The Econometrics Journal

Econometrics Journal (2021), volume **00**, pp. 1–20.
doi: 10.1093/ectj/utaa035

Panel kink threshold regression model with a covariate-dependent threshold

Lixiong Yang[*], Chunli Zhang[†], Chingnun Lee[‡] and I-Po Chen[§]

[*]*School of Management, Lanzhou University, 222 South Tianshui Road, Lanzhou 730000, China.*
Email: ylx@lzu.edu.cn

[†]*School of Economics, Lanzhou University, 222 South Tianshui Road, Lanzhou 730000, China.*
Email: zhangchl20@lzu.edu.cn

[‡]*Institute of Economics, National Sun Yet-sen University, 70 Lien-hai Road, Kaohsiung 80424, Taiwan.*
Email: lee_econ@mail.nsysu.edu.tw

[§]*Institute of Economics, National Sun Yet-sen University, 70 Lien-hai Road, Kaohsiung 80424, Taiwan.*
Email: j63007@gmail.com

First version received: 6 August 2019; final version accepted: 11 November 2020.

Summary: This article extends the kink threshold regression model with a constant threshold to a panel data framework with a covariate-dependent threshold, where the threshold is modeled as a function of informative covariates. We suggest an estimator based on the within-group transformation and propose test statistics for kink threshold effect and threshold constancy. We establish the asymptotic joint normality of the slope and threshold estimators and derive the limiting distributions of the test statistics. Our asymptotic results show that the inclusion of a covariate-dependent threshold does not affect the asymptotic joint normality of the slope and threshold estimates in the kink threshold regression model. Monte Carlo simulations show that the finite-sample proprieties of the proposed estimator and test statistics are generally satisfactory.

Keywords: *Panel data*, *kink threshold regression*, *covariate-dependent threshold*, *Monte Carlo simulations*.

JEL codes: *C12*, *C13*, *C33*.

1. INTRODUCTION

Since their introduction, threshold models have received much attention in econometrics, applied economics, and other fields; see Tong (1990) and Hansen (2000), among others. The threshold literature often assumes that the regression function has either a jump or a kink at a threshold point. For instance, Chan (1993) and Hansen (2000) have focused on the threshold models with a jump at the threshold point, whereas Chan and Tsay (1998), Hansen (2017), and Zhang et al. (2017) have focused on kink threshold regression (KTR) models. The kink threshold model is a subclass of threshold models subject to the requirement that the regression function is everywhere continuous, but the slope has a kink at a threshold point.

It is worth noting that the KTR models differ from the classical regression kink design studied by Nielsen et al. (2010) and Card et al. (2015). Regression kink design was popularised by Card et al. (2015) as a modification of the widely applied regression discontinuity design. The thresholds in the articles on regression discontinuity design and regression kink design are used to determine the treatment variable, whereas the thresholds in the KTR models specify a nonlinear effect, and hence KTR is not a quasi-experimental variation. In the classical regression kink design, the parameter of interest is the change in the slopes at the policy threshold; however, the parameter of interest is the slope coefficient itself in the KTR model investigated in the present article.[1]

Hansen (2017) proposed a KTR model with an unknown threshold for time-series data. Zhang et al. (2017) extended the KTR model with an unknown threshold to the panel data framework. However, all these papers assume a constant threshold, which might be impractical in applications. To illustrate this, consider the classical application of the KTR model in the nonlinear effect of public debt on economic growth in Hansen (2017). As noted by the literature, including Reinhart and Rogoff (2010), Cochrane (2011), and Krause and Moyen (2016), (high) inflation can erode the real value of public debt burden, and thus, higher inflation might plausibly ease negative growth effects of debt. On the contrary, other studies emphasise that the opposite might be true (e.g., Akitoby et al., 2017, and Kriwoluzky et al., 2019)—that is, that inflation, as a form of sovereign default, could increase the risk premium and affect the debt tolerance of investors (and hence increase public debt costs), which might reinforce the negative growth effects of debt. Overall, the two aforementioned effects can lead to the debt threshold depending on inflation; nevertheless, given the two opposing forces of inflation, the net effect of inflation on the debt threshold depends on which effect dominates and hence remains unclear. Thus, a covariate-dependent threshold model appears to be more suitable and more useful than a constant one in applications. However, in terms of a covariate-dependent threshold, only a few cases have appeared in the empirical literature so far.[2] Moreover, an asymptotic distribution theory for estimation and testing in the covariate-dependent threshold models is still invalid.

In this article, we propose a panel KTR model with a covariate-dependent threshold (PKTR-CDT), where the threshold is modeled as a function of informative covariates. The proposed model is an extension of the KTR model with an unknown constant threshold described by Hansen (2017) and Zhang et al. (2017). In estimation and model specification testing, we suggest an estimator based on the within-group transformation (following Hansen, 1999) and propose test statistics for kink effect and threshold constancy. Then, assuming a fixed threshold effect, we derive the asymptotic properties of the proposed estimator and test statistics when the number of individuals, N, tends to infinity under a fixed time period T. As in the KTR models with a constant threshold described by Chan and Tsay (1998), Hansen (2017), and Zhang et al. (2017), we show that the slope and threshold estimators are jointly asymptotically normal with $\sqrt{N}$ convergence rate and a non-zero asymptotic covariance—that is, that the inclusion of a covariate-dependent threshold does not affect the asymptotic joint normality of the slope and threshold estimates in the KTR model. We also establish the limiting distributions of the proposed test statistics, and

[1] We highly appreciate a referee's raising this point with us.

[2] For example, Dueker et al. (2013) extended the classical smooth-transition autoregressive models by allowing for a state-dependent threshold and applied their model to forecast US short-term interest rates; more recently, Yang and Su (2018) introduced a flexible KTR model with a covariate-dependent threshold for time-series data and applied the model to investigate the relationship between debt and growth.

we assess the finite-sample performances of the estimator and test statistics through Monte Carlo simulations.

The remainder of this article is organised as follows. Section 2 introduces the PKTR-CDT, describes least-squares estimation of the model parameters, and proposes test statistics for threshold effect and threshold constancy. In Sections 3, the Monte Carlo simulation study is described, and simulation results are reported. Section 4 concludes. We present a detailed and formal proof of the asymptotic results of the proposed estimator and test statistics in the Appendix, and we extend the PKTR-CDT to the dynamic panel context in the online Appendix.

2. PANEL KTR WITH A COVARIATE-DEPENDENT THRESHOLD

Consider the following panel KTR with a covariate-dependent threshold,

$$y_{it} = \beta_1^-(x_{it} - \gamma_{it})_- + \beta_1^+(x_{it} - \gamma_{it})_+ + \boldsymbol{\beta}_2' \boldsymbol{z}_{it} + \alpha_i + \varepsilon_{it}, \tag{2.1}$$

for $i = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$, where y_{it}, x_{it}, and ε_{it} are scalars, and $\boldsymbol{z}_{it}$ is an l-dimensional vector of regressors that include the covariates $\boldsymbol{q}_{it}$ (defined in the next equation).[3] α_i represents the unobserved individual heterogeneity, which can be correlated with x_{it} and $\boldsymbol{z}_{it}$. $(x_{it} - \gamma_{it})_- = \min[x_{it} - \gamma_{it}, 0]$ and $(x_{it} - \gamma_{it})_+ = \max[x_{it} - \gamma_{it}, 0]$ denote the negative part and positive part of $x_{it} - \gamma_{it}$, respectively. The slope with respect to x_{it} equals β_1^- if $x_{it} \leq \gamma_{it}$ and equals β_1^+ if $x_{it} > \gamma_{it}$. Thus, the regression function has a kink at $x_{it} = \gamma_{it}$. γ_{it} is a covariate-dependent threshold, which is specified as a linear combination of informative covariates $\boldsymbol{q}_{it} = (q_{1,it}, \ldots, q_{k,it})'$ explaining variation in thresholds over i and t; that is,

$$\gamma_{it} = \gamma_0 + \boldsymbol{\gamma}_1' \boldsymbol{q}_{it}, \tag{2.2}$$

where γ_0 represents an unknown threshold intercept, and $\boldsymbol{\gamma}_1 = (\gamma_{11}, \ldots, \gamma_{1k})'$ is a vector of unknown slope parameters. Here, $\boldsymbol{q}_{it}$ cannot include the variables in x_{it} contemporaneously because of the problem of perfect multicollinearity.

Note that the model defined in (2.1) is a panel version of Yang and Su's (2018) KTR with a covariate-dependent threshold. The model can also be treated as an extension of Zhang et al.'s (2017) panel KTR with an unknown constant threshold, which is a special case of our model when $\boldsymbol{\gamma}_1 = \mathbf{0}_{k\times 1}$.

As in Chan and Tsay (1998) and Hansen (2017), we restrict our model to the context in which the slope with respect to x_{it} has a kink, but the regression function is continuous in the regressors x_{it} and $\boldsymbol{z}_{it}$. Considering the slope change has the advantage of capturing inverted V-shaped relationships, which is important given that the economic literature often suggests an inverted V-shaped relationship between economic variables. As in Zhang et al. (2017), we restrict attention to the estimation and inference in the PKTR-CDT model when $N \to \infty$ and T is fixed. Furthermore, we also restrict the model to the case in which the regression segments and threshold setting are linear rather than nonparametric (or nonlinear). This linear setting is useful and reasonable in cases of moderate sample sizes, where nonparametric methods may work poorly.[4]

[3] It is important to include q_{it} in z_{it} for testing the kink effect against the linearity, which ensures that the linear model is nested in the kink threshold model (2.1) under the null $\beta_1^- = \beta_1^+$. However, this does not affect the consistency of the proposed estimator of PKTR-CDT given the presence of the kink threshold effect, because whether or not to include q_{it} in z_{it} does not affect the conditions under which Theorem 2.1 holds.

[4] It would be useful to extend our model to a nonparametric threshold specification. We thank an anonymous referee for raising this point with us.

2.1. The estimates and asymptotic properties

For convenience, we rewrite the model in a more compact form. Denote $\boldsymbol{\beta} = [\beta_1^-, \beta_1^+, \boldsymbol{\beta}_2']'$, $\boldsymbol{\gamma} = (\gamma_0, \boldsymbol{\gamma}_1')'$, and $\boldsymbol{x}_{it}(\boldsymbol{\gamma}) = [(x_{it} - \gamma_{it})_-, (x_{it} - \gamma_{it})_+, z_{it}']'$. Then, model (2.1) can be rewritten as

$$y_{it} = \boldsymbol{\beta}'\boldsymbol{x}_{it}(\boldsymbol{\gamma}) + \alpha_i + \varepsilon_{it}. \tag{2.3}$$

To eliminate the individual effect α_i, we take averages of (2.3) over the time

$$\bar{y}_i = \boldsymbol{\beta}'\bar{\boldsymbol{x}}_i(\boldsymbol{\gamma}) + \alpha_i + \bar{\varepsilon}_i, \tag{2.4}$$

where $\bar{y}_i = T^{-1}\sum_{t=1}^T y_{it}$, $\bar{\boldsymbol{x}}_i(\boldsymbol{\gamma}) = T^{-1}\sum_{t=1}^T \boldsymbol{x}_{it}(\boldsymbol{\gamma})$, and $\bar{\varepsilon}_i = T^{-1}\sum_{t=1}^T \varepsilon_{it}$. Then, we remove individual-specific means by taking the difference between (2.3) and (2.4),

$$\ddot{y}_{it} = \boldsymbol{\beta}'\ddot{\boldsymbol{x}}_{it}(\boldsymbol{\gamma}) + \ddot{\varepsilon}_{it}, \tag{2.5}$$

in which $\ddot{y}_{it} = y_{it} - \bar{y}_i$, $\ddot{\boldsymbol{x}}_{it}(\boldsymbol{\gamma}) = \boldsymbol{x}_{it}(\boldsymbol{\gamma}) - \bar{\boldsymbol{x}}_i(\boldsymbol{\gamma})$, and $\ddot{\varepsilon}_{it} = \varepsilon_{it} - \bar{\varepsilon}_i$.

Let $\ddot{y}_i = (\ddot{y}_{i1}, \ldots, \ddot{y}_{it})'$, $\ddot{\boldsymbol{x}}_i(\boldsymbol{\gamma}) = (\ddot{\boldsymbol{x}}_{i1}(\boldsymbol{\gamma}), \ldots, \ddot{\boldsymbol{x}}_{it}(\boldsymbol{\gamma}))'$, and $\ddot{\varepsilon}_i = (\ddot{\varepsilon}_{i1}, \ldots, \ddot{\varepsilon}_{it})'$. Let $\ddot{\boldsymbol{Y}}$, $\ddot{\boldsymbol{X}}(\boldsymbol{\gamma})$, and $\ddot{\varepsilon}$ denote the data stacked over all individuals, i.e., $\ddot{\boldsymbol{Y}} = (\ddot{y}_1', \ldots, \ddot{y}_N')'$, $\ddot{\boldsymbol{X}}(\boldsymbol{\gamma}) = (\ddot{\boldsymbol{x}}_1'(\boldsymbol{\gamma}), \ldots, \ddot{\boldsymbol{x}}_N'(\boldsymbol{\gamma}))'$, and $\ddot{\varepsilon} = (\ddot{\varepsilon}_1', \ldots, \ddot{\varepsilon}_N')'$. Thus, (2.5) can be rewritten as $\ddot{\boldsymbol{Y}} = \ddot{\boldsymbol{X}}(\boldsymbol{\gamma})\boldsymbol{\beta} + \ddot{\varepsilon}$. For any given γ_0 and $\boldsymbol{\gamma}_1'$, we can estimate the slope coefficient $\boldsymbol{\beta}$ as follows:

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\gamma}) = \left[\ddot{\boldsymbol{X}}'(\boldsymbol{\gamma})\ddot{\boldsymbol{X}}(\boldsymbol{\gamma})\right]^{-1}\ddot{\boldsymbol{X}}'(\boldsymbol{\gamma})\ddot{\boldsymbol{Y}}. \tag{2.6}$$

The parameters of the covariate-dependent threshold can be estimated as

$$\hat{\boldsymbol{\gamma}} = (\hat{\gamma}_0, \hat{\boldsymbol{\gamma}}_1')' = \arg\min_{\boldsymbol{\gamma}\in\boldsymbol{\Gamma}} \widetilde{SSR}_{NT}(\boldsymbol{\gamma}), \tag{2.7}$$

where $\widetilde{SSR}_{NT}(\boldsymbol{\gamma}) = \frac{1}{N}(\ddot{\boldsymbol{Y}} - \ddot{\boldsymbol{X}}(\boldsymbol{\gamma})\hat{\boldsymbol{\beta}}(\boldsymbol{\gamma}))'(\ddot{\boldsymbol{Y}} - \ddot{\boldsymbol{X}}(\boldsymbol{\gamma})\hat{\boldsymbol{\beta}}(\boldsymbol{\gamma}))$, $\boldsymbol{\Gamma} = \Gamma_0 \times \Gamma_1$, and Γ_0 and $\Gamma_1 = \Gamma_{11} \times \Gamma_{12} \times \cdots \times \Gamma_{1k}$ are the parameter spaces and are assumed to be compact. Once the estimates $\hat{\boldsymbol{\gamma}} = (\hat{\gamma}_0, \hat{\boldsymbol{\gamma}}_1')'$ are obtained, a natural estimate for $\boldsymbol{\beta}$ is given by $\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\gamma}})$. The least-squares estimator has been widely used in the threshold literature; see, e.g., Hansen (1999), Hansen (2017), Zhang et al. (2017), and Yu and Fan (2020), among others. Recently, Yu and Fan (2020) documented that the threshold parameters may not be identified uniquely in the discontinuous threshold models with a threshold boundary, leading to a difficulty in calculating the least-squares estimator of the threshold parameters, but in the continuous PKTR-CDT model this difficulty disappears.[5]

In applications, one may specify Γ_0 as $[x_{(0.15N)}, x_{(0.85N)}]$, with $x_{(\eta)}$ being the ηth-order statistic of x_{it}, and set Γ_{1j} as $[-r_{max}, r_{max}]$, in which $r_{max} = \max\{|x_{(0.15N)}|, |x_{(0.85N)}|\}$ for $j = 1, 2, \ldots, k$. Furthermore, to implement the minimisation in (2.7), we suggest a two-step approach based on concentration and grid search. First, for any given γ_0, we compute the sum of squared errors $\widetilde{SSR}_{NT}(\boldsymbol{\gamma}) = \widetilde{SSR}_{NT}(\gamma_0, \boldsymbol{\gamma}_1')$ for each $\boldsymbol{\gamma}_1 \in \Gamma_1$. Second, we compute the minimum $\widetilde{SSR}_{NT}(\gamma_0, \tilde{\boldsymbol{\gamma}}_1'(\gamma_0)) = \arg\min_{\boldsymbol{\gamma}_1} \widetilde{SSR}_{NT}(\gamma_0, \boldsymbol{\gamma}_1')$, and we estimate the threshold parameters by $(\hat{\gamma}_0, \hat{\boldsymbol{\gamma}}_1') \equiv (\hat{\gamma}_0, \tilde{\boldsymbol{\gamma}}_1'(\hat{\gamma}_0)) = \arg\min_{\gamma_0} \widetilde{SSR}_{NT}(\gamma_0, \tilde{\boldsymbol{\gamma}}_1'(\gamma_0))$. As suggested by Hansen (1999), it might be undesirable to select values for the threshold parameters $(\hat{\gamma}_0, \hat{\boldsymbol{\gamma}}_1')$ that sort too few data into one or the other regime. In application, we suggest that the aforementioned grid search be restricted to values of $(\hat{\gamma}_0, \hat{\boldsymbol{\gamma}}_1')$ such that a minimal percentage of the observations (say, 10% or 15%) lies in both regimes; the robustness of the results may be verified by using different choices of the percentage.

[5] This is illustrated by Monte Carlo simulations in an unreported appendix. We thank an anonymous referee for raising this point with us.

The estimation procedure described above works well in the case in which a few covariates are used to model the covariate-dependent threshold. There may be a practical difficulty in the case of large-dimensional $\boldsymbol{q}_{it}$, and the present article does not discuss how to select relevant covariates $\boldsymbol{q}_{it}$ among many potential covariates. For a discussion of large-dimensional $\boldsymbol{q}_{it}$ and how to select $\boldsymbol{q}_{it}$, please refer to Lee et al. (2018).

