Journal of Econometrics 258 (2026) 106337

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Panel threshold regression with unobserved individual-specific threshold effects

Ping Yu [a,*], Shengjie Hong [b], Peter C.B. Phillips [c]

[a] *HKU Business School, The University of Hong Kong, Pokfulam Road, Hong Kong*
[b] *School of Economics, Renmin University of China, Beijing, 100872, China*
[c] *Cowles Foundation for Research in Economics, Yale University, PO Box 208281, New Haven, CT, USA*

ARTICLE INFO

JEL classification:
C23
C24

Keywords:
Panel threshold regression
Unobserved threshold effects
Correlated random effects
Model selection
Sequential testing
Model misspecification
Adaptive bonferroni inference

ABSTRACT

This paper provides a new approach to estimation and inference in multiple regime panel threshold regression with unobserved individual-specific threshold effects. Such effects are practically relevant and have features that distinguish them from traditional linear panel data models. It is shown that within-regime demeaning in the static model or within-regime first-differencing in the dynamic model both fail to deliver consistent threshold estimators. Instead, correlated random effects models are suggested to address endogeneity in such panel threshold systems. The paper develops a unified framework of estimation and inference for both static and dynamic models that applies irrespective of whether the unobserved individual-specific threshold effects exist or whether the number of regimes is correctly specified. The approach involves model selection based on sequential testing which allows for underestimation of the number of regimes, develops asymptotic theory for least squares estimation that is robust to such model misspecification, and proposes new inferential methods for the model parameters which have improved asymptotic and empirical properties over existing methods. Simulations combined with an empirical application illustrate the practical utility of the new methodology.

1. Introduction

Panel data models provide an efficient way to eliminate endogeneity without the use of external instruments and dynamic panel models help to capture temporal effects that cross-sectional data models cannot. Panel threshold regression (PTR) models capture a further dimension in practical work by modeling individual behavior heterogeneity through the introducion of threshold effects. The PTR model in the literature usually assumes the form

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

where the parameter of interest is $\theta = \left(\gamma, \beta'\right)'$ with $\beta = \left(\beta_1', \beta_2'\right)'$ or equivalently, $\theta = \left(\gamma, \beta_2', \delta_\beta'\right)'$ with $\delta_\beta = \beta_1 - \beta_2$ being the threshold effect in the conditional mean of y_{it}, the observable time-variant covariates $\mathbf{x}_{it} \in \mathbb{R}^d$ can generally include a full set of time dummies, other aggregate time variables or lagged $\mathbf{x}_{it}$'s, α_i is the unobserved individual-specific, time-invariant effect (or unobserved heterogeneity), and u_{it} is the idiosyncratic time-varying shock. This setup is similar to the traditional linear panel data model except that

* Corresponding author.
E-mail addresses: pingyu@hku.hk (P. Yu), hongshj@ruc.edu.cn (S. Hong), peter.phillips@yale.edu (P.C.B. Phillips).

https://doi.org/10.1016/j.jeconom.2026.106337
Received 21 September 2025; Received in revised form 28 May 2026; Accepted 2 September 2026

the regression coefficients depend on whether the threshold variable q_{it} crosses γ, where $\gamma \in \Gamma$ is the threshold point. In static PTR (SPTR), we usually assume $\mathbf{x}_{it}$ is strictly exogenous with respect to u_{it}, while in dynamic PTR (DPTR), $\mathbf{x}_{it}$ also contains lagged y_{it} variables with the benchmark case where only $y_{i,t-1}$ appears in $\mathbf{x}_{it}$. Often, q_{it} is also included in $\mathbf{x}_{it}$. So in SPTR, we write $\mathbf{x}_{it}' = \left(x_{it}', q_{it}\right)$, and in DPTR, $\mathbf{x}_{it}' = \left(y_{i,t-1}, \underline{\mathbf{x}}_{it}'\right)$, where $\underline{\mathbf{x}}_{it}' = \left(x_{it}', q_{it}\right)$ if q_{it} is not $y_{i,t-1}$, and $\underline{\mathbf{x}}_{it} = x_{it}$ otherwise.

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

The second strand of literature transforms (1) into nonparametric PTR and applies the integrated difference kernel estimator (IDKE) of Yu and Phillips (2018) (YP hereafter) to identify γ. This idea is proposed at the end of YP and can be applied to both SPTR and DPTR. YP use the fixed-threshold-effect framework of Chan (1993), and Yu et al. (2018) extends this to the small-threshold-effect framework. Later, Gørgens and Ẃurtz (2019) combine the ideas of the two strands in DPTR; they first take differences to eliminate α_i and then apply the IDKE to estimate γ and GMM to estimate β. They compare the performance of $\hat{\gamma}$ and $\hat{\beta}$ with that of Seo and Shin (2016) in simulations and show that their approach is much more efficient in finite samples; this is because the convergence rate of the IDKE of γ is much faster than that of the FD-GMM estimator of γ, which makes the estimation of β in the second stage less contaminated. The IDKE may suffer from the curse of dimensionality when the dimension of x_{it} is large so the approach is not always practical. Finally, Wang and Lin (2010) extends the control function approach of Kourtellos et al. (2016) to SPTR, but as shown in Yu et al. (2024), the resulting estimate $\hat{\gamma}$ is not generically consistent; also, this approach requires external instruments, whereas the other methods mentioned above do not require such instruments for identification.

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

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

where a popular special case of u_{1it} and u_{2it} is that $u_{1it} = \sigma_1 u_{it}$ and $u_{2it} = \sigma_2 u_{it}$ with $\mathbb{E}\left[u_{it}^2\right] = 1$. To the best of our knowledge, this is the first paper to consider such generalizations in the literature. Note that the generalizations of (2) relative to (1) are not trivial in both practice and estimation. In practice, if $\beta_1 \neq \beta_2$ it is very unlikely that $\alpha_{1i} = \alpha_{2i}$ or that the same individual characteristics persist in decisions (i.e., the determination of y_{it}) across regimes. Of course, α_{1i} and α_{2i} should be correlated because the same individual is making decisions in the two regimes; for example, $\alpha_{1i} = \alpha_i + \eta_{1i}$ and $\alpha_{2i} = \alpha_i + \eta_{2i}$ with η_{1i} and η_{2i} being independent. Similarly, if there is a threshold effect in the conditional mean of y_{it}, it is natural to assume there is also a threshold effect in the conditional variance of y_{it}, i.e., $\sigma_1 \neq \sigma_2$. In estimation, the threshold effect in the variance of u_{it} is allowed by extending both strands of literature, but the threshold effect in α_i can be fixed only in the second strand. More especially, the demeaning or first-differencing methods mentioned above in the setup (1) cannot generate a consistent estimator of θ since α_{1i} and α_{2i} cannot be eliminated; for example, even in the simple setup of α_{1i} and α_{2i} above, usual demeaning or first-differencing operations cannot eliminate endogeneity because η_{1i} and η_{2i} are correlated with $\mathbf{x}_{it}$.[3]

A natural way to eliminate α_{1i} and α_{2i} might seem to take within-regime demeaning (WRD) in SPTR and within-regime first-differencing in DPTR. But, as Section 2 shows, such operations followed by least squares regression cannot generate consistent estimation. To obtain consistent estimators of θ in both SPTR and DPTR, this paper instead suggests the use of correlated random effects (CRE) models. Although CRE models are more restrictive in forms of endogeneity compared with fixed effects models, they provide a unified method valid for both SPTR and DPTR regardless of whether $\alpha_{1i} = \alpha_{2i}$. The present paper therefore provides an interesting example where the fixed effects estimator is not consistent whereas the CRE estimator is, which illustrates an unnoticed trade-off in traditional linear panel modeling between these two kinds of estimators – the former allows for more general forms of endogeneity but is less robust to the method of eliminating the endogeneity than the latter. Thus, because the setup (1) allows for both fixed-effects models and CRE models (just as traditional linear panel models), whereas the setup (2) allows for only CRE models, this paper points out a distinguishing feature of PTR compared to traditional linear panel data models, a feature as yet unnoticed in

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

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

[3] See Appendix G for finite sample evidence that the demeaning estimator in SPTR is biased when $\alpha_{1i} \neq \alpha_{2i}$.

the literature. As in Hansen (1999) and Ramírez-Rondán (2020), we use the small-threshold-effect framework of Hansen (2000) in developing our asymptotic results.

Section 3 discusses model selection (i.e., determination of the number of regimes) in a multiple regime PTR model. Our model selection is based on sequential testing and a score test statistic is used. Compared with the likelihood ratio (LR) statistic in the literature, our score test statistic is computationally more efficient. For consistency in model selection, we need to deliberately choose critical values such that the probabilities of both type-I and type-II errors shrink to zero. The literature usually specifies only an expansion rate for the critical values (e.g., $\log n$), whereas here a simulation method is suggested to obtain critical values. Since controlling both types of errors is not practical, our approach controls only the type-I error and is flexible regarding type-II error, so that the number of regimes may be underestimated. Section 4 develops asymptotic theory for the least squares estimator (LSE) that is robust to such potential misspecification; and the limit theory is new in the literature. We also provide a consistent estimator of the nuisance parameters in the asymptotic distribution and show that a popular estimator used in correctly specified models is invalid in this setting. A corollary of the asymptotic theory provides a test for unobserved individual-specific threshold effects, i.e., whether $\alpha_{1i} = \alpha_{2i}$ in (2).

Section 5 considers methods of inference for γ and β which are robust to the above-mentioned model misspecification. For γ, we invert the LR statistic to construct a confidence interval (CI), as suggested in Hansen (1999, 2000). For β, although the t-statistic could be inverted to construct a CI as in Hansen (1999), such a CI neglects the impact of γ estimation in finite samples (asymptotically, $\widehat{\beta}$ has the same distribution as in the case where γ_0 is known, where use of the subscript 0 on a parameter signifies its true value). Hansen (2000) suggests a Bonferroni-type CI where the coverage for γ is arbitrarily chosen and the coverage for β is fixed at the target level. We suggest a projection CI based on the LR statistic of $(\gamma, \beta')'$, which can be interpreted as an adaptive Bonferroni CI with the coverages for γ and β adaptively chosen. We also suggest two alternative LR statistics to construct CIs for γ and β; these LR statistics exclude the indirect effects of the null hypotheses. It turns out that the new LR statistics for γ have the same asymptotic null distribution as the original one, while those for β do not, which shows the sharp difference in the nature of γ and β. Section 6 applies the model selection, estimation and inference methodology developed in this paper to an empirical application to illustrate its usefulness in practice, and Section 7 concludes. Because the discussions on DPTR are very similar to those in SPTR, we collect them in Appendix A to avoid repetition in the main text. Proofs, calculation details, and simulation results are collected in the other six appendices.

For convenient reference we detail some further notation here. In SPTR, we observe $\{(y_{it}, \mathbf{x}'_{it})_{i=1}^N\}_{t=1}^T$ and $\{\underline{\mathbf{z}}_i\}_{i=1}^N$, and in DPTR, we observe $\{(y_{it}, \mathbf{x}'_{it})_{i=1}^N\}_{t=0}^T$ and $\{\underline{\mathbf{z}}_i\}_{i=1}^N$, i.e., the panel is balanced without missing data, where $\underline{\mathbf{z}}_i$ contains the time-invariant variables such as the constant 1. The number of observations used in estimation is $n = NT$. Throughout the paper, we let N diverge to infinity and T is fixed, i.e., the panel is short. In SPTR, $\mathbf{X}_i := (\mathbf{x}'_{i1}, \cdots, \mathbf{x}'_{iT}, \underline{\mathbf{z}}'_i)'$, and in DPTR, $\mathbf{X}_i := (\underline{\mathbf{x}}'_{i0}, \underline{\mathbf{x}}'_{i1}, \cdots, \underline{\mathbf{x}}'_{iT}, \underline{\mathbf{z}}'_i)'$ and $\mathbf{X}^t_i := (\mathbf{X}'_i, y_{i,t-1}, y_{i,t-2}, \cdots, y_{i0})'$, $t = 1, \cdots, T$. For any vector $\mathbf{z}$, $d_{\mathbf{z}}$ is the dimension of $\mathbf{z}$. For any matrix or vector A, $\|A\|$ denotes its Euclidean norm, and $A > 0$ means the matrix A is positive definite. For two real numbers a and b, $a \vee b := \max(a, b)$ and $a \wedge b := \min(a, b)$. $\approx$ means higher order terms are omitted. A bold zero, $\mathbf{0}$, represents a zero vector of suitable dimension for the context. The symbol ℓ is used to indicate the two regimes in (2) or the regimes in a more general $(m+1)$-regime model with $m \geq 1$ and, to simplify notation in what follows, the explicit values "$\ell = 1, 2$" or "$\ell = 1, 2, \cdots, m+1$" are often omitted.

2. Difficulties in applying WRD and setup of the CRE model

We show that the estimator based on WRD in SPTR is not consistent in general and explain the reason for inconsistency, thereby motivating the framework for a CRE model that enables consistent estimation.

2.1. Inconsistency of the WRD estimator

In the fixed effects SPTR model (2), the least squares objective function is

$$\sum_{i=1}^{n} \sum_{t=1}^{T} \left[\left(y_i - \mathbf{x}'_{it}\beta_1 - \alpha_{1i}\right)^2 1\left(q_{it} \leq \gamma\right) + \left(y_i - \mathbf{x}'_{it}\beta_2 - \alpha_{2i}\right)^2 1\left(q_{it} > \gamma\right) \right], \tag{3}$$

and, conditional on γ, we run two fixed effects linear panel regressions. As a result, concentrating out α_{1i} and α_{2i} is equivalent to the WRD operation. Specifically, we use

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

to eliminate $\alpha_{\ell i}$, where $\widetilde{y}^-_{it}(\gamma)$ is the residual of y_{it} regressed on $1\left(q_{it} \leq \gamma\right)$ among the observations with $q_{it} \leq \gamma$, and $\widetilde{y}^+_{it}(\gamma)$ can be similarly understood. Note here that only if $D^-_i(\gamma) := \sum_{t=1}^T 1\left(q_{it} \leq \gamma\right) \geq 1$ and $D^+_i(\gamma) := \sum_{t=1}^T 1\left(q_{it} > \gamma\right) \geq 1$, WRD can be implemented, and when $D^{\pm}_i(\gamma) = 1$, $\widetilde{y}^{\pm}_{it}(\gamma) = 0$. Notice also that only when $\gamma = \gamma_0$,

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

with

$$\widetilde{y}_{it}\left(\gamma_0\right) - \mathbb{E}\left[\widetilde{y}_{it}\left(\gamma_0\right) | \mathbf{X}_i\right] = \widetilde{u}^-_{1it}\left(\gamma_0\right) 1\left(q_{it} \leq \gamma_0\right) + \widetilde{u}^+_{2it}\left(\gamma_0\right) 1\left(q_{it} > \gamma_0\right).$$

As a result, we would expect the extremum estimator

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

with

$$S_n(\gamma,\beta) = \sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\left[\widetilde{y}_{it}^{-}(\gamma) - \widetilde{\mathbf{x}}_{it}^{-}(\gamma)'\beta_1\right]^2 1\left(q_{it} \le \gamma\right) + \left[\widetilde{y}_{it}^{+}(\gamma) - \widetilde{\mathbf{x}}_{it}^{+}(\gamma)'\beta_2\right]^2 1\left(q_{it} > \gamma\right)\right\} \tag{4}$$

to be a consistent estimator of $\left(\gamma_0, \beta_0\right)$, where $\mathbf{X}_i = \left(\mathbf{x}_{i1}', \cdots, \mathbf{x}_{iT}'\right)'$ with $\mathbf{x}_{it} = \left(x_{it}', q_{it}\right)'$, and $\widetilde{\mathbf{x}}_{it}^{\pm}(\gamma)$ and $\widetilde{u}_{\ell it}^{\pm}(\gamma)$ are defined in the same way as $\widetilde{y}_{it}^{\pm}(\gamma)$. Interestingly, the response variable $\widetilde{y}_{it}(\gamma)$ involves the unknown parameter γ, which makes estimation similar to that involving a Box-Cox transformation of the original response variable. The new aspect here is that γ is the parameter of interest rather than a nuisance parameter as it is in the Box-Cox transformation. This introduction of γ into the demeaning operation is what makes $\hat{\gamma}$ (and consequently $\hat{\beta}$) inconsistent.

For illustration, consider the following simple example

$$\begin{aligned} y_{it} &= \left(\alpha_{1i} + \sigma_1 u_{it}\right)1\left(q_{it} \le \gamma\right) + \left(\alpha_{2i} + \sigma_2 u_{it}\right)1\left(q_{it} > \gamma\right) \\ &=: \alpha_{2i} + \sigma_2 u_{it} + \left(\delta_{\alpha i} + \delta_\sigma u_{it}\right)1\left(q_{it} \le \gamma\right), \\ i &= 1, \cdots, N, t = 1, \cdots, T, \end{aligned} \tag{5}$$

where $u_{i1}, \cdots, u_{iT}, q_{i1}, \cdots, q_{iT}$ and $\left\{\alpha_{\ell i}\right\}_{\ell=1}^{2}$ are independent of each other, $\delta_{\alpha i} = \alpha_{1i} - \alpha_{2i}$, and $\delta_\sigma = \sigma_1 - \sigma_2$. To be more specific, assume the u_{it} follow $N(0,1)$, all q_{it} follow $U[0,1]$, α_{2i} may be correlated with $\left\{q_{it}\right\}_{t=1}^{T}$ and can follow any distribution, $\alpha_{1i} = \delta_\alpha + \alpha_{2i}$, $\gamma_0 = 0.7$, $\sigma_{20} = 1$ and $T = 5$. We vary δ_σ and δ_α to check the variation of the probability limit of the objective function. In this simple model, the only unknown parameter is γ; also, the error variance depends only on $1(q_{it} \le \gamma)$, but all testing and estimation procedures throughout the paper allow any form of conditional heteroskedasticity.

When $\gamma < \gamma_0$,

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

and after some tedious calculations detailed in Appendix D,

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

where $\widetilde{1}_{it}\left(\gamma, \gamma_0\right) = 1(\gamma < q_{it} \le \gamma_0) - \frac{\sum_{\tau=1}^{T} 1(\gamma < q_{i\tau} \le \gamma_0)}{\sum_{\tau=1}^{T} 1\left(q_{i\tau} > \gamma\right)}$, and $\widetilde{u}_{it}\left(\gamma, \gamma_0\right) = u_{it}1(\gamma < q_{it} \le \gamma_0) - \frac{\sum_{\tau=1}^{T} u_{i\tau}1(\gamma < q_{i\tau} \le \gamma_0)}{\sum_{\tau=1}^{T} 1\left(q_{i\tau} > \gamma\right)}$. Because $\mathbf{x}_{it}$ is not present in (5), the objective function (4) reduces to

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

In Appendix D, we show that the probability limit of $S_n(\gamma)/n$ is

$$S(\gamma) = T_1(\gamma) + T_2(\gamma) + T_3(\gamma),$$

where $T_1(\gamma)$ is from the $\delta_\alpha \widetilde{1}_{it}\left(\gamma, \gamma_0\right)$ term in $\widetilde{y}_{it}^{+}(\gamma)$, $T_2(\gamma)$ is from the $\widetilde{u}_{it}^{\pm}(\gamma)$ term in $\widetilde{y}_{it}^{\pm}(\gamma)$, and $T_3(\gamma)$ is from the $\delta_\sigma \widetilde{u}_{it}\left(\gamma, \gamma_0\right)$ in $\widetilde{y}_{it}^{+}(\gamma)$. The appearance of $T_2(\gamma)$ and $T_3(\gamma)$ is due completely to the dependence of the WRD on γ. $T_1(\gamma)$ would disappear if $\delta_\alpha = 0$, and $T_3(\gamma)$ would disappear if $\delta_\sigma = 0$. When $\gamma \ge \gamma_0$, $S(\gamma)$ has a similar structure.

