# Supplementary Material to "Uniform Convergence Rates of Kernel-Based Nonparametric Estimators for Continuous Time Diffusion Processes: A Damping Function Approach"

SHIN KANAYA

AARHUS UNIVERSITY, CREATES, AND IER

JUNE 2016

## S.1 Proof of Lemma 2 (Covering Number of Kernel Transformations)

Terminologies in this proof follow those of van der Vaart and Wellner (1996; hereafter referred to as VW96). First, observe that $K(\cdot)$ can be decomposed into

$$K(z) = K_1(z) - K_2(z), \tag{S.1}$$

since it is of bounded variation, where $K_1$ and $K_2$ are bounded monotone functions. By (S.1) and re-parametrization, we have

$$\mathcal{K} \subset \{ K_1 - K_2 \mid K_1 \in \mathcal{K}_1 \text{ and } K_2 \in \mathcal{K}_2 \}, \tag{S.2}$$

where $\mathcal{K}_l := \{ K_l(ap+b) \mid a \in (0,\infty);\ b \in \mathbb{R} \}$ for $l = 1, 2$. Since $K(\cdot)$ is bounded by $\bar{K}$, there exists a (uniformly) bounded envelope function $M_l(\cdot)$ for $\mathcal{K}_l$ $(l = 1, 2)$.

Note that the set of subgraphs of functions $\{ ap + b \mid a \in (0,\infty);\ b \in \mathbb{R} \}$ is a VC-class with the VC index 3, which is shown by an argument that is similar to the proof of Lemma 2.6.16 of VW96 (the set of subgraphs of functions $\{ ap + b \mid a \in (0,\infty) \text{ and } b \in \mathbb{R} \}$ cannot shatter any three-point set while it can shatter some two-point set in $\mathbb{R}^2$). Then, since $K_l(\cdot)$ is monotone, each $\mathcal{K}_l$ is also a VC-class with the VC-index at most 3, which follows from Lemma 9.9 (viii) of Kosorok (2008).

Using Theorem 2.6.7 of VW96, we can find the upper bound of the uniform covering number of $\mathcal{K}_l$, i.e.,

$$\sup_Q N(\epsilon \|M_l\|_{Q,r}, \mathcal{K}_l, \mathcal{L}_r(Q)) \leq \Lambda_l \epsilon^{-2r} \quad \text{for } \epsilon \in (0,1), \tag{S.3}$$

for some constant $\Lambda_l\,(>0)$ that is independent of $Q$. It then holds that for each $Q$,

$$\begin{aligned} N(\epsilon 4\left\|M_1+M_2\right\|_{Q,r},\mathcal{K},\mathcal{L}_r(Q)) &\leq N(\epsilon 4\left\|M_1+M_2\right\|_{Q,r},\mathcal{K}_1-\mathcal{K}_2,\mathcal{L}_r(Q)) \\ &\leq N(\epsilon\left\|M_1\right\|_{Q,r},\mathcal{K}_1,\mathcal{L}_r(Q))N(\epsilon\left\|M_2\right\|_{Q,r},-\mathcal{K}_2,\mathcal{L}_r(Q)) \\ &= \prod\nolimits_{l=1,2} N(\epsilon\left\|M_l\right\|_{Q,r},\mathcal{K}_l,\mathcal{L}_r(Q)), \end{aligned} \tag{S.4}$$

where the first inequality holds by (S.2); the second holds by Lemma 16 of Nolan and Pollard (1987); and the last equality holds by the fact that $-\mathcal{K}_2$ and $\mathcal{K}_2$ have the same covering number. From (S.3) and (S.4), it holds that

$$\begin{aligned} \sup\nolimits_Q N(\epsilon 4\left\|M_1+M_2\right\|_{Q,r},\mathcal{K},\mathcal{L}_r(Q)) &\leq \prod\nolimits_{l=1,2} \sup\nolimits_Q N(\epsilon\left\|M_l\right\|_{Q,r},\mathcal{K}_l,\mathcal{L}_r(Q)) \\ &\leq \Lambda_1\Lambda_2\epsilon^{-4r}. \end{aligned}$$

Since $4\left\|M_1+M_2\right\|_{Q,r}\leq 8\bar{K}$, we have shown that the inequality (43) holds with some constant $\Lambda=\Lambda_1\Lambda_2$, completing the proof. □

## S.2 Proof of Theorem 5 (Uniform Convergence of the Diffusion Estimator)

Using Ito's lemma:

$$\begin{aligned} \left[X_{(j+1)\Delta}-X_{j\Delta}\right]^2-\Delta\sigma^2(x) &= 2\int_{j\Delta}^{(j+1)\Delta}\left[X_s-X_{j\Delta}\right]\mu(X_s)\,ds \\ +2\int_{j\Delta}^{(j+1)\Delta}\left[X_s-X_{j\Delta}\right]\sigma(X_s)\,dW_s &+ \int_{j\Delta}^{(j+1)\Delta}\left[\sigma^2(X_s)-\sigma^2(x)\right]ds, \end{aligned} \tag{S.5}$$

we have the following decomposition:

$$\sup\nolimits_{x\in\mathbb{R}}|\hat{\Psi}_{\sigma^2}(x)-\hat{\Pi}(x)\,\sigma^2(x)|\leq\sum\nolimits_{i=1}^5\mathcal{V}_i,$$

where

$$\begin{aligned} \mathcal{V}_1 &:= 2\sup_{x\in\mathbb{R}}\left|(1/Th)\sum\nolimits_{j=1}^{n-1}K\left(\tfrac{X_{j\Delta}-x}{h}\right)B(X_{j\Delta})\int_{j\Delta}^{(j+1)\Delta}[X_s-X_{j\Delta}]\mu(X_s)\,ds\right|; \\ \mathcal{V}_2 &:= 2\sup_{x\in\mathbb{R}}\left|(1/Th)\sum\nolimits_{j=1}^{n-1}K\left(\tfrac{X_{j\Delta}-x}{h}\right)B(X_{j\Delta})\int_{j\Delta}^{(j+1)\Delta}[X_s-X_{j\Delta}]\sigma(X_s)\,dW_s\right|; \\ \mathcal{V}_3 &:= \sup_{x\in\mathbb{R}}\left|(1/Th)\sum\nolimits_{j=1}^{n-1}K\left(\tfrac{X_{j\Delta}-x}{h}\right)B(X_{j\Delta})\int_{j\Delta}^{(j+1)\Delta}[\sigma^2(X_s)-\sigma^2(X_{j\Delta})]ds\right|; \\ \mathcal{V}_4 &:= \sup_{x\in\mathbb{R}}\left|\frac{n-1}{nh}\int_{-\infty}^{\infty}K\left(\frac{p-x}{h}\right)B(p)\,\pi(p)\,[\sigma^2(p)-\sigma^2(x)]dp\right|; \\ \mathcal{V}_5 &:= \sup_{x\in\mathbb{R}}(1/nh)\left|\sum\nolimits_{j=1}^{n-1}\{\Gamma_{j\Delta}(x)-E[\Gamma_{j\Delta}(x)]\}\right|; \end{aligned}$$

and

$$\Gamma_{j\Delta}(x) := K\left(\tfrac{X_{j\Delta}-x}{h}\right) B(X_{j\Delta})\left[\sigma^2(X_{j\Delta}) - \sigma^2(x)\right].$$

Below, we verify the following results:

$$\mathcal{V}_1 = O_p(\Delta \log(1/\Delta) + \sqrt{(\log n)/nh}); \tag{S.6}$$

$$\mathcal{V}_2 = O_p(\sqrt{(\log n)/nh}); \tag{S.7}$$

$$\mathcal{V}_3 = O_p(\Delta + \sqrt{(\log n)/nh}); \tag{S.8}$$

$$\mathcal{V}_4 = O(h^2); \tag{S.9}$$

$$\mathcal{V}_5 = O_p(\sqrt{(\log n)/nh}). \tag{S.10}$$

These imply the first part of the theorem. The convergence of $\hat{\mu}(x)$ can be verified in the same way as in the proof of Theorem 8 of Hansen (2008), and hence, its proof is omitted.

**Proof of (S.6).** Since $X_s - X_{j\Delta} = \int_{j\Delta}^{s} \mu(X_u)\, du + \int_{j\Delta}^{s} \sigma(X_u)\, dW_u$, we obtain

$$\begin{aligned}
B(X_{j\Delta})\int_{j\Delta}^{(j+1)\Delta}[X_s - X_{j\Delta}]\,\mu(X_s)\,ds &= \int_{j\Delta}^{(j+1)\Delta} B^{1/2}(X_{j\Delta})\,[X_s - X_{j\Delta}]\,B^{1/2}(X_{j\Delta})[\mu(X_s) - \mu(X_{j\Delta})]ds \\
&\quad + B^{1/2}(X_{j\Delta})\mu(X_{j\Delta})\int_{j\Delta}^{(j+1)\Delta}\int_{j\Delta}^{s} B^{1/2}(X_{j\Delta})\,\mu(X_u)\,du ds \\
&\quad + B(X_{j\Delta})\,\mu(X_{j\Delta})\int_{j\Delta}^{(j+1)\Delta}\int_{j\Delta}^{s}\sigma(X_u)\,dW_u ds.
\end{aligned}$$

By Theorem 1 and (19), the first and second terms on the RHS are $O_{a.s.}(\Delta^2 \log(1/\Delta))$ and $O_{a.s.}(\Delta^2)$ uniformly over $j$, respectively, where we note that $\mu(\cdot)$ is at most of polynomial growth (as assumed in (A3.i)). Therefore,

$$\begin{aligned}
\mathcal{V}_1 &\le 2\sup_{x\in\mathbb{R}}(1/nh)\sum_{j=1}^{n-1}\left|K\left(\tfrac{X_{j\Delta}-x}{h}\right)\right| \times \left[O_{a.s.}(\Delta\log(1/\Delta)) + O_{a.s.}(\Delta)\right] \\
&\quad + 2\sup_{x\in\mathbb{R}}(1/nh)\sum_{j=1}^{n-1}K\left(\tfrac{X_{j\Delta}-x}{h}\right) B(X_{j\Delta})\,\mu(X_{j\Delta})\int_{j\Delta}^{(j+1)\Delta}\int_{j\Delta}^{s}\sigma(X_u)\,dW_u ds \\
&=: 2\mathcal{V}_{11} + 2\mathcal{V}_{12}.
\end{aligned}$$

By the uniform boundedness of $\pi(\cdot)$, we have

$$\sup_{x\in\mathbb{R}}(1/nh)\sum_{j=1}^{n-1}\left|K\left(\tfrac{X_{j\Delta}-x}{h}\right)\right| = O_p(1),$$

which is derived in the proof of Theorem 2 (see the term $R_{11}$), and therefore, $\mathcal{V}_{11} = O_p(\Delta\log(1/\Delta))$. To derive the rate of $\mathcal{V}_{12}$, observe that

$$\int_{j\Delta}^{(j+1)\Delta}\int_{j\Delta}^{s}\sigma\left(X_u\right)dW_u ds = \int_{j\Delta}^{(j+1)\Delta}\int_{u}^{(j+1)\Delta} ds\sigma\left(X_u\right)dW_u$$
$$= \int_{j\Delta}^{(j+1)\Delta}\left[(j+1)\Delta - u\right]\sigma\left(X_u\right)dW_u,$$

which holds by changing the order of (stochastic) integrals. Therefore, $\mathcal{V}_{12}$ can be written as the sum of martingale differences. It can be represented by a continuous martingale with index $r \in [0,1]$ (in the same way as the term $U_3$ in the proof of Theorem 4), and we can show that $\mathcal{V}_{12} = O_p(\sqrt{(\log n)/nh})$ in the same way as $\mathcal{V}_2$ below (and we omit details for brevity).

