Ann Inst Stat Math (2014) 66:141–163
DOI 10.1007/s10463-013-0409-x

A truncated estimation method with guaranteed accuracy

Vyacheslav A. Vasiliev

Received: 20 July 2012 / Revised: 13 February 2013 / Published online: 1 June 2013

Abstract This paper presents a truncated estimation method of ratio type functionals by dependent sample of finite size. This method makes it possible to obtain estimators with guaranteed accuracy in the sense of the L_m-norm, $m \geq 2$. As an illustration, the parametric and non-parametric estimation problems on a time interval of a fixed length are considered. In particular, parameters of linear (autoregressive) and non-linear discrete-time processes are estimated. Moreover, the parameter estimation problem of non-Gaussian Ornstein-Uhlenbeck process by discrete-time observations and the estimation problem of a multivariate logarithmic derivative of a noise density of an autoregressive process with guaranteed accuracy are solved. In addition to non-asymptotic properties, the limit behavior of presented estimators is investigated. It is shown that all the truncated estimators have asymptotic properties of basic estimators. In particular, the asymptotic efficiency in the mean square sense of the truncated estimator of the dynamic parameter of a stable autoregressive process is established.

Keywords Ratio estimation · Truncated estimation method · Fixed sample size · Multivariate autoregression · AR-ARCH model · Non-Gaussian Ornstein-Uhlenbeck process · Non-parametric multivariate logarithmic density derivative estimation

1 Introduction

Modern evolution of mathematical statistics is directed toward development of data processing methods by dependent sample of finite size. One of such possibilities gives a well-known sequential estimation method, which was successfully applied to

V. A. Vasiliev (✉)
Department of Applied Mathematics and Cybernetics, Tomsk State University, Lenina 36, 634050 Tomsk, Russia
e-mail: vas@mail.tsu.ru

parametric and non-parametric problems. This approach for various statistical problems for a scheme of independent observations has been primarily proposed by Wald (1947). Then this idea has been applied to parameter estimation problem of continuous- and discrete-time dynamic systems in many papers and books (see Dobrovidov et al. 2012; Galtchouk and Konev 2001; Konev 1985; Konev and Pergamenshchikov 1985, 1992; Küchler and Vasiliev 2010; Liptser and Shiryaev 1977; Novikov 1971 among others). Sequential approach has been also applied to non-parametric estimation problems, for example, for estimation of regression, autoregression and density function as well (see, e.g., Arkoun 2011; Arkoun and Pergamenchtchikov 2008; Dobrovidov et al. 2012; Efroimovich 2007).

To obtain sequential estimators with an arbitrary accuracy one needs to have a sample of unbounded size. However, in practice the observation time of a system is usually not only finite but fixed. 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 sequential estimation. The truncated sequential estimation method was developed in Fourdrinier et al. (2009), Konev and Pergamenshchikov (1990a,b) and others for parameter estimation problems in discrete-time dynamic models. Using a sequential approach, estimators of dynamic system parameters with known variance by sample of fixed size were constructed in these papers.

Non-parametric truncated sequential estimators of a regression function by dependent observations were presented by Politis and Vasiliev (2012a,b) on the basis of Nadaraya–Watson estimators calculated at a special stopping time. These estimators have known mean square errors as well. The duration of observations is also random but bounded from above by a non-random fixed number.

Results in non-asymptotic parametric and non-parametric problems can be found in Mikulski and Monsour (1991), Roll et al. (2002, 2005), Shiryaev and Spokoiny (2000) among others.

The main purpose of this paper was to obtain a truncated modification of ratio type estimator from a wide class, having guaranteed accuracy by dependent sample of fixed size. When estimating the ratio type functionals one uses as a rule the substitution statistics, see Borovkov (1997), that is ratio of some estimators. Studying the properties of such estimators, we face certain difficulties that are associated with finding the dominant sequences, see Cramér (1999). In some cases, for instance, in reconstruction of the multivariate logarithmic derivative of a distribution density one can use estimators for which an exact asymptotic expression of the mean square error (MSE) is available (see Dobrovidov et al. 2012; Novak 1996).

The theory of smoothing can also be used for this problem. It makes it possible to find the principal term of the MSE of the ratio estimators with an improved rate of convergence, similar to the case of independent observations. Moreover, the rate of convergence of the estimators of their ratio in metric $L_m,\ m \geq 2$, can be obtained (see, e.g., Dobrovidov et al. 2012; Penskaya 1990).

In this paper the truncated estimation method of ratio type functionals by dependent sample of fixed size is presented. This method allows to obtain estimators with guaranteed accuracy in the sense of the L_m-norm, $m \geq 2$. Examples of parametric and non-parametric estimation problems on a time interval of a fixed length are considered. It is shown that truncated estimators may keep asymptotic properties of basic estimators. In particular, the asymptotic efficiency in the mean square sense is derived for the truncated estimator of the parameter of a stable autoregression. Early similar results for scalar ratio type estimators were published in Vasiliev (2012).

2 Statement of the problem and main result

Let $(\Omega, \mathcal{F}, \mathsf{P})$ be a probability space with a filtration $\{\mathcal{F}_n\}_{n\geq 0}$ and let $(f_n)_{n\geq 1}$ and $(g_n)_{n\geq 1}$ be $\{\mathcal{F}_n\}$–adapted sequences of random $s \times q$ matrices and numbers, respectively.

Let

$$\Psi_N = f_N/g_N, \quad N \geq 1 \tag{1}$$

be an estimator of a matrix Ψ. For instance, the matrix Ψ can be a ratio

$$\Psi = f/g$$

and f_N and g_N are estimators of matrix f and number $g \neq 0$, respectively.

Consider the following modification of the estimator Ψ_N :

$$\tilde{\Psi}_N(H) = \Psi_N \cdot \chi(|g_N| \geq H), \quad N \geq 1, \tag{2}$$

where H is a positive number or sequence $H = (H_N)_{N\geq 1}$, defined below and the notation $\chi(A)$ means the indicator function of set A.

Our main aim was to formulate general conditions on the sequences (f_N), (g_N) and on the parameter H giving a possibility to estimate Ψ with a guaranteed accuracy in the sense of the L_m-norm, $m \geq 2$.

Define for some $\varphi_N(m)$, $w_N(\mu)$, H and g, the function

$$V_N(m, \mu, H) = \frac{1}{H^{2m}}\varphi_N(m) + \frac{||\Psi||^{2m}}{(|g| - H)^{2\mu}} w_N(\mu),$$

as well as for positive integer $p < m, \beta \in (0, 1)$ and positive numbers H_N, the function

$$\begin{aligned}V_N(p) = {}& 2^{p-1} g^{-2p} \varphi_N(p) + 2^{2p-1} g^{-2p} H_N^{-p} \varphi_N^{p/m}(m) w_N^{p/2\mu}(\mu) \\ &+ 4^{p-1} g^{-2p} H_N^{-2p} \varphi_N^{p/m}(m) w_N^{p/\mu}(\mu) + ||\Psi||^{2p} \cdot \left[(\beta|g|)^{-2\mu} w_N(\mu) + \gamma_N\right],\end{aligned}$$

where $\gamma_N = \chi(H_N > (1 - \beta)|g|)(= 0$ for N large enough).

Theorem 1 *Assume for some integers $m \geq 1$ and $\mu \geq 1$ there exist sequences of positive numbers $(\varphi_N(m))_{N\geq 1}$ and $(w_N(\mu))_{N\geq 1}$, decreasing to zero, as well as a number $g \neq 0$ such that for every $N \geq 1$ the following assumptions hold:*

(A1) $E||f_N - \Psi g_N||^{2m} \leq \varphi_N(m)$;

(A2) $E(g_N - g)^{2\mu} \leq w_N(\mu)$.

