ISSN 0278-6419, Moscow University Computational Mathematics and Cybernetics, 2016, Vol. 40, No. 2, pp. 71–78. © Allerton Press, Inc., 2016.
Original Russian Text © A.P. Motrenko, K.V. Rudakov, V.V. Strijov, 2016, published in Vestnik Moskovskogo Universiteta, Seriya 15: Vychislitel'naya Matematika i Kibernetika, 2016, No. 2, pp. 20–27.

Combining Endogenous and Exogenous Variables in a Special Case of Non-Parametric Time Series Forecasting Model[#]

A. P. Motrenko[1*], K. V. Rudakov[2], and V. V. Strijov[3***]**

[1]*Department of Control and Applied Mathematics, Moscow Institute of Physics and Technology (State University), Institutskii per. 9, Dolgoprudnyi, Moscow oblast, 141700 Russia*

[2]*Department of Computational Mathematics and Cybernetics, Moscow State University, Moscow, 119899 Russia Dorodnicyn Computing Center, Russian Academy of Sciences, ul. Vavilova 40, GSP-1, Moscow, 119333 Russia*

[3]*Dorodnicyn Computing Center, Russian Academy of Sciences, ul. Vavilova 40, GSP-1, Moscow, 119333 Russia*

Abstract—We address a problem of increasing quality of forecasting time series by taking into account the information about exogenous time series. We aim to improve a non-parametric forecasting algorithm that minimizes the convolution of a histogram of time series with the loss function. We propose to adjust the histogram, using mixtures of conditional histograms as a less sparse alternative to multidimensional histogram and in some cases demonstrate the decrease of loss compared to the basic forecasting algorithm. To the extent of our knowledge, such approach to combining endogenous and exogenous time series is original and has not been proposed yet. The suggested method is illustrated with the data from the Russian Railways.

Keywords: time series, forecasting, exogenous time series, hist algorithm, mixture model.

DOI: 10.3103/S0278641916020072

1. INTRODUCTION

Considering information about exogenous time series is a way to increase time series forecasting quality, since they provide additional information, or the "context" for the forecasted (endogenous) time series [1, 2]. Model structure defines the way to include exogenous time series into the model. For example, if the structure is linear, and inclusion of the exogenous factor in linear case consists in addition of several values (or their transformations) of the exogenous time series to the model. This is the way how ARMA (autoregressive moving average) model [1,3], which is widely used in short-time forecasting [4], is extended to its exogenous version ARMAX (eXogenous autoregressive moving average) [5]. The standard ARMA consists of three additive components: the autoregressive average, moving average and error term. The ARMAX model also includes a combination of exogenous time series.

We aim to improve performance of the *hist* algorithm, proposed in [6]. The forecasting procedure is based on obtaining a histogram of endogenous time series. The forecast, produced by *hist*, is equal to the center of a bin of the histogram, that corresponds to the optimal value of convoluting the histogram with the loss function. To increase the forecasting quality of *hist*, we propose to adjust the histogram using exogenous time series. Most straightforward way is to specify a histogram of endogenous time series, conditional on the values of exogenous time series. This can be done by using specific rows, correspondent to latest realisations of exogenous time series, of a multidimensional histogram, which approximates joint probability density function of the endogenous and exogenous time series [7]. The feasibility of such approach is limited with the length of studied time series. This limitation comes from

[*]E-mail: anastasiya.motrenko@phystech.edu.
[**]E-mail: rudakov@ccas.ru.
[***]E-mail: strijov@ccas.ru.
[#]The research was supported by the Ministry of Education and Science of the Russian Federation, RFMEFI60414X0041.

