2-2021

Panel Threshold Regression with Unobserved Individual-Specific Threshold Effects

Ping Yu

Peter C. B. Phillips

Shengjie Hong

PANEL THRESHOLD REGRESSION WITH UNOBSERVED INDIVIDUAL-SPECIFIC THRESHOLD EFFECTS

By

Ping Yu, Shengjie Hong and Peter C. B. Phillips

Month Year

COWLES FOUNDATION DISCUSSION PAPER NO. 2352

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

Panel Threshold Regression with Unobserved Individual-Specific Threshold Effects*

Ping Yu† — University of Hong Kong
Shengjie Hong‡ — Tsinghua University

Peter C. B. Phillips§
Yale University, University of Auckland
University of Southampton & Singapore Management University

First Version: February 2021
This Version: October 2022

Abstract

This paper studies the estimation and inferences in panel threshold regression with unobserved individual-specific threshold effects which is important from the practical perspective and is a distinguishing feature from traditional linear panel data models. It is shown that the within-regime differencing in the static model or the within-regime first-differencing in the dynamic model cannot generate consistent estimators of the threshold, so the correlated random effects models are suggested to handle the endogeneity in such general panel threshold models. We provide a unified estimation and inference framework that is valid for both the static and dynamic models and regardless of whether the unobserved individual-specific threshold effects exist or not. Especially, we propose alternative inference methods for the model parameters, which have better theoretical properties than the existing methods. Simulation studies and an empirical application illustrate the usefulness of our new estimation and inference methodology in practice.

Keywords: panel threshold regression, unobserved threshold effects, correlated random effects, adaptive Bonferroni inference
JEL-Classification: C23, C24

---

*We thank Bruce Hansen and seminar participants at ESAM 2022 for helpful comments. We also want to thank N.R. Ramírez-Rondán for providing his dataset to us. Phillips and Yu acknowledge support from the GRF of Hong Kong Government under Grant No. 17500719.

†HKU Business School, The University of Hong Kong, Pokfulam Road, Hong Kong; corresponding author email: pingyu@hku.hk.

‡SEM, Tsinghua University, Beijing, China 100084; email: hongshj@sem.tsinghua.edu.cn.

§Cowles Foundation for Research in Economics, Yale University, POBox 208281, New Haven, CT, USA; email: peter.phillips@yale.edu

1 Introduction

Panel data models provide an efficient way to eliminate endogeneity without the help of external instruments, and dynamic panel data models can further model the dynamic effects that the cross-sectional data models cannot handle. Panel threshold regression (PTR) models offer another dimension to panel data models as it can model the individual behaviour heterogeneity by introducing the threshold effect. The PTR model in the literature usually assumes

$$\begin{aligned} &y_{it} = \mathbf{x}_{it}'\beta_1 1\left(q_{it} \leq \gamma\right) + \mathbf{x}_{it}'\beta_2 1\left(q_{it} > \gamma\right) + \alpha_i + u_{it}, \\ &i = 1, \cdots, N, t = 1, \cdots, T, \end{aligned} \tag{1}$$

where the parameter of interest is $\theta = \left(\gamma, \beta'\right)'$ with $\beta = \left(\beta_1', \beta_2'\right)'$ or equivalently, $\theta = \left(\gamma, \beta_2', \delta_\beta'\right)'$ with $\delta_\beta = \beta_1 - \beta_2$ being the threshold effect in conditional mean of y_{it}, the observable time-variant covariates $\mathbf{x}_{it}$ can generally include a full set of time dummies, other aggregate time variables or lagged $\mathbf{x}_{it}$'s, α_i is the unobserved individual-specific, time-invariant effect (or unobserved heterogeneity), and u_{it} is the idiosyncratic time-varying shocks. This setup is similar to the traditional liner panel data model except that the regression coefficients depend on whether the threshold variable q_{it} crosses γ, where $\gamma \in \Gamma$ is the threshold point. In static PTR (SPTR), we usually assume $\mathbf{x}_{it}$ is strictly exogenous with respect to u_{it}, while in dynamic PTR (DPTR), $\mathbf{x}_{it}$ also contains the lagged y_{it}'s and the benchmark case is that only $y_{i,t-1}$ appears in $\mathbf{x}_{it}$. Often, q_{it} is also included in $\mathbf{x}_{it}$. So in SPTR, we write $\mathbf{x}_{it}' = (x_{it}', q_{it})$, and in DPTR, $\mathbf{x}_{it}' = (y_{i,t-1}, \underline{\mathbf{x}}_{it}')$, where $\underline{\mathbf{x}}_{it}' = (x_{it}', q_{it})$ if q_{it} is not $y_{i,t-1}$, and $\underline{\mathbf{x}}_{it} = x_{it}$ otherwise.

All the existing literature takes the setup (1), and we divide the literature into two strands. The first strand assumes α_i's are fixed effects and applies differencing to eliminate the endogeneity introduced by α_i. In SPTR, Hansen (1999) uses the usual fixed-effects transformation, i.e., the demeaning operation, to eliminate α_i, and then applies the least squares to estimate θ. In DPTR, both Seo and Shin (2016) and Ramírez-Rondán (2020) apply the first-difference transformation to eliminate α_i, but the former extends the FD-GMM approach of Arellano and Bond (1991) and the latter extends the maximum likelihood method of Hsiao et al. (2002) in linear dynamic panel models to the threshold case. Both Hansen (1999) and Ramírez-Rondán (2020) take the small-threshold-effect framework of Hansen (2000), so their convergence rate of $\widehat{\gamma}$ is $N^{1-2\kappa}$ and their asymptotic distributions involve a two-sided Brownian motion, where $\delta_\beta = O\left(N^{-\kappa}\right)$ with $0 < \kappa < 1/2$, and a hat over a parameter indicates its estimator. However, the convergence rate of $\widehat{\gamma}$ in Seo and Shin (2016) is $N^{1/2-\kappa}$, slower than $N^{1-2\kappa}$, and the estimation may suffer the identification failure when q_{it} is independent of the rest of the system as pointed out in Yu et al. (2018a). On the other hand, Seo and Shin (2016) allow the endogeneity of $\underline{\mathbf{x}}_{it}$ with respect to u_{it} and $\kappa = 0$ but the former two papers do not; [1] also, Seo and Shin (2016) use the unconditional moments to identify γ while the other two papers use the conditional moments.

The second strand of literature transforms (1) into nonparametric PTR and applies the integrated difference kernel estimator (IDKE) of Yu and Phillips (2018) to identify γ. This idea is proposed at the end of Yu and Phillips (2018) and can be applied to both SPTR and DPTR. Yu and Phillips (2018) use the fixed-threshold-effect framework of Chan (1993), and Yu et al. (2018a) extends to the small-threshold-effect framework. Later, Gøgens and Würtz (2019) combine the ideas of the two strands in DPTR; they first take differences to eliminate α_i and then apply the IDKE to estimate γ and GMM to estimate β. They compare the performance of $\widehat{\gamma}$ and $\widehat{\beta}$ with that of Seo and Shin (2016) by simulations and show that their approach is much more efficient in finite samples; this is because the convergence rate of the IDKE of γ is

[1] For other methods that do not allow the endogeneity of q_{it} and/or x_{it} in DPTR, see Shin and Kim (2011), Dang et al. (2012), Kremer et al. (2013) and Appendix A of Seo and Shin (2016).

much faster than that of the FD-GMM estimator of γ, which makes the β estimation in the second stage less contaminated. Anyway, the IDKE may suffer from the curse of dimensionality when the dimension of x_{it} is large so may not be practical. Finally, Wang and Lin (2010) extends the control function approach of Kourtellos et al. (2016) to SPTR, but as shown in Yu et al (2018b), the $\widehat{\gamma}$ in this approach is not generically consistent; also, this approach requires external instruments, while other methods mentioned above do not require such instruments for identification.

The setup (1) has two limitations. First, α_i is the same across the two regimes. This at least implies the intercepts in the two regimes must be the same, so if they are not, then the differencing method cannot generate a consistent estimator of γ.[2] Second, u_{it} does not experience a threshold effect in its distribution, e.g., in its variance. To fix these two limitations, we consider the following PTR model in this paper:

$$\begin{aligned} & y_{it} = \left(\mathbf{x}_{it}'\beta_1 + \alpha_{1i} + \sigma_1 u_{it}\right) 1\left(q_{it} \le \gamma\right) + \left(\mathbf{x}_{it}'\beta_2 + \alpha_{2i} + \sigma_2 u_{it}\right) 1\left(q_{it} > \gamma\right) \\ & i = 1, \cdots, N, t = 1, \cdots, T, \end{aligned} \tag{2}$$

where $E\left[u_{it}^2\right] = 1$. To the best of our knowledge, this is the first paper to consider such generalizations in the literature. Note that the generalizations of (2) relative to (1) are not trivial in both practice and estimation. In practice, if $\beta_1 \neq \beta_2$ it is very unlikely that $\alpha_{1i} = \alpha_{2i}$ or that the same individual characteristics persist in decisions (i.e., the determination of y_{it}) across regimes. Of course, α_{1i} and α_{2i} should be correlated because the same individual is making decision in the two regimes; for example, $\alpha_{1i} = \alpha_i + \eta_{1i}$ and $\alpha_{2i} = \alpha_i + \eta_{2i}$ with η_{1i} and η_{2i} being independent. Similarly, if there is a threshold effect in the conditional mean of y_{it}, it is natural to assume there is also a threshold effect in the conditional variance of y_{it}, i.e., $\sigma_1 \neq \sigma_2$. In estimation, the threshold effect in the variance of u_{it} is allowed by extending both strands of literature, but the threshold effect in α_i can be fixed only in the second strand. Especially, the differencing methods mentioned above in the setup (1) cannot generate a consistent estimator of θ since α_{1i} and α_{2i} cannot be eliminated; for example, even in the simple setup of α_{1i} and α_{2i} above, the usual (first-)differencing operation will not eliminate endogeneity because η_{1i} and η_{2i} are correlated with $\mathbf{x}_{it}$.

A natural idea to eliminate α_{1i} and α_{2i} is to take within-regime differencing in SPTR and within-regime first-differencing in DPTR. However, Section 2 shows that such differencings followed by the least squares cannot generate consistent estimation. To obtain consistent estimators of θ in both SPTR and DPTR, we suggest the correlated random effects (CRE) models in Section 3. Although the CRE models are more restrictive in the form of endogeneity compared with the fixed effects models, they provide a unified method that is valid for both the SPTR and DPTR and regardless of $\alpha_{1i} = \alpha_{2i}$ or not. In other words, this paper provides an interesting example where the fixed effects estimator is not consistent while the CRE estimator is, while it is commonly believed that the former is more robust to model specification than the latter. Because the setup (1) allows both fixed-effects models and CRE models, just like the traditional linear panel data models, while the setup (2) allows only CRE models, this paper points out a distinguishing feature of PTR from traditional linear panel data models which is not noticed in the existing literature. As Hansen (1999) and Ramírez-Rondán (2020), we take the small-threshold-effect framework of Hansen (2000) to develop our asymptotic results.

Section 4 considers the inference on γ and β. For γ, we invert the likelihood ratio (LR) statistic to construct the confidence interval (CI) as suggested in Hansen (1999, 2000). For β, although we can invert the t-statistic to construct the CI as in Hansen (1999), such a CI neglects the impact of γ estimation in finite samples (asymptotically, $\widehat{\beta}$ has the same distribution as in the case where γ_0 is known, where the subscript 0 of a parameter indicates its true value). Hansen (2000) suggests a Bonferroni-type CI where the

[2] Seo and Shin (2016) allow for the threshold effect in the intercept but can only identify the threshold effect of intercept rather than the levels of two intercepts.

coverage for γ is arbitrarily chosen and the coverage for β is fixed at the target level. We suggest a projection CI based on the LR statistic of $\left(\gamma, \beta'\right)'$, which can be interpreted as an adaptive Bonferroni CI with the coverages for γ and β adaptively chosen. We also suggest two alternative LR statistics to construct CIs for γ and β; these LR statistics exclude the indirect effects of the null hypotheses. It turns out that the new LR statistics for γ have the same asymptotic null distribution as the original one, while those for β do not, which shows the sharp difference in the nature of γ and β. Section 5 contains two auxiliary tests for our estimation procedure. The first one is to determine whether the threshold effect exists, and the second one is to test whether $\alpha_{1i} = \alpha_{2i}$ or whether the unobserved individual-specific threshold effect exists. Section 6 discusses two extensions of our basic CRE models: $\mathbf{x}_{it}$ contains some variables without threshold effects and there are multiple (instead of one) thresholds in (2). Section 7 applies the estimation and inference methodology developed in this paper to an empirical application to illustrate its usefulness in practice, and Section 8 concludes. Because the discussions on DPTR are very similar to those in SPTR, we collect them in Appendix A to avoid repetition in the main text. Proofs, calculation details, and simulation results are collected in the other six appendices.

We here collect further notations for future references. In SPTR, we observe $\left\{(y_{it}, \mathbf{x}'_{it})_{i=1}^{N}\right\}_{t=1}^{T}$ and $\{\underline{\mathbf{z}}_i\}_{i=1}^{N}$, and in DPTR, we observe $\left\{(y_{it}, \mathbf{x}'_{it})_{i=1}^{N}\right\}_{t=0}^{T}$ and $\{\underline{\mathbf{z}}_i\}_{i=1}^{N}$, i.e., the panel is balanced without missing data, where $\underline{\mathbf{z}}_i$ contains the time-invariant variables such as the constant 1. $n = NT$ is the number of observations used in estimation. Throughout the paper, we let N diverge to infinity and T fixed, i.e., the panel is short. In SPTR, $X_i := (\mathbf{x}'_{i1}, \cdots, \mathbf{x}'_{iT}, \underline{\mathbf{z}}'_i)'$, and in DPTR, $X_i := (\underline{\mathbf{x}}'_{i0}, \underline{\mathbf{x}}'_{i1}, \cdots, \underline{\mathbf{x}}'_{iT}, \underline{\mathbf{z}}'_i)'$ and $X_i^t := (X'_i, y_{i,t-1}, y_{i,t-2}, \cdots, y_{i0})'$, $t = 1, \cdots, T$. For any vector $\mathbf{z}$, $d_{\mathbf{z}}$ is the dimension of $\mathbf{z}$. For any matrix or vector A, $\|A\|$ denotes its Euclidean norm. For two real numbers a and b, $a \vee b := \max(a, b)$ and $a \wedge b := \min(a, b)$. The symbol ℓ is used to indicate the two regimes in (2) and, to simplify notation in what follows, the explicit values "$\ell = 1, 2$" are often omitted.

2 Difficulties in Applying Differencing in Fixed Effects Models

In this section, we show that the estimator based on the differencing method in SPTR is not consistent in general. Implicitly, we take the fixed-effect framework in this section.

We eliminate $\alpha_{\ell i}$ in each regime by a within-regime differencing. Specifically, we use

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

to remove $\alpha_{\ell i}$, where $\widetilde{y}_{it}^{-}(\gamma)$ is the residual of y_{it} regressed on $1(q_{it} \le \gamma)$ among the observations with $q_{it} \le \gamma$, and $\widetilde{y}_{it}^{+}(\gamma)$ is similarly understood. Only when $\gamma = \gamma_0$,

$$\mathbb{E}\left[\widetilde{y}_{it}(\gamma) | X_i\right] = \widetilde{\mathbf{x}}_{it}^{-}(\gamma)' \beta_1 1(q_{it} \le \gamma) + \widetilde{\mathbf{x}}_{it}^{+}(\gamma)' \beta_2 1(q_{it} > \gamma)$$

with

$$\begin{aligned}\widetilde{y}_{it}(\gamma_0) - \mathbb{E}\left[\widetilde{y}_{it}(\gamma_0) | X_i\right] \quad &= \quad \sigma_1 \widetilde{u}_{it}^{-}(\gamma_0) 1(q_{it} \le \gamma_0) + \sigma_2 \widetilde{u}_{it}^{+}(\gamma_0) 1(q_{it} > \gamma_0) \\ &= \quad : e_{it}^{-}(\gamma_0) 1(q_{it} \le \gamma_0) + e_{it}^{+}(\gamma_0) 1(q_{it} > \gamma_0) =: e_{it},\end{aligned}$$

where $\mathbf{x}_{it} = (x'_{it}, q_{it})'$, $X_i = (\mathbf{x}'_{i1}, \cdots, \mathbf{x}'_{iT})'$, and $\widetilde{\mathbf{x}}_{it}^{\pm}(\gamma)$ and $\widetilde{u}_{it}^{\pm}(\gamma)$ are similarly defined as $\widetilde{y}_{it}^{\pm}(\gamma)$, so we expect

$$\left(\widehat{\gamma}, \widehat{\beta}\right) = \arg\min_{\gamma,\beta} S_n(\gamma, \beta)$$

with

$$S_n(\gamma, \beta) = \sum_{i=1}^{N} \sum_{t=1}^{T} \left\{ \left[\widetilde{y}_{it}^{-}(\gamma) - \widetilde{\mathbf{x}}_{it}^{-}(\gamma)' \beta_1\right]^2 1(q_{it} \le \gamma) + \left[\widetilde{y}_{it}^{+}(\gamma) - \widetilde{\mathbf{x}}_{it}^{+}(\gamma)' \beta_2\right]^2 1(q_{it} > \gamma) \right\} \tag{3}$$

would be a consistent estimator of (γ_0, β_0). Interestingly, the response variable $\widetilde{y}_{it}(\gamma)$ involves an unknown parameter γ which is similar to estimation involving a Box-Cox transformation of the original response variable. The new aspect here is that γ appears also in the regression function $\mathbb{E}[\widetilde{y}_{it}(\gamma)|X_i]$ and appears in discontinuous forms in both $\widetilde{y}_{it}(\gamma)$ and $\mathbb{E}[\widetilde{y}_{it}(\gamma)|X_i]$. It turns out that these new characteristics imply an inconsistent estimator of γ and β.

For illustration, consider the following simple example,

$$\begin{aligned} y_{it} &= (\alpha_{1i} + \sigma_1 u_{it}) 1(q_{it} \le \gamma) + (\alpha_{2i} + \sigma_2 u_{it}) 1(q_{it} > \gamma) \\ &=: \alpha_{2i} + \sigma_2 u_{it} + (\delta_{\alpha i} + \delta_\sigma u_{it}) 1(q_{it} \le \gamma), \\ i &= 1, \cdots, N, t = 1, \cdots, T, \end{aligned} \tag{4}$$

where $u_{i1}, \cdots, u_{iT}, q_{i1}, \cdots, q_{iT}$ and $\{\alpha_{\ell i}\}_{\ell=1}^{2}$ are independent of each other. To be more specific, assume all u_{it} follows $N(0,1)$, all q_{it} follows $U[0,1]$, α_{1i} follows $N(0,1)$, $\alpha_{2i} = \delta_\alpha + \alpha_{1i}$, $\gamma_0 = 0.3$, $\sigma_{10} = 1$ and $T = 5$. We vary σ_{20} and δ_α to check the variation of the probability limits of the objective function. In this simple model, the only unknown parameter is γ.

When $\gamma < \gamma_0$,

$$\widetilde{y}_{it}^{-}(\gamma) = \sigma_{10} \widetilde{u}_{it}^{-}(\gamma),$$

and by some tedious calculation detailed in Appendix D,

$$\widetilde{y}_{it}^{+}(\gamma) = \sigma_{20} \widetilde{u}_{it}^{+}(\gamma) + \left[\delta_{\alpha i} \widetilde{1}_{it}(\gamma, \gamma_0) + \delta_\sigma \widetilde{u}_{it}(\gamma, \gamma_0)\right],$$

where $\widetilde{1}_{it}(\gamma, \gamma_0) = 1(\gamma < q_{it} \le \gamma_0) - \frac{\sum_{\tau=1}^{T} 1(\gamma < q_{i\tau} \le \gamma_0)}{\sum_{\tau=1}^{T} 1(q_{i\tau} > \gamma)}$, and $\widetilde{u}_{it}(\gamma, \gamma_0) = u_{it} 1(\gamma < q_{it} \le \gamma_0) - \frac{\sum_{\tau=1}^{T} u_{i\tau} 1(\gamma < q_{i\tau} \le \gamma_0)}{\sum_{\tau=1}^{T} 1(q_{i\tau} > \gamma)}$. As a result, when $\gamma < \gamma_0$,

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

and the probability limit of $S_n(\gamma)/n$ is

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

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

where $\overline{A}(\gamma,\gamma_0), A(\gamma,\gamma_0)$ and $B(\gamma,\gamma_0)$ are defined in Appendix D, $p^-(\gamma)=1-(1-\gamma)^T$ and $p^+(\gamma)=1-\gamma^T$. Similarly, when $\gamma>\gamma_0$, the probability limit of $S_n(\gamma)/n$ is

$$
\begin{aligned}
S(\gamma)\ =\ &\frac{1}{T}\left[\left(\frac{T\gamma}{1-(1-\gamma)^T}-1\right)p^-(\gamma)+\sigma_{20}^2\left(\frac{T(1-\gamma)}{1-\gamma^T}-1\right)p^+(\gamma)\right]\\
&+\frac{\delta_\alpha^2}{T}\left[\overline{A}(\gamma_0,\gamma)-B(\gamma_0,\gamma)\right]p^-(\gamma)+\frac{\sigma_{20}^2-1}{T}\left[\overline{A}(\gamma_0,\gamma)-A(\gamma_0,\gamma)\right]p^-(\gamma)\\
=\ &:T_1(\gamma)+T_2(\gamma)+T_3(\gamma),
\end{aligned}
$$

where where $\overline{A}(\gamma_0,\gamma), A(\gamma_0,\gamma)$ and $B(\gamma_0,\gamma)$ are defined in Appendix D. In the typical case where the dependent variable does not depend on γ, only $T_2(\gamma)$ would appear to indicate the threshold effect in $\alpha_{\ell i}$, and the minimizer is indeed γ_0.[3] Because $\widetilde{y}_{it}^{\pm}(\gamma)$ depends on γ, $\widetilde{u}_{it}^{\pm}(\gamma)$ and $\widetilde{u}_{it}(\gamma,\gamma_0)$ depend on γ, and T_1 and T_3 would not disappear, where T_3 is due to the threshold effect in error variance, while T_1 is attributed to $\widetilde{u}_{it}^{\pm}(\gamma)$ which comes completely from the construction of $\widetilde{y}_{it}^{\pm}(\gamma)$.

