C Supplementary Material

In order to complete the proof of Theorem 4.3, we still need to show that equations (45) – (47) are fulfilled for the terms (D_c), $(D_{V,j})$ and $(D_{B,j})$ and that (A) given in (44) is asymptotically negligible. In what follows, we establish these results in a series of lemmas.

Lemma C.1. *It holds that*

$$(D_{V,j}) = \frac{1}{\sqrt{T}} \sum_{s=1}^{T} g_{j,D}\Big(\frac{s}{T}, X_s\Big) u_s + o_p(1)$$

with

$$g_{j,D}\Big(\frac{s}{T}, X_s\Big) = g_{j,D}^{NW}(X_s^j) + g_{j,D}^{SBF}\Big(\frac{s}{T}, X_s\Big)$$

for $j = 0, \ldots, d$. The functions $g_{j,D}^{NW}$ and $g_{j,D}^{SBF}$ are absolutely bounded. Their exact form is given in the proof (see (54) and (57) – (59)).

Proof. We start by giving a detailed exposition of the proof for $j \neq 0$. By Theorem A.1, the stochastic part $\tilde{m}_j^A$ of the smooth backfitting estimate $\tilde{m}_j$ has the expansion

$$\tilde{m}_j^A(x_j) = \hat{m}_j^A(x_j) + \frac{1}{T} \sum_{s=1}^{T} r_{j,s}(x_j) u_s + o_p\Big(\frac{1}{\sqrt{T}}\Big)$$

uniformly in x_j, where $\hat{m}_j^A$ is the stochastic part of the Nadaraya-Watson pilot estimate and the function $r_{j,s}(\cdot) = r_j(\frac{s}{T}, X_s, \cdot)$ is Lipschitz continuous and absolutely bounded.
With this result, we can decompose $(D_{V,j})$ as follows:

$$\begin{aligned}
(D_{V,j}) &= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\partial v_t^2}{\partial \phi_i} \frac{1}{\sigma_t^2 \sigma_t^2} \sum_{k=1}^{t-1} a b^{k-1} \varepsilon_{t-k}^2 \tilde{m}_j^A(X_{t-k}^j) \\
&= \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \sum_{k=1}^{t-1} a b^{k-1} \varepsilon_{t-k}^2 \frac{\partial v_t^2}{\partial \phi_i} \frac{1}{\sigma_t^2 \sigma_t^2} \hat{m}_j^A(X_{t-k}^j) \\
&\quad + \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \sum_{k=1}^{t-1} a b^{k-1} \varepsilon_{t-k}^2 \frac{\partial v_t^2}{\partial \phi_i} \frac{1}{\sigma_t^2 \sigma_t^2} \Big[\frac{1}{T} \sum_{s=1}^{T} r_{j,s}(X_{t-k}^j) u_s\Big] + o_p(1) \\
&=: (D_{V,j}^{NW}) + (D_{V,j}^{SBF}) + o_p(1).
\end{aligned}$$

In the following, we will give the exact arguments needed to treat $(D_{V,j}^{NW})$. The line of argument for $(D_{V,j}^{SBF})$ is essentially identical although some of the steps are easier due to the properties of the $r_{j,s}$ functions.

W.l.o.g. set $\phi_i = a$ and let $m_{i,k} = \max\{k+1, i+1\}$. Using $\partial v_t^2 / \partial a = \sum_{i=1}^{t-1} b^{i-1} \varepsilon_{t-i}^2$ and $\hat{m}_j^A(x_j) = \frac{1}{T} \sum_{s=1}^{T} K_h(x_j, X_s^j) u_s / \frac{1}{T} \sum_{v=1}^{T} K_h(x_j, X_v^j)$, we get

$$(D_{V,j}^{NW}) = \sum_{k=1}^{T-1} a b^{k-1} \sum_{i=1}^{T-1} b^{i-1} \Big[\frac{1}{\sqrt{T}} \sum_{s=1}^{T} \frac{1}{T} \sum_{t=m_{i,k}}^{T} \frac{K_h(X_{t-k}^j, X_s^j)}{\frac{1}{T} \sum_{v=1}^{T} K_h(X_{t-k}^j, X_v^j)} \frac{\varepsilon_{t-k}^2 \varepsilon_{t-i}^2}{\sigma_t^2 \sigma_t^2} u_s\Big]. \tag{50}$$

In a first step, we replace the sum $\frac{1}{T}\sum_{v=1}^{T} K_h(X_{t-k}^j, X_v^j)$ in (50) by a term which only depends on X_{t-k}^j and show that the resulting error is asymptotically negligible. Let $q_j(x_j) = \int_0^1 K_h(x_j, w)dw\ p_j(x_j)$. Furthermore define

$$\begin{aligned} B_j(x_j) &= \frac{1}{T}\sum_{v=1}^{T} \mathbb{E}[K_h(x_j, X_v^j)] - q_j(x_j) \\ V_j(x_j) &= \frac{1}{T}\sum_{v=1}^{T} \big(K_h(x_j, X_v^j) - \mathbb{E}[K_h(x_j, X_v^j)]\big). \end{aligned}$$

Notice that $\sup_{x_j\in[0,1]} |B_j(x_j)| = O_p(h)$ and $\sup_{x_j\in[0,1]} |V_j(x_j)| = O_p(\sqrt{\log T/Th})$. From the identity $\frac{1}{T}\sum_{v=1}^{T} K_h(x_j, X_v^j) = q_j(x_j)+B_j(x_j)+V_j(x_j)$ and a second order Taylor expansion of $(1+x)^{-1}$ we arrive at

$$\begin{aligned} \frac{1}{\frac{1}{T}\sum_{v=1}^{T} K_h(x_j, X_v^j)} &= \frac{1}{q_j(x_j)}\Big(1 + \frac{B_j(x_j)+V_j(x_j)}{q_j(x_j)}\Big)^{-1} \\ &= \frac{1}{q_j(x_j)}\Big(1 - \frac{B_j(x_j)+V_j(x_j)}{q_j(x_j)} + O_p(h^2)\Big) \end{aligned} \tag{51}$$

uniformly in x_j. Plugging this decomposition into (50), we obtain

$$\begin{aligned} (D_{V,j}^{NW}) = &\sum_{k=1}^{T-1} ab^{k-1}\sum_{i=1}^{T-1} b^{i-1}\Big[\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\frac{1}{T}\sum_{t=m_{i,k}}^{T}\frac{K_h(X_{t-k}^j, X_s^j)}{q_j(X_{t-k}^j)}\frac{1}{\sigma_t^2\sigma_t^2}\varepsilon_{t-k}^2\varepsilon_{t-i}^2 u_s\Big] \\ &- (D_{V,j}^{NW,B}) - (D_{V,j}^{NW,V}) + o_p(1) \end{aligned}$$

with

$$\begin{aligned} (D_{V,j}^{NW,B}) &= \sum_{k=1}^{T-1} ab^{k-1}\sum_{i=1}^{T-1} b^{i-1}\Big[\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\frac{1}{T}\sum_{t=m_{i,k}}^{T} K_h(X_{t-k}^j, X_s^j)\frac{B_j(X_{t-k}^j)}{q_j^2(X_{t-k}^j)}\frac{1}{\sigma_t^2\sigma_t^2}\varepsilon_{t-k}^2\varepsilon_{t-i}^2 u_s\Big] \\ (D_{V,j}^{NW,V}) &= \sum_{k=1}^{T-1} ab^{k-1}\sum_{i=1}^{T-1} b^{i-1}\Big[\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\frac{1}{T}\sum_{t=m_{i,k}}^{T} K_h(X_{t-k}^j, X_s^j)\frac{V_j(X_{t-k}^j)}{q_j^2(X_{t-k}^j)}\frac{1}{\sigma_t^2\sigma_t^2}\varepsilon_{t-k}^2\varepsilon_{t-i}^2 u_s\Big]. \end{aligned}$$

