



# Estimating variances in time series kriging using convex optimization and empirical BLUPs

Martina Hančová<sup>1</sup> · Andrej Gajdoš<sup>1</sup> · Jozef Hanč<sup>2</sup> · Gabriela Vozáriková<sup>3</sup>

Received: 22 May 2019 / Revised: 17 December 2019 / Published online: 22 February 2020  
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

## Abstract

We revisit and update estimating variances, fundamental quantities in a time series forecasting approach called kriging, in time series models known as FDSLRLMs, whose observations can be described by a linear mixed model (LMM). As a result of applying the convex optimization, we resolved two open problems in FDSLRLM research: (1) theoretical existence and equivalence between two standard estimation methods—least squares estimators, non-negative (M)DOOLSE, and maximum likelihood estimators, (RE)MLE, (2) and a practical lack of free available computational implementation for FDSLRLM. As for computing (RE)MLE in the case of  $n$  observed time series values, we also discovered a new algorithm of order  $\mathcal{O}(n)$ , which at the default precision is  $10^7$  times more accurate and  $n^2$  times faster than the best current Python(or R)-based computational packages, namely CVXPY, CVXR, nlme, sommer and mixed. The LMM framework led us to the proposal of a two-stage estimation method of variance components based on the empirical (plug-in) best linear unbiased predictions of unobservable random components in FDSLRLM. The method, providing non-negative invariant estimators with a simple explicit analytic form and performance comparable with (RE)MLE in the Gaussian case, can be used for any absolutely continuous probability distribution of time series data. We illustrate our results via applications and simulations on three real data sets (electricity consumption, tourism and cyber security), which are easily available, reproducible, sharable and modifiable in the form of interactive Jupyter notebooks.

**Keywords** Finite discrete spectrum linear regression model · Linear mixed model · Least squares estimators · Maximum likelihood estimators · Efficient computational algorithms · Jupyter

**Mathematics Subject Classification** 62M10 · 62J12 · 91B84 · 90C25

---

This work was supported by the Slovak Research and Development Agency under the contract No. APVV-17-0568, the Scientific Grant Agency of the Slovak Republic (VEGA), VEGA Grant No.1/0311/18 and the Internal Research Grant System of Faculty of Science, P. J. Šafárik University in Košice (VVGS PF UPJŠ)—project VVGS-PF-2018-792.

Extended author information available on the last page of the article

## 1 Introduction

The need to obtain sufficiently accurate predictions for facilitating and improving decision making becomes an integral part not only of science, industry or economy but also in many other human activities. The key idea of the so-called *kriging* as a time series prediction approach consists in modeling the given time series data in an appropriate general class of linear regression models (LRMs) and subsequently finding *the best linear unbiased predictor*—the BLUP which minimizes the mean squared error (MSE) of prediction among all linear unbiased predictors, see e.g. Štulajter (2002), Kedem and Fokianos (2005), Christensen (2011) and Brockwell and Davis (2016).

In the frame of kriging, we investigate theoretical features and econometric applications of a class of time series models called *finite discrete spectrum linear regression models* or shortly FDSLRLMs. The FDSLRLM class was introduced in 2002–2003 by Štulajter (Štulajter 2002, 2003) as a direct extension of classical (ordinary) linear regression models (see e.g. Christensen 2011; Hyndman and Athanasopoulos 2018).

The FDSLRLM has mean values (trend) given by linear regression and random components (error terms) are represented as a sum of a linear combination of uncorrelated zero-mean random variables and white noise which together can be interpreted in terms of the finite discrete spectrum (Priestley 2004).

Formally the FDSLRLM can be presented as

$$X(t) = \sum_{i=1}^k \beta_i f_i(t) + \sum_{j=1}^l Y_j v_j(t) + Z(t); t \in \mathbb{T}, \quad (1.1)$$

where<sup>1</sup>

$\mathbb{T}$  representing the time domain is a countable subset of the real line  $\mathbb{R}$ ,

$k$  and  $l$  are some fixed non-negative integers, i.e.  $k, l \in \mathbb{N}_0$ ,

$\beta = (\beta_1, \beta_2, \dots, \beta_k)' \in \mathbb{R}^k$  is a vector of regression parameters,

$\mathbf{Y} = (Y_1, Y_2, \dots, Y_l)'$  is an unobservable  $l \times 1$  random vector with zero mean vector

$E\{\mathbf{Y}\} = \mathbf{0}_l$ , and with  $l \times l$  diagonal covariance matrix  $Cov\{\mathbf{Y}\} = \text{diag}\{\sigma_j^2\}$ , where

$\sigma_j^2 \in \mathbb{R}_+$  are non-negative real numbers,

$f_i(\cdot); i = 1, 2, \dots, k$  and  $v_j(\cdot); j = 1, 2, \dots, l$  are real functions defined on  $\mathbb{R}$ ,

$Z(\cdot)$  stands for white noise uncorrelated with  $\mathbf{Y}$  and having a positive dispersion

$D\{Z(t)\} = \sigma_0^2 \in \mathbb{R}_{++}$ .

Typically, due to the nature of time series data collection, the most frequently considered time domain  $\mathbb{T}$  is the set of natural numbers  $\mathbb{N} = \{1, 2, \dots\}$ . FDSLRLM variance parameters, which are fundamental quantities in kriging, are commonly described by one vector  $\mathbf{v} \equiv (v_0, v_1, \dots, v_l)' = (\sigma_0^2, \sigma_1^2, \dots, \sigma_l^2)'$ , an element of the parametric space  $\Upsilon = (0, \infty) \times [0, \infty)^l$  or  $\Upsilon = \mathbb{R}_{++} \times \mathbb{R}_+^l$  for short.

<sup>1</sup> As it is common practice, we use subindices as an indication of size for identity matrix  $\mathbf{I}_n$ , zero matrix  $\mathbf{0}_{l \times n}$  or zero vector  $\mathbf{0}_l$ . Dimension of other matrices will be indicated directly in the text or by a matrix space in which they belong.

Our recent article (Gajdoš et al. 2017) summarizes the FDSLRLM kriging procedure, guiding methodology and corresponding references dealing with FDSLRLM kriging for time series econometric forecasting as one of the advanced alternative approaches to the most popular Box-Jenkins methodology (Box et al. 2015; Shumway and Stoffer 2017).

The principal goal of this paper is to revisit and update theory, computational and implementation aspects of estimation procedures for variances in FDSLRLM fitting (Štulajter 2002; Štulajter and Witkovský 2004; Hančová 2008), the first part of FDSLRLM kriging, in the light of recent advances in closely related linear mixed modeling (LMM), convex optimization and modern computational technology. As a result of such investigation, we would also like to present relations of existing FDSLRLM estimators with the well-established standard variance estimators in the field of linear mixed models. Moreover, the alternative views via convex optimization and the theory of empirical BLUPs lead us to new theoretical results, computational algorithms and tools in estimating FDSLRLM variances.

The structure of the paper is organized as follows. Since the time series analysis and prediction based on FDSLRLM is still a relatively new approach and therefore not broadly known, brief review Sect. 2 recapitulates key basic features and benefits of FDSLRLM together with important notes about corresponding theoretical methods and computational tools offered by closely connected mathematical branches and computational research. This background became the strong basis for our investigation focusing especially on non-negative versions of the estimators.

Section 3 is devoted to standard estimation methods of variances  $\mathbf{v}$  in FDSLRLM leading to non-negative estimates of  $\mathbf{v}$  in the frame of convex optimization. Since the methods are based on maximum likelihood or least squares (Gajdoš et al. 2017), we also demonstrate their relations to the well-known maximum likelihood estimators in LMM—MLE, residual MLE (REMLE) requiring specification of the time series distribution, and noniterative distribution-free quadratic estimators—method of moments (MM), minimum norm quadratic estimation (MINQE) and variance least squares (VLS) estimation (Demidenko 2013).

Section 4 describes new alternative estimators for FDSLRLM variances whose definition becomes a consequence of applying the general theory of the BLUP in LMM. Our approach also represents an improved two-stage modification of the previously developed method of natural estimators (NE) (Hančová 2008) which is now based on the idea of empirical (plug-in) BLUPs. Therefore, we will refer to our new method as EBLUP-NE for short. The section also develops the computational form and some basic statistical properties of EBLUP-NE using two special matrices known in linear algebra as the Schur complement and Gram matrix (Zhang 2005). The method is distribution-free which means that a finite FDSLRLM observation  $\mathbf{X} = (X(1), \dots, X(n))'$ ,  $n \in \mathbb{N}$  must have only an absolutely continuous probability distribution with respect to some  $\sigma$ -finite measure.

In the following fifth section containing numerical simulations, we illustrate theoretical results of the paper and the performance of EBLUP-NE on three real time series data sets—*electricity consumption*, *tourism* and *cyber security*. Final Sect. 6 presents the conclusions of the paper. For the sake of paper readability, we moved some proofs to the Appendix. In the Appendix, we report current computational tools found or

developed by us for FDSLRLM kriging. Finally for a better orientation, the Appendix also contains a list of acronyms and abbreviations used in the paper (Table 6).

Since the paper is aimed at statisticians, analysts, econometricians and data scientists applying time series and forecasting methods, our notation is standard for time series analysis and prediction using linear regression models (Štulajter 2002; Kedem and Fokianos 2005; Brockwell and Davis 2009). Sets like  $\mathbb{R}$ ,  $\mathbb{R}_+$ ,  $\mathbb{R}_{++}$  are labeled as it is usual in convex optimization (Boyd and Vandenberghe 2009).

## 2 Theoretical and computational background for FDSLRLM kriging

Let us recall some important notes about time series FDSLRLM kriging, its connection to other math branches and current computational technology.

Any finite FDSLRLM observation  $\mathbf{X} = (X(1), \dots, X(n))'$ ,  $n \in \mathbb{N}$  satisfies a special type of linear mixed model (LMM) of the form<sup>2</sup>

$$\mathbf{X} = \mathbf{F}\boldsymbol{\beta} + \mathbf{V}\mathbf{Y} + \mathbf{Z}, \quad (2.1)$$

where

$$E\{\mathbf{Z}\} = \mathbf{0}_n, \quad Cov\{\mathbf{Z}\} = \sigma_0^2 \mathbf{I}_n, \quad Cov\{\mathbf{Y}, \mathbf{Z}\} = \mathbf{0}_{l \times n}, \\ Cov\{\mathbf{Y}\} = \text{diag}\{\sigma_j^2\} \equiv \mathbf{D}_v, \quad Cov\{\mathbf{X}\} = \sigma_0^2 \mathbf{I}_n + \mathbf{V}\mathbf{D}_v\mathbf{V}' \equiv \boldsymbol{\Sigma}_v.$$

Matrices  $\mathbf{F} = \{F_{ii}\} = \{f_i(t)\}$ ,  $\mathbf{V} = \{V_{ij}\} = \{v_j(t)\}$  for  $t = 1, 2, \dots, n$ ;  $i = 1, 2, \dots, k$ ;  $j = 1, 2, \dots, l$  are LMM design matrices and random vector  $\mathbf{Z} = (Z(1), Z(2), \dots, Z(n))'$  is a finite  $n$ -dimensional white noise observation. In the language of LMM terminology  $\boldsymbol{\beta}$  represents the  $k \times 1$  vector of fixed effects and  $\mathbf{Y}$  the  $l \times 1$  vector of the random effects (Witkovský 2012; Demidenko 2013).

Thanks to the fundamental property (2.1), we can apply convenient LMM mathematical techniques for FDSLRLM fitting and forecasting (Jiang 2007; Pinheiro and Bates 2009; Christensen 2011; Witkovský 2012; Demidenko 2013; Rao and Molina 2015; Covarrubias-Pazaran 2016; Singer et al. 2017).

Another characteristic feature of LMM (2.1) is that it also belongs to linear models with linear covariance structure or shortly linear covariance models (Demidenko 2013, Sect. 4.3)

$$Cov\{\mathbf{X}\} = \sum_{j=0}^l \nu_j \mathbf{V}_j \quad \text{with } \mathbf{V}_0 = \mathbf{I}_n, \quad \mathbf{V}_j = \mathbf{v}_j \mathbf{v}_j', \quad j = 1, \dots, l \quad (2.2)$$

where vectors  $\mathbf{v}_j$ ,  $j = 1, 2, \dots, l$  are columns of design matrix  $\mathbf{V}$ . According to Wu and Xiao (2012), such type of a linear model was introduced just in time series domain by Anderson (1970). From the perspective of the variance parameter estimation, it was

<sup>2</sup> To easily follow if any matrix (or vector) contains variance parameters, elements of vector  $\mathbf{v}$ , we usually use vector  $\mathbf{v}$  as its subscript.

studied extensively e.g. in Rao and Kleffe (1988) and it still remains the subject of intensive research (Zwiernik et al. 2017).

From a time series modeling point of view, if we consider any time series model  $X(\cdot)$  in the additive form  $X(t) = m(t) + \varepsilon(t); t \in \mathbb{T}$  whose mean value function  $m(\cdot)$  is a real function on  $\mathbb{R}$  expressible by a functional series (e.g. Taylor or Fourier series) and error term  $\varepsilon(\cdot)$  is a mean-zero stationary process, then according to the spectral representation theory of time series (Priestley 2004; Percival and Walden 2009)  $X(\cdot)$  can be approximated arbitrarily closely by FDSLRLM. Therefore, from a practice perspective, FDSLRLMs can be potentially applied in many practical situations.

In econometric applications of FDSLRLM kriging, we have some restrictions given by character of econometric data. There is typically only one realization of time series  $X(\cdot)$  and we do not know the mean-value parameters  $\beta \in \mathbb{R}^k$  nor the variance parameters  $\nu \in \Upsilon$ . From the view of time series analysis, the problem of estimating variances  $\nu$  belongs to the class of problems concerning covariance matrix estimation with one realization (Wu and Xiao 2012).

Simultaneously, in time series econometric analysis using FDSLRLMs (Štulajter 2002; Gajdoš et al. 2017), econometric datasets almost always show some periodic patterns as they are influenced by seasons or regularly repeating events. Therefore the fundamental modeling procedure is based on spectral analysis of time series (Štulajter 2002; Priestley 2004; Brockwell and Davis 2009) using trigonometric functions  $\cos, \sin$  as  $f_i, v_j$  regressors in FDLRLM (1.1).

If we identify the significant Fourier frequencies in time series data by periodogram (Štulajter and Witkovský 2004; Gajdoš et al. 2017), then it restricts the form of design matrices  $\mathbf{F}$  and  $\mathbf{V}$  leading to the orthogonality condition for FDSLRLM (Štulajter and Witkovský 2004)

$$\mathbf{F}'\mathbf{V} = 0 \text{ and } \mathbf{V}'\mathbf{V} = \text{diag} \left\{ \|v_j\|^2 \right\}, \tag{2.3}$$

where  $v_j, j = 1, 2, \dots, l$  is  $j$ -th column of  $\mathbf{V}$  and  $\|\bullet\|^2 = (\bullet)'(\bullet)$  is the square of the standard Euclidean vector norm. Such FDSLRLM, satisfying condition (2.3), is called *orthogonal* (Štulajter and Witkovský 2004). If we consider Gaussian FDSLRLM, then with respect to the model identifiability (Demidenko 2013, Sect. 3.2) the orthogonal model (2.3) is always identifiable. So in econometric practice, we mainly work with the orthogonal and identifiable version of FDSLRLM. Our interactive supplementary materials (Gajdoš et al. 2019c), dealing with real data sets in Sect. 5, demonstrate every step of the iterative econometric FDLRLM building that can be followed in every detail with respect to given restrictions.

Finally, it is worth to mention that from the viewpoint of prediction quality (the size of MSE), any FDSLRLM model with  $k, l > 0$  is always better in applications than corresponding classical LRM with the same regressors (containing  $\alpha_j \in \mathbb{R}$  instead of  $Y_j$  in (1.1)), because MSE of the FDSLRLM is always less or equal and conditions of the same MSE practically never hold. It was directly proved in Hančová (2007) using block matrix algebra and the Löwner partial order relation.

On the other hand, the standard maximum likelihood or least squares estimates of  $\nu$  in FDSLRLM are optimization problems. Our motivation to explore estimating FDSLRLM variances in the frame of convex optimization lays on the fact that its mathe-

mathematical tools and efficient, very reliable computational interior-point methods became important or fundamental tools in many other branches of mathematics (Boyd and Vandenberghe 2009) like the design of experiments, high-dimensional data, machine learning or data mining.

Inspired by essential and recent works on convex optimization (Boyd and Vandenberghe 2009; Bertsekas 2009; Koenker et al. 2014; Cornuéjols et al. 2018; Agrawal et al. 2018), we prove new theoretical relations among existing FDSLRLM estimators in the extended parametric space  $\Upsilon = \mathbb{R}_+^{l+1}$  for the orthogonal version of FDSLRLM. Obtained theoretical results bring us also to the development of new fast and accurate computational algorithms for estimating variances in FDSLRLM and can be effectively used in time series simulations (Monte Carlo and bootstrap).

One of the reasons, why time series FDSLRLM kriging is not widely used, seems to be also the lack of available computer implementation of FDSLRLM approach.<sup>3</sup> Therefore all our computational algorithms, developed functions, packages and results, confronted with standard LMM and convex optimization computational tools, are easily readable, sharable, reproducible, and modifiable, thanks to open-source Jupyter technology (Kluyver et al. 2016). In particular, we present our results in the form of Jupyter notebooks, dynamic HTML documents integrating prose, code (Python or R language) and results similarly as Mathematica notebooks, stored in our free available GitHub (Lima et al. 2014) repositories devoted to our FDSLRLM research (Gajdoš et al. 2019a). This easily reproducible way and presentation of our computing with real data were inspired by works (Brielle et al. 2019; Weiss 2017). More details dealing with the computer implementation and tools are described in the Appendix A.3.

### 3 Estimating FDSLRLM variances via convex optimization

#### 3.1 Mathematical approaches in estimating FDSLRLM variances

In the LMM framework (Searle et al. 2009; Demidenko 2013), generally in variance estimation problems, the basic nonlinear optimization theory with the standard matrix calculus or advanced matrix (tensor) calculus<sup>4</sup> are applied. Recently, however, we can also see the use of mathematical apparatus of convex optimization in the aforementioned work, Zwiernik et al. (2017), dealing with MLE in linear covariance models.

In time series analysis and forecasting, one of the traditional approaches in finding least squares or maximum likelihood estimators of unknown parameters in linear regression models, initiated by Kolmogorov in 1941 and presented in the well-known reference books (Brockwell and Davis 1991, 2009; Pourahmadi 2001; Galántai 2004; Christensen 2011), is represented by geometrical projection methods in the frame of

<sup>3</sup> More than ten years ago it was simply assumed that a potential user of FDSLRLM will use one of the well-known commercial environments for scientific computing—MATLAB or Mathematica where he will directly apply formulas from the corresponding theory. But now analyses (see e.g. Frederickson (2019, Sect. Scientific Languages)) show that the situation is dramatically changing, favoring open, free technology like Jupyter notebooks or R.

<sup>4</sup> The matrix calculus using the Kronecker product, vec and vech operators or commutation and duplication matrices.

