Proceedings of the 17th World Congress
The International Federation of Automatic Control
Seoul, Korea, July 6-11, 2008

# On sequential parameter estimation of a linear regression process ⋆

Uwe Küchler * Vyacheslav A. Vasiliev **

* Institute of Mathematics, Humboldt University Berlin, Unter den Linden 6, D-10099, Berlin, Germany (Tel: +49 30 2093 5851; e-mail: kuechler@mathematik.hu-berlin.de).
** Department of Applied Mathematics and Cybernetics, Tomsk State University, Lenina 36, 634050 Tomsk, Russia (e-mail: vas@mail.tsu.ru).

**Abstract:** This paper presents a sequential estimation procedure for unknown parameters of a stochastic linear regression. As examples the sequential estimation problem of two dynamic parameters in stochastic linear systems with memory and in autoregressive processes is solved. The estimation procedure is based on the least squares method with weights and yields estimators with guaranteed accuracy in the sense of the $L_q-$norm ($q \geq 2$). The proposed procedure works in the mentioned examples for all possible values of the unknown dynamic parameters on the plane $\mathcal{R}^2$ with the exception of some lines. The asymptotic behavior of the duration of observations is investigated.
It is shown, that the proposed general procedure may be applied to the sequential parameter estimation problem for affine stochastic delay differential equations as well as autoregressive stochastic differential equations of arbitrary order.

## 1. PRELIMINARIES

Affine and more general stochastic differential equations with time delay are widely used to model phenomena in economics, biology, technics and other sciences incorporating time delay, see e.g. Kolmanovskii and Myshkis (1992) and Mohammed (1996) for references. Often one has to estimate underlying parameters of the model from the observations of the running process.

This paper presents a sequential estimator for unknown dynamic parameters in stochastic linear systems including stochastic differential equations (SDE's) with and without memory. In the sequel $W = (W(t))_{t\geq 0}$ denotes a real-valued standard Wiener process on some probability space $(\Omega, \mathcal{F}, P)$ with respect to a filtration $\underline{\mathcal{F}} = (\mathcal{F})_{t\geq 0}$ from $\mathcal{F}$.

We have in mind the following types of equations.

Example I. Consider the stochastic delay differential equation (SDDE) given by

$$dX(t) = \sum_{i=0}^{p} \vartheta_i X(t - r_i)dt + dW(t),\ t \geq 0, \qquad (1)$$

$$X(s) = X_0(s), s \in [-r, 0].$$

The parameters $r_i,\ \vartheta_i,\ i = 0, \ldots, p$ are real numbers with $0 = r_0 < r_1 < \ldots < r_p =: r$ if $p \geq 1$ and $r_0 = r = 0$ if $p = 0$. The initial process $(X_0(s),\ s \in [-r, 0])$ also defined on $(\Omega, \mathcal{F}, P)$, is supposed to be cadlag and all $X_0(s),\ s \in [-r, 0]$ are assumed to be $\mathcal{F}_0-$measurable. Moreover assume that

$$E\int_{-r}^{0} X_0^2(s)ds < \infty.$$

Example II. Consider the stochastic differential equation of an autoregressive type given by

$$dx_t^{(p)} = \sum_{i=0}^{p} \vartheta_i x_t^{(p-i)}dt + dW(t),\ t \geq 0, \qquad (2)$$

$$x_0^{(p-i)} = x^{(p-i)}(0),\ E(x^{(p-i)}(0))^2 < \infty,\ i = \overline{0,p}.$$

We assume that in our examples the parameter $\vartheta$ belongs to some fixed set $\Theta \subset \mathcal{R}^{p+1}$ which coincides with the whole space $\mathcal{R}^{p+1}$ except of some lines, which we shall specify below.

We shall study the problem of estimating the parameters $(\vartheta_i,\ i = 0, \ldots, p)$ in a sequential way, based on continuous observation of $(X(t))$ and $(x_t)$ respectively. Before we shall consider a problem being slightly more general.

For time-delayed systems of the type (1) a sequential estimation procedure for some special chosen sets $\Theta$ of vector parameter $\vartheta = (\vartheta_0, \vartheta_1)'$ has been constructed by Küchler and Vasiliev (2001, 2003, 2005, 2006).

Sequential parameter estimation problems for the drift of diffusions without time delay (2) have been studied e.g. by Novikov (1971), Liptzer and Shiryaev (1977) (case $p = 0$) and for some special chosen sets $\Theta$ ($p \geq 0$) by Konev and Pergamenshchikov (1985, 1992).

⋆ Research was supported by RFBR - DFG 05-01-04004 Grant

Define the $L_q$-norm on the space of random vectors as $||\cdot||_q = (E_\vartheta||\cdot||^q)^{\frac{1}{q}}$, where $||a|| = (\sum_{i=0}^{p} a_i^2)^{\frac{1}{2}}$ and $E_\vartheta$ denotes the expectation under the distribution $P_\vartheta$ with the given parameter $\vartheta \in \Theta$.

We shall construct for every $\varepsilon > 0$ and arbitrary but fixed $q \geq 2$ a sequential estimator $\vartheta^*_\varepsilon$ of $\vartheta \in \Theta$ with $\varepsilon-$accuracy in the sense

$$||\vartheta^*_\varepsilon - \vartheta||_q^2 \leq \varepsilon. \qquad (3)$$

The estimators with such property may be used in various adaptive procedures (control, prediction, filtration).

## 2. PROBLEM SETTING

In this section we shall consider the linear regression model, which is of general character and will in the sequel be used to more specific studies. It is oriented, first of all, on the parameter estimation problem of the linear differential equation with and without time delay. This aim determines specific assumptions below for this model. Note that the asymptotic properties of processes (1) and (2) are essentially different in different subsets of $\Theta$ (see Tables 1-4 below). That is the reason of complexity of the presented below general sequential estimation procedure which works for regressions with similar properties.

