Вестник БГУ. Сер. 1. 2016. № 3. С. 108–119

УДК 519.2

STATISTICAL FORECASTING OF TIME SERIES: OPTIMALITY AND ROBUSTNESS[1]

Yu. S. KHARIN [a]

[a] *Research Institute for Applied Problems of Mathematics and Informatics, Belarusian State University, Nezavisimosti avenue, 4, 220030, Minsk, Republic of Belarus*

The statistical forecasting (prediction) problem of time series is considered for the typical in practice situation where the underlying hypothetical data model is distorted. We present a review of our solutions for the following topical problems of optimality and robustness in statistical forecasting: mathematical description of distortions for typical hypothetical models of time series; quantitative evaluation of the risk-robustness (sensitivity analysis) under distortions for traditional forecasting statistics (that are optimal under hypothetical models); evaluation of critical distortion levels; construction of new robust forecasting statistics.

Key words: statistical forecasting; prediction; time series; distortion; robustness; forecasting statistic.

СТАТИСТИЧЕСКОЕ ПРОГНОЗИРОВАНИЕ ВРЕМЕННЫХ РЯДОВ: ОПТИМАЛЬНОСТЬ И РОБАСТНОСТЬ

Ю. С. ХАРИН [1)]

[1)] *Учреждение БГУ «Научно-исследовательский институт прикладных проблем математики и информатики», пр. Независимости, 4, 220030, г. Минск, Республика Беларусь*

Рассматривается проблема статистического прогнозирования временных рядов для типичной на практике ситуации, когда гипотетическая модель данных искажена. Представлен обзор полученных автором результатов для следующих актуальных проблем оптимальности и робастности в статистическом прогнозировании: математическое описание искажений типовых гипотетических моделей временных рядов; количественная оценка устойчивости риска прогнозирования при наличии искажений для традиционно применяемых прогнозирующих статистик (оптимальных для гипотетических моделей); оценивание критических уровней искажений; построение новых робастных прогнозирующих статистик.

Ключевые слова: статистическое прогнозирование; временной ряд; искажение; робастность; прогнозирующая статистика.

Many significant applied problems in economics, finance, medicine and other fields lead to the global problem in mathematical statistics: statistical forecasting of random processes with discrete time, that are usually called random sequences or time series [1–6].

Mathematical substance of the Forecasting (Prediction) Problem is very simple: to estimate (evaluate) the future value of the random process $x_{T+\tau} \in \mathbb{R}^d$ in $\tau \geq 1$ steps ahead, based on the before observed data. We can clearly detect two stages in the history of attacking the Forecasting Problem [1–6].

First stage (up to the year 1974): construction of optimal forecasting statistics $\hat{x}_{T+\tau} = f_0(x_1, \ldots, x_T) : \mathbb{R}^{dT} \to \mathbb{R}^d$ for various hypothetical models M_0 w.r.t. to minimization of some forecast risk functional (e. g., mean square error of forecasting). A. N. Kolmogorov was the first who considered the forecasting problem in strict mathematical form [1].

[1] Статья публикуется в авторской редакции.

Образец цитирования:
Харин Ю. С. Статистическое прогнозирование временных рядов: оптимальность и робастность // Вестн. БГУ. Сер. 1, Физика. Математика. Информатика. 2016. № 3. С. 108–119.

For citation:
Kharin Yu. S. Statistical forecasting of time series: optimality and robustness. *Vestnik BGU. Ser. 1, Fiz. Mat. Inform.* 2016. No. 3. P. 108–119 (in Engl.).

Автор:
Юрий Семенович Харин – член-корреспондент НАН Беларуси, доктор физико-математических наук, профессор; директор.

Author:
Yuriy Kharin, corresponding member of the National Academy of Sciences of Belarus, doctor of science (physics and mathematics), full professor; director.
kharin@bsu.by

Second stage (from the year 1974). It was detected and announced by P. Huber at his lecture on the Vancouver International Congress of Mathematicians [2] that «statistical inferences (including forecasts) are based only in part upon the observations; an equally important base is formed by prior assumptions about the underlying situation (that is the hypothetical model M_0)». In practice, the hypothetical models are usually distorted, and the risk of optimal forecasting statistic (that was constructed under M_0) becomes much more than the hypothetical risk. It was proposed by P. Huber to construct *robust* statistical forecasts that «are weakly sensitive w. r. t. small distortions of the hypothetical model M_0».

A list of researchers influencing the field of robust statistical analysis of time series is given here: J. Tukey, P. Huber, F. Hampel, C. Croux, R. Dahlhaus, P. Filzmoser, R. Fried, R. Dutter, V. Gather, M. Genton, Yu. Kharin, R. Maronna, R. D. Martin, S. Morgethaler, C. Mueller, D. Pena, G. Tiao, H. Rieder, E. Ronchetti, P. J. Rousseeuw, R. Tsay, W. Wefelmeyer, V. J. Yohai. The majority of publications on robustness in statistical time series analysis are concentrated on estimation of parameters and hypotheses testing. Although these problems are fundamentals, they do not completely cover the problem of robustness in statistical forecasting of distortions that includes the following topical tasks considered in this paper: mathematical description for typical hypothetical models of time series; quantitative evaluation of the risk-robustness (sensitivity analysis) under distortions for traditional forecasting statistics (that are optimal under hypothetical models); evaluation of critical distortion levels; construction of new robust forecasting statistics.

Distortions of hypothetical models

Introduce the notation: $x_{T+\tau} \in \mathbb{R}^d$ is an observed d-variate time series with discrete time $t \in \mathbb{Z}$, $X = \left(x_1', \ldots, x_T'\right)' \in \mathbb{R}^{dT}$ is the composed vector of observations for T successive time points (prime means transposition), $x_{T+\tau} \in \mathbb{R}^d$ is the non-observable random vector to be predicted at the future time point $T+\tau$, $\tau \in \mathbb{N}$ (in economic applications the value T is called «the base of forecasting», τ – «the horizon of forecasting»). The probability model of the observed time series under distortions is determined by a family of probability measures

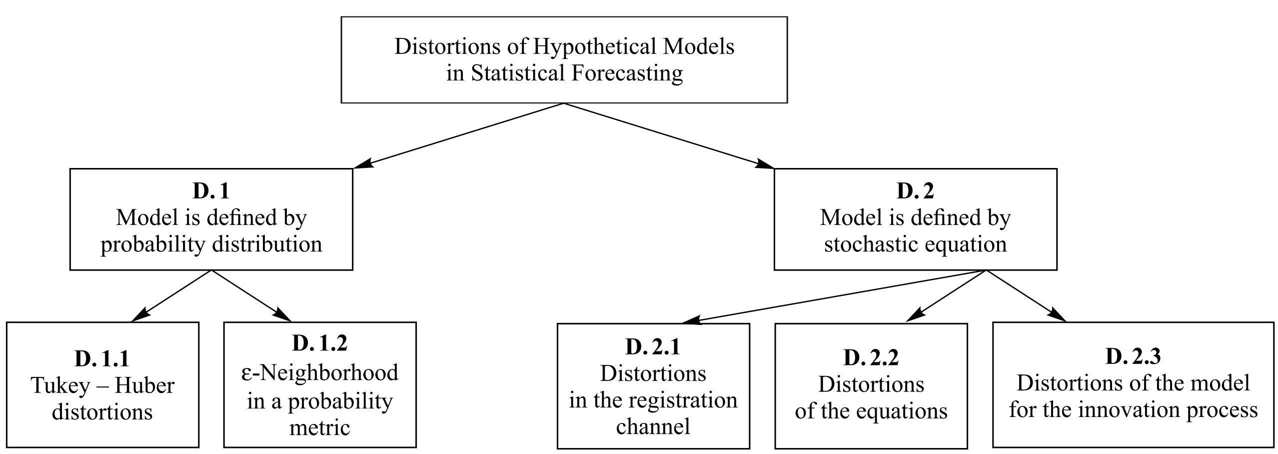
$$\left\{P_{T,\,\theta^0}^{\varepsilon}(A),\ A \in B^{Td} : T \in \mathbb{N},\ \theta^0 \in \Theta \subseteq \mathbb{R}^m,\ \varepsilon \in [0, \varepsilon_+]\right\}, \tag{1}$$

