Mathematical Methods In Economics
And Industry 2007
June 3 – 7, 2007, Herl'any, Slovak Republic

The prediction quality in time series forecasting*

Martina Hančová
P. J. Šafárik University in Košice,
Faculty of Science, Institute of Mathematics,
Jesenná 5, 041 54 Košice
martina.hancova@upjs.sk

Abstract

Time series analysis with the prediction theory belong to powerful tools providing important information for decision making not only in the field of the natural or social sciences, but also e.g. in economics, insurance, engineering, telecommunications or traffic. In the contribution we introduce main ideas of the best linear unbiased prediction (BLUP) — one of the most important approaches in the time series forecasting. Simultaneously we apply this method to a general class of time series models called finite discrete spectrum linear regression models (FDSLRM). The derived form of the mean squared error for the BLUP in the most general FDSLRM by us provides the criterion of the prediction quality and can be also used for a useful classification of various FDSLRM models.

Keywords: time series forecasting, prediction quality, best linear unbiased predictor, mean squared error, finite discrete spectrum linear regression model

1 *INTRODUCTION*

The need to obtain sufficiently accurate predictions from observed data can be found not only in all scientific disciplines, but also in many human activities like industry, economics or business. Therefore forecasting future values of a time series belongs to the most important problems of the statistical inference from time series data.

Some of the most popular prediction approaches, the Box-Jenkins methods (Box et al [1]) are based on ARMA, ARIMA and SARIMA models. An alternative theory, also one of the most important approaches in time series theory, using linear regression models is the best linear unbiased prediction theory (see e.g. Brockwell and Davis [2], Stein [9], Christensen [3] or Štulajter [10]).

Historical records acknowledge American econometrician A. S. Goldberger [4] as the first discoverer (1962) of a general form of the best linear unbiased predictor (BLUP) for linear models. At this time BLUP was seen by econometricians primarily as a prediction tool for time series. However mathematically identical approach, but with different name *kriging* was independently suggested approximately at the same time in geostatistics during solving a mining engineering problem how to predict quality of an ore deposit from known sample spatial data. Identical mathematical ideas were also developed in meteorology.

The ideas of best linear prediction and best linear unbiased prediction are very important, because it has important application in standard linear models, mixed models, and the analysis of spatial data. The theory is also a part of multivariate analysis, is significant for general stochastic processes, time series, principal component analysis and is the basis for linear Bayesian methods (see detailed references e.g. in Christensen [3]).

*This work was supported by the grant 1/3001/06 of the Slovak Scientific Grant Agency (VEGA).

In the paper we shall introduce key ideas of the best linear unbiased prediction in the framework of time series theory and our application will concentrate on a recently introduced and investigated class of linear regression time series models called *the finite discrete spectrum linear regression models* (Štulajter et al [11, 12, 13]). In the rest of the introduction we establish notation and recapitulate basic results and used models from mentioned references, which provide a starting point for our further considerations. In kriging we assume that we observe a time series $X(.)$ satisfying a linear regression model (LRM):

$$X(t) = \sum_{i=1}^{k} \beta_i f_i(t) + \varepsilon(t); t = 1, 2, \ldots \tag{1}$$

where

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

$f_1(.), f_2(.), \ldots, f_k(.)$ are given known real functions defined on $\mathbb{E}$; and

$\varepsilon(.)$ is a mean-zero time series ($E\{\varepsilon(t)\} = 0$) with finite covariance functions $\mathrm{Cov}\{\varepsilon(s), \varepsilon(t)\} = R(s,t); s, t = 1, 2, \ldots$

As we mentioned above, FDSLRM models are also linear regression models and are defined in the following way:

Definition 1 *A model of time series $X(.)$ is said to be* the finite discrete spectrum linear regression model *(FDSLRM), if $X(.)$ satisfies*

$$X(t) = \sum_{i=1}^{k} \beta_i f_i(t) + \sum_{j=1}^{l} Y_j v_j(t) + w(t); t = 1, 2, \ldots \tag{2}$$

where

k and l are fixed known nonnegative integers, i.e. $k, l \in \mathbb{N}_0$;

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

$Y = (Y_1, Y_2, \ldots, Y_l)'$ is a $l \times 1$ random vector with zero mean value, $E\{Y\} = 0 \in \mathbb{E}^l$, and with covariance matrix $\mathrm{Cov}\{Y\} = diag(\sigma_j^2)$ *of size $l \times l$, where variances $\sigma_j^2 \geq 0$ for all $j = 1, 2, \ldots, l$;*

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

$w(.)$ is white noise time series with the dispersion $D[w(t)] = \sigma^2 > 0$ and it is uncorrelated with Y.

We denote the unknown variance parameters of Y and $w(.)$, which are also variance parameters of the FDSLRM, by $\nu = (\sigma^2, \sigma_1^2, ..., \sigma_l^2)'$. Under the FDSLRM assumptions direct computation applied to the standard definition of the time series covariance function $R(s,t)$ yields its expression in the form $R_\nu(s,t) = \sigma^2 \delta_{s,t} + \sum_{j=1}^{l} \sigma_j^2 v_j(s) v_j(t); s, t = 1, 2, ...$ with the parameter ν belonging to the parametric space $\Upsilon = (0, \infty) \times \langle 0, \infty)^l$.

