Tatra Mt. Math. Publ. **26** (2003), 103–113

TATRA MOUNTAINS
Mathematical Publications

MIDDLE TERM PREDICTIONS OF TIME SERIES IN THE MODELS WITH EXTERNAL INFORMATION ABOUT FUTURE

JÁN BÁBELA

ABSTRACT. Predictions in the models with external information are investigated. Bayesian approach is used to determine the best predictors. New notion is introduced—optimal prediction with maximum posterior likelihood method. In various models for time series the optimal prediction is then computed.

Introduction

In the present paper we will try to make the predictions of time series with some prior external information about a future. Predictions of time series play an important role in the theory of random processes, because they are applicable in the economic theory and also in other fields of current life.

Adequacy of chosen theoretical model for time series is often measured by comparing the predicted values with the real ones, taken later. Problem of predicting the time series is therefore found in every time series model. Models for the time series are well known and studied in the basic literature about the time series (Anděl (1976), Brockwell, Davis (1987), Štulajter (2001)), but in the economic life new phenomena arise very often. For these phenomena statisticians seek new models for making predictions, which are optimal under some reasonable criterion. A main criterion for optimality of predictions is the mean square error (MSE) criterion in the chosen model. A practical criterion for model adequacy is a later confrontation of predicted and real values. The accuracy of predictions, optimal under some theoretical criterion depends on the accuracy of criterion. Making the reasonable theoretical criterions, building models, which can be applicable in new situations and predicting the time series

2000 Mathematics Subject Classification: 60G25.
Keywords: predictions, time series, Bayesian forecasting, external information.
Supported by the grant VEGA 1/7295/20.

is an actual field of theory of time series and this paper tries to contribute to this theme.

Observation of the time series is a set of the values $x_1, x_2, \dots, x_t, \dots x_n$ taken in the time $t = 1, 2, \dots, n$. From the observation of past values of time series, a prediction to the future is made. The problem of making predictions from the past values in general is complicated, so some model for time series is supposed. We suppose in the models, that observed values are realizations of random variables. Theoretical characteristics of random variables (such as mean value, dispersion, distribution,...) follows from the chosen model. The models mostly used are models for mean, covariance matrix and for stationary time series, autoregressive models, describing the dependence of random variable X_t on random variables $X_{t-1}, X_{t-2}, \dots$ Mainly, the linear regression models are used, because of their well known theoretical properties.

Problem of analysis of the time series and making the prediction can be in general written as follows: From the observed values $x_1, x_2, \dots, x_n$ of time series till the time $t = n$ to make a theoretical model and to make a prediction of time series values $x_{n+1}, x_{n+2}, \dots, x_{n+k}$, optimal according to the given criterion of optimality.

Predictions of real economic time series based on the past data, obtained by the statistical analysis, are not accurate in long terms, because of unceasing changes of economic environment. Predictions based on mathematical models, (such as ARMA, ARIMA, ARCH, GARCH) are therefore made only for short terms. It is too difficult to predict time series in mid, or long terms, so economists and model-makers are satisfied with the rough trend estimates and basic properties of observed processes.

In a real situation there might be cases, when it is necessary to make a middle-term prediction based on past data and also on some external information. For example a company may want to make a prediction of incomes, having a promises from some business partners to pay their debts in specified dates. Another example is a state, making a prediction of tax incomes with reports from tax offices about future payments of tax earnest. Predictions made from past data with regard to external information may be then accurate also in middle-terms.

Let $x_p = (x_1, x_2, \dots, x_n)$ be an observation of time series $X(t)$, $t = 1, 2, \dots, n, \dots$ from the past, let $z = (z_{n+1}, z_{n+2}, \dots, z_{n+k})$ be an external information about a future values of time series X. Our purpose is to find a prediction of time series X; $y = (y_{n+1}, y_{n+2}, \dots, y_{n+k})$, based on the past of time series and also on external information. We will make some assumptions on time series, from that we will build a model describing the time series and relations between time series and external information and finally find a prediction optimal in the chosen model.

We use mainly standard Bayesian theory in our models. From West–Harrison (1997) we adopted an approach how to include an external information to the models. All theorems are new and also the definition of optimal predictor resulting from models is new, through we tried to make it reasonable. Properties of normal and Student distribution used in proof are described and proved in standard books on Bayesian theory, also in West–Harrison (1997).

