Serdica

Mathematical Journal

Сердика

Математическо списание

Serdica Math. J. **43** (2017), 221–266

Serdica
Mathematical Journal

Bulgarian Academy of Sciences
Institute of Mathematics and Informatics

TRUNCATED ESTIMATION METHOD AND APPLICATIONS

Tatiana V. Dogadova, Marat I. Kusainov, Vyacheslav A. Vasiliev

Communicated by P. Jagers

Abstract. This paper presents an estimation method of ratio type functionals by dependent sample of fixed size. This method makes it possible to obtain estimators with guaranteed accuracy in the sense of the L_{2m}-norm, $m \geq 1$.

As an illustration, some parametric estimation problems on a time interval of a fixed length are considered. In particular, parameters of linear continuous-time and non-linear discrete-time processes are estimated. Moreover, the parameter estimation problem of non-Gaussian Ornstein–Uhlenbeck process by discrete-time observations with guaranteed accuracy is solved.

In addition to non-asymptotic properties, the limit behavior of presented estimators is investigated. It is shown that all the truncated estimators have rates of convergence of the estimators they are based upon. These estimators are used for the construction of adaptive predictors for dynamical systems with unknown parameters.

2010 *Mathematics Subject Classification:* 60G25, 62M20, 60G52, 60J60, 62F12, 93E10.

Key words: Adaptive optimal prediction, dynamic systems, autoregressive processes, stochastic differential equations, stochastic differential equations with time delay, risk function; guaranteed parameter estimation.

The problem of asymptotic efficiency of adaptive one-step predictors for stable discrete- and continuous-time processes with unknown parameters is considered. The proposed criteria of optimality are based on the loss function, defined as a linear combination of sample size and squared prediction error's sample mean. As a rule, the optimal sample size is a special stopping time.

1. Introduction. The main purpose of this paper is to present applications of the truncated estimation method in order to construct optimal adaptive predictors for the stochastic processes related with discrete and continuous-time dynamical systems. The proposed procedures are based on the so-called truncated estimators which have been developed in order to estimate ratio type functionals from a wide class by dependent observations and by samples of fixed size so that they had guaranteed accuracy in the sense of the L_{2m}-norm, $m \geq 1$. Examples of parameter estimation problems of discrete and continuous time systems on a time interval of a fixed length are considered.

It is shown that truncated estimators may keep asymptotic properties of the estimators they are based upon. One of the many useful applications of estimators with the said quality is adaptive prediction for dynamical systems with unknown parameters. It is then possible to optimize the risk function which is a linear combination of sample mean of mean-square deviation of predictors and sample size. The risk function of such structure was proposed in [3], see also references therein.

According to Ljung's concept the prediction is a crucial part in constructing complete probabilistic models of dynamical systems (see [24, 25]). A model is considered to be useful if it allows to make predictions of high statistical quality. Models of dynamical systems often have unknown parameters, which demand estimation in order to build adaptive predictors. The quality of adaptive prediction explicitly depends on the chosen estimators of model parameters. Possible estimation methods include the classic stochastic approximation, maximum likelihood, least squares and sequential estimation methods among others. The first three methods provide estimators with given statistical properties under asymptotic assumptions, when the duration of observations tends to infinity (see, e.g., [1, 36]). The sequential estimation method makes it possible to obtain estimators with guaranteed accuracy by samples of finite but random and unbounded size (see, e.g., [4, 9, 15, 16, 17, 18, 23, 27, 29, 30, 33, 34, 35, 37] among others).

Both approaches do not guarantee prescribed estimation accuracy when using samples of non-random finite size and lead up to complicated analytical problems in adaptive procedures.

Results in non-asymptotic parametric and non-parametric problems can be found in [28, 37] among others. In particular, they investigated non-asymptotic properties of the LSE-estimator for the scalar first-order autoregressive process.

At the same time, the more modern truncated sequential estimation method yields estimators with prescribed accuracy by samples of random but bounded size (see, e.g., [5, 8, 13, 14, 34, 35]). However, at the moment this approach is developed for scalar dynamic systems only. The truncated estimation method was introduced in [40] as a modification of the truncated sequential estimation one. Truncated estimators were constructed for ratio type multivariate functionals by samples of fixed size and have guaranteed accuracy in the sense of the L_{2m}-norm, $m \geq 1$ (see also [41]).

The requirement of both good prediction quality and reasonable duration of observations is formulated as a risk efficiency problem. The criterion is given by certain loss functions and optimization is performed based on it. The loss function describing sample mean of squared prediction errors and sample size as well as the corresponding risk was examined in [38, 39] in application to scalar AR(1). Later the results of those papers were refined and extended to other stochastic models in [11]. There was considered a risk function defined on the basis of squared estimation error of sequential estimator of the dynamic parameter. A modified stopping rule was proposed, enhancing the result of [38]. In the two papers on risk efficiency problems mentioned above the least squares estimators and sequential estimators of unknown parameters were used.

In this paper we construct and investigate real-time predictors which only use past values of the process. Such an approach leads to some technical difficulties but is more closely related to real applications. We consider the problem of minimization of the risk function associated with predictors of values of the process and size of a sample. It should be noted that first truncated parameter estimation method was applied for construction of adaptive optimal predictors of VAR(1) in [22]. Here we apply this method for more complicated stochastic systems. Among the processes considered are stable multivariate discrete time AR(1), ARMA$(1,1)$ and RCA(1), as well as continuous time diffusion and time delayed processes. The proposed procedure is shown to be asymptotically risk efficient as the cost of prediction error tends to infinity.

2. Truncated estimation method. General results. Let $(\Omega, \mathcal{F}, \mathsf{P})$ be a probability space with a filtration $\{\mathcal{F}_t\}$ in discrete or continuous time and let f_t and g_t be $\{\mathcal{F}_t\}$-adapted random processes, where f_t and g_t is $s \times q$-dimensional matrix and scalar function respectively.

Let

$$\Psi_T = f_T / g_T \tag{1}$$

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

$$\Psi = f/g$$

and f_T and g_T are estimators of matrix f and number g respectively.

Consider the following modification of the estimator Ψ_T:

$$\tilde{\Psi}_T(H) = \Psi_T \cdot \chi(|g_T| \geq H), \tag{2}$$

where H is a function $H = H_T$, defined below and the notation $\chi(A)$ means the indicator function of set A.

Our main aim is to formulate general conditions on the processes f_T, g_T and on the parameter H giving a possibility to estimate Ψ with a guaranteed accuracy in the sense of the L_{2m}-norm, $m \geq 1$.

Define for some $\varphi_T(m)$, $w_T(\mu)$, H and g, the function

$$V_T(m, \mu, H) = \frac{1}{H^{2m}} \varphi_T(m) + \frac{\|\Psi\|^{2m}}{(|g| - H)^{2\mu}} w_T(\mu),$$

as well as for positive integer $p < m$ and a positive monotonously decreasing function H_T, the function

$$V_T(p) = 2^{2p-1} g^{-2p} \cdot \left(\varphi_T(p) + H_T^{-2p} \cdot \varphi_T^{\frac{p}{m}}(m) \cdot w_T^{\frac{p}{\mu}}(\mu) \right) + \|\Psi\|^{2p} \cdot (g - H_T)^{-2\mu} \cdot w_T(\mu)$$

and the time $T_0 = \inf\{T \geq 0 : \ H_T \leq |g|\}$.

Theorem 2.1. *Assume for some values m and μ there exist positive functions $\varphi_T(m)$ and $w_T(\mu)$, decreasing to zero, as well as a value g such that the following assumptions hold*

(i) $\mathsf{E}\|f_T - \Psi g_T\|^{2m} \leq \varphi_T(m)$;

(ii) $\mathsf{E}(g_T - g)^{2\mu} \leq w_T(\mu)$.

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

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

$$\mathsf{E}\|\tilde{\Psi}_T(H) - \Psi\|^{2m} \leq V_T(m, \mu, H);$$

(*b*) *in the case of unknown g for every (possibly slowly decreasing to zero) positive function* $H = H_T$ *and every positive integer* p*, satisfying for some* $m > 1$ *and* $\mu > 1$

$$\frac{mp}{m-p} \le \mu$$

it holds

$$\mathsf{E}\|\tilde{\Psi}_T(H) - \Psi\|^{2p} \le V_T(p), \quad T > T_0.$$

Remark 1. If the number g in Theorem 2.1 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 $V_T(m, \mu, H)$ should be replaced by g_*.

Proof of Theorem 2.1 is similar to the proof of Theorem 1 from [40] formulated for the discrete-time case.

3. Parameter estimation. Examples.

3.1. Discrete-time systems.

3.1.1. Estimation of parameters of a stable first order scalar autoregression. Consider the process satisfying the following equation

$$x_n = \lambda x_{n-1} + \xi_n, \quad n \ge 1, \tag{3}$$

where noises $\xi_n,\ n \ge 1$ are i.i.d. zero mean random variables with finite (for some even number $\gamma \ge 2$) moments $\sigma^{2\gamma} = \mathsf{E}\xi_n^{2\gamma}$, as well as $\mathsf{E}x_0^{2\gamma} < \infty$ and $|\lambda| < 1$.

Consider the estimation problem of λ and $\sigma^2 = \mathsf{E}\xi_n^2$ with a guaranteed accuracy.

In what follows, C will denote a generic non-negative constant whose value is not critical (and not necessarily the same throughout the paper).

a) Non-asymptotic estimation of λ

We define the estimator of the type (2) with $T = N$ on the basis of the least squares estimator (LSE) of the form (1)

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

According to general notation, in this case we have

$$\Psi = \lambda, \quad \Psi_N = \hat{\lambda}_N,$$

$$f_N = \frac{1}{N}\sum_{n=1}^{N} x_n x_{n-1}, \quad g_N = \frac{1}{N}\sum_{n=1}^{N} x_{n-1}^2$$

and $\tilde\Psi_N = \tilde\lambda_N$,

(4)
$$\tilde\lambda_N = \hat\lambda_N \cdot \chi(g_N \geq H).$$

Using the equality

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

which can be obtained from (3), we can find the limit (see [37, 40])

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

All the conditions of Theorem 2.1 hold, hence
– for the case of known σ^2 and $0 < H < \sigma^2$

(5)
$$\mathsf{E}_\theta(\tilde\lambda_N - \lambda)^{2m} \leq \frac{C(\theta)}{N^m} + \frac{C(\theta)}{N^{2m}}, \quad N \geq 1;$$

– for the case of unknown σ^2 we put $H = (H_N)$ (e.g., slowly decreasing function) from Theorem 2.1 in the definition (4) of the estimator $\tilde\lambda_N$ and for N large enough, we have

$$\mathsf{E}_\theta(\tilde\lambda_N - \lambda)^{2m} \leq \frac{C(\theta)H_N^{2m}}{N^m}.$$

Here $\theta = (\lambda,\ \sigma^2,\ \sigma^{2\gamma})$.

For the parameter estimation with a guaranteed accuracy we have to know that, e.g., $\theta \in \Theta$, where $\Theta = \{\theta = (\lambda,\ \sigma^2,\ \sigma^{2\gamma}) :\ |\lambda| \leq r < 1,\ 0 < \underline{\sigma}^2 \leq \sigma^2 \leq \overline{\sigma}^2\}$.

In this case we can find the known functions

$$\overline{\varphi}_N(m) = \sup_{\theta\in\Theta}\varphi_N(m,\theta) \quad \text{and} \quad \overline{w}_N(m) = \sup_{\theta\in\Theta} w_N(m,\theta)$$

such that

$$\sup_{\Theta}\mathsf{E}_\theta(f_N - \lambda g_N)^{2m} \leq \overline{\varphi}_N(m),$$

$$\sup_{\Theta}\mathsf{E}_\theta(g_N - g)^{2m} \leq \overline{w}_N(m).$$

In general, for $0 < H < \underline{\sigma}^2$, we have

$$\sup_{\Theta} \mathsf{E}_\theta(\tilde\lambda_N - \lambda)^{2m} \le \frac{C}{N^m} + \frac{C}{N^{2m}}, \quad N \ge 1. \tag{6}$$

In particular, for $\gamma = 2$ and $m = 1$,

$$\sup_{\Theta} \mathsf{E}_\theta(\tilde\lambda_N - \lambda)^2 \le \left[\frac{(\overline{\sigma}^2)^2}{(1-r^2)H^2} + \frac{r^2 C}{(\underline{\sigma}^2 - H)^2}\right]\frac{1}{N} + \frac{r^2 C}{(\underline{\sigma}^2 - H)^2}\frac{1}{N^2}.$$

b) Non-asymptotic estimation of σ^2

Consider the estimation problem of the noise variance σ^2 in the model (3) under the assumption $\gamma = 4$ ($\sigma^8 < \infty$, $\mathsf{E}x_0^8 < \infty$).

In the definition of the LSE type estimator $\hat\sigma_N^2$ defined as

$$\hat\sigma_N^2 = \frac{1}{N}\sum_{n=1}^{N}(x_n - \tilde\lambda_N x_{n-1})^2, \quad N \ge 1,$$

we use the estimator $\tilde\lambda_N$ of λ, which is defined in (4) and has known non-asymptotic properties (6) for $m = 1$ and $m = 2$.

Thus, we have obtained estimator of σ^2 with a guaranteed accuracy:

$$\sup_{\Theta} \mathsf{E}_\theta(\hat\sigma_N^2 - \sigma^2)^2 \le \frac{C}{N}, \quad N \ge 1.$$

It should be noted, that this estimator is asymptotically equivalent to the corresponding LSE. In particular, it has optimal rate of convergence as $N \to \infty$.

Full proofs of results of this section can be found in [40].

