7 Additional Proofs for: Crump, Hotz, Imbens and Mitnik, "Nonparametric Tests for Treatment Effect Heterogeneity"

Proof of Lemma A.1: We will generalize the proof in Imbens, Newey and Ridder (2006). For (i) we will show

$$\mathbb{E}\left[\left\|\hat{\Omega}_{w,K} - \Omega_{w,K}\right\|^2\right] \le C \cdot \zeta(K)^2 K/N$$

so that the result follows by Markov's inequality.

$$\begin{aligned}
& \mathbb{E}\left[\left\|\hat{\Omega}_{w,K} - \Omega_{w,K}\right\|^2\right] \\
= \ & \mathbb{E}\left[\left\|\left(R'_{w,K}R_{w,K}/N_w\right) - \Omega_{w,K}\right\|^2\right] \\
= \ & \mathbb{E}\left[\operatorname{tr}\left(\left(R'_{w,K}R_{w,K}/N_w\right) - \Omega_{w,K}\right)'\left(\left(R'_{w,K}R_{w,K}/N_w\right) - \Omega_{w,K}\right)\right] \\
= \ & \mathbb{E}\left[\operatorname{tr}\left(R'_{w,K}R_{w,K}R'_{w,K}R_{w,K}/N_w^2 - \Omega_{w,K}\left(R'_{w,K}R_{w,K}/N_w\right) - \left(R'_{w,K}R_{w,K}/N_w\right)\Omega_{w,K} + \Omega^2_{w,K}\right)\right] \\
= \ & \operatorname{tr}\left(\mathbb{E}\left[R'_{w,K}R_{w,K}R'_{w,K}R_{w,K}/N_w^2\right] - \Omega_{w,K}\mathbb{E}\left[R'_{w,K}R_{w,K}/N_w\right] - \mathbb{E}\left[R'_{w,K}R_{w,K}/N_w\right]\Omega_{w,K} + \Omega^2_{w,K}\right) \\
= \ & \operatorname{tr}\left(\mathbb{E}\left[R'_{w,K}R_{w,K}R'_{w,K}R_{w,K}/N_w^2\right] - 2\cdot\Omega^2_{w,K} + \Omega^2_{w,K}\right) \\
= \ & \operatorname{tr}\left(\mathbb{E}\left[R'_{w,K}R_{w,K}R'_{w,K}R_{w,K}/N_w^2\right]\right) - \operatorname{tr}\left(\Omega^2_{w,K}\right)
\end{aligned}$$

The second term is

$$\operatorname{tr}(\Omega^2_{w,K}) = \sum_{k=1}^{K}\sum_{l=1}^{K}\left(\mathbb{E}\left[R_{kK}(X)R_{lK}(X)|W=w\right]\right)^2 \tag{B.1}$$

The first term is

$$\begin{aligned}
& \operatorname{tr}\left(\mathbb{E}\left[R'_{w,K}R_{w,K}R'_{w,K}R_{w,K}\right]/N_w^2\right) \\
= \ & \mathbb{E}\left[\sum_{k=1}^{K}\sum_{l=1}^{K}\left(\sum_{i|W_i=w}^{N} R_{kK}(X_i)R_{lK}(X_i)\right)^2\right]/N_w^2 \\
= \ & \sum_{k=1}^{K}\sum_{l=1}^{K}\sum_{i|W_i=w}^{N}\sum_{j|W_j=w}^{N}\mathbb{E}\left[R_{kK}(X_i)R_{lK}(X_i)R_{lK}(X_j)R_{kK}(X_j)|W=w\right]/N_w^2
\end{aligned}$$

We can then partition this expression into terms with $i = j$,

$$\sum_{k=1}^{K}\sum_{l=1}^{K}\sum_{i|W_i=w}^{N}\mathbb{E}\left[R_{kK}(X_i)^2R_{lK}(X_i)^2|W=w\right]/N_w^2 \tag{B.2}$$

and with terms $i \neq j$,

$$N_w(N_w-1)\sum_{k=1}^{K}\sum_{l=1}^{K}\left(\mathbb{E}\left[R_{kK}(X)R_{lK}(X)|W=w\right]\right)^2/N_w^2 \tag{B.3}$$

Combining equations $(B.1)$, $(B.2)$ and $(B.3)$ yields,

$$\begin{aligned}
\mathbb{E}\left[\left\|\hat{\Omega}_{w,K}-\Omega_{w,K}\right\|^2\right] &= \sum_{k=1}^{K}\sum_{l=1}^{K}\sum_{i|W_i=w}^{N}\mathbb{E}\left[R_{kK}(X_i)^2R_{lK}(X_i)^2|W=w\right]/N_w^2 \\
&\quad +N_w(N_w-1)\sum_{k=1}^{K}\sum_{l=1}^{K}\left(\mathbb{E}\left[R_{kK}(X)R_{lK}(X)|W=w\right]\right)^2/N_w^2 \\
&\quad -\sum_{k=1}^{K}\sum_{l=1}^{K}\left(\mathbb{E}\left[R_{kK}(X)R_{lK}(X)|W=w\right]\right)^2 \qquad \text{(B.4)} \\
&= \sum_{k=1}^{K}\sum_{l=1}^{K}\sum_{i|W_i=w}^{N}\mathbb{E}\left[R_{kK}(X_i)^2R_{lK}(X_i)^2|W=w\right]/N_w^2 \\
&\quad -\sum_{k=1}^{K}\sum_{l=1}^{K}\left(\mathbb{E}\left[R_{kK}(X)R_{lK}(X)|W=w\right]\right)^2/N_w \qquad \text{(B.5)} \\
&< \sum_{k=1}^{K}\sum_{l=1}^{K}\sum_{i|W_i=w}^{N}\mathbb{E}\left[R_{kK}(X_i)^2R_{lK}(X_i)^2|W=w\right]/N_w^2 \qquad \text{(B.6)} \\
&= \frac{1}{N_w^2}\sum_{i|W_i=w}^{N}\mathbb{E}\left[\sum_{k=1}^{K}R_{kK}(X_i)^2\sum_{l=1}^{K}R_{lK}(X_i)^2|W=w\right] \qquad \text{(B.7)} \\
&\leq \frac{1}{N_w^2}\sum_{i|W_i=w}^{N}\zeta(K)^2\cdot\mathbb{E}\left[\sum_{l=1}^{K}R_{lK}(X_i)^2|W=w\right] \qquad \text{(B.8)} \\
&= \frac{1}{N_w^2}\sum_{i|W_i=w}^{N}\zeta(K)^2\cdot\sum_{l=1}^{K}\mathbb{E}\left[R_{lK}(X_i)^2|W=w\right] \qquad \text{(B.9)} \\
&= \frac{1}{N_w}\zeta(K)^2\cdot\operatorname{tr}\left(\Omega_{w,K}\right) \qquad \text{(B.10)} \\
&\leq \frac{1}{N_w}\zeta(K)^2\cdot K\cdot\lambda_{max}\left(\Omega_{w,K}\right) \qquad \text{(B.11)} \\
&\leq C\cdot K\zeta(K)^2/N \qquad \text{(B.12)}
\end{aligned}$$

where $(B.11)$ follows by

$$\zeta(K)=\sup_x\|R_K(x)\|=\sup_x\left(\sum_{k=1}^{K}R_{kK}^2(x)\right)^{\frac{1}{2}}$$

which then implies that

$$\sum_{k=1}^{K}R_{kK}^2(x)\leq\zeta(K)^2.$$

$(B.12)$ follows since the maximum eigenvalue of $\Omega_{w,K}$ is $O(1)$ (see below).

