Pak. J. Statist.
2010, Vol. 26(1), 127-134

A NOTE ON THE INTERPRETATION OF THE EQUALITY OF OLSE AND BLUE

Augustyn Markiewicz[1]**, Simo Puntanen**[2]**,** and **George P.H. Styan**[3]

[1] Department of Mathematical and Statistical Methods, Poznan University of Life Sciences, Poznan, Poland
Email: amark@au.poznan.pl
[2] Department of Mathematics and Statistics, University of Tampere, Tampere, Finland
Email: simo.puntanen@uta.fi
[3] Department of Mathematics and Statistics, McGill University, Montreal, Canada
Email: styan@math.mcgill.ca

ABSTRACT

In this paper we consider the linear model $\{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{V}\}$, where $\mathbf{X}$ has full column rank. Denote the ordinary least squares estimator (OLSE) of $\boldsymbol{\beta}$ as $\hat{\boldsymbol{\beta}}$ and the best linear unbiased estimators (BLUE) of $\boldsymbol{\beta}$ as $\tilde{\boldsymbol{\beta}}$. Then in the statistical literature it is common to write $\hat{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}$ when we mean that the OLS method gives the BLU estimator for $\boldsymbol{\beta}$. In this note we discuss some unusual and possibly confusing interpretations of the equality $\hat{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}$.

KEYWORDS

Best linear unbiased estimator; BLUE; Gauss–Markov model; Generalized inverse: OLSE; Ordinary least squares; Orthogonal projector.

2000 Mathematics Subject Classification: 15A42, 62J05, 62F10, 62H12, 62H20

1 INTRODUCTION

In this paper we consider the general linear model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \text{or shortly} \quad \mathcal{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{V}\}, \tag{1.1}$$

where $\mathbf{X}$ is a known $n \times p$ model matrix, the vector $\mathbf{y}$ is an observable n-dimensional random vector, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters and $\boldsymbol{\varepsilon}$ is an unobservable vector of the random errors with the expectation $\mathrm{E}(\varepsilon) = \mathbf{0}$ and the covariance matrix $\mathrm{cov}(\varepsilon) = \sigma^2\mathbf{V}$, where $\sigma^2 > 0$ is an unknown constant while the nonnegative definite (possibly singular) matrix $\mathbf{V}$ is known.

Before proceeding, let us introduce some notation. We will use the symbols $\mathbf{A}'$, $\mathbf{A}^-$, $\mathbf{A}^+$, $\mathcal{C}(\mathbf{A})$, $\mathcal{C}(\mathbf{A})^\perp$, and $\mathcal{N}(\mathbf{A})$ to denote, respectively, the transpose, a generalized inverse, the Moore–Penrose inverse, the column space, the orthogonal complement of the column space, and the null space, of the matrix $\mathbf{A}$. By $(\mathbf{A} : \mathbf{B})$ we denote the partitioned matrix with $\mathbf{A}$ and $\mathbf{B}$ as submatrices. By $\mathbf{A}^\perp$ we denote any matrix satisfying $\mathcal{C}(\mathbf{A}^\perp) = \mathcal{N}(\mathbf{A}') = \mathcal{C}(\mathbf{A})^\perp$. Furthermore, we will write $\mathbf{P_A} = \mathbf{A}\mathbf{A}^+ = \mathbf{A}(\mathbf{A}'\mathbf{A})^-\mathbf{A}'$ [which is invariant with respect to the choice of $(\mathbf{A}'\mathbf{A})^-$] to denote the orthogonal projector (with respect to the standard inner product) onto $\mathcal{C}(\mathbf{A})$. In particular,

$$\mathbf{H} = \mathbf{P_X}, \qquad \mathbf{M} = \mathbf{I}_n - \mathbf{H}. \tag{1.2}$$

One choice for $\mathbf{X}^\perp$ is of course the projector $\mathbf{M}$.

