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

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

where the model variables $X_{it}$ and $\varepsilon_{it}$ 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 $\text{RSS}_i^{(j)}(h)$ for different pairs of indices $(i, j)$. More precisely, we define our bandwidth selector by

$$\hat{h} = \frac{1}{L} \sum_{1 \leq \ell \leq L} \hat{h}_{i2\ell-1}^{(i2\ell)},$$

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

1
To examine the finite sample behaviour of the bandwidth estimator \(\hat{h}\), 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 \(\hat{h}\) for each simulated sample. To do so, we define an equidistant grid \(G\) of step length 0.01 which spans the interval \([0.025, 0.5]\) and minimize the criterion functions \(\text{RSS}_j^{(2)}(h)\) over all bandwidth values \(h \in G\). 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 \(h^* - \hat{h}\) 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 \(h^* - \hat{h}\) takes a certain value. The plots of Figure 1 suggest that \(\hat{h}\) approximates the optimal value \(h^*\) reasonably well. They also make visible that the precision of the estimator \(\hat{h}\) 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 \(h^* - \hat{h}\) 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.](image-url)
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 $x |\hat{m}_i(x) - m_i(x)|$ 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

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

where

$$Q_{i,V}(x) = \frac{1}{T} \sum_{t=1}^{T} W_h(X_{it} - x) \epsilon_{it}$$

$$Q_{i,B}(x) = \frac{1}{T} \sum_{t=1}^{T} W_h(X_{it} - x) [m_i(X_{it}) - m_i(x)]$$

$$Q_{i,\gamma}(x) = \frac{1}{T} \sum_{t=1}^{T} W_h(X_{it} - x) \left( \frac{1}{n-1} \sum_{j=1 \atop j \neq i}^{n} [m_j(X_{jt}) + \epsilon_{jt}] \right)$$

$$\mathcal{Q}_i = \frac{1}{T} \sum_{t=1}^{T} [m_i(X_{it}) + \epsilon_{it}]$$

$$\overline{Q}_i = \frac{1}{(n-1)T} \sum_{j=1}^{n} \sum_{t=1 \atop j \neq i}^{T} [m_j(X_{jt}) + \epsilon_{jt}]$$

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

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

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

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

$$\max_{1 \leq i \leq n} \sup_{x \in [0,1]} |\hat{f}_i(x) - \mathbb{E}[\hat{f}_i(x)]| = O_p\left( \sqrt{\frac{\log T}{Th}} \right). \quad (S.5)$$
Moreover, standard bias calculations yield that \( \max_{1 \leq i \leq n} \sup_{x \in I_k} | \mathbb{E}[Q_{i,B}(x)] | = O(h^2) \) along with \( \max_{1 \leq i \leq n} \sup_{x \in [0,1] \setminus I_k} | \mathbb{E}[Q_{i,B}(x)] | = O(h) \). Analogously, \( \max_{1 \leq i \leq n} \sup_{x \in I_k} | \mathbb{E}[\hat{f}_i(x)] | = O(h^2) \) and \( \max_{1 \leq i \leq n} \sup_{x \in [0,1] \setminus I_k} | \mathbb{E}[\hat{f}_i(x)] | = O(h) \). Finally, a simplified version of the arguments for (S.2) shows that \( \max_{1 \leq i \leq n} |Q_i| = O_p(\alpha_{n,T}) \) as well as \( \max_{1 \leq i \leq n} |\overline{Q}_i| = O_p(\alpha_{n,T}) \). Lemma A.1 immediately follows upon combining (S.2)–(S.5) with these statements.

\[ \Box \]

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

\[
\begin{align*}
\varepsilon_{it}^\leq &= \varepsilon_{it} 1(\{\varepsilon_{it} \leq \psi_{n,T}\}) \\
\varepsilon_{it}^\geq &= \varepsilon_{it} 1(\{\varepsilon_{it} > \psi_{n,T}\}).
\end{align*}
\]

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

\[
Q_{i,V}(x) = \sum_{t=1}^{T} Z_{it,T}^\leq(x) + \sum_{t=1}^{T} Z_{it,T}^\geq(x),
\]

where

\[
egin{align*}
Z_{it,T}^\leq(x) &= (W_h(X_{it} - x)\varepsilon_{it}^\leq - \mathbb{E}[W_h(X_{it} - x)\varepsilon_{it}^\leq]) / T \\
Z_{it,T}^\geq(x) &= (W_h(X_{it} - x)\varepsilon_{it}^\geq - \mathbb{E}[W_h(X_{it} - x)\varepsilon_{it}^\geq]) / T.
\end{align*}
\]