**Table 1** Estimation methods with references in FDSLRLM

Least squares	Maximum likelihood
Double ordinary least squares estimators (DOOLSE) (Štulajter 2002; Štulajter and Witkovský 2004)	Maximum likelihood estimators (MLE) (Štulajter 2002; Štulajter and Witkovský 2004)
Modified (unbiased) DOOLSE (MDOOLSE) (Štulajter and Witkovský 2004)	Restricted (residual) MLE (REMLE) (Štulajter and Witkovský 2004)
Natural estimators (NE) (Hančová 2008)	

Hilbert (or inner product) spaces. Therefore the theory and computational aspects of the estimation procedures for  $\mathbf{v}$  in FDSLRLM fitting, have been developed using orthogonal and oblique projections (Štulajter 2002; Štulajter and Witkovský 2004; Hančová 2007).

In this section we investigate, revisit and update five variance estimation methods under different assumptions on the structure and distribution of FDSLRLM observation  $\mathbf{X}$ , summarized in Table 1, in the light of convex optimization, LMM and modern computational technology.

According to Agrawal et al. (2018) and Boyd and Vandenberghe (2009), to formulate any mathematical problem as an convex optimization problem, we need to identify three attributes of the problem: optimization variable, convex constraints that the variable must satisfy and the convex objective function depending on the variable whose optimal value we want to achieve.

In all our estimation methods, optimization variable is  $\mathbf{v} \in \mathbb{R}^{l+1}$  satisfying constraints  $\Upsilon = \mathbb{R}_{++} \times \mathbb{R}_+^l$ . In order to avoid any incompatibility problems in the convex optimization theoretical or computational framework, we extend constraints  $\Upsilon$  into the form of standard non-negativity constraints  $\Upsilon^* = \mathbb{R}_+^{l+1} = [0, \infty]^{l+1}$  or  $\mathbf{v} \geq 0$  using generalized inequality.

Finally, we also explain relations of the FDSLRLM estimators with the standard estimators in the LMM framework, where the problem of estimating variances has a long and rich history with many essential reference works (e.g. Rao and Kleffe 1988; Searle et al. 2009; Christensen 2011; Demidenko 2013 and Rao and Molina 2015, see also a review paper Witkovský 2012).

### 3.2 General case of the orthogonal FDSLRLM

First of all, we formulate the basic assumptions required in our investigations on the structure and distribution of FDSLRLM. As we mentioned in review Sect. 2, in econometric FDSLRLM analysis we mainly work with orthogonal FDSLRLMs (2.3). Unless otherwise stated, we will assume hereafter this type of FDSLRLMs.

The second assumption deals with a distribution of the FDSLRLM observation  $\mathbf{X} = (X(1), \dots, X(n))'$ ,  $n \in \mathbb{N}$ . We will assume  $\mathbf{X}$  (2.1), having any absolutely continuous probability distribution with respect to some  $\sigma$ -finite measure.

Under these two assumptions, we can consider three noniterative distribution-free quadratic estimation methods: *natural estimators* (NE), *double ordinary least squares estimators* (DOOLSE) and *modified (unbiased) double ordinary least squares estimators* (MDOOLSE).

### 3.2.1 Natural estimators—NE

In the paper Hančová (2008), we proposed the method of almost surely positive invariant quadratic estimators of FDSLRLM variances  $\mathbf{v}$ , called natural estimators (NE). The main idea behind NE came from the fact that  $\sigma_j^2 = Cov\{Y_j\} = E\{Y_j^2\}; j = 1, \dots, l$ . Therefore, if random vector  $\mathbf{Y}$  in (1.1) was known, the natural estimate of  $\sigma_j^2$  would be just  $Y_j^2$ . This initial consideration was identical with Rao’s estimates known as MINQE (Rao and Kleffe (1988, Sect. 5.1) and Christensen (2011, Sect. 12.7)).

To get sufficiently simple, explicit analytic expressions available for further theoretical study, we predicted the random vector  $\mathbf{Y}$  as fixed by the ordinary least squares method leading to the following form of linear predictor of  $\mathbf{Y}$  based on  $\mathbf{X}$

$$\check{\mathbf{Y}} = \mathbf{W}^{-1}\mathbf{V}'\mathbf{M}_F\mathbf{X}; \mathbf{W} = \mathbf{V}'\mathbf{M}_F\mathbf{V} \in \mathbb{R}^{l \times l}. \tag{3.1}$$

Matrix  $\mathbf{M}_F = (\mathbf{I}_n - \mathbf{F}(\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}') \in \mathbb{R}^{n \times n}$  represents (Hančová 2008) the orthogonal projector onto the orthogonal complement of the column space of  $\mathbf{F}$  and matrix<sup>5</sup>  $\mathbf{W} > 0$  is known in linear algebra and statistics (Zhang 2005, chap. 6) as the Schur complement of  $\mathbf{F}'\mathbf{F}$  in the block matrix  $\begin{pmatrix} \mathbf{F}'\mathbf{F} & \mathbf{F}'\mathbf{V} \\ \mathbf{V}'\mathbf{F} & \mathbf{V}'\mathbf{V} \end{pmatrix}$ .

The form (3.1) exists only in the case of FDSLRLM with the full column rank  $k + l$  of matrix  $(\mathbf{F} \ \mathbf{V})$ . Such FDSLRLM, not necessarily orthogonal, is called *full-rank* FDSLRLM. However, NE exist and can be calculated in FDSLRLMs without full-rank assumption. Moreover, if we assume extra condition  $n > k + l$  together with full-rankness, then from the viewpoint of LMM identifiability FDSLRLM (2.1) under the assumption of normal distribution becomes identifiable (Demidenko 2013, Sect. 3.2). As for  $\sigma_0^2$ , NE is simply the sum of residuals in the OLS method applied to (2.1) assuming both  $\boldsymbol{\beta}$  and  $\mathbf{Y}$  as fixed divided by degrees of freedom. Unlike NE for  $\sigma_j^2, j = 1, \dots, l$ , such estimator is not only almost surely positive, but also unbiased.

Employing results of Hančová (2008), orthogonality condition (2.3), also giving  $\mathbf{M}_F\mathbf{V} = \mathbf{V}$  in (3.1), leads to the following analytic form of NE

$$\check{\mathbf{v}}(\mathbf{e}) = \begin{pmatrix} \frac{1}{n-k-l} \mathbf{e}'\mathbf{M}_V\mathbf{e} \\ (\mathbf{e}'\mathbf{v}_1)^2 / \|\mathbf{v}_1\|^4 \\ (\mathbf{e}'\mathbf{v}_2)^2 / \|\mathbf{v}_2\|^4 \\ \vdots \\ (\mathbf{e}'\mathbf{v}_l)^2 / \|\mathbf{v}_l\|^4 \end{pmatrix} = \begin{pmatrix} \frac{1}{n-k-l} \left( \mathbf{e}'\mathbf{e} - \sum_{j=1}^l (\mathbf{e}'\mathbf{v}_j)^2 / \|\mathbf{v}_j\|^2 \right) \\ (\mathbf{e}'\mathbf{v}_1)^2 / \|\mathbf{v}_1\|^4 \\ (\mathbf{e}'\mathbf{v}_2)^2 / \|\mathbf{v}_2\|^4 \\ \vdots \\ (\mathbf{e}'\mathbf{v}_l)^2 / \|\mathbf{v}_l\|^4 \end{pmatrix}, \tag{3.2}$$

<sup>5</sup> As it is usual, for symmetric matrices  $> 0$  will mean positive definiteness.



where  $\mathbf{e} = \mathbf{x} - \mathbf{F}\boldsymbol{\beta}^* = \mathbf{M}_F\mathbf{x}$  ( $\boldsymbol{\beta}^* = (\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'\mathbf{x}$ ) is nothing else than the vector of ordinary least squares (OLS) residuals in FDSLRLM,  $\mathbf{x}$  is an arbitrary realization of the FDSLRLM observation  $\mathbf{X}$  and vectors  $\mathbf{v}_j, j = 1, 2, \dots, l$  are columns of design matrix  $\mathbf{V}$ . Matrix  $\mathbf{M}_V = \mathbf{I}_n - \mathbf{V}(\mathbf{V}'\mathbf{V})^{-1}\mathbf{V}'$  represents again the orthogonal projector onto the orthogonal complement of the column space of  $\mathbf{V}$ .

As for computational complexity of NE, using elementary theory of complexity (Boyd and Vandenberghe 2018), we get the complexity of  $2kn$  operations for the residual vector  $\mathbf{e}$  and  $4ln$  for NE (3.2). Therefore, computing NE (3.2) represents an algorithm with the complexity having order  $\mathcal{O}(n)$  with respect to the realization length or the number of time series data  $n$ .

Since NE employ the least-squares method, it is easy to show that in orthogonal FDSLRLMs, NE can be obtained as a unique, always existing, non-negative solution of the following convex optimization problem (Agrawal et al. 2018; Boyd and Vandenberghe 2009)

**NE**

$$\begin{aligned} & \text{minimize} \quad f_0(\boldsymbol{\nu}) = \|\mathbf{e}\mathbf{e}' - \sum_{j=1}^l \nu_j \mathbf{V}_j\|^2 + \|\mathbf{M}_V \mathbf{e} \mathbf{e}' \mathbf{M}_V - \nu_0 \mathbf{V}_0\|^2 \quad (3.3) \\ & \text{subject to} \quad \boldsymbol{\nu} = (\nu_0, \dots, \nu_l)' \in [0, \infty)^{l+1} \end{aligned}$$

where matrices in the objective function  $f_0(\boldsymbol{\nu})$  are defined by expressions  $\mathbf{V}_0 = \mathbf{M}_F \mathbf{M}_V$  and  $\mathbf{V}_j = \mathbf{v}_j \mathbf{v}_j', j = 1, \dots, l$ . Now for any symmetric matrix  $\mathbf{A}$

$$\|\mathbf{A}\| = \sqrt{\text{tr}(\mathbf{A}^2)} = \sqrt{\sum_{i,j} A_{ij}^2}$$

means the standard Euclidean matrix norm of  $\mathbf{A}$ .

In convex optimization, this kind of optimization problems belongs to the convex quadratic optimization problems or the norm approximation problems (Boyd and Vandenberghe 2009, Sects. 4.4, 6.1). Although we know the analytic form of NE, from the computational point of view, the representation as the convex optimization problem can be useful for the purposes of software cross-checking results and evaluating numerical precision of general convex optimization algorithms.

**Remark 1** NE vs. MM

From LMM perspective, NE method used in orthogonal FDSLRLMs can be viewed as a variation of the Henderson method 3 (method of fitting constants), whose generalization is the method of moments (MM) providing unbiased estimators (Christensen (2011, Sect. 12.9), Demidenko (2013, Sect. 3.11), Rao and Molina (2015, Sect. 5.1)).

If we applied the idea behind MM estimator in orthogonal FDSLRLMs, we would also calculate residuals  $\mathbf{e} = \mathbf{x} - \mathbf{F}\boldsymbol{\beta}^* = \mathbf{M}_F\mathbf{x}$  as in (3.2). Then we would orthogonally project  $\mathbf{e}$  onto column space of  $\mathbf{V}$  by OLS leading to empirical covariance matrix  $\check{\mathbf{Y}}\check{\mathbf{Y}}'$  with vector  $\check{\mathbf{Y}} = (\mathbf{V}'\mathbf{V})^{-1}\mathbf{V}'\mathbf{X}$ , the same as  $\check{\mathbf{Y}}$  (3.1) for orthogonal FDSLRLM ( $\mathbf{V}'\mathbf{M}_F = \mathbf{V}'$ ), from which we would finally removed its bias.

Comparing NE and MM, in orthogonal FDSLRLMs, gives us that NE for  $\nu_0$  is identical with the MM estimator and NE for each  $\nu_j, j = 1, \dots, l$  is the first non-negative term of MM estimator for  $\nu_j$  (or the term before removing bias). In other words, if we removed a bias from NE, we got MM estimator. But such unbiased NE, equivalent to MM estimates, generally do not hold non-negative constraints  $\mathbf{v} \geq 0$  required by convex optimization.

### 3.2.2 Double least squares estimators—DOOLSE, MDOOLSE

From the optimization viewpoint, assuming full-rank, not necessarily orthogonal FDSLRLM observation  $X$  (2.1), DOOLSE and MDOOLSE can be viewed as the following optimization problems for  $\mathbf{v}$  at given OLS residuals  $\mathbf{e}$

#### DOOLSE

$$\begin{aligned} & \text{minimize} && f_0(\mathbf{v}) = \|\mathbf{e}\mathbf{e}' - \boldsymbol{\Sigma}_{\mathbf{v}}\|^2 \\ & \text{subject to} && \mathbf{v} = (\nu_0, \dots, \nu_l)' \in [0, \infty)^{l+1} \end{aligned} \quad (3.4)$$

#### MDOOLSE

$$\begin{aligned} & \text{minimize} && f_0(\mathbf{v}) = \|\mathbf{e}\mathbf{e}' - \mathbf{M}_F \boldsymbol{\Sigma}_{\mathbf{v}} \mathbf{M}_F\|^2 \\ & \text{subject to} && \mathbf{v} = (\nu_0, \dots, \nu_l)' \in [0, \infty)^{l+1} \end{aligned} \quad (3.5)$$

where  $\mathbf{e}$  are again OLS residuals as in the case of NE and  $\boldsymbol{\Sigma}_{\mathbf{v}} = \sigma_0^2 \mathbf{I}_n + \mathbf{V}\mathbf{D}_{\mathbf{v}}\mathbf{V}'$  is the covariance matrix of  $X$ . We call empirical covariance matrix  $\mathbf{S}_{\mathbf{e}} \equiv \mathbf{e}\mathbf{e}'$  *matrix of residual products* or more compactly *the residual products matrix*. Using elementary properties of mean value, we can derive that  $E\{\mathbf{S}_{\mathbf{e}}\} = \mathbf{M}_F \boldsymbol{\Sigma}_{\mathbf{v}} \mathbf{M}_F$ .

#### *Remark 2 DOOLSE, MDOOLSE vs. VLS, MINQE*

In the LMM framework, DOOLSE and MDOOLSE without non-negativity constraints are identical with variance least squares (VLS) and unbiased VLS (UVLS) estimation methods (Demidenko 2013, Sect. 3.12) applied to FDSLRLM observation  $X$  described by the LMM model (2.1). The idea behind VLS is known more than 40 years (Amemiya 1977). The DOOLSE method, its name and acronym DOOLSE were independently reinvented and formulated in time series analysis by Štulajter (1991, 2002). Since the DOOLSE method consists in applying two OLS regressions, the first one producing residuals  $\mathbf{e}$  and the second one variance estimates, we consider name DOOLSE, even not well-known in statistical community, as more descriptive and better representing the idea of the method as name VLS.

In addition, using results of (Witkovský 2012), we can also prove by direct computation that in the orthogonal FDSLRLM invariant MINQE or shortly MINQE(I) are identical with DOOLSE and invariant unbiased MINQE, MINQE(U,I), are identical to MDOOLSE.

In the next text, some theoretical results can be written in one form for both of DOOLSE and MDOOLSE problems. To emphasize this fact, we will use one abbreviation with parenthesis ( ), e.g (M)DOOLSE, assuming that given results or considerations hold for both problems.

Using the geometrical language of Hilbert spaces (Brockwell and Davis 1991, 2009; Štulajter 2002), (M)DOOLSE determine an unknown variances  $\mathbf{v}$  stored in matrices  $\Sigma_{\mathbf{v}}$  or in  $\mathbf{M}_F \Sigma_{\mathbf{v}} \mathbf{M}_F$  in such way, that these matrices have the smallest Euclidean distance to the observable matrix  $\mathbf{S}_e$ . In such case, (M)DOOLSE for orthogonal FDSLRLMs could be computed geometrically as the orthogonal projection of  $\mathbf{S}_e$  onto the linear span  $\mathcal{L}(\mathbf{V}_0, \mathbf{V}_1, \dots, \mathbf{V}_l)$  using the Gram matrix  $\mathbf{G}$  of set  $\{\mathbf{V}_0, \mathbf{V}_1, \dots, \mathbf{V}_l\}$  generated by the inner product  $(\bullet, \bullet) = \text{tr}(\bullet \cdot \bullet)$ . Matrices  $\mathbf{V}_j, j = 1, \dots, l$  are the same as in (2.2), for DOOLSE:  $\mathbf{V}_0 = \mathbf{I}_n$  and for MDOOLSE:  $\mathbf{V}_0 = \mathbf{M}_F$ .

According to Štulajter and Witkovský (2004), the mentioned projection is given by

$$\tilde{\mathbf{v}}(e) = \mathbf{G}^{-1} \mathbf{q}, \tag{3.6}$$

where

$$\mathbf{G} = \begin{pmatrix} n^* & \|\mathbf{v}_1\|^2 & \|\mathbf{v}_2\|^2 & \dots & \|\mathbf{v}_l\|^2 \\ \|\mathbf{v}_1\|^2 & \|\mathbf{v}_1\|^4 & 0 & \dots & 0 \\ \|\mathbf{v}_2\|^2 & 0 & \|\mathbf{v}_2\|^4 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \|\mathbf{v}_l\|^2 & 0 & 0 & \dots & \|\mathbf{v}_l\|^4 \end{pmatrix} \succ 0, \quad \mathbf{q} = \begin{pmatrix} e'e \\ (e'\mathbf{v}_1)^2 \\ (e'\mathbf{v}_2)^2 \\ \vdots \\ (e'\mathbf{v}_l)^2 \end{pmatrix},$$

where for DOOLSE:  $n^* \equiv n$ , for MDOOLSE:  $n^* \equiv n - k$ ,  $e$  are again OLS residuals,  $\mathbf{v}_j$  are columns of  $\mathbf{V}$ . This formula gives the same results as formula (4.111) in Demidenko (2013) obtained by standard matrix calculus, applied to orthogonal FDSLRLMs in (2.1).

However, the formula (3.6) can produce estimates out of the parametric space  $\Upsilon$  at a relatively high probability (Gajdoš et al. 2017), i.e. in many practical cases of econometric time series analysis it frequently produces negative estimates. The projection-based (or standard matrix calculus) calculation (3.6) is not able to handle non-negative constraints given by  $\Upsilon$  reliably and in some simple way.

To distinguish the projection-based estimates (or VLS estimates from the viewpoint of LMM) from actual DOOLSE (3.4), MDOOLSE (3.5) which are always non-negative, we suggest to be more accurate and use full names for real DOOLSE and MDOOLSE: *non-negative* DOOLSE (NN-DOOLSE) and *non-negative* MDOOLSE (NN-MDOOLSE). Such specific way of notation is not rare and is similarly used e.g. in the case of Rao's MINQE (Rao and Kleffe 1988, chap. 5). For the projection-based DOOLSE and MDOOLSE (3.6), without considering non-negativity constraints, we leave the original acronyms DOOLSE, MDOOLSE coined by Štulajter.