To derive the asymptotic distribution for the proposed estimator of $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\gamma}')'$, define the true value

$$\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0', \boldsymbol{\gamma}_0')' = \underset{\boldsymbol{\beta}\in\mathbf{B}, \boldsymbol{\gamma}\in\boldsymbol{\Gamma}}{\arg\min}\, L(\boldsymbol{\beta}, \boldsymbol{\gamma}), \tag{2.8}$$

where $L(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \sum_{t=1}^{T} E[\ddot{y}_{it} - \boldsymbol{\beta}'\ddot{\boldsymbol{x}}_{it}(\boldsymbol{\gamma})]^2$ is the squared error loss. For any given $\boldsymbol{\gamma} \in \Gamma$, the concentrated squared loss is

$$L^c(\boldsymbol{\gamma}) = \sum_{t=1}^{T} E[\ddot{y}_{it} - \boldsymbol{\beta}'(\boldsymbol{\gamma})\ddot{\boldsymbol{x}}_{it}(\boldsymbol{\gamma})]^2, \tag{2.9}$$

in which $\boldsymbol{\beta}(\boldsymbol{\gamma}) = \left(\sum_{t=1}^{T} E\left[\ddot{\boldsymbol{x}}_{it}(\boldsymbol{\gamma})\ddot{\boldsymbol{x}}_{it}'(\boldsymbol{\gamma})\right]\right)^{-1} \sum_{t=1}^{T} E\left[\ddot{\boldsymbol{x}}_{it}(\boldsymbol{\gamma})\ddot{y}_{it}\right]$. By concentration, $\boldsymbol{\gamma}_0$ is the minimiser of $L^c(\boldsymbol{\gamma})$ and $\boldsymbol{\beta}_0 = \boldsymbol{\beta}(\boldsymbol{\gamma}_0)$.

Let $1_{it}^{-}(\boldsymbol{\gamma}) = 1(x_{it} \leq \gamma_{it})$, and $\ddot{1}_{it}^{-}(\boldsymbol{\gamma}) = 1_{it}^{-}(\boldsymbol{\gamma}) - \frac{1}{T}\sum_{t=1}^{T} 1_{it}^{-}(\boldsymbol{\gamma})$. Similarly, we can define $1_{it}^{+}(\boldsymbol{\gamma})$ and $\ddot{1}_{it}^{+}(\boldsymbol{\gamma})$. Let $\ddot{\boldsymbol{q}}_{it}^{-}(\boldsymbol{\gamma}) = \boldsymbol{q}_{it}1_{it}^{-}(\boldsymbol{\gamma}) - \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{q}_{it}1_{it}^{-}(\boldsymbol{\gamma})$, $\ddot{\boldsymbol{q}}_{it}^{+}(\boldsymbol{\gamma}) = \boldsymbol{q}_{it}1_{it}^{+}(\boldsymbol{\gamma}) - \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{q}_{it}1_{it}^{+}(\boldsymbol{\gamma})$, $e_{it}(\boldsymbol{\theta}) = \ddot{y}_{it} - \boldsymbol{\beta}'\ddot{\boldsymbol{x}}_{it}(\boldsymbol{\gamma})$, $\boldsymbol{h}_{it}(\boldsymbol{\theta}) = -\frac{\partial e_{it}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \begin{pmatrix} \ddot{\boldsymbol{x}}_{it}(\boldsymbol{\gamma}) \\ -\beta_1^{-}\ddot{1}_{it}^{-}(\boldsymbol{\gamma}) - \beta_1^{+}\ddot{1}_{it}^{+}(\boldsymbol{\gamma}) \\ -\beta_1^{-}\ddot{\boldsymbol{q}}_{it}^{-}(\boldsymbol{\gamma}) - \beta_1^{+}\ddot{\boldsymbol{q}}_{it}^{+}(\boldsymbol{\gamma}) \end{pmatrix}$, and $\boldsymbol{D}_{it}(\boldsymbol{\gamma}) = -\frac{\partial \boldsymbol{h}_{it}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} = \begin{pmatrix} 0 & 0 & \mathbf{0}_{1\times l} & \ddot{1}_{it}^{-}(\boldsymbol{\gamma}) & \ddot{\boldsymbol{q}}_{it}^{-'}(\boldsymbol{\gamma}) \\ 0 & 0 & \mathbf{0}_{1\times l} & \ddot{1}_{it}^{+}(\boldsymbol{\gamma}) & \ddot{\boldsymbol{q}}_{it}^{+'}(\boldsymbol{\gamma}) \\ \mathbf{0}_{l\times 1} & \mathbf{0}_{l\times 1} & \mathbf{0}_{l\times l} & \mathbf{0}_{l\times 1} & \mathbf{0}_{l\times k} \\ \ddot{1}_{it}^{-}(\boldsymbol{\gamma}) & \ddot{1}_{it}^{+}(\boldsymbol{\gamma}) & \mathbf{0}_{1\times l} & 0 & \mathbf{0}_{1\times k} \\ \ddot{\boldsymbol{q}}_{it}^{-}(\boldsymbol{\gamma}) & \ddot{\boldsymbol{q}}_{it}^{+}(\boldsymbol{\gamma}) & \mathbf{0}_{k\times l} & \mathbf{0}_{k\times 1} & \mathbf{0}_{k\times k} \end{pmatrix}$. Denote $\boldsymbol{h}_{it} = \boldsymbol{h}_{it}(\boldsymbol{\theta}_0)$, $e_{it} = e_{it}(\boldsymbol{\theta}_0)$, $\boldsymbol{D}_{it} = \boldsymbol{D}_{it}(\boldsymbol{\gamma}_0)$, $\boldsymbol{h}_i = (h_{i1}, h_{i2}, \ldots, h_{it})'$, and $\boldsymbol{e}_i = (e_{i1}, e_{i2}, \ldots, e_{it})'$.

To establish the asymptotic distribution of $\hat{\boldsymbol{\theta}}$, we need the following assumptions.

ASSUMPTION 1. *(i) For each t, $\boldsymbol{w}_{it} = (y_{it}, x_{it}, \boldsymbol{z}_{it}', \boldsymbol{q}_{it}')$ are independently identically distributed (i.i.d.) across i; (ii) for some $r > 1$, $E|y_{it}|^{4r} < \infty$, $E|x_{it}|^{4r} < \infty$, $E|\boldsymbol{z}_{it}|^{4r} < \infty$, and $E|\boldsymbol{q}_{it}|^{4r} < \infty$; (iii) $E[\varepsilon_{it}|(x_{is}, \boldsymbol{z}_{is}', \boldsymbol{q}_{is}', \alpha_i : 1 \leq s \leq T)] = 0$.*

ASSUMPTION 2. *(i) $\boldsymbol{w}_{it}$ has a probability density function $\boldsymbol{f}_{\boldsymbol{w},t}(\boldsymbol{w})$, where $\boldsymbol{w}_{it} = (y_{it}, x_{it}, \boldsymbol{z}_{it}', \boldsymbol{q}_{it}')$; (ii) $(\boldsymbol{w}_{it}, \boldsymbol{w}_{is})$ has a joint probability density function $\boldsymbol{f}_{\boldsymbol{w},ts}$; (iii) x_{it} has a conditional probability density function given $\boldsymbol{q}_{it} = \boldsymbol{q}$, satisfying $\max_{1\leq t\leq T} f_{\boldsymbol{q},t}(x|\boldsymbol{q}) \leq \bar{f}_{\boldsymbol{q}} < \infty$; and $F_{\boldsymbol{q}}$ is the corresponding conditional cumulative distribution function of x_{it} conditional on $\boldsymbol{q}_{it}$.*

ASSUMPTION 3. $\inf_{\boldsymbol{\gamma}\in\boldsymbol{\Gamma}} \det \boldsymbol{Q}(\boldsymbol{\gamma}) > 0$, *where $\boldsymbol{Q}(\boldsymbol{\gamma}) = \sum_{t=1}^{T} E[\ddot{\boldsymbol{x}}_{it}(\boldsymbol{\gamma})\ddot{\boldsymbol{x}}_{it}'(\boldsymbol{\gamma})]$ and $\boldsymbol{\Gamma}$ is compact.*

ASSUMPTION 4. *$\beta_1^{+} - \beta_1^{-}$ is a constant, and $\boldsymbol{\beta} \in \mathbf{B} \subset \mathbb{R}^{l+2}$, where $\mathbf{B}$ is compact.*

ASSUMPTION 5. *$\boldsymbol{\gamma}_0 = \arg\min_{\boldsymbol{\gamma}\in\boldsymbol{\Gamma}} L^c(\boldsymbol{\gamma})$ is unique.* □

These assumptions are almost the same as those in Hansen (2017) and Zhang et al. (2017). The difference is the inclusion of $\boldsymbol{q}_{it}$ in Assumptions 1–2, in which we assume the $4r$-th moment condition to ensure that the central limit theorem and the weak law of large numbers hold, and we impose strict exogeneity of the regressors and the covariates affecting the threshold. This rules out dynamic panel data models, which will be discussed in the online Appendix. Assumptions 3–5 are identification conditions. Assumptions 3 and 4 ensure that the projection coefficient $\boldsymbol{\beta}(\boldsymbol{\gamma})$ is well defined and the parameter spaces for $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ are compact. Assumption 5 rules out the case of multiple best-fitting threshold parameters $\boldsymbol{\gamma}$. Hansen (2017) and Zhang et al. (2017) showed that the slope and threshold estimators in KTR with an unknown constant threshold are jointly asymptotically normal for time-series data and panel data, respectively. We extend the asymptotic theory to panel KTR with a covariate-dependent threshold.

THEOREM 2.1. *Under Assumptions 1–5, as $N \to \infty$,*

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{d} N(0, \boldsymbol{V}), \tag{2.10}$$

where $\boldsymbol{V} = \boldsymbol{G}^{-1}\boldsymbol{S}\boldsymbol{G}^{-1}$, $\boldsymbol{S} = var(\boldsymbol{h}_i'\boldsymbol{e}_i)$, and $\boldsymbol{G} = E(\boldsymbol{h}_i'\boldsymbol{h}_i) + \sum_{t=1}^{T} E(\boldsymbol{D}_{it}\boldsymbol{e}_{it})$.

Proof. See the Appendix. □

Theorem 2.1 indicates that the inclusion of the covariates $\boldsymbol{q}_{it}$ in the covariate-dependent threshold does not affect the asymptotic joint normality of the slope and threshold estimates in the PKTR-CDT model. Following the proof in the Appendix, we can easily verify that such an asymptotic joint normality also holds in Yang and Su's (2018) KTR with a covariate-dependent threshold for time-series data.

As discussed in Hansen (2017), statistical inference based on the asymptotic distribution may perform poorly in small samples because the least-square criterion is nonquadratic with respect to the threshold parameters $\boldsymbol{\gamma}$, possibly leading to poor quadratic (e.g., normal) approximations unless sample sizes are quite large. In applications, we can construct confidence intervals for the threshold parameters by inverting the following test statistic for the null $H_0 : \boldsymbol{\gamma} = \boldsymbol{\gamma}_0$ against $H_1 : \boldsymbol{\gamma} \neq \boldsymbol{\gamma}_0$, given by

$$F_N(\boldsymbol{\gamma}) = \frac{\hat{\sigma}_N^2(\boldsymbol{\gamma}) - \hat{\sigma}_N^2(\hat{\boldsymbol{\gamma}})}{\hat{\sigma}_N^2(\hat{\boldsymbol{\gamma}})/N},$$

where $\hat{\sigma}_N^2(\boldsymbol{\gamma}) = \widetilde{SSR}_{NT}(\boldsymbol{\gamma})$. The null hypothesis is rejected for large values of $F_N(\boldsymbol{\gamma}_0)$. Following Hansen (2017) and Zhang et al. (2017), a bootstrapping procedure is proposed to compute the confidence intervals for the parameters.

Algorithm A. Confidence intervals for parameters

Step 1. Use the original sample $(y_{it}, x_{it}, \boldsymbol{z}_{it}', \boldsymbol{q}_{it}')$'s to estimate model (2.1), and obtain the parameter estimates $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}})$ and the residual $\hat{\varepsilon}_{it} = \hat{u}_{it} - \bar{\hat{u}}_i$, in which $\hat{u}_{it} = y_{it} - \hat{\boldsymbol{\beta}}'\boldsymbol{x}_{it}(\hat{\boldsymbol{\gamma}})$ and $\bar{\hat{u}}_i = \frac{1}{T}\sum_{t=1}^{T} \hat{u}_{it}$.

Step 2. Generate i.i.d. draws u_{it}^* from the $N(0, 1)$ distribution for $i = 1, \ldots, N$ and $t = 1, \ldots, T$, and set $\varepsilon_{it}^* = \hat{\varepsilon}_{it}u_{it}^*$ and $y_{it}^* = \hat{\boldsymbol{\beta}}'\boldsymbol{x}_{it}(\hat{\boldsymbol{\gamma}}) + \bar{\hat{u}}_i + \varepsilon_{it}^*$.

Step 3. Use the observations $(y_{it}^*, x_{it}, \boldsymbol{z}_{it}', \boldsymbol{q}_{it}')$'s to estimate the KTR model with a covariate-dependent threshold, yielding the parameter estimates $(\hat{\boldsymbol{\beta}}^*, \hat{\boldsymbol{\gamma}}^*)$ and $\hat{\sigma}_N^{*2}(\hat{\boldsymbol{\gamma}}^*) = \frac{1}{N}(\ddot{\boldsymbol{Y}}^* - \ddot{\boldsymbol{X}}(\hat{\boldsymbol{\gamma}}^*)\hat{\boldsymbol{\beta}}^*(\hat{\boldsymbol{\gamma}}^*))'(\ddot{\boldsymbol{Y}}^* - \ddot{\boldsymbol{X}}(\hat{\boldsymbol{\gamma}}^*)\hat{\boldsymbol{\beta}}^*(\hat{\boldsymbol{\gamma}}^*))$, in which $\ddot{\boldsymbol{Y}}^*$ denotes the data of y_{it}^* stacked over all individuals.

Step 4. Compute the F-type statistic $F_N^*(\hat{\boldsymbol{\gamma}}) = \frac{\hat{\sigma}_N^{*2}(\hat{\boldsymbol{\gamma}}) - \hat{\sigma}_N^{*2}(\hat{\boldsymbol{\gamma}}^*)}{\hat{\sigma}_N^{*2}(\hat{\boldsymbol{\gamma}}^*)/N}$, in which $\hat{\sigma}_N^{*2}(\hat{\boldsymbol{\gamma}}) = \frac{1}{N}(\ddot{\boldsymbol{Y}}^* - \ddot{\boldsymbol{X}}(\hat{\boldsymbol{\gamma}})\hat{\boldsymbol{\beta}}^*(\hat{\boldsymbol{\gamma}}))'(\ddot{\boldsymbol{Y}}^* - \ddot{\boldsymbol{X}}(\hat{\boldsymbol{\gamma}})\hat{\boldsymbol{\beta}}^*(\hat{\boldsymbol{\gamma}}))$.

Step 5. Repeat Steps 1–4 B times, and obtain a sample of simulated coefficient estimates and the test statistic $\{\hat{\boldsymbol{\beta}}^*(b), \hat{\boldsymbol{\gamma}}^*(b), F^*_{N,b}(\hat{\boldsymbol{\gamma}})\}_{b=1}^B$.

Step 6. Create $1-\alpha$ bootstrap confidence intervals for the slope parameters $\boldsymbol{\beta} = [\beta_1^-, \beta_1^+, \boldsymbol{\beta}_2']'$ by the symmetric percentile method: the estimates plus and minus the $(1-\alpha)$ quantile of the absolute centered bootstrap estimates. For example, for β_1^-, the confidence interval is $\hat{\beta}_1^- \pm q^*_{1-\alpha}$, where $q^*_{1-\alpha}$ is the $1-\alpha$ quantile of $|\hat{\beta}_1^{-*} - \hat{\beta}_1^-|$.

Step 7. If necessary, compute the bootstrap standard errors for the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ by using the sample of simulated coefficient estimates. For example, for β_1^-, the bootstrap standard error is $\hat{s}e^*_{\beta_1^-} = \sqrt{\frac{1}{B}\sum_{b=1}^B(\hat{\beta}_1^{-*}(b) - \bar{\hat{\beta}}_1^{-*})^2}$, where $\bar{\hat{\beta}}_1^{-*} = \frac{1}{B}\sum_{b=1}^B \hat{\beta}_1^{-*}(b)$.

Step 8. Calculate the $1-\alpha$ quantile $c^*_{1-\alpha}$ of the simulated F statistics $\{F^*_{N,b}(\hat{\boldsymbol{\gamma}})\}_{b=1}^B$. Then, one can create a $1-\alpha$ bootstrap confidence interval for $\boldsymbol{\gamma}$ as $C^*_{\boldsymbol{\gamma}} = \{\boldsymbol{\gamma} : F_N(\boldsymbol{\gamma}) \leq c^*_{1-\alpha}\}$.

Following Hansen (2017), we next consider inference on the KTR function in the proposed PKTR-CDT model:

$$g(\boldsymbol{\theta}|\boldsymbol{w}) = \boldsymbol{\beta}'\boldsymbol{x}(\boldsymbol{\gamma}) + \alpha, \tag{2.11}$$

where $\boldsymbol{w} = (x, \boldsymbol{z}', \boldsymbol{q}', \alpha)'$, $\boldsymbol{x}(\boldsymbol{\gamma}) = [(x-\gamma(\boldsymbol{\gamma}))_-, (x-\gamma(\boldsymbol{\gamma}))_+, \boldsymbol{z}']'$, and $\gamma(\boldsymbol{\gamma}) = \gamma_0 + \boldsymbol{\gamma}_1'\boldsymbol{q}$, and α denotes the unobserved individual heterogeneity stacked over all individuals. For clarity, the KTR function is conditional on $\boldsymbol{w}$, which is slightly different from Hansen (2017); however, as in Hansen (2017), our theory will also take $\boldsymbol{w}$ as fixed.

In the PKTR-CDT model, the KTR function is not differentiable at $x = \gamma_0 + \boldsymbol{\gamma}_1'\boldsymbol{q}$; i.e., the regression function is continuous but not differentiable. As discussed in the literature (e.g., Hirano and Porter, 2012; Woutersen and Ham, 2013; Fang and Santos, 2014; Fang, 2014; Hong and Li, 2015), the nondifferentiability implies that, even though Theorem 2.1 shows that $\hat{\boldsymbol{\theta}}$ is asymptotically normal, $g(\hat{\boldsymbol{\theta}}|\boldsymbol{w})$ will not be asymptotically normal at $x = \gamma_{00} + \boldsymbol{\gamma}_{01}'\boldsymbol{q}$, where $\boldsymbol{\gamma}_0 = (\gamma_{00}, \boldsymbol{\gamma}_{01}')'$ is the true value, and asymptotic normality is likely to be a poor approximation for $\boldsymbol{\gamma}_0$ close to $x = \gamma_0 + \boldsymbol{\gamma}_1'\boldsymbol{q}$. As can be seen, when $\boldsymbol{q} = \boldsymbol{0}$, this problem degenerates to the case in Section 6 of Hansen (2017).

