Panel Approaches to Econometric Analysis of Bubble Behaviour*

Yanbo Liu
Singapore Management University

November 5, 2019

Abstract

This study provides novel mechanisms for identifying explosive bubbles in mixed-roots panel autoregressions with latent group structures. Two post-classification panel data approaches are proposed to test the explosiveness in time-series data. The first approach, which is based on explosive panel autoregressions with latent group structures, applies recursive k-means clustering. The second approach, which is based on mixed-roots panel autoregressions with latent groups, employs modified k-means clustering. We establish the uniform consistency of the two k-means classification algorithms. The k-means procedures achieve the oracle properties so that the post-classification estimators are asymptotically equivalent to the infeasible estimators that employ correct group identities. We establish two Wald tests based on post-classification estimators for explosiveness detection. We suggest a panel recursive procedure to detect the origination of economic exuberance. Moreover, we provide a novel asymptotic theory for concentration inequalities, clustering algorithms, and Wald tests on mixed-roots panels. Extensive Monte Carlo simulations provide strong evidence that the proposed panel approaches result in substantial power gains. Finally, we apply the panel approach to China's real estate market.

1 Introduction

Financial bubbles, such as the dot-com bubble, are well recognized as explosive deviations of asset prices from their fundamental values. Unfortunately, neither bubbles nor fundamental values are observable, and it is not straightforward to detect bubbles. Interestingly, several time-series methods have been proposed recently to detect the presence of a bubble and, if one appears, to estimate the bubble's origination and termination dates. Although these augmented Dickey-Fuller (ADF)-type statistics can consistently identify bubbles, they have low power when a bubble is short lived or when a bubble grows slowly.

This study proposes an inference method to increase the power in identifying bubbles and We use panel data to aid the analysis. Intuitively, as long as there is some homogeneity

*Email: yanbo.liu.2015@phdecons.smu.edu.sg.

over cross-sectional units within groups, panel autoregressions based on pooling cross-sectional data within the same group should deliver power improvements to sharpen the statistical inferences on the common explosive root. In most applications, the true group structure is latent. Therefore, a data-driven clustering algorithm is needed. To identify the latent membership, in this study, we investigate several grouping approaches, namely, k-means clustering algorithms. Moreover, we aim to determine the true group number by the Bayesian information criterion (BIC).

Our new bubble inference procedure accommodates a mixed-roots panel autoregressive model with unobserved groups. We consider two k-means classifiers: the recursive k-means algorithm of Bonhomme and Manresa (2016), and the modified k-means algorithm of Lin and Ng (2012). We show the uniform consistency of both k-means clustering algorithms. Moreover, we apply the recursive k-means algorithm to an explosive panel autoregression model. We also employ the modified k-means algorithm for a mixed-roots panel autoregression model, which contains both explosive and stationary time series. This study derives the oracle property of two post-classification estimators under the joint asymptotic scheme $n \to \infty$ and $T \to \infty$. The oracle property reveals that the distance between the post-classification estimators and the oracle-within estimators is diminishing. The diminishing distance verifies the optimality of the k-means clustering approaches from the perspective of estimations. To the best of our knowledge, this is the first such attempted analysis of mixed-roots panels.

Based on the abovementioned two k-means classifications, we provide two Wald statistics for consistent detection of explosiveness. Under the null hypothesis of the panel unit root in a specific group, the proposed statistics converge to $\chi^2(1)$. They diverge under the alternative hypothesis of explosive roots. Our Wald statistics are superior to the ADF statistics in two aspects. First, the Wald statistics are more powerful than the ADF statistics. Our asymptotic theory shows that the Wald statistics diverge at a faster rate than the ADF statistics. Extensive Monte Carlo simulations demonstrate that the empirical power of the Wald statistics in realistic situations is nearly double that of the ADF statistics. Second, unlike the ADF statistics, whose limiting distribution is nonpivotal, our Wald statistics are asymptotically pivotal under the null hypothesis. Hence, it is much easier to find critical values in our Wald test than in the ADF test.

In the present study, we seek to provide a classified panel approach to the dating of bubble phenomena. Based on the supreme Wald statistics after k-means classifications, we develop a real-time detector for bubble origination dates and we derive a rigorous limit theory for estimators. We suggest recursive Wald tests and panel explosive autoregression asymptotics to evaluate explosive behaviour. We show that these methods are capable of detecting the existence of explosive behaviour based on k-means classifications. We use them to date the origination of real estate bubbles in China's large and medium-sized cities of the recent 2 decades. The empirical significance of housing bubbles corresponds to the conjecture by Chen and Wen (2017) that China's housing boom was a rational bubble.

Generally, we propose our inference procedure based on the established literature on bubble detection and classification algorithms. Our study adapts the real-time detector of bubble phenomena in Phillips et al. (2011, PWY hereafter), Harvey et al. (2013), and Phillips et al. (2015a, 2015b, PSY hereafter), which are bubble detection studies addressing the existence of explosive behaviour. Classification studies addressing the latent membership problem, namely, Lin and Ng (2012), Bonhomme and Manresa (2016), and Su et al. (2019), discuss the k-means algorithm, while Su et al. (2016) and Huang et al. (2019) employ the classifier LASSO algorithm.

PWY propose a recursive ADF procedure to detect the bubble when there is one bubble in a sample. PSY (2015ab) extend the algorithm to multiple bubbles. Both PWY and PSY procedures have consistency. However, these ADF-type statistics suffer from the problem of low power for a short-lived bubble or a slowly growing bubble; see Tables 4 and 6 in PSY (2015a) and Tables 2 and 6–10 in PSY (2015b). To improve the power, Harvey et al. (2013) propose a union of rejection strategies based on two individual tests and demonstrate that the union always has better power than the recursive ADF approach. This study proposes panel procedures to increase the power in identifying bubbles under the joint convergence framework $n \to \infty$ and $T \to \infty$. When the cross-sectional dimension is a constant, our panel approach collapses to the model averaging method in Harvey et al. (2013).

There is a growing body of literature on large dimensional panels with latent memberships. To recover true group structures, several grouping approaches have been proposed to identify the unknown memberships. One is k-means clustering; see Lin and Ng (2012), Sarafidis and Weber (2015), Bonhomme and Manresa (2016), and Su et al. (2019). The other is penalized classification; see Su et al. (2016) for classifier LASSO estimates and Ke et al. (2015) for data-driven segmentation (CARDS). A major concern for applying classification in bubble detection is whether there is evidence of panel homogeneity within groups. Fortunately, there is empirical evidence to support panel homogeneity. For example, common explosive behaviour over cross-sectional units within the same group are discussed in the literature (Pavlidis et al., 2016; Shi, 2016). Narayan et al. (2013) find strong evidence of group-specific heterogeneity when the stocks of 589 firms belong to nine sectors.

This study makes three methodological contributions. First, it contributes to the literature on bubble detection. Based on uniformly consistent classification of individuals, the proposed Wald statistics greatly enhance inferential powers as the solution to the power loss problem in the literature.

Second, this study contributes to the literature on data-driven classification. Although there are several existing clustering algorithms, most of them are developed for the stationary case only. An exception is the LASSO algorithm in Huang et al. (2019), which, however, does not apply to mixed-roots panel autoregressions. Intuitively, the classifier LASSO algorithm can provide classification and estimation by minimizing the penalized objective function. However, in the mixed-roots case, we cannot adjust the LASSO objective function without prior information on true membership. This is because the stationary time series needs polynomial rates, while the explosive time series requires exponential rates. Therefore, we contribute to the classification literature by searching for a suitable clustering algorithm on mixed-roots panels. In this study, we apply modified k-means algorithms for mixed-roots panels. The reason that the recursive k-means algorithm does not apply to mixed-roots panels is the adjustment rate pitfalls described above.

In addition, the study contributes to applications of concentration inequality. In the machine learning literature, concentration inequalities are the cornerstones of classification algorithms. Concentration inequality justifies the misclassification errors by controlling the decaying rates of sample moments at tails. In the stationary case, sample moments converge to the population moments. The uniform upper bounds of stationary sample moments depend on their population counterparts. However, the sample moments of explosive roots converge to a random variable. The uniform upper bound of explosive sample moments relies on the asymptotic distance between sample moments and population moments. Therefore, we introduce a novel decomposition of explosive sample moments to accommodate exponential inequality for the mixed-roots case.

The rest of this paper is organized as follows. Section 2 discusses the model setup. Section 3 introduces the classifications and estimations of mixed-roots panel autoregressions. Section 4 provides asymptotic properties of the k-means classifications and estimations. Section 5 presents inferential procedures and the model selection criterion. Section 6 reports simulated results. Section 7 applies our method to the data of China's real estate markets. Section 8 concludes.

We use the following notations throughout the study. I_d and $0_{d\times d}$ denote the $d \times d$ identity matrix and a $d \times d$ matrix of zeros, respectively. $\xrightarrow{p}$, $\xrightarrow{d}$, and $\Rightarrow$ denote convergence in probability, convergence in distribution, and convergence in functional space, respectively. Correspondingly, $(n, T) \to \infty$ denotes the joint limit, and $(n, T)_{seq} \to \infty$ denotes the sequential limit, where $T \to \infty$ followed by $n \to \infty$. Similarly, $(T, n)_{seq} \to \infty$ denotes the sequential limit, where $n \to \infty$ followed by $T \to \infty$. We denote $T_1 := T - 1$. $A >> B$ implies $B/A = o_p(1)$ as $(n, T) \to \infty$.

2 Model Setup

We formulate the panel autoregressive model as

$$y_{it} = \mu_i(\rho_i - 1) + \rho_i y_{i,t-1} + u_{it},\ i = 1, 2, ..., n;\ t = 1, 2, ..., T. \tag{1}$$

In the model, μ_i is an individual fixed effect for each i with $\sigma_\mu^2 = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n \mu_i^2 < \infty$. The error process $\{u_{it}\}$ is a martingale difference sequence with a conditional second moment σ^2 ($\mathbb{E}\left(u_{it}^2|\mathcal{F}_{i,t-1}\right) = \sigma^2$ for any i and t, where $\mathcal{F}_{i,t-1} := \sigma\{u_{i,t-1}, u_{i,t-2}, ...\}$) and finite finite q^{th} moments with $q \geq 4$ for all i and t. We assume $y_{i0} = 0$ almost certainly for all i, and y_{i0} is independent of u_{it} for all i and t. We further assume $u_{is} = 0$ for all $s \leq 0$.

For slopes, we assume

$$\rho_i = 1 + \frac{c_i}{T^\gamma}.$$

There exist $\underline{c}$ and $\bar{c}$ such that $\underline{c} < |c_i| < \bar{c}$ with $0 < \bar{c}$, $\underline{c} < \infty$ and $\gamma \in [0, 1)$. We formulate homogeneous scaling parameter γ and heterogeneous distance parameters, $\{c_i\}_{i=1}^n$. When $\gamma = 0$, we have standard explosive root case, $\rho_i = 1 + c_i$. For consistent estimation on the scaling parameter γ, see Phillips (2013).

On one hand, in conventional panel analysis, one typically assumes that $c_i = c$, leading to a panel with homogeneous slopes and heterogeneous individual fixed effects $\{\mu_i\}_{i=1}^n$. On the other hand, one can assume that $c_i \neq c_j$ for any $i \neq j$, leading to a panel with the maximum degree of heterogeneity in distance parameters.

In this study, we adopt a setup between the two extreme cases. In particular, we assume the following group structure as

$$c_i = \sum_{g=1}^{K^0} \alpha_g^0 \mathbf{1}\left\{i \in G_g^0\right\},$$

where $\alpha_g^0 \neq \alpha_l^0$ for any $g \neq l$, $\bigcup_{g=1}^{K_0} G_g^0 = \{1, 2, ..., n\}$, and $G_g^0 \cap G_l^0 = \varnothing$ for any $j \neq g$. Let $n_g := \# G_g^0$ represents the cardinality of the true group G_g^0. Let $\mathcal{A}$ be a set of arbitrary $K^0 \times 1$ vectors $\alpha\,(:= (\alpha_1, \alpha_2, ..., \alpha_{K^0}))$, and $\mathcal{C}$ be a set of group-specific distance parameters, so that $c_n\,(:= (c_{1n}, c_{2n}, ..., c_{K^0n})) \in \mathcal{C}$. In the bubble detection literature, K^0 corresponds to the number of market sectors for stocks or convergence clubs for real estate prices. Within the same market sector or convergence club, all cross-sectional units share the identical distance parameter α_g^0. To obtain the asymptotic properties of classification and inference, we first assume that K^0 is fixed and known while the true memberships are latent and unknown. We then propose using the BIC to select the true group number.

