Adaptive estimation for small diffusion processes

[1]Ryosuke Nomura, [2]Masayuki Uchida
[1]Graduate School of Mathematical Science, University of Tokyo and JST, CREST ,
[2]Graduate School of Engineering Science, Osaka University, MMDS and JST, CREST

Dynstoch 2016
at University Rennes 2, 9th June, 2016.

Plan of today's talk

1. Adaptive ML type estimation for small diffusion processes in the case when $\frac{1}{\epsilon\sqrt{n}} = O(1)$.

2. Adaptive Bayes type estimation and hybrid type estimation for small diffusion processes.

1. Adaptive ML type estimation for small diffusion processes

We consider a d-dimensional diffusion process defined by the stochastic differential equations

$$\begin{cases} dX_t = a(X_t, \alpha)dt + \epsilon b(X_t, \beta)dw_t, \quad t \in [0, T], \quad \epsilon \in (0, 1], \\ X_0 = x_0, \end{cases} \tag{1}$$

where

w is an r-dimensional standard Wiener process,

x_0 is a deterministic initial condition and

$\theta = (\alpha, \beta) \in \Theta_\alpha \times \Theta_\beta$ with Θ_α and Θ_β being compact convex subsets of $\mathbb{R}^p$ and $\mathbb{R}^q$, respectively.

Moreover, ϵ and T are known constants,

$a : \mathbb{R}^d \times \Theta_\alpha \to \mathbb{R}^d$ and $b : \mathbb{R}^d \times \Theta_\beta \to \mathbb{R}^d \otimes \mathbb{R}^r$.

$\theta^* = (\alpha^*, \beta^*)$ is the true value of θ.

We assume that $\theta^* \in \mathrm{Int}(\Theta)$ and the parameter spaces have locally Lipschitz boundaries, see Adam and Fournier (2003).

The data are discrete observations $\mathbf{X}_n = (X_{t_i^n})_{0 \le i \le n}$ with $t_i^n = ih_n$ and $h_n = T/n$.

The asymptotics will be considered for $\epsilon \to 0$, $n \to \infty$ and $\frac{1}{\epsilon\sqrt{n}} = O(1)$. Furthermore, there exists a constant $\gamma \in (0, 1]$ such that $\epsilon(\sqrt{n})^\gamma = O(1)$.

History

For statistics of small diffusion processes, there are many work.

A family of small diffusion processes defined by (1) is an important class of dynamical systems with small perturbations. For dynamical systems with small perturbations, see Azencott (1982) and Freidlin and Wentzell (1998).

Statistical inference for continuously observed small diffusion processes is well developed by Kutoyants (1984, 1994), Yoshida (1992a, 2003), Iacus (2000), Iacus and Kutoyants (2001), Uchida and Yoshida (2004a) and references therein.

Furthermore, there are a number of researches on parametric inference for discretely observed small diffusion processes, see Genon-Catalot (1990), Laredo (1990), Sørensen (2000, 2012), Sørensen and Uchida (2003), Uchida (2003, 2004, 2006, 2008), Gloter and Sørensen (2009) and Guy et. al. (2014).

For applications of small diffusion processes to mathematical finance, see Yoshida (1992b), Kunitomo and Takahashi (2001), Takahashi and Yoshida (2004), Uchida and Yoshida (2004b) and references therein.

Notation and Assumptions

Let X_t^0 be the solution of the ordinary differential equation corresponding to $\epsilon = 0$, i.e., $dX_t^0 = a(X_t^0, \alpha^*)dt$, $X_0^0 = x_0$.

Set $A^{\otimes 2} = AA^\star$ and $C[A] = \mathrm{tr}(CA^\star)$ for matrices A and C of the same size, where $\star$ means the transpose.

Let $B(x, \beta) = bb^\star(x, \beta)$, $\Delta X_i = X_{t_i} - X_{t_{i-1}}$, $a_{i-1}(\alpha) = a(X_{t_{i-1}}, \alpha)$ and $B_{i-1}(\beta) = B(X_{t_{i-1}}, \beta)$.

For a matrix A, we define $\|A\|^2 = \mathrm{tr}(AA^\star)$ and $|\cdot|$ denote the Euclidian norm.

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

Let $C_{\uparrow}^{k,l}(\mathbb{R}^d \times \Theta; \mathbb{R}^d)$ denote the space of all functions f satisfying the following conditions: (i) $f(x, \theta)$ is an $\mathbb{R}^d$-valued function on $\mathbb{R}^d \times \Theta$ that is continuously differentiable with respected to x and θ up to order k and l, respectively. (ii) for $|\boldsymbol{n}| = 0, 1, \ldots, k$ and $|\boldsymbol{\nu}| = 0, 1, \ldots, l$, there exists $C > 0$ such that $\sup_{\theta \in \Theta} |\delta^{\boldsymbol{\nu}} \partial^{\boldsymbol{n}} f| \leq C(1 + |x|)^C$ for all x. Here, $\boldsymbol{n} = (n_1, \ldots, n_d)$ and $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_l)$ are multi-indices, $l = \dim(\Theta)$, $|\boldsymbol{n}| = n_1 + \cdots + n_d$, $|\boldsymbol{\nu}| = \nu_1 + \cdots + \nu_l$, $\partial^{\boldsymbol{n}} = \partial_1^{n_1} \cdots \partial_d^{n_d}$, $\partial_i = \partial/\partial x_i$, $i = 1, \ldots, d$, $\partial^{\boldsymbol{\nu}} = \partial_1^{\nu_1} \cdots \partial_l^{\nu_l}$, $\partial_j = \partial/\partial \theta_j$, $j = 1, \ldots, l$.

We simply write $\|f\|_M \lesssim \|g\|_M$ when there exists a constant $C_M > 0$ such that $\|f\|_M \leq C_M \|g\|_M$ for $f, g \in L^M(P)$.

In this paper, we make the following assumptions.

[**A1**] **(i)** There exists $K > 0$ such that for all $x, y \in \mathbb{R}^d$,
$$\sup_{\alpha\in\Theta_\alpha}|a(x,\alpha)-a(y,\alpha)| + \sup_{\beta\in\Theta_\beta}\|b(x,\beta)-b(y,\beta)\| \le K|x-y|.$$

(ii) $\inf_{x,\beta} \det B(x,\beta) > 0$.

[**A2**] $a(x,\alpha) \in C_{\uparrow}^{6,4}(\mathbb{R}^d \times \Theta_\alpha; \mathbb{R}^d)$, $b(x,\beta) \in C_{\uparrow}^{6,4}(\mathbb{R}^d \times \Theta_\beta; \mathbb{R}^d \otimes \mathbb{R}^r)$.

Let

$$\begin{aligned}
\mathbb{Y}^{(1)}(\alpha) &= -\frac{1}{2}\int_{\mathbb{R}^d} \left|a(X_t^0,\alpha)-a(X_t^0,\alpha^*)\right|^2 dt,\\
\mathbb{Y}^{(2)}(\beta) &= -\frac{1}{2}\int_{\mathbb{R}^d} \left\{\mathrm{tr}\left[B(X_t^0,\beta)^{-1}B(X_t^0,\beta^*)-I_d\right] + \log\frac{\det B(X_t^0,\beta)}{\det B(X_t^0,\beta^*)}\right\} dt,\\
\mathbb{Y}^{(3)}(\alpha) &= -\frac{1}{2}\int_{\mathbb{R}^d} B(X_t^0,\beta^*)^{-1}\left[\left(a(X_t^0,\alpha)-a(X_t^0,\alpha^*)\right)^{\otimes 2}\right] dt.
\end{aligned}$$

[**A3**] There exist positive constants $\chi^{(1)}, \chi^{(2)}, \chi^{(3)}$ such that

$$\begin{aligned}
\mathbb{Y}^{(1)}(\alpha) &\le -\chi^{(1)}|\alpha-\alpha^*|^2,\\
\mathbb{Y}^{(2)}(\beta) &\le -\chi^{(2)}|\beta-\beta^*|^2,\\
\mathbb{Y}^{(3)}(\alpha) &\le -\chi^{(3)}|\alpha-\alpha^*|^2
\end{aligned}$$

for all $\alpha \in \Theta_\alpha$ and $\beta \in \Theta_\beta$.

Let

$$I(\theta) = \begin{pmatrix} \left(I_a^{ij}(\theta)\right)_{1\le i,j\le p} & 0 \\ 0 & \left(I_b^{ij}(\theta)\right)_{1\le i,j\le q} \end{pmatrix},$$

$$I_a^{ij}(\theta) = \int_{\mathbb{R}^d} \left(\partial_{\alpha_i} a(X_t^0, \alpha^*)\right)^\star B(X_t^0, \beta^*)\partial_{\alpha_j} a(X_t^0, \alpha^*)dt,$$

$$I_b^{ij}(\theta) = \frac{1}{2}\int_{\mathbb{R}^d} \mathrm{tr}\left\{B^{-1}(\partial_{\beta_i} B)B^{-1}(\partial_{\beta_j} B)(X_t^0, \beta^*)\right\} dt.$$

Joint estimation

The quasi-log likelihood function is defined as

$$U_{\epsilon,n}(\alpha,\beta) \;:=\; -\frac{1}{2}\sum_{i=1}^{n}\Big\{\log\det B_{i-1}(\beta) + (\epsilon^2 h_n)^{-1} B_{i-1}^{-1}(\beta)\left[(\Delta X_i - h_n a_{i-1}(\alpha))^{\otimes 2}\right]\Big\},$$

The joint estimators $\widehat{\alpha}_{\epsilon,n}^{(J)}$ and $\widehat{\beta}_{\epsilon,n}^{(J)}$ are defined as

$$U_{\epsilon,n}(\widehat{\alpha}_{\epsilon,n}^{(J)}, \widehat{\beta}_{\epsilon,n}^{(J)}) \;=\; \sup_{\alpha\in\Theta_\alpha,\beta\in\Theta_\beta} U_{\epsilon,n}(\alpha,\beta)$$

Theorem 1 (Sørensen and Uchida (2003)) *Under some regularity conditions, as* $\epsilon \to 0$, $n \to \infty$ *and* $\frac{1}{\epsilon\sqrt{n}} = O(1)$,

$$\left(\epsilon^{-1}(\widehat{\alpha}_{\epsilon,n}^{(J)} - \alpha^*), \sqrt{n}(\widehat{\beta}_{\epsilon,n}^{(J)} - \beta^*)\right) \xrightarrow{d} (\zeta_1,\zeta_2) \sim N_{p+q}(0, I(\theta^*)^{-1}).$$

Adaptive estimation

The quasi-log likelihood functions are defined as

$$
\begin{aligned}
U_{\epsilon,n}^{(1)}(\alpha) &:= -\frac{1}{2}\sum_{i=1}^{n} \left|\Delta X_i - h_n a_{i-1}(\alpha)\right|^2 (\epsilon^2 h_n)^{-1}, \\
U_{\epsilon,n}^{(2)}(\alpha,\beta) &:= -\frac{1}{2}\sum_{i=1}^{n} \left\{\log\det B_{i-1}(\beta) + (\epsilon^2 h_n)^{-1} B_{i-1}^{-1}(\beta)\left[(\Delta X_i - h_n a_{i-1}(\alpha))^{\otimes 2}\right]\right\}, \\
U_{\epsilon,n}^{(3)}(\alpha,\beta) &:= -\frac{1}{2}\sum_{i=1}^{n} (\epsilon^2 h_n)^{-1} B_{i-1}^{-1}(\beta)\left[(\Delta X_i - h_n a_{i-1}(\alpha))^{\otimes 2}\right].
\end{aligned}
$$

The initial estimator $\hat{\alpha}_{\epsilon,n}^{(1)}$ and the adaptive estimators $\hat{\alpha}_{\epsilon,n}$ and $\hat{\beta}_{\epsilon,n}$ are defined as

$$
\begin{aligned}
U_{\epsilon,n}^{(1)}(\hat{\alpha}_{\epsilon,n}^{(1)}) &= \sup_{\alpha\in\Theta_\alpha} U_{\epsilon,n}^{(1)}(\alpha), \\
U_{\epsilon,n}^{(2)}(\hat{\alpha}_{\epsilon,n}^{(1)}, \hat{\beta}_{\epsilon,n}) &= \sup_{\beta\in\Theta_\beta} U_{\epsilon,n}^{(2)}(\hat{\alpha}_{\epsilon,n}^{(1)}, \beta), \\
U_{\epsilon,n}^{(3)}(\hat{\alpha}_{\epsilon,n}, \hat{\beta}_{\epsilon,n}) &= \sup_{\alpha\in\Theta_\alpha} U_{\epsilon,n}^{(3)}(\alpha, \hat{\beta}_{\epsilon,n}).
\end{aligned}
$$