As $\sup_{x_j\in I_h} |B_j(x_j)| = O_p(h^2)$ and $\sup_{x_j\in I_h^c} |B_j(x_j)| = O_p(h)$, we can proceed similarly to the proof of Lemma C.3 later on to show that $(D_{V,j}^{NW,B}) = o_p(1)$. Next we will show that $(D_{V,j}^{NW,V}) = o_p(1)$. Let $\mathbb{E}_v[\cdot]$ denote the expectation with respect to the variables indexed by

v, then

$$
\begin{aligned}
\big|(D_{V,j}^{NW,V})\big| &= \Big|\sum_{k=1}^{T-1} ab^{k-1}\sum_{i=1}^{T-1} b^{i-1}\Big[\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\frac{1}{T}\sum_{t=m_{i,k}}^{T}\frac{K_h(X_{t-k}^j,X_s^j)}{q_j^2(X_{t-k}^j)}\frac{1}{\sigma_t^2\sigma_t^2}\varepsilon_{t-k}^2\varepsilon_{t-i}^2 \\
&\qquad\qquad \times\Big(\frac{1}{T}\sum_{v=1}^{T}(K_h(X_{t-k}^j,X_v^j)-\mathbb{E}_v[K_h(X_{t-k}^j,X_v^j)])\Big)u_s\Big]\Big| \\
&\le \sum_{k=1}^{T-1} ab^{k-1}\sum_{i=1}^{T-1} b^{i-1}\Big(\frac{1}{\sqrt{T}}\sum_{t=m_{i,k}}^{T}\Big|\frac{1}{q_j^2(X_{t-k}^j)}\frac{1}{\sigma_t^2\sigma_t^2}\varepsilon_{t-k}^2\varepsilon_{t-i}^2\Big| \\
&\qquad\qquad \times\sup_{x_j\in[0,1]}\Big|\frac{1}{T}\sum_{v=1}^{T}(K_h(x_j,X_v^j)-\mathbb{E}_v[K_h(x_j,X_v^j)])\Big| \\
&\qquad\qquad \times\sup_{x_j\in[0,1]}\Big|\frac{1}{T}\sum_{s=1}^{T}K_h(x_j,X_s^j)u_s\Big|\Big) \\
&= O_p\Big(\frac{\log T}{Th}\Big)\sum_{k=1}^{T-1} ab^{k-1}\sum_{i=1}^{T-1} b^{i-1}\underbrace{\Big(\frac{1}{\sqrt{T}}\sum_{t=m_{i,k}}^{T}\Big|\frac{1}{q_j^2(X_{t-k}^j)}\frac{1}{\sigma_t^2\sigma_t^2}\varepsilon_{t-k}^2\varepsilon_{t-i}^2\Big|\Big)}_{=O_p(\sqrt{T})\text{ by Markov's inequality}} \\
&= O_p\Big(\frac{\log T}{Th}\sqrt{T}\Big) = o_p(1).
\end{aligned}
$$

Together with the fact that $(D_{V,j}^{NW,B}) = o_p(1)$, this yields

$$
(D_{V,j}^{NW}) = \sum_{k=1}^{T-1} ab^{k-1}\sum_{i=1}^{T-1} b^{i-1}\Big[\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\frac{1}{T}\sum_{t=m_{i,k}}^{T}K_h(X_{t-k}^j,X_s^j)\mu_t^{i,k}u_s\Big] + o_p(1), \tag{52}
$$

where we use the shorthand $\mu_t^{i,k} = (q_j(X_{t-k}^j)\sigma_t^2\sigma_t^2)^{-1}\varepsilon_{t-k}^2\varepsilon_{t-i}^2$.
In the next step, we replace the inner sum over t in (52) by a term that only depends on X_s^j and show that the resulting error can be asymptotically neglected. Define

$$
\xi(X_{t-k}^j,X_s^j) := \xi_t^{i,k}(X_{t-k}^j,X_s^j) := K_h(X_{t-k}^j,X_s^j)\mu_t^{i,k} - \mathbb{E}_{-s}[K_h(X_{t-k}^j,X_s^j)\mu_t^{i,k}],
$$

where $\mathbb{E}_{-s}[\cdot]$ is the expectation with respect to all variables except for those depending on the index s. With the above notation at hand, we can write

$$
\begin{aligned}
(D_{V,j}^{NW}) &= \sum_{k=1}^{T-1} ab^{k-1}\sum_{i=1}^{T-1} b^{i-1}\Big[\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\frac{1}{T}\sum_{t=m_{i,k}}^{T}\mathbb{E}_{-s}[K_h(X_{t-k}^j,X_s^j)\mu_t^{i,k}]u_s\Big] \\
&\quad + (R_{V,j}^{NW}) + o_p(1),
\end{aligned}
$$

where

$$
\begin{aligned}
(R_{V,j}^{NW}) &= \sum_{k=1}^{T-1} ab^{k-1}\sum_{i=1}^{T-1} b^{i-1}\Big[\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\frac{1}{T}\sum_{t=m_{i,k}}^{T}\xi(X_{t-k}^j,X_s^j)u_s\Big] \qquad (53)\\
&= \sum_{k=1}^{C_2\log T} ab^{k-1}\sum_{i=1}^{C_2\log T} b^{i-1}\Big[\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\frac{1}{T}\sum_{t=m_{i,k}}^{T}\xi(X_{t-k}^j,X_s^j)u_s\Big] + o_p(1)
\end{aligned}
$$

for some sufficiently large constant $C_2 > 0$. Once we show that $(R^{NW}_{V,j}) = o_p(1)$, we are left with

$$\begin{aligned}
(D^{NW}_{V,j}) &= \sum_{k=1}^{T-1} ab^{k-1} \sum_{i=1}^{T-1} b^{i-1} \Big[\frac{1}{\sqrt{T}} \sum_{s=1}^{T} \frac{1}{T} \sum_{t=m_{i,k}}^{T} \mathbb{E}_{-s}[K_h(X^j_{t-k}, X^j_s)\mu^{i,k}_t] u_s \Big] + o_p(1) \\
&= \frac{1}{\sqrt{T}} \sum_{s=1}^{T} \Big(\sum_{k=1}^{T-1} ab^{k-1} \sum_{i=1}^{T-1} b^{i-1} \frac{T - m_{i,k}}{T} \mathbb{E}_{-s}[K_h(X^j_{-k}, X^j_s)\mu^{i,k}_0] \Big) u_s + o_p(1).
\end{aligned}$$

As the terms with $i, k \geq C_2 \log T$ are asymptotically negligible, we can expand the i and k sums to infinity, which yields

$$\begin{aligned}
(D^{NW}_{V,j}) &= \frac{1}{\sqrt{T}} \sum_{s=1}^{T} \Big(\sum_{k=1}^{\infty} ab^{k-1} \sum_{i=1}^{\infty} b^{i-1} \mathbb{E}_{-s}[K_h(X^j_{-k}, X^j_s)\mu^{i,k}_0] \Big) u_s + o_p(1) \qquad (54) \\
&=: \frac{1}{\sqrt{T}} \sum_{s=1}^{T} g^{NW}_{j,D}(X^j_s) u_s + o_p(1)
\end{aligned}$$

with

$$\begin{aligned}
\mu^{i,k}_0 &= \frac{1}{q_j(X^j_{-k})} \frac{1}{\sigma^2_0 \sigma^2_0} \varepsilon^2_{-k} \varepsilon^2_{-i} \\
q_j(X^j_{-k}) &= \int_0^1 K_h(X^j_{-k}, w) dw \; p_j(X^j_{-k}).
\end{aligned}$$

