On robust forecasting of autoregressive time series under censoring

Kharin Yu. [1], Badziahin I. [2]

1) Belarusian State University, kharin@bsu.by

2) Belarusian State University, bodiagin@cosmostv.by

Abstract: *Problems of robust statistical forecasting are considered for autoregressive time series observed under distortions generated by interval censoring. Three types of robust forecasting statistics are developed; mean-square risk is evaluated for the developed forecasting statistics. Numerical results are given.*

***Keywords*:** Autoregressive model, interval censoring, robust forecasting statistic, robustness.

1. INTRODUCTION

Autoregressive model AR(p) of order p developed by Box and Jenkins [1] is widely used to describe real time series with dependent observations in many fields, such as economy, finance, meteorology, astronomy, medicine [1, 2]. The case of "full data", where all observations are exactly known, is well studied [1–3]. Also the case, where some observations are missing and we haven't any information about them [3, 4], is studied. Nowadays the "middle" situation, where some observations are unknown but we have some additional information on their location, is becoming attractive and interesting direction of study [5, 6].

We study the situation where exact values of some observations are unknown, but they belong to certain intervals. This situation is usually called interval censoring [5–7]. Interval censoring can appear because of detection limits of measuring devices, high costs of measurement, disorders of equipment [5–7]. Censoring happens in physical science, engineering, business, economics and meteorology [5]. In this paper we solve the problems of forecasting of autoregressive time series under interval censoring. In practice the censored observations are usually replaced by some naive estimators, e.g. by the center of the interval or by its upper or lower bound. The risk of forecasting, based on such imputations, is usually high, so it is necessary to construct a robust forecasting statistic to improve the quality of prediction for time series containing censored observations.

2. MATHEMATICAL MODEL

Consider the AR(p) time series model [1]

$$x_t = \sum_{i=1}^{p} \theta_i x_{t-i} + u_t,\ t \in \mathbf{Z}, \qquad (1)$$

where $\mathbf{Z}$ is the set of integers, $\theta_1, \ldots, \theta_p$ are the coefficients of the autoregression, and all roots of the characteristic polynomial $z^p - \sum_{j=1}^{p} \theta_j z^{p-j} = 0$ are inside the unit circle, $\{u_t\}$ are i.i.d. normal random variables: $E\{u_t\} = 0$, $D\{u_t\} = \sigma^2 < +\infty$, $u_t \sim N(0, \sigma^2)$. So we have a strictly stationary process [1].

Instead of the observations $x_1, \ldots, x_T$ we observe only random events:

$$A_i^* = \{x_i \in A_i\},\ i \in \{1, \ldots, T\}, \qquad (2)$$

where $\{A_i\}$ are some known Borel sets, $T > p$ is the length of the observation process. If $A_i = \{x_i\}$ is a singleton, then the value of the *i*-th observation is known. If $A_i = (a_i, b_i)$ is an interval, then we have the case of interval censoring. In this paper we consider only these two cases.

A forecasting statistic for the future value $\hat{x}_{T+1}$ is a numerical function of the observed events:

$$\hat{x}_{T+1} = f\left(A_T^*, A_{T-1}^*, \ldots, A_1^*\right). \qquad (3)$$

Conditional risk of forecasting

$$r(T) = E\left\{\left(\hat{x}_{T+1} - x_{T+1}\right)^2 \middle| A_T^*, A_{T-1}^*, \ldots, A_1^*\right\} \qquad (4)$$

is the conditional mean-square error of forecasting under events $\{A_i^*\}_{i=1}^{T}$.

The main problem considered in this paper is to build the robust forecasting statistic (3) that minimizes the conditional risk (4) under censoring.

3. GENERAL RESULTS FOR THE AR(p) MODEL

Let the last $q\,(T - p \geq q \geq 1)$ observations are censored and other observations are exactly known:

$$A_T = (a_T, b_T), \ldots, A_{T-q+1} = (a_{T-q+1}, b_{T-q+1}),$$
$$A_{T-q} = \{x_{T-q}\}, \ldots,\ A_1 = \{x_1\}. \qquad (5)$$

Consider the problem of construction of the robust forecasting statistic $\hat{x}_{T+1}$ for the situation, where the parameters p, $\{\theta_i\}$ and σ of the model (1) are known.

