Supplement to "On the Choice of Test Statistic for Conditional Moment Inequalities"

Timothy B. Armstrong
Yale University

July 5, 2017

This supplementary appendix contains proofs of the results in the main text as well as auxiliary results. Section B contains auxiliary results used in the rest of this appendix. These results are restatements or simple extensions of well known results on uniform convergence, and do not constitute part of the main novel contribution of the paper. Section C of this appendix derives critical values for CvM statistics with variance weights. Section D contains proofs of the results in the body of the paper.

B Auxiliary Results

We state some results on uniform convergence that will be used in the proofs of the main results. The results in this section are essentially restatements of results used in Armstrong (2014b), which are in turn minor extensions of results in Pollard (1984). Throughout this section, we consider iid observations $Z_1, \ldots, Z_n$ and a sequence of classes of functions $\mathcal{F}_n$ on the sample space. Let $\sigma(f)^2 = Ef(Z_i)^2 - (Ef(Z_i))^2$ and let $\hat{\sigma}(f)^2 = E_n f(Z_i)^2 - (E_n f(Z_i))^2$.

Lemma B.1. *Suppose that* $|f(Z_i)| \leq \overline{f}$ *a.s. and that*

$$\sup_{n\in\mathbb{N}} \sup_{Q} N(\varepsilon, \mathcal{F}_n, L_1(Q)) \leq A\varepsilon^{-W}$$

for some A *and* W*, where* N *is the covering number defined in Pollard (1984) and the supremum over* Q *is over all probability measures. Let* σ_n *be a sequence of constants with* $\sigma_n\sqrt{n/\log n} \to \infty$*. Then, for some constant* C*,*

$$\frac{\sqrt{n}}{\sqrt{\log n}} \sup_{f\in\mathcal{F}_n} \left| \frac{(E_n - E)f(Z_i)}{\sigma(f) \vee \sigma_n} \right| \leq C$$

with probability approaching one and

$$\sup_{f\in\mathcal{F}_n}\left|\frac{(E_n-E)f(Z_i)}{\sigma(f)^2\vee\sigma_n^2}\right|\stackrel{p}{\to}0.$$

Proof. The first display follows by applying Lemma A.1 in Armstrong (2014b) to the sequence of classes of functions $\{f-E_Pf(Z_i)|f\in\mathcal{F}_n\}$, which satisfies the conditions of that lemma by Lemma A.5 in Armstrong (2014b). The second display follows from the first display since

$$\sup_{f\in\mathcal{F}_n}\left|\frac{(E_n-E)f(Z_i)}{\sigma(f)^2\vee\sigma_n^2}\right|\leq\frac{1}{\sigma_n}\sup_{f\in\mathcal{F}_n}\left|\frac{(E_n-E)f(Z_i)}{\sigma(f)\vee\sigma_n}\right|=\frac{\sqrt{\log n}}{\sigma_n\sqrt{n}}\frac{\sqrt{n}}{\sqrt{\log n}}\sup_{f\in\mathcal{F}_n}\left|\frac{(E_n-E)f(Z_i)}{\sigma(f)\vee\sigma_n}\right|$$

and $\sqrt{\log n}/(\sigma_n\sqrt{n})\to 0$. □

Lemma B.2. *Under the conditions of Lemma B.1,*

$$\sup_{f\in\mathcal{F}_n}\left|\frac{\hat{\sigma}(f)\vee\sigma_n}{\sigma(f)\vee\sigma_n}-1\right|\stackrel{p}{\to}0.$$

Proof. By continuity of $t\mapsto\sqrt{t}$ at 1, it suffices to prove that $\sup_{f\in\mathcal{F}_n}\left|\frac{\hat{\sigma}(f)^2\vee\sigma_n^2}{\sigma(f)^2\vee\sigma_n^2}-1\right|\stackrel{p}{\to}0$. We have

$$\sup_{f\in\mathcal{F}_n}\left|\frac{\hat{\sigma}(f)^2\vee\sigma_n^2}{\sigma(f)^2\vee\sigma_n^2}-1\right|=\sup_{f\in\mathcal{F}_n}\left|\frac{\hat{\sigma}(f)^2\vee\sigma_n^2-\sigma(f)^2\vee\sigma_n^2}{\sigma(f)^2\vee\sigma_n^2}\right|\leq\sup_{f\in\mathcal{F}_n}\left|\frac{\hat{\sigma}(f)^2-\sigma(f)^2}{\sigma(f)^2\vee\sigma_n^2}\right|.$$

Note that

$$\hat{\sigma}(f)^2-\sigma(f)^2=(E_n-E)[f(Z_i)-Ef(Z_i)]^2-[(E_n-E)f(Z_i)]^2. \tag{14}$$

Since $\sigma[(f-Ef(Z_i))^2]^2\leq E[f(Z_i)-Ef(Z_i)]^4\leq 4\overline{f}^2\sigma(f)^2$, we have

$$\sup_{f\in\mathcal{F}_n}\frac{|(E_n-E)[f(Z_i)-Ef(Z_i)]^2|}{\sigma(f)^2\vee\sigma_n^2}\leq\sup_{f\in\mathcal{F}_n}\frac{|(E_n-E)[f(Z_i)-Ef(Z_i)]^2|}{\sigma[(f-Ef(Z_i))^2]^2\vee\sigma_n^2}\cdot(4\overline{f}^2)\vee 1$$

which converges in probability to zero by Lemma B.1 (using Lemma A.5 in Armstrong, 2014b to verify that the sequence of classes of functions $\{[f-Ef(Z_i)]^2|f\in\mathcal{F}_n\}$ satisfies the conditions of the lemma). Since

$$\frac{[(E_n - E)f(Z_i)]^2}{\sigma(f)^2 \vee \sigma_n^2} \stackrel{p}{\to} 0$$

by Lemma B.1, the result now follows from this and the triangle inequality applied to (14). □

Lemma B.3. *Suppose that $|f(Z_i)| \leq \overline{f}$ and that $\sigma_n \sqrt{n} \geq 1$. Then*

$$E \left| \frac{\sqrt{n}(E_n - E)f(Z_i)}{\sigma(f) \vee \sigma_n} \right|^p \leq C_{p,\overline{f}}$$

for a constant $C_{p,\overline{f}}$ that depends only on p and $\overline{f}$.

