We consider the problem of delay estimation by the observations of the solutions of several SDEs. It is known that the MLEs for these models are consistent and asymptotically normal, but the likelihood ratio functions are not differentiable w.r.t. the parameter, and therefore the numerical calculation of the MLEs encounter certain difficulties. We propose One-step and Two-step MLEs, whose calculation has no such problems and provide an estimator asymptotically equivalent to the MLE. These constructions are realized in two or three steps. First we construct preliminary estimators which are consistent and asymptotically normal, but not asymptotically efficient. Then we use these estimators and a modified Fisher-score device to obtain One-step and Two-step MLEs. We suppose that its numerical realization is much more simple. Stochastic Pantograph equation is introduced and related statistical problems are discussed.

MSC 2010 subject classifications: Primary 62G05, 62M09; secondary 62G20.

Keywords: One-step MLE, Two-step MLE, One-step MDE, Stochastic Pantograph equation, Delay estimation.

1. Introduction

We consider two types of models of stochastic differential equation (SDE) with delays, which can be illustrated by the following equations

\[dX_t = S(X_{t-\vartheta}) \, dt + \varepsilon dW_t, \quad X_s = x_0, \quad 0 \leq t \leq T, \quad (1.1)\]

\[dX_t = S(X_{\vartheta t}) \, dt + \varepsilon dW_t, \quad X_0 = x_0, \quad 0 \leq t \leq T. \quad (1.2)\]

We call the delay in the equation (1.1) shift-type, and the one in the equation (1.2) scale-type. Here the functions \(S(x)\) are known and smooth, and the parameter (delay) \(\vartheta \in \Theta\). The set \(\Theta = (\alpha, \beta)\), \(0 < \alpha < \beta < T\) in the case of the equation (1.1), and \(\Theta = (\alpha, \beta)\), \(0 < \alpha < \beta < 1\) in the case of the equation (1.2). The parameter \(\vartheta\) has to be estimated by the observations \(X^T = (X_t, 0 \leq t \leq T)\), and the properties of estimators are studied in the asymptotic of small noise, i.e., as \(\varepsilon \to 0\). Such perturbed dynamical systems were studied in [5], and different statistical problems related to such models can be found in [14]. We suppose that the function \(S(x)\) is Lipschitz

\[|S(x) - S(y)| \leq L |x - y|.

Under this condition, the equations (1.1) and (1.2) have unique strong solutions, and all polynomial moments of the solutions are finite (see Theorem 4.6 in [21]).
Ordinary and stochastic differential equations with delay (feedback with delay) are widely used in many applied sciences (see [23], [2] (medicine), [4], [25], [26] (physics), [8] (neuron science), [10], [22], (population dynamics), as well as the references therein). At the same time, the statistical inference (estimation and testing concerning delays) for these systems does not attract sufficiently the attention of the statisticians.

We illustrate the possible difficulties in the study of these models of observation with the help of the linear Ornstein-Uhlenbeck process with delay

\[ dX_t = -\gamma X_{t-\vartheta}dt + \sigma dW_t, \quad X_s = x_0, \quad s \leq 0, \quad 0 \leq t \leq T. \]  

Consider the problem of estimation of \( \vartheta \in \Theta \) by continuous time observations \( X^T = (X_t, 0 \leq t \leq T) \) from (1.3). We suppose that the values \( \gamma > 0 \) and \( \sigma > 0 \) are known.

Let us write the equation (1.3) in the integral form and change the variables \( t \to t - \vartheta \). Then we obtain the representation

\[ X_{t-\vartheta} = x_0 - \gamma \int_{0}^{t-\vartheta} X_{s-\vartheta} ds + \sigma W_{t-\vartheta}, \quad t \geq \vartheta. \]  

We see that the drift \( S(X_{t-\vartheta}) = -\gamma X_{t-\vartheta} \) in the equation (1.3) is as smooth w.r.t. \( \vartheta \) as the Wiener process w.r.t. the time \( t \), i.e., even the first derivative of the drift \( \partial S(X_{t-\vartheta})/\partial \vartheta \) does not exist. The following question naturally arises: is the problem of estimation of \( \vartheta \) by observations (1.4) regular (family of measures is LAN, Fisher information is finite), or singular (family of measures is not LAN, Fisher information is infinite)? What is amusing with this model of observations, is that we have both these cases depending on the type of the asymptotics.

If the time of observations is fixed and \( \sigma \to 0 \), then the problem of estimation is regular, the family of measures is LAN, and the MLE \( \hat{\vartheta}_\sigma \) is asymptotically normal

\[ \sigma^{-1} \left( \hat{\vartheta}_\sigma - \vartheta \right) \Rightarrow N \left( 0, I(\vartheta)^{-1} \right), \quad I(\vartheta) = \gamma^4 \int_{\vartheta}^{T} x_{t-2\vartheta}^2 dt \]  

(see [13], [14]). Here \( I(\vartheta) \) plays the role of Fisher information and \( x_t = x_t(\vartheta) \) is solution of the limit \( (\sigma = 0) \) equation

\[ \frac{dx_t}{dt} = -\gamma x_{t-\vartheta}, \quad x_s = x_0, \quad s \leq 0, \quad 0 \leq t \leq T. \]  

Note that the observations \( X_t \) on the interval \([0, \vartheta]\) has no any information about the value of \( \vartheta \), and that is why the derivative of \( -\gamma X_{t-\vartheta} = -\gamma x_0 \) w.r.t. \( \vartheta \) is zero.

On the other hand, if we fix \( \sigma \) and consider the asymptotic \( T \to \infty \), then the problem of estimation becomes singular, the family of measures is no more LAN, and the MLE \( \hat{\vartheta}_T \) has another (non Gaussian) limit distribution

\[ Tc \left( \hat{\vartheta}_T - \vartheta \right) \Rightarrow \arg \sup_{u \in \mathcal{R}} \left( W(u) - \frac{|u|}{2} \right). \]
Here $c > 0$ is some constant, and $W(\cdot)$ is a two-sided Wiener process \cite{11}, \cite{15}. Recall that the MLE $\hat{\vartheta}_T$ has similar properties in the case of the observations

$$dX_t = -\gamma \text{sgn}(X_t - \varrho) dt + \sigma dW_t, \quad X_s = x_0, \quad 0 \leq s \leq 0, \quad 0 \leq t \leq T,$$

with discontinuous drift coefficient \cite{15}. Hence, here the delay is equivalent to a discontinuity (change-point in space) in the model. This is the only known to us model, which is at the same time regular and singular (statistically) depending on the type of the asymptotics. The proofs are based on the powerful general results in \cite{9}.

The properties of the MLE $\hat{\vartheta}_\varepsilon$ for the model (1.1) were described in the works \cite{13} (linear case) and \cite{1} (generalization of \cite{13} to the nonlinear case), see as well \cite{14}. It was shown that the MLE is consistent, asymptotically normal

$$\varepsilon^{-1} (\hat{\vartheta}_\varepsilon - \varrho) \Longrightarrow \mathcal{N}(0, 1(\varrho)^{-1}), \quad 1(\varrho) = \int_0^T S'(x_{t-\varrho})^2 S(x_{t-2\varrho})^2 dt,$$

and asymptotically efficient.

The likelihood ratio function for the model (1.1) is

$$L(\varrho, X^T) = \exp \left\{ \int_0^T \frac{S(X_{t-\varrho})}{\varepsilon^2} dX_t - \int_0^T \frac{S(X_{t-\varrho})^2}{2\varepsilon^2} dt \right\}, \quad \varrho \in \Theta,$$

and the MLE $\hat{\vartheta}_\varepsilon$ is defined by the equation

$$L(\hat{\vartheta}_\varepsilon, X^T) = \sup_{\varrho \in \Theta} L(\varrho, X^T). \quad (1.7)$$

As the function $L(\varrho, X^T)$ is not differentiable w.r.t. $\varrho$, the realization of numerical algorithms of calculation of $\hat{\vartheta}_\varepsilon$ can encounter certain difficulties.

The goal of this work is to construct estimators, whose calculation has no such difficulties, and which have the same asymptotic properties as the MLE for both models. This will be realized in two or three steps. First, we construct some preliminary estimators, which have sufficiently good rate of convergence, but are not asymptotically efficient. Then, using these estimators and Fisher-score device, we obtain Le Cam’s One-step MLEs $\vartheta_\varepsilon^\star$. The details and historic notes concerning the one-step MLE can be found in \cite{20} and \cite{3}. It is shown that the estimators $\vartheta_\varepsilon^\star$ are consistent and have the same asymptotic variance as the MLEs. The Fisher-score improvement of the preliminary estimators is a well-known approach in statistics, and we do not give here the review of the corresponding literature. The discussion of the works related with the diffusion processes can be found in \cite{18}.

Recall that the realization of the Fisher-score device requires two derivatives of the trend coefficient w.r.t. the unknown parameter. For example, suppose that the observed process is

$$dX_t = S(\varrho, t) dt + \varepsilon dW_t, \quad X_0 = 0, \quad 0 \leq t \leq T,$$
and that we have a preliminary estimator $\tilde{\vartheta}_\varepsilon$ such that $\varphi^{-1}_\varepsilon (\tilde{\vartheta}_\varepsilon - \vartheta_0)$ is bounded in probability. Then we can write the One-step MLE formally as follows

$$\vartheta^*_\varepsilon = \tilde{\vartheta}_\varepsilon + \int_0^T \frac{S(\tilde{\vartheta}_\varepsilon, t)}{I(\tilde{\vartheta}_\varepsilon)} \left[dX_t - S(\tilde{\vartheta}_\varepsilon, t) \, dt\right], \quad I(\vartheta) = \int_0^T \dot{S}(\vartheta, t)^2 \, dt.$$ 

Hence, using the Taylor formula (below $\tilde{\vartheta}_\varepsilon$ satisfies $|\tilde{\vartheta}_\varepsilon - \vartheta_0| \leq |\tilde{\vartheta}_\varepsilon - \vartheta_0|$)

$$S(\vartheta_0, t) - S(\tilde{\vartheta}_\varepsilon, t) = - (\tilde{\vartheta}_\varepsilon - \vartheta_0) \dot{S}(\tilde{\vartheta}_\varepsilon, t) - \frac{1}{2} (\tilde{\vartheta}_\varepsilon - \vartheta_0)^2 \ddot{S}(\tilde{\vartheta}_\varepsilon, t),$$

we obtain the representation

$$\frac{\vartheta^*_\varepsilon - \vartheta_0}{\varepsilon} = \varepsilon^{-1} (\tilde{\vartheta}_\varepsilon - \vartheta_0) + \frac{1}{\varepsilon} \int_0^T \dot{S}(\tilde{\vartheta}_\varepsilon, t) \, dW_t$$

$$+ \frac{1}{\varepsilon} \int_0^T \frac{\dot{S}(\tilde{\vartheta}_\varepsilon, t)}{I(\tilde{\vartheta}_\varepsilon)} \left[S(\vartheta_0, t) - S(\tilde{\vartheta}_\varepsilon, t)\right] \, dt$$

$$= \int_0^T \frac{\dot{S}(\vartheta_0, t)}{I(\vartheta_0)} \, dW_t - \frac{(\tilde{\vartheta}_\varepsilon - \vartheta_0)^2}{\varepsilon} \int_0^T \frac{\dot{S}(\vartheta_0, t)}{2I(\vartheta_0)} \, dW_t + o(1).$$

We see that if the second derivative is bounded and $\varepsilon^{-1} \varphi^2_\varepsilon \to 0$, then the One-step MLE is asymptotically normal and asymptotically efficient. We say formally, because there is a problem of definition of the stochastic integral. We remind this example to show that in the One-step MLE construction we need the second derivative. The stochastic models considered in the present work have no even first derivatives, and we have to avoid a similar problem with the stochastic integral too.

