

The efficiency factorization multiplier for the Watson efficiency in partitioned linear models: Some examples and a literature review

KA LOK CHU,¹ JARKKO ISOTALO,² SIMO PUNTANEN³
& GEORGE P. H. STYAN⁴

¹*Department of Mathematics, Dawson College,
McGill University, Montréal, Québec, Canada*
chu@math.mcgill.ca

^{2,3}*Department of Mathematics, Statistics & Philosophy,
University of Tampere, Tampere, Finland*
² jarkko.isotalo@uta.fi; ³ simo.puntanen@uta.fi

⁴*Department of Mathematics and Statistics,
McGill University, Montréal, Québec, Canada*
styan@math.mcgill.ca

We consider partitioned linear models where the model matrix $\mathbf{X} = (\mathbf{X}_1 : \mathbf{X}_2)$ has full column rank, and concentrate on the special case where $\mathbf{X}_1' \mathbf{X}_2 = \mathbf{0}$ when we say that the model is *orthogonally partitioned*. We assume that the underlying covariance matrix is positive definite and introduce the *efficiency factorization multiplier* which relates the total Watson efficiency of ordinary least squares to the product of the two subset Watson efficiencies. We illustrate our findings with several examples and present a literature review.

1 Introduction and mise-en-scène

In this paper we consider the general partitioned linear (or Gauß–Markov) model

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

or in another notation,

$$\mathcal{M}_{12} := \{\mathbf{y}, \mathbf{X}_{12} \boldsymbol{\beta}_{12}, \mathbf{V}\} := \{\mathbf{y}, \mathbf{X}_1 \boldsymbol{\beta}_1 + \mathbf{X}_2 \boldsymbol{\beta}_2, \mathbf{V}\}, \quad (1.2)$$

with

$$E(\mathbf{y}) = \mathbf{X}_{12}\boldsymbol{\beta}_{12}, \quad E(\boldsymbol{\varepsilon}) = \mathbf{0}, \quad \text{cov}(\mathbf{y}) = \text{cov}(\boldsymbol{\varepsilon}) = \mathbf{V}, \quad (1.3)$$

where $E(\cdot)$ denotes expectation (or expected value) and $\text{cov}(\cdot)$ denotes the covariance (or dispersion) matrix. The vector \mathbf{y} is an $n \times 1$ observable random vector, $\boldsymbol{\varepsilon}$ is an $n \times 1$ unobservable random error vector, and

$$\boldsymbol{\beta}_{12} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix} \quad (1.4)$$

is a $p_{12} \times 1$ vector of unknown parameters with $p_{12} = p_1 + p_2 = p$, with $\boldsymbol{\beta}_1$ ($p_1 \times 1$) and $\boldsymbol{\beta}_2$ ($p_2 \times 1$). The model (or design) matrix \mathbf{X}_{12} is $n \times p_{12}$ and is partitioned columnwise as

$$\mathbf{X}_{12} = (\mathbf{X}_1 : \mathbf{X}_2), \quad (1.5)$$

with $p_{12} = p_1 + p_2$, \mathbf{X}_1 ($n \times p_1$) and \mathbf{X}_2 ($n \times p_2$). Both the model matrix \mathbf{X}_{12} and the covariance matrix \mathbf{V} are known. Usually the model matrix \mathbf{X}_{12} is denoted by just \mathbf{X} and the number of its columns by just p .

When the model matrix \mathbf{X}_{12} has full column rank and the covariance matrix \mathbf{V} is positive definite, then as is well known, the vector $\boldsymbol{\beta}_{12}$ is estimable, and the OLSE (ordinary least squares estimator) and the BLUE (best linear unbiased estimator or Aitken estimator) of $\boldsymbol{\beta}_{12}$ under the full model \mathcal{M}_{12} are, respectively,

$$\text{OLSE}(\boldsymbol{\beta}_{12} | \mathcal{M}_{12}) = \hat{\boldsymbol{\beta}}_{12} = \begin{pmatrix} \hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_2 \end{pmatrix} = (\mathbf{X}'_{12}\mathbf{X}_{12})^{-1}\mathbf{X}'_{12}\mathbf{y} \quad (1.6a)$$

$$\text{BLUE}(\boldsymbol{\beta}_{12} | \mathcal{M}_{12}) = \tilde{\boldsymbol{\beta}}_{12} = \begin{pmatrix} \tilde{\boldsymbol{\beta}}_1 \\ \tilde{\boldsymbol{\beta}}_2 \end{pmatrix} = (\mathbf{X}'_{12}\mathbf{V}^{-1}\mathbf{X}_{12})^{-1}\mathbf{X}'_{12}\mathbf{V}^{-1}\mathbf{y}, \quad (1.6b)$$

with $(\cdot)'$ denoting transpose. The corresponding covariance matrices are, respectively,

$$\text{cov}(\hat{\boldsymbol{\beta}}_{12} | \mathcal{M}_{12}) = (\mathbf{X}'_{12}\mathbf{X}_{12})^{-1}\mathbf{X}'_{12}\mathbf{V}\mathbf{X}_{12}(\mathbf{X}'_{12}\mathbf{X}_{12})^{-1} \quad (1.7a)$$

$$\text{cov}(\tilde{\boldsymbol{\beta}}_{12} | \mathcal{M}_{12}) = (\mathbf{X}'_{12}\mathbf{V}^{-1}\mathbf{X}_{12})^{-1}, \quad (1.7b)$$

and hence from the Gauß–Markov theorem [43], we have the Löwner ordering, see, e.g., Wang & Chow [55, p. 207],

$$\text{cov}(\hat{\boldsymbol{\beta}}_{12} | \mathcal{M}_{12}) \geq_L \text{cov}(\tilde{\boldsymbol{\beta}}_{12} | \mathcal{M}_{12}), \quad (1.8)$$

or equivalently, the matrix difference between the two matrices in (1.8) is nonnegative definite.

There is no unique way to measure how “bad” the OLSE could be with respect to the BLUE. Almost certainly the most frequently used measure is the *Watson efficiency* which

is defined as the ratio of the generalized variances (determinants of the corresponding covariance matrices) in (1.7a) and (1.7b):

$$\text{eff}(\hat{\beta}_{12} | \mathcal{M}_{12}) = \frac{|\text{cov}(\tilde{\beta}_{12} | \mathcal{M}_{12})|}{|\text{cov}(\hat{\beta}_{12} | \mathcal{M}_{12})|} = \frac{|\mathbf{X}'_{12}\mathbf{X}_{12}|^2}{|\mathbf{X}'_{12}\mathbf{V}\mathbf{X}_{12}| \cdot |\mathbf{X}'_{12}\mathbf{V}^{-1}\mathbf{X}_{12}|}, \quad (1.9)$$

where $|\cdot|$ denotes determinant. We call $\text{eff}(\hat{\beta}_{12} | \mathcal{M}_{12})$ as defined by (1.9) the *total Watson efficiency* of the OLSE $\hat{\beta}_{12}$ in the model \mathcal{M}_{12} .

We define the *subset Watson efficiency* of the OLSE $\hat{\beta}_i$ ($i = 1, 2$) in the model \mathcal{M}_{12}

$$\text{eff}(\hat{\beta}_1 | \mathcal{M}_{12}) = \frac{|\text{cov}(\tilde{\beta}_1 | \mathcal{M}_{12})|}{|\text{cov}(\hat{\beta}_1 | \mathcal{M}_{12})|}, \quad \text{eff}(\hat{\beta}_2 | \mathcal{M}_{12}) = \frac{|\text{cov}(\tilde{\beta}_2 | \mathcal{M}_{12})|}{|\text{cov}(\hat{\beta}_2 | \mathcal{M}_{12})|} \quad (1.10)$$

and the *efficiency factorization multiplier* γ_{12} by

$$\gamma_{12} = \frac{\text{eff}(\hat{\beta}_{12} | \mathcal{M}_{12})}{\text{eff}(\hat{\beta}_1 | \mathcal{M}_{12}) \cdot \text{eff}(\hat{\beta}_2 | \mathcal{M}_{12})}, \quad (1.11)$$

or equivalently by

$$\text{eff}(\hat{\beta}_{12} | \mathcal{M}_{12}) = \gamma_{12} \cdot \text{eff}(\hat{\beta}_1 | \mathcal{M}_{12}) \cdot \text{eff}(\hat{\beta}_2 | \mathcal{M}_{12}). \quad (1.12)$$

We are interested in characterizing $\gamma_{12} > 1$, $\gamma_{12} = 1$, or $\gamma_{12} < 1$. When $\gamma_{12} = 1$, i.e.,

$$\text{eff}(\hat{\beta}_{12} | \mathcal{M}_{12}) = \text{eff}(\hat{\beta}_1 | \mathcal{M}_{12}) \cdot \text{eff}(\hat{\beta}_2 | \mathcal{M}_{12}), \quad (1.13)$$

then we say that the Watson efficiency *factorizes*. Since the Watson efficiency is nonnegative and can never exceed 1, it follows at once from (1.13) that

$$\text{eff}(\hat{\beta}_{12} | \mathcal{M}_{12}) = 1 \ \& \ \gamma_{12} = 1 \quad \Rightarrow \quad \text{eff}(\hat{\beta}_1 | \mathcal{M}_{12}) = 1 \ \& \ \text{eff}(\hat{\beta}_2 | \mathcal{M}_{12}) = 1. \quad (1.14)$$

But we may strengthen the result (1.14) to

$$\text{eff}(\hat{\beta}_{12} | \mathcal{M}_{12}) = 1 \quad \Rightarrow \quad \gamma_{12} = 1 \ \& \ \text{eff}(\hat{\beta}_1 | \mathcal{M}_{12}) = 1 \ \& \ \text{eff}(\hat{\beta}_2 | \mathcal{M}_{12}) = 1. \quad (1.15)$$

To prove (1.15), we note that from (1.8) it follows at once that

$$|\text{cov}(\hat{\beta}_{12} | \mathcal{M}_{12})| \geq |\text{cov}(\tilde{\beta}_{12} | \mathcal{M}_{12})| \quad (1.16)$$

with equality if and only if the covariance matrices in (1.16) are equal, see, e.g., Marshall & Olkin [42, (1979)], and then in the model \mathcal{M}_{12}

$$\hat{\beta}_{12} = \tilde{\beta}_{12} \quad (1.17)$$

with probability 1. This result together with the definition (1.11) establishes (1.15).