Let $X = (X(t))_{t\geq 0}$ be a scalar random process described by the stochastic differential equation

$$dX(t) = \vartheta' a(t)dt + dW(t),\ t \geq 0 \qquad (4)$$

with the initial condition $X(0) = X_0$, where $W = (W(t), \mathcal{F}_t)_{t\geq 0}$ is a real-valued standard Wiener process on $(\Omega, \mathcal{F}, P)$, $X_0$ is $\mathcal{F}_0$-measurable real-valued random variable and $a = (a(t))_{t\geq 0}$ an $\underline{\mathcal{F}}$-adapted $\mathcal{R}^{p+1}$-valued observable cadlag process on $(\Omega, \mathcal{F}, P)$, $\vartheta$ an unknown parameter vector from some non-void subset $\Theta \subseteq \mathcal{R}^{p+1}$, $\vartheta'$ denotes the transposed $\vartheta$.

Let for all $T > 0$ the following integrals be finite:

$$\int_0^T ||a(t)||^2 dt < \infty \quad \text{a.s.}$$

Then, in particular, the observation process $(X(t))$ is well defined for all $t > 0$.

The problem is to estimate the unknown vector $\vartheta$ with given accuracy in the sense (3) from the observation of $(X, a) = (X(t), a(t))_{t\geq 0}$.

In Galtchouk and Konev (2001) has been constructed the general sequential parameter estimation procedure for linear regression model, which can be applied for systems of the type (1) or (2) for some special chosen sets of unknown parameters. The proposed procedure enables them to estimate the parameters with any prescribed mean squares accuracy under appropriate conditions on the regressors $(a(t))$. Among conditions on the regressors there is one limiting the growth of the maximum eigenvalue of the information matrix $\int_0^T a(t)a'(t)dt$ with respect to its minimal eigenvalue if $T$ increase. This condition is slightly stronger than those usually imposed in asymptotic investigations and it is not possible to apply this estimation procedure to continuous-time models of the type (1) and (2) with essentially different behaviour of the eigenvalues (if, for example, the smallest eigenvalue growth linearly and the largest one - exponentially with the observation period). The sequential estimation procedure, presented in this paper works in more extended parametric sets $\Theta$ for SDE and SDDE in comparison of all mentioned above papers.

Let $(V(t))_{t\geq 0}$ be some $\underline{\mathcal{F}}-$adapted and observable matrix process of size $(p+1)\times(p+1)$, which we call weight matrix process and which can be interpreted as an estimator of some possibly unknown deterministic constant non-singular matrix $V$.

The weighted least squares estimator (LSE) of $\vartheta$ for the given observations from $S$ to $T$ has the form:

$$\hat{\vartheta}(S,T) = G^{-1}(S,T)\Phi(S,T),\ T > S > 0, \qquad (5)$$

where

$$\Phi(S,T) = \int_S^T b(s)dX(s), \quad G(S,T) = \int_S^T b(s)a'(s)ds,$$

$b(s) = V(s)a(s)$. Put $\bar{b}(s) = Va(s)$,

$$\Phi(T) = \Phi(0,T), \quad G(T) = G(0,T).$$

Our purpose is to construct sequential estimators of the parameters in the model (4) on the bases of weighted LSE's (5). The weights $(V(t))$ and $V$ will be chosen in such a way that the integrals of the squared components of the processes $(b(t))$ and $(\bar{b}(t))$ will have certain rates of increase (see Assumption (V) below).

For the construction of sequential plans and the investigation of their asymptotic properties the main problem consists in a possible essentially difference of the rates of convergence to infinity of the eigenvalues of the matrix $G(T)$ for $T \to \infty$. Our first steps are devoted to based on construct of $V$ and $(V(t))$, appropriate factors of normalization, which eliminate this deficiency.

Let the weight process $(V(t))_{t\geq 0}$ be such that for all $T > 0$ the following integrals are finite:

$$\int_0^T E||b(t)||^2 dt < \infty, \quad \vartheta \in \Theta.$$

We formulate some assumptions, which are fulfilled, in particular, in both examples mentioned above.

We shall write in the sequel $f(x) \simeq C$ as $x \to \infty$ instead of the limiting relations:

$$0 < \underline{\lim}_{x\to\infty} f(x) \leq \overline{\lim}_{x\to\infty} f(x) < \infty.$$

ASSUMPTION (V): *Fixed a non-void subset $\mathcal{A}$ from $\mathcal{R}^r$ for some $r \geq 1$. We assume, that for every $i = \overline{0,p}$ there exists a family of unboundedly increasing positive functions $(\varphi_i(\alpha, T),\ T > 0)$, $\alpha \in \mathcal{A}$, such that*

$$\overline{\lim_{T\in\infty}}\ \varphi_0(\alpha,T)/\varphi_i(\alpha,T) \leq 1, \quad i = \overline{1,p}, \quad \alpha \in \mathcal{A}$$

*and for some* $\alpha \in \mathcal{A}$ *and all the functions* $\tilde{b}(\cdot)$ *equal to* $b(\cdot)$ or $\overline{b}(\cdot)$ *the following relations*

$$\varphi_i^{-1}(\alpha, T) \int_0^T \tilde{b}_i^2(t)dt \simeq C \text{ as } T \to \infty, \ i = \overline{0,p}$$

*almost surely hold.*

Let $S$ and $T$ be two reals with $0 \le S < T$. The part of observations $(X(s), a(s), \ 0 \le s \le S)$ will be used to estimate $\alpha$, the part $(X(t), a(t), \ S \le t \le T)$ to estimate $\vartheta$.