Therefore the results presented in this work are the following: for the shift-type (1.1) and scale-type (1.2) models we propose One-step MLE’s of the parameter of delays and describe their asymptotic properties. Special attention is paid to a new model called stochastic pantograph equation and for scale-type delay parameter of this model we construct Two-step MLE. We show that this estimator is consistent and asymptotic normal.

Statistical problems of parameter estimation related to different generalizations of the model (1.3) with the asymptotics $\sigma = \varepsilon \to 0$ were treated in the works [1], [13], [14] (see a review in [16]). Some statistical problems for related SDEs with delay in the case of the asymptotics $T \to \infty$ were studied in the works [7], [11], [16].

2. Shift-type delay

As it was mentioned above, we solve the problem of delay estimation in two steps. First we construct a preliminary estimator, which is consistent and asymptotically normal, and then we use this estimator and the Fisher-score device to construct the One-step MLE, which is asymptotically equivalent to the MLE. In the next section we propose and study such a preliminary estimator.
MDE. We have the observations $X^T = (X_t, 0 \leq t \leq T)$ of the solution of the equation (1.1), and we have to estimate $\vartheta \in \Theta = (\alpha, \beta)$, $0 < \alpha < \beta < T$. The true value will be denoted as $\vartheta_0$. The solution $x_t$ of the equation (1.5) for $t \geq \vartheta$ is a function of $\vartheta$ and we write it as $x_t(\vartheta)$.

We define the MDE $\bar{\vartheta}_{\varepsilon}$ as solution of the equation

$$
\|X - x(\bar{\vartheta}_{\varepsilon})\| = \inf_{\vartheta \in \Theta} \|X - x(\vartheta)\|.
$$

(2.1)

Here $\|\cdot\|$ is $L_2[\alpha, T]$-norm

$$
\|X - x(\vartheta)\| = \left(\int_\alpha^T [X_t - x_t(\vartheta)]^2 dt\right)^{1/2}.
$$

Introduce Gaussian process $x^{(1)}_t = x^{(1)}_t(\vartheta_0)$ as solution of the equation

$$
dx^{(1)}_t = S'(x_{t-\vartheta_0}) x^{(1)}_t dt + dW_t, \quad x_0 = 0, \quad 0 \leq t \leq T.
$$

(2.2)

Note that $x^{(1)}_t = \partial X_t/\partial \vartheta|_{\varepsilon=0}$ (see [14]) and that $x^{(1)}_{t-\vartheta_0} = 0, 0 \leq t \leq \vartheta_0$. Hence $x^{(1)}_t = W_t, 0 \leq t \leq \vartheta_0$. The derivative $\dot{x}_t(\vartheta), t \geq \vartheta$, can be calculated as follows

$$
\dot{x}_t(\vartheta) = \frac{\partial x_t(\vartheta)}{\partial \vartheta} = -\int_0^t S'(x_{s-\vartheta}) S(x_{s-2\vartheta}) ds.
$$

(2.3)

Introduce as well the Gaussian random variable

$$
\zeta(\vartheta_0) = \left(\int_{\vartheta_0}^T \dot{x}_t(\vartheta_0)^2 dt\right)^{-1} \int_{\vartheta_0}^T x^{(1)}_t(\vartheta_0) \dot{x}_t(\vartheta_0) dt \sim \mathcal{N}(0, D(\vartheta_0)).
$$

Its variance $D(\vartheta_0)$ can be calculated with the help of the equation (2.2), but its particular value is not important.

**Theorem 1.** Suppose that the function $S(x), x \in \mathcal{R}$ is positive, has two continuous bounded derivatives, and there exists $x_* \in (x_0, x_{T-\beta})$ such that $S'(x_*) \neq 0$. Then the MDE is consistent, asymptotically normal

$$
\varepsilon^{-1} (\bar{\vartheta}_{\varepsilon} - \vartheta_0) \Rightarrow \mathcal{N}(0, D(\vartheta_0)),
$$

(2.4)

and for any $p > 0$ the moments converge

$$
\varepsilon^{-p} \mathbb{E}_{\vartheta_0} |\bar{\vartheta}_{\varepsilon} - \vartheta_0|^p \rightarrow \mathbb{E}_{\vartheta_0} |\zeta(\vartheta_0)|^p.
$$

(2.5)

**Proof.** First we verify the consistency of the MDE.

**Lemma 1.** For any $\nu > 0$, we have

$$
\sup_{\vartheta_0 \in \Theta} \mathbb{P}_{\vartheta_0} (|\bar{\vartheta}_{\varepsilon} - \vartheta_0| > \nu) \rightarrow 0.
$$


Proof. The proof follows the standard arguments used in such problems

\[ P_{\vartheta_0} \left( |\bar{\vartheta} - \vartheta_0| > \nu \right) = P_{\vartheta_0} \left( \inf_{|\vartheta - \vartheta_0| < \nu} \|X - x(\vartheta)\| > \inf_{|\vartheta - \vartheta_0| \geq \nu} \|X - x(\vartheta)\| \right) \]

\[ \leq P_{\vartheta_0} \left( \inf_{|\vartheta - \vartheta_0| < \nu} \left( \|X - x(\vartheta_0)\| + \|x(\vartheta_0) - x(\vartheta)\| \right) > \inf_{|\vartheta - \vartheta_0| \geq \nu} \left( \|x(\vartheta_0) - x(\vartheta)\| - \|X - x(\vartheta_0)\| \right) \right) \]

\[ = P_{\vartheta_0} \left( 2 \|X - x(\vartheta_0)\| > \inf_{|\vartheta - \vartheta_0| \geq \nu} \|x(\vartheta_0) - x(\vartheta)\| \right). \]

Here we used the obvious equality \( \inf_{|\vartheta - \vartheta_0| < \nu} \|x(\vartheta_0) - x(\vartheta)\| = 0 \) and the triangle inequality.

For any real \( \nu > 0 \) and \( \vartheta_0 \in \Theta \), we introduce the function

\[ g(\vartheta_0, \nu) = \inf_{|\vartheta - \vartheta_0| \geq \nu} \|x(\vartheta_0) - x(\vartheta)\| \]

and show that it satisfies the identifiability condition \( \inf_{\vartheta_0 \in \Theta} g(\vartheta_0, \nu) > 0 \). Suppose that \( g(\vartheta_0, \nu) = 0 \), then there exist \( \vartheta_1 \) and \( \vartheta_0 \) such that \( |\vartheta_1 - \vartheta_0| \geq \nu \) and \( \|x(\vartheta_0) - x(\vartheta_1)\| = 0 \). The functions \( x_t(\vartheta_0), x_t(\vartheta_1), 0 \leq t \leq T \) are continuous, and therefore we obtain the equality \( x_t(\vartheta_0) = x_t(\vartheta_1), 0 \leq t \leq T \). Hence, for \( \vartheta_1 > \vartheta_0 \), we have equality

\[ x_t(\vartheta_1) = x_t(\vartheta_0) = (\vartheta_1 - \vartheta_0) \hat{x}_t(\tilde{\vartheta}) \equiv 0, \quad 0 \leq t \leq T, \]

where \( \vartheta_0 \leq \tilde{\vartheta} \leq \vartheta_1 \). As the function \( S(x) \) is strictly positive, this equality implies (see (2.3)) \( S'(x_{s - \tilde{\vartheta}}) \equiv 0 \) for all \( s \in [\vartheta_0, T] \). It is possible to take such \( s \) that \( s - \tilde{\vartheta} = x_s \), but by the conditions of the theorem, we have \( S'(x_{s - \tilde{\vartheta}}) = S'(x_s) \neq 0 \). Hence, \( g(\vartheta_0, \nu) > 0 \).

Using Tchebychev inequality, we get

\[ P_{\vartheta_0} \left( |\bar{\vartheta} - \vartheta_0| > \nu \right) \leq \frac{4}{g(\vartheta_0, \nu)} \mathbb{E}_{\vartheta_0} \int_0^T \|X_t - x_t(\vartheta_0)\|^2 dt \leq \frac{C \varepsilon^2}{g(\vartheta_0, \nu)^2} \rightarrow 0. \]

Here we used the estimate (5.1) proved in the Appendix below. Therefore, the equation (2.5) is proved.

The proof of the asymptotic normality (2.4) follows the same steps as in [14].

The convergence of moments is obtained as follows. For \( |\vartheta - \vartheta_0| \leq \nu \) and sufficiently small \( \nu > 0 \), we have

\[ \|x(\vartheta) - x(\vartheta_0)\|^2 = (\vartheta - \vartheta_0)^2 \|\hat{x}(\vartheta_0)\|^2 (1 + o(\nu)) \geq \frac{1}{2} (\vartheta - \vartheta_0)^2 \|\hat{x}(\vartheta_0)\|^2. \]

This estimate together with \( g(\vartheta_0, \nu) > 0 \) allows us to conclude that there exists \( \kappa > 0 \) such that \( g(\vartheta_0, \nu) \geq \kappa \nu \).
Let us put \( \nu = \varepsilon y > 0 \). Then, for any integer \( N > 1 \), we can write
\[
P_{\vartheta_0} (\varepsilon^{-1} | \tilde{\vartheta}_\varepsilon - \vartheta_0 | > y) \leq \frac{2^{2N}}{\kappa^{2N} y^{2N}} E_{\vartheta_0} \left| \int_0^T |X_t - x_t (\vartheta_0)|^2 dt \right|^N
\]
\[
\leq \frac{2^{2N} (T - \alpha)^{N-1}}{\kappa^{2N} y^{2N}} \int_0^T E_{\vartheta_0} |X_t - x_t (\vartheta_0)|^{2N} dt \leq C_N y^{2N},
\]
where the constant \( C > 0 \) does not depend on \( \varepsilon \) and \( \vartheta_0 \).