We recall that an unbiased linear estimator $\mathbf{Gy}$ for $\mathbf{X}\beta$ is the best linear unbiased estimator (BLUE) for $\mathbf{X}\beta$ under $\mathcal{M}$ if

$$\mathrm{cov}(\mathbf{Gy}) \leq_{\mathrm{L}} \mathrm{cov}(\mathbf{Ly}) \quad \text{for all } \mathbf{L} : \mathbf{LX} = \mathbf{X}, \tag{1.3}$$

where "$\leq_{\mathrm{L}}$" refers to the Löwner partial ordering. It is well-known, cf., e.g., Rao (1967) and Zyskind (1967), that $\mathbf{Gy}$ is the BLUE for $\mathbf{X}\beta$ if and only if $\mathbf{G}$ satisfies the equation

$$\mathbf{G}(\mathbf{X} : \mathbf{V}\mathbf{X}^\perp) = (\mathbf{X} : \mathbf{0}). \tag{1.4}$$

The corresponding condition for $\mathbf{Ay}$ to be the BLUE of an estimable parametric function $\mathbf{K}'\beta$ is

$$\mathbf{A}(\mathbf{X} : \mathbf{V}\mathbf{X}^\perp) = (\mathbf{K}' : \mathbf{0}). \tag{1.5}$$

The vector $\mathbf{K}'\beta$ is said to be estimable if it has a linear unbiased estimator, which happens if and only if $\mathcal{C}(\mathbf{K}) \subset \mathcal{C}(\mathbf{X}')$.

Consider now two linear models

$$\mathcal{M}_1 = \{\mathbf{y}, \mathbf{X}\beta, \mathbf{V}_1\}, \quad \mathcal{M}_2 = \{\mathbf{y}, \mathbf{X}\beta, \mathbf{V}_2\}, \tag{1.6}$$

which differ only in their covariance matrices. For the proof of the following proposition and related discussion, see, e.g., Rao (1968, Lemma 5; 1971, Th. 5.2, Th. 5.5; 1973, p. 289), Mitra and Moore (1973, Th. 3.3, Th. 4.1–4.2), and Baksalary and Mathew (1986, Th. 3).

Proposition 1.1. *Consider the linear models $\mathcal{M}_1 = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}_1\}$ and $\mathcal{M}_2 = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}_2\}$, and let the notation*

$$\{\,\mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta} \mid \mathcal{M}_1)\,\} \subset \{\,\mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta} \mid \mathcal{M}_2)\,\} \tag{1.7}$$

mean that every representation of the BLUE *for $\mathbf{X}\boldsymbol{\beta}$ under $\mathcal{M}_1$ remains the* BLUE *for $\mathbf{X}\boldsymbol{\beta}$ under $\mathcal{M}_2$. Then the following statements are equivalent:*

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

(b) $\{\,\mathrm{BLUE}(\mathbf{K}'\boldsymbol{\beta} \mid \mathcal{M}_1)\,\} \subset \{\,\mathrm{BLUE}(\mathbf{K}'\boldsymbol{\beta} \mid \mathcal{M}_2)\,\}$ *for every estimable* $\mathbf{K}'\boldsymbol{\beta}$,

(c) $\mathcal{C}(\mathbf{V}_2\mathbf{X}^{\perp}) \subset \mathcal{C}(\mathbf{V}_1\mathbf{X}^{\perp})$.

Notice that obviously the following statements are equivalent:

$$\{\,\mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta} \mid \mathcal{M}_1)\,\} = \{\,\mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta} \mid \mathcal{M}_2)\,\}, \tag{1.8a}$$

$$\mathcal{C}(\mathbf{V}_2\mathbf{X}^{\perp}) = \mathcal{C}(\mathbf{V}_1\mathbf{X}^{\perp}). \tag{1.8b}$$

The ordinary least squares estimator, OLSE, for $\mathbf{K}'\boldsymbol{\beta}$ is defined as $\mathbf{K}'\hat{\boldsymbol{\beta}}$ where $\hat{\boldsymbol{\beta}}$ is an arbitrary solution to the normal equation

$$\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{y}. \tag{1.9}$$

The general solution to the normal equation (which is always consistent) can be expressed as

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y} + [\mathbf{I}_p - (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}]\mathbf{z}, \tag{1.10}$$

where $\mathbf{z} \in \mathbb{R}^p$ is free to vary and $(\mathbf{X}'\mathbf{X})^{-}$ is an arbitrary (but fixed) generalized inverse of $\mathbf{X}'\mathbf{X}$.