We thus split \( Q_{i,V}(x) \) into the “interior part” \( \sum_{t=1}^{T} Z_{it,T}^\leq(x) \) and the “tail part” \( \sum_{t=1}^{T} Z_{it,T}^\geq(x) \). 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

\[
\max_{1 \leq i \leq n} \sup_{x \in [0,1]} \left| \sum_{t=1}^{T} Z_{it,T}^\geq(x) \right| = O_p(\alpha_{n,T}). \tag{S.6}
\]

This can be achieved as follows:

\[
\begin{align*}
\mathbb{P} \left( \max_{1 \leq i \leq n} \sup_{x \in [0,1]} \left| \sum_{t=1}^{T} Z_{it,T}^\geq(x) \right| > a_{n,T} \right) \\
&\leq \sum_{i=1}^{n} \mathbb{P} \left( \sup_{x \in [0,1]} \left| \frac{1}{T} \sum_{t=1}^{T} W_h(X_{it} - x)\varepsilon_{it}^\geq \right| > \frac{a_{n,T}}{2} \right) \\
&+ \sum_{i=1}^{n} \mathbb{P} \left( \sup_{x \in [0,1]} \left| \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}[W_h(X_{it} - x)\varepsilon_{it}^\geq] \right| > \frac{a_{n,T}}{2} \right).
\end{align*}
\]
With the help of assumption (C3), we obtain that

\[
\sum_{i=1}^{n} P\left( \sup_{x \in [0,1]} \left| \frac{1}{T} \sum_{t=1}^{T} W_{h}(X_{ut} - x)\varepsilon_{it}^{>\gamma} \right| > \frac{a_{n,T}}{2} \right) \\
\leq \sum_{i=1}^{n} P\left( |\varepsilon_{it}| > \psi_{n,T} \text{ for some } 1 \leq t \leq T \right) \leq C(nT)^{1 - \frac{\theta}{2T}} = o(1).
\]

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

\[
\left| E\left[ W_{h}(X_{it} - x)\varepsilon_{it}^{>\gamma} \right] \right| \leq E\left[ W_{h}(X_{it} - x)E\left[ \frac{|\varepsilon_{it}|^{\theta}}{\psi_{n,T}^{\theta}} 1(|\varepsilon_{it}| > \psi_{n,T}) \right] |X_{it}\right] \\
\leq C(nT)^{-\frac{\theta+1}{2T-\theta}}
\]

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

\[
\sum_{i=1}^{n} P\left( \sup_{x \in [0,1]} \left| \frac{1}{T} \sum_{t=1}^{T} E\left[ W_{h}(X_{it} - x)\varepsilon_{it}^{>\gamma} \right] \right| > \frac{a_{n,T}}{2} \right) = 0
\]

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

We next have a closer look at the expression \( \sum_{t=1}^{T} Z_{it,T}^{\leq}(x) \). 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

\[
\max_{1 \leq i \leq n} \sup_{x \in [0,1]} \left| \sum_{t=1}^{T} Z_{it,T}^{\leq}(x) \right| \leq \max_{1 \leq i \leq n} \max_{1 \leq i \leq L} \left| \sum_{t=1}^{T} Z_{it,T}^{\leq}(x_{t}) \right| + C a_{n,T}.
\]

(S.7)

We can thus replace the supremum over \( x \) by a maximum over the grid points \( x_{t} \). Moreover, it holds that

\[
P\left( \max_{1 \leq i \leq n} \max_{1 \leq i \leq L} \left| \sum_{t=1}^{T} Z_{it,T}^{\leq}(x_{t}) \right| > C_0 a_{n,T} \right) \\
\leq \sum_{i=1}^{n} \sum_{t=1}^{L} P\left( \left| \sum_{t=1}^{T} Z_{it,T}^{\leq}(x_{t}) \right| > C_0 a_{n,T} \right),
\]

(S.8)

where \( C_0 \) is a sufficiently large constant to be specified later on. In what follows, we show that for each fixed \( x_{t} \),

\[
P\left( \sum_{t=1}^{T} Z_{it,T}^{\leq}(x_{t}) \right| > C_0 a_{n,T} \right) \leq C T^{-r},
\]

(S.9)
where the constants $C$ and $r$ are independent of $x_\ell$ 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

$$
\max_{1 \leq \ell \leq n} \sup_{x \in [0,1]} \left| \sum_{t=1}^{T} Z_{\ell,t,T}^c(x) \right| = O_p(a_{n,T}),
$$

(S.10)

which completes the proof.