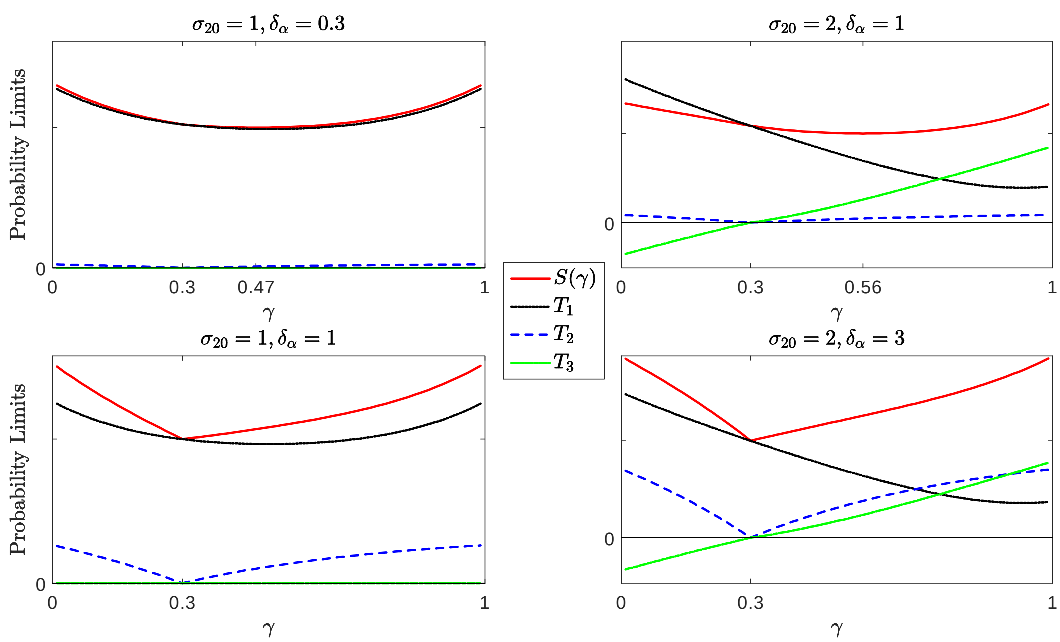

Figure 1: $S(\gamma)$ and Its Three Components for Various σ_{20} and δ_α Values: $\sigma_{10}=1, \gamma_0=0.3$, $T=5$

[3] When $\mathbf{x}_{it}$ is present, T_2 would also include a term indicating the threshold effect in β_ℓ, but the discussion and conclusion below still apply.

Figure 1 shows $S(\gamma)$ and its three components for various σ_{20} and δ_α values. First, when $\sigma_{20} = 1 = \sigma_{10}$, $T_3 = 0$, but T_1 still appear. When $\delta_\alpha = 0.3$ (i.e., the threshold effect in $\alpha_{\ell i}$ is small), T_2 is dominated by T_1 and the minimizer of $S(\gamma)$ is not γ_0, i.e., $\widehat{\gamma}$ is inconsistent. Second, when $\sigma_{20} = 1$ and $\delta_\alpha = 1$, although T_1 has a much larger value than T_2, it is quite flat such that the minimizer of $S(\gamma)$ is indeed γ_0. Third, when $\sigma_{20} = 2 \neq 1 = \sigma_{10}$, T_3 also appears; $\delta_\alpha = 1$ is not large enough to make the minimizer of $S(\gamma)$ to be γ_0 while $\delta_\alpha = 3$ is large enough to make so. In summary, for a larger threshold effect in error variance, we need a larger threshold effect in $\alpha_{\ell i}$ to make $\widehat{\gamma}$ consistent; in other words, $\widehat{\gamma}$ is generally inconsistent. In Appendix D, we also study the cases with $T = 2$ or $\gamma_0 = 0.5$; the conclusions here still apply. The case with $T = 2$ is especially interesting because $\widetilde{y}_{it}^-(\gamma) = 0$ and $\widetilde{y}_{it}^+(\gamma) = 0$ if $D_i^-(\gamma) := \sum_{t=1}^T 1(q_{it} \leq \gamma) = 0$ or 1 and $D_i^+(\gamma) := \sum_{t=1}^T 1(q_{it} > \gamma) = 0$ or 1, respectively, i.e., either $\widetilde{y}_{it}^-(\gamma)$ or $\widetilde{y}_{it}^+(\gamma)$ or both in $S_n(\gamma)$ is zero and all observations for each individual must fall in only one regime. In other words, the identification information of γ does not originate from the contract between the two regimes within group but from that between group. In the case with $\gamma_0 = 0.5$, $\arg\min_\gamma S(\gamma) = \gamma_0$ if $\sigma_{20} = 1 = \sigma_{10}$ because $\arg\min_\gamma T_1(\gamma) = \gamma_0$ due to the symmetricity of $f(q)$.

3 Correlated Random Effects Models

Due to the difficulties in applying differencing in a fixed effects model, we take the correlated random effects (CRE) model and use Chamberlain–Mundlak CRE device to control the endogeneity in this section. This device is a control function (CF) approach; actually, Yu et al. (2018) use the CF approach to handle the endogeneity in cross-sectional threshold regression. In the linear panel data model, correlations between $\mathbf{x}_{it}$ and α_i that are present at time t are also present at other times and can therefore be fully revealed by a linear function of $\overline{\mathbf{x}}_i$. In effect, the $\mathbf{x}_{it}$'s in other periods can serve as control variables in period t. Wooldridge (2010) shows how the CRE approach applies to commonly used models, such as unobserved effects probit, fractioanl response, tobit, and count models; this section shows that this approach can also apply to PTR.

3.1 Model Setup and Asymptotics

In SPTR, assume as in Mundlak (1978) that

$$\alpha_{\ell i} = \mathbf{z}_i'\psi_\ell + a_{\ell i} \text{ with } E[a_{\ell i}|X_i] = 0, \text{ and } E[u_{it}|X_i] = 0,$$

then

$$\begin{aligned} E[y_{it}|X_i] &= (\mathbf{x}_{it}'\beta_1 + \mathbf{z}_i'\psi_1)1(q_{it} \leq \gamma_0) + (\mathbf{x}_{it}'\beta_2 + \mathbf{z}_i'\psi_2)1(q_{it} > \gamma_0), \\ &=: \check{\mathbf{x}}_{it}'\theta_1 1(q_{it} \leq \gamma_0) + \check{\mathbf{x}}_{it}'\theta_2 1(q_{it} > \gamma_0) \end{aligned} \tag{5}$$

and the error term

$$\begin{aligned} e_{it}^0 &= (a_{1i} + \sigma_1 u_{it})1(q_{it} \leq \gamma_0) + (a_{2i} + \sigma_2 u_{it})1(q_{it} > \gamma_0) \\ &=: e_{1it}1(q_{it} \leq \gamma_0) + e_{2it}1(q_{it} > \gamma_0), \end{aligned} \tag{6}$$

where $\mathbf{z}_i' = (\overline{\mathbf{x}}_i', \underline{\mathbf{z}}_i')$ with $\overline{\mathbf{x}}_i = \frac{1}{T}\sum_{t=1}^T \mathbf{x}_{it}$, $\check{\mathbf{x}}_{it} := (\mathbf{x}_{it}', \mathbf{z}_i')'$, and $\theta_\ell = (\beta_\ell', \psi_\ell')'$.[4] A more flexible specification as in Chamberlain (1982, 1984) is possible, e.g., $\alpha_{\ell i} = \psi_\ell' X_i + a_{\ell i}$, but we maintain Mundlak's specification in this paper for simplicity. Note that a_{1i} and a_{2i} can be correlated, but need not the same. In other words, the correlation between α_{1i} and α_{2i} is through either $\mathbf{z}_i$ or a_{1i} and a_{2i}. When $\psi_1 = \psi_2$ and $a_{1i} = a_{2i}$, $\alpha_{1i} = \alpha_{2i}$. Also, $a_{\ell i}$ can be correlated with u_{it}.

[4] If $\mathbf{x}_{it}$ does not contain q_{it} in (2), $\overline{\mathbf{x}}_i$ should include $\overline{q}_i$ because $E[a_{\ell i}|X_i] = 0$ and X_i contains $\{q_{it}\}_{t=1}^T$. This comment applies also to DPTR where $q_{it} \neq y_{i,t-1}$ and $q_{it} \notin \underline{\mathbf{x}}_{it}$.

Now, the objective function is

$$S_n(\theta) = \sum_{i=1}^{N}\sum_{t=1}^{T}\left[y_{it} - \breve{\mathbf{x}}_{it}'\theta_1 1(q_{it}\le\gamma) - \breve{\mathbf{x}}_{it}'\theta_2 1(q_{it}>\gamma)\right]^2, \tag{7}$$

where $\theta = \left(\gamma, \beta', \psi'\right)'$ with $\beta' = \left(\beta_1', \beta_2'\right)$ and $\psi' = \left(\psi_1', \psi_2'\right)$, or $\theta = \left(\gamma, \underline{\theta}'\right)' := \left(\gamma, \theta_1', \theta_2'\right)'$. Denote the resulting estimator as $\widehat{\theta} = \left(\widehat{\gamma}, \widehat{\beta}', \widehat{\psi}'\right)'$ or $\left(\widehat{\gamma}, \widehat{\underline{\theta}}'\right)'$ and the residuals as

$$\widehat{e}_{it} = y_{it} - \breve{\mathbf{x}}_{it}'\widehat{\theta}_1 1(q_{it}\le\widehat{\gamma}) - \breve{\mathbf{x}}_{it}'\widehat{\theta}_2 1(q_{it}>\widehat{\gamma}) = \widehat{e}_{1it}1(q_{it}\le\widehat{\gamma}) + \widehat{e}_{2it}1(q_{it}>\widehat{\gamma}).$$

Usually, a two-step procedure is used to estimate γ. Specifically, for each $\gamma \in \Gamma$, run least squares of y_{it} on $\breve{\mathbf{x}}_{it}$ for (i,t)'s such that $q_{it} \le \gamma$ and $q_{it} > \gamma$ separately to obtain $\widehat{\theta}_1(\gamma)$ and $\widehat{\theta}_2(\gamma)$. The concentrated objective function

$$S_n(\gamma) = S_n\left(\gamma, \widehat{\theta}_1(\gamma), \widehat{\theta}_2(\gamma)\right) =: S_n\left(\gamma, \widehat{\underline{\theta}}(\gamma)\right);$$

then

$$\widehat{\gamma} = \arg\min_{\gamma\in\Gamma} S_n(\gamma)$$

and $\widehat{\theta}_\ell = \widehat{\theta}_\ell(\widehat{\gamma})$. As suggested in Section 3.2 of Hansen (1999), we need to search over $O(n)$ distinct q_{it} values (or just less quantiles of $\{q_{it}\}$) to estimate γ, where the smallest and largest ϵ% of q_{it}'s for some $\epsilon > 0$ (typically, 10 or 15) are excluded to guarantee that at least ϵ% observations lie in each regime. Usually, there is an interval to minimize $S_n(\gamma)$. Although in the small-threshold-effect framework assumed in this paper, any point on the minimizing interval has the same asymptotic distribution, we follow Yu (2012, 2015) to pick the middle-point as our estimator to improve the finite-sample performance of $\widehat{\gamma}$.

To study the asymptotic properties of $\widehat{\gamma}$ and $\widehat{\beta}$, we need to specify some assumptions. To ease the exposition, we define some further notations . $f_t(\gamma)$ is the density of q_{it} at γ. $f_{\tau|t}(\gamma_1|\gamma_2)$ is the conditional density of $q_{i\tau}$ given q_{it}.

$$\begin{aligned}
D(\gamma) &= \sum_{t=1}^{T} E\left[\breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{it}'|q_{it}=\gamma\right] f_t(\gamma) \text{ with } D = D(\gamma_0),\\
V_\ell(\gamma) &= \sum_{t=1}^{T} E\left[\breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{it}'e_{\ell it}^2|q_{it}=\gamma\right] f_t(\gamma) \text{ with } V_\ell = V_\ell(\gamma_0),\\
M &= \sum_{t=1}^{T} E\left[\breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{it}'\right], M(\gamma) := \sum_{t=1}^{T} E\left[\breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{it}'1(q_{it}\le\gamma)\right],\\
\Omega_1 &= \sum_{t=1}^{T}\sum_{\tau=1}^{T} E\left[\breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{i\tau}'e_{1it}e_{1i\tau}1(q_{it}\le\gamma_0)1(q_{i\tau}\le\gamma_0)\right],\\
\Omega_2 &= \sum_{t=1}^{T}\sum_{\tau=1}^{T} E\left[\breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{i\tau}'e_{2it}e_{2i\tau}1(q_{it}>\gamma_0)1(q_{i\tau}>\gamma_0)\right],\\
\Omega_{12} &= \sum_{t=1}^{T}\sum_{\tau\ne t} E\left[\breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{i\tau}'e_{1it}e_{2i\tau}1(q_{it}\le\gamma_0)1(q_{i\tau}>\gamma_0)\right];
\end{aligned}$$

compared with D and V_ℓ, Ω_1, Ω_2 and Ω_{12} contain cross terms (i.e., terms with $\tau \ne t$). Different from the D in Hansen (1999), the conditional means in our D involve information from other periods, i.e., $\breve{\mathbf{x}}_{it}$ contains information of $\mathbf{x}_{i\tau}$ with $\tau \ne t$.

Assumption SP:

(i) $\{\mathbf{x}_{it}, \underline{\mathbf{z}}_i, y_{it}\}_{t=1}^T$ are i.i.d. across i; T is fixed and $N \to \infty$.

(ii) For each i, $E[a_{\ell i}|X_i] = 0$ and $E[u_{it}|X_i] = 0$.

(iii) For each $j = 1, \cdots, d_{\mathbf{x}}$, $P\left(x_{i1}^j = \cdots = x_{iT}^j\right) < 1$, where x_{it}^j is the jth element of $\mathbf{x}_{it}$.

(iv) For $t = 1, \cdots, T$, $E\left[\|\check{\mathbf{x}}_{it}\|^4\right] < \infty$, $E\left[|e_{\ell it}|^4\right] < \infty$, $\sup_{\gamma \in \mathcal{N}} \mathbb{E}\left[(\|\check{\mathbf{x}}_{it}\| |e_{\ell it}|)^{2+\epsilon} |q_{it} = \gamma\right] < \infty$ for some $\epsilon > 0$ and some neighborhood $\mathcal{N}$ of γ_0; .

(v) For some fixed $c = \left(c_\beta', c_\psi'\right)'$, $\delta_N := \theta_1 - \theta_2 = cN^{-\kappa}$, where $0 < \kappa < 1/2$.

(vi) $D(\gamma)$ and $V_\ell(\gamma)$ are continuous at γ_0.

(vii) $c'Dc > 0$ and $c'V_\ell c > 0$.

(viii) For all $\gamma \in \Gamma$ and $t = 1, \cdots, T$, $f_t(\gamma) \leq \overline{f} < \infty$; for $\tau > t$, $f_{\tau|t}(\gamma_0|\gamma_0) < \infty$.

(ix) $\Omega_\ell > 0$.

(x) $M > M(\gamma) > 0$ for all $\gamma \in \Gamma$.

Condition (i) is standard in panel data models where the asymptotics are taken in the N dimension rather than the T dimension. In Condition (ii), we do not require $a_{1i}, a_{2i}, u_{i1}, \cdots, u_{iT}$ and X_i are independent of each other but only a conditional mean independence assumption, which implies $\Omega_{12} \neq \mathbf{0}$ due to $E[e_{1it}e_{2i\tau}|X_i] \neq 0$ as a result of the correlation between a_{1i} and a_{2i} and among $u_{i1}, \cdots, u_{iT}$.[5] Condition (iii) requires x_{it} to vary over t which is exactly how we partition $\mathbf{x}_{it}$ and $\underline{\mathbf{z}}_i$. This assumption avoids the multicollinear problem between $\mathbf{x}_{it}$ and $\overline{\mathbf{x}}_i$; a similar assumption is imposed in the differencing method of Hansen (1999) to avoid zero regressors. Condition (iv) includes some standard moment conditions, which implies $\Omega_\ell < \infty$, $M < \infty$, and (combining with the first part of Condition (viii)) $D(\gamma) < \infty$ and $V_\ell(\gamma) < \infty$ for any $\gamma \in \Gamma$. We can express this assumption in terms of $\mathbf{x}_{it}, u_{it}, \underline{\mathbf{z}}_i$ and $a_{\ell i}$, but the current formulation can simplify the statement. Note also that the Liapounov kind of condition $\sup_{\gamma \in \mathcal{N}} \mathbb{E}\left[(\|\check{\mathbf{x}}_{it}\| |e_{\ell it}|)^{2+\epsilon} |q_{it} = \gamma\right] < \infty$ is not implied by $E\left[\|\check{\mathbf{x}}_{it}\|^4\right] < \infty$ and $E\left[|e_{\ell it}|^4\right] < \infty$. Condition (v) indicates that we take the small-threshold-effect framework of Hansen (1999). The continuity of $D(\gamma)$ at γ_0 in Condition (vi) excludes the possibility that $\check{\mathbf{x}}_{it}$ has a discontinuous conditional distribution at $q_{it} = \gamma_0$ (if we assume $f_t(\gamma)$ is continuous at γ_0), but we can definitely relax this requirement.[6] Different from Hansen (1999), $V_1(\gamma)$ and $V_2(\gamma)$ are generally different due to $a_{1i} \neq a_{2i}$ and $\sigma_1 \neq \sigma_2$. Note that although $\psi_1 - \psi_2 \to 0$ as $N \to \infty$, $\alpha_{1i} - \alpha_{2i}$ need not converge to zero; in other words, our framework will not degenerate to Hansen (1999)'s in general. Condition (vii) excludes continuous threshold regression (see Chan and Tsay (1998) and Hansen (2017)) and guarantees that $f_t(\gamma_0) > 0$ for at least one t. The second part of Condition (viii) excludes the possibility that q_{it} lingers at γ_0 with a positive probability over time; it eliminates the cross terms in D and V_ℓ.[7] Condition (ix) is standard in stating the asymptotic distribution of $\widehat{\beta}$, and Condition (x) restricts Γ to a proper subset of the joint support of $\{q_{it}\}_{t=1}^T$.

The following theorem states the asymptotic distributions of $\widehat{\gamma}$ and $\widehat{\beta}$.

Theorem 1 *Under Assumption SP,*

$$N^{1-2\kappa}(\widehat{\gamma} - \gamma_0) \xrightarrow{d} \omega \cdot \zeta(\phi)$$

and

$$N^{1/2}\left(\widehat{\theta}_\ell - \theta_\ell\right) \xrightarrow{d} N(0, \Sigma_\ell),$$

[5] Of course, also because q_{it} is not fixed over $t = 1, \cdots, T$ so that $1(q_{it} \leq \gamma_0) 1(q_{i\tau} > \gamma_0) \neq 0$.

[6] Continuity of the conditional distribution of $\mathbf{x}_{it}$ at $q_{it} = \gamma_0$ is also an indicator for why all $\mathbf{x}_{it}$'s are used as control variables for $\alpha_{\ell i}$, not only those $\mathbf{x}_{it}$'s in each regime.

[7] The deeper reason that D and V_ℓ do not contain cross terms while Ω_ℓ and Ω_{12} do is because the γ estimation explores the local information around γ_0 while the $\underline{\theta}$ estimation explores global information as detailed in Yu (2012, 2015).

where

$$\omega = \frac{c'V_1c}{(c'Dc)^2} \text{ and } \zeta(\phi) = \arg\max_r \begin{cases} -\frac{|r|}{2} + B_1(-r), & \text{if } r \leq 0, \\ -\frac{r}{2} + \sqrt{\phi}B_2(r), & \text{if } r > 0, \end{cases}$$

with $\phi = c'V_2c/c'V_1c$ *and* $B_\ell(r)$, $\ell = 1, 2$, *being two independent standard Wiener processes on* $[0, \infty)$, *and*

$$\Sigma_\ell = M_\ell^{-1}\Omega_\ell M_\ell^{-1}$$

with $M_1 = M(\gamma_0)$ *and* $M_2 = M - M_1$. $\widehat{\gamma}$ *and* $\widehat{\underline{\theta}}$ *are asymptotically independent, while the asymptotic covariance of* $\widehat{\theta}_1$ *and* $\widehat{\theta}_2$ *is* $\Sigma_{12} := M_1^{-1}\Omega_{12}M_2^{-1}$.

