Supplement to "Asymptotically Exact Inference in Conditional Moment Inequality Models"

Timothy B. Armstrong
Yale University

December 10, 2014

This supplementary appendix contains auxiliary results and proofs for the main paper. Section A contains a proof of Theorem 3.1 in the case where $\ell = d_Y = 1$, which contains the main technical ideas of the general result, but requires less notation. Section B proposes an alternative way of obtaining critical values using the asymptotic distribution results in this paper. Section C contains proofs of the results from the main text and from Section B.

A Proof of Theorem 3.1 with a Single Contact Point

This section presents a proof of Theorem 3.1 in the special case where the conditional mean is minimized at a single point ($\ell = 1$, so that $\mathcal{X}_0 = \{x_1\}$) and the dimension of $m(W_i, \theta)$ (d_Y) is equal to one. Note that the dimension of X_i (d_X) is still allowed to be greater than one. This case contains the main technical aspects of the general proof, while requiring less notation. Section C.1 gives a proof in the general case. To focus on the main ideas, the proofs of some of the lemmas used in this section are omitted, with a reference to the correspinding lemma in Section C.1.

For notational convenience, let $Y_i = m(W_i, \theta)$ and $d = d_X$ throughout this section. Since $d_Y = \ell = 1$, we will always have $j = k = 1$ when referring to $\mathbb{G}_{P,x_k,j}$ and other quantities indexed by k and j, so I drop these subscripts and use the notation $\mathbb{G}_P(s,t)$ rather than $\mathbb{G}_{P,x_1,1}(s,t)$, etc.

The asymptotic distribution comes from the behavior of the objective function $E_n Y_i I(s < X_i < s + t)$ for s near x_1 and t near 0. The bulk of the proof involves showing that the objective function doesn't matter for (s,t) outside of a neighborhood of x_1 that shrinks at a fast enough rate. First, I derive the limiting distribution over such shrinking neighborhoods and the rate at which they shrink.

Theorem A.1. *Let $h_n = n^{-\alpha}$ for some $0 < \alpha < 1/d$. Let*

$$\mathbb{G}_n(s,t) = \frac{\sqrt{n}}{h_n^{d/2}}(E_n - E)Y_i I(h_n s < X_i - x_1 < h_n(s+t))$$

and

$$g_n(s,t) = \frac{1}{h_n^{d+2}} E Y_i I(h_n s < X_i - x_1 < h_n(s+t)).$$

Then, for any finite M, $\mathbb{G}_n(s,t) \overset{d}{\to} \mathbb{G}_P(s,t)$ taken as a random process on $\|(s,t)\| \le M$ with the supremum norm and $g_n(s,t) \to g_P(s,t)$ uniformly in $\|(s,t)\| \le M$ where $\mathbb{G}_P(s,t) = \mathbb{G}_{P,x_1,1}(s,t)$ and $g_P(s,t) = g_{P,x_1,1}(s,t)$ are defined as in Theorem 3.1 for m from 1 to ℓ.

Proof. The convergence in distribution in the first statement follows from verifying the conditions of Theorem 2.11.22 in van der Vaart and Wellner (1996). To derive the covariance kernel, note that

$$\begin{aligned}
&cov(\mathbb{G}_n(s,t), \mathbb{G}_n(s',t')) \\
&= h_n^{-d} E Y_i^2 I\left\{h_n(s \vee s') < X - x_1 < h_n\left[(s+t) \wedge (s'+t')\right]\right\} \\
&- h_n^{-d}\left\{E Y_i I\left[h_n s < X - x_1 < h_n(s+t)\right]\right\}\left\{E Y_i' I\left[h_n s' < X - x_1 < h_n(s'+t')\right]\right\}.
\end{aligned}$$

The second term goes to zero as $n \to \infty$. The first is equal to the claimed covariance kernel plus the error term

$$h_n^{-d} \int_{h_n(s\vee s') < x - x_1 < h_n[(s+t)\wedge(s'+t')]} \left[E(Y_i^2|X=x) f_X(x) - E(Y_i^2|X=x_1) f_X(x_1)\right] dx,$$

which is bounded by

$$\begin{aligned}
&\left\{\max_{\|x-x_1\| \le 2h_n M}\left[E(Y_i^2|X=x) f_X(x) - E(Y_i^2|X=x_1) f_X(x_1)\right]\right\} \\
&\times h_n^{-d} \int_{h_n(s\vee s') < x - x_1 < h_n[(s+t)\wedge(s'+t')]} dx \\
&= \left\{\max_{\|x-x_1\| \le 2h_n M}\left[E(Y_i^2|X=x) f_X(x) - E(Y_i^2|X=x_1) f_X(x_1)\right]\right\} \\
&\times \int_{(s\vee s') < x - x_1 < (s+t)\wedge(s'+t')} dx.
\end{aligned}$$

This goes to zero as $n \to \infty$ by continuity of $E(Y_i^2|X = x)$ and $f_X(x)$.

For the claim regarding $g_n(s,t)$, first note that the assumptions imply that the first derivative of $x \mapsto E(Y_i|X = x)$ at $x = x_1$ is 0, and that this function has a second order Taylor expansion:

$$E(Y_i|X = x) = \frac{1}{2}(x - x_1)'V(x_1)(x - x_1) + R_n(x)$$

where

$$R_n(x) = \frac{1}{2}(x - x_1)'V(x^*(x))(x - x_1) - \frac{1}{2}(x - x_1)'V(x_1)(x - x_1)$$

and $V(x^*)$ is the second derivative matrix evaluated at some $x^*(x)$ between x_1 and x.

We have

$$\begin{aligned} g_n(s,t) = &\frac{1}{2h_n^{d+2}} \int_{h_n s < x - x_1 < h_n(s+t)} (x - x_1)'V(x_1)(x - x_1) f_X(x_1)\, dx \\ &+ \frac{1}{2h_n^{d+2}} \int_{h_n s < x - x_1 < h_n(s+t)} (x - x_1)'V(x_1)(x - x_1)[f_X(x) - f_X(x_1)]\, dx \\ &+ \frac{1}{h_n^{d+2}} \int_{h_n s < x - x_1 < h_n(s+t)} R_n(x) f_X(x)\, dx. \end{aligned}$$

The first term is equal to $g_P(s,t)$ by a change of variable x to $h_n x + x_1$ in the integral. The second term is bounded by $g_P(s,t) \sup_{\|x - x_1\| \le 2h_n M} [f_X(x) - f_X(x_1)]/f_X(x_1)$, which goes to zero uniformly in $\|(s,t)\| \le M$ by continuity of f_X. The third term is equal to (using the same change of variables)

$$\frac{1}{2} \int_{s < x < s+t} [x'V(x^*(h_n x + x_1))x - x'V(x_1)x] f_X(h_n x + x_1)\, dx.$$

This is bounded by a constant times $\sup_{\|x\| \le M} |x'V(x^*(h_n x + x_1))x - x'V(x_1)x|$, which goes to zero as $n \to \infty$ by continuity of the second derivatives. $\square$

Thus, if we let h_n be such that $\sqrt{n}/h_n^{d/2} = 1/h_n^{d+2} \iff h_n = n^{-1/(d+4)}$ and scale up by $\sqrt{n}/h_n^{d/2} = 1/h_n^{d+2} = n^{(d+2)/(d+4)}$, we will have

$$n^{(d+2)/(d+4)} E_n Y_i I(h_n s < X - x_1 < h_n(s+t)) = \mathbb{G}_n(s,t) + g_n(s,t) \xrightarrow{d} \mathbb{G}_P(s,t) + g_P(s,t)$$

taken as stochastic processes in $\{\|(s,t)\| \le M\}$ with the supremum norm. From now on, let

$h_n = n^{-1/(d+4)}$ so that this will hold.

We would like to show that the infimum of $\mathbb{G}_n(s,t)$ process over all of $\mathbb{R}^{2d}$ converges to the infimum of the limiting process over all of $\mathbb{R}^{2d}$, but this does not follow immediately since we only have uniform convergence on compact sets. Another way of thinking about this problem is that convergence in distribution in $\{\|(s,t)\| \le M\}$ with the supremum norm for any M implies convergence in distribution in $\mathbb{R}^{2d}$ with the topology of uniform convergence on compact sets (see Kim and Pollard, 1990), but the infimum over all of $\mathbb{R}^{2d}$ is not a continuous mapping on this space since uniform convergence on all compact sets does not imply convergence of the infimum over all of $\mathbb{R}^{2d}$. To get the desired result, the following lemma will be useful. The idea is to show that values of (s,t) far away from zero won't matter for the limiting distribution, and then use convergence for fixed compact sets.

Lemma A.1. *Let $\mathbb{H}_n$ and $\mathbb{H}_P$ be random functions from $\mathbb{R}^d$ to $\mathbb{R}$ such that, (i) for all M, $\mathbb{H}_n \stackrel{d}{\to} \mathbb{H}_P$ when $\mathbb{H}_n$ and $\mathbb{H}_P$ are taken as random processes on $\{t \in \mathbb{R}^d | \|t\| \le M\}$ with the supremum norm, (ii) for all $r < 0$, $\varepsilon > 0$, there exists an M such that $P\left(\inf_{\|t\|>M} \mathbb{H}_P(t) \le r\right) < \varepsilon$ and an N such that $P\left(\inf_{\|t\|>M} \mathbb{H}_n(t) \le r\right) < \varepsilon$ for all $n \ge N$ and (iii) $\inf_t \mathbb{H}_n(t) \le 0$ and $\inf_t \mathbb{H}_P(t) \le 0$ with probability one. Then $\inf_{t\in\mathbb{R}^d} \mathbb{H}_n(t) \stackrel{d}{\to} \inf_{t\in\mathbb{R}^d} \mathbb{H}_P(t)$.*

Proof. It suffices to show that for all $r \in \mathbb{R}$, $\liminf_n P\left(\inf_{t\in\mathbb{R}^d} \mathbb{H}_n(t) < r\right) \ge P\left(\inf_{t\in\mathbb{R}^d} \mathbb{H}_P(t) < r\right)$ and $\limsup_n P\left(\inf_{t\in\mathbb{R}^d} \mathbb{H}_n(t) \le r\right) \le P\left(\inf_{t\in\mathbb{R}^d} \mathbb{H}_P(t) \le r\right)$ since, arguing along the lines of the Portmanteau Lemma, when r is a continuity point of the limiting distribution, we will have

$$P\left(\inf_{t\in\mathbb{R}^d} \mathbb{H}_P(t) \le r\right) = P\left(\inf_{t\in\mathbb{R}^d} \mathbb{H}_P(t) < r\right) \le \liminf_n P\left(\inf_{t\in\mathbb{R}^d} \mathbb{H}_n(t) < r\right)$$
$$\le \liminf_n P\left(\inf_{t\in\mathbb{R}^d} \mathbb{H}_n(t) \le r\right) \le \limsup_n P\left(\inf_{t\in\mathbb{R}^d} \mathbb{H}_n(t) \le r\right) \le P\left(\inf_{t\in\mathbb{R}^d} \mathbb{H}_P(t) \le r\right).$$

Given $\varepsilon > 0$, let M and N be as in the assumptions of the lemma. Then $P\left(\inf_{\|t\|\le M} \mathbb{H}_P(t) < r\right) + \varepsilon \ge P\left(\inf_{t\in\mathbb{R}^d} \mathbb{H}_P(t) < r\right)$ and, for $n \ge N$, $P\left(\inf_{\|t\|\le M} \mathbb{H}_n(t) \le r\right) + \varepsilon \ge P\left(\inf_{t\in\mathbb{R}} \mathbb{H}_n(t) \le r\right)$. Thus, by convergence in distribution of the infima over $\|t\| \le M$,

$$\liminf_n P\left(\inf_{t\in\mathbb{R}^d} \mathbb{H}_n(t) < r\right) \ge \liminf_n P\left(\inf_{\|t\|\le M} \mathbb{H}_n(t) < r\right) \ge P\left(\inf_{\|t\|\le M} \mathbb{H}_P(t) < r\right)$$
$$\ge P\left(\inf_{t\in\mathbb{R}^d} \mathbb{H}_P(t) < r\right) - \varepsilon$$

and

$$\limsup_n P\left(\inf_{t\in\mathbb{R}^d}\mathbb{H}_n(t)\le r\right)\le \limsup_n P\left(\inf_{\|t\|\le M}\mathbb{H}_n(t)\le r\right)+\varepsilon$$
$$\le P\left(\inf_{\|t\|\le M}\mathbb{H}_P(t)\le r\right)+\varepsilon\le P\left(\inf_{t\in\mathbb{R}^d}\mathbb{H}_P(t)\le r\right)+\varepsilon.$$

Since ε was arbitrary, this gives the desired result.

□

Part (i) of Lemma A.1 follows from Theorem A.1. Part (iii) follows since the processes involved are equal to zero when $t=0$. The main difficulty is in verifying part (ii).

The next two lemmas provide bounds that will be used to verify condition (ii) of Lemma A.1 for $\mathbb{G}_n(s,t)+g_n(s,t)$ and $\mathbb{G}_P(s,t)+g_P(s,t)$. To do this, the bounds in the lemmas are applied to sequences of sets of (s,t) where the norm of elements in the set increases with the sequence. The idea is similar to the "peeling" argument of, for example, Kim and Pollard (1990), but different arguments are required to deal with values of (s,t) for which, even though $\|s\|$ is large, $\prod_i t_i$ is small so that the objective function on average uses only a few observations, which may happen to be negative. To get bounds on the suprema of the limiting and finite sample processes where t may be small relative to s, the next two lemmas bound the supremum by a maximum over s in a finite grid of suprema over t with s fixed, and then use exponential bounds on suprema of the processes with fixed s.

Since some of the conditions only hold for X_i in a neighborhood of x_1 (e.g. conditions on the density of X_i), a different argument is used for $\mathbb{G}_n(s,t)+g_n(s,t)$ where $\|(s,t)\|\le\eta/h_n$ (which correspond to $E_nY_iI(s<X_i<s+t)$ with $\|(s-x_1,t)\|\le\eta$) and $\mathbb{G}_n(s,t)+g_n(s,t)$ with $\|(s,t)\|>\eta/h_n$. The proof for the ($\ell=d_Y=1$) case given in this section focuses on the former part of the argument (where $\|(s,t)\|\le\eta/h_n$), with the results used for the latter case (where $\|(s,t)\|>\eta/h_n$) stated here, but proved in Section C.1.

Lemma A.2. *For some $C>0$ that depends only on d, $f_X(x_1)$ and $E(Y_i^2|X=x_1)$, we have, for any $B\ge1$, $\varepsilon>0$, $w>0$,*

$$P\left(\sup_{\|(s,t)\|\le B,\prod_i t_i\le\varepsilon}|\mathbb{G}_P(s,t)|\ge w\right)\le 2\left\{3B[B^d/(\varepsilon\wedge1)]+2\right\}^{2d}\exp\left(-C\frac{w^2}{\varepsilon}\right)$$

for $\frac{w^2}{\varepsilon}$ greater than some constant that depends only on d, $f_X(x_m)$ and $E(Y_{i,j}^2|X=x_m)$.

Proof. We have, for any $s_0 \le s \le s+t \le s_0+t_0$,

$$\mathbb{G}_P(s,t) = \mathbb{G}_P(s_0, t+s-s_0)$$
$$+ \sum_{1\le j\le d} (-1)^j \sum_{1\le i_1<i_2<\ldots<i_j\le d}$$
$$\mathbb{G}_P(s_0, (t_1+s_1-s_{0,1}, \ldots, t_{i_1-1}+s_{i_1-1}-s_{0,i_1-1}, s_{i_1}-s_{0,i_1}, t_{i_1+1}+s_{i_1+1}-s_{0,i_1+1},$$
$$\ldots, t_{i_j-1}+s_{i_j-1}-s_{0,i_j-1}, s_{i_j}-s_{0,i_j}, t_{i_j+1}+s_{i_j+1}-s_{0,i_j+1}, \ldots, t_d+s_d-s_{0,d})).$$

Thus, since there are 2^d terms in the above display, each with absolute value bounded by $\sup_{t\le t_0} |\mathbb{G}_P(s_0,t)|$,

$$\sup_{s_0\le s\le s+t\le s_0+t_0} |\mathbb{G}_P(s,t)| \le 2^d \sup_{t\le t_0} |\mathbb{G}_P(s_0,t)| \stackrel{d}{=} 2^d \sup_{t\le t_0} |\mathbb{G}_P(0,t)|.$$

Let A be a grid of meshwidth $(\varepsilon\wedge 1)/B^d$ covering $[-B,2B]^d$. For any (s,t) with $\|(s,t)\| \le B$ and $\prod_i t_i \le \varepsilon$, there are s_0 and t_0 with $s_0, s_0+t_0 \in A$ such that $s_0 \le s \le s+t \le s_0+t_0$, and $\prod_i t_{0,i} \le \prod_i (t_i + (\varepsilon\wedge 1)/B^d) = \sum_{j=0}^d [(\varepsilon\wedge 1)/B^d]^j \sum_{I\in\{1,\ldots,d\},|I|=d-j} \prod_{i\in I} t_i \le \prod_i t_i + \sum_{j=1}^d [(\varepsilon\wedge 1)/B^d]^j \binom{d}{d-j} B^{d-j} \le \varepsilon + \varepsilon \sum_{j=1}^d B^{-dj}\binom{d}{d-j} B^{d-j} \le 2^d\varepsilon$. For this s_0, t_0, we will then have, by the above display, $|\mathbb{G}_P(s,t)| \le 2^d \sup_{t\le t_0} |\mathbb{G}_P(s_0,t)|$.

This gives

$$\sup_{\|(s,t)\|\le B, \prod_i t_i \le \varepsilon} |\mathbb{G}_P(s,t)| \le 2^d \max_{s_0, s_0+t_0\in A, \prod_i t_{0,i}\le 2^d\varepsilon} \sup_{t\le t_0} |\mathbb{G}_P(s_0,t)|,$$

so that

$$P\left(\sup_{\|(s,t)\|\le B, \prod_i t_i\le\varepsilon} |\mathbb{G}_P(s,t)| \ge w\right) \le |A|^2 \max_{s_0,s_0+t_0\in A, \prod_i t_{0,i}\le 2^d\varepsilon} P\left(2^d \sup_{t\le t_0} |\mathbb{G}_P(s_0,t)| \ge w\right)$$
$$= |A|^2 \max_{s_0,s_0+t_0\in A, \prod_i t_{0,i}\le 2^d\varepsilon} P\left(2^d \sup_{t\le 1}\left(\prod_i t_{0,i}\right)^{1/2} |\mathbb{G}_P(0,t)| \ge w\right)$$
$$\le |A|^2 P\left(\sup_{t\le 1} |\mathbb{G}_P(0,t)| \ge \frac{w}{2^d 2^{d/2}\varepsilon^{1/2}}\right).$$

The result then follows using the fact that $|A| \le \left\{3B[B^d/(\varepsilon\wedge 1)]+2\right\}^d$ and using Theorem 2.1 (p.43) in Adler (1990) to bound the probability in the last line of the display (the theorem in Adler (1990) shows that the probability in the above display is bounded by $2\exp(-K_1 w^2/\varepsilon + K_2 w/\varepsilon^{1/2} + K_3)$ for some constants K_1, K_2, and K_3 with $K_1 > 0$ that depend only on d, $f_X(x_m)$ and $E(Y_{i,j}^2|X = x_m)$ and this expression is less than $2\exp(-(K_1/2)w^2/\varepsilon)$ for w^2/ε greater than some constant that depends only on K_1, K_2, and K_3).

□

Lemma A.3. *For some $C > 0$ that depends only on the distribution of (X, Y) and some $\eta > 0$, we have, for any $1 \le B \le h_n^{-1}\eta$, $w > 0$ and $\varepsilon \ge n^{-4/(d+4)}(1 + \log n)^2$,*

$$P\left(\sup_{\|(s,t)\|\le B, \prod_i t_i \le \varepsilon} |\mathbb{G}_n(s,t)| \ge w\right) \le 2\left\{3B[B^d/(\varepsilon \wedge 1)] + 2\right\}^{2d} \exp\left(-C\frac{w}{\varepsilon^{1/2}}\right).$$

Proof. By the same argument as in the previous lemma with $\mathbb{G}$ replaced by $\mathbb{G}_n$, we have

$$\sup_{s_0 \le s \le s+t \le s_0+t_0} |\mathbb{G}_n(s,t)| \le 2^d \sup_{t \le t_0} |\mathbb{G}_n(s_0,t)|.$$

As in the previous lemma, let A be a grid of meshwidth $(\varepsilon \wedge 1)/B^d$ covering $[-B, 2B]^d$. Arguing as in the previous lemma, we have, for any (s,t) with $\|(s,t)\| \le B$ and $\prod_i t_i \le \varepsilon$, there exists some s_0, t_0 with $s_0, s_0 + t_0 \in A$ such that $\Pi_i t_{0,i} \le 2^d\varepsilon$ and $|\mathbb{G}_n(s,t)| \le 2^d \sup_{t \le t_0} |\mathbb{G}_n(s_0,t)|$. Thus,

$$\begin{aligned}
&\sup_{\|(s,t)\|\le B, \prod_i t_i \le \varepsilon} |\mathbb{G}_n(s,t)| \le 2^d \max_{s_0, s_0+t_0 \in A, \prod_i t_{0,i} \le 2^d\varepsilon} \sup_{t \le t_0} |\mathbb{G}_n(s_0,t)| \\
&= 2^d \max_{s_0, s_0+t_0 \in A, \prod_i t_{0,i} \le 2^d\varepsilon} \sup_{t \le t_0} \frac{\sqrt{n}}{h_n^{d/2}} |(E_n - E)Y_{i,j} I(h_n s_0 \le X_i - x_m \le h_n(s_0 + t))|.
\end{aligned}$$

This gives

$$\begin{aligned}
&P\left(\sup_{\|(s,t)\|\le B, \prod_i t_i \le 2^d\varepsilon} |\mathbb{G}_n(s,t)| \ge w\right) \\
&\le |A|^2 \max_{s_0, s_0+t_0 \in A, \prod_i t_{0,i} \le 2^d\varepsilon} P\left(2^d \sup_{t \le t_0} \frac{\sqrt{n}}{h_n^{d/2}} |(E_n - E)Y_{i,j} I(h_n s_0 \le X_i - x_m \le h_n(s_0 + t))| \ge w\right).
\end{aligned}$$

We have, for some universal constant K and all n with $\varepsilon \ge n^{-4/(d+4)}(1 + \log n)^2$, letting $\mathcal{F}_n = \{(x,y) \mapsto yI(h_n s_0 \le x - x_m \le h_n(s_0 + t)) | t \le t_0\}$ and defining $\|\cdot\|_{P,\psi_1}$ to be the Orlicz norm defined on p.90 of van der Vaart and Wellner (1996) for $\psi_1(x) = \exp(x) - 1$,

$$
\begin{aligned}
&\|2^d \sup_{f\in\mathcal{F}_n} |\sqrt{n}(E_n - E)f(X_i, Y_i)|\|_{P,\psi_1} \\
&\le K \left[E \sup_{f\in\mathcal{F}_n} |\sqrt{n}(E_n - E)f(X_i, Y_i)| + n^{-1/2}(1 + \log n)\| |Y_i| I(h_n s_0 \le X_i - x_1 \le h_n(s_0 + t_0))\|_{P,\psi_1} \right] \\
&\le K \left[J(1, \mathcal{F}_n, L^2) \left\{ E[|Y_i| I(h_n s_0 < X_i - x_1 < h_n(s_0 + t_0))]^2 \right\}^{1/2} + n^{-1/2}(1 + \log n)\|Y\|_{P,\psi_1} \right] \\
&\le K \left[J(1, \mathcal{F}_n, L^2) \overline{f}^{1/2} \overline{Y} h_n^{d/2} 2^{d/2} \varepsilon^{1/2} + n^{-1/2}(1 + \log n)\|Y_i\|_{P,\psi_1} \right] \\
&\le K \left[J(1, \mathcal{F}_n, L^2) \overline{f}^{1/2} \overline{Y} 2^{d/2} + \|Y_i\|_{P,\psi_1} \right] h_n^{d/2} \varepsilon^{1/2}.
\end{aligned}
$$

The first inequality follows by Theorem 2.14.5 in van der Vaart and Wellner (1996). The second uses Theorem 2.14.1 in van der Vaart and Wellner (1996). The fourth inequality uses the fact that $h_n^{d/2}\varepsilon^{1/2} = n^{-d/[2(d+4)]}\varepsilon^{1/2} \ge n^{-1/2}(1 + \log n)$ once $\varepsilon^{1/2} \ge n^{-1/2+d/[2(d+4)]}(1 + \log n) = n^{-2/(d+4)}(1 + \log n)$. Since each $\mathcal{F}_n$ is contained in the larger class $\mathcal{F} \equiv \{(x, y) \mapsto y_j I(s < x - x_1 < s + t) | (s, t) \in \mathbb{R}^{2d}\}$, we can replace $\mathcal{F}_n$ by $\mathcal{F}$ on the last line of this display. Since $J(1, \mathcal{F}, L^2)$ and $\|Y_i\|_{\psi_1}$ are finite ($\mathcal{F}$ is a VC class and Y_i is bounded), the bound is equal to $C^{-1}\varepsilon^{1/2}h_n^{d/2}$ for a constant C that depends only on the distribution of (X_i, Y_i).