Fig. 1 shows $S(\gamma)$ and its three components for various δ_σ and δ_α values. First, in the typical case of TR where the dependent variable does not depend on γ, only $T_1(\gamma)$ would appear, which has a kink due to the threshold effect in $\alpha_{\ell i}$ and is indeed minimized at γ_0. Second, when $\delta_\sigma = 0$, $T_3 = 0$ but T_1 still appears. When $\delta_\alpha = 0.3$ (i.e., the threshold effect in $\alpha_{\ell i}$ is small), T_1 is dominated by T_2 and the minimizer of $S(\gamma)$ is not γ_0, i.e., $\hat{\gamma}$ is inconsistent. Third, when $\delta_\sigma = 0$ and $\delta_\alpha = 1$, although T_2 has a much larger value than T_1, it is quite flat such that the minimizer of $S(\gamma)$ is indeed γ_0. Fourth, when $\delta_\sigma = 1$, T_3 also appears; $\delta_\alpha = 1$ is not large enough to make the minimizer of $S(\gamma)$ be γ_0 while $\delta_\alpha = 3$ is large enough to do so. In summary, the extra terms T_2 and T_3 introduced by the WRD can take unexpected shapes, which makes the WRD estimator inconsistent. In this simple example, for a larger threshold effect in error variance, we need a larger threshold effect in $\alpha_{\ell i}$ to make $\hat{\gamma}$ consistent and thus in general $\hat{\gamma}$ is inconsistent.

In Appendix D, cases with $T = 2$ or $\gamma_0 = 0.5$ are studied and the conclusions still apply. The case with $T = 2$ is especially interesting because $\widetilde{y}_{it}^{-}(\gamma) = 0$ ($\widetilde{y}_{it}^{+}(\gamma) = 0$) if $D_i^{-}(\gamma) = 0$ or 1 ($D_i^{+}(\gamma) = 0$ or 1), i.e., either $\widetilde{y}_{it}^{-}(\gamma)$ or $\widetilde{y}_{it}^{+}(\gamma)$ or both in $S_n(\gamma)$ is zero. In other words, the identification information of γ does not originate from the contrast between the two regimes within groups but from that between groups. In the case with $\gamma_0 = 0.5$, $\arg\min_\gamma S(\gamma) = \gamma_0$ if $\delta_\sigma = 0$ because $\arg\min_\gamma T_2(\gamma) = \gamma_0$ due to the symmetric density of q.

The arguments above hinge critically on the assumption that T is finite; if $T \to \infty$, denote the limit of $S(\gamma)$ as $S_\infty(\gamma)$, and then $\arg\min_\gamma S_\infty(\gamma)$ is indeed γ_0.[4] When $T \to \infty$, note that $P\left(D_i^{\pm}(\gamma) \ge 1\right) \to 1$, i.e., we can always conduct the WRD in both regimes for any individual. Fig. 2 shows $S_\infty(\gamma)$ and its three components $T_{1\infty}, T_{2\infty}$ and $T_{3\infty}$ in the same setup as Fig. 1. Interestingly, $T_2(\gamma) + T_3(\gamma)$ converges to a constant, so that $\arg\min_\gamma S_\infty(\gamma) = \arg\min_\gamma T_{1\infty}(\gamma) = \gamma_0$.

[4] Here, we implicitly assume that the sequential convergence and joint convergence limits as $N \to \infty$ and $T \to \infty$ are the same; for rigorous conditions to guarantee such an equivalence, see Phillips and Moon (1999).

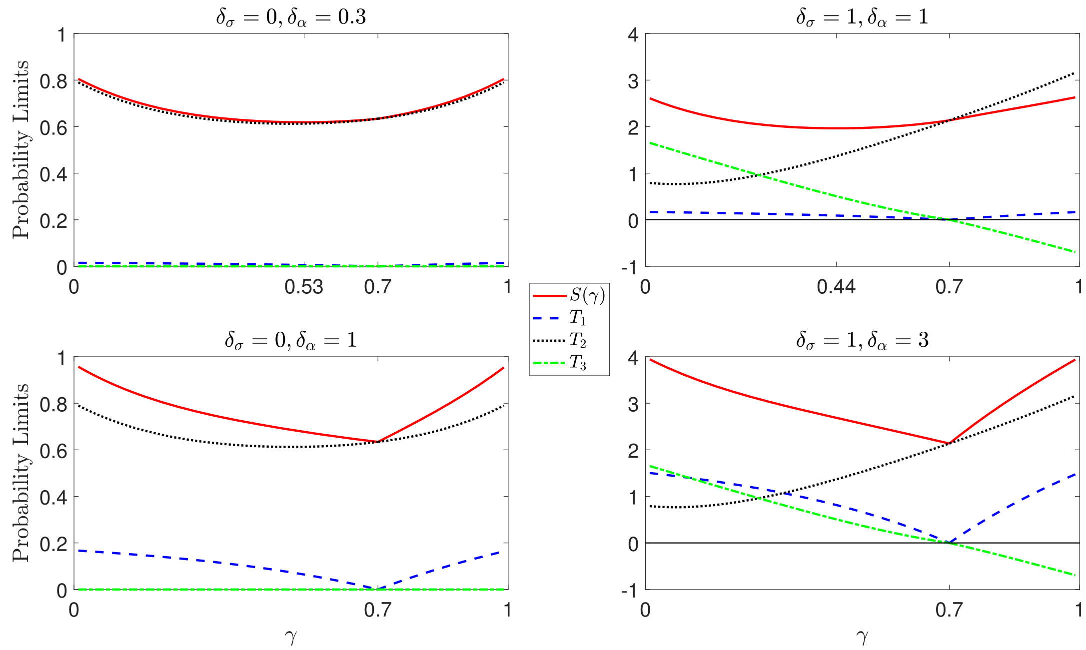

Fig. 1. $S(\gamma)$ and its three components for various δ_σ and δ_α values: $\sigma_{20} = 1$, $\gamma_0 = 0.7$, $T = 5$.

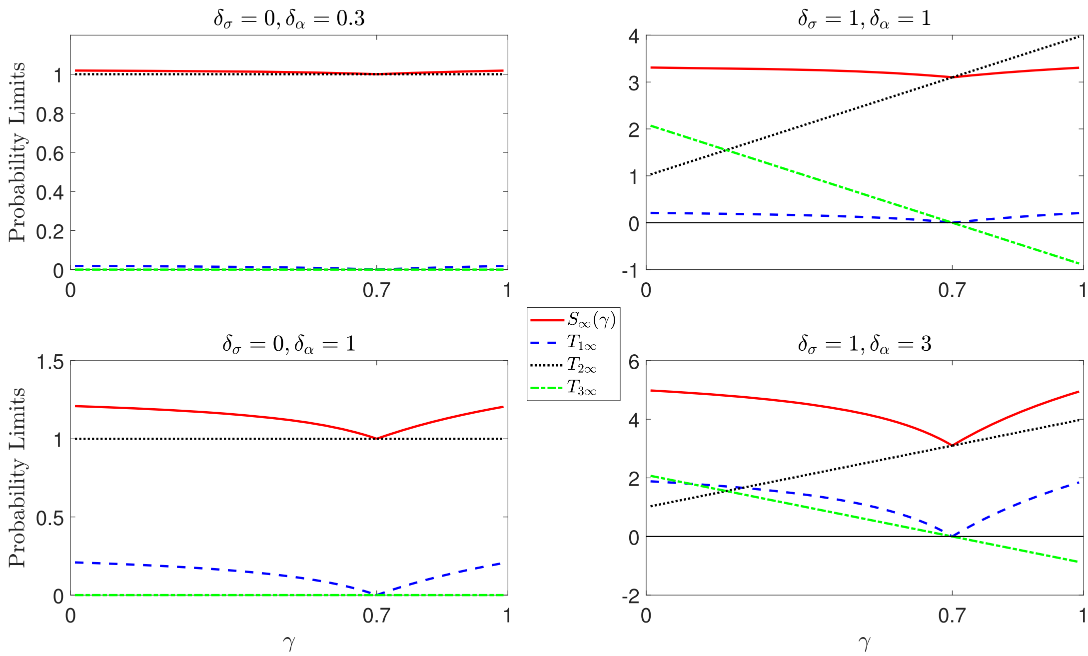

Fig. 2. $S_\infty(\gamma)$ and its three components for various δ_σ and δ_α values: $\sigma_{20} = 1$, $\gamma_0 = 0.7$.

2.2. Discussion

The panel structural change model can be treated as a special case of SPTR with $q_{it} = q_t \sim U[0,1]$ and independent of other covariates and the errors. In Appendix D, we take the same setup as in (5) and show that although $S(\gamma)$ can be greatly simplified, $\arg\min_\gamma S(\gamma)$ need not be γ_0 because $T_2(\gamma)$ and/or $T_3(\gamma)$ do not disappear. Actually, the setup that T is fixed is not appropriate in the panel structural change model because γ is defined only if $T \to \infty$. However, as $T \to \infty$, $S(\gamma)$ takes the same form as in the above SPTR model so that $\arg\min_\gamma S_\infty(\gamma) = \gamma_0$ indeed holds.

If the extra randomness of q_{it} in SPTR is not the reason for the failure of WRD, what is the true reason? We claim that if an M-estimation procedure after a demeaning operation, like the least squares estimation in the last subsection, is applied, then the demeaning cannot depend on unknown parameters, regardless of a particular parameter being nonregular like γ or regular like β. On the other hand, if a Z-estimation procedure (i.e., GMM) is applied after a demeaning operation, then the parameter can indeed be identified; see Bonhomme (2012). To justify this claim, we consider a simple linear panel data model

$$y_{it} = x_{it}\beta + \alpha_i + u_{it}, i = 1, \cdots, N, t = 1, \cdots, T, \tag{6}$$

wherein, like (5), it is assumed that $u_{i1}, \cdots, u_{iT}, x_{i1}, \cdots, x_{iT}$ are independent of each other, all the u_{it} follow $N(0,1)$, all the x_{it} follow $U[0,1]$, α_i may be correlated with $\{x_{it}\}_{t=1}^T$ and can follow any distribution, $T = 5$, and $\beta_0 = 0.7$. The parameter of interest is β, which is regular.

In Appendix D, we conduct a similar demeaning operation as in $\widetilde{y}_{it}(\gamma)$ but the operation depends on the regular parameter β instead of the nonregular parameter γ now. By setting up the least squares objective function $S_n(\beta)$, which is similar to $S_n(\gamma)$, we show that the probability limit of $S_n(\beta)/n$ as $N \to \infty$ can be written as

$$S(\beta) = T_1(\beta) + T_2(\beta) := D(\beta)\left(\beta - \beta_0\right)^2 + T_2(\beta),$$

for some functions $D(\beta)$ and $T_2(\beta)$. The left panel of Fig. 3 shows $S(\beta)$, $T_1(\beta)$ and $T_2(\beta)$. Although $\arg\min_\beta T_1(\beta) = 0.7 = \beta_0$, $\arg\min_\beta S(\beta) = 0.52 \neq \beta_0$. From Fig. 3, $\arg\min_\beta S(\beta)$ is driven mainly by $T_2(\beta)$, and the appearance of $T_2(\beta)$ arises for a similar reason as $T_2(\gamma)$ in the last subsection, i.e., the demeaning depends on the unknown parameter β.

The arguments above also clarify why Z-estimation works even if the demeaning involves unknown parameters while M-estimation does not. Suppose we denote the demeaning (or a general transformation) as $\Delta\left(\mathbf{y}_i, \mathbf{x}_i; \beta\right)$ in (6), and it satisfies

$$\mathbb{E}\left[\Delta\left(\mathbf{y}_i, \mathbf{x}_i; \beta_0\right)|\mathbf{x}_i\right] = \int \Delta\left(\mathbf{y}_i, \mathbf{x}_i; \beta_0\right) f\left(\mathbf{y}_i|\mathbf{x}_i; \beta_0\right) d\mathbf{y}_i = 0,$$

where $\mathbf{y}_i = \left(y_{i1}, \cdots, y_{iT}\right)'$, $\mathbf{x}_i = \left(x_{i1}, \cdots, x_{iT}\right)'$, and $f\left(\mathbf{y}_i|\mathbf{x}_i; \beta_0\right) = \int f\left(\mathbf{y}_i|\mathbf{x}_i, \alpha_i; \beta_0\right) f\left(\alpha_i|\mathbf{x}_i\right) d\alpha_i$ with $f\left(\alpha_i|\mathbf{x}_i\right)$ being the conditional density of α_i given $\mathbf{x}_i$. This moment condition implies that α_i is differenced out. It is clear that even if $\Delta\left(\mathbf{y}_i, \mathbf{x}_i; \beta\right)$ involves β, this moment condition can still identify β_0, noting that β_0 appears in both $\Delta\left(\mathbf{y}_i, \mathbf{x}_i; \beta_0\right)$ and $f\left(\mathbf{y}_i|\mathbf{x}_i; \beta_0\right)$. On the other hand, although

$$\frac{dT_1\left(\beta_0\right)}{d\beta} := \left.\frac{dT_1(\beta)}{d\beta}\right|_{\beta=\beta_0} = \left.\left[D'(\beta)\left(\beta_0 - \beta\right)^2 + 2D(\beta)\left(\beta - \beta_0\right)\right]\right|_{\beta=\beta_0} = 0,$$

$\frac{dT_2(\beta_0)}{d\beta}$ need not be zero so that $\frac{dS(\beta_0)}{d\beta}$ need not be zero, where in fact $\frac{dS(\beta)}{d\beta} = 0$ is the moment condition to identify β_0. So the problem here arises from an extra term introduced by demeaning that leads to the moment condition not holding at β_0. In PTR, $T_1(\gamma)$ is not differentiable, but the discussion here helps to shed light on the source of inconsistency.

Like $S_\infty(\gamma)$ in the last subsection, we also check the behavior of $S_\infty(\beta)$, the limit of $S(\beta)$ as $T \to \infty$, in Appendix D. Because the limit of $T_2(\beta)$ does not depend on β in the limit, $\arg\min_\beta S_\infty(\beta) = \beta_0$ indeed holds, which can be verified in the right panel of Fig. 3. We further make the demeaning independent of β but depend only on a constant c in Appendix D. In this case, $T_2(\beta)$ becomes a constant $T_2(c)$ such that $\arg\min_\beta S^c(\beta) = \beta_0$, as shown in the right panel of Fig. 3, where $S^c(\beta)$ is the counterpart of $S(\beta)$ when the demeaning depends only on c.

In summary, when the demeaning depends on unknown parameters, subsequent least squares (or general M-estimation procedure) will contain extra terms in the limit which render estimation inconsistent. In the linear panel model, the demeaning need not depend on β to eliminate α_i so that a consistent estimator can be easily constructed after the usual demeaning operation, while in PTR, the demeaning must depend on γ to eliminate α_{1i} and α_{2i} such that inconsistency is inevitable (we are not aware of any demeaning operation which can eliminate α_{1i} and α_{2i} simultaneously and does not depend on γ), which we believe is the key feature of PTR that distinguishes it from the linear panel model.

2.3. Setup of the CRE model

Given the inconsistency of the WRD estimator, we may consider the nonparametric PTR model as suggested in YP. Assuming $\mathbb{E}\left[\alpha_{\ell i}|\mathbf{X}_i\right] = h_\ell\left(\mathbf{X}_i\right)$, we have

$$\begin{aligned}\mathbb{E}\left[y_{it}|\mathbf{X}_i\right] &= \left(\mathbf{x}_{it}'\beta_1 + h_1\left(\mathbf{X}_i\right)\right)1\left(q_{it} \leq \gamma_0\right) + \left(\mathbf{x}_{it}'\beta_2 + +h_2\left(\mathbf{X}_i\right)\right)1\left(q_{it} > \gamma_0\right) \\ &=: \widetilde{h}_1\left(\mathbf{X}_i\right)1\left(q_{it} \leq \gamma_0\right) + \widetilde{h}_2\left(\mathbf{X}_i\right)1\left(q_{it} > \gamma_0\right),\end{aligned} \tag{7}$$

and can apply the IDKE to estimate γ, where $\mathbf{x}_{it}'\beta_\ell$ is absorbed in $\widetilde{h}_\ell\left(\mathbf{X}_i\right)$. Note that this method can identify γ but not β_ℓ; this is also why we cannot identify β using such a method in the linear panel data model. However, because the dimension of $\mathbf{X}_i$ is $T \cdot d_{\mathbf{x}} + d_{\underline{\mathbf{z}}}$ which is usually high, the formulation suffers from the curse of dimensionality: for example, in our empirical application of Section 6, $\dim\left(\mathbf{X}_i\right) = 42$ while $N = 565$ only. As a result, we must impose some constraints on $\widetilde{h}_\ell$ to make estimation practical. One approach is to impose sparsity on the series expansion of $\widetilde{h}_\ell$ and then use the lasso to identify signal terms. However, we adopt a different method, suggested by Chamberlain and Moreira (2009), to make our estimation and inferences feasible. Specifically, in the fixed effects model (2) with $u_{\ell it}$ independently and normally distributed, we impose an invariant prior distribution on the nuisance parameters $\left\{\alpha_{\ell i}\right\}_{i=1}^N$. Consequently, the integrated likelihood function coincides with the likelihood function for a normal CRE model. In other words, the fixed effects model simplifies to the CRE model under a family of special invariant priors on the nuisance parameters.

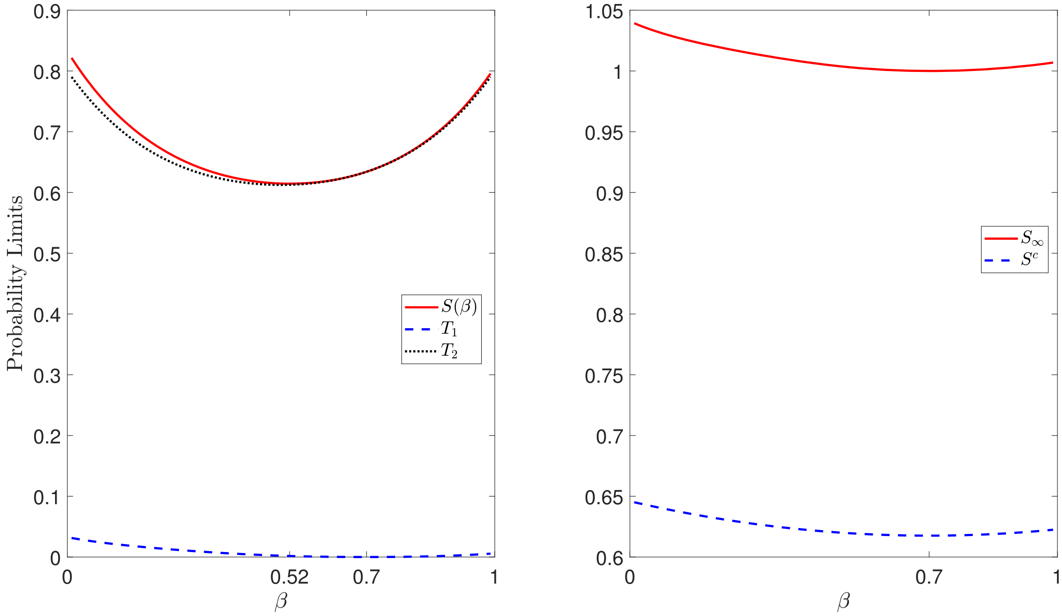

Fig. 3. $S(\beta)$, its two components $T_1(\beta)$ and $T_2(\beta)$, $S_\infty(\beta)$ and $S^c(\beta)$: $\beta_0 = 0.7$, $c = 0.4$, $T = 5$.

