Combined Identification Algorithms

Yury G. Dmitriev, Gennady M. Koshkin and Vadim Yu. Lukov
National Research Tomsk State University, Tomsk, Russia
e-mail: dmit@mail.tsu.ru, kgm@mail.tsu.ru, v-lukov@rambler.ru

Abstract

In many applied problems it is required to construct a mathematical model of the dependence of output variables on input variables of the stochastic object. To solve this problem, both parametric and nonparametric approaches are used. Each of these approaches has advantages and disadvantages. In the paper, we consider combined algorithms for the identification of stochastic objects using jointly nonparametric and parametric estimates of regression.

Keywords: Nadaraya–Watson statistic, parametric estimate, regression, combined algorithm, identification, bootstrap.

Introduction

Suppose that a stochastic object is described by a regression function

$$r(\vec{x}) = M(Y|\vec{X} = \vec{x}) = \int yp(y|\vec{x})dy = \frac{\int yp(\vec{x}, y)dy}{p(\vec{x})}, \tag{1}$$

where $\left(\vec{X}, Y\right) = \left(X^{(1)}, ..., X^{(p)}, Y\right)$ is a $(p+1)$-dimensional vector of p object's inputs and output, $p(\vec{x}, y)$ is their joint distribution density, $p(\vec{x})$ is a distribution density of inputs, and $p(y|\vec{x})$ is the conditional distribution density.

Let there be independent observations $\left(\vec{X}_i, Y_i\right) = \left(X_i^{(1)}, ..., X_i^{(p)}, Y_i\right)$, $i = 1, ..., n$, of the random vector $\left(\vec{X}, Y\right)$. Let us consider the nonparametric Nadaraya–Watson [2, 11] estimate of the regression function (1)

$$\hat{r}(\vec{x}) = \hat{r}(\vec{x}; \vec{X}_1, ..., \vec{X}_n) = \frac{\sum_{i=1}^{n} Y_i K\left(\frac{\vec{x} - \vec{X}_i}{\vec{h}_n}\right)}{\sum_{t=1}^{n} K\left(\frac{\vec{x} - \vec{X}_i}{\vec{h}_n}\right)}, \tag{2}$$

where $K\left(\frac{\vec{x} - \vec{X}_i}{\vec{h}_n}\right) = K\left(\frac{x^{(1)} - X_i^{(1)}}{h_n^{(1)}}\right) \cdots K\left(\frac{x^{(p)} - X_i^{(p)}}{h_n^{(p)}}\right)$ is a p-dimensional kernel (the product of p one-dimensional kernels), $\vec{h}_n = (h_n^{(1)}, \dots, h_n^{(p)})$ is a p-dimensional vector of bandwidth parameters.

Usually the researcher has some information about the nature of the dependence of the output of the object from the inputs. Suppose that he can express this knowledge in the form of a given function $\varphi(\vec{x},\vec{\theta})$, where $\vec{\theta}=(\theta^{(1)},\dots,\theta^{(s)})$ is the vector of the known parameters. This type of information we call as a prior guess.

Consider the task of sharing the nonparametric estimation of regression and a prior guess. The approach using combinations of different estimates was studied, for example, in [1]–[3],[9].

1 Combined estimators

1.1 Static model

As a combined regression estimate, we take [2, 9]

$$\hat{R}_\lambda(\vec{x})=(1-\lambda)\hat{r}(\vec{x})+\lambda\varphi(\vec{x},\vec{\theta}), \tag{3}$$

where λ is the weight coefficient determined from minimum of the criterion

$$M\{\hat{R}_\lambda(\vec{x})-r(\vec{x})\}^2. \tag{4}$$

So, from (4) we obtain the optimal λ:

$$\lambda(\vec{x})=\frac{M\{(\hat{r}(\vec{x})-r(\vec{x}))(\hat{r}(\vec{x})-\varphi(\vec{x},\vec{\theta}))\}}{M\{\hat{r}(\vec{x})-\varphi(\vec{x},\vec{\theta})\}^2}. \tag{5}$$

Substituting (5) into (4) and making the transformations, we get:

$$M\{\hat{R}_\lambda(\vec{x})-r(\vec{x})\}^2=M\{\hat{r}(\vec{x})-r(\vec{x})\}^2-\frac{[M\{(\hat{r}(\vec{x})-r(\vec{x}))(\hat{r}(\vec{x})-\varphi(\vec{x},\vec{\theta}))\}]^2}{M\{\hat{r}(\vec{x})-\varphi(\vec{x},\vec{\theta})\}^2}. \tag{6}$$