Proposition 1 *Assume [A1]–[A3]. Then for all* $M > 0$*, as* $\epsilon \to 0$*,* $n \to \infty$ *and* $\frac{1}{\epsilon\sqrt{n}} = O(1)$*,*

(i) $\sup_{\epsilon,n} E_{\theta^*}\left[\left|\epsilon^{-1}(\widehat{\alpha}_{\epsilon,n}^{(1)} - \alpha^*)\right|^M\right] < \infty.$

(ii) $\sup_{\epsilon,n} E_{\theta^*}\left[\left|\sqrt{n}(\widehat{\beta}_{\epsilon,n} - \beta^*)\right|^M\right] < \infty.$

(iii) $\sup_{\epsilon,n} E_{\theta^*}\left[\left|\epsilon^{-1}(\widehat{\alpha}_{\epsilon,n} - \alpha^*)\right|^M\right] < \infty.$

Theorem 2 *Assume [A1]–[A3]. Then, as* $\epsilon \to 0$*,* $n \to \infty$ *and* $\frac{1}{\epsilon\sqrt{n}} = O(1)$*,*

$$\left(\epsilon^{-1}(\widehat{\alpha}_{\epsilon,n} - \alpha^*), \sqrt{n}(\widehat{\beta}_{\epsilon,n} - \beta^*)\right) \xrightarrow{d} (\zeta_1, \zeta_2) \sim N_{p+q}(0, I(\theta^*)^{-1})$$

and

$$E_{\theta^*}\left[f\left(\epsilon^{-1}(\widehat{\alpha}_{\epsilon,n} - \alpha^*), \sqrt{n}(\widehat{\beta}_{\epsilon,n} - \beta^*)\right)\right] \to \mathbb{E}\left[f(\zeta_1, \zeta_2)\right]$$

for all continuous functions f *of at most polynomial growth.*

Example and simulation results

We consider the following one-dimensional diffusion process

$$\begin{cases} dX_t^\epsilon = (\alpha_1 - \alpha_2 X_t^\epsilon - 2\sin(\alpha_3 X_t^\epsilon))\,dt + \epsilon \dfrac{\beta_2 + (X_t^\epsilon)^2}{1 + \beta_1 (X_t^\epsilon)^2} dw_t, \quad t \in [0,1], \quad \epsilon \in (0,1], \\ X_0 = 2. \end{cases}$$

where $\theta = (\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2)$ is an unknown parameter, and

the true value is $\theta^* = (\alpha_1^*, \alpha_2^*, \alpha_3^*, \beta_1^*, \beta_2^*) = (3, 7, 5, 0.5, 5)$.

The parameter space is $\Theta^0 := [0.01, 20] \times [0.01, 20] \times [2.4, 7.6] \times [0.01, 20] \times [0.01, 20]$.

We examine the asymptotic behavior of the adaptive ML type estimator $\widehat{\theta} = (\widehat{\alpha}_{\epsilon,n}, \widehat{\beta}_{\epsilon,n})$ in Theorem 2 and the joint ML type estimator $\widehat{\theta}^{(J)} = (\widehat{\alpha}_{\epsilon,n}^{(J)}, \widehat{\beta}_{\epsilon,n}^{(J)})$ defined by $U_{\epsilon,n}^{(2)}(\widehat{\alpha}_{\epsilon,n}^{(J)}, \widehat{\beta}_{\epsilon,n}^{(J)}) = \sup_{\alpha,\beta} U_{\epsilon,n}^{(2)}(\alpha, \beta)$.

We set $\epsilon = 0.01, 0.05, 0.1, 0.15$ and $h_n = 1/100, 1/1000$.

The simulations were done for the two cases,
one is the case that the initial value is the true value, and
the other is the case that the initial value is derived from the uniform distribution on Θ^0.
Here we note that the initial value is used to calculate the ML type estimator with optim () in R.

We calculate the mean and the standard deviation of each estimator from 1000 independent sample paths based on the true model.

$$dX_t^\epsilon = (\alpha_1 - \alpha_2 X_t^\epsilon - 2\sin(\alpha_3 X_t^\epsilon))\,dt + \epsilon\frac{\beta_2 + (X_t^\epsilon)^2}{1+\beta_1(X_t^\epsilon)^2}dw_t \quad t\in[0,1], \quad \epsilon\in(0,1],$$

$X_0 = 2.$

Table 1. joint ML type estimator with the initial value being the true value

ϵ	n	$\widehat{\alpha}_1^{(J)}$	$\widehat{\alpha}_2^{(J)}$	$\widehat{\alpha}_3^{(J)}$	$\widehat{\beta}_1^{(J)}$	$\widehat{\beta}_2^{(J)}$
		3.0	7.0	5.0	0.5	5.0
0.15	100	3.4162 (1.6327)	7.5285 (1.5710)	4.8502 (0.8656)	0.6247 (0.2765)	4.9445 (0.7191)
	1000	3.5418 (1.7199)	7.8059 (1.6563)	4.9337 (0.8562)	0.5089 (0.0670)	4.9841 (0.1914)
0.1	100	3.1370 (0.9790)	7.1371 (1.0017)	4.8655 (0.5679)	0.6185 (0.2734)	4.9313 (0.6680)
	1000	3.2076 (1.0157)	7.3499 (1.0507)	4.9577 (0.5401)	0.5081 (0.0664)	4.9833 (0.1823)
0.05	100	2.9805 (0.4225)	6.8718 (0.4629)	4.8844 (0.1882)	0.5953 (0.2367)	4.8855 (0.5253)
	1000	3.0204 (0.4237)	7.0451 (0.4776)	4.9828 (0.1877)	0.5069 (0.0647)	4.9818 (0.1709)
0.01	100	2.9589 (0.0816)	6.8170 (0.0910)	4.8923 (0.0357)	0.5515 (0.2246)	4.8467 (0.4982)
	1000	2.9936 (0.0818)	6.9818 (0.0943)	4.9893 (0.0354)	0.5065 (0.0641)	4.9812 (0.1669)

When $\epsilon = 0.05$ or 0.01, and $n = 1000$, both $\widehat{\alpha}^{(J)}$ and $\widehat{\beta}^{(J)}$ have good behavior.

$$
\begin{aligned}
dX_t^\epsilon &= (\alpha_1 - \alpha_2 X_t^\epsilon - 2\sin(\alpha_3 X_t^\epsilon))\,dt + \epsilon \frac{\beta_2 + (X_t^\epsilon)^2}{1+\beta_1 (X_t^\epsilon)^2} dw_t \quad t \in [0,1], \quad \epsilon \in (0,1], \\
X_0 &= 2.
\end{aligned}
$$

Table 2. adaptive ML type estimator with the initial value being the true value

ϵ	n	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\beta}_1$	$\hat{\beta}_2$
		3.0	7.0	5.0	0.5	5.0
0.15	100	3.4605 (1.6985)	7.5722 (1.5909)	4.8136 (0.9786)	0.5921 (0.2441)	4.8809 (0.6398)
	1000	3.5562 (1.7363)	7.8194 (1.6617)	4.9033 (0.9728)	0.5070 (0.0665)	4.9803 (0.1905)
0.1	100	3.1489 (0.9838)	7.1589 (1.0007)	4.8393 (0.6385)	0.5831 (0.2407)	4.8709 (0.6081)
	1000	3.2169 (1.0229)	7.3667 (1.0517)	4.9283 (0.6230)	0.506 (0.0657)	4.9793 (0.1812)
0.05	100	2.9804 (0.4215)	6.8725 (0.4618)	4.8843 (0.1883)	0.5633 (0.2175)	4.8352 (0.5022)
	1000	3.0203 (0.4237)	7.0451 (0.4776)	4.9827 (0.1877)	0.5046 (0.0644)	4.9779 (0.1705)
0.01	100	2.9587 (0.0815)	6.8171 (0.0909)	4.8928 (0.0357)	0.5212 (0.2040)	4.7987 (0.4748)
	1000	2.9935 (0.0818)	6.9818 (0.0943)	4.9893 (0.0354)	0.5042 (0.0637)	4.9774 (0.1665)

Similarly as the joint estimator in Table 1, both the adaptive estimators $\hat{\alpha}$ and $\hat{\beta}$ also have good performance.

$$dX_t^\epsilon = (\alpha_1 - \alpha_2 X_t^\epsilon - 2\sin(\alpha_3 X_t^\epsilon))\,dt + \epsilon\frac{\beta_2 + (X_t^\epsilon)^2}{1+\beta_1(X_t^\epsilon)^2}dw_t \quad t\in[0,1], \quad \epsilon\in(0,1],$$
$$X_0 = 2.$$

Table 3. initial estimator of the adaptive ML type estimator with the initial value being the true value

ϵ	n	$\widehat{\alpha}_1^{(1)}$	$\widehat{\alpha}_2^{(1)}$	$\widehat{\alpha}_3^{(1)}$
		3.0	7.0	5.0
0.15	100	3.5081 (1.7313)	7.6533 (1.6300)	4.7968 (0.9616)
	1000	3.5921 (1.7857)	7.8896 (1.7200)	4.8817 (0.9533)
0.1	100	3.1742 (1.0122)	7.2051 (1.0244)	4.8189 (0.6075)
	1000	3.2292 (1.0553)	7.3999 (1.0879)	4.9076 (0.6089)
0.05	100	2.9863 (0.4293)	6.8895 (0.4723)	4.8825 (0.1940)
	1000	3.0206 (0.4363)	7.0541 (0.4955)	4.9755 (0.1960)
0.01	100	2.9610 (0.0827)	6.8225 (0.0931)	4.8978 (0.0365)
	1000	2.9933 (0.0832)	6.9824 (0.0963)	4.9902 (0.0367)

$$
\begin{aligned}
dX_t^\epsilon &= (\alpha_1 - \alpha_2 X_t^\epsilon - 2\sin(\alpha_3 X_t^\epsilon))\,dt + \epsilon \frac{\beta_2 + (X_t^\epsilon)^2}{1+\beta_1 (X_t^\epsilon)^2} dw_t \quad t \in [0,1], \quad \epsilon \in (0,1], \\
X_0 &= 2.
\end{aligned}
$$

Table 4. joint ML type estimator with the initial value being the random value

ϵ	n	$\hat{\alpha}_1^{(J)}$	$\hat{\alpha}_2^{(J)}$	$\hat{\alpha}_3^{(J)}$	$\hat{\beta}_1^{(J)}$	$\hat{\beta}_2^{(J)}$
		3.0	7.0	5.0	0.5	5.0
0.15	100	3.4545 (1.9742)	7.4284 (1.7703)	5.2094 (1.5162)	0.5901 (0.2533)	4.9144 (0.6432)
	1000	3.6910 (2.2078)	7.8343 (1.9058)	4.9913 (1.3962)	0.5067 (0.0670)	4.9824 (0.1907)
0.1	100	3.2023 (1.4402)	7.1338 (1.2781)	5.0704 (1.1601)	0.5885 (0.2755)	4.9260 (0.6705)
	1000	3.2593 (1.4132)	7.3344 (1.2769)	5.1569 (1.1301)	0.5044 (0.0661)	4.9814 (0.1815)
0.05	100	2.9726 (0.4274)	6.8559 (0.4778)	4.9102 (0.4249)	0.5861 (0.2444)	4.8888 (0.5251)
	1000	3.0110 (0.4256)	7.0273 (0.4928)	5.0597 (0.4851)	0.5039 (0.0654)	4.9819 (0.1709)
0.01	100	2.9589 (0.0816)	6.8170 (0.0910)	4.8923 (0.0357)	0.5515 (0.2246)	4.8467 (0.4982)
	1000	2.9926 (0.0847)	6.9775 (0.1321)	4.9841 (0.1210)	0.5055 (0.0678)	4.9837 (0.1767)