The basic result dealing with any finite observation of the FDSLRM time series — random vector $X = (X(1), \ldots, X(n))'$ — says that the observation X satisfies the following linear regression model (also called the FDSLRM model):

$$X = F\beta + \varepsilon, E\{\varepsilon\} = 0, \mathrm{Cov}\{\varepsilon\} = \sigma^2 I_n + \sum_{j=1}^{l} \sigma_j^2 V_j \text{ is a p.d. matrix,} \tag{3}$$

where

$F = (f_1 \ f_2 \ ... \ f_k) \in \mathbb{E}^{n\times k}$ is the design matrix with columns $f_i = (f_i(1), ..., f_i(n))'$; $i = 1, 2, ..., k$;

$V_j = v_j v_j' \in \mathbb{E}^{n\times n}; v_j = (v_j(1), v_j(2), ..., v_j(n))'; j = 1, 2, ..., l$ are matrices describing the structure of covariance matrix $Cov(\varepsilon) \equiv \Sigma_\nu$.

Model (3) is equivalent to a model belonging to the class of linear mixed models (see e.g. McCulloch & Searle [7], Christensen [3])[1]:

$$X = F\beta + VY + w, E\{w\} = 0, \mathrm{Cov}\{w\} = \sigma^2 I_n, \mathrm{Cov}\{Y, w\} = 0, \tag{4}$$

where $V = (v_1 \ v_2 \ ... \ v_l) \in \mathbb{E}^{n\times l}$ and random vector $w = (w(1), ..., w(n))'$ is a finite observation of white noise $w(.)$. Symbols $F, \beta, Y, w(.)$ and $v_j; j = 1, 2, ..., l$ have the same meaning as above. Properties of $R(s,t)$ in FDSLRM gives us also another important formula expressing covariance between observation X and some later value $X(n+d)$, which is used in kriging:

$$r_\nu \equiv Cov(X, X(n+d)) = \sum_{j=1}^{l} \sigma_j^2 v_j(n+d) v_j; t = 1, 2, ... \tag{5}$$

For getting more specific and better idea what FDSLRM really is we summarize various types of FDSLRM in Tab. 1.

Number of parameters	**Type of FDSLRM**
$k = 0,\ l = 0$	white noise (WN)
$k = 0,\ l \geq 1$	finite discrete spectrum models (FDSM)
$k \geq 1,\ l = 0$	classical linear regression models (CLRM)
$k, l \geq 1$	general finite discrete spectrum linear regression models (GFDSLRM)

Table 1: Classification of FDSLRMs with respect to k, l — numbers of model parameters.

[1]In this case unobservable vector β is frequently called a vector of "fixed effects" and Y is an unobservable vector of "random effects".

Two explicit, relatively simple examples of FDSLRM[2] are hourly observed electric consumption at a department store in a typical day described by a time series model:

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

where $\lambda_1 = 2\pi/24, \lambda_2 = 2\pi/12, \lambda_3 = 2\pi/8$ or weekly observed gasoline price[3] in Slovakia (2000):

$$X(t) = \alpha_1 + \alpha_2 t + \sum_{i=1}^{3} \beta_i \cos\lambda_i t + \sum_{j=1}^{3} \gamma_j \sin\lambda_j t + w(t); t = 1, 2, \ldots, 48$$

where $\lambda_1 = 2\pi/48, \lambda_2 = 2\pi/16, \lambda_3 = 2\pi/8$. In connection with Tab. 1 gasoline prices is an example of CLRM and electric consumption an example of GFDSLRM.

In this paper we shall assume that both matrices $F \in \mathbb{E}^{n\times k}$ and $V \in \mathbb{E}^{n\times l}$ are of full rank[4], i.e. $r(F, V) = k + l$ and number $k + l + 1$ of unknown parameters β and ν, which arise in the FDSLRM (2), is smaller than length n of a realization $x = (x_1, x_2, ..., x_n) \in \mathbb{E}^n$ of finite observation X. We shall investigate the most general type of FDSLRM for which $k, l \geq 1$ and $\mathrm{Cov}\{Y\}$ is nonsingular.

Finally it is worth to be aware of the close connection between prediction theory and time series modeling, because one of criteria for a model selection, e.g. what is better? using GFDRLSM or CLRM for gasoline prices?, can result from its influence on predictions of time series data and as we see below it is appropriate to choose such model which gives us "the least possible error" of prediction. So the paper also specify mathematically "the least possible error" and derives its particular and usable form for the class of FDSLRM models.

2 GEOMETRICAL LANGUAGE OF HILBERT SPACES IN STATISTICS

2.1 The Hilbert space of random variables $L^2(\Omega, \mathscr{F}, P)$

Although it is possible to study theory of time series without the geometrical language of Hilbert spaces and an explicit use of related mathematical techniques, there are great advantages to see concepts and algebraic results through eyes of the geometry. Powerful intuition gained from Euclidean spaces $\mathbb{E}^2$ and $\mathbb{E}^3$, the most familiar examples of

[2]Examples were taken from Štulajter [10, 12]), where a reader can find details why and how these models were selected.

