7-2019

Panel threshold regressions with latent group structures

Ke MIAO

Liangjun SU
Singapore Management University, ljsu@smu.edu.sg

Wendun WANG

SMU ECONOMICS &
STATISTICS

Panel Threshold Regressions with Latent Group Structures

Ke Miao, Liangjun Su, Wendun Wang

July 2019

Paper No. 13-2019

Panel Threshold Regressions with Latent Group Structures*

Ke Miao[a], Liangjun Su[a] and Wendun Wang[b]
[a] School of Economics, Singapore Management University
[b] Econometric Institute, Erasmus University Rotterdam and Tinbergen Institute

July 11, 2019

Abstract

In this paper, we consider the least squares estimation of a panel structure threshold regression (PSTR) model where both the slope coefficients and threshold parameters may exhibit latent group structures. We study the asymptotic properties of the estimators of the latent group structure and the slope and threshold coefficients. We show that we can estimate the latent group structure correctly with probability approaching 1 and the estimators of the slope and threshold coefficients are asymptotically equivalent to the infeasible estimators that are obtained as if the true group structures were known. We study likelihood-ratio-based inferences on the group-specific threshold parameters under the shrinking-threshold-effect framework. We also propose two specification tests: one tests whether the threshold parameters are homogenous across groups, and the other tests whether the threshold effects are present. When the number of latent groups is unknown, we propose a BIC-type information criterion to determine the number of groups in the data. Simulations demonstrate that our estimators and tests perform reasonably well in finite samples. We apply our model to revisit the relationship between capital market imperfection and the investment behavior of firms and to examine the impact of bank deregulation on income inequality. We document a large degree of heterogeneous effects in both applications that cannot be captured by conventional panel threshold regressions.

Key words: Classification, Dynamic panel, Latent group structures, Panel structure model, Panel threshold regression.

JEL Classification: C23, C24, C33

*The authors thank the co-editor Oliver Linton and an anonymous referee for their valuable comments on the paper. They also thank Mingli Chen, Michael Fan, Peter C.B. Phillips and Yichong Zhang for discussions on the subject matter and participants in the 2019 Asian Meeting of the Econometric Society (Xiamen) and the 2019 China Meeting of the Econometric Society (Guangzhou) for helpful comments. Su gratefully acknowledges the funding support provided by the Lee Kong Chian Fund for Excellence. Wang acknowledges the financial support by the EUR Fellowship. All errors are the authors' sole responsibility. Address correspondence to: Liangjun Su, School of Economics, Singapore Management University, 90 Stamford Road, Singapore, 178903; Phone: +65 6828 0386; e-mail: ljsu@smu.edu.sg.

1 Introduction

Threshold models have a wide variety of applications in economics; see Durlauf and Johnson (1995), Potter (1995), Kremer, Bick and Nautz (2013), and Arcand, Berkes and Panizza (2015), among others. In both the cross sectional and time series framework, asymptotic theory for estimation and inference in threshold models has been well developed. See, e.g., Chan (1993) and Hansen (2000) on asymptotic distribution theory for the threshold estimator in the fixed-threshold-effect and shrinking-threshold-effect frameworks, respectively, and Hansen (2011) for a review on the development and applications of threshold regression models in economics. Both Chan (1993) and Hansen (2000) require the exogeneity of the regressors. Endogeneity has been considered in some existing papers; see, e.g., Caner and Hansen (2004), Kourtellos, Stengos and Tan (2016), and Yu and Phillips (2018). In the panel setup, Hansen (1999) studies static panel threshold models with exogenous regressors and threshold variables; Seo and Shin (2016) propose a GMM method to estimate dynamic panel threshold models with additive fixed effects, where either the regressors or the threshold variables can be endogenous; and Miao, Li and Su (2018) study estimation and inference in dynamic panel threshold regression with interactive fixed effects.

All existing studies in panel threshold models assume that the slope coefficients and threshold parameters are common across all individual units. However, such an assumption of homogeneity is vulnerable in practice given that individual heterogeneity has been widely documented in empirical studies using panel data. See, e.g., Durlauf (2001) and Su and Chen (2013) for cross-country evidence and Browning and Carro (2007) for ample microeconomic evidences. In panel threshold regressions, heterogeneity can exist in not only the slopes but also the threshold coefficients. Neglecting latent heterogeneity in any aspect can lead to inconsistent estimation and misleading inferences. In particular, pooling individuals with different threshold values would bias the threshold and the slope coefficient estimation, and it can even lead to a failure in detecting any threshold effect in finite samples since heterogeneous threshold effects may offset each other. Even if all units share the same threshold coefficient, ignoring heterogeneity in the slopes would also lead to inconsistent estimates.

In this paper, we propose a new panel threshold model that allows the slope and threshold coefficients to vary across individual units. We model individual heterogeneity via a grouped pattern, such that all the members within the same group share the same slope and threshold coefficients, whereas these coefficients can differ across groups in an arbitrary manner. Hence, the latent group structure may result from two sources of heterogeneity: that in the slope coefficients and that in the threshold level coefficients. We allow the group membership structure (i.e., which individuals belong to which group) to be unknown and estimated from the data. We refer to our model as a *panel structure threshold regression* (PSTR) model.

Using a panel structure model that imposes a group pattern is a convenient way to model unobserved heterogeneity, and they have recently received much attention; see Lin and Ng (2012), Bonhomme and Manresa (2015), Ando and Bai (2016, 2017), Su, Shi, and Phillips (2016), Lu and Su (2017), Liu et al. (2018), Su and Ju (2018), Su, Wang and Jin (2019), and Okui and Wang

(2019), among others. An important advantage of the panel structure model is that it allows flexible forms of unobserved heterogeneity while remaining parsimonious at the same time. As group structure is latent in such a model, the determination of an individual's membership is the key question. Several approaches have been proposed to address this issue. Sun (2005), Kasahara and Shimotsu (2009), and Browning and Carro (2011) consider finite mixture models. Su, Shi, and Phillips (2016) propose a variant of the Lasso procedure (C-Lasso) to achieve a classification in this regard, and this method has been extended to allow for two-way component errors, interactive fixed effects, nonstationary regressors, and semiparametric specification, respectively, in Lu and Su (2017), Su and Ju (2018), Huang, Jin and Su (2019), and Su, Wang and Jin (2019). Lin and Ng (2012), Bonhomme and Manresa (2015), Sarafidis and Weber (2015), and Liu et al. (2018) extend the K-means algorithms to the panel regression framework. Wang, Phillips and Su (2018) and Wang and Su (2019) propose to identify the latent group structure based on the Lasso or spectral clustering techniques in the statistics literature. In the nonparametric literature, Vogt and Linton (2017, 2019) consider procedures to estimate the unknown group structures for nonparametric regression curves.

To estimate the PSTR model, we consider a least-squares-type estimator that minimizes the sum of squared errors. We choose the least-squares approach for classification because the group, slope, and threshold parameters can be estimated simultaneously, which facilitates the theory. The disadvantage is that we cannot allow for endogeneity in the regressors and threshold variables. Cases with endogenous regressors or threshold variables require different and more complicated analysis and will be left for future research. Due to the presence of the latent group structure and threshold parameters, we do not have an analytically closed-form solution to the problem. We propose to employ an EM-type iterative algorithm to find the solution with multiple starting values. Under some regularity conditions, we show that our estimators of the slope and threshold coefficients are asymptotically equivalent to the corresponding infeasible estimators of the group-specific parameters that are obtained by using individual group identity information.

To study the asymptotic properties of the estimators of the threshold coefficients, we follow the lead of Hansen (2000) and consider the shrinking-threshold-effect framework, where the threshold effect is diminishing as the sample size approaches infinity. In this framework, we can make inferences regarding each threshold parameter by constructing a likelihood ratio (LR) statistic. We show that the LR statistics are asymptotically pivotal in the case of conditional homoskedasticity and that they depend on a scale nuisance parameter otherwise. Such a scale parameter can be consistently estimated nonparametrically when conditional heteroskedasticity is suspected.

We also consider two specification test statistics. The first one is designed to test the homogeneity of the threshold parameters across each group via the LR principle. The corresponding LR test statistic is non-standard and involves a linear combination of two-sided Brownian motions. We show how one can obtain the simulated p-value with estimated parameters in our discussion. This test is useful since pooling units, if their threshold coefficients pass the homogeneity test, improves the efficiency of threshold estimation, especially in small samples. The second is designed to test the absence of the threshold effect under the null by adopting the method proposed by Hansen

(1996). In our latent group structure framework, one may suspect the presence of a subset of threshold effects among all groups, and we also need to take into account the uncertainty caused by the unknown group structure when studying the asymptotic behavior of the test.

We evaluate the finite-sample performance of the proposed tests and estimation methods via extensive simulation studies. First, the proposed information criterion can determine the correct number of groups with a large probability, regardless of whether any threshold effect is present. Given the number of groups, the next task is to test the existence of threshold effects. Our proposed test has an appropriate size and non-trivial power in detecting the threshold effect. The power is an increasing function of the strength of both the threshold effect and sample size. A nice feature of the test is that it performs well regardless of whether the threshold is heterogeneous across units. If the threshold effect is present, one can further test whether the threshold parameters differ across groups. We demonstrate that our test for the homogeneity of the threshold is also well-behaved in terms of size and its power improves as the degree of threshold heterogeneity and sample sizes increase. Finally, after the model and the number of groups are specified, we can proceed with parameter estimation. Our estimation method performs well in heterogeneous panels with threshold effects in finite samples. With this method, we can precisely estimate group membership, and the clustering accuracy improves as the number of time periods increases. Both the threshold parameters and slope coefficients can be precisely estimated. Moreover, we find that when the threshold parameters are homogeneous across groups, pooling observations with a common threshold does improve the efficiency of threshold estimation, which in turn highlights the importance of testing the homogeneity of the threshold parameters.

We illustrate the usefulness of our methods through two real-data examples. First, we revisit the relationship between capital market imperfections and firms' investment behavior. We document a large degree of heterogeneity in firms' investment behavior, which is bounded by various types of financial constraints, such as cash flow, Tobin's Q, and leverage. Such heterogenous threshold effects cannot be captured by the conventional panel threshold regressions. Next, we examine the impact of bank regulation, particularly branch deregulation, on income inequality in the US, allowing observed and unobserved heterogeneity in their impact. We find a group pattern of heterogeneity in the impact of deregulation across states even after controlling for the threshold effect. The group structure coincides with geographic locations to some extent but not perfectly, and the threshold effects appear to be salient in each group. This application again demonstrates the usefulness of the PSTR since it allows us to capture both observed heterogeneity through thresholds and unobserved heterogeneity through the latent group structure.

The remainder of the paper is organized as follows. In Section 2, we introduce our model and estimation method. In Section 3, we introduce the assumptions and examine the asymptotic properties of the estimators of the latent group structure and the slope and threshold coefficients. In Section 4, we introduce the inference procedure on the threshold parameters and propose a specification test for the homogeneity of the threshold parameters across groups. In Section 5, we consider the specification test for the presence of threshold effects. In Section 6, we propose a BIC-type information criterion to determine the number of groups. We conduct Monte Carlo experiments to evaluate the finite sample performance of our estimators and tests in Section 7. We apply our model to study the relationship between investment and financing constraints and the relationship between bank regulation and income distribution in Section 8. Section 9 concludes. The proofs of the main results in the paper are relegated to the Appendix. Further technical details can be found in the online Supplementary Materials.

To proceed, we adopt the following notation. The indicator function is denoted as $\mathbf{1}(\cdot)$. $\mathbf{0}_{a\times b}$ denotes an $a\times b$ matrix of zeros. For two constants a and b, we denote $\max(a,b)$ as $a\vee b$ and $\min(a,b)$ as $a\wedge b$. For an $m\times n$ real matrix A, we denote its transpose as A' and its Frobenius norm as $\|A\|$ $(\equiv[tr(AA')]^{1/2})$ where $\equiv$ means "is defined as". For a real symmetric matrix A, we denote its minimum eigenvalue as $\lambda_{\min}(A)$. The operators $\stackrel{p}{\to}$ and $\stackrel{d}{\to}$ denote convergence in probability and distribution, respectively. We use $(N,T)\to\infty$ to denote the joint convergence of N and T when N and T pass to infinity simultaneously. Alternatively, as the co-editor suggests, one can consider the pathwise asymptotics as in Phillips and Moon (1999) and Vogt and Linton (2009).

2 The Model and Estimates

In this section we first present the panel threshold model with latent group structures and then introduce the estimators of all the parameters in the model.

2.1 The Model

Let N denote the number of cross-sectional units and T the number of time periods. We consider the model

$$y_{it}=x_{it}'\beta^0_{g_i^0}+x_{it}'\delta^0_{g_i^0}\cdot d_{it}(\gamma^0_{g_i^0})+\mu_i+\varepsilon_{it},\quad i=1,...,N,\ t=1,...,T, \tag{2.1}$$

where x_{it} is a $K\times 1$ vector of observable regressors, $d_{it}(\gamma)\equiv\mathbf{1}(q_{it}\le\gamma)$, q_{it} is a scalar threshold variable, μ_i is the individual fixed effect and ε_{it} is the idiosyncratic error term. Note that we allow both the slope and threshold coefficients to be group specific: γ_g^0 is a scalar threshold coefficient, β_g^0 is a $K\times 1$ vector of regression coefficients that lies in a compact parameter space $\mathcal{B}$, and δ_g^0 is a $K\times 1$ vector of threshold-effect coefficients for $g\in\mathcal{G}\equiv\{1,...,G\}$, where G is a fixed integer known as the number of groups. The group-membership variable $g_i^0\in\mathcal{G}$ indicates to which group individual unit i belongs. This group-membership variable is unknown and has to be estimated from the data. All members in group g have the same coefficients $(\beta_g^{0\prime},\delta_g^{0\prime},\gamma_g^0)'$. We assume $\gamma_g^0\in\mathbf{\Gamma}=[\underline{\gamma},\overline{\gamma}]$ for all $g\in\mathcal{G}$, where $\underline{\gamma}$ and $\overline{\gamma}$ are two fixed constants. Following the lead of Hansen (2000), we will work in the shrinking-threshold-effect framework by assuming that $\delta_g^0\equiv\delta_{g,NT}^0\to 0$ as $(N,T)\to\infty$ for each $g\in\mathcal{G}$ unless specified otherwise.

Let $\mathbf{D}\equiv(\gamma_1,...,\gamma_G)'\in\Gamma^G$, $\mathbf{G}\equiv(g_1,...,g_N)'\in\mathcal{G}^N$ and $\mathbf{\Theta}\equiv(\theta_1',...,\theta_G')'\in\mathcal{B}^G$, where $\theta_g\equiv(\beta_g',\delta_g')'\in\mathcal{B}\subset\mathbb{R}^{2K}$. For any given group structure $\mathbf{G}$, we let $\mathbf{G}_g=\{i|\ g_i=g,\ 1\le i\le N\}$ be the index set of the members in group $g\in\mathcal{G}$. We denote the true parameters as $(\mathbf{\Theta}^0,\mathbf{D}^0,\mathbf{G}^0)$, where $\mathbf{\Theta}^0\equiv(\theta_1^{0\prime},...,\theta_G^{0\prime})'$, $\mathbf{D}^0\equiv(\gamma_1^0,...,\gamma_G^0)'$ and $\mathbf{G}^0\equiv(g_1^0,...,g_N^0)'$. Analogously, we denote the true members in group $g\in\mathcal{G}$ by $\mathbf{G}_g^0=\{i|\ g_i^0=g,\ 1\le i\le N\}$.

For the moment, we assume that the true number of groups G^0 is known and given by G. In Section 6, we will discuss how to determine G^0 in practice.

2.2 Estimation

To remove the individual-specific fixed effects μ_i, we employ the usual within-transformation which leads to

$$\tilde{y}_{it} = \tilde{x}_{it}'\beta^0_{g_i^0} + \tilde{x}_{it}(\gamma^0_{g_i^0})'\delta^0_{g_i^0} + \tilde{\varepsilon}_{it}, \quad i = 1, ..., N, \ t = 1, ..., T, \tag{2.2}$$

where $\tilde{x}_{it}(\gamma) \equiv x_{it}d_{it}(\gamma) - \frac{1}{T}\sum_{s=1}^{T} x_{is}d_{is}(\gamma)$, and $\tilde{x}_{it}$, $\tilde{y}_{it}$ and $\tilde{\varepsilon}_{it}$ are defined analogously. Let $z_{it}(\gamma) \equiv (x_{it}', x_{it}'d_{it}(\gamma))'$ and $\tilde{z}_{it}(\gamma) \equiv z_{it}(\gamma) - \frac{1}{T}\sum_{s=1}^{T} z_{is}(\gamma)$. Then the model in (2.2) can be rewritten as

$$\tilde{y}_{it} = \tilde{z}_{it}(\gamma^0_{g_i^0})'\theta^0_{g_i^0} + \tilde{\varepsilon}_{it}, \quad i = 1, ..., N, \ t = 1, ..., T. \tag{2.3}$$

Given G, we can obtain the following least squares estimator of $(\Theta, \mathbf{D}, \mathbf{G})$:

$$(\hat{\Theta}, \hat{\mathbf{D}}, \hat{\mathbf{G}}) = \underset{(\Theta,\mathbf{D},\mathbf{G})\in\mathcal{B}^G\times\Gamma^G\times\mathcal{G}^N}{\operatorname{argmin}} \mathcal{Q}(\Theta, \mathbf{D}, \mathbf{G}),$$

where

$$\mathcal{Q}(\Theta, \mathbf{D}, \mathbf{G}) = \sum_{i=1}^{N}\sum_{t=1}^{T} \left[\tilde{y}_{it} - \tilde{z}_{it}(\gamma_{g_i})'\theta_{g_i}\right]^2. \tag{2.4}$$

For any given threshold $\mathbf{D}$ and group structure $\mathbf{G}$, the slope coefficients θ_g, $g = 1, ..., G$, can be estimated by

$$\hat{\theta}_g(\mathbf{D}, \mathbf{G}) = \left(\sum_{i=1}^{N}\sum_{t=1}^{T} \mathbf{1}(g_i = g)\tilde{z}_{it}(\gamma_g)\tilde{z}_{it}(\gamma_g)'\right)^{-1} \sum_{i=1}^{N}\sum_{t=1}^{T} \mathbf{1}(g_i = g)\tilde{z}_{it}(\gamma_g)\tilde{y}_{it}.$$

Concentrating out Θ, we can estimate the threshold $\mathbf{D}$ and group structure $\mathbf{G}$ by

$$(\hat{\mathbf{D}}, \hat{\mathbf{G}}) = \underset{(\mathbf{D},\mathbf{G})\in\Gamma^G\times\mathcal{G}^N}{\operatorname{argmin}} \dot{\mathcal{Q}}(\mathbf{D}, \mathbf{G}), \tag{2.5}$$

where $\dot{\mathcal{Q}}(\mathbf{D}, \mathbf{G}) \equiv \mathcal{Q}(\hat{\Theta}(\mathbf{D}, \mathbf{G}), \mathbf{D}, \mathbf{G})$ and $\hat{\Theta}(\mathbf{D}, \mathbf{G}) = (\hat{\theta}_1(\mathbf{D}, \mathbf{G})', ..., \hat{\theta}_G(\mathbf{D}, \mathbf{G})')'$.

To find the solution to the above optimization problem, we need to search over the space of $(\mathbf{D}, \mathbf{G})$ to minimize the objective function in (2.5). We propose to employ the following EM-type iterative algorithm to conduct the searching process:

Algorithm 2.1 *Set* $\mathbf{G}^{(0)}$ *as a random initialization of the group structure* $\mathbf{G}$ *and let* $s = 0$.

Step 1 *For given* $\mathbf{G}^{(s)}$, *compute*

$$\mathbf{D}^{(s)} = \operatorname{argmin}_{\mathbf{D}\in\Gamma^G} \dot{\mathcal{Q}}(\mathbf{D}, \mathbf{G}^{(s)}).$$

Step 2 *For given* $\mathbf{D}^{(s)} = \{\gamma_g^{(s)},\ g = 1, ..., G\}$ *and* $\mathbf{G}^{(s)} = \{g_i^{(s)},\ i = 1, ..., N\}$*, compute the slope coefficients for each group* $g \in \mathcal{G}$

$$\hat{\theta}_g^{(s)} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{1}(g_i^{(s)} = g)\tilde{z}_{it}(\gamma_g^{(s)})\tilde{z}_{it}(\gamma_g^{(s)})'\right)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{1}(g_i^{(s)} = g)\tilde{z}_{it}(\gamma_g^{(s)})\tilde{y}_{it}.$$

Step 3 *Compute for all* $i \in \{1, \ldots, N\}$,

$$g_i^{(s+1)} = \underset{g\in\mathcal{G}}{\operatorname{argmin}}\sum_{t=1}^{T}[\tilde{y}_{it} - \tilde{z}_{it}(\gamma_g^{(s)})'\hat{\theta}_g^{(s)}]^2.$$

Step 4 *Set* $s = s + 1$*. Repeat Steps 1-3 until numerical convergence.*

The above algorithm is similar to Algorithm 1 in Bonhomme and Manresa (2015, BM hereafter) and it alternates among three steps. Steps 1 and 2 are the "update" steps where one updates the estimates of the threshold parameter and those of the slope coefficients in turn. Step 3 is an "assignment" step where each individual i is re-assigned to the group $g_i^{(s+1)}$. The objective function is non-increasing in the number of iterations and we find through simulations that numerical convergence is typically very fast. Nevertheless, it is hard to ensure that the obtained solution is globally optimal because it depends on the chosen starting values. In practice, one can start with multiple random starting values and select the solution that yields the lowest objective value.

3 Asymptotic Theory

In this section, we study the asymptotic properties of the estimators of the group structure, slope and threshold parameters. We first show the consistency of the group structure estimator and then establish the asymptotic properties of the estimators of the slope and threshold coefficients.

3.1 The estimator of the group structure

We establish the consistency of the group structure estimator in this subsection. Let $\mathcal{F}_{NT,t} \equiv \sigma(\{(x_{it}, q_{it}, \varepsilon_{i,t-1}), (x_{i,t-1}, q_{i,t-1}, \varepsilon_{i,t-2}), ...\}_{i=1}^{N})$ where $\sigma(A)$ denotes the minimal sigma-field generated from A. Let $X_i = (x_{i1}, ..., x_{iT})'$, $\varepsilon_i = (\varepsilon_{i1}, ..., \varepsilon_{iT})'$ and $q_i = (q_{i1}, ..., q_{iT})'$. We use N_g to denote the number of individuals belonging to group $g : N_g = \left|\mathbf{G}_g^0\right|$. That is, $\left|\mathbf{G}_g^0\right|$ denotes the cardinality of $\mathbf{G}_g^0$. For any group structure $\mathbf{G}$, let

$$M_{NT}(g, \tilde{g}, \mathbf{D}, \mathbf{G}) \equiv \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{1}(g_i^0 = g)\mathbf{1}(g_i = \tilde{g})\tilde{z}_{it}(\gamma_{\tilde{g}})\tilde{z}_{it}(\gamma_{\tilde{g}})'.$$

Let $0 < C < \infty$ denote a generic constant that may vary across places. Let $\max_i = \max_{1\leq i\leq N}$, $\max_t = \max_{1\leq t\leq T}$ and $\max_{i,t} = \max_{1\leq i\leq N}\max_{1\leq t\leq T}$. We first make the following assumptions.

Assumption A.1: (i.1) For each $i = 1, ..., N$, $t = 1, ..., T$, $E(\varepsilon_{it}|\mathcal{F}_{NT,t-1}) = 0$ a.s., or (i.2) for each $i = 1, ..., N$, $t = 1, ..., T$, $E(\varepsilon_{it}|X_i, q_i) = 0$ a.s.;

(ii) $\{(x_{it}, q_{it}, \varepsilon_{it}) : t = 1, 2, \ldots\}$ are mutually independent of each other across i;

(iii) The process $\{(x_{it}, q_{it}, \varepsilon_{it}), t \geq 1\}$ is a strong mixing process with mixing coefficients $\alpha_i[t]$ satisfying $\max_{1 \leq i \leq N} \alpha_i[t] \leq c_\alpha \rho^t$ for some constants $c_\alpha > 0$ and $\rho \in (0, 1)$.

(iv) The parameter space $\mathcal{B}$ and Γ are compact so that $\sup_{\theta \in \mathcal{B}} \|\theta\| \leq C$ and $\Gamma = [\underline{\gamma}, \overline{\gamma}]$;

(v) $\max_{i,t} E \|x_{it}\|^{8+\epsilon_0} \leq C$ and $\max_{i,t} E(\|\varepsilon_{it}\|^{8+\epsilon_0}) \leq C$ for some $\epsilon_0 > 0$;

(vi) The threshold effect satisfies $\delta_g^0 = (NT)^{-\alpha} C_g^0$ for some constants $\alpha \in (0, 1/2)$ and $C_g^0 \neq 0$ for all $g \in \mathcal{G}$.

Assumption A.2: There exists a constant $\underline{c}_\lambda > 0$ such that for all $g \in \mathcal{G}$,

$$\Pr\left(\inf_{(\mathbf{G},\mathbf{D}) \in \mathcal{G}^N \times \Gamma^G} \max_{\tilde{g} \in \mathcal{G}} \{\lambda_{\min}[M_{NT}(g, \tilde{g}, \mathbf{D}, \mathbf{G})]\} > \underline{c}_\lambda\right) \to 1 \text{ as } (N, T) \to 1.$$

Assumption A.3: (i) For all $g, \tilde{g} \in \mathcal{G}$ with $g \neq \tilde{g}$, we have $\left\|\beta_g^0 - \beta_{\tilde{g}}^0\right\| > \underline{c}_\beta$ for some constant $\underline{c}_\beta > 0$;

(ii) For any $g \neq \tilde{g}$ and $1 \leq i \leq N$, we have $E[\tilde{x}_{it}'(\beta_{\tilde{g}}^0 - \beta_g^0)]^2 \equiv \underline{c}_{g\tilde{g},i} \geq \underline{c}_{g\tilde{g}}$ for some constant $\underline{c}_{g\tilde{g}} > 0$;

(iii) For all $g \in \mathcal{G}$: $\lim_{N \to \infty} N_g/N = \pi_g > 0$.

(iv) $N = O(T^2)$ and $T = O(N^2)$ as $(N, T) \to \infty$.

Assumption A.1(i)–(iii) is similar to Assumption A.2(a)–(c) in Su and Chen (2013). The major differences lie in four aspects. First, Su and Chen (2013) consider linear panel data models with interactive fixed effects and the sigma-field $\mathcal{F}_{NT,t}$ there also incorporates the factors and factor loadings, whereas we consider the panel threshold regression models with a latent group structure and the additive fixed effects. Second, Su and Chen (2013) only consider Assumption A.1(i.1) and allow for lagged dependent variables to appear in the regressor vector. Here we consider both scenarios in Assumption A.1(i): the martingale difference sequence (m.d.s.) condition in A.1(i.1) and the strict exogeneity condition in A.1(i.2), where we allow for dynamic panels in the first scenario and assume strict exogeneity in the second scenario. In the second scenario, we allow for serial correlation of an unknown form in the error term. When A.1(i.1) holds, we have asymptotic biases for the estimators of the slope coefficients. When A.1(i.2) holds and serial correlation is likely to appear, we have to use the HAC estimator for the asymptotic variance of the slope estimators. Third, due to the potential appearance of the lagged dependent variables in the regression model, Su and Chen (2013) use the notion of conditional strong mixing for the process while we focus on the case of unconditional strong mixing in our model in Assumption A.1(iii). In other words, we follow Hahn and Kuersteiner (2011) and treat the fixed effects μ_i's to be nonrandom in our setting in the dynamic case. If μ_i's are random, we can modify the unconditional strong mixing conditions to the conditional strong mixing conditions as in Su and Chen (2013). Fourth, Su and Chen (2013) assume conditional cross-sectional independence whereas we assume cross-sectional independence in Assumption A.1(ii).

A.1(iv) is imposed to facilitate the proof as we do not have closed form solutions to our optimization problem. Assumption A.1(v) imposes some moment conditions on the regressors and error terms, which are weaker than the exponential tail assumption in BM (2015). Assumption A.1(vi) assumes shrinking threshold effect as in Hansen (2000). In this framework, the asymptotic distribution of the estimator of γ_g is pivotal up to a scale effect, which facilitates the inference procedure. In part E of the online supplement we study the asymptotic properties of our estimators in the fixed threshold effect framework. In the latter case, the inference becomes difficult in practice and one can consider extending the smoothed least squares estimation of Seo and Linton (2007) to our PSTR model.

Assumption A.2 is similar to Assumption 1(g) in BM (2015). Given any conjectured group structure $\mathbf{G}$ and for each $g \in \mathcal{G}$, we cannot assume $\lambda_{\min}[M_{NT}(g, \tilde{g}, \mathbf{D}, \mathbf{G})] > \underline{c}_\lambda$ for any $\tilde{g} \in \mathcal{G}$ due to the possibility of very few individuals assigned to be in group $\tilde{g}$. However, there exists some group $\tilde{g} \in \mathcal{G}$, in which a positive proportion of N members are assigned. As BM (2015) remark, such an assumption is reminiscent of the full rank condition in standard regression models.

Assumption A.3(i) and (iii) parallels Assumption A1(vi)–(vii) in Su, Shi, and Phillips (2016, SSP hereafter). A.3(i) requires that the group-specific slope coefficients be separated from each other, and it can be relaxed to allow the differences between the group-specific slope coefficients to shrink to zero at some slow rates at the cost of more lengthy arguments. It is worth emphasizing that the latent group structure is identified through the separation of group-specific slope coefficients and we find that the potential separation of the threshold parameters is not necessary; see the remarks after Theorem 3.1 for futher discussions. A.3(iii) implies that each group has an asymptotically non-negligible proportion of individuals as $N \to \infty$. Noting that $E[\tilde{x}_{it}'(\beta_{\tilde{g}}^0 - \beta_g^0)]^2 = (\beta_{\tilde{g}}^0 - \beta_g^0)' E(\tilde{x}_{it}\tilde{x}_{it}')(\beta_{\tilde{g}}^0 - \beta_g^0)$, A.3(ii) is automatically satisfied under A.3(i) provided that the minimum eigenvalue of $E(\tilde{x}_{it}\tilde{x}_{it}')$ is bounded away from zero. Apparently, x_{it} cannot contain time-invariant regressors under Assumption A.3(ii). Assumption A.3(iv) puts some restrictions on the relative magnitudes of N and T, which can be easily met in many macro and financial applications. If we follow BM (2015) and assume exponentially-decaying tails, we can relax the conditions on (N, T) to $N/T^v \to 0$ as $(N, T) \to \infty$ for some $v > 0$. If we follow Vogt and Linton (2019) and consider the pathwise asymptotics by setting $N = g(T)$ for some divergent function $g(\cdot)$ and passing $T \to \infty$. Then Assumption A.3(iv) can be satisfied when $g(T)/T^2 + T/g(T)^2$ converges to some nonnegative finite constant as $T \to \infty$.

The following theorem reports the consistency of the estimators of the group membership for all individuals.

Theorem 3.1 *Suppose that Assumptions A.1–A.3 hold. Then*

$$\Pr\left(\sup_{1 \le i \le N} \mathbf{1}(\hat{g}_i \neq g_i^0) = 1\right) \to 0 \text{ as } (N, T) \to \infty.$$

Theorem 3.1 is similar to Theorem 2 of BM (2015). This theorem states that as $(N, T) \to \infty$, we can correctly estimate the group structure with probability approaching one (w.p.a.1). From the proof of the above theorem, we can see that the identification of the true group structure highly hinges on Assumption A.3(i). In particular, since we permit $\delta_g^0 = \delta_{g,NT}^0 \to 0$ as $(N,T) \to \infty$ under the shrinking-threshold-effect framework, the proof of Theorem 3.1 mainly relies on the differences of β_g^0's across groups. In this case, as long as the slope coefficients in one regime are separate from each other across the G groups, they are also separate from each other asymptotically in the other regime and whether the threshold parameters in different groups differ from each other does not matter. In other words, the threshold parameters do not need to separate from each other. In the online Supplementary Material, we give a proof of Theorem 3.1 under the fixed-threshold-effect framework. We show that in that case, either the separation among θ_g^0's or that among γ_g^0's is sufficient for identifying the latent group structure under some regularity conditions. To stay focused, we will work in the shrinking-threshold-effect framework below.

3.2 The estimators of the slope and threshold coefficients

Given the fact that the latent group structure can be recovered from the data at a sufficiently fast rate (see Lemma A.3 in the appendix), we will show that the estimators of the slope and threshold coefficients are asymptotically equivalent to the infeasible estimators that are obtained as if the true group structure were known. Then we derive the asymptotic distributions of the coefficient estimators.

To establish the asymptotic equivalence, we add some notation. Let $\tilde{x}_{it}(\gamma,\gamma^*) = \tilde{x}_{it}(\gamma) - \tilde{x}_{it}(\gamma^*)$. Let $f_{it}(\cdot)$ denote the probability density function (PDF) of q_{it}. For all $g \in \mathcal{G}$, define

$$
\begin{aligned}
w_g(\gamma) &= \frac{1}{N_g T}\sum_{i\in \mathbf{G}_g^0}\sum_{t=1}^{T} \tilde{z}_{it}(\gamma)\tilde{z}_{it}(\gamma)', \\
\tilde{w}_g(\gamma) &= \frac{1}{N_g T}\sum_{i\in \mathbf{G}_g^0}\sum_{t=1}^{T} \tilde{x}_{it}(\gamma,\gamma_g^0)\tilde{x}_{it}(\gamma,\gamma_g^0)' \\
&\quad -\frac{1}{N_g T}\sum_{i\in \mathbf{G}_g^0}\sum_{t=1}^{T} \tilde{x}_{it}(\gamma,\gamma_g^0)\tilde{z}_{it}(\gamma)' \left[w_g(\gamma)\right]^{-1} \frac{1}{N_g T}\sum_{i\in \mathbf{G}_g^0}\sum_{t=1}^{T} \tilde{z}_{it}(\gamma)\tilde{x}_{it}(\gamma,\gamma_g^0)'. \\
M_{g,NT}(\gamma) &= \frac{1}{N_g T}\sum_{i\in \mathbf{G}_g^0}\sum_{t=1}^{T} E[x_{it}x_{it}' d_{it}(\gamma)], \\
D_{g,NT}(\gamma) &= \frac{1}{N_g T}\sum_{i\in \mathbf{G}_g^0}\sum_{t=1}^{T} E(x_{it}x_{it}' | q_{it} = \gamma) f_{it}(\gamma), \text{ and} \\
V_{g,NT}(\gamma) &= \frac{1}{N_g T}\sum_{i\in \mathbf{G}_g^0}\sum_{t=1}^{T} E(x_{it}x_{it}'\varepsilon_{it}^2 | q_{it} = \gamma) f_{it}(\gamma).
\end{aligned}
$$

Let $M_g(\gamma) = \lim_{(N,T)\to\infty} M_{g,NT}(\gamma)$, $D_g(\gamma) = \lim_{(N,T)\to\infty} D_{g,NT}(\gamma)$, $V_g(\gamma) = \lim_{(N,T)\to\infty} V_{g,NT}(\gamma)$, $D_g^0 = D_g(\gamma_g^0)$, and $V_g^0 = V_g(\gamma_g^0)$. We add the following two assumptions.

Assumption A.4: (i) There exists a constant $\tau > 0$ such that $\Pr\left(\min_{\gamma\in\Gamma}\lambda_{\min}\left[w_g(\gamma)\right] \geq \tau\right) \to 1$ as $(N,T) \to \infty$ for all $g \in \mathcal{G}$;

(ii) There exists a constant $\tau > 0$ such that $\min_{\gamma\in\Gamma}\left\{\Pr(\lambda_{\min}[\tilde{w}_g(\gamma)] \geq \tau \min[1, \left|\gamma - \gamma_g^0\right|])\right\} \to 1$ as $(N,T) \to \infty$ for each $g \in \mathcal{G}$.

Assumption A.5: (i) $\max_{\gamma\in\Gamma}\max_{i,t} E(\|\xi_{it}\|^4 \,|q_{it} = \gamma) \leq C$ for $\xi_{it} = x_{it}$ and $x_{it}\varepsilon_{it}$;

(ii) $f_{it}(\gamma)$ is continuous over Γ and $\max_{i,t}\sup_{\gamma\in\Gamma} f_{it}(\gamma) \leq c_f < \infty$.

(iii) For $g \in \mathcal{G}$, $D_g(\gamma)$ and $V_g(\gamma)$ are continuous at $\gamma = \gamma^0$;

(iv) There exists a constant $c > 0$ such that $\inf_{\gamma\in\Gamma}\lambda_{\min}[M_g(\gamma)] \geq c$ for all $g \in \mathcal{G}$.

Assumption A.4(i) is a non-colinearity assumption for the regressors and A.4(ii) holds because $E\left\|x_{it}(\gamma) - x_{it}(\gamma^*)\right\| \asymp |\gamma - \gamma^*|$ under some regularity conditions on $\{x_{it}, q_{it}\}$, where $a \asymp b$ means and both a/b and b/a are bounded away from zero. It's natural to expect that the first term in the definition of $\tilde{w}_g(\gamma)$ is of the same probability order as $\left|\gamma - \gamma_g^0\right|$. A.4(ii) requires that after projecting $\tilde{x}_{it}(\gamma, \gamma_g^0)$ onto $\tilde{z}_{it}(\gamma)$, the associated residual exhibits the same probability order of variations groupwise. Assumption A.5 imposes some conditions on the conditional PDF and moments of x_{it} and $x_{it}\varepsilon_{it}$. A.5(i) requires that the fourth order conditional moment of $x_{it}\varepsilon_{it}$ and x_{it} be well behaved; A.5(ii) requires that the PDF of q_{it} be uniformly bounded; A.5(iii)–(iv) requires the probability limits of some quantities associated with the asymptotic variance be well behaved.

To state the next theorem, we define the infeasible estimators of the slope and threshold coefficients that are obtained with known group structures:

$$(\check{\Theta}, \check{\mathbf{D}}) \equiv \underset{(\Theta,\mathbf{D})\in\mathcal{B}^G\times\Gamma^G}{\operatorname{argmin}} \check{\mathcal{Q}}(\Theta, \mathbf{D}), \tag{3.1}$$

where $\check{\mathcal{Q}}(\Theta, \mathbf{D}) \equiv \mathcal{Q}(\Theta, \mathbf{D}, \mathbf{G}^0)$. With the knowledge of the true group structure $\mathbf{G}^0$, we can split the N individuals into G groups perfectly and estimate the group-specific parameters for each group. Let $\check{\mathcal{Q}}_g(\theta, \gamma) = \sum_{i\in\mathbf{G}_g^0}\sum_{t=1}^T[\tilde{y}_{it} - \tilde{z}_{it}(\gamma)'\theta]^2$. Then we have

$$\check{\mathcal{Q}}(\Theta, \mathbf{D}) = \sum_{g=1}^G \check{\mathcal{Q}}_g(\theta, \gamma) \text{ and } (\check{\theta}_g, \check{\gamma}_g) = \underset{(\theta,\gamma)\in\mathcal{B}\times\Gamma}{\operatorname{argmin}} \check{\mathcal{Q}}_g(\theta, \gamma) \;\text{ for each } g \in \mathcal{G}.$$

The following theorem establishes the asymptotic equivalence between the feasible estimator $(\hat{\Theta}, \hat{\mathbf{D}})$ and the infeasible estimator $(\check{\Theta}, \check{\mathbf{D}})$.

Theorem 3.2 *Suppose that Assumptions A.1–A.5 hold with $\alpha \in (0, 1/3)$ in Assumption A.1(vi). Let $\alpha_{NT} = (NT)^{1-2\alpha}$. Then we have $(NT)^{1/2}\left\|\hat{\Theta} - \check{\Theta}\right\| \xrightarrow{p} 0$ and $\alpha_{NT}(\hat{\mathbf{D}} - \check{\mathbf{D}}) \xrightarrow{p} 0$.*

Theorem 3.2 shows that $\hat{\Theta} - \check{\Theta} = o_p((NT)^{-1/2})$ and $\hat{\mathbf{D}} - \check{\mathbf{D}} = o_p(\alpha_{NT}^{-1})$ by restricting $\alpha \in (0, 1/3)$ in Assumption A.1(vi). Under Assumptions A.1–A.5, we can show that $\check{\Theta} - \Theta^0 = O_p((NT)^{-1/2} + T^{-1})$ and $\check{\mathbf{D}}$ has α_{NT}-rate of convergence. Therefore, the estimator $(\hat{\Theta}, \hat{\mathbf{D}})$ has the same asymptotic distribution as that of $(\check{\Theta}, \check{\mathbf{D}})$. Then we can establish the asymptotic distribution of our least squares estimator.

To report the asymptotic distributions of $\hat{\theta}_g$ and $\hat{\gamma}_g$, we add some notation:

$$\begin{aligned}
\omega_{g,NT}(\gamma,\gamma^*) &\equiv \frac{1}{N_g T}\sum_{i\in\mathbf{G}_g^0}\sum_{t=1}^{T}\tilde{z}_{it}(\gamma)\tilde{z}_{it}(\gamma^*)',\\
\Omega_{g,NT1}(\gamma,\gamma^*) &\equiv \frac{1}{N_g T}\sum_{i\in\mathbf{G}_g^0}\sum_{t=1}^{T}\tilde{z}_{it}(\gamma)\tilde{z}_{it}(\gamma^*)'\varepsilon_{it}^2,\\
\Omega_{g,NT2}(\gamma,\gamma^*) &\equiv \frac{1}{N_g T}\sum_{i\in\mathbf{G}_g^0}\sum_{t=1}^{T}\sum_{s=1}^{T}\tilde{z}_{it}(\gamma)\tilde{z}_{is}(\gamma^*)'\varepsilon_{is}\varepsilon_{it}, \text{ and}\\
\mathbb{B}_{g,NT}(\gamma) &\equiv \frac{1}{N_g T}\sum_{i\in\mathbf{G}_g^0}\sum_{t=2}^{T}\sum_{s<t}E\left[z_{it}(\gamma)\varepsilon_{is}\right].
\end{aligned}$$

Assumption A.6: (i) For each $g\in\mathcal{G}$, the following probability limits exist and are finite: $\omega_g(\gamma,\gamma^*)=\text{plim}_{(N,T)\to\infty}\,\omega_{g,NT}(\gamma,\gamma^*)$, $\Omega_{g,\ell}(\gamma,\gamma^*)=\text{plim}_{(N,T)\to\infty}\,\Omega_{g,NT\ell}(\gamma,\gamma^*)$ for $\ell=1,2$, and $\mathbb{B}_g(\gamma)=\lim_{(N,T)\to\infty}\mathbb{B}_{g,NT}(\gamma)$.

