Kriging Methodology and Its Development in Forecasting Econometric Time Series

Andrej Gajdoš [1] | *Pavol Jozef Šafárik University in Košice, Slovak Republic*
Martina Hančová [2] | *Pavol Jozef Šafárik University in Košice, Slovak Republic*
Jozef Hanč [3] | *Pavol Jozef Šafárik University in Košice, Slovak Republic*

Abstract

One of the approaches for forecasting future values of a time series or unknown spatial data is kriging. The main objective of the paper is to introduce a general scheme of kriging in forecasting econometric time series using a family of linear regression time series models (shortly named as FDSLRM) which apply regression not only to a trend but also to a random component of the observed time series. Simultaneously performing a Monte Carlo simulation study with a real electricity consumption dataset in the R computational language and environment, we investigate the well-known problem of "negative" estimates of variance components when kriging predictions fail. Our following theoretical analysis, including also the modern apparatus of advanced multivariate statistics, gives us the formulation and proof of a general theorem about the explicit form of moments (up to sixth order) for a Gaussian time series observation. This result provides a basis for further theoretical and computational research in the kriging methodology development.[4]

Keywords	***JEL code***
Forecasting models, linear regression models, best linear unbiased prediction, approximation of mean squared error, moments of random vectors	*C10, C53, C59, C60, Q47*

INTRODUCTION

Data of many economic, financial, insurance or business variables can be generally considered as time series datasets – sets of observations tracking the same type of information at multiple points in time. Modern time-series econometrics, representing an interconnection of mathematical, statistical and computer methods, allows us to model, forecast, interpret and describe various real phenomena dealing with these types of data (Andersen et al., 2009; Box et al., 2008; Brockwell and Davis, 2006; Cipra, 2013; Enders, 2014; Tsay, 2010). Last twenty-five years brought notable advances in the time-series econometrics (Escobari, Ngo, 2014) and moreover, its applications and tools led to several Nobel Prize Awards

[1] Faculty of Science, Department of Economic and Financial Mathematics, Jesenná 5, 04001 Košice, Slovak Republic. Corresponding author: e-mail: andrej.gajdos@student.upjs.sk, phone: (+421)904002040.
[2] Faculty of Science, Department of Economic and Financial Mathematics, Jesenná 5, 04001 Košice, Slovak Republic. E-mail: martina.hancova@upjs.sk.
[3] Faculty of Science, Institute of Physics, Park Angelinum 9, 04001 Košice, Slovak Republic.
[4] This article is based on contribution at the conference *Robust 2016*.

in Economics – namely R. S. Shiller, E. F. Fama and L. P. Larsen in 2013, Ch. A. Sims in 2011 as well as R. F. Engle III and C. W. J. Granger in 2003.

One of the most important areas of time series theory application is the forecasting which solves a task how to predict future values of a time series from its current and past values (Hyndman and Athanasopoulos, 2014). From a practical point of view, information obtained by forecasting provides a crucial knowledge for effective and efficient planning or decision making. The Box-Jenkins methodology (Box et al., 2008; Cipra, 2013; Enders, 2014; Tsay, 2010), belonging to the most popular methodologies for modeling and forecasting econometric time series data (see current real econometric applications e.g. in Pošta, Pikhart, 2012; Salamaga, 2015; Šimpach, 2015) is based fundamentally on ARMA, ARIMA models or their vector counterparts (VARMA, VARIMA). But there exist other advanced and powerful forecasting alternatives such as exponential smoothing methods (Cipra, 2013; Hyndman and Athanasopoulos, 2014), neural networks models (Andersen et al., 2009; Crone et al., 2011; Fomby and Terrell, 2006), linear regression models (Brockwell and Davis, 2006; Chatterjee and Hadi, 2012; Cipra, 2013; Enders, 2014; Štulajter, 2002) or dynamic regression models (Pankratz, 1991; Shumway and Stoffer, 2011).

The prediction theory using linear regression models called kriging (Christensen, 2001; Cressie and Wikle, 2011; Moore, 2001; Stein, 1999; Štulajter, 2002) represents a process of finding the optimal linear prediction for random processes or random fields. The process is based on modeling in an appropriate general class of linear regression models where the following analytical or numerical optimization finds out the best unbiased linear predictor (BLUP) on a set of all linear unbiased predictors. The optimization criterion is a minimization of the mean squared error (MSE) among considered predictors.

Finally it is worth to mention that although the kriging was originally developed for predictions in spatial data (geostatistics and meteorology, Cressie, 1993), the idea of the BLUP brings fruitful results in a much broader set of problems (Harville, 2008; Murphy, 2012; Rao and Molina, 2015; Robinson, 1991), e.g. small-area estimation in economics, the prediction of breeding values in genetics, the estimation of treatment contrasts (e.g. in drug development, agriculture or manufacturing), the analysis of longitudinal data, insurance credibility theory, noise removing from images or machine learning.

Our paper deals with an application of kriging in forecasting econometric time series. Since kriging is not well-known in econometric journals and literature, the first section summarizes a general framework how the kriging methodology works. To not be distracted by many technical details and to focus on main ideas, we illustrate each step of kriging using a real econometric time series dataset dealing with electricity consumption, and reducing the number of used formulas as much as possible (an interested reader will find explicit references for all data and formulas). The illustrative example brings us naturally to a problem of a kriging failure when standard computational methods dealing with considered estimates of nonnegative variance parameters give us negative values. The second section continues in this generally well-known estimation problem. Here we numerically manifest the practical commonness (non-rareness) of this situation by a simulation study numerically quantifying a relative occurrence of explored cases. In the final third section of the paper, we analyze the mentioned problem in the broader context of theoretical developments in kriging methodology using appropriate advanced methods of multivariate statistics. Our analysis results in the formulation and proof of a general theorem about the explicit form of moments for a Gaussian time series observation.

As for numerical calculations, we carried out our computational research producing final results (tables and figures) of the paper in the R statistical computing language (<https://www.r-project.org>; Chambers, 2008; R Development Core Team, 2016) in a powerful integrated development environment called RStudio (<https://www.rstudio.com>; Verzani, 2011). At present, the free and open source R computational environment rapidly improves its capabilities (now there are almost 10 000 statistical packages) which R ranks as one of the best statistical tools for the high-quality computational time series research (McLeod et al., 2012).

1 FORECASTING TIME SERIES USING KRIGING

Forecasting time series within the framework of kriging consists of the following stages (Christensen, 2001; Stein, 1999; Štulajter, 2002): (i) selecting sufficiently general and broad class of linear regression models; (ii) obtaining an empirical realization of a given time series and its modeling; (iii) choosing predicted time series values and finding the BLUP for them; (iv) estimating model parameters on which the BLUP depends and using empirical ("plug-in") BLUP; and finally (v) exploring the impact of the estimation on properties, especially mean squared error, of the BLUP. Let us briefly illustrate this scheme in the case of a real econometric dataset which also brings us naturally to our research problem.

1.1 The first stage – a general class of models

As we mentioned above, in the first stage of kriging we select some general class of linear regression models. In our research, we are concerned with the so-called finite discrete spectrum linear regression models (FDSLRM) – a class of time series models whose mean values (trend) are given by linear regression and random components (error terms) are a linear combination of uncorrelated zero-mean random variables and white noise, which together can be interpreted in terms of finite discrete spectrum (Priestley, 2004).

This parametric family of time series models, a direct extension of classical regression models with many practical applications, was introduced in 2002–2003 by Štulajter (2002, 2003). Especially the monograph from 2002 focusing on forecasting econometric time series in terms of kriging has started a mathematical and statistical research of FDSLRM dealing with its properties and applications (Hančová, 2008, 2011; Hančová et al., 2015; Harman and Štulajter, 2010; Štulajter, 2007; Štulajter and Witkovský, 2004). The exact formal definition of FDSLRM is the following:

A model of time series $X(.)$ is said to be the finite discrete spectrum linear regression model (FDSLRM) iff $X(.)$ satisfies:

$$X(t) = \sum_{i=1}^{k} \beta_i f_i(t) + \sum_{j=1}^{l} Y_j v_j(t) + w(t);\ t \in \mathcal{T}, \tag{1}$$

where:

$\mathcal{T}$ representing a time domain is a countable subset of the real line $\mathbb{E}^1$,

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

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

$\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_l)'$ is a random vector with $E(\boldsymbol{Y}) = 0$ and with covariance matrix $Cov(\boldsymbol{Y}) = diag\{\sigma_j^2\}$ of size $l \times l$, where $\sigma_j^2 \geq 0, j = 1,2, \ldots, l$,

$f_i(.);\ i \in \{1, 2, \ldots, k\}$ and $v_j(.);\ j \in \{1, 2, \ldots, l\}$ are real functions defined on $\mathbb{E}^1$,

$w(.)$ is a white noise uncorrelated with $\boldsymbol{Y}$ and with dispersion $D[w(t)] = \sigma^2 > 0$.

