View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Electronic National Technical University "Kharkiv Polytechnic Institute"...

FORECASTING METHODS OF NON-STATIONARY STOCHASTIC PROCESSES THAT USE EXTERNAL CRITERIA

Igor V. Kononenko, Anton N. Ryepin

National Technical University "Kharkiv Polytechnic Institute", Ukraine

1. Introduction

While forecasting the development of socio-economic systems there often arise the problems of forecasting non-stationary stochastic processes having a scarce number of observations (5-30), while the repeated realizations of processes are impossible.

To solve such problems there have been suggested a number of methods, in which unknown parameters of the model are estimated not at all points of time series, but at a certain subset of points, called a learning sequence. At the remaining points not included in the learning sequence and called the check sequence, the suitability of the model for describing the time series is determined. These methods include the method of cross-validation and Group method of data handling (GMDH) proposed by Alexey G. Ivahnenko. The disadvantage of these methods is that a certain combination of data partitions is set in advance and it does not take into account the specifics of the task.

2. Purpose of work

The purpose of this work is to create and study an effective forecasting method of non-stationary stochastic processes in the case when observations in the base period are scarce.

3. H-criterion method (I. Kononenko, 1982)

The data including retrospective information can be presented in a form of matrix $\Gamma = \|\gamma_{r,i}\|$, $r = \overline{1,q}$, $i = \overline{1,n}$, where q – number of significant variables including

the predicted variable; n – number of points in the time base of forecast; $(\gamma_{1,1}, \gamma_{1,2}, ..., \gamma_{1,n})$ - vector of values of the predicted variable.

The list of elementary models is formed. It includes different mathematical models, which by hypothesis can be included in the final forecasting model. The elementary power, exponential, logarithmic, trigonometric, rational and other functions are used. From the models in the list the linear combinations of 1,2,...,M models are formed comprising the set of test models. For each test model the estimation of its suitability for forecasting is made.

The matrix Γ is further divided into two submatrices – learning submatrix Γ_L and check submatrix Γ_C. The division is made by means of selecting the first *n*/2 columns of matrix Γ as Γ_L and the remaining columns as Γ_C. If *n* is odd then (*n*–1)/2 columns should be selected. The parameters of all formed models are estimated using the learning submatrix

$$\hat{A}^{(j)} = \eta(\rho, G_L), \qquad (1)$$

where $A^{(j)}$ – the vector of estimated parameters for *j*-th model, $A^{(j)} = [a_1, a_2, \ldots, a_p]^T$; $\hat{A}^{(j)}$ – the vector of estimates for $A^{(j)}$; ρ – the vector of weighting coefficients considering the error variance or importance γ_{1i} for building the model, $i = \overline{1, N_L}$, $\rho = [\rho_1, \rho_2, \ldots, \rho_{N_L}]$; $\eta(\ldots)$ – function that is set analytically or algorithmically. The estimation of parameters is made by methods most appropriate for the situation at hand. When choosing a method the following criteria must be taken into account: the kind of test models, the existing assumptions about additivity and multiplicativity of errors, about the error distribution law, about the class it might belong to, about the error correlation and other information.

The loss-function *F*(ε) is selected according to the available information about the error distribution law or the class of such laws.

After the estimation of parameters of all test models according to formula (1) for each *j*-th model at all points of past history we calculate

$$\Delta_1 = \sum_{i=1}^{n} \rho_i F(\gamma_{1i} - \varphi_j(F_i^{(j)}, \hat{A}^{(j)})), \qquad (2)$$

where ρ_i – weighting coefficient; $F(\varepsilon)$ – loss-function, selected according to the available information about the distribution law of errors ε_i or the class of such laws. Then Γ_C is used as a learning submatrix, and Γ_L as a check one, and for all models the process of parameters estimation and calculation of Δ_2 values is repeated.

In the matrix Γ new learning and check submatrices are chosen. The number of rows in Γ_L is decreased by one. The process of estimation of model parameters and calculation of Δ_3 is repeated, the learning submatrix is used as the check one and the check submatrix - as the learning one, Δ_4 is calculated and similarly we continue using the bipartitioning. The process is stopped after a set number of iteration *g*. Among test models estimated by different methods the one with the minimum value of *H*-criterion is selected.

$$H = \Delta_1 + \Delta_2 + \ldots + \Delta_g . \qquad (3)$$

The obtained model is used for forecasting.

