Online Supplement for "Threshold Regression With A Threshold Boundary": Supporting Lemmas

Lemma 1 *Under Assumptions D1-D5 and $\beta_{10}-\beta_{20}$ is fixed, $\widehat{\theta}\xrightarrow{p}\theta_0$.*

Proof. We prove the Lemma by applying Theorem 2.1 of Newey and McFadden (1994). The objective function is $Q_n(\theta)=Q_n(\gamma,\beta)$. It is convenient to consider the recentered version of $Q_n(\gamma,\beta)$:

$$S_n(\gamma,\beta)=Q_n(\gamma,\beta)-Q_n(\gamma_0,\beta_0).$$

We need only show that (i) $\sup\limits_{\theta\in\Theta}|S_n(\gamma,\beta)-S(\gamma,\beta)|\xrightarrow{p}0$, where $\Theta=\Gamma\times B_1\times B_2$, and

$$\begin{aligned}
S(\gamma,\beta) &= \mathbb{E}\left[\left(y-\mathbf{x}'\beta_1 1(q\le\mathbf{z}'\gamma)-\mathbf{x}'\beta_2 1(q>\mathbf{z}'\gamma)\right)^2-\varepsilon^2\right]\\
&= (\beta_{10}-\beta_1)'\,\mathbb{E}\left[\mathbf{xx}'1(q\le\mathbf{z}'\gamma\wedge\mathbf{z}'\gamma_0)\right](\beta_{10}-\beta_1)\\
&\quad+(\beta_{20}-\beta_2)'\,\mathbb{E}\left[\mathbf{xx}'1(q>\mathbf{z}'\gamma\vee\mathbf{z}'\gamma_0)\right](\beta_{20}-\beta_2)\\
&\quad+(\beta_{10}-\beta_2)'\,\mathbb{E}\left[\mathbf{xx}'1(\mathbf{z}'\gamma\wedge\mathbf{z}'\gamma_0<q\le\mathbf{z}'\gamma_0)\right](\beta_{10}-\beta_2)\\
&\quad+(\beta_{20}-\beta_1)'\,\mathbb{E}\left[\mathbf{xx}'1(\mathbf{z}'\gamma_0<q\le\mathbf{z}'\gamma\vee\mathbf{z}'\gamma_0)\right](\beta_{20}-\beta_1)\,;
\end{aligned}$$

(ii) $S(\gamma,\beta)$ is continuous in θ and is uniquely minimized at θ_0. Given Assumption D5, (i) is straightforward by applying a Glivenko-Cantelli theorem, so we concentrate on (ii).

The continuity of $S(\gamma,\beta)$ is obvious given that $f(q|z)$ is bounded from Assumption D4. To show that $S(\gamma,\beta)$ is uniquely minimized at θ_0, we consider four cases. (i) $\gamma=\gamma_0$, $\beta\neq\beta_0$.

$$\begin{aligned}
S(\gamma,\beta)-S(\gamma_0,\beta_0)\ge\max\big\{&(\beta_{10}-\beta_1)'\,\mathbb{E}\left[\mathbf{xx}'1(q\le\mathbf{z}'\gamma_0)\right](\beta_{10}-\beta_1)\,,\\
&(\beta_{20}-\beta_2)'\,\mathbb{E}\left[\mathbf{xx}'1(q>\mathbf{z}'\gamma_0)\right](\beta_{20}-\beta_2)\big\}>0
\end{aligned}$$

Assumption D2 guarantees that $S(\gamma_0,\beta)$ is uniquely minimized at β_0. (ii) $\gamma\neq\gamma_0$, $\beta=\beta_0$.

$$\begin{aligned}
S(\gamma,\beta_0)-S(\gamma_0,\beta_0)&=(\beta_{10}-\beta_{20})'\,\mathbb{E}\left[\mathbf{xx}'1(\mathbf{z}'\gamma\wedge\mathbf{z}'\gamma_0<q\le\mathbf{z}'\gamma_0)\right](\beta_{10}-\beta_{20})\\
&\quad+(\beta_{20}-\beta_{10})'\,\mathbb{E}\left[\mathbf{xx}'1(\mathbf{z}'\gamma_0<q\le\mathbf{z}'\gamma\vee\mathbf{z}'\gamma_0)\right](\beta_{20}-\beta_{10})>0
\end{aligned}$$

by Assumptions D3 and D4. (iii) $\gamma\neq\gamma_0$, $\beta_1=\beta_{20}$ and/or $\beta_2=\beta_{10}$.

$$\begin{aligned}
S(\gamma,\beta)-S(\gamma_0,\beta_0)\ge\min\big\{&(\beta_{10}-\beta_{20})'\,\mathbb{E}\left[\mathbf{xx}'1(q\le\mathbf{z}'\gamma\wedge\mathbf{z}'\gamma_0)\right](\beta_{10}-\beta_{20})\,,\\
&(\beta_{20}-\beta_{10})'\,\mathbb{E}\left[\mathbf{xx}'1(q>\mathbf{z}'\gamma\vee\mathbf{z}'\gamma_0)\right](\beta_{20}-\beta_{10})\big\}>0
\end{aligned}$$

from Assumptions D1 and D2. (iv) $\gamma\neq\gamma_0$, $\beta\neq\beta_0,\beta_1\neq\beta_{20}$ and $\beta_2\neq\beta_{10}$.

$$\begin{aligned}
S(\gamma,\beta)-S(\gamma_0,\beta_0)\ge\max\big\{&(\beta_{10}-\beta_1)'\,\mathbb{E}\left[\mathbf{xx}'1(q\le\mathbf{z}'\gamma\wedge\mathbf{z}'\gamma_0)\right](\beta_{10}-\beta_1)\,,\\
&(\beta_{20}-\beta_2)'\,\mathbb{E}\left[\mathbf{xx}'1(q>\mathbf{z}'\gamma\vee\mathbf{z}'\gamma_0)\right](\beta_{20}-\beta_2)\big\}>0
\end{aligned}$$

by Assumption D2. Combining the previous cases proves the result. ■

Lemma 2 *Under Assumptions D1-D5 and $\|\delta_n\|\to 0$, $\sqrt{n}\,\|\delta_n\|\to\infty$, $\widehat{\beta}_\ell-\beta_{\ell 0}=o_p(\|\delta_n\|)$, and $\widehat{\gamma}-\gamma_0=o_p(1)$.*

Proof. We use the notations in the last lemma to prove this result. By result (i) of the Lemma 1, and $\|\delta_n\| \to 0$,

$$\sup_{\theta\in\Theta} |S_n(\gamma,\beta) - S(\gamma,\beta)| \xrightarrow{p} 0,$$

where $S(\gamma,\beta)$ is redefined as

$$\begin{aligned} S(\gamma,\beta) &= \mathbb{E}\left[\left(y - \mathbf{x}'\beta_1 1(q \le \mathbf{z}'\gamma) - \mathbf{x}'\beta_2 1(q > \mathbf{z}'\gamma)\right)^2 - \varepsilon^2\right] \\ &= (\beta_{10} - \beta_1)' \mathbb{E}\left[\mathbf{x}\mathbf{x}' 1(q \le \mathbf{z}'\gamma \wedge \mathbf{z}'\gamma_0)\right](\beta_{10} - \beta_1) \\ &\quad + (\beta_{20} - \beta_2)' \mathbb{E}\left[\mathbf{x}\mathbf{x}' 1(q > \mathbf{z}'\gamma \vee \mathbf{z}'\gamma_0)\right](\beta_{20} - \beta_2) \\ &\quad + (\beta_{20} - \beta_2)' \mathbb{E}\left[\mathbf{x}\mathbf{x}' 1(\mathbf{z}'\gamma \wedge \mathbf{z}'\gamma_0 < q \le \mathbf{z}'\gamma_0)\right](\beta_{20} - \beta_2) \\ &\quad + (\beta_{10} - \beta_1)' \mathbb{E}\left[\mathbf{x}\mathbf{x}' 1(\mathbf{z}'\gamma_0 < q \le \mathbf{z}'\gamma \vee \mathbf{z}'\gamma_0)\right](\beta_{10} - \beta_1) \end{aligned}$$

