \mathbf{BX}\boldsymbol{\beta} = \mathbf{K}\boldsymbol{\beta} \quad \text{for all } \boldsymbol{\beta} \in \mathbb{R}^p, \text{ i.e., } \mathbf{BX} = \mathbf{K}. \tag{3}$$

When $\mathscr{C}(\mathbf{K}') \subseteq \mathscr{C}(\mathbf{X}')$ holds, $\mathbf{K}\boldsymbol{\beta}$ is said to be estimable.

**Definition 1** The linear unbiased estimator $\mathbf{By}$ is the best linear unbiased estimator, BLUE, of estimable $\mathbf{K}\boldsymbol{\beta}$ if $\mathbf{By}$ has the smallest covariance matrix in the Löwner sense among all linear unbiased estimators of $\mathbf{K}\boldsymbol{\beta}$:

$$\mathrm{cov}(\mathbf{By}) \leq_{\mathrm{L}} \mathrm{cov}(\mathbf{B}_{\#}\mathbf{y}) \quad \text{for all } \mathbf{B}_{\#} : \mathbf{B}_{\#}\mathbf{X} = \mathbf{K}, \tag{4}$$

that is, $\mathrm{cov}(\mathbf{B}_{\#}\mathbf{y}) - \mathrm{cov}(\mathbf{By})$ is nonnegative definite for all $\mathbf{B}_{\#} : \mathbf{B}_{\#}\mathbf{X} = \mathbf{K}$.

Under the model $\{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$, the ordinary least squares estimator, OLSE, for $\boldsymbol{\beta}$ is the solution minimizing the quantity $\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2$ with respect to $\boldsymbol{\beta}$ yielding to the normal equation $\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{y}$. Thus, if $\mathbf{X}$ has full column rank, the OLSE of $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{X}^{+}\mathbf{y}$. In the general case, the set of *all* vectors $\hat{\boldsymbol{\beta}}$ satisfying $\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$, can be written as

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y} + [\mathbf{I}_p - (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}]\mathbf{t}, \tag{5}$$

where $(\mathbf{X}'\mathbf{X})^{-}$ is an arbitrary (but fixed) generalized inverse of $\mathbf{X}'\mathbf{X}$ and $\mathbf{t} \in \mathbb{R}^p$ is free to vary. On the other hand, every solution to the normal equations can be written as $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y}$ for some $(\mathbf{X}'\mathbf{X})^{-}$.

Of course, it is questionable whether it is quite correct to call $\hat{\boldsymbol{\beta}}$ an estimator when it is not unique (after $\mathbf{y}$ is being observed); it is merely a *solution* to the normal equations; "This point cannot be overemphasized", as stated by Searle (1971, p. 169). In this context we wish to cite also the following from Searle (2000, p. 26):

> One of the greatest contributions to understanding the apparent quirkiness of normal equations of non-full rank (as is customary with linear models), which have an infinity of solutions, is due to Rao (1962). Using the work of Moore (1920) and Penrose (1955), he showed how a generalized inverse matrix yields a solution to the normal equations and how that solution can be used to establish estimable functions and their estimators—and these results are invariant to whatever generalized inverse is being used. Although the arithmetic of generalized inverses is scarcely any less than that of regular inverses, the use of generalized inverses is of enormous help in understanding estimability and its consequences.

If $\mathbf{K}\boldsymbol{\beta}$ is estimable, then $\mathbf{K}\hat{\boldsymbol{\beta}} = \mathbf{K}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y}$, i.e., the OLSE of $\mathbf{K}\boldsymbol{\beta}$ is unique whatever choice of $\hat{\boldsymbol{\beta}}$, i.e., whatever $(\mathbf{X}'\mathbf{X})^{-}$ we use. This can be seen from Lemma

2.2.4 of Rao & Mitra (1971a) which states that for nonnull matrices $\mathbf{A}$ and $\mathbf{C}$ the following holds:

$$\mathbf{AB}^{-}\mathbf{C} = \mathbf{AB}^{+}\mathbf{C} \text{ for all } \mathbf{B}^{-} \iff \mathscr{C}(\mathbf{C}) \subseteq \mathscr{C}(\mathbf{B}) \;\&\; \mathscr{C}(\mathbf{A}') \subseteq \mathscr{C}(\mathbf{B}')\,. \tag{6}$$

In particular, choosing $(\mathbf{X}'\mathbf{X})^{-}$ as $(\mathbf{X}'\mathbf{X})^{+}$ and using $\mathbf{X}^{+} = (\mathbf{X}'\mathbf{X})^{+}\mathbf{X}'$, we can write $\mathbf{K}\hat{\boldsymbol{\beta}} = \mathbf{K}(\mathbf{X}'\mathbf{X})^{+}\mathbf{X}'\mathbf{y} = \mathbf{KX}^{+}\mathbf{y}$.

For $\mathbf{K} = \mathbf{X}$ we have

$$\mathrm{OLSE}(\mathbf{X}\boldsymbol{\beta}) = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y} = \mathbf{XX}^{+}\mathbf{y} = \mathbf{P_X y} = \mathbf{Hy} = \hat{\boldsymbol{\mu}}\,. \tag{7}$$

Obviously $\hat{\boldsymbol{\mu}} = \mathbf{Hy}$ is a LUE for $\mathbf{X}\boldsymbol{\beta}$. Let $\mathbf{By}$ be another LUE of $\mathbf{X}\boldsymbol{\beta}$, i.e., $\mathbf{B}$ satisfies $\mathbf{BX} = \mathbf{X}$ and thereby $\mathbf{BH} = \mathbf{H} = \mathbf{HB}'$. Thus, under the model $\{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{I}_n\}$:

$$\mathrm{cov}(\mathbf{By}) = \mathrm{cov}[\mathbf{Hy} - (\mathbf{H} - \mathbf{B})\mathbf{y}] = \mathrm{cov}(\mathbf{Hy}) + \mathrm{cov}[(\mathbf{H} - \mathbf{B})\mathbf{y}] \geq_{\mathrm{L}} \mathrm{cov}(\mathbf{Hy})\,, \tag{8}$$

and so we have proved a simple version of the *Gauss–Markov theorem*:

**Theorem 1** *Under the model* $\{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{I}_n\}$*, the* OLSE *of* $\mathbf{X}\boldsymbol{\beta}$ *is the* BLUE *of* $\mathbf{X}\boldsymbol{\beta}$*, or shortly*

$$\hat{\boldsymbol{\mu}} = \mathrm{OLSE}(\mathbf{X}\boldsymbol{\beta}) = \mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta}) = \tilde{\boldsymbol{\mu}}\,, \tag{9}$$

*and for any estimable* $\mathbf{K}\boldsymbol{\beta}$*,* $\mathrm{OLSE}(\mathbf{K}\boldsymbol{\beta}) = \mathrm{BLUE}(\mathbf{K}\boldsymbol{\beta})$.

When (9) holds, we will use phrases like "OLSE is BLUE". The claim concerning estimable $\mathbf{K}\boldsymbol{\beta}$ in Theorem 1 can be confirmed by observing that due to estimability, $\mathbf{K}\boldsymbol{\beta} = \mathbf{LX}\boldsymbol{\beta}$ for some $\mathbf{L}$ and thereby $\mathrm{OLSE}(\mathbf{K}\boldsymbol{\beta}) = \mathbf{L}\,\mathrm{OLSE}(\mathbf{X}\boldsymbol{\beta}) = \mathbf{LHy}$. Actually, under $\{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$, the statements $\mathrm{OLSE}(\mathbf{X}\boldsymbol{\beta}) = \mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta})$ and $\mathrm{OLSE}(\mathbf{K}\boldsymbol{\beta}) = \mathrm{BLUE}(\mathbf{K}\boldsymbol{\beta})$ for *all* estimable $\mathbf{K}\boldsymbol{\beta}$ are equivalent. It is clear that $\boldsymbol{\beta}$ is estimable if and only if $\mathbf{X}$ has full column rank.

Consider now the model $\mathscr{M}$ where $\mathbf{V}$ is positive definite, and suppose that $\mathbf{V}^{1/2}$ is the positive definite square root of $\mathbf{V}$. Premultiplying $\mathscr{M}$ by $\mathbf{V}^{-1/2}$ gives the transformed model $\mathscr{M}_{\#} = \{\mathbf{V}^{-1/2}\mathbf{y},\ \mathbf{V}^{-1/2}\mathbf{X}\boldsymbol{\beta},\ \mathbf{I}_n\}$. In light of Theorem 1, the BLUE of $\mathbf{X}\boldsymbol{\beta}$ under $\mathscr{M}_{\#}$ equals the OLSE under $\mathscr{M}_{\#}$ and thus

$$\mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta} \mid \mathscr{M}_{\#}) = \tilde{\boldsymbol{\mu}}(\mathscr{M}_{\#}) = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} = \mathbf{P}_{\mathbf{X};\mathbf{V}^{-1}}\mathbf{y}\,, \tag{10}$$

where $\mathbf{P}_{\mathbf{X};\mathbf{V}^{-1}}$ is the orthogonal projector onto $\mathscr{C}(\mathbf{X})$ when the inner product matrix is $\mathbf{V}^{-1}$. Here is a crucial question: is the BLUE of $\mathbf{X}\boldsymbol{\beta}$ under $\mathscr{M}_{\#}$ the same as under $\mathscr{M}$, in other words, has the transformation done via $\mathbf{V}^{-1/2}$ any effect on the BLUE of $\mathbf{X}\boldsymbol{\beta}$? The answer is that there is no effect and that

$$\mathbf{P}_{\mathbf{X};\mathbf{V}^{-1}}\mathbf{y} = \mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta} \mid \mathscr{M}) = \mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta} \mid \mathscr{M}_{\#}) = \mathrm{OLSE}(\mathbf{X}\boldsymbol{\beta} \mid \mathscr{M}_{\#})\,. \tag{11}$$

The result (11), sometimes referred to as the Aitken-approach, see Aitken (1935), Farebrother (1990, 1997) and Searle (1996), is well known in statistical textbooks. Farebrother (1990) points out that Aitken's contribution to the subject was to show that a least squares estimator of $\boldsymbol{\beta}$ minimizing $(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$ could be obtained by premultiplying the model $\{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$ by an $n \times n$ matrix $\mathbf{D}$ satisfying $\mathbf{DVD}' = \mathbf{I}_n$. However, Aitken did not show that this estimator was the best linear unbiased estimator.

The property that the transformation via $\mathbf{V}^{-1/2}$ has no effect on the BLUE is phrased as "$\mathbf{V}^{-1/2}\mathbf{y}$ is linearly sufficient for $\mathbf{X}\boldsymbol{\beta}$"; see, e.g., Baksalary & Kala (1981a, 1986) and Haslett et al. (2021).

**Example 1.** [Very simple model.] Let us consider a linear model $\mathbf{y} = \mathbf{x}\beta + \boldsymbol{\varepsilon}$, where $\operatorname{cov}(\mathbf{y}) = \mathbf{V}$. Then

$$\operatorname{OLSE}(\beta) = \hat{\beta} = (\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{y}, \qquad \operatorname{var}(\hat{\beta}) = (\mathbf{x}'\mathbf{x})^{-2}\mathbf{x}'\mathbf{V}\mathbf{x}, \tag{12a}$$

$$\operatorname{BLUE}(\beta) = \tilde{\beta} = (\mathbf{x}'\mathbf{V}^{-1}\mathbf{x})^{-1}\mathbf{x}'\mathbf{V}^{-1}\mathbf{y}, \quad \operatorname{var}(\tilde{\beta}) = (\mathbf{x}'\mathbf{V}^{-1}\mathbf{x})^{-1}, \tag{12b}$$

where "var" refers to the variance. Now we have $\operatorname{var}(\tilde{\beta}) \leq \operatorname{var}(\hat{\beta})$, i.e.,

$$(\mathbf{x}'\mathbf{V}^{-1}\mathbf{x})^{-1} \leq (\mathbf{x}'\mathbf{x})^{-2}\mathbf{x}'\mathbf{V}\mathbf{x}, \quad \text{i,e.,} \quad (\mathbf{x}'\mathbf{x})^2 \leq \mathbf{x}'\mathbf{V}^{-1}\mathbf{x} \cdot \mathbf{x}'\mathbf{V}\mathbf{x}, \tag{13}$$

which is a special case of the famous Cauchy–Schwarz inequality. It is well known that the equality in (13) holds if and only if

$$\mathbf{V}\mathbf{x} = \lambda\mathbf{x}, \quad \text{for some } \lambda \in \mathbb{R}, \tag{14}$$

and hence $\mathbf{x}$ is an eigenvector of $\mathbf{V}$ corresponding to eigenvalue $\lambda$. Condition (14) is just a version of Anderson's (1948) condition for the equality of OLSE and BLUE; see the beginning of Section 3 below. Notice that putting $\mathbf{x} = \mathbf{1}$, a vector of ones, shows that the arthmetic mean $\bar{y}$ is BLUE whenever $\mathbf{V}$ has its row totals equal, i.e., $\mathbf{V}\mathbf{1} = \lambda\mathbf{1}$ for some scalar $\lambda$.

