

MSE OF THE BEST LINEAR PREDICTOR IN NONORTHOGONAL MODELS

FRANTIŠEK ŠTULAJTER

(Communicated by Anatolij Dvurečenskij)

Dedicated to Professor B. Riečan on the occasion of his 70th birthday

ABSTRACT. The problem of computing the mean squared error (MSE) of the best linear predictor (BLP) in finite discrete spectrum with an additive white noise models (FDSWNMs) for an observed time series is considered. This is done under the assumption that the corresponding vectors in models for finite observation of this time series are not orthogonal.

©2007
Mathematical Institute
Slovak Academy of Sciences

1. Introduction

We shall consider the problem of prediction of time series which is based on modeling time series by *linear regression models*. In this approach the *best linear predictor* (BLP) minimizing the *mean squared error* (MSE) of prediction, can be found. This method is known in an engineering literature as *kriging*, see [2], [1], [5], and [6]. For a given time series data we can use different regression models and thus the problem of computation of the MSE of the BLP in different models arises.

We shall compute the MSE of the BLP in a *finite discrete spectrum with an additive white noise model* (FDSWNM). These models were already studied in [6] and [7], where it was assumed that the vectors which we get from functions generating these models are orthogonal. Models with orthogonal vectors can be

2000 Mathematics Subject Classification: Primary 62M10.

Keywords: time series, finite discrete spectrum with an additive white noise model, best linear predictor, mean squared error of predictor.

Supported by the VEGA grant of the Slovak Grant Agency No. 1/3023/06.

used in many practical applications of time series theory, but it is necessary to study also the situation where the model vectors are not orthogonal.

An FDSWNM for a time series $X(\cdot)$ is given by, see [6],

$$X(t) = \sum_{i=1}^l Y_i v_i(t) + w(t), \quad t = 1, 2, \dots, \quad (1)$$

where $Y = (Y_1, Y_2, \dots, Y_l)'$ is a random vector with $E[Y] = 0$ and with an covariance matrix $\text{Cov}(Y) = \text{diag}(\sigma_i^2)$. $v_i(\cdot)$, $i = 1, 2, \dots, l$, are given known functions, $w(\cdot)$ is a white noise with a variance $D[w(t)] = \sigma^2$ which is uncorrelated with random vector $Y = (Y_1, Y_2, \dots, Y_l)'$.

As an example we can consider a time series $Y(\cdot)$ with a finite discrete spectrum, see [4]. This can be written in the form

$$Y(t) = \sum_{i=1}^{l/2} (U_i \cos \lambda_i t + Z_i \sin \lambda_i t), \quad t = 1, 2, \dots,$$

where $Y = (U_1, U_2, \dots, U_{l/2}, Z_1, Z_2, \dots, Z_{l/2})'$ and where $\lambda_1, \dots, \lambda_{l/2}$ are some frequencies from $\langle 0, \pi \rangle$. Let $D[U_i] = \sigma_i^2$ and $D[Z_i] = \kappa_i^2$. Then the covariance function $R(\cdot, \cdot)$ of $Y(\cdot)$ is

$$R(s, t) = \sum_{i=1}^{l/2} (\sigma_i^2 \cos \lambda_i s \cos \lambda_i t + \kappa_i^2 \sin \lambda_i s \sin \lambda_i t), \quad s, t = 1, 2, \dots$$

Time series $Y(\cdot)$ is covariance stationary if $D[U_i] = D[Z_i] = \sigma_i^2$, $i = 1, 2, \dots, l/2$, and its covariance function in this case is given by

$$R(s, t) = \sum_{i=1}^{l/2} \sigma_i^2 \cos \lambda_i (s - t), \quad s, t = 1, 2, \dots$$

Realizations $y(\cdot)$ of time series in this example are simply linear combinations of goniometric functions, but this is not realistic in practise. More realistic situation is that a realization of a white noise is added to this linear combination and thus we get the FDSWNM. This approach is similar to that used in a classical linear regression model.