For (ii), let us first show that for any two positive semi-definite matrices A and B, and conformable vectors a and b, if $A\geq B$ in a positive semi-definite sense, then for

$$\lambda_{min}(A)=\min_{a'a=1}a'Aa=\underline{a}'A\underline{a},\quad \lambda_{min}(B)=\min_{b'b=1}b'Bb=\underline{b}'B\underline{b},$$

and

$$\lambda_{max}(A) = \max_{a'a=1} a'Aa = \bar{a}'A\bar{a}, \quad \lambda_{max}(B) = \max_{b'b=1} b'Bb = \bar{b}'B\bar{b},$$

we have that,

$$\lambda_{min}(A) = \underline{a}'A\underline{a} \geq \underline{a}'B\underline{a} \geq \underline{b}'B\underline{b} = \lambda_{min}(B) \tag{B.13}$$

and

$$\lambda_{max}(A) = \bar{a}'A\bar{a} \geq \bar{b}'A\bar{b} \geq \bar{b}'B\bar{b} = \lambda_{max}(B). \tag{B.14}$$

Now, let $f_w(x) = f_{X|W}(x|W=w)$ and recall that $\Omega_{w,K} = \mathbb{E}\left[R_K(X)R_K(X)'|W=w\right]$ where $\Omega_{1,K}$ is normalized to equal I_K. Next define

$$q(x) = f_0(x)/f_1(x)$$

and note that by Assumptions 2.3 and 3.1 we have that

$$0 < \underline{q} \leq q(x) \leq \bar{q} < \infty.$$

Thus we may define $q(x) \equiv \underline{q} + \tilde{q}(x)$ so that,

$$\begin{aligned}
\Omega_{0,K} &= \mathbb{E}\left[R_K(x)R_K(x)'|W=0\right] \\
&= \int R_K(x)R_K(x)'f_0(x)dx \\
&= \int R_K(x)R_K(x)'q(x)f_1(x)dx \\
&= \int R_K(x)R_K(x)'\left(\underline{q} + \tilde{q}(x)\right)f_1(x)dx \\
&= \underline{q}\int R_K(x)R_K(x)'f_1(x)dx + \int R_K(x)R_K(x)'\tilde{q}(x)f_1(x)dx \\
&= \underline{q}\cdot\Omega_{1,K} + \int R_K(x)R_K(x)'\tilde{q}(x)f_1(x)dx \\
&= \underline{q}\cdot\Omega_{1,K} + \tilde{Q}
\end{aligned}$$

$\tilde{Q}$ is a positive semi-definite matrix, which implies that $\Omega_{0,K} \geq \underline{q}\cdot\Omega_{1,K}$ in a positive semi-definite sense. Thus by $(B.13)$

$$\lambda_{min}\left(\Omega_{0,K}\right) \geq \underline{q}\cdot\lambda_{min}\left(\Omega_{1,K}\right) = \underline{q}$$

and the minimum eigenvalue of $\Omega_{0,K}$ is bounded away from zero. Also, since $\bar{q}\cdot\Omega_{1,K} \geq \tilde{Q}$ in a positive semi-definite sense, using $(B.14)$ we have

$$\begin{aligned}
\lambda_{\max}\left(\Omega_{0,K}\right) &= \max_{d'd=1} d'\left(\underline{q}\cdot\Omega_{1,K} + \tilde{Q}\right)d \\
&\leq \underline{q}\cdot\max_{d_1'd_1=1} d_1'\Omega_{1,K}d_1 + \max_{d_2'd_2=1} d_2'\tilde{Q}d_2 \\
&\leq \underline{q} + \bar{q}\cdot\max_{d_2'd_2=1} d_2'\Omega_{1,K}d_2 \\
&= \underline{q} + \bar{q}
\end{aligned}$$

and the maximum eigenvalue of $\Omega_{0,K}$ is bounded. Both the minimum and maximum eigenvalue of $\Omega_{1,K}$ are bounded away from zero and bounded, respectively, by construction.

For (iii) consider the minimum eigenvalue of $\hat{\Omega}_{w,K}$.

$$\begin{aligned}
\lambda_{min}\left(\hat{\Omega}_{w,K}\right) &= \min_{d'd=1} d'\left(\hat{\Omega}_{w,K}\right)d && \text{(B.15)}\\
&= \min_{d'd=1}\left(d'\left(\Omega_{w,K}\right)d + d'\left(\hat{\Omega}_{w,K}-\Omega_{w,K}\right)d\right) && \text{(B.16)}\\
&\geq \min_{d_1'd_1=1} d_1'\left(\Omega_{w,K}\right)d_1 + \min_{d_2'd_2=1} d_2'\left(\hat{\Omega}_{w,K}-\Omega_{w,K}\right)d_2 && \text{(B.17)}\\
&= \lambda_{min}\left(\Omega_{w,K}\right) + \lambda_{min}\left(\hat{\Omega}_{w,K}-\Omega_{w,K}\right) && \text{(B.18)}\\
&\geq \lambda_{min}\left(\Omega_{w,K}\right) - \left\|\hat{\Omega}_{w,K}-\Omega_{w,K}\right\| && \text{(B.19)}\\
&= \lambda_{min}\left(\Omega_{w,K}\right) - O_p\left(\zeta(K)K^{\frac{1}{2}}N^{-\frac{1}{2}}\right) && \text{(B.20)}
\end{aligned}$$

Where $(B.19)$ follows since for a symmetric matrix A

$$\|A\|^2 = \operatorname{tr}\left(A^2\right) \geq \lambda_{min}(A)^2,$$

and since the norm is nonnegative

$$\|A\| \geq -\lambda_{min}(A)$$

and

$$\|A\| \geq \lambda_{min}(A)$$

for all values of $\lambda_{min}(A)$. Finally, $(B.20)$ follows by part (i).