In FDSLRM applications (Štulajter, 2003, 2007; Štulajter and Witkovský, 2004) the most frequently considered time domain set $\mathcal{T}$ is the set of natural numbers $\mathbb{N} = \{1,2, \ldots\}$.

For further considerations, we remind one of basic properties of the FDSLRM (Štulajter, 2003), which says that a finite FDSLRM observation $\boldsymbol{X} = \big(X(1), X(2), \ldots, X(n)\big)'$, $n \in \mathbb{N}$ satisfies a linear mixed model (LMM) of the form:

$$\boldsymbol{X} = \mathrm{F}\boldsymbol{\beta} + \mathrm{V}\boldsymbol{Y} + \boldsymbol{w} \text{ with } E(\boldsymbol{w}) = \boldsymbol{0}, Cov(\boldsymbol{w}) = \sigma^2 \mathrm{I}_n, Cov(\boldsymbol{Y}, \boldsymbol{w}) = \boldsymbol{0}, \tag{2}$$

where matrices F (size $n \times k$) and V (size $n \times l$) are known design matrices given by values of functions $f_i(.), v_j(.)$ for times $t = 1,2,\ldots,n$ and $\boldsymbol{w} = (w(1),\ldots,w(n))'$ stands for a finite n-dimensional white noise observation. In the language of LMM terminology $\boldsymbol{\beta}$ would represent the k-vector of fixed effects and the random component would depend on l-vector $\boldsymbol{Y}$ of random effects and n-vector $\boldsymbol{w}$ of random errors. This fundamental FDSLRM property allows us to apply many results and mathematical techniques of LMM methodology (e.g. Demidenko, 2013; McCulloch et al., 2008; Searle et al., 2006; Witkovský, 2012).

The last remark in our formal introduction of FDSLRM deals with the variance parameters of $\boldsymbol{Y}$ and $w(.)$. It is common to describe the parameters by one vector $\boldsymbol{\nu} = (\sigma^2, \sigma_1^2, \ldots, \sigma_l^2)'$, so $\boldsymbol{\nu}$ becomes an element of the parametric space $\Upsilon = (0,\infty) \times [0,\infty)^l$. Because of several practical or theoretical reasons (similar as in the case of LMM; see Remark 1), it is common to work only with a restricted space $\Upsilon^* = (0,\infty)^{l+1}$.

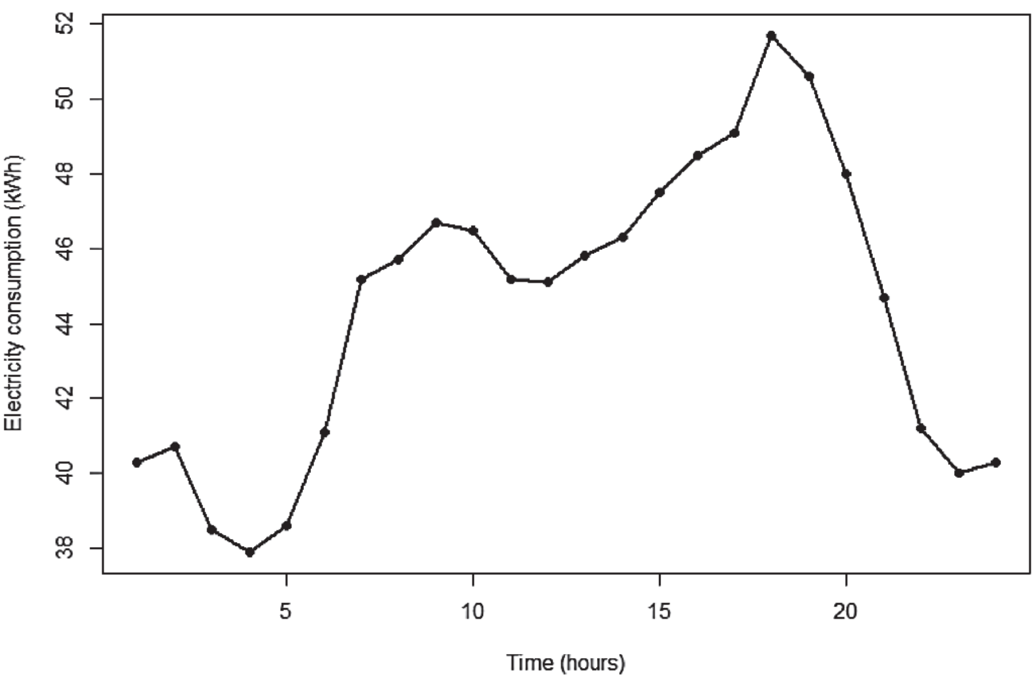
Remark 1 (Parametric space for variance parameters $\boldsymbol{\nu}$*)*

One practical reason for working only with the restricted space Υ^* is that any zero variance σ_j^2 implies almost sure zero random component Y_j (i.e. Y_j has a degenerate distribution with $P(Y_j = 0) = 1$), which in practice means ignoring this component in the model (1). Another research reason is to avoid technical or numerical problems dealing with zero variances in developing theory. However, there can be considerable interest not to reduce Υ, e.g. in testing (e.g. testing for overdispersion, where we would carry out testing for zero variance components and dropping them from the model) or in guaranteeing the existence of estimators and predictors (e.g. in some cases, estimates of $\boldsymbol{\nu}$ based on least-square minimization or likelihood maximization exist only in space Υ, but not in restricted space Υ^*).

1.2 The second stage – time series data

In the second stage of kriging we observe an empirical realization of finitely many values $\boldsymbol{X} \equiv (X(1), X(2), \ldots, X(n))'$ of time series $X(.)$. As a real data example, we use a microeconomic time series dataset (Figure 1) from Štulajter and Witkovský (2004).

Figure 1 Time series data of electricity consumption during 24 hours in a department store



Source: Authors' figure created in R software (R Development Core Team, 2016; real data from the table of Example 4.1, p.116, Štulajter, Witkovský, 2004; the dataset as a text file are available at: <https://goo.gl/tiJjvr>.)

This time series can be modeled by an adequate Gaussian FDSLRM, if we employ generally applicable empirical considerations commonly used in economics and business (Štulajter, 2002; Štulajter and Witkovský, 2004). First of all, economic or business data often show some periodic (seasonal) patterns as they are influenced by seasons or regularly repeating events. To identify significant frequencies describing the periodic behavior, we apply spectral time series analysis (Andersen et al., 2009; Brockwell and Davis, 2006; Priestley, 2004; Štulajter, 2002). Generally, there are more than one frequency. Lower frequencies appear in the trend, and higher frequencies are included in the random component. According to the periodogram,[5] the main tool of the spectral analysis, there are three most significant Fourier frequencies $\lambda_1 = 2\pi/24, \lambda_2 = 2\pi/8, \lambda_3 = 2\pi/6$. Considering and checking all mentioned facts in same way as in Štulajter and Witkovský (2004), we get the following FDSLRM (1) with $k = 3, l = 4$ for the explored consumption dataset:

$$X(t) = \beta_1 + \beta_2 \cos \lambda_1 t + \beta_3 \sin \lambda_1 t + Y_1 \cos \lambda_2 t + Y_2 \sin \lambda_2 t \\ + Y_3 \cos \lambda_3 t + Y_4 \sin \lambda_3 t + w(t)\,; t \in \{1, 2, \ldots, 24\}. \tag{3}$$

1.3 The third stage – the BLUP for a chosen future value

As for the third kriging stage, finding the BLUP, in this case, is straightforward. Mathematically, model (3) represents an orthogonal version of FDSLRM (Štulajter, 2003) for which exists a closed analytic form of the BLUP (theorem 2.1 in Štulajter, 2003, p. 129) for any future value $X(n + d), d \in \mathbb{N}$. This form denoted by $X^*(n + d)$ generally depends on variance parameters $\boldsymbol{\nu}$:

$$X^*(n + d) \equiv X^*_{\boldsymbol{\nu}}(n + d).$$

1.4 The fourth stage – estimation of models parameters and use of the EBLUP

In practical situations like this one, we need to estimate regression parameters $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)' \in \mathbb{E}^3$ and variance parameters $\boldsymbol{\nu} = (\sigma^2, \sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2)' \in \Upsilon^* = (0, \infty)^5$. Various standard and also nonstandard mathematical techniques for estimating $\boldsymbol{\beta}$ and $\boldsymbol{\nu}$ can be found in above mentioned references dealing with kriging and FDSLRM (Christensen, 2001; Hančová, 2008; Štulajter, 2002; Štulajter and Witkovský, 2004), but also in references dealing with the methodology of LMMs (e.g. Rao and Molina, 2015; Searle et al., 2006).