3.1.2. Estimation of parameters of a stable ARARCH(1,1). Consider the process satisfying the following equation

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

where noises $\xi_n,\ n \ge 1$ are i.i.d. zero mean random variables with the variance equal to one and finite fourth moment $\sigma^4 = \mathsf{E}\xi_1^4$, as well as $\mathsf{E}x_0^4 < \infty$.

Define the LSE $\hat\lambda_N$ of λ of the form:

$$\hat\lambda_N = \frac{\frac{1}{N}\sum\limits_{n=1}^{N} x_n x_{n-1}}{\frac{1}{N}\sum\limits_{n=1}^{N} x_{n-1}^2}, \quad N \ge 1,$$

which is strongly consistent under the following stability condition

$$\lambda^4 + 6\lambda^2\sigma_1^2 + (\sigma_1^2)^2\sigma^4 < 1. \tag{7}$$

a) Non-asymptotic estimation of λ

According to general notation, in this case we have

$$\Psi = \lambda, \quad \Psi_N = \hat{\lambda}_N,$$

$$f_N = \frac{1}{N}\sum_{n=1}^{N} x_n x_{n-1}, \quad g_N = \frac{1}{N}\sum_{n=1}^{N} x_{n-1}^2$$

and $\tilde{\Psi}_N = \tilde{\lambda}_N$,

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

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

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

Then for $0 < H < \dfrac{\underline{\sigma}_0^2}{1 - \underline{\sigma}_1^2}$ and every $N \geq 1$

$$\sup_{\Theta} \mathsf{E}_\theta(\tilde{\lambda}_N - \lambda)^2 \leq \frac{1}{H^2 N} + \frac{(1 - \underline{\sigma}_1^2)^2}{(\underline{\sigma}_0^2 - (1 - \underline{\sigma}_1^2)H)^2}\frac{1}{N^2}.$$

It should be noted, that the rate of convergence of the obtained upper bound is the same as the rate of the LSE and is optimal.

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) of σ_0^2 with known σ_1^2 :

$$\hat{\sigma}_0^2(N) = \frac{1}{N}\sum_{n=1}^{N}[x_n^2 - (\hat{\lambda}_N^2 + \sigma_1^2)x_{n-1}^2];$$

2b) of σ_1^2 with known σ_0^2 :

$$\hat{\sigma}_1^2(N) = \frac{\sum\limits_{n=1}^{N}(x_n^2 - \sigma_0^2)}{\sum\limits_{n=1}^{N} x_{n-1}^2} - \hat{\lambda}^2(N),$$

which are strongly consistent under the condition (7), see, e.g., [26].

Define estimators for considered cases

1b) $$\tilde{\sigma}_0^2(N) = \frac{1}{N}\sum_{n=1}^{N}[x_n^2 - ((\lambda_N^*)^2 + \sigma_1^2)x_{n-1}^2];$$

2b) $$\tilde{\sigma}_1^2(N) = \frac{\frac{1}{N}\sum\limits_{n=1}^{N}(x_n^2 - \sigma_0^2)}{\frac{1}{N}\sum\limits_{n=1}^{N}x_{n-1}^2}\chi(g_N \geq H) - (\lambda_N^*)^2,$$

where

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

$\tilde{\lambda}_N$ and g_N are defined in (8).

Similar to the previous sections, the upper bounds for the MSE's of these estimators with known constants C_0 and C_1 can be found

$$(i) \qquad \sup_{\Theta_0} \mathsf{E}_\theta(\tilde{\sigma}_0^2(N) - \sigma_0^2)^2 \leq \frac{C_0}{N},$$

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

$$(ii) \qquad \sup_{\Theta_1} \mathsf{E}_\theta(\tilde{\sigma}_1^2(N) - \sigma_1^2)^2 \leq \frac{C_1}{N},$$

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

Full proofs of results of this section can be found in [40].

3.1.3. Estimation of parameters of a stable first order VAR(1). We apply in this section the presented general truncated method for estimation of matrix parameters in multivariate systems.

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

$$x(n) = \Lambda x(n-1) + \xi(n), \quad n \geq 1, \tag{9}$$

where noises $\xi(n),\ n \geq 1$ are i.i.d. zero mean random column vectors with the variance matrix $\Sigma = \mathsf{E}\xi(n)\xi'(n)$ and finite moments of the order $8(p-1)$, as well as $\mathsf{E}\|x(0)\|^{8(p-1)} < \infty$ and the stability condition for the process (9) is satisfied, i.e. all the eigenvalues of the matrix Λ lie in the open unit circle.

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

$$\sup_{n,\Lambda_0} \mathsf{E}_\Lambda\|x(n)\|^{2m} \leq \sigma_x^{2m}, \quad 1 \leq m \leq 4(p-1),$$

where Λ_0 is a compact set from the stable region of $(x(n))$.

Consider the estimation problem of Λ with a guaranteed accuracy.

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

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

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

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

According to the general notation, in this case we have

$$\Psi = \Lambda, \quad \Psi_N = \hat{\Lambda}_N,$$

$$f_N = \overline{\Phi}_N\overline{G}_N^{+}, \quad g_N = \overline{\Delta}_N,$$

and $\tilde{\Psi}_N = \tilde{\Lambda}_N$,

$$\tilde{\Lambda}_N^{*} = \hat{\Lambda}_N \cdot \chi(g_N \geq H).$$

It is easy to verify that with P_Λ-probability one

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

where F is a positive definite $p \times p$-matrix (see, e.g., [1, 12]).

Then

$$f = \Lambda\Delta, \quad g = \Delta.$$

It can be shown, see [40] that there exists a given number C_0 such that for every $N \geq 1$,

$$\sup_{\Lambda\in\Lambda_0} \mathsf{E}_\Lambda\|\tilde{\Lambda}_N^{*} - \Lambda\|^2 \leq \frac{C_0}{N},$$

Λ_0 is a compact set from the stable region of the process (9).

Consider the case of unknown Λ_0.

Define the number $N_0 = \max\left\{p, \left\lfloor e^{\Delta^{-2}}\right\rfloor\right\}$.

Let the truncated estimators $\widetilde{\Lambda}_N$ be defined as follows

$$\widetilde{\Lambda}_N = \hat{\Lambda}_N \cdot \chi(\overline{\Delta}_N \geq H_N), \quad H_N = \log^{-1/2}(N+1). \tag{10}$$

Lemma 3.1. *Assume the model* (9) *and let for some integer* $m \geq 1$ *the conditions*

$$\mathsf{E}\|\xi(1)\|^{4mp} < \infty, \quad \mathsf{E}\|x(0)\|^{4mp} < \infty$$

be true. Then the truncated estimators $\widetilde{\Lambda}_N$ *satisfy*

(i) *for* $1 \leq N < N_0$

$$\mathsf{E}_\Lambda\|\widetilde{\Lambda}_N - \Lambda\|^{2m} \leq C;$$

(ii) *for* $N \geq N_0$

$$\mathsf{E}_\Lambda\|\widetilde{\Lambda}_N - \Lambda\|^{2m} \leq \frac{C \log^m N}{N^m}. \tag{11}$$

For Proof of Lemma 3.1 see Section 4.1 in [22].

Remark 2. Note that the truncated estimator of the parameter of the scalar first-order autoregression considered in Section 3.1.1 has simpler structure and a little stronger basic property compare (11) to (5).

3.1.4. Estimation of parameters of a stable first order multivariate random coefficient autoregressive model. Consider a stable p-dimensional vector random coefficient autoregressive process (VRCA(1)) satisfying the equation

$$x(k) = \Lambda_{k-1}x(k-1) + \xi(k), \quad k \geq 1, \tag{12}$$

$$\Lambda_k = \Lambda + \eta(k), \quad k \geq 0.$$

The parameter matrix Λ of size $p \times p$ is unknown, processes $(\xi(k))$ and $(\eta(k-1))$ are mutually independent and form sequences of i.i.d. random vectors and matrices respectively for which

$$\Sigma = \mathsf{E}\xi(1)\xi'(1) > 0, \quad \sigma_\xi^2 = \mathsf{E}\|\xi(1)\|^2 < \infty,$$

$$\Sigma_\eta = \mathsf{E}\eta'(0)\eta(0) > 0, \quad \mathsf{E}\xi(1) = \mathsf{E}\eta(0) = 0.$$

Here $\Sigma > 0$ denotes that Σ is positive definite in the sense of quadratic forms, i.e. $y^T\Sigma y > 0$ for every constant vector $y \neq 0$. For the process to be stable it is required that the matrix $\overline{\Lambda} = \mathsf{E}\Lambda_0^{\otimes 2} = \Lambda^{\otimes 2} + \mathsf{E}\eta^{\otimes 2}(0)$ be stable (i.e. its eigenvalues lie in the open unit circle), where $Y^{\otimes 2} = Y \otimes Y$.

We define the estimator of the type (2) as follows

$$\widetilde{\Lambda}_k = \hat{\Lambda}_k \chi\left(\overline{\Delta}_k \geq H_k\right), \quad k \geq 1, \tag{13}$$

where

$$\hat{\Lambda}_k = \overline{G}_k \overline{F}_k^{-1}, \quad k \geq p, \quad \text{and} \quad \hat{\Lambda}_i = 0, \quad i = \overline{0, p-1},$$

$$H_k = \log^{-1/2}(k+1), \quad \overline{G}_k = \frac{1}{k} \sum_{i=1}^{k} x(i) x'(i-1),$$

(14) $$\overline{F}_k = \frac{1}{k} \sum_{i=1}^{k} x(i-1) x'(i-1), \quad \overline{\Delta}_k = \det(\overline{F}_k).$$

Using formula (12) it can be shown that almost surely

$$\lim_{k \to \infty} \overline{F}_k = F \quad \text{and} \quad \lim_{k \to \infty} \overline{\Delta}_k = \Delta > 0,$$

where F is a positive definite $p \times p$ matrix.

Define the number $k_0 = \max \left\{ p, \ \left\lfloor e^{\Delta^{-2}} \right\rfloor \right\}$.

Lemma 3.2. *Assume the model* (12) *and let for some integer* $m \geq 1$ *the matrix* $\mathsf{E}\Lambda_0^{\otimes 4mp}$ *be stable, let also the following conditions hold*

$$\mathsf{E}\|\xi(1)\|^{4mp} < \infty, \quad \mathsf{E}\|x(0)\|^{4mp} < \infty.$$

Then for the truncated estimators $\widetilde{\Lambda}_k$ *it holds*

(i) *for* $1 \leq k < k_0$

(15) $$\mathsf{E}_\theta \|\widetilde{\Lambda}_k - \Lambda\|^{2m} \leq C;$$

(ii) *for* $k \geq k_0$

(16) $$\mathsf{E}_\theta \|\widetilde{\Lambda}_k - \Lambda\|^{2m} \leq \frac{C \log^m k}{k^m}.$$

Here E_θ denotes expectation under the distribution P_θ with the given parameter $\theta = (\lambda_{11}, \ ..., \ \lambda_{pp}, \ \Sigma_\eta, \ \sigma_\xi^2)$.

See Appendix for the proof of lemma.

3.1.5. Estimation of parameters of a stable first order multivariate autoregressive moving-average model. Consider a stable p-dimensional vector ARMA(1,1) process satisfying the equation

(17) $$x(k) = \Lambda x(k-1) + \xi(k) + \mathrm{M}\xi(k-1), \quad k \geq 1,$$

where Λ and M are $p \times p$ stable matrices. We assume the parameter Λ to be unknown and M to be known. The random vectors $\xi(k)$ for $k \geq 1$ are i.i.d. with zero mean and finite variance $\sigma^2 = \mathsf{E}\|\xi(1)\|^2$.

Let the truncated estimators of the autoregressive parameter Λ be based on the following Yule-Walker type estimators

$$\hat{\Lambda}_k = \overline{\Phi}_k \overline{G}_k^{-1}, \quad k \geq 2, \quad \Lambda_0 = \Lambda_1 = 0,$$

where

$$\overline{\Phi}_k = \frac{1}{k-1} \sum_{i=2}^{k} x(i)x'(i-2), \quad \overline{G}_k = \frac{1}{k-1} \sum_{i=2}^{k} x(i-1)x'(i-2),$$

and have the form

$$\widetilde{\Lambda}_k = \hat{\Lambda}_k \chi\left(|\overline{\Delta}_k| \geq H_k\right), \quad k \geq 2. \tag{18}$$

Here $\overline{\Delta}_k = \det(\overline{G}_k)$ and $H_k = \log^{-1/2} k$. It can be shown that the limit (in almost sure sense) $\Delta = \lim_{k\to\infty} \overline{\Delta}_k$ is nonzero if the matrix G defined as

$$G = \Lambda F + \mathrm{M}\Sigma, \quad F = \sum_{n\geq 0} \Lambda^n \widetilde{\Sigma} (\Lambda')^n,$$

$$\widetilde{\Sigma} = \Lambda\Sigma\mathrm{M}' + \mathrm{M}\Sigma\Lambda' + \Sigma + \mathrm{M}\Sigma\mathrm{M}', \quad \Sigma = \mathsf{E}\xi(1)\xi'(1) \tag{19}$$

is non-singular.

Define the number $k_0 = \max\left\{p, \ \left\lfloor e^{\Delta^{-2}} \right\rfloor\right\}$.

