Discriminant analysis for discretely observed ergodic diffusion processes

Masayuki Uchida (Osaka University and JST)
and
Nakahiro Yoshida (University of Tokyo)

Plan of today's talk

1. Introduction: discriminant analysis for SDE

2. asymptotic properties of discriminant functions when $nh_n^3 \to 0$

3. Examples and simulation results

Introduction

We consider d-dimensional diffusion models Π_k $(k = 0, 1, 2)$ defined by the following stochastic differential equations

$$\Pi_k : \quad dX_t^{(k)} = a^{(k)}(X_t^{(k)}, \alpha^{(k)})dt + b^{(k)}(X_t^{(k)}, \beta^{(k)})dw_t^{(k)}, \quad t \geq 0, \quad X_0^{(k)}, \tag{1}$$

where

$w^{(k)}$ is an r-dimensional standard Wiener process,

$w^{(0)}$, $w^{(1)}$, $w^{(2)}$, $X_0^{(0)}$, $X_0^{(1)}$ and $X_0^{(2)}$ are mutually independent and

$\theta^{(k)} = (\alpha^{(k)}, \beta^{(k)}) \in \Theta_\alpha^{(k)} \times \Theta_\beta^{(k)} = \Theta^{(k)}$ with $\Theta_\alpha^{(k)}$ and $\Theta_\beta^{(k)}$ being compact convex subsets of $\mathbf{R}^{p_k}$ and $\mathbf{R}^{q_k}$, respectively.

Let $k = 0, 1, 2$.

$a^{(k)} : \mathbf{R}^d \times \Theta_\alpha^{(k)} \to \mathbf{R}^d \otimes \mathbf{R}^r$ and $b^{(k)} : \mathbf{R}^d \times \Theta_\beta^{(k)} \to \mathbf{R}^d$.

$\theta_*^{(k)} = (\alpha_*^{(k)}, \beta_*^{(k)})$ is the true value of $\theta^{(k)}$ and we assume that $\theta_*^{(k)} \in \mathrm{Int}(\Theta^{(k)})$.

Moreover, the training data are discrete observations $\mathbf{X}_n^{(k)} = (X_{t_i^n}^{(k)})_{0 \leq i \leq n}$ obtained from the diffusion model Π_k, where $t_i^n = ih_n$ and $t_n^n = nh_n =: T_n$.

We assume that the data

$\mathbf{X}_n^{(0)} = (X_{t_i^n}^{(0)})_{0 \le i \le n}$ are obtained from either Π_1 or Π_2,

which means that the distribution of model Π_0 is the same as that of one of the two models Π_1 and Π_2.

We will consider the situation when $h_n \to 0$, $nh_n \to \infty$ and $nh_n^3 \to 0$ as $n \to \infty$.

First of all, in this paper, we treat a discriminant analysis for two discretely observed ergodic diffusion processes. For example, in Table I and Table II below, two sample paths $\mathbf{X}_n^{(1)}$ and $\mathbf{X}_n^{(2)}$ are obtained from the diffusion models Π_1 and Π_2 defined by the following stochastic differential equations

$$\begin{aligned}
\Pi_1: \quad dX_t^{(1)} &= (1 - 2X_t^{(1)})dt + 4dw_t^{(1)}, \quad t \geq 0, \quad X_0^{(1)} = 10, \\
\Pi_2: \quad dX_t^{(2)} &= (1.5 - 2.5X_t^{(2)} + 4\sin(X_t^{(2)}))dt + 4dw_t^{(2)}, \quad t \geq 0, \quad X_0^{(2)} = 10,
\end{aligned}$$

respectively. In table 3 below, a new sample path $\mathbf{X}_n^{(0)}$ is obtained from either Π_1 or Π_2.

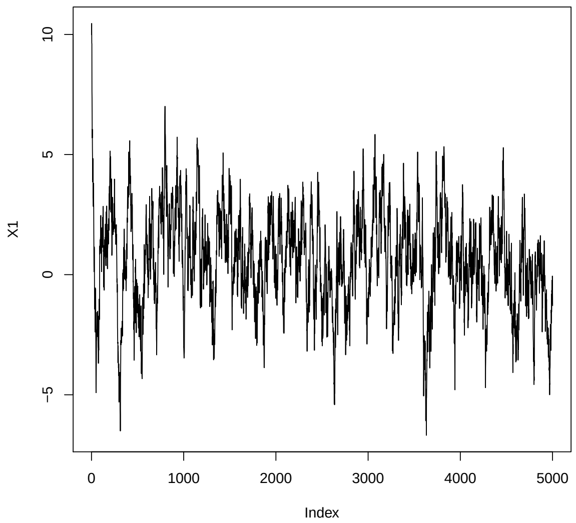

Table I. sample path of $\mathbf{X}_n^{(1)}$ from Π_1

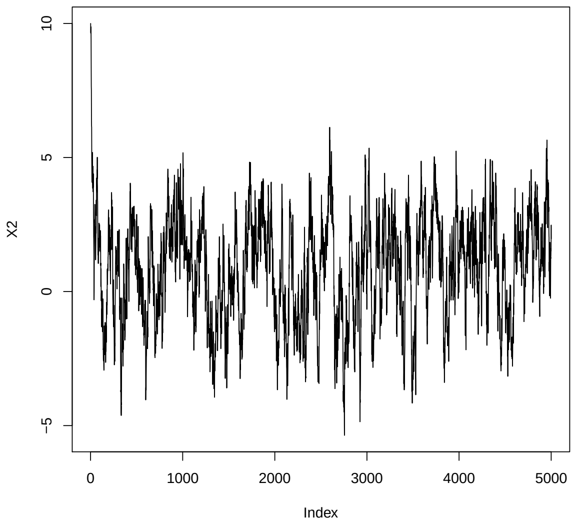

Table II. sample path of $\mathbf{X}_n^{(2)}$ from Π_2

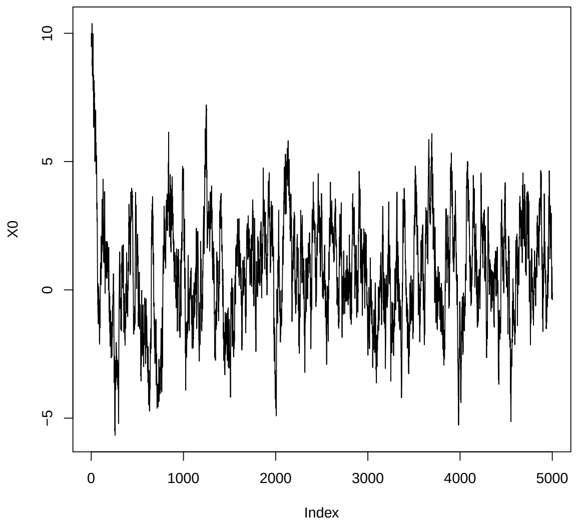

Table III. sample path of $\mathbf{X}_n^{(0)}$ from either Π_1 or Π_2

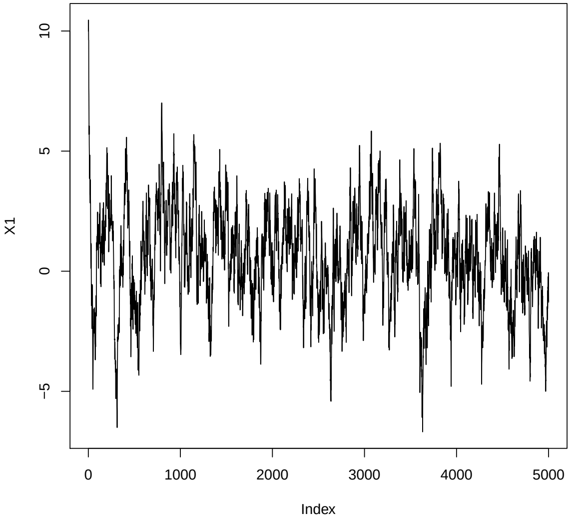

Table I. sample path of $\mathbf{X}_n^{(1)}$ from Π_1

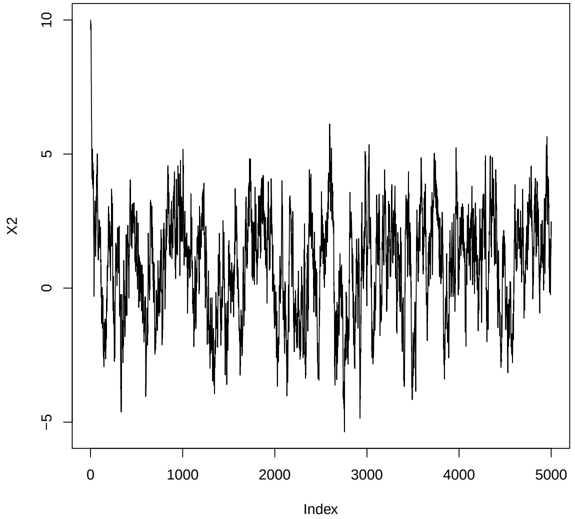

Table II. sample path of $\mathbf{X}_n^{(2)}$ from Π_2

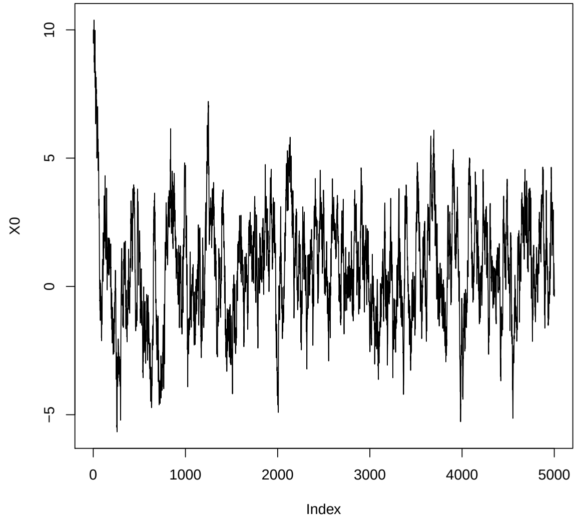

Table III. sample path of $\mathbf{X}_n^{(0)}$ from either Π_1 or Π_2

We treat a discriminant problem whether $\mathbf{X}_n^{(0)}$ is obtained from Π_1 or Π_2 by using training data $\mathbf{X}_n^{(k)}$ from Π_k for $k = 1, 2$.

