Tatra Mt. Math. Publ. **22** (2001), 159–170

TATRA MOUNTAINS
Mathematical Publications

REGRESSION MODELS AND TIME SERIES PREDICTIONS

František Štulajter

Dedicated to Lubomír Kubáček on the occasion of his 70th birthday

ABSTRACT. Predictions in time series using regression models are studied with respect to their mean squared errors. Methods of prediction are proposed in linear multivariate and some nonlinear models. A model with seasonal and cyclical components is also studied. The article is a survey of results of preceding papers.

1. Introduction

There are many possibilities for prediction of time series. The well known and commonly used are the methods based on Box-Jenkins (1976) methodology. The other approach is based on modeling time series by *linear regression models* and finding *the best linear unbiased predictor* (*the* BLUP). This method is known in an engineering literature as *kriging*, see Goldberger (1962) or Christensen (1991).

The kriging theory is based on the following assumption. Let $\mathbf{X} = \big(X(1), \dots, \dots, X(n)\big)'$ be a finite observation of a time series X and let U be a predicted random variable (usually $U = X(s)$; $s > n$). The prediction regression model is in the kriging theory given by

$$\mathbf{X} = \mathbf{F}\boldsymbol{\beta} + \boldsymbol{\varepsilon}; \quad E[\boldsymbol{\varepsilon}] = \mathbf{0}; \quad E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'] = \boldsymbol{\Sigma},$$
$$E[U] = \mathbf{f}'\boldsymbol{\beta}; \quad \mathrm{Cov}(U; \mathbf{X}) = \mathbf{r},$$

where $\mathbf{F}$ is a known $n \times k$ matrix, $\mathbf{f}$ a known $k \times 1$ vector and $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \dots, \beta_k)' \in E^k$ are unknown regression parameters. If $\mathbf{r}$ and $\boldsymbol{\Sigma}$ are known, then the BLUP U^* of U is given by

$$U^* = \mathbf{f}'\boldsymbol{\beta}^* + \mathbf{r}'\boldsymbol{\Sigma}^{-1}(\mathbf{X} - \mathbf{F}\boldsymbol{\beta}^*)$$

2000 Mathematics Subject Classification: Primary 62M10; Secondary 63J02, 62J05.
Keywords: linear regression model, nonlinear regression model, best predictor, mean squared error of predictor.

Supported by the VEGA grant of the Slovak Grant Agency No. 1/4196/97.

and has *the mean squared error* (*the* MSE)

$$E[U^* - U]^2 = D[U] - \mathbf{r}'\boldsymbol{\Sigma}^{-1}\mathbf{r} + \left\|\mathbf{f} - \mathbf{F}'\boldsymbol{\Sigma}^{-1}\mathbf{r}\right\|^2_{\Sigma_{\boldsymbol{\beta}^*}},$$

where $\|\mathbf{a}\|^2_{\mathbf{A}} = \mathbf{a}'\mathbf{A}\mathbf{a}$, $\boldsymbol{\beta}^* = (\mathbf{F}'\boldsymbol{\Sigma}^{-1}\mathbf{F})^{-1}\mathbf{F}'\boldsymbol{\Sigma}^{-1}\mathbf{X}$ is *the* BLUE of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\beta}^*}$ is a covariance matrix of $\boldsymbol{\beta}^*$.

If $\widehat{\boldsymbol{\beta}} = (\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'\mathbf{X}$ is *the* OLSE of $\boldsymbol{\beta}$ and

$$\widehat{U} = \mathbf{f}'\widehat{\boldsymbol{\beta}} + \mathbf{r}'\boldsymbol{\Sigma}^{-1}(\mathbf{X} - \mathbf{F}\widehat{\boldsymbol{\beta}}),$$

then

$$E[\widehat{U} - U]^2 = D[U] - \mathbf{r}'\boldsymbol{\Sigma}^{-1}\mathbf{r} + \left\|\mathbf{f} - \mathbf{F}'\boldsymbol{\Sigma}^{-1}\mathbf{r}\right\|^2_{\boldsymbol{\Sigma}_{\widehat{\beta}}},$$

where $\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\beta}}}$ is a covariance matrix of $\widehat{\boldsymbol{\beta}}$. Obviously we have

$$E[U^* - U]^2 \le E[\widehat{U} - U]^2.$$

The aim of this article is to study a method of prediction using multivariate and nonlinear regression models. This method is based on an idea to predict future values of observed time series using multivariate regression models with estimated and predicted regression parameters. A multivariate regression model with seasonal and cyclical component is also studied as a special case of a nonlinear regression model.

Let us assume, for example, that we observe a time series $X_i(t)$, $t = 1, 2, \dots, 12$ and $i = 1, 2, \dots, R$ which consists of monthly observations of some real time series during R years. Mainly in financial time series of observations of such kind we can see that the structure of observations is very similar but changes from year to year. Thus for modeling the whole observed time series by a regression model it is convenient to use a multivariate rather than a univariate regression model.