The s.d. of $\hat{\alpha}_3^{(J)}$ is larger than the one of $\hat{\alpha}_3^{(J)}$ in Table 1 since several estimators $\hat{\alpha}_3^{(J)}$ have considerable biases among 1000 sample paths.

$$
\begin{aligned}
dX_t^\epsilon &= \left(\alpha_1 - \alpha_2 X_t^\epsilon - 2\sin\left(\alpha_3 X_t^\epsilon\right)\right) dt + \epsilon \frac{\beta_2 + (X_t^\epsilon)^2}{1 + \beta_1 (X_t^\epsilon)^2} dw_t \quad t \in [0,1], \quad \epsilon \in (0,1], \\
X_0 &= 2.
\end{aligned}
$$

Table 5. adaptive ML type estimator with the initial value being the random value

ϵ	n	$\widehat{\alpha}_1$	$\widehat{\alpha}_2$	$\widehat{\alpha}_3$	$\widehat{\beta}_1$	$\widehat{\beta}_2$
		3.0	7.0	5.0	0.5	5.0
0.15	100	3.6011 (2.0459)	7.6240 (1.7440)	4.7496 (1.2698)	0.5803 (0.2349)	4.8658 (0.5965)
	1000	3.7217 (2.0956)	7.8983 (1.8084)	4.7904 (1.2382)	0.5060 (0.0661)	4.979 (0.1899)
0.1	100	3.2182 (1.2804)	7.1935 (1.1531)	4.8433 (0.9191)	0.5740 (0.2457)	4.8682 (0.6083)
	1000	3.3174 (1.4067)	7.4276 (1.2367)	4.8770 (0.8756)	0.5051 (0.0656)	4.9787 (0.1815)
0.05	100	2.9808 (0.4219)	6.8730 (0.4623)	4.8899 (0.2232)	0.5608 (0.2204)	4.8374 (0.5019)
	1000	3.0203 (0.4237)	7.0451 (0.4776)	4.9827 (0.1877)	0.5046 (0.0644)	4.9779 (0.1705)
0.01	100	2.9587 (0.0815)	6.8171 (0.0909)	4.8928 (0.0357)	0.5212 (0.2040)	4.7987 (0.4747)
	1000	2.9935 (0.0818)	6.9818 (0.0943)	4.9893 (0.0354)	0.5042 (0.0637)	4.9773 (0.1665)

The s.d. of the adaptive estimator is as small as the one of the adaptive estimator (with the initial value being the true value) is.

$$
\begin{aligned}
dX_t^\epsilon &= (\alpha_1 - \alpha_2 X_t^\epsilon - 2\sin(\alpha_3 X_t^\epsilon))\,dt + \epsilon \frac{\beta_2 + (X_t^\epsilon)^2}{1+\beta_1 (X_t^\epsilon)^2} dw_t \quad t \in [0,1], \quad \epsilon \in (0,1], \\
X_0 &= 2.
\end{aligned}
$$

Table 6. initial estimator of the adaptive ML type estimator with the initial value being the random value

ϵ	n	$\hat{\alpha}_1^{(1)}$	$\hat{\alpha}_2^{(1)}$	$\hat{\alpha}_3^{(1)}$
		3.0	7.0	5.0
0.15	100	3.6160 (2.0654)	7.6670 (1.7784)	4.7442 (1.2441)
	1000	3.7232 (2.1337)	7.9242 (1.8641)	4.7745 (1.2055)
0.1	100	3.2310 (1.2928)	7.2223 (1.1731)	4.8267 (0.8947)
	1000	3.3170 (1.4227)	7.4414 (1.2634)	4.8573 (0.8504)
0.05	100	2.9838 (0.4315)	6.8852 (0.4776)	4.8993 (0.2851)
	1000	3.0206 (0.4363)	7.0541 (0.4955)	4.9755 (0.1960)
0.01	100	2.9610 (0.0827)	6.8225 (0.0931)	4.8978 (0.0365)
	1000	2.9931 (0.0841)	6.9819 (0.0978)	4.9901 (0.0368)

Tables 1–3 are the results of the case with the initial value being the true value and
Tables 4–6 are the ones with the initial value being the random value.

In Tables 1–3, the values of the adaptive estimator seem to be quite similar to the ones of the joint estimator, that is, there is hardly any difference between the adaptive estimator and the joint estimator about the estimation accuracy in this model if we choose the true value as the initial value.

However in Table 4, the joint estimator does not have good performance for the drift parameter α_3 when the initial value is derived from the uniform distribution on Θ^0.

On the other hand, the adaptive estimator for α_3 works well in Tables 5–6 even if the initial value is derived from the uniform distribution on Θ^0.

When the case $\epsilon = 0.15$ both estimators show the bad performance. Moreover as ϵ becomes large, the performance of the estimators grows worse. This is because in these cases the diffusion term is not so small and there exists the bias in the first order estimators. Therefore, we need to consider the asymptotic expansion of the estimators with respect to ϵ and this is the future work.

2. Adaptive Bayes type estimation for small diffusion processes

Set $A^{\otimes 2} = AA^\star$ and $C[A] = \mathrm{tr}(CA^\star)$ for matrices A and C of the same size, where $\star$ means the transpose.

Let $B(x,\beta) = bb^\star(x,\beta)$, $\Delta X_i = X_{t_i} - X_{t_{i-1}}$, $a_{i-1}(\alpha) = a(X_{t_{i-1}},\alpha)$ and $B_{i-1}(\beta) = B(X_{t_{i-1}},\beta)$.

For a matrix A, we define $\|A\|^2 = \mathrm{tr}(AA^\star)$ and $|\cdot|$ denote the Euclidian norm.

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

We simply write $\|f\|_M \lesssim \|g\|_M$ when there exists a constant $C_M > 0$ such that $\|f\|_M \le C_M \|g\|_M$ for $f, g \in L^M(P)$.

Let X_t^0 be the solution of the ordinary differential equation corresponding to $\epsilon = 0$, i.e., $dX_t^0 = a(X_t^0, \alpha^*)dt$, $X_0^0 = x_0$.

Let $C_\uparrow^{k,l}(\mathbb{R}^d \times \Theta; \mathbb{R}^d)$ denote the space of all functions f satisfying the following conditions: (i) $f(x,\theta)$ is an $\mathbb{R}^d$-valued function on $\mathbb{R}^d \times \Theta$ that is continuously differentiable with respected to x and θ up to order k and l, respectively. (ii) for $|\boldsymbol{n}| = 0,1,\ldots,k$ and $|\boldsymbol{\nu}| = 0,1,\ldots,l$, there exists $C > 0$ such that $\sup_{\theta\in\Theta}|\delta^{\boldsymbol{\nu}}\partial^{\boldsymbol{n}} f| \le C(1+|x|)^C$ for all x. Here, $\boldsymbol{n} = (n_1,\ldots,n_d)$ and $\boldsymbol{\nu} = (\nu_1,\ldots,\nu_l)$ are multi-indices, $l = \dim(\Theta)$, $|\boldsymbol{n}| = n_1 + \cdots + n_d$, $|\boldsymbol{\nu}| = \nu_1 + \cdots + \nu_l$, $\partial^{\boldsymbol{n}} = \partial_1^{n_1}\cdots\partial_d^{n_d}$, $\partial_i = \partial/\partial x_i$, $i = 1,\ldots,d$, $\partial^{\boldsymbol{\nu}} = \partial_1^{\nu_1}\cdots\partial_l^{\nu_l}$, $\partial_j = \partial/\partial\theta_j$, $j = 1,\ldots,l$.

In this paper, we make the assumptions as follows.

[A1] **(i)** There exists $K > 0$ such that for all $x, y \in \mathbb{R}^d$,
$\sup_{\alpha\in\Theta_\alpha} |a(x,\alpha) - a(y,\alpha)| + \sup_{\beta\in\Theta_\beta} \|b(x,\beta) - b(y,\beta)\| \leq K|x-y|$.

(ii) $\inf_{x,\beta} \det B(x,\beta) > 0$.

[A2] $a(x,\alpha) \in C_{\uparrow}^{6,4}(\mathbb{R}^d \times \Theta_\alpha; \mathbb{R}^d)$, $b(x,\beta) \in C_{\uparrow}^{6,4}(\mathbb{R}^d \times \Theta_\beta; \mathbb{R}^d \otimes \mathbb{R}^r)$.

Let

$$\begin{aligned}
\mathbb{Y}^{(1)}(\alpha) &= -\frac{1}{2}\int_{\mathbb{R}^d} \left|a(X_t^0,\alpha) - a(X_t^0,\alpha^*)\right|^2 dt,\\
\mathbb{Y}^{(2)}(\beta) &= -\frac{1}{2}\int_{\mathbb{R}^d} \left\{\mathrm{tr}\left[B(X_t^0,\beta)^{-1}B(X_t^0,\beta^*) - I_d\right] + \log\frac{\det B(X_t^0,\beta)}{\det B(X_t^0,\beta^*)}\right\} dt,\\
\mathbb{Y}^{(3)}(\alpha) &= -\frac{1}{2}\int_{\mathbb{R}^d} B(X_t^0,\beta^*)^{-1}\left[\left(a(X_t^0,\alpha) - a(X_t^0,\alpha^*)\right)^{\otimes 2}\right] dt.
\end{aligned}$$

[A3] There exist positive constants $\chi^{(1)}, \chi^{(2)}, \chi^{(3)}$ such that for all $\alpha \in \Theta_\alpha$ and $\beta \in \Theta_\beta$,

$$\begin{aligned}
\mathbb{Y}^{(1)}(\alpha) &\leq -\chi^{(1)}|\alpha - \alpha^*|^2,\\
\mathbb{Y}^{(2)}(\beta) &\leq -\chi^{(2)}|\beta - \beta^*|^2,\\
\mathbb{Y}^{(3)}(\alpha) &\leq -\chi^{(3)}|\alpha - \alpha^*|^2.
\end{aligned}$$

Let

$$I(\theta^*) = \begin{pmatrix} \left(I_a^{ij}(\theta^*)\right)_{1\le i,j\le p} & 0 \\ 0 & \left(I_b^{ij}(\beta^*)\right)_{1\le i,j\le q} \end{pmatrix},$$

$$I_a^{ij}(\theta^*) = \int_{\mathbb{R}^d} \left(\partial_{\alpha_i} a(X_t^0, \alpha^*)\right)^\star B(X_t^0, \beta^*)\partial_{\alpha_j} a(X_t^0, \alpha^*)dt,$$

$$I_b^{ij}(\beta^*) = \frac{1}{2}\int_{\mathbb{R}^d} \mathrm{tr}\left\{B^{-1}(\partial_{\beta_i} B)B^{-1}(\partial_{\beta_j} B)(X_t^0, \beta^*)\right\} dt.$$

We assume that the prior densities $\pi_1(\alpha)$ and $\pi_2(\beta)$ are continuous and satisfy that $0 < \inf_{\alpha\in\Theta_\alpha} \pi_1(\alpha) \le \sup_{\alpha\in\Theta_\alpha} \pi_1(\alpha) < \infty$ and $0 < \inf_{\beta\in\Theta_\beta} \pi_2(\beta) \le \sup_{\beta\in\Theta_\beta} \pi_2(\beta) < \infty$.