Thus it remains to show that $(R^{NW}_{V,j}) = o_p(1)$, which requires a lot of care. We will prove that the term in square brackets in (53) is $o_p(1)$ uniformly over $i, k \leq C_2 \log T$, which yields the desired result. It is easily seen that

$$\begin{aligned}
P &:= \mathbb{P}\Big(\max_{i,k \leq C_2 \log T} \Big| \frac{1}{\sqrt{T}} \sum_{s=1}^{T} \frac{1}{T} \sum_{t=m_{i,k}}^{T} \xi(X^j_{t-k}, X^j_s) u_s \Big| > \delta \Big) \\
&\leq \sum_{k=1}^{C_2 \log T} \sum_{i=1}^{C_2 \log T} \underbrace{\mathbb{P}\Big(\Big| \frac{1}{\sqrt{T}} \sum_{s=1}^{T} \frac{1}{T} \sum_{t=m_{i,k}}^{T} \xi(X^j_{t-k}, X^j_s) u_s \Big| > \delta \Big)}_{=: P_{i,k}}
\end{aligned}$$

for a fixed $\delta > 0$. Then by Chebychev's inequality

$$\begin{aligned}
P_{i,k} &\leq \frac{1}{T^3 \delta^2} \sum_{s,s'=1}^{T} \sum_{t,t'=m_{i,k}}^{T} \mathbb{E}\Big[\xi(X^j_{t-k}, X^j_s) u_s \xi(X^j_{t'-k}, X^j_{s'}) u_{s'} \Big] \\
&= \frac{1}{T^3 \delta^2} \sum_{(s,s',t,t') \notin \Gamma_{i,k}} \mathbb{E}\Big[\xi(X^j_{t-k}, X^j_s) u_s \xi(X^j_{t'-k}, X^j_{s'}) u_{s'} \Big] \\
&\quad + \frac{1}{T^3 \delta^2} \sum_{(s,s',t,t') \in \Gamma_{i,k}} \mathbb{E}\Big[\xi(X^j_{t-k}, X^j_s) u_s \xi(X^j_{t'-k}, X^j_{s'}) u_{s'} \Big] =: P^1_{i,k} + P^2_{i,k},
\end{aligned}$$

where $\Gamma_{i,k}$ is the set of tuples (s, s', t, t') with $1 \leq s, s' \leq T$ and $m_{i,k} \leq t, t' \leq T$ such that one index is separated from the others. We say that an index, for instance t, is separated from the others if $\min\{|t-t'|,|t-s|,|t-s'|\} > C_3 \log T$, i.e. if it is further away from the other indices than $C_3 \log T$ for a constant C_3 to be chosen later on. We now analyse $P^1_{i,k}$ and $P^2_{i,k}$ separately.

(a) First consider $P^1_{i,k}$. If a tuple (s,s',t,t') is not an element of $\Gamma_{i,k}$, then no index can be separated from the others. Since the index t cannot be separated, there exists an index, say t', such that $|t-t'| \le C_3 \log T$. Now take an index different from t and t', for instance s. Then by the same argument, there exists an index, say s', such that $|s-s'| \le C_3 \log T$. As a consequence, the number of tuples $(s,s',t,t') \notin \Gamma_{i,k}$ is smaller than $CT^2(\log T)^2$ for some constant C. Using (A11), this suffices to infer that

$$\big|P^1_{i,k}\big| \le \frac{1}{T^3\delta^2} \sum_{(s,s',t,t')\notin\Gamma_{i,k}} \frac{C}{h^2} \le \frac{C}{\delta^2}\frac{(\log T)^2}{Th^2}.$$

Hence, $|P^1_{i,k}| \le C\delta^{-2}(\log T)^{-3}$ uniformly in i and k.

(b) The term $P^2_{i,k}$ is more difficult to handle. We start by taking a cover $\{I_m\}_{m=1}^{M_T}$ of the compact support $[0,1]$ of X^j_{t-k}. The elements I_m are intervals of length $1/M_T$ given by $I_m = [\frac{m-1}{M_T}, \frac{m}{M_T})$ for $m = 1,\dots,M_T-1$ and $I_{M_T} = [1-\frac{1}{M_T}, 1]$. The midpoint of the interval I_m is denoted by x_m. With this, we can write

$$K_h(X^j_{t-k}, X^j_s) = \sum_{m=1}^{M_T} I(X^j_{t-k} \in I_m) \times \big[K_h(x_m, X^j_s) + (K_h(X^j_{t-k}, X^j_s) - K_h(x_m, X^j_s))\big]. \tag{55}$$

Using (55), we can further write

$$\begin{aligned}\xi(X^j_{t-k}, X^j_s) &= \sum_{m=1}^{M_T} \Big\{ I(X^j_{t-k} \in I_m) K_h(x_m, X^j_s)\mu^{i,k}_t \\ &\qquad - \mathbb{E}_{-s}[I(X^j_{t-k} \in I_m) K_h(x_m, X^j_s)\mu^{i,k}_t] \Big\} \\ &\quad + \sum_{m=1}^{M_T} \Big\{ I(X^j_{t-k} \in I_m)(K_h(X^j_{t-k}, X^j_s) - K_h(x_m, X^j_s))\mu^{i,k}_t \\ &\qquad - \mathbb{E}_{-s}[I(X^j_{t-k} \in I_m)(K_h(X^j_{t-k}, X^j_s) - K_h(x_m, X^j_s))\mu^{i,k}_t] \Big\} \\ &=: \xi_1(X^j_{t-k}, X^j_s) + \xi_2(X^j_{t-k}, X^j_s)\end{aligned}$$

and

$$\begin{aligned}P^2_{i,k} &= \frac{1}{T^3\delta^2} \sum_{(s,s',t,t')\in\Gamma_{i,k}} \mathbb{E}\big[\xi_1(X^j_{t-k}, X^j_s) u_s \xi(X^j_{t'-k}, X^j_{s'}) u_{s'}\big] \\ &\quad + \frac{1}{T^3\delta^2} \sum_{(s,s',t,t')\in\Gamma_{i,k}} \mathbb{E}\big[\xi_2(X^j_{t-k}, X^j_s) u_s \xi(X^j_{t'-k}, X^j_{s'}) u_{s'}\big] =: P^{2,1}_{i,k} + P^{2,2}_{i,k}.\end{aligned}$$

We first consider $P^{2,2}_{i,k}$. Set $M_T = CT(\log T)^3 h^{-3}$ and exploit the Lipschitz continuity of the kernel K to get that $|K_h(X^j_{t-k}, X^j_s) - K_h(x_m, X^j_s)| \le \frac{C}{h^2}|X^j_{t-k} - x_m|$. This gives us

$$\big|\xi_2(X^j_{t-k}, X^j_s)\big| \le \frac{C}{h^2} \sum_{m=1}^{M_T} \Big(\underbrace{I(X^j_{t-k} \in I_m)|X^j_{t-k} - x_m|}_{\le I(X^j_{t-k} \in I_m) M_T^{-1}} \mu^{i,k}_t + \mathbb{E}\big[\underbrace{I(X^j_{t-k} \in I_m)|X^j_{t-k} - x_m|}_{\le I(X^j_{t-k} \in I_m) M_T^{-1}} \mu^{i,k}_t\big]\Big) \le \frac{C}{M_T h^2}\big(\mu^{i,k}_t + \mathbb{E}[\mu^{i,k}_t]\big). \tag{56}$$