(ii) $\omega_{g,NT}(\gamma,\gamma^*)\xrightarrow{p}\omega_g(\gamma,\gamma^*)$ and $\Omega_{g,NT\ell}(\gamma,\gamma^*)\xrightarrow{p}\Omega_{g,\ell}(\gamma,\gamma^*)$ for $\ell=1,2$ uniformly in $\gamma,\gamma^*\in\Gamma$.

Assumption A.6 imposes some conditions on the probability limits of random quantities that are associated with the asymptotic variance and bias of $\hat{\Theta}$. Here, we follow Hansen (2000) and assume directly that $\omega_{g,NT}$ and $\Omega_{g,NT\ell}$ for $\ell=1,2$ converge uniformly to some limits. The uniformity greatly facilitates the proofs of Theorem 3.3 below.

We establish the asymptotic distribution of our estimators in the following theorem.

Theorem 3.3 *Suppose that Assumptions A.1–A.6 hold with $\alpha\in(0,1/3)$ in Assumption A.1(vi). Let $\alpha_{N_gT}=(N_gT)^{1-2\alpha}$, $\omega_g^0=\omega_g(\gamma_g^0,\gamma_g^0)$, $\mathbb{B}_g^0=\mathbb{B}_g(\gamma_g^0)$, and $\Omega_{g,\ell}^0=\Omega_{g,\ell}(\gamma_g^0,\gamma_g^0)$ for $\ell=1,2$. Then for each $g\in\mathcal{G}$,*

(i)$\sqrt{N_gT}(\hat{\theta}_g-\theta_g^0)-(\omega_g^0)^{-1}\sqrt{\frac{N_g}{T}}\mathbb{B}_g^0\xrightarrow{d}\mathcal{N}(0,(\omega_g^0)^{-1}\Omega_{g,1}^0(\omega_g^0)^{-1})$ under Assumption A.1(i.1) and $\sqrt{N_gT}(\hat{\theta}_g-\theta_g^0)\xrightarrow{d}\mathcal{N}(0,(\omega_g^0)^{-1}\Omega_{g,2}^0(\omega_g^0)^{-1})$ under Assumption A.1(i.2);

(ii) $\alpha_{N_gT}(\hat{\gamma}_g-\gamma_g^0)\xrightarrow{d}\varpi_g\mathcal{T}_g$, where $\varpi_g=\frac{C_g^{0\prime}V_g^0C_g^0}{(C_g^{0\prime}D_g^0C_g^0)^2}$, $\mathcal{T}_g=\text{argmax}_{r\in\mathbb{R}}\left[-\frac{1}{2}|r|+W_g(r)\right]$, and $W_g(\cdot)$, $g\in\mathcal{G}$, are mutually independent two-sided Brownian motions.

Theorem 3.3 establishes the asymptotic distributions of the estimators of the slope and threshold coefficients. Note that we strengthen Assumption A.1(vi) slightly to require $\alpha\in(0,1/3)$. From the proof of Lemma A.7 that is used in the proof of the above theorem, we can easily find that such an extra condition is not needed if we only consider the case where $N/T\to\kappa$ for some $\kappa\in(0,\infty)$.

When we allow for dynamics in Assumption A.1(i.1), the estimator $\hat{\theta}_g$ of the group-specific slope coefficient θ_g^0 exhibits a bias term to be corrected as in standard dynamic panels. One can conduct the bias correction by estimating ω_g^0 and $\mathbb{B}_{g,0}$ consistently by

$$\hat{\omega}_g\equiv\frac{1}{\hat{N}_gT}\sum_{i\in\hat{\mathbf{G}}_g}\sum_{t=1}^{T}\tilde{z}_{it}(\hat{\gamma}_g)\tilde{z}_{it}(\hat{\gamma}_g)' \text{ and } \hat{\mathbb{B}}_g=\frac{1}{\hat{N}_gT}\sum_{i\in\hat{\mathbf{G}}_g}\sum_{t=2}^{T}\sum_{s<t}z_{it}(\hat{\gamma}_g)\hat{\varepsilon}_{is},$$

where $\hat{N}_g = \left|\hat{\mathbf{G}}_g\right|$ denotes the cardinality of $\hat{\mathbf{G}}_g$, $\hat{\mathbf{G}}_g \equiv \{i : \hat{g}_i = g\}$ for $g \in \mathcal{G}$, and $\hat{\varepsilon}_{it} = \tilde{y}_{it} - \tilde{z}_{it}(\hat{\gamma}_g)'\hat{\theta}_g$. Similarly, it is easy to show that a consistent estimator of the asymptotic variance of $\hat{\theta}_g$ in this case is given by $\hat{\omega}_g^{-1}\hat{\Omega}_{g,1}\hat{\omega}_g^{-1}$, where $\hat{\Omega}_{g,1} = \frac{1}{\hat{N}_g T}\sum_{i\in\hat{\mathbf{G}}_g}\sum_{t=1}^{T}\tilde{z}_{it}(\hat{\gamma}_g)\tilde{z}_{it}(\hat{\gamma}_g)'\hat{\varepsilon}_{it}^2$. When (X_i, q_i) is strictly exogenous in Assumption A.1(i.2), we allow for serial correlation in the error terms. In this case, we propose to estimate the asymptotic variance of $\hat{\theta}_g$ by $\hat{\omega}_g^{-1}\hat{\Omega}_{g,2}\hat{\omega}_g^{-1}$, where $\hat{\Omega}_{g,2}$ is a panel heteroskedasticity and autocorrelation consistent (HAC) estimator:

$$\hat{\Omega}_{g,2} = \frac{1}{\hat{N}_g}\sum_{i\in\hat{\mathbf{G}}_g}\left[\hat{\Lambda}_{i,0} + \sum_{s=1}^{J_T} w_{Ts}(\hat{\Lambda}_{is} + \hat{\Lambda}_{is}')\right],$$

where $w_{Ts} = 1 - |s|/J_T$, J_T satisfies $1/J_T + J_T^3/T \to 0$ as $T \to \infty$, and $\hat{\Lambda}_{is} = \frac{1}{T}\sum_{t=s+1}^{T}\tilde{z}_{it}(\hat{\gamma}_g)\tilde{z}_{i,t-s}(\hat{\gamma}_g)' \times\hat{\varepsilon}_{it}\hat{\varepsilon}_{i,t-s}$. Following Su and Jin (2012) and the results in Theorems 3.2–3.3, we can show that $\hat{\Omega}_{g,2}$ and $\hat{\omega}_g^{-1}\hat{\Omega}_{g,2}\hat{\omega}_g^{-1}$ are consistent estimators of $\Omega_{g,2}^0$ and $(\omega_g^0)^{-1}\Omega_{g,2}^0(\omega_g^0)^{-1}$, respectively.

Theorem 3.3(ii) indicates that the asymptotic distribution of $\hat{\gamma}_g$ is pivotal up to a scale parameter ϖ_g, which is similar to that given by Theorem 1 in Hansen (2000). It is well known that this result highly relies on the assumption that the threshold effect converges to zero as $(N, T) \to \infty$. Under the fixed-threshold-effect framework ($\alpha = 0$), it is possible to demonstrate $NT(\hat{\gamma}_g - \gamma_g^0) = O_p(1)$ but the asymptotic distribution of $\hat{\gamma}_g$ will not be asymptotically pivotal even after appropriate normalization. In addition, it is well known that the above scale parameter ϖ_g cannot be consistently estimated. To make inference on the threshold parameters, we propose to apply the likelihood ratio test in the next section.

4 Inference on the Threshold Parameter

In this section, we consider inference on the threshold parameter $\mathbf{D} = (\gamma_1, ..., \gamma_G)'$. We consider three cases. The first case is to test the null hypothesis on the threshold parameter γ_g for a single group $g \in \mathcal{G}$:

$$H_{01} : \gamma_g = \gamma_g^0 \text{ for some } \gamma_g^0 \in \Gamma.$$

Next, we consider testing the homogeneity of the threshold parameters:

$$H_{02} : \gamma_1^0 = ... = \gamma_G^0 = \gamma^0 \text{ for some } \gamma^0 \in \Gamma.$$

If one fails to reject the hypothesis of common threshold parameter for all groups, one can estimate the model with a common threshold parameter, γ, say. Then we can study the inference on the common threshold parameter

$$H_{03} : \gamma = \gamma^0 \text{ for some } \gamma^0 \in \Gamma.$$

4.1 Likelihood ratio test for a single γ_g

To test the null hypothesis $H_{01} : \gamma_g = \gamma_g^0$, a standard approach is to use the likelihood ratio (LR) test. If we know the true group structure, the likelihood ratio test statistic can be constructed as in Hansen (2000). In our framework, we need to construct the test statistic based on the estimated group structure $\{\hat{\mathbf{G}}_g, g \in \mathcal{G}\}$. Let $\bar{\theta}_g(\gamma) \equiv \operatorname{argmin}_{\theta \in \mathcal{B}} \bar{\mathcal{Q}}_g(\theta, \gamma)$, where $\bar{\mathcal{Q}}_g(\theta, \gamma) \equiv \sum_{i \in \hat{\mathbf{G}}_g} \sum_{t=1}^{T} [\tilde{y}_{it} - \tilde{z}_{it}(\gamma)'\theta]^2$. We follow the lead of Hansen (2000) and propose to employ the following LR test statistic for γ_g :

$$\mathcal{L}_{g,NT}(\gamma) \equiv \hat{N}_g T \frac{\bar{\mathcal{Q}}_g(\bar{\theta}_g(\gamma), \gamma) - \bar{\mathcal{Q}}_g(\hat{\theta}_g, \hat{\gamma}_g)}{\bar{\mathcal{Q}}_g(\hat{\theta}_g, \hat{\gamma}_g)}.$$

The major difference is that we consider the minimization of $\bar{\mathcal{Q}}_g(\theta, \gamma)$ instead of the infeasible version $\check{\mathcal{Q}}_g(\theta, \gamma)$. In the proof of Theorem 4.1 below, we show that $\bar{\mathcal{Q}}_g(\theta, \gamma)$ and $\check{\mathcal{Q}}_g(\theta, \gamma)$ are asymptotically equivalent so that we can study the asymptotic distribution of the LR test statistic based on the minimization of the infeasible objective function.

For each $g \in \mathcal{G}$, let $\sigma_g^2 = \lim_{(N,T)\to\infty} \frac{1}{N_g T} \sum_{i \in \mathbf{G}_g^0} \sum_{t=1}^{T} E(\varepsilon_{it}^2)$, $w_{g,V} = C_g^{0\prime} V_g^0 C_g^0$ and $w_{g,D} = C_g^{0\prime} D_g^0 C_g^0$. Let $\sigma^2 = \lim_{(N,T)\to\infty} \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} E(\varepsilon_{it}^2)$. The following theorem establishes the asymptotic null distribution of the above LR test statistic.

Theorem 4.1 *Suppose that Assumptions A.1–A.6 hold with $\alpha \in (0, 1/3)$ in Assumption A.1(vi). Then under $H_{01} : \gamma_g = \gamma_g^0$, we have*

$$\mathcal{L}_{g,NT}(\gamma_g^0) \xrightarrow{d} \eta_g^2 \xi_g \text{ for each } g \in \mathcal{G},$$

where $\eta_g^2 = \frac{w_{g,V}}{w_{g,D}\sigma_g^2}$ and $\xi_g = \max_{s \in \mathbb{R}}[2W_g(s) - |s|]$ has the distribution function characterized by $\Pr(\xi_g \le x) = (1 - e^{-x/2})^2$.

Theorem 4.1 indicates that the asymptotic distribution of the LR test statistic constructed from the estimated group structure is asymptotically equivalent to that of the infeasible test statistic obtained from the true group structure. Now, we still have a nuisance parameter η_g^2. In the special case where we have conditional homoskedasticity along both the cross-section and time dimensions, $\eta_g^2 = 1$ and the LR statistic is asymptotically free of any nuisance parameter. If heteroskedasticity is suspected, then we need to estimate η_g^2 consistently. Noting that

$$\eta_g^2 = \frac{\operatorname{plim}_{(N,T)\to\infty} \frac{1}{N_g T} \sum_{i \in \mathbf{G}_g^0} \sum_{t=1}^{T} E[(\delta_g^{0\prime} x_{it} \varepsilon_{it})^2 | q_{it} = \gamma_g^0] f_{it}\left(\gamma_g^0\right)}{\sigma_g^2 \operatorname{plim}_{(N,T)\to\infty} \frac{1}{N_g T} \sum_{i \in \mathbf{G}_g^0} \sum_{t=1}^{T} E[(\delta_g^{0\prime} x_{it})^2 | q_{it} = \gamma_g^0] f_{it}\left(\gamma_g^0\right)},$$

we propose to estimate η_g^2 by

$$\hat{\eta}_g^2 = \frac{\sum_{i \in \hat{\mathbf{G}}_g} \sum_{t=1}^{T} K_h(\hat{\gamma}_g - q_{it})(\hat{\delta}_g' x_{it} \hat{\varepsilon}_{it})^2}{\hat{\sigma}_g^2 \sum_{i \in \hat{\mathbf{G}}_g} \sum_{t=1}^{T} K_h(\hat{\gamma}_g - q_{it})(\hat{\delta}_g' x_{it})^2},$$

where $\hat{\sigma}_g^2 = \tilde{\mathcal{Q}}_g(\hat{\theta}_g, \hat{\gamma}_g)/(\hat{N}_g T)$, $K_h(u) = h^{-1}K(u/h)$, $h \to 0$ is the bandwidth parameter and $K(\cdot)$ is a kernel function. We can readily show that $\hat{\sigma}_g^2 = \sigma_g^2 + o_p(1)$ and $\hat{\eta}_g^2 = \eta_g^2 + o_p(1)$ under some standard weak conditions on h and $K(\cdot)$. Given the consistent estimate of η_g^2, we can consider the normalized LR statistic

$$\mathcal{NL}_{g,NT}(\gamma_g^0) = \mathcal{L}_{g,NT}(\gamma_g^0)/\hat{\eta}_g^2.$$

We can easily tabulate the asymptotic critical value for $\mathcal{NL}_{g,NT}(\gamma_g^0)$. We can also invert this statistic to obtain the asymptotic $1-a$ confidence interval for γ :

$$CI_{1-a} = \left\{\gamma \in \Gamma : \mathcal{NL}_{g,NT} \leq \xi_{1-a}\right\},$$

where ξ_{1-a} denotes the $1-a$ percentile of ξ. For example, $\xi_{1-\alpha} = 5.94$, 7.35, and 10.59 for $a = 0.10$, 0.05, and 0.01, respectively.

4.2 Test for a common threshold parameter

In applications, it is likely that all individuals share a common threshold parameter, although their slope coefficients may still vary across groups. In this case, estimating the model with the common-threshold restriction imposed improves the asymptotic efficiency of the threshold estimator. Thus motivated, one may wish to test the homogeneity of the threshold parameter prior to estimation. In this section, we consider testing the null hypothesis

$$H_{02} : \gamma_1^0 = ... = \gamma_G^0 = \gamma^0 \text{ for some } \gamma^0 \in \Gamma.$$

Let $\mathcal{D}_r = \{\mathbf{D} = (\gamma, ..., \gamma)', \gamma \in \Gamma\} \subseteq \Gamma^G$ be the restricted parameter space and $\mathbf{D}_{r,\gamma} \equiv (\gamma, ..., \gamma)' \in \mathcal{D}_r$. Then the null hypothesis can be equivalently rewritten as $H_{02} : \mathbf{D}^0 \in \mathcal{D}_r$. We can estimate the model by restricting $\mathbf{D} \in \mathcal{D}_r$ under H_{02} :

$$(\hat{\Theta}_r, \hat{\mathbf{D}}_r, \hat{\mathbf{G}}_r) = \underset{(\Theta,\mathbf{D},\mathbf{G})\in\mathcal{B}^G\times\mathcal{D}_r\times\mathcal{G}^N}{\operatorname{argmin}} \mathcal{Q}(\Theta, \mathbf{D}, \mathbf{G}).$$

Then we can construct the LR test statistic by

$$\mathcal{L}_{NT} = NT \frac{\mathcal{Q}(\hat{\Theta}_r, \hat{\mathbf{D}}_r, \hat{\mathbf{G}}_r) - \mathcal{Q}(\hat{\Theta}, \hat{\mathbf{D}}, \hat{\mathbf{G}})}{\mathcal{Q}(\hat{\Theta}, \hat{\mathbf{D}}, \hat{\mathbf{G}})}.$$

The following theorem studies the asymptotic distribution of $\mathcal{L}_{NT}$ under H_{02}.

Theorem 4.2 *Suppose that Assumptions A.1–A.6 hold with $\alpha \in (0, 1/3)$ in Assumption A.1(vi). Under the null hypothesis $H_{02} : \mathbf{D}^0 \in \mathcal{D}_r$, we have*

$$\mathcal{L}_{NT} \xrightarrow{d} \sum_{g=1}^{G} \tilde{\eta}_g^2 \max_{s_g\in\mathbb{R}} [2W_g(s_g) - |s_g|] - \max_{s\in\mathbb{R}} \left[\sum_{g=1}^{G} \tilde{\eta}_g^2 \left(2W_g(\rho_g s) - \left|\rho_g s\right|\right)\right] \equiv \Xi,$$

where $\rho_g = \frac{w_{g,D}}{w_{1,D}} \pi_g / \tilde{\eta}_g^2$, and $\tilde{\eta}_g^2 = w_{g,V}/(w_{g,D}\sigma^2)$.

Theorem 4.2 indicates that the limiting distribution Ξ of $\mathcal{L}_{NT}$ involves two sets of nuisance parameters, viz, $\tilde{\eta}_g^2$ and ρ_g for $g \in \mathcal{G}$. Under conditional homoskedasticity, we have $\tilde{\eta}_g^2 = 1$ for each g. If heteroskedasticity is suspected, then we need to estimate $\tilde{\eta}_g^2$ consistently. If ρ_g is homogeneous across g, we do not need to estimate it. However, ρ_g is generally not homogeneous across g and we need to estimate it via estimating $\tilde{\eta}_g^2$, $\frac{w_{g,D}}{w_{1,D}}$, and π_g. Using Theorem 3.1, it is easy to show that a consistent estimator of π_g is given by $\hat{\pi}_g = \hat{N}_g/N$. Noting that $\tilde{\eta}_g^2 = \frac{\sigma_g^2}{\sigma^2}\eta_g^2$ and

$$\frac{w_{g,D}}{w_{1,D}} = \frac{\text{plim}_{(N,T)\to\infty}\frac{1}{N_g T}\sum_{i\in\mathbf{G}_g^0}\sum_{t=1}^T E[(\delta_g^{0\prime}x_{it})^2|q_{it}=\gamma^0]f_{it}\left(\gamma^0\right)}{\text{plim}_{(N,T)\to\infty}\frac{1}{N_1 T}\sum_{i\in\mathbf{G}_1^0}\sum_{t=1}^T E[(\delta_1^{0\prime}x_{it})^2|q_{it}=\gamma^0]f_{it}\left(\gamma^0\right)},$$

we propose to estimate $\tilde{\eta}_g^2$ and $\frac{w_{g,D}}{w_{1,D}}$ respectively by

$$\widehat{\tilde{\eta}}_g^2 = \hat{\eta}_g^2\frac{\hat{\sigma}_g^2}{\hat{\sigma}^2} \quad \text{and} \quad \frac{\hat{w}_{g,D}}{\hat{w}_{1,D}} = \frac{\frac{1}{\hat{N}_g T}\sum_{i\in\hat{\mathbf{G}}_g}\sum_{t=1}^T K_h(\hat{\gamma}_g - q_{it})(\hat{\delta}_g^{\prime}x_{it})^2}{\frac{1}{\hat{N}_g T}\sum_{i\in\hat{\mathbf{G}}_1}\sum_{t=1}^T K_h(\hat{\gamma}_1 - q_{it})(\hat{\delta}_1^{\prime}x_{it})^2},$$

where $\hat{\sigma}^2 = \frac{1}{NT}\sum_{g=1}^G\sum_{i\in\hat{\mathbf{G}}_g}\sum_{t=1}^T[\tilde{y}_{it} - \tilde{z}_{it}(\hat{\gamma}_g)^\prime\hat{\theta}_g]^2$. It is easy to show that the above estimators are consistent under standard conditions and a consistent estimator of ρ_g is given by $\hat{\rho}_g = \frac{\hat{w}_{g,D}}{\hat{w}_{1,D}}\hat{\pi}_g/\widehat{\tilde{\eta}}_g^2$. To find out the p-value, we can simulate the asymptotic distribution with these estimates. Basically, we can simulate G independent two-sided Brownian motions $W_g(\cdot)$'s and construct the corresponding statistic where the nuisance parameters are replaced with their consistent estimates. Simulating a large number of times, we can mimic the asymptotic distribution sufficiently well. Then, we can reject the null hypothesis at the prescribed α level, if the test statistic is larger than $1-\alpha$ quantile of the simulated distribution.

4.3 Likelihood ratio test for a common threshold parameter

Suppose we have common threshold parameters, we can use the restricted estimator $(\hat{\Theta}_r, \hat{\mathbf{D}}_r, \hat{\mathbf{G}}_r)$ defined in the last subsection. Even in this case, the estimators of the group-specific slope coefficients share the same asymptotic distribution as the unrestricted estimators studied in the last section due to the asymptotic independence between the estimators of the slope coefficients and that of the threshold parameter.

To make inference on the common threshold parameter γ, we also consider an LR test for $H_{03}: \gamma = \gamma^0$. The LR test statistic is now defined by

$$\mathcal{L}_{NT}^c(\gamma) = NT\frac{\mathcal{Q}(\hat{\Theta}(\mathbf{D}_{r,\gamma}, \hat{\mathbf{G}}_r), \mathbf{D}_{r,\gamma}, \hat{\mathbf{G}}_r) - \mathcal{Q}(\hat{\Theta}_r, \hat{\mathbf{D}}_r, \hat{\mathbf{G}}_r)}{\mathcal{Q}(\hat{\Theta}_r, \hat{\mathbf{D}}_r, \hat{\mathbf{G}}_r)},$$

where $\hat{\Theta}(\mathbf{D}_{r,\gamma}, \hat{\mathbf{G}}_r)$ is defined as in Section 2.1 and the superscript c is an abbreviation for "common". Note that $H_{03}: \gamma = \gamma^0$ can be equivalently rewritten as $H_{03}: \mathbf{D}^0 = \mathbf{D}_{r,\gamma^0}$.

The next theorem establishes the asymptotic distribution of $\mathcal{L}_{NT}^c(\gamma)$ under H_{03}.

Theorem 4.3 *Suppose that Assumptions A.1–A.6 hold with $\alpha \in (0, 1/3)$ in Assumption A.1(vi). Under the null $H_{03} : \mathbf{D}^0 = \mathbf{D}_{r,\gamma^0}$, we have*

$$\mathcal{L}^c_{NT}(\gamma^0) \xrightarrow{d} \eta^2 \max_{s\in\mathbb{R}} [W(s) - |s|],$$

where $\eta^2 = \left(\sum_{g=1}^G \pi_g w_{g,V}\right) / \left(\sigma^2 \sum_{g=1}^G \pi_g w_{g,D}\right)$.

Like before, we can estimate the nuisance parameter η^2 consistently by the nonparametric estimator:

$$\hat{\eta}^2 = \frac{\sum_{i=1}^N \sum_{t=1}^T K_h(\hat{\gamma} - q_{it})(\hat{\delta}'_{\hat{g}_i} x_{it}\hat{\varepsilon}_{it})^2}{\hat{\sigma}^2 \sum_{i=1}^N \sum_{t=1}^T K_h(\hat{\gamma} - q_{it})(\hat{\delta}'_{\hat{g}_i} x_{it})^2},$$

where $\hat{\gamma}$ is the estimator of the common threshold parameter γ under H_{02}, $K_h(u) = h^{-1}K(u/h)$, $h \to 0$ is the bandwidth parameter and $K(\cdot)$ is a kernel function.

5 Test for the Presence of Threshold Effect

In application, one may suspect that a set of groups do not have the threshold effect. In this case, we can verify the existence of threshold effects for $P \leq G$ groups by testing the null hypothesis

$$\mathbb{H}_0 : \delta^0_{g_1} = ... = \delta^0_{g_P} = 0$$

versus the alternative hypothesis $\mathbb{H}_1 : \delta^0_{g_l} \neq 0$ for some $g_l \in \mathcal{G}_s$, where $\mathcal{G}_s \equiv \{g_l, l = 1, ..., P\} \subset \mathcal{G}$. To study the local power of our test, we consider the following sequence of Pitman local alternatives

$$\mathbb{H}_{1NT} : \delta^0_{g_l} = c_l/\sqrt{NT} \text{ for } g_l \in \mathcal{G}_s.$$

Let $\mathbf{c} \equiv (c'_1, ..., c'_P)'$ and $\mathbb{L} \equiv (e_{g_1}, ..., e_{g_P})' \otimes L$, where $\otimes$ denotes Kronecker product, $L \equiv [\mathbf{0}_{K\times K}, I_K]$ and e_{g_l} is a $G \times 1$ vector with g_lth entry being 1 and other entries equal to zero. Then we can rewrite the null and local alternative hypotheses respectively as

$$\mathbb{H}_0 : \mathbb{L}\Theta^0 = \mathbf{0}_{KP\times 1} \text{ and } \mathbb{H}_{1NT} : \mathbb{L}\Theta^0 = \mathbf{c}/\sqrt{NT}.$$

Note that $\mathbf{c} = \mathbf{0}_{KP\times 1}$ corresponds to the null hypothesis of no threshold effects and we allow $\delta^0_{g_l}$ for $g_l \in \mathcal{G}_s$ to shrink to zero at the $(NT)^{-1/2}$-parametric rate under the local alternative. Under $\mathbb{H}_{1NT}$, the early estimators of Θ^0 and $\mathbf{G}^0$ continue to be consistent with any $\mathbf{D} \in \Gamma^G$ despite the fact that we cannot identify $\mathbf{D}^0$.

As we do not know the true group structure, we need to rely on the estimated group structure $\hat{\mathbf{G}}$. For any fixed $\mathbf{D}$ and a preliminary estimate of group structure $\hat{\mathbf{G}}$, we can obtain the bias-corrected estimator $\bar{\Theta}^{\text{bc}}(\mathbf{D}, \hat{\mathbf{G}}) = (\bar{\theta}^{\text{bc}}_1 (\gamma_1)', ..., \bar{\theta}^{\text{bc}}_G (\gamma_G)')'$. Let

$$\hat{\Pi} = \text{diag}(\hat{\pi}_1, ..., \hat{\pi}_G) \otimes I_{2K} \text{ and } \hat{\mathbb{K}}_{NT}(\mathbf{D}) = \mathbb{L}\hat{\boldsymbol{\omega}}(\mathbf{D})^{-1}\hat{\boldsymbol{\Omega}}(\mathbf{D})\hat{\boldsymbol{\omega}}(\mathbf{D})^{-1}\mathbb{L}',$$

where

$$\hat{\boldsymbol{\omega}}(\mathbf{D}) = \begin{bmatrix} \hat{\omega}_1(\gamma_1,\gamma_1) & & \\ & \ddots & \\ & & \hat{\omega}_G(\gamma_G,\gamma_G) \end{bmatrix} \text{ and } \hat{\boldsymbol{\Omega}}(\mathbf{D}) = \begin{bmatrix} \hat{\Omega}_{1,1}(\gamma_1,\gamma_1) & & \\ & \ddots & \\ & & \hat{\Omega}_{G,1}(\gamma_G,\gamma_G) \end{bmatrix}.$$

We can construct the sup-Wald test statistic $\mathcal{W}_{NT} = \sup_{\mathbf{D}\in\Gamma^G} W_{NT}(\mathbf{D})$, where

$$W_{NT}(\mathbf{D}) = NT \cdot \bar{\Theta}^{\mathrm{bc}}(\mathbf{D},\hat{\mathbf{G}})'\hat{\Pi}^{1/2}\mathbb{L}'\left(\hat{\mathbb{K}}_{NT}(\mathbf{D})\right)^{-1}\mathbb{L}\hat{\Pi}^{1/2}\bar{\Theta}^{\mathrm{bc}}(\mathbf{D},\hat{\mathbf{G}}).$$

Let $S_{g,NT}(\gamma) = \frac{1}{\sqrt{N_gT}}\sum_{i\in\mathbf{G}_g^0}\sum_{t=1}^T\{z_{it}(\gamma) - \frac{1}{T}\sum_{s=1}^T E[z_{is}(\gamma)]\}\varepsilon_{it}$. Let $S_g(\gamma)$ be a zero mean Gaussian process with covariance kernel $\Omega_g(\gamma,\gamma^*)$. Let $\mathbb{K}(\mathbf{D}) = \mathbb{L}\boldsymbol{\omega}(\mathbf{D})^{-1}\boldsymbol{\Omega}(\mathbf{D})\boldsymbol{\omega}(\mathbf{D})^{-1}\mathbb{L}'$, $\overline{\mathbf{S}}(\mathbf{D}) = \mathbb{L}\boldsymbol{\omega}(\mathbf{D})^{-1}\mathbf{S}(\mathbf{D})$, $\mathbf{S}(\mathbf{D}) = (S_1(\gamma_1)',...,S_G(\gamma_G)')'$, and $\overline{\mathbf{Q}}(\mathbf{D}) = \mathbb{L}\boldsymbol{\omega}(\mathbf{D})^{-1}\mathbf{Q}(\mathbf{D})\Pi^{1/2}\mathbb{L}'$, where $\Pi = \mathrm{diag}(\pi_1, ..., \pi_G)\otimes I_{2K}$,

$$\mathbf{Q}(\mathbf{D}) = \begin{bmatrix} \omega_1(\gamma_1,\gamma_1^0) & & \\ & \ddots & \\ & & \omega_G(\gamma_G,\gamma_G^0) \end{bmatrix}, \boldsymbol{\Omega}(\mathbf{D}) = \begin{bmatrix} \Omega_{1,1}(\gamma_1,\gamma_1) & & \\ & \ddots & \\ & & \Omega_{G,1}(\gamma_G,\gamma_G) \end{bmatrix}, \text{ and}$$

$$\boldsymbol{\omega}(\mathbf{D}) = \begin{bmatrix} \omega_1(\gamma_1,\gamma_1) & & \\ & \ddots & \\ & & \omega_G(\gamma_G,\gamma_G) \end{bmatrix}.$$

To state the next theorem, we add one assumption.

Assumption A.7: For each $g\in\mathcal{G}$, $S_{g,NT}(\gamma)\Rightarrow S_g(\gamma)$ on the compact set Γ, where $\Rightarrow$ denotes the usual weak convergence.

The following theorem establishes the asymptotic distribution of our sup-Wald test statistic under $\mathbb{H}_{1NT}$.

Theorem 5.1 *Suppose that Assumptions A.1(i.1) and (ii)–(v), and A.2–A.7 hold. Then under* $\mathbb{H}_{1NT}: \mathbb{L}\Theta^0 = \mathbf{c}/\sqrt{NT}$, *we have*

$$\mathcal{W}_{NT} \xrightarrow{d} \sup_{\mathbf{D}\in\Gamma^G} W^{\mathbf{c}}(\mathbf{D}),$$

where $W^{\mathbf{c}}(\mathbf{D}) = \left[\overline{\mathbf{S}}(\mathbf{D}) + \overline{\mathbf{Q}}(\mathbf{D})\mathbf{c}\right]'[\mathbb{K}(\mathbf{D})]^{-1}\left[\overline{\mathbf{S}}(\mathbf{D}) + \overline{\mathbf{Q}}(\mathbf{D})\mathbf{c}\right]$.

Under $\mathbb{H}_0$, $\mathbf{c} = 0$ and $\overline{w}^0 \equiv \sup_{\mathbf{D}\in\Gamma^G} W^0(\mathbf{D}) = \sup_{\mathbf{D}\in\Gamma^G}\overline{\mathbf{S}}(\mathbf{D})'[\mathbb{K}(\mathbf{D})]^{-1}\overline{\mathbf{S}}(\mathbf{D})$. Clearly, the limiting null distribution of $\mathcal{W}_{NT}$ depends on the Gaussian process $\overline{\mathbf{S}}(\mathbf{D})$ and is not pivotal. We cannot tabulate the asymptotic critical values for the above sup-Wald statistic. Nevertheless, given the simple structure of $\overline{\mathbf{S}}(\mathbf{D})$, we can follow the literature (e.g., Hansen 1996) and simulate the critical values via the following procedure:

1. Generate $\{v_{it}, i = 1,...,N, t = 1,...,T\}$ independently from the standard normal distribution;

2. Calculate $\hat{\mathbf{S}}_{g,NT}(\mathbf{D}) = \frac{1}{\sqrt{\hat{N}_g T}} \sum_{i\in\hat{\mathbf{G}}_g} \sum_{t=1}^{T} \tilde{z}_{it}(\gamma_g)\hat{\varepsilon}_{it}(\gamma_g)v_{it}$;

3. Compute $\mathcal{W}_{NT}^{*} \equiv \sup_{\mathbf{D}\in\Gamma^G} \hat{\mathbf{S}}(\mathbf{D})'\boldsymbol{\omega}(\mathbf{D})^{-1}\mathbb{L}'[\hat{\mathbb{K}}_{NT}(\mathbf{D})]^{-1}\mathbb{L}\hat{\boldsymbol{\omega}}(\mathbf{D})^{-1}\hat{\mathbf{S}}(\mathbf{D})$;

4. Repeat Steps 1–3 B times and denote the resulting $\mathcal{W}_{NT}^{*}$ test statistics as $\mathcal{W}_{NT,j}^{*}$ for $j = 1,...,B$.

5. Calculate the simulated/bootstrap p-value for the $\mathcal{W}_{NT}$ test as $p_W^* = \frac{1}{B}\sum_{j=1}^{B} \mathbf{1}\{\mathcal{W}_{NT,j}^{*} \geq \mathcal{W}_{NT}\}$ and reject the null when p_W^* is smaller than some prescribed level of significance.

The above discussion was based on the m.d.s. condition in Assumption A.1(i.1). If we consider the case of static panels such that Assumption A.1(i.2) holds, then the covariance kernel is given by $\Omega_g(\gamma,\gamma^*) = \lim_{(N,T)\to\infty} \frac{1}{N_g T} \sum_{i\in\mathbf{G}_g^0} \sum_{s=1}^{T} \sum_{t=1}^{T} E[\tilde{z}_{is}(\gamma)\tilde{z}_{it}(\gamma^*)'\varepsilon_{it}\varepsilon_{is}]$ for $g \in \mathcal{G}$. Now, the above simulation procedure needs to be modified because $\hat{\mathbf{S}}_{g,NT}(\mathbf{D})$ constructed in Step 2 will not mimic the Gaussian process $\overline{\mathbf{S}}(\mathbf{D})$ in this case. Instead of generating the independently and identically distributed (i.i.d.) standard normal random variables $\{v_{it}\}$ in Step 1, we can generate $v_i = (v_{i1},...,v_{iT})'$ independently from a zero mean multivariate normal distribution with the variance-covariance matrix $\Sigma = \{\sigma_{ts}\}$ given by $\sigma_{ts} = [1 - (|t-s|/p_T)]\,\mathbf{1}(|t-s| \leq p_T)$ for some p_T such that $1/p_T + p_T^3/T \to 0$. Then

$$E_w\left[\hat{\mathbf{S}}_{g,NT}(\mathbf{D})\hat{\mathbf{S}}_{g,NT}(\mathbf{D})'\right] = \frac{1}{\hat{N}_g T}\sum_{i\in\hat{G}_g}\sum_{t=1}^{T}\sum_{s=1}^{T} k(\frac{t-s}{p_T})\tilde{z}_{it}(\gamma_g)\tilde{z}_{is}(\gamma_g)\hat{\varepsilon}_{it}(\gamma_g)\hat{\varepsilon}_{is}(\gamma_g),$$

where $E_w(\cdot)$ denotes the expectation conditional on the sample $w \equiv \{x_{it}, q_{it}, \varepsilon_{it},\ i = 1,...,N, t = 1,...,T\}$ and $k(s) = [1 - |s|/p_T]\,\mathbf{1}(|s| \leq p_T)$. Apparently, $E_w[\hat{\mathbf{S}}_{g,NT}(\mathbf{D})\hat{\mathbf{S}}_{g,NT}(\mathbf{D})']$ converges in probability to $\Omega_g(\gamma_g,\gamma_g)$ and the modified simulation procedure will generate statistics that follow the same asymptotic distribution as that of $\mathcal{W}_{NT}$.

In practice, we frequently consider testing the presence of threshold effects in all G groups, that is, testing $\mathbb{H}_0 : \delta_1^0 = ... = \delta_G^0 = 0$. In this case, $\mathbb{L} = I_G \otimes L$ and we can readily rewrite our Wald statistic $\mathcal{W}_{NT}$ as

$$\mathcal{W}_{NT} = \sup_{(\gamma_1,...,\gamma_G)\in\Gamma^G} \sum_{g=1}^{G} W_{gNT}\left(\gamma_g\right) \equiv \mathcal{W}_{NT}^{\mathrm{sum}},$$

where $W_{gNT}\left(\gamma_g\right) = \hat{N}_g T \cdot \bar{\delta}_g^{\mathrm{bc}}\left(\gamma_g\right)' [\hat{\mathbf{K}}_{gNT}(\gamma_g)]^{-1}\bar{\delta}_g^{\mathrm{bc}}\left(\gamma_g\right)$, $\hat{\mathbf{K}}_{gNT}(\gamma_g) = L\hat{\omega}_g(\gamma_g,\gamma_g)^{-1}\hat{\Omega}_{g,1}(\gamma_g,\gamma_g)$ $\times\hat{\omega}_g(\gamma_g,\gamma_g)^{-1}L'$, and $\bar{\delta}_g^{\mathrm{bc}}\left(\gamma_g\right) = L\bar{\theta}_g^{\mathrm{bc}}\left(\gamma_g\right)$. Here, $W_{gNT}\left(\gamma_g\right)$ is the Wald statistic used for testing whether $\delta_g^0 = 0$ for the gth group. For this reason, we can also refer to $\mathcal{W}_{NT}$ as a sup-sum-type of Wald statistic ($\mathcal{W}_{NT}^{\mathrm{sum}}$). Alternatively, we can also consider a sup-sup-type of Wald statistic:

$$\mathcal{W}_{NT}^{\mathrm{sup}} = \sup_{1\leq g\leq G}\sup_{\gamma_g\in\Gamma} W_{gNT}\left(\gamma_g\right).$$

Following the proof of Theorem 5.1, we can readily find the limiting null distribution of $\mathcal{W}_{NT}^{\mathrm{sup}}$. As before, when we allow for serial correlation in the error terms, we should use $\hat{\Omega}_{g,2}$ in place of $\hat{\Omega}_{g,1}$ and modify the simulation procedure correspondingly to obtain the simulated p-values. We will compare the performance of $\mathcal{W}_{NT}^{\mathrm{sum}}$ with that of $\mathcal{W}_{NT}^{\mathrm{sup}}$ via simulations in Section 7.

6 Determining the Number of Groups

In practice, the true number of groups G^0 is typically unknown. In this case, we can consider a BIC-type information criterion (IC) to determine the number of groups. Following BM (2015) and SSP (2016), we consider the following IC:

$$IC(G) = \ln(\hat{\sigma}^2(G)) + \lambda_{NT} GK, \tag{6.1}$$

where $\hat{\sigma}^2(G) = (NT)^{-1} \mathcal{Q}(\hat{\Theta}^{(G)}, \hat{\mathbf{D}}^{(G)}, \hat{\mathbf{G}}^{(G)})$, where we make the dependence of $\hat{\Theta}, \hat{\mathbf{D}}, \hat{\mathbf{G}}$ on the group number G explicit, and λ_{NT} is a tuning parameter that plays the role of $\ln(NT)/(NT)$ in the standard BIC for linear panel data models. The estimated number of groups is given by

$$\hat{G} = \underset{G \in \{1,\ldots,G_{\max}\}}{\operatorname{argmin}} IC(G),$$

where $G_{\max}$ is an upper bound for G^0 that does not grow with (N, T). Following the arguments in SSP (2016), we can readily show that $\Pr(\hat{G} < G^0) \to 0$ provided $\lambda_{NT} = o(1)$ under the standard condition that $\hat{\sigma}^2(G) \xrightarrow{p} \sigma^2(G) > \sigma^2$ whenever $G < G^0$. This implies that $\hat{G} \geq G^0$ w.p.a.1. As in BM (2015), it is difficult to further show that $\Pr(\hat{G} = G^0) \to 1$ as $(N, T) \to \infty$ without further restrictions given the use of the K-means-type iterative algorithm in our estimation procedure.

On the other hand, if we require each estimated group should contain a minimum proportion ν of individuals (e.g., $\nu = 0.05$),[1] then we can show that when $G > G^0$, the threshold parameters and slope coefficients can also be estimated consistently and it is possible to show that $\hat{\sigma}^2(G) - \hat{\sigma}^2(G^0) = O_p(T^{-1})$ under some conditions stated in the online supplement. In this case, a choice of λ_{NT} such that $T \cdot \lambda_{NT} \to \infty$ as $(N, T) \to \infty$ would help to eliminate the over-selected model. Then we can prove the following theorem.

Theorem 6.1 *Suppose that Assumptions A.1–A.5 hold. Suppose that Assumptions D.1-D.2 in the online supplement holds. Then* $\Pr(\hat{G} = G^0) \to 1$ *as* $(N, T) \to \infty$.

Theorem 6.1 shows that the use of the IC helps to determine the correct number of groups w.p.a.1. SSP and Liu et al. (2018) propose a similar IC to ours. SSP also require that $\lambda_{NT} \to 0$ and $\lambda_{NT} T \to \infty$ as $(N, T) \to \infty$ for general nonlinear models but remark that this condition can be relaxed substantially for linear panel data models. In contrast, Liu et al. (2018) require that $\lambda_{NT} \to 0$ and $\lambda_{NT} T^{\frac{1}{2(1+\epsilon)}} \to \infty$ for some $\epsilon > 0$, which is much stronger than our requirement on λ_{NT}. The main reason is that they consider general nonlinear regression models and do not explore the properties of their objective function. They suggests using the tuning parameter $\lambda_{NT} \asymp T^{-1/4}$, which satisfies our theoretical requirement but tends to be too large to be useful in practice. In the simulations in the next section, we find that by setting $\lambda_{NT} = 0.1 \ln(NT)/T$, the above IC works fairly well in determining the true number of groups.

[1] If a group contains less than $\lfloor \nu N \rfloor$ members, the members in this group can be merged into other groups.

7 Monte Carlo Simulations

In this section we evaluate the finite sample performance of our tests and estimates via a set of Monte Carlo experiments.