Let us denote \( F_\varepsilon (\vartheta_0, y) = P_{\vartheta_0} (|\tilde{u}_\varepsilon| < y) \). Then, for any \( p > 0 \), taking \( 2N > p + 1 \) we obtain
\[
E_{\vartheta_0} \left| \frac{\tilde{\vartheta}_\varepsilon - \vartheta_0}{\varepsilon} \right|^p \leq 1 + \int_1^{\infty} P_{\vartheta_0} (\varepsilon^{-1} | \tilde{\vartheta}_\varepsilon - \vartheta_0 | > y^{1/p}) dy \leq 1 + C_N \int_1^{\infty} y^{-2N/p} dy.
\]
Therefore, for any \( p > 1 \), the random variables \( \varepsilon^{-p} | \tilde{\vartheta}_\varepsilon - \vartheta_0 |^p \) are uniformly integrable and we obtain the convergence of moments (2.5).

Remark that if the value \( x_* \) mentioned in Theorem 1 does not exist and \( S' (x) = 0 \) for all \( x \in [x_0, x_{T_\beta}] \), then the function \( S (x) \), \( x \in [x_0, x_{T_\beta}] \), does not depend on \( x \) and consistent estimation of \( \vartheta_0 \) is impossible.

**One-step MLE.** Let us denote \( \left\{ P^{(\varepsilon)}_\vartheta, \vartheta \in \Theta \right\} \) the family of measures induced on the measurable space \( (\mathcal{C}_{(0,T)} \Theta) \) of continuous on \( [0,T] \) functions by the solutions of the equation (1.1) with different \( \vartheta \in \Theta \). These measures are equivalent with the likelihood ratio function \( L (\vartheta, X_T) \) [21]. Let us denote \( \vartheta_u = \vartheta_0 + \varepsilon u \) and introduce the normalized likelihood ratio process
\[
Z_\varepsilon (u) = \frac{L (\vartheta_0, X_T)}{L (\vartheta_0 + \varepsilon u, X_T)}, \quad u \in \mathbb{U}_\varepsilon,
\]
where \( u \in \mathbb{U}_\varepsilon = (\varepsilon^{-1} (\alpha - \vartheta_0), \varepsilon^{-1} (\beta - \vartheta_0)) \). We have (see (1.6))
\[
\ln Z_\varepsilon (u) = \int_0^T \frac{S(X_t - \vartheta_u)}{\varepsilon} S(X_t - \vartheta_0) dW_t - \int_0^T \frac{|S(X_t - \vartheta_u) - S(X_t - \vartheta_0)|^2}{2\varepsilon^2} dt.
\]
As it follows from the given below proof of the theorem (see also [1]), this likelihood ratio admits the representation
\[
Z_\varepsilon (u) = \exp \left\{ -u \int_0^T S'(x_t - \vartheta_0) S(x_t - 2\vartheta_0) dW_t - \frac{u^2}{2} I (\vartheta_0) + o (1) \right\}.
\]
Here \( x_t - \vartheta_0 = x_t - \vartheta_0 (\vartheta_0) \), \( x_t - 2\vartheta_0 = x_t - 2\vartheta_0 (\vartheta_0) \), and \( I (\vartheta) \) plays the role of the Fisher information. We have
\[
\int_0^T S'(x_t - \vartheta) S(x_t - 2\vartheta) dW_t \sim \mathcal{N} (0, I (\vartheta)), \quad I (\vartheta) = \int_0^T S'(x_t - \vartheta)^2 S(x_t - 2\vartheta)^2 dt.
\]

insart-bj ver. 2014/02/20 file: DB-1step_YK_v2.tex date: November 4, 2020
Hence, the family of measures \( \{ P_{\theta}^{(\epsilon)}, \theta \in \Theta \} \) is locally asymptotically normal (LAN) for all \( \theta_0 \in \Theta \). Therefore, we have the Hajek-Le Cam’s lower bound on the mean squared risk of all the estimators

\[
\lim_{\nu \to 0} \lim_{\epsilon \to 0} \sup_{|\theta - \theta_0| \leq \nu} \epsilon^{-2} E_{\theta} (\hat{\theta}_\epsilon - \theta)^2 \geq 1 (\theta_0)^{-1},
\]

where \( \hat{\theta}_\epsilon \) is an arbitrary estimator. As usual, we call an estimator \( \hat{\theta}_\epsilon^* \) asymptotically efficient, if for this estimator we have equality in (2.6) for all \( \theta_0 \in \Theta \).

Recall that the MLE \( \hat{\theta}_\epsilon \) for this model of observations is asymptotically efficient [1, 14].

The one-step MLE \( \hat{\theta}_\epsilon^* \) for this model of observations can formally be written as follows

\[
\hat{\theta}_\epsilon^* = \hat{\theta}_\epsilon + \epsilon^2 I (\hat{\theta}_\epsilon)^{-1} \frac{\partial \ln L (\theta, X^T)}{\partial \theta} \bigg|_{\theta = \hat{\theta}_\epsilon} \epsilon
\]

We say formally, because in this writing there are two difficulties. The first one: as the estimator \( \hat{\theta}_\epsilon \) depends on all the observations on the interval \([0, T]\), the Itô stochastic integral is not well defined. The second problem, of course, is the calculation of the derivative \( \partial X_{t-\theta} / \partial \theta \). Our goal is to find an alternative expression for the One-step MLE which does not have such problems.

Let us write the expressions for \( X_{t-\theta} \) and for its derivative (without Wiener part)

\[
\hat{X}_{t-\theta} = x_0 + \int_0^{t-\theta} S (X_{s-\theta_0}) \, ds, \quad \frac{\partial \hat{X}_{t-\theta}}{\partial \theta} = -S (X_{t-\theta-\theta_0}) \rho_{(t > \theta)}.
\]

First we replace the stochastic process \( X_{t-\theta} \) by \( x_{t-\theta} (\hat{\theta}_\epsilon) \) and write the integral as follows

\[
\int_0^{T} S' (x_{t-\theta}) \frac{\partial x_{t-\theta} (\hat{\theta}_\epsilon)}{\partial \theta} \bigg|_{\theta = \hat{\theta}_\epsilon} [dX_t - S (X_{t-\theta}) \, dt]
\]

\[
= \int_0^{T} S' (x_{t-\theta}) S (x_{t-2\theta}) \int_0^{T} S' (x_{t-\theta}) S (x_{t-2\theta}) \, dX_t.
\]

Note that \( X_{t-\theta} \big|_{\epsilon = 0} = x_{t-\theta} (\theta_0) \) and we differentiate on \( \theta \) only, but then replace the second value \( \theta_0 \) by \( \hat{\theta}_\epsilon \) too. Remind that the likelihood ratio depends on the true value \( \theta_0 \), but we never differentiate it on \( \theta_0 \).

Further, let us denote \( H (t, \theta) = S' (x_{t-\theta} (\theta)) S (x_{t-2\theta} (\theta)) \). Then

\[
d[H (t, \theta) X_t] = H (t, \theta) \, dX_t + X_t H_t' (t, \theta) \, dt
\]

and

\[
\int_0^{T} S' (x_{t-\theta}) S (x_{t-2\theta}) \, dX_t = H (T, \theta) X_T - H (\theta, \theta) X_\theta - \int_0^{T} H_t' (t, \theta) X_t \, dt.
\]
Of course, we have
\[ H^\prime_1(t, \vartheta) = S(x_{t-\vartheta} (\vartheta)) S(x_{t-2\vartheta} (\vartheta))^2 \]
\[ + S' (x_{t-\vartheta} (\vartheta)) S' (x_{t-2\vartheta} (\vartheta)) S (x_{t-3\vartheta} (\vartheta)) \mathbb{I}_{t \geq 2 \vartheta}. \]

Now we can put
\[ \Psi (\tilde{\vartheta}) = H \left( T, \tilde{\vartheta} \right) X_T - H \left( \vartheta, \tilde{\vartheta} \right) X_{\tilde{\vartheta}} - \int_{\vartheta}^{T} H^\prime_1 (t, \tilde{\vartheta}) X_t \, dt \]
and define the One-step MLE
\[ \hat{\vartheta}^*_1 = \tilde{\vartheta} + 1 (\tilde{\vartheta})^{-1} \left[ \int_{\vartheta}^{T} S' (x_{t-\vartheta}) S (x_{t-2\vartheta}) S (X_{t-\vartheta}) \, dt - \Psi (\tilde{\vartheta}) \right]. \]

Its properties are given in the next theorem.

**Theorem 2.** Suppose that the function \( S(\cdot) \) is positive, has four continuous bounded derivatives, and there exists \( x_0 \in (x_0, x_{T-\beta}) \) such that \( S'(x_0) \neq 0 \). Then the One-step MLE \( \hat{\vartheta}^*_1 \) is consistent, asymptotically normal
\[ \varepsilon^{-1} (\hat{\vartheta}^*_1 - \vartheta_0) \Rightarrow \mathcal{N} \left( 0, I(\vartheta_0)^{-1} \right) \]
and asymptotically efficient.

**Proof.** Let us study the expression
\[ R (\tilde{\vartheta}) = \int_{\vartheta}^{T} S' (x_{t-\vartheta}) S (x_{t-2\vartheta}) S (X_{t-\vartheta}) \, dt - \Psi (\tilde{\vartheta}) = H \left( \tilde{\vartheta}, \tilde{\vartheta} \right) X_{\tilde{\vartheta}} \]
\[ + \int_{\vartheta}^{T} \left[ H \left( t, \tilde{\vartheta} \right) S (X_{t-\vartheta}) - H^\prime_1 (t, \tilde{\vartheta}) X_t \right] \, dt - H \left( T, \tilde{\vartheta} \right) X_T. \]

We have to expand it at the vicinity of the point \( \tilde{\vartheta}_0 \). As \( \tilde{\vartheta} = \tilde{\vartheta}_0 + \varepsilon \tilde{u} \) and the function \( H(t, \vartheta) \) is sufficiently smooth w.r.t. \( \vartheta \) and \( t \), we can write the expansions for the function \( H(\cdot) \). Before, we note that there is a problem with the expansion of the term \( H(\vartheta, \tilde{\vartheta})X_{\tilde{\vartheta}} \). The process \( X_t \) has no derivative, but we have
\[ X_{\tilde{\vartheta}} = X_{\tilde{\vartheta}_0} + \int_{\vartheta_0}^{\vartheta} S (X_{t-\vartheta}) \, dt + \varepsilon \left[ W_{\tilde{\vartheta}} - W_{\vartheta_0} \right] \]
\[ = X_{\tilde{\vartheta}_0} + \varepsilon \tilde{u} S (x_0) (1 + o_m (1)) + \varepsilon^{3/2} O_m (1). \]
Here \( o_m (1) \) means the convergence of all the moments, i.e., \( E_{\tilde{\vartheta}_0} |o_m (1)|^p \to 0 \) as \( \varepsilon \to 0 \), and \( O_m (1) \) means \( E_{\tilde{\vartheta}_0} |O_m (1)|^p < C \) for any \( p > 0 \). The constant \( C > 0 \) here does not depend on \( \varepsilon \). Therefore, we can write
\[ H \left( T, \tilde{\vartheta} \right) = H \left( T, \vartheta_0 \right) + \varepsilon \tilde{u} H_0 (T, \vartheta_0) + \varepsilon^2 A(T, \tilde{\vartheta}), \]
\[ H \left( \vartheta, \tilde{\vartheta} \right) = H \left( \vartheta_0, \vartheta_0 \right) + \varepsilon \tilde{u} [H_0' (\vartheta_0, \vartheta_0) + H_0'' (\vartheta_0, \vartheta_0)] + \varepsilon^2 B(\tilde{\vartheta}, \tilde{\vartheta}), \]
}\]
and for the integral we have
\[ \int_{\tilde{\vartheta}}^{T} H'_t(t, \tilde{\vartheta}) X_t \, dt = \int_{\tilde{\vartheta}}^{T} H'_t(t, \vartheta_0) X_t \, dt \]
\[ + \varepsilon \bar{u}_{\varepsilon} \left[ \int_{\tilde{\vartheta}}^{T} H''_{t, \vartheta}(t, \vartheta_0) X_t \, dt - H'_t(t, \vartheta_0) X_{\vartheta_0} \right] + \varepsilon^{3/2} C(T, \tilde{\vartheta}_\varepsilon, \tilde{\vartheta}_\varepsilon). \]

Here we denoted \( A(\cdot), B(\cdot) \) and \( C(\cdot) \) the corresponding residuals. Note that all polynomial moments of these quantities are finite. Of course, the values of \( \tilde{\vartheta}_\varepsilon \) in these three expansions are different and this is just a symbolical writing.

