Tatra Mt. Math. Publ. **39** (2008), 175–182

TATRA MOUNTAINS
Mathematical Publications

COMPARISON OF PREDICTORS OF TIME SERIES IN ORTHOGONAL REGRESSION MODELS

František Štulajter

ABSTRACT. The problem of comparison of the best linear predictor defined in a finite discrete spectrum model, and of the best linear unbiased predictor defined in a simple linear regression model, is considered. Mean squared errors of these two predictors are computed and compared in both these models under the assumption that functions generating these two models are the same and that the corresponding vectors in these models for a finite observation are orthogonal.

1. Introduction.

We shall consider the problem of prediction of time series based on modeling time series by different regression models. One model will be a *finite discrete spectrum model* (FDSM) and the second a (simple) *linear regression model* (LRM). In this approach the *mean squared errors* (MSEs) of the *best linear predictor* (BLP) (which is defined in the FDSM) and *the best linear unbiased predictor* (BLUP) (defined in the LRM) will be compared in both FDSM and LRM. The general theory of best linear predictors in regression models is described in Goldberger (1962), Christensen (1991) Stein (1999) and Štulajter (2002).

For a given time series data both these two regression models, with the same regression functions, can be used. This is caused by the fact that on a base of time series data, it is not possible to decide what model generates the given data.

Some of the arising problems connected with the comparison of the mean squared errors of the BLP and the BLUP were studied in Štulajter (2002) and Štulajter (2003), where it was assumed that the vectors, which we get from functions generating the FDSM and the LRM, were orthogonal, and also

---

2000 Mathematics Subject Classification: 62M10.
Keywords: time series, finite discrete spectrum model, linear regression model, best linear unbiased predictors, mean squared errors of best linear unbiased predictors.
Supported by the VEGA grant of the Slovak Grant Agency No. 1/3023/06.

it was assumed that the parameters of the covariance function of time series in the FDSM were known.

But only the MSE of the BLP, computed in the FDSM, with the MSE of the BLUP, computed in the LRM, were compared in Š t u l a j t e r (2002). It is also necessary to compare the MSE of the BLP if we use this predictor in the LRM, and the MSE of the BLUP if we use this predictor in the case when the data are generated by the FDSM. This is done, for orthogonal regression models, in this contribution.

2. BLP in a finite discrete spectrum model.

An FDSM for time series $X(.)$ is given by Š t u l a j t e r (2002),

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

where $Y = (Y_1, Y_2, \dots, Y_l)'$ is a random vector with $E[Y] = 0$ and with mutually uncorrelated components Y_i with variances $D[Y_i] = \sigma_i^2;\ i = 1, 2, \dots, l$. It is assumed that $f_i(.);\ i = 1, 2, \dots, l$ are given known functions, $w(.)$ is a white noise with a variance $D\big[w(t)\big] = \sigma^2$ which is uncorrelated with random vector $Y = (Y_1, Y_2, \dots, Y_l)'$.

Time series $X(.)$, given by FDSM, have covariance functions $R_\nu(.,.)$ given by

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

where

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

A finite observation $X = \big(X(1), \dots, X(n)\big)'$ of $X(.)$ given by FDSM can be written as

$$X = FY + w, \tag{2.1}$$

where the $n \times l$ matrix $F = (f_1 \dots f_l)$ has columns, $n \times 1$ vectors,

$$f_i = \big(f_i(1), \dots, f_i(n)\big)'; \qquad i = 1, 2, \dots, l.$$

In this model $E[X] = 0$ and covariance matrices $\mathrm{Cov}_\nu(X) = \Sigma_\nu;\ \nu \in \Upsilon$ of X are positive definite and are given by

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

where $V_0 = I_n,\ V_i = f_i f_i'$, with ranks $r(V_i) = 1;\ i = 1, 2, \dots, l$ and $\sigma_0^2 = \sigma^2$. We call model (2.1.) also an FDSM and we call it *orthogonal* if $f_i \perp f_j$; for $i \neq j$.