**Proof of (S.7).** Let

$$\varrho_{s,j\Delta} := B\left(X_{j\Delta}\right)\left[\int_{j\Delta}^{s}\mu\left(X_u\right)du + \int_{j\Delta}^{s}\sigma\left(X_u\right)dW_u\right]\sigma\left(X_s\right);$$
$$\mathbf{e}_\Delta\left(s,j\right) := \mathbf{1}_{\left\{\sup_{v\in[j\Delta,s]} B^{1/2}(X_{j\Delta})\left[\left|\mu(X_v)-\mu(X_{j\Delta})\right|+\left|\sigma(X_v)-\sigma(X_{j\Delta})\right|\right]\leq 1\right\}},$$

where $\mathbf{e}_\Delta$ is an indicator function defined for each $(s,j)$ with $s \in [j\Delta,(j+1)\Delta]$. Using $\mathbf{e}_\Delta$, we also define

$$\begin{aligned}\bar{\varrho}_{s,j\Delta} := &\left\{\int_{j\Delta}^{s}\{B^{1/2}\left(X_{j\Delta}\right)\mu\left(X_{j\Delta}\right) + B^{1/2}\left(X_{j\Delta}\right)\left[\mu\left(X_u\right)-\mu\left(X_{j\Delta}\right)\right]\mathbf{e}_\Delta\left(u,j\right)\}du\right.\\ &\left.+\int_{j\Delta}^{s}\{B^{1/2}\left(X_{j\Delta}\right)\sigma\left(X_{j\Delta}\right) + B^{1/2}\left(X_{j\Delta}\right)\left[\sigma\left(X_u\right)-\sigma\left(X_{j\Delta}\right)\right]\}\mathbf{e}_\Delta\left(u,j\right)dW_u\right\}\\ &\times\left\{B\left(X_{j\Delta}\right)\sigma\left(X_{j\Delta}\right) + B\left(X_{j\Delta}\right)\left[\sigma\left(X_s\right)-\sigma\left(X_{j\Delta}\right)\right]\mathbf{e}_\Delta\left(s,j\right)\right\}.\end{aligned} \tag{S.11}$$

Then, we can write

$$\varrho_{s,j\Delta} = \bar{\varrho}_{s,j\Delta} + \tilde{\varrho}_{s,j\Delta},$$

where $\tilde{\varrho}_{s,j\Delta}$ is defined through the same form as $\bar{\varrho}_{s,j\Delta}$ with $\mathbf{e}_\Delta\left(u,j\right)$ and $\mathbf{e}_\Delta\left(s,j\right)$ replaced by $[1-\mathbf{e}_\Delta\left(u,j\right)]$ and $[1-\mathbf{e}_\Delta\left(s,j\right)]$, respectively. Therefore, we can also write

$$\begin{aligned}\mathcal{V}_2 &= 2\sup_{x\in\mathbb{R}}\left|(1/Th)\sum_{j=1}^{n-1}K\left(\tfrac{X_{j\Delta}-x}{h}\right)\int_{j\Delta}^{(j+1)\Delta}\bar{\varrho}_{s,j\Delta}dW_s\right|\\ &+2\sup_{x\in\mathbb{R}}\left|(1/Th)\sum_{j=1}^{n-1}K\left(\tfrac{X_{j\Delta}-x}{h}\right)\int_{j\Delta}^{(j+1)\Delta}\tilde{\varrho}_{s,j\Delta}dW_s\right|\\ &=: 2\mathcal{V}_{21} + 2\mathcal{V}_{22}.\end{aligned}$$

By Theorem 1, there exists some $\tilde{\Delta} > 0$ such that for any $\Delta \leq \tilde{\Delta}$,

$$\max\nolimits_{1\leq j\leq n-1}\sup\nolimits_{s\in[j\Delta,(j+1)\Delta]}\left|1-\mathbf{e}_\Delta\left(s,j\right)\right| = 0 \quad \text{almost surely,}$$

implying that $\mathcal{V}_{22} = 0$ almost surely for sufficiently small $\Delta$. Therefore, the convergence rate of $\mathcal{V}_2$ is determined by that of $\mathcal{V}_{21}$.

To derive the rate of $\mathcal{V}_{21}$, we note that

$$\bar{\varrho}_{s,j\Delta} \leq C_0 \left\{ \Delta + \int_{j\Delta}^{s} \{B^{1/2}(X_{j\Delta})\, \sigma(X_{j\Delta}) + B^{1/2}(X_{j\Delta})\, [\sigma(X_u) - \sigma(X_{j\Delta})]\} \mathbf{e}_\Delta(u,j)\, dW_u \right\} \tag{S.12}$$

for some constant $C_0 > 0$, which follows from the definition of $\bar{\varrho}_{s,j\Delta}$. Since the integrand of the stochastic integral on the RHS of (S.12) is uniformly bounded over $j$ and $u \in [j\Delta, s]$, we can apply the same argument as those for (15) and (19). That is, we let

$$\bar{\varrho}_{s,j\Delta} = \bar{\varrho}_{s,j\Delta} \mathbf{1}_{\left\{\bar{\varrho}_{s,j\Delta} \leq \sqrt{\Delta} \log \Delta^{-1}\right\}} + \bar{\varrho}_{s,j\Delta} \mathbf{1}_{\left\{\bar{\varrho}_{s,j\Delta} > \sqrt{\Delta} \log \Delta^{-1}\right\}},$$

where the second term is exactly zero for sufficiently small $\Delta$ (uniformly over $j$ and $s \in [j\Delta, (j+1)\Delta]$), implying that the rate of $\mathcal{V}_{21}$ is determined by that of

$$\bar{\mathcal{V}}_{21} := \sup_{x \in \mathbb{R}} \left| (1/Th) \sum_{j=1}^{n-1} K\left(\tfrac{X_{j\Delta} - x}{h}\right) \int_{j\Delta}^{(j+1)\Delta} \bar{\varrho}_{s,j\Delta} \mathbf{1}_{\left\{\bar{\varrho}_{s,j\Delta} \leq \sqrt{\Delta} \log \Delta^{-1}\right\}} dW_s \right|.$$

To derive the rate of $\bar{\mathcal{V}}_{21}$, we consider a finite covering $\{\mathcal{K}_k(h)\}_{k=1}^{\nu(h)}$ of the set of functions:

$$\mathcal{K}(h) := \left\{ K\left(\frac{p - x}{h}\right) \;\middle|\; x \in \mathbb{R} \right\} \quad \text{for each } h > 0.$$

By Lemma 2, we can find $\{\mathcal{K}_k(h)\}_{k=1}^{\nu(h)}$ such that each $\mathcal{K}_k(h)$ has the center $g_k(\cdot) := K\left(\frac{\cdot - x_k}{h}\right)$; for any probability measure $Q$,

$$\forall \varepsilon \in (0,1),\ \forall g \in \mathcal{K}_k(h),\ \{\textstyle\int |g - g_k|^{\bar{r}} dQ\}^{1/\bar{r}} \leq \varepsilon 8 \bar{K}, \text{ and } \nu(h) \leq \Lambda \varepsilon^{-4\bar{r}}, \tag{S.13}$$

for some constant $\Lambda\,(>0)$ (independent of $h$) and for any $\bar{r} \geq 1$. Then,

$$\begin{aligned} \bar{\mathcal{V}}_{21} &\leq \tfrac{n-1}{Th} \max_{k \in \{1,\dots,\upsilon(h)\}} \sup_{g \in \mathcal{K}_k(h)} \tfrac{1}{n-1} \sum_{j=1}^{n-1} |g_k(X_{j\Delta}) - g(X_{j\Delta})| \left| \int_{j\Delta}^{(j+1)\Delta} \bar{\varrho}_{s,j\Delta} \mathbf{1}_{\left\{\bar{\varrho}_{s,j\Delta} \leq \sqrt{\Delta} \log \Delta^{-1}\right\}} dW_s \right| \\ &\quad + (1/Th) \max_{k \in \{1,\dots,\upsilon(h)\}} \left| \sum_{j=1}^{n-1} g_k(X_{j\Delta}) \int_{j\Delta}^{(j+1)\Delta} \bar{\varrho}_{s,j\Delta} \mathbf{1}_{\left\{\bar{\varrho}_{s,j\Delta} \leq \sqrt{\Delta} \log \Delta^{-1}\right\}} dW_s \right| \\ &=: \bar{\mathcal{V}}_{211} + \bar{\mathcal{V}}_{212}. \end{aligned}$$

By the Hölder and Burkholder-Davis-Gundy (BDG) inequalities, we have for any $\bar{r} > 1$,

$$
\begin{aligned}
\bar{\mathcal{V}}_{211} &\le \tfrac{n-1}{Th}\left\{\max_{k\in\{1,\dots,v(h)\}}\sup_{g\in\mathcal{K}_k(h)}\tfrac{1}{n-1}\sum\nolimits_{j=1}^{n-1}\left|g_k\left(X_{j\Delta}\right)-g\left(X_{j\Delta}\right)\right|^{\bar{r}}\right\}^{1/\bar{r}} \\
&\quad\times\left\{\tfrac{1}{n-1}\sum\nolimits_{j=1}^{n-1}\left[\left|\int_{j\Delta}^{(j+1)\Delta}\bar{\varrho}_{s,j\Delta}\mathbf{1}_{\left\{\bar{\varrho}_{s,j\Delta}\le\sqrt{\Delta}\log\Delta^{-1}\right\}}dW_s\right|^{\bar{r}/(\bar{r}-1)}\right\}^{(\bar{r}-1)/\bar{r}} \\
&\le \tfrac{n-1}{Th}\varepsilon 8\bar{K}\times\left\{O_p(\Delta^{\bar{r}/(\bar{r}-1)}(\log\Delta^{-1})^{\bar{r}/(\bar{r}-1)})\right\}^{(\bar{r}-1)/\bar{r}} \\
&= O_p(h^{-1}\varepsilon\log\Delta^{-1}) \\
&= O_p(h^{-1}\varepsilon\log n) = O_p(\sqrt{(\log n)/nh}),
\end{aligned}
\tag{S.14}
$$

where the last two equalities have used the condition $\Delta^{-1} \le n^{\varkappa}$ (implying that $\log\Delta^{-1} = O(\log n)$) and

$$
\varepsilon = \sqrt{h/n(\log n)}. \tag{S.15}
$$

To find the probability bound of the second term $\bar{\mathcal{V}}_{212}$, we note that $\sum_{j=1}^{n-1} g_k\left(X_{j\Delta}\right)\int_{j\Delta}^{(j+1)\Delta}\bar{\varrho}_{s,j\Delta}dW_s$ can be written as a continuous martingale indexed by $r\in[0,1]$ in the same way as in the proof of Theorem 3 (see the expression for the term $U_3$), whose quadratic variation at $r=1$ is given by

$$
J_{n,T}\left(x_k,h\right) = \sum\nolimits_{j=1}^{n-1}K^2\left(\tfrac{X_{j\Delta}-x_k}{h}\right)\int_{j\Delta}^{(j+1)\Delta}\bar{\varrho}^2_{s,j\Delta}\mathbf{1}_{\left\{\bar{\varrho}_{s,j\Delta}\le\sqrt{\Delta}\log\Delta^{-1}\right\}}ds
$$

with $x_k$ being a point in $\mathbb{R}$ satisfying $g_k^2(\cdot) = K^2\left(\frac{\cdot-x_k}{h}\right)$. This $J_{n,T}(k)$ satisfies

$$
J_{n,T}\left(x_k,h\right) \le \bar{K}^2T\Delta[\log(1/\Delta)]^2 \quad \text{uniformly over } k,
$$

as well as

$$
\begin{aligned}
E[J_{n,T}\left(x_k,h\right)] &\le n\max_{1\le j\le n-1}E[K^2\left(\tfrac{X_{j\Delta}-x_k}{h}\right)\bar{\varrho}^2_{s,j\Delta}] \\
&\le C_J Th\Delta[\log(1/\Delta)]^2 \quad \text{uniformly over } k,
\end{aligned}
$$

for some constant $C_J > 0$, which follows from the BDG inequality (with the upper bound (S.12)) and the change-of-variable argument (with the uniform boundedness of $\pi$). Applying the exponential inequality for continuous martingales (Ex. 3.16 in Ch. IV of Revuz and Yor, 1999), for each $a > 0$ and each $y$, we have