Since the sample path of $\mathbf{X}_n^{(0)}$ is similar to the one of $\mathbf{X}_n^{(2)}$ compared with $\mathbf{X}_n^{(1)}$, from viewpoint of "similarity of sample path", one may classify $\mathbf{X}_n^{(0)}$ into Π_2.

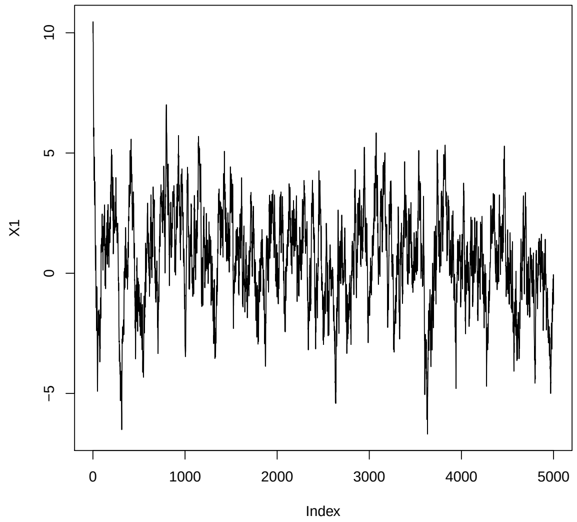

Table I. sample path of $\mathbf{X}_n^{(1)}$ from Π_1

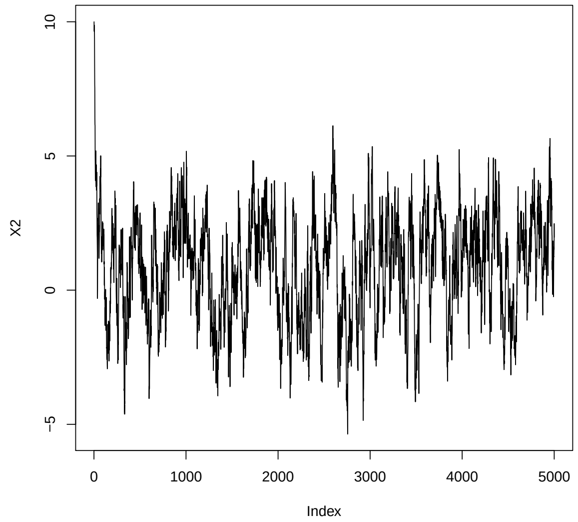

Table II. sample path of $\mathbf{X}_n^{(2)}$ from Π_2

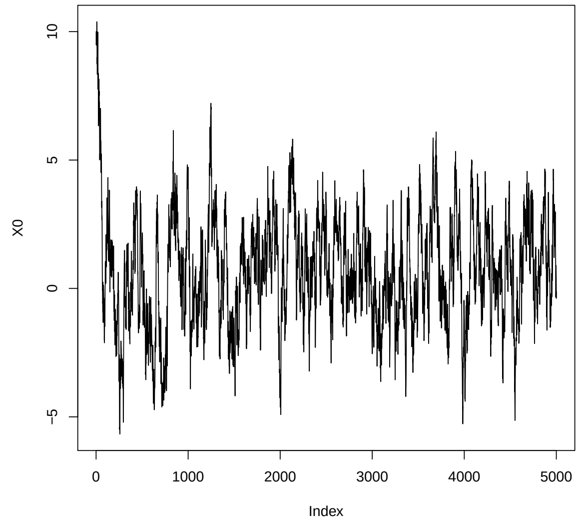

Table III. sample path of $\mathbf{X}_n^{(0)}$ from either Π_1 or Π_2

Unfortunately, the answer is wrong. In fact, $\mathbf{X}_n^{(0)}$ in Table 3 are the simulation data obtained from model Π_1.

This illustration cautions us about the risk of misclassification based on "similarity of sample path".

Therefore, we propose classification criteria based on discriminant functions for stochastic differential equations.

For the ergodic diffusion models, the asymptotic distributions of discriminant functions are obtained under the two situations where the volatility functions are same or not.

Discriminant analysis

Let $C_{\uparrow}^{k,l}(\mathbf{R}^d \times \Theta; \mathbf{R}^d)$ denote the space of all functions f satisfying the following conditions: (i) $f(x,\theta)$ is an $\mathbf{R}^d$-valued function on $\mathbf{R}^d \times \Theta$, (ii) $f(x,\theta)$ is continuously differentiable with respect to x up to order k for all θ. (iii) for $|\mathbf{n}| = 0, 1, \ldots, k$, $\partial^{\mathbf{n}} f(x,\theta)$ is continuously differentiable with respect to θ up to order l for all x. Moreover, for $|\nu| = 1, \ldots, l$ and $|\mathbf{n}| = 0, 1, \ldots, k$, $\delta^{\nu}\partial^{\mathbf{n}} f(x,\theta)$ is of polynomial growth in x uniformly in θ. Here $\mathbf{n} = (n_1, \ldots, n_d)$ and $\nu = (\nu_1, \ldots, \nu_m)$ are multi-indices, $m = \dim(\Theta)$, $|\mathbf{n}| = n_1 + \ldots + n_d$, $|\nu| = \nu_1 + \ldots + \nu_m$, $\partial^{\mathbf{n}} = \partial_1^{n_1} \cdots \partial_d^{n_d}$, $\partial_i = \partial/\partial x_i$, and $\delta^{\nu} = \delta_{\theta_1}^{\nu_1} \cdots \delta_{\theta_m}^{\nu_m}$, $\delta_{\theta_i} = \partial/\partial \theta_i$.

Let $\mathcal{F}_{\uparrow}(\mathbf{R}^d)$ be the space of all measurable functions f satisfying that $f(x)$ is an $\mathbf{R}$-valued function on $\mathbf{R}^d$ with polynomial growth in x.

$P_{\theta^{(k)}}$ denotes the law of the process defined by the equation (1).

Set $\Delta X_i^{(k)} = X_{t_i^n}^{(k)} - X_{t_{i-1}^n}^{(k)}$ and $B^{(k)}(x, \beta^{(k)}) = b^{(k)}(b^{(k)})^{\star}(x, \beta^{(k)})$, where $\star$ denotes the transpose.

Let $L_{\theta^{(0)}}$ be the infinitesimal generator of the diffusion (1) with $k = 0$: $L_{\theta^{(0)}} = \sum_{i=1}^{d} a_i^{(0)}(x, \alpha^{(0)})\partial_i + \frac{1}{2}\sum_{i,j=1}^{d} B_{ij}^{(0)}(x, \beta^{(0)})\partial_i\partial_j$.

Let $\stackrel{p}{\to}$ and $\stackrel{d}{\to}$ be the convergence in probability and the convergence in distribution, respectively.

For matrices A and B of the same size, we define $A^{\otimes 2} = AA^{\star}$ and $B[A] = tr(BA^{\star})$.

Let $k = 1, 2$. We make the following assumption.

[A1] (i) There exists $K > 0$ such that for all $x, y \in \mathbf{R}^d$,

$$\sup_{\alpha^{(k)} \in \Theta_\alpha^{(k)}} |a^{(k)}(x, \alpha^{(k)}) - a^{(k)}(y, \alpha^{(k)})| + \sup_{\beta^{(k)} \in \Theta_\beta^{(k)}} |b^{(k)}(x, \beta^{(k)}) - b^{(k)}(y, \beta^{(k)})| \leq K|x - y|.$$

(ii) $\inf_{x,\beta^{(k)}} \det(B^{(k)}(x, \beta^{(k)})) > 0$.
(iii) $a^{(k)} \in C_\uparrow^{2,2}(\mathbf{R}^d \times \Theta_\alpha^{(k)}; \mathbf{R}^d)$. $b^{(k)} \in C_\uparrow^{2,2}(\mathbf{R}^d \times \Theta_\beta^{(k)}; \mathbf{R}^d \otimes \mathbf{R}^r)$.
(iv) There exists a unique invariant probability measure $\mu_{\theta_*^{(k)}}$ of $X_t^{(k)}$ and for any $f \in \mathcal{F}_\uparrow(\mathbf{R}^d)$ satisfying $\int_{\mathbf{R}^d} |f(x)| \mu_{\theta_*^{(k)}}(dx) < \infty$, as $T \to \infty$,

$$\frac{1}{T} \int_0^T f(X_t^{(k)}) dt \xrightarrow{p} \int_{\mathbf{R}^d} f(x) \mu_{\theta_*^{(k)}}(dx).$$

Let $\widehat{\theta}_n^{(k)} := (\widehat{\alpha}^{(k)}(\mathbf{X}_n^{(k)}), \widehat{\beta}^{(k)}(\mathbf{X}_n^{(k)}))$ be an estimator of $\theta^{(k)} = (\alpha^{(k)}, \beta^{(k)})$ for model Π_k based on the data $\mathbf{X}_n^{(k)}$.

