Panel Threshold Mixed Data Sampling Models with a Covariate-Dependent Threshold*

December 2, 2024

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

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

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

Yihang Ye
School of Mathematics and Statistics, Lanzhou University, Lanzhou, China.
E-mail address: 220220935081@lzu.edu.cn

*The first author thanks the support of the National Natural Science Foundation of China (Grant No. 72273059).

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

Panel Threshold Mixed Data Sampling Models with a Covariate-Dependent Threshold

Abstract

This paper introduces a panel threshold mixed data sampling model with a covariate-dependent threshold and unobserved individual-specific threshold effects (PTMIDAS-CDT), in which we allow for a covariate-dependent threshold effect in the relationship between dependent and independent variables sampled at different frequencies, and allow for unobserved individual-specific threshold effects. Based on the Chamberlain-Mundlak correlated random effects (CRE) device and Markov chain Monte Carlo (MCMC) technique, we develop the estimator of model parameters, and suggest test statistics for threshold effect, threshold constancy, the equal weighting scheme, and unobserved individual-specific threshold effects. We establish the asymptotic properties of the proposed estimator in the small-threshold-effect framework and derive the limiting distributions of the suggested test statistics. Monte Carlo simulations are conducted to examine the performance properties of the estimation and testing procedures. The simulation results point out that the estimation procedure works well in finite samples, and the test statistics have good size and power properties. The model is illustrated with an application to the nexus between climate change and economic growth.

Keywords : Panel threshold model; Mixed data sampling (MIDAS); unobserved individual-specific threshold effects; Covariate-dependent threshold; Climate change.

JEL classification : C12, C13, C14, C23.

1 Introduction

Over the last few decades, two distinct strains of models – panel threshold models and mixed data sampling (MIDAS) models – have drawn considerable attention among econometricians and empirical researchers. Compared with linear models, threshold models are advocated due mainly to the fact that they have clearer interpretation (e.g., Chen, 2015; Hansen, 1999, 2000, 2017; Liu and Chen, 2020), and they have superior forecast performance (e.g., Hidalgo et al., 2019). However, the classical threshold models have been criticized for their assumption of a constant threshold setting, and hence a number of recent studies focus on relaxing this assumption and extending threshold models by allowing for a nonconstant threshold (e.g., Seo and Linton, 2007; Yang, 2020; Yu and Fan, 2021; Yang et al., 2021; Lee et al., 2021; Lee and Wang, 2023). On the other hand, by extracting information from high frequency variables, mixed data sampling regression (MIDAS) models can improve the efficiency of estimators and/or enhance the forecast accuracy of target variables observed at a lower frequency, and are first studied in the time series framework by Ghysels et al. (2004, 2006), Ghysels and Wright (2009), Galvao (2013), Miller (2014), Ghysels and Miller (2015), among others. More recently, a number of authors including Khalaf et al. (2021) and Yang et al. (2023) extend MIDAS models to the panel data framework, and tend to focus on the estimation and inference for model parameters unlike the usual MIDAS models focusing mainly on forecasting; they further empirically illustrate that different weighing schemes of an explanatory variable can drastically change a major result in applications, which is particularly important as an inappropriate choice of weighting scheme could result in biased estimates and unreliable inferences (e.g., Andreou et al., 2010; Khalaf et al., 2021; Yang et al., 2023). However, the existing panel threshold models often assume that the data are sampled at the same frequency, which is impractical in many applications.

In this paper, we contribute to the threshold and panel MIDAS literature by introducing a new model called panel threshold mixed data sampling regression with a covariate-dependent threshold and unobserved individual-specific threshold effects (PTMIDAS-CDT) by allowing for a threshold effect determined by a time-varying threshold modeled as a function of informative covariates explaining variation in thresholds, and allowing for some of independent variables to be sampled at a frequency higher than the dependent variable. Consequently, the PTMIDAS-CDT model can be treated as both an extension of the classical MIDAS model of Ghysels et al. (2004, 2006) to a panel data framework with a threshold effect, and an extension of threshold models with a time-varying threshold of Yu and Fan (2021) and Yang et al. (2024) to a mixed data sampling framework. Our model can be regarded as an extension of the panel threshold models of Hansen (1999) and Yu et al. (2022) to a MIDAS setting with a nonconstant threshold, and can also be treated as an extension of the panel MIDAS models of Khalaf et al. (2021) and Yang et al. (2023) to a threshold setting. Hence, the PTMIDAS-CDT model not only has the important advantage of the classical threshold model in nonlinear effects that widely exist in different fields of economics, but also enjoys the efficiency gains from extracting information from high frequency variables which is an important feature of MIDAS models. In addition, it is worth noting that another important feature of our model is the allowance for unobserved individual-specific threshold effects.[1] To the best of our knowledge, Yu et al. (2022) is the first work to introduce individual-specific threshold effects in panel threshold models; more recently, Yang et al. (2024) extend the model of Yu et al. (2022) to a framework with a covariate-dependent threshold.[2]

As emphasized by Yu et al. (2022), the allowance for unobserved individual-specific threshold effects implies that the intercept terms can be different in different regimes, which is a distinguishing feature from traditional panel data models. This leads to difficulty in estimation, as unobserved individual fixed effects can not be eliminated using the within-group transformation or first differencing. Following Yu et al. (2022), we overcome this difficulty by using the Chamberlain-Mundlak correlated random effects (CRE) device to control the endogeneity caused by individual fixed effects.[3] The main idea of this device is to project the individual fixed effects onto the time averages of the explanatory variables. Based on the CRE device and Markov chain Monte Carlo (MCMC) technique, we develop the estimation method of the suggested model, and propose test statistics for threshold effect, the equal weighting scheme, threshold constancy and unobserved individual-specific threshold effects. In estimation, we propose to use a two-step procedure. In the first step, one can estimate MIDAS models using non-linear least squares (NLS) approach for any given threshold value, and in the second step we estimate the threshold parameters using an MCMC-based algorithm which is more efficient than the grid-search method widely used in the threshold literature. In the specification testing, we propose four test statistics, i.e., a test statistic for testing the presence of threshold effect, a test statistic for testing the null hypothesis of equal weights (simple average) in aggregating higher-frequency time series data before estimating econometric models, a test statistic for testing the null of threshold constancy, and a test for unobserved individual-specific threshold effects. Then, employing the small threshold effect assumption which is conventional in threshold models since Hansen (2000), we derive the asymptotic properties of the proposed estimator and test statistics and establish the asymptotic distributions of the estimator and test statistics under the large-N and large-T setup where N and T denote the number of cross-sectional individuals and the number of time periods, respectively. Moreover, we conduct Monte Carlo simulations to examine the finite-sample performances of the estimation and testing procedures. Our

[1]We thank anonymous referees for raising this point with us.

[2]It is worth noting that our model differs from Yu et al. (2022) and Yang et al. (2024) as we consider a panel threshold setting with MIDAS regressors; moreover, our theoretical results are established under the large-N and large-T setup, while they derive the asymptotic results under the large-N and fixed-T setup.

[3]Yang et al. (2023) extend the CRE device to the linear panel MIDAS model.

simulation results point out that the estimation procedure works well, and the test statistics are correctly sized and have desired power properties in finite samples.

To illustrate the model, we investigate whether there is a covariate-dependent threshold effect in the relationship between climate change and economic growth. This relationship has attracted much attention among academics and policy makers in recent years, while the existing studies have obtained mixed results (e.g., Dell et al., 2012; Burke et al., 2015; Lee et al., 2020; Yang et al., 2023). In a recent study, Yang et al. (2023) find that using the MIDAS-type model to consider the information contained in temperature of different seasons can lead to very different empirical results. We add to the literature on the relationship between climate change and economic growth by considering a covariate-dependent threshold in the MIDAS setting to investigate whether there is a temperature threshold (optimal average annual temperature) above which temperature reduces growth and below which temperature is positively associated with growth, following the literature such as Burke et al. (2015) and Lee et al. (2020). Applying the new model, we find the evidence supporting a temperature threshold around 16°C for poor countries but 25°C for rich countries, and hence provide evidence supporting that the impact of future warming on growth would attenuate when countries become richer. This result is different from either Burke et al. (2015) who found a universal temperature threshold around 13°C, or Lee et al. (2020) who found a universal temperature threshold around 14.2°C. Moreover, we find that a temperature threshold effect is not supported by the data if we overlook the feature of threshold nonconstancy. Both simulation and empirical results demonstrate the usefulness of the proposed model.

The remainder of this paper is organized as follows: Section 2 introduces the threshold mixed data sampling model with a covariate-dependent threshold and unobserved individual-specific threshold effects (PTMIDAS-CDT), develops the estimation method of the model parameters, and constructs tests for threshold effect, equal weights, threshold constancy and unobserved individual-specific threshold effects. Sections 3 presents Monte Carlo experiments evaluating the finite-sample properties of the proposed estimator and the test statistics. Section 4 provides an empirical application. Section 5 concludes. All proofs are collected in Appendix A, and a simple forecasting exercise is conducted in Appendix B.

2 The model

Consider the mixed data sampling process $\{y_{it}, \mathbf{z}_{it}, \mathbf{x}_{it}^{(m)}, q_{it}, \mathbf{s}_{it}\}$, where $y_{it}, \mathbf{z}_{it}$, q_{it} and $\mathbf{s}_{it}$ are observed at $i = 1, 2, \ldots, N$, $t = 1, 2, \ldots, T$, and index t represents a low frequency, $\mathbf{x}_{it}^{(m)} = (x_{1,it}^{(m)}, ..., x_{p,it}^{(m)})'$ is a p-dimensional vector of higher frequency data, and the superscript m represents that the series are observed m times between $t-1$ and t. The panel threshold mixed data sampling regression with a covariate-dependent threshold and unobserved individual-specific threshold effects (PTMIDAS-CDT) is given by[4]

$$y_{it} = \begin{cases} \mathbf{z}_{it}'\boldsymbol{\alpha}_1 + \mathbf{x}_{it}'(\boldsymbol{\theta}_1)\boldsymbol{\beta}_1 + \mu_{1i} + \sigma_1 e_{it}, \text{ if } q_{it} \le s_{it}(\boldsymbol{\gamma}) \\ \mathbf{z}_{it}'\boldsymbol{\alpha}_2 + \mathbf{x}_{it}'(\boldsymbol{\theta}_2)\boldsymbol{\beta}_2 + \mu_{2i} + \sigma_2 e_{it}, \text{ if } q_{it} > s_{it}(\boldsymbol{\gamma}) \end{cases}, \tag{1}$$

where $y_{it}, \mathbf{z}_{it}, \mathbf{x}_{it}(\boldsymbol{\theta}_\ell)$, $\ell = 1, 2$ and q_{it} are assumed to be weakly dependent, q_{it} is the threshold variable and is used to split the sample into two subgroups. $\mathbf{z}_{it}$ is a vector of independent variables sampled at a low frequency. μ_{1i} and μ_{2i} denote the unobserved individual heterogeneity which may be correlated with $\mathbf{z}_{it}$ and $\mathbf{x}_{it}(\boldsymbol{\theta}_\ell)$. e_{it} is the disturbance term with $E(e_{it}^2) = 1$. σ_1 and σ_2 capture a threshold effect in the conditional variance of y_{it}. Moreover, following the recent literature (e.g., Yu and Fan, 2021; Yang et al, 2021), the covariate-dependent threshold $s_{it}(\boldsymbol{\gamma})$ is specified as

$$s_{it}(\boldsymbol{\gamma}) = \gamma_0 + \mathbf{s}_{it}'\boldsymbol{\gamma}_1, \tag{2}$$

where $\boldsymbol{\gamma} = (\gamma_0, \boldsymbol{\gamma}_1')' \in \Gamma \subset \mathbb{R}^{d+1}$ and there may be overlap between $\mathbf{z}_{it}$ and $\mathbf{s}_{it}$. In particular, our model allows that both $\mathbf{z}_{it}$ and $\mathbf{s}_{it}$ contain lagged dependent variables such as $y_{i,t-1}$.[5] In addition, $\mathbf{x}_{it}(\boldsymbol{\theta}_\ell) = (x_{1,it}^{(m)}(\boldsymbol{\theta}_{\ell 1}), x_{2,it}^{(m)}(\boldsymbol{\theta}_{\ell 2}), ..., x_{p,it}^{(m)}(\boldsymbol{\theta}_{\ell p}))'$ is a nonlinear function mapping the higher frequency data into a low frequency such that

$$x_{k,i,t}^{(m)}(\boldsymbol{\theta}_{\ell k}) = W(L^{1/m}, \boldsymbol{\theta}_{\ell k})x_{k,it}^{(m)} = \sum_{j=1}^{J} w_{j,k}(\boldsymbol{\theta}_{\ell k})L^{j/m}x_{k,it}^{(m)}, \text{ for } k = 1, 2, ..., p,$$

in which $L^{j/m}$ is the high frequency lag operator such that $L^{j/m}x_{k,it}^{(m)} = x_{k,i,t-(j/m)}^{(m)}$, and J is the number of high frequency lags used in the temporal aggregation of $\mathbf{x}_{it}^{(m)}$ such that $J \ge m$. To identify the slope parameters $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$, we assume that $0 < w_{j,k}(\boldsymbol{\theta}_{\ell k}) < 1$ and

[4] In our model, we assume that the dependent variable y_{it} is a scalar and exclude the case with a multivariate dependent variable. Future research can focus on the latter case.

[5] This is because the structures of nondynamic models and dynamic models in the CRE setting are similar, and thus, the estimation and inference methods developed in nondynamic models are valid in dynamic models with a slight modification following Yu et al. (2022) and Yang et al. (2024).

$\sum_{j=1}^{J} w_{j,k}(\boldsymbol{\theta}_{\ell k}) = 1.$

Following the MIDAS literature (see, e.g., Ghysels et al., 2007, and Andreou et al., 2010), we use the two parameter exponential Almon lag polynomial for the lag coefficients:[6]

$$w_{j,k}(\boldsymbol{\theta}_{\ell k}) = w_{j,k}(\theta_{\ell k,1}, \theta_{\ell k,2}) = \frac{\exp(\theta_{\ell k,1} j + \theta_{\ell k,2} j^2)}{\sum_{j=1}^{J} \exp(\theta_{\ell k,1} j + \theta_{\ell k,2} j^2)}. \tag{3}$$

When $\boldsymbol{\theta}_{\ell k} = (\theta_{\ell k,1}, \theta_{\ell k,2})' = (0,0)'$ in the weighting function (3), the PTMIDAS-CDT model simplifies to the model with high frequency data being aggregated using equal weights. This feature can be used to test model specification, which will be discussed later.

2.1 Model Estimation

For ease of manipulation, we express the model defined in (1) and (2) in a more compacted form. Define the following notation, for $\ell = 1, 2$, $k = 1, \ldots, p$, $\boldsymbol{\beta}_\ell = (\beta_{\ell 1}, \ldots, \beta_{\ell p})'$, $\boldsymbol{\theta}_\ell = (\boldsymbol{\theta}'_{\ell 1}, \ldots, \boldsymbol{\theta}'_{\ell p})'$, $\boldsymbol{\omega}(\boldsymbol{\theta}_{\ell k}) = (\omega_{1,k}(\boldsymbol{\theta}_{\ell k}), \ldots, \omega_{J,k}(\boldsymbol{\theta}_{\ell k}))'$, $\boldsymbol{\omega}_{\boldsymbol{\beta}}(\boldsymbol{\beta}_\ell, \boldsymbol{\theta}_\ell) = (\beta_{\ell 1}\boldsymbol{\omega}'(\boldsymbol{\theta}_{\ell 1}), \ldots, \beta_{\ell p}\boldsymbol{\omega}'(\boldsymbol{\theta}_{\ell p}))'$, $\boldsymbol{x}^{(m)}_{k,i,t} = \left(x^{(m)}_{k,i,t-1/m}, x^{(m)}_{k,i,t-2/m}, \ldots, x^{(m)}_{k,i,t-J/m}\right)'$. Then, the nonlinear function mapping higher frequency data can be rewritten as

$$\mathbf{x}'_{it}(\boldsymbol{\theta}_\ell)\boldsymbol{\beta}_\ell = \sum_{k=1}^{p} x^{(m)}_{k,i,t}(\boldsymbol{\theta}_{\ell k})\beta_{\ell k} = \sum_{k=1}^{p} \beta_{\ell k}\boldsymbol{\omega}'\left(\boldsymbol{\theta}_{\ell k}\right)\boldsymbol{x}^{(m)}_{k,it} = \mathbf{x}^{(m)'}_{it}\boldsymbol{\omega}_{\boldsymbol{\beta}}\left(\boldsymbol{\beta}_\ell, \boldsymbol{\theta}_\ell\right), \tag{4}$$

and the model defined in (1) and (2) can be rewritten as

$$y_{it} = \begin{cases} \mathbf{z}'_{it}\boldsymbol{\alpha}_1 + \mathbf{x}^{(m)'}_{it}\boldsymbol{\omega}_{\boldsymbol{\beta}}\left(\boldsymbol{\beta}_1, \boldsymbol{\theta}_1\right) + \mu_{1i} + \sigma_1 e_{it} & , \text{ if } q_{it} \le s_{it}(\boldsymbol{\gamma}) \\ \mathbf{z}'_{it}\boldsymbol{\alpha}_2 + \mathbf{x}^{(m)'}_{it}\boldsymbol{\omega}_{\boldsymbol{\beta}}\left(\boldsymbol{\beta}_2, \boldsymbol{\theta}_2\right) + \mu_{2i} + \sigma_2 e_{it} & , \text{ if } q_{it} > s_{it}(\boldsymbol{\gamma}) \end{cases}. \tag{5}$$

As illustrated by Yu et al. (2022), the inclusion of unobserved individual-specific threshold effects would lead to the failure of the traditional estimation methods such as the within-group transformation; therefore, Yu et al. (2022) suggest to take the correlated random effects (CRE) model and use Chamberlain-Mundlak CRE device to control the endogeneity. Following Mundlak (1978) and Yu et al. (2022), we assume that μ_{1i} and μ_{2i} in model (1) are given as follows

$$\mu_{\ell i} = \boldsymbol{\psi}'_\ell \mathbf{h}_i + a_{\ell i} \ (\ell = 1, 2) \text{ with } E[a_{\ell i}|\mathcal{F}_{NT,t-1}] = 0, \text{and } E[e_{it}|\mathcal{F}_{NT,t-1}] = 0, \tag{6}$$

[6] A more general polynomial function is given by $w_{j,k}(\boldsymbol{\theta}_k) = w_{j,k}(\theta_{k,1}, \ldots, \theta_{k,q}) = \frac{\exp(\theta_{k,1} j + \theta_{k,2} j^2 + \ldots + \theta_{k,q} j^q)}{\sum_{j=1}^{J} \exp(\theta_{k,1} j + \theta_{k,2} j^2 + \ldots + \theta_{k,q} j^q)}$. Ghysels et al. (2007) illustrate that, even with only two parameters, the function is flexible enough to mimic different weighting shapes.

where $\mathbf{h}_i = (\bar{\mathbf{z}}_i', \bar{\mathbf{x}}_i', \underline{\mathbf{h}}_i')'$ with $\bar{\mathbf{z}}_i = \frac{1}{T}\sum_{t=1}^T \mathbf{z}_{it}$ and $\bar{\mathbf{x}}_i = \frac{1}{T}\sum_{t=1}^T \mathbf{x}_{it}^{(m)}$,[7] the filter $\mathcal{F}_{NT,t} = \sigma\left(\{\mathbf{z}_{i1}, \ldots, \mathbf{z}_{iT}, \mathbf{x}_{i1}^{(m)}, \ldots, \mathbf{x}_{iT}^{(m)}, \underline{\mathbf{h}}_i, e_{it}, e_{i,t-1}, \ldots\}_{i=1}^N\right)$ in which $\underline{\mathbf{h}}_i$ contains the time-invariant variables such as the constant 1. Here, $\mathbf{h}_i$ is used to control the time-invariant effect, and we allow that $Cov(a_{1i}, a_{2i}) \neq 0$, and $a_{1i} \neq a_{2i}$; thus, the correlation between μ_{1i} and μ_{2i} is through $\mathbf{h}_i$ or (a_{1i}, a_{2i}). As in Yu et al. (2022), a_{1i} and a_{2i} can be correlated with e_{it}.

In the panel data literature, when the cross section dimension N is small and the time series dimension T is large, a popular approach is to treat the equations from the different cross section units as a system of seemingly unrelated regression equations (SURE) and then estimate the system by the generalized least squares (GLS) techniques (e.g., Pesaran, 2015). Thus, we can estimate the proposed model by treating the fixed effects μ_{1i} and μ_{2i} as parameters that are jointly estimated with other model parameters when T is large and N is small.[8] However, the application of the SURE-GLS approach to large N and T panels involves nuisance parameters that increase at a quadratic rate as the cross section dimension of the panel is allowed to rise (e.g., Pesaran, 2006). Therefore, the proposed CRE approach seems to be appropriate in the case where N is of the same order of magnitude or greater than T, which is often the case in cross-country studies including our empirical application in which our dataset contains 78 countries over the period from 1961 to 2003.

Substituting (6) into (5), we have

$$\begin{aligned} E[y_{it}|\mathcal{F}_{NT,t-1}] &= \left(\mathbf{z}_{it}'\boldsymbol{\alpha}_1 + \mathbf{x}_{it}^{(m)\prime}\boldsymbol{\omega}_{\boldsymbol{\beta}}(\boldsymbol{\beta}_1, \boldsymbol{\theta}_1) + \boldsymbol{\psi}_1'\mathbf{h}_i\right) I(q_{it} \leq s_{it}(\boldsymbol{\gamma})) \\ &\quad + \left(\mathbf{z}_{it}'\boldsymbol{\alpha}_2 + \mathbf{x}_{it}^{(m)\prime}\boldsymbol{\omega}_{\boldsymbol{\beta}}(\boldsymbol{\beta}_2, \boldsymbol{\theta}_2) + \boldsymbol{\psi}_2'\mathbf{h}_i\right) I(q_{it} > s_{it}(\boldsymbol{\gamma})), \end{aligned} \tag{7}$$

where $I(\cdot)$ is the indicator function. In addition, we define the error term as

$$\begin{aligned} u_{it} &= (a_{1i} + \sigma_1 e_{it}) I(q_{it} \leq s_{it}(\boldsymbol{\gamma})) + (a_{2i} + \sigma_2 e_{it}) I(q_{it} > s_{it}(\boldsymbol{\gamma})), \\ &=: e_{1it} I(q_{it} \leq s_{it}(\boldsymbol{\gamma})) + e_{2it} I(q_{it} > s_{it}(\boldsymbol{\gamma})). \end{aligned} \tag{8}$$

Define $\check{\mathbf{x}}_{it} = (\mathbf{z}_{it}', \mathbf{x}_{it}^{(m)\prime}, \mathbf{h}_i')'$, $\boldsymbol{\phi}_\ell = (\boldsymbol{\alpha}_\ell', \boldsymbol{\beta}_\ell', \boldsymbol{\theta}_\ell', \boldsymbol{\psi}_\ell')'$ and $\boldsymbol{g}(\boldsymbol{\phi}_\ell) = (\boldsymbol{\alpha}_\ell', \boldsymbol{\omega}_{\boldsymbol{\beta}}'(\boldsymbol{\beta}_\ell, \boldsymbol{\theta}_\ell), \boldsymbol{\psi}_\ell')'$ for

[7] If $\mathbf{z}_{it}$ does not contain q_{it} and $\mathbf{s}_{it}$, $\mathbf{h}_i$ should include $\bar{q}_i$ and $\bar{\mathbf{s}}_i$, while $\mathbf{h}_i$ should exclude the time averages of lagged dependent variables. In the dynamic case, $\mathbf{h}_i$ should include y_{i0} to control the initial condition effect; see Yu et al. (2022) for more details. Here, following Yang et al. (2023), we project the individual fixed effects onto the time averages of the high frequency series.

[8] We thank an anonymous referee for raising this point with us. Future research can focus on this issue.

$\ell = 1, 2$. Then, the objective function can be written as

$$
\begin{aligned}
& \widetilde{SSR}_{NT}(\boldsymbol{\phi}, \boldsymbol{\gamma}) \\
= & \sum_{i=1}^{N} \sum_{t=1}^{T} \left[y_{it} - \check{\mathbf{x}}_{it}' \boldsymbol{g}\left(\boldsymbol{\phi}_1\right) I(q_{it} \leq s_{it}(\boldsymbol{\gamma})) - \check{\mathbf{x}}_{it}' \boldsymbol{g}\left(\boldsymbol{\phi}_2\right) I(q_{it} > s_{it}(\boldsymbol{\gamma})) \right]^2 ,
\end{aligned}
\tag{9}
$$

where $\boldsymbol{\phi} = (\boldsymbol{\phi}_1', \boldsymbol{\phi}_2')'$.

Following the threshold literature (e.g., Hansen, 1999), a two-step procedure can be used to estimate $\boldsymbol{\gamma}$. First, as the minimizer in (9) is equal to the estimator of a standard MIDAS model for any fixed $\boldsymbol{\gamma}$, we thus employ the non-linear least squares (NLS) estimator $\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma})$ following the MIDAS literature. We assume $\boldsymbol{\gamma} \in \boldsymbol{\Gamma}$ and $\boldsymbol{\phi} \in \boldsymbol{\Phi}$, where the parameter spaces $\boldsymbol{\Gamma}$ and $\boldsymbol{\Phi}$ are bounded sets of the reals. Then NLS estimator is given as

$$
\widehat{\boldsymbol{\phi}}(\boldsymbol{\gamma}) = \underset{\boldsymbol{\phi} \in \boldsymbol{\Phi}}{\arg\min}\, \widetilde{SSR}_{NT}(\boldsymbol{\phi}, \boldsymbol{\gamma}). \tag{10}
$$

Define $\widetilde{SSR}_{NT}(\boldsymbol{\gamma}) = \widetilde{SSR}_{NT}(\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}), \boldsymbol{\gamma})$. Then, the threshold parameters can be estimated as

$$
\hat{\boldsymbol{\gamma}} = (\hat{\gamma}_0, \hat{\boldsymbol{\gamma}}_1')' = \underset{\boldsymbol{\gamma} \in \boldsymbol{\Gamma}}{\arg\min}\, \widetilde{SSR}_{NT}(\boldsymbol{\gamma}), \tag{11}
$$

where $\boldsymbol{\Gamma} = \Gamma_0 \times \Gamma_1$, Γ_0 and $\Gamma_1 = \Gamma_{11} \times \Gamma_{12} \times \cdots \times \Gamma_{1d}$ are the parameter spaces and assumed to be compact. Once $\hat{\boldsymbol{\gamma}}$ is obtained, the other parameters can be estimated as $\hat{\boldsymbol{\phi}} = \hat{\boldsymbol{\phi}}(\hat{\boldsymbol{\gamma}})$.

To implement the above minimization, one may use a two-step approach based on concentration and grid search as suggested by Hansen (1999). First, for any given γ_0, we compute the sum of squared errors $\widetilde{SSR}_{NT}(\gamma_0, \boldsymbol{\gamma}_1')$ for each $\boldsymbol{\gamma}_1 \in \boldsymbol{\Gamma}_1$. Second, fixed γ_0, we solve the minimum $\widetilde{SSR}_{NT}(\gamma_0, \hat{\boldsymbol{\gamma}}_1'(\gamma_0)) = \underset{\boldsymbol{\gamma}_1 \in \boldsymbol{\Gamma}_1}{\arg\min} \widetilde{SSR}_{NT}(\gamma_0, \boldsymbol{\gamma}_1')$, and then the threshold parameters can be estimated as $\widehat{\boldsymbol{\gamma}} = (\hat{\gamma}_0, \hat{\boldsymbol{\gamma}}_1')' = (\hat{\gamma}_0, \hat{\boldsymbol{\gamma}}_1'(\hat{\gamma}_0))' = \underset{\gamma_0 \in \Gamma_0}{\arg\min} \widetilde{SSR}_{NT}(\gamma_0, \hat{\boldsymbol{\gamma}}_1'(\gamma_0))$. As illustrated by Hansen (1999), we may restrict the grid search to values of $\widehat{\boldsymbol{\gamma}}$ such that a minimal percentage of the observations (say, 10% or 15%) lie in both regimes, as it is not desirable to sort too few data into one or the other regime. However, the grid search method may have a practical difficulty when the dimension of $\mathbf{s}_{it}$ becomes large; thus, following Yu and Fan (2021), we suggest to adopt an MCMC-based method to ease the computation burden in estimating threshold parameters.

Algorithm 1. Estimation based on the MCMC technique.

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

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

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

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

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

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

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

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

Following the spirit of Yu and Fan (2021), our MCMC-based algorithm is only auxiliary to the minimization problem in (11) by simulating the candidate minimizers. The MCMC-based procedure is more efficient than the grid search approach, as we can simulate $\boldsymbol{\gamma}$ with higher probability when $S_N(\boldsymbol{\gamma})$ is smaller. When B is large enough, the global minimizer of $S_N(\boldsymbol{\gamma})$ can be achieved as illustrated by Yu and Fan (2021). It is worth noting that the computational cost of the grid search method would increase with the sample size NT and the dimension of $\mathbf{s}_{it}$, while the computational cost of the MCMC algorithm increases slowly with the sample size and the dimension of $\mathbf{s}_{it}$, as the MCMC algorithm requires roughly B regressions as can be seen from Algorithm 1.[9]

Let $f_{it}(q_{it} \mid \mathbf{s}_{it})$ be the conditional distribution of the threshold variable q_{it} given $\mathbf{s}_{it}$ and $f_{k|t,\mathbf{s}}(q_{ik} \mid q_{it}, \mathbf{s}_{ik}, \mathbf{s}_{it})$ be the conditional distribution of the threshold variable q_{ik} given q_{it},

[9] In application, one could choose B using the convergence diagnostic of Gelman and Rubin (1992), who propose a statistic to monitor convergence of MCMC samples.

$\mathbf{s}_{ik}$ and $\mathbf{s}_{it}$. Define the moment functionals

$$
\begin{aligned}
\mathbf{M} &= E\left(\check{\mathbf{x}}_{it}\check{\mathbf{x}}_{it}'\right), \\
\mathbf{M}_1(\boldsymbol{\gamma}) &= E\left(\check{\mathbf{x}}_{it}\check{\mathbf{x}}_{it}' I\left(q_{it} \le s_{it}(\boldsymbol{\gamma})\right)\right), \mathbf{M}_2(\boldsymbol{\gamma}) = E\left(\check{\mathbf{x}}_{it}\check{\mathbf{x}}_{it}' I\left(q_{it} > s_{it}(\boldsymbol{\gamma})\right)\right), \\
\boldsymbol{\Omega}_1(\boldsymbol{\gamma}) &= E\left(\check{\mathbf{x}}_{it}\check{\mathbf{x}}_{it}' e_{1it} I\left(q_{it} \le s_{it}(\boldsymbol{\gamma})\right)\right), \boldsymbol{\Omega}_2(\boldsymbol{\gamma}) = E\left(\check{\mathbf{x}}_{it}\check{\mathbf{x}}_{it}' e_{2it} I\left(q_{it} > s_{it}(\boldsymbol{\gamma})\right)\right) \\
\mathbf{M}^*(\boldsymbol{\gamma}) &= \begin{bmatrix} \mathbf{M}_1(\boldsymbol{\gamma}) & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_2(\boldsymbol{\gamma}) \end{bmatrix}, \boldsymbol{\Omega}^*(\boldsymbol{\gamma}) = \begin{bmatrix} \boldsymbol{\Omega}_1(\boldsymbol{\gamma}) & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega}_2(\boldsymbol{\gamma}) \end{bmatrix}, \\
\mathbf{M}^* &= \mathbf{M}^*(\boldsymbol{\gamma}^0), \boldsymbol{\Omega}^* = \boldsymbol{\Omega}^*(\boldsymbol{\gamma}^0) \\
\mathbf{D}_{it}(\boldsymbol{\gamma}) &= E\left(\check{\mathbf{x}}_{it}\check{\mathbf{x}}_{it}' \,\middle|\, q_{it} = s_{it}(\boldsymbol{\gamma}), \mathbf{s}_{it}\right), \mathbf{V}_{\ell it}(\boldsymbol{\gamma}) = E\left(\check{\mathbf{x}}_{it}\check{\mathbf{x}}_{it}' e_{\ell it}^2 \,\middle|\, q_{it} = s_{it}(\boldsymbol{\gamma}), \mathbf{s}_{it}\right).
\end{aligned}
$$

Denote that $\check{\mathbf{x}}_t = (\check{\mathbf{x}}_{1t}', \ldots, \check{\mathbf{x}}_{Nt}')'$, $q_t = (q_{1t}, \ldots, q_{Nt})'$, $\mathbf{s}_t = (\mathbf{s}_{1t}', \ldots, \mathbf{s}_{Nt}')'$, $u_t = (u_{1t}, \ldots, u_{Nt})'$. To establish the asymptotic properties of the proposed estimator of the parameters, we make the following assumptions.

Assumption 1. *We assume that*

1. $\{\check{\mathbf{x}}_t, q_t, \mathbf{s}_t, u_t\}_{t=1}^T$ *is strictly stationary, ergodic and ρ-mixing, with ρ-mixing coefficients satisfying $\sum_{m=1}^{\infty} \rho_m^{1/2} < \infty$, and is i.i.d across i.*
2. $E\left(a_{\ell i} \,|\, \mathcal{F}_{NT,t}\right) = 0$ *and* $E\left(u_{it} \,|\, \mathcal{F}_{NT,t}\right) = 0$.[10]
3. *For each $j = 1, \ldots, p_*$, $P(x_{i1}^j = \cdots = x_{iT}^j) < 1$, where x_{it}^j is the jth element of $(\mathbf{z}_{it}', \mathbf{x}_{it}^{(m)\prime})'$ and p_* is the dimension of $(\mathbf{z}_{it}', \mathbf{x}_{it}^{(m)\prime})'$.*
4. $E\|\check{\mathbf{x}}_{it}\|^4 < \infty$ *and* $E|e_{\ell it}|^4 < \infty$.
5. *For any $\boldsymbol{\gamma} \in \boldsymbol{\Gamma}$, $E\left(\|\check{\mathbf{x}}_{it}\|^4 \,\middle|\, q_{it} = s_{it}(\boldsymbol{\gamma}), \mathbf{s}_{it}\right) \le C$, $E\left(\|\check{\mathbf{x}}_{it} e_{\ell it}\|^4 \,\middle|\, q_{it} = s_{it}(\boldsymbol{\gamma}), \mathbf{s}_{it}\right) \le C$ for some $C < \infty$, and $0 < f_{it}\left(q_{it} \,|\, \mathbf{s}_{it}\right) \le \bar{f} < \infty$.*
6. *$f_{it}\left(s_{it}(\boldsymbol{\gamma}) \,|\, \mathbf{s}_{it}\right)$, $\mathbf{D}_{it}(\boldsymbol{\gamma})$ and $\mathbf{V}_{lit}(\boldsymbol{\gamma})$ are continuous at $\boldsymbol{\gamma} = \boldsymbol{\gamma}^0$, where $\boldsymbol{\gamma}^0$ is the true value of $\boldsymbol{\gamma}$.*
7. *$\boldsymbol{\phi}_\ell = \boldsymbol{\phi}_c + \mathbf{c}_\ell (NT)^{-\alpha}$ with $\mathbf{c}_1 \neq \mathbf{c}_2$ and $0 < \alpha < 1/2$, where $\boldsymbol{\phi}_c$ and $\mathbf{c}_\ell$ are constant vectors.*
8. *$G(\boldsymbol{\omega}) > 0$ and $V_\ell(\boldsymbol{\omega}) > 0$ with $\boldsymbol{\omega} \neq 0$ (defined in Theorem 1).*

[10] To simplify the derivations of the asymptotic results, this paper also assumes that $a_{\ell i}$ for $\ell = 1, 2$ degenerate to a point mass at zero. Under this assumption, Yu et al. (2022) show that the objective function in (9) would be equivalent to a likelihood function. This assumption can be relaxed with more complicated asymptotic results, which are available on request from the authors.

9. $\det(\mathbf{M}) > \det(\mathbf{M}_\ell(\boldsymbol{\gamma})) > 0$ *and* $\det(\boldsymbol{\Omega}) > \det(\boldsymbol{\Omega}_\ell(\boldsymbol{\gamma})) > 0$ *for all* $\boldsymbol{\gamma} \in \boldsymbol{\Gamma}$.

10. *For* $k > t$, $f_{jk|it,\mathbf{s}}\left(s_{jk}(\boldsymbol{\gamma}^0) \,\middle|\, q_{it} = s_{it}(\boldsymbol{\gamma}^0), \mathbf{s}_{jk}, \mathbf{s}_{it}\right) < \infty$.

11. $(1, \mathbf{s}_{it}')'$ *is not multicollinear.*

12. $\boldsymbol{\phi}_\ell^0 \in \operatorname{int}(\Phi_\ell)$, $\ell = 1, 2$, *where* $\boldsymbol{\phi}_\ell^0$ *is the true value of* $\boldsymbol{\phi}_\ell$.