Note that different from the cross-sectional case, $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are not asymptotically independent now. In practice, researchers often assume the error components structure on $e_{\ell it}$, e.g., assume $\left(a_{1i}, a_{2i}, \{u_{it}\}_{t=1}^T\right)$ and X_i are independent and u_{it}, $t = 1, \cdots, T$, are i.i.d., and then V_ℓ, Ω_ℓ and Ω_{12} can be simplified. In such a case, $E\left[e_{\ell it}^2|X_i\right] = E\left[e_{\ell it}^2\right] = E\left[a_{\ell i}^2\right] + \sigma_\ell^2 := \varsigma_\ell^2$,[8] $E\left[e_{\ell it}e_{\ell i\tau}|X_i\right] = E\left[a_{\ell i}^2\right] =: c_\ell$, and $E\left[e_{1it}e_{2i\tau}|X_i\right] = E\left[a_{1i}a_{2i}\right] := c_{12}$ for $t, \tau = 1, \cdots, T$ and $\tau \neq t$ such that

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

which implies

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

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

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

Note that $\Sigma_\ell \neq \varsigma_\ell^2 M_\ell^{-1}$ as in cross sections.[9] When $\alpha_{1i} = \alpha_{2i}$, we need only set $c_\psi = \mathbf{0}$ in ω and ϕ and set $a_{1i} = a_{2i}$ in V_ℓ, Ω_ℓ and Ω_{12}.

3.2 Estimation of the Asymptotic Variance Matrix of $\widehat{\underline{\theta}}$

For future references, we collect the asymptotic variance matrix estimators of $\widehat{\underline{\theta}}$ in this subsection. In the general model,

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

[9] 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).

$$\widehat{\Sigma}_\ell = \widehat{M}_\ell^{-1}\widehat{\Omega}_\ell\widehat{M}_\ell^{-1} \text{ and } \widehat{\Sigma}_{12} = \widehat{M}_1^{-1}\widehat{\Omega}_{12}\widehat{M}_2^{-1}$$

where

$$\begin{aligned}
\widehat{M}_1 &= \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\check{\mathbf{x}}_{it}\check{\mathbf{x}}_{it}'1\left(q_{it}\leq\widehat{\gamma}\right), \\
\widehat{\Omega}_1 &= \frac{1}{N}\sum_{i=1}^{N}\left(\sum_{\tau=1}^{T}\check{\mathbf{x}}_{it}\widehat{e}_{1it}1\left(q_{it}\leq\widehat{\gamma}\right)\right)\left(\sum_{\tau=1}^{T}\check{\mathbf{x}}_{it}\widehat{e}_{1it}1\left(q_{it}\leq\widehat{\gamma}\right)\right)', \\
\widehat{\Omega}_{12} &= \frac{1}{N}\sum_{i=1}^{N}\left(\sum_{\tau=1}^{T}\check{\mathbf{x}}_{it}\widehat{e}_{1it}1\left(q_{it}\leq\widehat{\gamma}\right)\right)\left(\sum_{\tau=1}^{T}\check{\mathbf{x}}_{it}\widehat{e}_{2it}1\left(q_{it}>\widehat{\gamma}\right)\right)',
\end{aligned}$$

$\widehat{M}_2$ is the same as $\widehat{M}_1$ but replaces $q_{it}\leq\widehat{\gamma}$ by $q_{it}>\widehat{\gamma}$, and $\widehat{\Omega}_2$ is the same as $\widehat{\Omega}_1$ but replaces $\widehat{e}_{1it}1\left(q_{it}\leq\widehat{\gamma}\right)$ by $\widehat{e}_{2it}1\left(q_{it}>\widehat{\gamma}\right)$.

In the error components model, the formulae of $\widehat{\Omega}_\ell$ and $\widehat{\Omega}_{12}$ above 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 $\widehat{\Omega}_1$,

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

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

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

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

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

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

4 Inferences on γ and β

In this section, we propose some inference methods for γ and β. Following Hansen (1999, 2000), we invert the LR statistic for hypotheses concerning γ such as $H_0 : \gamma = \gamma_0$ to construct a CI for γ. As for β, Hansen (2000) suggests to use a Bonferroni-type CI to incorporate the randomness of $\widehat{\gamma}$ in finite samples (although $\widehat{\gamma}$ will not affect the distribution of $\widehat{\beta}$ asymptotically); the coverage for γ in such a CI is arbitrarily set at 80% and for each γ in its CI the coverage for β is fixed at $1-\alpha$. We propose a projection CI based on the LR statistic for hypotheses concerning both γ and β, which can be interpreted as an adaptive Bonferroni CI with the coverage for γ and the coverage for β at each γ adaptively chosen. We also propose two alternative forms of the two LR statistics. Comparing with the asymptotic null distributions of the original two LR statistics, we can easily see the difference in the nature of γ and β.

4.1 LR Inference on γ

As emphasized in Hansen (2000), the following LR-like statistic has better finite-sample performance than a typical t-like statistic when the threshold effect is small:

$$LR_n(\gamma) = \frac{S_n(\gamma) - S_n(\widehat{\gamma})}{\widehat{\eta}^2},$$

where $\widehat{\eta}^2$ is a consistent estimator of $\eta^2 = c'V_1c/c'Dc$, and $S_n(\gamma)$ is the concentrated objective function. This test statistic is a by-product of the two-step estimation procedure for γ.

Theorem 2 *Under Assumption SP,*

$$LR_n(\gamma_0) \xrightarrow{d} \xi(\phi),$$

where $\xi(\phi) = \sup_r \begin{cases} -|r| + 2B_1(-r), & \text{if } r \le 0, \\ -r + 2\sqrt{\phi}B_2(r), & \text{if } r > 0, \end{cases}$ *has the distribution* $P(\xi(\phi) \le x) = (1 - e^{-\frac{x}{2}})(1 - e^{-\frac{x}{2\phi}})$.

In the error components model, $\eta^2 = \varsigma_1^2$ and $\phi = \varsigma_2^2/\varsigma_1^2$ can be estimated from Section 3.2. In the general case, $\eta^2 = \frac{N^{-2\kappa}c'V_1c}{N^{-2\kappa}c'Dc}$ and $\phi = \frac{N^{-2\kappa}c'V_2c}{N^{-2\kappa}c'V_1c}$; we can estimate $N^{-2\kappa}c'Dc$ and $N^{-2\kappa}c'V_\ell c$ by

$$N^{-2\kappa}c'\widehat{D}c = \sum_{t=1}^T N^{-2\kappa}c'\widehat{D}_tc \text{ and } N^{-2\kappa}c'\widehat{V}_\ell c = \sum_{t=1}^T N^{-2\kappa}c'\widehat{V}_{\ell t}c,$$

where $N^{-2\kappa}c'\widehat{D}_tc$ and $N^{-2\kappa}c'\widehat{V}_{\ell t}c$ are estimators of $E\left[\left(\delta_N'\check{\mathbf{x}}_{it}\right)^2 | q_{it} = \gamma_0\right] f_t(\gamma_0)$ and $E\left[\left(\delta_N'\check{\mathbf{x}}_{it}\right)^2 e_{\ell it}^2 | q_{it} = \gamma_0\right] f_t(\gamma_0)$ using the data of all individuals in period t and substituting δ_N, γ_0 and $e_{\ell it}$ by $\widehat{\delta} := \widehat{\theta}_1 - \widehat{\theta}_2$, $\widehat{\gamma}$ and $\widehat{e}_{\ell it}$, respectively. Since such estimators are standard in the literature so will not be repeated here; see, e.g., Section 3.4 of Hansen (2000) or Section 3.4 of Yu et al. (2018b) for the details. Given the estimates $\widehat{\eta}^2$ and $\widehat{\phi}$, the $(1-\alpha)$ LR-CI for γ follows by inversion of the statistic from

$$\left\{\gamma | LR_n(\gamma) \le \widehat{c}_\alpha^0\right\},$$

where $\widehat{c}_\alpha^0$ is the $(1-\alpha)$ quantile of ξ with ϕ being replaced by $\widehat{\phi}$.

4.2 Adaptive Bonferroni Inference on β

Without loss of generality, suppose we are interested in β_{11}, the first element of β_1. Correspondingly, decompose β_1 as $\left(\beta_{11}, \beta_{12}'\right)'$, θ_1 as $\left(\beta_{11}, \theta_{12}'\right)'$, $\mathbf{x}_{it}$ as $\left(x_{it}^1, \mathbf{x}_{it}^{-1\prime}\right)'$, $\breve{\mathbf{x}}_{it}$ as $\left(x_{it}^1, \breve{\mathbf{x}}_{it}^{-1\prime}\right)'$, and $M_1 = \begin{pmatrix} S_{\beta_{11}\beta_{11}} & S_{\beta_{11}\theta_{12}} \\ S_{\theta_{12}\beta_{11}} & S_{\theta_{12}\theta_{12}} \end{pmatrix}$. The null hypothesis of interest is $H_0 : \gamma = \gamma_0$ and $\beta_{11} = \beta_{11}^0$, and the LR statistic is

$$LR_n\left(\gamma, \beta_{11}\right) = \frac{S_n\left(\gamma, \beta_{11}\right) - S_n\left(\widehat{\gamma}, \widehat{\beta}_{11}\right)}{\widehat{\eta}^2},$$

where the nuisance parameter η^2 can be estimated with either the null imposed or not. $S_n\left(\gamma, \beta_{11}\right)$ is the concentrated objective function on $\left(\gamma, \beta_{11}\right)$. Specifically, for each $\gamma \in \Gamma$, run least squares of $y_{it} - x_{it}^1\beta_{11}$ on $\left(\mathbf{x}_{it}^{-1\prime}, \mathbf{z}_i'\right)'$ for (i,t)'s such that $q_{it} \leq \gamma$ and y_{it} on $\left(\mathbf{x}_{it}', \mathbf{z}_i'\right)'$ for (i,t)'s such that $q_{it} > \gamma$ separately to obtain $\widehat{\theta}_{12}\left(\gamma, \beta_{11}\right)$ and $\widehat{\theta}_2\left(\gamma\right)$. The concentrated objective function

$$S_n\left(\gamma, \beta_{11}\right) = S_n\left(\gamma, \beta_{11}, \widehat{\theta}_{12}\left(\gamma, \beta_{11}\right), \widehat{\theta}_2\left(\gamma\right)\right).$$

Theorem 3 *Under Assumption SP,*

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

where

$$\varrho_{11} = \frac{\left(1, -S_{\beta_{11}\theta_{12}}S_{\theta_{12}\theta_{12}}^{-1}\right)\Omega_1\left(1, -S_{\beta_{11}\theta_{12}}S_{\theta_{12}\theta_{12}}^{-1}\right)'}{\widetilde{S}_{\beta_{11}\beta_{11}}\eta^2}$$

with $\widetilde{S}_{\beta_{11}\beta_{11}} = S_{\beta_{11}\beta_{11}} - S_{\beta_{11}\theta_{12}}S_{\theta_{12}\theta_{12}}^{-1}S_{\theta_{12}\beta_{11}}$, *and the* χ_1^2 *distribution and* $\xi(\phi)$ *are independent.*

In the error components model, we show in the proof of Theorem 3 that

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

and the estimator of ϱ_{11} can be constructed from Section 3.2, where note 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 critical value does not have an explicit form, but we can simulate independent χ_1^2 and $\xi(\phi)$ random numbers to obtain the critical value; especially, $\xi(\phi) = 2\max\left(\xi_1, \xi_2\left(\phi\right)\right)$, where ξ_1 follows the standard exponential distribution, $\xi_2\left(\phi\right)$ follows an exponential distribution with mean ϕ, and ξ_1 and $\xi_2\left(\phi\right)$ are independent. Suppose the level α critical value is $\widehat{c}_\alpha$, then the $(1-\alpha)$ CI of β_{11} is

$$\left\{\beta_{11} | LR_n\left(\gamma, \beta_{11}\right) \leq \widehat{c}_\alpha\right\}.$$

In practice, we can collect the intervals of β_{11} for each $q_{it} \in \Gamma$:

$$\bigcup\nolimits_{q_{it}\in\Gamma}\left\{\beta_{11} | LR_n\left(q_{it}, \beta_{11}\right) \leq \widehat{c}_\alpha\right\} =: \bigcup\nolimits_{q_{it}\in\Gamma}\mathrm{CI}\left(\beta_{11}|q_{it}\right), \quad (8)$$

where $\mathrm{CI}(\beta_{11}|q_{it})$ is either an interval or empty; see Appendix E for details on the construction of $\mathrm{CI}(\beta_{11}|q_{it})$. Note here that we do not preset the coverage for γ to construct the CI for β_{11}; rather, the actual coverage for γ depends on the dataset and is adaptively determined. Also, for each fixed γ in the joint confidence region, the coverage for β_{11} is not fixed at $1-\alpha$ as in Hansen (2000). Such a projection-based CI is typically conservative due to the correlation between the asymptotic distributions of the interested parameter and the nuisance parameter; however, because $\widehat{\beta}_{11}$ and $\widehat{\gamma}$ are asymptotically independent, the conservativeness here should not be severe. Our simulations in Appendix G confirm this intuition. Finally, for different elements of β, our test statistics are different, focusing information in the direction of the interested β element.

4.3 Alternative Inference Procedures

The two LR statistics, $LR_n(\gamma)$ and $LR_n(\gamma,\beta_{11})$, contain the indirect effects of fixing γ and (γ,β_{11}) at the hypothetical true values. Specifically,

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

where $\widehat{\underline{\theta}}(\widehat{\gamma})\neq\widehat{\underline{\theta}}(\gamma)$, and $\widehat{\underline{\theta}}_{-11}\left(\widehat{\gamma},\widehat{\beta}_{11}\right)\neq\widehat{\underline{\theta}}_{-11}(\gamma,\beta_{11})$ with $\widehat{\underline{\theta}}_{-11}(\gamma,\beta_{11})=\left(\widehat{\theta}_{12}(\gamma,\beta_{11})',\widehat{\theta}_2(\gamma)'\right)'$. We can exclude such indirect effects by defining the following alternative form of $LR_n(\gamma)$ and $LR_n(\gamma,\beta_{11})$:

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

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

$$
\widehat{\beta}_{11}\left(r,\widehat{\underline{\theta}}_{-11}(\gamma,\beta_{11})\right)=\frac{\sum_{i=1}^N\sum_{t=1}^T x_{it}^1\left(y_{it}-\breve{\mathbf{x}}_{it}^{-1\prime}\widehat{\theta}_{12}(\gamma,\beta_{11})\right)1(q_{it}\leq r)}{\sum_{i=1}^N\sum_{t=1}^T\left(x_{it}^1\right)^2 1(q_{it}\leq r)},
$$
[10]

and we then search r over Γ to minimize

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

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

Theorem 4 *Under Assumption SP,*

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

and

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

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

where ϱ_{11}, χ_1^2 and $\xi(\phi)$ have the same definitions as in Theorem 3.

The CIs for γ and β_{11} can be similarly constructed as in the last two subsections, but the calculations are more involved; see Appendix E for the details. 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 β_{11}'s (this is because $\{\beta_{11}|LR_{1n}(q_{it},\beta_{11})\leq\widehat{c}_{1\alpha}\}$ 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 $S_{\beta_{11}\beta_{11}}\geq\widetilde{S}_{\beta_{11}\beta_{11}}$, 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 3.

Comparing with Theorems 2 and 3, we can get 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 $\xi(\phi)$ 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 3 and 4, $\widehat{\beta}_{11}\left(\widehat{\underline{\theta}}_{-11}\left(\gamma_0,\beta_{11}^0\right)\right)$ in $LR_n\left(\gamma_0,\beta_{11}^0\right)$ is more efficient than $\widehat{\beta}_{11}$ in $LR_n(\gamma_0)$, 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 some contents. Third, note that $LR_{1n}(\gamma)\leq LR_n(\gamma)$ with $\min_\gamma LR_{1n}(\gamma)=\min_\gamma LR_n(\gamma)=0$ because $S_n\left(\widehat{\gamma}\left(\widehat{\underline{\theta}}(\gamma)\right),\widehat{\underline{\theta}}(\gamma)\right)\geq S_n\left(\widehat{\gamma},\widehat{\underline{\theta}}(\widehat{\gamma})\right)$ and $S_n\left(\gamma,\widehat{\underline{\theta}}(\gamma)\right)\geq S_n\left(\widehat{\gamma}\left(\widehat{\underline{\theta}}(\gamma)\right),\widehat{\underline{\theta}}(\gamma)\right)$ with both equalities hold when $\gamma=\widehat{\gamma}$, and similarly, $LR_{1n}(\gamma,\beta_{11})\leq LR_n(\gamma,\beta_{11})$ with $\min_{\gamma,\beta_{11}}LR_{1n}(\gamma,\beta_{11})=\min_{\gamma,\beta_{11}}LR_n(\gamma,\beta_{11})=0$. Since the critical values for $LR_{1n}(\gamma)$ and $LR_n(\gamma)$ are the same, $LR_{1n}(\gamma)$ would result in a wider CI for γ. On the other hand, the critical value for $LR_{1n}(\gamma,\beta_{11})$ is smaller than $LR_n(\gamma,\beta_{11})$, and thus the CI resulting from $LR_{1n}(\gamma,\beta_{11})$ need not be wider than that from $LR_n(\gamma,\beta_{11})$.

Due to the two drawbacks of CIs based on $LR_{1n}(\gamma)$ and $LR_{1n}(\gamma,\beta_{11})$, (i) time-consuming because of the computation of $\widehat{\gamma}\left(\widehat{\underline{\theta}}(\gamma)\right)$ and $\left(\widehat{\gamma}\left(\widehat{\underline{\theta}}_{-11}(\gamma,\beta_{11})\right),\widehat{\beta}_{11}\left(\widehat{\underline{\theta}}_{-11}(\gamma,\beta_{11})\right)\right)$, and (ii) less powerful because $LR_{1n}(\gamma)\leq LR_n(\gamma)$ and $LR_{1n}(\gamma,\beta_{11})\leq LR_n(\gamma,\beta_{11})$, we propose another alternative inference procedure. Specifically, define

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

This form of LR statistics have at least three advantages, (i) they shut down the indirect effects of the null; (ii) they need not calculate $\widehat{\gamma}\left(\widehat{\underline{\theta}}(\gamma)\right)$ and $\left(\widehat{\gamma}\left(\widehat{\underline{\theta}}_{-11}(\gamma,\beta_{11})\right),\widehat{\beta}_{11}\left(\widehat{\underline{\theta}}_{-11}(\gamma,\beta_{11})\right)\right)$; (iii) they are the most powerful among the three forms of LR statistics since $LR_{1n}(\gamma)\leq LR_n(\gamma)\leq LR_{2n}(\gamma)$ and $LR_{1n}(\gamma,\beta_{11})\leq LR_n(\gamma,\beta_{11})\leq LR_{2n}(\gamma,\beta_{11})$. The details of CI construction based on LR_{2n} are provided in Appendix E; it turns out that the complexity of CIs based on LR_{2n} is similar to that based on LR_n.

Theorem 5 *Under Assumption SP,*

$$LR_{2n}(\gamma_0)\xrightarrow{d}\xi(\phi),$$

and

$$LR_{2n}\left(\gamma_0,\beta_{11}^0\right)\xrightarrow{d}g(W_1)+\xi(\phi),$$

where $W_1 \sim N(\mathbf{0}, \Omega_1)$ is independent of $\xi(\phi)$ which is defined in Theorem 2, and the $g(\cdot)$ function is defined in the proof of the theorem.

Like LR_{1n}, the different formulation of LR_{2n} from LR_n does not affect the asymptotic component related to γ but only affect that related to β_{11}. Because the $g(\cdot)$ function does not take a quadratic form, $g(W_1)$ will not follow a scaled χ_1^2 distribution. Consequently, the critical value for $LR_{1n}\left(\gamma_0, \beta_{11}^0\right)$ depends on the simulation of $g(W_1)$, but this is not hard nowdays.

5 Two Hypothesis Tests

We develop two auxiliary tests in this section, which should be conducted before the estimation and inferences in the last two sections. The first test is to test whether there is threshold effect. Under the null, the model is linear, so this test is also termed as testing for linearity. The second test is to check whether there are unobserved individual-specific threshold effects, i.e., whether the new setup in this paper beyond those in the literature makes sense. These two tests can be conducted sequentially.

5.1 Testing for Linearity

The null hypothesis is $H_0: \theta_1 = \theta_2$ or $\delta_\theta = \mathbf{0}$ and the alternative is $H_1: \theta_1 \neq \theta_2$ or $\delta_\theta \neq 0$, where $\delta_\theta = \theta_1 - \theta_2$. Usually, the Wald-type or LR-type tests are suggested, but we will use the score test in Yu (2013) and Yu and Fan (2021) to test this hypothesis because it is much easier to implement.

