Parameter estimation with guaranteed accuracy for AR(1) by noised observations

Yulia B. Burkatovskaya[1,2] and Vyacheslav A. Vasiliev[1]
[1] *Tomsk State University, Tomsk, Russian Federation*
[2] *Tomsk Polytechnic University, Tomsk, Russian Federation*
e-mail: tracey@tpu.ru, vas@mail.tsu.ru

Abstract

This paper presents a truncated estimator of the dynamic parameter of a stable AR(1) process by observations with additive noise. The estimator is constructed by a sample of a fixed size and it has a known upper bound of the mean square deviation. Cases of known and unknown variance of observation noise are considered.

Keywords: autoregressive process, fixed sample size, guaranteed accuracy, observations with noise.

1 Introduction and problem statement

Development of parameter estimation methods of dynamic systems by samples of finite or fixed size is very important in statistical problems such that model construction and various adaptive problems (prediction, control, filtration etc.).

One of the possibilities for finding estimators with the guaranteed quality of inference using a sample of fixed size is provided by the approach of truncated estimation. Truncated estimators were constructed in [9] for ratio type multivariate functionals by a fixed-size sample. They have guaranteed accuracy in the sense of the L_{2m}-norm, $m \geq 1$. This fact allows one to obtain desired non-asymptotic and asymptotic properties of the estimators. The truncated estimation method was developed in [1] and others for parameter estimation problems in discrete-time dynamic models. Solutions of some non-asymptotic parametric and non-parametric problems can be found also in [4], [8], [5], [6], among others. In particular, [8] established the minimax optimality of the least-squares estimator of the dynamic parameter in AR(1) model.

In this paper, the truncated estimation method introduced in [9] is applied for the parameter estimation of AR(1) by additively-noised observations with unknown noise variance (another applications of this method can be found, e.g., in [2], [3]).

Consider the estimation problem of the parameter λ of the scalar first-order autoregressive process $(x_n)_{n\geq 0}$ satisfying the equation

$$x_n = \lambda x_{n-1} + \xi_n, \quad n \geq 1 \tag{1}$$

by observations

$$y_n = x_n + \eta_n, \quad n \geq 0. \tag{2}$$

Process (1) is supposed to be stable, i.e. $|\lambda| < 1$. Introduce the notation $\zeta = (x_0, \xi_1, \eta_0.)$ The processes (ξ_n), (η_n) and x_0 are supposed to be mutually independent;

noises ξ_n and η_n form sequences of i.i.d. random variables such that $E\zeta = 0, E||\zeta||^4 < \infty$. Denote $\sigma^2 = E\eta_0^2$. We assume that the variance of ξ_1 is known. Then without loss of generality we put $E\xi_1^2 = 1$.

The main aim of the paper is to construct truncated estimators of $\lambda \in (-1, 1)$ with guaranteed accuracy in the mean square sense by sample of fixed size. Cases of both known and unknown values of σ^2 will be considered.

A similar problem has been solved in, e.g., [10] on the basis of the sequential approach (when the sample size is a random value determined by a special stopping rule) for $\lambda \in (-1, 0) \cap (0, 1)$

2 Parameter estimation of AR(1) with known noise variance

To estimate the parameter λ, we use the correlation method. To this end, we obtain from the system (1), (2) the recurrent equation for the observed process $y = (y_n)_{n\geq 0}$:

$$\begin{aligned} y_n &= \lambda y_{n-1} + \delta_n, \quad n \geq 1, \\ \delta_n &= \xi_n + \eta_n - \lambda\eta_{n-1}. \end{aligned} \tag{3}$$

Due to the dependence of noises δ_n, the least squares estimator (LSE) of λ obtained from equation (3) is asymptotically biased, see, e.g., [7], [10]. Equation (3) implies the following formula for correlations of the process (y_n):

$$E_\lambda y_n y_{n-1} = \lambda E_\lambda (y_{n-1}^2 - \sigma^2), \quad n \geq 1.$$