$$\Pr\left[\bar{\mathcal{V}}_{212} \geq a\sqrt{(\log n)/nh}\right]$$
$$\leq v(h)\left\{\Pr\left[(1/Th)\left|\sum_{j=1}^{n-1} g_k(X_{j\Delta})\int_{j\Delta}^{(j+1)\Delta}\bar{\varrho}_{s,j\Delta}\mathbf{1}_{\left\{\bar{\varrho}_{s,j\Delta}\leq\sqrt{\Delta}\log\Delta^{-1}\right\}}dW_s\right| \geq a\sqrt{(\log n)/nh},\ J_{n,T}(k)\leq y\right]\right.$$
$$\left.+\Pr\left[|J_{n,T}(x_k,h)-E[J_{n,T}(x_k,h)]|\geq y/2\right]+\Pr\left[E[J_{n,T}(x_k,h)]\geq y/2\right]\right\}$$
$$\leq O(n^{2\bar{r}}h^{-2\bar{r}}(\log n)^{2\bar{r}})\left\{2\exp\left\{-\frac{a^2[(\log n)/nh]T^2h^2}{aTh\Delta}\right\}\right.$$
$$\left.+\Pr\left[|J_{n,T}(x_k,h)-E[J_{n,T}(x_k,h)]|\geq aTh\Delta\right]\right\}$$
$$= O(n^{2\bar{r}}h^{-2\bar{r}}(\log n)^{2\bar{r}}\times n^{-a})$$
$$+O(n^{2\bar{r}}h^{-2\bar{r}}(\log n)^{2\bar{r}})\times\Pr\left[|J_{n,T}(x_k,h)-E[J_{n,T}(x_k,h)]|\geq aTh\Delta\right], \tag{S.16}$$

where the second inequality holds with $y = 2aTh\Delta$ (for sufficiently large $a$) since $v(h) \leq \Lambda\varepsilon^{-4\bar{r}}$, $\varepsilon^{-4\bar{r}} = O\left(n^{2\bar{r}}h^{-2\bar{r}}(\log n)^{2\bar{r}}\right)$ (recall $\varepsilon = \sqrt{h/n(\log n)}$ in (S.15)) and

$$\Pr\left[E\left[J_{n,T}(x_k,h)\right]\geq y/2\right]\leq\Pr\left[C_J Th\Delta\geq aTh\Delta\right]=0.$$

The first term on the RHS of (S.16) approaches zero as $n\to\infty$ (for $n$ large enough). To find the bound of the second term on the RHS of (S.16), we apply the Bernstein-type inequality in Lemma 3. To this end, we write

$$J_{n,T}(x_k,h)-E\left[J_{n,T}(x_k,h)\right]=\sum\nolimits_{j=1}^{n-1}\left\{\mathcal{Z}_{j,n}-E\left[\mathcal{Z}_{j,n}\right]\right\},$$

where

$$\mathcal{Z}_{j,n}:=K^2\left(\tfrac{X_{j\Delta}-x_k}{h}\right)\int_{j\Delta}^{(j+1)\Delta}\bar{\varrho}^2_{s,j\Delta}\mathbf{1}_{\{\bar{\varrho}_{s,j\Delta}\leq\sqrt{\Delta}\log\Delta^{-1}\}}ds.$$

By the boundedness of $K$ and the definition of $\bar{\varrho}_{s,j\Delta}$ in (S.11), we can find some constant $C_{\mathcal{Z}}>0$ such that $\max_{1\leq j\leq n-1}|\mathcal{Z}_{j,n}|\leq C_{\mathcal{Z}}\bar{K}^2\Delta^2\left(\log\Delta^{-1}\right)^2$. By the change-of-variable argument and the BDG inequality, we can also find some constant $\varpi_{\mathcal{Z}}$ satisfying $E[|\sum_{j=1}^{m}\mathcal{Z}_{j,n}|^2]\leq\varpi_{\mathcal{Z}}m^2h\Delta^4$. Given these, we have for $m\leq\min\{aTh\Delta/4\bar{K}^2\Delta^2\left(\log\Delta^{-1}\right)^2, n-1\}$ and for each $a>0$, we have

$$\Pr\left[|J_{n,T}(k)-E\left[J_{n,T}(k)\right]|\geq aTh\Delta\right]$$
$$=\Pr\left[\left|\sum\nolimits_{j=1}^{n-1}\mathcal{Z}_{j,n}\right|\geq aTh\Delta\right]$$
$$\leq 4\exp\left\{-\frac{-a^2T^2h^2\Delta^2}{64n\left(\varpi_{\mathcal{Z}}m\Delta^4h\right)+(4/3)\left[C_{\mathcal{Z}}\bar{K}^2\Delta^2\left(\log\Delta^{-1}\right)^2\right](aTh\Delta)m}\right\}+4Anm^{-1-\beta}\Delta^{-\beta}$$
$$\leq 4\exp\left\{-\frac{-a^2(\log n)}{64\varpi_{\mathcal{Z}}/\left(\log\Delta^{-1}\right)^2+(4/3)C_{\mathcal{Z}}\bar{K}^2a}\right\}+O(h^{-1-\beta}n^{-\beta}(\log n)^{3(1+\beta)}\Delta^{-\beta})$$
$$=O\left(n^{-a}+h^{-1-\beta}n^{-\beta}(\log n)^{3(1+\beta)}\Delta^{-\beta}\right),$$

where we have set $m = nh/(\log n)(\log \Delta^{-1})^2$ and used $\log \Delta^{-1} = O(\log n)$ for the last two lines. Therefore, using $h^{-1} = O(n^{\vartheta}(\log n)^{-1})$ and $\Delta^{-1} = O(n^{\varkappa})$, we can write the second term on the RHS of (S.16) as

$$O(n^{2\bar{r}}h^{-2\bar{r}}(\log n)^{2\bar{r}})\left[n^{-a} + h^{-1-\beta}n^{-\beta}(\log n)^{3(1+\beta)}\Delta^{-\beta}\right]$$
$$= O\left(n^{-a+2\bar{r}+4\vartheta} + n^{-\beta(1-\vartheta-\varkappa)+2\bar{r}+\vartheta(2\bar{r}+1)}(\log n)^{4\bar{r}+4(1+\beta)}\right),$$

which approaches zero as $n \to \infty$ for $a$ large enough if

$$\frac{2\bar{r} + \vartheta(2\bar{r}+1)}{1-\vartheta-\kappa} < \beta.$$

Since we may pick any $\bar{r} > 1$ in (S.14), we have this inequality satisfied as long as

$$\frac{2+3\vartheta}{1-\vartheta-\kappa} < \beta.$$

Therefore, given this condition on $\beta$, we have shown that $\bar{\mathcal{V}}_{212} = O_p(\sqrt{(\log n)/nh})$, completing the proof of (S.7).

**Proof of (S.8).** The convergence rate of $\mathcal{V}_3$ can be derived in the same way as those of $\mathcal{V}_1$ and $\mathcal{V}_2$, and we here outline only the main points. Since $\sigma^2(\cdot)$ is twice continuously differentiable, we can apply Ito's lemma to $\sigma^2(X_s) - \sigma^2(X_{j\Delta})$ to obtain

$$\mathcal{V}_3 \leq \sup_{x\in\mathbb{R}}\left|(1/Th)\sum_{j=1}^{n-1}K\left(\tfrac{X_{j\Delta}-x}{h}\right)B(X_{j\Delta})\int_{j\Delta}^{(j+1)\Delta}\int_{j\Delta}^{s}m_3(X_u)\,duds\right|$$
$$+\sup_{x\in\mathbb{R}}\left|(1/Th)\sum_{j=1}^{n-1}K\left(\tfrac{X_{j\Delta}-x}{h}\right)B(X_{j\Delta})\int_{j\Delta}^{(j+1)\Delta}\int_{j\Delta}^{s}\partial\sigma^2(X_u)\,\sigma(X_u)\,dW_u ds\right|,$$

where $m_3(x) := \partial\sigma^2(x)\mu(x) + \partial^2\sigma^2(x)\sigma^2(x)/2$. Setting $\psi(x) = m_3(x)$, this $\psi(x)$ satisfies the conditions of Theorem 1 and thus, we have $B(X_{j\Delta})m_3(X_u) = O_{a.s.}(1)$ uniformly as discussed in deriving (19), and we can show that the first term on the RHS is $O_p(\Delta)$. The second term is $O_p(\sqrt{(\log n)/nh})$, which follows from the same arguments as those for $\mathcal{V}_{12}$ and $\mathcal{V}_2$ (we omit details for brevity).

**Proof of (S.9).** We look at

$$\mathcal{V}_4 = \frac{n-1}{n}\int_{-\infty}^{\infty}K(q)\left[H(qh+x) - H(x)\right]dq + \int_{-\infty}^{\infty}K(q)\left[l(qh+x) - l(x)\right]\sigma^2(x)\,dp$$
$$\leq (h^2/2)\int_{-\infty}^{\infty}q^2K(q)\,dq \times \sup_{x\in\mathbb{R}}|H''(x)| + (h^2/2)\int_{-\infty}^{\infty}q^2K(q)\,|l''(\lambda qh+x)|\,\sigma^2(x)\,dq,$$

where we have set $H(x) = B(x)\pi(x)\sigma^2(x)$ and $l(x) = B(x)\pi(x)$; the inequality follows from the usual Talyor-expansion argument with $\lambda \in [0,1]$ (which depends on $q$, $h$, and $x$). Then, we can check that

$\mathcal{V}_4 = O(h^2)$, since we have $\sup_{x\in\mathbb{R}} |H''(x)| < \infty$ (by the condition that $\sigma^2 \in \mathcal{D}(B,\pi)$), as well as the compactness of the support of $K$ and the following bound:

$$
\begin{aligned}
&|l''(\lambda q h + x)|\, \sigma^2(x)\, \mathbf{1}_{\{|q|\le c_K\}} \\
&\le 4C_B \left[\max_{k=0,1,2} \sup_{x\in\mathbb{R}} \left|\partial^k \pi(x)\right|\right] \times \sup_{x,y\in\mathbb{R}} B(y)\, \sigma^2(x)\, \mathbf{1}_{\{|x-y|\le c_K h\}} \\
&\le 4C_B \left[\max_{k=0,1,2} \sup_{x\in\mathbb{R}} \left|\partial^k \pi(x)\right|\right] \times \sup_{x,y\in\mathbb{R}} B(y)\, \bar{C}_0 [1+|x|^{\bar{q}_2}] \mathbf{1}_{\{|x-y|\le c_K h\}} \\
&\le \bar{C}_1 + \bar{C}_1 B(y) \left[|y| + |x-y|\right]^{\bar{q}_2+1} \mathbf{1}_{\{|x-y|\le c_K h\}} \\
&\le \bar{C}_1 + \bar{C}_1 B(y)\, 2^{\bar{q}_2} \left[|y|^{\bar{q}_2+1} + |x-y|^{\bar{q}_2+1}\right] \mathbf{1}_{\{|x-y|\le c_K h\}} \\
&\le \bar{C}_1 + \bar{C}_1 2^{\bar{q}_2} \left[\sup_{y\in\mathbb{R}} B(y)\, |y|^{\bar{q}_2+1} + \sup_{y\in\mathbb{R}} B(y)\, (c_K h)^{\bar{q}_2+1}\right] < \infty,
\end{aligned}
\tag{S.17}
$$

where the first inequality holds since $\max_{k=0,1,2} \sup_{x\in\mathbb{R}} \left|\partial^k \pi(x)\right| < \infty$ and $\max_{k=1,2} |\partial^k B(x)| \le C_B \times B(x)$; the second inequality holds since there exists some constant $\bar{C}_0 > 0$ such that $\sigma^2(x) \le \bar{C}_0 \left[1 + |x|^{\bar{q}_2+1}\right]$ (by the condition that $\sigma^2(x) = O(|x|^{\bar{q}_2})$ as $|x| \to \infty$ for $\bar{q}_2 \ge 0$); the third inequality holds with a constant $\bar{C}_1 = 4C_B \left[\max_{k=0,1,2} \sup_{x\in\mathbb{R}} \left|\partial^k \pi(x)\right|\right] \bar{C}_0 \in (0,\infty)$; and the fourth inequality follows from the Jensen inequality.