Introduce the notations:

$$\mu(T, p) = \theta_1 x_{T-1} + \ldots + \theta_p x_{T-p} = \sum_{i=1}^{p} \theta_i x_{T-i},$$

$$p_{T,q}(x_T, \ldots, x_{T-q+1}) = p(x_T, \ldots, x_{T-q+1} \mid x_{T-q}, \ldots, x_1).$$

Theorem 1. *Let values* $x_1, \ldots, x_{T-q}$ *and events* $A_{T-q+1}^* = \{x_{T-q+1} \in (a_{T-q+1}, b_{T-q+1})\}, \ldots, A_T^* = \{x_T \in (a_T, b_T)\}$ *be observed. Then the robust forecasting statistic is*

$$\hat{x}_{T+1} = E\{x_{T+1} \mid A_T^*, \ldots, A_{T-q+1}^*, x_{T-q}, \ldots, x_1\} = \qquad (6)$$

$$\frac{\int\limits_{a_T}^{b_T}\dots\int\limits_{a_{T-q+1}}^{b_{T-q+1}}\mu(T+1,p)\,p_{T,q}(x_T,\dots,x_{T-q+1})dx_{T-q+1}\dots dx_T}{\int\limits_{a_T}^{b_T}\dots\int\limits_{a_{T-q+1}}^{b_{T-q+1}}p_{T,q}(x_T,\dots,x_{T-q+1})dx_{T-q+1}\dots dx_T}$$

and its risk is the conditional variance:

$$r_0(T)=D\left\{x_{T+1}\middle|A_T^*,\dots,A_{T-q+1}^*,x_{T-q},\dots,x_1\right\}.$$

Corollary 1. *If* $a_T\to b_T$, ..., $a_{T-q+1}\to b_{T-q+1}$, *then the robust forecasting statistic and its risk are*

$$\hat{x}_{T+1}=\mu(T+1,p)=\sum_{i=1}^{p}\theta_i x_{T-i+1}\ ,\ r_0(T)=\sigma^2\,.$$

In Corollary 1 we consider the asymptotic case of "full data": $x_T=a_T=b_T$, …, $x_{T-q+1}=a_{T-q+1}=b_{T-q+1}$. The results indicated in Corollary 1 coincide with the well known results for this case in [1–3].

Introduce the notations:

$$\sigma_{ij}=Cov\{x_{T-i+1},x_{T-j+1}\},\ \Sigma=\begin{array}{cc} & \begin{array}{cc} q & p\end{array}\\ \begin{array}{c} q\\ p\end{array} & \begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{12}' & \Sigma_{22}\end{pmatrix}\end{array}=\left(\sigma_{ij}\right),$$

$i,j=1,2,\dots,p+q$.

Lemma 1. *Let values* $x_1,\dots,x_{T-q}$ *and events* $A_{T-q+1}^*=\{x_{T-q+1}\in(a_{T-q+1},b_{T-q+1})\}$, ..., $A_T^*=\{x_T\in(a_T,b_T)\}$ *be observed. Then the conditional probability density is*

$p(x_T,\dots,x_{T-q+1}\mid x_{T-q},\dots,x_{T-q-p+1})=n(x_T,\dots,x_{T-q+1}\mid\bar{\mu},\bar{\Sigma})$,

where $n(x_T,\dots,x_{T-q+1}\mid\bar{\mu},\bar{\Sigma})$ *is the* q*-variate normal probability density function with parameters* $\bar{\mu}=\Sigma_{12}\Sigma_{22}^{-1}\left(x_{T-q},\dots x_{T-q-p+1}\right)'$, $\bar{\Sigma}=\Sigma_{11}-\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}'$.