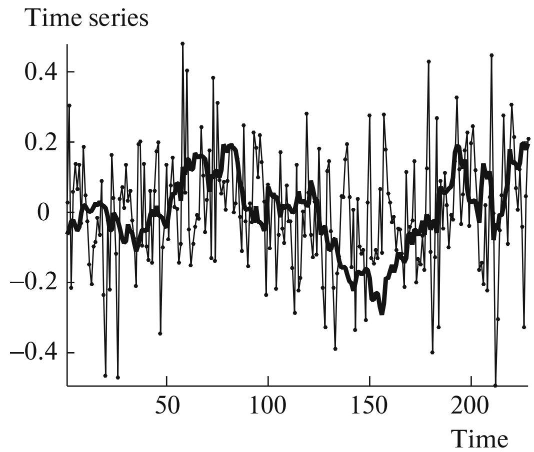

Fig. 1. Examples of endogenous $\mathbf{x}$ and exogenous $\mathbf{c}_j$ time series.

high sparsity of multidimensional histograms which grows rapidly with dimensionality (exponentially on the number of time series). The sparsity can be reduced in some special cases [8, 9]. We suggest to use a weighted sum of histograms, each conditional on one of exogenous time series. Compared to evaluation of one histogram, conditional on all exogenous time series, through a multidimensional joint histogram, the suggested approach is less demanding to the history length, but requires estimation of more parameters. Histogram adjustment based on histogram mixture requires estimation of n two-dimensional histograms instead of one $(n-1)$-diminsional histogram, which reduces the number of values to estimate from exponential to linear on n.

Mixture of histograms is an analogy for mixture models, a rather flexible method for density estimation from the class of Mixture Transition Distribution models [10]. Mixture model indicates the presence of several distributions in the sample, each represented with a component from the same parametric family distribution. According to [10], any probability density function an be approximated with a mixture of gaussian densities with arbitrary precision. For various modiffcations of see, for example [11–13]. The authors of [14–16] used mixtures of histograms as a more stable alternative to Gaussian mixture models in object detection in image and video processing. In mixtures of histograms each component is represented by a histogram instead of a parametric distribution as in classic mixture models. In the papers [14–16] the bins of component histogram are supposed to coincide, so that the weighted summation of histograms only effects the probability associated with each bin, but does not affect the fragmentation of the domain of endogenous variable. The papers [17,18] discuss algebraic operations with histograms, which allow to introduce a weighted sum of histograms with arbitrary division of variable domain into intervals.

2. PROBLEM STATEMENT

Let $\mathbf{x} = \{x(t)\}_{t=1}^{T-1}$ denote the endogenous time series, with $\mathbf{c}_j = \{c_j(t)\}_{t=1}^{T}$ denoting the j-th exogenous time series, $j = 1, \dots, n$. Fig. 1 exemplifies relationships between endogenous and endogenous time series. Endogenous time series $\mathbf{x}$ are depicted with a thin line, each time stamp $x(t)$ marked with a dot. Bold line represents exogenous time series $\mathbf{c}_j$.

2.1. The Hist Algorithm

A histogram $H = \langle \mathbf{X}, \mathbf{h} \rangle$, where $\mathbf{X} = [X_1, \dots, X_k, \dots, X_K]$ is a vector and of bin centers and $\mathbf{h} = [h_1, \dots, h_k, \dots, h_K]$ is a vector of associated probabilities, defines a probability distribution

$$h_k = \mathrm{P}\big(x = X_k\big), \quad \sum_{k=1}^{K} h_k = 1.$$

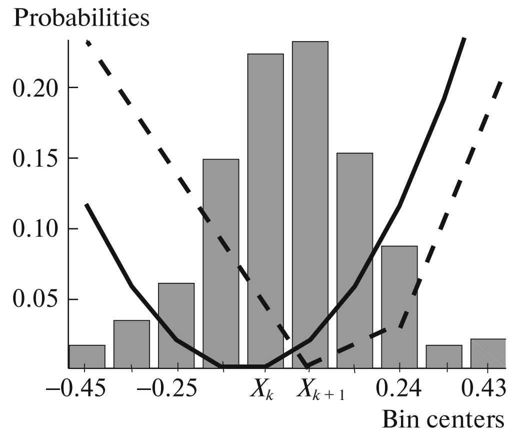

Fig. 2. Illustrations of the *hist* algorithm: convolution of histogram of time series with loss functions at various placements X_k.