A multivariate regression model can generally be described, see Christensen (1991), as a set of linear models

$$\mathbf{X}_i = \mathbf{F}\boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i; \quad i = 1, 2, \dots, R,$$

where we assume that $\mathbf{X}_i = \big(X_i(1), \dots, X_i(n)\big)'$, $i = 1, 2, \dots, R$, are $n \times 1$ random vectors of observations, $\mathbf{F}$ is a known $n \times k$ design matrix and $\boldsymbol{\beta}_i = (\beta_{i1}, \beta_{i2}, \dots, \beta_{ik})'$; $i = 1, 2, \dots, R$, are unknown $k \times 1$ regression parameters. The random $n \times 1$ error vectors $\boldsymbol{\varepsilon}_i$; $i = 1, 2, \dots, R$, are assumed to have $E[\boldsymbol{\varepsilon}_i] = 0$ for all $i = 1, 2, \dots, R$. The multivariate linear model consists of fitting R linear models simultaneously. This can also be done as follows.

Let $\mathbf{X} = (\mathbf{X}'_1, \dots, \mathbf{X}'_R)'$, $\boldsymbol{\beta} = (\boldsymbol{\beta}'_1, \dots, \boldsymbol{\beta}'_R)'$ and let $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}'_1, \dots, \boldsymbol{\varepsilon}'_R)'$. Then we can write the multivariate regression model as a univariate one

$$\mathbf{X} = (\mathbf{I} \otimes \mathbf{F})\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad E[\boldsymbol{\varepsilon}] = 0.$$

Next we shall assume that

$$E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'] = \boldsymbol{\Sigma}_R \otimes \boldsymbol{\Sigma}_n ,$$

where $\boldsymbol{\Sigma}_R$ is a positive definite covariance matrix and $\otimes$ denotes Kronecker product. The matrices $\boldsymbol{\Sigma}_R$ and $\boldsymbol{\Sigma}_n$ are assumed to be known in this article except the last paragraph. From the last assumption we have $\mathrm{Cov}(\mathbf{X}_i, \mathbf{X}_j) = \Sigma_{R,ij}\boldsymbol{\Sigma}_n$. If we assume that the random vectors $\mathbf{X}_i$; $i = 1, 2, \dots, R$, are independent then $E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'] = \mathbf{I}_R \otimes \boldsymbol{\Sigma}_n$.

Using the following well known properties of Kronecker product of matrices and vectors

$$(\mathbf{A} \otimes \mathbf{B})' = \mathbf{A}' \otimes \mathbf{B}', \qquad (\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$$
$$\text{and} \quad (\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC} \otimes \mathbf{BD})$$

it is easy to see that the BLUP $\boldsymbol{\beta}^* = (\boldsymbol{\beta}_1^*, \dots, \boldsymbol{\beta}_R^*)'$ of $\boldsymbol{\beta}$ is given by

$$\boldsymbol{\beta}_i^* = (\mathbf{F}'\boldsymbol{\Sigma}_n^{-1}\mathbf{F})^{-1}\mathbf{F}'\boldsymbol{\Sigma}_n^{-1}\mathbf{X}_i\,; \quad i = 1, 2, \dots, R$$

and thus we have that the OLSEs $\widehat{\boldsymbol{\beta}}_i$ are given by

$$\widehat{\boldsymbol{\beta}}_i = (\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'\mathbf{X}_i; \quad i = 1, 2, \dots, R.$$

A nonlinear regression model, see P á z m a n (1993), is given by

$$\mathbf{X} = \mathbf{m}_{\boldsymbol{\theta}} + \boldsymbol{\varepsilon}; \quad \boldsymbol{\theta} \in \boldsymbol{\Theta}; \quad E[\boldsymbol{\varepsilon}] = 0\,, \quad E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'] = \boldsymbol{\Sigma}\,,$$

where the mean value vector $\mathbf{m}_{\boldsymbol{\theta}}$ of the $n \times 1$ random vector $\mathbf{X}$ depends nonlinearly on $\boldsymbol{\theta}$. For a predicted random variable $U = X(s)$ we assume that

$$E_{\boldsymbol{\theta}}[U] = m_{\boldsymbol{\theta}}(s); \quad \boldsymbol{\theta} \in \boldsymbol{\Theta}, \quad \mathrm{Cov}(U\,; \mathbf{X}) = \mathbf{r}$$

for some time s, $s > n$. The OLSE of $\boldsymbol{\theta}$ will be denoted by $\widehat{\boldsymbol{\theta}}$.

2. Predictions in multivariate regression models

Let us assume that $\mathbf{X} = (\mathbf{X}_1', \dots, \mathbf{X}_R')'$ is the observed time series modelled by the multivariate regression model and let $\mathbf{X}_{R+1} = \big(X_{R+1}(1), \dots, X_{R+1}(n)\big)'$ is the future observation (in the $(R+1)$th year) which should be predicted. We shall assume that

$$E[\mathbf{X}_{R+1}] = \mathbf{F}\boldsymbol{\beta}_{R+1}\,.$$

This means that if U is some of $X_{R+1}(t)$ then we assume

$$E[U] = \mathbf{f}'\boldsymbol{\beta}_{R+1}$$

for some given vector $\mathbf{f} = (f_1, \dots, f_k)'$.

