Tatra Mt. Math. Publ. **17** (1999), 265–272

TATRA MOUNTAINS
Mathematical Publications

PREDICTIONS IN MULTIVARIATE REGRESSION MODELS

František Štulajter

ABSTRACT. Predictions in time series using multivariate regression models are studied with respect to their mean squared errors. A new method of prediction is proposed. The mean squared error of this prediction is computed. A model with seasonal and cyclical component is also studied.

1. Introduction

There are many possibilities for prediction of time series. The well known and commonly used are methods based on Box – Jenkins (1970) methodology.

The other approach is based on modelling time series by 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 (1990).

The aim of this paper is to study a method of prediction using multivariate regression models. This method is based on an idea to predict future values of observed time series using multivariate regression model with estimated and predicted regression parameters.

A multivariate regression model with seasonal and cyclical component is also studied.

Let us assume, for example, that we observe a time series

$$X_i(t), \qquad t = 1, 2, \dots, 12, \quad i = 1, 2, \dots, R,$$

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 in every year is similar, but changes from year to year.

2000 Mathematics Subject Classification: 62M10.
Keywords: multivariate regression model, seasonal and cyclical components, mean squared error of predictor.

Supported by grant No. 1/4196/97 of Slovak Grand Agency.

Thus for modelling the whole observed time series by a regression model it is convenient to use a multivariate, rather than an univariate regression model.

A multivariate regression model can generally be described, see Christensen [6], as a set of linear models

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

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 $\boldsymbol{\beta}_i = (\beta_{i1}, \dots, \beta_{ik})'$, $i = 1, 2, \dots, R$, 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 $\mathrm{E}[\boldsymbol{\varepsilon}_i] = \mathbf{0}$ for all $i = 1, 2, \dots, R$. The multivariate linear regression model consists of fitting the R linear models simultaneously. This can also be done as follows.

Let $\mathbf{X} = (\mathbf{X}_1', \mathbf{X}_2', \dots, \mathbf{X}_R')'$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \dots, \boldsymbol{\beta}_R')'$ and let $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1', \boldsymbol{\varepsilon}_2', \dots \dots, \boldsymbol{\varepsilon}_R')'$. Then we can write the multivariate regression model as an univariate one

$$\mathbf{X} = (\mathbf{I} \otimes \mathbf{F})\boldsymbol{\beta} + \boldsymbol{\varepsilon}\,, \quad E[\boldsymbol{\varepsilon}] = \mathbf{0}\,, \quad E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'] = \boldsymbol{\Sigma}_1 \otimes \boldsymbol{\Sigma}\,, \tag{1}$$

where $\boldsymbol{\Sigma}_1$ is a positive definite covariance matrix and $\otimes$ denotes Kronecker product. The matrices $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}$ are assumed to be known in this paper. From the last assumption we have $\mathrm{Cov}(\mathbf{X}_i\,; \mathbf{X}_j) = \Sigma_{1ij}\boldsymbol{\Sigma}$. If we assume that the random vectors $\mathbf{X}_i$, $i = 1, 2, \dots, R$, are independent, then $E[\boldsymbol{\varepsilon}\,\boldsymbol{\varepsilon}'] = \mathbf{I} \otimes \boldsymbol{\Sigma}$.

We shall use the following well known properties of Kronecker product of matrices and vectors:

$$(\mathbf{A} \otimes \mathbf{B})' = \mathbf{A}' \otimes \mathbf{B}', \qquad (\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}\,,$$

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC} \otimes \mathbf{BD})\,.$$

Using these properties it is easy to see that the BLUE $\boldsymbol{\beta}^* = \big(\boldsymbol{\beta}_1^{*\prime}, \boldsymbol{\beta}_2^{*\prime}, \dots, \boldsymbol{\beta}_{\mathbf{R}}^{*\prime}\big)'$ of $\boldsymbol{\beta}$ is given by $\boldsymbol{\beta}_i^* = \big(\mathbf{F}'\Sigma^{-1}\mathbf{F}\big)^{-1}\mathbf{F}'\Sigma^{-1}\mathbf{X}_i$, $i = 1, 2, \dots, R$, for any $\boldsymbol{\Sigma}_1$ and thus for the OLSE $\hat{\boldsymbol{\beta}} = (\hat{\boldsymbol{\beta}}_1', \hat{\boldsymbol{\beta}}_2', \dots, \hat{\boldsymbol{\beta}}_R')'$ of $\boldsymbol{\beta}$ we have $\hat{\boldsymbol{\beta}}_i = (\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'\mathbf{X}_i$, $i = 1, 2, \dots R$.

