ВЕСТНИК ТОМСКОГО ГОСУДАРСТВЕННОГО УНИВЕРСИТЕТА
2015 Управление, вычислительная техника и информатика № 3 (32)

УДК 519.233.22
DOI 10.17223/19988605/32/4

M.I. Kusainov

RISK EFFICIENCY OF ADAPTIVE ONE-STEP PREDICTION OF AUTOREGRESSION WITH PARAMETER DRIFT

A scalar stable autoregressive process with the dynamic parameter corrupted by additive noise is studied. The model parameters are assumed to be unknown. The truncated estimators of the parameters are used to build adaptive one-step predictors. The associated problem is to minimize a risk function of special form, defined to account for sample size and squared prediction error's sample mean. A sequential procedure is introduced to achieve the minimal risk.
Keywords: adaptive predictors; asymptotic risk efficiency; optimal sample size; scalar autoregression; stopping time; truncated parameter estimators.

When studying dynamic systems, the identification problem is often a major one to consider. A system's model is assumed to incorporate unknown parameters, estimation of which is vital for later research. Classic methods, such as maximum likelihood estimation, least squares fitting, etc., have known asymptotic properties. However, infinite samples do not occur in reality and efforts are being made to achieve non-asymptotic qualities of estimators. E.g., one may consider confidence regions for finite number of measurements (see [1, 2] among others).

Another approach are methods that employ sequential analysis. Among such methods is sequential estimation method (see, e.g., [3–12]), which has guaranteed accuracy by samples of random and finite, albeit unbounded, size. Its idea was further developed into the truncated sequential estimation method (see, e.g., [12–15]), which utilizes samples of bounded random size.

Recently, the truncated estimation method was suggested in [16] as a modification of the truncated sequential estimation method. Truncated estimators were constructed for ratio type functionals and only need samples of fixed non-random size to achieve guaranteed accuracy in the sense of the L_{2m}-norm, $m \ge 1$.

One possible use for estimated parameters is to predict future values of the modeled random process by existing observations. In order to control both the quality of predictions and the required sample size, a loss function dependent on the two is introduced. A risk efficiency problem arises, where the expected loss is minimized by choosing a certain duration of observations. Similar problems for autoregressive processes were examined in [17] and [18], the least squares estimators and sequential estimators of unknown parameters were used.

In this paper we consider a scalar stable AR(1) with parameter drift and construct real-time predictors based upon the truncated estimators of the parameters. All the parameters are assumed to be unknown. A similar model was studied in [14]. Sequential approach was, for the first time, applied to it to good effect. Resulting estimators were shown to have preassigned mean square accuracy and uniform in parameter asymptotic normality. These estimators, however, had different form than in this paper due to some parameters being known. We solve the optimization problem associated with the loss function of a special form. The proposed procedure is shown to be asymptotically risk efficient as the cost of prediction error tends to infinity. The simulation results confirm the result, but are not included for editorial reasons.

The scalar case without the drift was considered in [19], multivariate AR(1) in [20] and ARMA(1,1) in [21].

1. Problem statement

Consider the stable scalar autoregressive process satisfying the equation

$$x_k = \lambda_{k-1} x_{k-1} + \xi_k, \quad k \ge 1, \tag{1}$$

where

$$\lambda_k = \lambda + \eta_k, \quad k \ge 0,$$

the parameter λ is unknown, the condition $Ex_0^2 < \infty$ holds, ξ_k and η_k are sequences of independent identically distributed (i.i.d.) zero mean random variables, that are also independent of each other, with finite variances $\sigma_\xi^2 = E\xi_1^2, \sigma_\eta^2 = E\eta_1^2$. In addition, to guarantee stability of the process (1) we assume the following

$$\lambda^2 + \sigma_\eta^2 < 1. \tag{2}$$

It is known that the optimal in the mean square sense one-step predictor is the conditional expectation of the process with respect to its past, i.e.

$$x_k^{\text{opt}} = \lambda x_{k-1}, \quad k \ge 1.$$

Therefore, one needs an estimator $\hat{\lambda}_k$ for the unknown parameter λ to construct the adaptive predictors of the form

$$\hat{x}_k = \hat{\lambda}_{k-1} x_{k-1}, \quad k \ge 1. \tag{3}$$

Write the corresponding prediction errors

$$\hat{e}_k = x_k - \hat{x}_k = (\lambda - \hat{\lambda}_{k-1}) x_{k-1} + \eta_{k-1} x_{k-1} + \xi_k.$$

Let e_n^2 denote the sample mean of squared prediction error

$$e_n^2 = \frac{1}{n} \sum_{k=1}^{n} \hat{e}_k^{\,2}. \tag{4}$$

Define the loss function

$$L_n = \frac{A}{n} e_n^2 + n.$$

One way the parameter $A(>0)$ can be interpreted is being the cost of prediction error.

The corresponding risk function

$$R_n = E_\theta L_n = \frac{A}{n} E_\theta e_n^2 + n, \tag{5}$$

E_θ denotes expectation under the distribution P_θ with the given parameter $\theta = (\lambda, \sigma_\xi^2, \sigma_\eta^2)$. Define $\Theta = \left\{\theta : \lambda^2 + \sigma_\eta^2 < 1, \sigma_\xi^2 < \infty\right\}$ the process' stability parameter region.

The main aim is to minimize the risk R_n on the sample size n.