where B^{Td} is the Borel σ-algebra in $\mathbb{R}^{Td}$; θ^0 is an unknown true value of model parameters; ε is the distortion level; $\varepsilon_+ \geq 0$ is its maximal admissible value. If $\varepsilon_+ = 0$, then the distortions are absent, and we have the hypothetical model M_0.

Classification and mathematical description of typical for practice kinds of distortions (1) is given in [3]. In figure we present a short version of the classification scheme from [6].

Let us give short mathematical descriptions for classes of distortions in figure.

The class of distortions D.1 consists of two subclasses. *Tukey – Huber distortions D.1.1* for the observation vector X are described by the mixture: $p(X) = (1-\varepsilon)p^0(X) + \varepsilon \cdot h(X)$, $p^0(\cdot)$ is «non-distorted» (hypothetical) p. d. f., $h(\cdot)$ is so-called «contaminating» p. d. f., $\varepsilon \in [0, 1)$ is the distortion level. If $\varepsilon = 0$, then $p(\cdot) = p^0(\cdot)$, and distortions are absent.



Classification scheme for types of distortions

Distortions of the type D. 1.2 are described by ε-neighborhoods in some probability metric: $0 \le \rho\left(p(\cdot),\ p^0(\cdot)\right) \le \varepsilon,\ \rho(\cdot)$ is a probability metric, e. g. Kolmogorov and Hellinger probability metrics:

$$\rho\left(p,\ p^0\right) = \frac{1}{2}\int_{\mathbb{R}^{Td}} \left|p(X) - p^0(X)\right| dX \in [0, 1];$$

$$\rho\left(p,\ p^0\right) = \frac{1}{2}\int_{\mathbb{R}^{Td}} \left(\sqrt{p(X)} - \sqrt{p^0(X)}\right)^2 dX \in [0, 1].$$

The class of distortions D. 2 describes the hypothetical model by some stochastic equation:

$$x_t = G\left(x_{t-1}, \ldots, x_{t-s}, u_t, u_{t-1}, u_{t-L}; \theta^0\right),\ t \in \mathbb{Z},$$

where $u_t \in \mathbb{R}^\nu$ is innovation process on the probability space $(\Omega, \mathcal{F}, P)$; $s, L \in \mathbb{N}$ are some natural numbers indicating the memory depth; $\theta^0 \in \Theta \subseteq \mathbb{R}^m$ is vector of model parameters; $G(\cdot) : \mathbb{R}^{ds} \times \mathbb{R}^{\nu(L+1)} \times \Theta \to \mathbb{R}^d$ is some Borel function.

The class D. 2 consists of 3 subclasses. *The subclass D. 2.1* describes distortions in the observation channel: $X = H\left(X^0, U\right)$, where $X^0 \in \mathbb{R}^{Td}$ is «non-observable prehistory» of the process, $X \in \mathbb{R}^{Td}$ is observation results, that is the «observable prehistory», $U = \left(u_1', \ldots, u_T'\right) \in \mathbb{R}^{dT}$ is non-observable random vector of distortions (errors in the observation channel), $H(\cdot)$ is a function that describes the registration algorithm.

Subclass D. 2.1 includes following types of distortions: additive, multiplicative, ε-non-homogeneities, «outliers», «missing values», censoring, etc. *Subclass D. 2.2* describes distortions of the generating stochastic equation («misspecification»), includes two types of distortions: *parametric distortions*, when instead of the true parameter value θ^0 we get (estimate by statistical data) a different value $\tilde{\theta}$, with $\left|\tilde{\theta} - \theta^0\right| \le \varepsilon$, where ε is the distortion level; *functional distortions*, when instead of the true function $G(\cdot)$ we get a different function $\tilde{G}(\cdot)$, and in some metric $\left\|\tilde{G}(\cdot) - G(\cdot)\right\| \le \varepsilon$. *Subclass D. 2.3* describes distortions of the innovation process $u_t \in \mathbb{R}^\nu$, $t \in \mathbb{Z}$, in the generating stochastic equation and includes distortions of three types: ε-non-homogeneities, probabilistic dependence, «outliers».

Note, that in practice, the real data can be corrupted by distortions of two or more indicated classes simultaneously, e. g. outliers and missing values.

Functionals of robustness in statistical forecasting

Let $x_{T+\tau} = f_{T,\tau}(X) : \mathbb{R}^{Td} \to \mathbb{R}^d$ be a forecasting statistic (Borel function). Introduce the notation: $\mathbf{E}_\varepsilon\{\cdot\}$ is expectation w. r. t. to the measure $P^\varepsilon_{T,\theta^0}(\cdot)$ from (1); $\pi(\theta) : \Theta \to \mathbb{R}^1$ is some p. d. f. on Θ; $\hat{\theta} \in \mathbb{R}^m$ is some consistent estimator of θ^0 by *X*. To analyze optimality and robustness of forecasting statistics we will use the following functionals.

Point risk of forecasting: $\rho_\varepsilon = \rho_\varepsilon\left(f_{T,\tau}; \theta^0\right) = \mathbf{E}_\varepsilon\left\{\left\|\hat{x}_{T+\tau} - x_{T+\tau}\right\|^2\right\} \ge 0.$

Integral risk of forecasting: $r_\varepsilon = r_\varepsilon\left(f_{T,\tau}\right) = \int_\Theta \rho_\varepsilon\left(f_{T,\tau}; \theta\right)\pi(\theta)d\theta \ge 0.$

Guaranteed (upper) risk: $r_+ = r_+\left(f_{T,\tau}\right) = \sup_{0 \le \varepsilon \le \varepsilon_+} r_\varepsilon\left(f_{T,\tau}\right).$

Optimal forecasting statistic (for the hypothetical model M_0): $\hat{x}^0_{T,\tau} = f^0_{T,\tau}\left(X; \theta^0\right)$ with the hypothetical risk:

$$\rho_0\left(f_{T,\tau}; \theta\right) = \inf_{f_{T,\tau}} \rho_0\left(f_{T,\tau}; \theta^0\right);\ r_0 = \int_\Theta \rho_0\left(f^0_{T,\tau}; \theta\right)\pi(\theta)d\theta. \tag{2}$$

«Plug-in» forecasting statistic:

$$\hat{x}_{T,\tau} = f_{T,\tau}(X) = f^0_{T,\tau}\left(X;\ \hat{\theta}\right). \tag{3}$$

Coefficient of risk instability ($r_0 > 0$) [7]:

$$\kappa = \kappa\left(f_{T,\tau}\right) = \frac{\left(r_+\left(f_{T,\tau}\right) - r_0\right)}{r_0} \ge 0. \tag{4}$$