With regard to statistical properties, we remind some general results from the estimation theory. Standardly used estimators of $\boldsymbol{\beta}$ in linear regression models like least squares estimators (ordinary – OLSE or weighted – WELSE) or maximum likelihood estimators (MLE or REMLE) are linear with respect to a time series observation $\boldsymbol{X}$. For our FDSLRM (3), OLSE of the regression parameters $\boldsymbol{\beta}$ give us[6] $\widehat{\boldsymbol{\beta}} = (44.38, -3.15, -3.52)$.

In connection with estimating $\boldsymbol{\nu}$, standard least-squares methods of estimation in FDSLRM in many cases lead to quadratic estimators which are invariant quadratic forms[7] in $\boldsymbol{X}$. Variance parameters can be estimated e.g. by double ordinary least squares estimators (DOOLSE) or by their modified unbiased version (MDOOLSE) as it is described in Remark 2.

Remark 2 (DOOLSE and MDOOLSE)

The double least squares method is based on two following steps. First of all, we find OLSE $\widehat{\boldsymbol{\beta}}$ for $\boldsymbol{\beta}$, then we can compute empirical residuals $\widehat{\boldsymbol{\varepsilon}} \equiv \boldsymbol{X} - \mathrm{F}\widehat{\boldsymbol{\beta}}$. Then matrix $\widehat{\boldsymbol{\varepsilon}}\widehat{\boldsymbol{\varepsilon}}' = (\boldsymbol{X} - \mathrm{F}\widehat{\boldsymbol{\beta}})(\boldsymbol{X} - \mathrm{F}\widehat{\boldsymbol{\beta}})'$ represents

[5] The periodogram can be computed in the base R package e.g. by function `spec.pgram{}`.

[6] In the R environment, OLSE can be found via function `lm{}` in the base R package.

[7] It means that estimators of variances can be written as $\boldsymbol{X}'\mathrm{A}\boldsymbol{X}$, where A is some $n \times n$ real symmetric matrix and values of $\boldsymbol{X}'\mathrm{A}\boldsymbol{X}$ do not depend on $\boldsymbol{\beta}$. In FDSLRM, it is equivalent with the condition AF = 0.

the well-known estimation matrix S($\boldsymbol{X}$) for a covariance matrix Σ of $\boldsymbol{X}$ which is equal to $Cov(\boldsymbol{X}) \equiv \Sigma_\nu = \sigma^2 I_n + VDV'$, $D = diag\{\sigma_j^2\}$. Using ordinary least squares method a second time, i.e. minimizing the square of a norm (distance) between Σ_ν and S($\boldsymbol{X}$) with respect of $\boldsymbol{\nu} \in \Upsilon$, we find OLSE estimates $\hat{\boldsymbol{\nu}}$ which are called DOOLSE. As for calculation methods, two approaches are usually applied: multi variable calculus and geometrical projection theory, since the least squares problems can be expressed in terms of orthogonal projections.

Requiring unbiasedness, we can modify DOOLSE to unbiased estimators (MDOOLSE). The exact formal definition of these type of estimators in linear regression models can be found e.g. in Štulajter (2002, p. 25). Moreover, it can be shown (Štulajter, Witkovský, 2004) that in any orthogonal Gaussian FDSLRM as it is in our case (3), DOOLSE are identical with maximum likelihood estimators (MLE) and MDOOLSE are equal to restricted MLE (REMLE).

If we apply corresponding DOOLSE/MDOOLSE formulas based on the geometrical theory of orthogonal projectors (Štulajter and Witkovský; 2004, p. 107, last two formulas) in our FDSLRM, we get the following results:

projectors for DOOLSE $\quad \hat{\boldsymbol{\nu}} = (3.00, 0.12, 1.61, -0.24, 1.02) \notin \Upsilon^*$
projectors for MDOOLSE $\quad \tilde{\boldsymbol{\nu}} = (3.53, 0.08, 1.57, -0.29, 0.97) \notin \Upsilon^*$

We see that in both cases projection formulas for the standard estimation methods fail. Variance components can never be negative. However, it is very important to realize that the DOOLSE/MDOOLSE estimates can be based on the projection method only if the method provides values for $\boldsymbol{\nu}$ belonging to the parametric space Υ^* or Υ. Therefore in our case we had to use other methods of computation, e.g. numerical iterative methods (Štulajter, 2002), which indicate that DOOLSE give us an estimate of $\boldsymbol{\nu}$ with zero component $\tilde{\sigma}_3^2$ lying on the boundary of Υ or no estimate of $\boldsymbol{\nu}$, if we consider restricted space Υ^*. What action should be done in this situation?

In the framework of LMM, such estimation problem with negative or zero values of estimates for variance components has a long and rich history (Searle et al., 2006). It is a well-known problem at least 40 years, especially in using ANOVA, MLE and REMLE estimators for LMMs. Inspiring by section 4.4 in Searle at al. (2006, p.130), there are several possibilities how to solve it, if we speak about FDSLRM: (i) understand it as a consequence of insufficient data and collect more time series data; (ii) accept zero estimates and ignore the zero variance components in the model, if it is reasonable; (iii) interpret negative or zero results as indication of a wrong model and build a new, but still adequate FDSLRM model for considered data; (iv) use a modified or new method of estimation leading to positive estimates. In the case of FDSLRM, only last two possibilities were already studied.

Building a new adequate FDSLRM model was done in Štulajter and Witkovský (2004). During the spectral analysis authors replaced the third most significant Fourier frequency $2\pi/6$ by the fourth one $2\pi/12$. However, simulation results of the next section will show that this approach is not fully satisfactory since it does not work in relatively frequent circumstances.

The last above-mentioned solution (iv) is to use new estimators with always positive or almost sure positive values. Such new estimators also based on least squares (Remark 3), called natural estimators (NE), were proposed and studied in Hančová (2008). Statistically these estimators are biased invariant quadratic forms.

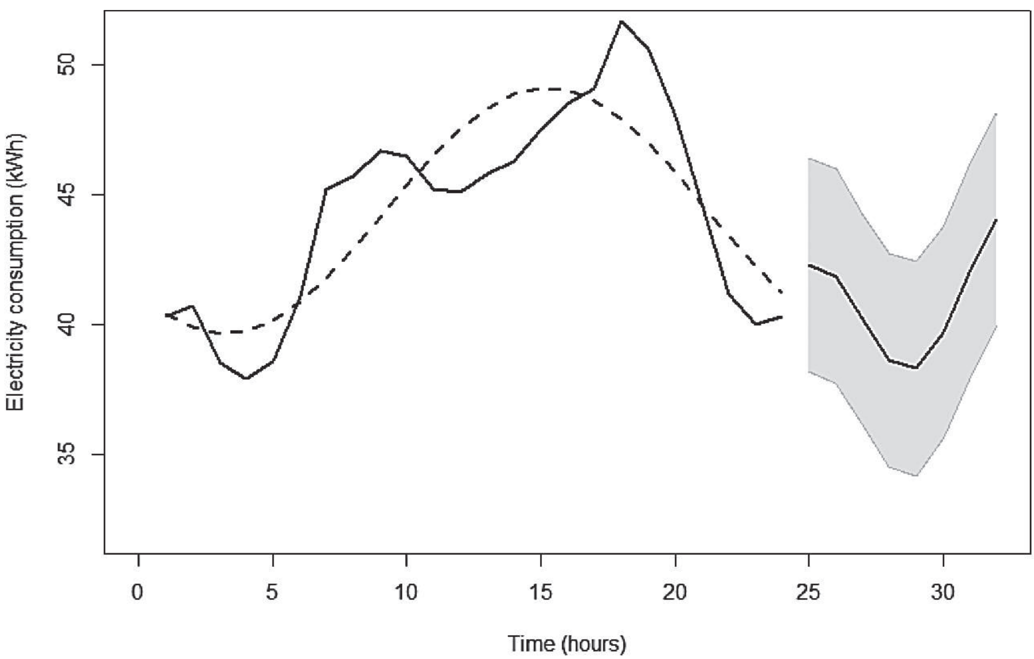
Remark 3 (Invariant quadratic biased NE)

The main idea behind NE comes from the fact that $\sigma_j^2 = E(Y_j^2), j = 1, \ldots, l.$ Then, it is reasonable to estimate an l-vector of unobservable realization $\boldsymbol{y}$ of $\boldsymbol{Y}$ in model (2) by ordinary least squares and use squares of these estimates $\tilde{y}_j^2, j = 1, \ldots, l$, as estimators of σ_j^2 (see more details and the exact formal definition in section 2.1, Hančová, 2008). Geometrically, these so-called natural estimators can be expressed by oblique projectors.