7.1 Data generation processes

We consider three main cases. The first two cases concern static panels with different error structures, and the third case examines the dynamic panel. In each case, we consider two subcases that differ regarding whether the threshold value is group specific or common across individual units. Thus, we have six data generating processes (DGPs) in total.

DGP 1: We generate the data from the following static panel structure model:

$$y_{it} = \mu_i + \beta_{1,g_i} x_{it} \mathbf{1}(q_{it} \leq \gamma_{g_i}) + \beta_{2,g_i} x_{it} \mathbf{1}(q_{it} > \gamma_{g_i}) + \varepsilon_{it}, \tag{7.1}$$

where $\mu_i = T^{-1} \sum_{t=1}^{T} x_{it}$, and we generate x_{it} from an i.i.d. standard normal distribution. The slope coefficient vector $\beta_{g_i} = (\beta'_{1,g_i}, \beta'_{2,g_i})'$ has a group pattern of heterogeneity with the number of groups $G = 3$, and it is specified as

$$(\beta_{1,1}, \beta_{1,2}, \beta_{1,3}) = (1, 1.75, 2.5), \quad \text{and} \quad (\beta_{2,1}, \beta_{2,2}, \beta_{2,3}) = (1, 1.75, 2.5) + c_1 (NT)^{-0.1},$$

where c_1 controls the size of the threshold effect and we set $c_1 = 1$ if not especially mentioned. Let π_g be the proportion of units in group g for $g = 1, 2, 3$, and we fix the ratio of units among groups such that $\pi_1 : \pi_2 : \pi_3 = 0.3 : 0.3 : 0.4$. The threshold variable q_{it} follows i.i.d. $N(1, 1)$. The error term ε_{it} is heteroskedastic, generated as $\varepsilon_{it} = \sigma_{it} e_{it}$, where $\sigma_{it} = (s + 0.1 x_{it}^2)^{1/2}$, with s controlling for the signal-to-noise ratio, and $e_{it} \sim$ i.i.d. $N(0, 1)$. We set $s = 0.5$, leading to R^2 of about 0.85. Let $\mathbf{D} = (\gamma_1, \gamma_2, \gamma_3)'$. We consider two subcases: group-specific and homogeneous threshold value, i.e.

$$\text{DGP 1.1}: \ \mathbf{D} = (0.5, 1, 1.5)', \qquad \text{DGP 1.2}: \ \mathbf{D} = (1, 1, 1)'.$$

DGP 2: This is the same as DGP 1 except that the error term is generated from an autoregressive process,

$$\varepsilon_{it} = 0.4 \varepsilon_{it-1} + e_{it}, \quad e_{it} \sim \text{ i.i.d. } N(0, 1).$$

As above, we consider two subcases, with group-specific and homogeneous threshold values, and we label these two subcases DGP 2.1 and DGP 2.2, respectively.

DGP 3: In this case, we consider dynamic panel data models,

$$y_{it} = \mu_i + (\alpha_{1,g_i}, \beta_{1,g_i}) X_{it} \mathbf{1}(q_{it} \leq \gamma_{g_i}) + (\alpha_{2,g_i}, \beta_{2,g_i}) X_{it} \mathbf{1}(q_{it} > \gamma_{g_i}) + \varepsilon_{it}, \tag{7.2}$$

where $X_{it} = (y_{i,t-1}, x_{it})'$ and $\mu_i = T^{-1} \sum_{t=1}^{T} x_{it}$. The slope coefficient of $y_{i,t-1}$ is set as

$$(\alpha_{1,1}, \alpha_{1,2}, \alpha_{1,3}) = (0.2, 0.4, 0.6), \quad \text{and} \quad (\alpha_{2,1}, \alpha_{2,2}, \alpha_{2,3}) = (0.2, 0.4, 0.6) + c_2 (NT)^{-0.1},$$

with $c_2 = 1/4$ if not especially mentioned. The slope coefficient β_{g_i}, the threshold variable q_{it}, and the error term ε_{it} are all generated in the same manner as that in DGP 1. We likewise consider two subcases with different types of threshold values, and we label them DGP 3.1 and DGP 3.2.

For each DGP, we consider two cross-sectional sample sizes, $N = (50, 100)$, and two time series periods, $T = (30, 60)$, leading to four combinations of cross-sectional and time series dimensions. The number of replications is set to 1000 for the estimation and 500 for the hypothesis testing.

7.2 Determining the number of groups

As both of our testing and estimation procedures require specifications of the number of groups, we first examine the accuracy of the IC in determining the number of groups, measured by the empirical probability of selecting a particular number. The proposed IC is calculated by assuming the presence of the threshold effect. Nevertheless, researchers typically do not have prior knowledge of the existence of the threshold effect, and tests for the threshold effect in turn require the input of the number of groups. Therefore, we examine the performance of IC for the PSTR model in both scenarios with and without the threshold effect ($c_1 = 1$ and $c_2 = 1/4$ in the former case and $c_1 = c_2 = 0$ in the latter). In practice, we need to choose an appropriate λ_{NT} for the information criterion. We experiment with many alternatives and find that $\lambda_{NT} = 0.1 \ln(NT)/T$ works fairly well.

TABLE 1 around here.

Table 1 displays the empirical probability of selecting a particular number of groups in the three DGPs, and the highest probability in each case is highlighted in bold. The left panel displays the selection frequency when there is no threshold effect but only group-specific slope coefficients, and the right panel considers the cases in the presence of the threshold effect. In both cases, our IC can select the correct number of groups with a large probability, more than 96% in all cases, and this probability increases as either N or T increases. This result suggests that the proposed IC can correctly determine the number of groups regardless whether the there is a threshold effect, and this further allows us to implement tests and estimation given the true number of groups.

7.3 Test for the existence of threshold effect

Next, we investigate the performance of the two Wald statistics ($\mathcal{W}_{NT}^{\text{sum}}$ and $\mathcal{W}_{NT}^{\text{sup}}$) to test the existence of a panel structure threshold effect at three conventional significance levels, namely, 1%, 5%, and 10%. These tests are evaluated given the correct number of groups, say $G^0 = 3$. We examine the performance of the tests under both homogeneous and heterogeneous threshold effects. However, prior to the test, one is typically ignorant whether the threshold is heterogeneous across groups. Hence, we implement our tests assuming that the threshold is group specific. To facilitate computation and avoid ill behavior for the test statistic, we truncate the top and bottom 10% of the threshold values and use the grid $\{11\%, 12\%, \dots, 89\%\}$. The critical values for the two test statistics are simulated based on $B = 600$ replications.

TABLEs 2 and 3 around here.

Table 2 presents the rejection frequency of the two tests when the threshold is group specific. The left panel presents the size of the test, i.e. the rejection frequency under the null hypothesis with $c_1 = 0$ in DGP 1 and 2 and $c_1 = c_2 = 0$ in DGP 3. Since the classification is based on the discrepancy of slope coefficients, heterogeneity in the threshold does not contribute to group separation. Hence, the size of both tests is generally well controlled. We find that both tests tend to be oversized when $N = 50$ and $T = 30$, but the sizes improve when either N or T increases. The middle panel shows the power of the tests in the presence of a weak threshold effect ($c_1 = 1/5$, $c_2 = 1/15$). Both tests demonstrate non-trivial power in detecting the threshold effect, and for the fixed DGP and nominal level, the power function monotonically increases as either dimension of the sample size grows. Finally, the right panel considers a stronger threshold effect with $c_1 = 1/2$ and $c_2 = 1/10$. We find that the rejection frequency of both tests increases as the threshold effect increases, and it reaches 1 with large samples.

Table 3 considers the case in which the threshold is homogeneous across groups. Again, both tests demonstrate reasonably good size and power properties. We find that both tests tend to over reject the null hypothesis when there are indeed no threshold effects, especially when $T = 30$. As T increases, the rejection frequency approaches the nominal level under the null. Under the alternative hypothesis, the rejection frequency in the presence of homogeneous thresholds seems to be higher than that in case of heterogeneous thresholds. This arises potentially because we estimate the threshold for each group, ignoring the feature of homogeneity. The inefficiency of threshold estimates may inflate the rejection frequency.

7.4 Test for homogeneity of threshold parameters across groups

If there exists a threshold effect, the next issue is whether the threshold is common for individuals. We test the homogeneity of the threshold using the LR-based statistic discussed in Section 4.2. As above, we use the grid $\{11\%, 12\%, \dots, 89\%\}$ to facilitate the computation. To estimate η_g^2, we employ the nonparametric method detailed in Section 4.2 and follow Hansen (2000) in using the Epanechnikov kernel and the bandwidth selected according to a minimum mean square error criterion. The rejection frequency is displayed in Table 4.

TABLE 4 around here.

The left panel of Table 4 presents the rejection frequency under the null hypothesis of homogeneous thresholds with $\mathbf{D} = (1, 1, 1)'$. The size of the test statistic is generally close to the nominal levels in all DGPs, except that it is undersized for the 10% level test in DGP 2 and 3. The middle panel reports the rejection frequency under the alternative hypothesis of weakly heterogeneous threshold values, i.e., $\mathbf{D} = (0.85, 1, 1.15)'$; the right panel considers the case in which the threshold is strongly heterogeneous, i.e., $\mathbf{D} = (0.5, 1, 1.5)'$. As the degree of heterogeneity increases, we observe a stable increase in the power function. The power is also increasing as either N or T increases for the fixed degree of heterogeneity and nominal level. This indicates that our test has reasonably good power in detecting the heterogeneity of threshold values.

7.5 Estimation results

Finally, we consider the estimation of the PSTR model in the case of both homogeneous and group-specific thresholds. When the thresholds are expected to be common across groups, we impose an equality restriction for threshold estimation, but we still allow group-specific slope coefficients. We evaluate the performance of the proposed method with respect to three aspects: clustering, slope coefficient estimates, and threshold estimates. The accuracy of classification is measured by the average of the misclassification frequency (MF) across replications, defined as

$$\text{MF} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}(\hat{g}_i \neq g_i^0).$$

For slope coefficient estimates, we focus on the bias, root mean squared error (RMSE), and coverage probability (CP) of the two-sided nominal 95% confidence interval, while the threshold parameter estimates are evaluated based on the bias, coverage probability, and average confidence interval length. In the dynamic panels (DGP 3), the evaluation is based on the bias-corrected slope coefficient estimates.

TABLE 5 around here.

Table 5 presents the average misclassification rate across replications. In general, the method can correctly estimate the group membership, and the misclassification rate decreases quickly as T increases. In the static panel with heteroskedastic error (DGP 1), PSTR can correctly classify at least 96% of individuals when $T = 30$ and roughly 99.7% when $T = 60$. When the errors are serially correlated (DGP 2), PSTR can correctly estimate the group membership for more than 90% of individuals in the worst case. Allowing for dynamics does not deteriorate the good performance of classification, and the misclassification rate remains low in all cases. Interestingly, we find that the misclassification rate is lower in the case of homogeneous threshold parameters than in the case of group-specific thresholds. This is consistent with our theoretical prediction that group identification requires the separation of group-specific slope coefficients instead of heterogeneity among the threshold parameters.

TABLEs 6–8 around here.

Next, we examine the estimates of the slope coefficients and threshold parameters, and the results are presented in Tables 6–8. In each DGP, the slope coefficients can be accurately estimated with a small bias, and the coverage probability is generally close to the 95% nominal level. Again, allowing for group-specific thresholds leads to poorer slope and threshold estimates. We find that when the threshold is group specific in DGP 2.1, the RMSE of the slope estimates sometimes decreases disproportionally faster than the speed of the increase in T. This occurs because the relatively large misclassification rate in DGP 2.1 is remarkably reduced by increasing T, and precise classification contributes to better slope estimates.

The threshold parameter is also estimated accurately in all cases, and the average length of the confidence interval shrinks as both N and T increase. We find that the average length of the confidence interval is generally much smaller in the case of a homogeneous threshold than the group specific threshold. This suggests that pooling does improve the efficiency of the threshold estimation for common threshold groups.

8 Empirical Applications

We illustrate our procedure through two empirical applications. Our first application examines the investment decision of firms in the presence of financing constraints using the popular data of Hansen (1999). As a second application, we examine the impact of bank deregulation on the distribution of income using the historical data of US states.

8.1 Investment and financing constraints

We first apply the proposed PSTR estimator to revisit the question whether capital market imperfections affect firms' investment behavior. An influential and seminal study by Fazzari et al. (1988) suggests that firms' investment is associated with its cash flow only when the firm is constrained by external financing. To investigate the threshold effect of financial constraints, Hansen (1999) examines three investment determinants, i.e., Tobin's Q, cash flow, and leverage, allowing the impact of cash flow to vary depending on whether a firm is financially constrained. This study assumes that firms are all homogeneous, such that they face the same threshold parameters and share a common effect of determinants. A number of evidence, however, has shown that firms behave heterogeneously in their financial activities, including investment decisions (see, for example, Spearot (2012), Bernard et al. (2007), and Foster et al. (2008)). Heterogeneity may occur not only in the effect of financial variables on investment (even after differentiating constrained and unconstrained firms), but also in threshold parameters. Firms with diversified characteristics may be subjected to distinct threshold levels.

Thus motivated, we revisit the determinants of investment and consider the following model

$$Inv_{it} = \alpha_i + \beta_{1,g_i} x_{i,t-1} \mathbf{1}(q_{i,t-1} \le \gamma_{g_i}) + \beta_{2,g_i} x_{i,t-1} \mathbf{1}(q_{i,t-1} > \gamma_{g_i}) + \varepsilon_{it}, \tag{8.1}$$

where Inv_{it} is the ratio of investment to capital and α_i denotes the firm fixed effect. We follow Lang et al. (1996) and Hansen (1999) to consider the potential determinants $x_{it} = (Q_{it}, CF_{it}, L_{it})$, where Q_{it} is Tobin's Q, CF_{it} is the ratio of cash flows to capital, and L_{it} denotes leverage. q_{it} is

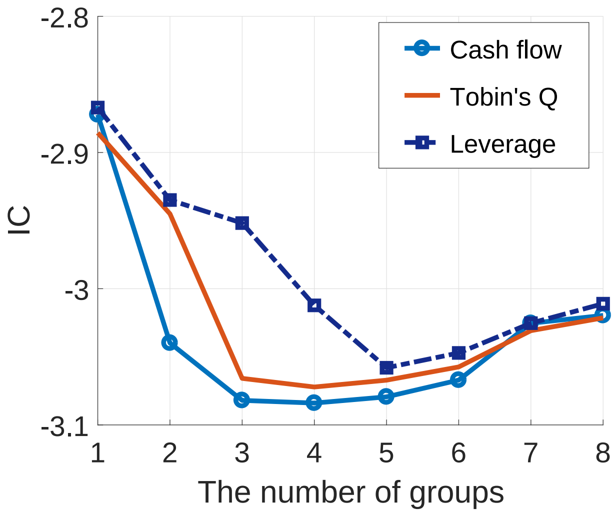

Figure 1: The information criterion for determining the number of groups in the investment and financial constraint application

the threshold variable, which we specify as Tobin's Q, cash flow, or leverage, all of which proxy for a certain degree of financial constraints. The lagged values of Q, CF, and L are used as regressors and threshold variables to avoid possible endogeneity (see also Hansen (1999) and González et al. (2005)). This model allows a *time-invariant* group pattern of heterogeneity in both slope coefficients and the threshold parameter as well as *time-varying* heterogeneity depending on the realization of the threshold variable. We use the same data set as Hansen (1999) that contains 565 firms over 15 years.

To estimate (8.1), we first determine the number of groups chosen based on the IC. Figure 1 displays the value of the IC when we choose the number of groups ranging from 1 to 8 under the three specifications of the threshold variable. For each given number of groups, we estimate the parameters in (8.1) based on 1000 initializations. The IC selects four groups when we use cash flow and Tobin's Q as the threshold variable, while it suggests five groups when leverage is used. We next test the existence of threshold effects using $\mathcal{W}_{NT}^{\text{sum}}$ and $\mathcal{W}_{NT}^{\text{sup}}$ defined in Section 5. Both tests (based on 600 bootstrap replications) suggest the presence of threshold effects for the three specifications of the threshold variable, and the common-threshold test tends to reject the null hypothesis of homogeneity in all cases.

TABLE 9 around here.

Table 9 summarizes the estimation results of (8.1) with three specifications of the threshold variable. When we specify the threshold variable as Tobin's Q, the estimates of the threshold are 10.721, 2.800, 0.854, and 0.282 for the four groups, such that 93%, 87%, 56%, and 15% of the sample fall below the threshold in each group, respectively. In most groups, both Tobin's Q and cash flow are positively associated with investment, as expected. Leverage generally has a negative impact on investment, and this impact is stronger for constrained firms than for unconstrained firms. This result supports the over-investment hypothesis that leverage serves as a disciplining device that prevents firms from over-investing (see, e.g., Jensen (1986) and Seo and Shin (2016)). Group 1 is characterized by relatively low average investment but high average Tobin's Q, while firms in Group 2 are mostly undervalued but still invest aggressively. Group 3 contains very "unsuccessful" firms with highest average leverage as well as lowest average cash flow and Tobin's Q. By contrast, Group 4 is featured by the highest average cash flow and Tobin's Q but lowest average leverage, indicating that firms in this group can be well operated and active in the market. The estimated thresholds for both Groups 1 and 2 occur at the upper quantiles, whereas the effects of cash flow and leverage differ remarkably across the two groups. The effect of cash flow is strongly and positively significant for overvalued firms in Group 2 but less clear for the same type of firms in Group 1. When Tobin's Q is below the threshold, the leverage effect is stronger for firms in Group 2 than for firms in Group 1. For the very "unsuccessful" firms in Group 3, investment is more sensitive to Tobin's Q and cash flow compared with Groups 1 and 2. This is in line with the expectation that the marginal benefit from extra cash and a high asset value is especially high for firms that lack financial resources. Most firms in Group 4 are "successful", with average Tobin's Q greater than 1. For a few firms in this group that are severely undervalued and thus financially constrained, both the positive impact of Tobin's Q and negative impact of leverage are pronounced.

Next, we examine the case in which we use cash flow as the threshold variable. Again, we find a large degree of heterogeneity in the estimates of threshold parameters and slope coefficients. Group 1 contains the burgeoning firms with the largest average cash flow and Tobin's Q. Most firms in this group fall below the lower threshold regime, with significantly positive effects of Tobin's Q and cash flow and a negative effect of leverage. The threshold effect in Group 2 is particularly prominent, since the impact of Tobin's Q and cash flow on investment is much stronger for cash-constrained firms than for unconstrained firms. We find that the effects of Tobin's Q and cash flow are both negative and sizable for extremely cash-constrained firms in Group 3. Further examination reveals that such firms may borrow money to expand, such that they still invest aggressively when they face a shortage of cash flow. This also explains a large positive effect of leverage when they are cash constrained.

Finally, we use leverage as the threshold variable. In this case, the IC suggests five groups. The first three groups share the same threshold at zero, but the slope coefficient estimates differ. Firms in Groups 1 and 2 generally have a low investment level, but firms in Group 1 are mostly overvalued, while those in Group 2 are often undervalued. When these firms have non-zero debt, their investment is positively affected by their cash flow and Tobin's Q. The investment behavior of Group 3 is more sensitive to cash flow than that of Groups 1 and 2. Group 4 contains a number of overvalued firms with large cash flow, and the negative effect of leverage on investment in this group is particularly strong in comparison with that of other groups. Group 5, as an extra group, emerges in this case because of seven firms with especially high investment. Such firms also have an abundance of cash and well-valued assets. These are possibly the aggressive firms, for which we find a strong and positive impacts of cash flow and leverage on investment.

In general, we find a large degree of heterogeneity across firms, which is potentially driven by unobserved firm characteristics, such as their market performance, investment strategy, and managerial risk-taking behavior. Such heterogeneity cannot be captured by conventional threshold regressions. The group pattern varies to some extent for different specifications of the threshold variable. This suggests that the three candidate threshold variables capture distinct aspects of financial constraints.

8.2 Bank regulation and income distribution

Our second application concerns the relationship between bank regulation and the distribution of income. Bank regulation plays a crucial role in governing the financial market. It subjects banks to certain restrictions and guidelines regarding, for example, bank mergers, acquisitions, and branching, in the hope of creating a transparent environment for banking institutions, individuals, and corporations. Bank regulations generally consist of two components: (1) licensing that sets requirements for starting a new bank and (2) governmental supervision of the bank's activities. Hence, with stiffer regulations, there could be fewer banks in operation in the market, and banking activities can be more restricted. In shaping regulation policies, income inequality is always one of the central concerns. There exists a theoretical debate on the impact of bank regulation on the distribution of income. On the one hand, imposing stiffer regulatory restrictions on bank mergers and branching is likely to create and protect local banking monopolies, which further leads to higher fixed fees that hurt the poor. Thus, the main motivation for deregulation is to intensify bank competition and improve bank performance. On the other hand, objection on deregulation is also raised due to the fears that centralized banking power would discriminatively curtail the financial opportunities of the poor (Kroszner and Strahan, 1999) and thus amplify inequality.

We revisit the relationship between bank regulation, particularly branch deregulation, and the distribution of income by applying the PSTR estimator. This analysis was first undertaken by Beck et al. (2010) using US state-level data in a standard (fixed effects) panel framework. We employ the same data set that covers 49 US states for 31 years from 1976 to 2006.[2] The impact of branch deregulation may vary remarkably across states depending on their financial market situations, economic performance, demographic features, and so forth. For example, Beck et al. (2010) suggested that the impact of bank deregulation is more prominent if bank performance prior to deregulations is more severely hurt by intrastate branching restrictions. Moreover, deregulation may disproportionately affect different income groups that are characterized by heterogeneous demographic features, and its impact on the distribution of income could also differ across states depending on their economic and financial market performance.

To model the heterogeneous impact of bank deregulation on the distribution of income, we consider the panel structure threshold model as follows:

$$Inc_{it} = \alpha_i + (\beta_{1,g_i} d_{it} + \beta_{1,g_i} x_{it})\mathbf{1}(q_{it} \leq \gamma_{g_i}) + (\beta_{2,g_i} d_{it} + \beta_{2,g_i} x_{it})\mathbf{1}(q_{it} > \gamma_{g_i}) + \varepsilon_{it}, \qquad (8.2)$$

[2]The dataset contains 50 US states and the District of Columbia but excludes Delaware and South Dakota.

where Inc_{it} represents the distribution of income, which is measured by the logistic transformation of the Gini coefficient following Beck et al. (2010) and α_i denotes the state fixed effect.[3] d_{it} is a dummy variable that equals one if a state has implemented deregulation and zero otherwise, and the date of deregulation refers to that on which a state permitted branching via mergers and acquisitions. The control variables in x_{it} include two salient and robust demographic determinants of income inequality based on the cornerstone study of Beck et al. (2010), namely, the percentage of high school dropouts (Dropout) and the unemployment rate (Unemp). We consider four specifications of the threshold variable q_{it}: the two demographic variables in the covariates (Dropout and Unemp), the initial share of small banks, and the initial share of small firms. Obviously, these two demographic variables allow us to examine the potentially heterogeneous impact of deregulation, which depends on the demographic features of the state. The initial share of small banks reflects the degree of bank competition at the date of deregulation, which may disproportionately determine the impact of deregulation. The initial share of small firms also plays a role in influencing the impact of deregulation because the barriers to obtaining credit from distant banks is greater for small firms than for larger firms, leading to a heterogeneous impact across states with different initial shares of small firms. To analyze the effect of the two share variables, we have to use a subsample of the data with 37 states if we wish to have a balanced panel. Detailed information on the dataset and its source can be found in Beck et al. (2010).

The moderate effect of the two initial share variables was first proposed and analyzed by Beck et al. (2010) in a difference-in-difference (DiD) framework. The advantages of (8.2) compared to the conventional DiD approach are as follows: (1) DiD can only report a positive or negative (linear) effect of the moderating variables, (e.g., the same value for all levels of the initial share of small firms), while PSTR provides information on how such an effect varies (possibly non-linearly) across different levels of these variables; (2) DiD captures only *observed* heterogeneity that is driven by the moderating variables, while PSTR allows us to model the *unobserved* heterogeneity as the group pattern is fully unrestricted.

We first examine the optimal number of groups chosen by the IC. Figure 2 displays the value of IC when we choose the number of groups ranging from 1 to 8 under four specifications of threshold variables. The IC robustly chooses two groups as the optimal specification in all cases. The p-values of $\mathcal{W}_{NT}^{\text{sup}}$ and $\mathcal{W}_{NT}^{\text{sum}}$ suggest that the impact of explanatory variables does exhibit threshold effects for all four specifications of the threshold variable, although to different extents.

TABLE 10 around here.

Table 10 presents the estimated threshold and effects of the explanatory variables. In general, we find a large degree of heterogeneity both across groups and across different levels of the threshold variables. We first examine the impact of deregulation if we specify the threshold variable as the rate of high school dropouts. In this case, the test for the common threshold rejects the null of

[3] We also consider alternative measures of the distribution of income, such as the logarithm of the Gini coefficients and Theil index, and the results are qualitatively unchanged.

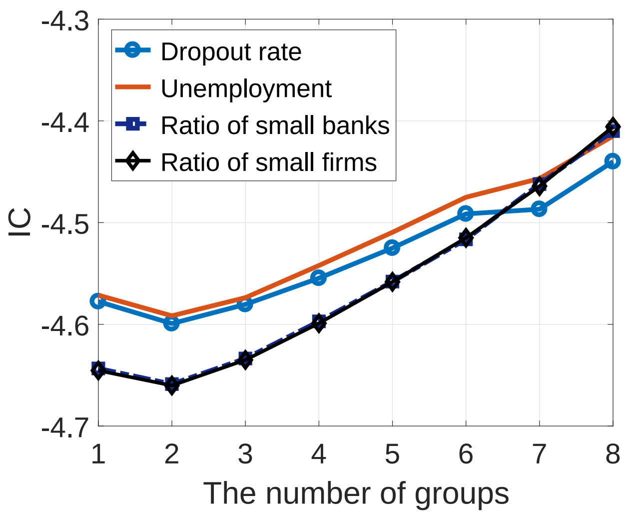

Figure 2: The information criterion for determining the number of groups in the bank deregulation application

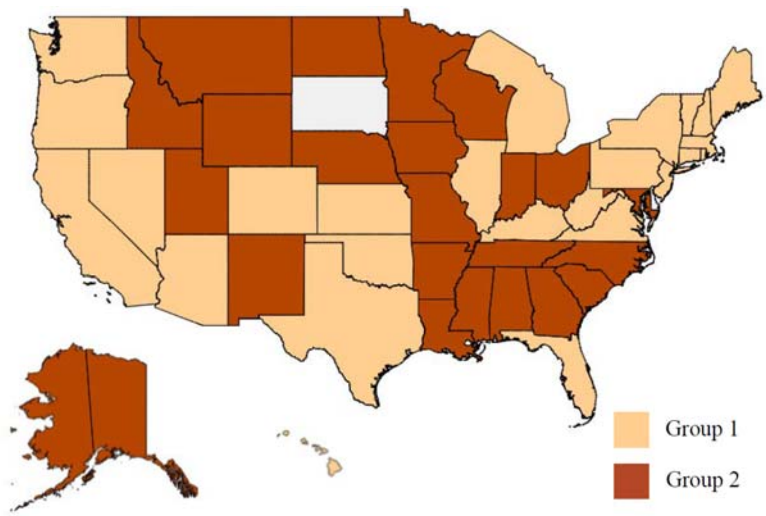

Figure 3: Estimates of the group memebership of US states ($G = 2$)

homogeneity with p-value 0.03; thus, we allow the threshold coefficient to vary across groups in our estimation. The estimation is based on 10000 initial values, and the same number of initializations is used for the estimation with other threshold variables below. Our method assigns 26 states into Group 1 and 23 states into Group 2. Interestingly, the classification coincides with the geographic location to some extent (see Figure 3). Group 1 contains mainly coastal states, such as Washington, Oregon, California, New York, New Jersey, Massachusetts, Vermont, Virginia, and Florida. These states are generally characterized by good economic performance and active financial markets. Group 2 contains states with less active financial markets, including mostly inland and Southeastern states, such as Montana, North Dakota, Minnesota, Nebraska, Iowa, North and South Carolina, and Georgia. The two groups are distinguished by the effects of covariates and the threshold. The estimated threshold of Group 1 is 0.295, such that 73% of observations fall below the threshold. The effect of deregulation on income inequality is significantly negative (-0.0291) when the dropout rate is below the threshold, and it is of a similar size as reported by Beck et al. (2010) (see column (1) of Table II of Beck et al. (2010)). Nevertheless, this effect becomes insignificant when the dropout rate is particularly high. For Group 2, the estimated threshold is much smaller with 1.5% of the sample in the lower threshold regime, and a majority of the sample in this group report a significantly negative impact of deregulation on inequality. Compared with Group 1, the inequality reduction induced by deregulation is much less sizeable in Group 2. This is possibly because bank competition is disproportionately intensified by deregulation in coastal states than in inland/south-eastern states, leading to better bank performance and further to a larger reduction in income inequality.

Next, we examine the deregulation effect when we specify the threshold variable as the unemployment rate. The p-value of the common-threshold test is 0.01, strongly favoring the hypothesis of the heterogeneous threshold coefficients. The group pattern estimated in this case is closely in line with the specification above, with only two states (Ohio and Wyoming) switching their group memberships. We again find a large degree of heterogeneity across the two groups. The estimated thresholds are 9.8 for Group 1 and 2.6 for Group 2, which leads to about 95% and 10% of the sample below the threshold, respectively. The impact of deregulation on inequality is significantly negative for the majority of the sample in both groups but insignificant for the minority. These results suggest that branching deregulation can reduce income inequality in most states, but the magnitude of reduction is bigger in Group 1. However, for the states with an extreme unemployment and dropout rate, deregulation does not significantly help reduce inequality and even enlarges inequality.

To explicitly examine how the degree of bank competition influences the impact of deregulation, we consider the threshold variable as the initial share of small bank. Owing to the unavailability of the initial share in some states, we employ a subsample of the data with 37 states. In this case, the test for the common threshold strongly suggests homogeneity; thus, we proceed with the estimation imposing the homogeneity restriction. The states are again classified into coastal and inland/south-eastern groups with only four states (Kentucky, New Hampshire, North Dakota, and West Virginia) switching their group memberships compared with the case of the dropout rate being the threshold variable. This confirms the heterogeneity of geographic locations and demonstrates the robustness of the estimated group pattern. The estimated threshold is 0.1723 for both groups (due to the common-threshold restriction), such that most observations are in the lower threshold regime. The impact of deregulation is negative in all groups and all regimes, but the magnitude of inequality reduction is larger when the share is beyond the threshold in both groups. This result is in line with the expectation that states with a comparatively high ratio of small banks benefit more from eliminating branching restrictions, as such restrictions that protect small banks from competition have been particularly harmful to bank operations. Since most states are in the lower threshold regime in both groups, we see that the magnitude of inequality reduction induced by deregulation is larger for the majority in Group 1 than the majority in Group 2 as in the previous states.

Finally, we consider the potential threshold effect induced by the initial share of small firms. Again, the test for the common threshold fails to reject the null of homogeneity; thus, we estimate the model restricting the two groups to share the same threshold. The estimated group pattern remains highly similar to the above case using the initial share of small banks as the threshold variable, with only one state changing its group membership. The estimated threshold in both groups is in the 0.783 quantile of the initial share of small firms. Interestingly, when we specify the threshold variable as the two initial-share variables, the estimated slope coefficients in Group 1 are close or even identical. This is, of course, due to the robustness of the classification; moreover, it implies that the two share variables result in similar sample thresholding for Group 1. However, sample thresholding by the two share variables differs in Group 2, and the impact of deregulation is not significant in Group 2 when we use the initial shares of small firms as the threshold variable. In both groups, the inequality reduction is more sizable when the initial share of small firms is beyond the threshold. This confirms the theoretical argument that the impact of deregulation is more pronounced in states with a large ratio of small firms before deregulation, since the existence of branching restrictions impedes the growth of small firms that typically face greater barriers to obtaining credit from distant banks and thus enlarges inequality (Beck et al., 2010).

To summarize, the PSTR estimates provide at least two new important insights that are not provided by standard panel data models with interaction terms. First, we find a large degree of heterogeneity between the two groups even after controlling for the threshold effect, and the impact of deregulation is more sizeable in the group containing most coastal states. This result is robust regardless of the way in which we specify the threshold variable. The group structure coincides with the geographic locations to some extent but not precisely, and this latent group pattern is difficult, if not impossible, to recover using standard panel data approaches. Second, we find a clear threshold effect in each of the two groups. The degree of inequality reduction induced by deregulation depends on the demographic features and the composition of financial markets. Such a group pattern heterogeneity and nonlinear feature of threshold effects can be simultaneously captured by our PSTR model but not by the conventional DiD approach.

9 Conclusion

In this paper, we consider the least squares estimation of a panel structure threshold regression (PSTR) model, where both the slope coefficients and threshold parameters may exhibit latent group structures. We summarize the practical procedure of using this model as follows. The procedure starts with selecting the right number of groups using the IC. With the number of groups given, we first test the presence of threshold effects using the two proposed Wald-type statistics. If there are threshold effects, we then need to test whether the threshold coefficients also vary across groups. Next, we can proceed with the estimation with or without the homogeneity of thresholds imposed, depending on the results of the common-threshold test. We show that we can consistently estimate the latent group structure and estimators of the slope and the threshold coefficients are asymptotically equivalent to the infeasible estimators that are obtained as if the true group structures were known. Moreover, the standard inference based on LR test statistic can provide a correct coverage for the group-specific threshold parameters.

There are several interesting topics for further research. First, we only allow individual fixed effects in our PSTR model. It is possible to also allow for fixed time effects in the model, but this will complicate the analysis to a great deal. Second, it is very interesting but challenging to study the PSTR model with interactive fixed effects, which can incorporate strong cross-sectional dependence in many macro or financial data. Third, we do not allow the latent group structures to change over time. It is interesting and extremely challenging to study PSTR models with a time-varying latent group structure. Fourth, as mentioned in the introduction, we can also consider a PSTR model with endogenous regressors and threshold variables and latent group structures, which would require the use of GMM-type estimation. Fifth, one can also consider a PSTR model with multiple thresholds or multiple threshold variables by extending the works of Li and Ling (2012) and Seo and Linton (2007) to the panel setup with or without latent group structures.

APPENDIX

In this appendix we prove the main results in the paper. The proofs rely on some technical lemmas whose proofs can be found in Appendix B of the online supplement. They also call on some other technical lemmas in Appendix C of the online supplement.

A Proofs of the main results

To prove Theorem 3.1, we first need three technical lemmas, viz, Lemmas A.1–A.3 below. To state these lemmas, we define some notation. First, we introduce the following auxiliary objective function:

$$\tilde{\mathcal{Q}}(\Theta, \mathbf{D}, \mathbf{G}) = \sum_{i=1}^{N} \sum_{t=1}^{T} \left[\tilde{x}_{it}'(\beta^0_{g_i^0} - \beta_{g_i}) + \tilde{x}_{it}(\gamma^0_{g_i^0})'\delta^0_{g_i^0} - \tilde{x}_{it}(\gamma_{g_i})\delta_{g_i} \right]^2 + \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{\varepsilon}_{it}^2. \tag{A.1}$$

Lemma A.1 shows that the distance between $\tilde{\mathcal{Q}}(\Theta, \mathbf{D}, \mathbf{G})$ and $\mathcal{Q}(\Theta, \mathbf{D}, \mathbf{G})$ is $o_p(1)$ uniformly in $(\Theta, \mathbf{D}, \mathbf{G})$ so that we can study the asymptotic properties of $\hat{\Theta}$ through $\tilde{\mathcal{Q}}(\Theta, \mathbf{D}, \mathbf{G})$ in Lemma A.2. Now, define the Hausdorff distance $d_H : \mathcal{B}^G \times \mathcal{B}^G \to R$ as follows

$$d_H(a, b) \equiv \max \left\{ \max_{g \in \mathcal{G}} \left(\min_{\tilde{g} \in \mathcal{G}} \|a_g - b_{\tilde{g}}\| \right), \ \max_{\tilde{g} \in \mathcal{G}} \left(\min_{g \in \mathcal{G}} \|a_g - b_{\tilde{g}}\| \right) \right\}.$$

Lemma A.1. Suppose that Assumption A.1 holds. Then $\sup_{(\Theta, \mathbf{D}, \mathbf{G}) \in \mathcal{B}^G \times \Gamma^G \times \mathcal{G}^N} \frac{1}{NT} |\mathcal{Q}(\Theta, \mathbf{D}, \mathbf{G}) - \tilde{\mathcal{Q}}(\Theta, \mathbf{D}, \mathbf{G})| = o_p(1)$.

Lemma A.2. Suppose that Assumptions A.1–A.3 hold. Then $d_H(\hat{\Theta}, \Theta^0) \xrightarrow{p} 0$ as $(N, T) \to \infty$.

Remark. The proof of Lemma A.2 shows that there exists a permutation $\sigma_{\hat{\Theta}}$ such that $\left\| \hat{\theta}_g - \theta^0_{\sigma_{\hat{\Theta}}(g)} \right\| = o_p(1)$. We can take $\sigma_{\hat{\Theta}}(g) = g$ by relabeling. In the following analysis, we shall write $\hat{\theta}_g - \theta^0_g = o_p(1)$ without referring to the relabeling any more.

Lemma A.3. Let $\hat{g}_i(\Theta, \mathbf{D}) = \text{argmin}_{g \in \mathcal{G}} \sum_{t=1}^{T} \left[\tilde{y}_{it} - \tilde{z}_{it}(\gamma_g)'\theta_g \right]^2$. Suppose Assumptions A.1–A.3 hold. For some $\eta > 0$ small enough and (N, T) large enough such that $\max_{g \in \mathcal{G}} \left\| \delta^0_g \right\| \leq \sqrt{\eta}$, we have

$$\Pr \left(\sup_{(\Theta, \mathbf{D}) \in \mathcal{N}_\eta \times \Gamma^G} \left[\frac{1}{N} \sum_{i=1}^{N} \mathbf{1}(\hat{g}_i(\Theta, \mathbf{D}) \neq g_i^0) \right] \right) = o(T^{-4}),$$

where $\mathcal{N}_\eta = \left\{ \Theta \in \mathcal{B}^G : \left\| \theta_g - \theta^0_g \right\|^2 < \eta, \ g \in \mathcal{G} \right\}$.

Proof of Theorem 3.1. By Lemma A.2, we have $(\hat{\Theta}, \hat{\mathbf{D}}) \in \mathcal{N}_\eta \times \Gamma^G$. Therefore, we can conclude that $\frac{1}{N} \sum_{i=1}^{N} \Pr(\hat{g}_i \neq g_i^0) = o(T^{-4})$ by Lemma A.3. Hence, we have

$$\Pr \left(\sup_i \mathbf{1} \left(\hat{g}_i \neq g_i^0 \right) = 1 \right) \leq \sum_{i=1}^{N} \Pr(\hat{g}_i \neq g_i^0) = N \cdot o(T^{-4}) = o\left(NT^{-4} \right). \blacksquare.$$

To prove Theorem 3.2, we need Lemmas A.4–A.7.

Lemma A.4. Suppose w_{it} is any random variable with $\frac{1}{NT} \sum_{i,t} E \|w_{it}\|^{3+\epsilon} \leq C$ for some constant $\epsilon > 0$ and $C > 0$. Suppose Assumptions A.1–A.5 hold. Then $\left\| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{1}(\hat{g}_i \neq g_i^0) w_{it} \right\| = o_p((NT)^{-1})$.

To state the next lemma, we define an auxiliary estimator $\check{\Theta}(\mathbf{D}) \equiv (\check{\theta}_1(\gamma_1)', ..., \check{\theta}_G(\gamma_G)')'$, which is the least squares estimator of Θ with fixed $\mathbf{D}$ and true group specification $\mathbf{G}^0$, that is,

$$\check{\theta}_g(\gamma) = \left(\sum_{i \in \mathbf{G}_g^0} \sum_{t=1}^{T} \tilde{z}_{it}(\gamma) \tilde{z}_{it}(\gamma)' \right)^{-1} \left(\sum_{i \in \mathbf{G}_g^0} \sum_{t=1}^{T} \tilde{z}_{it}(\gamma) \tilde{y}_{it} \right) \text{ for } g \in \mathcal{G}.$$

Then the infeasible estimator is given by $\check{\Theta} = \check{\Theta}(\check{\mathbf{D}})$ with $\check{\mathbf{D}} = \operatorname{argmin}_{\mathbf{D} \in \Gamma^G} \check{\mathcal{Q}}(\check{\Theta}(\mathbf{D}), \mathbf{D})$. See also (3.1) in Section 3.1. In the online Supplementary Material we derive the asymptotic properties of $\check{\Theta}$. The next lemma establishes the asymptotic equivalence by exploiting the properties of infeasible estimators.

Lemma A.5. Suppose that Assumptions A.1–A.5 hold. Then $(N,T) \to \infty$ we have $\hat{\theta}_g = \check{\theta}_g(\hat{\gamma}_g) + o_p((NT)^{-1})$ for all $g \in \mathcal{G}$.

Lemma A.6. Suppose that Assumptions A.1–A.5 hold and $\alpha \in (0, 1/3)$. Then $\alpha_{NT}(\hat{\gamma}_g - \gamma_g^0) = O_p(1)$ for all $g \in \mathcal{G}$.