Time series $X(\cdot)$, given by FDSWNM, have covariance functions $R_\nu(\cdot, \cdot)$ given by

$$R_\nu(s, t) = \sigma^2 \delta_{s,t} + \sum_{i=1}^l \sigma_i^2 v_i(s) v_i(t), \quad s, t = 1, 2, \dots, \quad (2)$$

where $\nu = (\sigma^2, \sigma_1^2, \dots, \sigma_l^2)' \in (0, \infty) \times \langle 0, \infty \rangle^l = \Upsilon$.

For this model we get for a finite observation $X = (X(1), \dots, X(n))'$ of $X(\cdot)$ the model

$$X = VY + w,$$

where the $n \times l$ matrix $V = (v_1, \dots, v_l)$ has columns, $n \times 1$ vectors, $v_i = (v_i(1), \dots, v_i(n))'$, $i = 1, 2, \dots, l$. In this model $E[X] = 0$ and covariance matrices Σ_ν , $\nu \in \Upsilon$ of X are positive definite and are given by

$$\Sigma_\nu = \sigma^2 I + \sum_{i=1}^l \sigma_i^2 v_i v_i' = \sum_{i=0}^l \sigma_i^2 V_i,$$

where $V_0 = I$, $V_i = v_i v_i'$, with ranks $r(V_i) = 1$, $i = 1, 2, \dots, l$, and $\sigma_0^2 = \sigma^2$.

It should be remarked that for time series $Y(\cdot)$ which was considered in the preceding example the vectors $v_i = (\cos \lambda_i 1, \dots, \cos \lambda_i n)'$ and $v_i = (\sin \lambda_i 1, \dots, \sin \lambda_i n)'$ in general are not be orthogonal.

According to the classical theory, see [2], the best linear predictor $X^*(n+d)$ of $X(n+d)$ is given by

$$X^*(n+d) = r'_\nu \Sigma_\nu^{-1} X, \tag{3}$$

where $r_\nu = \text{Cov}_\nu(X; X(n+d))$ and

$$\begin{aligned} MSE_\nu[X^*(n+d)] &= E_\nu[X^*(n+d) - X(n+d)]^2 \\ &= D_\nu[X(n+d)] - r'_\nu \Sigma_\nu^{-1} r_\nu. \end{aligned} \tag{4}$$

In [6] the explicit expressions, as functions of vectors v_i and variances σ^2 and σ_i^2 , $i = 1, 2, \dots, l$, for $X^*(n+d)$ and for $MSE_\nu[X^*(n+d)]$ are given under the assumption that the vectors v_i , $i = 1, 2, \dots, l$, are orthogonal. In this article we derive these expressions for $l = 2$ and under the assumption that the vectors v_1 , v_2 are not orthogonal.

2. Mean squared error of the BLP in a nonorthogonal finite discrete spectrum white noise model

The following lemma gives a basic result for FDSWNMs with two components in the case when the vectors v_1 , v_2 are not orthogonal. Some other useful results on matrix algebra can be found in [3].

LEMMA 1. *For any $n \times 1$ vectors v_1 , v_2 and any real positive numbers σ^2 , σ_1^2 and σ_2^2 we have*

$$(\sigma^2 I + \sigma_1^2 v_1 v_1' + \sigma_2^2 v_2 v_2')^{-1} = \frac{1}{\sigma^2} \left(I - \frac{d_1 V_1 + d_2 V_2 - d_1 d_2 (v_1, v_2) V_{12}}{1 - d_1 d_2 (v_1, v_2)^2} \right), \tag{5}$$

where

$$\begin{aligned} d_i &= (\sigma^2/\sigma_i^2 + \|v_i\|^2)^{-1}, \\ V_i &= v_i v_i', \quad V_{12} = v_1 v_2' + v_2 v_1', \\ (v_1, v_2) &= v_1' v_2 \quad \text{and} \quad \|v_i\|^2 = (v_i, v_i), \quad i = 1, 2. \end{aligned} \quad (6)$$

