JARKKO ISOTALO

Linear estimation and prediction in the general Gauss–Markov model

ACADEMIC DISSERTATION
To be presented, with the permission of
the Faculty of Information Sciences of the University of Tampere,
for public discussion in the Paavo Koli Auditorium, Kanslerinrinne 1,
Tampere, on September 7th, 2007, at 12 o'clock.

UNIVERSITY OF TAMPERE

ACADEMIC DISSERTATION
University of Tampere
Department of Mathematics, Statistics and Philosophy
Finland

Distribution
Bookshop TAJU
P.O. Box 617
33014 University of Tampere
Finland

Tel. +358 3 3551 6055
Fax +358 3 3551 7685
taju@uta.fi
www.uta.fi/taju
http://granum.uta.fi

Cover design by
Juha Siro

Acta Universitatis Tamperensis 1242
ISBN 978-951-44-7017-2 (print)
ISSN 1455-1616

Acta Electronica Universitatis Tamperensis 636
ISBN 978-951-44-7018-9 (pdf)
ISSN 1456-954X
http://acta.uta.fi

Tampereen Yliopistopaino Oy – Juvenes Print
Tampere 2007

Abstract

In this doctoral thesis we consider topics related to linear estimation and prediction in the general Gauss–Markov model. The thesis consists of eleven articles and an introduction to concepts considered in the articles. The main contributions of the thesis concern the concepts of the best linear unbiased estimator, BLUE, the best linear unbiased predictor, BLUP, linear sufficiency, linear prediction sufficiency, the ordinary least squares estimator, OLSE, and the Watson efficiency.

In this thesis we consider linear sufficiency and linear completeness in the context of estimating the given estimable parametric function. Some new characterizations for linear sufficiency and linear completeness in a case of estimation of the parametric function are given, and also a predictive approach for obtaining the BLUE of the estimable parametric function is considered.

In the context of predicting the value of new observation under the general Gauss–Markov model, a new concept—linear prediction sufficiency—is introduced, and some basic properties of linear prediction sufficiency are given.

Furthermore, in this thesis the equality of the OLSE and BLUE of the given estimable parametric function is considered, and properties of the Watson efficiency are investigated particularly under the partitioned linear model.

This thesis contains also an article concerning the best linear unbiased estimation under the linear mixed effects model, and an article considering a particular matrix decomposition useful in the theory of linear models.

Key Words: best linear unbiased estimation, best linear unbiased prediction, linear model, linear sufficiency, linear completeness, linear prediction sufficiency, ordinary least squares estimation, Watson efficiency.

Acknowledgements

I want to express my deepest gratitude to my supervisor Dr. Simo Puntanen for his extensive guidance and support. It has been fantastic five years working together with you, Simo, with a lot of funny memories to remember and quite a few research papers to show. All the time you have been a very cooperative research partner, an enthusiastic teacher that I have admired, an adventurous travel mate during our joint journeys, and in general, a good friend to me. You definitely have set the bar high for how to be an excellent supervisor. It has been a great privilege and pleasure to be your student.

I warmly thank my "grandsupervisor" Professor George P. H. Styan for his efforts in our joint research papers, and more generally, for sharing his wide expertise on linear statistical models with me. I am most indebted also to my other co-authors, Dr. Ka Lok Chu and Dr. Märt Möls, for their contributions to our joint research papers. My sincere thanks also to both Professor Styan and Dr. Chu of the warm hospitality I received from them during my visits to Montréal.

I would like to thank Professor Erkki P. Liski for his encouragement and support, and as the head of my home department, for taking care of all practical administrative tasks during my postgraduate studies. I thank all the personnel of my home department for creating friendly working environment. Specially, huge thanks to Dr. Tapio Nummi for hiring me as a researcher to his research project in 2003; that was a crucial step for this thesis to ever materialize. Also very warm thanks to Mrs. Raija Leppälä for letting me act as an assistant teacher in her courses during my studies.

I wish to thank the referees of the thesis, Professor Thomas Mathew and Professor Debasis Sengupta, for their constructive remarks on the thesis. Also warmest thanks to Professor Götz Trenkler for kindly agreeing to act as the opponent at the public defence of the thesis, and for providing completions to the articles (Section 6).

For financial support I am grateful to the Tampere Graduate School in Information Science and Engineering (TISE), the University of Tampere Foundation, the Academy of Finland, the Alfred Kordelin Foundation, and the Scientific Foundation of the City of Tampere.

Finally, I want to thank my friends for bringing many happy times into my life outside of work. I am also grateful to my near relatives for their support and encouragement over the years. And lastly, I want to thank my mother, father, and brother of the support and love they have given and shown to me during my whole life.

Tampere, August 2007 *Jarkko Isotalo*

Contents

Notations

$\mathbb{R}^{n\times m}$	set of $n \times m$ real matrices $\mathbf{A}$
$\mathbb{R}^{n}$	set of $n \times 1$ real column vectors $\mathbf{a}$
$\mathbf{A}'$	transpose of matrix $\mathbf{A}$
$\mathbf{A}^{-1}$	inverse of matrix $\mathbf{A}$
$\mathbf{A}^{-}$	generalized inverse of matrix $\mathbf{A}$
$\mathbf{A}^{+}$	Moore–Penrose inverse of matrix $\mathbf{A}$
$\lvert\mathbf{A}\rvert$	determinant of matrix $\mathbf{A}$, also denoted $\det(\mathbf{A})$
$\mathrm{r}(\mathbf{A})$	rank of matrix $\mathbf{A}$
$\mathrm{tr}(\mathbf{A})$	trace of matrix $\mathbf{A}$
$\mathbf{A} \geq \mathbf{0}$	$\mathbf{A}$ is nonnegative definite
$\mathbf{A} > \mathbf{0}$	$\mathbf{A}$ is positive definite
$\mathscr{C}(\mathbf{A})$	column space of matrix $\mathbf{A}$
$\mathscr{N}(\mathbf{A})$	null space of matrix $\mathbf{A}$
$\mathscr{C}(\mathbf{A})^{\perp}$	orthogonal complement of $\mathscr{C}(\mathbf{A})$
$\dim \mathscr{C}(\mathbf{A})$	dimension of $\mathscr{C}(\mathbf{A})$
$\mathbf{P}_{\mathbf{A}}$	orthogonal projector onto $\mathscr{C}(\mathbf{A})$ w.r.t. standard inner product
$\mathbf{A}^{\perp}$	matrix whose column space is $\mathscr{C}(\mathbf{A})^{\perp}$
$(\mathbf{A} : \mathbf{B})$	columnwise partitioned matrix with $\mathbf{A} \in \mathbb{R}^{n\times m}$ and $\mathbf{B} \in \mathbb{R}^{n\times k}$
$\mathbf{A} \geq \mathbf{B}$	$\mathbf{A} - \mathbf{B}$ is nonnegative definite (Löwner partial ordering)
$\mathbf{A} > \mathbf{B}$	$\mathbf{A} - \mathbf{B}$ is positive definite
$\mathbf{a}$	column vector $\mathbf{a} \in \mathbb{R}^{n}$
$\mathbf{I}$	identity matrix
$\mathbf{0}$	matrix of zeroes
$\mathbf{1}$	column vector of ones
$\mathrm{E}(\mathbf{x})$	expectation of a random vector $\mathbf{x}$
$\mathrm{cov}(\mathbf{x})$	covariance matrix of a random vector $\mathbf{x}$
$\mathrm{cov}(\mathbf{x}, \mathbf{y})$	(cross-)covariance matrix between random vectors $\mathbf{x}$ and $\mathbf{y}$

List of original articles

[1] Isotalo, J. & Puntanen, S. (2006). Linear sufficiency and completeness in the partitioned linear model. *Acta et Commentationes Universitatis Tartuensis de Mathematica*, 10, 53–67.

[2] Isotalo, J. & Puntanen, S. (2007). Linear sufficiency and completeness in the context of estimating the parametric function in the general Gauss–Markov model. *Journal of Statistical Planning and Inference*, 19 pp., submitted for publication.

