Panel threshold model with covariate-dependent thresholds and unobserved individual-specific threshold effects

January 18, 2024

Lixiong Yang†
School of Management, Lanzhou University, Lanzhou, China.
E-mail address: ylx@lzu.edu.cn

I-Po Chen
Institute of Economics, National Sun Yat-sen University, Kaohsiung, Taiwan.
E-mail address: j63007@gmail.com

Chingnun Lee
Institute of Economics, National Sun Yat-sen University, Kaohsiung, Taiwan.
E-mail address: lee_econ@mail.nsysu.edu.tw

Mingjian Ren
School of Management, Lanzhou University, Lanzhou, China.
E-mail address: renmj22@lzu.edu.cn.

†Corresponding author: Lixiong Yang. School of Management, Lanzhou University, 222 South Tianshui Road, Lanzhou 730000, China. E-mail address: ylx@lzu.edu.cn. Phone: +86 13669327501

Panel threshold model with covariate-dependent thresholds and unobserved individual-specific threshold effects

Abstract

This article introduces a panel threshold model with covariate-dependent and time-varying thresholds and unobserved individual-specific threshold effects (PTCDI). We develop methods for estimation and inference for threshold parameters in the proposed PTCDI model by employing the correlated random effects (CRE) device. We also suggest test statistics for linearity, threshold constancy, unobserved individual-specific threshold effects, and for determining the number of thresholds. We derive the asymptotic properties of the proposed estimator in the small-threshold-effect framework, and establish the limiting distributions of the suggested test statistics. We also investigate the extension to dynamic panels and show that both the static and dynamic models can be handled uniformly in the CRE framework. Monte Carlo simulation results indicate that the estimation, inference and testing procedures have desired performance in finite samples. The model is illustrated with two empirical applications to the relationship between cash flow and investment and the nexus between inflation and economic growth.

Keywords Panel threshold model, Multiple covariate-dependent thresholds, Unobserved threshold effects, Testing, Cash flow/investment relationship, Inflation/economic growth nexus.

JEL Classification C12, C13, C15, C33.

1 Introduction

Panel threshold models have become increasingly popular in applied econometric studies over the past two decades, due mainly to that they can capture many essential stylized features of modern economics including asymmetries, multiple equilibria, and nonlinear effects (e.g., Hansen, 2000). However, classical threshold models often assume the thresholds being constant, which has been criticized by the recent literature as the assumption is very restrictive in applications. Thus, a number of authors have extended threshold models by allowing for a nonconstant threshold (e.g., Yang and Su, 2018; Zhu et al., 2019; Yu and Fan, 2021; Yang et al., 2021a; Yang, 2020, 2021; Lee et al., 2021); while these models are promising, they are often limited to a cross-sectional data or time-series data modelling,[1] and are restricted to one

[1] Yang et al. (2021b) extend the kink threshold model of Hansen (2017) to a panel framework with one covariate-dependent threshold.

nonconstant threshold setting; furthermore, classical panel threshold models assume that the unobserved individual effects are the same across the subregimes and the disturbance term does not experience a threshold effect, which may be not suitable in many applications. If they are violated, the traditional estimation method based on the within-group transformation can be inconsistent, as illustrated by Yu et al. (2022).

In this article, we fill this gap in the threshold literature by proposing a panel threshold model with multiple covariate-dependent and time-varying thresholds and unobserved individual-specific threshold effects (PTCDI), in which we allow for unobserved individual-specific threshold effects and a covariate-dependent and time-varying threshold modeled as a function of informative covariates shaping the threshold. Our model can be regarded as an extension of Yu et al. (2022), who extend Hansen's (1999) panel threshold model by allowing for unobserved individual-specific threshold effects, to a time-varying threshold framework. This article first focuses on estimating the model, inference for threshold parameters, and testing for threshold effect and threshold constancy in the proposed model with one covariate-dependent and time-varying threshold and unobserved individual-specific threshold effects, and then discusses the extension to multiple covariate-dependent and time-varying thresholds. In estimating the proposed model, we face two important difficulties. First, as discussed by Yu et al. (2022), the allowance for unobserved individual-specific threshold effects implies the intercepts in the subregimes can be different, and hence, unobserved individual fixed effects can not be eliminated using first differencing or the within-group transformation. This leads to the difficulty in estimating panel models with unobserved individual-specific threshold effects. Second, although the covariate-dependent and time-varying threshold has the advantage of capturing a time-varying reference for assessing the relative magnitude of an economic variable in applications (e.g., Dueker et al., 2013; Yang and Su, 2018; Lee et al., 2021; Yu and Fan, 2021; Yang et al., 2021b),[2] the threshold estimation based on grid search, widely used in the classical threshold literature, is computational expensive, and even infeasible when the dimension of informative covariates shaping the threshold is large.[3]

To this end, following Yu et al. (2022), we overcome the first problem by taking the correlated random effects (CRE) model and use Chamberlain-Mundlak CRE device to control the endogeneity, as the involvement of unobserved individual-specific threshold effects results

[2]The covariate-dependent threshold setting can also be treated as a normalization of the classical threshold model with a linear index (e.g., Seo and Linton, 2007; Lee et al., 2021).

[3]We thank an anonymous referee for raising the issue of selecting the (optimal) set of covariates, which is important as the inappropriate choice of the covariates affecting the threshold can lead to biased estimates and distorted testing results. Future research can work on this issue.

in the failure of the traditional estimation methods such as differencing and the within-group transformation. To overcome the computational problem, we incorporate the MCMC algorithm to lighten the computational burden, and suggest an MCMC-based algorithm to construct the confidence intervals for threshold parameters. It is worth noting that a similar MCMC algorithm has been employed in the threshold literature (Yu and Fan, 2021). We also suggest test statistics for linearity, threshold constancy, and unobserved individual-specific threshold effects. Then, we derive the asymptotic properties of the proposed estimator in the small-threshold-effect framework, and establish the limiting distributions of the suggested test statistics. Moreover, Monte Carlo simulations are conducted to assess the finite sample properties of the proposed estimation procedure and test statistics, and the model is illustrated with two empirical applications to the relationship between cash flow and investment and the nexus between inflation and economic growth. Both simulation and empirical results demonstrate the usefulness of the proposed model.

The remainder of the article is organized as follows. Section 2 introduces the panel threshold model with one covariate-dependent and time-varying threshold and unobserved individual-specific threshold effects, describes the estimation, inference and testing methods for the proposed model, and establishes the asymptotic properties of the suggested estimator and test statistics. Section 3 extends the model to a panel data framework with multiple covariate-dependent and time-varying thresholds. Section 4 discusses the extension of the proposed model to dynamic panels. Section 5 presents Monte Carlo simulations evaluating the finite-sample properties of the estimation, inference and testing procedures. Section 6 provides two empirical applications and Section 7 concludes. In Appendix A, we present a detailed mathematical proof of the asymptotical results. In Appendix B, we report Monte Carlo simulation results to confirm the finite sample performances of the proposed estimation and testing procedures for the model with multiple covariate-dependent thresholds.

2 The model

The panel threshold model with a covariate-dependent and time-varying threshold and unobserved individual-specific threshold effects is given by[4]

$$y_{it} = (\boldsymbol{\beta}_1'\mathbf{x}_{it} + \alpha_{1i} + \sigma_1 u_{it})I(q_{it} \leq \boldsymbol{\gamma}'\mathbf{s}_{it}) + (\boldsymbol{\beta}_2'\mathbf{x}_{it} + \alpha_{2i} + \sigma_2 u_{it})I(q_{it} > \boldsymbol{\gamma}'\mathbf{s}_{it}), \tag{1}$$

for $i = 1, 2, ..., N$ and $t = 1, 2, ..., T$, where $E(u_{it}^2) = 1$, $\mathbf{s}_{it} = (1, \mathbf{s}_{1,it}')' \in \mathbb{R}^{k+1}, \boldsymbol{\gamma} = (\gamma_0, \boldsymbol{\gamma}_{\mathbf{s}}')'$, and there may be overlap between $\mathbf{x}_{it}$ and $\mathbf{s}_{1,it}$.[5] $\mathbf{x}_{it}$ is an $p_{\mathbf{x}}-$dimensional vector of regressors, $\mathbf{s}_{1,it}$ is a $k-$dimensional vector of covariates explaining variation in thresholds over time and/or over individuals, the dependent variable y_{it} and threshold variable q_{it} are scalar, and u_{it} is the disturbance term. σ_1 and σ_2 are used to capture a threshold effect in the conditional variance of y_{it}. α_{1i} and α_{2i} represent the unobserved individual heterogeneity which can be correlated with $\mathbf{x}_{it}$. The threshold setting can be treated as a normalization of the linear index threshold in Seo and Linton (2007). Specifically, the indictor function $I(q_{it} \leq \boldsymbol{\gamma}'\mathbf{s}_{it})$ can be generally modelled through the linear index specification as $I(\boldsymbol{\gamma}^{*'} q_{it}^* \leq 0)$, in which $q_{it}^* = (q_{it}, \mathbf{s}_{it}')'$ and $\boldsymbol{\gamma}^* = (\gamma_q, \gamma_0, \boldsymbol{\gamma}_{\mathbf{s}}')$. When γ_q is normalized to 1, the linear index threshold $I(\boldsymbol{\gamma}^{*'} q_{it}^* \leq 0)$ would degenerate to the covariate-dependent threshold setting.

It is worth noting that, when $\alpha_{1i} = \alpha_{2i}$ and $\sigma_1 = \sigma_2$, the model defined in (1) is actually a panel version of Yu and Fan's (2021) threshold model with a covariate-dependent threshold, and the model can also be treated as an extension of Hansen's (1999) panel threshold model with fixed effects to a nonconstant threshold setting with unobserved threshold effects. The proposed model can also be treated as an extension of Yu et al. (2022) by allowing for a time-varying threshold.

2.1 The estimates and asymptotic properties

As discussed by Yu et al. (2022), the involvement of unobserved individual-specific threshold effects results in the failure of the traditional estimation methods such as differencing and the within-group transformation; therefore they suggest to take the correlated random effects

[4]We first focus on the model with one covariate-dependent and time-varying threshold and unobserved individual-specific threshold effects, and then discuss the extension to multiple covariate-dependent and time-varying thresholds.

[5]In this article, we do not investigate how to select covariates (shaping the threshold) among many candidate variables. In applications, we can choose $\mathbf{s}_{1,it}$ based on economic intuition as argued by Yu and Fan (2021). It is interesting to develop a systematic approach to choose covariates shaping the threshold in the panel data framework with a covariate-dependent and time-varying threshold and unobserved individual-specific threshold effects.

(CRE) model and use Chamberlain-Mundlak CRE device to control the endogeneity. Following the same logic, the widely used estimation procedure based on the within-group transformation or the inferencing method would lead to inconsistent estimator for both threshold and slope parameters in the suggested model. Thus, we follow Yu et al. (2022) to control the endogeneity, and incorporate the MCMC algorithm to lighten the computational burden, and then describe an MCMC-based algorithm to construct the confidence intervals for threshold parameters.

Following Mundlak (1978) and Yu et al. (2022), we assume that α_{1i} and α_{2i} in model (1) are given as follows

$$\alpha_{li} = \boldsymbol{\psi}_l' \mathbf{z}_i + a_{li} \ (l = 1, 2) \text{ with } E[a_{li}|X_i] = 0, \text{and } E[u_{it}|X_i] = 0, \tag{2}$$

where $\mathbf{z}_i' = (\bar{\mathbf{x}}_i', \underline{\mathbf{z}}_i')$ with $\bar{\mathbf{x}}_i' = \frac{1}{T}\sum_{t=1}^{T} \mathbf{x}_{it}$. $X_i = (\mathbf{x}_{i1}', ..., \mathbf{x}_{iT}', \underline{\mathbf{z}}_i')'$ in which $\underline{\mathbf{z}}_i'$ contains the time-invariant variables such as the constant 1. We use $\mathbf{z}_i$ to control the time-invariant effect, and allow that $Cov(a_{1i}, a_{2i}) \neq 0$, and $a_{1i} \neq a_{2i}$; thus, the correlation between α_{1i} and α_{2i} is through $\mathbf{z}_i$ or a_{1i} and a_{2i}. As in Yu et al. (2022), a_{1i} and a_{2i} can be correlated with u_{it}.

Then, we have

$$E[y_{it}|X_i] = (\boldsymbol{\beta}_1' \mathbf{x}_{it} + \boldsymbol{\psi}_1' \mathbf{z}_i) I(q_{it} \leq \boldsymbol{\gamma}' \mathbf{s}_{it}) + (\boldsymbol{\beta}_2' \mathbf{x}_{it} + \boldsymbol{\psi}_2' \mathbf{z}_i) I(q_{it} > \boldsymbol{\gamma}' \mathbf{s}_{it}), \tag{3}$$

and the error term

$$\begin{aligned} e_{it} &= (a_{1i} + \sigma_1 u_{it}) I(q_{it} \leq \boldsymbol{\gamma}' \mathbf{s}_{it}) + (a_{2i} + \sigma_2 u_{it}) I(q_{it} > \boldsymbol{\gamma}' \mathbf{s}_{it}), \\ &=: e_{1it} I(q_{it} \leq \boldsymbol{\gamma}' \mathbf{s}_{it}) + e_{2it} I(q_{it} > \boldsymbol{\gamma}' \mathbf{s}_{it}). \end{aligned} \tag{4}$$

Define $\check{\mathbf{x}}_{it} = (\mathbf{x}_{it}', \mathbf{z}_i')$. Then, the objective function can be written as

$$\begin{aligned} & \widetilde{SSR}_{NT}(\boldsymbol{\gamma}, \boldsymbol{\theta}) \\ = & \sum_{i=1}^{N} \sum_{t=1}^{T} [y_{it} - \boldsymbol{\theta}_1' \check{\mathbf{x}}_{it} I(q_{it} \leq \boldsymbol{\gamma}' \mathbf{s}_{it}) - \boldsymbol{\theta}_2 \check{\mathbf{x}}_{it} I(q_{it} > \boldsymbol{\gamma}' \mathbf{s}_{it})]^2, \end{aligned} \tag{5}$$

where $\boldsymbol{\theta}' = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')$ and $\boldsymbol{\theta}_l' = (\boldsymbol{\beta}_l', \boldsymbol{\psi}_l')$ for $l = 1, 2$.

Following the threshold literature (e.g., Hansen, 1999; Yu and Fan, 2021), we suggest a two-step procedure to estimate $\boldsymbol{\gamma}$. In the first step, we run least squares of y_{it} on $\check{\mathbf{x}}_{it}$ to obtain $\hat{\boldsymbol{\theta}}(\boldsymbol{\gamma})$. Define $\widetilde{SSR}_{NT}(\boldsymbol{\gamma}) = \widetilde{SSR}_{NT}(\boldsymbol{\gamma}, \hat{\boldsymbol{\theta}}(\boldsymbol{\gamma}))$. Then, the threshold parameters can be estimated as

$$\hat{\boldsymbol{\gamma}} = (\hat{\gamma}_0, \hat{\boldsymbol{\gamma}}_{\mathbf{s}}')' = \arg\min_{\boldsymbol{\gamma}\in\boldsymbol{\Gamma}} \widetilde{SSR}_{NT}(\boldsymbol{\gamma}), \tag{6}$$

where $\boldsymbol{\Gamma} = \Gamma_0 \times \Gamma_1$, Γ_0 and $\Gamma_1 = \Gamma_{11} \times \Gamma_{12} \times \cdots \times \Gamma_{1k}$ are the parameter spaces and assumed to be compact. As illustrated by Yu and Fan (2021), the choice of $\hat{\boldsymbol{\gamma}}$ that is consistent with the data cannot be unique in the model with covariate-dependent thresholds. Thus, this article assumes that the arg min operator denotes the centroid of the minimizing set when this set includes more than one point as in Yu and Fan (2021). Once the estimates $\hat{\boldsymbol{\gamma}} = (\hat{\gamma}_0, \hat{\boldsymbol{\gamma}}_{\mathbf{s}}')'$ are obtained, a natural estimate for $\boldsymbol{\theta}$ is given by $\hat{\boldsymbol{\theta}} \equiv \hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\gamma}})$.

To implement the minimization in (6), we may use a two-step estimation procedure based on concentration and grid search, which is widely used in the classical threshold literature (e.g., Hansen, 1999). However, the mentioned grid-search based estimation procedure works poorly, especially when the dimension of $\boldsymbol{s}_{it}$ becomes large, as the inclusion of one more covariate would lead to the computational time increasing $N \times T$ times. Thus, in implementing the minimization in (6), following Yu and Fan (2021), we suggest an estimation procedure based on a Markov chain Monte Carlo (MCMC) to ease the computation burden in estimating threshold parameters.

Algorithm A. Parameter Estimation and confidence intervals based on the MCMC technique.

Step 1. Define $S_{NT} = \widetilde{SSR}_{NT}(\boldsymbol{\gamma})/NT$, and

$$p(\boldsymbol{\gamma}) = \frac{\exp\{-S_{NT}(\boldsymbol{\gamma})\} I(\boldsymbol{\gamma} \in \boldsymbol{\Gamma})}{\int_{\boldsymbol{\Gamma}} \exp\{-S_{NT}(\boldsymbol{\gamma})\} d\boldsymbol{\gamma}},$$

which is a quasi-posterior of $\boldsymbol{\gamma}$ with a uniform prior on $\boldsymbol{\Gamma}$.

Step 2. Use the MCMC method to draw a Markov chain

$$S = (\boldsymbol{\gamma}^{(1)}, \boldsymbol{\gamma}^{(2)}, ..., \boldsymbol{\gamma}^{(B)}),$$

whose marginal density is approximately given by $p(\boldsymbol{\gamma})$.

Step 3. For each $\boldsymbol{\gamma}^{(b)}$, $b = 1, 2, ..., B$, calculate $\widetilde{SSR}_{NT}(\boldsymbol{\gamma}^{(b)})$. Define the initial estimates as $\hat{\boldsymbol{\gamma}}_I = \arg\min_{\boldsymbol{\gamma}\in\boldsymbol{S}} \widetilde{SSR}_{NT}(\boldsymbol{\gamma})$, which may be a set of $\boldsymbol{\gamma}$ values.

Step 4. If desired, update the simulation set in Step 1 from $\boldsymbol{\Gamma}$ to a neighborhood of the initial estimates of $\hat{\boldsymbol{\gamma}}_I$. Repeat Steps 2 and 3 to obtain an updated set of $\boldsymbol{\gamma}$ estimation, say, $\hat{\boldsymbol{\gamma}}_U$. Then $\hat{\boldsymbol{\gamma}}$ is defined as the average of the points in $\hat{\boldsymbol{\gamma}}_U$, and $\hat{\boldsymbol{\theta}} \equiv \hat{\boldsymbol{\theta}}(\hat{\boldsymbol{\gamma}})$.

Step 5. Obtain the residuals $\hat{e}_{it} = y_{it} - \hat{\boldsymbol{\theta}}_1' \check{\mathbf{x}}_{it} I(q_{it} \leq \hat{\boldsymbol{\gamma}}' \mathbf{s}_{it}) - \hat{\boldsymbol{\theta}}_2 \check{\mathbf{x}}_{it} I(q_{it} > \hat{\boldsymbol{\gamma}}' \mathbf{s}_{it})$. Construct a uniformly consistent density estimator $\hat{f}^L$ of e^L based on $\hat{e}_{it}^L$, and a density estimator $\hat{f}^H$ of e^H based on $\hat{e}_{it}^H$ by kernel smoothing, where $\hat{e}_{it}^L$ is $\hat{e}_{it}$ when $q_{it} \leq \hat{\boldsymbol{\gamma}}' \mathbf{s}_{it}$, and $\hat{e}_{it}^H$ is $\hat{e}_{it}$ when $q_{it} > \hat{\boldsymbol{\gamma}}' \mathbf{s}_{it}$.

Step 6. Denote the estimated likelihood function as

$$\begin{aligned}
\widehat{\mathcal{L}}(\boldsymbol{\gamma}) &= \prod_{i=1}^{N} \prod_{t=1}^{T} \Big[\hat{f}^L \left(y_{it} - \hat{\boldsymbol{\theta}}_1' \check{\mathbf{x}}_{it} \right) I \left(q_{it} \leq \boldsymbol{\gamma}' \mathbf{s}_{it} \right) \Big) \\
&\quad + \hat{f}^H \left(y_{it} - \hat{\boldsymbol{\theta}}_2' \check{\mathbf{x}}_{it} \right) I \left(q_{it} > \boldsymbol{\gamma}' \mathbf{s}_{it} \right) \Big] \\
&= \exp \left\{ \sum_{i=1}^{N} \sum_{t=1}^{T} I \left(q_{it} \leq \boldsymbol{\gamma}' \mathbf{s}_{it} \right) \ln \hat{f}^L \left(y_{it} - \hat{\boldsymbol{\theta}}_1' \check{\mathbf{x}}_{it} \right) \right) \\
&\quad + \sum_{i=1}^{N} \sum_{t=1}^{T} I \left(q_{it} > \boldsymbol{\gamma}' \mathbf{s}_{it} \right) \ln \hat{f}^H \left(y_{it} - \hat{\boldsymbol{\theta}}_2' \check{\mathbf{x}}_{it} \right) \Big) \Big\} \\
&:= \exp \left\{ \hat{L}(\boldsymbol{\gamma}) \right\}.
\end{aligned}$$

Step 7. Use the MCMC technique to draw a Markov chain $S = (\boldsymbol{\gamma}^{(1)}, \boldsymbol{\gamma}^{(2)}, ..., \boldsymbol{\gamma}^{(B)})$, whose marginal density is approximately given as

$$\hat{\pi}(\boldsymbol{\gamma}) = \frac{\exp \left\{ \hat{L}(\boldsymbol{\gamma}) \right\} I(\boldsymbol{\gamma} \in \boldsymbol{\Gamma})}{\int_{\boldsymbol{\Gamma}} \exp \left\{ \hat{L}(\boldsymbol{\gamma}) \right\} d\boldsymbol{\gamma}}.$$

Step 8. Denote the jth component of S as $S_j := (\boldsymbol{\gamma}_j^{(1)}, \boldsymbol{\gamma}_j^{(2)}, ..., \boldsymbol{\gamma}_j^{(B)})$. Then the $(1-\alpha)$ CI for γ_j can be constructed as $[\gamma_{j(\alpha/2)}, \gamma_{j(1-\alpha/2)}]$, where $\gamma_{j(\tau)}$ is the τth quantile of S_j.

It is worth noting that the MCMC-based algorithm is only auxiliary to the minimization problem by simulating the possible minimizers. The MCMC-based estimation method is more efficient than the grid search method, as we can simulate $\boldsymbol{\gamma}$ with higher probability when $\widetilde{SSR}_{NT}(\boldsymbol{\gamma})$ is small, i.e., more $\boldsymbol{\gamma}$ values are drawn on (and around) the identified set. As illustrated by Yu and Fan (2021), when B is large enough, one can guarantee that the global minimizer of $\widetilde{SSR}_{NT}(\boldsymbol{\gamma})$ is achieved in Step 3. This is because when B goes to infinity, the density of $\{\boldsymbol{\gamma}^{(b)}\}_{b=1}^{B}$ from Step 2 would converge to $p(\boldsymbol{\gamma})$ in Step 1, and we have $p(\hat{\boldsymbol{\gamma}}_U) > 0$ for the minimizing set $\hat{\boldsymbol{\gamma}}_U$ of $\widetilde{SSR}_{NT}(\boldsymbol{\gamma})$. As the inference for $\boldsymbol{\theta}$ is standard, we focus on the inference for $\hat{\boldsymbol{\gamma}}$ in Algorithm A. As illustrated by Yu and Fan (2021), a by-product of Step 8 in Algorithm A is the semiparametric empirical Bayes estimator (SEBE) of threshold parameters, e.g., the posterior mean based on $\hat{\pi}(\boldsymbol{\gamma})$. Yu and Fan (2021) also provide simulation evidence for the efficiency improvement of SEBE relative to the least square estimator.

In practice, it is important to specify the parameter space for $\boldsymbol{\gamma} = (\gamma_0, \boldsymbol{\gamma}_1')'$. The parameter space can be specified as follows. Denote the least square estimator of q_{it} on $\mathbf{s}_{it}$ as $\hat{\boldsymbol{\gamma}}_1^L$, then the parameter space of $\boldsymbol{\gamma}_1$ can be set as $\boldsymbol{\Gamma}_1 = \bigotimes_{m=1}^{k-1}[\hat{\gamma}_m^L - C, \hat{\gamma}_m^L + C]$ for a large C such that $[\hat{\gamma}_m^L - C, \hat{\gamma}_m^L + C]$ contains zero for each m, ensuring that the estimation procedure works well even when the true model is a constant threshold model. As in the classical threshold model, we set the parameter space for γ_0 as $\Gamma_0 = \Gamma_0(\boldsymbol{\gamma}_1) \subseteq [q_{(\eta NT)}, q_{((1-\eta)NT)}]$ such that at least η portion of observations lie in each regime for some $\eta > 0$ (say, $\eta = 0.10$ or 0.15), in which $q_{(\alpha)}$ is the αth order statistic of q_{it}. When $k = 0$, $\boldsymbol{\Gamma}$ would degenerate to Γ_0, which is exactly the specification of the parameter space for threshold parameter in the classical constant threshold model.

We next study the asymptotic properties of the estimator. Let $f_{t|\mathbf{s}}(q_{it} \,|\, \mathbf{s}_{it})$ be the conditional distribution of the threshold variable q_{it} given $\mathbf{s}_{it}$ and $f_{k|t,\mathbf{s}}(q_{ik} \,|\, q_{it}, \mathbf{s}_{ik}, \mathbf{s}_{it})$ be the conditional distribution of the threshold variable q_{ik} given q_{it}, $\mathbf{s}_{ik}$ and $\mathbf{s}_{it}$. Define the moment functionals

$$
\begin{aligned}
\mathbf{M} &= \sum_{t=1}^{T} E(\check{\mathbf{x}}_{it}\check{\mathbf{x}}_{it}'), \ \mathbf{M}(\boldsymbol{\gamma}) = \sum_{t=1}^{T} E(\check{\mathbf{x}}_{it}\check{\mathbf{x}}_{it}' I(q_{it} \le \boldsymbol{\gamma}'\mathbf{s}_{it})), \\
\mathbf{M}_*(\boldsymbol{\gamma}) &= \begin{bmatrix} \mathbf{M}(\boldsymbol{\gamma}) & \mathbf{0} \\ \mathbf{0} & \mathbf{M} - \mathbf{M}(\boldsymbol{\gamma}) \end{bmatrix}, \ \mathbf{M}_* = \mathbf{M}_*(\boldsymbol{\gamma}^0) \\
\mathbf{D}_t(\boldsymbol{\gamma} \,|\, \mathbf{s}_{it}) &= E\left(\check{\mathbf{x}}_{it}\check{\mathbf{x}}_{it}' \,\middle|\, q_{it} = \boldsymbol{\gamma}'\mathbf{s}_{it}, \mathbf{s}_{it}\right), \\
\mathbf{V}_{lt}(\boldsymbol{\gamma} \,|\, \mathbf{s}_{it}) &= E\left(\check{\mathbf{x}}_{it}\check{\mathbf{x}}_{it}' e_{lit}^2 \,\middle|\, q_{it} = \boldsymbol{\gamma}'\mathbf{s}_{it}, \mathbf{s}_{it}\right). \\
\boldsymbol{\Omega}_1(\boldsymbol{\gamma}) &= \sum_{t=1}^{T}\sum_{\tau=1}^{T} E\left[\check{\mathbf{x}}_{it}\check{\mathbf{x}}_{i\tau}' e_{1it} e_{1i\tau} I(q_{it} \le \boldsymbol{\gamma}'\mathbf{s}_{it}) I(q_{i\tau} \le \boldsymbol{\gamma}'\mathbf{s}_{i\tau})\right] \\
\boldsymbol{\Omega}_2(\boldsymbol{\gamma}) &= \sum_{t=1}^{T}\sum_{\tau=1}^{T} E\left[\check{\mathbf{x}}_{it}\check{\mathbf{x}}_{i\tau}' e_{2it} e_{2i\tau} I(q_{it} > \boldsymbol{\gamma}'\mathbf{s}_{it}) I(q_{i\tau} > \boldsymbol{\gamma}'\mathbf{s}_{i\tau})\right] \\
\boldsymbol{\Omega}_{12}(\boldsymbol{\gamma}) &= \sum_{t=1}^{T}\sum_{\tau \ne t} E\left[\check{\mathbf{x}}_{it}\check{\mathbf{x}}_{i\tau}' e_{1it} e_{2i\tau} I(q_{it} \le \boldsymbol{\gamma}'\mathbf{s}_{it}) I(q_{i\tau} > \boldsymbol{\gamma}'\mathbf{s}_{i\tau})\right] \\
\boldsymbol{\Omega}_*(\boldsymbol{\gamma}) &= \begin{bmatrix} \boldsymbol{\Omega}_1(\boldsymbol{\gamma}) & \boldsymbol{\Omega}_{12}'(\boldsymbol{\gamma}) \\ \boldsymbol{\Omega}_{12}(\boldsymbol{\gamma}) & \boldsymbol{\Omega}_2(\boldsymbol{\gamma}) \end{bmatrix}, \ \boldsymbol{\Omega}_* = \boldsymbol{\Omega}_*(\boldsymbol{\gamma}^0)
\end{aligned}
$$

To establish the asymptotic properties of the estimator of the threshold parameter $\hat{\boldsymbol{\gamma}}$ and the slope $\hat{\boldsymbol{\theta}}$, we first introduce the following assumptions.