Note that the objective function $S_n(\theta)$ can be written as

$$S_n(\theta) = \sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it} - \check{\mathbf{x}}_{it}'\theta_o - \check{\mathbf{x}}_{it}'\delta_\theta 1(q_{it} \leq \gamma)\right)^2.$$

whose score function with respect to δ_θ and evaluated at $\mathbf{0}$ is

$$s_n(\gamma) = \sum_{i=1}^{N}\sum_{t=1}^{T}\check{\mathbf{x}}_{it}e_{it}^{o}1(q_{it} \leq \gamma)$$

after discarding the constant terms, where $\theta_o = \theta_2$, $e_{it}^o = e_{it}^0$ under H_0 but will include a bias under H_1, and can be estimated by $\widehat{e}_{it}^o = y_{it} - \check{\mathbf{x}}_{it}'\widehat{\theta}_o$ with $\widehat{\theta}_o$ being the coefficients in the regression of y_{it} on $\check{\mathbf{x}}_{it}$. So our score test is essentially testing the following moment conditions

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

Our test statistics are based on

$$T_n(\gamma) = \widehat{H}_n(\gamma)^{-1/2}\widehat{m}_n(\gamma), \tag{9}$$

where

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

and

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

with

$$\widehat{M}(\gamma) = N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\check{\mathbf{x}}_{it}\check{\mathbf{x}}_{it}'1(q_{it}\le\gamma) \text{ and } \widehat{M} = N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\check{\mathbf{x}}_{it}\check{\mathbf{x}}_{it}'.$$

The extra term $\widehat{M}(\gamma)\widehat{M}^{-1}\check{\mathbf{x}}_{it}$ in $\widehat{H}_n(\gamma)$ is to offset the effect of $\widehat{\theta}_o$ in $\widehat{e}_{it}^{o}$. We can also recenter $\check{\mathbf{x}}_{it}1(q_{it}\le\gamma)$ by $\widehat{M}(\gamma)\widehat{M}^{-1}\check{\mathbf{x}}_{it}$ in $\widehat{m}_n(\gamma)$ but since $\sum_{i=1}^{N}\sum_{t=1}^{T}\check{\mathbf{x}}_{it}\widehat{e}_{it}^{o} = \mathbf{0}$ this is not necessary. Given $T_n(\gamma)$, we can construct the Kolmogorov-Smirnov sup-type statistic

$$KS = \sup_{\gamma\in\Gamma}\|T_n(\gamma)\|^2$$

or the Cramér-von Mises average-type statistic

$$C_vM = \int_{\Gamma}\|T_n(\gamma)\|^2 d\gamma.$$

We will use $g_n = g(T_n)$ to denote either of these two functionals of $T_n(\gamma)$.

We next derive the asymptotic distribution of g_n under the local alternative,

$$H_1^c : \delta_\theta = N^{-1/2}c,$$

where using the same notation c as in Section 3 should not introduce any confusion. Our asymptotic results imply the asymptotic null distribution and the consistency of our tests.

Theorem 6 *Under* H_1^c*,*

$$g_n \xrightarrow{d} g_c := g(T^c),$$

where

$$T^c(\gamma) = H(\gamma)^{-1/2}\left\{\Xi(\gamma) + \left[M(\gamma\wedge\gamma_0) - M(\gamma)M^{-1}M(\gamma_0)\right]c\right\},$$

$\Xi(\gamma)$ *is a mean zero Gaussian process with covariance kernel*

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

and $H(\gamma) = H(\gamma,\gamma)$.

Because the asymptotic null distribution is not pivotal, we use the simulation method of Hansen (1996) to obtain the critical values or p-values. Specifically, the following procedure is conducted:

1. Generate i.i.d. $N(0,1)$ random variables $\{\xi_i^*\}_{i=1}^N$.
2. Set $T_n^*(\gamma) = \widehat{H}_n(\gamma)^{-1/2}\widehat{m}_n^*(\gamma)$ and $g_n^* = g(T_n^*)$, where

$$\widehat{m}_n^*(\gamma) = N^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\check{\mathbf{x}}_{it}1(q_{it}\le\gamma) - \widehat{M}(\gamma)\widehat{M}^{-1}\check{\mathbf{x}}_{it}\right)\widehat{e}_{it}^{o}\xi_i^*.$$

3. Repeat the first two steps J times to generate $\left\{g_n^{j*}\right\}_{j=1}^{J}$.

4. If $p_n^{J*} = J^{-1}\sum_{j=1}^{J} 1(g_n^{j*} \geq g_n) \leq \alpha$, we reject H_0; otherwise, accept H_0.

Step 2 deserves further explanations. First, the extra randomness introduced by the simulation only appears in $\widehat{m}_n^*(\gamma)$ but not in $\widehat{H}_n(\gamma)$; essentially, this is a wild bootstrap procedure. Second, the same $\{\xi_i^*\}_{i=1}^N$ are used for all $\gamma \in \Gamma$. In practice, we can replace Γ by a discrete approximation, say $\Gamma_n := \{q_{it}|q_{it} \in \Gamma\}$, which becomes dense in Γ as $N \to \infty$. Third, the same ξ_i^* is associated with $\widehat{e}_{it}^o$ for all $t = 1, \cdots, T$ to maintain the correlation structure between $\{e_{it}^o\}_{i=1}^T$. Fourth, the extra term $\widehat{M}(\gamma)\widehat{M}^{-1}\breve{\mathbf{x}}_{it}$ in $\widehat{m}_n^*(\gamma)$ is critical and cannot be omitted although it is unnecessary in $\widehat{m}_n(\gamma)$.

5.2 Testing for Unobserved Individual-Specific Threshold Effects

The null is $\alpha_{1i} = \alpha_{2i}$. If we pass the first test, i.e., there is indeed a threshold effect, then the null reduces to

$$H_0 : \psi_1 = \psi_2.$$

Although this is only an implication of $\alpha_{1i} = \alpha_{2i}$, it is the only relevant one to our estimation in the CRE framework. We will carry out the usual Wald test for this hypothesis; the LR test based on the S_n function in (7) is not easy to implement since its asymptotic null distribution is generally non-standard.[11]

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 3.2, and $\widehat{\Sigma}_{\psi_2}$ and $\widehat{C}_{\psi_2\psi_1}$ are similarly defined.[12] Specifically, we extract the corresponding submatrices from $\widehat{\Sigma}_1, \widehat{\Sigma}_2$ and $\widehat{\Sigma}_{12}$ there.

Quite often, some elements of ψ_ℓ are equal while the others are not; the Wald test, as an overall test, may not detect these details, so we suggest to conduct t-tests on each element of $\psi_1 - \psi_2$ if the conclusion of the Wald test is not sharp.

6 Extensions

In this section, we discuss two extensions of our CRE model which will be used in our empirical application. The first extension considers the case where some variables do not have threshold effects on y_{it}, i.e., their coefficients remain the same over different regimes. The second extension discusses the case with more-than-one thresholds.

6.1 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. Correspondingly, we decompose $\breve{\mathbf{x}}_{it}$ as $(\mathbf{x}'_{1it}, \breve{\mathbf{x}}'_{2it})' := (\mathbf{x}'_{1it}, \mathbf{x}'_{2it}, \mathbf{z}'_i)'$ and $\underline{\theta}$ as $\left(\beta', \psi'_1, \psi'_2\right)$ or $\left(\beta'_1, \theta'_{12}, \theta'_{22}\right)$ with $\beta' = (\beta'_1, \beta'_{12}, \beta'_{22})$ being the parameter of interest, where β_1 is the coefficient of $\mathbf{x}_{1it}$, $\left(\beta'_{12}, \beta'_{22}\right)'$ is the regime-specific coefficients

[11] We can of course construct the LR test statistics based on the GMM objective functions after writing out the moment conditions for $\underline{\theta}$, but we will not pursue this target here.

[12] 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.

of $\mathbf{x}_{2it}$, and $\theta_{\ell 2} = \left(\beta_{\ell 2}', \psi_\ell'\right)'$. 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 we will not pursue such an extension in this paper although it is straightforward.

Our objective function now changes to

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

In the asymptotic distributions of $\widehat{\gamma}$ and LR statistics for γ, set $c = \left(\mathbf{0}', c_{\theta_2}'\right)'$, where $c_{\theta_2} = N^{\kappa}(\theta_{12} - \theta_{22})$. Equivalently, replace c by c_{θ_2} and the $\breve{\mathbf{x}}_{it}$ in D and V_ℓ by $\breve{\mathbf{x}}_{2it}$. In the error components model, we still have $V_\ell(\gamma) = \varsigma_\ell^2 D(\gamma)$ so that $\eta^2 = \varsigma_1^2$ and $\phi = \varsigma_2^2/\varsigma_1^2$. The asymptotic variance matrix of $\underline{\widehat{\theta}}$ changes to

$$\Sigma = \overline{M}^{-1}\Omega\overline{M}^{-1}, \tag{10}$$

where

$$\overline{M} = \sum_{t=1}^{T} E\left[\begin{pmatrix}\mathbf{x}_{1it}\\ \breve{\mathbf{x}}_{2it}1(q_{it}\le\gamma_0)\\ \breve{\mathbf{x}}_{2it}1(q_{it}>\gamma_0)\end{pmatrix}\begin{pmatrix}\mathbf{x}_{1it}', & \breve{\mathbf{x}}_{2it}'1(q_{it}\le\gamma_0), & \breve{\mathbf{x}}_{2it}'1(q_{it}>\gamma_0)\end{pmatrix}\right]$$

and

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

Appendix F shows the relationship of $\overline{M}$ and Ω with M_ℓ, Ω_ℓ and Ω_{12} and simplifications of $\overline{M}$ and Ω in the error components model, which implies that the estimators of $\overline{M}$ and Ω can be derived from Section 3.2.

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 3 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 4 and 5 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.

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 is testing

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

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 (9) but redefines

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

and

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

with

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

where $\widehat{e}_{it}^{o}$ has the same definition as in Section 5.1. Under $H_1^c:\delta_{\theta_2} = N^{-1/2}c_{\theta_2}$, $T_n(\gamma)$ 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

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

$M_2(\gamma) = \sum_{t=1}^{T}E[\check{\mathbf{x}}_{2it}\check{\mathbf{x}}_{2it}'1(q_{it}\leq\gamma)]$, $\overline{M}_2(\gamma) = \sum_{t=1}^{T}E[\check{\mathbf{x}}_{2it}\check{\mathbf{x}}_{it}'1(q_{it}\leq\gamma)]$, $M = \sum_{t=1}^{T}E[\check{\mathbf{x}}_{it}\check{\mathbf{x}}_{it}']$, and $H(\gamma) = H(\gamma,\gamma)$. The simulation procedure in Section 5.1 can be easily adapted to the current scenario.

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

6.2 Multiple Thresholds

As in Hansen (1999), we consider only the double threshold model

$$\begin{aligned} y_{it} &= (\mathbf{x}_{it}'\beta_1 + \alpha_{1i} + \sigma_1 u_{it})1(q_{it}\leq\gamma_1) + (\mathbf{x}_{it}'\beta_2 + \alpha_{2i} + \sigma_2 u_{it})1(\gamma_1 < q_{it}\leq\gamma_2) \\ &+ (\mathbf{x}_{it}'\beta_3 + \alpha_{3i} + \sigma_3 u_{it})1(q_{it} > \gamma_2) \end{aligned}$$

as extensions to higher-order threshold models are straightforward. We first discuss the estimation and inference on (γ_1,γ_2) and $(\beta_1',\beta_2',\beta_3')$, and then discuss how to determine the number of regimes and test for unobserved individual-specific threshold effects.

As for estimation, although we can add augmented variables $\mathbf{z}_i$ to each regime and estimate (γ_1,γ_2) jointly using the concentrated objective function, say $S_n(\gamma_1,\gamma_2)$, it is computationally preferable to apply the sequential estimation procedure to estimate γ_1 and γ_2 because we can reduce the number of grid searches from $O(n^2)$ to $O(n)$. This estimation procedure is proposed by Bai (1997) in the structural change context and extended by Gonzalo and Pitarakis (2002) to threshold regression. In the first stage, we use the same objective function $S_n(\theta)$ as in (7). Although the estimator of γ is consistent to either γ_1 or γ_2 (depending on which effect is stronger), the asymptotic distributions in Theorem 1 and the LR inference in Theorem 2 are not valid as shown in Yu (2019). For notational convenience, denote this first-stage estimator of γ as $\widetilde{\gamma}_1$ and the corresponding concentrated objective function as $\widetilde{S}_{1n}(\gamma)$. Given $\widetilde{\gamma}_1$, define the second-stage objective function as

$$S_{2n}(\gamma_2) = \begin{cases} S_n(\widehat{\gamma}_1,\gamma_2), & \text{if } \gamma_2 > \widehat{\gamma}_1, \\ S_n(\gamma_2,\widehat{\gamma}_1), & \text{if } \gamma_2 < \widehat{\gamma}_1, \end{cases}$$

and the second-stage estimator of γ as

$$\widehat{\gamma}_2 = \arg\min_{\gamma_2} S_{2n}(\gamma_2),$$

where the search over γ_2 must guarantee a minimum number of observations to fall in each of the three regimes. Given $\widehat{\gamma}_2$, we can refine $\widetilde{\gamma}_1$ as

$$\widehat{\gamma}_1 = \arg\min_{\gamma_1} S_{1n}(\gamma_1),$$

where

$$S_{1n}(\gamma_1) = \begin{cases} S_n(\gamma_1, \widehat{\gamma}_2), & \text{if } \gamma_1 < \widehat{\gamma}_2, \\ S_n(\widehat{\gamma}_2, \gamma_1), & \text{if } \gamma_1 > \widehat{\gamma}_2. \end{cases}$$

Given $\widehat{\gamma}_1$ and $\widehat{\gamma}_2$, $\beta_\ell, \ell = 1, 2, 3$, can be estimated by the least squares in each regime.

As for inference, suppose $\widehat{\gamma}_1 < \widehat{\gamma}_2$ without loss of generality. Then the asymptotic distributions in Section 3 and inferences in Section 4 can apply to γ_1 and $(\beta_1', \beta_2')'$ but the data usage is restricted to the first and second regimes (as if γ_2 were known asymptotically). Note here that we use the objective function $S_n(\gamma)$ based on the data with $q_{it} \leq \widehat{\gamma}_2$ rather than $S_{1n}(\gamma_1)$ to construct the LR statistic for γ_1, which can ensure the CI for γ_1 not include any value greater than $\widehat{\gamma}_2$. Similarly, we restrict data usage to the second and third regimes for γ_2 and $(\beta_2', \beta_3')'$.

To determine the number of regimes, we can conduct sequential tests based on the test statistic in Section 5.1. Specifically, in the first stage, our hypotheses are $H_0 : \# = 0$ vs. $H_1 : \# = 1$, where # is the number of thresholds. Let our test statistic be $\sup_{\gamma\in\Gamma} \|T_n(\gamma)\|^2$; if we reject the null, then continue to the second stage. In the second stage, the hypotheses are $H_0 : \# = 1$ vs. $H_1 : \# = 2$ and our test statistic is $\max\left\{\sup_{\gamma\in\Gamma_1} \|T_{1n}(\gamma)\|^2, \sup_{\gamma\in\Gamma_2} \|T_{2n}(\gamma)\|^2\right\}$, where $T_{1n}(\gamma)$ and $T_{2n}(\gamma)$ are constructed in the same way as $T_n(\gamma)$ but use data with $q_{it} \leq \widetilde{\gamma}_1$ and $q_{it} > \widetilde{\gamma}_1$ respectively, and Γ_1 and Γ_2 are constructed similarly as Γ but based on the q_{it}'s in the two regimes. In the simulation method, the same ξ_i^* is used for $\{\widehat{e}_{it}^o\}_{t=1}^T$ regardless of $\widehat{e}_{it}^o$ falls in the first regime or the second regime. In the third stage, the hypotheses are $H_0 : \# = 2$ vs. $H_1 : \# = 3$ and our test statistic is $\max\left\{\sup_{\gamma\in\Gamma_1} \|T_{1n}(\gamma)\|^2, \sup_{\gamma\in\Gamma_2} \|T_{2n}(\gamma)\|^2, \sup_{\gamma\in\Gamma_3} \|T_{3n}(\gamma)\|^2\right\}$, where $T_{1n}(\gamma)$, $T_{2n}(\gamma)$ and $T_{3n}(\gamma)$ are constructed in the same way as $T_n(\gamma)$ but use data with $q_{it} \leq \widehat{\gamma}_1$, $\widehat{\gamma}_1 < q_{it} \leq \widehat{\gamma}_2$ and $q_{it} > \widehat{\gamma}_2$ respectively, and $\Gamma_\ell, \ell = 1, 2, 3$, are constructed similarly as Γ and may be different from Γ_1 and Γ_2 in the second stage. Continue this testing procedure until the null cannot be rejected.

In testing for unobserved individual-specific threshold effects, we can apply the tests in Section 5.2 to the data with $q_{it} \leq \widetilde{\gamma}_1$ and $q_{it} > \widetilde{\gamma}_1$ to check whether $\psi_1 = \psi_2$ and $\psi_2 = \psi_3$ separately, where ψ_1, ψ_2 and ψ_3 are the coefficients of augmented variables in the three regimes. Of course, we can test $\psi_1 = \psi_2$ and $\psi_2 = \psi_3$ jointly in the same way.

7 Empirical Application

We apply our testing and estimating procedures to an empirical application in this section. Our application is about firms' investment behaviour with financing constraints, which was analyzed in Hansen (1999) using the differencing method in a fixed effects model.

Fazzari et al. (1988) (FHP hereafter) argue that the effect of a firm's cash flow on its investment is different with and without financing constraints. Only if the firm faces constraints on external financial markets, its cash flow will positively influence its investment. This obviously suggests a threshold model to describe a firm's investment behaviour. FHP use a low dividend to income ratio to indicate the existence of financing constraints as a financially constrained firm will usually retain earnings instead of paying dividends. Actually, FHP consider a double-threshold (rather than one-threshold) model with the two threshold values arbitrarily chosen. Since we use the dataset in Hansen (1999), we choose the ratio of long-term debt to assets as our threshold variable. Now, a high value of this ratio indicates financial constraints.

Following Hansen (1999), we start with a SPTR model of three regimes to model the relationship between a firm's investment and its cash flow:[13]

$$\begin{aligned} I_{it} = \mathbf{x}_{1it}'\beta_1 &+ (CF_{i,t-1}\beta_{12} + \alpha_{1i} + \sigma_1 u_{it})\,1(D_{i,t-1} \le \gamma_1) \\ &+ (CF_{i,t-1}\beta_{22} + \alpha_{2i} + \sigma_2 u_{it})\,1(\gamma_1 < D_{i,t-1} \le \gamma_2) \\ &+ (CF_{i,t-1}\beta_{32} + \alpha_{3i} + \sigma_3 u_{it})\,1(D_{i,t-1} > \gamma_2), \\ &i = 1, \cdots, 565, t = 1, \cdots, 14, \end{aligned} \tag{11}$$

where $N = 565$, $T = 14$, I_{it} is the ratio of investment to capital, CF_{it} is the ratio of cash flow to assets, D_{it} is the ratio of long-term debt to assets, $\mathbf{x}_{1it}' = \left(Q_{i,t-1}, Q_{i,t-1}^2, Q_{i,t-1}^3, D_{i,t-1}, Q_{i,t-1}D_{i,t-1}\right)$ with Q_{it} being the ratio of total market value to assets, and all stock variables are measured at the end of year. The only difference of our model from Hansen's is that we allow α_i to be regime-specific. This model focuses attention on the threshold effects of $CF_{i,t-1}$, and maintains a constant effect of $Q_{i,t-1}$ and $D_{i,t-1}$ on I_{it} across regimes. Nonlinear terms such as $Q_{i,t-1}^2, Q_{i,t-1}^3$ and $Q_{i,t-1}D_{i,t-1}$ are introduced only to reduce spurious correlations due to omitted variable bias. As a result, in calculating $\overline{\mathbf{x}}_i$, we only include $Q_{i,t-1}, CF_{i,t-1}$ and $D_{i,t-1}$. Note also that the averaging in $\overline{\mathbf{x}}_i$ starts from $t = 0$ and ends at $t = T$. In summary, our $\mathbf{z}_i' = \left(\overline{Q}_i, \overline{CF}_i, \overline{D}_i, 1\right)$ with $\underline{\mathbf{z}}_i = 1$.