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

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

with $B_s = \sum_{t=(s-1)r_{n,T}+1}^{sr_{n,T}} Z_{\ell,t,T}^c(x_\ell)$, 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

$$
P \left( \left| \sum_{t=1}^{T} Z_{\ell,t,T}^c(x_\ell) \right| > C_0 a_{n,T} \right) \leq P \left( \left| \sum_{s=1}^{q_{n,T}} B_{2s-1} \right| > \frac{C_0}{2} a_{n,T} \right) + P \left( \left| \sum_{s=1}^{q_{n,T}} B_{2s} \right| > \frac{C_0}{2} a_{n,T} \right).
$$

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 $0 < \mu \leq \|B_{2s-1}\|_\infty$,

$$
P(|B_{2s-1} - B_{2s-1}^*| > \mu) \leq 18(\|B_{2s-1}\|_\infty/\mu)^{1/2} \alpha(r_{n,T}).
$$

With these variables, we obtain the bound

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

where

$$
P_1 = P \left( \left| \sum_{s=1}^{q_{n,T}} B_{2s-1}^* \right| > \frac{C_0}{4} a_{n,T} \right)
$$

and

$$
P_2 = P \left( \left| \sum_{s=1}^{q_{n,T}} (B_{2s-1} - B_{2s-1}^*) \right| > \frac{C_0}{4} a_{n,T} \right).
$$

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

(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
\[
\mathbb{E}\left[\exp\left(\pm \lambda_{n, T} B_{2s-1}\right)\right] \leq 1 + \lambda_{n, T}^2 \mathbb{E}\left[(B_{2s-1})^2\right] \leq \exp(\lambda_{n, T}^2 \mathbb{E}\left[(B_{2s-1})^2\right])
\]
along with
\[
\mathbb{E}\left[\exp\left(\pm \lambda_{n, T} B_{2s-1}^*\right)\right] \leq \exp(\lambda_{n, T}^2 \mathbb{E}\left[(B_{2s-1}^*)^2\right]).
\]

(iii) Standard calculations for kernel estimators yield that
\[
\sum_{s=1}^{q_{n, T}} \mathbb{E}[(B_{2s-1}^*)^2] \leq \frac{C}{Th}.
\]

Using (i)–(iii), we arrive at
\[
P\left(|\sum_{s=1}^{q_{n, T}} B_{2s-1}^*| > C_0 \frac{a_{n, T}}{4}\right)
\]
\[
\leq \mathbb{P}\left(\sum_{s=1}^{q_{n, T}} B_{2s-1}^* > C_0 \frac{a_{n, T}}{4}\right) + \mathbb{P}\left(- \sum_{s=1}^{q_{n, T}} B_{2s-1}^* > C_0 \frac{a_{n, T}}{4}\right)
\]
\[
\leq \exp\left(- \frac{C_0}{4} \lambda_{n, T} a_{n, T}\right) \left\{ \mathbb{E}\left[\exp\left(\lambda_{n, T} \sum_{s=1}^{q_{n, T}} B_{2s-1}^*\right)\right] + \mathbb{E}\left[\exp\left(- \lambda_{n, T} \sum_{s=1}^{q_{n, T}} B_{2s-1}^*\right)\right] \right\}
\]
\[
\leq \exp\left(- \frac{C_0}{4} \lambda_{n, T} a_{n, T}\right) \left\{ \prod_{s=1}^{q_{n, T}} \mathbb{E}\left[\exp\left(\lambda_{n, T} B_{2s-1}^*\right)\right] + \prod_{s=1}^{q_{n, T}} \mathbb{E}\left[\exp\left(- \lambda_{n, T} B_{2s-1}^*\right)\right] \right\}
\]
\[
\leq 2 \exp\left(- \frac{C_0}{4} \lambda_{n, T} a_{n, T}\right) \prod_{s=1}^{q_{n, T}} \mathbb{E}\left[\lambda_{n, T}^2 \mathbb{E}\left[(B_{2s-1}^*)^2\right]\right]
\]
\[
= 2 \exp\left(- \frac{C_0}{4} \lambda_{n, T} a_{n, T}\right) \exp\left(\lambda_{n, T}^2 \sum_{s=1}^{q_{n, T}} \mathbb{E}\left[(B_{2s-1})^2\right]\right)
\]
\[
\leq 2 \exp\left(- \frac{C_0}{4} \lambda_{n, T} a_{n, T} + \lambda_{n, T}^2 \frac{C}{Th}\right).
\]