This bound along with Lemma 8.1 in Kosorok (2008) implies

$$
\begin{aligned}
&P\left(2^d \sup_{t\le t_0} \frac{\sqrt{n}}{h_n^{d/2}} |(E_n - E)Y_i I(h_n s_0 \le X_i - x_1 \le h_n(s_0 + t))| \ge w \right) \\
&= P\left(2^d \sup_{f\in\mathcal{F}_n} |\sqrt{n}(E_n - E)f(X_i, Y_i)| \ge w h_n^{d/2} \right) \\
&\le 2\exp\left(-\frac{w h_n^{d/2}}{\|2^d \sup_{f\in\mathcal{F}_n} |\sqrt{n}(E_n - E)f(X_i, Y_i)|\|_{P,\psi_1}} \right) \\
&\le 2\exp\left(-\frac{w h_n^{d/2}}{C^{-1} h_n^{d/2} \varepsilon^{1/2}} \right) = 2\exp\left(-Cw/\varepsilon^{1/2}\right).
\end{aligned}
$$

The result follows using this and the fact that $|A| \le \left\{3B[B^d/(\varepsilon \wedge 1)] + 2\right\}^d$. □

The following theorem verifies the part of condition (ii) of Lemma A.1 concerning the limiting process $\mathbb{G}_P(s, t) + g_P(s, t)$.

Theorem A.2. *For any $r < 0$, $\varepsilon > 0$ there exists an M such that*

$$P\left(\inf_{\|(s,t)\|>M} \mathbb{G}_P(s,t) + g_P(s,t) \le r\right) < \varepsilon.$$

Proof. Let $S_k = \{k \le \|(s,t)\| \le k+1\}$ and let $S_k^L = S_k \cap \{\prod_i t_i \le (k+1)^{-\delta}\}$ for some fixed δ. By Lemma A.2,

$$P\left(\inf_{S_k^L} \mathbb{G}_P(s,t) + g_P(s,t) \le r\right) \le P\left(\sup_{S_k^L} |\mathbb{G}_P(s,t)| \ge |r|\right)$$
$$\le 2\left\{3(k+1)[(k+1)^d/k^{-\delta}] + 2\right\}^{2d} \exp\left(-Cr^2(k+1)^\delta\right)$$

for k large enough where C depends only on d. This bound is summable over k.

For any α and β with $\alpha < \beta$, let $S_k^{\alpha,\beta} = S_k \cap \{(k+1)^\alpha < \prod_i t_i \le (k+1)^\beta\}$. We have, for some $C_1 > 0$ that depends only on d and $V(x_1)$, $g(s,t) \ge C_1\|(s,t)\|^2 \prod_i t_i$. (To see this, note that $g(s,t)$ is greater than or equal to a constant times $\int_{s_1}^{s_1+t_1} \cdots \int_{s_d}^{s_d+t_d} \|x\|^2 dx_d \cdots dx_1 = \left(\Pi_{i=1}^d t_i\right) \sum_{i=1}^d (s_i^2 + t_i^2/3 + s_i t_i)$, and the sum can be bounded below by a constant times $\|(s,t)\|^2$ by minimizing over s_i for fixed t_i using calculus. The claimed expression for the integral follows from evaluating the inner integral to get an expression involving the integral for $d-1$, and then using induction.) Using this and Lemma A.2,

$$P\left(\inf_{S_k^{\alpha,\beta}} \mathbb{G}_P(s,t) + g_P(s,t) \le r\right) \le P\left(\sup_{S_k^{\alpha,\beta}} |\mathbb{G}_P(s,t)| \ge C_1 k^{2+\alpha}\right)$$
$$\le 2\left\{3(k+1)[(k+1)^d/((k+1)^\beta \wedge 1)] + 2\right\}^{2d} \exp\left(-CC_1^2 \frac{k^{4+2\alpha}}{(k+1)^\beta}\right).$$

This is summable over k if $4 + 2\alpha - \beta > 0$.

Now, note that, since $\prod_i t_i \le (k+1)^d$ on S_k, we have, for any $-\delta < \alpha_1 < \alpha_2 < \ldots < \alpha_{\ell-1} < \alpha_\ell = d$, $S_k = S_k^L \cup S_k^{-\delta,\alpha_1} \cup S_k^{\alpha_1,\alpha_2} \cup \ldots \cup S_k^{\alpha_{\ell-1},\alpha_\ell}$. If we choose $\delta < 3/2$ and $\alpha_i = i$ for $i \in \{1,\ldots,d\}$, the arguments above will show that the probability of the infimum being less than or equal to r over S_k^L, $S_k^{-\delta,\alpha_1}$ and each $S_k^{\alpha_i,\alpha_{i+1}}$ is summable over k, so that $P\left(\inf_{S_k} \mathbb{G}(s,t) + g(s,t) \le r\right)$ is summable over k, so setting M so that the tail of this sum past M is less than ε gives the desired result. □

The following theorem verifies condition (ii) of Lemma A.1 for the sequence of finite sample processes $\mathbb{G}_n(s,t) + g_n(s,t)$ with $\eta/h_n \ge \|(s,t)\|$. As explained above, the case where $\eta/h_n \le \|(s,t)\|$ is handled by a separate argument.

Theorem A.3. *There exists an $\eta > 0$ such that for any $r < 0$, $\varepsilon > 0$, there exists an M and N such that, for all $n \geq N$,*

$$P\left(\inf_{M<\|(s,t)\|\leq\eta/h_n} \mathbb{G}_n(s,t) + g_n(s,t) \leq r\right) < \varepsilon.$$

Proof. Let η be small enough that the assumptions hold for $\|x - x_m\| \leq 2\eta$ and that, for some constant C_2, $E(Y_i|X_i = x) \geq C_2\|x - x_1\|^2$ for $\|x - x_1\| \leq 2\eta$. This implies that, for $\|(s,t)\| \leq h_n^{-1}\eta$,

$$\begin{aligned}
g_n(s,t) &\geq \frac{C_2}{h_n^{d+2}} \int_{h_n s<x-x_1<h_n(s+t)} \|x - x_1\|^2 f_X(x)\, dx \\
&\geq \frac{C_2\underline{f}}{h_n^{d+2}} \int_{h_n s<x-x_1<h_n(s+t)} \|x - x_1\|^2\, dx = C_2\underline{f} \int_{s<x<s+t} \|x\|^2\, dx_d \cdots dx_1 \geq C_3\|(s,t)\|^2 \prod_i t_i
\end{aligned}$$

where C_3 is a constant that depends only on $\underline{f}$ and d and the last inequality follows from bounding the integral as explained in the proof of the previous theorem.

As in the proof of the previous theorem, let $S_k = \{k \leq \|(s,t)\| \leq k+1\}$ and let $S_k^L = S_k \cap \{\prod_i t_i \leq (k+1)^{-\delta}\}$ for some fixed δ. We have, using Lemma C.3,

$$\begin{aligned}
P\left(\inf_{S_k^L} \mathbb{G}_n(s,t) + g_n(s,t) \leq r\right) &\leq P\left(\sup_{S_k^L} |\mathbb{G}_n(s,t)| \geq |r|\right) \\
&\leq 2\left\{3(k+1)[(k+1)^d/k^{-\delta}] + 2\right\}^{2d} \exp\left(-C\frac{|r|}{(k+1)^{-\delta/2}}\right)
\end{aligned}$$

for $(k+1)^{-\delta} \geq n^{-4/(d+4)}(1+\log n)^2 \iff k+1 \leq n^{4/[\delta(d+4)]}(1+\log n)^{-2/\delta}$ so, if $\delta < 4$, this will hold eventually for all $(k+1) \leq h_n^{-1}\eta$ (once $h_n^{-1}\eta \leq n^{4/[\delta(d+4)]}(1+\log n)^{-2/\delta} \iff \eta \leq n^{(4/\delta-1)/(d+4)}(1+\log n)^{-2/\delta}$). The bound is summable over k for any $\delta > 0$.

Again following the proof of the previous theorem, for $\alpha < \beta$, define $S_k^{\alpha,\beta} = S_k \cap \{(k+1)^\alpha < \prod_i t_i \leq (k+1)^\beta\}$. We have, again using Lemma C.3,

$$\begin{aligned}
P\left(\inf_{S_k^{\alpha,\beta}} \mathbb{G}_n(s,t) + g_n(s,t) \leq r\right) &\leq P\left(\sup_{S_k^{\alpha,\beta}} |\mathbb{G}_n(s,t)| \geq C_3 k^{2+\alpha}\right) \\
&\leq 2\left\{3(k+1)[(k+1)^d/(k^\alpha \wedge 1)] + 2\right\}^{2d} \exp\left(-C\frac{C_3 k^{2+\alpha}}{(k+1)^{\beta/2}}\right)
\end{aligned}$$

for $(k+1)^\beta \geq n^{-4/(d+4)}$ (which will hold once the same inequality holds for δ for $-\delta < \beta$)

and $k+1 \leq h_n^{-1}\eta$. The bound is summable over k for any α, β with $4+2\alpha-\beta > 0$.

Thus, noting as in the previous theorem that, for any $-\delta < \alpha_1 < \alpha_2 < \ldots < \alpha_{\ell-1} < \alpha_\ell = d$, $S_k = S_k^L \cup S_k^{-\delta,\alpha_1} \cup S_k^{\alpha_1,\alpha_2} \cup \ldots \cup S_k^{\alpha_{\ell-1},\alpha_\ell}$, if we choose $\delta < 3/2$ and $\alpha_i = i$ for $i \in \{1, \ldots, d\}$ the probability of the infimum being less than or equal to r over the sets indexed by k for any $k \leq h_n^{-1}\eta$ is bounded uniformly in n by a sequence that is summable over k (once $\eta \leq n^{(4/\delta-1)/(d+4)}(1+\log n)^{-2/\delta}$). Thus, if we choose M such that the tail of this sum past M is less than ε and let N be large enough so that $\eta \leq N^{(4/\delta-1)/(d+4)}(1+\log N)^{-2/\delta}$, we will have the desired result.

□

To complete the proof, we need to show that

$$\inf_{\|(s,t)\|\geq \eta h_n^{-1}} \mathbb{G}_n(s,t) + g_n(s,t) = n^{(d_X+2)/(d_X+4)} \inf_{\|(s-x_1,t)\|\geq\eta} E_n Y_i I(s < X_i < s+t) \stackrel{p}{\to} 0.$$

Since some of the conditions of the theorem may not hold for x outside of a neighborhood of x_1 (e.g. X_i having a density bounded away from zero and infinity), we need a slightly different argument to show this. The result follows from the following lemmas, which are proved in the proof of the general result in Section C.1.

Lemma A.4. *Under Assumptions 3.1 and 3.2, for any $\eta > 0$, there exists some $\underline{B} > 0$ such that $EY_iI(s < X_i < s+t) \geq \underline{B}P(s < X_i < s+t)$ for all (s,t) with $\|(s-x_0,t)\| > \eta$.*

Proof. See Lemma C.4 in Section C.1. □

Lemma A.5. *Let S be any set in $\mathbb{R}^{2d}$ such that, for some $\underline{\mu} > 0$ and all $(s,t) \in S$, $EY_iI(s < X_i < s+t) \geq \underline{\mu}P(s < X_i < s+t)$. Then, under Assumption 3.2, for any sequence $a_n \to \infty$ and $r < 0$,*

$$\inf_{(s,t)\in S} \frac{n}{a_n \log n} E_n Y_{i,j} I(s < X_i < s+t) > r$$

with probability approaching 1.

Proof. See Lemma C.5 in Section C.1. □

By Lemma A.4, $\{(s,t)| \|(s-x_1,t)\| > \eta\}$ satisfies the conditions of Lemma A.5, so $E_nY_{i,j}I(s < X_i < s+t)$ converges to zero at a $n/(a_n \log n)$ rate for any $a_n \to \infty$, which can be made faster than the $n^{(d+2)/(d+4)}$ rate needed for the result. This completes the proof of Theorem 3.1 for the $\ell = d_Y = 1$ case.

B Alternative Method for Estimation of the Asymptotic Distribution

This section of the appendix describes a method for estimating the asymptotic distribution by estimating the unknown quantities that determine the distribution. This method can be used as an alternative to the subsampling based method described in the main text. Section B.1 shows how the asymptotic distribution can be estimated when Assumption 3.1 is known to hold, with known contact points $\{x_1, \ldots, x_\ell\}$. Section B.2 embeds this estimate in a procedure with a pre-test for Assumption 3.1 and estimation of the contact points. Proofs are given in Section C.6.

B.1 Estimation of the Asymptotic Distribution Under Assumption 3.1

As an alternative to subsampling based estimates, note that the asymptotic distribution in Theorem 3.1 depends on the underlying distribution only through the set $\mathcal{X}_0$ and, for points x_k in $\mathcal{X}_0$, the density $f_X(x_k)$, the conditional second moment matrix $E(m_{J(k)}(W_i,\theta)m_{J(k)}(W_i,\theta)'|X = x_k)$, and the second derivative matrix $V(x_k)$ of the conditional mean. Thus, with consistent estimates of these objects, we can estimate the distribution in Theorem 3.1 by replacing these objects with their consistent estimates and simulating from the corresponding distribution.

In order to accommodate different methods of estimating $f_X(x_k)$, $E(m_{J(k)}(W_i,\theta)m_{J(k)}(W_i,\theta)'|X = x_k)$, and $V(x_k)$, I state the consistency of these estimators as a high level condition, and show that the procedure works as long as these estimators are consistent. Since these objects only appear as $E(m_{J(k)}(W_i,\theta)m_{J(k)}(W_i,\theta)'|X = x_k)f_X(x_0)$ and $f_X(x_k)V(x_k)$ in the asymptotic distribution, we actually only need consistent estimates of these objects.

Assumption B.1. *The estimates $\hat{M}_k(x_k)$, $\hat{f}_X(x_k)$, and $\hat{V}(x_k)$ satisfy $\hat{f}_X(x_k)\hat{V}(x_k) \xrightarrow{p} f_X(x_k)V(x_k)$ and $\hat{M}_k(x_k)\hat{f}_X(x_k) \xrightarrow{p} E(m_{J(k)}(W_i,\theta)m_{J(k)}(W_i,\theta)'|X = x_k)f_X(x_k)$.*

For k from 1 to ℓ, let $\hat{\mathbb{G}}_{P,x_k}(s,t)$ and $\hat{g}_{P,x_k}(s,t)$ be the random process and mean function defined in the same way as $\mathbb{G}_{P,x_k}(s,t)$ and $g_{P,x_k}(s,t)$, but with the estimated quantities replacing the true quantities. We estimate the distribution of Z defined to have jth element

$$Z_j = \min_{m \text{ s.t. } j\in J(k)} \inf_{(s,t)\in\mathbb{R}^{2d}} \mathbb{G}_{P,x_k,j}(s,t) + g_{P,x_k,j}(s,t)$$

using the distribution of $\hat{Z}$ defined to have jth element

$$\hat{Z}_j = \min_{k \text{ s.t. } j \in J(k)} \inf_{\|(s,t)\| \leq B_n} \hat{\mathbb{G}}_{P,x_k,j}(s,t) + \hat{g}_{P,x_k,j}(s,t)$$

for some sequence B_n going to infinity. The convergence of the distribution $\hat{Z}$ to the distribution of Z is in the sense of conditional weak convergence in probability often used in proofs of the validity of the bootstrap (see, for example, Lehmann and Romano, 2005). From this, it follows that tests that replace the quantiles of $S(Z)$ with the quantiles of $S(\hat{Z})$ are asymptotically exact under the conditions that guarantee the continuity of the limiting distribution.

Theorem B.1. *Under Assumption B.1, $\rho(\hat{Z}, Z) \stackrel{p}{\to} 0$ where ρ is any metric on probability distributions that metrizes weak convergence.*

Corollary B.1. *Let $\hat{q}_{1-\alpha}$ be the $1-\alpha$ quantile of $S(\hat{Z})$. Then, under Assumptions 3.1, 3.2, 3.3, 4.1, 4.2, and B.1, the test that rejects when $n^{(d_X+2)/(d_X+4)} S(T_n(\theta)) > \hat{q}_{1-\alpha}$ and fails to reject otherwise is an asymptotically exact level α test.*

If the set $\mathcal{X}_0$ is known, the quantities needed to compute $\hat{Z}$ can be estimated consistently using standard methods for nonparametric estimation of densities, conditional moments, and their derivatives. However, typically $\mathcal{X}_0$ is not known, and the researcher will not even want to impose that this set is finite. In Section B.2, I propose methods for testing Assumption 3.1 and estimating the set $\mathcal{X}_0$ under weaker conditions on the smoothness of the conditional mean. These conditions allow for both the $n^{(d_X+2)/(d_X+4)}$ asymptotics that arise from Assumption 3.1 and the $\sqrt{n}$ asymptotics that arise from a positive probability contact set.

B.2 Pretest and Estimation with Unknown Contact Points

I make the following assumptions on the conditional mean and the distribution of X_i. These conditions are used to estimate the second derivatives of $\bar{m}(\theta, x) = E(m_j(W_i, \theta)|X_i = x)$, and the results are stated for local polynomial estimates. The conditions and results here are from Ichimura and Todd (2007). Other nonparametric estimators of conditional means and their derivatives and conditions for uniform convergence of such estimators could be used instead. The results in this section related to testing Assumption 3.1 are stated for $m_j(W_i, \theta)$ for a fixed index j. The consistency of a procedure that combines these tests for each j then follows from the consistency of the test for each j.

Assumption B.2. *The third derivatives of $\bar{m}_j(\theta, x)$ with respect to x are Lipschitz continuous and uniformly bounded.*

Assumption B.3. *X_i has a uniformly continuous density f_X such that, for some compact set $D \in \mathbb{R}^d$, $\inf_{x \in D} f_X(x) > 0$, and $E(m_j(W_i, \theta)|X_i)$ is bounded away from zero outside of D.*

Assumption B.4. *The conditional density of X_i given $m_j(W_i, \theta)$ exists and is uniformly bounded.*

Note that Assumption B.4 is on the density of X_i given $m_j(W_i, \theta)$, and not the other way around, so that, for example, count data for the dependent variable in an interval regression is okay.

Let $\mathcal{X}_0^j$ be the set of minimizers of $\bar{m}_j(\theta, x)$ if this function is less than or equal to 0 for some x and the empty set otherwise. In order to test Assumption 3.1, I first note that, if the conditional mean is smooth, the positive definiteness of the second derivative matrix on the contact set will imply that the contact set is finite. This reduces the problem to determining whether the second derivative matrix is positive definite on the set of minimizers of $\bar{m}_j(\theta, x)$, a problem similar to testing local identification conditions in nonlinear models (see Wright, 2003). I record this observation in the following lemma.

Lemma B.1. *Under Assumptions B.2 and B.3, if the second derivative matrix of $E(m_j(W_i, \theta)|X_i = x)$ is strictly positive definite on $\mathcal{X}_0^j$, then $\mathcal{X}_0^j$ must be finite.*

According to Lemma B.1, once we know that the second derivative matrix of $E(m_j(W_i, \theta)|X_i)$ is positive definite on the set of minimizers $E(m_j(W_i, \theta)|X_i)$, the conditions of Theorem 3.1 will hold. This reduces the problem to testing the conditions of the lemma. One simple way of doing this is to take a preliminary estimate of $\mathcal{X}_0^j$ that contains this set with probability approaching one, and then test whether the second derivative matrix of $E(m_j(W_i, \theta)|X_i)$ is positive definite on this set. In what follows, I describe an approach based on local polynomial regression estimates of the conditional mean and its second derivatives, but other methods of estimating the conditional mean would work under appropriate conditions. The methods require knowledge of a set D satisfying Assumption B.3. This set could be chosen with another preliminary test, an extension which I do not pursue.

Under the conditions above, we can estimate $\bar{m}_j(\theta, x)$ and its derivatives at a given point x with a local second order polynomial regression estimator defined as follows. For a kernel function K and a bandwidth parameter h, run a regression of $m_j(W_i, \theta)$ on a second order polynomial of X_i, weighted by the distance of X_i from x by $K((X-x)/h)$. That is, for each j and any x, define $\hat{\bar{m}}_j(\theta,x)$, $\hat{\beta}_j(x)$, and $\hat{V}_j(x)$ to be the values of m, β, and V that minimize

$$E_n\left\{\left[m_j(W_i,\theta)-\left(m+(X_i-x)'\beta+\frac{1}{2}(X_i-x)'V(X_i-x)\right)\right]^2\times K((X_i-x)/h)\right\}.$$

The pre-test uses $\hat{\bar{m}}_j(\theta,x)$ as an estimate of $\bar{m}_j(\theta,x)$ and $\hat{V}_j(x)$ as an estimate of $V_j(x)$.

The following theorem, taken from Ichimura and Todd (2007, Theorem 4.1), gives rates of convergence for these estimates of the conditional mean and its second derivatives that will be used to estimate $\mathcal{X}_0^j$ and $V_j(x)$ as described above. The theorem uses an additional assumption on the kernel K.

Assumption B.5. *The kernel function K is bounded, has compact support, and satisfies, for some C and for any $0\le j_1+\cdots+j_r\le 5$, $|u_1^{j_1}\cdots u_r^{j_r}K(u)-v_1^{j_1}\cdots v_r^{j_r}K(v)|\le C\|u-v\|$.*

Theorem B.2. *Under iid data and Assumptions 3.2, B.2, B.3, B.4, and B.5,*

$$\sup_{x\in D}\left|\hat{V}_{j,rs}(x)-V_{j,rs}(x)\right|=\mathcal{O}_p((\log n/(nh^{d_X+4}))^{1/2})+\mathcal{O}_p(h)$$

for all r and s, where $V_{j,rs}$ is the r,s element of V_j, and

$$\sup_{x\in D}\left|\hat{\bar{m}}_j(\theta,x)-\bar{m}_j(\theta,x)\right|=\mathcal{O}_p((\log n/(nh^{d_X}))^{1/2})+\mathcal{O}_p(h^3).$$

For both the conditional mean and the derivative, the first term in the asymptotic order of convergence is the variance term and the second is the bias term. The optimal choice of h sets both of these to be the same order, and is $h_n=(\log n/n)^{1/(d_X+6)}$ in both cases. This gives a $(\log n/n)^{1/(d_X+6)}$ rate of convergence for the second derivative, and a $(\log n/n)^{3/(d_X+6)}$ rate of convergence for the conditional mean. However, any choice of h such that both terms go to zero can be used.