Then, the estimator $\tilde{\Psi}_N(H)$ *defined in (2) has the following properties:*

(i) *in the case of known number* g *for every* $H \in (0, |g|)$

$$E||\tilde{\Psi}_N(H) - \Psi||^{2m} \leq V_N(m, \mu, H); \tag{3}$$

(ii) *in the case of unknown* g *for every (possibly slowly decreasing to zero) sequence* $H = (H_N)$ *of positive numbers and every positive integer* p, *satisfying*

$$\frac{mp}{m-p} \leq \mu, \quad m > 1 \quad \text{and} \quad \mu > 1,$$

it holds

$$E||\tilde{\Psi}_N(H_N) - \Psi||^{2p} \leq V_N(p). \tag{4}$$

Proof Proof of Theorem 1 is presented in Sect. 5. □

Corollary 1 *Assume that* $\Psi = f/g$ *for some matrix* f, *where* g *is defined in Theorem 1 and, instead of the assumption (A1), for some* $\nu \geq 1$ *there exists sequence* $(v_N(\nu))_{N\geq 1}$ *of non-negative numbers, decreasing to zero, such that*

$$E||f_N - f||^{2\nu} \leq v_N(\nu), \quad N \geq 1.$$

Then the assumption (A1) of Theorem 1 is fulfilled, where the function $\varphi_N(m)$ *should be replaced by the following one:*

$$\varphi_N(m) = \frac{2^{2m-1}}{g^{2m}} \left[g^{2m} v_N^{m/\nu}(m) + ||f||^{2\nu} w_N^{m/\mu}(\nu) \right], \quad m = \min(\nu, \mu).$$

Corollary 2 *If it is known that the matrix* Ψ *belongs to a bounded set* Q, *then the estimator (2) can be taken in the form*

$$\tilde{\Psi}_N^*(H) = \Psi_N \cdot \chi(|g_N| \geq H) + \overline{\Psi}\chi(|g_N| < H), \quad N \geq 1,$$

where $\overline{\Psi} = \operatorname{argmin}_{S\in Q} \sup\limits_{\Psi \in Q} ||S - \Psi||$. *In this case the number* $||\Psi||$ *in the definition of* $V_N(m, \mu, H)$ *and* $V_N(p)$ *should be replaced by the number* $||\overline{\Psi} - \Psi||$.

If, in particular, Ψ *is a scalar number and* $\Psi \in Q := [A, B]$, *then* $\overline{\Psi} = (A+B)/2$ *and the number* $||\Psi||$ *in* $V_N(m, \mu, H)$ *and* $V_N(p)$ *should be replaced by the number* $\Psi^* = (B - A)/2$.

Remark 1 If the number g in Theorem 1 is unknown but a positive lower bound g_* for $|g|$ is known, then the parameter H in the definition of the truncated estimator (2) should be taken from the interval $(0, g_*)$ and the number $|g|$ in the definition of the function $V_N(m, \mu, H)$ should be replaced by g_*.

Remark 2 The functions $V_N(m, \mu, H)$ and $V_N(p)$ may depend of unknown parameters. At the same time the knowledge of the rate of L_m-convergence of proposed estimators can be useful in various adaptive procedures (control, prediction, filtration etc.; see Sect. 3.5 below as well) and for the construction of pilot estimators (see, e.g., Dobrovidov et al. 2012; Vasiliev 1997; Vasiliev and Koshkin 1998).

Remark 3 The properties of estimators of the often encountered form $G_N^{-1}\Phi_N$ (G_N and Φ_N are random matrices) can be investigated using the presented method (see, e.g., Example 3.4 below).

Remark 4 Theorem 1 and all the corollaries and remarks can be similarly reformulated for the case of observations with continuous time.

3 Examples

We consider in this section applications of the presented method to parameter estimation problems (first four examples) and to a non-parametric one (fifth example). Moreover, the model AR(1) in the last two examples is multivariate.

3.1 Estimation of parameters of a stable first order scalar autoregression

Consider the process satisfying the following equation:

$$x_n = \lambda x_{n-1} + \xi_n, \quad n \geq 1, \tag{5}$$

where noises $\xi_n, n \geq 1$ are i.i.d. zero mean random variables with finite (for some even number $\gamma \geq 2$) moments $\sigma^{2\gamma} = E\xi_n^{2\gamma}$, as well as $Ex_0^{2\gamma} < \infty$ and $|\lambda| < 1$. It should be noted that under these conditions the process (5) is stable (see, e.g., Anderson 1971) and there exist functions $\sigma_x^{2\gamma}(\theta)$, $\theta = (\lambda, \sigma^2, \sigma^{2\gamma})$, such that

$$\sup_n E_\theta x_n^{2\gamma} \leq \sigma_x^{2\gamma}(\theta) < \infty, \tag{6}$$

where E_θ denotes the expectation under the distribution P_θ with the given parameter θ.

Consider the estimation problem of λ and σ^2 with a guaranteed accuracy. Note that similar results in the L_1-metrics can be found in Mikulski and Monsour (1991), Shiryaev and Spokoiny (2000).

(a) Non-asymptotic estimation of λ

We define the estimator of the type (2) on the basis of the least squares estimator (LSE) of the form (1)

$$\hat{\lambda}_N = \frac{\frac{1}{N}\sum_{n=1}^{N} x_n x_{n-1}}{\frac{1}{N}\sum_{n=1}^{N} x_{n-1}^2}, \quad N \geq 1. \tag{7}$$

According to general notation, in this case we set

$$\Psi = \lambda, \quad \Psi_N = \hat{\lambda}_N, \quad f_N = \frac{1}{N}\sum_{n=1}^{N} x_n x_{n-1}, \quad g_N = \frac{1}{N}\sum_{n=1}^{N} x_{n-1}^2.$$

Define $\tilde{\Psi}_N = \tilde{\lambda}_N$ (we omit H for simplicity), where

$$\tilde{\lambda}_N = \hat{\lambda}_N \cdot \chi(g_N \geq H). \tag{8}$$

Theorem 2 *Assume model (5). Then by $m = \mu = \gamma/2$*

(i) *for the case of known σ^2, $0 < H < \sigma^2$ and for some numbers $\tilde{C}_1(m,\theta)$, $\tilde{C}_2(m,\theta)$ we have*

$$E_\theta\left(\tilde{\lambda}_N - \lambda\right)^{2m} \leq \frac{\tilde{C}_1(m,\theta)}{N^m} + \frac{\tilde{C}_2(m,\theta)}{N^{2m}}, \quad N \geq 1; \tag{9}$$

(ii) *for the case of unknown σ^2, in the definition of the estimator (8) we put, for example, $H = (\log N)^{-1}$. Suppose the condition (6) for every $\gamma > 1$ holds (for simplicity). Then for some numbers $C_1(m,\theta), \ldots, C_4(m,\theta)$ and every $m > 1$ we have*

$$E_\theta(\tilde{\lambda}_N - \lambda)^{2m} \leq \tilde{V}_N(m,\theta), \quad N > 1,$$

where

$$\tilde{V}_N(m,\theta) = \frac{C_1(m,\theta)}{N^m} + C_2(m,\theta)\left(\frac{\log N}{N^{3/2}}\right)^m + C_3(m,\theta)\left(\frac{\log N}{N}\right)^{2m} + \frac{C_4(m,\theta)}{N^{2m}} + \gamma_N,$$

with the numbers $\gamma_N = 0$ for N large enough, by the definition.

Proof Proof of Theorem 2 is presented in Sect. 5.

Corollary 3 *In the case $m = 1$, the numbers $\tilde{C}_1(1,\theta)$ and $\tilde{C}_2(1,\theta)$ in the upper bound in (9) have the form*

$$\tilde{C}_1(1,\theta) = \frac{1}{(1-\lambda^2)^2(\sigma^2 - H)^2}\left[12\lambda^2\sigma^2\left(\lambda^2 E x_0^2 + \frac{\sigma^2}{1-\lambda^2}\right) + 3E\xi_1^4\right] + \frac{(\sigma^2)^2}{(1-\lambda^2)H^2},$$

$$\tilde{C}_2(1,\theta) = \frac{1}{(1-\lambda^2)^2(\sigma^2 - H)^2}\left[24\left(\lambda^4 E x_0^4 + \frac{4(\sigma^2)^2}{(1-\lambda^2)^2}\right) + E x_0^4\right].$$

Remark 5 From the proof of Theorem 2 (see, in particular, 39) it follows that

$$P_\theta\left(\tilde{\lambda}_N = \hat{\lambda}_N\right) = 1, \quad \theta \in \Theta$$

for N large enough.

Remark 6 By making use of the following representation for the deviation of $\tilde{\lambda}_N$

$$\tilde{\lambda}_N - \lambda = \hat{\lambda}_N - \lambda - \hat{\lambda}_N \chi(g_N < H)$$

and asymptotic properties of the estimator $\hat{\lambda}_N$, see Mikulski and Monsour (1991), Shiryaev and Spokoiny (2000), it is easy to establish the uniform asymptotic normality of the estimator $\tilde{\lambda}_N$ with the smallest asymptotic variance:

$$\lim_{N\to\infty} \sup_{\vartheta\in\tilde{\Theta}} \left| P_\vartheta\left(\sqrt{N}\left(\tilde{\lambda}_N - \lambda\right) \le x\right) - \Phi\left(x/\sqrt{1-\lambda^2}\right)\right| = 0,$$

where $\Phi(\cdot)$ is a standard Gaussian distribution function and $\tilde{\Theta} = \{\vartheta = (\lambda, x) : |\lambda| \le 1 - r,\ x \in \mathcal{R}^1\}, r \in (0, 1)$.