because $\beta_{10} - \beta_2 = \beta_{10} - \beta_{20} + \beta_{20} - \beta_2 = \delta_n + \beta_{20} - \beta_2 \to \beta_{20} - \beta_2$ and similarly $\beta_{20} - \beta_1 \to \beta_{10} - \beta_1$. From Assumption D2, $S(\gamma,\beta)$ is uniquely minimized at β_0 for any $\gamma \in \Gamma$, so by Theorem 2.1 of Newey and McFadden (1994), $\widehat{\beta}$ is consistent for any $\gamma \in \Gamma$. However, $S(\gamma,\beta)$ is not uniquely minimized at θ_0. For example, $S(\gamma,\beta_0) = 0$ for any $\gamma \in \Gamma$. To prove the consistency of $\widehat{\gamma}$, the normalization in $Q_n(\gamma,\beta)$ should be a_n^{-1} rather than n^{-1}, where $a_n = n\delta_n'\delta_n \to \infty$. We denote the uncentered and centered objective functions still as $Q_n(\gamma,\beta)$ and $S_n(\gamma,\beta)$. Specifically,

$$\begin{aligned} & S_n(\gamma,\beta) = Q_n(\gamma,\beta) - Q_n(\gamma_0,\beta_0) \\ &= \tfrac{1}{a_n}(\beta_{10} - \beta_1)' \sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i' 1(q_i \le \mathbf{z}_i'\gamma \wedge \mathbf{z}_i'\gamma_0)(\beta_{10} - \beta_1) \\ &\quad + \tfrac{1}{a_n}(\beta_{20} - \beta_2)' \sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i' 1(q_i > \mathbf{z}_i'\gamma \vee \mathbf{z}_i'\gamma_0)(\beta_{20} - \beta_2) \\ &\quad + \tfrac{1}{a_n}(\beta_{10} - \beta_2)' \sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i' 1(\mathbf{z}_i'\gamma \wedge \mathbf{z}_i'\gamma_0 < q_i \le \mathbf{z}_i'\gamma_0)(\beta_{10} - \beta_2) \\ &\quad + \tfrac{1}{a_n}(\beta_{20} - \beta_1)' \sum_{i=1}^n \mathbf{x}_i\mathbf{x}_i' 1(\mathbf{z}_i'\gamma_0 < q_i \le \mathbf{z}_i'\gamma \vee \mathbf{z}_i'\gamma_0)(\beta_{20} - \beta_1) \\ &:= T_1(\theta) + T_2(\theta) + T_3(\theta) + T_4(\theta). \end{aligned} \tag{12}$$

Also, without loss of generality, the parameter space for β_ℓ can be restricted as $\mathcal{N}_\ell$, a small neighborhood of $\beta_{\ell 0}$.

Now, the probability limit of $S_n(\gamma,\beta)$, $S(\gamma,\beta)$, is the same as that in Lemma 1 except that β_ℓ and $\beta_{\ell 0}$ are changed to $\beta_\ell / \|\delta_n\|$ and $\beta_{\ell 0} / \|\delta_n\|$, in other words, we rescale β_ℓ and $\beta_{\ell 0}$. We can similarly analyze the four cases to conclude that $(\beta_\ell - \beta_{\ell 0}) / \|\delta_n\| = o_p(1)$ and $\widehat{\gamma} - \gamma_0 = o_p(1)$, which implies the results we want. ■

Lemma 3 *Under Assumptions D1-D5 and $\beta_{10} - \beta_{20}$ is fixed, $n(\widehat{\gamma} - \gamma_0) = O_p(1)$, and $\sqrt{n}(\beta - \beta_0) = O_p(1)$.*

Proof. This proof uses Corollary 3.2.6 of van der Vaart and Wellner (1996). We follow the notations in Lemma 1.

First, $Q(\theta) - Q(\theta_0) \ge Cd^2(\theta,\theta_0)$, where $Q(\theta)$ is the probability limit of $Q_n(\theta)$, $d(\theta,\theta_0) = \|\beta - \beta_0\| + \sqrt{\|\gamma - \gamma_0\|}$ for $\theta \in \mathcal{N}$ with $\mathcal{N}$ being an open neighborhood of θ_0.

$$\begin{aligned} & Q(\theta) - Q(\theta_0) \\ = \; & \mathbb{E}\left[T(w|\beta_1,\beta_{10}) 1(q \le \mathbf{z}'\gamma \wedge \mathbf{z}'\gamma_0)\right] + \mathbb{E}\left[T(w|\beta_2,\beta_{20}) 1(q > \mathbf{z}'\gamma \vee \mathbf{z}'\gamma_0)\right] \\ & + \mathbb{E}\left[\overline{Z}_1(w|\beta_2,\beta_{10}) 1(\mathbf{z}'\gamma \wedge \mathbf{z}'\gamma_0 < q \le \mathbf{z}'\gamma_0)\right] + \mathbb{E}\left[\overline{Z}_2(w|\beta_1,\beta_{20}) 1(\mathbf{z}'\gamma_0 < q \le \mathbf{z}'\gamma \vee \mathbf{z}'\gamma_0)\right] \end{aligned}$$

$$
\begin{aligned}
&= (\beta_{10}-\beta_1)' \mathbb{E}\left[\mathbf{x}\mathbf{x}'1(q \le \mathbf{z}'\gamma \wedge \mathbf{z}'\gamma_0)\right](\beta_{10}-\beta_1) + (\beta_{20}-\beta_2)' \mathbb{E}\left[\mathbf{x}\mathbf{x}'1(q > \mathbf{z}'\gamma \vee \mathbf{z}'\gamma_0)\right](\beta_{20}-\beta_2) \\
&\quad + (\beta_{10}-\beta_2)' \mathbb{E}\left[\mathbf{x}\mathbf{x}'1(\mathbf{z}'\gamma \wedge \mathbf{z}'\gamma_0 < q \le \mathbf{z}'\gamma_0)\right](\beta_{10}-\beta_2) + (\beta_{20}-\beta_1)' \mathbb{E}\left[\mathbf{x}\mathbf{x}'1(\mathbf{z}'\gamma_0 < q \le \mathbf{z}'\gamma \vee \mathbf{z}'\gamma_0)\right](\beta_{20}-\beta_1) \\
&\ge C\left(\left\|\beta_{1,0}-\beta_1\right\|^2 + \left\|\beta_{2,0}-\beta_2\right\|^2 + \|\gamma-\gamma_0\|\right),
\end{aligned}
$$

where the last inequality is from Assumptions D1-D4.