We then extend the analysis in Hansen (2017, p. 235) to the PKTR-CDT model. As in Hansen (2017), although the KTR function is not differentiable at $x = \gamma_0 + \boldsymbol{\gamma}_1'\boldsymbol{q}$, it is directionally differentiable at all points, implying that both the left and right derivatives are well defined. The directional derivative of a function $\phi(\boldsymbol{\theta}) : \mathbb{R}^{l+k+3} \to \mathbb{R}$ in the direction $\boldsymbol{h} \in \mathbb{R}^{l+k+3}$ is

$$\phi_{\boldsymbol{\theta}}(\boldsymbol{h}) = \lim_{\varepsilon \to 0^+} \frac{\phi(\boldsymbol{\theta} + \boldsymbol{h}\varepsilon) - \phi(\boldsymbol{\theta})}{\varepsilon}. \tag{2.12}$$

Thus, we can calculate the directional derivative of $g(\boldsymbol{\theta}|\boldsymbol{w})$ in the direction $\boldsymbol{h} = (\boldsymbol{h}_{\boldsymbol{\beta}}', \boldsymbol{h}_{\boldsymbol{\gamma}}')'$ in which $\boldsymbol{h}_{\boldsymbol{\gamma}} = (h_{\gamma_0}, \boldsymbol{h}_{\boldsymbol{\gamma}_1}')'$ and $\boldsymbol{h}_{\boldsymbol{\gamma}_1} = (h_{\gamma_{11}}, \ldots, h_{\gamma_{1k}})'$:

$$g_{\boldsymbol{\theta}}(\boldsymbol{h}|\boldsymbol{w}) = \boldsymbol{x}(\boldsymbol{\gamma})'\boldsymbol{h}_{\boldsymbol{\beta}} + g_{\boldsymbol{\gamma}}(\boldsymbol{h}_{\boldsymbol{\gamma}}|\boldsymbol{w}), \tag{2.13}$$

where

$$g_{\boldsymbol{\gamma}}(\boldsymbol{h}_{\boldsymbol{\gamma}}|\boldsymbol{w}) = \begin{cases} -\beta_1^-\left(h_{\gamma_0} + \boldsymbol{q}'\boldsymbol{h}_{\boldsymbol{\gamma}_1}\right) & , \text{ if } x < \gamma_0 + \boldsymbol{\gamma}_1'\boldsymbol{q} \\ -\beta_1^-\left(h^+_{\gamma_0} + \sum_{i=1}^k q_i h^+_{\gamma_{1i}}\right) - \beta_1^+\left(h^-_{\gamma_0} + \sum_{i=1}^k q_i h^-_{\gamma_{1i}}\right) & , \text{ if } x = \gamma_0 + \boldsymbol{\gamma}_1'\boldsymbol{q} \\ -\beta_1^+\left(h_{\gamma_0} + \boldsymbol{q}'\boldsymbol{h}_{\boldsymbol{\gamma}_1}\right) & , \text{ if } x > \gamma_0 + \boldsymbol{\gamma}_1'\boldsymbol{q} \end{cases},$$

in which $h^-_{\gamma_i} = h_{\gamma_i}1(h_{\gamma_i} \leq 0)$ and $h^+_{\gamma_i} = h_{\gamma_i}1(h_{\gamma_i} > 0)$ for $i = 0, 11, \ldots, 1k$.

Because (2.13) is continuous in $\boldsymbol{h}$, on the basis of Theorem 2.1, we can immediately obtain the following asymptotic result following Hansen (2017) and Fang and Santos (2014, Theorem 2.1).

COROLLARY 2.1.

$$\sqrt{N}\left(g(\hat{\boldsymbol{\theta}}|\boldsymbol{w}) - g(\boldsymbol{\theta}_0|\boldsymbol{w})\right) \xrightarrow{d} g_{\boldsymbol{\theta}_0}(\boldsymbol{Z}|\boldsymbol{w}), \tag{2.14}$$

where $\boldsymbol{Z} \sim N(\boldsymbol{0}, \boldsymbol{V})$, *and* $g_{\boldsymbol{\theta}}(\boldsymbol{h}|\boldsymbol{w})$ *is defined in (2.13).* □

Note that Corollary 2.1, which is based on Theorem 2.1 and Fang and Santos (2014, Theorem 2.1), is a generalised version of Hansen (2017), and when $\boldsymbol{q} = \boldsymbol{0}$, our result degenerates to Hansen (2017). For $x \neq \gamma_{00} + \boldsymbol{\gamma}_{01}'\boldsymbol{q}$, we have a normal asymptotic distribution, as shown in Corollary 2.1; however, at $x = \gamma_{00} + \boldsymbol{\gamma}_{01}'\boldsymbol{q}$, the asymptotic distribution is a nonlinear transformation of a normal random vector and will be biased, with a bias depending on the covariate $\boldsymbol{q}$ and the relative magnitudes of β_1^- and β_1^+. For example, if $\beta_1^- = 1$, $\beta_1^- = -1$, and $\boldsymbol{q} = (1, \ldots, 1)'$, then $g_{\boldsymbol{\gamma}}(\boldsymbol{h}_{\boldsymbol{\gamma}}|\boldsymbol{w}) = -|h_{\gamma_0}| - \sum_{i=1}^{k} |h_{\gamma_{1i}}|$, leading to the asymptotic distribution in Corollary 2.1 having a negative mean.

In our PKTR-CDT model, as in Hansen (2017), we face the same problem in constructing the confidence intervals of the KTR function because of the non-normality in Corollary 2.1. According to Corollary 3.1 of Fang and Santos (2014), Corollary 2.1 implies that the conventional bootstrap is inconsistent, as illustrated by Hansen (2017). Therefore, Fang and Santos (2014) suggested that one can approximate the distribution of $\sqrt{N}(g(\hat{\boldsymbol{\theta}}|\boldsymbol{w}) - g(\boldsymbol{\theta}_0|\boldsymbol{w}))$ by that of $\hat{g}_{\boldsymbol{\theta}}(\sqrt{N}(\hat{\boldsymbol{\theta}}^* - \hat{\boldsymbol{\theta}})|\boldsymbol{w})$, where $\hat{\boldsymbol{\theta}}^*$ is the bootstrap distribution of $\hat{\boldsymbol{\theta}}$ and $\hat{g}_{\boldsymbol{\theta}}(\boldsymbol{h}|\boldsymbol{w})$ is an estimate of $g_{\boldsymbol{\theta}}(\boldsymbol{h}|\boldsymbol{w})$. When $g(\boldsymbol{\theta}|\boldsymbol{w})$ is Lipschitz continuous, Hong and Li (2015) suggested that $g_{\boldsymbol{\theta}}(\boldsymbol{h}|\boldsymbol{w})$ can be estimated as

$$\hat{g}_{\boldsymbol{\theta}}(\boldsymbol{h}|\boldsymbol{w}) = \boldsymbol{x}(\hat{\boldsymbol{\gamma}})'\boldsymbol{h}_{\boldsymbol{\beta}} + \frac{g(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}} + \boldsymbol{h}_{\boldsymbol{\gamma}}\varepsilon_N|\boldsymbol{w}) - g(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}|\boldsymbol{w})}{\sqrt{N}\varepsilon_N}, \tag{2.15}$$

in which we assume that, for some sequence $\varepsilon_N > 0$, we have $\varepsilon_N \to 0$ and $\sqrt{N}\varepsilon_N \to \infty$. The above method is known as "the numerical delta method."

2.2. Testing for the kink threshold effect and threshold constancy

One of the reasonable and important questions is to test whether the panel KTR model is significantly different from the linear panel model $y_{it} = \beta_1 x_{it} + \boldsymbol{\beta}_2'\boldsymbol{z}_{it} + \alpha_i + \varepsilon_{it}$, which is nested in (2.1) when $\boldsymbol{z}_{it}$ includes $\boldsymbol{q}_{it}$. Thus, we consider the null hypothesis of a no-kink threshold effect $H_0^1 : \beta_1^- = \beta_1^+$ against the alternative hypothesis $H_1^1 : \beta_1^- \neq \beta_1^+$. Denote the usual fixed-effect estimator of the linear model as $(\tilde{\beta}_1, \tilde{\boldsymbol{\beta}}_2')$, and obtain the residual $\tilde{\varepsilon}_{it} = \tilde{u}_{it} - \bar{\tilde{u}}_i$, in which $\tilde{u}_{it} = y_{it} - \tilde{\beta}_1 x_{it} - \tilde{\boldsymbol{\beta}}_2'\boldsymbol{z}_{it}$ and $\bar{\tilde{u}}_i = \frac{1}{T}\sum_{t=1}^{T} \tilde{u}_{it}$. Then, the error variance estimate in the linear model is $\tilde{\sigma}_N^2 = \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} \tilde{\varepsilon}_{it}^2$. On the other hand, denote the error variance estimate of the proposed covariate-dependent threshold model as $\hat{\sigma}_N^2(\hat{\boldsymbol{\gamma}}) = \frac{1}{N}(\ddot{\boldsymbol{Y}} - \ddot{\boldsymbol{X}}(\hat{\boldsymbol{\gamma}})\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\gamma}}))'(\ddot{\boldsymbol{Y}} - \ddot{\boldsymbol{X}}(\hat{\boldsymbol{\gamma}})\hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\gamma}}))$. Likewise, we denote the error variance estimate of Zhang et al.'s (2017) constant threshold model as $\hat{\sigma}_C^2(\hat{\gamma}_0)$.[6] Then, a standard test for the null hypothesis of linearity can be given

[6] In the panel KTR model with a constant threshold, to test the null hypothesis of no kink effect, one can directly extend Hansen's (2017) test statistic to the panel framework by constructing the following test statistic: $W_1^C = \frac{\tilde{\sigma}_N^2 - \hat{\sigma}_N^2(\hat{\gamma}_0)}{\hat{\sigma}_N^2(\hat{\gamma}_0)/N}$.

by

$$W_1 = \frac{\tilde{\sigma}_N^2 - \hat{\sigma}_N^2(\hat{\boldsymbol{\gamma}})}{\hat{\sigma}_N^2(\hat{\boldsymbol{\gamma}})/N} = \sup_{\boldsymbol{\gamma}\in\boldsymbol{\Gamma}} \frac{\tilde{\sigma}_N^2 - \hat{\sigma}_N^2(\boldsymbol{\gamma})}{\hat{\sigma}_N^2(\boldsymbol{\gamma})/N}. \tag{2.16}$$

Another important problem is to determine whether the threshold is constant. To this end, consider the following null hypothesis: $H_0^2 : \boldsymbol{\gamma}_1 = 0$. Then, a natural test for the null hypothesis of the constant threshold against the covariate-dependent threshold model (2.1) can be given by

$$W_2 = \frac{\hat{\sigma}_C^2(\hat{\gamma}_0) - \hat{\sigma}_N^2(\hat{\boldsymbol{\gamma}})}{\hat{\sigma}_N^2(\hat{\boldsymbol{\gamma}})/N}. \tag{2.17}$$

THEOREM 2.2. *Suppose Assumptions 1–5 hold. As $N \to \infty$, under H_0^1 we have*

$$W_1 \xrightarrow{d} \sup_{\boldsymbol{\gamma}\in\boldsymbol{\Gamma}} \frac{\boldsymbol{G}'(\boldsymbol{\gamma})\boldsymbol{Q}^{-1}(\boldsymbol{\gamma})\boldsymbol{R}_1\left[\boldsymbol{R}_1'\boldsymbol{Q}^{-1}(\boldsymbol{\gamma})\boldsymbol{R}_1\right]^{-1}\boldsymbol{R}_1'\boldsymbol{Q}^{-1}(\boldsymbol{\gamma})\boldsymbol{G}(\boldsymbol{\gamma})}{\sigma_T^2}, \tag{2.18}$$

and under H_0^2, we have

$$W_2 \xrightarrow{d} \frac{\boldsymbol{Z}'\boldsymbol{G}^{-1}\boldsymbol{R}_2'[\boldsymbol{R}_2\boldsymbol{G}^{-1}\boldsymbol{R}_2']^{-1}\boldsymbol{R}_2\boldsymbol{G}^{-1}\boldsymbol{Z}}{\sigma_T^2}, \tag{2.19}$$

where $\boldsymbol{Q}^{-1}(\boldsymbol{\gamma}) = (\sum_{t=1}^T E[\ddot{\boldsymbol{x}}_{it}(\boldsymbol{\gamma})\ddot{\boldsymbol{x}}_{it}'(\boldsymbol{\gamma})])^{-1}$, $\sigma_T^2 = \sum_{t=1}^T E(\ddot{\varepsilon}_{it}^2)$, $\boldsymbol{R}_1 = [1, -1, \boldsymbol{0}_{1\times l}]'$, $\boldsymbol{Z} \sim \boldsymbol{\mathcal{N}}(\boldsymbol{0}, \boldsymbol{S})$, $\boldsymbol{S} = var(\boldsymbol{h}_i'\boldsymbol{e}_i)$, $\boldsymbol{G} = E(\boldsymbol{h}_i'\boldsymbol{h}_i) + \sum_{t=1}^T E(\boldsymbol{D}_{it}\boldsymbol{e}_{it})$, and $\boldsymbol{R}_2 = [\boldsymbol{0}_{k\times(l+2)}, \mathbf{I}_k]$, and $\boldsymbol{G}(\boldsymbol{\gamma})$ is a zero-mean Gaussian process with covariance kernel

$$\boldsymbol{K}(\boldsymbol{\gamma}_{(1)}, \boldsymbol{\gamma}_{(2)}) = E\left(\boldsymbol{G}(\boldsymbol{\gamma}_{(1)})\boldsymbol{G}'(\boldsymbol{\gamma}_{(2)})\right) = \sum_{t=1}^T E\left(\ddot{\boldsymbol{x}}_{it}(\boldsymbol{\gamma}_{(1)})\ddot{\boldsymbol{x}}_{it}'(\boldsymbol{\gamma}_{(2)})\ddot{\varepsilon}_{it}^2\right).$$

Proof. See the Appendix. □

As discussed in Hansen (1996, 2017) and Zhang et al. (2017), the limiting distribution of the test statistic W_1 is the supremum of a quadratic form of the Gaussian process, and hence it is generally not straightforward to tabulate the critical values. In practice, we can follow Hansen (2017) and Zhang et al. (2017) to use a parametric bootstrap procedure to calculate the p values or critical values.

Algorithm B. Testing for kink effect and threshold constancy

Step 1. Use the original sample $(y_{it}, x_{it}, \boldsymbol{z}_{it}', \boldsymbol{q}_{it}')$'s to estimate the linear model $y_{it} = \beta_1 x_{it} + \boldsymbol{\beta}_2'z_{it} + \alpha_i + \varepsilon_{it}^l$ and the constant threshold model $y_{it} = \beta_1^-(x_{it} - \gamma)_- + \beta_1^+(x_{it} - \gamma)_+ + \boldsymbol{\beta}_2'z_{it} + \alpha_i + \varepsilon_{it}^c$, and obtain the residual $\hat{\varepsilon}_{it}^l = \hat{u}_{it}^l - \bar{\hat{u}}_i^l$ and $\hat{\varepsilon}_{it}^c = \hat{u}_{it}^c - \bar{\hat{u}}_i^c$, in which $\hat{u}_{it}^l = y_{it} - \hat{\beta}_1 x_{it} - \hat{\boldsymbol{\beta}}_2'z_{it}$, $\bar{\hat{u}}_i^l = \frac{1}{T}\sum_{t=1}^T \hat{u}_{it}^l$, and $\hat{u}_{it}^c = y_{it} - \hat{\beta}_1^-(x_{it} - \gamma)_- - \hat{\beta}_1^+(x_{it} - \gamma)_+ - \hat{\boldsymbol{\beta}}_2'z_{it}$, $\bar{\hat{u}}_i^c = \frac{1}{T}\sum_{t=1}^T \hat{u}_{it}^c$.

Step 2. Generate i.i.d. draws u_{it}^* from the $N(0, 1)$ distribution for $i = 1, \ldots, N$ and $t = 1, \ldots, T$, and set $\varepsilon_{it}^{l*} = \hat{\varepsilon}_{it}^l u_{it}^*$ and $y_{it}^{l*} = \hat{\beta}_1 x_{it} + \hat{\boldsymbol{\beta}}_2'z_{it} + \bar{\hat{u}}_i^l + \varepsilon_{it}^{l*}$. Set $\varepsilon_{it}^{c*} = \hat{\varepsilon}_{it}^c u_{it}^*$ and $y_{it}^{c*} = \hat{\beta}_1^-(x_{it} - \gamma)_- + \hat{\beta}_1^+(x_{it} - \gamma)_+ + \hat{\boldsymbol{\beta}}_2'z_{it} + \bar{\hat{u}}_i^c + \varepsilon_{it}^{c*}$.

Step 3. Use the observations $(y_{it}^{l*}, x_{it}, \boldsymbol{z}_{it}', \boldsymbol{q}_{it}')$'s and $(y_{it}^{c*}, x_{it}, \boldsymbol{z}_{it}', \boldsymbol{q}_{it}')$'s to estimate the linear model and the covariate-dependent threshold model (2.1), and compute the F-type statistics W_1 and W_2.

Step 4. Repeat Steps 1–3 B times so as to obtain two samples $W_1^*(1), W_1^*(2), \ldots, W_1^*(B)$ and $W_2^*(1), W_2^*(2), \ldots, W_2^*(B)$ of simulated W_1 and W_2 statistics.

Step 5. The empirical p values can be obtained by calculating the percentage of the simulated statistics that exceed actual value when the number of B is sufficiently large.