Next, consider the maximum eigenvalue of $\Omega_{w,K}$.

$$\begin{aligned}
\lambda_{max}\left(\hat{\Omega}_{w,K}\right) &= \max_{d'd=1} d'\left(\hat{\Omega}_{w,K}\right)d && \text{(B.21)}\\
&= \max_{d'd=1}\left(d'\left(\Omega_{w,K}\right)d + d'\left(\hat{\Omega}_{w,K}-\Omega_{w,K}\right)d\right) && \text{(B.22)}\\
&\leq \max_{d_1'd_1=1} d_1'\left(\Omega_{w,K}\right)d_1 + \max_{d_2'd_2=1} d_2'\left(\hat{\Omega}_{w,K}-\Omega_{w,K}\right)d_2 && \text{(B.23)}\\
&= \lambda_{max}\left(\Omega_{w,K}\right) + \lambda_{max}\left(\hat{\Omega}_{w,K}-\Omega_{w,K}\right) && \text{(B.24)}\\
&\leq \lambda_{max}\left(\Omega_{w,K}\right) + \left\|\hat{\Omega}_{w,K}-\Omega_{w,K}\right\| && \text{(B.25)}\\
&= \lambda_{max}\left(\Omega_{w,K}\right) + O_p\left(\zeta(K)K^{\frac{1}{2}}N^{-\frac{1}{2}}\right) && \text{(B.26)}
\end{aligned}$$

Where $(B.25)$ follows by the above discussion and $(B.26)$ follows by part (i). ■

Proof of Lemma A.2: For (i),

$$\begin{aligned}
\lambda_{\min}\left(V\right) &= \min_{d'd=1} d'\left(\frac{\sigma_{0,K}^2}{1-c}\cdot\Omega_{0,K}^{-1} + \frac{\sigma_{1,K}^2}{c}\cdot\Omega_{1,K}^{-1}\right)d\\
&\geq \frac{\sigma_{0,K}^2}{1-c}\cdot\min_{d_1'd_1=1} d_1'\Omega_{0,K}^{-1}d_1 + \frac{\sigma_{1,K}^2}{c}\cdot\min_{d_2'd_2} d_2'\Omega_{1,K}^{-1}d_2\\
&= \frac{\sigma_{0,K}^2}{1-c}\cdot\lambda_{\min}\left(\Omega_{0,K}^{-1}\right) + \frac{\sigma_{1,K}^2}{c}\cdot\lambda_{\min}\left(\Omega_{1,K}^{-1}\right)
\end{aligned}$$

So that $\lambda_{\min}(V)$ is bounded away from zero by Assumption 3.2 and by Lemma $A.1$. Also,

$$\lambda_{\max}(V) = \max_{d'd=1} d'\left(\frac{\sigma_{0,K}^2}{1-c}\cdot\Omega_{0,K}^{-1} + \frac{\sigma_{1,K}^2}{c}\cdot\Omega_{1,K}^{-1}\right)d$$

$$\leq \frac{\sigma_{0,K}^2}{1-c}\cdot\max_{d_1'd_1=1} d_1'\Omega_{0,K}^{-1}d_1 + \frac{\sigma_{1,K}^2}{c}\cdot\max_{d_2'd_2} d_2'\Omega_{1,K}^{-1}d_2$$

$$= \frac{\sigma_{0,K}^2}{1-c}\cdot\lambda_{\max}\left(\Omega_{0,K}^{-1}\right) + \frac{\sigma_{1,K}^2}{c}\cdot\lambda_{\max}\left(\Omega_{1,K}^{-1}\right)$$

So that $\lambda_{\max}(V)$ is bounded by Assumption 3.2 and by Lemma $A.1$.

For (ii),

$$\lambda_{\min}\left(\hat{V}\right) = \min_{d'd=1} d'\left(\frac{\hat{\sigma}_{0,K}^2}{1-\hat{c}}\cdot\hat{\Omega}_{0,K}^{-1} + \frac{\hat{\sigma}_{1,K}^2}{\hat{c}}\cdot\hat{\Omega}_{1,K}^{-1}\right)d$$

$$\geq \frac{\hat{\sigma}_{0,K}^2}{1-\hat{c}}\cdot\min_{d_1'd_1=1} d_1'\hat{\Omega}_{0,K}^{-1}d_1 + \frac{\hat{\sigma}_{1,K}^2}{\hat{c}}\cdot\min_{d_2'd_2} d_2'\hat{\Omega}_{1,K}^{-1}d_2$$

$$= \frac{\sigma_{0,K}^2}{1-c}\cdot\lambda_{\min}\left(\hat{\Omega}_{0,K}^{-1}\right) + \frac{\sigma_{1,K}^2}{c}\cdot\lambda_{\min}\left(\hat{\Omega}_{1,K}^{-1}\right) + o_p(1)$$

$$\geq \frac{\sigma_{0,K}^2}{1-c}\cdot\lambda_{\min}\left(\Omega_{0,K}^{-1}\right) + \frac{\sigma_{1,K}^2}{c}\cdot\lambda_{\min}\left(\Omega_{1,K}^{-1}\right) - O_p\left(\zeta(K)K^{\frac{1}{2}}N^{-\frac{1}{2}}\right)$$

Where the last line follows by $(B.20)$. Thus, $\lambda_{\min}\left(\hat{V}\right)$ is bounded away from zero with probability going to one by part (i) and Assumption 3.3. Finally,

$$\lambda_{\max}\left(\hat{V}\right) = \max_{d'd=1} d'\left(\frac{\hat{\sigma}_{0,K}^2}{1-\hat{c}}\cdot\hat{\Omega}_{0,K}^{-1} + \frac{\hat{\sigma}_{1,K}^2}{\hat{c}}\cdot\hat{\Omega}_{1,K}^{-1}\right)d$$

$$\leq \frac{\hat{\sigma}_{0,K}^2}{1-\hat{c}}\cdot\max_{d_1'd_1=1} d_1'\hat{\Omega}_{0,K}^{-1}d_1 + \frac{\hat{\sigma}_{1,K}^2}{\hat{c}}\cdot\max_{d_2'd_2} d_2'\hat{\Omega}_{1,K}^{-1}d_2$$

$$= \frac{\sigma_{0,K}^2}{1-c}\cdot\lambda_{\max}\left(\hat{\Omega}_{0,K}^{-1}\right) + \frac{\sigma_{1,K}^2}{c}\cdot\lambda_{\max}\left(\hat{\Omega}_{1,K}^{-1}\right) + o_p(1)$$

$$\leq \frac{\sigma_{0,K}^2}{1-c}\cdot\lambda_{\max}\left(\Omega_{0,K}^{-1}\right) + \frac{\sigma_{1,K}^2}{c}\cdot\lambda_{\max}\left(\Omega_{1,K}^{-1}\right) + O_p\left(\zeta(K)K^{\frac{1}{2}}N^{-\frac{1}{2}}\right)$$

Where the last line follows by $(B.26)$. Thus, $\lambda_{\max}\left(\hat{V}\right)$ is bounded with probability going to one by part (i) and Assumption 3.3. ■

Before proving Theorem (3.3) we need the following lemma.

