

Lixiong Yang*, Hujie Bai, I-Po Chen and Chingnun Lee

Maximum Likelihood Estimation of Dynamic Panel Threshold Model with a Covariate-Dependent Threshold

<https://doi.org/10.1515/snde-2025-0132>

Received September 22, 2025; accepted September 14, 2026; published online September 29, 2026

Abstract: This paper proposes a dynamic panel threshold model with a covariate-dependent and time-varying threshold (DPTCT), in which we allow for the threshold to depend on covariates that vary not only over time but also cross-sectionally. We develop a maximum likelihood estimator of the threshold and slope parameters based on the first difference transformation, and suggest test statistics for threshold effect and threshold constancy. We establish the asymptotic distribution for the threshold estimator and derive the limiting distributions of the suggested test statistics. Monte Carlo simulations are conducted to confirm the asymptotic results and show that the estimation and testing procedures work well in finite samples. The model is illustrated with an application to the inflation-growth nexus.

Keywords: dynamic panel threshold model; covariate-dependent thresholds; estimation; testing; inflation-growth nexus

JEL Classification: C12; C13; C15; C23

1 Introduction

Over the past several decades, threshold regression models have spawned a vast amount of literature in econometrics and applied econometric practice. The classical threshold models specify the sample splitting criteria based on the relative magnitude of a threshold variable and a fixed threshold value, which has been criticized for its lack of flexibility to model a time-varying sample splitting criteria; consequently, a growing number of recent studies have been devoted to extending threshold regression models to a framework with a nonconstant threshold (e.g., Yang and Su 2018; Yu and Fan 2021; Lee et al. 2021; Yang et al. 2021a, 2021b; Lee and Wang 2023). However, existing studies are limited to time-series data or non-dynamic panel data; hence, a dynamic panel threshold model with a nonconstant threshold is still unavailable. The main purpose of our paper is to fill this void in the literature.

In this paper, we introduce a new model called dynamic panel threshold model with covariate-dependent thresholds (DPTCT), in which the threshold varies not only over time but also cross-sectionally, and is modeled as a function of candidate variables that affect the separation of regimes, following the recent nonconstant threshold literature (e.g., Yang and Su 2018; Yu and Fan 2021; Lee et al. 2021; Yang et al. 2021a, 2021b). The new model can be regarded as an extension of Yu and Fan (2021) and Yang (2024), which are extensions of Hansen (2000) by

*Corresponding author: Lixiong Yang, School of Management, Lanzhou University, 222 South Tianshui Road, Lanzhou 730000, China, E-mail: ylx@lzu.edu.cn

Hujie Bai, School of Management, Lanzhou University, Lanzhou, China, E-mail: baihj21@lzu.edu.cn

I-Po Chen and Chingnun Lee, Institute of Economics, National Sun Yat-sen University, Kaohsiung, Taiwan, E-mail: j63007@gmail.com (I. Chen), lee_econ@mail.nsysu.edu.tw (C. Lee)

 Open Access. © 2026 the author(s), published by De Gruyter.  This work is licensed under the Creative Commons Attribution 4.0 International License.

allowing for a nonconstant threshold, to a dynamic panel framework with covariate-dependent thresholds. Our paper is also related to Yang et al. (2021b) who propose a panel kink threshold model with a covariate-dependent threshold, which can be treated as an extension of the kink regression model of Hansen (2017). The proposed model can also be viewed as an extension of the dynamic panel threshold model of Seo and Shin (2016) and Ramírez-Rondán (2020), who extend the static panel threshold model of Hansen (1999) to a dynamic setting, to a nonconstant threshold framework. Thus, our model is an important complement to the dynamic panel threshold models, as overlooking threshold nonconstancy can lead to biased estimates and distorted testing results, leading to misleading empirical results in applications (e.g., Yang and Su 2018; Yang et al. 2021a, 2021b).

To estimate the model, we suggest a maximum likelihood estimator following the tradition of Hsiao et al. (2002), who propose a maximum likelihood estimator (MLE) for dynamic linear panel models based on a first difference transformation. Unlike the estimation methods based on instrumental variables (IV) or the general method of moments (GMM), the maximum likelihood (ML) approach is advantageous because it does not require instruments but rather assumptions on the initial conditions. Hsiao et al. (2002) also show that the ML estimators dominate the GMM estimators in terms of the bias and root mean square error. As a consequence, Ramírez-Rondán (2020) proposes to use the ML approach to estimate the dynamic panel threshold model with a constant threshold. Compared to Ramírez-Rondán (2020), our paper has the following differences. First, the proposed model focuses on the nonconstant threshold setting, while Ramírez-Rondán (2020) focuses on a classical constant threshold setting. Second, in estimating the threshold, Ramírez-Rondán (2020) uses the grid search method, which may be infeasible in the proposed covariate-dependent threshold models due to serious computational costs, as illustrated by Yu and Fan (2021) and Lee et al. (2021). To overcome this problem, based on the first difference transformation and the maximum likelihood approach, we suggest an MCMC-based procedure to simultaneously estimate the threshold parameters and provide feasible confidence intervals for the parameters. In addition, our paper involves the issue of multiple thresholds, while existing literature does not study this problem in a dynamic panel setting. This paper also suggests test statistics for threshold effect, threshold constancy and the number of thresholds. Moreover, we derive the asymptotic properties of the proposed threshold estimator and test statistics in the large N (the number of individuals) small T (time period) framework, and show that the threshold estimator shares the convergence rate with the estimator in the constant threshold model. Hence, the inclusion of a time-varying threshold does not change the convergence rate of threshold parameters. However, it is obvious that the observations can not be correctly classified into subregimes, resulting in the true model parameters not being retrieved from the estimation method based on the constant threshold model if we misspecify a time-varying threshold as a constant one. Monte Carlo simulation results confirm our asymptotic theory and support that the estimation and testing procedures work well in finite samples. The usefulness of the model is illustrated by an application to investigate the inflation-growth nexus.

The remainder of the paper is organized as follows. Section 2 introduces the dynamic panel threshold model with a covariate-dependent threshold, suggests the estimation method, and establishes the asymptotic theory of the threshold estimator. Section 3 suggests the test statistics for threshold effect and threshold constancy, and derives the asymptotic distributions of the test statistics. Section 4 extends the model to a dynamic panel data framework with multiple covariate-dependent and time-varying thresholds. Section 5 provides an alternative estimation approach based on the generalized method of moments (GMM). Section 6 presents Monte Carlo simulations evaluating the finite-sample properties of the estimation and testing procedures. Section 7 provides an empirical application, and Section 8 concludes.

2 Dynamic Panel Threshold Models with a Covariate-Dependent Threshold

2.1 The Model

The dynamic panel threshold model with a covariate-dependent threshold we consider is given by

$$y_{it} = \alpha_i + \beta_1 y_{it-1} + \theta'_1 \mathbf{x}_{it} + (\beta_2 y_{it-1} + \theta'_2 \mathbf{x}_{it}) I(q_{it} > \gamma_{it}) + u_{it}, \quad (1)$$

for $i = 1, 2, \dots, N$ and $t = 1, 2, \dots, T$. $\mathbf{x}_{it}$ is an l_x -dimensional vector of exogenous regressors, the dependent variable y_{it} and threshold variable q_{it} are scalar, and u_{it} is the error term assumed to be independently and identically normally distributed with mean 0 and variance σ_{it}^2 . α_i is the fixed effect and represents the unobserved individual heterogeneity. $I(\cdot)$ is the indicator function. We assume that the initial values y_{i0} and $\mathbf{x}_{i0}$ are observable. γ_{it} is a covariate-dependent threshold specified as a linear combination of informative covariates $\mathbf{s}_{it}$, which can explain variation in thresholds over time and/or across individuals; that is,

$$\gamma_{it} = \boldsymbol{\gamma}' \mathbf{s}_{it}, \quad (2)$$

where $\mathbf{s}_{it} = (1, \mathbf{s}'_{1,it})' \in \mathbb{R}^{k+1}$, $\boldsymbol{\gamma} = (\gamma_0, \boldsymbol{\gamma}'_s)'$, and there may be overlap between $\mathbf{x}_{it}$ and $\mathbf{s}_{1,it}$. The threshold setting in (2) can be treated as an extension of the classical threshold model to a covariate-dependent and time-varying threshold (e.g., Dueker et al. 2013; Yang and Su 2018; Lee et al. 2021; Yu and Fan 2021; Yang et al. 2021b), and this nonconstant threshold setting is important in capturing a time-varying reference for assessing the relative magnitude of an economic variable in applications.

It is worth noting that the proposed model can be treated as an extension of the nonconstant threshold model to the dynamic setting. Specifically, the proposed dynamic model can be viewed as a direct extension of the model of Yu and Fan (2021). Moreover, the model can also be viewed as an extension of Seo and Shin's (2016) dynamic panel threshold model to a framework with a nonconstant threshold. It is also noted that, when $\boldsymbol{\gamma}_s = 0$, the model defined in (1) would degenerate to the classical dynamic panel threshold model studied by Ramírez-Rondán (2020).

2.2 Estimation

It is now well-known that the least-squares dummy variable (LSDV) estimator of the linear version of model (1) is inconsistent as N tends to infinity for a fixed T (e.g., Hsiao et al. 2002), due to the classical incidental parameters problem. To overcome this problem, following the literature including Hsiao et al. (2002) and Ramírez-Rondán (2020), we take the first difference to eliminate α_i in model (1):

$$\Delta y_{it} = \beta_1 \Delta y_{it-1} + \beta_2 \Delta y_{it-1}(\boldsymbol{\gamma}) + \theta'_1 \Delta \mathbf{x}_{it} + \theta'_2 \Delta \mathbf{x}_{it}(\boldsymbol{\gamma}) + \Delta u_{it}, \quad (3)$$

where $\Delta y_{it} = y_{it} - y_{it-1}$, $\Delta y_{it-1} = y_{it-1} - y_{it-2}$, $\Delta y_{it-1}(\boldsymbol{\gamma}) = y_{it-1} I(q_{it} > \gamma_{it}) - y_{it-2} I(q_{it-1} > \gamma_{it-1})$, $\Delta \mathbf{x}_{it} = \mathbf{x}_{it} - \mathbf{x}_{it-1}$, $\Delta \mathbf{x}_{it}(\boldsymbol{\gamma}) = \mathbf{x}_{it} I(q_{it} > \gamma_{it}) - \mathbf{x}_{it-1} I(q_{it-1} > \gamma_{it-1})$, and $\Delta u_{it} = u_{it} - u_{it-1}$.

It is important to note that for $t = 2, 3, \dots, T$, model (3) is well defined, but not for Δy_{i1} , as $y_{i,-1}$ and $\Delta \mathbf{x}_{i,1-j}$ are not available for $j = 0, 1, 2, \dots$, leading to that Δy_{i0} and $\Delta y_{i0}(\boldsymbol{\gamma})$ are missing. In the linear version of model (1) by assuming $\beta_2 = 0$ and $\theta_2 = \mathbf{0}$, starting from $\Delta y_{i,-m+1}$ (assumed given) and using continuous substitution, Hsiao et al. (2002) show that

$$\Delta y_{i1} = \beta_1^m \Delta y_{i,-m+1} + \theta'_1 \sum_{j=0}^{m-1} \beta_1^j \Delta \mathbf{x}_{i,1-j} + \sum_{j=0}^{m-1} \beta_1^j \Delta u_{i,1-j}, \quad i = 1, \dots, N. \quad (4)$$

As can be seen from (4), observations back to the initial period are not available. Thus, for $i = 1, \dots, N$, the mean of Δy_{i1} conditional on $\Delta y_{i,-m+1}$ and $\Delta \mathbf{x}_{i,1-j}$ for $j = 0, 1, 2, \dots$ is defined by

$$\eta_{i1} = E(\Delta y_{i1} | \Delta y_{i,-m+1}, \Delta \mathbf{x}_{i1}, \Delta \mathbf{x}_{i0}, \dots) = \beta_1^m \Delta y_{i,-m+1} + \theta'_1 \sum_{j=0}^{m-1} \beta_1^j \Delta \mathbf{x}_{i,1-j}. \quad (5)$$

As shown in (5), when $\theta_1 = 0$ and T is fixed, the MLE of the dynamic panel linear model depends on the assumption on the initial conditions, which may affect the consistency of the MLE estimator, as the covariance

matrix depends on the initial conditions.¹ When $\theta_1 \neq 0$, the mean η_{i1} is unknown even if $m \rightarrow \infty$. Treating η_{i1} as a parameter, the classical incidental parameters problem would appear. To address the above problem, we follow Hsiao et al. (2002) to assume the process has started from a finite period in the past (m is finite), and $E(\Delta y_{i-m+1} | \Delta \mathbf{x}_{i1}, \Delta \mathbf{x}_{i2}, \dots, \Delta \mathbf{x}_{iT})$ is the same across all individuals. Then, under the exogeneity assumption of $\mathbf{x}_{it}$ we have

$$\Delta y_{i1} = \delta_0 + \delta'_1 \Delta \mathbf{x}_{i1} + v_{i1}, \quad (6)$$

where $E(v_{i1} | \Delta \mathbf{x}_{i1}) = 0$, $E(v_{i1}^2) = \sigma_v^2$, $E(v_{i1} \Delta u_{i2}) = -\sigma_u^2$, and $E(v_{i1} \Delta u_{it}) = 0$ for $t = 3, 4, \dots, T$ and $i = 1, \dots, N$.

Although the derivation from (4)–(6) is based on the linear model of Hsiao et al. (2002), Ramírez-Rondán (2020) argues that similar results hold for dynamic panel threshold models, as the threshold model would become linear as the sample size increases under the “small threshold effect” assumption, which is widely used in the threshold literature (e.g., Hansen 2000). Based on a similar argument, we extend the maximum likelihood approach to estimate the proposed dynamic panel threshold model with a covariate-dependent threshold.

To develop a maximum likelihood approach for model estimation, we derive the likelihood function for any given threshold parameter γ . Define $\Delta u_i = (v_{i1}, \Delta u_{i2}, \dots, \Delta u_{iT})$ and $\Delta y_i = (\Delta y_{i1}, \Delta y_{i2}, \dots, \Delta y_{iT})$. Following Hsiao et al. (2002) and Ramírez-Rondán (2020), the joint probability distribution functions of Δu_i and Δy_i are the same, as the Jacobian of the transformation from Δu_i to Δy_i is unity. Thus, the covariance matrix of Δu_i takes the following form

$$\mathbf{\Omega} = \sigma_u^2 \begin{bmatrix} \omega & -1 & 0 & \dots & 0 \\ -1 & 2 & -1 & & \\ 0 & -1 & 2 & & \\ \vdots & & & \ddots & -1 \\ 0 & & & -1 & 2 \end{bmatrix} = \sigma_u^2 \mathbf{\Omega}^*,$$

where $\omega = \sigma_v^2 / \sigma_u^2$. Assume that u_{it} is independent normal. Then, the joint probability distribution of Δy_i conditional on x_i is therefore given by

$$L(\delta_0, \delta'_1, \beta', \theta', \gamma', \sigma_u^2, \omega) = \prod_{i=1}^N (2\pi)^{-T/2} |\mathbf{\Omega}|^{-1/2} \exp \left\{ -\frac{1}{2} \Delta u_i(\delta_0, \delta'_1, \beta', \theta', \gamma)' \mathbf{\Omega}^{-1} \Delta u_i(\delta_0, \delta'_1, \beta', \theta', \gamma') \right\}, \quad (7)$$

where $\Delta u_i(\delta_0, \delta'_1, \beta', \theta', \gamma) = [\Delta y_{i1} - \delta_0 - \delta'_1 \Delta \mathbf{x}_{i1}, \Delta y_{i2} - \beta_1 \Delta y_{i1} - \beta_2 \Delta y_{i1}(\gamma) - \theta'_1 \Delta \mathbf{x}_{i2} - \theta'_2 \Delta \mathbf{x}_{i2}(\gamma), \dots, \Delta y_{iT} - \beta_1 \Delta y_{iT-1} - \beta_2 \Delta y_{iT-1}(\gamma) - \theta'_1 \Delta \mathbf{x}_{iT} - \theta'_2 \Delta \mathbf{x}_{iT}(\gamma)]'$, $\beta' = (\beta_1, \beta_2)$, and $\theta' = (\theta'_1, \theta'_2)$.

Taking a logarithmic transformation for the likelihood function in (7), we obtain

$$\begin{aligned} \ln L(\delta_0, \delta'_1, \beta', \theta', \gamma', \sigma_u^2, \omega) &= -\frac{NT}{2} \ln(2\pi) - \frac{N}{2} \ln |\mathbf{\Omega}| \\ &\quad - \frac{1}{2} \sum_{i=1}^N \Delta u_i(\delta_0, \delta'_1, \beta', \theta', \gamma)' \mathbf{\Omega}^{-1} \Delta u_i(\delta_0, \delta'_1, \beta', \theta', \gamma'). \end{aligned} \quad (8)$$

As shown in Hsiao et al. (2002), $|\mathbf{\Omega}| = \sigma_u^{2T} [1 + T(\omega - 1)]$, and thus, the log-likelihood criterion in (8) depends on the unknown ω . To ensure that $|\mathbf{\Omega}|$ is positive, we assume $\omega > 1 - 1/T$. In threshold models, it is now well-understood that the threshold parameter can lead to that the objective function is non-convex. For any given threshold parameter γ , the threshold model would degenerate to the dynamic panel linear model of Hsiao et al. (2002), who show that the maximum likelihood estimator (MLE) in the linear case is consistent and asymptotically normal when $N \rightarrow \infty$, regardless of whether T tends to infinity or is fixed. Therefore, we suggest to use a two-step estimation procedure. First, we employ the maximum likelihood approach to estimate the parameters

¹ This problem disappears when $T \rightarrow \infty$.

$\delta_0, \delta'_1, \beta', \theta', \sigma_u^2$ and ω for a given γ . Second, we estimate the threshold parameter by plugging the MLE for a given $\gamma \in \Gamma$, which is the parameter space of γ .

Step 1: Estimation of $\delta_0, \delta'_1, \beta', \theta', \sigma_u^2, \omega$ for a given γ

Denote the matrix $\Delta y_{i,-1}(\gamma)$ as follows

$$\Delta y_{i,-1}(\gamma) = \begin{bmatrix} 1 & \Delta \mathbf{x}'_{i1} & 0 & 0 & 0 & 0 \\ 0 & 0 & \Delta y_{i1} & \Delta \mathbf{x}_{i2} & \Delta y_{i1}(\gamma) & \Delta \mathbf{x}_{i2}(\gamma) \\ 0 & 0 & \Delta y_{i2} & \Delta \mathbf{x}_{i3} & \Delta y_{i2}(\gamma) & \Delta \mathbf{x}_{i3}(\gamma) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \Delta y_{iT-1} & \Delta \mathbf{x}_{iT} & \Delta y_{iT-1}(\gamma) & \Delta \mathbf{x}_{iT}(\gamma) \end{bmatrix}.$$

For a fixed γ , we can compute the first-order derivative with respect to $\theta_\delta(\gamma) = (\delta_0, \delta'_1, \beta', \theta')'$ and set the partial derivative as zero, yielding to the estimator $\hat{\theta}_\delta(\gamma)$, and the residuals

$$\Delta \hat{u}_i(\gamma) = \Delta y_i - \Delta y_{i,-1}(\gamma) \hat{\theta}_\delta(\gamma). \quad (9)$$

Similarly, we can obtain the MLE estimators of σ_u^2 and ω for a given γ as follows

$$\hat{\sigma}_u^2(\gamma) = \frac{1}{NT} \sum_{i=1}^N [\Delta \hat{u}_i(\gamma)' \hat{\Omega}^*(\gamma)^{-1} \Delta \hat{u}_i(\gamma)], \quad (10)$$

and

$$\hat{\omega}(\gamma) = \frac{T}{T-1} + \frac{1}{\hat{\sigma}_u^2 NT^2} \sum_{i=1}^N [\Delta \hat{u}_i(\gamma)' \Phi \Delta \hat{u}_i(\gamma)], \quad (11)$$

in which

$$\Phi = \begin{bmatrix} T^2 & T(T-1) & T(T-2) & \dots & T \\ T(T-1) & (T-1)^2 & (T-1)(T-2) & \dots & (T-1) \\ \vdots & \vdots & \vdots & \dots & \vdots \\ T & (T-1) & (T-2) & \dots & 1 \end{bmatrix}.$$

Then, the log-likelihood function (8) can be written as

$$\ln L(\theta_\delta(\gamma), \gamma, \sigma_u^2, \omega) = -\frac{NT}{2} \ln(2\pi) - \frac{N}{2} \ln |\Omega| - \frac{1}{2} \sum_{i=1}^N [\Delta u_i(\gamma)' \Omega^{-1} \Delta u_i(\gamma)]. \quad (12)$$

As illustrated by Ramírez-Rondán (2020), the slope estimator $\hat{\theta}_\delta(\gamma)$ depends on σ_u^2 and ω , while σ_u^2 and ω depend on the slope parameters. Consequently, initial estimates are needed to efficiently implement the above maximum likelihood approach. In this paper, we suggest an iterative method, which goes as follows. We first obtain initial estimates of (β', θ') by an instrumental variable estimation (Anderson and Hsiao 1982). Specifically, let

$$\Delta \ddot{y}_i = \begin{bmatrix} \Delta y_{i3} \\ \Delta y_{i4} \\ \vdots \\ \Delta y_{iT} \end{bmatrix}, \quad \Delta \ddot{y}_{i,-1}(\gamma) = \begin{bmatrix} \Delta y_{i2} & \Delta x_{i3} & \Delta y_{i2}(\gamma) & \Delta x_{i3}(\gamma) \\ \Delta y_{i3} & \Delta x_{i4} & \Delta y_{i3}(\gamma) & \Delta x_{i4}(\gamma) \\ \vdots & \vdots & \vdots & \vdots \\ \Delta y_{iT-1} & \Delta x_{iT} & \Delta y_{iT-1}(\gamma) & \Delta x_{iT}(\gamma) \end{bmatrix}$$

and

$$\Delta \ddot{y}_{i,-2}(\gamma) = \begin{bmatrix} \Delta y_{i1} & \Delta x_{i3} & \Delta y_{i1}(\gamma) & \Delta x_{i3}(\gamma) \\ \Delta y_{i2} & \Delta x_{i4} & \Delta y_{i2}(\gamma) & \Delta x_{i4}(\gamma) \\ \vdots & \vdots & \vdots & \vdots \\ \Delta y_{iT-2} & \Delta x_{iT} & \Delta y_{iT-2}(\gamma) & \Delta x_{iT}(\gamma) \end{bmatrix}.$$

Thus, for any given $\gamma \in \Gamma$, the initial estimates of the slope parameters β and θ using instrumental variables are

$$(\hat{\beta}', \hat{\theta}')' = \left(\sum_{i=1}^n \Delta \ddot{y}_{i,-2}(\gamma)' \Delta \ddot{y}_{i,-1}(\gamma) \right)^{-1} \left(\sum_{i=1}^n \Delta \ddot{y}_{i,-2}(\gamma)' \Delta \ddot{y}_i \right).$$

Then, the initial estimates $\hat{\sigma}_u^2(\gamma)$ and $\hat{\omega}(\gamma)$ can be obtained by plugging the initial slope estimates into equations (10) and (11), respectively. The initial estimates of the external parameters (δ_0, δ_1') can be given by a least square regression of Δy_{it} on a constant and Δx_{it} . Finally, using these initial estimates, we can obtain the MLE by employing an iterative technique such as the Newton–Raphson procedure.

Step 2: Estimation of the threshold parameter γ

To obtain the threshold parameter γ , we plug the MLE of $(\delta_0(\gamma), \delta_1'(\gamma), \beta'(\gamma), \theta'(\gamma), \sigma_u^2(\gamma), \omega(\gamma))$ in step 1 into (8). Then, the threshold parameter γ can be estimated as

$$\hat{\gamma} = \arg \max_{\gamma \in \Gamma} \ln L(\gamma), \quad (13)$$

in which $\ln L(\gamma) = -\frac{NT}{2} \ln(2\pi) - \frac{N}{2} \ln |\hat{\Omega}| - \frac{1}{2} \sum_{i=1}^N \Delta \hat{u}_i(\gamma)' \hat{\Omega}^{-1} \Delta \hat{u}_i(\gamma)$.

As illustrated by Ramírez-Rondán (2020), because the loss function is not smooth, it is not suitable to use conventional gradient algorithms in its maximization. Due to the non-convexity in the criterion function (13), a typically used approach for the optimization in threshold models is the grid search method; however, this method may be infeasible, especially when the dimension of s_{it} is large, as the inclusion of one more covariate in s_{it} would lead to the computational cost increasing $N \times T$ times. This problem would cause a further deterioration of computational issues if we follow the threshold literature to compute confidence intervals (CI) based on a bootstrap procedure. We overcome this problem by suggesting a Markov Chain Monte Carlo (MCMC) algorithm to implement the maximization. The suggested algorithm can simultaneously estimate the threshold parameters and provide feasible confidence intervals for the parameters.

Algorithm A. *Estimation based on the MCMC technique.*

Step 1. Construct the quasi-posterior of γ with a uniform prior on Γ as follows

$$p(\gamma) = \frac{\exp\{L(\gamma)\} 1(\gamma \in \Gamma)}{\int_{\Gamma} \exp\{L(\gamma)\} d\gamma}.$$

Step 2. Use the MCMC method to draw a Markov chain from the quasi-posterior

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

Step 3. For each $\gamma^{(b)}$, $b = 1, 2, \dots, B$, calculate $\ln L(\gamma^{(b)})$, and obtain the threshold estimates as $\hat{\gamma} = \arg \max_{\gamma \in S} \ln L(\gamma^{(b)})$. When $\hat{\gamma}$ is a set of γ values, we define the threshold estimates as the mean of the points $\gamma^{(b)} \in S$ in the set.²

Step 4. Denote the j -th component of S as $S_j := (\gamma_j^{(1)}, \gamma_j^{(2)}, \dots, \gamma_j^{(B)})$. Then, the $(1 - \alpha)$ CI for γ_j is constructed as $[\gamma_{j(\alpha/2)}, \gamma_{j(1-\alpha/2)}]$ with $\gamma_{j(\tau)}$ being the τ -th order statistic of S_j .

This algorithm is most closely related to Yu and Fan (2021) who suggest a similar MCMC algorithm in the classical cross-sectional framework. Our MCMC algorithm extends Yu and Fan (2021) to a maximum likelihood framework in dynamic panel modelling. It is worth mentioning that the MCMC technique is only auxiliary to the maximization by simulating the candidate maximizers in (13). As explained by Yu and Fan (2021), the MCMC-based algorithm is more efficient than the grid search method which may be infeasible when the dimension of s_{it}

² If desired, one can update the simulation set in Step 1 from Γ to a neighborhood of the estimates of $\hat{\gamma}_j$ in Step 3, and repeat Steps 2 and 3 to obtain an updated estimate. In the simulations, we find that such an update procedure does not significantly increase the accuracy of the estimates in most cases; hence, we do not update the estimates to speed up computations in practice.

is large. An important advantage of our method is that it can simultaneously estimate the threshold parameters and provide feasible confidence intervals for the parameters, which can significantly save computational cost as the confidence interval is actually a by-product of the model estimation. For a general econometric theory for the consistency of the MCMC-based estimator and the validity of the confidence intervals, one can refer to Chernozhukov and Hong (2004). As emphasized by Yu and Fan (2021), the MCMC-based algorithm can achieve the global minimizer when B is sufficiently large; however, there is no general approach for selecting B (e.g., Li et al. 2018; Gallant 2023; Cross et al. 2023; Norets and Shimizu 2024). In applications, we thus suggest to choose B by employing the convergence diagnostic of Gelman and Rubin (1992), who provide a statistic R to monitor convergence of MCMC samples. The R statistic substantially above 1 provides the evidence of the lack of convergence.

Another important issue in applications is to specify the parameter space of $\gamma = (\gamma_0, \gamma_s')$. Denote parameter spaces of γ_0 and γ_s' as Γ_0 and Γ_s , respectively. Then, we suggest to set $\Gamma_0 = \Gamma_0(\gamma_0) \subseteq [q_{(\eta)NT}, q_{(1-\eta)NT}]$ to ensure that each regime contains at least η portion of data in which η is often set as 0.05, 0.10 or 0.15, and set $\Gamma_s = \bigotimes_{l=1}^k [\hat{\gamma}_l^L - C, \hat{\gamma}_l^L + C]$ for a large C so that $[\hat{\gamma}_l^L - C, \hat{\gamma}_l^L + C]$ can contain zero for each l , in which $\hat{\gamma}_l^L$ denotes the least square estimator of q_{it} on $\mathbf{s}_{it}$ (Yu and Fan 2021). When the dimension of γ_s is zero, the parameter space Γ_0 would be the same as in the classical constant threshold model (e.g., Hansen 1999; Ramírez-Rondán 2020).

2.3 Asymptotic Theory

In this section, we establish the asymptotic theory of the suggested estimator. Let $f_t(q_{it} | \mathbf{s}_{it})$ be the conditional distribution of the threshold variable q_{it} given $\mathbf{s}_{it}$ and $f_{k|t,s}(q_{ik} | q_{it}, \mathbf{s}_{ik}, \mathbf{s}_{it})$ be the conditional distribution of the threshold variable q_{ik} given q_{it} , $\mathbf{s}_{ik}$, and $\mathbf{s}_{it}$. Let

$$\Delta y_{i,-1}^* = \begin{bmatrix} 0 & 0 \\ \Delta y_{i1} & \Delta \mathbf{x}_{i2}' \\ \vdots & \vdots \\ \Delta y_{iT-1} & \Delta \mathbf{x}_{iT}' \end{bmatrix}, \quad \Delta y_{i,-1}^*(\gamma) = \begin{bmatrix} 0 & 0 \\ \Delta y_{i1}(\gamma) & \Delta \mathbf{x}_{i2}(\gamma)' \\ \vdots & \vdots \\ \Delta y_{iT-1}(\gamma) & \Delta \mathbf{x}_{iT}(\gamma)' \end{bmatrix}. \quad (14)$$

Define the moment functionals

$$\mathbf{M} = E(\Delta y_{i,-1}^{*'} \Omega_0^{*-1} \Delta y_{i,-1}^*), \quad (15)$$

$$\mathbf{M}(\gamma) = E(\Delta y_{i,-1}^{*'}(\gamma) \Omega_0^{*-1} \Delta y_{i,-1}^*(\gamma)), \quad (16)$$

$$D_{ij}(\gamma) = E(w_{ij}^2 | q_{ij+1} = \mathbf{s}_{ij+1}'\gamma, \mathbf{s}_{ij+1}), \quad (17)$$

$$V_{ijtk}(\gamma) = E(w_{ij}^2 e_{it} e_{ik} | q_{ij+1} = \mathbf{s}_{ij+1}'\gamma, \mathbf{s}_{ij+1}), \quad (18)$$

where $w_{ij} = y_{ij}c_{\beta_2} + \mathbf{x}_{ij+1}'c_{\theta_2}$ (c_{β_2} , c_{θ_2} define in Assumption 1), $e_{it} = v_{i1} + \sum_{j=2}^t a_{j-1} \Delta u_{ij}$, and $a_t = 1 + t(\omega_0 - 1)$. (ω_0, Ω_0^*) is the true value of (ω, Ω^*) .

