Supplementary Material for "A Semiparametric Model for Heterogeneous Panel Data with Fixed Effects"

Lena Körber
London School of Economics

Oliver Linton
University of Cambridge

Michael Vogt
University of Konstanz

January 17, 2015

Abstract

In this supplement, we investigate the finite sample performance of our estimators by means of a simulation study. In addition, we provide the technical details and proofs that are omitted in the paper.

1 Simulation study

To assess the small sample properties of our estimation methods, we simulate data from the following model setup: The regressors X_{it} are i.i.d. draws from a uniform distribution on the unit interval. Moreover, there are $K = 2$ common component functions defined by

$$\mu_1(x) = \sqrt{2}\sin(2\pi x) \qquad \text{and} \qquad \mu_2(x) = \sqrt{2}\cos(2\pi x).$$

These functions are orthonormal with respect to the standard scalar product on $[0, 1]$, i.e., $\int_0^1 \mu_1(x)\mu_2(x)dx = 0$ and $\int_0^1 \mu_k^2(x)dx = 1$ for $k = 1, 2$. We have chosen these functions as they indeed look similar to some of the estimated μ-functions from the application in Section 8 of the paper. As the regressors are uniformly distributed on $[0, 1]$, we obtain that $\mathbb{E}[\mu_k(X_{it})] = 0$ for $k = 1, 2$ and thus $\mathbb{E}[m_i(X_{it})] = 0$ with $m_i(x) = \beta_{i1}\mu_1(x)+\beta_{i2}\mu_2(x)$. Thus, the regression functions fulfill the normalization $\mathbb{E}[m_i(X_{it})] = 0$ that is assumed for identification.

The factor loadings β_{ik} ($i = 1, \dots, n,\ k = 1, \dots, K$) are generated deterministically according to

$$\beta_{i1} = 1 + \frac{i-1}{n-1} \qquad \text{and} \qquad \beta_{i2} = 2 - \frac{i-1}{n-1}.$$

With this choice, the coefficient β_{i1} of the function μ_1 linearly increases from 1 to 2 as the index i grows larger. Similarly, the loading β_{i2} of μ_2 decreases from 2 to 1. Hence, the component function μ_1 becomes more and more important as the index i gets larger and vice versa for the second component μ_2. The weighting matrix W is given by

$$W = \begin{pmatrix} 2/n & \dots & 2/n & 0 & \dots & 0 \\ 0 & \dots & 0 & 2/n & \dots & 2/n \end{pmatrix}.$$

Note that the coefficient matrix B and the weighting matrix W are chosen such that $S = WB$ has full rank. In addition, the μ-functions are orthonormal. Hence, the normalization conditions of Section 3 in the paper are fulfilled. In the simulations, S and μ are re-normalized such that they fulfill condition ($\mathrm{I_W}1$) of the paper.

The individual and time fixed effects α_i and γ_t are i.i.d. standard normal random variables. The model constant μ_0 is set to zero, and the disturbances ε_{it} are i.i.d. normal random variables with zero mean and standard deviation σ_ε. To vary the signal-to-noise ratio in the model, we choose two different values for σ_ε, in particular $\sigma_\varepsilon \in \{1, 2\}$. As can be seen, there is no time series dependence in the error terms and the regressors, and we have only included a very limited form of fixed effects. These simplifications allow us to get a clear picture of the performance of our estimation methods. It goes without saying that they may be relaxed, i.e., we may allow for time series dependence in the model variables and add some more complicated forms of fixed effects.

In what follows, we examine the performance of our estimators $\widehat{\mu}$ and $\widehat{\beta}_i$. Moreover, we assess the small sample behaviour of two estimators of the average regression function $m_{\text{av}}(x) = n^{-1} \sum_{i=1}^n m_i(x)$ defined by $\widehat{m}_{\text{av}}(x) = n^{-1} \sum_{i=1}^n \widehat{m}_i(x)$ and $\widehat{m}^e_{\text{av}}(x) = n^{-1} \sum_{i=1}^n \widehat{m}^e_i(x)$, where $\widehat{m}^e_i(x) = \widehat{\beta}_i^\intercal \widehat{\mu}(x)$ are the reconstructed regression functions. As performance measures, we employ the mean squared errors

$$\text{MSE}(\widehat{\mu}_k) = \int_0^1 \big[\widehat{\mu}_k(x) - \mu_k(x)\big]^2 dx$$

for $k = 1, 2$ along with

$$\text{MSE}(\widehat{m}_{\text{av}}) = \int_0^1 \big[\widehat{m}_{\text{av}}(x) - m_{\text{av}}(x)\big]^2 dx$$

$$\text{MSE}(\widehat{m}^e_{\text{av}}) = \int_0^1 \big[\widehat{m}^e_{\text{av}}(x) - m_{\text{av}}(x)\big]^2 dx.$$

Table 1: Small sample properties of the estimators in the design with $\sigma_\varepsilon = 1$.

a) MSE of $\widehat{m}_{av}$

$T \backslash n$	50	100	150	200
50	0.0449	0.0362	0.0331	0.032
100	0.0425	0.0347	0.0324	0.0311
150	0.042	0.0345	0.0321	0.0309
200	0.0418	0.0343	0.0321	0.0308

b) MSE of $\widehat{m}_{av}^e$

$T \backslash n$	50	100	150	200
50	0.017	0.0124	0.0107	0.0101
100	0.0109	0.0077	0.0071	0.0066
150	0.0092	0.0067	0.0061	0.0058
200	0.0085	0.0062	0.0057	0.0054

c) MSE of $\widehat{\mu}_1$

$T \backslash n$	50	100	150	200
50	0.0159	0.0099	0.0064	0.0052
100	0.008	0.004	0.003	0.0024
150	0.0053	0.0029	0.0021	0.0016
200	0.0041	0.0022	0.0016	0.0012

d) MSE of $\widehat{\mu}_2$

$T \backslash n$	50	100	150	200
50	0.0159	0.0092	0.0063	0.005
100	0.009	0.0051	0.0039	0.0035
150	0.0065	0.0043	0.0035	0.003
200	0.0054	0.0035	0.003	0.0027

e) L_1-norm of the coefficient estimates $\widehat{\beta}_{i1}$

$T \backslash n$	50	100	150	200
50	0.129	0.125	0.124	0.123
100	0.089	0.0853	0.0841	0.0837
150	0.072	0.0684	0.0679	0.0675
200	0.0627	0.0591	0.0583	0.0581

f) L_1-norm of the coefficient estimates $\widehat{\beta}_{i2}$

$T \backslash n$	50	100	150	200
50	0.136	0.13	0.128	0.128
100	0.0973	0.0914	0.0895	0.0886
150	0.0822	0.0752	0.0732	0.0721
200	0.0732	0.0658	0.0641	0.0629

Table 2: Small sample properties of the estimators in the design with $\sigma_\varepsilon = 2$.

a) MSE of $\widehat{m}_{av}$

$T\backslash n$	50	100	150	200
50	0.0512	0.039	0.0352	0.034
100	0.0456	0.0362	0.0334	0.0318
150	0.0442	0.0354	0.0327	0.0314
200	0.0428	0.035	0.0326	0.0312

b) MSE of $\widehat{m}^e_{av}$

$T\backslash n$	50	100	150	200
50	0.0233	0.0153	0.0129	0.0118
100	0.0144	0.00936	0.008	0.00749
150	0.0115	0.00779	0.00682	0.00632
200	0.00993	0.0069	0.00634	0.00579

c) MSE of $\widehat{\mu}_1$

$T\backslash n$	50	100	150	200
50	0.0343	0.019	0.0129	0.0103
100	0.0169	0.0089	0.00604	0.00465
150	0.0106	0.0057	0.00402	0.00294
200	0.00804	0.00418	0.00292	0.00225

d) MSE of $\widehat{\mu}_2$

$T\backslash n$	50	100	150	200
50	0.0339	0.0183	0.0125	0.00993
100	0.0171	0.00942	0.0071	0.00568
150	0.0117	0.007	0.00542	0.00429
200	0.00955	0.00555	0.00433	0.00366

e) L_1-norm of the coefficient estimates $\widehat{\beta}_{i1}$

$T\backslash n$	50	100	150	200
50	0.233	0.231	0.231	0.231
100	0.162	0.162	0.161	0.161
150	0.134	0.131	0.131	0.131
200	0.115	0.113	0.114	0.114

f) L_1-norm of the coefficient estimates $\widehat{\beta}_{i2}$

$T\backslash n$	50	100	150	200
50	0.237	0.234	0.234	0.234
100	0.169	0.166	0.164	0.164
150	0.138	0.135	0.134	0.134
200	0.12	0.118	0.116	0.116

The small sample behavior of the coefficient estimates $\widehat{\beta}_i$ is evaluated by the L_1-norm

$$\frac{1}{n}\sum_{i=1}^{n}|\widehat{\beta}_{ik}-\beta_{ik}|$$

for $k=1,2$. Throughout, we assume the number of components $K=2$ to be known and use the version of our method which is based on local linear estimators. Moreover, the bandwidth is set to $h=0.15$ and we use an Epanechnikov kernel. As a robustness check, we have varied the bandwidth. As this produces very similar results, we have however not reported them here. Finally, the number of replications is set to $N=1000$.