Define the following sets of positive functions:

$$\mathcal{P}_0 = \{f(\cdot) : \ \frac{f(y(x))}{f(x)} \simeq \ C \text{ if } \frac{y(x)}{x} \ \simeq C, x \uparrow \infty\},$$

$$\mathcal{P}_1 = \{y(\cdot) : \ \exists \ g(\cdot) \in \mathcal{P}_0 \ \text{ with } \varliminf_{T\to\infty} \ g(T) > 0 \text{ and}$$

$$y(S) = o(\overline{g}^{-1/2}(T)y(T)) \quad \text{if} \quad S = o(T), \ T \to \infty\},$$

where $\overline{g}(T) = g(\varphi_0(T))$,

$$\overline{\mathcal{P}}_1 = \{y(\cdot) : \ \exists \ g(\cdot) \in \mathcal{P}_0 \ \text{ with } \varliminf_{T\to\infty} \ g(T) > 0 \text{ and}$$

$$S = o(T) \quad \text{if} \quad y(S) = o(\overline{g}^{-1/2}(T)y(T)), \ T \to \infty\}$$

and introduce the matrices

$$\overline{\varphi}(T) = \text{diag}\{\varphi_0(\alpha, T), \varphi_1(\alpha, T), \dots, \varphi_p(\alpha, T)\},$$

$$\tilde{\varphi}^{\frac{1}{2}}(T) = \overline{\varphi}^{\frac{1}{2}}(T)(V')^{-1},$$

$$\overline{G}(S,T) = \overline{\varphi}^{-\frac{1}{2}}(T)G(S,T)\tilde{\varphi}^{-\frac{1}{2}}(T),$$

$$\tilde{G}(S,T) = \overline{\varphi}^{-\frac{1}{2}}(T)G(S,T)\varphi_0^{-\frac{1}{2}}(T),$$

$$\overline{G}(T) = \overline{G}(0,T), \quad \tilde{G}(T) = \tilde{G}(0,T),$$

$$\overline{\zeta}(S,T) = \overline{\varphi}^{-\frac{1}{2}}(T) \int_S^T b(t)dW(t).$$

Denote $\lambda_{min}\{A\}$ and $\lambda_{max}\{A\}$ the smallest and the largest eigenvalues of the matrix $A$.

For the construction of sequential estimators we shall use the following normalized representation for the deviation of the weighted LSE $\hat{\vartheta}(S,T)$ :

$$\varphi_0^{\frac{1}{2}}(T)(\hat{\vartheta}(S,T) - \vartheta) = \tilde{G}^{-1}(S,T)\overline{\zeta}(S,T). \qquad (6)$$

The matrices $\tilde{G}(S,T)$ and $\overline{G}(S,T)$ have similar asymptotic properties (in the sense of the Assumption (G) below) under the following condition on the functions $\varphi_i(T), \ i = \overline{0,p}$ and on the matrix V:

$$\varliminf_{T\to\infty} \ \lambda_{max}\{V'\overline{\varphi}^{-1}(T)\varphi_0(T)V\} > 0. \qquad (7)$$

ASSUMPTION (G): *Let the functions* $\overline{b}(t)$ and $b(t)$ *satisfy Assumption* $(V)$, *where functions* $\varphi_i(\cdot) \in \mathcal{P}_1, \ i = \overline{0,p}$. *We suppose that the following property for the matrix function* $\overline{G}(T)$ *holds:*

$$\varliminf_{T\to\infty} \overline{g}(T)\lambda_{min}\{\overline{G'}(T)\overline{G}(T)\} > 0 \quad P_\vartheta - \text{a.s.},$$

*where* $\overline{g}(T)$ *is the function from the definitions of the classes* $\mathcal{P}_1$ *and* $\overline{\mathcal{P}}_1$.

It can be shown that Assumption (G) and condition (7) give the possibility to control the behaviour of the matrix $\tilde{G}^{-1}(S,T)$ in the representation (6) of the deviation of the estimator $\hat{\vartheta}(S,T)$ from $\vartheta$ in the construction of sequential estimation plans.

From the properties of matrices $\overline{\zeta}(S,T)$ and $\tilde{G}^{-1}(S,T)$ it follows that the estimator $\hat{\vartheta}(S,T)$ has the rate of convergence $\overline{g}^{-1/2}(T)\varphi_0^{\frac{1}{2}}(T)$.

ASSUMPTION $(\varphi\Psi)$: *Assume* $\varphi_i(\alpha, T), \ i = \overline{0,p}, \ \alpha \in \mathcal{A}$ *are functions as described in Assumption* $(V)$. *We put* $\Psi_0(\alpha, x) = x$ *and suppose, that there exist so-called positive rate generating functions* $\Psi_i(\cdot, \cdot), \ i = \overline{1,p}$, *on* $\mathcal{A} \times (0, \infty)$, *increasing in the last variable* $x$, *and such that for all* $\alpha \in \mathcal{A}$

$$\varphi_i(\alpha, T)/\Psi_i(\alpha, \varphi_0(T)) \simeq C \ \text{as } T \to \infty, \ i = \overline{1,p}.$$

For example, the functions of the form

$$\Psi(\alpha, x) = x^{v_1} \mathrm{e}^{v_0 x}[c_1 \cos \xi x + c_2 \sin \xi x],$$

where $\alpha = (v_0, v_1) \in [\{0\} \times (0, +\infty)] \cup [(0, +\infty) \times (-\infty, +\infty)], \ \xi \in \mathcal{R}^1$ cover all the possible cases of asymptotic behaviour for solutions of linear SDE's and SDDE's.