When $\mathbf{X}$ does not have full column rank, then the vector $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y}$ is not unique and is not a *proper* estimator: it is merely a *solution* to the normal equations – "...this point cannot be overemphasized", as stated by Searle (1971, p. 169). The very same concerns trivially the parametric function $\mathbf{K}'\boldsymbol{\beta}$. However, as can be seen from (1.10), the OLSE of an estimable $\mathbf{K}'\boldsymbol{\beta}$ is unique; it is $\mathbf{K}'\mathbf{X}^{+}\mathbf{y}$. Notice that of course

$$\mathrm{OLSE}(\mathbf{X}\boldsymbol{\beta}) = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y} = \mathbf{X}\mathbf{X}^{+}\mathbf{y} = \mathbf{H}\mathbf{y}. \tag{1.11}$$

In view of (1.4), it is obvious that under $\mathcal{M}_I = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}\}$, the estimator $\mathbf{G}\mathbf{y}$ is the BLUE for $\mathbf{X}\boldsymbol{\beta}$ if and only if

$$\mathbf{G}(\mathbf{X} : \mathbf{X}^{\perp}) = (\mathbf{X} : \mathbf{0}), \tag{1.12}$$

whose only solution is $\mathbf{G} = \mathbf{H} = \mathbf{P_X}$, and so we have the well-known result:

$$\mathbf{Hy} = \mathrm{OLSE}(\mathbf{X}\boldsymbol{\beta}) \text{ is the BLUE for } \mathbf{X}\boldsymbol{\beta} \text{ under } \mathcal{M}_I = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}\}. \tag{1.13}$$

Let us now consider the equality

$$\hat{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}, \tag{1.14}$$

and let us present some interpretations for its real (statistical) meaning. First we observe that there is no possibility for (1.14) to hold unless $\boldsymbol{\beta}$ is estimable; of course $\boldsymbol{\beta}$ cannot have the BLUE if it does not have a linear unbiased linear estimator. So we necessarily need to assume, while considering (1.14), that $\mathbf{X}$ has full column rank, in which case $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{X}^{+}\mathbf{y}$.

Let us replace the models $\mathcal{M}_1$ and $\mathcal{M}_2$ of (1.6) with

$$\mathcal{M}_I = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{I}\}, \quad \text{and} \quad \mathcal{M}_V = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{V}\}, \tag{1.15}$$

and ask the following question:

$$\begin{gathered}\text{When does every BLUE of } \boldsymbol{\beta} \text{ under } \mathcal{M}_I \\ \text{continue to be the BLUE of } \boldsymbol{\beta} \text{ also under } \mathcal{M}_V \text{ and vice versa?}\end{gathered} \tag{1.16}$$

In view of Proposition 1.1, the answer follows from the equivalence of the following two statements:

$$\{\,\mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta} \mid \mathcal{M}_I)\,\} = \{\,\mathrm{BLUE}(\mathbf{X}\boldsymbol{\beta}) \mid \mathcal{M}_V)\,\}\,, \tag{1.17a}$$

$$\mathcal{C}(\mathbf{V}\mathbf{X}^{\perp}) = \mathcal{C}(\mathbf{X}^{\perp})\,, \tag{1.17b}$$

where the latter equality is obviously equivalent to

$$\mathcal{C}(\mathbf{VM}) = \mathcal{C}(\mathbf{M}), \quad \text{or equivalently,} \quad \mathcal{C}(\mathbf{VX}) = \mathcal{C}(\mathbf{X}). \tag{1.18}$$

However, the condition (1.18) is *not* exactly the same as the classic condition of Rao (1967) and Zyskind (1967) for $\mathrm{OLSE}(\boldsymbol{\beta})$ to be equal to the BLUE of $\boldsymbol{\beta}$ which can be written in the following equivalent forms:

$$\text{(a) } \mathcal{C}(\mathbf{VX}) \subset \mathcal{C}(\mathbf{X}), \quad \text{(b) } \mathcal{C}(\mathbf{VM}) \subset \mathcal{C}(\mathbf{M}), \quad \text{(c) } \mathbf{HVM} = \mathbf{0}. \tag{1.19}$$

For further equivalent conditions, see, e.g., Puntanen and Styan (1989).

What is now the explanation for the difference between the conditions (1.18) and (1.19)? To give a clarifying answer to this question is the main goal of this paper, and this is what we do in the next section.

