Online Supplement to: "Panel Threshold Regression with Unobserved Individual-Specific Threshold Effects

Ping Yu[a], Shengjie Hong[b] & Peter C. B. Phillips[c,d,e]

[a]University of Hong Kong, [b]Renmin University,
[c]Yale University, [d]University of Auckland, [e]Singapore Management University

Appendix A: Analysis for DPTR

This appendix considers a similar and possibly more severe problem in applying within-regime first-differencing in DPTR than that in Section 2. We then discuss how to build CRE models for DPTR.

Within-Regime First-Differencing in DPTR

In (4), the summation $\sum_{t=1}^{T}\left[\widetilde{y}_{it}^{-}(\gamma)-\widetilde{\mathbf{x}}_{it}^{-}(\gamma)'\beta_1\right]^2 1(q_{it}\leq\gamma)$ should be

$$\sum_{t=1}^{T}\left[\widetilde{y}_{it}^{-}(\gamma)-\widetilde{\mathbf{x}}_{it}^{-}(\gamma)'\beta_1\right]^2 1(q_{it}\leq\gamma)\,1\left(D_i^{-}(\gamma)\geq 1\right),$$

and $\sum_{t=1}^{T}\left[\widetilde{y}_{it}^{+}(\gamma)-\widetilde{\mathbf{x}}_{it}^{+}(\gamma)'\beta_2\right]^2 1(q_{it}>\gamma)$ should be

$$\sum_{t=1}^{T}\left[\widetilde{y}_{it}^{+}(\gamma)-\widetilde{\mathbf{x}}_{it}^{+}(\gamma)'\beta_2\right]^2 1(q_{it}>\gamma)\,1\left(D_i^{+}(\gamma)\geq 1\right)$$

since when $D_i^{\pm}(\gamma)=0$, $\widetilde{y}_{it}^{\pm}(\gamma)$ is not well defined. Our calculation in Appendix D takes this into account explicitly.[18] Our construction of $S_n(\gamma,\beta)$ seems harmless since at least all data points are used in $S_n(\gamma,\beta)$; if $D_i^{\pm}(\gamma)=0$, all observations of individual i fall in one regime. However, in DPTR, different data points are used for different γ's, which implies that first differencing in the usual dynamic panel model is not even applicable.

Specifically, to cancel $\alpha_{\ell i}$, when $q_{it}\leq\gamma$, we need to subtract y_{it} by $y_{i\tau}$, where τ is the largest time index such that $\tau<t$ and $q_{i\tau}\leq\gamma$, and when $q_{it}>\gamma$, we need to subtract y_{it} by $y_{i\tau'}$ where τ' is the largest time index such that $\tau<t$ and $q_{i\tau}>\gamma$. So the response variable is

$$\Delta y_{it}(\gamma):=(y_{it}-y_{i\tau})\,1(q_{it}\leq\gamma)+(y_{it}-y_{i\tau'})\,1(q_{it}>\gamma)\,,t=2,\cdots,T.$$

Only when $\gamma=\gamma_0$,

$$\mathbb{E}\left[\Delta y_{it}(\gamma)\,|\mathbf{X}_i\right]=(\mathbf{x}_{it}-\mathbf{x}_{i\tau})'\beta_1 1(q_{it}\leq\gamma)+(\mathbf{x}_{it}-\mathbf{x}_{i\tau'})'\beta_2 1(q_{it}>\gamma)\,,$$

[18]In $S(\gamma)$, $p^{\pm}(\gamma)=P\left(D_i^{\pm}(\gamma)\geq 1\right)$.

so we might expect that the objective function

$$S_n(\beta,\gamma)=\sum_{i=1}^{N}\sum_{t=2}^{T}\left\{\left[y_{it}-y_{i\tau}-(\mathbf{x}_{it}-\mathbf{x}_{i\tau})'\beta_1\right]^2 1(q_{it}\le\gamma)+\left[y_{it}-y_{i\tau'}-(\mathbf{x}_{it}-\mathbf{x}_{i\tau'})'\beta_2\right]^2 1(q_{it}>\gamma)\right\}$$

would generate a consistent estimator of γ. However, given a γ value, for the i's such that $D_i^-(\gamma)=1$ or $D_i^+(\gamma)=1$, $\Delta y_{it}(\gamma)$ is not defined for some $t=2,\cdots,T$. In other words, $\sum_{t=2}^{T}\left[y_{it}-y_{i\tau}-(\mathbf{x}_{it}-\mathbf{x}_{i\tau})'\beta_1\right]^2 1(q_{it}\le\gamma)$ in $S_n(\beta,\gamma)$ should be

$$\sum_{t=2}^{T}\left[y_{it}-y_{i\tau}-(\mathbf{x}_{it}-\mathbf{x}_{i\tau})'\beta_1\right]^2 1(q_{it}\le\gamma)\,1\left(D_i^-(\gamma)\ge 2\right),$$

and $\sum_{t=2}^{T}\left[y_{it}-y_{i\tau'}-(\mathbf{x}_{it}-\mathbf{x}_{i\tau'})'\beta_2\right]^2 1(q_{it}>\gamma)$ should be

$$\sum_{t=2}^{T}\left[y_{it}-y_{i\tau'}-(\mathbf{x}_{it}-\mathbf{x}_{i\tau'})'\beta_2\right]^2 1(q_{it}>\gamma)\,1\left(D_i^+(\gamma)\ge 2\right),$$

where note that τ and τ' depend on γ (and t). The case with $D_i^{\pm}(\gamma)=0$ can be handled similarly as in SPTR. However, for any γ value, $P\left(D_i^{\pm}(\gamma)=1\right)>0$ if T is fixed. More importantly, for different γ values, the sets of i's such that $D_i^{\pm}(\gamma)=1$ are different; in other words, for different γ's, different observations are used in $S_n(\beta,\gamma)$. As a result, $S_n(\beta,\gamma)$ is not well defined, and the usual first-differencing method cannot be applied in DPTR. For the same reason, Seo and Shin (2016)'s FD-GMM method or Ramírez-Rondán (2016)'s ML method cannot be applied here either.

CRE Models for DPTR

In DPTR, assume

$$\alpha_{\ell i}=\mathbf{z}_i'\psi_\ell+\pi_\ell y_{i0}+a_{\ell i}\text{ with }\mathbb{E}\left[a_{\ell i}|\mathbf{X}_i^T\right]=0,\text{ and }\mathbb{E}\left[u_{\ell it}|\mathbf{X}_i^t\right]=0,$$

where $\mathbf{z}_i'=(\overline{\mathbf{x}}_i',\underline{\mathbf{z}}_i')$ with $\overline{\mathbf{x}}_i=\frac{1}{T+1}\sum_{t=0}^{T}\underline{\mathbf{x}}_{it}$ controls the time-invariant effect, y_{i0} controls the initial condition effect, and a typical case of $a_{\ell i}$ is that it is i.i.d. with mean zero. Such a specification of $\alpha_{\ell i}$ dates back at least to the dynamic, nonlinear panel data models in Wooldridge (2000, 2005). Now,

$$\begin{aligned}\mathbb{E}\left[y_{it}|\mathbf{X}_i^t\right]&=\textstyle\sum_{\ell=1}^{m}\left(\mathbf{x}_{it}'\beta_\ell+\mathbf{z}_i'\psi_\ell+\pi_\ell y_{i0}\right)1\left(\gamma_{\ell-1}<q_{it}\le\gamma_\ell\right),\\&=:\textstyle\sum_{\ell=1}^{m}\check{\mathbf{x}}_{it}'\theta_\ell 1\left(\gamma_{\ell-1}<q_{it}\le\gamma_\ell\right),t=1,\cdots,T,\end{aligned}\tag{25}$$

and the error term takes the same form as (10), where $\check{\mathbf{x}}_{it}:=(\mathbf{x}_{it}',\mathbf{z}_i',y_{i0})'$, and $\theta_\ell=(\beta_\ell',\psi_\ell',\pi_\ell)'$. When $t=1$, the y_{i0} term in $\mathbf{x}_{it}'\beta_1$ and the initial condition effect $\pi_\ell y_{i0}$ can be collected but we do not need to do so; there is the multicollinearity problem when $T=1$, but not when $T>1$.

All the results in Sections 3, 4 and 5 can be applied here, with $\check{\mathbf{x}}_{it}$ and θ_ℓ redefined as above and $(\mathbf{z}_i',y_{i0})'$ and $(\psi_\ell',\pi_\ell)'$ playing the roles of $\mathbf{z}_i$ and ψ_ℓ respectively, so we omit the details here. We provide more rationalization to support our objective function (13) in correctly specified models with $u_{\ell it}=\sigma_\ell u_{it}$. If $u_{it}|\mathbf{X}_i,y_{i,t-1},y_{i,t-2},\cdots,y_{i0}\sim N(0,1)$, $a_{\ell i}|\mathbf{X}_i,y_{i0}\sim N\left(0,\sigma_{a_\ell}^2\right)$ and u_{it}, a_{1i} and a_{2i} are independent of each other, then the likelihood function of $(y_{iT},\cdots,y_{i1})_{i=1}^{N}$ given $\mathbf{X}_i, y_{i0}$ is

$$
\begin{aligned}
L_n(\vartheta) &= \prod_{i=1}^{N} f\left(y_{iT},\cdots,y_{i1}|\mathbf{X}_i,y_{i0}\right)\\
&= \prod_{i=1}^{N}\int\int\prod_{t=1}^{T} f\left(y_{it}|\mathbf{X}_i,y_{i,t-1},y_{i,t-2},\cdots,y_{i0},a_{1i},a_{2i}\right)dF_{a_{1i}|\mathbf{X}_i,y_{i0}}\left(a_{1i}\right)dF_{a_{2i}|\mathbf{X}_i,y_{i0}}\left(a_{1i}\right)\\
&= \prod_{i=1}^{N}\int\int\prod_{t=1}^{T}\frac{1}{\sqrt{2\pi\left(\sigma_1^2 1(q_{it}\le\gamma)+\sigma_2^2 1(q_{it}>\gamma)\right)}}\\
&\quad\cdot\exp\left\{-\frac{\left(y_{it}-\left(\mathbf{x}_{it}'\beta_1+\psi_1'\mathbf{z}_i+\pi_1 y_{i0}+\sigma_{a_1}a_{1i}^*\right)1(q_{it}\le\gamma)-\left(\mathbf{x}_{it}'\beta_2+\psi_2'\mathbf{z}_i+\pi_2 y_{i0}+\sigma_{a_2}a_{2i}^*\right)1(q_{it}>\gamma)\right)^2}{2\left(\sigma_1^2 1(q_{it}\le\gamma)+\sigma_2^2 1(q_{it}>\gamma)\right)}\right\}\,d\Phi\left(a_{1i}^*\right)d\Phi\left(a_{2i}^*\right),
\end{aligned}
$$

where $a_{\ell i}=\sigma_{a_\ell}a_{\ell i}^*$, $\vartheta=(\theta',\sigma',\sigma_a')'$ with $\sigma'=(\sigma_1,\sigma_2)$ and $\sigma_a'=(\sigma_{a_1},\sigma_{a_2})$, and $\Phi(\cdot)$ is the cdf of $N(0,1)$. If $\sigma_1=\sigma_2=\sigma$ and $a_{1i}=a_{2i}=a_i=\sigma_a a_i^*$ with $a_i^*\sim N(0,1)$, then the likelihood function reduces to

$$
\begin{aligned}
L_n(\vartheta) &= \prod_{i=1}^{N}\int\prod_{t=1}^{T}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{\left(y_{it}-\sigma_a a_i^*-\left(\mathbf{x}_{it}'\beta_1+\psi_1'\mathbf{z}_i+\pi_1 y_{i0}\right)1(q_{it}\le\gamma)-\left(\mathbf{x}_{it}'\beta_2+\psi_2'\mathbf{z}_i+\pi_2 y_{i0}\right)1(q_{it}>\gamma)\right)^2}{2\sigma^2}\right\}d\Phi\left(a_i^*\right)\\
&= \prod_{i=1}^{N}\frac{1}{\sqrt{(2\pi)^T|\Sigma|}}\exp\left(-\frac{1}{2}\mathbf{e}_i'\Sigma^{-1}\mathbf{e}_i\right),
\end{aligned}
$$

where $\mathbf{e}_i=\mathbf{y}_i-\mathbf{1}_{i\le\gamma}\odot\left(X_i\beta_1+\mathbf{1}_T\left(\psi_1'\mathbf{z}_i+\pi_1 y_{i0}\right)\right)-\mathbf{1}_{i>\gamma}\odot\left(X_i\beta_2+\mathbf{1}_T\left(\psi_2'\mathbf{z}_i+\pi_2 y_{i0}\right)\right)$ with $\odot$ being the element-by-element product,

$$
\begin{aligned}
\Sigma_{T\times T} &= \begin{pmatrix}\sigma^2+\sigma_a^2 & \sigma^2 & \cdots & \sigma^2\\ \sigma^2 & \sigma^2+\sigma_a^2 & \cdots & \sigma^2\\ \vdots & \vdots & \ddots & \vdots\\ \sigma^2 & \sigma^2 & \cdots & \sigma^2+\sigma_a^2\end{pmatrix},\mathbf{y}_i=\begin{pmatrix}y_{i1}\\ y_{i2}\\ \vdots\\ y_{iT}\end{pmatrix},X_i=\begin{pmatrix}\mathbf{x}_{i1}\\ \mathbf{x}_{i2}\\ \vdots\\ \mathbf{x}_{iT}\end{pmatrix},\\
\mathbf{1}_{i\le\gamma} &= \begin{pmatrix}1(q_{i1}\le\gamma)\\ 1(q_{i2}\le\gamma)\\ \vdots\\ 1(q_{iT}\le\gamma)\end{pmatrix},\mathbf{1}_{i>\gamma}=\begin{pmatrix}1(q_{i1}>\gamma)\\ 1(q_{i2}>\gamma)\\ \vdots\\ 1(q_{iT}>\gamma)\end{pmatrix},\mathbf{1}_T=\begin{pmatrix}1\\ 1\\ \vdots\\ 1\end{pmatrix}_{T\times 1};
\end{aligned}
$$

and if a_i^* degenerates to a point mass at zero or $\sigma_a=0$, then the likelihood function further reduces to

$$
L_n(\vartheta)=\prod_{i=1}^{N}\prod_{t=1}^{T}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{\left(y_{it}-\left(\mathbf{x}_{it}'\beta_1+\psi_1'\mathbf{z}_i+\pi_1 y_{i0}\right)1(q_{it}\le\gamma)-\left(\mathbf{x}_{it}'\beta_2+\psi_2'\mathbf{z}_i+\pi_2 y_{i0}\right)1(q_{it}>\gamma)\right)^2}{2\sigma^2}\right\},
$$

which is equivalent to $S_n(\theta)$ in the estimation of θ. In other words, the extra efficiency introduced by the general likelihood function beyond $S_n(\theta)$ lies in $\sigma_1\neq\sigma_2$, $a_{1i}\neq a_{2i}$ and the nondegeneracy of $a_{\ell i}$. Our objective function loses some information but is more robust because it does not rely on any distributional assumptions concerning $a_{\ell i}$.

Appendix B: Proofs

Define $\lambda_N=N^{1-2\kappa}$. Let $\rightsquigarrow$ signify weak convergence over a compact metric space and $\overset{d}{=}$ mean equality in distribution.

Proof of Theorem 1. First, under $H_{1c}^{(1)}$,

$$
\begin{aligned}
&N^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}\breve{\mathbf{x}}_{it}\widehat{e}_{it}^{o}1\left(q_{it}\le\gamma\right)=N^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}\breve{\mathbf{x}}_{it}1\left(q_{it}\le\gamma\right)\left(y_{it}-\breve{\mathbf{x}}_{it}'\widehat{\theta}_{o}\right)\\
&=N^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}\breve{\mathbf{x}}_{it}1\left(q_{it}\le\gamma\right)\left(y_{it}-\breve{\mathbf{x}}_{it}'\theta_{o}\right)-N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{it}'1\left(q_{it}\le\gamma\right)\sqrt{N}\left(\widehat{\theta}_{o}-\theta_{o}\right)\\
&=N^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}\breve{\mathbf{x}}_{it}1\left(q_{it}\le\gamma\right)\left(y_{it}-\breve{\mathbf{x}}_{it}'\theta_{o}\right)-\widehat{M}\left(\gamma\right)\widehat{M}^{-1}\left[N^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}\breve{\mathbf{x}}_{it}\left(y_{it}-\breve{\mathbf{x}}_{it}'\theta_{o}\right)\right]\\
&=N^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\breve{\mathbf{x}}_{it}1\left(q_{it}\le\gamma\right)-\widehat{M}\left(\gamma\right)\widehat{M}^{-1}\breve{\mathbf{x}}_{it}\right]\left[\breve{\mathbf{x}}_{it}'\delta_{1}1\left(q_{it}\le\gamma_{0}\right)+e_{it}^{0}\right]\\
&=N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\breve{\mathbf{x}}_{it}1\left(q_{it}\le\gamma\right)-\widehat{M}\left(\gamma\right)\widehat{M}^{-1}\breve{\mathbf{x}}_{it}\right]\breve{\mathbf{x}}_{it}'1\left(q_{it}\le\gamma_{0}\right)\left(N^{1/2}\delta_{1}\right)\\
&\quad+N^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\breve{\mathbf{x}}_{it}1\left(q_{it}\le\gamma\right)-\widehat{M}\left(\gamma\right)\widehat{M}^{-1}\breve{\mathbf{x}}_{it}\right]e_{it}^{0}\\
&\rightsquigarrow\left[M\left(\gamma\wedge\gamma_{0}\right)-M\left(\gamma\right)M^{-1}M\left(\gamma_{0}\right)\right]c+\Xi\left(\gamma\right),
\end{aligned}
$$

where the last step needs further clarification. Under Assumption 2(ii), $\widehat{M}\xrightarrow{p}M$ by the WLLN because $\mathbb{E}\left[\left\|\sum_{t=1}^{T}\breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{it}'\right\|\right]<\sum_{t=1}^{T}\mathbb{E}\left[\left\|\breve{\mathbf{x}}_{it}\right\|^{2}\right]<\infty$. Under Assumption 2, $\widehat{M}\left(\gamma\right)\xrightarrow{p}M\left(\gamma\right)$ uniformly in $\gamma\in\Gamma$ by applying a Glivenko-Cantelli theorem because $\left\{\sum_{t=1}^{T}\breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{it}'1\left(q_{it}\le\gamma\right)|\gamma\in\Gamma\right\}$ is a VC class of function with the envelope function $\left\|\sum_{t=1}^{T}\breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{it}'\right\|$ whose L_1 norm is bounded. Similarly,

$$
N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\breve{\mathbf{x}}_{it}1\left(q_{it}\le\gamma\right)\breve{\mathbf{x}}_{it}'1\left(q_{it}\le\gamma_{0}\right)\xrightarrow{p}M\left(\gamma\wedge\gamma_{0}\right),
$$

uniformly in $\gamma\in\Gamma$. To show the weak convergence of $N^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\breve{\mathbf{x}}_{it}1\left(q_{it}\le\gamma\right)-\widehat{M}\left(\gamma\right)\widehat{M}^{-1}\breve{\mathbf{x}}_{it}\right]e_{it}^{0}$, note that

$$
N^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\breve{\mathbf{x}}_{it}1\left(q_{it}\le\gamma\right)-\widehat{M}\left(\gamma\right)\widehat{M}^{-1}\breve{\mathbf{x}}_{it}\right]e_{it}^{0}=\left(\begin{array}{cc} I & -\widehat{M}\left(\gamma\right)\widehat{M}^{-1}\end{array}\right)\left(\begin{array}{c} N^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}\breve{\mathbf{x}}_{it}e_{it}^{0}1\left(q_{it}\le\gamma\right)\\ N^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}\breve{\mathbf{x}}_{it}e_{it}^{0}\end{array}\right),
$$

so by Slutsky's theorem we need only show the weak convergence of $(N^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}\breve{\mathbf{x}}_{it}e_{it}^{0}1\left(q_{it}\le\gamma\right)$, $N^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}\breve{\mathbf{x}}_{it}e_{it}^{0})$. Because $\left\{\left(\sum_{t=1}^{T}\breve{\mathbf{x}}_{it}e_{it}^{0}1\left(q_{it}\le\gamma\right),\sum_{t=1}^{T}\breve{\mathbf{x}}_{it}e_{it}^{0}\right)|\gamma\in\Gamma\right\}$ is a VC class function with the envelope function $\left\|\sum_{t=1}^{T}\breve{\mathbf{x}}_{it}e_{it}^{0}\right\|$ which satisfies

$$
\mathbb{E}\left[\left\|\sum_{t=1}^{T}\breve{\mathbf{x}}_{it}e_{it}^{0}\right\|^{2}\right]\le2\sum_{t=1}^{T}\mathbb{E}\left[\left\|\breve{\mathbf{x}}_{it}e_{it}^{0}\right\|^{2}\right]\le2\sum_{t=1}^{T}\mathbb{E}\left[\left\|\breve{\mathbf{x}}_{it}\right\|^{4}\right]\mathbb{E}\left[\left|e_{it}^{0}\right|^{4}\right]<\infty
$$

by Assumption 2(ii), where the first equality is from the Bahr-Esseen inequality and the second one from the Cauchy-Schwarz inequality, the weak convergence follows by a Donsker theorem. In summary,

$$
N^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\breve{\mathbf{x}}_{it}1\left(q_{it}\le\gamma\right)-\widehat{M}\left(\gamma\right)\widehat{M}^{-1}\breve{\mathbf{x}}_{it}\right]e_{it}^{0}\rightsquigarrow\Xi\left(\gamma\right)
$$

with the covariance kernel being

$$
\mathbb{E}\left[\left(\sum_{t=1}^{T}(1\left(q_{it}\le\gamma_{1}\right)\breve{\mathbf{x}}_{it}-M\left(\gamma_{1}\right)M^{-1}\breve{\mathbf{x}}_{it})e_{it}^{0}\right)\left((1\left(q_{it}\le\gamma_{2}\right)\breve{\mathbf{x}}_{it}-M\left(\gamma_{2}\right)M^{-1}\breve{\mathbf{x}}_{it})e_{it}^{0}\right)'\right].
$$

Next, it is standard to show that under $H_{1c}^{(1)}$, $\widehat{\theta}_o$ is consistent to θ_o and, applying a Glivenko-Cantelli theorem, we have

$$
\widehat{H}_{n}(\gamma_{1},\gamma_{2})\quad=\quad N^{-1}\sum_{i=1}^{N}\left[\sum_{t=1}^{T}\left(\breve{\mathbf{x}}_{it}1(q_{it}\le\gamma_{1})-\widehat{M}\left(\gamma_{1}\right)\widehat{M}^{-1}\breve{\mathbf{x}}_{it}\right)\widehat{e}_{it}^{o}\right]\left[\sum_{t=1}^{T}\left(\breve{\mathbf{x}}_{it}1(q_{it}\le\gamma_{2})-\widehat{M}\left(\gamma_{2}\right)\widehat{M}^{-1}\breve{\mathbf{x}}_{it}\right)\widehat{e}_{it}^{o}\right]'
$$

$$\xrightarrow{p} H(\gamma_1, \gamma_2),$$

uniformly over $(\gamma_1, \gamma_2) \in \Gamma \times \Gamma$, which implies $\widehat{H}_n(\gamma, \gamma) \xrightarrow{p} H(\gamma, \gamma)$ uniformly over $\gamma \in \Gamma$ under $H_{1c}^{(1)}$, so the results of the theorem follow. ■

Proof of Corollary 1. Because $\Xi(\gamma)$ is tight, the upper α'th quantile of $g\left(T^0\right)$, $F_0^{(1)-1}(1-\alpha)$, is finite. Because $H(\gamma_0) > 0$ by Assumption 2(v), $M(\gamma_0) - M(\gamma_0) M^{-1} M(\gamma_0) > 0$ by Assumption 2(iv), we have $H(\gamma)^{-1/2}\left[M(\gamma \wedge \gamma_0) - M(\gamma) M^{-1} M(\gamma_0)\right]$ is nonsingular for $\gamma = \gamma_0$. Because $H(\gamma)^{-1/2}\left[M(\gamma \wedge \gamma_0) - M(\gamma) M^{-1} M(\gamma_0)\right]$ is continuous in γ by Assumption 2(iii), there is a neighborhood of γ_0 such that $H(\gamma)^{-1/2}\left[M(\gamma \wedge \gamma_0) - M(\gamma) M^{-1} M(\gamma_0)\right]$ is nonsingular for γ in that neighborhood. As a result, $g_n^{(1)} = O_p\left(N^{1-2\kappa}\right)$ if $\delta_1 = c_1 N^{-\kappa}$ from the proof of Theorem 1, so that

$$P\left(g_n^{(1)} > F_0^{(1)-1}(1-\alpha)\right) \to 1.$$

In other words, our test is consistent against the alternative $\delta_1 = c_1 N^{-\kappa}$ for any $c \neq \mathbf{0}$ and $0 < \kappa < 1/2$. ■

Proof of Theorem 2. Conditional on the original data, $\widehat{m}_n^*(\gamma)$ is a zero-mean Gaussian process with covariance kernel $\widehat{H}_n(\gamma_1, \gamma_2)$. From the proof of Theorem 1, $\widehat{H}_n(\gamma_1, \gamma_2) \xrightarrow{p} H(\gamma_1, \gamma_2)$ uniformly over $(\gamma_1, \gamma_2) \in \Gamma \times \Gamma$ under $H_{1c}^{(1)}$. Also, $\widehat{H}_n(\gamma) \xrightarrow{p} H(\gamma)$ uniformly over $\gamma \in \Gamma$ under $H_{1c}^{(1)}$. As a result, $T_n^*(\gamma) \overset{*}{\rightsquigarrow} H(\gamma)^{-1/2}\Xi(\gamma) = T^0(\gamma)$ in $\ell^\infty(\Gamma)$ such that $\sup_{t\in\mathbb{R}}\left|F_n^{(1)*}(t) - F_0^{(1)}(t)\right| \xrightarrow{p} 0$ by Pólya's Theorem, where $\overset{*}{\rightsquigarrow}$ signifies weak convergence in probability, see Yu (2014, Section 2.2) for the precise definition. It follows that $\left|F_n^{(1)*}\left(g_n^{(1)}\right) - F_0^{(1)}\left(g_n^{(1)}\right)\right| \xrightarrow{p} 0$ such that $p_n^{(1)*} = 1 - F_n^{(1)*}\left(g_n^{(1)}\right) \xrightarrow{d} 1 - F_0^{(1)}\left(g_c^{(1)}\right)$.

If $\delta_1 = c_1 N^{-\kappa}$ with $c_1 \neq \mathbf{0}$ and $0 < \kappa < 1/2$, then $\widehat{\theta}_o$ is still consistent to θ_o. As a result, it still holds that $T_n^*(\gamma) \overset{*}{\rightsquigarrow} T^0(\gamma)$ in $\ell^\infty(\Gamma)$. From the proof of Corollary 1, $g_n^{(1)} = O_p\left(N^{1-2\kappa}\right)$, so $\left|F_n^{(1)*}\left(g_n^{(1)}\right) - F_0^{(1)}\left(g_n^{(1)}\right)\right| \xrightarrow{p} 0$ implies $F_n^{(1)*}\left(g_n^{(1)}\right) \xrightarrow{p} 1$ because $F_0^{(1)}\left(g_n^{(1)}\right) \xrightarrow{p} 1$. ■

Proof of Lemma 1. The proof is similar to that of Proposition 2.2 of GP with proper notational adaptation, for example, that $\arg\min_{\gamma\in\{\gamma_1^0,\cdots,\gamma_m^0\}} S(\gamma)$ is unique and $\arg\min_{r\in\{\gamma_1^0,\cdots,\gamma_m^0\}} J_\infty(r)$ is unique are equivalent, $G(\cdot) = M(\cdot)$, $Z_1' Z_1 = \sum_{i=1}^N \sum_{t=1}^T \check{\mathbf{x}}_{it}\check{\mathbf{x}}_{it}' 1(q_{it} \leq \gamma)$, $\rho_\ell = \theta_\ell - \theta_{\ell-1}$, and the normalization rate for $S_n(\theta)$ (after subtracting sum of squared errors) is $N^{1-2\kappa}$ instead of N, etc. Note that it is key to assume $\mathbb{E}\left[\alpha_{\ell i}|\mathbf{X}_i\right] = \mathbf{z}_i'\psi_\ell$, taking a linear form; otherwise, the least squares regression is further misspecified and need not provide coverage one of the threshold points, as shown in Yu (2015b). ■

Proof of Corollary 2. The proof is a straightforward extension of those of Theorem 1 and Corollary 1. Because $\widehat{\gamma} \xrightarrow{p} \gamma_1^0$, it can be replaced by γ_1^0 throughout the proof without any asymptotic effect. Note that although $g_n^{(2,1)}$ and $g_n^{(2,2)}$ are applied to two nonoverlapping subsamples, they are not asymptotically independent. This is because

$$\begin{aligned} & \mathbb{E}\left[\zeta_i^{(2,1)}(\gamma_1)\,\zeta_i^{(2,2)}(\gamma_2)\right] \\ = & \sum_{t=1}^T \sum_{\tau\neq t} \mathbb{E}\left[\left[\check{\mathbf{x}}_{it} 1(q_{it} \leq \gamma_1) - M(\gamma_1) M_1^{-1} \check{\mathbf{x}}_{it} 1\left(q_{it} \leq \gamma_1^0\right)\right]\left[\check{\mathbf{x}}_{i\tau} 1\left(\gamma_1^0 < q_{i\tau} \leq \gamma_2\right) - M\left(\gamma_1^0, \gamma_2\right) M_2^{-1} \check{\mathbf{x}}_{i\tau} 1(q_{i\tau} > \gamma_1^0)\right] e_{it}^0 e_{i\tau}^0\right] \end{aligned}$$

need not be zero because $\left(\check{\mathbf{x}}_{it}', e_{it}^0\right)'$ and $\left(\check{\mathbf{x}}_{i\tau}', e_{i\tau}^0\right)'$ are not independent for $\tau \neq t$, where $\gamma_1 \in \Gamma_1$ and $\gamma_2 \in \Gamma_2$. Note also that $H_1^{(2)}$ assumes a threshold effect in the second regime, so only $T_c^{(2,2)}(\cdot)$ contains the drifting term $\left[M\left(\gamma_1^0, \gamma \wedge \gamma_0\right) - M\left(\gamma_1^0, \gamma\right) M_2^{-1} M\left(\gamma_1^0, \gamma_2^0\right)\right] c$. Under Assumption 1 with $m = 2$ and Assumption 2(iv), $g_n^{(2,1)} = O_p(1)$ while $g_n^{(2,2)} = O_p\left(N^{1-2\kappa}\right)$, so $g_n^{(2)} = \max\left\{g_n^{(2,1)}, g_n^{(2,2)}\right\} = O_p\left(N^{1-2\kappa}\right)$. As a result, $p_n^{(2)} = 1 - F_0^{(2)}\left(g_n^{(2)}\right) \xrightarrow{p} 0$. ■

Proof of Corollary 3. The proof is a straightforward extension of that of Theorem 2. The only new task is to show the covariance between $\widehat{m}_n^{(2,1)}(\gamma_1)$ and $\widehat{m}_n^{(2,2)}(\gamma_2)$ is maintained. Note that conditional on the original data,

$$Cov\left(\widehat{m}_n^{(2,1)*}(\gamma_1), \widehat{m}_n^{(2,2)*}(\gamma_2)\right) = N^{-1}\sum_{i=1}^{N}\widehat{\zeta}^{(2,1)}(\gamma_1)\,\widehat{\zeta}^{(2,2)}(\gamma_2)' \xrightarrow{p} \mathbb{E}\left[\zeta^{(2,1)}(\gamma_1)\,\zeta^{(2,2)}(\gamma_2)'\right],$$

as required. ■

Proof of Lemma 2. This proof is similar to that of Proposition 2.4 of GP after adapting notation. We only provide two comments here. First, to show the probability limit of $S_n\left(\gamma|\widehat{\gamma}^{(1)}\right)/N^{1-2\kappa}$ is the same as that of $S_n\left(\gamma|\gamma_0^{(1)}\right)/N^{1-2\kappa}$, we require only the consistency of $\widehat{\gamma}^{(1)}$ instead of the N-consistency of $\widehat{\gamma}^{(1)}$. Actually, we show in the future that $\widehat{\gamma}^{(1)}$ is only $N^{1-2\kappa}$ consistent. Second, the parameter spaces Γ_1 and Γ_2 should be the truncated version instead of the original version of $\left[\underline{\gamma}, \widehat{\gamma}^{(1)}\right]$ and $\left[\widehat{\gamma}^{(1)}, \overline{\gamma}\right]$, as mentioned in the main text. ■

Proof of Theorem 3. We first show that $P(\widehat{m} < m_0) \to 0$. From the proof of Corollary 1, we need only show that $g_n^{(l)} = O_p\left(N^{1-2\kappa}\right)$ for $l = 1, \cdots, m_0 - 1$. We take $l = 1$ to illustrate. From the Proof of Theorem 1, under ST,

$$\begin{aligned}
& N^{-1/2}\sum\nolimits_{i=1}^{N}\sum\nolimits_{t=1}^{T}\check{\mathbf{x}}_{it}\widehat{e}_{it}^{o}1(q_{it}\le\gamma) \\
= \ & N^{-1/2}\sum\nolimits_{i=1}^{N}\sum\nolimits_{t=1}^{T}\left[\check{\mathbf{x}}_{it}1(q_{it}\le\gamma) - \widehat{M}(\gamma)\,\widehat{M}^{-1}\check{\mathbf{x}}_{it}\right]\left[\sum\nolimits_{\ell=1}^{m}\check{\mathbf{x}}_{it}'\delta_\ell 1(q_{it}\le\gamma_\ell^0) + e_{it}^0\right] \\
= \ & N^{1/2-\kappa}O_p\left(\sum\nolimits_{\ell=1}^{m}\left[M\left(\gamma\wedge\gamma_\ell^0\right) - M(\gamma)\,M^{-1}M\left(\gamma_\ell^0\right)\right]c_\ell\right) + O_p(1).
\end{aligned}$$

So $g_n^{(1)} = O_p\left(N^{1-2\kappa}\right)$ only if

$$\sum\nolimits_{\ell=1}^{m}\left[M\left(\gamma\wedge\gamma_\ell^0\right) - M(\gamma)\,M^{-1}M\left(\gamma_\ell^0\right)\right]c_\ell \neq \mathbf{0} \text{ for some } \gamma\in\Gamma.$$

We know when $\gamma = \gamma_\ell^0$, $M\left(\gamma\wedge\gamma_\ell^0\right) - M(\gamma)\,M^{-1}M\left(\gamma_\ell^0\right)$ is nonsingular, so as long as there is some $c_\ell \neq \mathbf{0}$, $\sum_{\ell=1}^{m}\left[M\left(\gamma\wedge\gamma_\ell^0\right) - M(\gamma)\,M^{-1}M\left(\gamma_\ell^0\right)\right]c_\ell \neq \mathbf{0}$. We next show $P(\widehat{m} > m_0) \to 0$. Note that

$$P(\widehat{m} > m_0) \le \sum_{\ell=1}^{m_0+1} P\left(g_n^{(m_0+1,\ell)} > c_N^{(m_0+1)}\right).$$

But we know $g_n^{(m_0+1,\ell)} = O_p(1)$ if the true number of thresholds is m_0 and $c_N^{(m_0+1)} \to \infty$, so $P(\widehat{m} > m_0)$ indeed converges to zero. ■

Proof of Theorem 4. First, $\widehat{\theta} = \arg\min_\theta S_n(\theta)$ implies

$$\begin{aligned}
\widehat{h}_n \ &: \ = \left(\lambda_N(\widehat{\gamma}-\gamma_0), N^{1/2}\left(\widehat{\underline{\theta}}-\underline{\theta}_0\right)\right) \\
&= \ \arg\min_{(v,u)}\left\{S_n\left(\gamma_0+\frac{v}{\lambda_N}, \underline{\theta}_0+\frac{u}{N^{1/2}}\right) - S_n(\gamma_0,\underline{\theta}_0)\right\} \\
&= \ \arg\min_{h}\left\{\mathbb{M}_n(h) + o_p(1)\right\},
\end{aligned}$$

where from Lemma 5,

$$\mathbb{M}_n(h) = \frac{1}{2}u_1'M_1u_1 + \frac{1}{2}u_2'M_2u_2 - W_n(u) + C_n(v),$$

and $(W_n(u), C_n(v))$ is defined there. Now, we apply Theorem 2.7 of Kim and Pollard (1990) (KP hereafter) to derive the asymptotic distribution.

(i) $\mathbb{M}_n(h) \rightsquigarrow \mathbb{M}(h) = \frac{1}{2}u_1'M_1u_1 + \frac{1}{2}u_2'M_2u_2 - W(u) + C(v) \in \mathbf{C}_{\min}\left(\mathbb{R}^{d_\theta}\right)$, where

$$C(v) = \begin{cases} \frac{1}{2}c'D(c+2c_-)|v| + \sqrt{c'V_1c}B_1(|v|), \\ \frac{1}{2}c'D(c-2c_+)v + \sqrt{c'V_2c}B_2(v), \end{cases} =: \begin{cases} \frac{1}{2}\mu_-|v| + \sqrt{\varpi_-}B_1(|v|), & \text{if } \nu \le 0, \\ \frac{1}{2}\mu_+|v| + \sqrt{\varpi_+}B_2(v), & \text{if } \nu > 0, \end{cases} \tag{26}$$

and $W(u) = u'W$ with

$$W = \begin{pmatrix} W_1 \\ W_2 \end{pmatrix} \sim N\left(\mathbf{0}, \begin{pmatrix} \Omega_1 & \Omega_{12} \\ \Omega_{12}' & \Omega_2 \end{pmatrix}\right), \tag{27}$$

$\mathbf{C}_{\min}\left(\mathbb{R}^{d_\theta}\right)$ is defined as the subset of continuous functions $x(\cdot) \in \mathbf{B}_{\text{loc}}\left(\mathbb{R}^{d_\theta}\right)$ for which (i) $x(t) \to \infty$ as $\|t\| \to \infty$ and (ii) $x(t)$ achieves its minimum at a unique point in $\mathbb{R}^{d_\theta}$, and $\mathbf{B}_{\text{loc}}(\mathbb{R}^{d_\theta})$ is the space of all locally bounded real functions on $\mathbb{R}^{d_\theta}$, endowed with the uniform metric on compacta. The weak convergence is proved in Lemma 6. We now check $\mathbb{M}(h) \in \mathbf{C}_{\min}\left(\mathbb{R}^{d_\theta}\right)$. Because u and v are separable in $\mathbb{M}(h)$, we can check $\mathbb{M}_1(u) := \frac{1}{2}u_1'M_1u_1 + \frac{1}{2}u_2'M_2u_2 - u'W \in \mathbf{C}_{\min}\left(\mathbb{R}^{d_{\underline{\theta}}}\right)$ and $\mathbb{M}_2(v) := C(v) \in \mathbf{C}_{\min}(\mathbb{R})$ separately. First, $\mathbb{M}_1(u) \in \mathbf{C}_{\min}\left(\mathbb{R}^{d_{\underline{\theta}}}\right)$ because it is continuous, has a unique explicit minimizer and $\lim_{\|u\|\to\infty}\mathbb{M}_1(u) = \infty$ with probability one given that for each value of W, $\mathbb{M}_1(u)$ is a quadratic function in u. Second, $\mathbb{M}_2(v) \in \mathbf{C}_{\min}(\mathbb{R})$ because it is continuous, has a unique minimum (see Lemma 2.6 of KP), and $\lim_{|v|\to\infty}\mathbb{M}_2(v) = \infty$ almost surely which follows since $\lim_{|v|\to\infty} B_\ell(v)/|v| = 0$ almost surely by virtue of the law of the iterated logarithm for Brownian motion.