If we are interested in non-negative and simultaneously unbiased estimators, then generally there do not exist any non-negative unbiased estimators for $\sigma_j^2, j = 1, \dots, l.$ Twenty years ago, Ghosh (1996) formulated an elegant proof of two general important results connected with incompatibility between non-negativity and unbiasedness of random effects estimators in LMM: (i) in any LMM with $\boldsymbol{\nu} \in \Upsilon = (0,\infty) \times [0,\infty)^l$ and $\boldsymbol{X}$ having an absolutely continuous probability distribution with respect to some σ-finite measure, if $v_j^*(\boldsymbol{X})$ is an unbiased estimator of σ_j^2, then there is always non-zero probability for $v_j^*(\boldsymbol{X})$ to be negative with respect to some $\boldsymbol{\beta}$, $\boldsymbol{\nu}$; (ii) the same is true, $P\{v_j^*(\boldsymbol{X}) < 0 | \boldsymbol{\beta}, \boldsymbol{\nu}\} > 0$ for some $\boldsymbol{\beta}$ and $\boldsymbol{\nu}$, if we suppose $\boldsymbol{\nu} \in \Upsilon^* = (0,\infty)^{l+1}$ and $\boldsymbol{X}$ having a probability density function continuous in all $\boldsymbol{\nu} \in \Upsilon^*$.

Computing NE numerically (Hančová, 2008, p. 268, formulas 2.2, 2.3), we get $\breve{\boldsymbol{\nu}} = (3.53, 0.37, 1.86, 0.004, 1.27) \in \Upsilon$. In practice these estimates are suitable for computation of BLUPs (Štulajter, 2003 the first formula on p. 129) for future values $X(n+d), d \in \mathbb{N}$. These „plug-in" BLUPs are called empirical BLUPs (EBLUPs). At the same time NE estimates can be used for computation of corresponding „plug-in" MSEs (Štulajter, 2003, the second formula on p. 129) and 95% prediction intervals,[8] which are commonly used in displaying the uncertainty in time series forecasting (Hyndman and Athanasopoulos, 2014). Figure 2 is a summary graphical representation of obtained predictions for electricity consumption during the next eight hours.

Figure 2 Kriging forecasting of the electricity consumption for next 8 hours with 95% forecast intervals (solid line in the gray shaded region) and the time series trend (dashed line)



Source: Authors' figure based on their calculations, created in R software (R Development Core Team, 2016)

[8] A formula for $\mathcal{I}_{0,95}(n+d) = [X^*(n+d) - 1{,}96\sqrt{MSE\{X^*(n+d)\}}; X^*(n+d) + 1{,}96\sqrt{MSE\{X^*(n+d)\}}]$.

The plug-in step replacing true value of parameters (for which the BLUP was derived) by NE causes a certain deviation in the mean squared error. Therefore, the main task of the last fifth stage of kriging is to study statistical properties of EBLUPs based on NE which are invariant quadratic biased estimators. However, such research has not yet been made. So, the third section of the paper (after the simulation study) will be devoted to our analysis of the last stage of kriging using NE in the broader context of theoretical developments in kriging methodology.

2 SIMULATION STUDY

To answer our research question, how efficiently building a new FDSLRM can solve problems with negative values of projectors for standard variance estimates (DOOLSE, MDOOLSE), we planned four possible FDSLRM simulation designs whose structure can be based on three significant frequencies chosen by Štulajter and Witkovský (2004): $\lambda_1 = 2\pi/24, \lambda_2 = 2\pi/8, \lambda_3 = 2\pi/12$. The considered designs differ with respect to possible number $m \in \{0,1,2,3\}$ of given frequencies included in the FDSLRM trend (remaining 3-m frequencies are in the random component – shortly RC). Due to easier, more compatible notation with spectral analysis, we wrote their forms by the following compact formula ($m \in \{0,1,2,3\}$):

$$X_m(t) = \alpha + \sum_{i=1}^{m} (\beta_i \cos \lambda_i t + \gamma_i \sin \lambda_i t) + \sum_{j=1}^{3-m} \left(Y_j \cos \lambda_j t + Z_j \sin \lambda_j t\right) + w(t), \tag{4}$$

for $m = 1$ we get the identical model with the original one applied by Štulajter and Witkovský (2004). OLSE for regression parameters $\alpha, \beta_i, \gamma_j$ and NE for variance parameters calculated from the real dataset (Figure 1) were assigned as true parameters for simulation designs (Table 1). We also mention that in this case, NE values are evidently nonzero and they are also close to DOOLSE and MDOOLSE for the dataset.

Table 1 Vectors of regression and variance parameters for considered model designs

Model design	Regression parameters $\boldsymbol{\beta}$	Variance parameters $\boldsymbol{\nu}$
$m = 0$ (3 frequencies in RC)	(44.38)′	(1.09, 9.93, 12.43, 2.97, 1.76, 0.37, 1.86)′
$m = 1$ (2 frequencies in RC)	(44.38, –3.15, –3.52)′	(1.09, 2.97, 1.76, 0.37, 1.86)′
$m = 2$ (1 frequency in RC)	(44.38, –3.15, –3.52, –1.72, –1.33)′	(1.09, 0.37, 1.86)′
$m = 3$ (0 frequencies in RC)	(44.38, –3.15, –3.52, –1.72, –1.33, 0.61, 1.36)′	(1.09)′

Source: Authors' calculations based on real data from Štulajter, Witkovský (2004) using R (R Development Core Team, 2016)

Using R, we simulated N = 5 000 time series realizations for each design (values of $Y_j, Z_j, w(t)$ were generated from normal distributions with zero means and variance parameters given by Table 1). Then for each realization (a time series dataset) estimates via corresponding orthogonal or oblique projectors (for DOOLSE, MDOOLSE and NE) were computed and simultaneously a relative occurrence of the projections with negative values was counted. Complete results dealing with a relative occurrence of negative values in the four evaluated simulation designs are reported in Table 2 (NE are not included since they really led only to positive estimates).