When

$$\text{eff}(\hat{\beta}_{12} | \mathcal{M}_{12}) = \text{eff}(\hat{\beta}_1 | \mathcal{M}_{12}) \Leftrightarrow \gamma_{12} \cdot \text{eff}(\hat{\beta}_2 | \mathcal{M}_{12}) = 1, \quad (1.18)$$

or when

$$\text{eff}(\hat{\beta}_{12} | \mathcal{M}_{12}) = \text{eff}(\hat{\beta}_2 | \mathcal{M}_{12}) \Leftrightarrow \gamma_{12} \cdot \text{eff}(\hat{\beta}_1 | \mathcal{M}_{12}) = 1, \quad (1.19)$$

then we say that there is a *reduction of the Watson efficiency of type 1*. When

$$\text{eff}(\hat{\beta}_{12} | \mathcal{M}_{12}) = \gamma_{12} \cdot \text{eff}(\hat{\beta}_1 | \mathcal{M}_{12}) \Leftrightarrow \text{eff}(\hat{\beta}_2 | \mathcal{M}_{12}) = 1, \quad (1.20)$$

or when

$$\text{eff}(\hat{\beta}_{12} | \mathcal{M}_{12}) = \gamma_{12} \cdot \text{eff}(\hat{\beta}_2 | \mathcal{M}_{12}) \Leftrightarrow \text{eff}(\hat{\beta}_1 | \mathcal{M}_{12}) = 1, \quad (1.21)$$

then we say that there is a *reduction of the Watson efficiency of type 2*.

We illustrate these formulas with several examples and review the relevant literature. For further related results see Chu [16] and Chu, Isotalo, Puntanen & Styan [17, 18, 19].

2 Examples

Example 2.1: A simple example with $n = 4$ and $p = 2$

For our first example let us consider the model matrix

$$\mathbf{X}_{12} = (\mathbf{X}_1 : \mathbf{X}_2) = \begin{pmatrix} 1 & : & -1 \\ 1 & : & -2 \\ 1 & : & +2 \\ 1 & : & +1 \end{pmatrix} \quad (2.1)$$

and the covariance matrix

$$\mathbf{V} = \begin{pmatrix} 3 & 1 & 1 & 3\rho \\ 1 & 3 & 1 & 1 \\ 1 & 1 & 3 & 1 \\ 3\rho & 1 & 1 & 3 \end{pmatrix}. \quad (2.2)$$

Then the matrix \mathbf{V} in (2.2) is positive definite whenever

$$-\frac{2}{3} < \rho < +1. \quad (2.3)$$

To prove (2.3) we may use Haynsworth's inertia additivity formula [48, §0.10] which shows that the matrix \mathbf{V} in (2.2) is positive definite provided the Schur complement of the top left 3×3 submatrix

$$\begin{aligned} 3 - (3\rho \ 1 \ 1) \begin{pmatrix} 3 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 3 \end{pmatrix}^{-1} \begin{pmatrix} 3\rho \\ 1 \\ 1 \end{pmatrix} &= 3 - (3\rho \ 1 \ 1) \frac{1}{10} \begin{pmatrix} 4 & -1 & -1 \\ -1 & 4 & -1 \\ -1 & -1 & 4 \end{pmatrix} \begin{pmatrix} 3\rho \\ 1 \\ 1 \end{pmatrix} \\ &= \frac{6}{5}(1 - \rho)(2 + 3\rho) > 0, \end{aligned} \quad (2.4)$$

which establishes (2.3).

With the model matrix \mathbf{X}_{12} as in (2.1), and with the covariance matrix \mathbf{V} as in (2.2) and positive definite, we obtain the variances

$$\text{var}(\tilde{\beta}_1) = \frac{2(2 + 3\rho)}{3(1 + \rho)} \quad \text{and} \quad \text{var}(\hat{\beta}_1) = \frac{11 + 3\rho}{8} \quad (2.5a)$$

$$\text{var}(\tilde{\beta}_2) = \frac{3(1 - \rho)}{2(7 - 6\rho)} \quad \text{and} \quad \text{var}(\hat{\beta}_2) = \frac{11 - 3\rho}{50}. \quad (2.5b)$$

The covariances $\text{cov}(\tilde{\beta}_1, \tilde{\beta}_2) = \text{cov}(\hat{\beta}_1, \hat{\beta}_2) = 0$, and so the generalized variances are the products of the corresponding variances:

$$|\text{cov}(\tilde{\boldsymbol{\beta}}_{12})| = \frac{(2 + 3\rho)(1 - \rho)}{(1 + \rho)(7 - 6\rho)} \quad \text{and} \quad |\text{cov}(\hat{\boldsymbol{\beta}}_{12})| = \frac{(11 + 3\rho)(11 - 3\rho)}{400}. \quad (2.6)$$

To ease the notation, we write here

$$\text{var}(\tilde{\beta}_i) = \text{var}(\tilde{\beta}_i \mid \mathcal{M}_{12}), \quad \text{var}(\hat{\beta}_i) = \text{var}(\hat{\beta}_i \mid \mathcal{M}_{12}), \quad i = 1, 2, \quad (2.7a)$$

$$\text{cov}(\tilde{\beta}_1, \tilde{\beta}_2) = \text{cov}(\tilde{\beta}_1, \tilde{\beta}_2 \mid \mathcal{M}_{12}), \quad \text{cov}(\hat{\beta}_1, \hat{\beta}_2) = \text{cov}(\hat{\beta}_1, \hat{\beta}_2 \mid \mathcal{M}_{12}) \quad (2.7b)$$

$$\text{cov}(\tilde{\boldsymbol{\beta}}_{12}) = \text{cov}(\tilde{\boldsymbol{\beta}}_{12} \mid \mathcal{M}_{12}), \quad \text{cov}(\hat{\boldsymbol{\beta}}_{12}) = \text{cov}(\hat{\boldsymbol{\beta}}_{12} \mid \mathcal{M}_{12}). \quad (2.7c)$$

The subset Watson efficiencies are

$$\text{eff}(\hat{\beta}_1) = \frac{16(2 + 3\rho)}{3(1 + \rho)(11 + 3\rho)} \quad \text{and} \quad \text{eff}(\hat{\beta}_2) = \frac{75(1 - \rho)}{(7 - 6\rho)(11 - 3\rho)} \quad (2.8)$$

and the total Watson efficiency is

$$\text{eff}(\hat{\boldsymbol{\beta}}_{12}) = \frac{400(2 + 3\rho)(1 - \rho)}{(1 + \rho)(11 + 3\rho)(7 - 6\rho)(11 - 3\rho)} = \text{eff}(\hat{\beta}_1) \cdot \text{eff}(\hat{\beta}_2), \quad (2.9)$$

the product of the two subset Watson efficiencies in (2.8), and so in this example the efficiency factorization multiplier $\gamma_{12} = 1$ and the total Watson efficiency factorizes for all ρ such that the matrix \mathbf{V} in (2.2) is positive definite.

It is interesting to note that in this example with $\gamma_{12} = 1$, we have

$$\text{eff}(\hat{\boldsymbol{\beta}}_{12}) = 1 \Leftrightarrow \text{eff}(\hat{\beta}_1) = 1 \Leftrightarrow \text{eff}(\hat{\beta}_2) = 1 \Leftrightarrow \rho = \frac{1}{3} \quad (2.10)$$

and so there is then (with $\rho = 1/3$) a reduction of the total Watson efficiency here of both type 1 and type 2, see (1.18)–(1.21) above.

Let us now consider the model matrix

$$\mathbf{X}_{12} = (\mathbf{X}_1 : \mathbf{X}_2) = \begin{pmatrix} 1 & : & -1 \\ 1 & : & +1 \\ 1 & : & +2 \\ 1 & : & -2 \end{pmatrix}, \quad (2.11)$$

which is the matrix \mathbf{X} in (2.1) with rows 2 and 4 switched, and the covariance matrix

$$\mathbf{V} = \begin{pmatrix} 3 & 1 & 1 & 3\rho \\ 1 & 3 & 1 & 1 \\ 1 & 1 & 3 & 1 \\ 3\rho & 1 & 1 & 3 \end{pmatrix}, \quad (2.12)$$

which is the same as in (2.2) above. We recall that \mathbf{V} in (2.12) is positive definite provided that $-2/3 < \rho < 1$, see (2.3) above.