Lemma A.7. Suppose that Assumptions A.1–A.5 hold. For any $\gamma = \gamma_g^0 + O_p(1/\alpha_{NT})$ and $g \in \mathcal{G}$, the following statement holds:

$$\check{\theta}_g(\gamma) - \check{\theta}_g(\gamma_g^0) = o_p((NT)^{-1/2}) \text{ and } \check{Q}_g(\check{\theta}_g(\gamma), \gamma) - \check{Q}_g(\check{\theta}_g, \gamma) = o_p(1).$$

Proof of Theorem 3.2. For the first result, we can show $\sqrt{NT}[\check{\theta}_g(\hat{\gamma}_g) - \check{\theta}_g] \to 0$ by Lemmas A.5–A.7. It suffices to show the second result. Given Lemma A.6, we can denote $\hat{\gamma}_g \equiv \gamma_g^0 + \hat{v}_g/\alpha_{N_gT}$ and $\check{\gamma}_g \equiv \gamma_g^0 + \check{v}_g/\alpha_{N_gT}$. Let

$$\begin{aligned} Q_{g,NT}^{**}(v_g) &\equiv \check{Q}_g(\check{\theta}_g(\hat{\gamma}_g), \gamma_g^0) - \check{Q}_g(\check{\theta}_g(\hat{\gamma}_g), \gamma_g^0 + v_g/\alpha_{N_gT}) \text{ and} && \text{(A.2)} \\ Q_{g,NT}^{*}(v_g) &\equiv \check{Q}_g(\check{\theta}_g, \gamma_g^0) - \check{Q}_g(\check{\theta}_g, \gamma_g^0 + v_g/\alpha_{N_gT}). && \text{(A.3)} \end{aligned}$$

First we show that $Q_{g,NT}^{**}(v) - Q_{g,NT}^{*}(v) \xrightarrow{p} 0$ uniformly on any compact set Ψ. It is straightforward to calculate that

$$Q_{g,NT}^{**}(v) - Q_{g,NT}^{*}(v) = L_{g,NT}^{*}(v) - L_{g,NT}(v),$$

where $L_{g,NT}(v)$ is a remainder term that is defined in Lemma C.14 in the online Supplementary Material and $L_{g,NT}^{*}(v)$ can be defined analogously. We show in the proof of Lemma C.14 that $L_{g,NT}^{*}(v) \xrightarrow{p} 0$ uniformly on any compact set Ψ. Similar arguments can be used to show that $L_{g,NT}^{*}(v) \xrightarrow{p} 0$ uniformly on any compact set Ψ. Therefore, we have $Q_{g,NT}^{**}(v) - Q_{g,NT}^{*}(v) \xrightarrow{p} 0$ uniformly on any compact set Ψ.

Next, we have

$$\begin{aligned} Q_{g,NT}^{**}(\hat{v}_g) &= \check{Q}_g(\check{\theta}_g(\hat{\gamma}_g), \gamma_g^0) - \check{Q}_g(\check{\theta}_g(\hat{\gamma}_g), \gamma_g^0 + \hat{v}_g/\alpha_{N_gT}) \\ &= \check{Q}_g(\check{\theta}_g, \gamma_g^0) - \left[\check{Q}_g(\check{\theta}_g(\hat{\gamma}_g), \gamma_g^0 + \hat{v}_g/\alpha_{N_gT}) + o_p(1)\right] \\ &= \check{Q}_g(\check{\theta}_g, \gamma_g^0) - \check{Q}_g(\check{\theta}_g, \gamma_g^0 + \check{v}_g/\alpha_{N_gT}) + o_p(1) \\ &= Q_{g,NT}^{*}(\check{v}_g) + o_p(1) \\ &= \max_{v \in \mathbb{R}} Q_{g,NT}^{*}(v) + o_p(1), \end{aligned}$$

where the first and second equalities hold by (A.2) and Lemma A.7, respectively, the fourth equality holds by (A.3) and the fact that $\check{\theta}_g = \check{\theta}_g(\check{\gamma}_g)$, and the last equality follows from the definition of $\check{\gamma}_g$. On the other hand side, $Q^{**}_{g,NT}(\hat{v}_g) = Q^{*}_{g,NT}(\hat{v}_g) + o_p(1)$ by the uniform convergence of $Q^{**}_{g,NT}(v) - Q^{*}_{g,NT}(v)$ in probability to zero. It follows that

$$Q^{*}_{g,NT}(\hat{v}_g) = \max_{v\in\mathbb{R}} Q^{*}_{g,NT}(v) + o_p(1).$$

Noting that $Q^{*}_{g,NT}(\cdot)$ converges weakly to a continuous stochastic process that has a unique maximum and $\check{v}_g = \text{argmax}_{v\in\mathbb{R}}\, Q^{*}_{g,NT}(v)$, we must have

$$\hat{v}_g = \underset{v\in\mathbb{R}}{\text{argmax}}\, Q^{*}_{g,NT}(v) + o_p(1) = \check{v}_g + o_p(1),$$

which implies $\alpha_{N_gT}(\hat{\gamma}_g - \check{\gamma}_g) = o_p(1)$. ■

Lemma A.8. Suppose Assumptions A.1(ii)–(vi) and A.3–A.6 hold. Let $\mathbb{M}_0 = I_T - \frac{1}{T}\iota_T\iota_T'$ with ι_T being a $T\times 1$ vector of ones.

(i) Under Assumption A.1(i.1) we have $\frac{1}{\sqrt{N_gT}}\sum_{i\in\mathbf{G}_g^0} Z_i(\gamma_g^0)'\mathbb{M}_0\varepsilon_i + \sqrt{\frac{N_g}{T}}\mathbb{B}_{g,NT}(\gamma_g^0) \xrightarrow{d} N(0,\Omega^0_{g,1})$, where $\mathbb{B}_{g,NT}(\gamma_g^0) = \frac{1}{N_gT}\sum_{i\in\mathbf{G}_g^0}\sum_{t=1}^{T}\sum_{s<t} E\left[z_{it}(\gamma_g^0)\varepsilon_{is}\right]$ for each $g\in\mathcal{G}$;

(ii) Under Assumption A.1(i.2) we have: $\frac{1}{\sqrt{N_gT}}\sum_{i\in\mathbf{G}_g^0} Z_i(\gamma_g^0)'\mathbb{M}_0\varepsilon_i \xrightarrow{d} N(0,\Omega^0_{g,2})$, where $\Omega_{g,2}(\gamma_g^0,\gamma_g^0)$ is as defined in Assumption A.6.

Proof of Theorem 3.3. (i) By Theorem 3.2, we only need to consider the infeasible estimator $\check{\Theta}$. By Lemma A.7, we have that

$$\begin{aligned}\sqrt{N_gT}(\check{\theta}_g - \theta_g^0) &= \sqrt{N_gT}(\check{\theta}_g(\gamma_g^0) - \theta_g^0) + o_p(1)\\ &= \left(\frac{1}{N_gT}\sum_{i\in\mathbf{G}_g^0} Z_i(\gamma_g^0)'\mathbb{M}_0 Z_i(\gamma_g^0)\right)^{-1}\frac{1}{\sqrt{N_gT}}\sum_{i\in\mathbf{G}_g^0} Z_i(\gamma_g^0)'\mathbb{M}_0\varepsilon_i + o_p(1).\end{aligned}$$

Then the result follows from Lemma A.8 and Assumption A.6.

(ii) The result follows from Theorem 3.2 and Lemma C.14 in the online supplement. ■

Proof of Theorem 4.1. First, using Lemma A.3, we can readily show that $\hat{N}_g/N_g \xrightarrow{p} 1$ and $\check{Q}_g(\check{\theta}_g,\check{\gamma}_g)/(N_gT) \xrightarrow{p} \sigma_g^2$. Let $\bar{\theta}_g(\gamma)$ be the minimizer of $\bar{Q}_g(\theta,\gamma)$ that is defined in Section 4.1. Following the proof of Lemma A.5, we can also show that $\bar{\theta}_g(\gamma_g^0) = \check{\theta}_g\left(\gamma_g^0\right) + o_p((NT)^{-1})$. With this and using Lemma A.4, we can readily show that

$$\bar{Q}_g(\bar{\theta}_g(\gamma_g^0),\gamma_g^0) = \bar{Q}_g(\check{\theta}_g\left(\gamma_g^0\right),\gamma_g^0) + o_p(1) = \check{Q}_g(\check{\theta}_g(\gamma_g^0),\gamma_g^0) + o_p(1). \tag{A.4}$$

On the one hand, by the definitions of $(\check{\theta}_g,\check{\gamma}_g)$ and $\{(\hat{\theta}_g,\hat{\gamma}_g), g=1,...,G\}$, we have

$$\check{Q}_g(\check{\theta}_g,\check{\gamma}_g) \le \check{Q}_g(\hat{\theta}_g,\hat{\gamma}_g) \text{ and } \sum_{g=1}^{G}\bar{Q}_g(\hat{\theta}_g,\hat{\gamma}_g) \le \sum_{g=1}^{G}\check{Q}_g(\check{\theta}_g,\check{\gamma}_g).$$

On the other hand, we can apply Lemma A.4 to show that

$$\bar{Q}_g(\theta,\gamma) = \check{Q}_g(\theta,\gamma) + o_p(1). \tag{A.5}$$

This, in conjunction with the first inequality in the above displayed equation implies that $\check{Q}_g(\check{\theta}_g,\check{\gamma}_g) \le \bar{Q}_g(\hat{\theta}_g,\hat{\gamma}_g) + o_p(1)$ and hence $\sum_{g=1}^{G}\check{Q}_g(\check{\theta}_g,\check{\gamma}_g) \le \sum_{g=1}^{G}\bar{Q}_g(\hat{\theta}_g,\hat{\gamma}_g) + o_p(1)$. Combining this last inequality with the second inequality in the above displayed equation yields

$$\sum_{g=1}^{G}\check{Q}_g(\check{\theta}_g,\check{\gamma}_g) = \sum_{g=1}^{G}\bar{Q}_g(\hat{\theta}_g,\hat{\gamma}_g) + o_p(1),$$

which, in conjunction with $\check{Q}_g(\check{\theta}_g, \check{\gamma}_g) \leq \bar{Q}_g(\hat{\theta}_g, \hat{\gamma}_g) + o_p(1)$ for each $g \in \mathcal{G}$, implies that

$$\bar{Q}_g(\hat{\theta}_g, \hat{\gamma}_g) = \check{Q}_g(\check{\theta}_g, \check{\gamma}_g) + o_p(1). \tag{A.6}$$

Noting that $\check{\theta}_g(\gamma_g^0) - \theta_g^0 = [\sum_{i\in\mathbf{G}_g^0} Z_i(\gamma_g^0)'\mathbb{M}_0 Z_i(\gamma_g^0)]^{-1} \sum_{i\in\mathbf{G}_g^0} Z_i(\gamma_g^0)'\mathbb{M}_0\varepsilon_i$ and using the analysis of $\check{\theta}_g - \theta_g^0$ in the proof of Theorem 3.3, we can readily show that $\check{\theta}_g(\gamma_g^0) - \check{\theta}_g = o_p(1/\sqrt{NT})$. With this, we can also show that

$$\check{Q}_g(\check{\theta}_g(\gamma_g^0), \gamma_g^0) - \check{Q}_g(\check{\theta}_g, \gamma_g^0) = o_p(1). \tag{A.7}$$

Then we have

$$\begin{aligned}
\bar{Q}_g(\bar{\theta}_g(\gamma_g^0), \gamma_g^0) - \bar{Q}_g(\hat{\theta}_g, \hat{\gamma}_g) &= \check{Q}_g(\bar{\theta}_g(\gamma_g^0), \gamma_g^0) - \check{Q}_g(\check{\theta}_g, \check{\gamma}_g) + o_p(1) \\
&= \check{Q}_g(\check{\theta}_g(\gamma_g^0), \gamma_g^0) - \check{Q}_g(\check{\theta}_g, \check{\gamma}_g) + o_p(1) \\
&= [\check{Q}_g(\check{\theta}_g, \gamma_g^0) - \check{Q}_g(\check{\theta}_g, \check{\gamma}_g)] + [\check{Q}_g(\check{\theta}_g(\gamma_g^0), \gamma_g^0) - \check{Q}_g(\check{\theta}_g, \gamma_g^0)] + o_p(1) \\
&= \check{Q}_g(\check{\theta}_g, \gamma_g^0) - \check{Q}_g(\check{\theta}_g, \check{\gamma}_g) + o_p(1), \qquad (A.8)
\end{aligned}$$

where the first equality follows from (A.5) and (A.6), the second and last equalities hold by (A.4) and (A.7), respectively.

By Lemma C.14 in the online Supplementary Material, we have

$$\check{Q}_g(\check{\theta}_g, \gamma_g^0) - \check{Q}_g(\check{\theta}_g, v/\alpha_{N_gT} + \gamma_g^0) \Rightarrow -\pi_g^{2\alpha} w_{g,D} |v| + 2\sqrt{w_{g,V}\pi_g^{2\alpha}} W_g(v),$$

where $w_{g,D} \equiv C_g^{0\prime} D_g^0 C_g^0$ and $w_{g,V} \equiv C_g^{0\prime} V_g^0 C_g^0$. Then by the continuous mapping theorem (CMT),

$$\begin{aligned}
\check{Q}_g(\check{\theta}_g, \gamma_g^0) - \check{Q}_g(\check{\theta}_g, \check{v}_g/\alpha_{N_gT} + \gamma_g^0) &\Rightarrow \max_{v\in\mathbb{R}} \left[-\pi_g^{2\alpha} w_{g,D} |v| + 2\sqrt{w_{g,V}\pi_g^{2\alpha}} W_g(v)\right] \\
&= \frac{w_{g,V}}{w_{g,D}} \max_{v\in\mathbb{R}} \left[-\frac{w_{g,D}^2}{w_{g,V}} \left|\pi_g^{2\alpha} v\right| + 2\sqrt{\frac{w_{g,D}^2 \pi_g^{2\alpha}}{w_{g,V}}} W_g(v)\right] \\
&= \frac{w_{g,V}}{w_{g,D}} \max_{v\in\mathbb{R}} \left[-\left|\frac{w_{g,D}^2}{w_{g,V}} \pi_g^{2\alpha} v\right| + 2W_g(\frac{w_{g,D}^2}{w_{g,V}} \pi_g^{2\alpha} v)\right] \\
&= \frac{w_{g,V}}{w_{g,D}} \max_{r\in\mathbb{R}} \left[-|r| + 2W_g(r)\right], \qquad (A.9)
\end{aligned}$$

where the second equality holds by the distributional equality $aW_g(v) = W_g(a^2v)$ and the last equality follows from the change of variable (by setting $r \equiv \frac{w_{g,D}^2}{w_{g,V}} \pi_g^{2\alpha} v$).

Lastly, we have

$$\begin{aligned}
\mathcal{L}_{g,NT}\left(\gamma_g^0\right) &= \frac{\bar{Q}_g(\bar{\theta}_g(\gamma_g^0), \gamma_g^0) - \bar{Q}_g(\hat{\theta}_g, \hat{\gamma}_g)}{\bar{Q}_g(\hat{\theta}_g, \hat{\gamma}_g)/(N_gT)} = \frac{\check{Q}_g(\check{\theta}_g, \gamma_g^0) - \check{Q}_g(\check{\theta}_g, \check{v}_g/\alpha_{N_gT} + \gamma_g^0)}{\check{Q}_g(\check{\theta}_g, \check{\gamma}_g)/(N_gT)} + o_P(1) \\
&\xrightarrow{d} \frac{w_{g,V}}{\sigma_g^2 w_{g,D}} \max_{r\in\mathbb{R}} \left[-|r| + 2W_g(r)\right],
\end{aligned}$$

where the first equality holds by (A.8) and (A.6), and the convergence follows from (A.9) and the fact that $\check{Q}_g(\check{\theta}_g, \check{\gamma}_g)/(N_gT) = \sigma_g^2 + o_p(1)$. ■

Proof of Theorem 4.2. Under the null hypothesis, one can study the asymptotic property of $(\hat{\Theta}_r, \hat{\mathbf{D}}_r, \hat{\mathbf{G}}_r)$ similar to that of $(\hat{\Theta}, \hat{\mathbf{D}}, \hat{\mathbf{G}})$. Following the arguments as used in the proof of Lemma A.5, we can show that

$$Q(\hat{\Theta}_r, \hat{\mathbf{D}}_r, \hat{\mathbf{G}}_r) = \check{Q}(\check{\Theta}(\check{\mathbf{D}}_r), \check{\mathbf{D}}_r) + o_p(1),$$

where $\check{\mathbf{D}}_r = \text{argmin}_{\mathbf{D}\in\mathcal{D}_r} \check{Q}(\check{\Theta}(\mathbf{D}), \mathbf{D})$. This, in conjunction with the fact that $Q(\hat{\Theta}, \hat{\mathbf{D}}, \hat{\mathbf{G}}) = \check{Q}(\check{\Theta}, \check{\mathbf{D}}) + o_p(1)$, implies that

$$\begin{aligned}
Q(\hat{\Theta}_r, \hat{\mathbf{D}}_r, \hat{\mathbf{G}}_r) - Q(\hat{\Theta}, \hat{\mathbf{D}}, \hat{\mathbf{G}}) &= \check{Q}(\check{\Theta}(\check{\mathbf{D}}_r), \check{\mathbf{D}}_r) - \check{Q}(\check{\Theta}, \check{\mathbf{D}}) + o_p(1) \\
&= \left[\check{Q}(\check{\Theta}(\check{\mathbf{D}}_r), \check{\mathbf{D}}_r) - \check{Q}(\check{\Theta}(\check{\mathbf{D}}_r), \mathbf{D}^0)\right] + \left[\check{Q}(\check{\Theta}, \mathbf{D}^0) - \check{Q}(\check{\Theta}, \check{\mathbf{D}})\right] \\
&\quad + \left[\check{Q}(\check{\Theta}(\check{\mathbf{D}}_r), \mathbf{D}^0) - \check{Q}(\check{\Theta}, \mathbf{D}^0)\right] + o_p(1) \\
&= \left[\check{Q}(\check{\Theta}, \mathbf{D}^0) - \check{Q}(\check{\Theta}, \check{\mathbf{D}})\right] - \left[\check{Q}(\check{\Theta}(\check{\mathbf{D}}_r), \mathbf{D}^0) - \check{Q}(\check{\Theta}(\check{\mathbf{D}}_r), \check{\mathbf{D}}_r)\right] + o_p(1),
\end{aligned}$$

where we use the fact that $\check{Q}(\check{\Theta}(\check{\mathbf{D}}_r), \mathbf{D}^0) - \check{Q}(\check{\Theta}, \mathbf{D}^0) = o_p(1)$ that can be proved by following the same arguments as used to derive (A.7).

For $\check{Q}(\check{\Theta}, \mathbf{D}^0) - \check{Q}(\check{\Theta}, \check{\mathbf{D}})$, we have that under $H_{02} : \mathbf{D}^0 \in \mathcal{D}_r$ (i.e., $\gamma_1^0 = ... = \gamma_G^0 = \gamma^0$),

$$\begin{aligned}
\check{Q}(\check{\Theta}, \mathbf{D}^0) - \check{Q}(\check{\Theta}, \check{\mathbf{D}}) &= \sum_{g=1}^{G} \left[\check{Q}_g(\check{\theta}_g, \gamma^0) - \check{Q}(\check{\theta}_g, \check{v}_g/\alpha_{N_gT} + \gamma^0)\right] = \sum_{g=1}^{G} Q^*_{g,NT}(\check{v}_g) \\
&\Rightarrow \sum_{g=1}^{G} \frac{w_{g,V}}{w_{g,D}} \max_{v_g\in\mathbb{R}} \left[-\left|\frac{w_{g,D}^2}{w_{g,V}}\pi_g^{2\alpha} v_g\right| + 2W_g(\frac{w_{g,D}^2}{w_{g,V}}\pi_g^{2\alpha} v_g)\right] \\
&= \sum_{g=1}^{G} \frac{w_{g,V}}{w_{g,D}} \max_{v_g\in\mathbb{R}} \left[-|v_g| + 2W_g(v_g)\right]
\end{aligned}$$

by Lemma C.13 in the online supplement. Writing $\check{\mathbf{D}}_r = (\gamma^0 + \check{v}_r/\alpha_{NT}, ..., \gamma^0 + \check{v}_r/\alpha_{NT})'$, we have that under $H_{02} : \mathbf{D}^0 \in \mathcal{D}_r$,

$$\begin{aligned}
\check{Q}(\check{\Theta}(\check{\mathbf{D}}_r), \mathbf{D}^0) - \check{Q}(\check{\Theta}(\check{\mathbf{D}}_r), \check{\mathbf{D}}_r) &= \sum_{g=1}^{G} \left[\check{Q}_g(\check{\theta}_g(\check{\mathbf{D}}_r), \gamma^0) - \check{Q}(\check{\theta}_g(\check{\mathbf{D}}_r), \pi_g^{1-2\alpha}\check{v}_r/\alpha_{N_gT} + \gamma^0)\right] \\
&= \sum_{g=1}^{G} Q^*_{g,NT}(\pi_g^{1-2\alpha}\check{v}_r) \\
&\Rightarrow \max_{v\in\mathbb{R}} \left(\sum_{g=1}^{G} \frac{w_{g,V}}{w_{g,D}} \left[-\left|\frac{w_{g,D}^2}{w_{g,V}}\pi_g v\right| + 2W_g(\frac{w_{g,D}^2}{w_{g,V}}\pi_g v)\right]\right) \\
&= \max_{u\in\mathbb{R}} \left(\sum_{g=1}^{G} \frac{w_{g,V}}{w_{g,D}} \left[-\left|\frac{w_{g,D}}{w_{1,D}}\frac{\sigma^2 w_{g,D}}{w_{g,V}}\pi_g u\right| + 2W_g(\frac{w_{g,D}}{w_{1,D}}\frac{\sigma^2 w_{g,D}}{w_{g,V}}\pi_g u)\right]\right),
\end{aligned}$$

where the last equality is obtained by changing variable $u = v \cdot \sigma^2/w_{1,D}$. This completes our proof. ■

Proof of Theorem 4.3. This proof is analogous to the first half of that of Theorem 4.2 and thus omitted. ■

Proof of Theorem 5.1. Following the arguments as used in the proof of Theorem 3.2, the Wald test statistic is asymptotically equivalent to the infeasible Wald test statistic uniformly for $\mathbf{D}$. Therefore, we can focus on the study of the asymptotic property of the infeasible Wald test statistic. To avoid introducing new notations, we just assume $\hat{\mathbf{G}} = \mathbf{G}^0$, which occurs w.p.a.1. Then $\bar{\theta}_g^{\text{bc}}(\gamma_g) = \check{\theta}_g^{\text{bc}}(\gamma_g)$ w.p.a.1., where $\check{\theta}_g^{\text{bc}}(\gamma_g)$ is the bias-corrected version of $\check{\theta}_g(\gamma_g)$ when necessary (e.g., in the dynamic case) and $\check{\theta}_g(\gamma_g)$ is defined before Theorem 3.2. Similarly, let $\check{\Theta}^{\text{bc}}(\mathbf{D})$ be the bias corrected version of $\check{\Theta}(\mathbf{D})$ when necessary

For $g \in \mathcal{G}$, we can readily establish

$$\begin{aligned}\sqrt{N_g T}\left[\check{\theta}_g^{\mathrm{bc}}(\gamma_g) - \theta_g^0\right] &= \omega_{g,NT}(\gamma_g,\gamma_g)^{-1}\frac{1}{\sqrt{N_g T}}\sum_{i\in\mathbf{G}_g^0}\sum_{t=1}^{T}\tilde{z}_{it}(\gamma_g)[\tilde{x}_{it}(\gamma_g^0) - \tilde{x}_{it}(\gamma_g)]'\delta_g^0 \\ &\quad + \omega_{g,NT}(\gamma_g,\gamma_g)^{-1}S_{g,NT}(\gamma) + o_p(1).\end{aligned}$$

Note that $\omega_{g,NT}(\gamma_g,\gamma_g) \xrightarrow{p} \omega_g(\gamma_g,\gamma_g)$ uniformly in γ_g by Assumption A.6 and $S_{g,NT}(\gamma) \Rightarrow S_g(\gamma)$ on Γ by Assumption A.7. In addition, by Assumption A.6,

$$\begin{aligned}\frac{1}{\sqrt{N_g T}}\sum_{i\in\mathbf{G}_g^0}\sum_{t=1}^{T}\tilde{z}_{it}(\gamma_g)[\tilde{x}_{it}(\gamma_g^0) - \tilde{x}_{it}(\gamma_g)]'\delta_g^0 &= \sqrt{\pi_g}\frac{1}{N_g T}\sum_{i\in\mathbf{G}_g^0}\sum_{t=1}^{T}\tilde{z}_{it}(\gamma_g)[\tilde{z}_{it}(\gamma_g^0) - \tilde{z}_{it}(\gamma_g)]'L'c_g \\ &= \sqrt{\pi_g}\frac{1}{N_g T}\sum_{i\in\mathbf{G}_g^0}\sum_{t=1}^{T}\left[\tilde{z}_{it}(\gamma_g)\tilde{z}_{it}(\gamma_g^0)' - \tilde{z}_{it}(\gamma_g)\tilde{z}_{it}(\gamma_g)'\right]L'c_g \\ &\xrightarrow{p} \sqrt{\pi_g}\left[\omega_g(\gamma_g,\gamma_g^0) - \omega_g(\gamma_g,\gamma_g)\right]L'c_g,\end{aligned}$$

where the convergence follows by Assumption A.6. Then under $\mathbb{H}_{1NT} : \mathbb{L}\Theta^0 = \mathbf{c}/\sqrt{NT}$,

$$\begin{aligned}\sqrt{N_g T}L\check{\theta}_g^{\mathrm{bc}}(\gamma_g) &\Rightarrow \sqrt{N_g T}L\theta_g^0 + L\omega_g(\gamma_g,\gamma_g)^{-1}\left\{S_g(\gamma_g) + \sqrt{\pi_g}\left[\omega_g(\gamma_g,\gamma_g^0) - \omega_g(\gamma_g,\gamma_g)\right]L'c_g\right\} \\ &= \sqrt{\pi_g}c_g + L\omega_g(\gamma_g,\gamma_g)^{-1}[S_g(\gamma_g) + \sqrt{\pi_g}\omega_g(\gamma_g,\gamma_g^0)L'c_g] - \sqrt{\pi_g}LL'c_g \\ &= L\omega_g(\gamma_g,\gamma_g)^{-1}[S_g(\gamma_g) + \sqrt{\pi_g}\omega_g(\gamma_g,\gamma_g^0)L'c_g].\end{aligned}$$

Then by the CMT, we can conclude that

$$\sqrt{NT}\mathbb{L}\hat{\Pi}^{1/2}\check{\Theta}^{\mathrm{bc}}(\mathbf{D}) = \begin{pmatrix}\sqrt{N_{g_1}T}L\check{\theta}_{g_1}^{\mathrm{bc}}(\gamma_{g_1}) \\ \vdots \\ \sqrt{N_{g_P}T}L\check{\theta}_{g_P}^{\mathrm{bc}}(\gamma_{g_P})\end{pmatrix} \Rightarrow \mathbb{L}\boldsymbol{\omega}(\mathbf{D})^{-1}\left[\mathbf{S}(\mathbf{D}) + \mathbf{Q}(\mathbf{D})\Pi^{1/2}\mathbb{L}'\mathbf{c}\right].$$

It is standard to show that $\hat{\mathbb{K}}_{NT}(\mathbf{D}) \xrightarrow{p} \mathbb{L}\boldsymbol{\omega}(\mathbf{D})^{-1}\boldsymbol{\Omega}(\mathbf{D})\boldsymbol{\omega}(\mathbf{D})^{-1}\mathbb{L}'$ uniformly in $\mathbf{D}$. Then we have $W_{NT}(\gamma) \Rightarrow W^c(\gamma)$ by the CMT. ■

Proof of Theorem 6.1. Using Theorem 3.2 and the analysis of the infeasible estimators in Section C of the online supplement, we can readily show that $\hat{\sigma}^2\left(G^0\right) \xrightarrow{p} \sigma^2$ as $(N,T) \to \infty$. Then $IC(G^0) = \ln\left(\hat{\sigma}^2\left(G^0\right)\right) + \lambda_{NT}G^0K \to \sigma^2$ by Assumption D.2(ii) in the online supplement, where $\sigma^2 = \lim_{(N,T)\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}E(\varepsilon_{it}^2)$. When $1 \leq G < G^0$, by Assumption D.2(ii) we have that w.p.a.1. $IC(G) = \ln\left(\hat{\sigma}^2\left(G\right)\right) + \lambda_{NT}GK \geq \ln(\bar{\sigma}^2) > \ln(\sigma^2)$ as $(N,T) \to \infty$. So we have

$$\Pr(\hat{G} < G^0) = \Pr(\exists 1 \leq G < G^0,\ IC(G) < IC(G^0)) \to 0 \text{ as } (N,T) \to \infty. \tag{A.10}$$

Next, we consider the case where $G^0 < G \leq G_{\max}$. When $G > G^0$, we have by Proposition D.1 in the online supplement that $\max_{G^0<G\leq G_{\max}}[\hat{\sigma}^2\left(G\right) - \hat{\sigma}^2(G^0)] = O_p(T^{-1})$. It follows that

$$\begin{aligned}\Pr(\hat{G} > G^0) &= \Pr(\exists G^0 < G \leq G_{\max},\ IC(G) < IC(G^0)) \\ &= \Pr(\exists G^0 < G \leq G_{\max}, T[\ln(\hat{\sigma}^2\left(G\right)) - \ln(\hat{\sigma}^2(G^0))] > (G - G^0)T\lambda_{NT}) \\ &\to 0 \text{ as } (N,T) \to \infty,\end{aligned} \tag{A.11}$$

where the last line follows from the fact that $T[\ln(\hat{\sigma}^2(G)) - \ln(\hat{\sigma}^2(G^0))] = T\ln(1 + \frac{\hat{\sigma}^2(G)-\hat{\sigma}^2(G^0)}{\hat{\sigma}^2(G^0)}) = O(T(\hat{\sigma}^2(G) - \hat{\sigma}^2(G^0)) = O_p(1)$ and $T\lambda_{NT} \to \infty$ as $(N,T) \to \infty$ by Assumption D.2(ii). Combining (A.10) and (A.11), we have $\Pr(\hat{G} = G^0) \to 1$ as $(N,T) \to \infty$. ■

REFERENCES

Ando, T., Bai, J., 2016. Panel data models with grouped factor structure under unknown group membership. Journal of Applied Econometrics 31, 163-191.

Ando, T., Bai, J., 2017. Clustering huge number of financial time series: A panel data approach with high-dimensional predictors and factor structures. Journal of the American Statistical Association 112, 1182-1198.

Arcand, J. L., Berkes, E., Panizza, U., 2015. Too much finance? Journal of Economic Growth 20, 105-148.

Beck, T., Levine, R., and Levkov, A., 2010. Big bad banks? The winners and losers from bank deregulation in the United States. Journal of Finance, 65, 1637-1667.

Bernard, A.B., Jensen, J.B., Redding, S.J., Schott, P.K., 2007. Firms in international trade. Journal of Economic Perspectives, 21, 105-130.

Bonhomme, S., Manresa, E., 2015. Grouped patterns of heterogeneity in panel data. Econometrica 83, 1147–1184.

Browning, M., Carro, J., 2007. Heterogeneity and microeconometrics modeling. In Advances in Economics and Econometrics, Theory and Applications: Ninth World Congress of the Econometric Society, Vol. 3, ed. by R. Blundell, W. K. Newey, and T. Persson. New York: Cambridge University Press, 45-74.

Caner, M., Hansen, B.E., 2004. Instrumental variable estimation of a threshold model. Econometric Theory 20, 813-843.

Chan, K.S., 1993. Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model. Annals of Statistics, 520-533.

Durlauf, S.N., Johnson, P.A., 1995. Multiple regimes and cross-country growth behaviour. Journal of Applied Econometrics 10, 365-384.

Durlauf, S.N., 2001. Manifesto for a growth econometrics. Journal of Econometrics 100, 65-69.

Fazzari, S., Hubbard, R.G., Petersen, B.C., 1987. Financing constraints and corporate investment. NBER Working Paper.

Foster, L., Haltiwanger, J., Syverson, C., 2008. Reallocation, firm turnover, and efficiency: Selection on productivity or profitability? American Economic Review 98, 394-425.

González, A., Terävirta, T., Van Dijk, D., Yang, Y., 2017. Panel smooth transition regression models. Working Paper, Department of Statistics, Uppsala University.

Hahn, J., Kuersteiner, G., 2011. Reduction for dynamic nonlinear panel models with fixed effects. Econometric Theory 27, 1152-1191.

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, 345-368.

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

Hansen, B. E., 2011. Threshold autoregression in economics. Statistics and Its Inference 4, 123-127.

Huang, W., Jin, S., Su, L., 2019. Panel cointegration with latent group structures and an Application to the PPP theory. Econometric Theory, forthcoming.

Jensen, M. C., 1986. Agency costs of free cash flow, corporate finance, and takeovers. American Economic Review 76, 323-329.

Kasahara, H., Shimotsu, K., 2009. Nonparametric identification of finite mixture models of dynamic discrete choices. Econometrica 77, 135-175.

Kourtellos, A., Stengos, T., Tan, C.M., 2016. Structural threshold regression. Econometric Theory 32, 827-860.

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

Kroszner, R. S., Strahan, P. E., 1999. What drives deregulation? Economics and politics of the relaxation of bank branching restrictions. Quarterly Journal of Economics 114, 1437-1467.

Lang, L., Ofek, E., Stulz, R. M., 1996. Leverage, investment, and firm growth. Journal of Financial Economics, 40, 3-29.

Li, D., Ling, S., 2012. On the least squares estimation of multiple-regime threshold autoregressive models. Journal of Econometrics 167, 240-253.

Lin, C-C., Ng, S., 2012. Estimation of panel data models with parameter heterogeneity when group membership is unknown. Journal of Econometric Methods 1, 42-55.

Liu, R., Schick, A., Shang, Z., Zhang, Y., Zhou, Q., 2018. Identification and estimation in panel models with overspecified number of groups, Working Paper, Department of Economics, Louisiana State University.

Lu, X., Su, L., 2017. Determining the number of groups in latent panel structures with an application to income and democracy. Quantitative Economics 8, 729-760.

Miao, K., Li, K., Su, L., 2018. Panel threshold regression with interactive fixed effects. Working Paper, School of Economics, Singapore Management University.

Okui, R., Wang, W., 2019. Heterogeneous structural breaks in panel data models. Journal of Econometrics, conditionally accepted.

Phillips, P. C., Moon, H. R., 1999. Linear regression limit theory for nonstationary panel data. Econometrica 67, 1057-1111.

Potter, S.M., 1995. A nonlinear approach to US GNP. Journal of Applied Econometrics 10, 109-125.

Sarafidis, V., Weber, N., 2015. A partially heterogeneous framework for analyzing panel data. Oxford Bulletin of Economics and Statistics 77, 274-296.

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

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

Spearot, A. C., 2012. Firm heterogeneity, new investment and acquisitions. Journal of Industrial Economics 60, 1-45.

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

Su, L., Jin, S., 2012. Sieve estimation of panel data models with cross section dependence. Journal of Econometrics 169, 34-47.

Su, L., Ju, G., 2018. Identifying latent grouped patterns in panel data models with interactive fixed effects. Journal of Econometrics 206, 554-573.

Su, L., Shi, Z., Phillips, P.C., 2016. Identifying latent structures in panel data. Econometrica 84, 2215-2264.

Su, L., Wang, X. and Jin, S., 2019. Sieve estimation of time-varying panel data models with latent structures. Journal of Business & Economic Statistics 37, 334-349.

Sun, Y., 2005. Estimation and inference in panel structure models. Working Paper, Dept. of Economics, UCSD.

Vogt, M., Linton, O., 2017. Classification of non-parametric regression functions in longitudinal data models. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 79, 5-27.

Vogt, M., Linton, O., 2019. Multiscale clustering of nonparametric regression curves. arXiv preprint arXiv:1903.01459.

Wang, W., Phillips, P.C. and Su, L., 2018. Homogeneity pursuit in panel data models: Theory and application. Journal of Applied Econometrics, 33, 797-815.

Wang, W., Su, L., 2019. Identifying latent group structures in nonlinear panels. Journal of Econometrics, conditionally accepted.

Yu, P., Phillips, P.C., 2018. Threshold regression with endogeneity. Journal of Econometrics 203, 50-68.

Table 1: Group number selection frequency using IC when $G^0 = 3$

			No threshold effect					With threshold effect				
	N	T	1	2	3	4	5	1	2	3	4	5
DGP 1.1	50	30	0.000	0.000	**0.967**	0.033	0.000	0.000	0.000	**0.976**	0.024	0.000
	50	60	0.000	0.000	**0.972**	0.026	0.002	0.000	0.000	**0.997**	0.003	0.000
	100	30	0.000	0.000	**1.000**	0.000	0.000	0.000	0.000	**0.998**	0.002	0.000
	100	60	0.000	0.000	**1.000**	0.000	0.000	0.000	0.000	**1.000**	0.000	0.000
DGP 1.2	50	30	0.000	0.000	**0.974**	0.026	0.000	0.000	0.000	**0.976**	0.024	0.000
	50	60	0.000	0.000	**0.998**	0.002	0.000	0.000	0.000	**0.998**	0.002	0.000
	100	30	0.000	0.000	**0.996**	0.004	0.000	0.000	0.000	**1.000**	0.000	0.000
	100	60	0.000	0.000	**1.000**	0.000	0.000	0.000	0.000	**1.000**	0.000	0.000
DGP 2.1	50	30	0.000	0.000	**0.982**	0.018	0.000	0.000	0.000	**0.982**	0.016	0.002
	50	60	0.000	0.000	**0.996**	0.004	0.000	0.000	0.000	**0.992**	0.008	0.000
	100	30	0.000	0.000	**0.998**	0.002	0.000	0.000	0.000	**0.998**	0.002	0.000
	100	60	0.000	0.000	**1.000**	0.000	0.000	0.000	0.000	**1.000**	0.000	0.000
DGP 2.2	50	30	0.000	0.000	**0.994**	0.006	0.000	0.000	0.000	**0.946**	0.032	0.022
	50	60	0.000	0.000	**1.000**	0.000	0.000	0.000	0.000	**0.998**	0.002	0.000
	100	30	0.000	0.000	**0.996**	0.004	0.000	0.000	0.000	**0.997**	0.003	0.000
	100	60	0.000	0.000	**1.000**	0.000	0.000	0.000	0.000	**1.000**	0.000	0.000
DGP 3.1	50	30	0.000	0.000	**0.992**	0.008	0.000	0.000	0.000	**0.998**	0.002	0.000
	50	60	0.000	0.000	**1.000**	0.000	0.000	0.000	0.000	**1.000**	0.000	0.000
	100	30	0.000	0.000	**1.000**	0.000	0.000	0.000	0.000	**0.998**	0.000	0.000
	100	60	0.000	0.000	**1.000**	0.000	0.000	0.000	0.000	**1.000**	0.000	0.000
DGP 3.2	50	30	0.000	0.000	**0.998**	0.002	0.000	0.000	0.000	**0.994**	0.006	0.000
	50	60	0.000	0.000	**1.000**	0.000	0.000	0.000	0.000	**1.000**	0.000	0.000
	100	30	0.000	0.000	**1.000**	0.000	0.000	0.000	0.000	**0.998**	0.002	0.000
	100	60	0.000	0.000	**1.000**	0.000	0.000	0.000	0.000	**1.000**	0.000	0.000

Table 2: Rejection frequency of test for existence of threshold effect: Heterogeneous thresholds

			No threshold effect ($c_1 = 0$, $c_2 = 0$)			Weak threshold effect ($c_1 = 1/5$, $c_2 = 1/15$)			Strong threshold effect ($c_1 = 1/2$, $c_2 = 1/10$)		
	N	T	1%	5%	10%	1%	5%	10%	1%	5%	10%
						$\mathcal{W}_{NT}^{\text{sup}}$					
DGP 1.1	50	30	0.026	0.072	0.122	0.096	0.228	0.332	0.728	0.833	0.923
	50	60	0.006	0.044	0.088	0.160	0.304	0.496	0.918	0.985	1.000
	100	30	0.016	0.050	0.084	0.160	0.308	0.436	0.923	0.980	0.993
	100	60	0.010	0.044	0.080	0.276	0.512	0.606	1.000	1.000	1.000
DGP 2.1	50	30	0.036	0.094	0.138	0.108	0.202	0.308	0.533	0.755	0.878
	50	60	0.008	0.058	0.088	0.096	0.240	0.332	0.760	0.923	0.943
	100	30	0.024	0.074	0.120	0.126	0.294	0.332	0.788	0.930	0.968
	100	60	0.010	0.044	0.080	0.140	0.342	0.442	0.968	0.993	0.998
DGP 3.1	50	30	0.024	0.070	0.150	0.160	0.306	0.444	0.826	0.942	0.970
	50	60	0.012	0.050	0.106	0.260	0.526	0.642	0.992	1.000	1.000
	100	30	0.018	0.062	0.118	0.212	0.492	0.610	0.984	0.998	1.000
	100	60	0.006	0.058	0.086	0.520	0.770	0.868	1.000	1.000	1.000
						$\mathcal{W}_{NT}^{\text{sum}}$					
DGP 1.1	50	30	0.030	0.076	0.148	0.152	0.276	0.376	0.853	0.915	0.968
	50	60	0.012	0.042	0.086	0.224	0.358	0.544	0.980	1.000	1.000
	100	30	0.020	0.060	0.102	0.244	0.398	0.554	0.985	0.995	1.000
	100	60	0.016	0.044	0.080	0.378	0.622	0.686	1.000	1.000	1.000
DGP 2.1	50	30	0.042	0.106	0.154	0.148	0.260	0.342	0.673	0.855	0.928
	50	60	0.016	0.056	0.090	0.122	0.274	0.382	0.880	0.963	0.980
	100	30	0.032	0.112	0.186	0.216	0.418	0.450	0.925	0.973	0.980
	100	60	0.012	0.060	0.086	0.244	0.436	0.530	0.995	1.000	1.000
DGP 3.1	50	30	0.012	0.064	0.098	0.178	0.312	0.436	0.888	0.962	0.986
	50	60	0.004	0.030	0.080	0.302	0.574	0.668	0.996	1.000	1.000
	100	30	0.014	0.054	0.094	0.272	0.528	0.654	1.000	0.998	1.000
	100	60	0.004	0.036	0.068	0.596	0.798	0.886	1.000	1.000	1.000

Table 3: Rejection frequency of test for existence of threshold effect: Homogeneous thresholds