The true group-specific parameters are defined as $c_n^0 := \left(c_{1n}^0, c_{2n}^0, ..., c_{K^0n}^0\right) \in \mathcal{C}$, $\alpha^0 := \left(\alpha_1^0, ..., \alpha_{K^0}^0\right) \in \mathcal{A}$ and $c^0 := \left(c_1^0, ..., c_n^0\right) \in \Phi^n$, where $\Phi := [-\bar{c}, -\underline{c}] \cup [\underline{c}, \bar{c}]$. The true group membership variable $\left\{g_i^0\right\}_{i=1}^n$ maps individual units into groups. For each $i = 1, 2, ..., n$, and $g = 1, 2, ..., K^0$, the event '$g_i^0 = g$' is equivalent to '$i \in G_g^0$'. With any estimator $\{\widehat{g}_i\}_{i=1}^n$, the event '$\widehat{g}_i = g$' is equivalent to '$i \in \widehat{G}_g$' for each $i = 1, 2, ..., n$, and $g = 1, 2, ..., K^0$. We denote $\delta := (g_1, g_2, ..., g_n) \in \Delta_{K^0}$ as a particular grouping of n, where Δ_{K^0} unit is the set of all groupings of $\{1, 2, ..., n\}$ into at most K^0 groups. For the g-th group, we define slopes and the relevant estimators as $\rho_{gn}^0 := \exp\left(\alpha_g^0 / T^\gamma\right)$, $\rho_{\widehat{g}n}^0 := \exp\left(\alpha_{\widehat{g}}^0 / T^\gamma\right)$, $\widehat{\rho}_{\widehat{g}n} := \exp\left(\widehat{\alpha}_{\widehat{g}} / T^\gamma\right)$. For simplicity, we write $\rho_i^0 \left(:= \exp\left(\frac{c_i^0}{T^\gamma}\right)\right)$ as ρ_i.

With the existence of latent group structures, we consider two kinds of panel autoregressions in the present study: One is the mixed-roots panel, and the other is the pure explosive panel. For mixed-roots panel autoregressions, there are two classes of groups: one with explosive roots ($\alpha_g^0 > 0$), and the other with stationary roots ($\alpha_g^0 < 0$). For the pure explosive panel, $\alpha_g^0 > 0$ for $g = 1, 2, ..., K^0$.

3 Classification and Estimation

For explosiveness detection, we consider both cases: (I) explosive panel with latent memberships; and (II) mixed-roots panel with latent memberships. For both cases (I) and (II), we apply classification methods in the first step and attain post-classification estimators and relevant Wald statistics in the second step. The difference between (I) and (II) lies in the first step: for case (I), we employ the recursive k-means classification proposed in Bonhomme and Manresa (2016), and for case (II), we consider the modified k-means classification proposed in Lin and Ng (2012).

After clustering in the first step, we consider two pooled least square estimators for α^0, c_n^0 and c^0: oracle within estimator and post-classification within estimator. We define $\overline{y}_{i,-1} := \frac{1}{T}\sum_{t=1}^T y_{i,t-1}$, $\overline{u}_i := \frac{1}{T}\sum_{t=1}^T u_{it}$, $\widetilde{y}_{i,t-1} := y_{i,t-1} - \overline{y}_{i,-1}$ and $\widetilde{u}_{i,t} := u_{i,t} - \overline{u}_i$. The oracle within estimator for the slope in the g-th group is

$$\widetilde{\alpha}_g - \alpha_g^0 = T^{\gamma}\frac{\sum_{i\in G_g^0}\sum_{t=1}^T \widetilde{y}_{i,t-1}\widetilde{u}_{i,t}}{\sum_{i\in G_g^0}\sum_{t=1}^T \widetilde{y}_{i,t-1}^2}. \tag{2}$$

Similarly, the post-classification within estimator for the g-th group is

$$\widehat{\alpha}_{\widehat{g}} - \alpha_g^0 = T^{\gamma}\frac{\sum_{i\in \widehat{G}_g}\sum_{t=1}^T \widetilde{y}_{i,t-1}\widetilde{u}_{i,t}}{\sum_{i\in \widehat{G}_g}\sum_{t=1}^T \widetilde{y}_{i,t-1}^2}. \tag{3}$$

We define $\left\{\widehat{G}_g\right\}_{g=1}^{K^0}$ as any consistent classification estimates on $\left\{G_g^0\right\}_{g=1}^{K^0}$. Therefore, we can establish inference procedures based on (3) in the second step.

Based on the discussions in the online supplement, we find that incidental parameters have no presence in the asymptotics of sample moments. The reason this happens is that for each $i = 1, 2, ..., n$, the fixed effect $\mu_i(1-\rho_i)$ has lower order $O_p(T^{-\gamma})$, dominated by noise. To illustrate this point clearly, we perceive the intertemporal summations of $y_{i,t}$ as

$$y_{it} = \frac{\rho_i^t - 1}{\rho_i - 1}\mu_i(-\rho_i + 1) + \sum_{s=1}^t \rho_i^{t-s}u_{is}, \tag{4}$$

where $i = 1, 2, ..., n$ and $t = 1, 2, ..., T$. The terms $\frac{\rho_i^t-1}{\rho_i-1}\mu_i(-\rho_i+1) = O\left(\rho_i^t\right)$ and $\sum_{s=1}^t \rho_i^{t-s}u_{is} = O_p\left(\rho_i^t T^{\gamma}\right)$. We clearly find that incidental parameters have smaller orders for each $i = 1, 2, ..., n$. Moreover, noise, not fixed effects, determines the asymptotics uniformly across $i = 1, 2, ..., n$. Therefore, to determine uniform bounds of sample moments, we need only consider innovations.

3.1 Recursive k-means algorithm

In this subsection, we consider explosive panel autoregression with latent groups. When memberships are unobserved, two types of parameters are considered: the parameter vector $\{c_i\}_{i=1}^n \subseteq [\underline{c}, \overline{c}]$, and the group membership variable $\{g_i\}_{i=1}^n$, which maps cross-sectional units into groups. Note that group-specific distance parameters, $\{c_i\}_{i=1}^n$, are well separated with minimum distance $c_{g,\widetilde{g}} > 0$; otherwise, we cannot correctly allocate individuals into the true groups.

The grouped estimators for $\left\{c_i^0\right\}_{i=1}^n$, $\left\{g_i^0\right\}_{i=1}^n$ in (1) are defined as the solution to the following optimization problem:

$$(\widehat{c}, \widehat{\delta}) = \arg \min_{(c,\delta)\in\Phi^n\times\Delta_{K^0}} \frac{1}{nT^{2\gamma}} \sum_{i=1}^n \frac{1}{\overline{\rho}_i^{2T}} \sum_{t=1}^T \left(\widetilde{y}_{it} - \widetilde{y}_{i,t-1} \exp\left(\frac{c_{g_i n}}{T^\gamma}\right)\right)^2, \tag{5}$$

taken over all possible groupings $\delta = \{g_1, g_2, ..., g_n\}$ of the n units into K^0 groups, and group-specific parameters $\{c_i\}_{i=1}^n$. We employ $\{\overline{\rho}_i\}_{i=1}^n$ as the collection of least squares estimates for each individual time series. To eliminate fixed effects, we employ a demeaned process as $\widetilde{y}_{it} := y_{it} - \overline{y}_i$ and $\widetilde{y}_{i,t-1} := y_{it} - \overline{y}_{i,-1}$. For given values of $\{c_{gn}\}_{g=1}^{K^0}$, the optimal group clustering for each $i = 1, 2, ..., n$, is

$$\widehat{g}_i(c) = \arg \min_{g\in\{1,2,...,K^0\}} \frac{1}{T^{2\gamma}\overline{\rho}_i^{2T}} \sum_{t=1}^T \left[\widetilde{y}_{it} - \widetilde{y}_{i,t-1}\exp\left(\frac{c_{gn}}{T^\gamma}\right)\right]^2, \tag{6}$$

where the minimum g optimizes a k-means classification problem. The estimator $\{\widehat{c}_i\}_{i=1}^n$ of (5) optimizes the following objective function:

$$\widehat{c} = \arg \min_{c\in\Phi^n} \frac{1}{nT^{2\gamma}} \sum_{i=1}^n \frac{1}{\overline{\rho}_i^{2T}} \sum_{t=1}^T \left(\widetilde{y}_{it} - \widetilde{y}_{i,t-1}\exp\left(\frac{c_{\widehat{g}_i(c)n}}{T^\gamma}\right)\right)^2, \tag{7}$$

where $\widehat{g}_i(c)$ is derived by (6). The classification estimates of $\left\{g_i^0\right\}_{i=1}^n$ are simply $\widehat{g}_i(\widehat{c})$.

The following computational algorithm summarizes the recursive k-means procedure to minimize (5) in the following steps.

Step 1: Let $\left\{\overline{c}_i^{(0)}\right\}_{i=1}^n$ be the collection of individual least squares estimates for all $i \in \{1, 2, ..., n\}$;

Step 2: Compute for any $i = 1, 2, ..., n$,

$$g_i^{(s+1)} = \arg \min_{g\in\{1,2,...,K^0\}} \frac{1}{T^{2\gamma}\overline{\rho}_i^{2T}} \sum_{t=1}^T \left(\widetilde{y}_{it} - \widetilde{y}_{i,t-1} \exp\left(\frac{c_{gn}^{(s)}}{T^\gamma}\right)\right)^2; \tag{8}$$

Step 3: Compute

$$\widehat{c}^{(s+1)} = \arg \min_{c\in\Phi^n} \frac{1}{n} \sum_{i=1}^n \frac{1}{T^{2\gamma}\overline{\rho}_i^{2T}} \sum_{t=1}^T \left(\widetilde{y}_{it} - \widetilde{y}_{i,t-1} \exp\left(\frac{c_{g_i^{(s+1)}n}}{T^\gamma}\right)\right)^2; \tag{9}$$

Step 4: Set s = s +1 and go to Step 2 (until numerical convergence).

This computation algorithm consists of two iterated steps, 'assignment' as in Step 2 and 'update' as in Step 3. In the 'assignment' step, each cross-section unit i is assigned to the nearest group g_i based on the distance defined in (8). In the 'assignment' step, to re-allocate centres of groups $\{g_i\}_{i=1}^n$, we compute $\{\widehat{c}_i\}_{i=1}^n$ by minimizing (9).

3.2 Modified k-means algorithm

If (1) incorporates both explosive and stationary roots, the recursive k-means classification algorithm fails. Because of heterogeneity in adjustment rates, the sample moments of stationary individuals are asymptotically unstable. This instability leads to severe misclassification. Therefore, there is a need for new methods.

We introduce the modified k-means classification approach to accommodate mixed roots. We follow the clustering approach in Lin and Ng (2012). In the present study, we establish the asymptotic property of modified k-means in mixed-roots panel autoregressions for the first time.

We summarize the computational algorithm in the following steps.

Step 1: We derive least squares estimates $\{\widehat{c}_i\}_{i=1}^n$ as

$$\widehat{c}_i - c_i = T^{\gamma} \frac{\sum_{t=1}^T \widetilde{y}_{i,t-1}\widetilde{u}_{it}}{\sum_{t=1}^T \widetilde{y}_{i,t-1}^2} = T^{\gamma} \frac{\sum_{t=1}^T \left(y_{i,t-1} - \overline{y}_{i,-1}\right)\left(u_{it} - \overline{u}_i\right)}{\sum_{t=1}^T \left(y_{i,t-1} - \overline{y}_{i,-1}\right)^2}. \tag{10}$$

Step 2: To recover latent memberships, it is straightforward to apply the k-means cluster algorithm for $\{\widehat{c}_i\}_{i=1}^n$. Specifically, with $\alpha = (\alpha_1, \alpha_2, ..., \alpha_{K^0}) \in \mathcal{A}$ being any arbitrary $K^0 \times 1$ vector for $\alpha_1, \alpha_2, ..., \alpha_{K^0}$, we define

$$\widehat{Q}_n(\alpha) = \frac{1}{n}\sum_{i=1}^n \min_{1\le l\le k} \left(\widehat{c}_i - \alpha_l\right)^2,$$

and $\widehat{\alpha} = (\widehat{\alpha}_1, \widehat{\alpha}_2, ..., \widehat{\alpha}_{K^0})$ with $\widehat{\alpha} := \arg\min_{\mathcal{A}} \widehat{Q}_n(\alpha)$. Therefore, we further compute the estimated cluster identity as

$$\widehat{g}_i = \arg\min_{1\le l\le K} \left|\widehat{c}_i - \widehat{\alpha}_l\right|,$$

where if there are multiple l's that achieve the minimum, $\widehat{g}_i$ takes the value of the smallest one.

Step 3: We apply pooled least squares in each estimated group.

When $c_i > 0$, we have $\widehat{c}_i - c_i = O_p\left(\frac{1}{\rho_i^T}\right)$. When $c_i < 0$, we have $\widehat{c}_i - c_i = O_p\left(\frac{1}{T^{\frac{1-\gamma}{2}}}\right)$. However, a pointwise convergence rate is insufficient for theoretical derivations on the classifications. Classification algorithms require controlling the uniform upper bound of sample moments. Therefore, we need to verify the uniform convergence rate for $\widehat{c}_i$ over $i = 1, 2, ..., n$.