Proof. By a direct computation we can verify that, see [6],

$$(A + \sigma_2^2 v_2 v_2')^{-1} = A^{-1} - \frac{\sigma_2^2 A^{-1} v_2 v_2' A^{-1}}{1 + \sigma_2^2 v_2' A^{-1} v_2}$$

for any positive definite matrix A . Let $A = \sigma^2 I + \sigma_1^2 v_1 v_1'$, then we have

$$A^{-1} = (\sigma^2 I + \sigma_1^2 v_1 v_1')^{-1} = \frac{1}{\sigma^2} (I - d_1 V_1),$$

and

$$\begin{aligned} & \sigma^2 (\sigma^2 I + \sigma_1^2 v_1 v_1' + \sigma_2^2 v_2 v_2')^{-1} \\ &= I - d_1 V_1 - \frac{(\sigma_2^2/\sigma^2)(I - d_1 V_1)v_2 v_2'(I - d_1 V_1)}{1 + (\sigma_2^2/\sigma^2)v_2'(I - d_1 V_1)v_2} \\ &= I - d_1 V_1 - \frac{(\sigma_2^2/\sigma^2)(V_2 - d_1(v_1, v_2)V_{12} + d_1^2(v_1, v_2)^2 V_1)}{1 + (\sigma_2^2/\sigma^2)(\|v_2\|^2 - d_1(v_1, v_2)^2)}. \end{aligned}$$

After some computation we get

$$\begin{aligned} & 1 + (\sigma_2^2/\sigma^2)(\|v_2\|^2 - d_1(v_1, v_2)^2) \\ &= 1 + (\sigma_2^2/\sigma^2)\|v_2\|^2 - (\sigma_2^2/\sigma^2)\frac{\sigma_1^2/\sigma^2}{1 + (\sigma_1^2/\sigma^2)\|v_1\|^2}(v_1, v_2)^2 \\ &= (1 + \sigma_2^2/\sigma^2)\|v_2\|^2 - d_1(\sigma_2^2/\sigma^2)(v_1, v_2)^2 \\ &= (1 + \sigma_2^2/\sigma^2)\|v_2\|^2 (1 - d_1 d_2(v_1, v_2)^2) \end{aligned}$$

and thus

$$\begin{aligned} & \frac{(\sigma_2^2/\sigma^2)(V_2 - d_1(v_1, v_2)V_{12} + d_1^2(v_1, v_2)^2 V_1)}{1 + (\sigma_2^2/\sigma^2)(\|v_2\|^2 - d_1(v_1, v_2)^2)} \\ &= \frac{d_2 V_2 - d_1 d_2(v_1, v_2)V_{12} + d_1^2 d_2(v_1, v_2)^2 V_1}{1 - d_1 d_2(v_1, v_2)^2}. \end{aligned}$$

Using this result we get

$$\begin{aligned} & \sigma^2(\sigma^2 I + \sigma_1^2 v_1 v_1' + \sigma_2^2 v_2 v_2')^{-1} \\ &= I - d_1 V_1 - \frac{d_2 V_2 - d_1 d_2(v_1, v_2) V_{12} + d_1^2 d_2(v_1, v_2)^2 V_1}{1 - d_1 d_2(v_1, v_2)^2} \\ &= I - \frac{d_1 V_1 + d_2 V_2 - d_1 d_2(v_1, v_2) V_{12}}{1 - d_1 d_2(v_1, v_2)^2} \end{aligned}$$

and the lemma is proved. \square

This result can be used for computing the BLP, $X^*(n+d)$, and its MSE in an FDSWNM. In such a model we have, using (2) and (5),

$$\begin{aligned} & r'_\nu \Sigma_\nu^{-1} = \\ &= \left(\sum_{i=1}^2 \frac{\sigma_i^2}{\sigma^2} v_i(n+d)v_i' \right)' \left(I - \frac{d_1 v_1 v_1' + d_2 v_2 v_2' - d_1 d_2(v_1, v_2)(v_1 v_2' + v_2 v_1')}{1 - d_1 d_2(v_1, v_2)^2} \right) \end{aligned}$$