1. Model with prior distribution of future of time series

1.1 Bayesian approach

The first problem, which arises, is a problem of mathematical formalization. There might be more approaches, one of them we suggest:

We assume, that time series X_t, $t = 1, 2, \dots$, has a normal distribution. Let random vector X_p be a past of time series until the time $t = n$, $X_p = (X_1, X_2, \dots, X_n)'$ and let random vector $Y = (Y_{n+1}, Y_{n+2}, \dots, Y_{n+k})'$ be a future of time series in times $t = n+1, n+2, \dots, n+k$. We assume that $(X_p, Y)'$ has a multivariate normal distribution with known parameters: $E\big((X_p, Y)'\big) = (\mu_{X_p}, \mu_Y)'$, $\operatorname{Cov}(X_p) = \Sigma_{X_pX_p}$, $\operatorname{Cov}(Y) = \Sigma_{YY}, \operatorname{Cov}(X_p, Y) = \Sigma_{X_pY}$, from that we obtain the conditional distribution $X_p/Y \sim N\big(\mu_{X_p/Y}, \Sigma_{X_p/Y}\big)$:

$$\begin{aligned}\mu_{X_p/Y} &= \mu_{X_p} + \Sigma_{X_pY}\Sigma_{YY}^{-1}(Y - \mu_Y)\,,\\ \Sigma_{X_p/Y} &= \Sigma_{X_pX_p} - \Sigma_{X_pY}\Sigma_{YY}^{-1}\Sigma_{YX_p}\,.\end{aligned}$$

It is a well known result from the theory of conditional normal distribution.

Next we assume, that we have the external prediction $Z = (Z_{n+1}, Z_{n+2}, \dots, Z_{n+k})'$ together with a measure of accuracy of predictions, e.g., we have the confidence intervals, where it is supposed, that with high probability future values of time series will be included in.

We may assume, that external prediction is a realization of random vector Z with some distribution. Information, which is included in external prediction is represented by conditional distribution of future of time series with respect to given information. We assume, that this conditional distribution is normal, with mean values equal to external predictions:

$$Y/Z \sim N(Z, \Sigma_{Y/Z})\,.$$

The distribution Y/Z is in some sense a prior distribution of Y with respect to given external information. It is not really a prior distribution, because we assume, that without any other information $Y \sim N(\mu_Y, \Sigma_{YY})$.

We assume that external information is some random estimate of future of time series. As a measure of accuracy of this random estimate we have the covariance matrix $\Sigma_{Y/Z}$. In practice, it is necessary to determine this matrix. It is obvious that we have to know how good the external prediction is.

Next we assume, that X_p and Z are independent. External source of predictions does not care about past of time series, but has some other relevant information about future of time series. We will discuss more about this assumption later.

From the Bayes theorem we get the distribution Y conditional on X_p and Z:

$$\begin{aligned} f_{Y/X_p,Z}(y|x_p,z) &= f_{X_p/Y,Z}(x_p|y,z)\cdot f_{Y/Z}(y|z)/f_{X_p/Z}(x_p|z) \\ &= f_{X_p/Y}(x_p|y)\cdot f_{Y/Z}(y|z)/f_{X_p}(x_p)\,. \end{aligned}$$

As a prediction of Y by maximum likelihood method we can take $\operatorname{argmax} f_{Y/X_p,Z}(y|x_p,z)$.

Definition 1. Let $X_p = (X_1, X_2, \dots, X_n)'$ be a past of time series X, let $Y = (Y_{n+1}, Y_{n+2}, \dots, Y_{n+k})'$ be it's future. Let $f_{X_p/Y}(x_p|y)$ be density of distribution X_p/Y and let $f_{Y/Z}(y|z)$ be a density of Y conditional on Z. The optimal prediction by maximum likelihood method is

$$\widehat{Y} = \operatorname{argmax} f_{Y/X_p,Z}(y|x_p,z) = \operatorname{argmax} f_{X_p/Y}(x_p|y)\cdot f_{Y/Z}(y|z)/f_{X_p}(x_p)\,.$$