The quasi log-likelihood functions $U^{(1)}_{\epsilon,n}(\alpha)$, $U^{(2)}_{\epsilon,n}(\alpha,\beta)$ and $U^{(3)}_{\epsilon,n}(\alpha,\beta)$ are as follows.

$$
\begin{aligned}
U^{(1)}_{\epsilon,n}(\alpha) &= -\frac{1}{2}\sum_{i=1}^{n} \left|\Delta X_i - h_n a_{i-1}(\alpha)\right|^2 (\epsilon^2 h_n)^{-1}, \\
U^{(2)}_{\epsilon,n}(\alpha,\beta) &= -\frac{1}{2}\sum_{i=1}^{n} \left\{\log\det B_{i-1}(\beta) + (\epsilon^2 h_n)^{-1} B^{-1}_{i-1}(\beta)\left[(\Delta X_i - h_n a_{i-1}(\alpha))^{\otimes 2}\right]\right\}, \\
U^{(3)}_{\epsilon,n}(\alpha,\beta) &= -\frac{1}{2}\sum_{i=1}^{n} (\epsilon^2 h_n)^{-1} B^{-1}_{i-1}(\beta)\left[(\Delta X_i - h_n a_{i-1}(\alpha))^{\otimes 2}\right].
\end{aligned}
$$

The initial Bayes type estimator $\tilde{\alpha}^{(1)}_{\epsilon,n}$ and the adaptive Bayes type estimators $\tilde{\alpha}_{\epsilon,n}$ and $\tilde{\beta}_{\epsilon,n}$ are defined as

$$
\tilde{\alpha}^{(1)}_{\epsilon,n} = \frac{\int_{\Theta_\alpha} \alpha \exp\left\{U^{(1)}_{\epsilon,n}(\alpha)\right\}\pi_1(\alpha)d\alpha}{\int_{\Theta_\alpha} \exp\left\{U^{(1)}_{\epsilon,n}(\alpha)\right\}\pi_1(\alpha)d\alpha},
$$

$$
\tilde{\beta}_{\epsilon,n} = \frac{\int_{\Theta_\beta} \beta \exp\left\{U^{(2)}_{\epsilon,n}(\tilde{\alpha}^{(1)}_{\epsilon,n},\beta)\right\}\pi_2(\beta)d\beta}{\int_{\Theta_\beta} \exp\left\{U^{(2)}_{\epsilon,n}(\tilde{\alpha}^{(1)}_{\epsilon,n},\beta)\right\}\pi_2(\beta)d\beta},
$$

$$
\tilde{\alpha}_{\epsilon,n} = \frac{\int_{\Theta_\alpha} \alpha \exp\left\{U^{(3)}_{\epsilon,n}(\alpha,\tilde{\beta}_{\epsilon,n})\right\}\pi_1(\alpha)d\alpha}{\int_{\Theta_\alpha} \exp\left\{U^{(3)}_{\epsilon,n}(\alpha,\tilde{\beta}_{\epsilon,n})\right\}\pi_1(\alpha)d\alpha}.
$$

Proposition 2 *Assume [A1]–[A3]. Then for all $M > 0$, as $\frac{1}{\epsilon\sqrt{n}} = O(1)$,*

(i) $\sup_{\epsilon,n} E_{\theta^*}\left[\left|\epsilon^{-1}(\tilde{\alpha}^{(1)}_{\epsilon,n} - \alpha^*)\right|^M\right] < \infty.$

(ii) $\sup_{\epsilon,n} E_{\theta^*}\left[\left|\sqrt{n}(\tilde{\beta}_{\epsilon,n} - \beta^*)\right|^M\right] < \infty.$

(iii) $\sup_{\epsilon,n} E_{\theta^*}\left[\left|\epsilon^{-1}(\tilde{\alpha}_{\epsilon,n} - \alpha^*)\right|^M\right] < \infty.$

Theorem 3 *Assume [A1]–[A3]. Then, as $\frac{1}{\epsilon\sqrt{n}} = O(1)$,*

$$\left(\epsilon^{-1}(\tilde{\alpha}_{\epsilon,n} - \alpha^*), \sqrt{n}(\tilde{\beta}_{\epsilon,n} - \beta^*)\right) \xrightarrow{d} (\zeta_1, \zeta_2) \sim N_{p+q}(0, I(\theta^*)^{-1})$$

and

$$E_{\theta^*}\left[f\left(\epsilon^{-1}(\tilde{\alpha}_{\epsilon,n} - \alpha^*), \sqrt{n}(\tilde{\beta}_{\epsilon,n} - \beta^*)\right)\right] \to \mathbb{E}\left[f(\zeta_1, \zeta_2)\right]$$

for all continuous functions f of at most polynomial growth.

Hybrid type estimation

It is possible to consider the hybrid estimation as follows. Let $r_1, r_2 \in (0,1]$ and $r_2 \le 2r_1\gamma$. Note that $\frac{1}{\epsilon\sqrt{n}} = O(1)$ and $\epsilon(\sqrt{n})^\gamma = O(1)$ as $n \to \infty$ and $\epsilon \to 0$.

$$\mathbb{H}^{(1)}_{\epsilon,n,r_1}(\alpha) = \epsilon^{2-2r_1} U^{(1)}_{\epsilon,n}(\alpha),$$
$$\mathbb{H}^{(2)}_{\epsilon,n,r_2}(\alpha,\beta) = \frac{1}{(\sqrt{n})^{2-2r_2}} U^{(2)}_{\epsilon,n}(\alpha,\beta).$$

The initial Bayes type estimators $\tilde{\alpha}^{(1)}_{\epsilon,n,r_1}$ and $\tilde{\beta}^{(2)}_{\epsilon,n,r_2}$ are defined by

$$\tilde{\alpha}^{(1)}_{\epsilon,n,r_1} = \frac{\int_{\Theta_\alpha} \alpha \exp\left\{\mathbb{H}^{(1)}_{\epsilon,n,r_1}(\alpha))\right\} \pi_1(\alpha) d\alpha}{\int_{\Theta_\alpha} \exp\left\{\mathbb{H}^{(1)}_{\epsilon,n,r_1}(\alpha)\right\} \pi_1(\alpha) d\alpha},$$
$$\tilde{\beta}^{(2)}_{\epsilon,n,r_2} = \frac{\int_{\Theta_\beta} \beta \exp\left\{\mathbb{H}^{(2)}_{\epsilon,n,r_2}(\tilde{\alpha}^{(1)}_{\epsilon,n,r_1},\beta)\right\} \pi_2(\beta) d\beta}{\int_{\Theta_\beta} \exp\left\{\mathbb{H}^{(2)}_{\epsilon,n,r_2}(\tilde{\alpha}^{(1)}_{\epsilon,n,r_1},\beta)\right\} \pi_2(\beta) d\beta}.$$

Hybrid estimators $\check{\alpha}_{\epsilon,n}$ and $\check{\beta}_{\epsilon,n}$ are defined by

$$\begin{aligned} U^{(3)}_{\epsilon,n}(\check{\alpha}_{\epsilon,n}, \tilde{\beta}^{(2)}_{\epsilon,n,r_2}) &= \sup_{\alpha\in\Theta_\alpha} U^{(3)}_{\epsilon,n}(\alpha, \tilde{\beta}^{(2)}_{\epsilon,n,r_2}), \\ U^{(2)}_{\epsilon,n}(\check{\alpha}_{\epsilon,n}, \check{\beta}_{\epsilon,n}) &= \sup_{\beta\in\Theta_\beta} U^{(2)}_{\epsilon,n}(\check{\alpha}_{\epsilon,n}, \beta), \end{aligned}$$

where

$$\begin{aligned} U^{(1)}_{\epsilon,n}(\alpha) &= -\frac{1}{2}\sum_{i=1}^n |\Delta X_i - h_n a_{i-1}(\alpha)|^2 (\epsilon^2 h_n)^{-1}, \\ U^{(2)}_{\epsilon,n}(\alpha,\beta) &= -\frac{1}{2}\sum_{i=1}^n \left\{\log\det B_{i-1}(\beta) + (\epsilon^2 h_n)^{-1} B^{-1}_{i-1}(\beta)\left[(\Delta X_i - h_n a_{i-1}(\alpha))^{\otimes 2}\right]\right\}, \\ U^{(3)}_{\epsilon,n}(\alpha,\beta) &= -\frac{1}{2}\sum_{i=1}^n (\epsilon^2 h_n)^{-1} B^{-1}_{i-1}(\beta)\left[(\Delta X_i - h_n a_{i-1}(\alpha))^{\otimes 2}\right]. \end{aligned}$$

Proposition 3 *Let* $r_1, r_2 \in (0,1]$. *Assume [A1]–[A3]. Then, for all* $M > 0$, *as* $\frac{1}{\epsilon\sqrt{n}} = O(1)$,

(i) $\sup_{\epsilon,n} E_{\theta^*}\left[\left|\epsilon^{-r_1}(\tilde{\alpha}^{(1)}_{\epsilon,n,r_1} - \alpha^*)\right|^M\right] < \infty.$

(ii) $\sup_{\epsilon,n} E_{\theta^*}\left[\left|(\sqrt{n})^{r_2}(\tilde{\beta}^{(2)}_{\epsilon,n,r_2} - \beta^*)\right|^M\right] < \infty.$

(iii) $\sup_{\epsilon,n} E_{\theta^*}\left[\left|\epsilon^{-1}(\breve{\alpha}_{\epsilon,n} - \alpha^*)\right|^M\right] < \infty.$

(iv) $\sup_{\epsilon,n} E_{\theta^*}\left[\left|\sqrt{n}(\breve{\beta}_{\epsilon,n} - \beta^*)\right|^M\right] < \infty.$

Theorem 4 *Assume [A1]–[A3]. Then, as* $\frac{1}{\epsilon\sqrt{n}} = O(1)$,

$$\left(\epsilon^{-1}(\breve{\alpha}_{\epsilon,n} - \alpha^*), \sqrt{n}(\breve{\beta}_{\epsilon,n} - \beta^*)\right) \xrightarrow{d} (\zeta_1, \zeta_2)$$

and

$$E_{\theta^*}\left[f\left(\epsilon^{-1}(\breve{\alpha}_{\epsilon,n} - \alpha^*), \sqrt{n}(\breve{\beta}_{\epsilon,n} - \beta^*)\right)\right] \to \mathbb{E}\left[f(\zeta_1, \zeta_2)\right]$$

for all continuous functions f *of at most polynomial growth.*

Remark 1 *When* $r_1 < 1$ *and* $r_2 = 1$, *it is easy to show that*

$$E_{\theta^*}\left[f\left(\epsilon^{-1}(\breve{\alpha}_{\epsilon,n} - \alpha^*), \sqrt{n}(\tilde{\beta}^{(2)}_{\epsilon,n,r_2} - \beta^*)\right)\right] \to \mathbb{E}\left[f(\zeta_1, \zeta_2)\right]$$

for all continuous functions f *of at most polynomial growth.*

Example and simulation results

We consider one-dimensional diffusion process as follows

$$\begin{cases} dX_t^\epsilon = (\alpha_1 - \alpha_2 X_t^\epsilon - 2\sin(\alpha_3 X_t^\epsilon))\,dt + \epsilon \dfrac{\beta_2 + (X_t^\epsilon)^2}{1+\beta_1 (X_t^\epsilon)^2} dw_t \quad t\in[0,1], \quad \epsilon \in (0,1], \\ X_0 = 2. \end{cases}$$

where $\theta = (\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2)$ is an unknown parameter, and

the true value is $\theta^* = (\alpha_1^*, \alpha_2^*, \alpha_3^*, \beta_1^*, \beta_2^*) = (3, 7, 5, 0.5, 5)$, and

the parameter space is $\Theta = [0.01, 20]^5$.

The data (X_{t_i}) are generated for $\epsilon = 0.01, 0.05, 0.1$ and $n = 100, 1000$.

We do simulations for the hybrid estimator $\check{\theta} = (\check{\alpha}_{\epsilon,n}, \check{\beta}_{\epsilon,n})$ in Theorem 4 and

the adaptive ML type estimator $\widehat{\theta} = (\widehat{\alpha}_{\epsilon,n}, \widehat{\beta}_{\epsilon,n})$ in Theorem 2, which is defined as

$$\begin{aligned} U_{\epsilon,n}^{(1)}(\widehat{\alpha}_{\epsilon,n}^{(1)}) &= \sup_{\alpha\in\Theta_\alpha} U_{\epsilon,n}^{(1)}(\alpha), \\ U_{\epsilon,n}^{(2)}(\widehat{\alpha}_{\epsilon,n}^{(1)}, \widehat{\beta}_{\epsilon,n}) &= \sup_{\beta\in\Theta_\beta} U_{\epsilon,n}^{(2)}(\widehat{\alpha}_{\epsilon,n}^{(1)}, \beta), \\ U_{\epsilon,n}^{(3)}(\widehat{\alpha}_{\epsilon,n}, \widehat{\beta}_{\epsilon,n}) &= \sup_{\alpha\in\Theta_\alpha} U_{\epsilon,n}^{(3)}(\alpha, \widehat{\beta}_{\epsilon,n}). \end{aligned}$$

We do not deal with the adaptive Bayes type estimator in Theorem 3 because
the estimator in Theorem 4 is better than the estimator in Theorem 3 in view of computational cost.

We calculate the mean and the standard deviation of each estimator from 1000 independent sample paths based on the true model.