**Proof of (S.10).** To bound the term $\mathcal{V}_5$, we consider a compact set $[-\mathcal{T}_n, \mathcal{T}_n] \subset \mathbb{R}$ with $\mathcal{T}_n \to \infty$, whose growing rate is specified below, and its finite covering $\{\mathcal{I}_k\}_{k=1}^{\nu(n)}$ such that $[-\mathcal{T}_n, \mathcal{T}_n] \subset \cup_{k=1}^{\nu(n)} \mathcal{I}_k$, each $\mathcal{I}_k$ is a closed ball in $\mathbb{R}$ with its center $x_k$ and radius $r_n$, and $\nu(n) = \mathcal{T}_n / r_n$. Then, we can write

$$
\begin{aligned}
\mathcal{V}_5(x) \le{}& \sup_{|x|>\mathcal{T}_n} (1/nh) \sum_{j=1}^{n-1} |\Gamma_{j\Delta}(x) - E[\Gamma_{j\Delta}(x)]| \\
&+ \max_{k\in\{1,\ldots,\nu(h)\}} \sup_{x\in\mathcal{I}_k} (1/nh) \sum_{j=1}^{n-1} \{|\Gamma_{j\Delta}(x) - \Gamma_{j\Delta}(x_k)| + |E[\Gamma_{j\Delta}(x)] - E[\Gamma_{j\Delta}(x_k)]|\} \\
&+ \max_{k\in\{1,\ldots,\nu(h)\}} \left|(1/nh) \sum_{j=1}^{n-1} \{\Gamma_{j\Delta}(x_k) - E[\Gamma_{j\Delta}(x_k)]\}\right| \\
=:{}& \mathcal{V}_{51} + \mathcal{V}_{52} + \mathcal{V}_{53},
\end{aligned}
$$

where we consider the bounds of these three terms below.

To find the bound of $\mathcal{V}_{51}$, we let

$$
\bar{\Gamma}_{j\Delta}(x) := \Gamma_{j\Delta}(x)\, \mathbf{1}_{\{|X_{j\Delta}|\le \mathcal{T}_n/2\}}; \quad \text{and} \quad \tilde{\Gamma}_{j\Delta}(x) := \Gamma_{j\Delta}(x)\, \mathbf{1}_{\{|X_{j\Delta}|> \mathcal{T}_n/2\}},
$$

and observe that

$$
\mathcal{V}_{51} \le \sup_{|x|>\mathcal{T}_n} (1/nh) \sum_{j=1}^{n-1} \left|\bar{\Gamma}_{j\Delta}(x) - E[\bar{\Gamma}_{j\Delta}(x)]\right| + \sup_{|x|>\mathcal{T}_n} (1/nh) \sum_{j=1}^{n-1} \left|\tilde{\Gamma}_{j\Delta}(x) - E[\tilde{\Gamma}_{j\Delta}(x)]\right|. \tag{S.18}
$$

For $|x| > \mathcal{T}_n$ and $|X_{j\Delta}| \leq \mathcal{T}_n/2$, it holds that $(X_{j\Delta} - x)/h \geq \mathcal{T}_n/2h$. Therefore, for sufficiently large $n$ with $\mathcal{T}_n/2h \geq c_K$, $K((X_{j\Delta} - x)/h) = 0$ and the first term on the RHS of (S.18) is zero. To find the bound of the second term, we also observe that

$$\begin{aligned}
&\sup_{|x-y|\leq c_K h} B^{1/2}(y)\left|\sigma^2(y) - \sigma^2(x)\right| \\
&= \sup_{|x-y|\leq c_K h} B^{1/2}(y)\left|\partial\sigma^2(y+\lambda(x-y))\right||x-y| \\
&\leq \sup_{|x-y|\leq c_K h} B^{1/2}(y)\bar{C}_2\left[1+|y+\lambda(x-y)|^{\bar{q}_2+1}\right]c_K h \\
&\leq O(h) + \bar{C}_2 \sup_{|x-y|\leq c_K h} B^{1/2}(y)\, 2^{\bar{q}_2}\left[|y|^{\bar{q}_2+1} + |\lambda(x-y)|^{\bar{q}_2+1}\right]c_K h \\
&= O(h), \qquad \text{(S.19)}
\end{aligned}$$

for some constant $\bar{C}_2 > 0$, where the first equality follows from the mean-value theorem with some $\lambda \in [0,1]$ (which depends on $x$ and $y$); the two inequalities use the polynomial growth condition of $\partial\sigma^2$ and the Jensen inequality. Thus, we have

$$\begin{aligned}
&\sup_{|x|>\mathcal{T}_n}(1/nh)\sum_{j=1}^{n-1}\left|\tilde{\Gamma}_{j\Delta}(x)\right| \\
&\leq (1/nh)\bar{K}\sum_{j=1}^{n-1}\mathbf{1}_{\{|X_{j\Delta}-x|\leq c_K h\}}B^{1/2}(X_{j\Delta})\left|\sigma^2(X_{j\Delta})-\sigma^2(x)\right| \times B^{1/2}(X_{j\Delta})\mathbf{1}_{\{|X_{j\Delta}|>\mathcal{T}_n/2\}} \\
&\leq (1/nh)\bar{K}\sum_{j=1}^{n-1}\sup_{|x-y|\leq c_K h} B^{1/2}(y)\left|\sigma^2(y)-\sigma^2(x)\right| \times B^{1/2}(X_{j\Delta})|X_{j\Delta}|^{\bar{d}}/(\mathcal{T}_n/2)^{\bar{d}} \\
&= O(1/\mathcal{T}_n^{\bar{d}}),
\end{aligned}$$

for any $\bar{d} > 0$, where the last equality holds since $B^{1/2}(x)|x|^d$ is uniformly bounded. We can also show that

$$\sup_{|x|>\mathcal{T}_n}(1/nh)\sum_{j=1}^{n-1}\left|E[\tilde{\Gamma}_{j\Delta}(x)]\right| = O(1/\mathcal{T}_n^{\bar{d}})$$

in the same way. From these, we can conclude that $\mathcal{V}_{51} = O(1/\mathcal{T}_n^{\bar{d}})$.

To bound the term $\mathcal{V}_{52}$, note that for any $x \in \mathcal{I}_k$, $|x - x_k| \leq r_n$. This implies that for an event

$$E_{n,j\Delta}(x, x_k) := \{\max\{|X_{j\Delta} - x|, |X_{j\Delta} - x_k|\} \leq c_K h + r_n\},$$

which is defined in $\Omega$ for each $(n, j\Delta, x, x_k)$, we have

$$\begin{aligned}
\{|X_{j\Delta} - x| \leq c_K h\} &= \{|X_{j\Delta} - x| \leq c_K h \;\&\; |X_{j\Delta} - x_k| \leq c_K h + r_n\} \\
&\subset E_{n,j\Delta}(x, x_k); \\
\{|X_{j\Delta} - x_k| \leq c_K h\} &\subset E_{n,j\Delta}(x, x_k).
\end{aligned}$$

Therefore, for any $x \in \mathcal{I}_k$,

$$
\begin{aligned}
&\left|\Gamma_{j\Delta}(x) - \Gamma_{j\Delta}(x_k)\right| \\
&\leq B(X_{j\Delta})\,\sigma^2(X_{j\Delta}) \left| K\left(\tfrac{X_{j\Delta}-x}{h}\right) - K\left(\tfrac{X_{j\Delta}-x_k}{h}\right) \right| + K\left(\tfrac{X_{j\Delta}-x}{h}\right) B(X_{j\Delta})\,\mathbf{1}_{E_{n,j}(x,x_k)} \left|\sigma^2(x) - \sigma^2(x_k)\right| \\
&+ B(X_{j\Delta})\,\sigma^2(x_k)\,\mathbf{1}_{\{|X_{j\Delta}-x| \leq c_K h \text{ or } |X_{j\Delta}-x| \leq c_K h\}} \left| K\left(\tfrac{X_{j\Delta}-x}{h}\right) - K\left(\tfrac{X_{j\Delta}-x_k}{h}\right) \right| \\
&\leq \left\{\sup_{x\in\mathbb{R}} B(x)\,\sigma^2(x)\right\} \bar{K}\,|x_k - x|/h + \bar{K} B(X_{j\Delta})\,\mathbf{1}_{E_{n,j}(x,x_k)} \left|\sigma^2(x) - \sigma^2(x_k)\right| \\
&+ B(X_{j\Delta})\,\sigma^2(x_k)\,\mathbf{1}_{E_{n,j}(x,x_k)} \bar{K}\,|x_k - x|/h \\
&\leq O(r_n/h) + O(r_n) + O(r_n/h) \\
&= O(r_n/h),
\end{aligned}
$$

uniformly over $x$, $x_k$, and $j$, where the first inequality uses the triangle inequalities as well as the fact that if $|X_{j\Delta} - x| > c_K h$ and $|X_{j\Delta} - x| > c_K h$, then

$$
K\left(\tfrac{X_{j\Delta}-x}{h}\right) = K\left(\tfrac{X_{j\Delta}-x_k}{h}\right) = 0 \quad \text{and} \quad \mathbf{1}_{\{|X_{j\Delta}-x| \leq c_K h \text{ or } |X_{j\Delta}-x| \leq c_K h\}} \leq \mathbf{1}_{E_{n,j}(x,x_k)};
$$

the second inequality uses the Lipschitz continuity and uniform boundedness of $K$; and the third inequality holds because we have

$$
\begin{aligned}
&B(X_{j\Delta})\,\mathbf{1}_{E_{n,j}(x,x_k)} \left|\sigma^2(x) - \sigma^2(x_k)\right| \\
&\leq B(X_{j\Delta})\,\mathbf{1}_{E_{n,j}(x,x_k)} \left|\partial\sigma^2(\lambda x + (1-\lambda)x_k)\right| \times |x - x_k| \\
&\leq \left\{B(X_{j\Delta})\left|\partial\sigma^2(X_{j\Delta})\right| + B(X_{j\Delta})\left|\partial\sigma^2(X_{j\Delta}) - \partial\sigma^2(\lambda x + (1-\lambda)x_k)\right|\right\} \mathbf{1}_{E_{n,j}(x,x_k)} \times |x - x_k| \\
&\leq \left\{\sup_{x\in\mathbb{R}} B(x)\left|\partial\sigma^2(x)\right| + \sup_{|x-y|\leq c_K h + r_n} B(x)\left|\partial\sigma^2(x) - \partial\sigma^2(y)\right|\right\} \times r_n \\
&= O(r_n) \quad \text{and}
\end{aligned}
$$

$$
B(X_{j\Delta})\,\sigma^2(x_k)\,\mathbf{1}_{E_{n,j}(x,x_k)} = O(1),
$$

uniformly over $x$, $x_k$, and $j$, which follow from the same arguments as those for deriving (S.17) and (S.19). From these, we can conclude that $\mathcal{V}_{52} = O(r_n/h^2)$. Now, by setting $\mathcal{T}_n = [nh/(\log n)]^{1/2\bar{d}}$ and $r_n = \sqrt{h^3(\log n)/n}$, we can let both $\mathcal{V}_{51}(= (1/\mathcal{T}_n^{\bar{d}}))$ and $\mathcal{V}_{52}$ be $O_p(\sqrt{(\log n)/nh})$.