Second, $\mathbb{E}\left[\sup_{d(\theta,\theta_0)<\delta} \left|\mathbb{G}_n\left(m\left(w|\theta\right)-m\left(w|\theta_0\right)\right)\right|\right] \le C\delta$ for any sufficiently small δ. From Pakes and Pollard (1989) or Andrews (1994), $\{A\left(w|\theta\right): d\left(\theta,\theta_0\right)<\delta\}$, $\{B\left(w|\theta\right): d\left(\theta,\theta_0\right)<\delta\}$, $\{C\left(w|\theta\right): d\left(\theta,\theta_0\right)<\delta\}$, and $\{D\left(w|\theta\right): d\left(\theta,\theta_0\right)<\delta\}$ are all VC subgraph. From Theorem 2.14.1 of Van der Vaart and Wellner (1996),

$$
\mathbb{E}\left[\sup_{d(\theta,\theta_0)<\delta} \left|\mathbb{G}_n\left(m\left(w|\theta\right)-m\left(w|\theta_0\right)\right)\right|\right] \le C\sqrt{PF^2},
$$

where F is the envelope of $\{m\left(w|\theta\right)-m\left(w|\theta_0\right): d\left(\theta,\theta_0\right)<\delta\}$ and can take the form of, e.g.,

$$
\begin{aligned}
F &= \left(\delta^2\|\mathbf{x}\|^2 + 2\delta\|\mathbf{x}\varepsilon_1\|\right)\sup_{d(\theta,\theta_0)<\delta} 1\left(q \le \mathbf{z}'\gamma \wedge \mathbf{z}'\gamma_0\right) + \left(\delta^2\|\mathbf{x}\|^2 + 2\delta\|\mathbf{x}\varepsilon_2\|\right)\sup_{d(\theta,\theta_0)<\delta} 1\left(q > \mathbf{z}'\gamma \vee \mathbf{z}'\gamma_0\right) \\
&\quad + \sup_{d(\theta,\theta_0)<\delta}\left(\|\beta_{10}-\beta_2\|^2\|\mathbf{x}\|^2 + 2\|\beta_{10}-\beta_2\|\|\mathbf{x}\varepsilon_1\|\right) 1\left(\mathbf{z}'\gamma \wedge \mathbf{z}'\gamma_0 < q \le \mathbf{z}'\gamma_0\right) \\
&\quad + \sup_{d(\theta,\theta_0)<\delta}\left(\|\beta_{20}-\beta_1\|^2\|\mathbf{x}\|^2 + 2\|\beta_{20}-\beta_1\|\|\mathbf{x}\varepsilon_2\|\right) 1\left(\mathbf{z}'\gamma_0 < q \le \mathbf{z}'\gamma \vee \mathbf{z}'\gamma_0\right).
\end{aligned}
$$

By Assumptions D4 and D5, $\sqrt{PF^2} \le C\delta$ for $\delta < 1$. So $\phi\left(\delta\right)=\delta$ in Corollary 3.2.6 of Van der Vaart and Wellner (1996) and δ/δ^{α} is decreasing for all $1<\alpha<2$. Since $r_n^2\phi\left(\frac{1}{r_n}\right)=r_n$, $\sqrt{n}d\left(\widehat{\theta}-\theta_0\right)=O_P\left(1\right)$. By the definition of d, the result follows. ■

Lemma 4 *Under Assumptions D1-D5 and* $\|\delta_n\|\to 0$, $\sqrt{n}\|\delta_n\|\to\infty$, $a_n\left(\widehat{\gamma}-\gamma_0\right)=O_p(1)$, *and* $\sqrt{n}\left(\beta-\beta_0\right)=O_p(1)$.

Proof. Since δ_n depends on n, Corollary 3.2.6 of van der Vaart and Wellner (1996) cannot be used. Nevertheless, we can apply the proof idea of Theorem 3.2.5 in van der Vaart and Wellner (1996) to prove this result. Define $d_n\left(\theta,\theta_0\right)=\|\beta-\beta_0\|+\|\delta_n\|\sqrt{\|\gamma-\gamma_0\|}$ for θ in a neighborhood of θ_0. For each n, the parameter space (minus θ_0) can be partitioned into the "shells" $S_{j,n}=\left\{\theta: 2^{j-1}<\sqrt{n}d_n\left(\theta,\theta_0\right)\le 2^j\right\}$ with j ranging over the integers. Given an integer J,

$$
\begin{aligned}
P\left(d_n\left(\widehat{\theta},\theta_0\right)>2^J\right) \le & \sum_{j\ge J,\|\beta-\beta_0\|<M\|\delta_n\|,\|\gamma-\gamma_0\|<\eta} P\left(\inf_{\theta\in S_{j,n}}\left(Q_n(\theta)-Q_n(\theta_0)\right)\le 0\right) \\
& + P\left(2\|\beta-\beta_0\|\ge M\|\delta_n\|, 2\|\gamma-\gamma_0\|\ge\eta\right),
\end{aligned} \tag{13}
$$

where $Q_n(\theta)-Q_n(\theta_0)=S_n(\theta)$ is defined in (12), and M and η are small positive numbers. The second term on the right hand side of (13) converges to zero as $n\to\infty$ for every $\eta>0$ and $M>0$ by Lemma 2, so we can concentrate on the first term.

$$
\begin{aligned}
& P\left(\inf_{\theta\in S_{j,n}}\left(Q_n(\theta)-Q_n(\theta_0)\right)\le 0\right) \\
\le\ & P\left(\sup_{\theta\in S_{j,n}}\left|Q_n(\theta)-Q_n(\theta_0)-\mathbb{E}\left[Q_n(\theta)-Q_n(\theta_0)\right]\right| \ge \inf_{\theta\in S_{j,n}}\left|\mathbb{E}\left[Q_n(\theta)-Q_n(\theta_0)\right]\right|\right) \\
\le\ & \mathbb{E}\left[\sup_{\theta\in S_{j,n}}\left|Q_n(\theta)-Q_n(\theta_0)-\mathbb{E}\left[Q_n(\theta)-Q_n(\theta_0)\right]\right|\right] \Big/ \inf_{\theta\in S_{j,n}}\left|\mathbb{E}\left[Q_n(\theta)-Q_n(\theta_0)\right]\right|
\end{aligned}
$$

$$\le \sum_{k=1}^{4} \mathbb{E}\left[\sup_{\theta\in S_{j,n}} \left|T_k(\theta)-\mathbb{E}\left[T_k(\theta)\right]\right|\right] \Big/ \inf_{\theta\in S_{j,n}} \left|\mathbb{E}\left[Q_n(\theta)-Q_n(\theta_0)\right]\right|,$$

where the last equality is from Markov's inequality, and $T_k(\theta)$, $k=1,2,3$, is defined in (12).