Therefore, we have the presentation \( \Psi(\tilde{\vartheta}_\varepsilon) = \Psi(\vartheta_0) + \varepsilon \bar{u}_{\varepsilon} \hat{\Psi}(\vartheta_0) + \varepsilon^{3/2} O_m(1) \), where
\[ \hat{\Psi}(\vartheta_0) = \int_{\vartheta_0}^{T} S'(x_{t-\vartheta_0}) S(x_{t-2\vartheta_0}) \, dX_t \]
\[ = \int_{\vartheta_0}^{T} H(t, \vartheta_0) S(X_{t-\vartheta_0}) \, dt + \varepsilon \int_{\vartheta_0}^{T} H(t, \vartheta_0) \, dW_t \]

and
\[ \hat{\Psi}(\vartheta_0) = H'_0(T, \vartheta_0) X_T - [H'_0(\vartheta_0, \vartheta_0) + H'_t(\vartheta_0, \vartheta_0)] X_{\vartheta_0} - H(\vartheta_0, \vartheta_0) S(x_0) \]
\[ + H'_t(\vartheta_0, \vartheta_0) X_{\vartheta_0} = \int_{\vartheta_0}^{T} H''_{t, \vartheta}(t, \vartheta_0) X_t \, dt \]
\[ = H'_0(T, \vartheta_0) X_T - H'_0(\vartheta_0, \vartheta_0) X_{\vartheta_0} - H(\vartheta_0, \vartheta_0) S(x_0) - \int_{\vartheta_0}^{T} H''_{t, \vartheta}(t, \vartheta_0) X_t \, dt \]
\[ = \int_{\vartheta_0}^{T} H'_0(t, \vartheta_0) \, dX_t - H(\vartheta_0, \vartheta_0) S(x_0) \]
\[ = \int_{\vartheta_0}^{T} H'_0(t, \vartheta_0) S(X_{t-\vartheta_0}) \, dt + \varepsilon \int_{\vartheta_0}^{T} H'_0(t, \vartheta_0) \, dW_t - H(\vartheta_0, \vartheta_0) S(x_0). \]

Further, we obtain
\[ \int_{\tilde{\vartheta}}^{T} H(t, \tilde{\vartheta}) S(X_{t-\tilde{\vartheta}}) \, dt = \int_{\vartheta_0}^{T} H(t, \vartheta_0) S(X_{t-\vartheta_0}) \, dt - \varepsilon \bar{u}_{\varepsilon} H(\vartheta_0, \vartheta_0) S(x_0) \]
\[ + \varepsilon \bar{u}_{\varepsilon} \int_{\vartheta_0}^{T} H'_0(t, \vartheta_0) S(X_{t-\vartheta_0}) \, dt + \varepsilon^{3/2} O_m(1). \]
Therefore, we have
\[
R (\bar{\vartheta}_\varepsilon) = \int_{\vartheta_0}^{T} H (t, \vartheta_0) [S (X_{t-\vartheta_0}) - S (X_{t-\vartheta_0})] \, dt - \varepsilon \int_{\vartheta_0}^{T} H (t, \vartheta_0) \, dW_t + \varepsilon^{3/2} O_m (1)
\]
\[
= -\varepsilon \bar{u}_\varepsilon \int_{\vartheta_0}^{T} H (t, \vartheta_0) S' (x_{t-\vartheta_0}) S (x_{t-\vartheta_0}) \, dt - \varepsilon \int_{\vartheta_0}^{T} H (t, \vartheta_0) \, dW_t + \varepsilon^{3/2} O_m (1)
\]
\[
= -\varepsilon \bar{u}_\varepsilon I (\vartheta_0) - \varepsilon \int_{\vartheta_0}^{T} H (t, \vartheta_0) \, dW_t + \varepsilon^{3/2} O_m (1).
\]
The obtained relations allow us to write
\[
\varepsilon \vartheta_\star - \varepsilon \vartheta_0 = \bar{u}_\varepsilon - I (\vartheta_\varepsilon) \int_{\vartheta_0}^{T} H (t, \vartheta_0) \, dW_t - \varepsilon \int_{\vartheta_0}^{T} H (t, \vartheta_0) \, dW_t + \varepsilon^{1/2} O_m (1)
\]
\[
= -I (\vartheta_0) \int_{\vartheta_0}^{T} H (t, \vartheta_0) \, dW_t \left( 1 + \varepsilon^{1/2} O_m (1) \right).
\]
Recall that
\[
-I (\vartheta_0) \int_{\vartheta_0}^{T} H (t, \vartheta_0) \, dW_t \sim \mathcal{N} \left( 0, I (\vartheta_0)^{-1} \right).
\]
Note that the Fisher information is separated from zero, i.e.,
\[
\inf_{\vartheta \in \Theta} I (\vartheta) > 0,
\]
and we can easily verify the convergence of all polynomial moments of \( o (\varepsilon^{1/2}) \) above.
Moreover, in all the estimates of the type
\[
E_{\vartheta_0} \left| \varepsilon^{1/2} O_m (1) \right|^p \leq C \varepsilon^{p/2}
\]
(with any \( p > 0 \)), the corresponding constants \( C = C (K) > 0 \) does not depend on \( \vartheta_0 \in K \) for any compact set \( K \subset \Theta \). This provides us the uniform convergence of moments of \( u_\star = \varepsilon^{-1} (\vartheta_\star - \vartheta_0) \). For the second moment, we have
\[
\sup_{|\vartheta - \vartheta_0| \leq \nu} E_{\vartheta} \left| \frac{\vartheta_\star - \vartheta_0}{\varepsilon} \right|^2 \varepsilon \to 0 \sup_{|\vartheta - \vartheta_0| \leq \nu} I (\vartheta)^{-1} \nu \to 0 \to I (\vartheta_0)^{-1}.
\]
Therefore, the One-step MLE \( \vartheta_\star \) is asymptotically efficient.

**Example 1.** Suppose that we have the Ornstein-Uhlenbeck process with delay
\[
dX_t = -\gamma X_{t-\vartheta} \, dt + \varepsilon \, dW_t, \quad X_s = x_0, \quad 0 \leq t \leq T.
\]
Here \( \gamma > 0 \) is supposed to be known, \( x_0 > 0 \) and \( \vartheta \in (\alpha, \beta) \), \( 0 < \alpha < \beta < T \). The solution \( x_t = x_t (\vartheta_0) \) of the limit equation
\[
\frac{dx_t}{dt} = -\gamma x_{t-\vartheta}, \quad x_s = x_0, \quad 0 \leq t \leq T,
\]
can be written in explicit form as follows
\[ x_t(\vartheta_0) = x_0 y_t(\vartheta_0) - \gamma x_0 \int_{-\vartheta_0}^{0} y_{t-s-\vartheta_0}(\vartheta_0) \, ds, \quad t \geq 0, \]
where
\[ y_t(\vartheta_0) = \left\lfloor \frac{t}{\vartheta_0} \right\rfloor \sum_{k=0}^{\left\lfloor \frac{t}{\vartheta_0} \right\rfloor} (-1)^k \frac{\gamma^k}{k!} (t - k\vartheta_0)^k, \]
and \( \lfloor a \rfloor \) is the integer part of \( a \) (see [12] for details).

The preliminary MDE \( \vartheta_{\varepsilon} \) and the Fisher information are
\[ \vartheta_{\varepsilon} = \operatorname{arg\,inf}_{\alpha<\vartheta<\beta} \int_{\alpha}^{T} [x_t - x_t(\vartheta)]^2 \, dt, \quad I(\vartheta) = \gamma^4 \int_{\vartheta}^{T} x_{t-2\vartheta}(\vartheta)^2 \, dt. \]
The One-step MLE \( \vartheta^*_{\varepsilon} \) has the following expression
\[
\vartheta^*_{\varepsilon} = \vartheta_{\varepsilon} + \gamma^4 \left[ \gamma^2 x_0 X_{\vartheta_{\varepsilon}} - \gamma^2 x_{T-2\vartheta_{\varepsilon}}(\vartheta_{\varepsilon}) X_{T} - \gamma^3 \int_{\vartheta_{\varepsilon}}^{T} x_{t-2\vartheta_{\varepsilon}}(\vartheta_{\varepsilon}) X_{t} \, dt - \gamma^3 \int_{2\vartheta_{\varepsilon}}^{T} x_{t-3\vartheta_{\varepsilon}}(\vartheta_{\varepsilon}) X_{t} \, dt \right].
\]
By Theorem 2, we have
\[ \varepsilon^{-1} (\vartheta^*_{\varepsilon} - \vartheta_0) \xrightarrow{\mathcal{D}} N\left(0, I(\vartheta)^{-1}\right). \]

3. Scale-type delay

Below we consider the models of type (1.2). We start with a linear model with an equation, which can be called Stochastic Pantograph Equation, and only afterwards we study the nonlinear model.

Stochastic Pantograph Equation. Recall that the deterministic Pantograph equation
\[ \frac{dx_t}{dt} = ax_t + bx_{\vartheta t}, \quad x_0, \quad t \geq 0, \quad (3.1) \]
was introduced by Ockendon and Tayler in 1971 [24]. This equation describes the special construction on the electric locomotive of the British Railways, which allowed to collect the current from an overhead wire. Here the parameter \( 0 < \vartheta < 1 \) defines the delay \( d = t - \vartheta t \). The further study of this equation can be found in the thesis [6].