We might ask how "bad" the OLSE could be with respect to the BLUE. One natural measure for the relative efficiency of OLSE is the ratio of their variances:

$$\phi = \operatorname{eff}(\hat{\beta}) = \frac{\operatorname{var}(\tilde{\beta})}{\operatorname{var}(\hat{\beta})} = \frac{(\mathbf{x}'\mathbf{V}^{-1}\mathbf{x})^{-1}}{(\mathbf{x}'\mathbf{x})^{-1}\mathbf{x}'\mathbf{V}\mathbf{x}(\mathbf{x}'\mathbf{x})^{-1}} = \frac{(\mathbf{x}'\mathbf{x})^2}{\mathbf{x}'\mathbf{V}\mathbf{x} \cdot \mathbf{x}'\mathbf{V}^{-1}\mathbf{x}}\,. \tag{15}$$

Clearly we have $0 < \phi \leq 1$, where the upper bound is obtained if and only if OLSE equals BLUE. The lower bound of $\phi$ can be obtained from the Kantorovich inequality; see, e.g., Watson et al. (1997),

$$\tau_1^2 := \frac{4\lambda_1\lambda_n}{(\lambda_1 + \lambda_n)^2} \leq \frac{(\mathbf{x}'\mathbf{x})^2}{\mathbf{x}'\mathbf{V}\mathbf{x} \cdot \mathbf{x}'\mathbf{V}^{-1}\mathbf{x}} = \operatorname{eff}(\hat{\beta}) = \phi, \tag{16}$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n > 0$ are the eigenvalues of $\mathbf{V}$. The lower bound is obtained when $\mathbf{x}$ is proportional either to $\mathbf{t}_1 + \mathbf{t}_n$ or to $\mathbf{t}_1 - \mathbf{t}_n$; in short, $\mathbf{x}$ is proportional to $\mathbf{x}_{\text{bad}} = \mathbf{t}_1 \pm \mathbf{t}_n$, where $\mathbf{T} = (\mathbf{t}_1 : \mathbf{t}_2 : \ldots : \mathbf{t}_n)$ is the matrix with $\mathbf{t}_i$ being the orthonormal eigenvectors of $\mathbf{V}$ corresponding to eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. □

Consider then the covariance matrices of OLSE and BLUE when $\mathbf{X}$ has a full column rank and $\mathbf{V}$ is positive definite. Then under $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$,

$$\operatorname{cov}(\hat{\boldsymbol\beta}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}, \quad \operatorname{cov}(\tilde{\boldsymbol\beta}) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}, \tag{17}$$

and we have the Löwner ordering

$$(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \leq_{\mathrm{L}} (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}. \tag{18}$$

If $\mathbf{X}$ does not have a full column rank then $\mathbf{X}\tilde{\boldsymbol\beta} = \tilde{\boldsymbol\mu} = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}$ and

$$\operatorname{cov}(\tilde{\boldsymbol\mu}) = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}' \leq_{\mathrm{L}} \mathbf{HVH} = \operatorname{cov}(\hat{\boldsymbol\mu})\,. \tag{19}$$

What is now interesting is that the difference $\operatorname{cov}(\hat{\boldsymbol\beta}) - \operatorname{cov}(\tilde{\boldsymbol\beta})$ has an alternative representation, expressed in Theorem 2 below. Among the first places where Theorem 2 occurs are probably the papers by Khatri (1966, Lemma 1) and Rao (1967, Lemmas 2a–2c); see also Rao (1973a, Problem 33, p. 77).

**Theorem 2** *Consider the linear model* $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol\beta, \mathbf{V}\}$*, where* $\mathbf{X}$ *has full column rank and* $\mathbf{V}$ *is positive definite, and denote* $\mathbf{H} = \mathbf{P}_{\mathbf{X}}$*,* $\mathbf{M} = \mathbf{I}_n - \mathbf{H}$*. Then*

(a) $\operatorname{cov}(\tilde{\boldsymbol\beta}) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} = (\mathbf{X}'\mathbf{X})^{-1}[\mathbf{X}'\mathbf{V}\mathbf{X} - \mathbf{X}'\mathbf{VM}(\mathbf{MVM})^{-}\mathbf{MVX}](\mathbf{X}'\mathbf{X})^{-1}$
$= \mathbf{X}^{+}[\mathbf{V} - \mathbf{VM}(\mathbf{MVM})^{-}\mathbf{MV}](\mathbf{X}^{+})'$
$= \operatorname{cov}(\hat{\boldsymbol\beta}) - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{VM}(\mathbf{MVM})^{-}\mathbf{MVX}(\mathbf{X}'\mathbf{X})^{-1}$,
(b) $\operatorname{cov}(\hat{\boldsymbol\beta}) - \operatorname{cov}(\tilde{\boldsymbol\beta}) = \mathbf{X}^{+}\mathbf{VM}(\mathbf{MVM})^{-}\mathbf{MV}(\mathbf{X}^{+})'$,
(c) $\operatorname{cov}(\hat{\boldsymbol\mu}) - \operatorname{cov}(\tilde{\boldsymbol\mu}) = \mathbf{HVH} - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}' = \mathbf{HVM}(\mathbf{MVM})^{-}\mathbf{MVH}$,
(d) $\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1} = \mathbf{I}_n - \mathbf{VM}(\mathbf{MVM})^{-}\mathbf{M} = \mathbf{H} - \mathbf{HVM}(\mathbf{MVM})^{-}\mathbf{M}$.

*In* (c) *and* (d) *the matrix* $\mathbf{X}$ *does not need to have full column rank.*

Actually, instead of $\mathbf{M}$, Rao (1967, Lemmas 2a–2c) used a full column rank matrix $\mathbf{Z}$ spanning $\mathscr{C}(\mathbf{X})^{\perp}$. Thus, for example,

$$(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1} = \mathbf{X}^{+} - \mathbf{X}^{+}\mathbf{VZ}(\mathbf{Z}'\mathbf{VZ})^{-1}\mathbf{Z}'. \tag{20}$$

It is noteworthy that for a positive definite $\mathbf{V}$ we have

$$\begin{aligned}\dot{\mathbf{M}} := \mathbf{M}(\mathbf{MVM})^{-}\mathbf{M} &= \mathbf{V}^{-1/2}\mathbf{P}_{\mathbf{V}^{1/2}\mathbf{M}}\mathbf{V}^{-1/2} \\ &= \mathbf{V}^{-1/2}(\mathbf{I}_n - \mathbf{P}_{\mathbf{V}^{-1/2}\mathbf{X}})\mathbf{V}^{-1/2} = \mathbf{V}^{-1}(\mathbf{I}_n - \mathbf{P}_{\mathbf{X};\mathbf{V}^{-1}})\,. \end{aligned} \tag{21}$$

Thus the BLUE's residual can be expressed as $\mathbf{y} - \tilde{\boldsymbol\mu} = \mathbf{V}\dot{\mathbf{M}}\mathbf{y}$ and the "weighted sum of squares of errors" is

$$\operatorname{SSE}(\mathbf{V}) = (\mathbf{y} - \tilde{\boldsymbol\mu})'\mathbf{V}^{-1}(\mathbf{y} - \tilde{\boldsymbol\mu}) = \mathbf{y}'\dot{\mathbf{M}}\mathbf{y}\,, \tag{22}$$

while the corresponding quantity in model $\{\mathbf{y}, \mathbf{X}\boldsymbol\beta, \mathbf{I}_n\}$ is $\operatorname{SSE}(\mathbf{I}) = \mathbf{y}'\mathbf{M}\mathbf{y}$.

What does it mean to have the equality OLSE = BLUE? There is no problem if this equality is interpreted as a short version of the phrase "$\mathbf{Hy}$ has the minimal covariance matrix in the sense of Definition 1". On the other hand, for example in the full rank model, we might ask what is the "real meaning" of the equality

$$\mathbf{Hy} = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} = \mathbf{P}_{\mathbf{X};\mathbf{V}^{-1}}\mathbf{y}\,? \tag{23}$$

Estimator $\mathbf{Hy}$ is a random vector as is $\mathbf{P}_{\mathbf{X};\mathbf{V}^{-1}}\mathbf{y}$; this is so before observing $\mathbf{y}$. The equality of two random vectors requires a specific definition and the essential matter is how the set of possible realized values of the response variable $\mathbf{y}$ is defined. Notice that we do not make a notational difference between the random vector $\mathbf{y}$ and its realized value.

The model $\mathscr{M}$ is said to be *consistent* if the observed value of $\mathbf{y}$ lies in $\mathscr{C}(\mathbf{X} : \mathbf{V})$:

$$\mathbf{y} \in \mathscr{C}(\mathbf{X} : \mathbf{V}) = \mathscr{C}(\mathbf{X} : \mathbf{V}\mathbf{X}^{\perp}) = \mathscr{C}(\mathbf{X} : \mathbf{VM}) = \mathscr{C}(\mathbf{X}) \oplus \mathscr{C}(\mathbf{VM}), \tag{24}$$

where $\oplus$ refers to the direct sum. For the equality $\mathscr{C}(\mathbf{X} : \mathbf{V}) = \mathscr{C}(\mathbf{X} : \mathbf{VM})$, see, e.g., (Rao, 1974, Lemma 2.1). Let $\mathbf{A}$ and $\mathbf{B}$ be $m \times n$ matrices. Then, in the consistent model $\mathscr{M}$, the estimators $\mathbf{Ay}$ and $\mathbf{By}$ are said to be equal with probability 1 if

$$\mathbf{Ay} = \mathbf{By} \quad \text{for all } \mathbf{y} \in \mathscr{C}(\mathbf{X} : \mathbf{V}), \quad \text{i.e.,} \quad \mathbf{A}(\mathbf{X} : \mathbf{V}) = \mathbf{B}(\mathbf{X} : \mathbf{V}), \tag{25}$$

which further can be written as $\mathbf{A}(\mathbf{X} : \mathbf{VM}) = \mathbf{B}(\mathbf{X} : \mathbf{VM})$. Often we drop off the phrase "with probability 1". In (23) the vector $\mathbf{y}$ varies through $\mathscr{C}(\mathbf{X} : \mathbf{V}) = \mathbb{R}^n$ and thus (23) becomes the equality between the multipliers $\mathbf{H}$ and $\mathbf{P}_{\mathbf{X},\mathbf{V}^{-1}}$. For the consistency of the linear model, see, e.g., Rao (1973a, p. 297), and Baksalary et al. (1992).

We will use short notations like $\mathbf{Ay} = \mathrm{BLUE}(\boldsymbol{\mu} \mid \mathscr{M}) = \tilde{\boldsymbol{\mu}} = \tilde{\boldsymbol{\mu}}(\mathscr{M})$. Thus the equality $\mathrm{OLSE}(\boldsymbol{\mu}) = \mathrm{BLUE}(\boldsymbol{\mu})$, i.e., $\hat{\boldsymbol{\mu}} = \tilde{\boldsymbol{\mu}}$ means that

$$\mathbf{Hy} = \mathbf{Ay} \quad \text{for all } \mathbf{y} \in \mathscr{C}(\mathbf{X} : \mathbf{V}), \tag{26}$$

where $\mathbf{A}$ is a matrix providing the $\tilde{\boldsymbol{\mu}}$. There is an infinite number of such matrices $\mathbf{A}$ when $\operatorname{rank}(\mathbf{X} : \mathbf{V}) < n$, but under a consistent model the realized value of $\mathbf{Ay}$ is unique.