δ-*Admissible (critical) distortion level ($\delta > 0$) [7]:*

$$\varepsilon^* = \varepsilon^*(\delta) = \sup\left\{\varepsilon : \kappa\left(f_{T,\tau}\right) \le \delta\right\}. \tag{5}$$

Minimax risk-robust forecasting statistic $\hat{x}_{T+\tau} = f^*_{T,\tau}(X)$ *[7]:*

$$\kappa\left(f^*_{T,\tau}\right) = \inf_{f_{T,\tau}(\cdot)} \kappa\left(f_{T,\tau}\right). \tag{6}$$

Optimality and robustness of forecasting for regression time series

The case of additive «outliers» D. 2.1. Hypothetical model (multiple linear regression) M_0 is determined by equation:

$$x_t = \theta^{0\prime}\psi\left(z_t\right) + u_t,\ t \in \mathbb{N},\ E\left\{u_t\right\} = 0,\ D\left\{u_t\right\} = \sigma^2 < +\infty, \tag{7}$$

where $x_t \in \mathbb{R};\ z_t \in Z \subseteq \mathbb{R}^M;\ \psi(z) = \left(\psi_i(z)\right) \in \mathbb{R}^m;\ \left\{\psi_i\right\}$ are linearly independent functions.

LS-forecasting statistic $\left(\left|\Psi_T'\Psi_T\right| \ne 0\right)$:

$$\hat{x}_{T+\tau} = \hat{\theta}\psi\left(z_{T+\tau}\right),\quad \hat{\theta} = \left(\Psi_T'\Psi_T\right)^{-1}\Psi_T' X_T,$$

$$\Psi_T = \left(\psi_j\left(z_t\right)\right) \in \mathbb{R}^{T\times m},\ X_T = \left(x_1, \ldots, x_T\right)' \in \mathbb{R}^T. \tag{8}$$

Distortions of the hypothetical model (7):

$$x_T = \theta^{0\prime}\psi\left(z_t\right) + u_t + \xi_t\nu_t,\ P\left\{\xi_t = 1\right\} = 1 - P\left\{\xi_t = 0\right\} = \varepsilon \le \varepsilon_+ < \frac{1}{2}, \tag{9}$$

where $\left\{\nu_t\right\}$ are independent identically distributed random variables (i. i. d. r. v.), $\mathbf{E}\left\{\nu_t\right\} = a,\ \mathbf{D}\left\{\nu_t\right\} = K\cdot\sigma^2$, $K \ge 0$, and random variables $\left\{u_t\right\},\ \left\{\xi_t\right\},\ \left\{\nu_t\right\}$ are jointly independent.

Introduce the notation: $g(T,\tau) = \Psi_T\left(\Psi_T'\Psi_T\right)^{-1}\psi\left(z_{T+\tau}\right) \in \mathbb{R}^T;\ (z)_+ = \max(z, 0);\ \mathbf{1}_T = (1, \ldots, 1)' \in \mathbb{R}^T.$

Theorem 1 [7, 8]. *Guaranteed upper risk of the LS-forecasting statistic (8) is*

$$r_+(T,\tau) = \left(\sigma^2 + \varepsilon_+\left(a^2 + K\sigma^2\right)\right)\left(1 + \left\|g(T,\tau)\right\|^2\right) + \varepsilon_+^2 a^2\left(\left(1 - \mathbf{1}_T' g(T,\tau)\right)^2 - \left\|g(T,\tau)\right\|^2 - 1\right).$$

Corollary 1. *In the «case of outliers in variance» ($a = 0,\ K > 0$) we have:*

$$\kappa(T,\tau) = \left\|g(T,\tau)\right\|^2 + \varepsilon_+ K\left(1 + \left\|g(T,\tau)\right\|^2\right),$$

$$\varepsilon^*(\delta, T, \tau) = \min\left\{\frac{1}{2}, \left(\delta - \left\|g(T,\tau)\right\|^2\right)_+ K^{-1}\left(1 + \left\|g(T,\tau)\right\|^2\right)^{-1}\right\}.$$

Corollary 2. *In the «case of outliers in mean» ($K = 0,\ a \ne 0$) we get:*

$$\kappa(T,\tau) = \left\|g(T,\tau)\right\|^2 + \varepsilon_+\left(\frac{a}{\sigma}\right)^2\left(1 + \left\|g(T,\tau)\right\|^2\right) + \varepsilon_+^2\left(\frac{a}{\sigma}\right)^2\left(\left(1 - \mathbf{1}_T' g(T,\tau)\right)^2 - \left\|g(T,\tau)\right\|^2 - 1\right).$$

To construct robust forecasting statistic under distorted model with outliers (9) we have developed [3] the so-called the local-median forecasting method using (2)–(6).

Local-median (LM) forecasting method

Introduce the notation: $N_T = \{1, \ldots, T\} \subset \mathbb{N}$, $\gamma^{(l)} = \{t_1^{(l)}, \ldots, t_n^{(l)}\} \subseteq N_T$, is an l-th version of subset of n time points from the set N_T, $l = 1, \ldots, L$, $m \le n \le T$, $m \le L \le C_T^n$; $\Psi_n^{(l)} = \left(\psi_j\left(z_{t_i^{(l)}}\right)\right) \in \mathbb{R}^{n\times n}$, $\det\left(\Psi_n^{(l)\prime}\Psi_n^{(l)}\right) \ne 0$; $X_n^{(l)} = \left(x_{t_1^{(l)}}, \ldots, x_{t_n^{(l)}}\right)' \in \mathbb{R}^n$ is a correspondent l-th version of the subsample of size n from the sample of size T; $g^{(l)} = \left(g_k^{(l)}\right) = \Psi_n^{(l)}\left(\Psi_n^{(l)\prime}\Psi_n^{(l)}\right)^{-1}\psi\left(z_{T+\tau}\right) \in \mathbb{R}^n$, $G^{(n,\,r)(l)} = \sum_{k=r}^{n}\left(g_k^{(l)}\right)^2 \ge 0$, $r = 1, \ldots, n$; $\hat\theta^{(l)} = \left(\Psi_n^{(l)\prime}\Psi_n^{(l)}\right)^{-1}\Psi_n^{(l)\prime}X_n^{(l)}$ $(l = 1, \ldots, L)$ is the l-th version of the LS-estimator (for θ^0) based on $X_n^{(l)}$.

Using $\left\{\hat\theta^{(l)}\right\}$ let us construct local «plug-in» forecasting statistics:

$$\hat x_{T+\tau}^{(l)} = \hat\theta^{(l)\prime}\psi\left(z_{T+\tau}\right),\ l = 1, \ldots, L. \tag{10}$$

LM-forecasting statistic is determined by L local forecasting statistics (10):

$$\hat x_{T+\tau} = \operatorname{med}\left\{\hat x_{T+\tau}^{(1)}, \ldots, \hat x_{T+\tau}^{(L)}\right\}. \tag{11}$$

Introduce the *«Hampel breakdown point»* for the statistic $S(\cdot)$:

$$\varepsilon^{**} = \max\left\{\varepsilon \in [0, 1] : \forall X_{(\varepsilon)} \Rightarrow \left|S\left(X_{(\varepsilon)}\right)\right| \le C < +\infty\right\}. \tag{12}$$

Theorem 2 [3]. *For the model (9) with «outliers», at* $L = C_T^n$ *the «breakdown point» (12) of the LM-forecast (10), (11) is* $\varepsilon^{**} \in \left[0, 1 - nT^{-1}\right]$, *and on the interval* $\left[0, 1 - nT^{-1}\right]$ *it is the unique root of the equation:*

$$\prod_{t=0}^{n-1}\left(1 - \varepsilon - tT^{-1}\right) = (1 - \alpha)\prod_{t=0}^{n-1}\left(1 - tT^{-1}\right),\ \ \alpha = \frac{\lfloor (L-1)/2 \rfloor}{L} = \frac{1}{2} + O\left(\frac{1}{C_T^n}\right).$$

Corollary 1. *If* $n \le \alpha T$, *then* $\varepsilon^{**} \ge T^{-1}$.