According to the classical theory, Goldberger (1962), the BLP, $X_\nu^*(n+d)$, of $X(n+d)$ is given by

$$X_\nu^*(n+d) = r_\nu' \Sigma_\nu^{-1} X, \tag{2.2}$$

where

$$r_\nu = \mathrm{Cov}_\nu\big(X; X(n+d)\big),$$

and

$$\mathrm{MSE}_\nu\big[X_\nu^*(n+d)\big] = D_\nu\big[X(n+d)\big] - r_\nu' \Sigma_\nu^{-1} r_\nu. \tag{2.3}$$

In an orthogonal FDSM we have

$$r_\nu = \sum_{i=1}^{l} \sigma_i^2 f_i(n+d) f_i \tag{2.4}$$

and, see Štulajter (2002),

$$\Sigma_\nu^{-1} = \frac{1}{\sigma^2}\left(I_n - \sum_{i=1}^{l}\left(\sigma^2/\sigma_i^2 + \|f_i\|^2\right)^{-1} f_i f_i'\right). \tag{2.5}$$

The following lemma easily follows from (2.2.)–(2.5.).

Lemma 2.1. *In an orthogonal FDSM*

$$X = FY + w$$

the BLPs, $X_\nu^*(n+d)$, of $X(n+d)$ are for every $\nu \in \Upsilon$ given by

$$X_\nu^*(n+d) = \sum_{i=1}^{l} \frac{f_i(n+d)}{\sigma^2/\sigma_i^2 + \|f_i\|^2} f_i' X,$$

and

$$\mathrm{MSE}_\nu\big[X_\nu^*(n+d)\big] = \sigma^2\left(1 + \sum_{i=1}^{l} \frac{f_i^2(n+d)}{\sigma^2/\sigma_i^2 + \|f_i\|^2}\right).$$

Remark. In Štulajter (2007), similar results as above, for $l = 2$, under the assumption that the vectors f_1, f_2 are not orthogonal, are derived.

3. Comparison of predictors.

Now let us consider for time series $X(.)$ the model

$$X(t) = \sum_{i=1}^{l} \beta_i f_i(t) + w(t); t = 1, 2, \dots$$

where $w(.)$ is a white noise with variance $D\big[w(t)\big] = \sigma^2$.

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

$$X = F\beta + w; \qquad (3.1)$$
$$\mathrm{Cov}_{\sigma^2}(X) = \Sigma_{\sigma^2} = \sigma^2 I_n,$$

where $\beta = (\beta_1, \beta_2, \dots, \beta_l)' \in E^k$, where the $n \times l$ matrix $F = (f_1 \dots f_l)$ has columns, $n \times 1$ vectors, f_i and where $\sigma^2 \in (0, \infty)$.

We call LRM (3.1.) *orthogonal* if $f_i \perp f_j$; for $i \neq j$.

The following lemma describes the BLUP and its MSE in an orthogonal LRM.

Lemma 3.1. *In an orthogonal LRM*

$$X = F\beta + w$$

the BLUP, $\hat{X}(n+d)$, of $X(n+d)$ is given by

$$\hat{X}(n+d) = \sum_{i=1}^{l} \frac{f_i(n+d)}{\|f_i\|^2} f_i' X$$

and

$$\mathrm{MSE}_{\sigma^2}\big[\hat{X}(n+d)\big] = \sigma^2 \left(1 + \sum_{i=1}^{l} \frac{f_i^2(n+d)}{\|f_i\|^2}\right); \qquad \sigma^2 \in (0, \infty).$$

Comparing the results given in Lemma 2.1. and in Lemma 3.1. we have directly the following statement.

Theorem 3.1. *Let $X_\nu^*(n+d)$ be the BLP in an orthogonal FDSM and let $\hat{X}(n+d)$ be the BLUP in an orthogonal LRM. Then*

$$\mathrm{MSE}_\nu\big[X_\nu^*(n+d)\big] \le \mathrm{MSE}_{\sigma^2}\big[\hat{X}(n+d)\big]$$

for every $\nu = (\sigma^2, \sigma_1^2, \dots \sigma_l^2)' \in \Upsilon$.