Introduce the condition
\( \mathcal{A}. \) The reals \( x_0 \neq 0, b \neq 0 \) and \( a + b \neq 0. \)
Note that if \( x_0 = 0 \), the solution of (3.1) is \( x_t \equiv 0 \), if \( b = 0 \), there is no delay, and if \( a + b = 0 \), we have once more a trivial solution \( x_t \equiv x_0 \).

The stochastic version of the Pantograph equation can be

\[
dX_t = a X_t \, dt + b X_{\sigma t} \, dt + \varepsilon \, dW_t, \quad X_0 = x_0, \quad t \geq 0.
\]

Consider the problem of estimation of \( \vartheta \in \Theta = (\alpha, \beta) \), \( 0 < \alpha < \beta < 1 \) by the observations \( X^T = (X_t, 0 \leq t \leq T) \). The reals \( a, b \neq 0 \), \( x_0 \neq 0 \) and \( \varepsilon \in (0, 1] \) are supposed to be known.

The likelihood ratio function is

\[
L(\vartheta, X^T) = \exp \left\{ \int_0^T \frac{b X_{\sigma t}}{\varepsilon^2} \, dX_t - \int_0^T \frac{b^2 X_{\sigma t}^2 - 2abX_tX_{\sigma t}}{2\varepsilon^2} \, dt \right\}, \quad \vartheta \in \Theta,
\]

and the numerical calculation of the MLE \( \hat{\vartheta}_\varepsilon \) defined by the equation (1.7) has the same problems as in the case of observations (1.1). Note that the asymptotic \( (\varepsilon \to 0) \) properties of the MLE in the case \( a = 0 \) were described in [13]. Here we consider the construction of the one-step MLE using the Fisher-score device similar to that given above.

The role of Fisher information is played by the quantity

\[
I(\vartheta) = b^2 \int_0^T t^2 [ax_{\sigma t}(\vartheta) + bx_{\sigma t}(\vartheta)]^2 \, dt.
\]

As preliminary estimator we can take the MDE \( \bar{\vartheta}_\varepsilon \) defined by the equation (2.1), where \( x(\vartheta) = (x_t(\vartheta), 0 \leq t \leq T) \) is the solution of the equation (3.1). It can be shown that this estimator is consistent and asymptotically normal \( \varepsilon^{-1} (\bar{\vartheta}_\varepsilon - \vartheta_0) \). Moreover, it is consistent and asymptotically normal even if we use the observations \( X^\tau = (X_t, 0 \leq t \leq \tau_\varepsilon) \) on the interval \( [0, \tau_\varepsilon] \) with \( \tau_\varepsilon \to 0 \).

**Estimator of substitution.** Note that the model (3.2) allows us to use another, more simple, estimator as preliminary one, which can be introduced as follows

\[
\bar{\vartheta}_\tau = \frac{2}{(a + b) b \tau_\varepsilon^2} [X_{\tau_\varepsilon} - x_0 - x_0 \tau_\varepsilon (a + b)] - \frac{a}{b}.
\]

Below we show that this is an estimator of substitution.

**Theorem 3.** Let the condition \( \mathcal{A} \) be fulfilled and \( \tau_\varepsilon = \varepsilon^\gamma \), where \( \gamma \in \left(\frac{2}{3}, 1\right) \). Then the estimator \( \bar{\vartheta}_\tau \) is consistent and asymptotically normal

\[
\frac{\tau_\varepsilon^{3/2}}{\varepsilon} (\bar{\vartheta}_\tau - \vartheta_0) \Rightarrow \mathcal{N}(0, D^2), \quad D^2 = \frac{4}{(a + b)^2 b^2}.
\]

Moreover, for any \( p > 0 \), the moments converge

\[
\varepsilon^{-p(1 - 3 \gamma)} \mathbf{E}_{\vartheta_0} \left| (\bar{\vartheta}_\tau - \vartheta_0) \right|^p \to D^p \mathbf{E} |\xi|^p, \quad \xi \sim \mathcal{N}(0, 1).
\]
Proof. The function $x_t(\vartheta)$ can be expanded at the vicinity of the point $t = 0$

$$x_t(\vartheta) = x_0 + \int_0^t [ax_s + bx_{\vartheta s}] \, ds$$

$$= x_0 + x_0 t [a + b] + \frac{1}{2} x_0 t^2 [a + b \vartheta] + t^3 O(1),$$

and we can write

$$\vartheta_{\tau} = \frac{2}{(a + b) b \tau^2} [x_{\tau} (\vartheta) - x_0 - x_0 \tau (a + b)] - \frac{a}{b} \tau + \tau O(1).$$

Remind that the process $X_t$ admits the representation

$$X_t = x_t(\vartheta_0) + \varepsilon x_t^{(1)}(\vartheta_0) \, dt + \varepsilon^2 O_m(1),$$

where the Gaussian process $x_t^{(1)}(\vartheta_0)$ (derivative of $X_t$ w.r.t. $\varepsilon$) has the stochastic differential

$$dx_t^{(1)}(\vartheta_0) = \left[ax_t^{(1)}(\vartheta_0) + bx_{\vartheta t}^{(1)}(\vartheta_0)\right] dt + dW_t, \quad x_0^{(1)}(\vartheta_0) = 0,$$

and its expansion is $x_t^{(1)}(\vartheta_0) = W_{\tau} + \tau \varepsilon^2 O_m(1)$. Hence, we have

$$\frac{\gamma \varepsilon}{\tau_{\varepsilon}} (\vartheta_{\tau_{\varepsilon}} - \vartheta_0) = \frac{2 W_{\tau_{\varepsilon}}}{(a + b) b \tau_{\varepsilon}^2} + \tau \varepsilon O_m(1) = \vartheta_0 + \frac{2 W_{\tau_{\varepsilon}}}{(a + b) b \tau_{\varepsilon}^2} + \tau \varepsilon O_m(1).$$

Here $W_{\tau} \sim N(0,1)$. Therefore, we get

$$\frac{\gamma^{3/2}}{\varepsilon} (\vartheta_{\tau_{\varepsilon}} - \vartheta_0) = \frac{2 W_{\tau_{\varepsilon}}}{(a + b) b + \frac{\gamma^{5/2}}{\varepsilon} O_m(1),}$$

and if we suppose that

$$\frac{\gamma^{3/2}}{\varepsilon} \rightarrow \infty, \quad \frac{\gamma^{5/2}}{\varepsilon} \rightarrow 0, \quad (3.4)$$

we obtain the asymptotic normality (3.3). Note that the choice $\tau_{\varepsilon} = \varepsilon^\gamma$ with $\gamma \in (\frac{2}{5}, \frac{2}{3})$ provides (3.4).

The rate of convergence of this estimator is $\varepsilon^\kappa$ with $0 < \kappa < \frac{1}{2}$, i.e.,

$$\vartheta_{\tau_{\varepsilon}} - \vartheta_0 \varepsilon^\kappa \rightarrow N(0, D^2).$$

This rate of convergence is not sufficient for the construction of the One-step MLE. We need an estimator with a rate $\kappa > \frac{1}{2}$. That is why we propose below another estimator which uses $\vartheta_{\tau_{\varepsilon}}$ and has the rate of convergence grater than $\frac{1}{2}$. 

One-step MDE. Let us consider the equation (2.1), where $X^\tau_t = (X_t, 0 \leq t \leq \tau_\varepsilon)$. The solution $\hat{\theta}_\tau$ of this equation satisfies the following minimum distance equation

$$
\int_0^{\tau_\varepsilon} [X_t - x_t (\hat{\theta}_\tau)] \dot{x}_t (\hat{\theta}_\tau) \, dt = 0. \tag{3.5}
$$

Here $\dot{x}_t (\theta) = \partial x_t (\theta) / \partial \theta$ is solution of the equation

$$
\frac{d\dot{x}_t (\theta)}{dt} = a\dot{x}_t (\theta) + bx_{\theta t} (\theta) + bt [ax_{\theta t} (\theta) + bx_{\theta^2 t} (\theta)], \quad \dot{x}_0 (\theta) = 0. \tag{3.6}
$$

Let us write $x_t (\hat{\theta}_\tau) = x_t (\hat{\theta}_0) + (\hat{\theta}_\tau - \hat{\theta}_0) \dot{x}_t (\hat{\theta}_\tau)$, where the value $\hat{\theta}_\tau$ satisfies the relation $|\hat{\theta}_\tau - \hat{\theta}_0| \leq |\hat{\theta}_\tau - \hat{\theta}_0|$. From the equation (3.5), we obtain the representation

$$
\hat{\theta}_\tau = \hat{\theta}_0 + \left( \int_0^{\tau_\varepsilon} \dot{x}_t (\hat{\theta}_\tau) \dot{x}_t (\hat{\theta}_\tau) \, dt \right)^{-1} \int_0^{\tau_\varepsilon} [X_t - x_t (\hat{\theta}_0)] \dot{x}_t (\hat{\theta}_\tau) \, dt.
$$

Let us denote

$$
Q_{\tau_\varepsilon} (\theta) = \int_0^{\tau_\varepsilon} \dot{x}_t (\theta)^2 \, dt.
$$

We introduce the One-step MDE $\hat{\theta}_\tau^*$ as follows

$$
\hat{\theta}_\tau^* = \hat{\theta}_\tau + Q_{\tau_\varepsilon} (\hat{\theta}_\tau)^{-1} \int_0^{\tau_\varepsilon} [X_t - x_t (\hat{\theta}_\tau)] \dot{x}_t (\hat{\theta}_\tau) \, dt.
$$

Theorem 4. Let the conditions of Theorem 3 be fulfilled. Then, for any $p > 0$, there exists a constant $C > 0$ such that

$$
\varepsilon^{-p} E_{\theta_0} \left| \hat{\theta}_\tau^* - \hat{\theta}_0 \right|^p \leq C,
$$

where $\delta \in (0, \frac{2}{3})$.

