IJSS (2007), Vol. 6, pp. 193-201

Manuscript for the 16 July 2006
International Journal of Statistical Sciences
Special Issue in Felicitation of Professor Mir Masoom Ali

Effect of adding regressors on the equality of the OLSE and BLUE

Jarkko Isotalo
University of Tampere, Tampere, Finland
Simo Puntanen
University of Tampere, Tampere, Finland
and
George P. H. Styan
McGill University, Montréal (Québec), Canada

ABSTRACT. We consider the estimation of regression coefficients in a partitioned linear model, shortly denoted as $\mathscr{M}_{12} = \{\mathbf{y}, \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2, \mathbf{V}\}$. We call $\mathscr{M}_{12}$ a full model, and correspondingly, $\mathscr{M}_1 = \{\mathbf{y}, \mathbf{X}_1\boldsymbol{\beta}_1, \mathbf{V}\}$ a small model. We introduce a necessary and sufficient condition for the equality between the ordinary least squares estimator (OLSE) of $\boldsymbol{\beta}_1$ and the best linear unbiased estimator (BLUE) of $\boldsymbol{\beta}_1$ under the full model $\mathscr{M}_{12}$ assuming that they are equal under the small model $\mathscr{M}_1$. This condition can then be applied to generalize some results of Nurhonen and Puntanen (1992) concerning the effect of deleting an observation on the equality of OLSE and BLUE.

1 Introduction

In this paper we consider the partitioned linear model

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}, \tag{1.1}$$

or shortly,

$$\mathscr{M}_{12} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\} = \{\mathbf{y}, \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2, \mathbf{V}\}, \tag{1.2}$$

where $\mathrm{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$, $\mathrm{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$, $\mathrm{cov}(\mathbf{y}) = \mathrm{cov}(\boldsymbol{\varepsilon}) = \mathbf{V}$. We denote the expectation vector and covariance matrix, respectively, by $\mathrm{E}(\cdot)$ and $\mathrm{cov}(\cdot)$.

In the model $\mathscr{M}_{12}$ the vector $\mathbf{y}$ is an $n \times 1$ observable random vector, $\boldsymbol{\varepsilon}$ is an $n \times 1$ random error vector, $\mathbf{X}$ is a known $n \times p$ matrix, partitioned columnwise as $\mathbf{X} = (\mathbf{X}_1 : \mathbf{X}_2)$ with $\mathbf{X}_1$ $(n \times p_1)$ and $\mathbf{X}_2$ $(n \times p_2)$, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters, and $\mathbf{V}$ is a known $n \times n$ nonnegative definite

MSC (2000) *subject classification.* 62J05, 62H12, 62H20.
Keywords. Best linear unbiased estimation, BLUE, Frisch–Waugh–Lovell theorem, Gauß–Markov model, OLSE, ordinary least squares, orthogonal projector, partitioned linear model, reduced linear model, Watson efficiency, weakly singular linear model, Zyskind–Martin model.

matrix. We omit the variance multiplier σ^2 from the covariance matrix of $\mathbf{y}$ since our main focus lies in the efficiency of ordinary least squares (OLS) estimator of $\boldsymbol{\beta}$ and therein σ^2 has no role.

We will use the symbols $\mathbf{A}'$, $\mathbf{A}^-$, $\mathbf{A}^+$, $\mathscr{C}(\mathbf{A})$, $\mathscr{C}(\mathbf{A})^\perp$ and $\mathrm{r}(\mathbf{A})$ to denote, respectively, the transpose, a generalized inverse, the Moore–Penrose inverse, the column space, the orthogonal complement of the column space and the rank of the matrix $\mathbf{A}$. Furthermore we will write $\mathbf{P_A} = \mathbf{AA}^+ = \mathbf{A}(\mathbf{A}'\mathbf{A})^-\mathbf{A}'$ to denote the orthogonal projector (with respect to the standard inner product) onto $\mathscr{C}(\mathbf{A})$. In particular,

$$\mathbf{P}_i = \mathbf{P}_{\mathbf{X}_i}, \quad \mathbf{M}_i = \mathbf{I} - \mathbf{P}_i, \quad i = 1, 2; \qquad \mathbf{H} = \mathbf{P_X}, \quad \mathbf{M} = \mathbf{I} - \mathbf{H}. \tag{1.3}$$