Tables 1 and 2 report the simulation results. Overall, our estimators perform well even for the moderate sample sizes $n=T=50$. The accuracy of the estimators increases steadily as the dimensions n and T grow larger, the only exception being the estimates of the factor loadings which improve above all in T but not so much in n. This is a very natural phenomenon as the factor loadings are estimated from individual time series regressions. Hence, their quality should depend above all on the time series dimension and not so much on the length of the cross-section. It is also worth mentioning that the MSE of the reconstructed average $\widehat{m}^e_{\text{av}}$ is smaller and converges faster to zero than the MSE of $\widehat{m}_{\text{av}}$. This observation is consistent with the asymptotic properties of the estimators $\widehat{m}_i$ and $\widehat{m}^e_i$: While $\widehat{m}_i$ converges at the rate $(Th)^{-1/2}$, $\widehat{m}^e_i$ converges at the faster rate $T^{-1/2}$ (cp. Section 5.4 in the paper). Finally, when the standard deviation σ_ε of the disturbance terms is increased to 2, the signal-to-noise ratio in the model decreases. This makes it harder to estimate the functions and parameters of interest, which is reflected in higher values of the MSE and the L_1-norm as can be seen upon comparing Tables 1 and 2.

2 Technical details

In what follows, we prove the uniform convergence results which are stated in Appendix B of the paper.

Proof of Lemma B1. The proof proceeds by slightly modifying standard arguments to derive uniform convergence rates for kernel estimators. We are thus content with giving some remarks on the necessary modifications.

We start with the proof of (35). Write

$$\mathbb{P}\Big(\max_{1\le i\le n}\sup_{x\in[0,1]}\big|\Psi_i(x)-\mathbb{E}[\Psi_i(x)]\big|>Ca_T\Big)\le\sum_{i=1}^{n}\mathbb{P}\Big(\sup_{x\in[0,1]}\big|\Psi_i(x)-\mathbb{E}[\Psi_i(x)]\big|>Ca_T\Big)$$

with $a_T=\sqrt{\log T/Th}$. Going along the lines of the standard proving strategy, the probabilities on the right-hand side can be bounded by a null sequence $\{c_T\}$ which does not depend on i. Under our conditions, this sequence can be chosen such that $\{nc_T\}$ is a null sequence as well. This yields the result.

We now turn to (34). As the variables Z_{it} are not bounded, we have to replace them by truncated versions $Z_{it}^{\leq} = Z_{it} I(|Z_{it}| \leq \tau_{n,T})$ in a first step. Since we maximize over i, the truncation sequence $\tau_{n,T}$ must be chosen to go to infinity much faster than in the standard case where i is fixed. In particular, we take $\tau_{n,T} = (nT)^{1/(\theta-\delta)}$ for some small $\delta > 0$. Applying the same proving strategy as for (35) to the truncated version of $\Psi_i(x)$, one can see that the arguments still go through. However, as the truncation points $\tau_{n,T}$ diverge much faster than in the standard case with fixed i, the convergence rate turns out to be slower than the standard rate $\sqrt{\log T/Th}$. □

Proof of Lemma B2. As the proof closely follows standard arguments, we only provide a short sketch: Let $a_{n,T} = \sqrt{\log nT/nTh}$ and write $\Psi(x) = \Psi^{\leq}(x) + \Psi^{>}(x)$ with

$$\Psi^{\leq}(x) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} K_h(X_{it}-x) Z_{it} I(|Z_{it}| \leq \tau_{n,T})$$
$$\Psi^{>}(x) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} K_h(X_{it}-x) Z_{it} I(|Z_{it}| > \tau_{n,T}),$$

where the truncation sequence $\tau_{n,T}$ is given by $\tau_{n,T} = (nT)^{1/(\theta-\delta)}$ with some small $\delta > 0$. We thus have

$$\Psi(x) - \mathbb{E}[\Psi(x)] = (\Psi^{\leq}(x) - \mathbb{E}[\Psi^{\leq}(x)]) + (\Psi^{>}(x) - \mathbb{E}[\Psi^{>}(x)]).$$

Straightforward arguments show that $\sup_{x\in[0,1]} |\Psi^{>}(x) - \mathbb{E}[\Psi^{>}(x)]| = O_p(a_{n,T})$. To analyze the term $\sup_{x\in[0,1]} |\Psi^{\leq}(x) - \mathbb{E}[\Psi^{\leq}(x)]|$, we cover the unit interval by a grid of points $G_{n,T}$ that gets finer and finer as the sample size increases. We then replace the supremum over x by the maximum over the grid points $x \in G_{n,T}$ and show that the resulting error is negligible. To complete the proof, we write

$$\mathbb{P}\Big(\max_{x\in G_{n,T}} \big|\Psi^{\leq}(x) - \mathbb{E}[\Psi^{\leq}(x)]\big| > Ca_{n,T}\Big) \leq \sum_{x\in G_{n,T}} \mathbb{P}\Big(\big|\Psi^{\leq}(x) - \mathbb{E}[\Psi^{\leq}(x)]\big| > Ca_{n,T}\Big)$$

and bound the probabilities $\mathbb{P}(|\Psi^{\leq}(x) - \mathbb{E}[\Psi^{\leq}(x)]| > Ca_{n,T})$ for each grid point with the help of an exponential inequality. To do so, let

$$\Psi^{\leq}(x) - \mathbb{E}[\Psi^{\leq}(x)] = \sum_{i=1}^{n}\sum_{t=1}^{T} W_{it}(x)$$

with $W_{it}(x) = \frac{1}{nT}\{K_h(X_{it}-x)Z_{it}I(|Z_{it}| \leq \tau_{n,T}) - \mathbb{E}[K_h(X_{it}-x)Z_{it}I(|Z_{it}| \leq \tau_{n,T})]\}$ and split up the expression $\sum_{t=1}^{T} W_{it}(x)$ into a growing number of blocks of increasing size. Using Bradley's lemma (see Lemma 1.2 in Bosq (1998)), we can replace these blocks by independent versions and apply an exponential inequality. □

Proof of Lemma B3. Throughout the proof, we use the following notation. Let

$\mathcal{C}_T$: the event that $\max_i \sup_x |V_i(x)^{1/\nu}| \le C\sqrt{\log T/Th}$ and $\max_i \sup_x T^{-1}\sum_{t=1}^T K_h(X_{it}-x) \le C$