(b) Efficiency of $\tilde{\lambda}_N$

Analogously to Theorem 2, by $\gamma = 4, m = \mu = 2$ according to the assertion (ii) of Theorem 1 we can get the following inequality for the estimator (8):

$$E_\theta\left(\tilde{\lambda}_N - \lambda\right)^2 \le \frac{1-\lambda^2}{N} + C_1\frac{\log N}{N^{3/2}} + C_2\left(\frac{\log N}{N}\right)^2 + \frac{C_3}{N^2} + \gamma_N, \quad N > 1, \quad (10)$$

where C_1, C_2 and C_3 are some numbers. According to (10) for every $|\lambda| < 1$ we have

$$\overline{\lim_{N\to\infty}} N E_\theta\left(\tilde{\lambda}_N - \lambda\right)^2 \le 1 - \lambda^2. \quad (11)$$

From (11) it follows that the truncated estimator $(\tilde{\lambda}_N)_{N\ge1}$ is optimal (see, e.g., Ibragimov and Khasminskii 1981; Shiryaev and Spokoiny 2000) in the asymptotic minimax sense

$$\underline{\lim_{N\to\infty}} R_{r,N}(\lambda_N) \ge \underline{\lim_{N\to\infty}} \inf_{\lambda_N} R_{r,N}(\lambda_N) = \lim_{N\to\infty} R_{r,N}\left(\tilde{\lambda}_N\right) = 1, \quad (12)$$

where

$$R_{r,N}(\lambda_N) = \sup_{\mathcal{P}} \sup_{|\lambda|\le 1-r} I(\lambda, f) N E_\lambda(\lambda_N - \lambda)^2, \quad r \in (0, 1)$$

and the infimum is taken over the class of all (non-randomized) estimators λ_N of the parameter λ. Here $\mathcal{P}$ is the class of all densities $f(\cdot)$ of the noises (ξ_n) having finite second moments and the Fisher information $I(\lambda, f) = gI(f)(= (1-\lambda^2)^{-1}$ for the case of Gaussian densities $f(\cdot)$),

$$I(f) = \int \left(\frac{f'(x)}{f(x)}\right)^2 f(x)\mathrm{d}x.$$

(c) Guaranteed estimation of λ

For the parameter estimation with a guaranteed accuracy we assume that, e.g., $\theta \in \Theta$, where $\Theta = \{\theta = (\lambda, \sigma^2, \sigma^{2\gamma}) : |\lambda| \le r < 1, 0 < \underline{\sigma}^2 \le \sigma^2 \le \overline{\sigma}^2, \sigma^{2\gamma} \le \overline{\sigma}^{2\gamma}\}$.
In this case, for $0 < H < \underline{\sigma}^2$ we can find the numbers

$$\tilde{C}_1(m) = \sup_{\theta\in\Theta} \tilde{C}_1(m,\theta) < \infty, \quad \tilde{C}_2(m) = \sup_{\theta\in\Theta} \tilde{C}_2(m,\theta) < \infty,$$

and then, according to (9), we have

$$\sup_{\theta\in\Theta} E_\theta \left(\tilde{\lambda}_N - \lambda\right)^{2m} \le \frac{\tilde{C}_1(m)}{N^m} + \frac{\tilde{C}_2(m)}{N^{2m}}, \quad N \ge 1. \tag{13}$$

In particular, for $m = 1$, according to Corollary 3 it holds

$$\tilde{C}_1(1) = \frac{1}{(1-r^2)^2(\underline{\sigma}^2 - H)^2}\left[12r^2\overline{\sigma}^2\left(r^2 Ex_0^2 + \frac{\overline{\sigma}^2}{1-r^2}\right) + 3\overline{\sigma}^4\right] + \frac{(\overline{\sigma}^2)^2}{(1-r^2)H^2},$$
$$\tilde{C}_2(1) = \frac{1}{(1-r^2)^2(\underline{\sigma}^2 - H)^2}\left[24\left(r^4 Ex_0^4 + \frac{4(\overline{\sigma}^2)^2}{(1-r^2)^2}\right) + Ex_0^4\right].$$

(d) Non-asymptotic estimation of σ^2

For the estimation problem of the noise variance σ^2 in the model (5).
We consider the LSE-type estimator $\hat{\sigma}_N^2$ defined as

$$\hat{\sigma}_N^2 = \frac{1}{N}\sum_{n=1}^{N}\left(x_n - \lambda_N^* x_{n-1}\right)^2, \quad N \ge 1, \tag{14}$$

where

$$\lambda_N^* = \mathrm{proj}_{[-1,1]}\tilde{\lambda}_N$$

and we use the estimator $\tilde{\lambda}_N$ of λ, defined in (8) with non-asymptotic properties (13) for $m = 1$ and $m = 2$. It should be noted that this estimator is asymptotically equivalent to the corresponding LSE.

Define the function

$$\overline{V}_N = \sup_{\theta\in\Theta} V_N^*(\theta) < \infty, \quad N \ge 1,$$

where

$$V_N^*(\theta) = 3\left[2\left(\tilde{C}_1(2,\theta)\sigma_x^8(\theta)\right)^{1/2} + \sigma^4 - (\sigma^2)^2\right]\frac{1}{N} + 6\left[\left(\tilde{C}_2(2,\theta)\sigma_x^8(\theta)\right)^{1/2}\right.$$
$$\left. +2\left(\tilde{C}_1(2,\theta)\sigma^4\sigma_x^4(\theta)\right)^{1/2}\right]\frac{1}{N^2} + 12\left(\tilde{C}_2(2,\theta)\sigma^4\sigma_x^4(\theta)\right)^{1/2}\frac{1}{N^3}.$$

Corollary 4 *Assume the model (5) with $\gamma = 4$. Then*

$$\sup_{\theta\in\Theta} E_\theta\left(\hat{\sigma}_N^2 - \sigma^2\right)^2 \le \overline{V}_N, \quad N \ge 1. \tag{15}$$

Proof Proof of Corollary 4 is presented in Sect. 5.

3.2 AR-ARCH(1,1)

Consider the process satisfying the following equation:

$$x_n = \lambda x_{n-1} + \sqrt{\sigma_0^2 + \sigma_1^2 x_{n-1}^2}\cdot\xi_n, \quad n \ge 1, \tag{16}$$

where noises $\xi_n,\ n \ge 1$ are i.i.d. zero mean random variables with the variance equal to one and finite fourth moments $\sigma^4 = E\xi_1^4$, as well as $Ex_0^4 < \infty$.

We consider the estimation problem of $\lambda,\ \sigma_0^2$ and σ_1^2 with a guaranteed accuracy.

(a) Non-asymptotic estimation of λ

Define the LSE $\hat{\lambda}_N$ of λ of the following form:

$$\hat{\lambda}_N = \frac{\frac{1}{N}\sum_{n=1}^N x_n x_{n-1}}{\frac{1}{N}\sum_{n=1}^N x_{n-1}^2}, \quad N \ge 1,$$

which is strongly consistent (see, e.g., Malyarenko 2010) under the following stability condition:

$$\lambda^4 + 6\lambda^2\sigma_1^2 + \left(\sigma_1^2\right)^2\sigma^4 < 1. \tag{17}$$

According to general notation, in this case we set

$$\Psi = \lambda, \quad \Psi_N = \hat{\lambda}_N, \quad f_N = \frac{1}{N}\sum_{n=1}^N x_n x_{n-1}, \quad g_N = \frac{1}{N}\sum_{n=1}^N x_{n-1}^2.$$

Define $\tilde{\Psi}_N = \tilde{\lambda}_N$ (we omit H for simplicity), where

$$\tilde{\lambda}_N = \hat{\lambda}_N \cdot \chi(g_N \ge H). \tag{18}$$

We assume that $\theta = (\lambda, \sigma_0^2, \sigma_1^2) \in \Theta$, where

$$\Theta = \left\{ \theta = \left(\lambda, \sigma_0^2, \sigma_1^2\right) : \lambda^4 + 6\lambda^2\sigma_1^2 + \left(\sigma_1^2\right)^2 \sigma^4 \le r, \underline{\sigma}_0^2 \le \sigma_0^2 \le \overline{\sigma}_0^2, \quad \underline{\sigma}_1^2 \le \sigma_1^2 \le \overline{\sigma}_1^2 \right\}$$

for some numbers $r \in (0, 1)$, $\underline{\sigma}_0^2$, $\overline{\sigma}_0^2$, $\underline{\sigma}_1^2$, and $\overline{\sigma}_1^2$.