In addition to $\mathscr{M}_{12}$, which we call the *full model*, we will consider the *small model*

$$\mathscr{M}_1 = \{\mathbf{y},\, \mathbf{X}_1\boldsymbol{\beta}_1,\, \mathbf{V}\}, \tag{1.4}$$

and the *reduced model*

$$\mathscr{M}_{12\cdot 1} = \{\mathbf{M}_1\mathbf{y},\, \mathbf{M}_1\mathbf{X}_2\boldsymbol{\beta}_2,\, \mathbf{M}_1\mathbf{V}\mathbf{M}_1\}. \tag{1.5}$$

The model $\mathscr{M}_{12\cdot 1}$ is obtained by premultiplying the full model equation (1.1) by the orthogonal projector $\mathbf{M}_1$. We define the models $\mathscr{M}_2$ and $\mathscr{M}_{12\cdot 2}$ similarly to the models $\mathscr{M}_1$ and $\mathscr{M}_{12\cdot 1}$.

We assume the model to be consistent in that

$$\mathbf{y} \in \mathscr{C}(\mathbf{X} : \mathbf{V}) = \mathscr{C}(\mathbf{X} : \mathbf{VM}). \tag{1.6}$$

Note that whenever we have a statement in this paper that is related to the random vector $\mathbf{y}$, such a statement holds with probability 1, i.e., the statement holds for all $\mathbf{y}$ satisfying (1.6). In particular, if the column space inclusion $\mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{V})$ holds, then (1.6) becomes $\mathbf{y} \in \mathscr{C}(\mathbf{V})$.

When $\mathbf{X}$ has full column rank, then the vector $\boldsymbol{\beta}$ is estimable, and the ordinary least squares estimator (OLSE) and the best linear unbiased estimator (BLUE) of $\boldsymbol{\beta}$ under the full model $\mathscr{M}_{12}$ are, respectively,

$$\mathrm{OLSE}(\boldsymbol{\beta}) = \hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_2 \end{pmatrix} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \hat{\boldsymbol{\beta}}(\mathscr{M}_{12}), \tag{1.7}$$

$$\mathrm{BLUE}(\boldsymbol{\beta}) = \tilde{\boldsymbol{\beta}} = \begin{pmatrix} \tilde{\boldsymbol{\beta}}_1 \\ \tilde{\boldsymbol{\beta}}_2 \end{pmatrix} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} = \tilde{\boldsymbol{\beta}}(\mathscr{M}_{12}), \tag{1.8}$$

when $\mathbf{V}$ is positive definite. The corresponding covariance matrices are

$$\mathrm{cov}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}, \quad \mathrm{cov}(\tilde{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}, \tag{1.9}$$

and hence we have the Löwner ordering

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \geq (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}, \tag{1.10}$$

i.e., the matrix $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} - (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}$ is nonnegative definite.

In this paper we consider the so-called *weakly singular model* (or Zyskind–Martin model) which means that $\mathbf{V}$ may be singular but then the column space inclusion

$$\mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{V}) \tag{1.11}$$

must hold; see, e.g., Zyskind and Martin (1969). Under this model, the BLUE($\boldsymbol{\beta}$) can be expressed as

$$\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{+}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{+}\mathbf{y} \tag{1.12}$$

($\mathbf{V}^{+}$ being replaceable with any $\mathbf{V}^{-}$), and its covariance matrix can be written as $(\mathbf{X}'\mathbf{V}^{+}\mathbf{X})^{-1}$ and hence, the Watson efficiency (Watson 1955, p. 330) becomes

$$\phi_{12} = \mathrm{eff}(\hat{\boldsymbol{\beta}} \mid \mathscr{M}_{12}) = \frac{|\mathrm{cov}(\tilde{\boldsymbol{\beta}})|}{|\mathrm{cov}(\hat{\boldsymbol{\beta}})|} = \frac{|\mathbf{X}'\mathbf{X}|^2}{|\mathbf{X}'\mathbf{V}\mathbf{X}| \cdot |\mathbf{X}'\mathbf{V}^{+}\mathbf{X}|}. \tag{1.13}$$

All expressions in (1.12) and (1.13) are invariant for all choices of generalized inverses; see, e.g., Rao and Mitra (1971, Lemma 2.2.4).