[3] Isotalo, J. & Puntanen, S. (2006). Linear prediction sufficiency for new observations in the general Gauss–Markov model. *Communications in Statistics – Theory and Methods*, 35, 1011–1023.

[4] 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: The Indian Journal of Statistics*, 66, 634–651.

[5] 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: The Indian Journal of Statistics*, 67, 74–89.

[6] Chu, K. L., Isotalo, J., Puntanen, S. & Styan, G. P. H. (2007). The efficiency factorization multiplier for the Watson efficiency in partitioned linear models: Some examples and a literature review. *Journal of Statistical Planning and Inference*, 137, 3336–3351.

[7] Isotalo, J., Puntanen, S. & Styan, G. P. H. (2007). Effect of adding regressors on the equality of the OLSE and BLUE. *International Journal of Statistical Sciences*, 9 pp., in press.

[8] Isotalo, J. & Puntanen, S. (2007). A note on the equality of the OLSE and the BLUE of the parametric function in the general Gauss–Markov model. *Statistical Papers*, 9 pp., in press, doi: 10.1007/s00362-007-0055-6.

[9] Isotalo, J., Puntanen, S. & Styan, G. P. H. (2006). Some comments on the Watson efficiency of the ordinary least squares estimator under the Gauss–Markov model. *Calcutta Statistical Association Bulletin*, 14 pp., submitted for publication.

[10] Isotalo, J., Möls, M. & Puntanen, S. (2006). Invariance of the BLUE under the linear fixed and mixed effects models. *Acta et Commentationes Universitatis Tartuensis de Mathematica*, 10, 69–76.

[11] Isotalo, J., Puntanen, S. & Styan, G. P. H. (2007). A useful matrix decomposition and its statistical applications in linear regression. *Communications in Statistics – Theory and Methods*, 22 pp., conditionally accepted.

1 Introduction

1.1 The general Gauss–Markov model

In this thesis we consider the general Gauss–Markov model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{1}$$

where $\mathbf{y}$ is an $n \times 1$ observable random vector, $\mathbf{X}$ is a known $n \times p$ model matrix, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters, and $\boldsymbol{\varepsilon}$ is an $n \times 1$ random error vector. The expectation and the covariance matrix of random vector $\mathbf{y}$ are $\mathrm{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ and $\mathrm{cov}(\mathbf{y}) = \sigma^2\mathbf{V}$, respectively, where $\sigma^2 > 0$ is an unknown scalar and $\mathbf{V}$ is a known nonnegative definite matrix. In short, we use the notation

$$\mathscr{M} = \{\mathbf{y},\, \mathbf{X}\boldsymbol{\beta},\, \sigma^2\mathbf{V}\} \tag{2}$$

to describe the general Gauss–Markov model.

When the model matrix $\mathbf{X}$ is partitioned columnwise as $\mathbf{X} = (\mathbf{X}_1 : \mathbf{X}_2)$ with $\mathbf{X}_1$ $(n \times p_1)$ and $\mathbf{X}_2$ $(n \times p_2)$, and correspondingly $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$ with $\boldsymbol{\beta}_1$ $(p_1 \times 1)$ and $\boldsymbol{\beta}_2$ $(p_2 \times 1)$, then the Gauss–Markov model $\mathscr{M}$ is called the partitioned linear model

$$\mathscr{M}_{12} = \{\mathbf{y},\, \mathbf{X}\boldsymbol{\beta},\, \sigma^2\mathbf{V}\} = \{\mathbf{y},\, \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2,\, \sigma^2\mathbf{V}\}. \tag{3}$$

1.2 The best linear unbiased estimator

Let us now consider estimation of a linear parametric function $\mathbf{K}'\boldsymbol{\beta}$, $\mathbf{K} \in \mathbb{R}^{p\times k}$, under the model $\mathscr{M}$. A given linear parametric function $\mathbf{K}'\boldsymbol{\beta}$ is said to be estimable under the model $\mathscr{M}$ if there exists a linear unbiased estimator for $\mathbf{K}'\boldsymbol{\beta}$, i.e., there exists a linear statistic $\mathbf{G}\mathbf{y}$ such that

$$\mathrm{E}(\mathbf{G}\mathbf{y}) = \mathbf{G}\mathbf{X}\boldsymbol{\beta} = \mathbf{K}'\boldsymbol{\beta} \quad \text{for all } \boldsymbol{\beta} \in \mathbb{R}^p, \tag{4}$$

or equivalently, if there exists a matrix $\mathbf{G}$ such that $\mathbf{G}\mathbf{X} = \mathbf{K}'$; in other words, $\mathscr{C}(\mathbf{K}) \subset \mathscr{C}(\mathbf{X}')$. Note that in our definition of the estimability of $\mathbf{K}'\boldsymbol{\beta}$ we allow $\boldsymbol{\beta}$ to vary freely over the whole $\mathbb{R}^p$. Thus we have not considered so-called natural restrictions on the parameter vector $\boldsymbol{\beta}$ which may arise under the model $\mathscr{M}$. For more about estimation under the natural restrictions, see, Groß (2004, Section 1), Sengupta & Jammalamadaka (2003, Section 7.2), and Baksalary, Rao & Markiewicz (1992).

Based on (4), it is clear that the vector of expectation $\mathbf{X}\boldsymbol{\beta}$ is always estimable, and, on the other hand, the vector of unknown parameters $\boldsymbol{\beta}$ is itself estimable if and only if $\mathscr{C}(\mathbf{X}') = \mathbb{R}^p$, or equivalently $\mathrm{r}(\mathbf{X}) = p$. In this thesis we are only interested in linear estimation of estimable parametric functions, i.e., we are only interested in linear parametric functions that have a linear unbiased estimator.

Let $\mathbf{K}'\boldsymbol{\beta}$ be a given estimable parametric function under the model $\mathscr{M}$. Then a linear unbiased estimator $\mathbf{Gy}$ of $\mathbf{K}'\boldsymbol{\beta}$ is said to be the best linear unbiased estimator, BLUE, of $\mathbf{K}'\boldsymbol{\beta}$ under the model $\mathscr{M}$ if, for any other unbiased linear estimator $\mathbf{Fy}$, the difference $\operatorname{cov}(\mathbf{Fy}) - \operatorname{cov}(\mathbf{Gy})$ is a nonnegative definite matrix, i.e.,

$$\operatorname{cov}(\mathbf{Fy}) - \operatorname{cov}(\mathbf{Gy}) \geq \mathbf{0} \qquad \text{for all } \mathbf{Fy} \text{ such that } \operatorname{E}(\mathbf{Fy}) = \mathbf{K}'\boldsymbol{\beta}. \tag{5}$$

The so-called fundamental equation of the BLUE states that a linear statistic $\mathbf{Gy}$ is the BLUE of $\mathbf{K}'\boldsymbol{\beta}$ under the model $\mathscr{M}$ if and only if $\mathbf{G}$ satisfies equation

$$\mathbf{G}(\mathbf{X} : \mathbf{V}\mathbf{X}^{\perp}) = (\mathbf{K}' : \mathbf{0}). \tag{6}$$

A proof of (6) is given, e.g., by Drygas (1970, p. 50), or Rao (1973, p. 282), see also more recent proofs of Baksalary (2004) and Puntanen, Styan & Werner (2000).

Throughout this thesis we assume the model $\mathscr{M}$ being the correct one from the modelling point of view. That is, we assume that the model $\mathscr{M}$ represents the structure of the random vector $\mathbf{y}$ accurately, and thus it always holds that

$$\mathbf{y} \in \mathscr{C}(\mathbf{X} : \mathbf{V}) = \mathscr{C}(\mathbf{X} : \mathbf{V}\mathbf{X}^{\perp}) \tag{7}$$

almost surely. The property (7) now guarantees that the BLUE of $\mathbf{K}'\boldsymbol{\beta}$ is unique. That is, if $\mathbf{G}_*$ is any other matrix satisfying the equation (6), then $\operatorname{BLUE}(\mathbf{K}'\boldsymbol{\beta} \mid \mathscr{M}) = \mathbf{Gy} = \mathbf{G}_*\mathbf{y}$ almost surely [see Groß (2004, Corollary 3)].