Hence, the consistent correlation estimator $\hat{\lambda}_n$ of λ has the following form (see [7])

$$\hat{\lambda}_{n,\sigma} = \frac{\sum\limits_{k=1}^{n} y_k y_{k-1}}{\sum\limits_{k=1}^{n} (y_{k-1}^2 - \sigma^2)}, \quad n \geq 1. \tag{4}$$

It is easy to verify that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} (y_{k-1}^2 - \sigma^2) = \frac{1}{1-\lambda^2} > 1 \quad P_\lambda - a.s. \tag{5}$$

Thus, according to the general procedure described in [9], it is reasonable to construct the truncated estimator $\tilde{\lambda}_n$ of λ as follows:

$$\tilde{\lambda}_n = \hat{\lambda}_n \cdot \chi(\sum_{k=1}^{n} (y_{k-1}^2 - \sigma^2) \geq hn), \quad n \geq 1, \tag{6}$$

where $h \in (0, 1)$ and $\chi(A)$ is the indicator of the set A.

The following theorem gives the first main result of this paper.

Theorem 1. *Assume model (1), (2). Then for every $|\lambda| < 1$ and $n \geq 1$, estimator (7) has the property*

$$E_\lambda(\tilde{\lambda}_n - \lambda)^2 \leq \frac{C}{n}. \tag{7}$$

The proofs of theorems and lemmas are given in Section 5.

3 Parameter estimation of AR(1) with unknown noise variance

To estimate $\lambda \in (-1, 1)$, we use an adaptive modification of estimator (5):

$$\lambda_n^* = \frac{\frac{1}{n}\sum_{k=1}^{n} y_k y_{k-1}}{\frac{1}{n}\sum_{k=1}^{n} y_{k-1}^2 - \sigma_n^2}, \quad n > 1. \tag{8}$$

Taking into account (6), we construct the estimator σ_n^2 of σ^2 as follows

$$\sigma_n^2 = \frac{1}{n}\sum_{k=1}^{n} y_{k-1}^2 - \frac{1}{1 - \lambda_n^2}, \quad n > 1 \tag{9}$$

where λ_n is the pilot estimator of λ

$$\lambda_n = \text{proj}_{[-1,1]}\breve{\lambda}_n, \quad n > 1, \tag{10}$$

$$\breve{\lambda}_n = \frac{\sum_{k=2}^{n} y_k y_{k-2}}{\sum_{k=2}^{n} y_{k-1} y_{k-2}} \cdot \chi(|\sum_{k=2}^{n} y_{k-1} y_{k-2}| \geq H_n), \quad n > 1. \tag{11}$$

Here we put $H_n = n(\log n)^{-1}$. According to the general truncated estimation method [9], the multiplier $(\log n)^{-1}$ in the definition of H_n can be any other slowly-decreasing function.

It should be noted that the estimator (10) is constructed on the bases of the correlation (Yule-Walker type) estimator which can not be used if $\lambda = 0$ (see Lemma 1 below). Our main aim is to construct an estimator of λ without this restriction.

Taking into account (10), estimator (9) can be written in the form

$$\lambda_n^* = (1 - \lambda_n^2)\frac{1}{n}\sum_{k=1}^{n} y_k y_{k-1}, \quad n > 1. \tag{12}$$

Lemma 1. *Assume that in model (1), (2), $E||\zeta||^8 < \infty$. Then estimator (10) for every $\lambda \in (-1, 0) \cup (0, 1)$ and $n > 1$ has the following property*

$$E_\lambda(\lambda_n - \lambda)^2 \leq \frac{C_1}{n} + C_2 \frac{\log^4 n}{n^2}.$$

This lemma makes possible to obtain the main result of the section.