From Lemma 2, it is not hard to see that

$$\begin{aligned}
&\inf_{\theta\in S_{j,n}} \left|\mathbb{E}\left[Q_n(\theta)-Q_n(\theta_0)\right]\right| = \inf_{\theta\in S_{j,n}} \left|\sum\nolimits_{k=1}^{4} \mathbb{E}\left[T_k(\theta)\right]\right| \\
&= \inf_{\theta\in S_{j,n}} C\left|\tfrac{n}{a_n}\left\|\beta-\beta_0\right\|^2+\tfrac{n}{a_n}\left\|\beta_{10}-\beta_2\right\|^2\left\|\gamma-\gamma_0\right\|+\tfrac{n}{a_n}\left\|\beta_{20}-\beta_1\right\|^2\left\|\gamma-\gamma_0\right\|\right| \\
&= \inf_{\theta\in S_{j,n}} C\left|\tfrac{n}{a_n}\left\|\beta-\beta_0\right\|^2+\tfrac{n}{a_n}\left\|\delta_n\right\|^2\left\|\gamma-\gamma_0\right\|\right| = \inf_{\theta\in S_{j,n}} C\tfrac{n}{a_n}d_n(\theta,\theta_0)^2 \ge C\tfrac{2^{2j-2}}{a_n} = C\tfrac{2^{2j}}{a_n},
\end{aligned}$$

where the third equality is because $\beta_{10}-\beta_{20}=\delta_n$ and $\left\|\beta_\ell-\beta_{\ell 0}\right\|<M\left\|\delta_n\right\|$ so that $\left\|\beta_1-\beta_{20}\right\|=O(\left\|\delta_n\right\|)$ and $\left\|\beta_{20}-\beta_1\right\|=O(\left\|\delta_n\right\|)$. From Lemma 3, for $k=1,2$,

$$\sum_{k=1}^{2}\mathbb{E}\left[\sup_{\theta\in S_{j,n}}\left|T_k(\theta)-\mathbb{E}\left[T_k(\theta)\right]\right|\right]\le C\frac{\sup_{\theta\in S_{j,n}}\left\|\beta-\beta_0\right\|}{\sqrt{n}\delta_n'\delta_n}.$$

As to $T_3(\theta)$, applying a maximal inequality (e.g., Theorem 2.14.1 of van der Vaart and Wellner (1996)) we can show that

$$\mathbb{E}\left[\sup_{\theta\in S_{j,n}}\left|T_3(\theta)-\mathbb{E}\left[T_3(\theta)\right]\right|\right]\le C\frac{\sup_{\theta\in S_{j,n}}\sqrt{\left\|\beta_{10}-\beta_2\right\|^2}\sqrt{\left|\gamma-\gamma_0\right|}}{\sqrt{n}\delta_n'\delta_n}=\frac{\sup_{\theta\in S_{j,n}}\left\|\delta_n\right\|\sqrt{\left|\gamma-\gamma_0\right|}}{\sqrt{n}\delta_n'\delta_n}.$$

Similarly, $\mathbb{E}\left[\sup_{\theta\in S_{j,n}}\left|T_3(\theta)-\mathbb{E}\left[T_3(\theta)\right]\right|\right]\le C\frac{\sup_{\theta\in S_{j,n}}\left\|\delta_n\right\|\sqrt{\left|\gamma-\gamma_0\right|}}{\sqrt{n}\delta_n'\delta_n}$. So

$$\sum_{k=1}^{4}\mathbb{E}\left[\sup_{\theta\in S_{j,n}}\left|T_k(\theta)-\mathbb{E}\left[T_k(\theta)\right]\right|\right]\le C\frac{\sup_{\theta\in S_{j,n}} d_n(\theta,\theta_0)}{\sqrt{n}\delta_n'\delta_n}\le C\frac{2^j/\sqrt{n}}{\sqrt{n}\delta_n'\delta_n}=C\frac{2^j}{a_n}.$$

In summary,

$$\sum_{j\ge J,\left\|\beta-\beta_0\right\|<M\left\|\delta_n\right\|,\left|\gamma-\gamma_0\right|<\eta} P\left(\sup_{\theta\in S_{j,n}}\left(Q_n(\theta)-Q_n(\theta_0)\right)\ge 0\right)\le C\sum_{j\ge J}\left(\frac{2^j}{a_n}\Big/\frac{2^{2j}}{a_n}\right)\le C\sum_{j\ge J}\frac{1}{2^j},$$

which can be made arbitrarily small by letting J large enough. So $\sqrt{n}d_n\left(\widehat{\theta},\theta_0\right)=O_p(1)$, which implies $a_n(\widehat{\gamma}-\gamma_0)=O_p(1)$, and $\sqrt{n}(\beta-\beta_0)=O_p(1)$. ■

Lemma 5 *Under Assumption D, uniformly for $h=(u',v)'$ in any compact set of $\mathbb{R}^{2(d+1)+1}$,*

$$\begin{aligned}
&nP_n\left(m\left(\cdot\left|\beta_0+\frac{u}{\sqrt{n}},\gamma_0+\frac{v}{n}\right.\right)-m\left(\cdot\left|\beta_0,\gamma_0\right.\right)\right) \\
&= u_1'\mathbb{E}\left[\mathbf{x}\mathbf{x}'1(\epsilon\le 0)\right]u_1+u_2'\mathbb{E}\left[\mathbf{x}\mathbf{x}'1(\epsilon>0)\right]u_2-2W_n(u)+D_n(v)+o_p(1),
\end{aligned}$$

where $u = (u_1', u_2')' \in \mathbb{R}^{2(d+1)}$, *and*

$$\begin{aligned} D_n(v) &= \sum_{i=1}^n \overline{Z}_{1i} 1(\mathbf{z}_i' v < n\epsilon_i \le 0) + \sum_{i=1}^n \overline{Z}_{2i} 1(0 < n\epsilon_i \le \mathbf{z}_i' v), \\ W_n(u) &= W_{1n}(u_1) + W_{2n}(u_2), \end{aligned}$$

with

$$W_{1n}(u_1) = u_1' \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i \varepsilon_{1i} 1(\epsilon_i \le 0) \right), \ W_{2n}(u_2) = u_2' \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i \varepsilon_{2i} 1(\epsilon_i > 0) \right).$$

Proof. First note that

$$\begin{aligned} & nP_n \left(m\left(\cdot \middle| \beta_0 + \tfrac{u}{\sqrt{n}}, \gamma_0 + \tfrac{v}{n} \right) - m(\cdot|\beta_0, \gamma_0) \right) \\ = & \sum_{i=1}^n \left(u_1' \tfrac{\mathbf{x}_i \mathbf{x}_i'}{n} u_1 - u_1' \tfrac{2}{\sqrt{n}} \mathbf{x}_i \varepsilon_{1i} \right) 1(q_i \le \mathbf{z}_i' \left(\gamma_0 + \tfrac{v}{n}\right) \wedge \mathbf{z}_i'\gamma_0) \\ + & \sum_{i=1}^n \left(u_2' \tfrac{\mathbf{x}_i \mathbf{x}_i'}{n} u_2 - u_2' \tfrac{1}{\sqrt{n}} \mathbf{x}_i \varepsilon_{2i} \right) 1(q_i > \mathbf{z}_i'\gamma_0 \vee \mathbf{z}_i' \left(\gamma_0 + \tfrac{v}{n}\right)) \\ + & \sum_{i=1}^n \left[\left(\beta_{10} - \beta_{20} - \tfrac{u_2}{\sqrt{n}} \right)' \mathbf{x}_i \mathbf{x}_i' \left(\beta_{10} - \beta_{20} - \tfrac{u_2}{\sqrt{n}} \right) + 2\mathbf{x}_i' \left(\beta_{10} - \beta_{20} - \tfrac{u_2}{\sqrt{n}} \right) \varepsilon_{1i} \right] 1\left(\mathbf{z}_i' \left(\gamma_0 + \tfrac{v}{n}\right) < q_i \le \mathbf{z}_i'\gamma_0\right) \\ + & \sum_{i=1}^n \left[\left(\beta_{1,0} + \tfrac{u_1}{\sqrt{n}} - \beta_{2,0} \right)' \mathbf{x}_i \mathbf{x}_i' \left(\beta_{1,0} + \tfrac{u_1}{\sqrt{n}} - \beta_{2,0} \right) - 2\mathbf{x}_i' \left(\beta_{10} + \tfrac{u_1}{\sqrt{n}} - \beta_{20} \right) \varepsilon_{2i} \right] 1\left(\mathbf{z}_i'\gamma_0 < q_i \le \mathbf{z}_i' \left(\gamma_0 + \tfrac{v}{n}\right)\right) \\ := & T_1(h) + T_2(h) + T_3(h) + T_4(h). \end{aligned}$$

We take T_1 and T_3 as examples since T_2 and T_4 can be similarly analyzed.