1.2 Predictions in the model with prior distribution of future

Theorem 1. *Let distribution of X_p/Y be $N(\mu_{X_p/Y}, \Sigma_{X_p/Y})$,*

$$\begin{aligned} \mu_{X_p/Y} &= \mu_{X_p} + \Sigma_{X_pY}\Sigma_{YY}^{-1}(Y-\mu_Y)\,, \\ \Sigma_{X_p/Y} &= \Sigma_{X_pX_p} - \Sigma_{X_pY}\Sigma_{YY}^{-1}\Sigma_{YX_p}\,. \end{aligned}$$

Let distribution of Y/Z be $N(Z, \Sigma_{Y/Z})$, matrices $\Sigma_{X_pX_p}, \Sigma_{YY}, \Sigma_{Y/Z}$ are symmetric, positive definite. The distribution of $Y/X_p, Z$ is then $N(\mu,\Sigma)$, with μ,Σ satisfying:

$$\begin{aligned} \mu &= \Sigma\Big(\Sigma_{YY}^{-1}\Sigma_{YX_p/Y}^{-1}\big(X_p - \mu_{X_p} + \Sigma_{X_pY}\Sigma_{YY}^{-1}\mu_Y\big) + \Sigma_{Y/Z}^{-1}Z\Big)\,, \\ \Sigma &= \big(\Sigma_{YY}^{-1}\Sigma_{YX_p}\Sigma_{X_p/Y}^{-1}\Sigma_{X_pY}\Sigma_{YY}^{-1} + \Sigma_{Y/Z}^{-1}\big)^{-1}\,. \end{aligned}$$

Proof. From the Bayes theorem $f_{Y/X_p,Z}(y|x_p,z) = K(x_p)f_{Y/Z}(y|z)$ $f_{X_p/Y}(x_p|y)$, then

$$\begin{aligned} -2\ln f_{Y/X_p,Z}(y|x_p,z) &= K_1(x_p) + \big(x_p - \mu_{X_p} - \Sigma_{X_pY}\Sigma_{YY}^{-1}(y-\mu_Y)\big)'\Sigma_{X_p/Y}^{-1}\big(x_p - \mu_{X_p} \\ &\quad - \Sigma_{X_pY}\Sigma_{YY}^{-1}(y-\mu_Y)\big) + (y-z)'\Sigma_{Y/Z}^{-1}(y-z) \\ &= y'\Sigma^{-1}y + y'\Sigma^{-1}\mu + \mu'\Sigma^{-1}y + K_2(x_p)\,, \end{aligned}$$

as can be easily checked by substituting, constant $K_2(x_p)$ does not depend on Y. That implies:

$$f_{Y/X_p,Z}(y\,|\,x_p,z) = K_3(x_p)\exp\left(-\frac{1}{2}(y-\mu)'\Sigma^{-1}(y-\mu)\right). \qquad \square$$

The optimal prediction by the maximum likelihood method is then $\widehat{Y} = \mu$. Let's examine some basic properties of predictor $\widehat{Y}$:

a) If there is "no relationship" between X_p, Y, that means $\Sigma_{X_pY} = \Sigma'_{YX_p} = 0$, then $\widehat{Y} = \mu_z$.

It corresponds to the intuitive expectations: if there is no linear dependence between observed and predicted values of time series, from the past, we can say nothing about the future and prediction is based only on the external information.

b) If we take distribution $Y \sim N(\mu_Y, \Sigma_{YY})$ as a distribution Y/Z (we don't have any external information), then $\widehat{Y} = \mu_Y + \Sigma_{YX_p}\Sigma^{-1}_{X_pX_p}(X_p - \mu_{X_p})$. Prediction is the same as prediction minimizing mean square error (or maximizing likelihood in case of normal distribution) in the model without external information.

c) Let $\Sigma_{Y/Z} = \delta V$. If $\delta \to +0$ we have:

$$\widehat{Y} = \left(\Sigma^{-1}_{YY}\Sigma_{YX_p}\Sigma^{-1}_{X_p/Y}\Sigma_{X_pY}\Sigma^{-1}_{YY} + \delta^{-1}V^{-1}\right)^{-1}$$

$$\cdot\left(\Sigma^{-1}_{YY}\Sigma_{YX_p}\Sigma^{-1}_{X_p/Y}\left(X_p - \mu_{X_p} + \Sigma_{X_pY}\Sigma^{-1}_{YY}\mu_Y\right) + \delta^{-1}V^{-1}Z\right) \to Z\,.$$

With growing accuracy of external information, $\overline{Y}$ tends to the external estimate of mean value. Also this property mirrors the intuitive expectations.