Assumption 1. *Assume that*

1. *For each t, $(y_{it}, \mathbf{x}_{it}, \underline{\mathbf{z}}_i, q_{it}, \mathbf{s}_{it})$ are independent and identically distributed (iid) across i; T is fixed and $N \to \infty$.*

2. *For each i, $E(a_{li} \mid X_i) = 0$ and $E(u_{it} \mid X_i) = 0$.*

3. *For each $j = 1, \ldots, p_{\mathbf{x}}$, $P(x_{i1}^j = \cdots = x_{iT}^j) < 1$, where x_{it}^j is the jth element of $\mathbf{x}_{it}$ and $p_{\mathbf{x}}$ is the dimension of $\mathbf{x}_{it}$.*

4. *For $t = 1, \ldots, T$, $E\|\check{\mathbf{x}}_{it}\|^4 < \infty$ and $E|e_{lit}|^4 < \infty$.*

5. *For any $\boldsymbol{\gamma} \in \boldsymbol{\Gamma}$ and $t = 1, \ldots, T$, $E\left(\|\check{\mathbf{x}}_{it}\|^4 \,\middle|\, q_{it} = \boldsymbol{\gamma}'\mathbf{s}_{it}, \mathbf{s}_{it}\right) \le C$, $E\left(\|\check{\mathbf{x}}_{it} e_{lit}\|^4 \,\middle|\, q_{it} = \boldsymbol{\gamma}'\mathbf{s}_{it}, \mathbf{s}_{it}\right) \le C$ for some $C < \infty$, and $0 < f_{t|\mathbf{s}}(q_{it} \mid \mathbf{s}_{it}) \le \bar{f} < \infty$. For $k > t$, $f_{k|t,\mathbf{s}}\left(\boldsymbol{\gamma}^{0\prime}\mathbf{s}_{it} \,\middle|\, q_{it} = \mathbf{s}_{it}(\boldsymbol{\gamma}^0), \mathbf{s}_{ik}, \mathbf{s}_{it}\right) < \infty$.*

6. *$f_{t|\mathbf{s}}(\boldsymbol{\gamma}'\mathbf{s}_{it} \mid \mathbf{s}_{it})$, $\mathbf{D}_t(\boldsymbol{\gamma}|\mathbf{s}_{it})$ and $\mathbf{V}_{lt}(\boldsymbol{\gamma}|\mathbf{s}_{it})$ are continuous at $\boldsymbol{\gamma} = \boldsymbol{\gamma}^0$, where $\boldsymbol{\gamma}^0$ is the true value of $\boldsymbol{\gamma}$.*

7. *$\mathbf{s}_{it}$ is not multicollinear. $\sum_{t=1}^T E\left(\check{\mathbf{x}}_{it}\check{\mathbf{x}}_{it}' \left| I(q_{it} \le \boldsymbol{\gamma}'\mathbf{s}_{it}) - I\left(q_{it} \le \boldsymbol{\gamma}^{0\prime}\mathbf{s}_{it}\right)\right|\right) = \mathbf{0}$ and $\sum_{t=1}^T E\left(\check{\mathbf{x}}_{it}\check{\mathbf{x}}_{it}' I(q_{it} \le \boldsymbol{\gamma}\mathbf{s}_{it}) \left[I(q_{it} \le \boldsymbol{\gamma}'\mathbf{s}_{it}) - I\left(q_{it} \le \boldsymbol{\gamma}^{0\prime}\mathbf{s}_{it}\right)\right]\right) = \mathbf{0}$ if and only if $\boldsymbol{\gamma} = \boldsymbol{\gamma}_0$.*

8. *$\boldsymbol{\theta}_1^0 - \boldsymbol{\theta}_2^0 = \boldsymbol{\delta}_0 = \mathbf{c}N^{-\alpha}$ with $\mathbf{c} \neq \mathbf{0}$ and $0 < \alpha < 1/2$, where $\mathbf{c}$ is fixed.*

9. *$G_T(\boldsymbol{\omega}) > 0$ and $V_{lT}(\boldsymbol{\omega}) > 0$ with $\boldsymbol{\omega} \neq 0$ (defined in Theorem 1). $\det(\mathbf{M}) > \det(\mathbf{M}(\boldsymbol{\gamma})) > 0$ for all $\boldsymbol{\gamma} \in \boldsymbol{\Gamma}$.*

Assumption 1.1 restricts us in the large N small T panels, which is widely used in the panel data models. Assumption 1.2 is the condition of conditional mean independence, while we do not require a_{1i}, a_{2i}, $u_{i1}, ..., u_{iT}$ and $\mathbf{X}_i$ are independent of each other. Assumption 1.3 requires $\mathbf{x}_{it}$ to vary over t to avoid the multicollinear problem. Assumptions 1.4 and 1.5 restrict unconditional and conditional moment bounds to be finite, ensuring that the central limit theorem and the weak law of large numbers hold. Assumption 1.6 requires the distributions of the threshold variable and the conditional moments are continuous, which are typically used in the threshold literature. Assumption 1.7 can ensure that the asymptotic distribution of the threshold estimator is well defined, and ensure that Theorem 2.1 of Newey and McFadden (1994) is applicable in proving the consistency of the estimator as in Yu and Fan (2021). The moment conditions in Assumption 1.7 can be relaxed when T goes to infinity or in a time series framework. It is worth noting that this assumption can allow for $\hat{\gamma}$ being not unique. Assumption 1.8 is the small threshold effect assumption, which is conventional in the literature of threshold models. Assumption 1.9 is full rank condition needed to have nondegenerate asymptotic distributions. These assumptions are typically used in the threshold literature, including Hansen (2017), Yu and Fan (2021), Yang et al. (2021a), and among others.

Theorem 1. *Under Assumption 1, as $N \to \infty$,*

$$N^{1-2\alpha}\left(\hat{\gamma} - \gamma^0\right) \xrightarrow{d} \arg\min_{\boldsymbol{\omega} \in \mathbb{R}^{k+1}} \left[\frac{1}{2} G_T(\boldsymbol{\omega}) - R_T(\boldsymbol{\omega})\right], \tag{7}$$

$$N^{1/2}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0\right) \xrightarrow{d} \mathcal{Z}, \tag{8}$$

where $R_T(\boldsymbol{\omega}) = R_{1T}(\boldsymbol{\omega}) + R_{2T}(\boldsymbol{\omega})$, $R_{lT}(\boldsymbol{\omega})$ is a Gaussian process with a positive variance $V_{lT}(\boldsymbol{\omega})$ when $\boldsymbol{\omega} \neq \mathbf{0}$, which $R_{1T}(\boldsymbol{\omega})$ and $R_{2T}(\boldsymbol{\omega})$ are independent. $G_T(\boldsymbol{\omega})$, $V_{1T}(\boldsymbol{\omega})$ and $V_{2T}(\boldsymbol{\omega})$ are defined as

$$\begin{aligned} G_T(\boldsymbol{\omega}) &= \sum_{t=1}^{T} \mathbf{c}' E\Big(\mathbf{D}_t\left(\gamma^0 \,|\, \mathbf{s}_{it}\right) f_{t|\mathbf{s}}\left(\gamma^{0\prime}\mathbf{s}_{it} \,\middle|\, \mathbf{s}_{it}\right) \left|\mathbf{s}_{it}'\boldsymbol{\omega}\right| \Big)\mathbf{c}, \\ V_{1T}(\boldsymbol{\omega}) &= \sum_{t=1}^{T} \mathbf{c}' E\Big(\mathbf{V}_{1t}\left(\gamma^0 \,|\, \mathbf{s}_{it}\right) f_{t|\mathbf{s}}\left(\gamma^{0\prime}\mathbf{s}_{it} \,\middle|\, \mathbf{s}_{it}\right) \left|\mathbf{s}_{it}'\boldsymbol{\omega}\right| I(\mathbf{s}_{it}'\boldsymbol{\omega} \leq 0)\Big)\mathbf{c}, \\ V_{2T}(\boldsymbol{\omega}) &= \sum_{t=1}^{T} \mathbf{c}' E\Big(\mathbf{V}_{2t}\left(\gamma^0 \,|\, \mathbf{s}_{it}\right) f_{t|\mathbf{s}}\left(\gamma^{0\prime}\mathbf{s}_{it} \,\middle|\, \mathbf{s}_{it}\right) \left|\mathbf{s}_{it}'\boldsymbol{\omega}\right| I(\mathbf{s}_{it}'\boldsymbol{\omega} > 0)\Big)\mathbf{c}, \end{aligned} \tag{9}$$

in which $V_{1T}(\boldsymbol{\omega})$ and $V_{2T}(\boldsymbol{\omega})$are both positive when $\boldsymbol{\omega} \neq \mathbf{0}$. $\mathcal{Z}$ is a Gaussian process with variance $\mathbf{M}_*^{-1}\boldsymbol{\Omega}_*\mathbf{M}_*^{-1}$.

Proof of Theorem 1. See the Appendix. □

Theorem 1 can be viewed as a generalization of the asymptotic result of Theorem 2 in Yu and Fan (2021). It is worth noting that the asymptotic results in Theorem 1 can not be expressed in the form of the two-sided Brownian motion as in the classical threshold literature (e.g., Hansen, 2000; Yu et al., 2022), which is because of the involvement of a covariate-dependent threshold as illustrated by Yu and Fan (2021).

2.2 Model specification testing

In this section, we suggest test statistics for linearity, threshold constancy, and unobserved individual-specific threshold effects, and then establish their limiting distributions.

Before using the proposed panel threshold model with a covariate-dependent threshold and unobserved individual-specific threshold effects, it is desirable to first test whether there is a threshold effect. Consider the null hypothesis of no threshold effect

$$H_0^1 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$$

against the alternative hypothesis

$$H_1^1 : \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2.$$

Under the null, the model (1) degenerates to the linear panel model $y_{it} = \boldsymbol{\theta}_1' \check{\mathbf{x}}_{it} + a_i + \varepsilon_{it}$. Denote the usual least square estimator of the linear panel model as $\tilde{\boldsymbol{\theta}}_1'$, and obtain the residuals $\tilde{e}_{it} = y_{it} - \tilde{\boldsymbol{\theta}}_1' \check{\mathbf{x}}_{it}$, yielding the sum of squared errors $SSR_0 = \sum_{i=1}^N \sum_{t=1}^T \tilde{e}_{it}^2$. In addition, we denote the sum of squared errors of the proposed model as $\widetilde{SSR}_{NT}(\hat{\boldsymbol{\gamma}})$. Likewise, we denote the sum of squared errors of Hansen's (1999) constant threshold model as $SSR_C(\tilde{\boldsymbol{\gamma}})$, where $\tilde{\boldsymbol{\gamma}} = (\tilde{\gamma}, \tilde{\boldsymbol{\gamma}}_{\mathbf{s}}')' = (\tilde{\gamma}, \mathbf{0}')'$. Then, a natural test statistic for the null hypothesis of no threshold effect can be constructed as

$$F_1 = \frac{SSR_0 - \widetilde{SSR}_{NT}(\hat{\boldsymbol{\gamma}})}{\widetilde{SSR}_{NT}(\hat{\boldsymbol{\gamma}})/NT} = \sup_{\boldsymbol{\gamma} \in \boldsymbol{\Gamma}} \frac{SSR_0 - \widetilde{SSR}_{NT}(\boldsymbol{\gamma})}{\widetilde{SSR}_{NT}(\boldsymbol{\gamma})/NT}. \tag{10}$$

If we pass the test for linearity, it is reasonable to determine whether or not the threshold is constant, and whether there are unobserved individual-specific threshold effects. The null hypothesis of threshold constancy can be written as $H_0^2 : \boldsymbol{\gamma}_{\mathbf{s}} = 0$. Then, a test statistic for threshold constancy can be given as

$$F_C = \frac{SSR_C(\tilde{\boldsymbol{\gamma}}) - \widetilde{SSR}_{NT}(\hat{\boldsymbol{\gamma}})}{\widetilde{SSR}_{NT}(\hat{\boldsymbol{\gamma}})/NT}. \tag{11}$$

If we reject the null H_0^1, then there is a threshold effect. Then, it is also natural to test for unobserved individual-specific threshold effects. The null is represented as $\alpha_{1i} = \alpha_{2i}$; in the CRE setting (2), this null for no unobserved threshold effect reduces to $H_0^3 : \boldsymbol{\psi}_1 = \boldsymbol{\psi}_2$. Under the null H_0^3, the model (1) degenerates to $y_{it} = (\boldsymbol{\beta}_1' \mathbf{x}_{it}) I(q_{it} \leq \boldsymbol{\gamma}' \mathbf{s}_{it}) + (\boldsymbol{\beta}_2' \mathbf{x}_{it}) I(q_{it} > \boldsymbol{\gamma}' \mathbf{s}_{it}) + \alpha_i + \varepsilon_{it} = (\boldsymbol{\beta}_1' \mathbf{x}_{it}) I(q_{it} \leq \boldsymbol{\gamma}' \mathbf{s}_{it}) + (\boldsymbol{\beta}_2' \mathbf{x}_{it}) I(q_{it} > \boldsymbol{\gamma}' \mathbf{s}_{it}) + \boldsymbol{\psi}' \mathbf{z}_i + a_i + \varepsilon_{it}$. Denote the sum of squared errors of the above null model as SSR^I. Then, a natural test can be given as follows:

$$F_I = \frac{SSR^I - \widetilde{SSR}_{NT}(\hat{\gamma})}{\widetilde{SSR}_{NT}(\hat{\gamma})/NT}. \tag{12}$$

We next derive the asymptotic distributions of the proposed test statistics for linearity, threshold constancy, and unobserved individual-specific threshold effects.[6]

Theorem 2. *Suppose Assumption 1 holds. Under $H_0^1 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$, we have*

$$F_1 \xrightarrow{d} \frac{1}{\sigma^2} \sup_{\gamma \in \boldsymbol{\Gamma}} \mathcal{Z}'(\gamma)\mathbf{R}_*' \left[\mathbf{R}_*\mathbf{M}_*^{-1}(\gamma)\mathbf{R}_*'\right]^{-1} \mathbf{R}_*\mathcal{Z}(\gamma), \tag{13}$$

under $H_0^2 : \boldsymbol{\gamma}_{\mathbf{s}} = \mathbf{0}$, we have

$$F_C \xrightarrow{d} \frac{1}{\sigma^2} \left[\left(\min_{\boldsymbol{\omega} \in \mathbb{R}_c^{k+1}} G_T(\boldsymbol{\omega}) - 2R_T(\boldsymbol{\omega})\right) - \left(\min_{\boldsymbol{\omega} \in \mathbb{R}^{k+1}} G_T(\boldsymbol{\omega}) - 2R_T(\boldsymbol{\omega})\right)\right], \tag{14}$$

and under $H_0^3 : \boldsymbol{\psi}_1 = \boldsymbol{\psi}_2$, we have

$$F_I \xrightarrow{d} \frac{1}{\sigma^2} \mathcal{Z}'\mathbf{R}_I' \left[\mathbf{R}_I\mathbf{M}_*^{-1}\mathbf{R}_I'\right]^{-1} \mathbf{R}_I\mathcal{Z}, \tag{15}$$

where $\mathbf{R}_*$ and $\mathbf{R}_I$ are matrices such that $\mathbf{R}_*\boldsymbol{\theta} = \boldsymbol{\theta}_1 - \boldsymbol{\theta}_2 = \mathbf{0}$ and $\mathbf{R}_I\boldsymbol{\theta} = \boldsymbol{\psi}_1 - \boldsymbol{\psi}_2 = \mathbf{0}$, $\sigma^2 = T^{-1}\sum_{t=1}^T E(e_{it}^2)$, $\mathbb{R}_c^{k+1} = \{\boldsymbol{\omega} \in \mathbb{R}^{k+1} : \boldsymbol{\omega}_{\mathbf{s}} = \mathbf{0}\}$, $\mathcal{Z}(\gamma)$ is a Gaussian process with variance $\mathbf{M}_*^{-1}(\gamma)\boldsymbol{\Omega}_*(\gamma)\mathbf{M}_*^{-1}(\gamma)$, and $G_T(\boldsymbol{\omega})$, $R_T(\boldsymbol{\omega})$ and $\mathcal{Z}$ are defined as in Theorem 1.

Proof of Theorem 2. See the Appendix. □

As is now well known, threshold models are not identified under the null of linearity, due to the famous Davies problem (Davies, 1977, 1987). Thus, the limiting distribution of the test statistic for linearity is the supremum of a quadratic form of Gaussian process; hence, it is generally not straightforward to tabulate the critical values. Therefore, following the threshold literature (e.g., Hansen, 2017; Yang et al., 2021b), we suggest a parametric bootstrap procedure to calculate the p-values or critical values in implementing the above test statistics.

Algorithm B. Testing for linearity, threshold constancy, and unobserved individual-specific threshold effects

Step 1. Use the original sample $(y_{it}, \mathbf{x}_{it}', q_{it}, \mathbf{s}_{it}')$'s to estimate the linear panel model $y_{it} = \boldsymbol{\theta}_1'\check{\mathbf{x}}_{it} + a_i + u_{it}$ and the constant threshold model $y_{it} = \boldsymbol{\theta}_1'\check{\mathbf{x}}_{it}I(q_{it} \leq \gamma) + \boldsymbol{\theta}_2'\check{\mathbf{x}}_{it}I(q_{it} >$

[6] It is worth noting that $SSR(\gamma)/NT$ is not asymptotically equivalent to $SSR(\gamma)/(NT - N - 2p_{\mathbf{x}})$ under the framework with large N and fixed T, as $\frac{T}{T-1} \neq 1$. Thus, the limiting distributions of the proposed test statistics would differ if we replace the denominator $SSR(\gamma)/NT$ with $SSR(\gamma)/(NT - N - 2p_{\mathbf{x}})$. We thank an anonymous referee for raising this point with us.

$\gamma)+a_i+u^c_{it}$, and the threshold model with no unobserved threshold effects $y_{it} = (\boldsymbol{\beta}'_1\mathbf{x}_{it})I(q_{it} \leq \boldsymbol{\gamma}'\mathbf{s}_{it}) + (\boldsymbol{\beta}'_2\mathbf{x}_{it})I(q_{it} > \boldsymbol{\gamma}'\mathbf{s}_{it}) + \boldsymbol{\psi}'\mathbf{z}_i + a_i + u^I_{it}$, obtain the residual $\hat{e}^l_{it} = y_{it} - \boldsymbol{\theta}'_1\check{\mathbf{x}}_{it}$, $\hat{e}^c_{it} = y_{it} - \boldsymbol{\theta}'_1\check{\mathbf{x}}_{it}I(q_{it} \leq \gamma) - \boldsymbol{\theta}'_2\check{\mathbf{x}}_{it}I(q_{it} > \gamma)$ and $\hat{e}^I_{it} = y_{it} - (\hat{\boldsymbol{\beta}}'_1\mathbf{x}_{it})I(q_{it} \leq \hat{\boldsymbol{\gamma}}'\mathbf{s}_{it}) - (\hat{\boldsymbol{\beta}}'_2\mathbf{x}_{it})I(q_{it} > \hat{\boldsymbol{\gamma}}'\mathbf{s}_{it}) - \hat{\boldsymbol{\psi}}'\mathbf{z}_i$.

Step 2. Generate i.i.d draws ε^*_i from the $N(0,1)$ distribution for $i = 1, ..., N$, and set $e^{l*}_{it} = \hat{e}^l_{it}\varepsilon^*_i$ and $y^{l*}_{it} = \hat{\boldsymbol{\theta}}'_1\check{\mathbf{x}}_{it} + e^{l*}_{it}$; Set $e^{c*}_{it} = \hat{e}^c_{it}\varepsilon^*_i$ and $y^{c*}_{it} = \hat{\boldsymbol{\theta}}'_1\check{\mathbf{x}}_{it}I(q_{it} \leq \hat{\gamma}) + \hat{\boldsymbol{\theta}}'_2\check{\mathbf{x}}_{it}I(q_{it} > \hat{\gamma}) + e^{c*}_{it}$; Set $e^{I*}_{it} = \hat{e}^I_{it}\varepsilon^*_i$ and $y^{I*}_{it} = (\hat{\boldsymbol{\beta}}'_1\mathbf{x}_{it})I(q_{it} \leq \hat{\boldsymbol{\gamma}}'\mathbf{s}_{it}) + (\hat{\boldsymbol{\beta}}'_2\mathbf{x}_{it})I(q_{it} > \hat{\boldsymbol{\gamma}}'\mathbf{s}_{it}) + \hat{\boldsymbol{\psi}}'\mathbf{z}_i + e^{I*}_{it}$.

Step 3. Use the bootstrap sample $(y^{l*}_{it}, \mathbf{x}'_{it}, q_{it}, \mathbf{s}'_{it})$'s, $(y^{c*}_{it}, \mathbf{x}'_{it}, q_{it}, \mathbf{s}'_{it})$'s and $(y^{I*}_{it}, \mathbf{x}'_{it}, q_{it}, \mathbf{s}'_{it})$'s to compute the F-type statistics F_1, F_C and F_I.

Step 4. Repeat Steps 1-3 B times to obtain three samples $F^*_1(1), F^*_1(2), ..., F^*_1(B)$, $F^*_C(1)$, $F^*_C(2), ..., F^*_C(B)$ and $F^*_I(1), F^*_I(2), ..., F^*_I(B)$ of simulated F_1, F_C and F_I statistics.

Step 5. Calculate the empirical p-values by the percentage of the simulated statistics that exceed actual values when the number of B is sufficiently large.

3 Extension to multiple covariate-dependent and time-varying thresholds

In this section, we extend the model to the framework with multiple covariate-dependent and time-varying thresholds. Consider the following panel threshold model with two covariate-dependent thresholds and unobserved individual-specific threshold effects given as[7]

$$\begin{aligned} y_{it} &= (\boldsymbol{\beta}'_1\mathbf{x}_{it} + \alpha_{1i} + \sigma_1\varepsilon_{it})I(q_{it} \leq \boldsymbol{\gamma}'_1\mathbf{s}_{it}) + (\boldsymbol{\beta}'_2\mathbf{x}_{it} + \alpha_{2i} + \sigma_2\varepsilon_{it})I(\boldsymbol{\gamma}'_1\mathbf{s}_{it} < q_{it} \leq \boldsymbol{\gamma}'_2\mathbf{s}_{it}) \\ &\quad +(\boldsymbol{\beta}'_3\mathbf{x}_{it} + \alpha_{3i} + \sigma_3\varepsilon_{it})I(q_{it} > \boldsymbol{\gamma}'_2\mathbf{s}_{it}), \end{aligned} \tag{16}$$

for $i = 1, 2, ..., N$ and $t = 1, 2, ..., T$, where $\mathbf{s}_{it} = (1, \mathbf{s}'_{1,it})' \in \mathbb{R}^{k+1}, \boldsymbol{\gamma}_1 = (\gamma_{10}, \boldsymbol{\gamma}'_{11})', \boldsymbol{\gamma}_2 = (\gamma_{20}, \boldsymbol{\gamma}'_{21})'$. α_{li} $(l = 1, 2, 3)$, defined as in (2), represent the unobserved individual heterogeneity which can be correlated with $\mathbf{x}_{it}$. Other notations are defined as in model (1).

We focus on the model with two covariate-dependent thresholds, as extensions to higher-order threshold models are straightforward following Hansen (1999). We first discuss the estimation and inference for model parameters, and then consider the testing for determining the number of thresholds.

[7]In this article, we assume $\boldsymbol{\gamma}'_1\mathbf{s}_{it} < \boldsymbol{\gamma}'_2\mathbf{s}_{it}$ for all $\mathbf{s}_{it}$'s, and exclude the case in which $\boldsymbol{\gamma}'_1\mathbf{s}_{it} < \boldsymbol{\gamma}'_2\mathbf{s}_{it}$ for some $\mathbf{s}_{it}$'s while $\boldsymbol{\gamma}'_1\mathbf{s}_{it} \geq \boldsymbol{\gamma}'_2\mathbf{s}_{it}$ for other $\mathbf{s}_{it}$'s. Future research can focus on the latter case.

3.1 Estimation

In estimation, we suggest a two-stage procedure. Denote the concentrated sum of squared errors in (16) as $SSR_{NT}(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$. One can estimate $(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$ jointly by minimizing $SSR_{NT}(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$, while solving the joint minimization is often computational expensive. Following the multiple changepoint literature (e.g., Bai, 1997) and the threshold literature (e.g., Hansen, 1999), we can use a sequential estimation procedure to estimate the threshold parameters. In the first stage, we use the objective function defined in (5), and denote the estimator as $\hat{\boldsymbol{\gamma}}_1$. In the constant threshold case, Yu (2015) has showed that such a least square estimator is estimating one of the true threshold parameters under suitable assumptions. Thus, we expect a similar result to hold in the covariate-dependent model, as our model can be treated as a normalization of the classical threshold model with a linear index.[8]

Given the first-stage estimate $\hat{\boldsymbol{\gamma}}_1$, we then define the second-stage objective function as

$$SSR^r_{NT}(\boldsymbol{\gamma}_2) = \begin{cases} SSR_{NT}(\hat{\boldsymbol{\gamma}}_1, \boldsymbol{\gamma}_2), \hat{\boldsymbol{\gamma}}_1'\mathbf{s}_{it} < \boldsymbol{\gamma}_2'\mathbf{s}_{it} \\ SSR_{NT}(\boldsymbol{\gamma}_2, \hat{\boldsymbol{\gamma}}_1), \boldsymbol{\gamma}_2'\mathbf{s}_{it} < \hat{\boldsymbol{\gamma}}_1'\mathbf{s}_{it} \end{cases}, \tag{17}$$

and the second-stage estimator is

$$\hat{\boldsymbol{\gamma}}_2 = \arg\min_{\boldsymbol{\gamma}_2} SSR^r_{NT}(\boldsymbol{\gamma}_2). \tag{18}$$

In the second stage, we estimate the threshold parameter based on Algorithm A with $\widetilde{SSR}_{NT}(\boldsymbol{\gamma})$ being replaced by $SSR^r_{NT}(\boldsymbol{\gamma}_2)$. Given $\hat{\boldsymbol{\gamma}}_2$, we can update $\hat{\boldsymbol{\gamma}}_1$ by replacing $\widetilde{SSR}_{NT}(\boldsymbol{\gamma})$ with $SSR^r_{NT}(\boldsymbol{\gamma}_1) = \begin{cases} SSR_{NT}(\boldsymbol{\gamma}_1, \hat{\boldsymbol{\gamma}}_2), \boldsymbol{\gamma}_1'\mathbf{s}_{it} < \hat{\boldsymbol{\gamma}}_2'\mathbf{s}_{it} \\ SSR_{NT}(\hat{\boldsymbol{\gamma}}_2, \boldsymbol{\gamma}_1), \hat{\boldsymbol{\gamma}}_2'\mathbf{s}_{it} < \boldsymbol{\gamma}_1'\mathbf{s}_{it} \end{cases}$. Given $\hat{\boldsymbol{\gamma}}_1$ and $\hat{\boldsymbol{\gamma}}_2$, $\boldsymbol{\beta}_l$ can be estimated as $\hat{\boldsymbol{\beta}}_l = \hat{\boldsymbol{\beta}}_l(\hat{\boldsymbol{\gamma}}_1, \hat{\boldsymbol{\gamma}}_2)$.

