Comparison of prediction quality of the best linear unbiased predictors in time series linear regression models

Martina Hančová
Institute of Mathematics, P. J. Šafárik University in Košice
Jesenná 5, 040 01 Košice, Slovakia

Keywords: Time series. Finite discrete spectrum linear regression models. Best linear unbiased prediction. Mean squared error. Löwner partial ordering.
AMS: 62M10, 62M20

Abstract

Using a matrix-block version of the mean squared error of the best linear unbiased predictor (BLUP) and the Löwner partial order relation for symmetric matrices, we derive and present our results dealing with a comparison of prediction quality of BLUP in different types of time series linear regression models which belong to a general class of the so-called finite discrete spectrum linear regression models (FDSLRM). Some of the results are completely new and some of them are generalization of recent results achieved in special — orthogonal cases of FDSLRM.

1. Introduction

Forecasting future values of a time series from its observation up to the present described by prediction theory belongs to classical and crucial problems of the statistical inference from time series data.

One of very important predicting approaches, known as kriging, is based on modeling time series by linear regression models (standard or mixed) and the idea of the best linear unbiased predictor (BLUP), which minimizes the mean squared error (MSE) of prediction among all linear unbiased predictors (see e.g. Goldberger 1962, Brockwell & Davis 1991, Stein 1999, Christensen 2001 or Štulajter 2002).

Since a general scheme of kriging is well described in detail e.g. in Stein (1999), Christensen (2001) or Štulajter (2002, 2007) we only briefly outline basic steps of the theory. In the framework of kriging prediction theory we choose for a given time series $X(.) = \{X(t); t \in T\}; T = \{1, 2, \ldots\}$ some general class of linear regression models describing a mean value function $m(.) = \{E[X(t)]; t \in T\}$ and also a covariance function $R(., .) = \{Cov(X(s), X(t));\ s, t \in T\}$ of time series $X(.)$. Then on the basis of time series observation $X = (X(1), X(2), ..., X(n))$ we find and study a form of the BLUP $X^*(n + d)$ of a future value $X(n + d)$ of $X(.)$ and its mean squared error to identify the most appropriate model for prediction from the considered class of models.

My research is connected to a recently introduced and investigated general class of linear regression time series models called *the finite discrete spectrum linear regression models* (see e.g. Štulajter 2007). The definition of the models is the following: A model of time series $X(.)$ is said to be *the finite discrete spectrum linear regression model* (FDSLRM), if $X(.)$ satisfies

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

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 regression parameters;

$Y = (Y_1, Y_2, \ldots, Y_l)'$ is a $l \times 1$ random vector with $E[Y] = 0$, $Cov(Y) = diag(\sigma_j^2)$ of size $l \times l$, where variances $\sigma_j^2 \geq 0$ for all $j = 1, 2, \ldots, l$;

$f_i(.)$ and $v_j(.)$ are known real functions defined on $\mathbb{E}$;

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

In the paper my attention is given to the general solution of the particular problem resolved recently by me during Ph.D. study — comparison MSEs or in other words prediction quality of BLUPs in various models belonging to the class of FDSLRM models like the *classical linear regression* models (CLRM) — FDSLRM with $l = 0, k > 0$, *finite discrete spectrum* models (FDSM)— FDSLRM with $k = 0, l > 0$, or remaining *general* FDSLRMs with $k, l > 0$.

In practice the problem is important. If we have time series dataset recorded over time which we want to describe by FDSLRM with known regression functions $f_i(.)$ $v_j(.)$, it is impossible to decide what model — CLRM, FDSM or GFDSLRM — generates the data, which can be viewed as observed values of any mentioned model.

2. Comparing prediction quality of BLUPs for various FDSLRMs

Although the general theory of best linear predictors gives us explicit expression for BLUP of $X(n+d)$ in any linear regression model (see e.g. in Christensen 2001 or Štulajter 2007), the BLUP and its MSE depends on $Cov_\nu(X, X(n+d))$ and the inverse of $Cov_\nu(X) \equiv \Sigma_\nu$, where ν represents variance-covariance parameters, in our case we consider $\nu = (\sigma^2, \sigma_1^2, \ldots, \sigma_l^2) \in (0, \infty)^{l+1}$.

This dependence makes effective comparison of the BLUP quality for different types of FDSLRM impossible. However our research led to 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.