Proof. By Bernstein's inequality,

$$P\left(\left| \frac{\sqrt{n}(E_n - E)f(Z_i)}{\sigma(f) \vee \sigma_n} \right| > t \right) \leq \exp\left(-\frac{1}{2} \frac{n[\sigma(f) \vee \sigma_n]^2 t^2}{n\sigma^2(f) + \frac{1}{3} \cdot 2\overline{f} \cdot \sqrt{n}[\sigma(f) \vee \sigma_n] t} \right)$$
$$\leq \exp\left(-\frac{1}{2} \frac{t^2}{1 + \frac{1}{3} \cdot 2\overline{f} \cdot \frac{t}{\sqrt{n}[\sigma(f) \vee \sigma_n]}} \right) \leq \exp\left(-\frac{1}{2} \frac{t^2}{1 + \frac{1}{3} \cdot 2\overline{f} \cdot t} \right) \leq \exp\left(-\frac{1}{2} \frac{t^2}{1 + \frac{1}{3} \cdot 2\overline{f} \cdot t} \right).$$

For $t \geq 1$, this is bounded by $\exp\left(-\frac{t}{2 + \frac{2}{3} \cdot 2\overline{f}} \right)$. Thus,

$$E \left| \frac{\sqrt{n}(E_n - E)f(Z_i)}{\sigma(f) \vee \sigma_n} \right|^p = \int_{t=0}^{\infty} P\left(\left| \frac{\sqrt{n}(E_n - E)f(Z_i)}{\sigma(f) \vee \sigma_n} \right|^p > t \right) dt$$
$$\leq 1 + \int_{t=1}^{\infty} \exp\left(-\frac{t^{1/p}}{2 + \frac{2}{3} \cdot 2\overline{f}} \right) dt$$

which is finite and depends only on p and $\overline{f}$ as claimed. □

C Critical Values for CvM Statistics with Variance Weights

For bounded choices of ω (which corresponds to σ_n bounded away from zero when a truncated variance weighting is used), Kim (2008) and Andrews and Shi (2013) derive a $\sqrt{n}$ rate of convergence to an asymptotic distribution that may be degenerate. Armstrong (2014b) shows that letting σ_n go to zero generally decreases the rate of convergence to $\sqrt{n/\log n}$ for the KS statistic $T_{n,\infty,\omega}$. In contrast to the KS case, CvM statistics do not behave much differently if the variance is allowed to go to zero, although some additional arguments are needed to show this.

To deal with the behavior of the CvM statistic for small variances, I place the following condition on the measure over which the sample means are integrated.

Assumption C.1. *$\mu(\{g|\sigma_j(\theta,g)\leq\delta\})\to 0$ as $\delta\to 0$ for all j.*

This condition will hold for the choices of $\mathcal{G}$ and μ used in the body of the paper, and also allow for more general choices of $\mathcal{G}$ and μ. I also make the following assumption on the complexity of the class of functions $\mathcal{G}$, which is also satisfied by the class used in the paper.

Assumption C.2. *For some constants A and ε, the covering number $N(\varepsilon,\mathcal{G},L_1(Q))$ defined in Pollard (1984) satisfies*

$$\sup_Q N(\varepsilon,\mathcal{G},L_1(Q))\leq A\varepsilon^{-W},$$

whre the supremum is over all probability measures.

The following condition imposes a bounded distribution of the function m.

Assumption C.3. *For some nonrandom constant $\overline{Y}$, $|m_j(W_i,\theta)|\leq\overline{Y}$ for each j with probability one.*

Theorem C.1. *Suppose that $\sigma_n\sqrt{n/\log n}\to\infty$ and that Assumptions C.1, C.2 and C.3 hold. Then, for $\theta\in\Theta_0$,*

$$n^{1/2}T_{n,p,(\hat{\sigma}\vee\sigma_n)^{-1},\mu}(\theta)\leq\left[\int\sum_{j=1}^{d_Y}\left|\frac{\sqrt{n}(E_n-E)m_j(W_i,\theta)g(X_i)}{\hat{\sigma}_j(\theta,g)\vee\sigma_n}\right|_-^p\,d\mu(g)\right]^{1/p}$$

$$\stackrel{d}{\to}\left[\int\sum_{j=1}^{d_Y}|\mathbb{G}_j(g,\theta)/\sigma_j(\theta,g)|_-^p\,d\mu(g)\right]^{1/p}$$

where $\mathbb{G}(g,\theta)$ is a vector of Gaussian processes with covariance function

$$\rho(g,\tilde{g})=E[m(W_i,\theta)g(X_i)-Em(W_i,\theta)g(X_i)][m(W_i,\theta)\tilde{g}(X_i)-Em(W_i,\theta)\tilde{g}(X_i)]'.$$

Proof. The result with the integral truncated over $\{\sigma_j(\theta,g)\leq\delta|\text{all } j\}$ follows immediately from standard arguments using functional central limit theorems. This, along with Lemma C.1 below gives, letting $Z_n(\delta)$ be the integral truncated at $\{\sigma_j(\theta,g)\leq\delta|\text{all } j\}$ and $Z(\delta)$ be the limiting variable with this truncation,

$$P(Z_n(\delta)-\varepsilon\leq t)-\varepsilon\leq P(n^{1/2}T_{n,p,\omega,\mu}(\theta)\leq t)\leq P(Z_n(\delta)\leq t)$$

for large enough n for any $\varepsilon>0$. The lim inf of the left hand size is greater than $P(Z(\delta)\leq t-2\varepsilon)-2\varepsilon$, and the lim sup of the right hand side is less than $P(Z(\delta)\leq t+\varepsilon)+\varepsilon$. We can bound $P(Z(\delta)\leq t-2\varepsilon)-2\varepsilon$ from below by $P(Z\leq t-2\varepsilon)-2\varepsilon$, and we can bound $P(Z(\delta)\leq t+\varepsilon)+\varepsilon$ from above by $P(Z\leq t+2\varepsilon)+2\varepsilon$ by making δ small enough by a version of Lemma C.1 for the limiting process. Since ε was arbitrary, this gives the result. □

The proof of the theorem above uses the following auxiliary lemma, which shows that functions g with low enough variance have little effect on the integral asymptotically.

Lemma C.1. *Fix j and suppose that Assumptions C.1, C.2 and C.3 hold, and that the null hypothesis holds under θ. Then, for every $\varepsilon>0$, there exists a $\delta>0$ such that*

$$P\left(\sqrt{n}\left[\int_{\sigma_j(\theta,g)\leq\delta}|E_nm_j(W_i,\theta)g(X_i)/(\hat{\sigma}_j(\theta,g)\vee\sigma_n)|_-^p\,d\mu(g)\right]^{1/p}>\varepsilon\right)\leq\varepsilon.$$