$\mathcal{C}_{iT}$: the event that $\sup_x |V_i(x)^{1/\nu}| \le C\sqrt{\log T/Th}$ and $\sup_x T^{-1}\sum_{t=1}^T K_h(X_{it}-x) \le C$

for a fixed large constant C. Moreover, write $\mathcal{C}_T^c$ and $\mathcal{C}_{iT}^c$ to denote the complements of $\mathcal{C}_T$ and $\mathcal{C}_{iT}$, respectively. Inspecting the proof of Lemma B1, it is easily seen that $P(\mathcal{C}_T^c) = o(1)$ and $P(\mathcal{C}_{iT}^c) = o(1)$, given that the constant C in the definition of the events $\mathcal{C}_T$ and $\mathcal{C}_{iT}$ is chosen sufficiently large. With this notation at hand, we obtain that

$$
\begin{aligned}
\mathbb{P}\Big(\sup_{x\in[0,1]}\big|\Psi(x)\big| > Ma_{n,T}\Big) &\le \mathbb{P}\Big(\sup_{x\in[0,1]}\big|\Psi(x)\big| > Ma_{n,T}, \mathcal{C}_T\Big) \\
&\quad + \mathbb{P}\Big(\sup_{x\in[0,1]}\big|\Psi(x)\big| > Ma_{n,T}, \mathcal{C}_T^c\Big) \\
&= \mathbb{P}\Big(\sup_{x\in[0,1]}\big|\Psi(x)\big| > Ma_{n,T}, \mathcal{C}_T\Big) + o(1),
\end{aligned}
$$

where $a_{n,T} = (\log nT\sqrt{nTh})^{-1}$ and M is a large positive constant. Moreover,

$$
\begin{aligned}
\mathbb{P}\Big(\sup_{x\in[0,1]}\big|\Psi(x)\big| > Ma_{n,T}, \mathcal{C}_T\Big) &= \mathbb{P}\Big(\sup_{x\in[0,1]}\Big|\frac{1}{n}\sum_{i=1}^n V_i(x)W_i(x)\Big| > Ma_{n,T}, \mathcal{C}_T\Big) \\
&= \mathbb{P}\Big(\sup_{x\in[0,1]}\Big|\frac{1}{n}\sum_{i=1}^n I(\mathcal{C}_T)V_i(x)W_i(x)\Big| > Ma_{n,T}\Big) \\
&\le \mathbb{P}\Big(\sup_{x\in[0,1]}\Big|\frac{1}{n}\sum_{i=1}^n I(\mathcal{C}_{iT})V_i(x)W_i(x)\Big| > Ma_{n,T}\Big).
\end{aligned}
$$

Now write

$$
\frac{1}{n}\sum_{i=1}^n I(\mathcal{C}_{iT})V_i(x)W_i(x) = Q^{\le}(x) + Q^{>}(x)
$$

with the two terms on the right-hand side being defined as

$$
\begin{aligned}
Q^{\le}(x) &= \frac{1}{n}\sum_{i=1}^n I(\mathcal{C}_{iT})V_i(x)W_i^{\le}(x) \\
Q^{>}(x) &= \frac{1}{n}\sum_{i=1}^n I(\mathcal{C}_{iT})V_i(x)W_i^{>}(x).
\end{aligned}
$$

Here, $W_i(x) = W_i^{\le}(x) + W_i^{>}(x)$ with

$$
W_i^{\le}(x) = \frac{1}{T}\sum_{t=1}^T K_h(X_{it}-x)Z_{it}^{\le}
$$

$$W_i^>(x) = \frac{1}{T}\sum_{t=1}^T K_h(X_{it}-x)Z_{it}^>$$

and $Z_{it} = Z_{it}^{\leq} + Z_{it}^>$ with

$$\begin{aligned}
Z_{it}^{\leq} &= Z_{it}I(|Z_{it}| \leq \tau_{n,T}) - \mathbb{E}[Z_{it}I(|Z_{it}| \leq \tau_{n,T})|X_{it}] \\
Z_{it}^{>} &= Z_{it}I(|Z_{it}| > \tau_{n,T}) - \mathbb{E}[Z_{it}I(|Z_{it}| > \tau_{n,T})|X_{it}],
\end{aligned}$$

where the truncation sequence $\tau_{n,T}$ is chosen to equal $\tau_{n,T} = (nT)^{1/(\theta-\delta)}$ for some small $\delta > 0$. We now arrive at

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{x\in[0,1]}\Big|\frac{1}{n}\sum_{i=1}^n I(\mathcal{C}_{iT})V_i(x)W_i(x)\Big| > Ma_{n,T}\Big) \\
&\qquad \leq \mathbb{P}\Big(\sup_{x\in[0,1]}|Q^{\leq}(x)| > \frac{M}{2}a_{n,T}\Big) + \mathbb{P}\Big(\sup_{x\in[0,1]}|Q^{>}(x)| > \frac{M}{2}a_{n,T}\Big).
\end{aligned}$$

In the remainder of the proof, we show that the two terms on the right-hand side converge to zero as the sample size goes to infinity. To do so, we proceed in several steps.

Step 1. We start by considering the term $Q^>(x)$. It holds that

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{x\in[0,1]}\Big|\frac{1}{n}\sum_{i=1}^n I(\mathcal{C}_{iT})V_i(x)\Big(\frac{1}{T}\sum_{t=1}^T K_h(X_{it}-x)Z_{it}I(|Z_{it}|>\tau_{n,T})\Big)\Big| > Ca_{n,T}\Big) \\
&\leq \mathbb{P}\Big(|Z_{it}| > \tau_{n,T} \text{ for some } 1 \leq i \leq n \text{ and } 1 \leq t \leq T\Big) \\
&\leq \sum_{i=1}^n\sum_{t=1}^T \mathbb{P}(|Z_{it}| > \tau_{n,T}) \leq \sum_{i=1}^n\sum_{t=1}^T \mathbb{E}\Big[\frac{|Z_{it}|^\theta}{\tau_{n,T}^\theta}\Big] \leq C\frac{nT}{\tau_{n,T}^\theta} \to 0.
\end{aligned}$$

In addition,

$$\begin{aligned}
&\sup_{x\in[0,1]}\Big|\frac{1}{n}\sum_{i=1}^n I(\mathcal{C}_{iT})V_i(x)\Big(\frac{1}{T}\sum_{t=1}^T K_h(X_{it}-x)\mathbb{E}[Z_{it}I(|Z_{it}|>\tau_{n,T})|X_{it}]\Big)\Big| \\
&\leq C\sqrt{\frac{\log T}{Th}}\max_{1\leq i\leq n}\max_{1\leq t\leq T}\mathbb{E}\big[|Z_{it}|I(|Z_{it}|>\tau_{n,T})|X_{it}\big] \\
&\leq C\sqrt{\frac{\log T}{Th}}\frac{1}{\tau_{n,T}^{\theta-1}} \leq Ca_{n,T},
\end{aligned}$$

where the third line follows by (A4'). As a result,

$$\mathbb{P}\Big(\sup_{x\in[0,1]}|Q^>(x)| > \frac{M}{2}a_{n,T}\Big) = o(1)$$

for M sufficiently large.

Step 2. We now turn to the analysis of the term $Q^{\leq}(x)$. Cover the region $[0,1]$ with open intervals J_l ($l = 1, \ldots, L_{n,T}$) of length $C/L_{n,T}$ and let x_l be the midpoint of the interval J_l. Then

$$\sup_{x\in[0,1]} |Q^{\leq}(x)| \leq \max_{1\leq l\leq L_{n,T}} |Q^{\leq}(x_l)| + \max_{1\leq l\leq L_{n,T}} \sup_{x\in J_l} |Q^{\leq}(x) - Q^{\leq}(x_l)|.$$

For any point $x \in J_l$, we have

$$I(\mathcal{C}_{iT})\big|V_i(x)W_i^{\leq}(x) - V_i(x_l)W_i^{\leq}(x_l)\big| \leq \frac{C\tau_{n,T}}{h^2}|x - x_l| \leq \frac{C\tau_{n,T}}{h^2 L_{n,T}}.$$

Therefore,

$$\max_{1\leq l\leq L_{n,T}} \sup_{x\in J_l} |Q^{\leq}(x) - Q^{\leq}(x_l)| \leq \frac{C\tau_{n,T}}{h^2 L_{n,T}}.$$

Choosing $L_{n,T} \to \infty$ with $L_{n,T} = C\tau_{n,T}/a_{n,T}h^2$, we obtain that

$$\max_{1\leq l\leq L_{n,T}} \sup_{x\in J_l} |Q^{\leq}(x) - Q^{\leq}(x_l)| \leq C a_{n,T}.$$