Now let $\widetilde{\boldsymbol{\beta}}_1, \dots, \widetilde{\boldsymbol{\beta}}_R$ be some estimators of $\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_R$. Since in practice we do not know the covariance matrix of errors, we usually take $\widetilde{\boldsymbol{\beta}}_i = \widehat{\boldsymbol{\beta}}_i$; $i = 1, 2, \dots, R$. Let $\widetilde{\boldsymbol{\beta}}_i = (\widetilde{\beta}_{i1}, \dots, \widetilde{\beta}_{ik})'$; $i = 1, 2, \dots, R$, and let

$$\widetilde{\boldsymbol{\beta}}_{.j} = (\widetilde{\beta}_{1j}, \dots, \widetilde{\beta}_{Rj})'$$

be the $R \times 1$ vector of estimators of the jth component of $\widetilde{\boldsymbol{\beta}}_1, \dots, \widetilde{\boldsymbol{\beta}}_R$ for $j = 1, 2, \dots, k$.

Let us assume that

$$E[\widetilde{\boldsymbol{\beta}}_{.j}] = \mathbf{H}\boldsymbol{\gamma}_j; \quad j = 1, 2, \dots, k,$$

where $\mathbf{H}$ is a known $R \times l$ matrix $\boldsymbol{\gamma}_j = (\gamma_{j1}, \dots, \gamma_{jl})'$; $j = 1, 2, \dots, k$, are unknown regression parameters.

Define $\widetilde{\boldsymbol{\beta}}_{R+1} = (\widetilde{\beta}_{R+11}, \dots, \widetilde{\beta}_{R+1k})'$ by

$$\widetilde{\beta}_{R+1j} = \mathbf{h}'\widetilde{\gamma}_j,$$

where $\mathbf{h} = (h_1, \dots, h_l)'$ is a given $l \times 1$ vector and

$$\widetilde{\boldsymbol{\gamma}}_j = (\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'\widetilde{\boldsymbol{\beta}}_{.j}; \quad j = 1, 2, \dots, k.$$

EXAMPLE 1. *Double linear trend prediction.* Let us assume that

$$E\big[X_i(t)\big] = \beta_{i1} + \beta_{i2}t; \qquad t = 1, 2, \dots, n, \quad i = 1, 2, \dots, R+1.$$

Then

$$\mathbf{F} = \begin{pmatrix} 1 & 1 & \cdot & \cdot & \cdot & 1 \\ 1 & 2 & \cdot & \cdot & \cdot & n \end{pmatrix}$$

and for $U = X_{R+1}(t)$ we have $\mathbf{f} = (1, t)'$; $t = 1, 2, \dots, n$.

Let $\widetilde{\boldsymbol{\beta}}_i = (\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'\mathbf{X}_i$; $i = 1, 2, \dots, R$, and let $\widetilde{\beta}_{.j} = (\widetilde{\beta}_{1j}, \dots, \widetilde{\beta}_{Rj})'$; $j = 1, 2$. If we take

$$\mathbf{H} = \begin{pmatrix} 1 & 1 & \cdot & \cdot & \cdot & 1 \\ 1 & 2 & \cdot & \cdot & \cdot & R \end{pmatrix}$$

and $\mathbf{h} = (1, R+1)'$, then both components $\widetilde{\beta}_{R+1j}$; $j = 1, 2$, are predicted by prolonged regression lines with parameters estimated by the OLSE.

Using the assumptions about mean values of U and $\mathbf{X}$ it is possible to show that we can write

$$E[U] = (\mathbf{f} \otimes \mathbf{h})'\boldsymbol{\gamma}$$

and

$$\mathbf{F}\boldsymbol{\beta}_i = (\mathbf{F} \otimes \mathbf{H}_{i.})\boldsymbol{\gamma},$$

where $\mathbf{H}_{i.}$ denotes the ith row of $\mathbf{H}$.

Thus we have
$$E[\mathbf{X}] = \mathbf{C}\boldsymbol{\gamma}\,,$$
where
$$\mathbf{C} = (\mathbf{C}_1', \dots \mathbf{C}_R')'; \quad \mathbf{C}_i = \mathbf{F} \otimes \mathbf{H}_{i.}; \qquad i = 1, 2, \dots, R$$
and
$$\operatorname{Cov}\Big((\mathbf{X}';\ \mathbf{X}_{R+1}')' \Big) = \begin{pmatrix} \boldsymbol{\Sigma}_1 \otimes \boldsymbol{\Sigma} & \mathbf{r}_1 \otimes \boldsymbol{\Sigma} \\ \mathbf{r}_1' \otimes \boldsymbol{\Sigma} & r_{1,R+1}\boldsymbol{\Sigma} \end{pmatrix},$$
where $\mathbf{r}_1$ is a given vector and $r_{1,R+1}$ is a given positive number. In this model we can find the BLUP U^* of U. We can see, after some computation, that
$$X_{R+1}^*(t) = \sum_{j=1}^{k} f_j \mathbf{h}'(\mathbf{H}'\boldsymbol{\Sigma}_R^{-1}\mathbf{H})^{-1}\mathbf{H}'\boldsymbol{\Sigma}_R^{-1}\boldsymbol{\beta}_{.j}^* + \sum_{i=1}^{R} \mathbf{r}_1'\boldsymbol{\Sigma}_R^{.i} \left[X_i(t) - \sum_{j=1}^{R} P_{ij}(\mathbf{F}\beta_j^*)(t) \right],$$
where $\mathbf{P} = \mathbf{H}(\mathbf{H}'\boldsymbol{\Sigma}_R^{-1}\mathbf{H})^{-1}\mathbf{H}'\boldsymbol{\Sigma}_R^{-1}$ and $\boldsymbol{\Sigma}_R^{.i}$ denotes the ith column of $\boldsymbol{\Sigma}_R^{-1}$.

