ISSN 0321-3900

Theory of Stochastic Processes

5(21)•1999 № 3-4

scientific
TBiMC
publishers

Theory of Stochastic Processes
Vol. 5 (21), no. 3-4, 1999, pp. 84-90

UDC 519.21

YURIJ S. KHARIN AND DMITRIJ V. ZENEVICH

ROBUSTNESS OF STATISTICAL FORECASTING BY AUTOREGRESSION MODEL UNDER DISTORTIONS

The paper deals with autoregressive model $AR(p)$ under two types of distortions: 1) parameter specification errors; 2) unknown initial values $X_0 \in R^p$. By the method of asymptotic expansion we construct new estimates of robustness characteristics (mean-square risk, guaranteed upper risk, δ-admissible distortion level) for the traditional forecasting procedures and also a new statistical estimator of initial values X_0. The results are illustrated by computer modelling.

1. Introduction

Let a time series x_t, $t \in N$ defined on the probability space $(\Omega, F, \mathbf{P})$ is observed for T time moments: $t = 1, 2, \ldots, T$ and is forecasted for the time moment $t = T + \tau$ $(\tau \geq 1)$. The hypothetical model of the time series used for calculation of the forecast $\hat{x}_{T+\tau}$ is the autoregression model $AR(p)$:

$$x_t = \theta^{0'} X_{t-1} + \xi_t, \tag{1}$$

where $X_{t-1} = (x_{t-1}, \ldots, x_{t-p})' \in R^p$, $\theta^0 \in R^p$ is a vector of coefficients, $\{\xi_t : t \in N\}$ are i.i.d. random variables, $\mathbf{E}\{\xi_t\} = 0$, $D\{\xi_t\} = \sigma^2$, initial values $x_0, x_{-1}, \ldots, x_{1-p}$ are known. The traditional autoregressive forecasting procedure has the form [1]:

$$\hat{x}_{T+j} = \theta' \hat{X}_{T+j-1}, \quad j = \overline{1, \tau}, \tag{2}$$

where $\hat{X}_{T+j-1} = (\hat{x}_{T+j-1}, \ldots, \hat{x}_{T+j-q})' \in R^q$, $\hat{X}_T = (x_T, \ldots, x_{T-q+1})' \in R^q$; $\theta \in R^q$ is a vector of coefficients used in forecasting, τ is the forecasting horizon.

In many applied problems of time series forecasting is used hypothetical autoregression model [1], [6]. In practice, however, the hypothetical model is often distorted. For example, parameters q and θ in (2) may differ from real values p and θ^0, i.e there are parameter specification errors. This type of distortions are investigated in [2],[3],[4] and others. These papers are concentrated on the problem of parameter θ estimation under distortions, but the problem of robustness of forecasting under distortions are not discussed there. The next section is devoted to investigations of *robustness of forecasting* in case of parametric specification errors ($\theta \neq \theta^0$, $p \neq q$, see [7] for details). Uncertainty in initial values X_0 also generates some distortions in the hypothetical model (1). This type of distortions will be discussed in the sections 3, 4.

1991 *AMS Mathematics Subject Classification.* 62F35,62J05.
Key words and phrases. Autoregressive model $AR(p)$, robustness characteristics, forecasting procedure.

2. Parameter specification errors

The case of deterministic specification errors. At first consider influence of deterministic error $\theta - \theta^0$ on forecasting risk in the case of known AR order: $q = p$. Introduce p-vector U_t and some $(p \times p)$-matrices $(k \in N)$:

(3)
$$U_t = \begin{pmatrix} \xi_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad B_0 = \begin{pmatrix} \theta_1^0 & \dots & \theta_{p-1}^0 & \theta_p^0 \\ 1 & \dots & 0 & 0 \\ \vdots & & \vdots & \vdots \\ 0 & \dots & 1 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} \theta_1 & \dots & \theta_{p-1} & \theta_p \\ 1 & \dots & 0 & 0 \\ \vdots & & \vdots & \vdots \\ 0 & \dots & 1 & 0 \end{pmatrix},$$
$$D = diag(\sigma^2, 0, \dots, 0), \quad S_k = \sum_{i=0}^{k-1} B_0^i D (B_0^i)', \quad C_k = S_k + B_0^k X_0 (B_0^k X_0)',$$
$$R(\theta^0, \theta; \tau) = \left(R_{ij}(\theta^0, \theta; \tau)\right) = S_\tau + (B^\tau - B_0^\tau) C_T (B^\tau - B_0^\tau)', \quad i, j = \overline{1, p}.$$

We use the mean-square risk of prediction for horizon τ: $r(\theta^0, \theta, \tau) ::= \mathbf{E}\{(\hat{x}_{T+\tau} - x_{T+\tau})^2\}$ as a characteristic of forecasting performance.

Theorem 1. *If the forecasting procedure (2) is used, then* $r(\theta^0, \theta; \tau) = R_{11}(\theta^0, \theta; \tau)$.

This result follows from the expressions $X_t = B_0 X_{t-1} + U_t$, $\hat{X}_t = B \hat{X}_{t-1}$.

Corollary 1. *Minimal w.r.t. θ risk is obtained at $\theta = \theta^0$:* $r_{min}(\theta^0; \tau) = R_{11}(\theta^0, \theta^0; \tau)$.

Corollary 2. *Let* $\alpha = \theta - \theta^0$ *and* $\beta(\tau) = \left(\sum_{i=0}^{\tau-1} (B_0^i)_{11} B_0^{\tau-1-i}\right) C_T \left(\sum_{j=0}^{\tau-1} (B_0^j)_{11} B_0^{\tau-1-j}\right)'$.
Then the following asymptotic expansion w.r.t. α is valid:

(4)
$$r(\theta^0, \theta; \tau) = r_{min}(\theta^0; \tau) + \sum_{k,l=1}^{p} \beta_{lk}(\tau) \alpha_l \alpha_k + O(|\alpha|^3).$$

Let us represent $B = B_0 + \Delta$, where $\Delta' = (\alpha \vdots O_p \vdots \cdots \vdots O_p)$ and O_p is p-vector of zeros. Then by the Theorem 1 we come to the result.

Let us denote the guaranteed upper risk: $r_+(\theta^0; \tau; \varepsilon) = \sup_{|\alpha| \le \varepsilon} r(\theta^0, \theta^0 + \alpha; \tau)$, where $\varepsilon \ge 0$ is the known admissible level of the specification error $|\alpha|$. Let us say that error α satisfies the condition of δ-admissibility (for any $\delta \ge 0$) if $r_+(\theta^0; \tau; \varepsilon) \le (1 + \delta) r_{min}(\theta^0; \tau)$.

Theorem 2. *Under Theorem 1 conditions for any fixed $\delta \ge 0$ and critical level $\varepsilon \to 0$:*
1) $r_+(\theta^0; \tau; \varepsilon) = r_{min}(\theta^0; \tau) + \varepsilon^2 \lambda_{max} + O(\varepsilon^3)$;
2) the set of δ-admissible errors α is the ellipsoid: $\sum_{i=1}^{p} \lambda_i \mu_i^2 \le \delta r_{min}(\theta^0; \tau)$,
where $\{\lambda_i \ge 0, \ i = \overline{1, p}\}$ *are the eigenvalues of the matrix* $\beta(\tau)$, λ_{max} *is the maximal eigenvalue;* $\mu = T\alpha$, T *is the orthogonal matrix such that* $\beta(\tau) = T' diag\{\lambda_1, \dots, \lambda_p\} T$.

This result is a consequence of using expansion (4) without remainder $O(|\alpha|^3)$ and indicated form of symmetrical non-negatively defined matrix $\beta(\tau)$.

Corollary 3. *The error α is δ-admissible if* $|\alpha| \le \sqrt{\delta r_{min}(\theta^0; \tau) / \lambda_{max}(\beta(\tau))}$.