Corollary 2. *For the fixed n, if* $T \to \infty$, *then* $\varepsilon^{**} \to 1 - 2^{-1/n}$. *Optimal size of subsamples turns out to be equal to the number of unknown regression coefficients:* $n^* = \arg\max_n \varepsilon^{**} = m$.

Theorem 3 [3]. *If the model (9) with «outliers in variance» takes place, LM-forecasting statistic (10), (11) is used,* $L\{u_t\} = N_1\left(0, \sigma^2\right)$, *local forecasts (10) are independent,* $\left(n^{-1}\Psi_n^{(l)\prime}\Psi_n^{(l)}\right)^{-1} = Q_n$, $G^{(n,\,r)(l)} = G^{(n,\,r)}$, $l = 1, \ldots, L$, $r = 1, 2, \ldots, |Q_n| \ne 0$, *then at* $\varepsilon \in [0, \varepsilon_+]$, $K \in [0, K_+]$, $T \to \infty$, $L = L(T) \to \infty$ *the following asymptotic expansion holds:*

$$r_+(T, \tau) = \sigma^2 + \varepsilon_+K_+\sigma^2 + \frac{\pi\sigma^2}{2L}\left(\sum_{r=0}^{n}\frac{\left(1-\varepsilon_+\right)^r\varepsilon_+^{n-r}C_n^r}{\sqrt{G^{(n,1)} + K_+G^{(n,\,r+1)}}}\right) + o\left(L^{-1}\right). \tag{13}$$

Case of functional distortions (FD) D.2.2. In addition to (13) consider now multiple regression time series under FD determined by the equation:

$$x_t = \sum_{i=1}^{m}\theta_i^0\psi_i\left(z_t\right) + \lambda\left(z_t\right) + u_t,\ t \in \mathbb{N},\ x_t \in \mathbb{R},\ z_t \in Z \subseteq \mathbb{R}^M, \tag{14}$$

where $\left\{\psi_i(\cdot) : Z \to \mathbb{R}\right\}$; $\theta^0 = \left(\theta_i^0\right) \in \mathbb{R}^m$; $\lambda(\cdot) : Z \to \mathbb{R}$ is an unknown FD function.

Types of functional distortions for the model (14)

FD-1 (interval distortions): $\varepsilon_-(z) \le \lambda(z) \le \varepsilon_+(z),\ z \in Z,\ \varepsilon_\pm(z) : \mathbb{R}^M \to \mathbb{R};$ the special case $\varepsilon_\pm(z) = \pm\varepsilon : -\varepsilon \le \lambda(z) \le +\varepsilon,\ \varepsilon \ge 0.$

FD-2 (relative distortions): $\dfrac{|\lambda(z)|}{|\theta^{0\prime}\psi(z)|} \le \varepsilon,\ z \in Z,\ \varepsilon \ge 0.$

FD-3 (distortions in l_p-metric): $\left(\sum_{t=1}^{T} |\lambda(z_t)|^p + |\lambda(z_{T+\tau})|^p\right)^{1/p} \le \varepsilon,\ p \ge 1,\ \tau \in \mathbb{N},\ \varepsilon \ge 0.$

FD-4 (distortions described by orthogonal expansions): $\lambda(z) = \sum_{j \in J} S_j \eta_j(z),\ \left(\sum_{j \in J} |S_j|^p\right)^{1/p} \le \varepsilon,$ where $z \in Z;$ $p \ge 1;\ J \subset \mathbb{N};\ J = \{j_1, \ldots, j_q\};\ q \in \mathbb{N};\ \varepsilon \ge 0;\ \{S_j : j \in J\};\ \{\eta_j(\cdot) : Z \to \mathbb{R}\}.$

Introduce the notation: $g(T, \tau) = \Psi_T (\Psi_T' \Psi_T)^{-1} \psi(z_{T+\tau}) \in \mathbb{R}^T,\ (z)_+ = \max(z, 0);\ \|x\| = \sqrt{x'x},\ x \in \mathbb{R},$ $r_0(T, \tau) = \sigma^2 \left(1 + \|g(T, \tau)\|^2\right)$ is the risk under the hypothetical model.

Theorem 4 [9]. *In the case of FD-1 the guaranteed upper risk*

$$r_+(T, \tau) = r_0(T, \tau) + \max\left(\sum_{t=1}^{T} \left((g_T(T, \tau))_+ \varepsilon_\pm(z_t) - (-g_T(T, \tau))_+ \varepsilon_\pm(z_t)\right) - \varepsilon_\pm(z_{T+\tau})\right)^2.$$

Theorem 5 [9]. *In the case of FD-2 the guaranteed upper risk*

$$r_+(T, \tau) = r_0(T, \tau) + \varepsilon^2 \left(\left|\psi'(z_{T+\tau})\theta^0\right| + \sum_{t=1}^{T} \left|g_t(T, \tau)\theta^{0\prime}\psi(z_t)\right|\right)^2.$$

Theorem 6 [9]. *In the case of FD-3 the guaranteed upper risk*

$$r_+(T, \tau) = r_0(T, \tau) + \varepsilon^2 \max\{1,\ g_t^2(T, \tau) : t = 1, \ldots, T\},\ p = 1,$$

$$r_+(T, \tau) = r_0(T, \tau) + \varepsilon^2 \left(\sum_{t=1}^{T} |g_t(T, \tau)|^{p/p-1} + 1\right)^{2(p-1)/p},\ p > 1.$$

Theorem 7 [9]. *In the case of FD-4 the guaranteed upper risk*

$$r_+(T, \tau) = r_0(T, \tau) + \varepsilon^2 \max_{j \in J} \left(\sum_{t=1}^{T} g_t(T, \tau)\eta_j(z_t) - \eta_j(z_{T+\tau})\right)^2,\ p = 1,$$

$$r_+(T, \tau) = r_0\left(T, \tau + \varepsilon^2 \left(\sum_{j \in J} \left|\sum_{t=1}^{T} g_t(T, \tau)\eta_j(z_t) - \eta_j(z_{T+\tau})\right|^{p/p-1}\right)^{2(p-1)/p}\right),\ p > 1.$$

Robust forecasting statistic under FD-1 distortion is based on the so-called *M*-estimator $\hat{\theta}$ with special choice of the loss function $\rho(\cdot)$ [9, 10]:

$$\hat{x}(T, \tau) = \hat{\theta}'\psi(z_{T+\tau}),\quad \hat{\theta} = \arg\min_{\theta} \sum_{t=1}^{T} \rho(x_t - \theta'\psi(z_t)), \tag{15}$$

where $\delta_\varepsilon \ge 0;\ \rho(z) = \frac{1}{2} \cdot \mathbf{I}(|z| - \delta_\varepsilon)(z - \delta_\varepsilon \operatorname{sign}(z))^2;\ I(\cdot)$ is indicator function.