Next we have
$$\begin{aligned} E\big[X_{R+1}^*(t) - X_{R+1}(t)\big]^2 &= (r_{1,R+1} - \mathbf{r}_1'\boldsymbol{\Sigma}_R^{-1}\mathbf{r}_1)\Sigma_{n,tt} \\ &\quad + \|\mathbf{f}\|^2_{(\mathbf{F}'\boldsymbol{\Sigma}_n^{-1}\mathbf{F})^{-1}} \big\|\mathbf{h} - \mathbf{H}'\boldsymbol{\Sigma}_R^{-1}\mathbf{r}_1\big\|^2_{(\mathbf{H}'\boldsymbol{\Sigma}_R^{-1}\mathbf{H})^{-1}}. \end{aligned}$$

For more details we refer to Štulajter (1999) and Štulajter (2001).

3. Predictions in nonlinear regression models

In *a nonlinear prediction regression model*
$$\mathbf{X} = \mathbf{m}_{\boldsymbol{\theta}} + \boldsymbol{\varepsilon}; \quad E[\boldsymbol{\varepsilon}] = \mathbf{0}\,, \quad E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'] = \boldsymbol{\Sigma}$$
and
$$E_{\boldsymbol{\theta}}[U] = m_{\boldsymbol{\theta}}(s); \quad \boldsymbol{\theta} \in \boldsymbol{\Theta}\,, \quad \operatorname{Cov}(U;\mathbf{X}) = \mathbf{r}$$
we shall consider the predictor $\widetilde{U}$ given by
$$\widetilde{U} = \mathbf{m}_{\widetilde{\boldsymbol{\theta}}}(s) + \mathbf{r}'\boldsymbol{\Sigma}^{-1}(\mathbf{X} - \mathbf{m}_{\widetilde{\theta}})\,,$$
where $\widetilde{\boldsymbol{\theta}}$ is some estimator of $\boldsymbol{\theta}$.

In Štulajter (1997) there are given some approximations to the $\mathrm{MSE}_{\boldsymbol{\theta}}[\widetilde{U}]$ for different estimators $\widetilde{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$. It is shown that for a nonlinear prediction regression model with Gaussian errors we can write the approximation
$$E_{\boldsymbol{\theta}_0}[\widetilde{U} - U]^2 = D[U] - \mathbf{r}'\boldsymbol{\Sigma}^{-1}\mathbf{r} + \big\|\mathbf{f}_0 - \mathbf{F}_0'\boldsymbol{\Sigma}^{-1}\mathbf{r}\big\|^2_{\widetilde{\mathbf{A}}_0\Sigma\widetilde{\mathbf{A}}_0' + \mathbf{V}_0}\,,$$

where

$$\mathbf{f}_0 = \frac{\partial \mathbf{m}_{\boldsymbol{\theta}}(s)}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$$

is a $k \times 1$ vector,

$$\mathbf{F}_0 = \frac{\partial \mathbf{m}_{\boldsymbol{\theta}}}{\partial \boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$$

is a $n \times k$ matrix and $\widetilde{\mathbf{A}}_0$ depends on $\boldsymbol{\theta}_0$ and on the type of estimator $\widetilde{\boldsymbol{\theta}}$. $\mathbf{V}_0$ is a nonnegative definite matrix depending on $\boldsymbol{\theta}_0$. For OLSE $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ we have $\widetilde{\mathbf{A}}_0 = \widehat{\mathbf{A}}_0 = (\mathbf{F}_0'\mathbf{F}_0)^{-1}\mathbf{F}_0'$ and thus we have

$$E_{\boldsymbol{\theta}_0}[\widehat{U} - U]^2 = D[U] - \mathbf{r}'\boldsymbol{\Sigma}^{-1}\mathbf{r} + \left\| \mathbf{f}_0 - \mathbf{F}_0'\boldsymbol{\Sigma}^{-1}\mathbf{r} \right\|^2_{\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\theta}}}+\mathbf{V}_0},$$

where $\boldsymbol{\Sigma}_{\widehat{\boldsymbol{\theta}}}$ is the "asymptotic" covariance matrix of $\widehat{\boldsymbol{\theta}}$ at $\boldsymbol{\theta}_0$.