We approximate the distribution of $x(t)$, specifying H as $\langle \mathbf{X}, \mathbf{h} \rangle$

$$h_k = \frac{1}{T}\sum_{t=1}^{T}[x \text{ belongs to } k\text{-th bin}],$$

where $[\,\cdot\,]$ is the indicator function, and the bins with centers X_k split the range of $\mathbf{x}$ into K intervals of equal width.

Given a loss function $L(x,\hat{x}) : \mathbb{R} \times \mathbb{R} \mapsto \mathbb{R}_+$ and a histogram estimation H, the forecast $hist(L,H)$ is equal to a bin center X_k that minimises the convolution

$$\sum_{x \in X_1,\dots,X_K} h_k L(X_k, x)$$

of histogram H with predefined loss function L at each bin center X_k, $k = 1,\dots,K$. The bin center X_k that corresponds to the minimum value of convolution is then chosen as a forecast value $\hat{x}$. Fig. 2 illustrates the the convolution procedure. Two kinds of loss function, a squared error (solid line) and a linear asymmetric error (dashed line) are convoluted with histogram H at different bin centers X_k and X_{k+1} respectively. The forecast $\hat{x}$ of the *hist* algorithm is equal to a solution of the following optimization problem:

$$\hat{x} = \text{hist}(H,L) = \underset{x \in X_1,\dots,X_K}{\text{argmin}} \sum_{k=1}^{K} h_k L(x, X_k). \tag{1}$$

2.2. Histogram Adjustment

Our goal is to construct an adjusted histogram H, considering observations of both endogenous and $\mathbf{x}$ and exogenous time series $\mathbf{c}_1,\dots,\mathbf{c}_n$ to minimize the loss function L. Given L, the forecast depends on the properties $\mathbf{X}$ and $\mathbf{h}$ of H. To adjust H, we fix the bin centers X_k (supposing $x(t)$ is stationary) and vary h_k to minimize loss function:

$$L\big(x(T), \hat{x}\big) \to \min_{\mathbf{h} \in [0,1]^K} \quad \text{with } \sum_{k=1}^{K} h_k = 1. \tag{2}$$

We look for a solution of the problem (2) in a form of mixture of conditional histograms. To formulate the expected solution, we introduce the notation $H(t)$ for the adjusted histogram, evaluated at time t. We suppose that

$$H(t) \equiv H(\mathbf{x}^{1:t-1}, \mathbf{c}_1^{1:t}, \dots, \mathbf{c}_n^{1:t}),$$

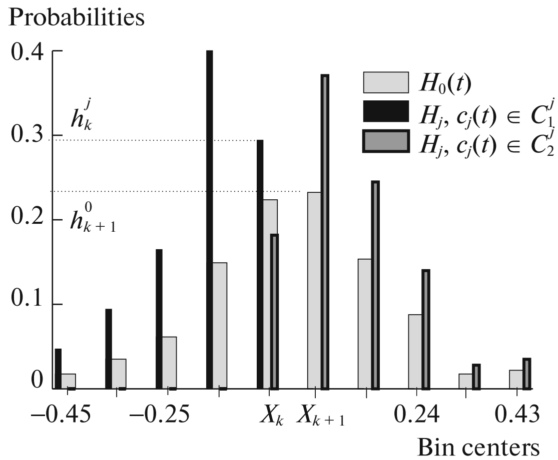

Fig. 3. Conditional and marginal histograms.

This notation means that first $t-1$ values $\mathbf{x}^{1:t-1}$ of endogenous time series $\mathbf{x}$ and t values $\mathbf{c}_n^{1:t}$ of exogenous time series $\mathbf{c}_j$, $1 \le j \le n$ are used to compute probability estimations $\mathbf{h}$. Similarly, $H_j(t) = \langle \mathbf{X}, \mathbf{h}^j \rangle$, $j = 1, \ldots, n$ denotes a histogram, conditional on $\mathbf{c}_j$, that corresponds to the value $c_j(t)$. Finally, let $H_0(t)$ denote marginal histogram of $\mathbf{x}^{1:t-1}$.