Our CRE model details are formally summarized in the following assumption.

Assumption 0. The response y_{it} satisfies

$$y_{it} = \sum_{\ell=1}^{m+1} \left(\mathbf{x}_{it}'\beta_\ell + \alpha_{\ell i} + u_{\ell it}\right)1(\gamma_{\ell-1} < q_{it} \le \gamma_\ell) \quad \text{with } \alpha_{\ell i} = \mathbf{z}_i'\psi_\ell + a_{\ell i}, \tag{8}$$

where $\mathbf{z}_i' = \left(\overline{\mathbf{x}}_i', \underline{\mathbf{z}}_i'\right)$, $\overline{\mathbf{x}}_i = \frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_{it}$, $\mathbb{E}\left[a_{\ell i}|\mathbf{X}_i\right] = 0$ and $\mathbb{E}\left[u_{\ell it}|\mathbf{X}_i\right] = 0$ with $\mathbf{X}_i := \left(\mathbf{x}_{i1}', \cdots, \mathbf{x}_{iT}', \underline{\mathbf{z}}_i'\right)'$, $0 \le m \le M$ for some integer $M \ge 1$, $\gamma_0 = -\infty$, $\gamma_{m+1} = \infty$, and

$$(\gamma_1, \cdots, \gamma_m) \in \Gamma^m := \left\{(\gamma_1, \cdots, \gamma_m) | -\infty < \underline{\gamma} \le \gamma_1 < \cdots < \gamma_m \le \overline{\gamma} < \infty\right\}$$

with $\Gamma^1 = \Gamma := [\underline{\gamma}, \overline{\gamma}]$.

First, (2) is generalized to have $m+1$ regimes where m can exceed unity and with $m = 0$ reducing the system to the linear panel data model. Second, we use Mundlak (1978)'s CRE device to control the endogeneity, and a more flexible specification as in Chamberlain (1982, 1984) is possible, e.g., $\alpha_{\ell i} = \psi_\ell'\mathbf{X}_i + a_{\ell i}$, although Mundlak's specification is maintained in this paper for simplicity. Essentially, $h_\ell\left(\mathbf{X}_i\right)$ in (7) is reduced to a linear form under the invariant prior, which greatly alleviates the curse of dimensionality and the computational burden of (7).[5] Third, the $a_{\ell i}$ can be correlated, but need not the same. In other words, correlation among the $\alpha_{\ell i}$ may be through either $\mathbf{z}_i$ or $a_{\ell i}$. When all the ψ_ℓ are equal and all the $a_{\ell i}$ are equal, our model reduces to the setup in Hansen (1999), i.e., all $\alpha_{\ell i}$'s are the same. Also, $a_{\ell i}$ can be correlated with $u_{\ell' it}$, where ℓ need not equal ℓ'. Fourth, in practice, we must guarantee that each regime contains at least ϵ% of the whole dataset for some $\epsilon > 0$ (typically, 10 or 15), so $\underline{\gamma}$ and $\overline{\gamma}$ are usually set as the lower and upper ϵ% quantile of q_{it}, respectively; also, $\left[\gamma_\ell, \gamma_{\ell+1}\right]$ contains at least ϵ% q_{it}'s for any $\ell = 1, \cdots, m-1$.

This device is a control function (CF) approach and in this respect relates to Yu et al. (2024) where a CF approach is used to handle endogeneity in cross-sectional threshold regression. In the linear panel data model, correlations between $\mathbf{x}_{it}$ and α_i that are present at time t are also present at other times and can therefore be fully revealed by a linear function of $\overline{\mathbf{x}}_i$. In effect, the $\mathbf{x}_{it}$'s in other periods can serve as control variables in period t. Wooldridge (2010) shows how the CRE approach applies to commonly used models, such as unobserved effects probit, fractional response, tobit, and count models. We show here that this approach can also apply to PTR with unobserved individual-specific threshold effects.

In the CRE model (8),

$$\mathbb{E}\left[y_{it}|\mathbf{X}_i\right] = \sum_{\ell=1}^{m+1}\left(\mathbf{x}_{it}'\beta_\ell + \mathbf{z}_i'\psi_\ell\right)1(\gamma_{\ell-1}^0 < q_{it} \le \gamma_\ell^0) =: \sum_{\ell=1}^{m+1}\breve{\mathbf{x}}_{it}'\theta_\ell 1(\gamma_{\ell-1}^0 < q_{it} \le \gamma_\ell^0) \tag{9}$$

[5] Even if $\mathbb{E}\left[\alpha_{\ell i}|\mathbf{X}_i\right] - \mathbf{z}_i'\psi_\ell$ is not zero, as long as it is $o(N^{-\kappa})$ in some sense, where κ is defined in the following Assumption 1, Yu (2019) shows that its effect on the LSE of γ can be neglected asymptotically. Note that the order $o(N^{-\kappa})$ may be still larger than $O\left(N^{-1/2}\right)$, which implies that γ estimation is more robust to misspecification than β estimation.

and the error term

$$e_{it}^0 = \sum_{\ell=1}^{m+1} \left(a_{\ell i} + u_{\ell it}\right) 1(\gamma_{\ell-1}^0 < q_{it} \le \gamma_\ell^0) =: \sum_{\ell=1}^{m+1} \varepsilon_{\ell it} 1(\gamma_{\ell-1}^0 < q_{it} \le \gamma_\ell^0), \tag{10}$$

where $\breve{\mathbf{x}}_{it} := \left(\mathbf{x}_{it}', \mathbf{z}_i'\right)'$, $\theta_\ell = \left(\beta_\ell', \psi_\ell'\right)'$, and we use the superscript 0 to signify the true value of γ_ℓ.[6] Often, we rewrite (9) in the following form,

$$\mathbb{E}\left[y_{it}|\mathbf{X}_i\right] = \sum_{\ell=1}^{m} \breve{\mathbf{x}}_{it}' \delta_\ell 1(q_{it} \le \gamma_\ell^0) + \breve{\mathbf{x}}_{it}' \theta_{m+1}, \tag{11}$$

where $\delta_\ell = \theta_\ell - \theta_{\ell+1}$. For estimation, a small-threshold-effect assumption is used, c.f., Hansen (2000).

Assumption 1. For some fixed $c_\ell = \left(c_{\ell\beta}', c_{\ell\psi}'\right)' \neq \mathbf{0}$, $\delta_\ell = c_\ell N^{-\kappa}$, $\ell = 1, \cdots, m$, where $0 < \kappa < 1/2$.

Note that (11) implies

$$\mathbb{E}\left[y_{it}|\mathbf{X}_i\right] = \sum_{\ell=1}^{m} \breve{\mathbf{x}}_{it}' \delta^\ell 1(\gamma_{\ell-1}^0 < q_{it} \le \gamma_\ell^0) + \breve{\mathbf{x}}_{it}' \theta_{m+1},$$

where $\delta^\ell = \sum_{l=\ell}^m \delta_\ell = N^{-\kappa} \sum_{l=\ell}^m c_\ell =: N^{-\kappa} c^\ell$.

The following three sections provide a formal analysis of the CRE model (8). Section 3 provides a model selection method to determine m based on sequential testing. Section 4 develops the asymptotic theory for the LSE when m could be underestimated. Section 5 gives new misspecification-robust inference methods for both the γ_ℓ and β_ℓ parameters.

3. Model selection based on sequential testing

First consider testing $m = 0$ versus $m = 1$, which is the usual test for linearity. To lighten the computational burden in the panel data scenario, we suggest using the score test in Yu (2013) and Yu and Fan (2021) instead of the usual Wald-type or LR-type tests. This test serves as the building block for general $m = l$ versus $m = l + 1$ tests, $l = 1, \cdots, M - 1$. The estimated m, say $\hat{m}$, is set as the first l that cannot be rejected in such a sequential testing procedure.

3.1. *Testing for linearity:* $m = 0$ *versus* $m = 1$

Under H_1, the CRE model can be expressed as

$$\mathbb{E}\left[y_{it}|\mathbf{X}_i\right] = \breve{\mathbf{x}}_{it}' \theta_1 1(q_{it} \le \gamma_0) + \breve{\mathbf{x}}_{it}' \theta_2 1(q_{it} > \gamma_0) = \breve{\mathbf{x}}_{it}' \theta_o + \breve{\mathbf{x}}_{it}' \delta_1 1\left(q_{it} \le \gamma_0\right),$$

where $\gamma_0 = \gamma_1^0$ in (9), $\theta_o = \theta_2$, and $\delta_1 = \theta_1 - \theta_2$. Now, $H_0 : m = 0$ is equivalent to $H_0 : \theta_1 = \theta_2$ or $\delta_1 = \mathbf{0}$ and $H_1 : m = 1$ is the negation of H_0.

If we estimate $\theta = \left(\gamma, \theta_o', \delta_1'\right)'$ based on least squares, the objective function is

$$S_n(\theta) = \sum_{i=1}^{N} \sum_{t=1}^{T} \left(y_{it} - \breve{\mathbf{x}}_{it}' \theta_o - \breve{\mathbf{x}}_{it}' \delta_1 1\left(q_{it} \le \gamma\right)\right)^2.$$

The corresponding score function with respect to δ_1, evaluated at $\mathbf{0}$, is

$$s_n(\gamma) = \sum_{i=1}^{N} \sum_{t=1}^{T} \breve{\mathbf{x}}_{it} e_{it}^o 1\left(q_{it} \le \gamma\right),$$

after discarding the constant terms, where e_{it}^o is the random error in projecting y_{it} on $\breve{\mathbf{x}}_{it}$, which equals e_{it}^0 under H_0 but includes a bias under H_1. e_{it}^o can be estimated by $\hat{e}_{it}^o = y_{it} - \breve{\mathbf{x}}_{it}' \hat{\theta}_o$, where $\hat{\theta}_o$ is the vector of coefficients in the linear regression of y_{it} on $\breve{\mathbf{x}}_{it}$, so it does not converge to θ_2 under H_1. In sum, our score test essentially tests the following moment conditions

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

The test statistics are based on

$$T_n(\gamma) = \hat{H}_n(\gamma)^{-1/2} \hat{m}_n(\gamma), \tag{12}$$

[6] If $\mathbf{x}_{it}$ does not contain q_{it} in (8), $\overline{\mathbf{x}}_i$ should include $\overline{q}_i$ because $\mathbb{E}\left[a_{\ell i}|\mathbf{X}_i\right] = 0$ and $\mathbf{X}_i$ contains $\left\{q_{it}\right\}_{t=1}^T$. This comment applies also to DPTR where $q_{it} \neq y_{i,t-1}$ and $q_{it} \notin \underline{\mathbf{x}}_{it}$.

where

$$\widehat{H}_n(\gamma) = N^{-1}\sum_{i=1}^{N}\widehat{\zeta}_i(\gamma)\widehat{\zeta}_i(\gamma)' \text{ and } \widehat{m}_n(\gamma) = N^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}\check{\mathbf{x}}_{it}\widehat{e}^{o}_{it}1\left(q_{it}\le\gamma\right),$$

with

$$\widehat{\zeta}_i(\gamma) = \sum_{t=1}^{T}\left(1\left(q_{it}\le\gamma\right)\check{\mathbf{x}}_{it} - \widehat{M}(\gamma)\widehat{M}^{-1}\check{\mathbf{x}}_{it}\right)\widehat{e}^{o}_{it} =: \sum_{t=1}^{T}\widehat{\zeta}_{it}(\gamma),$$

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

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

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

or the Cramér-von Mises average-type statistic

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

We will use $g_n^{(1)} = g\left(T_n\right)$ to denote either of these two functionals of $T_n(\gamma)$. One key advantage of our score test is that we need only run one least squares regression to construct the test statistic, while the usual Wald-type or LR-type tests need to run $O(n)$ regressions if we approximate the supremum in KS or the integral in C_vM using a discrete approximation of Γ, say $\Gamma_n := \left\{q_{it}|q_{it}\in\Gamma\right\}$. Because n is usually large in PTR, the score test greatly saves computational time.

We next derive the asymptotic distribution of $g_n^{(1)}$ under the local alternative, $H_{1c}^{(1)}: \delta_1 = N^{-1/2}c$. For this purpose, we impose the following assumptions.

Assumption 2.

(i) $\left\{\mathbf{x}_{it},\underline{\mathbf{z}}_i,y_{it}\right\}_{t=1}^{T}$ are i.i.d. across $i\in\{1,\cdots,N\}$; T is fixed and $N\to\infty$.
(ii) For $t=1,\cdots,T$, $\mathbb{E}\left[\left\|\check{\mathbf{x}}_{it}\right\|^4\right]<\infty$, and $\mathbb{E}\left[\left|e^0_{it}\right|^4\right]<\infty$, where $e^0_{it} = \varepsilon_{1it}1(q_{it}\le\gamma_0)+\varepsilon_{2it}1(q_{it}>\gamma_0)$ under H_1^c and which degenerates to $e_{it}(=\varepsilon_{1it}=\varepsilon_{2it})$ under H_0.
(iii) For $t=1,\cdots,T$, q_{it} has a continuous distribution;
(iv) $M > M(\gamma) > 0$ for all $\gamma\in\Gamma$, and $M(\gamma)$ is strictly increasing for $\gamma\in\Gamma$, where $M=\sum_{t=1}^{T}\mathbb{E}\left[\check{\mathbf{x}}_{it}\check{\mathbf{x}}_{it}'\right]$, and $M(\gamma) = \sum_{t=1}^{T}\mathbb{E}\left[\check{\mathbf{x}}_{it}\check{\mathbf{x}}_{it}'1\left(q_{it}\le\gamma\right)\right]$.
(v) $H(\gamma)=\mathbb{E}\left[\zeta_i(\gamma)\zeta_i(\gamma)'\right]>0$ for all $\gamma\in\Gamma$, where $\zeta_i(\gamma)=\sum_{t=1}^{T}(1\left(q_{it}\le\gamma\right)\check{\mathbf{x}}_{it}-M(\gamma)M^{-1}\check{\mathbf{x}}_{it})e^0_{it} =: \sum_{t=1}^{T}\zeta_{it}(\gamma)$.

Condition (i) is standard in panel data modeling where the asymptotics are taken over N in the cross section dimension rather than over T. This condition is assumed in all the literature mentioned in the Introduction. In Condition (ii), because $\check{\mathbf{x}}_{it} = \left(\mathbf{x}_{it}',\overline{\mathbf{x}}_i',\underline{\mathbf{z}}_i'\right)'$, $\mathbb{E}[\left\|\check{\mathbf{x}}_{it}\right\|^4]<\infty$ is implied by $\mathbb{E}[\left\|\mathbf{x}_{it}\right\|^4]<\infty$ and $\mathbb{E}[\left\|\underline{\mathbf{z}}_i\right\|^4]<\infty$. Under H_1^c, $\mathbb{E}[\left|e^0_{it}\right|^4]<\infty$ is implied by $\mathbb{E}\left|\varepsilon_{\ell it}\right|^4]<\infty$ which is in turn implied by $\mathbb{E}[\left|a_{\ell i}\right|^4]<\infty$ and $\mathbb{E}[\left|u_{\ell it}\right|^4]<\infty$. Under H_0, γ_0 is not identifiable, so the two error terms ε_{1it} and ε_{2it} are combined into one. Condition (iii) is a regularity condition to ensure that Glivenko-Cantelli and Donsker theorems can be applied. The assumption $M > M(\gamma) > 0$ in Condition (iv) excludes multicollinearity of $\check{\mathbf{x}}_{it}$ and restricts Γ to a proper subset of the joint support of $\left\{q_{it}\right\}_{t=1}^{T}$. The strict monotonicity of $M(\gamma)$ in Condition (iv) means that for any $\gamma_2>\gamma_1$, $M\left(\gamma_2\right)-M\left(\gamma_1\right)>0$, which requires $\sum_{t=1}^{T}f_t(\gamma)>0$ for all $\gamma\in\Gamma$, where $f_t(\gamma)$ is the density of q_{it} at γ; although this part of Condition (iv) is not required for all results of this subsection, it is indeed required for the results in the next subsection. Condition (v) ensures $\widehat{H}_n(\gamma)^{-1/2}$ is well defined when N is large enough.

Theorem 1. *Under Assumption 2 and $H_{1c}^{(1)}$,*

$$g_n^{(1)} \xrightarrow{d} g_c^{(1)} := g\left(T_c\right),$$

where

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

and $\Xi(\cdot)$ is a mean zero Gaussian process with covariance kernel

$$H(\gamma_1,\gamma_2) = \mathbb{E}\left[\zeta_i\left(\gamma_1\right)\zeta_i\left(\gamma_2\right)'\right].$$

When $c=\mathbf{0}$, we obtain the asymptotic null distribution of $g_n^{(1)}$ as $g_0^{(1)} := g\left(T_0\right)$, where $T_0(\gamma)=H(\gamma)^{-1/2}\Xi(\gamma)$. To show that our test rejects the null with probability approaching one under the small-threshold-effect assumption, we give the following corollary.

Corollary 1. *Under Assumption 1 with $m = 1$ and Assumption 2, $P\left(g_n^{(1)} > F_0^{(1)-1}(1-\alpha)\right) \to 1$ or $p_n^{(1)} \xrightarrow{p} 0$ for any $c_1 \neq \mathbf{0}$ and $0 < \kappa < 1/2$, where $F_0^{(1)}$ is the distribution of $g_0^{(1)}$, $F_0^{(1)-1}$ is its inverse function, α is the significance level, and $p_n^{(1)} = 1 - F_0^{(1)}\left(g_n^{(1)}\right)$.*

Because the asymptotic null distribution is not pivotal, we use the simulation method of Hansen (1996) to obtain the critical values or p -values. Specifically, let $\{\xi_i^*\}_{i=1}^N$ be i.i.d. $N(0,1)$ random variables, and set

$$T_n^*(\gamma) = \widehat{H}_n(\gamma)^{-1/2}\widehat{m}_n^*(\gamma) \text{ and } g_n^{(1)*} = g\left(T_n^*\right),$$

where

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

We reject H_0 if $p_n^{(1)*} = 1 - F_n^{(1)*}\left(g_n^{(1)}\right) \le \alpha$, and $F_n^{(1)*}$ is the conditional distribution of $g_n^{(1)*}$ given the original data. The construction of $T_n^*(\gamma)$ deserves further explanation. First, the extra randomness introduced by the simulation appears only in $\widehat{m}_n^*(\gamma)$ but not in $\widehat{H}_n(\gamma)$; essentially, this is a wild bootstrap procedure. Second, the same $\{\xi_i^*\}_{i=1}^N$ are used for all $\gamma \in \Gamma$. In practice, we can replace Γ by Γ_n mentioned above, which becomes dense in Γ as $N \to \infty$. Third, the same ξ_i^* is associated with $\widehat{e}_{it}^o$ for all $t = 1, \cdots, T$ to maintain the correlation structure between $\{e_{it}^o\}_{i=1}^T$. Fourth, the extra term $\widehat{M}(\gamma)\widehat{M}^{-1}\check{\mathbf{x}}_{it}$ in $\widehat{m}_n^*(\gamma)$ is critical and cannot be omitted although it is unnecessary in $\widehat{m}_n(\gamma)$. The following theorem shows asymptotic validity of this simulation procedure.