We may obtain explicit representations for the BLUE of $\mathbf{K}'\boldsymbol{\beta}$ from the equation (6). One often used representation for the BLUE of $\mathbf{K}'\boldsymbol{\beta}$ is

$$\operatorname{BLUE}(\mathbf{K}'\boldsymbol{\beta} \mid \mathscr{M}) = \widetilde{\mathbf{K}'\boldsymbol{\beta}} = \mathbf{K}'\tilde{\boldsymbol{\beta}} = \mathbf{K}'(\mathbf{X}'\mathbf{W}^{-}\mathbf{X})^{-}\mathbf{X}'\mathbf{W}^{-}\mathbf{y}, \tag{8}$$

where $\mathbf{W} = \mathbf{V} + \mathbf{XUX}'$ with $\mathbf{U}$ being an arbitrary matrix such that $\mathscr{C}(\mathbf{W}) = \mathscr{C}(\mathbf{X} : \mathbf{V})$.

1.3 The ordinary least squares estimator

Another linear unbiased estimator for the given estimable parametric function $\mathbf{K}'\boldsymbol{\beta}$ is the ordinary least squares estimator, OLSE. The ordinary least squares estimator of the given estimable parametric function $\mathbf{K}'\boldsymbol{\beta}$ under the model $\mathscr{M}$ is defined as

$$\operatorname{OLSE}(\mathbf{K}'\boldsymbol{\beta} \mid \mathscr{M}) = \widehat{\mathbf{K}'\boldsymbol{\beta}} = \mathbf{K}'\hat{\boldsymbol{\beta}} = \mathbf{K}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y}, \tag{9}$$

where $\hat{\boldsymbol{\beta}}$ is any vector satisfying the normal equation $\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$. Since $\mathbf{K}'\boldsymbol{\beta}$ is estimable and thereby $\mathbf{K}'\hat{\boldsymbol{\beta}}$ is independent of the choice of $(\mathbf{X}'\mathbf{X})^{-}$, we can express $\mathrm{OLSE}(\mathbf{K}'\boldsymbol{\beta} \mid \mathscr{M})$ also (using the Moore–Penrose inverse) as

$$\mathbf{K}'\hat{\boldsymbol{\beta}} = \mathbf{K}'\mathbf{X}^{+}\mathbf{y}. \tag{10}$$

Because the $\mathbf{K}'\hat{\boldsymbol{\beta}}$ is independent of the covariance matrix $\mathbf{V}$, $\mathbf{K}'\hat{\boldsymbol{\beta}}$ is in many situations an attractive alternative to the BLUE. For example, in many practical situations the covariance matrix $\mathbf{V}$ may not be completely known. That is, only the structure of $\mathbf{V}$ is known but its elements remain to be functions of some unknown parameters, presented in vector $\boldsymbol{\theta}$, say. Hence, in such situations, one may rather consider using the OLSE as an estimator to $\mathbf{K}'\boldsymbol{\beta}$ than using the estimated covariance matrix $\hat{\mathbf{V}} = \mathbf{V}(\hat{\boldsymbol{\theta}})$ as a plug-in estimator for $\mathbf{V}$ in any of the BLUE representation.

1.4 The best linear unbiased predictor

In addition to estimation of linear parametric functions, in this thesis we are also interested in prediction of new observations in the general Gauss–Markov model. That is, let $\mathbf{y}_f$ denote an $m \times 1$ unobservable random vector containing new observations (observable in future). New observations $\mathbf{y}_f$ are assumed to follow linear model

$$\mathbf{y}_f = \mathbf{X}_f\boldsymbol{\beta} + \boldsymbol{\varepsilon}_f, \tag{11}$$

where $\mathbf{X}_f$ is a known $m \times p$ model matrix associated with new observations, $\boldsymbol{\beta}$ is the same vector of unknown parameters as in (1), and $\boldsymbol{\varepsilon}_f$ is an $m \times 1$ random error vector associated with new observations. The expectation vector and the covariance matrix of $(\mathbf{y}' : \mathbf{y}_f')'$ are now assumed to have forms

$$\mathrm{E}\begin{pmatrix}\mathbf{y}\\ \mathbf{y}_f\end{pmatrix} = \begin{pmatrix}\mathbf{X}\\ \mathbf{X}_f\end{pmatrix}\boldsymbol{\beta} \quad \text{and} \quad \mathrm{cov}\begin{pmatrix}\mathbf{y}\\ \mathbf{y}_f\end{pmatrix} = \sigma^2\begin{pmatrix}\mathbf{V}_y & \mathbf{V}_{yf}\\ \mathbf{V}_{yf}' & \mathbf{V}_f\end{pmatrix} = \sigma^2\boldsymbol{\Omega}, \tag{12}$$

respectively, where $\sigma^2 > 0$ is an unknown scalar and $\boldsymbol{\Omega}$ is a known nonnegative definite matrix. Again, we use the short notation

$$\mathscr{M}_f = \left\{\begin{pmatrix}\mathbf{y}\\ \mathbf{y}_f\end{pmatrix}, \begin{pmatrix}\mathbf{X}\boldsymbol{\beta}\\ \mathbf{X}_f\boldsymbol{\beta}\end{pmatrix}, \sigma^2\begin{pmatrix}\mathbf{V}_y & \mathbf{V}_{yf}\\ \mathbf{V}_{yf}' & \mathbf{V}_f\end{pmatrix}\right\} \tag{13}$$

to describe the general Gauss–Markov model in a case of containing new (not yet observed) observations.

Under the model $\mathscr{M}_f$, we may consider the problem of predicting the values of new observations $\mathbf{y}_f$ based on the observable random vector $\mathbf{y}$. In this thesis, we only consider linear predictors of $\mathbf{y}_f$. A linear statistic $\mathbf{Gy}$ is said to be a linear unbiased predictor of $\mathbf{y}_f$ if

$$\mathrm{E}(\mathbf{Gy}) = \mathrm{E}(\mathbf{y}_f) = \mathbf{X}_f\boldsymbol{\beta} \qquad \text{for all } \boldsymbol{\beta} \in \mathbb{R}^p, \tag{14}$$

i.e., if the expected prediction error is zero. Clearly (14) is equivalent to $\mathbf{GX} = \mathbf{X}_f$, we say that $\mathbf{y}_f$ is unbiasedly predictable if there exists a matrix $\mathbf{G}$ with property $\mathbf{GX} = \mathbf{X}_f$, i.e., $\mathscr{C}(\mathbf{X}_f') \subset \mathscr{C}(\mathbf{X}')$.

Moreover, a linear unbiased predictor $\mathbf{Gy}$ is said to be the best linear unbiased predictor, BLUP, of $\mathbf{y}_f$ under the model $\mathscr{M}_f$, if for any other unbiased linear predictor $\mathbf{Fy}$ the difference $\mathrm{cov}(\mathbf{Fy} - \mathbf{y}_f) - \mathrm{cov}(\mathbf{Gy} - \mathbf{y}_f)$ is a nonnegative definite matrix, i.e.,

$$\mathrm{cov}(\mathbf{Fy} - \mathbf{y}_f) - \mathrm{cov}(\mathbf{Gy} - \mathbf{y}_f) \geq \mathbf{0} \quad \text{for all } \mathbf{Fy} \text{ such that } \mathrm{E}(\mathbf{Fy}) = \mathbf{X}_f\boldsymbol{\beta}. \tag{15}$$

Goldberger (1962) showed, that if $\mathbf{V}_y$ is positive definite, then the BLUP of $\mathbf{y}_f$ has a form

$$\mathrm{BLUP}(\mathbf{y}_f \mid \mathscr{M}_f) = \mathbf{X}_f\tilde{\boldsymbol{\beta}} + \mathbf{V}_{yf}'\mathbf{V}_y^{-1}(\mathbf{y} - \mathbf{X}\tilde{\boldsymbol{\beta}}), \tag{16}$$

where $\mathbf{X}_f\tilde{\boldsymbol{\beta}}$ and $\mathbf{X}\tilde{\boldsymbol{\beta}}$ are the best linear unbiased estimators, BLUEs, of $\mathbf{X}_f\boldsymbol{\beta}$ and $\mathbf{X}\boldsymbol{\beta}$, respectively. A more general representation of the BLUP of $\mathbf{y}_f$, which is also applicable to the case of singular covariance matrix $\mathbf{V}_y$, is given, e.g., in Sengupta & Jammalamadaka (2003, Section 7.13).