Theorem 2. *Let values* $x_1,\dots,x_{T-1}$ *and the event* $A_T^*=\{x_T\in(a_T,b_T)\}$ *be observed (* $q=1$ *). Then the robust forecasting statistic is*

$$\hat{x}_{T+1}=\theta_1\mu(T,p)+\sum_{i=2}^{p}\theta_i x_{T-i+1}+\theta_1\sigma\frac{\varphi\left(\frac{a-\mu(T,p)}{\sigma}\right)-\varphi\left(\frac{b-\mu(T,p)}{\sigma}\right)}{\Phi\left(\frac{b-\mu(T,p)}{\sigma}\right)-\Phi\left(\frac{a-\mu(T,p)}{\sigma}\right)},$$

where $\Phi(x)$ *is the standard normal probability distribution function and* $\varphi(x)$ *is the standard normal probability density function.*

4. THE CASE OF THE AR(1) MODEL

Consider now the AR(1) model

$$x_t=\theta\,x_{t-1}+u_t,\ t\in\mathbf{Z}, \qquad (7)$$

for the situation (5). As for the AR(p) model, we assume that the parameters θ and σ of the model (7) are known.

Introduce the notations:

$$I_1(l,m)=\int_{a_T}^{b_T}\dots\int_{a_{T-m+1}}^{b_{T-m+1}}x_T^l\,p(x_T,\dots,x_{T-m+1}\mid x_{T-m})dx_{T-m+1}\dots dx_T,$$

$$I_2(l,m)=\int_{a_T}^{b_T}\dots\int_{a_{T-m+1}}^{b_{T-m+1}}x_T^l\,p(x_T,\dots,x_{T-m+1})dx_{T-m+1}\dots dx_T\ .$$

Theorem 3. *Let values* $x_1,\dots,x_{T-q}$ *and events* $A_{T-q+1}^*=\{x_{T-q+1}\in(a_{T-q+1},b_{T-q+1})\}$, ..., $A_T^*=\{x_T\in(a_T,b_T)\}$ *be observed. Then the robust forecasting statistic is*

$$\hat{x}_{T+1}=\theta E\{x_T\mid A_T^*,\dots,A_{T-q+1}^*,x_{T-q}\}=\theta\frac{I_1(1,q)}{I_1(0,q)}, \qquad (8)$$

and its risk is

$$r_0(T)=\sigma^2+\theta^2\times\left(\frac{I_1(2,q)}{I_1(0,q)}-\left(\frac{I_1(1,q)}{I_1(0,q)}\right)^2\right).$$

Analyze the properties of the forecasting statistic (8).

Corollary 1. *If* $a_T\to-\infty$, ..., $a_{T-q+1}\to-\infty$, $b_T\to+\infty$, ..., $b_{T-q+1}\to+\infty$, *then the robust forecasting statistic and its risk are*

$$\hat{x}_{T+1}=\theta^{q+1}x_{T-q}\ ,\ r_0(T)=\sigma^2\sum_{i=0}^{q}\theta^{2i}\ .$$

Conditions of Corollary 1 mean that at the time moments *T, ..., T-q+1* the observations x_T, …, x_{T-q+1} "become" missing. The results indicated in Corollary 1 coincide with the known results for this case in [3, 4].

Corollary 2. *If* $a_T\to b_T$, ..., $a_{T-q+1}\to b_{T-q+1}$, *then the robust forecasting statistic and its risk are*

$$\hat{x}_{T+1}=\theta\ x_T\ ,\ r_0(T)=\sigma^2\,.$$

Theorem 4. *Let values* $x_1,\dots,x_{T-q}$ *and events* $A_{T-q+1}^*=\{x_{T-q+1}\in(a_{T-q+1},b_{T-q+1})\}$, ..., $A_T^*=\{x_T\in(a_T,b_T)\}$ *be observed and there exists* $k\in\{0,1,\dots,q-1\}$ *that the length of the* k*-th censoring interval tends to zero:* $\tau_{T-k}=b_{T-k}-a_{T-k}\to 0$. *Then the robust forecasting statistic is*

$$\hat{x}_{T+1}=\theta E\{x_T\mid A_T^*,\dots,A_{T-q+1}^*,x_{T-q}\}=$$
$$\theta E\{x_T\mid A_T^*,\dots,A_{T-k+1}^*,x_{T-k}\}.$$

It follows from this theorem that for prediction in the case of the AR(1) model we need to know only the last observed value x_{T-q} and all censoring intervals after it.

Consider two special cases: $q=1$ and $q=T$.