These properties show that our approach to the problem of making predictions with some external information is reasonable.

Assumption of independence of external prediction Z and past of time series X_p might look as a quite limiting for use in practical applications. If we want to abandon this assumption, we have to know the relationship between Z and X_p. Assuming a normal distribution, we have to know the covariance matrix Σ_{X_pZ}. We have to know, how external prediction maker used an information about a past of time series in his prediction. Indeed, if he does his predictions well, we have no chance to make it better. We must have some advantage from the external prediction maker to make judgements about his predictions and to make a better predictions. This advantage is the knowledge of the past of time series.

Even with the assumption of independence of X_p and Z, predictions maximizing $f_{Y/X_p,Z}(y\,|\,x_p,z)$ might be applicable. They may be used by a company comparing the evolution of some of internal economical index with the prediction made by an external prediction maker for all the economic sphere, in which the company is included, from the external factors. Predictions may be also used in the case, when external prediction was made before the realization of $X_p = (X_1, \dots, X_n)$ and in time $t = n$ it is not possible to get another external prediction. Predictions are applicable for a state, when it makes the predictions of filling the incomes to the end of the year in 5th month, if it has prediction from the beginning of the year and also knows the development from the 1st to the 5th month.

In the previous theorem, there are the parameters μ_{X_p} and μ_Y, which are usually unknown in a real situation. We may, of course, make an estimate for mean value using some linear model and take these estimated mean values as real ones. The next theorem brings the result that enables the parameters estimation together with the prediction determination by maximum likelihood method.

Theorem 2. *Provided that $(X_p, Y)'$ fulfils the linear regression model:*

$$(X_p, Y)' = (F_{X_p}, F_Y)'\beta + \varepsilon\,,$$
$$\varepsilon \sim N(0, \Sigma_\varepsilon)\,,$$

(matrix Σ_ε consists of blocks $\Sigma_{X_pX_p}, \Sigma_{X_pY}, \Sigma_{YX_p}, \Sigma_{YY}$) we get a distribution $X_p/(Y, \beta)$.

Let $X_p/(Y,\beta)$ has the distribution $N(\mu_{X_p/Y,\beta}, \Sigma_{X_p/Y})$, with

$$\begin{aligned}\mu_{X_p/Y,\beta} &= F_{X_p}\beta + \Sigma_{X_pY}\Sigma_{YY}^{-1}(Y - F_Y\beta)\,,\\ \Sigma_{X_p/Y} &= \Sigma_{X_pX_p} - \Sigma_{X_pY}\Sigma_{YY}^{-1}\Sigma_{YX_p}\,.\end{aligned}$$

Let the distribution of Y/Z be $N(Z, \Sigma_{Y/Z})$, let matrices $\Sigma_{X_pX_p}, \Sigma_{YY}, \Sigma_{Y/Z}$ be symmetric, positive definite. Estimates of Y and β by the maximum likelihood method are then:

$$\begin{pmatrix}\hat{Y}\\ \beta\end{pmatrix} = \begin{pmatrix}\Sigma_1 & \Sigma_2\\ \Sigma_4 & \Sigma_3\end{pmatrix}^{-1}\begin{pmatrix}v_1\\ v_2\end{pmatrix}$$

with:

$$\begin{aligned}\Sigma_1 &= \left(\Sigma_{YY}^{-1}\Sigma_{YX_p}\Sigma_{X_p/Y}^{-1}\Sigma_{X_pY}\Sigma_{YY}^{-1} + \Sigma_{Y/Z}^{-1}\right),\\ \Sigma_2 &= \Sigma_{YY}^{-1}\Sigma_{YX_p}\Sigma_{X_p/Y}^{-1}\left(F_{X_p} - \Sigma_{X_pY}\Sigma_{YY}^{-1}F_Y\right),\\ \Sigma_3 &= \left(F_{X_p} - \Sigma_{X_pY}\Sigma_{YY}^{-1}F_Y\right)'\Sigma_{X_p/Y}^{-1}\left(F_{X_p} - \Sigma_{X_pY}\Sigma_{YY}^{-1}F_Y\right),\\ \Sigma_4 &= \left(F_{X_p} - \Sigma_{X_pY}\Sigma_{YY}^{-1}F_Y\right)'\Sigma_{X_p/Y}^{-1}\Sigma_{X_pY}\Sigma_{YY}^{-1},\\ v_1 &= \Sigma_{YY}^{-1}\Sigma_{YX_p}\Sigma_{X_p/Y}^{-1}X_p + \Sigma_{Y/Z}^{-1}Z\,,\\ v_2 &= \left(F_{X_p} - \Sigma_{X_pY}\Sigma_{YY}^{-1}F_Y\right)'\Sigma_{X_p/Y}^{-1}X_p\,.\end{aligned}$$

Proof. From the Bayes theorem

$$f_{Y,\beta/X_p,Z}(y,\beta|x_p,z) = k(x_p) f_{X_p/(Y,\beta),Z}(x_p|y,\beta,z)\cdot f_{Y,\beta/Z}(y,\beta|z)\,.$$

Density function $f_{X_p/(Y,\beta),Z} = f_{X_p/(Y,\beta)}$ is known, density function $f_{Y,\beta/Z}$ can be determined as a product (we assume independence) of $f_{Y/Z}$ and 1. We assume, that β has a uniform distribution on some reasonable subset of $\mathbb{R}^m$, from that an improper density 1 for β follows. We have

$$\begin{aligned} f_{Y,\beta/X_p,Z}(y,\beta|x_p,z) &= k\cdot\exp\Big(-\frac{1}{2}\big(x_p-\mu_{X_p/(Y,\beta)}\big)'\Sigma^{-1}_{X_p/Y}\big(x_p-\mu_{X_p/(Y,\beta)}\big) \\ &\qquad -\frac{1}{2}(y-z)'\Sigma^{-1}_{Y/Z}(y-z)\Big)\,, -2\ln f_{Y,\beta/X_p,Z}(y,\beta|x_p,z) \\ &= K_1+\big(x_p-F_{X_p}\beta-\Sigma_{X_pY}\Sigma^{-1}_{YY}(Y-F_Y\beta)\big)' \\ &\qquad \cdot\Sigma^{-1}_{X_p/Y}\big(x_p-F_{X_p}\beta-\Sigma_{X_pY}\Sigma^{-1}_{YY}(Y-F_Y\beta)\big) \\ &\quad +(y-z)'\,\Sigma^{-1}_{Y/Z}(y-z)\,. \end{aligned}$$

Minimum of $-2\ln f_{Y,\beta/X_p,Z}(y,\beta|x_p,z)$ we can find by partial differencing:

$$\begin{aligned} &\partial\big(-2\ln f_{Y,\beta/X_p,Z}(y,\beta|x_p,z)\big)/\partial Y = 0 \;\Rightarrow \\ &\qquad(\Sigma^{-1}_{YY}\Sigma_{YX_p}\Sigma^{-1}_{X_p/Y}\Sigma_{X_pY}\Sigma^{-1}_{YY}+\Sigma^{-1}_{Y/Z})Y \\ &\quad=\Sigma^{-1}_{YY}\Sigma_{YX_p}\Sigma^{-1}_{X_p/Y}(x_p-F_{X_p}\beta+\Sigma_{X_pY}\Sigma^{-1}_{YY}F_Y\beta)+\Sigma^{-1}_{Y/Z}z\,, \end{aligned}$$

$$\begin{aligned} &\partial\big(-2\ln f_{Y,\beta/X_p,Z}(y,\beta|x_p,z)\big)/\partial\beta = 0 \;\Rightarrow \\ &\big(F_{X_p}-\Sigma_{X_pY}\Sigma^{-1}_{YY}F_Y\big)'\Sigma^{-1}_{X_p/Y}\big((F_{X_p}-\Sigma_{X_pY}\Sigma^{-1}_{YY}F_Y)\beta-(X_p-\Sigma_{X_pY}\Sigma^{-1}_{YY}Y)\big)=0\,. \end{aligned}$$