4 Statistical Properties

In this section, we study the asymptotic properties of the recursive k-means and modified k-means algorithms.

Given that we can recover the latent group structure consistently, we can demonstrate that the estimators of the slope are asymptotically equivalent to the oracle estimators that are derived as if the true group structure were known. Then, we provide the asymptotic distributions of the slope estimators.

We denote $\widehat{g}_i := \widehat{g}_i(\widehat{c})$ as any k-means classification estimates for g_i^0 for each $i = 1, 2, ...n$. To demonstrate consistency of the classifications, we intend to establish the consistency of $\widehat{c}$ in terms of the Hausdorff distance. The Hausdorff distance measures how far two compact subsets of a metric space are separated from each other:

$$d_H(a,b) = \left\{ \max_{g \in \{1,2,...,K^0\}} \left(\min_{\widetilde{g} \in \{1,2,...,K^0\}} \left(a_{\widetilde{g}} - b_g\right)^2 \right), \max_{\widetilde{g} \in \{1,2,...,K^0\}} \left(\min_{g \in \{1,2,...,K^0\}} \left(a_{\widetilde{g}} - b_g\right) \right)^2 \right\}.$$

To establish consistency of estimation and classification, we assume the following for the model.

Assumption 1 *(i) There exist $\overline{c}$ and $\underline{c}$ such that for each $i \in \{1, 2, ..., n\}$ we have $0 < \underline{c} \leq \left|c_i^0\right| \leq \overline{c} < \infty$.*

(ii) There exists a constant $c_{g,g'} \in (0, \infty)$ such that $\inf_{1 \leq g \leq g' \leq K} \left|\alpha_g^0 - \alpha_{g'}^0\right| \geq c_{g,g'}$.

(iii) Let $\{n_g\}_{g=1}^{K^0}$ denote the cardinality of latent groups. For each $g \in \left\{1, 2, ..., K^0\right\}$, $\frac{n_g}{n} \to \pi_g < \infty$. Moreover, $\inf_{1 \leq g \leq K^0} \pi_g \geq \underline{M}$ for some $\underline{M} > 0$.

(iv) The following rate restrictions hold: $\sqrt{n}/\log(n) = o\left(T^{\frac{3\gamma-1}{2}}\right)$, $n = o\left(T^{1-\gamma}\right)$, $n^{\frac{1}{4}}/\log(n) = o\left(T^{\frac{1}{4}}\right)$.

(v) The relationship $\delta_{nT} < \frac{\underline{M} c_{g,\widetilde{g}}^2}{15\overline{c}}$ holds for each n and T, where $\delta_{nT} := \frac{n^{\frac{1}{4}+\epsilon}\sqrt{\log(n)}}{T^{\frac{\gamma}{2}+\frac{3}{4}}}$ with arbitrary $\epsilon > 0$.

(vi) The individual fixed effect μ_i is $O_p(1)$ for each i with $\sigma_\mu^2 = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^n \mu_i^2 < \infty$.

(vii) The error process $\{u_{it}\}$ is a martingale difference sequence with a homogeneous conditional second moment σ^2 ($\mathbb{E}\left(u_{it}^2 | \mathcal{F}_{i,t-1}\right) = \sigma^2$ for any i and t, where $\mathcal{F}_{i,t-1} := \sigma\left\{u_{i,t-1}, u_{i,t-2}, ...\right\}$) and finite q^{th} moments with $q \geq 4$ for all i and t.

(viii) Initial conditions: Assume that $y_{i0} = 0$ almost surely for all i, y_{i0} is independent of u_{it} for all i and t, and $u_{is} = 0$ for all $s \leq 0$.

Assumption 1(i) imposes an identification condition for c^0 as one interior value of the compact space. The identification condition for Assumption 2(ii) is that the group-specific parameters are well separated from each other. Assumption 1(iii) implies that each group size is increasing proportionally to the dimension of individuals. Assumption 1(iv) imposes rate restrictions to apply exponential inequalities. Assumption 1(v) provides necessary conditions for uniform consistency in the modified k-means clustering. Assumption 1(vi) provides restrictions on individual fixed effects. Assumption 1(vii) assumes the martingale property for innovations. Assumption 1(viii) imposes restrictions on initial values.

4.1 Recursive k-means algorithm

We establish the consistency of the coefficient estimates of the recursive k-means algorithm in the explosive panel autoregressions.

Theorem 4.1 *(Estimation Consistency) Assumption 1(i)˜(iii) and 1(vi)˜(viii) hold, and* $c_i^0 > 0$ *for each i. Under* $(n, T) \to \infty$,

$$d_H(c^0, \widehat{c}) \xrightarrow{p} 0.$$

By Theorem 4.1 there is one permutation $\tau : \{1, 2, ...K^0\} \to \{1, 2, ..., K^0\}$ such that parameter estimates converge to the true values,

$$\left(\widehat{c}_{\tau(g)n} - c_{gn}^0\right)^2 \xrightarrow{p} 0.$$

By relabelling $\widehat{c}$, we take $\tau(g) = g$. For any $\eta > 0$, we define

$$\mathcal{N}_\eta := \left\{c_n \in [\underline{c}, \overline{c}]^{K^0} : \left(c_{gn}^0 - c_{gn}\right)^2 < \eta, \forall g \in \left\{1, 2, ..., K^0\right\}\right\}. \tag{11}$$

Let $\widehat{g}_i(\widehat{c}) = \arg\min_{g \in \{1,2,...,K^0\}} \sum_{t=1}^T \left(y_{it} - y_{i,t-1} \exp\left(\frac{\widehat{c}_{gn}}{T^\gamma}\right)\right)^2$. After verifying the consistency of $\widehat{c}$ for c^0, we provide the individual and uniform consistency of recursive k-means classifications in the following theorems.

Theorem 4.2 *(Individual Consistency of Classification) Let Assumption 1 hold and* $c_i^0 > 0$ *for each* $i = 1, 2, ..., n$. *With joint convergence* $(n, T) \to \infty$,

$$\Pr\left(\max_{1 \le i \le n} \left|\widehat{g}_i - g_i^0\right| > 0\right) = o(1).$$

Theorem 4.2 clearly justifies the individual consistency of the recursive k-means classifications. Theorem 4.2 is similar to Theorem 2 of Bonhonmme and Manresa (2016). This theorem states that under joint convergence framework $(n, T) \to \infty$, we can correctly estimate the group structure. From Theorem 4.2, we observe that correct classifications strongly rely on Assumption 1(ii). In our discussion, as long as the slope coefficients are separate across groups, the misclassification errors are asymptotically negligible.

Assume that $\widehat{G}_g$ represents the g-th estimated classification group and G_g^0 denotes the g-th true group in the population. To rigorously state the uniform consistency of the recursive k-means classifications, we define the following sequences of events as

$$\widehat{E}_{g,i} := \left\{i \notin \widehat{G}_g | i \in G_g^0\right\} \text{ and } \widehat{F}_{g,i} := \left\{i \notin \ G_g^0 | i \in \widehat{G}_g\right\}, \tag{12}$$

for $g = 1, 2, ..., K^0$ and $i = 1, 2, ..., n$. Let $\widehat{E}_{g,nT} := \bigcup_{i \in G_g^0} \widehat{E}_{g,i}$ and $\widehat{F}_{g,nT} := \bigcup_{i \in \widehat{G}_g} \widehat{F}_{g,i}$. $\widehat{E}_{g,nT}$ demonstrates the error event of not classifying the individual unit of G_g^0 into $\widehat{G}_g$ as a type-I classification error; $\widehat{E}_{g,nT}$ demonstrates the error event of not clustering the cross-sectional unit of $\widehat{G}_g$ into G_g^0 as a type-II classification error. Furthermore, we establish the uniform consistency of the recursive k-means classifier.

Theorem 4.3 *(Uniform Consistency of Classification) Let Assumption 1 hold and* $c_i^0 > 0$ *for each* $i = 1, 2, ..., n$. *Under joint convergence* $(n, T) \to \infty$,

(i) $\Pr\left(\bigcup_{g=1}^{K^0} \widehat{E}_{g,nT}\right) \le \sum_{g=1}^{K^0} \Pr\left(\widehat{E}_{g,nT}\right) \to 0$*; and*

(ii) $\Pr\left(\bigcup_{g=1}^{K^0} \widehat{F}_{g,nT}\right) \le \sum_{g=1}^{K^0} \Pr\left(\widehat{F}_{g,nT}\right) \to 0.$

Theorem 4.3 illustrates that for all $g \in \{1, 2, ..., K^0\}$ all cross-sectional units belonging to group G_g^0 are clustered into the same clustered group $\widehat{G}_g$ asymptotically. Meanwhile, all cross-sectional agents classified into the same group $\widehat{G}_g$ for all $g \in \{1, 2, ..., K^0\}$ belong to the same group G_g^0 in the probability limit. These observations show that the summarized classification errors are diminishing asymptotically.

The following theorem indicates that the post-classification within estimator $\widehat{\alpha}_{\widehat{g}}$ is asymptotically equivalent to the oracle within estimator $\widehat{\alpha}_g$ for each $g \in \{1, 2, ...K^0\}$. With uniform consistency of the clustering algorithm, we can verify the following argument:

$$\sqrt{n_g}\left(\rho_{gn}^0\right)^T\left(\widehat{\alpha}_{\widehat{g}} - \alpha_g^0\right) = \sqrt{n_g}\left(\rho_{gn}^0\right)^T\left(\widehat{\alpha}_g - \alpha_g^0\right) + o_p(1).$$

In this sense, the post-classification estimator $\widehat{\alpha}_{\widehat{g}}$ has an asymptotic oracle property for $\widehat{g} \in \{1, 2, ...K^0\}$. The following theorem shows the limiting distribution of $\widehat{\alpha}_{\widehat{g}}$.

Theorem 4.4 *Suppose Assumption 1 holds and $c_i^0 > 0$ for $i = 1, 2, ..., n$. Under joint convergence $(n, T) \to \infty$ and rate restriction $\frac{n}{T^{2-2\gamma}} = o(1)$,*

$$\sqrt{n_g}\left(\rho_{gn}^0\right)^T\left(\widehat{\alpha}_{\widehat{g}} - \alpha_g^0\right) \xrightarrow{d} N(0, 2\left(\alpha_g^0\right)),$$

where $\rho_{gn}^0 := \exp\left(\alpha_g^0/T^\gamma\right)$.

We provide two remarks. First, we define the estimator minimizing the objective function (5) as $\widehat{\alpha}_{\widehat{g}}^{WY}$. Although $\widehat{\alpha}_{\widehat{g}}^{WY}$ is a consistent estimator for each g, it does not enjoy the oracle property: the distance between the oracle estimator $\widehat{\alpha}_g$ and the true value α_g^0 is $\widehat{\alpha}_g - \alpha_g^0 = O_p\left(\frac{1}{\sqrt{n}\left(\rho_{gn}^0\right)^T}\right)$ while the misclassification error decays at a slower rate and is smaller than $O_p\left(\frac{1}{\sqrt{n}\left(\rho_{gn}^0\right)^T}\right)$. Therefore, the misclassification error dominates the limiting distribution part and ruins the oracle property.

Second, we prefer the within estimator to the first-differenced estimator. We define the first-differenced least squares estimates for homogeneous panels with explosive roots as $\widehat{\rho}^{FD}$ and $\widehat{c}^{FD}$. Under $(n, T) \to \infty$ and $\frac{\sqrt{n}T}{\rho^T} = o(1)$, we have

$$\sqrt{n}\left(\rho^0\right)^T\left(\widehat{\rho}^{FD} - \rho^0\right) \xrightarrow{p} 0;$$

while $\frac{\sqrt{n}T}{\rho^T} = o(1)$ restriction fails, we have

$$\frac{\left(\rho^0\right)^{2T}}{T}\left(\widehat{\rho}^{FD} - \rho^0\right) \xrightarrow{p} -4.$$

Due to the dominance of exponential rate, we usually have the first case. We collect the technical details in the online supplement. We perceive that the first-differenced estimator suffers from degenerate distributions. Therefore, no valid inference procedures are available for the first-differenced estimator. This phenomenon again justifies the superiority of our choice for the within estimators.

4.2 Modified k-means algorithm

We establish the uniform consistency of the coefficient estimates in the mixed-roots panel autoregressions.

Lemma 4.1 *Assume rate restriction* $\frac{3}{13} < \gamma < \frac{2}{3}$. *If Assumption 1 holds,*

$$d_H\left(c_n^0, \widehat{\alpha}\right) = O\left(\delta_{nT}^{\frac{1}{2}}\right),$$

where $\delta_{nT} := \frac{n^{\frac{1}{4}+\epsilon}\sqrt{\log(n)}}{T^{\frac{\gamma}{2}+\frac{3}{4}}}$ *with arbitrary* $\epsilon > 0$.

When n grows more slowly than T, the Hausdorff distance between c_n^0 and $\widehat{\alpha}$ is asymptotically diminishing. This lemma shows the uniform consistency of parameter estimations, based on which the following theorems establish the individual and uniform consistency of the modified k-means algorithm for mixed-roots panel autoregressions.