This result shows that the choice of the predictor $\hat{X}(n+d)$, assuming that the data follows an LRM, is not as good as the choice of the predictor $X_\nu^*(n+d)$, that is the BLP, assuming that observation X is given by an FDSM if parameter ν of the FDSM, the variances σ^2 and $\sigma_j^2; j = 1, 2, \dots, l$, are known. The reason for this result is that in an FDSM we have some limitations on random "regression parameters" consisting in the fact that they have finite variances and thus realizations of these "parameters" are in some sense bounded, while in the classical LRM we have no restrictions on regression parameters.

Now we shall consider the case when time series data come from an LRM, but we use the predictor $X_\nu^*(n+d)$ with some arbitrary, but fixed, value of parameter ν. It is easily to deduce the following lemma.

Lemma 3.2. *For the predictors $X^*_\nu(n+d)$ and $\hat{X}(n+d)$ the following equality holds*

$$X^*_\nu(n+d) = \hat{X}(n+d) - \sum_{i=1}^{l} \frac{\sigma^2/\sigma_i^2}{\|f_i\|^2\left(\sigma^2/\sigma_i^2 + \|f_i\|^2\right)} f_i(n+d) f_i' X \, . \tag{3.2}$$

Using this lemma in an orthogonal LRM, we get the expression

$$\begin{aligned}
\mathrm{MSE}_{(\beta,\sigma^2)}\left[X^*_\nu(n+d)\right] = {} & \mathrm{MSE}_{\sigma^2}\left[\hat{X}(n+d)\right] \\
& + D_{\sigma^2}\left[\sum_{i=1}^{l} \frac{f_i(n+d)\sigma^2/\sigma_i^2}{\|f_i\|^2\left(\sigma^2/\sigma_i^2 + \|f_i\|^2\right)} f_i' X\right] \\
& + \left(E_\beta\left[\sum_{i=1}^{l} \frac{f_i(n+d)\sigma^2/\sigma_i^2}{\|f_i\|^2\left(\sigma^2/\sigma_i^2 + \|f_i\|^2\right)} f_i' X\right]\right)^2 \\
& - 2\sum_{i=1}^{l} \frac{f_i(n+d)\sigma^2/\sigma_i^2}{\|f_i\|^2\left(\sigma^2/\sigma_i^2 + \|f_i\|^2\right)} \\
& \times \mathrm{Cov}_{\sigma^2}\left(\hat{X}(n+d) - X(n+d); f_i' X\right).
\end{aligned}$$

Since $f_i' X; i = 1, 2, \ldots, l$ are mutually uncorrelated with variances $D_{\sigma^2}\left[f_i' X\right] = \sigma^2 \|f_i\|^2$ if X is generated by an orthogonal LRM, we get

$$D_{\sigma^2}\left[\sum_{i=1}^{l} \frac{f_i(n+d)\sigma^2/\sigma_i^2}{\|f_i\|^2\left(\sigma^2/\sigma_i^2 + \|f_i\|^2\right)} f_i' X\right] = \sigma^4 \sum_{i=1}^{l} \frac{f_i^2(n+d)\sigma^2}{\|f_i\|^2\left(\sigma^2 + \sigma_i^2\|f_i\|^2\right)^2},$$

$$E_\beta\left[\sum_{i=1}^{l} \frac{f_i(n+d)\sigma^2/\sigma_i^2}{\|f_i\|^2\left(\sigma^2/\sigma_i^2 + \|f_i\|^2\right)} f_i' X\right] = \sigma^2 \sum_{i=1}^{l} \frac{f_i(n+d)}{\sigma^2 + \sigma_i^2\|f_i\|^2}\beta_i$$

and

$$\begin{aligned}
\sum_{i=1}^{l} \frac{f_i(n+d)\sigma^2/\sigma_i^2}{\|f_i\|^2\left(\sigma^2/\sigma_i^2 + \|f_i\|^2\right)} & \mathrm{Cov}_{\sigma^2}\left(\hat{X}(n+d) - X(n+d); f_i' X\right) \\
& = \sigma^4 \sum_{i=1}^{l} \frac{f_i^2(n+d)}{\|f_i\|^2\left(\sigma^2 + \sigma_i^2\|f_i\|^2\right)},
\end{aligned}$$

since, as it can be easily seen,

$$\mathrm{Cov}_{\sigma^2}\left(\hat{X}(n+d) - X(n+d); f_i' X\right) = \sigma^2 f_i(n+d).$$

The above derived results are collected in the following theorem.