3.2 Confidence Intervals

In this section, we extend the method for constructing confidence intervals for model (1) to the framework with two covariate-dependent thresholds, and thus provide an algorithm to construct confidence intervals for the threshold parameters $(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2)$. The algorithm is as follows.

Algorithm C. Confidence intervals for threshold parameters.

[8] We provide simulation evidence supporting the consistency of the estimator in the case with multiple covariate-dependent and time-varying thresholds. Future research can focus on providing a rigorous proof for this result, i.e., $\hat{\boldsymbol{\gamma}}_1$ would be consistent for either $\boldsymbol{\gamma}_1$ or $\boldsymbol{\gamma}_2$ depending on which effect is stronger. We thank an anonymous referee for raising this point to us.

Step 1. Use the original sample $(y_{it}, \mathbf{x}'_{it}, q_{it}, \mathbf{s}'_{it})$'s to estimate model (16), yielding $(\hat{\boldsymbol{\theta}}, \hat{\gamma}_1, \hat{\gamma}_2)$ and the residuals $\hat{e}_{it} = y_{it} - \hat{\boldsymbol{\theta}}'_1 \check{\mathbf{x}}_{it} I(q_{it} \leq \hat{\gamma}'_1 \mathbf{s}_{it}) - \hat{\boldsymbol{\theta}}_2 \check{\mathbf{x}}_{it} I(\hat{\gamma}'_1 \mathbf{s}_{it} < q_{it} \leq \hat{\gamma}'_2 \mathbf{s}_{it}) - \hat{\boldsymbol{\theta}}_3 \check{\mathbf{x}}_{it} I(q_{it} > \hat{\gamma}'_2 \mathbf{s}_{it})$.

Step 2. Obtain a uniformly consistent density estimator $\hat{f}^L$ of e^L based on $\hat{e}^L_{it}$, $\hat{f}^M$ of e^M based on $\hat{e}^M_{it}$, and a density estimator $\hat{f}^H$ of e^H based on $\hat{e}^H_{it}$ by kernel smoothing, where $\hat{e}^L_{it}$ is $\hat{e}_{it}$ when $q_{it} \leq \hat{\gamma}'_1 \mathbf{s}_{it}$, $\hat{e}^m_{it}$ is $\hat{e}_{it}$ when $\hat{\gamma}'_1 \mathbf{s}_{it} < q_{it} \leq \hat{\gamma}'_2 \mathbf{s}_{it}$, and $\hat{e}^H_{it}$ is $\hat{e}_{it}$ when $q_{it} > \hat{\gamma}'_2 \mathbf{s}_{it}$.

Step 3. Define the estimated likelihood function as

$$\begin{aligned}
\widehat{\mathcal{L}}(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) &= \prod_{i=1}^{N} \prod_{t=1}^{T} \Big[\hat{f}^L \left(y_{it} - \hat{\boldsymbol{\theta}}_1 \check{\mathbf{x}}_{it} \right) I \left(q_{it} \leq \boldsymbol{\gamma}'_1 \mathbf{s}_{it} \right) \Big) \\
&\quad + \hat{f}^M \left(y_{it} - \hat{\boldsymbol{\theta}}_2 \check{\mathbf{x}}_{it} \right) I \left(\boldsymbol{\gamma}'_1 \mathbf{s}_{it} < q_{it} \leq \boldsymbol{\gamma}'_2 \mathbf{s}_{it} \right) \\
&\quad + \hat{f}^H \left(y_{it} - \hat{\boldsymbol{\theta}}_3 \check{\mathbf{x}}_{it} \right) I \left(q_{it} > \boldsymbol{\gamma}'_2 \mathbf{s}_{it} \right) \Big] \\
&= \exp \left\{ \sum_{i=1}^{N} \sum_{t=1}^{T} I \left(q_{it} \leq \boldsymbol{\gamma}'_1 \mathbf{s}_{it} \right) \ln \hat{f}^L \left(y_{it} - \hat{\boldsymbol{\theta}}_1 \check{\mathbf{x}}_{it} \right) \right) \\
&\quad + \sum_{i=1}^{N} \sum_{t=1}^{T} I \left(\boldsymbol{\gamma}'_1 \mathbf{s}_{it} < q_{it} \leq \boldsymbol{\gamma}'_2 \mathbf{s}_{it} \right) \ln \hat{f}^M \left(y_{it} - \hat{\boldsymbol{\theta}}_2 \check{\mathbf{x}}_{it} \right) \Big) \\
&\quad + \sum_{i=1}^{N} \sum_{t=1}^{T} I \left(q_{it} > \boldsymbol{\gamma}'_2 \mathbf{s}_{it} \right) \ln \hat{f}^H \left(y_{it} - \hat{\boldsymbol{\theta}}_3 \check{\mathbf{x}}_{it} \right) \Big) \bigg\} \\
&:= \exp \left\{ \hat{L}(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) \right\}.
\end{aligned}$$

Step 4. Use the MCMC method to draw a Markov chain $S = (\boldsymbol{\gamma}_k^{(1)}, \boldsymbol{\gamma}_k^{(2)}, ..., \boldsymbol{\gamma}_k^{(B)})$ with $k = 1, 2$, whose marginal density is approximately given by

$$\hat{p}(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) = \frac{\exp \left\{ \hat{L}(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) \right\} I(\boldsymbol{\gamma} \in \boldsymbol{\Gamma})}{\int_{\boldsymbol{\Gamma}} \exp \left\{ \hat{L}(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) \right\} d\boldsymbol{\gamma}_1 d\boldsymbol{\gamma}_2}.$$

Step 5. Take out the jth component of S, denoted as $S_j := (\boldsymbol{\gamma}_j^{(1)}, \boldsymbol{\gamma}_j^{(2)}, ..., \boldsymbol{\gamma}_j^{(B)})$. Then the $(1-\alpha)$ CI for γ_j can be constructed as $[\gamma_{j(\alpha/2)}, \gamma_{j(1-\alpha/2)}]$, where $\gamma_{j(\tau)}$ is the τth quantile of S_j.

3.3 Determining the number of thresholds

To determine the number of thresholds, we conduct sequential tests based on the test statistic F_1, which tests the null of no threshold against one threshold. If the null is rejected, then we continue to test the null of one threshold against two thresholds by employing the following test statistic

$$F_2 = \frac{SSR(\hat{\gamma}) - SSR_2(\hat{\gamma}_1, \hat{\gamma}_2)}{SSR_2(\hat{\gamma}_1, \hat{\gamma}_2)/NT}, \tag{19}$$

in which $SSR_2(\hat{\gamma}_1, \hat{\gamma}_2)$ is the sum of squared errors of the model with two thresholds, and $SSR(\hat{\gamma})$ is the sum of squared errors of the model with one threshold.

To implement the above test F_2, we can use a parametric bootstrap procedure based on the null model with one covariate-dependent threshold and unobserved individual-specific threshold effects by employing an algorithm similar with Algorithm B. Based on a similar logic as above, a test statistic, say F_3, for two thresholds against three thresholds can be easily constructed.

4 Extension to dynamic panels

In this section, we briefly discuss an extension of the proposed model to a dynamic setting, i.e., dynamic panel threshold model with a covariate-dependent and time-varying threshold and unobserved individual-specific threshold effects.

Following Yu et al. (2022), we consider the dynamic panel data model given by

$$\begin{aligned} y_{it} &= (\boldsymbol{\beta}_1' \mathbf{x}_{it} + \rho_1 y_{i,t-1} + \alpha_{1i} + \sigma_1 \varepsilon_{it}) I(q_{it} \leq \boldsymbol{\gamma}' \mathbf{s}_{it}) \\ &\quad + (\boldsymbol{\beta}_2' \mathbf{x}_{it} + \rho_2 y_{i,t-1} + \alpha_{2i} + \sigma_2 \varepsilon_{it}) I(q_{it} > \boldsymbol{\gamma}' \mathbf{s}_{it}), \end{aligned} \tag{20}$$

for $i = 1, \ldots, N$ and $t = 0, 1, \ldots, T$. $y_{i,t-1}$ is the lag of the dependent variable y_{it}, and other notations are defined as in (1).

Following Yu et al. (2022) and Wooldridge (2000, 2005), for unobserved heterogeneity α_{li} we assume

$$\alpha_{li} = \boldsymbol{\psi}_l' \mathbf{z}_i + \pi_l y_{i0} + a_{li} \ (l = 1, 2) \text{ with } E[a_{lt} | X_i^T] = 0, \text{and } E[\varepsilon_{it} | X_i^t] = 0, \tag{21}$$

where $\mathbf{z}_i = (\bar{\underline{\mathbf{x}}}_i', \underline{\mathbf{z}}_i')'$ with $\bar{\underline{\mathbf{x}}}_i = \frac{1}{T+1} \sum_{t=0}^{T} \underline{\mathbf{x}}_{it}$, $\underline{\mathbf{x}}_{it} = (\mathbf{x}_{it}', q_{it})'$, $X_i^t = (\underline{\mathbf{x}}_{i0}', \underline{\mathbf{x}}_{i1}', ..., \underline{\mathbf{x}}_{iT}', \underline{\mathbf{z}}_i', y_{i,t-1}, \ldots, y_{i0})'$, $t = 1, ..., T$, and y_{i0} controls the initial condition effect. Then, we have

$$\begin{aligned} E[y_{it} | X_i^t] &= (\boldsymbol{\beta}_1' \mathbf{x}_{it} + \rho_1 y_{i,t-1} + \boldsymbol{\psi}_1' \mathbf{z}_i + \pi_1 y_{i0}) I(q_{it} \leq \boldsymbol{\gamma}' \mathbf{s}_{it}) \\ &\quad + (\boldsymbol{\beta}_2' \mathbf{x}_{it} + \rho_2 y_{i,t-1} + \boldsymbol{\psi}_2' \mathbf{z}_i + \pi_2 y_{i0}) I(q_{it} > \boldsymbol{\gamma}' \mathbf{s}_{it}) \\ &=: \check{\mathbf{x}}_{it}' \boldsymbol{\theta}_1 I(q_{it} \leq \boldsymbol{\gamma}' \mathbf{s}_{it}) + \check{\mathbf{x}}_{it}' \boldsymbol{\theta}_2 I(q_{it} > \boldsymbol{\gamma}' \mathbf{s}_{it}) \end{aligned} \tag{22}$$

for $t = 2, \ldots, T$, and the error term takes the same form as (4), where $\check{\mathbf{x}}_{it} = (\mathbf{x}'_{it}, y_{i,t-1}, \mathbf{z}'_i, y_{i0})'$ and $\boldsymbol{\theta}_l = (\boldsymbol{\beta}'_l, \rho_l, \boldsymbol{\psi}'_l, \pi_l)'$ for $l = 1, 2$. Then, the objective function can be written as

$$\begin{aligned} & \widetilde{SSR}_{NT}(\boldsymbol{\gamma}, \boldsymbol{\theta}) \\ = & \sum_{i=1}^{N} \sum_{t=1}^{T} \left[y_{it} - \boldsymbol{\theta}'_1 \check{\mathbf{x}}_{it} I(q_{it} \leq \boldsymbol{\gamma}' \mathbf{s}_{it}) - \boldsymbol{\theta}'_2 \check{\mathbf{x}}_{it} I(q_{it} > \boldsymbol{\gamma}' \mathbf{s}_{it}) \right]^2, \end{aligned} \tag{23}$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}'_1, \boldsymbol{\theta}'_2)'$. Clearly, the structure of the estimated parameters in dynamic models is the same as in non-dynamic models except for new definitions of $\check{\mathbf{x}}_{it}$ and $\boldsymbol{\theta}_l$. Therefore, the estimation procedure in the non-dynamic model discussed in Section 2 can be directly applied in the dynamic setting.

Denote the estimators of $\boldsymbol{\gamma}$ and $\boldsymbol{\theta}$ as $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\theta}}$. We next study the asymptotic properties of $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\theta}}$. Let $\mathbf{M}$, $\mathbf{M}(\boldsymbol{\gamma})$, $\mathbf{M}_*(\boldsymbol{\gamma})$, $\mathbf{D}_t(\boldsymbol{\gamma} \,|\, \mathbf{s}_{it})$, $\mathbf{V}_{lt}(\boldsymbol{\gamma} \,|\, \mathbf{s}_{it})$, $\boldsymbol{\Omega}_1(\boldsymbol{\gamma})$, $\boldsymbol{\Omega}_2(\boldsymbol{\gamma})$, $\boldsymbol{\Omega}_{12}(\boldsymbol{\gamma})$, $\boldsymbol{\Omega}_*(\boldsymbol{\gamma})$ and $\boldsymbol{\Omega}_*$ take the same form as in non-dynamic models with the new definition of $\check{\mathbf{x}}_{it}$, and $f_{t|\mathbf{s}}(q_{it} \,|\, \mathbf{s}_{it})$ and $f_{k|t,\mathbf{s}}(q_{ik} \,|\, q_{it}, \mathbf{s}_{ik}, \mathbf{s}_{it})$ have the same definition as in non-dynamic models.

Assumption 2. *Assume Assumption 1.4-1.9 hold. We also assume*

1. *For each t, $(y_{it}, \mathbf{x}_{it}, y_{i,t-1}, \underline{\mathbf{z}}_i, y_{i0}, q_{it}, \mathbf{s}_{it})$ are independent and identically distributed (iid) across i.*
2. *For each i, $E\left(a_{li} \,\middle|\, X_i^T\right) = 0$ and $E\left(u_{it} \,\middle|\, X_i^t\right) = 0$.*
3. *For each $j = 1, \ldots, p_*$, $P(x_{i1}^j = \cdots = x_{iT}^j) < 1$, where x_{it}^j is the jth element of $(\mathbf{x}'_{it}, y_{i,t-1})'$ and p_* is the dimension of $(\mathbf{x}'_{it}, y_{i,t-1})'$.*

Assumption 2 is essentially similar to Assumption 1. The following theorem establishes the asymptotic distribution of $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\theta}}$ in the dynamic model.

Theorem 3. *Under Assumption 2, $N^{1-2\alpha}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^0)$ and $N^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0)$ have the same form of asymptotic distributions as in Theorem 1 except that $\mathbf{D}_t(\boldsymbol{\gamma} \,|\, \mathbf{s}_{it})$, $\mathbf{V}_{lt}(\boldsymbol{\gamma} \,|\, \mathbf{s}_{it})$, $\mathbf{M}_*$ and $\boldsymbol{\Omega}_*$ are adjusted with new $\check{\mathbf{x}}_{it}$.*

Proof of Theorem 3. It is noted that the dynamic model can be treated as a non-dynamic model with the addition of variables $y_{i,t-1}$ and y_{i0}. Therefore, under fixed T and $N \to \infty$, the proof of Theorem 3 is the same as Theorem 1 after substituting the new $\check{\mathbf{x}}_{it}$ defined as in (22). $\square$

As the structures of non-dynamic models and dynamic models in the CRE setting are similar, the estimation and inference methods discussed in non-dynamic models are valid in dynamic models.

5 Monte Carlo simulations

In this section, we conduct Monte Carlo simulations to examine the finite sample performances of the proposed estimation and inference procedures and the test statistics. To this end, we consider the following data generating process (DGP):

$$\begin{aligned} y_{it} &= \left(\beta_1 x_{it} + \beta_{11} q_{it} + \alpha_{1i} + \sigma_1 u_{it}\right) I\left(q_{it} \leq \gamma_{it}\right) \\ &+ \left(\beta_2 x_{it} + \beta_{22} q_{it} + \alpha_{2i} + \sigma_2 u_{it}\right) I\left(q_{it} > \gamma_{it}\right), \end{aligned} \tag{24}$$

where $\alpha_{\ell i} = \psi_{\ell 1}\bar{x}_i + \psi_{\ell 2}\bar{q}_i + \psi_{\ell 0} + a_{\ell i}$ for $l = 1, 2$, $\bar{x}_i = \frac{1}{T}\sum_{t=1}^{T} x_{it}$, $\bar{q}_i = \frac{1}{T}\sum_{t=1}^{T} q_{it}$, $\gamma_{it} = \gamma_0 + \gamma_1 s_{it}$, and $u_{1it} = u_{2it} \sim N(0, 1)$, $\sigma_1 = \sigma_2 = 0.5$, a_{1i} and a_{2i} follow $N(0, 1)$. q_{it}, x_{it} and s_{it} follow $i.i.d. N(0, 1)$, and are independent of each other.

In examining the finite sample performance of the estimator, we set the true parameters as $(\gamma_0, \gamma_1) = (0.2, 0.5)$ and $(\beta_2, \beta_{22}, \psi_{20}, \psi_{21}, \psi_{22})$ = (-0.2, -0.2, 0, -0.2, -0.2). For the case with $\alpha_{1i} = \alpha_{2i}$, we set $(\beta_{11}, \psi_{10}, \psi_{11}, \psi_{12}) = (\beta_{22}, \psi_{20}, \psi_{21}, \psi_{22})$, but vary $\beta_1 \in \{0.2, 0.5, 1\}$ to assess the effect of threshold effects on estimation. For the case with $\alpha_{1i} \neq \alpha_{2i}$, we set $\psi_{11} = \{0.2, 0.5, 1\}$, and other parameters are the same as in the $\alpha_{1i} = \alpha_{2i}$ case. In all simulations, we set the number of replications at 1000. We run simulation experiments on a range of sample sizes $T = 2, 5, 10$, and $N = 250, 500, 1000$. To implement the MCMC-based algorithms, our estimation and inference softwares are written based on the runMCMC function in the R package BayesianTools.[9] The simulation results are reported in Tables 1 and 2, in which we report the empirical mean and standard deviation (Std.dev) for each parameter. The simulation results show that the mean of each parameter is fairly close to its true value for all combinations of T and N, and the standard deviations decrease to zero as either T or N increases. When N changes from 500 to 1000 (for a fixed T), the standard deviations of γ_0 and γ_1 decrease by almost half, which is consistent with the $N^{1-2\alpha} = N$ ($\alpha = 0$) convergence rate implied by Theorem 1,[10] especially in the case with $\alpha_{1i} \neq \alpha_{2i}$. Comparing Table 1 (the case with $\alpha_{1i} = \alpha_{2i}$) with Table 2 (the case with $\alpha_{1i} \neq \alpha_{2i}$), we can

[9] For the details, please see the appendix.
[10] In the simulations, we set the magnitude of the threshold effect being fixed, which implies $\alpha = 0$ in Theorem 1.

see that the estimator of the threshold parameters in Table 1 is more accurate than that in Table 2 in terms of standard deviation, which could be contributed to an extra unobserved threshold effect contained in the case considered in Table 2.

In investigating the performance of the test statistics, we set $(\gamma_0, \gamma_1) = (0.2, 0.5)$, $\{(\beta_2, \beta_{22}, \psi_{20}, \psi_{21}, \psi_{22}), (\beta_{12}, \psi_{10}, \psi_{12})\} = \{(-0.2, -0.2, 0, -0.2, -0.2), (-0.2, 0, -0.2)\}$, and vary the parameters $(\beta_1, \psi_{11}) = \{(-0.2, -0.2), (0.2, 0.2), (0.5, 0.5)\}$ to assess the size and power of the F_1 test for threshold effect; we set $\{(\beta_2, \beta_{22}, \psi_{20}, \psi_{21}, \psi_{22}), (\beta_1, \beta_{12}, \psi_{10}, \psi_{11}, \psi_{12})\} = \{(-0.2, -0.2, 0, -0.2, -0.2), (1, -0.2, -0.2, 1, -0.2)\}$, and vary $(\gamma_0, \gamma_1) = \{(0.2, 0), (0.2, 0.2), (0.2, 0.5)\}$ to assess the size and power of the F_C test for threshold constancy. In addition, we set $(\gamma_0, \gamma_1) = (0.2, 0.5)$ and $\{(\beta_2, \beta_{22}, \psi_{20}, \psi_{21}, \psi_{22}), (\beta_1, \beta_{12}, \psi_{10}, \psi_{11}, \psi_{12})\} = \{(-0.2, -0.2, 0, -0.2, -0.2), (1, 1, 0, -0.2, -0.2)\}$ to assess the size of the F_I test for unobserved threshold effect, and vary $(\psi_{11}, \psi_{12}) = \{(0.2, 0.2), (0.5, 0.5)\}$ to assess the power of the F_I test. As reported in Table 3, the simulation results show that the empirical size is close to the 5% nominal level for all the three tests, and the power for the test statistic F_1 (F_C or F_I) increases as either the magnitude of threshold effect (threshold constancy or unobserved threshold effect) or the sample size becomes larger. These results indicate that the proposed test statistics have desired performance in finite samples.

For the case of multiple covariate-dependent thresholds, we also conduct simulations to investigate the performance of the estimation and testing procedures suggested in Section 3. The simulation results support that the estimation and testing procedures work well in finite samples. These simulation results are reported in the appendix.

6 Empirical applications

In this section, we provide two empirical applications to illustrate the proposed model. One is about the famous empirical relationship between firms' cash flow and capital investment spending, following Hansen (1999) who illustrates his well-known panel threshold model with the cash flow/investment relationship. For comparison, the data we use are the same as in Hansen (1999), available from Hansen's website.[11] The other is about the relationship between inflation and economic growth, following Yu et al. (2022) who further capture the unobserved individual-specific threshold effects based on Ramírez-Rondán (2020). For comparison, the data we use in this application are the same as in Yu et al. (2022) and

[11] https://www.ssc.wisc.edu/bhansen. We are grateful to Bruce Hansen for making his data and code publicly available.

Table 1: Estimates for Model parameters ($\alpha_{1i} = \alpha_{2i}$)

T	N		$\gamma_0 = 0.2$	$\gamma_1 = 0.5$	$\beta_1 = 0.2$	β_{11} = -0.2	ψ_{11} =-0.2	ψ_{12} =-0.2	$\psi_{10} = 0$	β_2 = -0.2	β_{22} = -0.2	ψ_{21} = -0.2	ψ_{22} = -0.2	$\psi_{20} = 0$
2	250	Mean	0.181	0.470	0.216	-0.202	-0.193	-0.202	-0.010	-0.212	-0.190	-0.202	-0.206	-0.011
		Std.dev	0.431	0.441	0.113	0.148	0.177	0.193	0.172	0.135	0.192	0.196	0.221	0.236
2	500	Mean	0.200	0.501	0.212	-0.205	-0.209	-0.194	-0.006	-0.217	-0.201	-0.202	-0.205	0.004
		Std.dev	0.243	0.244	0.063	0.091	0.108	0.115	0.100	0.082	0.111	0.130	0.136	0.119
2	1000	Mean	0.204	0.506	0.204	-0.198	-0.201	-0.199	0.001	-0.207	-0.206	-0.205	-0.194	0.004
		Std.dev	0.106	0.124	0.037	0.054	0.067	0.077	0.053	0.048	0.069	0.080	0.087	0.073
5	250	Mean	0.194	0.500	0.210	-0.199	-0.197	-0.204	-0.001	-0.210	-0.199	-0.196	-0.218	0.000
		Std.dev	0.185	0.199	0.042	0.067	0.170	0.178	0.089	0.053	0.078	0.178	0.192	0.108
5	500	Mean	0.202	0.506	0.203	-0.201	-0.192	-0.200	-0.001	-0.205	-0.198	-0.198	-0.205	-0.003
		Std.dev	0.082	0.100	0.026	0.040	0.115	0.123	0.057	0.033	0.053	0.122	0.134	0.073
5	1000	Mean	0.199	0.501	0.202	-0.199	-0.198	-0.200	0.001	-0.203	-0.200	-0.195	-0.198	0.000
		Std.dev	0.031	0.045	0.018	0.027	0.081	0.083	0.041	0.023	0.037	0.085	0.093	0.050
10	250	Mean	0.204	0.509	0.203	-0.197	-0.195	-0.200	0.006	-0.206	-0.200	-0.200	-0.206	0.000
		Std.dev	0.086	0.105	0.025	0.039	0.208	0.225	0.075	0.030	0.051	0.220	0.232	0.085
10	500	Mean	0.202	0.502	0.200	-0.201	-0.200	-0.202	-0.001	-0.201	-0.198	-0.203	-0.203	-0.002
		Std.dev	0.033	0.046	0.017	0.027	0.150	0.149	0.049	0.022	0.035	0.151	0.154	0.057
10	1000	Mean	0.200	0.500	0.201	-0.199	-0.200	-0.197	-0.002	-0.201	-0.201	-0.202	-0.192	-0.003
		Std.dev	0.015	0.023	0.011	0.019	0.110	0.109	0.036	0.015	0.025	0.118	0.118	0.039
T	N		$\gamma_0 = 0.2$	$\gamma_1 = 0.5$	$\beta_1 = 0.5$	β_{11} = -0.2	ψ_{11} =-0.2	ψ_{12} =-0.2	$a_{1i} = 0$	β_2 = -0.2	β_{22} = -0.2	ψ_{21} = -0.2	ψ_{22} = -0.2	$a_{2i} = 0$
2	250	Mean	0.211	0.506	0.510	-0.201	-0.201	-0.193	-0.003	-0.220	-0.197	-0.204	-0.208	-0.003
		Std.dev	0.156	0.185	0.078	0.115	0.136	0.161	0.112	0.101	0.157	0.168	0.197	0.163
2	500	Mean	0.199	0.501	0.504	-0.200	-0.201	-0.199	-0.003	-0.204	-0.198	-0.199	-0.193	-0.006
		Std.dev	0.068	0.090	0.052	0.071	0.096	0.103	0.074	0.067	0.092	0.110	0.127	0.099
2	1000	Mean	0.199	0.500	0.505	-0.199	-0.204	-0.202	0.001	-0.205	-0.200	-0.196	-0.202	0.001
		Std.dev	0.029	0.037	0.034	0.050	0.068	0.074	0.051	0.045	0.064	0.080	0.086	0.064
5	250	Mean	0.201	0.502	0.505	-0.200	-0.191	-0.202	-0.003	-0.206	-0.200	-0.195	-0.202	-0.003
		Std.dev	0.045	0.061	0.036	0.057	0.163	0.166	0.080	0.044	0.072	0.173	0.180	0.103
5	500	Mean	0.199	0.501	0.503	-0.202	-0.204	-0.204	-0.002	-0.203	-0.199	-0.206	-0.200	0.000
		Std.dev	0.025	0.033	0.025	0.040	0.119	0.116	0.057	0.032	0.049	0.121	0.129	0.068
5	1000	Mean	0.200	0.500	0.500	-0.200	-0.203	-0.198	0.002	-0.201	-0.198	-0.198	-0.205	0.000
		Std.dev	0.012	0.016	0.017	0.027	0.082	0.083	0.042	0.021	0.036	0.090	0.090	0.050
10	250	Mean	0.201	0.500	0.502	-0.200	-0.206	-0.203	0.000	-0.203	-0.200	-0.200	-0.199	0.001
		Std.dev	0.023	0.029	0.023	0.038	0.207	0.208	0.075	0.030	0.050	0.217	0.229	0.085
10	500	Mean	0.200	0.501	0.501	-0.198	-0.206	-0.211	0.000	-0.201	-0.201	-0.206	-0.211	-0.001
		Std.dev	0.011	0.015	0.017	0.026	0.150	0.154	0.051	0.022	0.034	0.155	0.159	0.058
10	1000	Mean	0.200	0.500	0.500	-0.200	-0.200	-0.207	0.001	-0.200	-0.200	-0.199	-0.207	0.001
		Std.dev	0.006	0.008	0.012	0.018	0.105	0.104	0.036	0.015	0.024	0.113	0.112	0.040
T	N		$\gamma_0 = 0.2$	$\gamma_1 = 0.5$	$\beta_1 = 1$	β_{11} = -0.2	ψ_{11} =-0.2	ψ_{12} =-0.2	$a_{1i} = 0$	β_2 = -0.2	β_{22} = -0.2	ψ_{21} = -0.2	ψ_{22} = -0.2	$a_{2i} = 0$
2	250	Mean	0.200	0.504	1.007	-0.198	-0.200	-0.202	0.001	-0.211	-0.208	-0.188	-0.198	0.007
		Std.dev	0.052	0.065	0.071	0.105	0.133	0.149	0.103	0.090	0.137	0.158	0.180	0.139
2	500	Mean	0.200	0.501	1.003	-0.199	-0.200	-0.198	0.004	-0.205	-0.202	-0.196	-0.191	0.001
		Std.dev	0.021	0.028	0.049	0.070	0.099	0.102	0.070	0.063	0.089	0.108	0.114	0.094
2	1000	Mean	0.201	0.501	1.000	-0.201	-0.201	-0.195	-0.002	-0.201	-0.199	-0.201	-0.199	-0.005
		Std.dev	0.012	0.015	0.035	0.049	0.065	0.074	0.049	0.043	0.062	0.076	0.087	0.064
5	250	Mean	0.199	0.500	1.003	-0.198	-0.204	-0.207	0.003	-0.203	-0.201	-0.203	-0.204	-0.001
		Std.dev	0.019	0.025	0.035	0.054	0.161	0.168	0.080	0.043	0.070	0.170	0.180	0.094
5	500	Mean	0.200	0.500	1.000	-0.198	-0.197	-0.204	0.005	-0.201	-0.199	-0.199	-0.207	0.003
		Std.dev	0.010	0.012	0.025	0.037	0.114	0.119	0.056	0.031	0.050	0.125	0.128	0.068
5	1000	Mean	0.200	0.500	1.000	-0.199	-0.201	-0.200	0.002	-0.200	-0.201	-0.198	-0.198	0.002
		Std.dev	0.005	0.007	0.017	0.027	0.080	0.083	0.038	0.022	0.035	0.089	0.090	0.047
10	250	Mean	0.200	0.500	1.000	-0.199	-0.202	-0.203	0.000	-0.200	-0.201	-0.199	-0.198	-0.001
		Std.dev	0.009	0.012	0.024	0.037	0.219	0.222	0.069	0.029	0.048	0.228	0.235	0.082
10	500	Mean	0.200	0.500	0.999	-0.199	-0.203	-0.200	0.001	-0.199	-0.199	-0.200	-0.200	-0.003
		Std.dev	0.005	0.006	0.016	0.026	0.146	0.159	0.048	0.021	0.034	0.155	0.163	0.058
10	1000	Mean	0.200	0.500	0.999	-0.199	-0.200	-0.199	0.000	-0.199	-0.199	-0.200	-0.200	-0.001
		Std.dev	0.003	0.004	0.012	0.018	0.106	0.108	0.036	0.015	0.024	0.112	0.113	0.038