[3]The basic idea of model building in case of gasoline prices consists in the following: (1) due to not fully known physical mechanism of the economic phenomenon like a development of gasoline prices, we have to combine the incomplete theoretical knowledge – in this economic time series application a seasonal effect is expected – with (2) the mathematical-empirical approach, where on a basis of the corresponding data graph the scheme of classical decomposition methodology (representing data as a sum of a trend component, a seasonal component and a stationary random noise component) was suggested. (3) According to the spectral theory of time series (generalized Fourier analysis) the season effect can be always modeled by sine and cosine functions. This theory also provides a number of sine and cosine functions and estimations of frequencies. (4) In trend modeling the central role is played by the principle of parsimony – we try to employ the smallest possible number of parameters for adequate regression representation of a data trend. For gasoline prices a linear function is such most parsimonious function which with the modeled seasonal component leads to a residual random component not rejecting the assumption that it can be regarded as a white noise time series – the simplest version of time series.

[4]To have no problems in distinguishing between a matrix product (FV) and $\begin{pmatrix} F & V \end{pmatrix}$ as matrix F augmented by V, we will frequently write the matrix $\begin{pmatrix} F & V \end{pmatrix}$ as (F, V).

Hilbert spaces, allows us "natural" geometrical understanding and connections and makes frequently complicated theoretical results more obvious.

From the theoretical viewpoint the concept of Hilbert space allows us to study simultaneously properties described by linear algebra (e.g. linear independence, dimension, bases), metric properties (e.g. orthogonality, norm, angle) typical for geometry and problems of convergence, continuity or differentiability belonging to mathematical analysis.

Very briefly, a *Hilbert space* $\mathscr{H}$ is a complete inner product space, or a linear space possessing an inner product $(\cdot,\cdot)$ and containing all of its limit points under the norm $\|\cdot\|$ defined in terms of the inner product by the expression $\|\cdot\|^2 = (\cdot,\cdot)$. The detailed explanation of the Hilbert space definition can be found e.g. in Brockwell [2].

Concerning statistics, an important example of a Hilbert space is a Hilbert space $L^2(\Omega, \mathscr{F}, P)$. Notion of this space is based on the following procedure. Consider a collection $\mathscr{H}$ of all random variables with finite second moments ($E\{U^2\} < \infty$) defined on the same probability space $(\Omega, \mathscr{F}, P)$. If we consider an equivalence relation saying that random variables T and U are equivalent (or "equal") if they are equal almost everywhere (Symbolically: $T \equiv U \Leftrightarrow P(T = U) = 1$), then this relation partitions $\mathscr{H}$ into classes of random variables such that any two random variables in the same class are equal with probability 1.

Since each class is uniquely determined by specifying any one of the random variables in it, we shall use the same notation T, U for elements of $\mathscr{H}$ and to call them random variables although it is sometimes worth to remember that U stands for the collection of all random variables which are equivalent to U.

If we define an inner product as $(T, U)_{\mathscr{H}} = E\{TU\}; T, U \in \mathscr{H}$, then it can be shown (Brockwell [2]) that space $\mathscr{H}$ is a Hilbert space. The inner product easily leads to a useful geometrical interpretation of key statistical concepts as it is summarized in Tab.2.

Statistical concept	**Geometrical concept**
mean value of a random variable $E\{U\}$	$(U, 1)_{\mathscr{H}}$
mean value of a product $E\{TU\}$	$(T, U)_{\mathscr{H}}$
mean value of a squared difference $E\{(T-U)^2\}$	$\|T-U\|^2_{\mathscr{H}}$
dispersion $D\{U\} \equiv E\{(U - E\{U\})^2\}$	$\|U - E\{U\}\|^2_{\mathscr{H}}$
standard deviation $\sqrt{D\{U\}}$	$\|U - E\{U\}\|_{\mathscr{H}}$
covariance $\mathrm{Cov}\{T, U\} \equiv E\{(T - E\{T\})(U - E\{U\})\}$	$(T - E\{T\}, U - E\{U\})_{\mathscr{H}}$

Table 2: Key statistical concepts expressed in terms of geometrical concepts – inner product and norm defined in Hilbert space $\mathscr{H} = L^2(\Omega, \mathscr{F}, P)$. The mean value of any random variable from $\mathscr{H}$ is a statistical concept (in the left column), but it can be also considered as a constant random variable - an element of $\mathscr{H}$ (in the right column).

2.2 Geometrical proofs of some basic formulas

We illustrate advantage of this powerful geometrical viewpoint in deriving some basic statistical formulas that we prove using mentioned geometrical concepts.

Mean value. Linearity of inner product says that $(aT+bU,1)_{\mathscr{H}} = a(T,1)_{\mathscr{H}}+b(U,1)_{\mathscr{H}}$; $T, U \in \mathscr{H}$ and $a, b \in \mathbb{E}$, which according to Tab. 2 corresponds to the well-known assertion $E\{aT + bU\} = aE\{T\} + bE\{U\}$. It is worth to notice that for any two constant random variables $t, u \in \mathscr{H} : (t, u)_{\mathscr{H}} = E\{tu\} = tu$.