Recalling that $n/T \leq C$ and $T^{2/5} h \to \infty$ by assumption, setting $\theta$ to a value slightly larger than 4 and supposing that $a_{n, T} = T^{-1/10}$, it follows that
\[
\exp\left(- \frac{C_0}{4} \lambda_{n, T} a_{n, T} + \lambda_{n, T}^2 \frac{C}{Th}\right) \leq T^{-r}, \quad \text{(S.11)}
\]
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 \( r_{n,T} = \sqrt{(T/h)/(\psi_{n,T}^2 \log T)} \), (S.11) even holds for \( a_{n,T} = \sqrt{\log T/(Th)} \). From (S.11), it immediately follows that \( P_1 \leq CT^{-r} \), 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 \( Z_{it} = (n-1)^{-1} \sum_{j=1, j \neq i}^n (m_j(X_{jt}) + \varepsilon_{jt}) \) and write

\[
Q_{i,\gamma}(x) = \frac{1}{T} \sum_{t=1}^T W_h(X_{it} - x) Z_{it}.
\]  

(S.12)

By construction, the time series processes \( \{X_{it} : 1 \leq t \leq T\} \) and \( \{Z_{it} : 1 \leq t \leq T\} \) are independent of each other. Moreover, by Theorem 5.2 in Bradley (2005), the process \( \{Z_{it} : 1 \leq t \leq T\} \) 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 \( \varepsilon_{it} \) have unbounded support in general, we have introduced the truncation sequence \( \psi_{n,T} \) and have split \( \varepsilon_{it} \) into the two parts \( \varepsilon_{it}^\leq \) and \( \varepsilon_{it}^\geq \). 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 \( i, j \in G_k \),

\[
\hat{\Delta}_{ij} = \int \left( \frac{Q_{i,V}(x) + Q_{i,B}(x)}{\hat{f}_i(x)} - \frac{Q_{j,V}(x) + Q_{j,B}(x)}{\hat{f}_j(x)} \right)^2 \pi(x) dx
\]

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

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

\[
\max_{i \in G_k} \sup_{x \in [0,1]} |Q_{i,B}(x) - \mathbb{E}[Q_{i,B}(x)]| = O_p\left(h \frac{\log T}{Th}\right)
\]
\[ \max_{i \in G_k} \sup_{x \in I_h} \left| \frac{\hat{f}_i(x)}{f_i(x)} - 1 \right| = O_p \left( \sqrt{\frac{\log T}{Th}} + h^2 \right) \]

and

\[ \mathbb{E}[Q_{i,B}(x)] = h^2 \left( \int W(x) \varphi^2 dx \right) \left( m'_i(x) f'_i(x) + \frac{m''_i(x) f_i(x)}{2} \right) + O(h^3) \]

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

\[ \max_{i,j \in G_k} \hat{\Delta}_{ij} = \max_{i,j \in G_k} \int \left( \frac{Q_{i,V}(x)}{f_i(x)} - \frac{Q_{j,V}(x)}{f_j(x)} \right)^2 \pi(x) dx + o_p \left( \frac{1}{T h^{1/2}} \right). \]

Next define

\[ U_{i,T} = \sum_{s,t=1}^T a^{(i)}_{st} \varepsilon_{is} \varepsilon_{it}, \]

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

\[ U_{ij,T} = \sum_{s,t=1}^T a^{(ij)}_{st} \varepsilon_{is} \varepsilon_{jt}, \]

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

\[ B_{i,T} = \int \frac{1}{T^2} \sum_{t=1}^T W_h^2(X_{it} - x) \varepsilon_{it}^2 f^2_i(x) dx. \]

With these definitions at hand, we can write

\[ \max_{i,j \in G_k} \int \left( \frac{Q_{i,V}(x)}{f_i(x)} - \frac{Q_{j,V}(x)}{f_j(x)} \right)^2 \pi(x) dx = \max_{i,j \in G_k} \left\{ U_{i,T} - 2U_{ij,T} + U_{j,T} + B_{i,T} + B_{j,T} \right\}. \]

Below we show that

\[ \max_{1 \leq i \leq n} \left| U_{i,T} \right| = O_p \left( \frac{\log T}{TH^{1/2}} \right) \quad \text{(S.13)} \]

\[ \max_{1 \leq i < j \leq n} \left| U_{ij,T} \right| = O_p \left( \frac{\log T}{TH^{1/2}} \right). \quad \text{(S.14)} \]