We will call ϕ_{12} the *total Watson efficiency.* Clearly we have

$$0 < \phi_{12} \leq 1, \tag{1.14}$$

where the upper bound is attained if and only if the OLSE equals the BLUE; see, e.g., Puntanen and Styan (1989). There are numerous equivalent characterizations—originating from Anderson (1948), Rao (1967) and Zyskind (1967)—for the equality between OLSE and BLUE, i.e., equality in (1.10). For example, each of the following conditions is a necessary and sufficient condition for the equality between the OLSE and BLUE:

$$\mathscr{C}(\mathbf{V}\mathbf{X}) \subset \mathscr{C}(\mathbf{X}), \quad \mathbf{H}\mathbf{V} = \mathbf{V}\mathbf{H}, \quad \mathbf{H}\mathbf{V}\mathbf{M} = \mathbf{0}, \tag{1.15}$$

where $\mathbf{H}$ and $\mathbf{V}$ are replaceable with $\mathbf{M}$ and $\mathbf{V}^{+}$, respectively.

Chu *et al.* (2004, 2005) introduced a new decomposition for the total Watson efficiency ϕ_{12}. According to this decomposition, the total efficiency ϕ_{12} can be expressed as a product

$$\mathrm{eff}(\hat{\boldsymbol{\beta}} \mid \mathscr{M}_{12}) = \mathrm{eff}(\hat{\boldsymbol{\beta}}_1 \mid \mathscr{M}_1) \cdot \mathrm{eff}(\hat{\boldsymbol{\beta}}_2 \mid \mathscr{M}_{12}) \cdot \alpha_1, \tag{1.16}$$

where $\mathrm{eff}(\cdot \mid \cdot)$ refers to the Watson efficiency of a particular parameter vector under a particular model, and α_1 is a specific determinant ratio. We will not utilize this decomposition in this paper, but we will use the notation $\mathrm{eff}(\cdot \mid \cdot)$ for the Watson efficiency.

Taking a look at the models ($\mathbf{X}$ having full column rank), we can immediately conclude that the OLS estimators of $\boldsymbol{\beta}_2$ under the models $\mathscr{M}_{12}$ and $\mathscr{M}_{12\cdot 1}$ coincide:

$$\hat{\boldsymbol{\beta}}_2(\mathscr{M}_{12}) = \hat{\boldsymbol{\beta}}_2(\mathscr{M}_{12\cdot 1}) = (\mathbf{X}_2'\mathbf{M}_1\mathbf{X}_2)^{-1}\mathbf{X}_2'\mathbf{M}_1\mathbf{y}. \tag{1.17}$$

The equality in (1.17) is just the result that Davidson and MacKinnon (2004, §2.4) call the Frisch–Waugh–Lovell Theorem. Correspondingly, it can be shown (Chu *et al.* 2004, p. 640) that the BLUEs of $\boldsymbol{\beta}_2$ under the full model $\mathscr{M}_{12}$ and the reduced model $\mathscr{M}_{12\cdot 1}$ are equal, i.e.,

$$\tilde{\boldsymbol{\beta}}_2(\mathscr{M}_{12}) = \tilde{\boldsymbol{\beta}}_2(\mathscr{M}_{12\cdot 1}) = (\mathbf{X}_2'\dot{\mathbf{M}}_1\mathbf{X}_2)^{-1}\mathbf{X}_2'\dot{\mathbf{M}}_1\mathbf{y}, \tag{1.18}$$

where $\dot{\mathbf{M}}_1 = \mathbf{M}_1(\mathbf{M}_1\mathbf{V}\mathbf{M}_1)^{-}\mathbf{M}_1$. Now, in view of (1.17) and (1.18) we can, following Groß and Puntanen (2000, p. 142) and Chu *et al.* (2004, p. 641), conclude the following lemma:

Lemma 1.1. *Consider a partitioned linear model $\mathscr{M}_{12}$, where $\mathbf{X}_2$ has full column rank and the disjointness property*

$$\mathscr{C}(\mathbf{X}_1) \cap \mathscr{C}(\mathbf{X}_2) = \{\mathbf{0}\} \tag{1.19}$$

holds. Then the following statements are equivalent:

(a) $\hat{\boldsymbol{\beta}}_2(\mathscr{M}_{12}) = \tilde{\boldsymbol{\beta}}_2(\mathscr{M}_{12})$,

(b) $\hat{\boldsymbol{\beta}}_2(\mathscr{M}_{12\cdot 1}) = \tilde{\boldsymbol{\beta}}_2(\mathscr{M}_{12\cdot 1})$,

(c) $\mathscr{C}(\mathbf{M}_1\mathbf{V}\mathbf{M}_1\mathbf{X}_2) \subset \mathscr{C}(\mathbf{M}_1\mathbf{X}_2)$,

(d) *the column space $\mathscr{C}(\mathbf{M}_1\mathbf{X}_2)$ has a basis comprising p_2 orthonormal eigenvectors of $\mathbf{M}_1\mathbf{V}\mathbf{M}_1$.*

We may note that the disjointness condition (1.19) means that $\mathbf{X}_2\boldsymbol{\beta}_2$ (and thereby $\mathbf{X}_1\boldsymbol{\beta}_1$) is estimable under $\mathscr{M}_{12}$. The disjointness together with $\mathrm{r}(\mathbf{X}_2) = p_2$ guarantee the estimability of $\boldsymbol{\beta}_2$ under $\mathscr{M}_{12}$. Moreover, using the rank rule of the matrix product (Marsaglia and Styan 1974, p. 276),

$$\mathrm{r}(\mathbf{AB}) = \mathrm{r}(\mathbf{A}) - \dim \mathscr{C}(\mathbf{A}') \cap \mathscr{C}(\mathbf{B})^{\perp}, \tag{1.20}$$

we can conclude that $\mathrm{r}(\mathbf{X}_2'\mathbf{M}_1) = p_2$ holds if and only if (1.19) holds and $\mathrm{r}(\mathbf{X}_2) = p_2$.

Later we will need also the following result, see, e.g.; Sengupta and Jammalamadaka (2003, Ch. 9):

Lemma 1.2. *Consider a weakly singular partitioned linear model $\mathscr{M}_{12}$ where $\mathbf{X}$ has full column rank. Then*

$$\tilde{\boldsymbol{\beta}}_1(\mathscr{M}_{12}) = \tilde{\boldsymbol{\beta}}_1(\mathscr{M}_1) - (\mathbf{X}_1'\mathbf{V}^+\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{V}^+\mathbf{X}_2\tilde{\boldsymbol{\beta}}_2(\mathscr{M}_{12}). \tag{1.21}$$

2 Main results

In this section we shall pay particular attention to the case when $\hat{\boldsymbol{\beta}}_1$ is fully efficient under the small model $\mathscr{M}_1 = \{\mathbf{y}, \mathbf{X}_1\boldsymbol{\beta}_1, \mathbf{V}\}$. We then add new explanatory variables into the model, or delete observations from the model and study the consequences. Our particular aim is to find conditions under which $\hat{\boldsymbol{\beta}}_1$ remains fully efficient in the transformed model.

We begin with the main theorem of this paper.

Theorem 2.1. *Consider a weakly singular partitioned linear model $\mathscr{M}_{12} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$, where $\mathbf{X}$ has full column rank. Let us assume that $\hat{\boldsymbol{\beta}}_1$ is fully efficient in the small model $\mathscr{M}_1 = \{\mathbf{y}, \mathbf{X}_1\boldsymbol{\beta}_1, \mathbf{V}\}$, i.e.,*

$$\mathrm{eff}(\hat{\boldsymbol{\beta}}_1 \mid \mathscr{M}_1) = 1, \textit{ or equivalently, } \mathbf{M}_1\mathbf{V} = \mathbf{V}\mathbf{M}_1. \tag{2.1}$$