As for distributions, Figure 3 presents typical results in the form of histograms for projectors dealing with MDOOLSE and NE (as examples) in the case of simulation design $m = 2$. Table 2 clearly manifests that in all designs projection methods for computing estimations (DOOLSE = MLE, MDOOLSE = REMLE) give

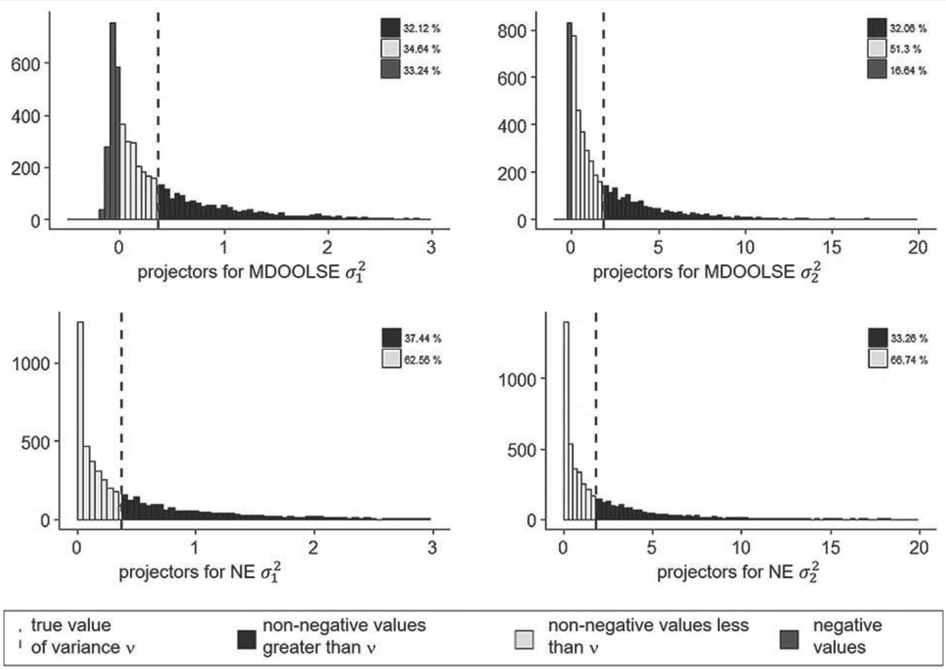
Table 2 Relative occurrence of negative values (as results of projection methods) for estimation of variance parameters $\boldsymbol{\nu}$ in N = 5 000 simulated replications for each model design

Model design	Projectors for estimators	Relative occurrence of negative values for:					
m = 0 (3 frequencies in RC)	DOOLSE	6.48 %	6.76 %	13.44 %	17.16 %	32.68 %	16.98 %
	MDOOLSE	6.62 %	6.92 %	13.84 %	17.64 %	33.60 %	17.30 %
m = 1 (2 frequencies in RC)	DOOLSE	13.16 %	16.36 %	32.10 %	15.54 %	x	x
	MDOOLSE	14.14 %	17.82 %	34.56 %	16.60 %	x	x
m = 2 (1 frequency in RC)	DOOLSE	29.54 %	14.66 %	x	x	x	x
	MDOOLSE	33.24 %	16.64 %	x	x	x	x
m = 3 (0 frequencies in RC)	DOOLSE	x	x	x	x	x	x
	MDOOLSE	x	x	x	x	x	x

Source: Authors' simulation results based on parameters in Table 1 using R (R Development Core Team, 2016)

us from 6% up to 35% of wrong negative values for estimates which are not small, insignificant numbers. If we are interested in actual values of the estimators, alternative numerical methods (Štulajter 2002; also applied by us in R) show that zero values of DOOLSE = MLE and MDOOLSE = REMLE for $\boldsymbol{\nu}$ correspond to the negative values of projectors. Since $\boldsymbol{\nu}$ has only nonzero components, from a theoretical point of view, zero estimates also mean a failure. Therefore, we can conclude that changing or rebuilding the model (3) as it was done by Štulajter and Witkovský (2004) will not work in relatively frequent circumstances.

Figure 3 Typical results of the simulation study showing relative occurrence of values (as results of projection methods) for estimation of variance parameters in model design m = 2



Source: Authors' figure based on their calculations, created in R software (R Development Core Team, 2016)

3 KRIGING IN PRACTICE – MSE OF EMPIRICAL BLUPS

During the fourth stage of kriging, we saw that in the next stage, there is a need to study effects of estimating unknown variance parameters of FDSLRM on statistical properties of BLUPs, if these true parameters (to avoid any misunderstanding in this section we denote them $\boldsymbol{\nu}_T$ instead of $\boldsymbol{\nu}$ are replaced by variance estimates $\check{\boldsymbol{\nu}}$ (e.g. invariant quadratic biased NE). Special attention must be paid to MSE of EBLUPs which determines not only quality of the obtained empirical predictors but allows us to express corresponding forecast intervals or to test statistical hypotheses.

However, general theory of empirical linear unbiased predictors (Harville, 2008; Štulajter, 2002; Witkovský, 2012; Żądło, 2009) says that explicit expression for the MSE of an empirical predictor (EBLUP) is not known. One of the reasons why it is still open research problem consists of the fact that EBLUP is a nonlinear function of the observation $\boldsymbol{X}$. Therefore, finding such expression is a very difficult mathematical task. On the other hand, the theory gives us an approximation for the correction (adjustment) of MSE of EBLUPs with respect to the original BLUP using Taylor's series (see e.g. Harville, 2008; or Štulajter, 2002):

$$E\left[X^*_{\check{\boldsymbol{\nu}}}(n+d) - X^*_{\boldsymbol{\nu}_T}(n+d)\right]^2 \approx E\left[\frac{\partial X^*_{\boldsymbol{\nu}}(n+d)}{\partial \boldsymbol{\nu}'}\Big|_{\boldsymbol{\nu}=\boldsymbol{\nu}_T} \cdot (\check{\boldsymbol{\nu}} - \boldsymbol{\nu}_T)\right]^2 = \\ = \sum_{a,b=0}^{l} E\left[\frac{\partial X^*_{\boldsymbol{\nu}}(n+d)}{\partial \nu_a}(\check{\nu}_a - \nu_a)(\check{\nu}_b - \nu_b)\frac{\partial X^*_{\boldsymbol{\nu}}(n+d)}{\partial \nu_b}\right] \tag{5}$$

Although we know explicit forms of $\check{\nu}_a, a = 0,1,\dots,l$ for NE (Hančová, 2008), the direct use of these quadratic forms in theoretical and corresponding computational study of the approximation (5) would lead to cumbersome and uselessly complicated mathematical work. In this case, more abstract and general approach paradoxically makes the problem more tractable and understandable, stripping away non-essential features. In addition, such generalization allows us to use a new arsenal of mathematical techniques.

Therefore mathematically, it is more useful to describe NE $\check{\nu}_a$ in the approximation (5) only as invariant quadratic biased estimators of the general form $Q_A \equiv \boldsymbol{X}'\mathrm{A}\boldsymbol{X}, \mathrm{AF} = 0$. Partial derivatives $\partial X^*_{\boldsymbol{\nu}}(n+d)/\partial \nu_a$ have the general form $\boldsymbol{c}'\boldsymbol{X} + d,\ \boldsymbol{c}' \in \mathbb{E}^n, d \in \mathbb{E}^1$. If we introduce a concept of the so-called parameter centered quadratic form $Q^*_A \equiv Q_A - \nu_a = \boldsymbol{X}'\mathrm{A}\boldsymbol{X} - \nu_a = \tilde{\nu}_a - \nu_a,$ then it is easy to see that the approximation (5) depends on expressions such as $E(Q^*_A), E(\boldsymbol{X}Q^*_A\boldsymbol{X}'), E(Q^*_A Q^*_B), E(\boldsymbol{X}Q^*_A Q^*_B), E(\boldsymbol{X}Q^*_A Q^*_B \boldsymbol{X}')$.

These moments are up to sixth order with respect to $\boldsymbol{X}$. However, as our next theoretical results demonstrate, if the finite time series observation $\boldsymbol{X}$ (model (2)) comes from Gaussian FDSLRM (1) and consequently has a multivariate normal distribution $\boldsymbol{X} \sim N(\mathrm{F}\boldsymbol{\beta}, \Sigma)$ with the positive definite covariance $n \times n$ matrix Σ ($\Sigma > 0$), then all moments up to sixth order can be expressed as functions depending only on the second-order (not higher) properties of $\boldsymbol{X}$ given by mean value parameters $\boldsymbol{\beta}$ and variance parameters $\boldsymbol{\nu}$. Under the assumption of normality for $\boldsymbol{X}$, using the modern algebraic apparatus of advanced multivariate statistics (Ghazal and Neudecker, 2000; Kollo and Rosen, 2005) which includes vectorization, commutation matrices, the Kronecker product and relations among them, we derived the explicit form of mentioned expressions. Our results are summarized by the following general theorem which contains the moments for any invariant quadratic biased estimators (NE are a special case). Due to higher mathematical sophistication and technicalities, its proof is explained in the Appendix.