Table 2: Estimates for Model parameters ($\alpha_{1i} \neq \alpha_{2i}$)

T	N		$\gamma_0 = 0.2$	$\gamma_1 = 0.5$	$\beta_1 = 0.2$	β_{11} = -0.2	ψ_{11} =0.2	ψ_{12} =-0.2	$\psi_{10} = 0$	β_2 = -0.2	β_{22} = -0.2	ψ_{21} = -0.2	ψ_{22} = -0.2	$\psi_{20} = 0$
2	250	Mean	0.188	0.513	0.211	-0.198	0.204	-0.197	-0.001	-0.213	-0.191	-0.209	-0.203	-0.015
		Std.dev	0.201	0.201	0.084	0.115	0.142	0.156	0.117	0.102	0.158	0.170	0.189	0.171
2	500	Mean	0.198	0.509	0.203	-0.202	0.206	-0.197	0.001	-0.204	-0.196	-0.205	-0.206	0.001
		Std.dev	0.078	0.102	0.051	0.074	0.094	0.105	0.073	0.067	0.094	0.112	0.130	0.099
2	1000	Mean	0.201	0.501	0.204	-0.199	0.200	-0.201	0.000	-0.206	-0.197	-0.200	-0.202	-0.005
		Std.dev	0.036	0.049	0.035	0.051	0.066	0.073	0.050	0.046	0.065	0.081	0.084	0.067
5	250	Mean	0.207	0.507	0.204	-0.200	0.203	-0.203	0.000	-0.206	-0.202	-0.205	-0.194	0.002
		Std.dev	0.095	0.130	0.038	0.058	0.167	0.171	0.082	0.047	0.079	0.187	0.195	0.105
5	500	Mean	0.200	0.503	0.202	-0.200	0.204	-0.205	0.001	-0.204	-0.200	-0.194	-0.201	-0.001
		Std.dev	0.044	0.057	0.025	0.041	0.113	0.121	0.056	0.032	0.048	0.122	0.132	0.067
5	1000	Mean	0.200	0.500	0.201	-0.199	0.199	-0.203	-0.001	-0.202	-0.199	-0.201	-0.201	-0.003
		Std.dev	0.021	0.028	0.018	0.027	0.082	0.084	0.039	0.023	0.036	0.084	0.091	0.048
10	250	Mean	0.200	0.502	0.204	-0.200	0.206	-0.206	0.003	-0.205	-0.198	-0.205	-0.204	0.000
		Std.dev	0.055	0.078	0.024	0.040	0.209	0.214	0.074	0.030	0.049	0.217	0.227	0.084
10	500	Mean	0.201	0.501	0.201	-0.198	0.206	-0.202	0.002	-0.201	-0.200	-0.197	-0.196	0.000
		Std.dev	0.028	0.033	0.017	0.025	0.151	0.155	0.052	0.021	0.035	0.155	0.161	0.056
10	1000	Mean	0.200	0.500	0.200	-0.201	0.203	-0.193	-0.001	-0.201	-0.199	-0.201	-0.194	-0.002
		Std.dev	0.013	0.017	0.011	0.018	0.110	0.105	0.036	0.015	0.024	0.112	0.110	0.040
T	N		$\gamma_0 = 0.2$	$\gamma_1 = 0.5$	$\beta_1 = 0.5$	β_{11} = -0.2	ψ_{11} =0.5	ψ_{12} =-0.2	$a_{1i} = 0$	β_2 = -0.2	β_{22} = -0.2	ψ_{21} = -0.2	ψ_{22} = -0.2	$a_{2i} = 0$
2	250	Mean	0.197	0.505	0.504	-0.202	0.501	-0.200	-0.004	-0.203	-0.197	-0.209	-0.202	-0.005
		Std.dev	0.057	0.073	0.072	0.102	0.139	0.144	0.102	0.094	0.136	0.160	0.180	0.137
2	500	Mean	0.202	0.501	0.503	-0.197	0.497	-0.204	0.005	-0.207	-0.202	-0.197	-0.201	0.000
		Std.dev	0.026	0.033	0.050	0.071	0.092	0.100	0.073	0.066	0.090	0.110	0.122	0.094
2	1000	Mean	0.200	0.500	0.500	-0.201	0.503	-0.200	0.001	-0.202	-0.202	-0.199	-0.198	0.003
		Std.dev	0.012	0.017	0.035	0.049	0.065	0.074	0.049	0.044	0.066	0.075	0.087	0.066
5	250	Mean	0.202	0.499	0.501	-0.200	0.502	-0.202	0.000	-0.204	-0.197	-0.199	-0.206	-0.002
		Std.dev	0.031	0.040	0.035	0.054	0.164	0.170	0.079	0.045	0.071	0.175	0.179	0.096
5	500	Mean	0.200	0.501	0.501	-0.200	0.500	-0.201	-0.001	-0.202	-0.202	-0.202	-0.203	0.001
		Std.dev	0.016	0.020	0.025	0.037	0.112	0.121	0.056	0.031	0.050	0.125	0.130	0.068
5	1000	Mean	0.200	0.500	0.500	-0.198	0.500	-0.201	0.001	-0.200	-0.201	-0.203	-0.194	-0.001
		Std.dev	0.007	0.009	0.018	0.027	0.080	0.083	0.040	0.022	0.036	0.087	0.092	0.048
10	250	Mean	0.199	0.501	0.502	-0.199	0.501	-0.204	0.004	-0.201	-0.202	-0.207	-0.205	0.008
		Std.dev	0.020	0.027	0.024	0.037	0.224	0.213	0.071	0.030	0.047	0.225	0.224	0.081
10	500	Mean	0.200	0.501	0.501	-0.200	0.505	-0.197	-0.002	-0.200	-0.201	-0.195	-0.197	-0.002
		Std.dev	0.010	0.013	0.016	0.026	0.144	0.156	0.051	0.021	0.035	0.156	0.161	0.059
10	1000	Mean	0.200	0.500	0.500	-0.201	0.503	-0.201	-0.001	-0.200	-0.200	-0.199	-0.200	0.000
		Std.dev	0.005	0.007	0.011	0.018	0.110	0.107	0.036	0.014	0.024	0.113	0.113	0.040
T	N		$\gamma_0 = 0.2$	$\gamma_1 = 0.5$	$\beta_1 = 1$	β_{11} = -0.2	ψ_{11} =1	ψ_{12} =-0.2	$a_{1i} = 0$	β_2 = -0.2	β_{22} = -0.2	ψ_{21} = -0.2	ψ_{22} = -0.2	$a_{2i} = 0$
2	250	Mean	0.201	0.501	0.998	-0.198	0.999	-0.202	0.004	-0.200	-0.199	-0.206	-0.203	-0.001
		Std.dev	0.024	0.031	0.073	0.099	0.141	0.147	0.101	0.091	0.132	0.160	0.173	0.130
2	500	Mean	0.200	0.499	0.999	-0.202	1.004	-0.200	0.002	-0.200	-0.199	-0.200	-0.207	0.001
		Std.dev	0.012	0.015	0.050	0.073	0.093	0.103	0.071	0.061	0.092	0.103	0.122	0.092
2	1000	Mean	0.200	0.500	0.999	-0.197	1.003	-0.203	0.004	-0.199	-0.202	-0.202	-0.201	0.001
		Std.dev	0.006	0.008	0.034	0.047	0.066	0.074	0.050	0.043	0.061	0.079	0.083	0.062
5	250	Mean	0.200	0.501	1.000	-0.198	1.008	-0.206	-0.002	-0.200	-0.198	-0.201	-0.197	-0.005
		Std.dev	0.012	0.017	0.034	0.054	0.163	0.173	0.081	0.043	0.071	0.177	0.191	0.095
5	500	Mean	0.200	0.500	0.998	-0.200	0.993	-0.205	0.000	-0.199	-0.202	-0.205	-0.205	0.002
		Std.dev	0.006	0.008	0.025	0.038	0.117	0.116	0.056	0.031	0.048	0.123	0.127	0.067
5	1000	Mean	0.200	0.500	0.999	-0.200	1.003	-0.201	0.000	-0.199	-0.199	-0.196	-0.203	-0.002
		Std.dev	0.004	0.006	0.018	0.027	0.081	0.080	0.041	0.022	0.036	0.085	0.090	0.048
10	250	Mean	0.200	0.500	1.000	-0.201	0.999	-0.196	-0.003	-0.200	-0.201	-0.195	-0.194	-0.001
		Std.dev	0.007	0.010	0.023	0.035	0.209	0.216	0.072	0.029	0.048	0.215	0.229	0.078
10	500	Mean	0.200	0.500	1.000	-0.199	1.000	-0.204	0.001	-0.200	-0.203	-0.199	-0.203	0.004
		Std.dev	0.004	0.006	0.016	0.026	0.151	0.154	0.050	0.021	0.034	0.155	0.163	0.060
10	1000	Mean	0.200	0.500	1.000	-0.200	1.000	-0.201	0.000	-0.200	-0.202	-0.199	-0.201	0.002
		Std.dev	0.003	0.004	0.012	0.019	0.105	0.112	0.036	0.015	0.023	0.111	0.118	0.040

Table 3: Finite-sample size and power of the F-type test statistics.

		F_1			F_C			F_I		
		size	power		size	power		size	power	
			(β_1, ψ_{11})			(γ_0, γ_1)			(ψ_{11}, ψ_{12})	
T	N	$(-0.2, -0.2)$	$(0.2, 0.2)$	$(0.5, 0.5)$	$(0.2, 0)$	$(0.2, 0.2)$	$(0.2, 0.5)$	$(-0.2, -0.2)$	$(0.2, 0.2)$	$(0.5, 0.5)$
2	250	0.048	0.725	0.998	0.041	0.999	1.000	0.046	0.531	0.833
	500	0.051	1.000	1.000	0.047	1.000	1.000	0.052	0.930	1.000
	1000	0.053	1.000	1.000	0.048	1.000	1.000	0.049	1.000	1.000
5	250	0.051	1.000	1.000	0.056	1.000	1.000	0.048	0.925	1.000
	500	0.050	1.000	1.000	0.051	1.000	1.000	0.051	1.000	1.000
	1000	0.049	1.000	1.000	0.050	1.000	1.000	0.050	1.000	1.000
10	250	0.050	1.000	1.000	0.050	1.000	1.000	0.051	1.000	1.000
	500	0.050	1.000	1.000	0.050	1.000	1.000	0.050	1.000	1.000
	1000	0.050	1.000	1.000	0.050	1.000	1.000	0.050	1.000	1.000

Ramírez-Rondán (2020).[12]

6.1 Investment and financing constraints

According to the literature (e.g, Fazzari et al., 1988; Hansen, 1999), the impact of a firm's cash flow on its investment varies with financing constraints; thus, a positive impact would be observed only when the firm faces financing constraints, and the relationship between firms' cash flow and investment is enhanced with the increasing of the degree of financing constraints. Based on the classical panel threshold model, Hansen (1999) concludes that his empirical results "are somewhat consistent with the theory of financing constraints", and hence partly consistent with Fazzari et al. (1988). In this section, we show that the empirical results will be completely consistent with Fazzari et al. (1988) by allowing for covariate-dependent thresholds and unobserved individual-specific threshold effects.

Following Hansen (1999), threshold models are employed to distinguish constrained and unconstrained firms by using the ratio of long-term debt to assets as a threshold variable. In doing so, a higher value of this ratio is associated with a higher degree of financing constraints. The reason why we consider a nonconstant threshold can be justified as follows. In distinguishing constrained and unconstrained firms, "banks will be reluctant to lend money to debt-heavy firms" as argued by Hansen (1999); however, a reasonable logic is that, if a debt-heavy firm has a very good investment opportunity, banks may be willing to lend money to such a firm, which will ease financial constraints faced by the firm. Moreover, a firm may have different individual-specific and time-invariant effects when facing different degrees of financing constraints, thus it is reasonable to expect unobserved individual-specific threshold effects in investigating the cash flow/investment relationship. This motivates us to modify

[12] We thank N.R. Ramírez-Rondán for sharing his dataset to us.

Hansen's (1999) empirical model as follows

$$
\begin{aligned}
I_{it} &= \theta_1 Q_{it-1} + \theta_2 Q_{it-1}^2 + \theta_3 Q_{it-1}^3 + \theta_4 D_{it-1} + \theta_5 Q_{it-1} D_{it-1} \\
&\quad + (\beta_1 CF_{it-1} + \alpha_{1i} + \sigma_1 u_{it}) I(D_{it-1} \leq \boldsymbol{\gamma}_1' \mathbf{S}_{it}) \\
&\quad + (\beta_2 CF_{it-1} + \alpha_{2i} + \sigma_2 u_{it}) I(\boldsymbol{\gamma}_1' \mathbf{S}_{it} < D_{it-1} \leq \boldsymbol{\gamma}_2' \mathbf{S}_{it}) \\
&\quad + (\beta_3 CF_{it-1} + \alpha_{3i} + \sigma_3 u_{it}) I(D_{it-1} > \boldsymbol{\gamma}_2' \mathbf{S}_{it}),
\end{aligned} \tag{25}
$$

where I_{it} is the ratio of investment to capital, Q_{it} is the ratio of total market value to assets (known as Tobin's Q), CF_{it} is the ratio of cash flow to assets, and D_{it} is the ratio of long term debt to assets as in Hansen (1999). $\mathbf{S}_{it}$ is set as $(1, Q_{it-1})$ reflecting that the threshold may vary with firms' investment opportunity. Compared with Hansen's (1999) empirical model, the main differences of our empirical model are as follows. First, we allow for time-varying/covariate-dependent thresholds. Second, we allow for unobservable threshold effects. When $\mathbf{S}_{it} \equiv 1$, the model will degenerate to Hansen's (1999) empirical model. The time-varying/covariate-dependent thresholds ($\boldsymbol{\gamma}_1' \mathbf{S}_{it}$ and $\boldsymbol{\gamma}_2' \mathbf{S}_{it}$) can be explained as time-varying references to separate the debt level into three regimes, i.e., low debt, moderate debt and high debt.

Table 4: Empirical application of investment and financing constraints.

	Hansen's (1999) model		PTCDI	
	Estimate	95%CI	Estimate	95%CI
Q_{it-1}	0.010	[0.006,0.014]	0.010	[0.005,0.014]
$Q_{it-1}^2/10^3$	-0.198	[-0.323,-0.073]	-0.182	[-0.423,0.060]
$Q_{it-1}^3/10^6$	1.047	[0.169,1.925]	0.947	[-2.241,4.135]
D_{it-1}	-0.016	[-0.034,0.002]	-0.011	[-0.023,0.002]
$Q_{it-1}D_{i-1}$	0.001	[-0.003,0.005]	-0.001	[-0.009,0.007]
$CF_{it-1}I\left(D_{it-1} \leq \hat{\gamma}_1' \mathbf{S}_{it}\right)$	0.063	[0.036,0.090]	0.062	[0.045,0.079]
$CF_{it-1}I\left(\hat{\gamma}_1' \mathbf{S}_{it} < D_{it-1} \leq \hat{\gamma}_2' \mathbf{S}_{it}\right)$	0.098	[0.078,0.118]		
$CF_{it-1}I\left(D_{it-1} > \hat{\gamma}_2' \mathbf{S}_{it}\right)$	0.039	[-0.022,0.100]	0.112	[0.033,0.191]
Threshold1	0.0157	[0.014,0.018]	-0.363	[-0.485,-0.311]
			1.373	[1.266,1.659]
Threshold2	0.5362	[0.531,0.563]		
	Statistic	p-value	Statistic	p-value
F_1	32.649	0.003	143.206	0.029
F_2	25.799	0.013	353.532	0.167
F_3	4.181	0.736		
F_C			62.352	0.008
F_I			121.155	0.031

Table 4 reports the empirical results, in which 95% confidence intervals for threshold parameters are based on Algorithm A and $p-$values of the tests are computed based on Algorithm B with 1000 bootstrap replications. The confidence intervals for slope parameters are based on cluster-robust standard errors. For the sake of comparison, we also replicate and report the empirical results of Hansen (1999) in Column 1 using the R codes and data available from Hansen's website. We first focus on the testing results. The test statistics F_1 and F_2 are employed to determine the number of thresholds by using 1000 replications in simulating the p-values. Different from Hansen (1999), our testing results support one threshold in the empirical model. Therefore, in Table 4 we only report the slope estimates based on the PTCDI model with one threshold. Compared with Hansen's (1999) results of regime-dependent coefficients ($\hat{\beta}_1 = 0.063$, $\hat{\beta}_2 = 0.098$, $\hat{\beta}_3 = 0.039$), which are not increasing with the degree of financing constraints, and hence not expected from the theory of financing constraints, our results show that regime-dependent coefficients ($\hat{\beta}_1 = 0.062$, $\hat{\beta}_2 = 0.112$) are completely consistent with theory of financing constraints, as they are strictly increasing with the degree of financing constraints.

According to the F_C test statistic for threshold constancy and the F_I test statistic for unobserved threshold effects, both the null of constant threshold and the null of no unobserved threshold effect are rejected. Thus, the difference in the empirical results based on the PTCDI model and Hansen's (1999) can be explained by the nonconstant threshold and unobserved threshold effects. Therefore, we conclude that the PTCDI model seems to be a meaningful complement of the classical model of Hansen (1999) in empirical studies.

6.2 Inflation and economic growth

A vast amount of literature argues that low inflation has no effect on economic growth, while high inflation is harmful to economic growth (e.g., Dornbusch and Fischer, 1993; Bruno and Easterly, 1998). Thus, a number of empirical literature focuses on the inflation threshold under which inflation has an effect on economic growth (e.g., Khan and Senhadji, 2001; Vaona and Schiavo, 2007; Kremer et al., 2013). More recently, Ramírez-Rondán (2020) suggests a dynamic panel threshold model to study this question on the basis of Kremer et al. (2013). To further capture the unobserved individual-specific threshold effects, Yu et al. (2022) revisit the same empirical question by extending Ramírez-Rondán's (2020) empirical model to allow for heterogenous individual fixed effects.

In this section, we add the literature by considering a covariate-dependent inflation threshold. As industrialized and nonindustrialized countries have different inflation thresholds (Khan and Senhadji, 2001; Kremer et al., 2013; Ramírez-Rondán, 2020), the literature mentioned above conducts separate analyses for industrialized and nonindustrialized countries to achieve this goal to some extent. However, if the sample size of a subsample is too small relative to the number of parameters, the analysis of this subsample would become inaccurate; moreover, it is difficult to test whether the threshold difference is statistically significant. Therefore, based on the work of Yu et al. (2022), to further demonstrate the usefulness of the proposed model, we extend their empirical model by allowing for a covariate-dependent threshold given by

$$\begin{aligned} y_{it} - y_{i,t-1} = \mathbf{x}'_{1it}\beta_1 \quad &+ \quad (\log(Inf_{it})\beta_{12} + \alpha_{1i} + \sigma_1 u_{it})I(Inf_{it} \leq \gamma_0 + \gamma_1 Ind_i) \\ &+ \quad (\log(Inf_{it})\beta_{22} + \alpha_{2i} + \sigma_2 u_{it})I(Inf_{it} > \gamma_0 + \gamma_1 Ind_i), \end{aligned} \tag{26}$$

for $i = 1, 2, ..., 74$ and $t = 1, 2, ..., 11$, where $y_{it} - y_{i,t-1}$ is the average growth rate of real GDP per capita over 5 years, Inf_{it} is the average inflation rate, Ind_i is a dummy variable that equals one if the country is industrialized and zero otherwise, and $\mathbf{x}'_{1it} = (y_{i,t-1}, \underline{\mathbf{x}}'_{1it})$ with $\underline{\mathbf{x}}'_{1it}$ including Ind_i and other determinants of economic growth as listed in Ramírez-Rondán (2020). $y_{i,t-1}$ is also included as an explanatory variable to account for the initial position of the economy. Among the slope parameters, we mainly focus on $\beta_{1,y_{i,t-1}}$ (the coefficient of $y_{i,t-1}$), β_{12} and β_{22}, as in Yu et al. (2022).

In Ramírez-Rondán (2020), three specifications of $\mathbf{x}'_{1it}$ are considered as control variables. The first specification includes only $y_{i,t-1}$ as a regressor, the second also includes another six determinants of economic growth, and the third includes further 10 time dummy variables. Since the third (most general) specification loses too many degrees of freedom, Yu et al. (2022) only consider the second one for illustration. Following Yu et al. (2022), we also only consider the second specification, but further include Ind_i in $\mathbf{x}'_{1it}$ as a control variable. In addition, our $\mathbf{z}_i$ also includes the within-group averages of the six determinants of economic growth and the constant 1 as in Yu et al. (2022). On the other hand, Ramírez-Rondán (2020) also conducts separate analyses for industrialized and nonindustrialized countries, while Yu et al. (2022) only consider the analysis for all countries because the sample size of nonindustrialized countries is too small relative to the number of parameters. We follow Yu et al. (2022), however, our covariate-dependent threshold model can intrinsically discern the disparate impacts of inflation on economic growth for industrialized and nonindustrialized countries. This capability stems from the fact that the threshold in (26) depends on the development level of a country, i.e., γ_0 for nonindustrialized countries and $\gamma_0 + \gamma_1$ for industrialized countries.

Table 5: Empirical application of inflation and economic growth.

	Yu et al.'s (2022) model		PTCDI	
	Estimate	95% CI	Estimate	95% CI
$\beta_{1,y_{i,t-1}}$	-0.0023	[-0.0081, 0.0035] [-0.0120, 0.0075] [-72.494, 72.489]	-0.0020	[-0.0084, 0.0043]
β_{12}	0.256	[0.0335, 0.478] [-0.175, 1.816] [-15.030, 16.971]	0.224	[-0.063, 0.512]
β_{22}	-0.767	[-1.161, -0.372] [-1.457, -0.0762] [-5.416, 3.941]	-0.700	[-1.053, -0.347]
γ_0	15.947	[2.172, 19.432] [1.212, 51.888] [14.128, 16.871]	15.923	[15.737, 16.917]
γ_1			-13.418	[-14.511, -12.927]
	Statistic	p-value	Statistic	p-value
F_1	28.756	0.014	90.471	0.017
F_2	20.367	0.284	33.232	0.337
F_C			9.350	0.019
F_I	55.355	0.000	36.856	0.057

Notes: The three confidence intervals of Yu et al. (2022) are, in order, (LR_n, LR_{1n}, LR_{2n}) for γ_0, and (t, LR_n, LR_{2n}) for $\beta_{1,y_{i,t-1}}, \beta_{12}$ and β_{22}.

Table 5 reports the empirical results, in which 95% confidence intervals for threshold parameters are based on Algorithm A and $p-$values of the tests are computed based on Algorithm B with 1000 bootstrap replications. The confidence intervals for slope parameters are based on cluster-robust standard errors. For comparison, we also replicate and report the empirical results of Yu et al. (2022) in Column 1 of Table 1. First, we determine the number of thresholds according to the test statistics F_1 and F_2. Consistent with Ramírez-Rondán (2020) and Yu et al. (2022), our testing results also support only one threshold in the empirical model, as can be seen from Table 5. Therefore, in Table 5 we only report the slope estimates based on the PTCDI model with one threshold.

Turning now to the estimation and inferences on the threshold parameters, Ramírez-Rondán (2020) finds a threshold inflation of 3.1-3.5% and 16.0-16.1% for industrialized and nonindustrialized countries, respectively, and the estimate for all countries is the same as that for nonindustrialized countries. Yu et al. (2022) obtain a roughly same estimate for the worldwide sample, but do not conduct separate analyses for industrialized and nonindustrialized countries. Our estimate for nonindustrialized countries (γ_0) is close to that of Ramírez-Rondán (2020), and thus similar to the 17% threshold estimated by Kremer et al. (2013). For industrialized countries, our estimate is 2.505% ($\gamma_0+\gamma_1$), which is lower than that of Ramírez-Rondán (2020) but close to the 2.5% threshold estimated by Kremer et al. (2013). Our 95% confidence intervals for threshold parameters are tight, indicating low uncertainty regarding the estimates, and the confidence interval for γ_0 is close to that of Ramírez-Rondán (2020) and the one based on LR_{2n} of Yu et al. (2022).

In the next, we focus on the estimation and inferences on the slope parameters. The estimates and confidence intervals of Yu et al. (2022) are very different from those of Ramírez-Rondán (2020) due to the difference in methodologies. Our estimates are close to those of Yu et al. (2022) and thus also consistent with the economic theory, indicating that rich countries have slower growth rates ($\widehat{\beta}_{1,y_{i,t-1}} < 0$), and high inflation is detrimental to economic growth ($\widehat{\beta}_{12} > 0, \widehat{\beta}_{22} < 0$). Consistent with Yu et al. (2022), our confidence intervals suggest that $\beta_{1,y_{i,t-1}}$ and β_{12} are not significant, and β_{22} is significant.

Finally, we test threshold constancy and unobserved individual-specific threshold effects. According to the F_C test statistic for threshold constancy, the null of a constant threshold is rejected with the $p-$value 0.019. As for the F_I test statistic for unobserved threshold effects, the simulated $p-$value is 0.057. Therefore, the rejection conclusion for no unobserved threshold effects is not as clear as that for threshold constancy, which may be attributed to some elements of $\mathbf{z}_i$ blurring the test, as in the empirical case of investment and financing constraints in Yu et al. (2022). In summary, these empirical results demonstrate the existence of a covariate-dependent inflation threshold, indicating that the inflation threshold for industrialized countries is statistically significantly different from that for nonindustrialized countries. Hence, we conclude that the PTCDI model is a meaningful complement to the models proposed by Ramírez-Rondán's (2020) and Yu et al. (2022) in empirical studies.

7 Conclusion

This article studies the estimation and inferences of panel threshold models with covariate-dependent thresholds and unobserved individual-specific threshold effects. Based on the CRE device and the MCMC technique, we suggest an estimation procedure to estimate the model parameters. We also propose test statistics to test for threshold effect, threshold constancy, and unobserved individual-specific threshold effects. Asymptotic results are established for both the suggested estimator and test statistics. Monte Carlo simulations are conducted and the simulation results show that the suggested estimation, inference and testing methods have desired performance in finite samples. We apply the PTCDI model to revisit two empirical relationships, and both applications demonstrate the usefulness of the proposed model.

Acknowledgements We would like to thank the Editor Prof. Esfandiar Maasoumi, an associate editor and three anonymous referees for very constructive comments and suggestions. We also thank Hujie Bai for research assistance. Any remaining errors are solely our responsibility.

Funding Lixiong Yang acknowledges the financial support from the National Natural Science Foundation of China (Grant No. 72273059).