We obtain the variances

$$\text{var}(\tilde{\beta}_1) = \frac{317 - 198\rho - 99\rho^2}{3(77 - 70\rho - 3\rho^2)} \quad \text{and} \quad \text{var}(\hat{\beta}_1) = \frac{11 + 3\rho}{8} \quad (2.13a)$$

$$\text{var}(\tilde{\beta}_2) = \frac{12(1 - \rho^2)}{77 - 70\rho - 3\rho^2} \quad \text{and} \quad \text{var}(\hat{\beta}_2) = \frac{4 + 3\rho}{25}, \quad (2.13b)$$

and hence the subset Watson efficiencies are

$$\text{eff}(\hat{\beta}_1) = \frac{8(317 - 198\rho - 99\rho^2)}{3(77 - 70\rho - 3\rho^2)(11 + 3\rho)} \quad \text{and} \quad \text{eff}(\hat{\beta}_2) = \frac{300(1 - \rho^2)}{(77 - 70\rho - 3\rho^2)(4 + 3\rho)}. \quad (2.14)$$

The generalized variances are

$$|\text{cov}(\tilde{\boldsymbol{\beta}}_{12})| = \frac{8(2 + 3\rho)(1 - \rho)}{77 - 70\rho - 3\rho^2} \quad \text{and} \quad |\text{cov}(\hat{\boldsymbol{\beta}}_{12})| = \frac{343 + 414\rho - 9\rho^2}{1600} \quad (2.15)$$

and so the total Watson efficiency is

$$\text{eff}(\hat{\boldsymbol{\beta}}_{12}) = \frac{12800(3\rho + 2)(1 - \rho)}{(3\rho^2 + 70\rho - 77)(343 + 414\rho - 9\rho^2)}. \quad (2.16)$$

The efficiency factorization multiplier

$$\gamma = \frac{\text{eff}(\hat{\beta}_{12})}{\text{eff}(\hat{\beta}_1) \cdot \text{eff}(\hat{\beta}_2)} = \frac{16(2 + 3\rho)(4 + 3\rho)(11 + 3\rho)(77 - 70\rho - 3\rho^2)}{(1 + \rho)(317 - 198\rho - 99\rho^2)(343 + 414\rho - 9\rho^2)}, \quad (2.17)$$

which seemingly does not simplify (further). We plot γ vs. ρ in Figure 2-1a and note that γ increases monotonically with ρ , and that $\gamma < 1$ for $\rho < 0$. Interestingly when $\rho = 0$, then

$$\gamma = \frac{108,416}{108,731} \approx 0.9971, \quad (2.18)$$

and $\gamma = 1$ when $\rho = 1/3$, that is, \mathbf{V} has intraclass correlation.

TABLE 2-2a: Coded data for an experiment with a lathe due to M. R. Delozier.

<i>S</i>	<i>F</i>	<i>Y</i>	<i>S</i>	<i>F</i>	<i>Y</i>
-1	-1	54.5	$-\sqrt{2}$	0	20.1
-1	-1	66.0	$+\sqrt{2}$	0	2.9
1	-1	11.8	0	0	3.8
1	-1	14.0	0	0	2.2
-1	1	5.2	0	0	3.2
-1	1	3.0	0	0	4.0
1	1	0.8	0	0	2.8
1	1	0.5	0	0	3.2
0	$-\sqrt{2}$	86.5	0	0	4.0
0	$+\sqrt{2}$	0.4	0	0	3.5

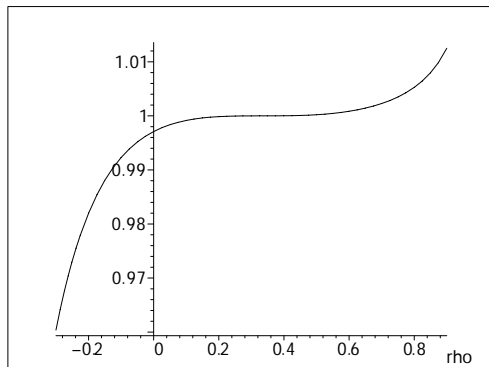


FIGURE 2-1a: Plot of the efficiency factorization multiplier γ (vertical axis) vs. the correlation ρ (horizontal axis) for the example with \mathbf{X} given by (2.11) and \mathbf{V} given by (2.12).

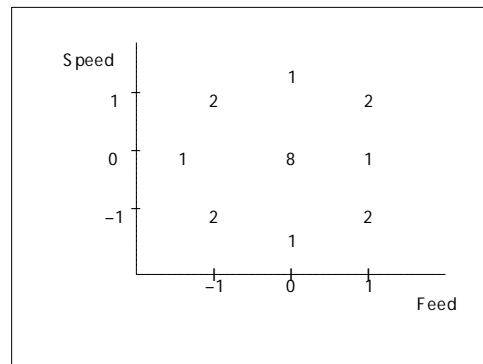


FIGURE 2-2a: Scatter plot of S vs. F for the lathe data in Table 2-2a. Numbers in the plots give the numbers of runs at a given combination of S and F .

$$\begin{pmatrix} 1 & 1 & 1 & : & -1 & -1 & 1 \\ 1 & 1 & 1 & : & -1 & -1 & 1 \\ 1 & 1 & 1 & : & 1 & -1 & -1 \\ 1 & 1 & 1 & : & 1 & -1 & -1 \\ 1 & 1 & 1 & : & -1 & 1 & -1 \\ 1 & 1 & 1 & : & -1 & 1 & -1 \\ 1 & 1 & 1 & : & 1 & 1 & 1 \\ 1 & 1 & 1 & : & 1 & 1 & 1 \\ 1 & 0 & 2 & : & 0 & -\sqrt{2} & 0 \\ 1 & 0 & 2 & : & 0 & \sqrt{2} & 0 \\ 1 & 2 & 0 & : & -\sqrt{2} & 0 & 0 \\ 1 & 2 & 0 & : & \sqrt{2} & 0 & 0 \\ 1 & 0 & 0 & : & 0 & 0 & 0 \\ 1 & 0 & 0 & : & 0 & 0 & 0 \\ 1 & 0 & 0 & : & 0 & 0 & 0 \\ 1 & 0 & 0 & : & 0 & 0 & 0 \\ 1 & 0 & 0 & : & 0 & 0 & 0 \\ 1 & 0 & 0 & : & 0 & 0 & 0 \\ 1 & 0 & 0 & : & 0 & 0 & 0 \\ 1 & 0 & 0 & : & 0 & 0 & 0 \end{pmatrix}$$

FIGURE 2-2b: The model matrix $\mathbf{X} = (\mathbf{X}_1 : \mathbf{X}_2)$ associated with the lathe data.

Example 2.2: Delozier's lathe data

For our second example we consider the lathe data of M. R. Delozier as discussed by Weisberg [60, pp. 166–167 (1985)]:

The data [in Table 2-2a] are the results of an experiment to characterize the performance of a cutting-tool material in cutting steel on a lathe. A completely randomized experiment in 20 runs was used, with two factors, cutting

speed (in feet per minute) and feed rate (in thousandths of an inch per revolution). For convenience, the levels of the two factors are coded and centered to give predictors $S=(\text{speed}-900)/300$ and $F=(\text{feed rate}-13)/6$. The response was $Y = \text{tool life (in minutes)}$. Figure 2-2a is a scatter plot of S vs. F ; the numbers on the plot correspond to the number of runs at each of the experimental settings. This layout of points is called a *central composite design* and is useful when the fitting of polynomials, and possibly interactions, is anticipated.

Weisberg [60, p. 167, eq. (7.9)] suggests fitting the full second-order model in S and F to $\log(Y)$ which we set up as

$$\log(Y) = \beta_0 + \beta_1 S^2 + \beta_2 F^2 + \beta_3 S + \beta_4 F + \beta_5 S \times F + \varepsilon \tag{2.19}$$

so that the associated 20×6 model matrix $\mathbf{X} = (\mathbf{X}_1 : \mathbf{X}_2)$, with \mathbf{X}_1 and \mathbf{X}_2 both 20×3 , is orthogonally partitioned, i.e., $\mathbf{X}'_1 \mathbf{X}_2 = \mathbf{0}$. The matrix \mathbf{X} is shown in Figure 2-2b and we find

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} \mathbf{X}'_1 \mathbf{X}_1 & \mathbf{X}'_1 \mathbf{X}_2 \\ \mathbf{X}'_2 \mathbf{X}_1 & \mathbf{X}'_2 \mathbf{X}_2 \end{pmatrix} = \begin{pmatrix} 20 & 12 & 12 & 0 & 0 & 0 \\ 12 & 16 & 8 & 0 & 0 & 0 \\ 12 & 8 & 16 & 0 & 0 & 0 \\ 0 & 0 & 0 & 12 & 0 & 0 \\ 0 & 0 & 0 & 0 & 12 & 0 \\ 0 & 0 & 0 & 0 & 0 & 8 \end{pmatrix}, \tag{2.20}$$

and so $\mathbf{X}'_1 \mathbf{X}_2 = \mathbf{0}$ and \mathbf{X} in Figure 2-2b is orthogonally partitioned.