2.3.1 Assumptions

To obtain the asymptotic properties of the estimator of the threshold $\hat{\gamma}$, we introduce the following assumptions.

Assumption 1. Assume that

1. y_{i0} and $\mathbf{x}_{i0}$ are observable.
2. For each t , $(\mathbf{x}_{it}, q_{it}, \mathbf{s}_{it}, u_{it}, v_{i1})$ are independent and identically distributed (iid) across i .
3. For each i , u_{it} is iid over t , is independent of $\{(y_{ij})_{j=1}^{t-1}, (\mathbf{x}_{it}, q_{it}, \mathbf{s}_{it})_{t=1}^T\}$, and v_{i1} is independent of $\{(y_{i0}), (\mathbf{x}_{it}, q_{it}, \mathbf{s}_{it})_{t=1}^T\}$.

4. u_{it} normally distributed with zero mean and variance σ_u^2 , v_{it} normally distributed with zero mean and variance σ_v^2 and $\omega > 1 - 1/T$.
5. $E|y_{i0}|^4 < \infty$ and $E\|\mathbf{x}_{it}\|^4 < \infty$.
6. For any $\boldsymbol{\gamma} \in \boldsymbol{\Gamma}$ and $t = 1, \dots, T$, $E(\|\mathbf{x}_{it}\|^4 | q_{it} = \mathbf{s}_{it}'\boldsymbol{\gamma}, \mathbf{s}_{it}) \leq C$ for some $C < \infty$, and $0 < f_t(q_{it} | \mathbf{s}_{it}) \leq \bar{f} < \infty$. For $k > t$, $f_{k|t,s}(\mathbf{s}_{ik}'\boldsymbol{\gamma}_0 | q_{it} = \mathbf{s}_{it}'\boldsymbol{\gamma}_0, \mathbf{s}_{ik}, \mathbf{s}_{it}) < \infty$.
7. $f_t(\mathbf{s}_{it}'\boldsymbol{\gamma} | \mathbf{s}_{it})$, $D_{ij}(\boldsymbol{\gamma})$ and $V_{ijk}(\boldsymbol{\gamma})$ are continuous at $\boldsymbol{\gamma} = \boldsymbol{\gamma}_0$, where $\boldsymbol{\gamma}_0$ is the true value of $\boldsymbol{\gamma}$.
8. $\mathbf{s}_{it}$ is not multicollinear. $E(\nabla \Delta y_{i-1}^*(\boldsymbol{\gamma})' \boldsymbol{\Omega}_0^{*-1} \nabla \Delta y_{i-1}^*(\boldsymbol{\gamma})) = \mathbf{0}$ and
9. $E(\nabla \Delta y_{i-1}^*(\boldsymbol{\gamma})' \boldsymbol{\Omega}_0^{*-1} \Delta y_{i-1}(\boldsymbol{\gamma})) = \mathbf{0}$ if and only if $\boldsymbol{\gamma} = \boldsymbol{\gamma}_0$, where $\nabla \Delta y_{i-1}^*(\boldsymbol{\gamma}) = \Delta y_{i-1}^*(\boldsymbol{\gamma}) - \Delta y_{i-1}^*(\boldsymbol{\gamma}_0)$.
10. $\beta_{20} = c_{\beta_2} N^{-\alpha}$ and $\theta_{20} = c_{\theta_2} N^{-\alpha}$ with $\mathbf{c} = (c_{\beta_2}, c_{\theta_2})' \neq \mathbf{0}$ and $0 < \alpha < 1/2$, where $\mathbf{c}$ is a vector of constant and $(\beta_{20}, \theta_{20})'$ is the true value of $(\beta_2, \theta_2)'$.
11. $G(\mathbf{v}) > 0$ and $V(\mathbf{v}) > 0$ with $\mathbf{v} \neq \mathbf{0}$ (define in Theorem 1).
12. $\det(\mathbf{M}) > \det(\mathbf{M}(\boldsymbol{\gamma})) > 0$ for all $\boldsymbol{\gamma} \in \boldsymbol{\Gamma}$.
13. $(\theta'_{\delta 0}, \sigma_{u0}^2, \omega_0)' \in \text{int}(\boldsymbol{\Theta})$, where $(\theta'_{\delta 0}, \sigma_{u0}^2, \omega_0)'$ is the true value of $(\theta'_{\delta}, \sigma_u^2, \omega)'$ and $\boldsymbol{\Theta}$ is the parameter space of $(\theta'_{\delta}, \sigma_u^2, \omega)'$. $\square$

Assumptions 1.1–1.5 are natural extensions of Assumptions 1.1–1.5 in Ramírez-Rondán (2020). Assumptions 1.1–1.4 are also similar to the assumptions of Hsiao et al. (2002) for the dynamic fixed effect panel linear model with exogenous regressors. Assumption 1.2 is standard in panel data models. Assumptions 1.3–1.4 require the model to have exogenous regressors. Assumptions 1.5–1.6 restrict unconditional and conditional moment bounds to be finite. Assumption 1.7 requires the threshold variable to have a continuous distribution. Assumption 1.8 ensures that the asymptotic distribution of $\hat{\boldsymbol{\gamma}}$ is well-defined and implies that the limit of $\ln L(\boldsymbol{\gamma})$ has a unique maximum at the true value $\boldsymbol{\gamma}_0$. Assumption 1.9 is well-known as the small threshold effect assumption, which is commonly used in the literature for threshold models (e.g., Hansen 2000). Assumptions 1.10–1.11 are full rank conditions needed to have nondegenerate asymptotic distributions. Assumption 1.12 ensures that condition (i) of Theorem 10.1 in Wooldridge (1994) is satisfied.

2.3.2 Asymptotic Distribution

Theorem 1. Under Assumption 1, as $N \rightarrow \infty$, we have

$$N^{1-2\alpha}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) \xrightarrow{d} \arg \min_{\mathbf{v} \in \mathbb{R}^{k+1}} \left[\frac{1}{2} G(\mathbf{v}) - R(\mathbf{v}) \right], \quad (19)$$

where $R(\mathbf{v})$ is a Gaussian process with a positive variance $V(\mathbf{v})$ when $\mathbf{v} \neq \mathbf{0}$. $G(\mathbf{v})$ and $V(\mathbf{v})$ are defined as

$$G(\mathbf{v}) = \sum_{t=2}^T [a_t a_{t-1}]^{-1} D_{1t}(\mathbf{v}),$$

$$V(\mathbf{v}) = \sum_{t=2}^T [a_t a_{t-1}]^{-1} V_{1t}(\mathbf{v}) + 2 \sum_{t=2}^{T-1} \sum_{k=t+1}^T [a_t a_{t-1} a_k a_{k-1}]^{-1} V_{2tk}(\mathbf{v}),$$

are positive when $\mathbf{v} \neq \mathbf{0}$, where $a_0 = 1$, $a_1 = \omega$, $a_{t+1} - 2a_t + a_{t-1} = 0$ for $t = 1, 2, \dots, T$. $a_0, \dots, a_T$ are constants of the matrix $\mathbf{H}$ defined in the Appendix A.

$$D_{1t}(\mathbf{v}) = \sum_{j=1}^{t-1} a_j^2 \left[E(D_{ij}^f(\mathbf{v})) + E(D_{ij-1}^f(\mathbf{v})) \right] - 2 \sum_{j=1}^{t-2} a_j a_{j+1} E(D_{ij}^f(\mathbf{v})),$$

$$V_{1t}(\mathbf{v}) = \sum_{j=1}^{t-1} a_j^2 \left[E(V_{ijt}^f(\mathbf{v})) + E(V_{ij-1tt}^f(\mathbf{v})) \right] - 2 \sum_{j=1}^{t-2} a_j a_{j+1} E(V_{ijt}^f(\mathbf{v})),$$

$$V_{2tk}(\mathbf{v}) = \sum_{j=1}^{t-1} a_j^2 \left[E(V_{ijt}^f(\mathbf{v})) + E(V_{ij-1tk}^f(\mathbf{v})) \right] - 2 \sum_{j=1}^{t-2} a_j a_{j+1} E(V_{ijt}^f(\mathbf{v})) \\ - a_t a_{t-1} E(V_{it-1tk}^f(\mathbf{v})),$$

in which

$$D_{ij}^f(\mathbf{v}) = D_{ij}(\gamma_0) f_{j+1|s}(\mathbf{s}'_{ij+1} \gamma_0 | \mathbf{s}_{ij+1}) |\mathbf{s}'_{ij+1} \mathbf{v}|, \\ V_{ijt}^f(\mathbf{v}) = V_{ijt}(\gamma_0) f_{j+1|s}(\mathbf{s}'_{ij+1} \gamma_0 | \mathbf{s}_{ij+1}) |\mathbf{s}'_{ij+1} \mathbf{v}|.$$

Proof of Theorem 1. See the Appendix A. □

As can be seen from Theorem 1, the asymptotic distribution of the threshold estimator is not related to the two-sided Brownian motion as in the classical threshold literature (e.g., Hansen 2000), due to the involvement of a covariate-dependent threshold as illustrated by Yu and Fan (2021). Our asymptotic distribution shares a similar form with Yu and Fan (2021) who derive the asymptotic distribution of the threshold estimator in a time-series framework with a covariate-dependent threshold. Compared with the threshold literature, Theorem 1 indicates that the inclusion of a covariate-dependent threshold does not change the convergence rate of the threshold estimator, which further implies that the distribution of the maximum likelihood slope estimator can also be approximated by the conventional normal approximation as if γ were known with certainty in the dynamic panel threshold model with a covariate-dependent threshold, as confirmed by Lemma 9 in the Appendix A.

3 Model Specification Testing

In applications, it is often useful to test whether there is a threshold effect and whether the threshold is constant. In this section, we suggest two test statistics for threshold effect and threshold constancy in the dynamic panel threshold model with a covariate-dependent threshold.

To test whether there is a threshold effect, we consider the null hypothesis of linearity $H_0^1: \beta_2 = 0$ and $\theta_2 = \mathbf{0}$. Under the null, the model degenerates to the dynamic panel linear model of Hsiao et al. (2002):

$$y_{it} = \alpha_i + \beta y_{it-1} + \theta_1' \mathbf{x}_{it} + u_{it}. \quad (20)$$

The linear model can be estimated using a maximum likelihood estimation procedure as discussed in Section 2. By taking the first difference transformation, we can obtain $\Delta y_{it} = \beta \Delta y_{it-1} + \theta_1' \Delta \mathbf{x}_{it} + \Delta u_{it}$, and we assume that the model in the first period is $\Delta y_{i1} = \delta_0 + \delta_1' \Delta \mathbf{x}_{i1} + v_{i1}$. Define $S(\hat{\gamma}) = -\ln L(\hat{\theta}_\delta(\hat{\gamma}), \hat{\gamma}, \hat{\sigma}_u^2, \hat{\omega})$, which is the negative log-likelihood function defined as in (12). Similarly, we denote the negative log-likelihood function of the linear dynamic model as $S_0 = -\ln L(\hat{\theta}_\delta^l, \hat{\sigma}_u^2, \hat{\omega}^l)$. Then, the likelihood test statistics of H_0 can be defined as

$$F_1 = \frac{S_0 - S(\hat{\gamma})}{S(\hat{\gamma})/N(T-1)} = \sup_{\gamma \in \Gamma} \frac{S_0 - S(\hat{\gamma})}{S(\hat{\gamma})/N(T-1)}. \quad (21)$$

Rejection of the null hypothesis provides the evidence supporting the existence of threshold effect. Given the existence of threshold effect, it is desired to test whether the threshold is constant. To this end, we consider the null hypothesis of constant threshold given as $H_0^2: \gamma_s = \mathbf{0}$. Under the null of constant threshold, the model will degenerate to the dynamic panel threshold model with a constant threshold of Ramírez-Rondán (2020). Denote the constant threshold model as

$$y_{it} = \alpha_i + \beta_1 y_{it-1} + \theta_1' \mathbf{x}_{it} + (\beta_2 y_{it-1} + \theta_2' \mathbf{x}_{it}) I(q_{it} > \gamma_0) + u_{it}. \quad (22)$$

Again, the constant threshold model can be estimated using a maximum likelihood approach as suggested in Ramírez-Rondán (2020). Denote the negative log-likelihood function of the constant threshold as $S_c = -\ln L(\hat{\theta}_\delta(\hat{\gamma}_0), \hat{\gamma}_0, \hat{\sigma}_u^2, \hat{\omega}^2)$. Then, the likelihood ratio test statistic for threshold constancy can be constructed as

$$F_c = \frac{S_c - S(\hat{\gamma})}{S(\hat{\gamma})/N(T-1)}. \quad (23)$$

We next establish the asymptotic distributions of the above test statistics for threshold effect and threshold constancy.

Theorem 2. Suppose Assumption 1 holds. Under $H_0^1: \beta_2 = 0$ and $\theta_2 = 0$, we have

$$F_1 \xrightarrow{d} \frac{1}{\sigma_T^2} \sup_{\gamma \in \Gamma} K(\gamma)' H(\gamma)^{-1} R' [RH(\gamma)^{-1} R']^{-1} RH(\gamma)^{-1} K(\gamma), \quad (24)$$

under $H_0^2: \gamma_s = 0$, we have

$$F_c \xrightarrow{d} \frac{-2}{\sigma_T^2} \left[\left(\min_{v \in \mathbb{R}^{d+1}} G(v) - 2R(v) \right) - \left(\min_{v \in \mathbb{R}^{d+1}} G(v) - 2R(v) \right) \right], \quad (25)$$

where $\sigma_T^2 = [T \ln(2\pi) + \ln |\Omega_0| + T]/(T-1)$, $\Omega\Omega_0$ is the true value of $\Omega\Omega$, R is a matrix such that $R(\theta'_\delta, \sigma_u^2, \omega)' = (\beta_2, \theta_2)' = 0$, $\mathbb{R}^{k+1} = \{v \in \mathbb{R}^{k+1}: v_s = 0\}$. $G(v)$ and $R(v)$ are defined in Theorem 1. $K(\gamma)$ is a zero-mean Gaussian process with covariance kernel $E(h_i(\gamma)h_i(\gamma)')$, in which $h_i(\gamma) = (h_{1i}(\gamma)', h_{2i}, h_{3i})'$, $h_{1i}(\gamma) = \Delta y_{i,-1}(\gamma)' \Omega_0^{-1} \Delta u_i$, $h_{2i} = (\Delta u_i' \Omega_0^{\star-1} \Delta u_i - T)/(2\sigma_{u0}^2)$, $h_{3i} = (\Delta u_i' \Phi \Delta u_i - T\sigma_{u0}^2 \varpi_0)/(2\sigma_{u0}^2 \varpi_0^2)$, $\varpi_0 = 1 + T(\omega_0 - 1)$. $H(\gamma)$ is defined as follows

$$H(\gamma) = \begin{bmatrix} H_{\theta_\delta \theta'_\delta}(\gamma) & H_{\theta_\delta \sigma_u^2}(\gamma) & H_{\theta_\delta \omega}(\gamma) \\ H_{\theta_\delta \sigma_u^2}(\gamma)' & H_{\sigma_u^2 \sigma_u^2} & H_{\sigma_u^2 \omega} \\ H_{\theta_\delta \omega}(\gamma)' & H_{\sigma_u^2 \omega} & H_{\omega \omega} \end{bmatrix},$$

where

$$H_{\theta_\delta \theta'_\delta}(\gamma) = E(\Delta y_{i,-1}(\gamma)' \Omega_0^{-1} \Delta y_{i,-1}(\gamma)),$$

$$H_{\theta_\delta \sigma_u^2}(\gamma) = \frac{1}{\sigma_{u0}^4} E(\Delta y_{i,-1}(\gamma)' \Omega_0^{\star-1} \Delta u_i),$$

$$H_{\theta_\delta \omega}(\gamma) = \frac{1}{\sigma_{u0}^2 \varpi_0^2} E(\Delta y_{i,-1}(\gamma)' \Phi \Delta u_i),$$

$$H_{\sigma_u^2 \sigma_u^2} = \frac{1}{\sigma_{u0}^6} E(\Delta u_i' \Phi \Delta u_i) - \frac{T}{2\sigma_{u0}^4},$$

$$H_{\sigma_u^2 \omega} = \frac{1}{\sigma_{u0}^4 \varpi_0^2} E(\Delta u_i' \Phi \Delta u_i),$$

$$H_{\omega \omega} = \frac{1}{\sigma_{u0}^3 \varpi_0^3} E(\Delta u_i' \Phi \Delta u_i) - \frac{T^2}{2\varpi_0^2}.$$

Proof of Theorem 2. See the Appendix A. □

According to the well-known Davies problem (Davies 1977, 1987), threshold models are not identified under the null of linearity, leading to the distribution of the test statistic for threshold effect being nonstandard. Thus,

the threshold literature often employs a bootstrap method to calculate the p -values or critical values of test statistics. We follow this tradition to suggest using a parametric bootstrap, which goes as follows.

Algorithm B. *Testing for threshold effect and threshold constancy*

Step 1. Use the original sample $(y_{it}, \mathbf{x}'_{it}, \mathbf{z}'_{it}, q_{it}, \mathbf{s}'_{it})'$'s to estimate the linear dynamic panel model $y_{it} = \alpha_i + \beta_1 y_{it-1} + \theta'_1 \mathbf{x}_{it} + u_{it}$, and the dynamic panel threshold model with a constant threshold $y_{it} = \alpha_i + \beta_1 y_{it-1} + \theta'_1 \mathbf{x}_{it} + (\beta_2 y_{it-1} + \theta'_2 \mathbf{x}_{it}) I(q_{it} > \gamma_0) + \psi \mathbf{z}'_{it} + u_{it}$, yielding to the fitted values $\Delta \hat{y}_{it}^l$ and $\Delta \hat{y}_{it}^c$, the residuals Δu_{it}^l and Δu_{it}^c .

Step 2. Group Δu_{it}^l and Δu_{it}^c by individuals. Draw by individuals from the regrouped residuals to generate the bootstrapping residuals Δu_{it}^{l*} and Δu_{it}^{c*} . Then, generate $\Delta y_{it}^{l*} = \Delta \hat{y}_{it}^l + \Delta u_{it}^{l*}$ and $\Delta y_{it}^{c*} = \Delta \hat{y}_{it}^c + \Delta u_{it}^{c*}$.

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

Step 4. Repeat Steps 1–3 B times to obtain two samples $F_1^*(1), F_1^*(2), \dots, F_1^*(B)$ and $F_C^*(1), F_C^*(2), \dots, F_C^*(B)$ of simulated F_1 and F_C statistics.

Step 5. Calculate the bootstrap p -values by the frequency of the simulated statistics that exceed actual value for a sufficiently large B .

4 Extension to Multiple Thresholds

In this section, we extend the proposed model to the case with multiple covariate-dependent thresholds. We begin our analysis with the following dynamic panel threshold model with two covariate-dependent thresholds

$$y_{it} = \alpha_i + \beta_1 y_{it-1} + \theta'_1 \mathbf{x}_{it} + (\beta_2 y_{it-1} + \theta'_2 \mathbf{x}_{it}) I(\gamma'_1 \mathbf{s}_{it} \leq q_{it} < \gamma'_2 \mathbf{s}_{it}) + (\beta_3 y_{it-1} + \theta'_3 \mathbf{x}_{it}) I(q_{it} > \gamma'_2 \mathbf{s}_{it}) + u_{it}, \quad (26)$$

for $i = 1, 2, \dots, N$ and $t = 1, 2, \dots, T$, where $\mathbf{s}_{it} = (1, s'_{1,it})' \in \mathbb{R}^{k+1}$, $\gamma_1 = (\gamma_{10}, \gamma'_{11})'$, $\gamma_2 = (\gamma_{20}, \gamma'_{21})'$, and $\gamma'_1 \mathbf{s}_{it} < \gamma'_2 \mathbf{s}_{it}$.³ Other nations are defined as in model (1).

This section focuses on the model with two covariate-dependent thresholds, because the methods discussed can be directly extended to high-order threshold models as explained by Hansen (1999). As in Hansen (1999), we focus on estimating the model, inference for threshold parameters, and testing for a double threshold effect in the proposed model.

4.1 Estimation

For given γ_1 and γ_2 , the double threshold model would degenerate to a dynamic panel linear model in Hsiao et al. (2002), so we can use a maximum likelihood approach to estimate the slope parameters after the first difference transformation. Therefore, for any given threshold parameters γ_1 and γ_2 , we can calculate the concentrated log-likelihood value $\ln L(\gamma_1, \gamma_2)$ as in the single threshold case. As such, we can define the MLE estimator of (γ_1, γ_2) by maximizing $\ln L(\gamma_1, \gamma_2)$; unfortunately, solving the joint maximization may have a high computational cost.

Motivated by the multiple changepoint literature (e.g., Bai 1997) and the threshold literature (e.g., Hansen 1999; Yang 2024), we suggest a sequential procedure to estimate the threshold parameters in the suggested model. In the first stage, we denote the single threshold concentrated log-likelihood as $\ln L(\gamma_1)$, and assume that $\hat{\gamma}_1$ is the threshold estimate maximizing $\ln L(\gamma_1)$. Then, $\hat{\gamma}_1$ will be consistent for either γ_1 or γ_2 depending on which is stronger. As discussed in panel threshold models by Hansen (1999) and Yang (2024), the reason for this

³ It is worth noting that the assumption of $\gamma'_1 \mathbf{s}_{it} < \gamma'_2 \mathbf{s}_{it}$ for all $\mathbf{s}_{it}$'s excludes the case in which $\gamma'_1 \mathbf{s}_{it} < \gamma'_2 \mathbf{s}_{it}$ for some $\mathbf{s}_{it}$'s while $\gamma'_1 \mathbf{s}_{it} \geq \gamma'_2 \mathbf{s}_{it}$ for other $\mathbf{s}_{it}$'s. A further investigation of this issue is worthwhile, but it will not be pursued in this paper.

consistency is that the single threshold concentrated log-likelihood $\ln L(\gamma_1)$ asymptotically converges to a limit function with two local minima at γ_1 and γ_2 .

After obtaining the first-stage estimate $\hat{\gamma}_1$, we then define the second-stage concentrated log-likelihood criterion as

$$\ln L(\gamma_2) = \begin{cases} \ln L(\hat{\gamma}_1, \gamma_2), \hat{\gamma}_1' s_{it} < \gamma_2' s_{it} \\ \ln L(\gamma_2, \hat{\gamma}_1), \gamma_2' s_{it} < \hat{\gamma}_1' s_{it} \end{cases}. \quad (27)$$

Then, we can estimate the threshold parameters by an MCMC-based algorithm similar to Algorithm A, where $\ln L(\gamma)$ is replaced by $\ln L(\gamma_2)$. The previous procedure can be repeated until convergence. To construct confidence intervals for the two threshold parameters (γ_1, γ_2) , we can use a procedure similar to Algorithm A. In detail, we can construct the confidence interval of γ_1 by the $\alpha/2$ and the $1 - \alpha/2$ quantiles of the MCMC sample of the first stage, and then construct the confidence interval of γ_2 by the $\alpha/2$ and the $1 - \alpha/2$ quantiles of the MCMC sample of the second stage.

4.2 Determining the Number of Thresholds

If the null of linearity is rejected, it is often desirable to determine the number of thresholds. To this end, we denote the negative log-likelihood function of the model with two thresholds as $S_2(\hat{\gamma}_1, \hat{\gamma}_2) = -\ln L(\hat{\theta}_\delta(\hat{\gamma}_1, \hat{\gamma}_2), \hat{\gamma}_1, \hat{\gamma}_2, \hat{\sigma}_u^2, \hat{\omega})$, and denote the negative log-likelihood function of the model of one threshold as $S_1(\hat{\gamma}) = -\ln L(\hat{\theta}_\delta(\hat{\gamma}), \hat{\gamma}, \hat{\sigma}_u^2, \hat{\omega})$. Then, a natural test of one threshold against two thresholds can be constructed as

$$F_2 = \frac{S_1(\hat{\gamma}) - S_2(\hat{\gamma}_1, \hat{\gamma}_2)}{S_2(\hat{\gamma}_1, \hat{\gamma}_2)/NT}. \quad (28)$$

To implement the above test F_2 for one threshold against two thresholds, we suggest using a parametric bootstrap procedure based on the null model, i.e., the dynamic panel threshold model with one covariate-dependent threshold. Moreover, a bootstrap procedure similar to Algorithm B can be used to implement the test F_2 . Following a similar logic as above, a test statistic, say F_3 , for two thresholds against three thresholds can be easily constructed, as discussed in Hansen (1999).

5 Extension to a GMM Framework

In this section, we suggest an alternative estimator based on the generalized method of moments (GMM) to complement the proposed MLE estimator due to the potential limitations of the MLE approach. First, the MLE approach requires the error terms $(v_{it}$ and Δu_{it} for $t = 2, \dots, T$) to be normally distributed, and this assumption may be too restrictive in many applications.⁴ Second, the MLE approach imposes strong assumptions on the initial conditions. Specifically, it requires that v_{i1} must be independent of y_{i0} and the covariates $\{\mathbf{x}_{it}, q_{it}, \mathbf{s}_{it}\}_{t=1}^T$. In contrast, the GMM approach is generally based on moment conditions (or orthogonality conditions), and thus does not require a specific parametric distribution or initial conditions. Hence, we can expect that the GMM approach is more robust to non-normal errors and does not require restrictive assumptions about initial conditions.⁵

In the DPTCT model with strictly exogenous variables $(\mathbf{x}_{it}, q_{it}, \mathbf{s}_{1,it})$, following Seo and Shin (2016) we have the following moment conditions

$$E(y_{is} \Delta u_{it}) = 0, \text{ for } t = 2, 3, \dots, T, s = 0, 1, \dots, t-2, \quad (29)$$

⁴ We thank an anonymous referee for raising this point to us.

⁵ Without the assumptions on errors and initial conditions, the GMM estimator may be less efficient.

and

$$E(\mathbf{x}_{is}\Delta u_{it}) = E(q_{is}\Delta u_{it}) = E(\mathbf{s}_{1,is}\Delta u_{it}) = 0, \text{ for } t = 2, 3, \dots, T, \text{ } s = 1, 2, \dots, T. \quad (30)$$

Let $\mathbf{z}_{it} = (y_{i0}, y_{i1}, \dots, y_{i,t-2}, \mathbf{x}'_{i1}, \dots, \mathbf{x}'_{iT}, q_{i1}, \dots, q_{iT}, \mathbf{s}'_{1,i1}, \dots, \mathbf{s}'_{1,iT})$, and $\mathbf{z}_i$ is a block diagonal matrix

$$\mathbf{z}_i = \begin{bmatrix} \mathbf{z}'_{i2} & 0 & \cdots & 0 \\ 0 & \mathbf{z}'_{i3} & & 0 \\ \vdots & & \ddots & \\ 0 & 0 & & \mathbf{z}'_{iT} \end{bmatrix}, \quad (31)$$

thus, the moment equations in (29) and (30) can be written as $E(\mathbf{z}'_i\Delta\dot{u}_i) = 0$, where $\Delta\dot{u}_i = (\Delta u_{i2}, \dots, u_{iT})'$.

5.1 Estimation

Let

$$\Delta\dot{y}_i = \begin{bmatrix} \Delta y_{i2} \\ \Delta y_{i3} \\ \vdots \\ \Delta y_{iT} \end{bmatrix}, \Delta\dot{y}_{i-1}(\gamma) = \begin{bmatrix} \Delta y_{i1} & \Delta y_{i1}(\gamma) & \Delta \mathbf{x}_{i2} & \Delta \mathbf{x}_{i2}(\gamma) \\ \Delta y_{i2} & \Delta y_{i2}(\gamma) & \Delta \mathbf{x}_{i3} & \Delta \mathbf{x}_{i3}(\gamma) \\ \vdots & \vdots & \vdots & \vdots \\ \Delta y_{i,T-1} & \Delta y_{i,T-1}(\gamma) & \Delta \mathbf{x}_{iT} & \Delta \mathbf{x}_{iT}(\gamma) \end{bmatrix}. \quad (32)$$

The model can be written in the matrix form as

$$\Delta\dot{y}_i = \Delta\dot{y}_{i-1}(\gamma)\boldsymbol{\varphi} + \Delta\dot{u}_i, \quad (33)$$

where $\boldsymbol{\varphi} = (\beta_1, \beta_2, \boldsymbol{\theta}'_1, \boldsymbol{\theta}'_2)'$.