			No threshold effect ($c_1 = 0$, $c_2 = 0$)			Weak threshold effect ($c_1 = 1/5$, $c_2 = 1/15$)			Strong threshold effect ($c_1 = 1/2$, $c_2 = 1/10$)		
	N	T	1%	5%	10%	1%	5%	10%	1%	5%	10%
						$\mathcal{W}_{NT}^{\text{sup}}$					
DGP 1.2	50	30	0.024	0.072	0.118	0.126	0.356	0.434	0.818	0.964	0.990
	50	60	0.006	0.044	0.088	0.164	0.408	0.526	0.984	0.996	1.000
	100	30	0.016	0.050	0.095	0.208	0.400	0.512	0.978	0.996	1.000
	100	60	0.010	0.044	0.085	0.412	0.635	0.734	1.000	1.000	1.000
DGP 2.2	50	30	0.032	0.076	0.138	0.090	0.220	0.360	0.692	0.926	0.948
	50	60	0.016	0.066	0.118	0.140	0.282	0.404	0.906	0.986	0.994
	100	30	0.020	0.068	0.116	0.122	0.330	0.440	0.908	0.982	0.996
	100	60	0.012	0.052	0.096	0.264	0.474	0.620	0.998	0.998	0.998
DGP 3.2	50	30	0.024	0.094	0.174	0.256	0.474	0.626	0.940	0.990	1.000
	50	60	0.008	0.066	0.118	0.454	0.700	0.804	1.000	1.000	1.000
	100	30	0.012	0.086	0.134	0.398	0.670	0.730	1.000	1.000	1.000
	100	60	0.007	0.056	0.104	0.740	0.906	0.966	1.000	1.000	1.000
						$\mathcal{W}_{NT}^{\text{sum}}$					
DGP 1.2	50	30	0.029	0.076	0.140	0.198	0.400	0.454	0.962	0.992	0.996
	50	60	0.012	0.042	0.086	0.300	0.508	0.672	0.998	1.000	1.000
	100	30	0.018	0.060	0.114	0.362	0.540	0.652	0.998	1.000	1.000
	100	60	0.015	0.044	0.086	0.620	0.780	0.881	1.000	1.000	1.000
DGP 2.2	50	30	0.034	0.076	0.154	0.146	0.322	0.408	0.912	0.970	0.980
	50	60	0.008	0.070	0.124	0.190	0.400	0.548	0.986	1.000	1.000
	100	30	0.041	0.099	0.156	0.298	0.442	0.566	0.990	0.996	1.000
	100	60	0.014	0.056	0.096	0.394	0.628	0.734	1.000	1.000	1.000
DGP 3.2	50	30	0.012	0.068	0.138	0.324	0.520	0.626	0.990	1.000	1.000
	50	60	0.006	0.036	0.070	0.560	0.760	0.816	1.000	1.000	1.000
	100	30	0.010	0.064	0.090	0.480	0.734	0.816	1.000	1.000	1.000
	100	60	0.008	0.044	0.088	0.860	0.982	0.992	1.000	1.000	1.000

Table 4: Rejection frequency for the test of homogeneous thresholds

Threshold			Homogeneous $\gamma = [1, 1, 1]$			Weakly heterogeneous $\gamma = [0.85, 1, 1.15]$			Strongly heterogeneous $\gamma = [0.5, 1, 1.5]$		
	N	T	1%	5%	10%	1%	5%	10%	1%	5%	10%
DGP 1	50	30	0.013	0.076	0.110	0.810	0.904	0.960	0.968	0.980	0.994
	50	60	0.018	0.061	0.114	0.994	0.986	0.996	1.000	1.000	1.000
	100	30	0.014	0.064	0.096	0.990	0.998	1.000	1.000	1.000	1.000
	100	60	0.012	0.046	0.112	1.000	1.000	1.000	1.000	1.000	1.000
DGP 2	50	30	0.014	0.034	0.052	0.344	0.592	0.690	0.116	0.312	0.408
	50	60	0.010	0.038	0.056	0.862	0.950	0.948	0.498	0.710	0.808
	100	30	0.010	0.052	0.058	0.844	0.932	0.956	0.498	0.714	0.794
	100	60	0.008	0.042	0.050	0.994	0.998	1.000	0.920	0.970	0.994
DGP 3	50	30	0.006	0.040	0.064	0.936	0.972	0.900	0.692	0.856	0.900
	50	60	0.010	0.046	0.048	1.000	1.000	1.000	0.968	0.994	0.998
	100	30	0.006	0.042	0.066	0.998	1.000	1.000	0.972	0.992	0.998
	100	60	0.010	0.036	0.040	1.000	1.000	1.000	1.000	1.000	1.000

Table 5: Average misclassification rate

	$N = 50$		$N = 100$	
	$T = 30$	$T = 60$	$T = 30$	$T = 60$
DGP 1.1	0.0365	0.0032	0.0316	0.0026
DGP 1.2	0.0203	0.0011	0.0179	0.0013
DGP 2.1	0.0963	0.0141	0.0697	0.0124
DGP 2.2	0.0509	0.0076	0.0470	0.0075
DGP 3.1	0.0041	0.0001	0.0028	0.0000
DGP 3.2	0.0011	0.0000	0.0015	0.0002

Table 6: Estimates of coefficients and threshold values: Heteroskedastic error (DGPs 1.1-1.2)

		β_1			β_2			γ		
		DGP 1.1: D^0= $(0.5, 1, 1.5)'$								
		Bias	RMSE	CP	Bias	RMSE	CP	Bias	CP	Length
$N = 50$	Group 1	−0.001	0.078	0.908	−0.002	0.056	0.915	0.009	0.958	0.549
$T = 30$	Group 2	0.003	0.097	0.895	0.015	0.107	0.893	0.018	0.923	0.373
	Group 3	0.002	0.078	0.920	0.004	0.103	0.890	0.001	0.960	0.545
$N = 50$	Group 1	−0.004	0.052	0.925	0.000	0.035	0.940	0.002	0.963	0.214
$T = 60$	Group 2	−0.001	0.042	0.925	−0.001	0.042	0.928	0.002	0.965	0.202
	Group 3	−0.003	0.037	0.948	−0.001	0.055	0.913	0.000	0.973	0.246
$N = 100$	Group 1	0.001	0.055	0.922	−0.002	0.038	0.898	−0.003	0.966	0.245
$T = 30$	Group 2	0.004	0.045	0.920	0.000	0.048	0.904	−0.003	0.948	0.207
	Group 3	0.007	0.035	0.928	−0.003	0.057	0.922	0.001	0.968	0.240
$N = 100$	Group 1	0.003	0.037	0.944	−0.002	0.024	0.938	0.000	0.972	0.125
$T = 60$	Group 2	0.003	0.030	0.938	−0.001	0.029	0.942	−0.002	0.970	0.108
	Group 3	0.000	0.025	0.920	−0.004	0.036	0.946	−0.004	0.962	0.119
		DGP 1.2: D^0= $(1, 1, 1)'$								
		Bias	RMSE	CP	Bias	RMSE	CP	Bias	CP	Length
$N = 50$	Group 1	0.001	0.057	0.938	−0.010	0.060	0.928	−0.002	0.928	0.073
$T = 30$	Group 2	−0.002	0.064	0.903	−0.010	0.060	0.923			
	Group 3	0.006	0.062	0.923	−0.011	0.061	0.913			
$N = 50$	Group 1	0.004	0.038	0.960	−0.003	0.040	0.943	0.001	0.933	0.049
$T = 60$	Group 2	0.001	0.043	0.927	−0.003	0.043	0.917			
	Group 3	0.004	0.042	0.940	−0.003	0.038	0.957			
$N = 100$	Group 1	0.001	0.042	0.930	−0.009	0.041	0.947	0.001	0.940	0.051
$T = 30$	Group 2	0.006	0.045	0.913	−0.006	0.041	0.933			
	Group 3	0.006	0.040	0.930	−0.011	0.039	0.963			
$N = 100$	Group 1	0.003	0.026	0.963	−0.002	0.028	0.950	−0.001	0.947	0.027
$T = 60$	Group 2	0.005	0.029	0.933	−0.002	0.029	0.940			
	Group 3	0.004	0.027	0.953	−0.004	0.029	0.943			

Table 7: Estimates of coefficients and threshold values: Autoregressive error (DGPs 2.1-2.2)

		β_1			β_2			γ		
		DGP 2.1: D^0= $(0.5, 1, 1.5)'$								
		Bias	RMSE	CP	Bias	RMSE	CP	Bias	CP	Length
$N = 50$	Group 1	−0.014	0.153	0.834	0.015	0.163	0.874	0.048	0.932	0.797
$T = 30$	Group 2	−0.008	0.198	0.812	0.032	0.225	0.802	−0.010	0.848	0.605
	Group 3	−0.024	0.140	0.858	0.001	0.203	0.856	−0.034	0.936	0.924
$N = 50$	Group 1	−0.008	0.092	0.914	−0.001	0.043	0.930	−0.006	0.966	0.374
$T = 60$	Group 2	−0.003	0.051	0.924	0.002	0.050	0.942	0.004	0.964	0.291
	Group 3	−0.005	0.050	0.922	0.005	0.073	0.892	−0.014	0.958	0.433
$N = 100$	Group 1	−0.021	0.080	0.894	−0.009	0.050	0.882	−0.015	0.960	0.380
$T = 30$	Group 2	−0.002	0.076	0.840	0.000	0.073	0.856	0.003	0.918	0.302
	Group 3	0.006	0.057	0.880	0.013	0.075	0.910	−0.003	0.946	0.331
$N = 100$	Group 1	0.002	0.045	0.944	0.002	0.031	0.932	0.001	0.980	0.195
$T = 60$	Group 2	−0.003	0.037	0.930	0.001	0.037	0.934	0.002	0.950	0.158
	Group 3	−0.002	0.031	0.942	0.000	0.046	0.940	0.000	0.972	0.181
		DGP 2.2: D^0= $(1, 1, 1)'$								
		Bias	RMSE	CP	Bias	RMSE	CP	Bias	CP	Length
$N = 50$	Group 1	−0.002	0.067	0.937	−0.008	0.074	0.920	0.001	0.960	0.181
$T = 30$	Group 2	0.012	0.108	0.877	0.000	0.091	0.923			
	Group 3	0.009	0.091	0.917	−0.008	0.097	0.927			
$N = 50$	Group 1	0.005	0.048	0.965	0.000	0.050	0.938	−0.001	0.985	0.079
$T = 60$	Group 2	0.001	0.053	0.918	−0.002	0.048	0.945			
	Group 3	0.004	0.051	0.930	−0.004	0.049	0.955			
$N = 100$	Group 1	−0.004	0.053	0.928	−0.017	0.061	0.851	−0.001	0.950	0.099
$T = 30$	Group 2	0.001	0.056	0.914	−0.002	0.057	0.910			
	Group 3	0.019	0.053	0.914	−0.002	0.051	0.932			
$N = 100$	Group 1	0.001	0.033	0.950	−0.005	0.036	0.930	0.000	0.980	0.051
$T = 60$	Group 2	−0.001	0.031	0.965	−0.001	0.033	0.965			
	Group 3	0.004	0.033	0.965	−0.001	0.036	0.920			

Table 8: Estimates of coefficients and threshold values: Dynamic panel (DGPs 3.1-3.2)

		β_1			β_2			γ		
		DGP 3.1: D^0= $(0.5, 1, 1.5)'$								
		Bias	RMSE	CP	Bias	RMSE	CP	Bias	CP	Length
$N = 50$	Group 1	-0.007	0.035	0.923	-0.010	0.025	0.920	-0.006	0.940	0.184
$T = 30$	Group 2	-0.003	0.017	0.963	-0.007	0.020	0.907	0.003	0.970	0.161
	Group 3	-0.002	0.012	0.923	-0.007	0.019	0.877	-0.008	0.947	0.147
$N = 50$	Group 1	-0.003	0.025	0.930	-0.005	0.017	0.950	0.001	0.973	0.095
$T = 60$	Group 2	-0.002	0.013	0.953	-0.003	0.012	0.943	-0.002	0.940	0.073
	Group 3	-0.001	0.007	0.960	-0.001	0.011	0.950	0.000	0.947	0.073
$N = 100$	Group 1	-0.007	0.027	0.927	-0.009	0.019	0.917	0.000	0.973	0.110
$T = 30$	Group 2	-0.003	0.014	0.937	-0.007	0.015	0.933	0.000	0.943	0.082
	Group 3	-0.002	0.008	0.940	-0.005	0.013	0.907	-0.002	0.947	0.074
$N = 100$	Group 1	-0.005	0.019	0.923	-0.004	0.013	0.927	-0.001	0.947	0.059
$T = 60$	Group 2	-0.002	0.009	0.930	-0.002	0.009	0.953	0.000	0.960	0.044
	Group 3	-0.001	0.005	0.950	-0.003	0.009	0.920	0.000	0.967	0.038
		DGP 3.2: D^0= $(1, 1, 1)'$								
		Bias	RMSE	CP	Bias	RMSE	CP	Bias	CP	Length
$N = 50$	Group 1	-0.008	0.029	0.957	-0.014	0.032	0.910	0.001	0.977	0.050
$T = 30$	Group 2	-0.004	0.019	0.930	-0.005	0.019	0.910			
	Group 3	0.000	0.012	0.937	-0.005	0.013	0.927			
$N = 50$	Group 1	-0.003	0.018	0.957	-0.005	0.021	0.937	0.000	0.950	0.024
$T = 60$	Group 2	-0.002	0.013	0.940	-0.002	0.012	0.940			
	Group 3	-0.001	0.008	0.953	-0.002	0.009	0.923			
$N = 100$	Group 1	-0.008	0.020	0.957	-0.010	0.025	0.897	0.001	0.983	0.029
$T = 30$	Group 2	-0.002	0.013	0.950	-0.006	0.014	0.933			
	Group 3	0.000	0.009	0.953	-0.005	0.010	0.893			
$N = 100$	Group 1	-0.006	0.017	0.948	-0.004	0.018	0.938	0.000	0.964	0.201
$T = 60$	Group 2	-0.002	0.011	0.942	-0.001	0.012	0.944			
	Group 3	-0.002	0.007	0.958	-0.003	0.008	0.922			

Table 9: Investment and financial constraint: Estimated threshold and slope coefficients

Threshold variable		Tobin's Q				
		Group 1	Group 2	Group 3	Group 4	
γ (Lower regime %)		10.721 (93%)	2.800 (87%)	0.854 (56%)	0.282 (15%)	
β_1	Q	0.0081***	0.0716***	0.1537***	1.3450***	
		(0.0008)	(0.0029)	(0.0146)	(0.0631)	
	CF	0.0918***	0.0977***	0.3278***	−4.6433***	
		(0.0051)	(0.0121)	(0.0366)	(0.1563)	
	L	−0.0158***	−0.0671***	0.0206	−0.8063***	
		(0.0039)	(0.0068)	(0.0204)	(0.1025)	
β_2	Q	0.0086***	0.0134***	0.0553***	−0.0004	
		(0.0010)	(0.0052)	(0.0084)	(0.0003)	
	CF	−0.0194*	0.3007***	−0.4886***	−0.0161***	
		(0.0116)	(0.0579)	(0.0617)	(0.0080)	
	L	0.0668	0.0798	0.1251***	−0.0143***	
		(0.0803)	(0.0830)	(0.0270)	(0.0061)	
Threshold variable		Cash flow				
		Group 1	Group 2	Group 3	Group 4	
γ (Lower regime %)		0.853 (98%)	0.279 (66%)	−0.084 (1.6%)	−0.343(0.2%)	
β_1	Q	0.0013***	0.1447***	−0.4135***	−0.0034***	
		(0.0004)	(0.0075)	(0.0295)	(0.0009)	
	CF	0.0684***	0.1545***	−2.0022***	−0.1496***	
		(0.0052)	(0.0216)	(0.0697)	(0.0411)	
	L	−0.0096*	0.0203*	52.6850*	−0.2208***	
		(0.0041)	(0.0106)	(0.1284)	(0.0435)	
β_2	Q	−0.0010***	0.0068***	0.0468***	0.0117***	
		(0.0005)	(0.0013)	(0.0028)	(0.0013)	
	CF	0.0806***	0.0054	−0.0835***	0.2958***	
		(0.0081)	(0.0138)	(0.0128)	(0.0110)	
	L	−0.0996***	0.1644***	−0.0399***	−0.0730***	
		(0.0193)	(0.0256)	(0.0060)	(0.0083)	
Threshold variable		Leverage				
		Group 1	Group 2	Group 3	Group 4	Group 5
γ (Lower regime %)		0 (8.5%)	0 (8.5%)	0 (8.5%)	0.002 (8.9%)	0.806 (98%)
β_1	Q	−0.0003	0.0957***	0.0014	0.0107***	0.0538***
		(0.0003)	(0.0109)	(0.0027)	(0.0012)	(0.0131)
	CF	0.0584***	−0.0047	0.2276***	−0.0519***	−0.8202***
		(0.0097)	(0.0509)	(0.0247)	(0.0132)	(0.1507)
	L	−0.0083	−0.0297	0.0816	−0.8464***	0.1648***
		(0.0165)	(0.0566)	(0.0549)	(0.0637)	(0.0337)
β_2	Q	0.0039***	0.0804***	0.0117***	0.0003	1.2055***
		(0.0008)	(0.0033)	(0.0021)	(0.0005)	(0.1284)
	CF	0.0304***	0.0423***	0.3854***	0.1164***	−4.3237***
		(0.0054)	(0.0133)	(0.0163)	(0.0086)	(0.2463)
	L	0.0024	−0.0535***	0.0258***	−0.1168***	0.2734***
		(0.0042)	(0.0072)	(0.0078)	(0.0099)	(0.0383)

Table 10: Impact of bank deregulation: Estimated threshold and slope coefficients

Threshold variable		Dropout rate		Unemployment rate	
		Group 1	Group 2	Group 1	Group 2
γ (Lower regime %)		0.295 (73%)	0.041 (1.5%)	9.80 (95%)	2.60 (10%)
β_1	Dereg	-0.0291^{***}	0.2444^{***}	-0.0316^{***}	-0.0228
		(0.0082)	(0.0576)	(0.0080)	(0.0427)
	Dropout	-0.6749^{***}	3.3793^{***}	-0.6959^{***}	-5.2629
		(0.0778)	(0.7635)	(0.0805)	(3.0658)
	Unemp	0.0032^{*}	0.0390^{*}	0.0007	0.1566^{***}
		(0.0020)	(0.0198)	(0.0022)	(0.0489)
β_2	Dereg	-0.1672^{*}	-0.0199^{***}	0.0339	-0.0197^{***}
		(0.0779)	(0.0086)	(0.0415)	(0.0088)
	Dropout	-1.1961^{***}	-0.2286^{***}	-0.4149	-0.2125^{***}
		(0.2666)	(0.0614)	(0.6825)	(0.0629)
	Unemp	0.0626^{***}	0.0263^{***}	0.0212^{***}	0.0263^{***}
		(0.0118)	(0.0021)	(0.0051)	(0.0022)
Threshold variable		Ratio of small banks		Ratio of small firms	
		Group 1	Group 2	Group 1	Group 2
γ (Lower regime %)		0.1723 (94.5%)		0.8943 (78.3%)	
β_1	Dereg	-0.0291^{***}	-0.0067	-0.0354^{***}	0.0003
		(0.0092)	(0.0091)	(0.0091)	(0.0117)
	Dropout	-0.7805^{***}	-0.2432^{***}	-0.8015^{***}	-0.3306^{***}
		(0.0933)	(0.0791)	(0.0924)	(0.0968)
	Unemp	0.0038	0.0253^{***}	0.0030	0.0244^{***}
		(0.0026)	(0.0022)	(0.0025)	(0.0026)
β_2	Dereg	-0.0655	-0.1555^{***}	-0.0655	-0.0089
		(0.0455)	(0.0479)	(0.0455)	(0.0141)
	Dropout	0.5417^{***}	-1.7011^{***}	0.5417^{***}	-0.0295
		(0.2723)	(0.4793)	(0.2723)	(0.1294)
	Unemp	0.0573^{***}	-0.0008	0.0573^{***}	0.0303^{***}
		(0.0179)	(0.0092)	(0.0179)	(0.0042)

Online Supplement to
"Panel Threshold Regressions With Latent Group Structures"
(NOT for Publication)

Ke Miao[a], Liangjun Su[a] and Wendun Wang[b]

[a] School of Economics, Singapore Management University, Singapore
[b] Econometric Institute, Erasmus University Rotterdam and Tinbergen Institute, The Netherlands

This supplement is composed of four parts. Section B contains the proofs of Lemmas A.-A.8 in the above paper. Section C contains the full analysis of the infeasible estimators. Section D provides some additional assumptions for the determination of the true number of groups and a new proposition. Section E studies the consistency of the panel threshold estimators in the framework of fixed threshold effects.

B Proof of Lemmas A.1-A.8 in Appendix A

Proof of Lemma A.1. Note that

$$\begin{aligned}
&\frac{1}{NT}\left[\mathcal{Q}(\Theta,\mathbf{D},\mathbf{G})-\tilde{\mathcal{Q}}(\Theta,\mathbf{D},\mathbf{G})\right]\\
=\;&\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[(\beta^0_{g_i^0}-\beta_{g_i})'\tilde{x}_{it}+\delta^{0\prime}_{g_i^0}\tilde{x}_{it}(\gamma^0_{g_i^0})-\delta'_{g_i}\tilde{x}_{it}(\gamma_{g_i})\right]\tilde{\varepsilon}_{it}\\
=\;&\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\theta^{0\prime}_{g_i^0}\tilde{z}_{it}(\gamma^0_{g_i^0})\varepsilon_{it}-\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\theta'_{g_i}\tilde{z}_{it}(\gamma_{g_i})\varepsilon_{it}\\
=\;&\sum_{g=1}^{G}\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{1}(g_i^0=g)\theta^{0\prime}_{g}\tilde{z}_{it}(\gamma^0_{g})\varepsilon_{it}-\sum_{g=1}^{G}\frac{2}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{1}(g_i=g)\theta'_{g}\tilde{z}_{it}(\gamma_{g})\varepsilon_{it}.
\end{aligned}$$

It suffices to show that the second term in the last line is $o_p(1)$ uniformly in $(\Theta,\mathbf{D},\mathbf{G})\in\mathcal{B}^G\times\Gamma^G\times\mathcal{G}^N$. For each $g\in\mathcal{G}$, we have

$$\begin{aligned}
\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{1}(g_i=g)\theta'_g\tilde{z}_{it}(\gamma_g)\varepsilon_{it}&=\frac{1}{NT}\sum_{i=1}^{N}\mathbf{1}(g_i=g)\sum_{t=1}^{T}\theta'_g z_{it}(\gamma_g)\varepsilon_{it}\\
&\quad-\frac{1}{NT^2}\sum_{i=1}^{N}\mathbf{1}(g_i=g)\sum_{t,s=1}^{T}\theta'_g z_{is}(\gamma_g)\varepsilon_{it}\equiv A_1\left(\theta_g,\gamma_g\right)-A_2\left(\theta_g,\gamma_g\right),
\end{aligned}$$

where $\sum_{t,s=1}^{T}=\sum_{t=1}^{T}\sum_{s=1}^{T}$. Then by the compactness of $\mathcal{B}$ in Assumption A.1(iv) and the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
\sup_{(\theta,\gamma)\in\mathcal{B}\times\Gamma}\left|\frac{1}{NT}\sum_{i=1}^{N}\mathbf{1}(g_i=g)\sum_{t=1}^{T}\theta' z_{it}(\gamma)\varepsilon_{it}\right|&\leq C\sup_{\gamma\in\Gamma}\left\|\frac{1}{NT}\sum_{i=1}^{N}\mathbf{1}(g_i=g)\sum_{t=1}^{T}z_{it}(\gamma)\varepsilon_{it}\right\|\\
&\leq C\left\{\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}(g_i=g)\right\}^{1/2}\frac{1}{N}\left\{\sum_{i=1}^{N}\sup_{\gamma\in\Gamma}\left\|\frac{1}{T}\sum_{t=1}^{T}z_{it}(\gamma)\varepsilon_{it}\right\|^2\right\}^{1/2}.
\end{aligned}$$

Following similar arguments used in the proof of Lemma A.3 in Hansen (2000), we can show that $\max_{1\leq i\leq N}\sup_{\gamma\in\Gamma}\left\|\frac{1}{T}\sum_{t=1}^{T}z_{it}(\gamma)\varepsilon_{it}\right\|^2=o_p(1)$. It follows that $\sup_{(\theta,\gamma)\in\mathcal{B}\times\Gamma}|A_1(\theta,\gamma)|=o_p(1)$. Similarly, by the repeated use of Cauchy-Schwarz inequality

$$\begin{aligned}\sup_{(\theta,\gamma)\in\mathcal{B}\times\Gamma}\left|\frac{1}{NT^2}\sum_{i=1}^N\mathbf{1}(g_i=g)\sum_{t,s=1}^T\theta' z_{is}(\gamma)\varepsilon_{it}\right| &\le C\left\{\sup_{\gamma\in\Gamma}\frac{1}{N}\sum_{i=1}^N\left\|\frac{1}{T}\sum_{s=1}^T z_{is}(\gamma)\right\|^2\right\}^{1/2}\left\{\frac{1}{N}\sum_{i=1}^N\left|\frac{1}{T}\sum_{t=1}^T\varepsilon_{it}\right|^2\right\}^{1/2}\\ &\le 2C\left\{\frac{1}{NT}\sum_{i=1}^N\sum_{s=1}^T\|x_{is}\|^2\right\}^{1/2}\left\{\frac{1}{N}\sum_{i=1}^N\left|\frac{1}{T}\sum_{t=1}^T\varepsilon_{it}\right|^2\right\}^{1/2}\\ &= O_P(T^{-1/2}).\end{aligned}$$

where we use the fact that $E|\frac{1}{T}\sum_{t=1}^T\varepsilon_{it}|^2=O\left(T^{-1}\right)$ under Assumption A.1(i.1) or A.1(i.2) and Assumption A.1(iii). Then $\sup_{(\theta,\gamma)\in\mathcal{B}\times\Gamma}|A_2(\theta,\gamma)|=o_p(1)$. Consequently, $\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\mathbf{1}(g_i=g)\theta_g'\tilde{z}_{it}(\gamma_g)\varepsilon_{it}=o_p(1)$ uniformly in $(\Theta,\mathbf{D},\mathbf{G})\in\mathcal{B}^G\times\Gamma^G\times\mathcal{G}^N$. ■

Proof of Lemma A.2. It suffices to show (i) $\max_{g\in\mathcal{G}}\left(\min_{\tilde{g}\in\mathcal{G}}\left\|\theta_g^0-\hat{\theta}_{\tilde{g}}\right\|\right)=o_p(1)$ and (ii) $\max_{\tilde{g}\in\mathcal{G}}\left(\min_{g\in\mathcal{G}}\left\|\theta_g^0-\hat{\theta}_{\tilde{g}}\right\|\right)=o_p(1)$.

We first show (i). By Lemma A.1, we have

$$\begin{aligned}\frac{1}{NT}\tilde{\mathcal{Q}}(\hat{\Theta},\hat{\mathbf{D}},\hat{\mathbf{G}}) &= \frac{1}{NT}\mathcal{Q}(\hat{\Theta},\hat{\mathbf{D}},\hat{\mathbf{G}})+o_p(1)\le\frac{1}{NT}\mathcal{Q}(\Theta^0,\mathbf{D}^0,\mathbf{G}^0)+o_p(1)\\ &= \frac{1}{NT}\tilde{\mathcal{Q}}(\Theta^0,\mathbf{D}^0,\mathbf{G}^0)+o_p(1),\end{aligned}$$

where the inequality holds by the definition of least squares estimator. On the other hand, noting that $\tilde{\mathcal{Q}}(\Theta,\mathbf{D},\mathbf{G})$ is minimized at $(\Theta^0,\mathbf{D}^0,\mathbf{G}^0)$, we have $\frac{1}{NT}[\tilde{\mathcal{Q}}(\hat{\Theta},\hat{\mathbf{D}},\hat{\mathbf{G}})-\tilde{\mathcal{Q}}(\Theta^0,\mathbf{D}^0,\mathbf{G}^0)]\ge 0$. It follows that $\frac{1}{NT}[\tilde{\mathcal{Q}}(\hat{\Theta},\hat{\mathbf{D}},\hat{\mathbf{G}})-\tilde{\mathcal{Q}}(\Theta^0,\mathbf{D}^0,\mathbf{G}^0)]=o_p(1)$. By direct calculation, we have uniformly in $(\Theta,\mathbf{D},\mathbf{G})$,

$$\begin{aligned}&\frac{1}{NT}\left[\tilde{\mathcal{Q}}(\Theta,\mathbf{D},\mathbf{G})-\tilde{\mathcal{Q}}(\Theta^0,\mathbf{D}^0,\mathbf{G}_g^0)\right]\\ =\ &\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\left\{\theta_{g_i^0}^{0\prime}\tilde{z}_{it}(\gamma_{g_i^0}^0)-\theta_{g_i}'\tilde{z}_{it}(\gamma_{g_i})\right\}^2\\ =\ &\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\left\{(\theta_{g_i^0}^0-\theta_{g_i})'\tilde{z}_{it}(\gamma_{g_i})+\delta_{g_i^0}^{0\prime}[\tilde{x}_{it}(\gamma_{g_i^0}^0)-\tilde{x}_{it}(\gamma_{g_i})]\right\}^2\\ =\ &\sum_{g=1}^G\sum_{\tilde{g}=1}^G\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\mathbf{1}(g_i^0=g)\mathbf{1}(g_i=\tilde{g})\left[(\theta_g^0-\theta_{\tilde{g}})'\tilde{z}_{it}(\gamma_{\tilde{g}})\right]^2+o_p(1)\end{aligned}$$

where the last equality follows from the fact that $\sup_i\left\|\delta_{g_i^0}^0\right\|=o(1)$ under Assumption A.1(vi),

$$\begin{aligned}\sup_{\gamma\in\Gamma}\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\|\tilde{x}_{it}(\gamma)\|^2 &\le \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\|\tilde{x}_{it}\|^2=O_p(1),\ \text{and}\\ \sup_{\gamma,\gamma^*\in\Gamma}\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\|\tilde{z}_{it}(\gamma)\|\,\|\tilde{x}_{it}(\gamma^*)\| &\le \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\|\tilde{x}_{it}\|^2=O_p(1).\end{aligned}$$

By the definition of $M_{NT}(g,\tilde{g},\mathbf{D},\mathbf{G})$ in section 3.1, we have

$$\begin{aligned}
o_p(1) &= \frac{1}{NT}\left[\widetilde{\mathcal{Q}}(\hat{\Theta},\hat{\mathbf{D}},\hat{\mathbf{G}})-\widetilde{\mathcal{Q}}(\Theta^0,\mathbf{D}^0,\mathbf{G}_g^0)\right] \\
&= \sum_{g=1}^{G}\sum_{\tilde{g}=1}^{G}(\theta_g^0-\hat{\theta}_{\tilde{g}})'M_{NT}(g,\tilde{g},\hat{\mathbf{D}},\hat{\mathbf{G}})(\theta_g^0-\hat{\theta}_{\tilde{g}})+o_p(1) \\
&\geq \max_{g\in\mathcal{G}}\sum_{\tilde{g}=1}^{G}(\theta_g^0-\hat{\theta}_{\tilde{g}})'M_{NT}(g,\tilde{g},\hat{\mathbf{D}},\hat{\mathbf{G}})(\theta_g^0-\hat{\theta}_{\tilde{g}})+o_p(1) \\
&\geq \max_{g\in\mathcal{G}}\left(\min_{\tilde{g}\in\mathcal{G}}\left\|\theta_g^0-\hat{\theta}_{\tilde{g}}\right\|^2\right)\sum_{\tilde{g}=1}^{G}\lambda_{\min}[M_{NT}(g,\tilde{g},\hat{\mathbf{D}},\hat{\mathbf{G}})]+o_p(1) \\
&\geq \max_{g\in\mathcal{G}}\left(\min_{\tilde{g}\in\mathcal{G}}\left\|\theta_g^0-\hat{\theta}_{\tilde{g}}\right\|^2\right)\underline{c}_\lambda+o_p(1).
\end{aligned}$$

where the last equality follows from Assumption A.2 which says that there exists a group $\tilde{g}^*\in\mathcal{G}$ such that $\lambda_{\min}[M_{NT}(g,\tilde{g}^*,\hat{\mathbf{D}},\hat{\mathbf{G}})]>\underline{c}_\lambda>0$ with probability approaching 1. Consequently, we have $\max_{g\in\mathcal{G}}\left(\min_{\tilde{g}\in\mathcal{G}}\left\|\theta_g^0-\hat{\theta}_{\tilde{g}}\right\|\right)$ $=o_p(1)$.

To show (ii), let $\sigma(g)\equiv\sigma_{\hat{\Theta}}(g)\equiv\text{argmin}_{\tilde{g}\in\mathcal{G}}\left\|\theta_g^0-\hat{\theta}_{\tilde{g}}\right\|$. Then by the triangle inequality, we have for any $\tilde{g}\neq g$

$$\left\|\hat{\theta}_{\sigma(g)}-\hat{\theta}_{\sigma(\tilde{g})}\right\|\geq\|\theta_g^0-\theta_{\tilde{g}}^0\|-\left\|\hat{\theta}_{\sigma(\tilde{g})}-\theta_{\tilde{g}}^0\right\|-\left\|\hat{\theta}_{\sigma(g)}-\theta_g^0\right\|.$$

The first term on the right hand side (RHS) of the last inequality is larger than $c_{g,\tilde{g}}$ by Assumption A.3(a) and the second and third terms are $o_p(1)$ by the above arguments. Then we can conclude that $\sigma_{\hat{\Theta}}(g)\neq\sigma_{\hat{\Theta}}(\tilde{g})$ w.p.a.1, implying that $\sigma_{\hat{\Theta}}(\cdot)$ is bijective and has the inverse $\sigma_{\hat{\Theta}}^{-1}$. Thus, we have for all $\tilde{g}\in\mathcal{G}$,

$$\min_{g\in\mathcal{G}}\left\|\theta_g^0-\hat{\theta}_{\tilde{g}}\right\|\leq\left\|\theta_{\sigma^{-1}(\tilde{g})}^0-\hat{\theta}_{\tilde{g}}\right\|=\min_{h\in\mathcal{G}}\left\|\theta_{\sigma^{-1}(\tilde{g})}^0-\hat{\theta}_h\right\|=o_p(1).$$

Therefore we have $\max_{\tilde{g}\in\mathcal{G}}\left(\min_{g\in\mathcal{G}}\left\|\theta_g^0-\hat{\theta}_{\tilde{g}}\right\|\right)=o_p(1)$. This completes the proof of Lemma A.2. ■.

Proof of Lemma A.3. For all $g\in\mathcal{G}$, we have

$$\mathbf{1}(\hat{g}_i(\Theta,\mathbf{D})=g)\leq\mathbf{1}\left\{\sum_{t=1}^{T}[\tilde{y}_{it}-\tilde{z}_{it}(\gamma_g)'\theta_g]^2\leq\sum_{t=1}^{T}[\tilde{y}_{it}-\tilde{z}_{it}(\gamma_{g_i^0})'\theta_{g_i^0}]^2\right\}.$$

Thus we have

$$\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}(\hat{g}_i(\Theta,\mathbf{D})\neq g_i^0)=\sum_{g=1}^{G}\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}(\hat{g}_i(\Theta,\mathbf{D})=g)\mathbf{1}(g_i^0\neq g)\leq\sum_{g=1}^{G}\frac{1}{N}\sum_{i=1}^{N}\mathcal{Z}_{ig}(\Theta,\mathbf{D}),$$

where

$$\mathcal{Z}_{ig}(\Theta,\mathbf{D})\equiv\mathbf{1}(g_i^0\neq g)\mathbf{1}\left\{\sum_{t=1}^{T}[\tilde{y}_{it}-\tilde{z}_{it}(\gamma_g)'\theta_g)]^2\leq\sum_{t=1}^{T}[\tilde{y}_{it}-\tilde{z}_{it}(\gamma_{g_i^0})'\theta_{g_i^0}]^2\right\}.$$

For $\mathcal{Z}_{ig}(\Theta,\mathbf{D})$, we have

$$\begin{aligned}
\mathcal{Z}_{ig}(\Theta,\mathbf{D}) &= \mathbf{1}\left(g_i^0\neq g\right)\mathbf{1}\left\{\sum_{t=1}^{T}\left[\tilde{z}_{it}(\gamma_{g_i^0})'(\theta_{g_i^0}-\theta_g)+[\tilde{x}_{it}(\gamma_{g_i^0})-\tilde{x}_{it}(\gamma_{g_i})]'\delta_g\right]\times\right. \\
&\qquad \left.\left[\tilde{z}_{it}(\gamma_{g_i^0}^0)'\theta_{g_i^0}^0+\tilde{\varepsilon}_{it}-\frac{1}{2}[\tilde{z}_{it}(\gamma_g)'\theta_g+\tilde{z}_{it}(\gamma_{g_i^0})'\theta_{g_i^0}]\right]\leq 0\right\} \\
&\leq \max_{\tilde{g}\in\mathcal{G}\setminus\{g\}}\mathbf{1}\left(L_i(g,\tilde{g})\leq 0\right),
\end{aligned}$$

where

$$L_i(g,\tilde{g}) = \sum_{t=1}^{T} \left\{ \tilde{z}_{it}(\gamma_{\tilde{g}})'(\theta_{\tilde{g}} - \theta_g) + [\tilde{x}_{it}(\gamma_{\tilde{g}}) - \tilde{x}_{it}(\gamma_g)]'\delta_g \right\} \{ \tilde{z}_{it}(\gamma_{\tilde{g}}^0)'\theta_{\tilde{g}}^0 + \tilde{\varepsilon}_{it} - \frac{1}{2}[\tilde{z}_{it}(\gamma_g)'\theta_g + \tilde{z}_{it}(\gamma_{\tilde{g}})'\theta_{\tilde{g}}] \}.$$

By adding and subtracting some terms, we have

$$L_i(g,\tilde{g}) = (\beta_{\tilde{g}}^0 - \beta_g^0)' \sum_{t=1}^{T} \tilde{x}_{it} \left[\frac{1}{2} \tilde{x}_{it}' (\beta_{\tilde{g}}^0 - \beta_g^0) + \tilde{\varepsilon}_{it} \right] + A_{iT}(g,\tilde{g}) + B_{iT}(g,\tilde{g}) + C_{iT}(g,\tilde{g}),$$

where

$$\begin{aligned}
A_{iT}(g,\tilde{g}) &= [(\beta_{\tilde{g}} - \beta_{\tilde{g}}^0) - (\beta_g - \beta_g^0)]' \sum_{t=1}^{T} \tilde{x}_{it} \left\{ \tilde{z}_{it}(\gamma_{\tilde{g}}^0)'\theta_{\tilde{g}}^0 + \tilde{\varepsilon}_{it} - \frac{1}{2}[\tilde{z}_{it}(\gamma_g)'\theta_g + \tilde{z}_{it}(\gamma_{\tilde{g}})'\theta_{\tilde{g}}] \right\}, \\
B_{iT}(g,\tilde{g}) &= \sum_{t=1}^{T} \left(\delta_{\tilde{g}}' \tilde{x}_{it}(\gamma_{\tilde{g}}) - \delta_g' \tilde{x}_{it}(\gamma_g) \right)' \left\{ \tilde{z}_{it}(\gamma_{\tilde{g}}^0)'\theta_{\tilde{g}}^0 + \tilde{\varepsilon}_{it} - \frac{1}{2}[\tilde{z}_{it}(\gamma_g)'\theta_g + \tilde{z}_{it}(\gamma_{\tilde{g}})'\theta_{\tilde{g}}] \right\}, \text{ and} \\
C_{iT}(g,\tilde{g}) &= (\beta_{\tilde{g}}^0 - \beta_g^0)' \sum_{t=1}^{T} \tilde{x}_{it} \left\{ \tilde{z}_{it}(\gamma_{\tilde{g}}^0)'\theta_{\tilde{g}}^0 + \tilde{\varepsilon}_{it} - \frac{1}{2}[\tilde{z}_{it}(\gamma_g)'\theta_g^0 + \tilde{z}_{it}(\gamma_{\tilde{g}})'\theta_{\tilde{g}}^0] \right\} \\
&\quad - (\beta_{\tilde{g}}^0 - \beta_g^0)' \sum_{t=1}^{T} \tilde{x}_{it} (\beta_{\tilde{g}}^0 - \beta_g^0) \left[\frac{1}{2} \tilde{x}_{it}' (\beta_{\tilde{g}}^0 - \beta_g^0) + \tilde{\varepsilon}_{it} \right].
\end{aligned}$$

For A_{iT}, have

$$\begin{aligned}
|A_{iT}(g,\tilde{g})| &\leq \left| [(\beta_{\tilde{g}} - \beta_{\tilde{g}}^0) - (\beta_g - \beta_g^0)]' \sum_{t=1}^{T} \tilde{x}_{it}\tilde{\varepsilon}_{it} \right| + \left| [(\beta_{\tilde{g}} - \beta_{\tilde{g}}^0) - (\beta_g - \beta_g^0)]' \sum_{t=1}^{T} \tilde{x}_{it}\tilde{z}_{it}(\gamma_{\tilde{g}}^0)'\theta_{\tilde{g}}^0 \right| \\
&\quad + \frac{1}{2} \left| [(\beta_{\tilde{g}} - \beta_{\tilde{g}}^0) - (\beta_g - \beta_g^0)]' \sum_{t=1}^{T} \tilde{x}_{it} (\tilde{z}_{it}(\gamma_g)'\theta_g + \tilde{z}_{it}(\gamma_{\tilde{g}})'\theta_{\tilde{g}}) \right| \\
&\equiv A_{iT,1}(g,\tilde{g}) + A_{T,2}(g,\tilde{g}) + A_{T,3}(g,\tilde{g}).
\end{aligned}$$

For $A_{T,1}$, we have

$$\begin{aligned}
A_{iT,1}(g,\tilde{g}) &\leq (\|\theta_{\tilde{g}} - \theta_{\tilde{g}}^0\| + \|\theta_g - \theta_g^0\|) \left\| \sum_{t=1}^{T} \tilde{x}_{it}\tilde{\varepsilon}_{it} \right\| \\
&\leq 2\sqrt{\eta} \left(\left\| \sum_{t=1}^{T} x_{it}\varepsilon_{it} \right\| + \frac{1}{T} \left\| \sum_{t=1}^{T} \sum_{s=1}^{T} x_{it}\varepsilon_{is} \right\| \right) \\
&\leq 2\sqrt{\eta} T \left\{ \left\| \frac{1}{T} \sum_{t=1}^{T} x_{it}\varepsilon_{it} \right\| + \left\| \frac{1}{T} \sum_{t=1}^{T} x_{it} \right\| \left\| \frac{1}{T} \sum_{s=1}^{T} \varepsilon_{is} \right\| \right\} \\
&\leq 4\sqrt{\eta} T \left(\frac{1}{T} \sum_{t=1}^{T} \|x_{it}\|^2 + \frac{1}{T} \sum_{t=1}^{T} \varepsilon_{it}^2 \right)
\end{aligned}$$

where we used the fact that $\left\|\theta_g - \theta_g^0\right\| \leq \sqrt{\eta}$ for all $g \in \mathcal{G}$. Similarly, we have

$$\begin{aligned} A_{iT,2}(g,\tilde{g}) &\leq 2\sqrt{\eta}T\left\|\theta_{\tilde{g}}^0\right\|\left\|\frac{1}{T}\sum_{t=1}^{T}\tilde{x}_{it}\tilde{z}_{it}(\gamma_{\tilde{g}}^0)'\right\| \leq 4\sqrt{\eta}T\left\|\theta_{\tilde{g}}^0\right\|\left(\frac{1}{T}\sum_{t=1}^{T}\|\tilde{x}_{it}\|^2\right) \\ &\leq 4\sqrt{\eta}T\left\|\theta_{\tilde{g}}^0\right\|\left(\frac{1}{T}\sum_{t=1}^{T}\|x_{it}\|^2\right), \text{ and} \\ A_{iT,3}(g,\tilde{g}) &\leq 4T\sqrt{\eta}(\|\theta_{\tilde{g}}\| + \|\theta_g\|)\frac{1}{T}\sum_{t=1}^{T}\|\tilde{x}_{it}\|^2 \leq 4T\sqrt{\eta}(\|\theta_{\tilde{g}}\| + \|\theta_g\|)\frac{1}{T}\sum_{t=1}^{T}\|x_{it}\|^2. \end{aligned}$$

Thus, for any $\theta \in \mathcal{N}_\eta$,

$$|A_{iT}(g,\tilde{g})| \leq C_1\sqrt{\eta}T\left(\frac{1}{T}\sum_{t=1}^{T}(\|x_{it}\|^2 + \varepsilon_{it}^2)\right) \equiv H_{1,iT},$$

where C_1 is a positive constant independent of η and T.