We now suppose that the 20 observations have been made sequentially in time with serial correlation coefficient ρ and hence that the covariance matrix

$$\mathbf{V} = \frac{1}{1 - \rho^2} \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{19} \\ \rho & 1 & \rho & \dots & \rho^{18} \\ \rho^2 & \rho & 1 & \dots & \rho^{17} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{19} & \rho^{18} & \rho^{17} & \dots & 1 \end{pmatrix} = \mathbf{A}'\mathbf{A}, \tag{2.21}$$

where

$$\mathbf{A} = \begin{pmatrix} \lambda & \lambda\rho & \lambda\rho^2 & \dots & \lambda\rho^{19} \\ 0 & 1 & \rho & \dots & \rho^{18} \\ 0 & 0 & 1 & \dots & \rho^{17} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}, \tag{2.22}$$

with $\lambda = 1/\sqrt{1 - \rho^2}$, see, e.g., Ward [56, pp. 25–26 (1973)]. It follows at once that the determinant

$$|\mathbf{V}| = \lambda^2 = \frac{1}{1 - \rho^2} \tag{2.23}$$

and the inverse

$$\mathbf{V}^{-1} = \begin{pmatrix} 1 & -\rho & 0 & 0 & \dots & \dots & 0 \\ -\rho & 1 + \rho^2 & -\rho & 0 & \dots & \dots & 0 \\ 0 & -\rho & 1 + \rho^2 & -\rho & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -\rho & 1 + \rho^2 & -\rho \\ 0 & 0 & 0 & \dots & 0 & -\rho & 1 \end{pmatrix}, \quad (2.24)$$

see, e.g., Chipman [15, p. 4 (1965)].

We plot the subset Watson efficiencies $\text{eff}(\hat{\beta}_1)$ and $\text{eff}(\hat{\beta}_2)$ (vertical axis) against the serial correlation coefficient ρ (horizontal axis) in Figure 2-2c and observe that for $-0.8 \leq \rho \leq +0.8$, the subset Watson efficiencies follow a parabola with maximum efficiency equal to 1 at $\rho = 0$ and minimum efficiencies $\text{eff}(\hat{\beta}_1) \approx 0.329$ and $\text{eff}(\hat{\beta}_2) \approx 0.099$ and $\text{eff}(\hat{\beta}_1) \approx 0.502$ at $\rho = -0.8$ and $\text{eff}(\hat{\beta}_2) \approx 0.159$ at $\rho = +0.8$.

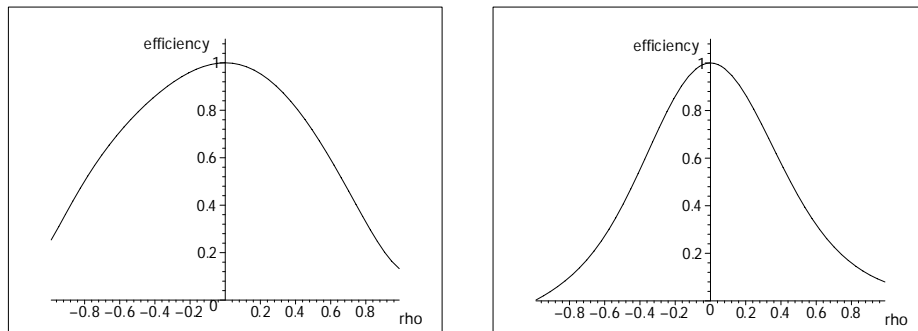


FIGURE 2-2c: Subset Watson efficiencies $\text{eff}(\hat{\beta}_1)$ (left panel) and $\text{eff}(\hat{\beta}_2)$ (right panel) for the lathe data plotted vs. the serial correlation coefficient ρ .

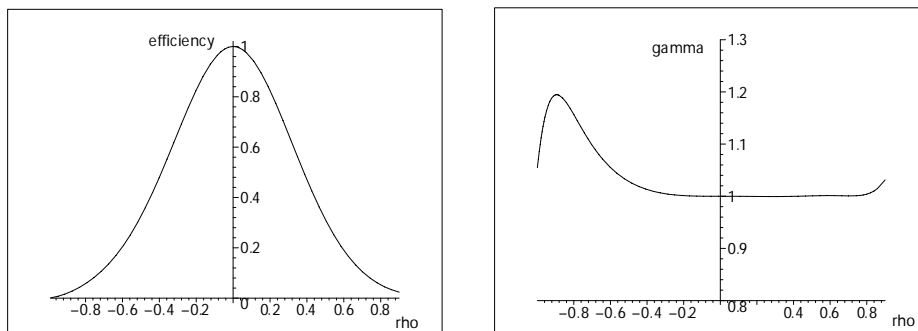


FIGURE 2-2d: Total Watson efficiencies for the lathe data: $\text{eff}(\hat{\beta}_{12})$ (left panel) and the efficiency factorization multiplier γ_{12} (right panel) plotted against the serial correlation coefficient ρ .

In Figure 2-2d, we plot (in the left panel) the total Watson efficiency $\text{eff}(\hat{\beta}_{12})$ and (in the right panel) the efficiency factorization multiplier γ_{12} both against the serial correlation coefficient ρ ; the efficiency is on the vertical axis and ρ on the horizontal axis. From the left panel of Figure 2-2d, we see that the total Watson efficiency follows a parabola with maximum efficiency equal to 1 at $\rho = 0$ and minimum efficiencies $\text{eff}(\hat{\beta}_{12}) \approx 0.053$ at $\rho = -0.8$ and $\text{eff}(\hat{\beta}_{12}) \approx 0.058$ at $\rho = +0.8$.

From the right panel of Figure 2-2d, we see that the efficiency factorization multiplier $\gamma_{12} \geq 1$. For positive ρ the efficiency factorization multiplier γ_{12} is very close to 1 and only equal to 1.004 at $\rho = +0.8$. For negative ρ the efficiency factorization multiplier γ_{12} behaves rather differently with a maximum of approximately 1.195 at $\rho = -0.893$. It is interesting to note, however, that $\gamma_{12} = 1$ (exactly) if and only if $\rho = 0$.

Example 2.3: Worsley’s fMRI data

For our third example we consider a set of fMRI (functional Magnetic Resonance Imaging) data kindly provided by Keith Worsley. As observed by Worsley *et al.* [62]:

The combined effect of k different stimuli types on data in scan i , denoted by k different responses x_{i1}, \dots, x_{ik} is often assumed to be additive but with different multiplicative coefficients β_1, \dots, β_k that vary from voxel to voxel. The combined fMRI response is modeled as the linear model [27] $x_{i1}\beta_1 + \dots + x_{ik}\beta_k$.

Some voxels in fMRI time series data show considerable drift over time. Drift can be either linear, or a more general slow variation. If drift is not removed, then it can either be confounded with the fMRI response, particularly if the stimuli vary slowly over time, or it can add to the estimate of the random noise. The first causes bias in estimates of the effect of the stimuli, the second causes bias in the estimate of the error of the estimated effect. Drift can be removed either by high-pass filtering or by introducing low frequency drift terms, such as cosines, polynomials, or splines, into the linear model ... we use a polynomial drift of order q . To do this we add extra “responses” w_{i1}, \dots, w_{im} at time i to the linear model. For example, a polynomial drift of order q can be removed by adding to the linear model $w_{ij} = t_i^{j-1}$; $j = 1, \dots, m = q + 1$. Finally a random error ε_i is added to obtain the observed fMRI data, Y_i , at time index i :

$$Y_i = \underbrace{w_{i1}\alpha_1 + \dots + w_{im}\alpha_m}_{\text{drift}} + \underbrace{x_{i1}\beta_1 + \dots + x_{ik}\beta_k}_{\text{fMRI response}} + \varepsilon_i \tag{2.25}$$

The simplest model of the temporal correlation structure is the first order autoregressive model [13], in which the scans are equally spaced in time and we suppose that the error from the previous scan is combined with fresh noise to produce the error for the

current scan: $\varepsilon_i = \rho\varepsilon_{i-1} + \xi_{i1}$, where $|\rho| < 1$ and ξ_{i1} is a sequence of independent and identically distributed normal random variables with mean 0 and standard deviation σ_1 , i.e., $\xi_{i1} \sim N(0, \sigma_1^2)$ (known as “white noise”). The resulting autocorrelation at lag h is $\text{corr}(\varepsilon_i, \varepsilon_{i-h}) = \rho^{|h|}$.

For our data set we have $k = 2$ stimuli in the fMRI response and we fit a cubic trend to the drift, and so $m = 4$ and $q = 3$. There are $n = 117$ observations in the data set of which the first 10 are given in Table 2-3a below; for all 117 observations, see Chu [16, pp. 14–16]. The first $m = 4$ columns in Tables 2-3a refer to the drift and the last $k = 2$ to the fMRI response. In our notation the model matrix \mathbf{X} is 117×6 , with the elements of the first column all equal to 1; the elements in column 2 represent a linear trend, the elements in column 3 a quadratic trend, and the elements in column 4 a cubic trend. Column 5 of \mathbf{X} represents a hot stimulus, whereas column 6 represents a warm stimulus. Columns 1–4 are considered to represent nuisance variables, the drift; columns 5 and 6 are of real interest, the fMRI response. We partition $\mathbf{X} = (\mathbf{X}_1 : \mathbf{X}_2)$, with $\mathbf{X}_1 : 117 \times 4$ and $\mathbf{X}_2 : 117 \times 2$. To orthogonally partition \mathbf{X} , we replace \mathbf{X}_2 with $\mathbf{M}_1\mathbf{X}_2$, where $\mathbf{M}_1 = \mathbf{I} - \mathbf{X}_1(\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{X}'_1$.