Theorem (the explicit form of moments)

Let a random vector $\boldsymbol{X} \sim N(\mathrm{F}\boldsymbol{\beta}, \Sigma)$ be a given finite observation of time series, where $\mathrm{F} \in \mathbb{E}^{n\times k}, \boldsymbol{\beta} \in \mathbb{E}^k$ and $\Sigma \in \mathbb{E}^{n\times n}, \Sigma > 0$. Let $Q_A \equiv \boldsymbol{X}'\mathrm{A}\boldsymbol{X}, Q_B \equiv \boldsymbol{X}'\mathrm{B}\boldsymbol{X}, \mathrm{A}, \mathrm{B} \in \mathbb{E}^{n\times n}$ be the invariant quadratic forms, i.e. $\mathrm{AF} = \mathrm{BF} = 0, E(Q_A) = \mathrm{tr}(\mathrm{A}\Sigma), E(Q_B) = \mathrm{tr}(\mathrm{B}\Sigma)$ and $Cov(Q_A, Q_B) = 2\mathrm{tr}(\mathrm{A}\Sigma\mathrm{B}\Sigma)$. Then for parameter-centered quadratic forms $Q^*_A \equiv Q_A - \nu_A$ and $Q^*_B \equiv Q_B - \nu_B$; $\nu_A, \nu_B \in \mathbb{E}^1$ the following properties hold:

(i) $E(Q_{\mathrm{A}}^{*}) = E(Q_{\mathrm{A}}) - \nu_{\mathrm{A}}, E(\boldsymbol{X}Q_{\mathrm{A}}^{*}) = E(Q_{\mathrm{A}}^{*})\mathrm{F}\boldsymbol{\beta}$,

(ii) $E(\boldsymbol{X}Q_{\mathrm{A}}^{*}\boldsymbol{X}') = E(Q_{\mathrm{A}}^{*})[\Sigma + \mathrm{F}\boldsymbol{\beta}(\mathrm{F}\boldsymbol{\beta})'] + 2\Sigma\mathrm{A}\Sigma$,

(iii) $E(Q_{\mathrm{A}}^{*}Q_{\mathrm{B}}^{*}) = Cov(Q_{\mathrm{A}}, Q_{\mathrm{B}}) + E(Q_{\mathrm{A}}^{*})E(Q_{\mathrm{B}}^{*})$,

(iv) $E(\boldsymbol{X}Q_{\mathrm{A}}^{*}Q_{\mathrm{B}}^{*}) = E(Q_{\mathrm{A}}^{*}Q_{\mathrm{B}}^{*})\mathrm{F}\boldsymbol{\beta}$,

(v) $$E(\boldsymbol{X}Q_{\mathrm{A}}^{*}Q_{\mathrm{B}}^{*}\boldsymbol{X}') = 2\Sigma[E(Q_{\mathrm{A}}^{*})\mathrm{B} + E(Q_{\mathrm{B}}^{*})\mathrm{A} + 2\mathrm{A}\Sigma\mathrm{B} + 2\mathrm{B}\Sigma\mathrm{A}]\Sigma + \\ + E(Q_{A}^{*}Q_{B}^{*})[\Sigma + \mathrm{F}\boldsymbol{\beta}(\mathrm{F}\boldsymbol{\beta})'].$$

CONCLUSIONS AND FURTHER DEVELOPMENTS

One of the most important areas of time series theory application is forecasting time series providing a crucial knowledge for effective and efficient planning or decision making. In the paper, we have presented a general framework of forecasting methodology for econometric time series, called kriging. Our kriging application deals with a recently introduced family of linear regression time series models named FDSLRM, which apply regression not only to a trend, but also to a random component of the observed time series.

Using a real data example dealing with electricity consumption, we have also investigated one of the current research problems of kriging – a problem of negative or zero estimates which leads to kriging failures in empirical prediction. Performing a simulation study, we manifested that this problem occurs in relatively frequent circumstances and therefore cannot be neglected. Simultaneously we pointed out inadequacy of rebuilding the model as used problem solution. If computational methods using a dataset of time series observation give failing negative or zero values for standard estimates, then we can apply one of possible solutions – using alternative estimators like natural estimators (NE) which are invariant quadratic biased estimators.

Our consequent analysis in the broader context of kriging methodology developments allowed us to derive explicitly moments of a finite Gaussian time series observation for any invariant quadratic biased estimators of time series variances. Confronting with other research, we have found that our theoretical results were a direct extension of the results of the previous research (Prasad and Rao, 1990; Srivastava and Tiwari, 1976). In comparison with these references, our use of the matrix approach of advanced modern multivariate statistics in proving our results seems more elegant and conceptually simpler than the original cumbersome multiple use of sums with many indices.

As for further research and kriging developments, these moments will allow a theoretical study of properties of empirical predictors and corresponding approximations of MSE based on any invariant quadratic biased estimators (e.g. according to Harville, 2008; Štulajter, 2007). Since our results are written in the recurrent matrix form, they are also very suitable for checking or conducting an effective computational research (statistical computing environments like R are essentially matrix algebra processors) with real empirical data using simulations or bootstrap methods for time series and kriging (Kreiss and Lahiri, 2012; Schelin and Sjöstedt-de Luna, 2010; Sjöstedt-de Luna and Young, 2003). Such computational research could also be applied to study effects of MLE and REMLE, in general FDSLRM not expressible in a closed analytic form, on statistical properties of BLUPs and their MSEs.

Our last conclusion deals with a corresponding implementation of FDSLRM in R. Although any finite FDSLRM observation satisfies a linear mixed model (LMM), according to our inspection it seems that no current package in R for LMM methodology[9] is directly suitable for FDSLRM. Therefore, one

[9] There are many packages in R fitting various forms of LMM, e.g. amer, gamm, glmmAK, lme4.0, lmec, lmm, MASS, MCMCglmm, nlme or PSM (more details in Galecki and Burzykowski, 2013).

of the tasks of future computational FDSLRM research is to create a fully functioning R package using the current object-oriented programming. We also assume that the O-O programming approach which is now standard in the context of statistical modeling (Galecki and Burzykowski, 2013) allows us to use some classes of objects and methods operating on them from existing R packages for LMM.

ACKNOWLEDGMENTS

Our research has been supported by the Scientific Grant Agency of the Slovak Republic (VEGA) – grants VEGA 1/0073/15, VEGA 1/0344/14 and the Internal Research Grant System of Faculty of Science, P. J. Šafárik University in Košice (VVGS PF UPJŠ) – project VVGS-PF-2016-72616.

References

ANDERSEN, T. G., DAVIS, R. A., KREISS, J.-P., MIKOSCH, T. V. eds. *Handbook of Financial Time Series.* Berlin: Springer, 2009.

BOX, G. E. P., JENKINS, G. M., REINSEL, G. C. *Time Series Analysis: Forecasting and Control.* 4th Ed. Hoboken, NJ: Wiley, 2008.

BROCKWELL, P. J. AND DAVIS, R. A. *Time Series: Theory and Methods.* 2nd Ed. New York: Springer-Verlag, 2006.

CHAMBERS, J. M. *Software for Data Analysis: Programming with R.* 1st Ed. New York: Springer, 2008.

CHATTERJEE, S. AND HADI, A. S. *Regression Analysis by Example.* 5th Ed. Hoboken, NJ: John Wiley & Sons, 2012.

CHRISTENSEN, R. *Advanced Linear Modeling: Multivariate, Time Series, and Spatial Data; Nonparametric Regression and Response Surface Maximization.* 2nd Ed. New York: Springer, 2001.

CIPRA, T. *Finanční ekonometrie.* 2nd Ed. Prague: Ekopress, 2013.

CRESSIE, N. *Statistics for Spatial Data.* Rev. Ed. New York: Wiley-Interscience, 1993.

CRESSIE, N. AND WIKLE, C. K. *Statistics for Spatio-Temporal Data.* 1st Ed. Hoboken, N.J: Wiley, 2011.

CRONE, S. F., HIBON, M., NIKOLOPOULOS, K. Advances in Forecasting with Neural Networks? Empirical Evidence from the NN3 Competition on Time Series Prediction. *International Journal of Forecasting,* 2011, 27(3), pp. 635–660.

DEMIDENKO, E. *Mixed Models: Theory and Applications with R.* 2nd Ed. Hoboken, NJ: Wiley, 2013.

ENDERS, W. *Applied Econometric Time Series.* 4th Ed. Hoboken, NJ: Wiley, 2014.

ESCOBARI, D. AND NGO, T. Preface: Special Issue on Time Series Econometric Applications in Finance. *American Journal of Economics,* 2014, 4(2A), pp. 0–0.