**Example 2.** [Equality of OLSE and $\mathbf{P}_{\mathbf{X};\mathbf{V}^{+}}\mathbf{y}$.] Denoting $\mathbf{P}_{\mathbf{X};\mathbf{V}^{+}} = \mathbf{X}(\mathbf{X}'\mathbf{V}^{+}\mathbf{X})^{+}\mathbf{X}'\mathbf{V}^{+}$, one might be curious to know under which condition $\mathbf{P}_{\mathbf{X};\mathbf{V}^{+}}\mathbf{y}$ equals $\hat{\boldsymbol{\mu}}$ (w.p. 1). This happens if and only if $\mathbf{X}(\mathbf{X}'\mathbf{V}^{+}\mathbf{X})^{+}\mathbf{X}'\mathbf{V}^{+}(\mathbf{X} : \mathbf{V}) = \mathbf{H}(\mathbf{X} : \mathbf{V})$, i.e.,

$$\text{(i)}\ \mathbf{X}(\mathbf{X}'\mathbf{V}^{+}\mathbf{X})^{+}\mathbf{X}'\mathbf{V}^{+}\mathbf{X} = \mathbf{X}, \quad \text{and} \quad \text{(ii)}\ \mathbf{X}(\mathbf{X}'\mathbf{V}^{+}\mathbf{X})^{+}\,\mathbf{X}'\mathbf{P}_{\mathbf{V}} = \mathbf{HV}. \tag{27}$$

Postmultiplying (ii) in (27) by $\mathbf{V}^{+}\mathbf{X}$ and using (i) yields $\mathbf{X} = \mathbf{HP}_{\mathbf{V}}\mathbf{X}$ and thereby $\mathbf{X}'\mathbf{X} = \mathbf{X}'\mathbf{P}_{\mathbf{V}}\mathbf{X}$, i.e., $\mathbf{X}'\mathbf{Q}_{\mathbf{V}}\mathbf{X} = \mathbf{0}$, so that $\mathscr{C}(\mathbf{X}) \subseteq \mathscr{C}(\mathbf{V})$. This is the result obtained by Baksalary & Kala (1983). It becomes particularly interesting if we utilize the following fact, see Zyskind & Martin (1969) and Mitra & Rao (1968, p. 286):

$$\mathbf{X}(\mathbf{X}'\mathbf{V}^{+}\mathbf{X})^{+}\mathbf{X}'\mathbf{V}^{+}\mathbf{y} = \mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta}) \iff \mathscr{C}(\mathbf{X}) \subseteq \mathscr{C}(\mathbf{V}). \tag{28}$$

This means that $\mathbf{P}_{\mathbf{X};\mathbf{V}^{+}}\mathbf{y}$ equals $\hat{\boldsymbol{\mu}}$ only if they equal $\tilde{\boldsymbol{\mu}}$. The model where $\mathscr{C}(\mathbf{X}) \subseteq \mathscr{C}(\mathbf{V})$, is often called a *weakly singular* linear model. Actually then $\mathbf{P}_{\mathbf{X};\mathbf{V}^{+}}\mathbf{y}$ is invariant for any choice of generalized inverses involved. □

## 3 Year 1967: a good one for the OLSE = BLUE

As noted by Puntanen & Styan (1989, p. 154), the first condition for the equality between OLSE and BLUE of $\mathbf{X}\boldsymbol{\beta}$ was obtained by Anderson (1948, p. 92):

> Let $\mathbf{X}$ and $\mathbf{V}$ have full rank. If the $p$ columns of the $n \times p$ matrix $\mathbf{X}$ are linear combinations of $p$ of the eigenvectors of $\mathbf{V}$, then OLSE is BLUE.

Anderson's result was published in *Skandinavisk Aktuarietidskrift* (from 1973: *Scandinavian Actuarial Journal*), and as Anderson says in his interview in *Statistical Science* (DeGroot, 1986, p. 102): "As a result it did not get a great deal of attention . . . So from time to time people discover that paper."

That Anderson's condition is also necessary may be deduced from results obtained by Watson (1951, 1955). Watson was discussing the efficiency of OLSE showing the necessity of Anderson's condition for $p = 1$.

Magness & McGuire (1962) appear to be the first to show that this condition is both necessary and sufficient, though Anderson (1972, p. 472) mentioned that sufficiency was essentially given in his 1948 paper. Interestingly, Magness & McGuire (1963) published the following "Acknowledgment of priority":

> Theorem 2 of the authors' paper 1962 is a special case of Equation (3.5) of Watson (1955). Also, the fact that least squares and minimum variance estimates are equally efficient when the regression vectors are eigenvectors of the noise covariance matrix is apparently known and is referred to by Watson. The authors regret having overlooked Professor Watson's outstanding prior contribution.

It seems that Zyskind, in an invited paper presented at the 1962 Institute of Mathematical Statistics Annual Meeting, was the first author to consider the equality of the OLSE and BLUE when $\mathbf{X}$ has rank less than $p$, see Zyskind (1962) which is an abstract of his talk. The covariance matrix was still assumed positive definite.

Goldman & Zelen (1964) allowed the covariance matrix $\mathbf{V}$ to be possibly singular; they obtained a similar eigenvector condition to that of Anderson (1948), namely $\mathscr{C}(\mathbf{X}) = \mathscr{C}(\mathbf{T}_{[r]})$ where $\mathbf{T}_{[r]}$ is an $n \times r$ matrix whose columns are the $r$ eigenvectors corresponding to $r$ nonzero eigenvalues of $\mathbf{V}$ with $r = \operatorname{rank}(\mathbf{X})$. As shown later by Zyskind (1967, p. 1098) the nonzero requirement is not needed.

Rao (1967, 1968) appears to be the first to provide further necessary and sufficient conditions for the general case when both $\mathbf{X}$ and $\mathbf{V}$ are possibly deficient in rank. In 1965 at the Fifth Berkeley Symposium, Rao (1967, p. 364) presented the following two conditions, each of which is both necessary and sufficient for the equality of the OLSE($\mathbf{X}\boldsymbol{\beta}$) and the BLUE($\mathbf{X}\boldsymbol{\beta}$):

$$\text{(i)}\;\; \mathbf{X}'\mathbf{V}\mathbf{X}^{\perp} = \mathbf{0}, \;\text{i.e.,}\; \mathbf{HVM} = \mathbf{0}, \quad \text{(ii)}\;\; \mathbf{V} = \alpha\mathbf{I}_n + \mathbf{XAX}' + \mathbf{X}^{\perp}\mathbf{B}(\mathbf{X}^{\perp})', \tag{29}$$

for some scalar $\alpha$ and some symmetric $\mathbf{A}$ and $\mathbf{B}$ so that $\mathbf{V}$ is nonnegative definite. It is clear that (i) is equivalent to $\mathscr{C}(\mathbf{VX}) \subseteq \mathscr{C}(\mathbf{X})$, which becomes equality if $\mathbf{V}$ is positive definite. Rao (1968, p. 68) emphasized that "the basis of the proof is the following: the necessary and sufficient condition that a statistic is a minimum variance unbiased estimator is that it has zero covariance with statistics whose expectation is identically zero (Rao, 1965, pp. 185, 257) [referring to the first edition of Rao (1973a)]."

Notice that the case (ii) of (29), which is sometimes called "Rao's structure", occurs in the mixed linear model, where $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Xu} + \mathbf{X}^{\perp}\mathbf{v} + \mathbf{e}$, and $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{e}$ are uncorrelated random vectors with zero expectations and covariance matrices $\boldsymbol{\Gamma}$, $\boldsymbol{\Theta}$, and $\sigma^2\mathbf{I}_n$, respectively. See also Mitra & Rao (1969) for properties of specific structures of $\mathbf{V}$.

Zyskind (1967, Th. 2) who referred to Rao (1967), gave, without any restrictive rank assumptions, eight necessary and sufficient conditions under which OLSE is BLUE, thus extending the results he gave in 1962. Kempthorne (1975), in the obituary for Zyskind (1929–1974), writes:

> Zyskind's interest in the method of least squares led to a rather remarkable set of papers. . . . A subsequent (1967) paper laid out the bulk of the story with respect to equality of OLSE and BLUE. . . . The importance of this whole line of work is underscored by the occurrence of related work by W.H. Kruskal, C.R. Rao and G.S. Watson, as well as others.

Because Rao gave his paper in 1965 and Zyskind (1967) refers to it, we credit Rao as the first author to have established $\mathbf{HVM} = \mathbf{0}$; see Baksalary & Kala (1983, pp. 119 and 240) and Kempthorne (1976, p. 217). In his Acknowledgements Zyskind (1967) writes: "I wish to thank Professor C.R. Rao for permitting me to see a copy of the final version of his manuscript (1967) before its publication."

We might cite an interesting piece from Rao's interview by DeGroot (1987, p. 59):

> Rao: There was another incident recently in which somebody claimed priority because he had mentioned a result slightly less general than mine in an abstract in the *Annals*. You can say anything in an abstract. If it is right, you can claim credit and priority.
>
> DeGroot: Yes. Take a chance; maybe it will be right. There is no serious screening of abstracts. I think that's OK, as long as everyone realizes that the results are not necessarily correct or original.
>
> Rao: Actually when that person wrote the full paper on the basis of the abstract, I was a referee and it turned out that this result was also not correct as stated.

Rao (1968, p. 259) further commented on Zyskind (1967) and Watson (1967) in the following way (in our notation):

> In my previous paper 1967, I gave details of the proof of (29) when $\mathbf{V}$ and $\mathbf{X}'\mathbf{X}$ are nonsingular and mentioned that the same proof holds more generally for singular $\mathbf{V}$ and $\mathbf{X}'\mathbf{X}$. I omitted the details in the latter as the extension was extremely simple, and not relevant to the main theme of the paper.
>
> In a recent paper, Zyskind (1967) thought that there may be some difficulty in proving the condition (29) in its widest generality when $\mathbf{V}$ is singular. Watson (1967) writes that, "Rao (1965) [referring to the Fifth Berkeley Symposium] remarks that his result is true for $\mathbf{V}$ singular and rank of $\mathbf{X}$ is below $p$. Some skill with generalized inverses might show the proof is still valid." In view of these remarks and other statements it seems necessary to elaborate the earlier proof.

An early description of the coordinate-free approach to linear models was made by Kruskal (1961, 1968). Watson (1967, p. 1682) wrote:

> In some 1962 correspondence with Dr. M.E. Muller and the author, Professor W. Kruskal indicated a coordinate-free proof of the necessity and sufficiency when $\mathbf{X}$, but not $\mathbf{V}$, is possibly not of full rank. This result is particularly simple to prove because, instead of working with $\hat{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\beta}}$ he uses $\hat{\boldsymbol{\mu}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\mu}} = \mathbf{X}\tilde{\boldsymbol{\beta}}$. He states that "$\hat{\boldsymbol{\mu}} = \tilde{\boldsymbol{\mu}}$ if and only if $\mathscr{C}(\mathbf{VX}) = \mathscr{C}(\mathbf{X})$". The author hopes that Professor Kruskal's result will appear in the near future.

In an interview by Zabell (1994, p. 294), Kruskal mentions, referring to his 1968 paper: "That started out as an exercise, an exam exercise in the course I was giving, and then Geoff Watson came along with much the same material; he encouraged me to try for publication."

Herr (1980, p. 46), in an interesting article "On the history of the use of geometry in the general linear model", commented on various approaches to handle linear models. About Kruskal he writes the following:

> These two papers. Kruskal (1968, 1975), are elegant examples of the analytic geometric approach to linear models. In Kruskal (1968), the question of equality of simple least squares and best linear unbiased estimates, which was considered in Zyskind (1967) and Watson (1967), is treated using a coordinate-free approach. The comparison of the parts of the three papers dealing with this question is very instructive. The simplicity and beauty of the coordinate-free approach is clearly demonstrated by such a comparison.
>
> In Kruskal (1975), an analytic geometric approach is used with such skill and grace that the paper ought to be required reading for anyone who might be tempted to deal with generalized inverses.