Using the auxiliary matrices $\Sigma_1,\Sigma_2,\Sigma_3,\Sigma_4$ and vectors v_1,v_2, we get the equations

$$\begin{aligned} \Sigma_1 Y+\Sigma_2\beta &= v_1\,,\\ \Sigma_3\beta+\Sigma_4 Y &= v_2\,. \end{aligned}$$

Assuming, that $\begin{pmatrix}\Sigma_1 & \Sigma_2\\ \Sigma_3 & \Sigma_4\end{pmatrix}$ is a square, full ranked matrix, we have:

$$\begin{pmatrix}\widehat{Y}\\ \widehat{\beta}\end{pmatrix} = \begin{pmatrix}\Sigma_1 & \Sigma_2\\ \Sigma_4 & \Sigma_3\end{pmatrix}^{-1}\begin{pmatrix}v_1\\ v_2\end{pmatrix}.$$

□

Another generalization of the first result is possible, if we weaken the assumption of known covariance matrices. Let's assume, that we know only the linear relations between the individual time observations and linear relations between externally predicted future values of time series and also some measure of accuracy of prior predictions. The model is:

$$X_p/Y,\phi \sim N\big(\mu_{X_p/Y},\phi^{-1}C_{X_p/Y}\big)\,,$$
$$Y/Z,\phi \sim N(Z,\phi^{-1}\alpha C_Z)\,,$$
$$\phi \sim \Gamma(n_0/2,d_0/2)\,.$$

Matrices $C_{X_p/Y}, C_Z$ are known correlation matrices, $\big(C_{X_p/Y} = C_{X_pX_p} - C_{X_pY}C_{YY}^{-1}C_{YX_p}\big)$, α is a known measure of accuracy of external prediction compared to past observations, ϕ is an unknown parameter of dispersion, it's prior distribution is Gamma distribution with parameters n_0, d_0. From the theory of Gamma distribution it is known:

$$E(\phi) = n_0/d_0,\ \ d_0\phi \sim \chi^2_{n_0}\,,$$
$$f(\phi) = k\cdot\phi^{n_0-1}\exp(-\phi d_0)\,.$$

THEOREM 3. *Let $X_p/Y,\ \phi \sim N(\mu_{X_p/Y}\phi^{-1}C_{X_p/Y})$, $Y/Z,\phi \sim N(Z,\phi^{-1}\alpha C_Z)$, $\phi \sim \Gamma(n_0/2,d_0/2)$. Matrices $C_{X_p/Y}, C_Z$ are known, positive definite correlation matrices, $C_{X_p/Y} = C_{X_pX_p} - C_{X_pY}C_{YY}^{-1}C_{YX_p}$, parameter α is known. Distribution of predictions is then*

$$Y/\phi \sim N(\mu,\phi^{-1}C)\,,$$
$$Y \sim T_{n_0}(\mu, d_0/n_0 C)\,,$$

with

$$\mu = C\Big(C_{YY}^{-1}C_{YX_p}C_{X_p/Y}^{-1}\big(X_p - \mu_{X_p} + C_{X_pY}C_{YY}^{-1}\mu_Y\big) + (\alpha C_Z)^{-1}Z\Big)\,,$$
$$C = \Big(C_{YY}^{-1}C_{YX_p}C_{X_p/Y}^{-1}C_{X_pY}C_{YY}^{-1} + \big(\alpha C_Z\big)^{-1}\Big)^{-1}\,,$$

distribution T_n is multidimensional Student's distribution, if V has dimension p, $V \sim T_n(m,\mathbb{R})$, then it's density function is

$$f_{T_n}(V) = k\big(n + (V-m)'R^{-1}(V-m)\big)^{-(n+p)/2}.$$

P r o o f . The standard theory of normal/gamma distribution implies that if $V/\phi \sim N(m,\phi^{-1}\mathbb{R})$, $\phi \sim \Gamma(n/2,d/2)$, then the distribution V unconditional on ϕ is $V \sim T_n(m,\mathbb{R})$. From the theorem there is

$$Y/\phi \sim N(\mu,\phi^{-1}C)\,.$$

It is now obvious that

$$Y \sim T_{n_0}(\mu, d_0/n_0 C)\,.$$

□

In the models described in previous theorems we always had a prior information in the form of prediction for every time value of future of time series. In practice we may find the cases when such predictions are not available, but there exists another partial information. In a current life it very often happens that the future values have to reach some value in aggregate: $Y_{n+1}+Y_{n+2}+\cdots+Y_{n+k}=a$. A company usually specifies the plan for year spending, or budget organization may also specify the year budget to spend. If we have monthly data from the past years, we may be concerned in predicting the monthly spending for the next year.