Lemma B.1 *Recall that we partitioned* $\hat{V}$ *as*

$$\hat{V} = \begin{pmatrix} \hat{V}_{00} & \hat{V}_{01} \\ \hat{V}_{10} & \hat{V}_{11} \end{pmatrix}$$

and

$$V = \begin{pmatrix} V_{00} & V_{01} \\ V_{10} & V_{11} \end{pmatrix}$$

where $\hat{V}_{00}$ and V_{00} are scalars, $\hat{V}_{01}$ and V_{01} are $1 \times (K-1)$ vectors, $\hat{V}_{10}$ and V_{10} are $(K-1) \times 1$ vectors and $\hat{V}_{11}$ and V_{11} are $(K-1) \times (K-1)$ matrices. Then,

$$\lambda_{max}\left(\hat{V}^{-1}\right) \geq \lambda_{max}\left(\hat{V}_{11}^{-1}\right) \qquad \lambda_{max}\left(V^{-1}\right) \geq \lambda_{max}\left(V_{11}^{-1}\right)$$

and

$$\lambda_{min}\left(\hat{V}^{-1}\right) \leq \lambda_{min}\left(\hat{V}_{11}^{-1}\right) \qquad \lambda_{min}\left(V^{-1}\right) \leq \lambda_{min}\left(V_{11}^{-1}\right)$$

Proof The proof follows by the interlacing theorem, see Li-Mathias (2002):

If A is an $n \times n$ positive semi-definite Hermitian matrix with eigenvalues $\lambda_1 \geq ... \geq \lambda_n$, B is a $k \times k$ principal submatrix of A with eigenvalues $\tilde{\lambda}_1 \geq ... \geq \tilde{\lambda}_k$, then

$$\lambda_i \geq \tilde{\lambda}_i \geq \lambda_{i+n-k}, \quad i = 1, ..., k.$$

In our case, $\hat{V}$ and V are positive definite, symmetric and thus positive definite, Hermitian. So then, by the interlacing theorem

$$\lambda_{min}\left(\hat{V}\right) \leq \lambda_{min}\left(\hat{V}_{11}\right) \Longrightarrow \lambda_{max}\left(\hat{V}^{-1}\right) \geq \lambda_{max}\left(\hat{V}_{11}^{-1}\right)$$

$$\lambda_{min}\left(V\right) \leq \lambda_{min}\left(V_{11}\right) \Longrightarrow \lambda_{max}\left(V^{-1}\right) \geq \lambda_{max}\left(V_{11}^{-1}\right)$$

$$\lambda_{max}\left(\hat{V}\right) \geq \lambda_{max}\left(\hat{V}_{11}\right) \Longrightarrow \lambda_{min}\left(\hat{V}^{-1}\right) \leq \lambda_{min}\left(\hat{V}_{11}^{-1}\right)$$

$$\lambda_{max}\left(V\right) \geq \lambda_{max}\left(V_{11}\right) \Longrightarrow \lambda_{min}\left(V^{-1}\right) \leq \lambda_{min}\left(V_{11}^{-1}\right)$$

■

Proof of Theorem (3.3)**:** When the conditional average treatment effect is constant we may choose the two approximating sequences, $\gamma_{0,K}^0$ and $\gamma_{1,K}^0$, to differ only by way of the first element (the coefficient of the constant term in the approximating sequence). In other words, if $\mu_1(x) - \mu_0(x) = \tau_0$ for all $x \in \mathbb{X}$, then the coefficients of the power series terms involving x^r such that $r > 0$ should be identical for $w = 0, 1$, so that their difference no longer varies with x.

Thus, a natural strategy to test the null hypothesis of a constant conditional average treatment effect is to compare the last $K-1$ elements of $\hat{\gamma}_{1,K}$ and $\hat{\gamma}_{0,K}$ and to reject the null hypothesis when these elements are sufficiently different.

First, note that by equations $(A.10)$ and $(A.11)$ and the consistency of $\hat{V}$ we have that

$$\hat{V}_{11}^{-1} \cdot \sqrt{N}\left(\hat{\gamma}_{11,K} - \hat{\gamma}_{01,K} - \left(\gamma_{11,K}^* - \gamma_{01,K}^*\right)\right) \xrightarrow{d} \mathcal{N}\left(0, I_{K-1}\right) \tag{B.27}$$

To simplify notation, define

$$\hat{\delta} = \hat{\gamma}_{11,K} - \hat{\gamma}_{01,K}$$

and

$$\delta^* = \gamma_{11,K}^* - \gamma_{01,K}^*.$$

We may again follow the logic of Lemmas $(A.3)$, $(A.4)$, and $(A.5)$ to conclude that

$$T^{*\prime} \equiv \left(N \cdot \left(\hat{\delta} - \delta^*\right)' \cdot \hat{V}_{11}^{-1} \cdot \left(\hat{\delta} - \delta^*\right) - (K-1)\right) / \sqrt{2\left(K-1\right)}$$

converges in distribution to a $\mathcal{N}(0,1)$ random variable. We need only show that $|T^{*\prime} - T'| = o_p(1)$ to complete the proof. First, we will again use results from Lemma $A.6$. Specifically, note that

$$\begin{aligned}
\left\|\gamma^*_{w1,K} - \gamma^0_{w1,K}\right\|^2 &= \sum_{i=2}^{K}\left(\gamma^*_{w1,K,i} - \gamma^0_{w1,K,i}\right)^2 \\
&\leq \sum_{i=2}^{K}\left(\gamma^*_{w1,K,i} - \gamma^0_{w1,K,i}\right)^2 + \left(\gamma^*_{w0,K} - \gamma^0_{w0,K}\right)^2 \\
&= \left\|\gamma^*_{w,K} - \gamma^0_{w,K}\right\|^2 \\
&= O\left(\left(K^{\frac{1}{2}}K^{-\frac{s}{d}}\right)^2\right) \qquad \text{(B.28)}
\end{aligned}$$

by Lemma $A.6$ (iii) and

$$\begin{aligned}
\left\|\hat{\gamma}_{w1,K} - \gamma^0_{w1,K}\right\|^2 &= \sum_{i=2}^{K}\left(\hat{\gamma}_{w1,K,i} - \gamma^0_{w1,K,i}\right)^2 \\
&\leq \sum_{i=2}^{K}\left(\hat{\gamma}_{w1,K,i} - \gamma^0_{w1,K,i}\right)^2 + \left(\hat{\gamma}_{w0,K} - \gamma^0_{w0,K}\right)^2 \\
&= \left\|\hat{\gamma}_{w,K} - \gamma^0_{w,K}\right\|^2 \\
&= O_p\left(\left(K^{\frac{1}{2}}N^{-\frac{1}{2}} + K^{-\frac{s}{d}}\right)^2\right). \qquad \text{(B.29)}
\end{aligned}$$