2. Main result

To solve the stated problem we shall use the truncated estimation method introduced in [16]. This method makes it possible to obtain the ratio type estimators with guaranteed accuracy using a sample of fixed size. According to the method, the truncated estimator of the autoregressive parameter λ is based on a ratio type estimator, the least-squares type in this case

$$\hat{\lambda}_k = \frac{\sum_{i=1}^{k} x_{i-1} x_i}{\sum_{i=1}^{k} x_{i-1}^2}, \quad k \ge 1, \tag{6}$$

and has the form

$$\tilde{\lambda}_k = \hat{\lambda}_k \chi\left(\bar{\Delta}_k \ge H_k\right), \quad k \ge 1, \tag{7}$$

where $\bar{\Delta}_k = \frac{1}{k} \sum_{i=1}^{k} x_{i-1}^2$, the notation $\chi(B)$ means the indicator function of the set B and

$$H_k = \log^{-1/2}(k+1). \tag{8}$$

It should be noted, that according to [16], H_k can be taken as any decreasing slowly changing positive function.

A model similar to (1) was studied in [14], but variances of the noises ξ_i and η_i, $i \geq 1$ were assumed to be known. This information allowed construction of true sequential least-squares estimators. However, it is absent in our case, forcing one to use estimators of the form (7).

Rewrite the formulae (3)–(5) with $\hat{\lambda}_k$ replaced by $\tilde{\lambda}_k$ as follows

$$\tilde{x}_k = \tilde{\lambda}_{k-1} x_{k-1}, \tag{9}$$

$$\tilde{e}_k = (\lambda - \tilde{\lambda}_{k-1}) x_{k-1} + \eta_{k-1} x_{k-1} + \xi_k,$$

$$\overline{e_n^2} = \frac{1}{n} \sum_{k=1}^{n} \tilde{e}_k^2,$$

$$R_n = \frac{A}{n} E_\theta \overline{e_n^2} + n. \tag{10}$$

To minimize the risk R_n we rewrite the risk function (10) using the definition of $\overline{e_n^2}$

$$R_n = \frac{A}{n}\left(\sigma_\xi^2 + \sigma_\eta^2 \sigma_x^2 + D_n\right) + n, \tag{11}$$

where

$$\sigma_x^2 = \frac{\sigma_\xi^2}{1-(\lambda^2+\sigma_\eta^2)}, \quad D_n = \frac{1}{n}\sum_{k=1}^{n} E_\theta (x_k^{\text{opt}} - \tilde{x}_k)^2 + \frac{\sigma_\eta^2}{n}\sum_{k=1}^{n}(\sigma_{x,k-1}^2 - \sigma_x^2), \quad \sigma_{x,k}^2 = E_\theta x_k^2.$$

From here on C denotes those non-negative constants, the values of which are not critical.

Further we show that

$$D_n = o(1) \quad \text{as} \quad n \to \infty. \tag{12}$$

To this end we establish the two following estimates

$$\sum_{k=1}^{n} \left|\sigma_{x,k-1}^2 - \sigma_x^2\right| \leq C, \tag{13}$$

$$\sum_{k=1}^{n} E_\theta (x_k^{\text{opt}} - \tilde{x}_k)^2 \leq C \log^2 n. \tag{14}$$

In order to prove (13), write the solution of x_k using its definition (1) as follows

$$x_k = \sum_{i=0}^{k-1} \xi_{k-i} \prod_{j=1}^{i} (\lambda + \eta_{k-j}) + x_0 \prod_{j=1}^{k} (\lambda + \eta_{k-j}).$$

Squaring and taking the expectation of both sides, we use independence and zero mean of ξ_i and η_i, $i \geq 1$ to obtain

$$\sigma_{x,k}^2 = \sum_{i=0}^{k-1} \sigma_\xi^2 \prod_{j=1}^{i} E(\lambda + \eta_{k-j})^2 + E x_0^2 \prod_{j=1}^{k} E(\lambda + \eta_{k-j})^2 =$$

$$= \sigma_\xi^2 \sum_{i=0}^{k-1} (\lambda^2 + \sigma_\eta^2)^i + E x_0^2 (\lambda^2 + \sigma_\eta^2)^k.$$

Using (2) one gets

$$\sigma_\xi^2 \sum_{i=0}^{k-1} (\lambda^2 + \sigma_\eta^2)^i = \frac{\sigma_\xi^2}{1-(\lambda^2+\sigma_\eta^2)}(1-(\lambda^2+\sigma_\eta^2)^{k-1}) = \sigma_x^2 (1-(\lambda^2+\sigma_\eta^2)^{k-1})$$

and thus,

$$\sigma_{x,k}^2 - \sigma_x^2 = -\sigma_x^2 (\lambda^2 + \sigma_\eta^2)^{k-1} + E x_0^2 (\lambda^2 + \sigma_\eta^2)^k.$$

From (2) it follows, that

$$\sum_{k=1}^{\infty} \left|\sigma_{x,k-1}^2 - \sigma_x^2\right| \leq C,$$

hence (13).

To prove (14) rewrite its left-hand side

$$\sum_{k=1}^{n} E_\theta (x_k^{opt} - \tilde{x}_k)^2 = \sum_{k=1}^{n} E_\theta (\tilde{\lambda}_{k-1} - \lambda)^2 x_{k-1}^2. \quad (15)$$

We establish the properties of the estimators $\tilde{\lambda}_k$.

Define $k_0 = \max\left\{1, [e^{(\sigma_x^2)^{-2}}]_1\right\}$, where $[a]_1$ denotes the integer part of a.

Lemma 1. Assume the model (1) and let for some integer $m \ge 1$ the conditions

$$E\xi_1^{4m} < \infty, \quad Ex_0^{4m} < \infty, \quad E(\lambda + \eta_1)^{4m} < 1 \quad (16)$$

be true. Then the truncated estimators $\tilde{\lambda}_k$ satisfy

(i) for $1 \le k < k_0$

$$E_\theta (\tilde{\lambda}_k - \lambda)^{2m} \le C; \quad (17)$$

(ii) for $k \ge k_0$

$$E_\theta (\tilde{\lambda}_k - \lambda)^{2m} \le \frac{C \log^m k}{k^m}. \quad (18)$$

The proof of Lemma 1 is presented in Section 4.