By the construction of our sequential plans we divide the set of functions $\varphi_0(\alpha, T), \dots, \varphi_p(\alpha, T)$ into some groups $\mathcal{G}_j = \{\varphi_i(\alpha, T), \ i = \overline{s_{j-1} + 1, s_j}\}, \ j = \overline{0,m}$, $-1 = s_{-1} < s_0 \le \dots \le s_m = p, \ 0 \le m \le p$ of the functions which rates of increase do not differ essentially in some sense. Then we shall define $m + 1$ systems of stopping times on the bases of the sums of appropriately normalized integrals $\int_0^T b_i^2(t)dt$, having the rates of increase $\varphi_i(T), \ i = \overline{s_{j-1} + 1, s_j}, \ j = \overline{0,m}, \ 0 \le m \le p$ respectively. To take this aim into account we introduce a "multidimensional time scale" $\overline{T} \in \mathcal{R}^{m+1}$ :

$$\overline{T} = (\underbrace{T_0, \dots, T_0}_{l_0+1}, \underbrace{T_1, \dots, T_1}_{l_1}, \dots, \underbrace{T_m, \dots, T_m}_{l_m}),$$

$l_0 = s_0, l_i = s_i - s_{i-1}, i = \overline{1,m}$ if $m > 0$, $\overline{T} = (\underbrace{T_0, \dots, T_0}_{p+1})$ if $m = 0$. We shall substitute in the following the components of the vector $\overline{T}$ by special stopping times.

Denote $T_{max} = \max_{i=\overline{0,m}} T_i$ and $T_{min} = \min_{i=\overline{0,m}} T_i$. We shall construct our sequential estimation plans on the bases of the estimator $\hat{\vartheta}(S, T_{min})$, which has the rate of convergence equals to $\overline{g}^{-1/2}(T_{min}) \cdot \varphi_0^{\frac{1}{2}}(T_{min})$ if $T_{min} \to \infty$. At the same time for calculation of $\hat{\vartheta}(S, T_{min})$ in the case of random time scales $\overline{T}$ we use the observations of the length $T_{max}$. Then the rate of convergence of $\hat{\vartheta}(S, T_{min})$ equals to $\overline{g}^{-1/2}(T_{min})\varphi_0^{\frac{1}{2}}(T_{max})$. To keep the order of the convergence rate $\overline{g}^{-1/2}(T_{min}) \cdot \varphi_0^{\frac{1}{2}}(T_{min})$ of estimators $\hat{\vartheta}(S, T_{min})$ by the random time change, we introduce the following admissible set for the time-scales $\overline{T}$ :

$$\Upsilon := \{(\overline{T}) : \ \varlimsup_{\overline{T}\to\infty} \varphi_0(T_{max})/\varphi_0(T_{min}) < \infty\}.$$

Now we return to our basic examples.

Example I. Define $p = 1,\ a_0(t) = X(t),\ a_1(t) = X(t-1)$. Then the equation (4) has the form (1):

$$dX(t) = \vartheta_0 X(t)dt + \vartheta_1 X(t-1)dt + dW(t). \tag{8}$$

To introduce the admissible parametric set $\Theta$ for $\vartheta = (\vartheta_0, \vartheta_1)'$ we need following notation.

Let $s = u(r)$ $(r < 1)$ and $s = w(r)$ $(r \in \mathcal{R}^1)$ be the functions given by the parametric representation $(r(\xi), s(\xi))$ in $\mathcal{R}^2$ :

$$r(\xi) = \xi \cot \xi,\ s(\xi) = -\xi / \sin \xi$$

with $\xi \in (0, \pi)$ and $\xi \in (\pi, 2\pi)$ respectively.

Consider the set $\Lambda$ of all (real or complex) roots of the so-called characteristic equation corresponding to (8)

$$\lambda - \vartheta_0 - \vartheta_1 \mathrm{e}^{-\lambda} = 0$$

and put $v_0 = v_0(\vartheta) = \max\{Re\lambda | \lambda \in \Lambda\}$,

$$v_1 = v_1(\vartheta) = \max\{Re\lambda | \lambda \in \Lambda,\ Re\lambda < v_0\}.$$

It can be easily shown that $-\infty < v_1 < v_0 < \infty$. By $m(\lambda)$ we denote the multiplicity of the solution $\lambda \in \Lambda$. Note that $\lambda = \vartheta_0 - 1 \in \Lambda$ if and only if $\vartheta_1 = -e^{\vartheta_0 - 1}$. In this case we have $m(\lambda) = 2$, in all the other cases it holds $m(\lambda) = 1$.

The general estimation procedure will be applied for all parameters $\vartheta$ from the set $\Theta$ defined by

$$\Theta = \Theta_1 \cup \Theta_2 \cup \Theta_3,$$

where

$$\Theta_1 = \Theta_{11} \cup \Theta_{12} \cup \Theta_{13}, \quad \Theta_2 = \Theta_{21} \cup \Theta_{22},$$

with

$$\Theta_{11} = \{\vartheta \in \mathcal{R}^2 |\ v_0(\vartheta) < 0\},$$
$$\Theta_{12} = \{\vartheta \in \mathcal{R}^2 |\ v_0(\vartheta) > 0 \text{ and } v_0(\vartheta) \notin \Lambda\},$$
$$\Theta_{13} = \{\vartheta \in \mathcal{R}^2 |\ v_0(\vartheta) > 0;\ v_0(\vartheta) \in \Lambda,\ m(v_0) = 2\},$$
$$\Theta_{21} = \{\vartheta \in \mathcal{R}^2 |\ v_0(\vartheta) > 0, v_0(\vartheta) \in \Lambda,\ m(v_0) = 1,\ v_1(\vartheta) > 0 \text{ and } v_1(\vartheta) \in \Lambda\},$$
$$\Theta_{22} = \{\vartheta \in \mathcal{R}^2 |\ v_0(\vartheta) > 0, v_0(\vartheta) \in \Lambda,\ m(v_0) = 1,\ v_1(\vartheta) > 0 \text{ and } v_1(\vartheta) \notin \Lambda\},$$
$$\Theta_3 = \{\vartheta \in \mathcal{R}^2 |\ v_0(\vartheta) > 0,\ \ v_0(\vartheta) \in \Lambda,\ \ m(v_0) = 1 \text{ and } v_1(\vartheta) < 0\},$$

and denote, in addition, the sets

$$\Theta_{41} = \{\vartheta \in \mathcal{R}^2 |\ \ v_0(\vartheta) = 0,\ v_0(\vartheta) \in \Lambda,\ m(v_0) = 1\},$$
$$\Theta_{42} = \{\vartheta \in \mathcal{R}^2 |\ \ v_0(\vartheta) = 0,\ v_0(\vartheta) \in \Lambda,\ m(v_0) = 2\},$$
$$\Theta_{43} = \{\vartheta \in \mathcal{R}^2 |\ \ v_0(\vartheta) > 0,\ v_0(\vartheta) \in \Lambda,\ m(v_0) = 1,\ v_1(\vartheta) = 0 \text{ and } v_1(\vartheta) \in \Lambda\}.$$

Note, that this decomposition is very related to the classification used in Gushchin and Küchler (1999), where can be found a figure giving an imagination of these sets.