13. *The third derivative matrix of* $\boldsymbol{g}(\boldsymbol{\phi}_\ell)$ *exists and is bounded in a neighborhood of* $\boldsymbol{\phi}_\ell^0$, $\ell = 1, 2$. □

Assumptions 1.1-1.10 above are natural extensions of Assumption in Hansen (1999, 2000), which are very conventional in the literature of threshold models. Assumption 1.1 is similar to Assumption CR (a) of Massacci (2017) under the large-N and large-T setup, and is used in Lemma 3.6 of Peligrad (1982). Assumption 1.2 requires the error term u_{it} to be a martingale difference sequence with respect to the filter $\mathcal{F}_{NT,t}$, allows conditional heteroskedasticity but excludes serial correlation and cross-section dependence in u_{it}. As discussed by Su and Chen (2013) and Miao et al. (2020), if u_{it} has serial dependence, we can add lagged dependent or independent variables into the model to eliminate serial correlation in u_{it}, so this assumption is not restrictive. Assumption 1.3 avoids the multicollinear problem between $(\mathbf{z}_{it}', \mathbf{x}_{it}^{(m)'})'$ and $(\bar{\mathbf{z}}_i', \bar{\mathbf{x}}_i')'$. Assumptions 1.4-1.5 restrict unconditional and conditional moment bounds of $\check{\mathbf{x}}_{it}$ and $\check{\mathbf{x}}_{it} e_{lit}$ to be finite. Assumption 1.6 requires the threshold variable to have a continuous distribution. Assumption 1.7 is well-known as the small threshold effect assumption ($\boldsymbol{\phi}_1 - \boldsymbol{\phi}_2 = (\mathbf{c}_1 - \mathbf{c}_2)(NT)^{-\alpha}$), which is very conventional in the literature (e.g., Hansen, 2000). Assumptions 1.8-1.9 are full rank conditions needed to have nondegenerate asymptotic distributions. Assumption 1.10 excludes the possibility that $q_{it} = s_{it}(\boldsymbol{\gamma}^0)$ for $t = 1, \ldots, T$. Assumption 1.11 implies that the limit of $\widetilde{SSR}_{NT}(\boldsymbol{\gamma})$ has a unique minimum at the true value $\boldsymbol{\gamma}^0$ and ensures that the asymptotic distribution of $\hat{\boldsymbol{\gamma}}$ is well defined, as in Yu and Fan (2021) and Yang et al. (2024). For nonlinear function $\boldsymbol{g}(\boldsymbol{\phi}_\ell)$ we make Assumption 1.12-1.13 following Miller (2014), and these assumptions ensure that conditions (i) and (ii) of Theorem 10.1 of Wooldridge (1994) are satisfied.

We next establish the convergence rates of the threshold estimator and the slope estimator. In the next, $\xrightarrow{d}$ stands for convergence in distribution. $\xrightarrow{p}$ stands for convergence in probability. $(N, T) \to (\infty, \infty)$ denotes that N and T go to infinity jointly.

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

$$(NT)^{1-2\alpha}\left(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}^0\right) \xrightarrow{d} \underset{\boldsymbol{\omega}\in\mathbb{R}^{d+1}}{\arg\min}\left[\frac{1}{2}G(\boldsymbol{\omega})-R(\boldsymbol{\omega})\right], \tag{12}$$

where $R(\boldsymbol{\omega}) = R_1(\boldsymbol{\omega}) + R_2(\boldsymbol{\omega})$, $R_l(\boldsymbol{\omega})$ *is a Gaussian process with a positive variance* $V_l(\boldsymbol{\omega})$ *when* $\boldsymbol{\omega} \neq \mathbf{0}$, *which* $R_1(\boldsymbol{\omega})$ *and* $R_2(\boldsymbol{\omega})$ *are independent.* $G(\boldsymbol{\omega})$ *and* $V_l(\boldsymbol{\omega})$ *are defined as*

$$\begin{aligned}
G(\boldsymbol{\omega}) &= \mathbf{c}_g' E\Big(\mathbf{D}_{it}(\boldsymbol{\gamma}^0) f_{it|\mathbf{s}}\left(s_{it}(\boldsymbol{\gamma}^0)\,|\,\mathbf{s}_{it}\right)\left|\mathbf{s}_{it,1}'\boldsymbol{\omega}\right|\Big)\mathbf{c}_g,\\
V_1(\boldsymbol{\omega}) &= \mathbf{c}_g' E\Big(\mathbf{V}_{1it}(\boldsymbol{\gamma}^0) f_{it|\mathbf{s}}\left(s_{it}(\boldsymbol{\gamma}^0)\,|\,\mathbf{s}_{it}\right)\left|\mathbf{s}_{it,1}'\boldsymbol{\omega}\right| I\left(\mathbf{s}_{it,1}'\boldsymbol{\omega}\le 0\right)\Big)\mathbf{c}_g,\\
V_2(\boldsymbol{\omega}) &= \mathbf{c}_g' E\Big(\mathbf{V}_{2it}(\boldsymbol{\gamma}^0) f_{it|\mathbf{s}}\left(s_{it}(\boldsymbol{\gamma}^0)\,|\,\mathbf{s}_{it}\right)\left|\mathbf{s}_{it,1}'\boldsymbol{\omega}\right| I\left(\mathbf{s}_{it,1}'\boldsymbol{\omega} > 0\right)\Big)\mathbf{c}_g,
\end{aligned}$$

which are positive when $\boldsymbol{\omega} \neq \mathbf{0}$, *where* $\mathbf{c}_g = [\partial \boldsymbol{g}(\phi_c)/\partial\phi'](\mathbf{c}_1-\mathbf{c}_2)$ *and* $\mathbf{s}_{it,1} = (1, \mathbf{s}_{it}')'$.

Proof of Theorem 1. See the Appendix. □

Theorem 2. *Under Assumption 1, as* $(N,T) \to (\infty,\infty)$,

$$\sqrt{NT}\left(\hat{\boldsymbol{\phi}}-\boldsymbol{\phi}^0\right) \xrightarrow{d} \mathcal{Z}, \tag{13}$$

where $\boldsymbol{\phi}^0 = (\boldsymbol{\phi}_1^{0\prime}, \boldsymbol{\phi}_2^{0\prime})'$ *is the true value of* $\boldsymbol{\phi}$, *and* $\mathcal{Z}$ *is a Gaussian process with variance* $\left(\boldsymbol{J_g}'\mathbf{M}^*\boldsymbol{J_g}\right)^{-1}\boldsymbol{J_g}'\boldsymbol{\Omega}^*\boldsymbol{J_g}\left(\boldsymbol{J_g}'\mathbf{M}^*\boldsymbol{J_g}\right)^{-1}$, *in which*

$$\boldsymbol{J_g} = \begin{bmatrix} \left.\frac{\partial \boldsymbol{g}(\boldsymbol{\phi}_1)}{\partial \boldsymbol{\phi}_1'}\right|_{\boldsymbol{\phi}_1=\boldsymbol{\phi}_c} & \mathbf{0} \\ \mathbf{0} & \left.\frac{\partial \boldsymbol{g}(\boldsymbol{\phi}_2)}{\partial \boldsymbol{\phi}_2'}\right|_{\boldsymbol{\phi}_2=\boldsymbol{\phi}_c} \end{bmatrix}. \tag{14}$$

Proof of Theorem 2. See the Appendix. □

Theorems 1 and 2 show that the estimators of $\boldsymbol{\gamma}$ and $\boldsymbol{\phi}$ are consistent, and the convergence rates of $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\phi}}$ are $(NT)^{1-2\alpha}$ and $\sqrt{NT}$, respectively. Our asymptotic results can be viewed as a generalization of Yu and Fan (2021) to the panel MIDAS framework in the large-N and large-T setup. Furthermore, we find that the limiting distribution of $\hat{\boldsymbol{\gamma}}$ and $\hat{\boldsymbol{\phi}}$ has the similar form as in Yu and Fan (2021), except for an additional nonlinear component $\partial \boldsymbol{g}(\boldsymbol{\phi})/\partial\boldsymbol{\phi}'$. Moreover, as in Yu and Fan (2021), the involvement of a covariate-dependent threshold lead to that the limiting distribution of $\hat{\boldsymbol{\gamma}}$ cannot be represented in the form of two-sided Brownian motion as in Hansen (2000). In our setting, the limiting distribution of $\hat{\boldsymbol{\gamma}}$ depends on the nonlinear component and the covariates $\{\mathbf{z}_{it}, \mathbf{x}_{it}^{(m)}, q_{it}, \mathbf{s}_{it}\}$.

To construct confidence intervals for the model parameters, we invert the following statistic for the null hypothesis $\boldsymbol{\gamma} = \boldsymbol{\gamma}_0$, given by

$$LR_{NT}(\boldsymbol{\gamma}) = \frac{S_N(\boldsymbol{\gamma}) - S_N(\hat{\boldsymbol{\gamma}})}{S_N(\hat{\boldsymbol{\gamma}})/(NT)}, \tag{15}$$

The null hypothesis is rejected for large values of $LR_{NT}(\boldsymbol{\gamma}^0)$. For the above null hypothesis, we have the following asymptotic result.

Theorem 3. *Suppose Assumption 1 holds. Under $H_0 : \boldsymbol{\gamma} = \boldsymbol{\gamma}^0$, as $(N, T) \to (\infty, \infty)$, we have*

$$LR_{NT}(\boldsymbol{\gamma}^0) \xrightarrow{d} \frac{1}{\sigma_u^2} \min_{\boldsymbol{\omega} \in \mathbb{R}^{d+1}} [2R(\boldsymbol{\omega}) - G(\boldsymbol{\omega})], \tag{16}$$

where $\sigma_u^2 = E\left(u_{it}^2\right)$, and $G(\boldsymbol{\omega})$ and $R(\boldsymbol{\omega})$ are defined as in Theorem 1.

Proof of Theorem 3. See the Appendix. □

Theorem 3 is a generalized version of Theorem 1 in Hansen (1999). Moreover, we can see that even with homoscedasticity, we still have nuisance parameters, and can not be tabulated. Thus, following the literature (e.g., Hansen, 2017; Yang et al., 2021), we suggest to compute the confidence intervals using a bootstrap procedure.

Algorithm 2. Confidence intervals for parameters

Step 1. Generate i.i.d draws v_{it}^* from the $N(0, 1)$ distribution for $i = 1, \ldots, N$ and $t = 1, \ldots, T$, and set $u_{it}^* = \hat{u}_{it}(\hat{\boldsymbol{\gamma}}) v_{it}^*$, where $\hat{u}_{it}(\hat{\boldsymbol{\gamma}})$ is the residuals from the fitted PTMIDAS-CDT model (1).

Step 2. Set $y_{it}^* = \check{\mathbf{x}}_{it}' \boldsymbol{g}(\hat{\boldsymbol{\phi}}_1) I\left(q_{it} \le s_{it}\left(\hat{\boldsymbol{\gamma}}\right)\right) + \check{\mathbf{x}}_{it}' \boldsymbol{g}(\hat{\boldsymbol{\phi}}_2) I\left(q_{it} > s_{it}\left(\hat{\boldsymbol{\gamma}}\right)\right) + u_{it}^*$, where $(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\gamma}})$ are estimates based on the original sample $\{y_{it}, \mathbf{z}_{it}, \mathbf{x}_{it}^{(m)}, q_{it}, \mathbf{s}_{it}\}$.

Step 3. Use the bootstrap sample $\{y_{it}^*, \mathbf{z}_{it}, \mathbf{x}_{it}^{(m)}, q_{it}, \mathbf{s}_{it}\}$ to estimate the PTMIDAS-CDT model and obtain the parameter estimates $(\hat{\boldsymbol{\phi}}^*, \hat{\boldsymbol{\gamma}}^*)$, and the sum of squared errors $S_N^*(\hat{\boldsymbol{\gamma}}^*)$, in which $S_N^*(\boldsymbol{\gamma}) = \sum_{i=1}^{N} \sum_{t=1}^{T} [y_{it}^* - \check{\mathbf{x}}_{it}' \boldsymbol{g}(\hat{\phi}_1^*(\boldsymbol{\gamma})) I\left(q_{it} \le s_{it}\left(\boldsymbol{\gamma}\right)\right) - \check{\mathbf{x}}_{it}' \boldsymbol{g}(\hat{\phi}_2^*(\boldsymbol{\gamma})) I\left(q_{it} > s_{it}\left(\boldsymbol{\gamma}\right)\right)]^2$, $\widehat{\boldsymbol{\phi}}^*(\boldsymbol{\gamma}) = \arg\min_{\boldsymbol{\phi} \in \boldsymbol{\Phi}} \sum_{i=1}^{N} \sum_{t=1}^{T} [y_{it}^* - \check{\mathbf{x}}_{it}' \boldsymbol{g}\left(\phi_1\right) I(q_{it} \le s_{it}(\boldsymbol{\gamma})) - \check{\mathbf{x}}_{it}' \boldsymbol{g}\left(\phi_2\right) I(q_{it} > s_{it}(\boldsymbol{\gamma}))]^2$.

Step 4. Compute the statistic for $\hat{\boldsymbol{\gamma}}$

$$LR_{NT}^*(\hat{\boldsymbol{\gamma}}) = \frac{S_N^*(\hat{\boldsymbol{\gamma}}) - S_N^*(\hat{\boldsymbol{\gamma}}^*)}{S_N^*(\hat{\boldsymbol{\gamma}}^*)/(NT)}. \tag{17}$$

Step 5. Repeat Steps 1-4 B times, and obtain a sample of simulated coefficient estimates $(\hat{\boldsymbol{\phi}}^*, \hat{\boldsymbol{\gamma}}^*)$ and a sample of $LR_{NT}^*(\hat{\boldsymbol{\gamma}})$. Construct $1 - a$ bootstrap confidence intervals for the estimates $(\hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\gamma}})$ by the symmetric percentile method: the estimates plus and minus the $(1-a)$ quantile of the absolute centered bootstrap estimates. For example, the confidence interval of $\hat{\gamma}_0$ is $\hat{\gamma}_0 \pm q^*_{1-a}$, where q^*_{1-a} is the $1-a$ quantile of $|\hat{\gamma}^*_0 - \hat{\gamma}_0|$.

2.2 Model Specification Testing

In applying the PTMIDAS-CDT model, we may be interested in investigating whether or not there is a threshold effect in the model, whether the simple aggregation using equal weights is supported by empirical data, whether the threshold is constant, and whether there are unobserved individual-specific threshold effects.

To test for the existence of the threshold effect, we consider the null hypothesis $H_0^1 : \phi_1 = \phi_2$. Under this null, the model shrinks to a panel MIDAS model:

$$y_{it} = \check{\mathbf{x}}'_{it}\boldsymbol{g}(\phi_1) + a_{1i} + \sigma_1 e_{it},\ i = 1, 2, \ldots, N,\ t = 1, 2, ..., T. \tag{18}$$

Here, the threshold $\boldsymbol{\gamma}$ is not identified under the null of linearity; hence, the null distribution of test statistic is non-standard due to the well-known Davies's problem, and the limiting distribution can be typically explored by taking the supremum of all possible values of unidentified parameters (see, e.g., Davies, 1987; Hansen, 1996). Denote the sum of squared errors of the MIDAS model (18) as S_0^L. Then the sup-test statistic can be defined as follows:

$$LR_1 = \sup_{\boldsymbol{\gamma}\in\boldsymbol{\Gamma}} \frac{S_0^L - \widetilde{SSR}_{NT}(\boldsymbol{\gamma})}{\widetilde{SSR}_{NT}(\boldsymbol{\gamma}))/(NT)} \equiv \frac{S_0^L - \widetilde{SSR}_{NT}(\hat{\boldsymbol{\gamma}})}{\widetilde{SSR}_{NT}(\hat{\boldsymbol{\gamma}})/(NT)}. \tag{19}$$

In addition, as shown by Andreou et al. (2010), aggregating the high frequency data using equal weights (simple average) would generally lead to an inconsistent estimator; hence, it is also important to test whether or not using equal weights is suitable in applications. To this end, consider the null $H_0^2 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 = \mathbf{0}$ under which the weighting function in (3) becomes flat,[11] leading to high frequency data being aggregated using equal weights, and thus model (1) becomes a panel threshold regression with a covariate-dependent threshold (PTR-CDT):

$$y_{it} = \begin{cases} \mathbf{z}'_{it}\boldsymbol{\alpha}_1 + \mathbf{x}^{*\prime}_{it}\boldsymbol{\beta}_1 + \mu_{1i} + \sigma_1 e_{it}, \text{ if } q_{it} \leq s_{it}(\boldsymbol{\gamma}) \\ \mathbf{z}'_{it}\boldsymbol{\alpha}_2 + \mathbf{x}^{*\prime}_{it}\boldsymbol{\beta}_2 + \mu_{2i} + \sigma_2 e_{it}, \text{ if } q_{it} > s_{it}(\boldsymbol{\gamma}) \end{cases}, \tag{20}$$

where $\mathbf{x}^*_{it}$ is taken as the simple average of the high-frequency data $\mathbf{x}^{(m)}_{it}$. We estimate the

[11] In applications, one can also test whether the Almon lag polynomial parameters are threshold-dependent by considering the null: $H_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$. A test for this null can be easily constructed based on a similar logic as LR_2.

PTR-CDT model and denote the sum of squared errors as S_0^{TH}. Then, a test statistic for the equal weighting scheme can be constructed as:

$$LR_2 = \frac{S_0^{TH} - \widetilde{SSR}_{NT}(\hat{\gamma})}{\widetilde{SSR}_{NT}(\hat{\gamma})/(NT)}. \tag{21}$$

Moreover, to determine whether the threshold is constant, we consider the null hypothesis $H_0^3 : \boldsymbol{\gamma}_1 = \mathbf{0}$. Under this null, the PTMIDAS-CDT model shrinks to the PTMIDAS model with a constant threshold given as:

$$y_{it} = \begin{cases} \mathbf{z}_{it}'\boldsymbol{\alpha}_1 + \mathbf{x}_{it}'(\boldsymbol{\theta}_1)\boldsymbol{\beta}_1 + \mu_{1i} + \sigma_1 e_{it}, \text{ if } q_{it} \leq \gamma \\ \mathbf{z}_{it}'\boldsymbol{\alpha}_2 + \mathbf{x}_{it}'(\boldsymbol{\theta}_2)\boldsymbol{\beta}_2 + \mu_{2i} + \sigma_2 e_{it}, \text{ if } q_{it} > \gamma \end{cases}. \tag{22}$$

Thus, the statistic for testing threshold constancy can be constructed as

$$LR_3 = \frac{S_N(\tilde{\boldsymbol{\gamma}}) - S_N(\hat{\boldsymbol{\gamma}})}{S_N(\hat{\boldsymbol{\gamma}})/(NT)}, \tag{23}$$

where $\tilde{\boldsymbol{\gamma}} = (\tilde{\gamma}, \mathbf{0}')'$, $\tilde{\gamma}$ is estimated from the PTMIDAS model with a constant threshold (22).

In applications, it is also natural to test for unobserved individual-specific threshold effects. The null can be represented as $\mu_{1i} = \mu_{2i}$; in the CRE setting (6), this null for no unobserved threshold effect reduces to $H_0^4 : \boldsymbol{\psi}_1 = \boldsymbol{\psi}_2$. Under the null H_0^4, we estimate the model (1) and denote the sum of squared errors of the above null model as SSR^I. Then, a natural test can be given as follows:

$$LR_4 = \frac{SSR^I - \widetilde{SSR}_{NT}(\hat{\gamma})}{\widetilde{SSR}_{NT}(\hat{\gamma})/(NT)}. \tag{24}$$

We next establish the asymptotic distributions of the above test statistics.

Theorem 4. *Suppose Assumption 1 holds. As $(N,T) \to (\infty, \infty)$, under $H_0^1 : \boldsymbol{\phi}_1 = \boldsymbol{\phi}_2$, we have*

$$LR_1 \xrightarrow{d} \frac{1}{\sigma_u^2} \sup_{\boldsymbol{\gamma} \in \boldsymbol{\Gamma}} \mathcal{Z}'(\boldsymbol{\gamma})\mathbf{R}_1' \left[\mathbf{R}_1 \left(\boldsymbol{J}_{\boldsymbol{g}}'\mathbf{M}^*(\boldsymbol{\gamma})\boldsymbol{J}_{\boldsymbol{g}}\right)^{-1} \mathbf{R}_1'\right]^{-1} \mathbf{R}_1 \mathcal{Z}(\boldsymbol{\gamma}), \tag{25}$$

under $H_0^2 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 = \mathbf{0}$, we have

$$LR_2 \xrightarrow{d} \frac{1}{\sigma_u^2} \mathcal{Z}'\mathbf{R}_2' \left[\mathbf{R}_2 \left(\boldsymbol{J}_{\boldsymbol{g}}'\mathbf{M}^*\boldsymbol{J}_{\boldsymbol{g}}\right)^{-1} \mathbf{R}_2'\right]^{-1} \mathbf{R}_2 \mathcal{Z}, \tag{26}$$

under $H_0^3 : \boldsymbol{\gamma}_1 = \mathbf{0}$, we have

$$LR_3 \xrightarrow{d} \frac{1}{\sigma_u^2} \left[\left(\min_{\boldsymbol{\omega} \in \mathbb{R}_c^{d+1}} G(\boldsymbol{\omega}) - 2R(\boldsymbol{\omega}) \right) - \left(\min_{\boldsymbol{\omega} \in \mathbb{R}^{d+1}} G(\boldsymbol{\omega}) - 2R(\boldsymbol{\omega}) \right) \right], \tag{27}$$

and under $H_0^4 : \boldsymbol{\psi}_1 = \boldsymbol{\psi}_2$, *we have*

$$LR_4 \xrightarrow{d} \frac{1}{\sigma_u^2} \mathcal{Z}' \mathbf{R}_I' \left[\mathbf{R}_I \left(\boldsymbol{J}_{\boldsymbol{g}}' \mathbf{M}^* \boldsymbol{J}_{\boldsymbol{g}} \right)^{-1} \mathbf{R}_I' \right]^{-1} \mathbf{R}_I \mathcal{Z}, \tag{28}$$

where $\mathbf{R}_1$, $\mathbf{R}_2$ *and* $\mathbf{R}_I$ *are matrices such that* $\mathbf{R}_1 \boldsymbol{\phi} = \phi_1 - \phi_2 = \mathbf{0}$, $\mathbf{R}_2 \boldsymbol{\phi} = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')' = \mathbf{0}$ *and* $\mathbf{R}_I \boldsymbol{\phi} = \boldsymbol{\psi}_1 - \boldsymbol{\psi}_2 = \mathbf{0}$, *respectively.* $\mathbb{R}_c^{d+1} = \{\boldsymbol{\omega} \in \mathbb{R}^{d+1} : \boldsymbol{\omega}_1 = \mathbf{0}\}$, $\mathcal{Z}(\gamma)$ *is a Gaussian process with variance* $\left(\boldsymbol{J}_{\boldsymbol{g}}' \mathbf{M}^* (\gamma) \boldsymbol{J}_{\boldsymbol{g}} \right)^{-1} \boldsymbol{J}_{\boldsymbol{g}}' \boldsymbol{\Omega}^* (\gamma) \boldsymbol{J}_{\boldsymbol{g}} \left(\boldsymbol{J}_{\boldsymbol{g}}' \mathbf{M}^* (\gamma) \boldsymbol{J}_{\boldsymbol{g}} \right)^{-1}$, *and* $G(\boldsymbol{\omega})$ *and* $R(\boldsymbol{\omega})$ *are defined as in Theorem 1.* $\mathcal{Z}$ *is defined as in Theorem 2.*

Proof of Theorem 4. See the Appendix. □

The limiting distributions of LR_j are nonstandard for $j = 1, 2, 3, 4$. To implement the test statistics, we propose a wild bootstrap algorithm following the classical threshold literature such as Hansen (1999) and Seo and Shin (2016). The bootstrap procedure goes as follows.

Algorithm 3. Testing for threshold effect and threshold constancy

Step 1. Generate i.i.d draws v_{it}^* from the $N(0,1)$ distribution for $i = 1, \ldots, N$ and $t = 1, \ldots, T$, and set $u_{j,it}^* = \hat{u}_{j,it} v_{it}^*$, $j = 1, 2, 3, 4$, where $\hat{u}_{1,it}$ are the residuals from the standard MIDAS model (18), $\hat{u}_{2,it}$ are the residuals from the PTR-CDT model (20), $\hat{u}_{3,it}$ are the residuals from the PTMIDAS model with a constant threshold (22), and $\hat{u}_{4,it}$ are the residuals from the PTMIDAS-CDT model with no unobserved threshold effect.

Step 2. For $i = 1, 2, \ldots, N$, $t = 1, 2, \ldots, T$, set

$$\begin{aligned}
y_{1,it}^* &= \mathbf{z}_{it}' \hat{\boldsymbol{\alpha}}_{1,1} + \mathbf{x}_{it}'(\hat{\boldsymbol{\theta}}_{1,1}) \hat{\boldsymbol{\beta}}_{1,1} + \mathbf{h}_i' \hat{\boldsymbol{\psi}}_{1,1} + u_{1,it}^*, \\
y_{2,it}^* &= \begin{cases} \mathbf{z}_{it}' \hat{\boldsymbol{\alpha}}_{2,1} + \mathbf{x}_{it}^{*\prime} \hat{\boldsymbol{\beta}}_{2,1} + \mathbf{h}_i' \hat{\boldsymbol{\psi}}_{2,1} + u_{2,it}^*, & \text{if } q_{it} \leq s_{it}(\hat{\boldsymbol{\gamma}}_2) \\ \mathbf{z}_{it}' \hat{\boldsymbol{\alpha}}_{2,2} + \mathbf{x}_{it}^{*\prime} \hat{\boldsymbol{\beta}}_{2,2} + \mathbf{h}_i' \hat{\boldsymbol{\psi}}_{2,2} + u_{2,it}^*, & \text{if } q_{it} > s_{it}(\hat{\boldsymbol{\gamma}}_2) \end{cases}, \\
y_{3,it}^* &= \begin{cases} \mathbf{z}_{it}' \hat{\boldsymbol{\alpha}}_{3,1} + \mathbf{x}_{it}'(\hat{\boldsymbol{\theta}}_{3,1}) \hat{\boldsymbol{\beta}}_{3,1} + \mathbf{h}_i' \hat{\boldsymbol{\psi}}_{3,1} + u_{3,it}^*, & \text{if } q_{it} \leq \tilde{\gamma} \\ \mathbf{z}_{it}' \hat{\boldsymbol{\alpha}}_{3,2} + \mathbf{x}_{it}'(\hat{\boldsymbol{\theta}}_{3,2}) \hat{\boldsymbol{\beta}}_{3,2} + \mathbf{h}_i' \hat{\boldsymbol{\psi}}_{3,2} + u_{3,it}^*, & \text{if } q_{it} > \tilde{\gamma} \end{cases}, \\
y_{4,it}^* &= \begin{cases} \mathbf{z}_{it}' \hat{\boldsymbol{\alpha}}_{4,1} + \mathbf{x}_{it}'(\hat{\boldsymbol{\theta}}_{4,1}) \hat{\boldsymbol{\beta}}_{4,1} + \mathbf{h}_i' \hat{\boldsymbol{\psi}}_4 + u_{4,it}^*, & \text{if } q_{it} \leq s_{it}(\hat{\boldsymbol{\gamma}}_4) \\ \mathbf{z}_{it}' \hat{\boldsymbol{\alpha}}_{4,2} + \mathbf{x}_{it}'(\hat{\boldsymbol{\theta}}_{4,2}) \hat{\boldsymbol{\beta}}_{4,2} + \mathbf{h}_i' \hat{\boldsymbol{\psi}}_4 + u_{4,it}^*, & \text{if } q_{it} > s_{it}(\hat{\boldsymbol{\gamma}}_4) \end{cases},
\end{aligned}$$

where $(\hat{\boldsymbol{\alpha}}_{1,1}, \hat{\boldsymbol{\beta}}_{1,1}, \hat{\boldsymbol{\theta}}_{1,1}, \hat{\boldsymbol{\psi}}_{1,1})$, $(\hat{\boldsymbol{\alpha}}_{2,1}, \hat{\boldsymbol{\beta}}_{2,1}, \hat{\boldsymbol{\psi}}_{2,1}, \hat{\boldsymbol{\alpha}}_{2,2}, \hat{\boldsymbol{\beta}}_{2,2}, \hat{\boldsymbol{\psi}}_{2,2}, \hat{\boldsymbol{\gamma}}_2)$, $(\hat{\boldsymbol{\alpha}}_{3,1}, \hat{\boldsymbol{\beta}}_{3,1}, \hat{\boldsymbol{\theta}}_{3,1}, \hat{\boldsymbol{\psi}}_{3,1}, \hat{\boldsymbol{\alpha}}_{3,2}, \hat{\boldsymbol{\beta}}_{3,2}, \hat{\boldsymbol{\theta}}_{3,2}, \hat{\boldsymbol{\psi}}_{3,2}, \hat{\gamma})$, and $(\hat{\boldsymbol{\alpha}}_{4,1}, \hat{\boldsymbol{\beta}}_{4,1}, \hat{\boldsymbol{\theta}}_{4,1}, \hat{\boldsymbol{\alpha}}_{4,2}, \hat{\boldsymbol{\beta}}_{4,2}, \hat{\boldsymbol{\theta}}_{4,2}, \hat{\boldsymbol{\psi}}_4, \hat{\boldsymbol{\gamma}}_4)$ are the estimates of the MIDAS model (18), the PTR-CDT model (20), the PTMIDAS model with a constant threshold (22), and the PTMIDAS-CDT model with no unobserved threshold effect based on the original sample $\{y_{it}, \mathbf{z}_{it}, \mathbf{x}_{it}^{(m)}, q_{it}, \mathbf{s}_{it}\}$, respectively.

Step 3. Use the bootstrap sample $\{y^*_{j,it}, \mathbf{z}_{it}, \mathbf{x}^{(m)}_{it}, q_{it}, \mathbf{s}_{it}\}$ for $j = 1,2,3,4$ to estimate the MIDAS model (18), the PTR-CDT model (20), the PTMIDAS model with a constant threshold (22), and the PTMIDAS-CDT model with no unobserved threshold effect, respectively. Compute the test statistics LR_j for $j = 1,2,3,4$.

Step 4. Repeat Steps 1-3 B times to obtain samples $LR^*_j(1), LR^*_j(2), ..., LR^*_j(B)$ of simulated LR_j statistics for $j = 1,2,3,4$.

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

3 Monte Carlo Simulations

In this section, we investigate the finite sample properties of the proposed estimator and test statistics proposed in Section 2. To do so, we consider the following data generating process (DGP):

$$y_{it} = \begin{cases} \beta_{11} q_{it} + \beta_{12} z_{it} + \beta_{13} x_{it}(\boldsymbol{\theta}_1) + \mu_{1i} + \sigma_1 e_{it}, \text{ if } q_{it} \le s_{it}(\boldsymbol{\gamma}) \\ \beta_{21} q_{it} + \beta_{22} z_{it} + \beta_{23} x_{it}(\boldsymbol{\theta}_2) + \mu_{2i} + \sigma_2 e_{it}, \text{ if } q_{it} > s_{it}(\boldsymbol{\gamma}) \end{cases}, \tag{29}$$

where $\mu_{\ell i} = \boldsymbol{\psi}'_\ell \mathbf{h}_i + a_{\ell i}$ for $\ell = 1,2, \mathbf{h}_i = (\bar{z}_i, \bar{q}_i, \bar{x}_i)', \bar{z}_i = \frac{1}{T}\sum_{t=1}^{T} z_{it}, \bar{q}_i = \frac{1}{T}\sum_{t=1}^{T} q_{it}, \bar{x}_i = \frac{1}{T}\sum_{t=1}^{T} x^{(m)}_{it}$. $x_{it}(\boldsymbol{\theta}_\ell) = \sum_{j=1}^{m} w_j(\boldsymbol{\theta}_\ell) L^{j/m} x^{(m)}_{it}$, $s_{it}(\boldsymbol{\gamma}) = \gamma_0 + \gamma_1 s_{it}$, $x^{(m)}_{it} = 0.7 x^{(m)}_{i,(t-1)/m} + u_{x,it}$, $z_{it} = 0.7 z_{i,t-1} + u_{z,it}$, $q_{it} \sim i.i.d.N(0.2, 1)$, and $s_{it} \sim i.i.d.N(0,1)$. The innovation processes e_{it}, $u_{x,it}$, $u_{z,it}$, $a_{\ell i}$ are independent of each other. e_{it} follows GARCH(1,1) process: $e_{it} = \sqrt{h_{it}}\varepsilon_{it}, h_{it} = 5 \times 10^{-4} + 0.9 h_{i,t-1} + 0.05 e^2_{i,t-1}, \varepsilon_{it} \sim i.i.d.N(0,1)$. $u_{x,it}$, $a_{\ell i}$ and $u_{z,it}$ follow $i.i.d.N(0,1)$. Similarly with Andreou et al. (2010), $x^{(m)}_{it}$ is sampled m times between $t-1$ and t such that the high frequency sample size is mT, while $z_{it}, \mu_i, q_{it}, s_{it}, e_{it}, u_{x,it}$ and $u_{z,it}$ are sampled at a low frequency with a sample size T. Furthermore, the high-frequency data $x^{(m)}_{it}$ are projected on to the low frequency data $x_{it}(\boldsymbol{\theta}_\ell)$, using the two-parameter exponential Almon lag polynomial given by

$$w_j(\boldsymbol{\theta}_\ell) = w_j(\theta_{\ell 1}, \theta_{\ell 2}) = \frac{\exp(\theta_{\ell 1} j + \theta_{\ell 2} j^2)}{\sum_{j=1}^{m} \exp(\theta_{\ell 1} j + \theta_{\ell 2} j^2)}. \tag{30}$$

in which we set $m = 3$ to match the case with a quarterly dependent variable and monthly independent variable.

We first assess the finite sample properties of the proposed estimator. In this simulation, we set $\boldsymbol{\beta}_1 = (\beta_{11}, \beta_{12}, \beta_{13}) = (-1,-1,-1)$, $\boldsymbol{\beta}_2 = (\beta_{21}, \beta_{22}, \beta_{23}) = (1,1,1)$, $\boldsymbol{\psi}_1 =$

$(-0.2, -0.2, -0.2, -0.2, -0.2, -0.2)$, $\boldsymbol{\psi}_2 = (0.2, 0.2, 0.2, 0.2, 0.2, 0.2)$, $\sigma_1 = 0.3$, $\sigma_2 = 0.4$, $\boldsymbol{\theta}_1 = (\theta_{11}, \theta_{12}) = (-0.5, 0.04)$, $\boldsymbol{\theta}_2 = (\theta_{21}, \theta_{22}) = (0, 0)$ and $(\gamma_0, \gamma_1) = (0.2, 0.5)$, and run experiments on a range of sample sizes $((T, N) = \{(10, 10), (20, 20), (50, 50)\})$. Each experiment is replicated 1,000 times to calculate the summary statistics (i.e., mean and standard deviation) for the parameter estimates. The simulation results are reported in Table 1. In Panel B of Table 1, we also report the estimation results based on the grid search method as a robustness check.[12] For all parameters, the mean of each parameter is fairly close to its true value for all combinations of T and N, and the standard deviations decrease to zero as either T or N increases; moreover, when the sample size changes from $N \times T = 10 \times 10$ to 20×20, the standard deviations of the threshold parameters decrease by four times as expected by the convergence rate $(NT)^{1-2\alpha}$ implied by Theorem 1,[13] while the standard deviations of the slope parameters almost decrease by half as expected by Theorem 2.[14] In addition, the estimation results based on the grid search are roughly similar with those based on the MCMC algorithm.[15] These results indicate that the estimation procedure works well in finite samples, and the convergence rates are consistent with Theorems 1-2.

We next conduct simulations to evaluate the size and power properties of the proposed test statistics for threshold effect, the equal weighting scheme, the threshold constancy and the unobserved individual-specific threshold effects. In doing so, we set $\{\boldsymbol{\beta}_1 = (-1, -1, -1), \boldsymbol{\psi}_1 = (-0.2, -0.2, -0.2, -0.2, -0.2, -0.2), \boldsymbol{\theta}_1 = (-0.5, 0.04), \sigma_1 = 0.3\}$. In examining the finite sample performance of the test for linearity, the rejection frequencies under GDP in (14) with $\{\boldsymbol{\beta}_2 = (-1, -1, -1), \boldsymbol{\psi}_2 = (-0.2, -0.2, -0.2, -0.2, -0.2, -0.2), \boldsymbol{\theta}_2 = (-0.5, 0.04), \sigma_2 = 0.3\}$ and $\{\boldsymbol{\beta}_2 = (1, 1, 1), \boldsymbol{\psi}_2 = (0.2, 0.2, 0.2, 0.2, 0.2, 0.2), \boldsymbol{\theta}_2 = (0, 0), \sigma_2 = 0.4\}$ are the size and power of the proposed test for threshold effect, respectively. Meanwhile, in evaluating the performance of the test for the equal weighting scheme, the rejection frequencies under the flat weighting setting with $\boldsymbol{\theta}_\ell = (\theta_{\ell 1}, \theta_{\ell 2}) = (0, 0), \ell = 1, 2$ and $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = (-0.5, 0.04, -1, -1)$ are the size and power of the proposed test for the equal weighting scheme, respectively. In assessing the finite sample performance of the test for threshold constancy, the rejection frequencies under DGP in (29) with $(\gamma_0, \gamma_1) = (0.2, 0)$ and $(\gamma_0, \gamma_1) = (0.2, 0.5)$ are the size and power of the proposed test for threshold constancy, respectively. Finally, in assessing the finite sample

[12] We thank an anonymous referee for raising this point with us.