Theorem 2. *Assume that in model (1), (2), $E||\zeta||^8 < \infty$. Then for every $|\lambda| < 1$ and $n > 1$, estimator (12) satisfies the following condition*

$$E_\lambda(\lambda_n^* - \lambda)^2 \leq \frac{C}{n} + C \frac{\log^4 n}{n^2}.$$

4 Simulation Results and Discussion

We conducted numerical simulation of the proposed estimation algorithm. For every set of the parameters, the experiment was performed 100 times, the number of observations is equal to 100, the parameter of the procedure $h = 0,5$. Table 1 presents the results of simulation. Here λ and σ are the parameters of model (1), $\tilde{\lambda}_n$ and λ_n^* are the mean estimators of the parameter λ when the noise variance σ^2 is supposed to be known and unknown, correspondingly; $\tilde{d}_n$ and d_n^* are sample standard errors of the corresponding estimators.

One can see that $\tilde{d}_n < d_n^*$ in all experiments; thus, if the noise variance is unknown then the standard error increases at least twice (if $\lambda = 0,5$); but d_n^* can be fully ten times larger than $\tilde{d}_n$, if $\lambda = 0,9$. Both deviations increase with the grow of σ^2, as one should expect; besides, $\tilde{d}_n$ decreases and d_n^* increases with the increase of λ.

5 Proofs

5.1 *Proof of Theorem 1*

To investigate the non-asymptotic properties of $\tilde{\lambda}_n$ we use the following representation of the deviation

$$\tilde{\lambda}_n - \lambda = \frac{f_n}{g_n} \cdot \chi(|g_n| \geq h) - \lambda \chi(|g_n| < h), \tag{13}$$

where

$$f_n = \frac{1}{n} \sum_{k=1}^{n} [y_{k-1}(\xi_k + \eta_k) - \lambda(y_{k-1}\eta_{k-1} - \sigma^2)],$$

$$g_n = \frac{1}{n} \sum_{k=1}^{n} (y_{k-1}^2 - \sigma^2).$$

It can be directly verified that for $|\lambda| < 1$

$$E_\lambda f_n^2 \leq \frac{I^{-1}(\lambda, \overline{\sigma})}{n}. \tag{14}$$

Table 1: Simulation results

λ	σ^2	$\tilde{\lambda}_n$	$\tilde{d}_n$	λ_n^*	d_n^*
0,5	0,09	0,477	0,0092	0,452	0,0294
0,5	0,25	0,492	0,0111	0,490	0,0314
0,5	0,49	0,487	0,0150	0,470	0,0488
0,5	0,81	0,475	0,0229	0,418	0,0795
0,5	1	0,473	0,0419	0,424	0,0953
0,8	0,09	0,786	0,0046	0,796	0,0465
0,8	0,25	0,794	0,0054	0,854	0,0793
0,8	0,49	0,786	0,0054	0,789	0,0737
0,8	0,81	0,772	0,0120	0,765	0,1435
0,8	1	0,788	0,0122	0,797	0,1590
0,9	0,09	0,876	0,0038	0,865	0,0772
0,9	0,25	0,889	0,0018	0,913	0,0596
0,9	0,49	0,888	0,0030	0,910	0,1044
0,9	0,81	0,874	0,0044	0,886	0,1822
0,9	1	0,891	0,0028	0,891	0,1780

Introduce the notation $g = 1/(1-\lambda^2)$. Then, using a representation

$$g_n - g = \frac{1}{n}\sum_{k=1}^{n}(x_{k-1}^2 - \sigma^2) + \frac{2}{n}\sum_{k=1}^{n} x_{k-1}\eta_{k-1} + \frac{1}{n}\sum_{k=1}^{n}(\eta_{k-1}^2 - \sigma^2)$$

and the following formula (see, e.g., the proof of Theorem 2 in [9])

$$\frac{1}{n}\sum_{k=1}^{n}(x_{k-1}^2 - g) = \frac{g}{n}\cdot[x_0^2 - x_n^2 + 2\lambda\sum_{k=1}^{n} x_{k-1}\xi_k + \sum_{k=1}^{n}(\xi_k^2 - 1)],$$

it is easy to prove that

$$E_\lambda(g_n - g)^2 \le \frac{C_0}{n}, \qquad n \ge 1. \tag{15}$$

Further, similar to [9] using the Chebyshev inequality we estimate

$$P_\lambda(|g_n| < h) \le P_\lambda(|g_n - g| > g - h) \le \frac{E_\lambda(g_n - g)^2}{(g-h)^2} \le \frac{C_0}{(1-h)^2 n}, \qquad n \ge 1. \tag{16}$$