H_0 vs. H_1	Test Statistic	p-value
$\# = 0$ vs. $\# = 1$	19.647	0.014
$\# = 1$ vs. $\# = 2$	11.063	0.630

Table I: Tests for Threshold Effects

We first determine the number of thresholds using the testing procedure in Section 6.2. The results are reported in Table I, where we use 500 replications in simulating the p-values.[14] Different from FHP and Hansen (1999), our tests find only one threshold rather than two. This may be surprising since we also explore the heterogeneity of α_i while Hansen (1999) assumes homogeneity. This can be understood from the Wald test where the power decreases as the number of restrictions increases. From Table 4 of Hansen (1999), the third regime contains much less observations than the other two. We actually absorb the third regime into the second one. Furthermore, $\widehat{\beta}_{12} = 0.063$, $\widehat{\beta}_{22} = 0.098$ and $\widehat{\beta}_{32} = 0.039$ in Hansen (1999) are not increasing, which is unexpected from the economic theory; actually, the third regime is a spurious outcome of assuming homogeneity of α_i. We will provide more intuitive evidences on the number of regimes below as we show the LR statistics for γ. In summary, our model is

$$\begin{aligned} I_{it} = \mathbf{x}_{1it}'\beta_1 &+ (CF_{i,t-1}\beta_{12} + \alpha_{1i} + \sigma_1 u_{it})\,1(D_{i,t-1} \le \gamma) \\ &+ (CF_{i,t-1}\beta_{22} + \alpha_{2i} + \sigma_2 u_{it})\,1(D_{i,t-1} > \gamma). \end{aligned}$$

We next estimate γ and construct CIs for it. The results are reported in Figure 2 and Table II. From Figure 2, it is obvious that $LR_{1n}(\gamma) \le LR_n(\gamma) \le LR_{2n}(\gamma)$ for any $\gamma \in \Gamma$, so the widths of CIs based

[13] Note that Seo and Shin (2016) model the same data using DPTR.

[14] As for Γ, Γ_1 and Γ_2, we approximate them by discrete quantile points between the 1% and 95% quantiles of unique q values (because q's distribution has a point mass at 0) on $[0, \infty)$, $[0, \widehat{\gamma}]$, and $[\widehat{\gamma}, \infty)$. If the number of q_i's on the respective range is greater than 400, we use 400 quantile points with equally spaced quantile indices for approximation, and if less than 400, we use the middle points of contiguous q_i's for approximation. The resulting approximation sets Γ_n, Γ_{1n} and Γ_{2n} contain 400, 226 and 400 points, respectively.

on $LR_{2n}(\gamma), LR_n(\gamma)$ and $LR_{1n}(\gamma)$ should be increasing, which is explicitly shown in Table II. The CIs based on $LR_n(\gamma)$ and $LR_{2n}(\gamma)$ are similar, but that based on $LR_{1n}(\gamma)$ is close to the whole Γ, and is not suggested in practice. The inset magnified portion of Figure 2 shows the subtle differences between $LR_n(\gamma)$ and $LR_{2n}(\gamma)$ around $\widehat{\gamma}$. Figure 2 also intuitively confirms the conclusion of Table I – the data imply only one threshold. Figure 1 of Hansen (1999) shows that his $LR_n(\gamma)$ has a second dip around 0.53 besides at his $\widehat{\gamma}_1 = 0.0157$ (which is close to our $\widehat{\gamma}$), but this does not happen in either of our LR statistics. Based on our $\widehat{\gamma}$, we report the percentage of firms in each regime by year in Table III. We are basically combining the "high debt" class and the "medium debt" class of Table 4 in Hansen (1999). We also see a decreasing trend in the number of the "low debt" class.

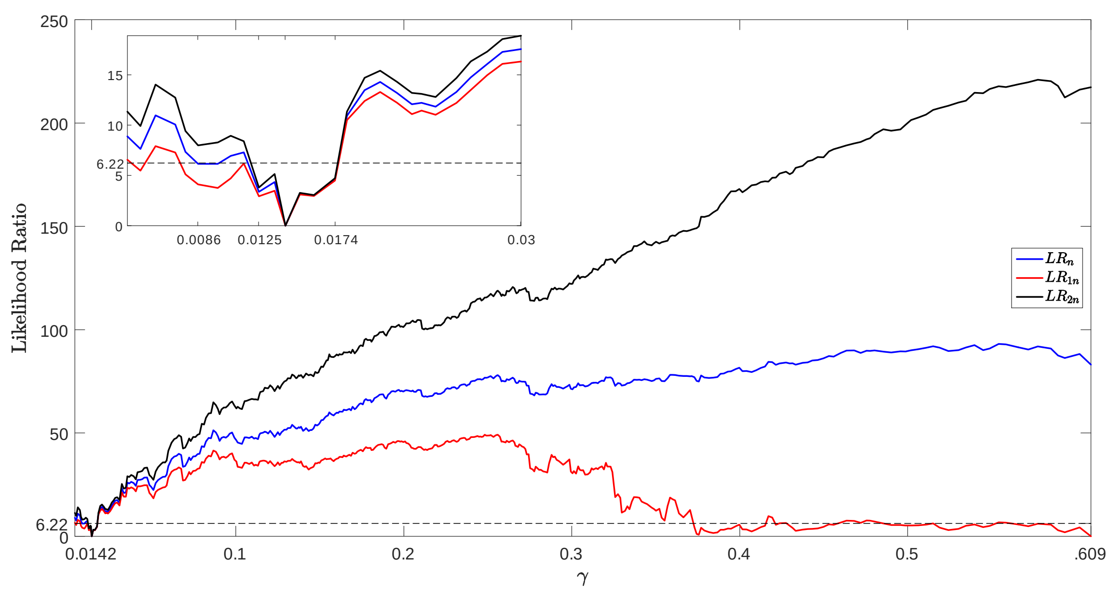

Figure 2: Three CIs for γ

Parameters	Our Estimates	Our CIs	Critical Values of LR Statistics
		$[0.0086, 0.0174]$	6.220
γ	0.0142	$[0.0049, 0.6091]$	6.220
		$[0.0125, 0.0174]$	6.220
		$[0.0268, 0.0778]$	SE: 0.0130
β_{12}	0.0523	$[0.0199, 0.0846]$	9.851
		$[0.0188, 0.0858]$	157.896
		$[0.0522, 0.1103]$	SE: 0.0148
β_{22}	0.0812	$[0.0487, 0.1138]$	15.578
		$[0.0525, 0.1099]$	88.712

Table II: Parameter Estimates and 95% CIs

Note: The three CIs are reported in the following order: for γ, (LR_n, LR_{1n}, LR_{2n}), for β_{12} and β_{22}, (t, LR_n, LR_{2n}), and SE means standard error

	Year													
Firm Class	1974	1975	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987
$D_{i,t-1} \leq 0.0142$	16	13	13	14	15	13	13	11	10	10	10	9	9	11
$D_{i,t-1} > 0.0142$	84	87	87	86	85	87	87	89	90	90	90	91	91	89

Table III: Percentage of Firms in Each Regime by Year

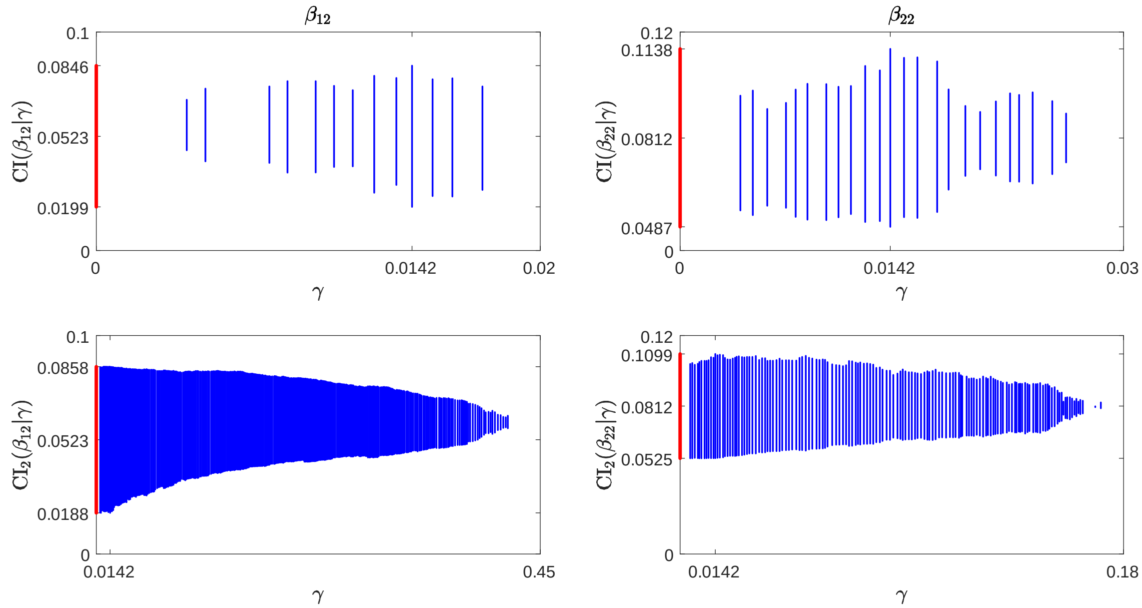

Figure 3: Construction of CIs for β_{12} and β_{22} Based on $LR_n(\cdot,\cdot)$ and $LR_{2n}(\cdot,\cdot)$

We then estimate β and conduct inferences on them. Since β_{12} and β_{22} are of main interest, we neglect β_1 in Figure 3 and Table II. Figure 3 shows how the CIs for β_{12} and β_{22} are constructed based on $LR_n(\cdot,\cdot)$ and $LR_{2n}(\cdot,\cdot)$. Combining Figure 3 and Table II, we can draw the following conclusions. First, our $\widehat{\beta}_{12}$ is close to Hansen's $\widehat{\beta}_{12}$ $(= 0.063)$, and $\widehat{\beta}_{22}$ is between his $\widehat{\beta}_{22}$ $(= 0.098)$ and $\widehat{\beta}_{32}$ $(= 0.039)$. Since our $\widehat{\beta}_{22}$ is greater than our $\widehat{\beta}_{12}$, FHP's theory is confirmed. Also, our standard error (SE) of $\widehat{\beta}_{12}$ is close to Hansen's White SE, 0.014, and the SE of our $\widehat{\beta}_{22}$ is between those of his $\widehat{\beta}_{22}$ and $\widehat{\beta}_{32}$ (0.010 and 0.031). Second, the number of γ values in Γ_n involved in inverting $LR_{2n}(\cdot,\cdot)$ is much more than that in inverting $LR_n(\cdot,\cdot)$. Although $LR_n(\cdot,\cdot) \leq LR_{2n}(\cdot,\cdot)$, Table II shows that the critical values associated with the latter are also much larger than those associated with the former. Third, there is no obvious trend between $\text{CI}(\cdot|\gamma)$ and γ, e.g., the centers of $\text{CI}(\cdot|\gamma)$ do not trend upward or downward as γ increases; the same conclusion applies to $\text{CI}_2(\cdot|\gamma)$, where $\text{CI}_2(\cdot|\gamma)$ is similarly defined as in (8) with LR_{2n} replacing LR_n and the new critical value $\widehat{c}_{2\alpha}$ replacing $\widehat{c}_\alpha$. This confirms the independence between $\widehat{\beta}$ and $\widehat{\gamma}$. As a result, only a few $\text{CI}(\cdot|\gamma)$ and $\text{CI}_2(\cdot|\gamma)$ intervals for γ around $\widehat{\gamma}$ are relevant to the ultimate CIs (as shown on the y-axis); actually, the CIs for β_{12} and β_{22} based on $LR_n(\cdot,\cdot)$ are exactly the same as $\text{CI}(\beta_{12}|\widehat{\gamma})$ and $\text{CI}(\beta_{22}|\widehat{\gamma})$. This is dramatically different from the regular case where the target parameter and the nuisance parameter are not statistically independent such that the projection CI is much longer than the CI with the nuisance parameter fixed at its estimate. Fourth, the projection CI need not be wider than the t-type CI, and the CI based on $LR_{2n}(\cdot,\cdot)$ need not be narrower than that based on $LR_n(\cdot,\cdot)$. As mentioned in the first point above, the critical values associated with the two LR statistics are different, which is the reason why the lengths of the two LR-CIs for β are not comparable. This is very different from the LR-CIs for γ, where the critical values for all three LR statistics are the same such that the lengths of the three LR-CIs are sortable. We think the LR-CI based on $LR_{2n}(\cdot,\cdot)$ is most preferable. When less data points are present such as the left regime which contains 936 data points, the CI is wide to indicate the uncertainty in the estimation of β and the impact of the uncertainty of $\widehat{\gamma}$, while when more data points are available such as the right regime with 6874 data points, the normality approximation is appropriate and the uncertainty in $\widehat{\gamma}$ is dominated, the CI is close to the t-type CI.

We finally use the procedure in Section 5.2 to test the existence of unobserved individual-specific threshold effects. The testing results are reported in Table IV. The p-value of the Wald test is 0.124, so the conclusion of rejection is not clear-cut. However, if we conduct t-tests on each element of $\psi_1 - \psi_2$, the whole picture is much clearer. Actually, three of four elements of $\psi_1 - \psi_2$ are significantly different from zero; in other words, the conclusion of the Wald test is blurred by only one element of $\psi_1 - \psi_2$. In summary, our CRE estimation and inferences above are justified, and the differencing method is not reliable in this application.

Parameters	Our Estimates	Test Statistics	p-values
$\overline{Q}_i$	-0.0013	-19.299	0
$\overline{CF}_i$	-0.0434	-2.679	0.0074
$\overline{D}_i$	0.0258	0.601	0.548
1	0.0130	9.542	0
Wald	-	7.232	0.124

Table IV: Tests for Unobserved Individual-Specific Threshold Effects

Note: parameters mean the associate elements of $\psi_1 - \psi_2$ with the listed variables

8 Conclusion and Discussions

This paper considers estimation and inferences of panel threshold regression with unobserved individual-specific threshold effects. A key observation is that within-regime differencing cannot eliminate the endogeneity problem induced by the fixed effects, so the CRE models are suggested as an alternative solution. This solution is valid for both the static and dynamic models and regardless of the presence of unobserved individual-specific threshold effects. Although the forms of endogeneity in the CRE models are less general than in the fixed effects models, they are more practical since no nonparametric components are involved in the estimation and inference procedures. Recall that the IDKE of Yu and Phillips (2018) can indeed provide a consistent estimator of the threshold point γ even if α_{1i} and α_{2i} were fixed effects, but when the dimension of x_{it} is large the IDKE is not practical (because we should condition on X_i and X_i^t in computing the conditional mean of y_{it} in SPTR and DPTR, respectively). Given a consistent estimator of γ, the within-regime differencing can be applied to generate consistent estimators of β. This estimation procedure based on the IDKE should serve as a benchmark for any future solution for the fixed effects models.

A natural extension of the CRE models in this paper is the flexible CRE models of Bester and Hansen (2007). In a likelihood framework, they replace the parametric form of CRE by a sieve form and provide many interesting identification results for regular parameters like β in this paper. However, identification is not an issue for PTR given that the estimation procedure based on the IDKE indeed provides identification. Also, when a sieve CRE is introduced, the estimation suffers from a similar curse-of-dimensionality problem as the IDKE.

Another possible solution to eliminate the effects of $\alpha_{\ell i}$ is to employ the functional differencing of Bonhomme (2012). However, the main focus of functional differencing is to provide identification for regular parameters by moment conditions in a likelihood framework. As mentioned above, identification for the nonregular parameter γ is not a problem, and as emphasized in Yu et al. (2018a), it is better not to identify γ by moment conditions. Of course, we can think of a generalized differencing procedure that estimates γ and β by an M-estimator instead of a Z-estimator and does not involve any nonparametric components in the objective function, but we are not aware of any such differencing scheme yet.

One possible remedy to the bias introduced by the incidental parameters problem in nonlinear panel models is to let T diverge to infinity (although at a lower rate than N) and then debias; see Arellano and Hahn (2007) for a summary of the literature. Indeed, when T goes to infinity, we show at the end of Appendix D that the within-regime-differencing estimator of γ in Section 2 is consistent. It is quite possible that this result can be extended to more general PTR models with covariates, but it is not clear how to debias $\widehat{\gamma}$ when T is relatively small since the existing literature all aims for debiasing estimators of regular parameters.

Finally, we do not consider interactive fixed effects in this paper. Ke et al. (2018) discuss such effects but assume they are invariant in the two regimes. It is possible to extend their estimation procedure to the more general case with regime-specific effects, and we reserve it as a promising future research topic. Since the setup here is more general than that in this paper and Ke et al. essentially estimate the incidental parameters directly, we must let T diverge to infinity to guarantee the consistency of γ and β.

References

Arellano, M. and S. Bond, 1991, Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations, *Review of Economic Studies*, 58, 277-297.

Arellano, M. and J. Hahn, 2007, Understanding Bias in Nonlinear Panel Models: Some Recent Developments, *Advances in Economics and Econometrics*, Theory and Applications, Ninth World Congress, Volume III, R. Blundell, W.K. Newey and T. Persson, eds., New York: Cambridge University Press, Chapter 12, 381-409.

Bai, J., 1997, Estimating Multiple Breaks One At A Time, *Econometric Theory*, 13, 315-352.

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

Bester, A. and C. Hansen, 2007, Flexible Correlated Random Effects Estimation in Panel Models with Unobserved Heterogeneity, mimeo.

Bonhomme, S., 2012, Functional Differencing, *Econometrica*, 80, 1337-1385.

Chamberlain, G., 1982, Multivariate Regression Models for Panel Data, *Journal of Econometrics*, 18, 5-46.

Chamberlain, G., 1984, Panel Data, *Handbook of Econometrics*, Vol. 2, Z. Griliches and M. Intriligator, eds., Elsevier Science B.V., Ch. 22, 1247-1318.

Chan, K.S., 1993, Consistency and Limiting Distribution of the Least Squares Estimator of a Threshold Autoregressive Model, *Annals of Statistics*, 21, 520-533.

Chan, K.S. and R.S. Tsay, 1998, Limiting Properties of the Least Squares Estimator of a Continuous Threshold Autoregressive Model, *Biometrika*, 85, 413-426.

Dang, V.A., M. Kim and Y. Shin, 2012, Asymmetric Capital Structure Adjustments: New Evidence from Dynamic Panel Threshold Models, *Journal of Empirical Finance*, 19, 465-482.

Fazzari, S.M., R. G. Hubbard and B.C. Petersen, 1988, Financing Constraints and Corporate Investment, *Brookings Papers on Economic Activity*, 1988, 141-195.

Gøgens, T. and A.H. Würtz, 2019, Threshold Regression with Endogeneity for Short Panels, *Econometrics*, 7, 23.

Gonzalo, J. and J.-Y. Pitarakis, 2002, Estimation and Model Selection Based Inference in Single and Multiple Threshold Models, *Journal of Econometrics*, 110, 319-352.

Hansen, B.E., 1996, Inference when a Nuisance Parameter is not Identified under the Null Hypothesis, *Econometrica*, 64, 413-430.

Hansen, B.E., 1999, Threshold Effects in Non-dynamic Panels: Estiamtion, Testing and Inference, *Journal of Econometrics*, 93, 345-368.

Hansen, B.E., 2000, Sample Splitting and Threshold Estimation, *Econometrica*, 575-603.

Hansen, B.E., 2017, Regression Kink With an Unknown Threshold, *Journal of Business & Economic Statistics*, 35, 228-240.

Hsiao, C., M. Pesaran and A.K. Tahmiscioglu, 2002, Maximum Likelihood Estimation of Fixed Effects Dynamic Panel Data Models Covering Short Time Periods, *Journal of Econometrics*, 109, 107-150.

Ke, M., K. Li and L. Su, 2018, Panel Threshold Models with Interactive Fixed Effects, *Journal of Econometrics*, 219: 137-170.

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

Kourtellos, A., T. Stengos and C.-M. Tan, 2016, Structural Threshold Regression, *Econometric Theory*, 32, 827-860.

Kremer, S., A. Bick and D. Nautz, 2013, Inflation and Growth: New Evidence from a Dynamic Panel Threshold Analysis, *Empirical Economics*, 44, 861-878.

Mundlak, Y., 1978, On the Pooling of Time Series and Cross Section Data, *Econometrica*, 46, 69-85.

Ramírez-Rondán, N.R., 2020, Maximum Likelihood Estimation of Dynamic Panel Threshold Models, *Econometric Reviews*, 39: 260-276.

Seo, M.H. and Y. Shin, 2016, Dynamic Panels with Threshold Effect and Endogeneity, *Journal of Econometrics*, 195, 169-186.

Shin, Y. and M. Kim, 2011, Bootstrap-based Bias Corrected Within Estimation of Threshold Regression Models in Dynamic Panels, mimeo.

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

Wang, C.-H. and E.S. Lin, 2010, Panel Threshold Regression Models with Endogenous Threshold Variables, mimeo.

Wooldridge, J.M., 2000, A Framework for Estimating Dynamic, Unobserved Effects Panel Data Models with Possible Feedback to Future Explanatory Variables, *Economics Letters*, 68, 245-250.

Wooldridge, J.M., 2005, Simple Solutions to the Initial Conditions Problem in Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity, *Journal of Applied Econometrics*, 20, 39-54.

Wooldridge, J.M., 2010, *Econometric Analysis of Cross Section and Panel Data*, 2nd ed., Cambridge, Mass.: MIT Press.

Yu, P., 2012, Likelihood Estimation and Inference in Threshold Regression, *Journal of Econometrics*, 2012, 167, 274-294.

Yu, P., 2013, Integrated Quantile Threshold Regression and Distributional Threshold Effects, mimeo.

Yu, P., 2015, Adaptive Estimation of the Threshold Point in Threshold Regression, *Journal of Econometrics*, 189, 83-100.