Setting $\bar{B}_{i-1}^{(k)}(\beta^{(k)}) = B^{(k)}(X_{t_{i-1}^n}^{(0)}, \beta^{(k)})$, $\bar{a}_{i-1}^{(k)}(\alpha^{(k)}) = a^{(k)}(X_{t_{i-1}^n}^{(0)}, \alpha^{(k)})$ and

$$\begin{aligned} u_n^{(k)}(\mathbf{X}_n^{(0)}, \theta^{(k)}) &= -\frac{1}{2}\sum_{i=1}^{n}\Big\{h_n^{-1}(\bar{B}_{i-1}^{(k)}(\beta^{(k)}))^{-1}[(\Delta X_i^{(0)} - h_n\bar{a}_{i-1}^{(k)}(\alpha^{(k)}))^{\otimes 2}] \\ &\quad + \log\det(\bar{B}_{i-1}^{(k)}(\beta^{(k)}))\Big\}, \end{aligned}$$

one uses the following discriminant function

$$U_n(\mathbf{X}_n^{(0)}, \widehat{\theta}_n^{(1)}, \widehat{\theta}_n^{(2)}) = u_n^{(1)}(\mathbf{X}_n^{(0)}, \widehat{\theta}_n^{(1)}) - u_n^{(2)}(\mathbf{X}_n^{(0)}, \widehat{\theta}_n^{(2)}).$$

We suggest a discriminant rule such that

$\mathbf{X}_n^{(0)}$ is classified into Π_1 if $U_n(\mathbf{X}_n^{(0)}, \widehat{\theta}_n^{(1)}, \widehat{\theta}_n^{(2)}) \geq 0$ and otherwise into Π_2.

Let two misclassification probabilities denote

$$\begin{aligned} p_n(1|2) &= P(U_n(\mathbf{X}_n^{(0)}, \widehat{\theta}_n^{(1)}, \widehat{\theta}_n^{(2)}) < 0 \mid \Pi_1), \\ p_n(2|1) &= P(U_n(\mathbf{X}_n^{(0)}, \widehat{\theta}_n^{(1)}, \widehat{\theta}_n^{(2)}) \geq 0 \mid \Pi_2). \end{aligned}$$

We make the following assumption.

$[A2]$ As $h_n \to 0$ and $nh_n \to \infty$, $\widehat{\theta}_n^{(k)} \to^p \theta_*^{(k)}$.

Proposition 1 *Assume* $[A1]$ *and* $[A2]$. *Then, as* $h_n \to 0$ *and* $nh_n \to \infty$,

$$p_n(1|2) + p_n(2|1) \to 0.$$

Set

$$\begin{aligned}
\Gamma_a^{(k)}(\theta_*^{(k)})_{ij} &= \int_{\mathbf{R}^d} (\partial_{\alpha_i} a^{(k)}(x,\alpha_*^{(k)}))^\star (B^{(k)}(x,\beta_*^{(k)}))^{-1} \partial_{\alpha_j} a^{(k)}(x,\alpha_*^{(k)}) \mu_{\theta_*^{(k)}}(dx), \\
\Gamma_b^{(k)}(\beta_*^{(k)})_{lm} &= \frac{1}{2}\int_{\mathbf{R}^d} tr\{(B^{(k)})^{-1}(\partial_{\beta_l} B^{(k)})(B^{(k)})^{-1}(\partial_{\beta_m} B^{(k)})(x,\beta_*^{(k)})\} \mu_{\theta_*^{(k)}}(dx)
\end{aligned}$$

for $i,j = 1,\dots,p_k$, and $l,m = 1,\dots,q_k$. Here we suppose that $\Gamma_a^{(k)}(\theta_*^{(k)})$ and $\Gamma_b^{(k)}(\beta_*^{(k)})$ are non-singular.

Moreover, in order to obtain the asymptotic distributions of the discriminant function below, we make the assumption as follows.

$[A3]$ As $nh_n^3 \to 0$,

$$(\sqrt{nh_n}(\widehat{\alpha}_n^{(k)} - \alpha_*^{(k)}), \sqrt{n}(\widehat{\beta}_n^{(k)} - \beta_*^{(k)})) \xrightarrow{d} N_{p_k+q_k}(0, \mathrm{diag}[(\Gamma_a^{(k)}(\theta_*^{(k)}))^{-1}, (\Gamma_b^{(k)}(\beta_*^{(k)}))^{-1}]).$$

Remark 1 *For an estimator satisfying* $[A3]$*, we can refer Yoshida (1992), Kessler (1995, 1997) and Uchida and Yoshida (2012). Let* $B^{(k)}_{i-1}(\beta)=B^{(k)}(X^{(k)}_{t^n_{i-1}},\beta)$ *and* $a^{(k)}_{i-1}(\alpha)=a^{(k)}(X^{(k)}_{t^n_{i-1}},\alpha)$ *for* $k=1,2$*. For example, we use the following utility functions.*

$$
\begin{aligned}
\mathcal{V}^{(k)}_{1,n}(\beta) &= -\frac{1}{2}\sum_{i=1}^{n}\left\{h_n^{-1}(B^{(k)}_{i-1}(\beta))^{-1}[(\Delta X^{(k)}_i)^{\otimes 2}]+\log\det(B^{(k)}_{i-1}(\beta))\right\}.\\
\mathcal{V}^{(k)}_{2,n}(\alpha,\beta) &= -\frac{1}{2}\sum_{i=1}^{n}h_n^{-1}(B^{(k)}_{i-1}(\beta))^{-1}[(\Delta X^{(k)}_i-h_n a^{(k)}_{i-1}(\alpha))^{\otimes 2}].\\
\mathcal{V}^{(k)}_{3,n}(\beta;\bar\theta) &= -\frac{1}{2}\sum_{i=1}^{n}\left\{h_n^{-1}(B^{(k)}_{i-1}(\beta))^{-1}[(\Delta X^{(k)}_i)^{\otimes 2}-h_n^2\bar D^{(k)}_{2,i-1}(\bar\theta)]+\log\det(B^{(k)}_{i-1}(\beta))\right\},
\end{aligned}
$$

where for $l,m=1,\ldots,d$*,*

$$
\begin{aligned}
\bar D^{(k)}_{2,i-1}(\theta) &= \gamma^{(k)}_2(X_{t^n_{i-1}},\theta)+(a^{(k)}(X_{t^n_{i-1}},\alpha))^{\otimes 2},\\
\gamma^{(k)}_2(x,\theta)_{lm} &= \frac{1}{2}\left\{L_{\theta^{(k)}}B^{(k)}(x,\beta)_{lm}+\sum_{j=1}^{d}\{(\partial_{x_j}a^{(k)}(x,\alpha)_l)B^{(k)}(x,\beta)_{jm}+(\partial_{x_j}a^{(k)}(x,\alpha)_m)B^{(k)}(x,\beta)_{jl}\}\right\}.
\end{aligned}
$$

The adaptive estimators $\tilde{\beta}_n^{(k)}$, $\hat{\alpha}_n^{(k)}$ and $\hat{\beta}_n^{(k)}$ are defined as

$$\begin{aligned}
\mathcal{V}_{1,n}^{(k)}(\tilde{\beta}_n^{(k)}) &= \sup_{\beta^{(k)}} \mathcal{V}_{1,n}^{(k)}(\beta^{(k)}), \\
\mathcal{V}_{2,n}^{(k)}(\hat{\alpha}_n^{(k)}, \tilde{\beta}_n^{(k)}) &= \sup_{\alpha^{(k)}} \mathcal{V}_{2,n}^{(k)}(\alpha^{(k)}, \tilde{\beta}_n^{(k)}), \\
\mathcal{V}_{3,n}^{(k)}(\hat{\beta}_n^{(k)}; \hat{\alpha}_n^{(k)}, \tilde{\beta}_n^{(k)}) &= \sup_{\beta^{(k)}} \mathcal{V}_{3,n}^{(k)}(\beta^{(k)}; \hat{\alpha}_n^{(k)}, \tilde{\beta}_n^{(k)}).
\end{aligned}$$

Then, under some regularity conditions, as $nh_n^2 \to 0$.

$$(\sqrt{nh_n}(\hat{\alpha}_n^{(k)} - \alpha_*^{(k)}), \sqrt{n}(\tilde{\beta}_n^{(k)} - \beta_*^{(k)})) \xrightarrow{d} N_{p_k+q_k}(0, \mathrm{diag}[(\Gamma_a^{(k)}(\theta_*^{(k)}))^{-1}, (\Gamma_b^{(k)}(\beta_*^{(k)}))^{-1}]).$$

Furthermore, under some regularity conditions, as $nh_n^3 \to 0$,

$$(\sqrt{nh_n}(\hat{\alpha}_n^{(k)} - \alpha_*^{(k)}), \sqrt{n}(\hat{\beta}_n^{(k)} - \beta_*^{(k)})) \xrightarrow{d} N_{p_k+q_k}(0, \mathrm{diag}[(\Gamma_a^{(k)}(\theta_*^{(k)}))^{-1}, (\Gamma_b^{(k)}(\beta_*^{(k)}))^{-1}]).$$

For details, see the adaptive estimator in Kessler (1995) and the Type III adaptive estimator in Uchida and Yoshida (2012).

Case I: $B^{(1)} \neq B^{(2)}$

In this section, we consider the asymptotic distribution of the discriminant function U_n for the case where the volatility function of Π_1 is different from the one of Π_2, that is, $\mu_{\theta_*^{(0)}}(B^{(1)}(x,\beta_*^{(1)}) \neq B^{(2)}(x,\beta_*^{(2)})) > 0$.