If we pick the constant M large enough, we thus arrive at

$$\mathbb{P}\Big(\sup_{x\in[0,1]} |Q^{\leq}(x)| > \frac{M}{2}a_{n,T}\Big) \leq \mathbb{P}\Big(\max_{1\leq l\leq L_{n,T}} |Q^{\leq}(x_l)| > \frac{M}{4}a_{n,T}\Big) + o(1).$$

Step 3. It remains to show that

$$\mathbb{P}\Big(\max_{1\leq l\leq L_{n,T}} |Q^{\leq}(x_l)| > \frac{M}{4}a_{n,T}\Big) = o(1)$$

for some large fixed constant M. To do so, we write

$$\mathbb{P}\Big(\max_{1\leq l\leq L_{n,T}} |Q^{\leq}(x_l)| > \frac{M}{4}a_{n,T}\Big) \leq P_1 + P_2$$

with

$$\begin{aligned} P_1 &= \mathbb{P}\Big(\max_{1\leq l\leq L_{n,T}} |Q^{\leq}(x_l) - \mathbb{E}Q^{\leq}(x_l)| > \frac{M}{8}a_{n,T}\Big) \\ P_2 &= \mathbb{P}\Big(\max_{1\leq l\leq L_{n,T}} |\mathbb{E}Q^{\leq}(x_l)| > \frac{M}{8}a_{n,T}\Big). \end{aligned}$$

First consider the term P_2. If $\nu \geq 3$, then

$$\begin{aligned} |\mathbb{E}Q^{\leq}(x_l)| &= \Big|\frac{1}{n}\sum_{i=1}^n \mathbb{E}[I(\mathcal{C}_{iT})V_i(x_l)W_i^{\leq}(x_l)]\Big| \\ &\leq \frac{1}{n}\sum_{i=1}^n \mathbb{E}\big[I(\mathcal{C}_{iT})V_i(x_l)^2\big]^{1/2}\mathbb{E}\big[W_i^{\leq}(x_l)^2\big]^{1/2} \end{aligned}$$

$$\leq \frac{C}{\sqrt{Th}}\Big(\frac{\log T}{Th}\Big)^{\nu/2} = o(a_{n,T}).$$

For $\nu \leq 2$, we write

$$\begin{aligned}
|\mathbb{E}Q^{\leq}(x_l)| &= \Big|\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[I(\mathcal{C}_{iT})V_i(x_l)W_i^{\leq}(x_l)]\Big| \\
&\leq \Big|\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[V_i(x_l)W_i^{\leq}(x_l)]\Big| + \Big|\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[I(\mathcal{C}_{iT}^c)V_i(x_l)W_i^{\leq}(x_l)]\Big|.
\end{aligned}$$

If $\nu = 1$, we have

$$\begin{aligned}
\big|\mathbb{E}[V_i(x_l)W_i^{\leq}(x_l)]\big| &= \Big|\frac{1}{T^2}\sum_{s,t=1}^{T}\mathbb{E}\big[(K_h(X_{is}-x_l) - \mathbb{E}[K_h(X_{is}-x_l)])K_h(X_{it}-x_l)Z_{it}^{\leq}\big]\Big| \\
&= \Big|\frac{1}{T^2}\sum_{\substack{s,t=1 \\ s\neq t}}^{T}\mathbb{E}\big[(K_h(X_{is}-x_l) - \mathbb{E}[K_h(X_{is}-x_l)])K_h(X_{it}-x_l)Z_{it}^{\leq}\big]\Big| \\
&\leq \frac{C\log T}{T} = o(a_{n,T}),
\end{aligned}$$

the last line following with the help of Davydov's inequality and (A4'). For $\nu = 2$, it holds that

$$\begin{aligned}
\big|\mathbb{E}[V_i(x_l)W_i^{\leq}(x_l)]\big| &= \Big|\frac{1}{T^3}\sum_{s,s',t=1}^{T}\mathbb{E}\big[(K_h(X_{is}-x_l) - \mathbb{E}[K_h(X_{is}-x_l)]) \\
&\qquad \times (K_h(X_{is'}-x_l) - \mathbb{E}[K_h(X_{is'}-x_l)])K_h(X_{it}-x_l)Z_{it}^{\leq}\big]\Big| \\
&\leq \frac{CT(\log T)^2}{T^3h^2} = C\Big(\frac{\log T}{Th}\Big)^2 = o(a_{n,T}),
\end{aligned}$$

the last line again following by Davydov's inequality and (A4'). In addition,

$$\mathbb{E}[I(\mathcal{C}_{iT}^c)V_i(x_l)W_i^{\leq}(x_l)] \leq \mathbb{E}[I(\mathcal{C}_{iT}^c)]^{1/2}\mathbb{E}[V_i(x_l)^2W_i^{\leq}(x_l)^2]^{1/2}.$$

Repeating the usual strategy to prove uniform convergence for kernel estimates, it can be shown that under our assumptions, $\mathbb{E}[I(\mathcal{C}_{iT}^c)] = \mathbb{P}(\mathcal{C}_{iT}^c) \leq T^{-C}$ for an arbitrarily large constant C. This yields that $\mathbb{E}[I(\mathcal{C}_{iT}^c)V_i(x_l)W_i^{\leq}(x_l)] = o(a_{n,T})$ uniformly over l, which in turn implies that $|\mathbb{E}Q^{\leq}(x_l)| = o(a_{n,T})$ uniformly over l for $\nu = 1, 2$. As a result, $P_2 = o(1)$ for any $\nu \geq 1$.

To cope with the term P_1, we apply the bound

$$P_1 \leq \sum_{l=1}^{L_{n,T}}\mathbb{P}\Big(|Q^{\leq}(x_l) - \mathbb{E}Q^{\leq}(x_l)| > \frac{M}{8}a_{n,T}\Big)$$

and consider the probability $\mathbb{P}(|Q^{\leq}(x_l) - \mathbb{E}Q^{\leq}(x_l)| > Ma_{n,T}/8)$ for an arbitrary fixed grid point x_l. Write

$$Q^{\leq}(x_l) - \mathbb{E}Q^{\leq}(x_l) = \sum_{i=1}^{n} \xi_i(x_l)$$

with $\xi_i(x_l) = n^{-1}\{I(\mathcal{C}_{iT})V_i(x_l)W_i^{\leq}(x_l) - \mathbb{E}[I(\mathcal{C}_{iT})V_i(x_l)W_i^{\leq}(x_l)]\}$. Recalling the definition of the events $\mathcal{C}_{iT}$, the variables $\xi_i(x_l)$ can be bounded as follows:

$$|\xi_i(x_l)| \leq C\sqrt{\frac{\log T}{Th}}\frac{\tau_{n,T}}{n} \leq \frac{C}{(nTh)^{1/2+\delta}} := \overline{C}_{n,T}$$

with some sufficiently large constant C and a small $\delta > 0$, given that $n \gg T^{2/3}$ and $\theta > 5$. With $\lambda_{n,T} = \overline{C}_{n,T}^{-1}/2$, we obtain that $\lambda_{n,T}|\xi_i(x_l)| \leq 1/2$. As $\exp(x) \leq 1 + x + x^2$ for $|x| \leq 1/2$,

$$\mathbb{E}\Big[\exp\big(\pm\lambda_{n,T}\xi_i(x_l)\big)\Big] \leq 1 + \lambda_{n,T}^2\mathbb{E}[\xi_i(x_l)^2] \leq \exp\Big(\lambda_{n,T}^2\mathbb{E}[\xi_i(x_l)^2]\Big).$$