Denote the sample moment condition as $g_N(\boldsymbol{\varphi}, \gamma) = N^{-1} \sum_{i=1}^N \mathbf{z}'_i \Delta\dot{u}_i(\boldsymbol{\varphi}, \gamma)$, where $\Delta\dot{u}_i(\boldsymbol{\varphi}, \gamma) = \Delta\dot{y}_i - \Delta\dot{y}_{i-1}(\gamma)\boldsymbol{\varphi}$. Then, the objective function of GMM is given by

$$Q_N(\boldsymbol{\varphi}, \gamma) = g_N(\boldsymbol{\varphi}, \gamma)' \mathbf{S}_N^{-1} g_N(\boldsymbol{\varphi}, \gamma), \quad (34)$$

where $\mathbf{S}_N$ is an optimal weighting matrix such that $\mathbf{S}_N \rightarrow_p \mathbf{S} \equiv E(\mathbf{z}'_i \Delta\dot{u}_i \Delta\dot{u}'_i \mathbf{z}_i) = \sigma_u^2 E(\mathbf{z}'_i \mathbf{V} \mathbf{z}_i)$.⁶ Here, we set

$$\mathbf{S}_N = \hat{\sigma}_u^2 \frac{1}{N} \sum_{i=1}^N \mathbf{z}'_i \mathbf{V} \mathbf{z}_i, \quad (35)$$

where $\hat{\sigma}_u^2$ is a consistent estimation for σ_u^2 and

$$\mathbf{V} = \begin{bmatrix} 2 & -1 & \cdots & 0 \\ -1 & 2 & & \\ \vdots & & \ddots & -1 \\ 0 & -1 & 2 & \end{bmatrix}. \quad (36)$$

For any given γ , the GMM estimator of $\boldsymbol{\varphi}$ is given by

$$\hat{\boldsymbol{\varphi}}(\gamma) = \left[\sum_{i=1}^N \Delta\dot{y}_{i-1}(\gamma)' \mathbf{z}_i \left(\sum_{i=1}^N \mathbf{z}'_i \mathbf{V} \mathbf{z}_i \right)^{-1} \sum_{i=1}^N \mathbf{z}'_i \Delta\dot{y}_{i-1}(\gamma) \right]^{-1} \quad (37)$$

⁶ Based on u_{it} follow i.i.d. Assumption, it is easy to show that $E(\mathbf{z}_{it} \mathbf{z}'_{it} \Delta u_{it}^2) = 2\sigma_u^2 E(\mathbf{z}_{it} \mathbf{z}'_{it})$ for $t = 2, 3, \dots, T$, $E(\mathbf{z}_{it} \mathbf{z}'_{i,t-1} \Delta u_{it} \Delta u_{i,t-1}) = -\sigma_u^2 E(\mathbf{z}_{it} \mathbf{z}'_{i,t-1})$ for $t = 3, 4, \dots, T$, and $E(\mathbf{z}_{it} \mathbf{z}'_{is} \Delta u_{it} \Delta u_{is}) = 0$ for $|t-s| > 1$.

$$\times \left[\sum_{i=1}^N \Delta \dot{y}_{i,-1}(\gamma)' \mathbf{z}_i \left(\sum_{i=1}^N \mathbf{z}_i' \mathbf{V} \mathbf{z}_i \right)^{-1} \sum_{i=1}^N \mathbf{z}_i' \Delta \dot{y}_i \right]. \quad (38)$$

Therefore, the GMM estimator of γ is given by

$$\hat{\gamma} = \arg \min_{\gamma \in \Gamma} Q_N(\gamma), \quad (39)$$

where $Q_N(\gamma) = Q_N(\boldsymbol{\varphi}(\gamma), \gamma)$. Once the estimates $\hat{\gamma}$ are obtained, the slop estimates can be computed via $\hat{\boldsymbol{\varphi}} = \hat{\boldsymbol{\varphi}}(\hat{\gamma})$. To implement the minimization of $Q_N(\gamma)$, we can use Algorithm A by replacing the objective function $\ln L(\gamma)$ with $-Q_N(\gamma)$.

5.2 Asymptotic Theory for the GMM Estimator

We next focus on the asymptotic theory for the GMM estimator. Define the moment functionals $\mathbf{G}_\varphi(\gamma) = E(\mathbf{z}_i' \Delta \dot{y}_{i,-1}(\gamma))$ and

$$\mathbf{G}_\gamma(\gamma) = \begin{bmatrix} \mathbf{G}_{\gamma,2,2}(\gamma) - \mathbf{G}_{\gamma,2,1}(\gamma) \\ \mathbf{G}_{\gamma,3,3}(\gamma) - \mathbf{G}_{\gamma,3,2}(\gamma) \\ \vdots \\ \mathbf{G}_{\gamma,T,T}(\gamma) - \mathbf{G}_{\gamma,T,T-1}(\gamma) \end{bmatrix}, \quad (40)$$

where

$$\mathbf{G}_{\gamma,t,j}(\gamma) = E \left(E \left(\mathbf{z}_{it} w_{i,j-1} \mid q_{ij} = \mathbf{s}_{ij}' \gamma, \mathbf{s}_{ij} \right) f_j \left(\mathbf{s}_{ij}' \gamma \mid \mathbf{s}_{ij} \right) \mathbf{s}_{ij}' \right), \quad (41)$$

in which w_{ij} and $f_t(q_{it} \mid \mathbf{s}_{it})$ have the same definition as in Section 2.

Assumption 2. Assume that

1. For each t , $(\mathbf{x}_{it}, q_{it}, \mathbf{s}_{it}, u_{it}, v_{it})$ are independent and identically distributed (iid) across i .
2. For each i , u_{it} is iid over t , is independent of $\left\{ (y_{ij})_{j=1}^{t-1}, (\mathbf{x}_{it}, q_{it}, \mathbf{s}_{it})_{t=1}^T \right\}$.
3. $E \|\mathbf{z}_{it} w_{i,j-1}\|^2 < \infty$, for $j = t, t-1$.
4. For any $\gamma \in \Gamma$ and $t = 2, \dots, T$, $j = t, t-1$, $E \left(\|\mathbf{z}_{it} w_{i,j-1}\|^2 \mid q_{ij} = \mathbf{s}_{ij}' \gamma, \mathbf{s}_{ij} \right) \leq C$ for some $C < \infty$, and $0 < f_j(q_{ij} \mid \mathbf{s}_{ij}) \leq \bar{f} < \infty$.
5. $f_j(\mathbf{s}_{ij}' \gamma \mid \mathbf{s}_{ij})$ and $\mathbf{G}_{\gamma,t,j}(\gamma)$ are continuous at $\gamma = \gamma_0$, where γ_0 is the true value of γ .
6. $\mathbf{s}_{it}$ is not multicollinear.
7. $\beta_{20} = c_{\beta_2} N^{-\alpha}$ and $\theta_{20} = c_{\theta_2} N^{-\alpha}$ with $\mathbf{c} = (c_{\beta_2}, c_{\theta_2})' \neq \mathbf{0}$ and $0 \leq \alpha < 1/2$, where $\mathbf{c}$ is a vector of constant and $(\beta_{20}, \theta_{20})'$ is the true value of $(\beta_2, \theta_2)'$.
8. Let $\mathbf{G}(\gamma) = (\mathbf{G}_\varphi(\gamma), \mathbf{G}_\gamma(\gamma))$, then $\mathbf{G}(\gamma)$ have full column rank for all $\gamma \in \Gamma$.
9. $\mathbf{S}$ is finite and positive definite. □

Assumption 2 is essentially similar to Assumption 1 without restrictions on error distributions and initial conditions. Here, we allow a fixed threshold setting following Seo and Shin (2016). We next establish the asymptotic distribution of the GMM estimators $\hat{\boldsymbol{\varphi}}$ and $\hat{\gamma}$.

Theorem 3. Under Assumption 2, as $N \rightarrow \infty$, we have

$$\begin{bmatrix} N^{1/2}(\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}_0) \\ N^{1/2-\alpha}(\hat{\gamma} - \gamma_0) \end{bmatrix} \xrightarrow{d} N \left(\mathbf{0}, (\mathbf{G}(\gamma_0)' \mathbf{S}^{-1} \mathbf{G}(\gamma_0))^{-1} \right). \quad (42)$$

Proof of Theorem 3. See the Appendix A. □

Theorem 3 can be viewed as a generalization of the asymptotic result of Theorem 1 in Seo and Shin (2016). Theorem 3 shows that the slope and threshold estimators are jointly asymptotically normal under both the fixed and diminishing threshold assumptions. Moreover, compared to the MLE threshold estimator, which is super-consistent and asymptotically independent of the slope estimator, the GMM threshold estimator has a slower convergence rate and is jointly dependent of the slope estimator.

6 Monte Carlo Simulations

In this section, we aim at examining the finite sample performances of the proposed estimation and testing procedures. To this end, we consider the data-generating process (DGP) given by

$$y_{it} = (\beta_1 y_{it-1} + \theta_1 q_{it})I(q_{it} \leq \gamma_{it}) + (\beta_2 y_{it-1} + \theta_2 q_{it})I(q_{it} > \gamma_{it}) + \psi z_{it} + e_{it}, \quad (43)$$

$$\gamma_{it} = \gamma_0 + \gamma_1 s_{it}, \quad (44)$$

$$\Delta y_{it} = \delta_0 + \delta_1 \Delta q_{it} + \delta_2 \Delta z_{it} + v_{it}, \quad (45)$$

in which $e_{it} = \alpha_i + u_{it}$, $z_{it} \sim N(0, 1)$, $q_{it} \sim N(0.2, 1)$, $s_{it} \sim N(0, 1)$, $u_{it} \sim N(0, 2^2)$, $v_{it} \sim N(0, 4^2)$. To allow for the correlation between the exogenous variable and the individual fixed effect, we set $\alpha_i = a_i + \bar{q}_i$, $a_i \sim N(2, 3^2)$, $\bar{q}_i = \sum_{t=0}^T q_{it}$, from $t = 0$ to $t = T$. To generate y_{it} , we also set $y_{i,0} = 0$.

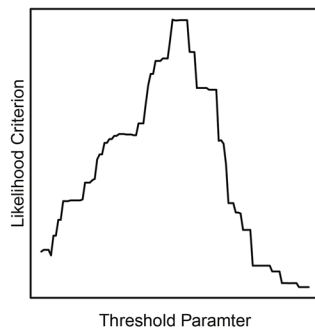
In the first experiment, we investigate the finite properties of the estimation procedure. To do so, we set $T = (5, 10)$, $N = (50, 100, 200)$, $\beta_1 = -0.3$, $\theta_1 = \psi = 1$, $\beta_2 = 0.7$, $\theta_2 = 2$, $\delta_0 = 0.5$, $\delta_1 = \delta_2 = 1$, $\gamma_0 = 0.2$ and $\gamma_1 = 0.5$. The number of replications is set as 1,000. To implement the MCMC-based algorithm described in Section 2, the estimation software (associated with Algorithm A) is written based on the **runMCMC** function in the **R** package **BayesianTools** (Hartig et al. 2023), which incorporates the so-called Differential Evolution Markov Chain (DREAM MCMC) sampling approach that can efficiently handle problems involving high dimensionality, multimodality, nonlinearity, and local optima, as explained by Vrugt et al. (2009). As can be seen from Figure 1, unlike constant threshold models, the likelihood criterion for the suggested nonconstant threshold model seems to contain problems such as multimodality, nonlinearity, and local optima, and thus, the DREAM MCMC approach may be more efficient than the classical MCMC methods such as the Metropolis-Hastings sampler.⁷ Table 1 reports the simulation results.⁸ The simulation results show that the bias of each parameter is very close to zero for all combinations of T and N , and the standard deviations decrease to zero as $T \times N$ increases. In particular, when N increases from 50 to 100 for a fixed T , the standard deviations of the threshold parameters γ_0 and γ_1 would decrease by almost half, which confirms our asymptotical results in Theorem 1. Also, the R statistics of Gelman and Rubin (1992) for the threshold parameters are close to 1, providing evidence of convergence of the MCMC samples for threshold parameters, which further demonstrates that the suggested estimation algorithm performs well in finite samples.

We next assess the size and power properties of the test statistics F_1 (the covariate-dependent threshold model against the linear model) and F_C (the covariate-dependent threshold model against the constant threshold model). We first examine the finite sample properties of the test F_1 . To this end, we set $\beta_1 = -0.3$, $\theta_1 = \psi = 1$, $\delta_0 = 0.5$, $\delta_1 = \delta_2 = 1$, $\gamma_0 = 0.2$, $\gamma_1 = 0.5$, and vary the parameters β_2 and θ_2 to assess the size and power of the F_1 test for threshold effect. Similarly, we set $\beta_1 = -0.3$, $\theta_1 = \psi = 1$, $\beta_2 = 0.7$, $\theta_2 = 2$, $\delta_0 = 0.5$, $\delta_1 = \delta_2 = 1$, $\gamma_0 = 0.2$, and vary γ_1 to assess the size and power of the F_C test for threshold constancy. As such, the rejection frequencies with $\beta_2 = -0.3$, $\theta_2 = 1$ and $(\beta_2, \theta_2) = \{(-0.1, 1.2), (0.7, 2)\}$ are the size and power of the first test statistic, and the rejection frequencies with $\gamma_1 = 0$ and $\gamma_1 = (0.2, 0.5)$ are the size and power of

⁷ In an unreported appendix, we, through simulations, show that the DREAM MCMC approach is more efficient than the Metropolis-Hastings sampler in our proposed model.

⁸ In the simulations, we draw 12,000 samples from the posterior after discarding the first 6,000 “burn-in” draws.

Constant Threshold Model



Nonconstant Threshold Model

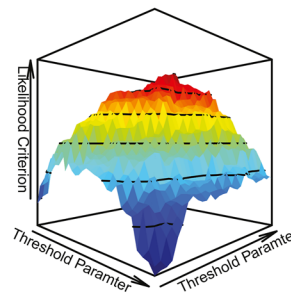


Figure 1: Simulated negative likelihood functions of threshold models.

Table 1: Estimates of the parameters.

T	N		$\gamma_0 = 0.2$	$\gamma_1 = 0.5$	$\beta_1 = -0.3$	$\beta_2 = 0.7$	$\theta_1 = 1$	$\theta_2 = 2$	$\psi = 1$	$\delta_0 = 0.5$	$\delta_1 = 1$	$\delta_2 = 1$	The R statistic	
													γ_0	γ_1
5	50	Bias	−0.001	−0.000	0.034	0.039	0.011	0.002	0.002	−0.006	−0.020	−0.023	1.140	1.120
		Std.dev	0.032	0.042	0.036	0.038	0.277	0.231	0.154	0.614	0.431	0.448	0.599	0.245
	100	Bias	0.000	0.001	0.034	0.040	0.010	−0.001	0.008	−0.006	−0.020	−0.018	1.146	1.157
		Std.dev	0.015	0.020	0.024	0.026	0.191	0.168	0.106	0.427	0.306	0.306	0.225	0.401
	200	Bias	0.000	0.000	0.034	0.039	0.007	0.002	0.007	−0.024	−0.000	−0.013	1.173	1.163
		Std.dev	0.007	0.009	0.017	0.018	0.143	0.118	0.074	0.292	0.221	0.206	0.285	0.255
10	50	Bias	−0.000	−0.000	0.015	0.016	−0.004	0.004	0.002	0.021	−0.009	0.007	1.160	1.155
		Std.dev	0.013	0.018	0.021	0.021	0.184	0.155	0.098	0.656	0.430	0.442	0.456	0.462
	100	Bias	0.000	0.000	0.015	0.016	−0.001	0.002	0.002	−0.018	−0.002	−0.027	1.194	1.194
		Std.dev	0.007	0.009	0.014	0.015	0.125	0.109	0.069	0.447	0.319	0.306	0.436	0.423
	200	Bias	−0.000	−0.000	0.015	0.017	−0.001	0.001	0.000	−0.013	0.004	−0.006	1.268	1.252
		Std.dev	0.003	0.005	0.010	0.011	0.094	0.074	0.048	0.316	0.221	0.219	0.477	0.429

Notes: this table presents the summary statistics (i.e., mean and standard deviation) for the estimates and the R statistics for threshold parameters. The number of replications is 1,000.

the second test statistic. All sizes and powers are constructed using 1,000 replications and bootstrap replications are set to 100 to save the simulation time. The simulation results are reported in Table 2. The simulation results show that the empirical sizes are close to the 5 % nominal level for the tests F_1 and F_C in most cases, except for the combination $(N, T) = (200, 10)$ for the test F_C , in which case the size is smaller than the 5 % nominal level; this may be explained by the small number of replications. Furthermore, the power for the test F_1 increases when the magnitude of threshold effect increases or the sample size becomes larger; and the power for the test F_C increases when the magnitude of the threshold constancy or the sample size increases. These results indicate that the proposed test statistics work well in finite samples.

Furthermore, in Appendix B we conduct Monte Carlo simulations comparing the proposed MLE with the GMM estimator discussed in Section 5 under the data-generating processes allowing the violations of the normal disturbances and initial-condition assumptions. The simulation results show that, when the sample size is doubled, the standard deviations of the threshold parameters would decrease by almost half when the MLE estimator is employed, while these standard deviations would decrease by roughly 30 % when the GMM approach is employed, confirming the super-consistent property in Theorem 1 for the MLE estimator and the normality in Theorem 3 for the GMM estimator. Overall, these simulation results indicate that the proposed MLE method works well in finite samples in the considered cases, and the GMM estimator serves as a useful complementary alternative under less restrictive assumptions on error distributions and initial conditions.

Table 2: Finite-sample size and power of the F -type test statistics.

T	N	Test for threshold effect			Test for threshold constancy		
		Size	Power		Size	Power	
			Small	Large		Small	Large
5	50	0.056	0.576	0.979	0.045	0.978	0.988
	100	0.040	0.956	1.000	0.048	0.999	1.000
	200	0.050	0.912	0.993	0.042	0.995	0.999
10	50	0.060	0.998	1.000	0.025	1.000	1.000
	100	0.055	1.000	1.000	0.028	0.999	1.000
	200	0.053	1.000	1.000	0.030	1.000	1.000

Notes: ‘small’ and ‘large’ in F_1 denote small threshold effect (e.g., $(\beta_2, \theta_2) = (-0.1, 1.2)$) and large threshold effect (e.g., $(\beta_2, \theta_2) = (0.7, 2)$), respectively; ‘small’ and ‘large’ in F_c denote the magnitude of threshold dependency, corresponding to $\gamma_1 = 0.2$ and 0.5 , respectively.

7 Empirical Application: The Inflation-Growth Nexus

Over the past several decades, the linkage between inflation and growth has attracted much attention among academics and policymakers. This linkage is important as the money neutrality assumption demonstrates that inflation does not affect economic growth. Much empirical literature has investigated the nexus between inflation and growth, nonetheless, reported conflicting and mixed results. As a result, a number of studies suggest to focus on the nonlinear relationship between inflation and growth based on threshold models, and provide evidence supporting a negative effect of inflation on growth for periods above the inflation threshold (e.g., Khan and Ssnhadji 2001; Kremer et al. 2013; Ramírez-Rondán 2020; Vaona and Schiavo 2007). Another important and closely related empirical issue is to estimate the inflation threshold above which inflation has a negative effect on growth. For example, Vaona and Schiavo (2007) use a nonparametric method to obtain an inflation threshold around 12 %. Kremer et al. (2013) find the inflation threshold at 2.5 % for industrialized countries and 17 % for nonindustrialized countries, respectively. More recently, Ramírez-Rondán (2020) proposes a dynamic panel threshold model and applies the model to the inflation-growth relationship, leading to an inflation threshold at around 3.1–3.5 % for industrialized countries and around 16.0–16.1 % for nonindustrialized countries.

In this section, we extend the analysis of Ramírez-Rondán (2020) by applying the proposed dynamic panel threshold model with a covariate-dependent threshold to estimate the inflation threshold. The data we used is the same as in Ramírez-Rondán (2020). This data set covers the observations from 1961 to 2015 over 5-year average periods for a sample of 74 countries.⁹ As emphasized in the literature, industrialized and nonindustrialized countries may have different inflation thresholds (e.g., Khan and Ssnhadji 2001; Kremer et al. 2013; Ramírez-Rondán 2020). Thus, the literature estimates the inflation thresholds based on separate analyses for industrialized and nonindustrialized countries by splitting the sample into subgroups. This leads to at least two problems. First, we have to face the small sample size problem, which results in inaccurate estimates and inferences. Second, it is difficult to test whether there is a significant difference between inflation thresholds for industrialized and nonindustrialized countries. To overcome these problems, we extend the empirical model of Ramírez-Rondán (2020) by allowing for a covariate-dependent threshold given by

$$\begin{aligned}
 y_{it} - y_{it-1} &= \alpha_i + \beta y_{it-1} + \theta_1 \log(x_{it})1(x_{it} \leq \gamma) + \theta_2 \log(x_{it})1(x_{it} > \gamma) \\
 &\quad + \pi' z_{it} + u_{it}, \\
 \gamma &= \gamma_0 + \gamma_1 s_{it},
 \end{aligned} \tag{46}$$

⁹ See Ramírez-Rondán (2020) for more details.

Table 3: Estimates based on different models.

Dependent variable: GDP per capita growth	Ramírez-Rondén's (2020) model					
	(1)		(2)		(3)	
	Estimate	95 %CI	Estimate	95 %CI	Estimate	95 %CI
Inflation rate (inflation rate $\leq \hat{\gamma}$)	−0.088	[−0.309, 0.133]	−0.016	[−0.227, 0.196]	−0.090	[−0.308, 0.127]
Inflation rate (inflation rate $> \hat{\gamma}$)	−0.693	[−0.850, −0.535]	−0.554	[−0.708, −0.400]	−0.591	[−0.747, −0.435]
Transitional convergence	−0.024	[−0.029, −0.020]	−0.026	[−0.032, −0.020]	−0.026	[−0.031, −0.020]
Human capital	−	−	−0.036	[−0.056, −0.015]	−0.015	[−0.041, 0.011]
Financial depth	−	−	−0.001	[−0.005, 0.003]	0.000	[−0.004, 0.004]
Governance	−	−	0.190	[−0.107, 0.487]	0.137	[−0.161, 0.434]
Public infrastructure	−	−	0.003	[0.001, 0.004]	0.003	[0.002, 0.005]
Trade openness	−	−	1.284	[0.711, 1.857]	1.492	[0.902, 2.083]
Economic instability	−	−	−0.241	[−0.310, −0.172]	−0.243	[−0.312, −0.175]
Threshold inflation estimate 0 ($\hat{\gamma}_0$)	16.1	[14.5, 16.6]	16.0	[14.6, 16.9]	16.1	[14.7, 16.9]
Threshold inflation estimate 1 ($\hat{\gamma}_1$)	−	−	−	−	−	−
Time shifts	No		No		Yes	
$F_1(p\text{-value})$	0.000		0.000		0.000	

Dependent variable: GDP per capita growth	DPTCT					
	(1)		(2)		(3)	
	Estimate	95 %CI	Estimate	95 %CI	Estimate	95 %CI
Inflation rate (inflation rate $\leq \hat{\gamma}$)	−0.103	[−0.353, 0.146]	−0.051	[−0.289, 0.187]	−0.051	[−0.289, 0.186]
Inflation rate (inflation rate $> \hat{\gamma}$)	−0.730	[−0.891, −0.569]	−0.608	[−0.765, −0.451]	−0.610	[−0.767, −0.454]
Transitional convergence	−0.021	[−0.025, −0.016]	−0.023	[−0.029, −0.017]	−0.023	[−0.029, −0.017]
Human capital	−	−	−0.033	[−0.054, −0.013]	−0.030	[−0.058, −0.003]
Financial depth	−	−	0.001	[−0.003, 0.005]	0.001	[−0.003, 0.005]
Governance	−	−	0.078	[−0.219, 0.376]	0.072	[−0.225, 0.369]
Public infrastructure	−	−	0.002	[0.000, 0.004]	0.002	[0.000, 0.004]
Trade openness	−	−	1.426	[0.859, 1.994]	1.445	[0.862, 2.028]
Economic instability	−	−	−0.243	[−0.311, −0.175]	−0.243	[−0.311, −0.175]
Threshold inflation estimate 0 ($\hat{\gamma}_0$)	16.1	[15.0, 16.2]	16.1	[15.0, 16.3]	16.1	[15.1, 16.2]
Threshold inflation estimate 1 ($\hat{\gamma}_1$)	−12.5	[−16.4, −3.3]	−13.0	[−16.2, −4.7]	−13.0	[−15.5, −5.7]
Time shifts	No		No		Yes	
$F_1(p\text{-value})$	0.000		0.002		0.000	
$F_2(p\text{-value})$	0.212		0.153		0.132	
$F_C(p\text{-value})$	0.007		0.017		0.019	

Notes: F_1 , F_2 and F_C are, in order, the test for threshold effect, the test for two thresholds effects and the test for threshold constancy.

where y_{it} is the log of real gross domestic product (GDP) per capita, x_{it} is the inflation rate; α_i is an unobserved country-specific effect; γ is a time-varying threshold level of the inflation rate; s_{it} is the dummy variable that equals one if the country is industrialized and zero otherwise; $I(\cdot)$ is an indicator variable; z_{it} is a set of other determinants of economic growth;¹⁰ i refers to the country, and t refers to the time period. All of the explanatory variables are in the form of logarithms, and the semi-log transformation is used when an explanatory variable has negative values. In addition, the threshold variable is the level of the inflation rate as in Ramírez-Rondán (2020).

¹⁰ Following Ramírez-Rondán (2020), z_{it} includes trade openness, financial depth, governance, public infrastructure, human capital investment, and economic instability which are usually posited as determinants of economic growth. In addition, the covariates in a time-varying threshold may bring endogeneity to a nonconstant threshold model, which can be solved by including the covariates as control. As a result, in our model we add an interaction between the covariate we used and time period shifts as a control to avoid this question.

Table 3 reports the empirical results. For the sake of comparison, we also replicate and report the empirical results of Ramírez-Rondán (2020) in Table 3. As in Ramírez-Rondán (2020), we consider three different specifications. In the first specification, we estimate the empirical model with only initial GDP per capita and without control variables. Then, we form the second specification and the third specification by adding control variables and time period shifts, respectively. From Table 3, we can see that the threshold estimate ($\hat{\gamma}_0$) for nonindustrialized countries is 16.0–16.1 %, which is similar to the Ramírez-Rondán (2020); this inflation threshold is close to the 17 % inflation threshold reported by Kremer et al. (2013), while different from the 11 % inflation threshold estimated by Khan and Ssnhadji (2001). For industrialized countries, our estimate is around 3.1–3.2 % ($\hat{\gamma}_0 + \hat{\gamma}_1$), which is also close to Ramírez-Rondán (2020) but higher than the 2.5 % threshold estimated by Kremer et al. (2013). Our 95 % confidence intervals for the threshold parameters are tight, indicating a higher estimation accuracy. We next focus on the estimation and inferences on the slope parameters. In the three specifications denoted by (1)–(3) in Table 3, the coefficients of the inflation rate have no effects on growth below the inflation threshold, as suggested by the money neutrality assumption. However, the inflation rate in the high inflation regime could significantly hurt growth at the 5 % level of significance. These results are also consistent with Khan and Ssnhadji (2001), Kremer et al. (2013), and Ramírez-Rondán (2020).

As is typical in the threshold literature, we use the proposed test statistics to test threshold effect and threshold constancy, and determine the number of thresholds. In the three specifications, according to F_1 , the null of linearity is rejected with the p -value 0.000, 0.002, and 0.000 for three specifications denoted by (1)–(3) in Table 3, respectively. According to F_2 , one threshold model is supported by the data according to the p -values. According to the F_c test statistic, the null of constant threshold is rejected with the p -value 0.016, 0.018, and 0.028 for three specifications. These testing results not only indicate that the DPTCT model with one covariate-dependent threshold is appropriate in the inflation-growth relationship, but also indicate that there is a significant difference between inflation thresholds for industrialized and nonindustrialized countries.

8 Conclusions

This paper proposes a dynamic panel threshold model with a covariate-dependent and time-varying threshold (DPTCT) modeled as a function of candidate variables affecting the threshold. We study the estimation, inference and specification testing of the proposed model. Following Hsiao et al. (2002) and Ramírez-Rondán (2020), we suggest an MLE approach to estimate the model. Also, we suggest test statistics for threshold effect and threshold constancy. The asymptotic properties of the estimator and the test statistics are established. Specifically, our asymptotic results indicate that the threshold estimator shares the convergence rate with the estimator in the constant threshold model. Hence, the inclusion of a time-varying threshold does not change the convergence rate of threshold parameters, and thus, the true model parameters can not be retrieved from the estimation method based on the constant threshold model if we misspecify a time-varying threshold as a constant one. Our Monte Carlo simulation results confirm the asymptotic theory, and indicate that the estimation and testing procedures work well in finite samples. We illustrate the model with an empirical application to investigate the inflation-growth relationship, and the empirical results reveal that there is a significant difference between inflation thresholds for industrialized and nonindustrialized countries.

Acknowledgments: The authors would like to thank the Editor Professor Jeremy Piger and two anonymous referees for their very constructive comments and suggestions. We also thank Yanli Xie for research assistance. Any remaining errors are solely our responsibility.

Research ethics: This article does not contain any studies with human participants or animals performed by the authors.

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

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

Data availability: Codes and data are available from the authors on request.

Appendix A: Mathematical Proofs

This appendix provides the detailed proof of Theorems 1–3 in the paper. To this end, we first prove the following Lemmas, which are useful in proving Theorems 1–3. For notational simplicity, we first clarify the following denotations.