(ii) $\lambda_N(\widehat{\gamma} - \gamma_0) = O_p(1)$ and $N^{1/2}\left(\widehat{\underline{\theta}} - \underline{\theta}_0\right) = O_p(1)$. This is shown in Lemma 4.

Then appealing to Theorem 2.7 of KP, we have

$$N^{1/2}\left(\widehat{\underline{\theta}} - \underline{\theta}_0\right) \xrightarrow{d} S_{\underline{\theta\theta}}^{-1}W = \begin{pmatrix} S_{\theta_1\theta_1}^{-1}W_1 \\ S_{\theta_2\theta_2}^{-1}W_2 \end{pmatrix},$$

and

$$\lambda_N(\widehat{\gamma} - \gamma_0) \xrightarrow{d} \arg\max_v \begin{cases} -\frac{1}{2}\mu_-|v| + \sqrt{\varpi_-}B_1(|v|), & \text{if } \nu \le 0, \\ -\frac{1}{2}\mu_+|v| + \sqrt{\varpi_+}B_2(v), & \text{if } \nu > 0, \end{cases} = \omega \cdot \zeta(\varphi, \phi),$$

where $S_{\underline{\theta\theta}}$ =diag$\{M_1, M_2\}$ =:diag$\{S_{\theta_1\theta_1}, S_{\theta_2\theta_2}\}$, and the last equality can be derived from the general results in Proposition 2(i) of Yu (2019). ■

Proof of Theorem 5. By the continuous mapping theorem (CMT),

$$\begin{aligned} S_n(\gamma_0) - S_n(\widehat{\gamma}) &= \left(S_n\left(\gamma_0, \widehat{\underline{\theta}}(\gamma_0)\right) - S_n(\gamma_0, \underline{\theta}_0)\right) - \left(S_n\left(\widehat{\gamma}, \widehat{\underline{\theta}}\right) - S_n(\gamma_0, \underline{\theta}_0)\right) \\ &\xrightarrow{d} \min_u\left\{\frac{1}{2}u'S_{\underline{\theta\theta}}u - u'W\right\} - \min_{u,v}\left\{\frac{1}{2}u'S_{\underline{\theta\theta}}u - u'W + C(v)\right\} \\ &= \max_v \{-C(v)\} \overset{d}{=} \max_v \begin{cases} -\frac{1}{2}\mu_-|v| + \sqrt{\varpi_-}B_1(|v|), & \text{if } v \le 0, \\ -\frac{1}{2}\mu_+|v| + \sqrt{\varpi_+}B_2(v), & \text{if } v > 0, \end{cases} \\ &= \eta^2\xi(\varphi, \phi), \end{aligned}$$

where the last equality can be derived from the general results in Proposition 2(ii) of Yu (2019), $\eta^2 = \varpi_-/\mu_-$, and the distribution of $\xi(\varphi, \phi)$ is derived in Proposition 2(iii) of Yu (2019). The required result follows by

Slutsky's theorem. ∎

Proof of Theorem 6. By the CMT,

$$\begin{aligned}
& S_n\left(\gamma_0,\beta_{11}^0\right)-S_n\left(\widehat{\gamma},\widehat{\beta}_{11}\right) \\
= & \left(S_n\left(\gamma_0,\beta_{11}^0,\widehat{\theta}_{12}\left(\gamma_0,\beta_{11}^0\right),\widehat{\theta}_2\left(\gamma\right)\right)-S_n\left(\gamma_0,\underline{\theta}_0\right)\right)-\left(S_n\left(\widehat{\gamma},\widehat{\underline{\theta}}\right)-S_n\left(\gamma_0,\underline{\theta}_0\right)\right) \\
& \xrightarrow{d} \min_{u_{12},u_2}\left\{\frac{1}{2}u_{12}'S_{\theta_{12}\theta_{12}}u_{12}-u_{12}'W_{12}+\frac{1}{2}u_2'S_{\theta_2\theta_2}u_2-u_2'W_2\right\}-\min_{u,v}\left\{\frac{1}{2}u'S_{\underline{\theta\theta}}u-u'W+C\left(v\right)\right\} \\
= & -\frac{1}{2}W_{12}'S_{\theta_{12}\theta_{12}}^{-1}W_{12}+\frac{1}{2}W_1'S_{\theta_1\theta_1}^{-1}W_1+\max_v\ \left\{-C\left(v\right)\right\} \\
& \overset{d}{=}\frac{1}{2}\frac{W_1'\left(1,-S_{\beta_{11}\theta_{12}}S_{\theta_{12}\theta_{12}}^{-1}\right)'\left(1,-S_{\beta_{11}\theta_{12}}S_{\theta_{12}\theta_{12}}^{-1}\right)W_1}{\widetilde{S}_{\beta_{11}\beta_{11}}}+\eta^2\xi(\varphi,\phi),
\end{aligned}$$

where

$$S_{\theta_1\theta_1}=\begin{pmatrix} S_{\beta_{11}\beta_{11}} & S_{\beta_{11}\theta_{12}} \\ S_{\theta_{12}\beta_{11}} & S_{\theta_{12}\theta_{12}}\end{pmatrix},W_1=\begin{pmatrix} W_{11} \\ W_{12}\end{pmatrix},\widetilde{S}_{\beta_{11}\beta_{11}}=S_{\beta_{11}\beta_{11}}-S_{\beta_{11}\theta_{12}}S_{\theta_{12}\theta_{12}}^{-1}S_{\theta_{12}\beta_{11}},$$

and

$$\widehat{u}_{11}=\frac{\left(1,-S_{\beta_{11}\theta_{12}}S_{\theta_{12}\theta_{12}}^{-1}\right)W_1}{\widetilde{S}_{\beta_{11}\beta_{11}}}.$$

Since $\left(1,-S_{\beta_{11}\theta_{12}}S_{\theta_{12}\theta_{12}}^{-1}\right)W_1\sim N\left(0,\left(1,-S_{\beta_{11}\theta_{12}}S_{\theta_{12}\theta_{12}}^{-1}\right)\Omega_1\left(1,-S_{\beta_{11}\theta_{12}}S_{\theta_{12}\theta_{12}}^{-1}\right)'\right)$, we have

$$W_1\left(1,-S_{\beta_{11}\theta_{12}}S_{\theta_{12}\theta_{12}}^{-1}\right)'\left(1,-S_{\beta_{11}\theta_{12}}S_{\theta_{12}\theta_{12}}^{-1}\right)W_1\sim\left(1,-S_{\beta_{11}\theta_{12}}S_{\theta_{12}\theta_{12}}^{-1}\right)\Omega_1\left(1,-S_{\beta_{11}\theta_{12}}S_{\theta_{12}\theta_{12}}^{-1}\right)'\chi_1^2.$$

As a result,

$$LR_n\left(\gamma_0,\beta_{11}^0\right)\xrightarrow{d}\frac{\left(1,-S_{\beta_{11}\theta_{12}}S_{\theta_{12}\theta_{12}}^{-1}\right)\Omega_1\left(1,-S_{\beta_{11}\theta_{12}}S_{\theta_{12}\theta_{12}}^{-1}\right)'}{2\widetilde{S}_{\beta_{11}\beta_{11}}\eta^2}\chi_1^2+\xi(\varphi,\phi)$$

which is sum of a scaled χ_1^2 distribution and $\xi(\varphi,\phi)$, where the χ_1^2 distribution and $\xi(\varphi,\phi)$ are independent.
∎

Proof of Theorem 7. By the CMT,

$$\begin{aligned}
& S_n\left(\gamma_0,\widehat{\underline{\theta}}\right)-S_n\left(\widehat{\gamma},\widehat{\underline{\theta}}\right) \\
= & \left(S_n\left(\gamma_0,\widehat{\underline{\theta}}\right)-S_n\left(\gamma_0,\underline{\theta}_0\right)\right)-\left(S_n\left(\widehat{\gamma},\widehat{\underline{\theta}}\right)-S_n\left(\gamma_0,\underline{\theta}_0\right)\right) \\
& \xrightarrow{d}\frac{1}{2}\widehat{u}'S_{\underline{\theta\theta}}\widehat{u}-\widehat{u}'W-\min_{u,v}\left\{\frac{1}{2}u'S_{\underline{\theta\theta}}u-u'W+C\left(v\right)\right\} \\
= & \max_v\ \left\{-C\left(v\right)\right\}\overset{d}{=}\eta^2\xi(\varphi,\phi),
\end{aligned}$$

and the rest of the proof is the same as in the proof of Theorem 5, where $\widehat{u}=\arg\min_u\left\{\frac{1}{2}u'S_{\underline{\theta\theta}}u-u'W+C\left(v\right)\right\}=\arg\min_u\left\{\frac{1}{2}u'S_{\underline{\theta\theta}}u-u'W\right\}$. Next,

$$\begin{aligned}
& S_n\left(\gamma_0,\beta_{11}^0,\widehat{\underline{\theta}}_{-11}\right)-S_n\left(\widehat{\gamma},\widehat{\beta}_{11},\widehat{\underline{\theta}}_{-11}\right) \\
= & \left(S_n\left(\gamma_0,\beta_{11}^0,\widehat{\underline{\theta}}_{-11}\right)-S_n\left(\gamma_0,\underline{\theta}_0\right)\right)-\left(S_n\left(\widehat{\gamma},\widehat{\beta}_{11},\widehat{\underline{\theta}}_{-11}\right)-S_n\left(\gamma_0,\underline{\theta}_0\right)\right)
\end{aligned}$$

$$
\begin{aligned}
&\xrightarrow{d} \frac{1}{2}\widehat{u}_{12}'S_{\theta_{12}\theta_{12}}\widehat{u}_{12} - \widehat{u}_{12}'W_{12} + \frac{1}{2}\widehat{u}_2'S_{\theta_2\theta_2}^2\widehat{u}_2 - \widehat{u}_2'W_2 - \min_{u,v}\left\{\frac{1}{2}u'S_{\underline{\theta\theta}}u - u'W + C(v)\right\} \\
=\ & -\frac{1}{2}S_{\beta_{11}\beta_{11}}\widehat{u}_{11}^2 - \widehat{u}_{11}S_{\beta_{11}\theta_{12}}\widehat{u}_{12} + \widehat{u}_{11}W_{11} + \max_v\ \{-C(v)\} \\
=\ & \frac{W_1'\left(1, -S_{\beta_{11}\theta_{12}}S_{\theta_{12}\theta_{12}}^{-1}\right)'}{\widetilde{S}_{\beta_{11}\beta_{11}}}\left[W_{11} - \frac{1}{2}\frac{\left(1, -S_{\beta_{11}\theta_{12}}S_{\theta_{12}\theta_{12}}^{-1}\right)W_1}{\widetilde{S}_{\beta_{11}\beta_{11}}/S_{\beta_{11}\beta_{11}}} - S_{\beta_{11}\theta_{12}}\widetilde{S}_{\theta_{12}\theta_{12}}^{-1}\left(-S_{\theta_{12}\beta_{11}}, \mathbf{I}\right)W_1\right] \\
&+ \max_v\ \{-C(v)\} \stackrel{d}{=} g(W_1) + \eta^2\xi(\varphi, \phi),
\end{aligned}
$$

where

$$
(\widehat{u}_1, \widehat{u}_2) = \arg\min_{u_1,u_2}\left\{\frac{1}{2}u_1'S_{\theta_1\theta_1}u_1 - u_1'W_1 + \frac{1}{2}u_2'S_{\theta_2\theta_2}u_2 - u_2'W_2\right\} = \left(S_{\theta_1\theta_1}^{-1}W_1, S_{\theta_2\theta_2}^{-1}W_2\right),
$$

with

$$
\begin{aligned}
\widehat{u}_{11} &= \frac{\left(1, -S_{\beta_{11}\theta_{12}}S_{\theta_{12}\theta_{12}}^{-1}\right)W_1}{\widetilde{S}_{\beta_{11}\beta_{11}}} \text{ and} \\
\widehat{u}_{12} &= \left(S_{\theta_{12}\theta_{12}} - S_{\theta_{12}\beta_{11}}S_{\beta_{11}\beta_{11}}^{-1}S_{\beta_{11}\theta_{12}}\right)^{-1}\left(-S_{\theta_{12}\beta_{11}}, \mathbf{I}\right)W_1 =: \widetilde{S}_{\theta_{12}\theta_{12}}^{-1}\left(-S_{\theta_{12}\beta_{11}}, \mathbf{I}\right)W_1
\end{aligned}
$$

$$
\begin{aligned}
=\ & \left(S_{\theta_{12}\theta_{12}}^{-1} + S_{\theta_{12}\theta_{12}}^{-1}S_{\theta_{12}\beta_{11}}\widetilde{S}_{\beta_{11}\beta_{11}}^{-1}S_{\beta_{11}\theta_{12}}S_{\theta_{12}\theta_{12}}^{-1}\right)\left(-S_{\theta_{12}\beta_{11}}, \mathbf{I}\right)W_1 \\
=\ & S_{\theta_{12}\theta_{12}}^{-1}W_{12} - S_{\theta_{12}\theta_{12}}^{-1}S_{\theta_{12}\beta_{11}}W_{11} - S_{\theta_{12}\theta_{12}}^{-1}S_{\theta_{12}\beta_{11}}\widetilde{S}_{\beta_{11}\beta_{11}}^{-1}S_{\beta_{11}\theta_{12}}S_{\theta_{12}\theta_{12}}^{-1}S_{\theta_{12}\beta_{11}}W_{11} \\
&+ S_{\theta_{12}\theta_{12}}^{-1}S_{\theta_{12}\beta_{11}}\widetilde{S}_{\beta_{11}\beta_{11}}^{-1}S_{\beta_{11}\theta_{12}}S_{\theta_{12}\theta_{12}}^{-1}W_{12} \\
=\ & S_{\theta_{12}\theta_{12}}^{-1}W_{12} + S_{\theta_{12}\theta_{12}}^{-1}S_{\theta_{12}\beta_{11}}\widetilde{S}_{\beta_{11}\beta_{11}}^{-1}S_{\beta_{11}\theta_{12}}S_{\theta_{12}\theta_{12}}^{-1}W_{12} - \left(\mathbf{I} + S_{\theta_{12}\theta_{12}}^{-1}S_{\theta_{12}\beta_{11}}\widetilde{S}_{\beta_{11}\beta_{11}}^{-1}S_{\beta_{11}\theta_{12}}\right)S_{\theta_{12}\theta_{12}}^{-1}S_{\theta_{12}\beta_{11}}W_{11}.
\end{aligned}
$$

The rest of the proof is similar to that of Theorem 6. ■

Appendix C: Lemmas

This appendix collects lemmas for consistency, convergence rates and local approximation. First, some notation is collected for subsequent reference. The letter C is used as a generic positive constant, which need not be the same in each location. The subscript 0 indicates the peudo-true value, e.g., $\theta_0 = \left(\gamma_0, \underline{\theta}_0'\right)' = \left(\gamma_0, \vartheta_1', \vartheta_2'\right)'$. P_N is the empirical probability measure, and $\mathbb{G}_N f = \sqrt{n}\left(P_N - P\right)f$ is the empirical process indexed by f. Define $w_i = (y_{it}, \breve{\mathbf{x}}_{it})_{t=1}^T$, and

$$
s(w_i|\theta) = \frac{1}{2}\sum_{t=1}^T\left(y_{it} - \breve{\mathbf{x}}_{it}'\theta_1 1(q_{it} \le \gamma) - \breve{\mathbf{x}}_{it}'\theta_2 1(q_{it} > \gamma)\right)^2;
$$

then

$$
S_n(\theta) = NP_N s(w_i|\theta).
$$

Since

$$
\begin{aligned}
s(w_i|\theta) &= \tfrac{1}{2}\textstyle\sum_{t=1}^T\left(g_1(\breve{\mathbf{x}}_{it}) - \breve{\mathbf{x}}_{it}'\theta_1 + e_{1it}\right)^2 1(q_{it} \le \gamma \wedge \gamma_0) + \tfrac{1}{2}\sum_{t=1}^T\left(g_2(\breve{\mathbf{x}}_{it}) - \breve{\mathbf{x}}_{it}'\theta_2 + e_{2it}\right)^2 1(q_{it} > \gamma \vee \gamma_0) \\
&+ \tfrac{1}{2}\textstyle\sum_{t=1}^T\left(g_1(\breve{\mathbf{x}}_{it}) - \breve{\mathbf{x}}_{it}'\theta_2 + e_{1it}\right)^2 1(\gamma \wedge \gamma_0 < q_{it} \le \gamma_0) + \tfrac{1}{2}\sum_{t=1}^T\left(g_2(\breve{\mathbf{x}}_{it}) - \breve{\mathbf{x}}_{it}'\theta_1 + e_{2it}\right)^2 1(\gamma_0 < q_{it} \le \gamma \vee \gamma_0),
\end{aligned}
$$

where $g_1(\breve{\mathbf{x}}_{it}) = \sum_{\ell=1}^{m_1} \breve{\mathbf{x}}_{it}'\delta^\ell 1(\gamma_{\ell-1}^0 < q_{it} \leq \gamma_\ell^0) + \breve{\mathbf{x}}_{it}'\theta_{m+1}$, and $g_2(\breve{\mathbf{x}}_{it}) = \sum_{\ell=m_1+1}^{m} \breve{\mathbf{x}}_{it}'\delta^\ell 1(\gamma_{\ell-1}^0 < q_{it} \leq \gamma_\ell^0) + \breve{\mathbf{x}}_{it}'\theta_{m+1}$, we have

$$\begin{aligned}
&s(w_i|\theta) - s(w_i|\theta_0) \\
&= \sum_{t=1}^{T}\left[\left(g_1(\breve{\mathbf{x}}_{it}) - \breve{\mathbf{x}}_{it}'\tfrac{\theta_1+\theta_{10}}{2}\right)\breve{\mathbf{x}}_{it}'(\theta_{10}-\theta_1) + (\theta_{10}-\theta_1)'\breve{\mathbf{x}}_{it}e_{1it}\right]1(q_{it} \leq \gamma\wedge\gamma_0) \\
&+ \sum_{t=1}^{T}\left[\left(g_2(\breve{\mathbf{x}}_{it}) - \breve{\mathbf{x}}_{it}'\tfrac{\theta_2+\theta_{20}}{2}\right)\breve{\mathbf{x}}_{it}'(\theta_{20}-\theta_2) + (\theta_{20}-\theta_2)'\breve{\mathbf{x}}_{it}e_{2it}\right]1(q_{it} > \gamma\vee\gamma_0) \\
&+ \sum_{t=1}^{T}\left[\left(g_1(\breve{\mathbf{x}}_{it}) - \breve{\mathbf{x}}_{it}'\tfrac{\theta_{10}+\theta_2}{2}\right)\breve{\mathbf{x}}_{it}'(\theta_{10}-\theta_2) + (\theta_{10}-\theta_2)\breve{\mathbf{x}}_{it}e_{1it}\right]1(\gamma\wedge\gamma_0 < q_{it} \leq \gamma_0) \\
&+ \sum_{t=1}^{T}\left[\left(g_2(\breve{\mathbf{x}}_{it}) - \breve{\mathbf{x}}_{it}'\tfrac{\theta_{20}+\theta_1}{2}\right)\breve{\mathbf{x}}_{it}'(\theta_{20}-\theta_1) + (\theta_{20}-\theta_1)\breve{\mathbf{x}}_{it}e_{2it}\right]1(\gamma_0 < q_{it} \leq \gamma\vee\gamma_0) \\
&=: \sum_{t=1}^{T} T(w_{it}|\theta_1,\theta_{10})1(q_{it} \leq \gamma\wedge\gamma_0) + \sum_{t=1}^{T} T(w_{it}|\theta_2,\theta_{20})1(q_{it} > \gamma\vee\gamma_0) \\
&+ \sum_{t=1}^{T} z_1(w_{it}|\theta_2,\theta_{10})1(\gamma\wedge\gamma_0 < q_{it} \leq \gamma_0) + \sum_{t=1}^{T} z_2(w_{it}|\theta_1,\theta_{20})1(\gamma_0 < q_{it} \leq \gamma\vee\gamma_0).
\end{aligned} \tag{28}$$

Here $g_1(\breve{\mathbf{x}}_{it}) - \breve{\mathbf{x}}_{it}'\frac{\theta_1+\theta_{10}}{2} = g_1(\breve{\mathbf{x}}_{it}) - \breve{\mathbf{x}}_{it}'\theta_{10} + \breve{\mathbf{x}}_{it}'\frac{\theta_{10}-\theta_1}{2}$, with $g_1(\breve{\mathbf{x}}_{it}) - \breve{\mathbf{x}}_{it}'\theta_{10}$ equal to

$$\breve{\mathbf{x}}_{it}'\delta_- =: \breve{\mathbf{x}}_{it}'\left(\delta^{m_1} - M(\gamma_0)^{-1}\sum_{\ell=1}^{m_1} M\left(\gamma_{\ell-1}^0,\gamma_\ell^0\right)\delta^\ell\right) = N^{-\kappa}\breve{\mathbf{x}}_{it}'c_-$$

for q_{it} in the left neighborhood of γ_0; and $g_2(\breve{\mathbf{x}}_{it}) - \breve{\mathbf{x}}_{it}'\frac{\theta_2+\theta_{20}}{2} = g_2(\breve{\mathbf{x}}_{it}) - \breve{\mathbf{x}}_{it}'\theta_{20} + \breve{\mathbf{x}}_{it}'\frac{\theta_{20}-\theta_2}{2}$, with $g_2(\breve{\mathbf{x}}_{it}) - \breve{\mathbf{x}}_{it}'\theta_{20}$ equal to

$$\breve{\mathbf{x}}_{it}'\delta_+ := \breve{\mathbf{x}}_{it}'\left(\delta^{m_1+1} - \overline{M}(\gamma_0)^{-1}\sum_{\ell=m_1+1}^{m} M\left(\gamma_{\ell-1}^0,\gamma_\ell^0\right)\delta^\ell\right) = N^{-\kappa}\breve{\mathbf{x}}_{it}'c_+$$

for q_{it} in the right neighborhood of γ_0; also $g_1(\breve{\mathbf{x}}_{it}) - \breve{\mathbf{x}}_{it}'\frac{\theta_{10}+\theta_2}{2} = g_1(\breve{\mathbf{x}}_{it}) - \breve{\mathbf{x}}_{it}'\theta_{10} + \breve{\mathbf{x}}_{it}'\frac{\theta_{10}-\theta_2}{2}$, and $g_2(\breve{\mathbf{x}}_{it}) - \breve{\mathbf{x}}_{it}'\frac{\theta_{20}+\theta_1}{2} = g_2(\breve{\mathbf{x}}_{it}) - \breve{\mathbf{x}}_{it}'\theta_{20} + \breve{\mathbf{x}}_{it}'\frac{\theta_{20}-\theta_1}{2}$.

Lemma 3 *Under Assumption 4,* $N^\kappa\left(\widehat{\theta}_\ell - \theta_{\ell 0}\right) = o_p(1)$.

Proof. Take $\widehat{\theta}_1$ as an example. Recall that

$$\begin{aligned}
\widehat{\theta}_1 &= \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{it}'1(q_{it} \leq \widehat{\gamma})\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\breve{\mathbf{x}}_{it}y_{it}1(q_{it} \leq \widehat{\gamma})\right) \\
&= \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{it}'1(q_{it} \leq \widehat{\gamma})\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\breve{\mathbf{x}}_{it}\left(\sum_{\ell=1}^{m}\breve{\mathbf{x}}_{it}'\delta^\ell 1(\gamma_{\ell-1}^0 < q_{it} \leq \gamma_\ell^0) + \breve{\mathbf{x}}_{it}'\theta_{m+1} + e_{it}^0\right)1(q_{it} \leq \widehat{\gamma})\right),
\end{aligned}$$

so

$$\begin{aligned}
N^\kappa\left(\widehat{\theta}_1 - \theta_{10}\right) &= \left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{it}'1(q_{it} \leq \widehat{\gamma})\right)^{-1}\frac{N^\kappa}{N^{1/2}}\left(\frac{1}{N^{1/2}}\sum_{i=1}^{N}\sum_{t=1}^{T}\breve{\mathbf{x}}_{it}e_{it}^0 1(q_{it} \leq \widehat{\gamma})\right) \\
&+ \left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{it}'1(q_{it} \leq \widehat{\gamma})\right)^{-1}\sum_{\ell=1}^{m}\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{it}'1(\gamma_{\ell-1}^0 < q_{it} \leq \gamma_\ell^0 \wedge \widehat{\gamma})\right)c^\ell \\
&- M(\gamma_0)^{-1}\sum_{\ell=1}^{m_1} M\left(\gamma_{\ell-1}^0,\gamma_\ell^0\right)c^\ell.
\end{aligned}$$

The first term is $o_p(1)$ because by Condition (iv) and (v), $\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{it}'1(q_{it} \leq \widehat{\gamma}) \xrightarrow{p} M(\gamma_0)$, $\frac{1}{N^{1/2}}\sum_{i=1}^{N}\sum_{t=1}^{T}\breve{\mathbf{x}}_{it}e_{it}^0 1(q_{it} \leq \widehat{\gamma}) = O_p(1)$, and $0 < \kappa < 1/2$, and the difference of the second and third terms can be shown to be $o_p(1)$ by applying a Glivenko-Cantelli theorem and using $\widehat{\gamma} - \gamma_0 = o_p(1)$. ■

Lemma 4 *Under Assumption 4,* $\lambda_N(\widehat{\gamma} - \gamma_0) = O_p(1)$ *and* $N^{1/2}\left(\underline{\widehat{\theta}} - \underline{\theta}_0\right) = O_p(1)$.

Proof. Since δ_N depends on N, we apply the proof idea in Theorem 3.2.5 in Van der Vaart and Wellner (2023) (VW hereafter) to establish this result. Define $d_N(\theta,\theta_0) = \|\underline{\theta} - \underline{\theta}_0\| + \|\delta_N\| \sqrt{\|\gamma-\gamma_0\|}$ for θ in a neighborhood of θ_0, and

$$\begin{aligned}
&Q_n(\theta) = \tfrac{1}{\lambda_N}\left(S_n(\theta) - S_n(\theta_0)\right) = \tfrac{1}{\lambda_N}\textstyle\sum_{i=1}^N \left(s(w_i|\theta) - s(w_i|\theta_0)\right) \\
&= \tfrac{1}{\lambda_N}\textstyle\sum_{i=1}^N \sum_{t=1}^T T(w_{it}|\theta_1,\theta_{10})\, 1(q_{it} \le \gamma\wedge\gamma_0) + \tfrac{1}{\lambda_N}\sum_{i=1}^N \sum_{t=1}^T T(w_{it}|\theta_2,\theta_{20})\, 1(q_{it} > \gamma\vee\gamma_0) \\
&\quad + \tfrac{1}{\lambda_N}\textstyle\sum_{i=1}^N \sum_{t=1}^T z_1(w_{it}|\theta_2,\theta_{10})\, 1(\gamma\wedge\gamma_0 < q_{it} \le \gamma_0) \\
&\quad + \tfrac{1}{\lambda_N}\textstyle\sum_{i=1}^N \sum_{t=1}^T z_2(w_{it}|\theta_1,\theta_{20})\, 1(\gamma_0 < q_{it} \le \gamma\vee\gamma_0) \\
&=: T_1(\theta) + T_2(\theta) + T_3(\theta) + T_4(\theta).
\end{aligned}$$

For each N, the parameter space (minus θ_0) can be partitioned into the "shells" $S_{j,N} = \left\{\theta : 2^{j-1} < \sqrt{N} d_N(\theta,\theta_0) \le 2^j\right\}$ with j ranging over the integers. Given an integer J,

$$\begin{aligned}
P\left(d_N\left(\widehat{\theta},\theta_0\right) > 2^J\right) \le &\sum_{j\ge J, \|\underline{\theta}-\underline{\theta}_0\| < M\|\delta_N\|, \|\gamma-\gamma_0\|<\eta} P\left(\inf_{\theta\in S_{j,N}} Q_n(\theta) \le 0\right) \\
&+P\left(2\|\underline{\theta}-\underline{\theta}_0\| \ge M\|\delta_N\|, 2\|\gamma-\gamma_0\| \ge \eta\right),
\end{aligned} \tag{29}$$

where M and η are small positive numbers. The second term on the right hand side of (29) converges to zero as $N\to\infty$ for every $\eta>0$ and $M>0$ by Lemma 3 and the consistency of $\widehat{\gamma}$, so we can concentrate on the first term.

$$\begin{aligned}
&P\left(\inf_{\theta\in S_{j,N}} Q_n(\theta) \le 0\right) \le P\left(\sup_{\theta\in S_{j,N}} |Q_n(\theta) - \mathbb{E}[Q_n(\theta)]| \ge \inf_{\theta\in S_{j,N}} |\mathbb{E}[Q_n(\theta)]|\right) \\
&\le \mathbb{E}\left[\sup_{\theta\in S_{j,N}} |Q_n(\theta)-\mathbb{E}[Q_n(\theta)]|\right] \Big/ \inf_{\theta\in S_{j,N}} |\mathbb{E}[Q_n(\theta)]| \le \sum_{k=1}^4 \mathbb{E}\left[\sup_{\theta\in S_{j,N}} |T_k(\theta) - \mathbb{E}[T_k(\theta)]|\right] \Big/ \inf_{\theta\in S_{j,N}} |\mathbb{E}[Q_n(\theta)]|,
\end{aligned}$$

where the second-to-last equality is from Markov's inequality.

Note that θ_{10} satisfies

$$\mathbb{E}\left[\sum\nolimits_{t=1}^T \breve{\mathbf{x}}_{it}\left(g_1(\breve{\mathbf{x}}_{it}) - \breve{\mathbf{x}}_{it}'\theta_{10}\right) 1(q_{it}\le\gamma_0)\right] = 0,$$

and similarly θ_{20} satisfies

$$\mathbb{E}\left[\sum\nolimits_{t=1}^T \breve{\mathbf{x}}_{it}\left(g_2(\breve{\mathbf{x}}_{it}) - \breve{\mathbf{x}}_{it}'\theta_{20}\right) 1(q_{it}>\gamma_0)\right] = 0.$$

So $\mathbb{E}[T_1(\theta)] \approx C\frac{N}{\lambda_N}\|\theta_1-\theta_{10}\|^2$, and $\mathbb{E}[T_2(\theta)] \approx C\frac{N}{\lambda_N}\|\theta_2-\theta_{20}\|^2$. From the analysis after (28) and the fact that γ_0 minimizes $\mathbb{E}[s(w_i|\theta)-s(w_i|\theta_0)]$, $\mathbb{E}[T_3(\theta)] \approx C\frac{N}{\lambda_N}\|\delta_N\|^2\|\gamma-\gamma_0\|$, and $\mathbb{E}[T_4(\theta)] \approx C\frac{N}{\lambda_N}\|\delta_N\|^2\|\gamma-\gamma_0\|$, where note that because $\theta_{10}-\theta_{20} = \delta_N$ and $\|\theta_\ell - \theta_{\ell 0}\| < M\|\delta_N\|$ so that $\|\theta_1-\theta_{20}\| = O(\|\delta_N\|)$ and $\|\theta_{20}-\theta_1\| = O(\|\delta_N\|)$. As a result,

$$\begin{aligned}
&\inf_{\theta\in S_{j,N}} |\mathbb{E}[Q_n(\theta)]| = \inf_{\theta\in S_{j,N}} \left|\textstyle\sum_{k=1}^4 \mathbb{E}[T_k(\theta)]\right| \\
&= \inf_{\theta\in S_{j,N}} C\left|\tfrac{N}{\lambda_N}\|\underline{\theta}-\underline{\theta}_0\|^2 + \tfrac{N}{\lambda_N}\|\delta_N\|^2\|\gamma-\gamma_0\|\right| = \inf_{\theta\in S_{j,N}} C\tfrac{N}{\lambda_N} d_N(\theta,\theta_0)^2 \ge C\tfrac{2^{2j-2}}{\lambda_N} = C\tfrac{2^{2j}}{\lambda_N}.
\end{aligned}$$

To bound $\sum_{k=1}^4 \mathbb{E}\left[\sup_{\theta\in S_{j,N}} |T_k(\theta)-\mathbb{E}[T_k(\theta)]|\right]$, we apply a maximal inequality (e.g., Theorem 2.14.1 of VW). For this purpose, we need to obtain an envelope F of $s(w_i|\theta) - s(w_i|\theta_0)$ over $S_{j,N}$. It is not hard to see that we can choose $F = \sup_{\theta \in S_{j,N}} \mathcal{F}$ with

$$\begin{aligned}
\mathcal{F} &= \textstyle\sum_{t=1}^{T} \left(\left| g_1(\breve{\mathbf{x}}_{it}) - \breve{\mathbf{x}}_{it}'\theta_{10} \right| \|\breve{\mathbf{x}}_{it}\| \|\theta_{10}-\theta_1\| + \frac{1}{2}\|\theta_{10}-\theta_1\|^2 \|\breve{\mathbf{x}}_{it}\|^2 + \|\theta_{10}-\theta_1\| \|\breve{\mathbf{x}}_{it} e_{1it}\| \right) 1(q_{it} \le \gamma \wedge \gamma_0) \\
&\quad + \textstyle\sum_{t=1}^{T} \left(\left| g_2(\breve{\mathbf{x}}_{it}) - \breve{\mathbf{x}}_{it}'\theta_{20} \right| \|\breve{\mathbf{x}}_{it}\| \|\theta_{20}-\theta_2\| + \frac{1}{2}\|\theta_{20}-\theta_2\|^2 \|\breve{\mathbf{x}}_{it}\|^2 + \|\theta_{20}-\theta_2\| \|\breve{\mathbf{x}}_{it} e_{2it}\| \right) 1(q_{it} > \gamma \vee \gamma_0) \\
&\quad + \textstyle\sum_{t=1}^{T} \left(\left| g_1(\breve{\mathbf{x}}_{it}) - \breve{\mathbf{x}}_{it}'\theta_{10} \right| \|\breve{\mathbf{x}}_{it}\| \|\theta_{10}-\theta_2\| + \frac{1}{2}\|\theta_{10}-\theta_2\|^2 \|\breve{\mathbf{x}}_{it}\|^2 + \|\theta_{10}-\theta_2\| \|\breve{\mathbf{x}}_{it} e_{1it}\| \right) 1(\gamma \wedge \gamma_0 < q_{it} \le \gamma_0) \\
&\quad + \textstyle\sum_{t=1}^{T} \left(\left| g_2(\breve{\mathbf{x}}_{it}) - \breve{\mathbf{x}}_{it}'\theta_{20} \right| \|\breve{\mathbf{x}}_{it}\| \|\theta_{20}-\theta_1\| + \frac{1}{2}\|\theta_{20}-\theta_1\|^2 \|\breve{\mathbf{x}}_{it}\|^2 + \|\theta_{20}-\theta_1\| \|\breve{\mathbf{x}}_{it} e_{2it}\| \right) 1(\gamma_0 < q_{it} \le \gamma \vee \gamma_0) \\
&=: F_1(\breve{\mathbf{x}}_i, e_{1i}|\theta_1) + F_2(\breve{\mathbf{x}}_i, e_{2i}|\theta_2) + F_3(\breve{\mathbf{x}}_i, e_{1i}|\theta_2,\gamma) + F_4(\breve{\mathbf{x}}_i, e_{2i}|\theta_1,\gamma),
\end{aligned}$$

so by Conditions (iv) and (viii), and letting N large enough such that $\|\theta_{\ell 0}-\theta_\ell\| < 1$ and $\|\theta_{10}-\theta_2\| < 1$ in (29), we have

$$\sum_{k=1}^{2} \mathbb{E}\left[\sup_{\theta \in S_{j,N}} |T_k(\theta) - \mathbb{E}[T_k(\theta)]|\right] \le C\frac{\sqrt{\mathbb{E}\left[\sup_{\theta \in S_{j,N}} F_1(\breve{\mathbf{x}}_i, e_{1i}|\theta_1)^2\right] + \mathbb{E}\left[\sup_{\theta \in S_{j,N}} F_2(\breve{\mathbf{x}}_i, e_{1i}|\theta_2)^2\right]}}{\sqrt{N}\|\delta_N\|^2} = C\frac{\sup_{\theta \in S_{j,N}} \left\|\underline{\theta}-\underline{\theta}_0\right\|}{\sqrt{N}\|\delta_N\|^2},$$

$$\mathbb{E}\left[\sup_{\theta \in S_{j,N}} |T_3(\theta) - \mathbb{E}[T_3(\theta)]|\right] \le C\frac{\sqrt{\mathbb{E}\left[\sup_{\theta \in S_{j,N}} F_3(\breve{\mathbf{x}}_i, e_{1i}|\theta_2,\gamma)^2\right]}}{\sqrt{N}\|\delta_N\|^2} \le C\frac{\sup_{\theta \in S_{j,N}} \sqrt{\|\theta_{10}-\theta_2\|^2}\sqrt{|\gamma-\gamma_0|}}{\sqrt{N}\|\delta_N\|^2} = C\frac{\sup_{\theta \in S_{j,N}} \|\delta_N\|\sqrt{|\gamma-\gamma_0|}}{\sqrt{N}\|\delta_N\|^2}.$$

Similarly, $\mathbb{E}\left[\sup_{\theta \in S_{j,N}} |T_4(\theta) - \mathbb{E}[T_4(\theta)]|\right] \le C\frac{\sup_{\theta \in S_{j,N}} \|\delta_N\|\sqrt{|\gamma-\gamma_0|}}{\sqrt{n}\|\delta_N\|^2}$. As a result,

$$\sum_{k=1}^{4} \mathbb{E}\left[\sup_{\theta \in S_{j,N}} |T_k(\theta) - \mathbb{E}[T_k(\theta)]|\right] \le C\frac{\sup_{\theta \in S_{j,N}} d_N(\theta,\theta_0)}{\sqrt{N}\|\delta_N\|^2} \le C\frac{2^j/\sqrt{N}}{\sqrt{N}\|\delta_N\|^2} = C\frac{2^j}{\lambda_N}.$$

In summary,

$$\sum_{j \ge J, \left\|\underline{\theta}-\underline{\theta}_0\right\| < M\|\delta_N\|, |\gamma-\gamma_0|<\eta} P\left(\sup_{\theta \in S_{j,N}} Q_n(\theta) \ge 0\right) \le C\sum_{j \ge J}\left(\frac{2^j}{\lambda_N}\Big/\frac{2^{2j}}{\lambda_N}\right) \le C\sum_{j \ge J}\frac{1}{2^j},$$

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 $\lambda_N(\widehat{\gamma}-\gamma_0) = O_p(1)$, and $N^{1/2}\left(\widehat{\underline{\theta}} - \underline{\theta}_0\right) = O_p(1)$. ■

Lemma 5 *Under Assumption 4, uniformly for $h := (v, u')' := (v, u_1', u_2')'$ in a compact set,*

$$S_n\left(\gamma_0 + \frac{v}{\lambda_N}, \underline{\theta}_0 + \frac{u}{N^{1/2}}\right) - S_n(\gamma_0,\beta_0) = \frac{1}{2}u_1'M_1u_1 + \frac{1}{2}u_2'M_2u_2 - W_n(u) + C_n(v) + o_P(1),$$

where $W_n(u) = W_{1n}(u_1) + W_{2n}(u_2)$ with

$$W_{1n}(u_1) = \frac{u_1'}{\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T}\breve{\mathbf{x}}_{it}e_{1it}1(q_{it} \le \gamma_0) \text{ and } W_{2n}(u_2) = \frac{u_2'}{\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T}\breve{\mathbf{x}}_{it}e_{2it}1(q_i > \gamma_0),$$

and

$$C_n(v) \quad = \quad S_n\left(\gamma_0 + \frac{v}{\lambda_N}, \beta_0\right) - S_n(\gamma_0,\beta_0)$$

$$= \sum_{i=1}^{N}\sum_{t=1}^{T} z_{1it} 1\left(\gamma_0 + \frac{v}{\lambda_N} < q_{it} \le \gamma_0\right) - \sum_{i=1}^{N}\sum_{t=1}^{T} z_{2it} 1\left(\gamma_0 < q_{it} \le \gamma_0 + \frac{v}{\lambda_N}\right)$$

with $z_{1it} = \frac{1}{2}\delta_N' \breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{it}'\delta_N + \delta_- \breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{it}'\delta_N + \delta_N' \breve{\mathbf{x}}_{it} e_{1it}$ *and* $z_{2it} = \frac{1}{2}\delta_N' \breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{it}'\delta_N - \delta_+ \breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{it}'\delta_N - \delta_N' \breve{\mathbf{x}}_{it} e_{2it}$.