Remark 1. The parameter H_k can be taken as a constant H, provided that $H \in (0, \sigma_\xi^2)$. The condition (2) then guarantees $H < \lim_{k\to\infty} \bar{\Delta}_k$, considering

$$\lim_{k\to\infty} \bar{\Delta}_k = \frac{\sigma_\xi^2}{1 - (\lambda^2 + \sigma_\eta^2)} \quad P_\theta - \text{a.s.}, \quad (19)$$

which follows from the ergodicity of the process $(x_k)_{k\ge 0}$ (see, e.g., [7]). Then the estimators $\tilde{\lambda}_k$ would satisfy

$$E_\theta (\tilde{\lambda}_k - \lambda)^{2m} \le \frac{C}{k^m}$$

for every $k \ge 1$, which can be proved similarly to Theorem 2 of [16]. This, though, requires knowledge of σ_ξ^2.

The Cauchy-Schwarz-Bunyakovsky inequality, (15) and (18) yield (14). The relation (12) follows directly from (13), (14).

Denote

$$\sigma^2 = \sigma_\xi^2 + \sigma_\eta^2 \sigma_x^2. \quad (20)$$

In view of (12) we minimize the principal term of R_n, analogously to [17]

$$R_n \approx \frac{A}{n}\sigma^2 + n \longrightarrow \min_n$$

to get the optimal sample size

$$n_A^o = A^{1/2}\sigma \quad (21)$$

and the corresponding approximate minimal risk value

$$R_{n_A^o} = 2A^{1/2}\sigma + O(\log^2 A) \quad \text{as} \quad A \to \infty. \quad (22)$$

Similarly to [17, 18, 20], we introduce the stopping time T_A as an estimator of n_A^o, replacing σ^2 in its definition with an estimator $\tilde{\sigma}_n^2$

$$T_A = \inf_{n \ge n_A} \left\{ n \ge A^{1/2} \tilde{\sigma}_n \right\}, \quad (23)$$

where n_A is the initial sample size depending on A and specified below (see Theorem 1),

$$\tilde{\sigma}_n^2 = \frac{1}{n} \sum_{k=1}^{n} (x_k - \tilde{\lambda}_n x_{k-1})^2. \quad (24)$$

We formulate a theorem to prove the asymptotic equivalence of T_A and n_A^o in the sense of almost sure and mean convergences (see respectively (27), (28) below) and the optimality of the adaptive prediction procedure in the sense of equivalence of $R_{n_A^o}$ and the modified risk

$$R_A = E_\theta L_{T_A} = AE_\theta \frac{1}{T_A}\overline{e_{T_A}^2} + E_\theta T_A, \qquad (25)$$

see (29).

Theorem 1. Assume that

$$E\xi_1^{16} < \infty, \quad Ex_0^{16} < \infty, \quad E(\lambda + \eta_1)^{16} < 1 \qquad (26)$$

and n_A in (23) is such that

$$n_A \geq \max\{k_0, A^r \log^2 A\}, \quad n_A \cdot A^{-1/2} \xrightarrow[A\to\infty]{} 0$$

with $r \in (2/5, 1/2)$. Let the predictors $\tilde{x}_k$ be defined by (9) and the risk functions defined by (5), (25). Then for every $\theta \in \Theta$ and $\sigma^2 > 0$

$$\frac{T_A}{n_A^o} \xrightarrow[A\to\infty]{} 1 \quad P_\theta - \text{a.s.}, \qquad (27)$$

$$\frac{E_\theta T_A}{n_A^o} \xrightarrow[A\to\infty]{} 1, \qquad (28)$$

$$\frac{R_A}{R_{n_A^o}} \xrightarrow[A\to\infty]{} 1. \qquad (29)$$

The proof of Theorem 1 is presented in Section 2.

3. Proofs

3.1. Proof of Lemma 1

It can be shown (see, e.g., [22], Lemma 1), that to guarantee $E_\theta x_k^{2m} \leq C, \ k, m \geq 1,$ the following suffices

$$E\xi_1^{2m} < \infty, \quad Ex_0^{2m} < \infty, \quad E(\lambda + \eta_1)^{2m} < 1.$$

Thus, from the conditions of Lemma 1 on noise moments for $\theta \in \Theta$ it follows

$$\sup_{k\geq 0} E_\theta x_k^{4m} \leq C. \qquad (30)$$

By the definition (7) of truncated estimators $\tilde{\lambda}_k$, their deviation has the form

$$\tilde{\lambda}_k - \lambda = (\hat{\lambda}_k - \lambda)\cdot\chi\left(\bar{\Delta}_k \geq H_k\right) - \lambda\cdot\chi\left(\bar{\Delta}_k < H_k\right). \qquad (31)$$

The definitions (6) and (1) yield

$$\hat{\lambda}_k - \lambda = \frac{\sum_{i=1}^k x_{i-1}\xi_i + \sum_{i=1}^k \eta_{i-1}x_{i-1}^2}{\sum_{i=1}^k x_{i-1}^2},$$

Hence, from (31)

$$(\tilde{\lambda}_k - \lambda)^{2m} = \frac{\left(\sum_{i=1}^k x_{i-1}\xi_i + \sum_{i=1}^k \eta_{i-1}x_{i-1}^2\right)^{2m}}{\left(\sum_{i=1}^k x_{i-1}^2\right)^{2m}}\chi(\bar{\Delta}_k \geq H_k) + \lambda^{2m}\cdot\chi(\bar{\Delta}_k < H_k).$$

