Supplement to "Classification of Nonparametric Regression Functions in Longitudinal Data Models"

Michael Vogt
University of Bonn

Oliver Linton
University of Cambridge

August 23, 2015

In this supplement, we investigate the finite sample performance of the bandwidth selection rule from Section 4.3 by means of a simulation study. Moreover, we provide the proofs and technical details that are omitted in the paper.

1 Additional Simulations

We now investigate the performance of the bandwidth selection procedure proposed in Section 4.3 of the paper. To do so, we pick one of the clusters from our simulation setup of Section 5 and simulate data from this cluster. In particular, we consider the cluster G_5 with $n_5 = |G_5| = 10$ and $g_5(x) = 1.75 \arctan(5(x - 0.6)) + 0.75$ and draw data from the model equation

$$Y_{it} = g_5(X_{it}) + \varepsilon_{it} \quad (1 \le i \le n_5,\, 1 \le t \le T), \tag{S.1}$$

where the model variables X_{it} and ε_{it} are generated in exactly the same way as in the simulations.

As discussed in Section 4.3 of the paper, our bandwidth selection procedure is based on minimizing the residual sum of squares criterion $\mathrm{RSS}_i^{(j)}(h)$ for different pairs of indices (i, j). More precisely, we define our bandwidth selector by

$$\widehat{h} = \frac{1}{L} \sum_{1 \le \ell \le L} \widehat{h}_{i_{2\ell-1}}^{(i_{2\ell})},$$

where $L = n_5/2$ and $\widehat{h}_i^{(j)} = \mathrm{argmin}_h \mathrm{RSS}_i^{(j)}(h)$. As already discussed in Section 4.3, $\widehat{h}$ can be regarded as an approximation to the optimal bandwidth h^* in a mean integrated squared error sense, which is defined as $h^* = \arg\min_h \mathrm{MISE}_i(h)$. Note that under the conditions of Section 4.3, $\mathrm{MISE}_i(h)$ is the same for all $1 \le i \le n_5$ and thus h^* is a group-wide optimal bandwidth independent of i.

To examine the finite sample behaviour of the bandwidth estimator $\widehat{h}$, we draw $N = 1000$ samples from the setting (S.1) for each time series length $T \in \{100, 150, 200, 500\}$ and compute the bandwidth $\widehat{h}$ for each simulated sample. To do so, we define an equidistant grid $\mathcal{G}$ of step length 0.01 which spans the interval $[0.025, 0.5]$ and minimize the criterion functions $\mathrm{RSS}_i^{(j)}(h)$ over all bandwidth values $h \in \mathcal{G}$. The optimal bandwidth h^* can be calculated to be approximately $0.225, 0.205, 0.195, 0.165$ for the time series lengths $T = 100, 150, 200, 500$, respectively.

Figure 1 summarizes the simulation results. Each panel shows the distribution of the differences $h^* - \widehat{h}$ for a specific time series length T. In particular, the bars in the plots give the number of simulations (out of total of 1000) in which the difference $h^* - \widehat{h}$ takes a certain value. The plots of Figure 1 suggest that $\widehat{h}$ approximates the optimal value h^* reasonably well. They also make visible that the precision of the estimator $\widehat{h}$ improves quite slowly as the sample size grows. This is not surprising as the convergence rate of standard bandwidth selectors (based on cross-validation or penalization techniques) is known to be very slow; see e.g. Härdle et al. (1988).

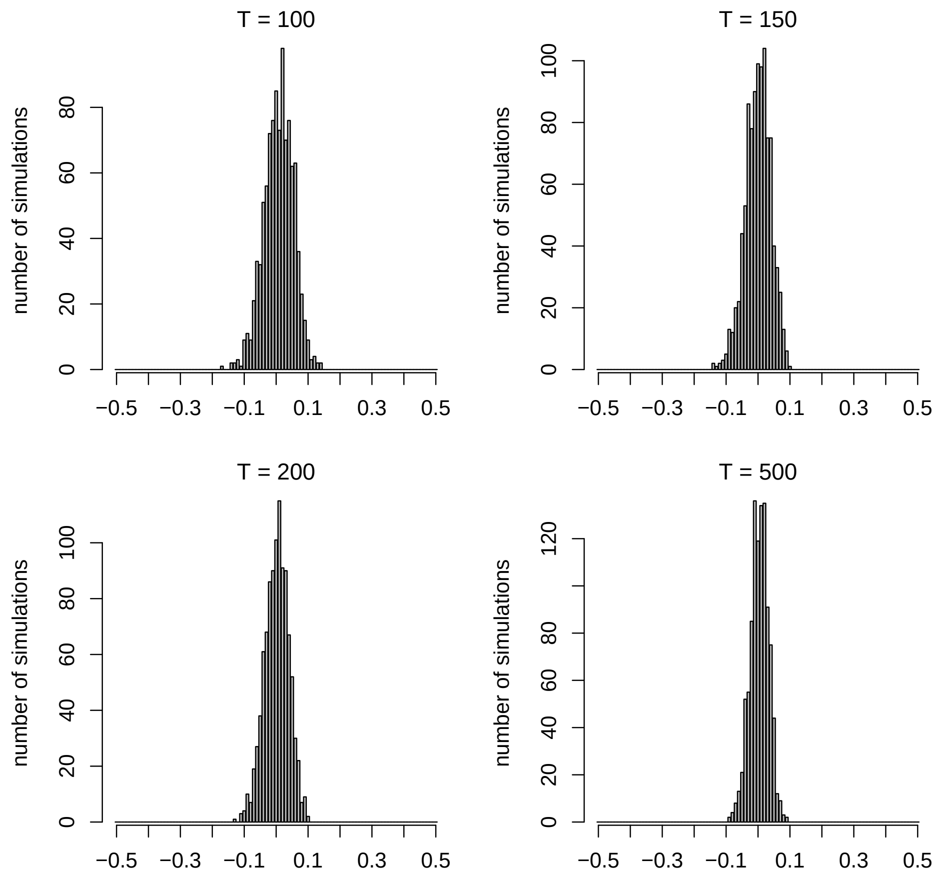

Figure 1: Simulation results for the bandwidth selection procedure from Section 4.3. Each panel depicts the distribution of the differences $h^* - \widehat{h}$ for a specific time series length T. The optimal bandwidth h^* is approximately $0.225, 0.205, 0.195, 0.165$ for $T = 100, 150, 200, 500$, respectively.

2 Technical Details

In this section, we provide the proofs and technical details omitted in the paper. Throughout the section, the symbol C denotes a universal real constant which may take a different value on each occurrence.