Theorem 4.5 *(Individual Classification Consistency) Assumption 1 and* $\frac{3}{13} < \gamma < \frac{2}{3}$ *hold. Under* $(n, T) \to \infty$,

$$\sup_{1 \le i \le n} \mathbf{1}\left\{\widehat{g}_i \ne g_i^0\right\} = o_p(1).$$

Theorem 4.5 shows that the modified k-means algorithm consistently recovers latent memberships for mixed-roots panel autoregressions. This machinery incorporates more general cases than the recursive classification algorithm. For the literature on bubble detection, it is too restrictive to impose homogeneous explosive roots. Thus, it is more realistic to argue that explosive roots exist in a proportion of individuals. The recursive k-means algorithm does not apply to the mixed-roots model. Therefore, under the general mixed roots case, the modified k-means algorithm is superior and powerful.

Similarly, to verify the oracle property on the post-classification estimator, we provide the uniform consistency of our modified k-means classification algorithm.

Theorem 4.6 *(Uniform Consistency of Classification) Assumption 1 and* $\frac{3}{13} < \gamma < \frac{2}{3}$ *hold. Under joint convergence* $(n, T) \to \infty$,

(i) $\Pr\left(\bigcup_{g=1}^{K^0} \widehat{E}_{g,nT}\right) \le \sum_{g=1}^{K^0} \Pr\left(\widehat{E}_{g,nT}\right) \to 0$; *and*

(ii) $\Pr\left(\bigcup_{g=1}^{K^0} \widehat{F}_{g,nT}\right) \le \sum_{g=1}^{K^0} \Pr\left(\widehat{F}_{g,nT}\right) \to 0$.

Once we attain the uniform consistency of the modified k-means classification, we can derive the asymptotic oracle property of (3). The following theorem reports the asymptotic distribution of $\widehat{\alpha}_{\widehat{g}}$.

Theorem 4.7 *Assume* $\frac{3}{13} < \gamma < \frac{2}{3}$ *and* $n/T^{2-2\gamma} = o(1)$ *hold and that Assumption 1 holds. Under joint asymptotics* $(n, T) \to \infty$,

$$\sqrt{n_g}\left(\rho_{gn}^0\right)^T\left(\widehat{\alpha}_{\widehat{g}} - \alpha_g^0\right) \xrightarrow{d} N(0, 2\left(\alpha_g^0\right)),$$

where $\rho_{gn}^0 = \exp\left(\frac{\alpha_g^0}{T^\gamma}\right)$ *for any* $g = 1, 2, \ldots, K^0$.

Based on the modified k-means algorithms, we still obtain the oracle property of the post-classification estimator. Therefore, we can develop reliable explosiveness tests for mixed-roots panels. The significant advantage of the post-classification estimates is that they employ both time-series and cross-sectional asymptotics. This strategy dramatically enhances the inferential power for explosiveness detection.

5 Inference Procedures and Model Selection

As argued in PWY, the recursive ADF test is a successful bubble identification strategy for explosive roots. However, as shown in Homm and Breitung (2012) and PSY, this recursive approach suffers power loss in finite samples. Therefore, the power inefficiency phenomenon in the bubble detection literature motivates the use of a panel data approach, whose main advantage is to obtain a more powerful test by clustering individual time series. In this section, we present two versions of panel Wald statistics for explosiveness detection: one is on the recursive k-means, and the other is on the modified k-means algorithm. This study justifies the consistency of both types of Wald statistics under the alternative hypothesis of explosive roots.

Furthermore, we aim to provide a panel approach to dating bubble phenomena. In particular, we propose a post-classification estimator of the origination date and its limit theory. By comparison, PWY employ recursive right-sided time-series statistics to assess empirical evidence for explosive behaviour. We use a similar recursive algorithm but build the estimator on panel data. Our method can detect the existence of mildly explosive episodes with better accuracy.

Moreover, we provide a BIC-type information criterion to select the correct group number. In practice, the exact number of groups is typically unknown. We assume that K^0 is bounded by upper bound $K_{\max}$ and derive the choice of the group number via BIC function. Our justifications demonstrate that this estimate on K^0 is consistent.

5.1 Inference procedure: explosiveness detection

Under the model (1) with latent memberships, we can estimate σ_g^2 consistently for each $g = 1, 2, ...K^0$. We define the estimator $\widetilde{\sigma}_{\widehat{g}}^2$ as

$$\widetilde{\sigma}_{\widehat{g}}^2 := \frac{1}{2n_{\widehat{g}}T} \sum_{i \in \widehat{G}_{\widehat{g}}} \sum_{t=1}^{T} \widehat{\widetilde{u}}_{it}^2, \tag{13}$$

where $\widehat{\widetilde{u}}_{it} := \widetilde{y}_{it} - \widehat{\rho}_{\widehat{g}n} \widetilde{y}_{i,t-1}$, $n_{\widehat{g}} := \#\widehat{G}_g$, and $\widehat{\rho}_{\widehat{g}n} := \exp\left(\frac{\widehat{\alpha}_{\widehat{g}}}{T^{\gamma}}\right)$. We define $\widehat{\alpha}_{\widehat{g}}$ as any post-classification within estimate based on either recursive k-means or modified k-means. The post-classification estimators on both classifiers are asymptotically equivalent to the infeasible within estimate.

Since we assume homoskedasticity over $i = 1, 2, ..., n$, we have

$$\widetilde{\sigma}^2 := \frac{1}{2n_{\widehat{g}}T} \sum_{i \in \widehat{G}_{\widehat{g}}} \sum_{t=1}^{T} \widehat{\widetilde{u}}_{it}^2, \tag{14}$$

for any $\widehat{g} = 1, 2, ...K^0$. We provide the consistency of both (13) and (14) in the following theorem.

Theorem 5.1 *Suppose Assumption 1 holds and rate restrictions $n/T^{2-2\gamma} = o(1)$ and $\frac{3}{13} < \gamma < \frac{2}{3}$ are valid. For any $g = 1, 2, ..., K^0$, under $(n, T) \to \infty$*

$$\widetilde{\sigma}_{\widehat{g}}^2 \xrightarrow{p} \sigma^2.$$

Consistency of $\widetilde{\sigma}_{\widehat{g}}^2$ is essential for theoretical derivations of inference procedures. Once we have consistent estimators for σ^2, we can properly scale the statistics. Based on (14), we propose the following Wald statistics:

$$W_{\widehat{g}} := \frac{(R\widehat{\rho}_{\widehat{g}n} - r)^2 D_{\widehat{g},nT}}{R^2 \widetilde{\sigma}_{\widehat{g}}^2}, \tag{15}$$

where $D_{\widehat{g},nT} := \sum_{i \in \widehat{G}_g} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}^2$ and $\widehat{g} = 1, 2, ...K^0$. We collect the details of (15) in the following theorem.

Theorem 5.2 *Suppose Assumption 1 holds and rate restrictions $n/T^{2-2\gamma} = o(1)$ and $\frac{3}{13} < \gamma < \frac{2}{3}$ are valid. Under $H_0 : R\rho_{gn}^0 = r$ and $(n, T) \to \infty$,*

$$W_{\widehat{g}} \xrightarrow{d} \chi^2(1),$$

where $\chi^2(1)$ is chi-square distribution with degree of freedom 1.

Under the null hypothesis, Theorem 5.2 provides confidence intervals for post-classification estimates. In the next subsection, extensive numerical simulations verify good probability coverage of (15) statistics.

One well-known application of explosiveness is bubble detection, since the bubble component satisfies the sub-martingale property. Previous detecting approaches consider only a single time series. When the bubble signal (growth rate, etc.) is not strong enough, the powers of the time-series procedure are not satisfactory. Therefore, a joint convergence framework significantly improves inferential powers for explosiveness detection.

Under the data-generating mechanism (1), we provide two consistent model specification tests for explosiveness detection. Under the alternative hypothesis, these statistics are proven to diverge at a faster rate under these model specification tests than under the ADF test. Therefore, our panel statistics are more powerful than the time-series approach. The model specification test is as follows:

$$\widetilde{W}_{\widehat{g}} := \frac{(\widehat{\rho}_{\widehat{g}n} - 1)^2 D_{\widehat{g},nT}}{\widetilde{\sigma}_{\widehat{g}}^2}, \tag{16}$$

where $\widehat{g} = 1, 2, ...K^0$. We assume $\widehat{\rho}_{\widehat{g}n}$ to be any post-classification estimator. The properties of $\widetilde{W}_g$ are provided in the following theorem.

Theorem 5.3 *Suppose Assumption 1 holds and rate restrictions $n/T^{2-2\gamma} = o(1)$ and $\frac{3}{13} < \gamma < \frac{2}{3}$ are valid. Under $(n, T) \to \infty$ and $H_0 : \rho_{gn}^0 = 1$,*

$$\widetilde{W}_{\widehat{g}} \xrightarrow{d} \chi^2(1),$$

where $\chi^2(1)$ is chi-square distribution with degree of freedom 1 and $g = 1, 2, ..., K^0$. Under the alternative $H_1 : \rho_{gn}^0 > 1$,

$$\widetilde{W}_{\widehat{g}} = O_p(\left(\rho_{gn}^0\right)^{2T} n),$$

where $\rho_{gn}^0 = \exp\left(\frac{\alpha_g^0}{T^\gamma}\right)$.

By comparison, the ADF test under the alternative hypothesis of explosive root diverges at the rate $O_p\left((\rho_{gn}^0)^{2T}\right)$. The divergence rate is slower than the test in (16), providing worse inferential powers. This theoretical finding again verifies the usefulness of panel-testing procedures compared to the time-series approach.

5.2 Inference procedure: detection of bubble origination

Obviously, $\widetilde{W}_{\widehat{g}}$ for $\widehat{g} = 1, 2, ..., K^0$ corresponds to the full-sample statistics on explosiveness detection. To accommodate the recursive bubble detection algorithm, we propose the following statistics:

$$\widetilde{W}_{\widehat{g}}(r) := \frac{\left(\widehat{\rho}_{\widehat{g}n}(r) - 1\right)^2 D_{\widehat{g},nT}(r)}{\widetilde{\sigma}_{\widehat{g}}^2(r)}, \tag{17}$$

where $\widehat{\rho}_{\widehat{g}n}(r)$ is the post-classification within estimator of ρ_{gn}^0 based on the first $\tau = [Tr]$ observations in the g-th estimated group, and $\widetilde{\sigma}_{\widehat{g}}^2(r)$ is the estimator of σ_g^2 based on the first $\tau = [Tr]$ observations in the g-th estimated group. $D_{\widehat{g},nT}(r)$ is the sample moment on the first $\tau = [Tr]$ observations in the g-th estimated group. As $\tau = [Tr] \to \infty$ for all $r \in [r_0, 1]$, we have $\widetilde{W}_{\widehat{g}}(r) \to \chi^2(1)$ under the null hypothesis of time-invariant unit root in the g-th estimated group.

Following PWY, we date the origination of an explosive episode as

$$\widehat{r}_g^e := \inf_{s \geq r_0} \left\{ s : \widetilde{W}_g(s) > cv_{\beta_{Tn}} \right\}, \tag{18}$$

and

$$\widehat{r}_{\widehat{g}}^e := \inf_{s \geq r_0} \left\{ s : \widetilde{W}_{\widehat{g}}(s) > cv_{\beta_{Tn}} \right\}, \tag{19}$$

where $cv_{\beta_{Tn}}$ is the right-side $100cv_{\beta_{Tn}}\%$ critical value of the limiting distribution of $\widetilde{W}_{\widehat{g}}$ and $\widetilde{W}_g$ statistics based on $\tau_s = [Ts]$ observations, and β_{Tn} is the size of the one-sided statistics. We allow $\beta_{Tn} \to 0$ as $(n, T) \to \infty$ because, in this event, $cv_{\beta_{Tn}} \to \infty$. This recursive method can apply in the same way to the PWY procedure based on the ADF statistics. Since there is no material change in the use of inference procedures, the analysis follows the same principle.

We model financial exuberance and bubble formulations under the alternative Hypothesis, in which the data-generating mechanism is with transmissions from unit root into explosiveness, as

$$y_{it} = \mu_i (1 - \rho_{it}) + \rho_{it} y_{i,t-1} + u_{it},\ t = 1, 2, ..., T,\ \text{and}\ i = 1, 2, ..., n, \tag{20}$$

where u_{it} is martingale difference sequences$\left(0, \sigma^2\right)$ and the initialization of the process is as $y_{i0} = 0$ for $i = 1, 2, ..., n$. The autoregressive coefficient is as $\rho_{it} = 1 + \frac{c_{it}}{T^\gamma}$ with $c_{it} = c_{1i}\mathbf{1}\left\{t < \tau^e_{g_i}\right\} + c_{2i}\mathbf{1}\left\{t \geq \tau^e_{g_i}\right\}$. Since we assume $\{c_{1i}\}_{i=1}^n = 0$ and $\{c_{2i}\}_{i=1}^n > 0$, model (20) allows for two regimes: a unit root regime and an explosive regime. The system switches from unity to explosiveness at $\tau^e_{g_i}$, for each $i = 1, 2, ..., n$. If individual i does not contain explosive roots, we have $r^e_{g_i} = 1$.