by Lemma $A.6$ (iv). We may choose the last $(K-1)$ elements of the approximating sequence to be equal, $\gamma^0_{11,K} = \gamma^0_{01,K}$. This allows us to bound $\hat{\delta}$ and δ^* by the following

$$\begin{aligned}
\left\|\hat{\delta}\right\| &= \left\|\hat{\gamma}_{11,K} - \hat{\gamma}_{01,K}\right\| \\
&= \left\|\hat{\gamma}_{11,K} - \gamma^0_{11,K} + \gamma^0_{01,K} - \hat{\gamma}_{01,K}\right\| \\
&\leq \left\|\hat{\gamma}_{11,K} - \gamma^0_{11,K}\right\| + \left\|\gamma^0_{01,K} - \hat{\gamma}_{01,K}\right\| \\
&= O_p\left(K^{\frac{1}{2}}N^{-\frac{1}{2}} + K^{-\frac{s}{d}}\right) \qquad \text{(B.30)}
\end{aligned}$$

by equation $(B.29)$. Also,

$$\begin{aligned}
\left\|\delta^*\right\| &= \left\|\gamma^*_{11,K} - \gamma^*_{01,K}\right\| \\
&= \left\|\gamma^*_{11,K} - \gamma^0_{11,K} + \gamma^0_{01,K} - \gamma^*_{01,K}\right\| \\
&\leq \left\|\gamma^*_{11,K} - \gamma^0_{11,K}\right\| + \left\|\gamma^0_{01,K} - \gamma^*_{01,K}\right\| \\
&= O\left(K^{\frac{1}{2}}K^{-\frac{s}{d}}\right) \qquad \text{(B.31)}
\end{aligned}$$

by equation $(B.28)$.

Next note that,

$$|T^{*\prime} - T'| = N\left\{\left(\hat{\delta} - \delta^*\right)'\hat{V}_{11}^{-1}\left(\hat{\delta} - \delta^*\right) - \hat{\delta}'\hat{V}_{11}^{-1}\hat{\delta}\right\} = N\left\{\delta^{*\prime}\hat{V}_{11}^{-1}\delta^* - 2\cdot\hat{\delta}'\hat{V}_{11}^{-1}\delta^*\right\}$$

Consider the first term,

$$\begin{aligned}
\left|\delta^{*\prime}\hat{V}_{11}^{-1}\delta^*\right| &= \left|\text{tr}\left(\delta^{*\prime}\hat{V}_{11}^{-1}\delta^*\right)\right| && \text{(B.32)}\\
&\leq \|\delta^*\|^2 \cdot \lambda_{max}\left(\hat{V}_{11}^{-1}\right) && \text{(B.33)}\\
&\leq \|\delta^*\|^2 \cdot \lambda_{max}(\hat{V}^{-1}) && \text{(B.34)}\\
&\leq C \cdot \|\delta^*\|^2 + o_p(1) && \text{(B.35)}\\
&= O\left(KK^{-\frac{2s}{d}}\right) && \text{(B.36)}
\end{aligned}$$

$(B.34)$ follows from Lemma $B.1$, $(B.35)$ follows from Lemma $A.2$ and Assumption 3.3, and $(B.36)$ follows from $(B.31)$. Now consider the second term,

$$\begin{aligned}
2 \cdot \left|\hat{\delta}'\hat{V}_{11}^{-1}\delta^*\right| &= 2 \cdot \left|\text{tr}\left(\hat{\delta}'\hat{V}_{11}^{-1}\delta^*\right)\right| && \text{(B.37)}\\
&\leq 2 \cdot \left\|\hat{\delta}\right\| \cdot \|\delta^*\| \cdot \lambda_{max}\left(\hat{V}_{11}^{-1}\right) && \text{(B.38)}\\
&\leq 2 \cdot \left\|\hat{\delta}\right\| \cdot \|\delta^*\| \cdot \lambda_{max}(\hat{V}^{-1}) && \text{(B.39)}\\
&\leq C \cdot \left\|\hat{\delta}\right\| \cdot \|\delta^*\| + o_p(1) && \text{(B.40)}\\
&= O_p\left(K^{\frac{1}{2}}N^{-\frac{1}{2}} + K^{-\frac{s}{d}}\right) \cdot O\left(K^{\frac{1}{2}}K^{-\frac{s}{d}}\right) && \text{(B.41)}
\end{aligned}$$

$(B.39)$ follows from Lemma $B.1$, $(B.40)$ follows from Lemma $A.2$ and Assumption 3.3, and $(B.41)$ follows from equations $(B.30)$ and $(B.31)$. Thus,

$$|T^{*\prime} - T'| = O(N) \cdot \left[O\left(KK^{-\frac{2s}{d}}\right) + O_p\left(K^{\frac{1}{2}}N^{-\frac{1}{2}} + K^{-\frac{s}{d}}\right) \cdot O\left(K^{\frac{1}{2}}K^{-\frac{s}{d}}\right)\right]$$

All three terms are $o_p(1)$ by Assumptions 3.2 and 3.3 and so we have

$$N \cdot (\hat{\gamma}_{11,K} - \hat{\gamma}_{01,K})' \hat{V}_{11}^{-1} (\hat{\gamma}_{11,K} - \hat{\gamma}_{01,K}) \xrightarrow{d} \chi^2 (K-1).$$

Finally, by Lemma $A.5$ replacing K with $(K-1)$, we have that

$$\frac{1}{\sqrt{2(K-1)}}\left[N \cdot \left((\hat{\gamma}_{11,K} - \hat{\gamma}_{01,K})' \hat{V}_{11}^{-1} (\hat{\gamma}_{11,K} - \hat{\gamma}_{01,K})\right) - (K-1)\right] \xrightarrow{d} \mathcal{N}(0,1)$$