Using this together with Markov's inequality, we arrive at

$$\begin{aligned}
&\mathbb{P}\Big(\Big|\sum_{i=1}^{n}\xi_i(x_l)\Big| > \frac{M}{8}a_{n,T}\Big) \\
&\leq \exp\Big(-\frac{M}{8}\lambda_{n,T}a_{n,T}\Big)\Big\{\mathbb{E}\Big[\exp\big(\lambda_{n,T}\sum_{i=1}^{n}\xi_i(x_l)\big)\Big] + \mathbb{E}\Big[\exp\big(-\lambda_{n,T}\sum_{i=1}^{n}\xi_i(x_l)\big)\Big]\Big\} \\
&\leq 2\exp\Big(-\frac{M}{8}\lambda_{n,T}a_{n,T}\Big)\prod_{i=1}^{n}\exp\big(\lambda_{n,T}^2\mathbb{E}[\xi_i(x_l)^2]\big) \\
&= 2\exp\Big(-\frac{M}{8}\lambda_{n,T}a_{n,T}\Big)\exp\Big(\lambda_{n,T}^2\sum_{i=1}^{n}\mathbb{E}[\xi_i(x_l)^2]\Big).
\end{aligned}$$

Now note that

$$\mathbb{E}[\xi_i(x_l)^2] \leq \frac{1}{n^2}\mathbb{E}[I(\mathcal{C}_{iT})V_i(x_l)^2W_i^{\leq}(x_l)^2] \leq \frac{C\log T}{n^2Th}\mathbb{E}[W_i^{\leq}(x_l)^2]$$

and

$$\begin{aligned}
\mathbb{E}[W_i^{\leq}(x_l)^2] &= \frac{1}{T^2}\sum_{s,t=1}^{T}\mathbb{E}\big[K_h(X_{is}-x_l)K_h(X_{it}-x_l)Z_{is}^{\leq}Z_{it}^{\leq}\big] \\
&= \frac{1}{T^2}\sum_{s,t=1}^{T}\mathrm{Cov}\big(K_h(X_{is}-x_l)Z_{is}^{\leq}, K_h(X_{it}-x_l)Z_{it}^{\leq}\big) \leq \frac{C}{Th}.
\end{aligned}$$

Hence, $\mathbb{E}[\xi_i(x_l)^2] \leq C\log T/(nTh)^2$ and

$$\lambda_{n,T}^2\sum_{i=1}^{n}\mathbb{E}[\xi_i(x_l)^2] \leq C(nTh)^{1+2\delta}\frac{\log T}{n(Th)^2} \leq C\frac{(nT)^{2\delta}}{Th} = o(1).$$

Moreover,

$$\lambda_{n,T} a_{n,T} = \frac{(nTh)^{1/2+\delta}}{\log nT(nTh)^{1/2}} \to \infty$$

at polynomial rate. As a result,

$$\mathbb{P}\Big(\Big| \sum_{i=1}^{n} \xi_i(x_l) \Big| > \frac{M}{8} a_{n,T} \Big) \leq CT^{-p},$$

where the constant $p > 0$ can be chosen arbitrarily large. This completes the proof. □

Proof of Lemma B4. The proof is similar to that of Lemma B3 with the roles of i and t being reversed. Let $a_{n,T} = (\log nT\sqrt{nTh})^{-1}$ and $\tau_{n,T} = (nT)^{1/(\theta-\delta)}$ for some small $\delta > 0$. Arguments analogous to those for Step 1 in the proof of Lemma B3 yield that $\Psi(x)$ can be replaced by the term

$$Q^{\leq}(x) = \frac{1}{T} \sum_{t=1}^{T} I(\mathcal{C}_{tn}) V_t(x) W_t^{\leq},$$

where $W_t^{\leq} = \frac{1}{n} \sum_{i=1}^{n} Z_{it}^{\leq}$ with $Z_{it}^{\leq} = Z_{it} I(|Z_{it}| \leq \tau_{n,T}) - \mathbb{E}[Z_{it} I(|Z_{it}| \leq \tau_{n,T})]$ and $\mathcal{C}_{tn}$ is the event that $\sup_x |V_t(x)| \leq C\sqrt{\log n/nh}$ for some sufficiently large constant C. Next cover the unit interval by a grid of $L_{n,T} = C\tau_{n,T}/a_{n,T}h^2$ points. As in the proof of Lemma B3, we can show that

$$\sup_{x\in[0,1]} |Q^{\leq}(x)| = \max_{1\leq l\leq L_{n,T}} |Q^{\leq}(x_l)| + O(a_{n,T}).$$

Moreover, again repeating the arguments from Lemma B3, we obtain that for some sufficiently large constant M,

$$\begin{aligned} \mathbb{P}\Big(\max_{1\leq l\leq L_{n,T}} |Q^{\leq}(x_l)| > M a_{n,T} \Big) &\leq \mathbb{P}\Big(\max_{1\leq l\leq L_{n,T}} |Q^{\leq}(x_l) - \mathbb{E}Q^{\leq}(x_l)| > \frac{M}{2} a_{n,T} \Big) + o(1) \\ &\leq \sum_{l=1}^{L_{n,T}} \mathbb{P}\Big(|Q^{\leq}(x_l) - \mathbb{E}Q^{\leq}(x_l)| > \frac{M}{2} a_{n,T} \Big) + o(1). \end{aligned}$$

To complete the proof, we bound the probability $\mathbb{P}(|Q^{\leq}(x) - \mathbb{E}Q^{\leq}(x)| > \frac{M}{2} a_{n,T})$ for an arbitrary point x by an exponential inequality. To do so, we must slightly vary the arguments for Lemma B3, taking into account the fact that $Q^{\leq}(x)$ is not a sum of independent terms any more. In particular, we write

$$Q^{\leq}(x) - \mathbb{E}Q^{\leq}(x) = \sum_{t=1}^{T} \xi_t(x)$$

with $\xi_t(x) = T^{-1}\{I(\mathcal{C}_{tn}) V_t(x) W_t^{\leq} - \mathbb{E}[I(\mathcal{C}_{tn}) V_t(x) W_t^{\leq}]\}$ and split up the expression

$\sum_{t=1}^{T} \xi_t(x)$ into blocks as follows:

$$\sum_{t=1}^{T} \xi_t(x) = \sum_{s=1}^{q_{n,T}} B_{2s-1}(x) + \sum_{s=1}^{q_{n,T}} B_{2s}(x)$$

with $B_s(x) = \sum_{t=(s-1)r_{n,T}+1}^{sr_{n,T}} \xi_t(x)$, where $2q_{n,T}$ is the number of blocks and $r_{n,T} = T/2q_{n,T}$ is the block length. We now get

$$\begin{aligned} \mathbb{P}\Big(\Big|\sum_{t=1}^{T} \xi_t(x)\Big| > \frac{M}{2} a_{n,T}\Big) \le\ & \mathbb{P}\Big(\Big|\sum_{s=1}^{q_{n,T}} B_{2s-1}(x)\Big| > \frac{M}{4} a_{n,T}\Big) \\ & + \mathbb{P}\Big(\Big|\sum_{s=1}^{q_{n,T}} B_{2s}(x)\Big| > \frac{M}{4} a_{n,T}\Big). \end{aligned}$$

In what follows, we restrict attention to the first term on the right-hand side of the above display. The second one can be analyzed by analogous arguments. We make use of the following two facts:

(1) Let $\mathcal{V}^{(i)} = \{\mathcal{V}_t^{(i)} : t = 1, \dots, T\} = \{(X_{it}, Z_{it}) : t = 1, \dots, T\}$ be the time series of the i-th individual and consider the time series $\mathcal{W} = \{\mathcal{W}_t : t = 1, \dots, T\}$ with $\mathcal{W}_t = h_t(\mathcal{V}_t^{(1)}, \dots, \mathcal{V}_t^{(n)}) = h_t(X_{1t}, Z_{1t}, \dots, X_{nt}, Z_{nt})$ for some Borel functions h_t. Then by Theorem 5.2 in Bradley (2005) and the comments thereafter, the mixing coefficients $\alpha^{\mathcal{W}}(k)$ of the time series $\mathcal{W}$ are such that $\alpha^{\mathcal{W}}(k) \le \sum_{i=1}^{n} \alpha_i(k) \le n\alpha(k)$ for each $k \in \mathbb{N}$. In particular, letting $\alpha^{\xi}(k)$ be the mixing coefficients of the time series $\{\xi_t(x)\}$, it holds that $\alpha^{\xi}(k) \le n\alpha(k)$.