2 Linear inference

2.1 Linear sufficiency

Linear sufficiency and linear prediction sufficiency are one of the most important concepts of this thesis, and are very much related to the best linear unbiased estimation and the best linear unbiased prediction defined in sections 1.2 and 1.4.

The concept of linear sufficiency was introduced by Barnard (1963), Baksalary & Kala (1981), and Drygas (1983)—who was the first to use the term linear sufficiency—while investigating those linear statistics $\mathbf{Ty}$, which are "sufficient" for estimation of the expected value $\mathbf{X}\boldsymbol{\beta}$ in the general Gauss–Markov model $\mathscr{M}$. Formally, a linear statistic $\mathbf{Ty}$ is defined to be linearly sufficient for $\mathbf{X}\boldsymbol{\beta}$ under the model $\mathscr{M}$ if there exists a matrix $\mathbf{A}$ such that

$\mathbf{ATy}$ is the BLUE of $\mathbf{X}\boldsymbol{\beta}$. Baksalary & Kala (1981) and Drygas (1983) showed that a linear statistic $\mathbf{Ty}$ is linearly sufficient for $\mathbf{X}\boldsymbol{\beta}$ under the model $\mathscr{M}$ if and only if the column space inclusion

$$\mathscr{C}(\mathbf{X}) \subset \mathscr{C}(\mathbf{WT}') \tag{17}$$

holds; here $\mathbf{W} = \mathbf{V} + \mathbf{XUX}'$ with $\mathbf{U}$ being an arbitrary nonnegative definite matrix such that $\mathscr{C}(\mathbf{W}) = \mathscr{C}(\mathbf{X} : \mathbf{V})$.

2.2 Linear minimal sufficiency and linear completeness

In addition to linear sufficiency, Drygas (1983) also considered related concepts of linear minimal sufficiency and linear completeness. A linearly sufficient statistic $\mathbf{Ty}$ is called linearly minimal sufficient for $\mathbf{X}\boldsymbol{\beta}$ under the model $\mathscr{M}$, if for any other linearly sufficient statistic $\mathbf{Sy}$, there exists a matrix $\mathbf{A}$ such that $\mathbf{Ty} = \mathbf{ASy}$ almost surely. Drygas (1983) showed that $\mathbf{Ty}$ is linearly minimal sufficient for $\mathbf{X}\boldsymbol{\beta}$ if and only if the equality

$$\mathscr{C}(\mathbf{X}) = \mathscr{C}(\mathbf{WT}') \tag{18}$$

holds.

Moreover, Drygas (1983) called a linear statistic $\mathbf{Ty}$ linearly complete if for every linear transformation of it, $\mathbf{LTy}$, such that $\mathrm{E}(\mathbf{LTy}) = \mathbf{0}$, it follows that $\mathbf{LTy} = \mathbf{0}$ almost surely. According to Drygas (1983), a linear statistic $\mathbf{Ty}$ is linearly complete if and only if

$$\mathscr{C}(\mathbf{TV}) \subset \mathscr{C}(\mathbf{TX}). \tag{19}$$

It was also shown by Drygas (1983) that a linear statistic $\mathbf{Ty}$ is linearly minimal sufficient for $\mathbf{X}\boldsymbol{\beta}$ if and only if it is simultaneously linearly sufficient and linearly complete for $\mathbf{X}\boldsymbol{\beta}$. Further properties on linear sufficiency, minimal sufficiency, and completeness in a case of estimation of $\mathbf{X}\boldsymbol{\beta}$ were then provided by Müller, Rao & Sinha (1984) and Müller (1987), see also Baksalary & Mathew (1986).

The notions of linear sufficiency and linear minimal sufficiency were extended to estimation of the given estimable parametric function $\mathbf{K}'\boldsymbol{\beta}$ by Baksalary & Kala (1986). Baksalary & Kala (1986) proved that $\mathbf{Ty}$ is linearly sufficient for $\mathbf{K}'\boldsymbol{\beta}$ under the model $\mathscr{M}$ if and only if the null space inclusion

$$\mathscr{N}(\mathbf{TX} : \mathbf{TVX}^{\perp}) \subset \mathscr{N}(\mathbf{K}' : \mathbf{0}) \tag{20}$$

holds, and $\mathbf{Ty}$ is linearly minimal sufficient for $\mathbf{K}'\boldsymbol{\beta}$ if and only if the null space equality

$$\mathscr{N}(\mathbf{TX} : \mathbf{TVX}^{\perp}) = \mathscr{N}(\mathbf{K}' : \mathbf{0}) \tag{21}$$

holds.

In article [1], we consider linear sufficiency in the context of estimating the parametric function $\mathbf{X}_1\boldsymbol{\beta}_1$ in the partitioned linear model $\mathscr{M}_{12} = \{\mathbf{y},\, \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2,\, \sigma^2\mathbf{V}\}$. In [1], we give some new characterizations for linear sufficiency and linear minimal sufficiency in a case of estimation of the parametric function $\mathbf{X}_1\boldsymbol{\beta}_1$. We also define and consider linear completeness in the context of estimating $\mathbf{X}_1\boldsymbol{\beta}_1$, and prove that a linear statistic $\mathbf{Ty}$ is simultaneously linearly sufficient and linearly complete for $\mathbf{X}_1\boldsymbol{\beta}_1$ under the partitioned model $\mathscr{M}_{12}$ if and only if it is a linearly minimal sufficient statistic for $\mathbf{X}_1\boldsymbol{\beta}_1$.

In article [2], we present further results on linear sufficiency in a case of estimation of the given estimable parametric function $\mathbf{K}'\boldsymbol{\beta}$ in the general Gauss–Markov model $\mathscr{M}$. Article [2] is closely connected to article [1]. The more general results given in article [2] have been obtained by first reparametrizing the general Gauss–Markov model $\mathscr{M}$ into particularly partitioned linear model and then by using the results given in article [1].

By a reparametrized model of $\mathscr{M}$ we mean the linear model

$$\mathscr{M}_{\boldsymbol{\gamma}} = \{\mathbf{y},\, \mathbf{X}_*\boldsymbol{\gamma},\, \sigma^2\mathbf{V}\}, \tag{22}$$

where $\mathbf{X}_*$ is any matrix such that $\mathscr{C}(\mathbf{X}) = \mathscr{C}(\mathbf{X}_*)$. That is, if $\mathbf{X}_*$ is a matrix with $\mathscr{C}(\mathbf{X}) = \mathscr{C}(\mathbf{X}_*)$, then there exists a matrix $\mathbf{A}$ such that $\mathbf{X} = \mathbf{X}_*\mathbf{A}$. Further if we define $\boldsymbol{\gamma} = \mathbf{A}\boldsymbol{\beta}$, then we can see that the models $\mathscr{M}$ and $\mathscr{M}_{\boldsymbol{\gamma}}$ are equivalent, and we can call the model $\mathscr{M}_{\boldsymbol{\gamma}}$ as a reparametrized model of $\mathscr{M}$; see, e.g., Peixoto (1993) and Sengupta & Jammalamadaka (2003, pp. 118–120).