Unconditional risk of forecasting. Now let us consider the situation where the vector θ is a random vector distributed near θ^0 and let us investigate the unconditional forecasting risk: $r(\theta^0; \tau) = \mathbf{E}\{r(\theta^0, \theta; \tau)\}$, where $\mathbf{E}\{\cdot\}$ is the expectation symbol.

Theorem 3. *Let the parametric error α is a random vector with covariance matrix Σ and restricted third order moments such that $\varepsilon_+ = \max_{ijk} \mathbf{E}\{|\alpha_i\alpha_j\alpha_k|\} \to 0$. Then the unconditional forecasting risk satisfies the following asymptotic expansion:*

$$r(\theta^0;\tau) = r_{min}(\theta^0;\tau) + tr(\beta(\tau)\Sigma) + O(\varepsilon_+).$$

For fixed θ, θ^0 and τ risk is a bounded polynomial w.r.t. $\{\theta_i, i = \overline{1,p}\}$. That is why all its derivatives are bounded in a closed region. Using the Taylor approximation and boundedness of derivatives we prove this Theorem.

The case of misspecification of AR order. Let us investigate the influence of the error ($q \neq p$) in assignment of AR order on the risk of forecasting. Let us evaluate the case $q < p$ (the case of $p > q$ is investigated in the same way). In this case last $p - q$ elements of the first row of the matrix B are zeros. Make the partition: $\theta^{0\prime} = (\theta^{0\prime}_{(1)} \vdots \theta^{0\prime}_{(2)})$, $\theta' = (\theta'_{(1)} \vdots \theta'_{(2)})$, where the first blocks of these p-vectors are q-vectors and $\theta_{(2)} = O_{p-q}$; $\beta(\tau) = \begin{pmatrix} \beta_{(11)} & \beta_{(12)} \\ \beta'_{(12)} & \beta_{(22)} \end{pmatrix}$, where $\beta_{(11)}$ is the $(q \times q)$-matrix.

Theorem 4. *Let the observed time series $\{x_t\}$ is described by the $AR(p)$-model and in the forecasting procedure (2) we use the $AR(q)$-model ($q < p$) where the first q coefficients are exactly estimated ($\theta_{(1)} = \theta^0_{(1)}$). Then the risk satisfies the expansion:*

$$r(\theta^0,\theta;\tau) = r_{min}(\theta^0;\tau) + \theta^{0\prime}_{(2)}\beta_{(22)}\theta^0_{(2)} + O(|\alpha|^3).$$

This expansion follows from (4).

Corollary 4. *The AR order error is δ-admissable if $\theta^{0\prime}_{(2)}\beta_{(22)}\theta^0_{(2)} < \delta r_{min}(\theta^0;\tau)$.*

All results of this section were verified by computer modelling [7].

3. The case of unknown initial value X_0

For estimating of unknown autoregressive coefficients the least squares method is often used [1]: $\hat{\theta} = \arg\min_{\theta} F(\theta, X_0)$, where $F(\theta, X_0) = \sum_{t=1}^{T}(x_t - \theta' X_{t-1})^2$ is the error function. The explicit form of this estimator for the hypothetical model (1) is

$$\hat{\theta} = A^{-1}a, \qquad A = \sum_{t=1}^{T} X_{t-1}X'_{t-1}, \; a = \sum_{t=1}^{T} x_t X_{t-1}. \tag{5}$$

It is seen from (5) that the matrix A and the vector a are dependent on the vector of initial values $X_0 = (x_0, x_{-1}, ..., x_{1-p})'$, i.e. we have a function $\hat{\theta} = \hat{\theta}(X_0)$. In theoretical analysis X_0 is usually assumed to be known but in practice this assumption is not valid. So for estimating of autoregressive coefficients we need to estimate initial values X_0. It is a significant problem for small samples.