Proof of Lemma A.1

To prove the lemma, we modify standard arguments to derive uniform convergence rates for kernel estimators, which can be found e.g. in Masry (1996), Bosq (1998) or Hansen (2008). These arguments are designed to derive the rate of $\sup_x |\widehat{m}_i(x) - m_i(x)|$ for a fixed individual i. They thus yield the rate which is uniform over x but pointwise in i. In contrast to this, we aim to derive the rate which is uniform both over x and i. To do so, we write

$$\widehat{m}_i(x) - m_i(x) = \big[Q_{i,V}(x) + Q_{i,B}(x) - Q_{i,\gamma}(x)\big] \big/ \widehat{f}_i(x) - \overline{Q}_i + \overline{\overline{Q}}_i,$$

where

$$\begin{aligned}
Q_{i,V}(x) &= \frac{1}{T}\sum_{t=1}^T W_h(X_{it} - x)\varepsilon_{it} \\
Q_{i,B}(x) &= \frac{1}{T}\sum_{t=1}^T W_h(X_{it} - x)\big[m_i(X_{it}) - m_i(x)\big] \\
Q_{i,\gamma}(x) &= \frac{1}{T}\sum_{t=1}^T W_h(X_{it} - x)\Big(\frac{1}{n-1}\sum_{\substack{j=1\\ j\neq i}}^n \big[m_j(X_{jt}) + \varepsilon_{jt}\big]\Big) \\
\overline{Q}_i &= \frac{1}{T}\sum_{t=1}^T \big[m_i(X_{it}) + \varepsilon_{it}\big] \\
\overline{\overline{Q}}_i &= \frac{1}{(n-1)T}\sum_{\substack{j=1\\ j\neq i}}^n \sum_{t=1}^T \big[m_j(X_{jt}) + \varepsilon_{jt}\big]
\end{aligned}$$

and $\widehat{f}_i(x) = T^{-1}\sum_{t=1}^T W_h(X_{it} - x)$. In what follows, we show that

$$\max_{1\le i\le n}\sup_{x\in[0,1]} \big|Q_{i,V}(x)\big| = O_p(a_{n,T}) \tag{S.2}$$

$$\max_{1\le i\le n}\sup_{x\in[0,1]} \big|Q_{i,B}(x) - \mathbb{E}[Q_{i,B}(x)]\big| = O_p(a_{n,T}) \tag{S.3}$$

$$\max_{1\le i\le n}\sup_{x\in[0,1]} \big|Q_{i,\gamma}(x)\big| = O_p(a_{n,T}) \tag{S.4}$$

$$\max_{1\le i\le n}\sup_{x\in[0,1]} \big|\widehat{f}_i(x) - \mathbb{E}[\widehat{f}_i(x)]\big| = O_p\Big(\sqrt{\frac{\log T}{Th}}\Big). \tag{S.5}$$

Moreover, standard bias calculations yield that $\max_{1\le i\le n}\sup_{x\in I_h}|\mathbb{E}[Q_{i,B}(x)]| = O(h^2)$ along with $\max_{1\le i\le n}\sup_{x\in[0,1]\setminus I_h}|\mathbb{E}[Q_{i,B}(x)]| = O(h)$. Analogously, $\max_{1\le i\le n}\sup_{x\in I_h}$ $|\mathbb{E}[\widehat{f}_i(x)]| = O(h^2)$ and $\max_{1\le i\le n}\sup_{x\in[0,1]\setminus I_h}|\mathbb{E}[\widehat{f}_i(x)]| = O(h)$. Finally, a simplified version of the arguments for (S.2) shows that $\max_{1\le i\le n}|\overline{Q}_i| = O_p(a_{n,T})$ as well as $\max_{1\le i\le n}|\overline{\overline{Q}}_i| = O_p(a_{n,T})$. Lemma A.1 immediately follows upon combining (S.2)–(S.5) with these statements. □

Proof of (S.2). Set $\psi_{n,T} = (nT)^{1/(\theta-\delta)}$, where θ is introduced in (C3) and $\delta > 0$ is a small positive number. Moreover, define

$$\begin{aligned}\varepsilon_{it}^{\le} &= \varepsilon_{it}1\big(|\varepsilon_{it}| \le \psi_{n,T}\big)\\ \varepsilon_{it}^{>} &= \varepsilon_{it}1\big(|\varepsilon_{it}| > \psi_{n,T}\big).\end{aligned}$$

With this notation at hand, we can rewrite the term $Q_{i,V}(x)$ as

$$Q_{i,V}(x) = \sum_{t=1}^T Z_{it,T}^{\le}(x) + \sum_{t=1}^T Z_{it,T}^{>}(x),$$

where

$$\begin{aligned}Z_{it,T}^{\le}(x) &= \big(W_h(X_{it}-x)\varepsilon_{it}^{\le} - \mathbb{E}\big[W_h(X_{it}-x)\varepsilon_{it}^{\le}\big]\big)/T\\ Z_{it,T}^{>}(x) &= \big(W_h(X_{it}-x)\varepsilon_{it}^{>} - \mathbb{E}\big[W_h(X_{it}-x)\varepsilon_{it}^{>}\big]\big)/T.\end{aligned}$$

We thus split $Q_{i,V}(x)$ into the "interior part" $\sum_{t=1}^T Z_{it,T}^{\le}(x)$ and the "tail part" $\sum_{t=1}^T Z_{it,T}^{>}(x)$. This parallels the standard arguments for deriving the convergence rate of $\sup_{x\in[0,1]}|Q_{i,V}(x)|$ for a fixed individual i. As we maximize over i, we however choose the truncation sequence $\psi_{n,T}$ to go to infinity much faster than in the standard case with a fixed i.

We now proceed in several steps. To start with, we show that

$$\max_{1\le i\le n}\sup_{x\in[0,1]}\Big|\sum_{t=1}^T Z_{it,T}^{>}(x)\Big| = O_p(a_{n,T}). \tag{S.6}$$

This can be achieved as follows:

$$\begin{aligned}&\mathbb{P}\Big(\max_{1\le i\le n}\sup_{x\in[0,1]}\Big|\sum_{t=1}^T Z_{it,T}^{>}(x)\Big| > a_{n,T}\Big)\\ &\le \sum_{i=1}^n \mathbb{P}\Big(\sup_{x\in[0,1]}\Big|\frac{1}{T}\sum_{t=1}^T W_h(X_{it}-x)\varepsilon_{it}^{>}\Big| > \frac{a_{n,T}}{2}\Big)\\ &\quad + \sum_{i=1}^n \mathbb{P}\Big(\sup_{x\in[0,1]}\Big|\frac{1}{T}\sum_{t=1}^T \mathbb{E}\big[W_h(X_{it}-x)\varepsilon_{it}^{>}\big]\Big| > \frac{a_{n,T}}{2}\Big).\end{aligned}$$

With the help of assumption (C3), we obtain that

$$\begin{aligned}
&\sum_{i=1}^{n} \mathbb{P}\Big(\sup_{x \in [0,1]} \Big| \frac{1}{T} \sum_{t=1}^{T} W_h(X_{it} - x) \varepsilon_{it}^{>} \Big| > \frac{a_{n,T}}{2} \Big) \\
&\le \sum_{i=1}^{n} \mathbb{P}\Big(|\varepsilon_{it}| > \psi_{n,T} \text{ for some } 1 \le t \le T \Big) \le C (nT)^{1-\frac{\theta}{\theta-\delta}} = o(1).
\end{aligned}$$

Once more applying (C3), it can be seen that

$$\begin{aligned}
\big| \mathbb{E}\big[W_h(X_{it} - x) \varepsilon_{it}^{>} \big] \big| &\le \mathbb{E}\Bigg[W_h(X_{it} - x) \mathbb{E}\Big[\frac{|\varepsilon_{it}|^{\theta}}{\psi_{n,T}^{\theta-1}} 1(|\varepsilon_{it}| > \psi_{n,T}) \Big| X_{it} \Big] \Bigg] \\
&\le C (nT)^{-\frac{\theta-1}{\theta-\delta}}
\end{aligned}$$

with some constant C independent of x. Since $C(nT)^{-\frac{\theta-1}{\theta-\delta}} < a_{n,T}/2$ as the sample size grows large, we arrive at

$$\sum_{i=1}^{n} \mathbb{P}\Big(\sup_{x \in [0,1]} \Big| \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}\big[W_h(X_{it} - x) \varepsilon_{it}^{>} \big] \Big| > \frac{a_{n,T}}{2} \Big) = 0$$

for sufficiently large sample sizes. This yields (S.6).