Proof. From the decomposition of $s(w_i|\theta) - s(w_i|\theta_0)$,

$$\begin{aligned}
& S_n\left(\gamma_0 + \tfrac{v}{\lambda_N}, \underline{\theta}_0 + \tfrac{u}{N^{1/2}}\right) - S_n(\gamma_0, \theta_0) \\
&= \sum_{i=1}^{N}\sum_{t=1}^{T} T\left(w_{it}|\theta_{10} + \tfrac{u_1}{N^{1/2}}, \theta_{10}\right) 1(q_{it} \le \gamma_0) + \sum_{i=1}^{n} T\left(w_{it}|\theta_{20} + \tfrac{u_2}{N^{1/2}}, \theta_{20}\right) 1(q_{it} > \gamma_0) \\
&\quad + \sum_{i=1}^{N}\sum_{t=1}^{T} z_1\left(w_{it}|\theta_{20} + \tfrac{u_2}{N^{1/2}}, \theta_{10}\right) 1(\gamma_0 + \tfrac{v}{\lambda_N} \wedge \gamma_0 < q_{it} \le \gamma_0) \\
&\quad - \sum_{i=1}^{N}\sum_{t=1}^{T} T\left(w_{it}|\theta_{10} + \tfrac{u_1}{N^{1/2}}, \theta_{10}\right) 1(\gamma_0 + \tfrac{v}{\lambda_N} \wedge \gamma_0 < q_{it} \le \gamma_0) \\
&\quad + \sum_{i=1}^{N}\sum_{t=1}^{T} z_2\left(w_{it}|\theta_{10} + \tfrac{u_1}{N^{1/2}}, \theta_{20}\right) 1(\gamma_0 < q_{it} \le \gamma_0 + \tfrac{v}{\lambda_N} \vee \gamma_0) \\
&\quad - \sum_{i=1}^{N}\sum_{t=1}^{T} T\left(w_{it}|\theta_{20} + \tfrac{u_2}{N^{1/2}}, \theta_{20}\right) 1(\gamma_0 < q_{it} \le \gamma_0 + \tfrac{v}{\lambda_N} \vee \gamma_0) \\
&=: T_1(u_1) + T_2(u_2) + T_3(u_2, v) - T_4(u_1, v) + T_5(u_1, v) - T_6(u_2, v).
\end{aligned}$$

Check each term in turn. The analyses for $T_2(u_2), T_5(u_1, v)$ and $T_6(u_2, v)$ are similar to $T_1(u_1), T_3(u_2, v)$ and $T_4(u_1, v)$, so we concentrate the latter three terms below.

First,

$$\begin{aligned}
& T_1(u_1) = \sum_{i=1}^{N}\sum_{t=1}^{T}\left[\tfrac{1}{2}\tfrac{u_1'}{N^{1/2}}\breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{it}'\tfrac{u_1}{N^{1/2}} + \left(g_1(\breve{\mathbf{x}}_{it}) - \breve{\mathbf{x}}_{it}'\theta_{10}\right)\breve{\mathbf{x}}_{it}'\tfrac{u_1}{N^{1/2}} - \tfrac{u_1'}{N^{1/2}}\breve{\mathbf{x}}_{it} e_{1it}\right] 1(q_{it} \le \gamma_0) \\
&= \tfrac{1}{2}u_1'\left[\tfrac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{it}' 1(q_{it} \le \gamma_0)\right] u_1 + \tfrac{u_1'}{N^{1/2}}\sum_{i=1}^{N}\sum_{t=1}^{T}\breve{\mathbf{x}}_{it}\left(g_1(\breve{\mathbf{x}}_{it}) - \breve{\mathbf{x}}_{it}'\theta_{10}\right) 1(q_{it} \le \gamma_0) - \tfrac{u_1'}{N^{1/2}}\sum_{i=1}^{N}\sum_{t=1}^{T}\breve{\mathbf{x}}_{it} e_{1it} 1(q_{it} \le \gamma_0) \\
&= \tfrac{1}{2}u_1' M_1 u_1 - \tfrac{u_1'}{N^{1/2}}\sum_{i=1}^{N}\sum_{t=1}^{T}\breve{\mathbf{x}}_{it} e_{1it} 1(q_{it} \le \gamma_0) + o_p(1),
\end{aligned}$$

where $o_p(1)$ is from the LLN, and the fact that

$$\begin{aligned}
\mathbb{E}\left[\sum_{t=1}^{T}\breve{\mathbf{x}}_{it}\left(g_1(\breve{\mathbf{x}}_{it}) - \breve{\mathbf{x}}_{it}'\theta_{10}\right) 1(q_{it} \le \gamma_0)\right] &= 0, \\
Var\left(\sum_{t=1}^{T}\breve{\mathbf{x}}_{it}\left(g_1(\breve{\mathbf{x}}_{it}) - \breve{\mathbf{x}}_{it}'\theta_{10}\right) 1(q_{it} \le \gamma_0)\right) &= O\left(N^{-2\kappa}\right) = o(1).
\end{aligned}$$

By a similar analysis, when $v < 0$,

$$\begin{aligned}
& T_4(u_1, v) = \tfrac{1}{2}u_1'\mathbb{E}\left[\textstyle\sum_{t=1}^{T}\breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{it}' 1(\gamma_0 + \tfrac{v}{\lambda_N} < q_{it} \le \gamma_0)\right] u_1 - \tfrac{u_1'}{N^{1/2}}\sum_{i=1}^{N}\sum_{t=1}^{T}\breve{\mathbf{x}}_{it} e_{1it} 1(\gamma_0 + \tfrac{v}{\lambda_N} < q_{it} \le \gamma_0) + o_p(1) \\
&= o_p(1) - u_1'\left[\begin{array}{c} \mathbb{G}_N\left(\sum_{t=1}^{T}\breve{\mathbf{x}}_{it} e_{1it} 1(\gamma_0 + \tfrac{v}{\lambda_N} < q_{it} \le \gamma_0)\right) \\ +\sqrt{N}\mathbb{E}\left[\sum_{t=1}^{T}\breve{\mathbf{x}}_{it} e_{1it} 1(\gamma_0 + \tfrac{v}{\lambda_N} < q_{it} \le \gamma_0)\right] \end{array}\right] + o_p(1) \\
&= o_p(1),
\end{aligned}$$

where the $o_p(1)$ in the first equality is from a Glivenko-Cantelli theorem, and

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{N^{1/2}}\sum_{i=1}^{N}\sum_{t=1}^{T}\breve{\mathbf{x}}_{it}\left(g_1\left(\breve{\mathbf{x}}_{it}\right)-\breve{\mathbf{x}}_{it}'\theta_{10}\right)1(\gamma_0+\frac{v}{\lambda_N}<q_{it}\leq\gamma_0)\right] &= O\left(N^{1/2}N^{-\kappa}/\lambda_N\right)=o\left(1\right), \\
Var\left(\frac{1}{N^{1/2}}\sum_{i=1}^{N}\sum_{t=1}^{T}\breve{\mathbf{x}}_{it}\left(g_1\left(\breve{\mathbf{x}}_{it}\right)-\breve{\mathbf{x}}_{it}'\theta_{10}\right)1(\gamma_0+\frac{v}{\lambda_N}<q_{it}\leq\gamma_0)\right) &= O\left(N^{-2\kappa}/\lambda_N\right)+o\left(N^{-2\kappa}/\lambda_N\right)=o\left(1\right)
\end{aligned}$$

by Conditions (iv), (viii) and (A9) of Hansen (1999), and the first $o_p(1)$ in the second equality is because

$$\mathbb{E}\left[\sum_{t=1}^{T}\breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{it}'1(\gamma_0+\frac{v}{\lambda_N}<q_{it}\leq\gamma_0)\right]=\frac{|v|}{\lambda_N}\sum_{t=1}^{T}\mathbb{E}\left[\breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{it}'|q_{it}=\overline{\gamma}\right]f_t\left(\overline{\gamma}\right),$$

for some $\overline{\gamma}$ between $\gamma_0+\frac{v}{\lambda_N}$ and γ_0, which is which is $o\left(1\right)$ by Conditions (iv) and (viii), and $o_p\left(1\right)$ in the third equality is because by the stochastic equicontinuity of $\gamma\mapsto\mathbb{G}_N\left(\sum_{t=1}^{T}\breve{\mathbf{x}}_{it}e_{1it}1(\gamma<q_{it}\leq\gamma_0)\right)$,

$$\mathbb{G}_N\left(\textstyle\sum_{t=1}^{T}\breve{\mathbf{x}}_{it}e_{1it}1(\gamma_0+\frac{v}{\lambda_N}<q_{it}\leq\gamma_0)\right)=\mathbb{G}_N\left(\textstyle\sum_{t=1}^{T}\breve{\mathbf{x}}_{it}e_{1it}1(\gamma_0<q_{it}\leq\gamma_0)\right)+o_p(1)=o_p(1)$$

and

$$\sqrt{N}\mathbb{E}\left[\sum\nolimits_{t=1}^{T}\breve{\mathbf{x}}_{it}e_{1it}1(\gamma_0+\frac{v}{\lambda_N}<q_{it}\leq\gamma_0)\right]=0.$$

Finally, when $v<0$,

$$\begin{aligned}
&T_3\left(u_2,v\right)=\sum_{i=1}^{N}\sum_{t=1}^{T}z_1\left(w_{it}|\theta_{20}+\tfrac{u_2}{N^{1/2}},\theta_{10}\right)1(\gamma_0+\tfrac{v}{\lambda_N}\wedge\gamma_0<q_{it}\leq\gamma_0)\\
&=\tfrac{1}{2}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\theta_{10}-\left(\theta_{20}+\tfrac{u_2}{N^{1/2}}\right)\right)'\breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{it}'\left(\theta_{10}-\left(\theta_{20}+\tfrac{u_2}{N^{1/2}}\right)\right)1(\gamma_0+\tfrac{v}{\lambda_N}<q_{it}\leq\gamma_0)\\
&+\sum_{i=1}^{N}\sum_{t=1}^{T}\left(g_1\left(\breve{\mathbf{x}}_{it}\right)-\breve{\mathbf{x}}_{it}'\theta_{10}\right)\breve{\mathbf{x}}_{it}'\left(\theta_{10}-\left(\theta_{20}+\tfrac{u_2}{N^{1/2}}\right)\right)1(\gamma_0+\tfrac{v}{\lambda_N}<q_{it}\leq\gamma_0)\\
&+\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\theta_{10}-\left(\theta_{20}+\tfrac{u_2}{N^{1/2}}\right)\right)\breve{\mathbf{x}}_{it}e_{1it}1(\gamma_0+\tfrac{v}{\lambda_N}<q_{it}\leq\gamma_0)\\
&=\tfrac{1}{2}\sum_{i=1}^{N}\sum_{t=1}^{T}\delta_N'\breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{it}'\delta_N1(\gamma_0+\tfrac{v}{\lambda_N}<q_{it}\leq\gamma_0)\\
&-\tfrac{1}{N^{1/2}}\sum_{i=1}^{N}\sum_{t=1}^{T}\delta_N'\breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{it}'u_21(\gamma_0+\tfrac{v}{\lambda_N}<q_{it}\leq\gamma_0)\\
&+\tfrac{1}{2}u_2'\left[\tfrac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{it}'1(\gamma_0+\tfrac{v}{\lambda_N}<q_{it}\leq\gamma_0)\right]u_2\\
&+\sum_{i=1}^{N}\sum_{t=1}^{T}\delta_-\breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{it}'\delta_N1(\gamma_0+\tfrac{v}{\lambda_N}<q_{it}\leq\gamma_0)-\tfrac{1}{N^{1/2}}\sum_{i=1}^{N}\sum_{t=1}^{T}\delta_-\breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{it}'u_21(\gamma_0+\tfrac{v}{\lambda_N}<q_{it}\leq\gamma_0)\\
&+\sum_{i=1}^{N}\sum_{t=1}^{T}\delta_N'\breve{\mathbf{x}}_{it}e_{1it}1(\gamma_0+\tfrac{v}{\lambda_N}<q_{it}\leq\gamma_0)-\tfrac{u_2'}{N^{1/2}}\sum_{i=1}^{N}\sum_{t=1}^{T}\breve{\mathbf{x}}_{it}e_{1it}1(\gamma_0+\tfrac{v}{\lambda_N}<q_{it}\leq\gamma_0)\\
&=\tfrac{1}{2}\sum_{i=1}^{N}\sum_{t=1}^{T}\delta_N'\breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{it}'\delta_N1(\gamma_0+\tfrac{v}{\lambda_N}<q_{it}\leq\gamma_0)+o_p(1)\\
&+\sum_{i=1}^{N}\sum_{t=1}^{T}\delta_-\breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{it}'\delta_N1(\gamma_0+\tfrac{v}{\lambda_N}<q_{it}\leq\gamma_0)+o_p(1)\\
&+\sum_{i=1}^{N}\sum_{t=1}^{T}\delta_N'\breve{\mathbf{x}}_{it}e_{1it}1(\gamma_0+\tfrac{v}{\lambda_N}<q_{it}\leq\gamma_0)+o_p(1)\\
&=\sum_{i=1}^{N}\sum_{t=1}^{T}z_{1it}1\left(\gamma_0+\tfrac{v}{\lambda_N}<q_{it}\leq\gamma_0\right)+o_p(1).
\end{aligned}$$

The $o_p(1)$ in the fourth equality needs careful analysis. The second term

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\sum_{t=1}^{T}\delta_N'\breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{it}'u_2 1(\gamma_0+\frac{v}{\lambda_N}<q_{it}\leq\gamma_0)=o_p(1)$$

by the stochastic equicontinuity of $\gamma\mapsto\mathbb{G}_n\left(\sum_{t=1}^{T}\breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{it}'u_2 1(\gamma<q\leq\gamma_0)\right)$ and

$$\sqrt{N}\delta_N'\mathbb{E}\left[\sum_{t=1}^{T}\breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{it}'u_2 1(\gamma_0+\frac{v}{\lambda_N}<q_{it}\leq\gamma_0)\right]\leq\frac{|v|\sqrt{N}\|\delta_N\|}{\lambda_N}\|u_2\|\sum_{t=1}^{T}\mathbb{E}\left[\|\breve{\mathbf{x}}_{it}\|^2|q_{it}=\overline{\gamma}\right]f_t(\overline{\gamma}),$$

which is $o(1)$ since $\sqrt{N}\|\delta_N\|/\lambda_N=\lambda_N^{-1/2}$, the third term

$$\begin{aligned}&u_2'\left[\tfrac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{it}'1(\gamma_0+\tfrac{v}{\lambda_N}<q_{it}\leq\gamma_0)\right]u_2\\&=u_2'\mathbb{E}\left[\sum_{t=1}^{T}\breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{it}'1(\gamma_0+\tfrac{v}{\lambda_N}<q_{it}\leq\gamma_0)\right]u_2+o_p(1)\\&=O\left(\lambda_N^{-1}\right)+o_p(1)=o_p(1),\end{aligned}$$

the fifth term

$$\frac{1}{N^{1/2}}\sum_{i=1}^{N}\sum_{t=1}^{T}\delta_-\breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{it}'u_2 1(\gamma_0+\frac{v}{\lambda_N}<q_{it}\leq\gamma_0)=o_p(1)$$

by similar analysis as in the second term, and the last term is $o_p(1)$ is by the stochastic equicontinuity of $\gamma\mapsto\mathbb{G}_n\left(\sum_{t=1}^{T}\breve{\mathbf{x}}_{it}e_{1it}1(\gamma<q_{it}\leq\gamma_0)\right)$. ■

Lemma 6 *Under Assumption 4,*

$$\left(W_n(u),C_n(v)\right)\rightsquigarrow\left(W(u),C(v)\right),$$

where $W(u)=u'W$ with W defined in (27) and $C(v)$ is defined in (26). Furthermore, W and $C(v)$ are independent of each other.

Proof. First, for any $v\in[-\bar{v},0]$ with $0<\bar{v}<\infty$,

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\delta_N'\breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{it}'\delta_N 1(\gamma_0+\frac{v}{\lambda_N}<q_{it}\leq\gamma_0)\xrightarrow{p}c'Dc\cdot|v|.$$

Because

$$\begin{aligned}&N\cdot Var\left(\sum_{t=1}^{T}\delta_N'\breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{it}'\delta_N 1(\gamma_0+\tfrac{v}{\lambda_N}<q_{it}\leq\gamma_0)\right)\\&\leq N\cdot C\cdot\mathbb{E}\left[\|\delta_N\|^4\sum_{t=1}^{T}\|\breve{\mathbf{x}}_{it}\|^4 1(\gamma_0+\tfrac{v}{\lambda_N}<q_{it}\leq\gamma_0)\right]\\&=O\left(\tfrac{N\|\delta_N\|^4}{\lambda_N}\right)=O\left(\|\delta_N\|^2\right)=o(1),\end{aligned}$$

we have

$$\left|\sum_{i=1}^{N}\sum_{t=1}^{T}\delta_N'\breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{it}'\delta_N 1(\gamma_0+\frac{v}{\lambda_N}<q_{it}\leq\gamma_0)-c'\left[N^{-2\kappa}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}\left[\breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{it}'1(\gamma_0+\frac{v}{\lambda_N}<q_{it}\leq\gamma_0)\right]\right]c'\right|\xrightarrow{p}0,$$

where

$$N^{-2\kappa}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbb{E}\left[\breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{it}'1(\gamma_0+\frac{v}{\lambda_N}<q_{it}\leq\gamma_0)\right]=\lambda_N\frac{|v|}{\lambda_N}\sum_{t=1}^{T}\mathbb{E}\left[\breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{it}'|q_{it}=\overline{\gamma}\right]f_t\left(\overline{\gamma}\right)\xrightarrow{p}c'Dc\cdot|v|$$

by Condition (vi). By the the same arguments as in Lemma A.10 of Hansen (2000), the convergence is uniform over $[-\bar{v},0]$. Similarly, uniformly over $v\in[-\bar{v},0]$,

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\delta_{-}\breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{it}'\delta_N 1(\gamma_0+\frac{v}{\lambda_N}<q_{it}\leq\gamma_0)\xrightarrow{p}c_{-}'Dc\cdot|v|\,,$$

and uniformly over $v\in[0,\bar{v}]$,

$$\begin{array}{c}\sum_{i=1}^{N}\sum_{t=1}^{T}\delta_N'\breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{it}'\delta_N 1(\gamma_0<q_{it}\leq\gamma_0+\frac{v}{\lambda_N})\xrightarrow{p}c'Dc\cdot v,\\ \sum_{i=1}^{N}\sum_{t=1}^{T}\delta_{+}'\breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{it}'\delta_N 1(\gamma_0<q_{it}\leq\gamma_0+\frac{v}{\lambda_N})\xrightarrow{p}c_{+}'Dc\cdot v,\end{array}$$

We next analyze the random parts in $W_n(u)$ and $C_n(v)$ by applying Theorem 2.11.24 of VW. First, for $v_1<0$ and $v_2>0$, define

$$\begin{array}{c}S_{1i}=N^{-\kappa}c'\sum\limits_{t=1}^{T}\breve{\mathbf{x}}_{it}e_{1it}\Delta_i\left(v_1\right),S_{2i}=N^{-\kappa}c'\sum\limits_{t=1}^{T}\breve{\mathbf{x}}_{it}e_{2it}\Delta_i\left(v_2\right),\\ S_{3i}=\frac{1}{\sqrt{N}}\left(\sum\limits_{t=1}^{T}\breve{\mathbf{x}}_{it}e_{1it}1\left(q_{it}\leq\gamma_0\right),\sum\limits_{t=1}^{T}\breve{\mathbf{x}}_{it}e_{2it}1\left(q_{it}>\gamma_0\right)\right)',\end{array}$$

where $\Delta_i(v)=\Delta_i(\gamma_0+v/\lambda_n)=1(q_i\leq\gamma_0+v/\lambda_n)-1(q_i\leq\gamma_0)$ and S_{3i} is the asymptotic random component in $\underline{\widehat{\theta}}$. For a fixed v_1,

$$\begin{array}{c}\mathbb{E}\left[\left(\sum_{i=1}^{N}S_{1i}\right)^2\right]=\lambda_N\mathbb{E}\left[\left(\sum_{t=1}^{T}c'\breve{\mathbf{x}}_{it}e_{1it}1(\gamma_0+\frac{v_1}{\lambda_N}<q_{it}\leq\gamma_0)\right)^2\right]\\ \to c'\sum_{t=1}^{T}\mathbb{E}\left[\breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{it}'e_{1it}^2|q_{it}=\gamma_0\right]f_t\left(\gamma_0\right)c\cdot|v_1|=c'V_1c\cdot|v_1|\,,\end{array}$$

where the cross terms with $t\neq\tau$

$$\begin{array}{l}\lambda_N\mathbb{E}\left[c'\breve{\mathbf{x}}_{it}e_{1it}1(\gamma_0+\frac{v_1}{\lambda_N}<q_{it}\leq\gamma_0)c'\breve{\mathbf{x}}_{i\tau}e_{2i\tau}1(\gamma_0+\frac{v_1}{\lambda_N}<q_{i\tau}\leq\gamma_0)\right]\\ =\lambda_N\mathbb{E}\left[c'\breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{i\tau}'ce_{1it}e_{1i\tau}1(\gamma_0+\frac{v_1}{\lambda_N}<q_{it}\leq\gamma_0)1(\gamma_0+\frac{v_1}{\lambda_N}<q_{i\tau}\leq\gamma_0)\right]\\ \leq\|c\|^2\left(\mathbb{E}\left[\|\breve{\mathbf{x}}_{it}\|^4\right]\mathbb{E}\left[\|\breve{\mathbf{x}}_{i\tau}\|^4\right]\mathbb{E}\left[|e_{1it}|^4\right]\mathbb{E}\left[|e_{1i\tau}|^4\right]\right)^{1/4}\\ \quad\cdot\lambda_N P\left(\gamma_0+\frac{v_1}{\lambda_N}<q_{it}\leq\gamma_0,\gamma_0+\frac{v_1}{\lambda_N}<q_{i\tau}\leq\gamma_0\right)\to 0\end{array}\tag{30}$$

by (A.9) of Hansen (1999). Similarly, we can show

$$\mathbb{E}\left[\left(\sum_{i=1}^{N}S_{2i}\right)^2\right]\to c'V_2c\cdot v_2.$$

Also,

$$\begin{aligned}&\mathbb{E}\left[\left(\sum_{i=1}^{N}S_{1i}\right)\left(\sum_{i=1}^{N}S_{2i}\right)\right]=\sum_{i=1}^{N}\sum_{j=1}^{N}\mathbb{E}\left[S_{1i}S_{2j}\right]\\&=-N^{-2\kappa}\sum_{i\neq j}\sum_{t=1}^{T}\sum_{\tau=1}^{T}\mathbb{E}\left[c'\breve{\mathbf{x}}_{it}e_{1it}1(\gamma_0+\tfrac{v_1}{\lambda_N}<q_{it}\leq\gamma_0)\right]\mathbb{E}\left[c'\breve{\mathbf{x}}_{it}e_{2it}1(\gamma_0<q_{i\tau}\leq\gamma_0+\tfrac{v_2}{\lambda_N})\right]\\&-N^{-2\kappa}\sum_{i=1}^{N}\sum_{t\neq\tau}\mathbb{E}\left[c'\breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{i\tau}'ce_{1it}e_{2i\tau}1(\gamma_0+\tfrac{v_1}{\lambda_N}<q_{it}\leq\gamma_0)1(\gamma_0<q_{i\tau}\leq\gamma_0+\tfrac{v_2}{\lambda_N})\right]\\&\to 0,\end{aligned}$$

where the first term is zero and the second term converges to zero by similar arguments as in (30). Finally,

$$\sum_{i=1}^{N}S_{3i}\xrightarrow{p}W,$$

by a CLT, and

$$\begin{aligned}\mathbb{E}\left[\sum_{i=1}^{N}S_{1i}\sum_{i=1}^{N}S_{3i}\right]&=-N^{-\kappa-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{\tau=1}^{T}\begin{pmatrix}\mathbb{E}\left[c'\breve{\mathbf{x}}_{it}e_{1it}1(\gamma_0+\tfrac{v_1}{\lambda_N}<q_{it}\leq\gamma_0)\breve{\mathbf{x}}_{i\tau}e_{1i\tau}1\left(q_{i\tau}\leq\gamma_0\right)\right]\\\mathbb{E}\left[c'\breve{\mathbf{x}}_{it}e_{1it}1(\gamma_0+\tfrac{v_1}{\lambda_N}<q_{it}\leq\gamma_0)\breve{\mathbf{x}}_{i\tau}e_{2i\tau}1\left(q_{i\tau}>\gamma_0\right)\right]\end{pmatrix}\\&=O\left(NN^{-\kappa-1/2}/\lambda_N\right)=O\left(N^{\kappa-1/2}\right)=o\left(1\right);\end{aligned}$$

similarly, $\mathbb{E}\left[\sum_{i=1}^{N}S_{2i}\sum_{i=1}^{N}S_{3i}\right]\to 0$. In summary, the finite dimensional distributions of $\left(W_n\left(u\right),C_n\left(v\right)\right)$ matches that of $\left(W\left(u\right),C\left(v\right)\right)$.

Second, we show stochastic equicontinuity of the random part of $C_n\left(v\right)$ since stochastic equicontinuity of $W_n\left(u\right)$ is obvious. Take $v\leq 0$ as an example:

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\delta_N'\breve{\mathbf{x}}_{it}e_{1it}1\left(\gamma_0+\frac{v}{\lambda_N}<q_{it}\leq\gamma_0\right)=\mathbb{G}_N\left(T_{3N}\left(v\right)\right)$$

where $T_{3N}\left(v\right)=\sqrt{N}\sum_{t=1}^{T}\delta_N'\breve{\mathbf{x}}_{it}e_{1it}1\left(\gamma_0+\frac{v}{\lambda_N}<q_{it}\leq\gamma_0\right)$. Since $\{T_{3N}(v):-\infty<-\overline{v}\leq v\leq 0\}$ is a VC-subgraph for each N and the VC-index bounded by some constant independent of N (see, e.g., Example 2.11.26 of VW), the uniform-entropy condition holds. It remains to show condition (2.11.23):

$$\text{(i) }P^*F_N^2=O\left(1\right),\text{ (ii) }P^*F_N^21\left(F_N>\eta\sqrt{N}\right)\to 0,\ \forall\eta>0,$$

and

$$\text{(iii)}\sup_{|v_1-v_2|<\eta_N}P\left(T_{3N}(v_1)-T_{3N}(v_2)\right)^2\to 0,\ \forall\eta_N\downarrow 0,$$

where P^* signifies outer probability, F_N is the envelope function of $\{T_{3N}(v):-\infty<-\overline{v}\leq v\leq 0\}$ and can be taken as

$$F_N=\sqrt{N}\sum_{t=1}^{T}\left\|\delta_N\right\|\left\|\breve{\mathbf{x}}_{it}\right\|\left|e_{1it}\right|1(\gamma_0-\frac{\overline{v}}{\lambda_N}<q_{it}\leq\gamma_0).$$

(i) First, we have

$$\begin{aligned}P^*F_N^2&\leq N\sum_{t=1}^{T}\int_{\gamma_0-\overline{v}/\lambda_N}^{\gamma_0}\mathbb{E}\left[\left\|\delta_N\right\|^2\left\|\breve{\mathbf{x}}_{it}\right\|^2e_{1it}^2|q_{it}\right]f_t\left(q_{it}\right)dq+o\left(1\right)\\&\leq C\frac{N\left\|\delta_N\right\|^2}{\lambda_N}\sum_{t=1}^{T}\sup_{\gamma\in\mathcal{N}}\left\{f_t(\gamma)\mathbb{E}\left[\left\|\breve{\mathbf{x}}_{it}\right\|^2e_{1it}^2|q_{it}=\gamma\right]\right\}=O\left(1\right),\end{aligned}$$

where the cross term is $o(1)$ by (30), and $\mathcal{N}$ is a neighborhood of γ_0. (ii) Second,

$$\begin{aligned}
& P^* F_N^2 1\left(F_N > \eta\sqrt{N}\right) \\
\leq\ & N \sum_{t=1}^{T} \mathbb{E}\left[\|\delta_N\|^2 \|\check{\mathbf{x}}_{it}\|^2 e_{1it}^2 1\left(\gamma_0 - \frac{\overline{v}}{\lambda_N} < q_{it} \leq \gamma_0\right) 1\left(\sum_{t=1}^{T} \|\check{\mathbf{x}}_{it}\| |e_{1it}| > \frac{\eta}{\|\delta_N\|}\right)\right] + o(1) \\
\leq\ & CN \|\delta_N\|^2 \sum_{t=1}^{T} \mathbb{E}\left[\left(\|\check{\mathbf{x}}_{it}\| e_{1it}\right)^{2+\epsilon} 1\left(\gamma_0 - \frac{\overline{v}}{\lambda_N} < q_{it} \leq \gamma_0\right)\right] / \left(\frac{\eta}{\|\delta_N\|}\right)^{\epsilon} \\
\leq\ & C \frac{N \|\delta_N\|^2}{\lambda_N} \sum_{t=1}^{T} \sup_{\gamma \in \mathcal{N}} \left\{f_t(\gamma) \mathbb{E}\left[\left(\|\check{\mathbf{x}}_{it}\| e_{1it}\right)^{2+\epsilon} | q_{it} = \gamma\right]\right\} / \left(\frac{\eta}{\|\delta_N\|}\right)^{\epsilon} \\
\to\ & 0,
\end{aligned}$$

where the convergence is from Conditions (iv), (vi) and (viii). (iii) Finally, suppose $v_1 < v_2 < 0$,

$$\begin{aligned}
& \sup_{|v_1 - v_2| < \eta_N} \mathbb{E}\left(T_{3N}(v_1) - T_{3N}(v_2)\right)^2 \\
=\ & \sup_{|v_1 - v_2| < \eta_N} N \sum_{t=1}^{T} \mathbb{E}\left[\|\delta_N\|^2 \|\check{\mathbf{x}}_{it}\|^2 e_{1it}^2 1\left(\gamma_0 + \frac{v_1}{\lambda_N} < q_{it} \leq \gamma_0 + \frac{v_2}{\lambda_N}\right)\right] + o(1) \\
\leq\ & C \sup_{|v_1 - v_2| < \eta_N} \left\{|v_1 - v_2| \frac{N \|\delta_N\|^2}{\lambda_N} \sum_{t=1}^{T} \sup_{\gamma \in \mathcal{N}} \left\{f_t(\gamma) \mathbb{E}\left[\|\check{\mathbf{x}}_{it}\|^2 e_{1it}^2 | q = \gamma\right]\right\}\right\} \\
\to\ & 0, \forall \eta_N \downarrow 0.
\end{aligned}$$

■

Appendix D: Details of Calculation in Section 2

Static Panel Threshold Regression

When $\gamma \leq \gamma_0$,

$$\begin{aligned}
&\widetilde{y}_{it}^{-}(\gamma) = \sigma_{10} \widetilde{u}_{it}^{-}(\gamma) 1(q_{it} \leq \gamma), \\
&\widetilde{y}_{it}^{+}(\gamma) = \left[\alpha_{1i} - \tfrac{\sum_{\tau=1}^{T} \alpha_{1i} 1(\gamma < q_{i\tau} \leq \gamma_0) + \sum_{\tau=1}^{T} \alpha_{2i} 1(q_{i\tau} > \gamma_0)}{\sum_{\tau=1}^{T} 1(q_{i\tau} > \gamma)}\right] 1(\gamma < q_{it} \leq \gamma_0) \\
&\quad + \left[\alpha_{2i} - \tfrac{\sum_{\tau=1}^{T} \alpha_{1i} 1(\gamma < q_{i\tau} \leq \gamma_0) + \sum_{\tau=1}^{T} \alpha_{2i} 1(q_{i\tau} > \gamma_0)}{\sum_{\tau=1}^{T} 1(q_{i\tau} > \gamma)}\right] 1(q_{it} > \gamma_0) \\
&\quad + \left[\sigma_{10} u_{it} - \tfrac{\sum_{\tau=1}^{T} \sigma_{10} u_{i\tau} 1(\gamma < q_{i\tau} \leq \gamma_0) + \sum_{\tau=1}^{T} \sigma_{20} u_{i\tau} 1(q_{i\tau} > \gamma_0)}{\sum_{\tau=1}^{T} 1(q_{i\tau} > \gamma)}\right] 1(\gamma < q_{it} \leq \gamma_0) \\
&\quad + \left[\sigma_{20} u_{it} - \tfrac{\sum_{\tau=1}^{T} \sigma_{10} u_{i\tau} 1(\gamma < q_{i\tau} \leq \gamma_0) + \sum_{\tau=1}^{T} \sigma_{20} u_{i\tau} 1(q_{i\tau} > \gamma_0)}{\sum_{\tau=1}^{T} 1(q_{i\tau} > \gamma)}\right] 1(q_{it} > \gamma_0) \\
&= \left[\delta_\alpha - \delta_\alpha \tfrac{\sum_{\tau=1}^{T} 1(\gamma < q_{i\tau} \leq \gamma_0)}{\sum_{\tau=1}^{T} 1(q_{i\tau} > \gamma)}\right] 1(\gamma < q_{it} \leq \gamma_0) - \delta_\alpha \tfrac{\sum_{\tau=1}^{T} 1(\gamma < q_{i\tau} \leq \gamma_0)}{\sum_{\tau=1}^{T} 1(q_{i\tau} > \gamma)} 1(q_{it} > \gamma_0) \\
&\quad + \left[\sigma_{20} \widetilde{u}_{it}^{+}(\gamma) + \delta_\sigma u_{it} - \delta_\sigma \tfrac{\sum_{\tau=1}^{T} u_{i\tau} 1(\gamma < q_{i\tau} \leq \gamma_0)}{\sum_{\tau=1}^{T} 1(q_{i\tau} > \gamma)}\right] 1(\gamma < q_{it} \leq \gamma_0) \\
&\quad + \left[\sigma_{20} \widetilde{u}_{it}^{+}(\gamma) - \delta_\sigma \tfrac{\sum_{\tau=1}^{T} u_{i\tau} 1(\gamma < q_{i\tau} \leq \gamma_0)}{\sum_{\tau=1}^{T} 1(q_{i\tau} > \gamma)}\right] 1(q_{it} > \gamma_0) \\
&= \sigma_{20} \widetilde{u}_{it}^{+}(\gamma) 1(q_{it} > \gamma) + \delta_\alpha \left[1(\gamma < q_{it} \leq \gamma_0) - \tfrac{\sum_{\tau=1}^{T} 1(\gamma < q_{i\tau} \leq \gamma_0)}{\sum_{\tau=1}^{T} 1(q_{i\tau} > \gamma)}\right] 1(q_{it} > \gamma) \\
&\quad + \delta_\sigma \left[u_{it} 1(\gamma < q_{it} \leq \gamma_0) - \tfrac{\sum_{\tau=1}^{T} u_{i\tau} 1(\gamma < q_{i\tau} \leq \gamma_0)}{\sum_{\tau=1}^{T} 1(q_{i\tau} > \gamma)}\right] 1(q_{it} > \gamma) \\
&=: \sigma_{20} \widetilde{u}_{it}^{+}(\gamma) 1(q_{it} > \gamma) + \left[\delta_\alpha \widetilde{1}_{it}(\gamma, \gamma_0) + \delta_\sigma \widetilde{u}_{it}(\gamma, \gamma_0)\right] 1(q_{it} > \gamma).
\end{aligned}$$

So

$$
\begin{aligned}
S_n(\gamma) &= \sum_{i=1}^{N}\sum_{t=1}^{T}\left[\widetilde{y}_{it}^{-}(\gamma)^2 1(q_{it}\leq\gamma)+\widetilde{y}_{it}^{+}(\gamma)^2 1(q_{it}>\gamma)\right]\\
&= \sum_{i=1}^{N}\sum_{t=1}^{T}\left[\sigma_{10}^2\widetilde{u}_{it}^{-}(\gamma)^2 1(q_{it}\leq\gamma)+\sigma_{20}^2\widetilde{u}_{it}^{+}(\gamma)^2 1(q_{it}>\gamma)\right]\\
&\quad+\left\{\left[\delta_\alpha\widetilde{1}_{it}(\gamma,\gamma_0)+\delta_\sigma\widetilde{u}_{it}(\gamma,\gamma_0)\right]^2+2\sigma_{20}\left[\delta_\alpha\widetilde{1}_{it}(\gamma,\gamma_0)+\delta_\sigma\widetilde{u}_{it}(\gamma,\gamma_0)\right]\widetilde{u}_{it}^{+}(\gamma)\right\}1(q_{it}>\gamma),
\end{aligned}
$$

whose probability limit is

$$
\begin{aligned}
S(\gamma) &= \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\sigma_{10}^2\widetilde{u}_{it}^{-}(\gamma)^2 1(q_{it}\leq\gamma)+\sigma_{20}^2\widetilde{u}_{it}^{+}(\gamma)^2 1(q_{it}>\gamma)\right]\\
&\quad+\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\delta_\alpha^2\right]\mathbb{E}\left[\widetilde{1}_{it}(\gamma,\gamma_0)^2 1(q_{it}>\gamma)\right]\\
&\quad+\frac{1}{T}\sum_{t=1}^{T}\left[\delta_\sigma^2\mathbb{E}\left[\widetilde{u}_{it}(\gamma,\gamma_0)^2 1(q_{it}>\gamma)\right]+2\sigma_{20}\delta_\sigma\mathbb{E}\left[\widetilde{u}_{it}(\gamma,\gamma_0)\widetilde{u}_{it}^{+}(\gamma)1(q_{it}>\gamma)\right]\right].
\end{aligned}
$$

Simiarly, when $\gamma>\gamma_0$,

$$
\begin{aligned}
&\widetilde{y}_{it}^{-}(\gamma)=\sigma_{10}\widetilde{u}_{it}^{-}(\gamma)1(q_{it}\leq\gamma)-\left[\delta_\alpha\widetilde{1}_{it}(\gamma_0,\gamma)+\delta_\sigma\widetilde{u}_{it}(\gamma_0,\gamma)\right]1(q_{it}\leq\gamma),\\
&\widetilde{y}_{it}^{+}(\gamma)=\sigma_{20}\widetilde{u}_{it}^{+}(\gamma)1(q_{it}>\gamma),
\end{aligned}
$$

and

$$
\begin{aligned}
S_n(\gamma) &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\widetilde{y}_{it}^{-}(\gamma)^2 1(q_{it}\leq\gamma)+\widetilde{y}_{it}^{+}(\gamma)^2 1(q_{it}>\gamma)\right]\\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\sigma_{20}^2\widetilde{u}_{it}^{+}(\gamma)^2 1(q_{it}>\gamma)+\sigma_{10}^2\widetilde{u}_{it}^{-}(\gamma)^2 1(q_{it}\leq\gamma)\right]\\
&\quad+\left\{\left[\delta_\alpha\widetilde{1}_{it}(\gamma_0,\gamma)+\delta_\sigma\widetilde{u}_{it}(\gamma_0,\gamma)\right]^2-2\sigma_{10}\left[\delta_\alpha\widetilde{1}_{it}(\gamma_0,\gamma)+\delta_\sigma\widetilde{u}_{it}(\gamma_0,\gamma)\right]\widetilde{u}_{it}^{-}(\gamma)\right\}1(q_{it}\leq\gamma),
\end{aligned}
$$

whose probability limit is

$$
\begin{aligned}
S(\gamma) &= \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\sigma_{10}^2\widetilde{u}_{it}^{-}(\gamma)^2 1(q_{it}\leq\gamma)+\sigma_{20}^2\widetilde{u}_{it}^{+}(\gamma)^2 1(q_{it}>\gamma)\right]\\
&\quad+\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\delta_\alpha^2\right]\mathbb{E}\left[\widetilde{1}_{it}(\gamma_0,\gamma)^2 1(q_{it}\leq\gamma)\right]\\
&\quad+\frac{1}{T}\sum_{t=1}^{T}\left\{\delta_\sigma^2\mathbb{E}\left[\widetilde{u}_{it}(\gamma_0,\gamma)^2 1(q_{it}\leq\gamma)\right]-2\sigma_{10}\delta_\sigma\mathbb{E}\left[\widetilde{u}_{it}(\gamma_0,\gamma)\widetilde{u}_{it}^{-}(\gamma)1(q_{it}\leq\gamma)\right]\right\},
\end{aligned}
$$

where $\widetilde{1}_{it}(\gamma_0,\gamma)=1(\gamma_0<q_{it}\leq\gamma)-\frac{\sum_{\tau=1}^{T}1(\gamma_0<q_{i\tau}\leq\gamma)}{\sum_{\tau=1}^{T}1(q_{i\tau}\leq\gamma)}$, and $\widetilde{u}_{it}(\gamma_0,\gamma)=u_{it}1(\gamma_0<q_{it}\leq\gamma)-\frac{\sum_{\tau=1}^{T}u_{i\tau}1(\gamma_0<q_{i\tau}\leq\gamma)}{\sum_{\tau=1}^{T}1(q_{i\tau}\leq\gamma)}$. Note that in the objective function we condition on $D_i^{-}(\gamma)\geq 1$ and $D_i^{+}(\gamma)\geq 1$ in the summation $\sum_{t=1}^{T}\widetilde{y}_{it}^{-}(\gamma)^2 1(q_{it}\leq\gamma)$ and $\sum_{t=1}^{T}\widetilde{y}_{it}^{+}(\gamma)^2 1(q_{it}>\gamma)$, where $D_i^{-}(\gamma)=\sum_{t=1}^{T}1(q_{it}\leq\gamma)$ and $D_i^{+}(\gamma)=\sum_{t=1}^{T}1(q_{it}>\gamma)$.