For the covariance matrix \mathbf{V} , we use the first-order autocorrelation structure $\{\rho^{|i-j|}\}$ as in (2.21) in Example 2-2 above, but with the covariance matrix now 117×117 . See also (3.12) below.

We tabulate the efficiency factorization multiplier γ vs. the serial correlation coefficient ρ in Table 2-3b below and note that $\gamma > 1$ for all $\rho < 0$. We find the behaviour of γ for $\rho > 0$ to be curious, see Figure 2-3 where we plot γ vs. ρ ; from Table 2-3b we see that $\gamma < 1$ for approximately $0.02 < \rho < 0.75$, with a minimum value of approximately $\gamma = 0.9989476$ at $\rho = 0.61$. We find also that $\gamma = 1$ when $\rho = 0$ and when $\rho \approx 0.751$.

TABLE 2-3a: First 10 observations of the fMRI data of Worsley *et al.* [62].

1	-1	1	-1	0.000590719	0
1	-0.982759	0.965815	-0.949163	0.348491	0
1	-0.965517	0.932224	-0.900078	1.26704	0
1	-0.948276	0.899227	-0.852715	1.50206	0
1	-0.931034	0.866825	-0.807044	0.926492	0
1	-0.913793	0.835018	-0.763034	-0.176833	0
1	-0.896552	0.803805	-0.720653	-0.481519	0.000590719
1	-0.87931	0.773187	-0.679871	-0.271114	0.348491
1	-0.862069	0.743163	-0.640658	-0.0896224	1.26704
1	-0.844828	0.713734	-0.602982	-0.0210606	1.50206

TABLE 2-3b: Serial correlation coefficient ρ (first column) and the efficiency factorization multiplier γ (second column) for the fMRI data of Worsley *et al.* [62].

-0.99	3.7036775	-0.49	1.0016779	0.01	1.0000000	0.51	0.9991626
-0.98	1.9251338	-0.48	1.0015487	0.02	0.9999999	0.52	0.9991308
-0.97	1.4808892	-0.47	1.0014285	0.03	0.9999998	0.53	0.9991002
-0.96	1.3041294	-0.46	1.0013166	0.04	0.9999994	0.54	0.9990712
-0.95	1.2146460	-0.45	1.0012123	0.05	0.9999989	0.55	0.9990442
-0.94	1.1619827	-0.44	1.0011152	0.06	0.9999981	0.56	0.9990194
-0.93	1.1277016	-0.43	1.0010249	0.07	0.9999970	0.57	0.9989974
-0.92	1.1037646	-0.42	1.0009407	0.08	0.9999955	0.58	0.9989787
-0.91	1.0861861	-0.41	1.0008624	0.09	0.9999937	0.59	0.9989637
-0.90	1.0727861	-0.40	1.0007896	0.10	0.9999913	0.60	0.9989532
-0.89	1.0622767	-0.39	1.0007219	0.11	0.9999885	0.61	0.9989476
-0.88	1.0538489	-0.38	1.0006590	0.12	0.9999851	0.62	0.9989478
-0.87	1.0469687	-0.37	1.0006005	0.13	0.9999812	0.63	0.9989545
-0.86	1.0412691	-0.36	1.0005463	0.14	0.9999766	0.64	0.9989686
-0.85	1.0364892	-0.35	1.0004960	0.15	0.9999713	0.65	0.9989911
-0.84	1.0324387	-0.34	1.0004494	0.16	0.9999653	0.66	0.9990230
-0.83	1.0289751	-0.33	1.0004062	0.17	0.9999585	0.67	0.9990655
-0.82	1.0259899	-0.32	1.0003664	0.18	0.9999510	0.68	0.9991200
-0.81	1.0233990	-0.31	1.0003295	0.19	0.9999425	0.69	0.9991878
-0.80	1.0211363	-0.30	1.0002956	0.20	0.9999332	0.70	0.9992707
-0.79	1.0191489	-0.29	1.0002643	0.21	0.9999230	0.71	0.9993704
-0.78	1.0173947	-0.28	1.0002356	0.22	0.9999119	0.72	0.9994891
-0.77	1.0158389	-0.27	1.0002093	0.23	0.9998997	0.73	0.9996290
-0.76	1.0144534	-0.26	1.0001852	0.24	0.9998866	0.74	0.9997929
-0.75	1.0132146	-0.25	1.0001632	0.25	0.9998724	0.75	0.9999837
-0.74	1.0121032	-0.24	1.0001432	0.26	0.9998571	0.76	1.0002049
-0.73	1.0111027	-0.23	1.0001250	0.27	0.9998408	0.77	1.0004603
-0.72	1.0101993	-0.22	1.0001085	0.28	0.9998233	0.78	1.0007545
-0.71	1.0093814	-0.21	1.0000936	0.29	0.9998047	0.79	1.0010926
-0.70	1.0086388	-0.20	1.0000803	0.30	0.9997850	0.80	1.0014805
-0.69	1.0079631	-0.19	1.0000683	0.31	0.9997642	0.81	1.0019252
-0.68	1.0073468	-0.18	1.0000577	0.32	0.9997422	0.82	1.0024345
-0.67	1.0067835	-0.17	1.0000483	0.33	0.9997191	0.83	1.0030176
-0.66	1.0062678	-0.16	1.0000400	0.34	0.9996949	0.84	1.0036851
-0.65	1.0057947	-0.15	1.0000327	0.35	0.9996696	0.85	1.0044490
-0.64	1.0053600	-0.14	1.0000264	0.36	0.9996432	0.86	1.0053230
-0.63	1.0049600	-0.13	1.0000210	0.37	0.9996157	0.87	1.0063221
-0.62	1.0045914	-0.12	1.0000165	0.38	0.9995872	0.88	1.0074627
-0.61	1.0042512	-0.11	1.0000126	0.39	0.9995578	0.89	1.0087616
-0.60	1.0039370	-0.10	1.0000094	0.40	0.9995274	0.90	1.0102345
-0.59	1.0036464	-0.09	1.0000068	0.41	0.9994962	0.91	1.0118931
-0.58	1.0033774	-0.08	1.0000048	0.42	0.9994643	0.92	1.0137407
-0.57	1.0031281	-0.07	1.0000032	0.43	0.9994316	0.93	1.0157642
-0.56	1.0028969	-0.06	1.0000020	0.44	0.9993984	0.94	1.0179239
-0.55	1.0026823	-0.05	1.0000011	0.45	0.9993648	0.95	1.0201378
-0.54	1.0024830	-0.04	1.0000006	0.46	0.9993308	0.96	1.0222654
-0.53	1.0022979	-0.03	1.0000002	0.47	0.9992967	0.97	1.0240946
-0.52	1.0021257	-0.02	1.0000001	0.48	0.9992627	0.98	1.0253455
-0.51	1.0019656	-0.01	1.0000000	0.49	0.9992288	0.99	1.0257114
-0.50	1.0018166	0.00	1.0000000	0.50	0.9991954		

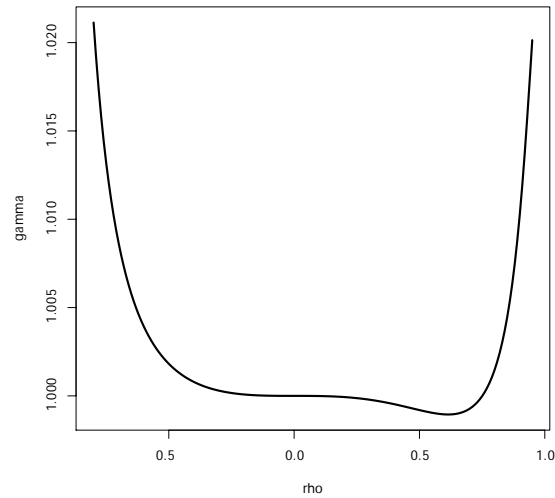


FIGURE 2-3a: A plot of the efficiency factorization multiplier γ (vertical axis) vs. the serial correlation coefficient ρ (horizontal axis).

3 Literature review

Consider the linear (or Gauß–Markov) model

$$E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}; \quad D(\mathbf{y}) = \sigma^2\mathbf{V}, \quad (3.1)$$

where \mathbf{X} is the $n \times p$ model (or design) matrix and \mathbf{V} is the $n \times n$ dispersion matrix. The $p \times 1$ vector $\boldsymbol{\beta}$ and the variance σ^2 are unknown and to be estimated based on a single realization of \mathbf{y} . The matrices \mathbf{X} and \mathbf{V} are assumed to be known.

When the dispersion matrix \mathbf{V} is positive definite, then the so-called “estimating equations”

$$\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\tilde{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \quad (3.2)$$

were introduced in 1935 by Aitken [1], and are known as the *Aitken equations*, see, e.g., Hinkelmann [31]. When \mathbf{X} has full column rank $p < n$, then $\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}$ is positive definite (invertible) and (3.2) has the unique solution

$$\tilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}, \quad (3.3)$$

say. The estimator $\tilde{\beta}$ is known as the generalized least squares (GLS) estimator of β , see, e.g., Seber & Lee [51, p. 67 (2003)], and the estimator

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{3.4}$$

is the well-known ordinary least squares (OLS) estimator of β .

Anderson [7, (1948)] seems to have been first to consider conditions for the equality of the ordinary least squares estimator and the generalized least squares estimator, and his contributions are discussed in some detail by Baksalary, Puntanen & Styan [10, (1990)]; see also Rao [50, (1967)], Zyskind [64, (1967)], and the survey by Puntanen & Styan [47, (1989)].