Theorem 5. *Let events* $A_1^*=\{x_1\in(a_1,b_1)\}$, ..., $A_T^*=\{x_T\in(a_T,b_T)\}$ *be observed (* $q=T$ *). Then the robust forecasting statistic and its risk are*

$$\hat{x}_{T+1}=\theta E\{x_T\mid A_T^*,\dots,A_1^*\}=\theta\frac{I_2(1,T)}{I_2(0,T)},$$

$$r_0(T)=\sigma^2+\theta^2\times\left(\frac{I_2(2,T)}{I_2(0,T)}-\left(\frac{I_2(1,T)}{I_2(0,T)}\right)^2\right).$$

The case $q=1$ means that the last observation is censored. To simplify our results let us write a and b

instead of a_T and b_T.

Theorem 6. *Let the value* x_{T-1} *and the event* $A_T^* = \{x_T \in (a,b)\}$ *be observed. Then the robust forecasting statistic is*

$$\hat{x}_{T+1} = \theta E\left\{x_T \left| A_T^*, x_{T-1}\right.\right\} \qquad (9)$$

and can be computed in the following way:

$$\hat{x}_{T+1} = \theta^2 x_{T-1} + \theta\sigma \frac{\varphi\left(\dfrac{a-\theta\, x_{T-1}}{\sigma}\right) - \varphi\left(\dfrac{b-\theta\, x_{T-1}}{\sigma}\right)}{\Phi\left(\dfrac{b-\theta\, x_{T-1}}{\sigma}\right) - \Phi\left(\dfrac{a-\theta\, x_{T-1}}{\sigma}\right)},$$

its risk is

$$r_0(T) = (1+\theta^2)\sigma^2 -$$

$$\theta^2\sigma^2 \left(\frac{\varphi\left(\dfrac{a-\theta\, x_{T-1}}{\sigma}\right) - \varphi\left(\dfrac{b-\theta\, x_{T-1}}{\sigma}\right)}{\Phi\left(\dfrac{b-\theta\, x_{T-1}}{\sigma}\right) - \Phi\left(\dfrac{a-\theta\, x_{T-1}}{\sigma}\right)} \right)^2 + \theta^2\sigma \times$$

$$\left(\frac{(a-\theta\, x_{T-1})\varphi\left(\dfrac{a-\theta\, x_{T-1}}{\sigma}\right) - (b-\theta\, x_{T-1})\varphi\left(\dfrac{b-\theta\, x_{T-1}}{\sigma}\right)}{\Phi\left(\dfrac{b-\theta\, x_{T-1}}{\sigma}\right) - \Phi\left(\dfrac{a-\theta\, x_{T-1}}{\sigma}\right)} \right).$$

Corollary 1. *If* $a \to -\infty$ *and* $b \to +\infty$*, then the robust forecasting statistic and its risk are* $\hat{x}_{T+1} = \theta\, x_T$, $r_0(T) = \sigma^2$.

Corollary 2. *If* $a \to b$*, then the robust forecasting statistic and its risk are* $\hat{x}_{T+1} = \theta^2 x_{T-1}$, $r_0(T) = \sigma^2\left(1+\theta^2\right)$.

Theorem 7. *Let the assumptions of Theorem 6 take place and* $\tau = b - a \to 0$*. Then the asymptotic expansion for the risk is*

$$r_0(T) = \sigma^2 + \frac{\theta^2}{4}\tau^2 + o\left(\tau^2\right). \qquad (10)$$

It is known [1] that for the case of "full data" the risk of the optimal forecasting statistic is $r_0 = \sigma^2$. To evaluate the sensitivity of the risk of forecasting to the length $\tau = b - a$ of the censoring interval (a,b) we will use the risk sensitivity coefficient [3]:

$$\chi = \frac{r - r_0}{r}. \qquad (11)$$

Corollary 1. *For the forecasting statistic (9) the risk sensitivity coefficient (11) has the following approximation:*

$$\chi_0 \approx \frac{\theta^2}{4\sigma^2}\tau^2.$$

According to [3] define the ε-admissible $(\varepsilon > 0)$ length of the censoring interval as the maximal length $\tau(\varepsilon)$, such that $\chi \le \varepsilon$ [3].