So we first condition on $D_i^-(\gamma) \geq 1$ and $D_i^+(\gamma) \geq 1$ in calculating the expectations, and then multiply $P\left(D_i^-(\gamma) \geq 1\right) = 1 - \overline{F}(\gamma)^T = 1 - (1-\gamma)^T =: p^-(\gamma)$ and $P\left(D_i^+(\gamma) \geq 1\right) = 1 - F(\gamma)^T = 1 - \gamma^T =: p^+(\gamma)$, respectively. Hence,

$$
\begin{aligned}
&\sum_{t=1}^T \mathbb{E}\left[\widetilde{u}_{it}^-(\gamma)^2 1(q_{it} \leq \gamma) | D_i^-(\gamma) \geq 1\right] \\
&= \mathbb{E}\left[\sum_{t=1}^T u_{it}^2 1(q_{it} \leq \gamma) | D_i^-(\gamma) \geq 1\right] + \mathbb{E}\left[\sum_{t=1}^T 1(q_{it} \leq \gamma) \frac{\left(\sum_{\tau=1}^T u_{i\tau} 1(q_{i\tau} \leq \gamma)\right)^2}{\left(\sum_{\tau=1}^T 1(q_{i\tau} \leq \gamma)\right)^2} \right| \left. D_i^-(\gamma) \geq 1\right] \\
&\quad - 2\mathbb{E}\left[\frac{\sum_{\tau=1}^T u_{i\tau} 1(q_{i\tau} \leq \gamma)}{\sum_{\tau=1}^T 1(q_{i\tau} \leq \gamma)} \sum_{t=1}^T u_{it} 1(q_{it} \leq \gamma) \right| \left. D_i^-(\gamma) \geq 1\right] \\
&= \mathbb{E}\left[\sum_{t=1}^T 1(q_{it} \leq \gamma) | D_i^-(\gamma) \geq 1\right] - \mathbb{E}\left[\frac{\sum_{\tau=1}^T 1(q_{i\tau} \leq \gamma)}{\sum_{\tau=1}^T 1(q_{i\tau} \leq \gamma)} \right| \left. D_i^-(\gamma) \geq 1\right] = \frac{TF(\gamma)}{1-\overline{F}(\gamma)^T} - 1 = \frac{T\gamma}{1-(1-\gamma)^T} - 1,
\end{aligned}
$$

and similarly,

$$
\sum_{t=1}^T \mathbb{E}\left[\widetilde{u}_{it}^+(\gamma)^2 1(q_{it} > \gamma) | D_i^+(\gamma) \geq 1\right] = \frac{T\overline{F}(\gamma)}{1 - F(\gamma)^T} - 1 = \frac{T(1-\gamma)}{1-\gamma^T} - 1,
$$

where $\overline{F}(\gamma) = 1 - F(\gamma)$. When $\gamma \leq \gamma_0$,

$$
\begin{aligned}
&\sum_{t=1}^T \mathbb{E}\left[\widetilde{1}_{it}(\gamma, \gamma_0)^2 1(q_{it} > \gamma) | D_i^+(\gamma) \geq 1\right] \\
&= \mathbb{E}\left[\sum_{t=1}^T 1(\gamma < q_{it} \leq \gamma_0) | D_i^+(\gamma) \geq 1\right] + \mathbb{E}\left[\sum_{t=1}^T 1(q_{it} > \gamma) \frac{\left(\sum_{\tau=1}^T 1(\gamma < q_{i\tau} \leq \gamma_0)\right)^2}{\left(\sum_{\tau=1}^T 1(q_{i\tau} > \gamma)\right)^2} \right| \left. D_i^+(\gamma) \geq 1\right] \\
&\quad - 2\mathbb{E}\left[\frac{\sum_{\tau=1}^T 1(\gamma < q_{i\tau} \leq \gamma_0)}{\sum_{\tau=1}^T 1(q_{i\tau} > \gamma)} \sum_{t=1}^T 1(\gamma < q_{it} \leq \gamma_0) \right| \left. D_i^+(\gamma) \geq 1\right] \\
&= \overline{A}(\gamma, \gamma_0) - \mathbb{E}\left[\frac{\left(\sum_{\tau=1}^T 1(\gamma < q_{i\tau} \leq \gamma_0)\right)^2}{\sum_{\tau=1}^T 1(q_{i\tau} > \gamma)} \right| \left. D_i^+(\gamma) \geq 1\right] = \overline{A}(\gamma, \gamma_0) - B(\gamma, \gamma_0),
\end{aligned}
$$

$$
\begin{aligned}
&\sum_{t=1}^T \mathbb{E}\left[\widetilde{u}_{it}(\gamma, \gamma_0)^2 1(q_{it} > \gamma) | D_i^+(\gamma) \geq 1\right] \\
&= \mathbb{E}\left[\sum_{t=1}^T u_{it}^2 1(\gamma < q_{it} \leq \gamma_0) | D_i^+(\gamma) \geq 1\right] + \mathbb{E}\left[\sum_{t=1}^T 1(q_{it} > \gamma) \frac{\left(\sum_{\tau=1}^T u_{i\tau} 1(\gamma < q_{i\tau} \leq \gamma_0)\right)^2}{\left(\sum_{\tau=1}^T 1(q_{i\tau} > \gamma)\right)^2} \right| \left. D_i^+(\gamma) \geq 1\right] \\
&\quad - 2\mathbb{E}\left[\frac{\sum_{\tau=1}^T u_{i\tau} 1(\gamma < q_{i\tau} \leq \gamma_0)}{\sum_{\tau=1}^T 1(q_{i\tau} > \gamma)} \sum_{t=1}^T u_{it} 1(\gamma < q_{it} \leq \gamma_0) \right| \left. D_i^+(\gamma) \geq 1\right] \\
&= \overline{A}(\gamma, \gamma_0) - \mathbb{E}\left[\frac{\sum_{\tau=1}^T 1(\gamma < q_{i\tau} \leq \gamma_0)}{\sum_{\tau=1}^T 1(q_{i\tau} > \gamma)} \right| \left. D_i^+(\gamma) \geq 1\right] = \overline{A}(\gamma, \gamma_0) - A(\gamma, \gamma_0),
\end{aligned}
$$

and

$$
\begin{aligned}
&\sum_{t=1}^T \mathbb{E}\left[\widetilde{u}_{it}(\gamma, \gamma_0) \widetilde{u}_{it}^+(\gamma) 1(q_{it} > \gamma) | D_i^+(\gamma) \geq 1\right] \\
&= \mathbb{E}\left[\sum_{t=1}^T u_{it}^2 1(\gamma < q_{it} \leq \gamma_0) | D_i^+(\gamma) \geq 1\right] + \mathbb{E}\left[\sum_{t=1}^T 1(q_{it} > \gamma) \frac{\left(\sum_{\tau=1}^T u_{i\tau} 1(\gamma < q_{i\tau} \leq \gamma_0)\right)\left(\sum_{\tau=1}^T u_{i\tau} 1(q_{i\tau} > \gamma)\right)}{\left(\sum_{\tau=1}^T 1(q_{i\tau} > \gamma)\right)^2} \right| \left. D_i^+(\gamma) \geq 1\right] \\
&\quad - \mathbb{E}\left[\frac{\sum_{\tau=1}^T u_{i\tau} 1(q_{i\tau} > \gamma)}{\sum_{\tau=1}^T 1(q_{i\tau} > \gamma)} \sum_{t=1}^T u_{it} 1(\gamma < q_{it} \leq \gamma_0) \right| \left. D_i^+(\gamma) \geq 1\right] - \mathbb{E}\left[\frac{\sum_{\tau=1}^T u_{i\tau} 1(\gamma < q_{i\tau} \leq \gamma_0)}{\sum_{\tau=1}^T 1(q_{i\tau} > \gamma)} \sum_{t=1}^T u_{it} 1(q_{it} > \gamma) \right| \left. D_i^+(\gamma) \geq 1\right] \\
&= \overline{A}(\gamma, \gamma_0) - \mathbb{E}\left[\frac{\sum_{\tau=1}^T 1(\gamma < q_{i\tau} \leq \gamma_0)}{\sum_{\tau=1}^T 1(q_{i\tau} > \gamma)} \right| \left. D_i^+(\gamma) \geq 1\right] = \overline{A}(\gamma, \gamma_0) - A(\gamma, \gamma_0).
\end{aligned}
$$

Here

$$
\begin{aligned}
&\overline{A}(\gamma, \gamma_0) = \mathbb{E}\left[C_T(\gamma, \gamma_0) | D_i^+(\gamma) \geq 1\right] = \sum_{k=1}^T \mathbb{E}\left[C_T(\gamma, \gamma_0) | D_i^+(\gamma) = k\right] P\left(D_i^+(\gamma) = k | D_i^+(\gamma) \geq 1\right) \\
&= \sum_{k=1}^T k \frac{F(\gamma_0) - F(\gamma)}{1 - F(\gamma)} \begin{pmatrix} T \\ k \end{pmatrix} \frac{(1-F(\gamma))^k F(\gamma)^{T-k}}{1 - F(\gamma)^T} = \sum_{k=1}^T k \frac{\gamma_0 - \gamma}{1-\gamma} \begin{pmatrix} T \\ k \end{pmatrix} \frac{(1-\gamma)^k \gamma^{T-k}}{1-\gamma^T} \\
&= \frac{\gamma_0 - \gamma}{1-\gamma} \frac{T(1-\gamma)}{1-\gamma^T} = T \frac{\gamma_0 - \gamma}{1-\gamma^T},
\end{aligned}
$$

with $C_T(\gamma,\gamma_0)=\sum_{\tau=1}^{T}1(\gamma<q_{i\tau}\le\gamma_0)$ following the Binomial$\left(k,\frac{F(\gamma_0)-F(\gamma)}{1-F(\gamma)}\right)$ distribution given $D_i^+(\gamma)=k$, and $\sum_{k=1}^{T}k\begin{pmatrix}T\\k\end{pmatrix}(1-\gamma)^k\gamma^{T-k}=T(1-\gamma)$,

$$
\begin{aligned}
&A(\gamma,\gamma_0)=\mathbb{E}\left[\left.\tfrac{C_T(\gamma,\gamma_0)}{D_i^+(\gamma)}\right|D_i^+(\gamma)\ge 1\right]=\textstyle\sum_{k=1}^{T}\mathbb{E}\left[\left.\tfrac{C_T(\gamma,\gamma_0)}{k}\right|D_i^+(\gamma)=k\right]P\left(D_i^+(\gamma)=k|D_i^+(\gamma)\ge 1\right)\\
&=\textstyle\sum_{k=1}^{T}\tfrac{F(\gamma_0)-F(\gamma)}{1-F(\gamma)}\begin{pmatrix}T\\k\end{pmatrix}\tfrac{(1-F(\gamma))^kF(\gamma)^{T-k}}{1-F(\gamma)^T}=\sum_{k=1}^{T}\tfrac{\gamma_0-\gamma}{1-\gamma}\begin{pmatrix}T\\k\end{pmatrix}\tfrac{(1-\gamma)^k\gamma^{T-k}}{1-\gamma^T}\\
&=\tfrac{\gamma_0-\gamma}{1-\gamma}\tfrac{(1-\gamma+\gamma)^T-\gamma^T}{1-\gamma^T}=\tfrac{\gamma_0-\gamma}{1-\gamma}
\end{aligned}
$$

with $\sum_{k=1}^{T}\begin{pmatrix}T\\k\end{pmatrix}(1-\gamma)^k\gamma^{T-k}=1-\gamma^T$, and

$$
\begin{aligned}
&B(\gamma,\gamma_0)=\mathbb{E}\left[\left.\tfrac{C_T(\gamma,\gamma_0)^2}{D_i^+(\gamma)}\right|D_i^+(\gamma)\ge 1\right]=\textstyle\sum_{k=1}^{T}\mathbb{E}\left[\left.\tfrac{C_T(\gamma,\gamma_0)^2}{k}\right|D_i^+(\gamma)=k\right]P\left(D_i^+(\gamma)=k|D_i^+(\gamma)\ge 1\right)\\
&=\textstyle\sum_{k=1}^{T}\left[\tfrac{F(\gamma_0)-F(\gamma)}{1-F(\gamma)}\tfrac{1-F(\gamma_0)}{1-F(\gamma)}+k\left(\tfrac{F(\gamma_0)-F(\gamma)}{1-F(\gamma)}\right)^2\right]\begin{pmatrix}T\\k\end{pmatrix}\tfrac{(1-F(\gamma))^kF(\gamma)^{T-k}}{1-F(\gamma)^T}\\
&=\textstyle\sum_{k=1}^{T}\left[\tfrac{\gamma_0-\gamma}{1-\gamma}\tfrac{1-\gamma_0}{1-\gamma}+k\left(\tfrac{\gamma_0-\gamma}{1-\gamma}\right)^2\right]\begin{pmatrix}T\\k\end{pmatrix}\tfrac{(1-\gamma)^k\gamma^{T-k}}{1-\gamma^T}\\
&=\tfrac{\gamma_0-\gamma}{1-\gamma}\tfrac{1-\gamma_0}{1-\gamma}+\left(\tfrac{\gamma_0-\gamma}{1-\gamma}\right)^2\tfrac{T(1-\gamma)}{1-\gamma^T}=\tfrac{(\gamma_0-\gamma)(1-\gamma_0)}{(1-\gamma)^2}+\tfrac{T(\gamma_0-\gamma)^2}{(1-\gamma)(1-\gamma^T)}.
\end{aligned}
$$

Note also that the first terms (i.e., $k=1$) of $\overline{A}(\gamma,\gamma_0)$, $A(\gamma,\gamma_0)$ and $B(\gamma,\gamma_0)$ are the same as expected. Since $\sum_{t=1}^{T}\mathbb{E}\left[\widetilde{u}_{it}(\gamma,\gamma_0)^2 1(q_{it}>\gamma)|D_i^+(\gamma)\ge 1\right]=\sum_{t=1}^{T}\mathbb{E}\left[\widetilde{u}_{it}(\gamma,\gamma_0)\widetilde{u}_{it}^+(\gamma)1(q_{it}>\gamma)|D_i^+(\gamma)\ge 1\right]$, we can collect terms by noting that $\delta_\sigma^2+2\sigma_{20}\delta_\sigma=\delta_\sigma^2+2\delta_\sigma$. In sum, when $\gamma\le\gamma_0$,

$$
\begin{aligned}
S(\gamma) &= \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\sigma_{10}^2\widetilde{u}_{it}^-(\gamma)^2 1(q_{it}\le\gamma)+\sigma_{20}^2\widetilde{u}_{it}^+(\gamma)^2 1(q_{it}>\gamma)\right]\\
&\quad+\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\delta_\alpha^2\right]\mathbb{E}\left[\widetilde{1}_{it}(\gamma,\gamma_0)^2 1(q_{it}>\gamma)\right]\\
&\quad+\frac{1}{T}\sum_{t=1}^{T}\left[\delta_\sigma^2\mathbb{E}\left[\widetilde{u}_{it}(\gamma,\gamma_0)^2 1(q_{it}>\gamma)\right]+2\sigma_{20}\delta_\sigma\mathbb{E}\left[\widetilde{u}_{it}(\gamma,\gamma_0)\widetilde{u}_{it}^+(\gamma)1(q_{it}>\gamma)\right]\right]\\
&= \frac{1}{T}\left[\sigma_{10}^2\left(\frac{T\gamma}{1-(1-\gamma)^T}-1\right)p^-(\gamma)+\left(\frac{T(1-\gamma)}{1-\gamma^T}-1\right)p^+(\gamma)\right]\\
&\quad+\frac{\delta_\alpha^2}{T}\left[\overline{A}(\gamma,\gamma_0)-B(\gamma,\gamma_0)\right]p^+(\gamma)+\frac{\delta_\sigma^2+2\delta_\sigma}{T}\left[\overline{A}(\gamma,\gamma_0)-A(\gamma,\gamma_0)\right]p^+(\gamma)\\
&= :T_2(\gamma)+T_1(\gamma)+T_3(\gamma).
\end{aligned}
$$

Similarly, when $\gamma>\gamma_0$,

$$
\begin{aligned}
\textstyle\sum_{t=1}^{T}\mathbb{E}\left[\widetilde{1}_{it}(\gamma_0,\gamma)^2 1(q_{it}\le\gamma)|D_i^-(\gamma)\ge 1\right]&=\overline{A}(\gamma_0,\gamma)-B(\gamma_0,\gamma),\\
\textstyle\sum_{t=1}^{T}\mathbb{E}\left[\widetilde{u}_{it}(\gamma_0,\gamma)^2 1(q_{it}\le\gamma)|D_i^-(\gamma)\ge 1\right]&=\overline{A}(\gamma_0,\gamma)-A(\gamma_0,\gamma),
\end{aligned}
$$

and

$$
\sum_{t=1}^{T}\mathbb{E}\left[\widetilde{u}_{it}(\gamma_0,\gamma)\widetilde{u}_{it}^-(\gamma)1(q_{it}\le\gamma)|D_i^-(\gamma)\ge 1\right]=\overline{A}(\gamma_0,\gamma)-A(\gamma_0,\gamma),
$$

where

$$
\begin{aligned}
\overline{A}(\gamma_0,\gamma) &= \textstyle\sum_{k=1}^{T} k\frac{\gamma-\gamma_0}{\gamma}\begin{pmatrix} T \\ k \end{pmatrix}\frac{\gamma^k(1-\gamma)^{T-k}}{1-(1-\gamma)^T} = \frac{T(\gamma-\gamma_0)}{1-(1-\gamma)^T},\\
A(\gamma_0,\gamma) &= \textstyle\sum_{k=1}^{T} \frac{\gamma-\gamma_0}{\gamma}\begin{pmatrix} T \\ k \end{pmatrix}\frac{\gamma^k(1-\gamma)^{T-k}}{1-(1-\gamma)^T} = \frac{\gamma-\gamma_0}{\gamma},\\
B(\gamma_0,\gamma) &= \textstyle\sum_{k=1}^{T}\left[\frac{\gamma-\gamma_0}{\gamma}\frac{\gamma_0}{\gamma}+k\left(\frac{\gamma-\gamma_0}{\gamma}\right)^2\right]\begin{pmatrix} T \\ k \end{pmatrix}\frac{\gamma^k(1-\gamma)^{T-k}}{1-(1-\gamma)^T}\\
&\qquad = \frac{\gamma-\gamma_0}{\gamma}\frac{\gamma_0}{\gamma}+\left(\frac{\gamma-\gamma_0}{\gamma}\right)^2\frac{T\gamma}{1-(1-\gamma)^T} = \frac{\gamma_0(\gamma-\gamma_0)}{\gamma^2}+\frac{T(\gamma-\gamma_0)^2}{\gamma\left[1-(1-\gamma)^T\right]}.
\end{aligned}
$$

Hence

$$
\begin{aligned}
S(\gamma) &= \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\sigma_{10}^2\widetilde{u}_{it}^{-}(\gamma)^2 1(q_{it}\le\gamma)+\sigma_{20}^2\widetilde{u}_{it}^{+}(\gamma)^2 1(q_{it}>\gamma)\right]\\
&\quad+\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\delta_\alpha^2\right]\mathbb{E}\left[\widetilde{1}_{it}(\gamma_0,\gamma)^2 1(q_{it}\le\gamma)\right]\\
&\quad+\frac{1}{T}\sum_{t=1}^{T}\left\{\delta_\sigma^2\mathbb{E}\left[\widetilde{u}_{it}(\gamma_0,\gamma)^2 1(q_{it}\le\gamma)\right]-2\sigma_{10}\delta_\sigma\mathbb{E}\left[\widetilde{u}_{it}(\gamma_0,\gamma)\,\widetilde{u}_{it}^{-}(\gamma)\,1(q_{it}\le\gamma)\right]\right\}\\
&= \frac{1}{T}\left[\sigma_{10}^2\left(\frac{T\gamma}{1-(1-\gamma)^T}-1\right)p^-(\gamma)+\left(\frac{T(1-\gamma)}{1-\gamma^T}-1\right)p^+(\gamma)\right]\\
&\quad+\frac{\delta_\alpha^2}{T}\left[\overline{A}(\gamma_0,\gamma)-B(\gamma_0,\gamma)\right]p^-(\gamma)-\frac{\delta_\sigma^2+2\delta_\sigma}{T}\left[\overline{A}(\gamma_0,\gamma)-A(\gamma_0,\gamma)\right]p^-(\gamma)\\
&= : T_2(\gamma)+T_1(\gamma)+T_3(\gamma).
\end{aligned}
$$

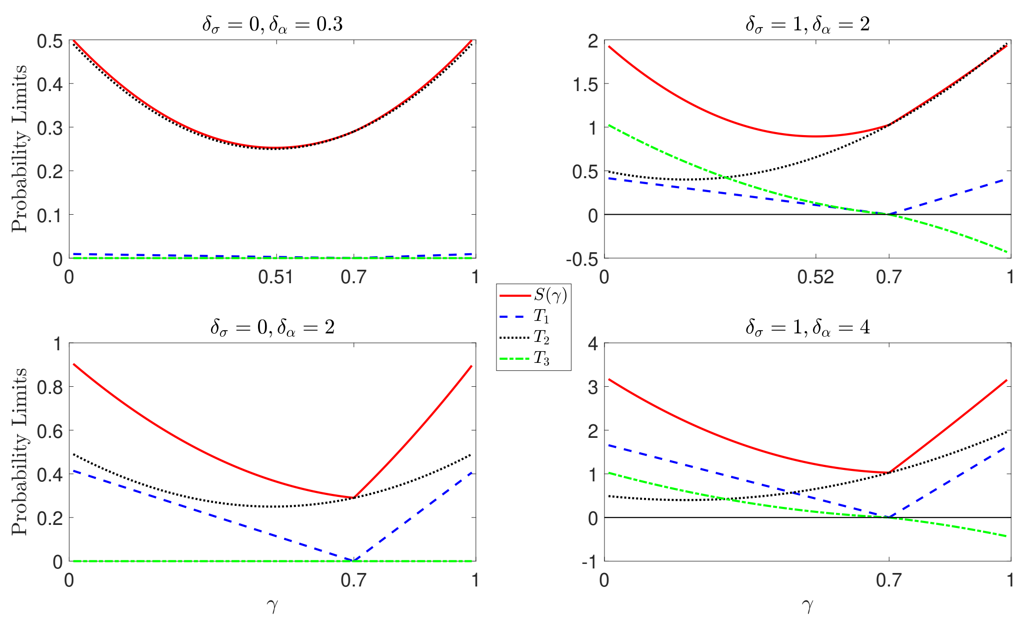

Figure 6: $S(\gamma)$ and its three components for various δ_σ and δ_α values: $\sigma_{20} = 1$, $\gamma_0 = 0.7$, $T = 2$

We next report the counterparts of Figure 1 when $T = 2$ or $\gamma_0 = 0.5$. Figure 6 shows the case with $T = 2$ and Figure 7 shows the case with $\gamma_0 = 0.5$. Obviously, the conclusions in the setting of the main text still apply here.

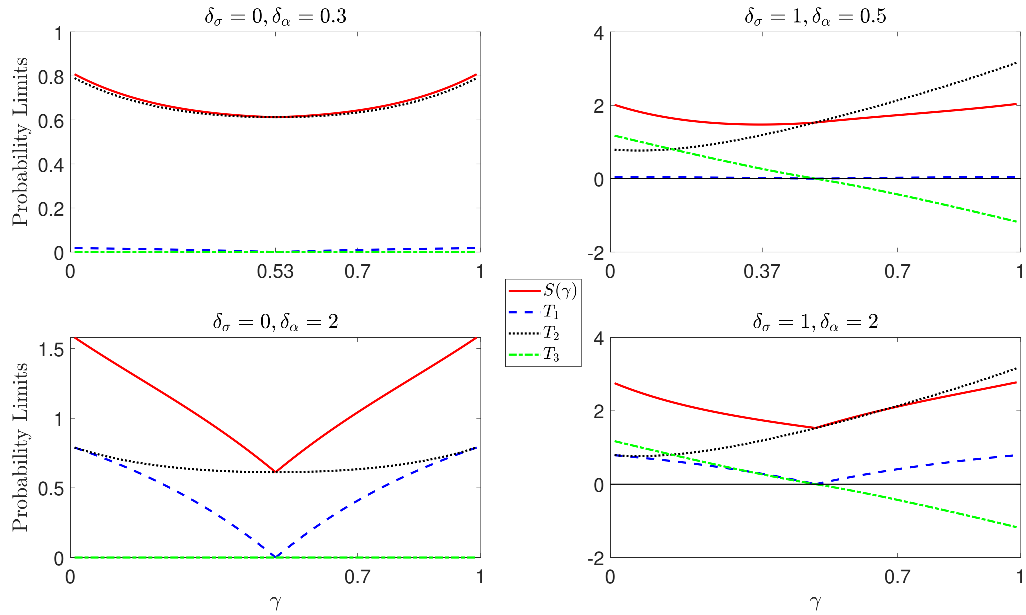

Figure 7: $S(\gamma)$ and its three components for various σ_{20} and δ_α values: $\sigma_{10} = 1$, $\gamma_0 = 0.5$, $T = 5$

When $T \to \infty$, both $p^+(\gamma)$ and $p^-(\gamma)$ converge to 1, $\frac{\overline{A}(\gamma,\gamma_0)}{T} \to \gamma_0 - \gamma$, $\frac{\overline{A}(\gamma_0,\gamma)}{T} \to \gamma - \gamma_0$, $\frac{B(\gamma,\gamma_0)}{T} \to \frac{(\gamma_0-\gamma)^2}{1-\gamma}$, $\frac{B(\gamma_0,\gamma)}{T} \to \frac{(\gamma-\gamma_0)^2}{\gamma}$, and both $\frac{A(\gamma,\gamma_0)}{T}$ and $\frac{A(\gamma_0,\gamma)}{T}$ converge to 0, so

$$
\begin{aligned}
S(\gamma) \to S_\infty(\gamma) &= \begin{cases} \sigma_{10}^2\gamma + (1-\gamma) + \delta_\alpha^2\left[(\gamma_0-\gamma) - \frac{(\gamma_0-\gamma)^2}{1-\gamma}\right] + \left(\delta_\sigma^2 + 2\delta_\sigma\right)(\gamma_0-\gamma), & \text{if } \gamma \le \gamma_0, \\ \sigma_{10}^2\gamma + (1-\gamma) + \delta_\alpha^2\left[(\gamma-\gamma_0) - \frac{(\gamma-\gamma_0)^2}{\gamma}\right] - \left(\delta_\sigma^2 + 2\delta_\sigma\right)(\gamma-\gamma_0), & \text{if } \gamma > \gamma_0, \end{cases} \\
&= C + \begin{cases} \delta_\alpha^2\left[(\gamma_0-\gamma) - \frac{(\gamma_0-\gamma)^2}{1-\gamma}\right], & \text{if } \gamma \le \gamma_0, \\ \delta_\alpha^2\left[(\gamma-\gamma_0) - \frac{(\gamma-\gamma_0)^2}{\gamma}\right], & \text{if } \gamma > \gamma_0, \end{cases} =: C + T_{1\infty}(\gamma),
\end{aligned}
$$

where the constant $C = 1 + \left(\delta_\sigma^2 + 2\delta_\sigma\right)\gamma_0$. As argued in the main text, $\arg\min_\gamma T_{1\infty}(\gamma) = \gamma_0$, so $\arg\min_\gamma S_\infty(\gamma) = \gamma_0$.

Panel Structural Change Models

We next consider panel structural change models where $q_{it} = q_t \sim U[0,1]$ and is fixed. Now, $P\left(D_i^-(\gamma) \ge 1\right) = P\left(D_i^+(\gamma) \ge 1\right) = 1$ by the assumption that γ must stay away from the boundary of $[0,1]$. So

$$
\begin{aligned}
&\textstyle\sum_{t=1}^T \mathbb{E}\left[\widetilde{u}_{it}^-(\gamma)^2 1(q_{it} \le \gamma)\right] = T\gamma - 1, \\
&\textstyle\sum_{t=1}^T \mathbb{E}\left[\widetilde{u}_{it}^+(\gamma)^2 1(q_{it} > \gamma)\right] = T(1-\gamma) - 1,
\end{aligned}
$$

and

$$
\begin{aligned}
&\overline{A}(\gamma,\gamma_0) = T(\gamma_0-\gamma), \overline{A}(\gamma_0,\gamma) = T(\gamma-\gamma_0), \\
&A(\gamma,\gamma_0) = \tfrac{\gamma_0-\gamma}{1-\gamma}, A(\gamma_0,\gamma) = \tfrac{\gamma-\gamma_0}{\gamma}, \\
&B(\gamma,\gamma_0) = \tfrac{T(\gamma_0-\gamma)^2}{1-\gamma}, B(\gamma_0,\gamma) = \tfrac{T(\gamma-\gamma_0)^2}{\gamma},
\end{aligned}
$$

which mimic the limit case in SPTR when $T \to \infty$. In sum, when $\gamma \le \gamma_0$,

$$\begin{aligned}S(\gamma) &= \tfrac{1}{T}\left[\sigma_{10}^2 (T\gamma - 1) + (T(1-\gamma)-1)\right] \\ &\quad + \tfrac{\delta_\alpha^2}{T}\left[T(\gamma_0-\gamma) - \tfrac{T(\gamma_0-\gamma)^2}{1-\gamma}\right] + \tfrac{\delta_\sigma^2+2\delta_\sigma}{T}\left[T(\gamma_0-\gamma) - \tfrac{\gamma_0-\gamma}{1-\gamma}\right] \\ &=: T_2(\gamma) + T_1(\gamma) + T_3(\gamma),\end{aligned}$$

and when $\gamma > \gamma_0$,

$$\begin{aligned}S(\gamma) &= \tfrac{1}{T}\left[\sigma_{10}^2 (T\gamma - 1) + (T(1-\gamma)-1)\right] \\ &\quad + \tfrac{\delta_\alpha^2}{T}\left[T(\gamma-\gamma_0) - \tfrac{T(\gamma-\gamma_0)^2}{\gamma}\right] - \tfrac{\delta_\sigma^2+2\delta_\sigma}{T}\left[T(\gamma-\gamma_0) - \tfrac{\gamma-\gamma_0}{\gamma}\right] \\ &=: T_2(\gamma) + T_1(\gamma) + T_3(\gamma).\end{aligned}$$

Figure 8 shows $S(\gamma)$ for the same specifications of δ_σ and δ_α as in Figure 1. When $\delta_\sigma = 0$, $T_2(\gamma)$ does not depend on γ, so $\arg\min_\gamma S(\gamma) = \gamma_0$. When $\delta_\sigma \neq 0$ and δ_α is small, $\arg\min_\gamma S(\gamma)$ is achieved at the boundary of Γ. When $T \to \infty$, $S(\gamma)$ has a same limit as in the PTR case, so $\arg\min_\gamma S_\infty(\gamma) = \gamma_0$ still holds.

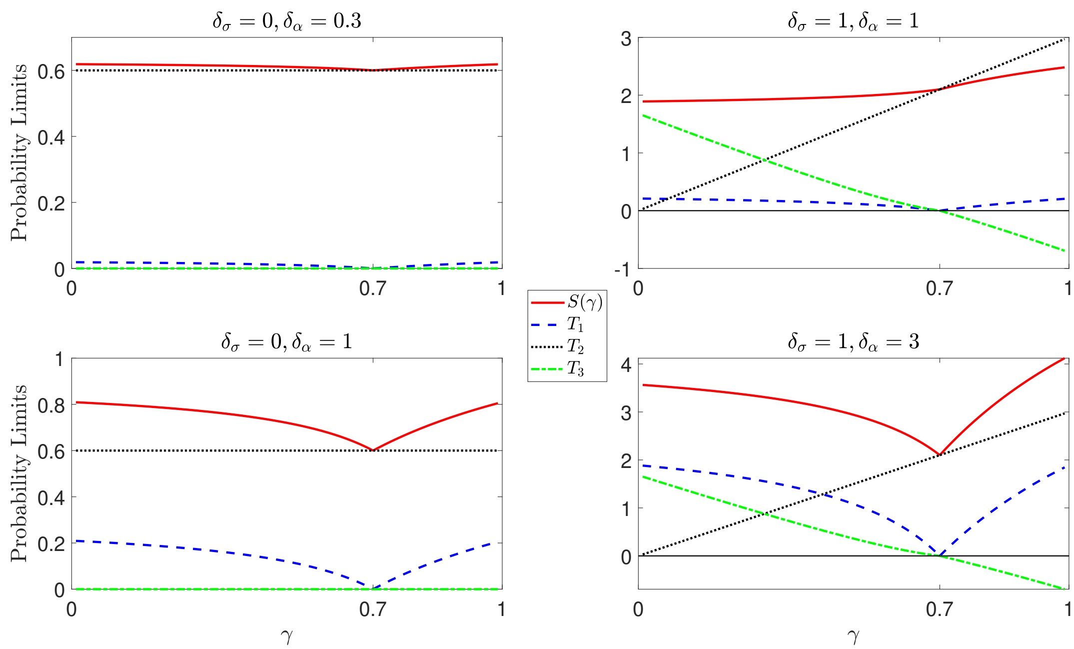

Figure 8: $S(\gamma)$ and its three components in the panel structural change model for various σ_{20} and δ_α values: $\sigma_{10} = 1$, $\gamma_0 = 0.3$, $T = 5$

Linear Panel Models

Finally, we consider the linear panel model in Section 2.2. There are many ways to eliminate α_i besides the usual demeaning operation. For example, let

$$\widetilde{y}_{it}(\beta) \quad : \quad = \left[y_{it} - \frac{\sum_{\tau=1}^T y_{i\tau} 1(x_{i\tau} \le \beta)}{\sum_{\tau=1}^T 1(x_{i\tau} \le \beta)}\right] 1(x_{it} \le \beta) + \left[y_{it} - \frac{\sum_{\tau=1}^T y_{i\tau} 1(x_{i\tau} > \gamma)}{\sum_{\tau=1}^T 1(x_{i\tau} > \gamma)}\right] 1(x_{it} > \beta)$$

$$= \; : \widetilde{y}_{it}^{-}(\beta) 1(x_{it} \leq \beta) + \widetilde{y}_{it}^{+}(\beta) 1(x_{it} > \beta),$$

which tries to mimic the WRD operation in the last subsection. This demeaning can eliminate α_i because α_i does not change over t. For any value of $\beta \in \mathcal{B} = [\epsilon, 1-\epsilon]$ with $\epsilon \in (0, 0.5)$,

$$\widetilde{y}_{it}(\beta) = \widetilde{x}_{it}^{-}(\beta)\beta_0 1(x_{it} \leq \beta) + \widetilde{x}_{it}^{+}(\beta)\beta_0 1(x_{it} > \beta) + \widetilde{u}_{it}^{-}(\beta) 1(x_{it} \leq \beta) + \widetilde{u}_{it}^{+}(\beta) 1(x_{it} > \beta).$$

So only when $\beta = \beta_0$

$$\mathbb{E}[\widetilde{y}_{it}(\beta)|\mathbf{X}_i] = \widetilde{x}_{it}^{-}(\beta)\beta 1(x_{it} \leq \beta) + \widetilde{x}_{it}^{+}(\beta)\beta 1(x_{it} > \beta) \tag{31}$$

with

$$\widetilde{y}_{it}(\beta_0) - \mathbb{E}[\widetilde{y}_{it}(\beta_0)|\mathbf{X}_i] = \widetilde{u}_{it}^{-}(\beta_0) 1(x_{it} \leq \beta_0) + \widetilde{u}_{it}^{+}(\beta_0) 1(x_{it} > \beta_0),$$

do we expect least squares based on the objective function

$$S_n(\beta) = \sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\left[\widetilde{y}_{it}^{-}(\beta) - \widetilde{x}_{it}^{-}(\beta)\beta\right]^2 1(x_{it} \leq \beta) + \left[\widetilde{y}_{it}^{+}(\gamma) - \widetilde{x}_{it}^{+}(\beta)\beta\right]^2 1(x_{it} > \beta)\right\}$$

to deliver a consistent estimate of β, where $\mathbf{X}_i = (x_{i1}, \cdots, x_{iT})'$, $\widetilde{x}_{it}^{\pm}(\beta)$ and $\widetilde{u}_{it}^{\pm}(\beta)$ are similarly defined as $\widetilde{y}_{it}^{\pm}(\beta)$, and β_0 is the true value of β.

The objective function of least squares

$$S_n(\beta) = \sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\left[\widetilde{x}_{it}^{-}(\beta)(\beta_0 - \beta) + \widetilde{u}_{it}^{-}(\beta)\right]^2 1(x_{it} \leq \beta) + \left[\widetilde{x}_{it}^{+}(\beta)(\beta_0 - \beta) + \widetilde{u}_{it}^{+}(\beta)\right]^2 1(x_{it} > \beta)\right\},$$

so the probability limit of $S_n(\beta)/n$ is

$$\begin{aligned} S(\beta) &= \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\widetilde{x}_{it}^{-}(\beta)^2 1(x_{it} \leq \beta)\right](\beta_0 - \beta)^2 + \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\widetilde{x}_{it}^{+}(\beta)^2 1(x_{it} > \beta)\right](\beta_0 - \beta)^2 \\ &\quad + \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\widetilde{u}_{it}^{-}(\beta)^2 1(x_{it} \leq \beta)\right] + \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\widetilde{u}_{it}^{+}(\beta)^2 1(x_{it} \leq \beta)\right]. \end{aligned}$$

As in the example of Section 2.1, we actually condition on $D_i^{-}(\beta) \geq 1$ and $D_i^{+}(\beta) \geq 1$ in the terms related to $\widetilde{y}_{it}^{-}(\beta)$ and $\widetilde{y}_{it}^{+}(\beta)$ respectively, where $D_i^{-}(\beta) = \sum_{t=1}^{T} 1(x_{it} \leq \beta)$, and $D_i^{+}(\beta) = \sum_{t=1}^{T} 1(x_{it} > \beta)$.