Proof. Let us write

$$
\frac{\hat{\theta}_\tau^* - \hat{\theta}_0}{\varepsilon^\delta} = \hat{\theta}_\tau - \hat{\theta}_0 + \varepsilon^{-\delta} Q_{\tau_\varepsilon} (\hat{\theta}_\tau)^{-1} \int_0^{\tau_\varepsilon} [X_t - x_t (\hat{\theta}_\tau)] \dot{x}_t (\hat{\theta}_\tau) \, dt.
$$

We have

$$
\varepsilon^{-\delta} [X_t - x_t (\hat{\theta}_\tau)] = \varepsilon^{-\delta} [X_t - x_t (\theta_0)] - \varepsilon^{-\delta} (\hat{\theta}_\tau - \hat{\theta}_0) \dot{x}_t (\hat{\theta}_\tau).
$$

Recall that

$$
E_{\theta_0} |X_t - x_t (\theta_0)|^p \leq C \varepsilon^p
$$

(see (5.2) below). Therefore, for any $\delta < 1$, we have $\varepsilon^{-\delta} [X_t - x_t (\theta_0)] = o_m (1)$. Further, we can write once more the expansion $\dot{x}_t (\hat{\theta}_\tau) = \dot{x}_t (\hat{\theta}_\tau) + (\hat{\theta}_\tau - \hat{\theta}_0) \dot{x}_t (\hat{\theta}_\tau)$. Hence, we
have

\[ \varepsilon^{-\delta} Q_{\tau_{\varepsilon}} (\bar{\vartheta}_{\tau_{\varepsilon}})^{-1} \int_{0}^{\tau_{\varepsilon}} \left[ X_t - x_t (\bar{\vartheta}_{\tau_{\varepsilon}}) \right] \dot{x}_t (\bar{\vartheta}_{\tau_{\varepsilon}}) \, dt \]

\[ = o_m (1) - \varepsilon^{-\delta} (\bar{\vartheta}_{\tau_{\varepsilon}} - \vartheta_0) Q_{\tau_{\varepsilon}} (\bar{\vartheta}_{\tau_{\varepsilon}})^{-1} \int_{0}^{\tau_{\varepsilon}} \dot{x}_t (\bar{\vartheta}_{\tau_{\varepsilon}})^2 \, dt \]

\[ + \varepsilon^{-\delta} (\bar{\vartheta}_{\tau_{\varepsilon}} - \vartheta_0) (\bar{\vartheta}_{\tau_{\varepsilon}} - \tilde{\vartheta}_{\tau_{\varepsilon}}) Q_{\tau_{\varepsilon}} (\bar{\vartheta}_{\tau_{\varepsilon}})^{-1} \int_{0}^{\tau_{\varepsilon}} \tilde{x}_t (\bar{\vartheta}_{\tau_{\varepsilon}}) \dot{x}_t (\bar{\vartheta}_{\tau_{\varepsilon}}) \, dt \]

\[ = o_m (1) - \frac{(\bar{\vartheta}_{\tau_{\varepsilon}} - \vartheta_0)}{\varepsilon^\delta} + \frac{(\bar{\vartheta}_{\tau_{\varepsilon}} - \vartheta_0) (\bar{\vartheta}_{\tau_{\varepsilon}} - \tilde{\vartheta}_{\tau_{\varepsilon}}) \int_{0}^{\tau_{\varepsilon}} \tilde{x}_t (\bar{\vartheta}_{\tau_{\varepsilon}}) \dot{x}_t (\bar{\vartheta}_{\tau_{\varepsilon}}) \, dt}{Q_{\tau_{\varepsilon}} (\bar{\vartheta}_{\tau_{\varepsilon}})} . \]

For small values of \( t \), the solution of the equation (3.6) can be written as follows

\[ \dot{x}_t (\vartheta) = \frac{1}{2} b t^2 x_0 (a + B) + O (t^3) . \]

For the second derivative we have a similar expansion

\[ \ddot{x}_t (\vartheta) = \frac{1}{2} a t^2 x_0 (a + b) + \frac{1}{2} b t^2 x_0 (a + b) + O (t^3) . \]

Hence, we have estimates

\[ Q_{\tau_{\varepsilon}} (\bar{\vartheta}_{\tau_{\varepsilon}}) \geq c \varepsilon^5 , \quad \int_{0}^{\tau_{\varepsilon}} \bar{x}_t (\bar{\vartheta}_{\tau_{\varepsilon}}) \dot{x}_t (\bar{\vartheta}_{\tau_{\varepsilon}}) \, dt \leq C \varepsilon^5 . \]

Note as well that

\[ \left| (\bar{\vartheta}_{\tau_{\varepsilon}} - \vartheta_0) (\bar{\vartheta}_{\tau_{\varepsilon}} - \tilde{\vartheta}_{\tau_{\varepsilon}}) \right| \leq 2 |\bar{\vartheta}_{\tau_{\varepsilon}} - \vartheta_0|^2 . \]

Therefore, we obtain the relation

\[ \left| \frac{(\bar{\vartheta}_{\tau_{\varepsilon}} - \vartheta_0) (\bar{\vartheta}_{\tau_{\varepsilon}} - \tilde{\vartheta}_{\tau_{\varepsilon}}) \int_{0}^{\tau_{\varepsilon}} \tilde{x}_t (\bar{\vartheta}_{\tau_{\varepsilon}}) \dot{x}_t (\bar{\vartheta}_{\tau_{\varepsilon}}) \, dt}{Q_{\tau_{\varepsilon}} (\bar{\vartheta}_{\tau_{\varepsilon}})} \right| \leq C \frac{|\bar{\vartheta}_{\tau_{\varepsilon}} - \vartheta_0|^2}{\varepsilon^\delta} = \frac{\varepsilon^{2\kappa}}{\varepsilon^{\delta}} |\bar{u}_\varepsilon|^2 , \]

where \( \bar{u}_\varepsilon = \varepsilon^{-\kappa} (\bar{\vartheta}_{\tau_{\varepsilon}} - \vartheta_0) \). If we take \( \delta < 2\kappa \), i.e., \( 0 < \delta < \frac{2}{3} \), then for the One-step MDE we obtain

\[ \frac{\vartheta^*_{\tau_{\varepsilon}} - \vartheta_0}{\varepsilon^\delta} = o_m (1) , \quad \mathbb{E}_{\vartheta_0} \left| \frac{\vartheta^*_{\tau_{\varepsilon}} - \vartheta_0}{\varepsilon^\delta} \right|^p = o_m (1) . \]

Now the rate of convergence of the preliminary estimator \( \vartheta^*_{\tau_{\varepsilon}} \) can be greater than \( \frac{1}{2} \), and it can be used for the construction of the One-step MLE.

**Two-step MLE.** Having this preliminary estimator, we construct the Two-step MLE as follows

\[ \vartheta^*_\tau = \vartheta^*_{\tau_{\varepsilon}} + \mathbb{I}_{\tau_{\varepsilon}} (\vartheta^*_{\tau_{\varepsilon}})^{-1} b \int_{\tau_{\varepsilon}}^{\tau} \left[ a x_t + b x_{\vartheta^*_\tau} \right] \left[ dX_t - \left( a X_t + b X_{\vartheta^*_\tau} \right) dt \right] , \]

\[ \text{insart-bj ver. 2014/02/20 file: DB-1step_YK_v2.tex date: November 4, 2020} \]
where
\[
I_r(\vartheta) = b^2 \int_{\tau}^{T} t^2 \left[ ax_{\vartheta t} (\vartheta) + bx_{\vartheta^2 t} (\vartheta) \right]^2 dt.
\]
Note that we have no more problems with the definition of the stochastic integral, because the estimator \( \vartheta^*_r \) depends on the observations \( X^r = (X_t, 0 \leq t \leq r) \), and the integral starts at the moment \( r \).

**Theorem 5.** Suppose that the condition \( \mathcal{A} \) is fulfilled and \( r = \varepsilon^\gamma \) with \( \gamma \in \left( \frac{2}{3}, \frac{1}{2} \right) \). Then the estimator \( \vartheta^*_r \) is consistent, asymptotically normal
\[
ev^{-1} (\vartheta^*_r - \vartheta_0) \Rightarrow N \left( 0, I(\vartheta_0)^{-1} \right),
\]
and we have the convergence of polynomial moments.

**Proof.** Below we substitute the observations (3.2), where \( \vartheta = \vartheta_0 \), in the stochastic integral
\[
ev^{-1} (\vartheta^*_r - \vartheta_0) = ev^{-1} (\vartheta^*_r - \vartheta_0) + I_r (\vartheta^*_r)^{-1} t \int_{\tau}^{T} t \left[ ax_t + bx_{\vartheta^*_r t} \right] dW_t
\]
\[
+ t^2 I_r (\vartheta^*_r)^{-1} b^2 \int_{\tau}^{T} t \left[ ax_t + bx_{\vartheta^*_r t} \right] \left( X_{\vartheta^*_r t} - X_{\vartheta^*_r t} \right) dt.
\]
From the consistency of the estimator \( \vartheta^*_r \) and the continuity of the corresponding functions, we obtain the convergences
\[
b^2 \int_{\tau}^{T} t^2 \left[ ax_t + bx_{\vartheta^*_r t} \right]^2 dt \rightarrow 1(\vartheta_0), \quad I_r (\vartheta^*_r) \rightarrow 1(\vartheta_0).
\]
Hence, by the central limit theorem for stochastic integrals, we have the asymptotic normality
\[
I_r (\vartheta^*_r)^{-1} t \int_{\tau}^{T} t \left[ ax_t + bx_{\vartheta^*_r t} \right] dW_t \Rightarrow N \left( 0, I(\vartheta_0)^{-1} \right).
\]
Remark that for this model of observations, we have the asymptotic normality “in probability” too
\[
I_r (\vartheta^*_r)^{-1} t \int_{\tau}^{T} t \left[ ax_t + bx_{\vartheta^*_r t} \right] dW_t \rightarrow 1(\vartheta_0)^{-1} b \int_{0}^{T} t \left[ ax_t + bx_{\vartheta_0 t} \right] dW_t,
\]
where
\[
1(\vartheta_0)^{-1} b \int_{0}^{T} t \left[ ax_t + bx_{\vartheta_0 t} \right] dW_t \sim N \left( 0, I(\vartheta_0)^{-1} \right).
\]
Further, uniformly on \( t \in [0, T] \), we have
\[
X_{\vartheta^*_r t} - X_{\vartheta_0 t} = x_{\vartheta^*_r t} (\vartheta_0) - x_{\vartheta_0 t} (\vartheta_0) + \varepsilon \left[ x_{\vartheta^*_r t}^{(1)} (\vartheta_0) - x_{\vartheta_0 t}^{(1)} (\vartheta_0) \right] + \varepsilon^2 O_m (1)
\]
\[
= (\vartheta^*_r - \vartheta_0) t \left[ ax_{\vartheta_0 t} + bx_{\vartheta_0^2 t} \right] + \varepsilon \Delta_x + \varepsilon^2 O_m (1) + (\vartheta^*_r - \vartheta_0)^2 O_m (1),
\]
Multi-step Estimation of Delay

where

\[ \Delta_\varepsilon = x_{\vartheta_{\varepsilon}^*}^{(1)}(\vartheta_0) - x_{\vartheta_0}^{(1)}(\vartheta_0) = \int_{\vartheta_0}^{\vartheta_{\varepsilon}^*} \left[ ax_{\vartheta_0}^{(1)}(\vartheta_0) + bx_{\vartheta_0}^{(1)}(\vartheta_0) \right] ds + W_{\vartheta_{\varepsilon}^*} - W_{\vartheta_0} \]

\[ = (\vartheta_{\varepsilon}^* - \vartheta_0)O_m(1) + \left| \vartheta_{\varepsilon}^* - \vartheta_0 \right|^{1/2}O_m(1) = \left| \vartheta_{\varepsilon}^* - \vartheta_0 \right|^{1/2}O_m(1). \]