Theorem 2. *Under $H_{1c}^{(1)}$ and Assumption 2, $p_n^{(1)*} - p_n^{(1)} = o_p(1)$, which implies $p_n^{(1)*} \xrightarrow{d} 1 - F_0^{(1)}\left(g_c^{(1)}\right)$ under $H_{1c}^{(1)}$ and $p_n^{(1)*} \xrightarrow{d} U[0,1]$ under $H_0^{(1)}$. If Assumption 1 with $m = 1$ instead of $H_{1c}^{(1)}$ holds, then $p_n^{(1)*} \xrightarrow{p} 0$.*

3.2. Sequential testing

If $m = 0$ is rejected, the threshold point can be estimated using least squares. From (9) with $m = 1$, the following objective function is minimized

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

where we impose a normalization constant $1/2$ on $S_n(\theta)$ for notational convenience, $\theta = \left(\gamma, \beta', \psi'\right)'$ with $\beta' = \left(\beta_1', \beta_2'\right)$ and $\psi' = \left(\psi_1', \psi_2'\right)$, or $\theta = \left(\gamma, \underline{\theta}'\right)' := \left(\gamma, \theta_1', \theta_2'\right)'$. Denote the resulting estimate as $\widehat{\theta} = \left(\widehat{\gamma}, \widehat{\beta}', \widehat{\psi}'\right)'$ or $\left(\widehat{\gamma}, \widehat{\underline{\theta}}'\right)'$ and the residuals as

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

A two-step procedure is normally used to estimate γ. Specifically, for each $\gamma \in \Gamma$, run least squares of y_{it} on $\check{\mathbf{x}}_{it}$ for the (i,t) such that $q_{it} \le \gamma$ and $q_{it} > \gamma$ separately to obtain $\widehat{\theta}_1(\gamma)$ and $\widehat{\theta}_2(\gamma)$. The concentrated objective function is

$$S_n(\gamma) = S_n\left(\gamma, \widehat{\theta}_1(\gamma), \widehat{\theta}_2(\gamma)\right) =: S_n\left(\gamma, \widehat{\underline{\theta}}(\gamma)\right); \tag{14}$$

and then

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

and $\widehat{\theta}_\ell = \widehat{\theta}_\ell\left(\widehat{\gamma}\right)$. As suggested in Hansen (1999, Section 3.2), we need to search over $O(n)$ distinct q_{it} values (or fewer quantiles of $\{q_{it}\}$ if Γ contains too many q_{it} values) to estimate γ. Usually, there is an interval to minimize $S_n(\gamma)$. Although in the small-threshold-effect framework assumed in this paper, any point on the minimizing interval has the same asymptotic distribution, we follow Yu (2012, 2015a) to pick the mid-point as our estimator to improve the finite sample performance of $\widehat{\gamma}$.

Note that the true m may be greater than one, so this objective function is misspecified. Nevertheless, we show that $\widehat{\gamma}$ will converge to one of the m thresholds that minimizes the probability limit of $S_n(\gamma)$ (after some normalization), say $S(\gamma)$, under the following identification assumption.

Assumption 3. $^{(1)}$: $\arg\min_{\gamma\in\left\{\gamma_1^0,\cdots,\gamma_m^0\right\}} S(\gamma) = \gamma_0^{(1)}$ is unique.

Lemma 1. *Under Assumptions 1, 2(i)-(iii) and $3^{(1)}$, $\widehat{\gamma} \xrightarrow{p} \gamma_0^{(1)}$.*

Although $m \ge 1$, setting $m = 1$ in the objective function (13) still generates a threshold that dominates other thresholds. This result generalizes corresponding results in cross sectional threshold models in Gonzalo and Pitarakis (2002) (GP hereafter) and in structural change models in Bai (1997b) to the panel threshold case. The most interesting part of Lemma 1 is that $\widehat{\gamma}$ cannot converge to some middle point between γ_ℓ^0 and $\gamma_{\ell+1}^0$, $\ell = 1, \cdots, m-1$. To aid intuition, let $m = 2$, and consider why $\widehat{\gamma}$ cannot converge to some point on $\left(\gamma_1^0, \gamma_2^0\right)$. The reason is that splitting by $\widehat{\gamma}$ tries to maximize the (weighted) difference between θ_1 and θ_2 (see, e.g., Eq. (8) of GP). When the splitting point falls on $\left(\gamma_1^0, \gamma_2^0\right)$, say γ_o, the data for $q_{it} \in (\gamma_1^0, \gamma_o]$ and for $q_{it} \in (\gamma_o, \gamma_2^0]$ share the same slope coefficients, while

θ_1 is a weighted average of this slope and that for $q_{it} \in (-\infty, \gamma_1^0]$ and θ_2 is a weighted average of this slope and that for $q_{it} \in (\gamma_2^0, \infty)$, i.e., θ_1 and θ_2 share a common component such that their difference is not as sharp as in the case where the splitting point is set at γ_1^0 or γ_2^0.

Given $\hat{\gamma}$, we can continue to test $H_0^{(2)}: m=1$ vs. $H_1^{(2)}: m=2$ by applying the test in Section 3.1 to the two regimes defined by $\hat{\gamma}$. More specifically, define $g_n^{(2,1)} = g\left(T_n^{(2,1)}\right)$ and $g_n^{(2,2)} = g\left(T_n^{(2,2)}\right)$ as the g_n statistics applied to $\left\{(y_i, \check{\mathbf{x}}_{it}')|q_{it} \le \hat{\gamma}\right\}$ and $\left\{(y_i, \check{\mathbf{x}}_{it}')|q_{it} > \hat{\gamma}\right\}$ separately, and the ultimate test statistic is then defined as $g_n^{(2)} = \max\left\{g_n^{(2,1)}, g_n^{(2,2)}\right\}$, where

$$
\begin{aligned}
T_n^{(2,1)} &= \hat{H}_n^{(2,1)}(\gamma)^{-1/2}\hat{m}_n^{(2,1)}(\gamma),\ T_n^{(2,2)} = \hat{H}_n^{(2,2)}(\gamma)^{-1/2}\hat{m}_n^{(2,2)}(\gamma),\\
\hat{H}_n^{(2,1)}(\gamma) &= N^{-1}\sum_{i=1}^N \hat{\zeta}_i^{(2,1)}(\gamma)\hat{\zeta}_i^{(2,1)}(\gamma)',\ \hat{H}_n^{(2,2)}(\gamma) = N^{-1}\sum_{i=1}^N \hat{\zeta}_i^{(2,2)}(\gamma)\hat{\zeta}_i^{(2,2)}(\gamma)',\\
\hat{m}_n^{(2,1)}(\gamma) &= N^{-1/2}\sum_{i=1}^N\sum_{t=1}^T \check{\mathbf{x}}_{it}\hat{e}_{it}^{(2,1)}1\left(q_{it} \le \gamma\right),\ \hat{m}_n^{(2,2)}(\gamma) = N^{-1/2}\sum_{i=1}^N\sum_{t=1}^T \check{\mathbf{x}}_{it}\hat{e}_{it}^{(2,2)}1\left(\hat{\gamma} < q_{it} \le \gamma\right)
\end{aligned}
$$

with

$$
\begin{aligned}
\hat{\zeta}_i^{(2,1)}(\gamma) &= \sum_{t=1}^T \left(1\left(q_{it} \le \gamma\right)\check{\mathbf{x}}_{it} - \hat{M}(\gamma)\hat{M}_1^{-1}\check{\mathbf{x}}_{it}1\left(q_{it} \le \hat{\gamma}\right)\right)\hat{e}_{it}^{(2,1)} =: \sum_{t=1}^T \hat{\zeta}_{it}^{(2,1)}(\gamma),\\
\hat{\zeta}_i^{(2,2)}(\gamma) &= \sum_{t=1}^T \left(1\left(\hat{\gamma} < q_{it} \le \gamma\right)\check{\mathbf{x}}_{it} - \hat{M}\left(\hat{\gamma}, \gamma\right)\hat{M}_2^{-1}\check{\mathbf{x}}_{it}1\left(q_{it} > \hat{\gamma}\right)\right)\hat{e}_{it}^{(2,2)} =: \sum_{t=1}^T \hat{\zeta}_{it}^{(2,2)}(\gamma),\\
\hat{M}_1 &= \hat{M}\left(\hat{\gamma}\right),\ \hat{M}_2 = \hat{M} - \hat{M}_1 \text{ and } \hat{M}\left(\hat{\gamma}, \gamma\right) = \hat{M}(\gamma) - \hat{M}\left(\hat{\gamma}\right).
\end{aligned}
$$

Here $\hat{e}_{it}^{(2,1)}$ and $\hat{e}_{it}^{(2,2)}$ are the least squares residuals in the first and second regimes, and the parameter spaces Γ_1 and Γ_2 for $g_n^{(2,1)}$ and $g_n^{(2,2)}$ are similarly constructed as Γ, i.e., $\Gamma_1 = \left[\underline{\gamma}, \overline{\gamma}_1\right]$ and $\Gamma_2 = \left[\underline{\gamma}_2, \overline{\gamma}\right]$ with $\overline{\gamma}_1$ being the upper $\epsilon\%$ quantile of q_{it} in $\left\{q_{it}|q_{it} \le \hat{\gamma}\right\}$ and $\underline{\gamma}_2$ being the lower $\epsilon\%$ quantile of q_{it} in $\left\{q_{it}|q_{it} > \hat{\gamma}\right\}$. Without loss of generality, assume $\hat{\gamma} \xrightarrow{p} \gamma_1^0$; then the local alternative is changed to

$$
H_{1c}^{(2)}: \delta_2 := \theta_2 - \theta_3 = N^{-1/2}c. \tag{15}
$$

Also, Assumption 2(v) should adapt from $H(\gamma) > 0$ for all $\gamma \in \Gamma$ to

$$
\begin{aligned}
H^{(2,1)}(\gamma) &= \mathbb{E}\left[\zeta_i^{(2,1)}(\gamma)\zeta_i^{(2,1)}(\gamma)'\right] > 0 \text{ for all } \gamma \in \Gamma_1,\\
H^{(2,2)}(\gamma) &= \mathbb{E}\left[\zeta_i^{(2,2)}(\gamma)\zeta_i^{(2,2)}(\gamma)'\right] > 0 \text{ for all } \gamma \in \Gamma_2,
\end{aligned}
$$

where $\zeta_i^{(2,1)}(\gamma) = \sum_{t=1}^T (1\left(q_{it} \le \gamma\right)\check{\mathbf{x}}_{it} - M(\gamma)M_1^{-1}\check{\mathbf{x}}_{it}1\left(q_{it} \le \gamma_1^0\right))e_{it}^0 =: \sum_{t=1}^T \zeta_{it}^{(2,1)}(\gamma)$ with $M_1 = M\left(\gamma_1^0\right)$, and $\zeta_i^{(2,2)}(\gamma) = \sum_{t=1}^T (1\left(\gamma_1^0 < q_{it} \le \gamma\right)\check{\mathbf{x}}_{it} - M\left(\gamma_1^0, \gamma\right)M_2^{-1}\check{\mathbf{x}}_{it}1\left(q_{it} > \gamma_1^0\right))e_{it}^0 =: \sum_{t=1}^T \zeta_{it}^{(2,2)}(\gamma)$ with $M_2 = M - M_1$ and $M\left(\gamma_1^0, \gamma\right) = M(\gamma) - M\left(\gamma_1^0\right)$. The following corollary extends Theorem 1 and Corollary 1.

Corollary 2. *Under $H_{1c}^{(2)}$ and Assumption 2,*

$$
g_n^{(2)} \xrightarrow{d} g_c^{(2)} := \max\left\{g_0^{(2,1)}, g_c^{(2,2)}\right\} = \max\left\{g\left(T_0^{(2,1)}\right), g\left(T_c^{(2,2)}\right)\right\},
$$

where

$$
\begin{aligned}
T_0^{(2,1)}(\gamma) &= H^{(2,1)}(\gamma)^{-1/2}\Xi^{(2,1)}(\gamma), \gamma \in \Gamma_1,\\
T_c^{(2,2)}(\gamma) &= H^{(2,2)}(\gamma)^{-1/2}\left\{\Xi^{(2,2)}(\gamma) + \left[M\left(\gamma_1^0, \gamma \wedge \gamma_2^0\right) - M\left(\gamma_1^0, \gamma\right)M_2^{-1}M\left(\gamma_1^0, \gamma_2^0\right)\right]c\right\}, \gamma \in \Gamma_2,
\end{aligned}
$$

and $\Xi^{(2)}(\gamma) := \begin{cases} \Xi^{(2,1)}(\gamma), \text{ if } \gamma \in \Gamma_1,\\ \Xi^{(2,2)}(\gamma), \text{ if } \gamma \in \Gamma_2,\end{cases}$ *is a mean zero Gaussian processes with covariance kernel*

$$
H^{(2)}(\gamma_1, \gamma_2) = \begin{cases}
H^{(2,1)}(\gamma_1, \gamma_2) = \mathbb{E}\left[\zeta_i^{(2,1)}\left(\gamma_1\right)\zeta_i^{(2,1)}\left(\gamma_2\right)'\right], \text{ if } (\gamma_1, \gamma_2) \in \Gamma_1 \times \Gamma_1,\\
H^{(2,2)}(\gamma_1, \gamma_2) = \mathbb{E}\left[\zeta_i^{(2,2)}\left(\gamma_1\right)\zeta_i^{(2,2)}\left(\gamma_2\right)'\right], \text{ if } (\gamma_1, \gamma_2) \in \Gamma_2 \times \Gamma_2,\\
C^{(1,2)}(\gamma_1, \gamma_2) = \mathbb{E}\left[\zeta_i^{(2,1)}\left(\gamma_1\right)\zeta_i^{(2,2)}\left(\gamma_2\right)'\right], \text{ if } \gamma_1 \in \Gamma_1 \text{ and } \gamma_2 \in \Gamma_2.
\end{cases}
$$

If Assumption 1 with $m = 2$ instead of $H_{1c}^{(2)}$ holds, then $p_n^{(2)} \xrightarrow{p} 0$ for any $c_2 \ne \mathbf{0}$ and $0 < \kappa < 1/2$, where $p_n^{(2)} = 1 - F_0^{(2)}\left(g_n^{(2)}\right)$ with $F_0^{(2)}$ being the distribution of $g_0^{(2)}$.

Two features in this corollary deserve further discussion. First, only $T_c^{(2,2)}(\gamma)$ involves c under $H_{1c}^{(2)}$. This is because the first regime does not contain any threshold effect, so $T_n^{(2,1)}(\cdot)$ converges to the null distribution $T_0^{(2,1)}(\cdot)$. Second, $\Xi^{(2,1)}(\cdot)$ and $\Xi^{(2,2)}(\cdot)$ are not independent although $T_n^{(2,1)}$ and $T_n^{(2,2)}$ are constructed using nonoverlapping subsamples. This is because although $\mathbb{E}[\zeta_{it}^{(2,1)}\left(\gamma_1\right)\zeta_{it}^{(2,2)}\left(\gamma_2\right)] = 0$,

$\mathbb{E}[\zeta_{it}^{(2,1)}(\gamma_1)\zeta_{i\tau}^{(2,2)}(\gamma_2)]$ need not be zero when $\tau \neq t$ as $\check{\mathbf{x}}_{it}$ and $\check{\mathbf{x}}_{i\tau}$ has a common component $\mathbf{z}_i$, q_{it} and $q_{i\tau}$ may be correlated, and a_{1i} and (a_{2i}, a_{3i}) in e_{it}^0 and $e_{i\tau}^0$ may be correlated.

To simulate the null distribution $F_0^{(2)}$, we generate $\{\xi_i^*\}_{i=1}^N$ as i.i.d. $N(0,1)$, and construct

$$g_n^{(2)*} := \max\left\{g_n^{*(2,1)}, g_n^{*(2,2)}\right\} = \max\left\{g\left(T_n^{(2,1)*}\right), g\left(T_n^{(2,2)*}\right)\right\},$$

where

$$T_n^{(2,1)*}(\gamma) = \widehat{H}_n^{(2,1)}(\gamma)^{-1/2}\widehat{m}_n^{(2,1)*}(\gamma) \text{ with } \widehat{m}_n^{(2,1)*}(\gamma) = N^{-1/2}\sum_{i=1}^N \widehat{\zeta}_i^{(2,1)}(\gamma)\xi_i^*$$

$$T_n^{(2,2)*}(\gamma) = \widehat{H}_n^{(2,2)}(\gamma)^{-1/2}\widehat{m}_n^{(2,2)*}(\gamma) \text{ with } \widehat{m}_n^{(2,2)*}(\gamma) = N^{-1/2}\sum_{i=1}^N \widehat{\zeta}_i^{(2,2)}(\gamma)\xi_i^*.$$

Note that for $T_n^{(2,1)*}$ and $T_n^{(2,2)*}$ we use the same ξ_i^* to mimic the correlation between $\Xi^{(2,1)}(\cdot)$ and $\Xi^{(2,2)}(\cdot)$. We reject $H_0^{(2)}$ if $p_n^{(2)*} = 1 - F_n^{(2)*}\left(g_n^{(2)}\right) \leq \alpha$, and $F_n^{(2)*}$ is the conditional distribution of $g_n^{(2)*}$ given the original data. The following corollary validates this simulation procedure.

Corollary 3. *Under $H_{1c}^{(2)}$ and Assumption 2, $p_n^{(2)*} - p_n^{(2)} = o_p(1)$, which implies $p_n^{(2)*} \xrightarrow{d} 1 - F_0^{(2)}\left(g_c^{(2)}\right)$ under $H_{1c}^{(2)}$ and $p_n^{(2)*} \xrightarrow{d} U[0,1]$ under $H_0^{(2)}$. If Assumption 1 with $m = 2$ instead of $H_{1c}^{(2)}$ holds, then $p_n^{(2)*} \xrightarrow{p} 0$.*

If we reject $H_0^{(2)}$, we can estimate another threshold based on least squares conditional on the first threshold $\widehat{\gamma}^{(1)} = \widehat{\gamma}$. Specifically, our objective function is

$$S_n\left(\gamma|\widehat{\gamma}^{(1)}\right) = S_{1n}\left(\gamma|\widehat{\gamma}^{(1)}\right)1(\gamma \in \Gamma_1) + S_{2n}\left(\gamma|\widehat{\gamma}^{(1)}\right)1(\gamma \in \Gamma_2),$$

where $S_{1n}\left(\gamma|\widehat{\gamma}^{(1)}\right)$ and $S_{2n}\left(\gamma|\widehat{\gamma}^{(1)}\right)$ are the concentrated least squares objective functions in the first and second regimes, respectively. Let $\arg\min_{\gamma\in\Gamma_1\cup\Gamma_2} S_n\left(\gamma|\widehat{\gamma}^{(1)}\right) = \widehat{\gamma}^{(2)}$. We show in the next theorem that $\widehat{\gamma}^{(2)}$ converges to another threshold that dominates the remaining thresholds under the following identification assumption.