We next have a closer look at the expression $\sum_{t=1}^{T} Z_{it,T}^{\le}(x)$. Let $0 = x_0 < x_1 < \ldots < x_L = 1$ be an equidistant grid of points covering the unit interval and set $L = L_{n,T} = \psi_{n,T}/(a_{n,T} h^2)$. Exploiting the Lipschitz continuity of the kernel W, straightforward calculations yield that

$$\max_{1 \le i \le n} \sup_{x \in [0,1]} \Big| \sum_{t=1}^{T} Z_{it,T}^{\le}(x) \Big| \le \max_{1 \le i \le n} \max_{1 \le \ell \le L} \Big| \sum_{t=1}^{T} Z_{it,T}^{\le}(x_\ell) \Big| + C a_{n,T}. \tag{S.7}$$

We can thus replace the supremum over x by a maximum over the grid points x_ℓ. Moreover, it holds that

$$\begin{aligned}
&\mathbb{P}\Big(\max_{1 \le i \le n} \max_{1 \le \ell \le L} \Big| \sum_{t=1}^{T} Z_{it,T}^{\le}(x_\ell) \Big| > C_0 a_{n,T} \Big) \\
&\qquad \le \sum_{i=1}^{n} \sum_{\ell=1}^{L} \mathbb{P}\Big(\Big| \sum_{t=1}^{T} Z_{it,T}^{\le}(x_\ell) \Big| > C_0 a_{n,T} \Big),
\end{aligned} \tag{S.8}$$

where C_0 is a sufficiently large constant to be specified later on. In what follows, we show that for each fixed x_ℓ,

$$\mathbb{P}\Big(\Big| \sum_{t=1}^{T} Z_{it,T}^{\le}(x_\ell) \Big| > C_0 a_{n,T} \Big) \le C T^{-r}, \tag{S.9}$$

where the constants C and r are independent of x_ℓ and $r > 0$ can be chosen arbitrarily large provided that C_0 is picked sufficiently large. Plugging (S.9) into (S.8) and combining the result with (S.7), we arrive at

$$\max_{1 \le i \le n} \sup_{x \in [0,1]} \Big| \sum_{t=1}^T Z_{it,T}^{\le}(x) \Big| = O_p(a_{n,T}), \tag{S.10}$$

which completes the proof.

It thus remains to prove (S.9). To do so, we split the term $\sum_{t=1}^T Z_{it,T}^{\le}(x_\ell)$ into blocks as follows:

$$\sum_{t=1}^T Z_{it,T}^{\le}(x_\ell) = \sum_{s=1}^{q_{n,T}} B_{2s-1} + \sum_{s=1}^{q_{n,T}} B_{2s}$$

with $B_s = \sum_{t=(s-1)r_{n,T}+1}^{sr_{n,T}} Z_{it,T}^{\le}(x_\ell)$, where $2q_{n,T}$ is the number of blocks and $r_{n,T} = T/(2q_{n,T})$ is the block length. In particular, we choose the block length such that $r_{n,T} = O(T^\eta)$ for some small $\eta > 0$. With this notation at hand, we get

$$\begin{aligned}\mathbb{P}\Big(\Big| \sum_{t=1}^T Z_{it,T}^{\le}(x_\ell)\Big| > C_0 a_{n,T}\Big) \le \mathbb{P}\Big(\Big| \sum_{s=1}^{q_{n,T}} B_{2s-1}\Big| > \frac{C_0}{2} a_{n,T}\Big) \\ + \mathbb{P}\Big(\Big| \sum_{s=1}^{q_{n,T}} B_{2s}\Big| > \frac{C_0}{2} a_{n,T}\Big).\end{aligned}$$

As the two terms on the right-hand side can be treated analogously, we focus attention to the first one. By Bradley's lemma (see Lemma 1.2 in Bosq (1998)), we can construct a sequence of random variables $B_1^*, B_3^*, \ldots$ such that (a) $B_1^*, B_3^*, \ldots$ are independent, (b) B_{2s-1} and B_{2s-1}^* have the same distribution for each s, and (c) for $0 < \mu \le \|B_{2s-1}\|_\infty$, $\mathbb{P}(|B_{2s-1}^* - B_{2s-1}| > \mu) \le 18(\|B_{2s-1}\|_\infty/\mu)^{1/2} \alpha(r_{n,T})$. With these variables, we obtain the bound

$$\mathbb{P}\Big(\Big| \sum_{s=1}^{q_{n,T}} B_{2s-1}\Big| > \frac{C_0}{2} a_{n,T}\Big) \le P_1 + P_2,$$

where

$$\begin{aligned}P_1 &= \mathbb{P}\Big(\Big| \sum_{s=1}^{q_{n,T}} B_{2s-1}^*\Big| > \frac{C_0}{4} a_{n,T}\Big) \\ P_2 &= \mathbb{P}\Big(\Big| \sum_{s=1}^{q_{n,T}} \big(B_{2s-1} - B_{2s-1}^*\big)\Big| > \frac{C_0}{4} a_{n,T}\Big).\end{aligned}$$

Using (c) together with the fact that the mixing coefficients $\alpha(\cdot)$ decay to zero exponentially fast, it is not difficult to see that P_2 converges to zero at an arbitrarily fast polynomial rate. To deal with P_1, we make use of the following three facts:

(i) For a random variable B and $\lambda > 0$, Markov's inequality yields that

$$\mathbb{P}\big(\pm B > \delta\big) \leq \frac{\mathbb{E}\exp(\pm\lambda B)}{\exp(\lambda\delta)}.$$

(ii) We have that $|B_{2s-1}| \leq C_B r_{n,T}\psi_{n,T}/(Th)$ for some constant $C_B > 0$. Define $\lambda_{n,T} = Th/(2C_B r_{n,T}\psi_{n,T})$, which implies that $\lambda_{n,T}|B_{2s-1}| \leq 1/2$. As $\exp(x) \leq 1 + x + x^2$ for $|x| \leq 1/2$, we get that

$$\mathbb{E}\Big[\exp\big(\pm \lambda_{n,T} B_{2s-1}\big)\Big] \leq 1 + \lambda_{n,T}^2 \mathbb{E}\big[(B_{2s-1})^2\big] \leq \exp\big(\lambda_{n,T}^2 \mathbb{E}\big[(B_{2s-1})^2\big]\big)$$

along with

$$\mathbb{E}\Big[\exp\big(\pm \lambda_{n,T} B^*_{2s-1}\big)\Big] \leq \exp\big(\lambda_{n,T}^2 \mathbb{E}\big[(B^*_{2s-1})^2\big]\big).$$