Let

$$
\begin{aligned}
M(x) &= -\frac{1}{2}\left\{\log\frac{\det B^{(1)}(x,\beta_*^{(1)})}{\det B^{(2)}(x,\beta_*^{(2)})} + \left\{(B^{(1)}(x,\beta_*^{(1)}))^{-1} - (B^{(2)}(x,\beta_*^{(2)}))^{-1}\right\}[B^{(0)}(x,\beta_*^{(0)})]\right\}, \\
\bar{M} &= \int_{\mathbf{R}^d} M(x)\mu_{\theta_*^{(0)}}(dx), \\
f(x) &= M(x) - \bar{M}.
\end{aligned}
$$

$[A4]$ There exist $G_f(x)$, $\partial_{x_i}G_f(x) \in \mathcal{F}_{\uparrow}(\mathbf{R})$ $(i = 1,\ldots,d)$ such that for all x,

$$L_{\theta_*^{(0)}}G_f(x) = f(x).$$

Remark 2 *(i) For a sufficient condition for* $[A4]$*, see Pardoux and Veretennikov (2001). For example, in addition to* $[A1]$-(i)-(ii)-(iii)*, we assume that* $\sup_{x,\beta^{(k)}} |B^{(k)}(x, \beta^{(k)})| < \infty$ *and that there exist* $c_0 > 0$*,* $M_0 > 0$ *and* $\gamma \geq 0$ *such that for all* $\alpha^{(k)}$*,*

$$\frac{x^* a^{(k)}(x, \alpha^{(k)})}{|x|} \leq -c_0 |x|^\gamma \qquad \text{for all } x \text{ satisfying } |x| \geq M_0.$$

Then, $[A4]$ *holds with* $[A1]$-(iv)-(v)*.*

(ii) In the case that $d = r = 1$*, under mild regularity conditions,* $\mu_{\theta_*^{(0)}}$ *has a density, that is,* $\mu_{\theta_*^{(0)}}(dx) = v(x, \theta_*^{(0)})dx$*, and* $\partial_x G_f(x)$ *has the following explicit form:*

$$\partial_x G_f(x) \;=\; \frac{2}{B^{(0)}(x, \beta_*^{(0)}) v(x, \theta_*^{(0)})} \int_{-\infty}^{x} f(y) v(y, \theta_*^{(0)}) dy.$$

Set

$$J = \int_{\mathbf{R}^d} (\partial_x G_f(x))^\star B^{(0)}(x, \beta_*^{(0)}) \partial_x G_f(x) \mu_{\theta_*^{(0)}}(dx).$$

Theorem 1 *Assume* $[A1]$, $[A3]$ *and* $[A4]$. *Then,*

$$\sqrt{nh_n}\left(\frac{1}{n}U_n(\mathbf{X}_n^{(0)}, \hat{\theta}_n^{(1)}, \hat{\theta}_n^{(2)}) - \bar{M}\right) \to^d N(0, J)$$

as $nh_n^3 \to 0$.

Remark 3 *In the case that* $nh_n^2 \to 0$, *we assume* $[A3]'$, *which is that* $[A3]$ *holds true as* $nh_n^2 \to 0$.

Under $[A1]$, $[A3]'$ *and* $[A4]$, *Theorem 1 holds true as* $nh_n^2 \to 0$.

In particular, for the estimator $\tilde{\theta}_n^{(k)} = (\hat{\alpha}_n^{(k)}, \tilde{\beta}_n^{(k)})$ *derived in Remark 1, one has that as* $nh_n^2 \to 0$,

$$\sqrt{nh_n}\left(\frac{1}{n}U_n(\mathbf{X}_n^{(0)}, \tilde{\theta}_n^{(1)}, \tilde{\theta}_n^{(2)}) - \bar{M}\right) \to^d N(0, J).$$

Case II: $B^{(1)} = B^{(2)}$

In this section, we consider the situation where the volatility function of Π_1 is the same as the one of Π_2 and the drift function of Π_1 is different from the one of Π_2, that is, $\mu_{\theta_*^{(0)}}(B^{(1)}(x, \beta_*^{(1)}) = B^{(2)}(x, \beta_*^{(2)})) = 1$ and $\mu_{\theta_*^{(0)}}(a^{(1)}(x, \alpha_*^{(1)}) \neq a^{(2)}(x, \alpha_*^{(2)})) > 0$.

Let

$$
\begin{aligned}
N(x) &= (B^{(1)}(x, \beta_*^{(1)}))^{-1}[a^{(0)}(x, \alpha_*^{(0)}), (a^{(1)}(x, \alpha_*^{(1)}) - a^{(2)}(x, \alpha_*^{(2)}))] \\
&\quad -\frac{1}{2}(B^{(1)}(x, \beta_*^{(1)}))^{-1}\left[\left(a^{(1)}(x, \alpha_*^{(1)})\right)^{\otimes 2} - \left(a^{(2)}(x, \alpha_*^{(2)})\right)^{\otimes 2}\right], \\
\bar{N} &= \int_{\mathbf{R}^d} N(x)\mu_{\theta_*^{(0)}}(dx), \\
g(x) &= N(x) - \bar{N}.
\end{aligned}
$$

[$A5$] There exist $G_g(x)$, $\partial_{x_i} G_g(x) \in \mathcal{F}_{\uparrow}(\mathbf{R})$ $(i = 1, \ldots, d)$ such that for all x,

$$L_{\theta_*^{(0)}} G_g(x) = g(x).$$

Set

$$
\begin{aligned}
\tilde{K} &= \int_{\mathbf{R}^d} \Big[(\partial_x G_g(x))^\star B^{(1)}(x,\beta_*^{(1)}) \partial_x G_g(x) \\
&\quad -2(\partial_x G_g(x))^\star (a^{(1)}(x,\alpha_*^{(1)}) - a^{(2)}(x,\alpha_*^{(2)})) \\
&\quad +(a^{(1)}(x,\alpha_*^{(1)}) - a^{(2)}(x,\alpha_*^{(2)}))^\star (B^{(1)}(x,\beta_*^{(1)}))^{-1} (a^{(1)}(x,\alpha_*^{(1)}) - a^{(2)}(x,\alpha_*^{(2)})) \Big] \mu_{\theta_*^{(0)}}(dx), \\
J^{(1)}(\alpha_*^{(1)}) &= \int_{\mathbf{R}^d} (\partial_\alpha a^{(1)}(x,\alpha_*^{(1)}))^\star (B^{(1)}(x,\beta_*^{(1)}))^{-1} (a^{(0)}(x,\alpha_*^{(0)}) - a^{(1)}(x,\alpha_*^{(1)})) \mu_{\theta_*^{(0)}}(dx), \\
J^{(2)}(\alpha_*^{(2)}) &= -\int_{\mathbf{R}^d} (\partial_\alpha a^{(2)}(x,\alpha_*^{(2)}))^\star (B^{(1)}(x,\beta_*^{(1)}))^{-1} (a^{(0)}(x,\alpha_*^{(0)}) - a^{(2)}(x,\alpha_*^{(2)})) \mu_{\theta_*^{(0)}}(dx), \\
K &= (J^{(1)}(\beta_*^{(1)}))^\star (\Gamma_a^{(1)}(\theta_*^{(1)}))^{-1} J^{(1)}(\beta_*^{(1)}) + (J^{(2)}(\beta_*^{(2)}))^\star (\Gamma_a^{(2)}(\theta_*^{(2)}))^{-1} J^{(2)}(\beta_*^{(2)}) + \tilde{K}.
\end{aligned}
$$

Theorem 2 *Assume* $[A1]$, $[A3]$ *and* $[A5]$. *Then,*

$$
\sqrt{nh_n}\left(\frac{1}{nh_n} U_n(\mathbf{X}_n^{(0)}, \hat{\theta}_n^{(1)}, \hat{\theta}_n^{(2)}) - \bar{N}\right) \to^d N(0,K)
$$

as $nh_n^3 \to 0$.

Remark 4 *Under* $[A1]$, $[A3]'$ *and* $[A5]$, *Theorem 2 holds true as* $nh_n^2 \to 0$. *In special, for* $\tilde{\theta}_n^{(k)} = (\hat{\alpha}_n^{(k)}, \tilde{\beta}_n^{(k)})$ *in Remark 1, we obtain that as* $nh_n^2 \to 0$,

$$
\sqrt{nh_n}\left(\frac{1}{nh_n} U_n(\mathbf{X}_n^{(0)}, \tilde{\theta}_n^{(1)}, \tilde{\theta}_n^{(2)}) - \bar{N}\right) \to^d N(0,K).
$$

Non-ergodic case

In this section, we consider d-dimensional stochastic differential equations Π_k $(k = 0, 1, 2)$ defined by

$$\Pi_k : \quad X_t^{(k)} = X_0^{(k)} + \int_0^t a_s^{(k)} ds + \int_0^t b^{(k)}(s, X_s^{(k)}, \theta^{(k)}) dw_s^{(k)}, \quad t \in [0, T], \tag{2}$$

where

$w^{(k)}$ is an r-dimensional standard Wiener process on a stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in[0,T]}, P)$,

$w^{(0)}$, $w^{(1)}$, $w^{(2)}$, $X_0^{(0)}$, $X_0^{(1)}$ and $X_0^{(2)}$ are mutually independent,

$a^{(k)}$ is progressively measurable processes with values in $\mathbf{R}^d$,

$b^{(k)}$ is an $\mathbf{R}^d \otimes \mathbf{R}^r$-valued function defined on $[0, T] \times \mathbf{R}^d \times \Theta^{(k)}$, and

$\Theta^{(k)}$ is a compact convex subset of $\mathbf{R}^{m_k}$. $\theta_*^{(k)}$ is the true value of $\theta^{(k)}$ and we assume that $\theta_*^{(k)} \in \mathrm{Int}(\Theta^{(k)})$.