4. Bootstrap evaluation method (I. Kononenko, 1990)

The data including retrospective information can be presented in a form of matrix $G = \|\gamma_{j,i}\|$, $j = \overline{1,q}$, $i = \overline{1,n}$, where q – number of significant variables including the predicted variable, n – volume of past history, $(\gamma_{1,1}, \gamma_{1,2}, \ldots, \gamma_{1,n})$ – vector of values of the predicted variable.

Let $L = 1$, where L – the number of a model in the set of test models. Let the model $f^L(\mathbf{N}_i, \mathbf{B})$ be tested for the description of the observed process, i.e. we get the expression $\gamma_{1,i} = f^L(\mathbf{N}_i, \mathbf{B}) + \xi_i$, where N_i – vector of independent variables, B – vector of estimated parameters, ξ_i – independent errors having the same and symmetrical density of distribution, $i = \overline{1,n}$.

1. The parameters of the model $f^L(\mathbf{N_i}, \mathbf{B})$ we estimate using matrix G basing on the condition

$$\hat{\mathbf{B}} = \arg\min_{\mathbf{B}} \sum_{i=1}^{n} F\left(\gamma_{1,i} - f^L(\mathbf{N}_i, \mathbf{B})\right),$$

where $F(\xi_i)$ – loss function, $\xi_i = \gamma_{1,i} - f^L(\mathbf{N}_i, \mathbf{B})$, $i = \overline{1,n}$. The loss function is selected depending on the available assumptions about the errors additively imposed on the true model. Thus, $F(\xi) = |\xi|$ or $F(\xi) = (\xi)^2$. For the model $f^L(\mathbf{N}_i, \hat{\mathbf{B}})$ we determine the deviation from points of G, $\text{bias}_i = \gamma_{1,i} - f^L(\mathbf{N}_i, \hat{\mathbf{B}})$, $i = \overline{1,n}$. Numbers bias_i form the **BIAS** vector.

2. We divide the matrix G into two submatrices – learning submatrix G_L and check submatrix G_C. We include first *n*-1 columns of matrix G in submatrix G_L and *n*-th column in G_C. The learning submatrix has the following form $G_L = \|\gamma_{j,i}\|$, $j = \overline{1,q}$, $i = \overline{1,n-1}$, and the check one – $G_c = \|\gamma_{j,n}\|$, $j = \overline{1,q}$.

Using the learning submatrix G_L we estimate the parameters **B** of the test model by the above mentioned method and obtain $\hat{\mathbf{B}}^{(0)}$ as a result. Basing on the check submatrix we calculate the deviation of the model from the statistics $D_0^L = \left(\gamma_{1,n} - f^L(\mathbf{N}_n, \hat{\mathbf{B}}^{(0)})\right)^2$.

Let k=1, where k – number of iteration, which performs the bootstrap evaluation.

3. We perform bootstrap evaluation, which consists in the following. We randomly (with equal probability) select numbers from the BIAS vector and add them to the values of model $f^L(\mathbf{N}_i, \hat{\mathbf{B}})$. As a result we obtain "new" statistics $\gamma_{1,i}^k$, $i = \overline{1,n}$, which looks like the following

$$\gamma_{1,i}^k = f^L(\mathbf{N}_i, \hat{\mathbf{B}}) + \text{bias}_s,\ i = \overline{1,n},\ s \in \{1,2,\ldots,n\}.$$

Then we divide the matrix G_k (a new one this time) into $G_{L,k}$ and $G_{C,k}$. We estimate the unknown parameters basing on $G_{L,k}$ as earlier and calculate the model deviation from $G_{C,k}$

$$D_k^L = \left(\gamma_{l,n}^k - f^L\left(\mathbf{N}_n, \hat{\mathbf{B}}^{(k)}\right)\right)^2.$$

4. If k<K-1 then we suppose that k:=k+1 and return to step 3 (where K – number of bootstrap iterations), otherwise proceed to step 5.

5. We evaluate

$$D^L = \sum_{k=0}^{N-1} D_k^L.$$

6. If L<z then we suppose that L:=L+1 and move to step 1 (where z – number of models in the list), otherwise we stop.

The model with minimal D^L is considered to be the best one.