(iii) Standard calculations for kernel estimators yield that

$$\sum_{s=1}^{q_{n,T}} \mathbb{E}\big[(B^*_{2s-1})^2\big] \leq \frac{C}{Th}.$$

Using (i)–(iii), we arrive at

$$\begin{aligned}
&\mathbb{P}\Big(\Big|\sum_{s=1}^{q_{n,T}} B^*_{2s-1}\Big| > \frac{C_0}{4} a_{n,T}\Big) \\
&\leq \mathbb{P}\Big(\sum_{s=1}^{q_{n,T}} B^*_{2s-1} > \frac{C_0}{4} a_{n,T}\Big) + \mathbb{P}\Big(- \sum_{s=1}^{q_{n,T}} B^*_{2s-1} > \frac{C_0}{4} a_{n,T}\Big) \\
&\leq \exp\Big(- \frac{C_0}{4}\lambda_{n,T} a_{n,T}\Big) \Big\{ \mathbb{E}\Big[\exp\Big(\lambda_{n,T} \sum_{s=1}^{q_{n,T}} B^*_{2s-1}\Big)\Big] + \mathbb{E}\Big[\exp\Big(- \lambda_{n,T} \sum_{s=1}^{q_{n,T}} B^*_{2s-1}\Big)\Big]\Big\} \\
&\leq \exp\Big(- \frac{C_0}{4}\lambda_{n,T} a_{n,T}\Big) \Big\{ \prod_{s=1}^{q_{n,T}} \mathbb{E}\Big[\exp\big(\lambda_{n,T} B^*_{2s-1}\big)\Big] + \prod_{s=1}^{q_{n,T}} \mathbb{E}\Big[\exp\big(- \lambda_{n,T} B^*_{2s-1}\big)\Big]\Big\} \\
&\leq 2\exp\Big(- \frac{C_0}{4}\lambda_{n,T} a_{n,T}\Big) \prod_{s=1}^{q_{n,T}} \exp\Big(\lambda_{n,T}^2 \mathbb{E}\big[(B^*_{2s-1})^2\big]\Big) \\
&= 2\exp\Big(- \frac{C_0}{4}\lambda_{n,T} a_{n,T}\Big) \exp\Big(\lambda_{n,T}^2 \sum_{s=1}^{q_{n,T}} \mathbb{E}\big[(B^*_{2s-1})^2\big]\Big) \\
&\leq 2\exp\Big(- \frac{C_0}{4}\lambda_{n,T} a_{n,T} + \lambda_{n,T}^2 \frac{C}{Th}\Big).
\end{aligned}$$

Recalling that $n/T \leq C$ and $T^{2/5}h \to \infty$ by assumption, setting θ to a value slightly larger than 4 and supposing that $a_{n,T} = T^{-1/10}$, it follows that

$$\exp\Big(- \frac{C_0}{4}\lambda_{n,T} a_{n,T} + \lambda_{n,T}^2 \frac{C}{Th}\Big) \leq T^{-r}, \tag{S.11}$$

where the constant $r > 0$ can be made arbitrarily large by picking C_0 large enough. If we strengthen (C3) to be satisfied for some $\theta > 20/3$ and choose the block length to be $r_{n,T} = \sqrt{(Th)/(\psi_{n,T}^2 \log T)}$, (S.11) even holds for $a_{n,T} = \sqrt{\log T/(Th)}$. From (S.11), it immediately follows that $P_1 \le CT^{-r}$, which in turn completes the proof of (S.9). $\square$

Proof of (S.3). The statement follows essentially by the same arguments as those for the proof of (S.2). $\square$

Proof of (S.4). Define $Z_{it} = (n-1)^{-1} \sum_{j=1, j \ne i}^{n} (m_j(X_{jt}) + \varepsilon_{jt})$ and write

$$Q_{i,\gamma}(x) = \frac{1}{T} \sum_{t=1}^{T} W_h(X_{it} - x) Z_{it}. \tag{S.12}$$

By construction, the time series processes $\{X_{it} : 1 \le t \le T\}$ and $\{Z_{it} : 1 \le t \le T\}$ are independent of each other. Moreover, by Theorem 5.2 in Bradley (2005), the process $\{Z_{it} : 1 \le t \le T\}$ is strongly mixing with mixing coefficients that are bounded by $n\alpha(k)$. (S.4) can thus be shown by applying the arguments from the proof of (S.2) to (S.12). $\square$

Proof of (S.5). The overall strategy is the same as that for the proof of (S.2). There is however one important difference: In the proof of (S.2), we have examined a kernel average of the form $T^{-1} \sum_{t=1}^{T} W_h(X_{it} - x) Z_{it}$ with $Z_{it} = \varepsilon_{it}$. As the variables ε_{it} have unbounded support in general, we have introduced the truncation sequence $\psi_{n,T}$ and have split ε_{it} into the two parts $\varepsilon_{it}^{\le}$ and $\varepsilon_{it}^{>}$. Here in contrast, we are concerned with the case $Z_{it} \equiv 1$. Importantly, the random variables $Z_{it} \equiv 1$ are bounded, implying that we do not have to truncate them at all. Keeping this in mind and going step by step along the proof of (S.2), we arrive at (S.5). $\square$

Proof of Lemma A.2

Under the conditions of the lemma, it holds that for any pair of indices $i, j \in G_k$,

$$\widehat{\Delta}_{ij} = \int \Big(\frac{Q_{i,V}(x) + Q_{i,B}(x)}{\widehat{f}_i(x)} - \frac{Q_{j,V}(x) + Q_{j,B}(x)}{\widehat{f}_j(x)} \Big)^2 \pi(x) dx$$

with $Q_{i,V}(x)$, $Q_{i,B}(x)$ and $\widehat{f}_i(x)$ defined as in the proof of Lemma A.1. Using the arguments from Lemma A.1, one can show that

$$\max_{i \in G_k} \sup_{x \in [0,1]} \big| Q_{i,V}(x) \big| = O_p\Big(\sqrt{\frac{\log T}{Th}} \Big)$$

$$\max_{i \in G_k} \sup_{x \in [0,1]} \big| Q_{i,B}(x) - \mathbb{E}[Q_{i,B}(x)] \big| = O_p\Big(h \sqrt{\frac{\log T}{Th}} \Big)$$

$$\max_{i\in G_k}\sup_{x\in I_h}\big|\widehat{f}_i(x)-f_i(x)\big| = O_p\Big(\sqrt{\frac{\log T}{Th}}+h^2\Big)$$

and

$$\mathbb{E}[Q_{i,B}(x)] = h^2\big(\textstyle\int W(\varphi)\varphi^2 d\varphi\big)\Big(m_i'(x)f_i'(x)+\frac{m_i''(x)f_i(x)}{2}\Big)+O(h^3)$$

uniformly for $i \in G_k$ and $x \in I_h$. Applying these uniform convergence results and noting that $\max_{i,j\in G_k}\widehat{\Delta}_{ij} = \max_{i,j\in G_k, i<j}\widehat{\Delta}_{ij}$, it is not difficult to see that

$$\max_{i,j\in G_k}\widehat{\Delta}_{ij} = \max_{\substack{i,j\in G_k\\ i<j}}\int\Big(\frac{Q_{i,V}(x)}{f_i(x)}-\frac{Q_{j,V}(x)}{f_j(x)}\Big)^2\pi(x)dx + o_p\Big(\frac{1}{Th^{1/2}}\Big).$$

Next define

$$U_{i,T} = \sum_{s,t=1}^{T} a_{st}^{(i)}\varepsilon_{is}\varepsilon_{it},$$

where $a_{st}^{(i)} = T^{-2}\int W_h(X_{is}-x)W_h(X_{it}-x)\pi(x)/f_i^2(x)dx$ for $s\neq t$ and $a_{st}^{(i)} = 0$ for $s=t$. Similarly, for $i \neq j$, let

$$U_{ij,T} = \sum_{s,t=1}^{T} a_{st}^{(ij)}\varepsilon_{is}\varepsilon_{jt}$$

with $a_{st}^{(ij)} = T^{-2}\int W_h(X_{is}-x)W_h(X_{jt}-x)\pi(x)/(f_i(x)f_j(x))dx$ and define

$$B_{i,T} = \int \frac{1}{T^2}\sum_{t=1}^{T} W_h^2(X_{it}-x)\varepsilon_{it}^2\frac{\pi(x)}{f_i^2(x)}dx.$$

With these definitions at hand, we can write

$$\max_{\substack{i,j\in G_k\\ i<j}}\int\Big(\frac{Q_{i,V}(x)}{f_i(x)}-\frac{Q_{j,V}(x)}{f_j(x)}\Big)^2\pi(x)dx = \max_{\substack{i,j\in G_k\\ i<j}}\big\{U_{i,T}-2U_{ij,T}+U_{j,T}+B_{i,T}+B_{j,T}\big\}.$$