FOMBY, T. B., TERRELL, D. eds. *Econometric Analysis of Financial and Economic Time Series Part B.* 1st Ed. Book series: Advances in Econometrics, Vol. 20, Oxford: Emerald Group Publishing Limited, 2006.

GALECKI, A. AND BURZYKOWSKI, T. *Linear Mixed-Effects Models Using R: A Step-by-Step Approach*, 2013 Ed. New York, NY: Springer, 2013.

GHAZAL, G. A. AND NEUDECKER, H. On Second-Order and Fourth-Order Moments of Jointly Distributed Random Matrices: A Survey. *Linear Algebra and Its Applications*, 2000, 321(1), pp. 61–93.

GHOSH, M. On the Nonexistence of Nonnegative Unbiased Estimators of Variance Components. *Sankhyā: The Indian Journal of Statistics, Series B (1960–2002)*, 1996, 58(3), pp. 360–362.

HANČOVÁ, M. Empirical Predictors in Finite Discrete Spectrum Linear Regression Models. In: HARMAN et al., eds. *PROBASTAT 2011 – Abstracts*, Bratislava, Slovak Republic: Institute of Measurement Science, Slovak Academy of Science, 2011, pp. 24–25.

HANČOVÁ, M. Natural Estimation of Variances in a General Finite Discrete Spectrum Linear Regression Model. *Metrika*, 2008, 67(3), pp. 265–276.

HANČOVÁ, M., HANČ, J., GAJDOŠ, J. A Simulation Study of Bootstrap Methods for Kriging in Time Series Forecasting. In: WITKOVSKÝ et al., eds. *PROBASTAT 2015 – Abstracts*, Bratislava, Slovak Republic: Institute of Measurement Science, Slovak Academy of Science, 2015, pp. 26–27.

HARMAN, R. AND ŠTULAJTER, F. Optimal Prediction Designs in Finite Discrete Spectrum Linear Regression Models. *Metrika*, 2010, 72(2), pp. 281–294.

HARVILLE, D. A. Accounting for the Estimation of Variances and Covariances in Prediction under a General Linear Model: An Overview. *Tatra Mountains Mathematical Publications*, 2008, 39(1), pp. 1–15.

HYNDMAN, R. J. AND ATHANASOPOULOS, G. *Forecasting: Principles and Practice*. Print Ed. Melbourne, Australia: OTexts, 2014.

KOLLO, T. AND ROSEN, D. VON. *Advanced Multivariate Statistics with Matrices*. Berlin: Springer, 2005.

KREISS, J.-P. AND LAHIRI, S. N. Bootstrap Methods for Time Series. In: RAO, RAO, RAO, eds. *Handbook of Statistics, Vol. 30: Time Series Analysis: Methods and Applications*, Amsterdam: Elsevier, 2012, pp. 3–26.

MCCULLOCH, C. E., SEARLE, S. R., NEUHAUS, J. M. *Generalized, Linear, and Mixed Models*. 2nd Ed. Hoboken, N.J: Wiley-Interscience, 2008.

MCLEOD, A. I., YU, H., MAHDI, E. Time Series with R. In: RAO, RAO, RAO, eds. *Handbook of Statistics, Vol. 30: Time Series Analysis: Methods and Applications*, Amsterdam: Elsevier, 2012, pp. 661–712.
MOORE, M. eds. *Spatial Statistics: Methodological Aspects and Applications.* New York: Springer Science & Business Media, 2001.
MURPHY, K. P. *Machine Learning: A Probabilistic Perspective.* 1st Ed. Cambridge, MA: The MIT Press, 2012.
PANKRATZ, A. *Forecasting with Dynamic Regression Models.* 1st Ed. New York: John Wiley & Sons, 1991.
POŠTA, V. AND PIKHART, Z. The Use of the Sentiment Economic Indicator for GDP Forecasting: Evidence from EU Economies [online]. *Statistika: Statistics and Economy Journal*, 2012, 92(1), pp. 41–55.
PRASAD, N. G. N. AND RAO, J. N. K. The Estimation of the Mean Squared Error of Small-Area Estimators. *Journal of the American Statistical Association*, 1990, 85(409), pp. 163–171.
PRIESTLEY, M. B. *Spectral Analysis and Time Series.* Amsterdam: Elsevier Acad. Press, 2004.
R DEVELOPMENT CORE TEAM. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing, 2016.
RAO, J. N. K. AND MOLINA, I. *Small Area Estimation*. 2nd Ed. Hoboken, New Jersey: Wiley, 2015.
ROBINSON, G. K. That BLUP Is a Good Thing: The Estimation of Random Effects. *Statist. Sci.*, 1991, 6(1), pp. 15–32.
SALAMAGA, M. Testing the Effectiveness of Some Macroeconomic Variables in Stimulating Foreign Trade in the Czech Republic, Hungary, Poland and Slovakia [online]. *Statistika: Statistics and Economy Journal,* 2015, 95(1), pp. 47–59.
SCHELIN, L. AND SJÖSTEDT-DE LUNA, S. Kriging Prediction Intervals Based on Semiparametric Bootstrap. *Mathematical Geosciences,* 2010, 42(8), pp. 985–1000.
SEARLE, S. R., CASELLA, G., MCCULLOCH, C. E. Variance Components. John Wiley & Sons, 2006.
SHUMWAY, R. H. AND STOFFER, D. S. *Time Series Analysis and Its Applications: With R Examples.* 3rd Ed. New York: Springer, 2011.
ŠIMPACH, O. Fertility of Czech Females Could Be Lower than Expected: Trends in Future Development of Age-Specific Fertility Rates up to the Year 2050 [online]. *Statistika: Statistics and Economy Journal,* 2015, 95(1), pp. 19–37.
SJÖSTEDT-DE LUNA, S. AND YOUNG, A. The Bootstrap and Kriging Prediction Intervals. *Scandinavian Journal of Statistics,* 2003, 30(1), pp. 175–192.
SRIVASTAVA, V. K. AND TIWARI, R. Evaluation of Expectations of Products of Stochastic Matrices. *Scandinavian Journal of Statistics,* 1976, 3(3), pp. 135–138.
STEIN, M. L. *Interpolation of Spatial Data: Some Theory for Kriging.* New York: Springer, 1999.
ŠTULAJTER, F. Mean Squared Error of the Empirical Best Linear Unbiased Predictor in an Orthogonal Finite Discrete Spectrum Linear Regression Model. *Metrika,* 2007, 65(3), pp. 331–348.
ŠTULAJTER, F. *Predictions in Time Series Using Regression Models.* New York: Springer, 2002.
ŠTULAJTER, F. The MSE of the BLUP in a Finite Discrete Spectrum LRM. *Tatra Mountains Mathematical Publications,* 2003, 26(1), pp. 125–131.
ŠTULAJTER, F. AND WITKOVSKÝ, V. Estimation of Variances in Orthogonal Finite Discrete Spectrum Linear Regression Models. *Metrika,* 2004, 60(2), pp. 105–118.
TSAY, R. S. *Analysis of Financial Time Series.* 3rd Ed. Cambridge, Mass.: Wiley, 2010.
VERZANI, J. *Getting Started with RStudio.* Sebastopol, Calif.: O'Reilly, 2011.
WITKOVSKÝ, V. Estimation, Testing, and Prediction Regions of the Fixed and Random Effects by Solving the Henderson's Mixed Model Equations. *Measurement Science Review,* 2012, 12(6), pp. 234–248.
ŻĄDŁO, T. On MSE of EBLUP. *Statistical Papers*, 2009, 50(1), pp. 101–118.