The most popular practical methods of estimating X_0 are [1]-[7]: 1) to rename indexes $x_{t-p} ::= x_t$, $t = \overline{1,T}$; 2) to assign $X_0 ::= (0, ..., 0)' \in R^p$; 3) to assign X_0 as mean value $X_0 ::= \mathbf{E}\{X_0\}$; 4) "back forecasting". Using of these methods in practice has empirical base only. Note that the most popular method is the first one, but it is not good for a small sample because of reducing of its size from T to $T - p$.

Least squares estimate of initial values. Let us apply the least squares method for estimation of the vector X_0 in the following way:

$$\hat{X}_0 = \arg\min_{X_0} F(\hat{\theta}, X_0). \tag{6}$$

Now find the explicit form of the LSE (6). At first, determine the functional dependence of the vector a, the matrix A on the vector X_0. For this purpose define shift $(p \times p)$-matrix S_k left lower $(p-k) \times (p-k)$-block of which is the unit matrix and other blocks are zeros: $S_k = \begin{pmatrix} O_{k\times(p-k)} & O_{k\times k} \\ I_{(p-k)\times(p-k)} & O_{(p-k)\times k} \end{pmatrix}$, $k = \overline{0,p}$ $(S_0 = I_p,\ S_p = O_p)$. Denote:

$$\begin{gathered} L = \sum_{t=0}^{p-1} x_{t+1} S_t, \quad G = G(X_0) = \sum_{t=0}^{p-1} (S_t X_0 X_0' S_t' + S_{p-t}' X_p X_0' S_t' + S_t X_0 X_p' S_{p-t}), \\ h = \sum_{t=0}^{p-1} x_{t+1} S_{p-t}' X_p + \sum_{t=p}^{T-1} x_{t+1} X_t, \quad K = \sum_{t=0}^{p-1} S_{p-t}' X_p X_p' S_{p-t} + \sum_{t=p}^{T-1} X_t X_t'. \end{gathered} \tag{7}$$

Using (7) and evident representation $X_t = S_{p-t}' X_p + S_t X_0$ we get the following result.

Lemma 1. *The vector a and the matrix A in (5) are $a = LX_0 + h$, $A = G(X_0) + K$.*

Let $\|\cdot\|_m$ is some matrix norm, $\|\cdot\|_v$ is some vector norm, and $w = \|K\|_m^{-3}\|h\|_v^2$. Separate "linear" (G_1) and "square" (G_2) parts of the matrix G: $G = G_2 + G_1$, $G_2 = \sum_{t=0}^{p-1} S_t X_0 X_0' S_t'$, $G_1 = \sum_{t=0}^{p-1} (S_{p-t}' X_p X_0' S_t' + S_t X_0 X_p' S_{p-t})$.

Lemma 2. *The error function $F(\hat{\theta}, X_0)$ satisfies the expansion:*

$$F(\hat{\theta}, X_0) = f_2(X_0, X_0') - f_1(X_0) + f_0 + O(w), \tag{8}$$

$$f_2(X_0, X_0') = h'K^{-1}G_2K^{-1}h + 2h'K^{-1}G_1K^{-1}LX_0 - X_0'L'K^{-1}LX_0 -$$

$$-K^{-1}G_1K^{-1}G_1K^{-1}, \ f_1(X_0) = 2h'K^{-1}LX_0 - h'K^{-1}G_1K^{-1}k, \ f_0 = \sum_{t=1}^{T} x_t^2 - h'K^{-1}h.$$

Using Lemma 1 we have $A^{-1} = (K+G)^{-1} = ((I_p + GK^{-1})K)^{-1} = K^{-1}(I + GK^{-1})^{-1}$. By applying Taylor approximation of $(I + GK^{-1})^{-1}$ we get the expansion: $A^{-1} = K^{-1}(I - GK^{-1} + GK^{-1}GK^{-1} + o(\|K^{-1}\|_m^2)) \cdot 1_p = K^{-1} - K^{-1}GK^{-1} + O(\|K\|_m^{-3}) \cdot 1_p$, where 1_p - $(p \times p)$-matrix of ones. Its substitution into the expression of the function $F(\hat{\theta}, X_0)$ proves the final results.