For $B_{iT}(g,\tilde{g})$, we have

$$|B_{iT}(g,\tilde{g})| \leq (\|\delta_{\tilde{g}}\| + \|\delta_g\|)\sup_{\gamma\in\Gamma}\left\|\sum_{t=1}^{T}\tilde{x}_{it}(\gamma)\tilde{\varepsilon}_{it}\right\| + 2(\|\delta_{\tilde{g}}\| + \|\delta_g\|)\sup_{\gamma\in\Gamma}\left\|\sum_{t=1}^{T}\tilde{x}_{it}(\gamma_{\tilde{g}}^0)\tilde{z}_{it}(\gamma)\right\|.$$

Due to the fact $\theta \in \mathcal{N}_\eta$, we have $\|\delta_g\| \leq \left\|\delta_g - \delta_g^0\right\| + \left\|\delta_g^0\right\| \leq 2\sqrt{\eta}$ for all $g \in \mathcal{G}$. Following the analysis of A_{iT}, we can show that

$$|B_{iT}(g,\tilde{g})| \leq C_2\sqrt{\eta}T\left(\frac{1}{T}\sum_{t=1}^{T}(\|x_{it}\|^2 + \varepsilon_{it}^2)\right) \equiv H_{2,iT},$$

where C_2 is a positive constant independent of η and T. Analogously, we can show that

$$\begin{aligned} |C_{iT}(g,\tilde{g})| &\leq \left|(\beta_{\tilde{g}}^0 - \beta_g^0)'\sum_{t=1}^{T}\tilde{x}_{it}\tilde{x}_{it}(\gamma_{\tilde{g}}^0)'\delta_{\tilde{g}}^0\right| + \frac{1}{2}\left|(\beta_{\tilde{g}}^0 - \beta_g^0)'\sum_{t=1}^{T}\tilde{x}_{it}\tilde{x}_{it}(\gamma_g)'\theta_g^0\right| + \frac{1}{2}\left|(\beta_{\tilde{g}}^0 - \beta_g^0)'\sum_{t=1}^{T}\tilde{x}_{it}\tilde{x}_{it}(\gamma_{\tilde{g}})'\theta_{\tilde{g}}^0\right| \\ &\leq C_3\sqrt{\eta}T\left(\frac{1}{T}\sum_{t=1}^{T}(\|x_{it}\|^2 + \varepsilon_{it}^2)\right) \equiv H_{3,iT}, \end{aligned}$$

where C_3 is a positive constant independent of η and T. It follows that

$$\mathcal{Z}_{ig}(\Theta,\mathbf{D}) \leq \max_{\tilde{g}\in\boldsymbol{G}\setminus\{g\}}\mathbf{1}\left((\beta_{\tilde{g}}^0 - \beta_g^0)'\sum_{t=1}^{T}\tilde{x}_{it}\left[\frac{1}{2}\tilde{x}_{it}'(\beta_{\tilde{g}}^0 - \beta_g^0) + \tilde{\varepsilon}_{it}\right] \leq H_{iT}\right) \equiv \tilde{\mathcal{Z}}_{ig},$$

where $H_{iT} = C\sqrt{\eta}T[\frac{1}{T}\sum_{t=1}^{T}(\|x_{it}\|^2 + \varepsilon_{it}^2)]$ with $C = C_1 + C_2 + C_3$. Hence, we can conclude that

$$\sup_{(\Theta,\mathbf{D})\in\mathcal{N}_\eta\times\Gamma^G}\frac{1}{N}\sum_{i=1}^{N}\mathbf{1}\left(\hat{g}_i(\Theta,\mathbf{D}) \neq g_i^0\right) \leq \frac{1}{N}\sum_{g=1}^{G}\sum_{i=1}^{N}\tilde{\mathcal{Z}}_{ig}.$$

Noting that $\tilde{\mathcal{Z}}_{ig}$ does not depend on $(\Theta,\mathbf{D})$ and G is fixed, we are left to bound $\Pr(\tilde{\mathcal{Z}}_{ig} = 1)$.

Observe that

$$\Pr(\tilde{\mathcal{Z}}_{ig} = 1) \leq \sum_{\tilde{g}\in\boldsymbol{G}\setminus\{g\}}\Pr\left\{\xi_{iT}(g,\tilde{g}) \leq H_{iT}\right\},$$

where $\xi_{iT}(g,\tilde{g})=(\beta^0_{\tilde{g}}-\beta^0_g)'\sum_{t=1}^T\tilde{x}_{it}\left[\frac{1}{2}\tilde{x}_{it}'(\beta^0_{\tilde{g}}-\beta^0_g)+\tilde{\varepsilon}_{it}\right]$. Letting $C_4=2\max_{i,t}E(\|x_{it}\|^2+\varepsilon^2_{it})$, we have

$$\begin{aligned}
\Pr(\tilde{\mathcal{Z}}_{ig} &= 1)\leq\sum_{\tilde{g}\in\mathcal{G}\backslash\{g\}}\Pr\left\{\xi_{iT}(g,\tilde{g})\leq H_{iT}\right\}\\
&\leq \sum_{\tilde{g}\in\mathcal{G}\backslash\{g\}}\Pr\left\{\xi_{iT}(g,\tilde{g})\leq H_{iT},H_{iT}\leq 2E\left(H_{iT}\right)\right\}+\sum_{\tilde{g}\in\mathcal{G}\backslash\{g\}}\Pr\left\{\xi_{iT}(g,\tilde{g})\leq H_{iT},H_{iT}>2E\left(H_{iT}\right)\right\}\\
&\leq \sum_{\tilde{g}\in\mathcal{G}\backslash\{g\}}\Pr\left\{\xi_{iT}(g,\tilde{g})\leq 2E\left(H_{iT}\right)\right\}+\sum_{\tilde{g}\in\mathcal{G}\backslash\{g\}}\Pr\left\{H_{iT}>2E\left(H_{iT}\right)\right\}\\
&\leq \sum_{\tilde{g}\in\mathcal{G}\backslash\{g\}}\Pr\left\{\xi_{iT}(g,\tilde{g})\leq CC_4\sqrt{\eta}T\right\}+\sum_{\tilde{g}\in\mathcal{G}\backslash\{g\}}\Pr\left(\frac{1}{T}\sum_{t=1}^T(\|x_{it}\|^2+\varepsilon^2_{it})\geq C_4\right).
\end{aligned}$$

Using the fact $(\beta^0_{\tilde{g}}-\beta^0_g)'\sum_{t=1}^T\tilde{x}_{it}\tilde{\varepsilon}_{it}=(\beta^0_{\tilde{g}}-\beta^0_g)'[\sum_{t=1}^T x_{it}\varepsilon_{it}-\frac{1}{T}\sum_{t=1}^T\sum_{s=1}^T x_{is}\varepsilon_{it}]$, we have

$$\begin{aligned}
\Pr(\tilde{\mathcal{Z}}_{ig} &= 1)\leq\sum_{\tilde{g}\in\mathcal{G}\backslash\{g\}}\left\{\Pr\left(\frac{1}{T}\sum_{t=1}^T(\|x_{it}\|^2+\varepsilon^2_{it})\geq C_4\right)\right.\\
&\quad+\Pr\left(\frac{1}{T}\sum_{t=1}^T[\tilde{x}_{it}'(\beta^0_{\tilde{g}}-\beta^0_g)]^2\leq\frac{c_{g,\tilde{g}}}{2}\right)\\
&\quad+\Pr\left((\beta^0_{\tilde{g}}-\beta^0_g)'\frac{1}{T}\sum_{t=1}^T x_{it}\varepsilon_{it}\leq-\frac{c_{g,\tilde{g}}}{4}+\frac{CC_4}{2}\sqrt{\eta}\right)\\
&\quad\left.+\Pr\left(-(\beta^0_{\tilde{g}}-\beta^0_g)'\frac{1}{T^2}\sum_{t=1}^T\sum_{s=1}^T x_{is}\varepsilon_{it}\leq-\frac{c_{g,\tilde{g}}}{4}+\frac{CC_4}{2}\sqrt{\eta}\right)\right\}. \qquad \text{(B.1)}
\end{aligned}$$

By Assumptions A.1 and A.3, we can use Lemma B.1 in the next section to show the first two terms to be $o(T^{-4})$. To study the third term on the RHS of the last inequality, we take η such that $\eta\leq\left[\min_{g\in\mathcal{G}}\left(\frac{\min_{\tilde{g}\in\mathcal{G}\backslash\{g\}}c_{g,\tilde{g}}}{8CC_4}\right)\right]^2$. Then we have that for any $g\neq\tilde{g}\in\mathcal{G}$,

$$\begin{aligned}
\Pr\left((\beta^0_{\tilde{g}}-\beta^0_g)'\frac{1}{T}\sum_{t=1}^T x_{it}\varepsilon_{it}\leq-\frac{c_{g,\tilde{g}}}{4}+\frac{CC_4}{2}\sqrt{\eta}\right) &\leq \Pr\left((\beta^0_{\tilde{g}}-\beta^0_g)'\frac{1}{T}\sum_{t=1}^T x_{it}\varepsilon_{it}\leq-\frac{c_{g,\tilde{g}}}{8}\right)\\
&\leq \Pr\left(\left|(\beta^0_{\tilde{g}}-\beta^0_g)'\frac{1}{T}\sum_{t=1}^T x_{it}\varepsilon_{it}\right|\geq\frac{c_{g,\tilde{g}}}{8}\right)\\
&= o(T^{-4}),
\end{aligned}$$

where the last equality follows by another application of Lemma B.1 and the fact that $\left\|\beta^0_{\tilde{g}}-\beta^0_g\right\|\geq\underline{c}_\beta>0$ under Assumption A.3(i). Similarly, we can show that the last term on the RHS of (B.1) is $o(T^{-4})$. Then we have

$$E\left(\sup_{(\Theta,\mathbf{D})\in\mathcal{N}_\eta\times\Gamma^G}\frac{1}{N}\sum_{i=1}^N\mathbf{1}\{\hat{g}_i(\Theta,\mathbf{D})\neq g^0_i\}\right)\leq E\left(\frac{1}{N}\sum_{g=1}^G\sum_{i=1}^N\tilde{\mathcal{Z}}_{ig}\right)=\frac{1}{N}\sum_{g=1}^G\sum_{i=1}^N\Pr(\tilde{\mathcal{Z}}_{ig}=1)=o(T^{-4}).$$

Lastly, by Markov inequality,

$$\begin{aligned}
\Pr\left(\sup_{(\Theta,\mathbf{D})\in\mathcal{N}_\eta\times\Gamma^G}\frac{1}{N}\sum_{i=1}^N\mathbf{1}(\hat{g}_i(\Theta,\mathbf{D})\neq g^0_i)>\epsilon T^{-4}\right) &\leq \Pr\left(\frac{1}{N}\sum_{g=1}^G\sum_{i=1}^N\Pr(\tilde{\mathcal{Z}}_{ig}=1)>\epsilon T^{-4}\right)\\
&\leq \frac{E\left(\frac{1}{N}\sum_{g=1}^G\sum_{i=1}^N\tilde{\mathcal{Z}}_{ig}\right)}{\epsilon T^{-4}}=o(1),
\end{aligned}$$

for any constant $\epsilon > 0$. This completes our proof. ■.

Proof of Lemma A.4. By Markov inequality, we have

$$\Pr(\sup_{i,t} \|w_{it}\| > \eta (NT)^{1/3}) \leq \frac{1}{\eta^{3+\epsilon} (NT)^{(3+\epsilon)/3}} \sum_{i,t} E \|w_{it}\|^{3+\epsilon} = o(1),$$

implying that $\sup_{i,t} \|w_{it}\| = o_p((NT)^{1/3})$. By Lemma A.3 and the order requirement on N and T, we have

$$\begin{aligned} \left\| \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{1}(\hat{g}_i \neq g_i^0) w_{it} \right\| &\leq \left(\sup_{(\Theta,\mathbf{D}) \in \mathcal{N}_\eta \times \Gamma^G} \frac{1}{N} \sum_{i=1}^{N} \mathbf{1}(\hat{g}_i(\Theta, \mathbf{D}) \neq g_i^0) \right) \sup_{i,t} \|w_{it}\| \\ &= o_p(T^{-4} (NT)^{1/3}) = o_p((NT)^{-1}). \blacksquare \end{aligned}$$

Proof of Lemma A.5. By direct calculations, we have

$$\begin{aligned} \sup_{\gamma \in \Gamma} \left(\frac{1}{NT} \sum_{i,t} E \|\tilde{z}_{it}(\gamma) \tilde{z}_{it}(\gamma)'\|^4 \right) &= \sup_{\gamma \in \Gamma} \left(\frac{1}{NT} \sum_{i,t} E \|\tilde{z}_{it}(\gamma)\|^8 \right) \leq \sup_{\gamma \in \Gamma} \left(\frac{1}{NT} \sum_{i,t} E \|\tilde{z}_{it}(\gamma)\|^8 \right) \\ &\leq \frac{256}{NT} \sum_{i,t} E \|x_{it}\|^8 \leq C < \infty \text{ by Assumption A.1(v).} \end{aligned}$$

Similarly, $\sup_{\gamma \in \Gamma} \left(\frac{1}{NT} \sum_{i,t} E \|\tilde{z}_{it}(\gamma) \tilde{y}_{it}\|^4 \right) \leq C < \infty$. Then we apply Lemma A.4 with $\epsilon = 1$ to obtain

$$\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{1}(\hat{g}_i = g) \tilde{z}_{it}(\hat{\gamma}_g) \tilde{y}_{it} = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{1}(g_i^0 = g) \tilde{z}_{it}(\hat{\gamma}_g) \tilde{y}_{it} + o_p((NT)^{-1}).$$

Analogously, we have

$$\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{1}(\hat{g}_i = g) \tilde{z}_{it}(\hat{\gamma}_g) \tilde{z}_{it}(\hat{\gamma}_g)' = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{1}(g_i^0 = g) \tilde{z}_{it}(\hat{\gamma}_g) \tilde{z}_{it}(\hat{\gamma}_g)' + o_p((NT)^{-1}).$$

To sum up, we have $\hat{\theta}_g = \check{\theta}_g(\hat{\gamma}_g) + o_p((NT)^{-1})$. ■.

Proof of Lemma A.6. Note that

$$\begin{aligned} \frac{1}{NT} \check{\mathcal{Q}}(\check{\Theta}(\hat{\mathbf{D}}), \hat{\mathbf{D}}) &= \frac{1}{NT} \check{\mathcal{Q}}(\hat{\Theta}, \hat{\mathbf{D}}) + o_p((NT)^{-1}) = \frac{1}{NT} \mathcal{Q}(\hat{\Theta}, \hat{\mathbf{D}}, \hat{\mathbf{G}}) + o_p((NT)^{-1}) \\ &\leq \frac{1}{NT} \check{\mathcal{Q}}(\check{\Theta}, \check{\mathbf{D}}) + o_p((NT)^{-1}), \end{aligned}$$

where the first and second equalities hold by Lemmas A.5 and A.4, respectively, and the inequality holds by the definition of least squares estimator $(\hat{\Theta}, \hat{\mathbf{D}}, \hat{\mathbf{G}})$. On the other hand,

$$\frac{1}{NT} \check{\mathcal{Q}}_g(\check{\Theta}, \check{\mathbf{D}}) = \frac{1}{NT} \check{\mathcal{Q}}_g(\check{\Theta}(\check{\mathbf{D}}), \check{\mathbf{D}}) \leq \frac{1}{NT} \check{\mathcal{Q}}_g(\check{\Theta}(\hat{\mathbf{D}}), \hat{\mathbf{D}})$$

by the fact that $\check{\mathcal{Q}}_g(\check{\theta}_g, \check{\gamma}_g) = \inf_{(\theta,\gamma)} \check{\mathcal{Q}}_g(\theta, \gamma))$. It follows that

$$\frac{1}{NT} \left[\check{\mathcal{Q}}_g(\check{\theta}_g(\check{\gamma}_g), \check{\gamma}_g) - \check{\mathcal{Q}}_g(\check{\theta}_g(\hat{\gamma}_g), \hat{\gamma}_g) \right] = o_p((NT)^{-1}), \text{ for all } g \in \mathcal{G}. \tag{B.2}$$

Following the analysis of the infeasible estimator $\check{\gamma}_g$ in Lemma C.10 in the online Supplementary Material, we can also show that $\hat{\gamma}_g - \gamma_g^0 = O_p(1/\alpha_{NT})$ based on (B.2). ■

Proof of Lemma A.7. For all $g \in \mathcal{G}$, we have

$$\check{\theta}_g(\gamma) - \theta_g^0 = [\Phi_{1g}(\gamma)]^{-1} \frac{1}{N_g T} \sum_{i \in \mathbf{G}_g^0} Z_i(\gamma)' \mathbb{M}_0 \varepsilon_i - [\Phi_{1g}(\gamma)]^{-1} \Phi_{2g}(\gamma) \delta_g^0,$$

where $\Phi_{1g}(\gamma) \equiv \frac{1}{N_g T} \sum_{i \in \mathbf{G}_g^0} Z_i(\gamma)' \mathbb{M}_0 Z_i(\gamma)$, $\Phi_{2g}(\gamma) \equiv \frac{1}{N_g T} \sum_{i \in \mathbf{G}_g^0} Z_i(\gamma)' \mathbb{M}_0 X_i(\gamma, \gamma_g^0)$ and $X_i(\gamma, \gamma_g^0) = X_i(\gamma) - X_i(\gamma_g^0)$. It is easy to show that

$$\begin{aligned} \Phi_{1g}(\gamma) &= \Phi_{1g}(\gamma_g^0) + O_p(\alpha_{N_g T}^{-1}) = O_p(1), \\ \frac{1}{N_g T} \sum_{i \in \mathbf{G}_g^0} Z_i(\gamma_g^0)' \mathbb{M}_0 \varepsilon_i &= O_p((NT)^{-1/2} + T^{-1}), \\ \frac{1}{N_g T} \sum_{i \in \mathbf{G}_g^0} \left[Z_i(\gamma) - Z_i(\gamma_g^0) \right]' \mathbb{M}_0 \varepsilon_i &= O_p(\alpha_{N_g T}^{-1} [(NT)^{-1/2} + T^{-1}]), \end{aligned}$$

and $\Phi_{2g}(\gamma)\delta_g^0 = (NT)^{-\alpha} \Phi_{2g}(\gamma) C_g^0 = (NT)^{-\alpha} O_p(\alpha_{N_g T}^{-1}) = O_p((NT)^{-1+\alpha})$, where we use the fact that $\alpha_{N_g T} = (N_g T)^{1-2\alpha}$ and $N_g/N \to \pi_g > 0$. With these results, we can readily show that

$$\begin{aligned} \check{\theta}_g(\gamma) - \check{\theta}_g(\gamma_g^0) &= \left\{ [\Phi_{1g}(\gamma)]^{-1} \frac{1}{N_g T} \sum_{i \in \mathbf{G}_g^0} Z_i(\gamma)' \mathbb{M}_0 \varepsilon_i - \left[\Phi_{1g}(\gamma_g^0) \right]^{-1} \frac{1}{N_g T} \sum_{i \in \mathbf{G}_g^0} Z_i(\gamma_g^0)' \mathbb{M}_0 \varepsilon_i \right\} \\ &\quad - [\Phi_{1g}(\gamma)]^{-1} \Phi_{2g}(\gamma) \delta_g^0 \\ &= \left\{ [\Phi_{1g}(\gamma)]^{-1} - \left[\Phi_{1g}(\gamma_g^0) \right]^{-1} \right\} \frac{1}{N_g T} \sum_{i \in \mathbf{G}_g^0} Z_i(\gamma_g^0)' \mathbb{M}_0 \varepsilon_i \\ &\quad + [\Phi_{1g}(\gamma)]^{-1} \frac{1}{N_g T} \sum_{i \in \mathbf{G}_g^0} \left[Z_i(\gamma) - Z_i(\gamma_g^0) \right] \mathbb{M}_0 \varepsilon_i - [\Phi_{1g}(\gamma)]^{-1} \Phi_{2g}(\gamma) \delta_g^0 \\ &= O_p(\alpha_{N_g T}^{-1} [(NT)^{-1/2} + T^{-1}]) + O_p(\alpha_{N_g T}^{-1} [(NT)^{-1/2} + T^{-1}]) + O_p((NT)^{-1+\alpha}) \\ &= O_p(\alpha_{N_g T}^{-1} [(NT)^{-1/2} + T^{-1}]) + O_p((NT)^{-1+\alpha}) = o_p((NT)^{-1/2}) \end{aligned}$$

where the last equality follows from the fact that $\alpha \in (0, 1/3)$ and $N = O(T^2)$. The above analysis also shows that $\check{\theta}_g(\gamma_g^0) - \theta_g^0 = O_p\left((NT)^{-1/2} + T^{-1}\right)$.

Next, noting that

$$\check{Q}_g(\theta, \gamma) = \sum_{i \in \mathbf{G}_g^0} \sum_{t=1}^{T} [\tilde{y}_{it} - \tilde{z}_{it}(\gamma)'\theta]^2 = \sum_{i \in \mathbf{G}_g^0} [Y_i - Z_i(\gamma)\theta]' \mathbb{M}_0 [Y_i - Z_i(\gamma)\theta],$$

$A_i' \mathbb{M}_0 A_i - B_i' \mathbb{M}_0 B_i = (A_i - B_i)' \mathbb{M}_0 (A_i - B_i) + 2(A_i - B_i)' \mathbb{M}_0 B_i$ for any two $T \times 1$ vectors A_i and B_i, and

$Y_i - Z_i(\gamma)\theta = [X_i(\gamma) - X_i(\gamma_g^0)]\delta_g^0 + \mu_i \iota_T + \varepsilon_i$ with ι_T being a $T \times 1$ vector of ones, we have

$$\begin{aligned}
\check{Q}_g(\check{\theta}_g(\gamma), \gamma) - \check{Q}_g(\check{\theta}_g, \gamma) &= \sqrt{NT}[\check{\theta}_g(\gamma) - \check{\theta}_g]' \frac{1}{NT} \sum_{i \in \mathbf{G}_g^0} Z_i(\gamma)' \mathbb{M}_0 Z_i(\gamma) \sqrt{NT}[\check{\theta}_g(\gamma) - \check{\theta}_g] \\
&\quad + 2NT[\check{\theta}_g(\gamma) - \check{\theta}_g]' \frac{1}{NT} \sum_{i \in \mathbf{G}_g^0} Z_i(\gamma)' \mathbb{M}_0 [X_i(\gamma) - X_i(\gamma_g^0)]\delta_g^0 \\
&\quad + 2NT[\check{\theta}_g(\gamma) - \check{\theta}_g]' \frac{1}{NT} \sum_{i \in \mathbf{G}_g^0} Z_i(\gamma)' \mathbb{M}_0 \varepsilon_i \\
&= o_p(1) + NT o_p((NT)^{-1/2}) O_p((NT)^{-1+\alpha}) \\
&\quad + NT[O_p(\alpha_{N_g T}^{-1}[(NT)^{-1/2} + T^{-1}]) + O_p((NT)^{-1+\alpha})] O_p((NT)^{-1/2} + T^{-1}) \\
&= o_p(1),
\end{aligned}$$

where the last equality follows from the fact that $\alpha \in (0, 1/3)$ and $N = O(T^2)$. ■

Proof of Lemma A.8. (i) Let $\mathbb{P}_0 = \frac{1}{T}\iota_T \iota_T'$. Note that

$$\begin{aligned}
\frac{1}{\sqrt{N_g T}} \sum_{i \in \mathbf{G}_g^0} Z_i(\gamma_g^0)' \mathbb{M}_0 \varepsilon_i &= \frac{1}{\sqrt{N_g T}} \sum_{i \in \mathbf{G}_g^0} [Z_i(\gamma_g^0) - \mathbb{P}_0 E(Z_i(\gamma_g^0))]' \varepsilon_i \\
&\quad - \frac{1}{T\sqrt{N_g T}} \sum_{i \in \mathbf{G}_g^0} \sum_{s,t=1}^{T} \{z_{it}(\gamma_g^0) - E\left[z_{it}(\gamma_g^0)\right]\} \varepsilon_{is} \equiv A_1 - A_2.
\end{aligned} \tag{B.3}$$

It suffices to show that (i1) $A_1 \xrightarrow{d} N(0, \Omega_{g,1}^0)$ and (i2) $A_2 = \sqrt{\frac{N_g}{T}} \mathbb{B}_{g,NT} + o_p(1)$. To prove (i1), we relabel the index $\mathbf{G}_g^0 = \{i_1, ..., i_{N_g}\}$ to $\{1, ..., N_g\}$. Let c denote a $2K \times 1$ nonrandom vector with $\|c\| = 1$. For $m = (i-1)T + t$ for $t = 1, ..., T$ and $i = 1, ..., N_g$, let $\zeta_m = \left[z_{it}(\gamma_g^0) - \frac{1}{T}\sum_{t=1}^{T} E(z_{it}(\gamma_g^0))\right] \varepsilon_{it}$. Let $M = N_g T$. Then we have

$$c' A_1 = \frac{1}{\sqrt{M}} \sum_{m=1}^{M} c' \zeta_m.$$

Immediately, $\{\zeta_m\}_{m=1}^M$ is a martingale difference sequence (m.d.s.) under the filtration $\mathcal{F}_m = \sigma(\{\zeta_n : 1 \le n \le m\})$, the minimal sigma-field generated from $\{\zeta_n : 1 \le n \le m\}$. Apparently, $\max_{1 \le m \le M} E\|\zeta_m\|^4 \le C$ for some $C < \infty$ under Assumption A.1. In addition,

$$\begin{aligned}
&\frac{1}{M} \sum_{m=1}^{M} c' \zeta_m \zeta_m' c \\
= \;& c' \frac{1}{N_g T} \sum_{i \in \mathbf{G}_g^0} \sum_{t=1}^{T} \tilde{z}_{it}(\gamma_g^0) \tilde{z}_{it}(\gamma_g^0)' c \varepsilon_{it}^2 \\
& + c' \frac{1}{N_g T} \sum_{i \in \mathbf{G}_g^0} \sum_{t=1}^{T} \varepsilon_{it}^2 \left[2 z_{it}(\gamma_g^0) - \frac{1}{T} \sum_{s=1}^{T} \{z_{is}(\gamma_g^0) - E[z_{is}(\gamma_g^0)]\}\right] \left[\frac{1}{T} \sum_{s=1}^{T} \{z_{is}(\gamma_g^0) - E[z_{is}(\gamma_g^0)\}'\right] c \\
\equiv \;& A_{1,1} + A_{1,2}.
\end{aligned}$$

By Assumption A.6, $A_{1,1} \xrightarrow{p} c'\Omega_{g,1}(\gamma_g^0, \gamma_g^0)c$. For $A_{1,2}$, we have by Cauchy-Schwarz and Markov inequalities

$$\begin{aligned}
|A_{1,2}| &\leq \|c\|^2 \left(\frac{1}{N_g T} \sum_{i \in G_g^0} \sum_{t=1}^{T} \varepsilon_{it}^4 \left\| \left[2z_{it}(\gamma_g^0) - \frac{1}{T} \sum_{s=1}^{T} \{ z_{is}(\gamma_g^0) + E[z_{is}(\gamma_g^0)] \} \right] \right\|^2 \right)^{1/2} \\
&\times \left(\frac{1}{N_g T} \sum_{i \in G_g^0} \sum_{t=1}^{T} \left\| \frac{1}{T} \sum_{s=1}^{T} \{ z_{is}(\gamma_g^0) - E[z_{is}(\gamma_g^0)] \} \right\|^2 \right)^{1/2} = O_P(1) \, O_p\left(T^{-1/2}\right) = o_p(1).
\end{aligned}$$

Then $A_1 \xrightarrow{d} N(0, \Omega_{g,1}(\gamma_g^0, \gamma_g^0))$ by the Cramér-Wold device and the martingale central limit theorem.

Next, we consider A_2. Note that

$$\begin{aligned}
A_2 &= \frac{1}{T\sqrt{N_g T}} \sum_{i \in \mathbf{G}_g^0} \sum_{s,t=1}^{T} E\left[z_{it}(\gamma_g^0) \varepsilon_{is} \right] + \frac{1}{T\sqrt{N_g T}} \sum_{i \in \mathbf{G}_g^0} \sum_{s,t=1}^{T} \left\{ \left(z_{it}(\gamma_g^0) - E[z_{it}(\gamma_g^0)] \right) \varepsilon_{is} - E\left[z_{it}(\gamma_g^0) \varepsilon_{is} \right] \right\} \\
&\equiv A_{2,1} + A_{2,2}.
\end{aligned}$$

For $A_{2,1}$, we have $A_{2,1} = \sqrt{\frac{N_g}{T}} \frac{1}{N_g T} \sum_{i \in \mathbf{G}_g^0} \sum_{t=1}^{T} \sum_{s<t} E\left[z_{it}(\gamma_g^0) \varepsilon_{is} \right] = \sqrt{\frac{N_g}{T}} \mathbb{B}_{g,NT}$. For $A_{2,2}$, we can easily verify that $E(A_{2,2}) = 0$ and

$$E \|A_{2,2}\|^2 = \frac{1}{N_g T^3} \sum_{i \in \mathbf{G}_g^0} E \left\| \sum_{s,t=1}^{T} \left[\{ z_{it}(\gamma_g^0) - E[z_{it}(\gamma_g^0)] \} \varepsilon_{is} - E\left[z_{it}(\gamma_g^0) \varepsilon_{is} \right] \right] \right\|^2 = O\left(T^{-1}\right)$$

by using the Davydov inequality for strong mixing processes. Then $A_{2,2} = O_p(T^{-1/2})$ and (i2) follows.

(ii) Now, let $u_i \equiv Z_i(\gamma_g^0)' \mathbb{M}_0 \varepsilon_i / \sqrt{N_g T}$. Then we have u_i independent across i and

$$\begin{aligned}
E \|u_i\|^2 &= \frac{1}{N_g T} \sum_{t=1}^{T} \sum_{s=1}^{T} E\left(\tilde{z}_{it}(\gamma_g^0) \tilde{z}_{is}(\gamma_g^0)' \varepsilon_{it} \varepsilon_{is} \right) \\
&\leq \frac{1}{N_g T} \sum_{t=1}^{T} \sum_{s=1}^{T} 6\alpha[|t-s|]^{1/2} \left\| \tilde{z}_{it}(\gamma_g^0) \varepsilon_{it} \right\|_4 \left\| \tilde{z}_{is}(\gamma_g^0) \varepsilon_{is} \right\|_4 \\
&= O(1/N_g).
\end{aligned}$$

By Theorem A of Yang (2016), we have $E \max_i \|u_i\|^{2+\delta} \leq C(N_g)^{-(2+\delta)/2} \max_i \max_{1 \leq t \leq T} \left\| \tilde{z}_{it}(\gamma_g^0) \varepsilon_{it} \right\|_{2+2\delta}$ for some $\delta > 0$ and $C < \infty$. Here $\|\cdot\|_r = \{E \|\cdot\|^r\}^{1/r}$. Then Lindeberg condition holds and we have the desired claim. ■

C Supplementary Lemmas

We first state a technical lemma that is also used in the proof the main results in the paper. Then we study the asymptotic properties of the infeasible estimators.

C.1 A technical lemma

Lemma C.1. Let ξ_t denote a $d_\xi \times 1$ random vector with mean zero and $E \|\xi_t\|^{8+\epsilon} < \infty$ for some $\epsilon > 0$. Suppose that $\{\xi_t, t = 1, ..., T\}$ is strong mixing process with mixing coefficients $\alpha[s] \leq c_\alpha \rho^s$ for some $c_\alpha > 0$ and $\rho \in (0, 1)$. Then as $T \to \infty$ and for any $c > 0$ we have

$$\Pr\left(\left\| \frac{1}{T} \sum_{t=1}^{T} \xi_t \right\| > c \right) = o(T^{-4}).$$

Proof of Lemma C.1. The proof is similar to and simpler than that of Lemma B.1(ii) in Wang, Phillips, and Su (2018) and thus omitted. ■

C.2 Asymptotic properties of the infeasible estimators

We present the analysis of infeasible estimator in this section.
Lemma C.2. Suppose Assumptions A.1, A.3(iv) and A.4 hold. For any $g \in \mathcal{G}$, we have that

$$\check{\gamma}_g - \gamma_g^0 = o_p(1), \text{ and } \check{\theta}_g - \theta_g^0 = o_p((NT)^{-\alpha}).$$

Proof of Lemma C.2. First, we show the convergence rate of $\check{\theta}_g(\gamma)$ for any $\gamma \in \Gamma$. Let $Z_i(\gamma) \equiv ([x_{i1}', x_{i1}' d_{i1}(\gamma)]', ..., [x_{iT}', x_{iT}' d_{iT}(\gamma)]')'$, a $T \times 2K$ matrix. Let $X_i(\gamma_1, \gamma_2) \equiv (x_{i1}[d_{i1}(\gamma_1) - d_{i1}(\gamma_2)], ..., x_{iT}[d_{iT}(\gamma_1) - d_{iT}(\gamma_2)])'$, a $T \times K$ matrix. By the definition of $\check{\theta}_g(\gamma)$, we have $\check{\theta}_g(\gamma) = [\sum_{i \in \mathbf{G}_g^0} Z_i(\gamma)' \mathbb{M}_0 Z_i(\gamma)]^{-1} \sum_{i \in \mathbf{G}_g^0} Z_i(\gamma)' \mathbb{M}_0 Y_i$. It follows that

$$\check{\theta}_g(\gamma) - \theta_g^0 = [\Phi_{1g}(\gamma)]^{-1} \frac{1}{N_g T} \sum_{i \in \mathbf{G}_g^0} Z_i(\gamma)' \mathbb{M}_0 \varepsilon_i - [\Phi_{1g}(\gamma)]^{-1} \Phi_{2g}(\gamma) \delta_g^0. \tag{C.1}$$

where $\Phi_{1g}(\gamma) \equiv \frac{1}{N_g T} \sum_{i \in \mathbf{G}_g^0} Z_i(\gamma)' \mathbb{M}_0 Z_i(\gamma)$, $\Phi_{2g}(\gamma) = \frac{1}{N_g T} \sum_{i \in \mathbf{G}_g^0} Z_i(\gamma)' \mathbb{M}_0 X_i(\gamma, \gamma_g^0)$. By Assumption A.4(i), $\Phi_{1g}(\gamma) = O_p(1)$ for all $\gamma \in \Gamma$. It is standard to show that $\frac{1}{N_g T} \sum_{i \in \mathbf{G}_g^0} Z_i(\gamma)' \mathbb{M}_0 \varepsilon_i = O_p((NT)^{-1/2} + T^{-1})$ and $\Phi_{2g}(\gamma) = O_p(1)$. Then we have $\check{\theta}_g(\gamma) - \theta_g^0 = O_p((NT)^{-\alpha} + T^{-1})$ by exploiting the fact that $\delta_g^0 = O(((NT)^{-\alpha})$. Given the fact that $\alpha < 1/3$ and $N = O(T^2)$, we can conclude from (C.1) that $\check{\theta}_g(\gamma) - \theta_g^0 = O_p((NT)^{-\alpha})$ and

$$\check{\theta}_g(\gamma) - \theta_g^0 = -[\Phi_1(\gamma)]^{-1} \Phi_2(\gamma) \delta_g^0 + o_p((NT)^{-\alpha}). \tag{C.2}$$

Next we show the consistency of $\check{\gamma}_g$. Let $\Phi_{3g}(\gamma) = \frac{1}{N_g T} \sum_{i \in \mathbf{G}_g^0} X_i(\gamma, \gamma_g^0)' \mathbb{M}_0 X_i(\gamma, \gamma_g^0)$. By direct calculations, we can show that

$$\begin{aligned} & \frac{1}{N_g T} \left(\check{\mathcal{Q}}_g(\check{\theta}_g, \check{\gamma}_g) - \check{\mathcal{Q}}_g(\theta_g^0, \gamma_g^0) \right) \\ = \; & \delta_g^{0\prime} \Phi_{3g}(\check{\gamma}_g) \delta_g^0 + (\check{\theta}_g - \theta_g^0)' \Phi_{1g}(\check{\gamma}_g)(\check{\theta}_g - \theta_g^0) + 2(\check{\theta}_g - \theta_g^0)' \Phi_{2g}(\check{\gamma}_g) \delta_g^0 \\ & -(\check{\theta}_g - \theta_g^0)' \frac{2}{N_g T} \sum_{i \in \mathbf{G}_g^0} Z_i(\check{\gamma}_g)' \mathbb{M}_0 \varepsilon_i - \check{\delta}_g' \frac{2}{N_g T} \sum_{i \in \mathbf{G}_g^0} X_i(\check{\gamma}_g, \gamma_g^0)' \mathbb{M}_0 \varepsilon_i. \end{aligned} \tag{C.3}$$

Note that the last two terms on the right hand side (RHS) of the above equation are $o_p((NT)^{-2\alpha})$. This, in conjunction with (C.2) and (C.3) implies that,

$$\frac{1}{N_g T} \left(\check{\mathcal{Q}}_g(\check{\theta}_g, \check{\gamma}_g) - \check{\mathcal{Q}}_g(\theta_g^0, \gamma_g^0) \right) = \delta_g^{0\prime} [\Phi_{3g}(\check{\gamma}_g) - \Phi_{2g}(\check{\gamma}_g)' \Phi_{1g}(\check{\gamma}_g)^{-1} \Phi_{2g}(\check{\gamma}_g)] \delta_g^0 + o_p((NT)^{-2\alpha}).$$

By Assumption A.4(ii), we have that

$$\Phi_3(\check{\gamma}_g) - \Phi_2(\check{\gamma}_g)' \Phi_1(\check{\gamma}_g)^{-1} \Phi_2(\check{\gamma}_g) = \tilde{w}_g(\check{\gamma}_g),$$

which is a $K \times K$ matrix with minimum eigenvalue $\lambda_{\min}[\tilde{w}_g(\check{\gamma}_g)] \geq \tau \min\{1, |\check{\gamma}_g - \gamma_g^0|\}$ w.p.a.1. Hence it follows that

$$(N_g T)^{2\alpha - 1} \left(\check{\mathcal{Q}}_g(\check{\theta}_g, \check{\gamma}_g) - \check{\mathcal{Q}}_g(\theta_g^0, \gamma_g^0) \right) \geq \pi_g^{2\alpha} \|C_g^0\| \tau \min\{1, |\check{\gamma}_g - \gamma_g^0|\} + o_p(1),$$

where we use the fact $\delta_g^0 = (NT)^{-\alpha}$ and $N_g/N \to \pi_g$ by Assumptions A.1(vi) and A.2(iii). On the other hand, we have $\check{\mathcal{Q}}_g(\check{\theta}_g, \check{\gamma}_g) - \check{\mathcal{Q}}_g(\check{\theta}_g, \gamma_g^0) \leq 0$. We can conclude that $\check{\gamma}_g - \gamma_g^0 = o_p(1)$.