[13] In our simulations, we set $\alpha = 0$.

[14] As pointed out by Wooldridge (2019), $\mathbf{h}_i$ is collinear with the explanatory variables by construction, which may lead to efficiency losses for estimates of the coefficient of the time averages of the explanatory variables.

[15] To save simulation time, we estimate the model through a crude grid search by spanning over 20×20 equispaced points on $\boldsymbol{\Gamma} = \Gamma_0 \times \Gamma_1$, which may lead to inaccurate estimates.

performance of the test for the unobserved individual-specific threshold effects, the rejection frequencies under DGP in (29) with $(\boldsymbol{\psi}_2, \sigma_2) = (-0.2, -0.2, -0.2, -0.2, -0.2, -0.2, 0.3)$ and $(\boldsymbol{\psi}_2, \sigma_2) = (0.2, 0.2, 0.2, 0.2, 0.2, 0.2, 0.4)$ are the size and power, respectively. The simulation results are reported in Table 2, in which the size and power for each experiment were constructed using 100 replications and the number of bootstrap replications was set as 100 to save the simulation time. The significance level is set at 5%. The simulation results indicate that the empirical sizes of the test statistics are close to 0.05, and the powers enhance as the sample size increases, indicating that the proposed test statistics have good size and power properties.

Table 1: Estimates of the parameters.

T	N		γ_0	γ_1	β_{11}	β_{12}	β_{13}	β_{21}	β_{22}	β_{23}	ψ_{11}	ψ_{12}	ψ_{13}	ψ_{14}	ψ_{15}	ψ_{16}	ψ_{21}	ψ_{22}	ψ_{23}	ψ_{24}	ψ_{25}	ψ_{26}	θ_{11}	θ_{12}	θ_{21}	θ_{22}
Panel A: the MCMC-based algorithm																										
10	10	Mean	0.200	0.501	-0.999	-1.000	-0.998	0.997	0.996	0.998	-0.197	-0.203	-0.168	-0.275	-0.187	-0.203	0.175	0.198	0.251	0.093	0.222	0.215	-0.556	0.052	-0.062	0.013
		Std	0.041	0.064	0.042	0.078	0.042	0.045	0.081	0.043	1.010	0.376	1.088	1.096	1.060	0.330	1.038	0.389	1.146	1.126	1.086	0.358	1.101	0.286	1.192	0.298
20	20	Mean	0.200	0.501	-1.000	-1.001	-0.999	1.000	0.999	0.998	-0.172	-0.198	-0.164	-0.204	-0.228	-0.207	0.206	0.211	0.217	0.210	0.185	0.202	-0.490	0.037	-0.027	0.006
		Std	0.011	0.015	0.020	0.038	0.020	0.020	0.040	0.020	0.647	0.213	0.741	0.767	0.773	0.181	0.647	0.213	0.744	0.767	0.774	0.184	0.538	0.140	0.546	0.136
50	50	Mean	0.200	0.500	-1.000	-1.000	-1.000	1.000	1.000	1.000	-0.177	-0.190	-0.197	-0.192	-0.223	-0.209	0.223	0.211	0.198	0.213	0.177	0.191	-0.507	0.042	-0.005	0.001
		Std	0.002	0.002	0.008	0.017	0.008	0.008	0.017	0.008	0.538	0.171	0.683	0.689	0.677	0.133	0.546	0.170	0.674	0.686	0.680	0.135	0.213	0.055	0.205	0.051
Panel B: the grid search method																										
10	10	Mean	0.179	0.492	-1.000	-1.001	-1.001	1.002	1.000	1.004	-0.212	-0.195	-0.112	-0.262	-0.178	-0.211	0.215	0.208	0.299	0.148	0.210	0.195	-0.439	0.022	0.080	-0.020
		Std	0.041	0.052	0.044	0.079	0.044	0.048	0.088	0.045	1.003	0.394	1.131	1.207	1.063	0.342	0.984	0.406	1.194	1.299	1.091	0.353	1.163	0.306	1.148	0.287
20	20	Mean	0.200	0.501	-0.998	-0.990	-0.997	0.998	0.999	0.996	-0.211	-0.207	-0.181	-0.255	-0.169	-0.184	0.195	0.195	0.228	0.127	0.234	0.205	-0.493	0.038	0.020	-0.005
		Std	0.017	0.025	0.023	0.046	0.022	0.021	0.046	0.022	0.667	0.220	0.772	0.799	0.768	0.193	0.684	0.226	0.784	0.784	0.761	0.205	0.642	0.168	0.593	0.148
50	50	Mean	0.199	0.499	-0.992	-0.991	-0.992	0.991	1.002	0.991	-0.227	-0.196	-0.190	-0.209	-0.193	-0.185	0.190	0.204	0.213	0.184	0.211	0.197	-0.487	0.036	0.011	-0.002
		Std	0.013	0.015	0.013	0.024	0.012	0.013	0.021	0.013	0.570	0.174	0.633	0.674	0.658	0.138	0.581	0.174	0.630	0.672	0.649	0.143	0.286	0.074	0.290	0.072

Table 2: Size and power of the proposed test statistics.

T	N	LR_1		LR_2		LR_3		LR_4	
		Size	Power	Size	Power	Size	Power	Size	Power
10	10	0.050	1.000	0.040	1.000	0.040	1.000	0.060	0.908
20	20	0.050	1.000	0.060	1.000	0.040	1.000	0.060	1.000
50	50	0.050	1.000	0.040	1.000	0.050	1.000	0.050	1.000

4 Empirical Application

In this section, we illustrate the proposed model by investigating whether there is a covariate-dependent threshold effect in the relationship between climate change and economic growth, which has attracted much attention among academics and policy makers in recent years. A number of studies are devoted to investigate the impact of weather conditions on economic growth, while they differ in whether there is a nonlinear nexus. Using linear models, Dell et al. (2012) document that a 1°C rise in average annual temperature reduces economic growth by about 1.3% in poor countries, while Yang et al. (2023) argue that calculating the annual temperature using simple average as in Dell et al. (2012) may result in misleading results. Meanwhile, Burke et al. (2015) argue that there is a nonlinear relationship between temperature and economic growth, and empirically find a 13°C threshold (optimal average annual temperature) above which temperature reduces growth and below which temperature is positively associated with growth, while Lee et al. (2020) find a temperature threshold not 13°C but 14.2°C using polynomial regression models, which might be misleading in identifying turning points as pointed by Lind and Mehlum (2010); moreover, as explained by the panel MIDAS literature (e.g., Khalaf et al., 2021; Yang et al., 2023), an inappropriate choice of weighting scheme could result in biased estimates and unreliable inferences, and hence, considering a panel threshold MIDAS model with a covariate-dependent threshold may be useful for accurately estimating a temperature threshold.

In this paper, we therefore apply the proposed panel threshold MIDAS model to estimate temperature thresholds by extending the empirical model of Burke et al. (2015) and Lee et al. (2020) to a panel threshold MIDAS framework. Our empirical model is given by

$$y_{it} = \begin{cases} \alpha_{11} Temp_{it}(\boldsymbol{\theta}_{11}) + \alpha_{12} Temp_{i,t-1}(\boldsymbol{\theta}_{12}) + \beta_1 Prec_{it} \\ \qquad + \lambda_t + region_{g(i)t} + \mu_{1i} + \sigma_1 e_{it}, \text{ if } Temp_{it} \leq s_{it}(\boldsymbol{\gamma}) \\ \alpha_{21} Temp_{it}(\boldsymbol{\theta}_{21}) + \alpha_{22} Temp_{i,t-1}(\boldsymbol{\theta}_{22}) + \beta_2 Prec_{it} \\ \qquad + \lambda_t + region_{g(i)t} + \mu_{2i} + \sigma_2 e_{it}, \text{ if } Temp_{it} > s_{it}(\boldsymbol{\gamma}) \end{cases} \tag{31}$$

in which y_{it} is the growth rate of per-capita output for i at year t, $Temp_{it}(\boldsymbol{\theta}_{11})$ and $Temp_{i,t-1}(\boldsymbol{\theta}_{12})$ are the weighted averages of the quarterly temperature for country i in year t and $t-1$ as in Yang et al. (2023). $Prec_{it}$ is the annual total precipitation for country i in year t. μ_i, λ_t, and $region_{g(i)t}$ denote country, year, and region-time fixed effects. Here, we control the region-time fixed effect $region_{g(i)t}$ using the interactions the region dummies and the time trend t and t^2 following Burke et al. (2015) and Lee et al. (2020). To investigate whether temperature threshold differs for poor and rich countries, we model the threshold as

$$s_{it}(\boldsymbol{\gamma}) = \gamma_0 + \gamma_1 Poor_i, \tag{32}$$

in which $Poor_i$ is equal to one for poor countries, and zero for rich countries, $\boldsymbol{\gamma} = (\gamma_0, \gamma_1)'$ is the vector of threshold parameters. Following Dell et al. (2012) and Yang et al. (2023), poor countries are defined as having below-median PPP-adjusted per capita GDP in the first year the country enters the dataset. Considering such a threshold setting is important as it can help us to answer whether the impacts of future warming could attenuate as countries become richer, i.e., whether countries adapted effectively to temperature as they became wealthier. The dataset we considered is the same as in Yang et al. (2023).[16] To match the proposed model, the analysis is performed for the balanced panel data from 1961 to 2003. The dataset contains 78 countries listed in Table 3. These countries cover most regions in the world and include advanced, emerging and developing economies.

The empirical results are reported in Table 4, in which 95% confidence intervals (95%CI) for parameters and p-values of the tests are computed by a bootstrap approach with 1000 bootstrap replications. For the sake of comparison, we also report the results based on the linear model with equal weights, linear MIDAS model, and the panel threshold MIDAS model with equal weights (which is a special case of the proposed model with $\boldsymbol{\gamma}_1 = 0$). As can be seen from Panel A of Table 4, in the linear setting, the null of the equal weight scheme can not be rejected (p-value=33%). According to Panel B, the null of the equal weight scheme can not be rejected at the 10% significant level in the constant threshold model (p-value=10.7%); however, the null of the equal weight scheme is rejected with the p-value 6.2% when a covariate-dependent threshold is considered. Furthermore, based on the constant threshold model, the null of linearity can not be rejected, while the proposed covariate-dependent threshold model can reject the null of linearity, indicating that there exists a nonconstant threshold in the relationship between climate change and economic growth. Moreover, the proposed model provides the evidence supporting a temperature threshold around 16°C for poor countries but 25°C for rich countries, and the difference between poor and rich countries is statistically significant at the 5% level (according to the test for threshold constancy), providing evidence of adaptation, i.e., the impact of future warming on growth would atten-

[16] We are very grateful to Yang et al. (2023) for sharing their data with us.

Table 3: List of the 78 countries in the sample.

Region	Country		
MENA	Algeria	Egypt	Israel
	Morocco	Oman	
LAC	Argentina	Belize	Brazil
	Chile	Colombia	Costa Rica
	Dominican Republic	Ecuador	El Salvador
	Guatemala	Guyana	Haiti
	Honduras	Mexico	Nicaragua
	Paraguay	Peru	Panama
	Uruguay		
WEOFF	Australia	Austria	Belgium
	Denmark	Ireland	Finland
	France	Greece	Iceland
	Italy	Luxembourg	Netherlands
	Norway	New Zealand	Portugal
	Spain	Sweden	Switzerland
SSAF	Botswana	Benin	Burundi
	Chad	Cameroon	Central African Republic
	Gabon	Ghana	Cote d'Ivoire
	Kenya	Liberia	Lesotho
	Madagascar	Malawi	Mauritania
	Niger	Nigeria	Rwanda
	Senegal	Sierra Leone	Sudan
	Togo	Burkina Faso	Zambia
	Zimbabwe		
SEAS	Sri Lanka	China	Indonesia
	India	Japan	Malaysia
	Nepal	Papua New Guinea	Philippines
	Thailand		
EECA	Hungary		

Notes: MENA denotes Middle East and North Africa. WEOFF denotes Western Europe and Offshoots. LAC denotes Latin American Countries. SSAF denotes Sub-Saharan Africa. SEAS denotes South-East Asia. EECA denotes Eastern Europe and Central Asia.

uate when countries become richer. This result is clearly different from either Burke et al. (2015) who found a universal temperature threshold around 13°C, or Lee et al. (2020) who found a universal temperature threshold around 14.2°C, and hence, "offering little evidence of adaptation". Moreover, the coefficient of $Temp_{it}$ is significantly negatively different from zero at the 5% level when the temperature is above the threshold, while it is positive but insignificant when the temperature is below the threshold. Interestingly, when we consider a covariate-dependent threshold, the estimated negative temperature effect on growth appears to be relatively stronger than that based on a constant threshold setting, which can be explained as overlooking the threshold nonconstancy would lead to misclassification of observations, and thus, the estimates of slope parameters may be biased; on the other hand, when the temperature is below the covariate-dependent threshold, the impact of temperature on growth is statistically insignificant, which can not be captured by the constant threshold setting. In addition, based on the linear panel MIDAS model, Yang et al. (2023) find that 1°C increase in temperature leads to a statistically significant reduction of growth rate, and “the negative effect of temperature persists only in the short run” according to the coefficient of temperature in year t. “As more lags are included, the growth effect (measured by as the summation of temperature coefficients) becomes statistically insignificant”. Our results confirm that the growth effect is statistically insignificant, as the summation of temperature coefficients is not statistically significantly different from zero as can be seen from Table 4. Moreover, unlike Yang et al. (2023) who support a universal negative effect in the short run, our results show that the short-run effect of temperature on growth may be positive when the temperature is below the threshold, and the negative short-run effect exists only when the temperature is above the threshold. Furthermore, as can be seen from Panel C, the estimates based on the grid search method are fairly similar with these based on the MCMC algorithm, which confirms the robustness of the results based on the suggested algorithm.[17] In the online appendix, we briefly compare our empirical results with these based on a polynomial model specification, and the results support that our main findings can not be captured by a simple polynomial model.[18] In summary, our results support that there is a covariate-dependent temperature threshold (optimal average annual temperature) above which temperature is negatively associated with growth. These results demonstrate the usefulness of the proposed panel threshold mixed data sampling model with a covariate-dependent threshold.

5 Conclusion

The relationship between economic variables is generally nonlinear, and the data are usually sampled at different frequencies. Panel threshold models are popular in applications, while the classical panel threshold models maintain two restrictive assumptions that (i) the data are sampled at the same frequency, and (ii) the threshold value is constant. This paper proposes a model called panel threshold mixed data sampling model with a covariate-dependent threshold and unobserved individual-specific threshold effects (PTMIDAS-CDT), which overcomes the two mentioned limitations by synergizing panel threshold models and MIDAS models. Based on the Chamberlain-Mundlak correlated random effects (CRE) device and Markov chain Monte Carlo (MCMC) technique, we develop a new estimation procedure to efficiently estimate the proposed model, and suggest test statistics for model specification. We derive the asymptotic properties of the proposed estimator in the small-threshold-effect framework,

[17] We thank the anonymous referee for raising this point to us. We do not report the testing results based on the grid search method to save time.

[18] We thank an anonymous referee for suggesting this comparison to us.

Table 4: Empirical results based on quarterly data.

Panel A: linear models	Linear model with equal weights		Linear MIDAS	
	Estimate	95%CI	Estimate	95%CI
$Temp_{it}$	-0.284	$[-0.675, 0.105]$	-0.429	$[-0.703, -0.155]$
$Temp_{i,t-1}$	0.238	$[-0.154, 0.629]$	0.407	$[0.119, 0.695]$
$Prec_{it}$	-0.041	$[-0.072, -0.009]$	-0.042	$[-0.074, -0.011]$
Testing	Statistic	p-value	Statistic	p-value
Test for the equal weight scheme			7.343	0.330
Sum of temperature coefficients	-0.047	$[-0.090, -0.003]$	-0.022	$[-0.067, 0.022]$
Panel B: threshold models	Constant threshold		Covariate-dependent threshold	
	Estimate	95%CI	Estimate	95%CI
$Temp_{it}\{Temp_{it} \leq s_{it}(\hat{\gamma})\}$	0.448	$[0.105, 0.790]$	0.147	$[-0.059, 0.353]$
$Temp_{i,t-1}\{Temp_{it} \leq s_{it}(\hat{\gamma})\}$	-0.398	$[-0.721, -0.074]$	-0.141	$[-0.337, 0.055]$
$Temp_{it}\{Temp_{it} > s_{it}(\hat{\gamma})\}$	-1.053	$[-1.483, -0.622]$	-1.393	$[-1.849, -0.936]$
$Temp_{i,t-1}\{Temp_{it} > s_{it}(\hat{\gamma})\}$	1.110	$[0.660, 1.559]$	1.470	$[0.972, 1.967]$
β_1	-0.072	$[-0.199, 0.054]$	0.012	$[-0.036, 0.059]$
β_2	-0.036	$[-0.069, -0.003]$	-0.122	$[-0.163, -0.081]$
$\hat{\gamma}_0$	15.992	$[11.229, 20.736]$	25.365	$[20.836, 30.307]$
$\hat{\gamma}_1$			-9.055	$[-19.094, -0.069]$
Testing	Statistic	p-value	Statistic	p-value
Test for threshold effect	33.865	0.360	63.714	0.040
Test for the equal weight scheme	13.130	0.107	14.853	0.062
Test for threshold constancy			25.527	0.010
Test for unobserved individual-specific threshold effects	3.411	0.690	1.573	0.680
Sum of temperature coefficients in $\{Temp_{it} \leq s_{it}(\hat{\gamma})\}$	0.049	$[-0.044, 0.143]$	0.006	$[-0.066, 0.078]$
Sum of temperature coefficients in $\{Temp_{it} > s_{it}(\hat{\gamma})\}$	0.057	$[-0.016, 0.131]$	0.077	$[-0.010, 0.164]$
Panel C: robust check by grid search	Constant threshold		Covariate-dependent threshold	
	Estimate	95%CI	Estimate	95%CI
$Temp_{it}\{Temp_{it} \leq s_{it}(\hat{\gamma})\}$	0.469	$[0.102, 0.836]$	0.033	$[-0.086, 0.152]$
$Temp_{i,t-1}\{Temp_{it} \leq s_{it}(\hat{\gamma})\}$	-0.418	$[-0.770, -0.067]$	-0.046	$[-0.121, 0.029]$
$Temp_{it}\{Temp_{it} > s_{it}(\hat{\gamma})\}$	-1.039	$[-1.465, -0.612]$	-1.413	$[-1.869, -0.957]$
$Temp_{i,t-1}\{Temp_{it} > s_{it}(\hat{\gamma})\}$	1.095	$[0.650, 1.539]$	1.489	$[0.997, 1.980]$
β_1	-0.071	$[-0.198, 0.055]$	0.011	$[-0.036, 0.058]$
β_2	-0.036	$[-0.069, -0.003]$	-0.119	$[-0.160, -0.077]$
$\hat{\gamma}_0$	15.996	$[11.250, 20.748]$	25.582	$[20.872, 30.304]$
$\hat{\gamma}_1$			-9.586	$[-19.031, -0.128]$
Sum of temperature coefficients in $\{Temp_{it} \leq s_{it}(\hat{\gamma})\}$	0.050	$[-0.045, 0.145]$	-0.013	$[-0.082, 0.057]$
Sum of temperature coefficients in $\{Temp_{it} > s_{it}(\hat{\gamma})\}$	0.056	$[-0.018, 0.130]$	0.076	$[-0.011, 0.162]$

and establish the limiting distributions of the suggested test statistics under the large-N and large-T setup. Monte Carlo simulations are conducted to confirm the finite sample properties of the proposed estimator and test statistics. We illustrate the new model by investigating the relationship between climate change and economic growth. Our results support the presence of a covariate-dependent threshold, and indicate that temperature thresholds differ for poor and rich countries. Both simulation and empirical results demonstrate the usefulness of the proposed model.

Acknowledgements We would like to thank the Editor and two anonymous referees for very constructive comments and suggestions. Any remaining errors are solely our responsibility. The first author (Lixiong Yang) acknowledges the financial support from the National Natural Science Foundation of China (Grant No. 72273059).

Data availability statement The codes and data are available on request from the authors.

References

Andreou, E., Ghysels, E., Kourtellos, A. (2010). Regression models with mixed sampling frequencies. *Journal of Econometrics*, 158(2), 246-261.

Burke, M., Hsiang, S. M., Miguel, E. (2015). Global non-linear effect of temperature on economic production. *Nature*, 527(7577), 235-239.

Chen, H. (2015). Robust estimation and inference for threshold models with integrated regressors. *Econometric Theory*, 31(4), 778-810.

Davies, R.B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, 74(1), 33-43.

Dell, M., Jones, B. F., Olken, B. A. (2012). Temperature shocks and economic growth: Evidence from the last half century. *American Economic Journal: Macroeconomics*, 4(3), 66-95.

Galvao, A. B. (2013). Changes in predictive ability with mixed frequency data. *International Journal of Forecasting*, 29(3), 395-410.

Gelman, A., Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7, 457-511.

Ghysels, J., Miller, J., (2015). Testing for Cointegration with Temporally Aggregated and Mixed-frequency Time Series. *Journal of Time Series Analysis*, 36(6), 797-816.

Ghysels, E., Santa-Clara, P., Valkanov, R., (2004). The MIDAS Touch: Mixed Data Sampling Regression Models. *CIRANO Working Paper*, 2004s-20.

Ghysels, E., Santa-Clara, P, Valkanov, R. (2006). Predicting volatility: Getting the Most out of Return Data Sampled at Different Frequencies. *Journal of Econometrics*, 131, 59-95.

Ghysels, E., Wright, J. (2009). Forecasting professional forecasters. *Journal of Business and Economic Statistics*, 27(4), 504-516

Ghysels, E., Sinko, A., Valkanov, R. (2007). Midas regressions: further results and new directions. *Econometric Reviews*, 26(1), 53-90.

Hidalgo, J., Lee, J., Seo, M. H. (2019). Robust inference for threshold regression models. *Journal of Econometrics*, 210(2), 291-309.

Hansen, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica*, 64, 413-430.

Hansen, B. E. (1999). Threshold effects in non-dynamic panels: Estimation, testing, and inference. *Journal of Econometrics*, 93(2), 345-368.

Hansen, B.E. (2000). Sample splitting and threshold estimation. *Econometrica*, 68(3): 575-603.

Hansen, B. E. (2017). Regression kink with an unknown threshold. *Journal of Business and Economic Statistics*, 35(2), 228-240.

Khalaf, L., Kichian, M., Saunders, C. J., Voia, M. (2021). Dynamic panels with MIDAS covariates: Nonlinearity, estimation and fit. *Journal of Econometrics*, 220(2), 589-605.

Lee, M., Gaspar, R., Villaruel, M. L. (2020). Effects of temperature shocks on economic growth and welfare in Asia. *Resources and Environmental Economics*, 2(2), 158-171.

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

Lee, Y., Wang, Y. (2023). Threshold regression with nonparametric sample splitting. *Journal of Econometrics*, 235(2), 816-842.

Lind, J. T., Mehlum, H. (2010). With or without U? The appropriate test for a U-shaped relationship. *Oxford Bulletin of Economics and Statistics*, 72(1), 109-118.

Liu, X., Chen, R. (2020). Threshold factor models for high-dimensional time series. *Journal of Econometrics*, 216:53-70.

Massacci, D. (2017). Least squares estimation of large dimensional threshold factor models. *Journal of Econometrics*, 197(1), 101-129.

Miao, K., Li, K., Su, L. (2020). Panel threshold models with interactive fixed effects. *Journal of Econometrics*, 219(1), 137-170.

Miller, J. I. (2014). Mixed-frequency cointegrating regressions with parsimonious distributed lag structures. *Journal of Financial Econometrics*, 12(3), 584-614.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica*, 46(1), 69-85.

Peligrad, M. (1982). Invariance Principles for Mixing Sequences of Random Variables. *Annals of Probability*, 10, 968-981.

Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure. Econometrica, 74(4), 967-1012.

Pesaran, M. H. (2015). Testing weak cross-sectional dependence in large panels. Econometric Reviews, 34(6-10), 1089-1117.

Seo, M. H., Linton, O. (2007). A smoothed least squares estimator for threshold regression models. *Journal of Econometrics*, 141(2), 704-735.

Seo, M. H., Shin, Y. (2016). Dynamic panels with threshold effect and endogeneity. *Journal of Econometrics*, 195(2), 169-186.

Su, L., Chen, Q. (2013). Testing homogeneity in panel data models with interactive fixed effects. *Econometric Theory*, 29(6), 1079-1135.

Wooldridge, J. M. (1994). Estimation and Inference for Dependent Processes. In R. F. Engle and D. L. McFadden (eds.), *Handbook of Econometrics*, Vol. IV, 2639-2738. Amsterdam: Elsevier.

Wooldridge, J. M. (2019). Correlated random effects models with unbalanced panels. *Journal of Econometrics*, 211(1), 137-150.

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

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

Yang, L., Chen, I. P., Lee, C., Ren, M. (2024). Panel threshold model with covariate-dependent thresholds and unobserved individual-specific threshold effects. *Econometric Reviews*, 43(7), 452-489.

Yang, Y., Jia, F., Li, H. (2023). Estimation of Panel Data Models with Mixed Sampling Frequencies. *Oxford Bulletin of Economics and Statistics*, 85(3), 514-544.

Yu, P., Fan, X. (2021). Threshold Regression With A Threshold Boundary. *Journal of Business and Economic Statistics*, 39(4), 1-59.

Yu, P., Hong, S., Phillips, P. C. B. (2022). Panel threshold regression with unobserved individual-specific threshold effects. *Cowles Foundation Discussion Papers 2737.* https://elischolar.library.yale.edu/cowles-discussion-paper-series/2737.

Appendix A: Mathematical proofs

This appendix contains the mathematical derivations of the theoretical results in the paper. Before proving Theorems 1-4 in the paper, we first provide some technical Lemmas which are useful in proving these theorems. For notational simplicity, we first clarify the following denotations.

1. $a_{NT} = (NT)^{1-2\alpha}$, $\gamma_{10} = \gamma_0$, $\boldsymbol{\gamma}_1 = (\gamma_{11}, \ldots, \gamma_{1k})'$,$\boldsymbol{\gamma} = (\gamma_0, \boldsymbol{\gamma}_1')' = (\gamma_{10}, \gamma_{11}, \ldots, \gamma_{1k})'$, $\boldsymbol{\omega} = (\omega_0, \omega_1')' = (\omega_{10}, \omega_{11}, \ldots, \omega_{1k})'$ and $\boldsymbol{\gamma} = \boldsymbol{\gamma}^0 + \boldsymbol{\omega}/a_{NT}$.

2. $\bar{\mathbf{c}}_g = \partial \boldsymbol{g}(\bar{\boldsymbol{\phi}})/\partial \boldsymbol{\phi}'(\mathbf{c}_1 - \mathbf{c}_2)$, where $\bar{\boldsymbol{\phi}}$ is a value between $\boldsymbol{\phi}_1^0$ and $\boldsymbol{\phi}_2^0$. $\mathbf{c}_g = \partial \boldsymbol{g}(\boldsymbol{\phi}_c)/\partial \boldsymbol{\phi}'(\mathbf{c}_1 - \mathbf{c}_2)$.

3. $I_{it}(\boldsymbol{\gamma}) = I(q_{it} \leq s_{it}(\boldsymbol{\gamma}))$, $\nabla I_{it}(\boldsymbol{\gamma}) = I_{it}(\boldsymbol{\gamma}) - I_{it}(\boldsymbol{\gamma}^0)$.

4. $J_{NT}(\boldsymbol{\gamma}) = (NT)^{-1/2} \sum_{i=1}^N \sum_{t=1}^T \bar{\mathbf{c}}_g' \check{\mathbf{x}}_{it} u_{it} I_{it}(\boldsymbol{\gamma})$.

5. $G_{NT}(\boldsymbol{\omega}, h) = (NTh)^{-1} \sum_{i=1}^N \sum_{t=1}^T \bar{\mathbf{c}}_g' \check{\mathbf{x}}_{it} \check{\mathbf{x}}_{it}' \bar{\mathbf{c}}_g \left| \nabla I_{it}\left(\boldsymbol{\gamma}^0 + \boldsymbol{\omega} h\right)\right|$.

6. $V_{lNT}(\boldsymbol{\omega}, h) = (NTh)^{-1} \sum_{i=1}^N \sum_{t=1}^T \bar{\mathbf{c}}_g' \check{\mathbf{x}}_{it} \check{\mathbf{x}}_{it}' e_{lit}^2 \bar{\mathbf{c}}_g \left| I_{lit}\left(\boldsymbol{\gamma}^0 + \boldsymbol{\omega} h\right)\right|$, where $I_{1it}(\boldsymbol{\gamma}) = -I\left(\mathbf{s}_{it}'\boldsymbol{\gamma} < q_{it} \leq \mathbf{s}_{it}'\boldsymbol{\gamma}^0\right)$ and $I_{2it}(\boldsymbol{\gamma}) = I\left(\mathbf{s}_{it}'\boldsymbol{\gamma}^0 < q_{it} \leq \mathbf{s}_{it}'\boldsymbol{\gamma}\right)$.

7. $R_{NT}(\boldsymbol{\omega}, h) = (NTh)^{-1/2} \sum_{i=1}^N \sum_{t=1}^T \bar{\mathbf{c}}_g' \check{\mathbf{x}}_{it} u_{it} \nabla I_{it}\left(\boldsymbol{\gamma}^0 + \boldsymbol{\omega} h\right)$.

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

9. $y_t = (y_{1t}, \ldots, y_{Nt})'$, $\check{\mathbf{x}}_t = (\check{\mathbf{x}}_{1t}, \ldots, \check{\mathbf{x}}_{Nt})'$, $\check{\mathbf{x}}_t(\boldsymbol{\gamma}) = (\check{\mathbf{x}}_{1t} I_{1t}(\boldsymbol{\gamma}), \ldots, \check{\mathbf{x}}_{Nt} I_{Nt}(\boldsymbol{\gamma}))'$, and $u_t = (u_{1t}, \ldots, u_{Nt})'$. Let $\mathbf{Y}$, $\check{\mathbf{X}}$, $\check{\mathbf{X}}_{\boldsymbol{\gamma}}$ and $\mathbf{u}$ denote the data stacked over all time, i.e., $\mathbf{Y} = (y_1', \ldots, y_T')'$, $\check{\mathbf{X}} = (\check{\mathbf{x}}_1', \ldots, \check{\mathbf{x}}_T)'$, $\check{\mathbf{X}}_{\boldsymbol{\gamma}} = (\check{\mathbf{x}}_1'(\boldsymbol{\gamma}), \ldots, \check{\mathbf{x}}_T'(\boldsymbol{\gamma}))'$, and $\mathbf{u} = (u_1', \ldots, u_N')'$.

10. $\check{\mathbf{X}}_{\boldsymbol{\gamma}}^* = (\check{\mathbf{X}}, \check{\mathbf{X}}_{\boldsymbol{\gamma}})$, $\check{\mathbf{X}}_0 = \check{\mathbf{X}}_{\boldsymbol{\gamma}^0}$, $\check{\mathbf{X}}_0^* = \check{\mathbf{X}}_{\boldsymbol{\gamma}^0}^*$, $\nabla\check{\mathbf{X}}_{\boldsymbol{\gamma}} = \check{\mathbf{X}}_{\boldsymbol{\gamma}} - \check{\mathbf{X}}_0$, $\nabla\check{\mathbf{X}}_{\boldsymbol{\gamma}}^* = \check{\mathbf{X}}_{\boldsymbol{\gamma}}^* - \check{\mathbf{X}}_0^*$, $\boldsymbol{G}(\boldsymbol{\phi}) = (\boldsymbol{g}'(\boldsymbol{\phi}_2), \boldsymbol{g}'(\boldsymbol{\phi}_1) - \boldsymbol{g}'(\boldsymbol{\phi}_2))'$, then the model can be rewritten as $\mathbf{Y} = \check{\mathbf{X}}_{\boldsymbol{\gamma}}^* \boldsymbol{G}(\boldsymbol{\phi}) + \mathbf{u}$.

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

12. $\mathscr{T}_{NT}(\boldsymbol{\gamma}) = (NT)^{\alpha} \left[\boldsymbol{G}\left(\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma})\right) - \boldsymbol{G}\left(\boldsymbol{\phi}^0\right)\right]$.

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

$$\sup_{\boldsymbol{\gamma}\in\mathbf{\Gamma}} \left| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \phi(x_{it})\{q_{it} \leq s_{it}(\boldsymbol{\gamma})\} - E\left(\phi(x_{it})\{q_{it} \leq s_{it}(\boldsymbol{\gamma})\}\right) \right| \xrightarrow{a.s.} 0 \tag{A1}$$

as $N \to \infty$, where $s_{it}(\boldsymbol{\gamma}) = \gamma_0 + \mathbf{s}_{it}'\boldsymbol{\gamma}_1$, $\boldsymbol{\gamma} = (\gamma_0, \boldsymbol{\gamma}_1')'$.

Lemma 2. *Under Assumption 1, for any random variable Z with finite first moment, set $\boldsymbol{\gamma} = \boldsymbol{\gamma}_1 + \boldsymbol{\omega} h$, $\boldsymbol{\omega} \neq \mathbf{0}$, $h > 0$, as $h \to 0$, we have*

$$\frac{P\left(A_{it}(\boldsymbol{\gamma}) \,|\, \mathbf{s}_{it}\right)}{s_{it}(\boldsymbol{\gamma}) - s_{it}(\boldsymbol{\gamma}_1)} \longrightarrow f_{it|\mathbf{s}}\left(s_{it}(\boldsymbol{\gamma}_1) \,|\, \mathbf{s}_{it}\right), \tag{A2}$$

$$\frac{P\left(A_{jk}(\boldsymbol{\gamma}) \,|\, A_{it}(\boldsymbol{\gamma}), \mathbf{s}_{jk}, \mathbf{s}_{it}\right)}{\mathbf{s}_{jk}(\boldsymbol{\gamma}) - s_{jk}(\boldsymbol{\gamma}_1)} \longrightarrow f_{jk|it,\mathbf{s}}\left(s_{jk}(\boldsymbol{\gamma}_1) \,|\, A_{it}^*\left(\boldsymbol{\gamma}_1\right), \mathbf{s}_{jk}, \mathbf{s}_{it}\right), \tag{A3}$$

and

$$\frac{E\left[Z \nabla I_{it}^2\left(\boldsymbol{\gamma}\right)\right]}{h} \longrightarrow E\left(E\left(Z \,|\, A_{it}^*\left(\boldsymbol{\gamma}_1\right), \mathbf{s}_{it}\right) f_{it|\mathbf{s}}\left(s_{it}\left(\boldsymbol{\gamma}_1\right) | \mathbf{s}_{it}\right) \left|\mathbf{s}_{it,1}'\boldsymbol{\omega}\right|\right), \tag{A4}$$

where $A_{it}(\boldsymbol{\gamma}) = \{s_{it}(\boldsymbol{\gamma}^0) < q_{it} \leq s_{it}(\boldsymbol{\gamma})\}$, $A_{it}^*(\boldsymbol{\gamma}) = \{q_{it} = s_{it}(\boldsymbol{\gamma})\}$.