The parameter set $\Theta$ equals the plane $\mathcal{R}^2$ without the bounds of the set $\Theta_{12} \cup \Theta_{13} \cup \Theta_3$. In particular, $\Theta_{11}$ is the set of parameters $\vartheta$ for which there exists a stationary solution of (8).

Obviously, all sets $\Theta_{11}, \Theta_{12}, \Theta_{13}, \Theta_{21}, \Theta_{22}, \Theta_3$ are pairwise disjoint and the closure of $\Theta$ is the whole $\mathcal{R}^2$. Moreover, the exceptional set is of Lebesque measure zero.

We shall consider the sequential estimation problem for the one-parametric set $\Theta_4 = \Theta_{41} \cup \Theta_{42} \cup \Theta_{43}$ as well. This case is of interest in view of that the set $\Theta_4$ is the bound of the following regions: $\Theta_{11},\ \Theta_{12}, \Theta_{21}, \Theta_3$. In this case $\vartheta_1 = -\vartheta_0$ and (8) can be written as a differential equation of the first order. We do not consider the scalar case $\Theta_4$ as an example of the general estimation procedure because our method is intend for two- or more-parametric models. Moreover, for similar one-parametric model a well-known sequential estimation procedure is constructed and investigated by Novikov (1971) and Liptzer and Shiryaev (1977). We shall use this procedure in Section 3 with applications to the case $\Theta_4$.

Now we define the process $Y_t = X(t) - \lambda_t X(t-1)$, as an estimator of $Y(t) = X(t) - \lambda X(t-1)$; the quantity

$$\lambda_t = \frac{\int_0^t X(s)X(s-1)ds}{\int_0^t X^2(s-1)ds}$$

is a strong consistent estimator of $\lambda = \mathrm{e}^{v_0}$, the parameter $\alpha = (v_0, v_1)$. Then all the Assumptions (V), (G) and $(\varphi\Psi)$ for $\vartheta \in \Theta$ are fulfilled and we have

| Table 1 | | | | |
|---|---|---|---|---|
| Region | $\varphi_0(\alpha, T)$ | $\varphi_1(\alpha, T)$ | $\Psi_1(\alpha, x)$ | $g(\varphi)$ |
| $\Theta_{11}$ | $T$ | $T$ | $x$ | 1 |
| $\Theta_{12}$ | $e^{2v_0T}$ | $e^{2v_0T}$ | $x$ | 1 |
| $\Theta_{13}$ | $T^2e^{2v_0T}$ | $T^2e^{2v_0T}$ | $x$ | $\ln^8 \varphi$ |
| $\Theta_2$ | $e^{2v_1T}$ | $e^{2v_0T}$ | $x^{v_0/v_1}$ | 1 |
| $\Theta_3$ | $T$ | $e^{2v_0T}$ | $\mathrm{e}^{2v_0x}$ | 1 |

if we put

| Table 2 | | | | |
|---|---|---|---|---|
| Region | $b_0(t)$ | $b_1(T)$ | $\bar{b}_0(t)$ | $\bar{b}_1(T)$ |
| $\Theta_{11}$ | $X(t)$ | $X(t-1)$ | $X(t)$ | $X(t-1)$ |
| $\Theta_{12}$ | $X(t)$ | $X(t-1)$ | $X(t)$ | $X(t-1)$ |
| $\Theta_{13}$ | $X(t)$ | $X(t)$ | $X(t)$ | $X(t)$ |
| $\Theta_2$ | $Y_t$ | $X(t)$ | $Y(t)$ | $X(t)$ |
| $\Theta_3$ | $Y_t$ | $X(t)$ | $Y(t)$ | $X(t)$ |

Example II. Define $p = 1,\ X(t) = a_0(t) = \dot{x}_t,\ a_1(t) = x_t$. Then the equation (4) has the form (2):

$$d\dot{x}_t = \vartheta_0 \dot{x}_t dt + \vartheta_1 x_t dt + dW(t),\ t \geq 0. \tag{9}$$

For this model we can define, similar to Example I, following parametric sets with differ asymptotic behaviour of the MLE:

$$\Theta_1^* = \Theta_{11}^* \cup \Theta_{12}^* \cup \Theta_{13}^* \cup \Theta_{14}^*,$$
$$\Theta_{11}^* = \{\vartheta \in \mathcal{R}^2 :\ \vartheta_0 < 0, \vartheta_1 < 0\},$$
$$\Theta_{12}^* = \{\vartheta \in \mathcal{R}^2 :\ \vartheta_0 > 0,\ \vartheta_1 < -(\vartheta_0/2)^2\},$$
$$\Theta_{13}^* = \{\vartheta \in \mathcal{R}^2 :\ \vartheta_0 = 0, \vartheta_1 < 0\},$$
$$\Theta_{14}^* = \{\vartheta \in \mathcal{R}^2 :\ \vartheta_0 > 0, \vartheta_1 = -(\vartheta_0/2)^2\},$$
$$\Theta_2^* = \{\vartheta \in \mathcal{R}^2 :\ \vartheta_0 > 0, -(\vartheta_0/2)^2 < \vartheta_1 < 0\},$$
$$\Theta_3^* = \{\vartheta \in \mathcal{R}^2 :\ \vartheta_1 > 0\}.$$

Then

$$\Theta^* = \Theta_1^* \cup \Theta_2^* \cup \Theta_3^* = \mathcal{R}^2 \setminus \{\vartheta \in \mathcal{R}^2 :\ \vartheta_1 = 0\}.$$

*Remark 2.1.* As usual, the condition $\vartheta_1 \neq 0$ means the knowledge of the order $(p = 1)$ of the process (9). It should be noted that the problem of sequential estimation for the case $\Theta^* \setminus \{\vartheta \in \mathcal{R}^2 : \vartheta_0 = 0\}$ has been solved, in principle, by Konev and Pergamenshchikov (1985, 1992).