Plugging (56) into the expression for $P^{2,2}_{i,k}$, we arrive at

$$\begin{aligned}
\big|P^{2,2}_{i,k}\big| &\le \frac{1}{T^3\delta^2} \sum_{(s,s',t,t') \in \Gamma_{i,k}} \mathbb{E}\Big[\big|\xi_2(X^j_{t-k}, X^j_s)\big|\big|u_s \xi(X^j_{t'-k}, X^j_{s'}) u_{s'}\big|\Big] \\
&\le \frac{1}{T^3\delta^2} \frac{C}{M_T h^2} \sum_{(s,s',t,t') \in \Gamma_{i,k}} \underbrace{\mathbb{E}\big[(\mu^{i,k}_t + \mathbb{E}[\mu^{i,k}_t])|u_s \xi(X^j_{t'-k}, X^j_{s'}) u_{s'}|\big]}_{\le C h^{-1}} \le \frac{C}{\delta^2} \frac{1}{(\log T)^3}.
\end{aligned}$$

We next turn to $P^{2,1}_{i,k}$. Write

$$P^{2,1}_{i,k} = \frac{1}{T^3\delta^2} \sum_{(s,s',t,t') \in \Gamma_{i,k}} \Big(\sum_{m=1}^{M_T} S_m \Big)$$

with

$$\begin{aligned}
S_m = \mathbb{E}\Big[&\big\{I(X^j_{t-k} \in I_m) K_h(x_m, X^j_s)\mu^{i,k}_t - \mathbb{E}_{-s}[I(X^j_{t-k} \in I_m) K_h(x_m, X^j_s)\mu^{i,k}_t]\big\} \\
&\times u_s \xi(X^j_{t'-k}, X^j_{s'}) u_{s'}\Big]
\end{aligned}$$

and assume that an index, w.l.o.g. t, can be separated from the others. Choosing $C_3 \gg C_2$, we get

$$\begin{aligned}
S_m &= \mathrm{Cov}\Big(I(X^j_{t-k} \in I_m)\mu^{i,k}_t - \mathbb{E}[I(X^j_{t-k} \in I_m)\mu^{i,k}_t], K_h(x_m, X^j_s) u_s \xi(X^j_{t'-k}, X^j_{s'}) u_{s'}\Big) \\
&\le \frac{C}{h^2}(\alpha([C_3 - C_2]\log T))^{1-\frac{2}{p}} \le \frac{C}{h^2}(a^{(C_3 - C_2)\log T})^{1-\frac{2}{p}} \le \frac{C}{h^2} T^{-C_4}
\end{aligned}$$

with some $C_4 > 0$ by Davydov's inequality, where p is chosen slightly larger than 2. Note that the above bound is independent of i and k and that we can make C_4 arbitrarily large by choosing C_3 large enough. This shows that $|P^{2,1}_{i,k}| \le C\delta^{-2}(\log T)^{-3}$ uniformly in i and k with some constant C.

Combining (a) and (b) yields that $P \to 0$ for each fixed $\delta > 0$. This implies that

$$(R^{NW,V}_{V,j}) = o_p(1),$$

which completes the proof for the term $(D^{NW}_{V,j})$.

As stated at the beginning of the proof, the term $(D^{SBF}_{V,j})$ can be treated in exactly the same way. Following analogous arguments as above and writing $\zeta^{i,k}_t = (\sigma^2_t\sigma^2_t)^{-1}\varepsilon^2_{t-k}\varepsilon^2_{t-i}$, one obtains

$$(D^{SBF}_{V,j}) = \sum_{k=1}^{T-1} ab^{k-1}\sum_{i=1}^{T-1} b^{i-1}\Big[\frac{1}{\sqrt{T}}\sum_{s=1}^{T}\frac{1}{T}\sum_{t=m_{i,k}}^{T}\mathbb{E}_{-s}[r_{j,s}(X^j_{t-k})\zeta^{i,k}_t]\ u_s\Big] + o_p(1) \tag{57}$$

$$= \frac{1}{\sqrt{T}}\sum_{s=1}^{T}\Big(\sum_{k=1}^{\infty} ab^{k-1}\sum_{i=1}^{\infty} b^{i-1}\mathbb{E}_{-s}[r_{j,s}(X^j_{-k})\zeta^{i,k}_0]\Big)u_s + o_p(1)$$

$$=: \frac{1}{\sqrt{T}}\sum_{s=1}^{T} g^{SBF}_{j,D}\Big(\frac{s}{T}, X_s\Big)u_s + o_p(1).$$

Finally, the proofs for $j = 0$ are very similar but somewhat simpler and are thus omitted here. For completeness we provide the functions $g^{NW}_{0,D}$ and $g^{SBF}_{0,D}$:

$$g^{NW}_{0,D}\Big(\frac{s}{T}\Big) = \Big(\sum_{k=1}^{\infty} ab^{k-1}\sum_{i=1}^{\infty} b^{i-1}\mathbb{E}\big[\frac{1}{\sigma^2_0\sigma^2_0}\varepsilon^2_{-k}\varepsilon^2_{-i}\big]\Big)\int_0^1 \frac{K_h(\frac{s}{T}, v)}{\int_0^1 K_h(v, w)dw}dv \tag{58}$$

$$g^{SBF}_{0,D}\Big(\frac{s}{T}, X_s\Big) = \Big(\sum_{k=1}^{\infty} ab^{k-1}\sum_{i=1}^{\infty} b^{i-1}\mathbb{E}\big[\frac{1}{\sigma^2_0\sigma^2_0}\varepsilon^2_{-k}\varepsilon^2_{-i}\big]\Big)\int_0^1 r_{0,s}(w)dw. \tag{59}$$

□

Lemma C.2. *It holds that*

$$(D_c) = \frac{1}{\sqrt{T}}\sum_{s=1}^{T} g_{c,D}u_s$$

with

$$g_{c,D} = \sum_{k=1}^{\infty} ab^{k-1}\sum_{i=1}^{\infty} b^{i-1}\mathbb{E}\Big[\frac{1}{\sigma^2_0\sigma^2_0}\varepsilon^2_{-i}\varepsilon^2_{-k}\Big].$$