Lemma 3.3. *Assume the model* (17) *and let for some integer* $m \geq 1$ *the conditions*

$$\mathsf{E}\|\xi(1)\|^{4pm} < \infty, \quad \mathsf{E}\|x(0)\|^{4pm} < \infty$$

be true. Assume also that the matrix G *is non-singular. Then the truncated estimators* $\widetilde{\Lambda}_k$ *satisfy*

(i) *for* $1 \leq k < k_0$

$$\mathsf{E}_\theta \|\widetilde{\Lambda}_k - \Lambda\|^{2m} \leq C; \tag{20}$$

(ii) *for* $k \geq k_0$

$$\mathsf{E}_\theta \|\widetilde{\Lambda}_k - \Lambda\|^{2m} \leq \frac{C \log^m k}{k^m}. \tag{21}$$

Here $\theta = (\lambda_{11}, \ldots, \lambda_{pp}, \mu_{11}, \ldots, \mu_{pp}, \sigma^2)$.
See Appendix for the proof of this and next lemmas.

Lemma 3.4. *Let for some integer $m \geq 1$ the following conditions hold*

$$\mathsf{E}\|\xi(0)\|^{4m(p+1)} < \infty, \quad \mathsf{E}\|x(0)\|^{4m(p+1)} < \infty,$$

also let the matrices Σ and G be non-singular. Then the following holds

(i) *for $1 \leq k < k_0$*

$$\mathsf{E}_\theta\|\widetilde{M}_k - \mathrm{M}\|^{2m} \leq C; \tag{22}$$

(ii) *for $k \geq k_0$*

$$\mathsf{E}_\theta\|\widetilde{M}_k - \mathrm{M}\|^{2m} \leq \frac{C \log^m k}{k^m}. \tag{23}$$

3.2. Continuous-time systems.

3.2.1. Parameter estimation of a stable Ornstein–Uhlenbeck process. Consider the model

$$dx_t = ax_t dt + dw_t, \quad t \geq 0 \tag{24}$$

with an unknown parameter a, where x_0 is zero mean random variable with variance σ_0^2 and finite moments of all order, (w_t) is a standard Wiener process, x_0 and (w_t) are mutually independent. Suppose that the process (24) is stable, i.e. the parameter $a < 0$. Note that in this case for every $m \geq 1$

$$\sup_{t\geq 0} \mathsf{E}x_t^{2m} < \infty.$$

We define the truncated estimators of the unknown parameter a

$$a_s = \frac{\int_0^s x_v dx_v}{\int_0^s x_v^2 dv} \chi\Big(\int_0^s x_v^2 dv \geq s \log^{-1} s\Big), \quad s > 0, \tag{25}$$

constructed similarly to the discrete-time case [40] on the basis of the maximum likelihood estimator.

The estimator a_t has the property

$$\mathsf{E}(a_t - a)^{2p} \leq \frac{C}{t^p}.$$

Proof of this property can be found in [6].

3.2.2. Parameter estimation of a multivariate diffusion process. Consider the model

$$dx(t) = \Lambda x(t)dt + dW_t, \quad t \geq 0 \tag{26}$$

with an unknown $p \times p$ matrix parameter Λ, where $x(0)$ is zero mean random vector with variance matrix Σ_0 and finite moments of all order, (W_t) is a multivariate Wiener process with independent components, $x(0)$ and (W_t) are mutually independent. Suppose that the process (26) is stable, i.e. all the eigenvalues of the matrix Λ have negative real parts. Note that in this case for every $m \geq 1$

$$\sup_{t\geq 0} \mathsf{E}\|x(t)\|^{2m} < \infty.$$

Consider the estimation problem of Λ with guaranteed accuracy.

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

$$\hat{\Lambda}'_T = \overline{G}_T^{-1}\overline{\Phi}_T, \quad T > 0,$$

where

$$\overline{G}_T = \frac{1}{T} G_T, \quad G_T = \int_0^T x(t)x'(t)dt,$$

$$\overline{\Phi}_T = \frac{1}{T}\Phi_T, \quad \Phi_T = \int_0^T x(t)dx'(t), \quad T > 0.$$

Define the matrix

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

According to the general notation, in this case we have

$$\Psi = \Lambda', \quad \Psi_T = \hat{\Lambda}'_T,$$

$$f_T = \overline{G}_T^{+}\overline{\Phi}_T, \quad g_T = \overline{\Delta}_T$$

and $\tilde{\Psi}_T = \tilde{\Lambda}'_T$,

$$\tilde{\Lambda}'_T = \hat{\Lambda}'_T \cdot \chi(\overline{\Delta}_T \geq H). \tag{27}$$

Using stability of the process (26) it is easy to verify (see, e.g. [42]) that with P_Λ-probability one

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

where F is a positive definite $p \times p$ matrix.

Then

$$f = \Lambda' \Delta, \quad g = \Delta.$$

We show in Appendix that there exists a given number C_Λ such that for every $T > 0$

$$\sup_{\Lambda \in \Lambda_0} \mathsf{E}_\Lambda \|\tilde{\Lambda}_T - \Lambda\|^2 \leq \frac{C_\Lambda}{T}, \tag{28}$$

where Λ_0 is a compact set from the stable region of the process $(x(n))$.

3.2.3. Parameter estimation of one-parameter delay differential equation. Assume $w = w_t$, $t \geq 0$ is a real-valued standard Wiener process, b is a real number and $x = (x_t,\ t \geq -r)$ is a solution of the stochastic delay differential equation

$$dx_t = bx_{t-r} + dw_t,\ t \geq 0 \tag{29}$$

with some fixed initial condition $x_t = X_0(t),\ t \in [-r,\ 0]$, where $X_0(\cdot)$ is a cadlag stochastic process independent of $w(\cdot)$. Note that the process (29) is stable when the parameter $b \in (-\pi/2r,\ 0)$, see [10].

The solution x of (29) exists, it is pathwise uniquely determined and can be represented as (see, e.g., [10, 31])

$$x_t = x_0(t)X_0(0) + b\int_{-r}^{0} x_0(t-s-r)X_0(s)ds + \int_0^t x_0(t-s)dw_s, \quad t \geq 0.$$

Obviously, it has continuous paths for $t \geq 0$ with probability one and, conditionally on $X_0(\cdot)$, x is a Gaussian zero mean process. Here $x_0(t),\ t \geq -r$ denotes the so-called fundamental solution of the deterministic equation

$$\dot{x}_0(t) = bx_0(t-r),\ x_0(0) = 1 \text{ and } x_0(t) = 0, \quad t \in [-r, 0).$$

The truncated estimator of the unknown parameter b can be defined on the basis of the MLE as follows

$$b_t = \frac{\int_r^t x_{v-r}dx_v}{\int_r^t x_{v-r}^2 dv} \chi\left(\int_r^t x_{v-r}^2 dv \geq t\log^{-1} t\right), \quad t > r. \tag{30}$$

Define the number $\sigma_0^2 = \int\limits_0^\infty x_0^2(v)dv$.

Estimators (30) have the properties (see Appendix)

$$\mathsf{E}(b_t - b)^{2m} \le \frac{C \log^{2m} t}{t^m}, \quad m \ge 1. \tag{31}$$

Estimation problems for stochastic differential equations with time delay have been considered using asymptotic and sequential approaches in a few papers up until now – see, e.g., [10, 17, 18] and the references therein.

3.2.4. Parameter estimation of a stable non-Gaussian Ornstein–Uhlenbeck process by discrete-time observations. The results presented below allow statistical inferences for continuous-time stochastic systems by finite sample size of observations. Moreover, one of the main assumptions is a discrete scheme of observations. It corresponds to numerous real situations, in particular in problems of financial mathematics.

Consider the following regression model

$$dx(t) = ax(t)dt + d\xi(t), \quad 0 \le t \le T \tag{32}$$

with an initial condition $x(0) = x_0$ having finite moments of all order. Here $\xi(t) = \rho_1 W(t) + \rho_2 Z(t)$, $\rho_1 \ne 0$ and ρ_2 are some constants, $(W(t),\ t \ge 0)$ is a standard Wiener process on a probability space $(\Omega,\ \mathcal{F},\ \{\mathcal{F}_t\}_{t\ge 0},\ P)$, adapted to a filtration $\{\mathcal{F}_t\}_{t\ge 0}$, $Z(t) = \sum\limits_{k=1}^{N_t} Y_k$, where $Y_k,\ k \ge 1$ are i.i.d.r.v.'s with finite moments of all order and (N_t) is a Poisson process with the intensity $\lambda > 0$.

It should be noted that for $\rho_2 = 0$ the process (32) is an Ornstein–Uhlenbeck process.

We assume that the unknown parameter lies within the interval $a \in [-\Delta, -\delta]$, where δ and $\Delta > \delta$ are known positive numbers.

The problem is to estimate the parameter a by observations of the discrete-time process $y = (y_k)$

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

Using the representation for the solution of the equation (32) we get the recurrent equation for the observations (y_k):

$$y_k = by_{k-1} + \eta_k, \quad k = \overline{1, n}, \tag{33}$$

where $b = \mathrm{e}^{aT/n}$, $\eta_k = \int\limits_{t_{k-1}}^{t_k} \mathrm{e}^{a(t_k - s)} d\xi(s)$ are i.i.d.r.v.'s with

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

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

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

$$\tilde{a}_n = \frac{n}{T} \log \tilde{b}_n, \quad n \ge 1, \tag{34}$$

where the estimator $\tilde{b}_n$ is constructed on the basis of the LSE $\hat{b}_n$, obtained using the equation (33)

$$\tilde{b}_n = \hat{b}_n \cdot \chi(g_n \ge H) + L\chi(g_n < H), \quad \hat{b}_n = \frac{f_n}{g_n}.$$

Here $L = [\mathrm{e}^{-\delta T/n} + \mathrm{e}^{-\Delta T/n}]/2$,

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

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 (4). In particular, according to Theorem 2.1, which holds for this model for all $m \ge 1$ and $\mu \ge 1$, the following inequalities

$$\sup_{a \le -\delta} \mathsf{E}_a (\tilde{b}_n - b)^{2m} \le \frac{C(m)}{n^m} + \frac{C(\mu)}{n^\mu}, \quad n \ge 1 \tag{35}$$

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

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

and numbers $C(m)$, $C(\mu)$ are known.

Using (34) and (35) it is easy to verify the following property of estimators $\tilde{a}_n$ for every $m \geq 1$ and $\mu > m$:

$$\sup_{a \in [-\Delta, -\delta]} \mathsf{E}_a (\tilde{a}_n - a)^{2m} \leq (nT^{-1} \mathrm{e}^{\Delta T/n})^{2m} \left\{ \frac{C(m)}{n^m} + \frac{C(\mu)}{n^\mu} \right\}, \quad n \geq 1.$$

Proofs of results of this section can be found in [40].

4. Optimal adaptive prediction.

4.1. Discrete-time systems.

4.1.1. Optimal adaptive prediction of VAR(1). Consider the problem of optimal adaptive one-step prediction for the vector process (9). It is well 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.

$$x^{opt}(k) = \Lambda x(k-1), \quad k \geq 1.$$

Substituting Λ with its estimator $\widetilde{\Lambda}_k$ defined in (10) one obtains the one-step predictors of the form

$$\widetilde{x}(k) = \widetilde{\Lambda}_{k-1} x(k-1), \quad k \geq 1,$$

for which the corresponding prediction errors have the following form

$$\widetilde{e}(k) = x(k) - \widetilde{x}(k) = (\Lambda - \widetilde{\Lambda}_{k-1}) x(k-1) + \xi(k).$$

Let $e^2(n)$ denote the sample mean of squared prediction error

$$e^2(n) = \frac{1}{n} \sum_{k=1}^{n} \|\widetilde{e}(k)\|^2.$$

Define the loss function

$$L_n = \frac{A}{n} e^2(n) + n,$$

where the parameter $A(> 0)$ is the cost of prediction error. Such a loss function formulates the problem of choosing between empirical mean-squared prediction accuracy versus costs of increasing the sample size. Define the risk function

$$R_n = \mathsf{E}_\theta L_n = \frac{A}{n} \mathsf{E}_\theta e^2(n) + n.$$

Here $\theta = (\lambda_{11}, \dots, \lambda_{pp}, \sigma^2)$, where $\sigma^2 = \mathsf{E}\|\xi(1)\|^2$.

Using the property (11) it can be shown that

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

if $\mathsf{E}\|\xi(1)\|^{8p} < \infty$, $\mathsf{E}\|x(0)\|^{8p} < \infty$. Minimization of this expression by n yields the optimal sample size of the form

$$n_A^o = A^{1/2}\sigma \tag{36}$$

and the corresponding approximate minimal risk value

$$R_{n_A^o} \approx 2A^{1/2}\sigma,$$

where $\sigma := \sqrt{\sigma^2}$.

The requirements can be further refined to $\mathsf{E}\|\xi(1)\|^{4p} < \infty$, $\mathsf{E}\|x(0)\|^{4p} < \infty$ by using $\mathrm{proj}_{\mathcal{B}}\widetilde{\Lambda}_k$ instead of $\widetilde{\Lambda}_k$, where $\mathcal{B}$ is a closed ball that contains the stability region of the matrix parameter Λ.