Dispersion. The definition of the norm and linearity of inner product immediately gives a formula

$$\|T - U\|^2_{\mathscr{H}} = \|T\|^2_{\mathscr{H}} + \|U\|^2_{\mathscr{H}} - 2(T, U)_{\mathscr{H}}. \tag{6}$$

which is nothing else as a general statement of *law of cosines* for two vectors and its difference (or a triangle formed by given vectors), known in $\mathbb{E}^2$ or $\mathbb{E}^3$ as $\|u - v\|^2 = \|u\|^2 + \|v\|^2 - 2\,\|u\|\,\|v\| \cos\theta$. (If u and v are orthogonal, then this reduces to the Pythagorean theorem).

Using fact that $E\{U\}$ is the constant we get according to previous section (see also Tab.2) $\|E\{U\}\|^2_{\mathscr{H}} = (E\{U\}, E\{U\})_{\mathscr{H}} = [E\{U\}]^2$ and $(U, E\{U\})_{\mathscr{H}} = E\{U\}(U, 1)_{\mathscr{H}} = [E\{U\}]^2$. Taking $T = E\{U\}$ in (6) we can write a basic formula for dispersion of any $U \in \mathscr{H}$

$$\|U - E\{U\}\|^2_{\mathscr{H}} = \|U\|^2_{\mathscr{H}} - [E\{U\}]^2 \text{ or } D\{U\} = E\{U^2\} - [E\{U\}]^2. \tag{7}$$

If we are interested in $D\{T - U\}$, then (7) directly yields $D\{T - U\} = E\{(T - U)^2\} - [E\{T - U\}]^2$. But $D\{T - U\}$ can be also expressed as $\|(T - E\{T\}) - (U - E\{U\})\|^2_{\mathscr{H}}$, so use of (6) leads to another basic formula

$$D\{T - U\} = \|T - E\{T\}\|^2_{\mathscr{H}} + \|U - E\{U\}\|^2_{\mathscr{H}} - 2(T - E\{T\}, U - E\{U\})_{\mathscr{H}}$$

or in terms of statistical concepts (Tab.2)

$$D\{T - U\} = D(T) + D(U) - 2\mathrm{Cov}\{T, U\}. \tag{8}$$

3 BEST LINEAR UNBIASED PREDICTION IN TIME SERIES THEORY

3.1 A criterion for prediction quality

Formulation of the prediction problem. Now we briefly formulate the problem of kriging prediction theory, assuming that given time series $X(.)$ satisfies a linear regression model.

We observe finitely many values $X = (X(1), ..., X(n))'$ of a time series $X(.)$. We would like to predict a future value of the time series: $U = X(n + d); d \in \mathbb{N}$ or in other words we find a random variable $\tilde{U} = \tilde{U}(X)$, *predictor*, based on observation X in such way that this predictor $\tilde{U}$ of U is the best in some sense. What should be effective criterion of the prediction quality of such predictor?

Heuristic considerations. To get a better intuitive idea, consider a specific situation, e.g. a random variable $U = X(n + d)$ describes an exchange rate between Slovak koruna and U.S. dollar at time $n + d$ (for example $d = 1$ means one week later) and we have two predictors $\tilde{U}_1$, $\tilde{U}_2$. We can take differences $\tilde{U}_1 - U$, $\tilde{U}_2 - U$ and compare their realizations after given week so we get an explicit result which predictor is better.[5] However it should

[5]Time series data of U.S. dollar to Slovak koruna exchange rate for last 5 years is available for example at <http://finance.yahoo.com/currency/convert?from=USD&to=SKK&amt=1&t=5y>.

be intuitively clear that one realization of random variables $\tilde{U}_1 - U$, $\tilde{U}_2 - U$ is not sufficient for making a general conclusion. Because the situation is quite analogical to making a decision who is taller? Slovakian or Austrian men? Comparison of two randomly chosen men from both nations cannot be sufficient for such decision. We need to take many (or all) men and compare them in average.

Statistical concepts for prediction quality. Therefore these considerations naturally lead to the following concepts: the difference $\tilde{U} - U$ is called *the prediction error* and $E\{(\tilde{U} - U)^2\}$ is termed *the mean squared error* of $\tilde{U}$.

In geometrical language of Hilbert spaces (see Tab. 1) the mean squared error is a squared distance between a predictor and predicted variable in space $\mathscr{H}$. Such interpretation of MSE offers the approach — finding the best predictor means finding a predictor whose squared distance from predicted variable is as small as possible. Or from the statistical viewpoint the best predictor $U^*(X)$ of U minimizes MSE of all predictors from a given class of predictors:

$$U^*(X) = \arg\min_{U} E\{[U - X(n+d)]^2\}.$$

In spite of fact that MSE was chosen as an effective measure of the quality of any predictor, it does not mean that prediction error has no meaning for prediction. For example unbiasedness of predictor $E\{\tilde{U}\} = E\{U\}$ can be stated as zero average prediction error $E\{\tilde{U} - U)\} = 0$. Or for unbiased prediction according to (7) $D\{\tilde{U} - U\} = E\{(\tilde{U} - U)^2\}$, so in that case MSE of predictor is given by dispersion of the prediction error.