Corollary 2. *For the forecasting statistic (9)* ε*-admissible length of the censoring interval has the following approximation:*

$$\tau_0(\varepsilon) \approx \sqrt{\varepsilon}\,\frac{2\sigma}{\theta}.$$

Let us compare now the constructed forecasting statistic (9) and widely used in practice forecasting statistics.

The first forecasting statistic is

$$\hat{x}_{T+1} = \theta E\left\{x_T \mid A_T^*\right\} = \theta E\left\{x_T \mid x_T \in (a,b)\right\}. \qquad (12)$$

Theorem 8. *Let values* $x_1, \dots, x_{T-1}$ *and the event* $A_T^* = \{x_T \in (a,b)\}$ *be observed. Then the forecasting statistic (12) can be calculated in the following way:*

$$\hat{x}_{T+1} = \frac{\theta\sigma}{\sqrt{(1-\theta^2)}} \frac{\varphi\left(\dfrac{a\sqrt{1-\theta^2}}{\sigma}\right) - \varphi\left(\dfrac{b\sqrt{1-\theta^2}}{\sigma}\right)}{\Phi\left(\dfrac{b\sqrt{1-\theta^2}}{\sigma}\right) - \Phi\left(\dfrac{a\sqrt{1-\theta^2}}{\sigma}\right)};$$

its risk is

$$r_1(T) = \frac{\sigma^2}{1-\theta^2} +$$

$$\frac{\theta^2\sigma}{\sqrt{(1-\theta^2)}} \frac{a\varphi\left(\dfrac{a\sqrt{1-\theta^2}}{\sigma}\right) - b\varphi\left(\dfrac{b\sqrt{1-\theta^2}}{\sigma}\right)}{\Phi\left(\dfrac{b\sqrt{1-\theta^2}}{\sigma}\right) - \Phi\left(\dfrac{a\sqrt{1-\theta^2}}{\sigma}\right)} -$$

$$\frac{\theta^2\sigma^2}{(1-\theta^2)} \left(\frac{\varphi\left(\dfrac{a\sqrt{1-\theta^2}}{\sigma}\right) - \varphi\left(\dfrac{b\sqrt{1-\theta^2}}{\sigma}\right)}{\Phi\left(\dfrac{b\sqrt{1-\theta^2}}{\sigma}\right) - \Phi\left(\dfrac{a\sqrt{1-\theta^2}}{\sigma}\right)} \right)^2.$$

If $\tau = b - a \to 0$*, then the risk has the following asymptotic expansion:*

$$r_1(T) = \sigma^2 + \frac{\theta^2}{4}\tau^2 + o\left(\tau^2\right);$$

the risk sensitivity coefficient and the ε*-admissible length of the censoring interval have the following approximations:*

$$\chi_1 \approx \frac{\theta^2}{4\sigma^2}\tau^2, \ \tau_1(\varepsilon) \approx \sqrt{\varepsilon}\,\frac{2\sigma}{\theta}.$$

The asymptotic expansions (under $\tau = b - a \to 0$) for the forecasting statistics (9) and (12) are similar. But we will see in Section 5 that if the length of the censoring interval increases, then the risk of the forecasting statistic (12) becomes higher than the risk of the robust forecasting statistic (9).

Corollary 1. *If* $a \to -\infty$ *and* $b \to +\infty$ *, then the risk of* *the forecasting statistic (12) is* $r_1(T) = \dfrac{\sigma^2}{1-\theta^2}$.

Another widely used in practice forecasting statistic is

$$\hat{x}_{T+1} = \theta \frac{a+b}{2}. \qquad (13)$$

Theorem 9. *Let values* $x_1, \ldots, x_{T-1}$ *and the event* $A_T^* = \{x_T \in (a,b)\}$ *be observed. Then the risk of the forecasting statistic (13) is*

$$r_2(T) = \frac{\sigma^2}{1-\theta^2} + \frac{\theta^2 (a+b)^2}{4} + \frac{\theta^2 \sigma}{\sqrt{(1-\theta^2)}} \times$$

$$\frac{a\varphi\left(\dfrac{b\sqrt{1-\theta^2}}{\sigma}\right) - b\varphi\left(\dfrac{a\sqrt{1-\theta^2}}{\sigma}\right)}{\Phi\left(\dfrac{b\sqrt{1-\theta^2}}{\sigma}\right) - \Phi\left(\dfrac{a\sqrt{1-\theta^2}}{\sigma}\right)}.$$

Corollary 1. *If* $\tau = b - a \to 0$, *then the risk of the forecasting statistic (13) is* $r_2(T) = \sigma^2$.