(2) By Bradley's lemma (see Lemma 1.2 in Bosq (1998)), we can construct a sequence of random variables $B_1^*(x), B_3^*(x), \dots$ such that (i) $B_1^*(x)$, $B_3^*(x)$, $\dots$ are independent, (ii) $B_{2s-1}^*(x)$ has the same distribution as $B_{2s-1}(x)$, and (iii) for $0 < \mu \le \|B_{2s-1}(x)\|_\infty$, it holds that

$$\mathbb{P}\big(|B_{2s-1}^*(x) - B_{2s-1}(x)| > \mu\big) \le 18\Big(\frac{\|B_{2s-1}(x)\|_\infty}{\mu}\Big)^{1/2} \alpha^{\xi}(r_{n,T}). \tag{S1}$$

Using fact (2), we can write

$$\mathbb{P}\Big(\Big|\sum_{s=1}^{q_{n,T}} B_{2s-1}(x)\Big| > \frac{M}{4} a_{n,T}\Big) \le P_1 + P_2$$

with

$$\begin{aligned} P_1 &= \mathbb{P}\Big(\Big|\sum_{s=1}^{q_{n,T}} B_{2s-1}^*(x)\Big| > \frac{M}{8} a_{n,T}\Big) \\ P_2 &= \mathbb{P}\Big(\Big|\sum_{s=1}^{q_{n,T}} \big(B_{2s-1}(x) - B_{2s-1}^*(x)\big)\Big| > \frac{M}{8} a_{n,T}\Big). \end{aligned}$$

We first consider P_1. Picking the block length to equal $r_{n,T} = (nT)^\eta$ for some small $\eta > 0$, it holds that $|B_{2s-1}(x)| \le C\sqrt{\frac{\log n}{nh}}\frac{\tau_{n,T} r_{n,T}}{T} \le \frac{C}{(nTh)^{1/2+\delta}} =: \overline{C}_{n,T}$ with some sufficiently large constant C and a small $\delta > 0$. Choosing $\lambda_{n,T} = \overline{C}_{n,T}^{-1}/2$ and applying Markov's inequality, the same arguments as in Lemma B3 yield that

$$P_1 \le 2\exp\Big(-\frac{M}{8}\lambda_{n,T}a_{n,T} + \lambda_{n,T}^2 \sum_{s=1}^{q_{n,T}} \mathbb{E}[B_{2s-1}^*(x)^2]\Big).$$

Since $\sum_{s=1}^{q_{n,T}} \mathbb{E}[B_{2s-1}^*(x)^2] \le C \log n \log T / n^2 Th$, we finally arrive at

$$P_1 \le 2\exp\Big(-\frac{M}{8}\lambda_{n,T}a_{n,T} + C\lambda_{n,T}^2 \frac{\log n \log T}{n^2 Th}\Big).$$

Direct calculations show that $\lambda_{n,T}a_{n,T} \to \infty$, whereas $\lambda_{n,T}^2 \frac{\log n \log T}{n^2 Th} = o(1)$. This implies that P_1 converges to zero at an arbitarily fast polynomial rate. Moreover, using (S1) together with the fact that $\alpha^\xi(k) \le n\alpha(k)$ and recalling that the coefficients $\alpha(k)$ decay exponentially fast to zero, it immediately follows that P_2 converges to zero at an arbitrarily fast polynomial rate as well. From this, the result easily follows. $\square$

Proof of Lemma B5. Let $\mathcal{C}_T$ be the event that $\max_{1\le i\le n}\sup_{x\in[0,1]}|\widehat{\phi}_i(x)| \le Cb_{n,T}$ and $\mathcal{C}_{iT}$ the event that $\sup_{x\in[0,1]}|\widehat{\phi}_i(x)| \le Cb_{n,T}$. Moreover, write $\mathcal{C}_T^c$ and $\mathcal{C}_{iT}^c$ to denote the complements of $\mathcal{C}_T$ and $\mathcal{C}_{iT}$, respectively. By assumption, $P(\mathcal{C}_T^c) = o(1)$ and $P(\mathcal{C}_{iT}^c) = o(1)$. With this notation at hand, we have

$$\begin{aligned}
\mathbb{P}\Big(\sup_{x\in[0,1]}\big|\Psi(x)\big| > Ma_{n,T}\Big) &\le \mathbb{P}\Big(\sup_{x\in[0,1]}\big|\Psi(x)\big| > Ma_{n,T}, \mathcal{C}_T\Big) \\
&\quad + \mathbb{P}\Big(\sup_{x\in[0,1]}\big|\Psi(x)\big| > Ma_{n,T}, \mathcal{C}_T^c\Big) \\
&\le \mathbb{P}\Big(\sup_{x\in[0,1]}\big|\Psi(x)\big| > Ma_{n,T}, \mathcal{C}_T\Big) + o(1),
\end{aligned}$$

where $a_{n,T} = \sqrt{\frac{\log nT}{nTh(nT)^\eta}}$ and M is a positive constant. Moreover,

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{x\in[0,1]}\big|\Psi(x)\big| > Ma_{n,T}, \mathcal{C}_T\Big) \\
&= \mathbb{P}\Big(\sup_{x\in[0,1]}\Big|\frac{1}{n}\sum_{i=1}^n\Big(\frac{1}{nT}\sum_{j\ne i}\sum_{t=1}^T I(\mathcal{C}_T)\varphi_{it}(x)Z_{jt}\Big)\Big| > Ma_{n,T}\Big) \\
&\le \mathbb{P}\Big(\sup_{x\in[0,1]}\Big|\frac{1}{n}\sum_{i=1}^n\Big(\frac{1}{nT}\sum_{j\ne i}\sum_{t=1}^T I(\mathcal{C}_{iT})\varphi_{it}(x)Z_{jt}\Big)\Big| > Ma_{n,T}\Big).
\end{aligned}$$

Defining

$$Z_{jt}^{\le} = Z_{jt}I(|Z_{jt}| \le \tau_{n,T}) - \mathbb{E}\big[Z_{jt}I(|Z_{jt}| \le \tau_{n,T})\big]$$

$$Z_{jt}^{>} = Z_{jt} I(|Z_{jt}| > \tau_{n,T}) - \mathbb{E}\big[Z_{jt} I(|Z_{jt}| > \tau_{n,T})\big]$$

with $\tau_{n,T} = (nT)^{1/(\theta-\delta)}$ for some small $\delta > 0$, we further get that

$$\frac{1}{n}\sum_{i=1}^{n}\Big(\frac{1}{nT}\sum_{j\neq i}\sum_{t=1}^{T} I(\mathcal{C}_{iT})\varphi_{it}(x)Z_{jt}\Big) = Q^{\leq}(x) + Q^{>}(x)$$

with

$$\begin{aligned}
Q^{\leq}(x) &= \frac{1}{n}\sum_{i=1}^{n}\Big(\frac{1}{nT}\sum_{j\neq i}\sum_{t=1}^{T} I(\mathcal{C}_{iT})\varphi_{it}(x)Z_{jt}^{\leq}\Big)\\
Q^{>}(x) &= \frac{1}{n}\sum_{i=1}^{n}\Big(\frac{1}{nT}\sum_{j\neq i}\sum_{t=1}^{T} I(\mathcal{C}_{iT})\varphi_{it}(x)Z_{jt}^{>}\Big).
\end{aligned}$$

Hence,

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{x\in[0,1]}\Big|\frac{1}{n}\sum_{i=1}^{n}\Big(\frac{1}{nT}\sum_{j\neq i}\sum_{t=1}^{T} I(\mathcal{C}_{iT})\varphi_{it}(x)Z_{jt}\Big)\Big| > M a_{n,T}\Big)\\
&\quad\leq \mathbb{P}\Big(\sup_{x\in[0,1]}\big|Q^{\leq}(x)\big| > \frac{M}{2}a_{n,T}\Big) + \mathbb{P}\Big(\sup_{x\in[0,1]}\big|Q^{>}(x)\big| > \frac{M}{2}a_{n,T}\Big).
\end{aligned}$$

In what follows, we show that the two terms on the right-hand side converge to zero as the sample size increases. The proof splits up into several steps.