Assumption 4. $^{(2)}$: $\arg\min_{\gamma\in\left\{\gamma_1^0,\cdots,\gamma_m^0\right\}\setminus\left\{\gamma_0^{(1)}\right\}} S\left(\gamma|\gamma_0^{(1)}\right) = \gamma_0^{(2)}$ is unique, where $S\left(\gamma|\gamma_0^{(1)}\right)$ is the probability limit of $S_n\left(\gamma|\widehat{\gamma}^{(1)}\right)$.

Lemma 2. *Under Assumptions 1, 2(i)-(iii) and $3^{(2)}$, $\widehat{\gamma}^{(2)} \xrightarrow{p} \gamma_0^{(2)}$.*

We can continue the first-test-then-estimate procedure until in the test of $H_0^{(l+1)}: m = l$ vs. $H_1^{(l+1)}: m = l+1$ the null cannot be rejected. Our estimate of m, $\widehat{m}$, is the first such l. Denote the first $l-1$ threshold estimates as $\left\{\widehat{\gamma}^{(1)}, \cdots, \widehat{\gamma}^{(l-1)}\right\}$, and their ordered counterpart as $\left\{\widehat{\gamma}_{(1)}, \cdots, \widehat{\gamma}_{(l-1)}\right\}$. Similarly, denote the ordered version of $\left\{\gamma_0^{(1)}, \cdots, \gamma_0^{(l-1)}\right\}$ as $\left\{\gamma_{(1)}^0, \cdots, \gamma_{(l-1)}^0\right\}$, which are probability limits of $\left\{\widehat{\gamma}_{(1)}, \cdots, \widehat{\gamma}_{(l-1)}\right\}$. To estimate $\gamma_0^{(l)}$, the objective function is

$$S_n\left(\gamma|\widehat{\gamma}_{(1)}, \cdots, \widehat{\gamma}_{(l-1)}\right) = \sum_{\ell=1}^{l} S_{\ell n}\left(\gamma|\widehat{\gamma}_{(1)}, \cdots, \widehat{\gamma}_{(l-1)}\right)1(\gamma \in \Gamma_\ell),$$

where Γ_ℓ is the truncated version of $\left[\widehat{\gamma}_{(\ell-1)}, \widehat{\gamma}_{(\ell)}\right]$ with $\widehat{\gamma}_{(0)} = -\infty$, and $\widehat{\gamma}_{(l)} = \infty$, and $S_{\ell n}\left(\gamma|\widehat{\gamma}_{(1)}, \cdots, \widehat{\gamma}_{(l-1)}\right)$ is the concentrated least squares objective function applied to the ℓth regime determined by $\left\{\widehat{\gamma}_{(1)}, \cdots, \widehat{\gamma}_{(l-1)}\right\}$. Under the following identification assumption, $\widehat{\gamma}^{(l)} = \arg\min_{\gamma\in\cup_{\ell=1}^{l}\Gamma_\ell} S_n\left(\gamma|\widehat{\gamma}_{(1)}, \cdots, \widehat{\gamma}_{(l-1)}\right)$ converges to $\gamma_0^{(l)}$.

Assumption 5. $^{(l)}$: $\arg\min_{\gamma\in\left\{\gamma_1^0,\cdots,\gamma_m^0\right\}\setminus\left\{\gamma_{(1)}^0,\cdots,\gamma_{(l-1)}^0\right\}} S\left(\gamma|\gamma_{(1)}^0, \cdots, \gamma_{(l-1)}^0\right) = \gamma_0^{(l)}$ is unique, where $S\left(\gamma|\gamma_{(1)}^0, \cdots, \gamma_{(l-1)}^0\right)$ is the probability limit of $S_n\left(\gamma|\widehat{\gamma}_{(1)}, \cdots, \widehat{\gamma}_{(l-1)}\right)$.

Given $\left\{\widehat{\gamma}_{(1)}, \cdots, \widehat{\gamma}_{(l)}\right\}$, in the test of $H_0^{(l+1)}$ vs. $H_1^{(l+1)}$, we construct our test statistic as

$$g_n^{(l+1)} = \max\left\{g_n^{(l+1,1)}, \cdots, g_n^{(l+1,l+1)}\right\},$$

where the $g_n^{(l+1,\ell)}$ are the g_n statistics applied to the ℓ th regimes determined by $\left\{\widehat{\gamma}_{(1)}, \cdots, \widehat{\gamma}_{(l)}\right\}$. We then use the simulation technique to obtain the critical value or the p-value. The results in Corollaries 2 and 3 can be naturally extended to this general case with more complicated notations.

We now state a consistency result for $\widehat{m}$, where we use the critical value $c_N^{(l)}$ instead of the simulated critical value in the test of $H_0^{(l)}$ vs. $H_1^{(l)}$.

Theorem 3. *Under Assumptions 1 and 2, if $c_N^{(l)} \to \infty$ and $\frac{c_N^{(l)}}{N^{1-2\kappa}} \to 0$, $l = 1, \cdots, M$, then $P\left(\widehat{m} = m_0\right) \to 1$ as $N \to \infty$, where m_0 is the true m, and Assumption 2(v) in the test of $H_0^{(l)}$ vs. $H_1^{(l)}$ should be adjusted similarly as in the test of $H_0^{(2)}$ vs. $H_1^{(2)}$.*

In this theorem, $c_N^{(l)} \to \infty$ guarantees the probability that the type-I error converges to zero in the test of $H_0^{(l)}$ vs. $H_1^{(l)}$, and $\frac{c_N^{(l)}}{N^{1-2\kappa}} \to 0$ guarantees the probability that the type-II error converges to zero under Assumption 1. One option of $c_N^{(l)}$ is $c_N^{(l)} = O(\log N)$ for any $0 < \kappa < 1/2$ in Assumption 1.

3.3. Discussion

The $F_T(l+1|l)$ test employed in Bai and Perron (1998) (BP hereafter) and the sequential information criterion (SIC) employed in GP essentially apply the likelihood ratio test of no threshold effect versus one threshold over the $(l+1)$ regimes under the null and then take a supremum. The difference is that the former uses a critical value from a pivotal distribution (given the truncation constant ϵ and the dimension of regessors d) while the latter presets the critical value, such as $O(\log T)$ in BIC. In both the $F_T(l+1|l)$ test and SIC, we need to run $O(n)$ least squares regressions in testing $H_0^{(l+1)}$ vs. $H_1^{(l+1)}$ for any $l = 0, 1, \cdots, M-1$, while our score test need only run $O(l+1)$ least squares regressions overall. The computational advantage is particularly pronounced in the panel data setting compared to the cross-sectional case considered by Yu and Fan (2021). Whereas the procedures of BP and GP need in total to run $O(Mn)$ regressions, our method need only run $O\left(M^2\right)$ regressions, and $n = NT$ is typically substantially larger than M. Also, unlike in the structural change models, the asymptotic null distributions in TR are not pivotal, so simulation methods are used to obtain the critical values in all tests, but such methods can be easily carried out and are not time consuming. In sum, our sequential procedure can be treated as a computationally less intensive alternative of those in BP and GP, and so is better suited in the PTR context. The simulations in Bai and Perron (2006) show that their sequential testing method performs better than traditional ICs in selecting m;[7] and we expect the same conclusion to apply in our case because structural change models can be treated as special TR models.

We often make decisions based on comparing the p-value with the significance level α, so it is helpful to state the corresponding conditions on α_N instead of $c_N^{(l)}$, where we use the subscript N to emphasize the dependency of α on N. These conditions depend on the tail probability of the null distributions. Because the asymptotic null distribution, as the supremum of one or more chi-square processes, is not pivotal, we consider the special case of structural change models. In such models, Proposition 7 of BP shows that the null distribution under $H_0^{(l+1)}$ is $G_{d,\epsilon}(x)^{l+1}$, where $G_{d,\epsilon}(x)$ is the distribution function of $\sup_{\epsilon\le\mu\le1-\epsilon}\left\|W_d(\mu)-\mu W_d(1)\right\|^2/(\mu(1-\mu))$ with $W_d(\cdot)$ being a d-vector of independent Wiener processes on $[0,1]$. As shown in De Long (1981), $1-G_{d,\epsilon}(z)$ decays to zero no faster than an exponential rate as $z\to\infty$. If such a rate also holds for our case, then $\alpha_N\to0$ and $-\log\alpha_N/N^{1-2\kappa}\to0$ are the required conditions, e.g., $\alpha_N = O\left(N^{-1}\right)$ satisfies these conditions for any $0<\kappa<1/2$.

The rate conditions on $c_N^{(l)}$ and α_N try to ensure the probabilities of both the type-I and type-II errors shrink to zero. However, because the null distributions in PTR are not pivotal, the constants in $c_N^{(l)}$ and α_N are not easy to determine; at least there is no literature that discusses how to choose the constant. For example, although we know $c_N^{(l)} = O(\log N)$, $c_N^{(l)} = \log N$ or $c_N^{(l)} = 2\log N$ might make a big difference – see the simulations in GP; a poorly chosen constant could be such that neither the type-I error nor type-II error is close to zero for a given N and T in practice. As a result, we take a more practical perspective – make the type-I error (which is easy to control, e.g., by the p-value) shrink to zero and accept the possibility that the type-II error (which is hard to control) does not shrink to zero. In other words, we allow $\hat{m}\le m_0$ with probability approaching one, and then make our inferences robust to such a misspecification.

The simulations in Bai and Perron (2006) indeed show that when the threshold effect is small, their sequential procedure tends to under-estimate m. This is particularly relevant to our setup since we consider shrinking threshold effects. BP suggest to use a double maximum test to ascertain if any break exists at all; if the null is rejected, then conduct the sequential tests starting from $l=1$ (i.e., neglect $F_T(1|0)$ because it may have low power when $m>1$). This can alleviate the underestimation problem for some DGPs but is not designed for the small-threshold-effect scenario in this paper. Consequently, we take an alternative route to handle this problem, i.e., admit that m can be underestimated, and conduct inferences under such a potential misspecification of m . This solution can also avoid the computational burden of the double maximum test in PTR.

4. Asymptotics in the CRE model

Given $\left(\hat{\gamma}_{(1)},\cdots,\hat{\gamma}_{(\hat{m})}\right)$, we can refine the estimation of $\gamma_{(\ell)}^0$ by using the data with $q_{it}\in\left[\hat{\gamma}_{(\ell-1)},\hat{\gamma}_{(\ell+1)}\right]$, where $\ell=1,\cdots,\hat{m}$, $\hat{\gamma}_{(0)}=-\infty$, and $\hat{\gamma}_{(\hat{m}+1)}=\infty$. Specifically, our updated least squares estimator (LSE) of $\gamma_{(\ell)}^0$ is

$$\hat{\gamma}_\ell = \arg\min_{\gamma\in\Gamma_\ell} S_{\ell n}(\gamma), \tag{16}$$

where $S_{\ell n}(\gamma)$ is the concentrated $S_{\ell n}(\theta)$ with

$$S_{\ell n}(\theta) = \frac{1}{2}\sum_{i=1}^N\sum_{t=1}^T\left[y_{it}-\check{\mathbf{x}}_{it}'\theta_1 1\left(q_{it}\le\gamma\right)-\check{\mathbf{x}}_{it}'\theta_2 1\left(q_{it}>\gamma\right)\right]^2 1\left(\hat{\gamma}_{(\ell-1)}\le q_{it}\le\hat{\gamma}_{(\ell+1)}\right), \tag{17}$$

and Γ_ℓ is the truncated version of $\left[\hat{\gamma}_{(\ell-1)},\hat{\gamma}_{(\ell+1)}\right]$.[8] If $\hat{m}$ is consistent for m_0, then we are effectively estimating $\left(\gamma_1^0,\cdots,\gamma_{m_0}^0\right)$. If $\hat{m}$ underestimates m_0, then some interval $\left[\hat{\gamma}_{(\ell-1)},\hat{\gamma}_{(\ell+1)}\right]$ may contain more than one threshold. Obviously, the estimation procedure for all the $\gamma_{(\ell)}^0$ is similar, and mirrors the estimation of $\gamma_0^{(1)}$ in (13) where the least squares objective function might be misspecified. As a result, we study the asymptotic properties of $\hat{\theta}$ from (13) to simplify notation, and these properties can be applied to the above estimator of $\gamma_{(\ell)}^0$ for any ℓ with notation adapted to the data segment $\left[\hat{\gamma}_{(\ell-1)},\hat{\gamma}_{(\ell+1)}\right]$.

[7] As argued in Bai and Perron (2006), traditional ICs do not take into account the heterogeneity of error variances across regimes.

[8] We can repeat the procedure to update $\left\{\hat{\gamma}_\ell\right\}_{\ell=1}^{\hat{m}}$ recursively but do not systematically investigate convergence properties of this recursive procedure in our simulations or empirical application.

4.1. Asymptotics for the LSE

To study the asymptotic properties of $\hat{\gamma}$ and $\hat{\beta}$, we need some assumptions. To facilitate the exposition, we introduce additional notation and for convenience collect some previously defined notation. Denote $\gamma_0^{(1)}$ as γ_0, and assume there are m_1 regimes below γ_0 and m_2 regimes above γ_0, where $1 \le m_1, m_2 \le m$, and $m_1 + m_2 = m + 1$. Recall that $f_t(\gamma)$ is the density of q_{it} at γ. Denote $f_{\tau|t}(\gamma_1|\gamma_2)$ as the conditional density of $q_{i\tau}$ given q_{it}. Define

$$\begin{aligned}
\vartheta_1 &= \theta_{m+1} + M(\gamma_0)^{-1} \sum_{\ell=1}^{m_1} M(\gamma_{\ell-1}^0, \gamma_\ell^0)\delta^\ell = M(\gamma_0)^{-1} \sum_{\ell=1}^{m_1} M(\gamma_{\ell-1}^0, \gamma_\ell^0)\theta_\ell, \\
\vartheta_2 &= \theta_{m+1} + \overline{M}(\gamma_0)^{-1} \sum_{\ell=m_1+1}^{m} M(\gamma_{\ell-1}^0, \gamma_\ell^0)\delta^\ell = \overline{M}(\gamma_0)^{-1} \sum_{\ell=m_1+1}^{m+1} M(\gamma_{\ell-1}^0, \gamma_\ell^0)\theta_\ell, \\
c &= M(\gamma_0)^{-1} \sum_{\ell=1}^{m_1} M(\gamma_{\ell-1}^0, \gamma_\ell^0)c^\ell - \overline{M}(\gamma_0)^{-1} \sum_{\ell=m_1+1}^{m} M(\gamma_{\ell-1}^0, \gamma_\ell^0)c^\ell, \\
c_- &= c^{m_1} - M(\gamma_0)^{-1} \sum_{\ell=1}^{m_1} M(\gamma_{\ell-1}^0, \gamma_\ell^0)c^\ell, c_+ = c^{m_1+1} - \overline{M}(\gamma_0)^{-1} \sum_{\ell=m_1+1}^{m} M(\gamma_{\ell-1}^0, \gamma_\ell^0)c^\ell,
\end{aligned}$$

and

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

where

$$\begin{aligned}
M &= \sum_{t=1}^{T} \mathbb{E}\left[\check{\mathbf{x}}_{it}\check{\mathbf{x}}_{it}'\right], M(\gamma) := \sum_{t=1}^{T} \mathbb{E}\left[\check{\mathbf{x}}_{it}\check{\mathbf{x}}_{it}' 1(q_{it} \le \gamma)\right], \\
\overline{M}(\gamma) &= M - M(\gamma), M(\gamma_1, \gamma_2) = M(\gamma_2) - M(\gamma_1) \text{ if } \gamma_1 < \gamma_2 \text{ and } \mathbf{0} \text{ otherwise}, \\
e_{1it} &= \sum_{\ell=1}^{m_1} \varepsilon_{\ell it} 1(\gamma_{\ell-1}^0 < q_{it} \le \gamma_\ell^0) \text{ and } e_{2it} = \sum_{\ell=m_1+1}^{m+1} \varepsilon_{\ell it} 1(\gamma_{\ell-1}^0 < q_{it} \le \gamma_\ell^0).
\end{aligned}$$

The pseudo-true value ϑ_1 is a weighted average of $\{\theta_\ell\}_{\ell=1}^{m_1}$ and similarly ϑ_2 is a weighted average of $\{\theta_\ell\}_{\ell=m_1+1}^{m+1}$; both ϑ_1 and ϑ_2 depend on N and T. Under Assumption 1, $\delta_N := \vartheta_1 - \vartheta_2 = cN^{-\kappa}$, $\theta_{m_1} - \vartheta_1 = c_- N^{-\kappa}$, and $\theta_{m_1+1} - \vartheta_2 = c_+ N^{-\kappa}$, and using the same notation c as in Section 3 should not introduce any confusion; note that although δ_N is standard in the literature, the other two terms here are new but are relevant in the asymptotic distribution of $\hat{\gamma}$. Also, compared with D and V_ℓ, Ω_1, Ω_2 and Ω_{12} contain cross terms (i.e., terms with $\tau \ne t$); and, different from the D in Hansen (1999), the conditional means in our specification of D involve information from other periods, i.e., $\check{\mathbf{x}}_{it}$ contains information of $\mathbf{x}_{i\tau}$ with $\tau \ne t$.

Because we assume $\hat{\gamma}$ is consistent to γ_0, the following assumptions that are relevant to the derivation of the asymptotic distribution of $\hat{\gamma}$ are all stated in the neighborhood of γ_0.

Assumption 6.

(i) $\{\mathbf{x}_{it}, \underline{\mathbf{z}}_i, y_{it}\}_{t=1}^{T}$ are i.i.d. across i, $i = 1, \cdots, N$; T is fixed and $N \to \infty$.
(ii) For each i, $\mathbb{E}\left[a_{\ell i}|\mathbf{X}_i\right] = 0$ and $\mathbb{E}\left[u_{it}|\mathbf{X}_i\right] = 0$.
(iii) For each $j = 1, \cdots, d_{\mathbf{x}}$, $P\left(x_{i1}^j = \cdots = x_{iT}^j\right) < 1$, where x_{it}^j is the jth element of $\mathbf{x}_{it}$.
(iv) For $t = 1, \cdots, T$, $i = 1, \cdots, N$, $\mathbb{E}[\|\check{\mathbf{x}}_{it}\|^4] < \infty$, $\mathbb{E}[|e_{\ell it}|^4] < \infty$, $\sup_{\gamma\in\mathcal{N}} \mathbb{E}[\|\check{\mathbf{x}}_{it}\|^4 | q_{it} = \gamma] < \infty$, $\sup_{\gamma\in\mathcal{N}} \mathbb{E}[|e_{\ell it}|^4 | q_{it} = \gamma] < \infty$, and $\sup_{\gamma\in\mathcal{N}} \mathbb{E}(\|\check{\mathbf{x}}_{it}\| |e_{\ell it}|)^{2+\epsilon} | q_{it} = \gamma] < \infty$ for some $\epsilon > 0$ and some neighborhood $\mathcal{N}$ of γ_0.