Define

$$\begin{aligned}
\overline{\varphi}_N &= \left(\overline{\sigma}_0^2 \sigma_x^2 + \overline{\sigma}_1^2 \sigma_x^4\right) \frac{1}{N}, \\
\overline{w}_N &= \left\{ 12 \left(\overline{\sigma}_0^2 \sigma_x^2 + \overline{\sigma}_1^2 \sigma_x^4\right) + \sigma_\xi^4 \cdot \left(\left(\overline{\sigma}_0^2\right)^2 + 2\overline{\sigma}_0^2 \overline{\sigma}_1^2 \sigma_x^2 + \left(\overline{\sigma}_1^2\right)^2 \sigma_x^4 \right) \right\} \frac{1}{N} \\
&\quad + \left\{ E x_0^4 + 3\sigma_x^4 4 \right\} \frac{1}{N^2},
\end{aligned}$$

where σ_ξ^4 is an upper bound for $E(\xi_1^2 - 1)^2$.

Theorem 3 *Assume the model (16) and in the definition (18) the number* $0 < H < \underline{\sigma}_0^2/(1 - \underline{\sigma}_1^2)$. *Then*

$$\sup_{\theta \in \Theta} E_\theta \left(\tilde{\lambda}_N - \lambda\right)^2 \le \frac{1}{H^2} \overline{\varphi}_N + \frac{\left(1 - \underline{\sigma}_1^2\right)^2}{\left(\underline{\sigma}_0^2 - \left(1 - \underline{\sigma}_1^2\right) H\right)^2} \overline{w}_N, \quad N \ge 1. \tag{19}$$

Proof Proof of Theorem 3 is presented in Sect. 5. □

Remark 7 It should be noted that in the absence of a priori information on parameters of the system the inequalities of the type (10) can be obtained as well.

(b) Non-asymptotic estimation of σ_0^2 and σ_1^2

We will construct estimators with guaranteed accuracy on the basis of correlation estimators:

(1b) σ_0^2 with known σ_1^2 :

$$\hat{\sigma}_{0N}^2 = \frac{1}{N} \sum_{n=1}^N \left[x_n^2 - \left(\hat{\lambda}_N^2 + \sigma_1^2\right) x_{n-1}^2 \right];$$

(2b) σ_1^2 with known σ_0^2 :

$$\hat{\sigma}_{1N}^2 = \frac{\sum_{n=1}^N \left(x_n^2 - \sigma_0^2\right)}{\sum_{n=1}^N x_{n-1}^2} - \hat{\lambda}_N^2,$$

which are strongly consistent under the condition (17), see, e.g., Malyarenko (2010).

Define estimators for considered cases

(1b)

$$\tilde{\sigma}_{0N}^2 = \frac{1}{N}\sum_{n=1}^{N}\left[x_n^2 - \left(\left(\lambda_N^*\right)^2 + \sigma_1^2\right)x_{n-1}^2\right];$$

(2b)

$$\tilde{\sigma}_{1N}^2 = \frac{\frac{1}{N}\sum_{n=1}^{N}\left(x_n^2 - \sigma_0^2\right)}{\frac{1}{N}\sum_{n=1}^{N}x_{n-1}^2}\chi(g_N \geq H) - \left(\lambda_N^*\right)^2,$$

where

$$\lambda_N^* = \mathrm{proj}_{[-1,1]}\tilde{\lambda}_N,$$

$\tilde{\lambda}_N$ and g_N are defined in (18).

Similar to Sects. 3.1 and 3.2a, the upper bounds for the MSE's of these estimators with known constants C_0 and C_1 can be found:

$$\text{(i)} \quad \sup_{\theta\in\Theta_0} E_\theta\left(\tilde{\sigma}_{0N}^2 - \sigma_0^2\right)^2 \leq \frac{C_0}{N}, \tag{20}$$

where $\Theta_0 = \{\theta = (\lambda, \sigma_0^2) : \lambda^4 + 6\lambda^2\sigma_1^2 + (\sigma_1^2)^2\sigma^4 \leq r,\ \underline{\sigma}_0^2 \leq \sigma_0^2 \leq \overline{\sigma}_0^2\}$ and

$$\text{(ii)} \quad \sup_{\theta\in\Theta_1} E_\theta\left(\tilde{\sigma}_{1N}^2 - \sigma_1^2\right)^2 \leq \frac{C_1}{N}, \tag{21}$$

where $\Theta_1 = \{\theta = (\lambda, \sigma_1^2) : \lambda^4 + 6\lambda^2\sigma_1^2 + (\sigma_1^2)^2\sigma^4 \leq r, \underline{\sigma}_1^2 \leq \sigma_1^2 \leq \overline{\sigma}_1^2\}, \quad r \in (0, 1)$.

3.3 Non-Gaussian Ornstein-Uhlenbeck process by discrete–time observations

The results presented below make it possible to do the statistical inferences for continuous-time stochastic systems by fixed sample size of observations. Moreover, one of the main assumption is a discrete scheme of observations. It corresponds to numerous real situations, in particular, in problems of financial mathematics.

Consider the following regression model:

$$\mathrm{d}x(t) = ax(t)\mathrm{d}t + \mathrm{d}\xi(t), 0 \leq t \leq T \tag{22}$$

with an initial condition $x(0) = x_0$, having all the moments. Here $\xi(t) = \rho_1 W(t) + \rho_2 Z(t)$, $\rho_1 \neq 0$ and ρ_2 are some constants, $(W(t), t \geq 0)$ is a standard Wiener process, given on a probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, P)$, adapted to a filtration $\{\mathcal{F}_t\}_{t\geq 0}$, $Z(t) = \sum_{k=1}^{N_t} Y_k$ is a compound Poisson process, where $Y_k,\ k \geq 1$ are i.i.d. random variables having all the moments and (N_t) is a Poisson process with the intensity $\lambda > 0$. It should be noted that for $\rho_2 = 0$ the process (22) is a standard Ornstein-Uhlenbeck process.

We suppose that $a \in [-\Delta, -\delta]$, where δ and $\Delta > \delta$ are known positive numbers. The problem is to estimate the parameter a by observations of the discrete-time process $y = (y_k)$,

$$y_k = x(t_k), \quad t_k = \frac{k}{n}T, \quad k = 0, \ldots, n.$$

Using the following representation for the solution of the Eq. (22)

$$x(t) = \mathrm{e}^{at} x_0 + \int_0^t \mathrm{e}^{a(t-z)} \mathrm{d}\xi(z), \quad 0 \le t \le T\,,$$

we get the recurrent equation for the observations (y_k) :

$$y_k = b y_{k-1} + \eta_k, \quad k = 0, \ldots, n, \tag{23}$$

where $b = \mathrm{e}^{aT/n}$, $\eta_k = \int_{t_{k-1}}^{t_k} \mathrm{e}^{a(t_k - s)} \mathrm{d}\xi(s)$ are i.i.d. random variables with

$$E_a \eta_k = 0, \quad \sigma^2 := D_a \eta_k = \frac{1}{2a}\left(\rho_1^2 + \lambda \rho_2^2\right)[b^2 - 1].$$

Moreover, for this model all the moments $\sigma^{2m} = E_a \eta_k^{2m}$ are finite and there exist their upper bounds $\overline{\sigma}^{2m} = \sup_{a \le -\delta} \sigma^{2m}$, $m \ge 1$.