2. Prediction in multivariate regression models

Let us assume that $\mathbf{X} = (\mathbf{X}_1', \mathbf{X}_2', \dots, \mathbf{X}_R')'$ is the observed time series modelled by the multivariate regression model given by (1) and let $\mathbf{X}_{R+1} = \big(X_{R+1}(1), X_{R+1}(2), \dots, X_{R+1}(n)\big)'$ be 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 $X_{R+1}(t)$, $t = 1,2,\dots,n$, then we assume $E[U] = \boldsymbol{f}'\boldsymbol{\beta}_{R+1}$ for some given $k \times 1$ vector $\boldsymbol{f} = (f_1, f_2, \dots, f_k)'$.

Now, let $\tilde{\boldsymbol{\beta}}_1, \dots, \tilde{\boldsymbol{\beta}}_R$ 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 $\tilde{\boldsymbol{\beta}}_i = \hat{\boldsymbol{\beta}}_i$, $i = 1,2,\dots,R$. Let $\boldsymbol{\beta}_i = (\beta_{i1}, \beta_{i2}, \dots, \beta_{ik})'$, $i = 1,2,\dots,R$, and let $\tilde{\boldsymbol{\beta}}_{.j} = (\tilde{\beta}_{1j}, \tilde{\beta}_{2j}, \dots, \tilde{\beta}_{Rj})'$ be the $R \times 1$ vector of estimators of the jth component of parameters $\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_R$ for $j = 1,2,\dots,k$.

Let us assume that

$$E[\tilde{\boldsymbol{\beta}}_{.j}] = \mathbf{H}\gamma_j\,, \qquad j = 1,2,\dots,k\,,$$

where $\mathbf{H}$ is a known $R \times l$ matrix and $\boldsymbol{\gamma}_j = (\gamma_{j1}, \gamma_{j2}, \dots, \gamma_{jl})'$, $j = 1,2,\dots,k$, are unknown $l \times 1$ regression parameters. Define $\tilde{\boldsymbol{\beta}}_{R+1} = (\tilde{\beta}_{R+1,1}, \dots, \tilde{\beta}_{R+1,k})'$ and

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

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

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

As a predictor $\tilde{U}$ of U we take

$$\tilde{U} = \boldsymbol{f}'\tilde{\boldsymbol{\beta}}_{R+1}\,. \tag{2}$$

To make the assumptions clear let us consider the following example.

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 $\mathbf{F} = \begin{pmatrix} 1 & 1 & . & . & . & 1 \\ 1 & 2 & . & . & . & n \end{pmatrix}'$ and for $U = X_{R+1}(t)$ we have $\boldsymbol{f} = (1,t)'$, $t = 1,2,\dots,n$. Let $\tilde{\boldsymbol{\beta}}_i = (\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'\mathbf{X}_i$, $i = 1,2,\dots,R$, and let

$$\tilde{\boldsymbol{\beta}}_{.j} = (\tilde{\beta}_{1j}, \dots, \tilde{\beta}_{Rj})'\,, \qquad j = 1,2\,.$$

If we take $H = \begin{pmatrix} 1 & 1 & . & . & . & 1 \\ 1 & 2 & . & . & . & R \end{pmatrix}'$ and $\boldsymbol{h} = (1, R+1)'$ then both components $\beta_{R+1,j}$, $j = 1,2$, are predicted from $\tilde{\boldsymbol{\beta}}_{.j}$, $j = 1,2$, by prolonged regression line with parameters estimated by the OLS method. The model indicates that the time series has a trend from year to year but that it also has a trend from January to December.

Now we shall compute the variance of the predictor $\tilde{U} = \boldsymbol{f}'\tilde{\boldsymbol{\beta}}_{R+1}$ where $\tilde{\boldsymbol{\beta}}_i = \hat{\boldsymbol{\beta}}_i$, $i = 1,2,\dots,R$, are the OLSE's. Let a_{si}, $s = 1,2,\dots,l$, $i = 1,2,\dots,R$, denote the elements of the matrix $(H'\mathbf{H})^{-1}\mathbf{H}'$. Then, using $\tilde{U} = \boldsymbol{f}'\tilde{\boldsymbol{\beta}}_{R+1}$, $\tilde{\boldsymbol{\beta}}_{R+1} = \boldsymbol{h}'\tilde{\gamma}_j$, $\tilde{\gamma}_j = (\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'\tilde{\boldsymbol{\beta}}_{.j}$, we can write

$$\tilde{U} = \sum_{j=1}^{k} f_j \sum_{s=1}^{l} h_s \sum_{i=1}^{R} a_{si}\tilde{\beta}_{ij}\,.$$