Optim in R is used to calculate the adaptive ML type estimator $\hat{\theta}$ and the hybrid estimator $\check{\theta}$ and

the initial Bayes type estimators $\tilde{\alpha}$ and $\tilde{\beta}$ are calculated with one of the MCMC method, the mixed preconditioned Crank-Nicolson (MpCN) method proposed by Kamatani (2014) for 5×10^4 Markov chains and 5×10^3 burn-in iterations.

The initial value $(\alpha_1^0, \alpha_2^0, \alpha_3^0, \beta_1^0, \beta_2^0)$ is the true value or is derived from the uniform distribution on Θ.

In Tables 7–9 the performance of the adaptive ML type estimator and the hybrid estimator seems to be almost the same and

both the adaptive ML type estimator and the hybrid estimator are unstable when $\epsilon = 0.1$.

In Table 10 however, the adaptive ML type estimator with the initial value being the random number from the uniform distribution on Θ has a considerable bias for many cases.

On the other hand, both the initial Bayes type estimator and the hybrid estimator show the good performance even if the initial value is derived from the uniform distribution on Θ in Tables 11–12.

Noting the standard deviations of the estimators $\tilde{\alpha}_3$ and $\check{\alpha}_3$ in the case that $\epsilon = 0.05$ and $n = 1000$, there exist the differences between the standard deviations of the estimator $\tilde{\alpha}_3$ in Tables 8 and 11 or those of the estimator $\check{\alpha}_3$ in Tables 9 and 12.

This is because the initial Bayes type estimator in Table 11 fails to estimate the parameter α_3^* for some cases, and so does the hybrid estimator in Table 12.

On the other hand, the standard deviations of the initial Bayes type estimator $\tilde{\alpha}_3$ in Table 13 are close to those of Table 8 and
the standard deviations of the hybrid estimator $\check{\alpha}_3$ in Table 14 are very close to those of Table 9 when $r_1 = 0.2$ or 0.3.

This means that the initial Bayes type estimator obtained by adjusting r_1 in Table 13 has good behavior and the hybrid estimator with $r_1 = 0.2$ or 0.3 and $r_2 = 1$ in Table 14 is very good as well as the adaptive ML type estimator with initial value being the true value in Table 13.

In Table 15, for all $r_2 = 0.2, \ldots, 0.8$, the intial Bayes type estimator with $r_1 = 0.3$ are worse than the best estimator (the hybrid estimator with $r_1 = 0.3$, $r_2 = 1$).

$$
\begin{aligned}
dX_t^\epsilon &= (\alpha_1 - \alpha_2 X_t^\epsilon - 2\sin(\alpha_3 X_t^\epsilon))\,dt + \epsilon\frac{\beta_2 + (X_t^\epsilon)^2}{1+\beta_1(X_t^\epsilon)^2}dw_t \quad t\in[0,1], \quad \epsilon\in(0,1],\\
X_0 &= 2.
\end{aligned}
$$

Table 7. adaptive ML type estimator $\widehat{\theta}$ with the initial value being the true value

ϵ	n	$\widehat{\alpha}_1$	$\widehat{\alpha}_2$	$\widehat{\alpha}_3$	$\widehat{\beta}_1$	$\widehat{\beta}_2$
		3.0	7.0	5.0	0.5	5.0
0.1	100	3.1522 (0.98947)	7.1602 (1.00240)	4.8413 (0.65782)	0.5835 (0.24111)	4.8717 (0.60907)
	1000	3.2158 (1.02660)	7.3645 (1.05284)	4.9263 (0.64899)	0.5060 (0.06575)	4.9793 (0.18120)
0.05	100	2.9804 (0.42147)	6.8725 (0.46175)	4.8843 (0.18826)	0.5633 (0.21750)	4.8352 (0.50216)
	1000	3.0204 (0.42371)	7.0451 (0.47758)	4.9827 (0.18772)	0.5046 (0.06437)	4.9779 (0.17050)
0.01	100	2.9587 (0.08148)	6.8171 (0.09088)	4.8928 (0.03566)	0.5212 (0.20404)	4.7987 (0.47476)
	1000	2.9935 (0.08183)	6.9818 (0.09426)	4.9893 (0.03540)	0.5042 (0.06365)	4.9774 (0.16647)

$$
\begin{aligned}
dX_t^\epsilon &= \left(\alpha_1 - \alpha_2 X_t^\epsilon - 2\sin\left(\alpha_3 X_t^\epsilon\right)\right) dt + \epsilon \frac{\beta_2 + (X_t^\epsilon)^2}{1 + \beta_1 (X_t^\epsilon)^2} dw_t \quad t \in [0,1], \quad \epsilon \in (0,1], \\
X_0 &= 2.
\end{aligned}
$$

Table 8. initial Bayes type estimators $\tilde{\alpha}^{(1)}$ and $\tilde{\beta}^{(2)}$ with the initial value being the true value

ϵ	n	$\tilde{\alpha}_1^{(1)}$	$\tilde{\alpha}_2^{(1)}$	$\tilde{\alpha}_3^{(1)}$	$\tilde{\beta}_1^{(2)}$	$\tilde{\beta}_2^{(2)}$
		3.0	7.0	5.0	0.5	5.0
0.1	100	3.1649 (1.20757)	7.1480 (1.11918)	4.7996 (1.48767)	0.6732 (0.29173)	5.1323 (0.76265)
	1000	3.2260 (1.24486)	7.3494 (1.17456)	4.8652 (1.42835)	0.5125 (0.06691)	5.0004 (0.18440)
0.05	100	2.9846 (0.43101)	6.8875 (0.47242)	4.8773 (0.24526)	0.6373 (0.23818)	5.0509 (0.54857)
	1000	3.0209 (0.43633)	7.0544 (0.49534)	4.9753 (0.19661)	0.5104 (0.06483)	4.9963 (0.17158)
0.01	100	2.9610 (0.08270)	6.8225 (0.09308)	4.8978 (0.03647)	0.5884 (0.21740)	4.9974 (0.50036)
	1000	2.9931 (0.08406)	6.9819 (0.09780)	4.9901 (0.03677)	0.5098 (0.06414)	4.9952 (0.16728)

$$\begin{aligned} dX_t^\epsilon &= \left(\alpha_1 - \alpha_2 X_t^\epsilon - 2\sin\left(\alpha_3 X_t^\epsilon\right)\right) dt + \epsilon \frac{\beta_2 + (X_t^\epsilon)^2}{1 + \beta_1 (X_t^\epsilon)^2} dw_t \quad t \in [0,1], \quad \epsilon \in (0,1], \\ X_0 &= 2. \end{aligned}$$

Table 9. hybrid estimator $\check{\theta}$ with the initial value being the initial Bayes type estimators $\tilde{\alpha}^{(1)}$ and $\tilde{\beta}^{(2)}$ in Table 8

ϵ	n	$\check{\alpha}_1$	$\check{\alpha}_2$	$\check{\alpha}_3$	$\check{\beta}_1$	$\check{\beta}_2$
		3.0	7.0	5.0	0.5	5.0
0.1	100	3.1550 (1.19584)	7.1218 (1.10431)	4.8144 (1.53128)	0.6176 (0.27304)	4.9223 (0.66270)
	1000	3.2141 (1.17337)	7.3304 (1.11897)	4.8795 (1.45665)	0.5081 (0.06666)	4.9827 (0.18295)
0.05	100	2.9797 (0.42335)	6.8707 (0.46139)	4.8786 (0.24277)	0.5958 (0.23516)	4.8868 (0.52298)
	1000	3.0205 (0.42371)	7.0452 (0.47757)	4.9827 (0.18772)	0.5069 (0.06472)	4.9818 (0.17092)
0.01	100	2.9582 (0.08138)	6.8159 (0.09071)	4.8916 (0.03542)	0.5525 (0.22260)	4.8489 (0.49509)
	1000	2.9935 (0.08183)	6.9818 (0.09426)	4.9893 (0.03540)	0.5065 (0.06409)	4.9812 (0.16688)

$$
\begin{aligned}
dX_t^\epsilon &= (\alpha_1 - \alpha_2 X_t^\epsilon - 2\sin(\alpha_3 X_t^\epsilon))\,dt + \epsilon \frac{\beta_2 + (X_t^\epsilon)^2}{1+\beta_1 (X_t^\epsilon)^2} dw_t \quad t \in [0,1], \quad \epsilon \in (0,1], \\
X_0 &= 2.
\end{aligned}
$$

Table 10. adaptive ML type estimator $\hat{\theta}$ with the initial value being the random value

ϵ	n	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\beta}_1$	$\hat{\beta}_2$
		3.0	7.0	5.0	0.5	5.0
0.1	100	2.6085 (1.37309)	6.5868 (1.26410)	9.3934 (5.99408)	0.5274 (0.21488)	4.9402 (0.53280)
	1000	2.7294 (1.43034)	6.8550 (1.26317)	9.3538 (6.00919)	0.5005 (0.06586)	4.9864 (0.18174)
0.05	100	2.0566 (0.85040)	5.8877 (1.01678)	9.2164 (5.18264)	0.3671 (0.28195)	5.1388 (0.70839)
	1000	2.2951 (1.00509)	6.3334 (1.13409)	9.0811 (5.91092)	0.4701 (0.24035)	4.9944 (0.56477)
0.01	100	2.0081 (0.98288)	5.7267 (1.12613)	8.4996 (5.28508)	0.2530 (0.29049)	10.4713 (5.67697)
	1000	1.9570 (1.06001)	5.7185 (1.29344)	8.7171 (5.27892)	0.2544 (0.24249)	5.6897 (0.76937)

The adaptive estimator is quite unstable.

The optimization of QLL fails because of the wrong initial value.

$$
\begin{aligned}
dX_t^\epsilon &= (\alpha_1 - \alpha_2 X_t^\epsilon - 2\sin(\alpha_3 X_t^\epsilon))\,dt + \epsilon \frac{\beta_2 + (X_t^\epsilon)^2}{1+\beta_1 (X_t^\epsilon)^2} dw_t \quad t \in [0,1], \quad \epsilon \in (0,1], \\
X_0 &= 2.
\end{aligned}
$$

Table 11. initial Bayes type estimators $\tilde{\alpha}^{(1)}$ and $\tilde{\beta}^{(2)}$ with the initial value being the random value

ϵ	n	$\tilde{\alpha}_1^{(1)}$	$\tilde{\alpha}_2^{(1)}$	$\tilde{\alpha}_3^{(1)}$	$\tilde{\beta}_1^{(2)}$	$\tilde{\beta}_2^{(2)}$
		3.0	7.0	5.0	0.5	5.0
0.1	100	3.0860 (1.27922)	7.0710 (1.17129)	5.3797 (2.99609)	0.6693 (0.29384)	5.1246 (0.75495)
	1000	3.1539 (1.31686)	7.2856 (1.21481)	5.4848 (2.99470)	0.5123 (0.06677)	4.9996 (0.18449)
0.05	100	2.9674 (0.46493)	6.8703 (0.50537)	4.9606 (1.10689)	0.6348 (0.24042)	5.0523 (0.54990)
	1000	3.0052 (0.47404)	7.0380 (0.52699)	5.0437 (1.06207)	0.5100 (0.06509)	4.9965 (0.17177)
0.01	100	2.9565 (0.12124)	6.8174 (0.13517)	4.9236 (0.47845)	0.5877 (0.21938)	5.0364 (0.83245)
	1000	2.9869 (0.13426)	6.9748 (0.15316)	5.0244 (0.54853)	0.5081 (0.07048)	5.0015 (0.19432)

Compared with the adaptive estimator in Table 10, the initial Bayes estimator in Table 11 is stable.

$$dX_t^\epsilon = (\alpha_1 - \alpha_2 X_t^\epsilon - 2\sin(\alpha_3 X_t^\epsilon))\,dt + \epsilon \frac{\beta_2 + (X_t^\epsilon)^2}{1+\beta_1 (X_t^\epsilon)^2} dw_t \quad t \in [0,1], \quad \epsilon \in (0,1],$$
$$X_0 = 2.$$

Table 12. hybrid estimator $\check{\theta}$ with the initial value being the initial Bayes type estimators $\tilde{\alpha}^{(1)}$ and $\tilde{\beta}^{(2)}$ in Table 11

ϵ	n	$\check{\alpha}_1$	$\check{\alpha}_2$	$\check{\alpha}_3$	$\check{\beta}_1$	$\check{\beta}_2$
		3.0	7.0	5.0	0.5	5.0
0.1	100	3.0989 (1.27256)	7.0785 (1.15228)	5.3928 (3.00612)	0.6157 (0.27360)	4.9194 (0.66030)
	1000	3.1639 (1.30059)	7.2916 (1.18150)	5.5051 (2.99580)	0.5080 (0.06659)	4.9822 (0.18311)
0.05	100	2.9628 (0.45765)	6.8543 (0.49362)	4.9617 (1.10814)	0.5935 (0.23745)	4.8885 (0.52359)
	1000	3.0068 (0.45617)	7.0332 (0.49930)	5.0506 (1.05550)	0.5066 (0.06494)	4.9821 (0.17108)
0.01	100	2.9535 (0.12169)	6.8106 (0.13577)	4.9175 (0.47933)	0.5515 (0.22425)	4.8856 (0.81001)
	1000	2.9862 (0.14533)	6.9729 (0.17245)	5.0241 (0.55603)	0.5047 (0.07110)	4.9871 (0.18999)

The initial value is quite important for optimization of QLL.