In order to test the conditions of Lemma B.1, we can use the following procedure. For some sequence a_n growing to infinity such that $a_n[(\log n/(nh^{d_X}))^{1/2}\vee h^3]$ converges to zero, let $\hat{\mathcal{X}}_0^j=\{x\in D|\hat{\bar{m}}_j(\theta,x)-(\inf_{x'\in D}\hat{\bar{m}}_j(\theta,x')\wedge 0)|\le[a_n(\log n/(nh^{d_X}))^{1/2}\vee h^3]\}$. By Theorem B.2, $\hat{\mathcal{X}}_0^j$ will contain $\mathcal{X}_0^j$ with probability approaching one. Thus, if we can determine that $V_j(x)$ is positive definite on $\hat{\mathcal{X}}_0^j$, then, asymptotically, we will know that $V_j(x)$ is positive definite on $\mathcal{X}_0^j$. Note that $\hat{\mathcal{X}}_0^j$ is an estimate of the set of minimizers of $\overline{m}_j(x,\theta)$ over x if the moment inequality binds or fails to hold, and is eventually equal to the empty set if the moment inequality is slack.

Since the determinant is a differentiable map from $\mathbb{R}^{d_X^2}$ to $\mathbb{R}$, the $\mathcal{O}_p((\log n/(nh^{d_X+4}))^{1/2})+\mathcal{O}_p(h)$ rate of uniform convergence for $\hat{V}_j(x)$ translates to the same (or faster) rate of convergence for $\det \hat{V}_j(x)$. If, for some $x_0 \in \mathcal{X}_0^j$, $V_j(x_0)$ is not positive definite, then $V_j(x_0)$ will be singular (the second derivative matrix at an interior minimum must be positive semidefinite if the second derivatives are continuous in a neighborhood of x_0), and $\det V_j(x_0)$ will be zero. Thus, $\inf_{x\in\hat{\mathcal{X}}_0^j} \det \hat{V}_j(x) \leq \det \hat{V}_j(x_0) = \mathcal{O}_p((\log n/(nh^{d_X+4}))^{1/2}) + \mathcal{O}_p(h)$ where the inequality holds with probability approaching one. Thus, letting b_n be any sequence going to infinity such that $b_n[(\log n/(nh^{d_X+4}))^{1/2} \vee h]$ converges to zero, if $V_j(x_0)$ is not positive definite for some $x_0 \in \mathcal{X}_0^j$, we will have $\inf_{x\in\hat{\mathcal{X}}_0^j} \det \hat{V}_j(x) \leq b_n[(\log n/(nh^{d_X+4}))^{1/2} \vee h]$ with probability approaching one (actually, since we are only dealing with the point x_0, we can use results for pointwise convergence of the second derivative of the conditional mean, so the $\log n$ term can be replaced by a constant, but I use the uniform convergence results for simplicity).

Now, suppose $V_j(x)$ is positive definite for all $x \in \mathcal{X}_0^j$. By Lemma B.1, we will have, for some $B > 0$, $\det V_j(x) \geq B$ for all $x \in \mathcal{X}_0^j$. By continuity of $V_j(x)$, we will also have, for some $\varepsilon > 0$, $\det V_j(x) \geq B/2$ for all $x \in \mathcal{X}_0^{j\varepsilon}$ where $\mathcal{X}_0^{j\varepsilon} = \{x | \inf_{x'\in\mathcal{X}_0^j} \|x - x'\| \leq \varepsilon\}$ is the ε-expansion of $\mathcal{X}_0^j$. Since $\hat{\mathcal{X}}_0^j \subseteq \mathcal{X}_0^{j\varepsilon}$ with probability approaching one, we will also have $\inf_{x\in\hat{\mathcal{X}}_0^j} \det V_j(x) \geq B/2$ with probability approaching one. Since $\det \hat{V}_j(x) \to \det V_j(x)$ uniformly over D, we will then have $\inf_{x\in\hat{\mathcal{X}}_0^j} \det \hat{V}_j(x) \geq b_n[(\log n/(nh^{d_X+4}))^{1/2} \vee h]$ with probability approaching one.

This gives the following theorem.

Theorem B.3. *Let $\hat{V}_j(x)$ and $\hat{\bar{m}}_j(\theta, x)$ be the local second order polynomial estimates defined with some kernel K with h such that the rate of convergence terms in Theorem B.2 go to zero. Let $\hat{\mathcal{X}}_0^j$ be defined as above with $a_n[(\log n/(nh^{d_X}))^{1/2} \vee h^3]$ going to zero and a_n going to infinity, and let b_n be any sequence going to infinity such that $b_n[(\log n/(nh^{d_X+4}))^{1/2} \vee h]$ goes to zero. Suppose that Assumptions 3.2, B.2, B.3, B.4, and B.5, hold, and the null hypothesis holds with $E(m(W_i, \theta)m(W_i, \theta)'|X_i = x)$ continuous and the data are iid. Then, if Assumption 3.1 holds, we will have $\inf_{x\in\hat{\mathcal{X}}_0^j} \det \hat{V}_j(x) > b_n[(\log n/(nh^{d_X+4}))^{1/2} \vee h]$ for each j with probability approaching one. If Assumption 3.1 does not hold, we will have $\inf_{x\in\hat{\mathcal{X}}_0^j} \det \hat{V}_j(x) \leq b_n[(\log n/(nh^{d_X+4}))^{1/2} \vee h]$ for some j with probability approaching one.*

The purpose of this test of Assumption 3.1 is as a preliminary consistent test in a procedure that uses the asymptotic approximation in Theorem 3.1 if the test finds evidence in favor of Assumption 3.1, and uses the methods that are robust to different types of contact sets, but possibly conservative, such as those described in Andrews and Shi (2013), otherwise. It follows from Theorem B.3 that such a procedure will have the correct size asymptotically.

Consider the following test. For some $b_n \to \infty$ and $h \to 0$ satisfying the conditions of Theorem B.3, perform a pre-test that finds evidence in favor of Assumption 3.1 iff. $\inf_{x \in \hat{\mathcal{X}}_0} \det \hat{V}_j(x) \geq b_n[(\log n/(nh^{d_X+4}))^{1/2} \vee h]$ for each j. If $\hat{\mathcal{X}}_0 = \emptyset$, do not reject the null hypothesis that $\theta \in \Theta_0$. If $\inf_{x \in \hat{\mathcal{X}}_0} \det \hat{V}_j(x) > b_n[(\log n/(nh^{d_X+4}))^{1/2} \vee h]$ for each j, reject the null hypothesis that $\theta \in \Theta_0$ if $n^{(d_X+2)/(d_X+4)} S(T_n(\theta)) > \hat{q}_{1-\alpha}$ where $\hat{q}_{1-\alpha}$ is an estimate of the $1-\alpha$ quantile of the distribution of $S(Z)$ formed using one of the methods in Section 4 or Section B.1. If $\inf_{x \in \hat{\mathcal{X}}_0} \det \hat{V}_j(x) \leq b_n[(\log n/(nh^{d_X+4}))^{1/2} \vee h]$ for some j, perform any (possibly conservative) asymptotically level α test.

In the statement of the following theorem, it is understood that Assumptions 4.1 and B.1, which refer to objects in Assumption 3.1, do not need to hold if the data generating process is such that Assumption 3.1 does not hold.

Theorem B.4. *Suppose that Assumptions 3.2, 3.3, 4.1, 4.2, B.2, B.3, B.4, and B.5 hold, $E(m(W_i, \theta)m(W_i, \theta)'|X_i = x)$ is continuous, and the data are iid. Then the test given above provides an asymptotically level α test of $\theta \in \Theta_0$ if the subsampling procedure is used or if Assumption B.1 holds and the procedure based on estimating the asymptotic distribution directly is used. If Assumption 3.1 holds, this test is asymptotically exact.*

The estimates used for this pre-test can also be used to construct estimates of the quantities in Assumption B.1 that satisfy the consistency requirements of this assumption. Suppose that we have estimates $\hat{M}(x)$, $\hat{f}_X(x)$, and $\hat{V}(x)$ of $E(m(W_i, \theta)m(W_i, \theta)'|X = x)$, $f_X(x)$, and $V(x)$ that are consistent uniformly over x in a neighborhood of $\mathcal{X}_0$. Then, if we have estimates of $\mathcal{X}_0$ and $J(k)$, we can estimate the quantities in Assumption B.1 using $\hat{M}_k(x_k)$, $\hat{f}_X(x_k)$, and $\hat{V}(x_k)$ for each x_k in the estimate of $\mathcal{X}_0$, where $\hat{M}_k(x_k)$ is a sparse version of $\hat{M}(x_k)$ with elements with indices not in the estimate of $J(k)$ set to zero.

The estimate $\hat{\mathcal{X}}_0$ contains infinitely many points, so it will not work for this purpose. Instead, define the estimate $\tilde{\mathcal{X}}_0$ of $\mathcal{X}_0$ and the estimate $\hat{J}(k)$ of $J(k)$ as follows. Let a_n be as in Theorem B.3, and let $\varepsilon_n^2 \to 0$ more slowly than $a_n[(\log n/(nh^{d_X}))^{1/2} \vee h^3]$. Let $\hat{\ell}_j$ be the smallest number such that $\hat{\mathcal{X}}_0^j \subseteq \cup_{k=1}^{\hat{\ell}_j} B_{\varepsilon_n}(\hat{x}_{j,k})$ for some $\hat{x}_{j,1}, \ldots, \hat{x}_{j,\hat{\ell}_j}$. Define an equivalence relation $\sim$ on the set $\{(j,k)|1 \leq j \leq d_Y, 1 \leq k \leq \hat{\ell}_j\}$ by $(j,k) \sim (j',k')$ iff. there is a sequence $(j,k) = (j_1,k_1), (j_2,k_2), \ldots, (j_r,k_r) = (j',k')$ such that $B_{\varepsilon_n}(\hat{x}_{j_s,k_s}) \cap B_{\varepsilon_n}(\hat{x}_{j_{s+1},k_{s+1}}) \neq \emptyset$ for s from 1 to $r-1$. Let $\hat{\ell}$ be the number of equivalence classes, and, for each equivalence class, pick exactly one (j,k) in the equivalence class and let $\tilde{x}_r = \hat{x}_{j,k}$ for some r between 1 and $\hat{\ell}$. Define the estimate of the set $\mathcal{X}_0$ to be $\tilde{\mathcal{X}}_0 \equiv \{\tilde{x}_1, \ldots, \tilde{x}_{\hat{\ell}}\}$, and define the estimate $\hat{J}(r)$ for r from 1 to $\hat{\ell}$ to be the set of indices j for which some (j,k) is in the same equivalence class as $\tilde{x}_r$.

Although these estimates of $\mathcal{X}_0$, ℓ, and $J(1), \ldots, J(\ell)$ require some cumbersome notation to define, the intuition behind them is simple. Starting with the initial estimates $\hat{\mathcal{X}}_j$, turn these sets into discrete sets of points by taking the centers of balls that contain the sets $\hat{\mathcal{X}}_j$ and converge at a slower rate. This gives estimates of the points at which the conditional moment inequality indexed by j binds for each j, but to estimate the asymptotic distribution in Theorem 3.1, we also need to determine which components, if any, of $\bar{m}(\theta, x)$ bind at the same value of x. The procedure described above does this by testing whether the balls used to form the estimated contact points for each index of $\bar{m}(\theta, x)$ intersect across indices.

The following theorem shows that this is a consistent estimate of the set $\mathcal{X}_0$ and the indices of the binding moments.

Theorem B.5. *Suppose that Assumptions 3.1, B.2, B.3, B.4, and B.5 hold. For the estimates $\tilde{\mathcal{X}}_0$, $\hat{\ell}$ and $\hat{J}(r)$, $\hat{\ell} = \ell$ with probability approaching one and, for some labeling of the indices of $\tilde{x}_1, \ldots, \tilde{x}_{\hat{\ell}}$ we have, for k from 1 to ℓ, $\tilde{x}_k \stackrel{p}{\to} x_k$ and, with probability approaching one, $\hat{J}(k) = J(k)$.*

An immediate consequence of this is that this estimate of $\mathcal{X}_0$ can be used in combination with consistent estimates of $E(m(W_i, \theta)m(W_i, \theta)'|X = x)$, $f_X(x)$, and $V(x)$ to form estimates of these functions evaluated at points in $\mathcal{X}_0$ that satisfy the assumptions needed for the procedure for estimating the asymptotic distribution described in Section 4.

Corollary B.2. *If the estimates $\hat{M}_k(x)$, $\hat{f}_X(x)$, and $\hat{V}(x)$ are consistent uniformly over x in a neighborhood of $\mathcal{X}_0$, then, under Assumptions 3.1, B.2, B.3, B.4, and B.5, the estimates $\hat{M}_k(\tilde{x}_k)$, $\hat{f}_X(\tilde{x}_k)$, and $\hat{V}_j(\tilde{x}_k)$ satisfy Assumption B.1.*

C Proofs

This section of the appendix contains proofs of the theorems in this paper. The proofs are organized into subsections according to the section containing the theorem in the body of the paper. In cases where a result follows immediately from other theorems or arguments in the body of the paper, I omit a separate proof. Statements involving convergence in distribution in which random elements in the converging sequence are not measurable with respect to the relevant Borel sigma algebra are in the sense of outer weak convergence (see van der Vaart and Wellner, 1996). For notational convenience, I use $d = d_X$ throughout this section of the appendix.

C.1 Asymptotic Distribution of the KS Statistic

In this subsection of the appendix, I prove Theorem 3.1. This generalizes the proof for the $\ell = d_Y = 1$ case in Section A, and much of this proof is taken verbatim from the proof in that section, with appropriate notational changes. For notational convenience, let $Y_i = m(W_i, \theta)$ and $Y_{i,J(m)} = m_{J(m)}(W_i, \theta)$ and let $d = d_X$ and $k = d_Y$ throughout this subsection.

The asymptotic distribution comes from the behavior of the objective function $E_n Y_{i,j} I(s < X_i < s + t)$ for (s, t) near x_m such that $j \in J(m)$. The bulk of the proof involves showing that the objective function doesn't matter for (s, t) outside of neighborhoods of x_m with $j \in J(m)$ where these neighborhoods shrink at a fast enough rate. First, I derive the limiting distribution over such shrinking neighborhoods and the rate at which they shrink.

Theorem C.1. *Let $h_n = n^{-\alpha}$ for some $0 < \alpha < 1/d$. Let*

$$\mathbb{G}_{n,x_m}(s,t) = \frac{\sqrt{n}}{h_n^{d/2}}(E_n - E)Y_{i,J(m)} I(h_n s < X_i - x_m < h_n(s+t))$$

and let $g_{n,x_m}(s,t)$ have jth element

$$g_{n,x_m,j}(s,t) = \frac{1}{h_n^{d+2}} E Y_{i,j} I(h_n s < X_i - x_m < h_n(s+t))$$

if $j \in J(m)$ and zero otherwise. Then, for any finite M, $(\mathbb{G}_{n,x_1}(s,t), \ldots, \mathbb{G}_{n,x_\ell}(s,t)) \xrightarrow{d} (\mathbb{G}_{P,x_1}(s,t), \ldots, \mathbb{G}_{P,x_\ell}(s,t))$ taken as random processes on $\|(s,t)\| \le M$ with the supremum norm and $g_{n,x_m}(s,t) \to g_{P,x_m}(s,t)$ uniformly in $\|(s,t)\| \le M$ where $\mathbb{G}_{P,x_m}(s,t)$ and $g_{P,x_m}(s,t)$ are defined as in Theorem 3.1 for m from 1 to ℓ.

Proof. The convergence in distribution in the first statement follows from verifying the conditions of Theorem 2.11.22 in van der Vaart and Wellner (1996). To derive the covariance kernel, note that

$$\begin{aligned}
&cov(\mathbb{G}_{n,x_m}(s,t), \mathbb{G}_{n,x_m}(s',t')) \\
&= h_n^{-d} E Y_{i,J(m)} Y_{i,J(m)}' I\left\{h_n(s \vee s') < X - x_m < h_n\left[(s+t) \wedge (s'+t')\right]\right\} \\
&- h_n^{-d}\left\{E Y_{i,J(m)} I\left[h_n s < X - x_m < h_n(s+t)\right]\right\}\left\{E Y_{i,J(m)}' I\left[h_n s' < X - x_m < h_n(s'+t')\right]\right\}.
\end{aligned}$$

The second term goes to zero as $n \to \infty$. The first is equal to the claimed covariance kernel plus the error term

$$h_n^{-d}\int_{h_n(s\vee s')<x-x_m<h_n[(s+t)\wedge(s'+t')]}\left[E(Y_{i,J(m)}Y_{i,J(m)}'|X=x)f_X(x)-E(Y_{i,J(m)}Y_{i,J(m)}'|X=x_m)f_X(x_m)\right]\,dx,$$

which is bounded by

$$\begin{aligned}&\left\{\max_{\|x-x_m\|\le 2h_nM}\left[E(Y_{i,J(m)}Y_{i,J(m)}'|X=x)f_X(x)-E(Y_{i,J(m)}Y_{i,J(m)}'|X=x_m)f_X(x_m)\right]\right\}\\&\times h_n^{-d}\int_{h_n(s\vee s')<x-x_m<h_n[(s+t)\wedge(s'+t')]}dx\\&=\left\{\max_{\|x-x_m\|\le 2h_nM}\left[E(Y_{i,J(m)}Y_{i,J(m)}'|X=x)f_X(x)-E(Y_{i,J(m)}Y_{i,J(m)}'|X=x_m)f_X(x_m)\right]\right\}\\&\times\int_{(s\vee s')<x-x_m<(s+t)\wedge(s'+t')}dx.\end{aligned}$$

This goes to zero as $n\to\infty$ by continuity of $E(Y_{i,J(m)}Y_{i,J(m)}'|X=x)$ and $f_X(x)$. For $m\neq r$ and $\|(s,t)\|\le M$, $\|(s',t')\|\le M$, $cov(\mathbb{G}_{n,x_m}(s,t),\mathbb{G}_{n,x_r}(s',t'))$ is eventually equal to

$$-h_n^{-d}\left\{EY_{i,J(m)}I\left[h_ns<X-x_m<h_n(s+t)\right]\right\}\left\{EY_{i,J(r)}'I\left[h_ns'<X-x_r<h_n(s'+t')\right]\right\},$$

which goes to zero, so the processes for different elements of $\mathcal{X}_0$ are independent as claimed.

For the claim regarding $g_{n,x_m}(s,t)$, first note that the assumptions imply that, for $j\in J(m)$, the first derivative of $x\mapsto E(Y_{i,j}|X=x)$ at $x=x_m$ is 0, and that this function has a second order Taylor expansion:

$$E(Y_{i,j}|X=x)=\frac{1}{2}(x-x_m)'V_j(x_m)(x-x_m)+R_n(x)$$

where

$$R_n(x)=\frac{1}{2}(x-x_m)'V_j(x^*(x))(x-x_m)-\frac{1}{2}(x-x_m)'V_j(x_m)(x-x_m)$$

and $V_j(x^*)$ is the second derivative matrix evaluated at some $x^*(x)$ between x_m and x.

We have

$$g_{n,x_m,j}(s,t) = \frac{1}{2h_n^{d+2}} \int_{h_n s < x - x_m < h_n(s+t)} (x - x_m)' V_j(x_m)(x - x_m) f_X(x_m)\, dx$$
$$+ \frac{1}{2h_n^{d+2}} \int_{h_n s < x - x_m < h_n(s+t)} (x - x_m)' V_j(x_m)(x - x_m)[f_X(x) - f_X(x_m)]\, dx$$
$$+ \frac{1}{h_n^{d+2}} \int_{h_n s < x - x_m < h_n(s+t)} R_n(x) f_X(x)\, dx.$$

The first term is equal to $g_{P,x_m,j}(s,t)$ by a change of variable x to $h_n x + x_m$ in the integral. The second term is bounded by $g_{P,x_m,j}(s,t) \sup_{\|x - x_m\| \le 2h_n M} [f_X(x) - f_X(x_m)]/f_X(x_m)$, which goes to zero uniformly in $\|(s,t)\| \le M$ by continuity of f_X. The third term is equal to (using the same change of variables)

$$\frac{1}{2} \int_{s < x < s+t} [x' V_j(x^*(h_n x + x_m))x - x' V_j(x_m) x] f_X(h_n x + x_m)\, dx.$$

This is bounded by a constant times $\sup_{\|x\| \le M} |x' V_j(x^*(h_n x + x_m))x - x' V_j(x_m)x|$, which goes to zero as $n \to \infty$ by continuity of the second derivatives. □

Thus, if we let h_n be such that $\sqrt{n}/h_n^{d/2} = 1/h_n^{d+2} \iff h_n = n^{-1/(d+4)}$ and scale up by $\sqrt{n}/h_n^{d/2} = 1/h_n^{d+2} = n^{(d+2)/(d+4)}$, we will have

$$n^{(d+2)/(d+4)}(E_n Y_{i,J(1)} I(h_n s < X - x_1 < h_n(s+t)), \ldots, E_n Y_{i,J(\ell)} I(h_n s < X - x_\ell < h_n(s+t))$$
$$= (\mathbb{G}_{n,x_1}(s,t) + g_{n,x_1}(s,t), \ldots, \mathbb{G}_{n,x_\ell}(s,t) + g_{n,x_\ell}(s,t))$$
$$\stackrel{d}{\to} (\mathbb{G}_{P,x_1}(s,t) + g_{P,x_1}(s,t), \ldots, \mathbb{G}_{P,x_m}(s,t) + g_{P,x_m}(s,t))$$

taken as stochastic processes in $\{\|(s,t)\| \le M\}$ with the supremum norm. From now on, let $h_n = n^{-1/(d+4)}$ so that this will hold.

We would like to show that the infimum of these stochastic processes over all of $\mathbb{R}^{2d}$ converges to the infimum of the limiting process over all of $\mathbb{R}^{2d}$, but this does not follow immediately since we only have uniform convergence on compact sets. Another way of thinking about this problem is that convergence in distribution in $\{\|(s,t)\| \le M\}$ with the supremum norm for any M implies convergence in distribution in $\mathbb{R}^{2d}$ with the topology of uniform convergence on compact sets (see Kim and Pollard, 1990), but the infimum over all of $\mathbb{R}^{2d}$ is not a continuous mapping on this space since uniform convergence on all compact sets does not imply convergence of the infimum over all of $\mathbb{R}^{2d}$. To get the desired result, the following lemma will be useful. The idea is to show that values of (s, t) far away from zero won't matter for the limiting distribution, and then use convergence for fixed compact sets.