Proof. We have

$$\begin{aligned}&E\int_{\sigma_j(\theta,g)\leq\delta}|\sqrt{n}E_nm_j(W_i,\theta)g(X_i)/(\sigma_j(\theta,g)\vee\sigma_n)|_-^p\,d\mu(g)\\&=\int_{\sigma_j(\theta,g)\leq\delta}E|\sqrt{n}E_nm_j(W_i,\theta)g(X_i)/(\sigma_j(\theta,g)\vee\sigma_n)|_-^p\,d\mu(g)\\&\leq\int_{\sigma_j(\theta,g)\leq\delta}E|\sqrt{n}(E_n-E)m_j(W_i,\theta)g(X_i)/(\sigma_j(\theta,g)\vee\sigma_n)|^p\,d\mu(g)\leq\mu\left(\{g|\sigma_j(\theta,g)\leq\delta\}\right)\cdot C_{p,\overline{Y}}\end{aligned}$$

for $C_{p,\overline{Y}}$ given in Lemma B.3. Applying Markov's inequality and using Assumption C.1, it follows that, for any $\varepsilon>0$, there exists a δ such that

$$P\left(\sqrt{n}\left[\int_{\sigma_j(\theta,g)\leq\delta}|E_nm_j(W_i,\theta)g(X_i)/(\sigma_j(\theta,g)\vee\sigma_n)|_-^p\,d\mu(g)\right]^{1/p}>\varepsilon/2\right)\leq\varepsilon/2.$$

The result follows since

$$\sqrt{n}\left[\int_{\sigma_j(\theta,g)\leq\delta} |E_n m_j(W_i,\theta)g(X_i)/(\hat{\sigma}_j(\theta,g)\vee\sigma_n)|_-^p \, d\mu(g)\right]^{1/p}$$

$$\leq \sqrt{n}\left[\int_{\sigma_j(\theta,g)\leq\delta} |E_n m_j(W_i,\theta)g(X_i)/(\sigma_j(\theta,g)\vee\sigma_n)|_-^p \, d\mu(g)\right]^{1/p} \cdot \sup_g(\sigma_j(\theta,g)\vee\sigma_n)/(\hat{\sigma}_j(\theta,g)\vee\sigma_n)$$

and $\sup_g(\sigma_j(\theta,g)\vee\sigma_n)/(\hat{\sigma}_j(\theta,g)\vee\sigma_n) \leq 2$ with probability approaching one by Lemma B.2. □

D Proofs

This section contains proofs of the results in the body of the paper. The proofs use a number of auxiliary lemmas, which are stated and proved first. In the following, θ_n is always assumed to be a sequence converging to θ_0.

Lemma D.1. *Under the assumptions of Theorem 4.5, there exists a constant C such that*

$$\sup_{x\in\mathbb{R}^{d_X}} \frac{\sqrt{n}}{\sqrt{h^{d_X}\log n}} |(E_n - E)m(W_i,\theta_n)k((X_i-x)/h)| \leq C$$

and

$$\sup_{x\in\mathbb{R}^{d_X}} \frac{\sqrt{n}}{\sqrt{h^{d_X}\log n}} |(E_n - E)k((X_i-x)/h)| \leq C$$

with probability approaching one. In addition,

$$\sup_{\{x|\omega_j(\theta_n,x)>0 \text{ some } j\}} \left|\frac{E_n k((X_i-h)/h)}{Ek((X_i-h)/h)} - 1\right| \xrightarrow{p} 0.$$

Proof. The first two displays follow from Lemma B.1 after noting that

$$var(m(W_i,\theta_n)k((X_i-x)/h)) \leq \overline{Y}^2\overline{k}^2\overline{f}_X B^{d_X}h^{d_X}$$

where $\overline{k}$ and $\overline{f}_X$ are bounds for k and f_X, and B is such that $k(u)=0$ whenever $\max_{1\leq j\leq d_X}|u_j| > B/2$, and similarly for $var(k((X_i-x)/h))$, and that $\sqrt{h^{d_X}}\sqrt{n}/\sqrt{\log n}\to\infty$ under these assumptions.

For the last display, note that, for x such that $\omega_j(\theta_n, x) > 0$ for some j, $Ek((X_i - x)/h) \geq \underline{f}_X h^{d_X} \int k(u)\, du$ for large enough n, where $\underline{f}_X$ is a lower bound for the density of X_i (which can be taken to be ε in Assumption 3.7). Thus,

$$\sup_{\{x|\omega_j(\theta_n,x)>0 \text{ some } j\}} \left| \frac{E_n k((X_i - h)/h)}{Ek((X_i - h)/h)} - 1 \right| \leq \sup_{x \in \mathbb{R}^{d_X}} \left| \frac{(E_n - E)k((X_i - h)/h)}{\underline{f}_X h^{d_X} \int k(u)\, du} \right|$$
$$= \sup_{x \in \mathbb{R}^{d_X}} \frac{\sqrt{n}}{\sqrt{h^{d_X} \log n}} |(E_n - E)k((X_i - h)/h)| \cdot \frac{\sqrt{h^{d_X} \log n}}{\sqrt{n} \underline{f}_X h^{d_X} \int k(u)\, du}.$$

The result then follows from the second display, since $\frac{\sqrt{\log n}}{\sqrt{nh^{d_X}}} \to 0$. □

Let

$$\tilde{T}_{n,p,(\hat{\sigma} \vee \sigma_n)^{-1},\mu}(\theta) = \left[\int_{h>0} \int_x \sum_{j=1}^{d_Y} \left| \frac{E_n m(W_i, \theta) k((X_i - x)/h)}{\sigma_j(\theta, x, h) \vee \sigma_n} \right|_-^p f_\mu(x, h)\, dx\, dh \right]^{1/p}$$

and let

$$\tilde{T}_{n,p,\text{kern}}(\theta) = \left[\int_x \sum_{j=1}^{d_Y} \left| \frac{E_n m(W_i, \theta) k((X_i - x)/h)}{Ek((X_i - x)/h)} \right|_-^p \omega_j(\theta, x)\, dx\, dh \right]^{1/p}.$$

The notation $\sigma_j(\theta, \tilde{x}, h)$ is used to denote $\sigma_j(\theta, g)$ where $g(x) = k((x - \tilde{x})/h)$.

Lemma D.2. *Under Assumptions 3.3, 3.4, 3.5 and 3.6,*

$$\sqrt{n} T_{n,p,(\hat{\sigma} \vee \sigma_n)^{-1},\mu}(\theta_n) = \sqrt{n} \tilde{T}_{n,p,(\hat{\sigma} \vee \sigma_n)^{-1},\mu}(\theta_n)(1 + o_P(1))$$

for any sequence $\theta_n \to \theta_0$. If Assumption 3.7 holds as well, then

$$(nh^{d_X})^{1/2} T_{n,p,kern}(\theta_n) = (nh^{d_X})^{1/2} \tilde{T}_{n,p,kern}(\theta_n)(1 + o_P(1))$$

for any sequence $\theta_n \to \theta_0$.