Theorem 2.1. (Block matrix form of the BLUP for a general FDSLRM)
Let us consider a general FDSLRM (1), $k, l \geq 1$, *with its corresponding observation* X:

$$X = F\beta + \varepsilon, \quad E[\varepsilon] = 0, \quad Cov_\nu(X) = \Sigma_\nu = \sigma^2 I + VDV',$$

where $\beta \in \mathbb{E}^k, F = (f_1\ f_2\ ...\ f_k)$ *is the design matrix of the model with columns* $f_i = (f_i(1), ..., f_i(n))'$, $V = (v_1\ v_2\ ...\ v_l)$ *is the matrix describing structure of* $Cov_\nu(X)$ *with columns* $v_j = (v_j(1), v_j(2), ..., v_j(n))'$ *and* $D = diag(\sigma_j^2) > 0$. *Let*

$$E[X(n+d)] = f'\beta, \quad f = (f_1(n+d),\ ..., f_k(n+d))',$$
$$Cov_\nu(X, X(n+d)) = VDv, \quad v = (v_1(n+d),\ ..., v_l(n+d))'.$$

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

$$X_\nu^*(n+d) = z'\mathbb{G}^{-1}Z'X \tag{2}$$

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

$$MSE[X_\nu^*(n+d)] = \sigma^2(1 + z'\mathbb{G}^{-1}z). \tag{3}$$

Substitution for z, Z and $\mathbb{G}$ yields these expressions for the BLUP and its MSE in the general FDSLRM:

$$\begin{aligned} X_\nu^*(n+d) &= \begin{pmatrix} f \\ v \end{pmatrix}' \begin{pmatrix} F'F & F'V \\ V'F & V'V + \sigma^2 D^{-1} \end{pmatrix}^{-1} \begin{pmatrix} F'X \\ V'X \end{pmatrix}, \\ MSE[X_\nu^*(n+d)] &= \sigma^2 \left[1 + \begin{pmatrix} f \\ v \end{pmatrix}' \begin{pmatrix} F'F & F'V \\ V'F & V'V + \sigma^2 D^{-1} \end{pmatrix}^{-1} \begin{pmatrix} f \\ v \end{pmatrix} \right]. \end{aligned} \tag{4}$$

Due to the limited number of pages for the paper we can present only very briefly what basic idea of the theorem's proof is. There are at least two ways of proof: one using Henderson's mixed model equations and some results for linear mixed models (McLean et al 1991) or the second one is in terms of the Banachiewicz formula and properties of Schur complements (Zhang 2005).

The derived block matrix form (4) allows us to employ properties of the Löwner partial ordering, an ordering relation for symmetric matrices, defined in the following way[1]:

$$A \geq B \ (A > B) \text{ means nonegative (positive) definite } A - B. \tag{5}$$

[1] An introduction to basic properties of the relation sufficient for our statistical purposes is e.g. in Wang & Chow (1993)

Or equivalently $A \geq B$ $(A > B)$, iff for any nonzero vector f of the adequate size $f'Af \geqq f'Bf$ $(f'Af > f'Bf)$.

Finally, to write out effectively our main result — general comparison theorem, which compares all considered FDSLRM models, and of which assertions directly results from our block matrix form of BLUP in FDSLRM and Löwner partial ordering, we need to establish a new notation. Let for now k and l be fixed and satisfying $1 \leqq k < l$, let F have l columns, $f \in \mathbb{E}^l$ and $D_* = diag(\sigma^2/\sigma_j^2)$ be of size $l \times l$. We partition matrix $F = (f_1\ f_2\ ...\ f_l)$ and D_* in such way that the following block matrices are formed:

$$D_* = \begin{pmatrix} D_k & 0 \\ 0 & D_{k+1} \end{pmatrix}, \quad F'F = \begin{pmatrix} F_k'F_k & F_k'F_{k+1} \\ F_{k+1}'F_k & F_{k+1}'F_{k+1} \end{pmatrix}, \quad f = \begin{pmatrix} f_{(k)} \\ f_{(k+1)} \end{pmatrix}$$

where matrices $F_k = (f_1\ ...\ f_k), F_{k+1} = (f_{k+1}\ ...\ f_l)$, $f = (f_{(k)}', f_{(k+1)}')'$ with $f_{(k)} = (f_1(n+d), ..., f_k(n+d))'$ and $f_{(k+1)} = (f_{k+1}(n+d), ..., f_l(n+d))'$ and finally diagonal D_k has elements $(D_k)_{jj} = \sigma^2/\sigma_j^2, j = 1, ..., k$ and diagonal D_{k+1} elements $(D_{k+1})_{jj} = \sigma^2/\sigma_j^2, j = k+1, ..., l$.