Define the estimator $\tilde{a}_n$ of a using an estimator $\tilde{b}_n$ of b as follows:

$$\tilde{a}_n = \frac{n}{T} \log \tilde{b}_n, \quad n \ge 1, \tag{24}$$

where

$$\tilde{b}_n = \hat{b}_n \cdot \chi(g_n \ge H) + L\chi(g_n < H)$$

is constructed using Corollary 2. Here $\hat{b}_n = f_n / g_n$ is the LSE of b, obtained from the Eq. (23) with

$$f_n = \frac{1}{n}\sum_{k=1}^n y_k y_{k-1}, \quad g_n = \frac{1}{n}\sum_{k=1}^n y_{k-1}^2,$$

$L = [\mathrm{e}^{-\delta T/n} + \mathrm{e}^{-\Delta T/n}]/2$ and the number g is defined as

$$g = \frac{\sigma^2}{1 - b^2}.$$

Then the estimator $\tilde{b}_n$ has all the properties of the estimator $\tilde{\lambda}_N$, defined in (8). In particular, according to Theorem 1, which holds for this model for all $m \geq 1$ and $\mu \geq 1$, the following inequalities

$$\sup_{a \leq -\delta} E_a \left(\tilde{b}_n - b\right)^{2m} \leq \frac{C_1^*(m)}{n^m} + \frac{C_2^*(\mu)}{n^\mu}, \quad n \geq 1 \tag{25}$$

for an arbitrary $0 < H \leq \underline{\sigma}^2$ hold, where

$$\underline{\sigma}^2 = \frac{1}{2\delta}\left(\varrho_1^2 + \lambda\varrho_2^2\right)\left[1 - \mathrm{e}^{-2\delta}\right]$$

and numbers $C_1^*(m)$, $C_2^*(\mu)$ are known.

From (24) and (25) it is easy to verify the following property of estimators $\tilde{a}_n$ for every $m \geq 1$ and $\mu > m$:

$$\begin{aligned}\sup_{a\in[-\Delta,-\delta]} E_a(\tilde{a}_n - a)^{2m} &= (nT^{-1})^{2m} \sup_{a\in[-\Delta,-\delta]} E_a\left[\log\left(1 + \frac{\tilde{b}_n - b}{b}\right)\right]^{2m} \\ &\leq (nT^{-1}\mathrm{e}^{\Delta T/n})^{2m}\left\{\frac{C_1^*(m)}{n^m} + \frac{C_2^*(\mu)}{n^\mu}\right\}, \quad n \geq 1. \end{aligned} \tag{26}$$

3.4 Vector AR(1)

We show in this section a possibility to apply the presented general truncated method for guaranteed estimation of matrix parameters in multivariate systems.

Consider the s-dimensional process ($s > 1$) satisfying the following equation:

$$x(n) = Ax(n-1) + \xi(n), \quad n \geq 1, \tag{27}$$

where noises $\xi(n), n \geq 1$ are i.i.d. zero mean random vectors with finite moments of the order $8(s-1)$, as well as $E||x(0)||^{8(s-1)} < \infty$ and the stability condition for the process (27) is satisfied, i.e. all the eigenvalues of the matrix A lie within the unit circle (see, e.g., Anderson 1971). Define the number $\sigma_\xi^4 = E||\xi(1)||^4$. We suppose that the matrix parameter A to be estimated belongs to a compact set Λ from the stable region.

It should be noted that under these conditions there exist finite numbers σ_x^{2m}, such that

$$\sup_{A\in\Lambda,n} E_A||x(n)||^{2m} \leq \sigma_x^{2m}, \quad 1 \leq m \leq 4(s-1). \tag{28}$$

Consider the estimation problem of A with a guaranteed accuracy.

We define the estimator of the type (2) on the basis of the LSE of the form (1)

$$\hat{A}_N = \overline{\Phi}_N \overline{G}_N^{-1}, \quad N \geq 1,$$

where

$$\overline{G}_N = \frac{1}{N} G_N, \quad G_N = \sum_{n=1}^{N} x(n-1)x'(n-1),$$

$$\overline{\Phi}_N = \frac{1}{N} \Phi_N, \quad \Phi_N = \sum_{n=1}^{N} x(n)x'(n-1), \quad N \geq 1.$$

Define the matrix

$$\overline{G}_N^+ = \overline{\Delta}_N \overline{G}_N^{-1}, \quad \overline{\Delta}_N = \det\left(\overline{G}_N\right).$$

According to the general notation, in this case we set

$$\Psi = A, \quad \Psi_N = \hat{A}_N, \quad f_N = \overline{\Phi}_N \overline{G}_N^+, \quad g_N = \overline{\Delta}_N.$$

Using formula (27) it is easy to verify that with P_A-probability one it holds

$$\lim_{N\to\infty} \overline{G}_N = F \quad \text{and} \quad \lim_{N\to\infty} \overline{\Delta}_N = \Delta > 0,$$

where F is a positive definite $s \times s$-matrix (see, e.g., Anderson 1971), such that $\Delta_* = \inf\limits_{A\in\Lambda} \Delta > 0$.

Then

$$f = A\Delta, \quad g = \Delta$$

and $\tilde{\Psi}_N = \tilde{A}_N$ (we omit H for simplicity), where

$$\tilde{A}_N = \hat{A}_N \cdot \chi\left(\overline{\Delta}_N \geq H\right), \tag{29}$$

and $H \in (0, \Delta_*)$.

In Sect. 5, it is shown that there exists a given number C_Λ such that for every $N \geq 1$,

$$\sup_{A\in\Lambda} E_A ||\tilde{A}_N - A||^2 \leq \frac{C_\Lambda}{N}. \tag{30}$$

3.5 Logarithmic density derivative

Consider the problem of estimating the multivariate logarithmic derivative ($q = 1$ in the general problem statement),

$$\Psi(t) = \nabla f(t)/f(t)$$

($\nabla f(t)$ is a $s \times 1$-vector of the first-order partial derivatives of $f(t)$) of a distribution density $f(t)$ of the i.i.d. vector noises $\xi(n) = (\xi_1(n), \ldots, \xi_s(n))'$ in the model (27) considered in Sect. 3.4. Noises $\xi(n), n \geq 1$ are zero mean random vectors with finite moments of the order 4α, where $\alpha = \max\{4(s-1), \nu + 1 + \delta\}$ for some $\delta > 0$, as well as $E||x(0)||^{4\alpha} < \infty$ (the number ν will be defined below in Assumption (f)) and the stability condition is satisfied. It is supposed that the matrix parameter A to be estimated belongs to a compact set Λ from the stable region.

ASSUMPTION (f) We suppose that the function $f(\cdot)$ satisfies the following condition:

$$\sup_{z \in \mathcal{R}^s} f(z) \leq C_f,$$

and, for some even $\nu \geq 2$, as well as $\mathcal{L} > 0$ and $\gamma \in (0, 1]$, for all the partial derivatives of the order $1 + \nu$ the Lipshitz condition

$$\left| f^{(1+\nu)}(x) - f^{(1+\nu)}(y) \right| \leq \mathcal{L}||x - y||^{\gamma}$$

holds.

The knowledge of $\Psi(t)$ is important in various statistical problems, e.g. for constructing the algorithm of optimal control of an autoregressive process, estimating of a regression curve, and testing close hypotheses. These problems are of particular interest in the case of dependent observations, for example, where the logarithmic derivative of a density is used for designing the optimal algorithms of nonlinear filtering and adaptive control of random processes (see, e.g., Dobrovidov et al. 2012 and references therein).

We will construct estimators of $f(t)$ and $\nabla f(t)$ using the following estimators $\tilde{\xi}(n)$ of noises $\xi(n)$ in (27):

$$\tilde{\xi}(n) = x(n) - A^*_{n-1}x(n-1), \quad n = 1, \ldots, N, \tag{31}$$

where $A^*_{n-1} = \text{proj}_{\Lambda} \tilde{A}_{n-1}$, $\tilde{A}_n$ is the estimator defined in (29).

As a non-parametric estimator of a density $f(t) = f^{(0)}(t)$ satisfying Assumption (f) and its partial derivative $f^{(1)}(t) = \partial f(t)/\partial t_j$, we use the combined statistic of the form

$$\widehat{f_N^{(r)}}(t) = \frac{1}{N h_{r,N}^{s+r}} \sum_{i=1}^{N} K^{(r)}\left(\frac{t - \tilde{\xi}(i)}{h_{r,N}}\right), \quad r = 0, 1, \tag{32}$$

where $K^{(0)}(u) = K(u) = \prod_{k=1}^{s} K(u_k)$ is a s-dimensional multiplicative kernel which, generally speaking, does not necessarily possess the characterizing properties of density (nonnegativity and normalization to 1), $K^{(1)}(u) = \partial K(u)/\partial u_j$, sequences of numbers $h_{r,N} \downarrow 0, N \to \infty$.

Then, as an estimator of the ratio $\Psi(t)$ from the observations $(x(n))_{n \geq 1}$, one can use the ratio

$$\hat{\Psi}_N(t) = \widehat{\nabla f}_N(t) / \hat{f}_N(t)$$

of statistics defined in (32).