Proof. Using the fact that

$$\tilde{m}_c = \frac{1}{T}\sum_{s=1}^{T} Z_{s,T} = m_c + \frac{1}{T}\sum_{s=1}^{T} m_0\Big(\frac{s}{T}\Big) + \sum_{j=1}^{d}\frac{1}{T}\sum_{s=1}^{T} m_j(X^j_s) + \frac{1}{T}\sum_{s=1}^{T} u_s,$$

we arrive at

$$(D_c) = -\Big(\frac{1}{T}\sum_{t=1}^{T} G_t\sum_{k=1}^{t-1} ab^{k-1}\varepsilon^2_{t-k}\Big)\Big(\frac{1}{\sqrt{T}}\sum_{s=1}^{T} u_s\Big)$$

with $G_t = \frac{\partial v^2_t}{\partial \phi_i}(\sigma^2_t\sigma^2_t)^{-1}$. Now let $m_{i,k} = \max\{k+1, i+1\}$ and assume w.l.o.g. that $\phi_i = a$. Then

$$\frac{1}{T}\sum_{t=1}^{T} G_t\sum_{k=1}^{t-1} ab^{k-1}\varepsilon^2_{t-k} = \frac{1}{T}\sum_{t=1}^{T}\Big(\sum_{i=1}^{t-1} b^{i-1}\varepsilon^2_{t-i}\Big)\frac{1}{\sigma^2_t\sigma^2_t}\sum_{k=1}^{t-1} ab^{k-1}\varepsilon^2_{t-k}$$

$$= \sum_{k=1}^{C_2\log T} ab^{k-1}\sum_{i=1}^{C_2\log T} b^{i-1}\frac{1}{T}\sum_{t=m_{i,k}}^{T}\frac{1}{\sigma^2_t\sigma^2_t}\varepsilon^2_{t-i}\varepsilon^2_{t-k} + o_p(1)$$

with some sufficiently large constant C_2. Using Chebychev's inequality and exploiting the mixing properties of the variables involved, one can show that

$$\max_{i,k \leq C_2 \log T} \frac{1}{T} \sum_{t=m_{i,k}}^{T} \Big(\frac{1}{\sigma_t^2 \sigma_t^2} \varepsilon_{t-i}^2 \varepsilon_{t-k}^2 - \mathbb{E}\Big[\frac{1}{\sigma_t^2 \sigma_t^2} \varepsilon_{t-i}^2 \varepsilon_{t-k}^2 \Big] \Big) = o_p(1).$$

This allows us to infer that

$$\begin{aligned}
\frac{1}{T} \sum_{t=1}^{T} G_t \sum_{k=1}^{t-1} ab^{k-1} \varepsilon_{t-k}^2 &= \sum_{k=1}^{C_2 \log T} ab^{k-1} \sum_{i=1}^{C_2 \log T} b^{i-1} \frac{1}{T} \sum_{t=m_{i,k}}^{T} \mathbb{E}\Big[\frac{1}{\sigma_t^2 \sigma_t^2} \varepsilon_{t-i}^2 \varepsilon_{t-k}^2 \Big] + o_p(1) \\
&= \sum_{k=1}^{\infty} ab^{k-1} \sum_{i=1}^{\infty} b^{i-1} \mathbb{E}\Big[\frac{1}{\sigma_0^2 \sigma_0^2} \varepsilon_{-i}^2 \varepsilon_{-k}^2 \Big] + o_p(1),
\end{aligned}$$

which completes the proof. □

Lemma C.3. *It holds that*

$$(D_{B,j}) = o_p(1)$$

for $j = 0, \ldots, d$.

Proof. We start by considering the case $j = 0$: Define

$$\begin{aligned}
J_h &= \{t \in \{1, \ldots, T\} : C_1 h \leq \frac{t}{T} \leq 1 - C_1 h\} \\
J_{h,c}^u &= \{t \in \{1, \ldots, T\} : 1 - C_1 h < \frac{t}{T}\} \\
J_{h,c}^l &= \{t \in \{1, \ldots, T\} : \frac{t}{T} < C_1 h\},
\end{aligned}$$

where $[-C_1, C_1]$ is the support of K. Using the uniform convergence rates from Theorem A.2 and assuming w.l.o.g. that $\phi_i = a$, we get

$$\begin{aligned}
|(D_{B,0})| &= \Big| \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \frac{\partial v_t^2}{\partial a} \frac{1}{\sigma_t^2 \sigma_t^2} \sum_{k=1}^{t-1} ab^{k-1} \varepsilon_{t-k}^2 \Big[m_0\Big(\frac{t-k}{T}\Big) - \tilde{m}_0^B\Big(\frac{t-k}{T}\Big) - \frac{1}{T} \sum_{s=1}^{T} m_0\Big(\frac{s}{T}\Big) \Big] \Big| \\
&\leq O_p(h) \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \sum_{i=1}^{t-1} b^{i-1} \sum_{k=1}^{t-1} ab^{k-1} \varepsilon_{t-i}^2 \varepsilon_{t-k}^2 I(t-k \in J_{h,c}^l) \\
&\quad + O_p(h) \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \sum_{i=1}^{t-1} b^{i-1} \sum_{k=1}^{t-1} ab^{k-1} \varepsilon_{t-i}^2 \varepsilon_{t-k}^2 I(t-k \in J_{h,c}^u) \\
&\quad + O_p(h^2) \frac{C}{\sqrt{T}} \sum_{t=1}^{T} \sum_{i=1}^{t-1} b^{i-1} \sum_{k=1}^{t-1} ab^{k-1} \varepsilon_{t-i}^2 \varepsilon_{t-k}^2 I(t-k \in J_h) \\
&=: (D_{B,0}^{J_{h,c}^l}) + (D_{B,0}^{J_{h,c}^u}) + (D_{B,0}^{J_h}).
\end{aligned}$$

By Markov's inequality, $(D_{B,0}^{J_h}) = O_p(h^2 \sqrt{T}) = o_p(1)$. Recognizing that

(i) $I(t-k \in J_{h,c}^u) \leq I(t \in J_{h,c}^u)$ for all $k \in \{0, \ldots, t-1\}$

(ii) $\sum_{t=1}^{T} I(t \in J_{h,c}^u) \leq C_1 T h$,

we get $(D_{B,0}^{J_{h,c}^u}) = O_p(h^2\sqrt{T}) = o_p(1)$ by another appeal to Markov's inequality. This just leaves $(D_{B,0}^{J_{h,c}^l})$, which is a bit more tedious. By a change of variable $j = t - k$,

$$
\begin{aligned}
(D_{B,0}^{J_{h,c}^l}) &\leq O_p(h)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{i=1}^{t-1} b^{i-1}\varepsilon_{t-i}^2 \sum_{j=1}^{t-1} ab^{t-j-1}\varepsilon_j^2 I(j \in J_{h,c}^l) \\
&= O_p(h)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{i=1}^{t-1} b^{i-1}\varepsilon_{t-i}^2 I\Big(\Big[\frac{t}{2}\Big] \in J_{h,c}^l\Big) \sum_{j=1}^{t-1} ab^{t-j-1}\varepsilon_j^2 I(j \in J_{h,c}^l) \\
&\quad + O_p(h)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{i=1}^{t-1} b^{i-1}\varepsilon_{t-i}^2 I\Big(\Big[\frac{t}{2}\Big] \notin J_{h,c}^l\Big) \sum_{j=1}^{t-1} ab^{t-j-1}\varepsilon_j^2 I(j \in J_{h,c}^l) \\
&=: (A) + (B),
\end{aligned}
$$

where $[x]$ denotes the smallest integer larger than x. Realizing that $[t/2] \in J_{h,c}^l$ only if $t < 2C_1 hT$, we get $(A) = O_p(h^2\sqrt{T}) = o_p(1)$ once again by Markov's inequality. In (B) we can truncate the summation over j at $[t/2]-1$, as $I(j \in J_{h,c}^l) = 0$ for $j \geq [t/2]$ if $[t/2] \notin J_{h,c}^l$. We thus obtain

$$
\begin{aligned}
(B) &\leq O_p(h)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{i=1}^{t-1} b^{i-1}\varepsilon_{t-i}^2 \sum_{j=1}^{[t/2]-1} ab^{t-j-1}\varepsilon_j^2 \\
&= O_p(h)\frac{1}{\sqrt{T}}\sum_{t=1}^{T} b^{[t/2]}\sum_{i=1}^{t-1} b^{i-1} \sum_{j=1}^{[t/2]-1} ab^{t-j-1-[t/2]}\varepsilon_{t-i}^2\varepsilon_j^2.
\end{aligned}
$$

By a final appeal to Markov's inequality we arrive at

$$
(B) = O_p(h)O_p\Big(\frac{1}{\sqrt{T}}\Big) = o_p(1),
$$

thus completing the proof for $j = 0$.