Then the following statement holds:

$$\mathrm{eff}(\hat{\boldsymbol{\beta}}_1 \mid \mathscr{M}_{12}) = 1 \iff \mathbf{X}_1'\mathbf{X}_2\tilde{\boldsymbol{\beta}}_2(\mathscr{M}_{12}) = \mathbf{X}_1'\mathbf{X}_2\hat{\boldsymbol{\beta}}_2(\mathscr{M}_{12}). \tag{2.2}$$

Proof. We first note that Lemma 1.2 implies that

$$\tilde{\boldsymbol{\beta}}_1(\mathscr{M}_{12}) = \tilde{\boldsymbol{\beta}}_1(\mathscr{M}_1) - (\mathbf{X}_1'\mathbf{V}^+\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{V}^+\mathbf{X}_2\tilde{\boldsymbol{\beta}}_2(\mathscr{M}_{12}), \tag{2.3}$$

$$\hat{\boldsymbol{\beta}}_1(\mathscr{M}_{12}) = \hat{\boldsymbol{\beta}}_1(\mathscr{M}_1) - (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2\hat{\boldsymbol{\beta}}_2(\mathscr{M}_{12}). \tag{2.4}$$

Assuming that $\tilde{\boldsymbol{\beta}}_1(\mathscr{M}_1) = \hat{\boldsymbol{\beta}}_1(\mathscr{M}_1)$, we have the equality

$$(\mathbf{X}_1'\mathbf{V}^+\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{V}^+ = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1', \tag{2.5}$$

and hence can rewrite (2.3) as

$$\tilde{\boldsymbol{\beta}}_1(\mathscr{M}_{12}) = \hat{\boldsymbol{\beta}}_1(\mathscr{M}_1) - (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{X}_2\tilde{\boldsymbol{\beta}}_2(\mathscr{M}_{12}). \tag{2.6}$$

Combining (2.4) and (2.6) yields

$$\tilde{\boldsymbol{\beta}}_1(\mathscr{M}_{12}) = \hat{\boldsymbol{\beta}}_1(\mathscr{M}_{12}) \iff \mathbf{X}_1'\mathbf{X}_2[\tilde{\boldsymbol{\beta}}_2(\mathscr{M}_{12}) - \hat{\boldsymbol{\beta}}_2(\mathscr{M}_{12})] = \mathbf{0}, \tag{2.7}$$

and thus (2.2) is proved. □

The equality (2.5) may deserve a further comment. Namely, even though it is clear that $\tilde{\boldsymbol{\beta}}_1(\mathscr{M}_1) = \hat{\boldsymbol{\beta}}_1(\mathscr{M}_1)$ means that

$$(\mathbf{X}_1'\mathbf{V}^+\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{V}^+\mathbf{y} := \mathbf{G}\mathbf{y} = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{y} = \mathbf{X}_1^+\mathbf{y} \tag{2.8}$$

holds, it does not necessarily mean that $\mathbf{G} = \mathbf{X}_1^+$. This is so since in view of the consistency condition (1.6), the equality (2.8) need to be valid only for all vectors $\mathbf{y} \in \mathscr{C}(\mathbf{V})$. However, using the commutativity of $\mathbf{P}_1\mathbf{V}^+$ [see (1.15)], we obtain

$$\begin{aligned}
(\mathbf{X}_1'\mathbf{V}^+\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{V}^+ &= (\mathbf{X}_1'\mathbf{V}^+\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{P}_1\mathbf{V}^+ \\
&= (\mathbf{X}_1'\mathbf{V}^+\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{V}^+\mathbf{P}_1 \\
&= (\mathbf{X}_1'\mathbf{V}^+\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{V}^+\mathbf{X}_1\mathbf{X}_1^+ \\
&= \mathbf{X}_1^+ = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'.
\end{aligned} \tag{2.9}$$

It is noteworthy that if the columns of $\mathbf{X}_1$ and $\mathbf{X}_2$ are orthogonal to each other, i.e., $\mathbf{X}_1'\mathbf{X}_2 = \mathbf{0}$, [and $\operatorname{eff}(\hat{\boldsymbol{\beta}}_1 \mid \mathscr{M}_1) = 1$] then adding new regressors (columns in $\mathbf{X}_2$) keeps $\hat{\boldsymbol{\beta}}_1$ fully efficient in $\mathscr{M}_{12}$.