Given the consistency of $\check{\gamma}_g$, we can easily show that $\frac{1}{N_g T}\sum_{i\in \mathbf{G}_g^0} Z_i(\check{\gamma}_g)'\mathbb{M}_0 X_i(\gamma_g^0,\check{\gamma}_g) = o_p(1)$. Then $\check{\theta}_g - \theta_g^0 = o_p((NT)^{-\alpha})$ follows. ■

Lemma C.3. Let $h_{it}(\gamma_1,\gamma_2) = \|x_{it}\varepsilon_{it}\|\,|d_{it}(\gamma_2) - d_{it}(\gamma_1)|$ and $k_{it}(\gamma_1,\gamma_2) = \|x_{it}\|\,|d_{it}(\gamma_2) - d_{it}(\gamma_1)|$. Suppose Assumptions A.1(v) and A.5 hold, there is a constant $C_1 < \infty$ such that for $\underline{\gamma} \leq \gamma_1 < \gamma_2 \leq \overline{\gamma}$ and $r \leq 4$,

$$\max_{i,t} E\left[h_{it}(\gamma_1,\gamma_2)\right]^r \leq C_1 \left|\gamma_2 - \gamma_1\right| \text{ and } \max_{i,t} E\left[k_{it}(\gamma_1,\gamma_2)\right]^r \leq C_1 \left|\gamma_2 - \gamma_1\right|.$$

Proof of Lemma C.3. For any random variable Z,

$$E\left[Zd_{it}(\gamma)\right] = E(Z\cdot 1\{q_{it}\leq\gamma\}) = E\left[1\{q_{it}\leq\gamma\}E(Z|q_{it})\right] = \int_{-\infty}^{\gamma} E(Z|q_{it})dF_{it}(q_{it}),$$

where $F_{it}(\cdot)$ is the cumulative distribution function (CDF) of q_{it} with the corresponding PDF $f_{it}(\cdot)$. Taking derivative with respect to γ on both sides yields

$$\frac{d}{d\gamma}E\left[Zd_{it}(\gamma)\right] = E(Z|q_{it}=\gamma)f_{it}(\gamma).$$

Then by the Hölder inequality and Assumptions A.1(v) and A.5

$$\begin{aligned}\frac{d}{d\gamma}E\left[\|x_{it}\varepsilon_{it}\|^r d_{it}(\gamma)\right] &= E(\|x_{it}\varepsilon_{it}\|^r|q_{it}=\gamma)f_{it}(\gamma) \leq \left[E(\|x_{it}\varepsilon_{it}\|^4|q_{it}=\gamma)\right]^{r/4}f_{it}(\gamma)\\ &\leq Cc_f \text{ for some } C<\infty\end{aligned}$$

This implies that

$$\max_{i,t} E\left[h_{it}(\gamma_1,\gamma_2)\right]^r \leq C_1\left|\gamma_2-\gamma_1\right| \text{ with } C_1 = Cc_f.$$

Analogously, we have $\max_{i,t} E\left[k_{it}(\gamma_1,\gamma_2)\right]^r \leq C_1\left|\gamma_2-\gamma_1\right|$. ■

Lemma C.4. Suppose Assumptions A.1, A.3(iii)–(iv) and A.4–A.5 hold. Then there exists a constant $C_2 < \infty$ such that for all $\underline{\gamma} \leq \gamma_1 < \gamma_2 \leq \overline{\gamma}$ and $g \in \mathcal{G}$

$$\begin{aligned}E\left|\frac{1}{\sqrt{N_g T}}\sum_{i\in\mathbf{G}_g^0}\sum_{t=1}^{T}\left(h_{it}^2(\gamma_1,\gamma_2) - Eh_{it}^2(\gamma_1,\gamma_2)\right)\right|^2 &\leq C_2\left|\gamma_2-\gamma_1\right|,\\ E\left|\frac{1}{\sqrt{N_g T}}\sum_{i\in\mathbf{G}_g^0}\sum_{t=1}^{T}\left(k_{it}^2(\gamma_1,\gamma_2) - Ek_{it}^2(\gamma_1,\gamma_2)\right)\right|^2 &\leq C_2\left|\gamma_2-\gamma_1\right|.\end{aligned}$$

Proof of Lemma C.4. For notational simplicity, let $h_{it}^r(\gamma_1,\gamma_2) = \left[h_{it}(\gamma_1,\gamma_2)\right]^r$ for $r \geq 0$. By the independence across i and strong mixing over t for $\{(x_{it},q_{it},\varepsilon_{it})\}$, there is a constant $C^\dagger$ such that

$$\begin{aligned}& E\left|\frac{1}{\sqrt{N_g T}}\sum_{i\in\mathbf{G}_g^0}\sum_{t=1}^{T}\left\{h_{it}^2(\gamma_1,\gamma_2) - E[h_{it}^2(\gamma_1,\gamma_2)]\right\}\right|^2\\ =\ & \frac{1}{N_g}\sum_{i\in\mathbf{G}_g^0} E\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left\{h_{it}^2(\gamma_1,\gamma_2) - E[h_{it}^2(\gamma_1,\gamma_2)]\right\}\right|^2\\ \leq\ & \frac{C^\dagger}{N_g}\sum_{i\in\mathbf{G}_g^0}\frac{1}{T}\sum_{t=1}^{T} E\left\{h_{it}^2(\gamma_1,\gamma_2) - E\left[h_{it}^2(\gamma_1,\gamma_2)\right]\right\}^2\\ \leq\ & \frac{C^\dagger}{N_g T}\sum_{i\in\mathbf{G}_g^0}\sum_{t=1}^{T} E\left[h_{it}^4(\gamma_1,\gamma_2)\right] \leq C^\dagger C_1\left|\gamma_2-\gamma_1\right|.\end{aligned}$$

The first result follows by setting $C_2 = C^{\dagger} C_1$. Analogously, we can prove the second result in the lemma. ■

Lemma C.5. Let $J_{g,NT}(\gamma) = N_g^{-1/2} T^{-1/2} \sum_{i \in \mathbf{G}_g^0} \sum_{t=1}^{T} x_{it} e_{it} d_{it}(\gamma)$. Suppose Assumptions A.1, A.3(iii)–(iv) and A.4–A.5 hold, there are constants K_1 and K_2 such that for all γ_g, $g \in \mathcal{G}$, $\epsilon > 0$, $\eta > 0$ and $\delta \geq (N_g T)^{-1}$, if $\sqrt{N_g T} \geq K_2/\eta$, then

$$\Pr\left(\sup_{\gamma' \leq \gamma \leq \gamma' + \delta} |J_{g,NT}(\gamma) - J_{g,NT}(\gamma')| > \eta \right) \leq \frac{K_1 \delta^2}{\eta^4}.$$

Proof of Lemma C.5. The proof is similar to that of Lemma A.3 in Hansen (2000). ■

Lemma C.6. Suppose Assumptions A.1, A.3(iii)–(iv) and A.4–A.6 hold, we have for $g \in \mathcal{G}$,

$$J_{g,NT}(\gamma) \Rightarrow J_g(\gamma),$$

a mean-zero Gaussian process with almost surely continuous sample paths.
Proof of Lemma C.6. The proof is similar to that of Lemma A.4 of Hansen (2000). ■

Lemma C.7. Let $G_{g,NT}(\gamma) = \frac{1}{N_g T} \sum_{i \in \mathbf{G}_g^0} \sum_{t=1}^{T} C_g^{0\prime} x_{it} x_{it}' C_g^0 [d_{it}(\gamma) - d_{it}(\gamma_g^0)]$ and $K_{g,NT}(\gamma) = \frac{1}{N_g T} \sum_{i \in \mathbf{G}_g^0} \sum_{t=1}^{T} \|x_{it}\| \left| d_{it}(\gamma) - d_{it}(\gamma_g^0) \right|$. Under Assumptions A.1, A.3(iii)–(iv) and A.4–A.5, there exist constants $B > 0$, $0 < d < \infty$, such that for all $\eta > 0$ and $\epsilon > 0$, there exists a $\overline{v} < \infty$ such that for all (N, T) and $g \in \mathcal{G}$,

$$\Pr\left(\inf_{\overline{v}/\alpha_{NT} \leq |\gamma - \gamma_g^0| \leq B} \frac{G_{g,NT}(\gamma)}{|\gamma_g - \gamma_g^0|} < (1 - \eta) d \right) \leq \epsilon,$$
$$\Pr\left(\sup_{\overline{v}/\alpha_{NT} \leq |\gamma - \gamma_g^0| \leq B} \frac{K_{g,NT}(\gamma)}{|\gamma_g - \gamma_g^0|} > (1 + \eta) k \right) \leq \epsilon.$$

Proof of Lemma C.7. The proof is similar to that of Lemma A.7 of Hansen (2000). ■

Lemma C.8: Under Assumptions A.1, A.3(iii)–(iv) and A.4–A.5, there exists some $\overline{v} < \infty$ such that for any $B < \infty$ and $g = 1, ..., G$,

$$\Pr\left(\sup_{\overline{v}/\alpha_{NT} \leq |\gamma - \gamma_g^0| \leq B} \frac{\left| J_{g,NT}(\gamma) - J_{g,NT}(\gamma_g^0) \right|}{\sqrt{\alpha_{NT}} \left| \gamma - \gamma_g^0 \right|} > \eta \right) \leq \epsilon.$$

Proof of Lemma C.8. The proof is similar to that of Lemma A.8 of Hansen (2000). ■

Lemma C.9. Let $\tilde{K}_{g,NT}(\gamma) = N_g^{-1} \sum_{i \in \mathbf{G}_g^0} \left[T^{-1} \sum_{t=1}^{T} \|x_{it}\| \left| d_{it}(\gamma) - d_{it}(\gamma_g^0) \right| \right]^2$ and $\tilde{J}_{g,NT}(\gamma) = N_g^{-1/2} T^{-3/2} \sum_{i \in \mathbf{G}_g^0} \sum_{t=1}^{T} \sum_{s=1}^{T} x_{is} e_{it} (d_{is}(\gamma) - d_{is}(\gamma_g^0))$. Suppose Assumptions A.1, A.3(iii)–(iv) and A.4–A.5 hold. Then there exists some $\overline{v} < \infty$ and $B > 0$ such that for any $\eta > 0$, $\epsilon > 0$ and $g \in \mathcal{G}$,

$$\Pr\left(\sup_{\overline{v}/\alpha_{NT} \leq |\gamma - \gamma_g^0| \leq B} \frac{\left| \tilde{J}_{g,NT}(\gamma) \right|}{\sqrt{\alpha_{N_g T}} \left| \gamma - \gamma_g^0 \right|} > \eta \right) \leq \epsilon \text{ and } \Pr\left(\sup_{\overline{v}/\alpha_{NT} \leq |\gamma - \gamma_g^0| \leq B} \frac{\left| \tilde{K}_{g,NT}(\gamma) \right|}{\left| \gamma - \gamma_g^0 \right|} > \eta \right) \leq \epsilon.$$

Proof of Lemma C.9. The analysis for the first result is analogous to that of Lemma C.8. For the second result, we consider the case $\gamma > \gamma_g^0$. Letting $k_{it}(\gamma) = k_{it}(\gamma, \gamma_g^0) = \|x_{it}\| \left|d_{it}(\gamma) - d_{it}(\gamma_g^0)\right|$, we have

$$\begin{aligned} E[\tilde{K}_{g,NT}(\gamma)] &= N_g^{-1} \sum_{i \in \mathbf{G}_g^0} E\left[T^{-1} \sum_{t=1}^{T} \|x_{it}\| \left|d_{it}(\gamma) - d_{it}(\gamma_g^0)\right|\right]^2 = N_g^{-1} \sum_{i \in \mathbf{G}_g^0} E\left[T^{-1} \sum_{t=1}^{T} k_{it}(\gamma)\right]^2 \\ &= N_g^{-1} \sum_{i \in \mathbf{G}_g^0} \mathrm{Var}\left[T^{-1} \sum_{t=1}^{T} k_{it}(\gamma)\right]^2 + N_g^{-1} \sum_{i \in \mathbf{G}_g^0} \left(\frac{1}{T} \sum_{t=1}^{T} E[k_{it}(\gamma)]\right)^2 \\ &\leq C^{\dagger} N_g^{-1} \sum_{i \in \mathbf{G}_g^0} T^{-2} \sum_{t=1}^{T} E\left[k_{it}(\gamma)^2\right] + N_g^{-1} \sum_{i \in \mathbf{G}_g^0} \left(\frac{1}{T} \sum_{t=1}^{T} E[k_{it}(\gamma)]\right)^2 \\ &\leq \frac{C^{\dagger} C_1}{T} \left|\gamma - \gamma_g^0\right| + C_1^2 \left|\gamma - \gamma_g^0\right|^2, \end{aligned}$$

where the first inequality follow from the fact that $\mathrm{Var}\left[T^{-1} \sum_{t=1}^{T} k_{it}(\gamma)\right]^2 \leq C^{\dagger} T^{-2} \sum_{t=1}^{T} \mathrm{Var}[k_{it}(\gamma)] \leq C^{\dagger} T^{-2} \sum_{t=1}^{T} E[k_{it}(\gamma)]^2$ for some $C^{\dagger} < \infty$ by using the fact that $\{k_{it}(\gamma), t \geq 1\}$ is also a strong mixing process, and the last inequality follows from Lemma C.2.

First we consider the case $\gamma - \gamma_g^0 > 0$. Choose a $b > 1$, $B < \epsilon(b-1)\eta/(4C_1^2 b^3)$ and $\overline{v}$ such that $\overline{v}/\alpha_{NT} < B$. We set $\gamma_j = \gamma_g^0 + b^{j-1}\overline{v}/\alpha_{NT}$ for $j = 1, ..., n+1$ such that $\gamma_n + 1 \geq B$ and $\gamma_n \leq B$. Since $\frac{b^n \overline{v}}{\alpha_{NT}} \leq B$, $n \leq \log_b(B\alpha_{NT}/\overline{v})$. When (N,T) is large enough,, we can have $\frac{C^{\dagger} C_1 b}{\eta} \frac{n}{T} \leq \epsilon/4$. Then we can calculate

$$\begin{aligned} \Pr\left(\sup_{1 \leq j \leq n} \frac{\tilde{K}_{g,NT}(\gamma_{j+1})}{\left|\gamma_j - \gamma_g^0\right|} > \eta\right) &\leq \sum_{j=1}^{n} \frac{E[\tilde{K}_{g,NT}(\gamma_{j+1})]}{\eta \left|\gamma_j - \gamma_g^0\right|} \\ &\leq \sum_{j=1}^{n} \frac{C^{\dagger} C_1 \left|\gamma_{j+1} - \gamma_g^0\right|/T}{\eta \left|\gamma_j - \gamma_g^0\right|} + \sum_{j=1}^{n} \frac{C_1^2 \left|\gamma_{j+1} - \gamma_g^0\right|^2}{\eta \left|\gamma_j - \gamma_g^0\right|} \\ &= \frac{C^{\dagger} C_1 b}{\eta} \frac{n}{T} + \frac{C_1^2 b^2}{\eta} \frac{\overline{v}(b^{n+1} - 1)}{\alpha_{NT}(b-1)} \\ &\leq \frac{C^{\dagger} C_1 b}{\eta} \frac{n}{T} + \frac{C_1^2 b^3}{\eta} \frac{B}{(b-1)} < \epsilon/2. \end{aligned}$$

For any $\gamma \in [\gamma_g^0 + \overline{v}/\alpha_{NT}, \gamma_g^0 + B]$, there exists a $j \in \{1, ..., n\}$ such that $\gamma_j \leq \gamma \leq \gamma_{j+1}$. In view of the fact that $\tilde{K}_{g,NT}(\gamma)$ is monotonic in γ, we have $\frac{\tilde{K}_{g,NT}(\gamma)}{\left|\gamma - \gamma_g^0\right|} \leq \frac{\tilde{K}_{g,NT}(\gamma_{j+1})}{\left|\gamma_j - \gamma_g^0\right|}$. It follows that

$$\Pr\left(\sup_{\overline{v}/\alpha_{NT} \leq \gamma_g - \gamma_g^0 \leq B} \frac{\left|\tilde{K}_{g,NT}(\gamma)\right|}{\left|\gamma - \gamma_g^0\right|} > \eta\right) \leq \Pr\left(\sup_{1 \leq j \leq n} \frac{\tilde{K}_{g,NT}(\gamma_{j+1})}{\left|\gamma_j - \gamma_g^0\right|} > \eta\right) \leq \epsilon/2.$$

A symmetric argument gives us the proof for the case $-B \leq \gamma - \gamma_g^0 \leq -\overline{v}/\alpha_{NT}$. This completes our proof. ■

Lemma C.10. Suppose that Assumptions A.1, A.3(iii)–(iv) and A.4–A.5 hold. Then we have $\alpha_{NT}(\check{\gamma}_g - \gamma_g^0) = O_p(1)$ for all $g \in \mathcal{G}$.

Proof of Lemma C.10: Let B, d, k be the coefficients defined in Lemma C.6-C.8 and $c = \left\|C_g^0\right\|$. Pick an η such that $\min\{1, c, k\} > \eta > 0$ and $\pi_g^{2\alpha}(1-\eta)d - 24\pi_g^{\alpha} c k \eta - \pi_g^{2\alpha}(6ck + 4c^2)\eta > 0$. Let $\mathbf{E}_{NT}$ be the joint event that, for all $g \in \mathcal{G}$: $\left|\check{\gamma}_g - \gamma_g^0\right| \leq B$, $(NT)^{\alpha} \left\|\check{\beta}_g - \beta_g^0\right\| \leq \eta$, $(NT)^{\alpha} \left\|\check{\delta}_g - \delta_g^0\right\| \leq \eta$,

$$\begin{aligned}
\inf_{\overline{v}/\alpha_{NT} \leq \left|\gamma - \gamma_g^0\right| \leq B} \frac{G_{g,NT}(\gamma)}{\left|\gamma - \gamma_g^0\right|} &\geq (1-\eta)d, \\
\sup_{\overline{v}/\alpha_{NT} \leq \left|\gamma - \gamma_g^0\right| \leq B} \frac{K_{g,NT}(\gamma)}{\left|\gamma - \gamma_g^0\right|} &\leq (1+\eta)k, \\
\sup_{\overline{v}/\alpha_{NT} \leq \left|\gamma - \gamma_g^0\right| \leq B} \frac{\left|J_{g,NT}(\gamma) - J_{g,NT}(\gamma_g^0)\right|}{\sqrt{\alpha_{N_gT}} \left|\gamma - \gamma_g^0\right|} &\leq \eta, \\
\sup_{\overline{v}/\alpha_{NT} \leq \left|\gamma - \gamma_g^0\right| \leq B} \frac{\left|\tilde{J}_{g,NT}(\gamma)\right|}{\sqrt{\alpha_{N_gT}} \left|\gamma - \gamma_g^0\right|} &\leq \eta, \\
\sup_{\overline{v}/\alpha_{NT} \leq \left|\gamma - \gamma_g^0\right| \leq B} \frac{\tilde{K}_{g,NT}(\gamma)}{\left|\gamma - \gamma_g^0\right|} &\leq \eta.
\end{aligned}$$

Then by Lemma C.7-C.9.Let $X_i(\gamma, \gamma_g^0) \equiv (x_{i1}[d_{i1}(\gamma) - d_{i1}(\gamma_g^0)], ..., x_{iT}[d_{iT}(\gamma) - d_{iT}(\gamma_g^0)])'$, a $T \times K$ matrix. Let $\Delta\mathbf{X}_{g,\gamma} \equiv \mathbf{X}_g(\gamma, \gamma_g^0) \equiv \left\{X_i(\gamma, \gamma_g^0), i \in \mathbf{G}_g^0\right\}$, which is an $N_gT \times K$ matrix. Let $Z_i(\gamma) \equiv ([x_{i1}', x_{i1}'d_{i1}(\gamma)]', ..., [x_{iT}', x_{iT}'d_{iT}(\gamma)]')'$, a $T \times 2K$ matrix. Let $\mathbf{Z}_g(\gamma) \equiv \left\{Z_i(\gamma), i \in \mathbf{G}_g^0\right\}$, which is an $N_gT \times 2K$ matrix. Let $\Delta\overline{\mathbf{X}}_{g,\gamma} = (I_{N_g} \otimes \mathbb{P}_0)\Delta\mathbf{X}_{g,\gamma}$, and $\overline{\mathbf{Z}}_g(\gamma_g^0) = (I_{N_g} \otimes \mathbb{P}_0)\mathbf{Z}_g(\gamma_g^0)$ where recall that $\mathbb{P}_0 = T^{-1}\iota_T\iota_T'$. Let $\varepsilon_i = (\varepsilon_{i1}, ..., \varepsilon_{iT})'$ and $\boldsymbol{\varepsilon}_g = \left\{\varepsilon_i, i \in \mathbf{G}_g^0\right\}$, an $N_gT \times 1$ vector.

$$\begin{aligned}
\check{Q}_{g,NT}(\theta, \gamma) - \check{Q}_{g,NT}(\theta, \gamma_g^0) &= \delta'\Delta\mathbf{X}_{g,\gamma}'(I_{N_g} \otimes \mathbb{M}_0)\Delta\mathbf{X}_{g,\gamma}\delta - 2\delta'\Delta\mathbf{X}_{g,\gamma}'(I_{N_g} \otimes \mathbb{M}_0)\mathbf{Z}_g(\gamma_g^0)(\theta - \theta_g^0) \\
&\quad + 2\delta'\Delta\mathbf{X}_{g,\gamma}'(I_{N_g} \otimes \mathbb{M}_0)\boldsymbol{\varepsilon}_g \\
&= \delta_g^{0\prime}\Delta\mathbf{X}_{g,\gamma}'\Delta\mathbf{X}_{g,\gamma}^0\delta_g^0 + (\delta - \delta_g^0)'\Delta\mathbf{X}_{g,\gamma}'\Delta\mathbf{X}_{g,\gamma}(\delta + \delta_g^0) - \delta'\Delta\overline{\mathbf{X}}_{g,\gamma}\Delta\overline{\mathbf{X}}_{g,\gamma}\delta \\
&\quad - 2\delta'\Delta\mathbf{X}_{g,\gamma}'\mathbf{Z}_g(\gamma_g^0)(\theta - \theta_g^0) + 2\delta'\Delta\overline{\mathbf{X}}_{g,\gamma}'\overline{\mathbf{Z}}_g(\gamma_g^0)(\theta - \theta_g^0) + 2\delta'\Delta\mathbf{X}_{g,\gamma}'\boldsymbol{\varepsilon}_g \\
&\quad - 2\delta'\Delta\overline{\mathbf{X}}_{g,\gamma}'\boldsymbol{\varepsilon}_g.
\end{aligned}$$

Let $\check{\delta}_g = (NT)^{-\alpha}C_g$ for some C_g such that $\left\|C_g - C_g^0\right\| \leq \kappa$, implied by $\mathbf{E}_{NT}$. Suppose that $\mathbf{E}_{NT}$ happens and for $\gamma \in [\gamma_g^0 + \overline{v}/\alpha_{NT}, \gamma_g^0 + B]$, we have

$$\begin{aligned}
& (N_gT)^{2\alpha-1}\frac{\check{Q}_{g,NT}(\check{\theta}_g, \gamma) - \check{Q}_{g,NT}(\check{\theta}_g, \gamma_g^0)}{\left|\gamma - \gamma_g^0\right|} \\
= \; & \frac{\pi_g^{2\alpha}C_g^{0\prime}\Delta\mathbf{X}_{g,\gamma}'\Delta\mathbf{X}_{g,\gamma}C_g^0}{N_gT\left|\gamma - \gamma_g^0\right|} + \frac{\pi_g^{2\alpha}(C_g + C_g^0)'\Delta\mathbf{X}_{g,\gamma}'\Delta\mathbf{X}_{g,\gamma}(C_g - C_g^0)}{(N_gT)\left|\gamma - \gamma_g^0\right|} - \frac{\pi_g^{2\alpha}C_g'\Delta\overline{\mathbf{X}}_{g,\gamma}\Delta\overline{\mathbf{X}}_{g,\gamma}C_g}{(N_gT)\left|\gamma - \gamma_g^0\right|} \\
& -2\frac{\pi_g^{\alpha}C_g'\left[\Delta\mathbf{X}_{g,\gamma}'\mathbf{Z}_g(\gamma_g^0) - \Delta\overline{\mathbf{X}}_{g,\gamma}'\overline{\mathbf{Z}}_g(\gamma_g^0)\right](N_gT)^{\alpha}(\check{\theta}_g - \theta_g^0)}{(N_gT)\left|\gamma - \gamma_g^0\right|} + 2\left\|\pi_g^{\alpha}C_g\right\|\frac{\Delta\mathbf{X}_{g,\gamma}'\boldsymbol{\varepsilon}_g - \Delta\overline{\mathbf{X}}_{g,\gamma}'\boldsymbol{\varepsilon}_g}{(N_gT)^{1-\alpha}\left|\gamma - \gamma_g^0\right|} \\
\geq \; & \frac{\pi_g^{2\alpha}G_{g,NT}(\gamma)}{\left|\gamma - \gamma_g^0\right|} - \pi_g^{2\alpha}(\left\|C_g^0\right\| + \left\|C_g^0\right\|)\left\|C_g - C_g^0\right\|\frac{K_{g,NT}(\gamma)}{\left|\gamma - \gamma_g^0\right|} - \pi_g^{2\alpha}\left\|C_g\right\|^2\frac{\tilde{K}_{g,NT}(\gamma)}{\left|\gamma - \gamma_g^0\right|} \\
& -4\pi_g^{\alpha}\left\|C_g\right\|(N_gT)^{\alpha}\left\|\check{\theta}_g - \theta_g^0\right\|\frac{K_{g,NT}(\gamma) + \tilde{K}_{g,NT}(\gamma)}{\left|\gamma - \gamma_g^0\right|} - 2\left\|\pi_g^{\alpha}C_g\right\|\frac{\left|J_{g,NT}(\gamma) - J_{g,NT}(\gamma_g^0)\right|}{\sqrt{\alpha_{N_gT}}\left|\gamma - \gamma_g^0\right|} \\
& -2\left\|\pi_g^{\alpha}C_g\right\|\frac{\left|\tilde{J}_{g,NT}(\gamma)\right|}{\sqrt{\alpha_{NT}}\left|\gamma - \gamma_g^0\right|}
\end{aligned}$$

$$
\begin{aligned}
&\geq \pi_g^{2\alpha}(1-\eta)d - \pi_g^{2\alpha}(2c+\eta)\eta(1+\eta)k - \pi_g^{2\alpha}(c+\eta)^2\eta - 4\pi_g^{\alpha}(c+\eta)\eta[(1+\eta)k+\eta] \\
&\quad -4\pi_g^{\alpha}(c+\eta)\eta \\
&> \pi_g^{2\alpha}(1-\eta)d - 24\pi_g^{\alpha}ck\eta - \pi_g^{2\alpha}(6ck+4c^2)\eta \\
&> 0,
\end{aligned}
$$

which indicates that $\check{\gamma}_g$ does not belong to $[\gamma_g^0+\overline{v}/\alpha_{NT}, \gamma_g^0+B]$. A symmetric argument shows that if $\mathbf{E}_{NT}$ happens $\check{\gamma}_g$ does not belong to $[\gamma_g^0-B, \gamma_g^0-\overline{v}/\alpha_{NT},]$. Hence, we have shown $\check{\gamma}_g-\gamma_g^0=O_p(1/\alpha_{NT})$ for all $g\in\mathcal{G}$. ■

Lemma C.11. Let $G_{g,NT}^*(v)=\alpha_{N_gT}G_{g,NT}(\gamma_g^0+v/\alpha_{N_gT})$ and $K_{NT}^*(v)=\alpha_{N_gT}K_{g,NT}(\gamma_g^0+v/\alpha_{N_gT})$. Suppose that Assumptions A.1, A.3(iii)–(iv) and A.4–A.5 hold. Then we have that uniformly in $v\in\Psi$,

$$
G_{g,NT}^*(v)\xrightarrow{p} w_{g,D}\left|v\right|, \text{ and } K_{g,NT}^*(v)\xrightarrow{p} D_g^0\left|v\right|
$$

where $w_{g,D}=C_g^{0\prime}D_g^0C_g^0$ for $g\in\mathcal{G}$ and Ψ is a compact set.

Proof of Lemma C.11. The proof is similar to that of Lemma A.10 in Hansen (2000). ■

Lemma C.12. Let $R_{g,NT}(v)=\sqrt{\alpha_{N_gT}}\left[J_{g,NT}(\gamma_g^0+v/\alpha_{N_gT})-J_{g,NT}(\gamma_g^0)\right]$. Suppose that Assumptions A.1, A.3(iii)–(iv) and A.4–A.5 hold. Then on any compact set Ψ,

$$
R_{g,NT}(v)\Rightarrow B_g(v)
$$

where $B_g(v)$ is a vector Brownian motion with covariance matrix $E\left[B_g(1)B_g(1)'\right]=V_g^0$.

Proof of Lemma C.12. First, we show the convergence of finite dimensional distribution: $R_{g,NT}(v)\xrightarrow{d} N(0,V_g^0)$. Let $u_{ni}(v)\equiv\frac{1}{\sqrt{N_gT}}\sum_{t=1}^T x_{it}\varepsilon_{it}\sqrt{\alpha_{N_gT}}[d_{it}(\gamma_g^0+v/\alpha_{N_gT})-d_{it}(\gamma_g^0)]$ and $\mathcal{F}_i=\sigma(\{u_{nj}(v),\, j\leq i\})$. By Assumption A.1(ii) and Liapunov's central limit theorem (e..g., Theorem 23.11 of Davidson (1994, pp.372-373), it suffices to verify that

$$
\sum_{i\in\mathbf{G}_g^0}u_{ni}(v)u_{ni}(v)'\xrightarrow{p}\left|v\right|V_g^0 \text{ and } \sum_{i\in\mathbf{G}_g^0}\left\|u_{ni}(v)\right\|^4=o_p(1).
$$

Note that

$$
\begin{aligned}
\sum_{i\in\mathbf{G}_g^0}u_{ni}(v)u_{ni}(v)' &= \frac{\alpha_{N_gT}}{N_gT}\sum_{i\in\mathbf{G}_g^0}\sum_{t=1}^T x_{it}x_{it}'\varepsilon_{it}^2\left|d_{it}(\gamma_g^0+v/\alpha_{N_gT})-d_{it}(\gamma_g^0)\right| \\
&\quad+\frac{\alpha_{N_gT}}{N_gT}\sum_{i\in\mathbf{G}_g^0}\sum_{1\leq s\neq t\leq T}x_{it}\varepsilon_{it}x_{is}'\varepsilon_{is}[d_{it}(\gamma_g^0+v/\alpha_{N_gT})-d_{it}(\gamma_g^0)][d_{is}(\gamma_g^0+v/\alpha_{N_gT})-d_{is}(\gamma_g^0)] \\
&\equiv A_{g,NT}+B_{g,NT}.
\end{aligned}
$$

For $A_{g,NT}$, we can conduct similar calculations as used in the proof of Lemma C.3 to obtain

$$
\frac{E\left[x_{it}x_{it}'\varepsilon_{it}^2\left|d_{it}(\gamma_g^0+v/\alpha_{N_gT})-d_{it}(\gamma_g^0)\right|\right]}{v/\alpha_{N_gT}}\to E(x_{it}x_{it}'\varepsilon_{it}^2|q_{it}=\gamma_g^0).
$$

Then we can readily show $A_{g,NT}\to\left|v\right|V_g^0$ by using the Chebyshev inequality and the fact that $\{(x_{it},q_{it},\varepsilon_{it})\}$ is independent across i and strong mixing along the time dimension. Let $\zeta_{it}=x_{it}\varepsilon_{it}[d_{it}(\gamma_g^0+v/\alpha_{N_gT})-$

$d_{it}(\gamma_g^0)]$. For $B_{g,NT}$, we have for any $K \times 1$ nonrandom vector c with $\|c\| = 1$, we have

$$
\begin{aligned}
|E[c'B_{g,NT}c]| &= \frac{\alpha_{N_gT}}{N_gT}\left|\sum_{i\in\mathbf{G}_g^0}\sum_{1\le s\ne t\le T}\mathrm{Cov}(c'\zeta_{it}, c'\zeta_{is})\right| \le \frac{\alpha_{N_gT}}{N_gT}\sum_{i\in\mathbf{G}_g^0}\sum_{s=1}^{T-1}\sum_{t=s+1}^{T_0}|\mathrm{Cov}(c'\zeta_{it}, c'\zeta_{is})| \\
&= \frac{\alpha_{N_gT}}{N_gT}\sum_{i\in\mathbf{G}_g^0}\sum_{0<|s-t|\le T_0}|\mathrm{Cov}(c'\zeta_{it}, c'\zeta_{is})| + \frac{\alpha_{N_gT}}{N_gT}\sum_{i\in\mathbf{G}_g^0}\sum_{|s-t|>T_0}|\mathrm{Cov}(c'\zeta_{it}, c'\zeta_{is})| \\
&\le 2T_0\alpha_{N_gT}\max_i\max_{0<|s-t|\le T_0}|\mathrm{Cov}(c_1'\zeta_{it}, c_2'\zeta_{is})| + \frac{T\alpha_{N_gT}}{N_g}\sum_{i\in\mathbf{G}_g^0}\{\alpha_i[T_0]\}^{(3+\epsilon_0)/(4+\epsilon_0)}\max_{i,t}\|\zeta_{it}\|_{8+\epsilon_0}^2 \\
&\le T_0\alpha_{N_gT}O((\alpha_{N_gT})^{-2}) + CT\alpha_{N_gT}\rho^{T_0(3+\epsilon_0)/(4+\epsilon_0)} = o(1)
\end{aligned}
$$

provided T_0 is chosen such that $T_0 = o(\alpha_{N_gT})$ and $T_0/(\ln T)^{c_0} \to \infty$ for some constant $c_0 > 1$. This implies that $E[B_{g,NT}] = o(1)$. In addition, it is easy to verify that $\mathrm{Var}[c'B_{g,NT}] = o(1)$. Then we have $B_{g,NT} = o_p(1)$. Consequently, $\sum_{i\in\mathbf{G}_g^0} u_{ni}(v)u_{ni}(v)' \xrightarrow{p} |v|\, V_g^0$.

Now, we verify that $\sum_{i\in\mathbf{G}_g^0}\|u_{ni}(v)\|^4 = o_p(1)$. Note that

$$
\begin{aligned}
\sum_{i\in\mathbf{G}_g^0} E[c'u_{ni}(v)]^4 &= \frac{\alpha_{N_gT}^2}{(N_gT)^2}\sum_{i\in\mathbf{G}_g^0} E\left|c'\sum_{t=1}^T\zeta_{it}\right|^4 \\
&= \frac{\alpha_{N_gT}^2}{(N_gT)^2}\sum_{i\in\mathbf{G}_g^0}\sum_{t=1}^T E(c'\zeta_{it})^4 + o(1) \\
&= O(\alpha_{N_gT}(N_gT)^{-1}) + o(1) = o(1),
\end{aligned}
$$

where the second equality follows from the simple application of the Davydov inequality for strong mixing processes and similar arguments as used in the analysis of $B_{g,NT}$. Then $\sum_{i\in\mathbf{G}_g^0}\|u_{ni}(v)\|^4 = o_p(1)$ by Markov inequality. Then the pointwise distributional result follows.

For the stochastic equicontinuity, the proof procedure is similar to that in Hansen (2000) and thus omitted. ■

Lemma C.13. Let $\tilde{K}_{g,NT}^*(v) = \alpha_{N_gT}\tilde{K}_{g,NT}(\gamma_g^0 + v/\alpha_{N_gT})$ and $\tilde{J}_{g,NT}^*(v) = \alpha_{N_gT}\tilde{J}_{g,NT}(\gamma_g^0 + v/\alpha_{N_gT})$. Suppose that Assumptions A.1, A.3(iii)-(iv) and A.4–A.5 hold. Then $\tilde{K}_{g,NT}^*(v) \xrightarrow{p} 0$ and $\tilde{J}_{g,NT}^*(v) \xrightarrow{p} 0$ uniformly in $v \in \Psi$, where Ψ is a compact set.

Proof of Lemma C.13. By the proof of Lemma C.9, we have

$$
E[\tilde{K}_{g,NT}^*(v)] = \alpha_{N_gT}O\left(\frac{1}{T}|v/\alpha_{N_gT}| + |v/\alpha_{N_gT}|^2\right) = o(1).
$$

Let $\kappa_i(v) = T^{-1}\sum_{t=1}^T\kappa_{it}(v)$, where $\kappa_{it}(v) = \|x_{it}\|\left|d_{it}(\gamma_g^0 + v/\alpha_{N_gT}) - d_{it}(\gamma_g^0)\right|$. Let $\tilde{\kappa}_{it}(v) = \kappa_{it}(v) -$

$E[\kappa_{it}(v)]$. Then

$$\begin{aligned}
\mathrm{Var}(\tilde{K}^*_{g,NT}(v)) &= \alpha^2_{N_gT} E\left[\left|N_g^{-1}\sum_{i\in G_g^0}\{\kappa_i(v)^2 - E[\kappa_i(v)^2]\}\right|^2\right] = \frac{\alpha^2_{N_gT}}{N_g^2}\sum_{i\in G_g^0} E\{\kappa_i(v)^2 - E[\kappa_i(v)^2]\}^2 \\
&\le \frac{\alpha^2_{N_gT}}{N_g^2}\sum_{i\in \mathbf{G}_g^0} E\left[T^{-1}\sum_{t=1}^T \kappa_{it}(v)\right]^4 \\
&\le \frac{8\alpha^2_{N_gT}}{N_g^2}\sum_{i\in \mathbf{G}_g^0} E\left[T^{-1}\sum_{t=1}^T \tilde{\kappa}_{it}(v)\right]^4 + \frac{8\alpha^2_{N_gT}}{N_g^2}\sum_{i\in G_g^0} E\left[T^{-1}\sum_{t=1}^T E\left[\kappa_{it}(v)\right]\right]^4 \\
&\le \frac{C\alpha^2_{N_gT}}{N_g^2T^4}\sum_{i\in \mathbf{G}_g^0}\left\{\sum_{t=1}^T E\left[\tilde{\kappa}_{it}(v)\right]^4 + \left(\sum_{t=1}^T E\left[\tilde{\kappa}_{it}(v)\right]^2\right)^2\right\} + o\left(1\right) + O(N_g^{-1}\alpha^{-2}_{N_gT}) \\
&= O\left(\alpha_{N_gT}N_g^{-1}T^{-3} + N_g^{-1}T^{-2}\right) + o\left(1\right) = o\left(1\right),
\end{aligned}$$

where the first equality follows from the Jensen inequality, the second inequality follows from the C_r inequality, the third one follows from the repeated application of Davydov inequality and the fact that $\max_{i,t} E\left[\kappa_{it}(v)\right] = O(\alpha^{-1}_{N_gT})$, and the next to last equality holds by the moment calculations. Then $\tilde{K}^*_{g,NT}(v) = o_p\left(1\right)$ for each $v\in\Psi$. This result, in conjunction with the monotonicity of $\tilde{K}^*_{g,NT}(v)$ in either the half line $[0,\infty)$ or the half line $(-\infty,0]$, implies that $\tilde{K}^*_{g,NT}(v) \xrightarrow{p} 0$ uniformly in $v\in\Psi$. See Hansen (2000, p. 598).

For $\tilde{J}^*_{g,NT}(v)$, we can follow the above arguments and show that $\tilde{J}^*_{g,NT}(v) = o_p\left(1\right)$ for each $v\in\Psi$. Following Lemma A.11 in Hansen (2000), we can readily show the tightness of the process $\{\tilde{J}^*_{g,NT}(v)\}$. As a result, we have $\tilde{J}^*_{g,NT}(v) \xrightarrow{p} 0$ uniformly in $v\in\Psi$. ■

Lemma C.14. Suppose that Assumptions A.1, A.3(iii)–(iv) and A.4–A.5 hold. Then on any compact set Ψ,

$$Q^*_{g,NT}(v) \Rightarrow -\pi_g^{2\alpha} w_{g,D}\left|v\right| + 2\sqrt{\pi_g^{2\alpha}w_{g,V}}W_g(v) = \frac{w_{g,V}}{w_{g,D}}\left(-\frac{\pi_g^{2\alpha}w^2_{g,D}}{w_{g,V}}\left|v\right| + 2W_g(\frac{\pi_g^{2\alpha}w^2_{g,D}}{w_{g,V}}v)\right),$$

where $w_{g,V} = C_g^{0\prime}V_g^0C_g^0$.

Proof of Lemma C.14. Let $X_i(\gamma_g^0 + v/\alpha_{N_gT},\gamma_g^0) = [x_{i1}[d_{i1}(\gamma_g^0 + v/\alpha_{N_gT}) - d_{i1}(\gamma_g^0)],...,x_{iT}[d_{iT}(\gamma_g^0 + v/\alpha_{N_gT}) - d_{iT}(\gamma_g^0)]]'$. We have

$$\begin{aligned}
Q^*_{g,NT}(v) &= \check{Q}_g(\check{\theta}_g,\gamma_g^0) - \check{Q}_g(\check{\theta}_g,\gamma_g^0 + v/\alpha_{N_gT}) \\
&= -\sum_{i\in\mathbf{G}_g^0}\delta_g^{0\prime}X_i(\gamma_g^0 + v/\alpha_{N_gT},\gamma_g^0)'X_i(\gamma_g^0 + v/\alpha_{N_gT},\gamma_g^0)\delta_g^0 + 2\sum_{i\in\mathbf{G}_g^0}\delta_g^{0\prime}X_i(\gamma_g^0 + v/\alpha_{N_gT},\gamma_g^0)'\varepsilon_i \\
&\quad + L_{g,NT}(v),
\end{aligned}$$

where

$$\begin{aligned}
L_{g,NT}(v) &= 2(N_gT)^\alpha(\check{\delta}_g - \delta_g^0)'R^*_{g,NT}(v) - 2(N_gT)^\alpha\check{\delta}_g'K^*_{g,NT}(v)(N_gT)^\alpha(\check{\beta}_g - \beta_g^0) \\
&\quad -(N_gT)^\alpha(\check{\delta}_g - \delta_g^0)'K^*_{g,NT}(v)(N_gT)^\alpha(\check{\delta}_g + \delta_g^0) + (N_gT)^\alpha\check{\delta}_g'\tilde{K}^*_{g,NT}(v)(N_gT)^\alpha\check{\delta}_g \\
&\quad +2(N_gT)^\alpha\check{\delta}_g'\tilde{J}^*_{g,NT}(v) - 2\check{\delta}_g'\sum_{i\in G_g^0}X_i(\gamma_g^0 + v/\alpha_{N_gT},\gamma_g^0)'\mathbb{M}_0Z_i(\gamma_g^0)(\check{\theta}_g - \theta_g^0) \\
&\equiv L_{1g,NT}(v) + ... + L_{6g,NT}(v).
\end{aligned}$$

By Lemma C.10, we have

$$\begin{aligned}
&\sum_{i\in\mathbf{G}_g^0}\delta_g^{0\prime}X_i(\gamma_g^0+v/\alpha_{N_gT},\gamma_g^0)'X_i(\gamma_g^0+v/\alpha_{N_gT},\gamma_g^0)\delta_g^0\\
=\;&(NT)^{-2\alpha}\sum_{i\in\mathbf{G}_g^0}C_g^{0\prime}X_i(\gamma_g^0+v/\alpha_{N_gT},\gamma_g^0)'X_i(\gamma_g^0+v/\alpha_{N_gT},\gamma_g^0)C_g^0\\
=\;&(\frac{N_g}{N})^{2\alpha}\frac{\alpha_{N_gT}}{N_gT}\sum_{i\in\mathbf{G}_g^0}C_g^{0\prime}X_i(\gamma_g^0+v/\alpha_{N_gT},\gamma_g^0)'X_i(\gamma_g^0+v/\alpha_{N_gT},\gamma_g^0)C_g^0\\
=\;&(\frac{N_g}{N})^{2\alpha}G_{g,NT}^*(v)\Rightarrow\pi_g^{2\alpha}w_{g,D}\left|v\right|.
\end{aligned}$$

By Lemma C.11, we have

$$\begin{aligned}
\sum_{i\in\mathbf{G}_g^0}\delta_g^{0\prime}X_i(\gamma_g^0+v/\alpha_{N_gT},\gamma_g^0)'\varepsilon_i &= (NT)^{-\alpha}\sum_{i\in\mathbf{G}_g^0}C_g^{0\prime}X_i(\gamma_g^0+v/\alpha_{N_gT},\gamma_g^0)'\varepsilon_i\\
&= (\frac{N_g}{N})^{\alpha}(N_gT)^{1/2-\alpha}C_g^{0\prime}\left[J_{g,NT}(\gamma_g^0+v/\alpha_{N_gT})-J_{g,NT}(\gamma_g^0)\right]\\
&= (\frac{N_g}{N})^{\alpha}R_{g,NT}(v)\\
&\Rightarrow \pi_gC_g^{0\prime}B_g(v)=\pi_g\sqrt{w_{g,V}}W_g(v).
\end{aligned}$$

By the fact that $(NT)^{\alpha}(\check{\theta}_g-\theta_g^0)=o_p\left(1\right)$, Assumption A.1(vi), and Lemma C.10, we have $L_{\ell g,NT}(v)=o_p(1)$ uniformly in v for $\ell=1,2,3,4$. By Lemma C.12 we have that $L_{5g,NT}(v)=o_p(1)$ uniformly in v. For $L_{6g,NT}(v)$, we have

$$\begin{aligned}
\left|L_{6g,NT}(v)\right| &\leq 2\left\{(N_gT)^{\alpha}\left\|\check{\delta}_g\right\|\right\}\left\{(N_gT)^{\alpha}\left\|\check{\theta}_g-\theta_g^0\right\|\right\}\frac{\alpha_{N_gT}}{N_gT}\left\|\sum_{i\in G_g^0}X_i(\gamma_g^0+v/\alpha_{N_gT},\gamma_g^0)'\mathbb{M}_0Z_i(\gamma_g^0)\right\|\\
&= O_p\left(1\right)o_p\left(1\right)O_p\left(1\right)=o_p\left(1\right)\text{ uniformly in } v\in\Psi
\end{aligned}$$

as we can follow the proofs of Lemmas C.10 and C.12 and show that $\frac{\alpha_{N_gT}}{N_gT}||\sum_{i\in G_g^0}X_i(\gamma_g^0+v/\alpha_{N_gT},\gamma_g^0)'\mathbb{M}_0Z_i(\gamma_g^0)||$ $=O_p\left(1\right)$ uniformly in $v\in\Psi$. Consequently, we have $Q_{g,NT}^*(v)\Rightarrow-\pi_g^{2\alpha}w_{g,D}\left|v\right|+2\sqrt{\pi_g^{2\alpha}w_{g,V}}W_g(v)$ on any compact set Ψ. ■

D Determination of the Number of Groups

Recall that $\hat{\sigma}^2(G)\equiv\frac{1}{NT}\mathcal{Q}(\hat{\Theta}^{(G)},\hat{\mathbf{D}}^{(G)},\hat{\mathbf{G}}^{(G)})$. Let $\bar{\sigma}_{NT}^2\equiv\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\varepsilon_{it}^2$. In the estimation, we require each group to contain at least $\lfloor\nu N\rfloor$ individuals. We denote the index set of members in group g as $\mathbf{G}_g$, where $\mathbf{G}_g\in\mathbb{G}_\nu=\{\mathbf{G}_{\tilde{g}},\,|\mathbf{G}_{\tilde{g}}|>\lfloor\nu N\rfloor\}$ for all $g\in\mathcal{G}$. Let $\hat{N}_g=\left|\mathbf{G}_g\right|$. We can define five empirical processes that depend on $\mathbf{G}_g$:

$$\begin{aligned}
J(\mathbf{G}_g,\gamma) &= \frac{1}{\hat{N}_gT}\sum_{i\in\mathbf{G}_g}Z_i(\gamma)'\mathbb{M}_0\varepsilon_i,\quad \Delta J(\mathbf{G}_g,\gamma,\gamma^*)=\frac{1}{\hat{N}_gT}\sum_{i\in\mathbf{G}_g}X_i(\gamma,\gamma^*)'\mathbb{M}_0\varepsilon_i,\\
\Phi_1(\mathbf{G}_g,\gamma) &= \frac{1}{\hat{N}_gT}\sum_{i\in\mathbf{G}_g}Z_i(\gamma)'\mathbb{M}_0Z_i(\gamma),\ \Phi_2(\mathbf{G}_g,\gamma,\gamma^*)=\frac{1}{\hat{N}_gT}\sum_{i\in\mathbf{G}_g}Z_i(\gamma)'\mathbb{M}_0X_i(\gamma,\gamma^*),\text{ and}\\
\Phi_3(\mathbf{G}_g,\gamma,\gamma^*) &= \frac{1}{\hat{N}_gT}\sum_{i\in\mathbf{G}_g}X_i(\gamma,\gamma^*)'\mathbb{M}_0X_i(\gamma,\gamma^*).
\end{aligned}$$

Let $\mathbf{G}^G$ be any possible group structure when the number of groups in $\{1, 2, ..., N\}$ is given by G. We assume the following conditions hold for the empirical processes.