5. The analysis of forecasting methods

A computational analysis of the suggested forecasting methods has been performed. The following mathematical models have been chosen for the analysis:

$y = x^2 + 2x + 3$, $y = -x^2 + 6x + 3$, $y = 2x^2 + 8x + 3$, $y = -x^2 + 16x + 3$,

$y = x^2 + 6x + 11$, $y = -x^2 - 2x + 11$, $y = 2x^2 - 16x + 27$, $y = -2x^2 - 8x + 27$,

hereinafter referred to as true models. On each of these models defined at points $x_i = 0{,}1 \cdot i$, $i = \overline{1,10}$ we imposed an additive noise $\xi \sim N(0, \sigma^2)$, where

$$\sigma = 0.3 \cdot \sqrt{\sum_{i=1}^{10} (y_i - \sum_{i=1}^{10} y_i / 10) \Big/ 9},$$

where y_i – value of the model at point x_i, $i = \overline{1,10}$, and then defined the best forecasting model by means of the suggested methods. The loss function of the form $F(\xi) = (\xi)^2$ was chosen as it is the most frequently used in practice.

During the analysis we considered all combinations of one, two, three functions from the list $x^{1/2}$, x, $x^{3/2}$, x^2, $x^{5/2}$, x^3, x^{-1}, $x^{-1/2}$, $x^{-3/2}$ in form of their linear combinations. We analyzed the properties of the method when forecasting on *d*

points, $d = 1,2,3,5,10$. For every forecasting model obtained we calculated the following characteristics:

- Relative percent mean absolute deviation (PMAD) evaluated at the estimation period

$$\theta = \sum_{i=1}^{10} |z_i - \hat{y}_i| \Big/ \sum_{i=1}^{10} |z_i|;$$

where $z_i = y_i + \xi_i$, $\hat{y}_i$ – value of the obtained forecasting model at point *i*, $i = \overline{1,10}$;

- Percent mean absolute deviation (PMAD) evaluated at the estimation period

$$E = \sum_{i=1}^{10} |y_i - \hat{y}_i| \Big/ \sum_{i=1}^{10} |y_i|;$$

- Percent mean absolute deviation (PMAD) evaluated at the forecasting period

$$E1_d = \sum_{i=11}^{10+d} |y_i - \hat{y}_i| \Big/ \sum_{i=11}^{d} |y_i|;$$

- Relative mean squared error (MSE) evaluated at the forecasting period

$$D_d^m = \frac{1}{d} \sum_{i=11}^{10+d} (z_i - \hat{y}_i)^2;$$

- Mean squared error (MSE) evaluated at the forecasting period

$$D_d^t = \frac{1}{d} \sum_{i=11}^{10+d} (y_i - \hat{y}_i)^2.$$

The analysis is performed on $N = 1000$ realizations of noise.

For the H-criterion method, the division of data into learning and checking submatrices was done in accordance with the rules determined by the matrices

$$R_1 = \begin{bmatrix} 12121212 \\ 12121212 \\ 12121212 \\ 12121221 \\ 12211221 \\ 21211221 \\ 21212121 \\ 21212121 \\ 21212121 \\ 21212121 \end{bmatrix}, \; R_2 = \begin{bmatrix} 21212121 \\ 21212121 \\ 21212121 \\ 12212121 \\ 12211221 \\ 12211212 \\ 12121212 \\ 12121212 \\ 12121212 \\ 12121212 \end{bmatrix}, \; R_3 = \begin{bmatrix} 1111111112 \\ 1111111121 \\ 1111111211 \\ 1111112111 \\ 1111121111 \\ 1111211111 \\ 1112111111 \\ 1121111111 \\ 1211111111 \\ 2111111111 \end{bmatrix}, R_4 = \begin{bmatrix} 2121211212 \\ 2112221122 \\ 1111112111 \\ 1222121112 \\ 2121222111 \\ 2121221212 \\ 1211111121 \\ 2121221121 \\ 1112121111 \\ 1221112112 \end{bmatrix}.$$

Every *j*-th column of the matrix R_d, $d = \overline{1,4}$ corresponds to the *j*-th method of data division, $j = \overline{1,8}$ for R_1, R_2 and $j = \overline{1,10}$ for R_3, R_4. Every $r_{ij}^{(d)}$-th element of matrix R_d $j = \overline{1,10}$ determines, into which submatrix – learning (G_L) or checking (G_C) – goes *i*-th point of history. Here $r_{ij}^{(d)} = 1$ means that the point is used in submatrix G_L, $r_{ij}^{(d)} = 2$ means that the point is used in submatrix G_C.