4. Predictions in multivariate regression models with seasonal and cyclical components

For economical time series or time series describing natural processes it is useful instead of a multivariate regression model to use a multivariate regression model with seasonal and cyclical components

$$\mathbf{X} = (\mathbf{I} \otimes \mathbf{F})\boldsymbol{\beta} + c_{\boldsymbol{\vartheta}} + \boldsymbol{\varepsilon} = \mathbf{m}_{\boldsymbol{\theta}} + \boldsymbol{\varepsilon}; \quad \boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\vartheta}')',$$

where the vector $\mathbf{c}_{\boldsymbol{\vartheta}}$ with parameters $\boldsymbol{\vartheta}$ describes *seasonal and cyclical components* of observed time series. These can be typically modelled by

$$\mathbf{c}_{\boldsymbol{\vartheta}}(t) = \sum_{i=1}^{p} \left(\alpha_i^{(1)} \cos \lambda_i t + \alpha_i^{(2)} \sin \lambda_i t\right); \quad t = 1, 2, \ldots, nR,$$

where $\boldsymbol{\vartheta} = \left(p, \alpha_1^{(1)}, \ldots, \alpha_p^{(1)},\ \alpha_1^{(2)}, \ldots, \alpha_p^{(2)},\ \lambda_1, \ldots, \lambda_p\right)' = (p, \boldsymbol{\alpha}', \boldsymbol{\lambda}')'$ is an unknown parameter.

For the predicted random variable U we assume that

$$E[U] = \mathbf{f}'\boldsymbol{\beta}_{R+1} + \mathbf{c}_{\boldsymbol{\vartheta}}(nR + t); \quad t = 1, 2, \ldots, n.$$

As a predictor $\widetilde{U}$ of U we now take

$$\widetilde{U} = \mathbf{f}'\widetilde{\boldsymbol{\beta}}_{R+1} + \mathbf{c}_{\widetilde{\boldsymbol{\vartheta}}}(nR + t) + \mathbf{r}'\boldsymbol{\Sigma}^{-1}(\mathbf{X} - \mathbf{m}_{\widetilde{\boldsymbol{\theta}}}),$$

where $\widetilde{\boldsymbol{\beta}}_{R+1}$ and $\widetilde{\boldsymbol{\vartheta}}$ are some estimators, usually *the* OLSE of unknown parameters $\boldsymbol{\beta}_{R+1}$ and $\boldsymbol{\vartheta}$.

We remark that our model is a nonlinear model for $\mathbf{m}_{\boldsymbol{\theta}}$ and initial values for $\boldsymbol{\beta}$ and $\boldsymbol{\vartheta}$ can be found as follows. Using $\mathbf{X}$ we find the OLSE $\widehat{\boldsymbol{\beta}}$ for $\boldsymbol{\beta}$. Then we can compute the periodogram of residuals $\mathbf{X} - (\mathbf{I} \otimes \mathbf{F})\widehat{\boldsymbol{\beta}}$ and according to a well known procedure we choose $\widehat{p}$ and $\widehat{\boldsymbol{\lambda}}$. Considering

$$\mathbf{X} - (\mathbf{I} \otimes \mathbf{F})\widehat{\boldsymbol{\beta}} = \mathbf{c}_{(\widehat{p},\boldsymbol{\alpha},\widehat{\boldsymbol{\lambda}})} + \mathbf{w}$$

as a linear model with parameter $\boldsymbol{\alpha}$ we can compute the OLSE $\widehat{\boldsymbol{\alpha}}$ of $\boldsymbol{\alpha}$. The estimates $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\vartheta}}$ can be used, together with $\widehat{p}$, as initial for some iterative procedure to find *the maximum likelihood estimator* (*the* MLE) $\widetilde{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$. Then we find $\widetilde{\boldsymbol{\beta}}_{R+1}$ using the procedure described for a multivariate regression model. The MSE of $\widetilde{U}$ can be approximated using the results of the preceding section.

On the following figures there are predictors $\widetilde{X}(37)$, $\widetilde{X}(38), \dots, \widetilde{X}(48)$ using a multivariate regression model with seasonal and cyclical components with two frequencies, $p = 2$, and with estimated covariance characteristics $\mathbf{r}$ and $\boldsymbol{\Sigma}$. The covariance functions of $X's$ were assumed to be stationary and were estimated according to the expression given in the following section.

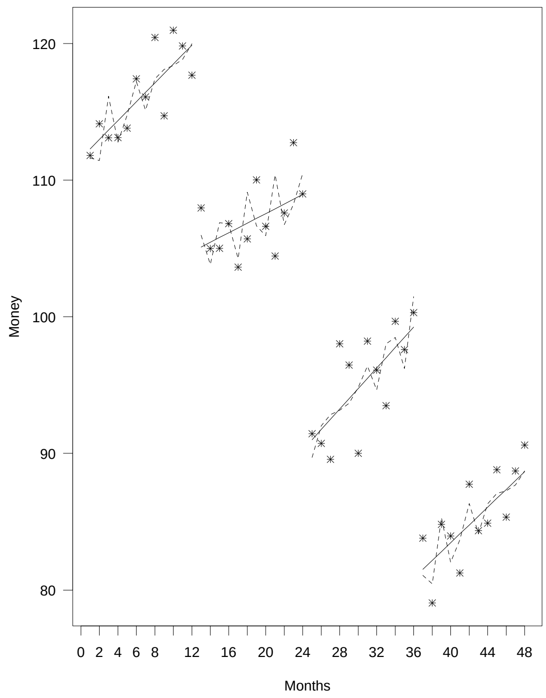

FIGURE 1. On Figure 1 there are simulated data which are considered as monthly observations during three years with white noise random errors. Predicted values were computed using also cyclical and seasonal components with two frequencies to illustrate the proposed method.