Finally, to investigate the rate of $\mathcal{V}_{53}$, we derive the bound

$$
\Sigma^2_{\Gamma,m} := E\left[\left|\sum_{j=1}^m \{\Gamma_{j\Delta}(x_k) - E[\Gamma_{j\Delta}(x_k)]\}\right|^2\right] \quad \text{for } m \leq (n-1),
$$

and then apply the exponential inequality. For this purpose, we observe that

$$
\begin{aligned}
&E[\{\Gamma_{j\Delta}(x_k) - E[\Gamma_{j\Delta}(x_k)]\}^2] \\
&\leq 2h\int_{-\infty}^{\infty} K^2(q)\,B^2 \cdot \pi(qh + x_k)\,[\sigma^2(qh + x_k) - \sigma^2(x_k)]^2 dq = O(h^3),
\end{aligned}
$$

uniformly over $x_k$ and $j\Delta$, where the last equality follows from (S.19). We can find some constant $\tilde{\omega} \in (0,\infty)$ such that $\Sigma^2_{\Gamma,m} \leq \tilde{\omega} m^2 h^3$, and also some constant $C_\Gamma \in (0,\infty)$ such that

$$\begin{aligned}\Gamma_{j\Delta}(x) &= K\left(\tfrac{X_{j\Delta}-x}{h}\right) B\left(X_{j\Delta}\right)\left|\partial\sigma^2\left(\lambda X_{j\Delta} - (1-\lambda)x\right)\right|\left|X_{j\Delta}-x\right| 1_{\left\{\left|X_{j\Delta}-x\right|\leq c_K h\right\}} \\ &\leq \bar{K} \sup_{x,y\in\mathbb{R};\ |\epsilon|\leq\bar{\epsilon}} B(x)\left|\partial\sigma^2(y)\right| c_K h \leq C_\Gamma h,\end{aligned}$$

which follows from the compactness of the support of $K$. Then, applying Lemma 3, we have for each $a > 0$,

$$\begin{aligned}&\Pr\left[\mathcal{V}_{53} \geq a\sqrt{(\log n)/nh}\right] \\ &\leq \sum_{k=1}^{\nu(n)} \Pr\left[\left|\sum_{j=1}^{n-1}\{\Gamma_{j\Delta}(x_k) - E[\Gamma_{j\Delta}(x_k)]\}\right| \geq a\sqrt{(\log n)/nh}nh\right] \\ &\leq \nu(n)\left\{4\exp\left\{-\frac{a^2(\log n)}{64\tilde{\omega}mh^2 + (8/3)(2C_\Gamma h)a\sqrt{(\log n)/nh}m}\right\} + \frac{4n}{m}A(m\Delta)^{-\beta}\right\} \\ &\leq (\mathcal{T}_n/r_n)\left\{4\exp\left\{-\frac{a^2(\log n)}{64\tilde{\omega}\sqrt{nh^5/(\log n)} + 16C_\Gamma a/3}\right\} + 4Anm^{-1-\beta}\Delta^{-\beta}\right\} \\ &\leq (\mathcal{T}_n/r_n) \times 4n^{-a^2/\left[64\tilde{\omega}\tilde{C}+16C_\Gamma a/3\right]} + 4A(\mathcal{T}_n/r_n)nm^{-1-\beta}\Delta^{-\beta}, \end{aligned} \tag{S.20}$$

where the third inequality holds by setting $m = \sqrt{n/h(\log n)}$ $(\leq \min\{(a\sqrt{(\log n)/nh}nh)/C_\Gamma h, n-1\}$ for $a$ large enough), and the last equality holds since $nh^5/(\log n) \leq \tilde{C}$ for some constant $\tilde{C} > 0$ (whose existence follows from (29)). Now, by the definitions of $\mathcal{T}_n (= [nh/(\log n)]^{1/2\bar{d}})$ and $r_n (= \sqrt{h^3(\log n)/n})$, we have

$$\nu(n) = \mathcal{T}_n/r_n = (\log n)^{-\left(1+1/\bar{d}\right)/2} h^{-\left(3-1/\bar{d}\right)/2} n^{\left(1+1/\bar{d}\right)/2},$$

which is a polynomial order of $n$, and the first term on the RHS of (S.20) approaches zero as $n\to\infty$ for $a$ large enough. Regarding the second term, recalling the definition of $m$ and the conditions $\Delta^{-1} \leq n^\kappa$, we have

$$(\mathcal{T}_n/r_n)nm^{-1-\beta}\Delta^{-\beta} \leq (\log n)^{-\left(\beta+2+1/\bar{d}\right)/2} h^{\left(\beta-2+1/\bar{d}\right)/2} n^{\left(-\beta+2\beta\kappa+2+1/\bar{d}\right)/2},$$

which approaches zero (as $n\to\infty$) if

$$-\beta + 2\beta\kappa + 2 < 0 \Longleftrightarrow 2/(1-2\kappa) < \beta,$$

where we note that $\bar{d}$ can be any arbitrarily large integer. Now, the proof of Theorem 5 is completed. □

## S.3 Convergence Results When Mixing Coefficients Decay Slowly

In this section, we present some results which complement the convergence results in Theorems 2-5, focusing on the case when the decay rate of the mixing coefficients in (10) is slow.

**General Convergence Results with Possibly Small $\beta$:** Theorem 2 requires at least $\beta > 5$ as in the condition (22), but the following theorem allows for any $\beta > 0$. At the price of possibly small $\beta$, we must have a slower convergence rate of $\sqrt{(\log T)/T^{\theta}h}$ (than that of $\sqrt{(\log T)/Th}$ in Theorem 2). We also note that smaller $\beta$ requires smaller $\theta$, implying a slower convergence rate of the bandwidth $h$ through (21).

**Theorem S.1.** *Suppose that the same conditions as in Theorem 2 hold with the condition (22) replaced by $\beta \geq 5\theta/(1-\theta)$. Then, it holds that as $n, T \to \infty$ and $\Delta, h \to 0$,*

$$\sup_{x \in I} |G_{n,T}(x) - E[G_{n,T}(x)]| = O_p(\sqrt{\Delta \log(1/\Delta)}) + O_p(\sqrt{(\log T)/T^{\theta}h}).$$

*Proof of Theorem S.1.* The proof proceeds in the same way as in that of Theorem 2. Since only the rate of the term $R_2$ differs, we omit details and outline only the main points for $R_2 \leq R_{21} + R_{22}$. To find the rate of $R_{21}$, we set

$$\varepsilon = h\sqrt{(\log T)/T^{\theta}h}, \tag{S.21}$$

instead of (51). This means that $R_{21} = O(\sqrt{(\log T)/T^{\theta}h})$. To derive the rate of $R_{22}$, we use the Bernstein-type inequality. To this end, observe the following moment bound:

$$E[|\textstyle\sum_{j=1}^{m} Y_{n,j}(k,h)|^2] \leq m \textstyle\sum_{j=1}^{m} E\left[Y_{n,j}^2(k,h)\right] \leq \varpi m^2 h,$$

uniformly over $k$ and $h$, where the first inequality holds by the Jensen inequality for $m \leq (n-1)$, and the second holds by the moment bound derived in (55).

Now, we apply Lemma 3 to $\sum_{j=1}^{n-1} Y_{n,j}(k,h)$ with

$$Z_{n,j} = Y_{n,j}(k,h) \quad \text{and} \quad \Sigma_m^2 = \Sigma_m^2(k,h) := E[|\textstyle\sum_{j=1}^{m} Y_{n,j}(k,h)|^2],$$

for each $(k,h)$. Let $\eta = a\left[(\log T)/T^{\theta}h\right]^{1/2} nh$ and $m = T^{(1-\theta)}/\Delta$ in (44), where $m \leq (n-1)$ and $m < \eta/4C_Y$ are satisfied for large $T$ (since $\theta \in (0,1)$ and $(\log T)/T^{\theta}h \to 0$). Then, it holds that for any $a > 0$,

$$\begin{aligned}
&\Pr\left[R_{22} \geq a\sqrt{(\log T)/T^{\theta}h}\right] \\
&\leq \sum_{k=1}^{\nu(h)} \Pr\left[\left|\sum_{j=1}^{n-1} Y_{n,j}(k,h)\right| \geq a\sqrt{(\log T)/T^{\theta}h}\,nh\right] \\
&\leq \nu(h)\left\{4\exp\left\{-\frac{a^2\left[(\log T)/T^{\theta}h\right]n^2h^2}{64n\varpi mh + (8/3)C_Y a\sqrt{(\log T)/T^{\theta}h}\,nhm}\right\} + \frac{4n}{m}\alpha(m\Delta)\right\} \\
&\leq 4\Lambda\varepsilon^{-4}\left\{\exp\left\{-\frac{a^2\log T}{64\varpi + (8/3)C_Y a\sqrt{(\log T)/T^{\theta}h}}\right\} + AT^{\theta-\beta(1-\theta)}\right\} \\
&\leq 4\Lambda(\log T)^{-4}\left\{T^{4\theta - a^2/[64\varpi + (8/3)C_Y a]} + AT^{5\theta-\beta(1-\theta)}\right\},
\end{aligned}$$

where the second inequality holds by (44) and (55); the third inequality uses (10) in (A2), "$\nu(h) \leq \Lambda\varepsilon^{-4}$," and "$m = T^{(1-\theta)}/\Delta$"; and the last inequality holds for large $T$ since $\varepsilon = h\sqrt{(\log T)/T^{\theta}h}$ defined in (S.21), $h^{-2} \leq T^{2\theta}(\log T)^{-2}$, and $\sqrt{(\log T)/T^{\theta}h} \leq 1$ (for large $T$). Therefore, for $a > 0$ large enough,

$$\Pr\left[R_{22} \geq a\sqrt{(\log T)/T^{\theta}h}\right] \to 0 \text{ as } T \to \infty,$$

if

$$5\theta - \beta(1-\theta) \leq 0 \Leftrightarrow \theta \leq \beta/(5+\beta).$$

Given this inequality and the rate of $R_{21}$, we obtain $R_{22} = O_p(\sqrt{(\log T)/T^{\theta}h})$, as desired. □

The next theorem also concerns the small-$\beta$ case. While Theorem 3 allows for any $\beta > 0$ (unlike Theorem 2), its probability bound is associated with the convergence rate of $\sqrt{(\log T)/Th}$. If we have a slower rate of $\sqrt{(\log T)/T^{\theta}h}$ (as in Theorem S.1), we can derive a sharper inequality for the probability bound of $M_{n,T}(x)$:

**Theorem S.2.** *Suppose that the same conditions as in Theorem 3 hold. Then, as $n, T \to \infty$ and $\Delta, h \to 0$, it holds that for each $a(> 0)$ and each $x \in I$,*

$$\Pr\left[M_{n,T}(x) \geq a\sqrt{(\log T)/T^{\theta}h}\right] \leq 2T^{-a^2/2} + 4\exp\left\{-aC_M(Th)^{1-\theta}\right\} + 4AT^{-\beta}h^{-\theta(\beta+1)},$$

*for each $\theta \in (0,1)$, where $C_M(> 0)$ is some constant independent of $x$.*

*Proof of Theorem S.2.* Since we use the same arguments as those for Theorem 3, we outline only the main points. Given the same notation as in the proof of Theorem 3, we have for any $\theta \in (0,1)$,

$$\begin{aligned}
&\Pr\left[M_{n,T}(x) \geq a\sqrt{(\log T)/T^{\theta}h}\right] \\
&\leq \Pr\left[|N_1(x,h)| \geq aTh\sqrt{(\log T)/T^{\theta}h}\right] \\
&\leq \Pr\left[|N_1(x,h)| \geq aTh\sqrt{(\log T)/T^{\theta}h},\ \langle N(x,h)\rangle_1 \leq y\right] + \Pr\left[\langle N(x,h)\rangle_1 > y\right] \\
&\leq 2\exp\left\{-a^2(\log T)/2\right\} + \Pr\left[\langle N(x,h)\rangle_1 > aT^{2-\theta}h\right] \\
&= 2T^{-a^2/2} + \Pr\left[\langle N(x,h)\rangle_1 > aT^{2-\theta}h\right], \qquad \text{(S.22)}
\end{aligned}$$

where the third inequality holds by (57) with $\eta = aTh\sqrt{(\log T)/T^{\theta}h}$ and $y = aT^{2-\theta}h$. By applying the Bernstein-type inequality for mixing arrays in Lemma 3 to $\langle N(x,h)\rangle_1$ and using arguments quite analogous to those for (59)-(61), we can also derive

$$\Pr\left[\langle N(x,h)\rangle_1 > aT^{2-\theta}h\right] \leq 4\exp\left\{-aC_M(Th)^{1-\theta}\right\} + 4AT^{-\beta}h^{-\theta(\beta+1)}.$$