Proof. We have

$$|\sqrt{n} T_{n,p,(\hat{\sigma} \vee \sigma_n)^{-1},\mu}(\theta_n) - \sqrt{n} \tilde{T}_{n,p,(\hat{\sigma} \vee \sigma_n)^{-1},\mu}(\theta_n)| \leq \sqrt{n} \tilde{T}_{n,p,(\hat{\sigma} \vee \sigma_n)^{-1},\mu}(\theta) \cdot \sup_{x,j} \left| \frac{\sigma_j(\theta_n, x, h) \vee \sigma_n}{\hat{\sigma}_j(\theta_n, x, h) \vee \sigma_n} - 1 \right|.$$

Thus, the first display follows from Lemma B.2.

Similarly, for the second display,

$$
\begin{aligned}
&|(nh^{d_X})^{1/2}T_{n,p,\text{kern}}(\theta_n) - (nh^{d_X})^{1/2}\tilde{T}_{n,p,\text{kern}}(\theta_n)| \\
&\leq (nh^{d_X})^{1/2}\tilde{T}_{n,p,\text{kern}}(\theta_n) \cdot \sup_{\{x|\omega_j(\theta,x)>0 \text{ some } j\}} \left| \frac{Ek((X_i - x)/h)}{E_n k((X_i - x)/h)} - 1 \right|,
\end{aligned}
$$

and the result follows from Lemma D.1. $\square$

Let

$$
\tilde{\tilde{T}}_{n,p,(\hat{\sigma}\vee\sigma_n)^{-1},\mu}(\theta) = \left[\int_{h>0} \int_x \sum_{j=1}^{d_Y} \left| \frac{Em(W_i,\theta)k((X_i - x)/h)}{\sigma_j(\theta,x,h)\vee\sigma_n} \right|_-^p f_\mu(x,h)\, dx\, dh \right]^{1/p}
$$

and let

$$
\tilde{\tilde{T}}_{n,p,\text{kern}}(\theta) = \left[\int_x \sum_{j=1}^{d_Y} \left| \frac{Em(W_i,\theta)k((X_i - x)/h)}{Ek((X_i - x)/h)} \right|_-^p \omega_j(\theta,x)\, dx\, dh \right]^{1/p}.
$$

Also define

$$
\tilde{\tilde{T}}_{n,p,1,\mu}(\theta) = \left[\int_{h>0} \int_x \sum_{j=1}^{d_Y} |Em(W_i,\theta)k((X_i - x)/h)|_-^p f_\mu(x,h)\, dx\, dh \right]^{1/p}.
$$

Lemma D.3. *Under Assumptions 3.3, 3.4, 3.5 and 3.6,*

$$
\sqrt{n}\tilde{T}_{n,p,(\hat{\sigma}\vee\sigma_n)^{-1},\mu}(\theta_n) = \sqrt{n}\tilde{\tilde{T}}_{n,p,(\hat{\sigma}\vee\sigma_n)^{-1},\mu}(\theta_n) + o_P(1).
$$

and

$$
\sqrt{n}T_{n,p,1,\mu}(\theta_n) = \sqrt{n}\tilde{\tilde{T}}_{n,p,1,\mu}(\theta_n) + o_P(1).
$$

Proof. Let $\tilde{\sigma}_n \to 0$ be such that $\tilde{\sigma}_n\sqrt{n/\log n} \to \infty$ and $\tilde{\sigma}_n/\sigma_n \to 0$ (i.e. $\tilde{\sigma}_n$ is chosen to be much smaller than σ_n, but such that the assumptions still hold for $\tilde{\sigma}_n$). Note that

$$
\begin{aligned}
&\sqrt{n}|\tilde{\tilde{T}}_{n,p,(\hat{\sigma}\vee\sigma_n)^{-1},\mu}(\theta_n) - \tilde{T}_{n,p,(\hat{\sigma}\vee\sigma_n)^{-1},\mu}(\theta_n)| \\
&\leq \left[\int \int_{(x,h)\in\hat{\mathcal{G}}} \sum_{j=1}^{d_Y} \left| \sqrt{n}\frac{(E_n - E)m(W_i,\theta_n)k((X_i - x)/h)}{\sigma_j(\theta,x,h)\vee\sigma_n} \right|^p f_\mu(x,h)\, dx\, dh \right]^{1/p}
\end{aligned}
$$

where $\hat{\mathcal{G}} = \{(x,h) | Em(W_i,\theta_n)k((X_i - x)/h) < 0 \text{ or } E_n(W_i,\theta_n)k((X_i - x)/h) < 0\}$.

For any $\varepsilon > 0$, there exists an $\eta > 0$ such that, for $h > \varepsilon$ and large enough n,

$$Em_j(W_i,\theta_n)k((X_i - x)/h) \geq \eta Ek((X_i - x)/h) \geq \eta \cdot var[m_j(W_i,\theta_n)k((X_i - x)/h)] \cdot \frac{1}{\overline{k}\overline{Y}^2}$$

where the second inequality follows since

$$var[m_j(W_i,\theta_n)k((X_i - x)/h)] \leq \overline{Y}^2 E[k((X_i - x)/h)^2] \leq \overline{Y}^2 \overline{k} Ek((X_i - x)/h).$$

Thus, for large enough n we will have

$$\begin{aligned} &E_n m_j(W_i,\theta_n)k((X_i - x)/h) \\ &\geq (E_n - E)m_j(W_i,\theta_n)k((X_i - x)/h) + var[m_j(W_i,\theta_n)k((X_i - x)/h)] \cdot \frac{\eta}{\overline{k}\overline{Y}^2}, \end{aligned}$$

and the last line is positive for all (x,h) with $\sigma_j(\theta_n,x,h) \geq \tilde{\sigma}_n$ with probability approaching one by Lemma B.1.

From this and the fact that $Em(W_i,\theta_n)k((X_i - x)/h) \geq 0$ for all $h > \varepsilon$ for large enough n, it follows that $\hat{\mathcal{G}} \subseteq \{(x,h) | h \leq \varepsilon \text{ or } \sigma_j(\theta_n,x,h) < \tilde{\sigma}_n\}$ with probability approaching one. Note that

$$\begin{aligned} &E \int\int_{\{(x,h)|h\leq\varepsilon\}} \sum_{j=1}^{d_Y} \left| \frac{\sqrt{n}(E_n - E)m(W_i,\theta_n)k((X_i - x)/h)}{\sigma_j(\theta,x,h) \vee \sigma_n} \right|^p f_\mu(x,h)\, dx\, dh \\ &= \int\int_{\{(x,h)|h\leq\varepsilon\}} \sum_{j=1}^{d_Y} E \left| \frac{\sqrt{n}(E_n - E)m(W_i,\theta_n)k((X_i - x)/h)}{\sigma_j(\theta,x,h) \vee \sigma_n} \right|^p f_\mu(x,h)\, dx\, dh \end{aligned}$$

by Fubini's theorem, and this can be made arbitrarily small by making ε small by Lemma B.3 and Assumption 3.4. Similarly,