Theorem 2.2. (Comparison of MSEs of BLUPs in CLRM, FDSM, GFDSLRM) *Consider the folllowing three models:*

$$\begin{aligned}
&\textit{CLRM:} && X(t) = \sum_{i=1}^{l} \beta_i f_i(t) + w(t), D(w(t)) = \sigma^2, \\
& && \textit{for which } X_\nu^*(n+d) = f'(F'F)^{-1}F'X, F = (f_1, f_2, ..., f_l). \\
&\textit{FDSM:} && X(t) = \sum_{j=1}^{l} Y_j f_j(t) + w(t), Cov(Y_j) = diag(\sigma_j^2) \\
& && \textit{for which } X_\nu^*(n+d) = f'\left[F'F + diag(\sigma^2/\sigma_j^2)\right]^{-1} F'X. \\
&\textit{GFDSLRM:} && X(t) = \sum_{i=1}^{k} \beta_i f_i(t) + \sum_{j=k+1}^{l} Y_j f_j(t) + w(t), 1 \leq k < l, \\
& && \textit{for which } X_\nu^*(n+d) = f' \begin{bmatrix} F_k'F_k & F_k'F_{k+1} \\ F_{k+1}'F_k & F_{k+1}'F_{k+1} + D_{k+1} \end{bmatrix}^{-1} F'X.
\end{aligned}$$

If matrix F is of full column rank, $f \neq 0$ and $k, l, n \in \mathbb{N} : 1 \leq k < l, k+l+1 < n$, then the following inequalities hold:

(i) $MSE_{FDSM} X_\nu^*(n+d) \leq MSE_{GFDSLRM} X_\nu^*(n+d)$ *and the equality holds if and only if* $f_{(k)} = F_k'F_{k+1}[F_{k+1}'F_{k+1} + D_{k+1}]^{-1} f_{(k+1)}$.

(ii) $MSE_{FDSLRM} X_\nu^*(n+d) \leq MSE_{CLRM} X_\nu^*(n+d)$, *and the equality holds if and only if* $f_{(k+1)} = F_{k+1}'F_k[F_k'F_k]^{-1} f_{(k)}$.

(iii) *for* $f_{(k+1)} \neq F_{k+1}'F_k[F_k'F_k]^{-1} f_{(k)}$, $f_{(k)} \neq F_k'F_{k+1}[F_{k+1}'F_{k+1} + D_{k+1}]^{-1} f_{(k+1)}$

$$MSE_{FDSM} X_\nu^*(n+d) < MSE_{GFDSLRM} X_\nu^*(n+d) < MSE_{CLRM} X_\nu^*(n+d)$$

At the end we briefly summarize the results and next important corollaries following from the theorem: (1) From the viewpoint of prediction any FDSLRM model with $k, l > 0$ is always better in applications than corresponding CLRM with the same regressors, because conditions of the same quality practically never holds;[2] (2) FDSM is always better than CLRM; (3) there is the upper bound for MSE of BLUPs in FDSLRM models equal to $\sigma^2(1 + z'G^{-1}z)$.

Acknowledgements: This work was supported by the grant 1/3001/06 of the Slovak Scientific Grant Agency (VEGA) and the grant 25/07-08 of VVGS at P. J. Šafárik University in Košice.

3. Bibliography

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

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

[3] Goldberger, A.S. (1962) Best linear unbiased prediction in the generalized linear regression models. *J. Am. Stat. Assoc.* **57**: 369-375.

[4] McLean, R.A., Sanders, W.L., Stroup W.W. (1991) A Unified Approach to Mixed Linear Models, *Am. Stat.* **45**: 54-64.

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

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

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

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

[9] Wang, S.– G., Chow, S.– Ch. (1993) *Advanced linear models: Theory and applications.* Marcel Dekker, New York.

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

[2]For example in the so-called orthogonal FDSLRM models for which $F'_{k+1}F_k = 0$ and $F'_kF_{k+1} = 0$ the condition is equivalent with the conditions $f_{(k)} = 0$ and $f_{(k+1)} = 0$ or shortly $f = 0$, which means that all regressors should have to give us zero value at time $n + d$.