Using (13–16), we estimate

$$E_\lambda(\tilde{\lambda} - \lambda)^2 \le \frac{1}{h^2}E_\lambda f_n^2 + P_\lambda(|g_n| < h) \le \frac{I^{-1}(\lambda, \overline{\sigma})}{h^2 n} + \frac{E_\lambda(g_n - g)^2}{(g-h)^2} \le \frac{C}{n}$$

and obtain assertion (7).

5.2 *Proof of Lemma 1*

The proof of Lemma 1 is similar to the proof of the second assertion of Theorem 1 in [9].

Definition (10) of λ_n implies

$$E_\lambda(\lambda_n - \lambda)^2 \le E_\lambda(\breve{\lambda}_n - \lambda)^2.$$

Introduce the following notations

$$f_n = \frac{1}{n}\sum_{k=2}^{n} y_{k-2}\delta_k, \quad g_n = \frac{1}{n}\sum_{k=1}^{n} y_{k-1}y_{k-2}, \quad g = \frac{\lambda}{1-\lambda^2}, \quad h_n = (\log n)^{-1}.$$

By the definition of $\breve{\lambda}$ in (11), its deviation has the form

$$\breve{\lambda}_n - \lambda = \frac{f_n}{g_n}\cdot\chi(|g_n| \ge h_n) - \lambda\cdot\chi(|g_n| < h_n) = \frac{f_n}{g}\cdot\chi(|g_n| \ge h_n) + \frac{f_n(g-g_n)}{gg_n}\cdot\chi(|g_n| \ge h_n) - \lambda\cdot\chi(|g_n| < h_n) = J_1 + J_2 + J_3.$$

Using the Cauchy-Schwarz-Bunyakovsky and Chebyshev's inequalities, estimate the second moments of these summands:

$$E_\lambda J_1^2 \le C E_\lambda f_n^2, \quad E_\lambda J_2^2 \le \frac{1}{g^2 h_n^2}\sqrt{E_\lambda f_n^4 E_\lambda (g_n - g)^4}, \quad E_\lambda J_3^2 \le h_n^{-4} E_\lambda (g_n - g)^4.$$

In view of the structure of the function f_n it is easy to verify that $E_\lambda f_n^4 \le C/n^2$.
By the definition of g_n we have

$$g_n - g = \frac{1}{n}\sum_{k=1}^{n} y_{k-1}y_{k-2} - \frac{\lambda}{1-\lambda^2} = \frac{1}{n}\sum_{k=1}^{n} x_{k-1}x_{k-2} - \frac{\lambda}{1-\lambda^2} + \frac{1}{n}\sum_{k=1}^{n} x_{k-1}\eta_{k-2} + \frac{1}{n}\sum_{k=1}^{n} \eta_{k-1}x_{k-2} + \frac{1}{n}\sum_{k=1}^{n}\eta_{k-1}\eta_{k-2} = \lambda\left(\frac{1}{n}\sum_{k=1}^{n} x_{k-2}^2 - \frac{1}{1-\lambda^2}\right) + \frac{1}{n}\sum_{k=1}^{n} x_{k-1}\eta_{k-2} + \frac{1}{n}\sum_{k=1}^{n}(\eta_{k-1}+\xi_{k-1})x_{k-2} + \frac{1}{n}\sum_{k=1}^{n}\eta_{k-1}\eta_{k-2}.$$