Unfortunately, the s.d. of the hybrid estimator in Table 12 is larger than the one of the adaptive estimator in Table 7.

Initial Bayes type estimator with tuning parameters r_1 and r_2

The initial Bayes type estimators $\tilde{\alpha}^{(1)}_{\epsilon,n,r_1}$ and $\tilde{\beta}^{(2)}_{\epsilon,n,r_2}$ are defined by

$$\tilde{\alpha}^{(1)}_{\epsilon,n,r_1} = \frac{\int_{\Theta_\alpha} \alpha \exp\left\{\epsilon^{2-2r_1} U^{(1)}_{\epsilon,n}(\alpha)\right\} \pi_1(\alpha) d\alpha}{\int_{\Theta_\alpha} \exp\left\{\epsilon^{2-2r_1} U^{(1)}_{\epsilon,n}(\alpha)\right\} \pi_1(\alpha) d\alpha},$$

$$\tilde{\beta}^{(2)}_{\epsilon,n,r_2} = \frac{\int_{\Theta_\beta} \beta \exp\left\{\frac{1}{(\sqrt{n})^{2-2r_2}} U^{(2)}_{\epsilon,n}(\tilde{\alpha}^{(1)}_{\epsilon,n,r_1}, \beta)\right\} \pi_2(\beta) d\beta}{\int_{\Theta_\beta} \exp\left\{\frac{1}{(\sqrt{n})^{2-2r_2}} U^{(2)}_{\epsilon,n}(\tilde{\alpha}^{(1)}_{\epsilon,n,r_1}, \beta)\right\} \pi_2(\beta) d\beta},$$

where

$$U^{(1)}_{\epsilon,n}(\alpha) = -\frac{1}{2} \sum_{i=1}^{n} \left|\Delta X_i - h_n a_{i-1}(\alpha)\right|^2 (\epsilon^2 h_n)^{-1},$$

$$U^{(2)}_{\epsilon,n}(\alpha,\beta) = -\frac{1}{2} \sum_{i=1}^{n} \left\{\log\det B_{i-1}(\beta) + (\epsilon^2 h_n)^{-1} B^{-1}_{i-1}(\beta) \left[(\Delta X_i - h_n a_{i-1}(\alpha))^{\otimes 2}\right]\right\}.$$

$$dX_t^\epsilon = (\alpha_1 - \alpha_2 X_t^\epsilon - 2\sin(\alpha_3 X_t^\epsilon))\,dt + \epsilon \frac{\beta_2 + (X_t^\epsilon)^2}{1+\beta_1 (X_t^\epsilon)^2} dw_t \quad t \in [0,1], \quad \epsilon \in (0,1],$$
$$X_0 = 2.$$

Table 13. initial Bayes type estimators $\tilde{\alpha}^{(1)}$ and $\tilde{\beta}^{(2)}$ with $r_2 = 1$, $\epsilon = 0.05$ and $n = 1000$

r_1	$\tilde{\alpha}_1^{(1)}$	$\tilde{\alpha}_2^{(1)}$	$\tilde{\alpha}_3^{(1)}$	$\tilde{\beta}_1^{(2)}$	$\tilde{\beta}_2^{(2)}$
	3.0	7.0	5.0	0.5	5.0
0.1	2.9386 (0.38420)	6.7986 (0.40359)	5.7185 (0.41996)	0.4730 (0.06666)	4.9692 (0.17111)
0.2	2.9787 (0.41936)	6.9292 (0.45566)	5.0128 (0.32924)	0.5051 (0.06572)	4.9913 (0.17195)
0.3	3.0141 (0.43359)	7.0188 (0.48016)	4.9092 (0.28671)	0.5075 (0.06498)	4.9934 (0.17171)
0.4	3.0181 (0.44859)	7.0421 (0.50395)	4.9585 (0.51720)	0.5089 (0.06491)	4.9948 (0.17162)
0.5	3.0107 (0.46482)	7.0396 (0.51930)	5.0079 (0.88014)	0.5094 (0.06491)	4.9957 (0.17151)
0.6	3.0049 (0.47892)	7.0365 (0.53252)	5.0423 (1.00453)	0.5096 (0.06507)	4.9960 (0.17140)
0.7	3.0026 (0.47721)	7.0349 (0.53111)	5.0508 (1.08994)	0.5098 (0.06512)	4.9962 (0.17148)
0.8	3.0020 (0.48074)	7.0341 (0.53613)	5.0581 (1.15824)	0.5097 (0.06507)	4.9961 (0.17145)
0.9	3.0064 (0.46961)	7.0394 (0.52317)	5.0264 (0.92606)	0.5101 (0.06500)	4.9964 (0.17166)
1	3.0052 (0.47404)	7.0380 (0.52699)	5.0437 (1.06207)	0.5100 (0.06509)	4.9965 (0.17177)

In this example, the initial Bayes estimators with the tuning parameters $r_1 = 0.2$ or 0.3 and $r_2 = 1$ have good behavior.

$$\begin{aligned} dX_t^\epsilon &= (\alpha_1 - \alpha_2 X_t^\epsilon - 2\sin(\alpha_3 X_t^\epsilon))\,dt + \epsilon \frac{\beta_2 + (X_t^\epsilon)^2}{1+\beta_1 (X_t^\epsilon)^2} dw_t \quad t \in [0,1], \quad \epsilon \in (0,1], \\ X_0 &= 2. \end{aligned}$$

Table 14. hybrid estimator $\check{\theta}$ with the initial value being the initial Bayes type estimators $\tilde{\alpha}^{(1)}$ and $\tilde{\beta}^{(2)}$ with $r_2 = 1$, $\epsilon = 0.05$ and $n = 1000$

r_1	$\check{\alpha}_1$	$\check{\alpha}_2$	$\check{\alpha}_3$	$\check{\beta}_1$	$\check{\beta}_2$
	3.0	7.0	5.0	0.5	5.0
0.1	3.0146 (0.43261)	7.0397 (0.48466)	4.9777 (0.31640)	0.5067 (0.06482)	4.9815 (0.17108)
0.2	3.0202 (0.42416)	7.0452 (0.47805)	4.9812 (0.19371)	0.5068 (0.06472)	4.9817 (0.17093)
0.3	3.0212 (0.42378)	7.0464 (0.47754)	4.9806 (0.19390)	0.5068 (0.06472)	4.9817 (0.17093)
0.4	3.0165 (0.43269)	7.0426 (0.48206)	4.9895 (0.49660)	0.5068 (0.06471)	4.9817 (0.17093)
0.5	3.0088 (0.45058)	7.0354 (0.49496)	5.0230 (0.88463)	0.5066 (0.06484)	4.9818 (0.17091)
0.6	3.0039 (0.46056)	7.0312 (0.50223)	5.0501 (0.99592)	0.5066 (0.06490)	4.9820 (0.17088)
0.7	3.0033 (0.45931)	7.0303 (0.50169)	5.0590 (1.09509)	0.5065 (0.06501)	4.9820 (0.17084)
0.8	3.0039 (0.46011)	7.0302 (0.50334)	5.0651 (1.15122)	0.5065 (0.06491)	4.9821 (0.17110)
0.9	3.0074 (0.45249)	7.0340 (0.49633)	5.0337 (0.92607)	0.5067 (0.06485)	4.9820 (0.17090)
1	3.0068 (0.45617)	7.0332 (0.49930)	5.0506 (1.05550)	0.5066 (0.06494)	4.9821 (0.17108)

The s.d. of the hybrid estimator is smaller than the one of the initial Bayes estimator in Table 13.

It is worth mentioning that the hybrid estimators with $r_1 = 0.2$ or $r_2 = 0.3$ has good performance similarly as the adaptive estimator in Table 7.

$$
\begin{aligned}
dX_t^\epsilon &= (\alpha_1 - \alpha_2 X_t^\epsilon - 2\sin(\alpha_3 X_t^\epsilon))\,dt + \epsilon\frac{\beta_2 + (X_t^\epsilon)^2}{1+\beta_1 (X_t^\epsilon)^2} dw_t \quad t\in[0,1], \quad \epsilon\in(0,1], \\
X_0 &= 2.
\end{aligned}
$$

Table 15. initial Bayes type estimators $\tilde{\alpha}^{(1)}$ with $r_1 = 1$ and $\tilde{\beta}^{(2)}$ for $\epsilon = 0.05$ and $n = 1000$

r_2	$\tilde{\alpha}_1^{(1)}$	$\tilde{\alpha}_2^{(1)}$	$\tilde{\alpha}_3^{(1)}$	$\tilde{\beta}_1^{(2)}$	$\tilde{\beta}_2^{(2)}$
	3.0	7.0	5.0	0.5	5.0
0.2	3.0141 (0.43359)	7.0188 (0.48016)	4.9092 (0.28671)	2.6938 (0.22217)	11.4409 (0.49244)
0.4	3.0141 (0.43359)	7.0188 (0.48016)	4.9092 (0.28671)	1.1431 (0.19775)	7.0083 (0.65410)
0.6	3.0141 (0.43359)	7.0188 (0.48016)	4.9092 (0.28671)	0.61550 (0.07644)	5.3287 (0.21264)
0.8	3.0141 (0.43359)	7.0188 (0.48016)	4.9092 (0.28671)	0.5263 (0.06655)	5.0531 (0.17611)
1	3.0141 (0.43359)	7.0188 (0.48016)	4.9092 (0.28671)	0.5075 (0.06498)	4.9934 (0.17171)

$$
\begin{aligned}
dX_t^\epsilon &= (\alpha_1 - \alpha_2 X_t^\epsilon - 2\sin(\alpha_3 X_t^\epsilon))\, dt + \epsilon \frac{\beta_2 + (X_t^\epsilon)^2}{1 + \beta_1 (X_t^\epsilon)^2} dw_t \quad t \in [0,1], \quad \epsilon \in (0,1], \\
X_0 &= 2.
\end{aligned}
$$

Table 16. hybrid estimator $\check{\theta}$ with the initial value being the initial Bayes type estimators $\tilde{\alpha}^{(1)}$ with $r_1 = 1$ and $\tilde{\beta}^{(2)}$ for $\epsilon = 0.05$ and $n = 1000$

r_2	$\check{\alpha}_1$	$\check{\alpha}_2$	$\check{\alpha}_3$	$\check{\beta}_1$	$\check{\beta}_2$
	3.0	7.0	5.0	0.5	5.0
0.2	3.0189 (0.55507)	7.0411 (0.59683)	4.9840 (0.25308)	0.5100 (0.06470)	4.9909 (0.17071)
0.4	3.0344 (0.44984)	7.0540 (0.50571)	4.9810 (0.20451)	0.5085 (0.06467)	4.9856 (0.17086)
0.6	3.0246 (0.42519)	7.0484 (0.47874)	4.9809 (0.19494)	0.5073 (0.06469)	4.9827 (0.17090)
0.8	3.0218 (0.42384)	7.0467 (0.47751)	4.9807 (0.19401)	0.5069 (0.06471)	4.9819 (0.17092)
1	3.0212 (0.42378)	7.0464 (0.47754)	4.9806 (0.19390)	0.5068 (0.06472)	4.9817 (0.17093)

3. Application to adaptive estimator for ergodic diffusions
(This is a joint work with Yuto Yoshida.)

We consider a d-dimensional ergodic diffusion process defined by the following stochastic differential equation

$$dX_t \quad = \quad b(X_t, \beta)dt + a(X_t, \alpha)dw_t, \qquad X_0 = x_0 \tag{2}$$

where w is an r-dimensional standard Wiener process,

x_0 is a deterministic initial condition,

$\theta = (\alpha, \beta) \in \Theta_\alpha \times \Theta_\beta = \Theta$ with Θ_α and Θ_β being compact convex subsets of $\mathbf{R}^{p_1}$ and $\mathbf{R}^{q_1}$, respectively.