Assumption D.1. (i) $\Pr\left(\inf_{(\mathbf{G}_g,\gamma)\in\mathbb{G}_\nu\times\Gamma} \lambda_{\min}\left[\Phi_1(\mathbf{G}_g,\gamma)\right] \geq c\right) \to 1$ as $(N,T)\to\infty$ for some $c>0$;

(ii) $\Pr\left(\inf_{\mathbf{G}\in\mathbb{G}_\nu} \inf_{|\gamma-\gamma^*|>\bar{v}/T} \lambda_{\min}\left[\Phi_3(\mathbf{G}_g,\gamma,\gamma^*) - \Phi_2(\mathbf{G}_g,\gamma,\gamma^*)'\Phi_1(\mathbf{G}_g,\gamma)^{-1}\Phi_2(\mathbf{G}_g,\gamma,\gamma^*)\right] / |\gamma-\gamma^*| \geq c\right) \to 1$ as $(N,T)\to\infty$ for some $c>0$ and $\bar{v}>0$;

(iii) $\Pr\left(\sup_{\mathbf{G}\in\mathbb{G}_\nu} \sup_{|\gamma-\gamma^*|>\bar{v}/T} \|\Phi_\ell(\mathbf{G}_g,\gamma,\gamma^*)\| / |\gamma-\gamma^*| \leq C\right) \to 1$ for $\ell = 2,3$ as $(N,T)\to\infty$ for some $C>0$;

(iv) $\Pr\left(\sup_{(\mathbf{G}_g,\gamma)\in\mathbb{G}_\nu\times\Gamma} \|J(\mathbf{G}_g,\gamma)\| \leq CT^{-1/2}\right) \to 1$ for some $C>0$;

(v) $\Pr\left(\sup_{\mathbf{G}_g\in\mathbb{G}_\nu} \sup_{|\gamma-\gamma^*|>\bar{v}/T} \|\Delta J(\mathbf{G}_g,\gamma,\gamma^*)\| / |\gamma-\gamma^*| \leq CT^{-1/2}\right) \to 1$ for some $C>0$ and $\bar{v}>0$.

Assumption D.2. (i) As $(N,T)\to\infty$, $\min_{1\leq G<G^0} \min_{\mathbf{G}^G} \hat{\sigma}^2_{\mathbf{G}^G} \xrightarrow{P} \bar{\sigma}^2 > \sigma^2$, where $\sigma^2 \equiv \lim_{(N,T)\to\infty}(NT)^{-1} \sum_{i=1}^N\sum_{t=1}^T E\left(\varepsilon_{it}^2\right)$.

(ii) $\lambda_{NT}\to 0$ and $T\lambda_{NT}\to\infty$ as $(N,T)\to\infty$.

Assumption D.1(i)-(iii) requires the sample covariance matrices are well behaved for any subset of individuals. Assumption D.1(iv) is the assumption that plays the most important role in our analysis. It requires $\sup_{(\mathbf{G}_g,\gamma)\in\mathbb{G}_\nu\times\Gamma} \|J(\mathbf{G}_g,\gamma)\| = O_p(T^{-1/2})$ for all $(\mathbf{G}_g,\gamma)\in\mathbb{G}_\nu\times\Gamma$. For the true group members $\mathbf{G}_g^0$, we can show that $J(\mathbf{G}_g^0,\gamma) = O_p((NT)^{-1/2})$ under some regularity conditions. However when we are estimating the model with $G>G^0$, it is possible that $\left\|J(\hat{\mathbf{G}}_g,\gamma)\right\| = O_p(T^{-1/2})$. Similar remarks hold for D.1(v). Assumption D.2 specifies the usual condition for the consistency of an information criterion. In particular, Assumption D.2(i) in conjunction with the first part of D.2(ii) helps to eliminate all underfitted models and the second part of D.2(ii) helps to eliminate the overfitted models.

Proposition D.1 *Suppose Assumptions A.1-A.5 in the text and Assumption D.1 hold. The following statement holds:*

$$\hat{\sigma}^2(G) - \bar{\sigma}^2_{NT} = O_p(T^{-1}) \text{ for any } G^0 \leq G \leq G_{\max}.$$

Remark. The probability order $O_p(T^{-1})$ in the above proposition is not a conservative order. To illustrate this point, we consider a simple regression where $y_{it} = \mu + \varepsilon_{it}$ so that there is only one group. If we estimate the model with $G=2$, we have

$$\begin{aligned} 0 \geq T\left[\hat{\sigma}^2(2) - \bar{\sigma}^2_{NT}\right] &= \frac{1}{N}\sum_{g=1}^2\sum_{i\in\hat{\mathbf{G}}_g}\sum_{t=1}^T\left(y_{it} - \frac{1}{\hat{N}_gT}\sum_{i\in\hat{\mathbf{G}}_g}\sum_{t=1}^T y_{it}\right)^2 - \frac{1}{N}\sum_{i=1}^N\sum_{t=1}^T\varepsilon_{it}^2 \\ &= \frac{1}{N}\sum_{g=1}^2\sum_{i\in\hat{\mathbf{G}}_g}\sum_{t=1}^T\left(\varepsilon_{it} - \frac{1}{\hat{N}_gT}\sum_{i\in\hat{\mathbf{G}}_g}\sum_{t=1}^T \varepsilon_{it}\right)^2 - \frac{1}{N}\sum_{i=1}^N\sum_{t=1}^T\varepsilon_{it}^2 \\ &= -T\sum_{g=1}^2\frac{\hat{N}_g}{N}\left(\frac{1}{\hat{N}_g}\sum_{i\in\hat{\mathbf{G}}_g}\bar{\varepsilon}_{i\cdot}\right)^2 \leq -\sum_{g=1}^2\frac{\tilde{N}_g}{N}\left(\frac{\sqrt{T}}{\tilde{N}_g}\sum_{i\in\tilde{\mathbf{G}}_g}\bar{\varepsilon}_{i\cdot}\right)^2, \end{aligned}$$

where $\bar{\varepsilon}_{i\cdot} = \frac{1}{T}\sum_{t=1}^T\varepsilon_{it}$, $\tilde{\mathbf{G}}_1 = \{i|\bar{\varepsilon}_{i\cdot}\leq 0\}$, $\tilde{\mathbf{G}}_2 = \{i|\bar{\varepsilon}_{i\cdot}>0\}$, $\tilde{N}_g = \left|\tilde{\mathbf{G}}_g\right|$, and the last inequality holds by the definitions of $\{\hat{\mathbf{G}}_g\}$ and $\hat{\sigma}^2(2)$. [Note that $\hat{\sigma}^2(2)$ is minimized at $(\hat{\mathbf{G}}_1, \hat{\mathbf{G}}_2)$.] Without loss of generality, we suppose that ε_{it} is i.i.d. $N(0,1)$ over both i and t. Then $v_{iT} \equiv \sqrt{T}\bar{\varepsilon}_{i\cdot} \backsim N(0,1)$ and by the strong law of large numbers

$$\left|\frac{\sqrt{T}}{\tilde{N}_1}\sum_{i\in\tilde{\mathbf{G}}_1}\bar{\varepsilon}_{i\cdot}\right| = \left|\frac{\sqrt{T}}{\tilde{N}_1}\sum_{i=1}^N\bar{\varepsilon}_{i\cdot}\mathbf{1}\left(\bar{\varepsilon}_{i\cdot}\leq 0\right)\right| = \frac{N}{\tilde{N}_1}\left|\frac{1}{N}\sum_{i=1}^N v_{iT}\mathbf{1}\left(v_{iT}\leq 0\right)\right| \xrightarrow{a.s.} 2\left|E\left[Z\mathbf{1}\left(Z\leq 0\right)\right]\right|,$$

where $Z \backsim N(0,1)$ and we use the fact that $\tilde{N}_1/N = \frac{1}{N}\sum_{i=1}^{N}\mathbf{1}(\bar{\varepsilon}_{i\cdot} \leq 0) \overset{a.s.}{\to} P(Z \leq 0) = \frac{1}{2}$. Similarly,

$$\left|\frac{\sqrt{T}}{\tilde{N}_2}\sum_{i\in\tilde{\mathbf{G}}_2}\bar{\varepsilon}_{i\cdot}\right| = \left|\frac{\sqrt{T}}{\tilde{N}_2}\sum_{i=1}^{N}\bar{\varepsilon}_{i\cdot}\mathbf{1}(\bar{\varepsilon}_{i\cdot} > 0)\right| = \frac{N}{\tilde{N}_2}\left|\frac{1}{N}\sum_{i=1}^{N}\upsilon_{iT}\mathbf{1}(\upsilon_{iT} > 0)\right| \overset{a.s.}{\to} 2\left|E\left[Z\mathbf{1}(Z>0)\right]\right|.$$

This calculation indicates that the negative value $T\left[\hat{\sigma}^2(2) - \bar{\sigma}^2_{NT}\right]$ has the probability order $O_p\left(T^{-1}\right)$ that cannot be $o_p\left(T^{-1}\right)$. In other words, the order $O_p(T^{-1})$ is a tight probability order for $\hat{\sigma}^2(2) - \bar{\sigma}^2_{NT}$.

Proof of Proposition D.1. Following similar arguments as used in the proofs of Lemmas A.1-A.3, we can show that individuals from the true group $\mathbf{G}^0_g$ would stay in the same estimated group w.p.a.1, i.e.,

$$\Pr\left[\sup_{1\leq i<j\leq N}\mathbf{1}(\hat{g}_i = \hat{g}_j,\ g^0_i \neq g^0_j) = 1\right] \to 0 \text{ as } (N,T)\to\infty.$$

We only consider the case where some true groups are further divided into several groups. For notational simplicity, we only consider the case $G^0 = 1$ where our true parameters can be rewritten as $(\theta^{0\prime},\gamma^0)' = (\beta^{0\prime},\delta^{0\prime},\gamma^0)'$ without the group-specific subscript. Since we still estimate a PSTR model with $G \geq 1$ groups, the estimators, e.g., $(\hat{\theta}^{(G)}_g, \hat{\gamma}^{(G)}_g)$, still have the group-specific subscript. But for notational brevity, we will denote $(\hat{\theta}^{(G)}_g, \hat{\gamma}^{(G)}_g)$ as $(\hat{\theta}_g, \hat{\gamma}_g)$. Then we can write $\mathcal{Q}(\hat{\Theta}^{(G)}, \hat{\mathbf{D}}^{(G)}, \hat{\mathbf{G}}^{(G)}) = \sum_{g=1}^{G}\bar{\mathcal{Q}}_g(\hat{\theta}_g, \hat{\gamma}_g)$, where $\bar{\mathcal{Q}}_g(\cdot,\cdot)$ is defined in Section 4.1. Following the analysis for (C.2) in the proof of Lemma C.2, we have

$$\hat{\theta}_g - \theta^0 = \bar{\Phi}_{1,g}(\hat{\gamma}_g)^{-1}\frac{1}{\hat{N}_g T}\sum_{i\in\hat{\mathbf{G}}_g} Z_i(\hat{\gamma}_g)'\mathbb{M}_0\varepsilon_i - \bar{\Phi}_{1,g}(\hat{\gamma}_g)^{-1}\bar{\Phi}_{2,g}(\hat{\gamma}_g)\delta^0, \tag{D.1}$$

where $\bar{\Phi}_{1,g}(\gamma) \equiv \frac{1}{\hat{N}_g T}\sum_{i\in\hat{\mathbf{G}}_g} Z_i(\gamma)'\mathbb{M}_0 Z_i(\gamma) = \Phi_1(\hat{\mathbf{G}}_g,\gamma)$ and $\bar{\Phi}_{2,g}(\gamma) \equiv \frac{1}{\hat{N}_g T}\sum_{i\in\hat{\mathbf{G}}_g} Z_i(\gamma)'\mathbb{M}_0 X_i(\gamma,\gamma^0) = \Phi_2(\hat{\mathbf{G}}_g,\gamma,\gamma^0)$. Following the similar analysis for (C.3) of Lemma C.2, we have

$$\begin{aligned}
& \frac{1}{\hat{N}_g T}\left[\bar{\mathcal{Q}}_g(\hat{\theta}_g,\hat{\gamma}_g) - \bar{\mathcal{Q}}_g(\theta^0,\gamma^0)\right] \\
= \;& \delta^{0\prime}\bar{\Phi}_{3g}(\hat{\gamma}_g)\delta^0 + (\hat{\theta}_g - \theta^0)'\bar{\Phi}_{1g}(\hat{\gamma}_g)(\hat{\theta}_g - \theta^0) + 2(\hat{\theta}_g - \theta^0)'\bar{\Phi}_{2g}(\hat{\gamma}_g)\delta^0 \\
& -(\hat{\theta}_g - \theta^0)'\frac{2}{N_g T}\sum_{i\in\hat{\mathbf{G}}_g} Z_i(\hat{\gamma}_g)'\mathbb{M}_0\varepsilon_i - \hat{\delta}_g'\frac{2}{N_g T}\sum_{i\in\hat{\mathbf{G}}_g} X_i(\hat{\gamma}_g,\gamma^0)'\mathbb{M}_0\varepsilon_i,
\end{aligned}$$

where $\bar{\Phi}_{3g}(\gamma) \equiv \frac{1}{\hat{N}_g T}\sum_{i\in\hat{\mathbf{G}}_g} X_i(\gamma,\gamma^0)'\mathbb{M}_0 X_i(\gamma,\gamma^0) = \Phi_3(\hat{\mathbf{G}}_g,\gamma,\gamma^0)$. Plugging (D.1) into the above equation, we have

$$\begin{aligned}
\frac{1}{\hat{N}_g T}\left[\bar{\mathcal{Q}}_g(\hat{\theta}_g,\hat{\gamma}_g) - \bar{\mathcal{Q}}_g(\theta^0,\gamma^0)\right] \;=\;& \delta^{0\prime}\left[\bar{\Phi}_{3g}(\hat{\gamma}_g) - \bar{\Phi}_{2g}(\hat{\gamma}_g)'\bar{\Phi}_{1g}(\hat{\gamma}_g)^{-1}\bar{\Phi}_{2g}(\hat{\gamma}_g)\right]\delta^0 \\
& -J(\hat{\mathbf{G}}_g,\hat{\gamma}_g)'\bar{\Phi}_{1g}(\hat{\gamma}_g)^{-1}J(\hat{\mathbf{G}}_g,\hat{\gamma}_g) + 2\delta^{0\prime}\bar{\Phi}_{2g}(\hat{\gamma}_g)'\bar{\Phi}_{1g}(\hat{\gamma}_g)^{-1}J(\hat{\mathbf{G}}_g,\hat{\gamma}_g) \\
& -2\left(\hat{\delta}_g - \delta^0\right)'\Delta J(\hat{\mathbf{G}}_g,\hat{\gamma}_g,\gamma^0) - 2\delta^{0\prime}\Delta J(\hat{\mathbf{G}}_g,\hat{\gamma}_g,\gamma^0) \\
\equiv\;& \Delta\bar{\mathcal{Q}}_{1,g} + \ldots + \Delta\bar{\mathcal{Q}}_{5,g}.
\end{aligned}$$

We discuss two cases: (1) $(NT)^{-\alpha}T^{1/2} = O(1)$ and (2) $(NT)^{-\alpha}T^{1/2} \to \infty$ as $(N,T)\to\infty$.

In Case (1), we have $\delta^0 = O(T^{-1/2})$. By Assumption D.1(iii) and equation (D.1), we can readily show that $\hat{\theta}_g - \theta^0 = O_p(T^{-1/2})$. With this result, then we can show that $\Delta\bar{\mathcal{Q}}_{l,g} = O_p(T^{-1})$ for $l = 1,\ldots,5$ by using Assumption D.1.

In Case (2), we have $\hat{\theta}_g - \theta^0 = O_p(T^{-1/2} + (NT)^{-\alpha} |\hat{\gamma}_g - \gamma^0|)$ by (D.1) and Assumption D.1(iii). Then we can apply Assumption D.1 to show that

$$\begin{aligned}
\Delta\bar{\mathcal{Q}}_{1,g} &= O_p((NT)^{-2\alpha} |\hat{\gamma}_g - \gamma^0|), \ \Delta\bar{\mathcal{Q}}_{2,g} = O_p(T^{-1}), \\
\Delta\bar{\mathcal{Q}}_{3,g} &= O_p(T^{-1/2}(NT)^{-\alpha} |\hat{\gamma}_g - \gamma^0|), \\
\Delta\bar{\mathcal{Q}}_{4,g} &= O_p(T^{-1/2}(NT)^{-\alpha} |\hat{\gamma}_g - \gamma^0|^2 + T^{-1} |\hat{\gamma}_g - \gamma^0|), \text{ and} \\
\Delta\bar{\mathcal{Q}}_{5,g} &= O_p(T^{-1/2}(NT)^{-\alpha} |\hat{\gamma}_g - \gamma^0|.
\end{aligned}$$

Because $\Delta\bar{\mathcal{Q}}_{1,g} > 0$ by Assumption D.1(ii) and $\frac{1}{\hat{N}_g T}\left[\bar{\mathcal{Q}}_g(\hat{\theta}_g, \hat{\gamma}_g) - \bar{\mathcal{Q}}_g(\theta^0, \gamma^0)\right] < 0$ by the definition of least squares estimation, we can conclude $\Delta\bar{\mathcal{Q}}_{1,g}$ should have at most the same order as $\sum_{l=2}^5 \Delta\bar{\mathcal{Q}}_{l,g}$. By comparison between these orders, we can show that $|\hat{\gamma}_g - \gamma^0| = O_p(T^{-1}(NT)^{2\alpha})$ and $\sum_{l=1}^5 \Delta\bar{\mathcal{Q}}_{l,g} = O_p(T^{-1})$ follows. Consequently,

$$\begin{aligned}
0 &\geq \hat{\sigma}^2(G) - \bar{\sigma}^2_{NT} = \frac{1}{NT}\sum_{g=1}^G \left[\bar{\mathcal{Q}}_g(\hat{\theta}_g, \hat{\gamma}_g) - \bar{\mathcal{Q}}_g(\theta^0, \gamma^0)\right] \\
&= \sum_{g=1}^G \frac{\hat{N}_g}{N} \frac{1}{\hat{N}_g T}\left[\bar{\mathcal{Q}}_g(\hat{\theta}_g, \hat{\gamma}_g) - \bar{\mathcal{Q}}_g(\theta^0, \gamma^0)\right] \\
&\geq \sum_{g=1}^G \frac{1}{\hat{N}_g T}\left[\bar{\mathcal{Q}}_g(\hat{\theta}_g, \hat{\gamma}_g) - \bar{\mathcal{Q}}_g(\theta^0, \gamma^0)\right] = O_p(T^{-1}).
\end{aligned}$$

This implies that $\hat{\sigma}^2(G) - \bar{\sigma}^2_{NT} = O_p(T^{-1})$ for any $G^0 \leq G \leq G_{\max}$. ■

E Consistency of group membership estimators in the fixed-threshold-effect framework

In this section, we discuss the asymptotic property of our least squares estimator under the constant threshold effect framework (i.e., $\alpha = 0$). Suppose Assumptions A.1-A.5 hold except that we now let $\alpha = 0$. Then one can follow the arguments as used in the proofs of Lemmas C.2-C.10 to show that $|\check{\gamma}_g - \gamma^0_g| = O_p((NT)^{-1})$ and $\check{\theta}_g = \check{\theta}(\gamma^0_g) + o_p((NT)^{-1/2})$, where $(\check{\theta}_g, \check{\gamma}_g)$ is infeasible estimator for $g \in \mathcal{G}$.

In the PSTR model, the major difficulty is to show the consistency of the estimator of the latent group structure as in Theorem 3.1. Once we establish a similar result as that of Lemma A.3, we can prove Theorem 3.1. In addition, we can prove Lemmas A.4-A.6 which confirms $|\hat{\gamma}_g - \gamma^0_g| = O_p((NT)^{-1})$ and $\hat{\theta}_g = \check{\theta}_g + o_p((NT)^{-1/2})$. In the following analysis, we give a sketch of the proof of Theorem 3.1 in the fixed-threshold-effect framework.

To proceed, we add some notations. Define

$$\tilde{M}_{NT}(g, \tilde{g}, \mathbf{G}) \equiv \frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T \mathbf{1}(g^0_i = g)\mathbf{1}(g_i = \tilde{g})E\left[(x'_{it}\delta^0_g)^2|\gamma^0_g\right],$$

where $E(\cdot|\gamma) \equiv E(\cdot|q_{it} = \gamma)$. We impose an additional identification condition:

Assumption E.1. As $(N, T) \to \infty$, the following statements hold: (i) For some constants $c > 0$ and $\bar{v} > 0$, we have

$$\sup_{1\leq i\leq N}\ \sup_{(\theta\prime,\theta^{*\prime})^{\prime}\in\mathcal{B}^2}\ \sup_{|\gamma-\gamma^*|>\bar{v}/T}\left\{\Pr\left[\sum_{t=1}^T \left[\tilde{z}_{it}(\gamma)'\theta - \tilde{z}_{it}(\gamma^*)'\theta^*\right]^2 \leq c\sum_{t=1}^T \left[(\theta-\theta^*)'\tilde{z}_{it}(\gamma^*)\right]^2 + |\gamma-\gamma^*|\, E\left[(x'_{it}\delta)^2|\gamma\right]\right]\right\} = o(T^{-4});$$

(ii) There exists a constant $\underline{c}_\lambda > 0$ such that for all $g \in \mathcal{G}$,

$$\Pr\left(\inf_{(\mathbf{G},\mathbf{D})\in\mathcal{G}^N\times\Gamma^G}\max_{\tilde{g}\in\mathcal{G}}\{\lambda_{\min}[M_{NT}(g,\tilde{g},\mathbf{D},\mathbf{G})]\wedge\tilde{M}_{NT}(g,\tilde{g},\mathbf{G})\} > \underline{c}_\lambda\right)\to 1;$$

(iii) For all $g, \tilde{g} \in \mathcal{G}$, where $g \neq \tilde{g}$, we have $\left\|(\theta_g^{0\prime},\gamma_g^0)' - (\theta_{\tilde{g}}^{0\prime},\gamma_{\tilde{g}}^0)'\right\| > \underline{c}_\beta$ for some constant $\underline{c}_\beta > 0$;
(iv) For any $g \neq \tilde{g}$ and $1 \le i \le N$, we have

$$\max\left(E[\tilde{z}_{it}(\gamma_g^0)'(\theta_{\tilde{g}}^0-\theta_g^0)]^2,\ \left|\gamma_{\tilde{g}}^0-\gamma_g^0\right| E\left[(x_{it}'\delta_{\tilde{g}}^0)^2|\gamma_{\tilde{g}}^0\right]\right)\equiv\underline{\tilde{c}}_{g\tilde{g},i}\ge\underline{\tilde{c}}_{g\tilde{g}},$$

for some constant $\underline{\tilde{c}}_{g\tilde{g}} > 0$.

Assumption E.1 (i) is a non-colinearity condition similar to Assumption A.4(ii) in the main text. However, it requires that the non-colinearity property should hold for each individual. Assumption E.1(ii) is modified from Assumption A.2. Assumption E.1(iii)-(iv) is modified from Assumption A.3(i)-(ii). As remarked in Section 3.1, E.1(iv) is redundant if we assume that $\lambda_{\min}(E[\tilde{z}_{it}(\gamma_g^0)\tilde{z}_{it}(\gamma_g^0)'])$ and $\delta_g^{0\prime}E(x_{it}x_{it}'|\gamma_g^0)\delta_g^0$ are bounded below from zero by a constant $\underline{c}$, say.

Below we prove Theorem 3.1 under Assumptions A.1, A.3(iii)-(iv) and E.1.

Proof of Theorem 3.1. Lemma A.1 still holds under the stated conditions. Lemmas A.2-A.3 are replaced by Lemmas E.1 and E.2 below. Combining Lemmas E.1-E.2 we have the desired claim. ■

Lemma E.1. Suppose that Assumptions A.1, A.3(iii)-(iv) and E.1 hold. Then we have $d_H((\hat{\Theta},\hat{\mathbf{D}}),(\Theta^0,\mathbf{D}^0))\xrightarrow{p}0$, where

$$d_H((\hat{\Theta},\hat{\mathbf{D}}),(\Theta^0,\mathbf{D}^0))=\max\left\{\max_{g\in\mathcal{G}}\left(\min_{\tilde{g}\in\mathcal{G}}\left\|\hat{\theta}_g-\theta_{\tilde{g}}^0\right\|^2+\left|\hat{\gamma}_g-\gamma_{\tilde{g}}^0\right|\right),\ \max_{\tilde{g}\in\mathcal{G}}\left(\min_{g\in\mathcal{G}}\left\|\hat{\theta}_g-\theta_{\tilde{g}}^0\right\|^2+\left|\hat{\gamma}_g-\gamma_{\tilde{g}}^0\right|\right)\right\}.$$

Proof of Lemma E.1. It suffices to show (i) $\max_{g\in\mathcal{G}}\left(\min_{\tilde{g}\in\mathcal{G}}\left\|\hat{\theta}_g-\theta_{\tilde{g}}^0\right\|+\left|\hat{\gamma}_g-\gamma_{\tilde{g}}^0\right|\right)=o_p(1)$ and (ii) $\max_{\tilde{g}\in\mathcal{G}}\left(\min_{g\in\mathcal{G}}\left\|\hat{\theta}_g-\theta_{\tilde{g}}^0\right\|+\left|\hat{\gamma}_g-\gamma_{\tilde{g}}^0\right|\right)=o_p(1)$.

We first show (i). By Lemma A.1, we have

$$\begin{aligned}\frac{1}{NT}\tilde{\mathcal{Q}}(\hat{\Theta},\hat{\mathbf{D}},\hat{\mathbf{G}}) &= \frac{1}{NT}\mathcal{Q}(\hat{\Theta},\hat{\mathbf{D}},\hat{\mathbf{G}})+o_p(1)\le\frac{1}{NT}\mathcal{Q}(\Theta^0,\mathbf{D}^0,\mathbf{G}^0)+o_p(1)\\ &= \frac{1}{NT}\tilde{\mathcal{Q}}(\Theta^0,\mathbf{D}^0,\mathbf{G}^0)+o_p(1),\end{aligned}$$

where the inequality holds by the definition of least squares estimator. On the other hand, noting that $\tilde{\mathcal{Q}}(\Theta,\mathbf{D},\mathbf{G})$ is minimized at $(\Theta^0,\mathbf{D}^0,\mathbf{G}^0)$, we have $\frac{1}{NT}[\tilde{\mathcal{Q}}(\hat{\Theta},\hat{\mathbf{D}},\hat{\mathbf{G}})-\tilde{\mathcal{Q}}(\Theta^0,\mathbf{D}^0,\mathbf{G}^0)]\ge 0$. It follows that $\frac{1}{NT}[\tilde{\mathcal{Q}}(\hat{\Theta},\hat{\mathbf{D}},\hat{\mathbf{G}})-\tilde{\mathcal{Q}}(\Theta^0,\mathbf{D}^0,\mathbf{G}^0)]=o_p(1)$. By direct calculation, we have uniformly in $(\Theta,\mathbf{D},\mathbf{G})$,

$$\begin{aligned}&\frac{1}{NT}\left[\tilde{\mathcal{Q}}(\Theta,\mathbf{D},\mathbf{G})-\tilde{\mathcal{Q}}(\Theta^0,\mathbf{D}^0,\mathbf{G}_g^0)\right]\\ =\ &\frac{1}{NT}\sum_{i=1}^N\sum_{t=1}^T\left\{\theta_{g_i^0}^{0\prime}\tilde{z}_{it}(\gamma_{g_i^0}^0)-\theta_{g_i}'\tilde{z}_{it}(\gamma_{g_i})\right\}^2\\ \ge\ &\frac{c}{NT}\sum_{i=1}^N\sum_{t=1}^T\left[(\theta_{g_i}-\theta_{g_i^0}^0)'\tilde{z}_{it}(\gamma_{g_i})\right]^2+\frac{c}{NT}\sum_{i=1}^N\sum_{t=1}^T\left|\gamma_{g_i}-\gamma_{g_i^0}^0\right|E\left[(x_{it}'\delta_{g_i^0}^0)^2|\gamma_{g_i^0}^0\right]+o_p(1)\\ =\ &\sum_{g=1}^G\sum_{\tilde{g}=1}^G\frac{c}{NT}\sum_{i=1}^N\sum_{t=1}^T\mathbf{1}(g_i^0=g)\mathbf{1}(g_i=\tilde{g})\left\{\left[(\theta_g^0-\theta_{\tilde{g}})'\tilde{z}_{it}(\gamma_{\tilde{g}})\right]^2+\left|\gamma_{\tilde{g}}-\gamma_g^0\right|E\left[(x_{it}'\delta_g^0)^2|\gamma_g^0\right]\right\}+o_p(1)\\ =\ &c\sum_{g=1}^G\sum_{\tilde{g}=1}^G\left[(\theta_g^0-\theta_{\tilde{g}})'M_{NT}(g,\tilde{g},\mathbf{D},\mathbf{G})(\theta_g^0-\theta_{\tilde{g}})+\left|\gamma_{\tilde{g}}-\gamma_g^0\right|\tilde{M}_{NT}(g,\tilde{g},\mathbf{G})\right]+o_p(1),\end{aligned}$$

where the inequality holds by Assumption E.1(i) and the last equation is by the definitions of $M_{NT}(g,\tilde{g},\mathbf{D},\mathbf{G})$ and $\tilde{M}_{NT}(g,\tilde{g},\mathbf{G})$. It follows that

$$\begin{aligned}
o_p(1) &= c\sum_{g=1}^{G}\sum_{\tilde{g}=1}^{G}\left[(\theta_g^0-\theta_{\tilde{g}})'M_{NT}(g,\tilde{g},\mathbf{D},\mathbf{G})(\theta_g^0-\theta_{\tilde{g}})+\left|\gamma_{\tilde{g}}-\gamma_g^0\right|\tilde{M}_{NT}(g,\tilde{g},\mathbf{G})\right]+o_p(1)\\
&\geq c\sum_{g=1}^{G}\sum_{\tilde{g}=1}^{G}\left\{\lambda_{\min}[M_{NT}(g,\tilde{g},\mathbf{D},\mathbf{G})]\wedge\tilde{M}_{NT}(g,\tilde{g},\mathbf{G})\right\}\left(\left\|\theta_g^0-\theta_{\tilde{g}}\right\|^2+\left|\gamma_{\tilde{g}}-\gamma_g^0\right|\right)+o_p(1)\\
&\geq c\max_{g\in\mathcal{G}}\sum_{\tilde{g}=1}^{G}\left\{\lambda_{\min}[M_{NT}(g,\tilde{g},\mathbf{D},\mathbf{G})]\wedge\tilde{M}_{NT}(g,\tilde{g},\mathbf{G})\right\}\left(\left\|\theta_g^0-\theta_{\tilde{g}}\right\|^2+\left|\gamma_{\tilde{g}}-\gamma_g^0\right|\right)+o_p(1)\\
&\geq c\max_{g\in\mathcal{G}}\left(\min_{\tilde{g}\in\mathcal{G}}\left\|\theta_g^0-\hat{\theta}_{\tilde{g}}\right\|^2+\left|\gamma_{\tilde{g}}-\gamma_g^0\right|\right)\sum_{\tilde{g}=1}^{G}\left\{\lambda_{\min}[M_{NT}(g,\tilde{g},\mathbf{D},\mathbf{G})]\wedge\tilde{M}_{NT}(g,\tilde{g},\mathbf{G})\right\}+o_p(1)\\
&\geq c\underline{c}_\lambda\max_{g\in\mathcal{G}}\left(\min_{\tilde{g}\in\mathcal{G}}\left\|\theta_g^0-\hat{\theta}_{\tilde{g}}\right\|^2+\left|\gamma_{\tilde{g}}-\gamma_g^0\right|\right)+o_p(1),
\end{aligned}$$

where the last inequality is by Assumption E.1(ii) which says that there exists a group $\tilde{g}^*\in\mathcal{G}$ such that $\left\{\lambda_{\min}[M_{NT}(g,\tilde{g},\mathbf{D},\mathbf{G})]\wedge\tilde{M}_{NT}(g,\tilde{g},\mathbf{G})\right\}>\underline{c}_\lambda>0$ w.p.a.1. Consequently, we have

$$\max_{g\in\mathcal{G}}\left(\min_{\tilde{g}\in\mathcal{G}}\left\|\theta_g^0-\hat{\theta}_{\tilde{g}}\right\|^2+\left|\gamma_{\tilde{g}}-\gamma_g^0\right|\right)=o_p\left(1\right).$$

To show (ii), we can follow a similar analysis given in the proof of Lemma A.2. The details are omitted here. ■.

Remark. The proof of Lemma E.1 shows that there exists a permutation $\sigma_{\hat{\Theta}}$ such that $\left\|\hat{\theta}_g-\theta^0_{\sigma_{\hat{\Theta}}(g)}\right\|^2+\left|\hat{\gamma}_g-\gamma^0_{\sigma_{\hat{\Theta}}(g)}\right|=o_p(1)$. We can take $\sigma_{\hat{\Theta}}(g)=g$ by relabeling. In the following analysis, we shall write $\left\|\hat{\theta}_g-\theta^0_{\sigma_{\hat{\Theta}}(g)}\right\|^2+\left|\hat{\gamma}_g-\gamma^0_{\sigma_{\hat{\Theta}}(g)}\right|=o_p(1)$ without referring to the relabeling any more.

Lemma E.2. Let $\hat{g}_i(\Theta,\mathbf{D})=\operatorname{argmin}_{g\in\mathcal{G}}\sum_{t=1}^T\left[\tilde{y}_{it}-\tilde{z}_{it}(\gamma_g)'\theta_g\right]^2$. Suppose that Assumptions A.1, A.3(iii)-(iv) and E.1 hold. Then we have that for some $\eta>0$,

$$\Pr\left(\sup_{(\Theta,\mathbf{D})\in\tilde{\mathcal{N}}_\eta}\left[\frac{1}{N}\sum_{i=1}^N\mathbf{1}(\hat{g}_i(\Theta,\mathbf{D})\neq g_i^0)\right]\right)=o(T^{-4}),$$

where $\tilde{\mathcal{N}}_\eta=\left\{(\Theta,\mathbf{D})\in\mathcal{B}^G\times\Gamma^G:\left\|\theta_g-\theta_g^0\right\|^2+\left|\gamma_g-\gamma_g^0\right|<\eta,\ g\in\mathcal{G}\right\}$

Proof of Lemma E.2. The proof is similar to that of Lemma A.3 except the details of bounding $\mathcal{Z}_{ig}(\Theta,\mathbf{D})$, where

$$\mathcal{Z}_{ig}(\Theta,\mathbf{D})\equiv\mathbf{1}(g_i^0\neq g)\mathbf{1}\left(\sum_{t=1}^T[\tilde{y}_{it}-\tilde{z}_{it}(\gamma_g)'\theta_g)]^2\leq\sum_{t=1}^T[\tilde{y}_{it}-\tilde{z}_{it}(\gamma_{g_i^0})'\theta_{g_i^0}]^2\right).$$

For $\mathcal{Z}_{ig}(\Theta,\mathbf{D})$, we have

$$\mathcal{Z}_{ig}(\Theta,\mathbf{D})\leq\max_{\tilde{g}\in\mathcal{G}\setminus\{g\}}\mathbf{1}\left(L_i(g,\tilde{g})\leq 0\right),$$

where

$$L_i(g,\tilde{g})=\sum_{t=1}^T[\tilde{z}_{it}(\gamma_{\tilde{g}})'\theta_{\tilde{g}}-\tilde{z}_{it}(\gamma_g)'\theta_g]\left\{\frac{1}{2}[\tilde{z}_{it}(\gamma_{\tilde{g}})'\theta_{\tilde{g}}-\tilde{z}_{it}(\gamma_g)'\theta_g]+\tilde{\varepsilon}_{it}+\tilde{z}_{it}(\gamma^0_{\tilde{g}})'\theta^0_{\tilde{g}}-\tilde{z}_{it}(\gamma_{\tilde{g}})'\theta_{\tilde{g}}\right\}.$$

Then we can follow the analysis of Lemma A.3 to show that

$$\mathcal{Z}_{ig}(\Theta,\mathbf{D}) \leq \max_{\tilde{g}\in\boldsymbol{G}\backslash\{g\}} \mathbf{1}\left\{\sum_{t=1}^{T}[\tilde{z}_{it}(\gamma_{\tilde{g}}^{0})'\theta_{\tilde{g}}^{0}-\tilde{z}_{it}^{0}(\gamma_{g}^{0})'\theta_{g}^{0}]\left\{\frac{1}{2}[\tilde{z}_{it}(\gamma_{\tilde{g}}^{0})'\theta_{\tilde{g}}^{0}-\tilde{z}_{it}^{0}(\gamma_{g}^{0})'\theta_{g}^{0}]+\tilde{\varepsilon}_{it}\right\}\leq H_{iT}\right\}\equiv\tilde{\mathcal{Z}}_{ig},$$

where $H_{iT} = C\sqrt{\eta}\sum_{t=1}^{T}(\|x_{it}\|^2+\varepsilon_{it}^2)$ for some constant $C>0$. Next, we can use the Assumption E.1(i) to show that

$$\begin{aligned}\Pr(\tilde{\mathcal{Z}}_{ig} &= 1) \leq \sum_{\tilde{g}\in\mathcal{G}\backslash\{g\}}\Pr\left\{\frac{c}{2}\sum_{t=1}^{T}\left[(\theta_{\tilde{g}}^{0}-\theta_{g}^{0})'\tilde{z}_{it}(\gamma_{g}^{0})\right]^{2}+\left|\gamma_{\tilde{g}}^{0}-\gamma_{g}^{0}\right|E\left[(x_{it}'\delta_{\tilde{g}}^{0})^{2}|\gamma_{\tilde{g}}^{0}\right]\right.\\ &\left.+\sum_{t=1}^{T}[\tilde{z}_{it}(\gamma_{\tilde{g}}^{0})'\theta_{\tilde{g}}^{0}-\tilde{z}_{it}^{0}(\gamma_{g}^{0})'\theta_{g}^{0}]\tilde{\varepsilon}_{it}\leq H_{iT}\right\}+o(T^{-4}).\end{aligned}$$

Then one can use Assumption E.1(iv) and similar arguments as used in the proof of Lemma A.3 to show that the leading term on the right hand side of the last inequality is $o(T^{-4})$. The result then follows from the Markov inequality as used in the proof of Lemma A.3. ■.

REFERENCES

Davidson, J., 1994. Stochastic Limit Theory: An Introduction for Econometricians. Oxford University Press, Oxford.

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

Wang, W., Phillips, P.C. and Su, L., 2018. Homogeneity pursuit in panel data models: Theory and application. Journal of Applied Econometrics 33, 797-815.

Yang, S.C., 2006. Maximal moment inequality for partial sums of strong mixing sequences and application. Acta Mathematica Sinica 23, 1013-1024.