Matrix R_3 corresponds to the cross-validation procedure that served as the source for comparison.

Matrix R_4 is the randomly generated matrix.

For the bootstrap evaluation method the number of bootstrap iterations was selected from 10, 20 to 50.

We calculated:

- Average (across noise realizations) relative percent mean absolute deviation (PMAD) evaluated at the estimation period

$$\overline{\theta} = \frac{1}{N} \sum_{k=1}^{N} \theta_k ;$$

- Average (across noise realizations) percent mean absolute deviation (PMAD) evaluated at the estimation period

$$\overline{E} = \frac{1}{N} \sum_{k=1}^{N} E_k ;$$

- Average (across noise realizations) percent mean absolute deviation (PMAD) evaluated at the forecasting period

$$\overline{E}1_d = \frac{1}{N}\sum_{k=1}^{N} E1_{dk};$$

- Average (across noise realizations) relative mean squared error (MSE) evaluated at the forecasting period

$$\overline{\mathrm{D}}_{\mathrm{d}}^{\mathrm{m}} = \frac{1}{\mathrm{N}}\sum_{\mathrm{k}=1}^{\mathrm{N}} \mathrm{D}_{\mathrm{dk}}^{\mathrm{m}};$$

- Average (across noise realizations) mean squared error (MSE) evaluated at the forecasting period

$$\overline{\mathrm{D}}_{\mathrm{d}}^{\mathrm{t}} = \frac{1}{\mathrm{N}}\sum_{\mathrm{k}=1}^{\mathrm{N}} \mathrm{D}_{\mathrm{dk}}^{\mathrm{t}},$$

where θ_k, E_k, $E1_{dk}$, D_{dk}^m and D_{dk}^t – error values for k-th realization of noise, $k = \overline{1,N}$. Confidence intervals of 95 percent have been estimated for $\overline{\theta}$, $\overline{E}$, $\overline{E}1_d$, $\overline{D}_d^m$ and $\overline{D}_d^t$.

The comparison of the efficiency of the suggested methods and cross-validation method has been made. Using the same analysis algorithm and the initial data as for analysis of the suggested methods, the investigation of cross-validation method has been conducted and the values of characteristics $\overline{\theta}$, $\overline{E}$, $\overline{E}1$, $\overline{D}^m$, $\overline{D}^t$ were obtained, also 95% confidence intervals for these characteristics have been built.

We compared the characteristics with the two-sample t-test assuming the samples were drawn from the normally distributed populations, which in the context of the considered problem has the following form

$$P\{|v| \geq V(N,Q)\} \approx 2 \cdot Q,$$

where $V(N,Q)$ – value defined by the table that corresponds to the significance level of Q, v – value calculated by the following formula

$$v = \eta \Big/ \sqrt{\frac{s_1^2 + s_2^2}{N}},$$

where $\eta = \bar{\xi}_1 - \bar{\xi}_2$, $\bar{\xi}_1$ and $\bar{\xi}_2$ – compared characteristics, s_1 and s_2 – estimates of root-mean-square differences of $\bar{\xi}_1$ and $\bar{\xi}_2$ correspondingly, $N = 1000$.

The significance level of Q is said to be equal to 2,5 %. For $Q = 2{,}5$ and $N = 1000$ $V(N,Q) = 1{,}96$.

The values of v calculated for pairs of compared characteristics (for H-criterion method matrices R_1, R_4, R_6 were selected) have been analyzed.

The values of characteristics of the suggested forecasting methods are significantly less (with the 95% of confidence probability) at the forecasting period than the values of characteristics of cross-validation method for all true models considered and intervals of the forecasting period.

Figure 1 depicts how the PMAD $\bar{E}1_d$, evaluated at the forecasting period changes for mathematical model $y = x^2 + 2x + 3$ depending on the number of partitions *g*. In the given case d=10, i.e. the forecasting is performed at 10 points.