Lemma 3. *Under Assumption 1, set $\boldsymbol{\gamma} = \boldsymbol{\gamma}_1 + \boldsymbol{\omega} h$, $\boldsymbol{\omega} \neq \mathbf{0}$, $h > 0$ and $r \leq 4$, as $h \to 0$*

$$E\left(b_{it}^r\left(\boldsymbol{\gamma}, \boldsymbol{\gamma}_1\right)\right) = O(h), \tag{A5}$$

$$E\left(h_{it}^r\left(\boldsymbol{\gamma}, \boldsymbol{\gamma}_1\right)\right) = O(h). \tag{A6}$$

Lemma 4. *Under Assumption 1, set $\boldsymbol{\gamma} = \boldsymbol{\gamma}_1 + \boldsymbol{\omega} h$, $\boldsymbol{\omega} \neq \mathbf{0}$, $h > 0$, as $h \to 0$*

$$E\left|\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[b_{it}^2(\boldsymbol{\gamma}, \boldsymbol{\gamma}_1) - E(b_{it}^2(\boldsymbol{\gamma}, \boldsymbol{\gamma}_1))\right]\right|^2 = O(h), \tag{A7}$$

$$E\left|\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[h_{it}^2(\boldsymbol{\gamma}, \boldsymbol{\gamma}_1) - E(h_{it}^2(\boldsymbol{\gamma}, \boldsymbol{\gamma}_1))\right]\right|^2 = O(h). \tag{A8}$$

Lemma 5. *There are finite constants K_1 and K_2 such that for all $\boldsymbol{\gamma}_1$, $\eta > 0$, and $\delta \geq (NT)^{-1}$, if $\sqrt{NT} \geq K_2/\eta$, then*

$$P\left(\sup_{\|\boldsymbol{\gamma}-\boldsymbol{\gamma}_1\|<\delta} \left\|J_{NT}(\boldsymbol{\gamma}) - J_{NT}(\boldsymbol{\gamma}_1)\right\| > \eta\right) \leq \frac{K_1\delta^2}{\eta^4}. \tag{A9}$$

Lemma 6. *Under Assumption 1, $J_{NT}(\boldsymbol{\gamma}) \to_d J(\boldsymbol{\gamma})$, a mean-zero Gaussian process with almost surely continuous sample paths.*

Lemma 7. *Under Assumption 1, for any fixed $\boldsymbol{\gamma} \in \mathbf{\Gamma}$,*

$$(NT)^{\alpha}\left[\hat{\boldsymbol{\phi}}\left(\boldsymbol{\gamma}\right) - \boldsymbol{\phi}^0\right] \xrightarrow{p} \mathscr{S}^{-1}(\boldsymbol{\phi}_c, \boldsymbol{\gamma})\mathscr{A}(\boldsymbol{\phi}_c, \boldsymbol{\gamma}), \tag{A10}$$

where $\boldsymbol{\phi}_c = (\phi_c', \phi_c')'$ *and*

$$
\begin{aligned}
\mathscr{A}(\boldsymbol{\phi},\boldsymbol{\gamma}) &= \frac{\partial \boldsymbol{G}'(\boldsymbol{\phi})}{\partial \boldsymbol{\phi}}\left(\begin{bmatrix} \mathbf{M}(\boldsymbol{\gamma}) \\ \mathbf{M}(\boldsymbol{\gamma}) \end{bmatrix} - \begin{bmatrix} \mathbf{M}(\boldsymbol{\gamma}^0) \\ \mathbf{M}(\boldsymbol{\gamma},\boldsymbol{\gamma}^0) \end{bmatrix}\right)\mathbf{c}_g, \\
\mathscr{S}(\boldsymbol{\phi},\boldsymbol{\gamma}) &= \left[\frac{\partial \boldsymbol{G}(\boldsymbol{\phi})}{\partial \boldsymbol{\phi}'}\right]' \mathbf{M}^*(\boldsymbol{\gamma}) \left[\frac{\partial \boldsymbol{G}(\boldsymbol{\phi})}{\partial \boldsymbol{\phi}'}\right].
\end{aligned} \tag{A11}
$$

Moreover, (A10) can be rewritten as

$$
(NT)^{\alpha}\left[\hat{\phi}_1(\boldsymbol{\gamma}) - \phi_1^0\right] \xrightarrow{p} \left[\boldsymbol{g}_c'\mathbf{M}(\boldsymbol{\gamma})\boldsymbol{g}_c\right]^{-1}\boldsymbol{g}_c'\left[\mathbf{M}(\boldsymbol{\gamma}) - \mathbf{M}\left(\boldsymbol{\gamma},\boldsymbol{\gamma}^0\right)\right]\mathbf{c}_g, \tag{A12}
$$

$$
(NT)^{\alpha}\left[\hat{\phi}_2(\boldsymbol{\gamma}) - \phi_2^0\right] \xrightarrow{p} \left[\boldsymbol{g}_c'\left(\mathbf{M} - \mathbf{M}(\boldsymbol{\gamma})\right)\boldsymbol{g}_c\right]^{-1}\boldsymbol{g}_c'\left[\mathbf{M}\left(\boldsymbol{\gamma},\boldsymbol{\gamma}^0\right) - \mathbf{M}\left(\boldsymbol{\gamma}^0\right)\right]\mathbf{c}_g, \tag{A13}
$$

where $\boldsymbol{g}_c = \partial \boldsymbol{g}(\phi)/\partial\phi'|_{\phi=\phi_c}$.

Lemma 8. *Under Assumption 1,* $\hat{\boldsymbol{\gamma}} \to_p \boldsymbol{\gamma}^0$.

Lemma 9. *Under Assumption 1, set* $\boldsymbol{\gamma} = \boldsymbol{\gamma}^0 + \boldsymbol{\omega}h$, $\boldsymbol{\omega} \neq \mathbf{0}$, $h > 0$, *and* $h \to 0$, $(N,T) \to \infty$, *then*

$$
(NT)^{\alpha}\left(\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}) - \boldsymbol{\phi}^0\right) \xrightarrow{p} \mathbf{0}, \tag{A14}
$$

$$
\mathscr{T}_{NT}(\boldsymbol{\gamma}) \xrightarrow{p} \mathbf{0}. \tag{A15}
$$

Lemma 10. *Under Assumption 1, on a compact set* $\boldsymbol{\Psi} = \times_{n=0}^{d}\Psi_n$, *we have the following uniform convergence results, set* $\boldsymbol{\gamma} = \boldsymbol{\gamma}^0 + \boldsymbol{\omega}h$, $\boldsymbol{\omega} \neq \mathbf{0}$, $h > 0$, *such that* $h \to 0$ *and* $NTh \to \infty$ *as* $(N,T) \to \infty$, *then*

$$
G_{NT}(\boldsymbol{\omega},h) \xrightarrow{p} G(\boldsymbol{\omega}), \tag{A16}
$$

$$
V_{lNT}(\boldsymbol{\omega},h) \xrightarrow{p} V_l(\boldsymbol{\omega}), \tag{A17}
$$

where

$$
\begin{aligned}
G(\boldsymbol{\omega}) &= \mathbf{c}_g'E\Big(\mathbf{D}_{it}(\boldsymbol{\gamma}^0)f_{it|\mathbf{s}}\left(s_{it}(\boldsymbol{\gamma}^0)\,|\,\mathbf{s}_{it}\right)\left|\mathbf{s}_{it,1}'\boldsymbol{\omega}\right|\Big)\mathbf{c}_g, \\
V_1(\boldsymbol{\omega}) &= \mathbf{c}_g'E\Big(\mathbf{V}_{1it}(\boldsymbol{\gamma}^0)f_{it|\mathbf{s}}\left(s_{it}(\boldsymbol{\gamma}^0)\,|\,\mathbf{s}_{it}\right)\left|\mathbf{s}_{it,1}'\boldsymbol{\omega}\right| I\left(\mathbf{s}_{it,1}'\boldsymbol{\omega} \le 0\right)\Big)\mathbf{c}_g, \\
V_2(\boldsymbol{\omega}) &= \mathbf{c}_g'E\Big(\mathbf{V}_{2it}(\boldsymbol{\gamma}^0)f_{it|\mathbf{s}}\left(s_{it}(\boldsymbol{\gamma}^0)\,|\,\mathbf{s}_{it}\right)\left|\mathbf{s}_{it,1}'\boldsymbol{\omega}\right| I\left(\mathbf{s}_{it,1}'\boldsymbol{\omega} > 0\right)\Big)\mathbf{c}_g,
\end{aligned}
$$

in which

$$
\begin{aligned}
\mathbf{D}_{it}(\boldsymbol{\gamma}) &= E\left(\check{\mathbf{x}}_{it}\check{\mathbf{x}}_{it}'\,|\,q_{it} = s_{it}(\boldsymbol{\gamma}), \mathbf{s}_{it}\right), \\
\mathbf{V}_{lit}(\boldsymbol{\gamma}) &= E\left(\check{\mathbf{x}}_{it}\check{\mathbf{x}}_{it}'e_{lit}^2\,|\,q_{it} = s_{it}(\boldsymbol{\gamma}), \mathbf{s}_{it}\right).
\end{aligned} \tag{A18}
$$

Lemma 11. *Under Assumption 1, on any compact set* $\boldsymbol{\Psi} = \times_{n=0}^{d}\Psi_n$, *set* $\boldsymbol{\gamma} = \boldsymbol{\gamma}^0 + \boldsymbol{\omega}h$, $\boldsymbol{\omega} \neq \mathbf{0}$, $h > 0$, *such that* $h \to 0$ *and* $NTh \to \infty$ *as* $(N,T) \to \infty$, *then*

$$
R_{NT}(\boldsymbol{\omega},h) \xrightarrow{d} R(\boldsymbol{\omega}), \tag{A19}
$$

where $R(\boldsymbol{\omega}) = R_1(\boldsymbol{\omega}) + R_2(\boldsymbol{\omega})$, $R_l(\boldsymbol{\omega})$ is a Gaussian process with a positive variance $V_l(\boldsymbol{\omega})$ when $\boldsymbol{\omega} \neq \mathbf{0}$, which $R_1(\boldsymbol{\omega})$ and $R_2(\boldsymbol{\omega})$ are independent.

Lemma 12. *Under Assumption 1, we have $a_{NT}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^0) = O_p(1)$.*

Lemma 13. *Under Assumption 1, set $\boldsymbol{\gamma} = \boldsymbol{\gamma}^0 + \boldsymbol{\omega}/a_{NT}$, then*

$$\sqrt{NT}\left[\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}) - \hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}^0)\right] \xrightarrow{p} \mathbf{0}, \text{ and } \sqrt{NT}\left[\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}^0) - \boldsymbol{\phi}^0\right] \xrightarrow{d} \mathcal{Z}, \tag{A20}$$

and

$$\begin{aligned} \sqrt{a_{NT}}\left[\mathscr{T}_{NT}(\boldsymbol{\gamma}) - \mathscr{T}_{NT}(\boldsymbol{\gamma}^0)\right] &\xrightarrow{p} \mathbf{0} &\text{(A21)}\\ \sqrt{a_{NT}}\mathscr{T}_{NT}(\boldsymbol{\gamma}) &\xrightarrow{d} \boldsymbol{A}^{-1}\boldsymbol{J_g}\mathcal{Z}. &\text{(A22)} \end{aligned}$$

where $\mathcal{Z}$ is a Gaussian process with variance $\left(\boldsymbol{J}_{\boldsymbol{g}}'\mathbf{M}^*\boldsymbol{J_g}\right)^{-1}\boldsymbol{J}_{\boldsymbol{g}}'\boldsymbol{\Omega}^*\boldsymbol{J_g}\left(\boldsymbol{J}_{\boldsymbol{g}}'\mathbf{M}^*\boldsymbol{J_g}\right)^{-1}$.

Proof of Lemmas 1-13. To save space, the proofs are skipped here and are provided in the online appendix. □

Proof of Theorem 1. From Lemma 12, the threshold parameters are consistent with convergence rate $a_{NT} = (NT)^{1-2\alpha}$. Thus we can study their asymptotic behavior in the neighborhood of the true thresholds, $\hat{\boldsymbol{\gamma}} = \boldsymbol{\gamma}^0 + \hat{\boldsymbol{\omega}}/a_{NT}$.

By the definition of the threshold estimator, we have

$$\begin{aligned} a_{NT}\left(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^0\right) &= \hat{\boldsymbol{\omega}} \\ &= \arg\min_{\boldsymbol{\omega}\in\boldsymbol{\Psi}} \widetilde{SSR}_{NT}\left(\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}^0 + \frac{\boldsymbol{\omega}}{a_{NT}}), \boldsymbol{\gamma}^0 + \frac{\boldsymbol{\omega}}{a_{NT}}\right) - \widetilde{SSR}_{NT}\left(\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}^0), \boldsymbol{\gamma}^0\right) \\ &= \arg\min_{\boldsymbol{\omega}\in\boldsymbol{\Psi}} Q_{NT}(\boldsymbol{\omega}). \qquad \text{(A23)} \end{aligned}$$

Note that we have $Q_{NT}(\boldsymbol{\omega}) = Q_1 + Q_2 - 2(Q_3 + Q_4 - Q_5 + Q_6 + Q_7)$. We next derive the limiting behavior of each Q_l ($l = 1, 2, ..., 7$), respectively. Thus, by Lemma 1-13, let

$\boldsymbol{\gamma} = \boldsymbol{\gamma}^0 + \boldsymbol{\omega}/a_{NT}$, we have

$$\begin{aligned}
Q_1 &= (NT)^{-2\alpha}\bar{\mathbf{c}}_g'\nabla\check{\mathbf{X}}_{\boldsymbol{\gamma}}'\nabla\check{\mathbf{X}}_{\boldsymbol{\gamma}}\bar{\mathbf{c}}_g \\
&= \frac{a_{NT}}{NT}\bar{\mathbf{c}}_g'\nabla\check{\mathbf{X}}_{\boldsymbol{\gamma}}'\nabla\check{\mathbf{X}}_{\boldsymbol{\gamma}}\bar{\mathbf{c}}_g = G_{NT}\left(\boldsymbol{\omega}, a_{NT}^{-1}\right), && \text{(A24)} \\
Q_2 &= \mathscr{T}_{NT}'(\boldsymbol{\gamma})\frac{a_{NT}}{NT}\left(\check{\mathbf{X}}_{\boldsymbol{\gamma}}^{*\prime}\check{\mathbf{X}}_{\boldsymbol{\gamma}}^{*} - \check{\mathbf{X}}_0^{*\prime}\check{\mathbf{X}}_0^{*}\right)\mathscr{T}_{NT}(\boldsymbol{\gamma}) \\
&= o_p(1)O_p(1)o_p(1) = o_p(1), && \text{(A25)} \\
Q_3 &= (NT)^{-\alpha}\bar{\mathbf{c}}_g'\nabla\check{\mathbf{X}}_{\boldsymbol{\gamma}}'\mathbf{u} \\
&= \sqrt{\frac{a_{NT}}{N}}\bar{\mathbf{c}}_g'\nabla\check{\mathbf{X}}_{\boldsymbol{\gamma}}'\mathbf{u} = R_{NT}\left(\boldsymbol{\omega}, a_{NT}^{-1}\right), && \text{(A26)} \\
Q_4 &= \mathscr{T}_{NT}'(\boldsymbol{\gamma})\sqrt{\frac{a_{NT}}{N}}\nabla\check{\mathbf{X}}_{\boldsymbol{\gamma}}^{*\prime}\mathbf{u} = o_p(1)O_p(1) = o_p(1), && \text{(A27)} \\
Q_5 &= \frac{a_{NT}}{NT}\bar{\mathbf{c}}_g'\nabla\check{\mathbf{X}}_{\boldsymbol{\gamma}}'\check{\mathbf{X}}_{\boldsymbol{\gamma}}^{*}\mathscr{T}_{NT}(\boldsymbol{\gamma}) \\
&= O_p(1)O_p(1)o_p(1) = o_p(1), && \text{(A28)} \\
Q_6 &= \sqrt{a_{NT}}\left[\mathscr{T}_{NT}(\boldsymbol{\gamma}) - \mathscr{T}_{NT}(\boldsymbol{\gamma}^0)\right]'\frac{1}{NT}\check{\mathbf{X}}_0'\check{\mathbf{X}}_0\sqrt{a_{NT}}\left[\mathscr{T}_{NT}(\boldsymbol{\gamma}) + \mathscr{T}_{NT}(\boldsymbol{\gamma}^0)\right] \\
&= o_p(1)O_p(1)O_p(1) = o_p(1), && \text{(A29)} \\
Q_7 &= \sqrt{a_{NT}}\left[\mathscr{T}_{NT}(\boldsymbol{\gamma}) - \mathscr{T}_{NT}(\boldsymbol{\gamma}^0)\right]'N^{-1/2}\check{\mathbf{X}}_0'\mathbf{u} = o_p(1)O_p(1) = o_p(1), && \text{(A30)}
\end{aligned}$$

and therefore we obtain

$$\begin{aligned}
Q_{NT}(\boldsymbol{\omega}) &= Q_1 - 2Q_3 + o_p(1) \\
&= G_{NT}\left(\boldsymbol{\omega}, a_{NT}^{-1}\right) - 2R_{NT}\left(\boldsymbol{\omega}, a_{NT}^{-1}\right) + o_p(1) \\
&\xrightarrow{d} G(\boldsymbol{\omega}) - 2R(\boldsymbol{\omega}). && \text{(A31)}
\end{aligned}$$

Moreover, we have

$$\begin{aligned}
a_{NT}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^0) &\xrightarrow{d} \arg\min_{\boldsymbol{\omega}\in\mathbb{R}^{d+1}} G(\boldsymbol{\omega}) - 2R(\boldsymbol{\omega}) \\
&= \arg\min_{\boldsymbol{\omega}\in\mathbb{R}^{d+1}} \frac{1}{2}G(\boldsymbol{\omega}) - R(\boldsymbol{\omega}). && \text{(A32)}
\end{aligned}$$

This completes the proof of Theorem 1.

□

Proof of Theorem 2. Using Lemma 13, we have

$$\begin{aligned}
\sqrt{NT}\left(\hat{\boldsymbol{\phi}}(\hat{\boldsymbol{\gamma}}) - \boldsymbol{\phi}^0\right) &= \sqrt{NT}\left(\hat{\boldsymbol{\phi}}(\hat{\boldsymbol{\gamma}}) - \hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}^0)\right) + \sqrt{NT}\left(\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}^0) - \boldsymbol{\phi}^0\right) \\
&\xrightarrow{d} \mathbf{0} + \mathcal{Z} = \mathcal{Z}. && \text{(A33)}
\end{aligned}$$

This completes the proof of Theorem 2.

□

Proof of Theorem 3. Under $H_0 : \boldsymbol{\gamma} = \boldsymbol{\gamma}^0$, from proof of Theorem 1, we have

$$\begin{aligned}
&\widetilde{SSR}_{NT}(\hat{\boldsymbol{\phi}}(\hat{\boldsymbol{\gamma}}), \hat{\boldsymbol{\gamma}}) - \widetilde{SSR}_{NT}(\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}^0), \boldsymbol{\gamma}^0) \\
&\quad = \min_{\boldsymbol{\gamma}\in\boldsymbol{\Gamma}} \left[\widetilde{SSR}_{NT}(\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}), \boldsymbol{\gamma}) - \widetilde{SSR}_{NT}(\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}^0), \boldsymbol{\gamma}^0)\right] \\
&\quad = \min_{\boldsymbol{\omega}\in\boldsymbol{\Psi}} \left[\widetilde{SSR}_{NT}(\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}^0 + \frac{\boldsymbol{\omega}}{a_{NT}}), \boldsymbol{\gamma}^0 + \frac{\boldsymbol{\omega}}{a_{NT}}) - \widetilde{SSR}_{NT}(\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}^0), \boldsymbol{\gamma}^0)\right] \\
&\quad \xrightarrow{d} \min_{\boldsymbol{\omega}\in\mathbb{R}^{d+1}} G(\boldsymbol{\omega}) - 2R(\boldsymbol{\omega}). \qquad \text{(A34)}
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
&\frac{1}{NT}\widetilde{SSR}_{NT}(\hat{\boldsymbol{\phi}}(\hat{\boldsymbol{\gamma}}), \hat{\boldsymbol{\gamma}}) = \frac{1}{NT}\widetilde{SSR}_{NT}(\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}^0), \boldsymbol{\gamma}^0) + O_p(\frac{1}{NT}) \\
&\quad = \frac{1}{NT}\left\|\mathbf{u} - \check{\mathbf{X}}_0^*\left[\boldsymbol{G}(\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}^0)) - \boldsymbol{G}(\boldsymbol{\phi}^0)\right]\right\|^2 + O_p(\frac{1}{NT}) \\
&\quad = \frac{1}{NT}\left\|\mathbf{u} - \check{\mathbf{X}}_0^*\mathscr{T}_{NT}(\boldsymbol{\gamma}^0)(NT)^{-\alpha}\right\|^2 + O_p(\frac{1}{NT}) \\
&\quad = \frac{1}{NT}\|\mathbf{u}\|^2 + \frac{1}{NT}\sqrt{a_{NT}}\mathscr{T}_{NT}'(\boldsymbol{\gamma}^0)\frac{1}{NT}\check{\mathbf{X}}_0^{*\prime}\check{\mathbf{X}}_0^*\sqrt{a_{NT}}\mathscr{T}_{NT}(\boldsymbol{\gamma}^0) \\
&\quad\quad - \frac{2}{NT}\sqrt{a_{NT}}\mathscr{T}_{NT}'(\boldsymbol{\gamma}^0)N^{-1/2}\check{\mathbf{X}}_0^*\mathbf{u} + O_p(\frac{1}{NT}) \\
&\quad \xrightarrow{p} E\left(u_{it}^2\right) = \sigma_u^2. \qquad \text{(A35)}
\end{aligned}$$

Finally, we have

$$\begin{aligned}
LR_{NT}(\boldsymbol{\gamma}^0) &= \frac{\widetilde{SSR}_{NT}(\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}^0), \boldsymbol{\gamma}^0) - \widetilde{SSR}_{NT}(\hat{\boldsymbol{\phi}}(\hat{\boldsymbol{\gamma}}), \hat{\boldsymbol{\gamma}})}{\widetilde{SSR}_{NT}(\hat{\boldsymbol{\phi}}(\hat{\boldsymbol{\gamma}}), \hat{\boldsymbol{\gamma}})/(NT)} \\
&\xrightarrow{d} \frac{-1}{\sigma_u^2}\min_{\boldsymbol{\omega}\in\mathbb{R}^{d+1}} G(\boldsymbol{\omega}) - 2R(\boldsymbol{\omega}). \qquad \text{(A36)}
\end{aligned}$$

□

Proof of Theorem 4. We derive the limit distributions of LR_1, LR_2, LR_3 and LR_4 sequentially.

First, we derive the limiting distribution of $\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma})$ by Theorem 10.1 of Wooldridge (1994) under $H_0 : \boldsymbol{\phi}_1 = \boldsymbol{\phi}_2$. Assumptions 1.11-1.12 ensure that conditions (i) and (ii) of Theorem 10.1 of Wooldridge (1994) are satisfied. We next verify condition (iv) and then condition (iii) last.

Under $H_0 : \phi_1 = \phi_2$, using the notation from Lemma 8 and similar proofs, we have

$$\begin{aligned}
\mathbf{Y} - \check{\mathbf{X}}_{\boldsymbol{\gamma}}^*\boldsymbol{G}(\boldsymbol{\phi}) &= \mathbf{u} + \check{\mathbf{X}}\boldsymbol{g}(\boldsymbol{\phi}_2^0) - \check{\mathbf{X}}\boldsymbol{g}(\boldsymbol{\phi}_2) - \check{\mathbf{X}}_{\boldsymbol{\gamma}}\left[\boldsymbol{g}(\boldsymbol{\phi}_1) - \boldsymbol{g}(\boldsymbol{\phi}_2)\right] \\
&= \mathbf{u} + \check{\mathbf{X}}\left[\boldsymbol{g}(\boldsymbol{\phi}_2^0) - \boldsymbol{g}(\boldsymbol{\phi}_2)\right] - \check{\mathbf{X}}_{\boldsymbol{\gamma}}\left[\boldsymbol{g}(\boldsymbol{\phi}_1) - \boldsymbol{g}(\boldsymbol{\phi}_2)\right]. \qquad \text{(A37)}
\end{aligned}$$

Therefore, by Lemma 1-6, for any fixed $\boldsymbol{\gamma} \in \boldsymbol{\Gamma}$, we have

$$\begin{aligned}
\mathbb{D}_{NT}^{-1}\mathbf{K}_{NT}(\boldsymbol{\phi}^0, \boldsymbol{\gamma}) &= \frac{\partial \boldsymbol{G}'(\boldsymbol{\phi}^0)}{\partial \boldsymbol{\phi}}\frac{1}{\sqrt{NT}}\check{\mathbf{X}}_{\boldsymbol{\gamma}}^{*\prime}\mathbf{u} \\
&\xrightarrow{d} \mathcal{Z}^*(\boldsymbol{\gamma}), \qquad \text{(A38)}
\end{aligned}$$

where $\mathscr{Z}_0(\boldsymbol{\gamma})$ is a Gaussian process with variance $\mathbf{M}(\boldsymbol{\gamma})$, and

$$\begin{aligned}
\mathbb{D}_{NT}^{-1}\mathbf{H}_{NT}(\boldsymbol{\phi}^0,\boldsymbol{\gamma})\mathbb{D}_{NT}^{-1} &= \frac{1}{NT}\mathbf{H}_{NT}(\boldsymbol{\phi}^0,\boldsymbol{\gamma}) \\
&= \left[\frac{\partial \boldsymbol{G}(\boldsymbol{\phi}^0)}{\partial \boldsymbol{\phi}'}\right]' \frac{1}{NT}\check{\mathbf{X}}_{\boldsymbol{\gamma}}^{*\prime}\check{\mathbf{X}}_{\boldsymbol{\gamma}}^{*}\left[\frac{\partial \boldsymbol{G}(\boldsymbol{\phi}^0)}{\partial \boldsymbol{\phi}'}\right] \\
&\quad -\left[\left(\frac{1}{NT}\mathbf{u}'\check{\mathbf{X}}_{\boldsymbol{\gamma}}^{*}\right)\otimes \boldsymbol{I}\right]\frac{\partial}{\partial \boldsymbol{\phi}'}\mathrm{vec}\left(\frac{\boldsymbol{G}'(\boldsymbol{\phi}^0)}{\partial \boldsymbol{\phi}}\right) \\
&\xrightarrow{p} \left[\frac{\partial \boldsymbol{G}(\boldsymbol{\phi}_c)}{\partial \boldsymbol{\phi}'}\right]' \mathbf{M}(\boldsymbol{\gamma})\left[\frac{\partial \boldsymbol{G}(\boldsymbol{\phi}_c)}{\partial \boldsymbol{\phi}'}\right]. \qquad \text{(A39)}
\end{aligned}$$

To establish condition (iii) of Wooldridge (1994, Theorem 10.1), we let $\mathbb{C}_{NT}=\mathbb{D}_{NT}^{1-2\delta}=\boldsymbol{I}(NT)^{1/2-\delta}$ for some $0<\delta<1/6$, and

$$\boldsymbol{\phi}_1=\boldsymbol{\phi}_1^0+\boldsymbol{\zeta}_1(NT)^{-1/2+\delta},\ \boldsymbol{\phi}_2=\boldsymbol{\phi}_2^0+\boldsymbol{\zeta}_2(NT)^{-1/2+\delta}, \qquad \text{(A40)}$$

such that $\|\boldsymbol{\zeta}\|\leq 1$, where $\boldsymbol{\zeta}=(\boldsymbol{\zeta}_1',\boldsymbol{\zeta}_2')'$.

Moreover, we have

$$\begin{aligned}
&\frac{1}{NT}\check{\mathbf{X}}_{\boldsymbol{\gamma}}^{*\prime}\left[\mathbf{Y}-\check{\mathbf{X}}_{\boldsymbol{\gamma}}^{*}\boldsymbol{G}(\boldsymbol{\phi})\right] \\
&\quad= \frac{1}{NT}\check{\mathbf{X}}_{\boldsymbol{\gamma}}^{*\prime}\left[\mathbf{u}-\check{\mathbf{X}}_{\boldsymbol{\gamma}}\left[\boldsymbol{g}(\boldsymbol{\phi}_2)-\boldsymbol{g}(\boldsymbol{\phi}_2^0)\right]-\check{\mathbf{X}}_{\boldsymbol{\gamma}}\left[\boldsymbol{g}(\boldsymbol{\phi}_1)-\boldsymbol{g}(\boldsymbol{\phi}_2)\right]\right] \\
&\quad= \frac{1}{NT}\check{\mathbf{X}}_{\boldsymbol{\gamma}}^{*\prime}\left[\mathbf{u}-\check{\mathbf{X}}_{\boldsymbol{\gamma}}\frac{\partial \boldsymbol{g}(\bar{\boldsymbol{\phi}}_2)}{\partial \boldsymbol{\phi}'}\zeta_2(NT)^{-1/2+\delta}-\check{\mathbf{X}}_{\boldsymbol{\gamma}}\frac{\partial \boldsymbol{g}(\bar{\boldsymbol{\phi}})}{\partial \boldsymbol{\phi}'}\left(\zeta_1-\zeta_2\right)(NT)^{-1/2+\delta}\right] \\
&\quad= o_p(1), \qquad \text{(A41)}
\end{aligned}$$

and

$$\begin{aligned}
&\frac{1}{NT}\check{\mathbf{X}}_{\boldsymbol{\gamma}}^{*\prime}\left[\mathbf{Y}-\check{\mathbf{X}}_{\boldsymbol{\gamma}}^{*}\mathbf{G}(\boldsymbol{\phi})\right]-\frac{1}{NT}\check{\mathbf{X}}_{\boldsymbol{\gamma}}^{*\prime}\left[\mathbf{Y}-\check{\mathbf{X}}_{\boldsymbol{\gamma}}^{*}\mathbf{G}(\boldsymbol{\phi}^0)\right] \\
&\quad= \frac{1}{NT}\check{\mathbf{X}}_{\boldsymbol{\gamma}}^{*\prime}\check{\mathbf{X}}_{\boldsymbol{\gamma}}^{*}\left[\boldsymbol{G}(\boldsymbol{\phi})-\boldsymbol{G}(\boldsymbol{\phi}^0)\right] \\
&\quad= \frac{1}{NT}\check{\mathbf{X}}_{\boldsymbol{\gamma}}^{*\prime}\check{\mathbf{X}}_{\boldsymbol{\gamma}}^{*}\frac{\partial \boldsymbol{G}(\bar{\boldsymbol{\phi}})}{\partial \boldsymbol{\phi}'}\boldsymbol{\zeta}(NT)^{-1/2+\delta} \\
&\quad= O_p(1)O(1)O((NT)^{-\alpha(1-2\delta)})=o_p(1). \qquad \text{(A42)}
\end{aligned}$$

Consequently , we have

$$\begin{aligned}
&\mathbb{C}_{NT}^{-1}\left[\mathbf{H}_{NT}(\boldsymbol{\phi},\boldsymbol{\gamma})-\mathbf{H}_{NT}(\boldsymbol{\phi}^0,\boldsymbol{\gamma})\right]\mathbb{C}_{NT}^{-1}\\
=\;& (NT)^{2\delta}\left[\frac{\partial \boldsymbol{G}(\boldsymbol{\phi})}{\partial \boldsymbol{\phi}'}-\frac{\partial \boldsymbol{G}(\boldsymbol{\phi}^0)}{\partial \boldsymbol{\phi}'}\right]'\frac{1}{NT}\check{\mathbf{X}}_{\gamma}^{*\prime}\check{\mathbf{X}}_{\gamma}^{*}\frac{\partial \boldsymbol{G}(\boldsymbol{\phi})}{\partial \boldsymbol{\phi}'}\\
&+(NT)^{2\delta}\frac{\partial \boldsymbol{G}'(\boldsymbol{\phi}^0)}{\partial \boldsymbol{\phi}}\frac{1}{NT}\check{\mathbf{X}}_{\gamma}^{*\prime}\check{\mathbf{X}}_{\gamma}^{*}\left[\frac{\partial \boldsymbol{G}(\boldsymbol{\phi})}{\partial \boldsymbol{\phi}'}-\frac{\partial \boldsymbol{G}(\boldsymbol{\phi}^0)}{\partial \boldsymbol{\phi}'}\right]\\
&-(NT)^{2\delta}\left[\left(\frac{1}{NT}\left[\mathbf{Y}-\check{\mathbf{X}}_{\gamma}^{*}\boldsymbol{G}(\boldsymbol{\phi})\right]'\check{\mathbf{X}}_{\gamma}^{*}\right)\otimes \boldsymbol{I}\right]\\
&\times\left[\frac{\partial}{\partial \boldsymbol{\phi}'}\mathrm{vec}\left(\frac{\partial \boldsymbol{G}'(\boldsymbol{\phi})}{\partial \boldsymbol{\phi}}\right)-\frac{\partial}{\partial \boldsymbol{\phi}'}\mathrm{vec}\left(\frac{\partial \boldsymbol{G}'(\boldsymbol{\phi}^0)}{\partial \boldsymbol{\phi}}\right)\right]\\
&-(NT)^{2\delta}\left[\left(\left[\boldsymbol{G}(\boldsymbol{\phi})-\boldsymbol{G}(\boldsymbol{\phi}^0)\right]\frac{1}{NT}\check{\mathbf{X}}_{\gamma}^{*\prime}\check{\mathbf{X}}_{\gamma}^{*}\right)\otimes \boldsymbol{I}\right]\frac{\partial}{\partial \boldsymbol{\phi}'}\mathrm{vec}\left(\frac{\partial \boldsymbol{G}'(\boldsymbol{\phi}^0)}{\partial \boldsymbol{\phi}}\right)\\
=\;& O(N^{2\delta})O(N^{-1/2+\delta})\left[O_p(1)O(1)+O_p(1)+o_p(1)+o_p(1)\right]\\
=\;& O(N^{-1/2+3\delta})O_p(1)=o_p(1), \qquad \text{(A43)}
\end{aligned}$$

where $0<\delta<1/6 \Leftrightarrow -1/2+3\delta<0$ is used, so that condition (iii) is satisfied.

Hence, by Theorem 10.1 of Wooldridge (1994), we can obtain the following result

$$\begin{aligned}
\sqrt{N}\left(\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma})-\boldsymbol{\phi}^0\right) &= \left[\frac{1}{NT}\mathbf{H}_{NT}(\boldsymbol{\phi}^0,\boldsymbol{\gamma})\right]^{-1}\frac{1}{\sqrt{N}}\mathbf{K}_{NT}(\boldsymbol{\phi}^0,\boldsymbol{\gamma})+o_p(1) &\text{(A44)}\\
&\xrightarrow{d} \left(\boldsymbol{J}_{\boldsymbol{g}}'\mathbf{M}^*(\gamma)\boldsymbol{J}_{\boldsymbol{g}}\right)^{-1}\mathcal{Z}^*(\gamma), &\text{(A45)}
\end{aligned}$$

where $\boldsymbol{\phi}^0$ is the true value of $\boldsymbol{\phi}$, and $\mathcal{Z}(\gamma)$ is a Gaussian process with variance $\boldsymbol{J}_{\boldsymbol{g}}'\boldsymbol{\Omega}^*(\gamma)\boldsymbol{J}_{\boldsymbol{g}}$.