Now we define the processes $\tilde{Z}_1(t) = X_1(t) - \hat{v}_0(t)X_2(t)$ and $\tilde{Z}_2(t) = X_1(t) - \hat{v}_1(t)X_2(t)$ estimating of $Z_1(t) = X_1(t) - v_0X_2(t)$ and $Z_2(t) = X_1(t) - v_1X_2(t)$ respectively, $\hat{v}_0(t)$ and $\hat{v}_1(t)$ are some strongly consistent estimators of $v_0 = Re\lambda_0$ and $v_1 = Re\lambda_1$, where $\lambda_i$ are the roots of characteristic polynomial of the observable process (9).

Define $\alpha = (v_0, v_1)$. Then all the Assumptions (V), (G) and $(\varphi\Psi)$ are fulfilled for the functions

Table 3

| Region | $\varphi_0(\alpha,T)$ | $\varphi_1(\alpha,T)$ | $\Psi_1(\alpha,x)$ | $g(\varphi)$ |
|---|---|---|---|---|
| $\Theta^*_{11}$ | $T$ | $T$ | $x$ | 1 |
| $\Theta^*_{12}$ | $e^{2v_0T}$ | $e^{2v_0T}$ | $x$ | 1 |
| $\Theta^*_{13}$ | $T^2$ | $T^2$ | $x$ | 1 |
| $\Theta^*_{14}$ | $T^2e^{2v_0T}$ | $T^2e^{2v_0T}$ | $x$ | $\ln^8 \varphi$ |
| $\Theta^*_2$ | $e^{2v_1T}$ | $e^{2v_0T}$ | $x^{v_0/v_1}$ | 1 |
| $\Theta^*_3$ | $T$ | $e^{2v_0T}$ | $e^{2v_0x}$ | 1 |

if we put

Table 4

| Region | $b_0(t)$ | $b_1(T)$ | $\bar{b}_0(t)$ | $\bar{b}_1(T)$ |
|---|---|---|---|---|
| $\Theta^*_{11}$ | $X_1(t)$ | $X_2(t)$ | $X_1(t)$ | $X_2(t)$ |
| $\Theta^*_{12}$ | $X_1(t)$ | $X_2(t)$ | $X_1(t)$ | $X_2(t)$ |
| $\Theta^*_{13}$ | $X_1(t)$ | $X_2(t)$ | $X_1(t)$ | $X_2(t)$ |
| $\Theta^*_{14}$ | $X_1(t)$ | $X_2(t)$ | $X_1(t)$ | $X_2(t)$ |
| $\Theta^*_2$ | $\tilde{Z}_1(t)$ | $\tilde{Z}_2(t)$ | $Z_1(t)$ | $Z_2(t)$ |
| $\Theta^*_3$ | $\tilde{Z}_1(t)$ | $X_2(t)$ | $Z_1(t)$ | $X_2(t)$ |

## 3. CONSTRUCTION OF SEQUENTIAL ESTIMATION PLANS

Let us return to the study of the equation (4) and assume that the Assumptions (V), (G) and $(\varphi\Psi)$ are valid.

Let $\varepsilon$ be any positive number being fixed in the sequel. Now we construct the sequential estimation plan $\mathrm{SEP}(\varepsilon) = (T(\varepsilon),\ \vartheta^*_\varepsilon)$ where $T(\varepsilon)$ and $\vartheta^*_\varepsilon$ are the duration of estimation and the estimator of $\vartheta$ with the $\varepsilon$-accuracy in the sense of $L_q$-norm respectively.

To construct a sequential estimator $\vartheta^*_\varepsilon$ of $\vartheta$ with preassigned accuracy $\varepsilon > 0$ firstly we introduce a random time substitution for the weighted least square estimator $\hat{\vartheta}(S,T)$ from (5). This enables us to control the moments of the process $\bar{\zeta}(S,T)$ in the representation (6) of its deviation. To do that, we have to take into account the fact, that the $L_q$–norms of the components of the vector $b(\cdot)$ may have different rates of increase. The knowledge of these rates gives the possibility to construct the system of stopping times from the admissible set $\Upsilon$.

For every positive $\varepsilon$ let us fix two unboundedly increasing sequences $(\nu_n(\varepsilon))_{n\ge1}$ and $(c_n)_{n\ge1}$ of positive $\underline{\mathcal{F}}$–stopping times (or real numbers) and real numbers respectively, satisfying the following conditions:

$$\varphi_0(\nu_n(\varepsilon)) = o(g^{-1/2}(\varepsilon^{-1}c_n)\varepsilon^{-1}c_n) \quad P_\vartheta - \text{a.s.} \tag{10}$$

as $n \to \infty$ or $\varepsilon \to 0$,

$$\sum_{n\ge1} c_n^{-q/2} < \infty \tag{11}$$

and for every fixed $\varepsilon > 0$

$$\sum_{n\ge1} g^{-q/2}(\varepsilon^{-1}c_n) = \infty, \tag{12}$$

where $g(\cdot)$ is the function from the definition of the function $\overline{g}(T) = g(\varphi_0(T))$ in Assumption (G).