APPENDIX: PROOF

Since used arguments are very similar in proofs of all items (i)–(v), we explain ideas of the proof only for the first two items (i), (ii). We achieve the first simplification, when we concentrate on deriving moments $E(Q_{\mathrm{A}}), E(\boldsymbol{X}Q_{\mathrm{A}}), E(\boldsymbol{X}Q_{\mathrm{A}}\boldsymbol{X}')$. The second, essential simplification of the proof arises from introducing the residual vector $\boldsymbol{\varepsilon} \equiv \boldsymbol{X} - E(\boldsymbol{X}) \sim N(0, \Sigma)$, where $\Sigma = E(\boldsymbol{\varepsilon\varepsilon}') = E(\boldsymbol{XX}') - \mathrm{F}\boldsymbol{\beta}(\mathrm{F}\boldsymbol{\beta})'$, using linearity of mean value $E(.)$, invariance of Q_{A} and rewriting considered moments as functions of $\boldsymbol{\varepsilon}$:

$$E(Q_{\mathrm{A}}) = E(\boldsymbol{\varepsilon}'\mathrm{A}\boldsymbol{\varepsilon}), E(\boldsymbol{X}Q_{\mathrm{A}}) = E(\boldsymbol{\varepsilon\varepsilon}'\mathrm{A}\boldsymbol{\varepsilon}) + \mathrm{F}\boldsymbol{\beta}E(\boldsymbol{\varepsilon}'\mathrm{A}\boldsymbol{\varepsilon})$$

$$E(\boldsymbol{X}Q_{\mathrm{A}}\boldsymbol{X}') = E(\boldsymbol{\varepsilon\varepsilon}'\mathrm{A}\boldsymbol{\varepsilon\varepsilon}') + \mathrm{F}\boldsymbol{\beta}E(\boldsymbol{\varepsilon}'\mathrm{A}\boldsymbol{\varepsilon})(\mathrm{F}\boldsymbol{\beta})'.$$

Assumptions of the theorem about Q_{A} give us immediately $E(\boldsymbol{\varepsilon}'\mathrm{A}\boldsymbol{\varepsilon}) = E(Q_{\mathrm{A}}) = \mathrm{tr}\,(\mathrm{A}\Sigma)$.

At this moment, we recall needed expressions and properties of multivariate statistics apparatus (Ghazal, Neudecker, 2000; Kollo, Rosen, 2005):

a) vec $\boldsymbol{a}' = \text{vec}\, \boldsymbol{a} = \boldsymbol{a}$ for any column vector $\boldsymbol{a}$ where vec is defined as follows:
let A be a $m \times n$ matrix and $A_{.j}$ the jth column of A; then vec A is the mn-column vector $\text{vec}\, A = \begin{pmatrix} A_{.1} \\ \vdots \\ A_{.n} \end{pmatrix}$,

b) vec $\boldsymbol{ab}' = \boldsymbol{b} \otimes \boldsymbol{a}$ for any pair of column vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ where $\otimes$ is the Kronecker product (also known as the direct or tensor product) defined in general for arbitrary $k \times l$ matrix A with elements A_{ij} and $m \times n$ matrix B by the formula:

$$A \otimes B = \begin{pmatrix} A_{11}B & \cdots & A_{1l}B \\ \vdots & \ddots & \vdots \\ A_{k1}B & \cdots & A_{kl}B \end{pmatrix},$$

c) vec (ABC) = $(C' \otimes A)\text{vec}\, B = \text{vec}[(C' \otimes A)\text{vec}\, B] = (\text{vec}\, B \otimes I_{mp})'\text{vec}(C' \otimes A)$ for compatible matrices A,B and C, where mp is the row order of $C' \otimes A$,
d) $\text{tr}\,(A'B) = (\text{vec}\, A)'\text{vec}\, B$ for compatible matrices A and B,
e) $K_{mn} := \Sigma_{ij}(E_{ij} \otimes E'_{ij})$, $i \in \{1,2,\ldots, m\}$, $j \in \{1,2,\ldots, n\}$ is called commutation matrix, where Σ_{ij} is a double summation symbol, E_{ij} is a $m \times n$ matrix with a unity in its i,j-th position and zeroes elsewhere,
f) $K_{mn}\, \text{vec}\, A = \text{vec}\, A'$,
g) $(A \otimes B)(C \otimes D) = AC \otimes BD$ for compatible matrices A,B,C and D,
h) $K'_{mn} = K^{-1}_{mn} = K_{nm}$,
i) $K_{m1} = K_{1m} = I_m$,
j) $\text{vec}\,(A \otimes B) = (I_n \otimes K_{qm} \otimes I_p)(\text{vec}\, A \otimes \text{vec}\, B)$ for $m \times n$ matrix A and $p \times q$ matrix B,
k) $K_{rs,m} = (I_r \otimes K_{sm})(K_{rm} \otimes I_s)$.

Now employing property c) we easily conclude about term $E(\boldsymbol{\varepsilon\varepsilon}' A\boldsymbol{\varepsilon})$ that:

$$E(\boldsymbol{\varepsilon\varepsilon}'A\boldsymbol{\varepsilon}) = E(\boldsymbol{\varepsilon}\ \text{vec}\ \boldsymbol{\varepsilon}'A\boldsymbol{\varepsilon}) = E[\boldsymbol{\varepsilon}(\boldsymbol{\varepsilon}' \otimes \boldsymbol{\varepsilon}')\text{vec}\, A] = E[\boldsymbol{\varepsilon}(\boldsymbol{\varepsilon}' \otimes \boldsymbol{\varepsilon}')]\text{vec}\, A\ .$$

An expression for $E[\boldsymbol{\varepsilon}(\boldsymbol{\varepsilon}' \otimes \boldsymbol{\varepsilon}')]$ (corollary 2.2.7.2 (ii) in Kollo, Rosen, 2005, p. 204) and $E(\boldsymbol{\varepsilon}) = 0$ definitively lead to:

$$E(\boldsymbol{\varepsilon\varepsilon}'A\boldsymbol{\varepsilon}) = 0.$$

The most sophisticated part of the proof is the calculation of $E(\boldsymbol{\varepsilon\varepsilon}'A\boldsymbol{\varepsilon\varepsilon}')$. Using properties c) and j), we can write:

$$\text{vec}\ \boldsymbol{\varepsilon\varepsilon}'A\boldsymbol{\varepsilon\varepsilon}' = (\text{vec}'A \otimes I_{n^2})(I_n \otimes K_{nn} \otimes I_n)(\boldsymbol{\varepsilon} \otimes \boldsymbol{\varepsilon} \otimes \boldsymbol{\varepsilon} \otimes \boldsymbol{\varepsilon})\ .$$

Taking the mean value of $\boldsymbol{\varepsilon} \otimes \boldsymbol{\varepsilon} \otimes \boldsymbol{\varepsilon} \otimes \boldsymbol{\varepsilon}$, using c) and the expression for $E[\boldsymbol{\varepsilon}(\boldsymbol{\varepsilon}' \otimes \boldsymbol{\varepsilon}' \otimes \boldsymbol{\varepsilon}')]$ (Corollary 2.2.7.2 (iii) in Kollo, Rosen, 2005, p. 204) we find that:

$$E(\boldsymbol{\varepsilon} \otimes \boldsymbol{\varepsilon} \otimes \boldsymbol{\varepsilon} \otimes \boldsymbol{\varepsilon}) = \text{vec}[\Sigma \otimes \text{vec}'\Sigma + (\text{vec}'\, \Sigma \otimes \Sigma)(I_{n^3} + I_n \otimes K_{nn})\ ]\ .$$

Then three expressions from the last equation need to be treated separately. Using appropriate relations from a)–k) and results from preceding steps, it is possible to show that the following equalities hold:

$(\text{vec}'\,A \otimes I_{n^2})(I_n \otimes K_{nn} \otimes I_n)\text{vec}(\Sigma \otimes \text{vec}'\Sigma) = \text{vec}\,\Sigma A \Sigma,$
$(\text{vec}'\,A \otimes I_{n^2})(I_n \otimes K_{nn} \otimes I_n)\text{vec}(\text{vec}'\,\Sigma \otimes \Sigma) = \text{vec}\,\Sigma A \Sigma,$
$(\text{vec}'\,A \otimes I_{n^2})(I_n \otimes K_{nn} \otimes I_n)\text{vec}[(\text{vec}'\,\Sigma \otimes \Sigma)(I_n \otimes K_{nn})] = \text{tr}(A\Sigma)\text{vec}\,\Sigma.$

All partial results together directly provide the final form for $E(\boldsymbol{\varepsilon\varepsilon}'A\boldsymbol{\varepsilon\varepsilon}')$:

$E(\boldsymbol{\varepsilon\varepsilon}'A\boldsymbol{\varepsilon\varepsilon}') = 2\Sigma A \Sigma + \text{tr}(A\Sigma)\Sigma.$

Combining obtained results for moments $E(Q_A), E(\boldsymbol{X}Q_A), E(\boldsymbol{X}Q_A\boldsymbol{X}')$ with $Q_A^* = Q_A - \nu_A$, we finally get required moments in (i), (ii).