1. $\boldsymbol{\phi} = (\boldsymbol{\theta}'_{\delta}, \sigma_u^2, \omega)'$, $\boldsymbol{\phi}_0 = (\boldsymbol{\theta}'_{\delta 0}, \sigma_{u0}^2, \omega_0)'$, $\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}) = (\hat{\boldsymbol{\theta}}_{\delta}(\boldsymbol{\gamma})', \hat{\sigma}_u^2(\boldsymbol{\gamma}), \hat{\omega}(\boldsymbol{\gamma}))'$, $\varpi = 1 + T(\omega - 1)$, $\varpi_0 = 1 + T(\omega_0 - 1)$.
2. $I_{it}(\boldsymbol{\gamma}) = I(q_{it} \leq \mathbf{s}'_{it}\boldsymbol{\gamma})$, $\nabla I_{it}(\boldsymbol{\gamma}) = I_{it}(\boldsymbol{\gamma}) - I_{it}(\boldsymbol{\gamma}_0)$.
3. $P_{it}(\boldsymbol{\gamma}) = P(q_{it} \leq \mathbf{s}'_{it}\boldsymbol{\gamma})$, $\nabla P_{it}(\boldsymbol{\gamma}) = P_{it}(\boldsymbol{\gamma}) - P_{it}(\boldsymbol{\gamma}_0)$.
4. $\lambda_N = N^{1-2\alpha}$, $\boldsymbol{\gamma} = (\gamma_0, \boldsymbol{\gamma}'_s)'$, $\boldsymbol{v} = (v_0, \boldsymbol{v}'_s)'$.
5. $\Delta \boldsymbol{w}_{it} = (\Delta y_{it}, \Delta \mathbf{x}'_{it+1})'$, $\nabla \Delta \boldsymbol{w}_{it}(\boldsymbol{\gamma}) = \Delta \boldsymbol{w}_{it} \nabla I_{it}(\boldsymbol{\gamma})$.
6. $\boldsymbol{M}^*(\boldsymbol{\gamma}, \boldsymbol{\gamma}_0) = E(\Delta y_{i,-1}(\boldsymbol{\gamma})' \boldsymbol{\Omega}_0^{-1} \Delta y_{i,-1}(\boldsymbol{\gamma}_0))$, $\boldsymbol{M}^*(\boldsymbol{\gamma}) = E(\Delta y_{i,-1}(\boldsymbol{\gamma})' \boldsymbol{\Omega}_0^{-1} \Delta y_{i,-1}(\boldsymbol{\gamma}))$.
7. $J_N(\boldsymbol{\gamma}) = N^{-1/2} \sum_{i=1}^N \Delta y_{i,-1}(\boldsymbol{\gamma})' \boldsymbol{\Omega}_0^{-1} \Delta u_i$.
8. $G_N(\boldsymbol{v}) = \lambda_N N^{-1} \sum_{i=1}^N \mathbf{c}' \nabla \Delta y_{i,-1}^*(\boldsymbol{\gamma}_0 + \boldsymbol{v}/\lambda_N)' \boldsymbol{\Omega}_0^{*-1} \nabla \Delta y_{i,-1}^*(\boldsymbol{\gamma}_0 + \boldsymbol{v}/\lambda_N) \mathbf{c}$.
9. $R_N(\boldsymbol{v}) = \sqrt{a_N} N^{-1/2} \sum_{i=1}^N \mathbf{c}' \nabla \Delta y_{i,-1}^*(\boldsymbol{\gamma}_0 + \boldsymbol{v}/\lambda_N) \boldsymbol{\Omega}_0^{*-1} \Delta u_i$.
10. $V_N(\boldsymbol{v}) = \lambda_N N^{-1} \sum_{i=1}^N (\mathbf{c}' \nabla \Delta y_{i,-1}^*(\boldsymbol{\gamma}_0 + \boldsymbol{v}/\lambda_N) \boldsymbol{\Omega}_0^{*-1} \Delta u_i)^2$.
11. $\boldsymbol{h}_i(\boldsymbol{\gamma}) = (\boldsymbol{h}_{1i}(\boldsymbol{\gamma})', h_{2i}, h_{3i})'$, where $\boldsymbol{h}_{1i}(\boldsymbol{\gamma}) = \Delta y_{i,-1}(\boldsymbol{\gamma})' \boldsymbol{\Omega}_0^{-1} \Delta u_i$, $h_{2i} = (\Delta u_i' \boldsymbol{\Omega}_0^{*-1} \Delta u_i - T)/(2\sigma_{u0}^2)$ and $h_{3i} = (\Delta u_i' \boldsymbol{\Phi} \Delta u_i - T\sigma_{u0}^2 \varpi_0)/(2\sigma_{u0}^2 \varpi_0)$.
12. $\boldsymbol{K}_N(\boldsymbol{\phi}, \boldsymbol{\gamma}) = N^{-1/2} \partial \ln L(\boldsymbol{\phi}, \boldsymbol{\gamma}) / \partial \boldsymbol{\phi}$, $\boldsymbol{H}_N(\boldsymbol{\phi}, \boldsymbol{\gamma}) = -N^{-1} \partial^2 \ln L(\boldsymbol{\phi}, \boldsymbol{\gamma}) / \partial \boldsymbol{\phi} \partial \boldsymbol{\phi}'$.
13. As shown in Hsiao et al. (2002), $\boldsymbol{\Omega}^{*-1} = \boldsymbol{H}' \boldsymbol{\Lambda}^{-1} \boldsymbol{H}$, where $a_{s+1} - 2a_s + a_{s-1} = 0$ for $s = 1, 2, \dots, T$, $a_0 = 1$, $a_1 = \omega$, $\boldsymbol{\Lambda} = \text{diag}(a_0 a_1, a_1 a_2, \dots, a_{T-1} a_T)$, and

$$\boldsymbol{H} = \begin{bmatrix} a_0 & 0 & \cdots & 0 \\ a_0 & a_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_0 & a_1 & \cdots & a_{T-1} \end{bmatrix}. \quad (\text{A1})$$

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

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

as $N \rightarrow \infty$.

Lemma 2. Under Assumption 1, $J_N(\boldsymbol{\gamma}) \rightarrow_d J(\boldsymbol{\gamma})$, a mean-zero Gaussian process with almost surely continuous sample paths.

Proof of Lemma 1–2. The proofs are similar to Hansen (1996, 2000) and Yang et al. (2021a). To save space, the proofs are skipped here. $\square$

Lemma 3. Under Assumption 1, we have

$$E(\Delta y_{i,-1}(\boldsymbol{\gamma})' \boldsymbol{\Omega}^{*-1} \Delta u_i) = \mathbf{0}, \quad (\text{A3})$$

$$E(\Delta u_i' \boldsymbol{\Phi} \Delta u_i) = T\sigma_u^2 \varpi. \quad (\text{A4})$$

Proof. We first prove (A3) and then (A4). Since $\Delta \mathbf{x}_{it}$ is independent of Δu_{it} , we next focus on Δy_{it} and $\Delta y_{it}(\gamma)$. To prove (A3), we need to prove the following equation

$$\sum_{t=2}^T [a_t a_{t-1}]^2 E \left[\left(\sum_{j=1}^{t-1} a_j \Delta y_{ij} \right) \left(v_{i1} + \sum_{j=2}^t a_{j-1} \Delta u_{ij} \right) \right] = 0, \quad (\text{A5})$$

$$\sum_{t=2}^T [a_t a_{t-1}]^2 E \left[\left(\sum_{j=1}^{t-1} a_j \Delta y_{ij}(\gamma) \right) \left(v_{i1} + \sum_{j=2}^t a_{j-1} \Delta u_{ij} \right) \right] = 0, \quad (\text{A6})$$

In the next, we only need to prove (A6), as the same argument can prove (A5).

Following Ramírez-Rondán (2020), denote

$$A_t = \left(\sum_{j=1}^{t-1} a_j \Delta y_{ij}(\gamma) \right) \left(v_{i1} + \sum_{j=2}^t a_{j-1} \Delta u_{ij} \right). \quad (\text{A7})$$

Then, we have

$$A_t = A_{t-1} + a_{t-1} \Delta y_{it-1}(\gamma) \left(v_{i1} + \sum_{j=2}^t a_{j-1} \Delta u_{ij} \right) + a_{t-1} \Delta u_{it} \sum_{j=1}^{t-2} a_j \Delta y_{ij}(\gamma). \quad (\text{A8})$$

Taking the expectation of (A8), we have

$$E(A_t) = E(A_{t-1}) + a_{t-1} E \left[\Delta y_{it-1}(\gamma) \left(v_{i1} + \sum_{j=2}^t a_{j-1} \Delta u_{ij} \right) \right]. \quad (\text{A9})$$

Through iteration, we only need to prove that the second term on the right side of (A9) is equal to 0. The proof of $E(A_2) = E(A_3) = E(A_4) = 0$ is similar as in Ramírez-Rondán (2020), and thus we next prove the case of $E(A_t)$ when $t > 4$. To do so, we prove that $E(A_{k+1}) = 0$ if $E(A_k) = 0$.

Suppose that for a particular k , $E(A_k) = 0$. Therefore, we have the following equation

$$E[\Delta y_{ik-1} v_{i1}] + \sum_{j=2}^k a_{j-1} E[y_{ik-1} \Delta u_{ij}] = 0, \quad (\text{A10})$$

$$E[\Delta y_{ik-1}(\gamma) v_{i1}] + \sum_{j=2}^k a_{j-1} E[y_{ik-1}(\gamma) \Delta u_{ij}] = 0. \quad (\text{A11})$$

We also have the following results

$$\begin{aligned} y_{it} &= y_{it-1} + \beta_1 \Delta y_{it-1} + \beta_2 \Delta y_{it-1}(\gamma) + \theta'_1 \Delta \mathbf{x}_{it} + \theta'_2 \Delta \mathbf{x}_{it}(\gamma) + \Delta u_{it}, \\ &= y_{i1} + \sum_{j=1}^{t-1} [\beta_1 \Delta y_{ij} + \beta_2 \Delta y_{ij}(\gamma) + \theta'_1 \Delta \mathbf{x}_{ij+1} + \theta'_2 \Delta \mathbf{x}_{ij+1}(\gamma) + \Delta u_{ij+1}], \end{aligned} \quad (\text{A12})$$

$$E(\Delta u_{ik} v_{i1}) = 0, \quad (\text{A13})$$

$$E(y_{i1} v_{i1}) = \sigma_v^2, \quad (\text{A14})$$

$$\sum_{s=1}^{k-2} E(\Delta u_{is+1} v_{i1}) = -\sigma_u^2, \quad (\text{A15})$$

$$\begin{aligned} \sum_{j=2}^k a_{j-1} E(\Delta u_{ik} \Delta u_{ij}) &= a_{k-1} E(\Delta u_{ik} \Delta u_{ik-1}) + a_{k-1} E(\Delta u_{ik}^2) \\ &= a_{k-2} (-\sigma_u^2) + 2a_{k-1} \sigma_u^2, \end{aligned} \quad (\text{A16})$$

$$\sum_{j=2}^k a_{j-1} E(y_{i1} \Delta u_{ij}) = a_1 E(y_{i1} \Delta u_{i2}) = a_1 (-\sigma_u^2), \quad (A17)$$

$$\begin{aligned} \sum_{j=2}^k a_{j-1} \sum_{s=1}^{t-2} E(\Delta u_{is+1} \Delta u_{ij}) &= \sum_{j=2}^{k-1} a_{j-1} E(\Delta u_{ij}^2) + \sum_{j=3}^k a_{j-1} E(\Delta u_{ij} \Delta u_{ij-1}) \\ &\quad + \sum_{j=2}^{t-2} a_{j-1} E(\Delta u_{ij} \Delta u_{ij+1}) \\ &= 2 \sum_{j=2}^{k-1} a_{j-1} \sigma_u^2 - \sum_{j=3}^k a_{j-1} \sigma_u^2 - \sum_{j=2}^{t-2} a_{j-1} \sigma_u^2 \\ &= a_0 \sigma_u^2, \end{aligned} \quad (A18)$$

which uses the result $a_{t-2} - 2a_{t-1} + a_t = 0$. Let

$$E(\Delta y_{ik}(\gamma) v_{i1}) = B_1(v_{i1}) + B_2(v_{i1}) + B_3(v_{i1}), \quad (A19)$$

$$E(\Delta y_{ik}(\gamma) \Delta u_{ij}) = B_1(\Delta u_{ij}) + B_2(\Delta u_{ij}) + B_3(\Delta u_{ij}), \quad (A20)$$

where

$$\begin{aligned} B_1(X) &= \beta_1 \left[P_{ik+1}(\gamma) E(\Delta y_{ik-1} X) + \Delta P_{ik+1}(\gamma) \sum_{s=1}^{t-2} E(\Delta y_{is} X) \right], \\ B_2(X) &= \beta_2 \left[P_{ik+1}(\gamma) E(\Delta y_{ik-1}(\gamma) X) + \Delta P_{ik+1}(\gamma) \sum_{s=1}^{t-2} E(\Delta y_{is}(\gamma) X) \right], \\ B_3(X) &= P_{ik+1}(\gamma) E(\Delta u_{ik} X) + \Delta P_{ik+1} \left[E(y_{i1} X) + \sum_{s=1}^{t-2} E(\Delta u_{is+1} X) \right]. \end{aligned} \quad (A21)$$

According to (A10) and (A11), it is easy to show that

$$B_1(v_{i1}) + \sum_{j=2}^k a_{j-1} B_1(\Delta u_{ij}) = 0,$$

$$B_2(v_{i1}) + \sum_{j=2}^k a_{j-1} B_2(\Delta u_{ij}) = 0,$$

and we have

$$\begin{aligned} B_3(v_{i1}) &= P_{ik+1}(\gamma) E(\Delta u_{ik} v_{i1}) + \Delta P_{ik+1}(\gamma) \left[E(y_{i1} v_{i1}) + \sum_{s=1}^{t-2} E(\Delta u_{is+1} v_{i1}) \right] \\ &= \Delta P_{ik+1}(\gamma) [\sigma_v^2 - \sigma_u^2], \end{aligned} \quad (A22)$$

$$\begin{aligned} \sum_{j=2}^k a_{j-1} B_3(\Delta u_{ij}) &= P_{ik+1}(\gamma) \sum_{j=2}^k a_{j-1} E(\Delta u_{ik} \Delta u_{ij}) + \Delta P_{ik+1}(\gamma) \left[\sum_{j=2}^k a_{j-1} E(y_{i1} \Delta u_{ij}) \right. \\ &\quad \left. + \sum_{j=2}^k a_{j-1} \sum_{s=1}^{t-2} E(\Delta u_{is+1} \Delta u_{ij}) \right] \\ &= P_{ik+1}(\gamma) (-\sigma_u^2) [a_{k-2} - 2a_{k-1}] + \Delta P_{ik+1}(\gamma) (-\sigma_u^2) (a_1 - a_0), \end{aligned} \quad (A23)$$

$$\begin{aligned}
E(\Delta y_{ik}(\gamma) \Delta u_{ik+1}) &= P_{ik+1}(\gamma) [\beta_1 E(\Delta y_{ik-1} \Delta u_{ik+1}) + \beta_2 E(\Delta y_{ik-1}(\gamma) \Delta u_{ik+1}) \\
&\quad + E(\Delta u_{ik} \Delta u_{ik+1})] + \Delta P_{ik+1}(\gamma) E(y_{ik-1} \Delta u_{ik+1}) \\
&= P_{ik+1}(\gamma) (-\sigma_u^2).
\end{aligned} \tag{A24}$$

Finally, we can obtain

$$\begin{aligned}
\frac{E(A_{k+1})}{a_{k+1}} &= E(\Delta y_{ik}(\gamma) v_{i1}) + \sum_{j=2}^k a_{j-1} E(\Delta y_{ik}(\gamma) \Delta u_{ij}) + a_k E(\Delta y_{ik}(\gamma) \Delta u_{ik+1}) \\
&= B_1(v_{i1}) + B_2(v_{i1}) + B_3(v_{i1}) + \sum_{j=2}^k a_{j-1} [B_1(\Delta u_{ij}) + B_2(\Delta u_{ij}) + B_3(\Delta u_{ij})] \\
&\quad + a_k E(\Delta y_{ik}(\gamma) \Delta u_{ik+1}) \\
&= B_1(v_{i1}) + \sum_{j=2}^k a_{j-1} B_3(\Delta u_{ij}) + a_k E(\Delta y_{ik}(\gamma) \Delta u_{ik+1}) \\
&= \Delta P_{ik+1}(\gamma) [\sigma_v^2 - \sigma_u^2] + P_{ik+1}(\gamma) (-\sigma_u^2) [a_{k-2} - 2a_{k-1}] + \Delta P_{ik+1}(\gamma) \\
&\quad \times (-\sigma_u^2) (a_1 - a_0) + P_{ik+1}(\gamma) (-\sigma_u^2) a_k \\
&= 0.
\end{aligned} \tag{A25}$$

That is, we have $E(A_{k+1}) = 0$. Through mathematical induction, $E(A_t) = 0$ is established for each t .

As shown in Hsiao et al. (2002), $\Phi = \mathfrak{g}\mathfrak{g}'$, where $\mathfrak{g} = (T, T-1, \dots, 1)'$. Therefore, we have

$$\begin{aligned}
E(\Delta u_i' \Phi \Delta u_i) &= E(\Delta u_i' \mathfrak{g})^2 \\
&= E \left[T v_{i1} + \sum_{j=2}^T (T-j+1) \Delta u_{ij} \right]^2 \\
&= T^2 E(v_{i1}^2) + \sum_{j=2}^T (T-j+1)^2 E(\Delta u_{ij}^2) + T(T-1) E(v_{i1} \Delta u_{i2}) \\
&\quad + 2 \sum_{j=3}^T (T-1+1)(T-j+2) E(\Delta u_{ij} \Delta u_{i,j-1}) \\
&= T^2 \sigma_v^2 + 2\sigma_u^2 \sum_{j=2}^T (T-j+1)^2 - 2T(T-1) \sigma_u^2 \\
&\quad - 2\sigma_u^2 \sum_{j=3}^T (T-1+1)(T-j+2) \\
&= T^2 \omega \sigma_u^2 - T(T-1) \sigma_u^2 \\
&= T \sigma_u^2 [1 + T(\omega - 1)].
\end{aligned} \tag{A26}$$

This completes the proof of Lemma 3. $\square$

Lemma 4. Under Assumption 1,

$$N^\alpha \left[\hat{\phi}(\gamma) - \phi_0 \right] \xrightarrow{p} H(\gamma)^{-1} K^*(\gamma) \text{ and } N^\alpha \left[\hat{\phi}(\gamma_0) - \phi_0 \right] \xrightarrow{p} 0, \tag{A27}$$

where $K^*(\gamma) = (K_{\theta_\delta}^*(\gamma)', K_{\sigma_u^2}^*(\gamma), K_\omega(\gamma))'$, in which

$$\begin{aligned}
K_{\theta_\delta}^*(\gamma) &= (M^*(\gamma, \gamma_0) - M^*(\gamma))c, \\
K_{\sigma_u^2}^*(\gamma) &= \frac{1}{\sigma_{u0}^4} E\left(c' \nabla \Delta y_{i,-1}(\gamma)' \Omega_0^{\star-1} \Delta u_i\right), \\
K_\omega^*(\gamma) &= \frac{1}{\sigma_{u0}^2 \varpi_0^2} E\left(c' \nabla \Delta y_{i,-1}^*(\gamma)' \Phi \Delta u_i\right).
\end{aligned}$$

Proof. First, we derive the limiting distribution of $\hat{\phi}(\gamma)$ using Theorem 10.1 in Wooldridge (1994). Assumption 1 ensures that condition (i) of Theorem 10.1 in Wooldridge (1994) is satisfied. Under the model setting, condition (ii) is also satisfied. We next verify condition (iv), and then check condition (iii).

For any fixed $\gamma \in \Gamma$, the first-order condition for $\hat{\phi}(\gamma)$ yields

$$\begin{aligned}
N^{\alpha-1} \frac{\partial \ln L}{\partial \theta_\delta}(\phi_0, \gamma) &= N^{\alpha-1} \sum_{i=1}^N \left[\Delta y_{i,-1}(\gamma)' \Omega_0^{-1} (\Delta y_i - \Delta y_{i,-1}(\gamma) \theta_{\delta 0}) \right] \\
&= \frac{-1}{N} \sum_{i=1}^N \Delta y_{i,-1}(\gamma)' \Omega_0^{-1} \nabla \Delta y_{i,-1}(\gamma) c \\
&\quad + N^{\alpha-1/2} \frac{1}{\sqrt{N}} \sum_{i=1}^N \Delta y_{i,-1}(\gamma)' \Omega_0^{-1} \Delta u_i \\
&\xrightarrow{p} (M^*(\gamma, \gamma_0) - M^*(\gamma))c \\
&= K_{\theta_\delta}^*(\gamma),
\end{aligned} \tag{A28}$$

$$\begin{aligned}
N^{\alpha-1} \frac{\partial \ln L}{\partial \sigma_u^2}(\phi_0, \gamma) &= N^{\alpha-1/2} \frac{1}{2\sigma_{u0}^4} \frac{1}{\sqrt{N}} \sum_{i=1}^N \left[\Delta u_i' \Omega_0^{\star-1} \Delta u_i - T \sigma_{u0}^2 \right] \\
&\quad + \frac{1}{N\sigma_{u0}^4} \sum_{i=1}^N c' \nabla \Delta y_{i,-1}(\gamma)' \Omega_0^{\star-1} \Delta u_i \\
&\quad + N^{-\alpha} \frac{1}{2N\sigma_{u0}^4} \sum_{i=1}^N c' \nabla \Delta y_{i,-1}(\gamma)' \Omega_0^{\star-1} \nabla \Delta y_{i,-1}(\gamma) c \\
&\xrightarrow{p} \frac{1}{\sigma_{u0}^4} E\left(c' \nabla \Delta y_{i,-1}(\gamma)' \Omega_0^{\star-1} \Delta u_i\right) \\
&= K_{\sigma_u^2}^*(\gamma),
\end{aligned} \tag{A29}$$

$$\begin{aligned}
N^{\alpha-1} \frac{\partial \ln L}{\partial \omega}(\phi_0, \gamma) &= N^{\alpha-1/2} \frac{1}{2\sigma_{u0}^2 \varpi_0^2} \frac{1}{\sqrt{N}} \sum_{i=1}^N \left[\Delta u_i' \Phi \Delta u_i - T \sigma_{u0}^2 \varpi_0 \right] \\
&\quad + \frac{1}{N\sigma_{u0}^2 \varpi_0^2} \sum_{i=1}^N c' \nabla \Delta y_{i,-1}^*(\gamma)' \Phi \Delta u_i \\
&\quad + N^{-\alpha} \frac{1}{2N\sigma_{u0}^2 \varpi_0^2} \sum_{i=1}^N c' \nabla \Delta y_{i,-1}^*(\gamma)' \Phi \nabla \Delta y_{i,-1}^*(\gamma) c \\
&\xrightarrow{p} \frac{1}{\sigma_{u0}^2 \varpi_0^2} E\left(c' \nabla \Delta y_{i,-1}^*(\gamma)' \Phi \Delta u_i\right) \\
&= K_\omega^*(\gamma).
\end{aligned} \tag{A30}$$

The second order condition for $\hat{\phi}(\gamma)$ yields

$$\begin{aligned} \frac{1}{N} \frac{\partial^2 \ln L}{\partial \theta_{\delta} \partial \theta'_{\delta}}(\phi_0, \gamma) &= \frac{-1}{N} \sum_{i=1}^N \Delta y_{i,-1}(\gamma)' \Omega_0^{-1} \Delta y_{i,-1}(\gamma) \\ &\xrightarrow{p} -E(\Delta y_{i,-1}(\gamma)' \Omega_0^{-1} \Delta y_{i,-1}(\gamma)) \\ &= -H_{\theta_{\delta} \theta'_{\delta}}(\gamma), \end{aligned} \quad (A31)$$

$$\begin{aligned} \frac{1}{N} \frac{\partial^2 \ln L}{\partial \theta_{\delta} \partial \sigma_u^2}(\phi_0, \gamma) &= \frac{-1}{\sigma_{u0}^4} \left[\frac{1}{N} \sum_{i=1}^N \Delta y_{i,-1}(\gamma)' \Omega_0^{-1} \nabla \Delta y_{i,-1}^*(\gamma) c N^{-\alpha} \right. \\ &\quad \left. + \frac{1}{N} \sum_{i=1}^N \Delta y_{i,-1}(\gamma)' \Omega_0^{\star-1} \Delta u_i \right] \\ &\xrightarrow{p} \frac{-1}{\sigma_{u0}^4} E(\Delta y_{i,-1}(\gamma)' \Omega_0^{\star-1} \Delta u_i) \\ &= -H_{\theta_{\delta} \sigma_u^2}(\gamma), \end{aligned} \quad (A32)$$

$$\begin{aligned} \frac{1}{N} \frac{\partial^2 \ln L}{\partial \theta_{\delta} \partial \omega}(\phi_0, \gamma) &= \frac{-1}{\sigma_{u0}^2 \varpi_0^2} \frac{1}{N} \sum_{i=1}^N [\Delta y_{i,-1}(\gamma)' \Phi(\Delta y_i - \Delta y_{i,-1}(\gamma) \theta_{\delta 0})] \\ &\xrightarrow{p} \frac{-1}{\sigma_{u0}^2 \varpi_0^2} E(\Delta y_{i,-1}(\gamma)' \Phi \Delta u_i) \\ &= -H_{\theta_{\delta} \omega}(\gamma), \end{aligned} \quad (A33)$$

$$\begin{aligned} \frac{1}{N} \frac{\partial^2 \ln L}{\partial \sigma_u^2 \partial \sigma_u^2}(\phi_0, \gamma) &= \frac{T}{2\sigma_{u0}^4} - \frac{1}{N\sigma_{u0}^6} \sum_{i=1}^N (\Delta y_i - \Delta y_{i,-1}(\gamma) \theta_{\delta 0})' \Omega_0^{\star-1} \\ &\quad \times (\Delta y_i - \Delta y_{i,-1}(\gamma) \theta_{\delta 0}) \\ &\xrightarrow{p} \frac{T}{2\sigma_{u0}^4} - \frac{1}{\sigma_{u0}^6} E(\Delta u_i' \Phi \Delta u_i) \\ &= -H_{\sigma_u^2 \sigma_u^2}, \end{aligned} \quad (A34)$$

$$\begin{aligned} \frac{1}{N} \frac{\partial^2 \ln L}{\partial \sigma_u^2 \partial \omega}(\phi_0, \gamma) &= \frac{-1}{\sigma_{u0}^4 \varpi_0^2} \frac{1}{N} \sum_{i=1}^N (\Delta y_i - \Delta y_{i,-1}(\gamma) \theta_{\delta 0})' \Phi (\Delta y_i - \Delta y_{i,-1}(\gamma) \theta_{\delta 0}) \\ &\xrightarrow{p} \frac{-1}{\sigma_{u0}^4 \varpi_0^2} E(\Delta u_i' \Phi \Delta u_i) \\ &= -H_{\sigma_u^2 \omega}, \end{aligned} \quad (A35)$$

$$\begin{aligned} \frac{1}{N} \frac{\partial^2 \ln L}{\partial \omega \partial \omega}(\phi_0, \gamma) &= \frac{T^2}{2\varpi_0^2} - \frac{1}{\sigma_{u0}^3 \varpi_0^3} \frac{1}{N} \sum_{i=1}^N (\Delta y_i - \Delta y_{i,-1}(\gamma) \theta_{\delta 0})' \Phi \\ &\quad \times (\Delta y_i - \Delta y_{i,-1}(\gamma) \theta_{\delta 0}) \\ &\xrightarrow{p} \frac{T^2}{2\varpi_0^2} - \frac{1}{\sigma_{u0}^3 \varpi_0^3} E(\Delta u_i' \Phi \Delta u_i) \\ &= -H_{\omega \omega}. \end{aligned} \quad (A36)$$

Next, condition (iii) of Wooldridge (1994, Theorem 10.1) requires that there exists a sequence of nonrandom positive definite diagonal matrices $\mathbb{C}_N$ and $\mathbb{A}_N$, such that $\mathbb{C}_N \mathbb{D}_N^{-1} = o(1)$, $\mathbb{A}_N \mathbb{B}_N^{-1} = o(1)$ and

$$\max_{\boldsymbol{\phi} \in \boldsymbol{\Phi}_N} \left\| \mathbb{C}_N^{-1} \left[\frac{\partial^2 \ln L}{\partial \boldsymbol{\phi} \partial \boldsymbol{\phi}'}(\boldsymbol{\phi}, \boldsymbol{\gamma}) - \frac{\partial^2 \ln L}{\partial \boldsymbol{\phi} \partial \boldsymbol{\phi}'}(\boldsymbol{\phi}_0, \boldsymbol{\gamma}) \right] \mathbb{C}_N^{-1} \right\| = o_p(1), \quad (\text{A37})$$

where $\boldsymbol{\Phi}_N = \{\boldsymbol{\phi} \in \boldsymbol{\Phi} : \|\mathbb{A}_N(\boldsymbol{\phi} - \boldsymbol{\phi}_0)\| \leq 1\}$ is a sequence of shrinking neighborhoods of $\boldsymbol{\phi}_0$.

To establish condition (iii), we let

$$\mathbb{C}_N = \mathbb{D}_N^{1-2\delta} = \mathbf{I}N^{-1/2+\delta}, \quad (\text{A38})$$

$$\mathbb{A}_N = \mathbb{B}_N^{1-2\delta} = \mathbf{I}N^{\alpha(1-2\delta)}, \quad (\text{A39})$$

for some $0 < \delta < \min\{\alpha/[2(1+\alpha)], 1/2\}$, and let $\boldsymbol{\phi} \in \boldsymbol{\Phi}_N$ be arbitrarily chosen. We set

$$\boldsymbol{\phi} = \boldsymbol{\phi}_0 + \boldsymbol{\zeta}N^{-\alpha(1-2\delta)}, \quad (\text{A40})$$

such that $\|\boldsymbol{\zeta}\| \leq 1$.