Next consider the case $j \neq 0$. Similarly to before, we have

$$
\begin{aligned}
|(D_{B,j})| &\leq O_p(h^2)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{i=1}^{t-1} b^{i-1}\sum_{k=1}^{t-1} ab^{k-1}\varepsilon_{t-i}^2\varepsilon_{t-k}^2 I(X_{t-k}^j \in I_h) \\
&\quad + O_p(h)\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{i=1}^{t-1} b^{i-1}\sum_{k=1}^{t-1} ab^{k-1}\varepsilon_{t-i}^2\varepsilon_{t-k}^2 I(X_{t-k}^j \notin I_h) \\
&= O_p(h^2\sqrt{T}) + O_p\Big(\frac{h}{\sqrt{T}}\Big)\underbrace{\sum_{t=1}^{T}\sum_{i=1}^{t-1} b^{i-1}\sum_{k=1}^{t-1} ab^{k-1}\varepsilon_{t-i}^2\varepsilon_{t-k}^2 I(X_{t-k}^j \notin I_h)}_{=:R_T}
\end{aligned}
$$

with $I_h = [2C_1h, 1-2C_1h]$ as defined in Theorem 4.1. Using (A11), it is easy to see that $R_T = O_p(h)$, which yields the result for $j \neq 0$. □

Lemma C.4. *It holds that*

$$(A) = -\frac{1}{\sqrt{T}} \sum_{t=1}^{T} \underbrace{\left(1 - \frac{\varepsilon_t^2}{\sigma_t^2}\right)}_{=(1-\eta_t^2)} \frac{1}{\sigma_t^2} \left(\frac{\partial \tilde{v}_t^2}{\partial \phi_i} - \frac{\partial v_t^2}{\partial \phi_i}\right) + o_p(1) = o_p(1).$$

Proof. W.l.o.g. let $\phi_i = a$. With the help of (G1) and a simple Taylor expansion, we get that

$$\begin{aligned}
\frac{\partial \tilde{v}_t^2}{\partial \phi_i} - \frac{\partial v_t^2}{\partial \phi_i} &= \sum_{k=1}^{t-1} b^{k-1} \left(\tilde{\varepsilon}_{t-k}^2 - \varepsilon_{t-k}^2\right) \\
&= \sum_{k=1}^{t-1} b^{k-1} \varepsilon_{t-k}^2 \left[\frac{\tau^2\left(\frac{t-k}{T}, X_{t-k}\right) - \tilde{\tau}^2\left(\frac{t-k}{T}, X_{t-k}\right)}{\tau^2\left(\frac{t-k}{T}, X_{t-k}\right)} + R_\varepsilon\left(\frac{t-k}{T}, X_{t-k}\right)\right] \\
&= \sum_{k=1}^{t-1} b^{k-1} \varepsilon_{t-k}^2 \left[\frac{\exp(\xi_{t-k})\left(m\left(\frac{t-k}{T}, X_{t-k}\right) - \tilde{m}\left(\frac{t-k}{T}, X_{t-k}\right)\right)}{\exp\left(m\left(\frac{t-k}{T}, X_{t-k}\right)\right)}\right] + O_p(h^2) \\
&= \sum_{k=1}^{t-1} b^{k-1} \varepsilon_{t-k}^2 \left[m\left(\frac{t-k}{T}, X_{t-k}\right) - \tilde{m}\left(\frac{t-k}{T}, X_{t-k}\right)\right] + O_p(h^2) \\
&= \sum_{k=1}^{t-1} b^{k-1} \varepsilon_{t-k}^2 \Bigg\{(m_c - \tilde{m}_c) - \tilde{m}_0^A\left(\frac{t-k}{T}\right) - \ldots - \tilde{m}_d^A\left(X_{t-k}^d\right) \\
&\quad + \left(m_0\left(\frac{t-k}{T}\right) - \tilde{m}_0^B\left(\frac{t-k}{T}\right)\right) + \ldots + \left(m_d\left(X_{t-k}^d\right) - \tilde{m}_d^B\left(X_{t-k}^d\right)\right)\Bigg\} \\
&\quad + O_p(h^2),
\end{aligned}$$

where ξ_{t-k} is an intermediate point between $m(\frac{t-k}{T}, X_{t-k})$ and $\tilde{m}(\frac{t-k}{T}, X_{t-k})$. Using this together with arguments similar to those for Lemma C.3 yields that

$$\begin{aligned}
(A) &= -\sum_{k=1}^{T-1} b^{k-1} \Bigg(\frac{1}{\sqrt{T}} \sum_{t=k+1}^{T} \left(1 - \eta_t^2\right) \frac{\varepsilon_{t-k}^2}{\sigma_t^2} \\
&\quad \times \left\{(m_c - \tilde{m}_c) - \tilde{m}_0^A\left(\frac{t-k}{T}\right) - \ldots - \tilde{m}_d^A\left(X_{t-k}^d\right)\right\}\Bigg) + o_p(1) \\
&=: (A_c) - (A_0) - (A_1) - \ldots - (A_d) + o_p(1).
\end{aligned}$$

It is straightforward to see that $(A_c) = o_p(1)$. In what follows, we further prove that $(A_j) = o_p(1)$ for $j = 0, \ldots, d$ as well, which completes the proof.
Consider a fixed $j \in \{0, \ldots, d\}$ and let $\delta > 0$ be an arbitrarily small but fixed constant. Write

$$(A_j) = \sum_{k=1}^{T-1} b^{k-1} \left(\frac{1}{\sqrt{T}} \sum_{t=k+1}^{T} \left(1 - \eta_t^2\right) \frac{\varepsilon_{t-k}^2}{\sigma_t^2} \tilde{m}_j^A(X_{t-k}^j)\right) =: (A_j^{\leq}) + (A_j^{>}),$$

where

$$(A_j^{\le}) = \sum_{k=1}^{T-1} b^{k-1} \left(\frac{1}{\sqrt{T}} \sum_{t=k+1}^{T} W_t^{\le} \frac{\varepsilon_{t-k}^2}{\sigma_t^2} \tilde{m}_j^A(X_{t-k}^j) \right)$$
$$(A_j^{>}) = \sum_{k=1}^{T-1} b^{k-1} \left(\frac{1}{\sqrt{T}} \sum_{t=k+1}^{T} W_t^{>} \frac{\varepsilon_{t-k}^2}{\sigma_t^2} \tilde{m}_j^A(X_{t-k}^j) \right)$$

with

$$W_t^{\le} = \left(1-\eta_t^2\right) I(|\eta_t| \le T^{1/48+\delta}) - \mathbb{E}[(1-\eta_t^2) I(|\eta_t| \le T^{1/48+\delta})]$$
$$W_t^{>} = \left(1-\eta_t^2\right) I(|\eta_t| > T^{1/48+\delta}) - \mathbb{E}[(1-\eta_t^2) I(|\eta_t| > T^{1/48+\delta})].$$