$$\begin{aligned} &E \int\int_{\{(x,h)|\sigma_j(\theta_n,x,h)<\tilde{\sigma}_n \text{ some } j\}} \sum_{j=1}^{d_Y} \left| \frac{\sqrt{n}(E_n - E)m(W_i,\theta_n)k((X_i - x)/h)}{\sigma_j(\theta,x,h) \vee \sigma_n} \right|^p f_\mu(x,h)\, dx\, dh \\ &\leq \mu(\mathbb{R}^{d_X} \times [0,\infty)) \cdot \sup_{\{(x,h,j)|\sigma_j(\theta_n,x,h)<\tilde{\sigma}_n\}} E \left| \frac{\sqrt{n}(E_n - E)m(W_i,\theta_n)k((X_i - x)/h)}{\sigma_j(\theta,x,h) \vee \sigma_n} \right|^p \\ &= \mu(\mathbb{R}^{d_X} \times [0,\infty)) \cdot \sup_{\{(x,h,j)|\sigma_j(\theta_n,x,h)<\tilde{\sigma}_n\}} E \left| \frac{\sqrt{n}(E_n - E)m(W_i,\theta_n)k((X_i - x)/h)}{\sigma_j(\theta,x,h) \vee \tilde{\sigma}_n} \right|^p \frac{\tilde{\sigma}_n}{\sigma_n}, \end{aligned}$$

which converges to zero by Lemma B.3. Using this and Markov's inequality, it follows that $\sqrt{n}|\tilde{\tilde{T}}_{n,p,(\hat\sigma\vee\sigma_n)^{-1},\mu}(\theta) - \tilde{T}_{n,p,(\hat\sigma\vee\sigma_n)^{-1},\mu}(\theta)|$ can be made arbitrarily small with probability approaching one by making ε small. This gives the first display of the lemma.

The second display follows by the same argument with σ_n set to the supremum of $\sigma_j(\theta, x, h)$ over x, h on the support of μ, θ in a neighborhood of θ_0 and all j. □

Lemma D.4. *Under Assumptions 3.3, 3.4, 3.5, 3.6 and 3.7,*

$$(nh^{d_X})^{1/2}\tilde{T}_{n,p,kern}(\theta_n) = (nh^{d_X})^{1/2}\tilde{\tilde{T}}_{n,p,kern}(\theta_n) + o_P(1).$$

Proof. For any $\varepsilon > 0$, there is an $\eta > 0$ such that $Em_j(W_i, \theta_n)k((X_i - x)/h) > \eta Ek((X_i - x)/h)$ for all $x \in \bar{\mathcal{X}}(\varepsilon)$ where $\bar{\mathcal{X}}(\varepsilon)$ is the set of x with $\|x - x_k\| \geq \varepsilon$ for all $k = 1, \ldots, \ell$ and $\omega_j(\theta_n, x) > 0$ for some j. Thus, arguing as in Lemma D.3 and using Lemma D.1, it follows that, with probability approaching one,

$$(nh^{d_X})^{1/2}|\tilde{T}_{n,p,\text{kern}}(\theta_n) - \tilde{\tilde{T}}_{n,p,\text{kern}}(\theta_n)|$$
$$\leq \left[\int_{x\notin\bar{\mathcal{X}}(\varepsilon)} \sum_{j=1}^{d_Y} \left| \frac{\sqrt{nh^{d_X}}(E_n - E)m_j(W_i, \theta_n)k((X_i - x)/h)}{Ek((X_i - x)/h)} \right|^p \omega_j(\theta_n, x)\, dx\right]^{1/p}.$$

Using Markov's inequality and Fubini's theorem along with the fact that $\int_{x\notin\bar{\mathcal{X}}(\varepsilon)} w_j(\theta_n x)\, dx$ can be made arbitrarily small by making ε small, the result follows so long as

$$E\left| \frac{\sqrt{nh^{d_X}}(E_n - E)m_j(W_i, \theta_n)k((X_i - x)/h)}{Ek((X_i - x)/h)} \right|^p$$

can be bounded uniformly over x such that $\omega_j(\theta_n, x) > 0$. But this follows from Lemma B.3, since, by Assumptions 3.3 and 3.7, for some $\delta > 0$, $Ek((X_i - x)/h) \geq \delta h^{d_X}$ for all x with $\omega_j(\theta_n, x) > 0$. □

For the following lemma, recall that $w_j(x_k) = (s_j^2(x_k, \theta_0)f_X(x_k)\int k(u)^2\, du)^{-1/2}$ and $s_j^2(x, \theta) = var(m(W_i, \theta)|X_i = x)$.

Lemma D.5. *Under Assumptions 3.3, 3.4, 3.5 and 3.6, for $k = 1, \ldots, \ell$*

$$\sup_{\|(x,h)-(x_k,0)\|\leq\varepsilon_n} \left|h^{-d_X/2}\sigma_j(\theta_n, x, h) - w_j(x_k)^{-1}\right| \to 0.$$

for any sequences $\varepsilon_n \to 0$ and $\theta_n \to \theta_0$.

Proof. By differentiability of the square root function at $w_j^{-2}(x_k)$, it suffices to show that $\sup_{\|(x,h)-(x_k,0)\|\le\varepsilon_n}\left|h^{-d_X}\sigma_j^2(\theta_n,x,h)-w_j^{-2}(x_k)\right|\to 0$. Note that

$$\begin{aligned}
&h^{-d_X}\sigma_j^2(\theta_n,x,h)=h^{-d_X}E[m(W_i,\theta_n)^2k((X_i-x)/h)^2]-h^{-d_X}\{E[m(W_i,\theta_n)k((X_i-x)/h)]\}^2\\
&=h^{-d_X}\int s_j^2(\tilde{x},\theta_n)k((\tilde{x}-x)/h)^2f_X(\tilde{x})\,d\tilde{x}\\
&+h^{-d_X}\int E[m(W_i,\theta_n)|X_i=\tilde{x}]^2k((\tilde{x}-x)/h)^2f_X(\tilde{x})\,d\tilde{x}\\
&-h^{-d_X}\left\{\int E[m(W_i,\theta_n)|X_i=\tilde{x}]k((\tilde{x}-x)/h)f_X(\tilde{x})\,d\tilde{x}\right\}^2.
\end{aligned}$$

By Assumption 3.3 and part (iii) of Assumption 3.5, the second term is bounded by a constant times $\sup_{\|(x,h)-(x_k,0)\|\le\varepsilon_n}E[m(W_i,\theta_n)|X_i=x]^2$, which converges to zero by continuity of $E[m(W_i,\theta)|X_i=x]$ at (θ_0,x_k). By Assumptions 3.3 and 3.5, the third term is bounded by a constant times $h^{-d_X}\cdot h^{2d_X}\le\varepsilon_n^{d_X}$ uniformly over (x,h) with $\|(x,h)-(x_k,0)\|\le\varepsilon_n$. Using a change of variables, the first term can be written as $\int s_j^2(x+uh,\theta_n)k(u)^2f_X(x+uh)\,du$, which converges to $w_j^{-2}(x_k)$ uniformly over $\|(x,h)-(x_k,0)\|\le\varepsilon_n$ by continuity of s_j and f_X, and by Assumption 3.3. □

Lemma D.6. *Suppose that Assumptions 3.3, 3.4, 3.5, 3.6 and 3.7 hold, and that* $\int k(u)\,du=1$. *Then*

$$\sup_{\|x-x_k\|\le\varepsilon}|h^{-d_X}Ek((X_i-x)/h)-f_X(x_k)|\to 0$$

as $h\to 0$ *and* $\varepsilon\to 0$ *for* $k=1,\ldots,\ell$.