It is easily seen that

$$\boldsymbol{\Omega}^{-1} = \frac{1}{\sigma_u^2} \boldsymbol{\Omega}^{\star-1} = \frac{1}{\sigma_u^2 \varpi} \times \begin{bmatrix} \xi_{11}(\omega) & \xi_{21}(\omega) & \cdots & \xi_{T1}(\omega) \\ \xi_{21}(\omega) & \xi_{22}(\omega) & \cdots & \xi_{T2}(\omega) \\ \vdots & \vdots & \ddots & \vdots \\ \xi_{T1}(\omega) & \xi_{T2}(\omega) & \cdots & \xi_{TT}(\omega) \end{bmatrix}, \quad (\text{A41})$$

where $\xi_{ij}(\omega) = [T - (i - 1)][(j - 1)\omega - (j - 2)]$, and

$$\xi_{ij}(\omega) - \xi_{ij}(\omega_0) = [T - (i - 1)](j - 1)(\omega - \omega_0) = O(N^{-\alpha(1-2\delta)}). \quad (\text{A42})$$

Therefore, we obtain $\boldsymbol{\Omega}^{-1} - \boldsymbol{\Omega}_0^{-1} = O(N^{-\alpha(1-2\delta)})$ and $\boldsymbol{\Omega}^{\star-1} - \boldsymbol{\Omega}_0^{\star-1} = O(N^{-\alpha(1-2\delta)})$. Furthermore, we have

$$\begin{aligned} & N^{2\delta-1} \left[\frac{\partial^2 \ln L}{\partial \boldsymbol{\theta}_\delta \partial \boldsymbol{\theta}_\delta'}(\boldsymbol{\phi}, \boldsymbol{\gamma}) - \frac{\partial^2 \ln L}{\partial \boldsymbol{\theta}_\delta \partial \boldsymbol{\theta}_\delta'}(\boldsymbol{\phi}_0, \boldsymbol{\gamma}) \right] \\ &= \frac{-1}{N} \sum_{i=1}^N \Delta y_{i-1}(\boldsymbol{\gamma})' [\boldsymbol{\Omega}^{-1} - \boldsymbol{\Omega}_0^{-1}] \Delta y_{i-1}(\boldsymbol{\gamma}) N^{2\delta} \\ &= O_p(N^{-\alpha(1-2\delta)}) O(N^{2\delta}) = O_p(N^{2\delta-\alpha(1-2\delta)}) = o_p(1), \end{aligned} \quad (\text{A43})$$

$$\begin{aligned} & N^{2\delta-1} \left[\frac{\partial^2 \ln L}{\partial \boldsymbol{\theta}_\delta \partial \sigma_u^2}(\boldsymbol{\phi}, \boldsymbol{\gamma}) - \frac{\partial^2 \ln L}{\partial \boldsymbol{\theta}_\delta \partial \sigma_u^2}(\boldsymbol{\phi}_0, \boldsymbol{\gamma}) \right] \\ &= N^{2\delta} \left[\frac{-1}{\sigma_u^2} + \frac{1}{\sigma_{u0}^2} \right] \frac{1}{N} \sum_{i=1}^N \Delta y_{i-1}(\boldsymbol{\gamma})' \boldsymbol{\Omega}^{\star-1} [\Delta y_i - \Delta y_{i-1}(\boldsymbol{\gamma}) \boldsymbol{\theta}_\delta] \\ &\quad + N^{2\delta} \frac{-1}{\sigma_{u0}^4 N} \sum_{i=1}^N \Delta y_{i-1}(\boldsymbol{\gamma})' [\boldsymbol{\Omega}^{\star-1} - \boldsymbol{\Omega}_0^{\star-1}] [\Delta y_i - \Delta y_{i-1}(\boldsymbol{\gamma}) \boldsymbol{\theta}_\delta] \\ &\quad + \frac{1}{\sigma_{u0}^4 N} \sum_{i=1}^N \Delta y_{i-1}(\boldsymbol{\gamma})' \boldsymbol{\Omega}_0^{\star-1} \Delta y_{i-1}(\boldsymbol{\gamma}) (\boldsymbol{\theta}_\delta - \boldsymbol{\theta}_{\delta 0}) \\ &= O(N^{2\delta}) O_p(N^{-\alpha(1-2\delta)}) = O_p(N^{2\delta-\alpha(1-2\delta)}) = o_p(1), \end{aligned} \quad (\text{A44})$$

$$\begin{aligned}
& N^{2\delta-1} \left[\frac{\partial^2 \ln L}{\partial \theta_\delta \partial \omega}(\boldsymbol{\phi}, \boldsymbol{\gamma}) - \frac{\partial^2 \ln L}{\partial \theta_\delta \partial \omega}(\boldsymbol{\phi}_0, \boldsymbol{\gamma}) \right] \\
&= N^{2\delta} \left[\frac{-1}{\sigma_u^2 \varpi^2} - \frac{-1}{\sigma_{u0}^2 \varpi_0^2} \right] \frac{1}{N} \sum_{i=1}^N \Delta y_{i,-1}(\boldsymbol{\gamma})' \boldsymbol{\Phi} [\Delta y_i - \Delta y_{i,-1}(\boldsymbol{\gamma}) \boldsymbol{\theta}_\delta] \\
&\quad + N^{2\delta} \frac{1}{\sigma_{u0}^2 \varpi_0^2 N} \sum_{i=1}^N \Delta y_{i,-1}(\boldsymbol{\gamma})' \boldsymbol{\Phi} \Delta y_{i,-1}(\boldsymbol{\gamma}) (\boldsymbol{\theta}_\delta - \boldsymbol{\theta}_{\delta 0}) \\
&= O(N^{2\delta}) O_p(N^{-\alpha(1-2\delta)}) = O_p(N^{2\delta-\alpha(1-2\delta)}) = o_p(1),
\end{aligned} \tag{A45}$$

$$\begin{aligned}
& N^{2\delta-1} \left[\frac{\partial^2 \ln L}{\partial \sigma_u^2 \partial \sigma_u^2}(\boldsymbol{\phi}, \boldsymbol{\gamma}) - \frac{\partial^2 \ln L}{\partial \sigma_u^2 \partial \sigma_u^2}(\boldsymbol{\phi}_0, \boldsymbol{\gamma}) \right] \\
&= N^{2\delta} \left[\frac{T}{2\sigma_u^4} - \frac{T}{\sigma_{u0}^4} \right] - N^{2\delta} \left[\frac{1}{\sigma_u^6} - \frac{1}{\sigma_{u0}^6} \right] \frac{1}{N} \sum_{i=1}^N [\Delta y_i - \Delta y_{i,-1}(\boldsymbol{\gamma}) \boldsymbol{\theta}_\delta]' \\
&\quad \times \boldsymbol{\Omega}^{\star-1} [\Delta y_i - \Delta y_{i,-1}(\boldsymbol{\gamma}) \boldsymbol{\theta}_\delta] \\
&\quad + N^{2\delta} \frac{1}{\sigma_{u0}^6 N} \sum_{i=1}^N (\boldsymbol{\theta}_\delta - \boldsymbol{\theta}_{\delta 0})' \Delta y_{i,-1}(\boldsymbol{\gamma})' \boldsymbol{\Omega}^{\star-1} [\Delta y_i - \Delta y_{i,-1}(\boldsymbol{\gamma}) \boldsymbol{\theta}_\delta] \\
&\quad - N^{2\delta} \frac{1}{\sigma_{u0}^6 N} \sum_{i=1}^N [\Delta y_i - \Delta y_{i,-1}(\boldsymbol{\gamma}) \boldsymbol{\theta}_\delta]' [\boldsymbol{\Omega}^{\star-1} - \boldsymbol{\Omega}_0^{\star-1}] [\Delta y_i - \Delta y_{i,-1}(\boldsymbol{\gamma}) \boldsymbol{\theta}_\delta] \\
&\quad + N^{2\delta} \frac{1}{\sigma_{u0}^6 N} \sum_{i=1}^N [\Delta y_i - \Delta y_{i,-1}(\boldsymbol{\gamma}) \boldsymbol{\theta}_\delta]' \boldsymbol{\Omega}^{\star-1} \Delta y_{i,-1}(\boldsymbol{\gamma}) (\boldsymbol{\theta}_\delta - \boldsymbol{\theta}_{\delta 0}) \\
&= O(N^{2\delta}) O_p(N^{-\alpha(1-2\delta)}) = O_p(N^{2\delta-\alpha(1-2\delta)}) = o_p(1).
\end{aligned} \tag{A46}$$

Similarly, it can be proved that

$$\begin{aligned}
& N^{2\delta-1} \left[\frac{\partial^2 \ln L}{\partial \sigma_u^2 \partial \omega}(\boldsymbol{\phi}, \boldsymbol{\gamma}) - \frac{\partial^2 \ln L}{\partial \sigma_u^2 \partial \omega}(\boldsymbol{\phi}_0, \boldsymbol{\gamma}) \right] = O_p(N^{2\delta-\alpha(1-2\delta)}) = o_p(1), \\
& N^{2\delta-1} \left[\frac{\partial^2 \ln L}{\partial \omega \partial \omega}(\boldsymbol{\phi}, \boldsymbol{\gamma}) - \frac{\partial^2 \ln L}{\partial \omega \partial \omega}(\boldsymbol{\phi}_0, \boldsymbol{\gamma}) \right] = O_p(N^{2\delta-\alpha(1-2\delta)}) = o_p(1).
\end{aligned}$$

Hence, the condition (iii) is satisfied. This completes the proof of Lemma 4. $\square$

Lemma 5. Under Assumption 1, $\hat{\boldsymbol{\gamma}} \rightarrow_p \boldsymbol{\gamma}_0$.

Proof. Note that

$$\ln L(\boldsymbol{\gamma}) = -\frac{NT}{2} \ln(2\pi) - \frac{N}{2} \ln |\hat{\boldsymbol{\Omega}}(\boldsymbol{\gamma})| - \frac{1}{2} \sum_{i=1}^N \Delta u_i(\boldsymbol{\gamma})' \hat{\boldsymbol{\Omega}}(\boldsymbol{\gamma})^{-1} \Delta u_i(\boldsymbol{\gamma}) \tag{A47}$$

$$+ \ln L^*(\boldsymbol{\gamma}), \tag{A48}$$

where

$$\begin{aligned}
\ln L^*(\gamma) &= \frac{-1}{2} \sum_{i=1}^N \mathbf{c}' \nabla \Delta y_{i,-1}^*(\gamma)' \hat{\Omega}(\gamma)^{-1} \nabla \Delta y_{i,-1}^*(\gamma) \mathbf{c} N^{-2\alpha} \\
&\quad \times \frac{-1}{2} \sum_{i=1}^N \left[\hat{\theta}_\delta(\gamma) - \theta_{\delta 0} \right]' \Delta y_{i,-1}(\gamma)' \hat{\Omega}(\gamma)^{-1} \Delta y_{i,-1}(\gamma) \left[\hat{\theta}_\delta(\gamma) - \theta_{\delta 0} \right] \\
&\quad - \sum_{i=1}^N \mathbf{c}' \nabla \Delta y_{i,-1}^*(\gamma)' \hat{\Omega}(\gamma)^{-1} \Delta u_i N^{-\alpha} \\
&\quad + \sum_{i=1}^N \mathbf{c}' \nabla \Delta y_{i,-1}^*(\gamma)' \hat{\Omega}(\gamma)^{-1} \Delta y_{i,-1}(\gamma) \left[\hat{\theta}_\delta(\gamma) - \theta_{\delta 0} \right] N^{-\alpha} \\
&\quad + \sum_{i=1}^N \Delta u_i' \hat{\Omega}(\gamma)^{-1} \Delta y_{i,-1}(\gamma) \left[\hat{\theta}_\delta(\gamma) - \theta_{\delta 0} \right].
\end{aligned} \tag{A49}$$

According to Lemma 4, the three terms on the right-hand side of (A47) are asymptotic constants. Therefore $\max \ln L(\gamma)$ and $\max \ln L^*(\gamma)$ are asymptotically equivalent. Next, we derive the distribution of $\max \ln L^*(\gamma)$.

We rewrite $\ln L^*(\gamma)$ as

$$2N^{2\alpha-1} \ln L^*(\gamma) = S_1 + S_2 - 2(S_3 - S_4 - S_5), \tag{A50}$$

where

$$\begin{aligned}
S_1 &= \frac{-1}{N} \sum_{i=1}^N \mathbf{c}' \nabla \Delta y_{i,-1}^*(\gamma)' \hat{\Omega}(\gamma)^{-1} \nabla \Delta y_{i,-1}^*(\gamma) \mathbf{c} = O_p(1) \\
S_2 &= \frac{-1}{N} \sum_{i=1}^N N^\alpha \left[\hat{\theta}_\delta(\gamma) - \theta_{\delta 0} \right]' \Delta y_{i,-1}(\gamma)' \hat{\Omega}(\gamma)^{-1} \Delta y_{i,-1}(\gamma) \left[\hat{\theta}_\delta(\gamma) - \theta_{\delta 0} \right] N^\alpha \\
&= O_p(1) \\
S_3 &= N^{\alpha-1/2} \frac{1}{\sqrt{N}} \sum_{i=1}^N \mathbf{c}' \nabla \Delta y_{i,-1}^*(\gamma)' \hat{\Omega}(\gamma)^{-1} \Delta u_i = o_p(1) \\
S_4 &= \frac{1}{N} \sum_{i=1}^N \mathbf{c}' \nabla \Delta y_{i,-1}^*(\gamma)' \hat{\Omega}(\gamma)^{-1} \Delta y_{i,-1}(\gamma) \left[\hat{\theta}_\delta(\gamma) - \theta_{\delta 0} \right] N^\alpha = O_p(1) \\
S_5 &= N^{\alpha-1/2} \frac{1}{\sqrt{N}} \sum_{i=1}^N \Delta u_i' \hat{\Omega}(\gamma)^{-1} \Delta y_{i,-1}(\gamma) \left[\hat{\theta}_\delta(\gamma) - \theta_{\delta 0} \right] N^\alpha = o_p(1).
\end{aligned} \tag{A51}$$

We also have following result

$$\begin{aligned}
S_1 + S_2 - 2S_4 &= \frac{-1}{N} \sum_{i=1}^N \left[\nabla \Delta y_{i,-1}^*(\gamma) \mathbf{c} - \Delta y_{i,-1}(\gamma) \left[\hat{\theta}_\delta(\gamma) - \theta_{\delta 0} \right] N^\alpha \right]' \hat{\Omega}(\gamma)^{-1} \\
&\quad \times \left[\nabla \Delta y_{i,-1}^*(\gamma) \mathbf{c} - \Delta y_{i,-1}(\gamma) \left[\hat{\theta}_\delta(\gamma) - \theta_{\delta 0} \right] N^\alpha \right] \\
&\leq 0.
\end{aligned} \tag{A52}$$

Therefore, $N^{2\alpha-1} \ln L^*(\gamma)$ will be asymptotically equal to or less than 0, and $N^{2\alpha-1} \ln L^*(\gamma)$ is asymptotically equal to 0 if and only if $\gamma = \gamma_0$, using Theorem 2.1 of Newey and McFadden (1994) we have $\hat{\gamma} \rightarrow_p \gamma_0$. $\square$

Lemma 6. Under Assumption 1, set $\gamma = \gamma_0 + \nu h$, and $h \rightarrow 0$ as $N \rightarrow \infty$,

$$\frac{P(A_{it}(\gamma) | \mathbf{s}_{it})}{\mathbf{s}_{it}'\gamma - \mathbf{s}_{it}'\gamma_0} \rightarrow f_{t|s}(\mathbf{s}_{it}'\gamma_0 | \mathbf{s}_{it}), \quad (\text{A53})$$

$$\frac{P(A_{ik}(\gamma) | A_{it}(\gamma), \mathbf{s}_{ik}, \mathbf{s}_{ik})}{\mathbf{s}_{ik}'\gamma - \mathbf{s}_{ik}'\gamma_0} \rightarrow f_{k|t,s}(\mathbf{s}_{ik}'\gamma_0 | A_{it}^0, \mathbf{s}_{ik}, \mathbf{s}_{it}). \quad (\text{A54})$$

and for some random variable X ,

$$h^{-1}E(X \nabla I_{it}(\gamma)^2 | \mathbf{s}_{it}) \rightarrow E(X | A_{it}^0, \mathbf{s}_{it}) f_{t|s}(\mathbf{s}_{it}'\gamma_0 | \mathbf{s}_{it}) |\mathbf{s}_{it}'\nu|, \quad (\text{A55})$$

$$\begin{aligned} h^{-2}E(X \nabla I_{it}(\gamma) \nabla I_{ik}(\gamma) | \mathbf{s}_{it}) &\rightarrow E(X | A_{ik}^0, A_{it}^0, \mathbf{s}_{ik}, \mathbf{s}_{it}) f_{t|s}(\mathbf{s}_{it}'\gamma_0 | \mathbf{s}_{it}) \mathbf{s}_{it}'\nu \\ &\quad \times f_{k|t,s}(\mathbf{s}_{ik}'\gamma_0 | A_{it}^0, \mathbf{s}_{ik}, \mathbf{s}_{it}) \mathbf{s}_{ik}'\nu, \end{aligned} \quad (\text{A56})$$

as $N \rightarrow \infty$, where $A_{it}(\gamma) = \{\mathbf{s}_{it}'\gamma_0 < q_{it} \leq \mathbf{s}_{it}'\gamma\}$, $A_{it}^0 = \{q_{it} = \mathbf{s}_{it}'\gamma_0\}$.

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

$$\begin{aligned} \mathbf{s}_{it}'\gamma - \mathbf{s}_{it}'\gamma_0 &= \mathbf{s}_{it}'(\gamma - \gamma_0) = \mathbf{s}_{it}'\nu h \\ &\rightarrow 0. \end{aligned}$$

Denote $h^* = \mathbf{s}_{it}'\gamma - \mathbf{s}_{it}'\gamma_0$, then we have $h^* \rightarrow 0$ as $N \rightarrow \infty$ and

$$\begin{aligned} \frac{P(A_{it}(\gamma) | \mathbf{s}_{it})}{\mathbf{s}_{it}'\gamma - \mathbf{s}_{it}'\gamma_0} &= \frac{P(q_{it} \leq \mathbf{s}_{it}'\gamma | \mathbf{s}_{it}) - P(q_{it} \leq \mathbf{s}_{it}'\gamma_0 | \mathbf{s}_{it})}{\mathbf{s}_{it}'\gamma - \mathbf{s}_{it}'\gamma_0} \\ &= \frac{P(q_{it} \leq \mathbf{s}_{it}'\gamma_0 + h^* | \mathbf{s}_{it}) - P(q_{it} \leq \mathbf{s}_{it}'\gamma_0 | \mathbf{s}_{it})}{h^*}. \end{aligned} \quad (\text{A57})$$

By the definition of a Differential, we have

$$\begin{aligned} \left. \frac{\partial P(a | \mathbf{s}_{it})}{\partial a} \right|_{a=\mathbf{s}_{it}'\gamma_0} &\equiv \lim_{h^* \rightarrow 0} \frac{P(q_{it} \leq \mathbf{s}_{it}'\gamma_0 + h^* | \mathbf{s}_{it}) - P(q_{it} \leq \mathbf{s}_{it}'\gamma_0 | \mathbf{s}_{it})}{h^*} \\ &= f_{t|s}(\mathbf{s}_{it}'\gamma_0 | \mathbf{s}_{it}). \end{aligned} \quad (\text{A58})$$

A similar argument can prove (A54).

Next, note that

$$I(q_{it} \leq \mathbf{s}_{it}'\gamma) \leq I(q_{it} \leq \mathbf{s}_{it}'\gamma_0) \Leftrightarrow \mathbf{s}_{it}'\gamma \leq \mathbf{s}_{it}'\gamma_0, \quad (\text{A59})$$

$$I(q_{it} \leq \mathbf{s}_{it}'\gamma) > I(q_{it} \leq \mathbf{s}_{it}'\gamma_0) \Leftrightarrow \mathbf{s}_{it}'\gamma > \mathbf{s}_{it}'\gamma_0, \quad (\text{A60})$$

(A59) and (A60) imply that

$$\begin{aligned} \mathbf{s}_{it}'\nu \times \text{sgn}(\nabla I_{it}(\gamma)) &= \mathbf{s}_{it}'\nu \times \text{sgn}(I_{it}(\gamma) - I_{it}(\gamma_0)) \\ &= \mathbf{s}_{it}'\nu \times \text{sgn}(\mathbf{s}_{it}'\gamma - \mathbf{s}_{it}'\gamma_0) \\ &= |\mathbf{s}_{it}'\nu|, \end{aligned} \quad (\text{A61})$$

where $\text{sgn}(x)$ is the sign function, i.e. $\text{sgn}(x) = I(x > 0) - I(x \leq 0)$.

Therefore, we have

$$\begin{aligned}
& h^{-1}E(X\nabla I_{it}(\gamma)^2 | \mathbf{s}_{it}) \\
&= E(X | A_{it}(\gamma), \mathbf{s}_{it}) \frac{P(A_{it}(\gamma) | \mathbf{s}_{it})}{\mathbf{s}_{it}'\gamma - \mathbf{s}_{it}'\gamma_0} \times \mathbf{s}_{it}'\nu \times \text{sgn}(\nabla I_{it}(\gamma)) \\
&\rightarrow E(X | A_{it}^0, \mathbf{s}_{it}) f_{t|s}(\mathbf{s}_{it}'\gamma_0 | \mathbf{s}_{it}) |\mathbf{s}_{it}'\nu|,
\end{aligned} \tag{A62}$$

and

$$\begin{aligned}
& h^{-2}E(X\nabla I_{it}(\gamma)\nabla I_{ik}(\gamma) | \mathbf{s}_{it}) \\
&= E(X | A_{ik}(\gamma), A_{it}(\gamma), \mathbf{s}_{ik}, \mathbf{s}_{it}) \times \frac{P(A_{it}(\gamma) | \mathbf{s}_{ik}, \mathbf{s}_{it})}{\mathbf{s}_{it}'\gamma - \mathbf{s}_{it}'\gamma_0} \times \mathbf{s}_{it}'\nu \\
&\quad \times \frac{P(A_{ik}(\gamma) | A_{it}(\gamma), \mathbf{s}_{ik}, \mathbf{s}_{it})}{\mathbf{s}_{ik}'\gamma - \mathbf{s}_{ik}'\gamma_0} \times \mathbf{s}_{ik}'\nu \\
&\rightarrow E(X | A_{ik}^0, A_{it}^0, \mathbf{s}_{ik}, \mathbf{s}_{it}) f_{t|s}(\mathbf{s}_{it}'\gamma_0 | \mathbf{s}_{it}) \times \mathbf{s}_{it}'\nu \\
&\quad \times f_{k|t,s}(\mathbf{s}_{ik}'\gamma_0 | A_{it}^0, \mathbf{s}_{ik}, \mathbf{s}_{it}) \times \mathbf{s}_{ik}'\nu.
\end{aligned} \tag{A63}$$

This completes the proof of Lemma 6. $\square$

Lemma 7. Under Assumption 1, on a compact set $\Psi = \times_{n=1}^{d+1} \Psi_n$, we have the following uniform convergence results,

$$G_N(\nu) \xrightarrow{p} G(\nu), \tag{A64}$$

$$V_N(\nu) \xrightarrow{p} V(\nu). \tag{A65}$$

Proof. By Lemma 6 (setting $h = \lambda_N^{-1}$), we can obtain

$$\begin{aligned}
& \lambda_N E(w_{ij}^2 \nabla I_{ij+1}(\gamma)^2 | \mathbf{s}_{ij+1}) \\
&\rightarrow E(w_{ij}^2 | A_{ij+1}^0, \mathbf{s}_{ij+1}) f_{j+1|s}(\mathbf{s}_{ij+1}'\gamma_0 | \mathbf{s}_{ij+1}) |\mathbf{s}_{ij+1}'\nu| \\
&= D_{ij}^f(\nu),
\end{aligned} \tag{A66}$$

$$\begin{aligned}
& \lambda_N E(w_{ij}^2 e_{it} e_{ik} \nabla I_{ij+1}(\gamma)^2 | \mathbf{s}_{ij+1}) \\
&\rightarrow E(w_{ij}^2 e_{it} e_{ik} | A_{ij+1}^0, \mathbf{s}_{ij+1}) f_{j+1|s}(\mathbf{s}_{ij+1}'\gamma_0 | \mathbf{s}_{ij+1}) |\mathbf{s}_{ij+1}'\nu| \\
&= V_{ijtk}^f(\nu),
\end{aligned} \tag{A67}$$

$$\begin{aligned}
\lambda_N E(w_{ij}^2 \nabla I_{ij+1}(\gamma)^2) &= \lambda_N E(E(w_{ij}^2 \nabla I_{ij+1}(\gamma)^2 | \mathbf{s}_{ij+1})) \\
&\rightarrow E(D_{ij}^f(\nu)),
\end{aligned} \tag{A68}$$

$$\begin{aligned}
\lambda_N E(w_{ij}^2 e_{it} e_{ik} \nabla I_{ij+1}(\gamma)^2) &= \lambda_N E(E(w_{ij}^2 e_{it} e_{ik} \nabla I_{ij+1}(\gamma)^2 | \mathbf{s}_{ij+1})) \\
&\rightarrow E(V_{ijtk}^f(\nu)).
\end{aligned} \tag{A69}$$

To prove (A64), then (A65), we note that

$$\begin{aligned}
& \lambda_N E \left(\sum_{j=1}^{t-1} a_j \nabla \Delta \mathbf{w}_{ij}(\boldsymbol{\gamma})' \mathbf{c} \right)^2 \\
&= \sum_{j=1}^{t-1} a_j^2 \left[\lambda_N E \left(w_{ij}^2 \nabla I_{ij+1}(\boldsymbol{\gamma})^2 \right) + \lambda_N E \left(w_{ij-1}^2 \nabla I_{ij}(\boldsymbol{\gamma})^2 \right) \right] \\
&\quad - 2 \sum_{j=1}^{t-2} a_j a_{j+1} \lambda_N E \left(w_{ij}^2 \nabla I_{ij+1}(\boldsymbol{\gamma})^2 \right) + O(\lambda_N^{-1}) \\
&\rightarrow \sum_{j=1}^{t-1} a_j \left[E \left(D_{ij}^f(\mathbf{v}) \right) + E \left(D_{ij-1}^f(\mathbf{v}) \right) \right] - 2 \sum_{j=1}^{t-2} a_j a_{j+1} E \left(D_{ij}^f(\mathbf{v}) \right) \\
&= D_{1t}(\mathbf{v}).
\end{aligned} \tag{A70}$$

Hence, we have

$$\begin{aligned}
E(G_N(\mathbf{v})) &= \lambda_N E \left(\mathbf{c}' \nabla \Delta y_{i,-1}^*(\boldsymbol{\gamma})' \boldsymbol{\Omega}^{\star-1} \nabla \Delta y_{i,-1}^*(\boldsymbol{\gamma}) \mathbf{c} \right) \\
&= \sum_{t=2}^T [a_t a_{t-1}]^{-1} \lambda_N E \left(\sum_{j=1}^{t-1} a_j \nabla \Delta \mathbf{w}_{is}(\boldsymbol{\gamma})' \mathbf{c} \right)^2 \\
&\rightarrow \sum_{t=2}^T [a_t a_{t-1}]^{-1} D_{1t}(\mathbf{v}) = G(\mathbf{v}).
\end{aligned} \tag{A71}$$

Note that

$$\begin{aligned}
E(V_N(\mathbf{v})) &= \lambda_N E \left(\mathbf{c}' \nabla \Delta y_{i,-1}^*(\boldsymbol{\gamma}) \boldsymbol{\Omega}_0^{\star-1} \Delta u_i \right)^2 \\
&= \lambda_N E \left[\sum_{t=2}^T [a_t a_{t-1}]^{-1} \left(\sum_{j=1}^{t-1} a_j \nabla \Delta \mathbf{w}_{ij}(\boldsymbol{\gamma})' \mathbf{c} \right) e_{it} \right]^2 \\
&= \lambda_N \sum_{t=2}^T [a_t a_{t-1}]^{-2} \left(\sum_{j=1}^{t-1} a_j \nabla \Delta \mathbf{w}_{ij}(\boldsymbol{\gamma})' \mathbf{c} e_{it} \right)^2 \\
&\quad + 2 \lambda_N \sum_{t=2}^{T-1} \sum_{k=t+1}^T [a_t a_{t-1}]^{-1} [a_k a_{k-1}]^{-1} \\
&\quad \times E \left[\left(\sum_{j=1}^{t-1} a_j \nabla \Delta \mathbf{w}_{ij}(\boldsymbol{\gamma})' \mathbf{c} e_{it} \right) \left(\sum_{j=1}^{k-1} a_j \nabla \Delta \mathbf{w}_{ij}(\boldsymbol{\gamma})' \mathbf{c} e_{ik} \right) \right].
\end{aligned} \tag{A72}$$

Thus, the first term on the right-hand-side of (A72)

$$\begin{aligned}
& \lambda_N E \left(\sum_{j=1}^{t-1} a_j \nabla \Delta \mathbf{w}_{ij}(\boldsymbol{\gamma})' \mathbf{c} e_{it} \right)^2 \\
&= \sum_{j=1}^{t-1} a_j^2 \left[\lambda_N E \left(w_{ij}^2 e_{it}^2 \nabla I_{ij+1}(\boldsymbol{\gamma})^2 \right) + \lambda_N E \left(w_{ij-1}^2 e_{it}^2 \nabla I_{ij}(\boldsymbol{\gamma})^2 \right) \right] \\
&\quad - 2 \sum_{j=1}^{t-2} a_j a_{j+1} \lambda_N E \left(w_{ij}^2 e_{it}^2 \nabla I_{ij+1}(\boldsymbol{\gamma})^2 \right) + O(\lambda_N^{-1})
\end{aligned}$$

$$\begin{aligned}
& \rightarrow \sum_{j=1}^{t-1} a_j^2 \left[E(V_{ijt}^f(\mathbf{v})) + E(V_{ij-1t}^f(\mathbf{v})) \right] - 2 \sum_{j=1}^{t-2} a_j a_{j+1} E(V_{ijt}^f(\mathbf{v})) \\
& = V_{1t}(\mathbf{v}).
\end{aligned} \tag{A73}$$

The second term on the right-hand-side of (A72),

$$\begin{aligned}
& \lambda_N E \left[\left(\sum_{j=1}^{t-1} a_j \nabla \Delta \mathbf{w}_{ij}(\boldsymbol{\gamma})' \mathbf{c} e_{it} \right) \left(\sum_{j=1}^{k-1} a_j \nabla \Delta \mathbf{w}_{ij}(\boldsymbol{\gamma})' \mathbf{c} e_{ik} \right) \right] \\
& = \lambda_N E \left[\left(\sum_{j=1}^{t-1} a_j \nabla \Delta \mathbf{w}_{ij}(\boldsymbol{\gamma})' \mathbf{c} e_{it} \right) \left(\sum_{j=1}^{t-1} a_j \nabla \Delta \mathbf{w}_{ij}(\boldsymbol{\gamma})' \mathbf{c} e_{ik} \right) \right] \\
& \quad - \lambda_N a_t a_{t-1} E(z_{it-1}^2 e_{it} e_{ik} \nabla I_{it}(\boldsymbol{\gamma})^2) + O(\lambda_N^{-1}) \\
& = \sum_{j=1}^{t-1} a_j^2 \left[\lambda_N E(w_{ij}^2 e_{it} e_{ik} \nabla I_{ij+1}(\boldsymbol{\gamma})^2) + \lambda_N E(w_{ij-1}^2 e_{it} e_{ik} \nabla I_{ij}(\boldsymbol{\gamma})^2) \right] \\
& \quad - 2 \sum_{j=1}^{t-2} a_j a_{j+1} \lambda_N E(w_{ij}^2 e_{it} e_{ik} \nabla I_{ij+1}(\boldsymbol{\gamma})^2) \\
& \quad - \lambda_N a_t a_{t-1} E(z_{it-1}^2 e_{it} e_{ik} \nabla I_{it}(\boldsymbol{\gamma})^2) + O(\lambda_N^{-1}) \\
& \rightarrow \sum_{j=1}^{t-1} a_j^2 \left[E(V_{ijt}^f(\mathbf{v})) + E(V_{ij-1t}^f(\mathbf{v})) \right] - 2 \sum_{j=1}^{t-2} a_j a_{j+1} E(V_{ijt}^f(\mathbf{v})) \\
& \quad - a_t a_{t-1} E(V_{it-1t}^f(\mathbf{v})) \\
& = V_{2tk}(\mathbf{v}).
\end{aligned} \tag{A74}$$

Finally, we can obtain below result

$$\begin{aligned}
E(V_N(\mathbf{v})) & = \lambda_N E \left(\mathbf{c}' \nabla \Delta y_{i,-1}^*(\boldsymbol{\gamma}) \boldsymbol{\Omega}_0^{\star-1} \Delta u_i \right)^2 \\
& \rightarrow \sum_{t=2}^T [a_t a_{t-1}]^{-1} V_{1t}(\mathbf{v}) + 2 \sum_{t=2}^{T-1} \sum_{k=t+1}^T [a_t a_{t-1} a_k a_{k-1}]^{-1} V_{2tk}(\mathbf{v}) \\
& = V(\mathbf{v}).
\end{aligned} \tag{A75}$$

Arguments similar to those in the proof of Lemma A.10 of Hansen (2000) show that (A71) and (A75) imply (A64) and (A65). This completes the proof of Lemma 7. $\square$

Lemma 8. Under Assumption 1, on any compact set $\boldsymbol{\Psi} = \times_{n=1}^{d+1} \Psi_n$,

$$R_N(\mathbf{v}) \xrightarrow{d} R(\mathbf{v}), \tag{A76}$$

where $R(\mathbf{v})$ is a Gaussian process with a positive variance $V(\mathbf{v})$ when $\mathbf{v} \neq \mathbf{0}$.