From the definition of $\bar{\Delta}_k$ (see (7)) and H_k (8) it follows

$$E_\theta(\tilde\lambda_k-\lambda)^{2m} \le \frac{\log^m k}{k^{2m}} E_\theta\left(\sum_{i=1}^k x_{i-1}\xi_i + \sum_{i=1}^k \eta_{i-1}x_{i-1}^2\right)^{2m} + \lambda^{2m} P_\theta(\bar\Delta_k < H_k). \tag{32}$$

It can be easily shown, that the sums $\sum_{i=1}^k x_{i-1}\xi_i$ and $\sum_{i=1}^k \eta_{i-1}x_{i-1}^2$ for $k \ge 1$ form martingales. Thus, by the Burkholder inequality and the Hölder inequality and (30), similarly to [16], Section 5.2, we get

$$\frac{\log^m k}{k^{2m}} E_\theta\left(\sum_{i=1}^k x_{i-1}\xi_i + \sum_{i=1}^k \eta_{i-1}x_{i-1}^2\right)^{2m} \le \frac{2^{2m-1}\log^m k}{k^{2m}}\left(E_\theta\left(\sum_{i=1}^k x_{i-1}\xi_i\right)^{2m} + E_\theta\left(\sum_{i=1}^k \eta_{i-1}x_{i-1}^2\right)^{2m}\right) \le$$

$$\le \frac{C\log^m k}{k^{2m}}\left(E_\theta\left(\sum_{i=1}^k x_{i-1}^2\xi_i^2\right)^m + E_\theta\left(\sum_{i=1}^k \eta_{i-1}^2 x_{i-1}^4\right)^m\right) \le \frac{C\log^m k}{k^m}. \tag{33}$$

The first assertion (17) of Lemma 1 follows from (32) and (33).

For the second summand of (6), using the Chebyshev inequality for $k \ge k_0$ one has

$$\lambda^{2m} P_\theta(\bar\Delta_k < H_k) \le C P_\theta\left(\left|\bar\Delta_k - \sigma_x^2\right| > \sigma_x^2 - H_k\right) \le C\frac{E_\theta(\bar\Delta_k - \sigma_x^2)^{2m}}{(\sigma_x^2 - H_k)^{2m}}. \tag{34}$$

Note that for $k \ge k_0$,

$$H_k = \frac{1}{\sqrt{\log(k+1)}} < \frac{1}{\sqrt{\log e^{(\sigma_x^2)^{-2}}}} = \sigma_x^2$$

and hence the difference $\sigma_x^2 - H_k > 0$.

To estimate (34) we rewrite σ_x^2

$$\sigma_x^2 = \frac{\sigma_\xi^2}{1-(\lambda^2+\sigma_\eta^2)} = \frac{\sigma_\xi^2(1-\lambda^2)}{(1-\lambda^2)(1-(\lambda^2+\sigma_\eta^2))} =$$

$$= \frac{\sigma_\xi^2}{1-\lambda^2} + \frac{\sigma_\xi^2\sigma_\eta^2}{(1-\lambda^2)(1-(\lambda^2+\sigma_\eta^2))} = \frac{1}{1-\lambda^2}(\sigma_\xi^2 + \sigma_\eta^2\sigma_x^2).$$

Using this and the definition of the process (1), one can write $\bar\Delta_k - \sigma_x^2$ as follows

$$\bar\Delta_k - \sigma_x^2 = \frac{1}{(1-\lambda^2)}\left(\lambda^2\cdot\frac{x_0^2 - x_k^2}{k} + \frac{2\lambda}{k}\sum_{i=1}^k \eta_{i-1}x_{i-1}^2 + \frac{2}{k}\sum_{i=1}^k \lambda_{i-1}\xi_i x_{i-1} + \right.$$

$$\left. + \frac{1}{k}\sum_{i=1}^k(\xi_i^2 - \sigma_\xi^2) + \frac{1}{k}\sum_{i=1}^k(\eta_{i-1}^2 x_{i-1}^2 - \sigma_\eta^2\sigma_{x,i-1}^2) + \frac{1}{k}\sum_{i=1}^k \sigma_\eta^2(\sigma_{x,i-1}^2 - \sigma_x^2)\right).$$

Then (34), the Burkholder inequality, (30) and (13) yield

$$\lambda^{2m} P_\theta(\bar\Delta_k < H_k) \le C\frac{1}{k^m}.$$

Together with (33) this proves the second assertion of Lemma 1.

3.1. Proof of Theorem 1

The conditions on noise moments (26) yield for $\theta \in \Theta$

$$\sup_{k\ge 0} E_\theta x_k^{16} \le C. \tag{35}$$

Note that assuming the distribution of η_1 is symmetrical, the condition $E(\lambda+\eta_1)^{16} < 1$ reduces to

$$\lambda^{16} + 120(\lambda^{14}\sigma_\eta^2 + \lambda^2\sigma_\eta^{14}) + 1820(\lambda^{12}\sigma_\eta^4 + \lambda^4\sigma_\eta^{12}) + 8008(\lambda^{10}\sigma_\eta^6 + \lambda^6\sigma_\eta^{10}) + 12870\lambda^8\sigma_\eta^8 + \sigma_\eta^{16} < 1,$$

where $\sigma_\eta^{2m} = E\eta_1^{2m}, m = \overline{2,8}$.