Estimators of type (32) of the density and its derivatives from observations (31) were considered in Dobrovidov et al. (2012), Chapter 4, where it was established, in particular, their asymptotic normality and convergence with probability one. The results on asymptotic ratio estimation of the partial derivatives of the noise distribution density in multivariate dynamic systems are given in Dobrovidov et al. (2012), Sect. 5.1.

To obtain estimators of $\Psi(t)$ with a known MSE we apply Theorem 1.

First we suppose in addition to Assumption (f) that there exists a known number c_f such that

$$0 < c_f \leq f(t).$$

Define the estimator

$$\tilde{\Psi}_N(t) = \hat{\Psi}_N(t)\chi\left(\hat{f}_N(t) \geq H\right), \quad N \geq 1$$

with a given number $H \in (0, c_f)$.

By the definition (31), the estimators $\tilde{\xi}(n)$ can be represented in the form

$$\tilde{\xi}(n) = \xi(n) + \left(A - A^*_{n-1}\right)x(n-1), \quad n = 1, \ldots, N.$$

Note that the matrices $A - A^*_{n-1}$ are uniformly bounded

$$\sup_{A\in\Lambda,n} ||A - A^*_{n-1}|| \leq C, \quad n \geq 1 \tag{33}$$

and, similar to Sect. 3.4, the following properties of the estimator A^*_{n-1} can be obtained:

$$\sup_{A\in\Lambda} E_A||A - A^*_{n-1}||^4 \leq \sup_{A\in\Lambda} E_A||A - \tilde{A}_{n-1}||^4 \leq \frac{C_\Lambda}{n^2}. \tag{34}$$

Using (33), (34) and the Cauchy-Schwarz inequality, we can find the known numbers C_1 and C_m, such that

$$\sup_{A\in\Lambda} \sum_{n=1}^{N} E_A||\left(A - A^*_{n-1}\right)x(n-1)||^{2m} \leq \begin{cases} C_1 \log N, & m = 1, \\ C_m, & 1 < m \leq \nu + 1. \end{cases} \tag{35}$$

Similar relations were obtained in Dobrovidov et al. (2012) (see Lemmas 5.1.3 and 5.1.5) for another type of estimators.

Then, using technique of Theorems 4.3.1 and 5.1.3 from Dobrovidov et al. (2012) and (35), by appropriate chosen kernels $K(\cdot)$ in (32), we can find known numbers $C_{i,1}$ and $C_{i,2}$, $i = 0, 1$, such that

$$\sup_{A\in\Lambda} E_A\left(\hat{f}_N(t)-f(t)\right)^2 \le \frac{C_{0,1}}{Nh_{0,N}^s}+C_{0,2}h_{0,N}^{2(\nu+1+\gamma)}, \tag{36}$$

$$\sup_{A\in\Lambda} E_A||\widehat{\nabla f}_N(t)-\nabla f(t)||^2 \le \frac{C_{1,1}}{Nh_{1,N}^{s+2}}+C_{1,2}h_{1,N}^{2(\nu+\gamma)}. \tag{37}$$

Thus, to minimize the obtained upper bounds, it is natural to put

$$h_{0,N}=\left(\frac{sC_{0,1}}{2(\nu+1+\gamma)C_{0,2}}\right)^{\frac{1}{2(\nu+1+\gamma)+s}} N^{-\frac{1}{2(\nu+1+\gamma)+s}},$$

$$h_{1,N}=\left(\frac{(s+3)C_{1,1}}{2(\nu+\gamma)C_{1,2}}\right)^{\frac{1}{2(\nu+1+\gamma)+s}} N^{-\frac{1}{2(\nu+1+\gamma)+s}}$$

and for obviously defined numbers $\tilde{C}_{0,1}$ and $\tilde{C}_{1,1}$ we have

$$\sup_{A\in\Lambda} E_A\left(\hat{f}_N(t)-f(t)\right)^2 \le \tilde{C}_{0,1}N^{-\frac{2(\nu+1+\gamma)}{2(\nu+1+\gamma)+s}},$$

$$\sup_{A\in\Lambda} E_A||\widehat{\nabla f}_N(t)-\nabla f(t)||^2 \le \tilde{C}_{1,1}N^{-\frac{2(\nu+\gamma)}{2(\nu+1+\gamma)+s}}.$$

It makes possible to apply Theorem 1 (taking into account Corollary 1) to estimation of $\Psi(t)$ for $H\in(0,c_f)$ with a known upper bound

$$\sup_{A\in\Lambda} E_A||\tilde{\Psi}_N(t)-\Psi(t)||^2 \le \tilde{C}_0N^{-\frac{2(\nu+\gamma)}{2(\nu+1+\gamma)+s}}+\tilde{C}_1N^{-\frac{2(\nu+1+\gamma)}{2(\nu+1+\gamma)+s}}. \tag{38}$$

For the case of unknown lower bound c_f for the number $f(t)$, Theorem 1 can be applied if noises $\xi(n)$ and the initial time $x(0)$ of the process (27) have moments of the order $8\max\{2(s-1),(\nu+1+\delta)\}$. Under these conditions, similar to (36) and (37), the following inequalities can be obtained:

$$\sup_{A\in\Lambda} E_A\left(\hat{f}_N(t)-f(t)\right)^4 \le C_0N^{-\frac{4(\nu+1+\gamma)}{2(\nu+1+\gamma)+s}},$$

$$\sup_{A\in\Lambda} E_A||\widehat{\nabla f}_N(t)-\nabla f(t)||^4 \le C_1N^{-\frac{4(\nu+\gamma)}{2(\nu+1+\gamma)+s}},$$

where C_0 and C_1 are some constants.

Put $H=(\log N)^{-1}$ in the definition of the estimator $\tilde{\Psi}_N(t)$. Then, according to (4) with $p=1$, $m=\mu=2$, for some numbers C_1^*, C_2^* and C_3^* for N large enough (to eliminate γ_N) and using Corollary 1 we have

$$\sup_{A\in\Lambda} E_A||\tilde{\Psi}_N(t)-\Psi(t)||^2 \le C_1^*N^{-\frac{2(\nu+\gamma)}{2(\nu+1+\gamma)+s}}$$
$$+C_2^*(\log N)^2N^{-\frac{4(\nu+\gamma)+2}{2(\nu+1+\gamma)+s}}+C_3^*N^{-\frac{4(\nu+1+\gamma)}{2(\nu+1+\gamma)+s}}.$$

It should be noted that in both considered cases (known and unknown number c_f) the estimator $\tilde{\Psi}_N(t)$ has equal rates of convergency in the mean square sense. Moreover, this rate is similar to the case of independent observations (see, e.g., Dobrovidov et al. 2012).

4 Summary

We have presented the truncated estimation method of ratio type functionals constructed by dependent samples of finite size. This method allows to obtain estimators with a guaranteed accuracy on a time interval of a fixed length.

As an illustration, parametric and non-parametric estimation problems are considered.The presented method was applied to estimation of parameters of a linear autoregressive and a non-linear AR-ARCH processes, as well as a non-Gaussian Ornstein-Uhlenbeck process by discrete-time observations (see properties 13, 15, 19, 20, 21, 26 and 30). Moreover, the estimators with a guaranteed accuracy in the mean square sense of a multivariate logarithmic derivative of the noise density of an autoregressive process with an unknown dynamic matrix parameter was investigated, see (38). The asymptotic efficiency in the sense (12) of the truncated estimator of the parameter of a stable autoregression is established.

The presented method can be similarly applied to samples from continuous-time models.