Smoothed version of the loss function $\left(\delta_\varepsilon \geq 0,\ d > 0,\ a \geq 0,\ b \geq 0\right)$ in (15):

$$\rho(z) = \begin{cases} az^2,\ 0 \leq z \leq \delta_\varepsilon - d, \\ \dfrac{b-a}{3d} z^3 + \dfrac{bd + \delta_\varepsilon(a-b)}{d} z^2 + \dfrac{(b-a)(\delta_\varepsilon - d)^2}{d} z + \dfrac{(a-b)(\delta_\varepsilon - d)^3}{3d},\ \delta_\varepsilon - d < z < \delta_\varepsilon, \\ b\left(z - \dfrac{2(b-a)\delta_\varepsilon + d(a-b)}{2b}\right)^2 + \dfrac{12\delta_\varepsilon(\delta_\varepsilon - d)a(b-a) + d^2(b+3a)(b-a)}{12b},\ z \geq \delta_\varepsilon, \\ \rho(-z),\ z < 0. \end{cases} \tag{16}$$

Let us denote for the smoothed loss function (16) ($t = 1, \ldots, T$):

$$\mu(z) = \frac{d\rho(z)}{dz},\ \nu(z) = \frac{d^2\rho(z)}{dz^2},\ \mu(\theta) = \left(\mu\left(x_t - \theta'\psi(z_t)\right)\right) \in \mathbb{R}^T,$$

$$\nu(\theta) = \operatorname{diag}\left(\nu\left(x_t - \theta'\psi(z_t)\right)\right),\ M(\theta) = \Psi_T'\mu(\theta),\ D(\theta) = \Psi_T'\nu(\theta)\Psi_T.$$

Theorem 8 [9]. *If* $\operatorname{rank}\left(\nu\left(\theta_{(n)}\right)\right) = k,\ m \leq k \leq T,$ *then the approximation* $\theta_{(n+1)}$ *of* $\hat\theta$ *at the* $(n+1)$*-th step is* $\theta_{(n+1)} = \theta_{(n)} + \left(D\left(\theta_{(n)}\right)\right)^{-1} M\left(\theta_{(n)}\right).$

Optimality criterion for the parameter δ_ε in (16) is determined by (4): $\kappa \to \min\limits_{\delta_\varepsilon}$.

Optimality and robustness of forecasting under autoregressive models

Case of non-homogeneity in mean. Consider the case when the underlying hypothetical model M_0 is $\mathrm{AR}(m)$:

$$x_t = \theta^{0\prime} X_{t-1} + u_t,\ X_{t-1} = \left(x_{t-1}, \ldots, x_{t-m}\right)',\ t \in \mathbb{N}, \tag{17}$$

where $X_0 = 0_m$; $\{u_t\}$ are i. i. d. r. v.; $L\{u_t\} = N_1\left(0, \sigma^2\right)$; $\theta^0 \in \mathbb{R}^m$ is unknown.

Optimal (in the mean square) hypothetical forecasting statistic is

$$\hat{x}^0_{T+j} = \theta^{0\prime}\hat{X}_{T+j-1},\ j = 1, \ldots, \tau,$$

$$\hat{X}_{T+j-1} = \left(\hat{x}^0_{T+j-1}, \ldots, \hat{x}^0_{T+j-m}\right)';\ \hat{x}^0_s = x_s \text{ for } s \leq T;\ \rho_0 = \sigma^2. \tag{18}$$

Forecasting statistic (under «misspecification error» $\theta - \theta^0$) is

$$\hat{x}_{T+j} = \hat\theta'\hat{X}_{T+j-1},\ j = 1, \ldots, \tau. \tag{19}$$

Consider the case when the model (17) is under D. 2.3 distortions [11–16].

Distortions D. 2.3 of the innovation process u_t (non-homogeneity in mean) are determined by an unknown distortion function $\lambda(\cdot): \mathrm{R}^1 \to \mathrm{R}^1$:

$$x_t = \theta^{0\prime} X_{t-1} + u_t + \lambda(t),\ t \in \mathbb{N}. \tag{20}$$

Introduce the matrix notation:

$$B_0 = \begin{pmatrix} & \theta^{0\prime} & \\ \cdots & \cdots & \cdots \\ \mathbf{I}_{m-1} & \vdots & O_{m-1} \end{pmatrix},\ B = \begin{pmatrix} & \theta' & \\ \cdots & \cdots & \cdots \\ \mathbf{I}_{m-1} & \vdots & O_{m-1} \end{pmatrix},$$

where $\mathbf{I}_n$ is the $(n \times n)$ identity matrix; O_n is the zero n-column.

Theorem 9 [11]. *Under the «misspecification error» (D. 2.2) $\theta - \theta^0$ in (20), the point risk for the traditional forecasting statistic (19) is* $\rho_0(\cdot) = \sigma^2\left(1+\sum_{j=1}^{\tau-1}\left(\left(B_0^j\right)_{11}\right)^2 + \sum_{t=1}^{T}\left(\left(\left(B^\tau - B_0^\tau\right)B_0^{t-1}\right)_{11}\right)^2\right).$

Theorem 10 [11, 12]. *Under the «misspecification error» (D. 2.2) and functional distortions (D. 2.3):* $T^{-1}\sum_{t=1}^{T}\lambda^2(t) \le \varepsilon_{(1)}^2,\ \tau^{-1}\sum_{t=T+1}^{T+\tau}\lambda^2(t) \le \varepsilon_{(2)}^2,$ *the guaranteed point risk of the forecasting statistic (19) is*

$$\rho_+\left(\theta^0, \theta, T, \tau\right) = \rho_0\left(\theta^0, \theta, T, \tau\right) + \left(\varepsilon_{(1)}\sqrt{T\sum_{i=0}^{T-1}\left(\left(\left(B^\tau - B_0^\tau\right)B_0^i\right)_{11}\right)^2} + \varepsilon_{(2)}\sqrt{\tau\sum_{i=0}^{T-1}\left(\left(B_0^i\right)_{11}\right)^2}\right)^2.$$

Theorem 11 [3, 12]. *If the distortions of the innovation process (D. 2.3) take place:*

$$T^{-1}\sum_{t=1}^{T}\lambda^2(t) \le \varepsilon_+^2,\ \lambda(t) = 0,\ t > T, \tag{21}$$

and the traditional LS-forecasting statistic is used:

$$\hat{x}_{T+j} = \hat{\theta}' X_{T+j-1},\ j = 1, \ldots, \tau;\ \ \hat{\theta} = \left(\sum_{t=1}^{T_0}\tilde{X}_{t-1}\tilde{X}'_{t-1}\right)^{-1}\sum_{t=1}^{T_0}\tilde{x}_t\tilde{X}_{t-1}, \tag{22}$$

where $\tilde{X} = \left(\tilde{x}_1, \ldots, \tilde{x}_{T_0}\right)'$ *does not depend on X, and the mean square error matrix for* $\hat{\theta}$ *is* $\Sigma = \mathbf{E}\left\{\left(\hat{\theta} - \theta^0\right) \times \left(\hat{\theta} - \theta^0\right)'\right\} = \left(T_0\right)^{-1}F,$ *then the guaranteed point risk is (for* $\tau = 1$*) equals to*

$$\rho_+\left(\theta^0; T, T_0\right) = \sigma^2 + \frac{\sigma^2}{T_0}\sum_{t=1}^{T}\left(B_0^{t-1}\right)'_{\cdot 1} F\left(B_0^{t-1}\right)'_{\cdot 1} + \varepsilon_+^2\frac{T}{T_0}\mu_{\max}\left(G_T\right),$$

where $\mu_{\max}\left(G_T\right)$ *is the maximal eigenvalue of the* $(T \times T)$*-matrix* $G_T = \left(\left(B_0^{i-1}\right)'_{\cdot 1} F\left(B_0^{j-1}\right)'_{\cdot 1}\right),$ *i, j = 1, …, T, and the dot intstead of the matrix index means summation on all its values.*