Rewrite formula (24) for $\tilde{\sigma}_n^2$ using the definition of the process (1)

$$\tilde{\sigma}_n^2 = \frac{1}{n}\sum_{k=1}^{n}(\xi_k + \eta_{k-1}x_{k-1} + (\lambda - \tilde{\lambda}_n)x_{k-1})^2 = \frac{1}{n}\sum_{k=1}^{n}(\xi_k^2 + \eta_{k-1}^2 x_{k-1}^2) + W_n + \nu_n, \quad (36)$$

where

$$W_n = \frac{(\tilde{\lambda}_n - \lambda)^2}{n}\sum_{k=1}^{n} x_{k-1}^2, \quad \nu_n = \frac{2}{n}\sum_{k=1}^{n}\xi_k\eta_{k-1}x_{k-1} - \frac{2}{n}\sum_{k=1}^{n}(\tilde{\lambda}_n - \lambda)x_{k-1}(\xi_k + \eta_{k-1}x_{k-1}).$$

Analogously to [17], we show that

$$\tilde{\sigma}_n^2 \xrightarrow[n\to\infty]{} \sigma^2 \quad P_\theta - \text{a.s.} \quad (37)$$

Consider W_n. Using the properties (18) and the Chebyshev inequality for any $\varepsilon > 0$ we get

$$P(|\tilde{\lambda}_n - \lambda| > \varepsilon) \le \frac{1}{\varepsilon^4}E_\theta(\tilde{\lambda}_n - \lambda)^4 \le Cn^{-2}\log^2 n.$$

From the Borel-Cantelli lemma it follows that

$$\tilde{\lambda}_n \xrightarrow[n\to\infty]{} \lambda \quad P_\theta - \text{a.s.}$$

Together with (19) and (35) this yields

$$W_n \xrightarrow[n\to\infty]{} 0 \quad P_\theta - \text{a.s.} \quad (38)$$

Similar arguments can be used to show

$$\nu_n \xrightarrow[n\to\infty]{} 0 \quad P_\theta - \text{a.s.} \quad (39)$$

At the same time, strong law of large numbers, (13) and the Borel-Cantelli lemma yield

$$\frac{1}{n}\sum_{k=1}^{n}(\xi_k^2 + \eta_{k-1}^2 x_{k-1}^2) \xrightarrow[n\to\infty]{} \sigma^2 \quad P_\theta - \text{a.s.} \quad (40)$$

Then (37) follows from the representation (36), (38)–(40).

From the definition (23) of T_A it follows that with P_θ-probability one $T_A \to \infty$ as $A \to \infty$. Therefore, by (11) we have $\tilde{\sigma}_{T_A}^2 \to \sigma^2$ P_θ-a.s. and hence

$$\frac{T_A}{A^{1/2}\sigma} \xrightarrow[A\to\infty]{} 1 \quad P_\theta - \text{a.s.}$$

To prove (28) we introduce for any positive A the auxiliary sequence of numbers $\gamma_{A,n}$

$$\gamma_{A,n} = n^2 A^{-1}\frac{1}{2\log A}, \quad n \ge 1.$$

Denote

$$m_n = \frac{1}{n}\sum_{k=1}^{n}\left(\xi_k^2 + \eta_{k-1}^2 x_{k-1}^2 - \sigma^2\right).$$

Observe that m_n can be represented as a sum of two martingales and decaying to zero as $O(n^{-1})$ sequence

$$m_n = \frac{1}{n}\sum_{k=1}^{n}(\xi_k^2 - \sigma_\xi^2) + \frac{1}{n}\sum_{k=1}^{n}(\eta_{k-1}^2 x_{k-1}^2 - \sigma_\eta^2\sigma_{x,k-1}^2) + \frac{1}{n}\sum_{k=1}^{n}\sigma_\eta^2(\sigma_{x,k-1}^2 - \sigma_x^2).$$

By the definition of T_A and (36) we have

$$E_\theta T_A \le n_A + \sum_{n\ge n_A} P_\theta\left(n^2 A^{-1} \le \frac{1}{n}\sum_{k=1}^{n}(\xi_k^2 + \eta_{k-1}^2 x_{k-1}^2) + W_n + \nu_n\right) \le$$

$$\le n_A + \sum_{n\ge n_A}\left\{P_\theta\left(n^2A^{-1} \le \sigma^2 + 2\gamma_{A,n}\right) + P_\theta(|\nu_n| > \gamma_{A,n}/2) + P_\theta(W_n > \gamma_{A,n}/2) + P_\theta(|m_n| > \gamma_{A,n})\right\}. \quad (41)$$

It can be shown, analogously to [19], that for $n \ge n_A$

$$P_\theta(|\nu_n| > \gamma_{A,n}/2) + P_\theta(W_n > \gamma_{A,n}/2) + P_\theta(|m_n| > \gamma_{A,n}) \le C\gamma_{A,n}^{-2}n^{-1} = 4CA^2\log^2 A \cdot n^{-5}$$

and at the same time

$$\frac{n_A + \sum_{n\ge n_A} P_\theta\left(n^2 A^{-1} \le \sigma^2 + 2\gamma_{A,n}\right)}{A^{1/2}\sigma} \xrightarrow[A\to\infty]{} 1. \tag{42}$$

Therefore, by assumptions on n_A

$$A^{-1/2} \sum_{n\ge n_A} \left\{P_\theta(|\nu_n| > \gamma_{A,n}/2) + P_\theta(W_n > \gamma_{A,n}/2) + P_\theta(|m_n| > \gamma_{A,n})\right\} \le$$

$$\le CA^{3/2}\log^2 A \sum_{n\ge n_A} n^{-5} \le CA^{3/2}\log^2 A \cdot n_A^{-4} \le CA^{-\frac{8r-3}{2}}\log^{-6} A \xrightarrow[A\to\infty]{} 0. \tag{43}$$

Then from (41) – (43) it follows that

$$\overline{\lim_{A\to\infty}} \frac{E_\theta T_A}{A^{1/2}\sigma} \le 1. \tag{44}$$

Analogously it can be shown that

$$\underline{\lim_{A\to\infty}} \frac{E_\theta T_A}{A^{1/2}\sigma} \ge 1$$

and thus, in view of (44) the assertion (28) holds.