Now,

$$\begin{aligned} &\textstyle\sum_{t=1}^{T}\mathbb{E}\left[\widetilde{x}_{it}^{-}(\beta)^2 1(x_{it} \leq \beta)|D_i^{-}(\beta) \geq 1\right] \\ &= \textstyle\sum_{t=1}^{T}\mathbb{E}\left[x_{it}^2 1(x_{it} \leq \beta)|D_i^{-}(\beta) \geq 1\right] + \mathbb{E}\left[D_i^{-}(\beta)\left(\frac{\sum_{\tau=1}^{T} x_{i\tau}1(x_{i\tau} \leq \beta)}{D_i^{-}(\beta)}\right)^2 \middle| D_i^{-}(\beta) \geq 1\right] \\ &\quad - 2\mathbb{E}\left[\textstyle\sum_{t=1}^{T} x_{it}1(x_{it} \leq \beta)\frac{\sum_{\tau=1}^{T} x_{i\tau}1(x_{i\tau} \leq \beta)}{D_i^{-}(\beta)} \middle| D_i^{-}(\beta) \geq 1\right] \\ &= \mathbb{E}\left[\textstyle\sum_{t=1}^{T} x_{it}^2 1(x_{it} \leq \beta)|D_i^{-}(\beta) \geq 1\right] - \mathbb{E}\left[\frac{\left(\sum_{t=1}^{T} x_{it}1(x_{it} \leq \beta)\right)^2}{D_i^{-}(\beta)} \middle| D_i^{-}(\beta) \geq 1\right] \\ &= \textstyle\sum_{k=1}^{T}\mathbb{E}\left[\sum_{t=1}^{T} x_{it}^2 1(x_{it} \leq \beta)|D_i^{-}(\beta) = k\right] P\left(D_i^{-}(\beta) = k|D_i^{-}(\beta) \geq 1\right) \\ &\quad - \textstyle\sum_{k=1}^{T}\mathbb{E}\left[\left(\sum_{t=1}^{T} x_{it}1(x_{it} \leq \beta)\right)^2 \middle| D_i^{-}(\beta) = k\right]\frac{P\left(D_i^{-}(\beta)=k|D_i^{-}(\beta) \geq 1\right)}{k} \\ &= \textstyle\sum_{k=1}^{T} k\frac{\beta^3}{3}\begin{pmatrix} T \\ k \end{pmatrix}\frac{\beta^k(1-\beta)^{T-k}}{1-(1-\beta)^T} - \sum_{k=1}^{T}\left[k\frac{\beta^3}{3} + k(k-1)\left(\frac{\beta^2}{2}\right)^2\right]\frac{1}{k}\begin{pmatrix} T \\ k \end{pmatrix}\frac{\beta^k(1-\beta)^{T-k}}{1-(1-\beta)^T} \\ &= \frac{\beta^3}{3}\frac{T\beta}{1-(1-\beta)^T} - \frac{\beta^3}{3} - \frac{\beta^4}{4}\left[\frac{T\beta}{1-(1-\beta)^T} - 1\right] =: A^{-}(\beta), \end{aligned}$$

and

$$\sum_{t=1}^{T}\mathbb{E}\left[\widetilde{u}_{it}^{-}(\beta)^{2}1(x_{it}\leq\beta)|D_{i}^{-}(\beta)\geq 1\right]=\frac{T\beta}{1-(1-\beta)^{T}}-1;$$

similarly,

$$\begin{aligned}
&\textstyle\sum_{t=1}^{T}\mathbb{E}\left[\widetilde{x}_{it}^{+}(\beta)^{2}1(x_{it}>\beta)|D_{i}^{+}(\beta)\geq 1\right]\\
&=\textstyle\sum_{k=1}^{T}k\frac{1-\beta^{3}}{3}\left(\begin{array}{c}T\\k\end{array}\right)\frac{(1-\beta)^{k}\beta^{T-k}}{1-\beta^{T}}-\sum_{k=1}^{T}\left[k\frac{1-\beta^{3}}{3}+k(k-1)\left(\frac{1-\beta^{2}}{2}\right)^{2}\right]\frac{1}{k}\left(\begin{array}{c}T\\k\end{array}\right)\frac{(1-\beta)^{k}\beta^{T-k}}{1-\beta^{T}}\\
&=\frac{1-\beta^{3}}{3}\frac{T(1-\beta)}{1-\beta^{T}}-\frac{1-\beta^{3}}{3}-\frac{\left(1-\beta^{2}\right)^{2}}{4}\left[\frac{T(1-\beta)}{1-\beta^{T}}-1\right]=:A^{+}(\beta),
\end{aligned}$$

and

$$\sum_{t=1}^{T}\mathbb{E}\left[\widetilde{u}_{it}^{+}(\beta)^{2}1(x_{it}>\beta)|D_{i}^{+}(\beta)\geq 1\right]=\frac{T(1-\beta)}{1-\beta^{T}}-1.$$

In summary,

$$\begin{aligned}
S(\beta)&=\tfrac{1}{T}\left[A^{-}(\beta)p^{-}(\beta)+A^{+}(\beta)p^{+}(\beta)\right](\beta_{0}-\beta)^{2}+\tfrac{1}{T}\left[\left(\tfrac{T\beta}{1-(1-\beta)^{T}}-1\right)p^{-}(\beta)+\left(\tfrac{T(1-\beta)}{1-\beta^{T}}-1\right)p^{+}(\beta)\right]\\
&=\tfrac{1}{T}\left[\begin{array}{c}\frac{\beta^{3}}{3}T\beta-\frac{\beta^{3}}{3}p^{-}(\beta)-\frac{\beta^{4}}{4}\left(T\beta-p^{-}(\beta)\right)\\+\frac{1-\beta^{3}}{3}T(1-\beta)-\frac{1-\beta^{3}}{3}p^{+}(\beta)-\frac{\left(1-\beta^{2}\right)^{2}}{4}\left(T(1-\beta)-p^{+}(\beta)\right)\end{array}\right](\beta_{0}-\beta)^{2}+1-\tfrac{p^{-}(\beta)+p^{+}(\beta)}{T}\\
&=\left[\begin{array}{c}\frac{\beta^{4}}{3}-\frac{\beta^{5}}{4}-\left(\frac{\beta^{3}}{3}-\frac{\beta^{4}}{4}\right)\frac{p^{-}(\beta)}{T}\\+\frac{\left(1-\beta^{3}\right)(1-\beta)}{3}-\frac{\left(1-\beta^{2}\right)^{2}(1-\beta)}{4}-\left(\frac{1-\beta^{3}}{3}-\frac{\left(1-\beta^{2}\right)^{2}}{4}\right)\frac{p^{+}(\beta)}{T}\end{array}\right](\beta_{0}-\beta)^{2}+1-\tfrac{p^{-}(\beta)+p^{+}(\beta)}{T}\\
&=:D(\beta)(\beta_{0}-\beta)^{2}+T_{2}(\beta),
\end{aligned}$$

where $p^{-}(\beta)=1-(1-\beta)^{T}$, and $p^{+}(\beta)=1-\beta^{T}$. $T_{2}(\beta)$ is generated from $\widetilde{u}_{it}^{\pm}(\beta)$, and this is similar to $T_{2}(\gamma)$ in the last subsection where we show that $T_{2}(\gamma)$ is generated from $\widetilde{u}_{it}^{\pm}(\gamma)$. Note that neither $\widetilde{u}_{it}^{\pm}(\beta)$ nor $\widetilde{u}_{it}^{\pm}(\gamma)$ is a part of the original model, and they are actually generated by the demeaning operation . It is only because the demeaning operation depends on β or γ that we have the extra term $T_{2}(\beta)$ or $T_{2}(\gamma)$, which biases the least squares estimation. Of course, we do not have the T_{3} term here because there is no threshold effect in the error variance, but the T_{2} term alone is enough to make our estimator inconsistent in both models.

When $T\to\infty$,

$$S(\beta)\to\left[\tfrac{\beta^{4}}{3}-\tfrac{\beta^{5}}{4}+\tfrac{\left(1-\beta^{3}\right)(1-\beta)}{3}-\tfrac{\left(1-\beta^{2}\right)^{2}(1-\beta)}{4}\right](\beta_{0}-\beta)^{2}+1=:C(\beta)(\beta_{0}-\beta)^{2}+1=:S_{\infty}(\beta),$$

where $C(\beta)=\lim_{T\to\infty}D(\beta)$, and $T_{2}(\beta)\to 1$. Since

$$\frac{dS_{\infty}(\beta)}{d\beta}=C'(\beta)(\beta_{0}-\beta)^{2}+2C(\beta)(\beta-\beta_{0}),$$

which achieves zero at $\beta=\beta_{0}$, we expect $\arg\min_{\beta}S_{\infty}(\beta)=\beta_{0}$, which is justified in the right panel of Figure 3.

If the demeaning does not depend on β but on a constant $c\in\mathcal{B}$, i.e., let $\beta=c$ in $\widetilde{y}_{it}(\beta)$, then least squares is based on

$$\mathbb{E}\left[\widetilde{y}_{it}(c)|\mathbf{X}_{i}\right]=\widetilde{x}_{it}^{-}(c)\beta 1(x_{it}\leq c)+\widetilde{x}_{it}^{+}(c)\beta 1(x_{it}>c),$$

and

$$S_n^c(\beta)=\sum_{i=1}^N\sum_{t=1}^T\left[\left(\widetilde{y}_{it}^-(c)-\widetilde{x}_{it}^-(c)\beta\right)^2 1(x_{it}\leq c)+\left(\widetilde{y}_{it}^+(c)-\widetilde{x}_{it}^+(c)\beta\right)^2 1(x_{it}>c)\right].$$

The probability limit of the objective function as $N\to\infty$ is

$$\begin{aligned}S^c(\beta)&=\tfrac{1}{T}\textstyle\sum_{t=1}^T\mathbb{E}\left[\widetilde{x}_{it}^-(c)^2 1(x_{it}\leq c)\right](\beta_0-\beta)^2+\tfrac{1}{T}\sum_{t=1}^T\mathbb{E}\left[\widetilde{x}_{it}^+(c)^2 1(x_{it}>c)\right](\beta_0-\beta)^2\\&\quad+\tfrac{1}{T}\textstyle\sum_{t=1}^T\mathbb{E}\left[\widetilde{u}_{it}^-(c)^2 1(x_{it}\leq c)\right]+\tfrac{1}{T}\sum_{t=1}^T\mathbb{E}\left[\widetilde{u}_{it}^+(c)^2 1(x_{it}\leq c)\right]\\&=\left[\begin{array}{c}\frac{c^4}{3}-\frac{c^5}{4}-\left(\frac{c^3}{3}-\frac{c^4}{4}\right)\frac{p^-(c)}{T}\\+\frac{(1-c^3)(1-c)}{3}-\frac{(1-c^2)^2(1-c)}{4}-\left(\frac{1-c^3}{3}-\frac{(1-c^2)^2}{4}\right)\frac{p^+(c)}{T}\end{array}\right](\beta_0-\beta)^2+1-\frac{p^-(c)+p^+(c)}{T}\\&=:D(c)(\beta-\beta_0)^2+T_2(c).\end{aligned}$$

Because the coefficient before $(\beta_0-\beta)^2$ is positive, $\arg\min_\beta S^c(\beta)=\beta_0$, as shown in the right panel of Figure 3.

Appendix E: Details of CI Construction for γ and β

This appendix provides more extensive discussion of the CI construction for γ and β. First, we give the LR_{1n} statistics more detailed analysis. Second, we provide algorithms for the CI construction based on the LR_n, LR_{1n} and LR_{2n} statistics.

Analysis of the LR_{1n} Statistics

First, the LR_{1n} statistics can be explicitly expressed as

$$\begin{aligned}LR_{1n}(\gamma)&=\tfrac{S_n\left(\gamma,\underline{\widehat{\theta}}(\gamma)\right)-S_n\left(\widehat{\gamma}\left(\underline{\widehat{\theta}}(\gamma)\right),\underline{\widehat{\theta}}(\gamma)\right)}{\widehat{\eta}^2},\\LR_{1n}(\gamma,\beta_{11})&=\tfrac{S_n\left(\gamma,\beta_{11},\underline{\widehat{\theta}}_{-11}(\gamma,\beta_{11})\right)-S_n\left(\widehat{\gamma}\left(\underline{\widehat{\theta}}_{-11}(\gamma,\beta_{11})\right),\widehat{\beta}_{11}\left(\underline{\widehat{\theta}}_{-11}(\gamma,\beta_{11})\right),\underline{\widehat{\theta}}_{-11}(\gamma,\beta_{11})\right)}{\widehat{\eta}^2},\end{aligned}$$

where $\widehat{\gamma}\left(\underline{\widehat{\theta}}(\gamma)\right)$ is the threshold estimator when $\underline{\theta}$ is set at $\underline{\widehat{\theta}}(\gamma)$, and $\left(\widehat{\gamma}\left(\underline{\widehat{\theta}}_{-11}(\gamma,\beta_{11})\right),\widehat{\beta}_{11}\left(\underline{\widehat{\theta}}_{-11}(\gamma,\beta_{11})\right)\right)$ are defined by a similar procedure. Specifically, given $r\in\Gamma$, the concentrated $\widehat{\beta}_{11}\left(\underline{\widehat{\theta}}_{-11}(\gamma,\beta_{11})\right)$ is

$$\widehat{\beta}_{11}\left(r,\underline{\widehat{\theta}}_{-11}(\gamma,\beta_{11})\right)=\frac{\sum_{i=1}^N\sum_{t=1}^T x_{it}^1\left(y_{it}-\check{\mathbf{x}}_{it}^{-1\prime}\widehat{\theta}_{12}(\gamma,\beta_{11})\right)1(q_{it}\leq r)}{\sum_{i=1}^N\sum_{t=1}^T\left(x_{it}^1\right)^2 1(q_{it}\leq r)},$$ [19]

and we then search r over Γ to minimize

$$S_n(r):=\sum_{i=1}^N\sum_{t=1}^T\left[y_{it}-\left[x_{it}^1\widehat{\beta}_{11}\left(r,\underline{\widehat{\theta}}_{-11}(\gamma,\beta_{11})\right)+\check{\mathbf{x}}_{it}^{-1\prime}\widehat{\theta}_{12}(\gamma,\beta_{11})\right]1(q_{it}\leq r)-\check{\mathbf{x}}_{it}'\widehat{\theta}_2(\gamma)1(q_{it}>r)\right]^2;$$

the minimizer $\widehat{r}$ is $\widehat{\gamma}\left(\underline{\widehat{\theta}}_{-11}(\gamma,\beta_{11})\right)$, and $\widehat{\beta}_{11}\left(\underline{\widehat{\theta}}_{-11}(\gamma,\beta_{11})\right)=\widehat{\beta}_{11}\left(\widehat{r},\underline{\widehat{\theta}}_{-11}(\gamma,\beta_{11})\right)$.

Theorem 8 *Under Assumption 4,*

$$LR_{1n}(\gamma_0)\xrightarrow{d}\xi(\varphi,\phi),$$

[19] Although $\widehat{\beta}_{11}\left(r,\underline{\widehat{\theta}}_{-11}(\gamma,\beta_{11})\right)$ depends only on $\widehat{\theta}_{12}(\gamma,\beta_{11})$ rather than $\underline{\widehat{\theta}}_{-11}(\gamma,\beta_{11})$, we maintain this notation to avoid confusion.

and

$$LR_{1n}\left(\gamma_0, \beta_{11}^0\right) \xrightarrow{d} \frac{1}{2}\frac{\widetilde{S}_{\beta_{11}\beta_{11}}}{S_{\beta_{11}\beta_{11}}}\varrho_{11}\chi_1^2 + \xi(\varphi,\phi),$$

where ϱ_{11}, χ_1^2 and $\xi(\varphi,\phi)$ have the same definitions as in Theorem 6.

Proof. By the CMT,

$$\begin{aligned}
& S_n\left(\gamma_0, \widehat{\underline{\theta}}(\gamma_0)\right) - S_n\left(\widehat{\gamma}\left(\widehat{\underline{\theta}}(\gamma_0)\right), \widehat{\underline{\theta}}(\gamma_0)\right) \\
&= \left(S_n\left(\gamma_0, \widehat{\underline{\theta}}(\gamma_0)\right) - S_n\left(\gamma_0, \underline{\theta}_0\right)\right) - \left(S_n\left(\widehat{\gamma}\left(\widehat{\underline{\theta}}(\gamma_0)\right), \widehat{\underline{\theta}}(\gamma_0)\right) - S_n\left(\gamma_0, \underline{\theta}_0\right)\right) \\
&\xrightarrow{d} \tfrac{1}{2}\widehat{u}'S_{\underline{\theta\theta}}\widehat{u} - \widehat{u}'W - \min_v \left\{\tfrac{1}{2}\widehat{u}'S_{\underline{\theta\theta}}\widehat{u} - \widehat{u}'W + C(v)\right\} = \max_v \ \{-C(v)\} \overset{d}{=} \eta^2\xi(\varphi,\phi),
\end{aligned}$$

which is the same as in $S_n(\gamma_0) - S_n(\widehat{\gamma})$, and the rest of the proof is the same as in the proof of Theorem 5, where $\widehat{u} = \arg\min_u \left\{\frac{1}{2}u'S_{\underline{\theta\theta}}u - u'W\right\}$. Next,

$$\begin{aligned}
& S_n\left(\gamma_0,, \beta_{11}^0, \widehat{\underline{\theta}}_{-11}\left(\gamma_0,, \beta_{11}^0\right)\right) - S_n\left(\widehat{\gamma}\left(\widehat{\underline{\theta}}_{-11}\left(\gamma_0,, \beta_{11}^0\right)\right), \widehat{\beta}_{11}\left(\widehat{\underline{\theta}}_{-11}\left(\gamma_0,, \beta_{11}^0\right)\right), \widehat{\underline{\theta}}_{-11}\left(\gamma_0,, \beta_{11}^0\right)\right) \\
= \ & \left(S_n\left(\gamma_0,, \beta_{11}^0, \widehat{\underline{\theta}}_{-11}\left(\gamma_0,, \beta_{11}^0\right)\right) - S_n\left(\gamma_0, \underline{\theta}_0\right)\right) \\
& - \left(S_n\left(\widehat{\gamma}\left(\widehat{\underline{\theta}}_{-11}\left(\gamma_0,, \beta_{11}^0\right)\right), \widehat{\beta}_{11}\left(\widehat{\underline{\theta}}_{-11}\left(\gamma_0,, \beta_{11}^0\right)\right), \widehat{\underline{\theta}}_{-11}\left(\gamma_0,, \beta_{11}^0\right)\right) - S_n\left(\gamma_0, \underline{\theta}_0\right)\right) \\
& \xrightarrow{d} \frac{1}{2}\widehat{u}_{12}'S_{\theta_{12}\theta_{12}}\widehat{u}_{12} - \widehat{u}_{12}'W_{12} + \frac{1}{2}\widehat{u}_2'S_{\theta_2\theta_2}\widehat{u}_2 - \widehat{u}_2'W_2 \\
& - \min_{u_{11},v}\left\{\frac{1}{2}S_{\beta_{11}\beta_{11}}u_{11}^2 + u_{11}S_{\beta_{11}\theta_{12}}\widehat{u}_{12} + \frac{1}{2}\widehat{u}_{12}'S_{\theta_{12}\theta_{12}}\widehat{u}_{12} - u_{11}W_{11} - \widehat{u}_{12}'W_{12} + \frac{1}{2}\widehat{u}_2'S_{\theta_2\theta_2}^2\widehat{u}_2 - \widehat{u}_2'W_2 + C(v)\right\} \\
= \ & -\frac{1}{2}S_{\beta_{11}\beta_{11}}\widehat{u}_{11}^2 - \widehat{u}_{11}S_{\beta_{11}\theta_{12}}\widehat{u}_{12} + \widehat{u}_{11}W_{11} + \max_v \ \{-C(v)\} \\
& \overset{d}{=} \frac{1}{2}\frac{W_1'\left(1, -S_{\beta_{11}\theta_{12}}S_{\theta_{12}\theta_{12}}^{-1}\right)'\left(1, -S_{\beta_{11}\theta_{12}}S_{\theta_{12}\theta_{12}}^{-1}\right)W_1}{S_{\beta_{11}\beta_{11}}} + \eta^2\xi(\varphi,\phi),
\end{aligned}$$

where

$$(\widehat{u}_{12}, \widehat{u}_2) = \arg\min_{u_{12},u_2}\left\{\frac{1}{2}u_{12}'S_{\theta_{12}\theta_{12}}u_{12} - u_{12}'W_{12} + \frac{1}{2}u_2'S_{\theta_2\theta_2}u_2 - u_2'W_2\right\} = \left(S_{\theta_{12}\theta_{12}}^{-1}W_{12}, S_{\theta_2\theta_2}^{-1}W_2\right),$$

and

$$\widehat{u}_{11} = \frac{\left(1, -S_{\beta_{11}\theta_{12}}S_{\theta_{12}\theta_{12}}^{-1}\right)W_1}{S_{\beta_{11}\beta_{11}}}.$$

The rest of the proof is the same as in the proof of Theorem 6. ■

The CIs for γ and β_{11} can be similarly constructed as those that are based on the LR_n statistics, but the calculations are more involved. Although inverting $LR_{1n}(\gamma)$ is still practically feasible, inverting $LR_{1n}(\gamma, \beta_{11})$ is too time-consuming since we need to grid search the β_{11}'s (this is because $\{\beta_{11} | LR_{1n}(q_{it}, \beta_{11}) \le \widehat{c}_{1\alpha}\}$ need not be an interval anymore, where $\widehat{c}_{1\alpha}$ is the new critical value); that is, the test statistic $LR_{1n}(\gamma, \beta_{11})$ only has theoretical value in this paper. Note that $S_{\beta_{11}\beta_{11}} \ge \widetilde{S}_{\beta_{11}\beta_{11}}$, where the equality holds only if $S_{\beta_{11}\theta_{12}} = 0$, which is generally impossible since $\breve{\mathbf{x}}_{it}^{-1}$ contains $\overline{x}_i^1 := \frac{1}{T}\sum_{t=1}^T x_{it}^1$. In other words, the coefficient before the χ_1^2 distribution is smaller than that in Theorem 6.

Comparing with Theorems 5 and 6, we can draw some interesting conclusions. First, replacing $\widehat{\underline{\theta}}(\widehat{\gamma})$ by $\widehat{\underline{\theta}}(\gamma_0)$ in $LR_n(\gamma)$ or $\widehat{\underline{\theta}}_{-11}\left(\widehat{\gamma}, \widehat{\beta}_{11}\right)$ by $\widehat{\underline{\theta}}_{-11}\left(\gamma_0, \beta_{11}^0\right)$ will not affect the asymptotic distribution of $\widehat{\gamma}$ and thus the $\xi(\varphi,\phi)$ component in the LR_n statistics. Second, replacing $\widehat{\underline{\theta}}_{-11}\left(\widehat{\gamma}, \widehat{\beta}_{11}\right)$ by $\widehat{\underline{\theta}}_{-11}\left(\gamma_0, \beta_{11}^0\right)$ indeed affects the asymptotic distribution of $\widehat{\beta}_{11}$ and thus the coefficient before the χ_1^2 distribution in the asymptotic distribution of $LR_n\left(\gamma_0,\beta_{11}^0\right)$. Actually, from the proofs of Theorems 6 and 8, $\widehat{\beta}_{11}\left(\widehat{\underline{\theta}}_{-11}\left(\gamma_0,\beta_{11}^0\right)\right)$ in $LR_n\left(\gamma_0,\beta_{11}^0\right)$ is more efficient than $\widehat{\beta}_{11}$ in $LR_n\left(\gamma_0\right)$, which is mainly because $\widehat{\beta}_{11}\left(\widehat{\underline{\theta}}_{-11}\left(\gamma_0,\beta_{11}^0\right)\right)$ uses the information $\beta_{11}=\beta_{11}^0$ in the null. This sharply contrasts the $\widehat{\gamma}$ case where the null information $\gamma=\gamma_0$ will not improve its first-order efficiency; see Banerjee and McKeague (2007) for a scenario where the null information $\gamma=\gamma_0$ indeed has material content. Third, note that $LR_{1n}\left(\gamma\right)\leq LR_n\left(\gamma\right)$ with $\min_\gamma LR_{1n}\left(\gamma\right)=\min_\gamma LR_n\left(\gamma\right)=0$ because $S_n\left(\widehat{\gamma}\left(\widehat{\underline{\theta}}\left(\gamma\right)\right),\widehat{\underline{\theta}}\left(\gamma\right)\right)\geq S_n\left(\widehat{\gamma},\widehat{\underline{\theta}}\left(\widehat{\gamma}\right)\right)$ and $S_n\left(\gamma,\widehat{\underline{\theta}}\left(\gamma\right)\right)\geq S_n\left(\widehat{\gamma}\left(\widehat{\underline{\theta}}\left(\gamma\right)\right),\widehat{\underline{\theta}}\left(\gamma\right)\right)$ with both equalities holding when $\gamma=\widehat{\gamma}$, and similarly, $LR_{1n}\left(\gamma,\beta_{11}\right)\leq LR_n\left(\gamma,\beta_{11}\right)$ with $\min_{\gamma,\beta_{11}}LR_{1n}\left(\gamma,\beta_{11}\right)=\min_{\gamma,\beta_{11}}LR_n\left(\gamma,\beta_{11}\right)=0$. Since the critical values for $LR_{1n}\left(\gamma\right)$ and $LR_n\left(\gamma\right)$ are the same, $LR_{1n}\left(\gamma\right)$ would result in a wider CI for γ. On the other hand, the critical value for $LR_{1n}\left(\gamma,\beta_{11}\right)$ is smaller than $LR_n\left(\gamma,\beta_{11}\right)$, and thus the CI resulting from $LR_{1n}\left(\gamma,\beta_{11}\right)$ need not be wider than that from $LR_n\left(\gamma,\beta_{11}\right)$.

Algorithms of CI Construction for β and γ

To simplify notation we neglect the constant $1/2$ in $S_n\left(\theta\right)$. Correspondingly, all critical values are multiplied by 2.

CI Construction Based on LR_n

In constructing the CI for β_{11} based on $LR_n\left(\gamma,\beta_{11}\right)$,

$$\begin{aligned}
& S_n\left(q_{it},\beta_{11}\right)-S_n\left(\widehat{\gamma},\widehat{\beta}_{11}\right)\\
= & \left(\mathbf{y}_{\leq q_{it}}-\mathbf{X}^1_{\leq q_{it}}\beta_{11}\right)'\mathbf{M}^2_{\leq q_{it}}\left(\mathbf{y}_{\leq q_{it}}-\mathbf{X}^1_{\leq q_{it}}\beta_{11}\right)+\mathbf{y}'_{>q_{it}}\mathbf{M}_{>q_{it}}\mathbf{y}_{>q_{it}}\\
& -\mathbf{y}'_{\leq\widehat{\gamma}}\mathbf{M}_{\leq\widehat{\gamma}}\mathbf{y}_{\leq\widehat{\gamma}}-\mathbf{y}'_{>\widehat{\gamma}}\mathbf{M}_{>\widehat{\gamma}}\mathbf{y}_{>\widehat{\gamma}}\\
= & \left(\mathbf{X}^{1\prime}_{\leq q_{it}}\mathbf{M}^2_{\leq q_{it}}\mathbf{X}^1_{\leq q_{it}}\right)\beta_{11}^2-2\left(\mathbf{X}^{1\prime}_{\leq q_{it}}\mathbf{M}^2_{\leq q_{it}}\mathbf{y}_{\leq q_{it}}\right)\beta_{11}\\
& +\mathbf{y}'_{\leq q_{it}}\mathbf{M}^2_{\leq q_{it}}\mathbf{y}_{\leq q_{it}}+\mathbf{y}'_{>q_{it}}\mathbf{M}_{>q_{it}}\mathbf{y}_{>q_{it}}-\mathbf{y}'_{\leq\widehat{\gamma}}\mathbf{M}_{\leq\widehat{\gamma}}\mathbf{y}_{\leq\widehat{\gamma}}-\mathbf{y}'_{>\widehat{\gamma}}\mathbf{M}_{>\widehat{\gamma}}\mathbf{y}_{>\widehat{\gamma}},
\end{aligned}$$

where $\mathbf{M}^2_{\leq\gamma}$ is the annihilator of $\left\{\breve{\mathbf{x}}^{-1}_{i\tau}1(q_{i\tau}\leq\gamma)\right\}_{i,\tau}$, $\mathbf{M}_{\leq\gamma}$ and $\mathbf{M}_{>\gamma}$ are annihilators of $\left\{\breve{\mathbf{x}}_{i\tau}1(q_{i\tau}\leq\gamma)\right\}_{i,\tau}$ and $\left\{\breve{\mathbf{x}}_{i\tau}1(q_{i\tau}>\gamma)\right\}_{i,\tau}$, $\mathbf{y}_{\leq\gamma}$ and $\mathbf{y}_{>\gamma}$ are vectors stacking $\left\{y_{i\tau}1(q_{i\tau}\leq\gamma)\right\}_{i,\tau}$ and $\left\{y_{i\tau}1(q_{i\tau}>\gamma)\right\}_{i,\tau}$, $\mathbf{X}^1_{\leq\gamma}$ is the vector stacking $\left\{x^1_{i\tau}1(q_{i\tau}\leq\gamma)\right\}_{i,\tau}$, and $i=1,\cdots,N,\tau=1,\cdots,T$ in all vectors and matrices. Note that $\mathbf{y}'_{\leq q_{it}}\mathbf{M}^2_{\leq q_{it}}\mathbf{y}_{\leq q_{it}}+\mathbf{y}'_{>q_{it}}\mathbf{M}_{>q_{it}}\mathbf{y}_{>q_{it}}$ is the SSR when the threshold is set at q_{it} and the regressors in the left regime are only $\breve{\mathbf{x}}^{-1}_{i\tau}$; we denote it as $S_{1n}\left(q_{it}\right)$, so the constant term of $S_n\left(q_{it},\beta_{11}\right)-S_n\left(\widehat{\gamma},\widehat{\beta}_{11}\right)$ is $S_{1n}\left(q_{it}\right)-S_n\left(\widehat{\gamma}\right)$. As a result,

$$\begin{aligned}
& \left\{\beta_{11}|LR_n\left(q_{it},\beta_{11}\right)\leq\widehat{c}_\alpha\right\}\\
= & \left\{\left\langle\mathbf{X}^1_{\leq q_{it}},\mathbf{X}^1_{\leq q_{it}}\right\rangle\beta_{11}^2-2\left\langle\mathbf{X}^1_{\leq q_{it}},\mathbf{y}_{\leq q_{it}}\right\rangle\beta_{11}+S_{1n}\left(q_{it}\right)-S_n\left(\widehat{\gamma}\right)\leq\widehat{c}_\alpha\widehat{\eta}^2\right\}\\
= & \begin{cases}\left[\dfrac{\left\langle\mathbf{X}^1_{\leq q_{it}},\mathbf{y}_{\leq q_{it}}\right\rangle-\sqrt{D(q_{it},\alpha)}}{\left\langle\mathbf{X}^1_{\leq q_{it}},\mathbf{X}^1_{\leq q_{it}}\right\rangle},\dfrac{\left\langle\mathbf{X}^1_{\leq q_{it}},\mathbf{y}_{\leq q_{it}}\right\rangle+\sqrt{D(q_{it},\alpha)}}{\left\langle\mathbf{X}^1_{\leq q_{it}},\mathbf{X}^1_{\leq q_{it}}\right\rangle}\right] & \text{if } D\left(q_{it},\alpha\right)\geq0,\\ \emptyset, & \text{otherwise,}\end{cases}
\end{aligned}$$

where the inner product $\langle\cdot,\cdot\rangle$ on $\mathbb{R}^n$ is defined as $\langle\mathbf{z},\mathbf{w}\rangle=\mathbf{z}'\mathbf{M}^2_{\leq q_{it}}\mathbf{w}$, and $D\left(q_{it},\alpha\right):=\left\langle\mathbf{X}^1_{\leq q_{it}},\mathbf{y}_{\leq q_{it}}\right\rangle^2-\left\langle\mathbf{X}^1_{\leq q_{it}},\mathbf{X}^1_{\leq q_{it}}\right\rangle\left(S_{1n}\left(q_{it}\right)-S_n\left(\widehat{\gamma}\right)-\widehat{c}_\alpha\widehat{\eta}^2\right)$.

When there are variables without threshold effects, we need only slightly adjust the procedure above.

For β_{11},

$$S_n\left(q_{it},\beta_{11}\right)=\left(\mathbf{y}-\mathbf{X}^1\beta_{11}\right)'\mathbf{M}^2\left(\mathbf{y}-\mathbf{X}^1\beta_{11}\right),$$

where $\mathbf{M}^2$ is the annihilator of $\left\{\left(\begin{array}{ccc}\check{\mathbf{x}}_{1i\tau}', & \check{\mathbf{x}}_{2i\tau}'1\left(q_{i\tau}\le q_{it}\right), & \check{\mathbf{x}}_{2i\tau}'1\left(q_{i\tau}>q_{it}\right)\end{array}\right)\right\}_{i,\tau}$ with $\check{\mathbf{x}}_{1i\tau}$ being $\mathbf{x}_{1i\tau}$ deleting $x_{11i\tau}$, $\mathbf{X}^1$ is the vector stacking $\{x_{11i\tau}\}_{i,\tau}$, and $\mathbf{y}$ is the vector stacking $\{y_{i\tau}\}_{i,\tau}$. Now,

$$\begin{aligned}
&\left\{\beta_{11}|LR_n\left(q_{it},\beta_{11}\right)\le\widehat{c}_\alpha\right\}\\
=\;&\left\{\left\langle\mathbf{X}^1,\mathbf{X}^1\right\rangle\beta_{11}^2-2\left\langle\mathbf{X}^1,\mathbf{y}\right\rangle\beta_{11}+\left\langle\mathbf{y},\mathbf{y}\right\rangle-S_n\left(\widehat{\gamma}\right)\le\widehat{c}_\alpha\widehat{\eta}^2\right\}\\
=\;&\begin{cases}\left[\frac{\left\langle\mathbf{X}^1,\mathbf{y}\right\rangle-\sqrt{D\left(q_{it},\alpha\right)}}{\left\langle\mathbf{X}^1,\mathbf{X}^1\right\rangle},\frac{\left\langle\mathbf{X}^1,\mathbf{y}\right\rangle+\sqrt{D\left(q_{it},\alpha\right)}}{\left\langle\mathbf{X}^1,\mathbf{X}^1\right\rangle}\right] & \text{if } D\left(q_{it},\alpha\right)\ge0,\\ \emptyset, & \text{otherwise,}\end{cases}
\end{aligned}$$

where the inner product $\langle\cdot,\cdot\rangle$ is defined as $\langle\mathbf{z},\mathbf{w}\rangle=\mathbf{z}'\mathbf{M}^2\mathbf{w}$, and $D\left(q_{it},\alpha\right):=\left\langle\mathbf{X}^1,\mathbf{y}\right\rangle^2-\left\langle\mathbf{X}^1,\mathbf{X}^1\right\rangle\left(\langle\mathbf{y},\mathbf{y}\rangle-S_n\left(\widehat{\gamma}\right)-\widehat{c}_\alpha\widehat{\eta}^2\right)$. For β_{121}, in $LR_n\left(q_{it},\beta_{121}\right)$, only redefine $\mathbf{M}^2$ as the annihilator of $\left\{\left(\begin{array}{ccc}\mathbf{x}_{1i\tau}', & \check{\mathbf{x}}_{2i\tau}'1\left(q_{i\tau}\le q_{it}\right), & \check{\mathbf{x}}_{2i\tau}'1\left(q_{i\tau}>q_{it}\right)\end{array}\right)\right\}_{i,\tau}$, and $\mathbf{X}^1$ as the vector stacking $\{x_{21i\tau}1\left(q_{i\tau}\le q_{it}\right)\}_{i,\tau}$, where $\check{\mathbf{x}}_{2i\tau}$ is $\check{\mathbf{x}}_{2i\tau}$ deleting $x_{21i\tau}$ whose coefficient in the left regime is β_{121}.

CI Construction Based on LR_{1n}

In constructing the CI for γ based on $LR_{1n}\left(\gamma\right)$,

$$\begin{aligned}
&S_n\left(\gamma,\widehat{\underline{\theta}}\left(\gamma\right)\right)-S_n\left(\widehat{\gamma}\left(\widehat{\underline{\theta}}\left(\gamma\right)\right),\widehat{\underline{\theta}}\left(\gamma\right)\right)\\
=&\sum_{i=1}^N\sum_{t=1}^T\left[\left(y_{it}-\check{\mathbf{x}}_{it}'\widehat{\theta}_1\left(\gamma\right)\right)^2-\left(y_{it}-\check{\mathbf{x}}_{it}'\widehat{\theta}_2\left(\gamma\right)\right)^2\right]1(\widehat{\gamma}\left(\widehat{\underline{\theta}}\left(\gamma\right)\right)<q_{it}\le\gamma)\\
+&\sum_{i=1}^N\sum_{t=1}^T\left[\left(y_{it}-\check{\mathbf{x}}_{it}'\widehat{\theta}_2\left(\gamma\right)\right)^2-\left(y_{it}-\check{\mathbf{x}}_{it}'\widehat{\theta}_1\left(\gamma\right)\right)^2\right]1(\gamma<q_{it}\le\widehat{\gamma}\left(\widehat{\underline{\theta}}\left(\gamma\right)\right)),
\end{aligned}$$

where note that $\widehat{\gamma}\left(\widehat{\underline{\theta}}\left(\gamma\right)\right)$ depends on γ. In practice, we can calculate $S_n\left(\gamma,\widehat{\underline{\theta}}\left(\gamma\right)\right)-S_n\left(\widehat{\gamma}\left(\widehat{\underline{\theta}}\left(\gamma\right)\right),\widehat{\underline{\theta}}\left(\gamma\right)\right)$ directly rather than based on this decomposition.