Proof. We have

$$h^{-d_X}Ek((X_i-x)/h)=h^{-d_X}\int k((\tilde{x}-x)/h)f_X(\tilde{x})\,d\tilde{x}=\int k(u)f_X(x+uh)\,du,$$

and $\int k(u)\,du=1$ and $f_X(x+uh)$ converges to $f_X(x_k)$ uniformly over $\|x-x_k\|\le\varepsilon$ and u in the support of k as $\varepsilon\to 0$ and $h\to 0$. □

For notational convenience in the following lemmas, define, for (j,k) with $j\in J(k)$,

$$\tilde{\psi}_{j,k}(x-x_k)=\frac{\bar{m}_j(\theta_0,x)-\bar{m}_j(\theta_0,x_k)}{\|x-x_k\|^{\gamma(j,k)}}$$

so that

$$\sup_{\|x-x_k\|<\delta}\left|\tilde{\psi}_{j,k}(x-x_k)-\psi_{j,k}\left(\frac{x-x_k}{\|x-x_k\|}\right)\right|\to 0$$

under Assumption 3.5.

Lemma D.7. *Under Assumptions 3.3, 3.4, 3.5 and 3.6, for any $a\in\mathbb{R}^{d_\theta}$,*

$$r^{-[d_X+p(d_X+\gamma)+1]/\gamma}\int\int\sum_{j=1}^{d_Y}|Em_j(W_i,\theta_0+ra)k((X_i-\tilde{x})/h)|_-^p f_\mu(\tilde{x},h)\,d\tilde{x}\,dh$$
$$\overset{r\to 0}{\to}\sum_{k=1}^{\mathcal{X}_0}\sum_{j\in\tilde{J}(k)}\lambda_{bdd}(a,j,k,p).$$

Proof. For simplicity, assume that $\gamma(j,k)=\gamma$ for all j,k. The general result follows from applying the same arguments to show that areas of (x,h) near (j,k) with $\gamma(j,k)<\gamma$ do not matter asymptotically.

For C large enough, the integrand will be zero unless $\max\{\|\tilde{x}-x_k\|,h\}<Cr^{1/\gamma}$ for some k with $j\in J(k)$. Thus, it suffices to prove the lemma for, fixing (j,k) with $j\in J(k)$,

$$\int\int|Em_j(W_i,\theta_0+ra)k((X_i-\tilde{x})/h)|_-^p f_\mu(\tilde{x},h)\,d\tilde{x}\,dh$$
$$=\int\int\left|\int\bar{m}_j(\theta_0+ra,x)k((x-\tilde{x})/h)f_X(x)\,dx\right|_-^p f_\mu(\tilde{x},h)\,d\tilde{x}\,dh$$
$$=\int\int\left|\int[\|x-x_k\|^\gamma\tilde{\psi}_{j,k}(x-x_k)+\bar{m}_{\theta,j}(\theta^*(r),x)ra]k((x-\tilde{x})/h)f_X(x)\,dx\right|_-^p f_\mu(\tilde{x},h)\,d\tilde{x}\,dh$$

where the integrals are taken over $\|\tilde{x}-x_k\|<Cr^{1/\gamma},h<Cr^{1/\gamma}$ and $\theta^*(r)$ is between θ_0 and θ_0+ra (we suppress the dependence of $\theta^*(r)$ on x in the notation). Using the change of variables $u=(x-x_k)/r^{1/\gamma}$, $v=(x-x_k)/r^{1/\gamma}$, $\tilde{h}=h/r^{1/\gamma}$, this is equal to

$$\int\int\left|\int[\|r^{1/\gamma}u\|^\gamma\tilde{\psi}_{j,k}(r^{1/\gamma}u)+\bar{m}_{\theta,j}(\theta^*(r),x_k+r^{1/\gamma}u)ra]k((u-v)/\tilde{h})f_X(x_k+r^{1/\gamma}u)r^{d_X/\gamma}\,du\right|_-^p$$
$$f_\mu(x_k+r^{1/\gamma}v,r^{1/\gamma}\tilde{h})r^{d_X/\gamma}\,dvr^{1/\gamma}\,d\tilde{h}$$
$$=r^{[d_X+1+p(\gamma+d_X)]/\gamma}\int\int\left|\int[\|u\|^\gamma\tilde{\psi}_{j,k}(r^{1/\gamma}u)+\bar{m}_{\theta,j}(\theta^*(r),x_k+r^{1/\gamma}u)a]k((u-v)/\tilde{h})f_X(x_k+r^{1/\gamma}u)\,du\right|_-^p$$
$$f_\mu(x_k+r^{1/\gamma}v,r^{1/\gamma}\tilde{h})\,dv\,d\tilde{h}$$

where the integrals are taken over $\|v\| < C, \tilde{h} < C$. The result now follows from the dominated convergence theorem (here, and in subsequent results involving sequences of the form $\int |\int g_n(z,w)\, d\mu(z)|_-^p\, d\nu(w)$, the dominated convergence theorem is applied to the inner integral for each w, and again to the outer integral).

$\square$

Lemma D.8. *Under the conditions of Theorem 4.3, for any $a \in \mathbb{R}^{d_\theta}$,*

$$r^{-[d_X+p(d_X/2+\gamma)+1]/\gamma} \int\int \sum_{j=1}^{d_Y} |Em_j(W_i,\theta_0+ra)k((X_i-\tilde{x})/h)/(\sigma_j(\theta_0+ra,\tilde{x},h)\vee\sigma_n)|_-^p f_\mu(\tilde{x},h)\, d\tilde{x}\, dh$$
$$\leq \sum_{k=1}^{\mathcal{X}_0} \sum_{j\in\tilde{J}(k)} \lambda_{var}(a,j,k,p) + o(1)$$

for any $r = r_n \to 0$. If, in addition, $\sigma_n r_n^{-d_X/(2\gamma)} \to 0$, the above display will hold with the inequality replaced by equality.