3. MONTE CARLO SIMULATIONS

In this section, we conduct Monte Carlo simulations to examine the finite-sample performances of the proposed estimation procedure and the test statistics for kink effect and threshold constancy. We adopt the following three data-generating processes (DGPs):

$$\begin{aligned}
\text{DGP } 1: &\quad y_{it} = 0.4(x_{it} - \gamma_{it})_- + 0.8(x_{it} - \gamma_{it})_+ + 2z_{it} + \alpha_i + u_{it} \\
\text{DGP } 2: &\quad y_{it} = 0.4(x_{it} - \gamma)_- + 0.8(x_{it} - \gamma)_+ + 2z_{it} + \alpha_i + u_{it} \\
\text{DGP } 3: &\quad y_{it} = 0.4x_{it} + 2z_{it} + \alpha_i + u_{it},
\end{aligned}$$

where $x_{it} = 0.25\alpha_i + u_{q,it} + u_{x,it}$, $z_{it} = 0.5\alpha_i + u_{q,it} + u_{z,it}$, and $\gamma_{it} = 0 + 0.5q_{it}$, in which $q_{it} = z_{it}$, $u_{q,it} \sim i.i.d.N(0.5, 1)$, and $\alpha_i \sim i.i.d.N(0, 1)$. The constant threshold γ is set as zero. The innovation processes u_{it}, $u_{x,it}$, and $u_{z,it}$ are independent of each other. u_{it} follows $i.i.d.N(0, 0.5^2)$. $u_{x,it}$ and $u_{z,it}$ follow $i.i.d.N(0, 1)$. The number of replications is 1,000.

We first evaluate the finite-sample properties of the proposed estimator, comparing with the constant threshold model proposed by Zhang et al. (2017). To this end, we focus on DGP 1 and DGP 2. Note that DGP 1 is accompanied by a covariate-dependent threshold, and hence using the constant threshold model is inappropriate and might lead to biased estimators; however, DGP 2 contains a constant threshold, and thus using the proposed covariate-dependent threshold model is not necessary but should be harmless.

Table 1 presents the summary statistics (i.e., mean and standard deviation) for the estimates based on the constant threshold model. From Table 1, it can be seen that the constant threshold model proposed by Zhang et al. (2017) works reasonably well when DGP 2 is used (i.e., the DGP does not contain a covariate-dependent threshold). However, the results show that if the covariate-dependent feature in the threshold is ignored, the estimates might be seriously biased and have large standard errors, caused by the misclassification of observations because of threshold misspecification.

From the summary statistics reported in Table 2, it can be seen that the proposed estimator seems to be unbiased, and the accuracy of the model improves as either N or T increases. When N changes from 50 to 200, the standard deviation decreases by almost half, which is consistent with the $\sqrt{N}$ convergence rate implied by Theorem 2.1. In sum, the proposed estimator has good properties in finite samples, even if the true DGP is without such covariate-dependent features in the threshold.

Table 3 reports the empirical sizes and powers of the test statistics. The empirical sizes of W_1 and W_2 are close to the nominal size 5%. When the sample size is small (i.e., $N \times T = 500$), the power of the test W_1 is low. However, the powers of the test W_1 are reasonably high when $N \times T \geq 1000$. Moreover, the empirical powers of W_2 are very high for all cases considered. In sum, the proposed tests perform well in finite samples.

Table 1. Estimates of the parameters on the basis of Zhang et al. (2017)'s constant threshold model.

DGP 1		$\gamma_0 = 0$		$\beta_1^- = 0.4$		$\beta_1^+ = 0.8$		$\beta_2 = 2$	
T	N	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
10	50	0.288	0.648	0.482	0.067	0.750	0.055	1.683	0.021
	100	0.209	0.496	0.484	0.052	0.739	0.038	1.684	0.014
	200	0.220	0.358	0.491	0.035	0.735	0.025	1.684	0.010
20	50	0.233	0.469	0.487	0.048	0.740	0.035	1.684	0.014
	100	0.229	0.357	0.492	0.035	0.736	0.025	1.684	0.010
	200	0.235	0.238	0.496	0.023	0.734	0.017	1.684	0.007
50	50	0.231	0.306	0.494	0.030	0.735	0.022	1.684	0.009
	100	0.230	0.229	0.495	0.021	0.734	0.016	1.684	0.006
	200	0.239	0.160	0.497	0.015	0.733	0.011	1.684	0.004
DGP 2		$\gamma_0 = 0$		$\beta_1^- = 0.4$		$\beta_1^+ = 0.8$		$\beta_2 = 2$	
T	N	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
10	50	0.009	0.266	0.393	0.057	0.805	0.037	2.000	0.019
	100	−0.007	0.172	0.394	0.039	0.802	0.024	2.000	0.013
	200	−0.001	0.117	0.398	0.027	0.801	0.017	2.000	0.010
20	50	0.000	0.168	0.396	0.038	0.802	0.024	2.000	0.013
	100	−0.001	0.112	0.398	0.026	0.801	0.017	2.000	0.010
	200	0.004	0.077	0.400	0.019	0.801	0.011	2.000	0.007
50	50	0.001	0.097	0.400	0.023	0.801	0.014	2.000	0.008
	100	0.004	0.073	0.400	0.016	0.801	0.010	2.000	0.006
	200	0.003	0.052	0.400	0.012	0.800	0.007	2.000	0.004

4. CONCLUSION

This article proposes a panel KTR model with a covariate-dependent threshold (PKTR-CDT). We suggest an estimator based on the within-group transformation, following Hansen (1999), and we construct test statistics for kink effect and threshold constancy. The asymptotic joint normality of the slope and threshold estimators is established, and the limiting distributions of the test statistics are derived. An interesting finding is that the inclusion of the covariates $\boldsymbol{q}_{it}$ in the covariate-dependent threshold does not affect the asymptotic joint normality of the slope and threshold estimates in the KTR models. Monte Carlo simulations show that the finite-sample proprieties of the estimator and test statistics are generally satisfactory.

ACKNOWLEDGEMENTS

We thank Petra Todd (editor) and two anonymous referees for constructive comments and suggestions. Any remaining errors are solely our responsibility. Lixiong Yang acknowledges the financial support of the National Natural Science Foundation of China (Grant No. 71803072).

Table 2. Estimates of the parameters obtained by using the estimator proposed in Section 2.

DGP1		$\gamma_0 = 0$		$\gamma_1 = 0.5$		$\beta_1^- = 0.4$		$\beta_1^+ = 0.8$		$\beta_2 = 2$	
T	N	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
10	50	−0.004	0.294	0.489	0.148	0.384	0.073	0.807	0.050	1.993	0.098
	100	0.000	0.189	0.494	0.101	0.396	0.048	0.803	0.034	1.996	0.067
	200	0.003	0.130	0.497	0.067	0.399	0.035	0.802	0.024	1.997	0.046
20	50	0.004	0.172	0.489	0.099	0.396	0.046	0.803	0.034	1.993	0.065
	100	0.004	0.126	0.498	0.068	0.400	0.034	0.802	0.023	1.998	0.046
	200	0.000	0.084	0.499	0.047	0.399	0.022	0.800	0.017	1.999	0.031
50	50	0.000	0.111	0.496	0.062	0.398	0.029	0.800	0.020	1.997	0.041
	100	0.003	0.074	0.497	0.044	0.399	0.019	0.801	0.014	1.998	0.030
	200	0.002	0.053	0.498	0.024	0.400	0.015	0.800	0.010	1.999	0.017
DGP2		$\gamma_0 = 0$		$\gamma_1 = 0$		$\beta_1^- = 0.4$		$\beta_1^+ = 0.8$		$\beta_2 = 2$	
T	N	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
10	50	−0.005	0.272	−0.016	0.191	0.388	0.066	0.803	0.039	1.990	0.129
	100	−0.003	0.185	−0.015	0.130	0.397	0.046	0.802	0.028	1.989	0.088
	200	−0.001	0.125	−0.008	0.092	0.399	0.032	0.801	0.020	1.994	0.063
20	50	0.005	0.176	−0.013	0.126	0.397	0.043	0.802	0.028	1.991	0.087
	100	−0.001	0.129	−0.008	0.087	0.400	0.032	0.801	0.020	1.994	0.059
	200	0.001	0.088	0.000	0.061	0.399	0.022	0.800	0.014	2.000	0.042
50	50	0.000	0.103	−0.004	0.080	0.399	0.027	0.800	0.016	1.997	0.055
	100	0.003	0.072	−0.003	0.057	0.400	0.017	0.801	0.012	1.998	0.039
	200	−0.001	0.054	−0.002	0.036	0.400	0.014	0.800	0.008	1.998	0.025

Table 3. Empirical sizes and powers of the test statistics.

		Test for kink effect (W_1)			Test for constancy (W_2)		
		Size (DGP 3)	Power (DGP 1)		Size (DGP 2)	Power (DGP 1)	
T	N	$\beta_1^+ = 0.4$	$\beta_1^+ = 0.6$	$\beta_1^+ = 0.8$	$\gamma_1 = 0.0$	$\gamma_1 = 0.3$	$\gamma_1 = 0.5$
10	50	0.053	0.759	1.000	0.051	0.770	0.978
	100	0.057	0.967	1.000	0.059	0.954	1.000
	200	0.065	1.000	1.000	0.048	1.000	1.000
20	50	0.053	0.980	1.000	0.056	0.970	1.000
	100	0.054	1.000	1.000	0.049	0.999	1.000
	200	0.050	1.000	1.000	0.056	1.000	1.000
50	50	0.056	1.000	1.000	0.048	1.000	1.000
	100	0.057	1.000	1.000	0.059	1.000	1.000
	200	0.042	1.000	1.000	0.057	1.000	1.000

REFERENCES

Akitoby, B., A. Binder and T. Komatsuzaki (2017). Inflation and public debt reversals in the G7 countries. *Journal of Banking and Financial Economics 1*, 5–27.

Card, D., D. Lee, Z. Pei and A. Weber (2015). Inference on causal effects in a generalized regression kink design. *Econometrica 83*, 2453–83.

Chan, K. S. (1993). Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. *Annals of Statistics 21*, 520–33.

Chan, K. S. and R. S. Tsay (1998). Limiting properties of the least squares estimator of a continuous threshold autoregressive model. *Biometrika 85*, 413–26.

Cochrane , J.H. (2011). Inflation and debt. *National affairs 9*, 56–78.

Dueker, M. J., Z. Psaradakis, M. Sola and F. Spagnolo (2013). State-dependent threshold smooth transition autoregressive models. *Oxford Bulletin of Economics and Statistics 75*, 835–54.

Fang, Z. (2014). Optimal plug-in estimators of directionally differentiable functionals. Working Paper, University of California, San Diego, California, USA.

Fang, Z. and A. Santos (2014). Inference on directionally differentiable functions. Working Paper, University of California, San Diego, California, USA.

Hansen, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica 64*, 413–30.

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

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

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

Hirano, K. and J. Porter (2012). Impossibility results for nondifferentiable functionals. *Econometrica 80*, 1769–90.

Hong, H. and J. Li (2015). The numerical directional delta method. Working Paper, Stanford University, Stanford, California, USA.

Krause, M. U. and S. Moyen (2016). Public debt and changing inflation targets. *American Economic Journal: Macroeconomics 8*, 142–76.

Kriwoluzky, A., .G Muller and M. Wolf (2019). Exit expectations and debt crises in currency unions. *Journal of International Economics 121*, 103253.

Lee, S., Y. Liao, M. H. Seo and Y. Shin (2018). Factor-driven two-regime regression. mimeo, arXiv preprint: arXiv:1810.11109.

Nielsen, H. S., T. Sorensen and C. Taber (2010). Estimating the effect of student aid on college enrollment: evidence from a government grant policy reform. *American Economic Journal: Economic Policy 2*, 185–215.

Reinhart, C. M. and K.S. Rogoff (2010). Growth in a Time of Debt. *American economic review 100*, 573–78.

Tong, H. (1990). *Non-Linear Time Series: A Dynamical System Approach*, pp. 96–120. Oxford: Oxford University Press.

Woutersen, T. and J. C. Ham (2013). Calculating confidence intervals for continuous and discontinuous functions of parameters. Working Paper, University of Arizona, Tucson, Arizona, USA.

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

Yu, P. and X. Fan (2020). Threshold regression with a threshold boundary. *Journal of Business and Economic Statistics*, DOI: 10.1080/07350015.2020.1740712.

Zhang, Y., Q. Zhou and L. Jiang (2017). Panel kink regression with an unknown threshold. *Economics Letters 157*, 116–21.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Online Supplement
Replication Package

Co-editor Petra Todd handled this manuscript.

APPENDIX: PROOFS OF THE MAIN RESULTS

This appendix provides the proofs of Theorems 2.1 and 2.2 in the paper. To this end, we first prove the following Lemma, which is used to prove Theorem 2.1.

LEMMA A.1. *Under Assumptions 1–2, we have*

$$E\left[1\left(\gamma_{1,it} \leq x_{it} \leq \gamma_{2,it}\right)\right] \leq \bar{f}_{\boldsymbol{q}} \left|\gamma_{02} - \gamma_{01}\right| + \boldsymbol{C} \bar{f}_{\boldsymbol{q}} \left\|\boldsymbol{\gamma}_{12} - \boldsymbol{\gamma}_{11}\right\|, \tag{A.1}$$

where $\gamma_{j,it} = \gamma_{0j} + \boldsymbol{\gamma}_{1j}'\boldsymbol{q}_{it}$ *for* $j = 1, 2$.

Proof of Lemma A.1. First, we consider the expectation conditional on $\boldsymbol{q}_{it}$,

$$E\left[1\left(\gamma_{1,it} \leq x_{it} \leq \gamma_{2,it}\right)\middle|\, \boldsymbol{q}_{it}\right] = F_{\boldsymbol{q}}(\gamma_{2,it}) - F_{\boldsymbol{q}}(\gamma_{1,it}), \tag{A.2}$$

where $F_{\boldsymbol{q}}$ is the conditional cumulative distribution function of x_{it} conditional on $\boldsymbol{q}_{it}$.

By the mean value theorem and Assumption 2, we have

$$\begin{aligned} F_{\boldsymbol{q}}(\gamma_{2,it}) - F_{\boldsymbol{q}}(\gamma_{1,it}) &\leq \bar{f}_{\boldsymbol{q}} \left|\gamma_{2,it} - \gamma_{1,it}\right| \\ &= \bar{f}_{\boldsymbol{q}} \left|\gamma_{02} + \boldsymbol{\gamma}_{12}'\boldsymbol{q}_{it} - \left(\gamma_{01} + \boldsymbol{\gamma}_{11}'\boldsymbol{q}_{it}\right)\right| \\ &\leq \bar{f}_{\boldsymbol{q}} \left|\gamma_{02} - \gamma_{01}\right| + \bar{f}_{\boldsymbol{q}} \left|\left(\boldsymbol{\gamma}_{12}' - \boldsymbol{\gamma}_{11}'\right)\boldsymbol{q}_{it}\right| \\ &\leq \bar{f}_{\boldsymbol{q}} \left|\gamma_{02} - \gamma_{01}\right| + \bar{f}_{\boldsymbol{q}} \left\|\boldsymbol{\gamma}_{12} - \boldsymbol{\gamma}_{11}\right\| \left\|\boldsymbol{q}_{it}\right\|. \end{aligned} \tag{A.3}$$

Taking the expectation over both sides of the inequality, we have

$$E[E\left[1\left(\gamma_{1,it} \leq x_{it} \leq \gamma_{2,it}\right)\middle|\, \boldsymbol{q}_{it}\right]] \leq \bar{f}_{\boldsymbol{q}} \left|\gamma_{02} - \gamma_{01}\right| + \bar{f}_{\boldsymbol{q}} \left\|\boldsymbol{\gamma}_{12} - \boldsymbol{\gamma}_{11}\right\| E\left\|\boldsymbol{q}_{it}\right\|. \tag{A.4}$$

By the law of iterated expectation and Assumption 1 ($E\|\boldsymbol{q}_{it}\| \leq \boldsymbol{C} < \infty$), we therefore have

$$E\left[1\left(\gamma_{1,it} \leq x_{it} \leq \gamma_{2,it}\right)\right] \leq \bar{f}_{\boldsymbol{q}} \left|\gamma_{02} - \gamma_{01}\right| + \boldsymbol{C} \bar{f}_{\boldsymbol{q}} \left\|\boldsymbol{\gamma}_{12} - \boldsymbol{\gamma}_{11}\right\|. \tag{A.5}$$

□

Proof of Theorem 2.1. Define $e_{it}(\boldsymbol{\theta}) = \ddot{y}_{it} - \boldsymbol{\beta}'\ddot{\boldsymbol{x}}_{it}(\boldsymbol{\gamma}) = \ddot{\varepsilon}_{it} + (\boldsymbol{\beta}'_0 - \boldsymbol{\beta}')\ddot{\boldsymbol{x}}_{it}(\boldsymbol{\gamma}) + \boldsymbol{\beta}_0'[\ddot{\boldsymbol{x}}_{it}(\boldsymbol{\gamma}_0) - \ddot{\boldsymbol{x}}_{it}(\boldsymbol{\gamma})]$. Then, we can write $\widetilde{SSR}_{NT}(\boldsymbol{\theta}) = \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} e_{it}^2(\boldsymbol{\theta})$.

The estimator $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}', \hat{\boldsymbol{\gamma}}')' = \arg\min_{\boldsymbol{\theta}\in\mathbf{B}\times\Gamma} \widetilde{SSR}_{NT}(\boldsymbol{\theta})$ solves the first-order condition—that is, $\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} h_{it}(\hat{\boldsymbol{\theta}})e_{it}(\hat{\boldsymbol{\theta}}) = 0$. The true value $\boldsymbol{\theta}_0 = (\boldsymbol{\beta}_0', \boldsymbol{\gamma}_0')'$ minimises $L(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \sum_{t=1}^{T} E[\ddot{y}_{it} - \boldsymbol{\beta}'\ddot{\boldsymbol{x}}_{it}(\boldsymbol{\gamma})]^2$, and hence we have $\sum_{t=1}^{T} E[h_{it}(\boldsymbol{\theta}_0)e_{it}(\boldsymbol{\theta}_0)] = 0$. □

Following Hansen (2017) and Zhang et al. (2017), we complete the proof of Theorem 2.1 by verifying the following conditions:

Condition 1. $\hat{\boldsymbol{\theta}} \to_p \boldsymbol{\theta}_0$.

Condition 2. $\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sum_{t=1}^{T} \boldsymbol{h}_{it} e_{it} \to_d N(\mathbf{0}, \boldsymbol{S})$.

Condition 3. $\boldsymbol{Q}(\boldsymbol{\theta}) = \sum_{t=1}^{T} E[\boldsymbol{h}_{it}(\boldsymbol{\theta})\boldsymbol{h}_{it}'(\boldsymbol{\theta})] + \sum_{t=1}^{T} E[(-\frac{\partial}{\partial \boldsymbol{\theta}'}\boldsymbol{h}_{it}(\boldsymbol{\theta}))e_{it}(\boldsymbol{\theta})]$ *is continuous in* $\boldsymbol{\theta}$, *and* $\boldsymbol{Q}(\boldsymbol{\theta}_0) = \boldsymbol{Q}$.