Eaton, also a great promoter of the coordinate-free approach, see his papers (1970; 1978), wrote in (2007, p. 265):

> The direct effect of Kruskal (1968), a marvelous paper, is relatively easy to describe. In coordinate-free language, here is a statement of the main result of that paper:
> The Gauss–Markov and least squares estimators are the same if and only if the linear manifold of the mean vector is an invariant subspace of the covariance.
> [$\mathscr{C}(\mathbf{VX}) \subseteq \mathscr{C}(\mathbf{X})$, in our notation.]

Anderson (1971, p. 563) gave a quite different rank criterion for the equality of OLSE and BLUE in the form of rank-additivity, assuming $\mathbf{V}$ and $\mathbf{X}$ be of full rank. George Styan (1973) extended this criterion by removing the restriction on the rank of $\mathbf{X}$. Below is George's description (in our notation) on this development, see O.M. Baksalary & Styan (2005, p. 16).

> I think that the first paper by Jerzy Baksalary I read was Baksalary & Kala (1977), which I reviewed for *Mathematical Reviews*. In that paper it is shown that in the linear model $\{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$ the best linear unbiased estimator of $\mathbf{X}\boldsymbol{\beta}$ equals the ordinary least-squares estimator if and only if
>
> $$\operatorname{rank}(\mathbf{X}'\mathbf{T}_1) + \operatorname{rank}(\mathbf{X}'\mathbf{T}_2) \cdots + \operatorname{rank}(\mathbf{X}'\mathbf{T}_s) = \operatorname{rank}(\mathbf{X}), \tag{30}$$
>
> where $\mathbf{V}$ has $s$ distinct eigenvalues and $\mathbf{T}_1, \mathbf{T}_2, \ldots, \mathbf{T}_s$ are matrices of corresponding orthonormalized eigenvectors. Here $\mathbf{X}$ can be less than full column rank and $\mathbf{V}$ may be singular. The result for $\mathbf{X}$ possibly of less than full column rank but with $\mathbf{V}$ positive definite was established by me in 1973, extending the earlier result with $\mathbf{X}$ of full column rank and $\mathbf{V}$ positive definite due to Anderson (1971, p. 561).

> The paper by Baksalary & Kala (1977) prompted me to read further papers by Baksalary and Kala, and [. . . *old reminiscences* . . . ] on 27 August 1980 both Jerzy Baksalary and Radosĺaw Kala met me at the main railway station in Poznań.

A very different modification of the problem of when the OLSE is the BLUE originates from McElroy (1967). We present it here in a generalized version due to Zyskind (1969) and Baksalary & van Eijnsbergen (1988): Given a matrix $\mathbf{U}$, such that $\operatorname{rank}(\mathbf{U}) \leq n - 1$, when does $\hat{\boldsymbol{\mu}} = \tilde{\boldsymbol{\mu}}$ hold for every model matrix $\mathbf{X}$ satisfying $\mathscr{C}(\mathbf{U}) \subseteq \mathscr{C}(\mathbf{X})$.

We have now more or less covered the development of the OLSE vs BLUE saga up till the end of the 1970s, having the role of Professor Rao in mind. There is a lot of interesting literature after that (as well as before) that we have no space to discuss. However, we might wish to mention

- Gouriéroux & Monfort (1980), Baltagi (1989), McAleer (1992) and Larocca (2005) providing econometric examples and references;
- Chapter 8 of Rao & Mitra (1971a), and Mathew & Bhimasankaram (1983a,b) reviewing conditions for optimality and validity of least-squares theory;
- Baksalary & Kala (1978, 1980) and Haberman (1975) who studied the Euclidean distance between OLSE and BLUE.

We may also mention Yongge Tian who in numerous papers has studied OLSE vs BLUE matters using so-called matrix rank methods, see, e.g., Tian (2013), Tian & Zhang (2016), and Puntanen, Styan & Tian (2005).

The model $\mathscr{M}$ can be extended to the case when we wish to predict a "new future" value of $\mathbf{y}_*$, assumed to be coming from $\mathbf{y}_* = \mathbf{X}_*\boldsymbol{\beta} + \boldsymbol{\varepsilon}_*$, where $\mathbf{X}_*$ is a known $q \times p$ matrix and $\boldsymbol{\varepsilon}_*$ is a $q$-dimensional random error vector. We assume that $\operatorname{cov}(\mathbf{y}, \mathbf{y}_*) = \mathbf{V}_{12}$ is known. For conditions of $\mathbf{Ay}$ being the best linear unbiased predictor, BLUP, for $\mathbf{y}_*$, minimizing the covariance matrix of the prediction error, see, e.g., Goldberger (1962), Christensen (2011, p. 294), and Isotalo & Puntanen (2006, p. 1015). For relations between OLSE, BLUE and BLUP, see, e.g., Watson (1972), Baksalary & Kala (1981b) and Haslett et al. (2014).

We feel it appropriate, though not fully related to OLSE vs BLUE, to complete this section (where the main year was 1967), by mentioning that on 5 April 1967, C.R. Rao left Calcutta for London and attended the induction ceremony to the Fellowship of the Royal Society, on 6 April 1967. This prompted Professor P.C. Mahalanobis to use the following words in his speech on 12 February 1968, see Mahalanobis (1969, p. 239):

> I should now like to say, briefly, how proud I feel that C.R. Rao was elected a Fellow of the Royal Society last year. He came to the Institute in January 1941 to learn statistics. I feel proud that my direct pupil is now in the Royal Society. In India, we have a saying *Putrat ichhyet parajayam*. One wishes for defeat by his son. I have no children. In India there is also the alternative version *Sishyat ichhyet parajayam*. One wishes for defeat by his pupil. It is my great happiness to admit defeat by my pupil.

## 4 OLSE = BLUE: conditions

**Theorem 3** *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 of the following equivalent conditions holds.* (*Note:* $\mathbf{V}$ *is replaceable with* $\mathbf{V}^{+}$ *and* $\mathbf{H}$ *and* $\mathbf{M}$ *can be interchanged.*)

(i) $\mathbf{HV} = \mathbf{VH}$, (ii) $\mathbf{HV} = \mathbf{HVH}$, (iii) $\mathbf{HVM} = \mathbf{0}$,
(ii) $\mathbf{X}'\mathbf{V}\mathbf{X}^{\perp} = \mathbf{0}$, (v) $\mathscr{C}(\mathbf{VX}) \subseteq \mathscr{C}(\mathbf{X})$, (vi) $\mathscr{C}(\mathbf{VX}) = \mathscr{C}(\mathbf{X}) \cap \mathscr{C}(\mathbf{V})$,
(vii) $\mathbf{HVH} \leq_{\mathrm{L}} \mathbf{V}$, *i.e.,* $\mathbf{V} - \mathbf{HVH}$ *is nonnegative definite,*
(viii) $\mathbf{HVH} \leq_{\mathrm{rs}} \mathbf{V}$, *i.e.,* $\mathrm{rank}(\mathbf{V} - \mathbf{HVH}) = \mathrm{rank}(\mathbf{V}) - \mathrm{rank}(\mathbf{HVH})$, *i.e.,* $\mathbf{HVH}$ *and* $\mathbf{V}$ *are rank-subtractive,*
(ix) $\mathscr{C}(\mathbf{X})$ *has a basis consisting of r eigenvectors of* $\mathbf{V}$, *where* $r = \mathrm{rank}(\mathbf{X})$,
(x) $\mathrm{rank}(\mathbf{T}_1'\mathbf{X}) + \cdots + \mathrm{rank}(\mathbf{T}_s'\mathbf{X}) = \mathrm{rank}(\mathbf{X})$, *where* $\mathbf{T}_i$ *is a matrix consisting of the orthogonal eigenvectors w.r.t. the* i*th largest eigenvalue* $\lambda_{(i)}$ *of* $\mathbf{V}$; $\lambda_{(1)} > \cdots > \lambda_{(s)}$,
(xi) $\mathbf{T}_i'\mathbf{H}\mathbf{T}_i = (\mathbf{T}_i'\mathbf{H}\mathbf{T}_i)^2$ *for all* $i = 1, 2, \ldots, s$,
(xii) $\mathbf{T}_i'\mathbf{H}\mathbf{T}_j = \mathbf{0}$ *for all* $i, j = 1, 2, \ldots, s$, $i \neq j$,
(xiii) $\mathscr{C}(\mathbf{T}_i) = \mathscr{C}(\mathbf{T}_i) \cap \mathscr{C}(\mathbf{X}) \oplus \mathscr{C}(\mathbf{T}_i) \cap \mathscr{C}(\mathbf{M})$ *for all* $i = 1, \ldots, s$,
(xiv) *the squared nonzero canonical correlations between* $\mathbf{y}$ *and* $\mathbf{Hy}$ *are the nonzero eigenvalues of* $\mathbf{V}^{-}\mathbf{HVH}$ *for all* $\mathbf{V}^{-}$,
(xv) $\mathbf{V} \in \mathcal{V}_1 = \{\, \mathbf{V} \geq_{\mathrm{L}} \mathbf{0} : \mathbf{V} = \mathbf{HAH} + \mathbf{MBM},\ \mathbf{A} \geq_{\mathrm{L}} \mathbf{0},\ \mathbf{B} \geq_{\mathrm{L}} \mathbf{0} \,\}$,
(xvi) $\mathbf{V} \in \mathcal{V}_2 = \{\, \mathbf{V} \geq_{\mathrm{L}} \mathbf{0} : \mathbf{V} = \mathbf{XCX}' + \mathbf{X}^{\perp}\mathbf{D}(\mathbf{X}^{\perp})',\ \mathbf{C} \geq_{\mathrm{L}} \mathbf{0},\ \mathbf{D} \geq_{\mathrm{L}} \mathbf{0} \,\}$,
(xvii) $\mathbf{V} \in \mathcal{V}_3 = \{\, \mathbf{V} \geq_{\mathrm{L}} \mathbf{0} : \mathbf{V} = \alpha\mathbf{I} + \mathbf{XKX}' + \mathbf{X}^{\perp}\mathbf{L}(\mathbf{X}^{\perp})',\ \alpha \in \mathbb{R},\ \mathbf{K} = \mathbf{K}',\ \mathbf{L} = \mathbf{L}' \,\}$.

Some sources for the above statements are given in Section 3. For collections of the proofs, see Alalouf & Styan (1984) and Puntanen et al. (2011, Ch. 10). Notice the somewhat peculiar statements (vii), (viii) and (xiv); they appear in Baksalary & Puntanen (1989, 1990a,b). Some further conditions are given by O.M. Baksalary et al. (2013, Th. 5).