Theorem 3.2. *In an orthogonal LRM*

$$X = F\beta + w$$

the MSEs of the predictor $X^*_\nu(n+d)$ are given for every $\beta \in E^l$ and every $\nu \in \Upsilon$ by

$$\begin{aligned} \mathrm{MSE}_{(\beta,\sigma^2)}\big[X^*_\nu(n+d)\big] = {} & \mathrm{MSE}_{\sigma^2}[\hat{X}(n+d)] \\ & - \sigma^4 \sum_{i=1}^{l} \frac{f_i^2(n+d)}{\|f_i\|^2 \left(\sigma^2 + \sigma_i^2 \|f_i\|^2\right)} \\ & - \sigma^4 \sum_{i=1}^{l} \sigma_i^2 \frac{f_i^2(n+d)}{\left(\sigma^2 + \sigma_i^2 \|f_i\|^2\right)^2} \\ & + \sigma^4 \left(\sum_{i=1}^{l} \frac{f_i(n+d)}{\sigma^2 + \sigma_i^2 \|f_i\|^2} \beta_i \right)^2 . \end{aligned}$$

We can see that the mean squared errors, $\mathrm{MSE}_{(\beta,\sigma^2)}\big[X^*_\nu(n+d)\big]$, of the predictor $X^*_\nu(n+d)$, used under the assumption that X is generated by an orthogonal LRM, depends also on parameter β, since $X^*_\nu(n+d)$ is a biased predictor in an LRM. The influence of values of β on the increase of the $\mathrm{MSE}_{(\beta,\sigma^2)}\big[X^*_\nu(n+d)\big]$ can be significant for a fixed (small) n, but in many regression models this influence is negligible for large values of n.

We can also rewrite (3.2.) as

$$\hat{X}(n+d) = X^*_\nu(n+d) + \sum_{i=1}^{l} \frac{f_i(n+d)\sigma^2/\sigma_i^2}{\|f_i\|^2 \left(\sigma^2/\sigma_i^2 + \|f_i\|^2\right)} f_i' X \,,$$

from which we have

$$\begin{aligned} \mathrm{MSE}_\nu\big[\hat{X}(n+d)\big] = {} & \mathrm{MSE}_\nu\big[X^*_\nu(n+d)\big] \\ & + E_\nu \left[\sum_{i=1}^{l} \frac{f_i(n+d)\sigma^2/\sigma_i^2}{\|f_i\|^2 \left(\sigma^2/\sigma_i^2 + \|f_i\|^2\right)} f_i' X \right]^2 \\ & + 2 \sum_{i=1}^{l} \frac{f_i(n+d)\sigma^2/\sigma_i^2}{\|f_i\|^2 \left(\sigma^2/\sigma_i^2 + \|f_i\|^2\right)} \\ & \times \mathrm{Cov}_\nu\big(X^*_\nu(n+d) - X(n+d); f_i' X\big). \end{aligned}$$

Since $f_i' X; i = 1, 2, \ldots, l$ are, in an orthogonal FDSM, mutually uncorrelated random variables with variances

$$D_\nu \left[f_i' X\right] = \sigma_i^2 \left\|f_i\right\|^2 \left(\sigma^2/\sigma_i^2 + \left\|f_i\right\|^2\right),$$

we can write

$$E_\nu \left[\sum_{i=1}^{l} \frac{f_i(n+d)\sigma^2/\sigma_i^2}{\left\|f_i\right\|^2 \left(\sigma^2/\sigma_i^2 + \left\|f_i\right\|^2\right)} f_i' X\right]^2 = \sigma^4 \sum_{i=1}^{l} \frac{f_i^2(n+d)}{\left\|f_i\right\|^2 \left(\sigma^2 + \sigma_i^2 \left\|f_i\right\|^2\right)}.$$

Next it is easy to verify that

$$\mathrm{Cov}_\nu\left(X_\nu^*(n+d) - X(n+d); f_i' X\right) = 0$$

and thus we have shown that the following theorem is true.