T_1: By a Glivenko-Cantelli theorem and Assumption D4,

$$\frac{1}{n} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i' 1\left(q_i \le \mathbf{z}_i' \left(\gamma_0 + \frac{v}{n} \right) \wedge \mathbf{z}_i'\gamma_0 \right) \xrightarrow{p} \mathbb{E}\left[\mathbf{x}_i \mathbf{x}_i' 1(q_i \le \mathbf{z}_i'\gamma_0) \right],$$

where $\mathbb{E}[\mathbf{x}_i \mathbf{x}_i' 1(q_i \le \mathbf{z}_i'\gamma_0)] = \mathbb{E}[\mathbf{x}_i \mathbf{x}_i' 1(\epsilon_i \le 0)]$ by the definition of ϵ_i. By stochastic equicontinuity of $\mathbb{G}_n \mathbf{x}_i \varepsilon_{1i} 1(q_i \le \mathbf{z}_i'\gamma)$ as a function of γ,

$$\mathbb{G}_n \mathbf{x}_i \varepsilon_{1i} 1\left(q_i \le \mathbf{z}_i' \left(\gamma_0 + \frac{v}{n} \right) \wedge \mathbf{z}_i'\gamma_0 \right) - \mathbb{G}_n \mathbf{x}_i \varepsilon_{1i} 1(q_i \le \mathbf{z}_i'\gamma_0) = o_p(1).$$

T_3: We need to show that

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i' 1\left(\mathbf{z}_i' \left(\gamma_0 + \frac{v}{n} \right) < q_i \le \mathbf{z}_i'\gamma_0 \right) &= o_p(1), \\ \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i' \varepsilon_{1i} 1\left(\mathbf{z}_i' \left(\gamma_0 + \frac{v}{n} \right) < q_i \le \mathbf{z}_i'\gamma_0 \right) &= o_p(1). \end{aligned}$$

The former follows from $n^{-1} \sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i' 1(\mathbf{z}_i'(\gamma_0 + v/n) < q_i \le \mathbf{z}_i'\gamma_0) = O_p(n^{-1})$ by a Glivenko-Cantelli theorem and Assumption D3. The latter follows the stochastic equicontinuity of $\mathbb{G}_n \mathbf{x}_i \varepsilon_{1i} 1(q_i \le \mathbf{z}_i'\gamma)$ as a function of γ. ■

Lemma 6 *Under Assumptions D1-D8 and* $\|\delta_n\| \to 0$, $\sqrt{n}\|\delta_n\| \to \infty$, *uniformly for* $h = (u', v)'$ *in any compact set of* $\mathbb{R}^{2(d+1)+1}$,

$$
\begin{aligned}
& nP_n \left(m \left(\cdot \left| \beta_0 + \frac{u}{\sqrt{n}}, \gamma_0 + \frac{v}{a_n} \right. \right) - m \left(\cdot \left| \beta_0, \gamma_0 \right. \right) \right) \\
& = u_1' \mathbb{E}\left[\mathbf{x}\mathbf{x}' 1\left(\epsilon \le 0 \right) \right] u_1 + u_2' \mathbb{E}\left[\mathbf{x}\mathbf{x}' 1\left(\epsilon > 0 \right) \right] u_2 - 2W_n\left(u\right) + 2C_n\left(v\right) + o_p\left(1\right),
\end{aligned}
$$

where $W_n(u)$ is defined in the last lemma, and

$$
C_n\left(v\right) = B_{1n}(v) - B_{2n}(v) + \frac{1}{2} I(v), \tag{14}
$$

with $I(v)$ defined in Theorem 2 and

$$
B_{1n}(v) = \delta_n' \left[\sum_{i=1}^n \mathbf{x}_i \varepsilon_{1i} 1\left(\mathbf{z}_i' v / a_n < \epsilon_i \le 0 \right) \right], B_{2n}(v) = \delta_n' \left[\sum_{i=1}^n \mathbf{x}_i \varepsilon_{2i} 1\left(0 < \epsilon_i \le \mathbf{z}_i' v / a_n \right) \right].
$$

Proof. We first decompose $nP_n \left(m \left(\cdot \left| \beta_0 + \frac{u}{\sqrt{n}}, \gamma_0 + \frac{v}{a_n} \right. \right) - m \left(\cdot \left| \beta_0, \gamma_0 \right. \right) \right)$ in a similar way as in the last lemma; the only difference is to replace n by a_n and $\beta_{10} - \beta_{20}$ by δ_n. The approximation for the regular parameter β is similar, so we concentrate on the approximation for the nonregular parameter γ here. Different from the fixed-threshold-effect case, we must combine $T_3(h)$ and $T_4(h)$ for the approximation since for a fixed v, some z_i's are included in $T_3(h)$ while the others are included in $T_4(h)$.

First note that

$$
\delta_n' \left[\sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i' 1 \left(\mathbf{z}_i' \left(\gamma_0 + \frac{v}{a_n} \right) < q_i \le \mathbf{z}_i' \gamma_0 \right) \right] \delta_n \xrightarrow{p} I_1(v)
$$

by calculating the mean (which converges to $I_1(v)$) and variance (which is $O(\delta_n' \delta_n)$). Similarly,

$$
\delta_n' \left[\sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i' 1 \left(\mathbf{z}_i' \gamma_0 < q_i \le \mathbf{z}_i' \left(\gamma_0 + \frac{v}{a_n} \right) \right) \right] \delta_n \xrightarrow{p} I_2(v).
$$

Next,

$$
\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i \varepsilon_{1i} 1 \left(\mathbf{z}_i' \left(\gamma_0 + \frac{v}{a_n} \right) < q_i \le \mathbf{z}_i' \gamma_0 \right) = o_p(1)
$$

by the stochastic equicontinuity of $\mathbb{G}_n \mathbf{x}_i \varepsilon_{1i} 1(q_i \le \mathbf{z}_i' \gamma)$ as a function of γ. Similarly,

$$
\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i \varepsilon_{2i} 1 \left(\mathbf{z}_i' \gamma_0 < q_i \le \mathbf{z}_i' \left(\gamma_0 + \frac{v}{n} \right) \right) = o_p(1).
$$

So the approximation is valid. ■

Lemma 7 *Under Assumption D, the finite-dimensional (fidi) weak limit of $D_n(v)$ is the same as $D(v)$.*

Proof. The behavior of $D_n(v)$ is determined by near-to-threshold observations, whose behavior is described using point processes. Our proof includes two steps. Step 1 constructs a point process and derives its limit. Step 2 applies Step 1 to obtain the fidi-limit of $D_n(v)$ and shows the asymptotic independence between $D_n(v)$ and $W_n(u)$.

Step 1. The intuition for Step 1 is provided after Theorem 1.

Define $E := \mathbb{R} \times \mathbb{Z}$ with the usual Euclidean topology. Define the point process of interest as follows: for any Borel subset A, $\widehat{\mathbf{N}}(A) = \sum_{i=1}^n 1\left((n\epsilon_i, z_i) \in A \right)$. Take $\widehat{\mathbf{N}}$ as a random element of $M_p(E)$, the metric space of nonnegative point measures on E, with the metric generated by the topology of vague convergence; cf. Resnick (1987, Chapter 3). We show that $\widehat{\mathbf{N}} \rightsquigarrow \mathbf{N}$ in $M_p(E)$, for $\mathbf{N}$ given in the comments after Theorem 1. This is done in the steps (a) and (b).