This, together with (S.22), implies the desired result. □

**Uniform Convergence Rates of Nadaraya-Watson Type Estimators with Possibly Small $\beta$:** The next two theorems are small-$\beta$ counterparts of Theorems 4 and 5. While Theorem S.4 on the convergence of the diffusion function estimator provides rates in terms of $T$ (slower than $\sqrt{(\log n)/nh}$), it imposes only minimal conditions, allowing for discontinuous kernels with unbounded support. It also relaxes the conditions on the derivatives of $\pi$ and $\sigma$:

**Theorem S.3** (Drift Function Estimation with Possibly Small $\beta$). *Suppose that the same conditions as in Theorem 4 hold but replace the condition on the exponent of the mixing coefficient $\beta$ by*

$$\beta \geq \max\left\{5\theta/\left(1-\theta\right), \left(4\theta+\theta^2+2\kappa\right)/\left(1-\theta^2\right)\right\}.$$

*Then the convergence results in (27)-(28) hold with $a^{\star}_{n,T}$ replaced by*

$$a^{*}_{n,T} := h^2 + \sqrt{\Delta \log\left(1/\Delta\right)} + \sqrt{\left(\log T\right)/T^{\theta}h}.$$

*Proof of Theorem S.3.* Proof arguments proceed in the same way as those for the proof of Theorem 4 while we employ convergence results of Theorems S.1-S.2, instead of Theorems 2-3. We omit details for brevity. □

**Theorem S.4** (Diffusion Function Estimation with Possibly Small $\beta$). *Suppose that Assumption 1 holds;* $\sup_{x\in\mathbb{R}} \pi\left(x\right) < \infty$*; the observation interval $\Delta$ and the bandwidth $h$ satisfy*

$$\Delta^{-1} = O(T^{\kappa}) \quad \textit{and} \quad \left(\log T\right)/T^{\theta}h \to 0,$$

*as $T \to \infty$ and $\Delta, h \to 0$, for some constants $\kappa > 0$ and $\theta \in (0,1)$;*

$$\sigma^2\left(\cdot\right) \in \mathcal{D}\left(B,\pi\right); \quad \left[\left|\partial\mu\left(x\right)\right| + \left|\partial\sigma\left(x\right)\right|\right] = O(|x|^{\tilde{q}_2}) \textit{ as } |x| \to \infty \textit{ for some } \tilde{q}_2 \geq 0.$$

*Let $c_{n,T}$, $\delta_{n,T}$, $a^{\star}_{n,T}$, and $a^{*}_{n,T}$ be sequences defined in Theorems 4 and S.3. Then, the following results hold (as $n, T \to \infty$ and $\Delta, h \to 0$):*

***(i-a)*** *If*

$$\beta \geq \max\left\{5\theta/\left(1-\theta\right), \left(4\theta+\theta^2\right)/\left(1-\theta^2\right)\right\},$$

*then,*

$$\sup_{x\in\mathbb{R}} \left|\hat{\Psi}_{\sigma^2}\left(x\right) - B\left(x\right)\sigma^2\left(x\right)\pi\left(x\right)\right| = O_p\left(a^{*}_{n,T}\right). \tag{S.23}$$

***(i-b)*** *Further if $a^{*}_{n,T}/\delta_{n,T} \to 0$,*

$$\sup_{|x|\leq c_{n,T}} \left|\hat{\sigma}^2\left(x\right) - \sigma^2\left(x\right)\right| = O_p\left(a^{*}_{n,T}/\delta_{n,T}\right).$$

***(ii)*** *If*

$$\beta \geq \max\left\{5\left(1+\theta\right)/\left(1-\theta\right), \left(2+3\theta\right)/\left(1-\theta\right)\right\},$$

*then, the convergence results in (i-a) and (i-b) hold with $a^{*}_{n,T}$ replaced by $a^{\star}_{n,T}$.*

*Proof of Theorem S.4.* Using (S.5), we split the LHS of (S.23) into three terms:

$$\sup_{x\in\mathbb{R}} |\hat{\Psi}_{\sigma^2}(x) - B(x)\sigma^2(x)\pi(x)| \leq \sum_{i=1}^{4} V_i,$$

where

$$V_1 := \sup_{x\in\mathbb{R}} \left| (2/Th)\sum_{j=1}^{n-1} K\left(\tfrac{X_{j\Delta}-x}{h}\right) B(X_{j\Delta}) \int_{j\Delta}^{(j+1)\Delta} [X_s - X_{j\Delta}]\,\mu(X_s)\,ds \right|;$$

$$V_2 := \sup_{x\in\mathbb{R}} \left| (2/Th)\sum_{j=1}^{n-1} K\left(\tfrac{X_{j\Delta}-x}{h}\right) B(X_{j\Delta}) \int_{j\Delta}^{(j+1)\Delta} [X_s - X_{j\Delta}]\,\sigma(X_s)\,dW_s \right|;$$

$$V_3 := \sup_{x\in\mathbb{R}} \Big\{ (1/Th)\sum_{j=1}^{n-1} K\left(\tfrac{X_{j\Delta}-x}{h}\right) B(X_{j\Delta}) \int_{j\Delta}^{(j+1)\Delta} \sigma^2(X_s)\,ds$$
$$- (1/h)\,E[K\left(\tfrac{X_{j\Delta}-x}{h}\right) B(X_{j\Delta})\,\sigma^2(X_{j\Delta})] \Big\};$$

$$V_4 := \sup_{x\in\mathbb{R}} \left| (1/h)\int_{-\infty}^{\infty} K\left(\frac{p-x}{h}\right) B(p)\,\sigma^2(p)\,\pi(p)\,dp - B(x)\,\sigma^2(x)\,\pi(x) \right|.$$

Below, we investigate these four terms. First, by Theorem 1 and (19), we have

$$\begin{aligned} B(X_{j\Delta})[X_s - X_{j\Delta}]\,\mu(X_s) &= B^{1/2}(X_{j\Delta})[X_s - X_{j\Delta}] \times B^{1/2}(X_{j\Delta})\,\mu(X_s) \\ &= O_{a.s.}(\sqrt{\Delta \log(1/\Delta)}), \end{aligned}$$

uniformly. This implies that $V_1 = O_p(\sqrt{\Delta \log(1/\Delta)})$, since

$$\sup_{x\in\mathbb{R}} (1/nh)\sum_{j=1}^{n-1} \left| K\left(\tfrac{X_{j\Delta}-x}{h}\right) \right| = O_p(1),$$

which is derived in the proof of Theorem 2. Next, applying Theorem S.1 (resp. Theorem 2) to $V_3$ with $\psi(\cdot) = \sigma^2(\cdot)$, we can immediately obtain $V_3 = O_p(a_T^*)$ (resp. $O_p(a_T^\star)$) under the condition on $\beta$ in part (i) (resp. part (ii)). We can also show that $V_4 = O(h^2)$ in the same way as for the term $U_2$ in the proof of Theorem 4. We subsequently show that

$$V_2 = O_p(\sqrt{(\log T)/T^\theta h}) \text{ for part (i); and } O_p(\sqrt{(\log T)/Th}) \text{ for part (ii),} \tag{S.24}$$

under the stated conditions. Given these, we can obtain the desired convergence results for $\sup_{x\in\mathbb{R}} |\hat{\Psi}_{\sigma^2}(x) - B(x)\sigma^2(x)\pi(x)|$, which in turn allow us to drive the desired convergence results for $|\hat{\sigma}^2(x) - \sigma^2(x)|$ in the same way as in the proof of Theorem 8 in Hansen (2008) (we omit details for brevity).

**Proof of (S.24).** We use arguments analogous to those for $U_2$ in Theorem 4, and thus only outline the main points. Look at

$$\begin{aligned} &B(X_{j\Delta})[X_s - X_{j\Delta}]\,\sigma(X_s) \\ &= B^{1/2}(X_{j\Delta})[\int_{j\Delta}^{s} \mu(X_u)\,du + \int_{j\Delta}^{s} \sigma(X_u)\,dW_u] \times B^{1/2}(X_{j\Delta})\,\sigma(X_{j\Delta}) \\ &+ B^{1/2}(X_{j\Delta})[\int_{j\Delta}^{s} \mu(X_u)\,du + \int_{j\Delta}^{s} \sigma(X_u)\,dW_u] \times B^{1/2}(X_{j\Delta})[\sigma(X_s) - \sigma(X_{j\Delta})], \end{aligned}$$

and define

$$\mathbf{f}_{\Delta}(s,j) := \mathbf{1}_{\left\{B^{1/2}(X_{j\Delta})[|\int_{j\Delta}^{s}\mu(X_u)du|+|\int_{j\Delta}^{s}\sigma(X_u)dW_u|+\left|\sigma(X_s)-\sigma(X_{j\Delta})\right|]\leq\Delta^{1/2}\log(1/\Delta)\right\}}.$$

Then,

$$\begin{aligned}
V_2 &= (2/Th)\sup_{x\in\mathbb{R}}\left|\sum_{j=1}^{n-1}K\left(\tfrac{X_{j\Delta}-x}{h}\right)\int_{j\Delta}^{(j+1)\Delta}B(X_{j\Delta})[X_s-X_{j\Delta}]\,\sigma(X_s)\,\mathbf{f}_{\Delta}(s,j)\,dW_s\right| \\
&\quad+(2/Th)\sup_{x\in\mathbb{R}}\left|\sum_{j=1}^{n-1}K\left(\tfrac{X_{j\Delta}-x}{h}\right)\int_{j\Delta}^{(j+1)\Delta}B(X_{j\Delta})[X_s-X_{j\Delta}]\,\sigma(X_s)\,[1-\mathbf{f}_{\Delta}(s,j)]dW_s\right| \\
&=: \bar{V}_2+\tilde{V}_2.
\end{aligned}$$

Using the same arguments as those used to derive the result (S.7) in the proof of Theorem 5, we have $\tilde{V}_2 = 0$ almost surely for sufficiently small $\Delta$. Therefore, the convergence rate of $V_2$ is determined by that of $\bar{V}_2$. Letting

$$q(s,j\Delta) := B(X_{j\Delta})[X_s-X_{j\Delta}]\,\sigma(X_s)\,\mathbf{f}_{\Delta}(s,j),$$

we have

$$\begin{aligned}
\bar{V}_2 &\leq (1/Th)\max_{k\in\{1,\dots,\nu(h)\}}\sup_{g\in\mathcal{K}_k(h)}\sum_{j=1}^{n-1}|g_k(X_{j\Delta})-g(X_{j\Delta})|\left|\int_{j\Delta}^{(j+1)\Delta}q(s,j\Delta)\,dW_s\right| \\
&\quad+\max_{k\in\{1,\dots,\nu(h)\}}\left|(1/Th)\sum_{j=1}^{n-1}K\left(\tfrac{X_{j\Delta}-x}{h}\right)\int_{j\Delta}^{(j+1)\Delta}q(s,j\Delta)\,dW_s\right| \\
&=: \bar{V}_{21}+\bar{V}_{22},
\end{aligned}$$

where $\{\mathcal{K}_k(h)\}_{k=1}^{\nu(h)}$ is the finite covering of $\mathcal{K}(h)$, as defined in the proof of Theorem 5, satisfying (S.13) with $\nu(h)\leq\Lambda\varepsilon^{-4\bar{r}}$ (for some constant $\Lambda>0$ and any $\bar{r}>1$). In the same way as in (S.14), we can show that

$$\bar{V}_{21} = O_p(h^{-1}\varepsilon\log\Delta^{-1}). \tag{S.25}$$

By Theorems S.2 and 3, for any $a>0$,

$$\Pr\left[\bar{V}_{22}\geq a\sqrt{(\log T)/T^{\theta}h}\right]\leq\Lambda\varepsilon^{-4\bar{r}}\left[2T^{-a^2/2}+4\exp\{-aC_M(Th)^{1-\theta}\}+4AT^{-\beta}h^{-\theta(\beta+1)}\right]; \tag{S.26}$$

$$\Pr\left[\bar{V}_{22}\geq a\sqrt{(\log T)/Th}\right]\leq\Lambda\varepsilon^{-4\bar{r}}\left[2T^{-a^2/2}+T^{-aC_M}+4AT^{-\beta}h^{-(\beta+1)}(\log T)^{1-\beta}\right]. \tag{S.27}$$

Now, we can derive the convergence result of $\bar{V}_2$ under the condition on $\beta$ of part (i). We let $\varepsilon=\sqrt{h/T^{\theta}(\log T)}$ and obtain $\bar{V}_{21}=O_p(\sqrt{(\log T)/T^{\theta}h})$, since $\Delta^{-1}=O(T^{\kappa})$ and $\log\Delta^{-1}=O(\log T)$. Then, by using (S.26) and the condition that $h^{-1}=O(T^{\theta}/(\log T))$, we can show that as $T\to\infty$,

$$\Pr\left[\bar{V}_{22}\geq a\sqrt{(\log T)/T^{\theta}h}\right]\to 0,$$

for any $a$ large enough if

$$\varepsilon^{-4\bar{r}}\times T^{-\beta}h^{-\theta(\beta+1)} = O((\log T)^{2\bar{r}-\theta(\beta+1)}\times T^{-\beta+4\bar{r}\theta+\theta^2(\beta+1)})$$

approaches zero, which occurs as long as

$$-\beta + 4\bar{r}\theta + \theta^2(\beta+1) < 0 \Longleftrightarrow 4\bar{r}\theta + \theta^2 < \beta\left(1-\theta^2\right).$$

Recalling that any $\bar{r} > 1$ can be selected, we can see that the last inequality is satisfied if $4\theta + \theta^2 < \beta(1-\theta^2)$. We now have $\bar{V}_2 = O_p(\sqrt{(\log T)/T^\theta h})$, as desired for the part (i) case.