To prove (29) we need the following properties

$$P_\theta(T_A < N') = O(A^{-r}), \quad P_\theta(T_A > N'') = O(A^{-1}), \tag{45}$$

$$N' = [(\sigma-\varepsilon)A^{1/2}]_1, \quad N'' = [(\sigma+\varepsilon)A^{1/2}]_1 + 1, \quad 0 < \varepsilon < \sigma,$$

which can be established similarly to (4.31) of [19].

Rewrite the left-hand side of (29) using (22) and (25)

$$\frac{R_A}{R_{n_A^o}} = \frac{AE_\theta \frac{1}{T_A}\overline{e_{T_A}^2} + E_\theta T_A}{2A^{1/2}\sigma + O(\log^2 A)}. \tag{46}$$

From (28) and (46) it follows that to prove (29) it suffices to show the convergence

$$A^{1/2} E_\theta \frac{1}{T_A\sigma}\overline{e_{T_A}^2} \xrightarrow[A\to\infty]{} 1. \tag{47}$$

To this end we show that

$$A^{1/2} E_\theta \frac{1}{T_A}\overline{e_{T_A}^2}\chi(T_A < N') \xrightarrow[A\to\infty]{} 0, \quad A^{1/2} E_\theta \frac{1}{T_A}\overline{e_{T_A}^2}\chi(T_A > N'') \xrightarrow[A\to\infty]{} 0, \tag{48}$$

$$A^{1/2} E_\theta \frac{1}{T_A\sigma}\overline{e_{T_A}^2}\chi(N' \le T_A \le N'') \xrightarrow[A\to\infty]{} 1. \tag{49}$$

All three relations are proved analogously to (4.39)–(4.41) of [19] using the definition of $\overline{e_n^2}$. E.g., for (49) we write

$$A^{1/2} E_\theta \frac{1}{T_A}\overline{e_{T_A}^2}\chi(N' \le T_A \le N'') = A^{1/2} E_\theta \frac{1}{T_A^2}\sum_{k=1}^{T_A}(\lambda - \tilde\lambda_{k-1})^2 x_{k-1}^2 \chi(N' \le T_A \le N'') +$$

$$+2A^{1/2} E_\theta \frac{1}{T_A^2}\sum_{k=1}^{T_A}\left(\xi_k \eta_{k-1} x_{k-1} - (\tilde\lambda_{k-1} - \lambda)\xi_k x_{k-1} - (\tilde\lambda_{k-1} - \lambda)\eta_{k-1} x_{k-1}^2\right)\chi(N' \le T_A \le N'') +$$

$$+A^{1/2} E_\theta \frac{1}{T_A^2}\sum_{k=1}^{T_A}(\xi_k^2 + \eta_{k-1}^2 x_{k-1}^2)\chi(N' \le T_A \le N''). \tag{50}$$

By the definitions of N' and N'', the Cauchy-Schwarz-Bunyakovsky inequality and Lemma 1, for the first summand one gets

$$A^{1/2} E_\theta \frac{1}{T_A^2\sigma}\sum_{k=1}^{T_A}(\lambda - \tilde\lambda_{k-1})^2 x_{k-1}^2 \chi(N' \le T_A \le N'') \le CA^{-1/2}\sum_{k=1}^{N''} E_\theta(\lambda - \tilde\lambda_{k-1})^2 x_{k-1}^2 \le CA^{-1/2}\log^2 A \xrightarrow[A\to\infty]{} 0.$$

For the second summand of (50) the Doob's maximal inequality for martingales (see, e.g., [3]) and the Cauchy-Schwarz-Bunyakovsky inequality can be used to show

$$A^{1/2}E_\theta \frac{1}{T_A^2}\sum_{k=1}^{T_A}\left(\xi_k\eta_{k-1}x_{k-1}-(\tilde{\lambda}_{k-1}-\lambda)\xi_k x_{k-1}-(\tilde{\lambda}_{k-1}-\lambda)\eta_{k-1}x_{k-1}^2\right)\chi(N'\le T_A\le N'')\le CA^{-1/4}\xrightarrow[A\to\infty]{}0.$$

Rewrite the left-hand side of the last summand of (50)

$$A^{1/2}E_\theta \frac{1}{T_A^2\sigma}\sum_{k=1}^{T_A}(\xi_k^2+\eta_{k-1}^2x_{k-1}^2)\chi(N'\le T_A\le N'')=$$

$$=A^{1/2}E_\theta \frac{1}{T_A\sigma}m_{T_A}\chi(N'\le T_A\le N'')+A^{1/2}\sigma E_\theta \frac{1}{T_A}\chi(N'\le T_A\le N'').$$

We show that the first summand converges to 0 and the second one converges to 1. By (13) the Doob's maximal inequality and the Cauchy-Schwarz-Bunyakovsky inequality

$$A^{1/2}E_\theta \frac{1}{T_A\sigma}\left|m_{T_A}\right|\chi(N'\le T_A\le N'')\le CA^{1/2}\frac{1}{(N')^2}\left(E_\theta \max_{1\le n\le N''}\left(\sum_{k=1}^{n}(\xi_k^2+\eta_{k-1}^2x_{k-1}^2-(\sigma_\xi^2+\sigma_\eta^2\sigma_{x,k-1}^2))\right)^2\right)^{1/2}\le$$

$$\le CA^{-1/2}\left(\sum_{k=1}^{N''}E_\theta(\xi_k^2+\eta_{k-1}^2x_{k-1}^2-(\sigma_\xi^2+\sigma_\eta^2\sigma_{x,k-1}^2))^2\right)^{1/2}\le CA^{-1/4}\xrightarrow[A\to\infty]{}0.$$

The almost sure convergence of the second summand to 1 follows from (27), boundedness of the family $\left\{A^{1/2}\frac{1}{T_A}\chi(N'\le T_A\le N'')\right\}_{A\ge1}$ and bounded convergence theorem.