■

Proof of Theorem 3.4 First, note that we may partition $R_K(x)$ as

$$R_K(x) = \begin{pmatrix} R_1 \\ R_{K-1}(x) \end{pmatrix}.$$

Next, consider

$$\begin{aligned}
\rho_N \cdot \sup_{x\in\mathbb{X}} |\Delta(x)| &= \sup_{x} |\mu_1(x) - \mu_0(x) - \tau|\\
&\leq \sup_{x\in\mathbb{X}} \left|R_K(x)'\gamma_{1,K}^0 - \mu_1(x)\right| + \sup_{x\in\mathbb{X}} \left|R_K(x)'\gamma_{0,K}^0 - \mu_0(x)\right|\\
&\quad + \sup_{x\in\mathbb{X}} \left|R_K(x)'\hat{\gamma}_{0,K} - R_K(x)'\gamma_{0,K}^0\right| + \sup_{x\in\mathbb{X}} \left|R_K(x)'\hat{\gamma}_{1,K} - R_K(x)'\gamma_{1,K}^0\right|\\
&\quad + \sup_{x\in\mathbb{X}} |R_{K-1}(x)'\hat{\gamma}_{11,K} - R_{K-1}(x)'\hat{\gamma}_{01,K}| + |R_1\hat{\gamma}_{10,K} - R_1\hat{\gamma}_{00,K} - \tau|\\
&\leq \sup_{x\in\mathbb{X}} \left|R_K(x)'\gamma_{0,K}^0 - \mu_0(x)\right| + \sup_{x\in\mathbb{X}} \left|R_K(x)'\gamma_{1,K}^0 - \mu_1(x)\right|
\end{aligned}$$

$$
\begin{aligned}
&+\sup_{x\in\mathbb{X}}\|R_K(x)\|\cdot\left\|\hat{\gamma}_{0,K}-\gamma^0_{0,K}\right\|+\sup_{x\in\mathbb{X}}\|R_K(x)\|\cdot\left\|\hat{\gamma}_{1,K}-\gamma^0_{1,K}\right\|\\
&+\sup_{x\in\mathbb{X}}\|R_{K-1}(x)\|\cdot\|\hat{\gamma}_{11,K}-\hat{\gamma}_{01,K}\|+|R_1\hat{\gamma}_{10,K}-R_1\hat{\gamma}_{00,K}-\tau|\\
\leq{}&\sup_{x\in\mathbb{X}}\left|R_K(x)'\gamma^0_{0,K}-\mu_0(x)\right|+\sup_{x\in\mathbb{X}}\left|R_K(x)'\gamma^0_{1,K}-\mu_1(x)\right|\\
&+\zeta(K)\cdot\left\|\hat{\gamma}_{0,K}-\gamma^0_{0,K}\right\|+\zeta(K)\cdot\left\|\hat{\gamma}_{1,K}-\gamma^0_{1,K}\right\|+\zeta(K)\cdot\|\hat{\gamma}_{11,K}-\hat{\gamma}_{01,K}\|\\
&+|R_1\hat{\gamma}_{10,K}-R_1\hat{\gamma}_{00,K}-\tau|
\end{aligned}
$$

Thus,

$$
\begin{aligned}
\|\hat{\gamma}_{11,K}-\hat{\gamma}_{01,K}\|\geq{}&\zeta^{-1}(K)\cdot\rho_N\cdot\sup_{x\in\mathbb{X}}|\Delta(x)|-\zeta^{-1}(K)\cdot\sup_{x\in\mathbb{X}}\left|R_K(x)'\gamma^0_{0,K}-\mu_0(x)\right|\\
&-\zeta^{-1}(K)\cdot\sup_{x\in\mathbb{X}}\left|R_K(x)'\gamma^0_{1,K}-\mu_1(x)\right|-\left\|\hat{\gamma}_{0,K}-\gamma^0_{0,K}\right\|-\left\|\hat{\gamma}_{1,K}-\gamma^0_{1,K}\right\|\\
&-\zeta^{-1}(K)\cdot|R_1\hat{\gamma}_{10,K}-R_1\hat{\gamma}_{00,K}-\tau|\\
\geq{}&\zeta^{-1}(K)\cdot\rho_N\cdot C_0\cdot\Bigg(1-\frac{\sup_{x\in\mathbb{X}}\left|R_K(x)'\gamma^0_{0,K}-\mu_0(x)\right|}{\rho_N\cdot C_0}-\frac{\sup_{x\in\mathbb{X}}\left|R_K(x)'\gamma^0_{1,K}-\mu_1(x)\right|}{\rho_N\cdot C_0}\\
&-\zeta(K)\cdot\frac{\left\|\hat{\gamma}_{0,K}-\gamma^0_{0,K}\right\|}{\rho_N\cdot C_0}-\zeta(K)\cdot\frac{\left\|\hat{\gamma}_{1,K}-\gamma^0_{1,K}\right\|}{\rho_N\cdot C_0}-\frac{|R_1\hat{\gamma}_{10,K}-R_1\hat{\gamma}_{00,K}-\tau|}{\rho_N\cdot C_0}\Bigg)
\end{aligned}
$$

By the results in the proof of Theorem 3.2 we need only consider,

$$
\frac{|R_1\hat{\gamma}_{10,K}-R_1\hat{\gamma}_{00,K}-\tau|}{\rho_N\cdot C_0}=O_p\left(N^{-1/2}\right)\cdot O\left(N^{1/2-3\nu/2-\varepsilon}\right)=o_p(1)
$$

Where $\sqrt{N}$-consistency follows since we are now in the case of the parametric term in a partially-linear model. Thus, we now have that

$$
\|\hat{\gamma}_{11,K}-\hat{\gamma}_{01,K}\|\geq\zeta^{-1}(K)\cdot\rho_N\cdot C_0
$$

and by following the steps in the proof of Theorem 3.2, it follows that for any M',

$$
\Pr\left(N^{1/2}K^{-1/2}\|\hat{\gamma}_{11,K}-\hat{\gamma}_{01,K}\|>M'\right)\longrightarrow 1. \tag{B.42}
$$

Next, we will show that this implies that

$$
\Pr\left(\frac{N\left(\hat{\gamma}_{11,K}-\hat{\gamma}_{01,K}\right)'V_{11}^{-1}\left(\hat{\gamma}_{11,K}-\hat{\gamma}_{01,K}\right)-(K-1)}{\sqrt{2\,(K-1)}}>2M\right)\longrightarrow 1.
$$

Let $\underline{\lambda}_{11}$ denote $\lambda_{\min}(V_{11}^{-1})$. Then by Lemma $B.1$, $\underline{\lambda}_{11}$ is bounded away from zero by at least the lower bound $\underline{\lambda}$. Thus,

$$
\begin{aligned}
&\Pr\left(\frac{N\left(\hat{\gamma}_{11,K}-\hat{\gamma}_{01,K}\right)'V_{11}^{-1}\left(\hat{\gamma}_{11,K}-\hat{\gamma}_{01,K}\right)-(K-1)}{\sqrt{2\,(K-1)}}>2M\right)\\
&\quad=\Pr\left(N\left(\hat{\gamma}_{11,K}-\hat{\gamma}_{01,K}\right)'V_{11}^{-1}\left(\hat{\gamma}_{11,K}-\hat{\gamma}_{01,K}\right)>M\sqrt{8}\,(K-1)^{1/2}+(K-1)\right)\\
&\quad\geq\Pr\left(N\underline{\lambda}\left(\hat{\gamma}_{11,K}-\hat{\gamma}_{01,K}\right)'\left(\hat{\gamma}_{11,K}-\hat{\gamma}_{01,K}\right)>M\sqrt{8}\,(K-1)^{1/2}+(K-1)\right)\\
&\quad=\Pr\left(NK^{-1}\|\hat{\gamma}_{11,K}-\hat{\gamma}_{01,K}\|^2>\underline{\lambda}^{-1}\left(1+M\sqrt{8}\,(K-1)^{1/2}K^{-1}-K^{-1}\right)\right)
\end{aligned}
$$

$$= \Pr\left(N^{1/2}K^{-1/2}\left\|\hat{\gamma}_{11,K}-\hat{\gamma}_{01,K}\right\| > \underline{\lambda}^{-1/2}\left(1+M\sqrt{8}\,(K-1)^{1/2}K^{-1}-K^{-1}\right)^{1/2}\right).$$