(v) Assumption 1 holds.
(vi) $f_t(\gamma)$, $t = 1, \cdots, T$, $D(\gamma)$ and $V_\ell(\gamma)$ are continuous at γ_0.
(vii) $c'D(c + 2c_-) > 0$, $c'D(c - 2c_+) > 0$ and $c'V_\ell c > 0$.
(viii) For $t = 1, \cdots, T$, $0 < f_t(\gamma_0) \le \overline{f} < \infty$, and for $\tau > t$, $f_{\tau|t}(\gamma_0|\gamma_0) < \infty$.
(ix) $M_1 = M(\gamma_0) > 0$, $M_2 = \overline{M}(\gamma_0) > 0$, and $\Omega_\ell > 0$.

Condition (i) reemphasizes that only the short panel is considered. In Condition (ii), we do not require $a_{1i}, a_{2i}, u_{i1}, \cdots, u_{iT}$ and $\mathbf{X}_i$ are independent of each other but only a conditional mean independence assumption, which implies $\Omega_{12} \neq \mathbf{0}$ due to $\mathbb{E}\left[e_{1it}e_{2i\tau}|\mathbf{X}_i\right] \neq 0$ as a result of the correlation among $a_{\ell i}$'s and among u_{it}'s.[9] Condition (iii) requires x_{it} to vary over t which is exactly how we partition $\mathbf{x}_{it}$ and $\underline{\mathbf{z}}_i$. This assumption avoids the multicollinearity problem between $\mathbf{x}_{it}$ and $\overline{\mathbf{x}}_i$; a similar assumption is imposed in the demeaning method of Hansen (1999) to avoid zero regressors. Condition (iv) includes some standard moment conditions, which implies $\Omega_\ell < \infty$, $M < \infty$, and (combining with Conditions (vi) and (viii)) $D(\gamma) < \infty$ and $V_\ell(\gamma) < \infty$ for γ in a neighborhood of γ_0. We can express this assumption in terms of $\mathbf{x}_{it}, u_{it}, \underline{\mathbf{z}}_i$ and $a_{\ell i}$, but the current formulation can simplify the statement. Note also that the Liapounov type of condition $\sup_{\gamma \in \mathcal{N}} \mathbb{E}[(\|\check{\mathbf{x}}_{it}\||e_{\ell it}|)^{2+\epsilon}|q_{it} = \gamma] < \infty$ is not implied by $\mathbb{E}[\|\check{\mathbf{x}}_{it}\|^4] < \infty$ and $\mathbb{E}[|e_{\ell it}|^4] < \infty$ or $\sup_{\gamma \in \mathcal{N}} \mathbb{E}[\|\check{\mathbf{x}}_{it}\|^4|q_{it} = \gamma] < \infty$ and $\sup_{\gamma \in \mathcal{N}} \mathbb{E}[|e_{\ell it}|^4|q_{it} = \gamma] < \infty$. Actually, we only require γ to be in a half neighborhood of γ_0 in stating the conditions involving $e_{\ell it}$ and these conditions can be expressed in terms of $\varepsilon_{m_1 it}$ and $\varepsilon_{m_1+1,it}$. Specifically, for γ in the left neighborhood of γ_0, $\mathbb{E}[|e_{1it}|^4|q_{it} = \gamma] = \mathbb{E}[\left|\varepsilon_{m_1 it}\right|^4|q_{it} = \gamma]$, and for γ in the right neighborhood of γ_0, $\mathbb{E}[|e_{1it}|^4|q_{it} = \gamma] = \mathbb{E}[\left|\varepsilon_{m_1+1,it}\right|^4|q_{it} = \gamma]$, and similarly for $\mathbb{E}[(\|\check{\mathbf{x}}_{it}\||e_{\ell it}|)^{2+\epsilon}|q_{it} = \gamma]$. This notational convention also applies to V_ℓ so that

$$V_1 = \sum_{t=1}^{T} \mathbb{E}\left[\check{\mathbf{x}}_{it}\check{\mathbf{x}}_{it}'\varepsilon_{m_1 it}^2|q_{it} = \gamma_0-\right] f_t(\gamma_0), \text{ and } V_2 = \sum_{t=1}^{T} \mathbb{E}\left[\check{\mathbf{x}}_{it}\check{\mathbf{x}}_{it}'\varepsilon_{m_1+1,it}^2|q_{it} = \gamma_0+\right] f_t(\gamma_0). \tag{18}$$

The continuity of $D(\gamma)$ at γ_0 in Condition (vi) excludes the possibility that $\check{\mathbf{x}}_{it}$ has a discontinuous conditional distribution at $q_{it} = \gamma_0$ (under Condition (vi), i.e., $f_t(\gamma)$ is continuous at γ_0).[10] Distinct from Hansen (1999), V_1 and V_2 generally differ due to $a_{m_1 i} \neq a_{m_1+1,i}$ and $u_{m_1 it} \neq u_{m_1+1,it}$. Note that although the differences between ψ_ℓ's converge to zero as $N \to \infty$, $\alpha_{\ell i}$ need not converge to a common α_i; in other words, our framework will not degenerate to Hansen (1999)'s in general. The first part of Condition (vii) is different from the counterpart in correctly specified models where we usually assume $c'Dc > 0$; the extra terms $2c'Dc_-$ and $-2c'Dc_+$ are due to model misspecification and are equal to zero when the model is correctly specified (i.e., $m_0 = 1$ and there are only two regimes, and hence $c_- = c_+ = \mathbf{0}$)[11]. The usual assumption $c'Dc > 0$ in cross-sectional models is an identification assumption for γ by excluding continuous threshold regression (see Chan and Tsay (1998) and Hansen (2017)), and the assumption here plays a similar role and is actually implied by Assumption 3$^{(1)}$. The second part of Condition (viii) excludes the possibility that q_{it} lingers at γ_0 with a positive probability over time; it eliminates the cross terms in D and V_ℓ.[12] Condition (ix) is standard in stating the asymptotic distribution of $\widehat{\beta}$.

The following result provides the limit theory of $\widehat{\gamma}$ and $\widehat{\beta}$.

Theorem 4. *Under Assumption 4,*

$$N^{1-2\kappa}\left(\widehat{\gamma} - \gamma_0\right) \xrightarrow{d} \omega \cdot \zeta(\varphi, \phi),$$
$$N^{1/2}\left(\widehat{\underline{\theta}} - \underline{\theta}\right) \xrightarrow{d} N(0, \Sigma),$$

and $\widehat{\gamma}$ and $\widehat{\underline{\theta}}$ are asymptotically independent, where

$$\omega = \frac{c'V_1 c}{\left[c'D\left(c + 2c_-\right)\right]^2} \text{ and } \zeta(\varphi, \phi) = \arg\max_r \begin{cases} -\frac{|r|}{2} + B_1(-r), & \text{if } r \le 0, \\ -\varphi\frac{r}{2} + \sqrt{\phi}B_2(r), & \text{if } r > 0, \end{cases}$$

with $\varphi = \frac{c'D(c-2c_+)}{c'D(c+2c_-)}$, $\phi = \frac{c'V_2 c}{c'V_1 c}$ and $B_\ell(r)$, $\ell = 1, 2$, being two independent standard Wiener processes on $[0, \infty)$, and

$$\Sigma = \begin{pmatrix} \Sigma_1 & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_2 \end{pmatrix}$$

with $\Sigma_\ell = M_\ell^{-1}\Omega_\ell M_\ell^{-1}$, and $\Sigma_{12} = M_1^{-1}\Omega_{12}M_2^{-1}$.

The asymptotic distribution of $\widehat{\gamma}$ is quite different from that in correctly specified models. First, $\omega \neq \frac{c'V_1 c}{(c'Dc)^2}$; second, there is an extra term φ in $\zeta(\varphi, \phi)$, i.e., $\varphi \neq 1$. This extra term usually appears because $\check{\mathbf{x}}_{it}$ has a discontinuous conditional distribution at $q_{it} =$

[9] This is also because q_{it} is not fixed over $t = 1, \cdots, T$ so that $1\left(q_{it} \le \gamma_0\right)1\left(q_{i\tau} > \gamma_0\right) \neq 0$.
[10] Continuity of the conditional distribution of $\mathbf{x}_{it}$ at $q_{it} = \gamma_0$ is also an indicator for why all the $\mathbf{x}_{it}$ are used as control variables for $\alpha_{\ell i}$, not only those $\mathbf{x}_{it}$'s in each regime.
[11] If only one regime is correctly specified, then the corresponding c_- or c_+ is zero; e.g., if $m_1 = m$, then $c_+ = \mathbf{0}$.
[12] The deeper reason that D and V_ℓ do not contain cross terms while Ω_ℓ and Ω_{12} do is because γ estimation explores the local information around γ_0 whereas $\underline{\theta}$ estimation explores global information as detailed in Yu (2012, 2015).

γ_0 and takes the form $c'D_+c/c'D_-c$ (see, e.g., Proposition 3 of Bai, 1997), where $D_\pm = \sum_{t=1}^T \mathbb{E}\left[\check{\mathbf{x}}_{it}\check{\mathbf{x}}_{it}'|q_{it} = \gamma_0\pm\right] f_t(\gamma_0)$. Interestingly, although we exclude the discontinuity in the distribution of $\check{\mathbf{x}}_{it}$, φ still appears in the misspecified model and takes a different form. The distribution of $\hat{\theta}_\ell$ also deserves comment. First, different from the cross-sectional case, $\hat{\theta}_1$ and $\hat{\theta}_2$ are no longer asymptotically independent. In Hansen (1999), $\hat{\beta}_1$ and $\hat{\beta}_2$ are also correlated, but his asymptotic variance does not involve the cross term as it does here. A key difference between the fixed effects estimator and the CRE estimator of β is that β_1 and β_2 cannot be estimated using separate data in the former, which greatly affects the formulae of the asymptotic variance estimator of $\hat{\beta}$ on pages 352–353 of Hansen (1999). Second, because the linear model in each regime is misspecified, we expect the asymptotic variance of $\hat{\theta}_1$ to include an extra term $Var\left(\sum_{t=1}^T \check{\mathbf{x}}_{it}\left(g_1(\check{\mathbf{x}}_{it}) - \check{\mathbf{x}}_{it}'\vartheta_1\right)1(q_{it} \le \gamma_0)\right)$, where $g_1(\check{\mathbf{x}}_{it}) = \sum_{\ell=1}^{m_1} \check{\mathbf{x}}_{it}'\delta^\ell 1(\gamma_{\ell-1}^0 < q_{it} \le \gamma_\ell^0) + \check{\mathbf{x}}_{it}'\theta_{m+1}$ is $\mathbb{E}[y_{it}|\mathbf{X}_i]$ in the left regime. However,

$$g_1(\check{\mathbf{x}}_{it}) - \check{\mathbf{x}}_{it}'\vartheta_1 = \sum_{\ell=1}^{m_1} \check{\mathbf{x}}_{it}'\left(\delta^\ell - M(\gamma_0)^{-1}\sum_{l=1}^{m_1} M(\gamma_{l-1}^0, \gamma_l^0)\delta^l\right)1(\gamma_{\ell-1}^0 < q_{it} \le \gamma_\ell^0) = O_p(N^{-\kappa}), \tag{19}$$

so this extra term can be neglected asymptotically. Similar analysis applies to $\hat{\theta}_2$.

4.2. Estimation of nuisance parameters in asymptotic distributions

For inferences on γ and β, we need to estimate the nuisance parameters in the asymptotic distributions of $\hat{\gamma}$ and $\hat{\theta}_\ell$. First, we discuss the estimation of ω, φ and ϕ. Because they all take the form of ratios, we need only estimate $N^{-2\kappa}c'D(c-2c_+)$, $N^{-2\kappa}c'D(c+2c_-)$, and $N^{-2\kappa}c'V_\ell c$. The estimation of the first two objects is not obvious since c_- and c_+ involve c^{m_1} and c^{m_1+1}, respectively, which are not estimated in least squares regression. In fact, they need not be estimated at all. From the proof of Theorem 4,

$$\begin{aligned} N^{-2\kappa}c'D(c+2c_-) &= 2\sum_{t=1}^T \mathbb{E}\left[\left(\delta_N'\check{\mathbf{x}}_{it}\right)\left(y_{it} - \check{\mathbf{x}}_{it}'\overline{\vartheta}\right)|q_{it} = \gamma_0-\right] f_t(\gamma_0), \\ N^{-2\kappa}c'D(c-2c_+) &= -2\sum_{t=1}^T \mathbb{E}\left[\left(\delta_N'\check{\mathbf{x}}_{it}\right)\left(y_{it} - \check{\mathbf{x}}_{it}'\overline{\vartheta}\right)|q_{it} = \gamma_0+\right] f_t(\gamma_0), \end{aligned}$$

where $\overline{\vartheta} = \frac{\vartheta_1+\vartheta_2}{2}$, so we can substitute δ_N, γ_0 and $\overline{\vartheta}$ by $\hat{\delta} := \hat{\theta}_1 - \hat{\theta}_2$, $\hat{\gamma}$ and $\hat{\overline{\vartheta}} = \frac{\hat{\theta}_1+\hat{\theta}_2}{2}$, and use the data of all individuals in period t to estimate $\mathbb{E}\left[\left(\delta_N'\check{\mathbf{x}}_{it}\right)\left(y_{it} - \check{\mathbf{x}}_{it}'\overline{\vartheta}\right)|q_{it} = \gamma_0\pm\right] f_t(\gamma_0)$. These estimators are standard in the literature – see, e.g., Hansen (2000, Section 3.4) or Yu et al. (2024, Section 3.4) for details. In correctly specified models, $c_\pm = \mathbf{0}$, and $N^{-2\kappa}c'Dc = \sum_{t=1}^T \mathbb{E}\left[\left(\delta_N'\check{\mathbf{x}}_{it}\right)^2|q_{it} = \gamma_0\right] f_t(\gamma_0)$. But in misspecified models estimation based on $\sum_{t=1}^T \mathbb{E}\left[\left(\delta_N'\check{\mathbf{x}}_{it}\right)^2|q_{it} = \gamma_0\right] f_t(\gamma_0)$ is not valid. On the other hand, estimation of $N^{-2\kappa}c'V_1c$ and $N^{-2\kappa}c'V_2c$ based on the formulae in the correctly specified models are indeed valid, i.e., we can estimate $N^{-2\kappa}c'V_1c$ and $N^{-2\kappa}c'V_2c$ based on $\sum_{t=1}^T \mathbb{E}\left[\left(\delta_N'\check{\mathbf{x}}_{it}\right)^2\left(y_{it} - \check{\mathbf{x}}_{it}'\vartheta_1\right)^2|q_{it} = \gamma_0-\right] f_t(\gamma_0)$ and $\sum_{t=1}^T \mathbb{E}\left[\left(\delta_N'\check{\mathbf{x}}_{it}\right)^2\left(y_{it} - \check{\mathbf{x}}_{it}'\vartheta_2\right)^2|q_{it} = \gamma_0+\right] f_t(\gamma_0)$, respectively. Details are given in Appendix F, which also provides simplified formulae in the homoskedastic model. In sum, estimation of ω, φ and ϕ is based on

$$\begin{aligned} N^{2\kappa}\omega &\approx \frac{1}{4}\frac{\sum_{t=1}^T \mathbb{E}\left[\left(\delta_N'\check{\mathbf{x}}_{it}\right)^2\left(y_{it} - \check{\mathbf{x}}_{it}'\vartheta_1\right)^2|q_{it} = \gamma_0-\right] f_t(\gamma_0)}{\left[\sum_{t=1}^T \mathbb{E}\left[\left(\delta_N'\check{\mathbf{x}}_{it}\right)\left(y_{it} - \check{\mathbf{x}}_{it}'\overline{\vartheta}\right)|q_{it} = \gamma_0-\right] f_t(\gamma_0)\right]^2}, \\ \varphi &= -\frac{\sum_{t=1}^T \mathbb{E}\left[\left(\delta_N'\check{\mathbf{x}}_{it}\right)\left(y_{it} - \check{\mathbf{x}}_{it}'\overline{\vartheta}\right)|q_{it} = \gamma_0+\right] f_t(\gamma_0)}{\sum_{t=1}^T \mathbb{E}\left[\left(\delta_N'\check{\mathbf{x}}_{it}\right)\left(y_{it} - \check{\mathbf{x}}_{it}'\overline{\vartheta}\right)|q_{it} = \gamma_0-\right] f_t(\gamma_0)}, \\ \phi &\approx \frac{\sum_{t=1}^T \mathbb{E}\left[\left(\delta_N'\check{\mathbf{x}}_{it}\right)^2\left(y_{it} - \check{\mathbf{x}}_{it}'\vartheta_2\right)^2|q_{it} = \gamma_0+\right] f_t(\gamma_0)}{\sum_{t=1}^T \mathbb{E}\left[\left(\delta_N'\check{\mathbf{x}}_{it}\right)^2\left(y_{it} - \check{\mathbf{x}}_{it}'\vartheta_1\right)^2|q_{it} = \gamma_0-\right] f_t(\gamma_0)}, \end{aligned}$$

which formulae are valid irrespective of whether the model is correctly specified. [13]

Estimation of Σ_ℓ and Σ_{12} is accomplished by means of their sample analogs: $\hat{\Sigma}_\ell = \widehat{M}_\ell^{-1}\hat{\Omega}_\ell\widehat{M}_\ell^{-1}$ and $\hat{\Sigma}_{12} = \widehat{M}_1^{-1}\hat{\Omega}_{12}\widehat{M}_2^{-1}$, where

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

and $\widehat{M}_2$ is the same as $\widehat{M}_1$ but replaces $q_{it} \le \hat{\gamma}$ by $q_{it} > \hat{\gamma}$, and $\hat{\Omega}_2$ is the same as $\hat{\Omega}_1$ but replaces $\hat{e}_{1it}1(q_{it} \le \hat{\gamma})$ by $\hat{e}_{2it}1(q_{it} > \hat{\gamma})$. From the discussion around (19), $\hat{e}_{\ell it}$ contains an extra approximation error $g_\ell(\check{\mathbf{x}}_{it}) - \check{\mathbf{x}}_{it}'\vartheta_\ell$, but this error can be neglected asymptotically. Appendix F shows how estimation of Σ_ℓ and Σ_{12} can be simplified in correctly specified error components models.

[13] Note that the LR inference below does not involve ω, so we do not need to know κ.

Based on these asymptotics a Wald test can be used for testing $\alpha_{\ell i} = \alpha_i$ for $\ell = 1, \cdots, m+1$, i.e., whether there is no unobserved individual-specific threshold effect. Such a test is standard in the literature and further discussion is given in Appendix F.

5. Inference on γ and β

This section proposes inferential methods for γ and β. Following Hansen (1999, 2000), we invert the LR statistic for hypotheses concerning γ such as $H_0 : \gamma = \gamma_0$ to construct a CI for γ. For β, Hansen (2000) suggests a Bonferroni-type CI to incorporate the randomness of $\widehat{\gamma}$ in finite samples (although $\widehat{\gamma}$ will not affect the distribution of $\widehat{\beta}$ asymptotically); the coverage for γ in such a CI is arbitrarily set at 80% and for each γ in its CI the coverage for β is fixed at $1-\alpha$. Instead, we propose a projection CI based on the LR statistic for hypotheses concerning both γ and β, which can be interpreted as an adaptive Bonferroni CI with the coverage for γ and the coverage for β at each γ adaptively chosen. We also propose alternative forms of the LR statistics. Our notation here follows the last section and similarly focuses on a multiple-regime model.