Table 1 illustrates relationships between H_0 and H_j. Each cell of Table 1 represents one bin p_{kg} of a multidimensional histogram which approximates the joint probability distribution of $\mathbf{x}$ and $\mathbf{c}_j$

$$p_{kg} \approx \mathrm{P}\big(x(t) \text{ belongstobin } X_k, c_j(t) \text{ belongstobin } C_g^j\big).$$

Multidimensional histogram is connected with H_0 and H_j through the following relations

$$h_k^0 = \sum_{g=1}^{N} p_{kg} \quad \text{and} \quad h_k^j = \frac{p_{kg}}{p_g}, \quad p_g = \sum_{k=1}^{K} p_{kg}.$$

Fig. 3 illustrates the case of $N = 2$. The histograms represented by thinner grey bars are conditional histograms $H_j(t)$ for $\mathbf{x}^{1:t-1}$ with $c_j(t)$ belonging to the first and the second bin. The light grey histogram depicts marginal histogram $H_0(t)$.

In this notation the expected solution of the problem (2) is given by

$$H(T) = w_0 H_0(T) + \sum_{j=1}^{n} w_j H_j(T), \tag{3}$$

where the vector of weights $\mathbf{w} = [w_0, \ldots, w_n]^T$ maximizes likelihood function $p(\mathbf{x}|\mathbf{w}, \mathbf{c}_1, \ldots, \mathbf{c}_n)$ of the model (3). We approximate $p(\mathbf{x}|\mathbf{w}, \mathbf{c}_1, \ldots, \mathbf{c}_n)$ with the probability given by histogram $H(t)$:

$$p\big(x(t)|\mathbf{w}, \mathbf{c}_1^{1:t}, \ldots, \mathbf{c}_n^{1:t}\big) \approx h_{k(t)},$$

where $k(t)$ corresponds to the histogram bin that contains $x(t)$:

$$\frac{1}{2}\big(X_{k(t)} + X_{k(t)-1}\big) \le x(t) < \frac{1}{2}\big(X_{k(t)+1} + X_{k(t)}\big).$$

Assuming $x(t)$ conditioned on $\mathbf{w}$ and $\mathbf{c}_j$ are independent

$$p\big(\mathbf{x}|\mathbf{w}, \mathbf{c}_1, \ldots, \mathbf{c}_n\big) = \prod_{t=t_{\min}}^{T} p\big(x(t)|\mathbf{w}, \mathbf{c}_1^{1:t}, \ldots, \mathbf{c}_n^{1:t}\big) \approx \prod_{t=t_{\min}}^{T} h_{k(t)},$$

where $t_{\min}$ is the minimum length of time series $\mathbf{x}$, required to compute a histogram. Since the number of exogenous time series we use might appear too large, we would like to be able to include only the most

Table 1. Relations between joint and conditional histograms

	C_1^j	$\dots$	C_g^j	$\dots$	C_N^j	$\sum_g$
X_1	p_{11}	$\dots$	$p_{1g} = h_1^j p_g$	$\dots$	p_{1N}	h_1^0
X_2	p_{21}	$\dots$	$p_{2g} = h_2^j p_g$	$\dots$	p_{2N}	h_2^0
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
X_K	p_{K1}	$\dots$	$p_{K2} = h_k^j p_g$	$\dots$	p_{KN}	h_K^0
$\sum_k$	p_1^j	$\dots$	p_g^j	$\dots$	p_N^j	1

informative time series $\mathbf{c}_j$, $j \in \mathcal{J} \subseteq \{0, 1, \dots, n\}$ into the model. Finally, we formulate the optimization problem as:

$$\mathbf{w} = \underset{\mathbf{w} \in [0,1]^{|\mathcal{J}|}}{\text{argmax}} \frac{1}{|\mathcal{J}|} \sum_{t=1}^{T} \log \left(\sum_{j \in \mathcal{J}} w_j h_k^j(t) \right), \quad \sum_{j \in \mathcal{J}} w_j = 1.$$

where the cost function accounts for the number of selected components $|\mathcal{J}|$.