We now consider the two terms $(A_j^{\le})$ and $(A_j^{>})$ separately. We start with $(A_j^{>})$. Standard arguments for kernel estimators show that $\sup_{x_j \in [0,1]} \left|\hat{m}_j^A(x_j)\right| = O_p(\sqrt{\log T/Th})$. This together with Theorem A.1 implies that $\sup_{x_j \in [0,1]} \left|\tilde{m}_j^A(x_j)\right| = O_p(\sqrt{\log T/Th})$ as well. As $\sqrt{\log T/Th} \le T^{-3/8+\delta}$, we can infer that

$$\begin{aligned} \left|(A_j^{>})\right| &\le O_p\left(\sqrt{\frac{\log T}{Th}}\right) \cdot \sum_{k=1}^{T-1} b^{k-1} \frac{1}{\sqrt{T}} \sum_{t=k+1}^{T} |W_t^{>}| \frac{\varepsilon_{t-k}^2}{\sigma_t^2} \\ &\le O_p(1) \underbrace{\sum_{k=1}^{T-1} b^{k-1} \frac{1}{T^{7/8-\delta}} \sum_{t=k+1}^{T} |W_t^{>}| \frac{\varepsilon_{t-k}^2}{\sigma_t^2}}_{:=(*)} . \end{aligned}$$

Moreover, since

$$\mathbb{E}\left[\left|1-\eta_t^2\right| I(|\eta_t| > T^{1/48+\delta})\right] \le \mathbb{E}\left[\left|1-\eta_t^2\right| \frac{\eta_t^6}{T^{6(1/48+\delta)}} I(|\eta_t| > T^{1/48+\delta})\right] \le \frac{C}{T^{1/8+6\delta}},$$

we get that $\mathbb{E}|W_t^{>}| \le C/T^{1/8+6\delta}$. From this and Markov's inequality, it follows that $(*) = o_p(1)$ and thus $(A_j^{>}) = o_p(1)$.
We next turn to the term $(A_j^{\le})$. Splitting $(A_j^{\le})$ into two parts with the help of the indicators $I(\varepsilon_{t-k}^2 \le T^{1/48+\delta})$ and $I(\varepsilon_{t-k}^2 > T^{1/48+\delta})$ and applying a similar truncation argument as above, we can show that

$$(A_j^{\le}) = \sum_{k=1}^{T-1} b^{k-1} \Big(\frac{1}{\sqrt{T}} \sum_{t=k+1}^{T} W_t^{\le} \frac{\varepsilon_{t-k}^2}{\sigma_t^2} I(|\varepsilon_{t-k}| \le T^{1/48+\delta}) \tilde{m}_j^A\big(X_{t-k}^j\big) \Big) + o_p(1).$$

Since the weights b^{k-1} decay exponentially fast to zero, we further obtain that

$$(A_j^{\le}) = \sum_{k=1}^{C_2 \log T} b^{k-1} \Big(\frac{1}{\sqrt{T}} \sum_{t=k+1}^{T} W_t^{\le} \frac{\varepsilon_{t-k}^2}{\sigma_t^2} I(|\varepsilon_{t-k}| \le T^{1/48+\delta}) \tilde{m}_j^A\big(X_{t-k}^j\big) \Big) + o_p(1)$$

with some sufficiently large constant C_2. By Theorem A.1, it holds that uniformly in x_j,

$$\tilde{m}_j^A(x_j) = \frac{1}{T} \sum_{s=1}^{T} \left(\frac{K_h(x_j, X_s^j)}{\frac{1}{T}\sum_{v=1}^{T} K_h(x_j, X_v^j)} + r_{j,s}(x_j) \right) u_s + o_p\left(\frac{1}{\sqrt{T}}\right).$$

By the same arguments as used in the proof of Lemma C.1, we can replace the term $\frac{1}{T}\sum_{v=1}^{T} K_h(x_j, X_v^j)$ by $q_j(x_j) = \int_0^1 K_h(x_j, w)dw\, p_j(x_j)$, which yields that

$$(A_j^{\leq}) = \sum_{k=1}^{C_2 \log T} b^{k-1}\Big(\frac{1}{\sqrt{T}} \sum_{t=k+1}^{T} W_t^{\leq} \frac{\varepsilon_{t-k}^2}{\sigma_t^2} I(|\varepsilon_{t-k}| \leq T^{1/48+\delta})\, \check{m}_j^A\big(X_{t-k}^j\big)\Big) + o_p(1)$$

with

$$\check{m}_j^A(x_j) = \frac{1}{T}\sum_{s=1}^{T}\left(\frac{K_h\left(x_j, X_s^j\right)}{q_j(x_j)} + r_{j,s}(x_j)\right) u_s.$$

We can thus write $(A_j^{\leq}) = \sum_{k=1}^{C_2 \log T} b^{k-1} \cdot (A_{j,k}^{\leq}) + o_p(1)$ with

$$(A_{j,k}^{\leq}) = \frac{1}{\sqrt{T}} \sum_{t=k+1}^{T} W_t^{\leq} \frac{\varepsilon_{t-k}^2}{\sigma_t^2} I(|\varepsilon_{t-k}| \leq T^{1/48+\delta})\, \check{m}_j^A(X_{t-k}^j).$$

In what follows, we prove that for any fixed $\varepsilon > 0$,

$$\max_{1\leq k\leq C_2 \log T} \mathbb{P}\left(\left|(A_{j,k}^{\leq})\right| > \varepsilon\right) \leq T^{-\kappa} \tag{60}$$

with some $\kappa > 0$. This implies that $\mathbb{P}(\max_{1\leq k\leq C_2 \log T} |(A_{j,k}^{\leq})| > \varepsilon) \leq \sum_{k=1}^{C_2 \log T} \mathbb{P}(|(A_{j,k}^{\leq})| > \varepsilon) = o(1)$, that is, $\max_{1\leq k\leq C_2 \log T} |(A_{j,k}^{\leq})| = o_p(1)$. Since $(A_j^{\leq}) = \sum_{k=1}^{C_2 \log T} b^{k-1} \cdot (A_{j,k}^{\leq}) + o_p(1) \leq C \max_{1\leq k\leq C_2 \log T} |(A_{j,k}^{\leq})| + o_p(1)$, we can conclude that $(A_j^{\leq}) = o_p(1)$.
It remains to prove (60). To do so, we embed the stochastic function $\check{m}_j^A$ into a class of Hölder functions: For any $\eta > 0$ and $x_j \neq x_j'$,

$$\begin{aligned}
&\left|\check{m}_j^A(x_j) - \check{m}_j^A(x_j')\right| \Big/ \left|x_j - x_j'\right|^{1/2+\eta} \\
&\leq \left|\frac{1}{T}\sum_{s=1}^{T} \frac{1}{q_j(x_j)} \left(K_h\left(x_j, X_s^j\right) - K_h\left(x_j', X_s^j\right)\right) u_s\right| \Big/ \left|x_j - x_j'\right|^{1/2+\eta} \\
&\quad + \left|\frac{1}{T}\sum_{s=1}^{T} K_h\left(x_j', X_s^j\right) \frac{q_j(x_j') - q_j(x_j)}{q_j(x_j')q_j(x_j)} u_s\right| \Big/ \left|x_j - x_j'\right|^{1/2+\eta} \\
&\quad + \left|\frac{1}{T}\sum_{s=1}^{T} \left(r_{j,s}(x_j) - r_{j,s}(x_j')\right) u_s\right| \Big/ \left|x_j - x_j'\right|^{1/2+\eta} \\
&=: \beta_1(x_j, x_j') + \beta_2(x_j, x_j') + \beta_3(x_j, x_j').
\end{aligned}$$