Lemma C.1. *Let $\mathbb{H}_n$ and $\mathbb{H}_P$ be random functions from $\mathbb{R}^{k_1}$ to $\mathbb{R}^{k_2}$ such that, (i) for all M, $\mathbb{H}_n \xrightarrow{d} \mathbb{H}_P$ when $\mathbb{H}_n$ and $\mathbb{H}_P$ are taken as random processes on $\{t \in \mathbb{R}^{k_1} | \|t\| \leq M\}$ with the supremum norm, (ii) for all $r < 0$, $\varepsilon > 0$, there exists an M such that $P\left(\inf_{\|t\|>M} \mathbb{H}_{P,j}(t) \leq r \text{ some } j\right) < \varepsilon$ and an N such that $P\left(\inf_{\|t\|>M} \mathbb{H}_{n,j}(t) \leq r \text{ some } j\right) < \varepsilon$ for all $n \geq N$ and (iii) $\inf_t \mathbb{H}_{n,j}(t) \leq 0$ and $\inf_t \mathbb{H}_{P,j}(t) \leq 0$ with probability one. Then $\inf_{t\in\mathbb{R}^{k_1}} \mathbb{H}_n(t) \xrightarrow{d} \inf_{t\in\mathbb{R}^{k_1}} \mathbb{H}_P(t)$.*

Proof. First, by the Cramer-Wold device, it suffices to show that, for all $w \in \mathbb{R}^{k_2}$, $w' \inf_{t\in\mathbb{R}^{k_1}} \mathbb{H}_n(t) \xrightarrow{d} w' \inf_{t\in\mathbb{R}^{k_1}} \mathbb{H}_P(t)$. For this, it suffices to show that for all $r \in \mathbb{R}$, $\liminf_n P\left(w' \inf_{t\in\mathbb{R}^{k_1}} \mathbb{H}_n(t) < r\right) \geq P\left(w' \inf_{t\in\mathbb{R}^{k_1}} \mathbb{H}_P(t) < r\right)$ and $\limsup_n P\left(w' \inf_{t\in\mathbb{R}^{k_1}} \mathbb{H}_n(t) \leq r\right) \leq P\left(w' \inf_{t\in\mathbb{R}^{k_1}} \mathbb{H}_P(t) \leq r\right)$ since, arguing along the lines of the Portmanteau Lemma, when r is a continuity point of the limiting distribution, we will have

$$P\left(w' \inf_{t\in\mathbb{R}^{k_1}} \mathbb{H}_P(t) \leq r\right) = P\left(w' \inf_{t\in\mathbb{R}^{k_1}} \mathbb{H}_P(t) < r\right) \leq \liminf_n P\left(w' \inf_{t\in\mathbb{R}^{k_1}} \mathbb{H}_n(t) < r\right)$$
$$\leq \liminf_n P\left(w' \inf_{t\in\mathbb{R}^{k_1}} \mathbb{H}_n(t) \leq r\right) \leq \limsup_n P\left(w' \inf_{t\in\mathbb{R}^{k_1}} \mathbb{H}_n(t) \leq r\right) \leq P\left(w' \inf_{t\in\mathbb{R}^{k_1}} \mathbb{H}_P(t) \leq r\right).$$

Given $\varepsilon > 0$, let M and N be as in the assumptions of the lemma, but with r replaced by $r/(k_2 \max_i |w_i|)$. Then

$$P\left(w' \inf_{\|t\|\geq M} \mathbb{H}_P(t) < r\right) \leq P\left((k_2 \max_i |w_i|) \inf_{\|t\|\geq M} \mathbb{H}_{P,j}(t) < r \text{ some } j\right) \leq \varepsilon$$

so that $P\left(w' \inf_{\|t\|\leq M} \mathbb{H}_P(t) < r\right) + \varepsilon \geq P\left(w' \inf_{t\in\mathbb{R}^{k_1}} \mathbb{H}_P(t) < r\right)$ and, for $n \geq N$,

$$P\left(w' \inf_{\|t\|\geq M} \mathbb{H}_n(t) \leq r\right) \leq P\left((k_2 \max_i |w_i|) \inf_{\|t\|\geq M} \mathbb{H}_{n,j}(t) \leq r \text{ some } j\right) \leq \varepsilon$$

so that $P\left(w' \inf_{\|t\|\leq M} \mathbb{H}_n(t) \leq r\right) + \varepsilon \geq P\left(w' \inf_{t\in\mathbb{R}} \mathbb{H}_n(t) \leq r\right)$. Thus, by convergence in distribution of the infima over $\|t\| \leq M$,

$$\liminf_n P\left(w' \inf_{t\in\mathbb{R}^{k_1}} \mathbb{H}_n(t) < r\right) \geq \liminf_n P\left(w' \inf_{\|t\|\leq M} \mathbb{H}_n(t) < r\right) \geq P\left(w' \inf_{\|t\|\leq M} \mathbb{H}_P(t) < r\right)$$
$$\geq P\left(w' \inf_{t\in\mathbb{R}^{k_1}} \mathbb{H}_P(t) < r\right) - \varepsilon$$

and

$$\limsup_n P\left(w' \inf_{t\in\mathbb{R}^{k_1}} \mathbb{H}_n(t) \le r\right) \le \limsup_n P\left(w' \inf_{\|t\|\le M} \mathbb{H}_n(t) \le r\right) + \varepsilon$$
$$\le P\left(w' \inf_{\|t\|\le M} \mathbb{H}_P(t) \le r\right) + \varepsilon \le P\left(w' \inf_{t\in\mathbb{R}^{k_1}} \mathbb{H}_P(t) \le r\right) + \varepsilon.$$

Since ε was arbitrary, this gives the desired result.

□

Technically, this lemma does not apply to

$$(\mathbb{G}_{n,x_1}(s,t) + g_{n,x_1}(s,t), \ldots, \mathbb{G}_{n,x_\ell}(s,t) + g_{n,x_\ell}(s,t))$$

since, for $m \neq r$, $\mathbb{G}_{n,x_m}(s,t) + g_{n,x_m}(s,t)$ evaluated at some increasing values of (s,t) may actually be equal to $\mathbb{G}_{n,x_r}(s',t') + g_{n,x_r}(s',t')$ for some small values of (s',t'), since, once the local indices are large enough, the original indices overlap. Instead, noting that, for any $\eta > 0$,

$$\begin{aligned}
&n^{(d+2)/(d+4)} \inf_{s,t} E_n Y_i I(s < X_i < s+t) \\
&= \Bigg(\min_{m \text{ s.t. } 1\in J(m)} \inf_{\|(s,t)\|\le\eta/h_n} \mathbb{G}_{n,x_m,1}(s,t) + g_{n,x_m,1}(s,t)), \ldots, \\
&\min_{m \text{ s.t. } k\in J(m)} \inf_{\|(s,t)\|\le\eta/h_n} \mathbb{G}_{n,x_m,k}(s,t) + g_{n,x_m,k}(s,t)\Bigg) \\
&\wedge \Bigg(n^{(d+2)/(d+4)} \inf_{\|(s-x_m,t)\|>\eta \text{ all } m \text{ s.t. } 1\,\in\, J(m)} E_n Y_{i,1} I(s < X_i < s+t), \ldots, \\
&n^{(d+2)/(d+4)} \inf_{\|(s-x_m,t)\|>\eta \text{ all } m \text{ s.t. } k\,\in\, J(m)} E_n Y_{i,k} I(s < X_i < s+t)\Bigg) \\
&\equiv Z_{n,1} \wedge Z_{n,2},
\end{aligned}$$

I show that, for some $\eta > 0$, $Z_{n,2} \stackrel{p}{\to} 0$ using a separate argument, and use Lemma C.1 to show that, for the same η,

$$\begin{aligned}
&(\inf_{s,t}[\mathbb{G}_{n,x_1}(s,t) + g_{n,x_1}(s,t)] I(\|(s,t)\| \le \eta/h_n), \ldots, \inf_{s,t}[\mathbb{G}_{n,x_\ell}(s,t) + g_{n,x_\ell}(s,t)] I(\|(s,t)\| \le \eta/h_n)) \\
&\stackrel{d}{\to} (\inf_{s,t} \mathbb{G}_{P,x_1}(s,t) + g_{P,x_1}(s,t), \ldots, \inf_{s,t} \mathbb{G}_{P,x_\ell}(s,t) + g_{P,x_\ell}(s,t)),
\end{aligned}$$

from which it follows that $Z_{n,1} \stackrel{d}{\to} Z$ for Z defined as in Theorem 3.1 by the continuous mapping theorem.

Part (i) of Lemma C.1 follows from Theorem C.1 (the $I(\|(s,t)\| \le \eta/h_n)$ term does not change this, since it is equal to one for $\|(s,t)\| \le M$ eventually). Part (iii) follows since the processes involved are equal to zero when $t = 0$. To verify part (ii), first note that it suffices to verify part (ii) of the lemma for $\mathbb{G}_{n,x_m,j}(s,t) + g_{n,x_m,j}(s,t)$ and $\mathbb{G}_{P,x_m,j}(s,t) + g_{P,x_m,j}(s,t)$ for each m and j individually. Part (ii) of the lemma holds trivially for m and j such that $j \notin J(m)$, so we need to verify this part of the lemma for m and j such that $j \in J(m)$.

The next two lemmas provide bounds that will be used to verify condition (ii) of Lemma C.1 for $\mathbb{G}_{n,x_m,j}(s,t) + g_{n,x_m,j}(s,t)$ and $\mathbb{G}_{P,x_m,j}(s,t) + g_{P,x_m,j}(s,t)$ for m and j with $j \in J(m)$. To do this, the bounds in the lemmas are applied to sequences of sets of (s,t) where the norm of elements in the set increases with the sequence. The idea is similar to the "peeling" argument of, for example, Kim and Pollard (1990), but different arguments are required to deal with values of (s,t) for which, even though $\|s\|$ is large, $\prod_i t_i$ is small so that the objective function on average uses only a few observations, which may happen to be negative. To get bounds on the suprema of the limiting and finite sample processes where t may be small relative to s, the next two lemmas bound the supremum by a maximum over s in a finite grid of suprema over t with s fixed, and then use exponential bounds on suprema of the processes with fixed s.

Lemma C.2. *Fix m and j with $j \in J(m)$. For some $C > 0$ that depends only on d, $f_X(x_m)$ and $E(Y_{i,j}^2|X = x_m)$, we have, for any $B \ge 1$, $\varepsilon > 0$, $w > 0$,*

$$P\left(\sup_{\|(s,t)\| \le B, \prod_i t_i \le \varepsilon} |\mathbb{G}_{P,x_m,j}(s,t)| \ge w\right) \le 2\left\{3B[B^d/(\varepsilon \wedge 1)] + 2\right\}^{2d} \exp\left(-C\frac{w^2}{\varepsilon}\right)$$

for $\frac{w^2}{\varepsilon}$ greater than some constant that depends only on d, $f_X(x_m)$ and $E(Y_{i,j}^2|X = x_m)$.

Proof. Let $\mathbb{G}(s,t) = \mathbb{G}_{P,x_m,j}(s,t)$. We have, for any $s_0 \le s \le s+t \le s_0+t_0$,

$$\begin{aligned}
&\mathbb{G}(s,t) = \mathbb{G}(s_0, t+s-s_0) \\
&+ \sum_{1 \le j \le d} (-1)^j \sum_{1 \le i_1 < i_2 < \ldots < i_j \le d} \\
&\mathbb{G}(s_0, (t_1 + s_1 - s_{0,1}, \ldots, t_{i_1-1} + s_{i_1-1} - s_{0,i_1-1}, s_{i_1} - s_{0,i_1}, t_{i_1+1} + s_{i_1+1} - s_{0,i_1+1}, \\
&\ldots, t_{i_j-1} + s_{i_j-1} - s_{0,i_j-1}, s_{i_j} - s_{0,i_j}, t_{i_j+1} + s_{i_j+1} - s_{0,i_j+1}, \ldots, t_d + s_d - s_{0,d})).
\end{aligned}$$

Thus, since there are 2^d terms in the above display, each with absolute value bounded by

$\sup_{t\le t_0} |\mathbb{G}(s_0,t)|$,

$$\sup_{s_0\le s\le s+t\le s_0+t_0} |\mathbb{G}(s,t)| \le 2^d \sup_{t\le t_0} |\mathbb{G}(s_0,t)| \overset{d}{=} 2^d \sup_{t\le t_0} |\mathbb{G}(0,t)|.$$

Let A be a grid of meshwidth $(\varepsilon\wedge 1)/B^d$ covering $[-B,2B]^d$. For any (s,t) with $\|(s,t)\|\le B$ and $\prod_i t_i \le \varepsilon$, there are s_0 and t_0 with $s_0, s_0+t_0\in A$ such that $s_0\le s\le s+t\le s_0+t_0$, and $\prod_i t_{0,i} \le \prod_i (t_i + (\varepsilon\wedge 1)/B^d) = \sum_{j=0}^d [(\varepsilon\wedge 1)/B^d]^j \sum_{I\in\{1,\ldots,d\},|I|=d-j} \prod_{i\in I} t_i \le \prod_i t_i + \sum_{j=1}^d [(\varepsilon\wedge 1)/B^d]^j \binom{d}{d-j} B^{d-j} \le \varepsilon + \varepsilon\sum_{j=1}^d B^{-dj}\binom{d}{d-j} B^{d-j} \le 2^d\varepsilon$. For this s_0, t_0, we will then have, by the above display, $|\mathbb{G}(s,t)| \le 2^d \sup_{t\le t_0} |\mathbb{G}(s_0,t)|$.

This gives

$$\sup_{\|(s,t)\|\le B, \prod_i t_i\le\varepsilon} |\mathbb{G}(s,t)| \le 2^d \max_{s_0,s_0+t_0\in A, \prod_i t_{0,i}\le 2^d\varepsilon} \sup_{t\le t_0} |\mathbb{G}(s_0,t)|,$$

so that

$$P\left(\sup_{\|(s,t)\|\le B, \prod_i t_i\le\varepsilon} |\mathbb{G}(s,t)| \ge w\right) \le |A|^2 \max_{s_0,s_0+t_0\in A, \prod_i t_{0,i}\le 2^d\varepsilon} P\left(2^d \sup_{t\le t_0} |\mathbb{G}(s_0,t)| \ge w\right)$$
$$= |A|^2 \max_{s_0,s_0+t_0\in A, \prod_i t_{0,i}\le 2^d\varepsilon} P\left(2^d \sup_{t\le 1}\left(\prod_i t_{0,i}\right)^{1/2} |\mathbb{G}(0,t)| \ge w\right)$$
$$\le |A|^2 P\left(\sup_{t\le 1} |\mathbb{G}(0,t)| \ge \frac{w}{2^d 2^{d/2}\varepsilon^{1/2}}\right).$$

The result then follows using the fact that $|A|\le \left\{3B[B^d/(\varepsilon\wedge 1)]+2\right\}^d$ and using Theorem 2.1 (p.43) in Adler (1990) to bound the probability in the last line of the display (the theorem in Adler (1990) shows that the probability in the above display is bounded by $2\exp(-K_1 w^2/\varepsilon + K_2 w/\varepsilon^{1/2} + K_3)$ for some constants K_1, K_2, and K_3 with $K_1>0$ that depend only on d, $f_X(x_m)$ and $E(Y_{i,j}^2|X=x_m)$ and this expression is less than $2\exp(-(K_1/2)w^2/\varepsilon)$ for w^2/ε greater than some constant that depends only on K_1, K_2, and K_3).

□

Lemma C.3. *Fix m and j with $j\in J(m)$. For some $C>0$ that depends only on the distribution of (X,Y) and some $\eta>0$, we have, for any $1\le B\le h_n^{-1}\eta$, $w>0$ and*

$\varepsilon \geq n^{-4/(d+4)}(1+\log n)^2$,

$$P\left(\sup_{\|(s,t)\|\leq B, \prod_i t_i \leq \varepsilon} |\mathbb{G}_{n,x_m,j}(s,t)| \geq w\right) \leq 2\left\{3B[B^d/(\varepsilon \wedge 1)]+2\right\}^{2d} \exp\left(-C\frac{w}{\varepsilon^{1/2}}\right).$$

Proof. Let $\mathbb{G}_n(s,t) = \mathbb{G}_{n,x_m,j}(s,t)$. By the same argument as in the previous lemma with $\mathbb{G}$ replaced by $\mathbb{G}_n$, we have

$$\sup_{s_0 \leq s \leq s+t \leq s_0+t_0} |\mathbb{G}_n(s,t)| \leq 2^d \sup_{t \leq t_0} |\mathbb{G}_n(s_0,t)|.$$

As in the previous lemma, let A be a grid of meshwidth $(\varepsilon \wedge 1)/B^d$ covering $[-B, 2B]^d$. Arguing as in the previous lemma, we have, for any (s,t) with $\|(s,t)\| \leq B$ and $\prod_i t_i \leq \varepsilon$, there exists some s_0, t_0 with $s_0, s_0+t_0 \in A$ such that $\Pi_i t_{0,i} \leq 2^d\varepsilon$ and $|\mathbb{G}_n(s,t)| \leq 2^d \sup_{t\leq t_0} |\mathbb{G}_n(s_0,t)|$. Thus,

$$\begin{aligned}
&\sup_{\|(s,t)\|\leq B, \prod_i t_i \leq \varepsilon} |\mathbb{G}_n(s,t)| \leq 2^d \max_{s_0, s_0+t_0 \in A, \prod_i t_{0,i} \leq 2^d\varepsilon} \sup_{t \leq t_0} |\mathbb{G}_n(s_0,t)| \\
&= 2^d \max_{s_0, s_0+t_0 \in A, \prod_i t_{0,i} \leq 2^d\varepsilon} \sup_{t \leq t_0} \frac{\sqrt{n}}{h_n^{d/2}} |(E_n - E) Y_{i,j} I(h_n s_0 \leq X_i - x_m \leq h_n(s_0+t))|.
\end{aligned}$$

This gives

$$\begin{aligned}
&P\left(\sup_{\|(s,t)\|\leq B, \prod_i t_i \leq 2^d\varepsilon} |\mathbb{G}_n(s,t)| \geq w\right) \\
&\leq |A|^2 \max_{s_0, s_0+t_0 \in A, \prod_i t_{0,i} \leq 2^d\varepsilon} P\left(2^d \sup_{t \leq t_0} \frac{\sqrt{n}}{h_n^{d/2}} |(E_n - E) Y_{i,j} I(h_n s_0 \leq X_i - x_m \leq h_n(s_0+t))| \geq w\right).
\end{aligned}$$

We have, for some universal constant K and all n with $\varepsilon \geq n^{-4/(d+4)}(1+\log n)^2$, letting $\mathcal{F}_n = \{(x,y) \mapsto y_j I(h_n s_0 \leq x - x_m \leq h_n(s_0+t)) | t \leq t_0\}$ and defining $\|\cdot\|_{P,\psi_1}$ to be the

Orlicz norm defined on p.90 of van der Vaart and Wellner (1996) for $\psi_1(x) = \exp(x) - 1$,

$$\begin{aligned}
&\|2^d \sup_{f\in\mathcal{F}_n} |\sqrt{n}(E_n - E)f(X_i, Y_i)|\|_{P,\psi_1} \\
&\le K\left[E \sup_{f\in\mathcal{F}_n} |\sqrt{n}(E_n - E)f(X_i, Y_i)| + n^{-1/2}(1+\log n)\||Y_{i,j}|I(h_n s_0 \le X_i - x_m \le h_n(s_0+t_0))\|_{P,\psi_1}\right] \\
&\le K\left[J(1,\mathcal{F}_n, L^2)\left\{E[|Y_{i,j}|I(h_n s_0 < X_i - x_m < h_n(s_0+t_0))]^2\right\}^{1/2} + n^{-1/2}(1+\log n)\|Y\|_{P,\psi_1}\right] \\
&\le K\left[J(1,\mathcal{F}_n, L^2)\overline{f}^{1/2}\overline{Y}h_n^{d/2}2^{d/2}\varepsilon^{1/2} + n^{-1/2}(1+\log n)\|Y_{i,j}\|_{P,\psi_1}\right] \\
&\le K\left[J(1,\mathcal{F}_n, L^2)\overline{f}^{1/2}\overline{Y}2^{d/2} + \|Y_{i,j}\|_{P,\psi_1}\right] h_n^{d/2}\varepsilon^{1/2}.
\end{aligned}$$

The first inequality follows by Theorem 2.14.5 in van der Vaart and Wellner (1996). The second uses Theorem 2.14.1 in van der Vaart and Wellner (1996). The fourth inequality uses the fact that $h_n^{d/2}\varepsilon^{1/2} = n^{-d/[2(d+4)]}\varepsilon^{1/2} \ge n^{-1/2}(1+\log n)$ once $\varepsilon^{1/2} \ge n^{-1/2+d/[2(d+4)]}(1+\log n) = n^{-2/(d+4)}(1+\log n)$. Since each $\mathcal{F}_n$ is contained in the larger class $\mathcal{F} \equiv \{(x,y) \mapsto y_j I(s < x - x_m < s+t)|(s,t) \in \mathbb{R}^{2d}\}$, we can replace $\mathcal{F}_n$ by $\mathcal{F}$ on the last line of this display. Since $J(1,\mathcal{F},L^2)$ and $\|Y_{i,j}\|_{\psi_1}$ are finite ($\mathcal{F}$ is a VC class and $Y_{i,j}$ is bounded), the bound is equal to $C^{-1}\varepsilon^{1/2}h_n^{d/2}$ for a constant C that depends only on the distribution of (X_i, Y_i).

This bound along with Lemma 8.1 in Kosorok (2008) implies

$$\begin{aligned}
&P\left(2^d \sup_{t\le t_0} \frac{\sqrt{n}}{h_n^{d/2}} |(E_n - E)Y_{i,j}I(h_n s_0 \le X_i - x_m \le h_n(s_0+t))| \ge w\right) \\
&= P\left(2^d \sup_{f\in\mathcal{F}_n} |\sqrt{n}(E_n - E)f(X_i,Y_i)| \ge w h_n^{d/2}\right) \\
&\le 2\exp\left(-\frac{w h_n^{d/2}}{\|2^d \sup_{f\in\mathcal{F}_n} |\sqrt{n}(E_n - E)f(X_i,Y_i)|\|_{P,\psi_1}}\right) \\
&\le 2\exp\left(-\frac{w h_n^{d/2}}{C^{-1}h_n^{d/2}\varepsilon^{1/2}}\right) = 2\exp\left(-Cw/\varepsilon^{1/2}\right).
\end{aligned}$$

The result follows using this and the fact that $|A| \le \left\{3B[B^d/(\varepsilon \wedge 1)] + 2\right\}^d$. □

The following theorem verifies the part of condition (ii) of Lemma C.1 concerning the limiting process $\mathbb{G}_{P,x_m,j}(s,t) + g_{P,x_m,j}(s,t)$.

Theorem C.2. *Fix m and j with $j \in J(m)$. For any $r < 0$, $\varepsilon > 0$ there exists an M such* *that*

$$P\left(\inf_{\|(s,t)\|>M} \mathbb{G}_{P,x_m,j}(s,t)+g_{P,x_m,j}(s,t)\le r\right)<\varepsilon.$$

Proof. Let $\mathbb{G}(s,t)=\mathbb{G}_{P,x_m,j}(s,t)$ and $g(s,t)=g_{P,x_m,j}(s,t)$. Let $S_k=\{k\le \|(s,t)\|\le k+1\}$ and let $S_k^L=S_k\cap\{\prod_i t_i\le (k+1)^{-\delta}\}$ for some fixed δ. By Lemma C.2,

$$\begin{aligned}P\left(\inf_{S_k^L}\mathbb{G}(s,t)+g(s,t)\le r\right)&\le P\left(\sup_{S_k^L}|\mathbb{G}(s,t)|\ge |r|\right)\\ &\le 2\left\{3(k+1)[(k+1)^d/k^{-\delta}]+2\right\}^{2d}\exp\left(-Cr^2(k+1)^{\delta}\right)\end{aligned}$$

for k large enough where C depends only on d. This bound is summable over k.

For any α and β with $\alpha<\beta$, let $S_k^{\alpha,\beta}=S_k\cap\{(k+1)^\alpha<\prod_i t_i\le (k+1)^\beta\}$. We have, for some $C_1>0$ that depends only on d and $V_j(x_m)$, $g(s,t)\ge C_1\|(s,t)\|^2\prod_i t_i$. (To see this, note that $g(s,t)$ is greater than or equal to a constant times $\int_{s_1}^{s_1+t_1}\cdots\int_{s_d}^{s_d+t_d}\|x\|^2dx_d\cdots dx_1=\left(\Pi_{i=1}^d t_i\right)\sum_{i=1}^d(s_i^2+t_i^2/3+s_it_i)$, and the sum can be bounded below by a constant times $\|(s,t)\|^2$ by minimizing over s_i for fixed t_i using calculus. The claimed expression for the integral follows from evaluating the inner integral to get an expression involving the integral for $d-1$, and then using induction.) Using this and Lemma C.2,

$$\begin{aligned}P\left(\inf_{S_k^{\alpha,\beta}}\mathbb{G}(s,t)+g(s,t)\le r\right)&\le P\left(\sup_{S_k^{\alpha,\beta}}|\mathbb{G}(s,t)|\ge C_1k^{2+\alpha}\right)\\ &\le 2\left\{3(k+1)[(k+1)^d/((k+1)^\beta\wedge 1)]+2\right\}^{2d}\exp\left(-CC_1^2\frac{k^{4+2\alpha}}{(k+1)^\beta}\right).\end{aligned}$$

This is summable over k if $4+2\alpha-\beta>0$.