The second term in (6) shows how much the MSE of the combined estimate $\hat{R}_\lambda(\vec{x})$, taking into account the prior guess $\varphi(\vec{x},\vec{\theta})$, decreases compared to $\hat{r}(\vec{x})$ for each $\vec{x}\in R^p$. Since the optimal $\lambda(\vec{x})$ (5) is usually unknown, it becomes necessary to construct an estimate $\hat{\lambda}(\vec{x})$ of this coefficient, which leads to an adaptive combined estimate

$$\hat{R}_{\hat{\lambda}}(\vec{x})=(1-\hat{\lambda}(\vec{x}))\hat{r}(\vec{x})+\hat{\lambda}(\vec{x})\varphi(\vec{x},\vec{\theta}). \tag{7}$$

Let us consider an estimate of a weight coefficient by a bootstrap method. We write (5) in the form:

$$\lambda(\vec{x})=\frac{M\psi_1(\vec{x})}{M\psi_2(\vec{x})}, \tag{8}$$

where

$$M\psi_1(\vec{x})=M[(\hat{r}(\vec{x})-r(\vec{x}))(\hat{r}(\vec{x})-\varphi(\vec{x},\vec{\theta}))],\quad M\psi_2(\vec{x})=M\{\hat{r}(\vec{x})-\varphi(\vec{x},\vec{\theta})\}^2.$$

Generate a bootstrap sample $\left(\vec{X}^*_j,Y^*_j\right)$, $\vec{X}^*_j=(\vec{X}^*_{1,j},...,\vec{X}^*_{n,j})$, $j=1,\dots,B$, for the numerator and denominator in (8). Then we have:

$$M\psi_1(\vec{x})\simeq\frac{1}{B}\sum_{j=1}^{B}[(\hat{r}(\vec{x};\vec{X}^*_j)-\hat{r}(\vec{x}))(\hat{r}(\vec{x};\vec{X}^*_j)-\varphi(\vec{x},\vec{\theta}))],$$

$$M\psi_2(\vec{x}) \simeq \frac{1}{B}\sum_{j=1}^{B}[\hat{r}(\vec{x};\vec{X}_j^*) - \varphi(\vec{x},\vec{\theta})]^2.$$

As a result, we obtain the following estimate of the weight coefficient (5):

$$\hat{\lambda}_B(\vec{x}) = \frac{\sum_{j=1}^{B}(\hat{r}(\vec{x};\vec{X}_j^*) - \hat{r}(\vec{x}))(\hat{r}(\vec{x};\vec{X}_j^*) - \varphi(\vec{x},\vec{\theta}))}{\sum_{j=1}^{B}[\hat{r}(\vec{x};\vec{X}_j^*) - \varphi(\vec{x},\vec{\theta})]^2}. \tag{9}$$

The usage of (9) in (7) leads to an adaptive combined estimate

$$\hat{R}_{\hat{\lambda}_B}(\vec{x}) = (1 - \hat{\lambda}_B(\vec{x}))\hat{r}(\vec{x}) + \hat{\lambda}_B(\vec{x})\varphi(\vec{x},\vec{\theta}). \tag{10}$$

If θ is evaluated by a sample, then the estimate (10) we will denote as $\tilde{R}_{\hat{\lambda}_B}(\vec{x})$. The properties of these estimates are illustrated below in section 3 by simulation.

1.2 Dynamic model

Consider the dynamic model (cf. [4]–[7],[10])

$$Y_t = f(\vec{X}_t) + \xi_t, \tag{11}$$

where Y_t is the output of the object at the time moment t, $\vec{X}_t = \left(X_t^{(1)}, \ldots, X_t^{(p)}\right)$ is the p-dimensional vector of the inputs at the time moment t, f is an unknown function, ξ_t is the sequence of the i.i.d. random variables with a nonnegative distribution density, $M\xi_t = 0$, $M\xi_t^2 < \infty$, $M\xi_t^3 = 0$, and $M\xi_t^4 < \infty$.