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

\[ \max_{1 \leq i \leq n} \left| B_{i,T} - \mathbb{E}[B_{i,T}] \right| = o_p \left( \frac{1}{T h^{1/2}} \right) \quad \text{(S.15)} \]

\[ \mathbb{E}[B_{i,T}] = \frac{1}{Th} \left( \int W^2(\varphi) d\varphi \right) \int \frac{\sigma_i(x) \pi(x)}{f_i(x)} dx + O\left( \frac{1}{T} \right) \quad \text{(S.16)} \]
uniformly in $i$. Combining (S.13)–(S.16) and noting that $\mathbb{E}[B_{i,T}] + \mathbb{E}[B_{j,T}] = B_{ij}/(Th) + O(T^{-1})$ uniformly in $i$ and $j$, we arrive at

$$\max_{i,j \in G_k} \hat{\Delta}_{ij} = \max_{i,j \in G_k} \frac{B_{ij}}{Th} + O_p\left(\frac{1}{Th^{1/2}}\right).$$

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

$$\max_{1 \leq i \leq n} \|A_T^{(i)}\| = O_p\left(\frac{1}{T}\right) \quad (S.17)$$

$$\max_{1 \leq i \leq n} \Lambda_T^{(i)} = O_p\left(\frac{1}{T^2 h}\right), \quad (S.18)$$

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

$$\lambda_T^{(i)} = \max_{1 \leq s \leq T} \sum_{t=1}^T |a_{st}^{(i)}|.$$  

Standard calculations yield that

$$\lambda_T^{(i)} \leq C \frac{1}{T} \max_{1 \leq s \leq T} \sum_{t=1}^T W_h(X_{is} - X_{it}),$$

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

$$\lambda_T^{(i)} \leq C \frac{1}{T} \sup_{x \in [0,1]} \frac{1}{T} \sum_{t=1}^T W_h(x - X_{it})$$

$$\leq C \frac{1}{T} \sup_{x \in [0,1]} \left| \frac{1}{T} \sum_{t=1}^T \mathbb{E}[W_h(x - X_{it})]\right|$$

$$+ C \frac{1}{T} \sup_{x \in [0,1]} \left| \frac{1}{T} \sum_{t=1}^T \{W_h(x - X_{it}) - \mathbb{E}[W_h(x - X_{it})]\}\right|$$

$$= O\left(\frac{1}{T}\right) + O_p\left(\frac{1}{T \sqrt{Th}}\right).$$
uniformly over $i$. As a result, we get that $\max_{1 \leq i \leq n} \|A_T^{(i)}\| \leq \max_{1 \leq i \leq n} \bar{X}^{(i)} = O_p(T^{-1})$, thus completing the proof of (S.17). To see (S.18), note that $\int W_h(X_is - x)W_h(X_it - x)\pi(x)/f_i^2(x)dx \leq C/h$. Keeping this in mind, we obtain that

$$\Lambda_T^{(i)} = \frac{1}{T^4} \sum_{s,t=1}^{T} \left\{ \int W_h(X_is - x)W_h(X_it - x)\frac{\pi(x)}{f_i^2(x)}dx \right\}^2 \leq \frac{C}{T^4h} \sum_{s,t=1}^{T} \int W_h(X_is - x)W_h(X_it - x)\frac{\pi(x)}{f_i(x)}dx \leq \frac{C}{T^2h} \int \left( \frac{1}{T} \sum_{s=1}^{T} W_h(X_is - x) \right) \left( \frac{1}{T} \sum_{t=1}^{T} W_h(X_it - x) \right)\frac{\pi(x)}{f_i(x)}dx = O_p\left(\frac{1}{T^2h}\right)$$

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 $X_{n,T} = (X_{11}, \ldots, X_{1T}, X_{21}, \ldots, X_{2T}, \ldots, X_{n1}, \ldots, X_{nT})$ be the vector of the regressors $X_{it}$ and define the event

$$E_T = \left\{ X_{n,T} : \max_{1 \leq i \leq n} \|A_T^{(i)}\| \leq \frac{\log T}{T} \text{ and } \max_{1 \leq i \leq n} \Lambda_T^{(i)} \leq \frac{\log T}{T^2h} \right\}.$$

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

$$\mathbb{P}\left(\max_{1 \leq i \leq n} |U_{i,T}| > C_U \frac{\log T}{Th^{1/2}}\right) = \mathbb{P}\left(\max_{1 \leq i \leq n} |U_{i,T}| > C_U \frac{\log T}{Th^{1/2}}, E_T\right) + o(1) \leq \sum_{i=1}^{n} \mathbb{P}\left(|U_{i,T}| > C_U \frac{\log T}{Th^{1/2}}, E_T\right) + o(1) = \sum_{i=1}^{n} \mathbb{P}\left(1(E_T) |U_{i,T}| > C_U \frac{\log T}{Th^{1/2}}\right) + o(1).$$