Theorem 5. *The LSE (6) of initial values X_0 satisfies the expansion:*

$$\hat{X}_0 = Z^{-1}z + O(w)1_p. \tag{9}$$

$$Z = \sum_{t=0}^{p-1} \Big(S_t'K^{-1}hh'K^{-1}S_t + h'K^{-1}S_{p-t}'X_pS_t'K^{-1}L + L'K^{-1}S_tX_p'S_{p-t}K^{-1}h +$$

$$+ S_t'K^{-1}hX_p'S_{p-t}K^{-1}L + L'K^{-1}S_{p-t}'X_ph'K^{-1}S_t \Big) - L'K^{-1}L,$$

$$z = LK^{-1}h - \sum_{t=0}^{p-1} h'K^{-1}S_{p-t}'X_pS_t'K^{-1}h.$$

The LSE must satisfy two conditions: $\nabla_{X_0} F(\hat{\theta}, X_0) = 0$ and $\nabla_{X_0}^2 F(\hat{\theta}, X_0) \succeq 0$. Differentiating of $F(\hat{\theta}, X_0)$ in the form (8) leads to the result (9).

Let us investigate some information properties of estimation of X_0. The estimate $\hat{X}_0$ is constructed using the information from the sample $\{x_t\}$, $t = \overline{1,T}$. Let us find the

Shannon information contained in the vector X_t about the vector X_0: $I\{X_0, X_t\} = \mathbf{E}\left\{\ln \frac{p_{X_0,X_t}(x,y)}{p_{X_0}(x)p_{X_t}(y)}\right\}$. Let us denote the covariance $\Sigma = Cov\{X_0, X_0\}$ and Σ_{X_0,X_t} is the covariance matrix of the composite vector $(X_0' \vdots X_t')' \in R^{2p}$.

Lemma 3. *For the model (1)* $Cov\{X_0, X_t\} = Cov\{X_0, X_0\} B_0^{t'}$.

Theorem 6. *Let an $AR(p)$ time series (1) is a Gaussian stationary time series. Then the Shannon information about X_0 contained in X_t is*

$$I\{X_0, X_t\} = \ln \sqrt{|\Sigma|/|\Sigma - B_0^t \Sigma B_0^{t'}|}.$$

The equation $|\Sigma_{X_0,X_t}| = |\Sigma||\Sigma - \Sigma'_{X_0,X_t}\Sigma^{-1}\Sigma_{X_0,X_t}|$ and Lemma 3 prove the result.

Corollary 5. *If (1) is a stable autoregression model, then $I\{X_0, X_t\} \to 0$ at $t \to \infty$.*

Let λ_{max} is the greatest in absolute eigenvalue of matrix B_0. Then according to [1] for every $\lambda > |\lambda_{max}|$ there exists a constant c such that all elements of the matrix B_0^t are less than $c\lambda^t$, $t = 0, 1, \ldots$. This result together with Theorem 6 and property of stable autoregression: $|\lambda_{max}| < 1$ gives the result.

Note, that Corollary 5 shows the *impossibility* of existence some consistent estimator of vector X_0 using unique sample $\{x_t\}$, $t = \overline{1,T}$.

4. Generalization of the model for the case of M samples

Let M independed samples of autoregressive processes $\{x_t^{(m)},\ t = \overline{1,T_m},\ m = \overline{1,M}\}$ of type (1) with the same initial vector X_0 are observed, i.e.

$$x_t^{(m)} = \theta^{0'} X_{t-1}^{(m)} + \xi_t^{(m)}, \quad X_0^{(m)} = X_0, \quad t = \overline{1,T_m}, \quad m = \overline{1,M},$$

where m is the sample index; every sample satisfies (1); the errors are uncorrelated: $\mathbf{E}\left\{\xi_t^{(i)}\xi_{t'}^{(j)}\right\} = \sigma^2\delta_{tt'}\delta_{ij}$. We have $2p+1$ unknown parameters: $\theta^0, X_0 \in R^p$, σ^2.