We will show first some elementary results about properties of distribution based on some external information.

Lemma 4. *Let $(V_1, V_2, \dots, V_k)'$ have a uniform distribution on the boundary of simplex $S = \{(v_1, v_2, \dots, v_k) \in \mathbb{R}^k \colon v_1 + v_2 + \cdots + v_k = a,\ v_1 \ge 0,\ v_2 \ge 0, \dots, v_k \ge 0\}$. The marginal distribution of individual random variables V_i, $i = 1, 2, \dots, k$, is then:*

The distribution function:

$$F_{V_i}(x) = 1-(1-x/a)^{k-1}, \quad \text{for } x \in (0,a),$$
$$F_{V_i}(x) = 0, \quad x \le 0, \qquad F_{V_i}(x) = 1, \quad x \ge a.$$

The density:

$$f_{V_i}(x) = (k-1)(1-x/a)^{k-2}/a, \quad \text{for } x \in (0,a),$$
$$f_{V_i}(x) = 0, \quad x \le 0,\ x \ge a.$$

The first and second moment:

$$E(V_i) = a/k, \quad D(V_i) = a^2\left(\frac{2}{k(k+1)} - \frac{1}{(k+1)^2}\right).$$

Proof. The symmetry implies that it is sufficient to consider only the random variable V_1. Let's determine $F_{V_1}(x) = P(V_1 < x)$, $x \in (0,a)$: Let $S_1 = \{(v_1, \dots, v_k) \in \mathbb{R}^k \colon v_1 + \cdots + v_k = a,\ x \ge v_1 \ge 0,\ v_2 \ge 0, \dots, v_k \ge 0\}$. From the uniform distribution of probability on S we have

$$P(V_1 < x) = \frac{\lambda(S_1)}{\lambda(S)},$$

where λ is the Lebesgue measure in $\mathbb{R}^{k-1}$. Geometrical shapes $S - S_1$ and S are similar to each other, index of similarity is $\big((a-x)/a\big)^{k-1}$, from that

$$P(V_1 < x) = \lambda(S_1)/\lambda(S) = 1 - \lambda(S - S_1)/\lambda(S) = 1-(1-x/a)^{k-1}.$$

From the distribution function the density, mean value and dispersion follow directly from the definitions. □

Let's assume that past and future of time series $(X_p, Y)' \in \mathbb{R}^{N+k}$ have a normal distribution. That implies $X_p/Y \sim N(\mu_{X_p/Y}, \Sigma_{X_p/Y})$. The distribution with respect to the external information is determined by function δ_S, which is not a density in $\mathbb{R}^k$. The function δ_S is density on a part of hyperplane with the dimension $k-1$. It means, that $Y = (Y_{n+1}, Y_{n+2}, \dots, Y_{n+k})'$ has a uniform distribution on a boundary of simplex S.

THEOREM 5. *Let $X_p/Y \sim N(\mu_{X_p/Y}, \Sigma_{X_p/Y})$, $Y/a \sim \delta_S$ The optimal prediction by maximum likelihood method is*

$$\widehat{Y} = \mu_{Y*} + \frac{KB^{-1}\cdot 1}{1'B^{-1}1}\,,$$

with

$$\begin{aligned} K &= a - 1'\mu_{Y*}\,, \\ \mu_{Y*} &= \mu_Y + B^{-1}\Sigma_{YY}^{-1}\Sigma_{YX_p}\Sigma_{X_p/Y}^{-1}(X_p - \mu_{X_p})\,, \\ B &= \Sigma_{YY}^{-1}\Sigma_{YX_p}\Sigma_{X_p/Y}^{-1}\Sigma_{X_pY}\Sigma_{YY}^{-1}\,, \end{aligned}$$

and 1 is a vector of k ones: $1 = (1, 1, \dots, 1)'$.