Step 1. We first consider $Q^{>}(x)$. Similarly to Lemma B3, it holds that

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{x\in[0,1]}\Big|\frac{1}{n}\sum_{i=1}^{n}\Big(\frac{1}{nT}\sum_{j\neq i}\sum_{t=1}^{T} I(\mathcal{C}_{iT})\varphi_{it}(x)Z_{jt}I(|Z_{jt}| > \tau_{n,T})\Big)\Big| > C a_{n,T}\Big)\\
&\leq \mathbb{P}\Big(|Z_{jt}| > \tau_{n,T} \text{ for some } 1 \leq j \leq n \text{ and } 1 \leq t \leq T\Big) \to 0
\end{aligned}$$

and

$$\begin{aligned}
&\sup_{x\in[0,1]}\Big|\frac{1}{n}\sum_{i=1}^{n}\Big(\frac{1}{nT}\sum_{j\neq i}\sum_{t=1}^{T} I(\mathcal{C}_{iT})\varphi_{it}(x)\mathbb{E}\big[Z_{jt}I(|Z_{jt}| > \tau_{n,T})\big]\Big)\Big|\\
&\leq \frac{Cb_{n,T}}{n^2Th}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{t=1}^{T}\mathbb{E}\big[|Z_{jt}|I(|Z_{jt}| > \tau_{n,T})\big] \leq \frac{Cb_{n,T}}{\tau_{n,T}^{\theta-1}h} \leq C a_{n,T}.
\end{aligned}$$

From this, it immediately follows that $\mathbb{P}(\sup_{x\in[0,1]}|Q^{>}(x)| > M a_{n,T}/2) = o(1)$ for M sufficiently large.

Step 2. We now turn to the analysis of $Q^{\leq}(x)$. Let $L_{n,T} \to \infty$ with $L_{n,T} = \max\{\frac{\tau_{n,T}c_{n,T}}{ha_{n,T}}, \frac{b_{n,T}\tau_{n,T}}{h^2a_{n,T}}, (nT)^{\delta}\}$ for some small $\delta > 0$. Cover the region $[0,1]$ with open intervals J_l

$(l = 1, \ldots, L_{n,T})$ of length $C/L_{n,T}$ and let x_l be the midpoint of the interval J_l. Then for $x \in J_l$,

$$
\begin{aligned}
\big|Q^{\leq}(x) - Q^{\leq}(x_l)\big| &\leq \frac{C\tau_{n,T}}{n} \sum_{i=1}^{n} \Big(\frac{1}{nT} \sum_{j \neq i} \sum_{t=1}^{T} I(\mathcal{C}_{iT}) |\varphi_{it}(x) - \varphi_{it}(x_l)| \Big) \\
&\leq \frac{C\tau_{n,T}}{n} \sum_{i=1}^{n} \Big(\frac{1}{nT} \sum_{j \neq i} \sum_{t=1}^{T} I(\mathcal{C}_{iT}) \big\{ K_h(X_{it} - x) |\widehat{\phi}_i(x) - \widehat{\phi}_i(x_l)| \\
&\qquad\qquad + |\widehat{\phi}_i(x_l)| |K_h(X_{it} - x) - K_h(X_{it} - x_l)| \big\} \Big) \\
&\leq C\tau_{n,T} \Big(\frac{c_{n,T}}{h} + \frac{b_{n,T}}{h^2} \Big) |x - x_l| \leq C \frac{\tau_{n,T}}{L_{n,T}} \Big(\frac{c_{n,T}}{h} + \frac{b_{n,T}}{h^2} \Big) \leq C a_{n,T}
\end{aligned}
$$

with probability tending to one. From this, it immediately follows that for sufficiently large M,

$$
\mathbb{P}\Big(\sup_{x \in [0,1]} \big|Q^{\leq}(x)\big| > \frac{M}{2} a_{n,T} \Big) \leq \mathbb{P}\Big(\max_{1 \leq l \leq L_{n,T}} \big|Q^{\leq}(x_l)\big| > \frac{M}{4} a_{n,T} \Big) + o(1).
$$

Step 3. It remains to show that

$$
\mathbb{P}\Big(\max_{1 \leq l \leq L_{n,T}} \big|Q^{\leq}(x_l)\big| > \frac{M}{4} a_{n,T} \Big) = o(1)
$$

for some sufficiently large constant M. Writing

$$
\max_{1 \leq l \leq L_{n,T}} \big|Q^{\leq}(x_l)\big| \leq \max_{\substack{1 \leq i \leq n \\ 1 \leq l \leq L_{n,T}}} \Big| \sum_{j \neq i} \sum_{t=1}^{T} I(\mathcal{C}_{iT}) \varphi_{it}(x_l) W_{jt} \Big|
$$

with $W_{jt} = \frac{1}{nT}\{Z_{jt} I(|Z_{jt}| \leq \tau_{n,T}) - \mathbb{E}[Z_{jt} I(|Z_{jt}| \leq \tau_{n,T})]\}$, we obtain

$$
\begin{aligned}
&\mathbb{P}\Big(\max_{1 \leq l \leq L_{n,T}} \big|Q^{\leq}(x_l)\big| > \frac{M}{4} a_{n,T} \Big) \\
&\leq \mathbb{P}\Big(\max_{\substack{1 \leq i \leq n \\ 1 \leq l \leq L_{n,T}}} \Big| \sum_{j \neq i} \sum_{t=1}^{T} I(\mathcal{C}_{iT}) \varphi_{it}(x_l) W_{jt} \Big| > \frac{M}{4} a_{n,T} \Big) \\
&\leq \sum_{i=1}^{n} \sum_{l=1}^{L_{n,T}} \mathbb{P}\Big(\Big| \sum_{j \neq i} \sum_{t=1}^{T} I(\mathcal{C}_{iT}) \varphi_{it}(x_l) W_{jt} \Big| > \frac{M}{4} a_{n,T} \Big).
\end{aligned}
$$

We now bound the probability $\mathbb{P}(|\sum_{j \neq i} \sum_{t=1}^{T} I(\mathcal{C}_{iT}) \varphi_{it}(x) W_{jt}| > M a_{n,T}/4)$ for an arbitrary point x with the help of an exponential inequality. To do so, we rewrite the expression $\sum_{j \neq i} \sum_{t=1}^{T} I(\mathcal{C}_{iT}) \varphi_{it}(x) W_{jt}$. In particular, we split up the inner sum over t

into blocks as follows:

$$\sum_{t=1}^{T} I(\mathcal{C}_{iT})\varphi_{it}(x)W_{jt} = \sum_{s=1}^{q_{n,T}} B_{j,2s-1}(x) + \sum_{s=1}^{q_{n,T}} B_{j,2s}(x)$$

with

$$B_{j,s}(x) = \sum_{t=(s-1)r_{n,T}+1}^{sr_{n,T}} I(\mathcal{C}_{iT})\varphi_{it}(x)W_{jt},$$

where as in Lemma B4, $2q_{n,T}$ is the number of blocks and $r_{n,T} = T/2q_{n,T}$ is the block length. We thus get

$$\mathbb{P}\Big(\Big|\sum_{j\neq i}\sum_{t=1}^{T} I(\mathcal{C}_{iT})\varphi_{it}(x)W_{jt}\Big| > \frac{M}{4}a_{n,T}\Big) \leq \mathbb{P}\Big(\Big|\sum_{j\neq i}\sum_{s=1}^{q_{n,T}} B_{j,2s-1}(x)\Big| > \frac{M}{8}a_{n,T}\Big) + \mathbb{P}\Big(\Big|\sum_{j\neq i}\sum_{s=1}^{q_{n,T}} B_{j,2s}(x)\Big| > \frac{M}{8}a_{n,T}\Big).$$

In what follows, we restrict attention to the first term on the right-hand side. The second one can be analyzed by similar arguments.