However the expression for n_A^o is of little practical use as it contains the unknown parameter σ. For this reason one replaces optimal sample size n_A^o with an estimate of the following form

$$T_A = \inf_{n \geq n_A} \left\{ n \geq A^{1/2}\widetilde{\sigma}_n \right\}, \tag{37}$$

where n_A is the initial sample size depending on A and specified below,

$$\widetilde{\sigma}_n^2 = \frac{1}{n}\sum_{k=1}^{n} \|x(k) - \widetilde{\Lambda}_n x(k-1)\|^2.$$

The modified risk takes the form

$$R_A = \mathsf{E}_\theta L_{T_A} = A\mathsf{E}_\theta \frac{1}{T_A} e^2(T_A) + \mathsf{E}_\theta T_A.$$

Theorem 4.1. *Let* $\mathsf{E}\|\xi(1)\|^{8p+4} < \infty$, $\mathsf{E}\|x(0)\|^{8p+4} < \infty$ *and* n_A *in* (37) *be such that*

$$n_A = o(A^{1/2}) \quad \text{as} \quad A \to \infty, \qquad n_A \geq \max\{k_0,\ A^r \log^2 A\}, \qquad r \in [2/5,\ 1/2).$$

Then the following holds

$$\frac{T_A}{n_A^o} \xrightarrow[A\to\infty]{} 1 \quad \mathsf{P}_\theta\text{-}a.s., \qquad \frac{\mathsf{E}_\theta T_A}{n_A^o} \xrightarrow[A\to\infty]{} 1, \qquad \frac{R_A}{R_{n_A^o}} \xrightarrow[A\to\infty]{} 1.$$

For proof of Theorem 4.1 see Section 4.2 in [22].

Assertions of Theorem 4.1 establish the asymptotic equivalence of T_A and n_A^o, as well as of R_A and $R_{n_A^o}$.

Remark 3. Note that similar result holds for the scalar first-order autoregression under lower restrictions to the model's parameters (see Remark 2).

Since in Section 3.2.3 the observed process satisfies the scalar equation of first-order autoregression then similar result on adaptive optimal prediction can be obtained by using the estimators presented in Sections 3.1.1 and 3.1.3.

4.1.2. Optimal adaptive prediction of VRCA(1). Consider the problem of optimal adaptive one-step prediction for the VRCA(1) process (12). The predictions and prediction errors are defined as follows

$$\widetilde{x}(k) = \widetilde{\Lambda}_{k-1} x(k-1), \quad k \geq 1,$$

where $\widetilde{\Lambda}_k$ is defined in (13),

$$\widetilde{e}(k) = x(k) - \widetilde{x}(k) = (\Lambda - \widetilde{\Lambda}_{k-1}) x(k-1) + \eta(k-1) x(k-1) + \xi(k),$$

$$e^2(n) = \frac{1}{n} \sum_{k=1}^{n} \|\widetilde{e}(k)\|^2.$$

The loss and risk functions are

$$L_n = \frac{A}{n} e^2(n) + n, \quad R_n = \frac{A}{n} \mathsf{E}_\theta e^2(n) + n.$$

Using the property (16) it can be shown that if $\mathsf{E}\|\xi(1)\|^{8p} < \infty$, $\mathsf{E}\|x(0)\|^{8p} < \infty$ and the matrix $\mathsf{E}\Lambda_0^{\otimes 8p}$ is stable then

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

where

$$\sigma^2 = \sigma_\xi^2 + \operatorname{tr}(\Sigma_\eta F).$$

The optimal sample size then has the form analogous to that of (36)

$$n_A^o = A^{1/2} \sigma.$$

Define the stopping time

$$T_A = \inf_{n \geq n_A} \left\{ n \geq A^{1/2} \widetilde{\sigma}_n \right\}, \tag{38}$$

where n_A is some function of A defined below and

$$\widetilde{\sigma}_n^2 = \frac{1}{n}\sum_{k=1}^{n} \|x(k) - \widetilde{\Lambda}_n x(k-1)\|^2.$$

Denote

$$R_A = \mathsf{E}_\theta L_{T_A} = A\mathsf{E}_\theta \frac{1}{T_A} e^2(T_A) + \mathsf{E}_\theta T_A.$$

Theorem 4.2. *Let* $\mathsf{E}\|\xi(1)\|^{8p+4} < \infty,\ \mathsf{E}\|x(0)\|^{8p+4} < \infty$. *Let also* n_A *in* (38) *be such that*

$$n_A = o(A^{1/2}) \ \ as \ \ A \to \infty, \quad n_A \geq \max\{k_0,\ A^r \log^2 A\}, \quad r \in [2/5,\ 1/2).$$

Then the following holds

$$\frac{T_A}{n_A^o} \xrightarrow[A\to\infty]{} 1 \quad \mathsf{P}_\theta\text{-}a.s., \qquad \frac{\mathsf{E}_\theta T_A}{n_A^o} \xrightarrow[A\to\infty]{} 1, \qquad \frac{R_A}{R_{n_A^o}} \xrightarrow[A\to\infty]{} 1 \tag{39}$$

for every $\theta \in \Theta_{8p+4}$, *where* $\Theta_m = \{\theta :\ \mathsf{E}\Lambda_0^{\otimes m}$ *is stable,* $0 < \sigma_\xi^2,\ \sigma_\eta^2 < \infty\}$.

See Appendix for the proof of theorem.

4.1.3. Optimal adaptive prediction of VARMA(1). Consider the problem of optimal adaptive one-step prediction for the VARMA(1,1) process (17). Assume that the matrix parameter M is known. Then the predictions and prediction errors are defined as follows

$$\widetilde{x}(k) = \widetilde{\Lambda}_{k-1} x(k-1) + \mathrm{M}\widetilde{\xi}(k-1), \quad k \geq 1,$$

$$\widetilde{e}(k) = x(k) - \widetilde{x}(k) = (\Lambda - \widetilde{\Lambda}_{k-1})x(k-1) + \mathrm{M}(\xi(k-1) - \widetilde{\xi}(k-1)) + \xi(k),$$

where the estimators $\widetilde{\Lambda}_k$ were defined in (18) and

$$\widetilde{\xi}(k) = \sum_{i=0}^{k-1} (-\mathrm{M})^i \big(x(k-i) - \widetilde{\Lambda}_k x(k-1-i)\big).$$

Define the loss and risk functions

$$L_n = \frac{A}{n} e^2(n) + n, \quad e^2(n) = \frac{1}{n}\sum_{k=1}^{n} \|\widetilde{e}(k)\|^2,$$

$$R_n = \frac{A}{n}\mathsf{E}_\theta e^2(n) + n.$$

Analogously to the previous subsections, if $\mathsf{E}\|\xi(0)\|^{8p} < \infty$, $\mathsf{E}\|x(0)\|^{8p} < \infty$, then it can be shown that

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

and the optimal sample size is

$$n_A^o = A^{1/2}\sigma.$$

The corresponding approximate minimal risk value is

$$R_{n_A^o} \approx 2A^{1/2}\sigma.$$

If σ is unknown one defines the stopping time of the form

$$T_A = \inf_{n \ge n_A} \left\{ n \ge A^{1/2}\widetilde{\sigma}_n \right\}, \tag{40}$$

where n_A is a function of A defined later, and the estimator of the parameter σ^2 is defined as follows

$$\widetilde{\sigma}_n^2 = J_p \frac{1}{n} \sum_{k=1}^{n} (I + \mathrm{M}^{\otimes 2})^{-1} \times \mathrm{vec}\big[\big(x(k) - \widetilde{\Lambda}_n x(k-1)\big)\big(x(k) - \widetilde{\Lambda}_n x(k-1)\big)'\big],$$

where

$$J_p = \langle j_i \rangle_{1\times p^2}, \qquad j_i = \Big\{1, \quad i = 1 + (l-1)(p+1) \text{ for } l = \overline{1,p};\ 0 \quad \text{otherwise}\}.$$

Here $\langle j_i \rangle_{1\times p^2}$ denotes a row vector of the length p^2 with elements j_i.

Define

$$R_A = A\mathsf{E}_\theta \frac{1}{T_A} e^2(T_A) + \mathsf{E}_\theta T_A.$$

Theorem 4.3. *Let* $\mathsf{E}\|\xi(1)\|^{8p+4} < \infty,\ \mathsf{E}\|x(0)\|^{8p+4} < \infty$. *Let also* n_A *in* (40) *be such that*

$$n_A = o(A^{1/2}) \quad as \quad A \to \infty, \quad n_A \ge \max\{k_0,\ A^r \log^2 A\}, \quad r \in [2/5,\ 1/2).$$

Then the following holds

$$\frac{T_A}{n_A^o} \xrightarrow[A\to\infty]{} 1 \quad \mathsf{P}_\theta\text{-}a.s., \qquad \frac{\mathsf{E}_\theta T_A}{n_A^o} \xrightarrow[A\to\infty]{} 1, \qquad \frac{R_A}{R_{n_A^o}} \xrightarrow[A\to\infty]{} 1.$$

Proof of Theorem 4.3 is analogous to that of Theorem 4.1. See also [20], where Theorem 4.3 is proved in a more specific case of uncorrelated and identically distributed components $\xi_j(k),\ j = \overline{1,p}$ of noises $\xi(k)$. This condition allows one to use a simpler form of the estimators $\widetilde{\sigma}_n^2$ and apart from that difference the proofs are identical.

4.2. Continuous-time systems.

4.2.1. Prediction of the Ornstein–Uhlenbeck process. Consider the model (24). The problem is to construct a predictor for x_t by observations $x^{t-u} = (x_s)_{0\le s\le t-u}$ which is optimal in the sense of the risk function introduced below. Here $u > 0$ is a fixed time delay.

Using the solution of (24) we obtain the following representation

$$x_t = \lambda x_{t-u} + \xi_{t,t-u}, \quad t \ge u, \tag{41}$$

where $\xi_{t,t-u} = \int\limits_{t-u}^{t} e^{a(t-s)} dw_s,\ \lambda = e^{au}$. Applying properties of the Ito integral it is easy to make sure that

$$\mathsf{E}\xi_{t,t-u} = 0, \qquad \sigma^2 := \mathsf{E}\xi^2_{t,t-u} = \frac{1}{2a}[\lambda^2 - 1].$$

Optimal in the mean square sense predictor x_t^0 for x_t is the conditional mathematical expectation of x_t under the condition of x^{t-u} which can be found by (41)

$$x_t^0 = \lambda x_{t-u}, \quad t \ge u.$$

Since the parameters a and λ are unknown, we define the adaptive predictor

$$\hat{x}_t = \lambda_{t-u} x_{t-u}, \quad t \ge u, \tag{42}$$

where $\lambda_s = e^{\hat{a}_s u},\ \hat{a}_s = \mathrm{proj}_{(-\infty,0]} a_s$, estimator a_s is defined in (25).

Define the prediction errors of x_t^0 and $\hat{x}_t$ as

$$e_t^0 = x_t - x_t^0 = \xi_{t,t-u},$$

$$e_t = x_t - \hat{x}_t = (\lambda - \lambda_{t-u})x_{t-u} + \xi_{t,t-u}, \quad t \ge u.$$

Now we define the loss function

$$L_t = \frac{A}{t} e^2(t) + t, \quad t \ge u,$$

where $e^2(t) = \dfrac{1}{t}\displaystyle\int_u^t e_s^2 ds$ and the parameter $A > 0$ is the cost of prediction error.

We also define the risk function $R_t = \mathsf{E}L_t$ which has the following form

$$R_t = \frac{A}{t}\mathsf{E}e^2(t) + t$$

and consider optimization problem

$$R_t \to \min_t .$$

For the optimal predictors x_t^0 it is possible to optimize the corresponding risk function

$$R_t^0 = \mathsf{E}\left(\frac{A}{t}(e^0(t))^2 + t\right) = \frac{A\sigma^2}{t} + t \to \min_t, \tag{43}$$

where $(e^0(t))^2 = \frac{1}{t}\int\limits_u^t (e_s^0)^2 ds.$

In this case the optimal duration of observations T_A^0 and the corresponding value of R_t^0 are respectively

$$T_A^0 = A^{\frac{1}{2}}\sigma, \quad R_{T_A^0}^0 = 2A^{\frac{1}{2}}\sigma, \tag{44}$$

where $\sigma := \sqrt{\sigma^2}$.

However, since a and consequently σ are unknown and both T_A^0 and $R_{T_A^0}^0$ depend on a, the optimal predictor can not be used. Then we define the estimator T_A of the optimal time T_A^0 as

$$T_A = \inf\{t \geq t_A : t \geq A^{1/2}\sigma_{t_A}\}, \tag{45}$$

where $t_A := A^{1/2} \cdot \log^{-1} A = o(A^{1/2})$. Here $\sigma_t := \sqrt{\sigma_t^2}$ is the estimator of unknown σ, where

$$\sigma_t^2 = \frac{1}{2}\theta_t \cdot [\lambda_t^2 - 1] \tag{46}$$

and θ_t is the truncated estimator of $\theta = a^{-1}$ defined as follows

$$\theta_t = a_t^{-1} \cdot \chi[a_t \leq -\log^{-1} t], \quad t > 0.$$

Estimators a_t, λ_t and σ_t, which are used in adaptive predictors, have the properties given in Lemma 4.1 below which will be proved in Appendix.

Lemma 4.1. *Assume the model* (24)*. Then the estimators* a_t, λ_t *and* σ_t *are strongly consistent. Moreover, for* $t - u > s_0 := \exp(2|a|)$ *the following properties hold*

$$\mathsf{E}(a_t - a)^{2p} \leq \frac{C}{t^p} \tag{47}$$

and

$$\mathsf{E}(\lambda_t - \lambda)^{2p} \le \frac{C}{t^p}, \quad p \ge 1, \tag{48}$$

$$\mathsf{E}(\sigma_t^2 - \sigma^2)^{2p} \le \frac{C \log^{2p} t}{t^p}, \quad p \ge 1. \tag{49}$$

Analogously to [22, 39] and [38], our purpose is to prove the asymptotic equivalence of T_A and T_A^0 in the almost surely and mean senses and the optimality of the presented adaptive prediction procedure in the sense of equivalence of R_A^0 and the obviously modified risk

$$\overline{R}_A = A \cdot \mathsf{E}\frac{1}{T_A} e^2(T_A) + \mathsf{E}T_A. \tag{50}$$

Theorem 4.4. *Assume the model* (24) *and* t_A *that is defined in* (45)*. Let the predictors* $\hat{x}_t$ *be defined by* (42)*, the times* T_A^0*,* T_A *and the risk functions* R_t^0*,* $\overline{R}_A$ *defined by* (44)*,* (45) *and* (43)*,* (50) *respectively. Then for every* $a < 0$

$$\frac{T_A}{T_A^0} \xrightarrow[A\to\infty]{} 1 \quad \mathsf{P}_\theta\text{-}a.s., \qquad \frac{\mathsf{E}T_A}{T_A^0} \xrightarrow[A\to\infty]{} 1, \qquad \frac{\overline{R}_A}{R_A^0} \xrightarrow[A\to\infty]{} 1.$$

Proof of Theorem 4.4 can be found in [7].