Denote: $e_t = \lambda(t)/\varepsilon_{(1)},\ t = 1, \ldots, T;\ L_t = \left(e_t, 0, \ldots, 0\right)' \in \mathbb{R},\ V_t = \sum_{t=0}^{t-1}B_0^i L_{t-i},\ d = (1, 0, \ldots, 0)' \in \mathbb{R}^m,\ A(\tau) = \sum_{i=0}^{\tau-1}B_0^i dg' F^{-1} \times B_0^{\tau-1-i},\ F = \sum_{i=0}^{\infty}B_0^i dd'\left(B_0^i\right)'.$

Theorem 12 [3]. *If the distortions (21) of the innovation process take place, the traditional LS-forecasting statistic (22) is used,* $\tilde{X} = X,$ *and the following limits exist:*

$$\lim_{T\to\infty}\frac{1}{T}\sum_{t=0}^{T-1}V_t V_t' = H,\quad \lim_{T\to\infty}\frac{1}{T}\sum_{t=0}^{T-1}e_{t+1}V_t = g,$$

then the asymptotic expansion of the guaranteed point risk holds:

$$\rho_+\left(\theta^0; T\right) = \sigma^2\left(1 + \sum_{i=1}^{\tau-1}\left(\left(B_0^i\right)_{11}\right)^2 + \left(\frac{\varepsilon_+}{\sigma}\right)^4\left(A(\tau)\right)_{1\cdot} F\left(A(\tau)\right)'_{1\cdot}\right) + o\left(\varepsilon^4 + T^{-1}\right).$$

Case of bilinear distortions D. 2.2. Let the observed time series $x_t \in \mathbb{R}^1$ satisfies the autoregressive equation of the order *m* with bilinear distortions [17]:

$$x_t = \left(\theta^0\right)' X_{t-1} + \varepsilon x_{t-1}u_{t-1} + u,\ t \in \mathbb{Z}, \tag{23}$$

where ε is the bilinearity coefficient. If ε = 0, then the model (23) coincides with $\mathrm{AR}(m)$ determined by (17).

The model (23) is called the bilinear model $BL(m, 0, 1, 1)$.

Introduce the matrix notation: $W = I_{m+1} - W_1 - W_2$,

$$W_1 = \begin{pmatrix} 0 & \theta_1^0 & \theta_2^0 & \dots & \theta_m^0 \\ 0 & \theta_2^0 & \theta_3^0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & \theta_m^0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}; \; W_2 = \begin{pmatrix} 0 & 0 & \dots & 0 & 0 \\ \theta_1^0 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ \theta_{m-1}^0 & \theta_{m-2}^0 & \dots & 0 & 0 \\ \theta_m^0 & \theta_{m-1}^0 & \dots & \theta_1^0 & 0 \end{pmatrix}; \; S = \begin{pmatrix} 1 & -\theta_1^0 & \dots & -\theta_{m-1}^0 \\ 0 & 1 & \dots & -\theta_{m-2} \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix},$$

where $\bar{S}(i, j)$ is the (i, j)-th element of the inverse matrix $\bar{S} = S^{-1}$.

Theorem 13 [17]. *If the bilinear model (23) is stationary,* $|W| \neq 0$, $|\varepsilon| \leq \varepsilon_+$, *and* $\varepsilon_+ \to 0$, *then the risk instability coefficient (4) of the traditional autoregressive forecasting statistic (18) satisfies the asymptotic expansion:*

$$\kappa(\tau) = \varepsilon_+^2 \sigma^2 \left(1 + W^{-1}(1, 1) + \frac{\left(\sum_{j=1}^{\tau} \bar{S}(1, j) \right)^2 + 4\left(1 - \sum_{j=1}^{m} \theta_j^0 \right)^{-1} \sum_{j=1}^{\tau-1} \bar{S}(1, j) \bar{S}(1, j+1)}{\sum_{j=1}^{\tau} \left(\bar{S}(1, j) \right)^2} \right) + o(\varepsilon_+^2).$$

To construct the robust forecasting statistic we use the following stochastic expansion (in the mean square sense) at $\varepsilon_+ \to 0$, $K \in \mathbb{N}$.

$$x_{T+1} = u_{T+1} + \theta^{0\prime} X_T + \sum_{k=1}^{K} (-1)^{k-1} \varepsilon^k \left(\prod_{i=0}^{k-1} x_{T-i} \right) \left(x_{T-k+1} - \theta^{0\prime} X_{T-r} \right) + O\left(\varepsilon_+^{K+1} \right).$$

The robust forecasting statistic for x_{T+1} is determined by the conditional mathematical expectation $\mathbf{E}\left\{ x_{T+1} \middle| x_T, x_{T-1}, \dots, x_1 \right\}$. This fact generates a family of asymptotically optimal $(K \to +\infty)$ forecasting statistics [3]:

$$\hat{x}_{T+1}^{(K)} = \theta^{0\prime} X_T + \sum_{k=1}^{K} (-1)^{k-1} \varepsilon^k \left(\prod_{i=0}^{k-1} x_{T-i} \right) \left(x_{T-k+1} - \theta^{0\prime} X_{T-k} \right), \; \kappa\left\{ \hat{x}_{T+1}^{(K)} \right\} \underset{T, K \to \infty}{\to} 0.$$

Forecasting statistics $\hat{x}_{T+\tau}^{(K)}$ for $\tau > 1$ are constructed iteratively w. r. t. τ.

Robust forecasting of AR under simultaneously influencing «outliers» and missing values. Let us consider the hypothetical $\mathrm{AR}(p)$-model (17) in the following form:

$$y_t + b_1 y_{t-1} + \dots + b_p y_{t-p} = u_t, \; t \in \mathbb{Z}, \tag{24}$$

where $b = (b_i) \in R^p$ is a vector of unknown autoregressive coefficients satisfying the stationarity condition; $\{u_t\}$ are i. i. d. r. v.; $u_t \sim N_1(0, \sigma^2)$; σ^2 is unknown variance. It is distorted by «outliers» (D. 2.1):

RO – replacement outliers:

$$z_t = (1 - \eta_t) y_t + \eta_t \nu_t, \tag{25}$$

AO – additive outliers:

$$z_t = y_t + \eta_t \nu_t, \tag{26}$$

where $\{\eta_t\}$ are Bernoulli i. i. d. r. v.; $\mathbf{P}\{\eta_t = 1\} = 1 - P\{\eta_t = 0\} = \varepsilon \in [0, \varepsilon_+]$; $\{\nu_t \in R\}$ are i. i. d. r. v. with some symmetric probability distribution.

One more simultaneously influencing type of distortions in (25), (26) is «missing values» [11, 12]:

$$x_k = z_{t_k},\ \ k = 1, 2, \ldots, K, \tag{27}$$

where $\{t_k\}$ is some known increasing sequence of observation time points: $o_{t_k} = 1;\ o_t = 0,\ t \neq t_k$.