For $k = 1, 2$, the training data are discrete observations $\mathbf{X}_n^{(k)} = (X_{t_i^n}^{(k)})_{0\le i\le n}$ obtained from the model Π_k, where $t_i^n = ih_n$ and $h = h_n = T/n$.

We assume that the data $\mathbf{X}_n^{(0)} = (X_{t_i^n}^{(0)})_{0\le i\le n}$ are obtained from either Π_1 or Π_2, which means that the distribution of model Π_0 is the same as that of one of the two models Π_1 and Π_2.

The asymptotics will be considered under $n \to \infty$.

Set $B^{(k)}(t, x, \theta^{(k)}) = (b^{(k)}(t, x, \theta^{(k)}))^{\otimes 2}$.

We denote by $\to^{d_s(\mathcal{F}_T)}$ the $\mathcal{F}_T$-stable convergence in distribution.

In this section, we consider the case where the volatility function of Π_1 is different from the one of Π_2, that is,

$P_{\theta_*^{(0)}}(B^{(1)}(t, X_t^{(0)}, \theta_*^{(1)}) = B^{(2)}(t, X_t^{(0)}, \theta_*^{(2)}) \text{ for all } t \in [0, T]\ ) = 0.$

We make the following assumption.

[B1] Let $k = 1, 2$.
(i) Equation (2) admits a non-exploding strong solution on $[0, T]$.
(ii) $t \to a_t^{(k)}$ is continuous.
(iii) The partial derivatives $\partial_t^{k_1} \partial_x^{k_2} \partial_{\theta^{(k)}}^{k_3} b^{(k)}$ exist and are continuous on $[0, T] \times \mathbf{R}^d \times \Theta^{(k)}$ for $k_1 = 0, 1$ and $k_2, k_3 = 0, 1, 2$, and $\inf_{t,x,\theta^{(k)}} \det B^{(k)}(t, x, \theta^{(k)}) > 0$.

Let $k = 1, 2$. Let $\widehat{\theta}_n^{(k)} := \widehat{\theta}^{(k)}(\mathbf{X}_n^{(k)})$ be an estimator for an unknown parameter $\theta^{(k)}$ of the model Π_k based on the data $\mathbf{X}_n^{(k)}$.

Set $\tilde{B}^{(k)}_{i-1}(\theta^{(k)}) = B^{(k)}(t^n_{i-1}, X^{(0)}_{t^n_{i-1}}, \theta^{(k)})$ and

$$\tilde{u}^{(k)}_n(\mathbf{X}^{(0)}_n, \theta^{(k)}) \;=\; -\frac{1}{2}\sum_{i=1}^n \left\{h_n^{-1}(\tilde{B}^{(k)}_{i-1}(\theta^{(k)}))^{-1}[(\Delta X^{(0)}_i)^{\otimes 2}] + \log\det(\tilde{B}^{(k)}_{i-1}(\theta^{(k)}))\right\}.$$

The following quadratic discriminant function for the non-ergodic case is used.

$$\begin{aligned} &\mathcal{U}_n(\mathbf{X}^{(0)}_n, \widehat{\theta}^{(1)}_n, \widehat{\theta}^{(2)}_n) = \tilde{u}^{(1)}_n(\mathbf{X}^{(0)}_n, \widehat{\theta}^{(1)}_n) - \tilde{u}^{(2)}_n(\mathbf{X}^{(0)}_n, \widehat{\theta}^{(2)}_n) \\ =\;& -\frac{1}{2}\sum_{i=1}^n \left\{\log\frac{\det B^{(1)}_{i-1}(\widehat{\theta}^{(1)}_n)}{\det B^{(2)}_{i-1}(\widehat{\theta}^{(2)}_n)} + h_n^{-1}\left\{(B^{(1)}_{i-1}(\widehat{\theta}^{(1)}_n))^{-1} - (B^{(2)}_{i-1}(\widehat{\theta}^{(2)}_n)^{-1}\right\}[(\Delta X^{(0)}_i)^{\otimes 2}]\right\}. \end{aligned}$$

We propose a discriminant rule such that

$\mathbf{X}^{(0)}_n$ is classified into Π_1 if $\mathcal{U}_n(\mathbf{X}^{(0)}_n, \widehat{\theta}^{(1)}_n, \widehat{\theta}^{(2)}_n) \geq 0$ and otherwise into Π_2.

Let two misclassification probabilities denote

$$\begin{aligned} q_n(1|2) &= P(\mathcal{U}_n(\mathbf{X}^{(0)}_n, \widehat{\theta}^{(1)}_n, \widehat{\theta}^{(2)}_n) < 0 \mid \Pi_1), \\ q_n(2|1) &= P(\mathcal{U}_n(\mathbf{X}^{(0)}_n, \widehat{\theta}^{(1)}_n, \widehat{\theta}^{(2)}_n) \geq 0 \mid \Pi_2). \end{aligned}$$

We make the following assumption.

$[B2]$ As $n \to \infty$, $\widehat{\theta}_n^{(k)} \to^p \theta_*^{(k)}$.

Proposition 2 *Assume* $[B1]$ *and* $[B2]$. *Then, as* $n \to \infty$,

$$q_n(1|2) + q_n(2|1) \to 0.$$

Set

$$\begin{aligned}
\gamma^{(k)}(t, \mathbf{X}_t^{(k)}, \theta_*^{(k)})_{ij} &= tr\{(B^{(k)})^{-1}(\partial_{\theta_i} B^{(k)})(B^{(k)})^{-1}(\partial_{\theta_j} B^{(k)})(t, X_t^{(k)}, \theta_*^{(k)})\}, \\
\Gamma^{(k)}(\theta_*^{(k)})_{ij} &= \frac{1}{2T}\int_0^T \gamma^{(k)}(t, \mathbf{X}_t^{(k)}, \theta_*^{(k)})_{ij} dt
\end{aligned}$$

for $i, j = 1, \ldots, m_k$. Here we suppose that $\Gamma^{(k)}(\theta_*^{(k)})$ is $P_{\theta_*^{(k)}}$-a.s. invertible. Moreover, we make the assumption as follows.

$[B3]$ $\sqrt{n}(\widehat{\theta}_n^{(k)} - \theta_*^{(k)}) \to^{d_s(\mathcal{F}_T)} (\Gamma^{(k)}(\theta_*^{(k)}))^{-1/2}\zeta^{(k)}$ as $n \to \infty$, where $\zeta^{(k)}$ is an m_k-dimensional standard normal random variable independent of $\Gamma^{(k)}(\theta_*^{(k)})$.

Remark 5 *For an estimator* $\widehat{\theta}_n^{(k)}$ *satisfying* $[B3]$, *we can refer Genon-Catalot and Jacod (1993) and Uchida and Yoshida (2011b).*

Let for $j = 1, \ldots, m_k$ and $k = 1, 2$,

$$\begin{aligned}
\mathcal{U}(t,x) &= -\frac{1}{2}\left\{\log\frac{\det B^{(1)}(t,x,\theta_*^{(1)})}{\det B^{(2)}(t,x,\theta_*^{(2)})}\right. \\
&\quad \left. + \left\{(B^{(1)}(t,x,\theta_*^{(1)}))^{-1} - (B^{(2)}(t,x,\theta_*^{(2)}))^{-1}\right\}[B^{(0)}(t,x,\theta_*^{(0)})]\right\}, \\
\bar{\mathcal{U}} &= \frac{1}{T}\int_0^T \mathcal{U}(t, X_t^{(0)})dt, \\
K^{(k)}(t,x,\theta_*^{(k)})_j &= -\frac{1}{2}\left\{tr\left((B^{(k)})^{-1}\partial_{\theta_j}B^{(k)}(t,x,\theta_*^{(k)})\right) + (\partial_{\theta_j}(B^{(k)})^{-1}(t,x,\theta_*^{(k)}))B^{(0)}(t,x,\theta_*^{(0)})\right\}, \\
\bar{K}_j^{(k)} &= \frac{1}{T}\int_0^T K^{(k)}(t, X_t^{(0)}, \theta_*^{(k)})_j dt, \\
J^{(0)}(t,x) &= \frac{1}{2}tr\left[\left(\left\{(B^{(1)}(t,x,\theta_*^{(1)}))^{-1} - (B^{(2)}(t,x,\theta_*^{(2)}))^{-1}\right\}B^{(0)}(t,x,\theta_*^{(0)})\right)^2\right], \\
\bar{J}^{(0)} &= \frac{1}{T}\int_0^T J^{(0)}(t, X_t^{(0)})dt.
\end{aligned}$$

Let

$$\Gamma = \left((\bar{K}^{(1)})^{\star}(\Gamma^{(1)}(\theta_*^{(1)}))^{-1/2}, -(\bar{K}^{(2)})^{\star}(\Gamma^{(2)}(\theta_*^{(2)}))^{-1/2}, (\bar{J}^{(0)})^{1/2}\right).$$

Theorem 3 *Assume* $[B1]$ *and* $[B3]$. *Then, as* $n \to \infty$,

$$\sqrt{n}\left(\frac{1}{n}\mathcal{U}_n(\mathbf{X}_n^{(0)}, \hat{\theta}_n^{(1)}, \hat{\theta}_n^{(2)}) - \bar{\mathcal{U}}\right) \longrightarrow^{d_s(\mathcal{F}_T)} \Gamma\zeta,$$

where ζ *is an* $(m_1 + m_2 + 1)$*-dimensional standard normal random variable independent of* Γ.