### 3.2.3 Non-negative double least squares estimators – NN-(M)DOOLSE

Now applying the basic convex quadratic optimization theory (Cornuéjols et al. 2018; Bertsekas 2009) in orthogonal FDSLRLMs, we can rewrite NN-(M)DOOLSE as strictly convex quadratic problems, whose solutions  $\tilde{\mathbf{v}}$  always exist and are unique global minimizers on  $\Upsilon^*$ . Here are details proving this conclusion.

Using the standard inner product of matrices defined as  $(\bullet, \bullet) = \text{tr}(\bullet \cdot \bullet)$ , generating the Euclidean norm  $\|\bullet\|$  of a matrix, and basic properties of the trace function, we can

easily rewrite expressions (3.4), (3.5) for the objective functions as quadratic forms. The particular form of these quadratic forms in optimization problems (M)DOOLSE is described by the following proposition.

**Proposition 1** (NN-(M)DOOLSE as quadratic optimization problems)

*In an orthogonal FDSLRLM the NN-(M)DOOLSE problems are convex quadratic optimization problems in the form*

$$\begin{aligned}
 &\text{minimize } f_0(\mathbf{v}) = \mathbf{v}'\mathbf{G}\mathbf{v} - 2\mathbf{q}'\mathbf{v} \\
 &\text{subject to } -\mathbf{I}_{l+1}\mathbf{v} \leq \mathbf{0}_{l+1}
 \end{aligned} \tag{3.7}$$

$$\mathbf{q} = \begin{pmatrix} \mathbf{e}'\mathbf{e} \\ (\mathbf{e}'\mathbf{v}_1)^2 \\ (\mathbf{e}'\mathbf{v}_2)^2 \\ \vdots \\ (\mathbf{e}'\mathbf{v}_l)^2 \end{pmatrix}, \quad \mathbf{G} = \begin{pmatrix} n^* & \|\mathbf{v}_1\|^2 & \|\mathbf{v}_2\|^2 & \dots & \|\mathbf{v}_l\|^2 \\ \|\mathbf{v}_1\|^2 & \|\mathbf{v}_1\|^4 & 0 & \dots & 0 \\ \|\mathbf{v}_2\|^2 & 0 & \|\mathbf{v}_2\|^4 & \dots & 0 \\ \vdots & \vdots & \vdots & \dots & \vdots \\ \|\mathbf{v}_l\|^2 & 0 & 0 & \dots & \|\mathbf{v}_l\|^4 \end{pmatrix} > 0,$$

where for NN-DOOLSE:  $n^* \equiv n$ , NN-MDOOLSE:  $n^* \equiv n - k$ ,

According to the fundamental KKT theorem of convex optimization (Boyd and Vandenberghe 2009, chap. 5) which handles convex optimization problems with constraints in the form of inequalities giving a list of the so-called *Karush–Kuhn–Tucker (KKT) optimality conditions*, we consider our optimization problem NN-(M)DOOLSE in the Lagrange multiplier form with Lagrangian  $L$  as a sum of the objective function  $f_0(\mathbf{v})$  and a linear combination of multipliers  $\lambda = (\lambda_0, \lambda_1, \dots, \lambda_l)' \geq \mathbf{0}_{l+1}$  and constraints  $-\mathbf{v} = (\nu_0, \nu_1, \dots, \nu_l)' \leq \mathbf{0}_{l+1}$ .

Then KKT optimality conditions for (3.7) become necessary and sufficient conditions of optimality, satisfied at any local optimal solution, being represented by three sets of conditions: (1) primal and dual feasibility:  $\mathbf{v} \geq \mathbf{0}_{l+1}$ ,  $\lambda \geq \mathbf{0}_{l+1}$ , (2) stationarity of the Lagrangian:  $\nabla_{\mathbf{v}}L(\mathbf{v}, \lambda) = \mathbf{0}_{l+1}$  and (3) complementary slackness:  $\nu_j\lambda_j = 0$ ,  $j \in \{0, 1, \dots, l\}$ . Computing the gradient of the Lagrangian, we arrive at the following proposition.

**Proposition 2** (KKT optimality conditions for NN-(M)DOOLSE)

*In an orthogonal FDSLRLM consider the non-negative NN-(M)DOOLSE problems in the Lagrange multiplier form with the Lagrangian*

$$L(\mathbf{v}, \lambda) = f_0(\mathbf{v}) - \sum_{j=0}^l \lambda_j \nu_j.$$

*Then, a necessary and sufficient condition for  $\tilde{\mathbf{v}}, \tilde{\lambda}$  to be problems' optimal solution is*

- (1)  $\mathbf{v} \geq \mathbf{0}_{l+1}$ ,  $\lambda \geq \mathbf{0}_{l+1}$ ,
- (2)  $(\mathbf{G} - \mathbf{I}_{l+1}) \begin{pmatrix} \mathbf{v} \\ \lambda \end{pmatrix} = \mathbf{q}$ ,

$$(3) \quad \boldsymbol{\nu} \circ \boldsymbol{\lambda} = \mathbf{0}_{l+1} \quad (\Leftrightarrow \nu_j \lambda_j = 0, \quad j \in \{0, 1, \dots, l\}).$$

As for the existence of optimal solutions, we can apply the well-known basic results for minimizing quadratic forms (Bertsekas (2009, chap. 3)). Since the Hessian of (3.7) equal to  $\nabla_{\boldsymbol{\nu}}^2 f_0(\boldsymbol{\nu}) = 2\mathbf{G}$  is a positive definite matrix, like  $\mathbf{G}$ , then  $f_0(\boldsymbol{\nu})$  is strictly convex and also coercive proper function. These two properties are sufficient conditions for NN-(M)DOOLSE (Bertsekas (2009, Weierstrass’ theorem, p. 119)) to have always a unique global optimal solution  $\tilde{\boldsymbol{\nu}}$ . Employing the familiar Bessel’s inequality (Brockwell and Davis 2009, Sect. 2.4)

$$\sum_{j=1}^l (\mathbf{e}' \mathbf{v}_j / \|\mathbf{v}_j\|)^2 \leq \|\mathbf{e}\|^2, \quad \text{where equality holds if and only if } \mathbf{e} \in \mathcal{L}(\mathbf{V}),$$

it is also not difficult to rewrite KKT conditions (1–3) as the following system

$$n^* \nu_0 + \sum_{j=1}^l \nu_j \|\mathbf{v}_j\|^2 - \lambda_0 = \|\mathbf{e}\|^2, \tag{3.8}$$

$$\nu_0 \|\mathbf{v}_j\|^2 + \nu_j \|\mathbf{v}_j\|^4 - \lambda_j = (\mathbf{e}' \mathbf{v}_j)^2; \quad j \in \{1, \dots, l\}, \tag{3.9}$$

$$\nu_j \lambda_j = 0, \quad \lambda_j \geq 0, \quad \nu_j \geq 0; \quad j \in \{0, 1, \dots, l\} \tag{3.10}$$

implying an optimal solution  $\tilde{\boldsymbol{\nu}}$  with  $\tilde{\nu}_0 = 0$  if and only if the vector of OLS residuals  $\mathbf{e}$  belongs to the column space of  $\mathbf{V}$ . Since probability of  $\mathbf{e} \in \mathcal{L}(\mathbf{V})$  is zero, our existence conclusions can be summarized in the next proposition.

**Proposition 3** (Solution existence of NN-(M)DOOLSE)

*In an orthogonal FDSLRLM the following holds*

- (1) NN-(M)DOOLSE problems are strictly quadratic optimization problems.
- (2) Their solutions  $\tilde{\boldsymbol{\nu}}$  always exist and they are unique global minimizers.
- (3)  $\tilde{\nu}_0 = 0 \Leftrightarrow \mathbf{e} \in \mathcal{L}(\mathbf{V}) \Leftrightarrow \mathbf{x} \in \mathcal{L}(\mathbf{F}, \mathbf{V})$ .
- (4)  $P(\tilde{\nu}_0 = 0) = P(\mathbf{e} \in \mathcal{L}(\mathbf{V})) = P(\mathbf{x} \in \mathcal{L}(\mathbf{F}, \mathbf{V})) = 0$ .

From the practical computational viewpoint, the most important result is fundamental KKT conditions because finding solutions of convex optimization problems is nothing else than solving the corresponding KKT conditions analytically or in general numerically.

In our case of NN-(M)DOOLSE, the KKT conditions (1–3) in Proposition 2 can be represented by a set of  $2^l$  nonsingular linear systems of equations, where each of them is given by a specific matrix  $\mathbf{K}$  derived from  $\mathbf{G}$  and by vector  $\mathbf{g}$  from (3.6). Details of such representation can be found in the Appendix A.2. Then our theoretical results in Proposition 3 dictate that only one of the linear systems always gives us all required non-negative elements of  $\tilde{\boldsymbol{\nu}}$ .

On this representation, we can establish a simple KKT optimization algorithm which run through given linear systems, compute their analytical solution and stops when the solution appears non-negative. In other words, the proposed KKT algorithm

**Table 2** Scheme of the KKT algorithm for NN-(M)DOOLSE  $\tilde{\mathbf{v}}$  in orthogonal FDSLRLM

**Input:** Form the matrix  $\mathbf{G}$ , the vector  $\mathbf{q}$  from vectors  $\mathbf{e}$  and  $\mathbf{v}_j$ ;  $j \in \{1, \dots, l\}$ .

**For** each auxiliary vector  $\mathbf{b} = (b_1, b_2, \dots, b_l)' \in \{0, 1\}^l$  **do**:

1. *Set the KKT-conditions matrix  $\mathbf{K}$ :*

$\mathbf{K} \leftarrow \mathbf{G}$ ;

**For**  $j = 1, 2, \dots, l$  **do**: **If**  $b_j = 0$  **then**  $K_{0j} \leftarrow 0, K_{jj} \leftarrow -1$ .

2. *Calculate the auxiliary vector  $\boldsymbol{\gamma}$ :*

$\boldsymbol{\gamma} \leftarrow \mathbf{K}^{-1}\mathbf{q}$ .

3. *Test non-negativity of  $\boldsymbol{\gamma}$ :*

**If**  $\boldsymbol{\gamma} \geq 0$  **then quit**.

**Output:** Use the last  $\mathbf{b}$ ,  $\boldsymbol{\gamma}$  to form NN-(M)DOOLSE  $\tilde{\mathbf{v}}$  of  $\boldsymbol{\nu}$ :

$\tilde{\mathbf{v}} \leftarrow \boldsymbol{\gamma}$ ;

**For**  $j = 1, 2, \dots, l$  **do**: **If**  $b_j = 0$  **then**  $\tilde{\nu}_j \leftarrow 0$ .

always finds the required NN-(M)DOOLSE  $\tilde{\mathbf{v}}$  in at most  $2^l$  steps. The algorithm is summarized in Table 2, whereas the proof is in the Appendix A.2.

As for computational complexity of the KKT optimization algorithm, all calculations of input have complexity  $\mathcal{O}(n)$ , whereas the body of the algorithm (steps 1–3) needs extra operations of order  $\mathcal{O}(l^2 \cdot 2^l)$ . Since fixed number  $l$  of variance parameters in matrix  $\mathbf{D}_{\boldsymbol{\nu}} = \text{Cov}\{\mathbf{Y}\}$  is usually much smaller than  $n$  (or  $l^2 \cdot 2^l$  is not bigger than  $n$ ), then the complete algorithm has the leading order  $\mathcal{O}(n)$  with respect to  $n$ . It is the same as in the case of NE and generally typical only for computationally fastest analytical solutions of the KKT conditions.

### 3.3 Gaussian orthogonal FDSLRLM

In the LMM framework, the most usual distribution of  $\mathbf{X}$  used in practice is represented by the multivariate normal (Gaussian) distribution. Therefore, in addition to the orthogonality of FDSLRLM, we will require in this section the distributional assumption  $\mathbf{X} \sim \mathcal{N}_n(\mathbf{F}\boldsymbol{\beta}, \boldsymbol{\Sigma}_{\boldsymbol{\nu}})$ . We will refer to FDSLRLM under these assumptions as to *Gaussian orthogonal FDSLRLM*.

In the case of Gaussian orthogonal FDSLRLM, we can add to our investigation last two always non-negative estimation methods from Table 1: *maximum likelihood estimators* (MLE) and *residual maximum likelihood estimators* (REMLE).

#### 3.3.1 Maximum likelihood estimators—MLE, REMLE

Both MLE and REMLE of variances  $\boldsymbol{\nu}$  provide estimates maximizing ML and REML log-likelihood functions (logarithms of likelihoods).

Using simple and clear arguments from Lamotte's paper (LaMotte 2007), we can easily form the ML log-likelihood function  $l_m(\boldsymbol{\nu}, \mathbf{x})$  and REML loglikelihood function

$l_r(\mathbf{v}, \mathbf{x})$ , assuming full-rank FDSLRLM observation  $\mathbf{X}$  (2.1), as

$$l_m(\mathbf{v}, \mathbf{x}) = \frac{1}{2} \left( \ln \det(\Sigma_{\mathbf{v}}^{-1}) - \|\mathbf{x} - \mathbf{F}\beta_{\mathbf{v}}^*\|_{\Sigma_{\mathbf{v}}^{-1}}^2 \right), \tag{3.11}$$

$$l_r(\mathbf{v}, \mathbf{x}) = \frac{1}{2} \left( \ln \det(\Sigma_{\mathbf{v}}^{-1}) - \ln \det(\mathbf{F}'\Sigma_{\mathbf{v}}^{-1}\mathbf{F}) - \|\mathbf{x} - \mathbf{F}\beta_{\mathbf{v}}^*\|_{\Sigma_{\mathbf{v}}^{-1}}^2 \right), \tag{3.12}$$

where  $\mathbf{e}_{\mathbf{v}} = \mathbf{x} - \mathbf{F}\beta_{\mathbf{v}}^*$  are FDSLRLM residuals, known as *marginal residuals* in LMM (Singer et al. 2017),  $\mathbf{x}$  is an arbitrary realization of the FDSLRLM observation  $\mathbf{X}$  with covariance matrix  $\Sigma_{\mathbf{v}}$ ,  $\beta_{\mathbf{v}}^* = (\mathbf{F}'\Sigma_{\mathbf{v}}^{-1}\mathbf{F})^{-1}\mathbf{F}'\Sigma_{\mathbf{v}}^{-1}\mathbf{x}$  is the best linear unbiased estimator (BLUE) of  $\beta$  and  $\|\bullet\|_{\Sigma_{\mathbf{v}}^{-1}}^2 = (\bullet)'\Sigma_{\mathbf{v}}^{-1}(\bullet)$  is the square of a generalized vector norm given by  $\Sigma_{\mathbf{v}}^{-1}$ . Moreover, in orthogonal FDSLRLM, BLUE  $\beta_{\mathbf{v}}^* = (\mathbf{F}'\Sigma_{\mathbf{v}}^{-1}\mathbf{F})^{-1}\mathbf{F}'\Sigma_{\mathbf{v}}^{-1}\mathbf{x}$  and marginal residuals  $\mathbf{e}_{\mathbf{v}}$  are identical with OLS estimate  $\beta^* = (\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'\mathbf{x}$  and OLS residuals  $\mathbf{e}$  respectively, not depending on  $\mathbf{v}$  (Štulajter and Witkovský 2004).

**Remark 3** — REML likelihood

In Lammote’s paper (LaMotte 2007), which presents a direct derivation of REML likelihood function in LMM using familiar linear algebra operations, we can find that the formulation of ML and REML likelihood in LMM under the assumption of normality was originally done in Harville’s works (Harville 1974, 1977).

However, LMM references up to 2007 show the explicit form of REML likelihood for LMM rarely. In Štulajter’s FDSLRLM reference works dealing with MLE (Štulajter 2002; Štulajter and Witkovský 2004) REML likelihood for FDSLRLM is also absent. According to Lammote (LaMotte 2007), the reason *why* seems to be Harville’s too sophisticated, and indirect derivation. As for Demidenko’s LMM monograph (Demidenko 2013), the FDSLRLM formulas (3.11), (3.12) are special cases of ML and REML log-likelihood functions presented in Sect. 2.2.6, where estimator  $\beta_{\mathbf{v}}^*$  is mentioned as the generalized least squares estimator. Assumption about full column rank of  $\mathbf{F}$  is identical with Demidenko’s assumption in Sect. 2.2.6.

Maximum likelihood and residual maximum likelihood estimation of variance components in LMMs produce, in general, no analytical expressions for the estimators (Searle et al. 2009, chap. 6). They exist only for  $\nu_0 > 0$ . According to Demidenko (2013, Sect. 2.5, Theorem 4) MLE and REMLE estimates do not exist in general FDSLRLM with LMM observation  $\mathbf{X}$  (2.1) if and only if a realization  $\mathbf{x}$  of  $\mathbf{X}$  belongs to  $\mathcal{L}(\mathbf{F}, \mathbf{V})$ , linear span of columns  $\mathbf{F}$  and  $\mathbf{V}$ . This condition is also equivalent with  $\mathbf{e} \in \mathcal{L}(\mathbf{V})$ .

In orthogonal FDSLRLM, the MLE and REMLE can be rewritten as the optimization problems for  $\mathbf{v}$  at given OLS residuals  $\mathbf{e}$

**MLE**

$$\begin{aligned} & \text{maximize} && f_0(\mathbf{v}) = \ln \det(\Sigma_{\mathbf{v}}^{-1}) - \|\mathbf{e}\|_{\Sigma_{\mathbf{v}}^{-1}}^2 \\ & \text{subject to} && \mathbf{v} = (\nu_0, \dots, \nu_l)' \in [0, \infty)^{l+1} \end{aligned} \tag{3.13}$$

**REMLE**

$$\begin{aligned} & \text{maximize} && f_0(\mathbf{v}) = \ln \det(\Sigma_{\mathbf{v}}^{-1}) - \ln \det(\mathbf{F}'\Sigma_{\mathbf{v}}^{-1}\mathbf{F}) - \|\mathbf{e}\|_{\Sigma_{\mathbf{v}}^{-1}}^2 \\ & \text{subject to} && \mathbf{v} = (\nu_0, \dots, \nu_l)' \in [0, \infty)^{l+1} \end{aligned} \tag{3.14}$$

where  $\|e\|_{\Sigma_v^{-1}}^2 = e' \Sigma_v^{-1} e = \text{tr}(\mathbf{S}_e \Sigma_v^{-1})$  and  $\mathbf{S}_e$  is the residual products matrix used in definitions (3.4), (3.5) of non-negative (M)DOOLSE.

Generally, (RE)MLE in FDSLRLM are not convex optimization problems with respect to  $v$  and they do not exist for  $v_0 = 0$  which occurs if and only if  $e \in \mathcal{L}(\mathbf{V})$ . If we apply the following bijective transformation defined on  $\Upsilon = (0, \infty) \times [0, \infty)^l$

$$d_0 = \frac{1}{v_0}, \quad d_j = \frac{v_j}{v_0(v_0 + \|v_j\|^2 v_j)}, \quad j \in \{1, \dots, l\}, \tag{3.15}$$

we can convert the (RE)MLE problems into the form of equivalent convex problems (Boyd and Vandenberghe 2009, Sect. 4.1.3) whose solutions can be readily converted into (RE)MLE solutions by the inverse transformation to (3.15). Here are more detailed steps of the conversion.