Now, note that, since $\prod_i t_i\le (k+1)^d$ on S_k, we have, for any $-\delta<\alpha_1<\alpha_2<\ldots<\alpha_{\ell-1}<\alpha_\ell=d$, $S_k=S_k^L\cup S_k^{-\delta,\alpha_1}\cup S_k^{\alpha_1,\alpha_2}\cup\ldots\cup S_k^{\alpha_{\ell-1},\alpha_\ell}$. If we choose $\delta<3/2$ and $\alpha_i=i$ for $i\in\{1,\ldots,d\}$, the arguments above will show that the probability of the infimum being less than or equal to r over S_k^L, $S_k^{-\delta,\alpha_1}$ and each $S_k^{\alpha_i,\alpha_{i+1}}$ is summable over k, so that $P\left(\inf_{S_k}\mathbb{G}(s,t)+g(s,t)\le r\right)$ is summable over k, so setting M so that the tail of this sum past M is less than ε gives the desired result. □

The following theorem verifies condition (ii) of Lemma C.1 for the sequence of finite sample processes $\mathbb{G}_{n,x_m,j}(s,t)+g_{n,x_m,j}(s,t)$ with $\eta/h_n\ge\|(s,t)\|$. As explained above, the case where $\eta/h_n\le\|(s,t)\|$ is handled by a separate argument.

Theorem C.3. *Fix m and j with $j \in J(m)$. There exists an $\eta > 0$ such that for any $r < 0$, $\varepsilon > 0$, there exists an M and N such that, for all $n \geq N$,*

$$P\left(\inf_{M<\|(s,t)\|\leq \eta/h_n} \mathbb{G}_{n,x_m,j}(s,t) + g_{n,x_m,j}(s,t) \leq r\right) < \varepsilon.$$

Proof. Let $\mathbb{G}_n(s,t) = \mathbb{G}_{n,x_m,j}(s,t)$ and $g_n(s,t) = g_{n,x_m,j}(s,t)$. Let η be small enough that the assumptions hold for $\|x - x_m\| \leq 2\eta$ and that, for some constant C_2, $E(Y_{i,j}|X_i = x) \geq C_2\|x - x_m\|^2$ for $\|x - x_m\| \leq 2\eta$. This implies that, for $\|(s,t)\| \leq h_n^{-1}\eta$,

$$g_n(s,t) \geq \frac{C_2}{h_n^{d+2}} \int_{h_n s<x-x_m<h_n(s+t)} \|x - x_m\|^2 f_X(x)\, dx$$
$$\geq \frac{C_2 \underline{f}}{h_n^{d+2}} \int_{h_n s<x-x_m<h_n(s+t)} \|x - x_m\|^2\, dx = C_2\underline{f} \int_{s<x<s+t} \|x\|^2\, dx_d \cdots dx_1 \geq C_3\|(s,t)\|^2 \prod_i t_i$$

where C_3 is a constant that depends only on $\underline{f}$ and d and the last inequality follows from bounding the integral as explained in the proof of the previous theorem.

As in the proof of the previous theorem, let $S_k = \{k \leq \|(s,t)\| \leq k+1\}$ and let $S_k^L = S_k \cap \{\prod_i t_i \leq (k+1)^{-\delta}\}$ for some fixed δ. We have, using Lemma C.3,

$$P\left(\inf_{S_k^L} \mathbb{G}_n(s,t) + g_n(s,t) \leq r\right) \leq P\left(\sup_{S_k^L} |\mathbb{G}_n(s,t)| \geq |r|\right)$$
$$\leq 2\left\{3(k+1)[(k+1)^d/k^{-\delta}] + 2\right\}^{2d} \exp\left(-C\frac{|r|}{(k+1)^{-\delta/2}}\right)$$

for $(k+1)^{-\delta} \geq n^{-4/(d+4)}(1+\log n)^2 \iff k+1 \leq n^{4/[\delta(d+4)]}(1+\log n)^{-2/\delta}$ so, if $\delta < 4$, this will hold eventually for all $(k+1) \leq h_n^{-1}\eta$ (once $h_n^{-1}\eta \leq n^{4/[\delta(d+4)]}(1+\log n)^{-2/\delta} \iff \eta \leq n^{(4/\delta-1)/(d+4)}(1+\log n)^{-2/\delta}$). The bound is summable over k for any $\delta > 0$.

Again following the proof of the previous theorem, for $\alpha < \beta$, define $S_k^{\alpha,\beta} = S_k \cap \{(k+1)^\alpha < \prod_i t_i \leq (k+1)^\beta\}$. We have, again using Lemma C.3,

$$P\left(\inf_{S_k^{\alpha,\beta}} \mathbb{G}_n(s,t) + g_n(s,t) \leq r\right) \leq P\left(\sup_{S_k^{\alpha,\beta}} |\mathbb{G}_n(s,t)| \geq C_3 k^{2+\alpha}\right)$$
$$\leq 2\left\{3(k+1)[(k+1)^d/(k^\alpha \wedge 1)] + 2\right\}^{2d} \exp\left(-C\frac{C_3 k^{2+\alpha}}{(k+1)^{\beta/2}}\right)$$

for $(k+1)^\beta \geq n^{-4/(d+4)}$ (which will hold once the same inequality holds for δ for $-\delta < \beta$)

and $k+1 \leq h_n^{-1}\eta$. The bound is summable over k for any α, β with $4+2\alpha-\beta>0$.

Thus, noting as in the previous theorem that, for any $-\delta<\alpha_1<\alpha_2<\ldots<\alpha_{\ell-1}<\alpha_\ell=d$, $S_k=S_k^L \cup S_k^{-\delta,\alpha_1} \cup S_k^{\alpha_1,\alpha_2} \cup \ldots \cup S_k^{\alpha_{\ell-1},\alpha_\ell}$, if we choose $\delta<3/2$ and $\alpha_i=i$ for $i \in \{1,\ldots,d\}$ the probability of the infimum being less than or equal to r over the sets indexed by k for any $k \leq h_n^{-1}\eta$ is bounded uniformly in n by a sequence that is summable over k (once $\eta \leq n^{(4/\delta-1)/(d+4)}(1+\log n)^{-2/\delta}$). Thus, if we choose M such that the tail of this sum past M is less than ε and let N be large enough so that $\eta \leq N^{(4/\delta-1)/(d+4)}(1+\log N)^{-2/\delta}$, we will have the desired result.

□

To complete the proof of Theorem 3.1, we need to show that

$$Z_{n,2} \equiv \Bigg(n^{(d+2)/(d+4)} \inf_{\|(s-x_m,t)\|>\eta \text{ all } m \text{ s.t. } 1 \in J(m)} E_n Y_{i,1} I(s<X_i<s+t), \ldots,$$
$$n^{(d+2)/(d+4)} \inf_{\|(s-x_m,t)\|>\eta \text{ all } m \text{ s.t. } k \in J(m)} E_n Y_{i,k} I(s<X_i<s+t)\Bigg) \xrightarrow{p} 0.$$

This follows from the next two lemmas.

Lemma C.4. *Under Assumptions 3.1 and 3.2, for any $\eta>0$, there exists some $\underline{B}>0$ such that $EY_{i,j}I(s<X_i<s+t) \geq \underline{B}P(s<X_i<s+t)$ for all (s,t) with $\|(s-x_m,t)\|>\eta$ for all m with $j \in J(m)$.*

Proof. Given $\eta>0$, we can make η smaller without weakening the result, so let η be small enough that $\|x_m-x_r\|_\infty>2\eta$ for all $m \neq r$ with $j \in J(m) \cap J(r)$ and f_X satisfies $0<\underline{f} \leq f_X(x) \leq \overline{f}<\infty$ for some $\overline{f}$ and $\underline{f}$ on $\{x | \|x-x_m\|_\infty \leq \eta\}$. If $\|(s-x_m,t)\|>\eta$, then $\|(s-x_m,s+t-x_m)\|_\infty>\eta/(4d)$, so it suffices to show that $EY_{i,j}I(s<X_i<s+t) \geq \underline{B}P(s<X_i<s+t)$ for all (s,t) with $\|(s-x_m,s+t-x_m)\|_\infty>\eta/(4d)$. Let $\underline{\mu}>0$ be such that $E(Y_{i,j}|X_i=x)>\underline{\mu}$ when $\|x-x_m\|_\infty \geq \eta/(8d)$ for m with $j \in J(m)$. For notational convenience, let $\delta=\eta/(4d)$.

For m with $j \in J(m)$, let $B(x_m,\delta)=\{x | \|x-x_m\|_\infty \leq \delta\}$ and $B(x_m,\delta/2)=\{x | \|x-x_m\|_\infty \leq \delta/2\}$. First, I show that, for any (s,t) with $\|(s-x_m,s+t-x_m)\|_\infty \geq \delta$, $P(\{s<X_i<s+t\} \cap B(x_m,\delta) \backslash B(x_m,\delta/2)) \geq (1/3)(\underline{f}/\overline{f})P(\{s<X_i<s+t\} \cap B(x_m,\delta/2))$. Intuitively, this holds because, taking any box with a corner outside of $B(x_m,\delta)$, this box has to intersect with a substantial proportion of $B(x_m,\delta) \backslash B(x_m,\delta/2)$ in order to intersect with $B(x_m,\delta/2)$.

Formally, we have $\{s<x<s+t\} \cap B(x_m,\delta)=\{s \vee (x_m-\delta)<x<(s+t) \wedge (x_m+\delta)\}$, so that, letting λ be the Lebesgue measure on $\mathbb{R}^d$, $\lambda(\{s<x<s+t\} \cap B(x_m,\delta))=\prod_i [(s_i+t_i) \wedge$

$(x_{m,i}+\delta)-s_i\vee(x_{m,i}-\delta)]$. Similarly, $\lambda(\{s<x<s+t\}\cap B(x_m,\delta/2))=\prod_i[(s_i+t_i)\wedge(x_{m,i}+\delta/2)-s_i\vee(x_{m,i}-\delta/2)]$. For all i, $[(s_i+t_i)\wedge(x_{m,i}+\delta/2)-s_i\vee(x_{m,i}-\delta/2)]\leq[(s_i+t_i)\wedge(x_{m,i}+\delta)-s_i\vee(x_{m,i}-\delta)]$. For some r, we must have $s_r\leq x_{m,r}-\delta$ or $s_r+t_r\geq x_{m,r}+\delta$. For this r, we will have $[(s_r+t_r)\wedge(x_{m,r}+\delta/2)-s_r\vee(x_{m,r}-\delta/2)]\leq 2[(s_r+t_r)\wedge(x_{m,r}+\delta)-s_r\vee(x_{m,r}-\delta)]/3$. Thus, $\lambda(\{s<x<s+t\}\cap B(x_m,\delta/2))\leq 2\lambda(\{s<x<s+t\}\cap B(x_m,\delta))/3$. It then follows that $\lambda(\{s<x<s+t\}\cap B(x_m,\delta)\backslash B(x_m,\delta/2))\geq(1/3)\lambda(\{s<x<s+t\}\cap B(x_m,\delta))$, so that $P(\{s<x<s+t\}\cap B(x_m,\delta)\backslash B(x_m,\delta/2))\geq(1/3)(\underline{f}/\overline{f})P(\{s<x<s+t\}\cap B(x_m,\delta))$.

Now, we use the fact that $E(Y_{i,j}|X_i)$ is bounded away from zero outside of $B(x_m,\delta/2)$, and that the proportion of $\{s<x<s+t\}$ that intersects with $B(x_m,\delta/2)$ can't be too large. We have, for any (s,t) with $\|(s-x_m,s+t-x_m)\|_\infty\geq\delta$,

$$
\begin{aligned}
&EY_{i,j}I(s<X_i<s+t)\geq\underline{\mu}P(\{s<X_i<s+t\}\backslash[\cup_m B(x_m,\delta/2)])\\
&=\underline{\mu}P(\{s<X_i<s+t\}\backslash[\cup_m B(x_m,\delta)])+\underline{\mu}\sum_m P(\{s<X_i<s+t\}\cap B(x_m,\delta)\backslash B(x_m,\delta/2))\\
&\geq\underline{\mu}P(\{s<X_i<s+t\}\backslash[\cup_m B(x_m,\delta)])+\underline{\mu}\sum_m(1/3)(\underline{f}/\overline{f})P(\{s<X_i<s+t\}\cap B(x_m,\delta))\\
&\geq\underline{\mu}(1/3)(\underline{f}/\overline{f})P(s<X_i<s+t)
\end{aligned}
$$

where the unions are taken over m such that $j\in J(m)$. The equality in the second line follows because the sets $B(x_m,\delta)$ are disjoint.

□

Lemma C.5. *Let S be any set in $\mathbb{R}^{2d}$ such that, for some $\underline{\mu}>0$ and all $(s,t)\in S$, $EY_{i,j}I(s<X_i<s+t)\geq\underline{\mu}P(s<X_i<s+t)$. Then, under Assumption 3.2, for any sequence $a_n\to\infty$ and $r<0$,*

$$\inf_{(s,t)\in S}\frac{n}{a_n\log n}E_nY_{i,j}I(s<X_i<s+t)>r$$

with probability approaching 1.

Proof. For $(s,t) \in S$,

$$
\begin{aligned}
&\frac{n}{a_n \log n} E_n Y_{i,j} I(s < X_i < s+t) \le r \\
&\Longrightarrow \frac{n}{a_n \log n}(E_n - E) Y_{i,j} I(s < X_i < s+t) \le r - \frac{n}{a_n \log n} E Y_{i,j} I(s < X_i < s+t) \\
&\le r - \frac{n}{a_n \log n} \underline{\mu} P(s < X_i < s+t) \le -\left\{ |r| \vee \left[\frac{n}{a_n \log n} \underline{\mu} P(s < X_i < s+t) \right] \right\} \\
&\Longrightarrow \left[\frac{\frac{a_n \log n}{n}}{\frac{a_n \log n}{n} \vee P(s < X_i < s+t)} \right]^{1/2} |(E_n - E) Y_{i,j} I(s < X_i < s+t)| \\
&\ge \left[\frac{\frac{a_n \log n}{n}}{\frac{a_n \log n}{n} \vee P(s < X_i < s+t)} \right]^{1/2} \left\{ \left[\frac{a_n \log n}{n} |r| \right] \vee \left[\underline{\mu} P(s < X_i < s+t) \right] \right\}.
\end{aligned}
$$

If $\frac{a_n \log n}{n} \ge P(s < X_i < s+t)$, then the last line is greater than or equal to $\frac{a_n \log n}{n}|r|$. If $\frac{a_n \log n}{n} \le P(s < X_i < s+t)$, the last line is greater than or equal to $\left[\frac{\frac{a_n \log n}{n}}{P(s<X_i<s+t)} \right]^{1/2} \underline{\mu} P(s < X_i < s+t) = \left(\frac{a_n \log n}{n} \right)^{1/2} \underline{\mu} \sqrt{P(s < X_i < s+t)} \ge \underline{\mu} \frac{a_n \log n}{n}$. Thus,

$$
\begin{aligned}
&P\left(\inf_{(s,t)\in S} \frac{n}{a_n \log n} E_n Y_{i,j} I(s < X_i < s+t) \le r \right) \\
&\le P\left(\sup_{(s,t)\in S} \left[\frac{\frac{a_n \log n}{n}}{\frac{a_n \log n}{n} \vee P(s < X_i < s+t)} \right]^{1/2} |(E_n - E) Y_{i,j} I(s < X_i < s+t)| \ge (|r| \wedge \underline{\mu}) \frac{a_n \log n}{n} \right).
\end{aligned}
$$

This converges to zero by Theorem 37 in Pollard (1984) with, in the notation of that theorem, $\mathcal{F}_n$ the class of functions of the form

$$
\left[\frac{\frac{a_n \log n}{n}}{\overline{Y}^2 \frac{a_n \log n}{n} \vee P(s < X_i < s+t)} \right]^{1/2} Y_{i,j} I(s < X_i < s+t)
$$

with $(s,t) \in S$, $\delta_n = \left(\frac{n}{a_n \log n} \right)^{1/2}$ and $\alpha_n = 1$. To verify the conditions of the lemma, the covering number bound holds because each $\mathcal{F}_n$ is contained in the larger class $\mathcal{F}$ of functions of the form $w Y_{i,j} I(s < X_i < s+t)$ where (s,t) ranges over S and w ranges over $\mathbb{R}$, and this larger class is a VC subgraph class. The supremum bound on functions in $\mathcal{F}_n$ holds by

Assumption 3.2. To verify the bound on the L^2 norm of functions in $\mathcal{F}_n$, note that

$$E\left\{\left[\frac{\frac{a_n \log n}{n}}{\overline{Y}^2 \frac{a_n \log n}{n} \vee P(s < X_i < s+t)}\right]^{1/2} Y_{i,j} I(s < X_i < s+t)\right\}^2$$
$$\leq \frac{\frac{a_n \log n}{n}}{\frac{a_n \log n}{n} \vee P(s < X_i < s+t)} P(s < X_i < s+T) \leq \frac{a_n \log n}{n} = \delta_n^2$$

since $ab/(a \vee b) \leq a$ for any $a, b > 0$.

□

By Lemma C.4, $\{\|(s - x_m, t)\| > \eta \text{ all } m \text{ s.t. } j \in J(m)\}$ satisfies the conditions of Lemma C.5, so $E_n Y_{i,j} I(s < X_i < s+t)$ converges to zero at a $n/(a_n \log n)$ rate for any $a_n \to \infty$, which can be made faster than the $n^{(d+2)/(d+4)}$ rate needed to show that $Z_{n,2} \xrightarrow{p} 0$. This completes the proof of Theorem 3.1.

C.2 Inference

I use the following lemma in the proof of Theorem 4.1

Lemma C.6. *Let $\mathbb{H}$ be a Gaussian random process with sample paths that are almost surely in the set $C(\mathbb{T}, \mathbb{R}^k)$ of continuous functions with respect to some semimetric on the index set $\mathbb{T}$ with a countable dense subset $\mathbb{T}_0$. Then, for any set $A \in \mathbb{R}^k$ with Lebesgue measure zero, $P(\inf_{t \in \mathbb{T}} \mathbb{H}(t) \in A) \leq P(\inf_{t \in \mathbb{T}, \det var(\mathbb{H}(t)) < \varepsilon} \mathbb{H}(t) \in A$ for all $\varepsilon > 0)$.*

Proof. First, note that, if the infimum over $\mathbb{T}$ is in A, then, since $\{t \in \mathbb{T} | \det var(\mathbb{H}(t)) \geq \varepsilon\}$ and $\{t \in \mathbb{T} | \det var(\mathbb{H}(t)) < \varepsilon\}$ partition T, the infimum over one of these sets must be in A. By Proposition 3.2 in Pitt and Tran (1979), the infimum of $\mathbb{H}(t)$ over the former set has a distribution that is continuous with respect to the Lebesgue measure, so the probability of the infimum of $\mathbb{H}(t)$ over this set being in A is zero. Thus, $P\left(\inf_{t \in \mathbb{T}} \mathbb{H}(t) \in A\right) \leq P\left(\inf_{t \in \mathbb{T}, \det var(\mathbb{H}(t)) < \varepsilon} \mathbb{H}(t) \in A\right)$. Taking ε to zero along a countable sequence gives the result.

□

Proof of Theorem 4.1. For m from 1 to ℓ, let $\{j_{m,1}, \ldots, j_{m,|J(m)|}\} = J(m)$. Then, letting

$$\tilde{Z} \equiv (\inf_{s,t} \mathbb{G}_{P,x_1,j_{1,1}}(s,t) + g_{P,x_1,j_{1,1}}(s,t), \ldots, \inf_{s,t} \mathbb{G}_{P,x_1,j_{1,|J(1)|}}(s,t) + g_{P,x_1,j_{1,|J(1)|}}(s,t), \ldots,$$
$$\inf_{s,t} \mathbb{G}_{P,x_\ell,j_{\ell,1}}(s,t) + g_{P,x_\ell,j_{\ell,1}}(s,t), \ldots, \inf_{s,t} \mathbb{G}_{P,x_\ell,j_{\ell,|J(\ell)|}}(s,t) + g_{P,x_\ell,j_{\ell,|J(\ell)|}}(s,t)),$$

each element of Z is the minimum of the elements of some subvector of $\tilde{Z}$, where the subvectors corresponding to different elements of Z do not overlap. Thus, it suffices to show that $\tilde{Z}$ has an absolutely continuous distribution. For this, it suffices to show that, for each m,

$$(\inf_{s,t} \mathbb{G}_{P,x_m,j_{m,1}}(s,t) + g_{P,x_m,j_{m,1}}(s,t), \ldots, \inf_{s,t} \mathbb{G}_{P,x_m,j_{m,|J(m)|}}(s,t) + g_{P,x_m,j_{m,|J(m)|}}(s,t))$$

has an absolutely continuous distribution, since these are independent across m.

To this end, fix m and let $\mathbb{H}(s,t)$ be the random process with sample paths in $C(\mathbb{R}^{2d}, \mathbb{R}^{|J(m)|})$ defined by

$$\mathbb{H}(s,t) = (\mathbb{G}_{P,x_m,j_{m,1}}(s,t) + g_{P,x_m,j_{m,1}}(s,t), \ldots, \mathbb{G}_{P,x_m,j_{m,|J(m)|}}(s,t) + g_{P,x_m,j_{m,|J(m)|}}(s,t)).$$

By Assumption 4.1 $var(\mathbb{H}(s,t)) = M \prod_i t_i$ for some positive definite matrix M, so that $\det var(\mathbb{H}(s,t)) = (\det M) \left(\prod_i t_i\right)^{|J(m)|}$. Thus, $\inf_{(s,t)\in\mathbb{R}^{2d}, \det var(\mathbb{H}(s,t))<\varepsilon} \mathbb{H}(s,t) \in A$ for all $\varepsilon > 0$ iff. $\inf_{(s,t)\in\mathbb{R}^{2d}, \prod_i t_i<\varepsilon} \mathbb{H}(s,t) \in A$ for all $\varepsilon > 0$ so, by Lemma C.6, $P(\inf_{(s,t)\in\mathbb{R}^{2d}} \mathbb{H}(s,t) \in A) \leq P(\inf_{(s,t)\in\mathbb{R}^{2d}, \prod_i t_i<\varepsilon} \mathbb{H}(s,t) \in A$ for all $\varepsilon > 0)$. For each j, $\prod_i t_i$ is equal to $var(\mathbb{H}_j(s,t)) = \rho_j(0,(s,t))$ times some constant, where ρ_j is the covariance semimetric for component j given by $\rho_j((s,t),(s',t')) = var(\mathbb{H}_j(s,t) - \mathbb{H}_j(s',t'))$. Thus, there exists a constant C such that $\prod_i t_i \leq \varepsilon$ implies $\rho_j(0,(s,t)) < C\varepsilon$ for all j, so that $P(\inf_{(s,t)\in\mathbb{R}^{2d}} \mathbb{H}(s,t) \in A) \leq P(\inf_{(s,t)\in\mathbb{R}^{2d}, \rho_j(0,(s,t))<C\varepsilon \text{ all } j} \mathbb{H}(s,t) \in A$ for all $\varepsilon > 0)$.