Let add index $^{(m)}$ to all results of previous section for indicating theirs membership to the m-th sample. For estimation of unknown parameters we use the least squares technique:

$$(10) \qquad \hat{\theta} = \arg\min_{\theta} F_M(\theta, X_0), \quad \hat{X}_0 = \arg\min_{X_0} F_M(\hat{\theta}, X_0), \quad s^2 = \frac{F_M(\hat{\theta}, \hat{X}_0)}{\sum_{m=1}^M T_m - 1},$$

where $F_M(\theta, X_0) = \sum_{m=1}^M F^{(m)}(\theta, X_0)$ is the total error function.

Explicit forms of these estimators are constructed in the same way as for the unique sample case.

Theorem 7. *The LSE (10) of initial values θ^0, X_0 are:*

$$(11) \qquad \hat{\theta} = A_M^{-1} a_M, \quad \hat{X}_0 = Z_M^{-1} z_M + O(w_M) 1_p$$

where $A_M = \sum_{m=1}^M A^{(m)}$, $a_M = \sum_{m=1}^M a^{(m)}$, $Z_M = \sum_{m=1}^M Z^{(m)}$, $z_M = \sum_{m=1}^M z^{(m)}$, $w_M = \sum_{m=1}^M w^{(m)}$ *and* $A^{(m)}$, $a^{(m)}$, $Z^{(m)}$, $z^{(m)}$, $w^{(m)}$ *are defined in (5) and Theorem 5.*

This result follows from the properties of the error function $F_M(\theta, X_0)$.