Consider next some further properties of (2.2). First, we observe that the implication

$$\mathbf{X}_1'\mathbf{X}_2[\tilde{\boldsymbol{\beta}}_2(\mathscr{M}_{12}) - \hat{\boldsymbol{\beta}}_2(\mathscr{M}_{12})] = \mathbf{0} \implies \tilde{\boldsymbol{\beta}}_2(\mathscr{M}_{12}) = \hat{\boldsymbol{\beta}}_2(\mathscr{M}_{12}) \tag{2.10}$$

holds whenever the matrix $\mathbf{X}_1'\mathbf{X}_2$ has full column rank:

$$\mathrm{r}(\mathbf{X}_1'\mathbf{X}_2) = \mathrm{r}(\mathbf{X}_2) = p_2. \tag{2.11}$$

Hence we have the following result.

Corollary 2.1. *Consider a weakly singular partitioned linear model* $\mathscr{M}_{12} = \{\mathbf{y},\, \mathbf{X}\boldsymbol{\beta},\, \mathbf{V}\}$, *where* $\mathbf{X}$ *has full column rank. Assume that*

$$\text{(i) } \operatorname{eff}(\hat{\boldsymbol{\beta}}_1 \mid \mathscr{M}_1) = 1, \text{ and (ii) } \mathrm{r}(\mathbf{X}_1'\mathbf{X}_2) = p_2. \tag{2.12}$$

Then the following statements are equivalent:

(a) $\tilde{\boldsymbol{\beta}}_1(\mathscr{M}_{12}) = \hat{\boldsymbol{\beta}}_1(\mathscr{M}_{12})$, *i.e.,* $\operatorname{eff}(\hat{\boldsymbol{\beta}}_1 \mid \mathscr{M}_{12}) = 1$,

(b) $\tilde{\boldsymbol{\beta}}_2(\mathscr{M}_{12}) = \hat{\boldsymbol{\beta}}_2(\mathscr{M}_{12})$, *i.e.,* $\operatorname{eff}(\hat{\boldsymbol{\beta}}_2 \mid \mathscr{M}_{12}) = 1$,

(c) $\tilde{\boldsymbol{\beta}}(\mathscr{M}_{12}) = \hat{\boldsymbol{\beta}}(\mathscr{M}_{12})$, *i.e.,* $\operatorname{eff}(\hat{\boldsymbol{\beta}} \mid \mathscr{M}_{12}) = 1$.

3 The situation when $\mathbf{X}_2 = \mathbf{x}_p$

It is of interest to comment on a special case of Corollary 2.1, where $\mathbf{X}_2$ comprises just one column and so $p_2 = 1$. We write $\mathbf{X}_2 = \mathbf{x}_p$. The result is presented in the following corollary.

Corollary 3.1. *Consider a weakly singular partitioned linear model* $\mathscr{M}_{12} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\}$, *where* $\mathbf{X} = (\mathbf{X}_1 : \mathbf{x}_p)$ *has full column rank. Assume that*

$$\text{(i) } \operatorname{eff}(\hat{\boldsymbol{\beta}}_1 \mid \mathscr{M}_1) = 1, \textit{ and } \text{ (ii) } \mathbf{X}_1'\mathbf{x}_p \neq \mathbf{0}. \tag{3.1}$$

Then the following statements are equivalent:

(a) $\tilde{\boldsymbol{\beta}}_1(\mathscr{M}_{12}) = \hat{\boldsymbol{\beta}}_1(\mathscr{M}_{12})$,

(b) $\tilde{\beta}_p(\mathscr{M}_{12}) = \hat{\beta}_p(\mathscr{M}_{12})$,

(c) $\tilde{\boldsymbol{\beta}}(\mathscr{M}_{12}) = \hat{\boldsymbol{\beta}}(\mathscr{M}_{12})$,

(d) $\mathbf{M}_1\mathbf{V}\mathbf{M}_1\mathbf{x}_p = \lambda^2\mathbf{M}_1\mathbf{x}_p$ *for some nonzero* $\lambda \in \mathbb{R}$.