Since the sample paths of $\mathbb{H}$ are almost surely continuous with respect to the semimetric $\max_j \rho_j((s,t),(s',t'))$ on the set $\|(s,t)\| \leq M$ for any finite M, $\inf_{\|(s,t)\|\leq M, \rho_j(0,(s,t))<C\varepsilon \text{ all } j} \mathbb{H}(s,t) \in A$ for all $\varepsilon > 0$ implies that $\mathbb{H}(0) = 0$ is a limit point of A on this probability one set. Thus, for any set A that does not have zero as a limit point, $P(\inf_{\|(s,t)\|\leq M} \mathbb{H}(s,t) \in A) = 0$ for any finite M. Applying this to $A \backslash B_\eta(0)$ where $B_\eta(0)$ is the η-ball around 0 in $\mathbb{R}^{|J(m)|}$, we have

$$P\left(\inf_{(s,t)\in\mathbb{R}^{2d}} \mathbb{H}(s,t) \in A\right) = P\left(\inf_{(s,t)\in\mathbb{R}^{2d}} \mathbb{H}(s,t) \in A \cap B_\eta(0)\right) + P\left(\inf_{(s,t)\in\mathbb{R}^{2d}} \mathbb{H}(s,t) \in A\backslash B_\eta(0)\right)$$
$$\leq P\left(\inf_{(s,t)\in\mathbb{R}^{2d}} \mathbb{H}(s,t) \in A \cap B_\eta(0)\right) + P\left(\inf_{\|(s,t)\|\leq M} \mathbb{H}(s,t) \in A\backslash B_\eta(0)\right)$$
$$+ P\left(\inf_{\|(s,t)\|>M} \mathbb{H}(s,t) \in A\backslash B_\eta(0)\right)$$
$$= P\left(\inf_{(s,t)\in\mathbb{R}^{2d}} \mathbb{H}(s,t) \in A \cap B_\eta(0)\right) + P\left(\inf_{\|(s,t)\|>M} \mathbb{H}(s,t) \in A\backslash B_\eta(0)\right).$$

Noting that $P\left(\inf_{\|(s,t)\|>M} \mathbb{H}(s,t) \in A\backslash B_\eta(0)\right)$ can be made arbitrarily small by making M

large, this shows that $P\left(\inf_{(s,t)\in\mathbb{R}^{2d}}\mathbb{H}(s,t)\in A\right) = P\left(\inf_{(s,t)\in\mathbb{R}^{2d}}\mathbb{H}(s,t)\in A\cap B_\eta(0)\right)$ Taking η to zero along a countable sequence, this shows that $P\left(\inf_{(s,t)\in\mathbb{R}^{2d}}\mathbb{H}(s,t)\in A\right) \leq P\left(\inf_{(s,t)\in\mathbb{R}^{2d}}\mathbb{H}(s,t)\in A\cap\{0\}\right)$ so that $\inf_{(s,t)\in\mathbb{R}^{2d}}\mathbb{H}(s,t)$ has an absolutely continuous distribution with a possible atom at zero.

To show that there can be no atom at zero, we argue as follows. Fix $j\in J(m)$. The component of $\mathbb{H}$ corresponding to this j is $\mathbb{G}_{P,x_m,j}(s,t)+g_{P,x_m,j}(s,t)$. For some constant K, for any $k\geq 0$, letting $s_{i,k}=(i/k,0,\ldots,0)$ and $t_k=(1/k,1,\ldots,1)$, we will have $g_{P,x_m,j}(s_{i,k},t_k)\leq K/k$ for $i\leq k$, so that

$$\begin{aligned}
&P\left(\inf_{(s,t)\in\mathbb{R}^{2d}}\mathbb{G}_{P,x_m,j}(s,t)+g_{P,x_m,j}(s,t)=0\right)=P\left(\inf_{(s,t)\in\mathbb{R}^{2d}}\mathbb{G}_{P,x_m,j}(s,t)+g_{P,x_m,j}(s,t)\geq 0\right)\\
&\leq P\left(\mathbb{G}_{P,x_m,j}(s_{i,k},t_k)+g_{P,x_m,j}(s_{i,k},t_k)\geq 0 \text{ all } i\in\{0,\ldots,k\}\right)\\
&\leq P\left(\mathbb{G}_{P,x_m,j}(s_{i,k},t_k)+K/k\geq 0 \text{ all } i\in\{0,\ldots,k\}\right)\\
&= P\left(\sqrt{k}\mathbb{G}_{P,x_m,j}(s_{i,k},t_k)+K/\sqrt{k}\geq 0 \text{ all } i\in\{0,\ldots,k\}\right)\\
&= P\left(\mathbb{G}_{P,x_m,j}(s_{i,1},t_1)+K/\sqrt{k}\geq 0 \text{ all } i\in\{0,\ldots,k\}\right).
\end{aligned}$$

The final line is the probability of $k+1$ iid normal random variables each being greater than or equal to $-K/\sqrt{k}$, which can be made arbitrarily small by making k large. □

proof of Theorem 4.2. This follows immediately from the continuity of the asymptotic distribution (see Politis, Romano, and Wolf, 1999). □

C.3 Other Shapes of the Conditional Mean

This section contains the proofs of the results in Section 5, which extend the results of Section 3 to other shapes of the conditional mean. First, I show how Assumption 3.1 implies Assumption 5.1 with $\gamma=2$. Next, I prove Theorem 5.1, which gives an interpretation of Assumption 5.2 in terms of conditions on the number of bounded derivatives in the one dimensional case. Finally, I prove Theorem 5.2, which derives the asymptotic distribution of the KS statistic under these assumptions. The proof is mostly the same as the proof of Theorem 3.1, and I present only the parts of the proof that differ, referring to the proof of Theorem 3.1 for the parts that do not need to be changed.

To see that, under part (ii) from Assumption 3.1, Assumption 5.1 will hold with $\gamma=2$, note that, by a second order Taylor expansion, for some $x^*(x)$ between x and x_k,

$$\frac{\bar{m}_j(\theta,x)-\bar{m}_j(\theta,x_k)}{\|x-x_k\|^2}=\frac{(x-x_k)V_j(x^*(x))(x-x_k)}{2\|x-x_k\|^2}=\frac{1}{2}\frac{x-x_k}{\|x-x_k\|}V_j(x^*(x))\frac{x-x_k}{\|x-x_k\|}.$$

Thus, letting $\psi_{j,k}(t)=\frac{1}{2}tV_j(x_k)t$ we have

$$\sup_{\|x-x_k\|\le\delta}\left\|\frac{\bar{m}_j(\theta,x)-\bar{m}_j(\theta,x_k)}{\|x-x_k\|^2}-\psi_{j,k}\left(\frac{x-x_k}{\|x-x_k\|}\right)\right\|$$
$$=\sup_{\|x-x_k\|\le\delta}\left\|\frac{1}{2}\frac{x-x_k}{\|x-x_k\|}V_j(x^*(x))\frac{x-x_k}{\|x-x_k\|}-\frac{1}{2}\frac{x-x_k}{\|x-x_k\|}V_j(x_k)\frac{x-x_k}{\|x-x_k\|}\right\|.$$

This goes to zero as $\delta\to 0$ by the continuity of the second derivative matrix.

The proof of Theorem 5.1 below shows that, in the one dimensional case, Assumption 3.1 follows more generally from conditions on higher order derivatives.

proof of Theorem 5.1. It suffices to consider the case where $d_Y=1$. First, suppose that $\mathcal{X}_0$ has infinitely many elements. Let $\{x_k\}_{k=1}^\infty$ be a nonrepeating sequence of elements in $\mathcal{X}_0$. Since $\mathcal{X}_0$ is compact, this sequence must have a subsequence that converges to some $\tilde{x}\in\mathcal{X}_0$. If $\bar{m}(\theta,x)$ had a nonzero rth derivative at $\tilde{x}$ for some $r<p$, then, by Lemma C.7 below, $\bar{m}(\theta,x)$ would be strictly greater than $\bar{m}(\theta,\tilde{x})$ for x in some neighborhood of $\tilde{x}$, a contradiction. Thus, a pth order taylor expansion gives, using the notation $D_r(x)=\delta^r/\delta x^r\bar{m}(\theta,x)$ for $r\le p$, $\bar{m}(\theta,x)-\bar{m}(\theta,\tilde{x})=D_p(x^*(x))(x-\tilde{x})^p/p!\le\bar{D}|x-\tilde{x}|^p/p!$ where $\bar{D}$ is a bound on the pth derivative and $x^*(x)$ is some value between x and $\tilde{x}$.

If $\mathcal{X}_0$ has finitely many elements, then, for each $x_0\in\mathcal{X}_0$, a pth order Taylor expansion gives $\bar{m}(\theta,x)-\bar{m}(\theta,x_0)=D_1(x_0)(x-x_0)+\frac{1}{2}D_2(x_0)(x-x_0)^2+\cdots+\frac{1}{p!}D_p(x^*(x))(x-x_0)^p$. If, for some $r<p$, $D_r(x_0)\neq 0$ and $D_{r'}(x_0)=0$ for $r'<r$, then Assumption 5.1 will hold at x_0 with $\gamma=r$. If not, we will have $\bar{m}(\theta,x)-\bar{m}(\theta,x_0)\le\bar{D}|x-x_0|^p/p!$ for all x. □

Lemma C.7. *Suppose that $g:[\underline{x},\overline{x}]\subseteq\mathbb{R}\to\mathbb{R}$ is minimized at some x_0. If the least nonzero derivative of g is continuous at x_0, then, for some $\varepsilon>0$, $g(x)>g(x_0)$ for $|x-x_0|\le\varepsilon$, $x\neq x_0$.*

Proof. Let p be the least integer such that the pth derivative $g^{(p)}(x_0)$ is nonzero. By a pth order Taylor expansion, $g(x)-g(x_0)=g^{(p)}(x^*(x))(x-x_0)^p$ for some $x^*(x)$ between x and x_0. By continuity of $g^{(p)}(x)$, $|g^{(p)}(x^*(x))-g^{(p)}(x_0)|>|g^{(p)}(x_0)|/2$ for x close enough to x_0, so that $g(x)-g(x_0)=g^{(p)}(x^*(x))(x-x_0)^p\ge|g^{(p)}(x_0)|/2|x-x_0|^p>0$ (the pth derivative must have the same sign as $x-x_0$ if p is odd in order for g to be minimized at x_0). □

I now prove Theorem 5.2. I prove the theorem under the assumption that $\gamma(j,k) = \gamma$ for all (j,k) with $j \in J(k)$. The general case follows from applying the argument to neighborhoods of each x_k, and getting faster rates of convergence for (j,k) such that $\gamma(j,k) < \gamma$. The proof is the same as the proof of Theorem 3.1 with the following modifications.

First, Theorem C.1 must be modified to the following theorem, with the new definition of $g_{P,x_k,j}(s,t)$.

Theorem C.4. *Let $h_n = n^{-\beta}$ for some $0 < \beta < 1/d_X$. Let*

$$\mathbb{G}_{n,x_m}(s,t) = \frac{\sqrt{n}}{h_n^{d/2}}(E_n - E)Y_{i,J(m)}I(h_n s < X_i - x_m < h_n(s+t))$$

and let $g_{n,x_m}(s,t)$ have jth element

$$g_{n,x_m,j}(s,t) = \frac{1}{h_n^{d_X+\gamma}} EY_{i,j}I(h_n s < X_i - x_m < h_n(s+t))$$

if $j \in J(m)$ and zero otherwise. Then, for any finite M, $(\mathbb{G}_{n,x_1}(s,t), \ldots, \mathbb{G}_{n,x_\ell}(s,t)) \xrightarrow{d} (\mathbb{G}_{P,x_1}(s,t), \ldots, \mathbb{G}_{P,x_\ell}(s,t))$ taken as random processes on $\|(s,t)\| \le M$ with the supremum norm and $g_{n,x_m}(s,t) \to g_{P,x_m}(s,t)$ uniformly in $\|(s,t)\| \le M$ where $\mathbb{G}_{P,x_m}(s,t)$ and $g_{P,x_m}(s,t)$ are defined as in Theorem 3.1 for m from 1 to ℓ.

Proof. The proof of the first display is the same. For the proof of the claim regarding $g_{n,x_m}(s,t)$, we have

$$\begin{aligned} g_{n,x_m,j}(s,t) = &\frac{1}{h_n^{d_X+\gamma}} \int_{h_n s < x - x_m < h_n(s+t)} \psi_{j,k}\left(\frac{x-x_m}{\|x-x_m\|}\right) \|x-x_m\|^\gamma f_X(x_m)\, dx \\ &+ \frac{1}{h_n^{d_X+\gamma}} \int_{h_n s < x - x_m < h_n(s+t)} \psi_{j,k}\left(\frac{x-x_m}{\|x-x_m\|}\right) \|x-x_m\|^\gamma [f_X(x) - f_X(x_m)]\, dx \\ &+ \frac{1}{h_n^{d_X+\gamma}} \int_{h_n s < x - x_m < h_n(s+t)} [\bar{m}_j(\theta,x) - \bar{m}_j(\theta,x_m) \\ &- \psi_{j,k}\left(\frac{x-x_m}{\|x-x_m\|}\right) \|x-x_m\|^\gamma \Bigg] f_X(x)\, dx. \end{aligned}$$

The first term is equal to $g_{P,x_m,j}(s,t)$ by a change of variable x to $h_n x + x_m$ in the integral. The second term is bounded by $g_{P,x_m,j}(s,t)\sup_{\|x-x_m\|\le 2h_n M}[f_X(x) - f_X(x_m)]/f_X(x_m)$, which goes to zero uniformly in $\|(s,t)\| \le M$ by continuity of f_X. The third term is equal to (using the same change of variables)

$$\int_{s<x<s+t}\left[\frac{\bar{m}_j(\theta,h_nx+x_m)-\bar{m}_j(\theta,x_m)}{h_n^\gamma}-\psi_{j,k}\left(\frac{x}{\|x\|}\right)\|x\|^\gamma\right]f_X(x)\,dx$$
$$=\int_{s<x<s+t}\|x\|^\gamma\left[\frac{\bar{m}_j(\theta,h_nx+x_m)-\bar{m}_j(\theta,x_m)}{\|h_nx\|^\gamma}-\psi_{j,k}\left(\frac{x}{\|x\|}\right)\right]f_X(x)\,dx.$$

For $\|(s,t)\|\leq M$, this is bounded by a constant times

$$\sup_{\|x\|\leq 2M}\left\|\frac{\bar{m}_j(\theta,h_nx+x_m)-\bar{m}_j(\theta,x_m)}{\|h_nx\|^\gamma}-\psi_{j,k}\left(\frac{x}{\|x\|}\right)\right\|,$$

which goes to zero as $n\to\infty$ by Assumption 5.1. □

The drift term and the mean zero term will be of the same order of magnitude if $\sqrt{n}/h_n^{d_X/2}=1/h_n^{d_X+\gamma}\Leftrightarrow h_n=n^{-1/(d_X+2\gamma)}$, so that

$$n^{(d_X+\gamma)/(d+2\gamma)}(E_nY_{i,J(1)}I(h_ns<X-x_1<h_n(s+t)),\ldots,E_nY_{i,J(\ell)}I(h_ns<X-x_\ell<h_n(s+t))$$
$$=(\mathbb{G}_{n,x_1}(s,t)+g_{n,x_1}(s,t),\ldots,\mathbb{G}_{n,x_\ell}(s,t)+g_{n,x_\ell}(s,t))$$
$$\stackrel{d}{\to}(\mathbb{G}_{P,x_1}(s,t)+g_{P,x_1}(s,t),\ldots,\mathbb{G}_{P,x_m}(s,t)+g_{P,x_m}(s,t))$$

taken as stochastic processes in $\{\|(s,t)\|\leq M\}$ with the supremum norm. From now on, let $h_n=n^{-1/(d+2\gamma)}$ so that this will hold.

Lemmas C.2 and C.3 hold as stated, except for the condition in Lemma C.3 that $\varepsilon\geq n^{-4/(d+4)}(1+\log n)^2$ must be replaced by $\varepsilon\geq n^{2\gamma/(d+2\gamma)}(1+\log n)^2$ so that $h_n^{d/2}2^{d/2}\varepsilon^{1/2}\geq n^{-1/2}(1+\log n)$, which implies the fourth inequality in the last display in the proof of this lemma, holds for the sequence h_n in the general case.

The next part of the proof that needs to be modified is the proofs of Theorems C.2 and C.3. For this, note that, for some constants C_1 and $\eta>0$

$$g_{P,x_m,j}(s,t)\geq C_1\|(s,t)\|^\gamma\prod_i t_i \tag{2}$$

and, for $\|(s,t)\|\leq\eta/h_n$,

$$g_{n,x_m,j}(s,t)\geq C_1\|(s,t)\|^\gamma\prod_i t_i \tag{3}$$

for all m and j. To see this, note that

$$g_{n,x_m,j}(s,t) = E\frac{1}{h_n^{d_X+\gamma}}EY_{i,j}I(h_n s < X_i - x_m < h_n(s+t))$$
$$= \frac{1}{h_n^{d_X+\gamma}}\int_{h_n s < x - x_m < h_n(s+t)} \bar{m}(\theta, x) f_X(x)\, dx = \int_{s<x<s+t} \frac{\bar{m}(\theta, h_n x + x_m)}{\|h_n x\|^\gamma}\|x\|^\gamma f_X(h_n x + x_m)\, dx$$

where the last equality follows from the change of variables x to $h_n x + x_m$. For small enough η, this is greater than or equal to $\frac{1}{2}\int_{s<x<s+t}\underline{\psi}\|x\|^\gamma f_X(x_m)\, dx$ for $\|(s,t)\| \leq \eta/h_n$ by Assumption 5.1 and the continuity of f_X. By definition, $g_{P,x_m,j}(s,t)$ is also greater than or equal to a constant times $\int_{s<x<s+t}\|x\|^\gamma\, dx$. To see that this is greater than or equal to a constant times $\|(s,t)\|^\gamma \prod_i t_i$, note that the Euclidean norm is equivalent to the norm $(s,t) \mapsto \max_i \max\{|s_i|, |s_i+t_i|\}$ and let i^* be an index such that $|s_{i^*}| = \max_i \max\{|s_i|, |s_i+t_i|\}$ or $|s_{i^*}+t_{i^*}| = \max_i \max\{|s_i|, |s_i+t_i|\}$. In the former case, we will have $\|x\| \geq |s_{i^*}|/2$ for x on the set $\{s_{i^*} \leq x_{i^*} \leq s_{i^*} + |s_{i^*}|/2\} \cap \{s < x < s+t\}$, which has Lebesgue measure $\left(\prod_{i\neq i^*} t_i\right)\cdot |s_{i^*}|/2 \geq \left(\prod_{i\neq i^*} t_i\right)\cdot t_{i^*}/4$, so that $\int_{s<x<s+t}\|x\|^\gamma\, dx \geq (\max_i \max\{|s_i|, |s_i+t_i|\}/2)^\gamma \prod_i t_i/4$, and a symmetric argument holds in the latter case.

With these inequalities in hand, the modified proofs of Theorems C.2 and C.3 are as follows.

proof of Theorem C.2 for general case. Let $\mathbb{G}(s,t) = \mathbb{G}_{P,x_m,j}(s,t)$ and $g(s,t) = g_{P,x_m,j}(s,t)$. Let $S_k = \{k \leq \|(s,t)\| \leq k+1\}$ and let $S_k^L = S_k \cap \{\prod_i t_i \leq (k+1)^{-\delta}\}$ for some fixed δ. By Lemma C.2,

$$P\left(\inf_{S_k^L} \mathbb{G}(s,t) + g(s,t) \leq r\right) \leq P\left(\sup_{S_k^L} |\mathbb{G}(s,t)| \geq |r|\right)$$
$$\leq \left\{3(k+1)[(k+1)^d/k^{-\delta}] + 2\right\}^{2d} \exp\left(-Cr^2(k+1)^\delta\right)$$

for k large enough where C depends only on d. Thus, the infimum over each S_k^L is summable over k.

For any $\underline{\beta}$ and $\overline{\beta}$ with $\underline{\beta} < \overline{\beta}$, let $S_k^{\underline{\beta},\overline{\beta}} = S_k \cap \{(k+1)^{\underline{\beta}} < \prod_i t_i \leq (k+1)^{\overline{\beta}}\}$. Using Lemma

C.2 and (2),

$$P\left(\inf_{S_k^{\underline{\beta},\overline{\beta}}} \mathbb{G}(s,t) + g(s,t) \le r\right) \le P\left(\sup_{S_k^{\underline{\beta},\overline{\beta}}} |\mathbb{G}(s,t)| \ge C_1 k^{\gamma+\underline{\beta}}\right)$$
$$\le \left\{3(k+1)[(k+1)^d/((k+1)^{\overline{\beta}} \wedge 1)] + 2\right\}^{2d} \exp\left(-CC_1^2 \frac{k^{2\gamma+2\underline{\beta}}}{(k+1)^{\overline{\beta}}}\right).$$

This is summable over k if $2\gamma + 2\underline{\beta} - \overline{\beta} > 0$.

Now, note that, since $\prod_i t_i \le (k+1)^d$ on S_k, we have, for any $-\delta < \beta_1 < \beta_2 < \ldots < \beta_{\ell-1} < \beta_\ell = d$, $S_k = S_k^L \cup S_k^{-\delta,\beta_1} \cup S_k^{\beta_1,\beta_2} \cup \ldots \cup S_k^{\beta_{\ell-1},\beta_\ell}$. If we choose $0 < \delta < \gamma$, $\beta_1 = 0$, $\beta_2 = \gamma$, and $\beta_{i+1} = (2\beta_i) \wedge d$ for $i \ge 2$, the arguments above will show that the probability of the infimum being less than or equal to r over S_k^L, $S_k^{-\delta,\beta_1}$ and each $S_k^{\beta_i,\beta_{i+1}}$ is summable over k, so that $P\left(\inf_{S_k} \mathbb{G}(s,t) + g(s,t) \le r\right)$ is summable over k, so setting M be such that the tail of this sum past M is less than ε gives the desired result. □

proof of Theorem C.3 for the general case. Let $\mathbb{G}_n(s,t) = \mathbb{G}_{n,x_m,j}(s,t)$ and $g_n(s,t) = g_{n,x_m,j}(s,t)$. Let η be small enough that (3) holds.

As in the proof of the previous theorem, let $S_k = \{k \le \|(s,t)\| \le k+1\}$ and let $S_k^L = S_k \cap \{\prod_i t_i \le (k+1)^{-\delta}\}$ for some fixed δ. We have, using Lemma C.3,

$$P\left(\inf_{S_k^L} \mathbb{G}_n(s,t) + g_n(s,t) \le r\right) \le P\left(\sup_{S_k^L} |\mathbb{G}_n(s,t)| \ge |r|\right)$$
$$\le \left\{6(k+1)[(k+1)^d/k^{-\delta}] + 2\right\}^{2d} \exp\left(-C\frac{|r|}{(k+1)^{-\delta/2}}\right)$$

for $(k+1)^{-\delta} \ge n^{-2\gamma/(d+2\gamma)}(1+\log n)^2 \iff k+1 \le n^{2\gamma/[\delta(d+2\gamma)]}(1+\log n)^{-2/\delta}$ so, if $\delta < 2\gamma$, this will hold eventually for all $(k+1) \le h_n^{-1}\eta$ (once $h_n^{-1}\eta \le n^{2\gamma/[\delta(d+2\gamma)]}(1+\log n)^{-2/\delta} \iff \eta \le n^{2\gamma/[\delta(d+2\gamma)]} n^{-1/(d+2\gamma)}(1+\log n)^{-2/\delta} = n^{(2\gamma/\delta-1)/(d+2\gamma)}(1+\log n)^{-2/\delta}$). The bound is summable over k for any $\delta > 0$.

Again following the proof of the previous theorem, for $\underline{\beta} < \overline{\beta}$, define $S_k^{\underline{\beta},\overline{\beta}} = S_k \cap \{(k+1)^{\underline{\beta}} <$

$\prod_i t_i \leq (k+1)^{\overline{\beta}}\}$. We have, again using Lemma C.3,

$$P\left(\inf_{S_k^{\underline{\beta},\overline{\beta}}} \mathbb{G}_n(s,t) + g_n(s,t) \leq r\right) \leq P\left(\sup_{S_k^{\underline{\beta},\overline{\beta}}} |\mathbb{G}_n(s,t)| \geq C_1 k^{\gamma+\underline{\beta}}\right)$$
$$\leq \left\{6(k+1)[(k+1)^d/(k^{\underline{\beta}} \wedge 1)] + 2\right\}^{2d} \exp\left(-C\frac{C_1 k^{\gamma+\underline{\beta}}}{(k+1)^{\overline{\beta}/2}}\right)$$

for $(k+1)^{\overline{\beta}} \geq n^{-2\gamma/(d+2\gamma)}(1+\log n)^2$ (which will hold once the same inequality holds for δ for $-\delta < \overline{\beta}$) and $k+1 \leq h_n^{-1}\eta$. The bound is summable over k for any $\underline{\beta}, \overline{\beta}$ with $2\gamma + 2\underline{\beta} - \overline{\beta} > 0$.