The assertion (29) follows from (48) and (49).

Summary

The problem of building optimal predictions for the values of scalar stable autoregressive process with parameter drift is considered. The predictors are constructed on the basis of the truncated estimators, which are shown to have prescribed mean-square accuracy on samples of fixed size. The optimal sample size is establish both theoretically and as a stopping time defined on observational data. Asymptotic equivalence of the two is proved in the sense of almost sure convergence and convergence in mean, as well as asymptotic equivalence of the corresponding risk functions.

Acknowledgements

The author wishes to thank his scientific adviser professor Vasiliev V.A., who provided valuable comments and ideas regarding the research theme.

REFERENCES

1. *Weyer E., Campi M.* Non-Asymptotic Confidence Ellipsoids for the Least-Squares Estimate // *Automatica*. 2002. V. 38, Is. 9. P. 1539–1547.
2. *Weyer E., Campi M.* Guaranteed Non-Asymptotic Confidence Regions in System Identification // *Automatica*. 2005. V. 41, Is. 10. P. 1751–1764.
3. *Liptser R., Shiryaev A.* Statistics of Random Processes. N.Y. : Springer, 1977.
4. *Konev V., Vorobeichikov S.* On Sequential Identification of Stochastic Systems // Izvestia of USSR Academy of Sciences, Technical Cybernetics. 1980. Is. 4. P. 176–182.
5. *Konev V.* Sequential Parameter Estimation of Stochastic Dynamical Systems. Tomsk University Press, Tomsk, 1985.
6. *Vasiliev V., Konev V.* The Sequential Parameter Identification of the Dynamic Systems in the Presence of Multiplicative and Additive Noises in Observations // Automation and Remote Control. 1985. V. 46. P. 706–716.
7. *Konev V., Pergamenshchikov S.* On the Duration of Sequential Estimation of Parameters of Stochastic Processes in Discrete Time // Stochastics. 1986. V. 18, Is. 2. P. 133–154.
8. *Galtchouk L., Konev V.* On Sequential Estimation of Parameters in Semimartingale Regression Models with Continuous Time Parameter // Annals of Statistics. 2001. V. 29, Is. 5. P. 1508–1536.
9. *Küchler U., Vasiliev V.* On Sequential Parameter Estimation for Some Linear Stochastic Differential Equations with Time Delay // Sequential Analysis. 2001. V. 20, Is. 3. P. 117–146.

10. *Küchler U., Vasiliev V.* On Guaranteed Parameter Estimation of a Multiparameter Linear Regression Process // Automatica. 2010. V. 46, Is. 4. P. 637–646.
11. *Malyarenko A., Vasiliev V.* On Parameter Estimation of Partly Observed Bilinear Discrete-Time Stochastic Systems // Metrika. 2012. V. 75, Is. 3. P. 403–424.
12. *Politis D., Vasiliev V.* Sequential Kernel Estimation of a Multivariate Regression Function // Proceedings of IX International Conference 'System Identification and Control Problems'. Moscow, 2012. P. 996–1009.
13. *Konev V., Pergamenshchikov S.* Truncated Sequential Estimation of the Parameters in Random Regression // Sequential Analysis. 1990. V. 9, Is. 1. P. 19–41.
14. *Konev V., Pergamenshchikov S.* On Truncated Sequential Estimation of the Drifting Parametermean in the First Order Autoregressive Models // Sequential Analysis. 1990. V. 9, Is. 2. P. 193–216.
15. *Fourdrinier D., Konev V., Pergamenshchikov S.* Truncated Sequential Estimation of the Parameter of a First Order Autoregressive Process with Dependent Noises // Mathematical Methods of Statistics. 2008. V. 18, Is. 1. P. 43–58.
16. *Vasiliev V.* A Truncated Estimation Method with Guaranteed Accuracy // Annals of Institute of Statistical Mathematics. 2014. V. 66. P. 141–163.
17. *Sriram T.* Sequential Estimation of the Autoregressive Parameter in a First Order Autoregressive Process // Sequential Analysis. 1988. V. 7, Is. 1. P. 53–74.
18. *Konev V., Lai T.* Estimators with Prescribed Precision in Stochastic Regression Models // Sequential Analysis. 1995. V. 14, Is. 3. P. 179–192.
19. *Vasiliev V., Kusainov M.* Asymptotic Risk-Efficiency of One-Step Predictors of a Stable AR(1) // Proceedings of XII All-Russian Conference on Control Problems. Moscow, 2014. P. 2619–2627.
20. *Kusainov M., Vasiliev V.* On Optimal Adaptive Prediction of Multivariate Autoregression // Sequential Analysis. 2015. V. 34, Is. 2. 23 p.
21. *Kusainov M.* On Optimal Adaptive Prediction of Multivariate ARMA(1,1) Process // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2015. № 1(30). P. 44–57.
22. *Malyarenko A.* Estimating the Generalized Autoregression Model Parameters for Unknown Noise Distribution // Automation and Remote Control. 2012. V. 71, Is. 2. P. 291–302.

Kusainov Marat Islambekovich. E-mail: rjrltsk@gmail.com
Tomsk State University, Tomsk, Russian Federation

Поступила в редакцию 22 апреля 2015 г.

Кусаинов М.И. (Томский государственный университет. Россия).

Риск-эффективность адаптивных одношаговых прогнозов авторегрессии с шумящим параметром.