Conflict of Interest The authors declare that they have no potential conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by the authors.

References

Bai, J., 1997. Estimating multiple breaks one at a time. *Econometric Theory*, 13(3), 315-352.

Bruno, M., Easterly, W., 1998. Inflation crises and long-run growth. *Journal of Monetary Economics*, 41(1), 3-26.

Davies, R.B., 1977. Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, 64, 247-254.

Davies, R.B., 1987. Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, 74, 33-43.

Dornbusch, R., Fischer, S., 1993. Moderate inflation. *The World Bank Economic Review*, 7(1), 1-44.

Dueker, M.J., Psaradakis, Z., Sola, M., 2013. State-dependent threshold smooth transition autoregressive models. *Oxford Bulletin of Economics and Statistics*, 75, 835-854.

Fazzari, S.M., Hubbard R.G., Petersen B.C., 1988. Financing constraints and corporate investment. *Brookings Papers on Economic Activity*, 141-195.

Hansen, B. E., 1996. Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica: Journal of the Econometric Society*, 64, 413-430.

Hansen, B. E., 1999. Threshold effects in non-dynamic panels: Estimation, testing, and inference. *Journal of Econometrics*, 93, 345-68.

Hansen, B.E., 2000. Sample splitting and threshold estimation. *Econometrica*, 68, 575-603.

Hansen, B. E., 2017. Regression kink with an unknown threshold. *Journal of Business and Economic Statistics*, 35, 228-40.

Khan, M., Senhadji, A., 2001. Threshold effects in the relationship between inflation and growth. *IMF Staff Papers*, 48(1), 1-21.

Kremer, S., Bick, A., Nautz, D., 2013. Inflation and growth: New evidence from a dynamic panel threshold analysis. *Empirical Economics*, 44(2), 861-878.

Lee, S., Liao, Y., Seo M. H., Shin Y., 2021. Factor-driven two-regime regression. *Annals of Statistics*, 49(3): 1656-1678.

Mundlak, Y., 1978. On the pooling of time series and cross section data. *Econometrica: Journal of the Econometric Society*, 46, 69-85.

Newey, W. K., McFadden, D., 1994. Large sample estimation and hypothesis testing. *Handbook of Econometrics*, 4, 2111-2245.

Ramírez-Rondán, N.R., 2020, Maximum likelihood estimation of dynamic panel threshold models. *Econometric Reviews*, 39, 260-276.

Seo, M. H., Linton, O., 2007. A smoothed least squares estimator for threshold regression models. *Journal of Econometrics*, 141, 704-735.

Vaona, A., Schiavo, S., 2007. Nonparametric and semiparametric evidence on the long-run effects of inflation on growth. *Economics Letters*, 94(3), 452-458.

Wooldridge, J. M., 2000. A framework for estimating dynamic, unobserved effects panel data models with possible feedback to future explanatory variables. *Economics Letters*, 68(3), 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(1), 39-54.

Yang, L., Su, J.J., 2018. Debt and growth: Is there a constant tipping point?. *Journal of International Money and Finance*, 87, 133-143.

Yang, L., Lee, C., Chen, I-P., 2021a. Threshold model with a time-varying threshold based on Fourier approximation. *Journal of Time Series Analysis*, 42(4): 406-430.

Yang, L., Zhang, C., Lee C., Chen I-P., 2021b. Panel kink threshold regression model with a covariate-dependent threshold. *Econometrics Journal*, 24(3): 462-481.

Yang, L., 2020. State-dependent biases and the quality of China's preliminary GDP announcements. *Empirical Economics*, 59(6): 2663-2687.

Yang, L., 2021. Time-varying threshold cointegration with an application to the Fisher hypothesis. *Studies in Nonlinear Dynamics and Econometrics*, DOI:10.1515/snde-2018-0101, forthcoming.

Yu, P., 2015. Consistency of the least squares estimator in threshold regression with endogeneity. *Economics Letters*, 131, 41-46.

Yu, P., Fan, X., 2021. Threshold regression with a threshold boundary. *Journal of Business and Economic Statistics*, 39(4): 1-59.

Yu, P., Hong, S., Phillips, P. C. B., 2022. Panel threshold regression with unobserved individual-specific threshold effects. *Working Paper*, https://pweb.fbe.hku.hk/~pingyu/WorkingPapers/PTR.pdf.

Zhu Y, Chen H, Lin M., 2019. Threshold models with time-varying threshold values and their application in estimating regime-sensitive Taylor rules. *Studies in Nonlinear Dynamics and Econometrics*, 23, 1-17.

Appendix A: Mathematical proofs

This appendix provides the proofs of Theorems 1-2 in the article. Before proving Theorem 1, we first prove the following Lemmas, which are useful to prove Theorem 1. For notational simplicity, we first clarify the following denotations.

1. $I(q_{it} \leq \mathbf{s}_{it}'\boldsymbol{\gamma}) = I_{it}(\boldsymbol{\gamma})$ and $\check{\mathbf{x}}_{it}(\boldsymbol{\gamma}) = \check{\mathbf{x}}_{it}I_{it}(\boldsymbol{\gamma})$.
2. $\text{sgn}\,(x) = 1(x > 0) - 1(x \leq 0)$ (i.e. the sign function), hence $|x| = x1(x > 0) - x1(x \leq 0) = x[1(x > 0) - 1(x \leq 0)] = x\,\text{sgn}\,(x)$.
3. $a_N = N^{1-2\alpha}$, $\gamma_{10} = \gamma_0$, $\boldsymbol{\gamma}_{\mathbf{s}} = (\gamma_{11}, \ldots, \gamma_{1k})'$,$\boldsymbol{\gamma} = (\gamma_0, \boldsymbol{\gamma}_{\mathbf{s}}')' = (\gamma_{10}, \ldots, \gamma_{1k})'$ and $\boldsymbol{\omega} = (\omega_0, \boldsymbol{\omega}_{\mathbf{s}}')' = (\omega_{10}, \ldots, \omega_{1k})'$.
4. $\nabla I_{it}(\boldsymbol{\gamma}) = I_{it}(\boldsymbol{\gamma}) - I_{it}(\boldsymbol{\gamma}^0)$.
5. $J_N(\boldsymbol{\gamma}) = N^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{c}'\check{\mathbf{x}}_{it}e_{it}I_{it}(\boldsymbol{\gamma})$.
6. $G_N(\boldsymbol{\omega}, h) = (Nh)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{c}'\check{\mathbf{x}}_{it}\check{\mathbf{x}}_{it}'\mathbf{c}|\nabla I_{it}(\boldsymbol{\gamma})|$.
7. $V_{lN}(\boldsymbol{\omega}, h) = (Nh)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{c}'\check{\mathbf{x}}_{it}\check{\mathbf{x}}_{it}'e_{lit}^2\mathbf{c}|I_{lit}(\boldsymbol{\gamma})|$, where $I_{1it}(\boldsymbol{\gamma}) = -I(\mathbf{s}_{it}'\boldsymbol{\gamma} < q_{it} \leq \mathbf{s}_{it}'\boldsymbol{\gamma}^0)$ and $I_{2it}(\boldsymbol{\gamma}) = I(\mathbf{s}_{it}'\boldsymbol{\gamma}^0 < q_{it} \leq \mathbf{s}_{it}'\boldsymbol{\gamma})$.
8. $R_N(\boldsymbol{\omega}, h) = (Nh)^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{c}'\check{\mathbf{x}}_{it}e_{it}\nabla I_{it}(\boldsymbol{\gamma})$.
9. $\boldsymbol{A\theta} = \boldsymbol{\theta}_{\boldsymbol{\delta}} = (\boldsymbol{\theta}_2', \boldsymbol{\delta}')'$, $\boldsymbol{\delta} = \boldsymbol{\theta}_1 - \boldsymbol{\theta}_2$, $\boldsymbol{A\theta}^0 = \boldsymbol{\theta}_{\boldsymbol{\delta}}^0 = (\boldsymbol{\theta}_2^{0\prime}, \boldsymbol{\delta}_0')'$ and $\boldsymbol{\delta}_0 = \boldsymbol{\theta}_1^0 - \boldsymbol{\theta}_2^0$, where

$$\boldsymbol{A} = \begin{bmatrix} \mathbf{0} & \boldsymbol{I} \\ \boldsymbol{I} & -\boldsymbol{I} \end{bmatrix}, \ \boldsymbol{A}^{-1} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{I} \\ \boldsymbol{I} & \mathbf{0} \end{bmatrix}. \tag{A1}$$

Lemma 1. *Under Assumption 1, $E|\phi(x_{it})|^2 < \infty$ and $E\|\mathbf{s}_{it}\| \le C < \infty$, $\mathbf{\Gamma}$ is a compact set, then*

$$\sup_{\gamma\in\mathbf{\Gamma}}\left|\frac{1}{N}\sum_{i=1}^{N}\phi(x_{it})I\left(q_{it}\le \mathbf{s}_{it}'\boldsymbol{\gamma}\right) - E\left(\phi(x_{it})I\left(q_{it}\le \mathbf{s}_{it}'\boldsymbol{\gamma}\right)\right)\right| \xrightarrow{a.s.} 0 \tag{A2}$$

as $N\to\infty$.

Lemma 2. *There is a $C_3<\infty$ such that for $\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2\in\mathbf{\Gamma}$, and $r\le 4$, we have*

$$E\big(\|\check{\mathbf{x}}_{it}\|^r\,|I_{it}(\boldsymbol{\gamma}_2)-I_{it}(\boldsymbol{\gamma}_1)|\big) \le C_3\,\|\boldsymbol{\gamma}_2-\boldsymbol{\gamma}_1\|, \tag{A3}$$

$$E\big(\|\check{\mathbf{x}}_{it}e_{it}\|^r\,|I_{it}(\boldsymbol{\gamma}_2)-I_{it}(\boldsymbol{\gamma}_1)|\big) \le C_3\,\|\boldsymbol{\gamma}_2-\boldsymbol{\gamma}_1\|, \tag{A4}$$

where $I_{it}(\boldsymbol{\gamma}) = I(q_{it}\le \mathbf{s}_{it}'\boldsymbol{\gamma})$.

Lemma 3. *There is a $K<\infty$ such that for $\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2\in\mathbf{\Gamma}$, we have*

$$E\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[b_{it}^2(\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2)-E(b_{it}^2(\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2))\right]\right|^2 \le K\,\|\boldsymbol{\gamma}_2-\boldsymbol{\gamma}_1\|, \tag{A5}$$

$$E\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[h_{it}^2(\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2)-E(h_{it}^2(\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2))\right]\right|^2 \le K\,\|\boldsymbol{\gamma}_2-\boldsymbol{\gamma}_1\|, \tag{A6}$$

where $b_{it}(\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2) = \|\check{\mathbf{x}}_{it}\||I_{it}(\boldsymbol{\gamma}_2)-I_{it}(\boldsymbol{\gamma}_1)|$ and $h_{it}(\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2) = \|\check{\mathbf{x}}_{it}e_{it}\||I_{it}(\boldsymbol{\gamma}_2)-I_{it}(\boldsymbol{\gamma}_1)|$.

Lemma 4. *There are finite constants K_1 and K_2 such that for all $\boldsymbol{\gamma}$, $\varepsilon>0$, $\eta>0$, and $\delta\ge N^{-1}$, if $\sqrt{N}\ge K_2/\eta$, we have*

$$P\left(\sup_{\|\boldsymbol{\gamma}-\boldsymbol{\gamma}_1\|<\delta}\big\|J_N(\boldsymbol{\gamma})-J_N(\boldsymbol{\gamma}_1)\big\|>\eta\right)\le\frac{K_1\delta^2}{\eta^4}. \tag{A7}$$

Lemma 5. *Under Assumption 1, $J_N(\boldsymbol{\gamma})\to_d J(\boldsymbol{\gamma})$, which is a mean-zero Gaussian process with almost surely continuous sample paths.*

proof of Lemma 1-5. The proofs are similar with Hansen (1996, 2000) and Yang et al. (2021a). To save space, the proofs are skipped here and is available from the authors upon request. □

Lemma 6. *Under Assumption 1, $\hat{\boldsymbol{\gamma}}\to_p\boldsymbol{\gamma}^0$.*

Proof. Let $y_i=(y_{i1},\ldots,y_{iT})'$, $\check{\mathbf{x}}_i=(\check{\mathbf{x}}_{i1},\ldots,\check{\mathbf{x}}_{iT})'$, $\check{\mathbf{x}}_i(\boldsymbol{\gamma})=(\check{\mathbf{x}}_{i1}(\boldsymbol{\gamma}),\ldots,\check{\mathbf{x}}_{iT}(\boldsymbol{\gamma}))'$, and $e_i=(e_{i1},\ldots,e_{iT})'$. Let $\mathbf{y}$, $\check{\mathbf{x}}$, $\check{\mathbf{x}}_{\boldsymbol{\gamma}}$ and $\mathbf{e}$ denote the data stacked over all individuals, i.e., $\mathbf{y}=(y_1',\ldots,y_N')'$, $\check{\mathbf{x}}=(\check{\mathbf{x}}_1',\ldots,\check{\mathbf{x}}_N')'$, $\check{\mathbf{x}}_{\boldsymbol{\gamma}}=(\check{\mathbf{x}}_1'(\boldsymbol{\gamma}),\ldots,\check{\mathbf{x}}_N'(\boldsymbol{\gamma}))'$, and $\mathbf{e}=(e_1',\ldots,e_N')'$. Thus, model can be rewritten as $\mathbf{y}=\check{\mathbf{x}}\boldsymbol{\theta}_2+\check{\mathbf{x}}_{\boldsymbol{\gamma}}\boldsymbol{\delta}+\mathbf{e}=\check{\mathbf{X}}_{\boldsymbol{\gamma}}\boldsymbol{\theta}_{\boldsymbol{\delta}}+\mathbf{e}$.

By Lemma 1-5, we have the following results

$$
\begin{aligned}
\frac{1}{N}\check{\mathbf{X}}_{\boldsymbol{\gamma}}'\check{\mathbf{X}}_{\boldsymbol{\gamma}} &= O_p(1),\\
\frac{1}{N}\nabla\check{\mathbf{x}}_{\boldsymbol{\gamma}}'\nabla\check{\mathbf{x}}_{\boldsymbol{\gamma}} &= O_p(1),\\
\frac{1}{N}\check{\mathbf{X}}_{\boldsymbol{\gamma}}'\nabla\check{\mathbf{x}}_{\boldsymbol{\gamma}} &= O_p(1),\\
\frac{1}{\sqrt{N}}\check{\mathbf{X}}_{\boldsymbol{\gamma}}'\mathbf{e} &= O_p(1),\\
\frac{1}{\sqrt{N}}\nabla\check{\mathbf{x}}_{\boldsymbol{\gamma}}'\mathbf{e} &= O_p(1),
\end{aligned}
\tag{A8}
$$

where $\nabla\check{\mathbf{x}}_{\boldsymbol{\gamma}} = \check{\mathbf{x}}_{\boldsymbol{\gamma}} - \check{\mathbf{x}}_0$ and $\check{\mathbf{x}}_0 = \check{\mathbf{x}}_{\boldsymbol{\gamma}^0}$.

First, we derive the convergence rate of $\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma})$. Note that

$$
\begin{aligned}
N^{\alpha}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}) - \boldsymbol{\theta}_{\boldsymbol{\delta}}^0\right) &= \left(\frac{1}{N}\check{\mathbf{X}}_{\boldsymbol{\gamma}}'\check{\mathbf{X}}_{\boldsymbol{\gamma}}\right)^{-1} N^{\alpha-1}\check{\mathbf{X}}_{\boldsymbol{\gamma}}'\left(\nabla\check{\mathbf{x}}_{\boldsymbol{\gamma}}\boldsymbol{\delta}_0 + \mathbf{e}\right)\\
&= \left(\frac{1}{N}\check{\mathbf{X}}_{\boldsymbol{\gamma}}'\check{\mathbf{X}}_{\boldsymbol{\gamma}}\right)^{-1}\left(\frac{1}{N}\check{\mathbf{X}}_{\boldsymbol{\gamma}}'\nabla\check{\mathbf{x}}_{\boldsymbol{\gamma}}\mathbf{c} + N^{\alpha-1}\check{\mathbf{X}}_{\boldsymbol{\gamma}}'\mathbf{e}\right)\\
&= O_p(1)\left(O_p(1) + O_p(N^{\alpha-1/2})\right)\\
&= O_p(1),
\end{aligned}
\tag{A9}
$$

and

$$
\begin{aligned}
N^{\alpha}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0) - \boldsymbol{\theta}_{\boldsymbol{\delta}}^0\right) &= \left(\frac{1}{N}\check{\mathbf{X}}_0'\check{\mathbf{X}}_0\right)^{-1}\left(N^{\alpha-1}\check{\mathbf{X}}_0'\mathbf{e}\right)\\
&= O_p(1)O_p(N^{\alpha-1/2})\\
&= o_p(1),
\end{aligned}
\tag{A10}
$$

where $\check{\mathbf{X}}_0 = \check{\mathbf{X}}_{\boldsymbol{\gamma}^0}$.

Next, we derive the convergence rate of $\widetilde{SSR}_{NT}(\boldsymbol{\gamma})$, which can be written as

$$
\begin{aligned}
N^{2\alpha-1}\left(\widetilde{SSR}_{NT}(\boldsymbol{\gamma}) - \mathbf{e}'\mathbf{e}\right) &= N^{2\alpha-1}\left(\hat{\mathbf{e}}'(\boldsymbol{\gamma})\hat{\mathbf{e}}(\boldsymbol{\gamma}) - \mathbf{e}'\mathbf{e}\right)\\
&= S_1 + S_2 - 2\left(S_3 + S_4 - S_5\right),
\end{aligned}
\tag{A11}
$$

where

$$
\begin{aligned}
S_1 &= \mathbf{c}'\frac{1}{N}\nabla\check{\mathbf{x}}_{\boldsymbol{\gamma}}'\nabla\check{\mathbf{x}}_{\boldsymbol{\gamma}}\mathbf{c} = O_p(1),\\
S_2 &= N^{\alpha}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}) - \boldsymbol{\theta}_{\boldsymbol{\delta}}^0\right)'\frac{1}{N}\check{\mathbf{X}}_{\boldsymbol{\gamma}}'\check{\mathbf{X}}_{\boldsymbol{\gamma}}N^{\alpha}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}) - \boldsymbol{\theta}_{\boldsymbol{\delta}}^0\right)\\
&= O_p(1)O_p(1)O_p(1) = O_p(1)\\
S_3 &= N^{\alpha-1/2}\frac{1}{\sqrt{N}}\mathbf{c}'\nabla\check{\mathbf{x}}_{\boldsymbol{\gamma}}'\mathbf{e} = O(N^{\alpha-1/2})O_p(1) = o_p(1)\\
S_4 &= N^{\alpha-1/2}N^{\alpha}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}) - \boldsymbol{\theta}_{\boldsymbol{\delta}}^0\right)\frac{1}{\sqrt{N}}\check{\mathbf{X}}_{\boldsymbol{\gamma}}'\mathbf{e}\\
&= O(N^{\alpha-1/2})O_p(1)O_p(1) = o_p(1)\\
S_5 &= N^{\alpha}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}) - \boldsymbol{\theta}_{\boldsymbol{\delta}}^0\right)'\frac{1}{N}\check{\mathbf{X}}_{\boldsymbol{\gamma}}'\nabla\check{\mathbf{x}}_{\boldsymbol{\gamma}}\mathbf{c} = O_p(1)O_p(1) = O_p(1).
\end{aligned}
\tag{A12}
$$

Thus, we have the following result:

$$\begin{aligned} S_1 - 2S_5 + S_2 &= \frac{1}{N}\left[\mathbf{c}'\nabla\check{\mathbf{x}}_{\boldsymbol{\gamma}}' - N^{\alpha}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}) - \boldsymbol{\theta}_{\boldsymbol{\delta}}^0\right)'\check{\mathbf{X}}_{\boldsymbol{\gamma}}'\right] \\ &\quad \times\left[\nabla\check{\mathbf{x}}_{\boldsymbol{\gamma}}\mathbf{c} - \check{\mathbf{X}}_{\boldsymbol{\gamma}}N^{\alpha}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}) - \boldsymbol{\theta}_{\boldsymbol{\delta}}^0\right)\right] \\ &\geq 0, \end{aligned} \tag{A13}$$

In addition, when $\boldsymbol{\gamma} = \boldsymbol{\gamma}^0$, we have $S_1 = S_3 = S_5 = 0$ and $S_2 = o_p(1)$. Combining the above results, we obtain

$$N^{2\alpha-1}\left(\widetilde{SSR}_{NT}(\boldsymbol{\gamma}) - \mathbf{e}'\mathbf{e}\right) \xrightarrow{p} b(\boldsymbol{\gamma}) \geq 0. \tag{A14}$$

It is easily seen that $b(\boldsymbol{\gamma}^0) = 0$ when $\boldsymbol{\gamma} = \boldsymbol{\gamma}^0$. Since $\boldsymbol{\gamma}$ minimizes $N^{2\alpha-1}(\widetilde{SSR}_{NT}(\boldsymbol{\gamma}) - \mathbf{e}'\mathbf{e})$, using Theorem 2.1 of Newey and McFadden (1994) we have $\hat{\boldsymbol{\gamma}} \to_p \boldsymbol{\gamma}^0$. This completes the proof of Lemma 6.

□

Lemma 7. *Under Assumption 1, set $\boldsymbol{\gamma} = \boldsymbol{\gamma}^0 + \boldsymbol{\omega}h$, and $h \to 0$ as $N \to \infty$, then*

$$N^{\alpha}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}) - \boldsymbol{\theta}_{\boldsymbol{\delta}}^0\right) \xrightarrow{p} \mathbf{0}. \tag{A15}$$

Proof. Note that, for any $\boldsymbol{\gamma}_1 \in \boldsymbol{\Gamma}$

$$\begin{aligned} \frac{1}{N}\check{\mathbf{X}}_{\boldsymbol{\gamma}_1}'\check{\mathbf{X}}_{\boldsymbol{\gamma}_1} &\xrightarrow{p} \begin{bmatrix} \mathbf{M} & \mathbf{M}(\boldsymbol{\gamma}_1) \\ \mathbf{M}(\boldsymbol{\gamma}_1) & \mathbf{M}(\boldsymbol{\gamma}_1) \end{bmatrix} \equiv \mathscr{S}(\boldsymbol{\gamma}_1), \\ \frac{1}{N}\check{\mathbf{X}}_{\boldsymbol{\gamma}_1}'\nabla\check{\mathbf{x}}_{\boldsymbol{\gamma}_1}\mathbf{c} &\xrightarrow{p} \begin{bmatrix} \mathbf{M}(\boldsymbol{\gamma}_1) - \mathbf{M}(\boldsymbol{\gamma}^0) \\ \mathbf{M}(\boldsymbol{\gamma}_1) - \mathbf{M}(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}^0) \end{bmatrix}\mathbf{c} \equiv \mathscr{A}(\boldsymbol{\gamma}_1), \end{aligned}$$

where $\mathbf{M}(\boldsymbol{\gamma}, \boldsymbol{\gamma}^0) = \sum_{t=1}^{T} E(\check{\mathbf{x}}_{it}\check{\mathbf{x}}_{it}' I(q_{it} \leq \min(\mathbf{s}_{it}'\boldsymbol{\gamma}, \mathbf{s}_{it}'\boldsymbol{\gamma}^0)))$, and therefore,

$$\begin{aligned} N^{\alpha}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}_1) - \boldsymbol{\theta}_{\boldsymbol{\delta}}^0\right) &= \left(\frac{1}{N}\check{\mathbf{X}}_{\boldsymbol{\gamma}_1}'\check{\mathbf{X}}_{\boldsymbol{\gamma}_1}\right)^{-1}\left(\frac{1}{N}\check{\mathbf{X}}_{\boldsymbol{\gamma}_1}'\nabla\check{\mathbf{x}}_{\boldsymbol{\gamma}_1}\mathbf{c} + N^{\alpha-1}\check{\mathbf{X}}_{\boldsymbol{\gamma}_1}'\mathbf{e}\right) \\ &\xrightarrow{p} \mathscr{S}(\boldsymbol{\gamma}_1)^{-1}\mathscr{A}(\boldsymbol{\gamma}_1). \end{aligned}$$

It is easily seen that $\mathscr{A}(\boldsymbol{\gamma}^0) = \mathbf{0}$. Since $\mathscr{S}(\boldsymbol{\gamma})$ and $\mathscr{A}(\boldsymbol{\gamma})$ are continuous, $\mathscr{S}^{-1}(\boldsymbol{\gamma})\mathscr{A}(\boldsymbol{\gamma})$ is continuous in $\boldsymbol{\Gamma}$. By setting $\boldsymbol{\gamma} \to \boldsymbol{\gamma}^0$, we can conclude that

$$N^{\alpha}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}) - \boldsymbol{\theta}_{\boldsymbol{\delta}}^0\right) \xrightarrow{p} \mathscr{S}^{-1}(\boldsymbol{\gamma}^0)\mathscr{A}(\boldsymbol{\gamma}^0) = \mathscr{S}^{-1}(\boldsymbol{\gamma}^0)\mathbf{0} = \mathbf{0}. \tag{A16}$$

□

Lemma 8. *Under Assumption 1, set $\boldsymbol{\gamma} = \boldsymbol{\gamma}^0 + \boldsymbol{\omega}h$, and $h \to 0$ as $N \to \infty$,*

$$\frac{P\left(A_{it}(\boldsymbol{\gamma}) \,|\, \mathbf{s}_{it}\right)}{\mathbf{s}_{it}'\boldsymbol{\gamma} - \mathbf{s}_{it}'\boldsymbol{\gamma}^0} \longrightarrow f_{t|\mathbf{s}}\left(\mathbf{s}_{it}'\boldsymbol{\gamma}^0 \,\middle|\, \mathbf{s}_{it}\right), \tag{A17}$$

$$\frac{P\left(A_{i\tau}(\boldsymbol{\gamma}) \,|\, A_{it}(\boldsymbol{\gamma}), \mathbf{s}_{i\tau}, \mathbf{s}_{it}\right)}{\mathbf{s}_{i\tau}'\boldsymbol{\gamma} - \mathbf{s}_{i\tau}'\boldsymbol{\gamma}^0} \longrightarrow f_{\tau|t,\mathbf{s}}\left(\mathbf{s}_{i\tau}'\boldsymbol{\gamma}^0 \,\middle|\, A_{it}^0, \mathbf{s}_{i\tau}, \mathbf{s}_{it}\right), \tag{A18}$$

$$\frac{\mathbf{M}_{itt}(\boldsymbol{\gamma})}{\mathbf{s}_{it}'\boldsymbol{\gamma}-\mathbf{s}_{it}'\boldsymbol{\gamma}^0} \longrightarrow E\left(\check{\mathbf{x}}_{it}\check{\mathbf{x}}_{it}' \,\middle|\, A_{it}^0, \mathbf{s}_{it}\right) f_{t|\mathbf{s}}\left(\mathbf{s}_{it}'\boldsymbol{\gamma}^0 \,\middle|\, \mathbf{s}_{it}\right) \operatorname{sgn}\left(\nabla I_{it}(\boldsymbol{\gamma})\right), \tag{A19}$$

$$\frac{\mathbf{V}_{litt}(\boldsymbol{\gamma})}{\mathbf{s}_{it}'\boldsymbol{\gamma}-\mathbf{s}_{it}'\boldsymbol{\gamma}^0} \longrightarrow E\left(\check{\mathbf{x}}_{it}\check{\mathbf{x}}_{it}'e_{lit}^2 \,\middle|\, A_{it}^0, \mathbf{s}_{it}\right) f_{t|\mathbf{s}}\left(\mathbf{s}_{it}'\boldsymbol{\gamma}^0 \,\middle|\, \mathbf{s}_{it}\right) \operatorname{sgn}\left(\nabla I_{it}(\boldsymbol{\gamma})\right), \tag{A20}$$

and for $\tau \neq t$

$$\mathbf{M}_{i\tau t}(\boldsymbol{\gamma}) = O_p(h^2), \ \ \mathbf{V}_{li\tau t}(\boldsymbol{\gamma}) = O_p(h^2), \tag{A21}$$

$$E\left(\check{\mathbf{x}}_{i\tau}\check{\mathbf{x}}_{it}'e_{1i\tau}e_{2it}I_{1i\tau}(\boldsymbol{\gamma})I_{2it}(\boldsymbol{\gamma})\right) = O_p(h^2) \tag{A22}$$

as $N \to \infty$*, where* $A_{it}(\boldsymbol{\gamma}) = \{\mathbf{s}_{it}'\boldsymbol{\gamma}^0 < q_{it} \leq \mathbf{s}_{it}'\boldsymbol{\gamma}\}$*,* $A_{it}^0 = \{q_{it} = \mathbf{s}_{it}'\boldsymbol{\gamma}^0\}$*,* $\mathbf{M}_{i\tau t}(\boldsymbol{\gamma}) = E\left(\nabla\check{\mathbf{x}}_{i\tau}(\boldsymbol{\gamma})\nabla\check{\mathbf{x}}_{it}'(\boldsymbol{\gamma}) \,|\, \mathbf{s}_{i\tau}, \mathbf{s}_{it}\right)$*,* $\mathbf{V}_{li\tau t}(\boldsymbol{\gamma}) = E\left(\nabla\check{\mathbf{x}}_{i\tau}(\boldsymbol{\gamma})\nabla\check{\mathbf{x}}_{it}'(\boldsymbol{\gamma})e_{li\tau}e_{lit} \,|\, \mathbf{s}_{i\tau}, \mathbf{s}_{it}\right)$*,* $\nabla\check{\mathbf{x}}_{it}(\boldsymbol{\gamma}) = \check{\mathbf{x}}_{it}(\boldsymbol{\gamma}) - \check{\mathbf{x}}_{it}(\boldsymbol{\gamma}^0)$*, and* $\nabla I_{it}(\boldsymbol{\gamma}) = I_{it}(\boldsymbol{\gamma}) - I_{it}(\boldsymbol{\gamma}^0)$*.*