In orthogonal FDSLRLM  $\Sigma_v^{-1}$  (Štulajter 2003, Lemma 2.1) and  $e' \Sigma_v^{-1} e$  are equal to

$$\begin{aligned} \Sigma_v^{-1} &= \frac{1}{v_0} \mathbf{I}_n - \sum_{j=1}^l \frac{v_j}{v_0(v_0 + \|v_j\|^2 v_j)} v_j v_j' = d_0 \mathbf{I}_n - \mathbf{V} \text{diag} \{d_j\} \mathbf{V}', \\ \|e\|_{\Sigma_v^{-1}}^2 &= e' \Sigma_v^{-1} e = d_0 e' e - e' \mathbf{V} \text{diag} \{d_j\} \mathbf{V}' e. \end{aligned} \tag{3.16}$$

Using orthogonality conditions (2.3) and determinant expression (Searle and Khuri 2017, Sect. 6.8)

$$\det(\lambda \mathbf{I}_n - \mathbf{A} \mathbf{B}) = \lambda^{n-l} \det(\lambda \mathbf{I}_l - \mathbf{B} \mathbf{A}) \text{ for } \lambda \neq 0, \mathbf{A} \in \mathbb{R}^{n \times l}, \mathbf{B} \in \mathbb{R}^{l \times n}, \tag{3.17}$$

we get for determinants in (RE)ML loglikelihoods (3.13), (3.14)

$$\begin{aligned} \det(\mathbf{F}' \Sigma_v^{-1} \mathbf{F}) &= \det(d_0 \mathbf{F}' \mathbf{F}) = d_0^k \det(\mathbf{F}' \mathbf{F}), \\ \det \Sigma_v^{-1} &= d_0^{n-l} \det(d_0 \mathbf{I}_l - \mathbf{V}' \mathbf{V} \text{diag} \{d_j\}) = d_0^{n-l} \prod_{j=1}^l (d_0 - d_j \|v_j\|^2). \end{aligned} \tag{3.18}$$

Since the chosen bijective transformation (3.15) also transforms convex constraints for  $v$  to convex constraints for  $d$ , substituting (3.16), (3.18) into objective functions (3.13), (3.14) of (RE)MLE, we can formulate the following proposition.

**Proposition 4** ((RE)MLE as equivalent convex problems)

Let assume  $e \notin \mathcal{L}(\mathbf{V})$  and consider a bijective transformation in the following form:

$$d_0 = \frac{1}{v_0}, \quad d_j = \frac{v_j}{v_0(v_0 + \|v_j\|^2 v_j)}, \quad j \in \{1, \dots, l\}.$$



Then, in a Gaussian orthogonal FDSLRLM the (RE)MLE problems are equivalent to convex problems:

$$\begin{aligned} &\text{minimize} && f_0(\mathbf{d}) = -(n^* - l) \ln d_0 - \sum_{j=1}^l \ln(d_0 - d_j \|\mathbf{v}_j\|^2) + d_0 \mathbf{e}' \mathbf{e} - \mathbf{e}' \mathbf{V} \text{diag} \{d_j\} \mathbf{V}' \mathbf{e} \\ &\text{subject to} && d_0 > \max\{d_j \|\mathbf{v}_j\|^2, j = 1, \dots, l\} \\ &&& d_j \geq 0, j = 1, \dots, l \end{aligned}$$

where for MLE:  $n^* \equiv n$ , and for REMLE:  $n^* \equiv n - k$ .

Similarly as for NN-(M)DOOLSE, we can write the equivalent (RE)MLE problem with  $2l$  constraints in the corresponding Lagrangian multiplier form based on the following Lagrangian with  $2l$  multipliers  $(\lambda_1, \dots, \lambda_l, \mu_1, \dots, \mu_l)' = (\boldsymbol{\lambda}', \boldsymbol{\mu}')' \in \mathbb{R}_+^{2l}$

$$L(\mathbf{d}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f_0(\mathbf{d}) - \sum_{j=1}^l \left[ \lambda_j d_j + \mu_j (d_0 - d_j \|\mathbf{v}_j\|^2) \right].$$

Then by direct computation, we obtain KKT conditions for equivalent (RE)MLE describing: (1) primal and dual feasibility for  $\mathbf{d}$  and  $(\boldsymbol{\lambda}', \boldsymbol{\mu}')'$ , (2) stationarity of the Lagrangian  $(\nabla_{\mathbf{d}} L(\mathbf{d}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = 0)$  and (3) complementary slackness, all described by the next proposition.

**Proposition 5** (KKT optimality conditions for equivalent (RE)MLE)

Consider equivalent convex optimization problems to (RE)MLE in the Lagrangian multiplier form

$$L(\mathbf{d}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f_0(\mathbf{d}) - \sum_{j=1}^l \left[ \lambda_j d_j + \mu_j (d_0 - d_j \|\mathbf{v}_j\|^2) \right].$$

Then, a **necessary and sufficient condition** for  $\hat{\mathbf{d}}, \hat{\boldsymbol{\lambda}}, \hat{\boldsymbol{\mu}}$  to be problems' optimal solution is

- (1)  $d_0 - d_j \|\mathbf{v}_j\|^2 > 0, d_j \geq 0, \lambda_j \geq 0, \mu_j \geq 0, j \in \{1, \dots, l\},$
- (2)  $\|\mathbf{e}\|^2 - \frac{n^* - l}{d_0} - \sum_{j=1}^l \left( \mu_j + \frac{1}{d_0 - d_j \|\mathbf{v}_j\|^2} \right) = 0,$   
 $\frac{\|\mathbf{v}_j\|^2}{d_0 - d_j \|\mathbf{v}_j\|^2} - (\mathbf{e}' \mathbf{v}_j)^2 - \lambda_j + \mu_j \|\mathbf{v}_j\|^2 = 0, j \in \{1, \dots, l\},$
- (3)  $-d_j \lambda_j = 0, -(d_0 - d_j \|\mathbf{v}_j\|^2) \mu_j = 0, j \in \{1, \dots, l\}.$

Similarly as in the case of NN-(M)-DOOLSE, the Hessian  $\mathbf{H} \equiv \nabla_{\mathbf{d}}^2 f_0(\mathbf{d})$  equal to

$$\begin{pmatrix} \frac{n^*-l}{d_0^2} + \sum_{j=1}^l \frac{1}{(d_j \|\mathbf{v}_j\|^2 - d_0)^2} & -\frac{\|\mathbf{v}_1\|^2}{(d_1 \|\mathbf{v}_1\|^2 - d_0)^2} & -\frac{\|\mathbf{v}_2\|^2}{(d_2 \|\mathbf{v}_2\|^2 - d_0)^2} & \cdots & -\frac{\|\mathbf{v}_l\|^2}{(d_l \|\mathbf{v}_l\|^2 - d_0)^2} \\ -\frac{\|\mathbf{v}_1\|^2}{(d_1 \|\mathbf{v}_1\|^2 - d_0)^2} & \frac{\|\mathbf{v}_1\|^4}{(d_1 \|\mathbf{v}_1\|^2 - d_0)^2} & 0 & \cdots & 0 \\ -\frac{\|\mathbf{v}_2\|^2}{(d_2 \|\mathbf{v}_2\|^2 - d_0)^2} & 0 & \frac{\|\mathbf{v}_2\|^4}{(d_2 \|\mathbf{v}_2\|^2 - d_0)^2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ -\frac{\|\mathbf{v}_l\|^2}{(d_l \|\mathbf{v}_l\|^2 - d_0)^2} & 0 & 0 & \cdots & \frac{\|\mathbf{v}_l\|^4}{(d_l \|\mathbf{v}_l\|^2 - d_0)^2} \end{pmatrix} \succ 0$$

leads to strict convexity of the problem. Since  $f_0(\mathbf{d})$  is also coercive, we can summarize the existence conditions of optimal solutions in the final proposition as we did for NN-(M)DOOLSE problems.

**Proposition 6** (Solution existence of equivalent (RE)MLE)

Let assume  $\mathbf{e} \notin \mathcal{L}(\mathbf{V})$ . Then in a Gaussian orthogonal FDSLRLM the following holds

- (1) Equivalent (RE)MLE problems are strictly convex optimization problems.
- (2) The objective function  $f_0(\mathbf{d})$  is coercive with respect to constraints.
- (3) Their solutions  $\hat{\mathbf{d}}$  always exist and they are unique global minimizers.

The bijection (3.15) implies  $d_0 - d_j \|\mathbf{v}_j\|^2 = (\nu_0 + \nu_j \|\mathbf{v}_j\|^2)^{-1} \neq 0$  which dictates  $\mu_j = 0$  to satisfy the complementary slackness conditions in KKT (3) and simultaneously forms the new version of KKT (1–3) of Proposition 5

$$\|\mathbf{e}\|^2 - (n^* - l)\nu_0 - \sum_{j=1}^l (\nu_0 + \nu_j \|\mathbf{v}_j\|^2) = 0, \tag{3.19}$$

$$\|\mathbf{v}_j\|^2(\nu_0 + \nu_j \|\mathbf{v}_j\|^2) - (\mathbf{e}'\mathbf{v}_j)^2 - \lambda_j = 0, \quad j \in \{1, \dots, l\}, \tag{3.20}$$

$$\nu_0 > 0, \quad -\nu_j \lambda_j = 0, \quad \nu_j \geq 0, \quad \lambda_j \geq 0, \quad j \in \{1, \dots, l\}. \tag{3.21}$$

But for  $\mathbf{e} \notin \mathcal{L}(\mathbf{V})$ , which occurs with probability 1, the system (3.19)–(3.21) is the same as the system (3.8)–(3.10).

In other words we just proved that for  $\nu_0 \neq 0$  (or  $\mathbf{e} \notin \mathcal{L}(\mathbf{V})$ ) the KKT conditions for NN-(M)DOOLSE are equivalent with the KKT conditions for (RE)MLE and the bijective transformation. It means that using the convex optimization approach in orthogonal FDSLRLM, we proved the equivalence of NN-(M)DOOLSE and (RE)MLE with probability 1.

**Theorem 1** (equivalence between NN-(M)DOOLSE and (RE)MLE)

In a Gaussian orthogonal FDSLRLM non-negative (M)DOOLSE are almost surely equal to (RE)MLE, i.e.

$$P \left( \begin{matrix} \text{non-negative} \\ \text{DOOLSE} \end{matrix} = \text{MLE} \right) = P \left( \begin{matrix} \text{non-negative} \\ \text{MDOOLSE} \end{matrix} = \text{REMLE} \right) = 1.$$

The important theoretical result of Theorem 1 is stronger and more general than in Štulajter and Witkovský (2004) proved only for interior points of  $\Upsilon$  and whose iterative computational algorithm could fail in calculating (RE)MLE when (M)DOOLSE, not satisfying non-negativity (equivalent with VLS estimates), are negative.

The very useful consequence of Theorem 1 is that we can practically compute MLE or REMLE in orthogonal FDSLRLMs as quadratic NN-(M)DOOLSE. Moreover, for  $e \in \mathcal{L}(\mathbf{V})$  NN-(M)DOOLSE do not fail as (RE)MLE, but naturally extend them. Therefore, we can use our KKT algorithm of order  $\mathcal{O}(n)$ . However, in computing (RE)MLE we can apply any of the current standard computational tools of convex optimization or LMM (all giving algorithms at least of the order  $\mathcal{O}(n^3)$ ) described in detail in the Appendix A.3.

## 4 Estimating FDSLRLM variances via EBLUPs

### 4.1 Definition and computational form of new estimators

The convex optimization approach brought us new theoretical relations between non-iterative distribution-free quadratic and maximum likelihood estimators of FDSLRLM variances and also a new fast computational algorithm for their computation. As we can see in this section, the LMM framework itself gives us a new improved version of the previously developed method of NE described in Sect. 3.2.

If we look at formula (3.1) for NE through eyes of the general theory of the BLUP in LMM (see e.g. Christensen (2011, chap. 12)), then in FDSLRLM the OLS predictor  $\check{Y}$  for  $Y$  is linear, unbiased, computationally and analytically simple as we required for further theoretical research, but not the best. Therefore, following the standard practice of LMMs in predicting  $Y$ , the random effect in LMM (2.1), we take the best linear unbiased predictor (BLUP) as the estimator of  $Y$ . This consideration leads to the following improved version of our original NE.

Let us consider FDSLRLM (1.1) and its observation  $X$  described by (2.1), not necessarily orthogonal. Then estimators in the form

$$\hat{\sigma}_j^2(X) \equiv (Y_{\mathbf{v}}^*)^2_j; Y_{\mathbf{v}}^* \text{ is the BLUP of } Y \text{ based on } X \tag{4.1}$$

are said to be *natural estimators of  $\nu$  based on the BLUP* or BLUP-NE for short.

The BLUP  $Y_{\mathbf{v}}^*$  of  $Y$  together with the best linear unbiased estimator  $\beta_{\mathbf{v}}^*$  (BLUE) of  $\beta$ , mentioned in (RE)ML log-likelihoods (3.11) and (3.12), can be obtained from celebrated Henderson’s mixed model equations, MME for short [for MME in the current framework of LMM see e.g. Christensen (2011, Sect. 12.3) or Witkovský (2012, Sect. 2)]. MME have the following form in the case of FDSLRLM

$$\begin{pmatrix} \mathbf{F}'\mathbf{F} & \mathbf{F}'\mathbf{V} \\ \mathbf{V}'\mathbf{F} & \mathbf{G}_{\mathbf{v}} \end{pmatrix} \cdot \begin{pmatrix} \beta_{\mathbf{v}}^* \\ Y_{\mathbf{v}}^* \end{pmatrix} = \begin{pmatrix} \mathbf{F}'X \\ \mathbf{V}'X \end{pmatrix}, \tag{4.2}$$

where  $\mathbf{G}_{\mathbf{v}} \equiv \mathbf{G}_{\mathbf{V}} + \sigma_0^2 \mathbf{D}_{\mathbf{v}}^{-1} = \mathbf{V}'\mathbf{V} + \sigma_0^2 \mathbf{D}_{\mathbf{v}}^{-1}$  and  $\mathbf{G}_{\mathbf{V}} \equiv \mathbf{V}'\mathbf{V} \in \mathbb{R}^{l \times l}$  is the Gram matrix (Boyd and Vandenberghe 2018, Sect. 10.1) for columns of matrix  $\mathbf{V}$ .

Assuming the full-ranked FDSLRLM with  $r(\mathbf{F} \mathbf{V}) = k + l$  as in the case of NE and using the so-called *Banachiewicz inversion formula* for the inverse of a  $2 \times 2$  partitioned (block) matrix (see e.g. Zhang 2005, Sect. 6.0.2) or (Hančová 2008, Sect. 2.1), we can easily prove the existence and the following form of the inverse to the block matrix in (4.2)

$$\begin{pmatrix} \mathbf{F}'\mathbf{F} & \mathbf{F}'\mathbf{V} \\ \mathbf{V}'\mathbf{F} & \mathbf{G}_v \end{pmatrix}^{-1} = \begin{pmatrix} (\mathbf{F}'\mathbf{F})^{-1} & \mathbf{0}_{k \times l} \\ \mathbf{0}_{l \times k} & \mathbf{0}_{l \times l} \end{pmatrix} + \begin{pmatrix} -(\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'\mathbf{V} \\ \mathbf{I}_l \end{pmatrix} \mathbf{W}_v^{-1} \begin{pmatrix} -\mathbf{V}'\mathbf{F}(\mathbf{F}'\mathbf{F})^{-1} & \mathbf{I}_l \end{pmatrix},$$

where  $\mathbf{W}_v \equiv \mathbf{W} + \sigma_0^2 \mathbf{D}_v^{-1} = \mathbf{V}'\mathbf{M}_F\mathbf{V} + \sigma_0^2 \mathbf{D}_v^{-1} > 0$ ,  $\mathbf{W}$  is the same Schur complement as in NE case (3.1). The matrix  $\mathbf{W}_v$  is also Schur complement of  $\mathbf{F}'\mathbf{F}$  but now in the block matrix  $\begin{pmatrix} \mathbf{F}'\mathbf{F} & \mathbf{F}'\mathbf{V} \\ \mathbf{V}'\mathbf{F} & \mathbf{G}_v \end{pmatrix}$ . We call matrix  $\mathbf{W}_v$  more specifically as *the extended Schur complement* determined by variances  $v$ .

Substituting the last result for the inverse into MME (4.2) and rearranging, we get for  $\mathbf{Y}_v^*$  (symbols  $\bullet$  denote blocks not needed in deriving)

$$\begin{pmatrix} \beta_v^* \\ \mathbf{Y}_v^* \end{pmatrix} = \begin{pmatrix} \mathbf{F}'\mathbf{F} & \mathbf{F}'\mathbf{V} \\ \mathbf{V}'\mathbf{F} & \mathbf{G}_v \end{pmatrix}^{-1} \cdot \begin{pmatrix} \mathbf{F}'\mathbf{X} \\ \mathbf{V}'\mathbf{X} \end{pmatrix} = \begin{pmatrix} \bullet & \bullet \\ -\mathbf{W}_v^{-1}\mathbf{V}'\mathbf{F}(\mathbf{F}'\mathbf{F})^{-1} & \mathbf{W}_v^{-1} \end{pmatrix} \cdot \begin{pmatrix} \mathbf{F}'\mathbf{X} \\ \mathbf{V}'\mathbf{X} \end{pmatrix},$$

$$\mathbf{Y}_v^* = \mathbf{W}_v^{-1}\mathbf{V}'(\mathbf{I}_n - \mathbf{F}(\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}')\mathbf{X} = \mathbf{W}_v^{-1}\mathbf{V}'\mathbf{M}_F\mathbf{X}. \tag{4.3}$$

Denoting matrix  $\mathbf{W}_v^{-1}\mathbf{V}'\mathbf{M}_F \in \mathbb{R}^{l \times n}$  as  $\mathbf{T}_v$ , we just derived the following computational form of new BLUP-NE estimators.

**Proposition 7** (the computational form of BLUP-NE)

Let us consider the following full-ranked LMM model for FDSLRLM observation  $\mathbf{X}$

$$\begin{aligned} \mathbf{X} &= \mathbf{F}\boldsymbol{\beta} + \mathbf{V}\mathbf{Y} + \mathbf{Z}, \quad E\{\mathbf{Z}\} = \mathbf{0}_n, \quad Cov\{\mathbf{Z}\} = \sigma_0^2 \mathbf{I}_n, \\ Cov\{\mathbf{Y}\} &= diag\{\sigma_j^2\}, \quad Cov\{\mathbf{Y}, \mathbf{Z}\} = \mathbf{0}_{l \times n}, \quad r(\mathbf{F} \mathbf{V}) = k + l. \end{aligned}$$

Then BLUP-NE estimators  $\hat{\sigma}_1^2, \dots, \hat{\sigma}_l^2$  of parameters  $\sigma_1^2, \dots, \sigma_l^2$  are given by

$$\hat{\sigma}_j^2(\mathbf{X}) = (\mathbf{Y}_v^*)_j^2 = (\mathbf{T}_v \mathbf{X})'_j (\mathbf{T}_v \mathbf{X})_j = \mathbf{X}' \boldsymbol{\tau}_j \boldsymbol{\tau}'_j \mathbf{X}; \quad j = 1, \dots, l, \tag{4.4}$$

where  $\boldsymbol{\tau}_j \equiv (T_{vj1}, T_{vj2}, \dots, T_{vjn})'$  are rows of matrix  $\mathbf{T}_v$  and  $\mathbf{T}_v = \mathbf{W}_v^{-1}\mathbf{V}'\mathbf{M}_F$  with  $\mathbf{W}_v = diag\{\sigma_0^2/\sigma_j^2\} + \mathbf{V}'\mathbf{M}_F\mathbf{V}$ .