Assume that f is bounded and its form does not change in the time interval under study. As an prior guess about the form of f, take the function $\varphi(\vec{x},\vec{\theta})$ and consider the following combined adaptive estimate:

$$\hat{R}_{\hat{\lambda}_B}(\vec{x}_t) = (1 - \hat{\lambda}_B(\vec{x}_t))\hat{r}(\vec{x}_t) + \hat{\lambda}_B(\vec{x}_t)\varphi(\vec{x}_t,\vec{\theta}), \tag{12}$$

where

$$\hat{\lambda}_B(\vec{x}_t) = \frac{\sum_{j=1}^{B}(\hat{r}(\vec{x}_t;\vec{X}_j^*) - \hat{r}(\vec{x}_t))(\hat{r}(\vec{x}_t;\vec{X}_j^*) - \varphi(\vec{x}_t,\vec{\theta}))}{\sum_{j=1}^{B}[\hat{r}(\vec{x}_t;\vec{X}_j^*) - \varphi(\vec{x}_t,\vec{\theta})]^2}.$$

The estimate (12) is applied in section 4 for the analysis of stock prices on real data.

2 Modeling

Consider an illustrative example. Let

$$Y(x) = 10 + 1.8x^2 + \xi, \quad \varphi(\vec{x};\vec{\theta}) = \theta^{(1)} + \theta^{(2)}x, \quad \theta^{(1)} = 8, \quad \theta^{(2)} = 10.8,$$

where ξ is normally distributed random variable with $M\xi = 0$ and $D\xi = \sigma^2$.

For different noise variances and samples sizes n, we investigate the behavior of the combined regression estimate for this model. The qualities of models identification and forecasting will be characterized using average relative errors:

$$\delta(\hat{r}) = \frac{1}{n}\sum_{i=1}^{n}\frac{|Y_i - \hat{r}(X_i)|}{|Y_i|}100\%.$$

For $n = 10$ and $\sigma = 1$, the plots of realizations $Y(x)$, $\varphi(x;\theta^{(1)},\theta^{(2)})$, $r(x)$ and combined estimate for different $x \in [0,1]$ are shown in Fig. 1. The behavior of the estimate of the weight coefficient (9) is shown in Fig. 2.

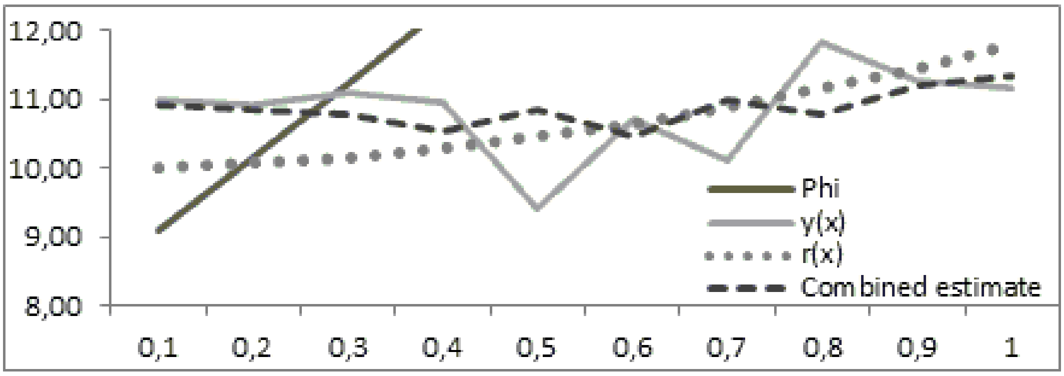

Figure 1: Plots of realizations $Y(x)$, $\varphi(x;\theta^{(1)},\theta^{(2)})$, $r(x)$ and combined estimate for $n = 10$ and $\sigma = 1$

Let $\varphi(x;\hat{\theta}^{(1)},\hat{\theta}^{(2)}) = \hat{\theta}^{(1)} + \hat{\theta}^{(2)}x$, where $\hat{\theta}^{(1)},\hat{\theta}^{(2)}$ are the least mean square (LMS) estimates. For this case, Table 1 gives average relative identification errors for various variances and samples sizes.