As we see bellow, for time series prediction it is important to know appropriate formula for $E\{(T-U)^2\} = \|T-U\|^2_{\mathscr{H}}$, so previous results (8) and (7) allow us to write for $T = \tilde{U}$ a U

$$E\{(\tilde{U} - U)^2\} = (E\{\tilde{U}\} - E\{U\})^2 + D\{\tilde{U}\} + D\{U\} - 2\mathrm{Cov}\{\tilde{U}, U\} \tag{9}$$

3.2 Definition and basic form of the best linear unbiased predictor

If we consider a class of linear unbiased predictors (such predictors have some theoretical and practical advantages and meaning which we mention in conclusion of the paper), then we talk about the best linear unbiased predictor.

Definition 2 (BLUP) *Let $X = (X(1), X(2), \dots, X(n))'$ be an observation of a time series $X(.)$ given by LRM with unknown regression parameter $\beta \in \mathbb{E}^k$. A random variable $X^*_\nu(n+d)$ is called the best linear unbiased predictor of $X(n+d)$, if it is:*

(a) *linear in X, i.e $X^*_\nu(n+d) = a'X + a_0; a \in \mathbb{E}^n, a_0 \in \mathbb{E}$,*

(b) *unbiased for all β, i.e. $E\{X^*_\nu(n+d)\} = E\{X(n+d)\}$ for all $\beta \in \mathbb{E}^k$,*

(c) *best, i.e. minimizing the MSE in the class of all linear unbiased predictors:*

$$X^*_\nu(n+d) = \underset{U=a'X+a_0; E\{U\}=E\{X(n+d)\}}{\arg\min} E\{[U - X(n+d)]^2\},$$

Using formula (9) and the well-known expression about covariance from statistics: $\mathrm{Cov}\{a'X_1 + a_0, b'X_2 + b_0\} = a'\mathrm{Cov}\{X_1, X_2\}b; a, b \in \mathbb{E}^n; a_0, b_0 \in \mathbb{E}$, we have for MSE of BLUP

$$E\{(X(n+d) - a'X - a_0)^2\} = (E\{X(n+d)\} - E\{a'X + a_0\})^2 + r_0 + a'\Sigma_\nu a - 2a'r_\nu,$$

where $r_0 = D\{X(n+d)\}$, $D\{a'X + a_0\} = a'\Sigma_\nu a$ and $r_\nu = \mathrm{Cov}\{X(n+d), X\}$.

Now our goal is to find a predictor satisfying all three conditions in the previous definition for a linear regression model of $X(.)$, whose observation X has a full rank design matrix F and a full rank covariance matrix Σ_ν (a special case of such observation is an observation of FDSLRM). We have adapted and prepared the complete, less traditional geometrical argument[6] originally used for spatial kriging (see Stein [9]).

Linearity and Unbiasedness. In particular, condition **(b)** in case of linear predictors: $E\{a'X + a_0\} = E\{X(n+d)\}$ has the form $a'F\beta + a_0 = f'\beta$ or $(F'a - f)'\beta = a_0$ and these equations must hold for all β. The case $\beta = 0$ gives $a_0 = 0$ and thus in case of $\beta \neq 0$ $F'a$ must be f. The unbiasedness restricting condition is therefore equivalent to

$$a_0 = 0 \text{ and } F'a = f. \tag{10}$$

Hence it is evident that a linear unbiased predictor exists and possess unbiasedness condition (10) if and only if f belongs to $\mathscr{L}(F') = \{F'x | x \in \mathbb{E}^n\}$, the row space of F, or in other words if vector f is a linear combination of F's rows. Such condition holds automatically for example in our case of GFDSLRM with a full column rank F [if $rank(F) = k$, then number of linearly independent rows in F is just k, so these rows constitute a basis for $\mathbb{E}^k$, which means that $f \in \mathscr{L}(F')$].

If a satisfies (10), then any linear unbiased predictor can be written as $(a+c)'X$ where $F'c = 0$ (vector c and columns of F are orthogonal). The collection $\mathscr{U}$ of all linear unbiased predictors U is

$$\mathscr{U} = \{U = (a+c)'X; a, c \in \mathbb{E}^n : F'a = f \text{ and } F'c = 0\}. \tag{11}$$

[($\supset$) if $U = (a+c)'X; F'a = f, F'c = 0$, then $F'(a+c) = f + 0 = f$, so according to (10) U is unbiased; ($\subset$) if $U = b'X + b_0$ is unbiased, i.e. $F'b = f, b_0 = 0$, let $c \equiv b - a$ where $F'a = f$, then $F'c = F'b - F'a = 0$ and $U = b'X = (a+c)'X$].