Denote $\widetilde{\boldsymbol{\phi}}=\arg\min_{\boldsymbol{\phi}\in\boldsymbol{\Phi}_{c1}}\widetilde{SSR}_{NT}(\boldsymbol{\phi},\boldsymbol{\gamma})$, where $\boldsymbol{\Phi}_{c1}=\{\boldsymbol{\phi}\in\boldsymbol{\Phi}:\boldsymbol{\phi}_1=\boldsymbol{\phi}_2\}$. Next, we derive the limit distribution of $\widetilde{\boldsymbol{\phi}}-\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma})$. Under $H_0:\boldsymbol{\phi}_1=\boldsymbol{\phi}_2$, the model degenerates into a MIDAS Regression model when estimating parameters using the restriction of $\boldsymbol{\phi}_1=\boldsymbol{\phi}_2$, in which case Andreou et al.(2010) show $\widetilde{\boldsymbol{\phi}}\to_p\boldsymbol{\phi}^0$. Hence, a Taylor's expansion of $\partial\widetilde{SSR}_{NT}(\widetilde{\boldsymbol{\phi}},\boldsymbol{\gamma})/\partial\boldsymbol{\phi}$ at $\boldsymbol{\phi}^0$ gives

$$\begin{aligned}
-2\mathbf{K}_{NT}(\widetilde{\boldsymbol{\phi}},\boldsymbol{\gamma}) &= -2\mathbf{K}_{NT}(\boldsymbol{\phi}^0,\boldsymbol{\gamma})+2\mathbf{H}_{NT}(\bar{\boldsymbol{\phi}}_{c1},\boldsymbol{\gamma})\left(\widetilde{\boldsymbol{\phi}}-\boldsymbol{\phi}^0\right)\\
&= -2\mathbf{K}_{NT}(\boldsymbol{\phi}^0,\boldsymbol{\gamma})+2\mathbf{H}_{NT}(\boldsymbol{\phi}^0,\boldsymbol{\gamma})\left(\widetilde{\boldsymbol{\phi}}-\boldsymbol{\phi}^0\right)+o_p((NT)^{1/2}), \qquad \text{(A46)}
\end{aligned}$$

where $\bar{\boldsymbol{\phi}}_{c1}$ is a value between $\widetilde{\boldsymbol{\phi}}$ and $\boldsymbol{\phi}^0$. Substituting (A46) in the first-order condition for $\widetilde{\boldsymbol{\phi}}$ yields

$$\left[\begin{array}{cc}\frac{1}{NT}\mathbf{H}_{NT}(\boldsymbol{\phi}^0,\boldsymbol{\gamma}) & \mathbf{R}_1'\\ \mathbf{R}_1 & \mathbf{0}\end{array}\right]\left[\begin{array}{c}\sqrt{NT}(\widetilde{\boldsymbol{\phi}}-\boldsymbol{\phi}^0)\\ (NT)^{-1/2}\widetilde{\boldsymbol{\lambda}}_1\end{array}\right]=\left[\begin{array}{c}\frac{1}{\sqrt{NT}}\mathbf{K}_{NT}(\boldsymbol{\phi}^0,\boldsymbol{\gamma})+o_p(1)\\ \mathbf{0}\end{array}\right], \qquad \text{(A47)}$$

where $\mathbf{R}_1$ is a matrix such that $\mathbf{R}_1\boldsymbol{\phi}=\boldsymbol{\phi}_1-\boldsymbol{\phi}_2=\mathbf{0}$ and $\widetilde{\boldsymbol{\lambda}}_1$ is the solution of the first-order conditions of Lagrangian multipliers. Thus, $\sqrt{NT}(\widetilde{\boldsymbol{\phi}}-\boldsymbol{\phi}^0)$ can be solved using the partition inverse rule

$$\sqrt{NT}\left(\widetilde{\boldsymbol{\phi}}-\boldsymbol{\phi}^0\right) = \Big[\widehat{\mathbf{H}}_{NT}^{-1}(\boldsymbol{\phi}^0,\boldsymbol{\gamma})-\widehat{\mathbf{H}}_{NT}^{-1}(\boldsymbol{\phi}^0,\boldsymbol{\gamma})\mathbf{R}_1'\left(\mathbf{R}_1\widehat{\mathbf{H}}_{NT}^{-1}(\boldsymbol{\phi}^0,\boldsymbol{\gamma})\mathbf{R}_1'\right)^{-1} \times \mathbf{R}_1\widehat{\mathbf{H}}_{NT}^{-1}(\boldsymbol{\phi}^0,\boldsymbol{\gamma})\Big]\widehat{\mathbf{K}}_{NT}(\boldsymbol{\phi}^0,\boldsymbol{\gamma})+o_p(1), \tag{A48}$$

where $\widehat{\mathbf{K}}_{NT}(\boldsymbol{\phi}^0,\boldsymbol{\gamma})=(NT)^{-1/2}\mathbf{K}_{NT}(\boldsymbol{\phi}^0,\boldsymbol{\gamma})$ and $\widehat{\mathbf{H}}_{NT}(\boldsymbol{\phi}^0,\boldsymbol{\gamma})=(NT)^{-1}\mathbf{H}_{NT}(\boldsymbol{\phi}^0,\boldsymbol{\gamma})$. From (A44) and (A48), we have

$$\begin{aligned}\sqrt{NT}\left(\widetilde{\boldsymbol{\phi}}-\boldsymbol{\phi}^0\right) &= \sqrt{NT}\left(\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma})-\boldsymbol{\phi}^0\right)+o_p(1)\\ &\quad -\widehat{\mathbf{H}}_{NT}^{-1}(\boldsymbol{\phi}^0,\boldsymbol{\gamma})\mathbf{R}_1'\left(\mathbf{R}_1\widehat{\mathbf{H}}_{NT}^{-1}(\boldsymbol{\phi}^0,\boldsymbol{\gamma})\mathbf{R}_1'\right)^{-1}\mathbf{R}_1\widehat{\mathbf{H}}_{NT}^{-1}(\boldsymbol{\phi}^0,\boldsymbol{\gamma})\\ &\quad \times\widehat{\mathbf{K}}_{NT}(\boldsymbol{\phi}^0,\boldsymbol{\gamma})+o_p(1),\end{aligned} \tag{A49}$$

it can be rewritten as

$$\begin{aligned}\sqrt{NT}\left(\widetilde{\boldsymbol{\phi}}-\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma})\right) &= -\widehat{\mathbf{H}}_{NT}^{-1}(\boldsymbol{\phi}^0,\boldsymbol{\gamma})\mathbf{R}_1'\left(\mathbf{R}_1\widehat{\mathbf{H}}_{NT}^{-1}(\boldsymbol{\phi}^0,\boldsymbol{\gamma})\mathbf{R}_1'\right)^{-1}\mathbf{R}_1\widehat{\mathbf{H}}_{NT}^{-1}(\boldsymbol{\phi}^0,\boldsymbol{\gamma})\\ &\quad \times\widehat{\mathbf{K}}_{NT}(\boldsymbol{\phi}^0,\boldsymbol{\gamma})+o_p(1).\end{aligned} \tag{A50}$$

Thus, we therefore also have $\widetilde{\boldsymbol{\phi}}-\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma})\rightarrow_p \mathbf{0}$.

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

$$\begin{aligned}\widetilde{SSR}_{NT}(\widetilde{\boldsymbol{\phi}},\boldsymbol{\gamma}) &= \widetilde{SSR}_{NT}(\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}),\boldsymbol{\gamma})-2\mathbf{K}_{NT}(\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}),\boldsymbol{\gamma})\left[\widetilde{\boldsymbol{\phi}}-\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma})\right]\\ &\quad +\frac{1}{2}\left[\widetilde{\boldsymbol{\phi}}-\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma})\right]'2\mathbf{H}_{NT}(\bar{\boldsymbol{\phi}}_{\dagger},\boldsymbol{\gamma})\left[\widetilde{\boldsymbol{\phi}}-\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma})\right],\end{aligned} \tag{A51}$$

where $\bar{\boldsymbol{\phi}}_{\dagger}$ is a value between $\widetilde{\boldsymbol{\phi}}$ and $\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma})$, and have the following result

$$\begin{aligned}\left\|\bar{\boldsymbol{\phi}}_{\dagger}-\boldsymbol{\phi}^0\right\| &\le \left\|\bar{\boldsymbol{\phi}}_{\dagger}-\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma})\right\|+\left\|\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma})-\boldsymbol{\phi}^0\right\|\\ &\le \left\|\widetilde{\boldsymbol{\phi}}-\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma})\right\|+\left\|\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma})-\boldsymbol{\phi}^0\right\|\\ &\xrightarrow{p} 0.\end{aligned} \tag{A52}$$

Consequently , we have

$$\begin{aligned}&\widetilde{SSR}_{NT}(\widetilde{\boldsymbol{\phi}},\boldsymbol{\gamma})-\widetilde{SSR}_{NT}(\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}),\boldsymbol{\gamma})\\ &\quad = \sqrt{NT}\left[\widetilde{\boldsymbol{\phi}}-\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma})\right]'\widehat{\mathbf{H}}_{NT}(\bar{\boldsymbol{\phi}}_{\dagger},\boldsymbol{\gamma})\sqrt{NT}\left[\widetilde{\boldsymbol{\phi}}-\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma})\right]\\ &\quad \xrightarrow{d} \mathcal{Z}'(\boldsymbol{\gamma})\mathbf{R}_1'\left[\mathbf{R}_1\left(\boldsymbol{J}_{\boldsymbol{g}}'\mathbf{M}^*(\boldsymbol{\gamma})\boldsymbol{J}_{\boldsymbol{g}}\right)^{-1}\mathbf{R}_1'\right]^{-1}\mathbf{R}_1\mathcal{Z}(\boldsymbol{\gamma}),\end{aligned} \tag{A53}$$

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

$$\begin{aligned}\widetilde{SSR}_{NT}(\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}),\boldsymbol{\gamma}) &= \widetilde{SSR}_{NT}(\boldsymbol{\phi}^0,\boldsymbol{\gamma})-2\mathbf{K}_{NT}(\boldsymbol{\phi}^0,\boldsymbol{\gamma})\left[\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma})-\boldsymbol{\phi}^0\right]\\ &\quad +\frac{1}{2}\left[\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma})-\boldsymbol{\phi}^0\right]'2\mathbf{H}_{NT}(\bar{\boldsymbol{\phi}}_{\ddagger},\boldsymbol{\gamma})\left[\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma})-\boldsymbol{\phi}^0\right]\\ &= \|\mathbf{u}\|^2+O_p(1)+O_p(1),\end{aligned} \tag{A54}$$

where $\bar{\boldsymbol{\phi}}_{\ddagger}$ is a value between $\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma})$ and $\boldsymbol{\phi}^0$, and we obtain

$$\begin{aligned} \frac{1}{NT}\widetilde{SSR}_{NT}(\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}),\boldsymbol{\gamma}) &= \frac{1}{NT}\left\|\mathbf{u}\right\|^2 + O_p(\frac{1}{NT}) + O_p(\frac{1}{NT}) \\ &\xrightarrow{p} \sigma_u^2. \end{aligned} \tag{A55}$$

Finally, we have

$$\begin{aligned} LR_1 &= \frac{\widetilde{SSR}_{NT}(\widetilde{\boldsymbol{\phi}},\hat{\boldsymbol{\gamma}}) - \widetilde{SSR}_{NT}(\hat{\boldsymbol{\phi}}(\hat{\boldsymbol{\gamma}}),\hat{\boldsymbol{\gamma}})}{\widetilde{SSR}_{NT}(\hat{\boldsymbol{\phi}}(\hat{\boldsymbol{\gamma}}),\hat{\boldsymbol{\gamma}})/(NT)} \\ &= \sup_{\boldsymbol{\gamma}\in\boldsymbol{\Gamma}} \frac{\widetilde{SSR}_{NT}(\widetilde{\boldsymbol{\phi}},\boldsymbol{\gamma}) - \widetilde{SSR}_{NT}(\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}),\boldsymbol{\gamma})}{\widetilde{SSR}_{NT}(\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}),\boldsymbol{\gamma})/\left(NT\right)} \\ &\xrightarrow{d} \frac{1}{\sigma_u^2}\sup_{\boldsymbol{\gamma}\in\boldsymbol{\Gamma}} \mathcal{Z}'(\boldsymbol{\gamma})\mathbf{R}_1'\left[\mathbf{R}_1\left(\boldsymbol{J}_{\boldsymbol{g}}'\mathbf{M}^*\left(\boldsymbol{\gamma}\right)\boldsymbol{J}_{\boldsymbol{g}}\right)^{-1}\mathbf{R}_1'\right]^{-1}\mathbf{R}_1\mathcal{Z}(\boldsymbol{\gamma}). \end{aligned} \tag{A56}$$

Second, we derive the limiting distribution of LR_2. Denote

$$\breve{\boldsymbol{\phi}}(\boldsymbol{\gamma}) = \arg\min_{\boldsymbol{\phi}\in\boldsymbol{\Phi}_{c2}} \widetilde{SSR}_{NT}(\boldsymbol{\phi},\boldsymbol{\gamma}), \tag{A57}$$

$$\breve{\boldsymbol{\gamma}} = \arg\min_{\boldsymbol{\gamma}\in\boldsymbol{\Gamma}} \widetilde{SSR}_{NT}(\breve{\boldsymbol{\phi}}(\boldsymbol{\gamma}),\boldsymbol{\gamma}), \tag{A58}$$

where $\boldsymbol{\Phi}_{c2} = \{\boldsymbol{\phi}\in\boldsymbol{\Phi} : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 = \mathbf{0}\}$.

Note that we have

$$\begin{aligned} &(NT)^{-1/2}\mathbf{K}_{NT}(\boldsymbol{\phi}^0,\hat{\boldsymbol{\gamma}}) \\ &\quad= \frac{\partial \boldsymbol{G}'(\boldsymbol{\phi}^0)}{\partial\boldsymbol{\phi}}\left\{\frac{1}{\sqrt{NT}}\breve{\mathbf{X}}_{\hat{\boldsymbol{\gamma}}}^{*\prime}\mathbf{u} - \frac{1}{\sqrt{a_{NT}}}\frac{a_{NT}}{NT}\breve{\mathbf{X}}_{\hat{\boldsymbol{\gamma}}}^{*\prime}\nabla\breve{\mathbf{X}}_{\hat{\boldsymbol{\gamma}}}\bar{\mathbf{c}}_g\right\} \\ &\quad\xrightarrow{d} \mathcal{Z}^*, \end{aligned} \tag{A59}$$

and

$$\begin{aligned} &\frac{1}{NT}\mathbf{H}_{NT}(\boldsymbol{\phi}^0,\hat{\boldsymbol{\gamma}}) \\ &\quad= \left[\frac{\partial\boldsymbol{G}(\boldsymbol{\phi}^0)}{\partial\boldsymbol{\phi}'}\right]'\frac{1}{NT}\breve{\mathbf{X}}_{\hat{\boldsymbol{\gamma}}}^{*\prime}\breve{\mathbf{X}}_{\hat{\boldsymbol{\gamma}}}^{*}\left[\frac{\partial\boldsymbol{G}(\boldsymbol{\phi}^0)}{\partial\boldsymbol{\phi}'}\right] \\ &\qquad -\left[\left(\frac{1}{NT}\mathbf{u}'\breve{\mathbf{X}}_{\hat{\boldsymbol{\gamma}}}^{*} - (NT)^{-\alpha}\bar{\mathbf{c}}_g'\frac{1}{NT}\nabla\breve{\mathbf{X}}_{\hat{\boldsymbol{\gamma}}}'\breve{\mathbf{X}}_{\hat{\boldsymbol{\gamma}}}^{*}\right)\otimes\boldsymbol{I}\right] \\ &\qquad \times\frac{\partial}{\partial\boldsymbol{\phi}'}\mathrm{vec}\left(\frac{\partial\boldsymbol{G}'(\boldsymbol{\phi}^0)}{\partial\boldsymbol{\phi}}\right) \\ &\quad\xrightarrow{p} \boldsymbol{J}_{\boldsymbol{g}}'\mathbf{M}^*\boldsymbol{J}_{\boldsymbol{g}}. \end{aligned} \tag{A60}$$

Under $H_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 = \mathbf{0}$, the model degenerates into a threshold models with a covariate-dependent threshold when we estimate parameters using the restriction of $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 = \mathbf{0}$, in which case a proof similar to Lemma 8 can be used, so we have $\breve{\boldsymbol{\gamma}} \to_p \boldsymbol{\gamma}^0$, By Lemma 13, we also have $\sqrt{NT}(\breve{\boldsymbol{\phi}}(\hat{\boldsymbol{\gamma}}) - \boldsymbol{\phi}^0) = O_p(1)$ as $\hat{\boldsymbol{\gamma}} \to_p \boldsymbol{\gamma}^0$.

A Taylor's expansion of $\partial\widetilde{SSR}_{NT}(\breve{\boldsymbol{\phi}}(\hat{\gamma}),\hat{\gamma})/\partial\boldsymbol{\phi}$ at $\boldsymbol{\phi}^0$ gives

$$\begin{aligned} -2\mathbf{K}_{NT}(\breve{\boldsymbol{\phi}}(\hat{\gamma}),\hat{\gamma}) &= -2\mathbf{K}_{NT}(\boldsymbol{\phi}^0,\hat{\gamma}) + 2\mathbf{H}_{NT}(\bar{\boldsymbol{\phi}}_{c2}),\hat{\gamma})\left(\breve{\boldsymbol{\phi}}(\hat{\gamma})-\boldsymbol{\phi}^0\right) \\ &= -2\mathbf{K}_{NT}(\boldsymbol{\phi}^0,\hat{\gamma}) + 2\mathbf{H}_{NT}(\boldsymbol{\phi}^0,\hat{\gamma})\left(\breve{\boldsymbol{\phi}}(\hat{\gamma})-\boldsymbol{\phi}^0\right) \\ &\quad + o_p((NT)^{1/2}), \end{aligned} \tag{A61}$$

where $\bar{\boldsymbol{\phi}}_{c2}$ is a value between $\breve{\boldsymbol{\phi}}(\hat{\gamma})$ and $\boldsymbol{\phi}^0$. Substituting (A61) in the first-order condition for $\breve{\boldsymbol{\phi}}(\hat{\gamma})$ yields

$$\begin{bmatrix} \frac{1}{NT}\mathbf{H}_{NT}(\boldsymbol{\phi}^0,\hat{\gamma}) & \mathbf{R}_2' \\ \mathbf{R}_2 & \mathbf{0} \end{bmatrix}\begin{bmatrix} \sqrt{NT}(\breve{\boldsymbol{\phi}}(\hat{\gamma})-\boldsymbol{\phi}^0) \\ (NT)^{-1/2}\widetilde{\boldsymbol{\lambda}}_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{NT}}\mathbf{K}_{NT}(\boldsymbol{\phi}^0,\hat{\gamma})+o_p(1) \\ \mathbf{0} \end{bmatrix} \tag{A62}$$

where $\mathbf{R}_2$ is a matrix such that $\mathbf{R}_2\boldsymbol{\phi} = (\boldsymbol{\theta}_1',\boldsymbol{\theta}_2')' = \mathbf{0}$ and $\widetilde{\boldsymbol{\lambda}}_2$ is the solution of the first-order conditions of Lagrangian multipliers. Thus, $\sqrt{NT}(\breve{\boldsymbol{\phi}}(\hat{\gamma})-\boldsymbol{\phi}^0)$ can be solved using the partition inverse rule

$$\begin{aligned} \sqrt{NT}\left(\breve{\boldsymbol{\phi}}(\hat{\gamma})-\boldsymbol{\phi}^0\right) &= \Big[\widehat{\mathbf{H}}_{NT}^{-1}(\boldsymbol{\phi}^0,\hat{\gamma}) - \widehat{\mathbf{H}}_{NT}^{-1}(\boldsymbol{\phi}^0,\hat{\gamma})\mathbf{R}_2'\left(\mathbf{R}_2\widehat{\mathbf{H}}_{NT}^{-1}(\boldsymbol{\phi}^0,\hat{\gamma})\mathbf{R}_2'\right)^{-1} \\ &\quad \times \mathbf{R}_2\widehat{\mathbf{H}}_{NT}^{-1}(\boldsymbol{\phi}^0,\hat{\gamma})\Big]\widehat{\mathbf{K}}_{NT}(\boldsymbol{\phi}^0,\hat{\gamma}) + o_p(1), \end{aligned} \tag{A63}$$

where $\widehat{\mathbf{K}}_{NT}(\boldsymbol{\phi}^0,\hat{\gamma}) = (NT)^{-1/2}\mathbf{K}_{NT}(\boldsymbol{\phi}^0,\hat{\gamma})$ and $\widehat{\mathbf{H}}_{NT}(\boldsymbol{\phi}^0,\hat{\gamma}) = \mathbf{H}_{NT}(\boldsymbol{\phi}^0,\hat{\gamma})/(NT)$. From Theorem 2 and (A63), we have

$$\begin{aligned} \sqrt{NT}\left(\breve{\boldsymbol{\phi}}(\hat{\gamma})-\boldsymbol{\phi}^0\right) &= \sqrt{NT}\left(\hat{\boldsymbol{\phi}}(\hat{\gamma})-\boldsymbol{\phi}^0\right) + o_p(1) \\ &\quad - \widehat{\mathbf{H}}_{NT}^{-1}(\boldsymbol{\phi}^0,\hat{\gamma})\mathbf{R}_2'\left(\mathbf{R}_2\widehat{\mathbf{H}}_{NT}^{-1}(\boldsymbol{\phi}^0,\hat{\gamma})\mathbf{R}_2'\right)^{-1}\mathbf{R}_2\widehat{\mathbf{H}}_{NT}^{-1}(\boldsymbol{\phi}^0,\hat{\gamma}) \\ &\quad \times \widehat{\mathbf{K}}_{NT}(\boldsymbol{\phi}^0,\hat{\gamma}) + o_p(1). \end{aligned} \tag{A64}$$

Noting that (A64) can be rewritten as

$$\begin{aligned} \sqrt{NT}\left(\breve{\boldsymbol{\phi}}(\hat{\gamma})-\hat{\boldsymbol{\phi}}(\hat{\gamma})\right) &= -\widehat{\mathbf{H}}_{NT}^{-1}(\boldsymbol{\phi}^0,\hat{\gamma})\mathbf{R}_2'\left(\mathbf{R}_2\widehat{\mathbf{H}}_{NT}^{-1}(\boldsymbol{\phi}^0,\hat{\gamma})\mathbf{R}_2'\right)^{-1}\mathbf{R}_2\widehat{\mathbf{H}}_{NT}^{-1}(\boldsymbol{\phi}^0,\hat{\gamma}) \\ &\quad \times \widehat{\mathbf{K}}_{NT}(\boldsymbol{\phi}^0,\hat{\gamma}) + o_p(1), \end{aligned} \tag{A65}$$

and hence, we have $\breve{\boldsymbol{\phi}}(\hat{\gamma}) - \hat{\boldsymbol{\phi}}(\hat{\gamma}) \to_p \mathbf{0}$.

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

$$\begin{aligned} \widetilde{SSR}_{NT}(\breve{\boldsymbol{\phi}}(\hat{\gamma}),\hat{\gamma}) &= \widetilde{SSR}_{NT}(\hat{\boldsymbol{\phi}}(\hat{\gamma}),\hat{\gamma}) - 2\mathbf{K}_{NT}(\hat{\boldsymbol{\phi}}(\hat{\gamma}),\hat{\gamma})\left[\breve{\boldsymbol{\phi}}(\hat{\gamma})-\hat{\boldsymbol{\phi}}(\hat{\gamma})\right] \\ &\quad + \frac{1}{2}\left[\breve{\boldsymbol{\phi}}(\hat{\gamma})-\hat{\boldsymbol{\phi}}(\hat{\gamma})\right]' 2\mathbf{H}_{NT}(\bar{\boldsymbol{\phi}}_\star,\hat{\gamma})\left[\breve{\boldsymbol{\phi}}(\hat{\gamma})-\hat{\boldsymbol{\phi}}(\hat{\gamma})\right], \end{aligned} \tag{A66}$$

where $\bar{\boldsymbol{\phi}}_\star$ is a value between $\breve{\boldsymbol{\phi}}(\hat{\gamma})$ and $\hat{\boldsymbol{\phi}}(\hat{\gamma})$.

Thus, we have

$$\begin{aligned} &\widetilde{SSR}_{NT}(\breve{\boldsymbol{\phi}}(\hat{\gamma}),\hat{\gamma}) - \widetilde{SSR}_{NT}(\hat{\boldsymbol{\phi}}(\hat{\gamma}),\hat{\gamma}) \\ &\quad = \sqrt{NT}\left[\breve{\boldsymbol{\phi}}(\hat{\gamma})-\hat{\boldsymbol{\phi}}(\hat{\gamma})\right]'\frac{1}{NT}\mathbf{H}_{NT}(\bar{\boldsymbol{\phi}}_\star,\hat{\gamma})\sqrt{NT}\left[\breve{\boldsymbol{\phi}}(\hat{\gamma})-\hat{\boldsymbol{\phi}}(\hat{\gamma})\right] \\ &\quad \xrightarrow{d} \mathcal{Z}'\mathbf{R}_2'\left[\mathbf{R}_2\left(\boldsymbol{J}_{\boldsymbol{g}}'\mathbf{M}^*\boldsymbol{J}_{\boldsymbol{g}}\right)^{-1}\mathbf{R}_2'\right]^{-1}\mathbf{R}_2\mathcal{Z}, \end{aligned} \tag{A67}$$

and

$$\begin{aligned}
&\left|\widetilde{SSR}_{NT}(\breve{\boldsymbol{\phi}}(\breve{\gamma}),\breve{\gamma})-\widetilde{SSR}_{NT}(\breve{\boldsymbol{\phi}}(\hat{\gamma}),\hat{\gamma})\right| \\
&\quad\leq \left|\widetilde{SSR}_{NT}(\breve{\boldsymbol{\phi}}(\breve{\gamma}),\breve{\gamma})-\widetilde{SSR}_{NT}(\boldsymbol{\phi}^{0},\boldsymbol{\gamma}^{0})\right| \\
&\qquad+\left|\widetilde{SSR}_{NT}(\boldsymbol{\phi}^{0},\boldsymbol{\gamma}^{0})-\widetilde{SSR}_{NT}(\breve{\boldsymbol{\phi}}(\hat{\gamma}),\hat{\gamma})\right| \\
&\quad\xrightarrow{p} 0+0=0,
\end{aligned} \tag{A68}$$

because the limit function of $\widetilde{SSR}_{NT}(\breve{\boldsymbol{\phi}}(\boldsymbol{\gamma}),\boldsymbol{\gamma})$ is continuous.

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

$$\begin{aligned}
\widetilde{SSR}_{NT}(\hat{\boldsymbol{\phi}}(\hat{\gamma}),\hat{\gamma}) &= \widetilde{SSR}_{NT}(\boldsymbol{\phi}^{0},\hat{\gamma})-2\mathbf{K}_{NT}(\boldsymbol{\phi}^{0},\hat{\gamma})\left[\hat{\boldsymbol{\phi}}(\hat{\gamma})-\boldsymbol{\phi}^{0}\right] \\
&\quad+\frac{1}{2}\left[\hat{\boldsymbol{\phi}}(\hat{\gamma})-\boldsymbol{\phi}^{0}\right]'2\mathbf{H}_{NT}(\bar{\boldsymbol{\phi}}_{*},\hat{\gamma})\left[\hat{\boldsymbol{\phi}}(\hat{\gamma})-\boldsymbol{\phi}^{0}\right] \\
&= \left\|\mathbf{u}-\nabla\check{\mathbf{X}}_{\hat{\gamma}}\bar{\mathbf{c}}_{g}(NT)^{-\alpha}\right\|^{2}+O_{p}(1)+O_{p}(1),
\end{aligned} \tag{A69}$$

where $\bar{\boldsymbol{\phi}}_{*}$ is a value between $\hat{\boldsymbol{\phi}}(\hat{\gamma})$ and $\boldsymbol{\phi}^{0}$. Thus, we obtain

$$\begin{aligned}
&\frac{1}{NT}\widetilde{SSR}_{NT}(\hat{\boldsymbol{\phi}}(\hat{\gamma}),\hat{\gamma}) \\
&\quad= \frac{1}{NT}\left\|\mathbf{u}\right\|^{2}+\bar{\mathbf{c}}_{g}'\frac{1}{NT}\nabla\check{\mathbf{X}}_{\hat{\gamma}}'\nabla\check{\mathbf{X}}_{\hat{\gamma}}\bar{\mathbf{c}}_{g}(NT)^{-2\alpha} \\
&\qquad-2\bar{\mathbf{c}}_{g}'\frac{1}{NT}\nabla\check{\mathbf{X}}_{\hat{\gamma}}'\mathbf{u}(NT)^{-\alpha}+O_{p}(\frac{1}{NT}) \\
&\quad\xrightarrow{p} \sigma_{u}^{2}+0+0=\sigma_{u}^{2}.
\end{aligned} \tag{A70}$$

Finally, we have

$$\begin{aligned}
LR_{2} &= \frac{\widetilde{SSR}_{NT}(\breve{\boldsymbol{\phi}}(\breve{\gamma}),\breve{\gamma})-\widetilde{SSR}_{NT}(\hat{\boldsymbol{\phi}}(\hat{\gamma}),\hat{\gamma})}{\widetilde{SSR}_{NT}(\hat{\boldsymbol{\phi}}(\hat{\gamma}),\hat{\gamma})/(NT)} \\
&= \frac{\widetilde{SSR}_{NT}(\breve{\boldsymbol{\phi}}(\breve{\gamma}),\breve{\gamma})-\widetilde{SSR}_{NT}(\breve{\boldsymbol{\phi}}(\hat{\gamma}),\hat{\gamma})}{\widetilde{SSR}_{NT}(\hat{\boldsymbol{\phi}}(\hat{\gamma}),\hat{\gamma})/(NT)} \\
&\quad+\frac{\widetilde{SSR}_{NT}(\breve{\boldsymbol{\phi}}(\hat{\gamma}),\hat{\gamma})-\widetilde{SSR}_{NT}(\hat{\boldsymbol{\phi}}(\hat{\gamma}),\hat{\gamma})}{\widetilde{SSR}_{NT}(\hat{\boldsymbol{\phi}}(\hat{\gamma}),\hat{\gamma})/(NT)} \\
&\xrightarrow{d} \frac{0}{\sigma_{u}^{2}}+\frac{1}{\sigma_{u}^{2}}\mathcal{Z}'\mathbf{R}_{2}'\left[\mathbf{R}_{2}\left(\boldsymbol{J}_{\boldsymbol{g}}'\mathbf{M}^{*}\boldsymbol{J}_{\boldsymbol{g}}\right)^{-1}\mathbf{R}_{2}'\right]^{-1}\mathbf{R}_{2}\mathcal{Z} \\
&= \frac{1}{\sigma_{u}^{2}}\mathcal{Z}'\mathbf{R}_{2}'\left[\mathbf{R}_{2}\left(\boldsymbol{J}_{\boldsymbol{g}}'\mathbf{M}^{*}\boldsymbol{J}_{\boldsymbol{g}}\right)^{-1}\mathbf{R}_{2}'\right]^{-1}\mathbf{R}_{2}\mathcal{Z}.
\end{aligned} \tag{A71}$$

Third, we derive the limiting distribution of LR_3. Denote

$$\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}) = \arg\min_{\boldsymbol{\phi}\in\boldsymbol{\Phi}}\widetilde{SSR}_{NT}(\boldsymbol{\phi},\boldsymbol{\gamma}), \tag{A72}$$

$$\widetilde{\boldsymbol{\gamma}} = \arg\min_{\boldsymbol{\gamma}\in\boldsymbol{\Gamma}_{c}}\widetilde{SSR}_{NT}(\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}),\boldsymbol{\gamma}), \tag{A73}$$

where $\boldsymbol{\Gamma}_c=\{\boldsymbol{\gamma}\in\boldsymbol{\Gamma}:\boldsymbol{\gamma}_1=\mathbf{0}\}$ and $\widetilde{\boldsymbol{\gamma}}=(\tilde{\gamma}_0,\tilde{\boldsymbol{\gamma}}_1')'=(\tilde{\gamma}_0,\mathbf{0}')'$.

Under $H_0:\boldsymbol{\gamma}_1=\mathbf{0}$, the model degenerates into a threshold MIDAS regression models with a constant threshold when estimating parameters using the restriction of $\boldsymbol{\gamma}_1=\mathbf{0}$, in which case a proof similar to Theorem 1 can be used, so we have following result

$$\begin{aligned}
&\widetilde{SSR}_{NT}(\hat{\boldsymbol{\phi}}(\widetilde{\boldsymbol{\gamma}}),\widetilde{\boldsymbol{\gamma}})-\widetilde{SSR}_{NT}(\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}^0),\boldsymbol{\gamma}^0)\\
&\quad=\min_{\boldsymbol{\gamma}\in\boldsymbol{\Gamma}_c}\left[\widetilde{SSR}_{NT}(\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}),\boldsymbol{\gamma})-\widetilde{SSR}_{NT}(\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}^0),\boldsymbol{\gamma}^0)\right]\\
&\quad=\min_{\boldsymbol{\omega}\in\boldsymbol{\Psi}_c}\left[\widetilde{SSR}_{NT}(\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}^0+\frac{\boldsymbol{\omega}}{a_{NT}}),\boldsymbol{\gamma}^0+\frac{\boldsymbol{\omega}}{a_{NT}})-\widetilde{SSR}_{NT}(\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}^0),\boldsymbol{\gamma}^0)\right]\\
&\quad\xrightarrow{d}\min_{\boldsymbol{\omega}\in\mathbb{R}_c^{d+1}}G(\boldsymbol{\omega})-2R(\boldsymbol{\omega}),
\end{aligned}\tag{A74}$$

where $\boldsymbol{\Psi}_c=\{\boldsymbol{\omega}\in\boldsymbol{\Psi}:\boldsymbol{\omega}_1=\mathbf{0}\}$, $\mathbb{R}_c^{d+1}=\{\boldsymbol{\omega}\in\mathbb{R}^{d+1}:\boldsymbol{\omega}_1=\mathbf{0}\}$, and from the proof of Theorem 3, we have

$$\begin{aligned}
&\widetilde{SSR}_{NT}(\hat{\boldsymbol{\phi}}(\hat{\boldsymbol{\gamma}}),\hat{\boldsymbol{\gamma}})-\widetilde{SSR}_{NT}(\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}^0),\boldsymbol{\gamma}^0)\\
&\quad\xrightarrow{d}\min_{\boldsymbol{\omega}\in\mathbb{R}^{d+1}}G(\boldsymbol{\omega})-2R(\boldsymbol{\omega}),
\end{aligned}\tag{A75}$$

and

$$\begin{aligned}
\frac{1}{NT}\widetilde{SSR}_{NT}(\hat{\boldsymbol{\phi}}(\hat{\boldsymbol{\gamma}}),\hat{\boldsymbol{\gamma}})&=\frac{1}{NT}\left\|\mathbf{u}-\check{\mathbf{X}}_0^*\mathscr{T}_{NT}\left(\boldsymbol{\gamma}^0\right)(NT)^{-\alpha}\right\|^2+O_p(\frac{1}{NT})\\
&=\frac{1}{NT}\left\|\mathbf{u}\right\|^2+\mathscr{T}_{NT}'\left(\boldsymbol{\gamma}^0\right)\frac{1}{NT}\check{\mathbf{X}}_0^{*\prime}\check{\mathbf{X}}_0^*\mathscr{T}_{NT}\left(\boldsymbol{\gamma}^0\right)(NT)^{-2\alpha}\\
&\quad+2\mathscr{T}_{NT}'\left(\boldsymbol{\gamma}^0\right)\frac{1}{NT}\check{\mathbf{X}}_0^{*\prime}\mathbf{u}(NT)^{-\alpha}+O_p(\frac{1}{NT})\\
&\xrightarrow{p}\sigma_u^2,
\end{aligned}\tag{A76}$$

Hence, we have

$$\begin{aligned}
LR_3&=\frac{\widetilde{SSR}_{NT}(\hat{\boldsymbol{\phi}}(\widetilde{\boldsymbol{\gamma}}),\widetilde{\boldsymbol{\gamma}})-\widetilde{SSR}_{NT}(\hat{\boldsymbol{\phi}}(\hat{\boldsymbol{\gamma}}),\hat{\boldsymbol{\gamma}})}{\widetilde{SSR}_{NT}(\hat{\boldsymbol{\phi}}(\hat{\boldsymbol{\gamma}}),\hat{\boldsymbol{\gamma}})/(NT)}\\
&=\frac{\widetilde{SSR}_{NT}(\hat{\boldsymbol{\phi}}(\widetilde{\boldsymbol{\gamma}}),\widetilde{\boldsymbol{\gamma}})-\widetilde{SSR}_{NT}(\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}^0),\boldsymbol{\gamma}^0)}{\widetilde{SSR}_{NT}(\hat{\boldsymbol{\phi}}(\hat{\boldsymbol{\gamma}}),\hat{\boldsymbol{\gamma}})/(NT)}\\
&\quad+\frac{\widetilde{SSR}_{NT}(\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}^0),\boldsymbol{\gamma}^0)-\widetilde{SSR}_{NT}(\hat{\boldsymbol{\phi}}(\hat{\boldsymbol{\gamma}}),\hat{\boldsymbol{\gamma}})}{\widetilde{SSR}_{NT}(\hat{\boldsymbol{\phi}}(\hat{\boldsymbol{\gamma}}),\hat{\boldsymbol{\gamma}})/(NT)}\\
&\xrightarrow{d}\frac{1}{\sigma_u^2}\left[\left(\min_{\boldsymbol{\omega}\in\mathbb{R}_c^{d+1}}G(\boldsymbol{\omega})-2R(\boldsymbol{\omega})\right)-\left(\min_{\boldsymbol{\omega}\in\mathbb{R}^{d+1}}G(\boldsymbol{\omega})-2R(\boldsymbol{\omega})\right)\right].
\end{aligned}\tag{A77}$$

Finally, the proof for LR_4 is almost identical to LR_2, as the null hypotheses in the two cases are all restrictions for the slope parameters of the PTMIDAS-CDT model. Thus, the distribution of LR_4 can be derived as LR_2 by replacing $\mathbf{R}_2$ with $\mathbf{R}_I$. □

Appendix B: A simple forecasting exercise

In this appendix, we conduct a simple forecasting exercise to compare the forecasting performance of different panel MIDAS models. To this end, we examine the out-of-sample forecasting performance of competing panel MIDAS models in forecasting economic growth. The considered models include linear panel MIDAS model, panel threshold MIDAS with a constant threshold and the PTMIDAS-CDT model (31) in Section 4; meanwhile, we employ the random walk model as a benchmark.

In forecasting economic growth, we split the sample into an estimation sample and an evaluation sample. In detail, we set the sample during the period from 1961 to 1993 as the estimation sample, and obtain one-step ahead forecasts until 2003 (the end of the sample) using an expanding window scheme. Table B1 reports the mean square forecast errors (MSFE) of different models, in which we report the MSEs for the full sample of 78 countries, and we also report the MSEs for the subsamples of 40 poor countries and 38 rich countries. As can be seen from the forecasting results in Table B1, while none of the panel MIDAS model is able to beat the random walk model in the subsample for rich countries, the PTMIDAS-CDT model slightly outperforms others models in the full sample and in the subsample for poor countries. Future work could extend our analysis to allow for more sophisticated empirical model settings.

Table B1: Mean square forecast errors (MSFE) of different models.