2 Conclusions

According to Mitra and Moore (1973, p. 139), the comparison of the BLUEs for $\mathbf{X}\boldsymbol{\beta}$ under the models $\mathcal{M}_1$ and $\mathcal{M}_2$, can be divided into three questions:

(a) When is a specific linear representation of the BLUE of $\mathbf{X}\boldsymbol{\beta}$ under $\mathcal{M}_1$ also a BLUE under $\mathcal{M}_2$?

(b) When does $\mathbf{X}\boldsymbol{\beta}$ have a common BLUE under $\mathcal{M}_1$ and $\mathcal{M}_2$?

(c) When is the BLUE of $\mathbf{X}\boldsymbol{\beta}$ under $\mathcal{M}_1$ irrespective of the linear representation used in its expression, also a BLUE under $\mathcal{M}_2$?

We may cite Mitra and Moore (1973):

> "When $\mathbf{V}_1$ is singular, it is conceivable that the BLUE of an estimable linear functional $\mathbf{p}'\boldsymbol{\beta}$ may have distinct linear representations which are equal with probability 1 under $\mathcal{M}_1$, but need not be so under $\mathcal{M}_2$. This shows, in such cases, it is important to recognize the existence of three separate problems as stated above. ... When
>
> $$\operatorname{rank}(\mathbf{V}_1\mathbf{X}^{\perp}) = n - \operatorname{rank}(\mathbf{X})\,, \tag{2.1}$$
>
> the BLUE of $\mathbf{p}'\boldsymbol{\beta}$ has a unique linear presentation. Here, naturally, the three problems merge into one."

Now we can see that the question posed in (1.16) is of type (c) of Mitra and Moore (1973), that is, we require that

$$\text{every BLUE of } \boldsymbol{\beta} \text{ under } \mathcal{M}_I \text{ remains the BLUE under } \mathcal{M}_V\,, \tag{2.2}$$

and

$$\text{every BLUE of } \boldsymbol{\beta} \text{ under } \mathcal{M}_V \text{ remains the BLUE under } \mathcal{M}_I. \tag{2.3}$$

As we know, the BLUE of $\boldsymbol{\beta}$ under $\mathcal{M}_I$ is nothing but the OLSE of $\boldsymbol{\beta}$, i.e.,

$$\operatorname{BLUE}(\boldsymbol{\beta} \mid \mathcal{M}_I) = \operatorname{OLSE}(\boldsymbol{\beta} \mid \mathcal{M}_I) = \hat{\boldsymbol{\beta}} = \mathbf{X}^{+}\mathbf{y} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}, \tag{2.4}$$

whose presentation is unique. Now (2.2) holds if and only if the classic condition (1.19), holds. On the other hand, (2.3) holds if and only if

$$\mathcal{C}(\mathbf{M}) \subset \mathcal{C}(\mathbf{VM}). \tag{2.5}$$

We observe that it is the requirement (2.3) that causes the unexpected equality in (1.18). Moreover, (2.5) implies that $\operatorname{rank}(\mathbf{M}) \leq \operatorname{rank}(\mathbf{VM}) \leq \operatorname{rank}(\mathbf{M})$, and so

$$\operatorname{rank}(\mathbf{M}) = \operatorname{rank}(\mathbf{VM}), \tag{2.6}$$

and thereby (2.5) can be equivalently written as

$$\mathcal{C}(\mathbf{M}) = \mathcal{C}(\mathbf{VM}). \tag{2.7}$$

Now $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ is the BLUE for $\boldsymbol{\beta}$ under $\mathcal{M}_V = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{V}\}$ if and only if

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X} : \mathbf{VM}) = (\mathbf{I}_p : \mathbf{0}), \tag{2.8}$$

which obviously is equivalent to any of the conditions in (1.19). Thus we have confirmed the classic Rao–Zyskind condition.