Minimizing MSE. The BLUP minimizes the MSE among all predictors from $\mathscr{U}$, where

$$\begin{aligned} E\{(X(n+d) - (a+c)'X)^2\} &= r_0 + (a+c)'\Sigma_\nu(a+c) - 2(a+c)'r_\nu \\ &= r_0 - 2a'r_\nu + a'\Sigma_\nu a + c'\Sigma_\nu c - 2(r_\nu - \Sigma_\nu a)'c \end{aligned} \tag{12}$$

If we choose $a^* \in \mathbb{E}^n$ satisfying simultaneously $F'a^* = f$ and $(r_\nu - \Sigma_\nu a^*) \in \mathscr{L}(F)$, i.e. it is a linear combination of F's columns [as we see bellow, in our case it is always possible], then there exists a vector b^* such that $r_\nu - \Sigma_\nu a^* = Fb^*$, $(r_\nu - \Sigma_\nu a^*)'c = b^{*\prime}F'c$ and the following expression for MSE holds

$$\mathrm{MSE}_\nu\{(a^* + c)'X)\} = r_0 - 2a^{*\prime}r_\nu + a^{*\prime}\Sigma_\nu a^* + c'\Sigma_\nu c;\ c \in \mathbb{E}^n : F'c = 0.$$

Since $c'\Sigma_\nu c = D\{c'X\} \geq 0$, the lower bound for MSEs of linear unbiased predictors is achieved for $c = 0$, so $a^{*\prime}X$ with $a^* \in \mathbb{E}^n$ satisfying $\Sigma_\nu a^* + Fb^* = r_\nu, F'a^* = f$ ($b^* \in \mathbb{E}^n$) is the required BLUP $X^*_\nu(n+d)$ of $X(n+d)$ with $\mathrm{MSE}_\nu\{X^*_\nu(n+d)\} = r_0 - 2a^{*\prime}r_\nu + a^{*\prime}\Sigma_\nu a^*$.

Existence of a^* and of the BLUP. The last equations for vectors a^*, b^* can be effectively written in a matrix form

$$\begin{pmatrix} \Sigma_\nu & F \\ F' & 0 \end{pmatrix} \begin{pmatrix} a^* \\ b^* \end{pmatrix} = \begin{pmatrix} r_\nu \\ f \end{pmatrix}, \tag{13}$$

[6]Finding BLUP is a minimization problem with an auxiliary condition, so that it is standardly solved by the Lagrange method of undetermined multipliers, see e.g. Štulajter [10].

which have just one solution, if Σ_ν and F are of full rank, because inverse of given block matrix exists. [If $rank(\Sigma_\nu) = n$ and $rank(F) = rank(F') = k$, then $(n+k)\times(n+k)$ block matrix $\left(\begin{smallmatrix}\Sigma_\nu & F\\ F' & 0\end{smallmatrix}\right)$ has also full rank $n+k$, which is necessity and sufficiency for the existence of its inverse matrix]

Form of the BLUP. The preceding geometrical argument proves not only that under our assumptions the BLUP always exists, but also provides a method how to construct this BLUP. Applying the well-known Banachiewicz formula (see e.g. Zhang [14])

$$\begin{pmatrix} A & B\\ B' & C\end{pmatrix}^{-1} = \begin{pmatrix} A^{-1} & 0\\ 0 & 0\end{pmatrix} + \begin{pmatrix} -A^{-1}B\\ I\end{pmatrix} W^{-1}\left(-B'A^{-1} \quad I\right),$$

where[7] $W = C - B'A^{-1}B$, after some arrangements we get finally for a^* and for BLUP of $X(n+d)$ with corresponding MSE:

$$\begin{aligned} a^* &= \Sigma_\nu^{-1}r_\nu - \beta_\nu^{*\prime}F'\Sigma_\nu^{-1}r_\nu + \beta_\nu^{*\prime}f,\\ X_\nu^*(n+d) &= f'\beta_\nu^* + r_\nu'\Sigma_\nu^{-1}(X - F\beta_\nu^*), \\ \mathrm{MSE}\{X_\nu^*(n+d)\} &= D_\nu\{X(n+d)\} - r_\nu'\Sigma_\nu^{-1}r_\nu + \left\|f - F'\Sigma_\nu^{-1}r_\nu\right\|^2_{\Sigma_{\beta_\nu^*}}, \end{aligned} \tag{14}$$

where

$f = (f_1(n+d), f_2(n+d), \ldots, f_k(n+d))' \in \mathbb{E}^k$,

$\beta_\nu^* = (F'\Sigma_\nu^{-1}F)^{-1}F'\Sigma_\nu^{-1}X$ is the best linear unbiased estimator (BLUE) of β with the covariance matrix $\mathrm{Cov}_\nu\{\beta_\nu^*\} = (F'\Sigma_\nu^{-1}F)^{-1}$ denoted $\Sigma_{\beta_\nu^*}$,

$\|x\|_A^2$ denotes a squared norm of a vector x defined as $x'Ax$.

3.3 A block matrix form of the best linear unbiased predictor in GFDSLRM

In this section we apply obtained general form (14) of BLUP in case of FDSLRM observation which is described by a linear mixed model. As we will see, in such case there exists a formally much simpler form of the BLUP written by means of a partitioned (block) matrix, which from the theoretical point of view is fundamental in deducing important general conclusions about the BLUP prediction quality in various FDSLRM models. This form results from the well-known Henderson's mixed model equations developed in Henderson et al [6] (see also [3, 8]) for linear mixed models, therefore we shall also call the form Henderson's form of the BLUP for general FDSLRM.