Proof.

$$\begin{aligned} \operatorname{argmax}\ k f_{X_p/Y}(x_p|y)\cdot\delta_S &= \operatorname{argmax} k f_{X_p/Y}(x_p|y) \\ &= \operatorname{argmin}\ k_1 - 2\ln f_{X_p/Y}(x_p|y)\,, -2\ln f_{X_p/Y}(x_p|y) \\ &= (X_p - \mu_{X_p/Y})'\Sigma_{X_p/Y}^{-1}(X_p - \mu_{X_p/Y})\,. \end{aligned}$$

We are seeking the minimum of $(X_p - \mu_{X_p/Y})'\Sigma_{X_p/Y}^{-1}(X_p - \mu_{X_p/Y})$, with the condition that sum of the elements of Y is a.

The set $\{Y \in \mathbb{R}^k\colon (X_p - \mu_{X_p/Y})'\Sigma_{X_p/Y}^{-1}(X_p - \mu_{X_p/Y}) = d\}$ is a $(k-1)$-dimensional ellipsoid in $\mathbb{R}^k$, the set $S = \{(Y_1, \dots, Y_k) \in \mathbb{R}^k\colon Y_1 + \dots + Y_k = a\}$ is a part of a hyperplane, it is a boundary of a simplex. The minimum $(X_p - \mu_{X_p/Y})'\Sigma_{X_p/Y}^{-1}(X_p - \mu_{X_p/Y})$ is reached if the ellipsoid touches S'. We will find it by the Lagrange multiplicator method:

$$\begin{gathered} f(Y) = (X_p - \mu_{X_p/Y})'\Sigma_{X_p/Y}^{-1}(X_p - \mu_{X_p/Y}) - \alpha(Y_1 + \dots + Y_k - a)\,, \\ \partial f/\partial Y: 2B(Y - \mu_y) - 2\Sigma_{YY}^{-1}\Sigma_{X_pY}\Sigma_{X_p/Y}^{-1}(X_p - \mu_{X_p}) - \alpha\cdot 1 = 0\,, \\ \partial f/\partial\alpha:\ Y'1 = a\,. \end{gathered}$$

The solution of the system is

$$Y = \mu_{Y*} + \frac{KB^{-1}\cdot 1}{1'B^{-1}1}\,.$$

□

REFERENCES

[1] ANDĚL, J.: *Statistická analýza časových řad*, SNTL, Praha, 1976.

[2] ANONYMOUS: *Splus4 Guide to Statistics*, Data Analysis Products Division MathSoft, Inc. Seatle, Washington, 1997.

[3] BROCKWELL, P. J.—DAVIS, R. A.: *Introduction to Time Series and Forecasting*, Springer-Verlag, New York, 1996.

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

[5] GODOLPHIN, E. J.—STONE, J. M.: *On the structural representation for polynomial-projecting predictor models based on Kalman filter*, J. Roy. Statist. Soc. (Ser.B) **42** (1980), 35–45.

[6] GOLDSTEIN, M.: *Bayesian analysis of regression problems*, Biometrika **63** (1976), 51–58.

[7] HARRISON, P. J.—STEVENS, C.: *Bayesian forecasting (with discussion)*, J. Roy. Statist. Soc. (Ser. B) **38** (1976), 205–247.

[8] HARTIGAN, J. A.: *Linear Bayesian methods*, J. Roy. Statist. Soc. (Ser. B) **31** (1969), 446–454.

[9] ŠTULAJTER, F.: *Random Processes and Time Series*, Comenius University, Bratislava, 2001.

[10] WEST, M.—HARRISON, J.: *Bayesian Forecasting and Dynamic Models*, Springer-Verlag, New York, 1997.

[11] VENABLES, W. N.—RIPLEY, B. D.: *Modern Applied Statistics with S+*, Springer-Verlag, New York, 2000.

Received March 28, 2002

Department of Probability and Statistics
Faculty of Mathematics, Physic and Informatics
Comenius University
Sk–842 48 Bratislava
SLOVAKIA

E-mail: katina@fmph.uniba.sk