We further write

$$\mathbb{P}\left(1(E_T) |U_{i,T}| > C_U \frac{\log T}{Th^{1/2}}\right) = \mathbb{E}\left[ \mathbb{P}\left(1(E_T) |U_{i,T}| > C_U \frac{\log T}{Th^{1/2}} \mid X_{n,T}\right) \right]$$

and derive an exponential bound on the conditional probability $\mathbb{P}(1(E_T) |U_{i,T}| > C_U \log T/(Th^{1/2}) \mid X_{n,T})$. 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

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

and suppose that the following conditions are satisfied:
(i) \( \{\eta_t : -T \leq t \leq T\} \) is a sequence of independent random variables with zero means. For some constants \( M, \gamma > 0 \), \( \mathbb{P}(|\eta_t| \geq c) \leq M \int_c^\infty \exp(-\gamma r^2)dr \) for all \(-T \leq t \leq T\) and all \( c \geq 0 \).

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

There exist constants \( C_a \) and \( C_b \) depending only on \( M \) and \( \gamma \) such that for every \( \delta > 0 \),

\[
\mathbb{P}(U_T > \delta) \leq \exp\left(-\min\left\{ \frac{C_a\delta}{\|A_T\|}, \frac{C_b\delta^2}{\Lambda_T}\right\}\right).
\]

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

\[
\mathbb{P}\left(1(E_T) |U_{i,T}| > C_U \frac{\log T}{Th^{1/2}} \right| X_n) \leq \mathbb{P}\left(- \min\left\{ \frac{C_a C_U \log T/(Th^{1/2})}{\log T/T}, \frac{C_b C_U^2 (\log T/(Th^{1/2}))^2}{\log T/(T^2 h)} \right\}\right) = \mathbb{P}\left(- \frac{C_b C_U^2 \log T}{T^{-C_a C_U^2}} \right)
\]

for sufficiently large sample sizes \( T \). As a result,

\[
\mathbb{P}\left(\max_{1 \leq i \leq n} |U_{i,T}| > C_U \frac{\log T}{Th^{1/2}} \right) \leq nT^{-C_a C_U^2} + o(1) = o(1)
\]

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

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

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

\[
\mathbb{P}\left(\max_{1 \leq i \leq j \leq n} |U_{ij,T}| > C_U \frac{\log T}{Th^{1/2}} \right) = \mathbb{P}\left(\max_{1 \leq i \leq j \leq n} |U_{ij,T}| > C_U \frac{\log T}{Th^{1/2}}, E_T \right) + o(1)
\]

\[
\leq \sum_{1 \leq i < j \leq n} \mathbb{P}\left(|U_{ij,T}| > C_U \frac{\log T}{Th^{1/2}}, E_T \right) + o(1)
\]
and
\[ \mathbb{P}
\left( |U_{ij,T}| > C_U \frac{\log T}{T^{h/2}}, E_T \right)
= \mathbb{P}
\left( 1(E_T) \left| U_{ij,T} \right| > C_U \frac{\log T}{T^{h/2}} \right)
\leq \mathbb{P}
\left( 1(E_{j,T}) \left| U_{ij,T} \right| > C_U \frac{\log T}{T^{h/2}} \right)
= \mathbb{P}
\left( \left| \frac{1}{T} \sum_{t=1}^{T} w_{ij,T} \varepsilon_{it} \right| > C_U \frac{\log T}{T^{h/2}} \right),
\]
where we set
\[ w_{ij,T} = \int W_h(X_{it} - x) \frac{Q_{j,V}(x)1(E_{j,T})}{f_i(x)} \pi(x) dx. \]
Noting that \( w_{ij,T} \leq C \sqrt{\log T/(Th)} \), one can show that
\[ \mathbb{P}
\left( \left| \frac{1}{T} \sum_{t=1}^{T} w_{ij,T} \varepsilon_{it} \right| > C_U \frac{\log T}{T^{h/2}} \right) \leq CT^{-r}, \]
where \( r > 0 \) can be made arbitrarily large by choosing \( C_U \) large enough. This implies
\[ \mathbb{P}
\left( \max_{1 \leq i < j \leq n} \left| U_{ij,T} \right| > C_U \frac{\log T}{T^{h/2}} \right) = o(1) \]
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 \( \hat{g}_k \) by the infeasible version
\[ \hat{g}_k^*(x) = \frac{1}{n_k} \sum_{i \in G_k} \hat{m}_i(x) \]
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
\[ \mathbb{P}
\left( \left| \hat{g}_k(x) - \hat{g}_k^*(x) \right| > a_{n,T} \right)
\leq \mathbb{P}
\left( \left| \hat{g}_k(x) - \hat{g}_k^*(x) \right| > a_{n,T}, \hat{G}_k = G_k \right) + \mathbb{P}(\hat{G}_k \neq G_k) = o(1), \]
since the first probability on the right-hand side is equal to zero by definition of \( \hat{g}_k \) and \( \hat{g}_k^* \) and the second one is of the order \( o(1) \) by Theorem 3.1. Hence, \( |\hat{g}_k(x) - \hat{g}_k^*(x)| = O_p(a_{n,T}) \)
for any null sequence \( \{a_{n,T}\} \) of positive numbers, which in turn implies that
\[ \sqrt{n_k T h} (\hat{g}_k(x) - g_k(x)) = \sqrt{n_k T h} (\hat{g}_k^*(x) - g_k(x)) + o_p(1). \]
The difference between \( \hat{g}_k \) and \( \hat{g}_k^* \) can thus be asymptotically ignored.