As observed in *The Oxford Dictionary of Statistical Terms* [21, p. 127 (2003)], see also David [20, (1995)],

The concept of efficiency in statistical estimation is due to Fisher [26, (1922)] and is an attempt to measure objectively the relative merits of several possible estimators. The criterion adopted by Fisher was that of variance, an estimator being regarded as more ‘efficient’ than another if it has smaller variance.

The results of Fisher [26] concerned the estimation of a scalar-valued parameter. Wilks [61, p. 476 (1932)] introduced the notion of generalized variance as the determinant of the dispersion matrix and was the first, we believe, to define [61, p. 478 (1932)] the ratio of generalized variances as a measure of efficiency in the estimation of a vector-valued parameter. It seems that Aitken [3, (1948)] was the first to consider the generalized variance of the ordinary least squares estimator, and that Watson, in his Ph.D. thesis [57, p. 66 (1951)], introduced the efficiency ϕ of the ordinary least squares estimator as the ratio of generalized variances and showed that

$$\phi = \frac{|D(\tilde{\beta})|}{|D(\hat{\beta})|} = \frac{|\mathbf{X}'\mathbf{X}|^2}{|\mathbf{X}'\mathbf{V}\mathbf{X}| \cdot |\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}|}, \tag{3.5}$$

where $|\cdot|$ denotes determinant. We call ϕ in (3.5) the *Watson efficiency*.

It is easy to show that the Watson efficiency

$$\phi \leq 1, \tag{3.6}$$

with equality if and only if $\hat{\beta} = \tilde{\beta}$ with probability 1, see also (1.17) above. A lower bound for the Watson efficiency (3.5) is provided by the Bloomfield–Watson–Knott inequality

$$\phi = \frac{|\mathbf{X}'\mathbf{X}|^2}{|\mathbf{X}'\mathbf{V}\mathbf{X}| \cdot |\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}|} \geq \prod_{i=1}^m \frac{4\lambda_i\lambda_{n-i+1}}{(\lambda_i + \lambda_{n-i+1})^2}, \tag{3.7}$$

where $m = \min(p, n - p)$ and $\lambda_1 \geq \dots \geq \lambda_n$ are the, necessarily positive, eigenvalues of \mathbf{V} .

The Bloomfield–Watson–Knott inequality (3.7) was originally conjectured by James Durbin, see Watson [57, p. 69 (1951)] and Watson [58, p. 331 (1955)], but was first established for $m = \min(p, n - p) \geq 2$ only much later by Bloomfield & Watson [12, (1975)] and Knott [38, (1975)]; see also Alpargu, Drury & Styan [4], Chu *et al.* [17, Lemma 4.1], Drury *et al.* [22], Khatri & Rao [36, 37], Yang [63],

When $p = 1$, the Bloomfield–Watson–Knott inequality (3.7) reduces, with the matrix \mathbf{X} replaced by the vector \mathbf{x} , to the well-known Kantorovich (or Frucht–Kantorovich) inequality:

$$\frac{(\mathbf{x}'\mathbf{x})^2}{\mathbf{x}'\mathbf{V}\mathbf{x} \cdot \mathbf{x}'\mathbf{V}^{-1}\mathbf{x}} \geq \frac{4\lambda_1\lambda_n}{(\lambda_1 + \lambda_n)^2}, \quad (3.8)$$

which was first established by Frucht [28, (1943)] and Kantorovich [35, (1948)]. Watson [57, Appendix 1, pp. 138–139 (1951)] and [58, Appendix, pp. 340–341 (1955)] used an inequality due to John William Scott Cassels to prove (3.8) and Watson, Alpargu & Styan [59] showed that the Cassels and Kantorovich inequalities were equivalent to each other and to four other named inequalities. See also Alpargu & Styan [5, 6], Drury *et al.* [22].

Hannan [30, p. 422 (1970)] considered the efficiency of the OLSE of a particular linear function $\mathbf{p}'\boldsymbol{\beta}$, where \mathbf{p} is a given $p \times 1$ column vector, and showed that

$$\frac{\text{var}(\mathbf{p}'\tilde{\boldsymbol{\beta}})}{\text{var}(\mathbf{p}'\hat{\boldsymbol{\beta}})} \geq \frac{4\lambda_1\lambda_n}{(\lambda_1 + \lambda_n)^2}, \quad (3.9)$$

see also Golub [29], Magness & McGuire [41], Sengupta & Jammalamadaka [52, §8.1.2, pp. 315–321], Wang & Chow [55, §5.6.2, pp. 211–215].

Bartmann & Bloomfield [11, (1981)] showed that the Watson efficiency

$$\begin{aligned} \phi &= \frac{|D(\hat{\boldsymbol{\beta}}_{\text{GLS}})|}{|D(\hat{\boldsymbol{\beta}}_{\text{OLS}})|} = \frac{|\mathbf{X}'\mathbf{X}|^2}{|\mathbf{X}'\mathbf{V}\mathbf{X}| \cdot |\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}|} \\ &= |\mathbf{I} - (\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\bar{\mathbf{X}}(\bar{\mathbf{X}}'\mathbf{V}\bar{\mathbf{X}})^{-1}\bar{\mathbf{X}}'\mathbf{V}\mathbf{X}| = \prod_{i=1}^m (1 - \rho_i^2), \end{aligned} \quad (3.10)$$

where $\bar{\mathbf{X}}$ is an $n \times (n - p)$ matrix such that $\mathbf{X}'\bar{\mathbf{X}} = \mathbf{0}$ and $\mathcal{C}(\bar{\mathbf{X}}) = \mathcal{N}(\mathbf{X})$, where $\mathcal{C}(\cdot)$ denotes column space (or range) and $\mathcal{N}(\cdot)$ denotes null (column) space.

In (3.10) the upper limit $m = \min(p, n - p)$ and the ρ_i^2 are the eigenvalues of

$$(\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\bar{\mathbf{X}}(\bar{\mathbf{X}}'\mathbf{V}\bar{\mathbf{X}})^{-1}\bar{\mathbf{X}}'\mathbf{V}\mathbf{X} \quad (3.11)$$

and so the ρ_i may be taken as the canonical correlations between the fitted values from OLS, i.e., the elements of the vector $\mathbf{X}\hat{\boldsymbol{\beta}}_{\text{OLS}}$, and the residuals from OLS, i.e., the elements of the vector $\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\text{OLS}}$; see also Puntanen [45]. It follows at once from (3.10) that the Watson efficiency $\phi = 1$ if and only if $\mathbf{X}'\mathbf{V}\bar{\mathbf{X}} = \mathbf{0}$, as shown by Rao [50] and Zyskind [64].

As observed by Durbin & Watson [23, (1950)],

A great deal of use has undoubtedly been made of least squares regression methods in circumstances in which they are known to be inapplicable. In particular, they have often been employed for the analysis of time series and similar data in which successive observations are serially correlated.

As noted in our examples in Section 2 above, the Watson efficiency ϕ may be used to measure the efficiency of ordinary least squares in such situations with the dispersion matrix

$$\mathbf{V} = \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-2} & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-3} & \rho^{n-2} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & \rho & 1 \end{pmatrix}, \quad (3.12)$$

where ρ is the serial correlation coefficient. Let $\mathbf{r} = \{r_i\}$ denote the vector of residuals from ordinary least squares. Then the well-known Durbin–Watson test is based on the statistic

$$D = \frac{\sum_{i=2}^n (r_i - r_{i-1})^2}{\sum_{i=1}^n r_i^2}. \quad (3.13)$$

As noted by Seber & Lee [51, p. 293], the Durbin–Watson test is “perhaps the most popular test for serial correlation” and “Durbin & Watson [25, (1971)] showed that the critical region $D < d_\varepsilon$ for testing the null hypothesis $\mathcal{H}_0 : \rho = 0$ against the one-sided alternative $\mathcal{H}_1 : \rho > 0$ has certain optimal properties; for example it is the locally most powerful invariant critical region.”

When \mathcal{H}_0 is true, the null distribution of D depends on the model matrix \mathbf{X} , so that d_ε has to be specially computed for each \mathbf{X} . Durbin & Watson [23, (1950)], however, obtained bounds D_L and D_U , say, so that $D_L \leq D \leq D_U$, and where the null distributions of D_L and D_U do not depend on \mathbf{X} . Significance points for this “Durbin–Watson bounds test” are given by Durbin & Watson [24, (1951)] and by Koerts & Abrahamse [39, pp. 176–178 (1969)].

The Watson efficiency is not the only measure of efficiency that has been proposed. Bloomfield & Watson [12, (1975)], see also Bartmann & Bloomfield [11, (1981)], introduced the measure of efficiency

$$\begin{aligned} \psi &= \frac{1}{2} \text{tr}(\mathbf{H}\mathbf{V} - \mathbf{V}\mathbf{H})'(\mathbf{H}\mathbf{V} - \mathbf{V}\mathbf{H}) \\ &= \text{tr} \mathbf{H}\mathbf{V}^2 - \text{tr}(\mathbf{H}\mathbf{V})^2 = \text{tr} \mathbf{H}\mathbf{V}\mathbf{M}\mathbf{V} = \text{tr} \mathbf{H}\mathbf{V}\mathbf{M}\mathbf{V}\mathbf{H}, \end{aligned} \quad (3.14)$$

where $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}$ and tr denotes trace. We will call ψ the *Bloomfield–Watson efficiency*. The matrix \mathbf{H} is often called the *hat matrix*, see, e.g., [40].