Consider now estimation of the given estimable parametric function $\mathbf{K}'\boldsymbol{\beta}$, $\mathbf{K} \in \mathbb{R}^{p\times k}$, under the model $\mathscr{M}$. Then the column space of the matrix $\mathbf{X}(\mathbf{K} : \mathbf{K}^{\perp})$ equals the column space of $\mathbf{X}$, and thus the partitioned model

$$\begin{aligned}\mathscr{M}_{\boldsymbol{\gamma}} &= \{\mathbf{y},\, \mathbf{X}(\mathbf{K} : \mathbf{K}^{\perp})\gamma,\, \sigma^2\mathbf{V}\} \\ &= \{\mathbf{y},\, \mathbf{XK}\boldsymbol{\gamma}_1 + \mathbf{XK}^{\perp}\boldsymbol{\gamma}_2,\, \sigma^2\mathbf{V}\},\end{aligned} \tag{23}$$

where $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1', \boldsymbol{\gamma}_2')'$, is equivalent to the original Gauss–Markov model $\mathscr{M}$. It is now shown in article [2] that a linearly sufficient statistic for $\mathbf{K}'\boldsymbol{\beta}$ in the general Gauss–Markov model $\mathscr{M}$ is also a linearly sufficient statistic for $\mathbf{XK}\boldsymbol{\gamma}_1$ in the partitioned reparametrized model $\mathscr{M}_{\boldsymbol{\gamma}}$, and thus the results obtained in article [1] can be used for characterizing linear sufficiency in a case of estimation of $\mathbf{K}'\boldsymbol{\beta}$.

In articles [1] and [2], we also consider a predictive approach for obtaining the best linear unbiased estimator of $\mathbf{X}_1\boldsymbol{\beta}_1$ and $\mathbf{K}'\boldsymbol{\beta}$, respectively. Sengupta

& Jammalamadaka (2003, Chapter 11) gave an interesting study on the linear version of the general estimation theory, including the linear analogues to the Rao–Blackwell and Lehmann–Scheffé Theorems when considering estimation of the expected value $\mathbf{X}\boldsymbol{\beta}$ under the model $\mathscr{M}$. In articles [1] and [2], we give the corresponding linear analogues of the Rao–Blackwell and Lehmann–Scheffé Theorems in the context of estimating $\mathbf{X}_1\boldsymbol{\beta}_1$ and $\mathbf{K}'\boldsymbol{\beta}$, respectively.

2.3 Linear prediction sufficiency

The concepts of linear prediction sufficiency and linear minimal prediction sufficiency are then considered in article [3]. Formally, we define a linear statistic $\mathbf{Ty}$ to be linearly prediction sufficient for new observations $\mathbf{y}_f$ under the general Gauss–Markov model $\mathscr{M}_f$ with new observations if there exists a matrix $\mathbf{A}$ such that $\mathbf{ATy}$ is the BLUP of $\mathbf{y}_f$, see [3, Definition 1].

In article [3], we give some equivalent characterizations for these new concepts of linear prediction sufficiency and linear minimal prediction sufficiency. These characterizations are similar to the characterizations of linear sufficiency given by Drygas (1983), Baksalary & Kala (1986), and Müller (1987), and to the characterizations of the concept of *linear error-sufficiency* introduced by Groß (1998). However, not all linear versions of the important concepts of mathematical statistics are considered in article [3]. For example, the notation of linear completeness in a case of prediction of $\mathbf{y}_f$ is not investigated in that article.

3 Efficiency of the OLSE

3.1 The Watson efficiency

One further important concept of this thesis is the Watson efficiency. Let the vector of parameters $\boldsymbol{\beta}$ itself be estimable under the model $\mathscr{M}$, and let us further assume that the covariance matrix $\mathbf{V}$ is positive definite. Then the ordinary least squares estimator, OLSE, and the best linear unbiased estimator, BLUE, of $\boldsymbol{\beta}$ under the model $\mathscr{M}$ are, respectively,

$$\mathrm{OLSE}(\boldsymbol{\beta} \mid \mathscr{M}) = \hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}, \tag{24}$$

$$\mathrm{BLUE}(\boldsymbol{\beta} \mid \mathscr{M}) = \tilde{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y}, \tag{25}$$

with the corresponding covariance matrices being

$$\operatorname{cov}(\hat{\boldsymbol{\beta}} \mid \mathscr{M}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}, \tag{26}$$

$$\operatorname{cov}(\tilde{\boldsymbol{\beta}} \mid \mathscr{M}) = \sigma^2(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}. \tag{27}$$

Both the OLSE and the BLUE are obviously unbiased estimators of $\boldsymbol{\beta}$ but

$$\operatorname{cov}(\hat{\boldsymbol{\beta}} \mid \mathscr{M}) \geq \operatorname{cov}(\tilde{\boldsymbol{\beta}} \mid \mathscr{M}). \tag{28}$$

Hence we may want to know how "bad" or inefficient the OLSE of $\boldsymbol{\beta}$ could be with respect to the BLUE of $\boldsymbol{\beta}$. Clearly, there is no unique way to measure the efficiency of the OLSE with respect to the BLUE. However, almost certainly the most frequently used measure is the Watson efficiency (Watson 1955, p. 330) defined as the ratio of the determinants of the covariance matrices of the BLUE($\boldsymbol{\beta} \mid \mathscr{M}$) and the OLSE($\boldsymbol{\beta} \mid \mathscr{M}$):

$$\operatorname{eff}(\hat{\boldsymbol{\beta}} \mid \mathscr{M}) = \frac{|\operatorname{cov}(\tilde{\boldsymbol{\beta}} \mid \mathscr{M})|}{|\operatorname{cov}(\hat{\boldsymbol{\beta}} \mid \mathscr{M})|} = \frac{|\mathbf{X}'\mathbf{X}|^2}{|\mathbf{X}'\mathbf{V}\mathbf{X}| \cdot |\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}|}\,. \tag{29}$$

It clearly holds that

$$0 < \operatorname{eff}(\hat{\boldsymbol{\beta}} \mid \mathscr{M}) \leq 1, \tag{30}$$

where the upper bound is attained if and only if the OLSE equals the BLUE, see, e.g., Puntanen & Styan (1989). For a lower bound of the efficiency, see Bloomfield & Watson (1975), Knott (1975); for the related geometry (and antieigenvalues), see, e.g., Gustafson (2002, 2006). In articles [4], [5], and [6], we now consider properties of the Watson efficiency in the partitioned linear model $\mathscr{M}_{12} = \{\mathbf{y},\, \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2,\, \sigma^2\mathbf{V}\}$. Particularly, we consider the Watson efficiency of the OLSE of the subvector $\boldsymbol{\beta}_2$ defined as the ratio

$$\operatorname{eff}(\hat{\boldsymbol{\beta}}_2 \mid \mathscr{M}_{12}) = \frac{|\operatorname{cov}(\tilde{\boldsymbol{\beta}}_2 \mid \mathscr{M}_{12})|}{|\operatorname{cov}(\hat{\boldsymbol{\beta}}_2 \mid \mathscr{M}_{12})|}. \tag{31}$$

In articles [4] and [5] we obtain some theoretical results on the relationship between the Watson efficiency of the OLSE of the whole parametric vector $\boldsymbol{\beta}$, $\operatorname{eff}(\hat{\boldsymbol{\beta}} \mid \mathscr{M}_{12})$, and the Watson efficiency of the OLSE of the subvector $\boldsymbol{\beta}_2$, $\operatorname{eff}(\hat{\boldsymbol{\beta}}_2 \mid \mathscr{M}_{12})$, and then in article [6] we present some real data demostrations about the relation between $\operatorname{eff}(\hat{\boldsymbol{\beta}} \mid \mathscr{M}_{12})$ and $\operatorname{eff}(\hat{\boldsymbol{\beta}}_2 \mid \mathscr{M}_{12})$.