Examples and simulation studies

Case I with different parametric models

Consider the one-dimensional diffusion processes defined by

$$\begin{aligned}\Pi_1: \quad dX_t^{(1)} &= \left(\alpha_1^{(1)} - \alpha_2^{(1)} X_t^{(1)}\right) dt + \beta^{(1)} dw_t^{(1)}, \quad X_0^{(1)} = 10, \\ \Pi_2: \quad dX_t^{(2)} &= (\alpha_1^{(2)} - \alpha_2^{(2)} X_t^{(2)} + \alpha_3^{(2)} \sin(X_t^{(2)}))dt + \frac{\beta_1^{(2)} + \beta_2^{(2)} (X_t^{(2)})^2}{1 + (X_t^{(2)})^2} dw_t^{(2)}, \ X_0^{(2)} = 10,\end{aligned}$$

where $\alpha^{(1)} = (\alpha_1^{(1)}, \alpha_2^{(1)})$, $\alpha^{(2)} = (\alpha_1^{(2)}, \alpha_2^{(2)}, \alpha_2^{(3)})$, $\beta^{(1)}$ and $\beta^{(2)} = (\beta_1^{(2)}, \beta_2^{(2)})$ are unknown parameters, and the true parameter values are $\alpha_*^{(1)} = (1, 2)$, $\alpha_*^{(2)} = (1.5, 2.5, 6)$, $\beta_*^{(1)} = 4$ and $\beta_*^{(2)} = (4, 3)$.

The simulations were done for each $(h_n, T) = (1/250, 20)$ and $(1/390, 250)$.

1000 independent sample paths $\mathbf{X}_n^{(1)}$ and $\mathbf{X}_n^{(2)}$ are generated from the models Π_1 and Π_2, respectively.

Moreover, 1000 independent sample paths $\mathbf{X}_n^{(0)}$ are generated from the model Π_1.

For each model Π_k, we use the type III adaptive estimators in Remark 1, that is, $\tilde{\theta}_n^{(k)} = (\hat{\alpha}_n^{(k)}, \tilde{\beta}_n^{(k)})$ for $nh_n^2 \to 0$, and $\hat{\theta}_n^{(k)} = (\hat{\alpha}_n^{(k)}, \hat{\beta}_n^{(k)})$ for $nh_n^3 \to 0$. Furthermore, the discriminant functions are $U_{3,n} := U_n(\mathbf{X}_n^{(0)}, \hat{\theta}_n^{(1)}, \hat{\theta}_n^{(2)})$ for $nh_n^3 \to 0$, and $U_{2,n} := U_n(\mathbf{X}_n^{(0)}, \tilde{\theta}_n^{(1)}, \tilde{\theta}_n^{(2)})$ for $nh_n^2 \to 0$ in Remark 3.

$$\Pi_1: \quad dX_t^{(1)} = \left(\alpha_1^{(1)} - \alpha_2^{(1)} X_t^{(1)}\right) dt + \beta^{(1)} dw_t^{(1)}, \quad X_0^{(1)} = 10,$$

$$\Pi_2: \quad dX_t^{(2)} = (\alpha_1^{(2)} - \alpha_2^{(2)} X_t^{(2)} + \alpha_3^{(2)} \sin(X_t^{(2)}))dt + \frac{\beta_1^{(2)} + \beta_2^{(2)}(X_t^{(2)})^2}{1 + (X_t^{(2)})^2} dw_t^{(2)}, \ X_0^{(2)} = 10,$$

Table 1. The number of selecting Π_1 for 1000 independent simulated sample paths with $\alpha_*^{(1)} = (1, 2)$, $\alpha_*^{(2)} = (1.5, 2.5, 6)$, $\beta_*^{(1)} = 4$ and $\beta_*^{(2)} = (4, 3)$.

h_n	T	$U_{2,n} \geq 0$	$U_{3,n} \geq 0$
1/250	20	1000	1000
1/390	250	1000	1000

Table 2. The mean and s.d. of estimators for 1000 independent simulated sample paths with $\alpha_*^{(1)} = (1, 2)$, $\alpha_*^{(2)} = (1.5, 2.5, 6)$, $\beta_*^{(1)} = 4$ and $\beta_*^{(2)} = (4, 3)$.

h_n	T	$\tilde{\beta}^{(1)}$	$\tilde{\beta}_1^{(2)}$	$\tilde{\beta}_2^{(2)}$	$\hat{\beta}^{(1)}$	$\hat{\beta}_1^{(2)}$	$\hat{\beta}_2^{(2)}$
1/250	20	3.99(0.038)	4.008(0.12)	2.98(0.05)	4.002(0.039)	4.00(0.13)	3.005(0.05)
1/390	250	3.99(0.009)	4.012(0.02)	2.98(0.01)	4.001(0.009)	3.99(0.02)	3.002(0.01)

h_n	T	$\hat{\alpha}_1^{(1)}$	$\hat{\alpha}_2^{(1)}$	$\hat{\alpha}_1^{(2)}$	$\hat{\alpha}_2^{(2)}$	$\hat{\alpha}_3^{(2)}$
1/250	20	1.052(1.001)	2.143(0.418)	1.901(1.379)	2.679(0.464)	5.860(1.228)
1/390	250	1.006(0.266)	2.010(0.125)	1.510(0.237)	2.494(0.119)	5.925(0.355)

Although the drift estimator $\hat{\alpha}^{(2)}$ has a bias when $T = 20$, the discriminant function works well. Also the discriminant function has a good behavior when $T = 250$.

$$\Pi_1: \quad dX_t^{(1)} = \left(\alpha_1^{(1)} - \alpha_2^{(1)} X_t^{(1)}\right) dt + \beta^{(1)} dw_t^{(1)}, \quad X_0^{(1)} = 10,$$

$$\Pi_2: \quad dX_t^{(2)} = (\alpha_1^{(2)} - \alpha_2^{(2)} X_t^{(2)} + \alpha_3^{(2)} \sin(X_t^{(2)}))dt + \frac{\beta_1^{(2)} + \beta_2^{(2)} (X_t^{(2)})^2}{1 + (X_t^{(2)})^2} dw_t^{(2)}, \ X_0^{(2)} = 10,$$

Furthermore, we consider the simulation study in the non-ergodic case for each

$(h_n, T) = (1/250, 1)$, $(1/390, 1)$ and $(1/1000, 1)$

with 1000 independent sample paths $\mathbf{X}_n^{(0)}$, $\mathbf{X}_n^{(1)}$ and $\mathbf{X}_n^{(2)}$ generated in the same way as the previous example.

For each model Π_k, we use the initial estimator $\tilde{\beta}_n^{(k)}$ in Remark 1, which is

the maximum likelihood type estimator and satisfies [B3], see Genon-Catalot and Jacod (1993) and Uchida and Yoshida (2011b).

The number of selecting the model Π_1 by using the discriminant functions $\mathcal{U}_n = \mathcal{U}_n(\mathbf{X}_n^{(0)}, \tilde{\beta}_n^{(1)}, \tilde{\beta}_n^{(2)})$ in Section 3, the mean and the s.d. for the estimators are computed and shown in Table 3 below.

Table 3. The number of selecting Π_1 and the mean and s.d. of estimators for 1000 independent simulated sample paths with $\alpha_*^{(1)} = (1, 2)$, $\alpha_*^{(2)} = (1.5, 2.5, 6)$, $\beta_*^{(1)} = 4$ and $\beta_*^{(2)} = (4, 3)$.

h_n	T	$\mathcal{U}_n \geq 0$	$\tilde{\beta}^{(1)}$	$\tilde{\beta}_1^{(2)}$	$\tilde{\beta}_2^{(2)}$
1/250	1	996	4.029(0.173)	3.265(2.057)	3.136(0.205)
1/390	1	999	4.022(0.134)	3.487(1.662)	3.097(0.174)
1/1000	1	1000	4.008(0.089)	3.773(1.043)	3.040(0.109)

In all cases, the volatility estimators $\hat{\beta}^{(2)}$ are biased, but the discriminant functions work well.

Case I for specified parametric models

Next, we treat the following one-dimensional diffusion processes. For $k = 1, 2$,

$$\Pi_k : \quad dX_t^{(k)} \quad = \quad (\alpha_1^{(k)} - \alpha_2^{(k)} X_t^{(k)} + \alpha_3^{(k)} \sin(X_t^{(k)}))dt + \frac{\beta_1^{(k)} + \beta_2^{(k)}(X_t^{(k)})^2}{1 + (X_t^{(k)})^2} dw_t^{(k)}, \ X_0^{(1)} = 10,$$

where $\alpha^{(1)} = (\alpha_1^{(1)}, \alpha_2^{(1)}, \alpha_3^{(1)})$, $\alpha^{(2)} = (\alpha_1^{(2)}, \alpha_2^{(2)}, \alpha_2^{(3)})$, $\beta^{(1)} = (\beta_1^{(1)}, \beta_2^{(1)})$ and $\beta^{(2)} = (\beta_1^{(2)}, \beta_2^{(2)})$ are unknown parameters, and the true parameter values are $\alpha_*^{(1)} = \alpha_*^{(2)} = (1.5, 2.5, 6)$, $\beta_*^{(1)} = (3.9, 2.9)$ and $\beta_*^{(2)} = (4, 3)$.