(a) For any $F \in \mathcal{T}$, the basis of relatively compact open sets in E (finite unions of bounded rectangles, cf. the remark after Proposition 3.22 of Resnick (1987)),

$$\lim_{n\to\infty} \mathbb{E}\left[\widehat{\mathbf{N}}(F)\right] := \lim_{n\to\infty} nP\left((n\epsilon_i, z_i) \in F\right) = \int_F f_{\epsilon|z}\left(0|z\right) dF_z(z) = m(F) < \infty,$$

where the measure m is defined as $dm(v,z) = f_{\epsilon|z}\left(0|z\right) dv dF_z(z)$. Since $\{(n\epsilon_i, z_i) \in F\}$ are independent across i, by Meyer's Theorem (cf. Meyer (1973))

$$\lim_{n\to\infty} P\left(\widehat{\mathbf{N}}(F) = 0\right) = e^{-m(F)}.$$

By Proposition 3.22 of Resnick (1987), $\widehat{\mathbf{N}} \rightsquigarrow \mathbf{N}$ in $M_p(E)$, where $\mathbf{N}$ is a Poisson point process with the mean intensity measure $m(\cdot)$.

(b) In this step, we show $\mathbf{N}$ has the same distribution as $\mathbf{N}$ given after Theorem 1. First, consider the canonical Poisson processes N_{10} and N_{20} with points $\{\mathcal{J}_{1i}\}$ and $\{\mathcal{J}_{2i}\}$ defined in Theorem 1. N_{10} has the mean measure $m_{10}(du) = du$ on $(-\infty, 0]$, and N_{20} has the mean measure $m_{20}(du) = du$ on $(0,\infty)$; see Resnick (1987, p. 138). Because N_{10} and N_{20} are independent, $N_{12}(\cdot) := N_{10}(\cdot) + N_{20}(\cdot)$ is a Poisson point process with mean measure $m_{12}(du) = du$ on $\mathbb{R}$ by definition of the Poisson process; see Resnick (1987, p. 130). Because $\{z_i, z_i'\}$ are i.i.d. and independent of $\{\mathcal{J}_{1i}, \mathcal{J}_{2i}\}$, by Proposition 3.8 in Resnick (1987), the composed process N_{12}' with points $(\{\mathcal{J}_{1i}, z_{1i}\}, \{\mathcal{J}_{2i}, z_{2i}\}, i \geq 1)$ is a Poisson process with the mean measure $m_{12}'(dv, dz) = dv dF_z(z)$ on E. Finally, $\mathbf{N}$ with the points $\{T(\mathcal{J}_{1i}, z_{1i}), T(\mathcal{J}_{2i}, z_{2i})\}$, where $T : (v,z) \mapsto \left(v/f_{\epsilon|z}\left(0|z\right), z\right)$, is a Poisson process with the desired mean measure $m(dv,dz) = m_{12}' \circ T^{-1}(dv,dz) = f_{\epsilon|z}\left(0|z\right) dv F_z(dz)$, by Proposition 3.7 in Resnick (1987).

Step 2. Because $D_n(v)$ cannot be written as a Lebesgue integral with respect to $\widehat{N}$, we cannot apply the continuous mapping theorem as in Chernozhukov and Hong (2004) to derive the weak limit of $D_n(v)$. Rather, we apply Theorem 7.6 of Billingsley (1968), i.e., use the convergence of characteristic functions, to prove our results. Define

$$\begin{aligned}
&T_1 = \tfrac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i \varepsilon_{1i} 1\left(\epsilon_i \leq 0\right) =: \tfrac{1}{\sqrt{n}} \sum_{i=1}^n S_{1i},\ T_2 = \tfrac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{x}_i \varepsilon_{1i} 1\left(\epsilon_i > 0\right) =: \tfrac{1}{\sqrt{n}} \sum_{i=1}^n S_{2i}, \\
&T_3 = \sum_{i=1}^n \overline{Z}_{1i} 1\left(\mathbf{z}_i' v_1 < n\epsilon_i \leq 0\right),\ T_4 = \sum_{i=1}^n \overline{Z}_{2i} 1\left(0 < n\epsilon_i \leq \mathbf{z}_i' v_2\right)
\end{aligned}$$

for some v_1 and v_2. We have

$$\begin{aligned}
&\mathbb{E}\left[\exp\left\{\sqrt{-1}\left[s_1' T_1 + s_2' T_2 + t_1 T_3 + t_2 T_4\right]\right\}\right] \\
&= \textstyle\prod_{i=1}^n \left\{1 + \frac{1}{n}\left[-\frac{1}{2} s_1' \mathbb{E}\left[S_{1i} S_{1i}'\right] s_1 - \frac{1}{2} s_2' \mathbb{E}\left[S_{2i} S_{2i}'\right] s_2\right] + o\left(\frac{1}{n}\right)\right\} \\
&\quad \cdot \mathbb{E}\left[\exp\left\{\sqrt{-1}\left[t_1 \sum_{i=1}^n \left(Z_{1i} 1\left(\mathbf{z}_i' v_1 < n\epsilon_i \leq 0\right) + o\left(\tfrac{1}{n}\right)\right) + t_2 \sum_{i=1}^n \left(Z_{2i} 1\left(0 < n\epsilon_i \leq \mathbf{z}_i' v_2\right) + o\left(\tfrac{1}{n}\right)\right)\right]\right\}\right] \\
&\to \exp\left\{-\tfrac{1}{2} s_1' \mathbb{E}\left[S_{1i} S_{1i}'\right] s_1 - \tfrac{1}{2} s_2' \mathbb{E}\left[S_{2i} S_{2i}'\right] s_2\right\} \cdot \mathbb{E}\left[\exp\left\{\sqrt{-1} t_1 \sum_{i=1}^n Z_{1i} 1\left(\mathbf{z}_i' v_1 < J_{1i} \leq 0\right)\right\}\right] \\
&\quad \cdot \mathbb{E}\left[\exp\left\{\sqrt{-1} t_2 \sum_{i=1}^n Z_{2i} 1\left(0 < J_{2i} \leq \mathbf{z}_i' v_2\right)\right\}\right],
\end{aligned}$$

where the equality is from Assumptions D4 and D5, the first part of the limit is standard, and the second part needs more explanation. Define the functional $T : \mathbb{R}^\infty \times M_p(E) \longmapsto \mathbb{R}^2$ as $(Z_{1i}, Z_{2i}, \cdots, N) \longmapsto$