3.2 The reduced model

One useful tool in our analyses concerning the Watson efficiency of $\hat{\boldsymbol{\beta}}_2$ is the so-called reduced model

$$\mathscr{M}_{12\cdot 1} = \{\mathbf{M}_1\mathbf{y},\, \mathbf{M}_1\mathbf{X}_2\boldsymbol{\beta}_2,\, \sigma^2\mathbf{M}_1\mathbf{V}\mathbf{M}_1\}, \tag{32}$$

where $\mathbf{M}_1 = \mathbf{I} - \mathbf{P}_{\mathbf{X}_1}$. The reduced model $\mathscr{M}_{12\cdot 1}$ is obtained by premultiplying the partitioned linear model $\mathscr{M}_{12}$ by the orthogonal projector $\mathbf{M}_1$. Usefulness of the reduced model $\mathscr{M}_{12\cdot 1}$ arises from the fact that both the OLSE and the BLUE of the subvector $\boldsymbol{\beta}_2$ under the reduced model $\mathscr{M}_{12\cdot 1}$ equal, respectively, the OLSE and the BLUE of the subvector $\boldsymbol{\beta}_2$ under the partitioned model $\mathscr{M}_{12}$, see Frisch & Waugh (1933), Lovell (1963), Groß & Puntanen (2000, Th. 4), and Bhimasankaram & Sengupta (1996, Th. 6.1). Hence also the Watson efficiency of $\hat{\boldsymbol{\beta}}_2$ under the reduced model $\mathscr{M}_{12\cdot 1}$ equals the Watson efficiency of $\hat{\boldsymbol{\beta}}_2$ under the partitioned model $\mathscr{M}_{12}$.

The Watson efficiency of the OLSE of the given estimable parametric function $\mathbf{K}'\boldsymbol{\beta}$ is then considered in article [9]. We have shown in article [9] that the Watson efficiency of $\mathbf{K}'\hat{\boldsymbol{\beta}}$ (when $\mathrm{r}(\mathbf{K}) = k$) under the model $\mathscr{M}$ equals to the Watson efficiency of the OLSE of the subvector $\boldsymbol{\gamma}_1$ under the reparametrized model $\mathscr{M}_{\boldsymbol{\gamma}}$. Since the reparametrized model $\mathscr{M}_{\boldsymbol{\gamma}}$ can be viewed as a certain partitioned model, the results presented in article [9] are hence based on the results established in articles [4] and [5].

3.3 The equality of the OLSE and BLUE

In articles [7] and [8] we consider the situation of the Watson efficiency being one, i.e., the situation when the ordinary least squares estimator equals the best linear unbiased estimator. In article [7] we investigate conditions for the equality between the OLSE and the BLUE of the subvector $\boldsymbol{\beta}_1$ under the partitioned model $\mathscr{M}_{12}$ assuming that they are first equal under the so-called small model

$$\mathscr{M}_1 = \{\mathbf{y},\, \mathbf{X}_1\boldsymbol{\beta}_1,\, \sigma^2\mathbf{V}\}. \tag{33}$$

Then in article [8] we give new characterizations for the equality between the OLSE and the BLUE of the given estimable parametric function $\mathbf{K}'\boldsymbol{\beta}$ under the general Gauss–Markov model $\mathscr{M}$.

4 Some other considerations

4.1 The linear mixed effects model

One particular type of the general Gauss–Markov model is the linear mixed effects model. That is, suppose in equation (1) we have a reason to model the random error term $\boldsymbol{\varepsilon}$ as

$$\boldsymbol{\varepsilon} = \mathbf{Z}\boldsymbol{\alpha} + \mathbf{u}, \tag{34}$$

where $\boldsymbol{\alpha}$ is a $q \times 1$ vector of random effects with corresponding a known $n \times q$ model matrix $\mathbf{Z}$, and where $\mathbf{u}$ is an $n \times 1$ random error vector uncorrelated with $\boldsymbol{\alpha}$, i.e., $\operatorname{cov}(\boldsymbol{\alpha}, \mathbf{u}) = \mathbf{0}$.

Since in the original model $\mathscr{M}$ we assume that $\mathrm{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$, we now make also an assumption that $\mathrm{E}(\boldsymbol{\alpha}) = \mathbf{0}$ and $\mathrm{E}(\mathbf{u}) = \mathbf{0}$. Let us further assume that the covariance matrices $\operatorname{cov}(\boldsymbol{\alpha}) = \mathbf{D}$ and $\operatorname{cov}(\mathbf{u}) = \mathbf{R}$ are both fully known, and thus the original model equation (1) becomes the mixed effects model equation

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\alpha} + \mathbf{u}, \tag{35}$$

where $\operatorname{cov}(\mathbf{y}) = \operatorname{cov}(\mathbf{Z}\boldsymbol{\alpha} + \mathbf{u}) = \mathbf{ZDZ}' + \mathbf{R} = \mathbf{V}$. In short, we may denote the linear mixed effects model as

$$\mathscr{M}_m = \{\mathbf{y},\, \mathbf{X}\boldsymbol{\beta},\, \mathbf{ZDZ}' + \mathbf{R}\} = \{\mathbf{y},\, \mathbf{X}\boldsymbol{\beta},\, \mathbf{V}\}. \tag{36}$$

Note that we now assume that the linear mixed effects model $\mathscr{M}_m$ does not include the unknown scalar σ^2.

In article [10] we consider estimation of the parametric function $\mathbf{X}_1\boldsymbol{\beta}_1$ under the following partitioned linear model

$$\mathscr{M}_{12} = \{\mathbf{y},\, \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2,\, \mathbf{R}\} \tag{37}$$

and under the following linear mixed effects model

$$\mathscr{M}_m = \{\mathbf{y},\, \mathbf{X}_1\boldsymbol{\beta}_1,\, \mathbf{X}_2\mathbf{D}\mathbf{X}_2' + \mathbf{R}\}. \tag{38}$$

The difference between models (37) and (38) is that in the mixed effects model (38) the parameters associated with the model matrix $\mathbf{X}_2$ are considered to be random where as in the partitioned model (37) they are considered to be fixed.

In article [10] we are interested in characterizing when the BLUE of $\mathbf{X}_1\boldsymbol{\beta}_1$ under the partitioned linear model (37) equals the corresponding BLUE under the linear mixed effects model (38). Interest for this research problem arises from the fact that in many practical situations it can be difficult to determine whether some of the fixed parameters in the partitioned model should actually be treated as random variables; see, e.g., Searle, Casella & McCulloch (1992, Section 1.4).

4.2 A useful matrix decomposition

In the last article of this thesis we take more thorough look at a particular matrix decomposition used in almost all other articles of this thesis. That is, in article [11] we consider properties and some statistical applications of the matrix $\dot{\mathbf{M}}$ defined as

$$\dot{\mathbf{M}} = \mathbf{M}(\mathbf{MVM})^{-}\mathbf{M}, \tag{39}$$

where $\mathbf{M} = \mathbf{I} - \mathbf{P}_{\mathbf{X}}$ and $\mathbf{V}$ is the covariance matrix from the general Gauss–Markov model $\mathscr{M}$. In general, the matrix $\dot{\mathbf{M}}$ is not necessarily unique with respect to the choice of $(\mathbf{MVM})^{-}$. However, when $\mathbf{V} > \mathbf{0}$, the matrix $\dot{\mathbf{M}}$ is unique and has a decomposition

$$\dot{\mathbf{M}} = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}\big(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\big)^{-}\mathbf{X}'\mathbf{V}^{-1}. \tag{40}$$

Thus, for example, the Watson efficiency of the OLSE of the parameter vector $\boldsymbol{\beta}$ under the general Gauss–Markov model $\mathscr{M}$ can be expressed as

$$\mathrm{eff}(\hat{\boldsymbol{\beta}} \mid \mathscr{M}) = \det\big[\mathbf{I} - \mathbf{X}'\mathbf{V}\dot{\mathbf{M}}\mathbf{V}\mathbf{X}(\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}\big]. \tag{41}$$

Related to the matrix $\dot{\mathbf{M}}$, we also consider in article [11] properties of the matrix $\dot{\mathbf{M}}_2$ defined as

$$\dot{\mathbf{M}}_2 = \mathbf{M}_2(\mathbf{M}_2\mathbf{V}\mathbf{M}_2)^{-}\mathbf{M}_2, \tag{42}$$

where $\mathbf{M}_2 = \mathbf{I} - \mathbf{P}_{\mathbf{X}_2}$. The matrix $\dot{\mathbf{M}}_2$ plays a major role in estimation of the parametric function $\mathbf{X}_1\boldsymbol{\beta}_1$ under the partitioned linear model $\mathscr{M}_{12}$. For example, a statistic $\mathbf{X}_1'\dot{\mathbf{M}}_2\mathbf{y}$ is linearly minimal sufficient statistic for $\mathbf{X}_1\boldsymbol{\beta}_1$ under the model $\mathscr{M}_{12}$ as shown in [1].