**Example 3.** [Centering the model.] Consider the partitioned linear model

$$\mathscr{M}_{12} = \{\mathbf{y},\ \mathbf{1}\alpha + \mathbf{X}_0\boldsymbol{\beta}_{\mathbf{x}},\ \mathbf{I}_n\}, \quad \text{where } \mathbf{X}_0 \in \mathbb{R}^{n\times k}, \tag{31}$$

and $\mathbf{1} \in \mathbb{R}^n$ is a vector of ones. Assume that $\mathbf{X} = (\mathbf{1} : \mathbf{X}_0)$ has full column rank. Premultiplying $\mathscr{M}_{12}$ by the centering matrix $\mathbf{C} = \mathbf{I}_n - \mathbf{P}_{\mathbf{1}}$ yields the centered model

$$\mathscr{M}_{12\cdot 1} = \{\mathbf{Cy},\ \mathbf{CX}_0\boldsymbol{\beta}_{\mathbf{x}},\ \mathbf{C}\}. \tag{32}$$

In this centered model we have a singular covariance matrix and hence it may seem that finding a BLUE would be problematic. However, corresponding to condition (v) of Theorem 3 we have now $\mathscr{C}(\mathbf{C}\cdot\mathbf{CX}_0) \subseteq \mathscr{C}(\mathbf{CX}_0)$ and thus

$$\mathrm{BLUE}(\boldsymbol{\beta}_{\mathbf{x}} \mid \mathscr{M}_{12\cdot 1}) = \mathrm{OLSE}(\boldsymbol{\beta}_{\mathbf{x}} \mid \mathscr{M}_{12\cdot 1}) = (\mathbf{X}_0'\mathbf{CX}_0)^{-1}\mathbf{X}_0'\mathbf{Cy} := \hat{\boldsymbol{\beta}}_{\mathbf{x}}. \tag{33}$$

On the other hand, it is standard textbook material that $\hat{\boldsymbol{\beta}}_{\mathbf{x}}$ is the BLUE for $\boldsymbol{\beta}_{\mathbf{x}}$ in the partitioned model $\mathscr{M}_{12}$. Thus centering has no effect on the BLUE of $\boldsymbol{\beta}_{\mathbf{x}}$, and so $\mathbf{Cy}$

is a linearly sufficient statistic for $\boldsymbol{\beta}_{\mathbf{x}}$ in $\mathscr{M}_{12}$. Centering is a simple example of the model reduction, i.e., premultiplying $\mathscr{M}_{12} = \{\mathbf{y}, \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2, \mathbf{V}\}$ by $\mathbf{M}_1 = \mathbf{I}_n - \mathbf{P}_{\mathbf{X}_1}$, yielding to

$$\mathscr{M}_{12\cdot 1} = \{\mathbf{M}_1\mathbf{y},\ \mathbf{M}_1\mathbf{X}_2\boldsymbol{\beta}_2,\ \mathbf{M}_1\mathbf{V}\mathbf{M}_1\}\,, \tag{34}$$

see, e.g., Groß & Puntanen (2000) and Chu et al. (2004, 2005). □

**Example 4.** [Intraclass correlation.] Consider $\mathbf{V}$ which has the intraclass correlation structure (of which the centering matrix $\mathbf{C}$ is an example), that is, $\mathbf{V}$ is of the type $\mathbf{V} = (1-\varrho)\mathbf{I}_n + \varrho\mathbf{1}\mathbf{1}'$, where $-\frac{1}{n-1} \leq \varrho \leq 1$. In this situation $\mathbf{H}\mathbf{V} = (1-\varrho)\mathbf{H} + \varrho\mathbf{H}\mathbf{1}\mathbf{1}'$, and thereby $\mathbf{H}\mathbf{V} = \mathbf{V}\mathbf{H}$ if and only if

$$\mathbf{H}\mathbf{1}\mathbf{1}' = \mathbf{1}\mathbf{1}'\mathbf{H}\,. \tag{35}$$

We can conclude that (35) holds if and only if $\mathbf{1}$ is an eigenvector of $\mathbf{H}$, i.e., $\mathbf{H}\mathbf{1} = \lambda\mathbf{1}$. The eigenvalues of $\mathbf{H}$ are 0 and 1, with multiplicities $n - \operatorname{rank}(\mathbf{X})$ and $\operatorname{rank}(\mathbf{X})$, respectively. Hence (35) holds, i.e., OLSE = BLUE, if and only if $\mathbf{1} \in \mathscr{C}(\mathbf{X})$ or $\mathbf{1} \in \mathscr{C}(\mathbf{X})^{\perp}$. □

## 5 The fundamental BLUE equation

Theorem 4 below provides so-called fundamental BLUE equations.

**Theorem 4** *Consider the linear model* $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$. *Then the linear estimator* $\mathbf{G}\mathbf{y}$ *is the* BLUE *for* $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ *if and only if* $\mathbf{G} \in \mathbb{R}^{n\times n}$ *satisfies the equation*

$$\mathbf{G}(\mathbf{X} : \mathbf{V}\mathbf{X}^{\perp}) = (\mathbf{X} : \mathbf{0})\,. \tag{36}$$

*Moreover, let* $\mathbf{K}\boldsymbol{\beta}$, *where* $\mathbf{K} \in \mathbb{R}^{q\times p}$, *be estimable so that* $\mathscr{C}(\mathbf{K}') \subseteq \mathscr{C}(\mathbf{X}')$. *Then* $\mathbf{B}\mathbf{y}$ *is the* BLUE *of* $\mathbf{K}\boldsymbol{\beta}$ *if and only if* $\mathbf{B} \in \mathbb{R}^{q\times n}$ *satisfies the equation*

$$\mathbf{B}(\mathbf{X} : \mathbf{V}\mathbf{X}^{\perp}) = (\mathbf{K} : \mathbf{0})\,. \tag{37}$$

For the proofs, see, e.g., (Rao, 1973b, p. 282) and for coordinate-free approach (Drygas, 1970, p. 55) and Zmyślony (1980). For further proofs see, for example, Groß (2004), Kala (1981, Th. 3.1), Puntanen, Styan & Werner (2011), and Puntanen et al. (2011, Th. 10). Baksalary (2004) provides a proof which he describes as follows: "From the algebraic point of view, the present development seems to be the simplest from among all accessible in the literature till now".

Of course, in (37) and (36) we can replace $\mathbf{X}^{\perp}$ with $\mathbf{M} = \mathbf{I}_n - \mathbf{P}_{\mathbf{X}}$. Equation (36) is always solvable for $\mathbf{G}$ while (37) is solvable whenever $\mathbf{K}\boldsymbol{\beta}$ is estimable. Solutions are unique if and only if $\operatorname{rank}(\mathbf{X} : \mathbf{V}) = n$. The solution for $\mathbf{G}$ satisfying $\mathbf{G}(\mathbf{X} : \mathbf{V}\mathbf{M}) = (\mathbf{X} : \mathbf{0})$ can be expressed, for example, in the following ways:

$$\mathbf{G}_1 = (\mathbf{X} : \mathbf{0})(\mathbf{X} : \mathbf{VM})^-, \qquad \mathbf{G}_2 = \mathbf{I}_n - \mathbf{VM}(\mathbf{MVM})^-\mathbf{M}, \tag{38a}$$

$$\mathbf{G}_3 = \mathbf{H} - \mathbf{HVM}(\mathbf{MVM})^-\mathbf{M}, \quad \mathbf{G}_4 = \mathbf{X}(\mathbf{X}'\mathbf{W}^-\mathbf{X})^-\mathbf{X}'\mathbf{W}^-, \tag{38b}$$

where $\mathbf{W}$ belongs to the class of matrices

$$\mathcal{W} = \{\mathbf{W} \in \mathbb{R}^{n\times n} : \mathbf{W} = \mathbf{V} + \mathbf{XUX}',\ \mathscr{C}(\mathbf{W}) = \mathscr{C}(\mathbf{X} : \mathbf{V})\}. \tag{39}$$

In (39) $\mathbf{U}$ can be any $p \times p$ matrix as long as $\mathscr{C}(\mathbf{W}) = \mathscr{C}(\mathbf{X} : \mathbf{V})$ is satisfied; see, e.g., Baksalary et al. (1990a, Th. 2), Baksalary & Mathew (1990, Th. 2) and Harville (1997, p. 468). The *general* solution to (36) can be expressed as $\mathbf{G}_i + \mathbf{N}_i\mathbf{Q}_\mathbf{W}$ where $\mathbf{N}_i \in \mathbb{R}^{n\times n}$ are free to vary. For the relations between the representations of the BLUEs, see, e.g., Albert (1973), Rao (1978, 1979), Rao & Mitra (1971b) and Searle (1994).

The covariance matrix of the $\tilde{\boldsymbol{\mu}} = \mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta})$ can be expressed as

$$\begin{aligned}\mathrm{cov}(\tilde{\boldsymbol{\mu}}) &= \mathbf{HVH} - \mathbf{HVM}(\mathbf{MVM})^-\mathbf{MVH} = \mathbf{V} - \mathbf{VM}(\mathbf{MVM})^-\mathbf{MV} \\ &= \mathbf{X}(\mathbf{X}'\mathbf{W}^-\mathbf{X})^-\mathbf{X}' - \mathbf{XUX}',\end{aligned} \tag{40}$$

where $\mathbf{W} = \mathbf{V} + \mathbf{XUX}' \in \mathcal{W}$; see, e.g., Baksalary et al. (1990a) and Isotalo et al. (2008a,b). Notice that

$$\mathrm{cov}(\hat{\boldsymbol{\mu}} - \tilde{\boldsymbol{\mu}}) = \mathrm{cov}(\hat{\boldsymbol{\mu}}) - \mathrm{cov}(\tilde{\boldsymbol{\mu}}) = \mathbf{HVM}(\mathbf{MVM})^-\mathbf{MVH}. \tag{41}$$

Corresponding to (22), the weighted sum of squares of errors in the general case is

$$\mathrm{SSE}(\mathbf{W}) = (\mathbf{y} - \tilde{\boldsymbol{\mu}})'\mathbf{V}^-(\mathbf{y} - \tilde{\boldsymbol{\mu}}) = \mathbf{y}'\dot{\mathbf{M}}\mathbf{y}, \quad \text{where } \dot{\mathbf{M}} = \mathbf{M}(\mathbf{MVM})^-\mathbf{M}. \tag{42}$$

Suppose that we have two models $\mathscr{A}_1 = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}_1\}$ and $\mathscr{A}_2 = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}_2\}$, which have different covariance matrices. Then we can ask, for example, what is needed that every representation of the BLUE of $\mathbf{X}\boldsymbol{\beta}$ under $\mathscr{A}_1$ remains BLUE under $\mathscr{A}_2$. Mitra & Moore (1973) give a very clear description of the different problems occurring. Let $\mathbf{G}$ be such a matrix that $\mathbf{Gy}$ is the BLUE for $\mathbf{X}\boldsymbol{\beta}$ under $\mathscr{A}_1$, Then we say that $\mathbf{Gy}$ remains BLUE under $\mathscr{A}_2$ if the following implication holds:

$$\mathbf{G}(\mathbf{X} : \mathbf{V}_1\mathbf{M}) = (\mathbf{X} : \mathbf{0}) \implies \mathbf{G}(\mathbf{X} : \mathbf{V}_2\mathbf{M}) = (\mathbf{X} : \mathbf{0}). \tag{43}$$

It appears that every representation of the BLUE for $\mathbf{X}\boldsymbol{\beta}$ under $\mathscr{A}_1$ remains BLUE under $\mathscr{A}_2$ and only if any of the following equivalent conditions hold:

$$\text{(i) } \mathscr{C}(\mathbf{V}_2\mathbf{M}) \subseteq \mathscr{C}(\mathbf{V}_1\mathbf{M}), \quad \text{(ii) } \mathbf{V}_2 = \alpha\mathbf{V}_1 + \mathbf{XAX}' + \mathbf{V}_1\mathbf{MBMV}_1, \tag{44}$$

for some $\alpha \in \mathbb{R}$, and $\mathbf{A}$ and $\mathbf{B}$ such that $\mathbf{V}_2$ is nonnegative definite. It is clear that even if (44) holds, the covariance matrices of $\mathbf{Gy}$ under $\mathscr{A}_1$ and $\mathscr{A}_2$ may be different; see, e.g., Rao & Mitra (1971b, Ch. 8). For the proof of (44) and related discussion, see, e.g., Mitra & Moore (1973, Th. 4.1–4.2), and Rao (1968, Lemma 5) and Rao (1971, Th. 5.2, Th. 5.5). For a special note on the interpretation of (i) for $\mathbf{V}_2 = \mathbf{I}_n$, see Markiewicz et al. (2010).