Let d_{jt}, $j = 1,2,\dots,k$; $t = 1,2,\dots,n$, denote the elements of the matrix $(\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'$, then

$$\begin{aligned}\tilde{U} &= \sum_{j=1}^{k} f_j \sum_{s=1}^{l} h_{si} \sum_{i=1}^{R} a_{si} \sum_{t=1}^{n} d_{jt} X_i(t) \\ &= \sum_{i=1}^{R}\sum_{t=1}^{n}\left(\sum_{s=1}^{l} h_s a_{si}\right)\left(\sum_{j=1}^{k} f_j d_{jt}\right) X_i(t)\,.\end{aligned}$$

Let $\boldsymbol{c} = (c_1, c_2, \dots, c_R)' = \mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}h$ and $\boldsymbol{b} = (b_1, b_2, \dots, b_n)' = \mathbf{F}(\mathbf{F}'\mathbf{F})^{-1}f$. Then we have

$$\tilde{U} = \sum_{i=1}^{R}\sum_{t=1}^{n} c_i b_t X_i(t)\,.$$

This expression can be written in a more convenient form

$$\tilde{U} = (\boldsymbol{c} \otimes \boldsymbol{b})'\mathbf{X}\,,$$

where $\mathbf{X}$ has the covariance matrix $\Sigma_1 \otimes \Sigma$. Thus we have for the variance of $\tilde{U}$

$$\begin{aligned} D[\tilde{U}] &= (\boldsymbol{c}' \otimes \boldsymbol{b}')(\Sigma_1 \otimes \Sigma)(\boldsymbol{c} \otimes \boldsymbol{b}) = \boldsymbol{c}'\Sigma_1 \boldsymbol{c}\, \boldsymbol{b}'\Sigma \boldsymbol{b} \\ &= \boldsymbol{h}'(\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'\Sigma_1\mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\boldsymbol{h}\boldsymbol{f}'(\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'\Sigma\,\mathbf{F}(\mathbf{F}'\mathbf{F})^{-1}\boldsymbol{f}\,.\end{aligned}$$

From our assumptions we get for *the mean squared error* (the MSE) of this prediction the expression

$$E[\tilde{U}-U]^2 = D[U] + \boldsymbol{h}'(\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'\Sigma_1\mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\boldsymbol{h}\,\boldsymbol{f}'(\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'\Sigma\,\mathbf{F}(\mathbf{F}'\mathbf{F})^{-1}\boldsymbol{f}\,. \quad (3)$$

In the special case when $\Sigma_1 = \sigma_1^2\mathbf{I}$ and $\Sigma = \sigma^2\mathbf{I}$ we get

$$E[\tilde{U} - U]^2 = D[U] + \sigma_1^2\sigma^2\,\boldsymbol{h}'(\mathbf{H}'\mathbf{H})^{-1}\boldsymbol{h}\,\boldsymbol{f}'(\mathbf{F}'\mathbf{F})^{-1}\boldsymbol{f}\,.$$

Using the kriging theory, see Christensen [6] or Goldberger [7], it is possible to show that $\tilde{U}$ is in this case *the best linear unbiased estimator* of U.

Remark. For the double linear trend prediction from Example 1 we get for $\tilde{U} = \tilde{X}_{R+1}(t)$

$$E[\tilde{U} - U]^2 = D[U] + \sigma_1^2\sigma^2\,\frac{2(2R+1)}{R(R-1)}\left(-\frac{12t}{n(n-1)} + \frac{12t^2}{R(R-1)}\right).$$

3. Prediction in multivariate regression models with seasonal and cyclical components

For economical time series or time series describing natural time series it is useful instead of model (1) to use a model

$$\mathbf{X} = (\mathbf{I} \otimes \mathbf{F})\boldsymbol{\beta} + \mathbf{m}_{\boldsymbol{\theta}} + \boldsymbol{\varepsilon}\,, \tag{4}$$

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

$$m_{\boldsymbol{\theta}}(s) = \sum_{i=1}^{p} \alpha_i^{(1)} \cos \lambda_i s + \alpha_i^{(2)} \sin \lambda_i s\,, \qquad s = 1, 2, \dots, nR\,,$$

where

$$\boldsymbol{\theta} = (\alpha_1^{(1)}, \dots, \alpha_p^{(1)}, \alpha_1^{(2)}, \dots, \alpha_p^{(2)}\,;\ \lambda_1, \dots, \lambda_p)' = (p, \boldsymbol{\alpha}'\boldsymbol{\lambda}')'$$

is an unknown parameter.