Proof. First, note that as $N \to \infty$, we have $h \to 0$ and

$$\begin{aligned} \mathbf{s}_{it}'\boldsymbol{\gamma}-\mathbf{s}_{it}'\boldsymbol{\gamma}^0 &= \mathbf{s}_{it}'\left(\boldsymbol{\gamma}-\boldsymbol{\gamma}^0\right) = \mathbf{s}_{it}'\boldsymbol{\omega}h \\ &\longrightarrow 0. \end{aligned}$$

Next, denote $h^* = \mathbf{s}_{it}'\boldsymbol{\gamma}-\mathbf{s}_{it}'\boldsymbol{\gamma}^0$, then we have $h^* \to 0$ as $N \to \infty$ and

$$\begin{aligned} \frac{P\left(A_{it}(\boldsymbol{\gamma}) \,|\, \mathbf{s}_{it}\right)}{\mathbf{s}_{it}'\boldsymbol{\gamma}-\mathbf{s}_{it}'\boldsymbol{\gamma}^0} &= \frac{P\left(q_{it} \leq \mathbf{s}_{it}'\boldsymbol{\gamma} \,|\, \mathbf{s}_{it}\right) - P\left(q_{it} \leq \mathbf{s}_{it}'\boldsymbol{\gamma}^0 \,\middle|\, \mathbf{s}_{it}\right)}{\mathbf{s}_{it}'\boldsymbol{\gamma}-\mathbf{s}_{it}'\boldsymbol{\gamma}^0} \\ &= \frac{P\left(q_{it} \leq \mathbf{s}_{it}'\boldsymbol{\gamma}^0 + h^* \,\middle|\, \mathbf{s}_{it}\right) - P\left(q_{it} \leq \mathbf{s}_{it}'\boldsymbol{\gamma}^0 \,\middle|\, \mathbf{s}_{it}\right)}{h^*}. \end{aligned} \tag{A23}$$

By the definition of a Differential, we have

$$\begin{aligned} \left.\frac{\partial P\left(a \,|\, \mathbf{s}_{it}\right)}{\partial a}\right|_{a=\mathbf{s}_{it}'\boldsymbol{\gamma}^0} &\equiv \lim_{h^* \to 0} \frac{P\left(q_{it} \leq \mathbf{s}_{it}'\boldsymbol{\gamma}^0 + h^* \,\middle|\, \mathbf{s}_{it}\right) - P\left(q_{it} \leq \mathbf{s}_{it}'\boldsymbol{\gamma}^0 \,\middle|\, \mathbf{s}_{it}\right)}{h^*} \\ &= f_{t|\mathbf{s}}\left(\mathbf{s}_{it}'\boldsymbol{\gamma}^0 \,\middle|\, \mathbf{s}_{it}\right). \end{aligned} \tag{A24}$$

A similar argument can prove (A18). Therefore, for $\tau \neq t$, we have

$$\begin{aligned} h^{-2}\mathbf{M}_{i\tau t}(\boldsymbol{\gamma}) &= E\left(\check{\mathbf{x}}_{i\tau}\check{\mathbf{x}}_{it}' \,\middle|\, A_{i\tau}(\boldsymbol{\gamma}), A_{it}(\boldsymbol{\gamma}), \mathbf{s}_{i\tau}, \mathbf{s}_{it}\right) \\ &\quad \times \frac{P\left(A_{it}(\boldsymbol{\gamma}) \,|\, \mathbf{s}_{i\tau}, \mathbf{s}_{it}\right)}{\mathbf{s}_{it}'\boldsymbol{\gamma}-\mathbf{s}_{it}'\boldsymbol{\gamma}^0} \times \mathbf{s}_{it}'\boldsymbol{\omega} \\ &\quad \times \frac{P\left(A_{i\tau}(\boldsymbol{\gamma}) \,|\, A_{it}(\boldsymbol{\gamma}), \mathbf{s}_{i\tau}, \mathbf{s}_{it}\right)}{\mathbf{s}_{i\tau}'\boldsymbol{\gamma}-\mathbf{s}_{i\tau}'\boldsymbol{\gamma}^0} \times \mathbf{s}_{i\tau}'\boldsymbol{\omega} \\ &\longrightarrow E\left(\check{\mathbf{x}}_{i\tau}\check{\mathbf{x}}_{it}' \,\middle|\, A_{i\tau}^0, A_{it}^0, \mathbf{s}_{i\tau}, \mathbf{s}_{it}\right) f_{t|\mathbf{s}}\left(\mathbf{s}_{it}'\boldsymbol{\gamma}^0 \,\middle|\, \mathbf{s}_{it}\right) \times \mathbf{s}_{it}'\boldsymbol{\omega} \\ &\quad \times f_{\tau|t,\mathbf{s}}\left(\mathbf{s}_{i\tau}'\boldsymbol{\gamma}^0 \,\middle|\, A_{it}^0, \mathbf{s}_{i\tau}, \mathbf{s}_{it}\right) \times \mathbf{s}_{i\tau}'\boldsymbol{\omega} \end{aligned} \tag{A25}$$

, and for $\tau = t$

$$\begin{aligned} \frac{\mathbf{M}_{itt}(\boldsymbol{\gamma})}{\mathbf{s}_{it}'\boldsymbol{\gamma}-\mathbf{s}_{it}'\boldsymbol{\gamma}^0} &= E\left(\check{\mathbf{x}}_{it}\check{\mathbf{x}}_{it}' \,\middle|\, A_{it}(\boldsymbol{\gamma}), \mathbf{s}_{it}\right) \frac{P\left(A_{it}(\boldsymbol{\gamma}) \,|\, \mathbf{s}_{it}\right)}{\mathbf{s}_{it}'\boldsymbol{\gamma}-\mathbf{s}_{it}'\boldsymbol{\gamma}^0} \operatorname{sgn}\left(\nabla I_{it}(\boldsymbol{\gamma})\right) \\ &\longrightarrow E\left(\check{\mathbf{x}}_{it}\check{\mathbf{x}}_{it}' \,\middle|\, A_{it}^0, \mathbf{s}_{it}\right) f_{t|\mathbf{s}}\left(\mathbf{s}_{it}(\boldsymbol{\gamma}^0) \,\middle|\, \mathbf{s}_{it}\right) \operatorname{sgn}\left(\nabla I_{it}(\boldsymbol{\gamma})\right). \end{aligned} \tag{A26}$$

A similar argument can prove (A20), (A21) and (A22). This completes the proof of Lemma 8.

□

Lemma 9. *Under Assumption 1, on a compact set* $\mathbf{\Psi} = \times_{n=0}^{d} \Psi_n$*, we have the following uniform convergence results, set* $\boldsymbol{\gamma} = \boldsymbol{\gamma}^0 + \boldsymbol{\omega} h$ *such that* $h \to 0$ *and* $Nh \to \infty$ *as* $N \to \infty$*, than*

$$
\begin{aligned}
G_N(\boldsymbol{\omega}, h) &\xrightarrow{p} G_T(\boldsymbol{\omega}), &\text{(A27)}\\
V_{lN}(\boldsymbol{\omega}, h) &\xrightarrow{p} V_{lT}(\boldsymbol{\omega}), &\text{(A28)}
\end{aligned}
$$

where

$$
\begin{aligned}
G_T(\boldsymbol{\omega}) &= \sum_{t=1}^{T} \mathbf{c}' E\Big(\mathbf{D}_t\left(\boldsymbol{\gamma}^0 \,\middle|\, \mathbf{s}_{it}\right) f_{t|\mathbf{s}}\left(\mathbf{s}_{it}'\boldsymbol{\gamma}^0 \,\middle|\, \mathbf{s}_{it}\right) \left|\mathbf{s}_{it}'\boldsymbol{\omega}\right| \Big)\mathbf{c},\\
V_{1T}(\boldsymbol{\omega}) &= \sum_{t=1}^{T} \mathbf{c}' E\Big(\mathbf{V}_{1t}\left(\boldsymbol{\gamma}^0 \,\middle|\, \mathbf{s}_{it}\right) f_{t|\mathbf{s}}\left(\mathbf{s}_{it}'\boldsymbol{\gamma}^0 \,\middle|\, \mathbf{s}_{it}\right) \left|\mathbf{s}_{it}'\boldsymbol{\omega}\right| I(\mathbf{s}_{it}'\boldsymbol{\omega} \le 0)\Big)\mathbf{c},\\
V_{2T}(\boldsymbol{\omega}) &= \sum_{t=1}^{T} \mathbf{c}' E\Big(\mathbf{V}_{2t}\left(\boldsymbol{\gamma}^0 \,\middle|\, \mathbf{s}_{it}\right) f_{t|\mathbf{s}}\left(\mathbf{s}_{it}'\boldsymbol{\gamma}^0 \,\middle|\, \mathbf{s}_{it}\right) \left|\mathbf{s}_{it}'\boldsymbol{\omega}\right| I(\mathbf{s}_{it}'\boldsymbol{\omega} > 0)\Big)\mathbf{c},
\end{aligned}
$$

in which

$$
\begin{aligned}
\mathbf{D}_t\left(\boldsymbol{\gamma}^0 \,\middle|\, \mathbf{s}_{it}\right) &= E\left(\check{\mathbf{x}}_{it}\check{\mathbf{x}}_{it}' \,\middle|\, q_{it} = \mathbf{s}_{it}'\boldsymbol{\gamma}, \mathbf{s}_{it}\right).\\
\mathbf{V}_{lt}\left(\boldsymbol{\gamma}^0 \,\middle|\, \mathbf{s}_{it}\right) &= E\left(\check{\mathbf{x}}_{it}\check{\mathbf{x}}_{it}' e_{lit}^2 \,\middle|\, q_{it} = \mathbf{s}_{it}'\boldsymbol{\gamma}, \mathbf{s}_{it}\right).
\end{aligned}
$$

Proof. Note that

$$
\begin{aligned}
\{q_{it} \le \mathbf{s}_{it}'\boldsymbol{\gamma}\} \le \{q_{it} \le \mathbf{s}_{it}'\boldsymbol{\gamma}^0\} &\Leftrightarrow \mathbf{s}_{it}'\boldsymbol{\gamma} \le \mathbf{s}_{it}'\boldsymbol{\gamma}^0 &\text{(A29)}\\
\{q_{it} \le \mathbf{s}_{it}'\boldsymbol{\gamma}\} > \{q_{it} \le \mathbf{s}_{it}'\boldsymbol{\gamma}^0\} &\Leftrightarrow \mathbf{s}_{it}'\boldsymbol{\gamma} > \mathbf{s}_{it}'\boldsymbol{\gamma}^0, &\text{(A30)}
\end{aligned}
$$

(A29) and (A30) imply that

$$
\operatorname{sgn}\left(I_{it}(\boldsymbol{\gamma}) - I_{it}(\boldsymbol{\gamma}^0)\right) = \operatorname{sgn}\left(\mathbf{s}_{it}'\boldsymbol{\gamma} - \mathbf{s}_{it}'\boldsymbol{\gamma}^0\right), \qquad \text{(A31)}
$$

and therefore,

$$
\begin{aligned}
\left[\mathbf{s}_{it}'\boldsymbol{\gamma} - \mathbf{s}_{it}'\boldsymbol{\gamma}^0\right] \operatorname{sgn}\left(I_{it}(\boldsymbol{\gamma}) - I_{it}(\boldsymbol{\gamma}^0)\right) &= \left[\mathbf{s}_{it}'\boldsymbol{\gamma} - \mathbf{s}_{it}'\boldsymbol{\gamma}^0\right] \operatorname{sgn}\left(\mathbf{s}_{it}'\boldsymbol{\gamma} - \mathbf{s}_{it}'\boldsymbol{\gamma}^0\right)\\
&= \left|\mathbf{s}_{it}'\boldsymbol{\gamma} - \mathbf{s}_{it}'\boldsymbol{\gamma}^0\right|. \qquad \text{(A32)}
\end{aligned}
$$

Denote $G_{Nt}(\boldsymbol{\omega}, h) = (Nh)^{-1} \sum_{i=1}^{N} \mathbf{c}'\check{\mathbf{x}}_{it}\check{\mathbf{x}}_{it}'\mathbf{c} \left|I_{it}(\boldsymbol{\gamma}) - I_{it}(\boldsymbol{\gamma}^0)\right|$. By Lemma 8 and (A32), we can obtain

$$\begin{aligned}
&E\left(G_{Nt}(\boldsymbol{\omega},h)\,|\,\mathbf{s}_{it}\right)\\
&\quad= E\left(\frac{1}{Nh}\sum_{i=1}^{N}\mathbf{c}'\check{\mathbf{x}}_{it}\check{\mathbf{x}}_{it}'\mathbf{c}\left|I_{it}(\boldsymbol{\gamma})-I_{it}(\boldsymbol{\gamma}^0)\right|\,\middle|\,\mathbf{s}_{it}\right)\\
&\quad= \frac{1}{h}\mathbf{c}'\mathbf{M}_{itt}(\boldsymbol{\gamma})\mathbf{c}\\
&\quad= \frac{1}{h}\mathbf{c}'\frac{\mathbf{M}_{itt}(\boldsymbol{\gamma})}{\mathbf{s}_{it}'\boldsymbol{\gamma}-\mathbf{s}_{it}'\boldsymbol{\gamma}^0}\mathbf{c}\left(\mathbf{s}_{it}'\boldsymbol{\gamma}-\mathbf{s}_{it}'\boldsymbol{\gamma}^0\right)\\
&\quad= \mathbf{c}'\frac{\mathbf{M}_{itt}(\boldsymbol{\gamma})}{\mathbf{s}_{it}'\boldsymbol{\gamma}-\mathbf{s}_{it}'\boldsymbol{\gamma}^0}\mathbf{c}\left(\mathbf{s}_{it}'\boldsymbol{\omega}\right)\\
&\quad\longrightarrow \mathbf{c}'E\left(\check{\mathbf{x}}_{it}\check{\mathbf{x}}_{it}'\,\middle|\,q_{it}=\mathbf{s}_{it}'\boldsymbol{\gamma}^0,\mathbf{s}_{it}\right)f_{t|\mathbf{s}}\left(\mathbf{s}_{it}'\boldsymbol{\gamma}^0\,\middle|\,\mathbf{s}_{it}\right)\mathbf{c}\,\mathrm{sgn}\left(\nabla I_{it}(\boldsymbol{\gamma})\right)\left(\mathbf{s}_{it}'\boldsymbol{\omega}\right)\\
&\quad= \mathbf{c}'\mathbf{D}_t\left(\boldsymbol{\gamma}^0\,\middle|\,\mathbf{s}_{it}\right)f_{t|\mathbf{s}}\left(\mathbf{s}_{it}'\boldsymbol{\gamma}^0\,\middle|\,\mathbf{s}_{it}\right)\mathbf{c}\left|\mathbf{s}_{it}'\boldsymbol{\omega}\right|, \qquad \text{(A33)}
\end{aligned}$$

in which we use the following result

$$\begin{aligned}
h^{-1}\left[\mathbf{s}_{it}'\boldsymbol{\gamma}-\mathbf{s}_{it}'\boldsymbol{\gamma}^0\right] &= h^{-1}\mathbf{s}_{it}'\left(\boldsymbol{\gamma}-\boldsymbol{\gamma}^0\right)\\
&= h^{-1}\mathbf{s}_{it}'\boldsymbol{\omega}h=\mathbf{s}_{it}'\boldsymbol{\omega}. \qquad \text{(A34)}
\end{aligned}$$

Taking the expectation over both sides of (A33), we have

$$\begin{aligned}
E\left(G_{Nt}(\boldsymbol{\omega},h)\right) &= E\left(E\left(G_{Nt}(\boldsymbol{\omega},h)\,|\,\mathbf{s}_{it}\right)\right)\\
&\longrightarrow \mathbf{c}'E\Big(\mathbf{D}_t\left(\boldsymbol{\gamma}^0\,\middle|\,\mathbf{s}_{it}\right)f\left(\mathbf{s}_{it}'\boldsymbol{\gamma}^0\,\middle|\,\mathbf{s}_{it}\right)\left|\mathbf{s}_{it0}'\boldsymbol{\omega}\right|\Big)\mathbf{c}. \qquad \text{(A35)}
\end{aligned}$$

Next, note that we have the following result

$$\begin{aligned}
&E\big\|G_{Nt}(\boldsymbol{\omega},h)-E(G_{Nt}(\boldsymbol{\omega},h))\big\|^2\\
&\quad\leq \|\mathbf{c}\|^2 E\left|\frac{1}{Nh}\sum_{i=1}^{N}\|\check{\mathbf{x}}_{it}\|^2\left|\nabla I_{it}(\boldsymbol{\gamma})\right|-E\Big(\|\check{\mathbf{x}}_{it}\|^2\left|\nabla I_{it}(\boldsymbol{\gamma})\right|\Big)\right|^2\\
&\quad= \frac{1}{Nh^2}\|\mathbf{c}\|^2 E\left|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}b_{it}^2(\boldsymbol{\gamma}^0,\boldsymbol{\gamma})-E(b_{it}^2(\boldsymbol{\gamma}^0,\boldsymbol{\gamma}))\right|^2\\
&\quad\leq \frac{1}{Nh^2}\|\mathbf{c}\|^2 K\left\|\boldsymbol{\gamma}-\boldsymbol{\gamma}^0\right\|=\frac{1}{Nh^2}\|\mathbf{c}\|^2K\|\boldsymbol{\omega}h\|\\
&\quad= \frac{1}{Nh}\|\mathbf{c}\|^2K\|\boldsymbol{\omega}\|\\
&\quad\longrightarrow 0. \qquad \text{(A36)}
\end{aligned}$$

The inequality is based on Lemma 3. Furthermore, by Markov's inequality, we have

$$\begin{aligned}
&P\left(\big\|G_{Nt}(\boldsymbol{\omega},h)-E(G_{Nt}(\boldsymbol{\omega},h))\big\|>\varepsilon\right)\\
&\quad\leq \frac{E\big\|G_{Nt}(\boldsymbol{\omega},h)-E(G_{Nt}(\boldsymbol{\omega},h))\big\|^2}{\varepsilon^2}\longrightarrow 0, \qquad \text{(A37)}
\end{aligned}$$

and

$$
\begin{aligned}
& P\left(\left\|G_N(\boldsymbol{\omega},h)-E(G_N(\boldsymbol{\omega},h))\right\|>\varepsilon\right) \\
& \quad \le \sum_{t=1}^{T} P\left(\left\|G_{Nt}(\boldsymbol{\omega},h)-E(G_{Nt}(\boldsymbol{\omega},h))\right\|>\varepsilon/T\right)\longrightarrow 0. \qquad \text{(A38)}
\end{aligned}
$$

Finally, we have

$$
G_N(\boldsymbol{\omega},h) \xrightarrow{p} \sum_{t=1}^{T}\mathbf{c}'E\Big(\mathbf{D}_t\left(\boldsymbol{\gamma}^0 \mid \mathbf{s}_{it}\right) f_{t|\mathbf{s}}\left(\mathbf{s}_{it}'\boldsymbol{\gamma}^0 \mid \mathbf{s}_{it}\right)\left|\mathbf{s}_{it}'\boldsymbol{\omega}\right|\Big)\mathbf{c}. \qquad \text{(A39)}
$$

The proof for (A28) is similar with (A27). This completes the proof of Lemma 9.

□

Lemma 10. *Under Assumption 1, on any compact set* $\boldsymbol{\Psi}=\times_{n=0}^{d}\Psi_n$*, set* $\boldsymbol{\gamma}=\boldsymbol{\gamma}^0+\boldsymbol{\omega}h$ *such that* $h\to 0$ *and* $Th\to\infty$ *as* $T\to\infty$*, then*

$$
R_N(\boldsymbol{\omega},h)\xrightarrow{d} R_T(\boldsymbol{\omega}), \qquad \text{(A40)}
$$

where $R_T(\boldsymbol{\omega})=R_{1T}(\boldsymbol{\omega})+R_{2T}(\boldsymbol{\omega})$*,* $R_{lT}(\boldsymbol{\omega})$ *is a Gaussian process with a positive variance* $V_{lT}(\boldsymbol{\omega})$ *when* $\boldsymbol{\omega}\neq\mathbf{0}$*, which* $R_{1T}(\boldsymbol{\omega})$ *and* $R_{2T}(\boldsymbol{\omega})$ *are independent.*

Proof. Denote

$$
\begin{aligned}
R_{1N}(\boldsymbol{\omega},h) &= \sqrt{\frac{1}{Nh}}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{c}'\check{\mathbf{x}}_{it}e_{1it}I_{1it}(\boldsymbol{\gamma}), \\
R_{2N}(\boldsymbol{\omega},h) &= \sqrt{\frac{1}{Nh}}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{c}'\check{\mathbf{x}}_{it}e_{2it}I_{2it}(\boldsymbol{\gamma}),
\end{aligned}
$$

where $I_{1it}(\boldsymbol{\gamma})=-I(\mathbf{s}_{it}'\boldsymbol{\gamma}<q_{it}\le\mathbf{s}_{it}'\boldsymbol{\gamma}^0)$ and $I_{2it}(\boldsymbol{\gamma})=I(\mathbf{s}_{it}'\boldsymbol{\gamma}^0<q_{it}\le\mathbf{s}_{it}'\boldsymbol{\gamma})$. According to the definition of e_{it}, we have the following result

$$
e_{it}\left(I_{it}(\boldsymbol{\gamma})-I_{it}(\boldsymbol{\gamma}^0)\right)=\begin{cases} e_{1it}I_{1it}(\boldsymbol{\gamma}) & \text{, if } \mathbf{s}_{it}'\boldsymbol{\gamma}\le\mathbf{s}_{it}'\boldsymbol{\gamma}^0 \\ e_{2it}I_{2it}(\boldsymbol{\gamma}) & \text{, if } \mathbf{s}_{it}'\boldsymbol{\gamma}>\mathbf{s}_{it}'\boldsymbol{\gamma}^0 \end{cases}, \qquad \text{(A41)}
$$

and hence $R_N(\boldsymbol{\omega},h)=R_{1N}(\boldsymbol{\omega},h)+R_{2N}(\boldsymbol{\omega},h)$.

First, we prove that $E(R_{lN}^2(\boldsymbol{\omega})) \to_p V_{lT}(\boldsymbol{\omega})$. Note that by Lemma 8, we have

$$\begin{aligned}
E(R_{lN}^2(\boldsymbol{\omega},h)) &= \frac{1}{Nh}E\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{c}'\check{\mathbf{x}}_{it}e_{lit}I_{lit}(\boldsymbol{\gamma})\right)^2 \\
&= \frac{1}{Nh}\sum_{i=1}^{N}E\left(\sum_{t=1}^{T}\mathbf{c}'\check{\mathbf{x}}_{it}e_{lit}I_{lit}(\boldsymbol{\gamma})\right)^2 \\
&= \frac{1}{h}\left(\mathbf{c}'\sum_{t=1}^{T}E\left(\check{\mathbf{x}}_{it}\check{\mathbf{x}}_{it}'e_{lit}^2|I_{lit}(\boldsymbol{\gamma})|\right)\mathbf{c}\right. \\
&\quad \left. + \mathbf{c}'\sum_{\tau\neq t}E\left(\check{\mathbf{x}}_{i\tau}\check{\mathbf{x}}_{it}'e_{li\tau}e_{lit}I_{li\tau}(\boldsymbol{\gamma})I_{lit}(\boldsymbol{\gamma})\right)\mathbf{c}\right) \\
&= \frac{1}{h}\sum_{t=1}^{T}\mathbf{c}'\mathbf{V}_{litt}(\boldsymbol{\gamma})\mathbf{c} + \frac{1}{h}\sum_{\tau\neq t}\mathbf{c}'\mathbf{V}_{lit\tau}(\boldsymbol{\gamma})\mathbf{c} \\
&\longrightarrow V_{lT}(\boldsymbol{\omega}) < \infty, \qquad \text{(A42)}
\end{aligned}$$

and

$$\begin{aligned}
&E(R_{1N}(\boldsymbol{\omega},h)R_{2N}(\boldsymbol{\omega},h)) \\
&= \frac{1}{Nh}E\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{c}'\check{\mathbf{x}}_{it}e_{1it}I_{1it}(\boldsymbol{\gamma})\right)E\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{c}'\check{\mathbf{x}}_{it}e_{2it}I_{2it}(\boldsymbol{\gamma})\right) \\
&= \frac{1}{Nh}\sum_{i=1}^{N}E\left(\sum_{t=1}^{T}\mathbf{c}'\check{\mathbf{x}}_{it}e_{1it}I_{1it}(\boldsymbol{\gamma})\right)\left(\sum_{t=1}^{T}\mathbf{c}'\check{\mathbf{x}}_{it}e_{2it}I_{2it}(\boldsymbol{\gamma})\right) \\
&= \frac{1}{h}\sum_{\tau\neq t}E\left(\mathbf{c}'\check{\mathbf{x}}_{i\tau}\check{\mathbf{x}}_{it}'\mathbf{c}e_{1i\tau}e_{2it}I_{1i\tau}(\boldsymbol{\gamma})I_{2it}(\boldsymbol{\gamma})\right) \\
&\longrightarrow 0. \qquad \text{(A43)}
\end{aligned}$$

Under Assumption 1, $\nabla\check{\mathbf{x}}_{it}e_{it}$ is an i.i.d. random sequence across i, so $R_N(\boldsymbol{\omega})$ converges pointwise to a Gaussian distribution with variance $V_{1T}(\boldsymbol{\omega}) + V_{2T}(\boldsymbol{\omega})$ by the CLT.

Next, given lemma 4, the proof of the tightness of $R_N(\boldsymbol{\omega},h)$ is same as Lemma A.11 in Hansen (2000). This completes the proof of Lemma 10.

□

Lemma 11. *Under Assumption 1, we have* $a_N(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^0) = O_p(1)$.

Proof. To prove Lemma 11, we need to prove that, for some $\bar{\omega} > 0$, we have

$$\lim_{T\to\infty} P\left(\|\boldsymbol{\gamma} - \boldsymbol{\gamma}^0\| \leq \frac{\bar{\omega}}{a_N}\right) = 1. \qquad \text{(A44)}$$

For any $B > 0$, define $V_B = \{\boldsymbol{\gamma} : \|\boldsymbol{\gamma} - \boldsymbol{\gamma}^0\| \leq B\}$. Then, when the sample size N is large enough, we have $\bar{\omega}/a_N \leq B$. By Lemma 6, we have $\hat{\boldsymbol{\gamma}} \to_p \boldsymbol{\gamma}^0$, and hence $\lim_{N\to\infty} P(\hat{\boldsymbol{\gamma}} \in V_B) = 1$. Therefore, we only need to exmine the limiting behavior in V_B.

Define a subset of $V_B : V_B' = \{\boldsymbol{\gamma} : \bar{\omega}/a_N < \|\boldsymbol{\gamma} - \boldsymbol{\gamma}^0\| \leq B\}$. To prove (A44), we just need to prove $\lim_{N\to\infty} P(\hat{\boldsymbol{\gamma}} \in V_B') = 0$.