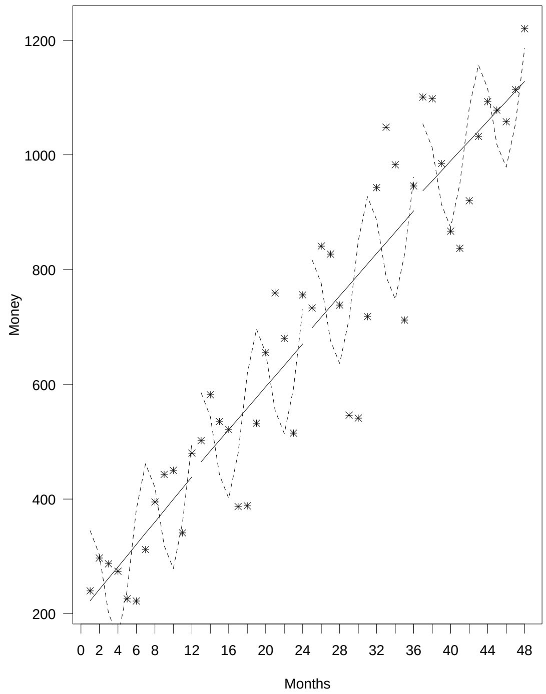

FIGURE 2. On Figure 2 there are real data of monthly observations during three years of a school office. Predictions for the fourth year using the described method are presented.

5. Some remarks on estimation of covariance parameters and kriging with estimated covariance parameters

The main problem by using the kriging theory in practice consist in estimation of covariance of the prediction regression model. For making estimation of covariance parameters it is necessary to use some model for covariance function. The possible model is (*a nonparametric*) *model of covariance stationarity*, where we assume that $R(s,t) = R(s-t)$; $s,t \in T$.

Then $\Sigma_{st} = R(s-t)$ and for $U = X(n+1)$ we have $\mathbf{r} = \big(R(n), R(n-1), \dots, R(1)\big)'$. For a regression model

$$E_{\boldsymbol{\theta}}[\mathbf{X}] = \mathbf{m}_{\boldsymbol{\theta}}; \quad \boldsymbol{\theta} \in \boldsymbol{\Theta},$$

we can consider the estimators $\widehat{R}_n(t)$ given by

$$\widehat{R}_n(t) = \frac{1}{n-t} \sum_{s=1}^{n-t} \big(X(s+t) - m_{\widehat{\boldsymbol{\theta}}}(s+t)\big)\big(X(s) - m_{\widehat{\boldsymbol{\theta}}}(s)\big); \quad t = 0, 1, \dots, n-1,$$

where $\widehat{\boldsymbol{\theta}}$ is the OLSE of $\boldsymbol{\theta}$. It is shown in Š t u l a j t e r (1989) that $\lim\limits_{n\to\infty} E\big[\widehat{R}_n(t)-R(t)\big]^2 = 0$ for any linear or nonlinear model for $\mathbf{m}_{\boldsymbol{\theta}}$ if $\lim\limits_{t\to\infty} R(t) = 0$ and X is a Gaussian time series.

Another possible model for $R(\cdot,\cdot)$ is to assume that $R(s,t) = R_{\boldsymbol{\nu}}(s,t)$; $\boldsymbol{\nu} \in \boldsymbol{\Upsilon}$, where the dependence of $R_{\boldsymbol{\nu}}$ on $\boldsymbol{\nu}$ is a nonlinear one. As an example we can consider AR(n) models for, especially for AR(1) we have

$$R_{\boldsymbol{\nu}}(t) = \frac{\sigma^2}{1-\rho^2}\rho^t; \quad t = 0,1,\dots$$

with $\boldsymbol{\nu} = (\sigma^2,\rho)' \in (0,\infty)\times(-1,1)$.

Under this assumption we have *a general model* for an observation $\mathbf{X}$:

$$\mathbf{X} = \mathbf{m}_{\boldsymbol{\theta}} + \boldsymbol{\varepsilon}; \quad E[\boldsymbol{\varepsilon}] = \mathbf{0},\ E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'] = \boldsymbol{\Sigma}_{\boldsymbol{\nu}}; \quad \boldsymbol{\theta} \in \boldsymbol{\Theta},\ \nu \in \Upsilon.$$

We can assume that $\mathbf{X}$ is a *gaussian vector* and find, for example, the MLE $(\widetilde{\boldsymbol{\theta}},\widetilde{\boldsymbol{\nu}})'$ of $(\boldsymbol{\theta},\boldsymbol{\nu})'$. We refer to A z a i s (1993) and H a r v i l l e (1977) for properties of such estimators.