So if the length of the censoring interval τ is close to zero, then the forecasting statistic (13) gives a prediction that is close to the optimal one. But as we will see in Section 5 if τ is increasing, then the quality of prediction of the forecasting statistic (13) becomes extremely bad (the risk grows very fast).

5. NUMERICAL RESULTS

Computer experiments are performed for the case of the AR(1) model (7) and $q = 1$ to compare the experimental risk of forecasting statistics (9), (12), (13) and the theoretical risk given by Theorems 6, 8, 9. For every fixed τ the Monte-Carlo experiments with 10000 simulations of time series are used to evaluate the experimental value of the risk and its 95%-confidence limits. For simulations the following values of parameters are used: $p = 1$, $\theta = 0.8$, $\sigma^2 = 1$, $T = 100$, $\tau = b - a \in \{0, 0.5, 1, \ldots, 15\}$. The last observation of the time series is replaced by the random censoring interval (a_T, b_T): the length of the interval (a_T, x_T) is $\alpha\tau$ and the length of the interval (x_T, b_T) is $(1-\alpha)\tau$, where α is the standard uniformly distributed random variable.

Fig.1 presents experimental values of the risk for three forecasting statistics (9), (12), (13). As we can see the robust forecasting statistic (9) has the smallest risk. The risk of the forecasting statistic (12) is approximately two times higher than for the optimal one. The risk of the forecasting statistic (13) grows very fast.

Fig.2–4 present experimental values of the risk and its 95%-confidence limits for each forecasting statistic. Theoretical values of the risk are also given. This figures show a sufficiently good fit of theoretical and experimental results.

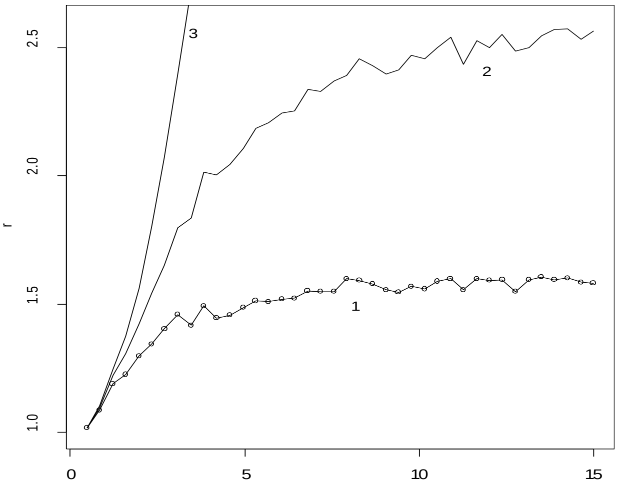

Fig.1 – Comparison of risks for all three forecasting statistics: 1 – risk of the robust forecasting statistic (9), 2 – risk of the forecasting statistic (12), 3 – risk of the forecasting statistic (13).

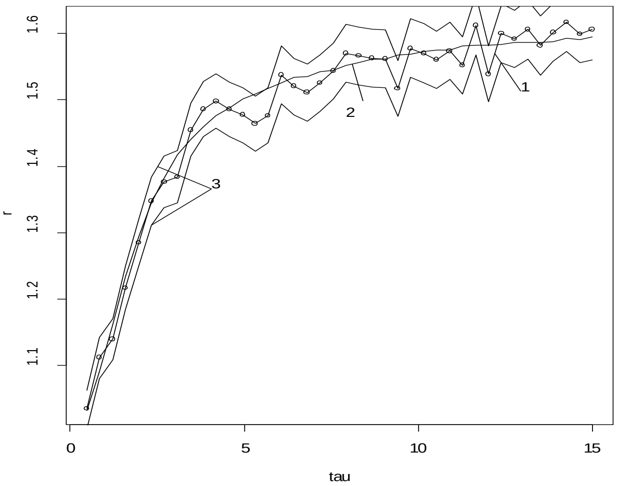

Fig.2 – Theoretical and experimental values of the risk for the robust forecasting statistic (9): 1 – experimental values of the risk, 2 – theoretical values of the risk, 3 – 95%-confidence limits.