Proof. Note that,

$$\begin{aligned}
E(R_N^2(\mathbf{v})) &= E\left(\sqrt{\lambda_N} \mathbf{c}' \nabla \Delta y_{i,-1}^*(\boldsymbol{\gamma}) \boldsymbol{\Omega}_0^{*-1} \Delta u_i\right)^2 \\
&= E(V_N(\mathbf{v})) \\
&\rightarrow V(\mathbf{v}),
\end{aligned} \tag{A77}$$

by Lemma 7. Under Assumption 1, $\sqrt{\lambda_N} \mathbf{c}' \nabla \Delta y_{i,-1}^*(\boldsymbol{\gamma}) \boldsymbol{\Omega}_0^{*-1} \Delta u_i$ is a i.i.d. random sequence across i , so $R_N(\mathbf{v})$ converges pointwise to a zero-mean Gaussian process with covariance kernel $V(\mathbf{v})$ by the CLT. Using arguments identical to those in the proof of Lemma A.11 of Hansen (2000) is sufficient to establish the tightness of $R_N(\mathbf{v})$. $\square$

Lemma 9. Under Assumption 1, set $\boldsymbol{\gamma} = \boldsymbol{\gamma}_0 + \mathbf{v}h$, and $h \rightarrow 0$ as $N \rightarrow \infty$,

$$\sqrt{N}[\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}) - \boldsymbol{\phi}_0] \xrightarrow{d} \mathbf{H}(\boldsymbol{\gamma}_0)^{-1} \mathbf{K}(\boldsymbol{\gamma}_0) \text{ and } \sqrt{N}[\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}) - \hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}_0)] \xrightarrow{p} \mathbf{0}, \tag{A78}$$

where $\mathbf{K}(\boldsymbol{\gamma})$ is a zero-mean Gaussian process with covariance kernel $E(\mathbf{h}_i(\boldsymbol{\gamma})\mathbf{h}_i(\boldsymbol{\gamma})')$.

Proof. Observe that

$$\begin{aligned}
N^{-1/2} \frac{\partial \ln L}{\partial \boldsymbol{\theta}_\delta}(\boldsymbol{\phi}_0, \boldsymbol{\gamma}) &= N^{-1/2} \sum_{i=1}^N \left[\Delta y_{i,-1}(\boldsymbol{\gamma})' \boldsymbol{\Omega}_0^{-1} (\Delta y_i - \Delta y_{i,-1}(\boldsymbol{\gamma}) \boldsymbol{\theta}_{\delta 0}) \right] \\
&= \frac{-1}{\sqrt{\lambda_N}} \frac{\lambda_N}{N} \sum_{i=1}^N \Delta y_{i,-1}(\boldsymbol{\gamma})' \boldsymbol{\Omega}_0^{-1} \nabla \Delta y_{i,-1}(\boldsymbol{\gamma}) \mathbf{c} \\
&\quad + \frac{1}{\sqrt{N}} \sum_{i=1}^N \Delta y_{i,-1}(\boldsymbol{\gamma})' \boldsymbol{\Omega}_0^{-1} \Delta u_i \\
&= \frac{1}{\sqrt{N}} \sum_{i=1}^N \mathbf{h}_{1i}(\boldsymbol{\gamma}) + O_p(\lambda_N^{-1/2}), \\
N^{-1/2} \frac{\partial \ln L}{\partial \sigma_u^2}(\boldsymbol{\phi}_0, \boldsymbol{\gamma}) &= \frac{1}{2\sigma_{u0}^4} \frac{1}{\sqrt{N}} \sum_{i=1}^N \left[\Delta u_i' \boldsymbol{\Omega}_0^{*-1} \Delta u_i - T\sigma_{u0}^2 \right] \\
&\quad + \frac{1}{\sqrt{\lambda_N}} \frac{\lambda_N}{N\sigma_{u0}^4} \sum_{i=1}^N \mathbf{c}' \nabla \Delta y_{i,-1}(\boldsymbol{\gamma})' \boldsymbol{\Omega}_0^{*-1} \Delta u_i \\
&\quad + \frac{1}{\sqrt{N}} \frac{\lambda_N}{2N\sigma_{u0}^4} \sum_{i=1}^N \mathbf{c}' \nabla \Delta y_{i,-1}(\boldsymbol{\gamma})' \boldsymbol{\Omega}_0^{*-1} \nabla \Delta y_{i,-1}(\boldsymbol{\gamma}) \mathbf{c} \\
&= \frac{1}{\sqrt{N}} \sum_{i=1}^N h_{2i} + O_p(\lambda_N^{-1/2}), \\
N^{-1/2} \frac{\partial \ln L}{\partial \omega}(\boldsymbol{\phi}_0, \boldsymbol{\gamma}) &= \frac{1}{2\sigma_{u0}^2 \varpi_0^2} \frac{1}{\sqrt{N}} \sum_{i=1}^N \left[\Delta u_i' \boldsymbol{\Phi} \Delta u_i - T\sigma_{u0}^2 \varpi_0 \right] \\
&\quad + \frac{1}{\sqrt{\lambda_N}} \frac{\lambda_N}{N\sigma_{u0}^2 \varpi_0^2} \sum_{i=1}^N \mathbf{c}' \nabla \Delta y_{i,-1}^*(\boldsymbol{\gamma})' \boldsymbol{\Phi} \Delta u_i \\
&\quad + \frac{1}{\sqrt{N}} \frac{\lambda_N}{2N\sigma_{u0}^2 \varpi_0^2} \sum_{i=1}^N \mathbf{c}' \nabla \Delta y_{i,-1}^*(\boldsymbol{\gamma})' \boldsymbol{\Phi} \nabla \Delta y_{i,-1}^*(\boldsymbol{\gamma}) \mathbf{c} \\
&= \frac{1}{\sqrt{N}} \sum_{i=1}^N h_{3i} + O_p(\lambda_N^{-1/2}),
\end{aligned}$$

A similar argument can be made to establish the limiting distribution of $\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma})$ through the proof of Lemma 4. Therefore, we can obtain

$$\begin{aligned}\sqrt{N}[\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}) - \boldsymbol{\phi}_0] &= \mathbf{H}_N(\boldsymbol{\phi}_0, \boldsymbol{\gamma})^{-1} \mathbf{K}_N(\boldsymbol{\phi}_0, \boldsymbol{\gamma}) + o_p(1) \\ &\xrightarrow{d} \mathbf{H}(\boldsymbol{\gamma}_0)^{-1} \mathbf{K}(\boldsymbol{\gamma}_0).\end{aligned}$$

Obviously, we have

$$\begin{aligned}\sqrt{N}[\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}) - \hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}_0)] &= \sqrt{N}[\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}) - \boldsymbol{\phi}_0] - \sqrt{N}[\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}_0) - \boldsymbol{\phi}_0] \\ &\xrightarrow{d} \mathbf{H}(\boldsymbol{\gamma}_0)^{-1} \mathbf{K}(\boldsymbol{\gamma}_0) - \mathbf{H}(\boldsymbol{\gamma}_0)^{-1} \mathbf{K}(\boldsymbol{\gamma}_0) \\ &= \mathbf{0}.\end{aligned}\tag{A79}$$

□

Proof of Theorem 1. Using the argument in the proof of Lemma A.9 in Hansen (2000) and Theorem 2 in Yang et al. (2021b), we can establish the convergence rate of the threshold estimator λ_N . Thus, we can study their asymptotic behavior in the neighborhood of the true thresholds, $\hat{\boldsymbol{\gamma}} = \boldsymbol{\gamma}_0 + \mathbf{v}/\lambda_N$.

Let $L_1(\boldsymbol{\gamma}) = \sum_{i=1}^N \Delta u_i \hat{\boldsymbol{\Omega}}(\boldsymbol{\gamma})^{-1} \Delta u_i / 2 - \sum_{i=1}^N \Delta \hat{u}_i(\boldsymbol{\gamma}_0)' \hat{\boldsymbol{\Omega}}(\boldsymbol{\gamma})^{-1} \Delta \hat{u}_i(\boldsymbol{\gamma}_0) / 2$ and $\hat{\mathbf{c}}(\boldsymbol{\gamma}) = N^\alpha \hat{\boldsymbol{\theta}}(\boldsymbol{\gamma})$. By the definition of the threshold estimator, we have

$$\lambda_N(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^0) = \hat{\mathbf{v}} = \arg \max_{\mathbf{v} \in \Psi} \ln L\left(\boldsymbol{\gamma}_0 + \frac{\mathbf{v}}{\lambda_N}\right),\tag{A80}$$

because $L_1(\boldsymbol{\gamma})$ is an asymptotic constant, $\max \ln L(\boldsymbol{\gamma}_0 + \mathbf{v}/\lambda_N)$ and $\max(\ln L^*(\boldsymbol{\gamma}_0 + \mathbf{v}/\lambda_N) - L_1(\boldsymbol{\gamma}))$ are asymptotically equivalent. Next, we derive the distribution of $\ln L^*(\boldsymbol{\gamma}_0 + \mathbf{v}/\lambda_N) - L_1(\boldsymbol{\gamma})$.

As above, we have $\ln L^*(\boldsymbol{\gamma}_0 + \mathbf{v}/\lambda_N) - L_1(\boldsymbol{\gamma}) = -[Q_1 + Q_2 - 2(Q_3 + Q_4 - Q_5)]/2$. We next derive the limiting behavior of each Q_i , respectively. Thus, by Lemmas 1–8, let $\boldsymbol{\gamma} = \boldsymbol{\gamma}_0 + \mathbf{v}/\lambda_N$, we have the following results

$$\begin{aligned}Q_1 &= \frac{\lambda_N}{N} \sum_{i=1}^N \hat{\mathbf{c}}(\boldsymbol{\gamma})' \nabla \Delta y_{i-1}^*(\boldsymbol{\gamma})' \hat{\boldsymbol{\Omega}}(\boldsymbol{\gamma})^{-1} \nabla \Delta y_{i-1}^*(\boldsymbol{\gamma}) \hat{\mathbf{c}}(\boldsymbol{\gamma}) \\ &\xrightarrow{p} \frac{1}{\sigma_{u0}^2} G(\mathbf{v}), \\ Q_2 &= \frac{1}{N} \sum_{i=1}^N \sqrt{N} [\hat{\boldsymbol{\theta}}_\delta(\boldsymbol{\gamma}) + \hat{\boldsymbol{\theta}}_\delta(\boldsymbol{\gamma}_0) - 2\boldsymbol{\theta}_{\delta 0}]' \Delta y_{i-1}(\boldsymbol{\gamma})' \hat{\boldsymbol{\Omega}}(\boldsymbol{\gamma})^{-1} \Delta y_{i-1}(\boldsymbol{\gamma}) \\ &\quad \times \sqrt{N} [\hat{\boldsymbol{\theta}}_\delta(\boldsymbol{\gamma}) - \hat{\boldsymbol{\theta}}_\delta(\boldsymbol{\gamma}_0)] \\ &= O_p(1) O_p(1) o_p(1) = o_p(1), \\ Q_3 &= \sqrt{N} [\hat{\boldsymbol{\theta}}_\delta(\boldsymbol{\gamma}) - \hat{\boldsymbol{\theta}}_\delta(\boldsymbol{\gamma}_0)] \frac{1}{\sqrt{N}} \sum_{i=1}^N \mathbf{c}' \Delta y_{i-1}(\boldsymbol{\gamma})' \hat{\boldsymbol{\Omega}}(\boldsymbol{\gamma})^{-1} \Delta u_i = o_p(1), \\ Q_4 &= \sqrt{\frac{\lambda_N}{N}} \sum_{i=1}^N \hat{\mathbf{c}}(\boldsymbol{\gamma})' \nabla \Delta y_{i-1}^*(\boldsymbol{\gamma})' \hat{\boldsymbol{\Omega}}(\boldsymbol{\gamma})^{-1} \Delta u_i \\ &\xrightarrow{d} \frac{1}{\sigma_{u0}^2} R(\mathbf{v}), \\ Q_5 &= \sqrt{N} [\hat{\boldsymbol{\theta}}_\delta(\boldsymbol{\gamma}) - \boldsymbol{\theta}_{\delta 0}] \frac{1}{N} \sum_{i=1}^N \Delta y_{i-1}(\boldsymbol{\gamma})' \hat{\boldsymbol{\Omega}}(\boldsymbol{\gamma})^{-1} \nabla \Delta y_{i-1}^*(\boldsymbol{\gamma}) \sqrt{N} \hat{\boldsymbol{\theta}}_\delta(\boldsymbol{\gamma})\end{aligned}$$

$$= O_p(1)o_p(1)O_p(1) = o_p(1),$$

and therefore,

$$\ln L^*(\gamma_0 + \nu/\lambda_N) - L_1(\gamma) \xrightarrow{d} \frac{-1}{2\sigma_{u0}^2} [G(\nu) - 2R(\nu)], \quad (\text{A81})$$

and

$$\begin{aligned} \lambda_N(\hat{\gamma} - \gamma_0) &\xrightarrow{d} \arg \max_{\nu \in \mathbb{R}^{d+1}} \frac{-1}{2\sigma_{u0}^2} [G(\nu) - 2R(\nu)] \\ &= \arg \min_{\nu \in \mathbb{R}^{k+1}} \left[\frac{1}{2} G(\nu) - R(\nu) \right]. \end{aligned} \quad (\text{A82})$$

This completes the proof of Theorem 1. $\square$

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

First, a similar argument in the proof of Lemma 4 can be made to establish the limiting distribution of $\hat{\phi}(\gamma)$. Under $H_0^1: \beta_2 = 0$ and $\theta_2 = \mathbf{0}$, we have

$$\Delta y_i = \Delta y_{i,-1} \theta_{\delta 0} + \Delta u_i. \quad (\text{A83})$$

Therefore, for any fixed $\gamma \in \Gamma$, we have

$$\begin{aligned} \frac{1}{\sqrt{N}} \frac{\partial \ln L}{\partial \theta_\delta}(\phi_0, \gamma) &= \frac{1}{\sqrt{N}} \sum_{i=1}^N \Delta y_{i,-1}(\gamma)' \Omega_0^{-1} \Delta u_i \\ &\equiv K_{N\theta_\delta}(\phi_0, \gamma), \end{aligned} \quad (\text{A84})$$

$$\begin{aligned} \frac{1}{\sqrt{N}} \frac{\partial \ln L}{\partial \sigma_u^2}(\phi_0, \gamma) &= \frac{1}{2\sigma_{u0}^4} \frac{1}{\sqrt{N}} \sum_{i=1}^N [\Delta u_i' \Omega_0^{\star-1} \Delta u_i - T\sigma_{u0}^2] \\ &\equiv K_{N\sigma_u^2}(\phi_0, \gamma), \end{aligned} \quad (\text{A85})$$

$$\begin{aligned} \frac{1}{\sqrt{N}} \frac{\partial \ln L}{\partial \omega}(\phi_0, \gamma) &= \frac{1}{2\sigma_{u0}^2 \varpi_0^2} \frac{1}{\sqrt{N}} \sum_{i=1}^N [\Delta u_i' \Phi \Delta u_i - T\sigma_{u0}^2 \varpi_0] \\ &\equiv K_{N\omega}(\phi_0, \gamma), \end{aligned} \quad (\text{A86})$$

The second-order condition for $\hat{\phi}(\gamma)$ yields

$$\begin{aligned} \frac{1}{N} \frac{\partial^2 \ln L}{\partial \theta_\delta \partial \theta_\delta'}(\phi_0, \gamma) &= \frac{-1}{N} \sum_{i=1}^N \Delta y_{i,-1}(\gamma)' \Omega_0^{-1} \Delta y_{i,-1}(\gamma) \\ &\equiv -H_{N\theta_\delta \theta_\delta'}(\phi_0, \gamma), \end{aligned} \quad (\text{A87})$$

$$\begin{aligned} \frac{1}{N} \frac{\partial^2 \ln L}{\partial \theta_\delta \partial \sigma_u^2}(\phi_0, \gamma) &= \frac{-1}{\sigma_{u0}^4} \frac{1}{N} \sum_{i=1}^N \Delta y_{i,-1}(\gamma)' \Omega_0^{\star-1} \Delta u_i \\ &\equiv -H_{N\theta_\delta \sigma_u^2}(\phi_0, \gamma), \end{aligned} \quad (\text{A88})$$

$$\begin{aligned} \frac{1}{N} \frac{\partial^2 \ln L}{\partial \theta_\delta \partial \omega}(\phi_0, \gamma) &= \frac{-1}{\sigma_{u0}^2 \varpi_0^2} \frac{1}{N} \sum_{i=1}^N \Delta y_{i,-1}(\gamma)' \Phi \Delta u_i \\ &\equiv -H_{N\theta_\delta \omega}(\phi_0, \gamma), \end{aligned} \quad (\text{A89})$$

$$\begin{aligned}\frac{1}{N} \frac{\partial^2 \ln L}{\partial \sigma_u^2 \partial \sigma_u^2}(\phi_0, \gamma) &= \frac{T}{2\sigma_{u0}^4} - \frac{1}{N\sigma_{u0}^6} \sum_{i=1}^N \Delta u_i' \Omega_0^{\star-1} \Delta u_i \\ &\equiv -H_{N\sigma_u^2 \sigma_u^2}(\phi_0, \gamma),\end{aligned}\quad (\text{A90})$$

$$\begin{aligned}\frac{1}{N} \frac{\partial^2 \ln L}{\partial \sigma_u^2 \partial \omega}(\phi_0, \gamma) &= \frac{-1}{\sigma_{u0}^4 \varpi_0^2} \frac{1}{N} \sum_{i=1}^N \Delta u_i' \Phi \Delta u_i \\ &\equiv -H_{N\sigma_u^2 \omega}(\phi_0, \gamma),\end{aligned}\quad (\text{A91})$$

$$\begin{aligned}\frac{1}{N} \frac{\partial^2 \ln L}{\partial \omega \partial \omega}(\phi_0, \gamma) &= \frac{T^2}{2\varpi_0^2} - \frac{1}{\sigma_{u0}^3 \varpi_0^3} \frac{1}{N} \sum_{i=1}^N \Delta u_i' \Phi \Delta u_i \\ &\equiv -H_{N\omega \omega}(\phi_0, \gamma).\end{aligned}\quad (\text{A92})$$

Hence, we can obtain

$$\begin{aligned}\sqrt{N}(\hat{\phi}(\gamma) - \phi_0) &= H_N(\phi_0, \gamma)^{-1} K_N(\phi_0, \gamma) + o_p(1), \\ &\xrightarrow{d} H(\gamma)^{-1} K(\gamma),\end{aligned}\quad (\text{A93})$$

where $K_N(\phi, \gamma) = (K_{N\theta_\delta}(\phi, \gamma)', K_{N\sigma_u^2}(\phi, \gamma), K_{N\omega}(\phi, \gamma))'$,

$$H_N(\phi, \gamma) = \begin{bmatrix} H_{N\theta_\delta \theta_\delta}(\phi, \gamma) & H_{N\theta_\delta \sigma_u^2}(\phi, \gamma) & H_{N\theta_\delta \omega}(\phi, \gamma) \\ H_{N\theta_\delta \sigma_u^2}(\phi, \gamma)' & H_{N\sigma_u^2 \sigma_u^2}(\phi, \gamma) & H_{N\sigma_u^2 \omega}(\phi, \gamma) \\ H_{N\theta_\delta \omega}(\phi, \gamma)' & H_{N\sigma_u^2 \omega}(\phi, \gamma) & H_{N\omega \omega}(\phi, \gamma) \end{bmatrix}. \quad (\text{A94})$$

Denote $\tilde{\phi} = \arg \max_{\phi \in \Phi_{c1}} \ln L(\phi, \gamma)$, where $\Phi_{c1} = \{\phi \in \Phi : \beta_2 = 0, \theta_2 = \mathbf{0}\}$. Next, we derive the limit distribution of $\tilde{\phi} - \hat{\phi}(\gamma)$. Under $H_0^1 : \beta_2 = 0, \theta_2 = \mathbf{0}$, the model degenerates into a dynamic panel linear model when estimating parameters using the restriction of $\beta_2 = 0, \theta_2 = \mathbf{0}$, in which case Hsiao et al. (2002) shows $\tilde{\phi} \xrightarrow{p} \phi_0$. Hence, a Taylor's expansion of $\partial \ln L(\tilde{\phi}, \gamma) / \partial \phi$ at ϕ_0 gives

$$K_N(\tilde{\phi}, \gamma) = K_N(\phi_0, \gamma) - H_N(\phi_0, \gamma) \sqrt{N}(\tilde{\phi} - \phi_0) + o_p(1). \quad (\text{A95})$$

Substituting (A95) into the first-order condition for $\tilde{\phi}$ yields

$$\begin{bmatrix} H_N(\phi_0, \gamma) & R' \\ R & \mathbf{0} \end{bmatrix} \begin{bmatrix} \sqrt{N}(\tilde{\phi} - \phi_0) \\ N^{-1/2} \tilde{\lambda}_1 \end{bmatrix} = \begin{bmatrix} K_N(\phi_0, \gamma) + o_p(1) \\ \mathbf{0} \end{bmatrix}, \quad (\text{A96})$$

where R is a matrix such that $R\phi = (\beta_2, \theta_2')' = \mathbf{0}$ and $\tilde{\lambda}_1$ is the solution of the first-order conditions of Lagrangian multipliers. $\sqrt{N}(\tilde{\phi} - \phi_0)$ can be solved using the partition inverse rule

$$\begin{aligned}\sqrt{N}(\tilde{\phi} - \phi_0) &= [H_N^{-1}(\phi_0, \gamma) - H_N^{-1}(\phi_0, \gamma) R' (R H_N^{-1}(\phi_0, \gamma) R')^{-1} \\ &\quad \times R H_N^{-1}(\phi_0, \gamma)] K_N(\phi_0, \gamma) + o_p(1).\end{aligned}\quad (\text{A97})$$

From (A93) and (A97), we have

$$\begin{aligned}\sqrt{N}(\tilde{\boldsymbol{\phi}} - \boldsymbol{\phi}_0) &= \sqrt{N}(\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}) - \boldsymbol{\phi}_0) + o_p(1) \\ &\quad \times \mathbf{H}_N^{-1}(\boldsymbol{\phi}_0, \boldsymbol{\gamma}) \mathbf{R}' (\mathbf{R} \mathbf{H}_N^{-1}(\boldsymbol{\phi}_0, \boldsymbol{\gamma}) \mathbf{R}')^{-1} \mathbf{R} \mathbf{H}_N^{-1}(\boldsymbol{\phi}_0, \boldsymbol{\gamma}) \\ &\quad \times \mathbf{K}_N(\boldsymbol{\phi}_0, \boldsymbol{\gamma}) + o_p(1).\end{aligned}\quad (\text{A98})$$

Note that the above equation can be rewritten as

$$\begin{aligned}\sqrt{N}(\tilde{\boldsymbol{\phi}} - \hat{\boldsymbol{\phi}}(\boldsymbol{\gamma})) &= \mathbf{H}_N^{-1}(\boldsymbol{\phi}_0, \boldsymbol{\gamma}) \mathbf{R}' (\mathbf{R} \mathbf{H}_N^{-1}(\boldsymbol{\phi}_0, \boldsymbol{\gamma}) \mathbf{R}')^{-1} \mathbf{R} \mathbf{H}_N^{-1}(\boldsymbol{\phi}_0, \boldsymbol{\gamma}) \\ &\quad \times \mathbf{K}_N(\boldsymbol{\phi}_0, \boldsymbol{\gamma}) + o_p(1),\end{aligned}\quad (\text{A99})$$

we therefore obtain $\tilde{\boldsymbol{\phi}} - \hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}) \xrightarrow{p} \mathbf{0}$.

Next, using the Taylor's expansion of $\text{SSR}_N(\tilde{\boldsymbol{\phi}}, \boldsymbol{\gamma})$ at $\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma})$ we have

$$\begin{aligned}\ln L(\tilde{\boldsymbol{\phi}}, \boldsymbol{\gamma}) &= \ln L(\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}), \boldsymbol{\gamma}) + \mathbf{K}_N(\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}), \boldsymbol{\gamma}) \sqrt{N}[\tilde{\boldsymbol{\phi}} - \hat{\boldsymbol{\phi}}(\boldsymbol{\gamma})] \\ &\quad - \frac{1}{2} \sqrt{N}[\tilde{\boldsymbol{\phi}} - \hat{\boldsymbol{\phi}}(\boldsymbol{\gamma})]' \mathbf{H}_N(\boldsymbol{\phi}_0, \boldsymbol{\gamma}) \sqrt{N}[\tilde{\boldsymbol{\phi}} - \hat{\boldsymbol{\phi}}(\boldsymbol{\gamma})] + o_p(1).\end{aligned}\quad (\text{A100})$$

Consequently, we have

$$\begin{aligned}\ln L(\tilde{\boldsymbol{\phi}}, \boldsymbol{\gamma}) - \ln L(\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}), \boldsymbol{\gamma}) &= -\sqrt{N}[\tilde{\boldsymbol{\phi}} - \hat{\boldsymbol{\phi}}(\boldsymbol{\gamma})]' \mathbf{H}_N(\boldsymbol{\phi}_0, \boldsymbol{\gamma}) \sqrt{N}[\tilde{\boldsymbol{\phi}} - \hat{\boldsymbol{\phi}}(\boldsymbol{\gamma})] / 2 + o_p(1) \\ &\xrightarrow{d} -\mathbf{K}(\boldsymbol{\gamma})' \mathbf{H}(\boldsymbol{\gamma})^{-1} \mathbf{R}' [\mathbf{R} \mathbf{H}(\boldsymbol{\gamma})^{-1} \mathbf{R}']^{-1} \mathbf{R} \mathbf{H}(\boldsymbol{\gamma})^{-1} \mathbf{K}(\boldsymbol{\gamma}) / 2,\end{aligned}\quad (\text{A101})$$

and using the Taylor's expansion of $\ln L(\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}), \boldsymbol{\gamma})$ at $\boldsymbol{\phi}_0$ we have

$$\begin{aligned}\ln L(\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}), \boldsymbol{\gamma}) &= \ln L(\boldsymbol{\phi}_0, \boldsymbol{\gamma}) + \mathbf{K}_N(\boldsymbol{\phi}_0, \boldsymbol{\gamma}) \sqrt{N}[\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}) - \boldsymbol{\phi}_0] \\ &\quad - \frac{1}{2} \sqrt{N}[\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}) - \boldsymbol{\phi}_0]' \mathbf{H}_N(\boldsymbol{\phi}_0, \boldsymbol{\gamma}) \sqrt{N}[\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}) - \boldsymbol{\phi}_0] + o_p(1) \\ &= -\frac{NT}{2} \ln(2\pi) - \frac{N}{2} \ln |\boldsymbol{\Omega}_0| - \frac{1}{2} \sum_{i=1}^N \Delta u_i' \boldsymbol{\Omega}_0^{-1} \Delta u_i + O_p(1),\end{aligned}\quad (\text{A102})$$

and we obtain

$$\begin{aligned}\frac{1}{N(T-1)} \ln L(\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}), \boldsymbol{\gamma}) &= -\frac{T}{2(T-1)} \ln(2\pi) - \frac{1}{2(T-1)} \ln |\boldsymbol{\Omega}_0| \\ &\quad - \frac{1}{2N(T-1)} \sum_{i=1}^N \Delta u_i' \boldsymbol{\Omega}_0^{-1} \Delta u_i + O_p(N^{-1}) \\ &\xrightarrow{p} -\frac{T}{2(T-1)} \ln(2\pi) - \frac{1}{2(T-1)} \ln |\boldsymbol{\Omega}_0| - \frac{T}{2(T-1)} \\ &\equiv -\sigma_T^2 / 2.\end{aligned}\quad (\text{A103})$$

Finally, we have the following result

$$\begin{aligned}
F_1 &= \frac{S_0 - S(\hat{\gamma})}{S(\hat{\gamma})/N(T-1)} = \frac{\ln L(\tilde{\phi}, \hat{\gamma}) - \ln L(\hat{\phi}(\hat{\gamma}), \hat{\gamma})}{\ln L(\hat{\phi}(\hat{\gamma}), \hat{\gamma})/[N(T-1)]} \\
&= \sup_{\gamma \in \Gamma} \frac{\ln L(\tilde{\phi}, \gamma) - \ln L(\hat{\phi}(\gamma), \gamma)}{\ln L(\hat{\phi}(\gamma), \gamma)/[N(T-1)]} \\
&\xrightarrow{d} \frac{1}{\sigma_T^2} \sup_{\gamma \in \Gamma} K(\gamma)' H(\gamma)^{-1} R' [RH(\gamma)^{-1} R']^{-1} RH(\gamma)^{-1} K(\gamma).
\end{aligned} \tag{A104}$$

Next, we derive the limiting distribution of F_C . Denote

$$\hat{\phi}(\gamma) = \arg \min_{\phi \in \Phi} \ln L(\phi, \gamma), \tag{A105}$$

$$\tilde{\gamma} = \arg \min_{\gamma \in \Gamma_c} \ln L(\hat{\phi}(\gamma), \gamma), \tag{A106}$$

where $\Gamma_c = \{\gamma \in \Gamma : \gamma_s = \mathbf{0}\}$ and $\tilde{\gamma} = (\tilde{\gamma}_0, \tilde{\gamma}_s')' = (\tilde{\gamma}_0, \mathbf{0}')'$.