Table 1: Average relative identification errors $\delta(\hat{r})$, $\delta(\hat{R})$, and $\delta(\tilde{R})$

σ^2	1			3			5		
n	10	50	100	10	50	100	10	50	100
$\delta(\hat{r})$	4.99	3.87	2.33	23.88	15.43	8.96	56.14	30.20	17.44
$\delta(\hat{R})$	4.53	2.99	2.12	22.86	14.99	8.78	49.90	29.60	16.98
$\delta(\tilde{R})$	4.18	2.69	2.15	19.72	14.07	8.68	48.72	30.11	17.22

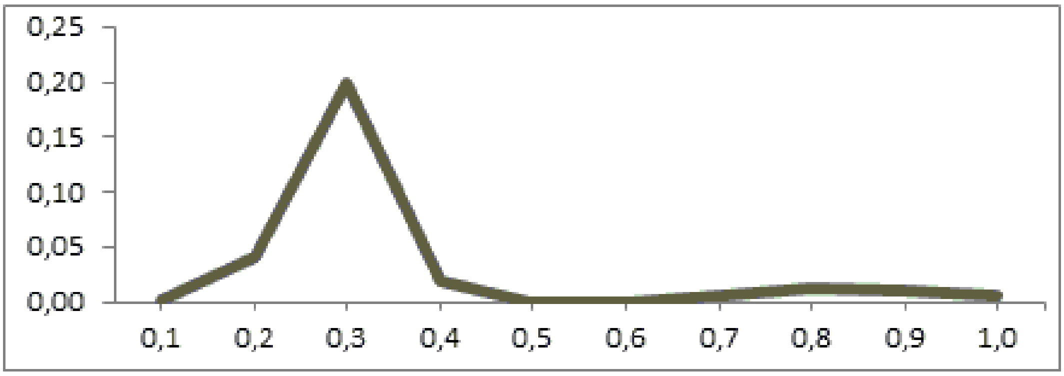

Figure 2: Plot of the dependence of the estimate of the weight coefficient (9) on x

In practice, from Table 1 it follows preferable applying a combined estimate in comparison with a nonparametric estimate in the case of small sample sizes and/or large noise variances.

3 Analysis of real data

The analysis of the prices of Gazprom's stocks for 2016 is carried out on the basis of the first-order autoregression. In this case, it is natural to take as the model the following modification of (11):

$$Y_t = f(Y_{t-1}) + \xi_t, \tag{13}$$

where $t = 2, ..., n$, Y_t is the stock price at the time moment t. We take the parametric function in the form $\varphi(Y_{t-1}; \theta^{(1)}, \theta^{(2)}) = \theta^{(1)} + \theta^{(2)} Y_{t-1}$, where for simplicity we set $\theta^{(1)} = 0$, $\theta^{(2)} = 1$, i.e. $\varphi(Y_{t-1}; 0, 1) = Y_{t-1}$.

As the nonparametric estimate of the interpolation forecast for Y_t, we take the following modification of estimate (2):

$$\hat{Y}_t = \hat{r}(Y_{t-1}) = \frac{\sum\limits_{j\geq 2, j\neq t} Y_t K\left(\dfrac{Y_{t-1} - Y_{j-1}}{h_n}\right)}{\sum\limits_{j\geq 2, j\neq t} K\left(\dfrac{Y_{t-1} - Y_{j-1}}{h_n}\right)}. \tag{14}$$

The combined estimate, for which $\varphi(Y_{t-1}; 0, 1) = Y_{t-1}$, takes the form

$$\bar{Y}_t = \hat{R}_{\hat{\lambda}_B}(Y_{t-1}) = (1 - \hat{\lambda}_B(Y_{t-1}))\hat{r}(Y_{t-1}) + \hat{\lambda}_B(Y_{t-1}) Y_{t-1}, \tag{15}$$

where

$$\hat{\lambda}_B(Y_{t-1}) =$$

$$= \frac{\sum_{j=1}^{B}(\hat{r}(Y_{t-1};Y_{j,1}^*,...,Y_{j,n-1}^*) - \hat{r}(Y_{t-1}))(\hat{r}(Y_{t-1};Y_{j,1}^*,...,Y_{j,n-1}^*) - Y_{t-1})}{\sum_{j=1}^{B}[\hat{r}(Y_{t-1};Y_{j,1}^*,...,Y_{j,n-1}^*) - Y_{t-1},)]^2}. \qquad (16)$$