5 Proofs

5.1 Proof of Theorem 1

From the definition of the estimator $\tilde{\Psi}_N(H)$ we find its deviation

$$\tilde{\Psi}_N(H) - \Psi = \frac{f_N - \Psi g_N}{g_N} \cdot \chi(|g_N| \geq H) - \Psi \cdot \chi(|g_N| < H).$$

Then, using the Chebyshev inequality and the definition of $V_N(m, \mu, H)$ we can estimate the desired moment

$$\begin{aligned}
E||\tilde{\Psi}_N(H) - \Psi||^{2m} &= E\frac{||f_N - \Psi g_N||^{2m}}{g_N^{2m}} \cdot \chi(|g_N| \geq H) \\
&+||\Psi||^{2m} \cdot P(|g_N| < H) \leq \frac{1}{H^{2m}} E||f_N - \Psi g_N||^{2m} \\
&+||\Psi||^{2m} \cdot P(|g_N - g| > |g| - H) \\
&\leq \frac{1}{H^{2m}}\varphi_N(m) + ||\Psi||^{2m} \cdot \frac{E(g_N - g)^{2\mu}}{(|g| - H)^{2\mu}} \leq V_N(m, \mu, H).
\end{aligned}$$

The inequalities (4) can be proved similarly. Let $\beta \in (0, 1)$ be the given number. Using the C_r-inequality $(\alpha + \beta)^r \leq 2^{r-1}(|\alpha|^r + |\beta|^r),\ r \geq 1$, Hölder's inequality and assumptions (A1), (A2) of the theorem, we have

$$
\begin{aligned}
&E||\tilde{\Psi}_N(H_N)-\Psi||^{2p} \le E||f_N-\Psi g_N||^{2p}\left[\frac{1}{|g|}+\left|\frac{1}{g_N}-\frac{1}{g}\right|\right]^{2p}\cdot\chi(|g_N|\ge H_N)\\
&+||\Psi||^{2p}\cdot P(|g_N|<H_N)\le g^{-2p}E||f_N-\Psi g_N||^{2p}\left[1+2\frac{|g_N-g|}{|g_N|}+\frac{|g_N-g|^2}{g_N^2}\right]^p\\
&+||\Psi||^{2p}\cdot P(||g|-|g_N-g||<H_N)\le 2^{p-1}g^{-2p}E||f_N-\Psi g_N||^{2p}\\
&+2^{2p-1}g^{-2p}H_N^{-p}\left(E||f_N-\Psi g_N||^{2m}\right)^{p/m}\left(E|g_N-g|^{mp/(m-p)}\right)^{(m-p)/m}\\
&+4^{p-1}g^{-2p}H_N^{-2p}\left(E||f_N-\Psi g_N||^{2m}\right)^{p/m}\left(E|g_N-g|^{2mp/(m-p)}\right)^{(m-p)/m}\\
&+||\Psi||^{2p}\cdot[P(|g_N-g|>\beta|g|)+\gamma_N]\le 2^{p-1}g^{-2p}\varphi_N(p)\\
&+2^{2p-1}g^{-2p}H_N^{-p}\varphi_N^{p/m}(m)w_N^{p/2\mu}(\mu)+4^{p-1}g^{-2p}H_N^{-2p}\varphi_N^{p/m}(m)w_N^{p/\mu}(\mu)\\
&+||\Psi||^{2p}\cdot\left[(\beta|g|)^{-2\mu}w_N(\mu)+\gamma_N\right]=V_N(p).
\end{aligned}
$$

5.2 Proof of Theorem 2

To proof Theorem 2 we verify assumptions of Theorem 1.

Using the equality

$$
g_N=\frac{\sigma^2}{1-\lambda^2}+\frac{1}{(1-\lambda^2)N}\left[x_0^2-x_N^2+2\lambda\sum_{n=1}^N x_{n-1}\xi_n+\sum_{n=1}^N\left(\xi_n^2-\sigma^2\right)\right],
$$

which can be obtained from (5), it is easy to verify (see, e.g., Shiryaev and Spokoiny 2000) that

$$
g=\lim_{N\to\infty}g_N=\frac{\sigma^2}{1-\lambda^2}\quad P_\theta-\text{a.s.} \tag{39}
$$

and $g\ge\sigma^2>0$. Then for $m=\gamma/2,\mu=m$, there exist constants $C_1(m,\theta)$ and $C_2(m,\theta)$ such that

$$
w_N(m,\theta)=\frac{C_1(m,\theta)}{N^m}+\frac{C_2(m,\theta)}{N^{2m}}
$$

and, using the Burkholder and Hölder inequalities, we have

$$
\begin{aligned}
E_\theta\left(f_N-\lambda g_N\right)^{2m}&=\frac{1}{N^{2m}}E_\theta\left(\sum_{n=1}^N x_{n-1}\xi_n\right)^{2m}\le\frac{B_{2m}^{2m}}{N^{2m}}E_\theta\left(\sum_{n=1}^N x_{n-1}^2\xi_n^2\right)^m\\
&\le\frac{B_{2m}^{2m}\sigma^{2m}}{N^{m+1}}\sum_{n=1}^N E_\theta x_{n-1}^{2m}\le B_{2m}^{2m}\sigma^{2m}\sigma_x^{2m}(\theta)\frac{1}{N^m}=:\varphi_N(m,\theta),
\end{aligned}
$$

where B_{2m} is the coefficient from the Burkholder inequality (see, e.g., Burkholder 1973; Liptser and Shiryaev 1989).

Thus all the assumptions of Theorem 1 hold.

5.3 Proof of Corollary 4

Using (5), (6), (9), (14) and the Cauchy-Schwarz inequality, we have

$$E_\theta\left(\hat{\sigma}_N^2-\sigma^2\right)^2 = E_\theta\left[\left(\lambda_N^*-\lambda\right)^2\frac{1}{N}\sum_{n=1}^N x_{n-1}^2 - 2\left(\lambda_N^*-\lambda\right)\frac{1}{N}\sum_{n=1}^N x_{n-1}\xi_n\right.$$
$$\left.+\frac{1}{N}\sum_{n=1}^N\left(\xi_n^2-\sigma^2\right)\right]^2 \le 3\left[2\left(E_\theta\left(\tilde{\lambda}_N-\lambda\right)^4 E_\theta\left(\frac{1}{N}\sum_{n=1}^N x_{n-1}^2\right)^4\right)^{1/2}\right.$$
$$\left.+4\left(E_\theta\left(\tilde{\lambda}_N-\lambda\right)^4 E_\theta\left(\frac{1}{N}\sum_{n=1}^N x_{n-1}\xi_n\right)^4\right)^{1/2}+E_\theta\left(\frac{1}{N}\sum_{n=1}^N\left(\xi_n^2-\sigma^2\right)\right)^2\right]$$
$$\le 3\left[2\left(\left(\frac{\tilde{C}_1(2,\theta)}{N^2}+\frac{\tilde{C}_2(2,\theta)}{N^4}\right)\frac{1}{N}\sum_{n=1}^N E_\theta x_{n-1}^8\right)^{1/2}\right.$$
$$\left.+4\left(\left(\frac{\tilde{C}_1(2,\theta)}{N^2}+\frac{\tilde{C}_2(2,\theta)}{N^4}\right)\frac{\sigma^4}{N^3}\sum_{n=1}^N E_\theta x_{n-1}^4\right)^{1/2}+\frac{\sigma^4-(\sigma^2)^2}{N}\right] \le V_N^*(\theta).$$

5.4 Proof of Theorem 3

Using the following equation for the process (x_n^2) :

$$x_n^2 = \left(\lambda^2+\sigma_1^2\right)x_{n-1}^2+\sigma_0^2+\varsigma_n, \quad n\ge 1,$$

where

$$\varsigma_n = 2\lambda x_{n-1}\sqrt{\sigma_0^2+\sigma_1^2 x_{n-1}^2}\cdot\xi_n+\sigma_0^2\left(\xi_n^2-1\right)+\sigma_1^2 x_{n-1}^2\left(\xi_n^2-1\right),$$

it is easy to find the supremums

$$\sigma_x^2 = \sup_{\theta\in\Theta,n} E_\theta x_n^2, \quad \sigma_x^4 = \sup_{\theta\in\Theta,n} E_\theta x_n^4$$

and for $\theta\in\Theta$ the limit (see, e.g., Malyarenko 2010)

$$g := \lim_{N\to\infty} g_N = \frac{\sigma_0^2}{1-\lambda^2-\sigma_1^2} \quad P_\theta - a.s.$$

For this model we can calculate

$$\sup_{\theta\in\Theta} E_\theta(g_N - g)^2 \le \overline{w}_N$$

and, by the definition of f_N and g_N we have

$$\sup_{\theta\in\Theta} E_\theta(f_N - \lambda g_N)^2 = \frac{1}{N^2}\sup_{\theta\in\Theta} E_\theta\left(\sum_{n=1}^{N} x_{n-1}\sqrt{\sigma_0^2 + \sigma_1^2 x_{n-1}^2}\cdot\xi_n\right)^2$$
$$= \frac{1}{N^2}\sum_{n=1}^{N}\sup_{\theta\in\Theta} E_\theta x_{n-1}^2\left(\sigma_0^2 + \sigma_1^2 x_{n-1}^2\right) \le \overline{\varphi}_N.$$