Assume that $\alpha$ is a parameter of the functions $\varphi_i(\alpha,t),\ i = \overline{0,p}$, which can be estimated consistently by observation of $(X,a)$. It is the case in all of our basic examples.

Denote by $\alpha_i(n,\varepsilon),\ i = \overline{1,r},\ n \ge 1$ an estimators of the parameters $\alpha_i,\ i = \overline{1,r}$, which we assumed to be constructed by the trajectory of the observation process $(X,a)$ of the duration $\nu_n(\varepsilon)$ and define

$$\Psi(\alpha,n,\varepsilon) = \mathrm{diag}\{\varepsilon^{-1}c_n, \Psi_1(\alpha,\varepsilon^{-1}c_n), \Psi_2(\alpha,\varepsilon^{-1}c_n),\ldots,\Psi_p(\alpha,\varepsilon^{-1}c_n)\},$$
$$\tilde{\Psi}(n,\varepsilon) = \Psi(\alpha(n,\varepsilon),n,\varepsilon),\quad \tilde{b}_n(t) = = \tilde{\Psi}^{-1/2}(n,\varepsilon)b(t) = (\tilde{b}_{0n}(t),\ldots,\tilde{b}_{pn}(t))'.$$

ASSUMPTION $(\alpha)$: *Let the condition* (10) *be fulfilled. The estimators $\alpha(n,\varepsilon)$ of the parameter $\alpha$ are supposed to have the properties:*

ASSUMPTION $(\alpha1)$: *for every $\varepsilon > 0$ and $i = \overline{1,p}$*

$$\frac{\tilde{\Psi}_{ii}(n,\varepsilon)}{\Psi_{ii}(\alpha,n,\varepsilon)} \simeq C \quad \textit{as } n \to \infty \quad P_\vartheta - \text{a.s.};$$

ASSUMPTION $(\alpha2)$: *for every $n \ge 1$ and $i = \overline{1,p}$*

$$\frac{\tilde{\Psi}_{ii}(n,\varepsilon)}{\Psi_{ii}(\alpha,n,\varepsilon)} \simeq C \quad \textit{as } \varepsilon \to 0 \quad P_\vartheta - \text{a.s.}$$

Assumption $(\alpha)$ can be verified for time delayed process (8) from Example I and for the autoregressive process (9), considered in Example II.

Let us define the sequences of stopping times $(\tau_j(n,\varepsilon), n \ge 1),\ j = \overline{0,m}$ as follows

$$\tau_j(n,\varepsilon) = \inf\{T > \nu_n(\varepsilon) : \sum_{i=s_{j-1}+1}^{s_j} \left(\int_{\nu_n(\varepsilon)}^{T} \tilde{b}^2_{in}(t)dt\right)^{q/2} = 1\},$$

where $\inf\{\oslash\} = \infty$ and denote

$$\tau_{min}(n,\varepsilon) = \min\{\tau_0(n,\varepsilon),\tau_1(n,\varepsilon),\ldots,\tau_m(n,\varepsilon)\},$$
$$\tau_{max}(n,\varepsilon) = \max\{\tau_0(n,\varepsilon),\tau_1(n,\varepsilon),\ldots,\tau_m(n,\varepsilon)\}.$$

The estimation of the parameter $\vartheta$ should be performed on the intervals $[\nu_n(\varepsilon),\tau_{min}(n,\varepsilon)]$ :

$$\vartheta(n,\varepsilon) = \hat{\vartheta}(\nu_n(\varepsilon),\tau_{min}(n,\varepsilon)),\ n \ge 1.$$

For the construction of sequential plan we put

$$\sigma(\varepsilon) = \inf\{N \ge 1 :\ S(N,\varepsilon) \ge \varrho\},$$

where

$$S(N,\varepsilon) = \sum_{n=1}^{N} \beta^q(n,\varepsilon),$$

with $\beta(n,\varepsilon)$ is defined as $\beta(n,\varepsilon) = ||G_{n,\varepsilon}^{-1}||^{-1}$ if the matrix

$$G_{n,\varepsilon} = (\varepsilon^{-1}c_n)^{-\frac{1}{2}}\tilde{\Psi}^{-\frac{1}{2}}(n,\varepsilon)G(\nu_n(\varepsilon),\tau_{min}(n,\varepsilon))$$

is invertible; 0 in the other case, $\varrho$ is known constant, defined, for example, in Küchler and Vasiliev (2003) and Küchler and Vasiliev (2005).

DEFINITION (D) *The sequential plan* $(T(\varepsilon),\ \vartheta^*_\varepsilon)$ of estimation of the vector $\vartheta \in \Theta$ *will be defined by the formulae*

$$T(\varepsilon) = \tau_{max}(\sigma(\varepsilon),\varepsilon)\ , \tag{13}$$

$$\vartheta^*_\varepsilon = S^{-1}(\sigma(\varepsilon),\varepsilon)\sum_{n=1}^{\sigma(\varepsilon)} \beta^q(n,\varepsilon)\vartheta(n,\varepsilon), \tag{14}$$

*where* $T(\varepsilon)$ *is the duration of estimation, and* $\vartheta^*_\varepsilon$ *is the estimator of* $\vartheta$ *with given accuracy* $\varepsilon > 0$.

The following theorem summarizes the main result concerning the sequential plan $(T(\varepsilon),\ \vartheta^*_\varepsilon)$.