Thus, noting as in the previous theorem that, for any $-\delta < \beta_1 < \beta_2 < \ldots < \beta_{\ell-1} < \beta_\ell = d$, $S_k = S_k^L \cup S_k^{-\delta,\beta_1} \cup S_k^{\beta_1,\beta_2} \cup \ldots \cup S_k^{\beta_{\ell-1},\beta_\ell}$, if we choose $0 < \delta < \gamma$, $\beta_1 = 0$, $\beta_2 = \gamma$, and $\beta_{i+1} = (2\beta_i) \wedge d$ for $i \geq 2$, the arguments above will show that the probability of the infimum being less than or equal to r over the sets indexed by k for any $k \leq h_n^{-1}\eta$ is bounded uniformly in n by a sequence that is summable over k (once $\eta \leq n^{(2\gamma/\delta-1)/(d+2\gamma)}(1+\log n)^{-2/\delta}$). Thus, if we choose M such that the tail of this sum past M is less than ε and let N be large enough so that $\eta \leq N^{(2\gamma/\delta-1)/(d+2\gamma)}(1+\log N)^{-2/\delta}$, we will have the desired result.

□

Lemmas C.4 and C.5 hold as stated with the same proofs, so the rest of the proof is the same as in the $\gamma = 2$ case. The $n/(a_n \log n)$ rate for $Z_{n,2}$ is still faster than the $n^{(d+\gamma)/(d+2\gamma)}$ rate for a_n increasing slowly enough.

The proof of Theorem 4.1 for the limiting process is the same as before. The only place the drift term is used is in ensuring that the inequality $g_{P,x_m,j}(s_{i,k}, t_k) \leq K/k$ holds in the last display in the proof of the theorem, which is still the case.

C.4 Testing Rate of Convergence Conditions

First, I collect results on the rate estimate $\hat{\beta}$ defined in (1). The next lemma bounds $\hat{\beta}$ when the statistic may not converge at a polynomial rate. Throughout the following, S_n is a statistic on $\mathbb{R}$ with cdf $J_n(x)$ and quantile function $J_n^{-1}(t)$. $L_{n,b}(x|\tau)$ and $\tilde{L}_{n,b}(x|\tau)$ are defined as in the body of the paper, with $S(T_n(\theta))$ replaced by S_n.

Lemma C.8. *Let S_n be a statistic such that, for some sequence τ_n and $x > 0$, $\tau_n J_n^{-1}(t) \geq x$ for large enough n. Then, if $\tau_b S_n \xrightarrow{p} 0$ and $b/n \to 0$, we will have, for any $\varepsilon > 0$, $L_{n,b}^{-1}(t+\varepsilon|\tau) \geq x - \varepsilon$ with probability approaching one.*

Proof. It suffices to show $L_{n,b}(x-\varepsilon|\tau) \leq t+\varepsilon$ with probability approaching one. On the event $E_n \equiv \{|\tau_b S_{\mathcal{S}}| \leq \varepsilon\}$, which has probability approaching one, $L_{n,b}(x-\varepsilon|\tau) \leq \tilde{L}_{n,b}(x|\tau)$. We also have $E[L_{n,b}(x|\tau)] = P(\tau_b S_{\mathcal{S}} \leq x) = J_b(x/\tau_b) \leq t$ by assumption. Thus,

$$
\begin{aligned}
&P(L_{n,b}(x-\varepsilon|\tau) \leq t+\varepsilon) \geq P\left(\left\{\tilde{L}_{n,b}(x|\tau) \leq t+\varepsilon\right\} \cap E_n\right) \\
&\geq P\left(\left\{\tilde{L}_{n,b}(x|\tau) \leq E[L_{n,b}(x|\tau)]+\varepsilon\right\} \cap E_n\right).
\end{aligned}
$$

This goes to one by standard arguments. □

Lemma C.9. *Let $\hat{\beta}_a$ be the estimator defined in Section 2.4, or any other estimator such that $\hat{\beta}_a = \frac{-\log L^{-1}_{n,b_1}(t|1)+\mathcal{O}_p(1)}{\log b_1 - \mathcal{O}_p(1)}$. Suppose that, for some $x_\ell > 0$ and β_u, $x_u n^{\beta_u} \leq J_n^{-1}(t-\varepsilon)$ eventually and $b_1^{\beta_u} S_n \xrightarrow{p} 0$. Then, for any $\varepsilon > 0$, we will have $\hat{\beta}_a \leq \hat{\beta}_u + \varepsilon$ with probability approaching one.*

Proof. We have

$$
\hat{\beta}_a = -\frac{\log L^{-1}_{n,b_1}(t|1)}{\log b_1} + o_p(1) = \frac{\beta_u \log b_1 - \log L^{-1}_{n,b_1}(t|b^{\beta_u})}{\log b_1} + o_p(1) \leq \beta_u - \frac{\log(x_u/2)}{\log b_1} + o_p(1) \xrightarrow{p} \beta_u
$$

where the inequality holds with probability approaching one by Lemma C.8. □

The following lemma shows that the asymptotic distribution of the KS statistic is strictly increasing on its support, which is needed for the estimates of the rate of convergence in Politis, Romano, and Wolf (1999) to converge at a fast enough rate that they can be used in the subsampling procedure.

Lemma C.10. *Under Assumptions 3.1, 3.2, 3.3, 4.1 and 4.2 with part (ii) of Assumption 3.1 replaced by Assumption 5.1, if S is convex, then the the asymptotic distribution $S(Z)$ in Theorem 5.2 satisfies $P(S(Z) \in (a,\infty)) = 1$ for some a, and the cdf of $S(Z)$ is strictly increasing on (a,∞).*

Proof. First, note that, for any concave functions $f_1,\ldots,f_{d_Y}$, $f_i : V_i \to \mathbb{R}$, for some vector space V_i, $x \mapsto S(f_1(x_1),\ldots,f_{d_Y}(x_{d_Y}))$ is convex, since, for any $\lambda \in (0,1)$,

$$
\begin{aligned}
&S(f_1(\lambda x_{a,1}+(1-\lambda)x_{b,1}),\ldots,f_k(\lambda x_{a,d_Y}+(1-\lambda)x_{b,d_Y})) \\
&\geq S(\lambda f_1(x_{a,1})+(1-\lambda)f_k(x_{b,1}),,\ldots,\lambda f_k(x_{a,d_Y})+(1-\lambda)f_k(x_{b,d_Y})) \\
&\geq \lambda S(f_1(x_{a,1}),\ldots,f_k(x_{a,d_Y}))+(1-\lambda)S(f_1(x_{b,1}),\ldots,f_k(x_{b,d_Y}))
\end{aligned}
$$

where the first inequality follows since S is decreasing in each argument and by concavity of the f_ks, and the second follows by convexity of S.

$S(Z)$ can be written as, for some random processes $\mathbb{H}_1(t), \ldots, \mathbb{H}_{d_Y}(t)$ with continuous sample paths and $\mathbb{T} \equiv \mathbb{R}^{|\mathcal{X}_0| \cdot 2d_X}$, $S(\inf_{t\in\mathbb{T}} \mathbb{H}_1(t), \ldots, \inf_{t\in\mathbb{T}} \mathbb{H}_{d_Y}(t))$. Since the infimum of a real valued function is a concave functional, this is a convex function of the sample paths of $(\mathbb{H}_1(t), \ldots, \mathbb{H}_{d_Y}(t))$. The result follows from Theorem 11.1 in Davydov, Lifshits, and Smorodina (1998) as long as the vector of random processes can be given a topology for which this function is lower semi-continuous. In fact, this step can be done away with by noting that, for $\mathbb{T}_0$ a countable dense subset of $\mathbb{T}$ and $\mathbb{T}_\ell$ the first ℓ elements of this subset, $S(\inf_{t\in\mathbb{T}_\ell} \mathbb{H}_1(t), \ldots, \inf_{t\in\mathbb{T}_\ell} \mathbb{H}_{d_Y}(t)) \xrightarrow{d} S(\inf_{t\in\mathbb{R}^{2d}} \mathbb{H}_1(t), \ldots, \inf_{t\in\mathbb{R}^{2d}} \mathbb{H}_{d_Y}(t))$ as $\ell \to \infty$, so, letting F_ℓ be the cdf of $S(\inf_{t\in\mathbb{T}_\ell} \mathbb{H}_1(t), \ldots, \inf_{t\in\mathbb{T}_\ell} \mathbb{H}_{d_Y}(t))$, applying Proposition 11.3 of Davydov, Lifshits, and Smorodina (1998) for each F_ℓ shows that $\Phi^{-1}(F_\ell(t))$ is concave for each ℓ, so, by convergence in distribution, this holds for $S(Z)$ as well. □

The same result in Davydov, Lifshits, and Smorodina (1998) could also be used in the proof of Theorem 4.1 to show that the distribution of $S(Z)$ is continuous except possibly at the infimum of its support, but an additional argument would be needed to show that, if such an atom exists, it would have to be at zero. In the proof of Theorem 4.1, this is handled by using the results of Pitt and Tran (1979) instead.

We are now ready to prove Theorem 6.1.

proof of Theorem 6.1. First, suppose that Assumption 3.1 holds with part (ii) of Assumption 3.1 replaced by Assumption 5.1 for some $\underline{\gamma} < \gamma < \overline{\gamma}$ and $\mathcal{X}_0$ nonempty. By Theorem 5.2, $n^{(d_X+\gamma)/(d_X+2\gamma)} S(T_n(\theta))$ converges in distribution to a continous distribution. Thus, by Lemma C.9, $\hat{\beta}_a \xrightarrow{p} (d_X + \gamma)/(d_X + 2\gamma)$, so $\hat{\beta}_a > \underline{\beta} = (d_X + \overline{\gamma})/(d_X + 2\overline{\gamma})$ with probability approaching one. On this event, the test uses the subsample estimate of the $1 - \alpha$ quantile with rate estimate $\hat{\beta} \wedge \overline{\beta}$. By Theorem 8.2.1 in Politis, Romano, and Wolf (1999), $\hat{\beta} \wedge \overline{\beta} = (d_X + \gamma)/(d_X + 2\gamma) + o_p((\log n)^{-1})$ as long as the asymptotic distribution of $n^{(d_X+\gamma)/(d_X+2\gamma)} S(T_n(\theta))$ is increasing on the smallest interval (k_0, k_1) on which the asymptotic distribution has probability one. This holds by Lemma C.10. By Theorem 8.3.1 in Politis, Romano, and Wolf (1999), the $o_p((\log n)^{-1})$ rate of convergence for the rate estimate $\hat{\beta} \wedge \overline{\beta}$ implies that the probability of rejecting converges to α.

Next, suppose that Assumption 3.1 holds with part (ii) of Assumption 3.1 replaced by Assumption 5.1 for $\gamma = \overline{\gamma}$. The test that compares $n^{1/2} S(T_n(\theta))$ to a positive critical value will fail to reject with probability approaching one in this case, so, on an event with probability approaching one, the test will reject only if $\hat{\beta}_a \geq \underline{\beta}$ and the subsampling test with rate $\hat{\beta} \wedge \overline{\beta}$ rejects. Thus, the probability of rejecting is asymptotically no greater than the probability of rejecting with the subsampling test with rate $\hat{\beta} \wedge \overline{\beta}$, which has asymptotic level α under these conditions by the argument above.

Now, consider the case where, for some $x_0 \in \mathcal{X}_0$ and $B < \infty$, $\bar{m}_j(\theta, x) \leq B\|x - x_0\|^\gamma$ for some $\gamma > \bar{\gamma}$. Let $\tilde{m}_j(W_i, \theta) = m_j(W_i, \theta) + (B\|x - x_0\|^\gamma - \bar{m}_j(\theta, x))$. Then $\tilde{m}_j(W_i, \theta) \geq m_j(W_i, \theta)$, and $\tilde{m}_j(W_i, \theta)$ satisfies the assumptions of Theorems 5.2 and 4.1, so

$$n^{(d_X+\gamma)/(d_X+2\gamma)} S(T_n(\theta)) \geq n^{(d_X+\gamma)/(d_X+2\gamma)} S(0, \ldots, 0, \inf_{s,t} E_n \tilde{m}_j(W_i, \theta) I(s < X_i < s + t), 0, \ldots, 0)$$

and the latter quantity converges in distribution to a continuous random variable that is positive with probability one. Thus, by Lemma C.9, for any $\varepsilon > 0$, $\hat{\beta}_a < (d_X+\gamma)/(d_X+2\gamma)+\varepsilon$ with probability approaching one. For ε small enough, this means that $\hat{\beta}_a < (d_X + \overline{\gamma})/(d_X + 2\overline{\gamma})$ with probability approaching one. Thus, the procedure uses an asymptotically level α test with probability approaching one.

The remaining case is where $\bar{m}_j(\theta, x)$ is bounded from below away from zero. If $m_j(W_i, \theta) \geq 0$ for all j with probability one, $S(T_n(\theta))$ and the estimated $1 - \alpha$ quantile will both be zero, so the probability of rejecting will be zero, so suppose that $P(m_j(W_i, \theta) < 0) > 0$ for some j. Then, for some $\eta > 0$, we have $nS(T_n(\theta)) > \eta$ with probability approaching one. From Lemma C.8 (applied with t less that $1 - \alpha$ and $\tau_b = b$), it follows that $L_{n,b}^{-1}(1 - \alpha | b^{\hat{\beta} \wedge \overline{\beta}}) = b^{\hat{\beta} \wedge \overline{\beta} - 1} L_{n,b}^{-1}(1 - \alpha | b) \geq b^{\hat{\beta} \wedge \overline{\beta} - 1} \eta / 2$ with probability approaching one. By Lemma C.5, $S(T_n(\theta))$ will converge at a $n \log n$ rate, so that $n^{\hat{\beta} \wedge \overline{\beta}} S(T_n(\theta)) < n^{\hat{\beta} \wedge \overline{\beta} - 1} (\log n)^2$ with probability approaching one. Thus, we will fail to reject with probability approaching one as long as $n^{\hat{\beta} \wedge \overline{\beta} - 1} (\log n)^2 \leq b^{\hat{\beta} \wedge \overline{\beta} - 1} \eta / 2 = n^{\chi_3(\hat{\beta} \wedge \overline{\beta} - 1)} \eta / 2$ for large enough n, and this holds since $\chi_3 < 1$. A similar argument holds for $\tilde{L}_{n,b}^{-1}(1 - \alpha | b^{\hat{\beta} \wedge \overline{\beta}})$.

□

C.5 Local Alternatives

proof of Theorem 7.1. Everything is the same as in the proof of Theorem 3.1, but with the following modifications.

First, in the proof of Theorem C.1, we need to show that, for all j,

$$\frac{\sqrt{n}}{\sqrt{h_n^d}} (E_n - E)[m_j(W_i, \theta_0 + a_n) - m_j(W_i, \theta_0)] I(h_n s < X_i - x_k < h_n(s + t))$$

converges to zero uniformly over $\|(s, t)\| < M$ for any fixed M. By Theorem 2.14.1 in van der Vaart and Wellner (1996), the L^2 norm of this is bounded up to a constant by $J(1, \mathcal{F}_n, L_2)\frac{1}{h_n^d}\sqrt{EF_n(X_i, W_i)^2}$, where $\mathcal{F}_n = \{(x, w) \mapsto [m_j(w, \theta_0 + a_n) - m_j(w, \theta_0)]I(h_n s < x - x_k < h_n(s + t))|(s, t) \in \mathbb{R}^{2d}\}$ and $F_n(x, w) = |m_j(w, \theta_0 + a_n) - m_j(w, \theta_0)|I(-h_n M\iota < x - x_k < 2h_n M\iota)$ is an envelope function for this class (here ι is a vector of ones). The covering numbers of the $\mathcal{F}_n$s are uniformly bounded by a polynomial, so that we just need to show that $\frac{1}{h_n^d}\sqrt{EF_n(X_i, W_i)^2}$ converges to zero. We have

$$\begin{aligned}
&\frac{1}{\sqrt{h_n^d}}\sqrt{EF_n(X_i, W_i)^2} \\
&= \frac{1}{\sqrt{h_n^d}}\sqrt{EE\{[m_j(W_i, \theta_0 + a_n) - m_j(W_i, \theta_0)]^2|X_i\}I(-h_n M\iota < X_i - x_k < 2h_n M\iota)} \\
&\leq \frac{1}{\sqrt{h_n^d}}\sqrt{EI(-h_n M\iota < X_i - x_k < 2h_n M\iota)} \sup_{\|x - x_k\| \leq \eta} E\{[m_j(W_i, \theta_0 + a_n) - m_j(W_i, \theta_0)]^2|X_i = x\}
\end{aligned}$$

where the first equality uses the law of iterated expectations and the second holds eventually with η chosen so that the convergence in Assumption 7.2 is uniform over $\|x - x_k\| < \eta$. The first term is bounded eventually by $\overline{f}\int_{-M\iota < x < 2M\iota} dx$ where $\overline{f}$ is a bound for the density of X_i in a neighborhood of x_k (this follows from the same change of variables as in other parts of the proof). The second term converges to zero by Assumption 7.2.

Next, in the proof of Theorem C.1, we need to show that

$$\frac{1}{h_n^{d+2}}E[\bar{m}_j(\theta_0 + a_n, X_i) - \bar{m}_j(\theta_0, X_i)]I(h_n s < X_i - x_k < h_n(s + t)) \to f_X(x_k)\bar{m}_{\theta,j}(\theta_0, x_k)a\prod_i t_i$$

uniformly in $\|(s, t)\| \leq M$. We have

$$\begin{aligned}
&\frac{1}{h_n^{d+2}}E[\bar{m}_j(\theta_0 + a_n, X_i) - \bar{m}_j(\theta_0, X_i)]I(h_n s < X_i - x_k < h_n(s + t)) - f_X(x_k)\bar{m}_{\theta,j}(\theta_0, x_k)a\prod_i t_i \\
&= \frac{1}{h_n^{d+2}}\int_{h_n s < x - x_k < h_n(s+t)} \left\{[\bar{m}_j(\theta_0 + a_n, x) - \bar{m}_j(\theta_0, x)]f_X(x) - h_n^2 f_X(x_k)\bar{m}_{\theta,j}(\theta_0, x_k)a\right\} dx \\
&= \int_{s < x < s+t} \left\{h_n^{-2}[\bar{m}_j(\theta_0 + a_n, h_n x + x_k) - \bar{m}_j(\theta_0, h_n x + x_k)]f_X(h_n x + x_k) - f_X(x_k)\bar{m}_{\theta,j}(\theta_0, x_k)a\right\} dx
\end{aligned}$$

where the second equality comes from the change of variable $x \mapsto h_n x + x_k$. This will go to zero uniformly in $\|(s, t)\| \leq M$ as long as $\sup_{\|x\| \leq 2M}\|f_X(h_n x + x_k) - f_X(x_k)\|$ and

$$\sup_{\|x\| \leq 2M}\|h_n^{-2}[\bar{m}_j(\theta_0 + a_n, h_n x + x_k) - \bar{m}_j(\theta_0, h_n x + x_k)] - \bar{m}_{\theta,j}(\theta_0, x_k)a\|$$

both go to zero. $\sup_{\|x\|\leq 2M}\|f_X(h_nx+x_k)-f_X(x_k)\|$ goes to zero by continuity of f_X at x_k. As for the other expression, since $ah_n^2=a_n$, the mean value theorem shows that this is equal to $\bar{m}_{\theta,j}(\theta^*(a_n),h_nx+x_k)a-\bar{m}_{\theta,j}(\theta_0,x_k)a$ for some $\theta^*(a_n)$ between θ_0 and θ_0+a_n. This goes to zero by Assumption 7.1.

In verifying the conditions of Lemma C.1, we need to make sure the bounds, $g_{P,x_k,j,a}(s,t)\geq C\|(s,t)\|^2\prod_i t_i$ and

$$g_{n,x_k,j,a}(s,t)\equiv\frac{1}{h_n^{d+2}}Em_j(W_i,\theta_0+a_n)I(h_ns<X_i<h_n(s+t))\geq C\|(s,t)\|^2\prod_i t_i$$

still hold for $\|(s,t)\|\geq M$ for M large enough and, for the latter function, $\|(s,t)\|\leq h_n^{-1}\eta$ for some $\eta>0$ and n greater than some N that does not depend on M. We have

$$g_{P,x_k,j,a}(s,t)=g_{P,x_k,j}(s,t)+\bar{m}_{\theta,j}(\theta_0,x_k)af_X(x_k)\prod_i t_i\geq C\|(s,t)\|^2\prod_i t_i+\bar{m}_{\theta,j}(\theta_0,x_k)af_X(x_k)\prod_i t_i$$
$$=\|(s,t)\|^2[C+\bar{m}_{\theta,j}(\theta_0,x_k)af_X(x_k)/\|(s,t)\|^2]\prod_i t_i$$

where the first inequality follows from the bound in the original proof. For $\|(s,t)\|\geq M$ for M large enough, this is greater than or equal to $K\|(s,t)\|^2\prod_i t_i$ for $K=C-|\bar{m}_{\theta,j}(\theta_0,x_k)a|f_X(x_k)/M^2>0$. For $g_{n,x_k,j,a}(s,t)$, we have

$$\|g_{P,x_k,j,a}(s,t)-g_{P,x_k,j}(s,t)\|=\|\frac{1}{h_n^{d+2}}E[m_j(W_i,\theta_0+a_n)-m_j(W_i,\theta_0)]I(h_ns<X_i<h_n(s+t))\|$$
$$\leq\sup_{\|x-x_k\|\leq\eta}\|\frac{1}{h_n^2}[\bar{m}_j(\theta_0+a_n,x)-\bar{m}_j(\theta_0,x)]\|\|\frac{1}{h_n^d}EI(h_ns<X_i<h_n(s+t))\|.$$

By the mean value theorem, $\bar{m}_j(\theta_0+a_n,x)-\bar{m}_j(\theta_0,x)=\bar{m}_{j,\theta}(\theta^*(a_n),x)a_n$ for some $\theta^*(a_n)$ between θ_0 and θ_0+a_n. By continuity of the derivative as a function of (θ,x), for small enough η and n large enough, $\bar{m}_{j,\theta}(\theta^*(a_n),x)$ is bounded from above, so that $\|\frac{1}{h_n^2}[\bar{m}_j(\theta_0+a_n,x)-\bar{m}_j(\theta_0,x)]\|$ is bounded by a constant times $\|a_n\|/h_n^2=\|a\|$. By continuity of f_X at x_k, $\|\frac{1}{h_n^d}EI(h_ns<X_i<h_n(s+t))\|$ is bounded by some constant times $\prod_i t_i$ for $\|(s,t)\|\leq h_n^{-1}\eta$. Thus, for $M\leq\|(s,t)\|\leq h_n^{-1}\eta$ for the appropriate M and η, we have, for some constant C_1,

$$g_{P,x_k,j,a}(s,t)\geq g_{P,x_k,j}(s,t)-C_1\prod_i t_i\geq C\|(s,t)\|^2\prod_i t_i-C_1\prod_i t_i$$
$$=\|(s,t)\|^2[C-C_1/\|(s,t)\|^2]\prod_i t_i$$

where the second inequality uses the bound from the original proof. For M large enough, this gives the desired bound with the constant equal to $C - C_1/M > 0$.