5.5 Proof of (30)

To prove (30) we verify the conditions of Theorem 1 by $m = \mu = 2$:

$$\sup_{A\in\Lambda} E_A||f_N - A\overline{\Delta}_N||^2 \le \varphi_N, \tag{40}$$
$$\sup_{A\in\Lambda} E_A\left(\overline{\Delta}_N - \Delta\right)^2 \le w_N. \tag{41}$$

Define the matrices

$$\overline{\zeta}_N = \frac{1}{N}\zeta_N, \quad \zeta_N = \sum_{n=1}^{N}\xi(n)x'(n-1), \quad N \ge 1.$$

We prove (40) using (28) and the Burkholder inequality (the number B_4 is the constant from the upper bound in the Burkholder inequality, see Burkholder 1973; Liptser and Shiryaev 1989 and Sect. 3.1):

$$\sup_{A\in\Lambda} E_A||f_N - A\overline{\Delta}_N||^2 = \sup_{A\in\Lambda} E_A||\overline{\zeta}_N\overline{G}_N^+||^2 \le \left(\sup_{A\in\Lambda} E_A||\overline{\zeta}_N||^4\right)^{1/2}$$
$$\times\left(\sup_{A\in\Lambda} E_A||\overline{G}_N^+||^4\right)^{1/2} \le sB_4^2\left(\sigma_\xi^4\sigma_x^4\right)^{1/2}\frac{1}{N}\cdot(s^5\sigma_x^{8(s-1)})^{1/2} =: \varphi_N.$$

Also, (41) follows from (28) and the inequalities

$$\sup_{A\in\Lambda} E_A||\overline{G}_N - F||^4 \le C_G N^{-2},$$

where C_G is a given number. This can be proved similar to Lemma 5.1.6 in Dobrovidov et al. (2012). Then the function w_N in (41) is inverse proportional to N :

$$w_N = C_w/N,$$

where C_w is a given number.

Acknowledgments The author is grateful to the anonymous reviewer, as well as to the editor and an associated editor for constructive criticism and clear guidelines. Finally, the author thanks S. Pergamenshchikov for the helpful comments and discussion on (12).

References

Anderson, T. W. (1971). *The statistical analysis of time series*. New York: Wiley.

Arkoun, O. (2011). Sequential adaptive estimators in nonparametric autoregressive models. *Sequential Analysis: Design Methods and Applications*, *30*(2), 228–246.

Arkoun, O., Pergamenchtchikov, S. (2008). Nonparametric estimation for an autoregressive models. *Vestnik of Tomsk State University, Series: Mathematics and Mechanics*, *2*(3), 20–30.

Borovkov, A. A. (1997). *Mathematical statistics*. Novosibirsk: Nauka and Sobolev Institute Press.

Burkholder, D. L. (1973). Distribution function inequalities for martingales. *Annals of Probability*, *1*, 19–42.

Cramér, G. (1999). *Mathematical methods of statistics*. Princeton: Princeton University Press.

Dobrovidov, A. V., Koshkin, G. M., Vasiliev, V. A. (2012). *Non-parametric state space models*. Heber City: Kendrick Press. http://www.kendrickpress.com/nonpara.html (Russian original: Dobrovidov, A.V., Koshkin, G.M., Vasiliev, V.A. (2004). *Nonparametric estimation of functionals of stationary sequences distributions*. Moscow: Nauka).

Efroimovich, S. (2007). Sequential design and estimation in heteroscedastic nonparametric regression. *Sequential Analysis: Design Methods and Applications*, *26*(1), 3–25. doi:10.1080/07474940601109670.

Fourdrinier, D., Konev, V., Pergamenshchikov, S. (2009). Truncated sequential estimation of the parameter of a first order autoregressive process with dependent noises. *Mathematical Methods of Statistics*, *18*(1), 43–58.

Galtchouk, L., Konev, V. (2001). On sequential estimation of parameters in semimartingale regression models with continuous time parameter. *The Annals of Statistics*, *29*(5), 1508–1536.

Ibragimov, I., Khasminskii, R. (1981). *Statistical estimation: Asymptotic theory*. New York: Springer.

Konev, V. V. (1985). *Sequential parameter estimation of stochastic dynamical systems*. Tomsk: Tomsk University Press (in Russian).

Konev, V. V., Pergamenshchikov, S. M. (1985). Sequential estimation of the parameters of diffusion processes. *Problems of Information Transmission*, *21*(1), 48–62 (in Russian).

Konev, V. V., Pergamenshchikov, S. M. (1990a). Truncated sequential estimation of the parameters in random regression. *Sequential Analysis*, *9*(1), 19–41.

Konev, V. V., Pergamenshchikov, S. M. (1990b). On truncated sequential estimation of the drifting parametermeant in the first order autoregressive models. *Sequential Analysis*, *9*(2), 193–216.

Konev, V. V., Pergamenshchikov, S. M. (1992). Sequential estimation of the parameters of linear unstable stochastic systems with guaranteed accuracy. *Problems of Information Transmission*, *28*(4), 35–48 (in Russian).

Küchler, U., Vasiliev, V. (2010). On guaranteed parameter estimation of a multiparameter linear regression process. Automatica. *Journal of IFAC*, *46*, 637–646.

Liptser, R. Sh., Shiryaev, A. N. (1977). *Statistics of random processes. I: General theory*. New York: Springer.

Liptser, R. Sh., Shiryaev, A. N. (1978). *Statistics of random processes. II: Applications* (1978). New York: Springer.

Liptser, R Sh., Shiryaev, A. N. (1989). *Theory of martingales*. Dordrecht: Kluwer.

Malyarenko, A. A. (2010). Estimating the generalized autoregression model parameters for unknown noise distribution. *Automation and Remote Control*, *71*(2), 291–302.

Malyarenko, A., Vasiliev, V. (2012). On parameter estimation of partly observed bilinear discrete-time stochastic systems. Metrika, 75(3), 403–424, April 2012. Online first 27 October 2010. doi:10.1007/s00184-010-0333-5.

Mikulski, P. W., Monsour, M. J. (1991). Optimality of the maximum likelihood estimator in first-order autoregressive processes. *Journal of Time Series Analysis*, *12*(3), 237–253. Online first 27 October 2010. doi:10.1007/s00184-010-0333-5.

Novak, S. Yu. (1996). On the distribution of the ratio of sums of random variables. *Theory of Probability and its Applications*, *41*(3), 533–560 (in Russian).

Novikov, A. A. (1971). Sequential estimation of the parameters of the diffusion processes. *Theory of Probability and its Applications*, *16*(2), 394–396 (in Russian).

Penskaya M.Ya. (1990). On stable estimation of the function of parameter. *Statistical Methods of Estimation and Testing Hypotheses*. Perm University Press, Perm, pp. 44–55 (in Russian).
Politis, D. N., Vasiliev, V. A. (2012a). Sequential kernel estimation of a multivariate regression function. In *Proceedings of the IX international conference 'System Identification and Control Problems'*, SICPRO'12. Moscow, 30 January–2 February, 2012, V. A. Tapeznikov Institute of Control Sciences, pp. 996–1009.
Politis, D. N., Vasiliev, V. A. (2012b). Non-parametric sequential estimation of a regression function based on dependent observations. *Sequential Analysis: Design Methods and Applications* 26 pages, 2013 (accepted).
Roll, J., Nazin, A., Ljung, L. (2002). A non-asymptotic approach to local modelling. In *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.
Roll, J., Nazin, A., Ljung, L. (2005). Non-linear system identification via direct weight optimization. Automatica. *Journal of IFAC*, *41*(3), 475–490. Predocumentation is available at http://www.control.isy.liu.se/research/reports/2005/2696.pdf (Special Issue on "Data-Based Modelling and System Identification").
Shiryaev, A. N., Spokoiny, V. G. (2000). *Statistical experiments and decisions. Asymptotic theory*. Singapore: World Scientific.
Vasiliev, V. A. (1997). On identification of dynamic systems of autoregressive type. *Automation and Remote Control*, *12*, 106–118.
Vasiliev, V. A. (2012). One investigation method of ratio type estimators. Berlin: 5 of the Mathematical Institute of Humboldt University, pp 1–15. http://www2.mathematik.hu-berlin.de/publ/pre/2012/p-list-12.html (Preprint).
Vasiliev, V. A., Koshkin, G. M. (1998). Nonparametric identification of autoregression. *Probability Theory and Its Applications*, *43*(3), 577–588.
Wald, A. (1947). *Sequential analysis*. New York: Wiley.