It is straightforward to extend our computational form to the case when some variances  $\sigma_j^2$  are zero or in other words when matrix  $\mathbf{D}_v$  is singular. For a singular  $\mathbf{D}_v$ , MME (4.2) have the alternative form (Witkovský 2012)

$$\begin{pmatrix} \mathbf{F}'\mathbf{F} & \mathbf{F}'\mathbf{V}\mathbf{D}_v \\ \mathbf{V}'\mathbf{F} & \mathbf{H}_v \end{pmatrix} \cdot \begin{pmatrix} \beta_v^* \\ \mathbf{Z}_v^* \end{pmatrix} = \begin{pmatrix} \mathbf{F}'\mathbf{X} \\ \mathbf{V}'\mathbf{X} \end{pmatrix}, \tag{4.5}$$

where  $\mathbf{H}_v \equiv \mathbf{G}_v \mathbf{D}_v + \sigma_0^2 \mathbf{I}_l = \mathbf{V}' \mathbf{V} \mathbf{D}_v + \sigma_0^2 \mathbf{I}_l$  and  $\mathbf{Y}_v^* = \mathbf{D}_v \mathbf{Z}_v^*$ .

Applying the same argument with the Banachiewicz inversion formula, we get the second version of the computational form of  $\mathbf{T}_v$  determining  $\mathbf{Y}_v^*$

$$\mathbf{T}_v = \mathbf{D}_v \mathbf{U}_v^{-1} \mathbf{V}' \mathbf{M}_F \text{ with } \mathbf{U}_v = \mathbf{W}_v \mathbf{D}_v + \sigma_0^2 \mathbf{I}_l = \mathbf{V}' \mathbf{M}_F \mathbf{V} \mathbf{D}_v + \sigma_0^2 \mathbf{I}_l. \tag{4.6}$$

As opposed to  $\mathbf{W}_v, \mathbf{G}_v$ , matrices  $\mathbf{U}_v, \mathbf{H}_v$  always exist under the assumption of  $r(\mathbf{F} \mathbf{V}) = k + l$ . Simultaneously both of  $\mathbf{U}_v, \mathbf{H}_v$  are also always invertible, since both of  $\det \mathbf{U}_v, \det \mathbf{H}_v$  are nonzero for any  $\mathbf{D}_v$  in our FDSLRLM as it is shown in the Appendix A.4. In the case of a nonsingular  $\mathbf{D}_v$  the mentioned matrices are connected via following relationships

$$\mathbf{H}_v = \mathbf{G}_v \mathbf{D}_v, \mathbf{G}_v^{-1} = \mathbf{D}_v \mathbf{H}_v^{-1}, \mathbf{U}_v = \mathbf{W}_v \mathbf{D}_v, \mathbf{W}_v^{-1} = \mathbf{D}_v \mathbf{U}_v^{-1}. \tag{4.7}$$

The version (4.5) of MME is also preferred in numerical calculations (Witkovský 2012), since it can handle not only a singular  $\mathbf{D}_v$  but also a very ill-conditioned  $\mathbf{D}_v$  appearing when  $v$  has very small positive components  $\sigma_j^2$ .

It is worth to mention that originally MME were derived under the normality assumptions [see the original work Henderson et al. (1959) or Witkovský (2012)], but from the viewpoint of least squares both versions (4.2), (4.5) of MME describe the BLUP  $\mathbf{Y}_v^*$  and BLUE  $\beta_v^*$  (Christensen 2011, Sect. 12.3) with no need to restrict distributions of  $\mathbf{Y}$  and  $\mathbf{Z}$  to be normal.

**Remark 4** —General FDSLRLM and BLUP-NE for  $\sigma_0^2$

Unlike our computational form in Proposition 7, the idea (4.1) behind BLUP-NE does not depend on the assumption of FDSLRLM full-rankness. As it is described in (Searle et al. 2009) or (Witkovský 2012) the inverse of block matrix in MME (4.2) can be calculated using any  $g$ -inverse of the block matrix. In this situation, we would have to abandon our analytic approach and explore the problem only numerically, but in econometric applications this is not the case.

As for the first component  $v_0 = \sigma_0^2$  of  $v$ , we could consider as the BLUP based natural estimator the sum of squares of white noise residuals  $\mathbf{Z}_v^* = \mathbf{X} - \mathbf{F}\beta_v^* - \mathbf{V}\mathbf{Y}_v^*$  divided by the number of degrees of freedom in FDSLRLM

$$\hat{\sigma}_0^2(\mathbf{X}) = \frac{1}{n - k - l} [\mathbf{X} - \mathbf{F}\beta_v^* - \mathbf{V}\mathbf{Y}_v^*]' [\mathbf{X} - \mathbf{F}\beta_v^* - \mathbf{V}\mathbf{Y}_v^*],$$

In the LMM framework, FDSLRLM white noise residuals  $\mathbf{Z}_v^*$  are also called *conditional residuals* (Singer et al. 2017). It is not difficult to show that such estimator would be biased even in orthogonal FDSLRLM. Therefore, as a natural estimator of  $\sigma_0^2$  we will keep the original unbiased invariant quadratic NE  $\check{\sigma}_0^2(\mathbf{X})$  of  $\sigma_0^2$ , the sum of residuals in the OLS method applied to (2.1) assuming both  $\beta$  and  $\mathbf{Y}$  as fixed divided by degrees of freedom [Eq. (2.2) in Hančová (2008)].

### 4.2 Statistical properties at known variance parameters

The derivation of theoretical properties of BLUP-NE estimates under the assumption of known  $\mathbf{v}$ , regarding the first and second order moment characteristics, can be significantly facilitated by the following lemma describing the properties of matrix  $\mathbf{T}_v$  determining the BLUP  $\mathbf{Y}_v^*$  in (4.3).

Its proof can be accomplished in a similar way as the proof of Lemma 3.1 in Hančová (2008) by a direct routine computation employing formulas (4.7),  $\mathbf{T}_v = \mathbf{W}_v^{-1}\mathbf{V}'\mathbf{M}_F$ ,  $\mathbf{W}_v = \mathbf{V}'\mathbf{M}_F\mathbf{V} + \sigma_0^2\mathbf{D}_v^{-1}$ , properties of orthogonal projectors (orthogonality, idempotency, symmetry) and Schur complements (symmetry, positive definiteness).

**Lemma 1** (basic properties of  $\mathbf{T}_v$ )

- (1)  $\mathbf{T}_v\mathbf{T}_v' = \mathbf{W}_v^{-1}(\mathbf{I}_l - \sigma_0^2\mathbf{D}_v^{-1}\mathbf{W}_v^{-1}) = \mathbf{D}_v\mathbf{U}_v^{-1}(\mathbf{I}_l - \sigma_0^2\mathbf{U}_v^{-1})$
- (2)  $\mathbf{T}_v\mathbf{F} = \mathbf{0}_{l \times n}$  and  $\mathbf{T}_v\mathbf{V} = \mathbf{I}_l - \sigma_0^2\mathbf{D}_v^{-1}\mathbf{W}_v^{-1} = \mathbf{I}_l - \sigma_0^2\mathbf{U}_v^{-1}$
- (3)  $\mathbf{T}_v\boldsymbol{\Sigma}_v\mathbf{T}_v' = \mathbf{D}_v - \sigma_0^2\mathbf{W}_v^{-1} = \mathbf{D}_v(\mathbf{I}_l - \sigma_0^2\mathbf{U}_v^{-1})$

The computational forms (4.4) of BLUP-NE are also quadratic forms of  $\mathbf{X}$ . In addition, result (2)  $\mathbf{T}_v\mathbf{F} = \mathbf{0}_{l \times n}$  implies that  $\boldsymbol{\tau}_j\mathbf{F} = \mathbf{0}$ . Such a condition leads to the same conclusion as in the case of NE. BLUP based natural estimators  $\hat{\sigma}_j^2(\mathbf{X})$  are always translation invariant or shortly invariant quadratic estimators (Štulajter (2002, Sect. 1.5) or Hančová (2008, Sect. 3.1)). The following theorem summarizes theoretical properties of BLUP-NE.

**Theorem 2** (statistical properties of  $\hat{\sigma}_j^2(\mathbf{X})$ )

Natural estimators  $\hat{\sigma}_j^2(\mathbf{X})$ ;  $j = 1, 2, \dots, l$  of  $\mathbf{v}$  based on the BLUP  $\mathbf{Y}_v^*$  of  $\mathbf{Y}$  are invariant quadratic estimators having the following properties

- (1)  $E_v \left\{ \hat{\sigma}_j^2(\mathbf{X}) \right\} = \sigma_j^2 - \sigma_0^2(\mathbf{W}_v^{-1})_{jj}$ ;  $j = 1, 2, \dots, l$ ,  
If  $\mathbf{X} \sim \mathcal{N}_n(\mathbf{F}\boldsymbol{\beta}, \boldsymbol{\Sigma}_v)$ , then
- (2)  $D_v\{\hat{\sigma}_j^2(\mathbf{X})\} = 2(\sigma_j^2 - \sigma_0^2(\mathbf{W}_v^{-1})_{jj})^2$ ;  $j = 1, 2, \dots, l$ ,
- (3)  $Cov_v \left\{ \hat{\sigma}_i^2(\mathbf{X}), \hat{\sigma}_j^2(\mathbf{X}) \right\} = 2(\sigma_0^2(\mathbf{W}_v^{-1})_{ij})^2$ ;  $i, j = 1, 2, \dots, l, i \neq j$ ,
- (4)  $MSE_v\{\hat{\sigma}_j^2(\mathbf{X})\} = 2(\sigma_j^2 - \sigma_0^2(\mathbf{W}_v^{-1})_{jj})^2 + \sigma_0^4(\mathbf{W}_v^{-1})_{jj}^2$ ;  
 $j = 1, 2, \dots, l$ .

**Proof** See the Appendix A.4. □

**Remark 5** Bias of BLUP-NE

If some components in  $\mathbf{v}$  are zero, there is a need to use expression  $\mathbf{D}_v\mathbf{U}_v^{-1}$  instead of  $\mathbf{W}_v^{-1}$ . The first property, biasedness of  $\hat{\sigma}_j^2(\mathbf{X})$ ,  $j = 1, 2, \dots, l$ , is in full accordance with the Ghosh theorem (Ghosh 1996; Gajdoš et al. 2017) about the incompatibility between simultaneous non-negativity and unbiasedness of estimators for variances of random components  $\mathbf{Y}$  in LMM. As for bias, defined as  $\Delta_v \left\{ \hat{\sigma}_j^2(\mathbf{X}) \right\} = E_v \left\{ \hat{\sigma}_j^2(\mathbf{X}) \right\} - \sigma_j^2$ , we have the following expression given by the extended Schur complement  $\mathbf{W}_v$

$$\Delta_v \left\{ \hat{\sigma}_j^2(\mathbf{X}) \right\} = -\sigma_0^2(\mathbf{W}_v^{-1})_{jj} \tag{4.8}$$

**Table 3** First and second order moment characteristics for BLUP-NE and NE

Characteristic	NE estimators $\check{\sigma}_j^2(\mathbf{X})$	BLUP-NE estimators $\hat{\sigma}_j^2(\mathbf{X})$
$\Delta_{\mathbf{v}}$	$\sigma_0^2(\mathbf{W}^{-1})_{jj}$	$-\sigma_0^2(\mathbf{W}_{\mathbf{v}}^{-1})_{jj}$
$D_{\mathbf{v}}$	$2(\sigma_j^2 + \sigma_0^2(\mathbf{W}^{-1})_{jj})^2$	$2(\sigma_j^2 - \sigma_0^2(\mathbf{W}_{\mathbf{v}}^{-1})_{jj})^2$
$MSE_{\mathbf{v}}$	$2(\sigma_j^2 + \sigma_0^2(\mathbf{W}^{-1})_{jj})^2 + \sigma_0^4(\mathbf{W}^{-1})_{jj}^2$	$2(\sigma_j^2 - \sigma_0^2(\mathbf{W}_{\mathbf{v}}^{-1})_{jj})^2 + \sigma_0^4(\mathbf{W}_{\mathbf{v}}^{-1})_{jj}^2$

Now we can theoretically compare quality of the BLUP natural estimators with original natural estimators (NE) proposed in our previous paper (Hančová 2008). As we can see from summarizing Table 3, their quality is determined by Schur complements  $\mathbf{W}$  and  $\mathbf{W}_{\mathbf{v}}$  where both of them are symmetric and positive definite matrices.

Using elementary properties of the Löwner partial ordering (Puntanen et al. 2013, chap. 24), we have immediately relations  $\mathbf{W}_{\mathbf{v}} > \mathbf{W}$ ,  $\mathbf{W}^{-1} > \mathbf{W}_{\mathbf{v}}^{-1}$ ,  $(\mathbf{W})_{jj} > (\mathbf{W}_{\mathbf{v}}^{-1})_{jj}$  which imply a smaller absolute value of bias, dispersion and MSE for BLUP-NE  $\hat{\sigma}_j^2(\mathbf{X})$  in comparison with original NE  $\check{\sigma}_j^2(\mathbf{X})$ . This conclusion directly supports our idea of the improvement leading to the BLUP-NE definition (4.1).

Under the condition of orthogonality (2.3), we can write for matrices  $\mathbf{M}_{\mathbf{F}}$ ,  $\mathbf{U}_{\mathbf{v}}$ ,  $\mathbf{W}_{\mathbf{v}}$ ,  $\mathbf{T}_{\mathbf{v}}$  and BLUP-NE  $\hat{\sigma}_j^2(\mathbf{X})$

$$\begin{aligned} \mathbf{M}_{\mathbf{F}}\mathbf{V} &= \mathbf{V}, \quad \mathbf{U}_{\mathbf{v}} = \mathbf{G}_{\mathbf{v}}\mathbf{D}_{\mathbf{v}} + \sigma_0^2\mathbf{I}_l = \text{diag} \left\{ \sigma_0^2 + \sigma_j^2 \|\mathbf{v}_j\|^2 \right\}, \\ \mathbf{W}_{\mathbf{v}}^{-1} &= \mathbf{D}_{\mathbf{v}}\mathbf{U}_{\mathbf{v}}^{-1} = \text{diag} \left\{ \rho_j / \|\mathbf{v}_j\|^2 \right\}, \quad \boldsymbol{\tau}_j = \rho_j \mathbf{v}'_j / \|\mathbf{v}_j\|^2, \\ \hat{\sigma}_j^2(\mathbf{X}) &= \rho_j^2 \mathbf{X}'\mathbf{v}_j\mathbf{v}'_j\mathbf{X} / \|\mathbf{v}_j\|^4, \quad j = 1, 2, \dots, l, \end{aligned}$$

where we introduced  $\rho_j \equiv \frac{\sigma_j^2 \|\mathbf{v}_j\|^2}{\sigma_0^2 + \sigma_j^2 \|\mathbf{v}_j\|^2} \in \mathbb{R}$ ,  $0 \leq \rho_j < 1$ ,  $j = 1, 2, \dots, l$ .

Applying these results, we get the following direct corollary of Theorem 2 for any orthogonal FDSLRLM.

**Corollary 1** (properties of  $\hat{\sigma}_j^2(\mathbf{X})$  in orthogonal FDSLRLM)

*In an orthogonal FDSLRLM natural estimators  $\hat{\sigma}_j^2$  based on the BLUP  $\mathbf{Y}_{\mathbf{v}}^*$  have the following properties*

$$(1) \quad E_{\mathbf{v}} \left\{ \hat{\sigma}_j^2(\mathbf{X}) \right\} = \rho_j \sigma_j^2; \quad j = 1, 2, \dots, l, \quad \text{where } \rho_j = \frac{\sigma_j^2 \|\mathbf{v}_j\|^2}{\sigma_0^2 + \sigma_j^2 \|\mathbf{v}_j\|^2} \in \mathbb{R}_+.$$

*If  $\mathbf{X} \sim \mathcal{N}_n(\mathbf{F}\boldsymbol{\beta}, \boldsymbol{\Sigma}_{\mathbf{v}})$ , then*

- (2)  $D_{\mathbf{v}}\{\hat{\sigma}_j^2(\mathbf{X})\} = 2\rho_j^2\sigma_j^4; \quad j = 1, 2, \dots, l,$
- (3)  $Cov_{\mathbf{v}} \left\{ \hat{\sigma}_i^2(\mathbf{X}), \hat{\sigma}_j^2(\mathbf{X}) \right\} = 0; \quad i, j = 1, 2, \dots, l, \quad i \neq j,$
- (4)  $MSE_{\mathbf{v}}\{\hat{\sigma}_j^2(\mathbf{X})\} = [2\rho_j^2 + (1 - \rho_j)^2]\sigma_j^4; \quad j = 1, 2, \dots, l.$

Finally, if we look at original natural estimators  $\check{\sigma}_j^2(\mathbf{X})$  from our paper (Hančová 2008), then in orthogonal FDSLRLM it can be easily shown that  $\hat{\sigma}_j^2(\mathbf{X}) = \rho_j^2 \check{\sigma}_j^2(\mathbf{X})$ ,

$j = 1, 2, \dots, l$ . If we introduce  $\rho_0 = 1$ , then we obtain the complete orthogonal version of (4.4) for computing BLUP-NE  $\hat{\sigma}_j^2(\mathbf{X})$ :

$$\hat{\sigma}_j^2(\mathbf{X}) = \rho_j^2 \check{\sigma}_j^2(\mathbf{X}), j = 0, 1, \dots, l. \tag{4.9}$$

### 4.3 Empirical version of BLUP-NE method

The computational form of our BLUP natural estimators, their first and second order moment characteristics were derived under the assumption of known variances  $\mathbf{v}$ , which is in practice rarely fulfilled. In such case, the simplest reasonable solution is to use an empirical version of BLUP (EBLUP), as it is usual in the general theory of empirical BLUPs in LMMs (Štulajter (2002, chap. 5), Witkovský (2012) and Rao and Molina 2015, chap. 5). In particular, this step means replacing unknown true parameters  $\mathbf{v}$  with other “initial” values or estimates of  $\mathbf{v}$  in FDSLRLM. In the light of iterative approaches, such empirical version of BLUP-NE can be viewed as a two-stage iterative estimation method with one step in the iteration.

From the theoretical perspective, we prefer initial estimates with the simplest explicit form and under least restrictive assumptions on the structure or distribution of FDSLRLM. On the other hand, from the practical point of view, it will be sufficient to have such initial estimates which can be obtained by reliable and time efficient computational methods. Therefore, we suggest natural estimators (NE) as initial estimates for BLUP-NE, although we could consider any estimates based on least squares or maximum likelihood.