Under $H_0^2 : \gamma_s = \mathbf{0}$, the model degenerates to a constant threshold model when estimating parameters using the restriction of $\gamma_s = \mathbf{0}$, in which case a proof similar to Theorem 1 can be used, so we have the following result

$$\begin{aligned}
\ln L(\tilde{\gamma}) &= \max_{\gamma \in \Gamma_c} \ln L(\gamma) \\
&= \max_{\nu \in \Psi_c} \ln L\left(\gamma_0 + \frac{\nu}{\lambda_N}\right) \\
&\xrightarrow{d} \min_{\nu \in \mathbb{R}_c^{d+1}} \left[\frac{1}{2} G(\nu) - R(\nu) \right],
\end{aligned} \tag{A107}$$

where $\Psi_c = \{\nu \in \Psi : \nu_s = \mathbf{0}\}$, $\mathbb{R}_c^{d+1} = \{\nu \in \mathbb{R}^{d+1} : \nu_s = \mathbf{0}\}$. From (A102), we have

$$\ln L(\hat{\gamma}) = -\frac{NT}{2} \ln(2\pi) - \frac{N}{2} \ln |\hat{\Omega}(\hat{\gamma})| - \frac{1}{2} \sum_{i=1}^N \Delta u_i(\hat{\gamma})' \hat{\Omega}(\hat{\gamma})^{-1} \Delta u_i(\hat{\gamma}) + o_p(1),$$

and

$$\frac{1}{N(T-1)} \ln L(\hat{\gamma}) \xrightarrow{p} -\sigma_T^2/2. \tag{A108}$$

Finally, we have the following result

$$\begin{aligned}
F_C &= \frac{S_c - S(\hat{\gamma})}{S(\hat{\gamma})/N(T-1)} = \frac{\ln L(\tilde{\gamma}) - \ln L(\hat{\gamma})}{\ln L(\hat{\gamma})/[N(T-1)]} \\
&= \frac{\ln L(\tilde{\gamma}) - \ln L(\hat{\phi}(\gamma_0), \gamma_0)}{\ln L(\hat{\gamma})/[N(T-1)]} + \frac{\ln L(\hat{\phi}(\gamma_0), \gamma_0) - \ln L(\hat{\gamma})}{\ln L(\hat{\gamma})/[N(T-1)]} \\
&\xrightarrow{d} \frac{-2}{\sigma_T^2} \left[\left(\min_{\nu \in \mathbb{R}_c^{d+1}} G(\nu) - 2R(\nu) \right) - \left(\min_{\nu \in \mathbb{R}^{d+1}} G(\nu) - 2R(\nu) \right) \right].
\end{aligned} \tag{A109}$$

This completes the proof of Theorem 2. $\square$

We next consider the asymptotic results of the GMM estimator. To this end, we denote

$$\Delta y_{i-1}^*(\gamma) = \begin{bmatrix} \Delta y_{i1}(\gamma) & \Delta \mathbf{x}_{i2}(\gamma) \\ \Delta y_{i2}(\gamma) & \Delta \mathbf{x}_{i3}(\gamma) \\ \vdots & \vdots \\ \Delta y_{iT-1}(\gamma) & \Delta \mathbf{x}_{iT}(\gamma) \end{bmatrix}, \tag{A110}$$

and $\mathbf{G}_{\varphi,2}(\gamma) = E(\mathbf{z}'_i \Delta \dot{\mathbf{y}}^*_{i,-1}(\gamma))$, $\tilde{\mathbf{G}}_{\varphi}(\gamma) = \mathbf{S}^{-1/2} \mathbf{G}_{\varphi}(\gamma)$, $\tilde{\mathbf{G}}_{\varphi,2}(\gamma) = \mathbf{S}^{-1/2} \mathbf{G}_{\varphi,2}(\gamma)$.

Lemma 10. Under Assumption 2, as $N \rightarrow \infty$

$$N^{\alpha}(\hat{\boldsymbol{\varphi}}(\gamma) - \boldsymbol{\varphi}_0) \xrightarrow{p} - [\mathbf{G}_{\varphi}(\gamma)' \mathbf{S}^{-1} \mathbf{G}_{\varphi}(\gamma)]^{-1} \mathbf{G}_{\varphi}(\gamma)' \mathbf{S}^{-1} \nabla \mathbf{G}_{\varphi,2}(\gamma) \mathbf{c}, \quad (\text{A111})$$

where $\nabla \mathbf{G}_{\varphi,2}(\gamma) = \mathbf{G}_{\varphi,2}(\gamma) - \mathbf{G}_{\varphi,2}(\gamma_0)$.

Proof. Note that

$$\begin{aligned} \Delta \dot{\mathbf{y}}_i &= \Delta \dot{\mathbf{y}}_{i,-1}(\gamma_0) \boldsymbol{\varphi}_0 + \dot{\mathbf{u}}_i \\ &= \Delta \dot{\mathbf{y}}_{i,-1}(\gamma) \boldsymbol{\varphi}_0 + \dot{\mathbf{u}}_i - [\Delta \dot{\mathbf{y}}_{i,-1}(\gamma) - \Delta \dot{\mathbf{y}}_{i,-1}(\gamma_0)] \boldsymbol{\varphi}_0 \\ &= \Delta \dot{\mathbf{y}}_{i,-1}(\gamma) \boldsymbol{\varphi}_0 + \dot{\mathbf{u}}_i - \nabla \Delta \dot{\mathbf{y}}^*_{i,-1}(\gamma) \mathbf{c} N^{-\alpha}, \end{aligned} \quad (\text{A112})$$

where $\nabla \Delta \dot{\mathbf{y}}^*_{i,-1}(\gamma) = \Delta \dot{\mathbf{y}}^*_{i,-1}(\gamma) - \Delta \dot{\mathbf{y}}^*_{i,-1}(\gamma_0)$.

By Lemma 1,

$$\mathbf{G}_{N,\varphi}(\gamma) \xrightarrow{p} \mathbf{G}_{\varphi}(\gamma), \quad \mathbf{G}_{N,\varphi,2}(\gamma) \xrightarrow{p} \mathbf{G}_{\varphi,2}(\gamma), \quad (\text{A113})$$

uniformly over $\gamma \in \Gamma$.

Therefore, we have

$$\begin{aligned} N^{\alpha}(\hat{\boldsymbol{\varphi}}(\gamma) - \boldsymbol{\varphi}_0) &= [\mathbf{G}_{N,\varphi}(\gamma)' \mathbf{S}_N^{-1} \mathbf{G}_{\varphi}(N, \gamma)]^{-1} \mathbf{G}_{N,\varphi}(\gamma)' \mathbf{S}_N^{-1} \\ &\quad \times \left[N^{\alpha-1/2} \frac{1}{\sqrt{N}} \sum_{i=1}^N \mathbf{z}'_i \Delta \dot{\mathbf{u}}_i - \nabla \mathbf{G}_{N,\varphi,2}(\gamma) \mathbf{c} \right] \\ &\xrightarrow{p} - [\mathbf{G}_{\varphi}(\gamma)' \mathbf{S}^{-1} \mathbf{G}_{\varphi}(\gamma)]^{-1} \mathbf{G}_{\varphi}(\gamma)' \mathbf{S}^{-1} \nabla \mathbf{G}_{\varphi,2}(\gamma) \mathbf{c}, \end{aligned} \quad (\text{A114})$$

uniformly over $\gamma \in \Gamma$. □

Lemma 11. Under Assumption 2, as $N \rightarrow \infty$, $\hat{\gamma} \rightarrow_p \gamma_0$.

Proof. Note that, by Lemma 10, for any $\gamma \in \Gamma$, we have

$$\begin{aligned} N^{\alpha} g_N(\gamma) &= N^{\alpha-1} \sum_{i=1}^N \mathbf{z}'_i [\Delta \dot{\mathbf{y}}_i - \Delta \dot{\mathbf{y}}_{i,-1}(\gamma) \hat{\boldsymbol{\varphi}}(\gamma)] \\ &= N^{\alpha-1/2} \frac{1}{\sqrt{N}} \sum_{i=1}^N \mathbf{z}'_i \Delta \dot{\mathbf{u}}_i - \mathbf{G}_{N,\varphi}(\gamma) N^{\alpha} [\hat{\boldsymbol{\varphi}}(\gamma) - \boldsymbol{\varphi}_0] - \nabla \mathbf{G}_{N,\varphi,2}(\gamma) \mathbf{c} \\ &\xrightarrow{p} \mathbf{G}_{\varphi}(\gamma) [\mathbf{G}_{\varphi}(\gamma)' \mathbf{S}^{-1} \mathbf{G}_{\varphi}(\gamma)]^{-1} \mathbf{G}_{\varphi}(\gamma)' \mathbf{S}^{-1} \nabla \mathbf{G}_{\varphi,2}(\gamma) \mathbf{c} - \nabla \mathbf{G}_{\varphi,2}(\gamma) \mathbf{c} \\ &\equiv g^{\dagger}(\gamma). \end{aligned} \quad (\text{A115})$$

Then, we have $N^{2\alpha} Q_N(\gamma) = N^{2\alpha} g_N(\gamma)' \mathbf{S}_N^{-1} g_N(\gamma) \rightarrow_p g^{\dagger}(\gamma)' \mathbf{S}^{-1} g^{\dagger}(\gamma)$. Note that it can be written as

$$g^{\dagger}(\gamma)' \mathbf{S}^{-1} g^{\dagger}(\gamma) = \mathbf{c}' \Delta \tilde{\mathbf{G}}_{\varphi,2}(\gamma)' \mathbf{M}_{\tilde{\mathbf{G}}_{\varphi}(\gamma)} \Delta \tilde{\mathbf{G}}_{\varphi,2}(\gamma) \mathbf{c}, \quad (\text{A116})$$

where $\mathbf{M}_{\tilde{\mathbf{G}}_{\varphi}(\gamma)} = \mathbf{I} - \tilde{\mathbf{G}}_{\varphi}(\gamma) [\tilde{\mathbf{G}}_{\varphi}(\gamma)' \tilde{\mathbf{G}}_{\varphi}(\gamma)]^{-1} \tilde{\mathbf{G}}_{\varphi}(\gamma)'$ is the orthogonal projection matrix and $\tilde{\mathbf{G}}_{\varphi,2}(\gamma)$ lies in the column space of $\tilde{\mathbf{G}}_{\varphi}(\gamma)$, then we have $\mathbf{M}_{\tilde{\mathbf{G}}_{\varphi}(\gamma)} \tilde{\mathbf{G}}_{\varphi,2}(\gamma) = \mathbf{0}$. Therefore, we have following equation

$$\mathbf{c}' \nabla \tilde{\mathbf{G}}_{\varphi,2}(\gamma)' \mathbf{M}_{\tilde{\mathbf{G}}_{\varphi}(\gamma)} \nabla \tilde{\mathbf{G}}_{\varphi,2}(\gamma) \mathbf{c} = \mathbf{c}' \tilde{\mathbf{G}}_{\varphi,2}(\gamma_0)' \mathbf{M}_{\tilde{\mathbf{G}}_{\varphi}(\gamma)} \tilde{\mathbf{G}}_{\varphi,2}(\gamma_0) \mathbf{c}. \quad (\text{A117})$$

Since $\mathbf{s}_{it}$ is not multicollinear, it implies that $\tilde{\mathbf{G}}_{\varphi,2}(\boldsymbol{\gamma}) = \tilde{\mathbf{G}}_{\varphi,2}(\boldsymbol{\gamma}_0)$ if and only if $\boldsymbol{\gamma} = \boldsymbol{\gamma}_0$. Consequently, $\tilde{\mathbf{G}}_{\varphi,2}(\boldsymbol{\gamma})$ and $\tilde{\mathbf{G}}_{\varphi}(\boldsymbol{\gamma})$ are linearly independent, which implies that $[\tilde{\mathbf{G}}_{\varphi,2}(\boldsymbol{\gamma}), \tilde{\mathbf{G}}_{\varphi}(\boldsymbol{\gamma})]$ has full column rank. By the Schur complement, for any $\boldsymbol{\gamma} \neq \boldsymbol{\gamma}_0$, $\tilde{\mathbf{G}}_{\varphi,2}(\boldsymbol{\gamma}_0)' \mathbf{M}_{\tilde{\mathbf{G}}_{\varphi}(\boldsymbol{\gamma})} \tilde{\mathbf{G}}_{\varphi,2}(\boldsymbol{\gamma}_0)$ is a positive definite matrix.

Therefore, by Assumption 2 $\mathbf{c} \neq \mathbf{0}$, we have $\mathbf{g}^\dagger(\boldsymbol{\gamma})' \mathbf{S}^{-1} \mathbf{g}^\dagger(\boldsymbol{\gamma}) = \mathbf{0}$ if and only if $\boldsymbol{\gamma} = \boldsymbol{\gamma}_0$. Since $\hat{\boldsymbol{\gamma}}$ minimizes $N^{2\alpha} Q_N(\boldsymbol{\gamma})$, by Theorem 2.1 of Newey and McFadden (1994), we have $\hat{\boldsymbol{\gamma}} \rightarrow_p \boldsymbol{\gamma}_0$. $\square$

Lemma 12. Under Assumption 2, set $\boldsymbol{\gamma} = \boldsymbol{\gamma}_0 + \mathbf{v}h$, and $h \rightarrow 0$ as $N \rightarrow \infty$, $\mathbf{G}_{N,\varphi,t,j}(\boldsymbol{\gamma}) \rightarrow_p \mathbf{G}_{\varphi,t,j}(\boldsymbol{\gamma}_0)$ for $j = t, t-1$.

Proof. By Lemma 6,

$$\begin{aligned} E(\mathbf{G}_{N,\varphi,t,j}(\boldsymbol{\gamma}) | \mathbf{s}_{ij}) &= E(\mathbf{z}_{it} w_{i,j-1} | A_{ij}(\boldsymbol{\gamma}), \mathbf{s}_{ij}) \frac{P(A_{ij}(\boldsymbol{\gamma}) | \mathbf{s}_{ij})}{\mathbf{s}_{ij}' \boldsymbol{\gamma} - \mathbf{s}_{ij}' \boldsymbol{\gamma}_0} \mathbf{s}_{ij}' \\ &\rightarrow E(\mathbf{z}_{it} w_{i,j-1} | A_{ij}^0, \mathbf{s}_{ij}) f_j(\mathbf{s}_{ij}' \boldsymbol{\gamma}_0 | \mathbf{s}_{ij}) \mathbf{s}_{ij}'. \end{aligned} \quad (\text{A118})$$

Using the law of iterated expectations, we have

$$E(\mathbf{G}_{N,\varphi,t,j}(\boldsymbol{\gamma})) = E(E(\mathbf{G}_{N,\varphi,t,j}(\boldsymbol{\gamma}) | \mathbf{s}_{ij})) \rightarrow \mathbf{G}_{\varphi,t,j}(\boldsymbol{\gamma}_0). \quad (\text{A119})$$

Using the arguments similar to those in the proof of Lemma A.10 in Hansen (2000), we can show that $\mathbf{G}_{N,\varphi,t,j}(\boldsymbol{\gamma}) - E(\mathbf{G}_{N,\varphi,t,j}(\boldsymbol{\gamma})) \rightarrow \mathbf{0}$. $\square$

Lemma 13. Under Assumption 2, set $\boldsymbol{\gamma} = \boldsymbol{\gamma}_0 + \mathbf{v}/\sqrt{\lambda_N}$, $N \rightarrow \infty$

$$\sqrt{N}[\hat{\boldsymbol{\varphi}}(\boldsymbol{\gamma}) - \boldsymbol{\varphi}_0] \xrightarrow{d} [\mathbf{G}_{\varphi}(\boldsymbol{\gamma}_0)' \mathbf{S}^{-1} \mathbf{G}_{\varphi}(\boldsymbol{\gamma}_0)]^{-1} \mathbf{G}_{\varphi}(\boldsymbol{\gamma}_0)' \mathbf{S}^{-1} [\mathbf{Z}_S - \mathbf{G}_{\boldsymbol{\gamma}}(\boldsymbol{\gamma}_0) \mathbf{v}], \quad (\text{A120})$$

where $\mathbf{Z}_S \sim N(\mathbf{0}, \mathbf{S})$.

Proof. Note that, for $t = 2, 3, \dots, T$

$$\begin{aligned} &\frac{\sqrt{\lambda_N}}{N} \sum_{i=1}^N \mathbf{z}_{it} \nabla \Delta y_{i,t-1}^*(\boldsymbol{\gamma}) \mathbf{c} \\ &= \frac{\sqrt{\lambda_N}}{N} \sum_{i=1}^N \mathbf{z}_{it} [\Delta y_{i,t-1}^*(\boldsymbol{\gamma}) - \Delta y_{i,t-1}^*(\boldsymbol{\gamma}_0)] \mathbf{c} \\ &= \frac{\sqrt{\lambda_N}}{N} \sum_{i=1}^N \mathbf{z}_{it} [y_{i,t-1}^* \nabla I_{it}(\boldsymbol{\gamma}) - y_{i,t-2}^* \nabla I_{i,t-1}(\boldsymbol{\gamma}_0)] \mathbf{c} \\ &= \frac{\sqrt{\lambda_N}}{N} \sum_{i=1}^N \left[\mathbf{z}_{it} y_{i,t-1}^* \mathbf{c} \frac{\nabla I_{it}(\boldsymbol{\gamma})}{\mathbf{s}_{it}'(\boldsymbol{\gamma} - \boldsymbol{\gamma}_0)} \frac{\mathbf{s}_{it}' \mathbf{v}}{\sqrt{\lambda_N}} - \mathbf{z}_{it} y_{i,t-2}^* \mathbf{c} \frac{\nabla I_{i,t-1}(\boldsymbol{\gamma})}{\mathbf{s}_{i,t-1}'(\boldsymbol{\gamma} - \boldsymbol{\gamma}_0)} \frac{\mathbf{s}_{i,t-1}' \mathbf{v}}{\sqrt{\lambda_N}} \right] \\ &= \frac{\sqrt{\lambda_N}}{N} \sum_{i=1}^N \left[\mathbf{z}_{it} y_{i,t-1}^* \mathbf{c} \frac{\nabla I_{it}(\boldsymbol{\gamma})}{\mathbf{s}_{it}'(\boldsymbol{\gamma} - \boldsymbol{\gamma}_0)} \mathbf{s}_{it}' - \mathbf{z}_{it} y_{i,t-2}^* \mathbf{c} \frac{\nabla I_{i,t-1}(\boldsymbol{\gamma})}{\mathbf{s}_{i,t-1}'(\boldsymbol{\gamma} - \boldsymbol{\gamma}_0)} \mathbf{s}_{i,t-1}' \right] \frac{\mathbf{v}}{\sqrt{\lambda_N}}, \\ &= [\mathbf{G}_{N,\boldsymbol{\gamma},t,t}(\boldsymbol{\gamma}) - \mathbf{G}_{N,\boldsymbol{\gamma},t,t-1}(\boldsymbol{\gamma})] \mathbf{v} \end{aligned} \quad (\text{A121})$$

Consequently, we have

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N \mathbf{z}_i' \nabla \Delta y_{-1}^*(\boldsymbol{\gamma}) N^{-\alpha} = \frac{\sqrt{\lambda_N}}{N} \sum_{i=1}^N \mathbf{z}_i' \nabla \Delta y_{-1}^*(\boldsymbol{\gamma})$$

$$= \frac{\sqrt{\lambda_N}}{N} \sum_{i=1}^N \begin{bmatrix} \mathbf{z}_{i2} \nabla \Delta y_{i,1}^*(\gamma) \mathbf{c} \\ \mathbf{z}_{i3} \nabla \Delta y_{i,2}^*(\gamma) \mathbf{c} \\ \vdots \\ \mathbf{z}_{iT} \nabla \Delta y_{i,T-1}^*(\gamma) \mathbf{c} \end{bmatrix} \quad (\text{A122})$$

$$= \begin{bmatrix} \mathbf{G}_{N,\gamma,2,2}(\gamma) - \mathbf{G}_{N,\gamma,2,1}(\gamma) \\ \mathbf{G}_{N,\gamma,3,3}(\gamma) - \mathbf{G}_{N,\gamma,3,2}(\gamma) \\ \vdots \\ \mathbf{G}_{N,\gamma,T,T}(\gamma) - \mathbf{G}_{N,\gamma,T,T-1}(\gamma) \end{bmatrix} \mathbf{v} \\ = \mathbf{G}_{N,\gamma}(\gamma) \mathbf{v}. \quad (\text{A123})$$

Therefore, by Lemma 12, we have

$$\begin{aligned} \sqrt{N}[\hat{\boldsymbol{\phi}}(\gamma) - \boldsymbol{\phi}_0] &= [\mathbf{G}_{N,\phi}(\gamma)' \mathbf{S}^{-1} \mathbf{G}_{N,\phi}(\gamma)]^{-1} \mathbf{G}_{N,\phi}(\gamma)' \mathbf{S}^{-1} \\ &\quad \times \left[\frac{1}{\sqrt{N}} \sum_{i=1}^N \mathbf{z}_i' (\Delta \dot{u}_i - \nabla \Delta y_{i-1}^*(\gamma) N^{-\alpha}) \right] \\ &= [\mathbf{G}_{N,\phi}(\gamma)' \mathbf{S}^{-1} \mathbf{G}_{N,\phi}(\gamma)]^{-1} \mathbf{G}_{N,\phi}(\gamma)' \mathbf{S}^{-1} \\ &\quad \times \left[\frac{1}{\sqrt{N}} \sum_{i=1}^N \mathbf{z}_i' \Delta \dot{u}_i - \mathbf{G}_{N,\gamma}(\gamma) \mathbf{v} \right] \\ &\xrightarrow{d} [\mathbf{G}_{\phi}(\gamma_0)' \mathbf{S}^{-1} \mathbf{G}_{\phi}(\gamma_0)]^{-1} \mathbf{G}_{\phi}(\gamma_0)' \mathbf{S}^{-1} [\mathbf{Z}_S - \mathbf{G}_{\gamma}(\gamma_0) \mathbf{v}]. \end{aligned} \quad (\text{A124})$$

□

Lemma 14. Under Assumption 2, as $N \rightarrow \infty$

$$N^{1/2-\alpha}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}_0) \xrightarrow{d} \left[\tilde{\mathbf{G}}_{\gamma}(\gamma_0)' \mathbf{M}_{\tilde{\mathbf{G}}_{\phi}(\gamma_0)} \tilde{\mathbf{G}}_{\gamma}(\gamma_0) \right]^{-1} \tilde{\mathbf{G}}_{\gamma}(\gamma_0)' \mathbf{M}_{\tilde{\mathbf{G}}_{\phi}(\gamma_0)} \tilde{\mathbf{Z}}_S, \quad (\text{A125})$$

where $\mathbf{Z}_S \sim N(\mathbf{0}, \mathbf{S})$ and $\tilde{\mathbf{Z}}_S = \mathbf{S}^{-1/2} \mathbf{Z}_S$.

Proof. Using the similar argument as in the proof of Lemma A.9 in Hansen (2000) and Theorem 2 in Yang et al. (2021b), we can establish the convergence rate of the threshold estimator $\sqrt{\lambda_N}$. Thus, we can study their asymptotic behavior in the neighborhood of the true thresholds, $\hat{\boldsymbol{\gamma}} = \boldsymbol{\gamma}_0 + \hat{\mathbf{v}}/\sqrt{\lambda_N}$.

Let $\boldsymbol{\gamma} = \boldsymbol{\gamma}_0 + \mathbf{v}/\sqrt{\lambda_N}$. Next, we derive the distribution of $NQ_N(\boldsymbol{\gamma})$. Note that

$$\begin{aligned} N^{1/2} \mathbf{S}_N^{-1/2} g_N(\boldsymbol{\gamma}) &= \mathbf{S}_N^{-1/2} \frac{1}{\sqrt{N}} \sum_{i=1}^N \mathbf{z}_i' [\Delta \dot{y}_i - \Delta \dot{y}_{i-1}(\boldsymbol{\gamma}) \hat{\boldsymbol{\phi}}(\boldsymbol{\gamma})] \\ &= \tilde{\mathbf{Z}}_{N,S} - \tilde{\mathbf{G}}_{N,\phi}(\boldsymbol{\gamma}) N^{1/2} [\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}) - \boldsymbol{\phi}_0] - \tilde{\mathbf{G}}_{N,\gamma}(\boldsymbol{\gamma}) \mathbf{v} \\ &= \tilde{\mathbf{G}}_{N,\phi}(\boldsymbol{\gamma}) \left[\tilde{\mathbf{G}}_{N,\phi}(\boldsymbol{\gamma})' \tilde{\mathbf{G}}_{N,\phi}(\boldsymbol{\gamma}) \right]^{-1} \tilde{\mathbf{G}}_{N,\phi}(\boldsymbol{\gamma}) [\tilde{\mathbf{Z}}_{N,S} - \tilde{\mathbf{G}}_{N,\gamma}(\boldsymbol{\gamma}) \mathbf{v}] \\ &\xrightarrow{d} \tilde{\mathbf{G}}_{\phi}(\gamma_0) \left[\tilde{\mathbf{G}}_{\phi}(\gamma_0)' \tilde{\mathbf{G}}_{\phi}(\gamma_0) \right]^{-1} \tilde{\mathbf{G}}_{\phi}(\gamma_0)' [\tilde{\mathbf{Z}}_S - \tilde{\mathbf{G}}_{\gamma}(\gamma_0) \mathbf{v}]. \end{aligned} \quad (\text{A126})$$

Therefore, we have

$$\begin{aligned} NQ_N(\gamma) &= N^{1/2}g_N(\gamma)'S_N^{-1}N^{1/2}g_N(\gamma) \\ &\xrightarrow{d} [\tilde{Z}_S - \tilde{G}_\gamma(\gamma_0)\nu]' \tilde{G}_\varphi(\gamma_0) [\tilde{G}_\varphi(\gamma_0)' \tilde{G}_\varphi(\gamma_0)]^{-1} \tilde{G}_\varphi(\gamma_0)' \\ &\quad \times [\tilde{Z}_S - \tilde{G}_\gamma(\gamma_0)\nu] \\ &\equiv Q(\nu), \end{aligned} \quad (A127)$$

Note that the minimization problem $\min_{\nu} Q(\nu)$ has the following explicit solution

$$\begin{aligned} \arg \min_{\nu} Q(\nu) &= [\tilde{G}_\gamma(\gamma_0)' M_{\tilde{G}_\varphi(\gamma_0)} \tilde{G}_\gamma(\gamma_0)]^{-1} \tilde{G}_\gamma(\gamma_0)' M_{\tilde{G}_\varphi(\gamma_0)} \tilde{Z}_S \\ &\equiv \nu_0. \end{aligned} \quad (A128)$$

Therefore, we obtain

$$\sqrt{\lambda_N}(\hat{\gamma} - \gamma_0) = \hat{\nu} \xrightarrow{d} \arg \min_{\nu} Q(\nu) = \nu_0. \quad (A129)$$

□

Proof of Theorem 3. By Lemma 13 and Lemma 14. Let $B_\gamma = \tilde{G}_\gamma(\gamma_0)' M_{\tilde{G}_\varphi(\gamma_0)} \tilde{G}_\gamma(\gamma_0)$ and $B_\varphi = \tilde{G}_\varphi(\gamma_0)' M_{\tilde{G}_\varphi(\gamma_0)} \tilde{G}_\varphi(\gamma_0)$. We have

$$\begin{aligned} \sqrt{N}[\hat{\varphi} - \varphi_0] &\xrightarrow{d} [\tilde{G}_\varphi(\gamma_0)' \tilde{G}_\varphi(\gamma_0)]^{-1} \tilde{G}_\varphi(\gamma_0)' \\ &\quad \times [\tilde{Z}_S - \tilde{G}_\gamma(\gamma_0) B_\gamma^{-1} \tilde{G}_\gamma(\gamma_0)' M_{\tilde{G}_\varphi(\gamma_0)} \tilde{Z}_S] \\ &= [\tilde{G}_\varphi(\gamma_0)' \tilde{G}_\varphi(\gamma_0)]^{-1} \tilde{G}_\varphi(\gamma_0)' \tilde{Z}_S \\ &\quad - [\tilde{G}_\varphi(\gamma_0)' \tilde{G}_\varphi(\gamma_0)]^{-1} \tilde{G}_\varphi(\gamma_0)' \tilde{G}_\gamma(\gamma_0) B_\gamma^{-1} \tilde{G}_\gamma(\gamma_0)' M_{\tilde{G}_\varphi(\gamma_0)} \tilde{Z}_S \\ &= B_\varphi^{-1} B_\varphi [\tilde{G}_\varphi(\gamma_0)' \tilde{G}_\varphi(\gamma_0)]^{-1} \tilde{G}_\varphi(\gamma_0)' \tilde{Z}_S \\ &\quad - B_\varphi^{-1} \tilde{G}_\varphi(\gamma_0)' \tilde{G}_\gamma(\gamma_0) [\tilde{G}_\gamma(\gamma_0)' \tilde{G}_\gamma(\gamma_0)]^{-1} \tilde{G}_\gamma(\gamma_0)' M_{\tilde{G}_\varphi(\gamma_0)} \tilde{Z}_S \\ &= B_\varphi^{-1} \tilde{G}_\varphi(\gamma_0)' [I - P_{\tilde{G}_\gamma(\gamma_0)} P_{\tilde{G}_\varphi(\gamma_0)}] \tilde{Z}_S \\ &\quad - B_\varphi^{-1} \tilde{G}_\varphi(\gamma_0)' P_{\tilde{G}_\gamma(\gamma_0)} M_{\tilde{G}_\varphi(\gamma_0)} \tilde{Z}_S \\ &= B_\varphi^{-1} \tilde{G}_\varphi(\gamma_0)' [I - P_{\tilde{G}_\gamma(\gamma_0)}] \tilde{Z}_S \\ &= [\tilde{G}_\varphi(\gamma_0)' M_{\tilde{G}_\varphi(\gamma_0)} \tilde{G}_\varphi(\gamma_0)]^{-1} \tilde{G}_\varphi(\gamma_0)' M_{\tilde{G}_\varphi(\gamma_0)} \tilde{Z}_S. \end{aligned} \quad (A130)$$

It is easily seen that the asymptotic distributions of the slope and threshold estimators are the partitioned matrices of $[G(\gamma_0)' S^{-1} G(\gamma_0)]^{-1} G(\gamma_0)' S^{-1} Z_S$. Therefore, we have

$$\left[\begin{array}{c} N^{1/2}(\hat{\varphi} - \varphi_0) \\ N^{1/2-\alpha}(\hat{\gamma} - \gamma_0) \end{array} \right] \xrightarrow{d} [G(\gamma_0)' S^{-1} G(\gamma_0)]^{-1} G(\gamma_0)' S^{-1} Z_S. \quad (A131)$$

□

Appendix B: Additional Monte Carlo Simulations

In this appendix, we investigate the finite sample performance of the proposed MLE estimator in comparison with the GMM estimator discussed in Section 5. Our main purpose is to examine the sensitivity of the MLE to the normality assumption of the error terms and the assumption on the initial conditions. To this end, we consider the following three cases.