For the predicted random variable $U = X_{R+1}(t)$ we assume that

$$E[U] = \boldsymbol{f}'\boldsymbol{\beta}_{R+1} + \mathbf{m}_{\boldsymbol{\theta}}(nR + t)\,, \qquad t = 1, 2, \dots, n\,.$$

As a predictor $\tilde{U}$ of $U = X_{R+1}(t)$ now take

$$\tilde{U} = \boldsymbol{f}'\tilde{\boldsymbol{\beta}}_{R+1} + m_{\tilde{\boldsymbol{\theta}}}(nR + t)\,, \tag{5}$$

where $\tilde{\boldsymbol{\beta}}_{R+1}$ and $\tilde{\boldsymbol{\theta}}$ are some estimators (usually *the ordinary* LSE) of unknown parameters $\boldsymbol{\beta}_{R+1}$ and $\boldsymbol{\theta}$. We remark that (4) is a nonlinear regression model for $\mathbf{X}$. Initial estimates for $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ can be found as follows. Using $\mathbf{X}$ we find OLSE's $\hat{\boldsymbol{\beta}}$ for $\boldsymbol{\beta}$. Then we can compute the periodogram of residuals $\mathbf{X} - \mathbf{F}\hat{\beta}$ and, according to well known procedure, choose $\hat{p}$ and $\hat{\boldsymbol{\lambda}}$. Considering

$$\mathbf{X} - \mathbf{F}\hat{\boldsymbol{\beta}} = \mathbf{m}_{(\hat{p}, \boldsymbol{\alpha}', \hat{\boldsymbol{\lambda}}')'} + \kappa$$

as a linear regression model with parameter $\boldsymbol{\alpha}$ we can compute the OLSE $\hat{\boldsymbol{\alpha}}$ of $\boldsymbol{\alpha}$. The estimates $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\theta}}$ can be used, together with $\hat{p}$, for some iterative procedure to find the MLE $\tilde{\boldsymbol{\beta}}$, $\tilde{\boldsymbol{\theta}}$ of $\boldsymbol{\beta}$, $\boldsymbol{\theta}$.

Then we find $\tilde{\boldsymbol{\beta}}_{R+1}$.

The MSE of $\tilde{U}$ given by (5) can be approximated using the results of Štulajter [8].

Example 1. On the following picture there are simulated data which are considered as monthly observations during three years with white noise random errors. Predicted values for the fourth year were found using the predictors given by (5).