$\left(\sum_{i=1}^{n} Z_{1i} 1\left(\left(1, \mathsf{z}_i'\right) v_1 < \mathsf{J}_i \leq 0\right), \sum_{i=1}^{n} Z_{2i} 1\left(0 < \mathsf{J}_i \leq \left(1, \mathsf{z}_i'\right) v_2\right)\right)$, where $N(\cdot) = \sum_{i=1}^{n} 1\left(\left(\mathsf{J}_i, \mathsf{z}_i\right) \in \cdot\right)$. By Proposition 3.13 of Resnick (1987), T is discontinuous at $\mathcal{D}(T) := \mathbb{R}^{\infty} \times \mathcal{D}(N)$, where $\mathcal{D}(N) = \{N \in M_p(E) : \mathsf{J}^N = \left(1, \mathsf{z}_i^{N\prime}\right) v_1$ or $\mathsf{J}^N = \left(1, \mathsf{z}_i^{N\prime}\right) v_2\}$ for some $i \geq 1$, and $\left(\mathsf{J}^N, \mathsf{z}_i^N, i \geq 1\right)$ denote the points of N. By definition of $\mathbf{N}$, $P\left(\left(Z_{1i}, Z_{2i}, \cdots, \mathbf{N}\right) \in D(T)\right) = P\left(\mathbf{N} \in \mathcal{D}(\mathbf{N})\right) = 0$. By the continuous mapping theorem (cf. Resnick (1987, p.152)), it follows that

$$\left(\sum_{i=1}^{n} Z_{1i} 1\left(\left(1, \mathsf{z}_i'\right) v_1 < \mathsf{J}_i \leq 0\right), \sum_{i=1}^{n} Z_{2i} 1\left(0 < \mathsf{J}_i \leq \left(1, \mathsf{z}_i'\right) v_2\right)\right) \rightsquigarrow \left(\sum_{i=1}^{n} Z_{1i} 1\left(\mathbf{z}_i' v_1 < J_{1i} \leq 0\right), \sum_{i=1}^{n} Z_{2i} 1\left(0 < J_{2i} \leq \mathbf{z}_i' v_2\right)\right)$$

if $N \rightsquigarrow \mathbf{N}$. From Step 1 and Theorem 7.6 of Billingsley (1968), the second part of the convergence holds. Finally, by Theorem 7.6 of Billingsley (1968), the convergence of the characteristic function implies $D_n(v) \rightsquigarrow D(v)$ and is asymptotically independent of $W_n(u)$. ■

Lemma 8 *Under Assumptions D1-D7 and* $\|\delta_n\| \to 0$, $\sqrt{n}\|\delta_n\| \to \infty$, $C_n(v) \rightsquigarrow C(v)$ *on any compact set of* $\mathbb{R}^{2(d+1)+1}$, *where* $C(v)$ *is defined in Theorem 2.*

Proof. This proof includes two parts: (i) the fidi weak limit of $C_n(v)$ are the same as $C(v)$; (ii) the process $C_n(v)$ is stochastically equicontinuous.

Part (i): This can be checked by direct calculation. Also, it can be checked that $C_n(v)$ is asymptotically independent of $W_n(u)$ by using the characteristic function as in the last lemma.

Part (ii): It is obvious that $I(v)$ in continuous in v, so we are left to prove $B_{1n}(v)+B_{2n}(v)$ is stochastically equicontinuous. Without loss of generality, we prove the result only for $B_{2n}(v)$ here. For v_1, v_2 in a compact set and for any $\eta > 0$,

$$P\left(\sup_{\|v_2 - v_1\| < \delta} \left|B_{2n}(v_2) - B_{2n}(v_1)\right| > \eta\right) \overset{(1)}{\leq} P\left(\sup_{\|v_2 - v_1\| < \delta} \left|\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \delta_n' \mathbf{x}_i \varepsilon_{2i} 1\left(\frac{\mathbf{z}_i' v_1 \wedge \mathbf{z}_i' v_2}{a_n} < \epsilon_i \leq \frac{\mathbf{z}_i' v_1 \vee \mathbf{z}_i' v_2}{a_n}\right)\right| > \frac{\eta}{\sqrt{n}}\right)$$
$$\overset{(2)}{\leq} C \sqrt{\delta_n' \delta_n \mathbb{E}\left[\mathbb{E}\left[\|\mathbf{x}\|^2 \varepsilon_2^2 \middle| z, \epsilon = 0\right] f_{\epsilon|z}(0|z) \|\mathbf{z}\|\right] \delta / a_n} \Big/ \frac{\eta}{\sqrt{n}} \overset{(3)}{\leq} \frac{C\sqrt{\delta}}{\eta},$$

where (1) is obvious, (2) is from Markov's inequality and a maximal inequality (e.g., Theorem 2.14.2 of van der Vaart and Wellner (1996)), and C in (3) is finite from Assumptions D4 and D5. So we can choose a small enough δ (which may depend on η) such that $P\left(\sup_{\|v_2 - v_1\| < \delta} \left|B_{2n}(v_2) - B_{2n}(v_1)\right| > \epsilon\right)$ is less than any specified level. ■

Lemma 9 (Bounding Moduli of Continuity) *Under Assumption D, for all* $n \geq n_0$, *where* n_0 *is sufficiently large, and any bounded rectangle* R,

$$P\left(\omega_{D_n}(R, \varphi) = 1\right) \leq C\,|R|\,\varphi$$

for all sufficiently small $\varphi > 0$, *where* $|R| = \sup\{\|v\| : v \in R\}$, *and* $\omega_{D_n}(R, \varphi) = 1\left(\inf_{v \in R} D_n(v) < \inf_{v \in \{v_{kj}\} \cap R} D_n(v)\right)$ *is a Skorohod-type modulus.*

Proof. This proof essentially follows from Lemma G.3 of Chernozhukov and Hong (2003b). To avoid duplication and show the main structure of this proof, we use a simple example with $\dim(z) = 1$ to mark the key points.

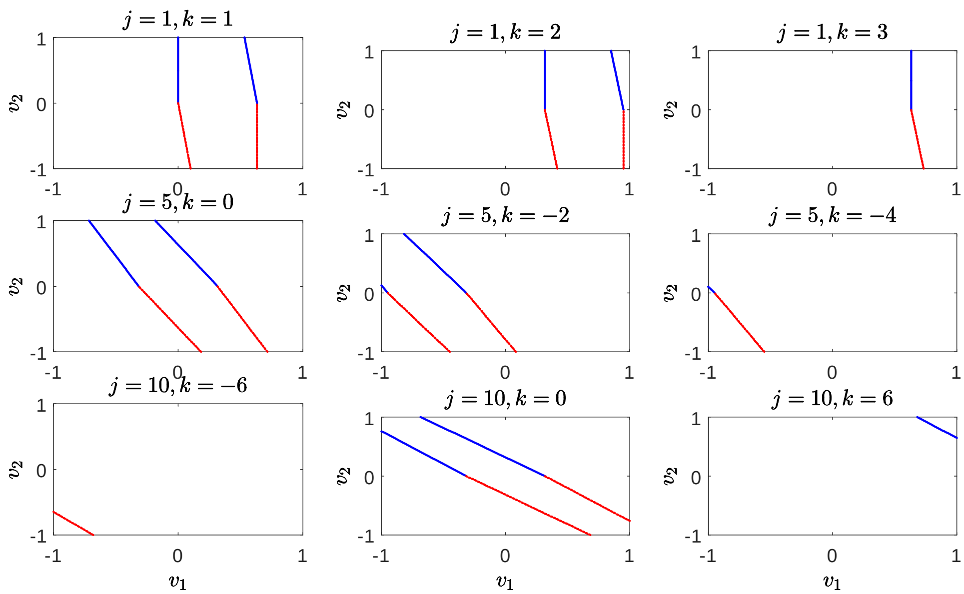

Figure 5: V_{kj} for Different Combinations of j and k When $C=1$, $\phi=0.1$ and $\varphi=\sqrt{0.1}$