We can add a fixed effect to the model formulation, but this modification has consequences for the asymptotic properties of $\{y_{it}\}_{i=1,t=1}^{n,T}$. If we impose μ_i rather than $\mu_i (1 - \rho_{it})$ as the fixed effect, the signal of the fixed effects dominates the signal of innovations when $t \geq \tau^e_{g_i}$. This property is empirically unrealistic for most economic and financial time series. Therefore, we assume asymptotically diminishing fixed effects with order $O\left(T^{-\gamma}\right)$.

Asymptotically the estimated group membership is identical to the real membership due to the uniform consistency of the modified k-means classifier. Therefore, we can directly denote the estimator on the origination date as $\widehat{r}^e_g$ rather than $\widehat{r}^e_{\widehat{g}}$. Based on $\widehat{r}^e_g$, we establish a limit theory for dating the origination of an explosive root under the null hypothesis.

Theorem 5.4 *Suppose Assumption 1 holds. Under the null hypothesis of no episode of explosiveness for each $g = 1, 2, ..., K^0$, and provided that $cv_{\beta_{Tn}} \to \infty$, the probability of detecting the origination of a bubble using $\widetilde{W}_g$ is zero as $(n, T) \to \infty$, so that*

$$\Pr\left(\widehat{r}^e_g \in [0, 1]\right) \to 0,$$

and

$$\Pr\left(\widehat{r}^e_{\widehat{g}} \in [0, 1]\right) \to 0,$$

where the estimators $\widehat{r}^e_g$, $\widehat{r}^e_{\widehat{g}}$ are from (18) (19) for $g = 1, 2, ..., K^0$.

Next, we show the consistency of the statistics under the alternative hypothesis for $g = 1, 2, ..., K^0$.

Theorem 5.5 *Suppose Assumption 1 holds. If $\frac{1}{cv_{\beta_{Tn}}} + \frac{cv_{\beta_{Tn}}}{P_{Tn}} \to 0$ with $P_{Tn} := nT^{2-\gamma}$, under the alternative hypothesis of an explosive root in the model (20),*

$$\widehat{r}^e_g \xrightarrow{p} r^e_g,$$

and

$$\widehat{r}^{e}_{\widehat{g}} \xrightarrow{p} r^{e}_{g},$$

for each $g = 1, 2, ..., K^0$. *The estimators* $\widehat{r}^{e}_{g}$, $\widehat{r}^{e}_{\widehat{g}}$ *are from (18) (19) for* $g = 1, 2, ..., K^0$.

We provide several remarks for detectors of origination dates. First, when we include observations from the explosive regimes, the signals from explosive roots dominate those from stationary regimes. This is the foundation for bubble detection. Second, under the alternative hypothesis, the statistics diverge at the rate of $O(P_{Tn})$. Once the critical value $cv_{\beta_{Tn}}$ increases at a slower rate than P_{Tn}, we can justify the consistency of the origination detector. Third, in the recursive algorithm for origination detection, we estimate model (20) repeatedly, using subsets of the sample data incremented by one observation at each pass for every individual $i \in \{1, 2, ..., n\}$. Under the null hypothesis of no bubble regime for the group $g \in \left\{1, 2, ..., K^0\right\}$, the supreme statistics have the following limiting distribution

$$\sup_{r\in[r_0,1]} \widetilde{W}_g(r) \xrightarrow{d} \chi^2(1),$$

where $(n, T) \to \infty$. Unlike the single time-series case in which the supreme statistics follow the limiting distribution comprised of stochastic integrals, and the proposed supreme statistics based on panel autoregression follow a pivotal chi-square distribution. The pivotal distribution means that no additional computational burdens are required for critical values. Therefore, our panel detector is easier to implement by empirical researchers on macroeconomics and finance markets.

5.3 Model sections: choosing K^0

To estimate the true number of groups K^0 consistently, we rely on the following information criterion:

$$\mathrm{IC}(K) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\widetilde{y}_{it} - \widetilde{y}_{i,t-1}\widehat{\rho}^{(K)}_{\widehat{g}_i n}\right)^2 + K h_{nT},$$

where $\left(\widehat{\rho}^{(K)}_{\widehat{g}_i n}\right)_{i=1}^{n}$ is the post-classification estimator assuming K groups. Post-classification estimators based on both recursive k-means and modified k-means are applicable. The estimated number of groups follows

$$\widehat{K} := \arg\min_{K=1,2,...,K_{\max}} \mathrm{IC}(K), \tag{21}$$

where $K_{\max}$ is a generic upper bound on K. The information criterion follows Bai and Ng (2002) and Bai (2003, 2009). The estimated number of groups $\widehat{K}$ is consistent for K^0 under the joint convergence framework $(n, T) \to \infty$.

Specifically, we follow the BIC in Bonhomme and Manresa (2016),

$$\mathrm{BIC}(K) = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\widetilde{y}_{it} - \widetilde{y}_{i,t-1}\widehat{\rho}^{(K)}_{\widehat{g}_i n}\right)^2 + \widetilde{\sigma}^2\frac{K+n}{nT}\ln(nT),$$

where $\widetilde{\sigma}^2 := \widetilde{\sigma}^2_{g(K)}$ for $g = 1, 2, ..., K$. The BIC represents a balance between model fitness and penalty of over-fitness. We demonstrate the consistency of this information criterion in the following theorem.

Theorem 5.6 *Suppose Assumption 1 holds and rate restrictions $T^{2-2\gamma} = o(n)$ and $\frac{3}{13} < \gamma < \frac{2}{3}$ are valid. Under $(n, T) \to \infty$,*

$$\min_{1 \leq K^* < K} \inf_{g(K^*)} \widetilde{\sigma}^2_{g(K^*)} \xrightarrow{p} \underline{\sigma}^2 > \sigma^2$$

for $g = 1, 2, ..., K^0$. With joint convergence of $(n, T) \to \infty$, we have

$$\widehat{K} \xrightarrow{p} K^0$$

where $\widehat{K}$ is derived by (21).

6 Monte Carlo Analysis

We present extensive simulation results on both the recursive k-means algorithm and the modified k-means classification algorithm. We also report numerical performances of the recursive detector for bubble origination using panel data.

6.1 Recursive k-means algorithm

Extensive numerical simulations are conducted to evaluate the empirical performance of the recursive k-means classification and its inference approaches. To verify classification accuracy, we consider only four data-generating processes on explosive roots. The observations in each data-generating mechanism are selected from three groups with equal proportion $n_1 : n_2 : n_3 = \frac{1}{3} : \frac{1}{3} : \frac{1}{3}$ or two groups with $n_1 : n_2 = \frac{1}{2} : \frac{1}{2}$. We choose cross-sectional dimension $n = 60, 120$ and time dimension $T = 150, 250, 350$. Here, the fixed effect $\{\mu_i\}_{i=1}^n$ is drawn from $N(0, 1)$ and innovation processes $\{u_{it}\}_{t=1}^T$ for each $i \in \{1, 2, ..., n\}$ follow independent and identically distributed normality $N(0, 2)$. Noise process $\{u_{it}\}_{t=1}^T$ is independent across $i \in \{1, 2, ..., n\}$. We have the following data-generating mechanisms: DGP1 ($c = [0.5, 1.5, 3]$, $\gamma = 0.6$); DGP2 ($c = [0.2, 0.7, 1.4]$, $\gamma = 0.4$); DGP3 ($c = [0.2, 0.5]$, $\gamma = 0.4$); and DGP3 ($c = [1, 2]$, $\gamma = 0.6$). We run 1,000 replications for each data-generating mechanism.

First, we run numerical simulations to verify the asymptotic validity of our proposed BIC procedure. We pick up the set of K^0 that minimizes the BIC objective function. Table 1 demonstrates the empirical frequency that K^0 is selected from 1 across 5 by applying the recursive k-means classification algorithm under DGP1~4. When T is larger than 150, the BIC objective function can choose the real group number with almost no errors.

Second, with correctly selected group numbers, we show the classification consistency of the recursive k-means procedure and provide the numerical results of its post-classification estimator. Tables 2~5 show the classification errors, RMSE, asymptotic bias, and probability coverage of the post-classification estimator and its oracle estimator, the infeasible within estimator when informed of the true group membership. Here, we provide several key definitions before detailed discussions. The classification error follows

$$\frac{1}{n}\sum_{g=1}^{K^0}\sum_{i\in\widehat{G}_g}\mathbf{1}\left\{\widehat{g}_i\neq g_i^0\right\}.$$

The RMSE is the square root of mean-squared error on the post-classification estimator. Bias is the averaged difference between our estimator and the true parameter. The probability coverage follows

$$\frac{1}{n}\sum_{g=1}^{K^0}\sum_{i\in\widehat{G}_g}\mathbf{1}\left\{\left|\frac{\widehat{c}_i-c_i^0}{\widetilde{\sigma}}\sqrt{\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\widetilde{y}_{i,t-1}\right)^2}\right|<1.96\right\},$$

where $\widetilde{\sigma}$ is the proposed consistent estimator for the standard deviations of the error process. As one comparison, we show the results for the oracle within estimator. Classification errors and probability coverage of oracle estimators similarly follow the abovementioned definitions. We only replace $\widehat{G}_g$ with G_g^0, since we can observe the true membership.

Tables 2~5 show extensive discussions on explosive roots. As shown here, the classification error approaches zero as the time horizon increases, and the RMSE and bias of the oracle estimator are smaller than the post-classification estimator. For post-classification estimators, the RMSE and bias generally decrease when $T\to\infty$. With $T>150$, the asymptotic difference between the oracle estimator and the post-classification estimator is almost negligible. The diminishing distance verifies the asymptotic equivalence between these two estimators. The diminishing distance is due to the uniform consistency of our recursive k-means classification technology. The recursive k-means induces much better finite sample performance than the estimation and classification results of the modified k-means algorithm shown later. This finding is not surprising, as the increase of the cross-sectional dimension helps improve the accuracy of the recursive algorithm while magnifying the classification error of the modified k-means approach. The main explanation for this phenomenon is that once we have larger n, the uniform convergence rate of individual least squares estimators slows down.

Last, we evaluate the performance of explosive detection statistics. The nominal level is set to 5%. We evaluate these tests with the correct number of groups of both $K^0=2$ and $K^0=3$. To demonstrate the superiority over the ADF test, we choose $n=30,60,90$ and $T=50,100,150$. With two groups, we assume $\pi_1=\pi_2=\frac{1}{2}$. With three groups, we assume $\pi_1=\pi_2=\pi_3=\frac{1}{3}$. We conduct all simulations with 1,000 iterations. We present the results in Tables 6 and 7.

The sizes of panel explosiveness tests are well controlled around nominal levels. The power of the test is greatly improved by the additional degree of cross-sectional asymptotics, n, when compared with the ADF statistics. The power improvement corresponds to our aim to create power-enhanced statistics for explosiveness detection. Based on the panel Wald statistics, we can construct a recursive real-time detector, like PWY.

6.2 Modified k-means classification

Extensive numerical simulations are conducted to evaluate the empirical performance of modified k-means classification and its inference approaches. For the classification and estimation, we consider seven data-generating processes that cover both explosive and stationary roots. We select observations in each data-generating mechanism from three groups with equal proportion $n_1 : n_2 : n_3 = \frac{1}{3} : \frac{1}{3} : \frac{1}{3}$ or two groups with $n_1 : n_2 = \frac{1}{2} : \frac{1}{2}$. We choose cross-sectional dimension $n = 60, 120$ and time dimension $T = 150, 250, 350$. For all seven processes, the fixed effect $\{\alpha_i\}_{i=1}^n$ is drawn from $N(0, 1)$ and innovation processes $\{u_{it}\}_{t=1}^T$ for each $i \in \{1, 2, ..., n\}$ follow independent and identically distributed normality $N(0, 2)$. The error processes $\{u_{it}\}_{t=1}^T$ are independent across $i \in \{1, 2, ..., n\}$.

The seven data-generating mechanisms are described as: DGP1 ($c = [0.5, 1.5, 3]$, $\gamma = 0.6$); DGP2 ($c = [0.2, 0.7, 1.4]$, $\gamma = 0.4$); DGP3 ($c = [0.2, 0.5]$, $\gamma = 0.4$); DGP3 ($c = [1, 2]$, $\gamma = 0.6$); DGP-A ($c = [-4, -2, 0.5]$, $\gamma = 0.6$); DGP-B ($c = [-4, -2, 1]$, $\gamma = 0.6$); and DGP-C ($c = [-1, -1]$, $\gamma = 0.4$). We run 1,000 replications for each data-generating mechanism.