MSFE	All countries	Poor countries	Rich countries
Random-Walk	49.984	45.322	4.662
Linear panel MIDAS	31.500	26.712	4.788
PTMIDAS	32.177	27.154	5.024
PTMIDAS-CDT	31.435	26.359	5.076

Online Appendix: Panel Threshold Mixed Data Sampling Models with a Covariate-Dependent Threshold

This online appendix consists 2 sections: Appendix A provides the detailed proofs of the Lemmas in the paper, and Appendix B compares the empirical results based on a simple polynomial specification with these based on our proposed model.

Appendix A: Detailed proofs

In this appendix, we provide the detailed proofs of the Lemmas in the paper.

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

$$\sup_{\boldsymbol{\gamma}\in\mathbf{\Gamma}} \left| \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\phi(x_{it})\{q_{it} \leq s_{it}(\boldsymbol{\gamma})\} - E\left(\phi(x_{it})\{q_{it} \leq s_{it}(\boldsymbol{\gamma})\}\right)\right| \xrightarrow{a.s.} 0 \tag{A1}$$

as $N \to \infty$, where $s_{it}(\boldsymbol{\gamma}) = \gamma_0 + \mathbf{s}_{it}'\gamma_1$, $\boldsymbol{\gamma} = (\gamma_0, \boldsymbol{\gamma}_1')'$.

Proof. The proof is similar in Hansen (1996, Lemma 1). For convenience, denote $\gamma_{10} = \gamma_0$, $\Gamma_{10} = \Gamma_0$, $s_{0it} = 1$ and $\mathbf{s}_{it} = (s_{1it}, \ldots, s_{dit})'$. Let $\Gamma_{1n} = [\underline{\gamma}_{1n}, \overline{\gamma}_{1n}]$ for $n = 0, 1, ..., k$, and set

$$J_\varepsilon = \left\lfloor E|\phi(x_t)|^2 \bar{f} C_1 \left(\sum_{n=0}^{d} \overline{\gamma}_{1n} - \underline{\gamma}_{1n}\right) \Big/ \varepsilon^2 \right\rfloor + 1$$

, $\gamma_{1nj} = \underline{\gamma}_{1n} + j(\overline{\gamma}_{1n} - \underline{\gamma}_{1n})/J_\varepsilon$ for $n = 0, 1, ..., k$, $j = 0, 1, ..., J_\varepsilon$, $\Gamma_{1nj} = [\gamma_{1nj}, \gamma_{1nj+1}]$ and $\mathbf{\Gamma}_j = \times_{n=0}^{d}\Gamma_{1nj}$, where $\lfloor a \rfloor$ is the greatest integer less than or equal to a, $C_1 = \max\{1, C\}$.

Note that for any $\boldsymbol{\gamma} \in \mathbf{\Gamma}_j$ for $j = 0, 1, ..., J_\varepsilon - 1$, we have

$$\gamma_{1nj} \leq \gamma_{1n} \leq \gamma_{1nj+1}, \text{ for } n = 0, 1, ..., k, \tag{A2}$$

times s_{nit}, we get

$$\gamma_{1nj}s_{nit}^{+} + \gamma_{1nj+1}s_{nit}^{-} \leq \gamma_{1n}s_{nit} \leq \gamma_{1nj+1}s_{nit}^{+} + \gamma_{1nj}s_{nit}^{-}, \tag{A3}$$

where $s_{nit}^{+} = s_{nit}\{s_{nit} > 0\}$ and $s_{nit}^{-} = s_{nit}\{s_{nit} \leq 0\}$. Take sum of n from 0 to d for (A3), we have

$$\gamma_{itj}^{l} \leq s_{it}(\boldsymbol{\gamma}) \leq \gamma_{itj}^{u} \text{ and } \{q_{it} \leq \gamma_{itj}^{l}\} \leq \{q_{it} \leq s_{it}(\boldsymbol{\gamma})\} \leq \{q_{it} \leq \gamma_{itj}^{u}\}, \tag{A4}$$

where $\gamma_{itj}^{l} = \sum_{n=0}^{d}[\gamma_{1nj}s_{nit}^{+} + \gamma_{1nj+1}s_{nit}^{-}]$ and $\gamma_{itj}^{u} = \sum_{n=0}^{d}[\gamma_{1nj+1}s_{nit}^{+} + \gamma_{1nj}s_{nit}^{-}]$.

Times $\phi(x_{it})$, we get

$$f_{\varepsilon,j}^{l}(x_{it}) \leq \phi(x_{it})\{q_{it} \leq s_{it}(\boldsymbol{\gamma})\} \leq f_{\varepsilon,j}^{u}(x_{it}), \tag{A5}$$

where $f^l_{\varepsilon,j}(x_{it}) = \phi(x_{it})\{q_{it} \leq \gamma^l_{itj}\}\{\phi(x_{it}) > 0\} + \phi(x_{it})\{q_{it} \leq \gamma^u_{itj}\}\{\phi(x_{it}) \leq 0\}$ and $f^u_{\varepsilon,j}(x_{it}) = \phi(x_{it})\{q_{it} \leq \gamma^u_{itj}\}\{\phi(x_{it}) > 0\} + \phi(x_{it})\{q_{it} \leq \gamma^l_{itj}\}\{\phi(x_{it}) \leq 0\}$.

Next, we show that $E|f^u_{\varepsilon,j}(x_{it}) - f^l_{\varepsilon,j}(x_{it})| \leq \varepsilon$, note that

$$\begin{aligned}
\left|f^u_{\varepsilon,j}(x_{it}) - f^l_{\varepsilon,j}(x_{it})\right| &= \Big|\phi(x_{it})\{\phi(x_{it}) > 0\}\{\gamma^l_{itj} < x_{it} \leq \gamma^u_{itj}\} \\
&\quad -\phi(x_{it})\{\phi(x_{it}) \leq 0\}\{\gamma^l_{itj} < x_{it} \leq \gamma^u_{itj}\}\Big| \\
&= \Big|\phi(x_{it})\{\gamma^l_{itj} < x_{it} \leq \gamma^u_{itj}\}[\{\phi(x_{it}) > 0\} - \{\phi(x_{it}) \leq 0\}]\Big| \\
&= \Big|\phi(x_{it})\{\gamma^l_{itj} < x_{it} \leq \gamma^u_{itj}\}\Big| \qquad \text{(A6)}
\end{aligned}$$

where $|\{\phi(x_{it}) > 0\} - \{\phi(x_{it}) \leq 0\}| = 1$, and

$$\begin{aligned}
E\left(\{\gamma^l_{tj} < x_{it} \leq \gamma^u_{tj}\}\,\Big|\,\mathbf{s}_{it}\right) &= P\left(\gamma^u_{tj}\,|\,\mathbf{s}_{it}\right) - P\left(\gamma^l_{tj}\,|\,\mathbf{s}_{it}\right) \\
&\leq \bar{f}\left|\gamma^u_{tj} - \gamma^l_{tj}\right| \\
&\leq \bar{f}\Big|\sum_{n=0}^{d} s_{nit}\{s_{nit} > 0\}[\gamma_{1nj+1} - \gamma_{1nj}] \\
&\quad -s_{nit}\{s_{nit} \leq 0\}[\gamma_{1nj+1} - \gamma_{1nj}]\Big| \\
&= \bar{f}\left|\sum_{n=0}^{d} s_{nit}[\gamma_{1nj+1} - \gamma_{1nj}][\{s_{nit} > 0\} - \{s_{nit} \leq 0\}]\right| \\
&\leq \bar{f}\sum_{n=0}^{d} |s_{nit}|\,[\gamma_{1nj+1} - \gamma_{1nj}], \\
&\leq \bar{f}\left[\gamma_{10j+1} - \gamma_{10j} + \sum_{n=1}^{d} |s_{nit}|\,[\gamma_{1nj+1} - \gamma_{1nj}]\right] \qquad \text{(A7)}
\end{aligned}$$

in which $|[\{s_{it} > 0\} - \{s_{it} \leq 0\}]| = 1$. Use the law of total expectation, Assumption 1 and we set partition, we get

$$\begin{aligned}
E\left(\{\gamma^l_{itj} < x_{it} \leq \gamma^u_{itj}\}\right) &= E\left(E\left(\{\gamma^l_{itj} < x_{it} \leq \gamma^u_{itj}\}\,\Big|\,\mathbf{s}_{it}\right)\right) \\
&\leq \bar{f}\left[\gamma_{10j+1} - \gamma_{10j} + \sum_{n=1}^{d} E\,|s_{nit}|\,[\gamma_{1nj+1} - \gamma_{1nj}]\right] \\
&\leq \bar{f}\left[\gamma_{10j+1} - \gamma_{10j} + C\sum_{n=1}^{d}[\gamma_{1nj+1} - \gamma_{1nj}]\right] \\
&\leq \bar{f}C_1\left(\sum_{n=0}^{d}\frac{\overline{\gamma}_{1n} - \underline{\gamma}_{1n}}{J_\varepsilon}\right). \qquad \text{(A8)}
\end{aligned}$$

Hence, by Holder's inequality,

$$\begin{aligned}
E\left|f_{\varepsilon,j}^{u}(x_{it})-f_{\varepsilon,j}^{l}(x_{it})\right| &= E\left|\phi(x_{it})\{\gamma_{itj}^{l}<x_{it}\le\gamma_{itj}^{u}\}\right| \\
&\le \left(E|\phi(x_{it})|^{2}\right)^{\frac{1}{2}}\left(E|\{\gamma_{itj}^{l}<x_{it}\le\gamma_{itj}^{u}\}|\right)^{\frac{1}{2}} \\
&\le \left(E|\phi(x_{it})|^{2}\right)^{\frac{1}{2}}\bar{f}C_{1}\left(\sum_{n=0}^{d}\overline{\gamma}_{1n}-\underline{\gamma}_{1n}\right)^{\frac{1}{2}}\Big/J_{\varepsilon}^{\frac{1}{2}} \\
&< \varepsilon \qquad\qquad (A9)
\end{aligned}$$

Finally, equations (A5) and (A9) satisfy conditions of Pollard (1984, Theorem II.2), so we can obtain the result of equation (A1). The proof of Lemma 1 is complete.

□

Lemma 2. *Under Assumption 1, for any random variable Z with finite first moment, set $\boldsymbol{\gamma}=\boldsymbol{\gamma}_1+\boldsymbol{\omega}h$, $\boldsymbol{\omega}\neq\mathbf{0}$, $h>0$, as $h\to 0$, we have*

$$\frac{P\left(A_{it}(\boldsymbol{\gamma})\,|\,\mathbf{s}_{it}\right)}{s_{it}(\boldsymbol{\gamma})-s_{it}(\boldsymbol{\gamma}_1)} \longrightarrow f_{it|\mathbf{s}}\left(s_{it}(\boldsymbol{\gamma}_1)\,|\,\mathbf{s}_{it}\right), \qquad (A10)$$

$$\frac{P\left(A_{jk}(\boldsymbol{\gamma})\,|\,A_{it}(\boldsymbol{\gamma}),\mathbf{s}_{jk},\mathbf{s}_{it}\right)}{\mathbf{s}_{jk}(\boldsymbol{\gamma})-s_{jk}(\boldsymbol{\gamma}_1)} \longrightarrow f_{jk|it,\mathbf{s}}\left(s_{jk}(\boldsymbol{\gamma}_1)\,|\,A_{it}^{*}\left(\boldsymbol{\gamma}_1\right),\mathbf{s}_{jk},\mathbf{s}_{it}\right), \qquad (A11)$$

and

$$\frac{E\left[Z\nabla I_{it}^{2}\left(\boldsymbol{\gamma}\right)\right]}{h} \longrightarrow E\left(E\left(Z\,|\,A_{it}^{*}\left(\boldsymbol{\gamma}_1\right),\mathbf{s}_{it}\right)f_{it|\mathbf{s}}\left(s_{it}\left(\boldsymbol{\gamma}_1\right)|\,\mathbf{s}_{it}\right)\left|\mathbf{s}_{it,1}^{\prime}\boldsymbol{\omega}\right|\right), \qquad (A12)$$

where $A_{it}(\boldsymbol{\gamma})=\{s_{it}(\boldsymbol{\gamma}^{0})<q_{it}\le s_{it}(\boldsymbol{\gamma})\}$, $A_{it}^{*}(\boldsymbol{\gamma})=\{q_{it}=s_{it}(\boldsymbol{\gamma})\}$.

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

$$\begin{aligned}
s_{it}(\boldsymbol{\gamma})-s_{it}(\boldsymbol{\gamma}_1) &= \mathbf{s}_{it,1}^{\prime}\left(\boldsymbol{\gamma}-\boldsymbol{\gamma}_1\right)=\mathbf{s}_{it,1}^{\prime}\boldsymbol{\omega}h \\
&\longrightarrow 0,
\end{aligned}$$

where $\mathbf{s}_{it,1}=(1,\mathbf{s}_{it}^{\prime})^{\prime}$.

Next, denote $h^{*}=s_{it}(\boldsymbol{\gamma})-s_{it}(\boldsymbol{\gamma}_1)$, then we have

$$\begin{aligned}
\frac{P\left(A_{it}(\boldsymbol{\gamma})\,|\,\mathbf{s}_{it}\right)}{s_{it}(\boldsymbol{\gamma})-s_{it}(\boldsymbol{\gamma}_1)} &= \frac{P\left(q_{it}\le s_{it}(\boldsymbol{\gamma})\,|\,\mathbf{s}_{it}\right)-P\left(q_{it}\le s_{it}(\boldsymbol{\gamma}_1)\,|\,\mathbf{s}_{it}\right)}{s_{it}(\boldsymbol{\gamma})-s_{it}(\boldsymbol{\gamma}_1)} \\
&= \frac{P\left(q_{it}\le s_{it}(\boldsymbol{\gamma}_1)+h^{*}\,|\,\mathbf{s}_{it}\right)-P\left(q_{it}\le s_{it}(\boldsymbol{\gamma}_1)\,|\,\mathbf{s}_{it}\right)}{h^{*}}. \qquad (A13)
\end{aligned}$$

By the definition of a Differential, we have

$$\begin{aligned}
\left.\frac{\partial P\left(q_{it}\le a\,|\,\mathbf{s}_{it}\right)}{\partial a}\right|_{a=s_{it}(\boldsymbol{\gamma}_1)} &\equiv \lim_{h^{*}\to 0}\frac{P\left(q_{it}\le s_{it}(\boldsymbol{\gamma}_1)+h^{*}\,|\,\mathbf{s}_{it}\right)-P\left(q_{it}\le s_{it}(\boldsymbol{\gamma}_1)\,|\,\mathbf{s}_{it}\right)}{h^{*}} \\
&= f_{it|\mathbf{s}}\left(s_{it}(\boldsymbol{\gamma}_1)\,|\,\mathbf{s}_{it}\right). \qquad (A14)
\end{aligned}$$

A similar argument can prove (A11).

Moreover, we have

$$I\left(q_{it} \leq s_{it}(\boldsymbol{\gamma})\right) \leq I\left(q_{it} \leq s_{it}(\boldsymbol{\gamma}_1)\right) \quad \Leftrightarrow \quad s_{it}(\boldsymbol{\gamma}) \leq s_{it}(\boldsymbol{\gamma}_1), \tag{A15}$$

$$I\left(q_{it} \leq s_{it}(\boldsymbol{\gamma})\right) > I\left(q_{it} \leq s_{it}(\boldsymbol{\gamma}_1)\right) \quad \Leftrightarrow \quad s_{it}(\boldsymbol{\gamma}) > s_{it}(\boldsymbol{\gamma}_1), \tag{A16}$$

(A15) and (A16) imply that

$$\operatorname{sgn}\left(I_{it}(\boldsymbol{\gamma}) - I_{it}(\boldsymbol{\gamma}_1)\right) = \operatorname{sgn}\left(s_{it}(\boldsymbol{\gamma}) - s_{it}(\boldsymbol{\gamma}_1)\right), \tag{A17}$$

where $\operatorname{sgn}(x)$ is the sign function, and therefore,

$$\begin{aligned} \left[s_{it}(\boldsymbol{\gamma}) - s_{it}(\boldsymbol{\gamma}_1)\right] \operatorname{sgn}\left(I_{it}(\boldsymbol{\gamma}) - I_{it}(\boldsymbol{\gamma}_1)\right) &= \left[s_{it}(\boldsymbol{\gamma}) - s_{it}(\boldsymbol{\gamma}_1)\right] \operatorname{sgn}\left(s_{it}(\boldsymbol{\gamma}) - s_{it}(\boldsymbol{\gamma}_1)\right) \\ &= \left|s_{it}(\boldsymbol{\gamma}) - s_{it}(\boldsymbol{\gamma}_1)\right|. \end{aligned} \tag{A18}$$

Finally, we can obtain the following result

$$\begin{aligned} \frac{E\left[Z \nabla I_{it}^2(\boldsymbol{\gamma}) \,\middle|\, \mathbf{s}_{it}\right]}{h} &= E\left(Z \mid A_{it}(\boldsymbol{\gamma}), \mathbf{s}_{it}\right) \frac{P\left(A_{it}(\boldsymbol{\gamma}) \mid \mathbf{s}_{it}\right)}{s_{it}(\boldsymbol{\gamma}) - s_{it}(\boldsymbol{\gamma}_1)} \left|\mathbf{s}_{it,1}' \boldsymbol{\omega}\right| \\ &\longrightarrow E\left(Z \mid A_{it}^*(\boldsymbol{\gamma}_1), \mathbf{s}_{it}\right) f_{it|\mathbf{s}}\left(s_{it}(\boldsymbol{\gamma}_1) \mid \mathbf{s}_{it}\right) \left|\mathbf{s}_{it,1}' \boldsymbol{\omega}\right|. \end{aligned} \tag{A19}$$

Taking the expectation over both sides of (A19), we have the desired result. This completes the proof of Lemma 2. □

Lemma 3. *Under Assumption 1, set* $\boldsymbol{\gamma} = \boldsymbol{\gamma}_1 + \boldsymbol{\omega} h$, $\boldsymbol{\omega} \neq \mathbf{0}$, $h > 0$ *and* $r \leq 4$, *, as* $h \to 0$

$$E\left(b_{it}^r(\boldsymbol{\gamma}, \boldsymbol{\gamma}_1)\right) = O(h), \tag{A20}$$

$$E\left(h_{it}^r(\boldsymbol{\gamma}, \boldsymbol{\gamma}_1)\right) = O(h),. \tag{A21}$$

Proof. According to Lemma 2, we obtain the following results

$$\frac{E\left(b_{it}^r(\boldsymbol{\gamma}, \boldsymbol{\gamma}_1)\right)}{h} \longrightarrow E\left(E\left(\|\check{\mathbf{x}}_{it}\|^r \mid q_{it} = s_{it}(\boldsymbol{\gamma}_1), \mathbf{s}_{it}\right) f_{t|\mathbf{s}}\left(s_{it}(\boldsymbol{\gamma}_1) \mid \mathbf{s}_{it}\right) \left|\mathbf{s}_{it,1}' \boldsymbol{\omega}\right|\right), \tag{A22}$$

The proofs of (A20) and (A21) are identical.

□

Lemma 4. *Under Assumption 1, set* $\boldsymbol{\gamma} = \boldsymbol{\gamma}_1 + \boldsymbol{\omega} h$, $\boldsymbol{\omega} \neq \mathbf{0}$, $h > 0$, *as* $h \to 0$

$$E\left|\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[b_{it}^2(\boldsymbol{\gamma}, \boldsymbol{\gamma}_1) - E(b_{it}^2(\boldsymbol{\gamma}, \boldsymbol{\gamma}_1))\right]\right|^2 = O(h), \tag{A23}$$

$$E\left|\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \left[h_{it}^2(\boldsymbol{\gamma}, \boldsymbol{\gamma}_1) - E(h_{it}^2(\boldsymbol{\gamma}, \boldsymbol{\gamma}_1))\right]\right|^2 = O(h). \tag{A24}$$

Proof. Lemma 3.4 of Peligrad (1982) show that for ρ-mixing sequences satisfy Assumption

1, there is a $K' < \infty$ such that

$$
\begin{aligned}
& E\left|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[b_{it}^2(\boldsymbol{\gamma},\boldsymbol{\gamma}_1)-E(b_{it}^2(\boldsymbol{\gamma},\boldsymbol{\gamma}_1))\right]\right|^2 \\
& \quad \leq \quad K'\left\{NT\times E\left[\frac{1}{\sqrt{NT}}\left(b_{it}^2(\boldsymbol{\gamma},\boldsymbol{\gamma}_1)-E(b_{it}^2(\boldsymbol{\gamma},\boldsymbol{\gamma}_1))\right)\right]^2\right. \\
& \qquad\qquad \left.+(NT)^2\left|E\left[\frac{1}{\sqrt{NT}}\left(b_{it}^2(\boldsymbol{\gamma},\boldsymbol{\gamma}_1)-E(b_{it}^2(\boldsymbol{\gamma},\boldsymbol{\gamma}_1))\right)\right]\right|^2\right\} \\
& \quad = \quad K'E\left[b_{it}^2(\boldsymbol{\gamma},\boldsymbol{\gamma}_1)-E(b_{it}^2(\boldsymbol{\gamma},\boldsymbol{\gamma}_1))\right]^2 \\
& \quad = \quad K'\left\{E(b_{it}^4(\boldsymbol{\gamma},\boldsymbol{\gamma}_1))-[E(b_{it}^2(\boldsymbol{\gamma},\boldsymbol{\gamma}_1))]^2\right\} \\
& \quad \leq \quad K'E(b_{it}^4(\boldsymbol{\gamma},\boldsymbol{\gamma}_1)) \\
& \quad \leq \quad K'O(h). \qquad \text{(A25)}
\end{aligned}
$$

The proofs of (A23) and (A24) are identical. □

Lemma 5. *There are finite constants K_1 and K_2 such that for all $\boldsymbol{\gamma}_1$, $\eta > 0$, and $\delta \geq (NT)^{-1}$, if $\sqrt{NT} \geq K_2/\eta$, then*

$$
P\left(\sup_{\|\boldsymbol{\gamma}-\boldsymbol{\gamma}_1\|<\delta}\left\|J_{NT}(\boldsymbol{\gamma})-J_{NT}(\boldsymbol{\gamma}_1)\right\|>\eta\right)\leq\frac{K_1\delta^2}{\eta^4}. \qquad \text{(A26)}
$$

Proof. Let m be an integer satisfying $NT\delta/2 \leq m \leq NT\delta$, which is possible since $NT\delta \geq 1$. Set $\boldsymbol{\gamma} = \boldsymbol{\gamma}_1 + \boldsymbol{\omega}\delta$ such that $\|\boldsymbol{\omega}\| \leq 1$, $\delta_m = \delta/m$ and $C_{10} = \sup\{\|\bar{\mathbf{c}}_g\| : N, T \in \mathbb{N}\}$. For $l = 1, 2, \ldots, m+1$, set $\boldsymbol{\gamma}_l = \boldsymbol{\gamma}_1 + \boldsymbol{\omega}\delta_m(l-1)$, $h_{it}(\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2) = \|\check{\mathbf{x}}_{it}u_{it}\||I_{it}(\boldsymbol{\gamma}_2) - I_{it}(\boldsymbol{\gamma}_1))|$, $h_{itl} = h_{it}(\boldsymbol{\gamma}_l, \boldsymbol{\gamma}_{l+1})$, $h_{itlj} = h_{it}(\boldsymbol{\gamma}_l, \boldsymbol{\gamma}_j)$ and $H_{NTl} = (NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} h_{itl}$.

Thus, By Lemma 2, there exists a constant $C > 0$ such that

$$
E(h_{itl}^r) \leq C\delta_m, \qquad \text{(A27)}
$$

and

$$
E(h_{itlj}^r) \leq C|l-j|\delta_m. \qquad \text{(A28)}
$$

Next, observe that

$$
\begin{aligned}
\|J_{NT}(\boldsymbol{\gamma})-J_{NT}(\boldsymbol{\gamma}_l)\| \quad & \leq \quad C_{10}\left\|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\check{\mathbf{x}}_{it}u_{it}I_{it}(\boldsymbol{\gamma})-\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\check{\mathbf{x}}_{it}u_{it}I_{it}(\boldsymbol{\gamma}_l)\right\| \\
& = \quad C_{10}\left\|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\check{\mathbf{x}}_{it}u_{it}\left[I_{it}(\boldsymbol{\gamma})-I_{it}(\boldsymbol{\gamma}_l)\right]\right\| \\
& \leq \quad \frac{C_{10}}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}h_{itl}=\sqrt{NT}H_{NTl} \\
& = \quad C_{10}\left[\sqrt{NT}H_{NTl}-\sqrt{NT}E(H_{NTl})+\sqrt{NT}E(H_{NTl})\right] \\
& \leq \quad C_{10}\sqrt{NT}\left|H_{NTl}-E(H_{NTl})\right|+C_{10}\sqrt{NT}E|H_{NTl}|. \qquad \text{(A29)}
\end{aligned}
$$

Therefore, for $\boldsymbol{\gamma}$ between $\boldsymbol{\gamma}_l$ and $\boldsymbol{\gamma}_{l+1}$, we have

$$\begin{aligned}
\|J_{NT}(\boldsymbol{\gamma}) - J_{NT}(\boldsymbol{\gamma}_1)\| &= \|J_{NT}(\boldsymbol{\gamma}) - J_{NT}(\boldsymbol{\gamma}_l) + J_{NT}(\boldsymbol{\gamma}_l) - J_{NT}(\boldsymbol{\gamma}_1)\| \\
&\leq \|J_{NT}(\boldsymbol{\gamma}) - J_{NT}(\boldsymbol{\gamma}_l)\| + \|J_{NT}(\boldsymbol{\gamma}_l) - J_{NT}(\boldsymbol{\gamma}_1)\| \\
&\leq \|J_{NT}(\boldsymbol{\gamma}_l) - J_{NT}(\boldsymbol{\gamma}_1)\| + C_{10}\sqrt{NT}\,|H_{NTl} - E(H_{NTl})| \\
&\quad + C_{10}\sqrt{NT}E|H_{NTl}|.
\end{aligned}$$

Hence,

$$\begin{aligned}
\sup_{\|\boldsymbol{\gamma}-\boldsymbol{\gamma}_1\|<\delta} \|J_{NT}(\boldsymbol{\gamma}) - J_{NT}(\boldsymbol{\gamma}_1)\| &\leq \max_{2\leq l\leq m+1} \|J_{NT}(\boldsymbol{\gamma}_l) - J_{NT}(\boldsymbol{\gamma}_1)\| \\
&\quad + C_{10} \max_{1\leq l\leq m} \sqrt{NT}\,|H_{NTl} - E(H_{NTl})| \\
&\quad + C_{10} \max_{1\leq l\leq m} \sqrt{NT}E|H_{NTl}|. \qquad \text{(A30)}
\end{aligned}$$

Noting that $E(h_{itl}^2) \leq C\delta_m$, (A24), $(NT)^{-1} \leq \delta_m$ and $|l-j|^{1/2} \leq |l-j|$, by Burkholder's inequality (see, e.g., Hall and Heyde (1980, p.23)) , Minkowski's inequality, for any $1 \leq j < l \leq m+1$ and some $C' < \infty$ and $K < \infty$ we have

$$\begin{aligned}
& E\,\|J_{NT}(\boldsymbol{\gamma}_l) - J_{NT}(\boldsymbol{\gamma}_j)\|^4 \\
&= E\left\|\frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T \bar{\mathbf{c}}_g'\check{\mathbf{x}}_{it}u_{it}\left[I_{it}(\boldsymbol{\gamma}_l) - I_{it}(\boldsymbol{\gamma}_j)\right]\right\|^4 \\
&\leq C_{10}^4C'E\left|\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \|\check{\mathbf{x}}_{it}u_{it}\|^2\,|I_{it}(\boldsymbol{\gamma}_l) - I_{it}(\boldsymbol{\gamma}_j)|\right|^2 \\
&\leq C_{10}^4C'E\left|\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T h_{itlj}^2\right|^2 \\
&= C_{10}^4C'E\left|\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T h_{itlj}^2 - \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T E(h_{itlj}^2) + \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T E(h_{itlj}^2)\right|^2 \\
&\leq C_{10}^4C'\left\{\frac{1}{\sqrt{NT}}\left[E\left|\frac{1}{\sqrt{NT}}\sum_{i=1}^N\sum_{t=1}^T\left(h_{itlj}^2 - E(h_{itlj}^2)\right)\right|^2\right]^{1/2} + \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T E(h_{itlj}^2)\right\}^2 \\
&\leq C_{10}^4C'\left(\frac{1}{\sqrt{NT}}\left[K|l-j|\delta_m\right]^{1/2} + C|l-j|\delta_m\right)^2 \\
&\leq C_{10}^4C'\left(\delta_m\sqrt{K}|l-j| + C|l-j|\delta_m\right)^2 \\
&= C_{10}^4C'\left(\sqrt{K}+C\right)^2(l-j)^2\delta_m^2, \qquad \text{(A31)}
\end{aligned}$$

where the first inequality is Burkholder's inequality with $p = 4$, the third inequality is Minkowski's inequality with $p = 2$, the fourth inequality is that

$$\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T E(h_{itlj}^r) \leq \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T C|l-j|\delta_m = C|l-j|\delta_m,$$

and Lemma 4, and the fifth inequality is $(NT)^{-1/2} \leq \delta_m^{1/2}$ and $|l-j|^{1/2} \leq (l-j)$.

Hence, using the bound (A31) and Theorem 12.2 of Billingsley (1968, p.94), we have

$$
\begin{aligned}
& P\left(\max_{2\leq l\leq m+1} \|J_{NT}(\boldsymbol{\gamma}_l) - J_{NT}(\boldsymbol{\gamma}_1)\| > \eta\right) \\
& \quad \leq \frac{K'_{4,2}}{\eta^4} C_{10}^4 C' \left(\sqrt{K} + C\right)^2 ((m+1)-1)^2 \delta_m^2 \\
& \quad = \frac{K'_{4,2}}{\eta^4} C_{10}^4 C' \left(\sqrt{K} + C\right)^2 \delta^2 \\
& \quad \equiv K'' \frac{\delta^2}{\eta^4}. \qquad \text{(A32)}
\end{aligned}
$$

We next consider the second term on the right-hand side of (A30). Lemma 3.6 of Peligrad (1982) shows that there is a $K^* < \infty$ such that

$$
\begin{aligned}
& E\left|\sqrt{NT}\left[H_{NTl} - E(H_{NTl})\right]\right|^4 \\
& \quad = E\left|\sum_{i=1}^N \left[\frac{1}{\sqrt{NT}}\left(h_{iktl} - E(h_{iktl})\right)\right]\right|^4 \\
& \quad \leq K^*\left[\frac{1}{NT} E(h_{itl} - E(h_{itl}))^4 + [E(h_{itl} - E(h_{itl}))^2]^2\right] \\
& \quad \leq K^*\left[\frac{1}{NT}\left[O\left(\delta_m\right) + O\left(\delta_m^2\right) + O\left(\delta_m^3\right) + O\left(\delta_m^4\right)\right] + \left[O\left(\delta_m\right) + O\left(\delta_m^2\right)\right]\right] \\
& \quad \leq K^*\left[\delta_m C \delta_m + C^2 \delta_m^2\right]^2 \\
& \quad = K^*[C + C^2]\delta_m^2. \qquad \text{(A33)}
\end{aligned}
$$

Hence, by Markov's inequality, we have

$$
\begin{aligned}
& P\left(\max_{1\leq l\leq m} \sqrt{NT}\left|H_{NTl} - E(H_{NTl})\right| > \eta/C_{10}\right) \\
& \quad \leq C_{10}^4 E \max_{1\leq l\leq m}\left|\sqrt{NT}\left[H_{NTl} - E(H_{NTl})\right]\right|^4 \Big/ \eta^4 \\
& \quad \leq C_{10}^4 \sum_{l=1}^m E\left|\sqrt{NT}\left[H_{NTl} - E(H_{NTl})\right]\right|^4 \Big/ \eta^4 \\
& \quad \leq C_{10}^4 \sum_{l=1}^m K^*[C + C^2]\delta_m^2 \Big/ \eta^4 = C_{10}^4 \frac{mK^*[C + C^2]\delta_m^2}{\eta^4} \\
& \quad = \frac{C_{10}^4 K^*[C + C^2]\delta_m \delta}{\eta^4} \leq \frac{C_{10}^4 K^*[C + C^2]\delta^2}{\eta^4} \qquad \text{(A34)}
\end{aligned}
$$

in which the final inequality holds because of $\delta_m \leq \delta$.

Finally, noting that for $l = 1, 2, ..., m$, we have

$$\begin{aligned}
C_{10}\sqrt{NT}E(H_{NTl}) &= C_{10}\sqrt{NT}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}E(h_{itl}) \\
&\leq C_{10}\sqrt{NT}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}C\delta_m = C_{10}\sqrt{NT}C\delta_m \\
&\leq \frac{2C_{10}C}{\sqrt{NT}}, \qquad \text{(A35)}
\end{aligned}$$

where the final inequality holds because $NT\delta/2 \leq m \leq NT\delta \Leftrightarrow \delta/m = \delta_m \leq 2/(NT)$. Hence, when $2C_{10}C/\sqrt{NT} \leq \eta$, we have

$$P\left(C_{10}\max_{1\leq l\leq m}\sqrt{NT}E(H_{NTl}) \leq \eta\right) = 1 \quad\Leftrightarrow\quad P\left(C_{10}\max_{1\leq l\leq m}\sqrt{NT}E(H_{NTl}) > \eta\right) = 0. \text{(A36)}$$

Based on (A30), (A32), (A34), and (A36), for $2C_{10}C/\sqrt{NT} \leq \eta$, we have

$$\begin{aligned}
&P\left(\sup_{\|\boldsymbol{\gamma}-\boldsymbol{\gamma}_1\|<\delta}\|J_{NT}(\boldsymbol{\gamma}) - J_{NT}(\boldsymbol{\gamma}_1)\| > 3\eta\right) \\
&\quad\leq P\left(\max_{2\leq l\leq m+1}\|J_{NT}(\boldsymbol{\gamma}_l) - J_{NT}(\boldsymbol{\gamma}_1)\| > \eta\right) \\
&\qquad + P\left(C_{10}\max_{1\leq l\leq m}\sqrt{NT}\,|H_{NTl} - E(H_{NTl})| > \eta\right) \\
&\qquad + P\left(C_{10}\max_{1\leq l\leq m}\sqrt{NT}E(H_{NTl}) > \eta\right),^{1} \\
&\quad\leq K''\frac{\delta^2}{\eta^4} + \frac{C_{10}^4K^*[C + C^2]\delta^2}{\eta^4} + 0 \\
&\quad= \left(K'' + C_{10}^4K^*[C + C^2]\right)\frac{\delta^2}{\eta^4}. \qquad \text{(A37)}
\end{aligned}$$

Letting $3\eta \equiv \eta'$, we have $6C_{10}C/\sqrt{NT} \leq \eta'$ and

$$P\left(\sup_{\|\boldsymbol{\gamma}-\boldsymbol{\gamma}_1\|<\delta}\|J_{NT}(\boldsymbol{\gamma}) - J_{NT}(\boldsymbol{\gamma}_1)\| > \eta'\right) \leq 3^4\left(K'' + C_{10}^4K^*[C + C^2]\right)\frac{\delta^2}{\eta'^4}. \qquad \text{(A38)}$$

Setting $K_1' = 3^4[K'' + C_{10}^4K^*(C + C^2)]$ and $K_2' = 6C_{10}C$, we can obtain the result. This completes the proof of Lemma 5.

□

Lemma 6. *Under Assumption 1, $J_{NT}(\boldsymbol{\gamma}) \to_d J(\boldsymbol{\gamma})$, a mean-zero Gaussian process with almost surely continuous sample paths.*

Proof. For each $\boldsymbol{\gamma}$, $\check{\mathbf{x}}_{it}u_{it}I_{it}(\boldsymbol{\gamma})$ is a martingale difference sequence, so $J_{NT}(\boldsymbol{\gamma})$ converges pointwise to a Gaussian distribution by a central limit theorem. This can be extended to any

[1] Consider these events $A = \{|X + Y| > 2\eta\}$, $B = \{|X| > \eta\}$ and $C = \{|Y| > \eta\}$, so we know that A implies $B \cup C$ (it can be proof by contradiction), then $A \subseteq B \cup C$ and $P(A) \leq P(B \cup C) \leq P(B) + P(C)$ by property of Probability Axioms.

finite collection of $\boldsymbol{\gamma}$ to yield the convergence of the finite dimensional distributions.

Fix $\varepsilon > 0$ and $\eta > 0$. Set $\delta = \varepsilon\eta^4/K_1$ and $\overline{NT} = \max\{\delta^{-1}, K_2^2/\eta^2\}$, where K_1 and K_2 are defined in Lemma 5. Then by Lemma 5, for any $\boldsymbol{\gamma}_1$, if $NT \geq \overline{NT}$,

$$P\left(\sup_{\|\boldsymbol{\gamma}-\boldsymbol{\gamma}_1\|<\delta} \|J_{NT}(\boldsymbol{\gamma}) - J_{NT}(\boldsymbol{\gamma}_1)\| > \eta\right) \leq \frac{K_1\delta^2}{\eta^4} = \delta\varepsilon. \tag{A39}$$

This establish the conditions for Theorem 15.5 of Billingsley (1968). It follows that $J_{NT}(\boldsymbol{\gamma})$ is tight, so $J_{NT}(\boldsymbol{\gamma}) \to_d J(\boldsymbol{\gamma})$.