Let us denote: $\sigma_\tau = \operatorname{cov}\{y_t, y_{t+\tau}\};\ \theta_\tau = \operatorname{corr}\{y_t, y_{t+\tau}\} = \dfrac{\sigma_\tau}{\sigma_0},\ \tau \in \mathbb{Z};\ \ \psi(\cdot) : \mathbb{R}^1 \to \mathbb{R}^1$ is a bounded odd function,

$$f_\psi(\theta_\tau) = \frac{\sqrt{1-\theta_\tau^2}}{\pi}\int_{-\infty}^{+\infty} \frac{\psi(z)}{1-\theta_\tau^2 + (z-\theta_\tau)^2}\,dz,\quad s_t = \psi\left(\frac{z_t}{z_{t+\tau}}\right),\quad \sigma_\psi(\nu) = \operatorname{cov}\{s_t, s_{t+\nu}\},$$

$$N_T = \sum_{t=1}^{T-\tau} o_t o_{t+\tau},\quad S_T = \sum_{t=1}^{T-\tau} \frac{s_t o_t o_{t+\tau}}{N_T},\quad \Sigma_\tau^\psi(\lambda) = \sum_{t\in Z} \sigma_\psi(t)\cos(\lambda t) \geq 0,\quad \lambda \in [-\pi, \pi]. \tag{28}$$

Theorem 14 [18]. *Under missing values (27) and RO-outliers (25), if* $\psi(\cdot)$ *is such that the function* $f_\psi(\cdot)$ *in (28) has the continuous inverse function* $f_\psi^{-1}(\cdot)$*, then under* $T \to \infty,\ N_T \to \infty$ *the statistic*

$$\hat{\theta}_\tau = f_\psi^{-1}\left(\frac{S_T}{(1-\varepsilon)^2}\right),\ \ 0 < \tau < T, \tag{29}$$

is a consistent estimator for the correlation coefficient $\theta_\tau : \hat{\theta}_\tau \xrightarrow{\mathbf{P}} \theta_\tau$.

Some examples of ψ-estimators determined by (29) are given in the table.

Examples of ψ-estimators

Notation	$\psi(x)$	$f_\psi(\theta)$	$f_\psi^{-1}(u)$
ψ*s*-Estimator	$\operatorname{sign}(x)$	$2/\pi \arcsin\theta$	$\sin(\pi u/2)$
ψ*a*-Estimator	$\arctan x$	$0{,}5 \arcsin\theta$	$\sin 2u$
ψ*t*-Estimator	$2x/(1+x^2)$	$\theta/\left(1+\sqrt{1-\theta^2}\right)$	$2u/(1+u^2)$

For example, ψ*s*-estimator for θ_τ has the form:

$$\hat{\theta}_\tau = \sin\left(\frac{0{,}5\pi}{(1-\varepsilon)^2}\sum_{t=1}^{T-\tau}\operatorname{sign}\left(\frac{z_t}{z_{t+\tau}}\right)\frac{o_t o_{t+\tau}}{N_T}\right).$$

Theorem 15 [18]. *Under theorem 14 conditions, if* $\min\limits_{\lambda\in[-\pi,\pi]} \Sigma_\tau^\psi(\lambda) > 0$*, then the ψ-estimator (29) is asymptotically normal:* $\dfrac{\left(\hat{\theta}_\tau - \theta_\tau\right)}{\sqrt{\mathbf{D}\{\hat{\theta}_\tau\}}} \underset{T\to\infty}{\xrightarrow{\mathbf{D}}} N_1(0, 1)$*, and the asymptotic mean square error (MSE) satisfies the asymptotic expression:* $E\left\{\left(\hat{\theta}_\tau - \theta_\tau\right)^2\right\} \underset{T\to\infty}{\sim} \dfrac{(1-\varepsilon)^{-4}\Sigma_T^\psi(\lambda_T^*)}{N_T\left(f_\psi'(\theta_\tau)\right)^2},\ \ \lambda_T^* \in [-\pi, \pi]$.

Corollary. *The optimal function* $\psi_*(\cdot)$ *in (29) for the estimation of* θ_τ *that minimizes the asymptotic MSE, has the form:* $\psi_*(x; \theta_\tau) = \dfrac{2x}{1+x^2}\left(\dfrac{2\left(1+x^2\right)^2\sqrt{1-\theta_\tau^2}}{1+x^4+2x^2\left(1-2\theta_\tau^2\right)} - 1\right)$.

Robust estimator $\hat{b} = (\hat{b}_i) \in R^p$ for autoregression coefficients in (24) is the solution of the system of $p + 1$ Yule – Walker equations:

$$\begin{cases} \hat{\theta}_0 + b_1\hat{\theta}_1 + \ldots + b_p\hat{\theta}_p = \dfrac{\sigma^2}{\sigma_0}, \\ \hat{\theta}_{-1} + b_1\hat{\theta}_0 + \ldots + b_p\hat{\theta}_{p-1} = 0, \\ \ldots\ \ldots\ \ldots \\ \hat{\theta}_{-p} + b_1\hat{\theta}_{1-p} + \ldots + b_p\hat{\theta}_0 = 0,\ \ \hat{\theta}_{-k} = \theta_k,\ \hat{\theta}_0 \equiv 1. \end{cases}$$

The algorithm Durbin – Levinson is suitable for solution of this system.

Theorem 16 [18]. *Under the conditions of theorem 15, the following convergence holds:*

$$\mathbf{P}\left\{\left\|\hat{b} - b\right\| \le Q \cdot \left\|\hat{\theta} - \theta\right\|\right\} \underset{T\to\infty}{\to} 1,$$

where $Q = \left(1 + \sqrt{\dfrac{2p}{\pi}\displaystyle\int_{-\pi}^{\pi}\left|B\left(e^{ix}\right)\right|^2 dx}\right) / \lambda_{\min}(\Theta)$*;* $B(\lambda) = \lambda^p + b_1\lambda^{p-1} + \ldots + b_p$ *is the generating characteristic polinom for the equation (24);* $\lambda_{\min}(\Theta)$ *is the minimal characteristic number of the matrix* $\Theta = \left(\theta_{i-j}\right)_{i,\ j=\overline{1,\ p}}$.

Statistical estimator for ε is constructed:

$$\hat{\varepsilon} = 1 - \sqrt{e}\,\hat{f}_z\left(\hat{\sigma}_2^{-1}\right),\ \ \hat{\sigma}_2 = \sqrt{\frac{2}{\lambda_1^2 - \lambda_2^2}\ln\frac{\hat{f}_z(\lambda_2)}{\hat{f}_z(\lambda_1)}},\ \ \hat{f}_z(\lambda) = \sum_{t=1}^{T} o_t \cos\left(\lambda_{z_t}\right) / \sum_{t=1}^{T} o_t$$

is the sample characteristic function,

$$\lambda_2 = 2\lambda_1,\ \ \lambda_1 = \sqrt{\sum_{t=1}^{T} o_t / \sum_{t=1}^{T} o_t z_t^2}.$$

Conclusion

The results presented in this paper provide a Statistician (a Forecaster) with quantitative estimates of the guaranteed upper risk, with the risk instability coefficient and with the δ-admissible (critical) distortion level for the «plug-in» statistical forecasting of time series under some typical distortions of the underlying hypothetical models, i. e. trend, regression and autoregressive models. These estimates reveal the influence of distortions on the risk, indicate the risk limits of the safe forecasting by traditional algorithms under distortions, and outline approaches to the robustification of forecasting procedures.