Using (4.9), these new distribution-free estimates which will be referred as empirical BLUP-NE or EBLUP-NE for short, are invariant and given in orthogonal FDSLRLM by

$$\hat{\sigma}_j^2(\mathbf{X}) = \check{\rho}_j^2 \check{\sigma}_j^2(\mathbf{X}), j = 0, 1, \dots, l, \quad \check{\sigma}_j^2(\mathbf{X}) \text{ are NE}, \tag{4.10}$$

where  $\check{\rho}_0 = 1, \check{\rho}_j = \frac{\check{\sigma}_j^2 \|v_j\|^2}{\check{\sigma}_0^2 + \check{\sigma}_j^2 \|v_j\|^2}$ .

As for computational complexity of EBLUP-NE, we get  $5l$  or  $\mathcal{O}(l)$  extra operations to compute  $\check{\rho}_j^2$  and product with  $\check{\sigma}_j^2(\mathbf{X})$  in comparison with NE from (3.2). Therefore, computing EBLUP-NE represents again an algorithm with the complexity having order  $\mathcal{O}(n)$ . It has the same order as our KKT algorithm, but in computational practice it will be a little quicker, since KKT algorithm extra operations are  $\mathcal{O}(l \cdot 2^l)$ , which should be observable in cases of smaller  $n$ , comparable with  $l \cdot 2^l$ .

**Remark 6** EBLUPs from NE

According to general theory of EBLUPs (Štulajter 2002; Żądło 2009; Witkovský 2012; Rao and Molina 2015), if initial estimates  $\tilde{\mathbf{v}}$  of  $\mathbf{v}$  are even, invariant and probability distribution of  $\mathbf{Y}$  and  $\mathbf{Z}$  in LMM (2.1) are both symmetric around zero with finite mean values, then empirical non-linear BLUP  $Y_{\tilde{\mathbf{v}}}^*$  based on  $\tilde{\mathbf{v}}$  is unbiased. Our quadratic natural estimators NE, which are even and invariant, belong to this type of initial estimates.

However, general theory of EBLUPs also says that an explicit expression for MSE of EBLUP is not known. Non-linearity of EBLUP causes that theoretical properties



of EBLUP can be usually explored only by approximate methods, using functional series, which will be our goal in future research.

## 5 Application in real data examples

### 5.1 Three real data sets and their FDSLRLMs

Generally, since EBLUPs are nonlinear functions of the observation  $\mathbf{X}$ , their theoretical study (especially their second order statistical properties) is considered as a very difficult mathematical task and it is still open research theoretical problem in LMM and FDSLRLM (Witkovský 2012; Gajdoš et al. 2017). Therefore, we will illustrate our theoretical results, performance and properties of the proposed EBLUP-NE method applying numerical and computer simulation approach.

We use three real data sets: *electricity consumption, tourism and cyber security*. For all data sets, we identified the most parsimonious structure of the FDSLRLM using an iterative process of the model building and selection based on exploratory tools of spectral analysis and LMM theory. All details of our FDSLRLM analysis and modeling can be found in our easy reproducible Jupyter notebooks freely available at our GitHub fdslrlm repository *EBLUP-NE* (Gajdoš et al. 2019a).

#### 5.1.1 Electricity consumption

As the first real data example, we have the econometric time series data set, representing hourly observations of the consumption of the electric energy (in kWh) in a department store. The number of time series observations is  $n = 24$ . The data and model was adapted from Štulajter and Witkovský (2004)

$$\begin{aligned}
 X(t) = & \beta_1 + \beta_2 \cos(\omega_1 t) + \beta_3 \sin(\omega_1 t) + \\
 & + Y_1 \cos(\omega_2 t) + Y_2 \sin(\omega_2 t) + \\
 & + Y_3 \cos(\omega_3 t) + Y_4 \sin(\omega_3 t) + Z(t), \quad t \in \mathbb{N}
 \end{aligned}
 \tag{5.1}$$

with  $k = 3, l = 4, \mathbf{Y} = (Y_1, Y_2, Y_3, Y_4)' \sim \mathcal{N}_4(\mathbf{0}_4, \mathbf{D}_\mathbf{v})$ ,  $Z(\cdot) \sim iid \mathcal{N}(0, \sigma_0^2)$  and  $\mathbf{v} = (\sigma_0^2, \sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2)' \in \mathbb{R}_+^5$ . Frequencies  $(\omega_1, \omega_2, \omega_3)' = 2\pi (1/24, 2/24, 3/24)'$  are significant Fourier frequencies from the time series periodogram in spectral analysis (Priestley 2004; Štulajter and Witkovský 2004; Gajdoš et al. 2017).

Since the time series data set contains only 24 observations and we have to estimate three regression and five variance parameters of the FDSLRLM, this real data FDSLRLM example should be considered only as a toy example. This model give us the opportunity to check our numerical results and to demonstrate how in principle FDSLRLM estimation methods and computational tools work.

### 5.1.2 Tourism

In this econometric FDSLRLM application, we consider the time series data set, called *visnights*, representing *total quarterly visitor nights* (in millions) from 1998 to 2016 in one of the regions of Australia—inner zone of Victoria state. The number of time series observations is  $n = 76$ . The data was adapted from Hyndman and Athanasopoulos (2018).

The Gaussian orthogonal FDSLRLM fitting the tourism data has the following form (for modeling details see our Jupyter notebook `tourism.ipynb`):

$$X(t) = \beta_1 + \beta_2 \cos\left(\frac{2\pi t}{76}\right) + \beta_3 \sin\left(\frac{2\pi t \cdot 2}{76}\right) + Y_1 \cos\left(\frac{2\pi t \cdot 19}{76}\right) + Y_2 \sin\left(\frac{2\pi t \cdot 19}{76}\right) + Y_3 \cos\left(\frac{2\pi t \cdot 38}{76}\right) + Z(t), \quad t \in \mathbb{N} \quad (5.2)$$

with  $k = 3, l = 3$ ,  $\mathbf{Y} = (Y_1, Y_2, Y_3)' \sim \mathcal{N}_3(\mathbf{0}_3, \mathbf{D}_\mathbf{v})$ ,  $Z(\cdot) \sim iid \mathcal{N}(0, \sigma_0^2)$  and  $\mathbf{v} = (\sigma_0^2, \sigma_1^2, \sigma_2^2, \sigma_3^2)' \in \mathbb{R}_+^4$ .

### 5.1.3 Cyber attacks

Our final FDSLRLM application describes the real time series data set representing the total weekly number of cyber attacks against a honeynet—an unconventional tool which mimics real systems connected to Internet, like business or school computers intranets, to study methods, tools and goals of cyber attackers. Data, taken from Sokol and Gajdoš (2018), were collected from November 2014 to May 2016 in CZ.NIC honeynet consisting of Kippo honeypots in medium-interaction mode. The number of time series observations is  $n = 72$ .

The suitable FDSLRLM, after a preliminary logarithmic transformation of data  $L(t) = \log X(t)$ , is again Gaussian orthogonal (for modeling details see our Jupyter notebook `cyberattacks.ipynb`) and in comparison with previous models (5.1), (5.2) has the simplest structure:

$$L(t) = \beta_1 + \beta_2 \cos\left(\frac{2\pi t}{72}\right) + \beta_3 \sin\left(\frac{2\pi \cdot 3t}{72}\right) + Y_1 \cos\left(\frac{2\pi \cdot 14t}{72}\right) + Y_2 \sin\left(\frac{2\pi \cdot 14t}{72}\right) + Z(t), \quad t \in \mathbb{N}, \quad (5.3)$$

with  $k = 3, l = 2$ ,  $\mathbf{Y} = (Y_1, Y_2)' \sim \mathcal{N}_2(\mathbf{0}_2, \mathbf{D}_\mathbf{v})$ ,  $Z(\cdot) \sim iid \mathcal{N}(0, \sigma_0^2)$  and  $\mathbf{v} = (\sigma_0^2, \sigma_1^2, \sigma_2^2)' \in \mathbb{R}_+^3$ .

### 5.1.4 Numerical results of estimating $\mathbf{v}$

For cross-checking purposes, we realized our numerical computations in both Python and R based software tools (or packages). As we describe in the Appendix A.3, we implemented own algorithms and methods in SciPy, SageMath and R, particularly analytical expressions (3.2) for NE and the KKT algorithm for NN-(M)DOOLSE (Table 2) in orthogonal FDSLRLM. Simultaneously, we confirmed the same estimations using CVXPY (or CVXR) package based on convex optimization and analogically using up-to-date standard LMM R packages `nlme`, `MMEinR` (mixed) and `sommer`.

Table 4 summarizes all types of considered estimates in all four models. Since computations also confirmed equivalency of NN-(M)DOOLSE and (RE)MLE described by Theorem 1, we wrote results for (RE)MLE only.

Detailed computational results for all three data sets using all corresponding relevant tools can be found and easily reproduced in our collection of Jupyter notebooks (asterisk \* represents a name of data and specific computational tool): `PY-estimation-*.ipynb` and `R-estimation-*.ipynb`, available in repository *EBLUP-NE* (Gajdoš et al. 2019c).

## 5.2 Simulation study of EBLUP-NE properties

In this section we present the results of a simulation study in the following scenario:

- (1) We take all three considered Gaussian FDSLRLM models (5.1), (5.2), (5.3);
- (2) We set NE of variances  $\check{\mathbf{v}}$  and OLSE of regression parameters  $\beta^*$ , computed from corresponding data sets (the first and last column in Table 4), as true values of models' parameters;
- (3) For each model we simulate  $N = 5000$  time series realizations of  $X$  by Monte Carlo from corresponding multivariate normal distributions  $\mathcal{N}_n(\mathbf{F}\beta^*, \Sigma_{\check{\mathbf{v}}})$ ;
- (4) Finally we calculate NE, (RE)MLE, EBLUP-NE and their statistical properties – Monte Carlo mean value (Mean) and Monte Carlo mean squared error (MSE).

Complete results of simulation summaries for all models are reported in Table 5.

As expected, NE exhibit the biggest biases, the biggest MSEs from all methods. Mean values of NE are always bigger than true variance parameters  $\mathbf{v}$  which is consistent with the positive theoretical bias in Table 3. As for the proposed method EBLUP-NE, results indicate that it is a really improvement of the NE method. Moreover, it seems that although we replaced in EBLUP-NE true parameters by NE, results indicate that absolute value of bias and MSE still remained smaller as we conclude from Table 3 for theoretical BLUP-NE at known  $\mathbf{v}$ .

Comparing the proposed method EBLUP-NE with MLE and REMLE, it is important for our further research that EBLUP-NE slightly outperforms both of them in majority of variance parameters (achieves smaller biases and smaller MSEs). But differences are really small, so results indicate that EBLUP-NE shows at least comparable performance as MLE and REMLE. Simulations also confirmed comparable runtimes of computational form (4.10) for EBLUP-NE and the KKT algorithm (Table 2) used for (RE)MLE, which is in full accordance with the same order  $\mathcal{O}(n)$  of theoretical computational complexity.

As it can be seen in our Jupyter notebooks, all results obtained by our procedures are fully consistent with the results of the best convex optimization and LMM computational tools (see Appendix A.3). However, we also noticed that unlike our procedures or the convex optimization tools, standard LMM tools failed in small number of realizations during simulations.

Both—the computational form for EBLUP-NE and the KKT algorithm for (RE)MLE working with double floating-point precision  $\epsilon < 10^{-15}$  as the default precision of outputs—significantly outperform these Python or R-based standard com-

**Table 4** Estimates for real datasets using different estimation methods

Variance parameter	Estimation				Regression parameter	Estimation OLSE
	NE	EBLUP-NE	MLE	REMLE		
<b>Electricity consumption—toy model</b>						
$\nu_0$	1.09	1.09	0.93	1.09	$\beta_1$	44.38
$\nu_1$	2.97	2.79	2.89	2.87	$\beta_2$	-3.15
$\nu_2$	1.76	1.59	1.68	1.67	$\beta_3$	-3.52
$\nu_3$	0.37	0.24	0.29	0.28		
$\nu_4$	1.86	1.69	1.79	1.77		
norm $\ \nu\ $	4.09	3.80	3.91	3.93		
<b>Tourism</b>						
$\nu_0$	0.1077	0.1077	0.1032	0.1077	$\beta_1$	4.2535
$\nu_1$	0.0039	0.0013	0.0012	0.0011	$\beta_2$	0.2556
$\nu_2$	0.2303	0.2247	0.2276	0.2275	$\beta_3$	-0.2474
$\nu_3$	0.0223	0.0197	0.0209	0.0209		
norm $\ \nu\ $	0.2552	0.2500	0.2508	0.2525		
<b>Cyber attacks</b>						
$\nu_0$	0.0593	0.0593	0.0560	0.0593	$\beta_1$	7.6833
$\nu_1$	0.0255	0.0225	0.0239	0.0238	$\beta_2$	-0.1448
$\nu_2$	0.0155	0.0127	0.0139	0.1380	$\beta_3$	0.2015
norm $\ \nu\ $	0.0664	0.0647	0.0624	0.0654	$\beta_4$	-0.1297

**Table 5** Simulation results comparing estimation methods in three models

Parameter (true)	NE		EBLUP-NE		MLE		REMLE	
	Mean	MSE	Mean	MSE	Mean	MSE	Mean	MSE
<b>Electricity consumption—toy model</b>								
$\gamma_0 = 1.09$	1.10	0.14	1.10	0.14	0.91	0.13	1.07	0.13
$\gamma_1 = 2.97$	3.01	17.53	2.87	17.32	2.94	17.49	2.93	17.48
$\gamma_2 = 1.76$	1.81	7.00	1.68	6.85	1.74	6.97	1.73	6.96
$\gamma_3 = 0.37$	0.47	0.45	0.38	0.39	0.41	0.43	0.41	0.43
$\gamma_4 = 1.86$	1.94	7.59	1.81	7.43	1.87	7.55	1.86	7.54
<b>Tourism</b>								
$\gamma_0 = 0.1077$	0.1078	0.0003	0.1078	0.0003	0.1027	0.0003	0.1071	0.0003
$\gamma_1 = 0.0039$	0.0065	0.0001	0.0043	0.0001	0.0047	0.0001	0.0046	0.0001
$\gamma_2 = 0.2303$	0.2321	0.1083	0.2273	0.1080	0.2295	0.1083	0.2294	0.1083
$\gamma_3 = 0.0223$	0.0244	0.0011	0.0224	0.0011	0.0232	0.0011	0.0232	0.0011
<b>Cyber attacks</b>								
$\gamma_0 = 0.0593$	0.0594	0.0001	0.0594	0.0001	0.0558	0.0001	0.0592	0.0001
$\gamma_1 = 0.0255$	0.0287	0.0017	0.0264	0.0017	0.0273	0.0017	0.0272	0.0017
$\gamma_2 = 0.0155$	0.0170	0.0006	0.0150	0.0006	0.0157	0.0006	0.0157	0.0006

putational tools in precision ( $10^7$  times more accurate) and also in run-times (10–100 times faster on the current standard PC).<sup>6</sup>

## 6 Conclusions

In this paper we revisited and updated theoretical and computational knowledge dealing with FDSLRLM least squares and maximum likelihoods estimation procedures, in the light of recent advances in closely related convex optimization, linear mixed modeling (LMM), and modern computational technology.

Specifically, applying the convex optimization theory, we reformulated all estimation methods as convex optimization problems in the so-called orthogonal FDSLRLM, the most usual form of FDSLRLM used in econometric practice. We formulated the KKT optimality conditions, which, unlike likelihood equations, are necessary and sufficient conditions for optimal solutions of (RE)MLE or NN-(M)DOOLSE on extended parametric space  $\Upsilon = [0, \infty)^{l+1}$ .

KKT optimality conditions dictate not only the exact existence conditions of estimates but they also solve the well-known problem dealing with standardly used likelihood equations (Christensen (2011, chap. 12) and Searle et al (2009, chap. 6)) in LMM where their solutions for (RE)MLE or NN-(M)DOOLSE may not be required estimates—they may be out of the parameter space or they may be other than the maximum or minimum.

Moreover, using KKT optimality conditions in orthogonal FDSLRLMs, we proved the equivalence of NN-(M)DOOLSE and (RE)MLE with probability 1. This important theoretical result of the paper is the stronger and more general result than in Štulajter and Witkovský (2004) proved only for interior points of  $\Upsilon$ .

Simultaneously, the convex optimization theory brought us to the new KKT algorithm for computing NN-(M)DOOLSE, equivalent to (RE)MLE, with double floating-point precision  $\epsilon < 10^{-15}$  as the default precision of outputs and with computational complexity  $\mathcal{O}(n)$ .

Such an algorithm which we implemented in SciPy, SageMath and R (all free and open-source) and which at the default precision level is  $10^7$  times more accurate and approximately  $n^2$  faster than the best current Python(or R)-based standard computational tools, can be used in effective computational time series research to study properties of FDSLRLM (Monte Carlo and bootstrap methods, Kreiss and Lahiri (2012)).

As a consequence of applying the general theory of BLUP in LMM, we suggested and investigated an alternative, new method EBLUP-NE based on empirical BLUPs for estimating variances in time series modeled by FDSLRLM. The estimation method can be also viewed, analogously like EBLUP (Rao and Molina 2015; Witkovský 2012), as a two-stage iterative method with one step in the iteration.

---

<sup>6</sup> It is worth to mention that the speed of our Python procedures is not final. Our Jupyter notebooks have been written in standard Python in SciPy. We assume that the rewriting of the code, using special Python tools (Cython or Numba), speeds up our procedures (10–100 times faster) to run-times comparable with C or C++ codes.

EBLUP-NE are almost surely positive invariant estimators whose simple explicit computational form are given by two special matrices—the Schur complement and Gram matrix (Zhang 2005). The method can be used not only in the case of normally distributed time series data, but for any absolutely continuous probability distribution of time series data.

As initial (plug-in) estimates for EBLUP-NE, we decided to employ natural estimators (NE) based on least squares. Applications of FDSLRLM with the EBLUP-NE on three real data sets (electricity consumption, tourism, cyber attacks) providing one toy and two real models, together with a corresponding simulation study, indicate that the method computationally gives at least comparable results with REMLE or NN-MDOOLSE, but in faster run-time ( $n^2$  times faster which in our examples meant approximately 10–1000 times faster real run-times on the standard PC). Unlike REMLE or NN-DOOLSE, EBLUP-NE have simple explicit analytic forms, which can be used in further theoretical study connected to kriging predictions dealing with EBLUPs theory.

Regarding computational aspects, we were also successful in the identification and demonstration of consistent results for the real data applications of FDSLRLM in several free, open-source current standard computational tools—namely CVXPY, CVXR (R version of CVXPY) packages for convex optimization and LMM R packages nlme, sommer and MMEinR (our R version of Witkovsky's MATLAB mixed function). These results and procedures can be freely viewed in our 15 Jupyter notebooks which are easily readable, sharable, reproducible and modifiable directly in our GitHub repository (Gajdoš et al. 2019a). Open-source Jupyter technology with Python and R packages also solved our problem stated in Gajdoš et al. (2017) that no current package in R was directly and effectively suitable for FDSLRLM.