Condition 4. $\boldsymbol{\nu}(\boldsymbol{\theta}) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \sum_{t=1}^{T} [\boldsymbol{h}_{it}(\boldsymbol{\theta})e_{it}(\boldsymbol{\theta}) - E[\boldsymbol{h}_{it}(\boldsymbol{\theta})e_{it}(\boldsymbol{\theta})]]$ *is stochastic equicontinuous.*

We first verify Condition 1. For any given x_{it}, $\boldsymbol{x}_{it}(\boldsymbol{\gamma})$ is continuous in $\boldsymbol{\gamma}$, and hence $\ddot{\boldsymbol{x}}_{it}(\boldsymbol{\gamma}) = \boldsymbol{x}_{it}(\boldsymbol{\gamma}) - \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{x}_{it}(\boldsymbol{\gamma})$ is also continuous in $\boldsymbol{\gamma}$. Furthermore, $e_{it}(\boldsymbol{\theta})$ and $e_{it}^2(\boldsymbol{\theta})$ are continuous in $\boldsymbol{\theta}$. By the triangle inequality, we have the following bound:

$$\begin{aligned}
&\|\ddot{\boldsymbol{x}}_{it}(\boldsymbol{\gamma})\|^2 \le \|\boldsymbol{x}_{it}(\boldsymbol{\gamma})\|^2 + \tfrac{1}{T}\textstyle\sum_{s=1}^{T} \|\boldsymbol{x}_{is}(\boldsymbol{\gamma})\|^2 \\
&\le \|\boldsymbol{z}_{it}\|^2 + x_{it}^2 + \bar{C}_{\Gamma_0}^2 + \|\boldsymbol{q}_{it}\|^2 \bar{\boldsymbol{C}}_{\Gamma_1}^2 + \tfrac{1}{T}\textstyle\sum_{s=1}^{T} \{\|\boldsymbol{z}_{is}\|^2 + x_{is}^2 + \bar{C}_{\Gamma_0}^2 + \|\boldsymbol{q}_{is}\|^2 \bar{\boldsymbol{C}}_{\Gamma_1}^2\},
\end{aligned}$$

in which $\bar{C}_{\Gamma_0} = \sup\{|\gamma_0| : \gamma_0 \in \Gamma_0\}$ and $\bar{\boldsymbol{C}}_{\Gamma_1} = \sup\{||\boldsymbol{\gamma}_1|| : \boldsymbol{\gamma}_1 \in \Gamma_1\}$. Furthermore, we have $e_{it}^2(\boldsymbol{\theta}) = (\ddot{y}_{it} - \boldsymbol{\beta}'\ddot{\boldsymbol{x}}_{it}(\boldsymbol{\gamma}))^2 \le 2\ddot{y}_{it}^2 + 2\bar{\boldsymbol{\beta}}^2\|\ddot{\boldsymbol{x}}_{it}(\boldsymbol{\gamma})\|^2$, where $\bar{\boldsymbol{\beta}} = \sup\{||\boldsymbol{\beta}|| : \boldsymbol{\beta} \in \mathbf{B}\}$. Thus, under Assumption 1, we have $E[e_{it}^2(\boldsymbol{\theta})] < \infty$. Then, by Lemma 2.4 of Newey and McFadden (1994), $E[e_{it}^2(\boldsymbol{\theta})]$ is continuous in $\boldsymbol{\theta}$ and

$$\sup_{\boldsymbol{\theta}\in\mathbf{B}\times\boldsymbol{\Gamma}} \left| \frac{1}{N}\sum_{i=1}^{N}\left\{\sum_{t=1}^{T} e_{it}^2(\boldsymbol{\theta}) - \sum_{t=1}^{T} E[e_{it}^2(\boldsymbol{\theta})]\right\}\right| = o_p(1).$$

Given the compactness of $\mathbf{B} \times \boldsymbol{\Gamma}$ and the uniqueness of the minimum true value $\boldsymbol{\theta}_0$ by assumption, Theorem 2.1 of Newey and MaFadden (1994) established Condition 1: $\hat{\boldsymbol{\theta}} \to_p \boldsymbol{\theta}_0$.

Condition 2 follows by the standard central limit theorem (Assumption 1).

We next establish Condition 3. From the expression of $\boldsymbol{Q}(\boldsymbol{\theta}) = \sum_{t=1}^{T} E[\boldsymbol{h}_{it}(\boldsymbol{\theta})\boldsymbol{h}_{it}'(\boldsymbol{\theta})] + \sum_{t=1}^{T} E[(-\frac{\partial}{\partial \boldsymbol{\theta}'}\boldsymbol{h}_{it}(\boldsymbol{\theta}))e_{it}(\boldsymbol{\theta})]$, we note that the elements of the matrix $E[\boldsymbol{h}_{it}(\boldsymbol{\theta})\boldsymbol{h}_{it}'(\boldsymbol{\theta})]$ are quadratic functions of $\boldsymbol{\beta}$, and $e_{it}(\boldsymbol{\theta})$ is continuous in $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma})$. Thus, $\boldsymbol{Q}(\boldsymbol{\theta})$ is continuous in $\boldsymbol{\beta}$. Next, we observe that $\boldsymbol{\gamma}$ enters $\boldsymbol{Q}(\boldsymbol{\theta})$ through one of the following forms (or its transpose): $E[\boldsymbol{x}_{it}(\boldsymbol{\gamma})\boldsymbol{x}_{is}(\boldsymbol{\gamma})]$, $E[\boldsymbol{x}_{it}(\boldsymbol{\gamma})\boldsymbol{z}_{is}]$, $E[1_{it}(\boldsymbol{\gamma})1_{is}(\boldsymbol{\gamma})]$, $E[\boldsymbol{w}_{it}1_{is}(\boldsymbol{\gamma})]$, and $E[\boldsymbol{q}_{it}'\boldsymbol{w}_{it}1_{is}(\boldsymbol{\gamma})]$. By Assumption 1, there is a $\boldsymbol{C}$ satisfying $(E||\boldsymbol{w}_{it}||^{2r})^{1/r} \le \boldsymbol{C} < \infty$. Thus, by Lemma A.1 and Holder's inequality, we obtain

$$\begin{aligned}
&E\left\|\boldsymbol{w}_{it}1(\gamma_{1,it} \le x_{it} \le \gamma_{2,it})\right\|^2 \le (E||\boldsymbol{w}_{it}||^{2r})^{1/r}(E|1(\gamma_{1,it} \le x_{it} \le \gamma_{2,it})|^{\tau})^{1/\tau} \\
&\le \boldsymbol{C}(\bar{f}_{\boldsymbol{q}}\,|\gamma_{02} - \gamma_{01}| + \boldsymbol{C}\bar{f}_{\boldsymbol{q}}\,\|\boldsymbol{\gamma}_{12} - \boldsymbol{\gamma}_{11}\|)^{1/\tau},
\end{aligned}$$

where $\tau = \frac{r}{r-1}$. Therefore, $E[\boldsymbol{w}_{it}1_{is}(\boldsymbol{\gamma})]$ is continuous in $\boldsymbol{\gamma}$. Similarly, we can show that $E[\boldsymbol{x}_{it}(\boldsymbol{\gamma})\boldsymbol{x}_{is}(\boldsymbol{\gamma})]$, $E[\boldsymbol{x}_{it}(\boldsymbol{\gamma})\boldsymbol{z}_{is}]$, $E[1_{it}(\boldsymbol{\gamma})1_{is}(\boldsymbol{\gamma})]$, and $E[\boldsymbol{q}_{it}'\boldsymbol{w}_{it}1_{is}(\boldsymbol{\gamma})]$ are also continuous in $\boldsymbol{\gamma}$. We therefore conclude that $\boldsymbol{Q}(\boldsymbol{\theta})$ is continuous in $\boldsymbol{\theta}$. Evaluated at $\boldsymbol{\theta}_0$, we find $\boldsymbol{Q}(\boldsymbol{\theta}_0) = \boldsymbol{Q}$. Thus we obtain Condition 3.

We next establish Condition 4. We write $\boldsymbol{m}_{it}(\boldsymbol{\theta}) = (\boldsymbol{m}_{it}^1(\boldsymbol{\theta})', m_{it}^2(\boldsymbol{\theta})', \boldsymbol{m}_{it}^3(\boldsymbol{\theta})')'$, in which $\boldsymbol{m}_{it}(\boldsymbol{\theta}) = \boldsymbol{h}_{it}(\boldsymbol{\theta})e_{it}(\boldsymbol{\theta})$, $\boldsymbol{m}_{it}^1(\boldsymbol{\theta})' = \ddot{\boldsymbol{x}}_{it}(\boldsymbol{\gamma})[\ddot{y}_{it} - \boldsymbol{\beta}'\ddot{\boldsymbol{x}}_{it}(\boldsymbol{\gamma})]$, $m_{it}^2(\boldsymbol{\theta})' = -[\beta_1^-\ddot{1}_{it}^-(\boldsymbol{\gamma}) + \beta_1^+\ddot{1}_{it}^+(\boldsymbol{\gamma})][\ddot{y}_{it} - \boldsymbol{\beta}'\ddot{\boldsymbol{x}}_{it}(\boldsymbol{\gamma})]$, and $\boldsymbol{m}_{it}^3(\boldsymbol{\theta})' = -[\beta_1^-\ddot{\boldsymbol{q}}_{it}^-(\boldsymbol{\gamma}) + \beta_1^+\ddot{\boldsymbol{q}}_{it}^+(\boldsymbol{\gamma})][\ddot{y}_{it} - \boldsymbol{\beta}'\ddot{\boldsymbol{x}}_{it}(\boldsymbol{\gamma})]$.

Noting that the first part is linear in $\boldsymbol{\beta}$, and the second and third terms are quadratic in $\boldsymbol{\beta}$, it suffices to show the stochastic equicontinuity with regard to $\boldsymbol{\gamma} = (\gamma_0, \boldsymbol{\gamma}_1')$. We thus simplify notation by writing $\boldsymbol{m}_{it}(\boldsymbol{\theta}) = \boldsymbol{m}_{it}^*(\boldsymbol{\gamma})$. Note that $\boldsymbol{\gamma}$ enters all the terms in $\boldsymbol{m}_{it}(\boldsymbol{\theta})$ through one of the following forms (or its transpose): $E[y_{it}\boldsymbol{x}_{is}(\boldsymbol{\gamma})]$, $E[\boldsymbol{x}_{it}(\boldsymbol{\gamma})\boldsymbol{x}_{is}(\gamma)]$, $E[\boldsymbol{x}_{it}(\gamma)\boldsymbol{z}_{is}]$, $E[\boldsymbol{w}_{it}1_{is}(\boldsymbol{\gamma})]$, $E[\boldsymbol{q}_{it}'\boldsymbol{w}_{it}1_{is}(\boldsymbol{\gamma})]$ for $t, s = 1, \ldots, T$. Under Assumption 1, $\boldsymbol{m}_{it}^*(\boldsymbol{\gamma})$ has a bounded $2r$-th moment, and the envelop condition holds. For any δ set $N(\delta) = \delta^{-2\tau}$ and set $\boldsymbol{\gamma}_{\cdot,k} = [\gamma_{0,k}, \boldsymbol{\gamma}_{1,k}']'$, $k = 1, \ldots, N_\delta$, to be an equally spaced grid on $\boldsymbol{\Gamma}$. Notice that the distance between the grid points is $O(\frac{1}{N_\delta})$. Define $\boldsymbol{m}_{itk}^* = \min[\boldsymbol{m}_{it}^*(\boldsymbol{\gamma}_{\cdot,k-1}), \boldsymbol{m}_{it}^*(\boldsymbol{\gamma}_{\cdot,k})]$ and $\boldsymbol{m}_{itk}^{**} = \max[\boldsymbol{m}_{it}^*(\boldsymbol{\gamma}_{\cdot,k-1}), \boldsymbol{m}_{it}^*(\boldsymbol{\gamma}_{\cdot,k})]$. Then, for each $\boldsymbol{\gamma} = [\gamma_0, \boldsymbol{\gamma}_1']'$, there exists $\boldsymbol{\gamma}_{\cdot,k} = [\gamma_{0,k}, \boldsymbol{\gamma}_{1,k}']'$ such that $\boldsymbol{m}_{itk}^* \le \boldsymbol{m}_{it}^*(\boldsymbol{\gamma}) \le \boldsymbol{m}_{itk}^{**}$. Thus, $[\boldsymbol{m}_{itk}^*, \boldsymbol{m}_{itk}^{**}]$ brackets $\boldsymbol{m}_{it}^*(\boldsymbol{\gamma})$. Using the bound of $E\left\|\boldsymbol{w}_{it}1(\gamma_{it,k-1} \le x_{it} \le \gamma_{it,k})\right\|^2$, we can obtain

$$\begin{aligned}
&E\left\|\boldsymbol{m}_{itk}^{**} - \boldsymbol{m}_{itk}^*\right\|^2 = E\left\|\boldsymbol{m}_{it}^*(\boldsymbol{\gamma}_{\cdot,k}) - \boldsymbol{m}_{it}^*(\boldsymbol{\gamma}_{\cdot,k-1})\right\|^2 \\
&\le \boldsymbol{C}\bar{f}_{\boldsymbol{q}}^{1/\tau}\{|\gamma_{0,k} - \gamma_{0,k-1}| + \boldsymbol{C}\|\boldsymbol{\gamma}_{1,k} - \boldsymbol{\gamma}_{1,k-1}\|\}^{1/\tau} \le O(N_\delta^{-\frac{1}{\tau}}) = O(\delta^2).
\end{aligned}$$

It follows that $N(\delta) = \delta^{-2\tau}$ are the L^2 bracketing numbers and $H_2(\delta) = \ln N(\delta) = O(|\ln \delta|)$ is the metric entropy with bracketing for the class $\{\boldsymbol{m}^*_{it}(\boldsymbol{\gamma}) : \boldsymbol{\gamma} \in \boldsymbol{\Gamma}\}$. Hence, Condition 4 holds by (2.17) of Doukhan et al. (1995). □

Proof of Theorem 2.2. For convenience, let $\boldsymbol{R}_1 = (1, -1, \boldsymbol{0}_{1\times l})'$, $\boldsymbol{z}_{it} = (\boldsymbol{q}'_{it}, \boldsymbol{z}'_{2it})'$, and $\boldsymbol{\beta}_2 = (\boldsymbol{\beta}'_{21}, \boldsymbol{\beta}'_{22})'$, and the true value $\boldsymbol{\theta}_0 = (\beta^-_{01}, \beta^+_{01}, \boldsymbol{\beta}'_{021}, \boldsymbol{\beta}'_{022}, \gamma_{00}, \boldsymbol{\gamma}'_{01})'$. Therefore, under $\beta^-_{01} = \beta^+_{01} \equiv \beta_{01}$ and for any fixed $\boldsymbol{\gamma} \in \boldsymbol{\Gamma}$, (2.1) can be rewritten as

$$
\begin{aligned}
y_{it} &= (x_{it} - \gamma_{it}(\boldsymbol{\gamma}_0))_-\, \beta_{01} + (x_{it} - \gamma_{it}(\boldsymbol{\gamma}_0))_+\, \beta_{01} + \boldsymbol{q}'_{it}\boldsymbol{\beta}_{021} + \boldsymbol{z}'_{2it}\boldsymbol{\beta}_{022} + \alpha_i + \varepsilon_{it} \\
&= (x_{it} - \gamma_{it}(\boldsymbol{\gamma}_0))\, \beta_{01} + \boldsymbol{q}'_{it}\boldsymbol{\beta}_{021} + \boldsymbol{z}'_{2it}\boldsymbol{\beta}_{022} + \alpha_i + \varepsilon_{it} \\
&= (x_{it} - \gamma_{it}(\boldsymbol{\gamma}))\, \beta_{01} + (\gamma_0 - \gamma_{00})\beta_{01} + \boldsymbol{q}'_{it}(\boldsymbol{\beta}_{021} + (\boldsymbol{\gamma}_1 - \boldsymbol{\gamma}_{01})\beta_{01}) + \boldsymbol{z}'_{2it}\boldsymbol{\beta}_{022} + \alpha_i + \varepsilon_{it} \\
&= (x_{it} - \gamma_{it}(\boldsymbol{\gamma}))_-\, \beta_{01} + (x_{it} - \gamma_{it}(\boldsymbol{\gamma}))_+\, \beta_{01} + (\gamma_0 - \gamma_{00})\beta_{01} \\
&\quad + \boldsymbol{q}'_{it}(\boldsymbol{\beta}_{021} + (\boldsymbol{\gamma}_1 - \boldsymbol{\gamma}_{01})\beta_{01}) + \boldsymbol{z}'_{2it}\boldsymbol{\beta}_{022} + \alpha_i + \varepsilon_{it} \\
&= \boldsymbol{x}'_{it}(\boldsymbol{\gamma})\boldsymbol{\beta}_0(\boldsymbol{\gamma}) + (\gamma_0 - \gamma_{00})\beta_{01} + \alpha_i + \varepsilon_{it},
\end{aligned} \tag{A.6}
$$

where $\boldsymbol{\beta}_0(\boldsymbol{\gamma}) = (\beta_{01}, \beta_{01}, \boldsymbol{\beta}'_{021} + (\boldsymbol{\gamma}_1 - \boldsymbol{\gamma}_{01})'\beta_{01}, \boldsymbol{\beta}'_{022})'$.