In constructing the CI for β_{11} based on $LR_{1n}\left(\gamma,\beta_{11}\right)$, we can still collect the intervals of β_{11} for each $q_{it}\in\Gamma$:

$$\bigcup\nolimits_{q_{it}\in\Gamma}\left\{\beta_{11}|LR_{1n}\left(q_{it},\beta_{11}\right)\le\widehat{c}_\alpha^1\right\},$$

where $\widehat{c}_\alpha^1$ is the level α critical value for $LR_{1n}\left(\gamma_0,\beta_{11}^0\right)$. For each $q_{it}\in\Gamma$,

$$\begin{aligned}
&S_n\left(q_{it},\beta_{11},\widehat{\underline{\theta}}_{-11}\left(q_{it},\beta_{11}\right)\right)-S_n\left(\widehat{\gamma}\left(\widehat{\underline{\theta}}_{-11}\left(q_{it},\beta_{11}\right)\right),\widehat{\beta}_{11}\left(\widehat{\underline{\theta}}_{-11}\left(q_{it},\beta_{11}\right)\right),\widehat{\underline{\theta}}_{-11}\left(q_{it},\beta_{11}\right)\right)\\
=&\left(\mathbf{y}_{\le q_{it}}-\mathbf{X}_{\le q_{it}}^1\beta_{11}\right)'\mathbf{M}_{\le q_{it}}^2\left(\mathbf{y}_{\le q_{it}}-\mathbf{X}_{\le q_{it}}^1\beta_{11}\right)+\mathbf{y}_{>q_{it}}'\mathbf{M}_{>q_{it}}\mathbf{y}_{>q_{it}}\\
-&\left(\mathbf{y}_{\le\widetilde{\gamma}}-\mathbf{X}_{\le\widetilde{\gamma}}^1\widehat{\theta}_{12}\left(q_{it},\beta_{11}\right)\right)'\mathbf{M}_{\le\widetilde{\gamma}}^1\left(\mathbf{y}_{\le\widetilde{\gamma}}-\mathbf{X}_{\le\widetilde{\gamma}}^1\widehat{\theta}_{12}\left(q_{it},\beta_{11}\right)\right)-\left(\mathbf{y}_{>\widetilde{\gamma}}-\mathbf{X}_{>\widetilde{\gamma}}\widehat{\theta}_2\left(q_{it}\right)\right)'\left(\mathbf{y}_{>\widetilde{\gamma}}-\mathbf{X}_{>\widetilde{\gamma}}\widehat{\theta}_2\left(q_{it}\right)\right),
\end{aligned}$$

where $\widetilde{\gamma}:=\widehat{\gamma}\left(\widehat{\underline{\theta}}_{-11}\left(q_{it},\beta_{11}\right)\right)$ depends on β_{11}, $\mathbf{M}_{\le\gamma}^1$ is the annihilator of $\left\{x_{i\tau}^11(q_{i\tau}\le\gamma)\right\}_{i,\tau}$, $\mathbf{X}_{>\gamma}$ is the matrix stacking $\{\check{\mathbf{x}}_{i\tau}1(q_{i\tau}>\gamma)\}_{i,\tau}$,

$$\widehat{\theta}_{12}\left(q_{it},\beta_{11}\right)=\left(\mathbf{X}_{\le q_{it}}^{2\prime}\mathbf{X}_{\le q_{it}}^2\right)^{-1}\left(\mathbf{X}_{\le q_{it}}^{2\prime}\left(\mathbf{y}_{\le q_{it}}-\mathbf{X}_{\le q_{it}}^1\beta_{11}\right)\right),$$

with $\mathbf{X}^2_{\leq\gamma}$ being the matrix stacking $\left\{\check{\mathbf{x}}_{i\tau}^{-1}1(q_{i\tau}\leq\gamma)\right\}_{i,\tau}$, and

$$\widehat{\theta}_2\left(q_{it}\right)=\left(\mathbf{X}'_{>q_{it}}\mathbf{X}_{>q_{it}}\right)^{-1}\left(\mathbf{X}'_{>q_{it}}\mathbf{y}_{>q_{it}}\right)$$

depends only on q_{it}. Because $\widetilde{\gamma}$ depends on β_{11} in a nonlinear way, we construct $\left\{\beta_{11}|LR_{1n}\left(q_{it},\beta_{11}\right)\leq\widehat{c}_\alpha^1\right\}$ by a grid search.

When there are variables without threshold effects, we need also to adjust the procedure above slightly. For β_{11}, $S_n\left(q_{it},\beta_{11},\underline{\widehat{\theta}}_{-11}\left(q_{it},\beta_{11}\right)\right)$ is the same as in $LR_n\left(\gamma,\beta_{11}\right)$, and

$$\begin{aligned}
& S_n\left(\widehat{\gamma}\left(\underline{\widehat{\theta}}_{-11}\left(q_{it},\beta_{11}\right)\right),\widehat{\beta}_{11}\left(\underline{\widehat{\theta}}_{-11}\left(q_{it},\beta_{11}\right)\right),\underline{\widehat{\theta}}_{-11}\left(q_{it},\beta_{11}\right)\right)\\
= & \left(\mathbf{y}-\left(\check{\mathbf{X}}^1,\check{\mathbf{X}}^2_{\leq\widetilde{\gamma}},\check{\mathbf{X}}^2_{>\widetilde{\gamma}}\right)\left(\widehat{\beta}^1_{-11}\left(q_{it},\beta_{11}\right)',\widehat{\theta}_{12}\left(q_{it},\beta_{11}\right)',\widehat{\theta}_{22}\left(q_{it},\beta_{11}\right)'\right)'\right)'\\
& \mathbf{M}^{11}\left(\mathbf{y}-\left(\check{\mathbf{X}}^1,\check{\mathbf{X}}^2_{\leq\widetilde{\gamma}},\check{\mathbf{X}}^2_{>\widetilde{\gamma}}\right)\left(\widehat{\beta}^1_{-11}\left(q_{it},\beta_{11}\right)',\widehat{\theta}_{12}\left(q_{it},\beta_{11}\right)',\widehat{\theta}_{22}\left(q_{it},\beta_{11}\right)'\right)'\right),
\end{aligned}$$

where $\widetilde{\gamma}:=\widehat{\gamma}\left(\underline{\widehat{\theta}}_{-11}\left(q_{it},\beta_{11}\right)\right)$ depends on β_{11}, $\mathbf{M}^{11}$ is the annihilator of $\{x_{11i\tau}\}_{i,\tau}$, $\left(\check{\mathbf{X}}^1,\check{\mathbf{X}}^2_{\leq\widetilde{\gamma}},\check{\mathbf{X}}^2_{>\widetilde{\gamma}}\right)$ is the matrix stacking $\left\{\left(\begin{array}{ccc}\check{\mathbf{x}}'_{1i\tau}, & \check{\mathbf{x}}'_{2i\tau}1\left(q_{i\tau}\leq\widetilde{\gamma}\right), & \check{\mathbf{x}}'_{2i\tau}1\left(q_{i\tau}>\widetilde{\gamma}\right)\end{array}\right)\right\}_{i,\tau}$, and β^1_{-11} is β_1 deleting β_{11}. For β_{121}, $S_n\left(q_{it},\beta_{121},\underline{\widehat{\theta}}_{-121}\left(q_{it},\beta_{121}\right)\right)$ is the same as in $LR_n\left(\gamma,\beta_{121}\right)$, and

$$\begin{aligned}
& S_n\left(\widehat{\gamma}\left(\underline{\widehat{\theta}}_{-121}\left(q_{it},\beta_{121}\right)\right),\widehat{\beta}_{121}\left(\underline{\widehat{\theta}}_{-121}\left(q_{it},\beta_{121}\right)\right),\underline{\widehat{\theta}}_{-121}\left(q_{it},\beta_{121}\right)\right)\\
= & \left(\mathbf{y}-\left(\mathbf{X}_1,\check{\mathbf{X}}^2_{\leq\widetilde{\gamma}},\check{\mathbf{X}}^2_{>\widetilde{\gamma}}\right)\left(\widehat{\beta}_1\left(q_{it},\beta_{121}\right)',\widehat{\theta}^{12}_{-121}\left(q_{it},\beta_{121}\right)',\widehat{\theta}_{22}\left(q_{it},\beta_{121}\right)'\right)'\right)'\\
& \mathbf{M}^{21}\left(\mathbf{y}-\left(\mathbf{X}_1,\check{\mathbf{X}}^2_{\leq\widetilde{\gamma}},\check{\mathbf{X}}^2_{>\widetilde{\gamma}}\right)\left(\widehat{\beta}_1\left(q_{it},\beta_{121}\right)',\widehat{\theta}^{12}_{-121}\left(q_{it},\beta_{121}\right)',\widehat{\theta}_{22}\left(q_{it},\beta_{121}\right)'\right)'\right),
\end{aligned}$$

where $\widetilde{\gamma}:=\widehat{\gamma}\left(\underline{\widehat{\theta}}_{-121}\left(q_{it},\beta_{121}\right)\right)$ depends on β_{121}, $\mathbf{M}^{21}$ is the annihilator of $\{x_{21i\tau}\}_{i,\tau}$, $\left(\mathbf{X}_1,\check{\mathbf{X}}^2_{\leq\widetilde{\gamma}},\check{\mathbf{X}}^2_{>\widetilde{\gamma}}\right)$ is the matrix stacking $\left\{\left(\begin{array}{ccc}\mathbf{x}'_{1i\tau}, & \check{\mathbf{x}}'_{2i\tau}1\left(q_{i\tau}\leq\widetilde{\gamma}\right), & \check{\mathbf{x}}'_{2i\tau}1\left(q_{i\tau}>\widetilde{\gamma}\right)\end{array}\right)\right\}_{i,\tau}$, and θ^{12}_{-121} is θ_{12} deleting β_{121}.

CI Construction Based on LR_{2n}

In constructing the CI for γ based on $LR_{2n}\left(\gamma\right)$,

$$\begin{aligned}
& S_n\left(\gamma,\underline{\widehat{\theta}}\right)-S_n\left(\widehat{\gamma},\underline{\widehat{\theta}}\right)\\
& =\textstyle\sum_{i=1}^N\sum_{t=1}^T\left[\left(y_{it}-\check{\mathbf{x}}'_{it}\widehat{\theta}_1\right)^2-\left(y_{it}-\check{\mathbf{x}}'_{it}\widehat{\theta}_2\right)^2\right]1(\widehat{\gamma}<q_{it}\leq\gamma)\\
& +\textstyle\sum_{i=1}^N\sum_{t=1}^T\left[\left(y_{it}-\check{\mathbf{x}}'_{it}\widehat{\theta}_2\right)^2-\left(y_{it}-\check{\mathbf{x}}'_{it}\widehat{\theta}_1\right)^2\right]1(\gamma<q_{it}\leq\widehat{\gamma}).
\end{aligned}$$

In practice, we can calculate $S_n(\gamma,\underline{\widehat{\theta}})-S_n(\widehat{\gamma},\underline{\widehat{\theta}})$ directly rather than based on this decomposition.

In constructing the CI for β_{11} based on $LR_{2n}\left(\gamma,\beta_{11}\right)$, we can still collect the intervals of β_{11} for each $q_{it}\in\Gamma$:

$$\bigcup\nolimits_{q_{it}\in\Gamma}\left\{\beta_{11}|LR_{2n}\left(q_{it},\beta_{11}\right)\leq\widehat{c}_\alpha^2\right\}.$$

For each $q_{it} \in \Gamma$,

$$
\begin{aligned}
& S_n\left(q_{it}, \beta_{11}, \widehat{\underline{\theta}}_{-11}\right) - S_n\left(\widehat{\gamma}, \widehat{\beta}_{11}, \widehat{\underline{\theta}}_{-11}\right) \\
= & \left(\mathbf{y}_{\leq q_{it}} - \mathbf{X}^1_{\leq q_{it}}\beta_{11} - \mathbf{X}^2_{\leq q_{it}}\widehat{\theta}^1_{-11}\right)' \left(\mathbf{y}_{\leq q_{it}} - \mathbf{X}^1_{\leq q_{it}}\beta_{11} - \mathbf{X}^2_{\leq q_{it}}\widehat{\theta}^1_{-11}\right) \\
& + \left(\mathbf{y}_{>q_{it}} - \check{\mathbf{X}}_{>q_{it}}\widehat{\theta}_2\right)' \left(\mathbf{y}_{>q_{it}} - \check{\mathbf{X}}_{>q_{it}}\widehat{\theta}_2\right) - S_n\left(\widehat{\gamma}\right) \\
= & \mathbf{X}^{1\prime}_{\leq q_{it}}\mathbf{X}^1_{\leq q_{it}}\beta^2_{11} - 2\beta_{11}\mathbf{X}^{1\prime}_{\leq q_{it}}\left(\mathbf{y}_{\leq q_{it}} - \mathbf{X}^2_{\leq q_{it}}\widehat{\theta}^1_{-11}\right) + \left(\mathbf{y}_{\leq q_{it}} - \mathbf{X}^2_{\leq q_{it}}\widehat{\theta}^1_{-11}\right)' \left(\mathbf{y}_{\leq q_{it}} - \mathbf{X}^2_{\leq q_{it}}\widehat{\theta}^1_{-11}\right) \\
& + \left(\mathbf{y}_{>q_{it}} - \check{\mathbf{X}}_{>q_{it}}\widehat{\theta}_2\right)' \left(\mathbf{y}_{>q_{it}} - \check{\mathbf{X}}_{>q_{it}}\widehat{\theta}_2\right) - S_n\left(\widehat{\gamma}\right) \\
=: & \mathbf{X}^{1\prime}_{\leq q_{it}}\mathbf{X}^1_{\leq q_{it}}\beta^2_{11} - 2\beta_{11}\mathbf{X}^{1\prime}_{\leq q_{it}}\left(\mathbf{y}_{\leq q_{it}} - \mathbf{X}^2_{\leq q_{it}}\widehat{\theta}^1_{-11}\right) + S_{1n}\left(q_{it}\right) - S_n\left(\widehat{\gamma}\right),
\end{aligned}
$$

where $\mathbf{X}^2_{\leq\gamma}$ is the matrix stacking $\left\{\check{\mathbf{x}}^{-1}_{i\tau}1(q_{i\tau} \leq \gamma)\right\}_{i,\tau}$, $\check{\mathbf{X}}_{>\gamma}$ is the matrix stacking $\{\check{\mathbf{x}}_{i\tau}1(q_{i\tau} > \gamma)\}_{i,\tau}$, and θ^1_{-11} is θ_1 deleting β_{11}. As a result,

$$
\begin{aligned}
& \left\{\beta_{11} | LR_{2n}\left(q_{it}, \beta_{11}\right) \leq \widehat{c}^2_\alpha\right\} \\
= & \left\{\left\langle \mathbf{X}^1_{\leq q_{it}}, \mathbf{X}^1_{\leq q_{it}}\right\rangle \beta^2_{11} - 2\left\langle \mathbf{X}^1_{\leq q_{it}}, \mathbf{y}_{\leq q_{it}} - \mathbf{X}^2_{\leq q_{it}}\widehat{\theta}^1_{-11}\right\rangle \beta_{11} + S_{1n}\left(q_{it}\right) - S_n\left(\widehat{\gamma}\right) \leq \widehat{c}^2_\alpha \widehat{\eta}^2\right\} \\
= & \begin{cases} \left[\frac{\left\langle \mathbf{X}^1_{\leq q_{it}}, \mathbf{y}_{\leq q_{it}} - \mathbf{X}^2_{\leq q_{it}}\widehat{\theta}^1_{-11}\right\rangle - \sqrt{D(q_{it},\alpha)}}{\left\langle \mathbf{X}^1_{\leq q_{it}}, \mathbf{X}^1_{\leq q_{it}}\right\rangle}, \frac{\left\langle \mathbf{X}^1_{\leq q_{it}}, \mathbf{y}_{\leq q_{it}} - \mathbf{X}^2_{\leq q_{it}}\widehat{\theta}^1_{-11}\right\rangle + \sqrt{D(q_{it},\alpha)}}{\left\langle \mathbf{X}^1_{\leq q_{it}}, \mathbf{X}^1_{\leq q_{it}}\right\rangle}\right] & \text{if } D\left(q_{it}, \alpha\right) \geq 0, \\ \emptyset, & \text{otherwise,} \end{cases}
\end{aligned}
$$

where $\langle \cdot, \cdot \rangle$ is the usual Euclidean inner product on $\mathbb{R}^n$, and $D\left(q_{it}, \alpha\right) := \left\langle \mathbf{X}^1_{\leq q_{it}}, \mathbf{y}_{\leq q_{it}} - \mathbf{X}^2_{\leq q_{it}}\widehat{\theta}^1_{-11}\right\rangle^2 - \left\langle \mathbf{X}^1_{\leq q_{it}}, \mathbf{X}^1_{\leq q_{it}}\right\rangle \left(S_{1n}\left(q_{it}\right) - S_n\left(\widehat{\gamma}\right) - \widehat{c}^2_\alpha \widehat{\eta}^2\right)$.

When there are variables without threshold effects, we need also to adjust the procedure above slightly. For β_{11},

$$
\begin{aligned}
& S_n\left(q_{it}, \beta_{11}, \widehat{\underline{\theta}}_{-11}\right) - S_n\left(\widehat{\gamma}, \widehat{\beta}_{11}, \widehat{\underline{\theta}}_{-11}\right) \\
= & \left(\mathbf{y} - \mathbf{X}^1\beta_{11} - \left(\check{\mathbf{X}}^1, \check{\mathbf{X}}^2_{\leq q_{it}}, \check{\mathbf{X}}^2_{>q_{it}}\right)\widehat{\underline{\theta}}_{-11}\right)' \left(\mathbf{y} - \mathbf{X}^1\beta_{11} - \left(\check{\mathbf{X}}^1, \check{\mathbf{X}}^2_{\leq q_{it}}, \check{\mathbf{X}}^2_{>q_{it}}\right)\widehat{\underline{\theta}}_{-11}\right) - S_n\left(\widehat{\gamma}\right),
\end{aligned}
$$

which is quadratic in β_{11}, so

$$
\begin{aligned}
& \left\{\beta_{11} | LR_{2n}\left(q_{it}, \beta_{11}\right) \leq \widehat{c}^2_\alpha\right\} \\
= & \begin{cases} \left[\frac{\left\langle \mathbf{X}^1, \mathbf{y} - \left(\check{\mathbf{X}}^1, \check{\mathbf{X}}^2_{\leq q_{it}}, \check{\mathbf{X}}^2_{>q_{it}}\right)\widehat{\underline{\theta}}_{-11}\right\rangle - \sqrt{D(q_{it},\alpha)}}{\left\langle \mathbf{X}^1, \mathbf{X}^1\right\rangle}, \frac{\left\langle \mathbf{X}^1, \mathbf{y} - \left(\check{\mathbf{X}}^1, \check{\mathbf{X}}^2_{\leq q_{it}}, \check{\mathbf{X}}^2_{>q_{it}}\right)\widehat{\underline{\theta}}_{-11}\right\rangle + \sqrt{D(q_{it},\alpha)}}{\left\langle \mathbf{X}^1, \mathbf{X}^1\right\rangle}\right] & \text{if } D\left(q_{it}, \alpha\right) \geq 0, \\ \emptyset, & \text{otherwise,} \end{cases}
\end{aligned}
$$

where $D\left(q_{it}, \alpha\right) = \left\langle \mathbf{X}^1, \mathbf{y} - \left(\check{\mathbf{X}}^1, \check{\mathbf{X}}^2_{\leq q_{it}}, \check{\mathbf{X}}^2_{>q_{it}}\right)\widehat{\underline{\theta}}_{-11}\right\rangle^2 - \left\langle \mathbf{X}^1, \mathbf{X}^1\right\rangle \left(S_{1n}\left(q_{it}\right) - S_n\left(\widehat{\gamma}\right) - \widehat{c}^2_\alpha \widehat{\eta}^2\right)$ with $S_{1n}\left(q_{it}\right) = \left(\mathbf{y} - \left(\check{\mathbf{X}}^1, \check{\mathbf{X}}^2_{\leq q_{it}}, \check{\mathbf{X}}^2_{>q_{it}}\right)\widehat{\underline{\theta}}_{-11}\right)' \left(\mathbf{y} - \left(\check{\mathbf{X}}^1, \check{\mathbf{X}}^2_{\leq q_{it}}, \check{\mathbf{X}}^2_{>q_{it}}\right)\widehat{\underline{\theta}}_{-11}\right)$. For β_{121}, in $S_n\left(q_{it}, \beta_{11}, \widehat{\underline{\theta}}_{-11}\right)$, only replace $\left(\check{\mathbf{X}}^1, \check{\mathbf{X}}^2_{\leq q_{it}}, \check{\mathbf{X}}^2_{>q_{it}}\right)$ by $\left(\mathbf{X}_1, \check{\mathbf{X}}^2_{\leq q_{it}}, \check{\mathbf{X}}^2_{>q_{it}}\right)$, and redefine $\mathbf{X}^1$ as the vector stacking $\left\{x_{21i\tau}1\left(q_{i\tau} \leq q_{it}\right)\right\}_{i,\tau}$ and $\widehat{\underline{\theta}}_{-11}$ as $\left(\widehat{\beta}'_1, \widehat{\theta}^{12\prime}_{-121}, \widehat{\theta}'_{22}\right)'$, i.e., $\widehat{\underline{\theta}}$ excluding $\widehat{\beta}_{121}$.

Appendix F: Analysis in Simplified Models

This appendix consisers two simplified versions of our CRE model. We first provide simplified formulae for the asymptotic distributions of $\widehat{\gamma}$, $\widehat{\theta}_\ell$, $LR_n\left(\gamma_0\right)$, and $LR_n\left(\gamma_0, \beta^0_{11}\right)$ in correctly specified error components models, and these simplified formulae can also be extended to the asymptotic distributions of $LR_{1n}\left(\gamma_0\right)$,

$LR_{1n}\left(\gamma_0, \beta_{11}^0\right)$, $LR_{2n}\left(\gamma_0\right)$, and $LR_{2n}\left(\gamma_0, \beta_{11}^0\right)$. We then consider the case where some variables do not have threshold effects on y_{it}, i.e., their coefficients remain the same over different regimes. This simplified case will be used in our empirical application. We finally discuss testing for unobserved individual-specific threshold effects based on the asymptotics in Section 4 and provide some detailed analysis for the claims in Section 4.2.

Formulae in Correctly Specified Error Components Models

When the model is correctly specified, $e_{\ell it} = \varepsilon_{\ell it} = \alpha_{\ell i} + u_{\ell it}$. In practice, researchers often assume the error components structure on $e_{\ell it}$, e.g., assume $(a_{1i}, a_{2i}, \{u_{it}\}_{t=1}^T)$ and $\mathbf{X}_i$ are independent and u_{it}, $t = 1, \cdots, T$, are independent across regimes and i.i.d. in each regime, and then $\eta^2, \varphi, \phi, V_\ell, \Omega_\ell$ and Ω_{12} can be simplified. First of all, because $c_\pm = \mathbf{0}$ in correctly specified models, $\omega = \frac{c'V_1c}{(c'Dc)^2}$, $\varphi = 1$, and $\eta^2 = \frac{c'V_1c}{c'Dc}$.

In such a case, $\mathbb{E}\left[e_{\ell it}^2|\mathbf{X}_i\right] = \mathbb{E}\left[e_{\ell it}^2\right] = \mathbb{E}\left[a_{\ell i}^2\right] + \sigma_\ell^2 := \varsigma_\ell^2$ with $\mathbb{E}\left[u_{\ell it}^2\right] = \sigma_\ell^2$,[20] $\mathbb{E}\left[e_{\ell it}e_{\ell i\tau}|\mathbf{X}_i\right] = \mathbb{E}\left[a_{\ell i}^2\right] =: c_\ell$, and $\mathbb{E}\left[e_{1it}e_{2i\tau}|\mathbf{X}_i\right] = \mathbb{E}\left[a_{1i}a_{2i}\right] := c_{12}$ for $t, \tau = 1, \cdots, T$ and $\tau \neq t$ such that

$$\begin{aligned} &V_\ell\left(\gamma\right) = \varsigma_\ell^2 D\left(\gamma\right), \\ &\Omega_\ell = \varsigma_\ell^2\left[\left(1-\rho_\ell\right)M_\ell + \rho_\ell\Psi_\ell\right], \\ &\Omega_{12} = c_{12}\Psi_{12}, \end{aligned}$$

which implies

$$\begin{aligned} \omega &= \varsigma_1^2/c'Dc,\ \phi = \varsigma_2^2/\varsigma_1^2,\ \eta^2 = \varsigma_1^2, \text{ and} \\ \Sigma_\ell &= \varsigma_\ell^2\left[\left(1-\rho_\ell\right)M_\ell^{-1} + \rho_\ell M_\ell^{-1}\Psi_\ell M_\ell^{-1}\right], \end{aligned}$$

where $\rho_\ell = c_\ell/\varsigma_\ell^2$ is the correlation between $e_{\ell it}$ and $e_{\ell i\tau}$, and

$$\begin{aligned} &\Psi_1 = \mathbb{E}\left[\left(\sum_{t=1}^T \check{\mathbf{x}}_{it}1\left(q_{it} \le \gamma_0\right)\right)\left(\sum_{t=1}^T \check{\mathbf{x}}_{it}1\left(q_{it} \le \gamma_0\right)\right)'\right], \\ &\Psi_2 = \mathbb{E}\left[\left(\sum_{t=1}^T \check{\mathbf{x}}_{it}1\left(q_{it} > \gamma_0\right)\right)\left(\sum_{t=1}^T \check{\mathbf{x}}_{it}1\left(q_{it} > \gamma_0\right)\right)'\right], \\ &\Psi_{12} = \sum_{t=1}^T\sum_{\tau\neq t}\mathbb{E}\left[\check{\mathbf{x}}_{it}\check{\mathbf{x}}_{i\tau}'1\left(q_{it} \le \gamma_0\right)1\left(q_{i\tau} > \gamma_0\right)\right]. \end{aligned}$$

Note that $\Sigma_\ell \neq \varsigma_\ell^2 M_\ell^{-1}$ as in cross sections.[21] When $\alpha_{1i} = \alpha_{2i}$, we need only set $c_\psi = \mathbf{0}$ in ω and set $a_{1i} = a_{2i}$ in V_ℓ, Ω_ℓ and Ω_{12}.

In the error components model, the formulae of $\widehat{\Omega}_\ell$ and $\widehat{\Omega}_{12}$ in Section 4.2 can be simplified:

$$\widehat{\Omega}_\ell = \widehat{\varsigma}_\ell^2\left[\left(1-\widehat{\rho}_\ell\right)\widehat{M}_\ell + \widehat{\rho}_\ell\widehat{\Psi}_\ell\right] \text{ and } \widehat{\Omega}_{12} = \widehat{c}_{12}\widehat{\Psi}_{12}.$$

Here, in $\widehat{\Omega}_1$,

$$\begin{aligned} &\widehat{\Psi}_1 = \frac{1}{N}\sum_{i=1}^N\left(\sum_{t=1}^T \check{\mathbf{x}}_{it}1\left(q_{it} \le \widehat{\gamma}\right)\right)\left(\sum_{t=1}^T \check{\mathbf{x}}_{it}1\left(q_{it} \le \widehat{\gamma}\right)\right)', \\ &\widehat{\varsigma}_1^2 = \frac{1}{n_1}\sum_{i=1}^N\sum_{t=1}^T \widehat{e}_{1it}^2 1(q_{it} \le \widehat{\gamma}),\ \widehat{\rho}_1 = \widehat{c}_1/\widehat{\varsigma}_1^2, \end{aligned}$$

[20] Because $\underline{\mathbf{z}}_i$ includes 1, $\mathbb{E}\left[a_{\ell i}\right] = 0$.

[21] In Hansen (1999), $\widehat{\beta}_1$ and $\widehat{\beta}_2$ are also correlated, but in the homoskedastic case, such a form of simplification is indeed available. A key difference between the fixed effects estimator and the CRE estimator of β is that β_1 and β_2 cannot be estimated using separate data in the former, which greatly affects the formulae of the asymptotic variance estimator of $\widehat{\beta}$ on pages 352-353 of Hansen (1999).

where $n_1 = \sum_{i=1}^{N} \sum_{t=1}^{T} 1(q_{it} \le \widehat{\gamma})$,

$$\begin{aligned} \widehat{c}_1 &= \tfrac{1}{N_1} \sum_{i=1}^{N} \tfrac{\sum_{\tau \neq t} \widehat{e}_{1it} \widehat{e}_{1i\tau} 1(q_{it} \le \widehat{\gamma}) 1(q_{i\tau} \le \widehat{\gamma})}{T_{1i}(T_{1i}-1)} 1\left(T_{1i} \ge 2\right) \\ &= \tfrac{1}{N_1} \sum_{i=1}^{N} \tfrac{\left(\sum_{t=1}^{T} \widehat{e}_{1it} 1(q_{it} \le \widehat{\gamma})\right)^2 - \sum_{t=1}^{T} \widehat{e}_{1it}^2 1(q_{it} \le \widehat{\gamma})}{T_{1i}(T_{1i}-1)} 1\left(T_{1i} \ge 2\right) \end{aligned}$$

with $N_1 = \sum_{i=1}^{N} 1(T_{1i} \ge 2)$ and $T_{1i} = \sum_{t=1}^{T} 1(q_{it} \le \widehat{\gamma})$. In $\widehat{\Omega}_2$, $\widehat{\Psi}_2$, $\widehat{\varsigma}_2^2$ and $\widehat{\rho}_2$ are similarly defined but replace $q_{it} \le \widehat{\gamma}$ by $q_{it} > \widehat{\gamma}$ and $\widehat{e}_{1it}$ by $\widehat{e}_{2it}$. In $\widehat{\Omega}_{12}$,

$$\begin{aligned} \widehat{c}_{12} &= \tfrac{1}{N_{12}} \sum_{i=1}^{N} \tfrac{\left(\sum_{t=1}^{T} \widehat{e}_{1it} 1(q_{it} \le \widehat{\gamma})\right)\left(\sum_{\tau=1}^{T} \widehat{e}_{2it} 1(q_{it} > \widehat{\gamma})\right)}{T_{1i} T_{2i}} 1\left(T_{1i} \ge 1, T_{2i} \ge 1\right), \\ \widehat{\Psi}_{12} &= \tfrac{1}{N} \sum_{i=1}^{N} \left(\sum_{t=1}^{T} \check{\mathbf{x}}_{it} 1\left(q_{it} \le \widehat{\gamma}\right)\right) \left(\sum_{t=1}^{T} \check{\mathbf{x}}_{it} 1\left(q_{it} > \widehat{\gamma}\right)\right)', \end{aligned}$$

where $N_{12} = \sum_{i=1}^{N} 1(T_{1i} \ge 1, T_{2i} \ge 1)$ with $T_{1i} = \sum_{t=1}^{T} 1(q_{it} \le \widehat{\gamma})$ and $T_{2i} = \sum_{t=1}^{T} 1(q_{it} > \widehat{\gamma})$.

In the proof of Theorem 6, if the model is correctly specified with an error components structure,

$$\begin{aligned} \left(1, -S_{\beta_{11}\theta_{12}} S_{\theta_{12}\theta_{12}}^{-1}\right) W_1 \quad &\sim \quad \varsigma_1 N\left(0, \left(1, -S_{\beta_{11}\theta_{12}} S_{\theta_{12}\theta_{12}}^{-1}\right) \left[\left(1-\rho_1\right) M_1 + \rho_1 \Psi_1\right] \left(1, -S_{\beta_{11}\theta_{12}} S_{\theta_{12}\theta_{12}}^{-1}\right)'\right) \\ &\overset{d}{=} \varsigma_1 N\left(0, \left[\left(1-\rho_1\right) \widetilde{S}_{\beta_{11}\beta_{11}} + \rho_1 \left(1, -S_{\beta_{11}\theta_{12}} S_{\theta_{12}\theta_{12}}^{-1}\right) \Psi_1 \left(1, -S_{\beta_{11}\theta_{12}} S_{\theta_{12}\theta_{12}}^{-1}\right)'\right]\right) \end{aligned}$$

and $\eta^2 = \varsigma_1^2$, so

$$LR_n\left(\gamma_0, \beta_{11}^0\right) \xrightarrow{d} \frac{1}{2} \varrho_{11} \chi_1^2 + \xi(\varphi, \phi),$$

where

$$\varrho_{11} = \left(1-\rho_1\right) + \rho_1 \frac{\left(1, -S_{\beta_{11}\theta_{12}} S_{\theta_{12}\theta_{12}}^{-1}\right) \Psi_1 \left(1, -S_{\beta_{11}\theta_{12}} S_{\theta_{12}\theta_{12}}^{-1}\right)'}{\widetilde{S}_{\beta_{11}\beta_{11}}},$$

and the estimator of ϱ_{11} can be constructed based on $\widehat{\rho}_1$ and $\widehat{\Psi}_1$ by noting that $S_{\beta_{11}\beta_{11}}, S_{\beta_{11}\theta_{12}}$ and $S_{\theta_{12}\theta_{12}}$ are components of M_1. Different from the cross-sectional case where $\Psi_1 = M_1$ such that $\varrho_{11} = 1$, the coefficient before χ_1^2 here is not 1. The same ϱ_{11} appears in the asymptotic distribution of $LR_{1n}\left(\gamma_0, \beta_{11}^0\right)$. In the asymptotic distribution of $LR_{2n}\left(\gamma_0, \beta_{11}^0\right)$, just note that the asymptotic variance matrix of W_1, Ω_1, can be simplified.

Variables Without Threshold Effects

We decompose $\mathbf{x}_{it}$ into $\mathbf{x}_{1it}$ and $\mathbf{x}_{2it}$, where $\mathbf{x}_{1it}$ do not have threshold effects and $\mathbf{x}_{2it}$ do. As a result,

$$\begin{aligned} \mathbb{E}\left[y_{it} | \mathbf{X}_i\right] &= \mathbf{x}_{1it}' \beta_1 + \textstyle\sum_{\ell=1}^{m+1} \left(\mathbf{x}_{2it}' \beta_{\ell 2}' + \mathbf{z}_i' \psi_\ell\right) 1(\gamma_{\ell-1}^0 < q_{it} \le \gamma_\ell^0), \\ &=: \mathbf{x}_{1it}' \beta_1 + \textstyle\sum_{\ell=1}^{m+1} \check{\mathbf{x}}_{2it}' \theta_{\ell 2} 1(\gamma_{\ell-1}^0 < q_{it} \le \gamma_\ell^0), \end{aligned} \tag{32}$$

where we decompose $\check{\mathbf{x}}_{it}$ as $(\mathbf{x}_{1it}', \check{\mathbf{x}}_{2it}')' := (\mathbf{x}_{1it}', \mathbf{x}_{2it}', \mathbf{z}_i')'$ and θ_ℓ as $(\beta_1', \theta_{\ell 2}') = (\beta_1', \beta_{\ell 2}', \psi_\ell')$, and we assume in estimation that

$$\theta_{\ell 2} = \theta_{m+1,2} + \delta_2^\ell \text{ with } \delta_2^\ell = N^{-\kappa} c_2^\ell, \ell = 1, \cdots, m.$$

Implicitly, we assume the augmented variables $\mathbf{z}_i$ have threshold effects. In general, we can consider the case where part of the augmented variables does not have any threshold effect, but this is straightforward and is not pursue here.

Sequential Testing

In testing for linearity, the null changes to $H_0 : \theta_{12} = \theta_{22}$ or $\delta_{\theta_2} = \mathbf{0}$, where $\delta_{\theta_2} = \theta_{12} - \theta_{22}$. Our score test has the null

$$H_0 : \mathbb{E}\left[\sum_{t=1}^{T} \breve{\mathbf{x}}_{2it} e_{it}^{o} 1\left(q_{it} \leq \gamma\right)\right] = \mathbf{0} \text{ for all } \gamma \in \Gamma,$$

where $e_{it}^{o} = y_{it} - \breve{\mathbf{x}}_{it}'\theta_o$ with θ_o being the population coefficient of y_{it} regressed on $\breve{\mathbf{x}}_{it}$. $T_n(\gamma)$ still takes the form of (12) but redefines

$$\widehat{m}_n(\gamma) = N^{-1/2} \sum_{i=1}^{N} \sum_{t=1}^{T} \breve{\mathbf{x}}_{2it} \widehat{e}_{it}^{o} 1\left(q_{it} \leq \gamma\right),$$

and

$$\widehat{H}_n(\gamma) = N^{-1} \sum_{i=1}^{N} \left[\sum_{t=1}^{T} \left(1\left(q_{it} \leq \gamma\right) \breve{\mathbf{x}}_{2it} - \widehat{M}_2(\gamma) \widehat{M}^{-1} \breve{\mathbf{x}}_{it}\right) \widehat{e}_{it}^{o}\right] \left[\sum_{t=1}^{T} \left(1\left(q_{it} \leq \gamma\right) \breve{\mathbf{x}}_{2it} - \widehat{M}_2(\gamma) \widehat{M}^{-1} \breve{\mathbf{x}}_{it}\right) \widehat{e}_{it}^{o}\right]'$$

with

$$\widehat{M}_2(\gamma) = N^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \breve{\mathbf{x}}_{2it} \breve{\mathbf{x}}_{it}' 1\left(q_{it} \leq \gamma\right) \text{ and } \widehat{M} = N^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \breve{\mathbf{x}}_{it} \breve{\mathbf{x}}_{it}',$$

where $\widehat{e}_{it}^{o}$ has the same definition as in Section 3.1. Under $H_{1c} : \delta_{\theta_2} = N^{-1/2} c_{\theta_2}$, $T_n(\gamma)$ converges weakly to

$$H(\gamma)^{-1/2} \left\{\Xi(\gamma) + \left[M_2(\gamma \wedge \gamma_0) - \overline{M}_2(\gamma) M^{-1} \overline{M}_2(\gamma_0)'\right] c_{\theta_2}\right\}$$

on $\gamma \in \Gamma$, where $\Xi(\gamma)$ is a mean zero Gaussian process with covariance kernel

$$H(\gamma_1, \gamma_2) = \mathbb{E}\left[\left(\sum\nolimits_{t=1}^{T} (1\left(q_{it} \leq \gamma_1\right) \breve{\mathbf{x}}_{2it} - \overline{M}_2(\gamma) M^{-1} \breve{\mathbf{x}}_{it}) e_{it}^{0}\right) \left(\sum\nolimits_{t=1}^{T} (1\left(q_{it} \leq \gamma_2\right) \breve{\mathbf{x}}_{2it} - \overline{M}_2(\gamma) M^{-1} \breve{\mathbf{x}}_{it}) e_{it}^{0}\right)'\right],$$

$M_2(\gamma) = \sum_{t=1}^{T} \mathbb{E}\left[\breve{\mathbf{x}}_{2it} \breve{\mathbf{x}}_{2it}' 1\left(q_{it} \leq \gamma\right)\right]$, $\overline{M}_2(\gamma) = \sum_{t=1}^{T} \mathbb{E}\left[\breve{\mathbf{x}}_{2it} \breve{\mathbf{x}}_{it}' 1\left(q_{it} \leq \gamma\right)\right]$, $M = \sum_{t=1}^{T} \mathbb{E}\left[\breve{\mathbf{x}}_{it} \breve{\mathbf{x}}_{it}'\right]$, and $H(\gamma) = H(\gamma, \gamma)$. The simulation procedure in Section 3.1 can be easily adapted to the current scenario.

In sequential testing, e.g., $H_0^{(2)} : m = 1$ vs. $H_1^{(2)} : m = 2$, we need only adjust the notations in $\widehat{m}_n^{(2,1)}(\gamma), \widehat{m}_n^{(2,2)}(\gamma), \widehat{\zeta}^{(2,1)}(\gamma)$, and $\widehat{\zeta}^{(2,2)}(\gamma)$ as above.