Theorem 3.3. *In an orthogonal FDSM*

$$X = FY + w$$

the MSEs of the predictor $\hat{X}(n+d)$ are given by

$$\mathrm{MSE}_\nu\left[\hat{X}(n+d)\right] = \mathrm{MSE}_\nu\left[X_\nu^*(n+d)\right] + \sigma^4 \sum_{i=1}^{l} \frac{f_i^2(n+d)}{\left\|f_i\right\|^2 \left(\sigma^2 + \sigma_i^2 \left\|f_i\right\|^2\right)}; \quad \nu \in \Upsilon.$$

Consequence. From the results derived above we have, for orthogonal models, for every $\beta \in E^k$ and every $\nu \in \Upsilon$, the following statements: if

$$\lim_{n\to\infty} \frac{f_i(n+d)}{\left\|f_i\right\|} = 0,$$

then

$$\begin{aligned}
\lim_{n\to\infty} \mathrm{MSE}_{\sigma^2}\left[\hat{X}(n+d)\right] &= \lim_{n\to\infty} \mathrm{MSE}_\nu\left[X_\nu^*(n+d)\right] \\
&= \lim_{n\to\infty} \mathrm{MSE}_\nu\left[\hat{X}(n+d)\right] \\
&= \lim_{n\to\infty} \mathrm{MSE}_{(\beta,\sigma^2)}\left[X_\nu^*(n+d)\right] \\
&= \sigma^2.
\end{aligned}$$

We see that in orthogonal models the choice of the predictor $\hat{X}(n+d)$ is, from the asymptotic point of view, equivalent with the choice of the predictor $X_\nu^*(n+d)$, since, under the weak conditions, these two predictors have in the both considered models the same asymptotic mean squared errors for all parameters β and ν.

In practical applications of the kriging theory we have only time series data, and thus all parameters of both regression models that we use for modeling the given time series data, should be estimated from these data. Thus it is necessary to study the mean squared errors in both regression models of the predictor $X^*_{\tilde{\nu}}(n+d)$, where the unknown parameter ν is replaced by an estimator $\tilde{\nu}(X)$ of ν, and to compare these mean squared errors with mean squared errors of the predictor $\hat{X}(n+d)$.

The predictor $X^*_{\tilde{\nu}}(n+d)$ is called the *empirical* BLUP. Some results on mean squared error of the empirical BLUP can be found in Christensen (1992), Harville, Jeske (1992), Štulajter (2002), Das, Jiang, Rao (2004), Štulajter (2007a) and others. Methods of estimation of variances are given in Harville (1977), Searle, Casella, McCulloch (1992), Štulajter, Witkovský (2004) and others.

REFERENCES

[1] CHRISTENSEN, R.: *Linear Models for Multivariate, Time Series and Spatial Data*, Springer–Verlag, New York, 1991.

[2] DAS, K.—JIANG, J.—RAO, J. N. K.: *Mean squared error of empirical predictor*, Ann. Statist., **32** (2004), 818–840.

[3] GOLDBERGER, A. S.: *Best linear unbiased prediction in the generalized linear regression model*, J. Amer. Statist. Assoc. **57** (1962), 369-375.

[4] HARVILLE, D. A.: *Maximum likelihood approach to variance components estimation and related problems*, J. Amer. Statist. Assoc. **72** (1977), 320–340.

[5] HARVILLE, D. A.—JESKE, D. R.: *Mean squared error of estimation or prediction under a general linear model*, J. Amer. Statist. Assoc. **87** (1992), 724–731.

[6] SEARLE, S. R.—CASELLA, G.—MCCULLOCH, C. E.: *Variance Components*, Wiley, New York, 1992.

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

[8] ŠTULAJTER, F.: *Predictions in Time Series Using Regression Models*, Springer–Verlag, New York, 2002.

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

[10] ŠTULAJTER, F.—WITKOVSKÝ, V.: *Estimation of variances in orthogonal* FDSLRMs, Metrika **60** (2004), 105–118.

[11] ŠTULAJTER, F.: MSE *of the best predictor in nonorthogonal models*, Math. Slovaca **57** (2007), 83–91.

[12] ŠTULAJTER, F.: *Mean sqaured error of the empirical best linear unbiased predictor in an orthogonal finite discrete spectrum linear regression model*, Metrika **65** (2007a), 331–348.

Received September 21, 2006

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