Thus, for any fixed $\boldsymbol{\gamma} \in \boldsymbol{\Gamma}$, by Theorem 1 in Hansen (1996), we have

$$
\begin{aligned}
\sqrt{N}\boldsymbol{R}'_1\left[\hat{\boldsymbol{\beta}}(\boldsymbol{\gamma}) - \boldsymbol{\beta}_0(\boldsymbol{\gamma})\right] &= \sqrt{N}\left(\hat{\beta}^-_1(\boldsymbol{\gamma}) - \hat{\beta}^+_1(\boldsymbol{\gamma})\right) \\
&= \boldsymbol{R}'_1\left[\frac{1}{N}\sum_{i=1}^N \ddot{\boldsymbol{x}}'_i(\boldsymbol{\gamma})\ddot{\boldsymbol{x}}_i(\boldsymbol{\gamma})\right]^{-1}\left[\frac{1}{\sqrt{N}}\sum_{i=1}^N \ddot{\boldsymbol{x}}'_i(\boldsymbol{\gamma})\ddot{\varepsilon}_i\right] \\
&\xrightarrow{d} \boldsymbol{R}'_1\boldsymbol{Q}^{-1}(\boldsymbol{\gamma})\boldsymbol{G}(\boldsymbol{\gamma}),
\end{aligned} \tag{A.7}
$$

and for any fixed $\boldsymbol{\gamma}_{(1)}, \boldsymbol{\gamma}_{(2)} \in \boldsymbol{\Gamma}$, as $N \to \infty$, we have

$$
\begin{aligned}
\hat{\boldsymbol{Q}}(\boldsymbol{\gamma}_{(1)}, \boldsymbol{\gamma}_{(2)}) &= \frac{1}{N}\sum_{i=1}^N \ddot{\boldsymbol{x}}'_i(\boldsymbol{\gamma}_{(1)})\ddot{\boldsymbol{x}}_i(\boldsymbol{\gamma}_{(2)}) \\
&\xrightarrow{a.s.} \boldsymbol{Q}(\boldsymbol{\gamma}_{(1)}, \boldsymbol{\gamma}_{(2)}) = \sum\nolimits_{t=1}^T E(\ddot{\boldsymbol{x}}_{it}(\boldsymbol{\gamma}_{(1)})\ddot{\boldsymbol{x}}'_{it}(\boldsymbol{\gamma}_{(2)})),
\end{aligned} \tag{A.8}
$$

noting that

$$
\begin{aligned}
\hat{\sigma}^2(\boldsymbol{\gamma}) &= \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} e_{it}^2(\boldsymbol{\gamma}) \\
&= \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} \ddot{\varepsilon}_{it}^2 - 2\left[\hat{\boldsymbol{\beta}}(\boldsymbol{\gamma}) - \boldsymbol{\beta}_0(\boldsymbol{\gamma})\right]' \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} \ddot{\boldsymbol{x}}_{it}(\boldsymbol{\gamma})\ddot{\varepsilon}_{it} \\
&\quad + \left[\hat{\boldsymbol{\beta}}(\boldsymbol{\gamma}) - \boldsymbol{\beta}_0(\boldsymbol{\gamma})\right]' \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} \ddot{\boldsymbol{x}}_{it}(\boldsymbol{\gamma})\ddot{\boldsymbol{x}}_{it}'(\boldsymbol{\gamma})\left[\hat{\boldsymbol{\beta}}(\boldsymbol{\gamma}) - \boldsymbol{\beta}_0(\boldsymbol{\gamma})\right] \\
&= \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} \ddot{\varepsilon}_{it}^2 - \frac{2}{\sqrt{N}}\sqrt{N}\left[\hat{\boldsymbol{\beta}}(\boldsymbol{\gamma}) - \boldsymbol{\beta}_0(\boldsymbol{\gamma})\right]' \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} \ddot{\boldsymbol{x}}_{it}(\boldsymbol{\gamma})\ddot{\varepsilon}_{it} \\
&\quad + \frac{1}{N}\sqrt{N}\left[\hat{\boldsymbol{\beta}}(\boldsymbol{\gamma}) - \boldsymbol{\beta}_0(\boldsymbol{\gamma})\right]' \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} \ddot{\boldsymbol{x}}_{it}(\boldsymbol{\gamma})\ddot{\boldsymbol{x}}_{it}'(\boldsymbol{\gamma})\sqrt{N}\left[\hat{\boldsymbol{\beta}}(\boldsymbol{\gamma}) - \boldsymbol{\beta}_0(\boldsymbol{\gamma})\right] \\
&= \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} \ddot{\varepsilon}_{it}^2 + O_p(N^{-1/2}) + O_p(N^{-1}) \\
&\xrightarrow{p} \sum_{t=1}^{T} E(\ddot{\varepsilon}_{it}^2) \triangleq \sigma_T^2. \qquad \text{(A.9)}
\end{aligned}
$$

Then, under $H_0^1 : \beta_{01}^- - \beta_{01}^+ = \boldsymbol{R}_1'\boldsymbol{\beta}_0(\boldsymbol{\gamma}) = 0$, we have

$$
\begin{aligned}
W_1 &= \sup_{\boldsymbol{\gamma}\in\boldsymbol{\Gamma}} \frac{N(\tilde{\sigma}^2 - \hat{\sigma}^2(\boldsymbol{\gamma}))}{\hat{\sigma}^2(\boldsymbol{\gamma})} \\
&= \sup_{\boldsymbol{\gamma}\in\boldsymbol{\Gamma}} \frac{N[\boldsymbol{R}_1'\hat{\boldsymbol{\beta}}(\boldsymbol{\gamma}) - 0]'[\boldsymbol{R}_1'(\ddot{X}'(\boldsymbol{\gamma})\ddot{X}(\boldsymbol{\gamma}))^{-1}\boldsymbol{R}_1]^{-1}[\boldsymbol{R}_1'\hat{\boldsymbol{\beta}}(\boldsymbol{\gamma}) - 0]}{\hat{\sigma}^2(\boldsymbol{\gamma})} \\
&= \sup_{\boldsymbol{\gamma}\in\boldsymbol{\Gamma}} \frac{\sqrt{N}[\hat{\boldsymbol{\beta}}(\boldsymbol{\gamma}) - \boldsymbol{\beta}_0(\boldsymbol{\gamma})]'\boldsymbol{R}_1[\boldsymbol{R}_1'\hat{\boldsymbol{Q}}^{-1}(\boldsymbol{\gamma})\boldsymbol{R}_1]^{-1}\boldsymbol{R}_1'[\hat{\boldsymbol{\beta}}(\boldsymbol{\gamma}) - \boldsymbol{\beta}_0(\boldsymbol{\gamma})]\sqrt{N}}{\hat{\sigma}^2(\boldsymbol{\gamma})} \\
&\xrightarrow{d} \sup_{\boldsymbol{\gamma}\in\boldsymbol{\Gamma}} \frac{\boldsymbol{G}'(\boldsymbol{\gamma})\boldsymbol{Q}^{-1}(\boldsymbol{\gamma})\boldsymbol{R}_1\left[\boldsymbol{R}_1'\boldsymbol{Q}^{-1}(\boldsymbol{\gamma})\boldsymbol{R}_1\right]^{-1}\boldsymbol{R}_1'\boldsymbol{Q}^{-1}(\boldsymbol{\gamma})\boldsymbol{G}(\boldsymbol{\gamma})}{\sigma_T^2}. \qquad \text{(A.10)}
\end{aligned}
$$

□

We turn next to the test statistic W_2. Denote $\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}\in\mathbf{B}\times\boldsymbol{\Gamma}} \widetilde{SSR}_{NT}(\boldsymbol{\theta})$ and $\hat{\boldsymbol{\theta}}_C = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Gamma}} \widetilde{SSR}_{NT}(\boldsymbol{\theta})$ subject to $\boldsymbol{R}_2\boldsymbol{\theta} = \mathbf{0}_{k\times 1}$, where $\boldsymbol{R}_2 = [\mathbf{0}_{k\times(l+2)}, \mathbf{I}_k]$. The first- and second-order conditions for $\hat{\boldsymbol{\theta}}$ yield

$$
\begin{aligned}
\frac{\partial \widetilde{SSR}_{NT}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} &= \frac{\partial}{\partial\boldsymbol{\theta}}\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} e_{it}^2(\boldsymbol{\theta}) = \frac{1}{N}\sum_{i=1}^{N}\frac{\partial}{\partial\boldsymbol{\theta}}\boldsymbol{e}_i'(\boldsymbol{\theta})\boldsymbol{e}_i(\boldsymbol{\theta}) \\
&= \frac{1}{N}\sum_{i=1}^{N}\frac{2}{-1}\left[-\frac{\partial\boldsymbol{e}_i'(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right]\boldsymbol{e}_i(\boldsymbol{\theta}) = \frac{-2}{N}\sum_{i=1}^{N}\boldsymbol{h}_i'(\boldsymbol{\theta})\boldsymbol{e}_i(\boldsymbol{\theta}) \\
&\equiv \frac{-2}{\sqrt{N}}\boldsymbol{Z}_N(\boldsymbol{\theta}), \qquad \text{(A.11)}
\end{aligned}
$$

$$\begin{aligned}\frac{\partial^2 \widetilde{SSR}_{NT}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'\partial\boldsymbol{\theta}} &= \frac{-2}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\frac{\partial}{\partial\boldsymbol{\theta}'}\boldsymbol{h}_{it}(\boldsymbol{\theta})e_{it}(\boldsymbol{\theta}) \\ &= \frac{2}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{-\boldsymbol{h}_{it}\frac{\partial}{\partial\boldsymbol{\theta}'}e_{it}(\boldsymbol{\theta}) + \left[-\frac{\partial\boldsymbol{h}_{it}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right]e_{it}(\boldsymbol{\theta})\right\} \\ &= \frac{2}{N}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\boldsymbol{h}_{it}(\boldsymbol{\theta})\boldsymbol{h}_{it}'(\boldsymbol{\theta}) + \boldsymbol{D}_{it}(\boldsymbol{\theta})e_{it}(\boldsymbol{\theta})\right] \\ &\equiv 2\boldsymbol{G}_N(\boldsymbol{\theta}),\end{aligned} \tag{A.12}$$

where $\boldsymbol{Z}_N(\boldsymbol{\theta}) = N^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}\boldsymbol{h}_{it}(\boldsymbol{\theta})e_{it}(\boldsymbol{\theta})$ and $\boldsymbol{G}_N(\boldsymbol{\theta}) = N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}[\boldsymbol{h}_{it}(\boldsymbol{\theta})\boldsymbol{h}_{it}'(\boldsymbol{\theta}) + \boldsymbol{D}_{it}(\boldsymbol{\theta})e_{it}(\boldsymbol{\theta})]$. The first-order condition for the solution of $\hat{\boldsymbol{\theta}}$ is

$$\left.\frac{\partial \widetilde{SSR}_{NT}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = \frac{-2}{\sqrt{N}}\boldsymbol{Z}_N(\hat{\boldsymbol{\theta}}) = \boldsymbol{0}_{(l+k+2)\times 1}. \tag{A.13}$$

Taking Taylor's expansion of the unconstrained first-order conditions at $\boldsymbol{\theta}_0$ gives

$$\frac{-2}{\sqrt{N}}\boldsymbol{Z}_N(\hat{\boldsymbol{\theta}}) = \boldsymbol{0} = \frac{-2}{\sqrt{N}}\boldsymbol{Z}_N(\boldsymbol{\theta}_0) + 2\boldsymbol{G}_N(\boldsymbol{\theta}^{\dagger})\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right), \tag{A.14}$$

where $\boldsymbol{\theta}^{\dagger}$ is a value between $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$. Therefore, from conditions of the proof of Theorem 2.1, as $N \to \infty$, we have

(1) $\hat{\boldsymbol{\theta}} \to_p \boldsymbol{\theta}_0$;
(2) $\boldsymbol{\theta}^{\dagger} \to_p \boldsymbol{\theta}_0$ as $\hat{\boldsymbol{\theta}} \to_p \boldsymbol{\theta}_0$;
(3) $\boldsymbol{Z}_N(\boldsymbol{\theta}_0) \to_d \boldsymbol{Z} \sim \boldsymbol{\mathcal{N}}(\boldsymbol{0}, \boldsymbol{S})$;
(4) $\boldsymbol{G}_N(\boldsymbol{\theta}_0) \to_p \boldsymbol{G} = E(\boldsymbol{h}_i'\boldsymbol{h}_i) + \sum_{t=1}^{T}E(\boldsymbol{D}_{it}\boldsymbol{e}_{it})$.

The equation (A.14) can be rewritten as

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) = \boldsymbol{G}_N^{-1}(\boldsymbol{\theta}_0)\boldsymbol{Z}_N(\boldsymbol{\theta}_0) + o_p(1). \tag{A.15}$$

Thus, by Theorem 2.1 we obtain

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \xrightarrow{d} \boldsymbol{G}^{-1}\boldsymbol{Z} \sim N(\boldsymbol{0}, \boldsymbol{G}^{-1}\boldsymbol{S}\boldsymbol{G}^{-1}). \tag{A.16}$$

Next, under $H_0^2 : \boldsymbol{R}_2\boldsymbol{\theta} = \boldsymbol{0}_{k\times 1} = \boldsymbol{R}_2\boldsymbol{\theta}_0$, using a similar argument in Condition 1 of the proof of Theorem 2.1, we have $\hat{\boldsymbol{\theta}}_C \to_p \boldsymbol{\theta}_0$. Hence, a Taylor's expansion of $\partial\widetilde{SSR}_{NT}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_C}$ at $\boldsymbol{\theta}_0$ gives

$$\begin{aligned}\frac{-2}{\sqrt{N}}\boldsymbol{Z}_N(\hat{\boldsymbol{\theta}}_C) &= \frac{-2}{\sqrt{N}}\boldsymbol{Z}_N(\boldsymbol{\theta}_0) + 2\boldsymbol{G}_N(\boldsymbol{\theta}^{\ddagger})\left(\hat{\boldsymbol{\theta}}_C - \boldsymbol{\theta}_0\right) \\ &= \frac{-2}{\sqrt{N}}\boldsymbol{Z}_N(\boldsymbol{\theta}_0) + 2\boldsymbol{G}_N(\boldsymbol{\theta}_0)\left(\hat{\boldsymbol{\theta}}_C - \boldsymbol{\theta}_0\right) + o_p(1),\end{aligned} \tag{A.17}$$

where $\boldsymbol{\theta}^{\ddagger}$ is a value between $\hat{\boldsymbol{\theta}}_C$ and $\boldsymbol{\theta}_0$. Substituting (A.17) in the first-order condition for $\hat{\boldsymbol{\theta}}_C$ yields

$$\begin{aligned}&\begin{bmatrix}\boldsymbol{Z}_N(\boldsymbol{\theta}_0)\\ \boldsymbol{0}_{k\times 1}\end{bmatrix} - \begin{bmatrix}\boldsymbol{G}_N(\boldsymbol{\theta}_0) & \boldsymbol{R}_2'\\ \boldsymbol{R}_2 & \boldsymbol{0}_{k\times k}\end{bmatrix}\begin{bmatrix}\sqrt{N}(\hat{\boldsymbol{\theta}}_C - \boldsymbol{\theta}_0)\\ \sqrt{N}\hat{\boldsymbol{\lambda}}_C\end{bmatrix} + \begin{bmatrix}o_p(1)\\ \boldsymbol{0}_{k\times 1}\end{bmatrix} \\ &\qquad = \begin{bmatrix}\boldsymbol{0}_{(l+k+2)\times 1}\\ \boldsymbol{0}_{k\times 1}\end{bmatrix},\end{aligned} \tag{A.18}$$

where $\hat{\boldsymbol{\lambda}}_C$ is the solution of the first-order conditions of Lagrangian multipliers. From (A.15) and (A.18), we obtain

$$\begin{aligned}\sqrt{N}\left(\hat{\boldsymbol{\theta}}_C-\boldsymbol{\theta}_0\right) &= \boldsymbol{G}_N^{-1}(\boldsymbol{\theta}_0)\boldsymbol{Z}_N(\boldsymbol{\theta}_0)+\boldsymbol{G}_N^{-1}(\boldsymbol{\theta}_0)\boldsymbol{R}_2'\left[\boldsymbol{R}_2\boldsymbol{G}_N^{-1}(\boldsymbol{\theta}_0)\boldsymbol{R}_2'\right]^{-1}\\ &\quad\times\boldsymbol{R}_2\boldsymbol{G}_N^{-1}(\boldsymbol{\theta}_0)\boldsymbol{Z}_N(\boldsymbol{\theta}_0)+o_p(1)\\ &=\sqrt{N}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)+\boldsymbol{G}_N^{-1}(\boldsymbol{\theta}_0)\boldsymbol{R}_2'\left[\boldsymbol{R}_2\boldsymbol{G}_N^{-1}(\boldsymbol{\theta}_0)\boldsymbol{R}_2'\right]^{-1}\\ &\quad\times\boldsymbol{R}_2\sqrt{N}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)+o_p(1),\end{aligned}\tag{A.19}$$

noting that (A.19) can be rewritten as

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}}_C-\hat{\boldsymbol{\theta}}\right)=\boldsymbol{G}_N^{-1}(\boldsymbol{\theta}_0)\boldsymbol{R}_2'\left[\boldsymbol{R}_2\boldsymbol{G}_N^{-1}(\boldsymbol{\theta}_0)\boldsymbol{R}_2'\right]^{-1}\boldsymbol{R}_2\sqrt{N}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right)+o_p(1).\tag{A.20}$$

Finally, a Taylor's expansion of $\widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}}_C)$ at $\hat{\boldsymbol{\theta}}$ gives

$$\begin{aligned}\widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}}_C) &= \widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}})+\frac{-2}{\sqrt{N}}\boldsymbol{Z}_N(\hat{\boldsymbol{\theta}})\left[\hat{\boldsymbol{\theta}}_C-\hat{\boldsymbol{\theta}}\right]\\ &\quad+\frac{1}{2}\left[\hat{\boldsymbol{\theta}}_C-\hat{\boldsymbol{\theta}}\right]'2\boldsymbol{G}_N(\boldsymbol{\theta}^{\S})\left[\hat{\boldsymbol{\theta}}_C-\hat{\boldsymbol{\theta}}\right]\\ &=\widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}})+\left[\hat{\boldsymbol{\theta}}_C-\hat{\boldsymbol{\theta}}\right]'\boldsymbol{G}_N(\boldsymbol{\theta}_0)\left[\hat{\boldsymbol{\theta}}_C-\hat{\boldsymbol{\theta}}\right]+o_p(1).\end{aligned}\tag{A.21}$$

Combining (A.20) and (A.21), we have

$$\begin{aligned}&\widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}}_C)-\widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}})\\ &=\left[\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right]'\boldsymbol{R}_2'\left[\boldsymbol{R}_2\boldsymbol{G}_N^{-1}(\boldsymbol{\theta}_0)\boldsymbol{R}_2'\right]^{-1}\boldsymbol{R}_2\left[\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right]+o_p(1),\end{aligned}\tag{A.22}$$

and a Taylor's expansion of $\widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}})$ at $\boldsymbol{\theta}_0$ gives

$$\begin{aligned}\widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}}) &= \widetilde{SSR}_{NT}(\boldsymbol{\theta}_0)+\frac{-2}{\sqrt{N}}\boldsymbol{Z}_N(\boldsymbol{\theta}_0)\left[\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right]\\ &\quad+\frac{1}{2}\left[\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right]'2\boldsymbol{G}_N(\boldsymbol{\theta}^*)\left[\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\right]\\ &= \widetilde{SSR}_{NT}(\boldsymbol{\theta}_0)+o_p(1)\\ &\xrightarrow{p}\sum_{t=1}^{T}E(\ddot{\varepsilon}_{it}^2)\equiv\sigma_T^2,\end{aligned}\tag{A.23}$$

where $\boldsymbol{\theta}^*$ is a value between $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$.