and, after some computation, we get

$$\begin{aligned} r'_\nu \Sigma_\nu^{-1} &= \frac{\sigma_1^2}{\sigma^2} v_1(n+d)v_1' + \frac{\sigma_2^2}{\sigma^2} v_2(n+d)v_2' \\ &\quad - \frac{\sigma_1^2}{\sigma^2} v_1(n+d) \frac{d_1 \|v_1\|^2 v_1' + d_2(v_1, v_2)v_2'}{1 - d_1 d_2(v_1, v_2)^2} \\ &\quad + \frac{\sigma_1^2}{\sigma^2} v_1(n+d) \frac{d_1 d_2(v_1, v_2)(\|v_1\|^2 v_2' + (v_1, v_2)v_1')}{1 - d_1 d_2(v_1, v_2)^2} \\ &\quad - \frac{\sigma_2^2}{\sigma^2} v_2(n+d) \frac{d_2 \|v_2\|^2 v_2' + d_1(v_1, v_2)v_1'}{1 - d_1 d_2(v_1, v_2)^2} \\ &\quad + \frac{\sigma_2^2}{\sigma^2} v_2(n+d) \frac{d_1 d_2(v_1, v_2)(\|v_2\|^2 v_1' + (v_1, v_2)v_2')}{1 - d_1 d_2(v_1, v_2)^2} \\ &= \frac{\sigma_1^2}{\sigma^2} v_1(n+d) \frac{1 - d_1 \|v_1\|^2}{1 - d_1 d_2(v_1, v_2)^2} v_1' \\ &\quad + \frac{\sigma_2^2}{\sigma^2} v_2(n+d) \frac{1 - d_2 \|v_2\|^2}{1 - d_1 d_2(v_1, v_2)^2} v_2'. \end{aligned}$$

Using the expression

$$1 - d_i \|v_i\|^2 = \frac{\sigma^2}{\sigma_i^2} d_i, \quad i = 1, 2, \quad (7)$$

which follows from (6), we get

$$\begin{aligned} r'_\nu \Sigma_\nu^{-1} &= \frac{d_1(v_1(n+d) + d_2(v_1, v_2)v_2(n+d))}{1 - d_1 d_2(v_1, v_2)^2} v'_1 \\ &+ \frac{d_2(v_2(n+d) + d_1(v_1, v_2)v_1(n+d))}{1 - d_1 d_2(v_1, v_2)^2} v'_2 \end{aligned} \quad (8)$$

and this expression can be used by computing the BLP, $X^*(n+d)$.

To compute the mean squared error of this predictor, which is given by (4), we use the expression

$$D_\nu[X(n+d)] = \sigma^2 + \sum_{i=1}^2 \sigma_i^2 v_i^2(n+d), \quad (9)$$

and for $r'_\nu \Sigma_\nu^{-1} r_\nu$ we get, using (2) and (8),

$$\begin{aligned} r'_\nu \Sigma_\nu^{-1} r_\nu &= \frac{d_1(v_1(n+d) + d_2(v_1, v_2)v_2(n+d))}{1 - d_1 d_2(v_1, v_2)^2} \left(\sum_{i=1}^2 \sigma_i^2 v_i(n+d)(v_1, v_i) \right) \\ &+ \frac{d_2(v_2(n+d) + d_1(v_1, v_2)v_1(n+d))}{1 - d_1 d_2(v_1, v_2)^2} \left(\sum_{i=1}^2 \sigma_i^2 v_i(n+d)(v_2, v_i) \right). \end{aligned}$$

After long and tedious computations we can write

$$\begin{aligned} &r'_\nu \Sigma_\nu^{-1} r_\nu = \\ &= \sigma^2 \frac{\sigma_1^4}{\sigma^4} v_1^2(n+d) \|v_1\|^2 \frac{1 - d_1 \|v_1\|^2 - d_2(v_1, v_2)^2 \|v_1\|^{-2} + d_1 d_2(v_1, v_2)^2}{1 - d_1 d_2(v_1, v_2)^2} \\ &+ 2\sigma^2 \frac{\sigma_1^2 \sigma_2^2}{\sigma^2 \sigma^2} v_1(n+d) v_2(n+d) (v_1, v_2) \times \\ &\quad \times \frac{1 - d_1 \|v_1\|^2 - d_2 \|v_2\|^2 + d_1 d_2 \|v_1\|^2 \|v_2\|^2}{1 - d_1 d_2(v_1, v_2)^2} \\ &+ \sigma^2 \frac{\sigma_2^4}{\sigma^4} v_2^2(n+d) \|v_2\|^2 \frac{1 - d_2 \|v_2\|^2 - d_1(v_1, v_2)^2 \|v_2\|^{-2} + d_1 d_2(v_1, v_2)^2}{1 - d_1 d_2(v_1, v_2)^2}. \end{aligned} \quad (10)$$