Based on the prices of stocks $Y_1,...,Y_n$, formulas (14) and (15), the estimates of forecasts by one step for price Y_{n+1} are defined as follows:

$$\hat{Y}_{n+1} = \hat{r}(Y_n;Y_1,...,Y_{n-1}) = \frac{\sum_{i=2}^{n} Y_i K\left(\frac{Y_n - Y_{i-1}}{h_n}\right)}{\sum_{i=2}^{n} K\left(\frac{Y_n - Y_{i-1}}{h_n}\right)}. \qquad (17)$$

$$\bar{Y}_{n+1} = \hat{R}_{\hat{\lambda}_B}(Y_n) = (1 - \hat{\lambda}_B(Y_n))\hat{r}(Y_n) + \hat{\lambda}_B(Y_n)Y_n, \qquad (18)$$

where

$$\hat{\lambda}_B(Y_n) =$$

$$= \frac{\sum_{j=1}^{B}(\hat{r}(Y_n;Y_{j,1}^*,...,Y_{j,n-1}^*) - \hat{r}(Y_n;Y_1,...,Y_{n-1})(\hat{r}(Y_n;Y_{j,1}^*,...,Y_{j,n-1}^*) - Y_n)}{\sum_{j=1}^{B}[\hat{r}(Y_n;Y_{j,1}^*,...,Y_{j,n-1}^*) - Y_{n-1},)]^2}. \qquad (19)$$

Let there be $n+L$ stock prices. Estimates of forecasts $\hat{Y}_{n+2}$ and $\bar{Y}_{n+2}$ will be constructed at n prices $Y_3,...,Y_{n+1}$ by formulas (17) and (18). Similarly, at n prices, shifting by the required number of steps, make forecasts $\hat{Y}_{n+3},...,\hat{Y}_{n+L}$ and $\bar{Y}_{n+3},...,\bar{Y}_{n+L}$.

The quality of identification and forecasting will be characterized by means of the average relative errors $\delta_{real}(\hat{r})$ and $\eta_{real}(\hat{r})$:

$$\delta_{real}(\hat{r}) = \frac{1}{n-1}\sum_{i=2}^{n}\frac{|Y_i - \hat{r}(Y_{i-1})|}{Y_i}100\%, \quad \eta_{real}(\hat{r}) = \frac{1}{L}\sum_{i=n+1}^{n+L}\frac{|Y_i - \hat{r}(Y_{i-1})|}{Y_i}100\%.$$

Consider the case $n = 100$. In Fig. 3 there are presented the results of identification and prediction for the combined model using estimates (15) and (18), and the behavior of the weight coefficients (16) and (19) are shown in Fig. 4.

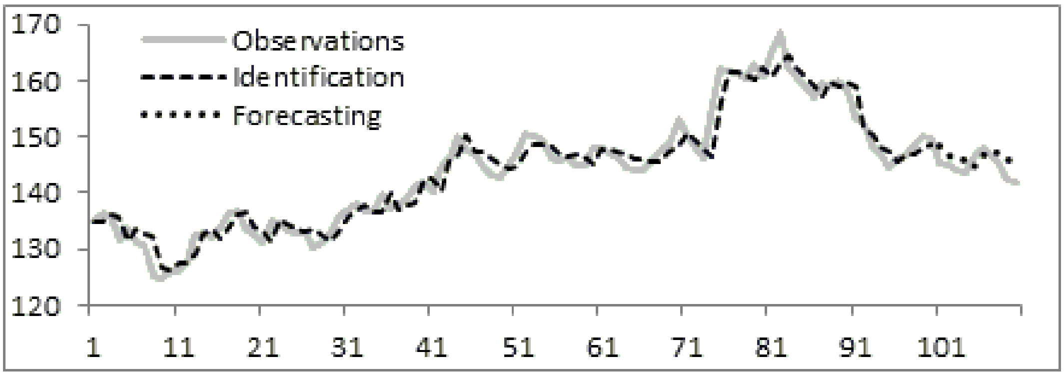

Figure 3: Identification and forecasting using combined estimates (15) and (18)

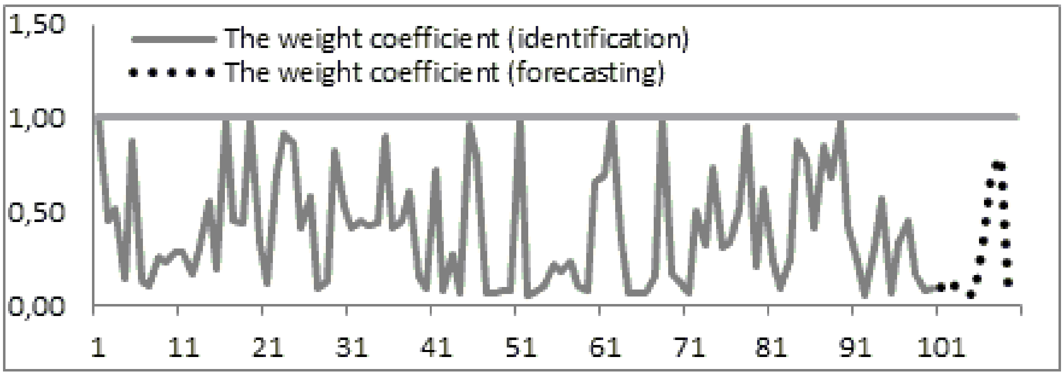

Figure 4: Plot of the dependence of the estimates of the weight coefficients (16) (identification) and (19) (forecasting)

For different volumes of observations, Table 2 gives the average relative errors of identification and prediction.