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

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

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

\[
\sqrt{n_k T} \left( \hat{g}_k(x) - g_k(x) \right) = \sqrt{n_k T} \left\{ \frac{1}{n_k} \sum_{i \in G_k} Q_{i,V}(x) / \hat{f}_i(x) + \frac{1}{n_k} \sum_{i \in G_k} Q_{i,B}(x) / \hat{f}_i(x) - \frac{1}{n_k} \sum_{i \in G_k} Q_{i,\gamma}(x) / \hat{f}_i(x) - \frac{1}{n_k} \sum_{i \in G_k} (\overline{Q}_i - \overline{Q}_i) \right\}
\]

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

\[
\frac{1}{n_k} \sum_{i \in G_k} Q_{i,\gamma}(x) / \hat{f}_i(x) = o_p \left( \frac{1}{\sqrt{n_k T}} \right)
\]

(S.19)

\[
\frac{1}{n_k} \sum_{i \in G_k} Q_{i,V}(x) / \hat{f}_i(x) = \frac{1}{n_k} \sum_{i \in G_k} Q_{i,V}(x) / \hat{f}_i(x) + o_p \left( \frac{1}{\sqrt{n_k T}} \right)
\]

(S.20)

\[
\frac{1}{n_k} \sum_{i \in G_k} Q_{i,B}(x) / \hat{f}_i(x) = \frac{1}{n_k} \sum_{i \in G_k} Q_{i,B}(x) / \hat{f}_i(x) + o_p \left( \frac{1}{\sqrt{n_k T}} \right).
\]

(S.21)

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

\[
\sqrt{n_k T} \left( \hat{g}_k(x) - g_k(x) \right) = \sqrt{n_k T} \left\{ \frac{1}{n_k} \sum_{i \in G_k} Q_{i,V}(x) / \hat{f}_i(x) + \frac{1}{n_k} \sum_{i \in G_k} Q_{i,B}(x) / \hat{f}_i(x) \right\} + o_p(1)
\]

\[
= \sqrt{n_k T} \left( \frac{1}{n_k T} \sum_{i \in G_k} \sum_{t=1}^T W_h(X_{it} - x) / \hat{f}_i(x) \right)
\]

\[
+ \sqrt{n_k T} \left( \frac{1}{n_k T} \sum_{i \in G_k} \sum_{t=1}^T W_h(X_{it} - x) / \hat{f}_i(x) [m_i(X_{it}) - m_i(x)] \right) + o_p(1).
\]

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. \( \square \)
Proof of (S.19). In a first step, we show that
\[
R_\gamma := \frac{1}{n_k} \sum_{i \in G_k} \frac{Q_{i,\gamma}(x)}{f_i(x)} - \frac{1}{n_k} \sum_{i \in G_k} \frac{Q_{i,\gamma}(x)}{E[f_i(x)]} = o_p\left(\frac{1}{\sqrt{n_kTh}}\right).
\] (S.22)

To do so, we write \(R_\gamma = R_{\gamma,1} + R_{\gamma,2}\), where
\[
R_{\gamma,1} = \frac{1}{n_k} \sum_{i \in G_k} \frac{\mathbb{E}[\hat{f}_i(x)] - \hat{f}_i(x)}{\mathbb{E}[f_i(x)]^2} Q_{i,\gamma}(x)
\]
\[
R_{\gamma,2} = \frac{1}{n_k} \sum_{i \in G_k} \frac{(\mathbb{E}[\hat{f}_i(x)] - \hat{f}_i(x))^2}{\mathbb{E}[f_i(x)]^2} Q_{i,\gamma}(x).
\]