Bloomfield & Watson [12, (1975)] showed that

$$\psi = \text{tr} \mathbf{H}\mathbf{V}\mathbf{M}\mathbf{V}\mathbf{H} \leq \frac{1}{4} \sum_{i=1}^m (\lambda_i - \lambda_{n-i+1})^2, \quad (3.15)$$

where $m = \min(p, n - p)$ and $\lambda_1 \geq \dots \geq \lambda_n$ are the, necessarily positive, eigenvalues of the covariance matrix \mathbf{V} . The inequality (3.15) with $p = 1$ was established by Styan [53, (1983)] and was called *Styan's Inequality* by Jia [33, (1996)] and Jiang [34, (1998)]. We will call (3.15) the *Bloomfield–Watson trace inequality*. We note that $\psi = 0$ if and only if $\text{HVM} = \mathbf{0}$ if and only if the Watson efficiency $\phi = 1$.

Our original motivation in this research was the paper [14, (1969)] by Canner, who studied the special case of simple linear regression with model matrix

$$\mathbf{X} = \begin{pmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{pmatrix} \quad (3.16)$$

and covariance matrix

$$\mathbf{V} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & r & r \\ r & 1 & r \\ r & r & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & c & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & cr & r \\ cr & c^2 & cr \\ r & cr & 1 \end{pmatrix}. \quad (3.17)$$

Canner [14] observed that with the model matrix \mathbf{X} and covariance matrix \mathbf{V} as given by (3.16) and (3.17), respectively, the generalized least squares (GLS) regression line lies completely outside the data set with probability 1 provided c and r fall into a certain region, see Chu [16, ch. I, §1.4, Fig. 1.2]. We observe that with the model matrix \mathbf{X} and covariance matrix \mathbf{V} as given by (3.16) and (3.17), respectively, the generalized least squares (GLS) and ordinary least squares (OLS) regression lines are always parallel with probability 1 for all c, r such that the covariance matrix \mathbf{V} is positive definite, and so then there is a reduction of the Watson efficiency of type 2, see (1.20) in our §1 above. This result is extended in Chu [16, ch. I, Th. 1] following results of Puntanen [46, (1996)].

Acknowledgements

We are most grateful to Keith J. Worsley for providing us with the set of data analyzed in Example 2-3. Thanks go also to Evelyn M. Styan for her help. This research was supported in part by the Natural Sciences and Engineering Research Council of Canada.

References

- [1] Aitken, A. C. On least squares and linear combination of observations. *Proceedings of the Royal Society of Edinburgh*, 55 (1935), 42–48.
- [2] Aitken, A. C. Studies in practical mathematics, I: The evaluation, with applications, of a certain triple product matrix. *Proceedings of the Royal Society of Edinburgh*, 57 (1937), 269–304.
- [3] Aitken, A. C. On a problem in correlated errors. *Proceedings of the Royal Society of Edinburgh, Section A*, 62 (1948), 369–377.

- [4] Alpargu, Gülhan; Drury, S. W.; Styan, George P. H. Some remarks on the Bloomfield–Watson–Knott inequality and on some other inequalities related to the Kantorovich inequality. In *Proceedings of the Conference in Honor of Shayle R. Searle, August 9–10, 1996*, Biometrics Unit, Cornell University, Ithaca, New York, pp. 125–148, 1998.
- [5] Alpargu, Gülhan; Styan, George P. H. Some remarks and a bibliography on the Kantorovich inequality. In *Multidimensional Statistical Analysis and Theory of Random Matrices: Proceedings of the Sixth Eugene Lukacs Symposium, Bowling Green, OH, USA, 29–30 March 1996* (Gupta, Arjun K.; Girko, Vyacheslav L., eds.), VSP International Science Publishers, Zeist (Utrecht), pp. 1–13, 1996.
- [6] Alpargu, Gülhan; Styan, George P. H. Some comments and a bibliography on the Frucht–Kantorovich and Wielandt inequalities. In *Innovations in Multivariate Statistical Analysis: A Festschrift for Heinz Neudecker* Heijmans, R. D. H.; Pollock, D. S. G.; Satorra, A., eds., Kluwer Academic Publishers, Dordrecht, pp. 1–38, 2000.
- [7] Anderson, T. W. On the theory of testing serial correlation. *Skandinavisk Aktuarietidskrift*, 31 (1948), 88–116; reprinted with corrections and commentary in *The Collected Papers of T. W. Anderson: Volume 1* (George P. H. Styan, ed.), pp. 61–89 (1990).
- [8] Anderson, T. W. *The Statistical Analysis of Time Series*, Wiley Classics Library edition. Wiley, New York, 1994. (Original version: 1971.)
- [9] Baksalary, Jerzy K.; Puntanen, Simo. Characterizations of the best linear unbiased estimator in the general Gauss-Markov model with the use of matrix partial orderings. *Linear Algebra and its Applications*, 127 (1990), 363–370.
- [10] Baksalary, Jerzy K.; Puntanen, Simo; Styan, George P. H. 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. In *The Collected Papers of T. W. Anderson: Volume 2* (George P. H. Styan, ed.), 1579–1591 (1990).
- [11] Bartmann, Flavio C.; Bloomfield, Peter. Inefficiency and correlation. *Biometrika*, 68 (1981), 67–71.
- [12] Bloomfield, Peter; Watson, Geoffrey S. The inefficiency of least squares. *Biometrika*, 62 (1975), 121–128.
- [13] Bullmore, E. T.; Brammer, M. J.; Williams, S. C. R.; Rabe-Hesketh, S.; Janot, N., David, A. S.; Mellers, J. D. C.; Howard, R.; Sham, P. Statistical methods of estimation and inference for functional MR image analysis. *Magnetic Resonance in Medicine*, 35 (1996), 261–277.
- [14] Canner, Paul L. Some curious results using minimum variance linear unbiased estimators. *The American Statistician*, 23, no. 5 (1969), pp. 39–40.
- [15] Chipman, John S. The problem of testing for serial correlation in regression analysis: the story of a dilemma. Technical Report no. 4, Dept. of Economics, University of Minnesota, Minneapolis, 35 pp., revised June 21, 1965.
- [16] Chu, Ka Lok. *Inequalities and Equalities Associated with Ordinary Least Squares and Generalized Least Squares in Partitioned Linear Models*. Ph. D. dissertation, Dept. of Mathematics and Statistics, McGill University, Montréal (Québec), Canada, December 2004.
- [17] Chu, K. L.; Isotalo, J.; Puntanen, S.; Styan, G. P. H. On decomposing the Watson efficiency of ordinary least squares in a partitioned weakly singular linear model. *Sankhyā: The Indian Journal of Statistics*, 66 (2004), 634–651.

- [18] Chu, Ka Lok; Isotalo, Jarkko; Puntanen, Simo; Styan, George P. H. Inequalities and equalities for the generalized efficiency function in orthogonally partitioned linear models. Invited paper for *Inequalities and Applications* (Themistocles M. Rassias, ed.), Springer, in press, 2005. [Expanded version of Report A 352, Dept. of Mathematics, Statistics & Philosophy, University of Tampere, Tampere, Finland, July 2004.]
- [19] Chu, K. L.; Isotalo, J.; Puntanen, S.; Styan, G. P. H. Some further results concerning the decomposition of the Watson efficiency in partitioned linear models. *Sankhyā: The Indian Journal of Statistics*, 67 (2005), 74–89.
- [20] David, H. A. First (?) occurrence of common terms in mathematical statistics. *The American Statistician*, 49 (1995), 121–133.
- [21] Dodge, Yadolah, ed. *The Oxford Dictionary of Statistical Terms*. Oxford University Press, 2003.
- [22] Drury, S. W.; Liu, Shuangzhe; Lu, Chang-Yu; Puntanen, Simo; Styan, George P. H. Some comments on several matrix inequalities with applications to canonical correlations: historical background and recent developments. *Sankhyā: The Indian Journal of Statistics, Series A*, 64 (2002), 453–507.
- [23] Durbin, J.; Watson, G. S. Testing for serial correlation in least squares regression, I. *Biometrika*, 37 (1950), 409–428. (Corrections in [24, pp. 177–178 (1951)].)
- [24] Durbin, J.; Watson, G. S. Testing for serial correlation in least squares regression II. *Biometrika*, 38 (1951), 159–178.
- [25] Durbin, J.; Watson, G. S. Testing for serial correlation in least squares regression III. *Biometrika*, 58 (1971), 1–42.
- [26] Fisher, R. A. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London, Series A: Mathematical and Physical Sciences*, 222 (1922), 309–368. (Reprinted in *Collected Papers of R. A. Fisher, Volume I: 1912–24* (Bennett, J. H., ed.), pp. 275–335 (1971).
- [27] Friston, K. J.; Holmes, A. P.; Worsley, K. J.; Poline, J.-B.; Frith, C. D.; Frackowiak, R. S. J. Statistical parametric maps in functional imaging: a general linear approach. *Human Brain Mapping*, 2 (1995), 189–210.
- [28] Frucht, Roberto. Sobre algunas desigualdades: observación relativa a la solución del Problema N° 21, indicada por el Ing. Ernesto M. Saleme [with an untitled appendix by Beppo Levi, all in Spanish]. *Mathematicæ Notæ–Boletín del Instituto de Matemática “Beppo Levi” (Rosario)*, 3 (1943), 41–46. (Translated into English as Appendix A in Watson, Alpargu & Styan [59, (1997)].)
- [29] Golub, Gene H. Comparison of the variance of minimum variance and weighted least squares regression coefficients. *The Annals of Mathematical Statistics*, 37 (1963), 984–991.
- [30] Hannan, E. *Multiple Time Series*. Wiley, New York, 1970.
- [31] Hinkelmann, Klaus. Aitken equations. In *Encyclopedia of Statistical Sciences: Update Volume 1* (Kotz, Samuel; Read, Campbell B.; Banks, David L., eds.), Wiley, New York, pp. 22–23, 1997.
- [32] Huuhtanen, Pentti; Liski, Erkki; Pukkila, Tarmo; Puntanen, Simo, eds. *Festschrift for Eino Haikala on his Seventieth Birthday*. Acta Universitatis Tamperensis Series 153, University of Tampere, Tampere, Finland.