Below we show that

$$\max_{1\le i\le n}\big|U_{i,T}\big| = O_p\Big(\frac{\log T}{Th^{1/2}}\Big) \tag{S.13}$$

$$\max_{1\le i<j\le n}\big|U_{ij,T}\big| = O_p\Big(\frac{\log T}{Th^{1/2}}\Big). \tag{S.14}$$

Moreover, similar arguments as those for the proof of Lemma A.1 yield that

$$\max_{1\le i\le n}\big|B_{i,T}-\mathbb{E}[B_{i,T}]\big| = o_p\Big(\frac{1}{Th^{1/2}}\Big) \tag{S.15}$$

$$\mathbb{E}[B_{i,T}] = \frac{1}{Th}\big(\textstyle\int W^2(\varphi)d\varphi\big)\int\frac{\sigma_i^2(x)\pi(x)}{f_i(x)}dx + O\Big(\frac{1}{T}\Big) \tag{S.16}$$

uniformly in i. Combining (S.13)–(S.16) and noting that $\mathbb{E}[B_{i,T}] + \mathbb{E}[B_{j,T}] = \mathcal{B}_{ij}/(Th)$ $+O(T^{-1})$ uniformly in i and j, we arrive at

$$\begin{aligned}\max_{i,j\in G_k} \widehat{\Delta}_{ij} &= \max_{\substack{i,j\in G_k\\ i<j}} \int \Big(\frac{Q_{i,V}(x)}{f_i(x)} - \frac{Q_{j,V}(x)}{f_j(x)}\Big)^2 \pi(x)dx + o_p\Big(\frac{1}{Th^{1/2}}\Big) \\ &= \max_{\substack{i,j\in G_k\\ i<j}} \frac{\mathcal{B}_{ij}}{Th} + O_p\Big(\frac{\log T}{Th^{1/2}}\Big).\end{aligned}$$

To complete the proof, it thus remains to verify (S.13) and (S.14).

Proof of (S.13). Define the matrix $A_T^{(i)} = (|a_{st}^{(i)}|)_{s,t=1}^T$ and $\Lambda_T^{(i)} = \sum_{s,t=1}^T (a_{st}^{(i)})^2$. We first show that

$$\max_{1\le i\le n} \|A_T^{(i)}\| = O_p\Big(\frac{1}{T}\Big) \tag{S.17}$$

$$\max_{1\le i\le n} \Lambda_T^{(i)} = O_p\Big(\frac{1}{T^2 h}\Big), \tag{S.18}$$

where $\|A_T^{(i)}\|$ denotes the spectral norm of $A_T^{(i)}$. By definition, $\|A_T^{(i)}\|$ is the largest absolute eigenvalue of $A_T^{(i)}$. As the diagonal elements $|a_{tt}^{(i)}|$ of $A_T^{(i)}$ are all zero, Gerschgorin's theorem says that the largest absolute eigenvalue of $A_T^{(i)}$ is bounded by

$$\overline{\lambda}^{(i)} = \max_{1\le s\le T} \sum_{t=1}^T |a_{st}^{(i)}|.$$

Standard calculations yield that

$$\overline{\lambda}^{(i)} \le \frac{C}{T} \max_{1\le s\le T} \frac{1}{T}\sum_{t=1}^T \mathcal{W}_h(X_{is} - X_{it}),$$

where $\mathcal{W}_h(x) = h^{-1}\mathcal{W}(x/h)$ and $\mathcal{W}(x) = \int_{-C_1}^{C_1} W(x+\varphi)d\varphi$. One can easily show that (a) $|\mathcal{W}(x)| \le C$, (b) $\mathcal{W}(x) = 0$ for all $|x| > 2C_1$, and (c) $|\mathcal{W}(x) - \mathcal{W}(x')| \le L|x - x'|$ for some constant L. Hence, similar arguments as those from Lemma A.1 yield that

$$\begin{aligned}\overline{\lambda}^{(i)} &\le \frac{C}{T} \sup_{x\in[0,1]} \frac{1}{T}\sum_{t=1}^T \mathcal{W}_h(x - X_{it}) \\ &\le \frac{C}{T} \sup_{x\in[0,1]} \Big|\frac{1}{T}\sum_{t=1}^T \mathbb{E}\big[\mathcal{W}_h(x - X_{it})\big]\Big| \\ &\quad + \frac{C}{T} \sup_{x\in[0,1]} \Big|\frac{1}{T}\sum_{t=1}^T \big\{\mathcal{W}_h(x - X_{it}) - \mathbb{E}\big[\mathcal{W}_h(x - X_{it})\big]\big\}\Big| \\ &= O\Big(\frac{1}{T}\Big) + O_p\Big(\frac{1}{T}\sqrt{\frac{\log T}{Th}}\Big)\end{aligned}$$

uniformly over i. As a result, we get that $\max_{1\le i\le n}\|A_T^{(i)}\| \le \max_{1\le i\le n}\overline{\lambda}^{(i)} = O_p(T^{-1})$, thus completing the proof of (S.17). To see (S.18), note that $\int W_h(X_{is}-x)W_h(X_{it}-x)\pi(x)/f_i^2(x)dx \le C/h$. Keeping this in mind, we obtain that

$$\begin{aligned}
\Lambda_T^{(i)} &= \frac{1}{T^4}\sum_{\substack{s,t=1\\ s\neq t}}^{T}\Big\{\int W_h(X_{is}-x)W_h(X_{it}-x)\frac{\pi(x)}{f_i^2(x)}dx\Big\}^2 \\
&\le \frac{C}{T^4 h}\sum_{s,t=1}^{T}\int W_h(X_{is}-x)W_h(X_{it}-x)\frac{\pi(x)}{f_i(x)}dx \\
&\le \frac{C}{T^2 h}\int\Big(\frac{1}{T}\sum_{s=1}^{T}W_h(X_{is}-x)\Big)\Big(\frac{1}{T}\sum_{t=1}^{T}W_h(X_{it}-x)\Big)\frac{\pi(x)}{f_i(x)}dx \\
&= O_p\Big(\frac{1}{T^2 h}\Big)
\end{aligned}$$

uniformly in i, taking into account that $T^{-1}\sum_{t=1}^T W_h(X_{it}-x) = O_p(1)$ uniformly over i and x.

We now let $\mathcal{X}_{n,T} = (X_{11},\dots,X_{1T},X_{21},\dots,X_{2T},\dots,X_{n1},\dots,X_{nT})$ be the vector of the regressors X_{it} and define the event

$$E_T = \Big\{\mathcal{X}_{n,T} : \max_{1\le i\le n}\|A_T^{(i)}\| \le \frac{\log T}{T} \text{ and } \max_{1\le i\le n}\Lambda_T^{(i)} \le \frac{\log T}{T^2 h}\Big\}.$$

By (S.17) and (S.18), it holds that $\mathbb{P}(E_T)\to 1$. Hence,

$$\begin{aligned}
\mathbb{P}\Big(\max_{1\le i\le n}|U_{i,T}| > C_U\frac{\log T}{Th^{1/2}}\Big) &= \mathbb{P}\Big(\max_{1\le i\le n}|U_{i,T}| > C_U\frac{\log T}{Th^{1/2}}, E_T\Big) + o(1) \\
&\le \sum_{i=1}^n \mathbb{P}\Big(|U_{i,T}| > C_U\frac{\log T}{Th^{1/2}}, E_T\Big) + o(1) \\
&= \sum_{i=1}^n \mathbb{P}\Big(1(E_T)\,|U_{i,T}| > C_U\frac{\log T}{Th^{1/2}}\Big) + o(1).
\end{aligned}$$

We further write

$$\mathbb{P}\Big(1(E_T)\,|U_{i,T}| > C_U\frac{\log T}{Th^{1/2}}\Big) = \mathbb{E}\Big[\mathbb{P}\Big(1(E_T)\,|U_{i,T}| > C_U\frac{\log T}{Th^{1/2}}\,\Big|\,\mathcal{X}_{n,T}\Big)\Big]$$

and derive an exponential bound on the conditional probability $\mathbb{P}(1(E_T)\,|U_{i,T}| > C_U\log T/(Th^{1/2})|\mathcal{X}_{n,T})$. To do so, we make use of the following result, which is immediately implied by the proof of the theorem in Wright (1973).