Thus, we have

$$
\begin{aligned}
W_2 &= \frac{\hat{\sigma}_C^2(\hat{\gamma}_0) - \hat{\sigma}_N^2(\hat{\boldsymbol{\gamma}})}{\hat{\sigma}_N^2(\hat{\boldsymbol{\gamma}})/N} = \frac{N[\widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}}_C) - \widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}})]}{\widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}})} \\
&= \frac{\sqrt{N}\left[\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right]' \boldsymbol{R}_2' \left[\boldsymbol{R}_2 \boldsymbol{G}_N^{-1}(\boldsymbol{\theta}_0)\boldsymbol{R}_2'\right]^{-1} \boldsymbol{R}_2 \sqrt{N}\left[\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right]}{\widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}})} + o_p(1) \\
&= \frac{\boldsymbol{Z}_N'(\boldsymbol{\theta}_0)\boldsymbol{G}_N^{-1}(\boldsymbol{\theta}_0)\boldsymbol{R}_2' \left[\boldsymbol{R}_2 \boldsymbol{G}_N^{-1}(\boldsymbol{\theta}_0)\boldsymbol{R}_2'\right]^{-1} \boldsymbol{R}_2 \boldsymbol{G}_N^{-1}(\boldsymbol{\theta}_0)\boldsymbol{Z}_N(\boldsymbol{\theta}_0)}{\widetilde{SSR}_{NT}(\hat{\boldsymbol{\theta}})} + o_p(1) \\
&\xrightarrow{d} \frac{\boldsymbol{Z}'\boldsymbol{G}^{-1}\boldsymbol{R}_2'[\boldsymbol{R}_2\boldsymbol{G}^{-1}\boldsymbol{R}_2']^{-1}\boldsymbol{R}_2\boldsymbol{G}^{-1}\boldsymbol{Z}}{\sigma_T^2}.
\end{aligned}
\tag{A.24}
$$

■

Online appendix of PKTR-CDT published in *The Econometrics Journal*, pp. 1–6.

Online Appendix for "Panel Kink Threshold Regression Model with a Covariate-Dependent Threshold"

APPENDIX: EXTENTION TO DYNAMIC PANEL CONTEXT

This appendix extends the panel kink threshold regression model with a covariate-dependent threshold (PKTR-CDT) to the dynamic panel context. In the dynamic panel data models, since lagged dependent variables enters as independent variables, it is well-known that the usual fixed effects estimator may be inconsistent. In this case, we can use instrument variable to establish a consistent estimator. In this appendix, we briefly discuss the dynamic panel kink threshold regression model with a covariate-dependent threshold. Despite the necessary technical modifcations, all the issues discussed in this paper can be extended to the dynamic panel data context.

B.1. Dynamic panel kink threshold regression model with a covariate-dependent threshold

Following Arellano and Bond (1991) and Seo and Shin (2016), we consider the AR(1) panel data model

$$y_{it} = \beta_1^-(x_{it} - \gamma_{it})_- + \beta_1^+(x_{it} - \gamma_{it})_+ + \boldsymbol{\beta}_2'\boldsymbol{q}_{it} + \beta_3 y_{i,t-1} + \alpha_i + \varepsilon_{it}, \tag{B.1}$$

for $i = 1, 2, \ldots, N$ and $t = 2, 3, \ldots, T$, where ε_{it} is assumed to be a martingale difference sequence. The regressors x_{it} and covariates q_{it} are predetermined or strictly exogenous variables. Here, to eliminate the individual effect α_i, we take the first-difference transformation of (B.1)

$$\Delta y_{it} = \beta_1^-\Delta(x_{it} - \gamma_{it})_- + \beta_1^+\Delta(x_{it} - \gamma_{it})_+ + \boldsymbol{\beta}_2'\Delta\boldsymbol{q}_{it} + \beta_3\Delta y_{i,t-1} + \Delta\varepsilon_{it}, \tag{B.2}$$

where Δ is the first difference operator, i.e., $\Delta(x_{it}-\gamma_{it})_- = (x_{it}-\gamma_{it})_- - (x_{i,t-1}-\gamma_{i,t-1})_-$ and $\Delta(x_{it}-\gamma_{it})_+ = (x_{it}-\gamma_{it})_+ - (x_{i,t-1}-\gamma_{i,t-1})_+$. It is easily seen that Δy_{it} is function of ε_{it} and $\varepsilon_{i,t-1}$ and $\Delta y_{i,t-1}$ is function of $\varepsilon_{i,t-1}$ and $\varepsilon_{i,t-2}$, hence $\Delta y_{i,t-1}$ is correlated with $\Delta\varepsilon_{i,t-1}$. To fix this problem, we follow Arellano and Bond (1991) to estimate the parameter $\boldsymbol{\theta} = (\beta_1^-, \beta_1^+, \boldsymbol{\beta}_2', \beta_3, \boldsymbol{\gamma}')'$ by the generalized method of moments (GMM). By construction, we have the following moment conditions

$$E(y_{i,t-j}\Delta\varepsilon_{it}) = 0, \text{ for } j = 2, \ldots, t-1 \text{ and } t = 3, \ldots, T, \tag{B.3}$$

and if $(x_{it}, \boldsymbol{q}_{it})$ are predetermined, we have

$$E(x_{is}\Delta\varepsilon_{it}) = E(\boldsymbol{q}_{is}'\Delta\varepsilon_{it}) = 0, \text{ for } s \leq t-1 \text{ and } t = 3, \ldots, T, \tag{B.4}$$

or if $(x_{it}, \boldsymbol{q}_{it})$ are strictly exogenous, we have

$$E(x_{is}\Delta\varepsilon_{it}) = E(\boldsymbol{q}_{is}'\Delta\varepsilon_{it}) = 0, \text{ for } s = 1, \ldots, T \text{ and } t = 3, \ldots, T. \tag{B.5}$$

Let $w_{it} = (y_{i1}, \ldots, y_{t-2}, x_{i1}, \ldots, x_{i,t-1}, \boldsymbol{q}_{i1}', \ldots, \boldsymbol{q}_{i,t-1}')'$ for $(x_{it}, \boldsymbol{q}_{it})$ are predetermined or $w_{it} = (y_{i1}, \ldots, y_{t-2}, x_{i1}, \ldots, x_{iT}, \boldsymbol{q}'_{i1}, \ldots, \boldsymbol{q}'_{iT})'$ for $(x_{it}, \boldsymbol{q}_{it})$ are strictly exogenous, $w_i = \mathrm{diag}(w_{i3}, \ldots, w_{iT})'$ (i.e. w_i is a block diagonal matrix), and $\Delta\varepsilon_i = (\Delta\varepsilon_{i3}, \ldots, \Delta\varepsilon_{iT})'$. Thus, the moment equations in (B.3), (B.4) and (B.5) can be written as

$$E(w'_i \Delta\varepsilon_i) = \mathbf{0}. \tag{B.6}$$

B.2. The estimates and asymptotic properties

Let $\boldsymbol{x}_{it}(\boldsymbol{\gamma}) = ((x_{it} - \gamma_{it})_-, (x_{it} - \gamma_{it})_+, \boldsymbol{q}'_{it}, y_{i,t-1})'$, $\boldsymbol{\beta} = (\beta_1^-, \beta_1^+, \boldsymbol{\beta}'_2, \beta_3)'$, $\Delta\boldsymbol{x}_{it}(\boldsymbol{\gamma}) = \boldsymbol{x}_{it}(\boldsymbol{\gamma}) - \boldsymbol{x}_{i,t-1}(\boldsymbol{\gamma})$, $\Delta y_i = (\Delta y_{i3}, \ldots, \Delta y_{iT})'$, and $\Delta\boldsymbol{x}_i(\boldsymbol{\gamma}) = (\Delta\boldsymbol{x}_{i3}(\boldsymbol{\gamma}), \ldots, \Delta\boldsymbol{x}_{iT}(\boldsymbol{\gamma}))'$. Let $\Delta\boldsymbol{Y}$, $\Delta\boldsymbol{X}(\boldsymbol{\gamma})$, $\Delta\boldsymbol{\varepsilon}$ and $\boldsymbol{W}$ denote the data stacked over all individuals, i.e., $\Delta\boldsymbol{Y} = (\Delta y'_1, \ldots, \Delta y'_N)'$, $\Delta\boldsymbol{X}(\boldsymbol{\gamma}) = (\Delta\boldsymbol{x}'_1(\boldsymbol{\gamma}), \ldots, \Delta\boldsymbol{x}'_N(\boldsymbol{\gamma}))'$, $\Delta\boldsymbol{\varepsilon} = (\Delta\varepsilon'_1, \ldots, \Delta\varepsilon'_N)'$ and $\boldsymbol{W} = (w'_1, \ldots, w'_N)'$. Thus, (B.2) can be rewritten as $\Delta\boldsymbol{Y} = \Delta\boldsymbol{X}(\boldsymbol{\gamma})\boldsymbol{\beta} + \Delta\boldsymbol{\varepsilon}$.

Thus, for any given $\boldsymbol{\gamma}$, the GMM estimator of $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\gamma}) = \left[\Delta\boldsymbol{X}'(\boldsymbol{\gamma})\boldsymbol{W}\boldsymbol{A}_N\boldsymbol{W}'\Delta\boldsymbol{X}(\boldsymbol{\gamma})\right]^{-1} \Delta\boldsymbol{X}'(\boldsymbol{\gamma})\boldsymbol{W}\boldsymbol{A}_N\boldsymbol{W}'\Delta\boldsymbol{Y}, \tag{B.7}$$

where $\boldsymbol{A}_N$ is the optimal weighting matrix such that $\boldsymbol{A}_N \to_p [E(\boldsymbol{W}'\Delta\boldsymbol{\varepsilon}\Delta\boldsymbol{\varepsilon}'\boldsymbol{W})]^{-1} = \boldsymbol{\Omega}^{-1}$, in which $\boldsymbol{A}_N$ and $\boldsymbol{\Omega}$ are assumed to be positive definite.

Therefore, we obtain the GMM estimator of $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ by

$$\hat{\boldsymbol{\gamma}} = \underset{\boldsymbol{\gamma}\in\boldsymbol{\Gamma}}{\mathrm{argmin}} \Delta\hat{\boldsymbol{\varepsilon}}'(\boldsymbol{\gamma})\boldsymbol{W}\boldsymbol{A}_N\boldsymbol{W}'\Delta\hat{\boldsymbol{\varepsilon}}(\boldsymbol{\gamma}), \text{ and } \hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}(\hat{\boldsymbol{\gamma}}), \tag{B.8}$$

where $\Delta\hat{\boldsymbol{\varepsilon}}(\boldsymbol{\gamma}) = \Delta\boldsymbol{Y} - \Delta\boldsymbol{X}(\boldsymbol{\gamma})\hat{\boldsymbol{\beta}}(\boldsymbol{\gamma})$.

Following Seo and Shin (2016), the two-step optimal GMM estimator can be obtained as follows:

Step 1. Estimate the parameter $\boldsymbol{\theta}$ by setting $\boldsymbol{A}_N = \boldsymbol{I}$ or $\boldsymbol{A}_N = (N^{-1}\sum_{i=1}^N w'_i\boldsymbol{H}w_i)^{-1}$, where $\boldsymbol{H}$ is a $(T-2)$ square matrix which has twos in the main diagonal, minus ones in the first subdiagonals and zeroes otherwise,[1] and collect residuals, $\widehat{\Delta\varepsilon}_{it}$.

Step 2. Estimate the parameter $\boldsymbol{\theta}$ by setting

$$\boldsymbol{A}_N = \left[\frac{1}{N}\sum_{i=1}^N w'_i\widehat{\Delta\varepsilon}_i\widehat{\Delta\varepsilon}'_i w_i - \left(\frac{1}{N}\sum_{i=1}^N w'_i\widehat{\Delta\varepsilon}_i\right)\left(\frac{1}{N}\sum_{i=1}^N \widehat{\Delta\varepsilon}'_i w_i\right)\right]^{-1}, \tag{B.9}$$

where $\widehat{\Delta\varepsilon}_i = (\widehat{\Delta\varepsilon}_{i3}, \ldots, \widehat{\Delta\varepsilon}_{iT})'$.

The true value of $\boldsymbol{\theta}$ is denoted by $\boldsymbol{\theta}_0$. Define $\boldsymbol{K}(\boldsymbol{\theta}) = (\boldsymbol{K}_{\boldsymbol{\beta}}(\boldsymbol{\gamma}), \boldsymbol{K}_{\boldsymbol{\gamma}}(\boldsymbol{\theta}))$, where

$$\boldsymbol{K}_{\boldsymbol{\beta}}(\boldsymbol{\gamma}) = -E\left(w'_i\Delta\boldsymbol{x}_i(\boldsymbol{\gamma})\right), \ \boldsymbol{K}_{\boldsymbol{\gamma}}(\boldsymbol{\theta}) = \begin{bmatrix} E(\beta_1^-\Delta 1_i^{-\prime}(\boldsymbol{\gamma})w_i - \beta_1^+\Delta 1_i^{+\prime}(\boldsymbol{\gamma})w_i) \\ E(\beta_1^-\Delta\boldsymbol{q}_i^{-\prime}(\boldsymbol{\gamma})w_i - \beta_1^+\Delta\boldsymbol{q}_i^{+\prime}(\boldsymbol{\gamma})w_i) \end{bmatrix}',$$

in which $\Delta 1_{it}^-(\boldsymbol{\gamma}) = 1_{it}^-(\boldsymbol{\gamma}) - 1_{i,t-1}^-(\boldsymbol{\gamma})$, $\Delta\boldsymbol{q}_{it}^-(\boldsymbol{\gamma}) = \boldsymbol{q}_{it}^-(\boldsymbol{\gamma}) - \boldsymbol{q}_{i,t-1}^-(\boldsymbol{\gamma})$, $\boldsymbol{q}_{it}^-(\boldsymbol{\gamma}) = \boldsymbol{q}_{it}1_{it}^-(\boldsymbol{\gamma})$, $\Delta 1_i^-(\boldsymbol{\gamma}) = (\Delta 1_{i3}^-(\boldsymbol{\gamma}), \ldots, \Delta 1_{i3}^-(\boldsymbol{\gamma}))'$ and $\Delta\boldsymbol{q}_i^-(\boldsymbol{\gamma}) = (\Delta\boldsymbol{q}_{i3}^-(\boldsymbol{\gamma}), \ldots, \Delta\boldsymbol{q}_{i3}^-(\boldsymbol{\gamma}))'$. Similarly, we can define $\Delta 1_{it}^+(\boldsymbol{\gamma})$, $\Delta\boldsymbol{q}_{it}^+(\boldsymbol{\gamma})$, $\boldsymbol{q}_{it}^+(\boldsymbol{\gamma})$, $\Delta 1_i^+(\boldsymbol{\gamma})$ and $\Delta\boldsymbol{q}_i^+(\boldsymbol{\gamma})$.

To obtain the asymptotic properties of the GMM estimator, we make the following assumption.

[1]For $T = 6$, $\boldsymbol{H} = \begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{bmatrix}$.

Assumption B1. (i) For each t, $\boldsymbol{w}_{it} = (y_{it}, x_{it}, \boldsymbol{q}'_{it})$ are independently identically distributed (i.i.d.) across i; (ii) For some $r > 1$, $E|y_{it}|^{4r} < \infty$, $E|x_{it}|^{4r} < \infty$, $E|\boldsymbol{q}_{it}|^{4r} < \infty$, and $E|\varepsilon_{it}|^{4r} < \infty$; (iii) ε_{it} is a martingale difference sequence.

Assumption B2. (i) $\boldsymbol{w}_{it}$ has a probability density function (PDF) $\boldsymbol{f}_{\boldsymbol{w},t}(\boldsymbol{w})$, where $\boldsymbol{w}_{it} = (y_{it}, x_{it}, \boldsymbol{q}'_{it})$; (ii) $(\boldsymbol{w}_{it}, \boldsymbol{w}_{is})$ has a joint PDF $\boldsymbol{f}_{\boldsymbol{w},ts}$; (iii) x_{it} has a conditional probability density function given $\boldsymbol{q}_{it} = \boldsymbol{q}$, satisfying $\max_{1\le t\le T} f_{\boldsymbol{q},t}(x|\boldsymbol{q}) \le \bar{f}_{\boldsymbol{q}} < \infty$; and $F_{\boldsymbol{q}}$ is the corresponding conditional cumulative distribution function of x_{it} conditional on $\boldsymbol{q}_{it}$.

Assumption B3. The true value of $\boldsymbol{\theta}$ is fixed at $\boldsymbol{\theta}_0$. $\boldsymbol{\theta}_0$ are interior points of $\boldsymbol{\Theta}$, where $\boldsymbol{\Theta}$ is compact. $\boldsymbol{\Omega}$ is finite and positive definite.

Assumption B4. Let $\boldsymbol{K} = \boldsymbol{K}(\boldsymbol{\theta}_0)$, then $\boldsymbol{K}$ is of full collumn rank.

Assumptions B1 and B2 are the same as in section 2, except that ε_{it} is a martingale difference sequence in the dynamic panel data model. Assumptions B3 and B4 are standard in the GMM framework for identification.

THEOREM B.1. *Under Assumption B1-B4, as $N \to \infty$*

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \xrightarrow{d} N\left(\mathbf{0}, \left(\boldsymbol{K}'\boldsymbol{\Omega}^{-1}\boldsymbol{K}\right)^{-1}\right). \tag{B.10}$$

In the GMM estimator, the slope and threshold estimators are also jointly asymptotically normal with the same convergence rate in Theorem 2.1.