5.1. LR inference on γ

As emphasized in Hansen (2000), the following LR-like statistic has better finite sample performance than a typical t-like statistic when the threshold effect is small: $LR_n(\gamma) = \frac{S_n(\gamma) - S_n(\widehat{\gamma})}{\widehat{\eta}^2}$. Here, $\widehat{\eta}^2$ is a consistent estimator of

$$\eta^2 = \frac{c' V_1 c}{c' D\left(c + 2c_-\right)} \approx \frac{\sum_{t=1}^T \mathbb{E}\left[\left(\delta_N' \check{\mathbf{x}}_{it}\right)^2 \left(y_{it} - \check{\mathbf{x}}_{it}' \vartheta_1\right)^2 | q_{it} = \gamma_0 -\right] f_t(\gamma_0)}{2\sum_{t=1}^T \mathbb{E}\left[\left(\delta_N' \check{\mathbf{x}}_{it}\right)\left(y_{it} - \check{\mathbf{x}}_{it}' \overline{\vartheta}\right) | q_{it} = \gamma_0 -\right] f_t(\gamma_0)},$$

and $S_n(\gamma)$ is the concentrated objective function (14). This test statistic is a by-product of the two-step estimation procedure for γ.

Theorem 5. *Under Assumption 4,*

$$LR_n\left(\gamma_0\right) \xrightarrow{d} \xi(\varphi, \phi), \tag{20}$$

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

Given the estimates $\widehat{\eta}^2$, $\widehat{\varphi}$ and $\widehat{\phi}$, the $(1-\alpha)$ LR-CI for γ follows by inversion of the statistic from

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

where $\widehat{c}_\alpha^0$ is the $(1-\alpha)$ quantile of $\xi(\varphi, \phi)$ with (φ, ϕ) being replaced by $\left(\widehat{\varphi}, \widehat{\phi}\right)$. In Appendix F, we show that the estimation of η^2, φ and ϕ can be simplified in correctly specified error components models.

5.2. Adaptive bonferroni inference on β

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

$$LR_n\left(\gamma, \beta_{11}\right) = \frac{S_n\left(\gamma, \beta_{11}\right) - S_n\left(\widehat{\gamma}, \widehat{\beta}_{11}\right)}{\widehat{\eta}^2}. \tag{21}$$

The nuisance parameter η^2 in (21) can be estimated with the null imposed but we suggest it be estimated without imposing the null to improve computational efficiency. $S_n\left(\gamma, \beta_{11}\right)$ is the concentrated objective function on $\left(\gamma, \beta_{11}\right)$. Specifically, for each $\gamma \in \Gamma$, run least squares of $y_{it} - x_{it}^1 \beta_{11}$ on $\left(\mathbf{x}_{it}^{-1\prime}, \mathbf{z}_i'\right)'$ for those (i,t) such that $q_{it} \le \gamma$ and y_{it} on $\left(\mathbf{x}_{it}', \mathbf{z}_i'\right)'$ for those (i,t) such that $q_{it} > \gamma$ separately to obtain $\widehat{\theta}_{12}\left(\gamma, \beta_{11}\right)$ and $\widehat{\theta}_2(\gamma)$. The concentrated objective function is $S_n\left(\gamma, \beta_{11}\right) = S_n\left(\gamma, \beta_{11}, \widehat{\theta}_{12}\left(\gamma, \beta_{11}\right), \widehat{\theta}_2(\gamma)\right)$, and notice that $S_n\left(\widehat{\gamma}, \widehat{\beta}_{11}\right)$ is the same as $S_n\left(\widehat{\gamma}\right)$ in $LR_n(\gamma)$.[14]

Theorem 6. *Under Assumption 4,*

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

[14] In our notation, $S_n(\theta) = S_n\left(\gamma, \theta_1, \theta_2\right) = S_n\left(\gamma, \beta_{11}, \theta_{12}, \theta_2\right)$ denotes the original objective function; whenever the argument list is shorter than θ, $S_n(\cdot)$ denotes the concentrated objective function.

where $\xi(\varphi,\phi)$ is defined in (20) and

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

with $\widetilde{S}_{\beta_{11}\beta_{11}}=S_{\beta_{11}\beta_{11}}-S_{\beta_{11}\theta_{12}}S^{-1}_{\theta_{12}\theta_{12}}S_{\theta_{12}\beta_{11}}$, and the χ^2_1 variate and $\xi(\varphi,\phi)$ are independent.

Note that $\varrho_{11}=1$ when $\Omega_1=\eta^2M_1$, which holds in correctly specified, homoskedastic (in each regime) cross-sectional models. But in our setup, $\varrho_{11}\neq 1$ in general, and then $2\cdot LR_n\left(\gamma_0,\beta^0_{11}\right)\overset{d}{\nrightarrow}\chi^2_1+2\xi(\varphi,\phi)$ as in the usual asymptotic distribution of LR statistics.

The critical value does not have an explicit form, but we can simulate independent χ^2_1 and $\xi(\varphi,\phi)$ random numbers to obtain the critical value; especially, $\xi(\varphi,\phi)=\max\left(\xi_1,\xi_2(\varphi,\phi)\right)$, where ξ_1 follows the standard exponential distribution, $\xi_2(\varphi,\phi)$ follows an exponential distribution with mean ϕ/φ, and ξ_1 and $\xi_2(\varphi,\phi)$ are independent. Suppose the level α critical value is $\widehat{c}_\alpha$, then the $(1-\alpha)$ CI of β_{11} is

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

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

$$\bigcup\nolimits_{q_{it}\in\Gamma}\left\{\beta_{11}|LR_n(q_{it},\beta_{11})\leq\widehat{c}_\alpha\right\}=:\bigcup\nolimits_{q_{it}\in\Gamma}\mathrm{CI}(\beta_{11}|q_{it}),\tag{22}$$

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

5.3. Alternative inference procedures

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

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

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

where $\widehat{\underline{\theta}}(\widehat{\gamma})\neq\widehat{\underline{\theta}}(\gamma)$, and $\widehat{\underline{\theta}}_{-11}\left(\widehat{\gamma},\widehat{\beta}_{11}\right)\neq\widehat{\underline{\theta}}_{-11}(\gamma,\beta_{11})$ with $\underline{\theta}_{-11}(\gamma,\beta_{11})=\left(\widehat{\theta}_{12}(\gamma,\beta_{11})',\widehat{\theta}_2(\gamma)'\right)'$. We can exclude such indirect effects by using the same $\underline{\theta}$ ($\underline{\theta}_{-11}$) estimator in $LR_n(\gamma)$ ($LR_n(\gamma,\beta_{11})$). We define $LR_{1n}(\gamma)$ and $LR_{1n}(\gamma,\beta_{11})$ as the LR statistics when $\widehat{\underline{\theta}}(\gamma)$ and $\widehat{\underline{\theta}}_{-11}(\gamma,\beta_{11})$ are used as common estimators of $\underline{\theta}$ and $\underline{\theta}_{-11}$, and $LR_{2n}(\gamma)$ and $LR_{2n}(\gamma,\beta_{11})$ when $\widehat{\underline{\theta}}(\widehat{\gamma})$ and $\widehat{\underline{\theta}}_{-11}\left(\widehat{\gamma},\widehat{\beta}_{11}\right)$ are used as common estimators. Because the LR_{1n} statistics suffer some shortcomings, such as time-consuming CI construction for γ and β_{11}, and tendency towards long CIs for γ and β_{11}, we leave analysis on this group of LR statistics to Appendix E, and concentrate here on the following LR_{2n} statistics:

$$LR_{2n}(\gamma)=\frac{S_n\left(\gamma,\widehat{\underline{\theta}}\right)-S_n\left(\widehat{\gamma},\widehat{\underline{\theta}}\right)}{\widehat{\eta}^2},$$

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

noting that $S_n\left(\widehat{\gamma},\widehat{\underline{\theta}}\right)=S_n\left(\widehat{\gamma},\widehat{\beta}_{11},\widehat{\underline{\theta}}_{-11}\right)$ is the same as $S_n(\widehat{\gamma})$ in $LR_n(\gamma)$ and $S_n\left(\widehat{\gamma},\widehat{\beta}_{11}\right)$ in $LR_n(\gamma,\beta_{11})$.

Before detailed analysis of LR_{2n}, we summarize the advantages and disadvantages of the three groups of LR statistics. First, LR_{1n} and LR_{2n} shut down the indirect effects of the null but LR_n does not. Second, the LR_{1n} statistics are computationally more demanding than the LR_n and LR_{2n} statistics. Third, the LR_{2n} statistics are the most powerful among the three groups of LR statistics since $LR_{1n}(\gamma)\leq LR_n(\gamma)\leq LR_{2n}(\gamma)$ and $LR_{1n}(\gamma,\beta_{11})\leq LR_n(\gamma,\beta_{11})\leq LR_{2n}(\gamma,\beta_{11})$. [15] The LR_{2n} statistics appear to be the most favored option from the standpoint of both theory and computation.

[15] Since the critical values for $LR_n(\gamma)$, $LR_{1n}(\gamma)$ and $LR_{2n}(\gamma)$ are the same, $LR_{2n}(\gamma)$ would result in the shortest CI for γ , and $LR_{1n}(\gamma)$ the longest. On the other hand, the critical values for $LR_n(\gamma,\beta_{11})$, $LR_{1n}(\gamma,\beta_{11})$ and $LR_{2n}(\gamma,\beta_{11})$ are different, and thus the CI for β_{11} resulting from $LR_{2n}(\gamma,\beta_{11})$ need not be the shortest.

Theorem 7. *Under Assumption 4,*

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

and

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

where $W_1 \sim N(\mathbf{0}, \Omega_1)$ is independent of $\xi(\varphi, \phi)$, and

$$g(W_1) = \frac{W_1'\left(1, -S_{\beta_{11}\theta_{12}}S_{\theta_{12}\theta_{12}}^{-1}\right)'}{\widetilde{S}_{\beta_{11}\beta_{11}}}\left[W_{11} - \frac{1}{2}\frac{\left(1, -S_{\beta_{11}\theta_{12}}S_{\theta_{12}\theta_{12}}^{-1}\right)W_1}{\widetilde{S}_{\beta_{11}\beta_{11}}/S_{\beta_{11}\beta_{11}}} - S_{\beta_{11}\theta_{12}}\widetilde{S}_{\theta_{12}\theta_{12}}^{-1}\left(-S_{\theta_{12}\beta_{11}}, \mathbf{I}\right)W_1\right],$$

with W_{11} being the first element of W_1, $\widetilde{S}_{\beta_{11}\beta_{11}}$ is defined in Theorem 6, $\widetilde{S}_{\theta_{12}\theta_{12}} = S_{\theta_{12}\theta_{12}} - S_{\theta_{12}\beta_{11}}S_{\beta_{11}\beta_{11}}^{-1}S_{\beta_{11}\theta_{12}}$, and $\mathbf{I}$ being the identity matrix.

In the asymptotic distribution of $LR_{2n}(\gamma_0, \beta_{11}^0)$, the $g(\cdot)$ function does not take a quadratic form, so $g(W_1)$ does not follow a scaled χ_1^2 distribution. Consequently, critical value computations for $LR_{2n}(\gamma_0, \beta_{11}^0)$ rely on simulations of $g(W_1)$, but this is not difficult and algorithms for CI construction based on LR_{2n} are given in Appendix E. In particular, the $(1-\alpha)$ CI of β_{11} can be constructed as

$$\bigcup_{q_{it}\in\Gamma}\{\beta_{11}|LR_{2n}(q_{it}, \beta_{11}) \le \hat{c}_\alpha^2\} =: \bigcup_{q_{it}\in\Gamma}\mathrm{CI}_2(\beta_{11}|q_{it}),$$

where $\hat{c}_\alpha^2$ is the level α critical value for $LR_{2n}(\gamma_0, \beta_{11}^0)$, and the complexity of $\mathrm{CI}_2(\beta_{11}|q_{it})$ is similar to that of $\mathrm{CI}(\beta_{11}|q_{it})$.

It is useful to compare the asymptotic distributions of the three groups of LR statistics. This is done in Appendix E which details the asymptotic null distributions of the LR_{1n} statistics and shows that the different formulations of the LR_n, LR_{1n} and LR_{2n} statistics do not affect the asymptotic component related to γ but only affect those related to β_{11}. This difference in impact on γ and β_{11} reveals how these two parameters γ and β have considerably different roles in the asymptotics.

6. Empirical application

This section reports an empirical implementation of our model selection, estimation and inference procedures. The application concerns about firm investment behavior with financing constraints, a topic analyzed in Hansen (1999) using the demeaning method in a fixed effects model.

Fazzari et al. (1988) (FHP hereafter) argue that the effect of a firm's cash flow on its investment is different with and without financing constraints. The implied hypothesis is that only when a firm faces constraints on external financial markets, will its cash flow positively influence its investment. The proposition suggests a threshold model to describe firm investment behavior. FHP use a low dividend to income ratio to indicate the existence of financing constraints as a financially constrained firm will usually retain earnings instead of paying dividends. In fact, FHP consider a double-threshold (rather than a single-threshold) model with the two threshold values arbitrarily chosen. We use the dataset in Hansen (1999) and choose the ratio of long-term debt to assets as the threshold variable, a high value of this ratio signifying the presence of financial constraints.

Following Hansen (1999), we start with the following SPTR model of three regimes to model the relationship between a firm's investment and its cash flow: [16]

$$\begin{aligned} I_{it} = \mathbf{x}_{1it}'\beta_1 &+ \left(CF_{i,t-1}\beta_{12} + \alpha_{1i} + u_{1it}\right)1(D_{i,t-1} \le \gamma_1) \\ &+ \left(CF_{i,t-1}\beta_{22} + \alpha_{2i} + u_{2it}\right)1(\gamma_1 < D_{i,t-1} \le \gamma_2) \\ &+ \left(CF_{i,t-1}\beta_{32} + \alpha_{3i} + u_{3it}\right)1(D_{i,t-1} > \gamma_2), \end{aligned} \tag{23}$$

$$\mathbf{x}_{1it}' = \left(Q_{i,t-1}, Q_{i,t-1}^2, Q_{i,t-1}^3, D_{i,t-1}, Q_{i,t-1}D_{i,t-1}\right) \tag{24}$$

with $i = 1, \cdots, N = 565$; $t = 1, \cdots, T = 14$. In (23) I_{it} is the ratio of investment to capital, CF_{it} is the ratio of cash flow to assets, D_{it} is the ratio of long-term debt to assets, and Q_{it} is the ratio of total market value to assets. All stock variables are measured at the end of year. The only difference between this model and Hansen's is that we allow α_i to be regime-specific. The model (23) focuses attention on the threshold effects of $CF_{i,t-1}$, and maintains a constant effect of $Q_{i,t-1}$ and $D_{i,t-1}$ on I_{it} across regimes. Nonlinear terms such as $Q_{i,t-1}^2, Q_{i,t-1}^3$ and $Q_{i,t-1}D_{i,t-1}$ are introduced to reduce spurious correlations due to omitted variable bias. So in calculating $\overline{\mathbf{x}}_i$, we only include the level variables $Q_{i,t-1}, CF_{i,t-1}$ and $D_{i,t-1}$. Note also that the averaging in $\overline{\mathbf{x}}_i$ starts from $t = 0$ and ends at $t = T$. In summary, our $\mathbf{z}_i' = \left(\overline{Q}_i, \overline{CF}_i, \overline{D}_i, 1\right)$ with $\underline{\mathbf{z}}_i = 1$, and $\mathbf{z}_i$ has different threshold effects on I_{it} in each regime. Due to the existence of variables without threshold effects in this model, the analysis in previous sections needs to be modified to accommodate the simplification – Appendix F details the required adjustments for this simplified model.

We start by determining the number of thresholds using the sequential testing procedure in Section 3. The results are reported in Table 1, where we use 500 replications in simulating the p-values.[17] Because in testing $m = 0$ vs. $m = 1$ the p-value is less than

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

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

Table 1
Tests for threshold effects.

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

Table 2
Parameter estimates and 95% CIs.

Parameters	Our Estimates	Our CIs	Critical Values of LR Statistics
γ	0.0142	[0.0125, 0.0165]	7.309
		[0.0086, 0.6091]	7.309
		[0.0125, 0.0165]	7.309
β_{12}	0.0523	[0.0268, 0.0778]	SE: 0.0130
		[0.0225, 0.0820]	14.313
		[0.0189, 0.0857]	265.558
β_{22}	0.0812	[0.0522, 0.1103]	SE: 0.0148
		[0.0499, 0.1126]	24.830
		[0.0526, 0.1098]	148.763

Note: the three CIs are reported in the following order: for γ, $\left(LR_n, LR_{1n}, LR_{2n}\right)$, for β_{12} and β_{22}, $\left(t, LR_n, LR_{2n}\right)$, SE means standard error, and other critical values are implied in Theorems 5–7

Table 3
Percentage of firms in each regime by year.

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

5% while in testing $m = 1$ vs. $m = 2$ it is much larger than 5%, we conclude that $m = 1$, i.e., there is only one threshold, a conclusion that differs from FHP and Hansen (1999) where two thresholds are detected. This may be surprising since we also explore the heterogeneity of α_i while Hansen (1999) assumes homogeneity. Because Hansen (1999) uses the LR test which involves the SSR after the demeaning operation, this is a good illustration of the unexpected effects of demeaning when the model is misspecified. In Appendix G, we conduct simulations to show that it is possible for Hansen's LR test to conclude $m = 2$ while our sequential testing procedure concludes $m = 1$, and also explain why this can occur using a simple example. From Table 4 of Hansen (1999), the third regime contains much fewer observations than the other two. We actually absorb the third regime into the second. Furthermore, $\hat{\beta}_{12} = 0.063$, $\hat{\beta}_{22} = 0.098$ and $\hat{\beta}_{32} = 0.039$ in Hansen (1999) are not increasing, which is unexpected from the relevant economic theory; in fact, the third regime appears to be a spurious outcome of assuming homogeneity of α_i. We provide more intuitive evidence on the number of regimes below as we show the LR statistics for γ in Fig. 4. While a type-II error in testing $H_0 : m = 1$ vs. $H_1 : m = 2$ is theoretically possible, the p-value of 0.63 suggests this is highly unlikely, given the inherent trade-off between type-I and type-II errors. Further support for the $m = 1$ outcome is that based on the formula in Section 4.2, with $\hat{\varphi} = 0.977$, very close to $\varphi = 1$ in the correctly specified models. In sum, our selected model is

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

Nonetheless, as noted earlier, our inferences on γ and β below remain robust to model misspecification.