Yu, P., 2019, Asymptotics for Threshold Regression Under Misspecification, mimeo.

Yu, P. and X. Fan, 2021, Threshold Regression With A Threshold Bounary, *Journal of Business and Economic Statistics*, 39, 953-971.

Yu, P. and P.C.B. Phillips, 2018, Threshold Regression with Endogeneity, *Journal of Econometrics*, 203, 50-68.

Yu. P., Q. Liao and P.C.B. Phillips, 2018a, Inferences and Specification Testing in Threshold Regression with Endogeneity, mimeo.

Yu. P., Q. Liao and P.C.B. Phillips, 2018b, New Control Function Approaches in Threshold Regression with Endogeneity, mimeo.

Supplementary Document

Appendix A: Analyses for DPTR

In this appendix, we first discuss a similar (maybe even more severe) problem as in Section 2 for DPTR in applying within-regime differencing, and then discuss how to build CRE models for DPTR.

Within-Regime First-Differencing in DPTR

In (3), 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.[15] Anyway, 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 the first differencing in the usual dynamic panel model is not even applicable (of course, inconsistent).

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 $\gamma=\gamma_0$,

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

so we expect the objective function

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

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, the summation $\sum_{t=2}^{T}\left[y_{it}-y_{i\tau}-(\mathbf{x}_{it}-\mathbf{x}_{i\tau})'\beta_1\right]^2 1($ in $S_n(\beta,\gamma)$ should be

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

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

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

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

where note that τ and τ' depend on γ (and t). The case with $D_i^{\pm}\left(\gamma\right)=0$ can be handled similarly as in SPTR. However, for any γ value, $P\left(D_i^{\pm}\left(\gamma\right)=1\right)>0$ if T is fixed. More importantly, for different γ values, the sets of i's such that $D_i^{\pm}\left(\gamma\right)=1$ are different; in other words, for different γ's, different observations are used in $S_n\left(\beta,\gamma\right)$. As a result, $S_n\left(\beta,\gamma\right)$ 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

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

where $\mathbf{z}_i'=\left(\overline{\mathbf{x}}_i',\underline{\mathbf{z}}_i'\right)$ 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,

$$\begin{aligned}E\left[y_{it}|X_i^t\right]&=\left(\mathbf{x}_{it}'\beta_1+\mathbf{z}_i'\psi_1+\pi_1y_{i0}\right)1\left(q_{it}\leq\gamma_0\right)+\left(\mathbf{x}_{it}'\beta_2+\mathbf{z}_i'\psi_2+\pi_2y_{i0}\right)1\left(q_{it}>\gamma_0\right),\\&=:\breve{\mathbf{x}}_{it}'\theta_1 1\left(q_{it}\leq\gamma_0\right)+\breve{\mathbf{x}}_{it}'\theta_2 1\left(q_{it}>\gamma_0\right),t=1,\cdots,T,\end{aligned}\tag{12}$$

and the error term takes the same form as (6), where $\breve{\mathbf{x}}_{it}:=\left(\mathbf{x}_{it}',\mathbf{z}_i',y_{i0}\right)'$, and $\theta_\ell=\left(\beta_\ell',\psi_\ell',\pi_\ell\right)'$. 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 multicollinear problem when $T=1$, but not when $T>1$. The objective function is

$$S_n\left(\theta\right)=\sum_{i=1}^{N}\sum_{t=1}^{T}\left[y_{it}-\breve{\mathbf{x}}_{it}'\theta_1 1\left(q_{it}\leq\gamma\right)-\breve{\mathbf{x}}_{it}'\theta_2 1\left(q_{it}>\gamma\right)\right]^2,\tag{13}$$

where $\theta=\left(\gamma,\beta',\psi',\pi'\right)'$with $\beta'=\left(\beta_1',\beta_2'\right)$, $\psi'=\left(\psi_1',\psi_2'\right)$, $\pi'=\left(\pi_1,\pi_2\right)$, or $\theta=\left(\gamma,\underline{\theta}'\right)':=\left(\gamma,\theta_1',\theta_2'\right)'$. Denote the resulting estimator as $\widehat{\theta}=\left(\widehat{\gamma},\widehat{\beta}',\widehat{\psi}',\widehat{\pi}'\right)'$ or $\left(\widehat{\gamma},\widehat{\underline{\theta}}'\right)'$ and the residuals as

$$\widehat{e}_{it}=y_{it}-\breve{\mathbf{x}}_{it}'\widehat{\theta}_1 1\left(q_{it}\leq\widehat{\gamma}\right)-\breve{\mathbf{x}}_{it}'\widehat{\theta}_2 1\left(q_{it}>\widehat{\gamma}\right).$$

Obviously, the structure of the estimation problem in DPTR is the same as that in SPTR except for new definitions of $\breve{\mathbf{x}}_{it}$ and θ_ℓ. As a result, the two-step estimation procedure in SPTR can be applied.

How to rationalize our objective function? If $u_{it}|X_i,y_{i,t-1},y_{i,t-2},\cdots,y_{i0}\sim N\left(0,1\right)$, $a_{\ell i}|X_i,y_{i0}\sim N\left(0,\sigma_{a_\ell}^2\right)$ and u_{it}, a_{1i} and a_{2i} are independent of each other, then the likelihood function of $\left(y_{iT},\cdots,y_{i1}\right)_{i=1}^N$

given X_i, y_{i0} is

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

where $a_{\ell i} = \sigma_{a_\ell} a_{\ell i}^*$, $\vartheta = \left(\theta', \boldsymbol{\sigma}', \boldsymbol{\sigma}_a'\right)'$ with $\boldsymbol{\sigma}' = (\sigma_1, \sigma_2)$ and $\boldsymbol{\sigma}_a' = (\sigma_{a_1}, \sigma_{a_2})$, 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

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

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

$$\Sigma_{T\times T} = \begin{pmatrix} \sigma^2+\sigma_a^2 & \sigma^2 & \cdots & \sigma^2 \\ \sigma^2 & \sigma^2+\sigma_a^2 & \cdots & \sigma^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma^2 & \sigma^2 & \cdots & \sigma^2+\sigma_a^2 \end{pmatrix}, \mathbf{y}_i = \begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{pmatrix}, \mathbf{X}_i = \begin{pmatrix} \mathbf{x}_{i1} \\ \mathbf{x}_{i2} \\ \vdots \\ \mathbf{x}_{iT} \end{pmatrix},$$

$$\mathbf{1}_{i\le\gamma} = \begin{pmatrix} 1(q_{i1}\le\gamma) \\ 1(q_{i2}\le\gamma) \\ \vdots \\ 1(q_{iT}\le\gamma) \end{pmatrix}, \mathbf{1}_{i>\gamma} = \begin{pmatrix} 1(q_{i1}>\gamma) \\ 1(q_{i2}>\gamma) \\ \vdots \\ 1(q_{iT}>\gamma) \end{pmatrix}, \mathbf{1}_T = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}_{T\times 1};$$

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 $\sigma_1 \ne \sigma_2$, $a_{1i} \ne a_{2i}$ 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 on $a_{\ell i}$.

As in SPTR, we define some further notations and specify some assumptions before stating the asymptotic distributions of $\widehat{\gamma}$ and $\widehat{\beta}$. Let $D(\gamma), V_\ell(\gamma), M, M(\gamma), \Omega_\ell, \Omega_{12}$ take the same form as in SPTR with the new definition of $\check{\mathbf{x}}_{it}$, and $f_t(\gamma)$ and $f_{\tau|t}(\gamma_1|\gamma_2)$ have the same definition as in SPTR.

Assumption DP:

Conditions (iv) and (vi)-(x) hold as in Assumption SP.
(i) $\{\underline{\mathbf{x}}_{it}, \underline{\mathbf{z}}_i, y_{it}\}_{t=0}^T$ are i.i.d. across i; T is fixed and $N \to \infty$.
(ii) For each i, $E\left[a_{\ell i}|X_i^T\right] = 0$ and $E\left[u_{it}|X_i^t\right] = 0$.
(iii) For each $j = 1, \cdots, d_{\underline{\mathbf{x}}}$, $P\left(x_{i1}^j = \cdots = x_{iT}^j\right) < 1$, where x_{it}^j is the jth element of $\underline{\mathbf{x}}_{it}$.
(v) For some fixed $c = \left(c_\beta', c_\psi', c_\pi\right)'$, $\delta_N := \theta_1 - \theta_2 = cN^{-\kappa}$, where $0 < \kappa < 1/2$.

As in Assumption SP, Condition (iv) can be expressed in terms of $\underline{\mathbf{x}}_{it}, u_{it}, y_{it}, \underline{\mathbf{z}}_i$ and $a_{\ell i}$, but the current form is more convenient. The comments on other conditions in Assumption SP can be applied here, so not repeated.

The following theorem states the asymptotic distributions of $\widehat{\gamma}$ and $\widehat{\beta}$.

Theorem 7 *Under Assumption DP, $N^{1-2\kappa}(\widehat{\gamma}-\gamma_0)$ and $N^{1/2}\left(\widehat{\theta}_\ell-\theta_\ell\right)$ have the same form of asymptotic distributions as in Theorem 1 except that $D, V_\ell, M_\ell, \Omega_\ell$ and Ω_{12} are adjusted with the new $\check{\mathbf{x}}_{it}$.*

The comments after Theorem 1 can be applied here except the following two small notes. First, when $\alpha_{1i}=\alpha_{2i}$, we now set $\left(c'_\psi, c_\pi\right)'=\mathbf{0}$ in the asymptotic distribution of $\widehat{\gamma}$. Second, the error components model implies $E\left[e^2_{\ell it}|X^t_i\right]=E\left[e^2_{\ell it}\right]=E\left[a^2_{\ell i}\right]+\sigma^2_\ell=\varsigma^2_\ell$, $E\left[e_{\ell it}e_{\ell i\tau}|X^{t\vee\tau}_i\right]=E\left[a^2_{\ell i}\right]=c_\ell$, and $E\left[e_{1it}e_{2i\tau}|X^{t\vee\tau}_i\right]=E\left[a_{1i}a_{2i}\right]=c_{12}$ for $t,\tau=1,\cdots,T$ and $\tau\neq t$ such that V_ℓ, Ω_ℓ and Ω_{12} can be simplified. Because the structures of SPTR and SPTR in the CRE models are similar, the analyses after Section 3.1 can also be applied to DPTR. For example, in constructing $LR_n(\gamma,\beta_{11})$, for each $\gamma\in\Gamma$, run least squares of $y_{it}-x^1_{it}\beta_{11}$ on $\left(\mathbf{x}^{-1\prime}_{it}, \mathbf{z}'_i, y_{i0}\right)'$ for (i,t)'s such that $q_{it}\le\gamma$ and y_{it} on $(\mathbf{x}'_{it},\mathbf{z}'_i)'$ for (i,t)'s such that $q_{it}>\gamma$ separately to obtain $\widehat{\theta}_{12}(\gamma,\beta_{11})$ and $\widehat{\theta}_2(\gamma)$, and the concentrated objective function

$$S_n(\gamma,\beta_{11})=S_n\left(\gamma,\beta_{11},\widehat{\theta}_{12}(\gamma,\beta_{11}),\widehat{\theta}_2(\gamma)\right).$$

Similarly, in Section 6.1, the augmented variables are $(\mathbf{z}'_i, y_{i0})'$ rather than $\mathbf{z}_i$ as in SPTR and the corresponding cofficients are changed from ψ_ℓ to $\left(\psi'_\ell,\pi_\ell\right)'$. For another example, in testing for unobserved individual-specific threshold effects, the null is

$$H_0:\psi_1=\psi_2 \text{ and } \pi_1=\pi_2.$$

Now, the Wald statistic is

$$W_n=N\left(\widehat{\psi}'_1-\widehat{\psi}'_2,\widehat{\pi}_1-\widehat{\pi}_2\right)\left(\widehat{\Sigma}_{11}-\widehat{C}_{12}-\widehat{C}_{21}+\widehat{\Sigma}_{22}\right)^{-1}\left(\widehat{\psi}'_1-\widehat{\psi}'_2,\widehat{\pi}_1-\widehat{\pi}_2\right)',$$

which converges in distribution to $\chi^2_{1+d_{\psi_\ell}}$ under the null, where $\widehat{\Sigma}_{11}$ and $\widehat{C}_{12}$ are consistent estimates of the asymptotic variance matrix of $\left(\widehat{\psi}'_\ell,\widehat{\pi}_\ell\right)'$ and the asymptotic covariance matrix between $\left(\widehat{\psi}'_1,\widehat{\pi}_1\right)'$ and $\left(\widehat{\psi}'_2,\widehat{\pi}_2\right)'$ in Section 3.2, and $\widehat{\Sigma}_{22}$ and $\widehat{C}_{21}$ are similarly defined.

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. The proof of Theorem 7 is similar to that of Theorem 1, so omitted.

Proof of Theorem 1. First, $\widehat{\theta}=\arg\min_\theta S_n(\theta)$ implies

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

where from Lemma 3,

$$\mathbb{M}_n(h)=u'_1M_1u_1+u'_2M_2u_2-2W_n(u)+C_n(v),$$

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

$$C(v) = \begin{cases} c'Dc\,|v| + 2\sqrt{c'V_1c}B_1(|v|), \\ c'Dcv + 2\sqrt{c'V_2c}B_2(v), \end{cases} =: \begin{cases} \mu\,|v| + 2\sqrt{\varpi_-}B_1(|v|), & \text{if } \nu \le 0, \\ \mu\,|v| + 2\sqrt{\varpi_+}B_2(v), & \text{if } \nu > 0, \end{cases} \tag{14}$$

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

$$W = \begin{pmatrix} W_1 \\ W_2 \end{pmatrix} \sim N\left(\mathbf{0}, \begin{pmatrix} \Omega_1 & \Omega_{12} \\ \Omega_{12}' & \Omega_2 \end{pmatrix}\right), \tag{15}$$

$\mathbf{C}_{\min}(\mathbb{R}^{d_\theta})$ is defined as the subset of continuous functions $x(\cdot) \in \mathbf{B}_{\text{loc}}(\mathbb{R}^{d_\theta})$ 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 4. We now check $\mathbb{M}(h) \in \mathbf{C}_{\min}(\mathbb{R}^{d_\theta})$. Because u and v are separable in $\mathbb{M}(h)$, we can check $\mathbb{M}_1(u) := u_1'M_1u_1 + u_2'M_2u_2 - 2u'W \in \mathbf{C}_{\min}(\mathbb{R}^{d_{\underline{\theta}}})$ and $\mathbb{M}_2(v) := C(v) \in \mathbf{C}_{\min}(\mathbb{R})$ separately. First, $\mathbb{M}_1(u) \in \mathbf{C}_{\min}(\mathbb{R}^{d_{\underline{\theta}}})$ 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 2.

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

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

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

Proof of Theorem 2. By the CMT,

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

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 $\xi(\phi)$ is derived in Proposition 2(iii) of Yu (2019). The required result follows by Slutsky's theorem. ■

Proof of Theorem 3. By the CMT,

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

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,

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

which is sum of a scaled χ_1^2 distribution and $\xi(\phi)$, where the χ_1^2 distribution and $\xi(\phi)$ are independent. In the error components model,

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

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

$$
LR_n\left(\gamma_0, \beta_{11}^0\right) \xrightarrow{d} \left[\left(1-\rho_1\right) + \rho_1 \frac{\left(1, -S_{\beta_{11}\theta_{12}} S_{\theta_{12}\theta_{12}}^{-1}\right) \Psi_1 \left(1, -S_{\beta_{11}\theta_{12}} S_{\theta_{12}\theta_{12}}^{-1}\right)'}{\widetilde{S}_{\beta_{11}\beta_{11}}}\right] \chi_1^2 + \xi(\phi).
$$

■

Proof of Theorem 4. By the CMT,

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

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 2, where $\widehat{u} = \arg\min_u \left\{u' S_{\underline{\theta\theta}} u - 2u'W\right\}$. Next,

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

where

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

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 3. ■

Proof of Theorem 5. By the CMT,

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

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

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

where

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

with

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

The rest of the proof is similar to that of Theorem 3. ■

Proof of Theorem 6. First, under H_1^c,

$$\begin{aligned}
&N^{-1/2} \sum_{i=1}^N \sum_{t=1}^T \breve{\mathbf{x}}_{it} \widehat{e}_{it}^o 1(q_{it} \le \gamma) = N^{-1/2} \sum_{i=1}^N \sum_{t=1}^T \breve{\mathbf{x}}_{it} 1(q_{it} \le \gamma) \left(y_{it} - \breve{\mathbf{x}}_{it}' \widehat{\theta}_o\right) \\
&= N^{-1/2} \sum_{i=1}^N \sum_{t=1}^T \breve{\mathbf{x}}_{it} 1(q_{it} \le \gamma) (y_{it} - \breve{\mathbf{x}}_{it}' \theta_o) - N^{-1} \sum_{i=1}^N \sum_{t=1}^T \breve{\mathbf{x}}_{it} \breve{\mathbf{x}}_{it}' 1(q_{it} \le \gamma) \sqrt{N} \left(\widehat{\theta}_o - \theta_o\right) \\
&= N^{-1/2} \sum_{i=1}^N \sum_{t=1}^T \breve{\mathbf{x}}_{it} 1(q_{it} \le \gamma) (y_{it} - \breve{\mathbf{x}}_{it}' \theta_o) - \widehat{M}(\gamma) \widehat{M}^{-1} \left[N^{-1/2} \sum_{i=1}^N \sum_{t=1}^T \breve{\mathbf{x}}_{it} (y_{it} - \breve{\mathbf{x}}_{it}' \theta_o)\right] \\
&= N^{-1/2} \sum_{i=1}^N \sum_{t=1}^T \left[\breve{\mathbf{x}}_{it} 1(q_{it} \le \gamma) - \widehat{M}(\gamma) \widehat{M}^{-1} \breve{\mathbf{x}}_{it}\right] \left[\breve{\mathbf{x}}_{it}' \delta_\theta 1(q_{it} \le \gamma_0) + e_{it}^0\right] \\
&= N^{-1} \sum_{i=1}^N \sum_{t=1}^T \left[\breve{\mathbf{x}}_{it} 1(q_{it} \le \gamma) - \widehat{M}(\gamma) \widehat{M}^{-1} \breve{\mathbf{x}}_{it}\right] \breve{\mathbf{x}}_{it}' 1(q_{it} \le \gamma_0) \left(N^{1/2} \delta_\theta\right) \\
&\quad + N^{-1/2} \sum_{i=1}^N \sum_{t=1}^T \left[\breve{\mathbf{x}}_{it} 1(q_{it} \le \gamma) - \widehat{M}(\gamma) \widehat{M}^{-1} \breve{\mathbf{x}}_{it}\right] e_{it}^0 \\
&\rightsquigarrow \left[M(\gamma \wedge \gamma_0) - M(\gamma) M^{-1} M(\gamma_0)\right] c + \Xi(\gamma),
\end{aligned}$$

where $N^{-1} \sum_{i=1}^N \sum_{t=1}^T \breve{\mathbf{x}}_{it} 1(q_{it} \le \gamma) \breve{\mathbf{x}}_{it}' 1(q_{it} \le \gamma_0) \xrightarrow{p} M(\gamma \wedge \gamma_0) c$ uniformly in $\gamma \in \Gamma$, $\widehat{M}(\gamma) \xrightarrow{p} M(\gamma)$ uniformly in $\gamma \in \Gamma$, $\widehat{M} \xrightarrow{p} M$, and the covariance kernel of $\Xi(\gamma)$ is

$$E\left[\left(\sum_{t=1}^T (1(q_{it} \le \gamma_1) \breve{\mathbf{x}}_{it} - M(\gamma) M^{-1} \breve{\mathbf{x}}_{it}) e_{it}^0\right) \left((1(q_{it} \le \gamma_2) \breve{\mathbf{x}}_{it} - M(\gamma) M^{-1} \breve{\mathbf{x}}_{it}) e_{it}^0\right)'\right].$$

Next, it is standard to show that under H_1^c, $\widehat{\theta}_o$ is consistent to θ_o and

$$\begin{aligned}
\widehat{H}_n(\gamma_1, \gamma_2) &= N^{-1} \sum_{i=1}^N \left[\sum_{t=1}^T \left(\mathbf{x}_{it} 1(q_{it} \le \gamma_1) - \widehat{M}(\gamma_1) \widehat{M}^{-1} \mathbf{x}_{it}\right) \widehat{e}_{it}^o\right] \left[\sum_{t=1}^T \left(\mathbf{x}_{it} 1(q_{it} \le \gamma_1) - \widehat{M}(\gamma_2) \widehat{M}^{-1} \mathbf{x}_{it}\right) \widehat{e}_{it}^o\right]' \\
&\xrightarrow{p} H(\gamma_1, \gamma_2)
\end{aligned}$$

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_1^c, so the results of the theorem follow. ■

Appendix C: Lemmas

This appendix collects lemmas for consistency, convergence rates and local approximation. First, some notations are collected for reference in all lemmas. The letter C is used as a generic positive constant, which need not be the same from line to line. The subscript 0 indicates the true value. 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

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

then

$$S_n\left(\theta\right) = NP_N s\left(w_i|\theta\right)$$

Since

$$\begin{aligned}&s\left(w_i|\theta\right) = \textstyle\sum_{t=1}^{T}\left(\breve{\mathbf{x}}_{it}'\left(\theta_{10}-\theta_1\right)+e_{1it}\right)^2 1(q_{it}\le\gamma\wedge\gamma_0) + \sum_{t=1}^{T}\left(\breve{\mathbf{x}}_{it}'\left(\theta_{20}-\theta_2\right)+e_{2it}\right)^2 1(q_{it}>\gamma\vee\gamma_0)\\ &+\textstyle\sum_{t=1}^{T}\left(\breve{\mathbf{x}}_{it}'\left(\theta_{10}-\theta_2\right)+e_{1it}\right)^2 1(\gamma\wedge\gamma_0<q_{it}\le\gamma_0) + \sum_{t=1}^{T}\left(\breve{\mathbf{x}}_{it}'\left(\theta_{20}-\theta_1\right)+e_{2it}\right)^2 1(\gamma_0<q_{it}\le\gamma\vee\gamma_0),\end{aligned}$$

we have

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

We will concentrate on the SPTR case because the proof for the DPTR is similar.