- [33] Jia, Zhongzhen. An extension of Styan's inequality (in Chinese, with abstract in English). *Gongcheng Shuxue Xuebao/Journal of Engineering Mathematics (Xi'an)*, 13, no. 1 (1996), 122–126.
- [34] Jiang, Yongquan. Comment: An extension of Styan's inequality (in Chinese, with abstract in English). *Xuzhou Shifan Daxue Xuebao: Ziran Kexue Ban (Journal of Xuzhou Normal University, Natural Science Edition)*, 16, no. 1 (1998), pp. 4–5.
- [35] Kantorovich, L. V. Funkcional'nyi analiz i prikladnaya matematika (in Russian). *Uspekhi Matematičeskii Nauk, Novaya Seriya*, 3, no. 6/28 (1948), pp. 89–185; see pp. 142–144. (Translated by C. D. Benster into English as *Functional Analysis and Applied Mathematics* and edited by G. E. Forsythe, NBS Report no. 1509 (1101-10-5100), U.S. Dept. of Commerce, National Bureau of Standards, Los Angeles, see pp. 106–109.)
- [36] Khatri, C. G.; Rao, C. Radhakrishna. Some extensions of the Kantorovich inequality and statistical applications. *Journal of Multivariate Analysis*, 11 (1981), 498–505.
- [37] Khatri, C. G.; Rao, C. Radhakrishna. Some generalizations of Kantorovich inequality. *Sankhyā: Sankhyā: The Indian Journal of Statistics, Series A* 44 (1982), 91–102; reprinted in *Selected Papers of C. R. Rao: Volume Four* (Das Gupta, S.; Ghosh, J. K.; Mitra, S. K.; Mukhopadhyay, A. C.; Prakasa Rao, B. L. S.; Rao, P. S. S. N. V. P.; Rao, S. B.; Sarma, Y. R., eds.), Indian Statistical Institute, Kolkata & New Age International (P) Limited, New Delhi, pp. 240–249, 1999.
- [38] Knott, M. On the minimum efficiency of least squares. *Biometrika*, 62 (1975), 129–132.
- [39] Koerts, J.; Abrahamse, A. P. J. *On the Theory and Application of the General Linear Model*. Rotterdam University Press, 1969.
- [40] Kotz, Samuel; Johnson, Norman L.; Read, Campbell B., eds. Hat matrix. In *Encyclopedia of Statistical Sciences, Volume 3: Faà di Bruno's Formula–Hypothesis Testing* (Kotz, Samuel; Johnson, Norman L.; Read, Campbell B., eds.), Wiley, New York, p. 580, 1983.
- [41] Magness, T. A.; McGuire, J. B. Comparison of least squares and minimum variance estimates of regression parameters. *The Annals of Mathematical Statistics*, 33 (1962), 462–470.
- [42] Marshall, Albert W.; Olkin, Ingram. *Inequalities: Theory of Majorization and its Applications* Academic Press, New York, 1979.
- [43] Odell, Patrick L. Gauss–Markov theorem. In *Encyclopedia of Statistical Sciences, Volume 3: Faà di Bruno's Formula–Hypothesis Testing* (Kotz, Samuel; Johnson, Norman L.; Read, Campbell B., eds.), Wiley, New York, pp. 314–316, 1983.
- [44] Pukkila, Tarmo; Puntanen, Simo, eds. *Proceedings of the First International Tampere Seminar on Linear Statistical Models and their Applications: University of Tampere, Tampere, Finland, August 30th to September 2nd, 1983*. Dept. of Mathematical Sciences/Statistics, University of Tampere, Tampere, Finland, 1985.
- [45] Puntanen, Simo. On the relative goodness of ordinary least squares and on some associated canonical correlations. Paper 1 in *On the Relative Goodness of Ordinary Least Squares Estimation in the General Linear Model* by Simo Puntanen, Ph.D. dissertation, Dept. of Mathematical Sciences, University of Tampere, Acta Universitatis Tamperensis Series A 216, University of Tampere, Tampere, Finland, 1987.
- [46] Puntanen, Simo. Some matrix results related to a partitioned singular linear model. *Communications in Statistics: Theory and Methods*, 25 (1996), 269–279.

- [47] Puntanen, Simo; Styan, George P. H. 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 (1989), 153–164.
- [48] Puntanen, Simo; Styan, George P. H. Historical introduction: Issai Schur and the early development of the Schur complement. In *The Schur Complement and Its Applications* (Fuzhen Zhang, ed.), Springer, New York, pp. 1–16, 259–288 (2005).
- [49] Puntanen, Simo; Styan, George P. H. Schur complements in statistics and probability. In *The Schur Complement and Its Applications* (Fuzhen Zhang, ed.), Springer, New York, pp. 163–226, 259–288 (2005).
- [50] Rao, C. Radhakrishna. Least squares theory using an estimated covariance matrix and its application to measurement of signals. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability held at the Statistical Laboratory, University of California, June 21–July 18, 1965 and December 27, 1965–January 7, 1966, Volume I: Statistics* (Le Cam, Lucien M.; Neyman, Jerzy, eds.), University of California Press, Berkeley, pp. 355–372, 1967; reprinted in *Selected Papers of C. R. Rao: Volume Three* (Das Gupta, S.; Ghosh, J. K.; Mitra, S. K.; Mukhopadhyay, A. C.; Prakasa Rao, B. L. S.; Rao, P. S. S. N. V. P.; Rao, S. B.; Sarma, Y. R., eds.), Indian Statistical Institute, Kolkata & New Age International (P) Limited, New Delhi, pp. 92–108, 1995.
- [51] Seber, George A. F.; Lee, Alan J. *Linear Regression Analysis*, 2nd edition. Wiley-Interscience, Hoboken, New Jersey, 2003. (Original version by George A. F. Seber, Wiley, New York, 1977.)
- [52] Sengupta, Debasis; Jammalamadaka, Sreenivasa Rao. *Linear Models: An Integrated Approach*. World Scientific, Singapore, 2003.
- [53] Styan, George P. H. On some inequalities associated with ordinary least squares and the Kantorovich inequality. In Huuhtanen *et al.* [32, pp. 158–166 (1983)].
- [54] Styan, George P. H. Schur complements and linear statistical models. In Pukkila & Puntanen [44, pp. 37–75 (1985)].
- [55] Wang, Song-Gui; Chow, Shein-Chung. *Advanced Linear Models: Theory and Applications*. Statistics: Textbooks and Monographs 141, Marcel Dekker, New York, 1994.
- [56] Ward, Laurel Lorraine. *Is Uncorrelating the Residuals Worth It?* M.Sc. thesis, Dept. of Mathematics, McGill University, March 1973.
- [57] Watson, Geoffrey S. *Serial Correlation in Regression Analysis*. Ph.D. thesis, Dept. of Experimental Statistics, North Carolina State College, Raleigh, 1951.
- [58] Watson, Geoffrey S. Serial correlation in regression analysis, I. *Biometrika*, 42 (1955), 327–341.
- [59] Watson, Geoffrey S.; Alpargu, Gülhan; Styan, George P. H. Some comments on six inequalities associated with the inefficiency of ordinary least squares with one regressor. *Linear Algebra and its Applications*, 264 (1997), 13–53.
- [60] Weisberg, Sanford. *Applied Linear Regression*, 2nd edition. Wiley, New York, 1985. (Original version: 1980; 3rd edition, expected February 2005.)
- [61] Wilks, S. S. Certain generalizations in the analysis of variance. *Biometrika*, 24 (1932), 471–494. (Reprinted in *S. S. Wilks: Collected Papers, Contributions to Mathematical Statistics* (Anderson, T. W., ed.), Wiley, New York, pp. 80–103, 1967.)

- [62] Worsley, K. J.; Liao, C. H.; Aston, J.; Petre, V.; Duncan, G. H.; Morales, F.; Evans, A. C. A general statistical analysis for fMRI data. *NeuroImage*, 15 (2002), 1–15.
- [63] Yang, Hu. A brief proof on the generalized variance bound of the relative efficiency in statistics. *Communications in Statistics–Theory and Methods*, 19 (1990), 4587–4590.
- [64] Zyskind, George. On canonical forms, non-negative covariance matrices and best and simple least squares linear estimators in linear models. *The Annals of Mathematical Statistics*, 38 (1967), 1092–1109.
- [65] Zyskind, George; Martin, Frank B. On best linear estimation and a general Gauss–Markov theorem in linear models with arbitrary nonnegative covariance structure. *SIAM Journal on Applied Mathematics*, 17 (1969), 1190–1202.