4.2.2. Prediction of the multivariate diffusion process. Consider the model (26). The problem is to construct a predictor for $x(t)$ defined in (26) by observations $x^{t-u} = (x(s))_{0\le s\le t-u}$ which is optimal in the sense of the risk function introduced below. Here $u > 0$ is a fixed time delay.

Using the solution of (26) we obtain the following representation

$$x(t) = Bx(t-u) + \xi_{t,t-u}, \quad t \ge u, \tag{51}$$

where $\xi_{t,t-u} = \int\limits_{t-u}^{t} e^{\Lambda(t-s)} dW_s$, $B = e^{\Lambda u}$. Applying properties of the Ito integral it is easy to verify that

$$\mathsf{E}\xi_{t,t-u} = 0, \qquad \sigma^2 := \mathsf{E}\|\xi_{t,t-u}\|^2 = \int\limits_0^u \|e^{\Lambda s}\|^2 ds.$$

Optimal in the mean square sense predictor $x^0(t)$ for $x(t)$ is the conditional mathematical expectation of $x(t)$ under the condition of x^{t-u} which can be found using (51)

$$x^0(t) = Bx(t-u), \quad t \geq u.$$

Since the parameters Λ and B are unknown, we define the adaptive predictor

$$\hat{x}(t) = B_{t-u}x(t-u), \quad t \geq u, \tag{52}$$

where B_{t-u} is the estimator of B defined as follows

$$B_t = \mathrm{e}^{\Lambda_t u},$$

where $\Lambda_t = \mathrm{proj}_{\Lambda_0}\tilde{\Lambda}_t$, Λ_0 is a compact set from the stability region of the process (26), $\tilde{\Lambda}_t$ is defined in (27).

Denote the prediction errors of x_t^0 and $\hat{x}_t$ as

$$e^0(t) = x(t) - x^0(t) = \xi_{t,t-u},$$

$$e(t) = x(t) - \hat{x}(t) = (B - B_{t-u})x_(t-u) + \xi_{t,t-u}, \quad t \geq u.$$

Now we define the loss function

$$L_t = \frac{A}{t}\overline{e}^2(t) + t, \quad t \geq u,$$

where

$$\overline{e}^2(t) = \frac{1}{t}\int_u^t \|e(s)\|^2 ds$$

and the parameter $A > 0$ is the cost of prediction error.

We also define the risk function $R_t = \mathsf{E}L_t$ which has the following form

$$R_t = \frac{A}{t}\mathsf{E}\overline{e}^2(t) + t$$

and consider optimization problem

$$R_t \to \min_t.$$

For the optimal predictors $x^0(t)$ it is possible to optimize the corresponding risk function

$$R_t^0 = \mathsf{E}\left(\frac{A}{t}(e^0(t))^2 + t\right) = \frac{A\sigma^2}{t} + t \to \min_t, \tag{53}$$

where $(e^0(t))^2 = \frac{1}{t}\int\limits_u^t (e_s^0)^2 ds.$

In this case the optimal duration of observations T_A^0 and the corresponding value of R_t^0 are respectively

$$T_A^0 = A^{\frac{1}{2}}\sigma, \quad R^0_{T_A^0} = 2A^{\frac{1}{2}}\sigma, \tag{54}$$

where $\sigma := \sqrt{\sigma^2}$.

However, since Λ and consequently σ are unknown and both T_A^0 and $R^0_{T_A^0}$ depend on a, the optimal predictor can not be used in practice. We define the estimator T_A of the optimal time T_A^0 as

$$T_A = \inf\{t \geq t_A : t \geq A^{1/2}\hat{\sigma}_{t_A}\}, \tag{55}$$

where $t_A := A^{1/2} \cdot \log^{-1} A = o(A^{1/2})$. Here $\hat{\sigma}_t := \sqrt{\sigma_t^2}$ is the estimator of unknown σ^2, where

$$\hat{\sigma}_t^2 = \int\limits_0^u \|\mathrm{e}^{\Lambda_t s}\|^2 ds.$$

Estimators Λ_t, B_t and $\hat{\sigma}_t$ that are used in the adaptive predictors have the properties given in Lemma 4.2 below which will be proved in Appendix.

Lemma 4.2. *Assume the model* (26)*. Then the estimators* Λ_t, B_t *and* $\hat{\sigma}_t$ *are strongly consistent. Moreover, for* t *large enough the following properties hold*

$$\mathsf{E}\|\Lambda_t - \Lambda\|^{2p} \leq \frac{C}{t^p} \tag{56}$$

and

$$\mathsf{E}\|B_t - B\|^{2p} \leq \frac{C}{t^p}, \quad p \geq 1, \tag{57}$$

$$\mathsf{E}(\hat{\sigma}_t^2 - \sigma^2)^{2p} \leq \frac{C\log^{2p} t}{t^p}, \quad p \geq 1. \tag{58}$$

Analogously to [22, 38, 39], our aim is to prove the asymptotic equivalence of T_A and T_A^0 in the almost surely and mean senses and the optimality of the presented adaptive prediction procedure in the sense of equivalence of R_A^0 and the obviously modified risk

$$\overline{R}_A = A \cdot \mathsf{E}\frac{1}{T_A} e^2(T_A) + \mathsf{E}T_A. \tag{59}$$

Theorem 4.5. *Assume the model* (26) *and* t_A *that is defined in* (55). *Let the predictors* $\hat{x}_t$ *be defined by* (52), *the times* T_A^0, T_A *and the risk functions* R_t^0, $\overline{R}_A$ *defined by* (54), (55) *and* (53), (59) *respectively. Then as for every* $a < 0$

$$\frac{T_A}{T_A^0} \xrightarrow[A\to\infty]{} 1 \qquad \mathsf{P}_\theta\text{-}a.s., \qquad \frac{\mathsf{E}T_A}{T_A^0} \xrightarrow[A\to\infty]{} 1, \qquad \frac{\overline{R}_A}{R_A^0} \xrightarrow[A\to\infty]{} 1.$$

The proof of Theorem 4.5 is similar to that of Theorem 4.4.

4.2.3. One-parameter delay differential equation. Consider the model (29). We construct optimal and adaptive predictors for the process (29). Optimal in the mean square sense predictor is the conditional mathematical expectation

$$z_t^{(k)}(t-u) = \mathsf{E}(x_t | x_{t-u}),$$

which satisfies the following equation

$$\begin{aligned} z_t^{(k)}(t-u) &= x_{t-u} + b \int\limits_{t-u}^{t-(u-r)\wedge t} x_{v-r} dv + b \int\limits_{t-(u-r)\wedge t}^{t} z_{v-kr}^{(0)} dv + \\ &+ b \sum_{i=1}^{k-1} \int\limits_{t}^{t-r} z_{v-(k-i)r}^{(i)} dv, \;\; kr < u \le (k+1)r, \;\; t > u. \end{aligned} \tag{60}$$

Here $\alpha \wedge \beta$ denotes the minimum between α and β.

Since the parameter b in the definition of the optimal predictors $z_t^{(k)}(t-u)$ is unknown, we define the adaptive predictor by formula (60) replacing the unknown b with $\hat{b}_{t-u}$, where $\hat{b}_{t-u}$ is the projection

$$\hat{b}_{t-u} = \mathrm{proj}_{[-\pi/2r,0]} b_{t-u}$$

of the truncated estimator of the parameter b which is proposed in (30).

Define the numbers $\sigma_0^2 = \int\limits_r^\infty x_0^2(v) dv$ and $s_0 = \max\{r, \exp(\sigma_0^{-2})\}$.

Denote the adaptive prediction error and rewrite it in the form

$$e_t^{(k)}(t-u) := x_t - \hat{z}_t^{(k)}(t-u) = e_t^0 + \hat{e}_t^{(k)}(t-u),$$

where $e_t^0(t-u) = x_t - \mathsf{E}(x_t|x_{t-u})$ and $\hat{e}_t^{(k)}(t-u) = z_t^{(k)}(t-u) - \hat{z}_t^{(k)}(t-u)$. Then for every fixed $k \geq 0$ the following limit inequalities hold

$$\overline{\lim_{t\to\infty}} t(\mathsf{E}(e_t^{(k)}(t-u))^2 - \sigma_0^2) \leq C.$$

If it is known that $b \in [b_0, b_1]$, $-\pi/2r < b_0 < b_1 < 0$, then for $t-u > s_1 = \max\{r, \exp(\sigma_1^{-2})\}$, where $\sigma_1^2 = \inf\limits_{b\in[b_0,b_1]} \sigma_0^2$ the non-asymptotic property is fulfilled

$$\mathsf{E}(e_t^{(k)}(t-u))^2 - \sigma_0^2 \leq \frac{C}{t}.$$

These properties mentioned in [6] can be used to prove optimality in the sense of considered above risk functions.

4.2.4. Stable non-Gaussian Ornstein–Uhlenbeck process by discrete-time observations. Results presented in Sections 3.2.4, 4.2.1 make it possible to solve the prediction problem for the process defined in (32) in every point.

For some $u \in (0,1]$ define the numbers $s_k = (k-1+u)h,\ h = T/n$. We introduce the process

$$z = (z_k)_{k\geq 0}, \quad z = x(s_k).$$

Using the representation for the solution of the equation (32) we get the equation for the observations (z_k, y_k)

$$z_k = b_u \cdot y_{k-1} + \eta_{k,u}, \quad \eta_{k,u} = \int_{t_{k-1}}^{s_k} e^{a(s_k-t)} d\xi(t), \tag{61}$$

$$b_u = e^{auh}, \qquad E\eta_{k,u} = 0, \qquad \sigma_u^2 = E\eta_{k,u}^2 = \frac{1}{2a}(\rho_1^2 + \lambda\rho_2^2)[b_u^2 - 1].$$

The adaptive optimal prediction problem can be solved similarly to Section 4.1.1 for predictors

$$\hat{z}_k = \tilde{b}_{u,k-1} y_{k-1}, \qquad \tilde{b}_{u,k-1} = \hat{b}_{u,k-1} \cdot \chi\left[\sum_{i=1}^k y_{i-1}^2 \geq k \log^{-1} k\right],$$

where $\hat{b}_{u,k-1}$ is the LSE obtained from the equation (61).

The risk function can be defined analogously to Section 4.1.1 with prediction errors

$$e_k = \hat{z}_k - z_k.$$

Properties presented in Theorem 4.1 hold for the predictors $\hat{z}_k$ as well.

Appendix.

P r o o f o f L e m m a 3.2 is similar to that of Lemma 3.1. For this reason below we present those proof parts that are essentially different between the two. See also [21] for the proof in scalar case.

For this proof we first need to establish the properties of $\mathsf{E}_\theta x(k)x'(k)$. Solving the equation (12) yields

$$x(k) = \sum_{i=0}^{k-1} \prod_{j=1}^{i} \Lambda_{k-j}\xi(k-i) + \prod_{j=1}^{k} \Lambda_{k-j}x(0)$$

and thus,

$$\begin{aligned}\mathsf{E}_\theta x(k)x'(k) = \mathsf{E}_\theta \sum_{i=0}^{k-1} \prod_{j=1}^{i} \Lambda_{k-j}\xi(k-i)\xi'(k-i) \prod_{j=i}^{1} \Lambda'_{k-j} + \\ + \mathsf{E}_\theta \prod_{j=1}^{k} \Lambda_{k-j}x(0)x'(0) \prod_{j=k}^{1} \Lambda'_{k-j},\end{aligned} \tag{62}$$

here products $\prod$ are ordered, i.e. $\prod_{j=k}^{1} \Lambda'_{k-j} = \Lambda'_0 \cdot \Lambda'_1 \cdot \ldots \cdot \Lambda'_{k-1} \neq \prod_{j=1}^{k} \Lambda'_{k-j}$.