MSE OF THE BEST LINEAR PREDICTOR IN NONORTHOGONAL MODELS

Thus, using (4), (9) and (10), the expression for the $MSE_\nu[X^*(n+d)]$ can be written as

$$\begin{aligned}
 MSE_\nu[X^*(n+d)] &= \\
 &= \sigma^2 + \sigma^2 \frac{\sigma_1^2}{\sigma^2} v_1^2(n+d) \times \\
 &\quad \times \left(1 - \frac{\sigma_1^2}{\sigma^2} \|v_1\|^2 \frac{1 - d_1 \|v_1\|^2 - d_2(v_1, v_2)^2 \|v_1\|^{-2} + d_1 d_2 (v_1, v_2)^2}{1 - d_1 d_2 (v_1, v_2)^2} \right) \\
 &\quad - 2\sigma^2 \frac{\sigma_1^2 \sigma_2^2}{\sigma^2 \sigma^2} v_1(n+d)v_2(n+d)(v_1, v_2) \times \\
 &\quad \quad \quad \times \frac{1 - d_1 \|v_1\|^2 - d_2 \|v_2\|^2 + d_1 d_2 \|v_1\|^2 \|v_2\|^2}{1 - d_1 d_2 (v_1, v_2)^2} \\
 &\quad - \sigma^2 \frac{\sigma_2^2}{\sigma^2} v_2^2(n+d) \left(1 - \frac{\sigma_2^2}{\sigma^2} \|v_2\|^2 \times \right. \\
 &\quad \quad \quad \left. \times \frac{1 - d_2 \|v_2\|^2 - d_1(v_1, v_2)^2 \|v_2\|^{-2} + d_1 d_2 (v_1, v_2)^2}{1 - d_1 d_2 (v_1, v_2)^2} \right). \tag{11}
 \end{aligned}$$

Next we have from (7)

$$\begin{aligned}
 &1 - \frac{\sigma_1^2}{\sigma^2} \|v_1\|^2 \frac{1 - d_1 \|v_1\|^2 - d_2(v_1, v_2)^2 \|v_1\|^{-2} + d_1 d_2 (v_1, v_2)^2}{1 - d_1 d_2 (v_1, v_2)^2} \\
 &= \frac{1 - (\sigma_1^2/\sigma^2) \|v_1\|^2 (1 - d_1 \|v_1\|^2)}{1 - d_1 d_2 (v_1, v_2)^2} \\
 &= \frac{1 - \|v_1\|^2 (\sigma^2/\sigma_1^2 + \|v_1\|^2)^{-1}}{1 - d_1 d_2 (v_1, v_2)^2} \\
 &= \frac{(\sigma^2/\sigma_1^2)d_1}{1 - d_1 d_2 (v_1, v_2)^2},
 \end{aligned}$$

and, by analogy,

$$\begin{aligned}
 &1 - \frac{\sigma_2^2}{\sigma^2} \|v_2\|^2 \frac{1 - d_2 \|v_2\|^2 - d_1(v_1, v_2)^2 \|v_2\|^{-2} + d_1 d_2 (v_1, v_2)^2}{1 - d_1 d_2 (v_1, v_2)^2} \\
 &= \frac{(\sigma^2/\sigma_2^2)d_2}{1 - d_1 d_2 (v_1, v_2)^2}.
 \end{aligned}$$

We can also derive that

$$\begin{aligned} & \frac{\sigma_1^2 \sigma_2^2}{\sigma^2 \sigma^2} \frac{1 - d_1 \|v_1\|^2 - d_2 \|v_2\|^2 + d_1 d_2 \|v_1\|^2 \|v_2\|^2}{1 - d_1 d_2 (v_1, v_2)^2} \\ &= \frac{\sigma_1^2 \sigma_2^2}{\sigma^2 \sigma^2} \frac{(1 - d_1 \|v_1\|^2)(1 - d_2 \|v_2\|^2)}{1 - d_1 d_2 (v_1, v_2)^2} \\ &= \frac{d_1 d_2}{1 - d_1 d_2 (v_1, v_2)^2}. \end{aligned}$$

Using these results and (11) we get, after some computation, the expression for the $MSE_\nu[X^*(n+d)]$, which is given in the following theorem.