5 Abstracts of original articles

[1] **Linear sufficiency and completeness in the partitioned linear model**

In this paper we consider the estimation of $\mathbf{X}_1\boldsymbol{\beta}_1$ under the partitioned linear model $\{\mathbf{y},\, \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2,\, \sigma^2\mathbf{V}\}$. In particular, we consider linear sufficiency and linear completeness of $\mathbf{X}_1\boldsymbol{\beta}_1$. We give new characterizations for linear sufficiency of $\mathbf{X}_1\boldsymbol{\beta}_1$, and define and characterize linear completeness in a case of the estimation of $\mathbf{X}_1\boldsymbol{\beta}_1$. We also introduce a predictive approach for obtaining the best linear unbiased estimator of $\mathbf{X}_1\boldsymbol{\beta}_1$, and subsequently, we give the linear analogues of the Rao–Blackwell and Lehmann–Scheffé Theorems in the context of estimating $\mathbf{X}_1\boldsymbol{\beta}_1$.

[2] **Linear sufficiency and completeness in the context of estimating the parametric function in the general Gauss–Markov model**

In this paper we consider linear sufficiency and linear completeness in the context of estimating the estimable parametric function $\mathbf{K}'\boldsymbol{\beta}$ under the general Gauss–Markov model $\{\mathbf{y},\, \mathbf{X}\boldsymbol{\beta},\, \sigma^2\mathbf{V}\}$. We give new characterizations for linear sufficiency, and define and characterize linear completeness in a case of the estimation of $\mathbf{K}'\boldsymbol{\beta}$. Also, we consider a predictive approach for obtaining the best linear unbiased estimator of $\mathbf{K}'\boldsymbol{\beta}$, and subsequently, we give the linear analogues of the Rao–Blackwell and Lehmann–Scheffé Theorems in the context of estimating $\mathbf{K}'\boldsymbol{\beta}$.

[3] **Linear prediction sufficiency for new observations in the general Gauss–Markov model**

We consider the prediction of new observations in a general Gauss–Markov model. We state the fundamental equations of the best linear unbiased prediction, BLUP, and consider some properties of the BLUP. Particularly, we focus on such linear statistics, which preserve enough information for obtaining the BLUP of new observations as a linear function of them. We call such statistics linearly prediction sufficient for new observations, and introduce some equivalent characterizations for this new concept.

[4] **On decomposing the Watson efficiency of ordinary least squares in a partitioned weakly singular linear model**

We consider the estimation of regression coefficients in a partitioned weakly singular linear model and focus on questions concerning the Watson efficiency of the ordinary least squares estimator of a subset of the parameters with respect to the best linear unbiased estimator. Certain submodels are also considered. The conditions under which the Watson efficiency in the full model splits into a function of some other Watson efficiencies is given special attention. In particular, a new decomposition of the Watson efficiency into a product of three particular factors appears to be very useful.

[5] **Some further results concerning the decomposition of the Watson efficiency in partitioned linear models**

While considering the estimation of regression coefficients in a partitioned weakly singular linear model, Chu, Isotalo, Puntanen and Styan (2004) introduced a particular decomposition for the Watson efficiency of the ordinary least squares estimator. This decomposition presents the "total" Watson efficiency as a product of three factors. In this paper we give new insight into the decomposition showing that all three factors are related to the efficiencies of particular submodels or their transformed versions. Moreover, we prove an interesting connection between a particular split of the Watson efficiency and the concept of linear sufficiency. We shortly review the relation between the efficiency and specific canonical correlations. We also introduce the corresponding decomposition for the Bloomfield–Watson commutator criterion, and give a necessary and sufficient condition for its specific split.

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

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.

[7] **Effect of adding regressors on the equality of the OLSE and BLUE**

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, $\mathscr{M}_1 = \{\mathbf{y}, \mathbf{X}_1\boldsymbol{\beta}_1, \sigma^2\mathbf{V}\}$ 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.

[8] **A note on the equality of the OLSE and the BLUE of the parametric function in the general Gauss–Markov model**

In this note we consider the equality of the ordinary least squares estimator (OLSE) and the best linear unbiased estimator (BLUE) of the estimable parametric function in the general Gauss–Markov model. Especially we consider the structures of the covariance matrix $\mathbf{V}$ for which the OLSE equals the BLUE. Our results are based on the properties of a particular reparametrized version of the original Gauss–Markov model.

[9] **Some comments on the Watson efficiency of the ordinary least squares estimator under the Gauss–Markov model**

We consider the estimation of a given estimable parametric function in the Gauss–Markov model, and focus on questions concerning the Watson efficiency of the ordinary least squares estimator (OLSE) of the given parametric function with respect to the best linear unbiased estimator (BLUE). We apply the Frisch–Waugh–Lovell Theorem for the estimation of the parametric function, and give an interesting decomposition of the total Watson efficiency with respect to the efficiency of the parametric function. Also, a relation between the Watson efficiency of the OLSE of the given parametric function and specific canonical correlations is established.

[10] **Invariance of the BLUE under the linear fixed and mixed effects models**

We consider the estimation of the parametric function $\mathbf{X}_1\boldsymbol{\beta}_1$ under the partitioned linear fixed effects model $\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}$ and the linear mixed effects model $\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\gamma}_2 + \boldsymbol{\varepsilon}$, where $\boldsymbol{\gamma}_2$ is considered to be a random vector. Particularly, we consider when the best linear unbiased estimator, BLUE, of $\mathbf{X}_1\boldsymbol{\beta}_1$ under the linear fixed effects model equals the corresponding BLUE under the linear mixed effects model.

[11] **A useful matrix decomposition and its statistical applications in linear regression**

It is well known that if $\mathbf{V}$ is a symmetric positive definite $n \times n$ matrix, and $(\mathbf{X} : \mathbf{Z})$ is a partitioned orthogonal $n \times n$ matrix, then

$$(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} = \mathbf{X}'\mathbf{V}\mathbf{X} - \mathbf{X}'\mathbf{V}\mathbf{Z}(\mathbf{Z}'\mathbf{V}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{V}\mathbf{X}. \qquad (*)$$

In this paper we show how useful we have found the formula $(*)$, and in particular its version

$$\mathbf{Z}(\mathbf{Z}'\mathbf{V}\mathbf{Z})^{-1}\mathbf{Z}' = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1} := \dot{\mathbf{M}}, \quad (**)$$

and present several related formulas, as well as some generalized versions. We also include several statistical applications.

6 Errata and completions to the articles

Article [1]:

– p. 56, line 13 ↑: Printed: $\mathbf{W}_1 = \mathbf{V} + \mathbf{X}_1\mathbf{X}_1$,
Should be: $\mathbf{W}_1 = \mathbf{V} + \mathbf{X}_1\mathbf{X}_1'$.

Article [3]:

– p. 1021, line 2 ↓: Printed: $\mathbf{BT}(\mathbf{X} : \mathbf{VX}^{\perp}) = \mathbf{S}(\mathbf{X} : \mathbf{VX}^{\perp})$,
Should be: $\mathbf{T}(\mathbf{X} : \mathbf{VX}^{\perp}) = \mathbf{BS}(\mathbf{X} : \mathbf{VX}^{\perp})$.

– p. 1021, line 1 ↑: Printed: $\mathbf{w} = \sigma^2(1, 2, 3, \ldots, T-1, T)'$,
Should be: $\mathbf{w} = (1, 2, 3, \ldots, T-1, T)'$.