Using this representation, it is easy to verify similarly the proof of Theorem 1 that

$$E_\lambda(g_n - g)^4 \le \frac{C}{n^2}.$$

Thus we have

$$E_\lambda J_1^2 \le C\frac{1}{n}, \quad E_\lambda J_2^2 \le C\frac{\log^2 n}{n^2}, \quad E_\lambda J_3^2 \le C\frac{\log^4 n}{n^2}.$$

5.3 *Proof of Theorem 2*

Introduce the following notations

$$\Delta_n = (\lambda^2 - \lambda_n^2)\frac{\lambda}{1-\lambda^2} + (1-\lambda_n^2)\lambda\frac{1}{1-\lambda^2}\left\{\frac{1}{n}(x_0^2 - \lambda^2 x_{k-1}^2) + \frac{2\lambda}{n}\sum_{k=2}^{n} x_{k-2}\xi_{k-1}\right.$$
$$\left. + \frac{1}{n}\sum_{k=2}^{n}(\xi_{k-1}^2 - 1) - \frac{1}{n}\right\} + (1-\lambda_n^2)\frac{1}{n}\sum_{k=1}^{n}[\lambda x_{k-1}\eta_{k-1} + y_{k-1}(\xi_k + \eta_k)].$$

Definition (12) of the estimator λ_n^* and equation (3) imply

$$\lambda_n^* = (1-\lambda_n^2)\frac{1}{n}\sum_{k=1}^{n}[\lambda y_{k-1}^2 + y_{k-1}(\xi_k+\eta_k) - \lambda y_{k-1}\eta_{k-1}]$$
$$= (1-\lambda_n^2)\left\{\lambda\frac{1}{n}\sum_{k=1}^{n} x_{k-1}^2 + \frac{1}{n}\sum_{k=1}^{n}[\lambda x_{k-1}\eta_{k-1} + y_{k-1}(\xi_k+\eta_k)]\right\}$$
$$= (1-\lambda_n^2)\frac{\lambda}{1-\lambda^2} + (1-\lambda_n^2)\lambda\left[\frac{1}{n}\sum_{k=1}^{n} x_{k-1}^2 - \frac{1}{1-\lambda^2}\right]$$
$$+(1-\lambda_n^2)\frac{1}{n}\sum_{k=1}^{n}[\lambda x_{k-1}\eta_{k-1} + y_{k-1}(\xi_k+\eta_k)] = \lambda + (\lambda^2 - \lambda_n^2)\frac{\lambda}{1-\lambda^2}$$
$$+(1-\lambda_n^2)\lambda\frac{1}{1-\lambda^2}\left\{\frac{1}{n}(x_0^2 - \lambda^2 x_{k-1}^2) + \frac{2\lambda}{n}\sum_{k=2}^{n} x_{k-2}\xi_{k-1} + \frac{1}{n}\sum_{k=2}^{n}(\xi_{k-1}^2 - 1) - \frac{1}{n}\right\}$$
$$+(1-\lambda_n^2)\frac{1}{n}\sum_{k=1}^{n}[\lambda x_{k-1}\eta_{k-1} + y_{k-1}(\xi_k+\eta_k)] = \lambda + \Delta_n.$$

Thus the mean square deviation of the estimator λ_n^* has the following form

$$E_\lambda(\lambda_n^* - \lambda)^2 = E_\lambda\Delta_n^2 \cdot \chi(\lambda = 0) + E_\lambda\Delta_n^2 \cdot \chi(\lambda \neq 0) =: I_1 + I_2,$$

where

$$I_1 = E_\lambda((1-\lambda_n^2)\frac{1}{n}\sum_{k=1}^{n}[y_{k-1}(\xi_k+\eta_k)])^2 \cdot \chi(\lambda = 0)$$
$$= E_\lambda((1-\lambda_n^2)\frac{1}{n}\sum_{k=1}^{n}(\xi_{k-1}+\eta_{k-1})(\xi_k+\eta_k))^2 \cdot \chi(\lambda = 0),$$
$$I_2 = E_\lambda\Delta_n^2 \cdot \chi(\lambda \neq 0).$$