By standard arguments to derive uniform convergence rates for kernel estimators which can be found for example in Bosq (1998), Masry (1996) or Hansen (2008), we can show that

$$\mathbb{P}\left(\sup_{x_j, x_j' \in [0,1], x_j \neq x_j'} \left|\beta_k(x_j, x_j')\right| > \frac{M a_T}{6}\right) = O(T^{-\kappa})$$

for all $k = 1, 2, 3$ and some $\kappa > 0$, where $a_T = \sqrt{\log T / T h^{2+\varsigma}}$ for some small $\varsigma > 0$ and M is a sufficiently large constant. From this, it immediately follows that

$$\mathbb{P}\left(\sup_{x_j, x_j' \in [0,1], x_j \neq x_j'} \frac{\left|\check{m}_j^A(x_j) - \check{m}_j^A(x_j')\right|}{\left|x_j - x_j'\right|^{1/2+\eta}} > \frac{M a_T}{2}\right) = O(T^{-\kappa}). \tag{61}$$

Similarly, it can be verified that

$$\mathbb{P}\left(\sup_{x_j\in[0,1]}\left|\check{m}_j^A(x_j)\right| > \frac{Ma_T}{2}\right) = O(T^{-\kappa}). \tag{62}$$

From (61) and (62), we can conclude that with probability $1-O(T^{-\kappa})$, the random function $\frac{1}{Ma_T}\check{m}_j^A$ is contained in the Hölder space $\mathcal{F} := C_1^{1/2+\eta}([0,1])$ which is defined as follows: For any $\alpha \in (0,1]$,

$$C_1^\alpha([0,1]) = \{f : [0,1] \to \mathbb{R} : f \text{ is continuous with } \|f\|_\alpha \le 1\}$$

with

$$\|f\|_\alpha = \sup_{x\in(0,1)}|f(x)| + \sup_{x,y\in(0,1),x\neq y}\frac{|f(x)-f(y)|}{|x-y|^\alpha}.$$

Let $\mathcal{N}(\delta, C_1^\alpha([0,1]), \|\cdot\|_\infty)$ be the δ-covering number of $C_1^\alpha([0,1])$ endowed with the supremum norm $\|\cdot\|_\infty$. By Theorem 2.7.1 in van der Vaart and Wellner (1996), we have the bound

$$\log\mathcal{N}\left(\delta, C_1^\alpha([0,1]), \|\cdot\|_\infty\right) \le K\delta^{-1/\alpha} \tag{63}$$

for any $\delta > 0$ with some fixed constant $K > 0$. We next define

$$Z_{T,k}(f) := \frac{Ma_T}{\sqrt{T}}\sum_{t=k+1}^T W_t^{\le}\frac{\varepsilon_{t-k}^2}{\sigma_t^2} I(|\varepsilon_{t-k}| \le T^{1/48+\delta})\, f\big(X_{t-k}^j\big)$$

and note that $(A_{j,k}^{\le}) = Z_{T,k}(\frac{1}{Ma_T}\check{m}_j^A)$. Since $\frac{1}{Ma_T}\check{m}_j^A$ is contained in the Hölder space $\mathcal{F} = C_1^{1/2+\eta}([0,1])$ with probability $1-O(T^{-\kappa})$, it follows that

$$\mathbb{P}\left(\left|(A_{j,k}^{\le})\right| > \varepsilon\right) \le P\left(\sup_{f\in\mathcal{F}}|Z_{T,k}(f)| > \varepsilon\right) + O(T^{-\kappa})$$

and it remains to show that

$$\mathbb{P}\left(\sup_{f\in\mathcal{F}}|Z_{T,k}(f)| > \varepsilon\right) \le CT^{-\kappa}. \tag{64}$$

To do so, define $Z_{T,k}^\gamma := T^\gamma Z_{T,k}$ with $\gamma > 0$ small and write

$$\begin{aligned}&\mathbb{P}\Big(\left|Z_{T,k}^\gamma(f) - Z_{T,k}^\gamma(g)\right| > \varepsilon\,||f-g||_\infty\Big)\\ &= \mathbb{P}\left(T^\gamma\left|\frac{Ma_T}{\sqrt{T}}\sum_{t=k+1}^T \underbrace{W_t^{\le}\frac{\varepsilon_{t-k}^2}{\sigma_t^2} I(|\varepsilon_{t-k}| \le T^{1/48+\delta})\left(f\big(X_{t-k}^j\big) - g\big(X_{t-k}^j\big)\right)}_{=:\psi_{t,j,k}}\right| > \varepsilon\,||f-g||_\infty\right).\end{aligned}$$

Using the trivial bound $|\psi_{t,j,k}| \le CT^{1/12+4\delta}||f-g||_\infty$ and noting that $\{\psi_{t,j,k} : t \in \mathbb{Z}\}$ is a martingale difference sequence for any $k \ge 1$, we can show that the process $Z_{T,k}^\gamma =$

$(Z^\gamma_{T,k}(f))_{f\in\mathcal{F}}$ has subgaussian increments. More specifically, we can apply an exponential inequality for martingale differences such as theorem 15.20 in Davidson (1994) to obtain that

$$\begin{aligned}
&\mathbb{P}\left(\left|Z^\gamma_{T,k}(f) - Z^\gamma_{T,k}(g)\right| > \varepsilon \, ||f-g||_\infty\right) \\
&\leq 2\exp\left(-\frac{\varepsilon^2}{2\sum_{t=k+1}^{T}\left(\frac{T^\gamma M a_T}{\sqrt{T}} C T^{1/12+4\delta}\right)^2}\right) \\
&\leq 2\exp\left(-\frac{\varepsilon^2}{2(CM)^2\left(T^\gamma a_T\right)^2 T^{1/6+8\delta}}\right) \leq 2\exp\left(-\frac{\varepsilon^2}{2}\right)
\end{aligned}$$

for T large enough. Next, let $\|\cdot\|_{\psi_0}$ denote the Orlicz norm corresponding to $\psi_0(x) = \exp(x^2) - 1$. Applying a maximal inequality such as theorem 2.2.4 in van der Vaart and Wellner (1996) along with the metric entropy bound (63), we obtain that

$$\begin{aligned}
\left\|\sup_{f\in\mathcal{F}}|Z^\gamma_{T,k}(f)|\right\|_{\psi_0} &\leq \int_0^C \sqrt{K\varepsilon^{-\frac{1}{1/2+\eta}}}\,d\varepsilon = \sqrt{K}\int_0^C \varepsilon^{-\frac{1}{1+2\eta}}\,d\varepsilon \\
&= \sqrt{K}\frac{1}{1-\frac{1}{1+2\eta}}\varepsilon^{1-\frac{1}{1+2\eta}}\bigg|_0^C \leq r_0 < \infty
\end{aligned}$$

with some sufficiently large C. Hence, by Markov's inequality,

$$\begin{aligned}
\mathbb{P}\left(\sup_{f\in\mathcal{F}}|Z_{T,k}(f)| > \varepsilon\right) &= \mathbb{P}\left(T^{-\gamma}\sup_{f\in\mathcal{F}}|Z^\gamma_{T,k}(f)| > \varepsilon\right) \\
&\leq \frac{\mathbb{E}\left[\psi_0\left(\sup_{f\in\mathcal{F}}|Z^\gamma_{T,k}(f)|/r_0\right)\right]}{\psi_0(\varepsilon T^\gamma/r_0)} \leq \frac{1}{\exp(\varepsilon^2 T^{2\gamma}/r_0^2) - 1},
\end{aligned}$$

which completes the proof of (64). □