Therefore, we get

\[ X_{\vartheta_{\varepsilon}^*} - X_{\vartheta_0} = (\vartheta_{\varepsilon}^* - \vartheta_0) t \left[ ax_{\vartheta_0} + bx_{\vartheta_0} \right] + \varepsilon \left| \vartheta_{\varepsilon}^* - \vartheta_0 \right|^{1/2}O_m(1), \]

and we can write

\[ \varepsilon^{-1}I_{\vartheta_0} \left( \vartheta_{\varepsilon}^* - \vartheta_0 \right)^2 \int_{\tau_0}^{\tau_0^*} t \left[ ax_t + bx_{\vartheta_{\varepsilon}^*} \right] \left( X_{\vartheta_0} - X_{\vartheta_{\varepsilon}^*} \right) dt \]

\[ = - \left( \vartheta_{\varepsilon}^* - \vartheta_0 \right)^2 \varepsilon^{-1}I_{\vartheta_0} \left( \vartheta_{\varepsilon}^* - \vartheta_0 \right)^2 \int_{\tau_0}^{\tau_0^*} t^2 \left[ ax_t + bx_{\vartheta_{\varepsilon}^*} \right] \left[ ax_t + bx_{\vartheta_0} \right] dt \]

\[ + \left| \vartheta_{\varepsilon}^* - \vartheta_0 \right|^{1/2}O_m(1) \]

\[ = - \left( \vartheta_{\varepsilon}^* - \vartheta_0 \right)^2 \varepsilon^{-1} \left[ 1 + \left( \vartheta_{\varepsilon}^* - \vartheta_0 \right) O_m(1) \right] + \left| \vartheta_{\varepsilon}^* - \vartheta_0 \right|^{1/2}O_m(1). \]

Here, by the conditions of the Theorem 5, we have

\[ \left( \vartheta_{\varepsilon}^* - \vartheta_0 \right)^2 = \varepsilon^{2\delta}O_m(1) = o_m(1). \]

Finally, we obtain

\[ \varepsilon^{-1} \left( \vartheta_{\varepsilon}^* - \vartheta_0 \right) = \frac{b}{\varepsilon I(\vartheta_0)} \int_{\tau_0}^{\tau_0^*} t \left[ ax_t + bx_{\vartheta_0} \right] dW_t + o_m(1). \]

It can be verified that we have the uniform on compacts \( K \subset \Theta \) convergence of moments too

\[ \varepsilon^{-2} \sup_{\vartheta_0 \in K} E_{\vartheta_0} \left| \vartheta_{\varepsilon}^* - \vartheta_0 \right|^2 \longrightarrow \sup_{\vartheta_0 \in K} I(\vartheta_0). \]

Therefore, for any (small) \( \nu > 0 \), we have

\[ \lim_{\varepsilon \to 0} \varepsilon^{-2} \sup_{|\vartheta - \vartheta_0| \leq \nu} E_{\vartheta} \left| \vartheta_{\varepsilon}^* - \vartheta \right|^2 = \sup_{|\vartheta - \vartheta_0| \leq \nu} I(\vartheta) \]

and

\[ \lim_{\nu \to 0} \sup_{|\vartheta - \vartheta_0| \leq \nu} I(\vartheta) = I(\vartheta_0). \]

Hence, the Two-step MLE \( \vartheta_{\varepsilon}^* \) is asymptotically efficient (see (2.6)).
The procedure of estimation is the following: first we calculate the estimator of substitution $\bar{\varphi}_\tau$, then with the help of this estimator we construct the One-step MDE $\varphi^*_\tau$, which has the “good rate” and can be used for the construction of the Two-step MLE $\varphi^*_\varepsilon$.

This last estimator is asymptotically equivalent to the MLE.

Let us write together all these estimators. The first one is the estimator of substitution $\bar{\varphi}_\tau = (a + b) b \tau \varepsilon \left( X_\tau - x_0 - x_0 \tau (a + b) \right) - \frac{a}{b}$.

Then, using this estimator, we calculate the One-step MDE

$$\varphi^*_\tau = \bar{\varphi}_\tau + Q_{\tau} (\bar{\varphi}_\tau)^{-1} \int_{\tau}^{\tau} \left[ X_t - x_t (\bar{\varphi}_\tau) \right] \tilde{x}_t (\bar{\varphi}_\tau) dt.$$}

Finally, we calculate the Two-step MLE

$$\varphi^*_\varepsilon = \varphi^*_\tau + I_{\tau} (\varphi^*_\tau)^{-1} b \int_{\tau}^{T} t \left[ ax_t + bx_{\varphi^*_\tau} \right] \left[ dX_t - (aX_t + bX_{\varphi^*_\tau}) dt \right].$$}

The advantage of this procedure is the absence of problems of optimization like (1.7) and (2.1).

**Remark.** Note that it is possible to introduce the Two-step MLE-process $\varphi^*_t, \tau_\varepsilon < t \leq T$, which can be used in the problems, where we need estimators for all $t \in (0, T]$. For example, such estimators are used in the approximation of the solutions of backward stochastic differential equations [17]. Here this estimator-process can be written as follows

$$\varphi^*_t = \varphi^*_\tau + I_{\tau} (\varphi^*_\tau)^{-1} b \int_{\tau}^{T} [ax_s + bx_{\varphi^*_\tau}] \left[ dX_s - (aX_s + bX_{\varphi^*_\tau}) ds \right].$$}

where $t \in (\tau_\varepsilon, T]$ and

$$I^*_t (\varphi) = b^2 \int_{\tau}^{T} \tilde{x}_s (\varphi)^2 ds.$$

It can be shown that

$$\varepsilon^{-1} (\varphi^*_\varepsilon, \varphi) \Rightarrow N \left( 0, I^*_t (\varphi) \right).$$

For the proof of this convergence we can use the proof of Theorem 5, where we just put $T = t$. Moreover, it can be shown that we have the weak convergence of the stochastic process $\eta_{\varepsilon, \tau} = I^*_t (\varphi)$ (see [19]).

### 3.1. Nonlinear scale-type equation

We consider the following nonlinear SDE

$$dX_t = S (X_{\varphi_t}) dt + \varepsilon dW_t, \quad X_0 = x_0, \quad 0 \leq t \leq T,$$}
Multi-step Estimation of Delay

where $\vartheta \in \Theta = (\alpha, \beta)$, $0 < \alpha < \beta < 1$. Suppose that the function $S(x)$ has four continuous bounded derivatives. Recall that this equation has a unique strong solution (see Theorem 4.6 in [21]).

The limit differential equation is

$$\frac{dx_t}{dt} = S(x_{\vartheta t}), \quad x_0, \quad 0 \leq t \leq T.$$  

Its solution $x_t = x_t(\vartheta)$ is a function of $\vartheta$.

The preliminary MDE $\hat{\vartheta}_\varepsilon$ for this model of observation is defined by the same equation as in (2.1). This estimator is consistent, asymptotically normal, and the convergence of moments hold. The proof is similar to that of Theorem 1.

The Fisher information is

$$I(\vartheta) = \int_0^T t^2 S'(x_{\vartheta t}(\vartheta))^2 S(x_{\vartheta t}(\vartheta))^2 dt.$$  

The properties of the MLE were studied in [1], [13] (linear case) and [14]. It was shown that the MLE $\hat{\vartheta}_\varepsilon$ is consistent, asymptotically normal

$$\varepsilon^{-1}(\hat{\vartheta}_\varepsilon - \vartheta_0) \Rightarrow \mathcal{N}(0, I(\vartheta_0)^{-1})$$  

and asymptotically efficient. Its numerical calculation can have the same difficulties as in the case of the observations (1.1).

To introduce the One-step MLE we recall the definition of the functions $H(\cdot)$ and $\Psi(\cdot)$ for this model: $H(t, \vartheta) = t S'(x_{\vartheta t}(\vartheta)) S(x_{\vartheta t}(\vartheta))$

$$\Psi(\vartheta) = H(T, \vartheta) X_T - H(0, \vartheta)x_0 - \int_0^T H'_t(t, \vartheta) X_t dt.$$  

Note that the function $H(\cdot)$ and its derivatives are bounded on the set $[0, T] \times \Theta$. Then we can define

$$\vartheta^*_\varepsilon = \hat{\vartheta}_\varepsilon + 1 (\hat{\vartheta}_\varepsilon)^{-1} \left[ \Psi(\hat{\vartheta}_\varepsilon) - \int_0^T t S'(x_{\hat{\vartheta}_\varepsilon t}(\hat{\vartheta}_\varepsilon)) S(x_{\hat{\vartheta}_\varepsilon t}(\hat{\vartheta}_\varepsilon)) S(X_{\hat{\vartheta}_\varepsilon t}) dt \right].$$

This One-step MLE is consistent and asymptotically normal

$$\varepsilon^{-1}(\vartheta^*_\varepsilon - \vartheta_0) \Rightarrow \mathcal{N}(0, I(\vartheta_0)^{-1}).$$  

The proof follows the same lines as that of Theorem 2.

4. Discussion

The proposed result can be generalized to some slightly more general models. For example, the cases of observations

$$dX_t = [S_1(X_t) + S_2(X_{t-\vartheta})]dt + \varepsilon \sigma(t, X_t)dW_t, \quad X_s = x_s, s \leq 0, \quad 0 \leq t \leq T,$$
\[ \dot{X}_t = S(\vartheta, t, X_{t-\vartheta}) \, dt + \varepsilon \sigma(t, X_t) \, dW_t, \quad X_s = x_s, s \leq 0, \quad 0 \leq t \leq T, \]
can be treated in exactly the same way as above. Here the functions \( S_1(\cdot), S_2(\cdot), S(\cdot) \) and \( \sigma(\cdot) \) are supposed to be deterministic and smooth.

Another model which can be treated in a similar way is
\[ \dot{X}_t = S(X_{t-f(\vartheta,t)}) \, dt + \varepsilon \, dW_t, \quad X_s = x_s, s \leq 0, \quad 0 \leq t \leq T. \]
Here \( f(\vartheta, t) \geq 0 \) is some known smooth function. If \( f(\vartheta, t) = \vartheta \) or \( f(\vartheta, t) = t - \vartheta t \), we obtain the models (1.1) and (3.7) respectively.

It is also possible to study the models with multiple delays like
\[ \dot{X}_t = \sum_{k=1}^{K} S_k(X_{t-\vartheta_k}) \, dt + \varepsilon \, dW_t, \quad X_s = x_s, s \leq 0, \quad 0 \leq t \leq T. \]
In the last model, the delays form a vector \( \vartheta = (\vartheta_1, \ldots, \vartheta_K) \), where \( \vartheta_1 < \vartheta_2 < \ldots < \vartheta_K \), and the problem of estimation of \( \vartheta \) become more complicated. However, the construction of the One-step procedure can be realized using an appropriate modification of the given above procedure.