Case 1 is the baseline design as in Section 6, in which both the normality assumption and the assumption on the initial conditions hold. Thus, the DGP in Case 1 is the same as in Section 6.

Case 2 follows the non-normal error design of Blundell and Bond (1998) by replacing the normal error term with a centered and standardized chi-square disturbance. As in Blundell and Bond (1998), Case 2 considers a departure from the normality assumption by generating the error term from a centered and standardized chi-square distribution. Specifically, we set

$$u_{it} = \sigma_u \frac{\chi_{it}^2(1) - 1}{\sqrt{2}}, \quad (\text{B1})$$

where $\sigma_u = 2$ denotes the standard deviation of the error term in the baseline design, $\chi_{it}^2(1)$ denotes a chi-square random variable with one degree of freedom, and $(\chi_{it}^2(1) - 1)/\sqrt{2}$ is centered and scaled to have mean zero and variance one. Other settings of the DGP in Case 2 are the same as in the baseline design described in Section 6 of the main text.

Case 3 modifies the first difference equation so that the initial condition disturbance is correlated with the history of the threshold variable, thereby violating the independence restriction imposed on v_{i1} by the likelihood approach. Case 3 considers a violation of the assumption on the initial conditions. In this case, the error term u_{it} remains normally distributed as in the baseline design, but the first difference equation is modified as

$$\Delta y_{it} = \delta_0 + \delta_1 \Delta q_{it} + \delta_2 \Delta z_{it} + v_{it}^*, \quad (\text{B2})$$

where $v_{it}^* = 0.5\bar{q}_i + v_{it}$ with $\bar{q}_i = \frac{1}{T+1} \sum_{t=0}^T q_{it}$. Except for this modification to the first difference equation, the DGP used in Case 3 is similar to the baseline design described in Section 6.

Table B.1 reports the simulation results for the three cases with $T \in \{5, 10\}$ and $N \in \{50, 100, 200\}$. The number of replications is set to 1,000. The simulation results can be summarized as follows. First, under the baseline design in Case 1, both the MLE and GMM estimators perform well. The biases of the threshold parameters γ_0 and γ_1 are close to zero for both estimators. For the slope parameters, the GMM estimator has smaller biases for the dynamic coefficients β_1 and β_2 , whereas the biases of the MLE estimator are moderate and decrease as T increases. On the other hand, the MLE estimator generally has smaller standard deviations than the GMM estimator, which is consistent with the efficiency advantage of the likelihood approach when the normality assumption and the assumption on the initial conditions are satisfied, and confirms our asymptotic results in Theorems 1 and 3. Second, the results in Case 2 show that the finite sample performance of both estimators does not substantially differ from that under normal errors when the normal error term is replaced by the centered and standardized chi-square disturbance. The biases and standard deviations under the non-normal error design are very close to those under the baseline design. Moreover, the R statistics for the threshold parameters are close to 1, indicating convergence of the MCMC samples. Third, Case 3 examines the effect of violating the assumption on initial conditions. The comparison of Cases 1 and 3 suggests that allowing the initial condition disturbance to be correlated with the history of the threshold variable has little impact on the biases and standard deviations of these estimators.

In Table B.2, we further examine the sensitivity of the MLE estimator to the normality assumption by considering three additional non-normal disturbance designs. In each design, we allow both the error term u_{it} and the initial condition disturbance v_{i1} to depart from normality. First, following Blundell and Bond (1998), the disturbances are drawn from a skewed chi-square distribution. Second, following Wang and Zhu (2018), the disturbances are drawn from a Student- t distribution with five degrees of freedom, which implies fatter tails than the normal distribution. Finally, in the spirit of the mixed-normal simulation design in Moral-Benito et al. (2019), disturbances are generated from a mixture of two normal distributions with different means. Specifically, we

Table B.1: Comparison of MLE and GMM estimators under different DGP designs.

T	N	Method		$\gamma_0 = 0.2$	$\gamma_1 = 0.5$	$\beta_1 = -0.3$	$\beta_2 = 0.7$	$\theta_1 = 1$	$\theta_2 = 2$	$\psi = 1$	The R statistic		
											γ_0	γ_1	
Case1: baseline with normal errors and valid initial conditions													
5	50	MLE	Bias	−0.001	0.000	0.034	0.039	0.011	0.002	0.002	1.041	1.042	
			Std.dev	0.032	0.041	0.036	0.038	0.278	0.231	0.154	0.066	0.064	
		GMM	Bias	0.001	−0.001	−0.006	−0.002	0.001	−0.003	−0.006	1.028	1.026	
			Std.dev	0.044	0.058	0.046	0.051	0.405	0.325	0.157	0.029	0.028	
		100	MLE	Bias	0.000	0.002	0.034	0.040	0.010	−0.001	0.008	1.052	1.050
				Std.dev	0.015	0.020	0.024	0.026	0.191	0.168	0.106	0.073	0.067
	200	GMM	Bias	0.000	0.002	−0.004	−0.002	0.001	0.002	−0.001	1.026	1.023	
			Std.dev	0.034	0.039	0.040	0.044	0.395	0.312	0.108	0.024	0.021	
		MLE	Bias	0.000	0.000	0.034	0.039	0.007	0.002	0.007	1.064	1.066	
			Std.dev	0.007	0.010	0.017	0.018	0.143	0.118	0.074	0.099	0.112	
		GMM	Bias	0.001	0.001	0.000	−0.003	−0.018	0.016	−0.002	1.015	1.020	
			Std.dev	0.026	0.034	0.035	0.039	0.354	0.278	0.076	0.011	0.017	
10	50	MLE	Bias	0.000	0.000	0.015	0.016	−0.004	0.004	0.002	1.049	1.050	
			Std.dev	0.013	0.018	0.021	0.021	0.184	0.155	0.098	0.070	0.069	
		GMM	Bias	−0.001	0.000	−0.002	−0.003	−0.002	0.002	0.000	1.031	1.029	
			Std.dev	0.014	0.020	0.022	0.023	0.197	0.164	0.098	0.036	0.034	
		100	MLE	Bias	0.000	0.000	0.015	0.016	−0.001	0.002	0.002	1.068	1.062
				Std.dev	0.007	0.009	0.014	0.015	0.125	0.109	0.069	0.113	0.094
	200	GMM	Bias	0.000	0.001	−0.002	−0.002	−0.004	0.004	0.000	1.026	1.025	
			Std.dev	0.010	0.013	0.019	0.020	0.182	0.150	0.069	0.027	0.027	
		MLE	Bias	0.000	0.000	0.015	0.017	−0.001	0.001	0.000	1.083	1.083	
			Std.dev	0.003	0.004	0.010	0.011	0.094	0.074	0.048	0.109	0.113	
		GMM	Bias	0.000	0.001	−0.001	0.000	−0.003	0.000	−0.001	1.023	1.023	
			Std.dev	0.007	0.010	0.016	0.016	0.175	0.136	0.048	0.023	0.024	
Case2: non-normal errors													
5	50	MLE	Bias	−0.003	−0.001	0.035	0.041	0.002	0.002	0.005	1.045	1.046	
			Std.dev	0.032	0.040	0.039	0.042	0.265	0.232	0.142	0.079	0.089	
		GMM	Bias	−0.003	0.000	−0.004	−0.002	−0.003	−0.004	−0.003	1.032	1.027	
			Std.dev	0.049	0.059	0.047	0.053	0.383	0.317	0.144	0.037	0.032	
		100	MLE	Bias	−0.001	0.000	0.033	0.039	−0.009	0.010	0.003	1.054	1.051
				Std.dev	0.013	0.019	0.025	0.029	0.193	0.164	0.104	0.095	0.079
	200	GMM	Bias	−0.001	0.001	−0.004	−0.001	−0.042	0.027	−0.006	1.027	1.023	
			Std.dev	0.033	0.047	0.040	0.046	0.372	0.302	0.106	0.025	0.022	
		MLE	Bias	−0.001	0.000	0.033	0.039	−0.001	0.004	0.009	1.065	1.064	
			Std.dev	0.007	0.010	0.019	0.019	0.139	0.114	0.069	0.098	0.113	
		GMM	Bias	−0.001	0.003	−0.003	−0.002	−0.002	0.002	0.001	1.015	1.021	
			Std.dev	0.028	0.033	0.037	0.040	0.377	0.288	0.072	0.011	0.017	
10	50	MLE	Bias	0.000	−0.001	0.015	0.016	0.000	0.009	0.003	1.053	1.050	
			Std.dev	0.013	0.018	0.021	0.022	0.175	0.153	0.097	0.096	0.079	
		GMM	Bias	0.000	−0.001	−0.002	−0.003	−0.003	0.008	0.001	1.029	1.028	
			Std.dev	0.015	0.020	0.022	0.024	0.198	0.165	0.097	0.030	0.026	
		100	MLE	Bias	0.000	0.000	0.015	0.016	0.003	−0.005	−0.002	1.064	1.066
				Std.dev	0.007	0.008	0.015	0.017	0.128	0.109	0.067	0.108	0.119
	200	GMM	Bias	0.000	0.000	−0.001	−0.002	0.001	−0.004	−0.004	1.026	1.024	
			Std.dev	0.010	0.013	0.018	0.021	0.190	0.154	0.067	0.025	0.025	
		MLE	Bias	0.000	0.000	0.015	0.017	−0.004	0.000	0.002	1.084	1.086	
			Std.dev	0.003	0.004	0.010	0.011	0.090	0.076	0.044	0.140	0.130	
		GMM	Bias	0.000	0.000	−0.001	0.000	0.001	−0.005	0.000	1.024	1.022	
			Std.dev	0.007	0.009	0.016	0.016	0.175	0.141	0.044	0.023	0.021	

Table B.1: (continued).

T	N	Method		$\gamma_0 = 0.2$	$\gamma_1 = 0.5$	$\beta_1 = -0.3$	$\beta_2 = 0.7$	$\theta_1 = 1$	$\theta_2 = 2$	$\psi = 1$	The R statistic	
												γ_0
Case3: violation of the assumption on initial conditions												
5	50	MLE	Bias	−0.001	0.000	0.033	0.039	0.010	0.003	0.002	1.053	1.049
			Std.dev	0.031	0.041	0.036	0.038	0.278	0.231	0.154	0.396	0.281
		GMM	Bias	0.001	−0.002	−0.006	−0.002	−0.003	−0.001	−0.005	1.032	1.028
			Std.dev	0.044	0.057	0.046	0.051	0.404	0.326	0.157	0.040	0.032
	100	MLE	Bias	0.000	0.001	0.034	0.039	0.010	−0.001	0.008	1.047	1.047
			Std.dev	0.015	0.019	0.024	0.025	0.191	0.169	0.106	0.060	0.063
		GMM	Bias	0.000	0.001	−0.004	−0.003	0.002	0.000	−0.001	1.027	1.023
			Std.dev	0.033	0.039	0.040	0.043	0.397	0.314	0.108	0.026	0.023
	200	MLE	Bias	0.000	0.000	0.034	0.038	0.007	0.002	0.006	1.062	1.063
			Std.dev	0.007	0.009	0.017	0.018	0.143	0.119	0.074	0.093	0.094
		GMM	Bias	0.001	0.000	−0.001	−0.002	−0.020	0.017	−0.002	1.015	1.020
			Std.dev	0.025	0.035	0.034	0.038	0.347	0.278	0.076	0.011	0.017
10	50	MLE	Bias	−0.001	0.000	0.015	0.016	−0.004	0.004	0.002	1.049	1.050
			Std.dev	0.013	0.018	0.021	0.021	0.185	0.155	0.098	0.069	0.066
		GMM	Bias	−0.001	0.000	−0.002	−0.003	−0.002	0.002	0.000	1.032	1.032
			Std.dev	0.014	0.020	0.022	0.023	0.197	0.164	0.098	0.037	0.047
	100	MLE	Bias	0.000	0.000	0.015	0.016	−0.001	0.002	0.002	1.061	1.057
			Std.dev	0.006	0.009	0.014	0.015	0.125	0.109	0.069	0.083	0.082
		GMM	Bias	0.000	0.000	−0.002	−0.002	−0.004	0.004	0.000	1.027	1.026
			Std.dev	0.010	0.013	0.018	0.020	0.181	0.151	0.069	0.029	0.026
	200	MLE	Bias	0.000	0.000	0.015	0.017	−0.001	0.001	0.000	1.080	1.079
			Std.dev	0.003	0.005	0.010	0.011	0.094	0.074	0.048	0.101	0.104
		GMM	Bias	0.000	0.000	−0.001	0.000	−0.003	0.000	−0.001	1.023	1.023
			Std.dev	0.007	0.009	0.016	0.016	0.175	0.137	0.048	0.022	0.024

Notes: this table presents the summary statistics (i.e., mean and standard deviation) for the estimates and the R statistics for threshold parameters. The number of replications is 1,000. “MLE” and “GMM” denote the maximum likelihood estimator and the generalized method of moments estimator, respectively.

Table B.2: Comparison of MLE and GMM estimators under non-normal disturbances.

T	N	Method		γ_0	γ_1	β_1	β_2	θ_1	θ_2	ψ	The R statistic	
											γ_0	γ_1
Case1: centered and scaled chi-square disturbances for both u_{it} and v_{it}												
5	50	MLE	Bias	−0.003	0.000	0.035	0.040	0.009	−0.001	0.006	1.047	1.043
			Std.dev	0.034	0.042	0.039	0.044	0.271	0.236	0.146	0.087	0.061
		GMM	Bias	−0.002	0.000	−0.004	−0.002	−0.004	−0.001	−0.003	1.029	1.026
			Std.dev	0.055	0.063	0.046	0.054	0.382	0.320	0.146	0.033	0.030
	100	MLE	Bias	−0.001	0.000	0.035	0.040	−0.010	0.012	0.001	1.058	1.059
			Std.dev	0.017	0.023	0.027	0.031	0.194	0.164	0.104	0.095	0.120
		GMM	Bias	−0.001	0.001	−0.004	−0.001	−0.036	0.023	−0.005	1.026	1.022
			Std.dev	0.038	0.051	0.041	0.047	0.374	0.298	0.106	0.023	0.021
	200	MLE	Bias	0.000	0.000	0.034	0.038	−0.001	0.005	0.009	1.068	1.072
			Std.dev	0.007	0.010	0.019	0.020	0.140	0.113	0.070	0.100	0.152
		GMM	Bias	0.000	0.003	−0.004	−0.001	−0.012	0.003	0.001	1.018	1.022
			Std.dev	0.030	0.041	0.035	0.037	0.366	0.279	0.072	0.016	0.020

Table B.2: (continued).

T	N	Method		γ_0	γ_1	β_1	β_2	θ_1	θ_2	ψ	The R statistic	
											γ_0	γ_1
10	50	MLE	Bias	0.000	−0.001	0.015	0.016	−0.001	0.007	0.004	1.062	1.058
			Std.dev	0.013	0.019	0.022	0.023	0.173	0.152	0.097	0.172	0.142
		GMM	Bias	0.000	−0.001	−0.003	−0.004	−0.003	0.008	0.001	1.029	1.029
			Std.dev	0.016	0.022	0.023	0.025	0.194	0.163	0.097	0.035	0.032
	100	MLE	Bias	0.000	0.000	0.015	0.016	0.002	−0.005	−0.002	1.065	1.065
			Std.dev	0.007	0.010	0.015	0.017	0.128	0.110	0.067	0.103	0.092
		GMM	Bias	0.000	0.000	−0.001	−0.002	0.005	−0.006	−0.003	1.026	1.026
			Std.dev	0.011	0.015	0.018	0.020	0.189	0.156	0.067	0.025	0.024
	200	MLE	Bias	0.000	0.000	0.015	0.017	−0.003	0.000	0.003	1.088	1.092
			Std.dev	0.003	0.005	0.011	0.012	0.091	0.077	0.045	0.119	0.126
		GMM	Bias	0.000	0.000	−0.001	−0.001	0.002	−0.005	0.001	1.025	1.025
			Std.dev	0.008	0.010	0.015	0.016	0.175	0.142	0.045	0.028	0.030
Case2: centered and scaled Student- t_5 disturbances for both u_{it} and v_{it}												
5	50	MLE	Bias	−0.001	0.001	0.035	0.042	0.012	0.012	0.008	1.043	1.042
			Std.dev	0.030	0.043	0.036	0.039	0.281	0.229	0.141	0.071	0.071
		GMM	Bias	−0.001	0.005	−0.004	−0.002	0.000	0.015	−0.002	1.030	1.028
			Std.dev	0.046	0.061	0.048	0.051	0.404	0.316	0.146	0.032	0.031
	100	MLE	Bias	0.000	0.001	0.034	0.039	0.005	0.004	0.009	1.052	1.052
			Std.dev	0.015	0.020	0.025	0.027	0.199	0.163	0.101	0.081	0.087
		GMM	Bias	0.001	0.002	−0.004	−0.002	−0.014	0.013	0.000	1.026	1.022
			Std.dev	0.032	0.046	0.041	0.044	0.385	0.296	0.102	0.024	0.022
	200	MLE	Bias	0.000	0.000	0.036	0.040	0.011	−0.002	0.009	1.067	1.064
			Std.dev	0.008	0.010	0.017	0.019	0.135	0.110	0.073	0.130	0.112
		GMM	Bias	−0.001	0.000	−0.001	0.000	0.005	−0.006	0.000	1.015	1.020
			Std.dev	0.028	0.036	0.037	0.038	0.357	0.277	0.077	0.012	0.017
10	50	MLE	Bias	0.000	−0.001	0.014	0.015	0.002	−0.003	0.002	1.051	1.049
			Std.dev	0.014	0.019	0.021	0.022	0.178	0.150	0.100	0.068	0.068
		GMM	Bias	0.000	−0.001	−0.004	−0.005	−0.005	0.001	0.000	1.031	1.030
			Std.dev	0.015	0.021	0.021	0.024	0.196	0.165	0.099	0.035	0.032
	100	MLE	Bias	0.000	0.000	0.015	0.016	0.000	0.006	0.003	1.066	1.064
			Std.dev	0.007	0.010	0.015	0.016	0.128	0.111	0.067	0.108	0.101
		GMM	Bias	0.000	0.000	−0.001	−0.002	0.006	0.001	0.001	1.025	1.024
			Std.dev	0.011	0.014	0.019	0.020	0.189	0.152	0.067	0.027	0.026
	200	MLE	Bias	0.000	0.000	0.015	0.017	0.000	−0.001	0.002	1.088	1.085
			Std.dev	0.003	0.004	0.011	0.011	0.086	0.075	0.048	0.116	0.118
		GMM	Bias	0.001	0.000	0.000	−0.001	−0.005	0.003	0.000	1.023	1.024
			Std.dev	0.007	0.009	0.016	0.016	0.172	0.139	0.048	0.022	0.024
Case3: centered and scaled mixture-normal disturbances for both u_{it} and v_{it}												
5	50	MLE	Bias	−0.002	0.002	0.034	0.038	−0.006	0.016	0.005	1.054	1.054
			Std.dev	0.031	0.044	0.037	0.041	0.286	0.245	0.148	0.214	0.206
		GMM	Bias	−0.002	0.006	−0.005	−0.006	−0.010	0.016	−0.003	1.028	1.025
			Std.dev	0.054	0.074	0.047	0.053	0.398	0.328	0.150	0.031	0.028
	100	MLE	Bias	0.000	−0.001	0.035	0.038	−0.003	0.013	0.002	1.058	1.060
			Std.dev	0.016	0.021	0.026	0.028	0.197	0.163	0.100	0.104	0.117
		GMM	Bias	0.000	0.001	−0.003	−0.002	−0.025	0.017	−0.006	1.027	1.022
			Std.dev	0.035	0.051	0.041	0.045	0.385	0.315	0.103	0.028	0.025
	200	MLE	Bias	−0.001	0.000	0.035	0.041	−0.001	0.002	0.007	1.073	1.071
			Std.dev	0.008	0.010	0.018	0.020	0.131	0.114	0.070	0.122	0.110
		GMM	Bias	0.000	0.001	−0.002	0.000	0.003	−0.009	0.000	1.016	1.021
			Std.dev	0.031	0.037	0.035	0.038	0.365	0.289	0.073	0.014	0.020

Table B.2: (continued).

T	N	Method		γ_0	γ_1	β_1	β_2	θ_1	θ_2	ψ	The R statistic	
											γ_0	γ_1
10	50	MLE	Bias	−0.001	−0.001	0.014	0.016	−0.013	0.006	0.007	1.063	1.060
			Std.dev	0.014	0.020	0.022	0.023	0.181	0.154	0.095	0.151	0.101
		GMM	Bias	−0.001	−0.001	−0.005	−0.003	−0.013	0.005	0.005	1.028	1.029
			Std.dev	0.016	0.022	0.022	0.024	0.197	0.165	0.095	0.029	0.032
	100	MLE	Bias	0.000	0.000	0.015	0.016	−0.004	0.003	−0.002	1.073	1.074
			Std.dev	0.007	0.009	0.015	0.016	0.127	0.109	0.063	0.102	0.115
		GMM	Bias	0.000	−0.001	−0.001	−0.003	−0.015	0.012	−0.004	1.025	1.025
			Std.dev	0.011	0.015	0.018	0.019	0.183	0.147	0.063	0.024	0.024
	200	MLE	Bias	0.000	0.000	0.016	0.017	−0.002	0.007	−0.001	1.090	1.085
			Std.dev	0.003	0.005	0.010	0.011	0.089	0.077	0.048	0.133	0.118
		GMM	Bias	0.000	0.001	−0.001	−0.001	−0.010	0.012	−0.002	1.024	1.025
			Std.dev	0.008	0.010	0.014	0.016	0.174	0.135	0.049	0.023	0.024

Notes: this table presents the summary statistics (i.e., mean and standard deviation) for the estimates and the R statistics for threshold parameters. The number of replications is 1,000. “MLE” and “GMM” denote the maximum likelihood estimator and the generalized method of moments estimator, respectively.

use $0.9N(0, 1) + 0.1N(20, 1)$, so that the resulting distribution is asymmetric. In all three cases, the disturbances are centered and scaled to have the same mean and variance as in the baseline design. Table B.2 reports the simulation results for these three non-normal disturbance designs. Overall, the results are very similar to those of the normal case reported in Table B.1. These results indicate that the finite sample performance of the proposed MLE estimator is robust to a more general error structure, including skewed chi-square, Student- t , and mixed-normal disturbances for both u_{it} and v_{it} .

In the cases considered above, the simulations show that, when the sample size is doubled, the standard deviations of the threshold parameters would decrease by almost half for the MLE estimator, while these standard deviations would decrease by roughly 30 % for the GMM estimator. These results confirm the super-consistent property in Theorem 1 for the MLE estimator and the normality in Theorem 3 for the GMM estimator. In sum, these simulations indicate that the proposed MLE performs well in finite samples when the normality and initial condition assumptions are violated, while the GMM estimator provides a useful alternative under less restrictive assumptions on error distributions and initial conditions.

References

- Anderson, T. W., and C. Hsiao. 1982. “Formulation and Estimation of Dynamic Models Using Panel Data.” *Journal of Econometrics* 18 (1): 47–82.
- Bai, J. 1997. “Estimating Multiple Breaks One at a Time.” *Econometric Theory* 13 (3): 315–52.
- Blundell, R., and S. Bond. 1998. “Initial Conditions and Moment Restrictions in Dynamic Panel Data Models.” *Journal of Econometrics* 87 (1): 115–43.
- Chernozhukov, V., and H. Hong. 2004. “Likelihood Estimation and Inference in a Class of Nonregular Econometric Models.” *Econometrica* 72 (5): 1445–80.
- Cross, J. L., C. Hou, G. Koop, and A. Poon. 2023. “Large Stochastic Volatility in Mean VARs.” *Journal of Econometrics* 236 (1): 105469.
- Davies, R. B. 1977. “Hypothesis Testing When a Nuisance Parameter is Present Only Under the Alternative.” *Biometrika* 64 (2): 247–54.
- Davies, R. B. 1987. “Hypothesis Testing When a Nuisance Parameter is Present Only Under the Alternative.” *Biometrika* 74 (1): 33–43.
- 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 (6): 835–54.
- Gallant, A. R. 2023. “Variance-Covariance from a Metropolis Chain on a Curved, Singular Manifold.” *Journal of Econometrics* 235 (2): 843–61.
- Gelman, A., and D. B. Rubin. 1992. “Inference from Iterative Simulation Using Multiple Sequences.” *Statistical Science* 7 (4): 457–511.

- Hansen, B. E. 1996. "Inference When a Nuisance Parameter is not Identified Under the Null Hypothesis." *Econometrica: Journal of the Econometric Society* 64 (2): 413–30.
- Hansen, B. E. 1999. "Threshold Effects in Non-Dynamic Panels: Estimation, Testing, and Inference." *Journal of Econometrics* 93 (2): 345–68.
- Hansen, B. E. 2000. "Sample Splitting and Threshold Estimation." *Econometrica* 68 (3): 575–603.
- Hansen, B. E. 2017. "Regression Kink with an Unknown Threshold." *Journal of Business and Economic Statistics* 35 (2): 228–40.
- Hartig, F., F. Minunno, and S. Paul. 2023. BayesianTools: General-Purpose MCMC and SMC Samplers and Tools for Bayesian Statistics. Available at <https://cran.r-project.org/web/packages/BayesianTools/index.html>.
- Hsiao, C., M. H. Pesaran, and A. K. Tahmiscioglu. 2002. "Maximum Likelihood Estimation of Fixed Effects Dynamic Panel Data Models Covering Short Time Periods." *Journal of Econometrics* 109 (1): 107–50.
- Khan, M. S., and A. S. Ssnhadji. 2001. "Threshold Effects in the Relationship Between Inflation and Growth." *IMF Staff Papers* 48 (1): 1–21.
- Kremer, S., A. Bick, and D. Nautz. 2013. "Inflation and Growth: New Evidence from a Dynamic Panel Threshold Analysis." *Empirical Economics* 44 (2): 861–78.
- Lee, S., Y. Liao, M. H. Seo, and Y. Shin. 2021. "Factor-Driven Two-Regime Regression." *Annals of Statistics* 49 (3): 1656–78.
- Lee, Y., and Y. Wang. 2023. "Threshold Regression with Nonparametric Sample Splitting." *Journal of Econometrics* 235 (2): 816–42.
- Li, Y., J. Yu, and T. Zeng. 2018. "Specification Tests Based on MCMC Output." *Journal of Econometrics* 207 (1): 237–60.
- Moral-Benito, E., P. Allison, and R. Williams. 2019. "Dynamic Panel Data Modelling Using Maximum Likelihood: An Alternative to Arellano-Bond." *Applied Economics* 51 (20): 2221–32.
- Newey, W. K., and D. McFadden. 1994. "Chapter 36 Large Sample Estimation and Hypothesis Testing." *Handbook of Econometrics* 4 (05): 2111–245.
- Norets, A., and K. Shimizu. 2024. "Semiparametric Bayesian Estimation of Dynamic Discrete Choice Models." *Journal of Econometrics* 238 (2): 105642.
- Ramírez-Rondán, N. R. 2020. "Maximum Likelihood Estimation of Dynamic Panel Threshold Models." *Econometric Reviews* 39 (3): 260–76.
- Seo, M. H., and Y. Shin. 2016. "Dynamic Panels with Threshold Effect and Endogeneity." *Journal of Econometrics* 195 (2): 169–86.
- Vaona, A., and S. Schiavo. 2007. "Nonparametric and Semiparametric Evidence on the Long-Run Effects of Inflation on Growth." *Economics Letters* 94 (3): 452–8.
- Vrugt, J. A., C. J. F. Ter Braak, C. G. H. Diks, B. A. Robinson, J. M. Hyman, and D. Higdon. 2009. "Accelerating Markov Chain Monte Carlo Simulation by Differential Evolution with Self-Adaptive Randomized Subspace Sampling." *International Journal of Nonlinear Sciences and Numerical Simulation* 10 (3): 273–90.
- Wang, K., and Y. Zhu. 2018. "M-Estimation in High-Dimensional Linear Model." *Journal of Inequalities and Applications* 2018 (1): 225.
- Wooldridge, J. M. 1994. "Estimation and Inference for Dependent Processes." *Handbook of Econometrics* 4: 2639–738.
- Yang, L. 2024. "Panel Threshold Model with Covariate-Dependent Thresholds and its Application to the Cash Flow/Investment Relationship." *Studies in Nonlinear Dynamics and Econometrics* 28 (4): 645–59.
- Yang, L., C. Lee, and I-P. Chen. 2021a. "Threshold Model with a Time-Varying Threshold Based on Fourier Approximation." *Journal of Time Series Analysis* 42 (4): 406–30.
- Yang, L., and J. J. Su. 2018. "Debt and Growth: Is There a Constant Tipping Point?" *Journal of International Money and Finance* 87: 133–43.
- Yang, L., C. Zhang, C. Lee, and I-P. Chen. 2021b. "Panel Kink Threshold Regression Model with a Covariate-Dependent Threshold." *The Econometrics Journal* 24 (3): 462–81.
- Yu, P., and X. Fan. 2021. "Threshold Regression with a Threshold Boundary." *Journal of Business and Economic Statistics* 39 (4): 953–71.