Henderson's mixed model equations in case of FDSLRM observation have the form

$$\begin{pmatrix} F'R^{-1}F & F'R^{-1}V\\ V'R^{-1}F & D^{-1}+V'R^{-1}V\end{pmatrix}\begin{pmatrix}\beta_\nu^*\\ Y_\nu^*\end{pmatrix} = \begin{pmatrix} F'R^{-1}X\\ V'R^{-1}X\end{pmatrix} \tag{15}$$

where $R = \mathrm{Cov}\{w\} = \sigma^2 I$, $D = \mathrm{Cov}\{Y\} = diag(\sigma_j^2)$, β_ν^* is the BLUE for β and Y_ν^* is BLUP of Y based on the time series observation X. This BLUP is defined in same way as BLUP for $X(n+d)$ and its derivation (again without assuming normality) can be found e.g. in Searle et al [8]. Substituting $\mathrm{Cov}\{w\} = \sigma^2 I$ into (15), we obtain a simplified form of mixed model equations:

$$\begin{pmatrix} F'F & F'V\\ V'F & \sigma^2D^{-1}+V'V\end{pmatrix}\begin{pmatrix}\beta_\nu^*\\ Y_\nu^*\end{pmatrix} = \begin{pmatrix} F'X\\ V'X\end{pmatrix}. \tag{16}$$

[7] W is called the Schur complement of A in $K = \left(\begin{smallmatrix}A & B\\ B' & C\end{smallmatrix}\right)$.

The form of the BLUP for Y has a similar structure as the BLUP for $X(n+d)$

$$\mathrm{BLUP}(Y) = Y^*_\nu = \mathrm{Cov}\{X,Y\}\Sigma_\nu^{-1}(X - F\beta^*_\nu),$$

where $\mathrm{Cov}\{X,Y\} = \mathrm{Cov}\{F\beta + VY + w, Y\} = VD$. Since it is not difficult to see that r_ν in FDSLRM equals VDv, the BLUP for X can be written as

$$X^*_\nu(n+d) = f'\beta^*_\nu + v'Y^*_\nu$$

or using (16)

$$X^*_\nu(n+d) = \begin{pmatrix} f \\ v \end{pmatrix}' \begin{pmatrix} \beta^*_\nu \\ Y^*_\nu \end{pmatrix} = \begin{pmatrix} f \\ v \end{pmatrix}' \begin{pmatrix} F'F & F'V \\ V'F & \sigma^2 D^{-1} + V'V \end{pmatrix}^{-1} \begin{pmatrix} F'X \\ V'X \end{pmatrix}$$

The obtained results are summarized in the following theorem.

Theorem 1 (Block matrix form of the BLUP in general FDSLRM)
Let us consider a general FDSLRM (2), $k, l \geq 1$, *with its corresponding observation* X:

$$X = F\beta + \varepsilon, \quad E\{\varepsilon\} = 0, \quad \mathrm{Cov}_\nu\{X\} = \Sigma_\nu = \sigma^2 I + VDV',$$

where $\beta \in \mathbb{E}^k, F = (f_1\ f_2\ ...\ f_k), V = (v_1\ v_2\ ...\ v_l)$ *a* $D = diag(\sigma_j^2)$. *Let*

$$E\{X(n+d)\} = f'\beta, \quad f = (f_1(n+d),\ ..., f_k(n+d))',$$
$$\mathrm{Cov}_\nu\{X, X(n+d)\} = VDv, \quad v = (v_1(n+d),\ ..., v_l(n+d))'.$$

Then BLUP $X^*_\nu(n+d)$ *of* $X(n+d)$ *is given by:*

$$X^*_\nu(n+d) = z'G^{-1}Z'X \tag{17}$$

where $z = \begin{pmatrix} f \\ v \end{pmatrix}, Z \equiv (F\ V), G = Z'Z + \begin{pmatrix} 0 & 0 \\ 0 & \sigma^2 D^{-1} \end{pmatrix}$.
For its MSE the following expression holds:

$$MSE[X^*_\nu(n+d)] = \sigma^2(1 + z'G^{-1}z). \tag{18}$$

First of all this block matrix version of BLUP is computationally more economic than original one (14), since original one requires finding inversion of Σ_ν of order n, whereas our block matrix expression requires inversion of G of order $k+l$. The structure of this new form of BLUP also leads to the following effective classification of FDSLRM models, which describes a mutual geometrical relationship between columns of F and V and influences significantly the structure of the MSE in the FDSLRM.

FDSLRM model is said to be:

(a) *full-orthogonal*, if $f_i \perp v_j$ for $i = 1, 2, \ldots, k$; $j = 1, 2, \ldots, l$, $v_i \perp v_j$ for $i, j = 1, 2, \ldots, l, i \neq j$ and $f_i \perp f_j$ for $i, j = 1, 2, \ldots, k, i \neq j$. Then $F'V = 0$, $F'F$ and $V'V$ are diagonal.