To further break down the resulting expression we will use matrix vectorization operator vec[·], which has the following property (see, e.g., [32])

$$\operatorname{vec}[VYZ] = (Z' \otimes V)\operatorname{vec}[Y]. \tag{63}$$

Applying vec[·] and its property (63) to (62), one obtains

$$\begin{aligned}\operatorname{vec}[\mathsf{E}_\theta x(k)x'(k)] = \sum_{i=0}^{k-1} \mathsf{E}_\theta \Bigg(\prod_{j=1}^{i} \Lambda_{k-j} \Bigg)^{\otimes 2} \operatorname{vec}[\Sigma] + \\ + \mathsf{E}_\theta \Bigg(\prod_{j=1}^{k} \Lambda_{k-j} \Bigg)^{\otimes 2} \operatorname{vec}[\mathsf{E}x(0)x'(0)].\end{aligned} \tag{64}$$

Using the Kronecker product's property $(S \otimes V) \cdot (Y \otimes Z) = S \cdot Y \otimes V \cdot Z$ we have $\left(\prod_{j=1}^{i} \Lambda_{k-j}\right)^{\otimes 2} = \prod_{j=1}^{i} \Lambda_{k-j}^{\otimes 2}$ and since $\eta(k),\ k \geq 1$, are independent, then

$$\mathsf{E}_\theta \left(\prod_{j=1}^{i} \Lambda_{k-j}\right)^{\otimes 2} = \prod_{j=1}^{i} \mathsf{E}_\theta \Lambda_{k-j}^{\otimes 2} = \overline{\Lambda}^i,$$

where $\overline{\Lambda} = \mathsf{E}\Lambda_0^{\otimes 2}$. Recall that $\overline{\Lambda}$ is stable (see conditions in 3.1.4). This allows us to simplify (64) in the following manner

$$\begin{aligned} \text{vec}[\mathsf{E}_\theta x(k)x'(k)] &= \sum_{i=0}^{k-1} \overline{\Lambda}^i \,\text{vec}[\Sigma] + \overline{\Lambda}^k \,\text{vec}[\mathsf{E}x(0)x'(0)] = \\ &= (I - \overline{\Lambda})^{-1}(I - \overline{\Lambda}^k) \,\text{vec}[\Sigma] + \overline{\Lambda}^k \,\text{vec}[\mathsf{E}x(0)x'(0)]. \end{aligned} \tag{65}$$

Letting $k \to \infty$ we get

$$\text{vec}[F] = (I - \overline{\Lambda})^{-1} \,\text{vec}[\Sigma]. \tag{66}$$

Using this equality as well as (65), it can be shown that

$$\sum_{k=0}^{\infty} \|\mathsf{E}_\theta x(k)x'(k) - F)\|^2 \leq C, \tag{67}$$

$$\frac{1}{n} \sum_{k=1}^{n} \left| \text{tr}(\Psi(\mathsf{E}_\theta x(k-1)x'(k-1) - F)) \right| = O(n^{-1}), \qquad n \to \infty. \tag{68}$$

Now we proceed to prove the first assertion (15) of Lemma 3.2. Denote $z_1(k) = \text{vec}[x(k)x'(k)],\ k \geq 1$. From (12) and (63) it follows that the equation for $z_1(k)$ can be written as

$$z_1(k) = \Lambda_{k-1}^{\otimes 2} z_1(k-1) + \epsilon_1(k),$$

$$\epsilon_1(k) = (\xi(k) \otimes \Lambda_{k-1})x(k-1) + (\Lambda_{k-1} \otimes \xi(k))x(k-1) + \text{vec}[\xi(k)\xi'(k)].$$

Examine $\|z_1(k)\|_1$, where $\|a\|_1 = \sum_{i=1}^{p} |a_i|$. The condition necessary for $\sup_{k \geq 1} \mathsf{E}_\theta \|z_1(k)\|_1 < \infty$ is stability of the matrix $\mathsf{E}\Lambda_0^{\otimes 2}$ as well as finiteness of $\xi(1)$

and $x(0)$ (see (65)). If we then consider the process $z_2(k) = \mathrm{vec}[z_1(k)z_1'(k)]$, which satisfies the equation

$$z_2(k) = \Lambda_{k-1}^{\otimes 4} z_2(k-1) + \epsilon_2(k),$$

it turns out that $\sup_{k\geq 1} \mathsf{E}_\theta \|z_2(k)\|_1 < \infty$, if the matrix $\Lambda_0^{\otimes 4}$ is stable and the fourth moments of $\xi(1)$, $x(0)$ are finite. Continuing in this manner and using the following obvious inequality

$$\sup_{k\geq 1} \mathsf{E}_\theta \|x(k)\|^2 \leq \sup_{k\geq 1} \mathsf{E}_\theta \|\mathrm{vec}[x(k)x'(k)]\|_1,$$

we find that conditions of the lemma guarantee

(69) $$\sup_{k\geq 0} \mathsf{E}_\theta \|x(k)\|^{4mp} \leq C, \quad \sup_{k\geq 0} \mathsf{E}_\theta \Delta_k^{2m} \leq C.$$

Using (69) one can verify the following inequality

(70) $$\mathsf{E}_\theta \|\widetilde{\Lambda}_k - \Lambda\|^{2m} \leq \frac{1}{H_k^{2m}} \mathsf{E}_\theta \|\overline{\zeta}_k \overline{F}_k^{+}\|^{2m} + \|\Lambda\|^{2m} \mathsf{P}_\theta(\overline{\Delta}_k < H_k),$$

where

$$\overline{F}_k^{+} = \overline{\Delta}_k \overline{F}_k^{-1}, \quad k \geq p,$$

$$\overline{\zeta}_k = \frac{1}{k}\sum_{i=1}^{k} \eta(i-1)x(i-1)x'(i-1) + \frac{1}{k}\sum_{i=1}^{k} \xi(i)x'(i-1).$$

Since $x(i-1)$ and $\eta(i-1)$ are mutually independent, the process $(\overline{\zeta}_k)_{k\geq 1}$ is, in fact, a sum of two martingales. Thus, analogously to (4.4) in [22], it can be shown that $\mathsf{E}_\theta \|\overline{\zeta}_k\|^{2mp} \leq C \cdot k^{-mp}$. Then for $k \geq p$ the inequality

(71) $$\frac{1}{H_k^{2m}} \mathsf{E}_\theta \|\overline{\zeta}_k \overline{F}_k^{+}\|^{2m} \leq \frac{C \ln^m k}{k^m}$$

follows from Hölder's inequality, and hence (15) holds.

To tackle $\mathsf{E}_\theta(\overline{\Delta}_k - \Delta)^{2m}$, where $\Delta = \det(F)$, one needs to determine the properties of $(\overline{F}_k - F)$. We will use the following identity

(72) $$\mathrm{vec}[F] = (I - \Lambda^{\otimes 2})^{-1}(\mathrm{vec}[\Sigma] + \overline{\eta}\,\mathrm{vec}[F]),$$

where $\overline{\eta} = \mathsf{E}\eta^{\otimes 2}(0)$. To obtain it we use the representation (66) as follows

$$(I - \Lambda^{\otimes 2})^{-1}(\mathrm{vec}[\Sigma] + \overline{\eta}\,\mathrm{vec}[F]) = (I - \Lambda^{\otimes 2})^{-1}(\mathrm{vec}[\Sigma] + \overline{\eta}(I - \overline{\Lambda})^{-1}\mathrm{vec}[\Sigma]) =$$
$$= (I - \Lambda^{\otimes 2})^{-1}\big((I - \overline{\Lambda})(I - \overline{\Lambda})^{-1} + \overline{\eta}(I - \overline{\Lambda})^{-1}\big)\mathrm{vec}[\Sigma] =$$
$$= (I - \Lambda^{\otimes 2})^{-1}\big(I - (\Lambda^{\otimes 2} + \overline{\eta}) + \overline{\eta}\big)(I - \overline{\Lambda})^{-1}\mathrm{vec}[\Sigma] = (I - \overline{\Lambda})^{-1}\mathrm{vec}[\Sigma] = \mathrm{vec}[F].$$

Using the definition (14) of $\overline{F}_k$ it can be shown that

$$\overline{F}_k = \Lambda\overline{F}_k\Lambda' + \Sigma + \frac{1}{k}\sum_{i=1}^{k}\eta(k-1)x(k-1)x'(k-1)\eta'(k-1) + S_k, \tag{73}$$

where

$$S_k = \frac{1}{k}(x(k)x'(k) - x(0)x'(0)) + m_{1,k} + m_{2,k} + m'_{2,k},$$
$$m_{1,k} = \frac{1}{k}\sum_{i=1}^{k}(\xi(i)\xi'(i) - \Sigma), \quad m_{2,k} = \frac{1}{k}\sum_{i=1}^{k}\xi(i)x'(i-1)\Lambda'_{k-1}.$$

Note that S_k has a martingale structure. One can show, analogously to (4.4) in [22], that

$$\mathsf{E}_\theta\|S_k\|^{2m} \le C\cdot k^{-m}. \tag{74}$$

Solving the equation (73) yields

$$\overline{F}_k = \sum_{n\ge 0}\Lambda^n\Big(\Sigma + \frac{1}{k}\sum_{i=1}^{k}\eta(k-1)x(k-1)x'(k-1)\eta'(k-1) + S_k\Big)(\Lambda')^n.$$

Applying vec[$\cdot$] to both sides and taking into account its properties as well as the identity $(\Lambda^n)^{\otimes 2} = (\Lambda^{\otimes 2})^n$, one gets

$$\mathrm{vec}[\overline{F}_k] = \sum_{n\ge 0}(\Lambda^{\otimes 2})^n\,\mathrm{vec}\Big[\Sigma + \frac{1}{k}\sum_{i=1}^{k}\eta(k-1)x(k-1)x'(k-1)\eta'(k-1) + S_k\Big] =$$
$$= (I-\Lambda^{\otimes 2})^{-1}\Big(\mathrm{vec}[\Sigma] + \frac{1}{k}\sum_{i=1}^{k}\eta^{\otimes 2}(k-1)\,\mathrm{vec}[x(k-1)x'(k-1)] + \mathrm{vec}[S_k]\Big).$$

From this representation and (72) it follows that

$$\mathrm{vec}[\overline{F}_k - F] = (I-\Lambda^{\otimes 2})^{-1}\Big(\frac{1}{k}\sum_{i=1}^{k}\big(\eta^{\otimes 2}(k-1)\,\mathrm{vec}[x(k-1)x'(k-1)] - \tag{75}$$
$$-\,\overline{\eta}\;\mathsf{E}_\theta\,\mathrm{vec}[x(k-1)x'(k-1)]\big) + \frac{1}{k}\sum_{i=1}^{k}\overline{\eta}\,\mathrm{vec}[\mathsf{E}_\theta x(k-1)x'(k-1) - F] + \mathrm{vec}[S_k]\Big).$$

The first summand inside the parentheses is a normalized martingale, for which the following holds, analogously to (74),

$$\frac{1}{k^{2m}}\mathsf{E}_\theta\Bigg(\sum_{i=1}^{k}\big(\eta^{\otimes 2}(k-1)\operatorname{vec}[x(k-1)x'(k-1)] - \\ - \mathsf{E}_\theta\eta^{\otimes 2}(k-1)\operatorname{vec}[x(k-1)x'(k-1)]\big)\Bigg)^{2m} \le \frac{C}{k^m}.$$

The second summand is non-random. From (65)-(67) we get

$$\frac{1}{k^{2m}}\Bigg(\sum_{i=1}^{k}\overline{\eta}\operatorname{vec}[\mathsf{E}_\theta x(k-1)x'(k-1) - F]\Bigg)^{2m} \le \frac{C}{k^m}.$$

Then from (75) as well as (69) and (74) it follows that

$$\mathsf{E}_\theta\|\overline{F}_k - F\|^{2m} \le \frac{C}{k^m}. \tag{76}$$

The second assertion of Lemma 3.2 follows from (70), (71), Chebyshev's inequality and (76). Lemma 3.2 is proven. □

P r o o f o f L e m m a 3.3 is similar to that of Lemma 3.1. For such k that G_k is non-singular we can write

$$\hat{\Lambda}_k - \Lambda = \frac{1}{\overline{\Delta}_k}\overline{\zeta}_k\overline{G}_k^+, \quad k \ge 2,$$

where

$$\overline{\zeta}_k = \frac{1}{k-1}\sum_{i=2}^{k}\big(\xi(i) + \mathrm{M}\xi(i-1)\big)x'(i-2), \quad \overline{G}_k^+ = \overline{\Delta}_k\overline{G}_k^{-1}.$$

Hence

$$\mathsf{E}_\theta\|\widetilde{\Lambda}_k - \Lambda\|^{2m} \le \frac{1}{H_k^{2m}}\mathsf{E}_\theta\|\overline{\zeta}_k\overline{G}_k^+\|^{2m} + \|\Lambda\|^{2m}\mathsf{P}_\theta(\overline{\Delta}_k < H_k). \tag{77}$$

The martingale structure of $\overline{\zeta}_k$ allows one to prove

$$\frac{1}{H_k^{2m}}\mathsf{E}_\theta\|\overline{\zeta}_k\overline{F}_k^+\|^{2m} \le \frac{C\ln^m k}{k^m}, \tag{78}$$

and this together with (77) proves (20).

To prove the second assertion (21) one needs to study the properties of $\|\overline{G}_k - G\|$. Using (17) we can write the equation for $\overline{G}_k$ as follows

$$\overline{G}_k = \Lambda \overline{F}_{k-1} + \mathrm{M}\Sigma + S_{1,k-1}, \qquad \overline{F}_k = \sum_{n\geq 0} \Lambda^n (\widetilde{\Sigma} + S_{2,k})(\Lambda')^n,$$

where $\widetilde{\Sigma}$ is defined in (19) and

$$S_{1,k} = \frac{1}{k}\sum_{i=2}^{k+1} \xi(i-1)x'(i-2) + \frac{1}{k}\sum_{i=2}^{k+1} \mathrm{M}\big(\xi(i-2)x'(i-2) - \Sigma\big),$$

$$S_{2,k} = \frac{1}{k}\big(x(k)x'(k) - x(0)x'(0)\big) + m_{1,k} + m'_{1,k} + m_{2,k} + m'_{2,k} + m_{3,k},$$

$$m_{1,k} = \frac{1}{k}\sum_{i=1}^{k} \big(\Lambda x(i-1) + \mathrm{M}\xi(i-1)\big)\xi'(i),$$

$$m_{2,k} = \frac{1}{k}\sum_{i=1}^{k} \Lambda\big(x(i-1)\xi'(i-1) - \Sigma\big)\mathrm{M}',$$

$$m_{3,k} = \frac{1}{k}\sum_{i=1}^{k} \big(\xi(i)\xi'(i) - \Sigma\big) + \frac{1}{k}\sum_{i=1}^{k} \mathrm{M}\big(\xi(i-1)\xi'(i-1) - \Sigma\big)\mathrm{M}'.$$

Then

$$\overline{G}_k - G = \Lambda(\overline{F}_{k-1} - F) + S_{1,k-1} = \Lambda \cdot \sum_{n\geq 0} \Lambda^n S_{2,k-1}(\Lambda')^n + S_{1,k-1}.$$

For $\|S_{j,k}\|$, $j = 1, 2$, analogously to (74), it can be shown that $\mathsf{E}_\theta\|S_{j,k}\|^{2m} \leq C \cdot k^{-m}$. Thus, since Λ is a stable matrix,

$$\mathsf{E}_\theta\|\overline{G}_k - G\|^{2m} \leq \frac{C}{k^m}\Big(\sum_{n\geq 0}\|\Lambda^n\|^2\Big)^{2m} + \frac{C}{k^m} \leq \frac{C}{k^m}. \tag{79}$$

Then (21) follows from (77)–(79).