Finally, suppose that the condition on $\beta$ of part (ii) holds. In this case, plugging $\varepsilon = \sqrt{h/T(\log T)}$ into (S.25) and (S.27), we have $\bar{V}_{21} = O_p(\sqrt{(\log T)/Th})$, and $\Pr\left[\bar{V}_{22} \geq a\sqrt{(\log T)/Th}\right] \to 0$ as $T \to \infty$ for any $a$ large enough if

$$\varepsilon^{-4\bar{r}} \times T^{-\beta} h^{-(\beta+1)} (\log T)^{1-\beta} = O((\log T)^{2\bar{r}-\theta(\beta+1)} \times T^{-\beta(1-\theta)+2\bar{r}+(2\bar{r}+1)\theta})$$

approaches zero, which occurs as long as

$$2\bar{r} + (2\bar{r}+1)\theta < \beta(1-\theta).$$

We can obtain this inequality if $2 + 3\theta < \beta(1-\theta)$, since any $\bar{r} > 1$ can be picked. The proof is now completed. □

## S.4 Effects of the Damping Function

In this section, we briefly investigate the effects of the damping function by presenting bias, variance, and mean-squared-error (MSE) expressions of the estimators (4) and (5) as well as by providing some graphical illustration.

**Bias and Variance Expressions:** The exact expressions are difficult to analyze, and we derive their approximations for each $x \in (l, r)$:

$$E\left[(\hat{\mu}(x) - \mu(x))^2\right] \simeq \mathcal{B}_\mu^2(x) + \mathcal{V}_\mu(x);$$
$$E\left[\left(\hat{\sigma}^2(x) - \sigma^2(x)\right)^2\right] \simeq \mathcal{B}_{\sigma^2}^2(x) + \mathcal{V}_{\sigma^2}(x);$$

where

$$\mathcal{B}_\mu(x) := h^2\left\{(d/dx)\left[B(x)\mu(x)\right] \times \frac{\pi'(x)}{B(x)\pi(x)} + \frac{(d^2/dx^2)\left[B(x)\mu(x)\right]}{2B(x)}\right\}\int_{-\infty}^{\infty} z^2 K(z)\,dz;$$
$$\mathcal{V}_\mu(x) := (1/Th)\left[\sigma^2(x)/\pi(x)\right]\int_{-\infty}^{\infty} K^2(z)\,dz;$$
$$\mathcal{B}_{\sigma^2}(x) := h^2\left\{(d/dx)\left[B(x)\sigma^2(x)\right] \times \frac{\pi'(x)}{B(x)\pi(x)} + \frac{(d^2/dx^2)\left[B(x)\sigma^2(x)\right]}{2B(x)}\right\}\int_{-\infty}^{\infty} z^2 K(z)\,dz;$$
$$\mathcal{V}_{\sigma^2}(x) := (1/nh)\,2\left[\sigma^4(x)/\pi(x)\right]\int_{-\infty}^{\infty} K^2(z)\,dz.$$

We can derive these approximations by using the standard method as in Pagan and Ullah (1999). For their validation, we require some conditions on the existence of moments, the decay rate of the mixing coefficients, and the shrinking rate of $\Delta$. For brevity, we omit such detailed conditions and derivations for the approximations, which are obtained analogously to the ones provided in Kanaya and Kristensen (2014), who also provide the precise meaning of "$\simeq$." $\mathcal{B}_{\mu}(x)$ and $\mathcal{B}_{\sigma^2}(x)$ correspond to the biases of the estimators. $\mathcal{V}_{\mu}(x)$ and $\mathcal{V}_{\sigma^2}(x)$ correspond to their variances, which are the same as the variances of the asymptotic normal distributions. Obviously, the damping function affects only the bias properties, and the variance components are of the same form as those of the standard NW estimators.

**Graphical Illustration of Effects of the Damping Function:** To see the effects of $B(\cdot)$ in finite samples, we compare the standard NW estimator $\tilde{\mu}(x)$ and its damped version $\hat{\mu}(x)$ with $B(x) = \exp\{-cx^2\}$, with $c = 0.1$ and 10. The following graphs are based on the same simulated path of the Ornstein-Uhlenbeck process $dX_s = \lambda(m - X_s)\,dt + \sigma dW_s$, where $(\lambda, m, \sigma^2) = (0.85837, 0.089102, 0.0021854)$, which is Aït-Sahalia's (1996a) estimate for short-term interest rates; $(T, \Delta, n) = (25, 1/52, 1300)$; $h = 4\hat{s}n^{-1/5}$ (this bandwidth has been used in Stanton, 1997; see p. 360 of Chapman and Pearson, 2000); and $\tilde{\mu}(x)$ and $\hat{\mu}(x)$ are evaluated over equally-spaced 50 grid points between 1 and 99 percentiles of the invariant distribution of the process (0.0061 and 0.1721, respectively). As we can see in the two figures, the NW estimates and their damped versions perfectly coincide for both $c = 0.1$ and 10, even with the one hundred times difference in the scale parameter $c$. While these are only based on one sample, it has been quite difficult to obtain some other samples/examples in which $\tilde{\mu}(x)$ and $\hat{\mu}(x)$ look significantly different for some other choices of date-generating-process, sample-size, and bandwidth settings. We have found a similar result for the diffusion function estimation. From these, we conclude that the effects of the damping function $B$ are not significant, where we again note that $B$'s effects are cancelled out between the numerator and denominator parts.

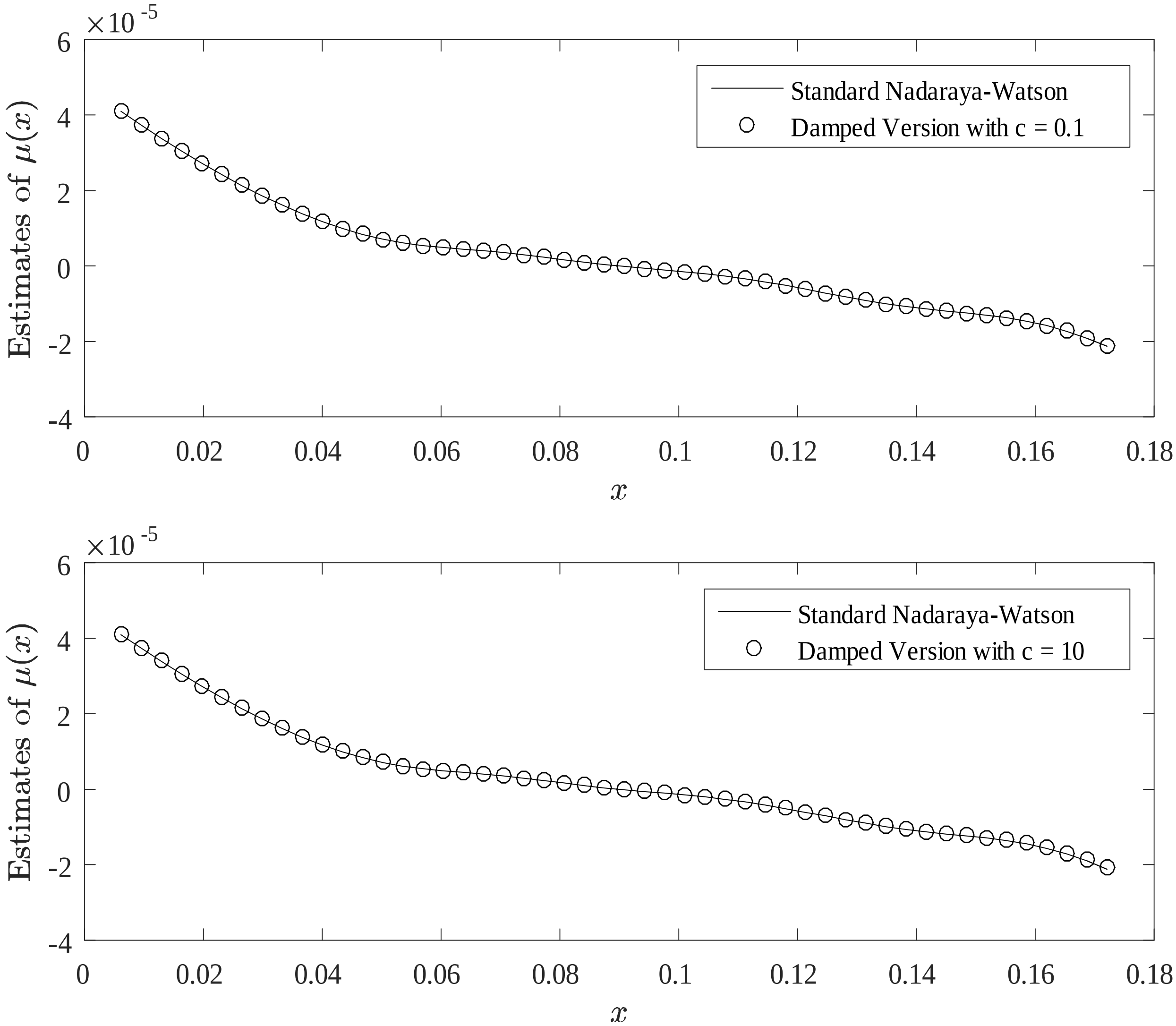


Figure: Comparison of Nadaraya-Watson estimates and their damped versions.

## S.5 Estimation of Non-Negative Valued Processes

In this section, we provide some discussion and results for processes with $I = (0, \infty)$ or $[0, \infty)$, as many parametric models for short-term interest rates have such a state space $I$. While we here focus on the case where the left end point of $I$ is 0, we can also think of some other cases (e.g., the left end point is finite and non-zero and/or the right one is also bounded), to which the results obtained for $I = (0, \infty)$ or $[0, \infty)$ carry over (if suitable modifications are made).

When $I$ has a finite end point, the choice of $B(x) = \exp\{-cx^2\}$, as considered in Sections 3, may not be sufficient. For example, Aït-Sahalia (1996b, 1999) considers a parametric diffusion model with

$I = (0, \infty)$ and the drift function

$$\bar{\mu}(x) := \alpha_0 + \alpha_1 x + \alpha_2^2 + \alpha_3 / x,$$

which diverges as $x \to 0$. To accommodate this kind of model, we can think of a damping function such as

$$B(x) = x^b \exp\{-x^2\}, \tag{S.28}$$

with some $b > 2$. This choice of $B$ allows us to establish the conclusion of Theorem 1, even with a drift function such as $\bar{\mu}(x)$.