Lemma 1 *Under Assumption SP,* $\widehat{\gamma}-\gamma_0 = o_p\left(1\right)$ *and* $N^{\kappa}\left(\widehat{\underline{\theta}}-\underline{\theta}\right) = o_p\left(1\right)$.

Proof. This proof is similar to that of Lemma A.5 and A.6 of Hansen (2000), so we need only point out the differences. Define $Y, X, X_\gamma, Z_\gamma, X_\gamma^*, e$ as the matrices stacking $y_{it}, \breve{\mathbf{x}}_{it}, \breve{\mathbf{x}}_{it}d_{it}\left(\gamma\right), \breve{\mathbf{x}}_{it}\left(1-d_{it}\left(\gamma\right)\right), \left(\breve{\mathbf{x}}_{it}, \breve{\mathbf{x}}_{it}d_{it}\left(\gamma\right)\right)$ and e_{it}^0 with $d_{it}\left(\gamma\right) = 1(q_{it}\le\gamma)$. Now, M, $M\left(\gamma\right)$ and $D\left(\gamma\right)$ defined in the main text plays the role of M, $M\left(\gamma\right)$ and $D\left(\gamma\right)f\left(\gamma\right)$ in Lemma A.5. ■

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

Proof. Since δ_N depends on N, we apply the proof idea of Theorem 3.2.5 in Van der Vaart and Wellner (1996) (VW hereafter) to prove this result. Define $d_N\left(\theta,\theta_0\right) = \left\|\underline{\theta}-\underline{\theta}_0\right\| + \left\|\delta_N\right\|\sqrt{\left\|\gamma-\gamma_0\right\|}$ for θ in a neighborhood of θ_0, and

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

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

$$P\left(d_N\left(\widehat{\theta}, \theta_0\right) > 2^J\right) \leq \sum_{j \geq J, \left\|\underline{\theta}-\underline{\theta}_0\right\| < M\|\delta_N\|, \|\gamma-\gamma_0\| < \eta} P\left(\inf_{\theta \in S_{j,N}} Q_n(\theta) \leq 0\right) + P\left(2\left\|\underline{\theta} - \underline{\theta}_0\right\| \geq M\left\|\delta_N\right\|, 2\left\|\gamma - \gamma_0\right\| \geq \eta\right), \tag{16}$$

where M and η are small positive numbers. The second term on the right hand side of (16) converges to zero as $N \to \infty$ for every $\eta > 0$ and $M > 0$ by Lemma 1, so we can concentrate on the first term.

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

where the last equality is from Markov's inequality.

From the form of $s(w_i|\theta) - s(w_i|\theta_0)$, it is not hard to see that

$$\begin{aligned} & \inf_{\theta \in S_{j,N}} \left|\mathbb{E}\left[Q_n(\theta)\right]\right| = \inf_{\theta \in S_{j,N}} \left|\sum\nolimits_{k=1}^{4} \mathbb{E}\left[T_k(\theta)\right]\right| \\ & = \inf_{\theta \in S_{j,N}} C\left|\tfrac{N}{\lambda_N}\left\|\underline{\theta} - \underline{\theta}_0\right\|^2 + \tfrac{N}{\lambda_N}\left[\left\|\theta_{10} - \theta_2\right\|^2 + \left\|\theta_{20} - \theta_1\right\|^2\right]\left\|\gamma - \gamma_0\right\|\right| \\ & = \inf_{\theta \in S_{j,N}} C\left|\tfrac{N}{\lambda_N}\left\|\underline{\theta} - \underline{\theta}_0\right\|^2 + \tfrac{N}{\lambda_N}\left\|\delta_N\right\|^2\left\|\gamma - \gamma_0\right\|\right| = \inf_{\theta \in S_{j,N}} C\tfrac{N}{\lambda_N} d_N(\theta, \theta_0)^2 \geq C\tfrac{2^{2j-2}}{\lambda_N} = C\tfrac{2^{2j}}{\lambda_N}, \end{aligned}$$

where the third equality is 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 $\|\theta_{20} - \theta_1\| = O(\|\delta_N\|)$. To bound $\sum_{k=1}^{4} \mathbb{E}\left[\sup_{\theta \in S_{j,N}} \left|T_k(\theta) - \mathbb{E}\left[T_k(\theta)\right]\right|\right]$, 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

$$\begin{aligned} \mathcal{F} = & \sum\nolimits_{t=1}^{T}\left(\left\|\theta_{10} - \theta_1\right\|^2\left\|\check{\mathbf{x}}_{it}\right\|^2 + 2\left\|\theta_{10} - \theta_1\right\|\left\|\check{\mathbf{x}}_{it} e_{1it}\right\|\right) 1\left(q_{it} \leq \gamma_0\right) \\ & + \sum\nolimits_{t=1}^{T}\left(\left\|\theta_{20} - \theta_2\right\|^2\left\|\check{\mathbf{x}}_{it}\right\|^2 + 2\left\|\theta_{20} - \theta_2\right\|\left\|\check{\mathbf{x}}_{it} e_{2it}\right\|\right) 1\left(q_{it} > \gamma_0\right) \\ & + \sum\nolimits_{t=1}^{T}\left(\left\|\theta_{10} - \theta_2\right\|^2\left\|\check{\mathbf{x}}_{it}\right\|^2 + 2\left\|\theta_{10} - \theta_2\right\|\left\|\check{\mathbf{x}}_{it} e_{1it}\right\|\right) 1\left(\gamma \wedge \gamma_0 < q_{it} \leq \gamma_0\right) \\ & + \sum\nolimits_{t=1}^{T}\left(\left\|\theta_{20} - \theta_1\right\|^2\left\|\check{\mathbf{x}}_{it}\right\|^2 + 2\left\|\theta_{20} - \theta_1\right\|\left\|\check{\mathbf{x}}_{it} e_{2it}\right\|\right) 1\left(\gamma_0 < q_{it} \leq \gamma \vee \gamma_0\right) \\ =: & F_1\left(\check{\mathbf{x}}_i, e_{1i}|\theta_1\right) + F_2\left(\check{\mathbf{x}}_i, e_{2i}|\theta_2\right) + F_3\left(\check{\mathbf{x}}_i, e_{1i}|\theta_2, \gamma\right) + F_4\left(\check{\mathbf{x}}_i, e_{2i}|\theta_1, \gamma\right)., \end{aligned}$$

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 (16), we have

$$\begin{aligned} & \sum_{k=1}^{2} \mathbb{E}\left[\sup_{\theta \in S_{j,N}} \left|T_k(\theta) - \mathbb{E}\left[T_k(\theta)\right]\right|\right] \leq C\frac{\sqrt{E\left[\sup_{\theta \in S_{j,N}} F_1(\check{\mathbf{x}}_i, e_{1i}|\theta_1)^2\right] + E\left[\sup_{\theta \in S_{j,N}} F_2(\check{\mathbf{x}}_i, e_{1i}|\theta_2)^2\right]}}{\sqrt{N}\|\delta_N\|^2} = C\frac{\sup_{\theta \in S_{j,N}}\left\|\underline{\theta} - \underline{\theta}_0\right\|}{\sqrt{N}\|\delta_N\|^2}, \\ & \mathbb{E}\left[\sup_{\theta \in S_{j,N}} \left|T_3(\theta) - \mathbb{E}\left[T_3(\theta)\right]\right|\right] \leq C\frac{\sqrt{E\left[\sup_{\theta \in S_{j,N}} F_3(\check{\mathbf{x}}_i, e_{1i}|\theta_2, \gamma)^2\right]}}{\sqrt{N}\|\delta_N\|^2} \leq C\frac{\sup_{\theta \in S_{j,N}} \sqrt{\|\theta_{10} - \theta_2\|^2}\sqrt{|\gamma - \gamma_0|}}{\sqrt{N}\|\delta_N\|^2} = \frac{\sup_{\theta \in S_{j,N}} \|\delta_N\|\sqrt{|\gamma - \gamma_0|}}{\sqrt{N}\|\delta_N\|^2}. \end{aligned}$$

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

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

In summary,

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

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

$$S_n\left(\gamma_0 + \frac{v}{\lambda_N}, \underline{\theta}_0 + \frac{u}{N^{1/2}}\right) - S_n(\gamma_0, \beta_0) = u_1' M_1 u_1 + u_2' M_2 u_2 - 2W_n(u) + C_n(v) + o_P(1),$$

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

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

and

$$\begin{aligned} C_n(v) &= S_n\left(\gamma_0 + \frac{v}{\lambda_N}, \beta_0\right) - S_n(\gamma_0, \beta_0) \\ &= \sum_{i=1}^{N} \sum_{t=1}^{T} z_{1it} 1\left(\gamma_0 + \frac{v}{\lambda_N} < q_{it} \leq \gamma_0\right) - \sum_{i=1}^{N} \sum_{t=1}^{T} z_{2it} 1\left(\gamma_0 < q_{it} \leq \gamma_0 + \frac{v}{\lambda_N}\right) \end{aligned}$$

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

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

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

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,

$$\begin{aligned} T_1(u_1) &= \sum_{i=1}^{N}\sum_{t=1}^{T}\left[\tfrac{u_1'}{N^{1/2}}\breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{it}'\tfrac{u_1}{N^{1/2}} - 2\tfrac{u_1'}{N^{1/2}}\breve{\mathbf{x}}_{it}e_{1it}\right]1(q_{it}\le\gamma_0) \\ &= u_1'M_1u_1 - \tfrac{2u_1'}{N^{1/2}}\sum_{i=1}^{N}\sum_{t=1}^{T}\breve{\mathbf{x}}_{it}e_{1it}1(q_{it}\le\gamma_0) + o_p(1), \end{aligned}$$

where $o_p(1)$ is from the LLN. By a similar analysis, when $v<0$,

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

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

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

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

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

and

$$\sqrt{N}\mathbb{E}\left[\sum_{t=1}^{T}\breve{\mathbf{x}}_{it}e_{1it}1(\gamma_0+\frac{v}{\lambda_N}<q_{it}\le\gamma_0)\right] = \frac{|v|}{\lambda_N}\sum_{t=1}^{T}\mathbb{E}\left[\breve{\mathbf{x}}_{it}e_{1it}|q_{it}=\overline{\gamma}\right]f_t(\overline{\gamma}) = 0.$$

Finally, when $v<0$,

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

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

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

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

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

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

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

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

Lemma 4 *Under Assumption SP,*

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

where $W\left(u\right)=u'W$ with W defined in (15) and $C\left(v\right)$ is defined in (14). Furthermore, W and $C\left(v\right)$ are independent of each other.

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

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

Because

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

we have

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

where

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

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 $v\in[0,\bar{v}]$, $\sum_{i=1}^{N}\sum_{t=1}^{T}\delta_N'\breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{it}'\delta_N 1(\gamma_0<q_{it}\le\gamma_0+\frac{v}{\lambda_N})\xrightarrow{p}c'Dc\cdot v$.

We next analyze the random parts in $W_n(u)$ and $C_n(v)$ by applying Theorem 2.11.22 of VW. First, for $v_1<0$ and $v_2>0$, define

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

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

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

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

$$\begin{aligned}
&\lambda_N E\left[c'\breve{\mathbf{x}}_{it}e_{1it}1(\gamma_0+\tfrac{v_1}{\lambda_N}<q_{it}\le\gamma_0)c'\breve{\mathbf{x}}_{i\tau}e_{2i\tau}1(\gamma_0+\tfrac{v_1}{\lambda_N}<q_{i\tau}\le\gamma_0)\right]\\
&=\lambda_N E\left[c'\breve{\mathbf{x}}_{it}\breve{\mathbf{x}}_{i\tau}'ce_{1it}e_{1i\tau}1(\gamma_0+\tfrac{v_1}{\lambda_N}<q_{it}\le\gamma_0)1(\gamma_0+\tfrac{v_1}{\lambda_N}<q_{i\tau}\le\gamma_0)\right]\\
&\le\|c\|^2\left(E\left[\|\breve{\mathbf{x}}_{it}\|^4\right]E\left[\|\breve{\mathbf{x}}_{i\tau}\|^4\right]E\left[|e_{1it}|^4\right]E\left[|e_{1i\tau}|^4\right]\right)^{1/4}\\
&\quad\cdot\lambda_N P\left(\gamma_0+\tfrac{v_1}{\lambda_N}<q_{it}\le\gamma_0,\gamma_0+\tfrac{v_1}{\lambda_N}<q_{i\tau}\le\gamma_0\right)\to 0
\end{aligned}\tag{17}$$

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

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

Also,

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

where the first term is zero and the second term converges to zero by similar arguments as in (17). Finally,

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

by a CLT, and

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

similarly, $E\left[\sum_{i=1}^{N} S_{2i} \sum_{i=1}^{N} S_{3i}\right] \to 0$. 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 the stochastic equicontinuity of $C_n\left(v\right)$ since the stochastic equicontinuity of of $W_n\left(u\right)$ is obvious. Note that

$$C_n\left(v\right) = \mathbb{G}_N\left(T_{3N}\left(v\right)\right)$$

where $T_{3N}\left(v\right) = \sqrt{N} \sum_{t=1}^{T} \delta_N' \check{\mathbf{x}}_{it} e_{1it} 1\left(\gamma_0 + \frac{v}{\lambda_N} < q_{it} \leq \gamma_0\right)$. Since $\{T_{3N}(v) : -\infty < -\overline{v} \leq v \leq 0\}$ is VC-subgraph for each N and the VC-index bounded by some constant independent of N (see, e.g., Example 2.11.24 of VW), the uniform-entropy condition holds. It remains to show condition (2.11.21):

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

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

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

(i)

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

where the cross term is $o(1)$ by (17), and $\mathcal{N}$ is a neighborhood of γ_0. (ii)

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

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

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

■

Appendix D: Details of Calculation in Section 2

When $\gamma < \gamma_0$,

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

Simiarly, when $\gamma > \gamma_0$,

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

and

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

whose probability limit is

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

where $\widetilde{1}_{it}(\gamma_0,\gamma) = 1(\gamma_0 < q_{it} \le \gamma) - \frac{\sum_{\tau=1}^{T}1(\gamma_0 < q_{i\tau} \le \gamma)}{\sum_{\tau=1}^{T}1(q_{i\tau} \le \gamma)}$, and $\widetilde{u}_{it}(\gamma_0,\gamma) = u_{it}1(\gamma_0 < q_{it} \le \gamma) - \frac{\sum_{\tau=1}^{T}u_{i\tau}1(\gamma_0 < q_{i\tau} \le \gamma)}{\sum_{\tau=1}^{T}1(q_{i\tau} \le \gamma)}$. Note that in the objective function, we actually condition on $D_i^{-}(\gamma) \ge 1$ and $D_i^{+}(\gamma) \ge 1$ in the summation $\sum_{t=1}^{T}\widetilde{y}_{it}^{-}(\gamma)^2\,1(q_{it} \le \gamma)$ and $\sum_{t=1}^{T}\widetilde{y}_{it}^{+}(\gamma)^2\,1(q_{it} > \gamma)$, where $D_i^{-}(\gamma) = \sum_{t=1}^{T}1(q_{it} \le \gamma)$ and $D_i^{+}(\gamma) = \sum_{t=1}^{T}1(q_{it} > \gamma)$. So we first condition on $D_i^{-}(\gamma) \ge 1$ and $D_i^{+}(\gamma) \ge 1$ in calculating the expectations, and then multiply $P\left(D_i^{-}(\gamma) \ge 1\right) = 1 - \overline{F}(\gamma)^T = 1 - (1-\gamma)^T =: p^{-}(\gamma)$ and $P\left(D_i^{+}(\gamma) \ge 1\right) = 1 - F(\gamma)^T = 1 - \gamma^T =: p^{+}(\gamma)$, respectively.

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

and similarly,

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

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

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

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

and

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

where

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

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

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

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

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

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 $\sum_{t=1}^{T}E\left[\widetilde{1}_{it}(\gamma,\gamma_0)^2 1(q_{it}>\gamma)|D_i^+(\gamma)\geq 1\right]=\sum_{t=1}^{T}E\left[\widetilde{u}_{it}(\gamma,\gamma_0)\widetilde{u}_{it}^+(\gamma)1(q_{it}>\gamma)|D_i^+(\gamma)\geq 1\right]$, we can collect terms by noting that $\delta_\sigma^2+2\sigma_{20}\delta_\sigma=1-\sigma_{20}^2$. Similarly, when $\gamma>\gamma_0$,

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

and

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

where

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

We next report the counterparts of Figure 1 when $T=2$ or $\gamma_0=0.5$. Figure 4 shows the case with $T=2$ and Figure 5 shows the case with $\gamma_0=0.5$. Obviously, the conclusions in the setting of the main text still apply here.

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

$$\begin{aligned}
S(\gamma)&\to\begin{cases}\gamma+\sigma_{20}^2(1-\gamma)+\delta_\alpha^2\left[(\gamma_0-\gamma)-\frac{(\gamma_0-\gamma)^2}{1-\gamma}\right]+\left(1-\sigma_{20}^2\right)(\gamma_0-\gamma), & \text{if } \gamma\leq\gamma_0,\\ \gamma+\sigma_{20}^2(1-\gamma)+\delta_\alpha^2\left[(\gamma-\gamma_0)-\frac{(\gamma-\gamma_0)^2}{\gamma}\right]+\left(\sigma_{20}^2-1\right)(\gamma-\gamma_0), & \text{if } \gamma>\gamma_0,\end{cases}\\
&=C+\begin{cases}\delta_\alpha^2\left[(\gamma_0-\gamma)-\frac{(\gamma_0-\gamma)^2}{1-\gamma}\right], & \text{if } \gamma\leq\gamma_0,\\ \delta_\alpha^2\left[(\gamma-\gamma_0)-\frac{(\gamma-\gamma_0)^2}{\gamma}\right], & \text{if } \gamma>\gamma_0,\end{cases}
\end{aligned}$$

where the constant $C=1+\left(1-\sigma_{20}^2\right)(\gamma_0-1)$. As argued in the main text, $\arg\min_\gamma T_2(\gamma)=\gamma_0$, so $\arg\min_\gamma S(\gamma)=\gamma_0$.