In the same way as in Section 4.1, the simulations were done for each

$(h_n, T) = (1/250, 20)$ and $(1/390, 250)$ with

1000 independent sample paths $\mathbf{X}_n^{(0)}$, $\mathbf{X}_n^{(1)}$ and $\mathbf{X}_n^{(2)}$ generated from the models Π_1, Π_1 and Π_2, respectively.

The type III adaptive estimators in Remark 1, $\tilde{\theta}_n^{(k)} = (\hat{\alpha}_n^{(k)}, \tilde{\beta}_n^{(k)})$ for $nh_n^2 \to 0$ and $\hat{\theta}_n^{(k)} = (\hat{\alpha}_n^{(k)}, \hat{\beta}_n^{(k)})$ for $nh_n^3 \to 0$, are used and the discriminant functions are $U_{3,n} := U_n(\mathbf{X}_n^{(0)}, \hat{\theta}_n^{(1)}, \hat{\theta}_n^{(2)})$ for $nh_n^3 \to 0$, and $U_{2,n} := U_n(\mathbf{X}_n^{(0)}, \tilde{\theta}_n^{(1)}, \tilde{\theta}_n^{(2)})$ for $nh_n^2 \to 0$ in Remark 3.

For $k = 1, 2$,

$$\Pi_k: \quad dX_t^{(k)} = (\alpha_1^{(k)} - \alpha_2^{(k)} X_t^{(k)} + \alpha_3^{(k)} \sin(X_t^{(k)}))dt + \frac{\beta_1^{(k)} + \beta_2^{(k)} (X_t^{(k)})^2}{1 + (X_t^{(k)})^2} dw_t^{(k)}, \ X_0^{(1)} = 10.$$

Table 4. The number of selecting Π_1 for 1000 independent simulated sample paths with $\alpha_*^{(1)} = \alpha_*^{(2)} = (1.5, 2.5, 6)$, $\beta_*^{(1)} = (3.9, 2.9)$ and $\beta_*^{(2)} = (4, 3)$.

h_n	T	$U_{2,n} \geq 0$	$U_{3,n} \geq 0$
1/250	20	841	875
1/390	250	1000	1000

Table 5. The mean and s.d. of estimators for 1000 independent simulated sample paths with $\alpha_*^{(1)} = \alpha_*^{(2)} = (1.5, 2.5, 6)$, $\beta_*^{(1)} = (3.9, 2.9)$ and $\beta_*^{(2)} = (4, 3)$.

h_n	T	$\tilde{\beta}_1^{(1)}$	$\tilde{\beta}_2^{(1)}$	$\hat{\beta}_1^{(1)}$	$\hat{\beta}_2^{(1)}$
1/250	20	3.907(0.152)	2.889(0.051)	3.897(0.125)	2.905(0.050)
1/390	250	3.911(0.027)	2.887(0.010)	3.897(0.027)	2.903(0.011)

h_n	T	$\tilde{\beta}_1^{(2)}$	$\tilde{\beta}_2^{(2)}$	$\hat{\beta}_1^{(2)}$	$\hat{\beta}_2^{(2)}$
1/250	20	4.008(0.127)	2.989(0.052)	4.000(0.131)	3.005(0.054)
1/390	250	4.012(0.027)	2.986(0.011)	3.998(0.027)	3.002(0.011)

h_n	T	$\hat{\alpha}_1^{(1)}$	$\hat{\alpha}_2^{(1)}$	$\hat{\alpha}_3^{(1)}$	$\hat{\alpha}_1^{(2)}$	$\hat{\alpha}_2^{(2)}$	$\hat{\alpha}_3^{(2)}$
1/250	20	1.79(1.23)	2.64(0.44)	5.84(1.23)	1.90(1.37)	2.67(0.46)	5.86(1.22)
1/390	250	1.51(0.24)	2.49(0.11)	5.91(0.35)	1.51(0.23)	2.49(0.11)	5.92(0.35)

When $T = 20$, the drift estimators $\hat{\alpha}^{(1)}$ and $\hat{\alpha}^{(2)}$ have biases, and the discriminant function has misclassification. On the other hand, when $T = 250$, the discriminant function works well.

For $k=1,2$,

$$\Pi_k: \quad dX_t^{(k)} \quad = \quad (\alpha_1^{(k)} - \alpha_2^{(k)} X_t^{(k)} + \alpha_3^{(k)} \sin(X_t^{(k)}))dt + \frac{\beta_1^{(k)} + \beta_2^{(k)} (X_t^{(k)})^2}{1+(X_t^{(k)})^2} dw_t^{(k)}, \ X_0^{(1)} = 10.$$

Moreover, we investigate the simulation study in the non-ergodic cases for each $(h_n, T) = (1/250, 1)$, $(1/390, 1)$ and $(1/1000, 1)$ with 1000 independent sample paths $\mathbf{X}_n^{(0)}$, $\mathbf{X}_n^{(1)}$ and $\mathbf{X}_n^{(2)}$ generated as in the previous subsection.

Table 6. The number of selecting Π_1 and the mean and s.d. of estimators for 1000 independent simulated sample paths with $\alpha_*^{(1)} = \alpha_*^{(2)} = (1.5, 2.5, 6)$, $\beta_*^{(1)} = (3.5, 2.5)$ and $\beta_*^{(2)} = (4, 3)$.

h_n	T	$\mathcal{U}_n \geq 0$	$\tilde{\beta}_1^{(1)}$	$\tilde{\beta}_2^{(1)}$	$\tilde{\beta}_1^{(2)}$	$\tilde{\beta}_2^{(2)}$
1/250	1	834	2.479(1.969)	2.677(0.190)	3.265(2.057)	3.136(0.205)
1/390	1	902	2.817(1.452)	2.619(0.150)	3.487(1.662)	3.097(0.174)
1/1000	1	968	3.233(0.816)	2.547(0.091)	3.773(1.043)	3.040(0.109)

Table 7. The number of selecting Π_1 and the mean and s.d. of estimators for 1000 independent simulated sample paths with $\alpha_*^{(1)} = (2.5, 3.5, 7)$, $\alpha_*^{(2)} = (1.5, 2.5, 6)$, $\beta_*^{(1)} = (3.5, 2.5)$ and $\beta_*^{(2)} = (4, 3)$.

h_n	T	$\mathcal{U}_n \geq 0$	$\tilde{\beta}_1^{(1)}$	$\tilde{\beta}_2^{(1)}$	$\tilde{\beta}_1^{(2)}$	$\tilde{\beta}_2^{(2)}$
1/250	1	805	2.294(1.496)	2.766(0.195)	3.265(2.057)	3.136(0.205)
1/390	1	877	2.697(1.194)	2.678(0.158)	3.487(1.662)	3.097(0.174)
1/1000	1	965	3.181(0.655)	2.572(0.097)	3.773(1.043)	3.040(0.109)

We can see from Tables 6 and 7 that the drift term does not strongly depend on the classification

Table 8. The number of selecting Π_1 and the mean and s.d. of estimators for 1000 independent simulated sample paths with $\alpha_*^{(1)} = (1.5, 2.5, 5)$, $\alpha_*^{(2)} = (1.5, 2.5, 6)$, $\beta_*^{(1)} = (3.5, 2.5)$ and $\beta_*^{(2)} = (4, 3)$.

h_n	T	$\mathcal{U}_n \geq 0$	$\tilde{\beta}_1^{(1)}$	$\tilde{\beta}_2^{(1)}$	$\tilde{\beta}_1^{(2)}$	$\tilde{\beta}_2^{(2)}$
1/250	1	820	2.529(1.857)	2.677(0.188)	3.265(2.057)	3.136(0.205)
1/390	1	900	2.857(1.409)	2.618(0.148)	3.487(1.662)	3.097(0.174)
1/1000	1	966	3.250(0.785)	2.547(0.090)	3.773(1.043)	3.040(0.109)

Case II for different parametric models

We consider the one-dimensional diffusion processes defined by

$$\begin{aligned} \Pi_1: \quad dX_t^{(1)} &= (\alpha_1^{(1)} - \alpha_2^{(1)} X_t^{(1)})dt + \frac{\beta_1^{(1)} + \beta_2^{(1)}(X_t^{(1)})^2}{1+(X_t^{(1)})^2} dw_t^{(1)}, \quad X_0^{(1)} = 10, \\ \Pi_2: \quad dX_t^{(2)} &= (\alpha_1^{(2)} - \alpha_2^{(2)} X_t^{(2)} + \alpha_3^{(2)} \sin(X_t^{(2)}))dt + \frac{\beta_1^{(2)} + \beta_2^{(2)}(X_t^{(2)})^2}{1+(X_t^{(2)})^2} dw_t^{(2)}, \ X_0^{(2)} = 10, \end{aligned}$$

where $\alpha^{(1)} = (\alpha_1^{(1)}, \alpha_2^{(1)})$, $\alpha^{(2)} = (\alpha_1^{(2)}, \alpha_2^{(2)}, \alpha_3^{(2)})$, $\beta^{(1)} = (\beta_1^{(1)}, \beta_2^{(1)})$ and $\beta^{(2)} = (\beta_1^{(2)}, \beta_2^{(2)})$ are unknown parameters, and the true parameter values are $\alpha_*^{(1)} = (1, 2)$, $\alpha_*^{(2)} = (1.5, 2.5, 6)$ and $\beta_*^{(1)} = \beta_*^{(2)} = (3, 4)$.

The simulations were done for each $(h_n, T) = (1/250, 20)$ and $(1/390, 250)$.