Inference on γ

We assume that there are only two regimes in our estimation, so our objective function now changes to

$$S_n(\theta) := \frac{1}{2} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[y_{it} - \mathbf{x}_{1it}'\beta_1 - \breve{\mathbf{x}}_{2it}'\theta_{12} 1\left(q_{it} \leq \gamma\right) - \breve{\mathbf{x}}_{2it}'\theta_{22} 1\left(q_{it} > \gamma\right)\right]^2,$$

where $\theta = \left(\gamma, \underline{\theta}'\right)'$, $\underline{\theta} = (\beta_1', \theta_{12}', \theta_{22}')$ or $(\beta', \psi_1', \psi_2')$ with $\beta' = (\beta_1', \beta_{12}', \beta_{22}')$ being the parameter of interest. Now, the pseudo-true value of $\underline{\theta}$ is

$$\vartheta = \sum_{\ell=1}^{m+1} \overline{M}^{-1} M^{\ell} \begin{pmatrix} \beta_1 \\ \theta_{\ell 2} \end{pmatrix} = \overline{M}^{-1} \left(\sum_{\ell=1}^{m+1} M^{\ell}\right) \begin{pmatrix} \beta_1 \\ \theta_{m+1,2} \end{pmatrix} + \overline{M}^{-1} \sum_{\ell=1}^{m} M^{\ell} \begin{pmatrix} \mathbf{0} \\ \delta_2^{\ell} \end{pmatrix}$$

$$
\begin{aligned}
&= \overline{M}^{-1}\begin{pmatrix} M_{11} & M_{12} \\ M_{12}^{-\prime} & M_{22}^{-} \\ M_{12}^{+\prime} & M_{22}^{+} \end{pmatrix}\begin{pmatrix} \beta_1 \\ \theta_{m+1,2} \end{pmatrix} + \overline{M}^{-1}\begin{pmatrix} \sum_{\ell=1}^{m} M_{12}^{\ell}\delta_2^{\ell} \\ \sum_{\ell=1}^{m_1} M_{22}^{\ell}\delta_2^{\ell} \\ \sum_{\ell=m_1+1}^{m} M_{22}^{\ell}\delta_2^{\ell} \end{pmatrix} \\
&= \begin{pmatrix} \beta_1 \\ \theta_{m+1,2} \\ \theta_{m+1,2} \end{pmatrix} + \overline{M}^{-1}\begin{pmatrix} \sum_{\ell=1}^{m} M_{12}^{\ell}\delta_2^{\ell} \\ \sum_{\ell=1}^{m_1} M_{22}^{\ell}\delta_2^{\ell} \\ \sum_{\ell=m_1+1}^{m} M_{22}^{\ell}\delta_2^{\ell} \end{pmatrix},
\end{aligned}
$$

where

$$
\begin{aligned}
\overline{M} &= \sum_{t=1}^{T}\mathbb{E}\left[\begin{pmatrix} \mathbf{x}_{1it} \\ \check{\mathbf{x}}_{2it}1(q_{it}\le\gamma_0) \\ \check{\mathbf{x}}_{2it}1(q_{it}>\gamma_0) \end{pmatrix}\begin{pmatrix} \mathbf{x}_{1it}', & \check{\mathbf{x}}_{2it}'1(q_{it}\le\gamma_0), & \check{\mathbf{x}}_{2it}'1(q_{it}>\gamma_0) \end{pmatrix}\right] =: \begin{pmatrix} M_{11} & M_{12}^{-} & M_{12}^{+} \\ M_{12}^{-\prime} & M_{22}^{-} & \mathbf{0} \\ M_{12}^{+\prime} & \mathbf{0} & M_{22}^{+} \end{pmatrix}, \\
M^{\ell} &= \sum_{t=1}^{T}\mathbb{E}\left[\begin{pmatrix} \mathbf{x}_{1it} \\ \check{\mathbf{x}}_{2it}1(q_{it}\le\gamma_0) \\ \check{\mathbf{x}}_{2it}1(q_{it}>\gamma_0) \end{pmatrix}\begin{pmatrix} \mathbf{x}_{1it}'1(\gamma_{\ell-1}^0<q_{it}\le\gamma_{\ell}^0), & \check{\mathbf{x}}_{2it}'1(\gamma_{\ell-1}^0<q_{it}\le\gamma_{\ell}^0) \end{pmatrix}\right] \\
&= \begin{cases} \begin{pmatrix} M_{11}^{\ell} & M_{12}^{\ell} \\ M_{12}^{\ell\prime} & M_{22}^{\ell} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} & \text{if } \ell=1,\cdots,m_1, \\ \begin{pmatrix} M_{11}^{\ell} & M_{12}^{\ell} \\ \mathbf{0} & \mathbf{0} \\ M_{12}^{\ell\prime} & M_{22}^{\ell} \end{pmatrix} & \text{if } \ell=m_1+1,\cdots,m+1 \end{cases}
\end{aligned}
$$

with $M_{11}^{\ell}=\sum_{t=1}^{T}\mathbb{E}\left[\check{\mathbf{x}}_{1it}\check{\mathbf{x}}_{1it}'1(\gamma_{\ell-1}^0<q_{it}\le\gamma_{\ell}^0)\right]$, $M_{22}^{\ell}=\sum_{t=1}^{T}\mathbb{E}\left[\check{\mathbf{x}}_{2it}\check{\mathbf{x}}_{2it}'1(\gamma_{\ell-1}^0<q_{it}\le\gamma_{\ell}^0)\right]$, $M_{11}=\sum_{\ell=1}^{m+1}M_{11}^{\ell}$, $M_{12}^{-}=\sum_{\ell=1}^{m_1}M_{12}^{\ell}$, $M_{12}^{+}=\sum_{\ell=m_1+1}^{m+1}M_{12}^{\ell}$, $M_{22}^{-}=\sum_{\ell=1}^{m_1}M_{22}^{\ell}$, $M_{22}^{+}=\sum_{\ell=m_1+1}^{m+1}M_{22}^{\ell}$, and $M_{12}=M_{12}^{-}+M_{12}^{+}$. Note that M^{ℓ} is not a square matrix, so $\overline{M}\neq\sum_{\ell=1}^{m+1}M^{\ell}$, and the β_1 component of ϑ need not be the true β_1. Note also that if we denote the pseudo-true value in the main text as $\widetilde{\beta}_{11},\widetilde{\beta}_{12},\widetilde{\theta}_{12}$, and $\widetilde{\theta}_{22}$, then $\widetilde{\beta}_{11}$ need not equal $\widetilde{\beta}_{12}$, neither of them need equal β_1, and $\widetilde{\theta}_{12}(\widetilde{\theta}_{22})$ need not equal $\theta_{12}(\theta_{22})$. This is substantially different from the scenario in correctly specified models where $\widetilde{\beta}_{11}=\widetilde{\beta}_{12}=\beta_1$, $\widetilde{\theta}_{12}=\theta_{12}$ and $\widetilde{\theta}_{22}=\theta_{22}$. The pseudo-true difference

$$
\theta_{12}-\theta_{22}=(\mathbf{0},\mathbf{I},-\mathbf{I})\,\overline{M}^{-1}\begin{pmatrix} \sum_{\ell=1}^{m} M_{12}^{\ell}\delta_2^{\ell} \\ \sum_{\ell=1}^{m_1} M_{22}^{\ell}\delta_2^{\ell} \\ \sum_{\ell=m_1+1}^{m} M_{22}^{\ell}\delta_2^{\ell} \end{pmatrix} =: N^{-\kappa}c_2,
$$

where

$$
c_2=(\mathbf{0},\mathbf{I},-\mathbf{I})\,\overline{M}^{-1}\begin{pmatrix} \sum_{\ell=1}^{m} M_{12}^{\ell}c_2^{\ell} \\ \sum_{\ell=1}^{m_1} M_{22}^{\ell}c_2^{\ell} \\ \sum_{\ell=m_1+1}^{m} M_{22}^{\ell}c_2^{\ell} \end{pmatrix},
$$

which differs from the c_2 component of c in Assumption 4. Furthermore, the new

$$
c_{-} = \begin{pmatrix} \mathbf{0} \\ c_2^{m_1} \end{pmatrix} - \begin{pmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} \end{pmatrix}\overline{M}^{-1}\begin{pmatrix} \sum_{\ell=1}^{m} M_{12}^{\ell}c_2^{\ell} \\ \sum_{\ell=1}^{m_1} M_{22}^{\ell}c_2^{\ell} \\ \sum_{\ell=m_1+1}^{m} M_{22}^{\ell}c_2^{\ell} \end{pmatrix},
$$

$$c_{+} = \begin{pmatrix} \mathbf{0} \\ c_2^{m_1+1} \end{pmatrix} - \begin{pmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{pmatrix} \overline{M}^{-1} \begin{pmatrix} \sum_{\ell=1}^{m} M_{12}^{\ell} c_2^{\ell} \\ \sum_{\ell=1}^{m_1} M_{22}^{\ell} c_2^{\ell} \\ \sum_{\ell=m_1+1}^{m} M_{22}^{\ell} c_2^{\ell} \end{pmatrix},$$

where note that the β_1 components of $c_{\pm}$ are not zero. In other words, $c'D\left(c-2c_{+}\right)$ and $c'D\left(c+2c_{-}\right)$ cannot be further simplified except $c=(\mathbf{0}', c_2')'$ now. On the other hand, $c'V_\ell c$ can be simplified with c and $\breve{\mathbf{x}}_{it}$ in V_ℓ replaced by c_2 and $\breve{\mathbf{x}}_{2it}$. Correspondingly, η^2, φ and ϕ can be estimated based on

$$\begin{aligned} \eta^2 &\approx \frac{\sum_{t=1}^{T} \mathbb{E}\left[\left(\delta_{2N}'\breve{\mathbf{x}}_{2it}\right)^2\left(y_{it}-\breve{\mathbf{x}}_{it}'\vartheta_1\right)^2 | q_{it}=\gamma_0-\right] f_t(\gamma_0)}{2\sum_{t=1}^{T} \mathbb{E}\left[\left(\delta_{2N}'\breve{\mathbf{x}}_{2it}\right)\left(y_{it}-\breve{\mathbf{x}}_{it}'\overline{\vartheta}\right) | q_{it}=\gamma_0-\right] f_t(\gamma_0)} \\ \varphi &= -\frac{\sum_{t=1}^{T} \mathbb{E}\left[\left(\delta_{2N}'\breve{\mathbf{x}}_{2it}\right)\left(y_{it}-\breve{\mathbf{x}}_{it}'\overline{\vartheta}\right) | q_{it}=\gamma_0+\right] f_t(\gamma_0)}{\sum_{t=1}^{T} \mathbb{E}\left[\left(\delta_{2N}'\breve{\mathbf{x}}_{2it}\right)\left(y_{it}-\breve{\mathbf{x}}_{it}'\overline{\vartheta}\right) | q_{it}=\gamma_0-\right] f_t(\gamma_0)}, \\ \phi &\approx \frac{\sum_{t=1}^{T} \mathbb{E}\left[\left(\delta_{2N}'\breve{\mathbf{x}}_{2it}\right)^2\left(y_{it}-\breve{\mathbf{x}}_{it}'\vartheta_2\right)^2 | q_{it}=\gamma_0+\right] f_t(\gamma_0)}{\sum_{t=1}^{T} \mathbb{E}\left[\left(\delta_{2N}'\breve{\mathbf{x}}_{2it}\right)^2\left(y_{it}-\breve{\mathbf{x}}_{it}'\vartheta_1\right)^2 | q_{it}=\gamma_0-\right] f_t(\gamma_0)}, \end{aligned}$$

where $\delta_N = (\mathbf{0}', \delta_{2N}')'$, $\vartheta_1 = \begin{pmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} \end{pmatrix}\vartheta$, $\vartheta_2 = \begin{pmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} \end{pmatrix}\vartheta$, and $\overline{\vartheta} = \begin{pmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}/2 & \mathbf{I}/2 \end{pmatrix}\vartheta$ can be estimated by their sample analogs.

In correctly specified models, $c'Dc$ can also be simplified with c and $\breve{\mathbf{x}}_{it}$ in D replaced by c_2 and $\breve{\mathbf{x}}_{2it}$. In the error components model, we still have $V_\ell(\gamma) = \varsigma_\ell^2 D(\gamma)$ so that $\eta^2 = \varsigma_1^2$ and $\phi = \varsigma_2^2/\varsigma_1^2$.

Inference on β

The asymptotic variance matrix of $\widehat{\underline{\theta}}$ changes to

$$\Sigma = \overline{M}^{-1}\Omega\overline{M}^{-1}, \tag{33}$$

where $\overline{M}$ is defined above, and

$$\Omega = \mathbb{E}\left[\left(\sum_{t=1}^{T}\begin{pmatrix} \mathbf{x}_{1it} \\ \breve{\mathbf{x}}_{2it}1(q_{it}\le\gamma_0) \\ \breve{\mathbf{x}}_{2it}1(q_{it}>\gamma_0) \end{pmatrix} e_{it}^0\right)\left(\sum_{t=1}^{T}\begin{pmatrix} \mathbf{x}_{1it} \\ \breve{\mathbf{x}}_{2it}1(q_{it}\le\gamma_0) \\ \breve{\mathbf{x}}_{2it}1(q_{it}>\gamma_0) \end{pmatrix} e_{it}^0\right)'\right].$$

We now show the relationship of $\overline{M}$ and Ω with M_ℓ, Ω_ℓ and Ω_{12} and simplifications of $\overline{M}$ and Ω in the correctly specified error components model, which implies that the estimators of $\overline{M}$ and Ω can be derived from Section 4.2 and the last subsection. First, write

$$\overline{M} = \begin{pmatrix} M_{11} & M_{12}^{-} & M_{12}^{+} \\ M_{12}^{-\prime} & M_{22}^{-} & \mathbf{0} \\ M_{12}' & \mathbf{0} & M_{22}^{+} \end{pmatrix}, M_1 = \begin{pmatrix} M_{11}^{-} & M_{12}^{-} \\ M_{12}^{-\prime} & M_{22}^{-} \end{pmatrix} \text{ and } M_2 = \begin{pmatrix} M_{11}^{+} & M_{12}^{+} \\ M_{12}^{+\prime} & M_{22}^{+} \end{pmatrix};$$

then $M_{11} = M_{11}^{-} + M_{11}^{+}$ with $M_{11}^{-} = \sum_{t=1}^{T}\mathbb{E}\left[\breve{\mathbf{x}}_{1it}\breve{\mathbf{x}}_{1it}'1(q_{it}\le\gamma_0)\right] = \sum_{\ell=1}^{m_1} M_{11}^{\ell}$ and $M_{11}^{+} = \sum_{t=1}^{T}\mathbb{E}\left[\breve{\mathbf{x}}_{1it}\breve{\mathbf{x}}_{1it}'1(q_{it}>\gamma_0)\right] = \sum_{\ell=m_1+1}^{m+1} M_{11}^{\ell}$. Second, write

$$\Omega=\begin{pmatrix}\Omega_{11} & \Omega_{12}^{-} & \Omega_{12}^{+}\\ \Omega_{21}^{-} & \Omega_{22}^{-} & \Omega_{22}^{\mp}\\ \Omega_{21}^{+} & \Omega_{22}^{\pm} & \Omega_{22}^{+}\end{pmatrix},\Omega_1=\begin{pmatrix}\Omega_{11}^{-} & \Omega_{12}^{--}\\ \Omega_{21}^{--} & \Omega_{22}^{-}\end{pmatrix}\text{ and }\Omega_2=\begin{pmatrix}\Omega_{11}^{+} & \Omega_{12}^{++}\\ \Omega_{21}^{++} & \Omega_{22}^{+}\end{pmatrix},$$

and

$$\Omega_{12}=\begin{pmatrix}\Omega_{11}^{\mp} & \Omega_{12}^{\mp}\\ \Omega_{21}^{\mp} & \Omega_{22}^{\mp}\end{pmatrix},\Omega_{21}=\Omega_{12}'=\begin{pmatrix}\Omega_{11}^{\pm} & \Omega_{12}^{\pm}\\ \Omega_{21}^{\pm} & \Omega_{22}^{\pm}\end{pmatrix};$$

then

$$\begin{aligned}&\Omega_{11}=\mathbb{E}\left[\left(\textstyle\sum_{t=1}^{T}\mathbf{x}_{1it}e_{it}^{0}\right)\left(\textstyle\sum_{t=1}^{T}\mathbf{x}_{1it}e_{it}^{0}\right)'\right]\\&=\mathbb{E}\left[\left(\textstyle\sum_{t=1}^{T}\mathbf{x}_{1it}1\left(q_{it}\le\gamma_0\right)e_{1it}+\sum_{t=1}^{T}\mathbf{x}_{1it}1\left(q_{it}>\gamma_0\right)e_{2it}\right)\left(\textstyle\sum_{t=1}^{T}\mathbf{x}_{1it}1\left(q_{it}\le\gamma_0\right)e_{1it}+\sum_{t=1}^{T}\mathbf{x}_{1it}1\left(q_{it}>\gamma_0\right)e_{2it}\right)'\right]\\&=\Omega_{11}^{-}+\Omega_{11}^{+}+\Omega_{11}^{\mp}+\Omega_{11}^{\pm},\end{aligned}$$

$$\begin{aligned}\Omega_{12}^{-}&=\mathbb{E}\left[\left(\textstyle\sum_{t=1}^{T}\mathbf{x}_{1it}e_{it}^{0}\right)\left(\textstyle\sum_{t=1}^{T}\check{\mathbf{x}}_{2it}1\left(q_{it}\le\gamma_0\right)e_{1it}\right)'\right]\\&=\mathbb{E}\left[\left(\textstyle\sum_{t=1}^{T}\mathbf{x}_{1it}1\left(q_{it}\le\gamma_0\right)e_{1it}+\sum_{t=1}^{T}\mathbf{x}_{1it}1\left(q_{it}>\gamma_0\right)e_{2it}\right)\left(\textstyle\sum_{t=1}^{T}\check{\mathbf{x}}_{2it}1\left(q_{it}\le\gamma_0\right)e_{1it}\right)'\right]\\&=\Omega_{12}^{--}+\Omega_{12}^{\pm},\end{aligned}$$

and

$$\begin{aligned}\Omega_{12}^{+}&=\mathbb{E}\left[\left(\textstyle\sum_{t=1}^{T}\mathbf{x}_{1it}e_{it}^{0}\right)\left(\textstyle\sum_{t=1}^{T}\check{\mathbf{x}}_{2it}1\left(q_{it}>\gamma_0\right)e_{2it}\right)'\right]\\&=\mathbb{E}\left[\left(\textstyle\sum_{t=1}^{T}\mathbf{x}_{1it}1\left(q_{it}\le\gamma_0\right)e_{1it}+\sum_{t=1}^{T}\mathbf{x}_{1it}1\left(q_{it}>\gamma_0\right)e_{2it}\right)\left(\textstyle\sum_{t=1}^{T}\check{\mathbf{x}}_{2it}1\left(q_{it}>\gamma_0\right)e_{2it}\right)'\right]\\&=\Omega_{12}^{\mp}+\Omega_{12}^{++}.\end{aligned}$$

In the error components model, write

$$\begin{aligned}\Psi_1&=\begin{pmatrix}\Psi_{11}^{-} & \Psi_{12}^{-}\\ \Psi_{12}^{-\prime} & \Psi_{22}^{-}\end{pmatrix},\Psi_2=\begin{pmatrix}\Psi_{11}^{+} & \Psi_{12}^{+}\\ \Psi_{12}^{+\prime} & \Psi_{22}^{+}\end{pmatrix},\\ \Psi_{12}&=\begin{pmatrix}\Psi_{11}^{\mp} & \Psi_{12}^{\mp}\\ \Psi_{12}^{\mp\prime} & \Psi_{22}^{\mp}\end{pmatrix}\text{ and }\Psi_{21}=\Psi_{12}'=\begin{pmatrix}\Psi_{11}^{\pm} & \Psi_{12}^{\pm}\\ \Psi_{12}^{\pm\prime} & \Psi_{22}^{\pm}\end{pmatrix};\end{aligned}$$

then

$$\begin{aligned}\Omega_{11}^{-}&=\varsigma_1^2\left[\left(1-\rho_1\right)M_{11}^{-}+\rho_1\Psi_{11}^{-}\right],\Omega_{11}^{+}=\varsigma_2^2\left[\left(1-\rho_2\right)M_{11}^{+}+\rho_2\Psi_{11}^{+}\right],\\ \Omega_{11}^{\mp}&=c_{12}\Psi_{11}^{\mp},\Omega_{11}^{\pm}=c_{12}\Psi_{11}^{\pm},\\ \Omega_{12}^{--}&=\varsigma_1^2\left[\left(1-\rho_1\right)M_{12}^{-}+\rho_1\Psi_{12}^{-}\right],\Omega_{12}^{\pm}=c_{12}\Psi_{12}^{\pm},\\ \Omega_{12}^{++}&=\varsigma_2^2\left[\left(1-\rho_2\right)M_{12}^{+}+\rho_2\Psi_{12}^{+}\right],\Omega_{12}^{\mp}=c_{12}\Psi_{12}^{\mp},\\ \Omega_{22}^{-}&=\varsigma_1^2\left[\left(1-\rho_1\right)M_{22}^{-}+\rho_1\Psi_{22}^{-}\right],\Omega_{22}^{\mp}=c_{12}\Psi_{22}^{\mp},\\ \Omega_{22}^{\pm}&=c_{12}\Psi_{22}^{\pm},\Omega_{22}^{+}=\varsigma_2^2\left[\left(1-\rho_2\right)M_{22}^{+}+\rho_2\Psi_{22}^{+}\right].\end{aligned}$$

In testing for unobserved individual-specific threshold effects, we need only pay attention to the changes in the asymptotic variance estimate of $\widehat{\psi}_1-\widehat{\psi}_2$.

In adaptive Bonferroni inference on β, suppose we are interested in β_{11} and β_{121}, the first element of

β_1 and β_{12} respectively. It is not hard to show that the result of Theorem 6 still holds, and we need only redefine

$$\varrho_{11} = \frac{\left(1, -S_{12}S_{22}^{-1}\right)\Omega\left(1, -S_{12}S_{22}^{-1}\right)'}{\widetilde{S}_{11}\eta^2},$$

where for β_{11}, S_{12} is the first row of $\overline{M}$ deleting the first element, S_{22} is the submatrix of $\overline{M}$ deleting the first row and the first column, $\widetilde{S}_{11} = S_{11} - S_{12}S_{22}^{-1}S_{21}$ with S_{11} being the $(1,1)$ element of $\overline{M}$, and for β_{121}, S_{12}, S_{22} and $\widetilde{S}_{11}$ are similarly defined but replace the position of β_{11} in $\overline{M}$ by that of β_{121}. In the two alternative inference procedures for β, the results in Theorems 8 and 7 still hold with notations properly adjusted as above. Appendix E details the construction of CIs for β_{11} and β_{121} based on the LR statistics in finite samples.

Testing for Unobserved Individual-Specific Threshold Effects

The null is $\alpha_{\ell i} = \alpha_i$ for $\ell = 1, \cdots, m+1$, i.e., there is no unobserved individual-specific threshold effect. This test aims to assess the validity of the new setup introduced in this paper, which goes beyond those in the existing literature. When m is consistently estimated, this null implies

$$H_0 : \psi_1 = \psi_2$$

in our two-regime model in Section 4.[22] In the general $(m+1)$-regime model, the null changes to $\psi_1 = \psi_2 = \cdots = \psi_{m+1}$, and the analysis is similar but more involved. So we concentrate on the two-regime model here. Although H_0 is only an implication of the null, it is the only relevant one to our estimation in the CRE framework. In other words, the null $\alpha_{\ell i} = \alpha_i$ is refutable but nonverifiable (see Breusch (1986)) in our framework. We will carry out the usual Wald test for this hypothesis; the LR test based on the S_n function in (13) is not easy to implement since its asymptotic null distribution is generally non-standard.[23]

The Wald statistic is

$$W_n = N\left(\widehat{\psi}_1 - \widehat{\psi}_2\right)'\left(\widehat{\Sigma}_{\psi_1} - \widehat{C}_{\psi_1\psi_2} - \widehat{C}_{\psi_2\psi_1} + \widehat{\Sigma}_{\psi_2}\right)^{-1}\left(\widehat{\psi}_1 - \widehat{\psi}_2\right),$$

which converges in distribution to $\chi^2_{d_{\psi_\ell}}$ under the null, where $\widehat{\Sigma}_{\psi_1}$ and $\widehat{C}_{\psi_1\psi_2}$ are consistent estimates of the asymptotic variance matrix of $\widehat{\psi}_1$ and the asymptotic covariance matrix between $\widehat{\psi}_1$ and $\widehat{\psi}_2$ in Section 4.2, and $\widehat{\Sigma}_{\psi_2}$ and $\widehat{C}_{\psi_2\psi_1}$ are similarly defined.[24] Because $\widehat{\psi}_1$ and $\widehat{\psi}_2$ are subvectors of $\widehat{\theta}_1$ and $\widehat{\theta}_2$, we simply extract the corresponding submatrices from $\widehat{\Sigma}_1, \widehat{\Sigma}_2$ and $\widehat{\Sigma}_{12}$.

Quite often, some elements of ψ_ℓ are equal while others are not. The Wald test, as an overall test, may not detect such nuances, so our suggestion is to conduct individual t-tests on each element of $\psi_1 - \psi_2$ if the Wald test is not conclusive.

Details Supporting the Claims in Section 4.2

Section 4.2 claims that estimation based on $\sum_{t=1}^{T}\mathbb{E}\left[\left(\delta_N'\breve{\mathbf{x}}_{it}\right)^2 | q_{it} = \gamma_0\right] f_t(\gamma_0)$ is not valid in misspecified models. On the other hand, estimation of $N^{-2\kappa}c'V_1c$ and $N^{-2\kappa}c'V_2c$ based on the formulae in the correctly specified models is indeed valid.

[22] When m is underestimated, the ψ-components of ϑ_1 and ϑ_2 need not be the same. This is because the ψ-component of ϑ_1 involves $\{\beta_\ell\}_{\ell=1}^{m_1}$, while the ψ-component of ϑ_2 involves $\{\beta_\ell\}_{\ell=m_1+1}^{m+1}$, and the β_ℓ's are not equal.

[23] We can certainly construct the LR test statistics based on the GMM objective functions after writing out the moment conditions for $\underline{\theta}$, but this is not pursued here.

[24] Usually, $\widehat{\Sigma}_{\psi_1} - \widehat{C}_{\psi_1\psi_2} - \widehat{C}_{\psi_2\psi_1} + \widehat{\Sigma}_{\psi_2}$ is invertible, so no generalized inverse is needed.

The first claim is because

$$\begin{aligned}
& 2\mathbb{E}\left[\left(\delta_N'\breve{\mathbf{x}}_{it}\right)\left(y_{it}-\breve{\mathbf{x}}_{it}'\overline{\vartheta}\right)|q_{it}=\gamma_0-\right]-\mathbb{E}\left[\left(\delta_N'\breve{\mathbf{x}}_{it}\right)^2|q_{it}=\gamma_0\right] \\
= \; & 2\mathbb{E}\left[\left(\delta_N'\breve{\mathbf{x}}_{it}\right)\left(y_{it}-\breve{\mathbf{x}}_{it}'\vartheta_1\right)|q_{it}=\gamma_0-\right] \\
= \; & 2\mathbb{E}\left[\left(\delta_N'\breve{\mathbf{x}}_{it}\right)\left(g_1\left(\breve{\mathbf{x}}_{it}\right)-\breve{\mathbf{x}}_{it}'\vartheta_1\right)|q_{it}=\gamma_0-\right]=O\left(N^{-2\kappa}\right),
\end{aligned}$$

and similarly for $-2\sum_{t=1}^T\mathbb{E}\left[\left(\delta_N'\breve{\mathbf{x}}_{it}\right)\left(y_{it}-\breve{\mathbf{x}}_{it}'\overline{\vartheta}\right)|q_{it}=\gamma_0+\right]f_t\left(\gamma_0\right)-\mathbb{E}\left[\left(\delta_N'\breve{\mathbf{x}}_{it}\right)^2|q_{it}=\gamma_0\right]$.

The second claim is because

$$\begin{aligned}
N^{-2\kappa}c'V_1c &= \sum\nolimits_{t=1}^T\mathbb{E}\left[\left(\delta_N'\breve{\mathbf{x}}_{it}\right)^2\left(y_{it}-\breve{\mathbf{x}}_{it}'\vartheta_1\right)^2|q_{it}=\gamma_0-\right]f_t\left(\gamma_0\right)+O\left(N^{-4\kappa}\right), \\
N^{-2\kappa}c'V_2c &= \sum\nolimits_{t=1}^T\mathbb{E}\left[\left(\delta_N'\breve{\mathbf{x}}_{it}\right)^2\left(y_{it}-\breve{\mathbf{x}}_{it}'\vartheta_2\right)^2|q_{it}=\gamma_0+\right]f_t\left(\gamma_0\right)+O\left(N^{-4\kappa}\right),
\end{aligned}$$

where

$$\begin{aligned}
& \sum\nolimits_{t=1}^T\mathbb{E}\left[\left(\delta_N'\breve{\mathbf{x}}_{it}\right)^2\left(y_{it}-\breve{\mathbf{x}}_{it}'\vartheta_1\right)^2|q_{it}=\gamma_0-\right]f_t\left(\gamma_0\right)-N^{-2\kappa}c'V_1c \\
= \; & \sum\nolimits_{t=1}^T\left(\mathbb{E}\left[\left(\delta_N'\breve{\mathbf{x}}_{it}\right)^2\left(y_{it}-\breve{\mathbf{x}}_{it}'\vartheta_1\right)^2|q_{it}=\gamma_0-\right]-\mathbb{E}\left[\left(\delta_N'\breve{\mathbf{x}}_{it}\right)^2e_{1it}^2|q_{it}=\gamma_0-\right]\right)f_t\left(\gamma_0\right) \\
= \; & \sum\nolimits_{t=1}^T\mathbb{E}\left[\left(\delta_N'\breve{\mathbf{x}}_{it}\right)^2\left(g_1\left(\breve{\mathbf{x}}_{it}\right)-\breve{\mathbf{x}}_{it}'\vartheta_1\right)^2|q_{it}=\gamma_0-\right]f_t\left(\gamma_0\right)=O\left(N^{-4\kappa}\right)=o\left(N^{-2\kappa}\right),
\end{aligned}$$

and similarly for $N^{-2\kappa}c'V_2c$. So estimation based on $\sum_{t=1}^T\mathbb{E}\left[\left(\delta_N'\breve{\mathbf{x}}_{it}\right)^2\left(y_{it}-\breve{\mathbf{x}}_{it}'\vartheta_\ell\right)^2|q_{it}=\gamma_0-\right]f_t\left(\gamma_0\right)$, which is valid in correctly specified models, remains valid in misspecified models.[25]

From (18), if the model is homoskedastic in the neighborhood of $q_{it}=\gamma_0$ in the sense that for all $t=1,\cdots,T$,

$$\mathbb{E}\left[\varepsilon_{m_1it}^2|\mathbf{X}_i,q_{it}=\gamma_0-\right]=\sigma_1^2\text{ and }\mathbb{E}\left[\varepsilon_{m_1+1,it}^2|\mathbf{X}_i,q_{it}=\gamma_0+\right]=\sigma_2^2,$$

then

$$\begin{aligned}
N^{-2\kappa}c'V_1c &= \sigma_1^2\sum\nolimits_{t=1}^T\mathbb{E}\left[\left(\delta_N'\breve{\mathbf{x}}_{it}\right)^2|q_{it}=\gamma_0-\right]f_t\left(\gamma_0\right), \\
N^{-2\kappa}c'V_2c &= \sigma_2^2\sum\nolimits_{t=1}^T\mathbb{E}\left[\left(\delta_N'\breve{\mathbf{x}}_{it}\right)^2|q_{it}=\gamma_0+\right]f_t\left(\gamma_0\right),
\end{aligned}$$

and

$$\begin{aligned}
\sigma_1^2 &= \mathbb{E}\left[\left(y_{it}-\breve{\mathbf{x}}_{it}'\vartheta_1\right)^2|q_{it}=\gamma_0-\right]+O\left(N^{-4\kappa}\right), \\
\sigma_2^2 &= \mathbb{E}\left[\left(y_{it}-\breve{\mathbf{x}}_{it}'\vartheta_2\right)^2|q_{it}=\gamma_0+\right]+O\left(N^{-4\kappa}\right)
\end{aligned}$$

for any $t=1,\cdots,T$. So estimation of $N^{-2\kappa}c'V_\ell c$ can be simplified.

Appendix G: Simulations

This appendix examines the performance of our estimation and inferential methods in finite samples with either $\alpha_{1i}\neq\alpha_{2i}$ or $\alpha_{1i}=\alpha_{2i}$. To evaluate estimation performance, following Yu et al. (2024), we use the

[25]Estimation of $N^{-2\kappa}c'V_1c$ can also be based on $\sum_{t=1}^T\mathbb{E}\left[\left(\delta_N'\breve{\mathbf{x}}_{it}\right)^2\left(y_{it}-\breve{\mathbf{x}}_{it}'\overline{\vartheta}\right)^2|q_{it}=\gamma_0-\right]f_t\left(\gamma_0\right)$, which is parallel to $\sum_{t=1}^T\mathbb{E}\left[\left(\delta_N'\breve{\mathbf{x}}_{it}\right)\left(y_{it}-\breve{\mathbf{x}}_{it}'\overline{\vartheta}\right)|q_{it}=\gamma_0-\right]f_t\left(\gamma_0\right)$ in $N^{-2\kappa}c'D\left(c+2c_-\right)$. Similar results apply to $N^{-2\kappa}c'V_2c$.

mean absolute deviation (MAD) to measure the risk for γ, and use the usual root-mean-square error (RMSE) for regular parameters. For comparison, we also show results for the standard fixed effects (FE) estimators as proposed in Hansen (1999). For inference on γ, we compare the CIs based on $LR_n(\gamma)$, $LR_{1n}(\gamma)$ and $LR_{2n}(\gamma)$; for inference on β, we compare the CIs by inverting the traditional t-statistic, $LR_n(\gamma, \beta_{11})$ and $LR_{2n}(\gamma, \beta_{11})$, where the CI based on $LR_{1n}(\gamma, \beta_{11})$ is excluded because it is not practical. Because the performance of the hypothesis test in Sections 7 in similar scenarios is available from the literature, we do not check its performance here. All simulations employ 1000 replications, $T = 2, 5, 10$, and $N = 100, 250, 500$.

We consider the following data generating process (DGP):

$$y_{it} = (q_{it}\beta_1 + \alpha_{1i})\,1(q_{it} \le \gamma) + (q_{it}\beta_2 + \alpha_{2i})\,1(q_{it} > \gamma) + u_{it}, \tag{34}$$

where

$$\alpha_{\ell i} = \overline{q}_i\psi_\ell + a_i,$$

$(q_{i1}, \cdots q_{iT}, a_i, u_{i1}, \cdots, u_{iT})$ are independent of each other and each follows a normal distribution with $q_{it} \sim N(0,1)$, $a_i \sim N(0,1)$, and $u_{it} \sim 0.5N(0,1)$. This DGP implies an error components model with $\varsigma_1^2 = \varsigma_2^2 = 1.25$, $c_1 = c_2 = c_{12} = 1$, and $\rho_1 = \rho_2 = 1/1.25 = 0.8$. We normalize $(\beta_2, \psi_2)' = -0.2 \cdot \mathbf{1}_2$ with $\mathbf{1}_m$ being a column of ones with length m. The parameters of interest are γ and β_1. When $\alpha_{1i} = \alpha_{2i}$, we set $\psi_1 = \psi_2 = -0.2$, $\gamma = 0$, and $\beta_1 = 0.2, 0.5$ and 1, corresponding to small, medium and large threshold effects. The simulation results are collected in Tables 1 and 2. When $\alpha_{1i} \neq \alpha_{2i}$, we set $(\beta_1, \psi_1)' = \Delta \cdot \mathbf{1}_2$, $\gamma = 0$, and $\Delta = \beta_1$ in the $\alpha_{1i} = \alpha_{2i}$ case. The simulation results are collected in Tables 3 and 4. Although the two DGPs contains some special structures, we do not use them in our estimation, i.e., we estimate the models as if $\psi_1 \neq \psi_2$. In our estimation of nuisance parameters in the asymptotic distributions of $\widehat{\gamma}$ and $\widehat{\beta}$, we use the error components structure as in Appendix F, but we do not employ $a_{1i} = a_{2i}$ or $\sigma_1 = \sigma_2$ for further simplification.

According to Tables 1 - 4, the performances of the FE estimators are as expected. Specifically, as shown in Tables 1 and 2, when $\alpha_{1i} = \alpha_{2i}$ so that the standard FE estimation method is valid, larger T and/or N induce smaller biases and risks (using MADs for $\widehat{\gamma}$ and RMSEs for $\widehat{\beta}_1$). The FE estimators and our CRE estimators show similar performances in terms of bias and risk. When $\alpha_{1i} \neq \alpha_{2i}$ so that the FE estimation method is no longer applicable, the FE estimators of γ perform fine and similarly to ours when the difference between α_{1i} and α_{2i} is relatively small (i.e., $\Delta = 0.2$), but perform poorly when the difference becomes larger (i.e., $\Delta = 0.5$, and 1.0) with bias and MAD being noticeably larger than those of ours, as shown in Table 3. The FE estimators of β suffer a more severe performance hit from the presence of heterogeneous individual effects as shown in Table 4. They perform poorly even when $\Delta = 0.2$, and show no sign of convergence as N increases under all settings of Δ.

As for our CRE estimators, two general results apply to all cases. First, larger T and/or N induce smaller biases and risks (MADs for $\widehat{\gamma}$ and RMSEs for $\widehat{\beta}_1$). Second, the lengths of the CIs match the coverages (i.e., higher coverages correspond to longer CIs), and also match the convergence rates of $\widehat{\gamma}$ and $\widehat{\beta}_1$ (i.e., the CIs for β_1 shrink at the rate $\sqrt{N}$ and the CIs for γ shrink faster for a larger Δ value). Thus we report only results that are unique to each case below.

From Tables 1 and 3 we draw the following conclusions. First, the convergence rate of $\widehat{\gamma}$ in Table 1 is usually smaller than N and that in Table 3 is close to N by inspecting the MADs when $N = 250$, 500, 1000. This is because the DGPs in Table 3 contain an extra threshold effect from ψ_1 and ψ_2. Second, among the three CIs for γ, LR_n-CI performs the best in coverage, LR_{1n}-CI tends to over-cover, while LR_{2n}-CI tends to under-cover in some cases. Overall, the coverage of LR_{2n}-CI is acceptable as long as $n = NT$ is not too small, but LR_{1n}-CI always over-covers with the longest width so is not recommended in practice.

From Tables 2 and 4 we draw the following conclusions. First, the convergence rate of $\widehat{\beta}_1$ is roughly $\sqrt{N}$ by inspecting the RMSEs when $N = 250$, 500, 1000. Second, among the three CIs for β_1, the t-CI occasionally under-covers – particularly when threshold effects are small and NT is limited – whereas both LR_n-CI and LR_{2n}-CI tend to over-cover as anticipated, though the extent of over-coverage remains modest. For small threshold effects and NT, the over-coverage of LR_{2n}-CI is less severe, while for large threshold effects or NT, the over-coverage of LR_n-CI is less severe.

As requested by a referee, we also report the simulated expected length of the CIs conditional on the true value of the corresponding parameter being covered in Table 5. As expected, for CIs for γ, the conditional expected lengths are overall slightly, yet still noticeably, larger than the unconditional ones reported in Tables 1-4. For the CIs for β_1, the conditional expected lengths are closer to their unconditional counterparts. In all cases, we observe that the conditional expected length becomes smaller as the the sample sizes N and T increase.

We also study the size and power performance of the sequential tests proposed in Section 3 (CRE tests, hereafter) for determining the number of regimes, i.e., $m + 1$. The simulated selection probabilities for m values under a variety of DGPs are collected in Table 6. For comparison, we also present results for the corresponding tests proposed in Hansen (1999) (FE tests hereafter). We set $\gamma = 0$, $\beta_1 = 0.1$ and 0.5, and set $\psi_1 = -0.4$ and 0.4 when $\alpha_{1i} \neq \alpha_{2i}$, while keeping all other specifications unchanged from the DGPs used in the previous simulations. Therefore, in truth $m = 1$, or, equivalently, the number of regimes is 2. As a result, probabilities recorded in the "$m = 0$" columns of Table 6 represent the simulated likelihood of Type II errors (i.e., 1−power) in testing $H_0 : m = 0$ vs. $H_1 : m = 1$. And probabilities recorded in the "$m = 1$" columns represent the size in testing $H_0 : m = 1$ vs. $H_1 : m = 2$. As shown in the left sector of Table 6, when $\alpha_{1i} = \alpha_{2i}$ so that the FE tests are valid, both the FE and CRE tests perform reasonably well, with the CRE tests providing better size control in some cases where the sample size is small (i.e., $\beta_1 = 0.1$, $T = 2$, $N = 250$ and 500). When $\alpha_{1i} \neq \alpha_{2i}$ so that the FE tests are no longer applicable, the CRE tests keep providing good size control, with most of the probabilities shown in the "$m = 0$" columns being very close to or equal to zero, indicating good power performance. In comparison, the FE tests show noticeable size distortion when $T = 2$. For example, in the particular case where $\beta_1 = 0.1$, $\psi_1 = -0.4$, and $T = 2$, there is severe size distortion (0.566 vs. a targeted size of 0.95) even under a relatively large cross section sample size $N = 1000$. Interestingly, when $T = 5$, the FE tests seem to perform fine, while the CRE tests still provide better size control.