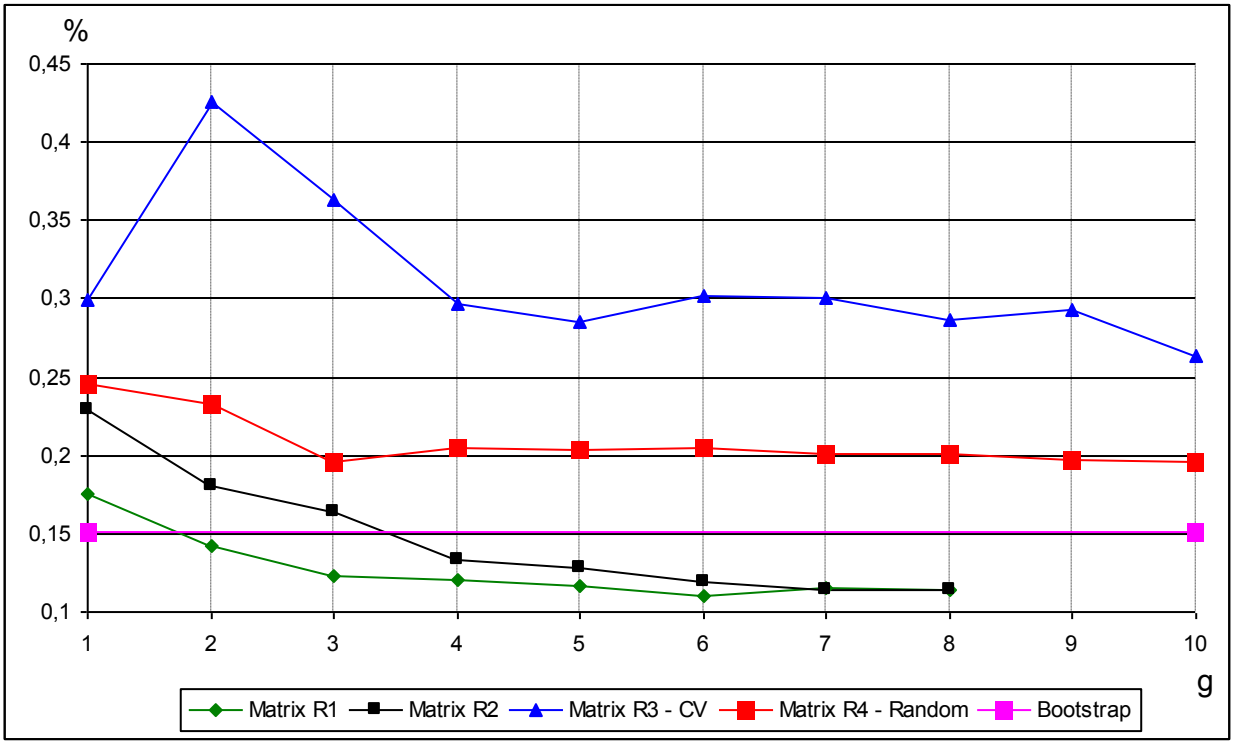

Figure 1 - Percent mean absolute deviation (PMAD) evaluated at the forecasting period

Having analyzed the given example we can draw a number of conclusions:

- when the number of partitions increases in case of using matrices R_1 and R_2 we observe the downward trend of $\overline{E}1_d$ with some fluctuations in this trend that depend on the ways of data partition;

- the partition according to the cross-validation procedure, in which the check points fall into the observation interval, produces significantly less accurate forecasts. The comparison of the efficiency of different partitions with randomly generated matrix R_6 has shown that the reasonable choice of partition sequences permits to get a more accurate longer-term forecast;

- the bootstrap evaluation method, which requires no learning or checking matrices, produces the more accurate forecast than the cross-validation procedure

- the comparison of two suggested methods enables to state that the bootstrap evaluation method makes it possible to obtain more accurate longer-term forecasts as compared with H-criterion method only in case of a small number of partitions. Otherwise the usage of selected matrices R_1 and R_2 permits to get more accurate forecasts. Nevertheless, the bootstrap evaluation method turned out to be more accurate than the H-criterion method when using matrix R_6.