As a predictor $\widetilde{U}$ for $U = X(s)$ we can use, according to *a generalized kriging theory*, the plug-in predictor $\widetilde{U}$ given by

$$\widetilde{U} = m_{\widetilde{\boldsymbol{\theta}}}(s) + \mathbf{r}'_{\widetilde{\boldsymbol{\nu}}}\boldsymbol{\Sigma}^{-1}_{\widetilde{\boldsymbol{\nu}}}(\mathbf{X} - \mathbf{m}_{\widetilde{\boldsymbol{\theta}}}),$$

where we assume that

$$E_{\boldsymbol{\theta}}[U] = m_{\boldsymbol{\theta}}(s); \quad \boldsymbol{\theta} \in \boldsymbol{\Theta},$$

and

$$\mathrm{Cov}(U;\mathbf{X}) = \mathbf{r}_{\boldsymbol{\nu}}; \quad \boldsymbol{\nu} \in \boldsymbol{\Upsilon},$$

also depends on $\boldsymbol{\nu}$. Some properties of this plug-in predictor for a linear model for $\mathbf{m}_{\boldsymbol{\theta}}$ are given in C h r i s t e n s e n (1991).

For an AR(1) errors we have for $U = X(s)$

$$\widetilde{U} = m_{\widetilde{\boldsymbol{\theta}}}(s) + \widetilde{\rho}\big(X(n) - m_{\widetilde{\boldsymbol{\theta}}}(n)\big).$$

Properties of this predictor for a linear model for $\mathbf{m}_{\boldsymbol{\theta}}$ were studied also in T o y o k a (1982). An approximation for the MSE of this predictor is given in Š t u l a j t e r (1994b).

Another possible model for errors is

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

where we assume that $\mathbf{Y} = (Y_1,\dots,Y_l)'$ has $E[\mathbf{Y}] = \mathbf{0}$, $\mathrm{Cov}(\mathbf{Y}) = \boldsymbol{\Sigma}^{\mathbf{Y}}$, g_i are known functions and $w(t)$ is a white noise independent of $\mathbf{Y}$ and of a variance σ^2. Then

$$\mathrm{Cov}(\mathbf{X}) = \mathbf{\Sigma} = G\mathbf{\Sigma}^{\mathbf{Y}}\mathbf{G}' + \boldsymbol{\sigma}^2\mathbf{I}$$

with $G_{ij} = g_j(i);\ i = 1,2,\dots,n;\ j = 1,2,\dots,l$.

In a special case when $Y_i;\ i = 1,2,\dots,l$, are uncorrelated with variances σ_i^2 we have

$$\mathbf{\Sigma} = \mathbf{\Sigma}_{\sigma^2} = \sum_{j=1}^{l+1} \sigma_j^2 \mathbf{V}_j\,,$$

where $\mathbf{g}_j = \big(g_j(1),\dots,g_j(n)\big)'$ and $\mathbf{V}_j = \mathbf{g}_j\mathbf{g}_j';\ j = 1,2,\dots,l+1$, are known symmetric matrices, and $\mathbf{V}_{l+1} = \mathbf{I}$. Thus in this case we have *a variance components model for* $\mathbf{\Sigma}$ with an unknown parameter $\boldsymbol{\sigma}^2 = \big(\sigma_1^2,\dots,\sigma_{l+1}^2\big)'$. Harville (1977) and Rao and Kleffe (1988) give methods of estimation of $\boldsymbol{\sigma}^2$.

An initial estimator $\widetilde{\boldsymbol{\sigma}}^2$ of $\boldsymbol{\sigma}^2$ can be found by a *double ordinary least squares method* which is given by

$$\widehat{\boldsymbol{\sigma}}^2 = \arg\min_{\boldsymbol{\sigma}^2} \left\| \widehat{\mathbf{\Sigma}} - \sum_{j=1}^{l+1} \sigma_j^2 \mathbf{V}_j \right\|^2,$$

where

$$\widehat{\mathbf{\Sigma}} = (\mathbf{X} - \mathbf{m}_{\widehat{\boldsymbol{\theta}}})(\mathbf{X} - \mathbf{m}_{\widehat{\boldsymbol{\theta}}})',$$

$\widehat{\boldsymbol{\theta}}$ is the OLSE of $\boldsymbol{\theta}$ and $\|\mathbf{A}\|^2 = \sum\limits_{i,j=1}^{n} A_{ij}^2$.

It is easy to see that

$$\widehat{\boldsymbol{\sigma}}^2 = \mathbf{H}^{-1} \begin{pmatrix} \mathrm{tr}(\mathbf{V}_1\widehat{\mathbf{\Sigma}}) \\ \vdots \\ \mathrm{tr}(\mathbf{V}_{l+1}\widehat{\mathbf{\Sigma}}) \end{pmatrix},$$

where $H_{ij} = \mathrm{tr}(\mathbf{V}_i\mathbf{V}_j);\ i,j = 1,2,\dots,l+1$.

Properties of this estimator were, for a linear model for $\mathbf{m}_{\boldsymbol{\theta}}$ and special matrices $\mathbf{V}_j$, studied in Štulajter (1989). Then we use a predictor $\widehat{X}(s)$ for some s given by

$$\widehat{X}(s) = m_{\widehat{\theta}}(s) + \mathbf{r}_{\widehat{\sigma}^2}\mathbf{\Sigma}_{\widehat{\sigma}^2}^{-1}(\mathbf{X} - \mathbf{m}_{\widehat{\theta}})\,,$$

where

$$\mathbf{\Sigma}_{\widehat{\boldsymbol{\sigma}}} = \sum_{j=1}^{l+1} \widehat{\sigma}_j^2 \mathbf{V}_j \qquad \text{and} \qquad r_{\widehat{\boldsymbol{\sigma}}}(t) = \sum_{j=1}^{l+1} \widehat{\sigma}_j^2 g_j(t) g_j(s);\quad t = 1,2,\dots,n\,.$$

The mean squared error of this predictor should be computed.