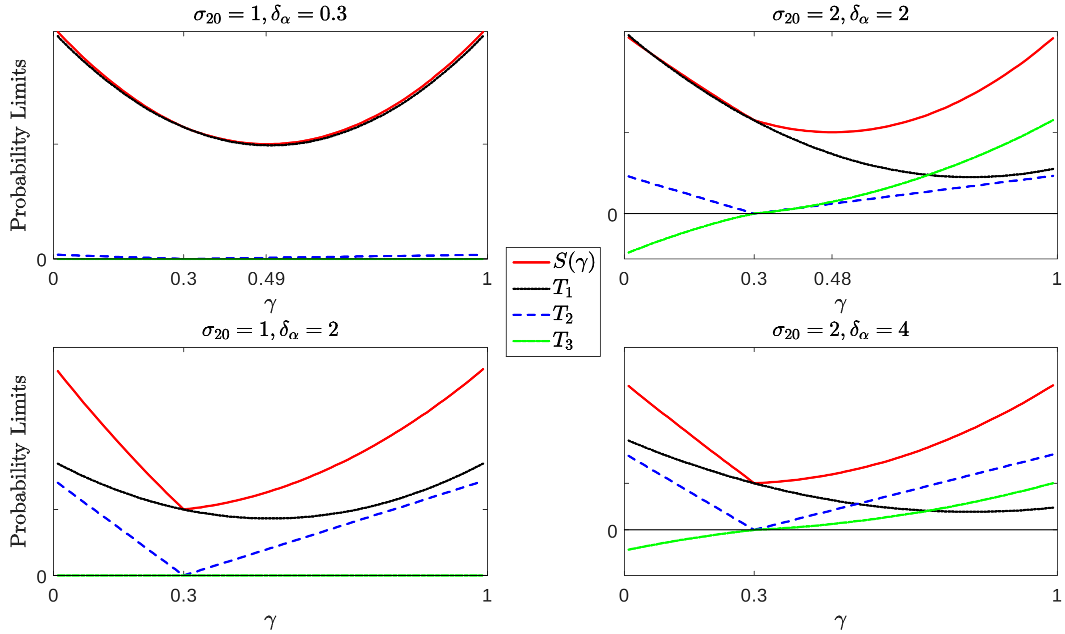

Figure 4: $S(\gamma)$ and Its Three Components for Various σ_{20} and δ_α Values: $\sigma_{10} = 1, \gamma_0 = 0.3$, $T = 2$

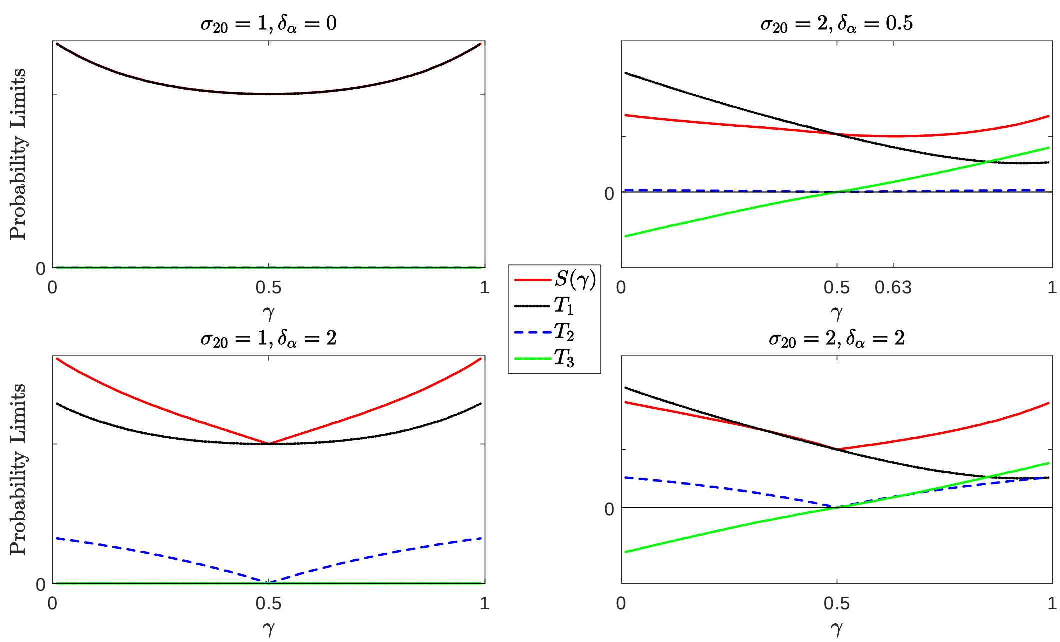

Figure 5: $S(\gamma)$ and Its Three Components for Various σ_{20} and δ_α Values: $\sigma_{10} = 1, \gamma_0 = 0.5$, $T = 5$

Appendix E: Details of CI Construction for β and γ

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

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

where $\mathbf{M}^2_{\leq\gamma}$ is the annihilator of $\left\{\check{\mathbf{x}}^{-1}_{i\tau}1(q_{i\tau} \leq \gamma)\right\}_{i,\tau}$, $\mathbf{M}_{\leq\gamma}$ and $\mathbf{M}_{>\gamma}$ are annihilators of $\{\check{\mathbf{x}}_{i\tau}1(q_{i\tau} \leq \gamma)\}_{i,\tau}$ and $\{\check{\mathbf{x}}_{i\tau}1(q_{i\tau} > \gamma)\}_{i,\tau}$, $\mathbf{y}_{\leq\gamma}$ and $\mathbf{y}_{>\gamma}$ are vectors stacking $\{y_{i\tau}1(q_{i\tau} \leq \gamma)\}_{i,\tau}$ and $\{y_{i\tau}1(q_{i\tau} > \gamma)\}_{i,\tau}$, $\mathbf{X}^1_{\leq\gamma}$ 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 $\check{\mathbf{x}}^{-1}_{i\tau}$; we denote it as $S_{1n}(q_{it})$, so the constant term of $S_n(q_{it}, \beta_{11}) - S_n\left(\widehat{\gamma}, \widehat{\beta}_{11}\right)$ is $S_{1n}(q_{it}) - S_n(\widehat{\gamma})$. As a result,

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

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 $D(q_{it}, \alpha) := \left\langle \mathbf{X}^1_{\leq q_{it}}, \mathbf{y}_{\leq q_{it}}\right\rangle^2 - \left\langle \mathbf{X}^1_{\leq q_{it}}, \mathbf{X}^1_{\leq q_{it}}\right\rangle \left(S_{1n}(q_{it}) - S_n(\widehat{\gamma}) - \widehat{c}_\alpha \widehat{\eta}^2\right)$.

When there are variables without threshold effects, we need only adjust the procedure above a little bit. For β_{11},

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

where $\mathbf{M}^2$ is the annihilator of $\left\{\left(\begin{array}{ccc} \check{\mathbf{x}}'_{1i\tau}, & \check{\mathbf{x}}'_{2i\tau}1(q_{i\tau} \leq q_{it}), & \check{\mathbf{x}}'_{2i\tau}1(q_{i\tau} > q_{it}) \end{array}\right)\right\}_{i,\tau}$ 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,

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

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

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

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

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

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

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

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

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

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

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

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

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

depends only on q_{it}. Because $\widetilde{\gamma}$ depends on β_{11} in a nonlinear way, it is hard to see whether $\{\beta_{11}|LR_{1n}(q_{it},\beta_{11})\le\widehat{c}_\alpha\}$ is an interval or not. As a result, we must construct $\{\beta_{11}|LR_{1n}(q_{it},\beta_{11})\le\widehat{c}_\alpha\}$ by grid search.

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

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

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

$S_n\left(q_{it},\beta_{121},\widehat{\underline{\theta}}_{-121}(q_{it},\beta_{121})\right)$ is the same as in $LR_n(\gamma,\beta_{121})$, and

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

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

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

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

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

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

$$\bigcup_{q_{it}\in\Gamma}\{\beta_{11}|LR_{2n}(q_{it},\beta_{11})\le\widehat{c}_\alpha\}.$$

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

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

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

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

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

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

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

which is quadratic in β_{11}, so

$$\begin{aligned}
& \left\{\beta_{11}|LR_{2n}\left(q_{it},\beta_{11}\right)\leq\widehat{c}_\alpha\right\} \\
=& \begin{cases}\left[\frac{\left\langle\mathbf{X}^1,\mathbf{y}-\left(\check{\mathbf{X}}^1,\check{\mathbf{X}}^2_{\leq q_{it}},\check{\mathbf{X}}^2_{>q_{it}}\right)\underline{\widehat{\theta}}_{-11}\right\rangle-\sqrt{D(q_{it},\alpha)}}{\left\langle\mathbf{X}^1,\mathbf{X}^1\right\rangle},\frac{\left\langle\mathbf{X}^1,\mathbf{y}-\left(\check{\mathbf{X}}^1,\check{\mathbf{X}}^2_{\leq q_{it}},\check{\mathbf{X}}^2_{>q_{it}}\right)\underline{\widehat{\theta}}_{-11}\right\rangle+\sqrt{D(q_{it},\alpha)}}{\left\langle\mathbf{X}^1,\mathbf{X}^1\right\rangle}\right] & \text{if } D\left(q_{it},\alpha\right)\geq0, \\ \emptyset, & \text{otherwise,}\end{cases}
\end{aligned}$$

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

Appendix F: Details of Calculation in Section 6.1

First, write

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

then $M_{11}=M_{11}^-+M_{11}^+$. Second, write

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

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

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

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

and

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

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

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

Appendix G: Simulations

In this appendix, we will examine the performances of our estimation and inference methods in finite samples with either $\alpha_{1i}\neq\alpha_{2i}$ or $\alpha_{1i}=\alpha_{2i}$.Following Yu et al. (2018b), we use the mean absolute deviation (MAD) to measure the risk for γ, and use the usual root-mean-square error (RMSE) for regular parameters. For inferences on γ, we will compare the CIs based on $LR_n(\gamma)$, $LR_{1n}(\gamma)$ and $LR_{2n}(\gamma)$; for inferences on β, we will 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 performances of the two hypothesis tests in Section 5 in similar scenarios are available from the literature, we will not check their performances here. In all simulations, we consider 1000 replications, $T = 2, 5, 10$, and $N = 100, 250, 500$.

We consider the following data generating process (DGP):

$$y_{it} = (q_{it}\beta_1 + \alpha_{1i})\, 1(q_{it} \leq \gamma) + (q_{it}\beta_2 + \alpha_{2i})\, 1(q_{it} > \gamma) + u_{it}, \tag{18}$$

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, $q_{it} \sim N(0,1)$, $a_i \sim N(0,1)$, 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 Section 3.2, but we do not employ $a_{1i} = a_{2i}$ or $\sigma_1 = \sigma_2$ for further simplification.

Two general results apply to all cases. First, larger T 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 require longer CIs, and also match the convergence rates of $\widehat{\gamma}$ and $\widehat{\beta}_1$. 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 DGP in Table 3 contains 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. Anyway, the performance 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 suggested 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 under-covers occasionally, while both LR_n-CI and LR_{2n}-CI over-cover. For small threshold effects and NT, the overcoverage of LR_{2n}-CI is less severe, while for large threshold effects or NT, the overcoverage of LR_n-CI is less severe.

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

		Estimation		CI: length			CI: coverage prob.		
T	N	Bias	MAD	LR_n	LR_{1n}	LR_{2n}	LR_n	LR_{1n}	LR_{2n}
				$\beta_1 = 0.2$					
2	250	0.0066	0.1785	0.6005	0.6761	0.5591	0.956	0.981	0.905
	500	0.0016	0.1482	0.4905	0.5052	0.4725	0.973	0.978	0.966
	1000	0.0018	0.1157	0.3858	0.3968	0.3755	0.966	0.977	0.962
5	250	-0.0032	0.1424	0.4479	0.4613	0.4336	0.967	0.972	0.958
	500	-0.0141	0.1160	0.3606	0.3674	0.3505	0.970	0.981	0.963
	1000	-0.0052	0.0933	0.2813	0.2861	0.2735	0.952	0.964	0.948
10	250	0.0088	0.115	0.3612	0.3723	0.3503	0.966	0.983	0.968
	500	0.0049	0.0873	0.2875	0.2938	0.2805	0.980	0.987	0.979
	1000	-0.0074	0.0714	0.2216	0.2253	0.2186	0.969	0.971	0.965
				$\beta_1 = 0.5$					
2	250	0.0017	0.1282	0.4117	0.4265	0.3981	0.955	0.969	0.934
	500	-0.0068	0.1026	0.3391	0.3472	0.3316	0.981	0.985	0.981
	1000	-0.0026	0.0791	0.2630	0.2679	0.2579	0.972	0.976	0.965
5	250	-0.0071	0.0979	0.3062	0.3121	0.3009	0.984	0.985	0.976
	500	-0.0051	0.0750	0.2486	0.2514	0.2442	0.975	0.973	0.975
	1000	0.0019	0.0613	0.1958	0.1982	0.1908	0.968	0.970	0.953
10	250	0.0057	0.0971	0.4148	0.4977	0.3513	0.956	0.963	0.931
	500	-0.0014	0.0470	0.2302	0.2729	0.2029	0.969	0.972	0.956
	1000	0.0009	0.0214	0.1154	0.1270	0.1072	0.945	0.956	0.943
				$\beta_1 = 1.0$					
2	250	0.0011	0.0877	0.2873	0.2954	0.2803	0.958	0.967	0.939
	500	-0.0004	0.0683	0.2320	0.2350	0.2281	0.989	0.985	0.978
	1000	-0.0003	0.0548	0.1872	0.1880	0.1848	0.988	0.990	0.984
5	250	0.0009	0.0643	0.2178	0.2210	0.2142	0.990	0.994	0.986
	500	-0.0043	0.0529	0.1707	0.1721	0.1678	0.985	0.984	0.976
	1000	0.0017	0.0411	0.1348	0.1357	0.1336	0.970	0.968	0.965
10	250	0.0038	0.0490	0.1664	0.1681	0.1652	0.980	0.982	0.980
	500	0.0005	0.0397	0.1342	0.1359	0.1332	0.977	0.990	0.976
	1000	-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 ($\alpha_{1i} = \alpha_{2i}$, static)

T	N	Estimation		CI: length			CI: coverage prob.		
		Bias	RMSE	t	LR_n	LR_{2n}	t	LR_n	LR_{2n}
					$\beta_1 = 0.2$				
2	250	0.0002	0.0210	0.0724	0.1235	0.0942	0.933	0.994	0.970
	500	-0.0004	0.0146	0.0514	0.0865	0.0668	0.928	1.000	0.969
	1000	0.0011	0.0094	0.0363	0.0608	0.0471	0.952	0.990	0.975
5	250	0.0008	0.0118	0.0472	0.0650	0.0703	0.945	1.000	1.000
	500	0.0003	0.0085	0.0334	0.0459	0.0500	0.955	0.983	0.990
	1000	0.0007	0.0062	0.0237	0.0324	0.0353	0.958	0.991	0.995
10	250	0.0005	0.0107	0.0427	0.0505	0.0630	0.976	0.981	1.000
	500	-0.0005	0.0077	0.0302	0.0356	0.0443	0.943	0.990	0.998
	1000	0.0003	0.0052	0.0214	0.0251	0.0316	0.960	0.993	0.995
					$\beta_1 = 0.5$				
2	250	-0.0003	0.0206	0.0721	0.1215	0.0940	0.946	0.997	0.970
	500	-0.0006	0.0145	0.0515	0.0856	0.0668	0.928	1.000	0.975
	1000	0.0010	0.0092	0.0363	0.0602	0.0471	0.950	0.988	0.981
5	250	0.0007	0.0117	0.0474	0.0648	0.0705	0.945	0.993	1.000
	500	0.0003	0.0084	0.0335	0.0457	0.0500	0.956	0.984	0.993
	1000	0.0006	0.0062	0.0237	0.0323	0.353	0.957	0.990	0.996
10	250	0.0005	0.0107	0.0428	0.0504	0.0631	0.970	0.979	0.996
	500	-0.0005	0.0078	0.0302	0.0356	0.0444	0.941	0.990	1.000
	1000	0.0002	0.0052	0.0214	0.0252	0.0316	0.958	0.989	0.999
					$\beta_1 = 1.0$				
2	250	-0.0004	0.0200	0.0721	0.1202	0.0939	0.936	0.998	0.975
	500	-0.0006	0.0142	0.0513	0.0849	0.0666	0.935	1.000	0.975
	1000	0.0009	0.0091	0.0362	0.0598	0.0470	0.955	0.988	0.983
5	250	0.0004	0.0115	0.0474	0.0646	0.0704	0.938	0.994	1.000
	500	0.0003	0.0084	0.0335	0.0457	0.0500	0.959	0.985	0.992
	1000	0.0006	0.0062	0.0237	0.0323	0.0353	0.957	0.990	1.000
10	250	0.0004	0.0107	0.0429	0.0504	0.0632	0.973	0.982	0.980
	500	-0.0005	0.0078	0.0303	0.0356	0.0445	0.941	0.995	0.996
	1000	-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 ($\alpha_{1i} \neq \alpha_{2i}$, static)

T	N	Estimation		CI: length			CI: coverage prob.		
		Bias	MAD	LR_n	LR_{1n}	LR_{2n}	LR_n	LR_{1n}	LR_{2n}
		$\Delta = 0.2$							
2	250	0.0476	0.3001	1.2512	2.8564	0.9020	0.946	1.000	0.837
	500	0.0052	0.2287	0.8471	1.7521	0.6386	0.965	0.992	0.881
	1000	0.0056	0.1322	0.5606	0.8500	0.4408	0.961	0.976	0.910
5	250	-0.0182	0.2325	0.9863	2.0291	0.7540	0.973	0.990	0.895
	500	-0.0273	0.1336	0.6525	1.0567	0.5020	0.982	1.000	0.933
	1000	0.0011	0.0658	0.3864	0.5049	0.3241	0.961	0.968	0.952
10	250	-0.0175	0.2447	0.8502	1.5980	0.6353	0.923	0.983	0.880
	500	0.0017	0.1366	0.5752	0.7633	0.4706	0.937	0.986	0.927
	1000	0.0073	0.0641	0.3359	0.4031	0.2805	0.981	0.988	0.953
		$\Delta = 0.5$							
2	250	0.0268	0.1507	0.6634	1.1759	0.5166	0.953	0.982	0.895
	500	0.0006	0.0809	0.4061	0.5369	0.3469	0.957	0.979	0.956
	1000	-0.0101	0.0417	0.2169	0.2538	0.1882	0.951	0.958	0.939
5	250	0.0018	0.0987	0.4843	0.6883	0.3914	0.967	0.992	0.925
	500	-0.0024	0.0475	0.2620	0.3229	0.2275	0.985	0.986	0.963
	1000	-0.0031	0.0259	0.1350	0.1484	0.1239	0.946	0.955	0.941
10	250	0.0057	0.0971	0.4148	0.4977	0.3513	0.956	0.963	0.931
	500	-0.0014	0.0470	0.2302	0.2729	0.2029	0.969	0.972	0.956
	1000	0.0009	0.0214	0.1154	0.1270	0.1072	0.945	0.956	0.943
		$\Delta = 1.0$							
2	250	0.0063	0.0547	0.2850	0.3426	0.2528	0.921	0.930	0.905
	500	0.0004	0.0314	0.1546	0.1670	0.1411	0.958	0.964	0.957
	1000	-0.0040	0.0187	0.0781	0.0839	0.0750	0.955	0.961	0.952
5	250	0.0039	0.0432	0.1876	0.2105	0.1702	0.967	0.978	0.938
	500	-0.0005	0.0164	0.0947	0.1015	0.0876	0.965	0.979	0.960
	1000	-0.0008	0.0102	0.0440	0.0455	0.0423	0.930	0.926	0.925
10	250	0.0038	0.0350	0.1710	0.1949	0.1554	0.963	0.980	0.961
	500	0.0022	0.0149	0.0854	0.0889	0.0779	0.953	0.962	0.939
	1000	-0.0011	0.0078	0.0376	0.0383	0.0364	0.910	0.913	0.910

Note: The confidence level is targeted at 0.95.

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

		Estimation		CI: length			CI: coverage prob.		
T	N	Bias	RMSE	t	LR_n	LR_{2n}	t	LR_n	LR_{2n}
		$\Delta = 0.2$							
2	250	-0.0014	0.1210	0.3687	0.6801	0.4748	0.896	0.991	0.958
	500	-0.0147	0.0783	0.2576	0.4432	0.3353	0.913	1.000	0.963
	1000	-0.0028	0.0473	0.1840	0.3052	0.2357	0.950	0.998	0.989
5	250	0.0103	0.0609	0.2307	0.3351	0.3538	0.938	0.996	1.000
	500	-0.0033	0.0422	0.1667	0.2311	0.2498	0.941	0.993	0.997
	1000	0.0009	0.0298	0.1180	0.1622	0.1763	0.952	0.995	0.990
10	250	-0.0041	0.0546	0.2063	0.2530	0.3111	0.945	0.973	0.995
	500	-0.0021	0.0370	0.1495	0.1777	0.2227	0.938	0.982	1.000
	1000	0.0024	0.0261	0.1063	0.1258	0.1573	0.940	0.980	0.993
		$\Delta = 0.5$							
2	250	-0.0049	0.1039	0.3623	0.6123	0.4687	0.922	0.995	0.972
	500	-0.0137	0.0707	0.2558	0.4266	0.3325	0.938	1.000	0.975
	1000	-0.0025	0.0442	0.1818	0.2991	0.2355	0.962	0.997	0.993
5	250	0.0081	0.0587	0.2339	0.3254	0.3532	0.943	0.994	1.000
	500	-0.0034	0.0421	0.1671	0.2277	0.2488	0.946	0.995	0.996
	1000	-0.0008	0.0302	0.1184	0.1615	0.1770	0.954	0.992	0.995
10	250	-0.0070	0.0571	0.2114	0.2509	0.3129	0.962	0.967	0.994
	500	-0.0023	0.0372	0.1515	0.1778	0.2243	0.957	0.985	0.995
	1000	0.0026	0.0262	0.1067	0.1258	0.1578	0.946	0.980	0.990
		$\Delta = 1.0$							
2	250	-0.0060	0.0956	0.3604	0.5947	0.4678	0.930	0.995	0.987
	500	-0.0104	0.0689	0.2557	0.4206	0.3325	0.937	1.000	0.985
	1000	0.0005	0.0441	0.1816	0.2971	0.2351	0.961	1.000	0.991
5	250	0.0093	0.0593	0.2354	0.3237	0.3537	0.962	0.986	1.000
	500	-0.0025	0.0418	0.1673	0.2272	0.2490	0.957	0.995	0.996
	1000	-0.0003	0.0301	0.1185	0.1613	0.1768	0.945	0.993	0.995
10	250	-0.0070	0.0574	0.2124	0.2510	0.3135	0.961	0.975	0.990
	500	-0.0019	0.0373	0.1516	0.1777	0.2241	0.952	0.983	0.990
	1000	0.0028	0.0263	0.1068	0.1258	0.1578	0.946	0.980	0.992

Note: The confidence level is targeted at 0.95.