3. COMPONENTS SELECTION AND WEIGHTS ESTIMATION

Suppose the centers $X_1, \dots, X_K$ and $C_1^j, \dots, C_N^j$ of bins of $H_0(t)$ and $H_j(t)$ are fixed. To estimate the weights w_j of mixture components we use a stochastic modification of EM-algorithm [19]. The algorithm iteratively repeats two steps: simulation step, or resampling of the time series according to the current estimation of $H(t)$, and EM-step, where the posteriori probabilities w_{jt} of components for each sample x_t are recalculated. The idea is to evaluate the ratio of samples, generated by each component by simulating the dataset, and to remove the components that generate too few samples.

Suppose we have fixed maximum number of components $n_{\max}$, minimum length $t_{\min}$ of time series $\mathbf{x}$; minimum ratio α of samples, generated by each component, and an initial approximation of w_j and $\mathcal{J} = 0, \dots, n$. The procedure of component selection consists of the following steps:

1. Generate a data sample $\tilde{\mathbf{x}}^{t_{\min}+1:T} = \{\tilde{x}(t)\}_{t=t_{\min}+1}^{T}$ according to the probability distribution, specified by w_{jt}:

$$\hat{x}_t \sim \sum_{j = \in \mathcal{J}} w_{jt} H_j(t).$$

with w_{jt} initially set equal to w_j for all t. Let $\tilde{x}(t)$ belong to $k(t)$-th bin of $H(t)$. On the basis of the data sample $\tilde{\mathbf{x}}$ compute the number T_j of samples, described by j-th component:

$$T_j = \sum_{t=t_{\min}}^{T} \left[\underset{j \in \mathcal{J}}{\text{argmax}}\, h_{k(t)}^j = j \right], \tag{4}$$

where $[\cdot]$ denotes the indicator function. Remove from the model all components that generate $T_j < \alpha(T - t_{\min})$ of samples:

$$\mathcal{J} = \mathcal{J} \setminus \{j : T_j < \alpha(T - t_{\min})\}.$$

For the rest of components recompute T_j according to (4) to estimate corresponding weights w_j as $w_j = T_j/T$.

2. Adjust the posteriori probabilities w_{jt}:

$$w_{jt} = \frac{w_j h_k^j(t)}{\sum_l w_l h_k^l(t)},$$

where $x(t)$ belongs to $k(t)$-th bin.

The steps are repeated until the number $\mathcal{J}$ of components is less than $n_{\max}$.

Following [19] we run the procedure of weight estimation with $\alpha = 0$ (no component selection) R times after termination to obtain more stable estimation:

$$\mathbf{w} = \frac{1}{r}\sum_{i=1}^{R}\mathbf{w}^{(i)},$$

where $\mathbf{w}^{(i)}$ is the estimation, given by i-th run.

4. COMPUTATIONAL EXPERIMENT

The dataset we used to test the algorithm consists of 38 endogenous time series, each corresponding to a type of good (such as coal, cox, petroleum and it's products, etc) transported via national railways. A point $x(t)$ of time series $\mathbf{x}$ corresponds to one day and is equal to the weight in tons of a certain type of goods, transported at the day t. The time series were measured from January, 2007 to May, 2008. The set of exogenous time series $\mathbf{c}_j$ consisted of time series for the prices for Sugar, Petrol, Cuprum, Zink, Aurum, Nikel, Wheat, Masut, Gas, Tin, Crude oil, Argentum and Plumbum observed during the period from January, 2007 to May, 2008. Due to the presence of missing values length of time series after processing equals 228. Since the presence of trend in the time series leads to inadequate results of forecast when the edges of $H_j(t)$ and categorization intervals are constant in time, we removed the trend from all time series from the dataset and additionally normalized the values of the time series to [0, 1].