Proof. Corollary 2.1 implies immediately the equivalence of (a), (b) and (c). The last statement (d) comes from part (d) of Lemma 1.1. This is so because in view of part (d) of Lemma 1.1, the vector $\mathbf{M}_1\mathbf{x}_p$ must be an eigenvector of $\mathbf{M}_1\mathbf{V}\mathbf{M}_1$, i.e., the equality

$$\mathbf{M}_1\mathbf{V}\mathbf{M}_1\mathbf{x}_p = \lambda^2\mathbf{M}_1\mathbf{x}_p \tag{3.2}$$

must hold for some λ. It is easy to see that the scalar λ in (3.2) is necessarily nonzero once we require $\mathbf{M}_1\mathbf{x}_p \neq \mathbf{0}$ and $\mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{V})$. Note that $\mathbf{M}_1\mathbf{x}_p \neq \mathbf{0}$ is equivalent to $\mathbf{x}_p \notin \mathscr{C}(\mathbf{X}_1)$ which of course holds since we assume $\mathbf{X}$ to have full column rank. □

There is one particular choice of $\mathbf{x}_p$ that deserves special attention,

$$\mathbf{x}_p = \mathbf{e}_i = i\text{th column of } \mathbf{I}_n. \tag{3.3}$$

Let us now consider three (weakly singular) linear models:

$$\mathscr{M} = \{\mathbf{y}, \mathbf{X}\boldsymbol{\beta}, \mathbf{V}\},\ \mathscr{M}_{(i)} = \{\mathbf{y}_{(i)}, \mathbf{X}_{(i)}\boldsymbol{\beta}, \mathbf{V}_{(i)}\},\ \mathscr{M}_Z = \{\mathbf{y}, \mathbf{Z}\boldsymbol{\gamma}, \mathbf{V}\}, \tag{3.4}$$

where

$$\mathbf{Z} = (\mathbf{X} : \mathbf{e}_i), \quad \boldsymbol{\gamma} = \begin{pmatrix} \boldsymbol{\beta} \\ \delta \end{pmatrix}. \tag{3.5}$$

By $\mathscr{M}_{(i)}$ we mean such a version of $\mathscr{M}$ in which the ith case (ith observation) is deleted; thus $\mathbf{y}_{(i)}$ has $n-1$ elements, $\mathbf{X}_{(i)}$ has $n-1$ rows, and $\operatorname{cov}(\mathbf{y}_{(i)}) = \mathbf{V}_{(i)}$. The partitioned model $\mathscr{M}_Z$ is an *extended* version of $\mathscr{M}$. The extended model $\mathscr{M}_Z$ appears to be very useful in calculating regression diagnostics; see, e.g., Beckman and Cook (1983), and Chatterjee and Hadi (1986).

We assume that the model $\mathscr{M}_Z$ is a weakly singular model and that $\mathbf{Z}$ has full column rank. Then, as pointed out by Puntanen (1996, Th. 3), the model $\mathscr{M}_{(i)}$ is a reduced version of $\mathscr{M}_Z$; in other words, if $\mathscr{M}_Z$ corresponds to $\mathscr{M}_{12}$, then $\mathscr{M}_{(i)}$ and $\mathscr{M}$ correspond to $\mathscr{M}_{12\cdot 2}$ and $\mathscr{M}_1$, respectively. Hence we have, in short notation,

$$\hat{\boldsymbol{\beta}}(\mathscr{M}_Z) = \hat{\boldsymbol{\beta}}(\mathscr{M}_{(i)}) = \hat{\boldsymbol{\beta}}_{(i)}, \quad \tilde{\boldsymbol{\beta}}(\mathscr{M}_Z) = \tilde{\boldsymbol{\beta}}(\mathscr{M}_{(i)}) = \tilde{\boldsymbol{\beta}}_{(i)}. \tag{3.6}$$

Using Corollary 3.1, we now obtain immediately the following result, which is a generalized version of a result of Nurhonen & Puntanen (1992, p. 133) who assumed $\mathbf{V}$ to be positive definite and used a different approach.