## 6 The relative efficiency of OLSE

In this section we follow closely Puntanen et al. (2011, Sec. 10.8) to take a brief look at the relative efficiency of OLSE with respect to the BLUE. Consider the linear model $\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$, where $\mathbf{X}$ has full column rank and $\mathbf{V}$ is positive definite. Then by Theorem 2,

$$\operatorname{cov}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}, \quad \operatorname{cov}(\tilde{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}, \tag{45a}$$

$$\operatorname{cov}(\hat{\boldsymbol{\beta}}) - \operatorname{cov}(\tilde{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} := \mathbf{D}\,. \tag{45b}$$

The relative efficiency, so-called Watson efficiency, see Watson (1955, p. 330), of OLSE vs BLUE is defined as the ratio of the determinants of the covariance matrices:

$$\operatorname{eff}(\hat{\boldsymbol{\beta}}) = \frac{|\operatorname{cov}(\tilde{\boldsymbol{\beta}})|}{|\operatorname{cov}(\hat{\boldsymbol{\beta}})|} = \frac{|\mathbf{X}'\mathbf{X}|^2}{|\mathbf{X}'\mathbf{V}\mathbf{X}| \cdot |\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}|} = \frac{|\operatorname{cov}(\hat{\boldsymbol{\beta}}) - \mathbf{D}|}{|\operatorname{cov}(\hat{\boldsymbol{\beta}})|}\,. \tag{46}$$

We have $0 < \operatorname{eff}(\hat{\boldsymbol{\beta}}) \leq 1$, with $\operatorname{eff}(\hat{\boldsymbol{\beta}}) = 1$ if and only if $\hat{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}$. Moreover, the efficiency can be expressed as

$$\begin{aligned}\operatorname{eff}(\hat{\boldsymbol{\beta}}) &= |\mathbf{I}_p - \mathbf{X}'\mathbf{V}\mathbf{M}(\mathbf{M}\mathbf{V}\mathbf{M})^{-}\mathbf{M}\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}| = |[\operatorname{cov}(\hat{\boldsymbol{\beta}})]^{-1} \cdot \operatorname{cov}(\tilde{\boldsymbol{\beta}})| \\ &= (1-\kappa_1^2)\cdots(1-\kappa_p^2) = \theta_1^2 \cdots \theta_p^2\,,\end{aligned} \tag{47}$$

where $\kappa_1 \geq \kappa_2 \geq \cdots \geq \kappa_p \geq 0$ and $\theta_1 \geq \theta_2 \geq \cdots \geq \theta_p > 0$ are the canonical correlations between $\mathbf{X}'\mathbf{y}$ and $\mathbf{M}\mathbf{y}$, and $\hat{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\beta}}$, respectively. Notice that

$$\operatorname{cov}\begin{pmatrix}\hat{\boldsymbol{\beta}} \\ \tilde{\boldsymbol{\beta}}\end{pmatrix} = \begin{pmatrix}\operatorname{cov}(\hat{\boldsymbol{\beta}}) & \operatorname{cov}(\tilde{\boldsymbol{\beta}}) \\ \operatorname{cov}(\tilde{\boldsymbol{\beta}}) & \operatorname{cov}(\tilde{\boldsymbol{\beta}})\end{pmatrix} = \begin{pmatrix}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} & (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \\ (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} & (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\end{pmatrix}, \tag{48}$$

and thus the squared canonical correlations between $\hat{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\beta}}$ are the eigenvalues of the matrix product $[\operatorname{cov}(\hat{\boldsymbol{\beta}})]^{-1}\operatorname{cov}(\tilde{\boldsymbol{\beta}})[\operatorname{cov}(\tilde{\boldsymbol{\beta}})]^{-1}\operatorname{cov}(\tilde{\boldsymbol{\beta}})$:

$$\{\theta_1^2, \ldots, \theta_p^2\} = \operatorname{ch}\big[[\operatorname{cov}(\hat{\boldsymbol{\beta}})]^{-1}\operatorname{cov}(\tilde{\boldsymbol{\beta}})\big]\,, \tag{49}$$

where $\operatorname{ch}(\cdot)$ denotes the set of the eigenvalues of the matrix argument. On account of (49), we see, as claimed in (47), that indeed

$$\theta_1^2 \cdots \theta_p^2 = |[\operatorname{cov}(\hat{\boldsymbol{\beta}})]^{-1} \cdot \operatorname{cov}(\tilde{\boldsymbol{\beta}})|\,. \tag{50}$$

The efficiency formulas (47) in terms of $\kappa_i$'s and $\theta_i$'s were first introduced by Watson (1967, p. 1686) and by Bartmann & Bloomfield (1981), respectively. It can be shown that the nonzero canonical correlations between $\mathbf{X}'\mathbf{y}$ and $\mathbf{M}\mathbf{y}$ are the same as those between $\mathbf{H}\mathbf{y}$ and $\mathbf{M}\mathbf{y}$. For further references on the relative efficiency and canonical correlations, see, e.g., Chu et al. (2004, 2005) and Drury et al. (2002).

As regards the lower bound of the OLSE's efficiency, Bloomfield & Watson (1975) and Knott (1975) proved the following inequality:

$$\text{eff}(\hat{\boldsymbol{\beta}}) \geq \frac{4\lambda_1\lambda_n}{(\lambda_1+\lambda_n)^2} \cdot \frac{4\lambda_2\lambda_{n-1}}{(\lambda_2+\lambda_{n-1})^2} \cdots \frac{4\lambda_p\lambda_{n-p+1}}{(\lambda_p+\lambda_{n-p+1})^2} = \tau_1^2\,\tau_2^2 \cdots \tau_p^2\,, \tag{51}$$

where $\lambda_i = \text{ch}_i(\mathbf{V}) =$ $i$th largest eigenvalue and $\tau_i =$ $i$th *antieigenvalue* of $\mathbf{V}$; it is assumed that $p \leq n/2$. The concept of antieigenvalue was introduced by Gustafson (1972); see also Gustafson (2006, 2012) and Rao (2007).

Assuming that $p \leq n/2$, the minimum of $\phi$ is attained when $\mathbf{X}$ is chosen as $(\mathbf{t}_1 \pm \mathbf{t}_n : \mathbf{t}_2 \pm \mathbf{t}_{n-1} : \ldots : \mathbf{t}_p \pm \mathbf{t}_{n-p+1})$, where $\mathbf{t}_i$ are the orthonormal eigenvectors of $\mathbf{V}$ with respect to $\lambda_i$. The inequality (51) was originally conjectured in 1955 by Durbin (see Watson 1955, p. 331), but first established (for $p > 1$) only twenty years later by Bloomfield & Watson (1975) and Knott (1975). For further proofs (and related considerations), see Khatri & Rao (1981, 1982).

Another measure of efficiency of OLSE, introduced by Bloomfield & Watson (1975), is based on the Frobenius norm of the commutator $\mathbf{HV} - \mathbf{VH}$:

$$\psi = \tfrac{1}{2}\|\mathbf{HV} - \mathbf{VH}\|_F^2 = \tfrac{1}{2}\,\text{trace}(\mathbf{HV} - \mathbf{VH})(\mathbf{HV} - \mathbf{VH})' = \|\mathbf{HVM}\|_F^2\,. \tag{52}$$

Bloomfield & Watson (1975) proved that $\psi \leq \frac{1}{4}\sum_{i=1}^{p}(\lambda_i - \lambda_{n-i-1})^2$, and that the equality is attained in the same situation as the minimum of $\phi$.

Rao (1985a) studied the trace of the difference between the covariance matrices of the OLSE and BLUE of $\mathbf{X}\boldsymbol{\beta}$:

$$\eta = \text{trace}[\text{cov}(\mathbf{X}\hat{\boldsymbol{\beta}}) - \text{cov}(\mathbf{X}\tilde{\boldsymbol{\beta}})] = \text{trace}[\mathbf{HVH} - \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}']\,. \tag{53}$$

Styan (1983) considered (53) when $p = 1$ and Liski et al. (1992) considered the upper bound of the trace of $\mathbf{HVM}(\mathbf{MVM})^{-}\mathbf{MVH}$.

We can now conveniently complete this section with a somewhat curious remark. Namely, in the references of Rao (1985a, p. 255), it is said: “Simo Puntanen, Personal communication, 1982.” The story behind is that Simo indeed was communicating with Professor Rao about the upper bound of (53) explaining it to be a tough problem. So, Professor Rao conveniently decided to solve the problem.

## 7 Personal glimpses and conclusions

The matrix algebra related to matters like the equality of OLSE and BLUE is not apparently everybody’s cup of tea and some discussion in the literature appears to be rather critical. One of the most critical is by Kempthorne (1989) writing as his comment on Puntanen & Styan (1989): “I suggest that Zyskind and Rao gave the bulk of the story and that the flood of papers since their work has added only trivially, arcanely, and (usually) uselessly.” Sengupta & Jammalamadaka (2003, p. 311) agreed with Kempthorne’s criticism. Searle (1989), however, gives some supporting remarks on the importance of the OLSE vs BLUE topic and according to Harville (1990), “Puntanen and Styan’s (1989) article should be very useful to anyone with an interest in linear-model theory”. Baksalary (1988, p. 98) states that

"The importance of this problem is due to the fact that such conditions characterize which (unknown) dispersion structures can be ignored without consequence to best linear unbiased estimation."

In any event, we still find this area offering unexpected and interesting matrix problems. Over the years, in our teaching (not only in research) we have attempted to make our students familiar with the matrix algebra related to OLSE and BLUE matters: we have found it very educational. One evidence of our interest is the book *Formulas Useful for Linear Regression Analysis and Related Matrix Theory: It's Only Formulas But We Like Them*, by Puntanen, Styan & Isotalo (2013).

When DeGroot (1987, p. 60) asked Professor Rao for his favourite publications, part of the answer was the following:

> RAO: . . . A second set of papers I like are mostly in the analysis of repeated measurements and in singular linear models, i.e., when the design and covariance matrices are deficient in rank. I developed generalized inverses of matrices for dealing with such problems. . . .

There is some important interesting personal history between the authors and Professor Rao that we briefly wish to mention. First, while spending his sabbatical in Finland from September 1975 to August 1976, the third author, George, visited C.R. Rao in spring 1976 in New Delhi. He sent a postcard to Simo:

> En route, Calcutta – Bombay, 27 March 1976. Unbelievable that we've been gone six weeks already & will be back in two. Spent a hectic month in Delhi; wrote two papers. Hope C.R. Rao will visit Helsinki in mid-June. Relaxed for three days in Kathmandu, . . . Greetings, George.

C.R. Rao did not come to Finland in 1976 but indeed he did so in 1983, 1985, 1987 (twice), and 1990. He attended three conferences organized in Tampere and in 1985 received an Honorary PhD. In June 1987 he attended a conference but in January he was an opponent on the thesis defence of Simo. The thesis was entitled "On the Relative Goodness of Ordinary Least Squares Estimation in the General Linear Model".

Actually Simo met C.R. Rao for the first time in Sheffield, UK, in August 1982, at the first ICOTS Conference (see the website), which George also attended. During this conference Simo invited C.R. Rao to be a keynote speaker in a statistical conference in Tampere in 1983. Rao replied: "I'll come if you George will come too." So they both certainly came—and wrote papers for the *Proceedings*, see Rao (1985b) and Styan (1985).

The first author of the present paper, Augustyn Markiewicz, has two joint articles with C.R. Rao, published in 1992 and 1995. The third coauthor of those papers was Jerzy K. Baksalary (1944–2005), a prolific Polish linear algebra and linear models lover. Those papers deal with the admissibility concept and the consistency of the linear model; see also C.R. Rao's comments on Jerzy's career in O.M. Baksalary & Styan (2005, p. 16).

## Acknowledgements

We would like to acknowledge first the many conversations we had with our colleagues who are no longer with us: T. W. Anderson (1918–2016), Jerzy K. Baksalary (1944–2005), Oscar Kempthorne (1919–2000), Sujit Kumar Mitra (1932–2004), Shayle R. Searle (1928–2013), Geoffrey S. Watson (1921–1998), and George Zyskind (1929–1974).

For lively discussions about 30 years ago, addressing in particular some least-squares historical problems and linear unbiasedness concepts, we would like to thank Ronald Christensen, Richard William Farebrother, and David A. Harville and in particular for their Letters to the Editor about Puntanen & Styan (1989), see Christensen (1990), Farebrother (1990), Harville (1990), and the "Reply" by Puntanen & Styan (1990).

We wish to thank the anonymous referee for valuable remarks improving our presentation. Thanks go also to Kimmo Vehkalahti for constructive comments and to Jarmo Niemelä for very helpful LaTeX assistance while preparing this article.

The quotations in this article from *Statistical Science* and *Annals of Mathematical Statistics* are reprinted with the permission of the Institute of Mathematical Statistics. Similarly, sincere thanks for corresponding permissions go to *IMAGE: The Bulletin of ILAS*, *Linear Algebra and its Applications*, *Sankhyā: The Indian Journal of Statistics*, and *The American Statistician*.

This research was essentially completed during the meetings of an International Research Group on Multivariate and Mixed Linear Models in the Mathematical Research and Conference Center, Będlewo, Poland, in November 2019 and February 2020, supported by the Stefan Banach International Mathematical Center.