□

Lemma 7. *Under Assumption 1, for any fixed $\boldsymbol{\gamma} \in \boldsymbol{\Gamma}$,*

$$(NT)^{\alpha}\left[\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}) - \boldsymbol{\phi}^0\right] \xrightarrow{p} \mathscr{S}^{-1}(\boldsymbol{\phi}_c, \boldsymbol{\gamma})\mathscr{A}(\boldsymbol{\phi}_c, \boldsymbol{\gamma}), \tag{A40}$$

where $\boldsymbol{\phi}_c = (\phi_c', \phi_c')'$ and

$$\begin{aligned}
\mathscr{A}(\boldsymbol{\phi}, \boldsymbol{\gamma}) &= \frac{\partial \boldsymbol{G}'(\boldsymbol{\phi})}{\partial \boldsymbol{\phi}}\left(\begin{bmatrix} \mathbf{M}(\boldsymbol{\gamma}) \\ \mathbf{M}(\boldsymbol{\gamma}) \end{bmatrix} - \begin{bmatrix} \mathbf{M}(\boldsymbol{\gamma}^0) \\ \mathbf{M}(\boldsymbol{\gamma}, \boldsymbol{\gamma}^0) \end{bmatrix}\right)\mathbf{c}_g, \\
\mathscr{S}(\boldsymbol{\phi}, \boldsymbol{\gamma}) &= \left[\frac{\partial \boldsymbol{G}(\boldsymbol{\phi})}{\partial \boldsymbol{\phi}'}\right]'\mathbf{M}^*(\boldsymbol{\gamma})\left[\frac{\partial \boldsymbol{G}(\boldsymbol{\phi})}{\partial \boldsymbol{\phi}'}\right].
\end{aligned} \tag{A41}$$

Moreover, (A40) can be rewritten as

$$(NT)^{\alpha}\left[\hat{\phi}_1(\boldsymbol{\gamma}) - \phi_1^0\right] \xrightarrow{p} \left[\boldsymbol{g}_c'\mathbf{M}(\boldsymbol{\gamma})\boldsymbol{g}_c\right]^{-1}\boldsymbol{g}_c'\left[\mathbf{M}(\boldsymbol{\gamma}) - \mathbf{M}(\boldsymbol{\gamma}, \boldsymbol{\gamma}^0)\right]\mathbf{c}_g, \tag{A42}$$

$$(NT)^{\alpha}\left[\hat{\phi}_2(\boldsymbol{\gamma}) - \phi_2^0\right] \xrightarrow{p} \left[\boldsymbol{g}_c'(\mathbf{M} - \mathbf{M}(\boldsymbol{\gamma}))\boldsymbol{g}_c\right]^{-1}\boldsymbol{g}_c'\left[\mathbf{M}(\boldsymbol{\gamma}, \boldsymbol{\gamma}^0) - \mathbf{M}(\boldsymbol{\gamma}^0)\right]\mathbf{c}_g, \tag{A43}$$

where $\boldsymbol{g}_c = \partial \boldsymbol{g}(\phi)/\partial \phi'|_{\phi=\phi_c}$.

Proof. Note that by Lemma 1-6, we have the following results

$$\begin{aligned}
\check{\mathbf{X}}_{\boldsymbol{\gamma}}^{*\prime}\check{\mathbf{X}}_0/(NT) &= \begin{bmatrix} \check{\mathbf{X}}'\check{\mathbf{X}}_0/(NT) \\ \check{\mathbf{X}}_{\boldsymbol{\gamma}}'\check{\mathbf{X}}_0/(NT) \end{bmatrix} \xrightarrow{p} \begin{bmatrix} \mathbf{M}(\boldsymbol{\gamma}^0) \\ \mathbf{M}(\boldsymbol{\gamma}, \boldsymbol{\gamma}^0) \end{bmatrix}, \\
\check{\mathbf{X}}_{\boldsymbol{\gamma}}^{*\prime}\check{\mathbf{X}}_{\boldsymbol{\gamma}}/(NT) &= \begin{bmatrix} \check{\mathbf{X}}'\check{\mathbf{X}}_{\boldsymbol{\gamma}}/(NT) \\ \check{\mathbf{X}}_{\boldsymbol{\gamma}}'\check{\mathbf{X}}_{\boldsymbol{\gamma}}/(NT) \end{bmatrix} \xrightarrow{p} \begin{bmatrix} \mathbf{M}(\boldsymbol{\gamma}) \\ \mathbf{M}(\boldsymbol{\gamma}) \end{bmatrix}, \\
\check{\mathbf{X}}_{\boldsymbol{\gamma}}^{*\prime}\check{\mathbf{X}}_{\boldsymbol{\gamma}}^{*}/(NT) &= \begin{bmatrix} \check{\mathbf{X}}'\check{\mathbf{X}}/(NT) & \check{\mathbf{X}}'\check{\mathbf{X}}_{\boldsymbol{\gamma}}/(NT) \\ \check{\mathbf{X}}_{\boldsymbol{\gamma}}'\check{\mathbf{X}}/(NT) & \check{\mathbf{X}}_{\boldsymbol{\gamma}}'\check{\mathbf{X}}_{\boldsymbol{\gamma}}/(NT) \end{bmatrix} \xrightarrow{p} \begin{bmatrix} \mathbf{M} & \mathbf{M}(\boldsymbol{\gamma}) \\ \mathbf{M}(\boldsymbol{\gamma}) & \mathbf{M}(\boldsymbol{\gamma}) \end{bmatrix} \equiv \mathbf{M}^*(\boldsymbol{\gamma}), \\
(NT)^{-1/2}\check{\mathbf{X}}_{\boldsymbol{\gamma}}^{*\prime}\mathbf{u} &\xrightarrow{d} N\left(\mathbf{0}, \boldsymbol{A}'\boldsymbol{\Omega}^*(\boldsymbol{\gamma})\boldsymbol{A}\right).
\end{aligned} \tag{A44}$$

First, we derive the limiting distribution of $\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma})$ by Theorem 10.1 of Wooldridge (1994). Aussumptions 1.11-1.12 ensure that conditions (i) and (ii) of Theorem 10.1 of Wooldridge (1994) are satisfied. We next verify condition (iv) and then condition (iii).

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

$$\begin{aligned}
\frac{\partial \widetilde{SSR}_{NT}(\boldsymbol{\phi},\boldsymbol{\gamma})}{\partial \boldsymbol{\phi}} &= -2\frac{\partial \boldsymbol{G}'(\boldsymbol{\phi})}{\partial \boldsymbol{\phi}} \check{\mathbf{X}}_{\boldsymbol{\gamma}}^{*\prime} \left[\mathbf{Y} - \check{\mathbf{X}}_{\boldsymbol{\gamma}}^{*} \boldsymbol{G}(\boldsymbol{\phi})\right] \\
&= 2\mathbf{K}_{NT}(\boldsymbol{\phi},\boldsymbol{\gamma}), \qquad \text{(A45)}\\
\frac{\partial^2 \widetilde{SSR}_{NT}(\boldsymbol{\phi},\boldsymbol{\gamma})}{\partial \boldsymbol{\phi}\partial \boldsymbol{\phi}'} &= 2\frac{\partial \mathbf{K}_{NT}(\boldsymbol{\phi},\boldsymbol{\gamma})}{\partial \boldsymbol{\phi}'} \\
&= 2\left[\check{\mathbf{X}}_{\boldsymbol{\gamma}}^{*} \frac{\partial \boldsymbol{G}(\boldsymbol{\phi})}{\partial \boldsymbol{\phi}'}\right]' \left[\check{\mathbf{X}}_{\boldsymbol{\gamma}}^{*} \frac{\partial \boldsymbol{G}(\boldsymbol{\phi})}{\partial \boldsymbol{\phi}'}\right] \\
&\quad -2\left[\left(\left[\mathbf{Y} - \check{\mathbf{X}}_{\boldsymbol{\gamma}}^{*} \boldsymbol{G}(\boldsymbol{\phi})\right]' \check{\mathbf{X}}_{\boldsymbol{\gamma}}^{*}\right) \otimes \boldsymbol{I}\right] \frac{\partial}{\partial \boldsymbol{\phi}'} \operatorname{vec}\left(\frac{\partial \boldsymbol{G}'(\boldsymbol{\phi})}{\partial \boldsymbol{\phi}}\right) \\
&= 2\mathbf{H}_{NT}(\boldsymbol{\phi},\boldsymbol{\gamma}), \qquad \text{(A46)}
\end{aligned}$$

where $\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}) = \arg\min_{\boldsymbol{\phi}\in\boldsymbol{\Phi}} \widetilde{SSR}_{NT}(\boldsymbol{\phi},\boldsymbol{\gamma})$.

Let $\mathbb{D}_{NT} = \boldsymbol{I}(NT)^{1/2}$, $\boldsymbol{I} = \operatorname{diag}(1,1,\ldots,1)$, $\sqrt{a_{NT}} = (NT)^{1/2-\alpha}$, $\mathbb{B}_{NT} = a_{NT}^{-1/2}\mathbb{D}_{NT} = \boldsymbol{I}(NT)^{\alpha}$, note that

$$\begin{aligned}
\mathbf{Y} - \check{\mathbf{X}}_{\boldsymbol{\gamma}}^{*} \boldsymbol{G}(\boldsymbol{\phi}^0) &= \mathbf{Y} - \left\{\check{\mathbf{X}}\boldsymbol{g}(\phi_2^0) + \check{\mathbf{X}}_{\boldsymbol{\gamma}}\left[\boldsymbol{g}(\phi_1^0) - \boldsymbol{g}(\phi_2^0)\right]\right\} \\
&= \mathbf{u} - \left[\check{\mathbf{X}}_{\boldsymbol{\gamma}} - \check{\mathbf{X}}_0\right]\left[\boldsymbol{g}(\phi_1^0) - \boldsymbol{g}(\phi_2^0)\right] \\
&= \mathbf{u} - \left[\check{\mathbf{X}}_{\boldsymbol{\gamma}} - \check{\mathbf{X}}_0\right]\frac{\partial \boldsymbol{g}(\bar{\phi})}{\partial \phi'}\left[\phi_1^0 - \phi_2^0\right] \\
&= \mathbf{u} - \left[\check{\mathbf{X}}_{\boldsymbol{\gamma}} - \check{\mathbf{X}}_0\right]\frac{\partial \boldsymbol{g}(\bar{\phi})}{\partial \phi'}\left(\mathbf{c}_1 - \mathbf{c}_2\right)(NT)^{-\alpha} \\
&= \mathbf{u} - \left[\check{\mathbf{X}}_{\boldsymbol{\gamma}} - \check{\mathbf{X}}_0\right]\bar{\mathbf{c}}_g\,(NT)^{-\alpha} \qquad \text{(A47)}
\end{aligned}$$

where $\bar{\phi}$ is a value between ϕ_1^0 and ϕ_2^0. Therefore, we have

$$\begin{aligned}
&a_{NT}^{-1/2}\mathbb{D}_{NT}^{-1}\mathbf{K}_{NT}(\boldsymbol{\phi}^0,\boldsymbol{\gamma}) \\
&\quad= T^{\alpha-1}\mathbf{K}_{NT}(\boldsymbol{\phi}^0,\boldsymbol{\gamma}) \\
&\quad= -\frac{\partial \boldsymbol{G}'(\boldsymbol{\phi}^0)}{\partial \boldsymbol{\phi}}\left\{\frac{1}{\sqrt{a_{NT}}}\frac{1}{\sqrt{NT}}\check{\mathbf{X}}_{\boldsymbol{\gamma}}^{*\prime}\mathbf{u} - \left[\frac{1}{NT}\check{\mathbf{X}}_{\boldsymbol{\gamma}}^{*\prime}\check{\mathbf{X}}_{\boldsymbol{\gamma}} - \frac{1}{NT}\check{\mathbf{X}}_{\boldsymbol{\gamma}}^{*\prime}\check{\mathbf{X}}_0\right]\bar{\mathbf{c}}_g\right\} \\
&\quad\equiv -\mathscr{A}_{NT}(\boldsymbol{\phi}^0,\boldsymbol{\gamma}) \\
&\quad\xrightarrow{p} -\frac{\partial \boldsymbol{G}'(\boldsymbol{\phi}_c)}{\partial \boldsymbol{\phi}}\left(\begin{bmatrix}\mathbf{M}(\boldsymbol{\gamma})\\ \mathbf{M}(\boldsymbol{\gamma})\end{bmatrix} - \begin{bmatrix}\mathbf{M}(\boldsymbol{\gamma}^0)\\ \mathbf{M}(\boldsymbol{\gamma},\boldsymbol{\gamma}^0)\end{bmatrix}\right)\mathbf{c}_g, \qquad \text{(A48)}
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{D}_{NT}^{-1}\mathbf{H}_{NT}(\boldsymbol{\phi}^0,\boldsymbol{\gamma})\mathbb{D}_{NT}^{-1} \\
&\quad= \frac{1}{NT}\mathbf{H}_{NT}(\boldsymbol{\phi}^0,\boldsymbol{\gamma}) \\
&\quad= \left[\frac{\partial \boldsymbol{G}(\boldsymbol{\phi}^0)}{\partial \boldsymbol{\phi}'}\right]' \frac{1}{NT}\check{\mathbf{X}}_{\gamma}^{*\prime}\check{\mathbf{X}}_{\gamma}^{*} \left[\frac{\partial \boldsymbol{G}(\boldsymbol{\phi}^0)}{\partial \boldsymbol{\phi}'}\right] \\
&\qquad - \left[\left(\frac{1}{NT}\mathbf{u}'\check{\mathbf{X}}_{\gamma}^{*} - (NT)^{-\alpha}\bar{\mathbf{c}}_g'\left[\frac{1}{NT}\check{\mathbf{X}}_{\gamma}'\check{\mathbf{X}}_{\gamma}^{*} - \frac{1}{NT}\check{\mathbf{X}}_{0}'\check{\mathbf{X}}_{\gamma}^{*}\right]\right)\otimes \boldsymbol{I}\right] \\
&\qquad \times \frac{\partial}{\partial \boldsymbol{\phi}'}\mathrm{vec}\left(\frac{\partial \boldsymbol{G}'(\boldsymbol{\phi}^0)}{\partial \boldsymbol{\phi}}\right) \\
&\quad\equiv \mathscr{S}_{NT}(\boldsymbol{\phi}^0,\boldsymbol{\gamma}) \\
&\quad\xrightarrow{p} \left[\frac{\partial \boldsymbol{G}(\boldsymbol{\phi}_c)}{\partial \boldsymbol{\phi}'}\right]' \mathbf{M}(\boldsymbol{\gamma}) \left[\frac{\partial \boldsymbol{G}(\boldsymbol{\phi}_c)}{\partial \boldsymbol{\phi}'}\right]. \qquad \text{(A49)}
\end{aligned}$$

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

$$\max_{\boldsymbol{\phi}\in\boldsymbol{\Phi}_{NT}} \left\|\mathbb{C}_{NT}^{-1}\left[\mathbf{H}_{NT}(\boldsymbol{\phi},\boldsymbol{\gamma}) - \mathbf{H}_{NT}(\boldsymbol{\phi}^0,\boldsymbol{\gamma})\right]\mathbb{C}_{NT}^{-1}\right\| = o_p(1), \qquad \text{(A50)}$$

where $\boldsymbol{\Phi}_{NT} = \{\boldsymbol{\phi}\in\boldsymbol{\Phi} : \|\mathbb{A}_{NT}(\boldsymbol{\phi}-\boldsymbol{\phi}^0)\| \le 1\}$ is a sequence of shrinking neighborhoods of $\boldsymbol{\phi}^0$.

To establish condition (iii), we let

$$\begin{aligned}
\mathbb{C}_{NT} &= \mathbb{D}_{NT}^{1-2\delta} = \boldsymbol{I}(NT)^{-1/2+\delta}, &\text{(A51)}\\
\mathbb{A}_{NT} &= \mathbb{B}_{NT}^{1-2\delta} = \boldsymbol{I}(NT)^{\alpha(1-2\delta)}, &\text{(A52)}
\end{aligned}$$

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

$$\boldsymbol{\phi} = \boldsymbol{\phi}^0 + \boldsymbol{\zeta}(NT)^{-\alpha(1-2\delta)}, \qquad \text{(A53)}$$

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

Therefore, by mean value theorem and Assumption 1, we have following results

$$\begin{aligned}
\boldsymbol{G}(\boldsymbol{\phi}) - \boldsymbol{G}(\boldsymbol{\phi}^0) &= O(\|\boldsymbol{\phi}-\boldsymbol{\phi}^0\|) = O((NT)^{-\alpha(1-2\delta)}), &\text{(A54)}\\
\frac{\partial \boldsymbol{G}(\boldsymbol{\phi})}{\partial \boldsymbol{\phi}'} - \frac{\partial \boldsymbol{G}(\boldsymbol{\phi}^0)}{\partial \boldsymbol{\phi}'} &= O(\|\boldsymbol{\phi}-\boldsymbol{\phi}^0\|) = O((NT)^{-\alpha(1-2\delta)}), &\text{(A55)}
\end{aligned}$$

and

$$\frac{\partial}{\partial \boldsymbol{\phi}'}\mathrm{vec}\left(\frac{\boldsymbol{G}'(\boldsymbol{\phi})}{\partial \boldsymbol{\phi}}\right) - \frac{\partial}{\partial \boldsymbol{\phi}'}\mathrm{vec}\left(\frac{\boldsymbol{G}'(\boldsymbol{\phi}^0)}{\partial \boldsymbol{\phi}}\right) = O(\|\boldsymbol{\phi}-\boldsymbol{\phi}^0\|) = O((NT)^{-\alpha(1-2\delta)}). \qquad \text{(A56)}$$

Moreover, we have

$$
\begin{aligned}
&\frac{1}{NT}\check{\mathbf{X}}_{\gamma}^{*\prime}\left[\mathbf{Y}-\check{\mathbf{X}}_{\gamma}^{*}\mathbf{G}(\boldsymbol{\phi})\right] \\
&\quad = \frac{1}{NT}\check{\mathbf{X}}_{\gamma}^{*\prime}\left[\mathbf{u}-\check{\mathbf{X}}_{\gamma}^{*}\left(\mathbf{G}(\boldsymbol{\phi})-\mathbf{G}(\boldsymbol{\phi}^{0})\right)-\nabla\check{\mathbf{X}}_{\boldsymbol{\gamma}}\left(\boldsymbol{g}(\phi_1^0)-\boldsymbol{g}(\phi_2^0)\right)\right] \\
&\quad = \frac{1}{NT}\check{\mathbf{X}}_{\gamma}^{*\prime}\left[\mathbf{u}-\check{\mathbf{X}}_{\gamma}^{*}\left(\mathbf{G}(\boldsymbol{\phi})-\mathbf{G}(\boldsymbol{\phi}^{0})\right)-\nabla\check{\mathbf{X}}_{\boldsymbol{\gamma}}\frac{\partial\boldsymbol{g}(\bar{\phi})}{\partial\phi'}\mathbf{c}(NT)^{-\alpha}\right] \\
&\quad = O_p((NT)^{-1/2})+O_p(1)O((NT)^{-\alpha(1-2\delta)})+O_p(1)O(1)O((NT)^{-\alpha}) \\
&\quad = o_p(1),
\end{aligned}
\tag{A57}
$$

and

$$
\begin{aligned}
&\frac{1}{T}\check{\mathbf{X}}_{\gamma}^{*\prime}\left[\mathbf{Y}-\check{\mathbf{X}}_{\gamma}^{*}\mathbf{G}(\boldsymbol{\phi})\right]-\frac{1}{T}\check{\mathbf{X}}_{\gamma}^{*\prime}\left[\mathbf{Y}-\check{\mathbf{X}}_{\gamma}^{*}\mathbf{G}(\boldsymbol{\phi}^{0})\right] \\
&\quad = \frac{1}{T}\check{\mathbf{X}}_{\gamma}^{*\prime}\check{\mathbf{X}}_{\gamma}^{*}\left[\boldsymbol{G}(\boldsymbol{\phi})-\boldsymbol{G}(\boldsymbol{\phi}^{0})\right] \\
&\quad = O_p(1)O((NT)^{-\alpha(1-2\delta)})=o_p(1),
\end{aligned}
\tag{A58}
$$

where $0<\delta<1/2$ and $0<\alpha<1/2 \Leftrightarrow 0<\alpha(1-2\delta)$ is used.

Consequently , we have

$$
\begin{aligned}
&\mathbb{C}_{NT}^{-1}\left[\mathbf{H}_{NT}(\boldsymbol{\phi},\boldsymbol{\gamma})-\mathbf{H}_{NT}(\boldsymbol{\phi}^{0},\boldsymbol{\gamma})\right]\mathbb{C}_{NT}^{-1} \\
&\quad = (NT)^{2\delta}\left[\frac{\partial\boldsymbol{G}(\boldsymbol{\phi})}{\partial\boldsymbol{\phi}'}-\frac{\partial\boldsymbol{G}(\boldsymbol{\phi}^{0})}{\partial\boldsymbol{\phi}'}\right]'\frac{1}{NT}\check{\mathbf{X}}_{\gamma}^{*\prime}\check{\mathbf{X}}_{\gamma}^{*}\frac{\partial\boldsymbol{G}(\boldsymbol{\phi})}{\partial\boldsymbol{\phi}'} \\
&\qquad +(NT)^{2\delta}\frac{\partial\boldsymbol{G}'(\boldsymbol{\phi}^{0})}{\partial\boldsymbol{\phi}}\frac{1}{NT}\check{\mathbf{X}}_{\gamma}^{*\prime}\check{\mathbf{X}}_{\gamma}^{*}\left[\frac{\partial\boldsymbol{G}(\boldsymbol{\phi})}{\partial\boldsymbol{\phi}'}-\frac{\partial\boldsymbol{G}(\boldsymbol{\phi}^{0})}{\partial\boldsymbol{\phi}'}\right] \\
&\qquad -(NT)^{2\delta}\left[\left(\frac{1}{NT}\left[\mathbf{Y}-\check{\mathbf{X}}_{\gamma}^{*}\boldsymbol{G}(\boldsymbol{\phi})\right]'\check{\mathbf{X}}_{\gamma}^{*}\right)\otimes\boldsymbol{I}\right] \\
&\qquad \times\left[\frac{\partial}{\partial\boldsymbol{\phi}'}\mathrm{vec}\left(\frac{\partial\boldsymbol{G}'(\boldsymbol{\phi})}{\partial\boldsymbol{\phi}}\right)-\frac{\partial}{\partial\boldsymbol{\phi}'}\mathrm{vec}\left(\frac{\partial\boldsymbol{G}'(\boldsymbol{\phi}^{0})}{\partial\boldsymbol{\phi}}\right)\right] \\
&\qquad -(NT)^{2\delta}\left[\left(\left[\boldsymbol{G}(\boldsymbol{\phi})-\boldsymbol{G}(\boldsymbol{\phi}^{0})\right]\frac{1}{NT}\check{\mathbf{X}}_{\gamma}^{*\prime}\check{\mathbf{X}}_{\gamma}^{*}\right)\otimes\boldsymbol{I}\right]\frac{\partial}{\partial\boldsymbol{\phi}'}\mathrm{vec}\left(\frac{\partial\boldsymbol{G}'(\boldsymbol{\phi}^{0})}{\partial\boldsymbol{\phi}}\right) \\
&\quad = O((NT)^{2\delta})O((NT)^{-\alpha(1-2\delta)})\left[O_p(1)O(1)+O_p(1)+o_p(1)+o_p(1)\right] \\
&\quad = O((NT)^{2\delta-\alpha(1-2\delta)})O_p(1)=o_p(1),
\end{aligned}
\tag{A59}
$$

where $0<\delta<\min\{\alpha/[2(1+\alpha)],1/2\}\Leftrightarrow 2\delta<\alpha(1-2\delta)$ is used. Hence, condition (iii) is satisfied.

Thus, Theorem 10.1 of Wooldridge (1994) holds for a scalar a_{NT}. Then, we can obtain below result

$$
\begin{aligned}
\mathbb{B}_{NT}\left[\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma})-\boldsymbol{\phi}^{0}\right] &= (NT)^{\alpha}\left[\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma})-\boldsymbol{\phi}^{0}\right] \\
&\xrightarrow{p} \mathscr{S}^{-1}(\boldsymbol{\phi}_c,\boldsymbol{\gamma})\mathscr{A}(\boldsymbol{\phi}_c,\boldsymbol{\gamma}).
\end{aligned}
\tag{A60}
$$

□

Lemma 8. *Under Assumption 1, $\hat{\boldsymbol{\gamma}} \to_p \boldsymbol{\gamma}^0$.*

Proof. Denoting $I_{it}^{-}(\boldsymbol{\gamma}) = I_{it}(\boldsymbol{\gamma})$ and $I_{it}^{+}(\boldsymbol{\gamma}) = 1 - I_{it}^{-}(\boldsymbol{\gamma})$, we have

$$\begin{aligned}
& y_{it} - \check{\mathbf{x}}_{it}'\boldsymbol{g}\left(\hat{\phi}_1(\boldsymbol{\gamma})\right) I_{it}^{-}(\boldsymbol{\gamma}) - \check{\mathbf{x}}_{it}'\boldsymbol{g}\left(\hat{\phi}_2(\boldsymbol{\gamma})\right) I_{it}^{+}(\boldsymbol{\gamma}) \\
&= \left\{e_{1it} - \check{\mathbf{x}}_{it}'\left[\boldsymbol{g}\left(\hat{\phi}_1(\boldsymbol{\gamma})\right) - \boldsymbol{g}\left(\phi_1^0\right)\right]\right\} I_{it}^{-}(\boldsymbol{\gamma}) I_{it}^{-}(\boldsymbol{\gamma}^0) \\
&\quad + \left\{e_{2it} - \check{\mathbf{x}}_{it}'\left[\boldsymbol{g}\left(\hat{\phi}_2(\boldsymbol{\gamma})\right) - \boldsymbol{g}\left(\phi_2^0\right)\right]\right\} I_{it}^{+}(\boldsymbol{\gamma}) I_{it}^{+}(\boldsymbol{\gamma}^0) \\
&\quad + \left\{e_{1it} - \check{\mathbf{x}}_{it}'\left[\boldsymbol{g}\left(\hat{\phi}_2(\boldsymbol{\gamma})\right) - \boldsymbol{g}\left(\phi_1^0\right)\right]\right\} I_{it}^{+}(\boldsymbol{\gamma}) I_{it}^{-}(\boldsymbol{\gamma}^0) \\
&\quad + \left\{e_{2it} - \check{\mathbf{x}}_{it}'\left[\boldsymbol{g}\left(\hat{\phi}_1(\boldsymbol{\gamma})\right) - \boldsymbol{g}\left(\phi_2^0\right)\right]\right\} I_{it}^{-}(\boldsymbol{\gamma}) I_{it}^{+}(\boldsymbol{\gamma}^0).
\end{aligned} \tag{A61}$$

Then, the objective function can be written as

$$\begin{aligned}
& (NT)^{2\alpha-1}\left(\widetilde{SSR}_{NT}\left(\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}), \boldsymbol{\gamma}\right) - \widetilde{SSR}_{NT}(\boldsymbol{\phi}^0, \boldsymbol{\gamma}^0)\right) \\
&= (NT)^{\alpha}\left[\boldsymbol{g}\left(\hat{\phi}_1(\boldsymbol{\gamma})\right) - \boldsymbol{g}\left(\phi_1^0\right)\right]' \mathbf{M}_{NT,1}(\boldsymbol{\gamma}) (NT)^{\alpha}\left[\boldsymbol{g}\left(\hat{\phi}_1(\boldsymbol{\gamma})\right) - \boldsymbol{g}\left(\phi_1^0\right)\right] \\
&\quad + (NT)^{\alpha}\left[\boldsymbol{g}\left(\hat{\phi}_2(\boldsymbol{\gamma})\right) - \boldsymbol{g}\left(\phi_2^0\right)\right]' \mathbf{M}_{NT,2}(\boldsymbol{\gamma}) (NT)^{\alpha}\left[\boldsymbol{g}\left(\hat{\phi}_2(\boldsymbol{\gamma})\right) - \boldsymbol{g}\left(\phi_2^0\right)\right] \\
&\quad + (NT)^{\alpha}\left[\boldsymbol{g}\left(\hat{\phi}_2(\boldsymbol{\gamma})\right) - \boldsymbol{g}\left(\phi_1^0\right)\right]' \mathbf{M}_{NT,3}(\boldsymbol{\gamma}) (NT)^{\alpha}\left[\boldsymbol{g}\left(\hat{\phi}_2(\boldsymbol{\gamma})\right) - \boldsymbol{g}\left(\phi_1^0\right)\right] \\
&\quad + (NT)^{\alpha}\left[\boldsymbol{g}\left(\hat{\phi}_1(\boldsymbol{\gamma})\right) - \boldsymbol{g}\left(\phi_2^0\right)\right]' \mathbf{M}_{NT,4}(\boldsymbol{\gamma}) (NT)^{\alpha}\left[\boldsymbol{g}\left(\hat{\phi}_1(\boldsymbol{\gamma})\right) - \boldsymbol{g}\left(\phi_2^0\right)\right] \\
&\equiv b_{NT,1}(\boldsymbol{\gamma}) + b_{NT,2}(\boldsymbol{\gamma}) + b_{NT,3}(\boldsymbol{\gamma}) + b_{NT,4}(\boldsymbol{\gamma}),
\end{aligned} \tag{A62}$$

where

$$\begin{aligned}
\mathbf{M}_{NT,1}(\boldsymbol{\gamma}) &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \check{\mathbf{x}}_{it}\check{\mathbf{x}}_{it}' I_{it}^{-}(\boldsymbol{\gamma}) I_{it}^{-}(\boldsymbol{\gamma}^0), \\
\mathbf{M}_{NT,1}(\boldsymbol{\gamma}) &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \check{\mathbf{x}}_{it}\check{\mathbf{x}}_{it}' I_{it}^{+}(\boldsymbol{\gamma}) I_{it}^{+}(\boldsymbol{\gamma}^0), \\
\mathbf{M}_{NT,1}(\boldsymbol{\gamma}) &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \check{\mathbf{x}}_{it}\check{\mathbf{x}}_{it}' I_{it}^{+}(\boldsymbol{\gamma}) I_{it}^{-}(\boldsymbol{\gamma}^0), \\
\mathbf{M}_{NT,1}(\boldsymbol{\gamma}) &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \check{\mathbf{x}}_{it}\check{\mathbf{x}}_{it}' I_{it}^{-}(\boldsymbol{\gamma}) I_{it}^{+}(\boldsymbol{\gamma}^0).
\end{aligned}$$

Therefore, by Lemma 7, we have following result

$$
\begin{aligned}
b_{NT,1}(\boldsymbol{\gamma}) &\xrightarrow{p} \mathscr{A}_1'(\boldsymbol{\phi}_c,\boldsymbol{\gamma})\mathscr{S}_1^{-1}(\boldsymbol{\phi}_c,\boldsymbol{\gamma})\boldsymbol{g}_c'\mathbf{M}_{\infty,1}(\boldsymbol{\gamma})\,\boldsymbol{g}_c\mathscr{S}_1^{-1}(\boldsymbol{\phi}_c,\boldsymbol{\gamma})\mathscr{A}_1(\boldsymbol{\phi}_c,\boldsymbol{\gamma}),\\
&\equiv b_1(\boldsymbol{\gamma}), \qquad \text{(A63)}\\
b_{NT,2}(\boldsymbol{\gamma}) &\xrightarrow{p} \mathscr{A}_2'(\boldsymbol{\phi}_c,\boldsymbol{\gamma})\mathscr{S}_2^{-1}(\boldsymbol{\phi}_c,\boldsymbol{\gamma})\boldsymbol{g}_c'\mathbf{M}_{\infty,2}(\boldsymbol{\gamma})\,\boldsymbol{g}_c\mathscr{S}_2^{-1}(\boldsymbol{\phi}_c,\boldsymbol{\gamma})\mathscr{A}_2(\boldsymbol{\phi}_c,\boldsymbol{\gamma}),\\
&\equiv b_2(\boldsymbol{\gamma}), \qquad \text{(A64)}\\
b_{NT,3}(\boldsymbol{\gamma}) &\xrightarrow{p} \left[\mathscr{A}_2'(\boldsymbol{\phi}_c,\boldsymbol{\gamma})\mathscr{S}_2^{-1}(\boldsymbol{\phi}_c,\boldsymbol{\gamma})+(\mathbf{c}_2-\mathbf{c}_1)'\right]\boldsymbol{g}_c'\mathbf{M}_{\infty,3}(\boldsymbol{\gamma})\,\boldsymbol{g}_c\\
&\quad\times\left[\mathscr{S}_2^{-1}(\boldsymbol{\phi}_c,\boldsymbol{\gamma})\mathscr{A}_2(\boldsymbol{\phi}_c,\boldsymbol{\gamma})+(\mathbf{c}_2-\mathbf{c}_1)\right],\\
&\equiv b_3(\boldsymbol{\gamma}), \qquad \text{(A65)}\\
b_{NT,4}(\boldsymbol{\gamma}) &\xrightarrow{p} \left[\mathscr{A}_1'(\boldsymbol{\phi}_c,\boldsymbol{\gamma})\mathscr{S}_1^{-1}(\boldsymbol{\phi}_c,\boldsymbol{\gamma})+(\mathbf{c}_1-\mathbf{c}_2)'\right]\boldsymbol{g}_c'\mathbf{M}_{\infty,4}(\boldsymbol{\gamma})\,\boldsymbol{g}_c\\
&\quad\times\left[\mathscr{S}_1^{-1}(\boldsymbol{\phi}_c,\boldsymbol{\gamma})\mathscr{A}_1(\boldsymbol{\phi}_c,\boldsymbol{\gamma})+(\mathbf{c}_1-\mathbf{c}_2)\right],\\
&\equiv b_4(\boldsymbol{\gamma}), \qquad \text{(A66)}
\end{aligned}
$$

where

$$
\begin{aligned}
\mathscr{A}_1(\boldsymbol{\phi},\boldsymbol{\gamma}) &= \left[\frac{\partial \boldsymbol{g}(\phi_1)}{\partial\phi_1}\right]'\left[\mathbf{M}(\boldsymbol{\gamma})-\mathbf{M}\left(\boldsymbol{\gamma},\boldsymbol{\gamma}^0\right)\right]\mathbf{c}_g,\\
\mathscr{A}_2(\boldsymbol{\phi},\boldsymbol{\gamma}) &= \left[\frac{\partial \boldsymbol{g}(\phi_2)}{\partial\phi_2}\right]'\left[\mathbf{M}\left(\boldsymbol{\gamma},\boldsymbol{\gamma}^0\right)-\mathbf{M}\left(\boldsymbol{\gamma}^0\right)\right]\mathbf{c}_g,\\
\mathscr{S}_1(\boldsymbol{\phi},\boldsymbol{\gamma}) &= \left[\frac{\partial \boldsymbol{g}(\phi_1)}{\partial\phi_1}\right]'\mathbf{M}(\boldsymbol{\gamma})\left[\frac{\partial \boldsymbol{g}(\phi_1)}{\partial\phi_1}\right],\\
\mathscr{S}_2(\boldsymbol{\phi},\boldsymbol{\gamma}) &= \left[\frac{\partial \boldsymbol{g}(\phi_2)}{\partial\phi_2}\right]'(\mathbf{M}-\mathbf{M}(\boldsymbol{\gamma}))\left[\frac{\partial \boldsymbol{g}(\phi_2)}{\partial\phi_2}\right],\\
\mathbf{M}_{\infty,p}(\boldsymbol{\gamma}) &= \lim_{(N,T)\to\infty} E\left(\mathbf{M}_{NT,p}(\boldsymbol{\gamma})\right),\ \text{for } p=1,2,3,4. \qquad \text{(A67)}
\end{aligned}
$$

Assumption 1 guarantees that $\mathbf{M}(\boldsymbol{\gamma})=\mathbf{M}\left(\boldsymbol{\gamma}^0\right)$ and $\mathbf{M}\left(\boldsymbol{\gamma},\boldsymbol{\gamma}^0\right)=\mathbf{M}\left(\boldsymbol{\gamma}^0,\boldsymbol{\gamma}^0\right)$ if and only if $\boldsymbol{\gamma}=\boldsymbol{\gamma}^0$, which means $\mathscr{A}_1(\boldsymbol{\phi},\boldsymbol{\gamma})=\mathscr{A}_2(\boldsymbol{\phi},\boldsymbol{\gamma})=\mathbf{0}$ if and only if $\boldsymbol{\gamma}=\boldsymbol{\gamma}^0$, then $b_1(\boldsymbol{\gamma})>0$ and $b_2(\boldsymbol{\gamma})>0$ if $\boldsymbol{\gamma}\neq\boldsymbol{\gamma}^0$.