However, taking an arbitrary representation of the BLUE for $\boldsymbol{\beta}$ under $\mathcal{M}_V$, say $\tilde{\boldsymbol{\beta}}$, this arbitrary $\tilde{\boldsymbol{\beta}}$ may not be a BLUE under the model $\mathcal{M}_I$; in view of Proposition 1.1, this holds if and only if (2.5) holds. We may confirm this quickly as follows. We first write the general representation for $\mathbf{G}_0$ such that $\mathbf{G}_0\mathbf{y}$ is the BLUE for $\mathbf{X}\boldsymbol{\beta}$ as

$$\mathbf{G}_0 = \mathbf{H} + \mathbf{HVM}(\mathbf{MVM})^{+}\mathbf{M} + \mathbf{F}(\mathbf{I}_n - \mathbf{P}_{(\mathbf{X}:\mathbf{V})}), \tag{2.9}$$

where $\mathbf{F}$ is free to vary; for the above expression, see, e.g., Rao (1973). Requesting that $\mathbf{G}_0$ also satisfies

$$\mathbf{G}_0(\mathbf{X} : \mathbf{M}) = (\mathbf{X} : \mathbf{0}) \tag{2.10}$$

for every $\mathbf{F}$, forces $\mathbf{G}_0 = \mathbf{H}$, which obviously holds if and only if

$$\mathcal{C}(\mathbf{X} : \mathbf{V}) = \mathbb{R}^n, \quad \text{and} \quad \mathbf{HVM} = \mathbf{0}, \tag{2.11}$$

i.e.,

$$\mathcal{C}(\mathbf{X} : \mathbf{V}) = \mathbb{R}^n, \quad \text{and} \quad \mathcal{C}(\mathbf{VM}) \subset \mathcal{C}(\mathbf{M}). \tag{2.12}$$

Because $\operatorname{rank}(\mathbf{X} : \mathbf{V}) = \operatorname{rank}(\mathbf{X}) + \operatorname{rank}(\mathbf{VM})$, the first condition in (2.12) implies that $\operatorname{rank}(\mathbf{VM}) = \operatorname{rank}(\mathbf{M})$, which together with the second condition means that $\mathcal{C}(\mathbf{VM}) = \mathcal{C}(\mathbf{M})$. Thus we have proved the equivalence of (2.3) and (2.7).

Notice that condition $\mathcal{C}(\mathbf{VM}) = \mathcal{C}(\mathbf{M})$, or equivalently $\mathcal{C}(\mathbf{VX}) = \mathcal{C}(\mathbf{X})$, implies that $\mathbf{V}$ is necessarily positive definite. This is seen from $\mathcal{C}(\mathbf{X} : \mathbf{V}) = \mathbb{R}^n$ by replacing $\mathbf{X}$ with $\mathbf{VX}$.

Our conclusion is that the classic condition

$$\mathcal{C}(\mathbf{VM}) \subset \mathcal{C}(\mathbf{M}) \tag{2.13}$$

is the answer to the following question: What is a necessary and sufficient condition that OLSE($\boldsymbol{\beta}$), i.e., $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$, is the BLUE for $\boldsymbol{\beta}$ under the model $\mathcal{M}_V = \{\mathbf{y}, \mathbf{X}\beta, \sigma^2\mathbf{V}\}$?

Moreover, the "somewhat unexpected" condition

$$\mathcal{C}(\mathbf{M}) \subset \mathcal{C}(\mathbf{VM}) \tag{2.14}$$

is the answer to the following question: What is a necessary and sufficient condition that every BLUE for $\boldsymbol{\beta}$ under $\mathcal{M}_V = \{\mathbf{y}, \mathbf{X}\beta, \sigma^2\mathbf{V}\}$ is also the BLUE for $\boldsymbol{\beta}$ under $\mathcal{M}_I = \{\mathbf{y}, \mathbf{X}\beta, \sigma^2\mathbf{I}\}$?

We find the latter question rather unusual and it is the former one that most statisticians have in mind when they ask for the conditions for the equality $\hat{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}$ to hold.

Acknowledgements

Thanks go to Jarkko Isotalo for helpful discussions. The research of the third author was supported in part by the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

1. Baksalary, J.K. and Mathew, T. (1986). Linear sufficiency and completeness in an incorrectly specified general Gauss–Markov model. *Sankhyā, Ser. A*, **48**, 169–180.

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

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

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

5. 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.

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

7. Rao, C.R. (1973). Representations of best linear unbiased estimators in the Gauss–Markoff model with a singular dispersion matrix. *Journal of Multivariate Analysis*, **3**, 276–292.

8. Searle, S.R. (1971). *Linear Models*. Wiley, New York. (Wiley Classics Library Reprint Edition, 1997.)

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