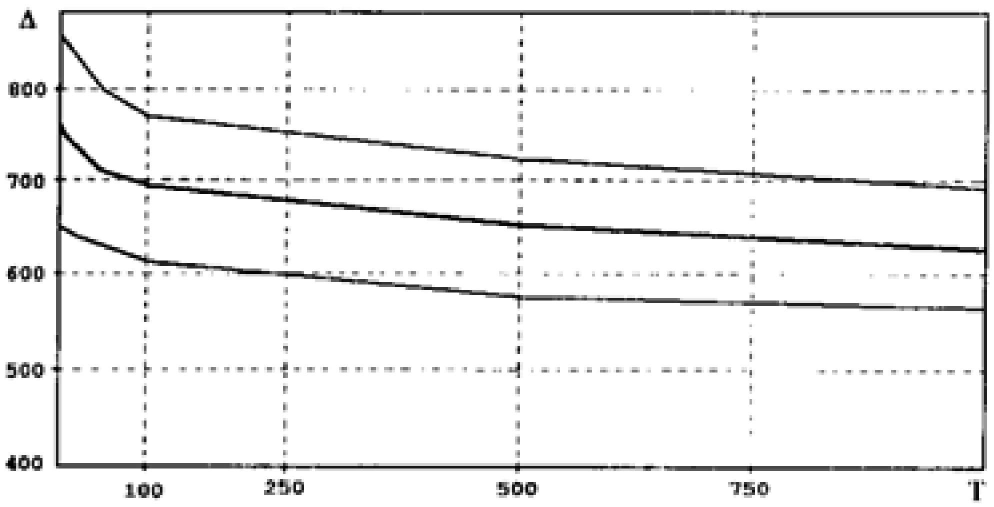

FIGURE 1. Plots of dependence of Δ on T with 95%-confidence limits

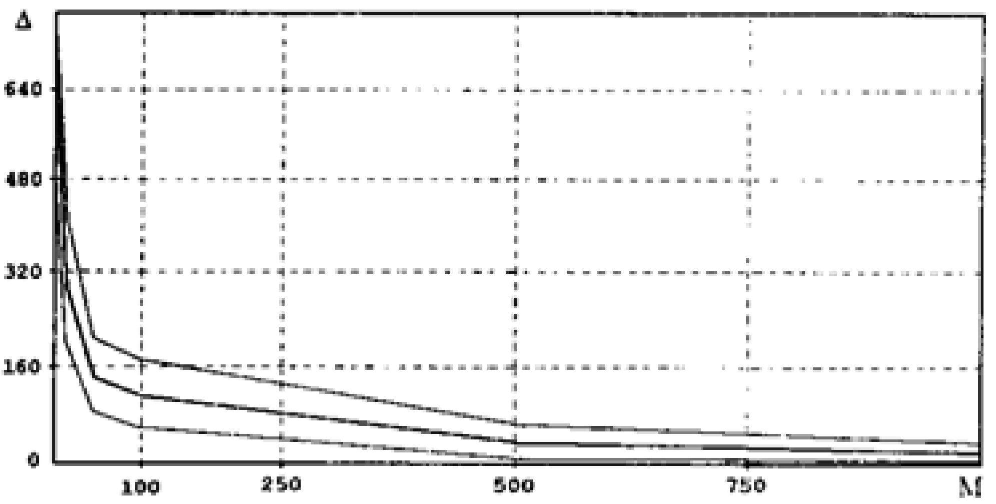

FIGURE 2. Plots of dependence of Δ on M with 95%-confidence limits

5. NUMERICAL RESULTS

The results of Theorems 5, 7 were verified by Monte-Carlo modelling. We generated N random samples by the model (1) with unique vector X_0 and estimated X_0 using (9) and (11). After that we calculated deviation $\Delta_i = \|\hat{X}_0^{(i)} - X_0\|^2$, where $\hat{X}_0^{(i)}$ is the estimate of X_0 in i-th experiment ($i = \overline{1,N}$) and averaged them: $\Delta = \frac{1}{N}\sum_{i=1}^{N}\Delta_i$. In simulations we used $AR(2)$-model with $\theta^0 = (1.14, -0.32)$; $X_0 = (6.69, 4.14)$; $\lambda_1 = 0.64$, $\lambda_2 = 0.50$; $\sigma^2 = 63.7$. Figure 1 illustrates inconsistency of LSE (6) using the only sample. Figure 2 demonstrates result of using M samples in estimating (size of each sample $T_m = 20$, $m = \overline{1,M}$); it illustrates the consistency property.

BIBLIOGRAPHY

1. Anderson T.W., *The Statistical Analysis of Time Series* (1971), John Wiley & Sons,Inc., New York/London/Sydney/Toronto.
2. Bhansali R.J., *Effects of not knowing the order of an autoregressive process on the mean squared error of prediction*, J. Amer. Statist. Assoc. **76** (1981), 588–597.
3. Chen Han-Fu, Denie C., *Parameter estimation for ARMA processes with errors in models.*, Stat. & Prob. Letters **20** (1994), 91–99.
4. Dahlhaus R. and Wefelmeyer W., *Asimptotically optimal estimation in misspecified time series models*, Ann. Statist. **24** (1996), no. 3, 952–974.
5. Hannan E.J., *Regression for time series with errors of measurement*, Biometrika **50** (1963), 293–302.

6. Kharin Yu., *Robustness in Statistical Pattern Recognition* (1996), Kluwer Academic Publishers, Dordrecht/Boston/London, 302.
7. Kharin Yu.S., Zenevich D.V., *Robastness of Autoregressive Forecasting under Misspecified Model*, Computer Data Analysis & Modeling. Proc. of the V-th Int. Conf. (1998), BSU, Minsk, 120–126.

DEPT. OF MATH. MODELLING & DATA ANALYSIS, FACULTY OF APPLIED MATH. & INFORMATICS, BELARUS STATE UNIVERSITY, 4 FR. SKORINY AVE., MINSK 220050, BELARUS
E-mail: kharin@fpm.bsu.unibel.by