To indicate the dependence of the block $B_{j,s}(x)$ on the i-th time series $\{X_{it}\}_{t=1}^{T}$, we use the notation $B_{j,s}(x) = B_{j,s}(x, \{X_{it}\}_{t=1}^{T})$. Moreover, we employ the shorthand $\overline{B}_{j,s}(x) = B_{j,s}(x, \{x_{it}\}_{t=1}^{T})$ to denote the s-th block for a fixed realization $\{x_{it}\}_{t=1}^{T}$ of $\{X_{it}\}_{t=1}^{T}$. With this notation at hand, we write

$$\begin{aligned}\mathbb{P}\Big(\Big|\sum_{j\neq i}\sum_{s=1}^{q_{n,T}} B_{j,2s-1}(x)\Big| > \frac{M}{8}a_{n,T}\Big) \\ = \mathbb{E}\Big[\mathbb{P}\Big(\Big|\sum_{j\neq i}\sum_{s=1}^{q_{n,T}} B_{j,2s-1}(x)\Big| > \frac{M}{8}a_{n,T}\Big|\{X_{it}\}_{t=1}^{T}\Big)\Big]\end{aligned}$$

and bound the term

$$\begin{aligned}\mathbb{P}\Big(\Big|\sum_{j\neq i}\sum_{s=1}^{q_{n,T}} B_{j,2s-1}(x)\Big| > \frac{M}{8}a_{n,T}\Big|\{X_{it}\}_{t=1}^{T} = \{x_{it}\}_{t=1}^{T}\Big) \\ = \mathbb{P}\Big(\Big|\sum_{j\neq i}\sum_{s=1}^{q_{n,T}} \overline{B}_{j,2s-1}(x)\Big| > \frac{M}{8}a_{n,T}\Big)\end{aligned}$$

for an arbitrary but fixed realization $\{x_{it}\}_{t=1}^{T}$. By Bradley's lemma, we can construct a sequence of random variables $\overline{B}^{*}_{j,1}(x), \overline{B}^{*}_{j,3}(x), \ldots$ such that (i) $\overline{B}^{*}_{j,1}(x)$, $\overline{B}^{*}_{j,3}(x)$, ... are independent, (ii) $\overline{B}^{*}_{j,2s-1}(x)$ has the same distribution as $\overline{B}_{j,2s-1}(x)$, and (iii) for $0 < \mu \leq$

$\|\overline{B}_{j,2s-1}(x)\|_\infty$,

$$\mathbb{P}\big(|\overline{B}^*_{j,2s-1}(x) - \overline{B}_{j,2s-1}(x)| > \mu\big) \le 18\Big(\frac{\|\overline{B}_{j,2s-1}(x)\|_\infty}{\mu}\Big)^{1/2}\alpha(r_{n,T}). \tag{S2}$$

This allows us to write

$$\mathbb{P}\Big(\Big|\sum_{j\neq i}\sum_{s=1}^{q_{n,T}}\overline{B}_{j,2s-1}(x)\Big| > \frac{M}{8}a_{n,T}\Big) \le P_1 + P_2$$

with

$$P_1 = \mathbb{P}\Big(\Big|\sum_{j\neq i}\sum_{s=1}^{q_{n,T}}\overline{B}^*_{j,2s-1}(x)\Big| > \frac{M}{16}a_{n,T}\Big)$$
$$P_2 = \mathbb{P}\Big(\Big|\sum_{j\neq i}\sum_{s=1}^{q_{n,T}}\big(\overline{B}_{j,2s-1}(x) - \overline{B}^*_{j,2s-1}(x)\big)\Big| > \frac{M}{16}a_{n,T}\Big).$$

First consider P_1. It holds that

$$|\overline{B}_{j,2s-1}(x)| \le \frac{C\tau_{n,T}r_{n,T}b_{n,T}}{nTh} \le \frac{C\tau_{n,T}r_{n,T}(b_{n,T}/\sqrt{h})}{nTh} \le \frac{C\tau_{n,T}r_{n,T}}{nTh(nT)^{\eta/2}} =: \overline{C}_{n,T}.$$

Choosing $\lambda_{n,T} = \overline{C}_{n,T}^{-1}/2$ and applying Markov's inequality, the same arguments as in Lemma B3 yield that

$$P_1 \le 2\exp\Big(-\frac{M}{16}\lambda_{n,T}a_{n,T} + \lambda_{n,T}^2\sum_{j\neq i}\sum_{s=1}^{q_{n,T}}\mathbb{E}[\overline{B}^*_{j,2s-1}(x)^2]\Big).$$

Noting that

$$\begin{aligned}
\sum_{j\neq i}\sum_{s=1}^{q_{n,T}}\mathbb{E}[\overline{B}^*_{j,2s-1}(x)^2] &= \sum_{j\neq i}\sum_{s=1}^{q_{n,T}}\mathbb{E}[\overline{B}_{j,2s-1}(x)^2] \\
&= \sum_{j\neq i}\sum_{s=1}^{q_{n,T}}\mathbb{E}[B_{j,2s-1}(x)^2|\{X_{it}\}_{t=1}^T = \{x_{it}\}_{t=1}^T] \\
&\le \sum_{j\neq i}\sum_{s,t=1}^{T} I(\mathcal{C}_{iT})|\varphi_{is}(x)\varphi_{it}(x)|\big|\mathbb{E}[W_{js}W_{jt}]\big| \\
&\le Cb_{n,T}^2\sum_{j\neq i}\sum_{s,t=1}^{T}K_h(x_{is}-x)K_h(x_{it}-x)\big|\mathbb{E}[W_{js}W_{jt}]\big| \\
&\le \frac{Cb_{n,T}^2}{h^2}\sum_{j\neq i}\Big(\sum_{t=1}^{T}|\mathbb{E}[W_{jt}^2]| + 2\sum_{l=1}^{T-1}\sum_{t=1}^{T-l}|\mathbb{E}[W_{jt}W_{jt+l}]|\Big)
\end{aligned}$$

$$\leq \frac{C}{nTh(nT)^\eta},$$

we arrive at

$$P_1 \leq C \exp\Big(-\frac{M}{16}\lambda_{n,T} a_{n,T} + C\frac{\lambda_{n,T}^2}{nTh(nT)^\eta}\Big).$$

Moreover, choosing

$$r_{n,T} = \sqrt{\frac{nTh}{\tau_{n,T}^2 \log nT}},$$

we obtain that $\frac{\lambda_{n,T}^2}{nTh(nT)^\eta} = \log(nT)$ and $\lambda_{n,T} a_{n,T} = \log(nT)$. As a result,

$$P_1 \leq C \exp\Big(\Big[C - \frac{M}{16}\Big] \log nT\Big) \leq C(nT)^{-p},$$

where p can be made arbitrarily large by choosing M large enough. We next turn to P_2. Using (S2), we obtain that

$$\begin{aligned} P_2 &\leq \sum_{j \neq i} \sum_{s=1}^{q_{n,T}} \mathbb{P}\Big(\big|\overline{B}_{j,2s-1}(x) - \overline{B}^*_{j,2s-1}(x)\big| > \frac{M a_{n,T}}{16 n q_{n,T}}\Big) \\ &\leq C \sum_{j \neq i} \sum_{s=1}^{q_{n,T}} \Big(\frac{\overline{C}_{n,T}}{a_{n,T}/n q_{n,T}}\Big)^{1/2} \alpha(r_{n,T}) \leq C(nT)^{-q}, \end{aligned}$$

where q can be chosen arbitrarily large as the α-coefficients decay exponentially fast.

Putting everything together, we arrive at

$$\begin{aligned} \mathbb{P}\Big(\max_{1 \leq l \leq L_{n,T}} \big|Q^{\leq}(x_l)\big| > \frac{M}{4} a_{n,T}\Big) &\leq \sum_{i=1}^{n} \sum_{l=1}^{L_{n,T}} \mathbb{P}\Big(\Big|\sum_{j \neq i} \sum_{t=1}^{T} I(\mathcal{C}_{iT}) \varphi_{it}(x_l) W_{jt}\Big| > \frac{M}{4} a_{n,T}\Big) \\ &\leq C n L_{n,T}\big[(nT)^{-p} + (nT)^{-q}\big]. \end{aligned}$$

If we choose the exponents p and q sufficiently large, then the right-hand side converges to zero at an arbitrarily fast polynomial rate. This completes the proof. □