1000 independent sample paths $\mathbf{X}_n^{(1)}$, $\mathbf{X}_n^{(2)}$ and $\mathbf{X}_n^{(0)}$ are generated from the models Π_1, Π_2 and Π_1, respectively.

For each model Π_k, the type III adaptive estimators in Remark 1, $\tilde{\theta}_n^{(k)} = (\hat{\alpha}_n^{(k)}, \tilde{\beta}_n^{(k)})$ for $nh_n^2 \to 0$, and $\hat{\theta}_n^{(k)} = (\hat{\alpha}_n^{(k)}, \hat{\beta}_n^{(k)})$ for $nh_n^3 \to 0$, are used and the discriminant functions are $U_{2,n} := U_n(\mathbf{X}_n^{(0)}, \tilde{\theta}_n^{(1)}, \tilde{\theta}_n^{(2)})$ for $nh_n^2 \to 0$ in Remark 4, and $U_{3,n} := U_n(\mathbf{X}_n^{(0)}, \hat{\theta}_n^{(1)}, \hat{\theta}_n^{(2)})$ for $nh_n^3 \to 0$.

We consider the one-dimensional diffusion processes defined by

$$\Pi_1: \quad dX_t^{(1)} = (\alpha_1^{(1)} - \alpha_2^{(1)} X_t^{(1)})dt + \frac{\beta_1^{(1)} + \beta_2^{(1)}(X_t^{(1)})^2}{1+(X_t^{(1)})^2} dw_t^{(1)}, \quad X_0^{(1)} = 10,$$

$$\Pi_2: \quad dX_t^{(2)} = (\alpha_1^{(2)} - \alpha_2^{(2)} X_t^{(2)} + \alpha_3^{(2)} \sin(X_t^{(2)}))dt + \frac{\beta_1^{(2)} + \beta_2^{(2)}(X_t^{(2)})^2}{1+(X_t^{(2)})^2} dw_t^{(2)}, \; X_0^{(2)} = 10.$$

Table 9. The number of selecting Π_1 for 1000 independent simulated sample paths with $\alpha_*^{(1)} = (1,2)$, $\alpha_*^{(2)} = (1.5, 2.5, 6)$ and $\beta_*^{(1)} = \beta_*^{(2)} = (3,4)$.

h_n	T	$U_{2,n} \geq 0$	$U_{3,n} \geq 0$
1/250	20	945	941
1/390	250	1000	1000

Table 10. The mean and s.d. of estimators for 1000 independent simulated sample paths with $\alpha_*^{(1)} = (1,2)$, $\alpha_*^{(2)} = (1.5, 2.5, 6)$ and $\beta_*^{(1)} = \beta_*^{(2)} = (3,4)$.

h_n	T	$\tilde{\beta}_1^{(1)}$	$\tilde{\beta}_2^{(1)}$	$\hat{\beta}_1^{(1)}$	$\hat{\beta}_2^{(1)}$
1/250	20	2.995(0.059)	3.999(0.065)	3.002(0.061)	4.002(0.066)
1/390	250	2.997(0.012)	3.996(0.015)	2.999(0.013)	4.002(0.015)

h_n	T	$\tilde{\beta}_1^{(2)}$	$\tilde{\beta}_2^{(2)}$	$\hat{\beta}_1^{(2)}$	$\hat{\beta}_2^{(2)}$
1/250	20	3.036(0.087)	3.973(0.060)	2.998(0.088)	4.007(0.061)
1/390	250	3.024(0.019)	3.978(0.013)	2.997(0.019)	4.003(0.013)

h_n	T	$\hat{\alpha}_1^{(1)}$	$\hat{\alpha}_2^{(1)}$	$\hat{\alpha}_1^{(2)}$	$\hat{\alpha}_2^{(2)}$	$\hat{\alpha}_3^{(2)}$
1/250	20	1.038(0.855)	2.145(0.439)	1.804(1.260)	2.685(0.480)	5.876(1.328)
1/390	250	1.004(0.229)	2.011(0.137)	1.511(0.253)	2.496(0.133)	5.923(0.379)

When $T = 20$, the drift estimator of model Π_2, $\hat{\alpha}^{(2)}$, is biased and the discriminant function has some misclassification. On the other hand, when $T = 250$, the discriminant function works very well.

Case II for specified parametric models

Next, we investigate the one-dimensional diffusion processes as follows. For $k = 1, 2$,

$$\Pi_k : \quad dX_t^{(k)} = (\alpha_1^{(k)} - \alpha_2^{(k)} X_t^{(k)} + \alpha_3^{(k)} \sin(X_t^{(k)}))dt + \frac{\beta_1^{(k)} + \beta_2^{(k)}(X_t^{(k)})^2}{1 + (X_t^{(k)})^2} dw_t^{(k)}, \ X_0^{(k)} = 10,$$

where $\alpha^{(1)} = (\alpha_1^{(1)}, \alpha_2^{(1)}, \alpha_3^{(1)})$, $\alpha^{(2)} = (\alpha_1^{(2)}, \alpha_2^{(2)}, \alpha_3^{(2)})$, $\beta^{(1)} = (\beta_1^{(1)}, \beta_2^{(1)})$ and $\beta^{(2)} = (\beta_1^{(2)}, \beta_2^{(2)})$ are unknown parameters, and the true parameter values are $\alpha_*^{(1)} = (0.5, 1.5, 5)$, $\alpha_*^{(2)} = (1.5, 2.5, 6)$ and $\beta_*^{(1)} = \beta_*^{(2)} = (4, 3)$.

In the same way as in the previous example, the simulations were done for each

$(h_n, T) = (1/250, 20)$ and $(1/390, 250)$ with

1000 independent sample paths $\mathbf{X}_n^{(1)}$, $\mathbf{X}_n^{(2)}$ and $\mathbf{X}_n^{(0)}$ generated from the models Π_1, Π_2 and Π_1, respectively.

We use the type III adaptive estimators in Remark 1, $\tilde{\theta}_n^{(k)} = (\hat{\alpha}_n^{(k)}, \tilde{\beta}_n^{(k)})$ for $nh_n^2 \to 0$, and $\hat{\theta}_n^{(k)} = (\hat{\alpha}_n^{(k)}, \hat{\beta}_n^{(k)})$ for $nh_n^3 \to 0$. Moreover, the discriminant functions are $U_{2,n} := U_n(\mathbf{X}_n^{(0)}, \tilde{\theta}_n^{(1)}, \tilde{\theta}_n^{(2)})$ for $nh_n^2 \to 0$ in Remark 4, and $U_{3,n} := U_n(\mathbf{X}_n^{(0)}, \hat{\theta}_n^{(1)}, \hat{\theta}_n^{(2)})$ for $nh_n^3 \to 0$.

For $k = 1, 2$,

$$\Pi_k: \quad dX_t^{(k)} = (\alpha_1^{(k)} - \alpha_2^{(k)} X_t^{(k)} + \alpha_3^{(k)} \sin(X_t^{(k)}))dt + \frac{\beta_1^{(k)} + \beta_2^{(k)} (X_t^{(k)})^2}{1 + (X_t^{(k)})^2} dw_t^{(k)}, \ X_0^{(k)} = 10.$$

Table 11. The number of selecting Π_1 for 1000 independent simulated sample paths with $\alpha_*^{(1)} = (0.5, 1.5, 5)$, $\alpha_*^{(2)} = (1.5, 2.5, 6)$ and $\beta_*^{(1)} = \beta_*^{(2)} = (4, 3)$.

h_n	T	$U_{2,n} \geq 0$	$U_{3,n} \geq 0$
1/250	20	830	829
1/390	250	1000	1000

Table 12. The mean and s.d. of estimators for 1000 independent simulated sample paths with $\alpha_*^{(1)} = (0.5, 1.5, 5)$, $\alpha_*^{(2)} = (1.5, 2.5, 6)$ and $\beta_*^{(1)} = \beta_*^{(2)} = (4, 3)$.

h_n	T	$\tilde{\beta}_1^{(1)}$	$\tilde{\beta}_2^{(1)}$	$\hat{\beta}_1^{(1)}$	$\hat{\beta}_2^{(1)}$
1/250	20	4.011(0.129)	2.989(0.047)	3.999(0.133)	3.003(0.049)
1/390	250	4.011(0.028)	2.989(0.010)	3.998(0.029)	3.002(0.010)

h_n	T	$\tilde{\beta}_1^{(2)}$	$\tilde{\beta}_2^{(2)}$	$\hat{\beta}_1^{(2)}$	$\hat{\beta}_2^{(2)}$
1/250	20	4.008(0.127)	2.989(0.052)	4.000(0.131)	3.005(0.054)
1/390	250	4.012(0.027)	2.986(0.011)	3.998(0.027)	3.002(0.011)

h_n	T	$\hat{\alpha}_1^{(1)}$	$\hat{\alpha}_2^{(1)}$	$\hat{\alpha}_3^{(1)}$	$\hat{\alpha}_1^{(2)}$	$\hat{\alpha}_2^{(2)}$	$\hat{\alpha}_3^{(2)}$
1/250	20	0.648(1.01)	1.634(0.32)	4.978(1.13)	1.901(1.37)	2.679(0.46)	5.860(1.22)
1/390	250	0.507(0.21)	1.505(0.08)	4.960(0.32)	1.510(0.23)	2.494(0.11)	5.925(0.35)

Similarly as the previous example, when $T = 20$, the discriminant function has some misclassification because the drift estimators $\hat{\alpha}^{(1)}$ and $\hat{\alpha}^{(2)}$ have considerable biases. On the other hand, when $T = 250$, the discriminant function works very well.