Some new minimax risk-robust forecasting statistics are constructed and compared to the traditional forecasting algorithms.

The theoretical results have been tested on simulated data, on real statistical data and applied in the development of the software package ROSTATFOR (Robust Statistical Forecasting) in the Belarusian State University.

The results are applied in statistical forecasting of macroeconomic time series, spatio-temporal processes [20] and in information protection [21].

REFERENCES

1. *Колмогоров А. Н.* Интерполирование и экстраполирование стационарных случайных последовательностей // Изв. Акад. наук СССР. Сер. матем. 1941. Т. 5, № 1. С. 3–14 [Kolmogorov A. N. Interpolation and extrapolation of stationary stochastic series. *Izv. Akad. Nauk SSSR. Ser. Mat.* 1941. Vol. 5, No. 1. P. 3–14 (in Russ.)].
2. *Huber P.* Some mathematical problems arising in robust statistics // Proceedings of the International Congress of Mathematicians. Vancouver, 1974. P. 821–824.
3. *Kharin Yu.* Robustness in statistical forecasting. Heidelberg ; N. Y. ; Dordrecht ; London, 2013.
4. *Kharin Yu.* Robustness in statistical pattern recognition. Dordrecht, 1996.

5. *Kharin Yu.* Optimality and robustness in statistical forecasting // International Encyclopedia of Statistical Sciences. N. Y., 2011. P. 1034–1037.

6. *Kharin Yu.* Robustness in statistical forecasting (chapter 14) // Robustness and Complex Data Structures. N. Y., 2013. P. 225–242.

7. *Kharin Yu.* Robustness analysis in forecasting of time series // Developments in Robust Statistics. N. Y., 2000. P. 180–193.

8. *Kharin Yu.* Robust forecasting of parametric trends // Studies in Classification and Data Analysis. 2000. Vol. 17. P. 197–206 [Kharin Yu. Robust forecasting of parametric trends. *Stud. Classif. Data Anal.* 2000. Vol. 17. P. 197–206 (in Engl.)].

9. *Kharin Yu., Maevskiy V.* Robust regressive forecasting under functional distortions in a model // Autom. Remote Control. 2002. Vol. 63, № 11. P. 1803–1820 [Kharin Yu., Maevskiy V. Robust regressive forecasting under functional distortions in a model. *Autom. Remote Control.* 2002. Vol. 63, No. 11. P. 1803–1820 (in Engl.)].

10. *Kharin Yu.* Robustness of mean square risk in forecasting of regression time series // Comm. Stat. Theory Methods. 2011. Vol. 40, № 16. P. 2893–2906 [Kharin Yu. Robustness of mean square risk in forecasting of regression time series. *Comm. Stat. Theory Methods.* 2011. Vol. 40, No. 16. P. 2893–2906 (in Engl.)].

11. *Kharin Yu., Zenevich D.* Robustness statistical forecasting by autoregressive model under distortions // Theory Stoch. Processes. 1999. Vol. 5. P. 84–91 [Kharin Yu., Zenevich D. Robustness statistical forecasting by autoregressive model under distortions. *Theory Stoch. Processes.* 1999. Vol. 5. P. 84–91 (in Engl.)].

12. *Kharin Yu., Zenevich D.* Robust forecasting of autoregressive time series for additive distortions // J. Math. Sci. 2005. Vol. 127, № 2. P. 163–174 [Kharin Yu., Zenevich D. Robust forecasting of autoregressive time series for additive distortions. *J. Math. Sci.* 2005. Vol. 127, No. 2. P. 163–174 (in Engl.)].

13. *Kharin Yu., Huryn A.* «Plug-in» statistical forecasting of vector autoregressive time series with missing values // Austrian J. Stat. 2005. Vol. 34, № 2. P. 163–174 [Kharin Yu., Huryn A. «Plug-in» statistical forecasting of vector autoregressive time series with missing values. *Austrian J. Stat.* 2005. Vol. 34, No. 2. P. 163–174 (in Engl.)].

14. *Kharin Yu., Huryn A.* Sensitivity analysis of the risk of forecasting for autoregressive time series with missing values // Pliska Studia Math. Bulgarica. 2005. Vol. 17. P. 137–146 [Kharin Yu., Huryn A. Sensitivity analysis of the risk of forecasting for autoregressive time series with missing values. *Pliska Studia Math. Bulgarica.* 2005. Vol. 17. P. 137–146 (in Engl.)].

15. *Kharin Yu. S., Pashkevich M. A.* Robust estimation and forecasting for beta-mixed hierachical modes of grouped binary data // Stat. Oper. Res. Trans. 2004. Vol. 28, № 2. P. 125–160 [Kharin Yu. S., Pashkevich M. A. Robust estimation and forecasting for beta-mixed hierachical modes of grouped binary data. *Stat. Oper. Res. Trans.* 2004. Vol. 28, No. 2. P. 125–160 (in Engl.)].

16. *Kharin Yu., Kostevich A.* Discriminant analysis of stationary Markov chains // Math. Meth. Stat. 2004. № 1. P. 235–252 [Kharin Yu., Kostevich A. Discriminant analysis of stationary Markov chains. *Math. Meth. Stat.* 2004. No. 1. P. 235–252 (in Engl.)].

17. *Kharin Yu., Radzievskaya O.* Robustness of forecasting for autoregressive time series with bilinear distortions // Austrian J. Stat. 2008. Vol. 37, № 1. P. 61–71 [Kharin Yu., Radzievskaya O. Robustness of forecasting for autoregressive time series with bilinear distortions. *Austrian J. Stat.* 2008. Vol. 37, No. 1. P. 61–71 (in Engl.)].

18. *Kharin Yu., Voloshko V.* Robust estimation of AR coefficients under simultaneously influencing outliers and missing values // J. Stat. Plan. Inference. 2011. Vol. 141, № 9. P. 3276–3288 [Kharin Yu., Voloshko V. Robust estimation of AR coefficients under simultaneously influencing outliers and missing values. *J. Stat. Plan. Inference.* 2011. Vol. 141, No. 9. P. 3276–3288 (in Engl.)].

19. *Kharin Yu., Voloshko V.* On asymptotic properties of the plug-in cepstrum estimator for Gaussian time series // Math. Meth. Stat. 2012. Vol. 21, № 1. P. 43–60 [Kharin Yu., Voloshko V. On asymptotic properties of the plug-in cepstrum estimator for Gaussian time series. *Math. Meth. Stat.* 2012. Vol. 21, No. 1. P. 43–60 (in Engl.)].

20. *Kharin Yu., Zhurak M.* Statistical analysis of spatio-temporal data based on Poisson conditional autoregressive model // Informatica. 2015. Vol. 26, № 1. P. 67–87 [Kharin Yu., Zhurak M. Statistical analysis of spatio-temporal data based on Poisson conditional autoregressive model. *Informatica.* 2015. Vol. 26, No. 1. P. 67–87 (in Engl.)].

21. *Kharin Yu., Vecherko E.* Statistical estimation of parameters for binary Markov chain models with embeddings // Discrete Math. Appl. 2013. Vol. 23, № 2. P. 153–169 [Kharin Yu., Vecherko E. Statistical estimation of parameters for binary Markov chain models with embeddings. *Discrete Math. Appl.* 2013. Vol. 23, No. 2. P. 153–169 (in Engl.)].

Received by editorial board 05.05.2016.