Proof. As in the previous lemma, the following argument assumes, for simplicity, that $\gamma(j,k) = \gamma$ for all (j,k) with $j \in J(k)$. Let $\tilde{s}_j(r,\tilde{x},h) = \sigma_j(\theta_0+ra,\tilde{x},h)/h^{d_X/2}$. As before, for large enough C, the integrand will be zero unless $\max\{\|\tilde{x}-x_k\|, h\} < Cr^{1/\gamma}$ for some k with $j \in J(k)$. Thus, it suffices to prove the result for, fixing (j,k) with $j \in J(k)$,

$$\int\int |Em_j(W_i,\theta_0+ra)k((X_i-\tilde{x})/h)(h^{-d_X/2}\tilde{s}_j^{-1}(r,\tilde{x},h)\wedge\sigma_n^{-1})|_-^p f_\mu(\tilde{x},h)\, d\tilde{x}\, dh$$
$$= \int\int \left|\int [\|x-x_k\|^\gamma \tilde{\psi}_{j,k}(x-x_k) + \bar{m}_{\theta,j}(\theta^*(r),x)ra]\right.$$
$$\left. k((x-\tilde{x})/h)(h^{-d_X/2}\tilde{s}_j^{-1}(r,\tilde{x},h)\wedge\sigma_n^{-1})f_X(x)\, dx\right|_-^p f_\mu(\tilde{x},h)\, d\tilde{x}\, dh$$

where the integral is taken over $\|\tilde{x}-x_k\| < Cr^{1/\gamma}$, $h < Cr^{1/\gamma}$ and $\theta^*(r)$ is between θ_0 and θ_0+ra. Using the change of variables $u = (x-x_k)/r^{1/\gamma}$, $v = (\tilde{x}-x_k)/r^{1/\gamma}$,$\tilde{h} = h/r^{1/\gamma}$, this is equal to

$$\int\int\left|\int r[\|u\|^\gamma\tilde{\psi}_{j,k}(r^{1/\gamma}u)+\bar{m}_{\theta,j}(\theta^*(r),x_k+ur^{1/\gamma})a]k((u-v)/\tilde{h})\right.$$
$$\left.(((r^{1/\gamma}\tilde{h})^{-d_X/2}\tilde{s}_j^{-1}(r,x_k+vr^{1/\gamma},r^{1/\gamma}\tilde{h}))\wedge\sigma_n^{-1})f_X(x_k+ur^{1/\gamma})r^{d_X/\gamma}\,du\right|_-^p$$
$$f_\mu(x_k+vr^{1/\gamma},r^{1/\gamma}\tilde{h})r^{d_X/\gamma}\,dvr^{1/\gamma}\,d\tilde{h}$$
$$=r^{[p(\gamma+d_X/2)+d_X+1]/\gamma}\int\int\left|\int[\|u\|^\gamma\tilde{\psi}_{j,k}(r^{1/\gamma}u)+\bar{m}_{\theta,j}(\theta^*(r),x_k+ur^{1/\gamma})a]k((u-v)/\tilde{h})\right.$$
$$\left.((\tilde{h}^{-d_X/2}\tilde{s}_j^{-1}(r,x_k+vr^{1/\gamma},r^{1/\gamma}\tilde{h}))\wedge(r^{d_X/(2\gamma)}\sigma_n^{-1}))f_X(x_k+ur^{1/\gamma})\,du\right|_-^p f_\mu(x_k+vr^{1/\gamma},r^{1/\gamma}\tilde{h})\,dv\,d\tilde{h}.$$

where the integral is taken over $\|v\|<C$, $h<C$. By Lemma D.5 and the dominated convergence theorem, this converges to $\lambda_{var}(a,j,k,p)$ if $\sigma_n r_n^{-d_X/(2\gamma)}\to 0$. If $\sigma_n r_n^{-d_X/(2\gamma)}$ does not converge to zero, the above display is bounded from above by the same expression with σ_n^{-1} replaced by ∞.

□

Lemma D.9. *Under the conditions of Theorem 4.5, for any $a\in\mathbb{R}^{d_\theta}$,*

$$r^{-(\gamma p+d_X)/\gamma}\int\sum_{j=1}^{d_Y}|[Em_j(W_i,\theta_0+ra)k((X_i-x)/h)/Ek((X_i-x)/h)]\omega_j(\theta_0+ra,x)|_-^p\,dx$$
$$\to\sum_{k=1}^{|\mathcal{X}_0|}\sum_{j\in J(k)}\lambda_{kern}(a,c_{h,r},j,k,p)$$

as $r\to 0$ with $h/r^{1/\gamma}\to c_{h,r}$ for $c_{h,r}>0$. If the limit is zero for $(a,c_{h,r})$ in a neighborhood of the given values, the sequence will be exactly equal to zero for large enough r.

If $h/r^{1/\gamma}\to 0$, then, as $r\to 0$,

$$r^{-(\gamma p+d_X)/\gamma}\int\sum_{j=1}^{d_Y}|[Em_j(W_i,\theta_0+ra)k((X_i-x)/h)/Ek((X_i-x)/h)]\omega_j(\theta_0+ra,x)|_-^p\,dx$$
$$\to\sum_{k=1}^{|\mathcal{X}_0|}\sum_{j\in J(k)}\tilde{\lambda}_{kern}(a,j,k,p).$$

Proof. As before, this proof treats the case where $J(k)=\tilde{J}(k)$ for ease of exposition. As with the proofs of Lemmas D.7 and D.8, it suffices to prove the result for, fixing (j,k) with

$j \in J(k)$,

$$\int |[Em_j(W_i,\theta_0+ra)k((X_i-\tilde{x})/h)/Ek((X_i-\tilde{x})/h)]\omega_j(\theta_0+ra,\tilde{x})|_-^p \, d\tilde{x}$$
$$= \int \left| \int [\|x-x_k\|^\gamma \tilde{\psi}_{j,k}(x-x_k) + \bar{m}_{\theta,j}(\theta^*(r),x)ra]k((x-\tilde{x})/h)f_X(x)\, dx h^{-d_X} b(\tilde{x})\omega_j(\theta_0+ra,\tilde{x}) \right|_-^p \, d\tilde{x}$$

where the integral is over $\|\tilde{x}-x_k\| < Cr^{1/\gamma}$ and $b(\tilde{x}) \equiv h^{d_X}/Ek((X_i-\tilde{x})/h)$ converges to $(f_X(x_k))^{-1}$ uniformly over $\tilde{x}$ in any shrinking neighborhood of x_k by Lemma D.6. Let $\tilde{h} = h/r^{1/\gamma}$. By the change of variables $u = (x-x_k)/r^{1/\gamma}$, $v = (\tilde{x}-x_k)/r^{1/\gamma}$, the above display is equal to