When the left end point of $I$ is 0, we can also think of a case where the invariant density $\pi$ may not be bounded around $x = 0$. Processes with this feature can be easily found. Among others, we can think of the CIR process:

$$dX_s = \kappa(\theta - X_s)ds + \sigma\sqrt{X_s}dW_s, \tag{S.29}$$

with $\kappa, \theta, \sigma > 0$. If $2\kappa\theta/\sigma^2 > 0$, the stationary solution to (S.29) exists and its invariant density $\pi(\cdot)$ is given by the gamma distribution with $2\kappa\theta/\sigma^2$ and $\sigma^2/2\kappa$ being the shape and scale parameters, respectively. Given that $2\kappa\theta/\sigma^2 \in (0, 1)$, the left boundary $l = 0$ is attainable. In this case, the process has the gamma distribution as its invariant distribution by making $l$ instantaneously reflecting (for construction of this kind of process, see discussions in Section 2 on the behavior of the process running over an infinite time horizon, after the hit on $l = 0$; see also discussions on p. 441 of Forman and Sørensen, 2008). In this case, $\pi(x) \to \infty$ as $x \to 0$. This process with $2\kappa\theta/\sigma^2 \in (0, 1)$ satisfies the mixing condition in (A2.ii) with a geometric decay rate (i.e. $\beta = \infty$).[1] This sort of process with unbounded $\pi(x)$ at the end point can also be handled through the choice of $B$ as in (S.28), which can ensure the uniform boundedness of $B(x)\pi(x)$. We note that the integrability of the density implies $\pi(x) \sim x^{-q}$ (with some $q \in (0, 1)$) in the neighborhood of 0 and thus, the divergence rates of $\pi'(x)$ and $\pi''(x)$ around zero are also at most of the polynomial order.

For the case $\pi(x) \to \infty$ as $x \to 0$, we can still verify the uniform convergence of $\hat{\mu}(x)$ and $\hat{\sigma}^2(x)$ over $I = (0, \infty)$, given the damping function as in (S.28). In Theorem 2, we have supposed the uniform boundedness of $\pi(x)$, but this condition can be removed if we slightly change the relevant conditions,

[1] This can be checked by noting the following facts: i) the process is conservative and reversible (see Sections 8-9 in Kent, 1978); ii) the spectrum of its (infinitesimal) generator is discrete and has a gap left to zero, which is given by $\{\lambda_j\}$ with $\lambda_j = -\kappa j$ (see, e.g., p. 334 of KT81); and iii) both i) and ii) imply that $\{X_s\}$ is geometrically $\rho$-mixing (see discussions on p. 799 of Hansen and Scheinkman, 1995, as well as those in Hansen *et al.*, 1998). Note that in the case $2\kappa\theta/\sigma^2 > 1$, neither of the boundaries is attracting, and we can also check the geometric mixing property of the process by the same argument or by using, for example, Corollary 5.5 of Chen *et al.* (2010).

such as replacing "$\psi(\cdot) \in \mathcal{D}(B,\pi)$" with

$$\psi(\cdot) \in \mathcal{D}(B^{1/2},\pi) \;\; \text{and} \;\; \sup_{x\in(0,\infty)} B^{1/2}(x)\,\pi(x) < \infty, \tag{S.30}$$

for example. As seen in the proof of Theorem 2, we need to show that the following object:

$$\sup_{x\in(0,\infty)} (\Delta/Th) \sum_{j=1}^{n-1} K\left(\tfrac{X_{j\Delta}-x}{h}\right) \tag{S.31}$$

is bounded (in probability). However, it may not be so if $\pi(x)$ is unbounded. In this case, instead of (S.31), we consider

$$\sup_{x\in(0,\infty)} (\Delta/Th) \sum_{j=1}^{n-1} K\left(\tfrac{X_{j\Delta}-x}{h}\right) B^{1/2}(X_{j\Delta}),$$

whose $O_p$-boundedness is guaranteed under the condition that $\sup_{x\in(0,\infty)} B^{1/2}(x)\,\pi(x) < \infty$, where we note that even for the case with $I = [0,\infty)$ (i.e., the process may attain the point 0), the supremum with respect to $x$ needs to be taken over $(0,\infty)$ instead of $[0,\infty)$ if $\pi(0)$ is unbounded. The same argument applies to Theorem 3 for which the conditions in (S.30) can be used to relax the uniform boundedness of $\pi(x)$.

Under the condition in (S.30), we can still verify the same convergence rates of the variance effect terms as in Theorems 2-3 even with an unbounded $\pi(x)$ around $x = 0$. However, the boundedness of the end point in general slows down the convergence rate of the smoothing bias. This observation is summarized in the following remark:

**Remark S.1. (i)** Let $I = (0,\infty)$ or $[0,\infty)$. Then, given the conditions in (S.30) and the kernel function $K$ satisfying (B2), it holds that as $h \to 0$,

$$\sup_{x\in(0,\infty)} \left|\bar{G}_{n,T}(x) - H(x)\right| = O(h), \tag{S.32}$$

where $H(x) = B(x)\,\psi(x)\,\pi(x)$. This slower convergence occurs because we cannot use the symmetricity property of the kernel in the neighborhood of zero (i.e., $\int_{-\infty}^{\infty} xK(x)\,dx = 0$) to kill the first-order term of the smoothing bias (therefore, if $I \neq \mathbb{R}$, the use of higher-order kernels does not improve the uniform convergence rate over $I$). This kind of phenomenon, the so-called boundary bias, is observed if the end point of the support is bounded and the symmetric kernel is used (see arguments in Bouezmarni and Scaillet, 2005) while the boundary bias may be avoided by using asymmetric kernels as in Bouezmarni *et al.* (2005) and Gospodinov and Hirukawa (2012). We note that the supremum is taken over the open set $(0,\infty)$ in (S.32), which avoids the indefiniteness at $x = 0$ when $I = [0,\infty)$ and $\pi(0)$ is unbounded (we may have $[0,\infty)$ in (S.32) if 0 is a point attainable by the process and if $\pi(0) < \infty$).

**(ii)** If we use some special kernel and restrict the domain of $x$, we can recover the smoothing-bias rate of $h^2$. That is, when we assume that $K(\cdot)$ is a non-negative valued kernel with bounded support (resp.

the normal kernel), then under the conditions in (S.30),

$$\sup_{x\in[r(h),\infty)} \left|\bar{G}_{n,T}(x) - H(x)\right| = O\left(h^2\right), \tag{S.33}$$

as $h \to 0$, where $r(h)\,(\to 0)$ is a trimming sequence with $r(h) = -c_K h$ and $c_K := \inf\{c < 0 : K(c) > 0\}$ (resp. $r(h) = 2h\sqrt{\log(1/h)}$).

We can apply the results (S.32)-(S.33) to obtain the uniform rates of $\hat{\mu}(x)$ and $\hat{\sigma}^2(x)$ when $I = (0,\infty)$ or $[0,\infty)$.

*Proof of the Statements in Remark S.1.* Let $I = (0,\infty)$ or $I = [0,\infty)$. (i) We prove (S.32) here:

$$\begin{aligned}
\left|\bar{G}_{n,T}(x) - H(x)\right| &= \sup_{x\in(0,\infty)} \left|\int_{-x/h}^{\infty} K(q)\left[H(qh+x) - H(x)\right] dq\right| \\
&= \sup_{x\in(0,\infty)} \left|\int_{-x/h}^{\infty} K(q)\, H'(\tilde{x})\, qh dq\right| \\
&\le \sup_{x\in(0,\infty)} |H'(x)|\, h \int_{-\infty}^{\infty} |qK(q)|\, dq = O(h),
\end{aligned}$$

where we note that $qh + x \in I$ (if $x \in I$ and $q \in (-x/h,\infty)$) for the second inequality; and the third equality holds by the Taylor expansion ($\tilde{x}$ is on the line segment connecting $x$ to $qh+x$ and the expansion is valid for any $x \in (0,\infty)$).

(ii) Suppose that $K(\cdot)$ is a non-negative valued kernel whose support is bounded or it is the normal kernel. We prove (S.33) here:

$$\begin{aligned}
&\left|\bar{G}_{n,T}(x) - H(x)\right| \\
&= \sup_{x\in[r(h),\infty)} \left|\int_{-x/h}^{\infty} K(q)\left[H'(x)\, qh + (1/2)\, H''(\tilde{x})\,(qh)^2\right] dq\right| \\
&\le h \sup_{x\in(0,\infty)} |H'(x)| \times \sup_{x\in[r(h),\infty)} \left|\int_{-x/h}^{\infty} qK(q)\, dq\right| + \left(h^2/2\right) \int_{-x/h}^{\infty} q^2 |K(q)|\, dq \sup_{x\in(0,\infty)} |H''(x)|, \qquad \text{(S.34)}
\end{aligned}$$

where we can easily check that the second term on the RHS is $O(h^2)$. If the support of $K(\cdot)$ is bounded and $r(h) = -c_K h(> 0)$ with $c_K = \inf\{c < 0 : K(c) > 0\}$, then the first-order term on the RHS is zero since $\int_{-x/h}^{\infty} qK(q)\, dq = \int_{c_K}^{\infty} qK(q)\, dq = 0$ for any $x \in [-c_K h,\infty)$. If $K(\cdot)$ is the normal kernel and $r(h) = 2h\sqrt{\log(1/h)}$, then

$$\begin{aligned}
\sup_{x\in[r(h),\infty)} \left|\int_{-x/h}^{\infty} qK(q)\, dq\right| &= \sup_{x\in[r(h),\infty)} \left|\int_{x/h}^{\infty} qK(q)\, dq\right| \\
&\le (2\pi)^{-1/2} \int_{r(h)/h}^{\infty} \left[q\exp\left\{-q^2/4\right\}\right] \exp\left\{-q^2/4\right\} dq \\
&\le (2\pi)^{-1/2} \int_{0}^{\infty} q\exp\left\{-q^2/4\right\} dq \times \exp\{-\left(r(h)/h\right)^2/4\} \\
&= O(h),
\end{aligned}$$

where the first equality holds since $\int_{-\infty}^{\infty} qK(q)\, dq = 0$. From these arguments, we can show that the RHS of (S.34) is $O(h^2)$ to complete the proof. □

# References for Supplementary Material


1. Aït-Sahalia, Y. (1996b) Testing continuous-time models of the spot interest rate, *Review of Financial Studies* 9, 385-426.

2. Aït-sahalia, Y. (1999) Transition densities for interest rate and other nonlinear diffusions, *Journal of Finance* 54, 1361-1395.

3. Bouezmarni, T. & O. Scaillet (2005) Consistency of asymmetric kernel density estimators and smoothed histograms with application to income data, *Econometric Theory* 21, 390-412.

4. Chapman, D.A. & Pearson, N.D. (2003) Is the short rate drift actually nonlinear? *Journal of Finance* 55, 355-388.

5. Forman, J.Y. & M. Sørensen (2008) The Pearson diffusions: A class of statistically tractable diffusion processes, *Scandinavian Journal of Statistics* 35, 438-465.

6. Gospodinov, N. & M. Hirukawa (2012) Nonparametric estimation of scalar diffusion processes of interest rates using asymmetric kernels, *Journal of of Empirical Finance* 19, 595-609.

7. Kanaya, S. & D. Kristensen (2014) Optimal sampling and bandwidth selection for kernel estimators of diffusion processes, Working paper, University of Aarhus and University College London.

8. Kosorok, M. (2008) *Introduction to Empirical Processes and Semiparametric Inference*, Springer.

9. Pagan, A. & A. Ullah (1999) *Nonparametric Econometrics*, Cambridge University Press.

10. Stanton, R. (1997) A nonparametric model of term structure dynamics and the market price of interest rate risk, *Journal of Finance* 52, 1973-2002.

11. van der Vaart, A.W. & J. A. Wellner (1996) *Weak Convergence and Empirical Processes: With Applications to Statistics*, Springer.