In case the noise covariance matrix Σ is known and the matrix parameter M is unknown, the estimators $\widetilde{\Lambda}_k$ can be used to estimate it as follows

$$\widetilde{M}_k = \big(\overline{\Gamma}'_{1,k} - \widetilde{\Lambda}_n\overline{\Gamma}_{0,k}\big)\Sigma^{-1}, \tag{80}$$

where

$$\overline{\Gamma}_{0,k} = \frac{1}{k}\sum_{i=1}^{k} x(i-1)x'(i-1), \quad \overline{\Gamma}_{1,k} = \frac{1}{k}\sum_{i=1}^{k} x(i-1)x'(i).$$

Lemma 3.3 is proven. $\square$

P r o o f o f L e m m a 3.4. Using the definition (80) of $\widetilde{\mathrm{M}}_k$ we can write its deviation as follows

$$\widetilde{\mathrm{M}}_k - \mathrm{M} = \Bigg(\frac{1}{k}\sum_{i=1}^{k}(\Lambda - \widetilde{\Lambda}_k)x(i-1)x'(i-1) + \frac{1}{k}\mathrm{M}\xi(0)x'(0) +$$

$$+\frac{1}{k}\sum_{i=1}^{k}\xi(i)x'(i-1) + \frac{1}{k}\sum_{i=2}^{k}\mathrm{M}\xi(i-1)x'(i-2)\Lambda' + \tag{81}$$

$$+\frac{1}{k}\sum_{i=2}^{k}\mathrm{M}\xi(i-1)\xi'(i-2)\mathrm{M}' + \frac{\mathrm{M}}{k}\sum_{i=1}^{k}\big(\xi(i-1)\xi'(i-1) - \Sigma\big)\Bigg)\Sigma^{-1}.$$

The last 4 summands in the parentheses (which we denote $S_{j,k},\ j = \overline{1,4}$) are normalized martingales, for which the following can easily be shown $\mathsf{E}_\theta\|S_{j,k}\|^{2m} \leq Ck^{-m}$. Observe also that by lemma's conditions $\mathsf{E}\left\|\frac{1}{k}\mathrm{M}\xi(0)x'(0)\right\|^{2m} \leq Ck^{-2m}$.

For the first summand in (81), using Hölder's inequality and properties of the estimators $\widetilde{\Lambda}_k$, we obtain for $k \geq k_0$

$$\mathsf{E}_\theta\left\|\frac{1}{k}\sum_{i=1}^{k}(\Lambda - \widetilde{\Lambda}_k)x(i-1)x'(i-1)\right\|^{2m} \leq$$

$$\leq \left(\mathsf{E}_\theta\|\Lambda - \widetilde{\Lambda}_k\|^{2m\frac{p+1}{p}}\right)^{\frac{p}{p+1}} \cdot \left(\mathsf{E}_\theta\left\|\frac{1}{k}\sum_{i=1}^{k}x(i-1)x'(i-1)\right\|^{4m(p+1)}\right)^{\frac{1}{p+1}} \leq \frac{C\ln^m k}{k^m},$$

which implies the second assertion (23) of lemma. When $1 \leq k < k_0$ the property (20) ensures that the boundary in the above inequality is some constant C, hence the first assertion (22) of lemma. Lemma 3.4 is proven. $\square$

P r o o f o f t h e p r o p e r t y (28). Consider the deviation of the estimator in the form

$$\tilde{\Lambda}_T' - \Lambda' = (\hat{\Lambda}_T' - \Lambda')\cdot\chi(g_T \geq H) - \Lambda'\cdot\chi(g_T < H) =$$

$$= \overline{G}_T^{-1}\cdot\overline{\zeta}_T\cdot\chi(g_T \geq H) - \Lambda'\cdot\chi(g_T < H) := I_1 + I_2,$$

where $\overline{\zeta}_T = \frac{1}{T}\int_0^T x(t)dW'(t)$.

We estimate the mathematical expectation using the Cauchy-Bunyakovskii inequality and the properties of the Ito integral

$$\mathsf{E}\|I_1\|^{2p} = \mathsf{E}\|\overline{G}_T^{+} \cdot \overline{\zeta}_T\|^{2p} \cdot g_T^{-2p} \cdot \chi(g_T \geq H) \leq H^{-2p} \cdot \mathsf{E}\|\overline{G}_T^{+}\|^{2p} \cdot \|\overline{\zeta}_T\|^{2p} \leq$$

$$\leq H^{-2p} \cdot \sqrt{\mathsf{E}\|\overline{G}_T^{+}\|^{4p} \cdot \mathsf{E}\|\overline{\zeta}_T\|^{4p}} \leq \frac{C}{T^p}.$$

Similarly to [4, 12] we get

$$\mathsf{E}\|I_2\|^{2p} = \|\Lambda\|^{2p} \cdot P(g_T < H) \leq \|\Lambda\|^{2p} \cdot P(|g_T - g| > g - H) \leq C \cdot \mathsf{E}(g_T - g)^{2p} \leq \frac{C}{T^p}.$$

Thus the property (28) is proven. □

P r o o f o f t h e p r o p e r t y (31). Denote

$$b_t - b = \left(\int_r^t x_s dw_s \Big/ \int_r^t x_s^2 ds \right) \cdot \chi \left(\int_r^t x_s^2 ds \geq t \log^{-1} t \right) -$$

$$-b \cdot \chi \left(\int_r^t x_s^2 ds < t \log^{-1} t \right) := I_1 + I_2.$$

It is obvious that

$$\mathsf{E}(b_s - b)^{2m} = \mathsf{E}I_1^{2m} + \mathsf{E}I_2^{2m}.$$

Consider each summand separately. Using Burckholder's (see, e.g. [2]) and Hölder's inequalities, we obtain

$$\mathsf{E}I_1^{2m} \leq \frac{\log^{2m} t}{t^{2m}} \mathsf{E} \left(\int_r^t x_s dw_s \right)^{2m} \leq \frac{\log^{2m} t}{t^{2m}} \mathsf{E} \left(\int_r^t x_s^2 ds \right)^{m} \leq C \frac{\log^{2m} t}{t^m}.$$

For the second term, by Chebyshev's inequality and for $t > \max(r, \exp^{(\sigma_0^2)^{\frac{m}{2}}})$ we obtain

$$\mathsf{E} I_2^{2m} \le b^{2m} P\left(\frac{1}{t}\int_r^t x_s^2 ds < \log^{-1} t\right)^{2m} =$$

$$= b^{2m} P\left(\sigma_0^2 - \frac{1}{t}\int_r^t x_s^2 ds > \sigma_0^2 - \log^{-1} t\right)^{2m} \le$$

$$\le \frac{b^{2m}}{(\sigma_0^2 - \log^{-1} t)^{2m}} \mathsf{E}\left(\frac{1}{t}\int_r^t x_s^2 ds - \sigma_0^2\right)^{2m} \le \frac{C}{t^m}.$$

The last inequality follows from formulae (5.9)–(5.11) in [19]. The property (31) is proven. □

P r o o f o f T h e o r e m 4.2 is similar to that of Theorem 4.1. Below are those proof parts that are essentially different between the two. See also [21] for the proof in scalar case.

From the conditions it follows that for $\theta \in \Theta_{8p+4}$

$$\sup_{k\ge 0} \mathsf{E}_\theta \|x(k)\|^{8p+4} \le C.$$

We now prove the first assertion in (39). Rewrite the estimators $\widetilde{\sigma}_n^2$ in the following form

$$\widetilde{\sigma}_n^2 = \frac{1}{n}\sum_{k=1}^n (\|\xi(k)\|^2 + \|\eta(k-1)x(k-1)\|^2) + W_n + \nu_n,$$

where

$$W_n = \frac{1}{n}\sum_{k=1}^n \|(\widetilde{\Lambda}_n - \Lambda)x(k-1)\|^2, \quad \nu_n = \frac{2}{n}\sum_{k=1}^n \xi'(k)\eta(k-1)x(k-1) -$$

$$-\frac{2}{n}\sum_{k=1}^n (\xi'(k) + x'(k-1)\eta'(k-1))(\widetilde{\Lambda}_n - \Lambda)x(k-1).$$

Next we show that

(82) $$\widetilde{\sigma}_n^2 \xrightarrow[n\to\infty]{} \sigma^2 \quad \mathsf{P}_\theta\text{-a.s.}$$

Using Chebyshev's inequality and (16) we have for some $0 < \beta \le 4+2p^{-1}$ and every $\epsilon > 0$

$$\mathsf{P}_\theta(\|\widetilde{\Lambda}_n - \Lambda\| \ge \epsilon) \le \frac{1}{\epsilon^{2+\beta}}\mathsf{E}_\theta\|\widetilde{\Lambda}_n - \Lambda\|^{2+\beta} \le C\ln^{1+\frac{\beta}{2}} n \cdot n^{-(1+\frac{\beta}{2})}.$$

Therefore, the Borel-Cantelli lemma guarantees that the estimators $\widetilde{\Lambda}_n$ are strongly consistent. Together with (76) this yields the convergence $W_n \xrightarrow[n\to\infty]{\mathsf{P}_\theta\text{-a.s.}} 0$. Analogously it can be shown that $\nu_n \xrightarrow[n\to\infty]{\mathsf{P}_\theta\text{-a.s.}} 0$.

At the same time, from Kolmogorov's strong law of large numbers, (67) and the Borel-Cantelli lemma it follows that

$$\frac{1}{n}\sum_{k=1}^{n}(\|\xi(k)\|^2 + \|\eta(k-1)x(k-1)\|^2) \xrightarrow[n\to\infty]{} \sigma^2 \quad \mathsf{P}_\theta\text{-a.s.}$$

From this and from the convergencies of W_n and ν_n we have (82). Since, by definition (38), $T_A \xrightarrow[A\to\infty]{\mathsf{P}_\theta\text{-a.s.}} \infty$, and thus $\widetilde{\sigma}^2_{T_A} \xrightarrow[A\to\infty]{\mathsf{P}_\theta\text{-a.s.}} \sigma^2$, then

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

the first assertion of the Theorem is proved.

The second and third assertions in (39) are proved analogously to those of Theorem 4.1 with m_n therein (see page 222 in [22]) having a different form, namely

$$m_n = \frac{1}{n}\sum_{k=1}^{n}(\|\xi\|^2 - \sigma_\xi^2) + \frac{1}{n}\sum_{k=1}^{n}\mathrm{tr}(\eta'(k-1)\eta(k-1)x(k-1)x'(k-1)-$$

$$-\mathsf{E}_\theta\eta(k-1)\eta'(k-1)x(k-1)x'(k-1)) + \frac{1}{n}\sum_{k=1}^{n}\mathrm{tr}\left(\Psi(\mathsf{E}_\theta x(k-1)x'(k-1) - F)\right),$$

which is a sum of two normalized martingales and a non-random function of n, decaying as $O(n^{-1})$ (see (68)). The martingale nature of m_n is what makes the two proofs analogous. Theorem 4.2 is proven. □

P r o o f o f L e m m a 4.1. Properties (47),(48) are proved in [6].