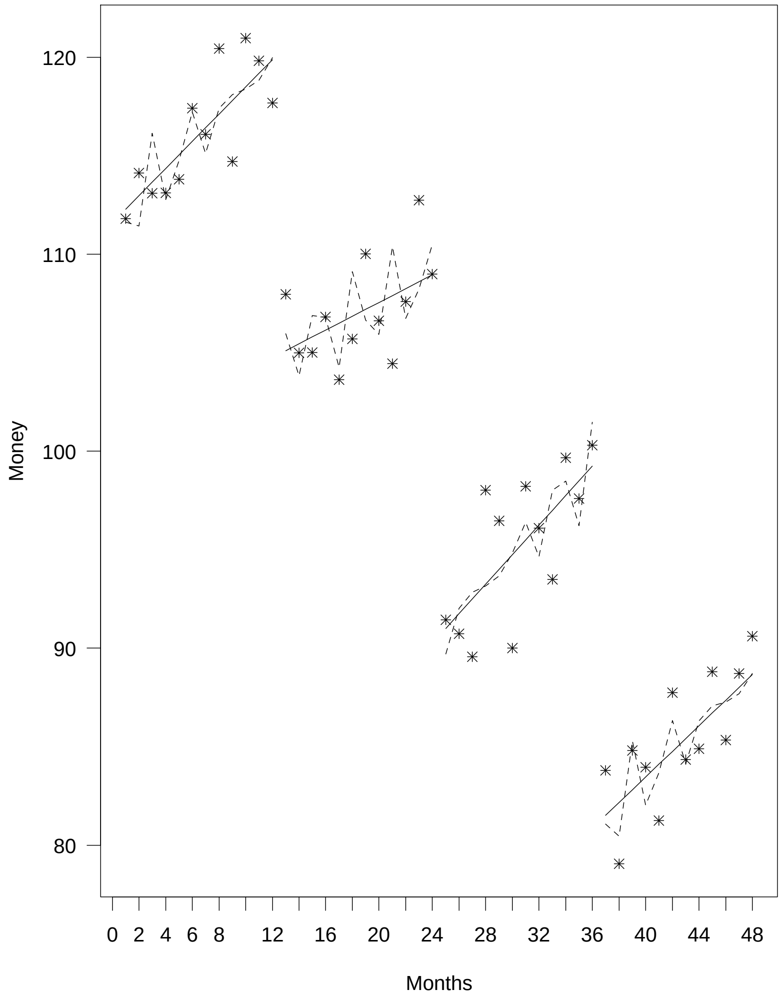

Figure 1

Example 2. On Figure 2 there are real data of monthly observation during three years of a school office. The prediction for the following year using (5) are also presented.

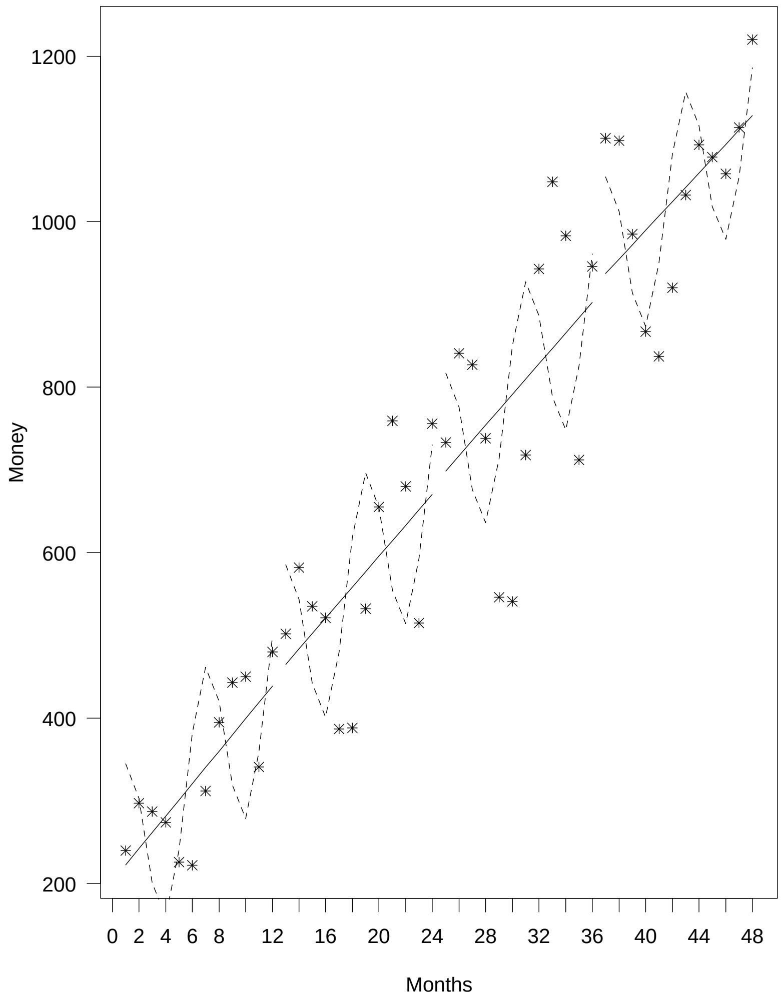

FIGURE 2

REFERENCES

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

[2] ANDERSON, T. W.: *An Introduction to Multivariate Statistical Analysis. Second Edition*, John Wiley and Sons, New York, 1984.

[3] ARNOLD, S. F.: *The Theory of Linear Models and Multivariate Analysis*, John Wiley and Sons, New York, 1981.

[4] BOX, G. E. P.—JENKINS, G. M.: *Time Series Analysis, Forecasting and Control.* (*Revised Edition, 1976*), Holden Day, San Francisco, 1970.

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

[6] CHRISTENSEN, R.: *Linear Models for Multivariate Time Series and Spatial Data*, Springer-Verlag, New York, 1990.

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

[8] ŠTULAJTER, F.: *Predictions in nonlinear regression models*, Acta Math. Univ. Comenian. **61** (1997), 71–81.

Received May 7, 1998

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

E-mail: stulajter@fmph.uniba.sk