The consistency of classifications and estimations relies heavily on correctly determining the exact group numbers, and the following numerical simulations verify the asymptotic behaviour of our proposed BIC procedure. We pick up the set of K^0 that minimizes the BIC objective function. Tables 8 and 9 demonstrate the performance of the identical information criterion under the modified k-means classification algorithm. Table 8 provides the pure explosive roots, while Table 9 presents cases accommodating both explosive roots and stationary roots. Under pure explosive roots and mixed roots cases, we observe rare selection errors in choosing group numbers when time observations are 150. When T is larger than 250, the BIC objective function can select the real group number with no classification errors. The mixed root cases have fewer selection errors than the pure explosive case. This result is not surprising, since adjustment rates for stationary and explosive roots are different and provide a better reference for differentiation.

Second, with correctly selected group numbers, we next show the classification consistency of the modified k-means procedure and provide the numerical results of the post-classification estimator based on the modified k-means procedure. Tables 10~16 show the classification errors, RMSE, asymptotic bias, and probability coverage of the post-classification estimator and its oracle estimator.

Tables 10~13 show extensive results on explosive roots. Tables 14~16 provide results on mixed-roots cases, in which both explosive and stationary roots appear. As illustrated in the numerical simulations, the classification errors approach zero as the time horizon increases, and the RMSE and bias of the oracle estimator are much smaller than those of the post-classification estimator based on the modified k-means classifications. For post-classification estimators, the RMSE and bias generally decrease and get closer to the RMSE and bias of the oracle estimators as $T \to \infty$. The decreasing RMSE and bias demonstrate that as the time horizon diverges, the asymptotic difference between the post-classification and oracle estimators is asymptotically diminishing. The diminishing distance is due to the uniform consistency of our modified k-means classification technology. Unfortunately, since classification errors of the modified k-means algorithm diminish at a slower rate than the recursive k-means algorithm, the RMSE and bias of post-classification estimators are reduced more slowly.

Next, we investigate the performance of the Wald statistics to detect the existence of explosive roots. The significance level is set to 5%. We evaluate these tests with the correct number of groups of both $K^0 = 2$ and $K^0 = 3$. We examine the performance of the proposed tests under a pure explosive-root case and a mixture case of both explosive and stationary roots. The sample sizes over the cross-sectional dimension and time horizon are chosen as $n = 30, 60, 90$ and $T = 50, 100, 150$, respectively. With two groups, we assume $\pi_1 = \pi_2 = \frac{1}{2}$. With three groups, we assume $\pi_1 = \pi_2 = \pi_3 = \frac{1}{3}$. We conduct all simulations with 1,000 iterations. We present the detailed results in Tables 17~21.

The sizes of the proposed panel Wald test are well controlled under both mixed roots and pure explosive roots. One exception is for the mixed roots case in Table 19, in which we observe mild size distortions for the panel statistics when $n \leq 30$ and $T \leq 150$. However, when the sample size increases, such phenomena quickly disappear. As illustrated here, empirical rejection frequency under the null hypothesis is very close to the nominal level when $T \geq 150$ and $n \geq 30$.

The most interesting observations are the significant power improvements brought by an additional degree of cross-sectional asymptotics. For example, in Tables 17~21, under the alternative hypothesis of the distance parameter as small as 0.3, the rejection rate of the Wald test is almost unity when $n = 10$ and $T = 100$. However, the counterpart, the ADF test, is able to detect explosive roots only with around a 50% chance. The power deficiency shows the high priority of the panel Wald test on explosiveness identifications. The increase of cross-sectional dimension may reduce the classification accuracy, since with the larger n, we have a smaller uniform convergence rate δ_{nT} and a larger classification error. However, the power enhancement of cross-sectional asymptotics strongly dominates the classification errors and contributes to the superiority of the proposed Wald statistics.

6.3 Detection of bubble origination

This subsection reports some brief simulations examining the performance of the dating estimation procedure and the accuracy of the asymptotic theory. We employ the panel recursive dating detector $\widehat{r}^e_{\widehat{g}}$ based on $\widetilde{W}_{\widehat{g}}$ for $\widehat{g} = 1, 2, ..., K^0$. We also conduct the time-series recursive detector $\widehat{r}^e$ proposed in PWY. We conduct the numerical experiments using 500 sample path replications. The observations are selected from three groups with equal proportion $n_1 : n_2 : n_3 = \frac{1}{3} : \frac{1}{3} : \frac{1}{3}$. We choose cross-sectional dimension $n = 60, 90$ and time dimension $T = 150, 250, 350$. The fixed effect $\{\mu_i\}_{i=1}^n$ is drawn from $N(0,1)$ and innovation processes $\{u_{it}\}_{t=1}^T$ for each $i \in \{1, 2, ..., n\}$ follow independent and identically distributed normality $N(0,2)$ or student's t distribution with degree of freedom 3 and variance 2. The noise processes $\{u_{it}\}_{t=1}^T$ are independent across $i \in \{1, 2, ..., n\}$. We have the following data-generating mechanisms: $c_1 = 0$, $c_2 = -3$, $c_3 = -9$, $\gamma = 0.5$ when

$r \le r_g^e$, and $c_1 = 1, 2, 3, 4$, $c_2 = -3$, $c_3 = -9$, $\gamma = 0.5$ when $r > r_g^e$. In all cases, we set the true values of the origination date as $r_g^e = 0.5$ for some $g = 1$. We use critical values as $\log(\log(nT))$ for the panel recursive algorithm and $\log(\log(T))$ for the time-series recursive algorithm proposed in PWY.

Tables 22~25 report results for both the panel recursive algorithm and the time-series recursive algorithm, giving means, standard errors, and RMSE for $\widehat{r}_{\widehat{g}}^e$. We can observe the following five patterns. First, both panel recursive detector $\widehat{r}_{\widehat{g}}^e$ and time-series recursive detector $\widehat{r}^e$ can estimate the true bubble origination date with high accuracy, reflected by small bias and a small standard error. When $T = 350$, for cases, the means of $\widehat{r}_{\widehat{g}}^e$ and $\widehat{r}^e$ are almost 0.5, the true value, with small standard errors. Second, panel recursive detector $\widehat{r}_{\widehat{g}}^e$ converges to the true value faster than the time-series detector $\widehat{r}^e$. For almost all cases, the means of $\widehat{r}_{\widehat{g}}^e$ are closer to 0.5 with smaller standard errors and RMSE. Especially when the bubble signal is small ($c_1 = 1$) or the bubble period is short ($T = 150$), the bias of $\widehat{r}_{\widehat{g}}^e$ is much smaller than 0.02 while the bias of $\widehat{r}^e$ can be bigger than 0.05. Third, when the explosive signal (c_1) is stronger, it is easier to estimate the true origination date for both detectors. In this case, bias, standard error, and RMSE of both recursive detectors are smaller. Fourth, when the sample size increases, it is easier to estimate r^e. Both the bias and standard errors become smaller, corroborating the consistency results. Specifically, the consistency of the panel recursive detector benefits from the increase in both n and T. However, the time-series recursive detector benefits only from the increase in T. Last, both the panel recursive detector and the time-series recursive detector are robust to heavy-tail innovations. From Tables 24 and 25, the consistency of both the panel recursive detector and the time-series recursive detector is obvious on the student's t distribution with degree of freedom 3.

7 Empirical Illustrations

We take housing data from Fang et al. (2016). We employ the monthly real estate price index (PI) on China's large and medium-sized cities. Our sample covers the period March 2003 to March 2013, and is comprised of 123 monthly observations from 113 cities. We apply the modified k-means classifications for the data set to study the persistence of housing prices.

The BIC objective function suggests $K^0 = 4$ in Table 26. With the known group number, $K^0 = 4$, the modified k-means algorithm provides the classified group structure in Table 27. Moreover, we show the time series plots for Groups 1–4 in Figure 1.

Table 27 reports the post-classification estimates and explosiveness statistics for each group. The housing prices in the second group are the most persistent, with the largest slope. We perceive explosive roots in Groups 2 and 4, while Groups 1 and 3 demonstrate stationary roots.

Furthermore, we apply the panel recursive detector of bubble origination dates to Groups 1~4. We present the results in Figure 2. For Groups 1 and 3, the statistics are below the critical value, illustrating that we cannot reject the null hypothesis of unit root behaviour. In Group 2, the supreme statistics are larger than the critical value after September 2009. In Group 4, the supreme statistics are larger than the critical value after February 2011. These observations show the existence of bubbles and the starting dates of bubbles at the same time.

To demonstrate the robustness of post-classification procedures, we apply the time-series recursive detector in PWY. We select four cities from Group 1, four cities from Group 2, two cities from Group 3, and two cities from Group 4. We plot the sequences of supreme statistics for the selected cities in Figures 3–5. The time-series supreme statistics verify the existence of bubbles in Groups 2 and 4, coinciding with the panel recursive detectors.

8 Conclusions

Explosiveness roots exist because of asset bubbles in various financial markets. These roots can differ across individuals. Therefore, these explosive trajectories can be heterogeneous, and any neglect of group-specific heterogeneity may lead to inconsistency of estimations and imprecision of related inference procedures.

In this paper, we propose a mixed-roots panel autoregression model and the relevant clustering algorithm. We model individual heterogeneity through latent group structures. The group patterns here represent the homogeneous slope coefficient within the identical group and heterogeneous autoregression coefficients across different groups. Under explosive panel autoregressions with latent group structures, we apply the recursive k-means classification, and illustrate the uniform consistency of the group clustering algorithm. Similarly, within the mixed-roots panel autoregressions, a modified k-means clustering algorithm is applied with uniform consistency of classification. With these two group classifiers, we can furthermore build two post-classification within estimators. Both post-classification estimators are asymptotically equivalent to the oracle estimators. Based on post-classification estimates, we provide two consistent model specification tests on explosiveness detection. From Monte Carlo simulations, we perceive that both classifiers work well, and the recursive algorithm tends to induce fewer misclassifications compared to the modified k-means algorithm. Besides, model specification tests based on two k-means clustering algorithms provide significant power enhancement compared to ADF statistics. The real number of latent groups can be consistently estimated using BIC information criteria. This framework amplifies the panel data modeling, and firstly introduces explosiveness into the literature.

APPENDIX

A Monte Carlo Analysis

This section displays simulated results for both recursive k-means and modified k-means algorithms. This section also demonstrates the simulated performance of the panel recursive algorithm for bubble detections.

A.1 Recursive k-means algorithm

We display the simulated results for the recursive k-means algorithm.

DGP1						
N	T	K=1	K=2	K=3	K=4	K=5
60	150	0	0	0.998	0.002	0
60	250	0	0	1	0	0
60	350	0	0	1	0	0
120	150	0	0	0.998	0.002	0
120	250	0	0	1	0	0
120	350	0	0	1	0	0
DGP2						
N	T	K=1	K=2	K=3	K=4	K=5
60	150	0	0	1	0	0
60	250	0	0	1	0	0
60	350	0	0	1	0	0
120	150	0	0	1	0	0
120	250	0	0	1	0	0
120	350	0	0	1	0	0
DGP3						
N	T	K=1	K=2	K=3	K=4	K=5
60	150	0	0.994	0.003	0.002	0.001
60	250	0	0.999	0.001	0	0
60	350	0	1	0	0	0
120	150	0	0.996	0.002	0.001	0
120	250	0	1	0	0	0
120	350	0	1	0	0	0
DGP4						
N	T	K=1	K=2	K=3	K=4	K=5
60	150	0	0.996	0.002	0.001	0.001
60	250	0	1	0	0	0
60	350	0	1	0	0	0
120	150	0	0.996	0.002	0.001	0.001
120	250	0	1	0	0	0
120	350	0	1	0	0	0

Table 1: Empirical Frequency of BIC by Applying Recursive K-means under DGP 1,2,3,4

	DGP 1			Post-Classification			Oracle		
	N	T	Cluster Error	RMSE	Bias	Coverage	RMSE	Bias	Coverage
Group 1	60	150	0.0009	0.1011	-0.0052	0.931	0.0058	-0.0029	0.9313
	60	250	0.0006	0.0794	-0.0039	0.909	0.003	-0.0004	0.9087
	60	350	0.0003	0.0703	-0.0036	0.962	0.0013	-0.0003	0.9618
	120	150	0.0003	0.0667	-0.0018	0.964	0.0041	-0.0003	0.9646
	120	250	0.0002	0.0636	-0.0024	0.914	0.002	-0.0002	0.9141
	120	350	0.0001	0.0636	-0.0032	0.95	0.0001	-0.0001	0.9496
Group 2	60	150	0.0009	0	0	0.931	0	0	0.931
	60	250	0.0006	0	0	0.935	0	0	0.935
	60	350	0.0003	0	0	0.949	0	0	0.949
	120	150	0.0003	0	0	0.943	0	0	0.943
	120	250	0.0002	0	0	0.913	0	0	0.913
	120	350	0.0001	0	0	0.918	0	0	0.918
Group 3	60	150	0.0009	0	0	0.943	0	0	0.943
	60	250	0.0006	0	0	0.925	0	0	0.925
	60	350	0.0003	0	0	0.947	0	0	0.947
	120	150	0.0003	0	0	0.96	0	0	0.96
	120	250	0.0002	0	0	0.928	0	0	0.928
	120	350	0.0001	0	0	0.904	0	0	0.904