THEOREM 2.1. *The BLP, $X^*(n+d)$, of $X(n+d)$ in an FDSWNM*

$$\begin{aligned} X(t) &= \sum_{i=1}^2 Y_i v_i(t) + w(t), \quad t = 1, 2, \dots, \\ E[Y] &= 0, \quad \text{Cov}(Y) = \text{diag}(\sigma_i^2) \end{aligned}$$

is given by

$$\begin{aligned} X^*(n+d) &= \frac{d_1 v_1(n+d) + d_1 d_2 (v_1, v_2) v_2(n+d)}{1 - d_1 d_2 (v_1, v_2)^2} v_1' X \\ &+ \frac{d_2 v_2(n+d) + d_1 d_2 (v_1, v_2) v_1(n+d)}{1 - d_1 d_2 (v_1, v_2)^2} v_2' X \end{aligned}$$

and

$$\begin{aligned} MSE_\nu[X^*(n+d)] &= \sigma^2 \left(1 + \frac{d_1 v_1^2(n+d) + d_2 v_2^2(n+d)}{1 - d_1 d_2 (v_1, v_2)^2} \right) \\ &- 2\sigma^2 \frac{d_1 d_2 v_1(n+d) v_2(n+d) (v_1, v_2)}{1 - d_1 d_2 (v_1, v_2)^2}, \end{aligned}$$

where

$$d_i = d_i(\nu) = (\sigma^2 / \sigma_i^2 + \|v_i\|^2)^{-1}, \quad i = 1, 2.$$

Remarks. It should be remarked that in the case when the vectors v_1 and v_2 are orthogonal, that means $(v_1, v_2) = 0$, we get

$$X^*(n+d) = d_1 v_1(n+d) v_1' X + d_2 v_2(n+d) v_2' X$$

and

$$MSE_\nu [X^*(n+d)] = \sigma^2 (1 + d_1 v_1^2(n+d) + d_2 v_2^2(n+d))$$

what is the result which is given in [6].

The expression for the $MSE_\nu [X^*(n+d)]$ can be used to find conditions on functions $v_1(\cdot)$ and $v_2(\cdot)$ by which

$$\lim_{n \rightarrow \infty} MSE_\nu [X^*(n+d)] = \sigma^2,$$

the variance of the white noise only.

All results derived above can also serve as a base for computing the mean squared error of the best linear unbiased predictor, see [6], in a linear regression model

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

textwhere

$$\varepsilon(t) = \sum_{j=1}^l Y_j v_j(t) + w(t), \quad t = 1, 2, \dots,$$

is given by the FDSWNM. But this is not the objective of this article.

REFERENCES

- [1] CHRISTENSEN, R.: *Linear Models for Multivariate, Time Series and Spatial Data*, Springer-Verlag, New York, 1991.
- [2] GOLDBERGER, A. S.: *Best linear unbiased prediction in the generalized linear regression model*, J. Amer. Statist. Assoc. **57** (1962), 369–375.
- [3] HARVILLE, D. A.: *Matrix Algebra From a Statistician's Perspective*, Springer-Verlag, New York, 1997.
- [4] PRIESTLEY, M. B.: *Spectral Analysis and Time Series*, Academic Press, London, 1981.
- [5] STEIN, M. L.: *Interpolation of Spatial Data. Some Theory for Kriging*, Springer-Verlag, New York, 1999.
- [6] ŠTULAJTER, F.: *Predictions in Time Series Using Regression Models*, Springer-Verlag, New York, 2002.

FRANTIŠEK ŠTULAJTER

- [7] ŠTULAJTER, F.: *The MSE of the BLUP in a finite discrete spectrum LRM*, Tatra Mt. Math. Publ. **26** (2003), 125–131.

Received 28. 11. 2005
Revised 26. 4. 2006

Department of Probability and Statistics
Faculty of Mathematics, Physics and Informatics
Comenius University
Mlynská Dolina
SK-842 48 Bratislava
SLOVAKIA
E-mail: stulajter@fmph.uniba.sk