## References

Aitken, A.C. (1935). On least squares and linear combination of observations. *Proc. Roy. Soc. Edinburgh Sect. A*, 55, 42–49. DOI

Alalouf, I.S., Styan, G.P.H. (1984). Characterizations of the conditions for the ordinary least squares estimator to be best linear unbiased. *Topics in Applied Statistics* (Proceedings of the Statistics '81 Canada Conference: Concordia University, Montréal, April–May 1981.) (Y.P. Chaubey, T.D. Dwivedi, eds.) Concordia University, Montréal, pp. 331–344.

Albert, A. (1973). The Gauss–Markov theorem for regression models with possibly singular covariances. *SIAM J. Appl. Math.*, 24, 182–187. DOI

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

Anderson, T.W. (1971). *The Statistical Analysis of Time Series*. Wiley, New York. DOI

Anderson, T.W. (1972). Efficient estimation of regression coefficients in time series. *Proc. Sixth Berkeley Symp. on Math. Statist. and Prob., Vol. 1*. (L.M. Le Cam, J. Neyman, E.L. Scott, eds.) Univ. of Calif. Press, pp. 471–482. DOI

Baksalary, J.K. (1988). Criteria for the equality between ordinary least squares and best linear unbiased estimators under certain linear models. *Can. J. Stat.*, 16, 97–102. DOI

Baksalary, J.K. (2004). An elementary development of the equation characterizing best linear unbiased estimators. *Linear Algebra Appl.*, 388, 3–6. DOI

Baksalary, J.K., van Eijnsbergen, A.C. (1988). A comparison of two criteria for ordinary-least-squares estimators to be best linear unbiased estimators. *Am. Stat.*, 42, 205–208. DOI

Baksalary, J.K., Kala, R. (1977). An extension of the rank criterion for the least squares estimator to be the best linear unbiased estimator. *J. Stat. Plann. Inference*, 1, 309–312. DOI

Baksalary, J.K., Kala, R. (1978). A bound for the Euclidean norm of the difference between the least squares and the best linear unbiased estimators. *Ann. Stat.*, 6, 1390–1393. DOI

Baksalary, J.K., Kala, R. (1980). A new bound for the Euclidean norm of the difference between the least squares and the best linear unbiased estimators. *Ann. Stat.*, 8, 679–681. DOI

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

Baksalary, J.K., Kala, R. (1981b). Simple least squares estimation versus best linear unbiased prediction. *J. Stat. Plann. Inference*, 5, 147–151. DOI

Baksalary, J.K., Kala, R. (1983). On equalities between BLUEs, WLSEs, and SLSEs. *Can. J. Stat.*, 11, 119–123. DOI [Corrigendum (1984), 12, p. 240. DOI]

Baksalary, J.K., Kala, R. (1986). Linear sufficiency with respect to a given vector of parametric functions. *J. Stat. Plann. Inference*, 14, 331–338. DOI

Baksalary, J.K., Mathew, T. (1990). Rank invariance criterion and its application to the unified theory of least squares. *Linear Algebra Appl.*, 127, 393–401. DOI

Baksalary, J.K., Puntanen, S. (1989). Weighted-least-squares estimation in the general Gauss–Markov model. *Statistical Data Analysis and Inference*. (Y. Dodge, ed.) Elsevier Science Publishers B.V., Amsterdam, 355–368. DOI

Baksalary, J.K., Puntanen, S. (1990a). Characterizations of the best linear unbiased estimator in the general Gauss–Markov model with the use of matrix partial orderings. *Linear Algebra Appl.*, 127, 363–370. DOI.

Baksalary, J.K., Puntanen, S. (1990b). Spectrum and trace invariance criterion and its statistical applications. *Linear Algebra Appl.*, 142, 121–128. DOI

Baksalary, J.K., Puntanen, S., Styan, G.P.H. (1990a). A property of the dispersion matrix of the best linear unbiased estimator in the general Gauss–Markov model. *Sankhyā Ser. A*, 52, 279–296. DOI

Baksalary, J.K., Puntanen, S., Styan, G.P.H. (1990b). On T.W. Anderson's contributions to solving the problem of when the ordinary least-squares estimator is best linear unbiased and to characterizing rank additivity of matrices. *The Collected Papers of T.W. Anderson: 1943–1985*. (G.P.H. Styan, ed.). Wiley, New York, pp. 1579–1591.

Baksalary, J.K., Rao, C.R., Markiewicz, A. (1992). A study of the influence of the "natural restrictions" on estimation problems in the singular Gauss–Markov model. *J. Stat. Plann. Inference*, 31, 335–351. DOI

Baksalary, J.K., Rao, C.R., Markiewicz, A. (1995). Admissible linear estimation in the general Gauss–Markov model with respect to an arbitrary quadratic risk function. *J. Stat. Plann. Inference*, 34, 341–347. DOI

Baksalary, O.M., Styan, G.P.H. (2005). Some comments on the life and publications of Jerzy K. Baksalary (1944–2005). *Linear Algebra Appl.*, 410, 3–53. DOI

Baksalary, O.M., Trenkler, G., Liski E. (2013). Let us do the twist again. *Stat. Pap.*, 54, 1109–1119. DOI

Baltagi, B.H. (1989). Applications of a necessary and sufficient condition for OLS to be BLUE. *Statist. Probab. Lett.*, 8, 457–461. DOI

Bartmann, F.C., Bloomfield, P. (1981). Inefficiency and correlation. *Biometrika*, 68, 67–71. DOI

Bloomfield, P., Watson, G.S. (1975). The inefficiency of least squares. *Biometrika*, 62, 121–128. DOI

Christensen, R. (1990). Comment on Puntanen and Styan (1989). [Letter to the Editor] *Am. Stat.*, 44, 191–192. DOI

Christensen, R. (2011). *Plane Answers to Complex Questions: the Theory of Linear Models*, 4th Ed. Springer, New York. DOI

Chu, K.L., Isotalo, J., Puntanen, S., Styan, G.P.H. (2004). On decomposing the Watson efficiency of ordinary least squares in a partitioned weakly singular linear model. *Sankhyā*, 66, 634–651. DOI

Chu, K.L., Isotalo, J., Puntanen, S., Styan, G.P.H. (2005). Some further results concerning the decomposition of the Watson efficiency in partitioned linear models. *Sankhyā*, 67, 74–89. DOI

DeGroot, M.H. (1986). A conversation with T.W. Anderson. *Statist. Sci.*, 1, 97–105. DOI

DeGroot, M,H. (1987). A conversation with C.R. Rao. *Statist. Sci.*, 2, 55–67. DOI

Drury, S.W., Liu, S., Lu, C.-Y., Puntanen, S., Styan, G.P.H. (2002). Some comments on several matrix inequalities with applications to canonical correlations: historical background and recent developments. *Sankhyā Ser. A*, 64, 453–507. DOI

Drygas, H. (1970). *The Coordinate-Free Approach to Gauss–Markov Estimation.* Springer, Berlin. DOI

Eaton, M.L. (1970). Gauss–Markov estimation for multivariate linear models: A coordinate free approach. *Ann. Math. Stat.*, 41, 528–538. DOI

Eaton, M.L. (1978). A note on the Gauss–Markov Theorem. *Ann. Inst. Stat. Math.*, 30, 181–184. DOI

Eaton, M.L. (2007). William H. Kruskal and the development of coordinate-free methods. *Statist. Sci.*, 22, 264–265. DOI

Farebrother, R.W. (1990). Weighted least squares and the Gauss–Markov model. [Letter to the Editor] *Am. Stat.*, 44, 191. DOI

Farebrother, R.W. (1997). A.C. Aitken and the consolidation of matrix theory. *Linear Algebra Appl.*, 264, 3–12. DOI

Goldberger, A.S. (1962). Best linear unbiased prediction in the generalized linear regression model. *J. Am. Stat. Assoc.*, 58, 369–375. DOI

Goldman, A.J., Zelen, M. (1964). Weak generalized inverses and minimum variance unbiased estimation. *J. Res. Nat. Bur. Standards Sect. B*, 68, 151–172. DOI

Gouriéroux, C., Monfort, A. (1980). Sufficient linear structures: econometric applications. *Econometrica*, 48, 1083–1097. DOI

Groß, J. (2004). The general Gauss–Markov model with possibly singular dispersion matrix. *Stat. Pap.*, 45, 311–336. DOI

Groß, J., Puntanen, S. (2000). Estimation under a general partitioned linear model. *Linear Algebra Appl.*, 321, 131–144. DOI

Gustafson, K. (1972). Antieigenvalue inequalities in operator theory. *Inequalities, III (Proc. Third Sympos., Univ. Calif., Los Angeles.* (O. Shisha, ed.) Academic Press, New York, pp. 115–119.

Gustafson, K. (2006). The trigonometry of matrix statistics. *Int. Stat. Rev.*, 74, 187–202. DOI

Gustafson, K. (2012). *Antieigenvalue Analysis.* World Scientific, Singapore. DOI

Haberman, S.J. (1975). How much do Gauss–Markov and least square estimates differ? A coordinate-free approach. *Ann. Stat.*, 3, 982–990. DOI

Harville, D.A. (1990). Comment on Puntanen and Styan (1989). [Letter to the Editor] *Am. Stat.*, 44, 192. DOI

Harville, D.A. (1997). *Matrix Algebra From a Statistician's Perspective.* Springer, New York. DOI

Haslett, S.J., Isotalo, J., Kala, R., Markiewicz, A., Puntanen, S. (2021). A review of the linear sufficiency and linear prediction sufficiency in the linear model with new observations. *Multivariate, Multilinear and Mixed Linear Models*. (K. Filipiak, A. Markiewicz & D. von Rosen, eds.) Springer International Publishing, Cham, pp. 265–318. DOI

Haslett, S.J., Isotalo, J., Liu, Y., Puntanen, S. (2014). Equalities between OLSE, BLUE and BLUP in the linear model. *Stat. Pap.*, 55, 543–561. DOI

Haslett, S.J., Welsh, A.H. (2019). Empirical best linear unbiased prediction (EBLUP). *Wiley StatsRef: Statistics Reference Online.* stat08180. Wiley, Chichester. 10 pp. DOI

Herr, D.G. (1980). On the history of the use of geometry in the general linear model. *Am. Stat.*, 34, 43–47. DOI

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

Isotalo, J., Puntanen, S., Styan, G.P.H. (2008a). A useful matrix decomposition and its statistical applications in linear regression. *Commun. Stat. Theory Methods*, 37, 1436–1457. DOI

Isotalo, J., Puntanen, S., Styan, G.P.H. (2008b). The BLUE's covariance matrix revisited: a review. *J. Stat. Plann. Inference*, 138, 2722–2737. DOI

Kala, R. (1981). Projectors and linear estimation in general linear models. *Commun. Stat. Theory Methods*, 10, 849–873. DOI

Kempthorne, O. (1975). George Zyskind 1929–1974 [Obituary]. *Am. Stat.*, 29, 106–107. DOI

Kempthorne, O. (1976). Best linear unbiased estimation with arbitrary variance matrix. *Essays in Probability and Statistics: A Volume in Honor of Professor Junjiro Ogawa*. (S. Ikeda, ed.), Shinko Tsusho, Tokyo, pp. 203–225.

Kempthorne, O. (1989). Comment [on Puntanen and Styan (1989)]. *Am. Stat.*, 43, 161–162. DOI

Khatri, C.G. (1966). A note on a MANOVA model applied to problems in growth curves. *Ann. Inst. Stat. Math.*, 18, 75–86. DOI

Khatri, C.G., Rao, C.R. (1981). Some extensions of the Kantorovich inequality and statistical applications. *J. Multivariate Anal.*, 11, 498–505. DOI

Khatri, C.G., Rao, C.R. (1982). Some generalizations of the Kantorovich inequality. *Sankhyā Ser. A*, 44, 91–102. DOI

Knott, M. (1975). On the minimum efficiency of least squares. *Biometrika*, 62, 129–132. DOI

Kruskal, W. (1961). The coordinate-free approach to Gauss–Markov estimation, and its application to missing and extra observations. *Proc. Fourth Berkeley Symp. on Math. Statist. and Prob., Vol. 1*. Univ. of Calif. Press, 435–451. DOI

Kruskal, W. (1968). When are Gauss–Markov and least squares estimators identical? A coordinate-free approach. *Ann. Math. Stat.*, 39, 70–75. [Corrigendum: (1986), 48, p. 258] DOI

Kruskal, W. (1975). The geometry of generalized inverses. *J. Roy. Statist. Soc. Ser. B*, 37, 272–283. [Corrigendum: (1986), 48, p. 258] DOI

Larocca, R. (2005). Reconciling conflicting Gauss–Markov conditions in the classical linear regression model. *Political Analysis*, 13, 188–207. DOI

Liski, E.P., Puntanen, S., Wang, S.-G. (1992). Bounds for the trace of the difference of the covariance matrices of the OLSE and BLUE. *Linear Algebra Appl.*, 176, 121–130. DOI.