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

$\theta^* = (\alpha^*, \beta^*)$ is the true value of θ and we assume that $\theta^* \in \mathrm{Int}(\Theta)$ and the parameter spaces have locally Lipschitz boundaries, see Adams and Fournier (2003).

The data are discrete observations $\mathbf{X}_n = (X_{t_i^n})_{0 \le i \le n}$ with $t_i^n = ih_n$ and $t_n^n = nh_n = T_n$.

We will consider the situation when $h_n \to 0$ and $nh_n^p \to 0$ as $n \to \infty$, and there exists $\epsilon_0 \in (0, (p-1)/p)$ such that $n^{\epsilon_0} \le nh_n$ for large n.

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, and their derivatives up to order k are of polynomial growth in x uniformly in 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.

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.

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

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

Set $A(x,\alpha) = aa^\star(x,\alpha)$, where $\star$ denotes the transpose. Let $\Delta X_i = X_{t_i^n} - X_{t_{i-1}^n}$, $A_{i-1}(\alpha) = A(X_{t_{i-1}^n},\alpha)$ and $b_{i-1}(\beta) = b(X_{t_{i-1}^n},\beta)$.

We make the following assumptions.

$[C1]$ (i) There exists $K > 0$ such that for all $x, y \in \mathbf{R}^d$,

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

(ii) $\inf_{x,\alpha} \det(A(x,\alpha)) > 0$.
(iii) There exists a unique invariant probability measure μ_{θ^*} of X_t and for any $f \in \mathcal{F}_\uparrow(\mathbf{R}^d)$ satisfying $\int_{\mathbf{R}^d} |f(x)|\mu_{\theta^*}(dx) < \infty$, as $T \to \infty$,

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

(iv) $\sup_t E[|X_t|^M] < \infty$ for all $M > 0$.
(v) For any $g \in \mathcal{F}_\uparrow(\mathbf{R}^d)$ satisfying $\int_{\mathbf{R}^d} g(x)\mu_{\theta^*}(dx) = 0$, there exists $G(x)$, $\partial_{x_i}G(x) \in \mathcal{F}_\uparrow(\mathbf{R}^d)$ $(i = 1,\ldots,d)$ such that for all x,

$$L_{\theta^*}G(x) = -g(x).$$

(vi) $\Gamma_\alpha(\theta^*) = (\Gamma_\alpha(\theta^*)_{i,j})_{i,j=1,\ldots,p_1}$ and $\Gamma_\beta(\theta^*) = (\Gamma_\beta(\theta^*)_{i,j})_{i,j=1,\ldots,q_1}$ are invertible, where

$$\begin{aligned}\Gamma_\alpha(\theta^*)_{ij} &= \frac{1}{2}\int_{\mathbf{R}^d} tr\{A^{-1}(\partial_{\alpha_i}A)A^{-1}(\partial_{\alpha_j}A)(x,\alpha^*)\}\mu_{\theta^*}(dx),\\ \Gamma_\beta(\theta^*)_{ij} &= \int_{\mathbf{R}^d} (\partial_{\beta_i}b(x,\beta^*))^\star A(x,\alpha^*)^{-1}\partial_{\beta_j}b(x,\beta^*)\mu_{\theta^*}(dx).\end{aligned}$$

$[C2](k,l)$ $b \in C_\uparrow^{k,4}(\mathbf{R}^d \times \Theta_\beta; \mathbf{R}^d)$. $a \in C_\uparrow^{l,4}(\mathbf{R}^d \times \Theta_\alpha; \mathbf{R}^d \otimes \mathbf{R}^r)$.

Let

$$I(\theta^*) = \begin{pmatrix} \Gamma_\alpha(\theta^*) & 0 \\ 0 & \Gamma_\beta(\theta^*) \end{pmatrix}$$

and

$$\begin{aligned}
\mathbb{Y}(\alpha) &= -\frac{1}{2}\int_{\mathbf{R}^d} \left\{ tr\left[A(x,\alpha)^{-1}A(x,\alpha^*) - I_d\right] + \log\frac{\det(A(x,\alpha))}{\det(A(x,\alpha^*))} \right\} \mu_{\theta^*}(dx), \\
\tilde{\mathbb{Y}}(\beta) &= -\frac{1}{2}\int_{\mathbf{R}^d} A(x,\alpha^*)^{-1}[(b(x,\beta)-b(x,\beta^*))^{\otimes 2}]\mu_{\theta^*}(dx).
\end{aligned}$$

Moreover, we make the following assumptions.

$[C3]$ There exists a positive constant χ such that $\mathbb{Y}(\alpha) \le -\chi|\alpha-\alpha^*|^2$ for all $\alpha \in \Theta_\alpha$.

$[C4]$ There exists a positive constant $\tilde{\chi}$ such that $\tilde{\mathbb{Y}}(\beta) \le -\tilde{\chi}|\beta-\beta^*|^2$ for all $\beta \in \Theta_\beta$.

Remark 2 *(i) For a sufficient condition for* $[A1]$-(v), *see Pardoux and Veretennikov (2001). For example, in addition to* $[A1]$-(i)-(ii), *we assume that* $\sup_{x,\alpha}|A(x,\alpha)| < \infty$ *and that there exist* $c_0 > 0$ *and* $M_0 > 0$ *such that for all* θ_2,

$$\frac{x^* b(x,\beta)}{|x|} \le -c_0 \qquad \text{for all } x \text{ satisfying } |x| \ge M_0.$$

Then, $[A1]$-(v) *holds with* $[A1]$-(iii)-(iv).
(ii) Let $\epsilon_1 = \epsilon_0/(2(p-1))$ *and* $f \in C^{1,1}_{\uparrow}(\mathbf{R}^d \times \Theta)$. *Under* $[A1]$,

$$\sup_{n\in\mathbf{N}} E_{\theta^*}\left[\sup_{\theta\in\Theta}\left(n^{\epsilon_1}\left|\frac{1}{n}\sum_{i=1}^n f(X_{t^n_{i-1}},\theta) - \int_{\mathbf{R}^d} f(x,\theta)\mu_{\theta^*}(dx)\right|\right)^M\right] < \infty$$

for all $M > 0$, *see Uchida (2010).*

Type III adaptive ML estimator for ergodic diffusion processes

In order to get the type III adaptive ML estimator (Kessler (1995), and U and Yoshida (2012)), we consider the following QLL. Let $k_0 = [p/2]$ and note that $nh_n^p \to 0$. Set $\overline{\theta} = (\overline{\alpha}, \overline{\beta})$.

$$\begin{aligned}
\mathcal{V}_n^{(1)}(\alpha) &= -\frac{1}{2}\sum_{i=1}^n \log\det(A_{i-1}(\alpha)) - \frac{1}{2h_n}\sum_{i=1}^n A_{i-1}^{-1}(\alpha)[(X_{t_i^n} - X_{t_{i-1}^n})^{\otimes 2}]. \\
\mathcal{V}_n^{(2)}(\beta, \overline{\alpha}) &= -\frac{1}{2}\sum_{i=1}^n h_n^{-1} A_{i-1}^{-1}(\overline{\alpha})[(X_{t_i^n} - X_{t_{i-1}^n} - h_n b_{i-1}(\beta))^{\otimes 2}],
\end{aligned}$$

and for $k = 1, \ldots, k_0$,

$$\begin{aligned}
\mathcal{V}_n^{(2k+1)}(\alpha \mid \overline{\theta}) &= -\frac{1}{2}\sum_{i=1}^n \left\{ h_n^{-1} A_{i-1}^{-1}(\alpha) \left[(X_{t_i^n} - X_{t_{i-1}^n})^{\otimes 2} - \sum_{j=2}^{k+1} h_n^j \overline{D}_{i-1}^{(j)}(\overline{\theta}) \right] + \log\det A_{i-1}(\alpha) \right\}, \\
\mathcal{V}_n^{(2k+2)}(\beta \mid \overline{\theta}) &= -\frac{1}{2}\sum_{i=1}^n h_n^{-1} A_{i-1}^{-1}(\overline{\alpha}) \left[\left(X_{t_i^n} - X_{t_{i-1}^n} - h_n b_{i-1}(\beta) - \sum_{j=2}^{k+1} h_n^j \overline{r}_{i-1}^{(j)}(\overline{\theta}) \right)^{\otimes 2} \right],
\end{aligned}$$

where for $l, q = 1, \ldots, d$, $f_l(x) = x_l$, $h_{lq}(x) = (x - X_{t_{i-1}^n})_l (x - X_{t_{i-1}^n})_q$,

$$\overline{D}_{i-1}^{(j)}(\overline{\theta})_{lq} = \frac{1}{j!} L_{\overline{\theta}}^j h_{lq}(X_{t_{i-1}^n}), \quad \overline{r}_{i-1}^{(j)}(\overline{\theta})_l = \frac{1}{j!} L_{\overline{\theta}}^j f_l(X_{t_{i-1}^n}).$$

Let $p \geq 2$, $k_0 = [p/2]$ and $l_0 = [(p-1)/2]$. In the same way as Kessler (1995) and U and Yoshida (2012), the adaptive ML type estimators $\breve{\alpha}_n^{(2k_0-1)}$, $\breve{\beta}_n^{(2k_0)}$, $\breve{\alpha}_n^{(2k_0+1)}$ are defined as follows. For $k = 1, 2, \ldots, k_0$,

$$\begin{aligned}
\mathcal{V}_n^{(1)}(\breve{\alpha}_n^{(1)}) &= \sup_{\alpha \in \Theta_\alpha} \mathcal{V}_n^{(1)}(\alpha), \\
\mathcal{V}_n^{(2k)}(\breve{\beta}_n^{(2k)} \mid \breve{\alpha}_n^{(2k-1)}, \breve{\beta}_{2,n}^{(2k-2)}) &= \sup_{\beta \in \Theta_\beta} \mathcal{V}_n^{(2k)}(\beta \mid \breve{\alpha}_n^{(2k-1)}, \breve{\beta}_n^{(2k-2)}), \\
\mathcal{V}_n^{(2k+1)}(\breve{\alpha}_n^{(2k+1)} \mid \breve{\alpha}_n^{(2k-1)}, \breve{\beta}_n^{(2k)}) &= \sup_{\alpha \in \Theta_\alpha} \mathcal{V}_n^{(2k+1)}(\alpha \mid \breve{\alpha}_n^{(2k-1)}, \breve{\beta}_n^{(2k)}),
\end{aligned}$$

where we define $\breve{\beta}_n^{(0)} = 0$.

Theorem 5 (U and Yoshida (2012)) *Let $p \geq 2$, $k_0 = [p/2]$ and $l_0 = [(p-1)/2]$. Assume $[C1]$, $[C2](2k_0, 2k_0+1)$, $[C3]$, $[C4]$. Then as $nh_n^p \to 0$,*

$$(\sqrt{n}(\breve{\alpha}_n^{(2l_0+1)} - \alpha^*), \sqrt{nh_n}(\breve{\beta}_n^{(2k_0)} - \beta^*)) \xrightarrow{d} (\zeta_1, \zeta_2) \sim N_{p_1+q_1}(0, I^{-1}(\theta^*))$$

and

$$E_{\theta^*}[f(\sqrt{n}(\breve{\alpha}_n^{(2l_0+1)} - \alpha^*), \sqrt{nh_n}(\breve{\beta}_n^{(2k_0)} - \beta^*))] \to \mathbb{E}[f(\zeta_1, \zeta_2)]$$

for all continuous functions f of at most polynomial growth.