Selecting histogram parameters N, K. Following [7] we choose to set $K = \lceil 3\sqrt[3]{T}\rceil = 15$. To select optimal number N of categorization intervals for $\mathbf{c}_j$, we have evaluated the probabilities of including each exogenous time series into the model and used MATLAB's implementation of Kruskal-Wallis test [20] to test the hypothesis that the probabilities do not differ for different values of N. We observed p-values around 0.95, which means that the data was not sufficient to reject hypothesis and the results of component selection most likely do not depend on N. Moreover, we found no significant dependance of loss change on N for the studied endogenous time series. Since minimum required length $t_{\min}$ of time series depends linearly on N we chose minimum value of $N = 2$.

Results of component selection. In addition to the original exogenous time series $\mathbf{c}_j$ we considered the derivative $\mathring{\mathbf{c}}_j$, $\mathring{c}(t)_j \in \{0, 1\}$ time series, virtually signalizing whether the time series $\mathbf{c}_j$ experiences period of grow on average on last $t_0 = 10$ time stamps.

$$\mathring{\mathbf{c}}_j(t) = \begin{cases} 1, \text{ if } \bar{\mathbf{c}}_j^{t:t-t_0} > 0, \\ 0, \text{ otherwise}, \end{cases} \quad \text{where} \quad \bar{\mathbf{c}}^{t:t-t_0,j} = \frac{1}{t_0}\sum_{\tau=0}^{t_0-1} c_j(t-\tau).$$

To add lagging to the model (3), the set of time series $\{\mathbf{c}_j, \mathring{\mathbf{c}}_j\}$ was extended with time series

$$\mathbf{c}_j^{1:T-l\tau}, \mathring{\mathbf{c}}_j^{1:T-l\tau}, j = 1, \dots n, \text{ and } \mathbf{x}^{1:T-l\tau-1}$$

for $l = 0, \dots, L-1$, where L is the maximum lag value. We set component selection parameter α to 0.07 the maximum number $n_{\max}$ of components to $n_{\max} = 5$ and the maximum lag L to $L = 3$.

Let $\hat{x}_0$ and $\hat{x}_{Ex}$ denote the forecast (1) of original and the adjusted *hist* algorithm respectively. To measure the quality of adjusted algorithm we computed decrease of loss

$$\Delta L = L_0 - L_{Ex}, \quad L\big(x, \hat{x}\big) = (x - \hat{x})^2$$

We used stochastic EM algorithm to select a set of exogenous time series $\mathbf{c}_j$, $j \in \mathcal{J}$ and evaluate corresponding weights w_j of corresponding mixture components. Then we computed adjusted histograms $H(t)$ at 50 historical points (control points) $x(t)$ using $\mathbf{x}^{1:t-1}$, $\mathbf{c}_j^{1:t}$, $j \in \mathcal{J}$ according to (3). Using $H(t)$ we forecasted $x(t)$ according to (1) and computed $\Delta L(t)$, thus obtaining a sample $\{\Delta L(t)\}$ of loss change values $\Delta L(t)$. We then tested null hypothesis that the expectancy $\mathsf{E}(\Delta L)$ of loss change ΔL is zero under alternative $\mathsf{E}(\Delta L) > 0$ using Student's t-test.

The experiments showed that "Plumbum costs" time series were the most informative in terms of forecasting with adjusted histogram $H(t)$. This suggests that the "Plumbum costs" time series were

Table 2. Results of including external time series from $\{\mathbf{c}_j, \mathbf{c}_j\}$ into model