In the empirical application of Section 6, the FE test selects $m = 2$ while the CRE test selects $m = 1$. The simulations here show that the CRE test has better size control when the truth is $m = 1$, so the conclusion that $m = 1$ is more reliable. Although the probability that $m \geq 2$ (1 minus the probabilities of $m = 0$ and 1) is small for both tests under the current DGP specifications, the simulations in Table 5 indicate that it is indeed possible for the CRE test to select $m = 1$ while the FE test select $m \geq 2$ when the individual effect is heterogeneous. To clarify why this occurs, we provide a brief explanation below based on the DGP (34) (which also explains why the FE estimator of γ is more biased than the CRE estimator in Table 3). Assume first $T = 2$ for simplicity. Among the N individuals, there are four cases. First, $q_{i1} \leq \gamma_0$, and $q_{i2} > \gamma_0$, then

$$
\begin{aligned}
\mathbb{E}\left[y_{i1}|\mathbf{X}_i\right] &= q_{i1}\beta_{10} + \overline{q}_i\psi_{10} = q_{i1}\left(\beta_{10} + \frac{\psi_{10}}{2}\right) + q_{i2}\frac{\psi_{10}}{2}, \\
\mathbb{E}\left[y_{i2}|\mathbf{X}_i\right] &= q_{i2}\beta_{20} + \overline{q}_i\psi_{20} = q_{i2}\left(\beta_{20} + \frac{\psi_{20}}{2}\right) + q_{i1}\frac{\psi_{20}}{2}.
\end{aligned}
$$

So following the notation of Hansen (1999),

$$\begin{aligned}
\mathbb{E}\left[y_{i1}^{*}|\mathbf{X}_{i}\right] &= \frac{\mathbb{E}\left[y_{i1}|\mathbf{X}_{i}\right]-\mathbb{E}\left[y_{i2}|\mathbf{X}_{i}\right]}{2}=\frac{q_{i1}}{2}\left(\beta_{10}+\frac{\psi_{10}-\psi_{20}}{2}\right)-\frac{q_{i2}}{2}\left(\beta_{20}-\frac{\psi_{10}-\psi_{20}}{2}\right), \\
&= \left(\beta_{10}+\frac{\psi_{10}-\psi_{20}}{2},\beta_{20}-\frac{\psi_{10}-\psi_{20}}{2}\right)x_{i1}^{*}\left(\gamma_{0}\right), \\
\mathbb{E}\left[y_{i2}^{*}|\mathbf{X}_{i}\right] &= \frac{\mathbb{E}\left[y_{i2}|\mathbf{X}_{i}\right]-\mathbb{E}\left[y_{i1}|\mathbf{X}_{i}\right]}{2}=-\frac{q_{i1}}{2}\left(\beta_{10}+\frac{\psi_{10}-\psi_{20}}{2}\right)+\frac{q_{i2}}{2}\left(\beta_{20}-\frac{\psi_{10}-\psi_{20}}{2}\right) \\
&= \left(\beta_{10}+\frac{\psi_{10}-\psi_{20}}{2},\beta_{20}-\frac{\psi_{10}-\psi_{20}}{2}\right)x_{i2}^{*}\left(\gamma_{0}\right),
\end{aligned}$$

where $x_{i1}^{*}\left(\gamma_{0}\right)=\left(\frac{q_{i1}}{2},-\frac{q_{i2}}{2}\right)'$, and $x_{i2}^{*}\left(\gamma_{0}\right)=\left(-\frac{q_{i1}}{2},\frac{q_{i2}}{2}\right)'$. Second, $q_{i1}>\gamma_{0}$, and $q_{i2}\leq\gamma_{0}$, then

$$\begin{aligned}
\mathbb{E}\left[y_{i1}|\mathbf{X}_{i}\right] &= q_{i1}\beta_{20}+\overline{q}_{i}\psi_{20}=q_{i1}\left(\beta_{20}+\frac{\psi_{20}}{2}\right)+q_{i2}\frac{\psi_{20}}{2}, \\
\mathbb{E}\left[y_{i2}|\mathbf{X}_{i}\right] &= q_{i2}\beta_{10}+\overline{q}_{i}\psi_{10}=q_{i2}\left(\beta_{10}+\frac{\psi_{10}}{2}\right)+q_{i1}\frac{\psi_{10}}{2},
\end{aligned}$$

so

$$\begin{aligned}
\mathbb{E}\left[y_{i1}^{*}|\mathbf{X}_{i}\right] &= \frac{\mathbb{E}\left[y_{i1}|\mathbf{X}_{i}\right]-\mathbb{E}\left[y_{i2}|\mathbf{X}_{i}\right]}{2}=\frac{q_{i1}}{2}\left(\beta_{20}+\frac{\psi_{20}-\psi_{10}}{2}\right)-\frac{q_{i2}}{2}\left(\beta_{10}-\frac{\psi_{20}-\psi_{10}}{2}\right), \\
&= \left(\beta_{10}-\frac{\psi_{20}-\psi_{10}}{2},\beta_{20}+\frac{\psi_{20}-\psi_{10}}{2}\right)x_{i1}^{*}\left(\gamma_{0}\right), \\
\mathbb{E}\left[y_{i2}^{*}|\mathbf{X}_{i}\right] &= \frac{\mathbb{E}\left[y_{i2}|\mathbf{X}_{i}\right]-\mathbb{E}\left[y_{i1}|\mathbf{X}_{i}\right]}{2}=-\frac{q_{i1}}{2}\left(\beta_{20}+\frac{\psi_{20}-\psi_{10}}{2}\right)+\frac{q_{i2}}{2}\left(\beta_{10}-\frac{\psi_{20}-\psi_{10}}{2}\right) \\
&= \left(\beta_{10}-\frac{\psi_{20}-\psi_{10}}{2},\beta_{20}+\frac{\psi_{20}-\psi_{10}}{2}\right)x_{i2}^{*}\left(\gamma_{0}\right),
\end{aligned}$$

where $x_{i1}^{*}\left(\gamma_{0}\right)=\left(-\frac{q_{i2}}{2},\frac{q_{i1}}{2}\right)'$, and $x_{i2}^{*}\left(\gamma_{0}\right)=\left(\frac{q_{i2}}{2},-\frac{q_{i1}}{2}\right)'$. Third, $q_{i1}\leq\gamma_{0}$, and $q_{i2}\leq\gamma_{0}$, then

$$\begin{aligned}
\mathbb{E}\left[y_{i1}|\mathbf{X}_{i}\right] &= q_{i1}\beta_{10}+\overline{q}_{i}\psi_{10}, \\
\mathbb{E}\left[y_{i2}|\mathbf{X}_{i}\right] &= q_{i2}\beta_{10}+\overline{q}_{i}\psi_{10},
\end{aligned}$$

and so

$$\begin{aligned}
\mathbb{E}\left[y_{i1}^{*}|\mathbf{X}_{i}\right] &= \frac{\mathbb{E}\left[y_{i1}|\mathbf{X}_{i}\right]-\mathbb{E}\left[y_{i2}|\mathbf{X}_{i}\right]}{2}=\frac{q_{i1}}{2}\beta_{10}-\frac{q_{i2}}{2}\beta_{10}=\left(\beta_{10},\beta_{20}\right)x_{i1}^{*}\left(\gamma_{0}\right), \\
\mathbb{E}\left[y_{i2}^{*}|\mathbf{X}_{i}\right] &= \frac{\mathbb{E}\left[y_{i2}|\mathbf{X}_{i}\right]-\mathbb{E}\left[y_{i1}|\mathbf{X}_{i}\right]}{2}=-\frac{q_{i1}}{2}\beta_{10}+\frac{q_{i2}}{2}\beta_{10}=\left(\beta_{10},\beta_{20}\right)x_{i2}^{*}\left(\gamma_{0}\right),
\end{aligned}$$

where $x_{i1}^{*}\left(\gamma_{0}\right)=\left(\frac{q_{i1}-q_{i2}}{2},0\right)'$, and $x_{i2}^{*}\left(\gamma_{0}\right)=\left(\frac{q_{i2}-q_{i1}}{2},0\right)'$. Fourth, $q_{i1}>\gamma_{0}$, and $q_{i2}>\gamma_{0}$, then

$$\begin{aligned}
\mathbb{E}\left[y_{i1}|\mathbf{X}_{i}\right] &= q_{i1}\beta_{20}+\overline{q}_{i}\psi_{20}, \\
\mathbb{E}\left[y_{i2}|\mathbf{X}_{i}\right] &= q_{i2}\beta_{20}+\overline{q}_{i}\psi_{20},
\end{aligned}$$

so

$$\mathbb{E}\left[y_{i1}^{*}|\mathbf{X}_{i}\right] = \frac{\mathbb{E}\left[y_{i1}|\mathbf{X}_{i}\right]-\mathbb{E}\left[y_{i2}|\mathbf{X}_{i}\right]}{2}=\frac{q_{i1}}{2}\beta_{20}-\frac{q_{i2}}{2}\beta_{20}=\left(\beta_{10},\beta_{20}\right)x_{i1}^{*}\left(\gamma_{0}\right),$$

$$\mathbb{E}\left[y_{i2}^*|\mathbf{X}_i\right] = \frac{\mathbb{E}\left[y_{i2}|\mathbf{X}_i\right]-\mathbb{E}\left[y_{i1}|\mathbf{X}_i\right]}{2} = -\frac{q_{i1}}{2}\beta_{20}+\frac{q_{i2}}{2}\beta_{20} = \left(\beta_{10},\beta_{20}\right)x_{i2}^*\left(\gamma_0\right),$$

where $x_{i1}^*\left(\gamma_0\right)=\left(0,\frac{q_{i1}-q_{i2}}{2}\right)'$, and $x_{i2}^*\left(\gamma_0\right)=\left(0,\frac{q_{i2}-q_{i1}}{2}\right)'$. In the third and fourth cases, the CREs are the same for y_{i1} and y_{i2} (i.e., $\alpha_{1i}=\alpha_{2i}$), which matches Hansen's setup. However, in the first and second cases, the coefficients of $x_{i1}^*\left(\gamma_0\right)$ and $x_{i2}^*\left(\gamma_0\right)$ are the same (which is good) but different from $\left(\beta_{10},\beta_{20}\right)'$ (which is bad) if $\psi_{10}\neq\psi_{20}$. In other words, the bias of Hansen's method comes from misspecifying the coefficients of $x_{i1}^*\left(\gamma_0\right)$ and $x_{i2}^*\left(\gamma_0\right)$ to be the same for all individuals. This implies that a γ value that is different from γ_0 may have a lower SSR; and in this simple example, different γ's indicate that the proportions of these four cases are different.

The $T=2$ case is too simple because there at least exists a common β such that $\mathbb{E}\left[y_{it}^*|\mathbf{X}_i\right]=\beta' x_{it}^*\left(\gamma_0\right)$ for all t in each case. Generally, this cannot hold. Consider $T=3$, and the case with $\max\left(q_{i1},q_{i2}\right)\leq\gamma_0$, and $q_{i3}>\gamma_0$. Then

$$\begin{aligned}
\mathbb{E}\left[y_{i1}|\mathbf{X}_i\right] &= q_{i1}\beta_{10}+\overline{q}_i\psi_{10} = q_{i1}\left(\beta_{10}+\frac{\psi_{10}}{3}\right)+q_{i2}\frac{\psi_{10}}{3}+q_{i3}\frac{\psi_{10}}{3},\\
\mathbb{E}\left[y_{i2}|\mathbf{X}_i\right] &= q_{i2}\beta_{10}+\overline{q}_i\psi_{10} = q_{i1}\frac{\psi_{10}}{3}+q_{i2}\left(\beta_{10}+\frac{\psi_{10}}{3}\right)+q_{i3}\frac{\psi_{10}}{3},\\
\mathbb{E}\left[y_{i3}|\mathbf{X}_i\right] &= q_{i3}\beta_{20}+\overline{q}_i\psi_{20} = q_{i1}\frac{\psi_{20}}{3}+q_{i2}\frac{\psi_{20}}{3}+q_{i3}\left(\beta_{20}+\frac{\psi_{20}}{3}\right),
\end{aligned}$$

so

$$\begin{aligned}
\mathbb{E}\left[y_{i1}^*|\mathbf{X}_i\right] &= \frac{2\mathbb{E}\left[y_{i1}|\mathbf{X}_i\right]-\mathbb{E}\left[y_{i2}|\mathbf{X}_i\right]-\mathbb{E}\left[y_{i3}|\mathbf{X}_i\right]}{3}\\
&= \frac{q_{i1}}{3}\left(2\beta_{10}+\frac{\psi_{10}-\psi_{20}}{3}\right)-\frac{q_{i2}}{3}\left(\beta_{10}-\frac{\psi_{10}-\psi_{20}}{3}\right)-\frac{q_{i3}}{3}\left(\beta_{20}-\frac{\psi_{10}-\psi_{20}}{3}\right),\\
&= \left(\beta_{10}-\frac{\psi_{10}-\psi_{20}}{3},\beta_{20}-\frac{\psi_{10}-\psi_{20}}{3}\right)x_{i1}^*\left(\gamma_0\right)+\frac{q_{i1}}{3}\left(\psi_{10}-\psi_{20}\right),\\
\mathbb{E}\left[y_{i2}^*|\mathbf{X}_i\right] &= \frac{2\mathbb{E}\left[y_{i2}|\mathbf{X}_i\right]-\mathbb{E}\left[y_{i1}|\mathbf{X}_i\right]-\mathbb{E}\left[y_{i3}|\mathbf{X}_i\right]}{3}\\
&= -\frac{q_{i1}}{3}\left(\beta_{10}-\frac{\psi_{10}-\psi_{20}}{3}\right)+\frac{q_{i2}}{3}\left(2\beta_{10}+\frac{\psi_{10}-\psi_{20}}{3}\right)-\frac{q_{i3}}{3}\left(\beta_{20}-\frac{\psi_{10}-\psi_{20}}{3}\right)\\
&= \left(\beta_{10}-\frac{\psi_{10}-\psi_{20}}{3},\beta_{20}-\frac{\psi_{10}-\psi_{20}}{3}\right)x_{i2}^*\left(\gamma_0\right)+\frac{q_{i2}}{3}\left(\psi_{10}-\psi_{20}\right),\\
\mathbb{E}\left[y_{i3}^*|\mathbf{X}_i\right] &= \frac{2\mathbb{E}\left[y_{i3}|\mathbf{X}_i\right]-\mathbb{E}\left[y_{i1}|\mathbf{X}_i\right]-\mathbb{E}\left[y_{i2}|\mathbf{X}_i\right]}{3}\\
&= -\frac{q_{i1}}{3}\left(\beta_{10}+2\frac{\psi_{10}-\psi_{20}}{3}\right)-\frac{q_{i2}}{3}\left(\beta_{10}+2\frac{\psi_{10}-\psi_{20}}{3}\right)+\frac{q_{i3}}{3}\left(2\beta_{20}-2\frac{\psi_{10}-\psi_{20}}{3}\right)\\
&= \left(\beta_{10}-\frac{\psi_{10}-\psi_{20}}{3},\beta_{20}-\frac{\psi_{10}-\psi_{20}}{3}\right)x_{i3}^*\left(\gamma_0\right)-\frac{q_{i1}+q_{i2}}{3}\left(\psi_{10}-\psi_{20}\right),
\end{aligned}$$

where $x_{i1}^*\left(\gamma_0\right)=\left(\frac{2q_{i1}-q_{i2}}{3},-\frac{q_{i3}}{3}\right)'$, $x_{i2}^*\left(\gamma_0\right)=\left(\frac{2q_{i2}-q_{i1}}{3},-\frac{q_{i3}}{3}\right)'$, and $x_{i3}^*\left(\gamma_0\right)=\left(-\frac{q_{i1}+q_{i2}}{3},\frac{2q_{i3}}{3}\right)'$. Obviously, besides $\beta' x_{it}^*\left(\gamma_0\right)$ with a common β, there are extra terms in $\mathbb{E}\left[y_{it}^*|\mathbf{X}_i\right]$, i.e., the endogeneity effects of $\alpha_{\ell i}$ cannot be completely absorbed by $\beta' x_{it}^*\left(\gamma_0\right)$ with a common β even in each case of $\left(q_{i1},q_{i2},q_{i3}\right)$ values. It is of course less likely that $\text{plim}(\widehat{\gamma})=\gamma_0$. Given that $\text{plim}(\widehat{\gamma})=\gamma\neq\gamma_0$, say, $\gamma>\gamma_0$, it seems that the data with $q_{it}\leq\widehat{\gamma}$ still contain threshold effects, so Hansen's model selection method may overestimate m. For a general T, there are in total 2^T cases, so analysis is inevitably messy, but the key point is the same as in this simple example.

Additional References:

Banerjee, M. and I.W. McKeague, 2007, Confidence Sets for Split Points in Decision Trees, *Annals of Statistics*, 35, 543-574.

Breusch, T.S., 1986, Hypothesis Testing in Unidentified Models, *Review of Economic Studies*, 53, 635-651.

Kim, J. and D. Pollard, 1990, Cube Root Asymptotics, *Annals of Statistics*, 18, 191-219.

Van der Vaart, A.W. and J.A. Wellner, 2023, *Weak Convergence and Empirical Processes: With Applications to Statistics*, 2nd edition, New York: Springer.

Yu, P., 2014, The Bootstrap in Threshold Regression, *Econometric Theory*, 30, 676-714.

Table 1: Estimation and CI for γ ($\alpha_{1i} = \alpha_{2i}$)

		Est. FE		Estimation		CI: length			CI: coverage prob.		
T	N	Bias	MAD	Bias	MAD	LR_n	LR_{1n}	LR_{2n}	LR_n	LR_{1n}	LR_{2n}
						$\beta_1 = 0.2$					
2	250	-0.0075	0.3290	0.0066	0.1785	0.6005	0.6761	0.5591	0.956	0.981	0.905
	500	-0.0170	0.2379	0.0016	0.1482	0.4905	0.5052	0.4725	0.973	0.978	0.966
	1000	0.0054	0.1740	0.0018	0.1157	0.3858	0.3968	0.3755	0.966	0.977	0.962
5	250	-0.0027	0.1781	-0.0032	0.1424	0.4479	0.4613	0.4336	0.967	0.972	0.958
	500	0.0093	0.1328	-0.0141	0.1160	0.3606	0.3674	0.3505	0.970	0.981	0.963
	1000	0.0063	0.0949	-0.0052	0.0933	0.2813	0.2861	0.2735	0.952	0.964	0.948
10	250	-0.0032	0.1149	0.0088	0.115	0.3612	0.3723	0.3503	0.966	0.983	0.968
	500	0.0045	0.0956	0.0049	0.0873	0.2875	0.2938	0.2805	0.980	0.987	0.979
	1000	-0.0047	0.0705	-0.0074	0.0714	0.2216	0.2253	0.2186	0.969	0.971	0.965
						$\beta_1 = 0.5$					
2	250	-0.0159	0.1951	0.0017	0.1282	0.4117	0.4265	0.3981	0.955	0.969	0.934
	500	0.0019	0.1390	-0.0068	0.1026	0.3391	0.3472	0.3316	0.981	0.985	0.981
	1000	0.0062	0.1061	-0.0026	0.0791	0.2630	0.2679	0.2579	0.972	0.976	0.965
5	250	-0.0075	0.1079	-0.0071	0.0979	0.3062	0.3121	0.3009	0.984	0.985	0.976
	500	-0.0018	0.0769	-0.0051	0.0750	0.2486	0.2514	0.2442	0.975	0.973	0.975
	1000	0.0028	0.0624	0.0019	0.0613	0.1958	0.1982	0.1908	0.968	0.970	0.953
10	250	-0.0053	0.0747	0.0057	0.0971	0.2514	0.2554	0.2482	0.956	0.963	0.931
	500	0.0018	0.0588	-0.0014	0.0470	0.1958	0.1983	0.1925	0.969	0.972	0.956
	1000	0.0027	0.0457	0.0009	0.0214	0.1509	0.1523	0.1498	0.945	0.956	0.943
						$\beta_1 = 1.0$					
2	250	0.0067	0.1224	0.0011	0.0877	0.2873	0.2954	0.2803	0.958	0.967	0.939
	500	0.0134	0.0906	-0.0004	0.0683	0.2320	0.2350	0.2281	0.989	0.985	0.978
	1000	-0.0023	0.0678	-0.0003	0.0548	0.1872	0.1880	0.1848	0.988	0.990	0.984
5	250	0.0089	0.0671	0.0009	0.0643	0.2178	0.2210	0.2142	0.990	0.994	0.986
	500	0.0029	0.0554	-0.0043	0.0529	0.1707	0.1721	0.1678	0.985	0.984	0.976
	1000	-0.0014	0.0435	0.0017	0.0411	0.1348	0.1357	0.1336	0.970	0.968	0.965
10	250	-0.0015	0.0481	0.0038	0.0490	0.1664	0.1681	0.1652	0.980	0.982	0.980
	500	0.0007	0.0389	0.0005	0.0397	0.1342	0.1359	0.1332	0.977	0.990	0.976
	1000	0.0022	0.0293	-0.0007	0.0311	0.1068	0.1076	0.1059	0.981	0.982	0.981

Note: The confidence level is targeted at 0.95.

Table 2: Estimation and CI for β_1 ($\alpha_{1i} = \alpha_{2i}$)

		Est. FE		Estimation		CI: length			CI: coverage prob.		
T	N	Bias	RMSE	Bias	RMSE	LR_n	LR_{1n}	LR_{2n}	LR_n	LR_{1n}	LR_{2n}
						$\beta_1 = 0.2$					
2	250	-0.0141	0.0299	0.0002	0.0210	0.0724	0.1235	0.0942	0.933	0.994	0.970
	500	-0.0097	0.0210	-0.0004	0.0146	0.0514	0.0865	0.0668	0.928	1.000	0.969
	1000	-0.0052	0.0154	0.0011	0.0094	0.0363	0.0608	0.0471	0.952	0.990	0.975
5	250	-0.0042	0.0145	0.0008	0.0118	0.0472	0.0650	0.0703	0.945	1.000	1.000
	500	-0.0030	0.0104	0.0003	0.0085	0.0334	0.0459	0.0500	0.955	0.983	0.990
	1000	-0.0024	0.0070	0.0007	0.0062	0.0237	0.0324	0.0353	0.958	0.991	0.995
10	250	-0.0027	0.0093	0.0005	0.0107	0.0427	0.0505	0.0630	0.976	0.981	1.000
	500	-0.0023	0.0066	-0.0005	0.0077	0.0302	0.0356	0.0443	0.943	0.990	0.998
	1000	-0.0013	0.0047	0.0003	0.0052	0.0214	0.0251	0.0316	0.960	0.993	0.995
						$\beta_1 = 0.5$					
2	250	-0.0120	0.0293	-0.0003	0.0206	0.0721	0.1215	0.0940	0.946	0.997	0.970
	500	-0.0075	0.0194	-0.0006	0.0145	0.0515	0.0856	0.0668	0.928	1.000	0.975
	1000	-0.0038	0.0145	0.0010	0.0092	0.0363	0.0602	0.0471	0.950	0.988	0.981
5	250	-0.0028	0.0133	0.0007	0.0117	0.0474	0.0648	0.0705	0.945	0.993	1.000
	500	-0.0018	0.0100	0.0003	0.0084	0.0335	0.0457	0.0500	0.956	0.984	0.993
	1000	-0.0021	0.0068	0.0006	0.0062	0.0237	0.0323	0.353	0.957	0.990	0.996
10	250	-0.0022	0.0087	0.0005	0.0107	0.0428	0.0504	0.0631	0.970	0.979	0.996
	500	0.0017	0.0060	-0.0005	0.0078	0.0302	0.0356	0.0444	0.941	0.990	1.000
	1000	-0.0011	0.0045	0.0002	0.0052	0.0214	0.0252	0.0316	0.958	0.989	0.999
						$\beta_1 = 1.0$					
2	250	-0.0097	0.0269	-0.0004	0.0200	0.0721	0.1202	0.0939	0.936	0.998	0.975
	500	-0.0062	0.0182	-0.0006	0.0142	0.0513	0.0849	0.0666	0.935	1.000	0.975
	1000	-0.0026	0.0137	0.0009	0.0091	0.0362	0.0598	0.0470	0.955	0.988	0.983
5	250	-0.0016	0.0133	0.0004	0.0115	0.0474	0.0646	0.0704	0.938	0.994	1.000
	500	-0.0016	0.0098	0.0003	0.0084	0.0335	0.0457	0.0500	0.959	0.985	0.992
	1000	-0.0019	0.0069	0.0006	0.0062	0.0237	0.0323	0.0353	0.957	0.990	1.000
10	250	-0.0017	0.0085	0.0004	0.0107	0.0429	0.0504	0.0632	0.973	0.982	0.980
	500	-0.0015	0.0060	-0.0005	0.0078	0.0303	0.0356	0.0445	0.941	0.995	0.996
	1000	-0.0009	0.0045	-0.0004	0.0051	0.0214	0.0252	0.0316	0.960	0.995	0.997

Note: The confidence level is targeted at 0.95.

Table 3: Estimation and CI for γ ($\alpha_{1i} \neq \alpha_{2i}$)

		Est. FE		Estimation		CI: length			CI: coverage prob.		
T	N	Bias	MAD	Bias	MAD	LR_n	LR_{1n}	LR_{2n}	LR_n	LR_{1n}	LR_{2n}
						$\Delta = 0.2$					
2	250	0.0109	0.3329	0.0476	0.3001	1.2683	2.7688	0.9014	0.946	1.000	0.837
	500	-0.0082	0.2684	0.0052	0.2287	0.8458	1.7535	0.6369	0.965	0.992	0.881
	1000	0.0057	0.1979	0.0056	0.1322	0.5434	0.8247	0.4522	0.961	0.976	0.910
5	250	0.0200	0.1648	-0.0182	0.2325	0.9989	2.1280	0.7364	0.973	0.990	0.895
	500	0.0159	0.1191	-0.0273	0.1336	0.6393	0.9978	0.5027	0.982	1.000	0.933
	1000	0.0041	0.0900	0.0011	0.0658	0.3961	0.5080	0.3355	0.961	0.968	0.952
10	250	-0.0181	0.2296	-0.0175	0.2447	0.8081	1.3728	0.6483	0.923	0.983	0.880
	500	0.0032	0.1550	0.0017	0.1366	0.5625	0.7390	0.4651	0.937	0.986	0.927
	1000	-0.0057	0.0691	0.0073	0.0641	0.3330	0.3896	0.2824	0.981	0.988	0.953
						$\Delta = 0.5$					
2	250	0.0040	0.2519	0.0268	0.1507	0.6347	1.1172	0.5038	0.953	0.982	0.895
	500	0.0074	0.1914	0.0006	0.0809	0.4060	0.5371	0.3469	0.957	0.979	0.956
	1000	0.0059	0.1032	-0.0101	0.0417	0.2251	0.2615	0.2032	0.951	0.958	0.939
5	250	0.0205	0.1127	0.0018	0.0987	0.4980	0.6848	0.4056	0.967	0.992	0.925
	500	0.0114	0.0848	-0.0024	0.0475	0.2559	0.2982	0.2229	0.985	0.986	0.963
	1000	0.096	0.0645	-0.0031	0.0259	0.1464	0.1655	0.1318	0.946	0.955	0.941
10	250	-0.0067	0.0768	0.0057	0.0971	0.3884	0.5177	0.3394	0.956	0.963	0.931
	500	0.0070	0.0598	-0.0014	0.0470	0.2426	0.2771	0.2076	0.969	0.972	0.956
	1000	-0.0065	0.0565	0.0009	0.0214	0.1144	0.1232	0.1074	0.945	0.956	0.943
						$\Delta = 1.0$					
2	250	0.0028	0.1076	0.0063	0.0547	0.2769	0.3697	0.2405	0.921	0.930	0.905
	500	0.0023	0.0848	0.0004	0.0314	0.1548	0.1672	0.1414	0.958	0.964	0.957
	1000	-0.0029	0.0699	-0.0040	0.0187	0.0770	0.0819	0.0732	0.955	0.961	0.952
5	250	0.0118	0.0875	0.0039	0.0432	0.1891	0.2221	0.1722	0.967	0.978	0.938
	500	0.0061	0.0685	-0.0005	0.0164	0.0943	0.1006	0.0884	0.965	0.979	0.960
	1000	0.0057	0.0615	-0.0008	0.0102	0.0469	0.0499	0.0447	0.930	0.926	0.925
10	250	-0.0039	0.0588	0.0038	0.0350	0.1662	0.1841	0.1499	0.963	0.980	0.961
	500	-0.0043	0.0445	0.0022	0.0149	0.0833	0.0883	0.0797	0.953	0.962	0.939
	1000	-0.0031	0.0364	-0.0011	0.0078	0.0403	0.0414	0.0391	0.910	0.913	0.910

Note: The confidence level is targeted at 0.95.

Table 4: Estimation and CI for β_1 ($\alpha_{1i} \neq \alpha_{2i}$)

		Est. FE		Estimation		CI: length			CI: coverage prob.		
T	N	Bias	RMSE	Bias	RMSE	LR_n	LR_{1n}	LR_{2n}	LR_n	LR_{1n}	LR_{2n}
						$\Delta = 0.2$					
2	250	0.0865	0.0911	-0.0014	0.1210	0.3687	0.6827	0.4784	0.896	0.991	0.958
	500	0.0920	0.0942	-0.0147	0.0783	0.2576	0.4430	0.3352	0.913	1.000	0.963
	1000	0.0955	0.0968	-0.0028	0.0473	0.1840	0.3081	0.2374	0.950	0.998	0.989
5	250	0.0556	0.0588	0.0103	0.0609	0.2307	0.3328	0.3509	0.938	0.996	1.000
	500	0.0575	0.0589	-0.0033	0.0422	0.1667	0.2321	0.2490	0.941	0.993	0.997
	1000	0.0577	0.0584	0.0009	0.0298	0.1180	0.1621	0.1763	0.952	0.995	0.990
10	250	0.0372	0.0397	-0.0041	0.0546	0.2063	0.2502	0.3121	0.945	0.973	0.995
	500	0.0379	0.0390	-0.0021	0.0370	0.1495	0.1781	0.2220	0.938	0.982	1.000
	1000	0.0386	0.0393	0.0024	0.0261	0.1063	0.1258	0.1570	0.940	0.980	0.993
						$\Delta = 0.5$					
2	250	0.1620	0.1655	-0.0049	0.1039	0.3623	0.6125	0.4712	0.922	0.995	0.972
	500	0.1681	0.1698	-0.0137	0.0707	0.2558	0.4268	0.3326	0.938	1.000	0.975
	1000	0.1715	0.1725	-0.0025	0.0442	0.1818	0.3002	0.2361	0.962	0.997	0.993
5	250	0.0863	0.0884	0.0081	0.0587	0.2339	0.3219	0.3511	0.943	0.994	1.000
	500	0.0880	0.0890	-0.0034	0.0421	0.1671	0.2279	0.2490	0.946	0.995	0.996
	1000	0.0880	0.0885	-0.0008	0.0302	0.1184	0.1613	0.1766	0.954	0.992	0.995
10	250	0.0528	0.0545	-0.0070	0.0571	0.2114	0.2498	0.3143	0.962	0.967	0.994
	500	0.0532	0.0540	-0.0023	0.0372	0.1515	0.1782	0.2232	0.957	0.985	0.995
	1000	0.0538	0.0542	0.0026	0.0262	0.1067	0.1254	0.1574	0.946	0.980	0.990
						$\Delta = 1.0$					
2	250	0.2872	0.2907	-0.0060	0.0956	0.3604	0.5976	0.4698	0.930	0.995	0.987
	500	0.2936	0.2954	-0.0104	0.0689	0.2557	0.4209	0.3322	0.937	1.000	0.985
	1000	0.2962	0.2971	0.0005	0.0441	0.1816	0.2976	0.2357	0.961	1.000	0.991
5	250	0.1164	0.1185	0.0093	0.0593	0.2354	0.3209	0.3523	0.962	0.986	1.000
	500	0.1182	0.1192	-0.0025	0.0418	0.1673	0.2274	0.2490	0.957	0.995	0.996
	1000	0.1181	0.1186	-0.0003	0.0301	0.1185	0.1610	0.1768	0.945	0.993	0.995
10	250	0.0579	0.0595	-0.0070	0.0574	0.2124	0.2502	0.3152	0.961	0.975	0.990
	500	0.0583	0.0591	-0.0019	0.0373	0.1516	0.1781	0.2232	0.952	0.983	0.990
	1000	0.0587	0.0591	0.0028	0.0263	0.1068	0.1254	0.1575	0.946	0.980	0.992

Note: The confidence level is targeted at 0.95.

Table 5: Average length of CI conditioned on the true value being covered

		$\alpha_{1i} = \alpha_{2i}$						$\alpha_{1i} \neq \alpha_{2i}$					
		CI for γ			CI for β_1			CI for γ			CI for β_1		
T	N	LR_n	LR_{1n}	LR_{2n}	LR_n	LR_{1n}	LR_{2n}	LR_n	LR_{1n}	LR_{2n}	LR_n	LR_{1n}	LR_{2n}
		$\beta_1 = 0.2$						$\Delta = 0.2$					
2	250	0.6199	0.6838	0.5914	0.0724	0.1235	0.0942	1.3105	2.7841	0.9762	0.3689	0.6827	0.4784
	500	0.4982	0.5115	0.4822	0.0515	0.0865	0.0668	0.8631	1.7676	0.6795	0.2577	0.4430	0.3348
	1000	0.3936	0.4022	0.3842	0.0365	0.0609	0.0471	0.5549	0.8407	0.4658	0.1840	0.3081	0.2374
5	250	0.4597	0.46716	0.4472	0.0472	0.0651	0.0703	1.0191	2.1444	0.7669	0.2307	0.3328	0.3509
	500	0.3663	0.3719	0.3585	0.0335	0.0459	0.0500	0.6490	1.0017	0.5153	0.1665	0.2322	0.2490
	1000	0.2907	0.2932	0.2848	0.0237	0.0325	0.0353	0.4053	0.5153	0.3446	0.1179	0.1621	0.1763
10	250	0.3696	0.3772	0.3594	0.0427	0.0505	0.0630	0.8254	1.3775	0.6804	0.2077	0.2507	0.3129
	500	0.2907	0.2958	0.2848	0.0303	0.0356	0.0443	0.5690	0.7382	0.4746	0.1493	0.1780	0.2220
	1000	0.2259	0.2289	0.2235	0.0214	0.0251	0.0316	0.3347	0.3901	0.2859	0.1063	0.1258	0.1570
		$\beta_1 = 0.5$						$\Delta = 0.5$					
2	250	0.4241	0.4354	0.4140	0.0720	0.1215	0.0940	0.6537	1.1316	0.5264	0.3626	0.6125	0.4707
	500	0.3441	0.3515	0.3369	0.0516	0.0856	0.0668	0.4132	0.5427	0.3536	0.2558	0.4268	0.3325
	1000	0.2672	0.2713	0.2623	0.0363	0.0602	0.0471	0.2272	0.2643	0.2051	0.1818	0.3002	0.2361
5	250	0.3106	0.3157	0.3058	0.0475	0.0648	0.0705	0.5037	0.6894	0.4104	0.2339	0.3219	0.3510
	500	0.2524	0.2553	0.2480	0.0335	0.0458	0.0500	0.2572	0.2999	0.2245	0.1672	0.2279	0.2490
	1000	0.2003	0.2021	0.1979	0.0237	0.0323	0.353	0.1453	0.1656	0.1306	0.1184	0.1614	0.1767
10	250	0.2541	0.2568	0.2514	0.0428	0.0504	0.0631	0.3946	0.5215	0.3463	0.2114	0.2498	0.3143
	500	0.1964	0.1990	0.1931	0.0302	0.0356	0.0444	0.2457	0.2787	0.2100	0.1515	0.1782	0.2232
	1000	0.1535	0.1549	0.1525	0.0214	0.0252	0.0316	0.1160	0.1250	0.1085	0.1068	0.1254	0.1574
		$\beta_1 = 1.0$						$\Delta = 1.0$					
2	250	0.2962	0.3021	0.2903	0.0721	0.1202	0.0940	0.2818	0.3741	0.2479	0.3609	0.5976	0.4699
	500	0.2338	0.2368	0.2308	0.0514	0.0849	0.0666	0.1571	0.1694	0.1437	0.2557	0.4209	0.3322
	1000	0.1876	0.1892	0.1857	0.0362	0.0598	0.0470	0.0777	0.0826	0.0740	0.1816	0.2976	0.2357
5	250	0.2187	0.2212	0.2159	0.0475	0.0646	0.0704	0.1907	0.2230	0.1730	0.2354	0.3209	0.3523
	500	0.1723	0.1736	0.1703	0.0335	0.0457	0.0500	0.0956	0.1014	0.0896	0.1673	0.2275	0.2490
	1000	0.1376	0.1385	0.1369	0.0237	0.0323	0.0353	0.0473	0.0504	0.0451	0.1185	0.1610	0.1768
10	250	0.1687	0.1704	0.1674	0.0429	0.0504	0.0632	0.1684	0.1870	0.1509	0.2125	0.2503	0.3152
	500	0.1359	0.1364	0.1349	0.0303	0.0356	0.0445	0.0837	0.0894	0.0801	0.1516	0.1781	0.2232
	1000	0.1075	0.1083	0.1066	0.0214	0.0252	0.0316	0.0414	0.0420	0.0402	0.1068	0.1254	0.1575

Note: The confidence level is targeted at 0.95.

Table 6: Selection Probabilities for m Values (i.e., # of regimes $= m+1$)

T	N	**Homogeneity DGPs (i.e., $\alpha_{1i} = \alpha_{2i}$)**				**Heterogeneity DGPs (i.e., $\alpha_{1i} \neq \alpha_{2i}$)**							
		FE		CRE		FE		CRE		FE		CRE	
		$m=0$	$m=1$	$m=0$	$m=1$	$m=0$	$m=1$	$m=0$	$m=1$	$m=0$	$m=1$	$m=0$	$m=1$
		$\beta_1 = 0.1$				$\beta_1 = 0.1$ and $\psi_1 = -0.4$				$\beta_1 = 0.1$ and $\psi_1 = 0.4$			
2	250	0.630	0.356	0.284	0.698	0.812	0.176	0.610	0.386	0.114	0.766	0.000	0.980
	500	0.310	0.656	0.034	0.960	0.710	0.282	0.256	0.850	0.000	0.896	0.000	0.962
	1000	0.062	0.894	0.000	0.946	0.406	0.566	0.028	0.938	0.000	0.920	0.000	0.946
5	250	0.052	0.922	0.034	0.950	0.134	0.834	0.066	0.908	0.000	0.966	0.000	0.952
	500	0.000	0.934	0.000	0.946	0.022	0.920	0.000	0.946	0.000	0.972	0.000	0.946
	1000	0.000	0.946	0.000	0.956	0.000	0.922	0.000	0.952	0.000	0.922	0.000	0.938
		$\beta_1 = 0.5$				$\beta_1 = 0.5$ and $\psi_1 = -0.4$				$\beta_1 = 0.5$ and $\psi_1 = 0.4$			
2	250	0.022	0.940	0.000	0.982	0.096	0.872	0.000	0.988	0.000	0.984	0.000	0.978
	500	0.000	0.940	0.000	0.972	0.000	0.936	0.000	0.964	0.000	0.930	0.000	0.960
	1000	0.000	0.938	0.000	0.938	0.000	0.936	0.000	0.938	0.000	0.926	0.000	0.952
5	250	0.000	0.966	0.000	0.958	0.000	0.954	0.000	0.950	0.000	0.916	0.000	0.958
	500	0.000	0.962	0.000	0.956	0.000	0.960	0.000	0.942	0.000	0.934	0.000	0.948
	1000	0.000	0.928	0.000	0.946	0.000	0.930	0.000	0.946	0.000	0.938	0.000	0.956

Note: The significance level is set at 0.05.