In connection with our current computational research in FDSLRLM, but also for real time series data analysis and forecasting, we started to build our own R package (see a fully functional version in our *fdslrm* storage *R-package* Gajdoš et al. (2019d)) on mentioned LMM R packages to manipulate comfortably and readily with FDSLRLM concepts and procedures.

Our investigation has also brought new questions for further research. There is definitely a need for more extended analysis based on a simulation study and general EBLUP theory (Žađlo 2009; Witkovský 2012; Rao and Molina 2015) focusing on EBLUP-NE quality with respect to previously used estimation methods and the performance of the EBLUP-NE method itself with different initial starting points (not only NE), using different computational tools in various probability distributions. These research questions are under our current intensive investigation.

Finally, our results in the paper can be seen reciprocally as contributions to convex optimization and LMM methodology. Particularly, our convex optimization application in the context of time series modeling has become another one from a wide variety of application areas of convex optimization (Boyd and Vandenberghe 2009). Since FDSLRLM describing  $n$  observed time series values is also a special type of LMM, our EBLUP-NE and the very fast, accurate KKT algorithm to compute (RE)MLE may have potential to be used in computational research and applications dealing with LMMs.

**Acknowledgements** We are very grateful especially to František Štulajter (Comenius University in Bratislava, SK) for his guidance during our early contact and study of FDSLRLM concepts and to Viktor Witkovský (Slovak Academy of Science, SK) for his recommendations and deep insights dealing with the LMM methodology and its computational tools connected to FDSLRLM. We would also like to thank Ivan Žezula and Daniel Klein (both from P. J. Šafárik University in Košice, SK) for a valuable feedback dealing with considered estimation methods. Concerning applied computational tools, we would also like to acknowledge involvement and recommendations of Giovanni Covarrubias-Pazarán (University of Wisconsin, US) in using sommer package, Steve Diamond and Stephen Boyd (Stanford University, US) in using CVXPY and CVXR. Finally, in connection with using SageMath, Jupyter and GitHub, our special thanks go namely to William Stein (University of Washington, US; now CoCalc company), Harald Schilly (Universität Wien, AT), Erik Bray (Université Paris-Sud, FR), Luca de Feo (Université de Versailles Saint-Quentin-en-Yvelines, FR) and Charles J. Weiss (Augustana University, US). Finally, we very much appreciate the helpful and constructive comments of the referees which improved clarity and consistency of the paper substantially.

## Appendix

### A.1 Acronyms and abbreviations

See Table 6

**Table 6** List of acronyms used in the paper

Acronym	Explanation	Specification
BLUE	Best linear unbiased estimator	p. 14, (4.2) p. 18
BLUP	Best linear unbiased predictor	p. 2, (4.2) p. 18
BLUP-NE	Natural estimators based on BLUP	(4.1) p. 18
DOOLSE	Double ordinary least squares estimator without non-negativity constraints	(3.6) p. 10
EBLUP	Empirical (plug-in) BLUP	p. 22
EBLUP-NE	Natural estimators based on EBLUP	p. 23
FDSLRLM	Finite discrete spectrum linear regression model	(1.1) p. 2
KKT	Karush–Kuhn–Tucker	p. 12
LMM	Linear mixed model	p. 4
LRM	Linear regression model	p. 2
MDOOLSE	Modified (unbiased) DOOLSE without non-negativity constraints	(3.6) p. 10
(M)DOOLSE	Considering both DOOLSE and MDOOLSE	Remark 2 p. 10
MINQE	Minimum norm quadratic estimators	p. 3
MINQE(I)	Invariant MINQE	Remark 2 p. 10
MINQE(U,I)	Invariant unbiased MINQE	Remark 2 p. 10
MLE	Maximum likelihood estimators	(3.13) p. 15
MM	Method of moments (Henderson method 3)	Remark 1 p. 9
MME	Henderson's mixed model equations	(4.2) p. 18



**Table 6** continued

Acronym	Explanation	Specification
MSE	Mean squared error	p. 2
NE	Natural estimators	p. 8, (3.2) p. 8
NN-DOOLSE	Non-negative DOOLSE	(3.4) p. 10
NN-MDOOLSE	Non-negative modified DOOLSE	(3.5) p. 10
NN-(M)DOOLSE	Considering both NN-DOOLSE and NN-MDOOLSE	Remark 2 p. 10
OLS	Ordinary least squares	p. 8
REMLE	Residual (restricted) maximum likelihood estimator	(3.14) p. 15
(RE)MLE	Considering both MLE and REMLE	Remark 2 p. 10
VLS	Variance least squares estimator	Remark 2 p. 10
UVLS	Unbiased VLS	Remark 2 p. 10

**A.2 KKT algorithm for NN-(M)DOOLSE**

If we consider  $\nu_0 = 0$  which occurs if and only if  $\mathbf{e} = \sum_{j=1}^l \alpha_j \mathbf{v}_j$  in KKT conditions (3.8)–(3.10) then the optimal solution  $\tilde{\mathbf{v}}$  for NN-(M)DOOLSE is trivial:  $\nu_0 = 0, \nu_j = \alpha_j^2, j = 1, \dots, l$ . For  $\nu_0 \neq 0$  implying  $\lambda_0 = 0, \nu_j$  and  $\lambda_j$  cannot be simultaneous zero otherwise it would lead to contradiction between (3.8) and (3.9). Therefore, in the case  $\nu_0 \neq 0$ , the complementary slackness condition (3)  $\nu_j \lambda_j = 0, j \in \{1, \dots, l\}$  can be rewritten in the form  $b_j \nu_j (1 - b_j) \lambda_j = 0$ , where  $b_j$  is an auxiliary indicator, which is zero if  $\nu_j = 0, \lambda_j \neq 0$  and one if  $\nu_j \neq 0, \lambda_j = 0$ .

Using vector  $\mathbf{b} = (b_1, b_2, \dots, b_l)' \in \{0, 1\}^l$ , the derived KKT conditions (2) in Proposition 2 can be described by a  $(l + 1) \times (l + 1)$  matrix function  $\mathbf{K}(\mathbf{b}) = \{K_{ij}(\mathbf{b})\}$  and a vector function  $\boldsymbol{\gamma}(\mathbf{b}) = (\gamma_0(\mathbf{b}), \gamma_1(\mathbf{b}), \dots, \gamma_l(\mathbf{b}))'$  as

$$\mathbf{K}(\mathbf{b})\boldsymbol{\gamma}(\mathbf{b}) = \mathbf{q},$$

where

$$K_{ij}(\mathbf{b}) = \begin{cases} 0, & \text{if } i = 0, b_j = 0, \\ -1, & \text{if } i = j \neq 0, b_j = 0, \\ G_{ij}, & \text{otherwise.} \end{cases} \quad \gamma_j(\mathbf{b}) = \begin{cases} \nu_0, & \text{if } j = 0 \\ b_j \nu_j + (1 - b_j) \lambda_j, & \text{otherwise.} \end{cases}$$

Applying the Banachiewicz formula, we can write for the inverse of  $\mathbf{K}(\mathbf{b})$  the following analytic expression

$$\mathbf{K}(\mathbf{b})^{-1} = \phi^{-1} \begin{pmatrix} 1 & -\mathbf{b}'\mathbf{G}_V\mathbf{D}_b^{-1} \\ -\mathbf{D}_b^{-1}\mathbf{G}_V\mathbf{j}_l & \phi\mathbf{D}_b^{-1} + \mathbf{D}_b^{-1}\mathbf{G}_V\mathbf{j}_l\mathbf{b}'\mathbf{G}_V\mathbf{D}_b^{-1} \end{pmatrix}$$

where

$$\mathbf{D}_b = \text{diag} \left\{ b_j \|\mathbf{v}_j\|^4 + b_j - 1 \right\}, \quad \mathbf{G}_v = \text{diag} \left\{ \|\mathbf{v}_j\|^2 \right\},$$

$$\mathbf{j}_l = (1, 1, \dots, 1)' \in \mathbb{R}^l, \quad \phi = n^* - \mathbf{b}' \mathbf{G}_v \mathbf{D}_b^{-1} \mathbf{G}_v \mathbf{j}_l.$$

Finally, Proposition 3 guarantees the existence of unique auxiliary vectors  $\tilde{\boldsymbol{\gamma}}, \tilde{\mathbf{b}}$ :

$$\tilde{\boldsymbol{\gamma}} = \{\boldsymbol{\gamma}; \boldsymbol{\gamma} = \mathbf{K}(\mathbf{b})^{-1} \mathbf{q} \geq 0, \mathbf{b} \in \{0, 1\}^l\}$$

$$\tilde{\mathbf{b}} = \{\mathbf{b}; \tilde{\boldsymbol{\gamma}} = \mathbf{K}(\mathbf{b})^{-1} \mathbf{q}, \mathbf{b} \in \{0, 1\}^l\}$$

Based on vectors  $\tilde{\boldsymbol{\gamma}}, \tilde{\mathbf{b}}$ , the NN-(M)DOOLSE  $\tilde{\mathbf{v}} = (\tilde{v}_0, \tilde{v}_1, \dots, \tilde{v}_l)'$  of  $\mathbf{v}$  as a solution of KKT conditions has the final form

$$\tilde{v}_j = \begin{cases} 0, & \text{if } j \neq 0, b_j = 0, \\ \tilde{\gamma}_j, & \text{otherwise.} \end{cases}$$

Thanks to Proposition 3, it is worth to mention that the given matrix system includes also solutions with  $v_0 = 0$ .

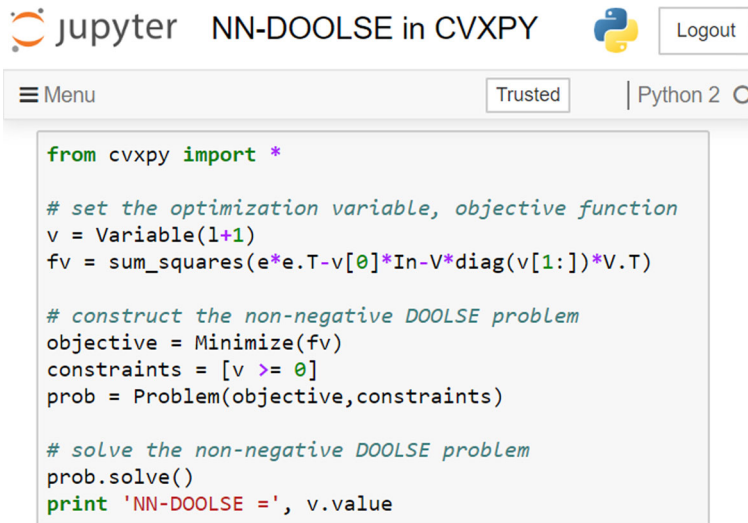
### A.3 Modern computational technology for FDSLRLM kriging

#### Our computational tools for FDSLRLM

As for computational technology, our time series calculations are carried out using free open-source software based on the R statistical language and packages (R Development Core Team 2019; McLeod et al. 2012), the Scientific Python with the SciPy ecosystem (Jones et al. 2001; Oliphant 2007) and free Python-based mathematics software SageMath (Stein and others 2019; Beezer et al. 2013; Zimmermann et al 2018), an open source alternative to the popular commercial computer algebra systems Mathematica or Maple. The simultaneous use of R, SciPy and SageMath provides us valuable cross-checking of our computational results.

Detailed records of our computing with explaining narratives have the form of Jupyter notebooks stored in our free available GitHub repositories devoted to FDSLRLM research, <https://github.com/fdslrm>; see especially a *fdslrm* repository *EBLUP-NE* dealing with this paper (Gajdoš et al. 2019c). All notebooks should be seen or studied as static HTML pages via *Jupyter nbviewer* (<https://nbviewer.jupyter.org/>, Kluyver et al. (2016)) or interactively as live HTML documents using *Binder* (<https://mybinder.org/>, Project Jupyter et al. (2018)) where the code is executable without any need to install or compile the software. To do that, it is only needed to choose and click appropriate links in the Software section of the README homepage in the mentioned repository *EBLUP-NE*. This easily reproducible way and presentation of our computing with real data were inspired by works (Briuelle et al. 2019; Weiss 2017).

We also started to develop our own R package for FDSLRLM, so any potential user can install and use a fully functional version stored in our *fdslrm* repository *R-package* (Gajdoš et al. 2019d), which allows the FDSLRLM modeling and predictions.



```

from cvxpy import *

# set the optimization variable, objective function
v = Variable(1+1)
fv = sum_squares(e*e.T-v[0]*In-V*diag(v[1:])*V.T)

# construct the non-negative DOOLSE problem
objective = Minimize(fv)
constraints = [v >= 0]
prob = Problem(objective,constraints)

# solve the non-negative DOOLSE problem
prob.solve()
print 'NN-DOOLSE =', v.value

```

Fig. 1 CVXPY code for computing DOOLSE in Jupyter environment

Its applications on various real data sets illustrating every step of our three-stage FDSLRLM-building approach—formulation (identification), estimation (fitting) and residual diagnostics (checking)—can be found in other *fdslrm* repository *applications* (Gajdoš et al. 2019b). All our results obtained from our own implementation were fully consistent with results of standard convex optimization and LMM computational tools which we applied to FDSLRLM and are described in the following subsections.

### Convex optimization tools for FDSLRLM

In nonlinear optimization, there is a variety of highly efficient, fast and reliable open-source and commercial software packages (Cornuéjols et al 2018, chap. 20). Our theoretical consideration and results allow us to apply for estimating  $\mathbf{v}$  directly the latest, one of the most well-known convex optimization packages CVXPY and CVXR based on disciplined convex programming, written in Python (Diamond and Boyd 2016) and R (Fu et al. 2019).

CVXPY and its R version CVXR are relatively unknown in statistical community. CVXPY is not only a scientific Python library, but also a language with very simple, readable syntax not requiring any expertise in convex optimization and its computer implementation. CVXPY allows the user to specify the mathematical optimization problem naturally following normal mathematical notation as we can see in computing NN-DOOLSE (Fig. 1) where a code easily mimics the non-negative DOOLSE mathematical formulation (3.4) with  $\Sigma_{\mathbf{v}} = \sigma_0^2 \mathbf{I}_n + \mathbf{V}\mathbf{D}_{\mathbf{v}}\mathbf{V}' = \nu_0 \mathbf{I}_n + \mathbf{V}\text{diag}\{\nu_j\}\mathbf{V}'$ .

CVXPY implements not only convex optimization solvers using interior-point numerical methods which are extremely reliable and fast, but also verifies convexity of the given problem using rules of disciplined convex programming (Grant et al. 2006). CVXPY was inspired by MATLAB optimization package CVX (Grant and

Boyd 2018) still used in many references, e.g. Cornuéjols et al. (2018). However, in CVXPY (CVXR) the user can easily combine convex optimization and simplicity of Python (R) language together with its high-level features such as object-oriented design or parallelism.

It is worth to mention that unlike our KKT algorithm, both CVXPY and CVXR are able to solve non-negative DOOLSE and MDOOLSE in a general FDSLRLM without the orthogonality condition. In FDSLRLM, interior-points numerical methods have complexity  $n^3$  for one iteration or  $\log(1/\varepsilon)$  times bigger for the complete computation with a precision  $\varepsilon$  of the required optimal solution (Boyd and Vandenberghe 2009).

### LMM computational tools for FDSLRLM

In our previous paper (Gajdoš et al. 2017), we stated that no current package in R is directly and effectively suitable for FDSLRLM. Thanks to a detailed study of (Demidenko 2013; Galecki and Burzykowski 2013) we were successfully directed and instructed how to implement the FDSLRLM variance structure into one of the best R packages for LMM known as *nlme* (Pinheiro and Bates 2009; Pinheiro et al 2018). After that, inspired by *nlme*, we found another R package called *sommer*, also suitable for FDSLRLM fitting (Covarrubias-Pazarán 2016; Cornuéjols et al. 2018).

Simultaneously, thanks to online computing environment *CoCalc* for SageMath (Stein and others 2019) with the possibility to run computations and codes for free in many other programming languages and open softwares, we are also able to run and test MATLAB function *mixed* (Witkovský 2018, 2012) primarily intended for estimating variances in LMMs. On the basis of successful tests, we wrote its R version called *MMEinR*, which is also available in our GitHub Repository (Gajdoš 2019).

The mentioned LMM packages can also handle (RE)MLE in the general FDSLRLM without orthogonality or full-rankness restriction. The packages use iterative methods based on EM algorithm or Henderson MME together with some version of the Newton-Raphson method whose complexity is generally at least  $n^3$ .

We are fully aware that another efficient implementation for estimating variances in LMMs provides *lme4* package (Bates et al. 2015) or SAS package *PROC MIXED* (Stroup et al. 2018; SAS Institute Inc. 2018). As for *lme4*, which is much faster than *nlme*, we found that the package does not allow implementing FDSLRLM covariance structure using *lme4* standard input procedure. As SAS laymans, we also tried a university edition of SAS, free from 2014, but we left it after running into initial problems with Windows installation process in a virtual machine environment and subsequently with a more sophisticated programming language than Python or R. Therefore our knowledge with SAS remained only at the theoretical level.

Finally, we point out that thanks to SageMath and our very fast KKT algorithm, we were able to compute (in real time) NN-(M)-DOOLSE in the toy example with infinite precision—as the exact closed-form (algebraic) numbers. To get an explicit idea, e.g. for our toy model, we got these results for NN-MDOOLSE  $\tilde{\mathbf{v}}$  identical with REMLE

$$\tilde{\mathbf{v}} = \begin{pmatrix} \tilde{\sigma}_0^2 \\ \tilde{\sigma}_1^2 \\ \tilde{\sigma}_2^2 \\ \tilde{\sigma}_3^2 \\ \sigma_4^2 \end{pmatrix} = \begin{pmatrix} -\frac{6569}{4080} \sqrt{3}\sqrt{2} - \frac{13207}{4080} \sqrt{3} - \frac{22533}{6800} \sqrt{2} + \frac{312727}{20400} \\ \frac{6569}{48960} \sqrt{3}\sqrt{2} + \frac{275509}{244800} \sqrt{3} + \frac{7511}{27200} \sqrt{2} + \frac{50461}{244800} \\ \frac{6569}{48960} \sqrt{3}\sqrt{2} + \frac{37687}{48960} \sqrt{3} + \frac{7511}{27200} \sqrt{2} - \frac{93427}{244800} \\ \frac{6569}{48960} \sqrt{3}\sqrt{2} + \frac{13207}{48960} \sqrt{3} + \frac{11081}{27200} \sqrt{2} - \frac{66779}{61200} \\ \frac{6569}{48960} \sqrt{3}\sqrt{2} + \frac{13207}{48960} \sqrt{3} + \frac{43637}{48960} \sqrt{2} - \frac{17377}{61200} \end{pmatrix} \approx \begin{pmatrix} 1.093 \\ 2.875 \\ 1.671 \\ 0.281 \\ 1.772 \end{pmatrix}.$$

It means that we can compute *real* numerical errors in results for all used computational tools to explore their quality from the viewpoint of numerical precision.