Type of load	Exoge-nous time series	Time lag τ	ΔL	$\Delta L / L_0$	p-value	Positive change	Negative change
Petrol and petroleum products	$\mathring{\mathbf{c}}$	14	0.15311	0.27873	0.0002305	0.31373	0.039216
Ferrous metals	$\mathbf{c}$	7	0.11195	0.29659	0.00099359	0.17647	0
Metal constructions	$\mathbf{c}$	7	0.066946	0.1044	0.039325	0.21569	0.058824
Metalwork	$\mathbf{c}$	7	0.2046	0.43858	0.00022156	0.64706	0.2549
Chemicals and soda	$\mathring{\mathbf{c}}$	14	0.26036	0.41155	4.9025×10^{-8}	0.47059	0.019608
Construction loads	$\mathbf{c}$	7	0.14043	0.34462	0.00011159	0.35294	0.098039
Granulated slag	$\mathbf{c}$	7	0.062775	0.17258	0.03783	0.15686	0.039216
Refractories	$\mathring{\mathbf{c}}$	14	0.035985	0.086756	0.099958	0.039216	0
Cement	$\mathbf{c}$	7	0.25782	0.31271	0.00010496	0.5098	0.058824
Fish	$\mathring{\mathbf{c}}$	14	0.25241	0.34342	1.0663×10^{-6}	0.43137	0.019608
Grain	$\mathbf{c}$	7	0.19695	0.29581	5.3108×10^{-5}	0.39216	0.13725

more sensitive to the changes of some hidden variable that influenced transportation loads than other exogenous time series from our set. The first column of Table 2 lists those types of load, for which plumbum time series were chosen as most informative. The second column of Table 2 indicates whether the original time series $\mathbf{c}$ or its derivative $\mathring{\mathbf{c}}$, was chosen. The third column lists lagging values τ. Lagging values were considered proportional to $\tau = 7$, correspondent to one week. Other columns contain numerical characteristics of performance of adjusted *hist*: the sample mean loss change ΔL, average relative loss change $\Delta L / L_0$, minimal confidence level required to reject $H_0 : \mathsf{E}(\Delta L) = 0$ in favor of $\mathsf{E}(\Delta L) > 0$ using Student's t-test. Lower values indicate better certainty that adjusted method improves forecasting quality. Only the time series with p-value <0.1 are shown in Table 2. Columns of Table 2 labeled "Positive change" and "Negative change" list the ratio of time points where forecasts were strictly improved $\Delta L(t) > 0$ and the ratio of cases where forecasting quality strictly worsened $\Delta L(t) < 0$ respectively. Note that the ratios of negative and positive $\Delta L(t)$ not only does not sum up to one, but is equally far less in sum than one: for most time series there was no change in forecasting quality in approximately half of control sample.

5. CONCLUSION AND DISCUSSION

We present a method of combining exogenous and endogenous time series to improve the forecasting quality of a special case of non-parametric algorithm *hist*. Our method is based on adjusting the histogram of endogenous time series by modelling it as a mixture of conditional histograms. We compare thus extended algorithm with the basic *hist* and demonstrate the decrease of loss functions in some points. The constant loss for the rest cases is related to insufficient length of forecasted time series, when the we had reside to the basic version of *hist*. We see several ways to extend the presented research. First is to consider more advanced ways of computing initial histogram, such as kernel estimations, which allow to use information about a single sample to evaluate several histogram cells, instead of assigning one sample strictly to one cell. Application of kernel methods may reduce sparsity and decrease the necessary length of time series, allowing to further improve performance of adjusted *hist*. Other ways include allowing time dependance for w_j and considering more types of derivative times series.

Alternative way is to extend the *hist* algorithm is through multivariate quantile regression. Though there is a considerable number of papers [21–23] devoted to the problem of extending the concept of quantile regression to a multivariate case, each extension focuses on maintaining some of the features on univariate quantile regression, and no dominant approach can be singled out. Since the original *hist* is a modification of quantile regression, designing a multivariate quantile regression model would provide a natural way to include exogenous time series into the model.

ACKNOWLEDGMENTS

This work was supported by the Ministry of Education and Science of the Russian Federation, project no. RFMEFI60414X0041.