Corollary 3.2. *Consider a weakly singular linear model*

$$\mathscr{M}_Z = \{\mathbf{y},\, \mathbf{Z}\boldsymbol{\gamma},\, \mathbf{V}\} = \{\mathbf{y},\, \mathbf{Z}\begin{pmatrix}\boldsymbol{\beta}\\ \delta\end{pmatrix},\, \mathbf{V}\}, \tag{3.7}$$

where $\mathbf{Z} = (\mathbf{X} : \mathbf{e}_i)$ *has full column rank, and denote* $\mathscr{M} = \{\mathbf{y},\, \mathbf{X}\boldsymbol{\beta},\, \mathbf{V}\}$, *and let* $\mathscr{M}_{(i)} = \{\mathbf{y}_{(i)},\, \mathbf{X}_{(i)}\boldsymbol{\beta},\, \mathbf{V}_{(i)}\}$ *be such a version of* $\mathscr{M}$ *in which the* ith *observation is deleted. Assume* $\mathbf{X}'\mathbf{e}_i \neq \mathbf{0}$ $(i = 1, \dots, n)$, *and that* OLSE($\boldsymbol{\beta}$) *equals* BLUE($\boldsymbol{\beta}$) *under* $\mathscr{M}$. *Then*

$$\hat{\boldsymbol{\beta}}(\mathscr{M}_{(i)}) = \tilde{\boldsymbol{\beta}}(\mathscr{M}_{(i)}) \text{ holds for all } i = 1, \dots, n, \tag{3.8}$$

if and only if $\mathbf{V}$ *satisfies*

$$\mathbf{MVM} = \lambda^2\mathbf{M}, \text{ for some nonzero scalar } \lambda. \tag{3.9}$$

Acknowledgements

The research of the third author was supported in part by the Natural Sciences and Engineering Research Council of Canada.

References

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

Beckman, R. J. & Cook, R. D. (1983). Outlier..............s (with discussion). *Technometrics*, 25, 119–163.

Chatterjee, S. & Hadi, A. S. (1986). Influential observations, high leverage points, and outliers in linear regression (with discussion). *Statistical Science*, 1, 379–416.

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.

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.

Davidson, R. & MacKinnon, J. G. (2004). *Econometric Theory and Methods.* Oxford University Press, New York.

Groß, J. & Puntanen, S. (2000). Estimation under a general partitioned linear model. *Linear Algebra and Its Applications*, 321, 131–144.

Marsaglia, G. & Styan, G. P. H. (1974). Equalities and inequalities for ranks of matrices. *Linear and Multilinear Algebra*, 2, 269–292.

Nurhonen, M. & Puntanen, S. (1992). Effect of deleting an observation on the equality of the OLSE and BLUE. *Linear Algebra and Its Applications*, 176, 131–136.

Puntanen, S. (1996). Some matrix results related to a partitioned singular linear model. *Communications in Statistics: Theory and Methods*, 25, 269–279.

Puntanen, S. & Styan, G. P. H. (1989). The equality of the ordinary least squares estimator and the best linear unbiased estimator (with discussion). *The American Statistician*, 43, 153–164.

Rao, C. R. (1967). Least squares theory using an estimated dispersion matrix and its application to measurement of signals. *Proc. Fifth Berkeley Symposium on Mathematical Statistics and Probability: Berkeley, California, 1965/1966*, vol. 1, L.M. Le Cam and J. Neyman, eds., Univ. of California Press, Berkeley, 355–372.

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

Sengupta, D. & Jammalamadaka, S. R. (2003). *Linear Models: an Integrated Approach.* World Scientific, Singapore.

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

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

Zyskind, G. & Martin, F. B. (1969). On best linear estimation and a general Gauss–Markov theorem in linear models with arbitrary nonnegative covariance structure. *SIAM Journal on Applied Mathematics*, 17, 1190–1202.

Jarkko Isotalo and Simo Puntanen
Department of Mathematics,
Statistics and Philosophy
FI-33014 University of Tampere
Finland

E-mail: jarkko.isotalo@uta.fi
E-mail: simo.puntanen@uta.fi

George P. H. Styan
Department of Mathematics and Statistics
McGill University
805 ouest, rue Sherbrooke Street West
Montréal (Québec), Canada H3A 2K6

E-mail: styan@math.mcgill.ca