5. Appendix

We suppose that the function \( S(x), x \in \mathcal{R} \), has two continuous bounded derivatives \( |S'(x)| \leq L \) and \( |S''(x)| \leq M \), and hence the equation (1.1) has a unique strong solution (see Theorem 4.6 in [21]). Moreover, we have the following estimates (with probability 1)
\[ |X_t - x_t| \leq \int_0^t |S(X_{s-\vartheta_0}) - S(x_{s-\vartheta_0})| \, ds + \varepsilon |W_t| \]
\[ \leq L \int_0^t |X_{s-\vartheta_0} - x_{s-\vartheta_0}| \, ds + \varepsilon |W_t| \leq L \int_0^t |X_s - x_s| \, ds + \varepsilon \sup_{0 \leq s \leq T} |W_s|. \]

Hence, by the Grönwall-Bellman lemma, we can write
\[ |X_t - x_t| \leq \varepsilon C \overline{W}_T, \]
where we denoted \( \overline{W}_T = \sup_{0 \leq s \leq T} |W_s| \).

For the second moment, we have
\[ \mathbb{E}_S |X_t - x_t|^2 \leq 2L^2 T \int_0^t \mathbb{E}_S |X_{s-\vartheta_0} - x_{s-\vartheta_0}|^2 \, ds + 2\varepsilon^2 T \]
\[ \leq 2L^2 T \int_0^t \mathbb{E}_S |X_s - x_s|^2 \, ds + 2\varepsilon^2 T, \]
and by the same lemma
\[ \mathbf{E}_s |X_t - x_t|^2 \leq C \varepsilon^2. \tag{5.1} \]
Recall that for \( s \leq 0 \), we have \( X_s - x_s = x_0 - x_0 = 0 \).

Writing the formal derivative \( X_t^{(1)} \) of \( X_t \) w.r.t. \( \varepsilon \), we obtain the equation
\[
dX_t^{(1)} = S'_x (X_{t-\vartheta}) \, X_{t-\vartheta}^{(1)} \, dt + dW_t, \quad X_s^{(1)} = 0, \quad s \leq 0, \quad 0 \leq t \leq T.
\]
The Gaussian process \( X_t^{(1)} \) satisfies the linear equation
\[
dX_t^{(1)} = S'_x (x_{t-\vartheta}) \, x_{t-\vartheta}^{(1)} \, dt + dW_t, \quad x_s^{(1)} = 0, \quad s \leq 0, \quad 0 \leq t \leq T.
\]
The proof of
\[
V_\varepsilon (t) = \mathbf{E}_{\vartheta_0} \left| \frac{X_t - x_t (\vartheta_0)}{\varepsilon} - x_t^{(1)} (\vartheta_0) \right|^2 \to 0
\]
can be carried out using the standard technique based on the Grönwall-Bellman lemma as follows. We write
\[
v_\varepsilon (t) = \frac{X_t - x_t (\vartheta_0)}{\varepsilon} - x_t^{(1)} = \int_0^t \left[ S (X_{s-\vartheta}) - S (x_{s-\vartheta}) - S' (x_{s-\vartheta}) \, x_{s-\vartheta}^{(1)} \right] ds
\]
\[
= \int_0^t \left[ S'(X_{s-\vartheta}) \frac{X_{s-\vartheta} - x_{s-\vartheta}}{\varepsilon} - S' (x_{s-\vartheta}) \right] ds
\]
\[
= \int_0^t S' (\tilde{X}_{s-\vartheta}) \left[ \frac{X_{s-\vartheta} - x_{s-\vartheta}}{\varepsilon} - x_{s-\vartheta}^{(1)} \right] ds
\]
\[
+ \int_0^t S' (\tilde{X}_{s-\vartheta}) \left[ S' (\tilde{X}_{s-\vartheta}) - S' (x_{s-\vartheta}) \right] x_{s-\vartheta}^{(1)} ds,
\]
where \( |\tilde{X}_{s-\vartheta} - x_{s-\vartheta}| \leq |X_{s-\vartheta} - x_{s-\vartheta}| \). Hence, we have
\[
\mathbf{E}_{\vartheta_0} |v_\varepsilon (t)|^2 \leq L^2 T \int_0^t \mathbf{E}_{\vartheta_0} |v_\varepsilon (s - \vartheta_0)|^2 \, ds + M^2 T \int_0^T \mathbf{E}_{\vartheta_0} |X_{s-\vartheta} - x_{s-\vartheta}|^2 \, ds
\]
and
\[
V_\varepsilon (t) \leq L^2 T \int_0^t V_\varepsilon (s - \vartheta_0) \, ds + M^2 T^2 C \varepsilon^2 \leq L^2 T \int_0^t V_\varepsilon (s) \, ds + M^2 T^2 C \varepsilon^2.
\]
Therefore, by the Grönwall-Bellman lemma, we get \( V_\varepsilon (t) \to 0 \).

Let us study the quantity \( R_\varepsilon (t) = \mathbf{E}_{\vartheta_0} |X_t - x_t (\vartheta_0)|^2 \), where \( x_t (\vartheta_0) \) and \( X_t \) are solutions of the equations (3.1) and (3.2) respectively. For the difference \( r_\varepsilon (t) = X_t - x_t (\vartheta_0) \), we have the relation
\[
r_\varepsilon (t) = \int_0^t [ar_\varepsilon (s) + br_\varepsilon (\vartheta_0 s)] \, ds + \varepsilon W_t.
\]
Therefore, we get
\[
\mathbb{E}_{\vartheta_0} |r_\varepsilon(t)|^p \leq C T^{p-1} |a|^p \int_0^t \mathbb{E}_{\vartheta_0} |r_\varepsilon(s)|^p \, ds + C |b|^p \vartheta_0^{-1} \int_0^{\vartheta_0 t} \mathbb{E}_{\vartheta_0} |r_\varepsilon(s)|^p \, ds + C \varepsilon^p T^\frac{p}{2}
\]
and
\[
R_\varepsilon(t) \leq \left[ C T^{p-1} |a|^p + C |b|^p \vartheta_0^{-1} \right] \int_0^t \mathbb{E}_{\vartheta_0} |r_\varepsilon(s)|^p \, ds + C \varepsilon^p T^\frac{p}{2}.
\]
Now, by the Grönwall-Bellman lemma, we obtain the estimate
\[
\mathbb{E}_{\vartheta_0} |X_t - x_t(\vartheta_0)|^p \leq C \varepsilon^p
\]
with a corresponding constant \( C > 0 \).

**Acknowledgments**

I would like to thank Reviewer and Associate Editor for many useful comments. This research was supported by the RSF project no. 20-61-47043.

**References**

[1] Apoyan G.T. (1986) Parameter estimation of non differentiable trend coefficient (in Russian), *Uchen. Zap. Erevan University* 1, 33–62.

[2] Bandrivskyy, A., Luchinsky, D.G., McClintock, P.V.E., Smelyanskiy, V.N. and Stefanovska, A. (2005) Inference of systems with delay and applications to cardiovascular dynamics, *Stochastics and Dynamics* 5, 2, 321–331.

[3] Davies, R.B. (1986) Asymptotic inference when the amount of information is random, in L.M. Le Cam & R.A. Olshen (eds), *Proceedings of the Berkeley symposium in honor of J. Neyman and J. Kiefer*, V. II, Wadsworth, 841–864.

[4] Frank, T.D., Friedrich, R. and Beek, P.J. (2005) Time series analysis of multivariate time-delayed systems with noise: applications to laser physics and human movement, *Stochastics and Dynamics* 5, 2, 297–306.

[5] Freidlin M.I. and Wentzell A.D. (1984) *Random Perturbations of Dynamical Systems*, Springer, New York.

[6] Griebel, T. (2017) *The Pantograph Equation in Quantum Calculus*, Missouri University of Science and Technology, Masters Theses.

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

[8] Hauptmann, C., Popovych, O. and Tass, P.A. (2005) Multisite coordinated delayed feedback for an effective desynchronization of neuronal networks, *Stochastics and Dynamics* 5, 2, 307–319.
[9] IBRAIMOV I.A. AND HAS’MINSKII R.Z. (1981) Statistical Estimation: Asymptotic Theory, Springer, Berlin.
[10] KUANG, Y. (1993) Delay Differential Equations: With Applications in Population Dynamics, Academic Press, London.
[11] KÜCHLER, U., KUTOYANTS, YU.A. (2000) Delay estimation for some stationary diffusion-type processes, Scandinavian Journal of Statistics 27, 3, 405–414.
[12] KÜCHLER, U., MENSCH, B. (1992) Langevins stochastic differential equation extended by a time-delayed term, Stochastics and Stochastic Reports 40, 23–42.
[13] KUTOYANTS, YU.A. (1988). An example of estimating a parameter of a nondifferentiable drift coefficient, Theory Probab. Appl. 33, 1, 175–179.
[14] KUTOYANTS, YU.A. (1994) Identification of Dynamical Systems with Small Noise, Kluwer, Dordrecht.
[15] KUTOYANTS, YU.A. (2004) Statistical Inference for Ergodic Diffusion Processes, Springer, London.
[16] KUTOYANTS, YU.A. (2005) On delay estimation for stochastic differential equations, Stochastics and Dynamics 5, 2, 333–342.
[17] KUTOYANTS, YU.A. (2014) Approximation of the solution of the backward stochastic differential equation. Small noise, large samples and high frequency cases, Proceedings of the Steklov Institute of Mathematics 287, 133–154.
[18] KUTOYANTS, YU.A. (2017) On the multi-step MLE-process for ergodic diffusion, Stochastic Processes and their Applications 127, 2243–2261.
[19] KUTOYANTS, Y.A. AND ZHOU, L. (2014) On approximation of the backward stochastic differential equation, J. Stat. Plann. Infer. 150, 111–123.
[20] LE CAM L. AND YANG, G.L. (2000) Asymptotics in Statistics: Some Basic Concepts, 2nd ed., Springer, New York.
[21] LIPTSER, R.S. AND SHIRYAEV, A.N. (2001). Statistics of Random Processes, 2-nd ed., vol. 1, Springer, New York.
[22] MAO, X. (2005) Delay population dynamics and environmental noise, Stochastics and Dynamics 5, 2, 149–162.
[23] MYSHKIS, A.D. (1951) Linear Differential Equations with Delayed Argument (in Russian), Nauka, Moscow.
[24] OCKENDON, J.R., TAYLER, A.B. (1971) The dynamics of a current collection system for an electric locomotive, in Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences 322, 447–468.
[25] OHIRA, T., YAMANE, T. (2000) Delayed stochastic systems, Physical Review. E, Statistical Physics, Plasmas, Fluids and Related Interdisciplinary Topics 61, 2, 1247–1257.
[26] ROSENBERG, M.L., TARJUS, G. AND MUNAKATA, T. (2018) Influence of time delay on information exchanges between coupled linear stochastic systems, Physical Review. E 98, 3, 032130.