(a) [Covering Sets] Cover $[0,1]$, the support of z, by the minimal number of closed equal-lengthed intervals $\{I_{\phi,j}, j\leq J(\phi)\}$ with $\phi<1$; e.g., $I_{\phi,j}=[(j-1)\phi, j\phi]$. Construct the (overlapping) sets

$$\{V_{kj}, k=-m,\cdots,m, j=1,\cdots,J(\phi)\}\subset\mathbb{R}^2$$

such that

$$V_{kj}:=\left\{v\in\mathbb{R}^2:\nu_k-\varphi\leq v_1+v_2z\leq\nu_k+\varphi \text{ such that } z\in I_{\phi,j}\right\},$$

where $\varphi>0$ and $\nu_k=k\varphi$, for $k=-m,\cdots,0,\cdots,m$. Suppose $R=[-C,C]\times[-C,C]$, then the range of v_1+v_2z is $O(C)$, and we can cover the range by $2m+1$ brackets of the form $[\nu_k-\varphi,\nu_k+\varphi]$ where $m=O(C/\varphi)$. Choose $\phi=O\left(\varphi^2\right)$ for all small φ. Hence the total number L of covering sets V_{kj} is bounded as $L\leq(2m+1)J(\phi)$ and grows at most at rate C/φ^3. Next, construct the "centers" v_{kj} in $V_{kj}\cap R$ so that,

$$\delta_{kj}\leq v_{1kj}+v_{2kj}z\leq\delta_{kj}+\eta \text{ for any } z\in I_{\phi,j},$$

where $\delta_{kj}=\inf_{v\in V_{kj}\cap R, z\in I_{\phi,j}}\{v_1+v_2z\}$, and $\eta=O\left(\varphi^2\right)$. For this special case, $\{V_{kj}\}_{k=-m}^{m}$ for any j can cover R, and η can take 0. Figure 5 shows the form of V_{kj} for different combinations of j and k when $C=1$, where V_{kj} is the area surrounded by the two (or one) lines. In this figure, we set $\phi=0.1$, so $J(\phi)=10$. $\varphi=\sqrt{0.1}$. When $j=1$, $m=3$; when $j=5$, $m=4$; when $j=10$, $m=6$. Obviously, v_{kj} may not be unique.

(b) [Characterization of Break-Points] Recall that

$$\begin{aligned} D_n(v) &= \sum_{i=1}^{n}\overline{Z}_{1i}1\left(\mathbf{z}_i'v<n\epsilon_i\leq 0\right)+\sum_{i=1}^{n}\overline{Z}_{2i}1\left(0<n\epsilon_i\leq\mathbf{z}_i'v\right)\\ &=: D_n^-(v)+D_n^+(v), \end{aligned}$$

where $\mathbf{z}_i'v = v_1 + v_2 z_i$. We will examine the nature of the discontinuities of $D_n(v)$ by examining those of $D_n^+(v)$ for an example since the case for $D_n^-(v)$ is similar.

Suppose we have $n\epsilon_i = v_1 + v_2 z_i$ for some $v \in V_{kj}$ and $z_i \in I_{\phi,j}$, then the pair $(n\epsilon_i, z_i)$ is said to induce a break-point in the set V_{kj} and in the bracket $[\nu_k - \varphi, \nu_k + \varphi]$ to which $v_1 + v_2 z_i$ belongs. Given that this is the only pair that induces a break-point in V_{kj} it follows that

$$\inf_{v \in V_{kj} \cap R^{\gamma}} D_n^+(v) \neq D_n^+(v_{kj}) \text{ only if } n\epsilon_i \in [\delta_{kj}, \delta_{kj} + \eta],$$

since $D_n^+(v)$ is piecewise-constant. We need to control the probability of two kinds of events. First, the errors $n\epsilon_i$ are not separated in non-overlapping brackets, which is the event

$$A_1(R) := \cup_{|k|<m} \{\text{there are } n\epsilon_i,\, n\epsilon_{i'} \in [\nu_k - \varphi, \nu_k + \varphi]\}.$$

Second, we need to control the probability that for all $n\epsilon_i$ that are separated into the brackets $[\nu_k - \varphi, \nu_k + \varphi]$, they do not fall into the "bad subset" $[\delta_{kj}, \delta_{kj} + \eta]$ of such brackets, given that $z_i \in I_{\phi,j}$. Formally, conditionally on the complement of $A_1(R)$, i.e., on $A_1^c(R)$ define the event $A_2(R)$ as the union of

$$A_{2i,k,j}(R) := \{n\epsilon_i \in [\delta_{kj}, \delta_{kj} + \eta] |\, n\epsilon_i \in [\nu_k - \varphi, \nu_k + \varphi], z_i \in I_{\phi,j}, v \in V_{kj}\}$$

across $i \leq n$, $|k| < m$, $j \leq J(\phi)$.

First,

$$\begin{aligned} P(A_1(R)) &\leq \sum_{|k|<m} \sum_{i'=1:i'\neq i}^{n} \sum_{i=1}^{n} P(n\epsilon_i,\, n\epsilon_i' \in [\nu_k - \varphi, \nu_k + \varphi]) \\ &\leq (2m+1)(2C\varphi)^2 \leq C|R|\varphi. \end{aligned}$$

Denote the total number of $n\epsilon_i$ that fall into brackets of the form $[\nu_k - \varphi, \nu_k + \varphi]$ by N_n. By a similar analysis in Chernozhukov and Hong (2003b),

$$P(A_2(R)|N_n, A_1^c(R)) \leq CN_n \sup_{i\leq n, |k|<m, j\leq J(\phi)} P(A_{2i,k,j}(R)) \leq CN_n \frac{\eta}{2\varphi}.$$

Since $\mathbb{E}[N_n] \leq n\mathbb{E}[|n\epsilon_i| \leq (1 + |\mathbb{Z}|)|R|] \leq C|R|$,

$$P(A_2(R)|A_1^c(R)) \leq C|R|\frac{\eta}{\varphi}.$$

Since η can take 0 in this special case and ϵ_i is absolutely continuous, $A_2(R)$ has probability zero. Hence,

$$\begin{aligned} P\left(\bigcup_{k,j}\left\{\inf_{v\in V_{kj}} D_n(v) \neq D_n(v_{kj})\right\}\right) &\leq P(A_2(R) \cap A_1^c(R)) + P(A_2(R) \cap A_1(R)) \\ &\leq P(A_2(R)|A_1^c(R)) + P(A_1(R)) \leq C|R|\left(\frac{\eta}{\varphi} + \varphi\right) \leq C|R|\varphi. \end{aligned}$$

Therefore,

$$P\left(\inf_{v\in R} D_n^+(v) \neq \inf_{\{v_{kj}\}} D_n^+(v_{kj})\right) \leq C|R|\varphi.$$

■

Additional References

Andrews, D.W.K., 1994, Empirical Process Methods in Econometrics, *Handbook of Econometrics*, Vol. 4, R.F. Eagle and D.L. McFadden, eds., New York: Elsevier Science B.V., Ch. 37, 2247- 2294.

Billingsley, P., 1968, *Convergence of Probability Measures*, New York: Wiley.

Meyer, R.M., 1973, A Poisson-Type Limit Theorem for Mixing Sequences of Dependent Rare Events, *The Annals of Probability*, 1, 480-483.

Newey, W.K. and D.L. McFadden, 1994, Large Sample Estimation and Hypothesis Testing, *Handbook of Econometrics*, Vol. 4, R.F. Eagle and D.L. McFadden, eds., Amsterdam: Elsevier Science B.V., Ch. 36, 2113-2245.

Pakes, A., and D. Pollard, 1989, Simulation and the Asymptotics of Optimization Estimators, *Econometrica*, 57, 1027-1057.

Van der Vaart, A.W. and J. A. Wellner, 1996, *Weak Convergence and Empirical Processes: With Applications to Statistics*, New York: Springer.