Table 2: Classification and Estimation by Recursive K-means Algorithm under DGP 1

	DGP 2			Post-Classification			Oracle		
	N	T	Cluster Error	RMSE	Bias	Coverage	RMSE	Bias	Coverage
Group 1	60	150	0	0.0017	0.0001	0.941	0.0017	0.0001	0.941
	60	250	0	0.0004	0	0.91	0.0004	0	0.91
	60	350	0	0.0001	0	0.951	0.0001	0	0.951
	120	150	0	0.0012	0	0.968	0.0012	0	0.968
	120	250	0	0.0003	0	0.903	0.0003	0	0.903
	120	350	0	0.0001	0	0.946	0.0001	0	0.946
Group 2	60	150	0	0	0	0.933	0	0	0.933
	60	250	0	0	0	0.938	0	0	0.938
	60	350	0	0	0	0.954	0	0	0.954
	120	150	0	0	0	0.946	0	0	0.946
	120	250	0	0	0	0.929	0	0	0.929
	120	350	0	0	0	0.92	0	0	0.92
Group 3	60	150	0	0	0	0.958	0	0	0.958
	60	250	0	0	0	0.943	0	0	0.943
	60	350	0	0	0	0.952	0	0	0.952
	120	150	0	0	0	0.965	0	0	0.965
	120	250	0	0	0	0.961	0	0	0.961
	120	350	0	0	0	0.956	0	0	0.956

Table 3: Classification and Estimation by Recursive K-means Algorithm under DGP 2

DGP 3				Post-Classification			Oracle		
	N	T	Cluster Error	RMSE	Bias	Coverage	RMSE	Bias	Coverage
Group 1	60	150	0.0013	0.0387	-0.0023	0.937	0.0017	0.0001	0.938
	60	250	0.0001	0.0236	-0.0014	0.907	0.0004	-0.0001	0.9042
	60	350	0	0.0001	0	0.951	0.0001	0	0.951
	120	150	0.0003	0.0227	-0.0006	0.967	0.0012	0.0001	0.9672
	120	250	0.0001	0.0201	-0.0001	0.901	0.0003	0	0.8978
	120	350	0	0.0001	0	0.946	0.0001	0	0.946
Group 2	60	150	0.0013	0	0	0.932	0	0	0.932
	60	250	0.0001	0	0	0.936	0	0	0.936
	60	350	0	0	0	0.952	0	0	0.952
	120	150	0.0003	0	0	0.937	0	0	0.937
	120	250	0.0001	0	0	0.913	0	0	0.913
	120	350	0	0	0	0.915	0	0	0.915

Table 4: Classification and Estimation by Recursive K-means Algorithm under DGP 3

DGP 4				Post-Classification			Oracle		
	N	T	Cluster Error	RMSE	Bias	Coverage	RMSE	Bias	Coverage
Group 1	60	150	0.0012	0.0966	-0.006	0.937	0.0003	0	0.9344
	60	250	0	0.0001	0	0.918	0.0001	0	0.938
	60	350	0	0	0	0.952	0	0	0.952
	120	150	0.0013	0.0966	-0.006	0.94	0.0002	0	0.9379
	120	250	0	0.0001	0	0.913	0.0001	0	0.913
	120	350	0	0	0	0.93	0	0	0.93
Group 2	60	150	0.0012	0	0	0.935	0	0	0.935
	60	250	0	0	0	0.938	0	0	0.938
	60	350	0	0	0	0.953	0	0	0.953
	120	150	0.0013	0	0	0.95	0	0	0.95
	120	250	0	0	0	0.924	0	0	0.924
	120	350	0	0	0	0.92	0	0	0.92

Table 5: Classification and Estimation by Recursive K-means Algorithm under DGP 4

N	Statistics		Size	
c1=0(c2=2,c3=4),gamma=0.5		T=50	T=100	T=150
1	ADF	0.054	0.068	0.082
10(30)	Post-Classification	0.044	0.054	0.042
20(60)	Post-Classification	0.052	0.048	0.052
30(90)	Post-Classification	0.056	0.038	0.054
N	Statistics		Power	
c1=0.3(c2=2,c3=4),gamma=0.5		T=50	T=100	T=150
1	ADF	0.566	0.774	0.884
10(30)	Post-Classification	1	1	1
20(60)	Post-Classification	1	1	1
30(90)	Post-Classification	1	1	1
N	Statistics		Power	
c1=0.5(c2=2,c3=4),gamma=0.5		T=50	T=100	T=150
1	ADF	0.828	0.958	0.986
10(30)	Post-Classification	1	1	1
20(60)	Post-Classification	1	1	1
30(90)	Post-Classification	1	1	1

Table 6: Sizes and Powers of ADF and Panel Wald Tests on Explosive Roots Detection (c1=0, 0.3, 0.5; c2=2, c3=4; gamma=0.5)

N	Statistics		Size	
c1=0(c2=2.5),gamma=0.6		T=50	T=100	T=150
1	ADF	0.054	0.068	0.082
10(30)	Post-Classification	0.044	0.054	0.042
20(60)	Post-Classification	0.052	0.048	0.052
30(90)	Post-Classification	0.056	0.038	0.054
N	Statistics		Power	
c1=0.3(c2=2.5),gamma=0.6		T=50	T=100	T=150
1	ADF	0.338	0.514	0.628
10(30)	Post-Classification	0.95	0.998	1
20(60)	Post-Classification	1	1	1
30(90)	Post-Classification	1	1	1
N	Statistics		Power	
c1=0.5(c2=2.5),gamma=0.6		T=50	T=100	T=150
1	ADF	0.63	0.816	0.888
10(30)	Post-Classification	1	1	1
20(60)	Post-Classification	1	1	1
30(90)	Post-Classification	1	1	1

Table 7: Sizes and Powers of ADF and Panel Wald Tests on Explosive Roots Detection (c1=0, 0.3, 0.5; c2=2.5; gamma=0.6)

A.2 Modified k-means algorithm

We display the simulated results for the modified k-means algorithm.

DGP1						
N	T	K=1	K=2	K=3	K=4	K=5
60	150	0	0.001	0.983	0.011	0.005
60	250	0	0	1	0	0
60	350	0	0	1	0	0
120	150	0	0	0.983	0.012	0.005
120	250	0	0	1	0	0
120	350	0	0	1	0	0
DGP2						
N	T	K=1	K=2	K=3	K=4	K=5
60	150	0	0	0.99	0.008	0.002
60	250	0	0	1	0	0
60	350	0	0	1	0	0
120	150	0	0	0.995	0.005	0
120	250	0	0	1	0	0
120	350	0	0	1	0	0
DGP3						
N	T	K=1	K=2	K=3	K=4	K=5
60	150	0	0.997	0.003	0	0
60	250	0	1	0	0	0
60	350	0	1	0	0	0
120	150	0	0.996	0.004	0	0
120	250	0	1	0	0	0
120	350	0	1	0	0	0
DGP4						
N	T	K=1	K=2	K=3	K=4	K=5
60	150	0	0.999	0.001	0	0
60	250	0	1	0	0	0
60	350	0	1	0	0	0
120	150	0	1	0	0	0
120	250	0	1	0	0	0
120	350	0	1	0	0	0

Table 8: Empirical Frequency of BIC by Applying Modified K-means under DGP 1,2,3,4

.

DGP-A						
N	T	K=1	K=2	K=3	K=4	K=5
60	150	0	0	0.999	0.001	0
60	250	0	0	1	0	0
60	350	0	0	1	0	0
120	150	0	0	1	0	0
120	250	0	0	1	0	0
120	350	0	0	1	0	0
DGP-B						
N	T	K=1	K=2	K=3	K=4	K=5
60	150	0	0	0.999	0.001	0
60	250	0	0	1	0	0
60	350	0	0	1	0	0
120	150	0	0	0.999	0.001	0
120	250	0	0	1	0	0
120	350	0	0	1	0	0
DGP-C						
N	T	K=1	K=2	K=3	K=4	K=5
60	150	0	1	0	0	0
60	250	0	1	0	0	0
60	350	0	1	0	0	0
120	150	0	1	0	0	0
120	250	0	1	0	0	0
120	350	0	1	0	0	0

Table 9: Empirical Frequency of BIC by Applying Modified K-means under DGP-A,-B,-C

	DGP 1			Post-Classification			Oracle		
	N	T	Cluster Error	RMSE	Bias	Coverage	RMSE	Bias	Coverage
Group 1	60	150	0.0169	0.2232	0.0215	0.899	0.0065	-0.0027	0.8922
	60	250	0.0096	0.1411	0.0132	0.928	0.0027	-0.0005	0.9315
	60	350	0.0081	0.1265	0.0159	0.932	0.0014	-0.0001	0.9252
	120	150	0.0174	0.2117	0.0216	0.908	0.0044	-0.0015	0.909
	120	250	0.0122	0.1556	0.0179	0.919	0.0018	-0.0002	0.9117
	120	350	0.0063	0.1265	0.016	0.928	0.0009	-0.0001	0.923
Group 2	60	150	0.0169	3.8867	-0.6829	0.95	0	0	0.9502
	60	250	0.0096	3.6665	-0.4499	0.943	0	0	0.9471
	60	350	0.0081	4.4413	-0.5587	0.945	0	0	0.9469
	120	150	0.0174	3.8267	-0.6551	0.95	0	0	0.9508
	120	250	0.0122	4.1986	-0.6007	0.953	0	0	0.953
	120	350	0.0063	4.4413	-0.5587	0.952	0	0	0.9487
Group 3	60	150	0.0169	2.0763	-0.1857	0.953	0	0	0.953
	60	250	0.0096	2.8901	-0.2742	0.951	0	0	0.951
	60	350	0.0081	1.6372	-0.0732	0.945	0	0	0.945
	120	150	0.0174	2.543	-0.2786	0.961	0	0	0.961
	120	250	0.0122	2.1541	-0.1523	0.949	0	0	0.949
	120	350	0.0063	1.6372	-0.0732	0.942	0	0	0.942

Table 10: Classification and Estimation by Modified K-means Algorithm under DGP 1

	DGP 2			Post-Classification			Oracle		
	N	T	Cluster Error	RMSE	Bias	Coverage	RMSE	Bias	Coverage
Group 1	60	150	0.0376	0.1629	0.0088	0.866	0.009	-0.0085	0.8685
	60	250	0.0163	0.0896	0.0097	0.92	0.0042	-0.0028	0.9223
	60	350	0.0113	0.0867	0.006	0.928	0.0024	-0.0023	0.9312
	120	150	0.0339	0.1278	0.0156	0.88	0.0062	-0.0046	0.8838
	120	250	0.0143	0.0851	0.0122	0.92	0.0029	-0.0017	0.9191
	120	350	0.0099	0.0749	0.0085	0.925	0.0016	-0.0009	0.9229
Group 2	60	150	0.0376	2.8978	-0.625	0.919	0.0001	0	0.9159
	60	250	0.0163	2.6637	-0.4195	0.935	0	0	0.9381
	60	350	0.0113	2.6761	-0.3625	0.942	0	0	0.9398
	120	150	0.0339	2.8391	-0.6012	0.938	0.0001	0	0.9396
	120	250	0.0143	2.7637	-0.456	0.945	0	0	0.9483
	120	350	0.0099	2.676	-0.3632	0.951	0	0	0.9475
Group 3	60	150	0.0376	2.4413	-0.4367	0.956	0	0	0.956
	60	250	0.0163	2.0365	-0.241	0.948	0	0	0.948
	60	350	0.0113	1.9076	-0.181	0.954	0	0	0.954
	120	150	0.0339	2.3241	-0.3958	0.957	0	0	0.957
	120	250	0.0143	1.6328	-0.1549	0.949	0	0	0.949
	120	350	0.0099	1.7985	-0.1609	0.941	0	0	0.941

Table 11: Classification and Estimation by Modified K-means Algorithm under DGP 2

DGP 3				Post-Classification			Oracle		
	N	T	Cluster Error	RMSE	Bias	Coverage	RMSE	Bias	Coverage
Group 1	60	150	0.0237	0.0873	0.012	0.872	0.0019	-0.0006	0.8688
	60	250	0.0103	0.0503	0.0083	0.918	0.0004	0	0.9205
	60	350	0.0018	0.0212	0.0015	0.937	0.0001	0	0.9391
	120	150	0.0262	0.0842	0.0172	0.873	0.0013	-0.0003	0.8727
	120	250	0.0066	0.0414	0.0057	0.935	0.0029	0	0.9343
	120	350	0.0048	0.0355	0.0042	0.924	0.0009	0	0.9172
Group 2	60	150	0.0237	1.8743	-0.4436	0.957	0	0	0.957
	60	250	0.0103	1.6069	-0.2689	0.952	0	0	0.952
	60	350	0.0018	0.7718	-0.0546	0.95	0	0	0.95
	120	150	0.0262	2.0348	-0.5228	0.963	0	0	0.963
	120	250	0.0066	1.3237	-0.1825	0.953	0	0	0.953
	120	350	0.0048	1.2914	-0.1528	0.952	0	0	0.952