Table 2: Average relative errors of identification and prediction

n	10	50	100
$\delta_{real}(\hat{r})$	1.87	1.40	1.34
$\delta_{real}(\hat{R})$	1.67	1.29	1.31
$\eta_{real}(\hat{r})$	1.90	2.80	1.16
$\eta_{real}(\hat{R})$	1.41	1.20	1.01

From the results obtained, in practice it is preferable using the combined evaluation in comparison with the nonparametric estimate, especially in the case of small sample sizes.

Conclusions

In this paper, the problem of identification of a stochastic object by means of a combined estimate is considered, which is a weighted sum of the nonparametric estimate of the regression and some function given by the researcher. Adaptive combined estimates are constructed on the basis of which algorithms for predicting static and dynamic objects are proposed.

Based on the results of numerical simulation, the advantage of adaptive combined estimates is shown in comparison with nonparametric regression estimates for small samples sizes and a large noise level. The expediency of applying the proposed approach in practice is illustrated in the analysis of the prices of Gazprom's stocks for 2016.

Acknowledgements

This work was supported by Tomsk State University Competitiveness Improvement Program for 2013-2020.

References

[1] Dmitriev Yu.G., Koshkin G.M. (1987). On the Use of a Priori Information in Nonparametric Regression Estimation. *IFAC Proceedings Series.* Vol. **2**, pp. 223–228.

[2] Dmitriev Yu.G., Koshkin G.M. (1987). Using Additional Information in Nonparametric Estimation of Density Functionals. *Automat. and Remote Control.* Vol. **48**, No. **10**, pp. 1307–1316.

[3] Dmitriev Yu.G., Skripin S.V. (2012). On the Combined Estimate of the Probability of Failure-free Operation for a Full Sample *Journal of Control and Computer Science. Tomsk State University.* No. **4(21)**, pp. 32–38.

[4] Kitaeva A.V., Koshkin G.M. (2009). Recurrent Nonparametric Estimation of Functions from Multidimensional Functional Density and their Derivatives. *Automat. and Remote Control.* Vol. **70**, No. **3**, pp. 389–407.

[5] Kitaeva A.V., Koshkin G.M. (2009). Kernel Estimates of Basic Functionals for Dependent Observations. *Bulletin of the Tomsk Polytechnic University.* Vol. **314**, No. **2**, pp. 26–31.

[6] Koshkin G.M., Lukov V.Yu., Piven I.G. (2016). Nonparametric Algorithms of Identification and Prediction in the ARX-Models. *Proceedings. The Second International Symposium on Stochastic Models, in Reliability Engineering, Life Science, and Operations Management (SMRLO 2016).* Beer Sheva, Israel. Conference Publishing Services The Institute of Electrical and Electronics Engineers, pp. 620–623.

[7] Koshkin G.M., Tarasenko F.P. (1988). Nonparametric Algorithms for Identifying and Control of Continuous-discrete Stochastic Objects. *8-th IFAC-IFORS Symposium on Identification and System Parameter Estimation*. Beijing. Pargamon Press, No. **2**, pp. 882–887.

[8] Nadaraya E. (1964). On Estimating of Regression. *Theory Probab. Appl.* Vol. **9**, No. **1**, pp. 141–142.

[9] Skripin S.V. (2008). Properties of a Combined Regression Estimate for Finite Sample Sizes. *Bulletin of the Tomsk Polytechnic University*. Vol. **313**, No. **5**, pp. 10–14.

[10] Vasiliev V.A., Koshkin G.M. (1998). Nonparametric Identification of Autoregressions. *Theory Probab. Appl.* Vol. **43**, No. **3**, pp. 507–517.

[11] Watson G.S. (1964). Smooth Regression Analysis. *Sankhya. Indian J. Statist.* Vol. **A26**, pp. 359–372.