$$\int \left| \int [\|ur^{1/\gamma}\|^\gamma \tilde{\psi}_{j,k}(ur^{1/\gamma}) + \bar{m}_{\theta,j}(\theta^*(r),x_k+ur^{1/\gamma})ra]k((u-v)/\tilde{h})f_X(x_k+ur^{1/\gamma})r^{d_X/\gamma}\, du \right.$$
$$\left. (r^{1/\gamma}\tilde{h})^{-d_X} b(x_k+vr^{1/\gamma})\omega_j(\theta_0+ra,x_k+r^{1/\gamma}v) \right|_-^p r^{d_X/\gamma}\, dv$$
$$= r^{p+d_X/\gamma} \int \left| \int [\|u\|^\gamma \tilde{\psi}_{j,k}(ur^{1/\gamma}) + \bar{m}_{\theta,j}(\theta^*(r),x_k+ur^{1/\gamma})a]k((u-v)/\tilde{h})f_X(x_k+ur^{1/\gamma})\, du \right.$$
$$\left. \tilde{h}^{-d_X} b(x_k+vr^{1/\gamma})\omega_j(\theta_0+ra,x_k+r^{1/\gamma}v) \right|_-^p \, dv \tag{15}$$

where the integral is over $v < C$. The first display of the lemma (the case where $h/r^{1/\gamma} \to c_{h,r}$ for $c_{h,r} > 0$) follows from this and the dominated convergence theorem.

To show that the sequence is exactly zero for small enough r when the limit is zero in a neighborhood of $(a, c_{h,r})$, note, that, if the limit is zero in a neighborhood of $(a, c_{h,r})$, we will have, for all $(\tilde{a}, \tilde{c}_{h,r})$ in this neighborhood and any v,

$$\int \left[\|u\|^\gamma \psi_{j,k}\left(\frac{u}{\|u\|}\right) + \bar{m}_{\theta,j}(\theta_0,x_k)\tilde{a} \right] k((u-v)/\tilde{c}_{h,r})\, du$$
$$= \int \left[\tilde{c}_{h,r}^\gamma \|\tilde{u}\|^\gamma \psi_{j,k}\left(\frac{u}{\|u\|}\right) + \bar{m}_{\theta,j}(\theta_0,x_k)\tilde{a} \right] k(\tilde{u}-\tilde{v})\, \tilde{c}_{h,r}^{d_X} d\tilde{u} \geq 0.$$

Evaluating this at $(\tilde{c}_{r,h}, \tilde{a})$ such that $\tilde{c}_{h,r}^\gamma \leq c_{h,r}^\gamma(1-\varepsilon)$ and (for the case where $\bar{m}_{\theta,j}(\theta_0,x_k)a$ is negative) $\bar{m}_{\theta,j}(\theta_0,x_k)\tilde{a} \leq (\bar{m}_{\theta,j}(\theta_0,x_k)a)(1+\varepsilon)$ shows that

$$\int \left[c_{h,r}^\gamma \|\tilde{u}\|^\gamma \psi_{j,k}\left(\frac{u}{\|u\|}\right) \cdot (1-\varepsilon) + (\bar{m}_{\theta,j}(\theta_0,x_k)a)(1+\varepsilon) \right] k(\tilde{u}-\tilde{v})\, d\tilde{u} \geq 0$$

for all v for some $\varepsilon > 0$. The above display is, for small enough r, a lower bound for the inner integral in (15) times a constant that does not depend on r, so that, for small enough

r, the inner integral in (15) will be nonnegative for all v and (15) will eventually be equal to zero.

For the case where $\tilde{h} = h/r^{1/\gamma} \to 0$, multiplying (15) by $r^{-(p+d_X/\gamma)}$ gives, after the change of variables $\tilde{u} = (u - v)/\tilde{h}$,

$$\int \left| \int [[\|\tilde{h}\tilde{u} + v\|^\gamma \tilde{\psi}_{j,k}((\tilde{h}\tilde{u} + v)r^{1/\gamma}) + \bar{m}_{\theta,j}(\theta^*(r), x_k + (\tilde{h}\tilde{u} + v)r^{1/\gamma})a]k(\tilde{u})f_X(x_k + (\tilde{u}\tilde{h} + v)r^{1/\gamma})\, d\tilde{u} \right.$$
$$\left. b(x_k + vr^{1/\gamma})\omega_j(\theta_0 + ra, x_k + r^{1/\gamma}v) \right|_-^p \, dv$$

which converges to

$$\int |[\|v\|^\gamma \psi_{j,k}(v/\|v\|) + \bar{m}_{\theta,j}(\theta_0, x_k)a]\omega_j(\theta_0, x_k)|_-^p \, dv$$

by the dominated convergence theorem, as required.

□

We are now ready for the proofs of the main results.

proof of Theorem 4.1. The result follows immediately from Lemmas D.3 and D.7 since $(n^{-\gamma/\{2[d_X+\gamma+(d_X+1)/p]\}})^{-[d_X+p(d_X+\gamma)+1]/(\gamma p)} = n^{1/2}$. □

proof of Theorem 4.3. The result follows immediately from Lemmas D.2, D.3 and D.8 since $(n^{-\gamma/\{2[d_X/2+\gamma+(d_X+1)/p]\}})^{-[d_X+p(d_X/2+\gamma)+1]/(\gamma p)} = n^{1/2}$. □

proof of Theorem 4.5. The result follows from Lemmas D.2, D.4 and D.9. Note that $(nh^{d_X})^{p/2}/(n^{1-d_X s})^{p/2}$ $c_h^{d_X p/2}$, and that, for the case where $s \geq 1/[2(\gamma + d_X/p + d_X/2)$,

$$(n^{-q})^{-(\gamma p+d_X)/(\gamma p)} = (n^{-(1-sd_X)/[2(1+d_X/(p\gamma))]})^{-(\gamma p+d_X)/(\gamma p)} = n^{(1-sd_X)/2}.$$

For the case where $s < 1/[2(\gamma + d_X/p + d_X/2)]$, it follows from Lemmas D.2, D.4 and D.9 that

$$n^{q(\gamma p+d_X)/(\gamma p)} T_n(\theta_0 + a_n) \overset{p}{\to} \left(\sum_{k=1}^{|\mathcal{X}_0|} \sum_{j \in J(k)} \lambda_{\text{kern}}(a, c_h, j, k, p) \right)^{1/p}$$

so that $(nh^{d_X})^{1/2} T_n(\theta_0 + a_n)$ will converge to ∞ in this case if the limit in the above display is strictly positive. If the limit in the above display is zero in a neighborhood of (a, c_h), it follows from Lemmas D.2 and D.4 that $(nh^{d_X})^{1/2}T_n(\theta_0 + a_n)$ is, up to $o_p(1)$, equal to a term that is zero for large enough n by Lemma D.9.

□