The similar chars can be observed for the remaining mathematical models used in the analysis.

The number of bootstrap iterations reasonable for using in the corresponding methods has been determined. In case of analyzed models the number of bootstrap iterations that allowed to reduce PMAD evaluated at the forecasting period was 40. By changing the number of bootstrap iterations from 10 to 40 the value of PMAD decreased and reached its minimum at 40, and than started to increase as the number of iterations reached 50.

Thus, we conclude that the suggested methods are more accurate in the forecasting period than the cross-validation method. Such conclusion permits to recommend them for forecasting of non-stationary stochastic processes when the number of points in the base period is small.

The suggested bootstrap evaluation method has helped in solving the tasks of forecasting the sales volume of wheel tractors in the USA and the production

volume of bread and bakery in Kharkiv region (Ukraine). The latest is shown on the figure 2. It should be noticed, that the forecast was made in 2002 and was not corrected since then. The mean relative error for the period 2003-2006 is 5,91 %.

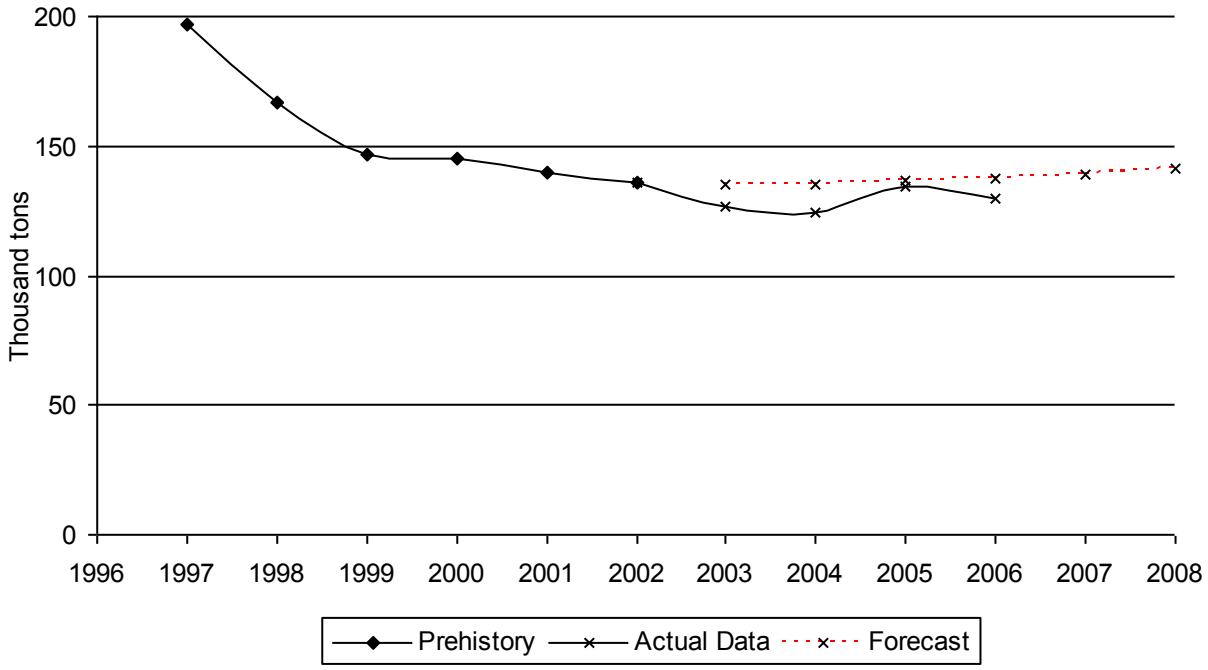

Figure 2 - Production volume of bread and bakery in Kharkiv region

When solving different forecasting problems it is important to determine the appropriateness of using one of the methods. The analysis results have shown that when the number of partitions is large the H-criterion method produces the more accurate longer-term forecasts than the bootstrap evaluation method. However, in the real-life problems the bootstrap evaluation method might turn out to be more accurate in the number of cases. That is why it is recommended to use the given methods together. In such case every result obtained by means of these methods must be assigned some weight on the basis of the a priori estimates of the methods accuracy. The final forecast will be received in the form the weighted average value of individual forecasts.