Let $\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}} = (\hat{\boldsymbol{\theta}}_2', \hat{\boldsymbol{\delta}}')'$ be the estimation of $y_{it} = \check{\mathbf{x}}_{it}\boldsymbol{\theta}_2 + \check{\mathbf{x}}_{it}(\boldsymbol{\gamma})\boldsymbol{\delta} + e_{it}$. Denote $\widetilde{SSR}^*_{NT}(\boldsymbol{\gamma}) = \widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\hat{\gamma}), \boldsymbol{\gamma})$ and $\widetilde{SSR}^*_{NT}(\boldsymbol{\gamma}^0) = \widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\hat{\gamma}), \boldsymbol{\gamma}^0)$, where $\widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\hat{\gamma}), \boldsymbol{\gamma}) = \|\hat{\mathbf{e}}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\hat{\gamma}), \boldsymbol{\gamma})\|^2$. From the estimation procedure of $\hat{\boldsymbol{\gamma}}$, we have $\widetilde{SSR}^*_{NT}(\hat{\boldsymbol{\gamma}}) \leq \widetilde{SSR}^*_{NT}(\boldsymbol{\gamma}^0)$. Thus it suffices to prove that for any $\boldsymbol{\gamma} \in V_B'$,

$$\lim_{N\to\infty} P\left(\widetilde{SSR}^*_{NT}(\boldsymbol{\gamma}) - \widetilde{SSR}^*_{NT}(\boldsymbol{\gamma}^0) > 0\right) = 1. \tag{A45}$$

Note that, the equation (A45) is equivalent to prove

$$\lim_{N\to\infty} P\left(\frac{\widetilde{SSR}^*_{NT}(\boldsymbol{\gamma}) - \widetilde{SSR}^*_{NT}(\boldsymbol{\gamma}^0)}{a_N\|\boldsymbol{\gamma} - \boldsymbol{\gamma}^0\|} > 0\right) = 1. \tag{A46}$$

Since the true model can be rewritten as $\mathbf{y} = \check{\mathbf{X}}_0\boldsymbol{\theta}^0_{\boldsymbol{\delta}} + \mathbf{e}$, thus we have

$$\begin{aligned}
\hat{\mathbf{e}}(\hat{\boldsymbol{\phi}}(\hat{\gamma}), \boldsymbol{\gamma}) &= \mathbf{y} - \check{\mathbf{X}}_{\boldsymbol{\gamma}}\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}} \\
&= \check{\mathbf{X}}_0\boldsymbol{\theta}^0_{\boldsymbol{\delta}} + \mathbf{e} - \check{\mathbf{X}}_{\boldsymbol{\gamma}}\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}} \\
&= \mathbf{e} - \nabla\check{\mathbf{x}}_{\boldsymbol{\gamma}}\boldsymbol{\delta}_0 - \check{\mathbf{X}}_{\boldsymbol{\gamma}}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}} - \boldsymbol{\theta}^0_{\boldsymbol{\delta}}).
\end{aligned} \tag{A47}$$

Therefore,

$$\begin{aligned}
\widetilde{SSR}^*_{NT}(\boldsymbol{\gamma}) - \widetilde{SSR}^*_{NT}(\boldsymbol{\gamma}^0) &= \left\|\hat{\mathbf{e}}(\hat{\boldsymbol{\theta}}(\hat{\gamma}), \boldsymbol{\gamma})\right\|^2 - \left\|\hat{\mathbf{e}}(\hat{\boldsymbol{\theta}}(\hat{\gamma}), \boldsymbol{\gamma}^0)\right\|^2 \\
&= S_1 + S_2 - 2(S_3 + S_4 - S_5),
\end{aligned} \tag{A48}$$

where

$$\begin{aligned}
S_1 &= T^{-2\alpha}\mathbf{c}'\nabla\check{\mathbf{x}}_{\boldsymbol{\gamma}}'\nabla\check{\mathbf{x}}_{\boldsymbol{\gamma}}\mathbf{c} \\
S_2 &= \left(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}} - \boldsymbol{\theta}^0_{\boldsymbol{\delta}}\right)'\left(\check{\mathbf{X}}_{\boldsymbol{\gamma}}'\check{\mathbf{X}}_{\boldsymbol{\gamma}} - \check{\mathbf{X}}_0'\check{\mathbf{X}}_0\right)\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}} - \boldsymbol{\theta}^0_{\boldsymbol{\delta}}\right) \\
S_3 &= T^{-\alpha}\mathbf{c}'\nabla\check{\mathbf{x}}_{\boldsymbol{\gamma}}'\mathbf{e} \\
S_4 &= \left(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}} - \boldsymbol{\theta}^0_{\boldsymbol{\delta}}\right)'\nabla\check{\mathbf{X}}_{\boldsymbol{\gamma}}'\mathbf{e} \\
S_5 &= T^{-\alpha}\mathbf{c}'\nabla\check{\mathbf{x}}_{\boldsymbol{\gamma}}'\check{\mathbf{X}}_{\boldsymbol{\gamma}}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}} - \boldsymbol{\theta}^0_{\boldsymbol{\delta}}\right).
\end{aligned}$$

Let $\boldsymbol{\gamma} = \boldsymbol{\gamma}^0 + \boldsymbol{\omega}h$, $h = N^{2\delta-1}$, where $\alpha < \delta < 2$, then we have $\lim_{N\to\infty} P(\boldsymbol{\gamma} \in V_B') = 1$.

Using Lemma 9 and Assumption 1 ($\boldsymbol{\theta}_1^0 - \boldsymbol{\theta}_2^0 = \mathbf{c}N^{-\alpha}$), we can show

$$\begin{aligned}
\frac{S_1}{a_N\|\boldsymbol{\gamma}-\boldsymbol{\gamma}^0\|} &= \|\boldsymbol{\omega}\|^{-1}G_N(\boldsymbol{\omega},h) = O_p(1) > 0 && \text{(A49)}\\
\frac{S_2}{a_N\|\boldsymbol{\gamma}-\boldsymbol{\gamma}^0\|} &= \left(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}} - \boldsymbol{\theta}_{\boldsymbol{\delta}}^0\right)' \left(\frac{\check{\mathbf{X}}_{\boldsymbol{\gamma}}'\check{\mathbf{X}}_{\boldsymbol{\gamma}} - \check{\mathbf{X}}_0'\check{\mathbf{X}}_0}{\|\boldsymbol{\omega}\|Nh}\right)\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}} - \boldsymbol{\theta}_{\boldsymbol{\delta}}^0\right) = o_p(1)O_p(1)o_p(1)\\
&= o_p(1), && \text{(A50)}\\
\frac{|S_3|}{a_N\|\boldsymbol{\gamma}-\boldsymbol{\gamma}^0\|} &= \left(a_N\left\|\boldsymbol{\gamma}-\boldsymbol{\gamma}^0\right\|\right)^{-1/2} \times \left|\|\boldsymbol{\omega}\|^{-1}R_N(\boldsymbol{\omega},h)\right|\\
&\lesssim \bar{\omega}^{-1/2} \times \left|\|\boldsymbol{\omega}\|^{-1}R_N(\boldsymbol{\omega},h)\right| = O(\bar{\omega}^{-1/2})O_p(1), && \text{(A51)}\\
\frac{|S_4|}{a_N\|\boldsymbol{\gamma}-\boldsymbol{\gamma}^0\|} &= \left(a_N\left\|\boldsymbol{\gamma}-\boldsymbol{\gamma}^0\right\|\right)^{-1/2} \times \left|\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}} - \boldsymbol{\theta}_{\boldsymbol{\delta}}^0\right)' \frac{1}{\sqrt{\|\boldsymbol{\omega}\|Nh}}\nabla\check{\mathbf{X}}_{\boldsymbol{\gamma}}'\mathbf{e}\right|\\
&\lesssim O(\bar{\omega}^{-1/2})o_p(1)O_p(1) = o_p(1), && \text{(A52)}\\
\frac{S_5}{a_N\|\boldsymbol{\gamma}-\boldsymbol{\gamma}^0\|} &= \frac{1}{\|\boldsymbol{\omega}\|Nh}\mathbf{c}'\nabla\check{\mathbf{x}}_{\boldsymbol{\gamma}}'\check{\mathbf{X}}_{\boldsymbol{\gamma}}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}} - \boldsymbol{\theta}_{\boldsymbol{\delta}}^0\right) = O_p(1)o_p(1) = o_p(1), && \text{(A53)}
\end{aligned}$$

where we denote $A_N \lesssim B_N$ (Asymptotic inequality), if $A_N \leq CB_N$ holds for all sufficiently large absolute constant N.

Hence, we can show that for any $\boldsymbol{\gamma} \in V_B'$, it is possible to find a $\bar{\omega} < \infty$ such that

$$\left|\frac{S_1}{a_N\|\boldsymbol{\gamma}-\boldsymbol{\gamma}^0\|}\right| > \sum_{k=2}^{5}\left|\frac{S_k}{a_N\|\boldsymbol{\gamma}-\boldsymbol{\gamma}^0\|}\right| \tag{A54}$$

holds for all sufficiently large N in probability. By (A54), we obtain (A45) and (A46). Therefore, we have established $\lim_{N\to\infty} P(\hat{\gamma} \in V_B') = 0$ and (A44), thus we obtain the convergence rate. This completes the proof of Lemma 11.

□

Lemma 12. *Under Assumption 1, set $\boldsymbol{\gamma} = \boldsymbol{\gamma}^0 + \boldsymbol{\omega}/a_N$, then*

$$\sqrt{N}\left[\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}) - \hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0)\right] \xrightarrow{p} \mathbf{0}, \text{ and } \sqrt{N}\left[\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0) - \boldsymbol{\theta}_{\boldsymbol{\delta}}^0\right] \xrightarrow{d} \mathcal{Z}_{**}, \tag{A55}$$

where $\mathcal{Z}_{**}$ is a Gaussian process with variance $\mathbf{M}_{**}^{-1}\boldsymbol{\Omega}_{**}\mathbf{M}_{**}^{-1}$, in which

$$\mathbf{M}_{**} = \begin{bmatrix} \mathbf{M} & \mathbf{M}_1 \\ \mathbf{M}_1 & \mathbf{M}_1 \end{bmatrix}, \ \boldsymbol{\Omega}_{**} = \begin{bmatrix} \boldsymbol{\Omega}_1 + \boldsymbol{\Omega}_2 + \boldsymbol{\Omega}_{12} + \boldsymbol{\Omega}_{12}' & \boldsymbol{\Omega}_1 + \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_1 + \boldsymbol{\Omega}_{12}' & \boldsymbol{\Omega}_1 \end{bmatrix}, \tag{A56}$$

$\mathbf{M}_1 = \mathbf{M}(\boldsymbol{\gamma}^0)$, $\boldsymbol{\Omega}_1 = \boldsymbol{\Omega}_1(\boldsymbol{\gamma}^0)$, $\boldsymbol{\Omega}_2 = \boldsymbol{\Omega}_2(\boldsymbol{\gamma}^0)$ *and* $\boldsymbol{\Omega}_{12} = \boldsymbol{\Omega}_{12}(\boldsymbol{\gamma}^0)$.

Proof. By Lemma 5, on any compact set $\boldsymbol{\Psi}$, we have

$$\frac{1}{\sqrt{N}}\check{\mathbf{X}}_{\boldsymbol{\gamma}}'\mathbf{e} \xrightarrow{d} \mathcal{Z}_{0**}, \tag{A57}$$

where $\mathcal{Z}_{0**}$ is a Gaussian process with variance $\boldsymbol{\Omega}_{**}$. Also,

$$\begin{aligned}
\sqrt{N}\left[\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}) - \boldsymbol{\theta}_{\boldsymbol{\delta}}^0\right] &= \left[\frac{1}{N}\check{\mathbf{X}}_{\boldsymbol{\gamma}}'\check{\mathbf{X}}_{\boldsymbol{\gamma}}\right]^{-1}\left[\frac{1}{\sqrt{N}}\check{\mathbf{X}}_{\boldsymbol{\gamma}}'\mathbf{e} - a_N^{-1/2}\frac{1}{N}\check{\mathbf{X}}_{\boldsymbol{\gamma}}\nabla\check{\mathbf{x}}_{\boldsymbol{\gamma}}\mathbf{c}\right] && \text{(A58)}\\
&\xrightarrow{d} \mathbf{M}_{**}^{-1}\mathcal{Z}_{0**} \equiv \mathcal{Z}_{**}. && \text{(A59)}
\end{aligned}$$

We have established $\sqrt{N}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\gamma) - \boldsymbol{\theta}^0_{\boldsymbol{\delta}}) \xrightarrow{d} \mathcal{Z}_{**}$, and it implies

$$\begin{aligned}
\sqrt{N}\left[\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\gamma) - \hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0)\right] &= \sqrt{N}\left[\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\gamma) - \boldsymbol{\theta}^0_{\boldsymbol{\delta}}\right] - \sqrt{N}\left[\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0) - \boldsymbol{\theta}^0_{\boldsymbol{\delta}}\right] \\
&\xrightarrow{d} \mathcal{Z}_{**} - \mathcal{Z}_{**} = \mathbf{0}. \qquad \text{(A60)}
\end{aligned}$$

This completes the proof of Lemma 12.

□

Proof of Theorem 1. Using Lemma 11-12, we can conclude that

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}} - \boldsymbol{\theta}^0_{\boldsymbol{\delta}}\right) \xrightarrow{d} \mathcal{Z}_{**}, \qquad \text{(A61)}$$

and

$$\begin{aligned}
\sqrt{N}\boldsymbol{A}^{-1}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}} - \boldsymbol{\theta}^0_{\boldsymbol{\delta}}\right) &= \sqrt{N}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0\right) \\
&\xrightarrow{d} \boldsymbol{A}^{-1}\mathcal{Z}_{**} = \mathcal{Z}. \qquad \text{(A62)}
\end{aligned}$$

From Lemma 11, the threshold estimator is consistent with convergence rate $a_N = N^{1-2\alpha}$; thus, we can study their asymptotic behavior in the neighborhood of the true thresholds, $\hat{\boldsymbol{\gamma}} = \boldsymbol{\gamma}^0 + \hat{\boldsymbol{\omega}}/a_N$.

By the definition of the threshold estimator, we have

$$\begin{aligned}
a_N\left(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^0\right) &= \hat{\boldsymbol{\omega}} \\
&= \underset{\boldsymbol{\omega}\in\boldsymbol{\Psi}}{\arg\min}\, \widetilde{SSR}_{NT}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0 + \frac{\boldsymbol{\omega}}{a_N}), \boldsymbol{\gamma}^0 + \frac{\boldsymbol{\omega}}{a_N}\right) - \widetilde{SSR}_{NT}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0), \boldsymbol{\gamma}^0\right) \\
&= \underset{\boldsymbol{\omega}\in\boldsymbol{\Psi}}{\arg\min}\, Q_N(\boldsymbol{\omega}). \qquad \text{(A63)}
\end{aligned}$$

As in Lemma 11, we can obtain $Q_N(\boldsymbol{\omega}) = Q_1+Q_2+Q_3-2(Q_4-Q_5+Q_6+Q_7+Q_8)$ (defined in the later). We next derive the limiting behavior of each Q_i (i=1,2,...,8), respectively. Thus, by Lemma 1-12, let $\boldsymbol{\gamma} = \boldsymbol{\gamma}^0 + \boldsymbol{\omega} h$ and $h = a_N^{-1}$, we have the following results

$$\begin{aligned}
Q_1 &= N^{-2\alpha}\mathbf{c}'\nabla\check{\mathbf{x}}_{\boldsymbol{\gamma}}'\nabla\check{\mathbf{x}}_{\boldsymbol{\gamma}}\mathbf{c} = G_N(\boldsymbol{\omega}, h), && \text{(A64)}\\
Q_2 &= \sqrt{N}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0) - \boldsymbol{\theta}_{\boldsymbol{\delta}}^0\right)' \frac{1}{N}\left(\check{\mathbf{X}}_{\boldsymbol{\gamma}}'\check{\mathbf{X}}_{\boldsymbol{\gamma}} - \check{\mathbf{X}}_0'\check{\mathbf{X}}_0\right)\sqrt{N}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0) - \boldsymbol{\theta}_{\boldsymbol{\delta}}^0\right)\\
&= O_p(1)o_p(1)O_p(1) = o_p(1), && \text{(A65)}\\
Q_3 &= \sqrt{N}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}) - \hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0)\right)' \frac{1}{N}\check{\mathbf{X}}_{\boldsymbol{\gamma}}'\check{\mathbf{X}}_{\boldsymbol{\gamma}}\sqrt{N}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}) - \hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0)\right)\\
&= o_p(1)O_p(1)o_p(1) = o_p(1), && \text{(A66)}\\
Q_4 &= N^{-\alpha}\mathbf{c}'\nabla\check{\mathbf{x}}_{\boldsymbol{\gamma}}'\mathbf{e} = R_N(\boldsymbol{\omega}, h), && \text{(A67)}\\
Q_5 &= \sqrt{N}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0) - \boldsymbol{\theta}_{\boldsymbol{\delta}}^0\right)' \frac{1}{\sqrt{N}}\nabla\check{\mathbf{X}}_{\boldsymbol{\gamma}}\mathbf{e} = O_p(1)o_p(1) = o_p(1), && \text{(A68)}\\
Q_6 &= \sqrt{N}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}) - \boldsymbol{\theta}_{\boldsymbol{\delta}}^0\right)' \frac{1}{\sqrt{Nh}}\check{\mathbf{X}}_{\boldsymbol{\gamma}}\nabla\check{\mathbf{x}}_{\boldsymbol{\gamma}}\mathbf{c}N^{-\alpha} = O_p(1)O_p(1)o_p(1) = o_p(1), && \text{(A69)}\\
Q_7 &= \sqrt{N}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}) - \hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0)\right)' \frac{1}{N}\check{\mathbf{X}}_{\boldsymbol{\gamma}}'\check{\mathbf{X}}_{\boldsymbol{\gamma}}\sqrt{N}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0) - \boldsymbol{\theta}_{\boldsymbol{\delta}}^0\right)\\
&= o_p(1)O_p(1)o_p(1) = o_p(1) && \text{(A70)}\\
Q_8 &= \sqrt{N}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}) - \hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0)\right)' \frac{1}{\sqrt{N}}\check{\mathbf{X}}_{\boldsymbol{\gamma}}\mathbf{e} = o_p(1)O_p(1) = o_p(1), && \text{(A71)}
\end{aligned}$$

and therefore,

$$\begin{aligned}
Q_N(\boldsymbol{\omega}) &= Q_1 - 2Q_4 + o_p(1)\\
&\xrightarrow{d} G_T(\boldsymbol{\omega}) - 2R_T(\boldsymbol{\omega}). && \text{(A72)}
\end{aligned}$$

Moreover, we have

$$\begin{aligned}
a_N(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^0) &\xrightarrow{d} \underset{\boldsymbol{\omega}\in\mathbb{R}^{k+1}}{\arg\min}\, G_T(\boldsymbol{\omega}) - 2R_T(\boldsymbol{\omega})\\
&= \underset{\boldsymbol{\omega}\in\mathbb{R}^{k+1}}{\arg\min}\left[\frac{1}{2}G_T(\boldsymbol{\omega}) - R_T(\boldsymbol{\omega})\right]. && \text{(A73)}
\end{aligned}$$

This completes the proof of Theorem 1.

□

Proof of Theorem 2. We first derive the limiting distribution for F_1, then the limiting distribution for F_C, and finally the limiting distribution for F_I.

First, we derive the limiting distribution of $\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma})$ under $H_0^1 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$.

Under $H_0^1 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$, by Lemma 1-5, for any fixed $\boldsymbol{\gamma} \in \boldsymbol{\Gamma}$, we have

$$\begin{aligned}
\sqrt{N}\left[\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}) - \boldsymbol{\theta}_{\boldsymbol{\delta}}^0\right] &= \left[\frac{1}{N}\check{\mathbf{X}}_{\boldsymbol{\gamma}}'\check{\mathbf{X}}_{\boldsymbol{\gamma}}\right]^{-1}\left[\frac{1}{\sqrt{N}}\check{\mathbf{X}}_{\boldsymbol{\gamma}}'\mathbf{e}\right] && \text{(A74)}\\
&\xrightarrow{p} \mathcal{Z}_{**}(\boldsymbol{\gamma}), && \text{(A75)}
\end{aligned}$$

where $\mathcal{Z}_{**}(\boldsymbol{\gamma})$ is a Gaussian process with variance $\mathbf{M}_{**}^{-1}(\boldsymbol{\gamma})\boldsymbol{\Omega}_{**}(\boldsymbol{\gamma})\mathbf{M}_{**}^{-1}(\boldsymbol{\gamma})$, in which

$$\begin{aligned}\mathbf{M}_{**}(\boldsymbol{\gamma}) &= \begin{bmatrix} \mathbf{M} & \mathbf{M}(\boldsymbol{\gamma}) \\ \mathbf{M}(\boldsymbol{\gamma}) & \mathbf{M}(\boldsymbol{\gamma}) \end{bmatrix} \\ \boldsymbol{\Omega}_{**}(\boldsymbol{\gamma}) &= \begin{bmatrix} \boldsymbol{\Omega}_1(\boldsymbol{\gamma}) + \boldsymbol{\Omega}_2(\boldsymbol{\gamma}) + \boldsymbol{\Omega}_{12}(\boldsymbol{\gamma}) + \boldsymbol{\Omega}_{12}'(\boldsymbol{\gamma}) & \boldsymbol{\Omega}_1(\boldsymbol{\gamma}) + \boldsymbol{\Omega}_{12}(\boldsymbol{\gamma}) \\ \boldsymbol{\Omega}_1(\boldsymbol{\gamma}) + \boldsymbol{\Omega}_{12}'(\boldsymbol{\gamma}) & \boldsymbol{\Omega}_1(\boldsymbol{\gamma}) \end{bmatrix}. \end{aligned} \tag{A76}$$

Denote $\widetilde{\boldsymbol{\theta}}_{\boldsymbol{\delta}} = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_{c1}} \widetilde{SSR}_{NT}(\boldsymbol{\theta}_{\boldsymbol{\delta}}, \boldsymbol{\gamma})$, where $\boldsymbol{\Theta}_{c1} = \{\boldsymbol{\theta} \in \boldsymbol{\Theta} : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2\}$. Next, we derive the limit distribution of $\tilde{\boldsymbol{\theta}}_{\boldsymbol{\delta}} - \hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma})$. Under $H_0^1 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$, the model degenerates into a linear regression model when estimating parameters using the restriction of $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$, in which case we have $\tilde{\boldsymbol{\theta}}_{\boldsymbol{\delta}} - \boldsymbol{\theta}_{\boldsymbol{\delta}}^0 \to_p 0$. This, $\sqrt{N}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{\delta}} - \boldsymbol{\theta}_{\boldsymbol{\delta}}^0)$ can be solved using the restricted least squared estimator

$$\begin{aligned}\sqrt{N}\left(\tilde{\boldsymbol{\theta}}_{\boldsymbol{\delta}} - \boldsymbol{\theta}_{\boldsymbol{\delta}}^0\right) &= \Big[\widehat{\mathbf{H}}_N^{-1}(\boldsymbol{\theta}_{\boldsymbol{\delta}}^0, \boldsymbol{\gamma}) - \widehat{\mathbf{H}}_N^{-1}(\boldsymbol{\theta}_{\boldsymbol{\delta}}^0, \boldsymbol{\gamma})\mathbf{R}_{**}'\left(\mathbf{R}_{**}\widehat{\mathbf{H}}_N^{-1}(\boldsymbol{\theta}_{\boldsymbol{\delta}}^0, \boldsymbol{\gamma})\mathbf{R}_{**}'\right)^{-1} \\ &\quad \times \mathbf{R}_{**}\widehat{\mathbf{H}}_N^{-1}(\boldsymbol{\theta}_{\boldsymbol{\delta}}^0, \boldsymbol{\gamma})\Big]\widehat{\mathbf{K}}_N(\boldsymbol{\theta}_{\boldsymbol{\delta}}^0, \boldsymbol{\gamma}),\end{aligned} \tag{A77}$$

where $\mathbf{R}_{**}$ is a matrix such that $\mathbf{R}_{**}\boldsymbol{\theta}_{\boldsymbol{\delta}} = \boldsymbol{\delta} = \mathbf{0}$, $\widehat{\mathbf{K}}_N(\boldsymbol{\theta}_{\boldsymbol{\delta}}^0, \boldsymbol{\gamma}) = N^{-1/2}\check{\mathbf{X}}_{\boldsymbol{\gamma}}'\mathbf{e}$ and $\widehat{\mathbf{H}}_N(\boldsymbol{\theta}_{\boldsymbol{\delta}}^0, \boldsymbol{\gamma}) = \check{\mathbf{X}}_{\boldsymbol{\gamma}}'\check{\mathbf{X}}_{\boldsymbol{\gamma}}/N$. From (A74) and (A77), we have

$$\begin{aligned}\sqrt{N}\left(\tilde{\boldsymbol{\theta}}_{\boldsymbol{\delta}} - \boldsymbol{\theta}_{\boldsymbol{\delta}}^0\right) &= \sqrt{N}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}) - \boldsymbol{\theta}_{\boldsymbol{\delta}}^0\right) - \widehat{\mathbf{H}}_N^{-1}(\boldsymbol{\theta}_{\boldsymbol{\delta}}^0, \boldsymbol{\gamma})\mathbf{R}_{**}'\left(\mathbf{R}_{**}\widehat{\mathbf{H}}_N^{-1}(\boldsymbol{\theta}_{\boldsymbol{\delta}}^0, \boldsymbol{\gamma})\mathbf{R}_{**}'\right)^{-1}\mathbf{R}_{**} \\ &\quad \times \widehat{\mathbf{H}}_N^{-1}(\boldsymbol{\theta}_{\boldsymbol{\delta}}^0, \boldsymbol{\gamma})\widehat{\mathbf{K}}_N(\boldsymbol{\theta}_{\boldsymbol{\delta}}^0, \boldsymbol{\gamma}),\end{aligned} \tag{A78}$$

it can be rewritten as

$$\begin{aligned}\sqrt{N}\left(\tilde{\boldsymbol{\theta}}_{\boldsymbol{\delta}} - \hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma})\right) &= -\widehat{\mathbf{H}}_N^{-1}(\boldsymbol{\theta}_{\boldsymbol{\delta}}^0, \boldsymbol{\gamma})\mathbf{R}_{**}'\left(\mathbf{R}_{**}\widehat{\mathbf{H}}_N^{-1}(\boldsymbol{\theta}_{\boldsymbol{\delta}}^0, \boldsymbol{\gamma})\mathbf{R}_{**}'\right)^{-1}\mathbf{R}_{**}\widehat{\mathbf{H}}_N^{-1}(\boldsymbol{\theta}_{\boldsymbol{\delta}}^0, \boldsymbol{\gamma}) \\ &\quad \times \widehat{\mathbf{K}}_N(\boldsymbol{\theta}_{\boldsymbol{\delta}}^0, \boldsymbol{\gamma}),\end{aligned} \tag{A79}$$

, we therefore also have $\tilde{\boldsymbol{\theta}}_{\boldsymbol{\delta}} - \hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}) \to_p \mathbf{0}$.