We can also use $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\sigma}}$ as initial estimators for computing *the* MLE $\widetilde{\boldsymbol{\theta}}$ and $\widetilde{\boldsymbol{\sigma}}^2$ of $\boldsymbol{\theta}$ and $\boldsymbol{\sigma}^2$ and to use a predictor

$$\widetilde{X}(s) = m_{\widetilde{\theta}}(s) + \mathbf{r}'_{\widetilde{\sigma}} \boldsymbol{\Sigma}_{\widetilde{\sigma}}^{-1} (\mathbf{X} - \mathbf{m}_{\widetilde{\theta}}) .$$

The mean squared error of this predictor should also be computed. This is an open problem also for other models on covariance structure of errors. The problem of estimation of parameters in a nonlinear covariance model was studied in Gompertz and Pantulla (1992). Using their results in prediction problems, mainly computing the MSE of predictors with estimated covariance parameters, is still an open problem.

REFERENCES

[1] ANDERSON, T. W.: *The Statistical Analysis of Time Series*, Wiley and Sons, New York, 1971.

[2] AZAIS, J. M.—BARDIN, A.—DHORNE, T.: MINQUE*, maximum likelihood estimation and Fisher Scoring Algorithm for non linear variance models*, Statistics **24** (1993), 205–213.

[3] BOX, G. E. P.—JENKINS, G. M.: *Time Series Analysis, Forecasting and Control*, Holden Day, San Francisco, 1976.

[4] BROCKWELL, P. J.—DAVIS, R. A.: *Time Series: Theory and Methods*, Springer-Verlag, Berlin, 1987.

[5] CHRISTENSEN, R.: *Plain Answers to Complex Questions: The Theory of Linear Models*, Springer-Verlag, Berlin, 1987.

[6] CHRISTENSEN, R.: *Linear Models for Multivariate, Time Series and Spatial Data*, Springer-Verlag, Berlin, 1991.

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

[8] GUMPERTZ, M. L.—PANTULLA, S. G.: *Nonlinear regression with variance components*, J. Amer. Statist. Assoc. **87** (1992), 201–209.

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

[10] HARVILLE, D. A.: BLUP *and beyond*, in: Advances in statistical methods for improvement of livestock, Springer-Verlag, Berlin, 1990.

[11] Mc GILCHRIST, C. A.: *Estimation in generalized mixed models*, J. Roy. Statist. Soc. Ser. B **56** (1994), 61–69.

[12] PÁZMAN, A.: *Nonlinear Statistical Models*, Kluwer, Dordrecht, 1993.

[13] RAO, C. R.—KLEFFE, J.: *Estimation of Variance Components and Applications*, North Holland, Amsterdam, 1998.

[14] ŠTULAJTER, F.: *Consistency of linear and quadratic* LSE *in regression models with covariance stationary errors*, Appl. Math. **36** (1989), 149–155.

[15] ŠTULAJTER, F.: *Mean squared error matrix of an approximate* LSE *in a nonlinear regression model with correlated errors*, Acta Math. Univ. Comenian. **LXI** (1992), 251–261.

[16] ŠTULAJTER, F.: *On estimation of a covariance function of stationary errors in a nonlinear regression model*, Acta Math. Univ. Comenian. **LXIII** (1994a), 107–116.

[17] ŠTULAJTER, F.: *Mean squared errors of prediction by kriging in linear models with AR(1) errors*, Acta Math. Univ. Comenian. **LXIII** (1994b), 247–254.

[18] ŠTULAJTER, F.—STANO, S.: *A comparison of some predictors in nonlinear regression models*, Tatra Mt. Math. Publ. **7** (1996), 305–309.

[19] ŠTULAJTER, F.: *Predictions in nonlinear regression models*, Acta Math. Univ. Comenian. **LXVI** (1997), 71–81.

[20] ŠTULAJTER, F.: *Predictions in multivariate regression models*, Tatra Mt. Math. Publ. **17** (1999), 265–272.

[21] ŠTULAJTER, F.: *Predictions in time series using multivariate regression models*, J. Time Ser. Anal. **22** (2001), 365–373.

[22] TOYOOKA, Y.: *Prediction error in a linear model with estimated parameters*, Biometrica **69** (1982), 453–459.

[23] USAMI, Y.—HUZII, M.: *Estimation of time series regression with a nonstationary error process*, J. Time Ser. Anal. **96** (1993), 105–118.

[24] ZIMMERMAN, D. L.—CRESSIE, N.: *Mean squared error in spatial linear model with estimated regression parameters*, Ann. Inst. Statist. Math. **44** (1992), 27–43.

Received February 5, 2001

Department of Probability
and Mathematical Statistics
Faculty of Mathematics,
Physics and Informatics
Comenius University
Mlynská dolina
SK–842 15 Bratislava
SLOVAKIA

E-mail: stulajter@fmph.uniba.sk