Again, for any M, for large enough N, we have

$$\underline{\lambda}^{-1/2}\left(1+M\sqrt{8}\,(K-1)^{1/2}K^{-1}-K^{-1}\right)^{1/2} < 2\underline{\lambda}^{-1/2},$$

it follows that this probability is for large N bounded from below by the probability

$$= \Pr\left(N^{1/2}K^{-1/2}\left\|\hat{\gamma}_{11,K}-\hat{\gamma}_{01,K}\right\| > 2\underline{\lambda}^{-1/2}\right),$$

which goes to one by (B.42). Finally, we show that this implies that

$$\Pr(T' > M) = \Pr\left(\frac{N\left(\hat{\gamma}_{11,K}-\hat{\gamma}_{01,K}\right)'\hat{V}_{11}^{-1}\left(\hat{\gamma}_{11,K}-\hat{\gamma}_{01,K}\right)-(K-1)}{\sqrt{2\,(K-1)}} > M\right) \longrightarrow 1.$$

Let $\hat{\underline{\lambda}}_{11} = \lambda_{\min}(\hat{V}_{11}^{-1})$ be the minimum eigenvalue of the matrix $\hat{V}_{11}^{-1}$. Lemmas A.1, A.2 and B.1, Assumption 3.3 and the consistency of $\hat{\sigma}^2_{0,K}$, $\hat{\sigma}^2_{1,K}$ and $\hat{c}$ imply that $\hat{\underline{\lambda}}_{11} - \underline{\lambda}_{11} = o_p(1)$. Since $\underline{\lambda}_{11}$ is bounded away from zero, it follows that $\hat{\underline{\lambda}}_{11}$ is bounded away from zero with probability going to one. Let B denote the event that $\lambda_{\min}(\hat{V}_{11}^{-1}) > \underline{\lambda}/2$ and $\frac{N(\hat{\gamma}_{11,K}-\hat{\gamma}_{01,K})'(\hat{\gamma}_{11,K}-\hat{\gamma}_{01,K})-(K-1)}{\sqrt{2(K-1)}} > 2M/\underline{\lambda}$. The probability of the event B converges to one since,

$$\Pr\left(\frac{N\left(\hat{\gamma}_{11,K}-\hat{\gamma}_{01,K}\right)'V_{11}^{-1}\left(\hat{\gamma}_{11,K}-\hat{\gamma}_{01,K}\right)-(K-1)}{\sqrt{2\,(K-1)}} > 2M\right)$$

$$\geq \Pr\left(\frac{N\left(\hat{\gamma}_{11,K}-\hat{\gamma}_{01,K}\right)\left(\hat{\gamma}_{11,K}-\hat{\gamma}_{01,K}\right)-(K-1)}{\sqrt{2\,(K-1)}} > 2M/\underline{\lambda}\right)$$

Also, B implies that

$$\frac{N\left(\hat{\gamma}_{11,K}-\hat{\gamma}_{01,K}\right)'\hat{V}_{11}^{-1}\left(\hat{\gamma}_{11,K}-\hat{\gamma}_{01,K}\right)-(K-1)}{\sqrt{2\,(K-1)}}$$

$$\geq \frac{N\hat{\underline{\lambda}}'\left(\hat{\gamma}_{11,K}-\hat{\gamma}_{01,K}\right)'\left(\hat{\gamma}_{11,K}-\hat{\gamma}_{01,K}\right)-(K-1)}{\sqrt{2\,(K-1)}}$$

$$> \frac{N\underline{\lambda}/2\left(\hat{\gamma}_{11,K}-\hat{\gamma}_{01,K}\right)'\left(\hat{\gamma}_{11,K}-\hat{\gamma}_{01,K}\right)-(K-1)}{\sqrt{2\,(K-1)}} > \underline{\lambda}/2\cdot 2M/\underline{\lambda} = M.$$

Hence $\Pr(T' > M)) \to 1$. ■

Lemma B.2 *Suppose Assumptions 2.1-2.3 and 3.1-3.3 hold. Then,*

$$K^{\frac{1}{2}}\cdot\left|\hat{\sigma}^2_{w,K}-\sigma^2_w\right| = o_p(1)$$

Proof

$$\left|\hat{\sigma}^2_{w,K}-\sigma^2_w\right| = \left|\frac{1}{N_w}\sum_{i|W_i=w}\left(Y_i-\hat{\mu}_{w,K}(X_i)\right)^2-\sigma^2_w\right|$$

$$= \left|\frac{1}{N_w}\sum_{i|W_i=w}\left(Y_i-\mu_w(X_i)\right)^2-\sigma^2_w+\frac{1}{N_w}\sum_{i|W_i=w}\left(\hat{\mu}_{w,K}(X_i)-\mu_w(X_i)\right)^2\right.$$

$$+\frac{1}{N_w}\sum_{i|W_i=w}\left(Y_i-\mu_w\left(X_i\right)\right)\left(\hat{\mu}_{w,K}\left(X_i\right)-\mu_w\left(X_i\right)\right)\Bigg|$$

$$\leq\left|\frac{1}{N_w}\sum_{i|W_i=w}\left(Y_i-\mu_w\left(X_i\right)\right)^2-\sigma_w^2\right| \tag{B.43}$$

$$+\frac{1}{N_w}\sum_{i|W_i=w}\left(\hat{\mu}_{w,K}\left(X_i\right)-\mu_w\left(X_i\right)\right)^2 \tag{B.44}$$

$$+\left|\frac{1}{N_w}\sum_{i|W_i=w}\left(Y_i-\mu_w\left(X_i\right)\right)\left(\hat{\mu}_{w,K}\left(X_i\right)-\mu_w\left(X_i\right)\right)\right| \tag{B.45}$$