REFERENCES

1. M. Roopaei, M. Zolghadri, and A. Emadi, *Economical forecasting by exogenous variables,* in *Proc. IEEE 17th Int. Conf. on Fuzzy Systems* (*FuzzIEEE '08*)*, Hong Kong, June 1–6, 2008* (IEEE, New York, 2008), p. 1491–1495.
2. P. Syrovátka and L. Grega, *Analysis of methodological approaches to evaluation of complementary and substitution relationship in consumer demand for food,* Agr. Econ. ***48,*** 456–462 (2002).
3. P. Whittle, *Prediction and Regulation by Linear Least-Square Methods* (Univ. Minnesota Press, Minneapolis, Minn, 1983).
4. J. G. De Gooijer and R. J. Hyndman, *25 years of time series forecasting,* Int. J. Forecasting, No. 22, 443–473 (2006).
5. D. Peña and I. Sánchez, *Multifold predictive validation in armax time series models,* J. Am. Stat. Assoc., No. 100, 135–146 (2005).
6. A. S. Valkov, E. M. Kozhanov, M. M. Medvednikova, et al., *Non-parametric forecasting of railroad stations occupancy according to historical data,* Machine Learning and Data Analysis ***1***, 448–465 (2012).
7. D. W. Scott, *On optimal and data-based histograms,* Biometrika **66**, 605–610 (1979).
8. W. E. Deming and F. F. Stephan, *On a least squares adjustment of a sampled frequency table when the expected marginal totals are known,* Ann. Math. Stat. **11**, 427–444 (1940).
9. Y. M. M. Bishop, S E. Fienberg, and P. W. Holland, *Discrete Multivariate Analysis: Theory and Practice* (MIT Press., Cambridge, 1976).
10. L. Rabiner, *A tutorial on hidden Markov models and selected applications in speech recognition,* PIEEE **77**, 257–286 (1989).
11. C. S. Wong and W. K. Li, *On a mixture autoregressive conditional heteroscedastic model,* J. Am. Stat. Assoc. No. 96, 982–995 (2001).
12. A. Berchtold, *Mixture transition distribution (mtd) modeling of heteroscedastic time series,* Comput. Stat. & Data Analysis, No. 41, 399–411 (2003).
13. F. Bartolucci and A. Farcomeni, *A note on the mixture transition distribution and hidden Markov models,* J. Time Ser. Anal. **31**, 132–138 (2010).
14. E. Aitnouri, S. Wang, D. Ziou, et al., *Estimation of multi-modal histogram's pdf using a mixture model,* Neural, Perallel and Scientific Comput., No. 7, 103–118 (1999).
15. M. Everingham and B. Thomas, *Supervised segmentation and tracking of nonrigid objects using a "mixture of histograms" model,* IEEE Image Proc., **1**, 62–65 (2001).
16. Yu-Ren Lai, Kuo-Liang Chung, Guei-Yin Lin, et al., *Gaussian mixture modeling of histograms for contrast enhancement,* Expert Syst. Appl., No. 39, 6720–6728 (2012).
17. J. M. Schopf, *A practical methodology for defining histograms for predictions and scheduling,* in *Proc. Int. Conf. ParCo99, Delft, Netherlands, Aug 17–20, 1999* (Imperial College Press, 2000), p. 664—671.
18. J. Arroyo, G. Gonzalez-Rivera, C. Maté, et al., *Smoothing methods for histogram-valued time series: an application to value at risk,* Stat. Analysis Data Mining. **4**, 216–228 (2011).
19. J. Diebolt and E. H. S. Ip, *Stochastic EM: method and application,* in *Chain Markov Monte Carlo in Practice* (Chapman and Hall, London, 1996), Chap. 15, p. 259–273.
20. J. D Spurrier, *On the null distribution of the Kruskal–Wallis statistic,* J. Nonparametr. Statist. **15**, 685–691 (2003).
21. V. Koltchinskii, *M-estimation, convexity and quantiles,* Ann. Stat. **25**, 435–477 (1997).
22. R. Serfling, *Quantile functions for multivariate analysis: Approaches and applications,* Stat. Neerl., No. 56, 214–232 (2002).
23. L. Kong and I. Mizera, *Quantile tomography: using quantiles with multivariate data,* Stat. Sinica, No. 22, 1589–1610 (2012).