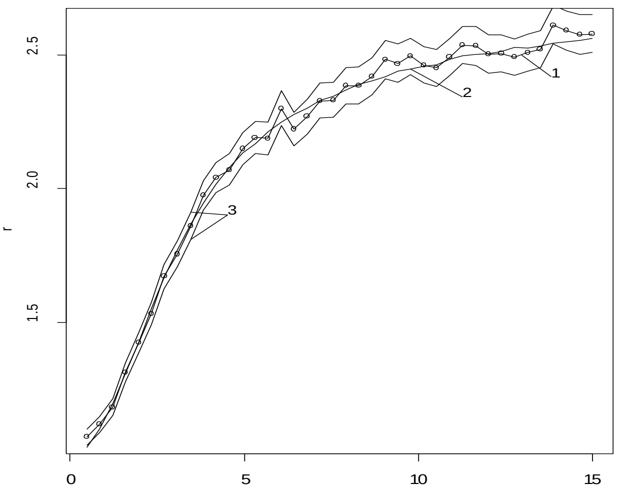

Fig.3 – Theoretical and experimental values of the risk for the forecasting statistic (12): 1 – experimental values of the risk, 2 – theoretical values of the risk, 3 – 95%-confidence limits.

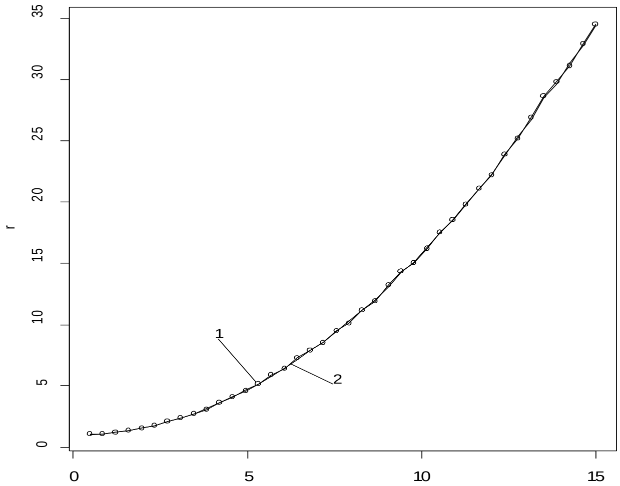

Fig.4 – Theoretical and experimental values of the risk for the forecasting statistic (13): 1 – experimental values of the risk, 2 – theoretical values of the risk, 3 – 95%-confidence limits.

6. REFERENCES

[1] G.E.P Box & G.M. Jenkins. *Time Series Analysis: Forecasting and Control*. Holden-Day. San-Francisco, 1970.

[2] T.W. Anderson. *The Statistical Analysis of Time Series*. Wiley. New-York. 1971.

[3] Yu.S. Kharin. *Optimality and robustness in statistical forecasting*. BSU. Minsk. 2008. (in Russian)

[4] R. Little, D. Rubin. *Statistical Analysis with Missing Data*. Wiley. New-York. 1987.

[5] J.W. Park, M.G. Genton, S.K. Ghosh. Censored Time Series Analysis with Autoregressive Moving Average Models. *The Canadian Journal of Statistics*. 35 (1) (2007). p. 151–168.

[6] P.K. Sen. Censoring in Theory and Practice: Statistical Perspectives and Controversies. *IMS Lecture Notes*. 27. (1995).

[7] G. Gomez, A. Espinal, W. Lagakos. Inference for a Linear Model with an Interval-censored Covariate. *Statistics in Medicine*. 22. (2003). p. 409–425.

[8] T.W. Anderson. *An Introduction to Multivariate Statistical Analysis*. Wiley. New-York. 1963.

[9] I.S. Gradshteyn, I.M. Ryzhik. *Table of Integrals, Series, and Products, seventh edition*. Academic Press, 2007.