Next, a Taylor's expansion of $\widetilde{SSR}_{NT}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{\delta}}, \boldsymbol{\gamma})$ at $\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma})$ gives

$$\begin{aligned}\widetilde{SSR}_{NT}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{\delta}}, \boldsymbol{\gamma}) &= \widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}), \boldsymbol{\gamma}) - 2\widehat{\mathbf{K}}_N(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}), \boldsymbol{\gamma})\sqrt{N}\left[\tilde{\boldsymbol{\theta}}_{\boldsymbol{\delta}} - \hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma})\right] \\ &\quad + \sqrt{N}\left[\tilde{\boldsymbol{\theta}}_{\boldsymbol{\delta}} - \hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma})\right]'\widehat{\mathbf{H}}_N(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}), \boldsymbol{\gamma})\sqrt{N}\left[\tilde{\boldsymbol{\theta}}_{\boldsymbol{\delta}} - \hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma})\right].\end{aligned} \tag{A80}$$

Consequently , we have

$$\begin{aligned}&\widetilde{SSR}_{NT}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{\delta}}, \boldsymbol{\gamma}) - \widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}), \boldsymbol{\gamma}) \\ &\quad = \sqrt{N}\left[\tilde{\boldsymbol{\theta}}_{\boldsymbol{\delta}} - \hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma})\right]'\widehat{\mathbf{H}}_N(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}), \boldsymbol{\gamma})\sqrt{N}\left[\tilde{\boldsymbol{\theta}}_{\boldsymbol{\delta}} - \hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma})\right] \\ &\quad \xrightarrow{d} \mathcal{Z}_{**}'(\boldsymbol{\gamma})\mathbf{R}_{**}'\left[\mathbf{R}_{**}\mathbf{M}_{**}^{-1}(\boldsymbol{\gamma})\mathbf{R}_{**}'\right]^{-1}\mathbf{R}_{**}\mathcal{Z}_{**}(\boldsymbol{\gamma}),\end{aligned} \tag{A81}$$

and a Taylor's expansion of $\widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\gamma), \gamma)$ at $\boldsymbol{\theta}^0_{\boldsymbol{\delta}}$ gives

$$\begin{aligned}
\widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\gamma), \gamma) &= \widetilde{SSR}_{NT}(\boldsymbol{\theta}^0_{\boldsymbol{\delta}}, \gamma) - 2\widehat{\mathbf{K}}_N(\boldsymbol{\theta}^0_{\boldsymbol{\delta}}, \gamma)\sqrt{N}\left[\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\gamma) - \boldsymbol{\theta}^0_{\boldsymbol{\delta}}\right] \\
&\quad + \sqrt{N}\left[\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\gamma) - \boldsymbol{\theta}^0_{\boldsymbol{\delta}}\right]' \widehat{\mathbf{H}}_N(\boldsymbol{\theta}^0_{\boldsymbol{\delta}}, \gamma)\sqrt{N}\left[\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\gamma) - \boldsymbol{\theta}^0_{\boldsymbol{\delta}}\right] \\
&= \|\mathbf{e}\|^2 + O_p(1) + O_p(1), \qquad \text{(A82)}
\end{aligned}$$

and we obtain

$$\begin{aligned}
\frac{1}{NT}\widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\gamma), \gamma) &= \frac{1}{NT}\|\mathbf{e}\|^2 + O_p(N^{-1}) + O_p(N^{-1}) \\
&\xrightarrow{p} \sigma^2. \qquad \text{(A83)}
\end{aligned}$$

Finally, combining the above results, we obtain

$$\begin{aligned}
F_1 &= \frac{\widetilde{SSR}_{NT}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{\delta}}, \hat{\gamma}) - \widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\hat{\gamma}), \hat{\gamma})}{\widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\hat{\gamma}), \hat{\gamma})/(NT)} \\
&= \sup_{\gamma \in \boldsymbol{\Gamma}} \frac{\widetilde{SSR}_{NT}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{\delta}}, \gamma) - \widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\gamma), \gamma)}{\widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\gamma), \gamma)/(NT)} \\
&\xrightarrow{d} \frac{1}{\sigma^2}\sup_{\gamma \in \boldsymbol{\Gamma}} \mathcal{Z}'_{**}(\gamma)\mathbf{R}'_{**}\left[\mathbf{R}_{**}\mathbf{M}^{-1}_{**}(\gamma)\mathbf{R}'_{**}\right]^{-1}\mathbf{R}_{**}\mathcal{Z}_{**}(\gamma). \qquad \text{(A84)}
\end{aligned}$$

The limiting distribution of F_1 is obtained, as it is easy to show that

$$\begin{aligned}
&\mathcal{Z}'_{**}(\gamma)\mathbf{R}'_{**}\left[\mathbf{R}_{**}\mathbf{M}^{-1}_{**}(\gamma)\mathbf{R}'_{**}\right]^{-1}\mathbf{R}_{**}\mathcal{Z}_{**}(\gamma) \\
&\quad = \mathcal{Z}'_{**}(\gamma)\boldsymbol{A}^{-1}\boldsymbol{A}\mathbf{R}'_{**}\left[\mathbf{R}_{**}\boldsymbol{A}\boldsymbol{A}^{-1}\mathbf{M}^{-1}_{**}(\gamma)\boldsymbol{A}^{-1}\boldsymbol{A}\mathbf{R}'_{**}\right]^{-1}\mathbf{R}_{**}\boldsymbol{A}\boldsymbol{A}^{-1}\mathcal{Z}_{**}(\gamma) \\
&\quad = \mathcal{Z}'(\gamma)\mathbf{R}'_{*}\left[\mathbf{R}_{*}\mathbf{M}^{-1}_{*}(\gamma)\mathbf{R}'_{*}\right]^{-1}\mathbf{R}_{*}\mathcal{Z}(\gamma). \qquad \text{(A85)}
\end{aligned}$$

Next, we derive the limiting distribution of F_C. Denote

$$\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}) = \arg\min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \widetilde{SSR}_{NT}(\boldsymbol{\theta}_{\boldsymbol{\delta}}, \boldsymbol{\gamma}), \qquad \text{(A86)}$$

$$\tilde{\boldsymbol{\gamma}} = \arg\min_{\boldsymbol{\gamma} \in \boldsymbol{\Gamma}_c} \widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}), \boldsymbol{\gamma}), \qquad \text{(A87)}$$

where $\boldsymbol{\Gamma}_c = \{\boldsymbol{\gamma} \in \boldsymbol{\Gamma} : \boldsymbol{\gamma}_{\mathbf{s}} = \mathbf{0}\}$ and $\tilde{\boldsymbol{\gamma}} = (\tilde{\gamma}_0, \tilde{\boldsymbol{\gamma}}'_{\mathbf{s}})' = (\tilde{\gamma}_0, \mathbf{0}')'$.

Under $H^2_0 : \boldsymbol{\gamma}_{\mathbf{s}} = \mathbf{0}$, the model degenerates into a panel threshold model with a constant threshold when estimating parameters using the restriction of $\boldsymbol{\gamma}_{\mathbf{s}} = \mathbf{0}$. Similarly with the proof of Theorem 1, we obtain

$$\begin{aligned}
&\widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\tilde{\gamma}), \tilde{\gamma}) - \widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0), \boldsymbol{\gamma}^0) \\
&\quad = \min_{\boldsymbol{\gamma} \in \boldsymbol{\Gamma}_c}\left[\widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}), \boldsymbol{\gamma}) - \widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0), \boldsymbol{\gamma}^0)\right] \\
&\quad = \min_{\boldsymbol{\omega} \in \boldsymbol{\Psi}_c}\left[\widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}_0 + \frac{\boldsymbol{\omega}}{a_N}), \boldsymbol{\gamma}_0 + \frac{\boldsymbol{\omega}}{a_N}) - \widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0), \boldsymbol{\gamma}^0)\right] \\
&\quad \xrightarrow{d} \min_{\boldsymbol{\omega} \in \mathbb{R}^{k+1}_c} G_T(\boldsymbol{\omega}) - 2R_T(\boldsymbol{\omega}), \qquad \text{(A88)}
\end{aligned}$$

where $\boldsymbol{\Psi}_c = \{\boldsymbol{\omega} \in \boldsymbol{\Psi} : \boldsymbol{\omega}_{\mathbf{s}} = \mathbf{0}\}$, $\mathbb{R}^{k+1}_c = \{\boldsymbol{\omega} \in \mathbb{R}^{k+1} : \boldsymbol{\omega}_{\mathbf{s}} = \mathbf{0}\}$. From the proof of Theorem 1, we have

$$\widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\hat{\gamma}),\hat{\gamma}) - \widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0),\boldsymbol{\gamma}^0) \xrightarrow{d} \min_{\boldsymbol{\omega}\in\mathbb{R}^{k+1}} G_T(\boldsymbol{\omega}) - 2R_T(\boldsymbol{\omega}), \tag{A89}$$

and

$$\begin{aligned}\frac{1}{NT}\widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\hat{\gamma}),\hat{\gamma}) &= \frac{1}{NT}\|\mathbf{e}\|^2 + \frac{1}{NT}\sqrt{N}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0) - \boldsymbol{\theta}^0_{\boldsymbol{\delta}}\right)'\frac{1}{N}\check{\mathbf{X}}_0'\check{\mathbf{X}}_0\sqrt{N}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0) - \boldsymbol{\theta}^0_{\boldsymbol{\delta}}\right)\\ &\quad -\frac{2}{NT}\sqrt{N}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0) - \boldsymbol{\theta}^0_{\boldsymbol{\delta}}\right)'\check{\mathbf{X}}_0'\mathbf{e} + O_p(N^{-1})\\ &\xrightarrow{p} \sigma^2, \end{aligned}\tag{A90}$$

Finally, we have the following result

$$\begin{aligned}F_C &= \frac{\widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\tilde{\gamma}),\tilde{\gamma}) - \widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\hat{\gamma}),\hat{\gamma})}{\widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\hat{\gamma}),\hat{\gamma})/(NT)}\\ &= \frac{\widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\tilde{\gamma}),\tilde{\gamma}) - \widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0),\boldsymbol{\gamma}^0)}{\widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\hat{\gamma}),\hat{\gamma})/(NT)}\\ &\quad - \frac{\widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\hat{\gamma}),\hat{\gamma}) - \widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0),\boldsymbol{\gamma}^0)}{\widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\hat{\gamma}),\hat{\gamma})/(NT)}\\ &\xrightarrow{d} \frac{1}{\sigma^2}\left[\left(\min_{\boldsymbol{\omega}\in\mathbb{R}^{k+1}_c} G_T(\boldsymbol{\omega}) - 2R_T(\boldsymbol{\omega})\right) - \left(\min_{\boldsymbol{\omega}\in\mathbb{R}^{k+1}} G_T(\boldsymbol{\omega}) - 2R_T(\boldsymbol{\omega})\right)\right]. \end{aligned}\tag{A91}$$

Next, we derive the limiting distribution of F_I. Denote

$$\tilde{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}) = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}_c} \widetilde{SSR}_{NT}(\boldsymbol{\theta}_{\boldsymbol{\delta}},\boldsymbol{\gamma}), \tag{A92}$$

$$\tilde{\gamma} = \arg\min_{\boldsymbol{\gamma}\in\boldsymbol{\Gamma}} \widetilde{SSR}_{NT}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}),\boldsymbol{\gamma}), \tag{A93}$$

where $\boldsymbol{\Theta}_c = \{\boldsymbol{\theta}\in\boldsymbol{\Theta} : \boldsymbol{\psi}_1 = \boldsymbol{\psi}_2\}$.

First, we derive the limiting distribution of $\sqrt{N}[\tilde{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0) - \hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0)]$ under $H_0^3 : \boldsymbol{\psi}_1 = \boldsymbol{\psi}_2$. Note that, $\sqrt{N}[\tilde{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0) - \boldsymbol{\theta}^0_{\boldsymbol{\delta}}]$ can be solved as a restricted least squared estimator

$$\begin{aligned}\sqrt{N}\left(\tilde{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0) - \boldsymbol{\theta}^0_{\boldsymbol{\delta}}\right) &= \Big[\widehat{\mathbf{H}}_N^{-1}(\boldsymbol{\theta}^0_{\boldsymbol{\delta}},\boldsymbol{\gamma}^0) - \widehat{\mathbf{H}}_N^{-1}(\boldsymbol{\theta}^0_{\boldsymbol{\delta}},\boldsymbol{\gamma}^0)\mathbf{R}_{I*}'\left(\mathbf{R}_{I*}\widehat{\mathbf{H}}_N^{-1}(\boldsymbol{\theta}^0_{\boldsymbol{\delta}},\boldsymbol{\gamma}^0)\mathbf{R}_{I*}'\right)^{-1}\\ &\quad \times \mathbf{R}_{I*}\widehat{\mathbf{H}}_N^{-1}(\boldsymbol{\theta}^0_{\boldsymbol{\delta}},\boldsymbol{\gamma}^0)\Big]\widehat{\mathbf{K}}_N(\boldsymbol{\theta}^0_{\boldsymbol{\delta}},\boldsymbol{\gamma}^0), \end{aligned}\tag{A94}$$

where $\mathbf{R}_{I*}$ is a matrix such that $\mathbf{R}_{I*}\boldsymbol{\theta}_{\boldsymbol{\delta}} = \boldsymbol{\psi}_1 - \boldsymbol{\psi}_2 = \mathbf{0}$, $\widehat{\mathbf{K}}_N(\boldsymbol{\theta}^0_{\boldsymbol{\delta}},\boldsymbol{\gamma}^0) = N^{-1/2}\check{\mathbf{X}}_0'(\mathbf{e} - \nabla\check{\mathbf{x}}\boldsymbol{\delta}_0)$ and $\widehat{\mathbf{H}}_N(\boldsymbol{\theta}^0_{\boldsymbol{\delta}},\boldsymbol{\gamma}^0) = \check{\mathbf{X}}_0'\check{\mathbf{X}}_0/N$. From (A58) and (A94), we have

$$\begin{aligned}\sqrt{N}\left(\tilde{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0) - \boldsymbol{\theta}^0_{\boldsymbol{\delta}}\right) &= \sqrt{N}\left(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0) - \boldsymbol{\theta}^0_{\boldsymbol{\delta}}\right) - \widehat{\mathbf{H}}_N^{-1}(\boldsymbol{\theta}^0_{\boldsymbol{\delta}},\boldsymbol{\gamma}^0)\mathbf{R}_{I*}'\left(\mathbf{R}_{I*}\widehat{\mathbf{H}}_N^{-1}(\boldsymbol{\theta}^0_{\boldsymbol{\delta}},\boldsymbol{\gamma}^0)\mathbf{R}_{I*}'\right)^{-1}\\ &\quad \times \mathbf{R}_{I*}\widehat{\mathbf{H}}_N^{-1}(\boldsymbol{\theta}^0_{\boldsymbol{\delta}},\boldsymbol{\gamma}^0)\widehat{\mathbf{K}}_N(\boldsymbol{\theta}^0_{\boldsymbol{\delta}},\boldsymbol{\gamma}^0). \end{aligned}\tag{A95}$$

(A95) can be rewritten as

$$\sqrt{N}\left(\tilde{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0)-\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0)\right) = -\widehat{\mathbf{H}}_N^{-1}(\boldsymbol{\theta}^0_{\boldsymbol{\delta}},\boldsymbol{\gamma}^0)\mathbf{R}'_{I*}\left(\mathbf{R}_{I*}\widehat{\mathbf{H}}_N^{-1}(\boldsymbol{\theta}^0_{\boldsymbol{\delta}},\boldsymbol{\gamma}^0)\mathbf{R}'_{I*}\right)^{-1}\mathbf{R}_{I*}\widehat{\mathbf{H}}_N^{-1}(\boldsymbol{\theta}^0_{\boldsymbol{\delta}},\boldsymbol{\gamma}^0) \times \widehat{\mathbf{K}}_N(\boldsymbol{\theta}^0_{\boldsymbol{\delta}},\boldsymbol{\gamma}^0). \tag{A96}$$

Next, a Taylor's expansion of $\widetilde{SSR}_{NT}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0),\boldsymbol{\gamma}^0)$ at $\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0)$ gives

$$\begin{aligned}\widetilde{SSR}_{NT}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0),\boldsymbol{\gamma}^0) &= \widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0),\boldsymbol{\gamma}^0)-2\widehat{\mathbf{K}}_N(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0),\boldsymbol{\gamma}^0)\sqrt{N}\left[\tilde{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma})-\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0)\right]\\ &\quad+\sqrt{N}\left[\tilde{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0)-\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0)\right]'\widehat{\mathbf{H}}_N(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0),\boldsymbol{\gamma}^0)\\ &\quad\times\sqrt{N}\left[\tilde{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0)-\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0)\right].\end{aligned} \tag{A97}$$

Thus, we have

$$\begin{aligned}&\widetilde{SSR}_{NT}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0),\boldsymbol{\gamma}^0)-\widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0),\boldsymbol{\gamma}^0)\\ &\quad= \sqrt{N}\left[\tilde{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0)-\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0)\right]'\widehat{\mathbf{H}}_N(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0),\boldsymbol{\gamma}^0)\sqrt{N}\left[\tilde{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0)-\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0)\right]\\ &\quad\xrightarrow{d} \mathscr{Z}'_{**}\mathbf{R}'_{I*}\left[\mathbf{R}_{I*}\mathbf{M}^{-1}_{**}\mathbf{R}'_{I*}\right]^{-1}\mathbf{R}_{I*}\mathscr{Z}_{**}.\end{aligned} \tag{A98}$$

From Theorem 1, we obtain

$$\begin{aligned}&\widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\hat{\gamma}),\hat{\gamma})-\widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0),\boldsymbol{\gamma}^0)\\ &\quad\xrightarrow{d} \min_{\boldsymbol{\omega}\in\mathbb{R}^{k+1}} G_T(\boldsymbol{\omega})-2R_T(\boldsymbol{\omega}),\end{aligned} \tag{A99}$$

and

$$\begin{aligned}&\widetilde{SSR}_{NT}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\tilde{\gamma}),\tilde{\gamma})-\widetilde{SSR}_{NT}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0),\boldsymbol{\gamma}^0)\\ &\quad= \min_{\boldsymbol{\gamma}\in\boldsymbol{\Gamma}}\left[\widetilde{SSR}_{NT}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}),\boldsymbol{\gamma})-\widetilde{SSR}_{NT}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0),\boldsymbol{\gamma}^0)\right]\\ &\quad= \min_{\boldsymbol{\omega}\in\boldsymbol{\Psi}}\left[\widetilde{SSR}_{NT}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}_0+\frac{\boldsymbol{\omega}}{a_N}),\boldsymbol{\gamma}_0+\frac{\boldsymbol{\omega}}{a_N})-\widetilde{SSR}_{NT}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0),\boldsymbol{\gamma}^0)\right]\\ &\quad\xrightarrow{d} \min_{\boldsymbol{\omega}\in\mathbb{R}^{k+1}} G_T(\boldsymbol{\omega})-2R_T(\boldsymbol{\omega}).\end{aligned} \tag{A100}$$

Hence, we have

$$\begin{aligned}&\left[\widetilde{SSR}_{NT}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\tilde{\gamma}),\tilde{\gamma})-\widetilde{SSR}_{NT}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0),\boldsymbol{\gamma}^0)\right]-\left[\widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\hat{\gamma}),\hat{\gamma})-\widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\boldsymbol{\gamma}^0),\boldsymbol{\gamma}^0)\right]\\ &\quad\xrightarrow{d} \min_{\boldsymbol{\omega}\in\mathbb{R}^{k+1}} G_T(\boldsymbol{\omega})-2R_T(\boldsymbol{\omega})-\min_{\boldsymbol{\omega}\in\mathbb{R}^{k+1}} G_T(\boldsymbol{\omega})-2R_T(\boldsymbol{\omega})=0.\end{aligned} \tag{A101}$$

From (A90), we have the following result

$$\frac{1}{NT}\widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\hat{\gamma}),\hat{\gamma}) \xrightarrow{p} \sigma^2. \tag{A102}$$

Combining the above results, we obtain

$$\begin{aligned}
F_I &= \frac{\widetilde{SSR}_{NT}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\tilde{\gamma}),\tilde{\gamma}) - \widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\hat{\gamma}),\hat{\gamma})}{S\tilde{S}R_{NT}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\hat{\gamma}),\hat{\gamma})/(NT)} \\
&= \frac{\widetilde{SSR}_{NT}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\tilde{\gamma}),\tilde{\gamma}) - \widetilde{SSR}_{NT}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\gamma^0),\gamma^0)}{S\tilde{S}R_{NT}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\hat{\gamma}),\hat{\gamma})/(NT)} \\
&\quad - \frac{\widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\hat{\gamma}),\hat{\gamma}) - \widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\gamma^0),\gamma^0)}{S\tilde{S}R_{NT}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\hat{\gamma}),\hat{\gamma})/(NT)} \\
&\quad + \frac{\widetilde{SSR}_{NT}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\gamma^0),\gamma^0) - \widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\gamma^0),\gamma^0)}{S\tilde{S}R_{NT}(\hat{\boldsymbol{\theta}}_{\boldsymbol{\delta}}(\hat{\gamma}),\hat{\gamma})/(NT)} \\
&\xrightarrow{d} \frac{0}{\sigma^2} + \frac{1}{\sigma^2}\mathcal{Z}_{**}'\mathbf{R}_{I*}'\left[\mathbf{R}_{I*}\mathbf{M}_{**}^{-1}\mathbf{R}_{I*}'\right]^{-1}\mathbf{R}_{I*}\mathcal{Z}_{**} \\
&= \frac{1}{\sigma^2}\mathcal{Z}_{**}'\mathbf{R}_{I*}'\left[\mathbf{R}_{I*}\mathbf{M}_{**}^{-1}\mathbf{R}_{I*}'\right]^{-1}\mathbf{R}_{I*}\mathcal{Z}_{**}. \qquad \text{(A103)}
\end{aligned}$$

The limiting distribution of F_I can be immediately obtained, as it is easy to see that

$$\begin{aligned}
&\mathcal{Z}_{**}'\mathbf{R}_{I*}'\left[\mathbf{R}_{I*}\mathbf{M}_{**}^{-1}\mathbf{R}_{I*}'\right]^{-1}\mathbf{R}_{I*}\mathcal{Z}_{**} \\
&\quad = \mathcal{Z}_{**}'\boldsymbol{A}^{-1}\boldsymbol{A}\mathbf{R}_{I*}'\left[\mathbf{R}_{I*}\boldsymbol{A}\boldsymbol{A}^{-1}\mathbf{M}_{**}^{-1}\boldsymbol{A}^{-1}\boldsymbol{A}\mathbf{R}_{I*}'\right]^{-1}\mathbf{R}_{I*}\boldsymbol{A}\boldsymbol{A}^{-1}\mathcal{Z}_{**} \\
&\quad = \mathcal{Z}'\mathbf{R}_I'\left[\mathbf{R}_I\mathbf{M}_*^{-1}\mathbf{R}_I'\right]^{-1}\mathbf{R}_I\mathcal{Z}. \qquad \text{(A104)}
\end{aligned}$$

□

Appendix B: Simulation results of the model with multiple covariate-dependent thresholds

In this section, we conduct Monte Carlo simulations to examine the finite sample performances of the proposed estimation and testing procedures for the model with multiple covariate-dependent thresholds. To this end, we consider the following data generating process (DGP) with double thresholds:

$$\begin{aligned} y_{it} &= (\beta_1 x_{it} + \beta_{11} q_{it} + \alpha_{1i} + \sigma_1 u_{it})\, I\,(q_{it} \leq \gamma_{1,it}) \\ &+ (\beta_2 x_{it} + \beta_{21} q_{it} + \alpha_{2i} + \sigma_2 u_{2it})\, I(\gamma_{1,it} < q_{it} \leq \gamma_{2,it}) \\ &+ (\beta_3 x_{it} + \beta_{31} q_{it} + \alpha_{3i} + \sigma_3 u_{it})\, I\,(q_{it} > \gamma_{2,it}) + \delta z_{it}, \qquad \text{(B1)} \\ \gamma_{1,it} &= \gamma_{10} + \gamma_{11} s_{it}, \qquad \text{(B2)} \\ \gamma_{2,it} &= \gamma_{20} + \gamma_{21} s_{it}, \qquad \text{(B3)} \end{aligned}$$

where $\alpha_{\ell i} = \psi_{\ell 1}\bar{x}_i + \psi_{\ell 2}\bar{q}_i + \psi_{\ell 3}\bar{z}_i + \psi_{\ell 0} + a_{\ell i}$ for $l = 1, 2, 3$, $\bar{x}_i = \frac{1}{T}\sum_{t=1}^{T} x_{it}$, $\bar{q}_i = \frac{1}{T}\sum_{t=1}^{T} q_{it}$, $\bar{z}_i = \frac{1}{T}\sum_{t=1}^{T} z_{it}$, $\gamma_{it} = \gamma_0 + \gamma_1 s_{it}$, and $u_{it} \sim N(0, 1)$, $\sigma_1 = \sigma_2 = \sigma_3 = 0.5$, a_{1i}, a_{2i} and a_{3i} follow $N(0, 1)$. q_{it}, x_{it}, z_{it} and s_{it} follow $\sim i.i.d.N(0, 1)$, and are independent of each other. The number of replications is set as 1000.

In examining the performance of the estimation procedure for the multiple threshold model, we set the true parameters as $(\beta_1, \beta_{11}, \beta_2, \beta_{21}, \beta_3, \beta_{31}, \delta) = (0.2, 0.2, -0.2, -0.2, -0.5, -0.5, 2)$, $(\gamma_{10}, \gamma_{11}, \gamma_{20}, \gamma_{21}) = (-0.5, 0.3, 0.2, 0.5)$, $(\psi_{11}, \psi_{12}, \psi_{13}, , \psi_{10}) = (-0.2, -0.2, 1, 0.3)$, $(\psi_{21}, \psi_{22}, \psi_{23}, , \psi_{20}) = (-0.5, -0.5, 2, 0.6)$, and $(\psi_{31}, \psi_{32}, \psi_{33}, , \psi_{30}) = (-1, -1, 3, 0.9)$. Table 5 reports the simulation results.[13] As expected, the simulation results show that the empirical mean of each parameter is close to its true value for all combinations of T and N, and the standard deviations decrease as the sample size increases. These simulation results support that the proposed estimation procedure for the PTCDI model with multiple threshold model works well in finite samples.

Table 6: Estimation for the PTCDI with two thresholds

T	N		γ_{10}	γ_{11}	γ_{20}	γ_{21}	β_1	β_{11}	β_2	β_{21}	β_3	β_{31}	δ
2	250	Mean	-0.496	0.309	0.205	0.501	0.201	0.192	-0.204	-0.195	-0.503	-0.512	2.001
		Std.dev	0.066	0.080	0.076	0.115	0.118	0.191	0.155	0.327	0.093	0.138	0.036
2	500	Mean	-0.500	0.301	0.203	0.504	0.201	0.195	-0.202	-0.210	-0.500	-0.502	1.999
		Std.dev	0.019	0.025	0.031	0.037	0.081	0.127	0.092	0.200	0.065	0.093	0.024
2	1000	Mean	-0.500	0.300	0.200	0.500	0.200	0.200	-0.197	-0.196	-0.501	-0.502	2.000
		Std.dev	0.009	0.011	0.013	0.017	0.056	0.091	0.063	0.134	0.043	0.063	0.016
5	250	Mean	-0.501	0.300	0.201	0.501	0.199	0.199	-0.199	-0.207	-0.499	-0.508	2.000
		Std.dev	0.016	0.020	0.025	0.031	0.055	0.098	0.063	0.170	0.044	0.072	0.016
5	500	Mean	-0.500	0.300	0.200	0.501	0.200	0.203	-0.200	-0.203	-0.501	-0.499	2.000
		Std.dev	0.008	0.010	0.012	0.016	0.036	0.070	0.044	0.117	0.031	0.050	0.011
5	1000	Mean	-0.500	0.300	0.200	0.500	0.200	0.203	-0.199	-0.198	-0.500	-0.498	2.000
		Std.dev	0.005	0.006	0.007	0.008	0.026	0.049	0.031	0.078	0.022	0.036	0.008
10	250	Mean	-0.500	0.300	0.201	0.501	0.200	0.199	-0.199	-0.198	-0.501	-0.498	2.000
		Std.dev	0.008	0.010	0.013	0.015	0.037	0.068	0.043	0.111	0.029	0.048	0.011
10	500	Mean	-0.500	0.300	0.200	0.500	0.200	0.204	-0.200	-0.197	-0.499	-0.499	2.000
		Std.dev	0.005	0.006	0.007	0.008	0.026	0.049	0.030	0.082	0.022	0.035	0.008
10	1000	Mean	-0.500	0.300	0.200	0.500	0.198	0.202	-0.200	-0.194	-0.499	-0.498	2.000
		Std.dev	0.004	0.005	0.005	0.006	0.017	0.033	0.022	0.058	0.015	0.024	0.005

[13] To save space, we do not report the simulation results for $(\psi_{l1}, \psi_{l2}, \psi_{l3}, , \psi_{l0})$ (l=1,2,3). The results show that the empirical mean of each parameter is also close to its true value for all combinations of T and N, and the standard deviations decrease as either T or N increases. These simulation results are available from the authors upon request.

To examine the size and power properties for the tests determining the number of thresholds is very time-consuming, while we conduct a small number of simulations and find that the tests have good size and power performance in small samples in an unreported appendix, which are available from the authors upon request.