Theorem. *Define*

$$U_T = \sum_{s,t=-T}^{T} a_{st}\big(\eta_s\eta_t - \mathbb{E}[\eta_s\eta_t]\big)$$

and suppose that the following conditions are satisfied:

(i) $\{\eta_t : -T \le t \le T\}$ *is a sequence of independent random variables with zero means. For some constants* $M, \gamma > 0$, $\mathbb{P}(|\eta_t| \ge c) \le M \int_c^\infty \exp(-\gamma r^2) dr$ *for all* $-T \le t \le T$ *and all* $c \ge 0$.

(ii) For $-T \le s, t \le T$, a_{st} *are real numbers with* $a_{st} = a_{ts}$ *and* $\Lambda_T = \sum_{s,t=-T}^T a_{st}^2 \le C < \infty$. *Let* $A_T = (|a_{st}|)_{s,t=-T}^T$ *and denote the spectral norm of* A_T *by* $\|A_T\|$.

There exist constants C_a *and* C_b *depending only on* M *and* γ *such that for every* $\delta > 0$,

$$\mathbb{P}\big(U_T > \delta\big) \le \exp\Big(- \min\Big\{ \frac{C_a \delta}{\|A_T\|}, \frac{C_b \delta^2}{\Lambda_T} \Big\}\Big).$$

Setting $a_{st}^{(i)} = 0$ whenever $s < 1$ or $t < 1$, we can write $U_{i,T} = \sum_{s,t=-T}^T a_{st}^{(i)} \varepsilon_{is} \varepsilon_{it}$ and directly apply the above theorem. This yields

$$\begin{aligned}
&\mathbb{P}\Big(1(E_T)\, |U_{i,T}| > C_U \frac{\log T}{T h^{1/2}} \,\Big|\, \mathcal{X}_{n,T} \Big) \\
&\le \exp\Big(- \min\Big\{ \frac{C_a C_U \log T / (T h^{1/2})}{\log T / T}, \frac{C_b C_U^2 (\log T / (T h^{1/2}))^2}{\log T / (T^2 h)} \Big\}\Big) \\
&= \exp\big(- C_b C_U^2 \log T \big) = T^{-C_b C_U^2}
\end{aligned}$$

for sufficiently large sample sizes T. As a result,

$$\mathbb{P}\Big(\max_{1 \le i \le n} |U_{i,T}| > C_U \frac{\log T}{T h^{1/2}} \Big) \le n T^{-C_b C_U^2} + o(1) = o(1)$$

for C_U chosen sufficiently large. $\square$

Proof of (S.14). First of all, note that we can write

$$U_{ij,T} = \int \frac{Q_{i,V}(x) Q_{j,V}(x)}{f_i(x) f_j(x)} \pi(x) dx$$

with $Q_{i,V}(x) = T^{-1} \sum_{t=1}^T W_h(X_{it} - x) \varepsilon_{it}$. The arguments from the proof of Lemma A.1 show that

$$\mathbb{P}\Big(\max_{1 \le i \le n} \sup_{x \in [0,1]} \big| Q_{i,V}(x) \big| > C_Q \sqrt{\frac{\log T}{Th}} \Big) = o(1)$$

for C_Q chosen sufficiently large. Now let E_T be the event that $\max_i \sup_x |Q_{i,V}(x)| \le C_Q \sqrt{\log T / (Th)}$ and $E_{j,T}$ the event that $\sup_x |Q_{j,V}(x)| \le C_Q \sqrt{\log T / (Th)}$. Then

$$\begin{aligned}
\mathbb{P}\Big(\max_{1 \le i < j \le n} \big| U_{ij,T} \big| > C_U \frac{\log T}{T h^{1/2}} \Big) &= \mathbb{P}\Big(\max_{1 \le i < j \le n} \big| U_{ij,T} \big| > C_U \frac{\log T}{T h^{1/2}}, E_T \Big) + o(1) \\
&\le \sum_{1 \le i < j \le n} \mathbb{P}\Big(\big| U_{ij,T} \big| > C_U \frac{\log T}{T h^{1/2}}, E_T \Big) + o(1)
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{P}\Big(\big|U_{ij,T}\big| > C_U \frac{\log T}{Th^{1/2}}, E_T\Big) &= \mathbb{P}\Big(1(E_T)\big|U_{ij,T}\big| > C_U \frac{\log T}{Th^{1/2}}\Big) \\
&\le \mathbb{P}\Big(1(E_{j,T})\big|U_{ij,T}\big| > C_U \frac{\log T}{Th^{1/2}}\Big) \\
&= \mathbb{P}\Big(\Big|\frac{1}{T}\sum_{t=1}^{T} w_{ij,T}\varepsilon_{it}\Big| > C_U \frac{\log T}{Th^{1/2}}\Big),
\end{aligned}$$

where we set

$$w_{ij,T} = \int \frac{W_h(X_{it}-x)}{f_i(x)} \frac{Q_{j,V}(x)1(E_{j,T})}{f_j(x)} \pi(x)dx.$$

Noting that $w_{ij,T} \le C\sqrt{\log T/(Th)}$, one can show that

$$\mathbb{P}\Big(\Big|\frac{1}{T}\sum_{t=1}^{T} w_{ij,T}\varepsilon_{it}\Big| > C_U \frac{\log T}{Th^{1/2}}\Big) \le CT^{-r},$$

where $r > 0$ can be made arbitrarily large by choosing C_U large enough. This implies that

$$\mathbb{P}\Big(\max_{1\le i<j\le n} \big|U_{ij,T}\big| > C_U \frac{\log T}{Th^{1/2}}\Big) = o(1)$$

for C_U sufficiently large. □

Proof of Theorem 3.3

We focus attention on the proof of the distribution result (3.2). The convergence result (3.1) follows by slightly modifying the arguments of the proof. In a first step, we replace the estimator $\widehat{g}_k$ by the infeasible version

$$\widehat{g}_k^*(x) = \frac{1}{n_k}\sum_{i\in G_k} \widehat{m}_i(x)$$

and show that the difference between the two estimators is asymptotically negligible: For any null sequence $\{a_{n,T}\}$ of positive numbers, it holds that

$$\begin{aligned}
&\mathbb{P}\Big(\big|\widehat{g}_k(x) - \widehat{g}_k^*(x)\big| > a_{n,T}\Big) \\
&\quad \le \mathbb{P}\Big(\big|\widehat{g}_k(x) - \widehat{g}_k^*(x)\big| > a_{n,T}, \widehat{G}_k = G_k\Big) + \mathbb{P}\big(\widehat{G}_k \ne G_k\big) = o(1),
\end{aligned}$$

since the first probability on the right-hand side is equal to zero by definition of $\widehat{g}_k$ and $\widehat{g}_k^*$ and the second one is of the order $o(1)$ by Theorem 3.1. Hence, $|\widehat{g}_k(x) - \widehat{g}_k^*(x)| = O_p(a_{n,T})$ for any null sequence $\{a_{n,T}\}$ of positive numbers, which in turn implies that

$$\sqrt{\widehat{n}_k Th}\big(\widehat{g}_k(x) - g_k(x)\big) = \sqrt{\widehat{n}_k Th}\big(\widehat{g}_k^*(x) - g_k(x)\big) + o_p(1).$$

The difference between $\widehat{g}_k$ and $\widehat{g}_k^*$ can thus be asymptotically ignored.