*A.4 The BLUP-NE method — proofs*

*Existence of inversions for  $\mathbf{U}_v, \mathbf{H}_v$*

To prove nonsingularity of  $\mathbf{U}_v, \mathbf{H}_v$ , it is sufficient to show that both matrices have non-zero determinants. Using idempotence of  $\mathbf{M}_F$  and determinant expression (3.17), we can write for determinants of  $\mathbf{U}_v, \mathbf{H}_v$

$$\det \mathbf{H}_v = \det(\sigma_0^2 \mathbf{I}_l + \mathbf{V}' \mathbf{V} \mathbf{D}_v) = (-\sigma_0^2)^{l-n} \det(\sigma_0^2 \mathbf{I}_l + \mathbf{V} \mathbf{D}_v \mathbf{V}'),$$

$$\det \mathbf{U}_v = \det(\sigma_0^2 \mathbf{I}_l + \mathbf{V}' \mathbf{M}_F \mathbf{M}_F \mathbf{V} \mathbf{D}_v) = (-\sigma_0^2)^{l-n} \det(\sigma_0^2 \mathbf{I}_l + \mathbf{M}_F \mathbf{V} \mathbf{D}_v \mathbf{V}' \mathbf{M}_F).$$

Now we can see that the sum of positive definite matrix  $\sigma_0^2 \mathbf{I}_l$  ( $\sigma_0^2 > 0$ ) and each of positive semidefinite matrices  $\mathbf{V} \mathbf{D}_v \mathbf{V}'$ ,  $\mathbf{M}_F \mathbf{V} \mathbf{D}_v \mathbf{V}' \mathbf{M}_F$  is a positive definite matrix, whose determinant is always positive.

**Proof (Theorem 2)** Employing Lemma 1 and the well-known standard expressions for mean values and covariances of invariant quadratic estimators [see e.g. Christensen (2011), Puntanen et al. (2013)]  $E_v \{ \mathbf{X}' \mathbf{A} \mathbf{X} \} = \text{tr}(\mathbf{A} \boldsymbol{\Sigma}_v)$  and if  $\mathbf{X} \sim \mathcal{N}_n(\mathbf{F} \boldsymbol{\beta}, \boldsymbol{\Sigma}_v)$  then  $Cov_v \{ \mathbf{X}' \mathbf{A} \mathbf{X}, \mathbf{X}' \mathbf{B} \mathbf{X} \} = 2 \text{tr}(\mathbf{A} \boldsymbol{\Sigma}_v \mathbf{B} \boldsymbol{\Sigma}_v)$ , we have

$$(1) \quad E_v \{ \hat{\sigma}_j^2 \} = E(\mathbf{X} \boldsymbol{\tau}_j \boldsymbol{\tau}_j' \mathbf{X}) = \text{tr}(\boldsymbol{\tau}_j \boldsymbol{\tau}_j' \boldsymbol{\Sigma}_v) = \text{tr}(\boldsymbol{\tau}_j' \boldsymbol{\Sigma}_v \boldsymbol{\tau}_j) = (\mathbf{T}_v \boldsymbol{\Sigma}_v \mathbf{T}_v')_{jj}.$$

According to (3) of Lema 1  $(\mathbf{T}_v \boldsymbol{\Sigma}_v \mathbf{T}_v')_{jj} = \mathbf{D}_{vjj} - \sigma^2 (\mathbf{W}_v^{-1})_{jj} = \sigma_j^2 - \sigma^2 (\mathbf{W}_v^{-1})_{jj}$ .

$$(2) \quad D_v \{ \hat{\sigma}_j^2 \} = 2 \text{tr}(\boldsymbol{\tau}_j \boldsymbol{\tau}_j' \boldsymbol{\Sigma}_v \boldsymbol{\tau}_j \boldsymbol{\tau}_j' \boldsymbol{\Sigma}_v) = 2 \text{tr}(\boldsymbol{\tau}_j' \boldsymbol{\Sigma}_v \boldsymbol{\tau}_j \boldsymbol{\tau}_j' \boldsymbol{\Sigma}_v \boldsymbol{\tau}_j) = 2(\mathbf{T}_v \boldsymbol{\Sigma}_v \mathbf{T}_v')_{jj}$$

As a consequence of (1)  $D_v \{ \hat{\sigma}_j^2 \} = 2(\sigma_j^2 - \sigma^2 (\mathbf{W}_v^{-1})_{jj})^2$ .

(3) In a similar way as in (2)

$$Cov_v \{ \hat{\sigma}_i^2, \hat{\sigma}_j^2 \} = 2 \text{tr}(\boldsymbol{\tau}_i \boldsymbol{\tau}_i' \boldsymbol{\Sigma}_v \boldsymbol{\tau}_j \boldsymbol{\tau}_j' \boldsymbol{\Sigma}_v) = 2 \text{tr}(\boldsymbol{\tau}_i' \boldsymbol{\Sigma}_v \boldsymbol{\tau}_j \boldsymbol{\tau}_j' \boldsymbol{\Sigma}_v \boldsymbol{\tau}_i) = 2(\boldsymbol{\tau}_i' \boldsymbol{\Sigma}_v \boldsymbol{\tau}_j)^2$$

$$= 2(\mathbf{T}_v \boldsymbol{\Sigma}_v \mathbf{T}_v')_{ij}^2 = 2(0 - \sigma^2 (\mathbf{W}_v^{-1})_{ij})^2 = 2(\sigma^2 (\mathbf{W}_v^{-1})_{ij})^2.$$

$$\begin{aligned}
 (4) \quad MSE_{\mathbf{v}}\{\hat{\sigma}_j^2\} &= E_{\mathbf{v}} \left\{ (\hat{\sigma}_j^2 - \sigma_j^2)^2 \right\} = (E_{\mathbf{v}} \left\{ \hat{\sigma}_j^2 \right\} - \sigma_j^2)^2 + D_{\mathbf{v}}\{\hat{\sigma}_j^2\} \\
 &= (\sigma_j^2 - \sigma^2(\mathbf{W}_{\mathbf{v}}^{-1})_{jj} - \sigma_j^2)^2 + 2(\sigma_j^2 - \sigma^2(\mathbf{W}_{\mathbf{v}}^{-1})_{jj})^2 \\
 &= \sigma^4(\mathbf{W}_{\mathbf{v}}^{-1})_{jj}^2 + 2(\sigma_j^2 - \sigma^2(\mathbf{W}_{\mathbf{v}}^{-1})_{jj})^2. \quad \square
 \end{aligned}$$

## References



- Agrawal A, Verschueren R, Diamond S, Boyd S (2018) A rewriting system for convex optimization problems. *J Control Decis* 5(1):42–60
- Amemiya T (1977) A note on a heteroscedastic model. *J Econom* 6(3):365–370
- Anderson T (1970) Estimation of covariance matrices which are linear combinations or whose inverses are linear combinations of given matrices. In: Bose RC, Roy SN (eds) *Essays in probability and statistics*. University of North Carolina Press, Chapel Hill, pp 1–24
- Bates D, Mächler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using lme4. *J Stat Softw* 67(1):1–48
- Beezer RA, Bradshaw R, Grout J, Stein WA (2013) Sage. In: Hogben L (ed) *Handbook of linear algebra*, 2nd edn. Chapman and Hall/CRC, Boca Raton, pp 91–1–91–26
- Bertsekas DP (2009) *Convex optimization theory*, 1st edn. Athena Scientific, Belmont
- Box GEP, Jenkins GM, Reinsel GC, Ljung GM (2015) *Time series analysis: forecasting and control*, 5th edn. Wiley, Hoboken
- Boyd S, Vandenberghe L (2009) *Convex optimization*, 7th edn. Cambridge University Press, Cambridge
- Boyd S, Vandenberghe L (2018) *Introduction to applied linear algebra: vectors, matrices, and least squares*. Cambridge University Press, Cambridge
- Briuelle L, De Feo L, Doliskani J, Flori JP, Schost É (2019) Computing isomorphisms and embeddings of finite fields. *Math Comp* 88(317):1391–1426
- Brockwell PJ, Davis RA (1991) *Time series: theory and methods*. Springer, New York
- Brockwell PJ, Davis RA (2009) *Time series: theory and methods*. Springer, New York
- Brockwell PJ, Davis RA (2016) *Introduction to time series and forecasting*, 3rd edn. Springer, New York
- Christensen R (2011) *Plane answers to complex questions: the theory of linear models*, 4th edn. Springer, New York
- Cornuéjols G, Peña J, Tütüncü R (2018) *Optimization methods in finance*. Cambridge University Press, Cambridge
- Covarrubias-Pazarán G (2016) Genome-assisted prediction of quantitative traits using the R package sommer. *PLoS ONE* 11(6):e0156
- Demidenko E (2013) *Mixed models: theory and applications with R*, 2nd edn. Wiley, Hoboken
- Diamond S, Boyd S (2016) CVXPY: a python-embedded modeling language for convex optimization. *J Mach Learn Res* 17(83):1–5
- Frederickson B (2019) Ranking Programming Languages by GitHub Users. <https://www.benfrederickson.com/ranking-programming-languages-by-github-users/>. Accessed 19 Feb 2020
- Fu A, Narasimhan B, Diamond S, Miller J, Boyd S, Rosenfield PK (2019) CVXR: disciplined Convex Optimization. <https://CRAN.R-project.org/package=CVXR>. Accessed 19 Feb 2020
- Gajdoš A (2019) MMEinR: R-version of Witkovský's MATLAB mixed function. <https://github.com/fdslrm/MMEinR>
- Gajdoš A, Hančová M, Hanč J (2017) Kriging methodology and its development in forecasting econometric time series. *Statistika* 97(1):59–73
- Gajdoš A, Hanč J, Hančová M (2019a) fdslrm. <https://github.com/fdslrm>. Accessed 19 Feb 2020
- Gajdoš A, Hanč J, Hančová M (2019b) fdslrm: applications. <https://github.com/fdslrm/applications>. Accessed 19 Feb 2020
- Gajdoš A, Hanč J, Hančová M (2019c) fdslrm: EBLUP-NE. <https://github.com/fdslrm/EBLUP-NE>. Accessed 19 Feb 2020
- Gajdoš A, Hanč J, Hančová M (2019d) fdslrm: R package. <https://github.com/fdslrm/R-package>. Accessed 19 Feb 2020
- Galántai A (2004) *Projectors and projection methods*. Springer, New York
- Galecki A, Burzykowski T (2013) *Linear mixed-effects models using R: a step-by-step approach*. Springer, New York

- Ghosh M (1996) On the nonexistence of nonnegative unbiased estimators of variance components. *Indian J Stat Ser B* (1960–2002) 58(3):360–362
- Grant M, Boyd S (2018) CVX: matlab software for disciplined convex programming. <http://cvxr.com/cvx/>. Accessed 19 Feb 2020
- Grant M, Boyd S, Ye Y (2006) Disciplined convex programming. In: Liberti L, Maculan N (eds) *Global optimization: from theory to implementation*. Springer, New York
- Hančová M (2007) Comparison of prediction quality of the best linear unbiased predictors in time series linear regression models. In: *Proceedings of 15th European Young Statisticians Meeting*, University of Extremadura, Castro Urdiales (Spain). <https://github.com/fdsrlm/EBLUP-NE>. Accessed 19 Feb 2020
- Hančová M (2008) Natural estimation of variances in a general finite discrete spectrum linear regression model. *Metrika* 67(3):265–276
- Harville DA (1974) Bayesian inference for variance components using only error contrasts. *Biometrika* 61(2):383–385
- Harville DA (1977) Maximum likelihood approaches to variance component estimation and to related problems. *J Am Stat Assoc* 72(358):320–338
- Henderson CR, Kempthorne O, Searle SR, von Krosigk CM (1959) The Estimation of environmental and genetic trends from records subject to culling. *Biometrics* 15(2):192–218
- Hyndman RJ, Athanasopoulos G (2018) *Forecasting: principles and practice*, 2nd edn. Monash University, OTexts
- Jiang J (2007) *Linear and generalized linear mixed models and their applications*. Springer, New York
- Jones E, Oliphant T, Peterson P et al (2001) SciPy: Open source scientific tools for Python. <http://www.scipy.org/>. Accessed 19 Feb 2020
- Kedem B, Fokianos K (2005) *Regression models for time series analysis*. Wiley, New York
- Kluyver T, Ragan-Kelley B, Perez F, Granger B, Bussonnier M, Frederic J, Kelley K, Hamrick J, Grout J, Corlay S, Ivanov P, Avila D, Abdalla S, Willing C (2016) Jupyter Notebooks—a publishing format for reproducible computational workflows. In: Loizides F, Schmidt B (eds) *Positioning and power in academic publishing: players, agents and agendas*. *Proceedings of the 20th ELPUB*. Ios Press, Amsterdam, pp 87–90
- Koenker R, Mizera I (2014) Convex optimization in R. *J Stat Softw* 60(5):1–23
- Kreiss JP, Lahiri SN (2012) Bootstrap methods for time series. In: Rao TS, Rao SS, Rao CR (eds) *Handbook of statistics, time series analysis: methods and applications*, vol 30. Elsevier, Amsterdam, pp 3–26
- LaMotte LR (2007) A direct derivation of the REML likelihood function. *Stat Pap* 48(2):321–327
- Lima A, Rossi L, Musolesi M (2014) Coding together at scale: GitHub as a collaborative social network. In: Adar E, Resnick P, Choudhury MD, Hogan B, Oh AH (eds) *Proceedings of the 8th ICWSM*, Ann Arbor, June 1–4, 2014. The AAAI Press, Palo Alto
- McLeod AI, Yu H, Mahdi E (2012) Time series with R. In: Rao TS, Rao SS, Rao CR (eds) *Time series analysis: methods and applications*, vol 30. Elsevier, Amsterdam, pp 661–712
- Oliphant TE (2007) *Python for scientific computing*. *Comput Sci Eng* 9(3):10–20
- Percival DB, Walden AT (2009) *Spectral analysis for physical applications: multitaper and conventional univariate techniques*. Cambridge University Press, Cambridge
- Pinheiro J, Bates D (2009) *Mixed-effects models in S and S-PLUS*. Springer, New York
- Pinheiro JC, Bates D, DebRoy S, Sarkar D, EISPACK authors, Heiserkamp S, Van Willigen B, R-core (2018) nlme: linear and nonlinear mixed effects models. <https://CRAN.R-project.org/package=nlme>. Accessed 19 Feb 2020
- Pourahmadi M (2001) *Foundations of time series analysis and prediction theory*. Wiley, New York
- Priestley M (2004) *Spectral analysis and time series*. Elsevier Academic Press, Amsterdam
- Project Jupyter, Bussonnier M, Forde J, Freeman J, Granger B, Head T, Holdgraf C, Kelley K, Nalvarte G, Osheroff A, Pacer M, Panda Y, Perez F, Ragan-Kelley B, Willing C (2018) Binder 2.0—reproducible, interactive, sharable environments for science at scale. *Proceedings of the 17th Python in Science Conference*, pp 113–120
- Puntanen S, Styan GPH, Isotalo J (2013) *Formulas useful for linear regression analysis and related matrix theory*. SpringerBriefs in statistics. Springer, Berlin
- R Development Core Team (2019) R: a language and environment for statistical computing. <http://www.r-project.org/>. Accessed 19 Feb 2020
- Rao CR, Kleffe J (1988) *Estimation of variance components and applications*. North-Holland, Amsterdam
- Rao JNK, Molina I (2015) *Small area estimation*, 2nd edn. Wiley, Hoboken
- SAS Institute Inc (2018) *SAS/STAT 15.1 user's guide: the MIXED procedure*. SAS Institute Inc., Cary

- Searle SR, Khuri AI (2017) Matrix algebra useful for statistics, 2nd edn. Wiley, Hoboken
- Searle SR, Casella G, McCulloch CE (2009) Variance components. Wiley, Hoboken
- Shumway RH, Stoffer DS (2017) Time series analysis and its applications: with R examples, 4th edn. Springer, New York
- Singer JM, Rocha FMM, Nobre JS (2017) Graphical Tools for detecting departures from linear mixed model assumptions and some remedial measures. *Int Stat Rev* 85(2):290–324
- Sokol P, Gajdoš A (2018) Prediction of attacks against honeynet based on time series modeling. In: Silhavy R, Silhavy P, Prokopova Z (eds) Applied computational intelligence and mathematical methods. Advances in intelligent systems and computing. Springer, New York, pp 360–371
- Stein WA et al (2019) Sage Mathematics Software—SageMath. <http://www.sagemath.org>. Accessed 19 Feb 2020
- Stroup WW, Milliken GA, Claassen EA, Wolfinger RD (2018) SAS for mixed models: introduction and basic applications. SAS Institute, Cary
- Štulajter F (1991) Consistency of linear and quadratic least squares estimators in regression models with covariance stationary errors. *Appl Math-Czech* 36(2):149–155
- Štulajter F (2002) Predictions in time series using regression models. Springer, New York
- Štulajter F (2003) The MSE of the BLUP in a finite discrete spectrum LRM. *Tatra Mt Math Publ* 26(1):125–131
- Štulajter F, Witkovský V (2004) Estimation of variances in orthogonal finite discrete spectrum linear regression models. *Metrika* 60(2):105–118
- Weiss CJ (2017) Scientific computing for chemists: an undergraduate course in simulations, data processing, and visualization. *J Chem Educ* 94(5):592–597
- Witkovský V (2012) Estimation, Testing, and prediction regions of the fixed and random effects by solving the Henderson's mixed model equations. *Meas Sci Rev* 12(6):234–248
- Witkovský V (2018) mixed—File Exchange—MATLAB Central. <https://www.mathworks.com/matlabcentral/fileexchange/200>. Accessed 19 Feb 2020
- Wu WB, Xiao H (2012) Covariance matrix estimation in time series. In: Rao TS, Rao SS, Rao CR (eds) Handbook of statistics, time series analysis: methods and applications, vol 30. North-Holland, Elsevier, Amsterdam, pp 187–209
- Żądło T (2009) On MSE of EBLUP. *Stat Pap* 50(1):101–118
- Zhang F (2005) The Schur complement and its applications. Springer, New York
- Zimmermann P, Casamayou A, Cohen N, Connan G, Dumont T, Fousse L, Maltey F, Meulien M, Mezzarobba M, Pernet C, Thiéry NM, Bray E, Cremona J, Forests M, Ghitza A, Thomas H (2018) Computational mathematics with SageMath. SIAM, Philadelphia
- Zwiernik P, Uhler C, Richards D (2017) Maximum likelihood estimation for linear Gaussian covariance models. *J R Stat Soc Series B Stat Methodol* 79(4):1269–1292

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Affiliations

Martina Hančová<sup>1</sup>  · Andrej Gajdoš<sup>1</sup> · Jozef Hanč<sup>2</sup>  · Gabriela Vozáriková<sup>3</sup>

✉ Martina Hančová  
martina.hancova@upjs.sk

<sup>1</sup> Institute of Mathematics, Pavol Jozef Šafárik University, Košice, Slovakia

<sup>2</sup> Institute of Physics, Pavol Jozef Šafárik University, Košice, Slovakia

<sup>3</sup> Institute of Computer Science, Pavol Jozef Šafárik University, Košice, Slovakia