THEOREM 1. *Suppose Assumptions* $(V)$, $(G)$, $(\varphi\Psi)$ and $(\alpha1)$ *hold and the conditions* (10)–(12) *are fulfilled. Then for every* $\varepsilon > 0$ *and every* $\vartheta \in \Theta$ *the sequential plan* $(T(\varepsilon),\vartheta^*_\varepsilon)$ *from Definition (D) is closed, i.e. it holds* $T(\varepsilon) < \infty \quad P_\vartheta - a.s.$

*Moreover, for every* $\vartheta \in \Theta$ *the following statements are true:*

1°. *for any* $\varepsilon > 0$ *it holds*

$$\|\vartheta^*_\varepsilon - \vartheta\|^2_q \le \varepsilon;$$

2°. *if Assumption* $(\alpha2)$ *holds, and if for some known positive function* $h(\cdot)$, such that $h(\varepsilon) = \varepsilon$ *if* $\overline{\lim\limits_{T\to\infty}} g(T) < \infty$ *and* $h(\varepsilon) = o(\varepsilon)$ as $\varepsilon \to 0$ *if* $\overline{\lim\limits_{T\to\infty}} g(T) = \infty$, *then*

$$a)\ \overline{\lim_{\varepsilon\to 0}}\ h(\varepsilon)\cdot\varphi_0(T(\varepsilon)) < \infty \quad P_\vartheta - \text{a.s.},$$

*and, moreover, if the condition* (7) *is valid, then*

$$b)\ \underline{\lim_{\varepsilon\to 0}}\ \varepsilon\cdot\varphi_0(T(\varepsilon)) > 0 \quad P_\vartheta - \text{a.s.};$$

3°. *if* $g(T) = o(T)$ *as* $T \to \infty$ *then the estimator* $\vartheta^*_\varepsilon$ *is strongly consistent:*

$$\lim_{\varepsilon\to 0}\vartheta^*_\varepsilon = \vartheta \quad P_\vartheta - a.s.$$

We have constructed, similarly to Küchler and Vasiliev (2006), sequential estimators for the parameters $\vartheta$ from the sets $\Theta$ and $\Theta^*$ in the both our examples as a linear combinations of estimators of the type (13), (14) (and of the type Novikov (1971), Liptzer and Shiryaev (1977) for the one-parametric case $\Theta_4$). Namely, we have defined the sequential estimation plans $(T_i(\varepsilon),\vartheta_i(\varepsilon)),\ i=\overline{1,4}$ for each of the regions $\Theta_1,\Theta_2,\Theta_3,\Theta_4$ separately in Example I and the sequential plans $(T_i^*(\varepsilon),\vartheta_i^*(\varepsilon)),\ i=\overline{1,3}$ for the regions $\Theta_1^*,\Theta_2^*,\Theta_3^*$ in Example II. Then, because in general it is unknown to which region the parameter $\vartheta$ belongs to, we define the sequential plans $(T(\varepsilon),\vartheta(\varepsilon))$ of estimation $\vartheta \in \Theta$ and $(T^*(\varepsilon),\vartheta^*(\varepsilon))$ of estimation $\vartheta \in \Theta^*$ as combinations of all defined above estimators respectively by the formulae

$$T(\varepsilon) = \min(T_1(\varepsilon),\ldots,T_4(\varepsilon)),$$

$$\vartheta(\varepsilon) = \chi_1(\varepsilon)\vartheta_1(\varepsilon) + \ldots + \chi_4(\varepsilon)\vartheta_4(\varepsilon),$$

where $\chi_i(\varepsilon) = \chi(T(\varepsilon) = T_i(\varepsilon)),\ i=\overline{1,4},\ \chi(a=b) = 1,\ a=b;\ 0,\ a\ne b$ and

$$T^*(\varepsilon) = \min(T_1^*(\varepsilon),T_2^*(\varepsilon)),T_3^*(\varepsilon)),$$

$$\vartheta^*(\varepsilon) = \chi_1(\varepsilon)\vartheta_1^*(\varepsilon) + \chi_2(\varepsilon)\vartheta_2^*(\varepsilon) + \chi_3(\varepsilon)\vartheta_3^*(\varepsilon).$$

Both the obtained general estimators have the properties of estimators from the Definition (D). Moreover, all the rates of increase of the observation periods $T(\varepsilon)$ and $T^*(\varepsilon)$ in the almost sure sense are obtained.

## REFERENCES


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

A. A. Gushchin, and U. Küchler. Asymptotic inference for a linear stochastic differential equation with time delay. *Bernoulli*, volume5, 6, pages 1059–1098, 1999.

V. Kolmanovskii, and A. Myshkis. *Applied Theory of Functional Differential Equations*, Kluwer Acad. Pabl., 1992.

V. V. Konev, and S. M. Pergamenshchikov. Sequential estimation of the parameters of diffusion processes. *Problems of Inform. Trans.*, volume 21, 1, pages 48–62, 1985 (in Russian).

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

U. Küchler, and V. Vasiliev. On sequential parameter estimation for some linear stochastic differential equations with time delay. *Sequential Analysis*, volume 20, 3, pages 117–146, 2001.

U. Küchler, and V. Vasiliev. On sequential identification of a diffusion type process with memory. *Proceedings Symp. Int. Fed. Autom. Contr. SYSID-2003*, Rotterdam, Holland, 27-29 August, pages 1217–1221, 2003.

U. Küchler, and V. Vasiliev. Sequential identification of linear dynamic systems with memory. *Statist. Inference for Stochastic Processes*, volume 8, pages 1–24, 2005.

U. Küchler, and V. Vasiliev. On Sequential Estimators for an Affine Stochastic Delay Differential Equations. *Algorithms for Approximation. Proceedengs of the 5th International Conference*, Chester, July 2005, Editors: A. Iske, J. Levesley. Springer-Verlag, Berlin, Heidelberg, pages 287–296, 2006.

R. S. Liptzer, and A. N. Shiryaev. *Statistics of Random Processes.* Springer-Verlag, New York, Heidelberg, 1977.

S. E-A. Mohammed. *Stochastic differential systems with memory: Theory, examples and applications.* Probability and Statistics. Birkhäuser, volume 42, pages 1–77. 1996.

A. A. Novikov. *The sequential parameter estimation in the process of diffusion type.* Probab. Theory and its Appl., volume 16, 2, pages 394–396, 1971 (in Russian).