We next estimate γ and construct corresponding CIs. The results are reported in Fig. 4 and Table 2. From Fig. 4, it is obvious that $LR_{1n}(\gamma) \le LR_n(\gamma) \le LR_{2n}(\gamma)$ for any $\gamma \in \Gamma$, so the widths of CIs based on $LR_{2n}(\gamma)$, $LR_n(\gamma)$ and $LR_{1n}(\gamma)$ should be increasing, which is shown explicitly in Table 2. The CIs based on $LR_n(\gamma)$ and $LR_{2n}(\gamma)$ are the same, but that based on $LR_{1n}(\gamma)$ is close to the whole Γ, and is not suggested in practice. The inset magnified portion of Fig. 4 shows the subtle differences between $LR_n(\gamma)$ and $LR_{2n}(\gamma)$ around $\hat{\gamma}$. Fig. 4 also intuitively confirms the conclusion of Table 1 - - the data imply only one threshold. Fig. 1 of Hansen (1999) shows that his $LR_n(\gamma)$ has a second dip around 0.53 besides at his $\hat{\gamma}_1 = 0.0157$ (which is close to our $\hat{\gamma}$), but this does not happen in either of our LR statistics. Based on our $\hat{\gamma}$, we report the percentage of firms in each regime by year in Table 3. These findings may be interpreted as combining the "high debt" class and the "medium debt" class of Table 4 in Hansen (1999). We also see a decreasing trend in the number of the "low debt" class.

We then estimate β and draw inferences accordingly. Since β_{12} and β_{22} are of primary interest, we neglect β_1 in Fig. 5 and Table 2. Fig. 5 shows how the CIs for β_{12} and β_{22} are constructed based on $LR_n(\cdot,\cdot)$ and $LR_{2n}(\cdot,\cdot)$. Combining the findings from Fig. 5 and Table 2, we draw the following conclusions. First, our $\hat{\beta}_{12}$ is close to Hansen's $\hat{\beta}_{12}$ (= 0.063), and $\hat{\beta}_{22}$ is between his $\hat{\beta}_{22}$ (= 0.098) and $\hat{\beta}_{32}$ (= 0.039). Since our $\hat{\beta}_{22}$ is greater than our $\hat{\beta}_{12}$, FHP's theory is supported. Also, our standard error (SE) of $\hat{\beta}_{12}$ is close to Hansen's

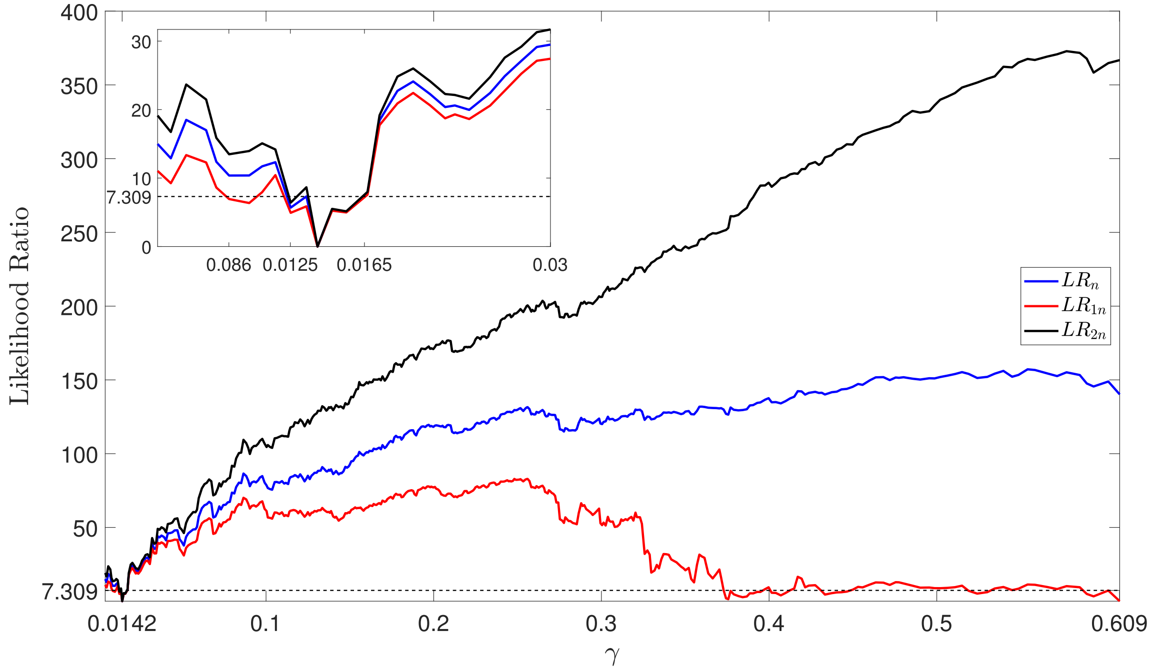

Fig. 4. Three CIs for γ: In the graphic 7.309 is the LR critical value at the 5% level for all three LR tests. The smaller inset figure provides a magnification of the main figure graphics over the shorter interval to highlight details of the LR statistics over the range $\gamma \in [0, 0.03]$ which are critical for CI construction.

Table 4
Tests for unobserved individual-specific threshold effects.

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

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

White-SE, 0.014, and the SE of our $\hat{\beta}_{22}$ is between those of his $\hat{\beta}_{22}$ and $\hat{\beta}_{32}$ (0.010 and 0.031). Second, the number of γ values in Γ_n involved in inverting $LR_{2n}(\cdot,\cdot)$ is much more than that in inverting $LR_n(\cdot,\cdot)$. This is because although $LR_n(\cdot,\cdot) \le LR_{2n}(\cdot,\cdot)$, the critical values associated with the latter are also much larger than those associated with the former as shown in Table 2. Third, there is no obvious trend between $\mathrm{CI}(\cdot|\gamma)$ and γ, e.g., the centers of $\mathrm{CI}(\cdot|\gamma)$ do not trend upward or downward as γ increases; the same conclusion applies to $\mathrm{CI}_2(\cdot|\gamma)$. This confirms the independence between $\hat{\beta}$ and $\hat{\gamma}$. As a result, only a few $\mathrm{CI}(\cdot|\gamma)$ and $\mathrm{CI}_2(\cdot|\gamma)$ intervals for γ around $\hat{\gamma}$ are relevant to the ultimate CIs (as shown on the y-axis); actually, the CIs for β_{12} and β_{22} based on $LR_n(\cdot,\cdot)$ are exactly the same as $\mathrm{CI}\left(\beta_{12}|\hat{\gamma}\right)$ and $\mathrm{CI}\left(\beta_{22}|\hat{\gamma}\right)$. This outcome differs markedly from the regular case where the target parameter and the nuisance parameter are not statistically independent such that the projection CI is much longer than the CI with the nuisance parameter fixed at its estimate. Fourth, the projection CI need not be wider than the t-type CI, and the CI based on $LR_{2n}(\cdot,\cdot)$ need not be narrower than that based on $LR_n(\cdot,\cdot)$. As mentioned in the first point above, the critical values associated with the two LR statistics are different, which is the reason why the lengths of the two LR-CIs for β are not comparable. This is very different from the LR-CIs for γ, where the critical values for all three LR statistics are the same such that the lengths of the three LR-CIs are sortable. We think the LR-CI based on $LR_{2n}(\cdot,\cdot)$ is most preferable. When less data points are present such as the left regime which contains 936 data points, the CI is wide to indicate the uncertainty in the estimation of β and the impact of the uncertainty of $\hat{\gamma}$, while when more data points are available such as the right regime which contains 6874 data points, the normal approximation is appropriate and the uncertainty in $\hat{\gamma}$ is dominated, the CI is close to the t-type CI.

Finally, the procedure in Appendix F is used to test the existence of unobserved individual-specific threshold effects. The test results are reported in Table 4. The p-value of the Wald test is 0.124, so the conclusion of rejection is not clear-cut. However, if we conduct t-tests on each element of $\psi_1 - \psi_2$, the overall picture is much clearer. Actually, three out of the four elements of $\psi_1 - \psi_2$ are significantly different from zero. This implies that the ambiguity in the Wald test result stems from just one element. Overall, our CRE estimation and the associated inferences are well-supported, whereas the demeaning approach proves unreliable in this application.

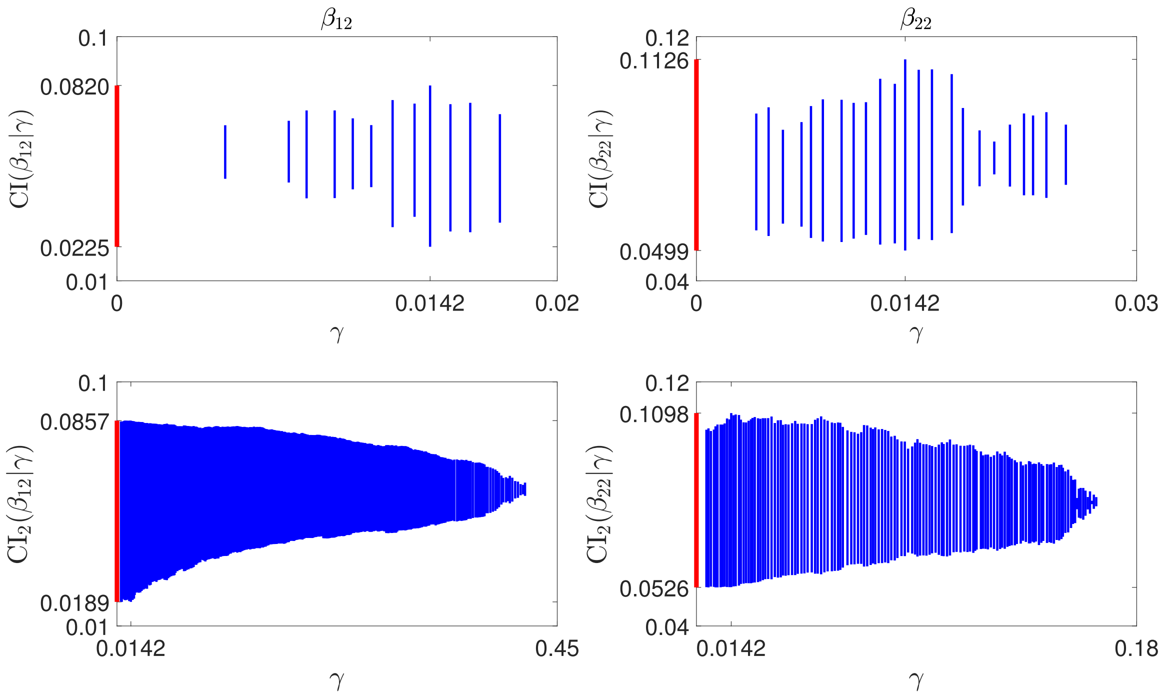

Fig. 5. Construction of CIs for β_{12} and β_{22} based on $LR_n(\cdot,\cdot)$ and $LR_{2n}(\cdot,\cdot)$.

7. Conclusion and discussion

This paper develops methods of model selection, estimation and inference in panel threshold regression with unobserved individual-specific threshold effects. Importantly, the new estimation and inference procedures are robust to underestimation of the number of regimes. Since WRD methods do not eliminate the endogeneity induced by fixed effects, CRE models are suggested as a valid approach for static and dynamic models, irrespective of the presence of unobserved individual-specific threshold effects. Forms of endogeneity in CRE models are less general than in fixed effects models, but are more practical since nonparametric components are not involved in estimation and inference. The IDKE method of Yu and Phillips (2018) can provide a consistent estimator of the threshold point γ in fixed effects models but when the regressor x_{it} dimension is large this method is impractical. On the other hand, given a consistent estimator of γ, WRD can be applied to obtain consistent estimators of β and this approach based on IDKE may serve as a benchmark in any future solution to fixed effects modeling with threshold effects.

A natural extension of the CRE models in this paper is the flexible CRE models of Bester and Hansen (2007). In a likelihood framework, they replace the parametric form of CRE by a sieve form and provide identification results for regular parameters like β in this paper. However, identification is not an issue for PTR given that the estimation procedure based on the IDKE indeed provides identification. Moreover, introducing a sieve CRE leads to an estimation challenge akin to the curse-of-dimensionality encountered with the IDKE.

Another possible solution to eliminate the effects of $\alpha_{\ell i}$ is to employ the functional differencing approach of Bonhomme (2012). The main focus of functional differencing is to deliver identification for regular parameters by moment conditions in a likelihood framework. But as mentioned above, identification for the nonregular parameter γ is not a problem, and as emphasized in Yu et al. (2018), it is better not to identify γ by moment conditions. One could envision a generalized differencing procedure whereby γ and β are estimated via an M -estimator instead of a Z-estimator, while avoiding nonparametric components in the objective function. However, to our knowledge, no such differencing scheme currently exists.

A potential remedy for the bias arising from the incidental parameters problem in nonlinear panel models is to allow T to grow large (when long panel data conditions are envisaged), albeit at a slower rate than N, followed by a debiasing step. For a comprehensive overview, see Arellano and Hahn (2007). Indeed, as T goes to infinity, we show in Appendix D that the WRD estimator of γ in Section 2 is consistent. While this result may plausibly extend to more general PTR models with covariates, the challenge of debiasing $\hat{\gamma}$ when T is relatively small remains unresolved, as existing literature primarily focuses on bias correction for regular parameters.

Finally, we do not consider interactive fixed effects in this paper. Miao et al. (2020) discuss such effects but assume they are invariant in the two regimes. Extending their estimation procedure to accommodate regime-specific interactive effects represents a promising direction for future research. Given that the setup is more general than that of this paper and Miao et al. (2020) essentially estimate the incidental parameters directly, we must correspondingly allow T to diverge to infinity to ensure the consistency of γ and β.

Funding

We thank Co-Editors Michael Jansson and Xiaohong Chen, an associate editor and three referees for helpful comments that have improved the paper. We also thank Bruce Hansen, Nan Liu, Yulong Wang, Yimeng Xie and seminar participants at AMES2022, AMES2023 and Xiamen University for helpful comments. We want to further thank N.R. Ram írez-Rondán for generously sharing his dataset with us. Phillips and Yu acknowledge support from the GRF of Hong Kong Government under Grant No. 17500719. Hong acknowledges support from the National Natural Science Foundation of China (NSFC) under Grant No. 72373175.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplementary material

Supplementary material associated with this article can be found in the online version at 10.1016/j.jeconom.2026.106337.

References

Arellano, M., Bond, S., 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. Rev. Econ. Stud. 58 (2), 277–297.
Arellano, M., Hahn, J., 2007. Understanding bias in nonlinear panel models: some recent developments. Econom. Soc. Monogr. 43, 381–409.
Bai, J., 1997. Estimation of a change point in multiple regression models. Rev. Econ. Stat. 79 (4), 551–563.
Bai, J., 1997b. Estimating multiple breaks one at a time. Econom. Theory 13 (3), 315–352.
Bai, J., Perron, P., 1998. Estimating and testing linear models with multiple structural changes. Econometrica, 66 (1), 47–78.
Bai, J., Perron, P., 2006. Multiple structural change models: a simulation analysis. Econom. Theory Pract.: Front. Anal. Appl. Res. 1, 212–38.
Bester, A., Hansen, C., 2007. Flexible correlated random effects estimation in panel models with unobserved heterogeneity. Unpublished manuscript, Grad. Sch. Bus., Univ. Chicago
Bonhomme, S., 2012. Functional differencing. Econometrica 80 (4), 1337–1385.
Chamberlain, G., 1982. Multivariate regression models for panel data. J. Econom. 18 (1), 5–46.
Chamberlain, G., 1984. Panel data. Handb. Econom. 2, 1247–1318.
Chamberlain, G., Moreira, M.J., 2009. Decision theory applied to a linear panel data model. Econometrica 77 (1), 107–133.
Chan, K.S., 1993. Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. Ann. Stat. , 520–533.
Chan, K.S., Tsay, R.S., 1998. Limiting properties of the least squares estimator of a continuous threshold autoregressive model. Biometrika 85 (2), 413–426.
Dang, V.A., Kim, M., Shin, Y., 2012. Asymmetric capital structure adjustments: new evidence from dynamic panel threshold models. J. Empir. Finance 19 (4), 465–482.
De Long, D.M., 1981. Crossing probabilities for a square root boundary by a bessel process. Commun. Stat.-Theory Methods 10 (21), 2197–2213.
Fazzari, S.M., Hubbard, R.G., Petersen, B.C., 1988. Financing constraints and corporate investment. Brookings Papers on Economic Activity, 141–195.
Gonzalo, J., Pitarakis, J.-Y., 2002. Estimation and model selection based inference in single and multiple threshold models. J. Econom. 110 (2), 319–352.
Gørgens, T., Ẃurtz, A.H., 2019. Threshold regression with endogeneity for short panels. Econometrics 7 (2), 23.
Hansen, B.E., 1996. Inference when a nuisance parameter is not identified under the null hypothesis. Econom.: J. Econom. Soc. 64 (2), 413–430.
Hansen, B.E., 1999. Threshold effects in non-dynamic panels: estimation, testing, and inference. J. Econom. 93 (2), 345–368.
Hansen, B.E., 2000. Sample splitting and threshold estimation. Econometrica 68 (3), 575–603.
Hansen, B.E., 2017. Regression kink with an unknown threshold. J. Bus. Econ. Stat. 35 (2), 228–240.
Hsiao, C., Pesaran, M.H., Tahmiscioglu, A.K., 2002. Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods. J. Econom. 109 (1), 107–150.
Kourtellos, A., Stengos, T., Tan, C.M., 2016. Structural threshold regression. Econom. Theory 32 (4), 827–860.
Kremer, S., Bick, A., Nautz, D., 2013. Inflation and growth: new evidence from a dynamic panel threshold analysis. Empir. Econ. 44 (2), 861–878.
Miao, K., Li, K., Su, L., 2020. Panel threshold models with interactive fixed effects. J. Econom. 219 (1), 137–170.
Mundlak, Y., 1978. On the pooling of time series and cross section data. Econom.: J. Econom. Soc. 46 (1), 69–85.
Phillips, P. C.B., Moon, H.R., 1999. Linear regression limit theory for nonstationary panel data. Econometrica 67 (5), 1057–1111.
Ramírez-Rondán, N.R., 2020. Maximum likelihood estimation of dynamic panel threshold models. Econom. Rev. 39 (3), 260–276.
Seo, M.H., Shin, Y., 2016. Dynamic panels with threshold effect and endogeneity. J. Econom. 195 (2), 169–186.
Shin, Y., Kim, M., 2011. Bootstrap-Based Bias Corrected within Estimation of Threshold Regression Models in Dynamic Panels. Technical Report. mimeo.
Wang, C.-H., Lin, E.S., 2010. Panel Threshold Regression Models with Endogenous Threshold Variables. Technical Report. mimeo.
Wooldridge, J.M., 2010. Econometric Analysis of Cross Section and Panel Data. MIT press.
Yu, P., 2012. Likelihood estimation and inference in threshold regression. J. Econom. 167 (1), 274–294.
Yu, P., 2013. Integrated Quantile Threshold Regression and Distributional Threshold Effects. Technical Report. mimeo.
Yu, P., 2015. Adaptive estimation of the threshold point in threshold regression. J. Econom. 189 (1), 83–100.
Yu, P., 2019. Asymptotics for Threshold Regression Under Misspecification. Technical Report. mimeo.
Yu, P., Fan, X., 2021. Threshold regression with a threshold boundary. J. Bus. Econ. Stat. 39 (4), 953–971.
Yu, P., Liao, Q., Phillips, P. C.B., 2018. Inferences and Specification Testing in Threshold Regression with Endogeneity. Technical Report. mimeo.
Yu, P., Liao, Q., Phillips, P. C.B., 2024. New control function approaches in threshold regression with endogeneity. Econom. Theory 40 (5), 1065–1119.
Yu, P., Phillips, P. C.B., 2018. Threshold regression with endogeneity. J. Econom. 203 (1), 50–68.