To complete the proof of Theorem 3.3, we derive the limit distribution of the term $\sqrt{\widehat{n}_k Th}(\widehat{g}_k^*(x) - g_k(x))$: Since $\mathbb{P}(\widehat{n}_k \neq n_k) = o(1)$ by Theorem 3.1, it holds that $\sqrt{\widehat{n}_k Th}(\widehat{g}_k^*(x) - g_k(x)) = \sqrt{n_k Th}(\widehat{g}_k^*(x) - g_k(x)) + o_p(1)$. It thus suffices to compute the limit distribution of $\sqrt{n_k Th}(\widehat{g}_k^*(x) - g_k(x))$. To do so, write

$$\widehat{m}_i(x) - m_i(x) = \big[Q_{i,V}(x) + Q_{i,B}(x) - Q_{i,\gamma}(x)\big] / \widehat{f}_i(x) - \overline{Q}_i + \overline{\overline{Q}}_i,$$

where $Q_{i,V}(x)$, $Q_{i,B}(x)$, $Q_{i,\gamma}(x)$ along with $\overline{Q}_i$, $\overline{\overline{Q}}_i$ and $\widehat{f}_i(x)$ are defined in the proof of Lemma A.1. With this notation at hand, we obtain that

$$\begin{aligned}
&\sqrt{n_k Th}\,\big(\widehat{g}_k^*(x) - g_k(x)\big) \\
&= \sqrt{n_k Th}\Big\{\frac{1}{n_k}\sum_{i\in G_k}\frac{Q_{i,V}(x)}{\widehat{f}_i(x)} + \frac{1}{n_k}\sum_{i\in G_k}\frac{Q_{i,B}(x)}{\widehat{f}_i(x)} - \frac{1}{n_k}\sum_{i\in G_k}\frac{Q_{i,\gamma}(x)}{\widehat{f}_i(x)} - \frac{1}{n_k}\sum_{i\in G_k}\big(\overline{Q}_i - \overline{\overline{Q}}_i\big)\Big\} \\
&= \sqrt{n_k Th}\,\Big\{\frac{1}{n_k}\sum_{i\in G_k}\frac{Q_{i,V}(x)}{\widehat{f}_i(x)} + \frac{1}{n_k}\sum_{i\in G_k}\frac{Q_{i,B}(x)}{\widehat{f}_i(x)} - \frac{1}{n_k}\sum_{i\in G_k}\frac{Q_{i,\gamma}(x)}{\widehat{f}_i(x)}\Big\} + o_p(1),
\end{aligned}$$

the last line following by standard calculations. In the sequel, we show that

$$\frac{1}{n_k}\sum_{i\in G_k}\frac{Q_{i,\gamma}(x)}{\widehat{f}_i(x)} = o_p\Big(\frac{1}{\sqrt{n_k Th}}\Big) \tag{S.19}$$

$$\frac{1}{n_k}\sum_{i\in G_k}\frac{Q_{i,V}(x)}{\widehat{f}_i(x)} = \frac{1}{n_k}\sum_{i\in G_k}\frac{Q_{i,V}(x)}{f_i(x)} + o_p\Big(\frac{1}{\sqrt{n_k Th}}\Big) \tag{S.20}$$

$$\frac{1}{n_k}\sum_{i\in G_k}\frac{Q_{i,B}(x)}{\widehat{f}_i(x)} = \frac{1}{n_k}\sum_{i\in G_k}\frac{Q_{i,B}(x)}{f_i(x)} + o_p\Big(\frac{1}{\sqrt{n_k Th}}\Big). \tag{S.21}$$

(S.19)–(S.21) allow us to conclude that

$$\begin{aligned}
&\sqrt{n_k Th}\,\big(\widehat{g}_k^*(x) - g_k(x)\big) \\
&= \sqrt{n_k Th}\Big\{\frac{1}{n_k}\sum_{i\in G_k}\frac{Q_{i,V}(x)}{f_i(x)} + \frac{1}{n_k}\sum_{i\in G_k}\frac{Q_{i,B}(x)}{f_i(x)}\Big\} + o_p(1) \\
&= \sqrt{n_k Th}\Big(\frac{1}{n_k T}\sum_{i\in G_k}\sum_{t=1}^{T}\frac{W_h(X_{it}-x)}{f_i(x)}\varepsilon_{it}\Big) \\
&\quad + \sqrt{n_k Th}\Big(\frac{1}{n_k T}\sum_{i\in G_k}\sum_{t=1}^{T}\frac{W_h(X_{it}-x)}{f_i(x)}\big[m_i(X_{it}) - m_i(x)\big]\Big) + o_p(1).
\end{aligned}$$

With the help of a standard central limit theorem, the first term on the right-hand side can be shown to weakly converge to a normal distribution with mean zero and variance $V_k(x)$. Moreover, standard bias calculations yield that the second term converges in probability to the bias expression $B_k(x)$. This completes the proof. □

Proof of (S.19). In a first step, we show that

$$R_\gamma := \frac{1}{n_k} \sum_{i \in G_k} \frac{Q_{i,\gamma}(x)}{\widehat{f}_i(x)} - \frac{1}{n_k} \sum_{i \in G_k} \frac{Q_{i,\gamma}(x)}{\mathbb{E}[\widehat{f}_i(x)]} = o_p\Big(\frac{1}{\sqrt{n_k T h}}\Big). \tag{S.22}$$

To do so, we write $R_\gamma = R_{\gamma,1} + R_{\gamma,2}$, where

$$\begin{aligned} R_{\gamma,1} &= \frac{1}{n_k} \sum_{i \in G_k} \frac{\mathbb{E}[\widehat{f}_i(x)] - \widehat{f}_i(x)}{\mathbb{E}[\widehat{f}_i(x)]^2} Q_{i,\gamma}(x) \\ R_{\gamma,2} &= \frac{1}{n_k} \sum_{i \in G_k} \frac{(\mathbb{E}[\widehat{f}_i(x)] - \widehat{f}_i(x))^2}{\mathbb{E}[\widehat{f}_i(x)]^2 \widehat{f}_i(x)} Q_{i,\gamma}(x). \end{aligned}$$

Defining $Z_{it}(x) = \mathbb{E}[W_h(X_{it} - x)] - W_h(X_{it} - x)$, the first term $R_{\gamma,1}$ can be expressed as

$$\begin{aligned} R_{\gamma,1} = \frac{1}{n_k} \sum_{i \in G_k} & \frac{1}{\mathbb{E}[\widehat{f}_i(x)]^2} \Big\{ \frac{1}{T} \sum_{t=1}^T Z_{it}(x) \Big\} \\ & \times \Big\{ \frac{1}{T} \sum_{t=1}^T W_h(X_{it} - x) \Big(\frac{1}{n-1} \sum_{\substack{j=1 \\ j \neq i}}^n \big[m_j(X_{jt}) + \varepsilon_{jt} \big] \Big) \Big\}. \end{aligned}$$