Note first that equation (B.43) is $O_p\left(N^{-\frac{1}{2}}\right)$ by the weak law of large numbers. Now consider equation (B.44),

$$\frac{1}{N_w}\sum_{i|W_i=w}\left(\hat{\mu}_{w,K}\left(X_i\right)-\mu_w\left(X_i\right)\right)^2$$

$$\leq\sup_x\left|\hat{\mu}_{w,K}\left(x\right)-\mu_w\left(x\right)\right|^2$$

$$=\left(O_p\left(\zeta\left(K\right)K^{-\frac{s}{d}}\right)+O_p\left(\zeta\left(K\right)K^{\frac{1}{2}}N^{-\frac{1}{2}}\right)\right)^2$$

where the last line follows from Lemma A.6 (iv), since $\zeta(K)=O(K)$, and by Assumption 3.3. Finally, consider equation (B.45). Note first that,

$$\left|\frac{1}{N_w}\sum_{i|W_i=w}\left(Y_i-\mu_w\left(X_i\right)\right)\left(\hat{\mu}_{w,K}\left(X_i\right)-\mu_w\left(X_i\right)\right)\right|$$

$$\leq\left|\frac{1}{N_w}\sum_{i|W_i=w}\left(Y_i-\mu_w\left(X_i\right)\right)\left(\hat{\mu}_{w,K}\left(X_i\right)-\mu^*_{w,K}\left(X_i\right)\right)\right| \tag{B.46}$$

$$+\left|\frac{1}{N_w}\sum_{i|W_i=w}\left(Y_i-\mu_w\left(X_i\right)\right)\left(\mu^*_{w,K}\left(X_i\right)-\mu_w\left(X_i\right)\right)\right| \tag{B.47}$$

We will first work with equation (B.47). Note that the individual summands have mean zero conditional on $\mathbf{X}$. Thus,

$$\mathbb{V}\left[\frac{1}{N_w}\sum_{i|W_i=w}\left(Y_i-\mu_w\left(X_i\right)\right)\left(\mu^*_{w,K}\left(X_i\right)-\mu_w\left(X_i\right)\right)\right]$$

$$=\mathbb{E}\left[\mathbb{V}\left[\frac{1}{N_w}\sum_{i|W_i=w}\left(Y_i-\mu_w\left(X_i\right)\right)\left(\mu^*_{w,K}\left(X_i\right)-\mu_w\left(X_i\right)\right)\Big|\mathbf{X}\right]\right]$$

$$=\mathbb{E}\left[\frac{1}{N_w}\mathbb{V}\left[Y_w|\mathbf{X}\right]\left(\mu^*_{w,K}\left(X\right)-\mu_w\left(X\right)\right)^2\right]$$

$$=\sigma_w^2\frac{1}{N_w}\cdot\mathbb{E}\left[\left(\mu^*_{w,K}\left(X\right)-\mu_w\left(X\right)\right)^2\right]$$

$$
\begin{aligned}
&= \sigma_w^2 \frac{1}{N_w} \cdot \mathbb{E}\left[\left(\mu^*_{w,K}(X) - \mu_w(X)\right)^2\right] \\
&\leq C \cdot N^{-1} \cdot \sup_x \left|\mu^*_{w,K}(x) - \mu_w(x)\right|^2 \\
&= C \cdot N^{-1} \zeta(K)^2 K K^{-\frac{2s}{d}} \\
&= O\left(N^{-1} \zeta(K)^2 K K^{-\frac{2s}{d}}\right)
\end{aligned}
$$

where the penultimate line follows by Lemma A.6 (i) and (ii). So finally,

$$
\begin{aligned}
&\mathbb{E}\left[\left|\frac{1}{N_w} \sum_{i|W_i=w} \left(Y_i - \mu_w(X_i)\right)\left(\mu^*_{w,K}(X_i) - \mu_w(X_i)\right)\right|\right] \\
&\quad \leq \left(\mathbb{V}\left[\frac{1}{N_w} \sum_{i|W_i=w} \left(Y_i - \mu_w(X_i)\right)\left(\mu^*_{w,K}(X_i) - \mu_w(X_i)\right)\right]\right)^{\frac{1}{2}} \\
&\quad = O\left(N^{-\frac{1}{2}} \zeta(K) K^{\frac{1}{2}} K^{-\frac{s}{d}}\right)
\end{aligned}
$$

so then by Markov's inequality, equation (B.47) is $O\left(N^{-\frac{1}{2}} \zeta(K) K^{\frac{1}{2}} K^{-\frac{s}{d}}\right)$. Now consider equation (B.46),

$$
\begin{aligned}
&\left|\frac{1}{N_w} \sum_{i|W_i=w} \left(Y_i - \mu_w(X_i)\right)\left(\hat{\mu}_{w,K}(X_i) - \mu^*_{w,K}(X_i)\right)\right| \\
&\leq \frac{1}{N_w} \sum_{i|W_i=w} \left|\left(Y_i - \mu_w(X_i)\right)\left(\hat{\mu}_{w,K}(X_i) - \mu^*_{w,K}(X_i)\right)\right| \\
&\leq \sup_x \left|\hat{\mu}_{w,K}(x) - \mu^*_{w,K}(x)\right| \cdot 2 \cdot |Y| \\
&= O_p\left(\zeta(K)^3 N^{-\frac{1}{2}}\right)
\end{aligned}
$$

where the last line follows by Markov's inequality and Assumption 3.2 and Imbens, Newey and Ridder (2006). Finally, combining terms yields,

$$
\begin{aligned}
\left|\hat{\sigma}^2_{w,K} - \sigma^2_w\right| &= O_p\left(N^{-\frac{1}{2}}\right) + O_p\left(\zeta(K)^2 K^{-\frac{2s}{d}}\right) + O_p\left(\zeta(K)^2 K N^{-1}\right) + O_p\left(\zeta(K)^2 K^{\frac{1}{2}} N^{-\frac{1}{2}} K^{-\frac{s}{d}}\right) \\
&\quad + O\left(N^{-\frac{1}{2}} \zeta(K) K^{\frac{1}{2}} K^{-\frac{s}{d}}\right) + O_p\left(\zeta(K)^3 N^{-\frac{1}{2}}\right)
\end{aligned}
$$

and so,

$$
\begin{aligned}
K^{\frac{1}{2}} \cdot \left|\hat{\sigma}^2_{w,K} - \sigma^2_w\right| &= O_p\left(K^{\frac{1}{2}} N^{-\frac{1}{2}}\right) + O_p\left(\zeta(K)^2 K^{\frac{1}{2}} K^{-\frac{2s}{d}}\right) + O_p\left(\zeta(K)^2 K^{\frac{3}{2}} N^{-1}\right) + O_p\left(\zeta(K)^2 K N^{-\frac{1}{2}} K^{-\frac{s}{d}}\right) \\
&\quad O\left(N^{-\frac{1}{2}} \zeta(K) K K^{-\frac{s}{d}}\right) + O_p\left(\zeta(K)^3 K^{\frac{1}{2}} N^{-\frac{1}{2}}\right)
\end{aligned}
$$

and all five terms are $o_p(1)$ by Assumptions 3.2 and 3.3. ■