It is easily seen that $b_1(\boldsymbol{\gamma})+b_2(\boldsymbol{\gamma})+b_3(\boldsymbol{\gamma})+b_4(\boldsymbol{\gamma})=0$ if and only if $\boldsymbol{\gamma}=\boldsymbol{\gamma}^0$. Since $\boldsymbol{\gamma}$ minimizes $(NT)^{2\alpha-1}[\widetilde{SSR}_{NT}(\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}),\boldsymbol{\gamma})-\widetilde{SSR}_{NT}(\boldsymbol{\phi}^0,\boldsymbol{\gamma}^0)]$, using Theorem 2.1 of Newey and McFadden (1994) we have $\hat{\boldsymbol{\gamma}}\to_p\boldsymbol{\gamma}^0$. This completes the proof of Lemma 8.

□

Lemma 9. *Under Assumption 1, set* $\boldsymbol{\gamma}=\boldsymbol{\gamma}^0+\boldsymbol{\omega}h$, $\boldsymbol{\omega}\neq\mathbf{0}$, $h>0$, *and* $h\to 0$, $(N,T)\to\infty$, *then*

$$
\begin{aligned}
(NT)^{\alpha}\left(\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma})-\boldsymbol{\phi}^0\right) &\xrightarrow{p} \mathbf{0}, \qquad \text{(A68)}\\
\mathscr{T}_{NT}(\boldsymbol{\gamma}) &\xrightarrow{p} \mathbf{0}. \qquad \text{(A69)}
\end{aligned}
$$

Proof. Using (A60), for any $\boldsymbol{\gamma}_1 \in \boldsymbol{\Gamma}$

$$\begin{aligned}(NT)^{\alpha}\left[\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}_1)-\boldsymbol{\phi}^0\right] &= \mathscr{S}_{NT}^{-1}(\boldsymbol{\phi}^0,\boldsymbol{\gamma}_1)\mathscr{A}_{NT}(\boldsymbol{\phi}^0,\boldsymbol{\gamma}_1)+o_p(1)\\ &\xrightarrow{p} \mathscr{S}^{-1}(\boldsymbol{\phi}_c,\boldsymbol{\gamma}_1)\mathscr{A}(\boldsymbol{\phi}_c,\boldsymbol{\gamma}_1). \end{aligned}\tag{A70}$$

Note that $\mathscr{A}(\boldsymbol{\phi}_c,\boldsymbol{\gamma}^0)=\mathbf{0}$ from (A48). Since $\mathscr{S}(\boldsymbol{\phi}_c,\boldsymbol{\gamma})$ and $\mathscr{A}(\boldsymbol{\phi}_c,\boldsymbol{\gamma})$ are continuous, $\mathscr{S}^{-1}(\boldsymbol{\phi}_c,\boldsymbol{\gamma})\mathscr{A}(\boldsymbol{\phi}_c,\boldsymbol{\gamma})$ is continuous in $\boldsymbol{\Gamma}$. By setting $\boldsymbol{\gamma}\to\boldsymbol{\gamma}^0$, we can conclude that

$$(NT)^{\alpha}\left(\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma})-\boldsymbol{\phi}^0\right) \xrightarrow{p} \mathscr{S}^{-1}(\boldsymbol{\phi}_c,\boldsymbol{\gamma}^0)\mathscr{A}(\boldsymbol{\phi}_c,\boldsymbol{\gamma}^0)=\mathscr{S}^{-1}(\boldsymbol{\phi}_c,\boldsymbol{\gamma}^0)\mathbf{0}=\mathbf{0}.\tag{A71}$$

By mean value theorem, we have

$$\begin{aligned}\mathscr{T}_{NT}(\boldsymbol{\gamma}) &= \frac{\partial \boldsymbol{G}\left(\bar{\boldsymbol{\phi}}\right)}{\partial \boldsymbol{\phi}'}(NT)^{\alpha}\left(\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma})-\boldsymbol{\phi}^0\right)\\ &\xrightarrow{p} \frac{\partial \boldsymbol{G}\left(\boldsymbol{\phi}_c\right)}{\partial \boldsymbol{\phi}'}\mathbf{0}=\mathbf{0},\end{aligned}\tag{A72}$$

where $\bar{\boldsymbol{\phi}}$ is a value between $\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma})$ and $\boldsymbol{\phi}^0$. □

Lemma 10. *Under Assumption 1, on a compact set $\boldsymbol{\Psi}=\times_{n=0}^{d}\Psi_n$, we have the following uniform convergence results, set $\boldsymbol{\gamma}=\boldsymbol{\gamma}^0+\boldsymbol{\omega}h$, $\boldsymbol{\omega}\neq\mathbf{0}$, $h>0$, such that $h\to 0$ and $NTh\to\infty$ as $(N,T)\to\infty$, then*

$$G_{NT}(\boldsymbol{\omega},h) \xrightarrow{p} G(\boldsymbol{\omega}),\tag{A73}$$

$$V_{lNT}(\boldsymbol{\omega},h) \xrightarrow{p} V_l(\boldsymbol{\omega}),\tag{A74}$$

where

$$\begin{aligned}G(\boldsymbol{\omega}) &= \mathbf{c}_g' E\Big(\mathbf{D}_{it}(\boldsymbol{\gamma}^0) f_{it|\mathbf{s}}\left(s_{it}(\boldsymbol{\gamma}^0)\,\middle|\,\mathbf{s}_{it}\right)\left|\mathbf{s}_{it,1}'\boldsymbol{\omega}\right|\Big)\mathbf{c}_g,\\ V_1(\boldsymbol{\omega}) &= \mathbf{c}_g' E\Big(\mathbf{V}_{1it}(\boldsymbol{\gamma}^0) f_{it|\mathbf{s}}\left(s_{it}(\boldsymbol{\gamma}^0)\,\middle|\,\mathbf{s}_{it}\right)\left|\mathbf{s}_{it,1}'\boldsymbol{\omega}\right| I\left(\mathbf{s}_{it,1}'\boldsymbol{\omega}\le 0\right)\Big)\mathbf{c}_g,\\ V_2(\boldsymbol{\omega}) &= \mathbf{c}_g' E\Big(\mathbf{V}_{2it}(\boldsymbol{\gamma}^0) f_{it|\mathbf{s}}\left(s_{it}(\boldsymbol{\gamma}^0)\,\middle|\,\mathbf{s}_{it}\right)\left|\mathbf{s}_{it,1}'\boldsymbol{\omega}\right| I\left(\mathbf{s}_{it,1}'\boldsymbol{\omega}> 0\right)\Big)\mathbf{c}_g,\end{aligned}$$

in which

$$\begin{aligned}\mathbf{D}_{it}(\boldsymbol{\gamma}) &= E\left(\check{\mathbf{x}}_{it}\check{\mathbf{x}}_{it}'\,\middle|\,q_{it}=s_{it}(\boldsymbol{\gamma}),\mathbf{s}_{it}\right),\\ \mathbf{V}_{lit}(\boldsymbol{\gamma}) &= E\left(\check{\mathbf{x}}_{it}\check{\mathbf{x}}_{it}' e_{lit}^2\,\middle|\,q_{it}=s_{it}(\boldsymbol{\gamma}),\mathbf{s}_{it}\right).\end{aligned}\tag{A75}$$

Proof. By Lemma 2 and (A18), we can obtain the following result

$$\begin{aligned} & E\left(G_{NT}(\boldsymbol{\omega}, h)\right) \\ = \quad & E\left(\frac{1}{NTh}\sum_{i=1}^{N}\sum_{t=1}^{T}\bar{\mathbf{c}}_g'\check{\mathbf{x}}_{it}\check{\mathbf{x}}_{it}'\bar{\mathbf{c}}_g\left|I_{it}(\boldsymbol{\gamma}) - I_{it}(\boldsymbol{\gamma}^0)\right|\right) \\ = \quad & \bar{\mathbf{c}}_g'\frac{E\left(\check{\mathbf{x}}_{it}\check{\mathbf{x}}_{it}'\nabla I_{it}^2(\boldsymbol{\gamma})\right)}{h}\bar{\mathbf{c}}_g \\ \longrightarrow \quad & G(\boldsymbol{\omega}). \end{aligned} \tag{A76}$$

Next, By Lemma 4, there exists a constant $K > 0$ such that

$$\begin{aligned} & E\left\|G_{NT}(\boldsymbol{\omega}, h) - E(G_{NT}(\boldsymbol{\omega}, h))\right\|^2 \\ \leq \quad & \left\|\bar{\mathbf{c}}_g\right\|^2 E\left|\frac{1}{NTh}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\|\check{\mathbf{x}}_{it}\right\|^2\left|\nabla I_{it}(\boldsymbol{\gamma})\right| - E\Big(\left\|\check{\mathbf{x}}_{it}\right\|^2\left|\nabla I_{it}(\boldsymbol{\gamma})\right|\Big)\right|^2 \\ = \quad & \frac{1}{NTh^2}\left\|\bar{\mathbf{c}}_g\right\|^2 E\left|\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}b_{it}^2(\boldsymbol{\gamma}^0, \boldsymbol{\gamma}) - E(b_{it}^2(\boldsymbol{\gamma}^0, \boldsymbol{\gamma}))\right|^2 \\ \leq \quad & \frac{1}{NTh^2}\left\|\bar{\mathbf{c}}_g\right\|^2 Kh \\ = \quad & \frac{1}{NTh}\left\|\bar{\mathbf{c}}_g\right\|^2 K \\ \longrightarrow \quad & 0. \end{aligned} \tag{A77}$$

The inequality uses Lemma 4. Furthermore, by Markov's inequality, we have

$$\begin{aligned} & P\left(\left\|G_{NT}(\boldsymbol{\omega}, h) - E(G_{NT}(\boldsymbol{\omega}, h))\right\| > \varepsilon\right) \\ \leq \quad & \frac{E\left\|G_{NT}(\boldsymbol{\omega}, h) - E(G_{NT}(\boldsymbol{\omega}, h))\right\|^2}{\varepsilon^2} \longrightarrow 0. \end{aligned} \tag{A78}$$

The proof for (A74) is similar on (A73). This completes the proof of Lemma 10.

□

Lemma 11. *Under Assumption 1, on any compact set $\boldsymbol{\Psi} = \times_{n=0}^{d}\Psi_n$, set $\boldsymbol{\gamma} = \boldsymbol{\gamma}^0 + \boldsymbol{\omega}h$, $\boldsymbol{\omega} \neq \mathbf{0}$, $h > 0$, such that $h \to 0$ and $NTh \to \infty$ as $(N, T) \to \infty$, then*

$$R_{NT}(\boldsymbol{\omega}, h) \xrightarrow{d} R(\boldsymbol{\omega}), \tag{A79}$$

where $R(\boldsymbol{\omega}) = R_1(\boldsymbol{\omega}) + R_2(\boldsymbol{\omega})$, $R_l(\boldsymbol{\omega})$ is a Gaussian process with a positive variance $V_l(\boldsymbol{\omega})$ when $\boldsymbol{\omega} \neq \mathbf{0}$, which $R_1(\boldsymbol{\omega})$ and $R_2(\boldsymbol{\omega})$ are independent.

Proof. Denote

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

where $I_{1it}(\gamma) = -I(s_{it}(\gamma) < q_{it} \leq s_{it}(\gamma^0))$ and $I_{2it}(\gamma) = I(s_{it}(\gamma^0) < q_{it} \leq s_{it}(\gamma))$. According to the definition of u_{it}, we have the following result

$$u_{it}\left(I_{it}(\gamma) - I_{it}(\gamma^0)\right) = \begin{cases} e_{1it}I_{1it}(\gamma) & , \text{ if } s_{it}(\gamma) \leq s_{it}(\gamma^0) \\ e_{2it}I_{2it}(\gamma) & , \text{ if } s_{it}(\gamma) > s_{it}(\gamma^0) \end{cases}, \tag{A80}$$

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

First, we prove that $E(R^2_{lNT}(\boldsymbol{\omega})) \to_p V_l(\boldsymbol{\omega})$. Note that by Lemma 2 and Lemma 10, we have

$$\begin{aligned} E(R^2_{lN}(\boldsymbol{\omega}, h)) &= \frac{1}{NTh}E\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\bar{\mathbf{c}}'_g\check{\mathbf{x}}_{it}e_{lit}I_{lit}(\gamma)\right)^2 \\ &= \frac{1}{NTh}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{j=1}^{N}\sum_{k=1}^{T}\bar{\mathbf{c}}'_g E\left(\check{\mathbf{x}}_{it}\check{\mathbf{x}}'_{jk}e_{lit}e_{ljk}I_{lit}(\gamma)I_{ljk}(\gamma)\right)\bar{\mathbf{c}}_g \\ &= \frac{1}{NTh}\sum_{i=1}^{N}\sum_{t=1}^{T}\bar{\mathbf{c}}'_g E\left(\check{\mathbf{x}}_{it}\check{\mathbf{x}}'_{it}e^2_{lit}|I_{lit}(\gamma)|\right)\bar{\mathbf{c}}_g \\ &= E\left(V_{lNT}(\boldsymbol{\omega}, h)\right) \\ &= \bar{\mathbf{c}}'_g\frac{E\left(\check{\mathbf{x}}_{it}\check{\mathbf{x}}'_{it}e^2_{lit}|I_{lit}(\gamma)|\right)}{h}\bar{\mathbf{c}}_g \\ &\longrightarrow V_l(\boldsymbol{\omega}), \end{aligned} \tag{A81}$$

and

$$\begin{aligned} &E(R_{1NT}(\boldsymbol{\omega}, h)R_{2NT}(\boldsymbol{\omega}, h)) \\ &= \frac{1}{NTh}E\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\bar{\mathbf{c}}'_g\check{\mathbf{x}}_{it}e_{1it}I_{1it}(\gamma)\right)E\left(\sum_{j=1}^{N}\sum_{k=1}^{T}\bar{\mathbf{c}}'_g\check{\mathbf{x}}_{jk}e_{2jk}I_{2jk}(\gamma)\right) \\ &= \frac{1}{NTh}\sum_{i\neq j}\sum_{t\neq k}\bar{\mathbf{c}}'_g E\left(\check{\mathbf{x}}_{it}\check{\mathbf{x}}'_{jk}e_{1it}e_{2jk}I_{1it}(\gamma)I_{2jk}(\gamma)\right)\bar{\mathbf{c}}_g \\ &= 0, \end{aligned} \tag{A82}$$

where $I_{1it}(\gamma)I_{2it}(\gamma) = 0$ is used.

Under Assumption 1, $\nabla\check{\mathbf{x}}_{it}u_{it}$ is a martingale difference sequence, so $R_{NT}(\gamma, h)$ converges pointwise to a Gaussian distribution with variance $V_1(\boldsymbol{\omega})+V_2(\boldsymbol{\omega})$ by the central limit theorem.

Next, given lemma 5, the proof of the tightness of $R_{NT}(\boldsymbol{\omega}, h)$ is same as Lemma A.11 of Hansen (2000). This completes the proof of Lemma 11.

□

Lemma 12. *Under Assumption 1, we have $a_{NT}(\hat{\gamma} - \gamma^0) = O_p(1)$.*

Proof. To prove Lemma 12, we need to prove that, for some $\bar{\omega} > 0$, we have

$$\lim_{(N,T)\to\infty} P\left(\|\gamma - \gamma^0\| \leq \frac{\bar{\omega}}{a_{NT}}\right) = 1. \tag{A83}$$

For any $B>0$, define $V_B=\{\boldsymbol{\gamma}:\|\boldsymbol{\gamma}-\boldsymbol{\gamma}^0\|\le B\}$. Then when the sample size NT is large enough, we have $\bar{\omega}/a_{NT}\le B$. By Lemma 8, we have $\hat{\boldsymbol{\gamma}}\to_p\boldsymbol{\gamma}^0$, and hence $\lim_{(N,T)\to\infty}P(\hat{\boldsymbol{\gamma}}\in V_B)=1$. Therefore, we only need to exmine the limiting behavior in V_B.

Define a subset of $V_B: V_B'=\{\boldsymbol{\gamma}:\bar{\omega}/a_{NT}<\|\boldsymbol{\gamma}-\boldsymbol{\gamma}^0\|\le B\}$. To prove (A83), we just need to prove $\lim_{(N,T)\to\infty}P(\hat{\boldsymbol{\gamma}}\in V_B')=0$.

Let $\hat{\boldsymbol{\phi}}(\hat{\boldsymbol{\gamma}})$ be the estimation of $y_{it}=\check{\mathbf{x}}_{it}'\boldsymbol{g}(\phi_2)+\check{\mathbf{x}}_{it}'(\boldsymbol{\gamma})[\boldsymbol{g}(\phi_1)-\boldsymbol{g}(\phi_2)]+u_{it}$. Denote $\widetilde{SSR}^*_{NT}(\boldsymbol{\gamma})=\widetilde{SSR}_{NT}(\hat{\boldsymbol{\phi}}(\hat{\boldsymbol{\gamma}}),\boldsymbol{\gamma})$ and $\widetilde{SSR}^*_{NT}(\boldsymbol{\gamma}^0)=\widetilde{SSR}_{NT}(\hat{\boldsymbol{\phi}}(\hat{\boldsymbol{\gamma}}),\boldsymbol{\gamma}^0)$, where $\widetilde{SSR}_{NT}(\hat{\boldsymbol{\phi}}(\hat{\boldsymbol{\gamma}}),\boldsymbol{\gamma})=\|\hat{\mathbf{u}}(\hat{\boldsymbol{\phi}}(\hat{\boldsymbol{\gamma}}),\boldsymbol{\gamma})\|^2$. From the estimation procedure of $\hat{\boldsymbol{\gamma}}$, we have $\widetilde{SSR}^*_{NT}(\hat{\boldsymbol{\gamma}})\le\widetilde{SSR}^*_{NT}(\boldsymbol{\gamma}^0)$. Thus it suffices to prove that for any $\boldsymbol{\gamma}\in V_B'$,

$$\lim_{(N,T)\to\infty}P\left(\widetilde{SSR}^*_{NT}(\boldsymbol{\gamma})-\widetilde{SSR}^*_{NT}(\boldsymbol{\gamma}^0)>0\right)=1. \tag{A84}$$

Note that, the equation (A84) is equivalent to prove

$$\lim_{(N,T)\to\infty}P\left(\frac{\widetilde{SSR}^*_{NT}(\boldsymbol{\gamma})-\widetilde{SSR}^*_{NT}(\boldsymbol{\gamma}^0)}{a_{NT}\|\boldsymbol{\gamma}-\boldsymbol{\gamma}^0\|}>0\right)=1. \tag{A85}$$

Since the true model can be rewritten as $\mathbf{Y}=\check{\mathbf{X}}_0\boldsymbol{G}(\boldsymbol{\phi}^0)+\mathbf{u}$, thus we have

$$\begin{aligned}\hat{\mathbf{u}}(\hat{\boldsymbol{\phi}}(\hat{\boldsymbol{\gamma}}),\boldsymbol{\gamma}) &= \mathbf{Y}-\check{\mathbf{X}}^*_{\boldsymbol{\gamma}}\boldsymbol{G}(\hat{\boldsymbol{\phi}}(\hat{\boldsymbol{\gamma}}))\\ &= \mathbf{u}+\check{\mathbf{X}}_0\boldsymbol{G}(\boldsymbol{\phi}^0)-\check{\mathbf{X}}^*_{\boldsymbol{\gamma}}\boldsymbol{G}(\hat{\boldsymbol{\phi}}(\hat{\boldsymbol{\gamma}}))\\ &= \mathbf{u}-\nabla\check{\mathbf{X}}^*_{\boldsymbol{\gamma}}\boldsymbol{G}(\boldsymbol{\phi}^0)-\check{\mathbf{X}}^*_{\boldsymbol{\gamma}}[\boldsymbol{G}(\hat{\boldsymbol{\phi}}(\hat{\boldsymbol{\gamma}}))-\boldsymbol{G}(\boldsymbol{\phi}^0)]\\ &= \mathbf{u}-\nabla\check{\mathbf{X}}_{\boldsymbol{\gamma}}\bar{\mathbf{c}}_g(NT)^{-\alpha}-\check{\mathbf{X}}^*_{\boldsymbol{\gamma}}\mathscr{T}_{NT}(\hat{\boldsymbol{\gamma}})(NT)^{-\alpha}.\end{aligned} \tag{A86}$$

Therefore,

$$\begin{aligned}\widetilde{SSR}^*_{NT}(\boldsymbol{\gamma})-\widetilde{SSR}^*_{NT}(\boldsymbol{\gamma}^0) &= \left\|\hat{\mathbf{u}}(\hat{\boldsymbol{\phi}}(\hat{\boldsymbol{\gamma}}),\boldsymbol{\gamma})\right\|^2-\left\|\hat{\mathbf{u}}(\hat{\boldsymbol{\phi}}(\hat{\boldsymbol{\gamma}}),\boldsymbol{\gamma}^0)\right\|^2\\ &= S_1+S_2-2(S_3+S_4-S_5),\end{aligned} \tag{A87}$$

where

$$\begin{aligned}S_1 &= (NT)^{-2\alpha}\bar{\mathbf{c}}_g'\nabla\check{\mathbf{X}}_{\boldsymbol{\gamma}}'\nabla\check{\mathbf{X}}_{\boldsymbol{\gamma}}\bar{\mathbf{c}}_g\\ S_2 &= (NT)^{-2\alpha}\mathscr{T}_{NT}'(\hat{\boldsymbol{\gamma}})\left(\check{\mathbf{X}}^{*\prime}_{\boldsymbol{\gamma}}\check{\mathbf{X}}^*_{\boldsymbol{\gamma}}-\check{\mathbf{X}}^{*\prime}_0\check{\mathbf{X}}^*_0\right)\mathscr{T}_{NT}(\hat{\boldsymbol{\gamma}})\\ S_3 &= (NT)^{\alpha}\bar{\mathbf{c}}_g'\nabla\check{\mathbf{X}}_{\boldsymbol{\gamma}}'\mathbf{u}\\ S_4 &= (NT)^{-\alpha}\mathscr{T}_{NT}'(\hat{\boldsymbol{\gamma}})\nabla\check{\mathbf{X}}^{*\prime}_{\boldsymbol{\gamma}}\mathbf{u}\\ S_5 &= (NT)^{-2\alpha}\bar{\mathbf{c}}_g'\nabla\check{\mathbf{X}}_{\boldsymbol{\gamma}}'\check{\mathbf{X}}^*_{\boldsymbol{\gamma}}\mathscr{T}_{NT}(\hat{\boldsymbol{\gamma}}).\end{aligned}$$

Let $\boldsymbol{\gamma}=\boldsymbol{\gamma}^0+\boldsymbol{\omega}h$, $h=(NT)^{2\delta-1}$, where $\alpha<\delta<2$, then we have $\lim_{(N,T)\to\infty}P(\boldsymbol{\gamma}\in V_B')=1$. Using Lemma 10 and Assumption 1 ($\phi_1^0-\phi_2^0=(\mathbf{c}_1-\mathbf{c}_2)(NT)^{-\alpha}$). Thus, we can show

$$
\begin{aligned}
\frac{S_1}{a_{NT}\|\boldsymbol{\gamma}-\boldsymbol{\gamma}^0\|} &= \frac{1}{\|\boldsymbol{\omega}\|}\frac{1}{NTh}\sum_{i=1}^{N}\sum_{t=1}^{T}\bar{\mathbf{c}}_g'\check{\mathbf{x}}_{it}\check{\mathbf{x}}_{it}'\bar{\mathbf{c}}_g\left|\nabla I_{it}(\boldsymbol{\gamma})\right| \\
&= \frac{1}{\|\boldsymbol{\omega}\|}G_{NT}(\boldsymbol{\omega},h) = O_p(1) > 0, && \text{(A88)}\\
\frac{S_2}{a_{NT}\|\boldsymbol{\gamma}-\boldsymbol{\gamma}^0\|} &= \frac{1}{\|\boldsymbol{\omega}\|}\mathscr{T}_{NT}'(\hat{\gamma})\frac{1}{NTh}\left(\check{\mathbf{X}}_{\gamma}^{*\prime}\check{\mathbf{X}}_{\gamma}^{*}-\check{\mathbf{X}}_{0}^{*\prime}\check{\mathbf{X}}_{0}^{*}\right)\mathscr{T}_{NT}(\hat{\gamma}) \\
&= o_p(1)O_p(1)o_p(1) = o_p(1), && \text{(A89)}\\
\frac{|S_3|}{a_{NT}\|\boldsymbol{\gamma}-\boldsymbol{\gamma}^0\|} &= \frac{1}{\|\boldsymbol{\omega}\|}\left(a_{NT}\left\|\boldsymbol{\gamma}-\boldsymbol{\gamma}^0\right\|\right)^{-1/2}\times\left|(NTh)^{-1/2}\bar{\mathbf{c}}_g'\nabla\check{\mathbf{X}}_{\gamma}'\mathbf{u}\right| \\
&\lesssim \frac{1}{\|\boldsymbol{\omega}\|}\bar{\omega}^{-1/2}\times|R_{NT}(\boldsymbol{\omega},h)| = O(1)O(\bar{\omega}^{-1/2})O_p(1), && \text{(A90)}\\
\frac{|S_4|}{a_{NT}\|\boldsymbol{\gamma}-\boldsymbol{\gamma}^0\|} &= \frac{1}{\|\boldsymbol{\omega}\|}\left(a_{NT}\left\|\boldsymbol{\gamma}-\boldsymbol{\gamma}^0\right\|\right)^{-1/2}\times\left|\mathscr{T}_{NT}'(\hat{\gamma})\times(NTh)^{-1/2}\nabla\check{\mathbf{X}}_{\gamma}^{*\prime}\mathbf{u}\right| \\
&\lesssim O(1)O(\bar{\omega}^{-1/2})o_p(1)O_p(1) = o_p(1), && \text{(A91)}\\
\frac{S_5}{a_{NT}\|\boldsymbol{\gamma}-\boldsymbol{\gamma}^0\|} &= \frac{1}{\|\boldsymbol{\omega}\|}\bar{\mathbf{c}}_g'\frac{1}{NTh}\nabla\check{\mathbf{X}}_{\gamma}'\check{\mathbf{X}}_{\gamma}^{*}\mathscr{T}_{NT}(\hat{\gamma}) \\
&= O(1)O_p(1)o_p(1) = o_p(1), && \text{(A92)}
\end{aligned}
$$

where we denote $A_{NT} \lesssim B_{NT}$ (Asymptotic inequality), if $A_{NT} \leq B_{NT}$ holds for all sufficiently large NT.

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

$$
\left|\frac{S_1}{a_{NT}\|\boldsymbol{\gamma}-\boldsymbol{\gamma}^0\|}\right| > \sum_{k=2}^{5}\left|\frac{S_k}{a_{NT}\|\boldsymbol{\gamma}-\boldsymbol{\gamma}^0\|}\right| \tag{A93}
$$

holds for all sufficiently large N in probability. By (A93), we obtain (A84) and (A85). Therefore, we have established $\lim_{(N,T)\to\infty} P(\hat{\gamma} \in V_B') = 0$ and (A83), thus we obtain the convergence rate.

□

Lemma 13. *Under Assumption 1, set $\boldsymbol{\gamma} = \boldsymbol{\gamma}^0 + \boldsymbol{\omega}/a_{NT}$, then*

$$
\sqrt{NT}\left[\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}) - \hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}^0)\right] \xrightarrow{p} \mathbf{0}, \text{ and } \sqrt{NT}\left[\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}^0) - \boldsymbol{\phi}^0\right] \xrightarrow{d} \mathcal{Z}, \tag{A94}
$$

and

$$
\begin{aligned}
\sqrt{a_{NT}}\left[\mathscr{T}_{NT}(\boldsymbol{\gamma}) - \mathscr{T}_{NT}(\boldsymbol{\gamma}^0)\right] &\xrightarrow{p} \mathbf{0} && \text{(A95)}\\
\sqrt{a_{NT}}\mathscr{T}_{NT}(\boldsymbol{\gamma}) &\xrightarrow{d} \boldsymbol{A}^{-1}\boldsymbol{J_g}\mathcal{Z}. && \text{(A96)}
\end{aligned}
$$

where $\mathcal{Z}$ is a Gaussian process with variance $\left(\boldsymbol{J_g}'\mathbf{M}^*\boldsymbol{J_g}\right)^{-1}\boldsymbol{J_g}'\boldsymbol{\Omega}^*\boldsymbol{J_g}\left(\boldsymbol{J_g}'\mathbf{M}^*\boldsymbol{J_g}\right)^{-1}$.

Proof. By Lemma 6, on any compact set $\boldsymbol{\Psi}$, we have

$$
\frac{1}{\sqrt{NT}}\check{\mathbf{X}}_{\gamma}^{*\prime}\mathbf{u} \xrightarrow{d} N\left(\mathbf{0}, \boldsymbol{A}'\boldsymbol{\Omega}^*\left(\boldsymbol{\gamma}^0\right)\boldsymbol{A}\right) \equiv \mathcal{Z}^*, \tag{A97}
$$

Also,

$$\begin{aligned}
\mathscr{S}_{NT}(\bar{\boldsymbol{\phi}},\boldsymbol{\gamma}) &\xrightarrow{p} \mathscr{S}(\boldsymbol{\phi}_c,\boldsymbol{\gamma}^0), \\
\sqrt{a_{NT}}\mathscr{A}_{NT}(\boldsymbol{\phi}^0,\boldsymbol{\gamma}) &= \frac{\partial \boldsymbol{G}'(\boldsymbol{\phi}^0)}{\partial \boldsymbol{\phi}}\left\{\frac{1}{\sqrt{NT}}\check{\mathbf{X}}_{\boldsymbol{\gamma}}^{*}{}'\mathbf{u} - \sqrt{a_{NT}}\frac{1}{NT}\check{\mathbf{X}}_{\boldsymbol{\gamma}}^{*}{}'\nabla\check{\mathbf{X}}_{\boldsymbol{\gamma}}\bar{\mathbf{c}}_g\right\}. \\
&\xrightarrow{d} \frac{\partial \boldsymbol{G}'(\boldsymbol{\phi}_c)}{\partial \boldsymbol{\phi}}\mathcal{Z}^*, \qquad \text{(A98)}
\end{aligned}$$

where $a_{NT}^{-1/2} \times a_{NT}\check{\mathbf{X}}_{\boldsymbol{\gamma}}^{*}{}'\nabla\check{\mathbf{X}}_{\boldsymbol{\gamma}}/(NT) = O(a_{NT}^{-1/2})O_p(1) = o_p(1)$.

Hence,

$$\begin{aligned}
\sqrt{NT}\left[\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}) - \boldsymbol{\phi}^0\right] &= \mathscr{S}_{NT}^{-1}(\bar{\boldsymbol{\phi}},\boldsymbol{\gamma})\sqrt{a_{NT}}\mathscr{A}_{NT}(\boldsymbol{\phi}^0,\boldsymbol{\gamma}) \\
&\xrightarrow{d} \mathscr{S}^{-1}(\boldsymbol{\phi}_c,\boldsymbol{\gamma}^0)\frac{\partial \boldsymbol{G}'(\boldsymbol{\phi}_c)}{\partial \boldsymbol{\phi}}\mathcal{Z}^* = \mathcal{Z}. \qquad \text{(A99)}
\end{aligned}$$

We have established $\sqrt{NT}[\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}^0) - \boldsymbol{\phi}^0] \xrightarrow{d} \mathcal{Z}$, and (A99) implies

$$\begin{aligned}
\sqrt{NT}\left[\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}) - \hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}^0)\right] &= \sqrt{NT}\left[\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}) - \boldsymbol{\phi}^0\right] - \sqrt{NT}\left[\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}^0) - \boldsymbol{\phi}^0\right] \\
&\xrightarrow{d} \mathcal{Z} - \mathcal{Z} = \mathbf{0}. \qquad \text{(A100)}
\end{aligned}$$

Moreover, we have

$$\begin{aligned}
\sqrt{a_{NT}}\left[\mathscr{T}_{NT}(\boldsymbol{\gamma}) - \mathscr{T}_{NT}(\boldsymbol{\gamma}^0)\right] &= \sqrt{NT}\left[\boldsymbol{G}(\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma})) - \boldsymbol{G}(\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}^0))\right] \\
&= \frac{\partial \boldsymbol{G}(\bar{\boldsymbol{\phi}})}{\partial \boldsymbol{\phi}'}\sqrt{NT}\left[\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}) - \hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}^0)\right] \\
&\xrightarrow{p} \mathbf{0} \qquad \text{(A101)} \\
\sqrt{a_{NT}}\mathscr{T}_{NT}(\boldsymbol{\gamma}) &= \frac{\partial \boldsymbol{G}(\bar{\boldsymbol{\phi}})}{\partial \boldsymbol{\phi}'}\sqrt{NT}\left[\hat{\boldsymbol{\phi}}(\boldsymbol{\gamma}) - \boldsymbol{\phi}^0\right] \\
&\xrightarrow{d} \frac{\partial \boldsymbol{G}'(\boldsymbol{\phi}_c)}{\partial \boldsymbol{\phi}}\mathcal{Z} \qquad \text{(A102)}
\end{aligned}$$

This completes the proof of Lemma 13.

□

Appendix B: Empirical comparison

In this appendix, we compare our empirical results based on the proposed threshold model with these based on a simple polynomial specification (e.g., all quadratics and interactions) using the same data as in our paper. The empirical results are reported in Table B1. First, we focus on the benchmark model without dummy variable for the poor countries. The results, reported in Column 1 of Table B1, indicate that an inverted-U relationship between temperature and growth with a universal temperature threshold around 10°C, which is lower than these reported by the related studies such as Lee et al. (2020). Second, according to the results reported in Column 2 of Table B1, the polynomial model supports a temperature threshold around 12.7°C for rich countries, while our threshold model indicates a 25°C threshold in this case. Moreover, for poor countries, the results based on the polynomial model imply a temperature threshold around 50°C, which seems to be unreasonable, while our threshold model supports a temperature threshold around 16°C for poor countries. As such, our proposed model can provide evidence supporting that the impact of future warming on growth would attenuate when countries become richer, while the polynomial model can not find such empirical evidence. Thus, we conclude that our proposed model may capture economic phenomena that can not be captured by a simple polynomial model.

Table B1: Empirical results based on a simple polynomial specification.

	(1)	(2)
$Temp_{it}$	0.757	0.709
	$[0.348, 1.167]$	$[0.191, 1.226]$
$Temp_{i,t-1}$	-0.891	-0.815
	$[-1.363, -0.419]$	$[-1.338, -0.293]$
$Temp^2_{it}$	-0.037	-0.028
	$[-0.051, -0.022]$	$[-0.041, -0.015]$
$Temp^2_{i,t-1}$	0.039	0.032
	$[0.023, 0.055]$	$[0.019, 0.046]$
$Prec_{it}$	-0.035	-0.032
	$[-0.067, -0.004]$	$[-0.077, 0.013]$
$Temp_{it} \times Poor\ dummy$		-3.264
		$[-4.217, -2.311]$
$Temp_{i,t-1} \times Poor\ dummy$		3.374
		$[2.405, 4.344]$
$Temp^2_{it} \times Poor\ dummy$		0.053
		$[0.036, 0.071]$
$Temp^2_{i,t-1} \times Poor\ dummy$		-0.058
		$[-0.077, -0.039]$
$Prec_{it} \times Poor\ dummy$		-0.059
		$[-0.120, 0.003]$

Notes: Numbers in square brackets are the 95% confidence intervals based on cluster robust standard errors.

References

Burke, M., Hsiang, S. M., Miguel, E. (2015). Global non-linear effect of temperature on economic production. *Nature*, 527(7577), 235-239.

Hall, P., Heyde, C. C. (1980). Martingale Limit Theory and Its Application. New York: Academic Press.

Hansen, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica*, 64, 413-430.

Hansen, B. E. (2000). Sample splitting and threshold estimation. *Econometrica*, 68(3): 575-603.

Lee, M., Gaspar, R., Villaruel, M. L. (2020). Effects of temperature shocks on economic growth and welfare in Asia. *Resources and Environmental Economics*, 2(2), 158-171.

Newey, W. K., McFadden, D. L. (1994). Large Sample Estimation and Hypothesis Testing. *Handbook of Econometrics*, Vol. IV, ed. by R. F. Engle and D. L. McFadden. New York: Elsevier, pp. 2113-2245.

Peligrad, M. (1982). Invariance Principles for Mixing Sequences of Random Variables. *The Annals of Probability*, 10, 968-981.

Pollard, D. (1984). *Convergence of Stochastic Processes.* NY: Springer-Verlag.

Wooldridge, J. M. (1994). Estimation and Inference for Dependent Processes. In R. F. Engle and D. L. McFadden (eds.), *Handbook of Econometrics*, Vol. IV, pp. 2639-2738. Amsterdam: Elsevier.