Magness, T.A., McGuire, J.B. (1962). Comparison of least squares and minimum variance estimates of regression parameters. *Ann. Math. Stat.*, 33, 462–470. DOI.

Magness, T.A., McGuire, J.B. (1963). Acknowledgement of priority. *Ann. Math. Stat.*, 34, 352. DOI

Mahalanobis, P.C. (1969). Speech by Professor P.C. Mahalanobis. Indian Statistical Institute, Thirty-Sixth Annual Report: April 1967–March 1968. *Sankhyā Ser. B*, 31, 153–319. DOI

Markiewicz, A., Puntanen, S., Styan, G.P.H. (2010). A note on on the interpretation of the equality of OLSE and BLUE. *Pakistan Journal of Statistics*, 26, 127–134. PDF.

Mathew, T., Bhimasankaram, P. (1983a). On the robustness of the LRT with respect to specification errors in a linear model. *Sankhyā Ser. A*, 45, 212–225. DOI

Mathew, T., Bhimasankaram, P. (1983b). On the robustness of LRT in singular linear models. *Sankhyā Ser. A*, 45, 301–312. DOI

McAleer, M. (1992). Efficient estimation: the Rao–Zyskind condition, Kruskal's theorem and ordinary least squares. *Economic Record*, 68, 65–72. DOI

McElroy, F.W. (1967), A necessary and sufficient condition that ordinary least-squares estimators be best linear unbiased. *J. Am. Stat. Assoc.*, 62, 1302–1304. DOI

Mitra, S.K., Moore, B.J. (1973). Gauss–Markov estimation with an incorrect dispersion matrix. *Sankhyā Ser. A*, 35, 139–152. DOI

Mitra, S.K., Rao, C.R. (1968). Some results in estimation and lests of linear hypotheses under the Gauss–Markoff model. *Sankhyā Ser. A*, 30, 281–290. DOI

Mitra, S.K., Rao, C.R. (1969). Conditions for optimality and validity of simple least squares theory. *Ann. Math. Stat.*, 40, 1617–1624. DOI

Moore, E.H. (1920). On the reciprocal of the general algebraic matrix (Abstract). *Bull. Am. Math. Soc.*, 26, 890–891.

Neel, J.V., Schull, W.J. (1956). The effect of exposure to the atomic bomb on pregnancy termination in Hiroshima and Nagasaki. National Academy of Sciences National Research Council, Washington, D.C., Publ. No. 461.

Penrose, R. (1955). A generalized inverse for matrices. *Math. Proc. Camb. Philos. Soc.*, 51, 406–413. DOI

Puntanen, S. (1987). On the relative goodness of ordinary least squares estimation in the general linear model. (Ph.D. Dissertation) *Acta Universitatis Tamperensis, Ser. A*, vol. 216.

Puntanen, S., Styan, G.P.H. (1989). The equality of the ordinary least squares estimator and the best linear unbiased estimator [with comments by O. Kempthorne and by S.R. Searle and with "Reply" by the authors]. *Am. Stat.*, 43, 153–164. DOI

Puntanen, S., Styan, G.P.H. (1990). Reply [to Letters to the Editor by R.W. Farebrother, R. Christensen and D.A. Harville on Puntanen & Styan (1989)] *Am. Stat.*, 44, 192–193. DOI

Puntanen, S., Styan, G.P.H. (1996). A brief biography and appreciation of Calyampudi Radhakrishna Rao, with a bibliography of books and papers. *Linear Algebra Appl.*, 237/238, 1–40. DOI.

Puntanen, S., Styan, G.P.H. (2012). A conversation with Sujit Kumar Mitra in 1993. *Lectures on Matrix and Graph Methods*. (R.B. Bapat, S. Kirkland, K.M. Prasad, S. Puntanen, eds.) Manipal Universal Press, pp. 221–244. PDF

Puntanen, S., Styan, G.P.H., Isotalo, J. (2011). *Matrix Tricks for Linear Statistical Models: Our Personal Top Twenty.* Springer, Heidelberg. DOI

Puntanen, S., Styan, G.P.H., Isotalo, J. (2013). *Formulas Useful for Linear Regression Analysis and Related Matrix Theory: It's Only Formulas But We Like Them*. Springer, Heidelberg. DOI

Puntanen, S., Styan, G.P.H., Tian, Y. (2005). Three rank formulas associated with the covariance matrices of the BLUE and the OLSE in the general linear model. *Econ. Theory*, 21, 659–663. DOI.

Puntanen, S., Styan, G.P.H., Werner, H.J. (2000). Two matrix-based proofs that the linear estimator $Gy$ is the best linear unbiased estimator. *J. Stat. Plann. Inference*, 88, 173–179. DOI.

Rao, C.R. (1955). Analysis of dispersion for multiply classified data with unequal numbers in cells. *Sankhyā*, 15, 253–280. DOI

Rao. C.R. (1962). A note on a generalized inverse of a matrix with applications to problems in mathematical statistics. *J. Roy. Statist. Soc. Ser. B*, 24, 152–158. DOI

Rao, C.R. (1967). Least squares theory using an estimated dispersion matrix and its application to measurement of signals. *Proc. Fifth Berkeley Symp. on Math. Statist. and Prob., Vol. 1*. (L.M. Le Cam & J. Neyman, eds.) Univ. of Calif. Press, Berkeley, pp. 355–372. DOI

Rao, C.R. (1968). A note on a previous lemma in the theory of least squares and some further results. *Sankhyā Ser. A*, 30, 259–266. DOI

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

Rao, C.R. (1973a). *Linear Statistical Inference and its Applications*, 2nd Ed. Wiley, New York. (First edition 1965.) DOI

Rao, C.R. (1973b). Representations of best linear estimators in the Gauss–Markoff model with a singular dispersion matrix. *J. Multivariate Anal.*, 3, 276–292. DOI

Rao, C.R. (1973c). Unified theory of least squares. *Commun. Stat.*, 1, 1–8. DOI

Rao, C.R. (1974). Projectors, generalized inverses and the BLUEs. *J. Roy. Statist. Soc. Ser. B*, 36, 442–448. DOI

Rao, C.R. (1978). Choice of best linear in the Gauss–Markoff model with a singular dispersion matrix. *Commun. Stat. Theory Methods*, 7, 1119–1208. DOI

Rao, C.R. (1979). Estimation of parameters in the singular Gauss–Markoff model. *Commun. Stat. Theory Methods*, 8, 1353–1358. DOI

Rao, C.R. (1985a). The inefficiency of least squares: extensions of the Kantorovich inequality. *Linear Algebra Appl.*, 70, 249–255. DOI

Rao, C.R. (1985b). A unified approach to inference from linear models. *Proceedings of the First International Tampere Seminar on Linear Statistical Models and their Applications*. (T. Pukkila, S. Puntanen, eds.) Dept. of Math. Sciences, Univ. of Tampere, Finland, pp. 9–36. PDF

Rao, C.R. (2007). Antieigenvalues and antisingularvalues of a matrix and applications to problems in statistics. *Math. Inequal. Appl.*, 10, 471–489. DOI

Rao, C.R., Mitra, S.K. (1971a). *Generalized Inverse of Matrices and Its Applications.* Wiley, New York.

Rao, C.R., Mitra, S.K. (1971b). Further contributions to the theory of generalized inverse of matrices and its applications. *Sankhyā Ser. A*, 33, 289–300. [Corrigenda: (1972) 34, p. 477] DOI

Searle, S.R. (1971). *Linear Models*. Wiley, New York.

Searle, S.R. (1989). Comment [on Puntanen and Styan (1989)]. *Am. Stat.*, 43, 162–163. DOI

Searle, S.R. (1994). Extending some results and proofs for the singular linear model. *Linear Algebra Appl.*, 210, 139–151. DOI

Searle, S.R. (1996). Some follow-up on Aitken's least squares equations. *Proc. of the A.C. Aitken Centenary Conference*. (L. Kavalieris, F.C. Lam, L.A. Roberts, J.A. Shanks, eds.) Univ. of Otago Press, pp. 299–309.

Searle, S.R. (2000). The infusion of matrices into statistics. *IMAGE: The Bulletin of ILAS*, 24, 25–32.

Sengupta, D. & Jammalamadaka, S.R. (2003). *Linear Models: An Integrated Approach.* World Scientific, River Edge.

Styan, G.P.H. (1973). When does least squares give the best linear unbiased estimate? *Proceedings of the Research Seminar at Dalhousie University, Halifax, Nova Scotia, March 23–25, 1972*. (D.G. Kabe, R.P. Gupta, eds.) North-Holland, Amsterdam, pp. 241–246.

Styan, G.P.H. (1983). On some inequalities associated with ordinary least squares and the Kantorovich inequality. *Festschrift for Eino Haikala on his Seventieth Birthday*. (P. Huuhtanen, E.P. Liski, T. Pukkila, S. Puntanen, eds.) *Acta Universitatis Tamperensis, Ser. A*, vol. 153, pp. 158–166. PDF

Styan, G.P.H. (1985). Schur complements and linear statistical models. *Proceedings of the First International Tampere Seminar on Linear Statistical Models and their Applications*. (T. Pukkila, S. Puntanen, eds.) Dept. of Math. Sciences, Univ. of Tampere, Finland, pp. 37–75. PDF

Tian, Y. (2013). On properties of BLUEs under general linear regression models. *J. Stat. Plann. Inference*, 143, 771–782. DOI

Tian, Y., Zhang, X. (2016). On connections among OLSEs and BLUEs of whole and partial parameters under a general linear model. *Statist. Probab. Lett.*, 112, 105–112. DOI

Watson, G.S. (1951). Serial correlation in regression analysis. Ph.D. Dissertation. Dept. of Experimental Statistics, North Carolina State College, Raleigh. DOI

Watson, G.S. (1955). Serial correlation in regression analysis, I. *Biometrika*, 42, 327–341. DOI

Watson, G.S. (1967). Linear least squares regression. *Ann. Math. Stat.*, 38, 1679–1699. DOI

Watson, G.S. (1972). Prediction and the efficiency of least squares. *Biometrika*, 59, 91–98. DOI

Watson, G.S., Alpargu, G., Styan, G.P.H. (1997). Some comments on six inequalities associated with the inefficiency of ordinary least squares with one regressor. *Linear Algebra Appl.*, 264, 13–54. DOI

Zabell, S. (1994). A conversation with William Kruskal. *Statist. Sci.*, 9, 285–303. DOI

Zmyślony, R. (1980). A characterization of best linear unbiased estimators in the general linear model. *Mathematical Statistics and Probability Theory: Proc. Sixth Int. Conf. (Wisła, Poland, 1978)*. (W. Klonecki, A. Kozek, J. Rosiński, eds.) Springer, New York, pp. 365–373. DOI

Zyskind, G. (1962). On conditions for equality of best and simple linear least squares estimators. (Abstract) *Ann. Math. Statist.*, 33, 1502–1503. DOI

Zyskind, G. (1967). On canonical forms, non-negative covariance matrices and best and simple least squares linear estimators in linear models. *Ann. Math. Stat.*, 38, 1092–1109. DOI

Zyskind, G. (1969). Parametric augmentations and error structures under which certain simple least squares and analysis of variance procedures are also best. *J. Am. Stat. Assoc.*, 64, 1353–1368. DOI

Zyskind, G., Martin, F.B. (1969). On best linear estimation and general Gauss–Markov theorem in linear models with arbitrary nonnegative covariance structure. *SIAM J. Appl. Math.*, 17, 1190–1202. DOI