B.3. Testing for kink threshold effect and threshold constancy

In this section, we propose test statistics for kink threshold effect and threshold constancy in the dynamic panel kink threshold regression model with a covariate-dependent threshold. Under the linear nulll hypothesis $H_0^1 : \beta_1^- = \beta_1^+$, for any given $\boldsymbol{\gamma} \in \boldsymbol{\Gamma}$, we consider the sup-Wald statistic for testing kink effect given by

$$W_3 = \sup_{\boldsymbol{\gamma}\in\boldsymbol{\Gamma}} \frac{N[\hat{\beta}_1^-(\boldsymbol{\gamma}) - \hat{\beta}_1^+(\boldsymbol{\gamma})]^2}{\widehat{\mathrm{var}}(\sqrt{N}(\hat{\beta}_1^-(\boldsymbol{\gamma}) - \hat{\beta}_1^+(\boldsymbol{\gamma})))}, \tag{B.11}$$

where $\widehat{\mathrm{var}}(\sqrt{N}(\hat{\beta}_1^-(\boldsymbol{\gamma}) - \hat{\beta}_1^+(\boldsymbol{\gamma}))) = \boldsymbol{R}'_3[\hat{\boldsymbol{K}}'_{\boldsymbol{\beta}}(\boldsymbol{\gamma})\hat{\boldsymbol{\Omega}}^{-1}(\hat{\boldsymbol{\theta}}(\boldsymbol{\gamma}))\hat{\boldsymbol{K}}_{\boldsymbol{\beta}}(\boldsymbol{\gamma})]^{-1}\boldsymbol{R}_3$, in which $\boldsymbol{R}_3 = (1, -1, \mathbf{0}_{1\times(k+1)})'$, $\hat{\boldsymbol{\theta}}(\boldsymbol{\gamma}) = (\hat{\boldsymbol{\beta}}'(\boldsymbol{\gamma}), \boldsymbol{\gamma}')'$, $\hat{\boldsymbol{K}}(\hat{\boldsymbol{\theta}}(\boldsymbol{\gamma})) = (\hat{\boldsymbol{K}}_{\boldsymbol{\beta}}(\boldsymbol{\gamma}), \hat{\boldsymbol{K}}_{\boldsymbol{\gamma}}(\hat{\boldsymbol{\theta}}(\boldsymbol{\gamma})))$,

$$\hat{\boldsymbol{K}}_{\boldsymbol{\beta}}(\boldsymbol{\gamma}) = -\frac{1}{N}\sum_{i=1}^{N} w_i'\Delta\boldsymbol{x}_i(\boldsymbol{\gamma}),$$

$$\hat{\boldsymbol{K}}_{\boldsymbol{\gamma}}(\hat{\boldsymbol{\theta}}(\boldsymbol{\gamma})) = \begin{bmatrix} N^{-1}\sum_{i=1}^{N}\left(\hat{\beta}_1^-(\boldsymbol{\gamma})\Delta 1_i^{-\prime}(\boldsymbol{\gamma})w_i - \hat{\beta}_1^+(\boldsymbol{\gamma})\Delta 1_i^{+\prime}(\boldsymbol{\gamma})w_i\right) \\ N^{-1}\sum_{i=1}^{N}\left(\hat{\beta}_1^-(\boldsymbol{\gamma})\Delta\boldsymbol{q}_i^{-\prime}(\boldsymbol{\gamma})w_i - \hat{\beta}_1^+(\boldsymbol{\gamma})\Delta\boldsymbol{q}_i^{+\prime}(\boldsymbol{\gamma})w_i\right) \end{bmatrix}',$$

and

$$\hat{\boldsymbol{\Omega}}(\hat{\boldsymbol{\theta}}(\boldsymbol{\gamma})) = \frac{1}{N}\sum_{i=1}^{N} w_i'\Delta\hat{\varepsilon}_i(\boldsymbol{\gamma})\Delta\hat{\varepsilon}_i(\boldsymbol{\gamma})'w_i - \left(\frac{1}{N}\sum_{i=1}^{N} w_i'\Delta\hat{\varepsilon}_i(\boldsymbol{\gamma})\right)\left(\frac{1}{N}\sum_{i=1}^{N}\Delta\hat{\varepsilon}_i(\boldsymbol{\gamma})'w_i\right),$$

where $\Delta\hat{\varepsilon}_i(\boldsymbol{\gamma}) = \Delta y_i - \Delta\boldsymbol{x}_i(\boldsymbol{\gamma})\hat{\beta}(\boldsymbol{\gamma})$.

Under the null $H_0^2 : \boldsymbol{\gamma}_1 = \mathbf{0}$, a test statistic for threshold constancy can be given by

$$W_4 = N\hat{\boldsymbol{\theta}}'\boldsymbol{R}_4\left(\boldsymbol{R}_4'\hat{\boldsymbol{K}}'\hat{\boldsymbol{\Omega}}^{-1}\hat{\boldsymbol{K}}\boldsymbol{R}_4\right)^{-1}\boldsymbol{R}_4'\hat{\boldsymbol{\theta}}, \tag{B.12}$$

where $\boldsymbol{R}_4 = (\mathbf{0}_{k\times(k+3)}, \boldsymbol{I}_k)$, $\hat{\boldsymbol{K}} = \hat{\boldsymbol{K}}(\hat{\boldsymbol{\theta}}(\hat{\gamma}))$ and $\hat{\boldsymbol{\Omega}} = \hat{\boldsymbol{\Omega}}(\hat{\boldsymbol{\theta}}(\hat{\gamma}))$.

The limiting distributions of the statistics in W_3 and W_4 are given as follows.

THEOREM B.2. *Let $\boldsymbol{V}(\gamma) = \boldsymbol{K}_{\boldsymbol{\beta}}'(\gamma)\boldsymbol{\Omega}^{-1}\boldsymbol{K}_{\boldsymbol{\beta}}(\gamma)$. Suppsose that $\inf_{\boldsymbol{\gamma}\in\boldsymbol{\Gamma}}\det(\boldsymbol{V}(\gamma)) > 0$ and Assumptions B1-B4 hold. As $N \to \infty$, under H_0^1 we have*

$$W_3 \xrightarrow{d} \sup_{\boldsymbol{\gamma}\in\boldsymbol{\Gamma}} \boldsymbol{Z}_1'\boldsymbol{K}_{\boldsymbol{\beta}}(\gamma)\boldsymbol{V}^{-1}(\gamma)\boldsymbol{R}_3\left[\boldsymbol{R}_3'\boldsymbol{V}^{-1}(\gamma)\boldsymbol{R}_3\right]^{-1}\boldsymbol{R}_3'\boldsymbol{V}^{-1}(\gamma)\boldsymbol{K}_{\boldsymbol{\beta}}'(\gamma)\boldsymbol{Z}_1, \tag{B.13}$$

and under H_0^2, we have

$$W_4 \xrightarrow{d} \chi_k^2, \tag{B.14}$$

where $\boldsymbol{Z}_1 \sim N(\mathbf{0}, \boldsymbol{\Omega}^{-1})$.

The limiting distribution of W_3 is not straightforward to pivotalize the statistic and tabulate the critical values. Thus, the p-values can be simulated following the similar bootstrap procedure in Section 2.2.

Proof of Theorem B.1: Let $\Delta\varepsilon_i(\boldsymbol{\theta}) = \Delta y_i - \Delta\boldsymbol{x}_i(\boldsymbol{\gamma})\boldsymbol{\beta}$, $g_i(\boldsymbol{\theta}) = w_i'\Delta\varepsilon_i(\boldsymbol{\theta})$, $\xi_i(\boldsymbol{\gamma}) = w_i'\Delta\boldsymbol{x}_i(\boldsymbol{\gamma})$, $g_i = g_i(\boldsymbol{\theta}_0) = w_i'\Delta\varepsilon_i$ and $\xi_i = \xi_i(\boldsymbol{\gamma}_0)$. Then, we can rewrite the moment indicator $g_i(\boldsymbol{\theta})$ as

$$\begin{aligned} g_i(\boldsymbol{\theta}) &= g_i + \xi_i\boldsymbol{\beta}_0 - \xi_i(\boldsymbol{\gamma})\boldsymbol{\beta} && \text{(B.15)}\\ &= g_i - \xi_i(\boldsymbol{\gamma})\left[\boldsymbol{\beta} - \boldsymbol{\beta}_0\right] - \left[\xi_i(\boldsymbol{\gamma}) - \xi_i\right]\boldsymbol{\beta}_0. && \text{(B.16)} \end{aligned}$$

We follow the proof of Seo and Shin (2016) to verify that asymptotically normal also hold for our model. To this end, we first establish consistency and then derive the asymptotic normality.

As shown in Seo and Shin (2016), the rank condition in Assumption B4 is sufficient to show that $E(g_i(\boldsymbol{\theta}_0)) = \mathbf{0}$ if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ for consistency.

By the linearity in the slope parameters for a fixed $\boldsymbol{\gamma}$, we have

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\gamma}) - \boldsymbol{\beta}_0 = \left[\bar{\xi}_N'(\boldsymbol{\gamma})\boldsymbol{A}_N\bar{\xi}_N(\boldsymbol{\gamma})\right]^{-1}\bar{\xi}_N'(\boldsymbol{\gamma})\boldsymbol{A}_N\left[\bar{g}_N + \frac{1}{N}\sum_{i=1}^{N}\left(\xi_i - \xi_i(\boldsymbol{\gamma})\right)\boldsymbol{\beta}_0\right], \tag{B.17}$$

where $\bar{\xi}_N(\boldsymbol{\gamma}) = N^{-1}\sum_{i=1}^N \xi_i(\boldsymbol{\gamma})$ and $\bar{g}_N = N^{-1}\sum_{i=1}^N g_i$. We can show that $\boldsymbol{A}_N \to_p \boldsymbol{\Omega}^{-1}$, $\bar{g}_N \to_p E(g_i) = \mathbf{0}$ and $\bar{\xi}_N(\boldsymbol{\gamma}) \to_p \xi(\boldsymbol{\gamma}) = E(\xi_i(\boldsymbol{\gamma}))$ uniformly by the standard weak law of large number (WLLN), and thus we have

$$\hat{\boldsymbol{\beta}}(\boldsymbol{\gamma}) - \boldsymbol{\beta}_0 \xrightarrow{p} \left[\xi'(\boldsymbol{\gamma})\boldsymbol{\Omega}^{-1}\xi(\boldsymbol{\gamma})\right]^{-1}\xi'(\boldsymbol{\gamma})\boldsymbol{\Omega}^{-1}\left[\xi - \xi(\boldsymbol{\gamma})\right]\boldsymbol{\beta}_0, \tag{B.18}$$

where $\xi = E(\xi_i)$. Since $\bar{g}_N(\boldsymbol{\theta})$ is continuous in $\boldsymbol{\beta}$ for any given $\boldsymbol{\gamma}$, using the continuous mapping theorem , (B.16) and (B.18) yields that

$$\begin{aligned}\bar{g}_N(\hat{\boldsymbol{\beta}}(\boldsymbol{\gamma}),\boldsymbol{\gamma}) &= \bar{g}_N - \bar{\xi}_N(\boldsymbol{\gamma})\left[\hat{\boldsymbol{\beta}}(\boldsymbol{\gamma}) - \boldsymbol{\beta}_0\right] + \left[\bar{\xi}_N - \bar{\xi}_N(\boldsymbol{\gamma})\right]\boldsymbol{\beta}_0 \\ &\xrightarrow{p} \left(\boldsymbol{I} - \xi(\boldsymbol{\gamma})\left[\xi'(\boldsymbol{\gamma})\boldsymbol{\Omega}^{-1}\xi(\boldsymbol{\gamma})\right]^{-1}\xi'(\boldsymbol{\gamma})\boldsymbol{\Omega}^{-1}\right)\left[\xi - \xi(\boldsymbol{\gamma})\right]\boldsymbol{\beta}_0. \quad \text{(B.19)}\end{aligned}$$

The term in the first brackets in the right hand side is positive definite and $\xi(\boldsymbol{\gamma}) = \xi$ if and only if $\boldsymbol{\gamma} = \boldsymbol{\gamma}_0$. Therefore, $\text{plim}_{N\to\infty}\bar{g}_N'(\hat{\boldsymbol{\beta}}(\boldsymbol{\gamma}),\boldsymbol{\gamma})\boldsymbol{A}_N\bar{g}_N(\hat{\boldsymbol{\beta}}(\boldsymbol{\gamma}),\boldsymbol{\gamma})$ is continuous and uniquely minimized at $\boldsymbol{\gamma} = \boldsymbol{\gamma}_0$ and the convergence is uniform, which implies the consistency of the estimator.

We next establish the asymptotic distribution. Let $\bar{J}_N(\boldsymbol{\theta}) = \bar{g}_N'(\boldsymbol{\theta})\boldsymbol{A}_N\bar{g}_N(\boldsymbol{\theta})$, $J_N(\boldsymbol{\theta}) = E(g_i'(\boldsymbol{\theta}))\boldsymbol{A}_N E(g_i(\boldsymbol{\theta}))$ and $\boldsymbol{D}_N = 2\boldsymbol{K}'\boldsymbol{A}_N\bar{g}_N$. We first show that for any $h_N \to 0$

$$\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\le h_N} \frac{\sqrt{N}R_N(\boldsymbol{\theta})}{1+\sqrt{N}\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|} = o_p(1), \tag{B.20}$$

where $R_N(\boldsymbol{\theta}) = \bar{J}_N(\boldsymbol{\theta}) - \bar{J}_N(\boldsymbol{\theta}_0) - J_N(\boldsymbol{\theta}) - \boldsymbol{D}_N'(\boldsymbol{\theta}-\boldsymbol{\theta}_0)$. To show this, we need to prove that

$$\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\le h_N} \|\varepsilon_N(\boldsymbol{\theta})\| = o_p(1), \tag{B.21}$$

where $\varepsilon_N(\boldsymbol{\theta}) = \sqrt{N}(\bar{g}_N(\boldsymbol{\theta}) - E(g_i(\boldsymbol{\theta})) - \bar{g}_N)$ is a centered empirical process. Note that, if the empirical process $\sqrt{N}(\bar{g}_N(\boldsymbol{\theta}) - E(g_i(\boldsymbol{\theta})))$ is stochastic equicontinuous, then (B.21) holds. However, $g_i(\boldsymbol{\theta})$ is a sum of three terms in (B.15), of which the first and second terms are free of $\boldsymbol{\theta}$. For the last term, note that $\boldsymbol{\beta}$ is bounded and the same argument in condition 4 of proof of Theorem 2.1 is sufficient to show that $\sqrt{N}(\bar{\xi}_N(\boldsymbol{\gamma}) - E(\xi_i(\boldsymbol{\gamma})))$ is stochastic equicontinuous. Thus, we obtain (B.20). Therefore, as in Seo and Shin (2016), using Theroem 7.1 of Newey and McFadden (1994) we have $\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \to_d N(\mathbf{0}, (\boldsymbol{K}'\boldsymbol{\Omega}^{-1}\boldsymbol{K})^{-1})$.

■

Proof of Theorem B.2: Under $H_0^1 : \beta_{01}^- = \beta_{01}^+ = \beta_{01}$, (B.17) can be rewritten as

$$\sqrt{N}\left(\hat{\boldsymbol{\beta}}(\boldsymbol{\gamma}) - \boldsymbol{\beta}_0(\boldsymbol{\gamma})\right) = \left[\bar{\xi}_N'(\boldsymbol{\gamma})\boldsymbol{A}_N\bar{\xi}_N(\boldsymbol{\gamma})\right]^{-1}\bar{\xi}_N'(\boldsymbol{\gamma})\boldsymbol{A}_N\sqrt{N}\bar{g}_N, \tag{B.22}$$

where $\boldsymbol{\beta}_0(\boldsymbol{\gamma}) = [\beta_{01}, \beta_{01}, \boldsymbol{\beta}_{02}' + (\boldsymbol{\gamma}_1 - \boldsymbol{\gamma}_{01})'\beta_{01}, \beta_{03}]'$. Thus, we have

$$\begin{aligned}\sqrt{N}\boldsymbol{R}_3\left(\hat{\boldsymbol{\beta}}(\boldsymbol{\gamma}) - \boldsymbol{\beta}_0(\boldsymbol{\gamma})\right) &= \sqrt{N}\left(\hat{\beta}_1^-(\boldsymbol{\gamma}) - \hat{\beta}_1^+(\boldsymbol{\gamma})\right) \\ &= \boldsymbol{R}_3\left[\bar{\xi}_N'(\boldsymbol{\gamma})\boldsymbol{A}_N\bar{\xi}_N(\boldsymbol{\gamma})\right]^{-1}\bar{\xi}_N'(\boldsymbol{\gamma})\boldsymbol{A}_N\sqrt{N}\bar{g}_N \\ &\xrightarrow{d} \boldsymbol{R}_3\left[\xi'(\boldsymbol{\gamma})\boldsymbol{\Omega}^{-1}\xi(\boldsymbol{\gamma})\right]^{-1}\xi'(\boldsymbol{\gamma})\boldsymbol{Z}_1 \\ &= \boldsymbol{R}_3\boldsymbol{V}^{-1}(\boldsymbol{\gamma})\boldsymbol{K}_{\boldsymbol{\beta}}'(\boldsymbol{\gamma})\boldsymbol{Z}_1, \quad \text{(B.23)}\end{aligned}$$

where $\boldsymbol{V}(\boldsymbol{\gamma}) = \boldsymbol{K}_{\boldsymbol{\beta}}'(\boldsymbol{\gamma})\boldsymbol{\Omega}^{-1}\boldsymbol{K}_{\boldsymbol{\beta}}(\boldsymbol{\gamma})$ and $\boldsymbol{Z}_1 \sim N(\mathbf{0}, \boldsymbol{\Omega}^{-1})$. Applying the standard WLLN and the continuous mapping theorem, we obtain

$$\begin{aligned} W_3 &= \sup_{\boldsymbol{\gamma}\in\boldsymbol{\Gamma}} \sqrt{N}\hat{\boldsymbol{\beta}}'(\gamma)\boldsymbol{R}_3\left(\boldsymbol{R}_3'\left[\hat{\boldsymbol{K}}'_{\boldsymbol{\beta}}(\gamma)\hat{\boldsymbol{\Omega}}^{-1}(\hat{\boldsymbol{\theta}}(\gamma))\hat{\boldsymbol{K}}_{\boldsymbol{\beta}}(\gamma)\right]^{-1}\boldsymbol{R}_3\right)^{-1}\boldsymbol{R}_3'\hat{\boldsymbol{\beta}}(\gamma)\sqrt{N} \\ &\xrightarrow{d} \sup_{\boldsymbol{\gamma}\in\boldsymbol{\Gamma}} \boldsymbol{Z}_1'\boldsymbol{K}_{\boldsymbol{\beta}}(\gamma)\boldsymbol{V}^{-1}(\gamma)\boldsymbol{R}_3\left[\boldsymbol{R}_3'\boldsymbol{V}^{-1}(\gamma)\boldsymbol{R}_3\right]^{-1}\boldsymbol{R}_3'\boldsymbol{V}^{-1}(\gamma)\boldsymbol{K}'_{\boldsymbol{\beta}}(\gamma)\boldsymbol{Z}_1. \end{aligned} \tag{B.24}$$

Next, given the presence of kink threshold effect, the asymptotic distribution of W_4 is χ^2_k by the normality proved in Theorem B.1. The proof is standard and we omit it to save space.

■