Application of adaptive estimation for small diffusions to ergodic diffusions

We consider the following QLL. Let $k_0 = [p/2]$ and note that $nh_n^p \to 0$. Set $\overline{\theta} = (\overline{\alpha}, \overline{\beta})$.

$$\begin{aligned} \mathcal{U}_n^{(1)}(\alpha) &= -\frac{1}{2h_n}\sum_{i=1}^{n} \left|\left|(X_{t_i^n} - X_{t_{i-1}^n})^{\otimes 2} - h_n A_{i-1}(\alpha)\right|\right|^2, \\ \mathcal{U}_n^{(2)}(\beta) &= -\frac{1}{2h_n}\sum_{i=1}^{n} \left|X_{t_i^n} - X_{t_{i-1}^n} - h_n b_{i-1}(\beta)\right|^2, \end{aligned}$$

and for $k = 1, \ldots, k_0$,

$$\begin{aligned} \mathcal{V}_n^{(2k+1)}(\alpha \mid \overline{\theta}) &= -\frac{1}{2}\sum_{i=1}^{n}\left\{h_n^{-1}A_{i-1}^{-1}(\alpha)\left[(X_{t_i^n} - X_{t_{i-1}^n})^{\otimes 2} - \sum_{j=2}^{k+1} h_n^j \overline{D}_{i-1}^{(j)}(\overline{\theta})\right] + \log\det A_{i-1}(\alpha)\right\}, \\ \mathcal{V}_n^{(2k+2)}(\beta \mid \overline{\theta}) &= -\frac{1}{2}\sum_{i=1}^{n} h_n^{-1}A_{i-1}^{-1}(\overline{\alpha})\left[\left(X_{t_i^n} - X_{t_{i-1}^n} - h_n b_{i-1}(\beta) - \sum_{j=2}^{k+1} h_n^j \overline{r}_{i-1}^{(j)}(\overline{\theta})\right)^{\otimes 2}\right], \end{aligned}$$

where for $l, q = 1, \ldots, d$, $f_l(x) = x_l$, $h_{lq}(x) = (x - X_{t_{i-1}^n})_l (x - X_{t_{i-1}^n})_q$,

$$\overline{D}_{i-1}^{(j)}(\overline{\theta})_{lq} = \frac{1}{j!}L_{\overline{\theta}}^j h_{lq}(X_{t_{i-1}^n}), \quad \overline{r}_{i-1}^{(j)}(\overline{\theta})_l = \frac{1}{j!}L_{\overline{\theta}}^j f_l(X_{t_{i-1}^n}).$$

Let $k_0 \in \mathbf{N}$. The new adaptive ML type estimators $\breve{\alpha}_n^{(2k_0-1)}$, $\breve{\beta}_n^{(2k_0)}$, $\breve{\alpha}_n^{(2k_0+1)}$ are defined as follows. For $k = 1, 2, \ldots, k_0$,

$$\begin{aligned}
\mathcal{U}_n^{(1)}(\breve{\alpha}_n^{(1)}) &= \sup_{\alpha \in \Theta_\alpha} \mathcal{U}_n^{(1)}(\alpha), \\
\mathcal{U}_n^{(2)}(\breve{\beta}_n^{(2)}) &= \sup_{\beta_m \in \Theta_\beta} \mathcal{U}_n^{(2)}(\beta), \\
\mathcal{V}_n^{(2k+1)}(\breve{\alpha}_n^{(2k+1)} \mid \breve{\alpha}_n^{(2k-1)}, \breve{\beta}_n^{(2k)}) &= \sup_{\alpha \in \Theta_\alpha} \mathcal{V}_n^{(2k+1)}(\alpha \mid \breve{\alpha}_n^{(2k-1)}, \breve{\beta}_n^{(2k)}), \\
\mathcal{V}_n^{(2k+2)}(\breve{\beta}_n^{(2k+2)} \mid \breve{\alpha}_n^{(2k+1)}, \breve{\beta}_n^{(2k)}) &= \sup_{\beta \in \Theta_\beta} \mathcal{V}_n^{(2k+2)}(\beta \mid \breve{\alpha}_n^{(2k+1)}, \breve{\beta}_n^{(2k)}).
\end{aligned}$$

Theorem 6 *Let $p \geq 2$, $k_0 = [p/2]$ and $l_0 = [(p-1)/2]$. Assume $[C1]$, $[C2](2k_0, 2k_0+1)$, $[C3]$, $[C4]$. Then, as $nh_n^p \to 0$,*

$$(\sqrt{n}(\breve{\alpha}_n^{(2l_0+1)} - \alpha^*), \sqrt{nh_n}(\breve{\beta}_n^{(2k_0)} - \beta^*)) \xrightarrow{d} (\zeta_1, \zeta_2) \sim N_{p_1+q_1}(0, I^{-1}(\theta^*))$$

and

$$E_{\theta^*}[f(\sqrt{n}(\breve{\alpha}_n^{(2l_0+1)} - \alpha^*), \sqrt{nh_n}(\breve{\beta}_n^{(2k_0)} - \beta^*))] \to \mathbb{E}[f(\zeta_1, \zeta_2)]$$

for all continuous functions f of at most polynomial growth.

Hybrid estimation for ergodic diffusions

Let $q = 1/p$ and $r = \min\{1/2, 1/(p-1)\}$.

$$\breve{\alpha}^{(1)}_{p,n} = \frac{\int_{\Theta_\alpha} \alpha \exp\left\{\frac{1}{n^{1-2q}}\mathcal{U}^{(1)}_n(\alpha)\right\}\pi_1(\alpha)d\alpha}{\int_{\Theta_\alpha} \exp\left\{\frac{1}{n^{1-2q}}\mathcal{U}^{(1)}(\alpha)\right\}\pi_1(\alpha)d\alpha},$$

$$\breve{\beta}^{(2)}_{p,n} = \frac{\int_{\Theta_\beta} \beta \exp\left\{\frac{1}{(nh_n)^{1-2r}}\mathcal{U}^{(2)}_n(\beta)\right\}\pi_2(\beta)d\beta}{\int_{\Theta_\beta} \exp\left\{\frac{1}{(nh_n)^{1-2r}}\mathcal{U}^{(2)}_n(\beta)\right\}\pi_2(\beta)d\beta},$$

where

$$\mathcal{U}^{(1)}_n(\alpha) = -\frac{1}{2h_n}\sum_{i=1}^n \left|\left|(X_{t_i^n} - X_{t_{i-1}^n})^{\otimes 2} - h_n A_{i-1}(\alpha)\right|\right|^2$$

$$\mathcal{U}^{(2)}_n(\beta) = -\frac{1}{2h_n}\sum_{i=1}^n \left|X_{t_i^n} - X_{t_{i-1}^n} - h_n b_{i-1}(\beta)\right|^2,$$

Let $k_0 = [p/2]$ and $l_0 = [(p-1)/2]$. The hybrid type estimators $\breve{\alpha}_n^{(2k_0-1)}$, $\breve{\beta}_n^{(2k_0)}$, $\breve{\alpha}_n^{(2k_0+1)}$ are defined as follows. For $k = 1, 2, \ldots, k_0 - 1$,

$$\begin{aligned}
\mathcal{U}_n^{(2k+1)}(\breve{\alpha}_n^{(2k+1)} \mid \breve{\alpha}_n^{(2k-1)}, \breve{\beta}_n^{(2k)}) &= \sup_{\alpha \in \Theta_\alpha} \mathcal{U}_n^{(2k+1)}(\alpha \mid \breve{\alpha}_n^{(2k-1)}, \breve{\beta}_n^{(2k)}), \\
\mathcal{U}_n^{(2k+2)}(\breve{\beta}_n^{(2k+2)} \mid \breve{\alpha}_n^{(2k+1)}, \breve{\beta}_n^{(2k)}) &= \sup_{\beta \in \Theta_\beta} \mathcal{U}_n^{(2k+2)}(\beta \mid \breve{\alpha}_n^{(2k+1)}, \breve{\beta}_n^{(2k)}),
\end{aligned}$$

where

$$\begin{aligned}
\mathcal{U}_n^{(2k+1)}(\alpha \mid \bar{\theta}) &= -\frac{1}{2h_n} \sum_{i=1}^n \left\| (X_{t_i^n} - X_{t_{i-1}^n})^{\otimes 2} - h_n A_{i-1}(\alpha) - \sum_{j=2}^{k+1} h_n^j \overline{D}_{i-1}^{(j)}(\bar{\theta}) \right\|^2, \\
\mathcal{U}_n^{(2k+2)}(\beta \mid \bar{\theta}) &= -\frac{1}{2h_n} \sum_{i=1}^n \left| X_{t_i^n} - X_{t_{i-1}^n} - h_n b_{i-1}(\beta) - \sum_{j=2}^{k+1} h_n^j \bar{r}_{i-1}^{(j)}(\bar{\theta}) \right|^2,
\end{aligned}$$

Finally,

$$\begin{aligned}\mathcal{V}_n^{(2l_0+1)}(\breve{\alpha}_n^{(2l_0+1)} \mid \breve{\alpha}_n^{(2l_0-1)}, \breve{\beta}_n^{(2l_0)}) &= \sup_{\alpha\in\Theta_\alpha} \mathcal{V}_n^{(2l_0+1)}(\alpha \mid \breve{\alpha}_n^{(2l_0-1)}, \breve{\beta}_n^{(2l_0)}),\\ \mathcal{V}_n^{(2k_0)}(\breve{\beta}_n^{(2k_0)} \mid \breve{\alpha}_n^{(2k_0-1)}, \breve{\beta}_n^{(2k_0-2)}) &= \sup_{\beta\in\Theta_\beta} \mathcal{V}_n^{(2k_0)}(\beta \mid \breve{\alpha}_n^{(2k_0-1)}, \breve{\beta}_n^{(2k_0-2)}),\end{aligned}$$

where

$$\begin{aligned}\mathcal{V}_n^{(2l_0+1)}(\alpha \mid \overline{\theta}) &= -\frac{1}{2}\sum_{i=1}^n \left\{ h_n^{-1} A_{i-1}^{-1}(\alpha)\left[(X_{t_i^n} - X_{t_{i-1}^n})^{\otimes 2} - \sum_{j=2}^{l_0+1} h_n^j \overline{D}_{i-1}^{(j)}(\overline{\theta})\right] + \log\det A_{i-1}(\alpha)\right\},\\ \mathcal{V}_n^{(2k_0)}(\beta \mid \overline{\theta}) &= -\frac{1}{2}\sum_{i=1}^n h_n^{-1} A_{i-1}^{-1}(\overline{\alpha})\left[\left(X_{t_i^n} - X_{t_{i-1}^n} - h_n b_{i-1}(\beta) - \sum_{j=2}^{k_0} h_n^j \overline{r}_{i-1}^{(j)}(\overline{\theta})\right)^{\otimes 2}\right],\end{aligned}$$

where for $l, q = 1, \ldots, d$, $f_l(x) = x_l$, $h_{lq}(x) = (x - X_{t_{i-1}^n})_l (x - X_{t_{i-1}^n})_q$,

$$\overline{D}_{i-1}^{(j)}(\overline{\theta})_{lq} = \frac{1}{j!} L_{\overline{\theta}}^j h_{lq}(X_{t_{i-1}^n}), \quad \overline{r}_{i-1}^{(j)}(\overline{\theta})_l = \frac{1}{j!} L_{\overline{\theta}}^j f_l(X_{t_{i-1}^n}).$$

Theorem 7 *Let $p \geq 2$. Assume $[C1]$ and $[C2](2k_0, 2k_0+1)$, $[C3]$, $[C4]$. Then as $nh_n^p \to 0$,*

$$(\sqrt{n}(\breve{\alpha}_n^{(2l_0+1)} - \alpha^*), \sqrt{nh_n}(\breve{\beta}_n^{(2k_0)} - \beta^*)) \xrightarrow{d} (\zeta_1, \zeta_2) \sim N_{p_1+q_1}(0, I^{-1}(\theta^*))$$

and

$$E_{\theta^*}[f(\sqrt{n}(\breve{\alpha}_n^{(2l_0+1)} - \alpha^*), \sqrt{nh_n}(\breve{\beta}_n^{(2k_0)} - \beta^*))] \to \mathbb{E}[f(\zeta_1, \zeta_2)]$$

for all continuous functions f of at most polynomial growth.

Thank you for your kind attention