Ключевые слова: адаптивные прогнозы; асимптотическая риск-эффективность; оптимальный размер выборки; скалярная авторегрессия; момент остановки; усечённое оценивание.

DOI 10.17223/19988605/32/4

Изучается скалярный устойчивый процесс авторегрессии с аддитивным шумом в параметре динамики. Параметры модели предполагаются неизвестными. Усечённые оценки параметров используются для построения адаптивных одношаговых прогнозов. Сопутствующая проблема заключается в минимизации функции риска специального вида, описывающей размер выборки и выборочное среднее квадрата ошибки прогноза. Последовательная процедура вводится для достижения минимального риска.

REFERENCES

1. Weyer, E. & Campi, M. (2002) Non-Asymptotic Confidence Ellipsoids for the Least-Squares Estimate. *Automatica.* 38(9). pp. 1539-1547. DOI: 10.1109/CDC.2000.914211
2. Weyer, E. & Campi, M. (2005) Guaranteed Non-Asymptotic Confidence Regions in System Identification. *Automatica.* 41 (10). pp. 1751-1764.
3. Liptser, R. & Shiryaev, A. (1977) *Statistics of Random Processes*. New York: Springer.
4. Konev, V. & Vorobeichikov, S. (1980) On Sequential Identification of Stochastic Systems. *Izvestia of USSR Academy of Sciences, Technical Cybernetics.* 4. pp. 176-182.
5. Konev, V. (1985) *Sequential Parameter Estimation of Stochastic Dynamical Systems*. Tomsk: Tomsk State University Press.
6. Vasiliev, V. & Konev, V. (1985) The Sequential Parameter Identification of the Dynamic Systems in the Presence of Multiplicative and Additive Noises in Observations. *Automation and Remote Control.* 46. pp. 706-716. (In Russian).
7. Konev, V. & Pergamenshchikov, S. (1986) On the Duration of Sequential Estimation of Parameters of Stochastic Processes in Discrete Time. *Stochastics.* 18 (2). pp. 133-154. DOI: 10.1080/17442508608833405
8. Galtchouk, L. & Konev, V. (2001) On sequential estimation of parameters in semimartingale regression models with continuous time parameter. *Annals of Statistics.* 29 (5). pp. 1508-1536. DOI: 10.1214/aos/1013203463
9. Küchler, U. & Vasiliev, V. (2001) On Sequential Parameter Estimation for Some Linear Stochastic Differential Equations with Time Delay. *Sequential Analysis.* 20 (3). pp. 117-146. DOI: 10.1081/SQA-100106052

10. Küchler, U. & Vasiliev, V. (2001) On Guaranteed Parameter Estimation of a Multiparameter Linear Regression Process. *Automatica*. 46 (4). pp. 637-646. DOI: 10.1016/j.automatica.2010.01.003
11. Malyarenko, A. & Vasiliev, V. (2012) On Parameter Estimation of Partly Observed Bilinear Discrete-Time Stochastic Systems. *Metrika*. 75 (3). pp. 403-424. DOI: 10.1007/s00184-010-0333-5
12. Politis, D. & Vasiliev, V. (2012) Sequential Kernel Estimation of a Multivariate Regression Function. *Proceedings of IX International Conference 'System Identification and Control Problems'*. Moscow. pp. 996-1009.
13. Konev, V. & Pergamenshchikov, S. (1990) Truncated Sequential Estimation of the Parameters in Random Regression. *Sequential Analysis*. 9 (1). pp. 19-41. DOI: 10.1080/07474949008836194
14. Konev, V. & Pergamenshchikov, S. (1990) On Truncated Sequential Estimation of the Drifting Parametermean in the First Order Autoregressive Models. *Sequential Analysis*. 9 (2). pp. 193-216. DOI: 10.1080/07474949008836205
15. Fourdrinier, D., Konev, V. & Pergamenshchikov, S. (2008) Truncated Sequential Estimation of the Parameter of a First Order Autoregressive Process with Dependent Noises. *Mathematical Methods of Statistics*. 18 (1). pp. 43-58. DOI: 10.3103/S1066530709010037
16. Vasiliev, V. (2014) A Truncated Estimation Method with Guaranteed Accuracy. *Annals of Institute of Statistical Mathematics*. 66. pp. 141-163. DOI: 10.1007/s10463-013-0409-x.
17. Sriram, T.(1988) Sequential Estimation of the Autoregressive Parameter in a First Order Autoregressive Process. *Sequential Analysis*. 7 (1). pp. 53-74. DOI: 10.1080/07474948808836142
18. Konev, V. & Lai, T. (1995) Estimators with Prescribed Precision in Stochastic Regression Models. *Sequential Analysis*. 14 (3). pp. 179-192. DOI: 10.1080/07474949508836330
19. Vasiliev, V. & Kusainov, M. (2004) Asymptotic Risk-Efficiency of One-Step Predictors of a Stable AR(1). *Proceedings of XII All-Russian Conference on Control Problems*. Moscow. pp. 2619-2627. (In Russian).
20. Kusainov, M. & Vasiliev, V. (2015) On Optimal Adaptive Prediction of Multivariate Autoregression. *Sequential Analysis*. 34 (2). pp. 211-234. DOI: 10.1080/07474946.2015.1030977
21. Kusainov, M. (2015) On Optimal Adaptive Prediction of Multivariate ARMA (1,1) Process. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitel'naya tekhnika i informatika – Tomsk State University Journal of Control and Computer Science*. 1(30). pp. 44-57. (In Russian).
22. Malyarenko, A. (2012) Estimating the Generalized Autoregression Model Parameters for Unknown Noise Distribution. *Automation and Remote Control*. 71 (2). pp. 291-302.