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
\[
R_{\gamma,1} = \frac{1}{n_k} \sum_{i \in G_k} \frac{1}{\mathbb{E}[f_i(x)]^2} \left\{ \frac{1}{T} \sum_{t=1}^{T} Z_{it}(x) \right\} \times \left\{ \frac{1}{T} \sum_{t=1}^{T} W_h(X_{it} - x) \left( \frac{1}{n-1} \sum_{j=1}^{n} \left[ m_j(X_{jt}) + \varepsilon_{jt} \right] \right) \right\}.
\]

We thus obtain that
\[
\mathbb{E}[R_{\gamma,1}^2] = \frac{1}{n_k^2(n-1)^2} \sum_{i,i' \in G_k} \sum_{j \neq j'} \frac{1}{\mathbb{E}[f_i(x)]^2 \mathbb{E}[f_{i'}(x)]^2} \times \left( \frac{1}{T^4} \sum_{t,t',s,s'} \Psi_{i,i',j,j',t,t',s,s'}(x) \right),
\] (S.23)

where we use the shorthand
\[
\Psi_{i,i',j,j',t,t',s,s'}(x) = \mathbb{E}[Z_{it}(x)W_h(X_{is} - x)\{m_j(X_{js}) + \varepsilon_{js}\}] \times Z_{i't'}(x)W_h(X_{i't'} - x)\{m_{j'}(X_{j's'}) + \varepsilon_{j's'}\}.
\]

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), \(|T^{-4} \sum_{t,t',s,s'}^T \psi_{i,i',j,j',t,t',s,s'}(x)| \leq C(\log T)^2/(Th)^2\) and in case (b),
\[
\left| \frac{1}{T^4} \sum_{t,t',s,s'}^T \psi_{i,i',j,j',t,t',s,s'}(x) \right| \leq \begin{cases} 
C(\log T)^3/(T^3h^2) & \text{for } i \neq i' \\
C(\log T)^2/(T^2h^3) & \text{for } i = i'.
\end{cases}
\]
Plugging these bounds into (S.23), we immediately arrive at $R_{\gamma,1} = o_p(1/\sqrt{n_kT\hbar})$. Furthermore, with the help of Hölder’s inequality and (S.5), we obtain that

$$\begin{align*}
R_{\gamma,2} &\leq \left\{ \max_{1 \leq i \leq n} \sup_{x \in [0,1]} \left\{ \frac{\mathbb{E}[\hat{f}_i(x)] - \hat{f}_i(x)^2}{\mathbb{E}[\hat{f}_i(x)]^2} \right\} \left\{ \frac{1}{n_k} \sum_{i \in G_k} \left( \frac{1}{T} \sum_{t=1}^{T} W_{\gamma}^{4/3}(X_{it} - x) \right)^{3/4} \right\} \\
& \quad \times \left( \frac{1}{T} \sum_{t=1}^{T} \left( \frac{1}{n-1} \sum_{j=1}^{n} \left[ m_j(X_j \hbar) + \varepsilon_j \right] \right)^{4/4} \right) \\
& = O_p\left( \sqrt{\frac{\log T}{T\hbar}} \right)^2 \frac{1}{h^{1/4}(n-1)^{1/2}} = o_p\left( \frac{1}{\sqrt{n_kT\hbar}} \right),
\end{align*}$$

which completes the proof of (S.22).

In the next step, we show that

$$\frac{1}{n_k} \sum_{i \in G_k} \frac{Q_{i,\gamma}(x)}{\mathbb{E}[\hat{f}_i(x)]} = o_p\left( \frac{1}{\sqrt{n_kT\hbar}} \right). \quad (S.24)$$

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

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

where $\Psi_{i,i',j,j',t,t'}(x) = \mathbb{E}[W_h(X_{it} - x)\{m_j(X_j \hbar) + \varepsilon_j\}W_h(X_{i't'} - x)\{m_{j'}(X_{j'} \hbar) + \varepsilon_{j'}\}]$. 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),

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

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),

$$\begin{align*}
\frac{1}{n_k} \sum_{i \in G_k} \frac{Q_{i,\ell}(x)}{\hat{f}_i(x)} - \frac{1}{n_k} \sum_{i \in G_k} \frac{Q_{i,\ell}(x)}{\mathbb{E}[\hat{f}_i(x)]} = o_p\left( \frac{1}{\sqrt{n_kT\hbar}} \right), \quad (S.26)
\end{align*}$$
for $\ell \in \{V, B\}$. With the help of standard bias calculations, we further obtain that

$$\frac{1}{n_k} \sum_{i \in G_k} Q_{i,\ell}(x) - \frac{1}{n_k} \sum_{i \in G_k} Q_{i,\ell}(x) = o_p \left( \frac{1}{\sqrt{n_kT_h}} \right).$$  \hspace{1cm} \text{(S.27)}$$

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

\section*{References}