(b) *orthogonal*, if $f_i \perp v_j$ for $i = 1, 2, \ldots, k$; $j = 1, 2, \ldots, l$ and $v_i \perp v_j$ for $i, j = 1, 2, \ldots, l, i \neq j$. Then $F'V = 0$, $V'V$ is diagonal.

(c) *semi-orthogonal*, if $f_i \perp v_j$ for all i, j. It means $F'V = 0$.

(d) *nonorthogonal*, if $f_i \not\perp v_j$ for some i, j or if $F'V \neq 0$.

4 CONCLUSION

In the paper we have presented main ideas of the best linear unbiased predictor (BLUP) and its particular form in case of the so called finite discrete spectrum linear regression models (FDLSRM).

Since selecting of the time series model for measured data is usually left on user, e.g. gasoline price model from introduction could be also described by GFDSLRM, our computationally economic matrix-block version of MSE for BLUP in FDSLRM, subject to no assumption[8] imposed on the joint distribution of $(X, X(n+d))'$, can be effectively used in comparing selected models from the viewpoint of the prediction.

Finally we summarize advantages of the best linear unbiased prediction. BLUP does not depend on unknown mean value regression parameter β (this property was obtained by requiring unbiasedness of predictors). It depends only on parameters of covariance functions in its calculation, so there is no need to know the joint distribution of $(X, X(n+d))'$. Even if the joint distribution is known, BLUP is much easier to calculate than the conditional expectation $E\{X(n+d)|X\}$ (having always smaller MSE), whose explicit form is known only for a few distributions (moreover generally $E\{X(n+d)|X\}$ is a nonlinear function of X causing difficult theoretical study of its statistical properties).

However there are examples (see e.g. Stein [9]) when BLUP can be poor predictor. Finally we have to remind that all derived results are based on the assumption that we know $\Sigma = \text{Cov}\{X\}$ and $r = \text{Cov}\{X, X(n+d)\}$. However in practise this conditions almost never holds. In such case we need to solve a problem of its estimation. In the orthogonal FDSLRM the problem was studied and resolved by Štulajter & Witkovský [12] using modified DOOLSE estimators. For this type of the FDSLRM there were also studied effects of such estimating on the MSE with the result that the suggested estimates give asymptotically the same MSE as it would be given by unknown real parameters (Štulajter [13]).

Concerning a general FDSLRM one (practically-minded) solution of the problem of estimation ν was recently given in Hančová [5] using the so-called natural estimators. Results of the paper are needed as a theoretical basis for a similar study of effects of estimating ν on MSE of the BLUP in the semi-orthogonal and nonorthogonal FDSLRM as it was done for orthogonal models.

REFERENCES

[1] G.E.P. Box, G.M. Jenkins, G.C. Reinsel, *Time Series Analysis, Forecasting and Control*, 3rd ed., Prentice Hall, New Jersey, 1994.

[2] P. J. Brockwell and R. A. Davis, *Time Series: Theory and Methods*, 2nd ed., Springer–Verlag, New York, 1991

[3] R. Christensen, *Advanced Linear Modeling: Multivariate, Time series and Spatial Data*, 2nd ed., Springer–Verlag, New York, 2001

[4] A. S. Goldberger, Best linear unbiased prediction in the generalized linear regression models, *J. Am. Stat. Assoc.* **57**, pp. 369–375, 1962

[8] e.g. without assuming normality.

[5] M. Hančová, Natural estimation of variances in a general finite discrete spectrum linear regression model, *Metrika*, DOI 10.1007/s00184-007-0132-9, published online first, www.springerlink.com, 2007

[6] C. R. Henderson, S. R. Searle, O. Kempthorne, C. M. von Krosigk, Estimation of Enviromental and Genetic Trends from Records Subject to Culling, *Biometrics* **15**, pp. 192–218, 1959

[7] Ch. E. McCulloch and S. R. Searle, *Generalized, Linear, and Mixed Models*, Wiley–Interscience, New York, 2001

[8] S. R. Searle, G. Cassella and C. E. McCulloch, *Variance Components*, Wiley–Interscience, New York, 1992.

[9] M. L. Stein, *Interpolation of Spatial Data: Some Theory of Kriging*, Springer–Verlag, New York, 1999.

[10] F. Štulajter, *Predictions in Time Series Using Regression Models*, Springer–Verlag, New York, 2002

[11] F. Štulajter, The MSE of the BLUP in a finite discrete spectrum LRM. *Tatra Mt. Math. Publ.* **26**, pp. 125–131, 2003

[12] F. Štulajter and V. Witkovský, Estimation of variances in orthogonal finite discrete spectrum linear regression models, *Metrika* **60**, pp. 105–118, 2004.

[13] F. Štulajter, Mean squared error of the empirical best linear unbiased predictor in an orthogonal finite discrete spectrum linear regression model, *Metrika* **65**, pp. 331–348, 2007

[14] F. Zhang, et al, *The Schur Complement and Its Applications*, Springer Science & Business Media, New York, 2005