Article [5]:

– p. 83, equation (3.2): Equation (3.2) holds, since (under a weakly singular model)

$$\begin{aligned}\operatorname{cov}(\hat{\boldsymbol{\beta}}, \tilde{\boldsymbol{\beta}}) &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}\mathbf{V}^{+}\mathbf{X}(\mathbf{X}'\mathbf{V}^{+}\mathbf{X})^{-1}\\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{V}^{+}\mathbf{X})^{-1}\\ &= \sigma^2(\mathbf{X}'\mathbf{V}^{+}\mathbf{X})^{-1} = \operatorname{cov}(\tilde{\boldsymbol{\beta}}).\end{aligned}$$

Article [8]:

– p. 7, line 2 ↑: Typing error. Should be: "yet at the same time".

Article [11]:

– p. 7, Lemma 2.1: Some important conditions can be added, e.g.,

(i) $\mathbf{HP_V} = \mathbf{P_V H}$,

(ii) $\mathscr{C}(\mathbf{P_V H}) = \mathscr{C}(\mathbf{P_V}) \cap \mathscr{C}(\mathbf{H})$,

(iii) $\mathrm{r}(\mathbf{HP_V}) = \dim[\mathscr{C}(\mathbf{V}) \cap \mathscr{C}(\mathbf{X})]$.

7 Author's contribution to the articles

All eleven articles in this thesis are joint research between myself and different co-authors. First three articles [1], [2], and [3] dealing with linear sufficiency and linear prediction sufficiency were accomplished in collaboration between me and Dr. Simo Puntanen (University of Tampere, Finland).

Since Dr. Puntanen is actually co-author in all of the eleven articles, and more importantly the supervisor of this thesis, the collaboration between us has appeared in many levels, both formal and informal, during this study. As the supervisor of this thesis, Dr. Puntanen has guided and conducted my research. He has given new research problems, commented the results achieved by me, and provided references to earlier research articles, etc. Also all eleven articles has been proofread and fine-tuned together by me and Dr. Puntanen.

In case of articles [1], [2], and [3], I proposed the general themes of the articles. That is, I started to explore the subjects of linear sufficiency and linear prediction sufficiency, and later on carried out initial proofs of new results, and organized the contents of the articles. The initial manuscripts prepared by me were then substantially improved and clarified jointly with Dr. Puntanen.

Articles [4], [5], and [6] concerning the Watson efficiency are joint research together with Dr. Ka Lok Chu (Dawson College, Montréal, Canada), Dr. Puntanen, and Prof. George P. H. Styan (McGill University, Montréal, Canada). Article [4] is based on the technical reports Chu & Styan (2003) and Isotalo & Puntanen (2003). During the 12th International Workshop on Matrices and Statistics (Dortmund, Germany, August 2003), it was decided to combine some of the results established in the reports Chu & Styan (2003) and Isotalo & Puntanen (2003); this yielded article [4]. Articles [5] and [6] then extend the results given in [4]. The initial version of article [5] was prepared jointly by me and Dr. Puntanen, and similarly, the initial version of article [6] was prepared by Dr. Chu and Prof. Styan. The final versions of articles [5] and [6] were then accomplished in collaboration between all authors.

Article [7] is related to article [4], and is a result of straightforward collaboration between me, Dr. Puntanen, and Prof. Styan. In case of article [8], I proposed the considered research problem to the co-author Dr. Puntanen. The process of writing the manuscript was then joint effort from me and Dr. Puntanen. Similar pattern holds also for article [9]; I proposed the theme of the article, and then the manuscript was prepared jointly by me, Dr. Puntanen, and Prof. Styan.

Article [10] is collaboration between me, Dr. Märt Möls (University of Tartu, Estonia), and Dr. Puntanen. The article was initiated during the research visit of Dr. Möls to University Tampere at January 2006. During the visit, I introduced an open problem to Dr. Möls concerning the linear mixed effects model. Initial results for the problem were then obtained by Dr. Möls which were then further extended and generalized jointly by me and Dr. Puntanen. The final manuscript was prepared jointly by the authors of the article.

Last article [11] is based on the results encountered and considered by Dr. Puntanen and Prof. Styan while studying linear statistical models over the years. My contributions to the article appear specially in sections 3.3, 3.4, and 3.7, where the main results of the article are applied to the partitioned linear model.

Tampere, May 2007 *Jarkko Isotalo*

References

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

Baksalary, J. K. & Kala, R. (1981). Linear transformations preserving best linear unbiased estimators in a general Gauss–Markoff model. *The Annals of Statistics*, 4, 913–916.

Baksalary, J. K. & Kala, R. (1986). Linear sufficiency with respect to a given vector of parametric functions. *Journal of Statistical Planning and Inference*, 14, 331–338.

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

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

Barnard, G. A. (1963). The logic of least squares. *Journal of the Royal Statistical Society, Series B (Methodological)*, 25, 124–127.

Bhimasankaram, P. & Sengupta, D. (1996). The linear zero functions approach to linear models. *Sankhyā, Series B*, 58, 338–351.

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

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: The Indian Journal of Statistics*, 66, 634–651.

Chu, K. L. & Styan, G. P. H. (2003). Some remarks on the efficiency of OLS in simple linear regression, in particular when the OLS and GLS lines are parallel. *McGill University, Dept. of Mathematics and Statistics*, Report 2003-04.

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

Drygas, H. (1983). Sufficiency and completeness in the general Gauss–Markov model. *Sankhyā, Series A*, 45, 88–98.

Frisch, R. & Waugh, F. V. (1933). Partial time regressions as compared with individual trends. *Econometrica*, 1, 387–401.

Goldberger, A. S. (1962). Best linear unbiased prediction in the generalized linear regression model. *Journal of American Statistical Association*, 58, 369–375.

Groß, J. (1998). A note on the concepts of linear and quadratic sufficiency. *Journal of Statistical Planning and Inference*, 70, 69–76.

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

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

Gustafson, K. (2002). Operator trigonometry of statistics and econometrics. *Linear Algebra and Its Applications*, 354, 141–158.

Gustafson, K. (2006). The trigonometry of matrix statistics. *International Statisistical Review*, 74, 187–202.

Isotalo, J. & Puntanen, S. (2003). On the efficiency of OLS in a partitioned linear model. *University of Tampere, Dept. of Mathematics, Statistics & Philosophy*, Report A 343, 43 pp.

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

Lovell, M. C. (1963). Seasonal adjustment of economic time series and multiple regression analysis. *Journal of the American Statistical Association*, 58, 993–1010.

Müller, J. (1987). Sufficiency and completeness in the linear model. *Journal of Multivariate Analysis*, 21, 312–323.

Müller, J., Rao, C. R. & Sinha, B. K. (1984). Inference on parameters in a linear model: a review of recent results. In *Experimental Design, Statistical Models, and Genetic Statistics*, ed. K. Hinkelmann, Statistics 50, Dekker, New York, 277–295.

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.

Peixoto, J. L. (1993). Four equivalent definitions of reparametrizations and restrictions in linear model. *Communications in Statistics: Theory and Methods*, 22, 283–299.

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.

Puntanen, S., Styan, G. P. H. & Werner, H. J. (2000). Two matrix-based proofs that the linear estimator Gy is the best linear unbiased estimator. *Journal of Statistical Planning and Inference*, 88, 173–179.

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.

Searle, S. R., Casella, G. & McCulloch, C. E. (1992). *Variance Components*. Wiley.

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

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

8 Original articles

[1] Linear sufficiency and completeness in the partitioned linear model

[2] Linear sufficiency and completeness in the context of estimating the parametric function in the general Gauss–Markov model

[3] Linear prediction sufficiency for new observations in the general Gauss-Markov model

[4] On decomposing the Watson efficiency of ordinary least squares in a partitioned weakly singular linear model

[5] Some further results concerning the decomposition of the Watson efficiency in partitioned linear models

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

[7] Effect of adding regressors on the equality of the OLSE and BLUE

[8] A note on the equality of the OLSE and the BLUE of the parametric function in the general Gauss–Markov model

[9] Some comments on the Watson efficiency of the ordinary least squares estimator under the Gauss–Markov model

[10] Invariance of the BLUE under the linear fixed and mixed effects models

[11] A useful matrix decomposition and its statistical applications in linear regression