Now we prove the assertion (49). To this end we consider the deviation of the estimator (46)

$$|\sigma_t^2 - \sigma^2| = \left|\frac{1}{2}\theta_t \cdot [\lambda_t^2 - 1] - \frac{1}{2}\theta[\lambda^2 - 1]\right| = \left|\frac{1}{2}[\theta_t - \theta] \cdot [\lambda_t^2 - 1] +\right.$$

$$+\frac{1}{2}\theta[\lambda_t^2-1]-\frac{1}{2}\theta[\lambda^2-1]\Big|\le\frac{1}{2}|\theta_t-\theta|+\frac{1}{2}|\theta||\lambda_t^2-\lambda^2|\le\frac{1}{2}|\theta_t-\theta|+|\theta||\lambda_t-\lambda|\le$$

$$\le\left|\frac{1}{a_t}+\frac{1}{a}\right|\cdot\chi[a_t\le-\log^{-1}t]+\frac{1}{|a|}\cdot\chi[a_t>-\log^{-1}t]+|\theta||\lambda_t-\lambda|=I_1+I_2+I_3.$$

Considering each summand separately we obtain

$$|I_1|=\left|\frac{a-a_t}{a_ta}\right|\cdot\chi[a_t\le-\log^{-1}t]\le\frac{\log t}{|a|}|a-a_t|.$$

Whence by property (47)

$$\mathsf{E}I_1^{2p}\le C\log^{2p}t\,\mathsf{E}|a-a_t|^{2p}\le C\frac{\log^{2p}t}{t^p}.$$

By Chebyshev's inequality, we get

$$\mathsf{E}I_2^{2p}\le C\cdot P(a_t>-\log^{-1}t)\le C\cdot P(a_t-a>-\log^{-1}t-a)\le$$

$$\le C\cdot P(|a_t-a|>|a|-\log^{-1}t)\le C\cdot\frac{1}{(|a|-\log^{-1}t)^{2p}}\mathsf{E}(a_t-a)^{2p}\le C\cdot\frac{\log^{2p}t}{t^p}.$$

To finish the prove of Lemma 4.1 it is enough to apply (48) for estimation of $\mathsf{E}I_3^{2p}$. □

Proof of Lemma 4.2. The property (56) is proved above. Let us prove (57). Estimate the norm

$$\|B_t-B\|=\|e^{\Lambda_tu}-e^{\Lambda u}\|\le\|e^{\Lambda u}\|\cdot\|e^{(\Lambda_t-\Lambda)u}-I\|=$$

$$=\|e^{\Lambda u}\|\cdot\|\sum_{k=1}^{\infty}\frac{u^k}{k!}(\Lambda_t-\Lambda)^k\|\le\|e^{\Lambda u}\|\cdot\sum_{k=1}^{\infty}\frac{u^k\|\Lambda_t-\Lambda\|^k}{k!}=$$

$$=\|e^{\Lambda u}\|\cdot u\cdot\|\Lambda_t-\Lambda\|\cdot\sum_{k\ge1}\frac{u^{k-1}\|\Lambda_t-\Lambda\|^{k-1}}{(k-1)!\cdot k}\le$$

$$\le\|e^{\Lambda u}\|\cdot u\cdot\|\Lambda_t-\Lambda\|\cdot e^{u\|\Lambda_t-\Lambda\|}\le C\cdot\|\tilde\Lambda_t-\Lambda\|.$$

Then the mathematical expectation

$$\mathsf{E}\|B_t-B\|^{2p}\le C\cdot\mathsf{E}\|\tilde\Lambda_t-\Lambda\|^{2p}\le\frac{C}{t^p}.$$

We prove now the property (58). The deviation of the estimator has the form

$$\hat{\sigma}_t^2 - \sigma^2 = \int_0^u \left[\|e^{\Lambda_t s}\|^2 - \|e^{\Lambda s}\|^2\right] ds.$$

Consider the integrand

$$\|e^{\Lambda_t s}\|^2 - \|e^{\Lambda s}\|^2 = \operatorname{tr}\left[e^{(\Lambda_t + \tilde{\Lambda}_t^{'})s} - e^{(\Lambda+\Lambda^{'})s}\right]$$

$$= \operatorname{tr}\left\{e^{(\Lambda+\Lambda^{'})s} \cdot \left[e^{(\Lambda_t - \Lambda)s + (\Lambda_t^{'} - \Lambda^{'})s} - I\right]\right\}.$$

Using the trace property $\operatorname{tr}(AB) \leq \|A\| \cdot \|B\|$ and by the definition of the matrix exponent, we obtain

$$\left|\|e^{\Lambda_t s}\|^2 - \|e^{\Lambda s}\|^2\right| \leq \|e^{(\Lambda+\Lambda^{'})s}\| \cdot \|e^{(\Lambda_t-\Lambda)s+(\Lambda_t^{'}-\Lambda^{'})s} - I\| \leq$$

$$\leq \|e^{(\Lambda+\Lambda^{'})s}\| \cdot \|\sum_{k=1}^{\infty} \frac{1}{k!}\left[(\Lambda_t - \Lambda) + (\Lambda_t^{'} - \Lambda^{'})\right]^k s^k\| \leq$$

$$\leq \|e^{(\Lambda+\Lambda^{'})s}\| \cdot \sum_{k=1}^{\infty} \frac{1}{k!} s^k \cdot \|(\Lambda_t - \Lambda) + (\Lambda^{'} - \Lambda^{'})\|^k \leq$$

$$\leq \|e^{(\Lambda+\Lambda^{'})s}\| \cdot 2 \cdot s \cdot \|\Lambda_t - \Lambda\| \cdot \sum_{k=1}^{\infty} \frac{s^{k-1} \cdot 2^k}{(k-1)!} \cdot \|(\Lambda_t - \Lambda\|^{k-1} \leq$$

$$\leq 4s \cdot e^{2\|\Lambda\|s} \cdot \|\Lambda_t - \Lambda\| \cdot e^{\|\Lambda_t - \Lambda\|} \leq C \cdot \|\Lambda_t - \Lambda\|.$$

The last inequality uses boundedness of Λ_t. Thus, we obtain

$$\mathsf{E}(\hat{\sigma}_t^2 - \sigma^2)^{2p} \leq C \cdot E\|\Lambda_t - \Lambda\|^{2p} \leq \frac{C}{t^p}.$$

Lemma 4.2 is proven. □

Acknowledgment. The work I. of T. V. Dogadova, M. I. Kusainov and V. A. Vasiliev, as well as 3.1.3, 4.1.1 of M. I. Kusainov and V. A. Vasiliev, and 3.1.4, 3.1.5, 4.1.2, 4.1.3 of M. I. Kusainov is supported by The Ministry of Education and Science of the Russian Federation, Goszadanie No 2.3208.2017/4.6, also II, 3.1.1, 3.1.2, 3.2.4 of V. A. Vasiliev, 3.2.2, 3.2.3, 4.2 of T. V. Dogadova and V. A. Vasiliev is supported by the Russian Science Foundation under grant No. 17-11-01049 and performed in National Research Tomsk State University.

REFERENCES

[1] T. W. Anderson. The Statistical Analysis of Time Series. New York-London-Sydney, John Wiley & Sons, Inc., 1971.

[2] E. Carlen, P. Kree. L_p-estimates on iterated stochastic integrals. *Ann. Probab.* **19**, *1* (1991), 354–368.

[3] H. Chernoff. Sequential analysis and optimal design. Conference Board of the Mathematical Sciences Regional Conference Series in Applied Mathematics, No. 8, Philadelphia, SIAM, 1972

[4] A. V. Dobrovidov, G. M. Koshkin, V. A. Vasiliev. Non-parametric State Space Models. Heber City, UT, Kendrick Press, 2012.

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

[6] T. V. Dogadova, V. A. Vasiliev. Adaptive prediction of stochastic differential equations with unknown parameters. *Tomsk State University Journal of Control and Computer Science* **38** (2017), 17–23.

[7] T. V. Dogadova, V. A. Vasiliev. On adaptive optimal prediction of Ornstein–Uhlenbeck process. *Appl. Math. Sci.*, **11**, *12* (2017), 591–600.

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

[9] L. Galtchouk, V. Konev. On sequential estimation of parameters in semimartingale regression models with continuous time parameter. *Ann. Statist.* **29**, *5* (2001), 1508–1536.

[10] A. A. Guschin, U. Küchler. Asymptotic inference for a linear stochastic differential equation with time delay. *Bernoulli* **5**, *6* (1999), 1059–1098.

[11] V. V. Konev, T. L. Lai. Estimators with prescribed precision in stochastic regression models. *Sequential Anal.* **14**, *3* (1995), 179–192.

[12] V. V. Konev, S. M. Pergamenshchikov. On the duration of sequential estimation of parameters of stochastic processes in discrete time. *Stochastics* **18**, *2* (1986), 133–154.

[13] V. V. Konev, S. M. Pergamenshchikov. Truncated Sequential Estimation of the Parameters in Random Regression. *Sequential Anal.* **9**, *1* (1990), 19–41.

[14] V. V. Konev, S. M. Pergamenshchikov. On Truncated sequential estimation of the drifting parameter mean in the first order autoregressive model. *Sequential Anal.* **9**, *2* (1990), 193–216.

[15] V. V. Konev, V. A. Vasiliev. Sequential estimation of parameters of dynamic systems in the presence of multiplicative and additive noise in observations. *Avtomat. i Telemekh. 6* (1985), 33–43, (in Russian).

[16] S. E. Vorobeichikov, V. V. Konev. On sequential identification of stochastic systems. *Izv. Akad. Nauk SSSR Tekhn. Kibernet. 4* (1980), 176–182, 223 (in Russian).

[17] U. Küchler, V. A. Vasiliev. On sequential parameter estimation for some linear stochastic differential equations with time delay. *Sequential Anal.* **20**, *3* (2001), 117–146.

[18] U. Küchler, V. A. Vasiliev. On guaranteed parameter estimation of a multiparameter linear regression process. *Automatica J. IFAC* **46**, *4* (2010), 637-646.

[19] U. Küchler, V. A. Vasiliev. On a certainty equivalence design of continuous-time stochastic systems. *SIAM J. Control Optim.* **51**, *2* (2013), 938–964

[20] M. I. Kusainov. On optimal adaptive prediction of multivariate ARMA(1,1) process. *Tomsk State University Journal of Control and Computer Science* **30** (2015), 44–57.

[21] M. I. Kusainov. Risk efficiency of adaptive one-step prediction of autoregression with parameter drift. *Tomsk State University Journal of Control and Computer Science* **32** (2015), 33–43.

[22] M. I. Kusainov, V. A. Vasiliev. On optimal adaptive prediction of multivariate autoregression.*Sequential Anal.* **34**, *2* (2015), 211–234.

[23] R. S. , A. N. Shiryaev. Statistics of Random Processes. I. General Theory. Translated by A. B. Aries. Applications of Mathematics, vol. **5**. New York-Heidelberg, Springer-Verlag, 1977.

[24] L. Ljung. System Identification: Theory for the User. Prentice Hall Information and System Sciences Series. Englewood Cliffs, NJ, Prentice Hall, Inc., 1987.

[25] L. Ljung, T. Söderström. Theory and Practice of Recursive Identification. MIT Press Series in Signal Processing, Optimization, and Control, vol. **4**. Cambridge, Mass.-London, MIT Press, 1983.

[26] A. A. Malyarenko. Estimating the generalized autoregression model parameters for unknown noise distribution. *Avtomat. i Telemekh. 2* (2010), 128–140 (in Russian); English translation in *Autom. Remote Control* **71**, *2* (2010), 291–302.

[27] A. A. Malyarenko, V. A. Vasiliev. On parameter estimation of partly observed bilinear discrete-time stochastic systems. *Metrika* **75**, *3* (2012), 403–424.

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

[29] N. Mukhopadhyay, T. N. Sriram. On sequential comparisons of means of first-order autoregressive models. *Metrika* **39**, *3–4* (1992), 155–164.

[30] N. Mukhopadhyay, B. M. de Silva. Sequential Methods and Their Applications. Boca Raton, Fl, Chapman & Hall/CRC Press, 2009.

[31] A. D. Myschkis, *Linear Differential Equations with Delayed Argument.* Moscow, Russia, Nauka, 1972 (in Russian).

[32] D. F. Nicholls, B. G. Quinn. The estimation of random coefficient autoregressive models. I. *J. Time Ser. Anal.* **1**, *1* (1980), 37–46.

[33] A. A. Novikov. Sequential estimation of the parameters of the diffusion processes. In: Résumé of Reports Presented at Sessions of the Seminar on the theory of Probability and Mathematical Statistics of the Mathematical Institute of the Academy of Sciences of the USSR. *Teor. Veroyatnost. i Primenen.* **16**, *2* (1971), 394–396 (in Russian); English translation in *Theory of Probability and its Applications* **16**, *2* (1971).

[34] D. N. Politis, V. A. Vasiliev. "Sequential kernel estimation of a multivariate regression function," in Proc. of the IX International Conference

"System Identification and Control Problems" SICPRO'12, Moscow, 2012, 996–1009.

[35] D. N. Politis, V. A. Vasiliev. Non-parametric sequential estimation of a regression function based on dependent observations. *Sequential Anal.* **32**, *3* (2013), 243–266.

[36] H. Robbins, S. Monro. A stochastic approximation method. *Ann. Math. Statistics* **22** (1951), 400–407.

[37] A. N. Shiryaev, V. G. Spokoiny. Statistical experiments and decisions. Asymptotic theory. Advanced Series on Statistical Science & Applied Probability, vol. **8**. River Edge, NJ, World Scientific Publishing Co., Inc., 2000.

[38] T. N. Sriram. Sequential estimation of the autoregressive parameter in a first order autoregressive process. *Sequential Anal.* **7**, *2* (1988), 53–74.

[39] T. N. Sriram, R. Iaci. Sequential estimation for time series models. *Sequential Anal.* **33**, 2 (2014), 136–157.

[40] V. A. Vasiliev. A truncated estimation method with guaranteed accuracy. *Ann. Inst. of Statist. Math.* **66**, *1* (2014), 141–163.

[41] V. A. Vasiliev. Guaranteed estimation of logarithmic density derivative by dependent observations. In: Topics in Nonparametric Statistics (Eds M. G. Akritas et al.). Springer Proc. Math. Stat., vol. **74**, New York, Springer, 2014, 347-356.

[42] V. A. Vasiliev, V. V. Konev. On sequential identification of linear dynamic systems in continuous time by noisy in observations. *Problems Control Inform. Theory/Problemy Upravlen. Teor. Inform.* **16**, *2* (1987), 101–112.

International Laboratory of Statistics of Stochastic Processes
and Quantitative Finance
Tomsk State University
36, Lenin ave.
634050 Tomsk, Russia
e-mail: `aurora1900@mail.ru` *(Tatiana V. Dogadova)*
e-mail: `rjrltsk@gmail.com` *(Marat I. Kusainov)*
e-mail: `vas@mail.tsu.ru` *(Vyacheslav A. Vasiliev)*

Received July 3, 2017
Revised February 12, 2018