We thus obtain that

$$\begin{aligned} \mathbb{E}[R_{\gamma,1}^2] = \frac{1}{n_k^2 (n-1)^2} & \sum_{i,i' \in G_k} \sum_{\substack{j \neq i \\ j' \neq i'}} \frac{1}{\mathbb{E}[\widehat{f}_i(x)]^2} \frac{1}{\mathbb{E}[\widehat{f}_{i'}(x)]^2} \\ & \times \Big(\frac{1}{T^4} \sum_{t,t',s,s'=1}^T \Psi_{i,i',j,j',t,t',s,s'}(x) \Big), \end{aligned} \tag{S.23}$$

where we use the shorthand

$$\begin{aligned} \Psi_{i,i',j,j',t,t',s,s'}(x) = \mathbb{E}\big[& Z_{it}(x) W_h(X_{is} - x) \big\{ m_j(X_{js}) + \varepsilon_{js} \big\} \\ & \times Z_{i't'}(x) W_h(X_{i's'} - x) \big\{ m_{j'}(X_{j's'}) + \varepsilon_{j's'} \big\} \big]. \end{aligned}$$

Importantly, the expressions $\Psi_{i,i',j,j',t,t',s,s'}(x)$ in (S.23) have the following property: $\Psi_{i,i',j,j',t,t',s,s'}(x) \neq 0$ only if (a) $i = j'$ and $i' = j$ or (b) $j = j'$. Exploiting the mixing conditions of (C1) by means of Davydov's inequality (see Corollary 1.1 in Bosq (1998)), we can show that in case (a), $|T^{-4} \sum_{t,t',s,s'=1}^T \psi_{i,i',j,j',t,t',s,s'}(x)| \leq C (\log T)^2 / (Th)^2$ and in case (b),

$$\Big| \frac{1}{T^4} \sum_{t,t',s,s'=1}^T \psi_{i,i',j,j',t,t',s,s'}(x) \Big| \leq \begin{cases} C (\log T)^3 / (T^3 h^2) & \text{for } i \neq i' \\ C (\log T)^2 / (T^2 h^3) & \text{for } i = i'. \end{cases}$$

Plugging these bounds into (S.23), we immediately arrive at $R_{\gamma,1} = o_p(1/\sqrt{n_k T h})$. Furthermore, with the help of Hölder's inequality and (S.5), we obtain that

$$
\begin{aligned}
R_{\gamma,2} &\leq \Big\{ \max_{1\leq i\leq n} \sup_{x\in[0,1]} \frac{(\mathbb{E}[\widehat{f}_i(x)] - \widehat{f}_i(x))^2}{\mathbb{E}[\widehat{f}_i(x)]^2 \widehat{f}_i(x)} \Big\} \Big\{ \frac{1}{n_k} \sum_{i\in G_k} \Big(\frac{1}{T} \sum_{t=1}^{T} W_h^{4/3}(X_{it} - x) \Big)^{3/4} \\
&\qquad \times \Big(\frac{1}{T} \sum_{t=1}^{T} \Big(\frac{1}{n-1} \sum_{\substack{j=1 \\ j\neq i}}^{n} \big[m_j(X_{jt}) + \varepsilon_{jt} \big] \Big)^4 \Big)^{1/4} \Big\} \\
&= O_p\Big(\Big(\sqrt{\frac{\log T}{Th}} \Big)^2 \frac{1}{h^{1/4}(n-1)^{1/2}} \Big) = o_p\Big(\frac{1}{\sqrt{n_k T h}} \Big),
\end{aligned}
$$

which completes the proof of (S.22).

In the next step, we show that

$$
\frac{1}{n_k} \sum_{i\in G_k} \frac{Q_{i,\gamma}(x)}{\mathbb{E}[\widehat{f}_i(x)]} = o_p\Big(\frac{1}{\sqrt{n_k T h}} \Big). \tag{S.24}
$$

To do so, we derive the convergence rate of the second moment

$$
\begin{aligned}
\mathbb{E}\left[\Big\{ \frac{1}{n_k} \sum_{i\in G_k} \frac{Q_{i,\gamma}(x)}{\mathbb{E}[\widehat{f}_i(x)]} \Big\}^2 \right] &= \frac{1}{n_k^2 (n-1)^2} \sum_{i,i'\in G_k} \sum_{\substack{j\neq i \\ j'\neq i'}} \frac{1}{\mathbb{E}[\widehat{f}_i(x)]} \frac{1}{\mathbb{E}[\widehat{f}_{i'}(x)]} \\
&\qquad \times \Big(\frac{1}{T^2} \sum_{t,t'=1}^{T} \Psi_{i,i',j,j',t,t'}(x) \Big), \qquad \text{(S.25)}
\end{aligned}
$$

where $\Psi_{i,i',j,j',t,t'}(x) = \mathbb{E}[W_h(X_{it} - x)\{m_j(X_{jt}) + \varepsilon_{jt}\} W_h(X_{i't'} - x)\{m_{j'}(X_{j't'}) + \varepsilon_{j't'}\}]$. Similarly as above, $\Psi_{i,i',j,j',t,t'}(x) \neq 0$ only if (a) $i = j'$ and $i' = j$ or (b) $j = j'$. Applying Davydov's inequality once again, we get that in case (a), $|T^{-2} \sum_{t,t'=1}^{T} \Psi_{i,i',j,j',t,t'}(x)| \leq C \log T / T$ and in case (b),

$$
\Big| \frac{1}{T^2} \sum_{t,t'=1}^{T} \Psi_{i,i',j,j',t,t'}(x) \Big| \leq \begin{cases} C/T & \text{for } i \neq i' \\ C/(Th) & \text{for } i = i'. \end{cases}
$$

Plugging these bounds into (S.25), we easily arrive at (S.24). The statement (S.19) now follows upon combining (S.22) with (S.24). □

Proof of (S.20) and (S.21). By arguments similar to those for (S.19),

$$
\frac{1}{n_k} \sum_{i\in G_k} \frac{Q_{i,\ell}(x)}{\widehat{f}_i(x)} - \frac{1}{n_k} \sum_{i\in G_k} \frac{Q_{i,\ell}(x)}{\mathbb{E}[\widehat{f}_i(x)]} = o_p\Big(\frac{1}{\sqrt{n_k T h}} \Big) \tag{S.26}
$$

for $\ell \in \{V, B\}$. With the help of standard bias calculations, we further obtain that

$$\frac{1}{n_k} \sum_{i \in G_k} \frac{Q_{i,\ell}(x)}{\mathbb{E}[\widehat{f}_i(x)]} - \frac{1}{n_k} \sum_{i \in G_k} \frac{Q_{i,\ell}(x)}{f_i(x)} = o_p\Big(\frac{1}{\sqrt{n_k T h}}\Big). \tag{S.27}$$

Combining (S.26) and (S.27) completes the proof. □

References

Bosq, D. (1998). *Nonparametric statistics for stochastic processes.* New York, Springer.

Bradley, R. C. (2005). Basic properties of strong mixing conditions. a survey and some open questions. *Probability Surveys*, **2** 107–144.

Hansen, B. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory*, **24** 726–748.

Härdle, W., Hall, P. and Marron, J. S. (1988). How far are automatically chosen regression smoothing parameters from their optimum? *Journal of the American Statistical Association*, **83** 86–95.

Masry, E. (1996). Multivariate local polynomial regression for time series: uniform strong consistency and rates. *Journal of Time Series Analysis*, **17** 571–599.

Wright, F. T. (1973). A bound on tail probabilities for quadratic forms in independent random variables whose distributions are not necessarily symmetric. *Annals of Probability*, **1** 1068–1070.