From assumptions of Theorem 2 it follows $I_1 \leq C/n$. In view of Lemma 1 and the property $|\lambda_n + \lambda| \leq 2$, we have

$$I_2 \leq E_\lambda\left(\frac{2|\lambda_n - \lambda|}{1-\lambda^2}\chi(\lambda \neq 0) + \frac{1}{1-\lambda^2}\left\{\frac{1}{n}(x_0^2 + x_{k-1}^2) + \frac{2}{n}\left|\sum_{k=2}^{n} x_{k-2}\xi_{k-1}\right|\right.\right.$$
$$\left.\left. + \frac{1}{n}\left|\sum_{k=2}^{n}(\xi_{k-1}^2 - 1)\right| + \frac{1}{n}\right\} + \frac{1}{n}\left|\sum_{k=1}^{n} x_{k-1}\eta_{k-1}\right| + \frac{1}{n}\left|\sum_{k=1}^{n} y_{k-1}(\xi_k+\eta_k)\right|\right)^2$$
$$\leq CE_\lambda(\lambda_n - \lambda)^2\chi(\lambda \neq 0) + \frac{C}{n} + \frac{C}{n^2} \leq \frac{C}{n} + C\frac{\log^4 n}{n^2}.$$

Acknowledgment

This study is supported by the RSF (project 17-11-01049).

References

[1] Dogadova, T. V., and V. A. Vasiliev. 2014. Guaranteed parameter estimation of stochastic linear regression by sample of fixed size. *Tomsk State University Journal of Control and Computer Science.* Vol. **26(1)**, pp. 39–52.

[2] Tatiana V. Dogadova, Marat I. Kusainov and Vyacheslav V. Vasiliev. 2017. *Truncated estimation method and applications. Serdica. Mathematical Journal Bulgarian Academy of Sciences Institute of Mathematics and Informatics.* Vol. **43**, pp. 221–266.

[3] Kusainov, M. I., and V. A. Vasiliev. 2015. On optimal adaptive prediction of multivariate autoregression. *Sequential Analysis: Design Methods and Applications.* Vol. **35(2)**, pp. 211–234.

[4] Mikulski, P. W., and M. J. Monsour. 1991. Optimality of the maximum likelihood estimator in first-order autoregressive processes. *Journal of Time Series Analysis.* Vol. **12(3)**, pp. 237–253.

[5] Roll, J., A. Nazin, and L. Ljung. 2002. A Non-Asymptotic Approach to Local Modelling. *The 41st IEEE CDC, Las Vegas, Nevada, USA, 10-13 Dec. 2002 (Regular paper).* Predocumentation is available at http://www.control.isy.liu.se/research/reports/2003/2482.pdf

[6] Roll J, Nazin A, Ljung L. 2005. Non-linear System Identification Via Direct Weight Optimization. *Automatica, Journal of IFAC. Special Issue on "Data-Based Modelling and System Identification".* Vol. **41(3)**, pp. 475–490. Predocumentation is available at http://www.control.isy.liu.se/research/reports/2005/2696.pdf

[7] Shneeweiss, H. 1976. Consistent estimation of a regression with error in the variables. *Metrika.* Physica-Verlag, Wien. Vol. **23**, pp. 101–115.

[8] Shiryaev, A. N., and V. G. Spokoiny. 2000. *Statistical Experiments and Decisions. Asymptotic Theory.* Singapore, New Jersey, London, Hong Kong: World Scientific.

[9] Vasiliev, V. 2014. A Truncated Estimation Method with Guaranteed Accuracy. *Annals of Institute of Statistical Mathematics.* Vol. **66**, pp. 141–163.

[10] Vasiliev V. A., and V. V. Konev 1982. The Sequential Parameters Estimation of Dynamic Systems with Noisy Observations. *Izvestia Acad. USSR, Technical Cybernetics.* Vol. **6**, pp. 145–154.