In verifying the conditions of Lemma C.1, we also need to make sure the argument in Lemma C.3 still goes through when $m(W_i, \theta_0)$ is replaced by $m(W_i, \theta_0 + a_n)$. To get the lemma to hold (with the constant C depending only on the distribution of X and the $\overline{Y}$ in Assumption 7.3), we can use the same proof, but with the classes of functions $\mathcal{F}_n$ defined to be $\mathcal{F}_n = \{(x, w) \mapsto m_j(w, \theta_0 + a_n)I(h_n s_0 < x - x_k < h_n(s_0 + t))|t \leq t_0\}$ ($J(1, \mathcal{F}_n, L^2)$ is bounded uniformly for these classes because the covering number of each $\mathcal{F}_n$ is bounded by the same polynomial), and using the envelope function $F_n(x, w) = \overline{Y}I(h_n s_0 < x - x_k < h_n(s_0 + t_0))$ when applying Theorem 2.14.1 in van der Vaart and Wellner (1996).

□

proof of Theorem 7.2. First, note that, for any neighborhoods $B(x_k)$ of the elements of $\mathcal{X}_0$, $\sqrt{n} \inf_{s,t} E_n m_j(W_i, \theta_0 + a_n)I(s < X < s + t) = \sqrt{n} \inf_{(s,s+t) \in \cup_{k \text{ s.t. } j \in J(k)} B(x_k)} E_n m_j(W_i, \theta_0 + a_n)I(s < X_i < s + t) + o_p(1)$ since, if these neighborhoods are made small enough, we will have, for any $(s, s+t)$ not in one of these neighborhoods, $E m_j(W_i, \theta_0 + a_n)I(s < X_i < s+t) \geq \underline{B}P(s < X_i < s + t)$ by an argument similar to the one in Lemma C.4, so that an argument similar to the one in Lemma C.5 will show that $\inf_{(s,s+t) \in \cup_{k \text{ s.t. } j \notin J(k)} B(x_k)} E_n m_j(W_i, \theta_0 + a_n)I(s < X_i < s + t)$ converges to zero at a faster than $\sqrt{n}$ rate (Assumption 7.1 guarantees that $E[m_j(W_i, \theta_0 + a_n)|X]$ is eventually bounded away from zero outside of any neighborhood of $\mathcal{X}_0$ so that a similar argument applies).

Thus, the result will follow once we show that, for each j and k such that $j \in J(k)$,

$$\sqrt{n} \inf_{(s,s+t) \in B(x_k)} E_n m_j(W_i, \theta_0 + a_n)I(s < X_i < s + t)$$
$$\xrightarrow{p} \inf_{s,t} f_X(x_k) \int_{s<x<s+t} \left(\frac{1}{2} x' V x + \overline{m}_{\theta,j}(\theta_0, x_k) a \right) dx.$$

With this in mind, fix j and k with $j \in J(k)$.

Let (s_n^*, t_n^*) minimize $E_n m_j(W_i, \theta_0 + a_n)I(s < X < s+t)$ over $B(x_k)^2$ (and be chosen from the set of minimizers in a measurable way). First, I show that $\rho(0, (s_n^*, t_n^*)) \xrightarrow{p} 0$ where ρ is the covariance semimetric $\rho((s, t), (s', t')) = var(m_j(W_i, \theta_0)I(s < x < s + t) - m_j(W_i, \theta_0)I(s' < x < s' + t'))$. To show this, note that, for any $\varepsilon > 0$, $E m_j(W_i, \theta_0 + a_n)I(s < X_i < s + t)$ is bounded from below away from zero for $\rho(0, (s, t)) \geq \varepsilon$ for large enough n. To see this, note that, for $\rho(0, (s, t)) \geq \varepsilon$, $\prod_i t_i \geq K$ for some constant K, so that $\|(s, t)\| \geq K^{1/d}$ and, for some constant C and a bound $\overline{f}$ for f_X on $B(x_k)$,

$$\begin{aligned}
&Em_j(W_i,\theta_0+a_n)I(s<X_i<s+t) \\
&= Em_j(W_i,\theta_0)I(s<X_i<s+t) + E[\bar{m}_j(\theta_0+a_n,X_i)-\bar{m}_j(\theta_0,X_i)]I(s<X_i<s+t) \\
&\geq C_1\|(s,t)\|^2\left(\prod_i t_i\right) - \sup_{x\in B(x_k)}\|\bar{m}_j(\theta_0+a_n,x)-\bar{m}_j(\theta_0,x)\|\overline{f}\left(\prod_i t_i\right) \\
&\geq \left[C_1\|(s,t)\|^2 - \sup_{x\in B(x_k)}\|\bar{m}_j(\theta_0+a_n,x)-\bar{m}_j(\theta_0,x)\|\overline{f}\right]K.
\end{aligned}$$

By Assumption 7.2, $\sup_{x\in B(x_k)}\|\bar{m}_j(\theta_0+a_n,x)-\bar{m}_j(\theta_0,x)\|$ converges to zero, so the last term in this display will be positive and bounded away from zero for large enough n. Thus, we can write $\sqrt{n}E_n m_j(W_i,\theta_0+a_n)I(s<X_i<s+t)$ as the sum of $\sqrt{n}(E_n-E)m_j(W_i,\theta_0+a_n)I(s<X_i<s+t)$, which is $\mathcal{O}_p(1)$ uniformly in (s,t), and $\sqrt{n}Em_j(W_i,\theta_0+a_n)I(s<X<s+t)$, which is bounded from below uniformly in $\rho(0,(s,t))\geq\varepsilon$ by a sequence of constants that go to infinity. Thus, $\inf_{\rho(0,(s,t))\geq\varepsilon}\sqrt{n}E_n m_j(W_i,\theta_0+a_n)I(s<X<s+t)$ is greater than zero with probability approaching one, so $\rho(0,(s^*,t^*))\xrightarrow{p}0$.

Thus, for some sequence of random variables $\varepsilon_n\xrightarrow{p}0$,

$$\begin{aligned}
&\sqrt{n}\inf_{s,t}E_n m_j(W_i,\theta_0+a_n)I(s<X<s+t) \\
&=\sqrt{n}\inf_{\rho(0,(s^*,t^*))\leq\varepsilon_n,(s,s+t)\in B(x_k)}E_n m_j(W_i,\theta_0+a_n)I(s<X<s+t).
\end{aligned}$$

This is equal to $\sqrt{n}\inf_{\rho(0,(s^*,t^*))\leq\varepsilon_n,(s,s+t)\in B(x_k)}Em_j(W_i,\theta_0+a_n)I(s<X<s+t)$ plus a term that is bounded by $\sqrt{n}\sup_{\rho(0,(s^*,t^*))\leq\varepsilon_n,(s,s+t)\in B(x_k)}|(E_n-E)E_n m_j(W_i,\theta_0+a_n)I(s<X<s+t)|$. By Assumption 7.2 and an argument using the maximal inequality in Theorem 2.14.1 in van der Vaart and Wellner (1996), $\sqrt{n}\sup_{(s,s+t)\in B(x_k)}|(E_n-E)[m_j(W_i,\theta_0+a_n)-m_j(W_i,\theta_0)]I(s<X_i<s+t)|$ converges in probability to zero. $\sqrt{n}(E_n-E)m_j(W_i,\theta_0)I(s<X_i<s+t)$ converges in distribution under the supremum norm to a mean zero Gaussian process $\mathbb{H}(s,t)$ with covariance kernel $cov(\mathbb{H}(s,t),\mathbb{H}(s',t'))=cov(m_j(W_i,\theta_0)I(s<X_i<s+t),m_j(W_i,\theta_0)I(s'<X_i<s'+t'))$ and almost sure ρ continuous sample paths. Since $(z,\varepsilon)\mapsto\sup_{\rho(0,(s,t))\leq\varepsilon}|z(s,t)|$ is continuous in $C(\mathbb{R}^{2d_X},\rho)\times\mathbb{R}$ (where $C(\mathbb{R}^{2d_X},\rho)$ is the space of ρ continuous functions on $\mathbb{R}^{2d}$) under the product norm of the supremum norm and the Euclidean norm, by the continuous mapping theorem, $\sup_{\rho(0,(s,t))\leq\varepsilon_n}|\sqrt{n}(E_n-E)m_j(W_i,\theta_0)I(s<X_i<s+t)|\xrightarrow{d}\sup_{\rho(0,(s,t))\leq 0}\mathbb{H}(s,t)=0$ (the last step follows since $var(\mathbb{H}(s,t))=0$ whenever $\rho(0,(s,t))=0$).

Thus,

$$\begin{aligned}
&\sqrt{n} \inf_{(s,s+t)\in B(x_k)} E_n m_j(W_i,\theta_0+a_n) I(s<X_i<s+t) \\
&= \sqrt{n} \inf_{\rho(0,(s,t))<\varepsilon_n,(s,s+t)\in B(x_k)} E m_j(W_i,\theta_0+a_n) I(s<X_i<s+t) + o_p(1) \\
&= \sqrt{n} \inf_{\rho(0,(s,t))<\varepsilon_n,(s,s+t)\in B(x_k)} \int_{s<x<s+t} \bar{m}_j(\theta_0+a_n,x) f_X(x)\, dx + o_p(1).
\end{aligned}$$

By Assumption 7.1, the integrand is positive eventually for $\|(s-x_k,t)\| \geq \eta$ for any $\eta>0$, and once this holds, the infimum will be achieved on $\|(s-x_k,t)\|<\eta$. Using a first order Taylor expansion in the first argument of $\bar{m}_j(\theta_0+a_n,x)$ and a second order Taylor expansion in the second argument the integrand is equal to

$$\left[\frac{1}{2}(x-x_k)V(x^*(x))(x-x_k)' + \bar{m}_{\theta,j}(\theta^*(a_n),x)a_n\right] f_X(x)$$

for some $x^*(x)$ between x and x_k and $\theta^*(a_n)$ between θ_0 and θ_0+a_n. For η small enough, continuity of the derivatives at (θ_0,x_k) guarantees that this is bounded from below by $C_1\|x-x_k\|^2 - C_2 a_n$ for some constants C_1 and C_2, so the integrand is positive for x greater than $C\sqrt{\|a_n\|}$ for some large C, so that the infimum will be taken on $\|(s,s+t)\| < C\sqrt{\|a_n\|}$. Thus, we have

$$\begin{aligned}
&\sqrt{n} \inf_{(s,s+t)\in B(x_k)} E_n m_j(W_i,\theta_0+a_n) I(s<X_i<s+t) \\
&= \sqrt{n} \inf_{\rho(0,(s,t))<\varepsilon_n,\|(s-x_k,t)\|<C\sqrt{\|a_n\|}} \int_{s<x<s+t} \bar{m}_j(\theta_0+a_n,x) f_X(x)\, dx + o_p(1).
\end{aligned}$$

This will be equal up to $o(1)$ to the infimum of

$$\sqrt{n} \int_{s<x<s+t} \left[\frac{1}{2}(x-x_k)V_j(x_k)(x-x_k)' + \bar{m}_{\theta,j}(\theta_0,x_k)a_n\right] f_X(x_k)\, dx$$

once we show that the difference between this expression and $\sqrt{n}\int_{s<x<s+t} \bar{m}_j(\theta_0+a_n,x) f_X(x)\, dx$ goes to zero uniformly over $\|(s-x_k,t)\| \leq C\sqrt{\|a_n\|}$ (the infimum of this last display will be taken at a sequence where $\|(s-x_k,t)\| \leq C\sqrt{\|a_n\|}$ anyway, so that the infimum can be taken over all of $\mathbb{R}^{2d}$).

The difference between these terms is

$$\sqrt{n}\int_{s<x<s+t}\left[\frac{1}{2}(x-x_k)V_j(x_k)(x-x_k)'+\bar{m}_{\theta,j}(\theta_0,x_k)a_n\right][f_X(x)-f_X(x_k)]\,dx$$
$$+\sqrt{n}\int_{s<x<s+t}\frac{1}{2}\left[(x-x_k)V_j(x^*(x))(x-x_k)'-(x-x_k)V_j(x_k)(x-x_k)'\right]f_X(x)\,dx$$
$$+\sqrt{n}\int_{s<x<s+t}\left[\bar{m}_{\theta,j}(\theta^*(a_n),x)-\bar{m}_{\theta,j}(\theta_0,x_k)\right]a_nf_X(x)\,dx.$$

These can all be bounded using the change of variables $u=(x-x_k)n^{1/(2(d+2))}$ and the continuity of densities, conditional means, and their derivatives. The first term is

$$\sqrt{n}\int_{n^{1/(2(d+2))}(s-x_k)<u<(s+t-x_k)n^{1/(2(d+2))}}\left[\frac{1}{2}uV_j(x_k)u'n^{-1/(d+2)}+\bar{m}_{\theta,j}(\theta_0,x_k)an^{-1/(d+2)}\right]$$
$$\times[f_X(n^{-1/(2(d+2))}u+x_k)-f_X(x_k)]n^{-d/(2(d+2))}\,du$$
$$=\int_{n^{1/(2(d+2))}(s-x_k)<u<(s+t-x_k)n^{1/(2(d+2))}}\left[\frac{1}{2}uV_j(x_k)u'+\bar{m}_{\theta,j}(\theta_0,x_k)a\right]$$
$$\times[f_X(n^{-1/(2(d+2))}u+x_k)-f_X(x_k)]\,du.$$

The integrand converges to zero uniformly over u in any bounded set by the continuity of f_X at x_k, and the area of integration is bounded by $\|u\|\leq 2n^{1/(2(d+2))}\|(s-x_k,t)\|\leq 2Cn^{1/(2(d+2))}\sqrt{\|a\|}n^{-1/(2(d+2))}=2C\sqrt{\|a\|}$ on $\|(s-x_k,t)\|\leq C\sqrt{\|a_n\|}$. Using the same change of variables, the second term is bounded by the integral of

$$\frac{1}{2}\left[u'V_j(x^*(n^{-1/(2(d+2))}u+x_k))u'-uV_j(x_k)u'\right]f_X(n^{-1/(2(d+2))}u+x_k)$$

over a bounded region, and this converges to zero uniformly in any bounded region by continuity of the second derivative matrix. The last term is, by the same change of variables, bounded by the integral of

$$\left[\bar{m}_{\theta,j}(\theta^*(a_n),n^{-1/(2(d+2))}u+x_k)-\bar{m}_{\theta,j}(\theta_0,x_k)\right]af_X(n^{-1/(2(d+2))}u+x_k)$$

over a bounded region, and this converges to zero by continuity of $m_{\theta,j}(\theta,x)$ at (θ_0,x_k).

Thus,

$$
\begin{aligned}
&\sqrt{n} \inf_{(s,s+t)\in B(x_k)} E_n m_j(W_i, \theta_0 + a_n) I(s < X_i < s+t) \\
&= \inf_{\|(s-k,t)\|\le C\sqrt{\|a_n\|}} \sqrt{n} \int_{s<x<s+t} \left[\frac{1}{2}(x-x_k)V_j(x_k)(x-x_k)' + \bar{m}_{\theta,j}(\theta_0, x_k)a_n\right] f_X(x_k)\, dx + o_p(1) \\
&= \inf_{\|(s-x_k,t)\|\le C\sqrt{\|a\|}} \int_{(s-x_k)<u<(s-x_k+t)} \left[\frac{1}{2}uV_j(x_k)u' + \bar{m}_{\theta,j}(\theta_0, x_k)a_n\right] f_X(x_k)\, du + o_p(1)
\end{aligned}
$$

where the last equality follows from the same change of variables and a change of coordinates in (s,t). The result follows since, for large enough C, the unconstrained infimum is taken on $\|(s-x_k,t)\| \le C\sqrt{\|a\|}$, and C can be chosen arbitrarily large. □

C.6 Alternative Method for Estimating the Asymptotic Distribution

proof of Theorem B.1. It suffices to show that, for every subsequence, there exists a further subsequence along which the distribution of $\hat{Z}$ converges weakly to the distribution of Z. Given a subsequence, let the further subsequence be such that the convergence in probability in Assumption B.1 is with probability one.

For any fixed $B > 0$, the processes

$$\left[\hat{\mathbb{G}}_{P,x_k}(s,t) + \hat{g}_{P,x_k}(s,t)\right] I(\|(s,t)\| \le B_n)$$

are, along this subsequence, Gaussian processes with mean functions and covariance kernels converging with probability one to those of the distribution being estimated uniformly in $\|(s,t)\| \le B$. Thus, with probability one, the distributions of these processes converge weakly to the distribution of the process being estimated along this subsequence taken as random processes on $\|(s,t)\| \le B$. Thus, to get the weak convergence of the elementwise infimum, we just need to verify part (ii) of Lemma C.1. To this end, note that, along the further subsequence, the infimum of

$$\left[\hat{\mathbb{G}}_{P,x_k,j}(s,t) + \hat{g}_{P,x_k,j}(s,t)\right] I(\|(s,t)\| \le B_n)$$

is eventually bounded from below (in the stochastic dominance sense) by the infimum of a process defined the same way as

$$\mathbb{G}_{P,x_k,j}(s,t) + g_{P,x_k,j}(s,t),$$

but with $E(m_{J(k)}(W_i,\theta)m_{J(k)}(W_i,\theta)'|X = x_k)$ replaced by $2E(m_{J(k)}(W_i,\theta)m_{J(k)}(W_i,\theta)'|X = x_k)$, and $V(x_k)$ replaced by $V(x_k)/2$. Once n is large enough that this holds along this further subsequence, part (ii) of Lemma C.1 will hold by Lemma C.2 applied to this process. $\square$

proof of Corollary B.1. By Theorem B.1, the distribution of $S(\hat{Z})$ converges weakly conditionally in probability to the distribution of $S(Z)$, and by Theorem 3.1, $n^{(d_X+2)/(d_X+4)}S(T_n(\theta)) \xrightarrow{d} S(Z)$. $S(Z)$ has a continuous distribution by Theorem 4.1, so the result follows by standard arguments. $\square$

proof of Lemma B.1. Let $h(x) = \bar{m}_j(\theta,x) - \min_{x' \in D} \bar{m}_j(\theta,x)$ where $\bar{m}_j(\theta,x) = E(m_j(W_i,\theta)|X_i = x)$ for a continuous version of the conditional mean function. First, note that $\mathcal{X}_0^j$ is compact. Since each $x \in \mathcal{X}_0^j$ is a local minimizer of $h(x)$ such that the second derivative matrix is strictly positive definite at x, there is an open set $A(x)$ containing each $x \in \mathcal{X}_0^j$ such that $h(x) > 0$ on $A(x)\backslash x$. The sets $A(x)$ with x ranging over $\mathcal{X}_0^j$ form a covering of $\mathcal{X}_0^j$ with open sets. Thus, there is a finite subcover $A(x_1), \ldots A(x_\ell)$ of $\mathcal{X}_0^j$. Since the only elements in $A(x_1) \cup \cdots \cup A(x_\ell)$ that are also in $\mathcal{X}_0^j$ are $x_1, \ldots, x_\ell$, this means that $\mathcal{X}_0^j = \{x_1, \ldots, x_\ell\}$. $\square$

proof of Theorem B.5. By the next lemma, we will have $\mathcal{X}_0^j \subseteq \hat{\mathcal{X}}_0^j \subseteq \cup_{k=1}^{\hat{\ell}_j} B_{\varepsilon_n}(\hat{x}_{j,k})$ and $\mathcal{X}_0^j \subseteq \hat{\mathcal{X}}_0^j \subseteq \cup_{k \text{ s.t. } j \in J(k)} B_{\varepsilon_n}(x_k)$ with probability approaching one. When this holds, we will have $\hat{\ell} \leq |\{k|j \in J(k)\}|$ by construction and, once ε_n is less than the smallest distance between any two points in $\mathcal{X}_0^j$, we will also have $\hat{\ell}_j = |\{k|j \in J(k)\}|$ and, for each k from 1 to $\hat{\ell}_j$, we will have, for some function $r(j,k)$ such that $r(j,\cdot)$, is bijective from $\{1, \ldots, \hat{\ell}_j\}$ to $\{k|j \in J(k)\}$, $x_{r(j,k)} \in B_{\varepsilon_n}(\hat{x}_{j,k})$ for each j,k. When this holds, all of the $\hat{x}_{j,k}$s with $r(j,k)$ equal will be in the same equivalence class, since the corresponding ε_n neighborhoods will intersect. When ε_n is small enough that ε_n neighborhoods containing x_r and ε_n neighborhoods containing x_s do not intersect for $r \neq s$, there will be exactly ℓ equivalence classes, each one corresponding to the (j,k) indices such that $r(j,k)$ is the same. Let the labeling of the $\tilde{x}_s$s be such that, for all s, $\tilde{x}_s = \hat{x}_{j,k}$ for some (j,k) such that $r(j,k) = s$. Then, for each s, we have, for some (j,k) such that $r(j,k) = s$, $x_s = x_{r(j,k)} \in B_{\varepsilon_n}(\hat{x}_{j,k}) = B_{\varepsilon_n}(\tilde{x}_s)$ with probability approaching one so that $\tilde{x}_s \xrightarrow{p} x_s$. To verify that $\hat{J}(s) = J(s)$ with probability approaching one, note that, for $j \in J(s)$, we will have $x_s \in \mathcal{X}_0^j \subseteq \cup_k B_{\varepsilon_n}(\hat{x}_{j,k})$ and $x_s \in B_{\varepsilon_n}(\tilde{x}_s)$ eventually, and, when this holds, $[\cup_k B_{\varepsilon_n}(\hat{x}_{j,k})] \cap B_{\varepsilon_n}(\tilde{x}_s) \neq \emptyset$ so that $j \in \hat{J}(s)$. For $j \notin J(s)$, each $\hat{x}_{j,k}$ will eventually be within ε_n of some x_r with $r \neq s$, while indices (j', k') in the equivalence class associated with s will eventually have $\hat{x}_{j',k'}$ within 2ε of x_s, so that (j, k) will not be in the equivalence class associated with s for any k, and $j \notin \hat{J}(s)$. $\square$

Lemma C.11. *Suppose that* $\sup_{x \in D} \|\hat{\bar{m}}_j(\theta, x) - \bar{m}_j(\theta, x)\| = \mathcal{O}(a_n)$ *for some sequence* $a_n \to 0$*. Then, under Assumption 3.1, for any sequence* $b_n \to \infty$ *with* $b_n a_n \to 0$ *and* ε_n *with* $\varepsilon_n \to 0$ *more slowly than* $\sqrt{b_n a_n}$*, the set* $\hat{\mathcal{X}}_0^j \equiv \{x | \hat{\bar{m}}_j(\theta, x) \leq b_n a_n\}$ *satisfies*

$$\mathcal{X}_0^j \subseteq \hat{\mathcal{X}}_0^j \subseteq \cup_{k \text{ s.t. } j \in J(k)} B_{\varepsilon_n}(x_k)$$

Proof. We will have $\mathcal{X}_0^j \subseteq \hat{\mathcal{X}}_0^j$ as soon as $\sup_{x \in D} \|\hat{\bar{m}}_j(\theta, x) - \bar{m}_j(\theta, x)\| \leq b_n a_n$, which happens with probability approaching one. To show that $\hat{\mathcal{X}}_0^j \subseteq \cup_{k \text{ s.t. } j \in J(k)} B_{\varepsilon_n}(x_k)$ eventually, suppose that, for some $\hat{x} \in \hat{\mathcal{X}}_0^j$, $\hat{x} \notin B_{\varepsilon_n}(x_k)$ for any k. Let C and η be such that $\bar{m}_j(\theta, x) \geq C \min_k \|x - x_k\|^2$ when $\|x - x_k\| \leq \eta$ for some k (such a C and η exist by Assumption 3.1). Then, for any $\hat{x}$ such that $\hat{\bar{m}}_j(\theta, \hat{x}) \leq b_n a_n$, we must have, with probability approaching one,

$$C \min_k \|x - x_k\|^2 \leq \bar{m}_j(\theta, \hat{x}) \leq b_n a_n + \bar{m}_j(\theta, \hat{x}) - \hat{\bar{m}}_j(\theta, \hat{x}) \leq 2 b_n a_n$$

where the first inequality follows since $\hat{\mathcal{X}}_0^j$ is contained in $\{x | \|x - x_k\| \leq \eta \text{ some } k \text{ s.t. } j \in J(k)\}$ eventually. Since $\varepsilon_n \geq \sqrt{2 b_n a_n / C}$ eventually, the first claim follows. $\square$