Table 12: Classification and Estimation by Modified K-means Algorithm under DGP 3

DGP 4				Post-Classification			Oracle		
	N	T	Cluster Error	RMSE	Bias	Coverage	RMSE	Bias	Coverage
Group 1	60	150	0.0056	0.1265	0.016	0.915	0.0038	0	0.9139
	60	250	0.0007	0.0447	0.002	0.945	0.0007	0	0.9442
	60	350	0.0004	0.0316	0.001	0.939	0.0001	0	0.939
	120	150	0.0038	0.1049	0.011	0.941	0.0026	0	0.9384
	120	250	0.0003	0.0316	0.001	0.954	0.0005	0	0.955
	120	350	0.0002	0.0315	0.001	0.936	0.0001	0	0.9367
Group 2	60	150	0.0056	2.8099	-0.3554	0.953	0	0	0.953
	60	250	0.0007	1.3177	-0.0589	0.953	0	0	0.953
	60	350	0.0004	1.126	-0.0356	0.946	0	0	0.946
	120	150	0.0038	2.3298	-0.2444	0.956	0	0	0.956
	120	250	0.0003	0.9317	-0.0295	0.964	0	0	0.964
	120	350	0.0002	1.126	-0.0356	0.952	0	0	0.952

Table 13: Classification and Estimation by Modified K-means Algorithm under DGP 4

DGP-A				Post-Classification			Oracle		
	N	T	Cluster Error	RMSE	Bias	Coverage	RMSE	Bias	Coverage
Group 1	60	150	0.1198	0.613	0.0424	0.772	0.1672	-0.006	0.7386
	60	250	0.0822	0.5188	0.0496	0.824	0.1521	-0.0075	0.7887
	60	350	0.0664	0.4691	0.0484	0.848	0.1378	-0.0023	0.7973
	120	150	0.1171	0.5779	0.0436	0.743	0.1155	-0.0023	0.6536
	120	250	0.0836	0.5063	0.0466	0.799	0.1097	-0.0087	0.7076
	120	350	0.0623	0.4269	0.0353	0.843	0.0989	-0.0023	0.7645
Group 2	60	150	0.1198	3.9075	-1.5016	0.896	0.1249	-0.0084	0.8774
	60	250	0.0822	4.3259	-1.3592	0.921	0.1128	-0.0026	0.8805
	60	350	0.0664	4.6975	-1.3302	0.915	0.1021	-0.0073	0.869
	120	150	0.1171	3.8117	-1.4226	0.894	0.0877	-0.0024	0.8795
	120	250	0.0836	4.3902	-1.4004	0.915	0.0799	-0.0022	0.8875
	120	350	0.0623	4.4528	-1.1962	0.912	0.0725	-0.0062	0.8622
Group 3	60	150	0.1198	0.4031	-0.0128	0.963	0.0006	0	0.963
	60	250	0.0822	0.0001	0	0.941	0.0001	0	0.941
	60	350	0.0664	0	0	0.961	0	0	0.961
	120	150	0.1171	0.5701	-0.0255	0.954	0.0004	0	0.954
	120	250	0.0836	0.0001	0	0.944	0.0001	0	0.944
	120	350	0.0623	0	0	0.946	0	0	0.946

Table 14: Classification and Estimation by Modified K-means Algorithm under DGP-A

DGP-B				Post-Classification			Oracle		
	N	T	Cluster Error	RMSE	Bias	Coverage	RMSE	Bias	Coverage
Group 1	60	150	0.1174	0.6063	0.036	0.77	0.1672	-0.006	0.7346
	60	250	0.0823	0.5181	0.0514	0.823	0.1521	-0.0075	0.7841
	60	350	0.0649	0.4591	0.044	0.851	0.1378	-0.0023	0.8029
	120	150	0.1148	0.5764	0.0409	0.737	0.1155	-0.0023	0.645
	120	250	0.0833	0.5096	0.0494	0.798	0.1097	-0.0087	0.708
	120	350	0.061	0.418	0.0324	0.845	0.0989	-0.0023	0.7681
Group 2	60	150	0.1174	3.8791	-1.4798	0.903	0.1249	-0.0084	0.8903
	60	250	0.0823	4.3259	-1.3578	0.92	0.1128	-0.0026	0.8842
	60	350	0.0649	4.6377	-1.2965	0.916	0.1021	-0.0073	0.8691
	120	150	0.1148	3.7824	-1.407	0.897	0.0877	-0.0024	0.8797
	120	250	0.0833	4.3684	-1.3862	0.918	0.0799	-0.0022	0.8937
	120	350	0.061	4.3896	-1.1622	0.913	0.0725	-0.0062	0.867
Group 3	60	150	0.1174	0	0	0.95	0	0	0.95
	60	250	0.0823	0	0	0.943	0	0	0.943
	60	350	0.0649	0	0	0.952	0	0	0.952
	120	150	0.1148	0	0	0.955	0	0	0.955
	120	250	0.0833	0	0	0.947	0	0	0.947
	120	350	0.061	0	0	0.936	0	0	0.936

Table 15: Classification and Estimation by Modified K-means Algorithm under DGP-B

	DGP-C			Post-Classification			Oracle		
	N	T	Cluster Error	RMSE	Bias	Coverage	RMSE	Bias	Coverage
Group 1	60	150	0.021	0.1345	-0.0087	0.927	0.0907	-0.0023	0.933
	60	250	0.0105	0.1002	-0.0069	0.937	0.0825	-0.0043	0.9352
	60	350	0.0093	0.0905	-0.0023	0.945	0.0722	-0.0006	0.947
	120	150	0.0149	0.0827	-0.0009	0.934	0.0648	-0.0032	0.9364
	120	250	0.0138	0.0798	-0.005	0.927	0.0576	-0.0032	0.9328
	120	350	0.0124	0.0752	-0.0024	0.926	0.0528	-0.0006	0.9294
Group 2	60	150	0.021	0.429	0.092	0.914	0	0	0.9096
	60	250	0.0105	0.2966	0.044	0.929	0	0	0.9314
	60	350	0.0093	0.2828	0.04	0.928	0	0	0.9256
	120	150	0.0149	0.3521	0.062	0.92	0	0	0.9302
	120	250	0.0138	0.3406	0.058	0.934	0	0	0.9367
	120	350	0.0124	0.3347	0.056	0.93	0	0	0.9209

Table 16: Classification and Estimation by Modified K-means Algorithm under DGP-C

N	Statistics		Size	
c1=0 (c2=-4), gamma=0.6		T=50	T=100	T=150
1	ADF	0.054	0.068	0.082
15(30)	Post-Classification	0.088	0.074	0.068
30(60)	Post-Classification	0.076	0.07	0.068
45(90)	Post-Classification	0.08	0.072	0.06
N	Statistics		Power	
c1=0.3 (c2=-4), gamma=0.6		T=50	T=100	T=150
1	ADF	0.338	0.514	0.628
15(30)	Post-Classification	0.994	1	1
30(60)	Post-Classification	1	1	1
45(90)	Post-Classification	1	1	1
N	Statistics		Power	
c1=0.5 (c2=-4), gamma=0.6		T=50	T=100	T=150
1	ADF	0.63	0.816	0.888
15(30)	Post-Classification	1	1	1
30(60)	Post-Classification	1	1	1
45(90)	Post-Classification	1	1	1

Table 17: Sizes and Powers of ADF and Panel Wald Tests on Explosive Roots Detection (c1=0, 0.3, 0.5; c2=-4; gamma=0.6)

N	Statistics		Size	
c1=0 (c2=2.5), gamma=0.6		T=50	T=100	T=150
1	ADF	0.054	0.068	0.082
10(30)	Post-Classification	0.044	0.056	0.062
20(60)	Post-Classification	0.05	0.054	0.052
30(90)	Post-Classification	0.048	0.064	0.048
N	Statistics		Power	
c1=0.3 (c2=2.5), gamma=0.6		T=50	T=100	T=150
1	ADF	0.338	0.514	0.628
10(30)	Post-Classification	0.994	1	1
20(60)	Post-Classification	1	1	1
30(90)	Post-Classification	1	1	1
N	Statistics		Power	
c1=0.5 (c2=2.5), gamma=0.6		T=50	T=100	T=150
1	ADF	0.63	0.816	0.888
10(30)	Post-Classification	1	1	1
20(60)	Post-Classification	1	1	1
30(90)	Post-Classification	1	1	1

Table 18: Sizes and Powers of ADF and Panel Wald Tests on Explosive Roots Detection (c1=0, 0.3, 0.5; c2=2.5; gamma=0.6)

N	Statistics		Size	
c1=0 (c2=3, c3=-3), gamma=0.6		T=50	T=100	T=150
1	ADF	0.054	0.068	0.082
10(30)	Post-Classification	0.116	0.11	0.106
20(60)	Post-Classification	0.146	0.096	0.082
30(90)	Post-Classification	0.122	0.076	0.06
N	Statistics		Power	
c1=0.3 (c2=3, c3=-3), gamma=0.6		T=50	T=100	T=150
1	ADF	0.338	0.514	0.628
10(30)	Post-Classification	0.936	0.998	1
20(60)	Post-Classification	1	1	1
30(90)	Post-Classification	1	1	1
N	Statistics		Power	
c1=0.5 (c2=3, c3=-3), gamma=0.6		T=50	T=100	T=150
1	ADF	0.63	0.816	0.888
10(30)	Post-Classification	1	1	1
20(60)	Post-Classification	1	1	1
30(90)	Post-Classification	1	1	1

Table 19: Sizes and Powers of ADF and Panel Wald Tests on Explosive Roots Detection (c1=0, 0.3, 0.5; c2=3; c3=-3; gamma=0.6)

N	Statistics		Size	
c1=0 (c2=-3, c3=-6), gamma=0.5		T=50	T=100	T=150
1	ADF	0.054	0.068	0.082
10(30)	Post-Classification	0.08	0.068	0.044
20(60)	Post-Classification	0.09	0.054	0.058
30(90)	Post-Classification	0.08	0.04	0.062
N	Statistics		Power	
c1=0.3 (c2=-3, c3=-6), gamma=0.5		T=50	T=100	T=150
1	ADF	0.566	0.774	0.884
10(30)	Post-Classification	1	1	1
20(60)	Post-Classification	1	1	1
30(90)	Post-Classification	1	1	1
N	Statistics		Power	
c1=0.5 (c2=-3, c3=-6), gamma=0.5		T=50	T=100	T=150
1	ADF	0.828	0.958	0.986
10(30)	Post-Classification	1	1	1
20(60)	Post-Classification	1	1	1
30(90)	Post-Classification	1	1	1

Table 20: Sizes and Powers of ADF and Panel Wald Tests on Explosive Roots Detection (c1=0, 0.3, 0.5; c2=-3; c3=-6; gamma=0.5)

N	Statistics		Size	
c1=0 (c2=2, c3=4), gamma=0.5		T=50	T=100	T=150
1	ADF	0.054	0.068	0.082
10(30)	Post-Classification	0.05	0.056	0.044
20(60)	Post-Classification	0.058	0.048	0.054
30(90)	Post-Classification	0.062	0.04	0.054
N	Statistics		Power	
c1=0.3 (c2=2, c3=4), gamma=0.5		T=50	T=100	T=150
1	ADF	0.566	0.774	0.884
10(30)	Post-Classification	1	0.998	1
20(60)	Post-Classification	1	1	1
30(90)	Post-Classification	1	1	1
N	Statistics		Power	
c1=0.5 (c2=2, c3=4), gamma=0.5		T=50	T=100	T=150
1	ADF	0.828	0.958	0.986
10(30)	Post-Classification	1	1	1
20(60)	Post-Classification	1	1	1
30(90)	Post-Classification	1	1	1

Table 21: Sizes and Powers of ADF and Panel Wald Tests on Explosive Roots Detection (c1=0, 0.3, 0.5; c2=2; c3=4; gamma=0.5)

A.3 Detection of bubble origination

We collect results on estimations of bubble origination date based on panel recursive detector and time-series recursive detector. We simulate 500 sample paths. Each group has an equal proportion. The true value of the bubble origination date is 0.5. The critical value for panel recursive method is $\log(\log(nT))$. The critical value for the time-series recursive method is $\log(\log(T))$. "Panel" denotes the panel recursive detector. "TS" denotes the time-series recursive detector.

c1=(c2=-3,c3=-9)	1		2		3		4	
N=60, T=150								
Detector	Panel	TS	Panel	TS	Panel	TS	Panel	TS
Mean	0.5151	0.5859	0.4979	0.525	0.4925	0.5068	0.4903	0.498
Std	0.0703	0.1529	0.0683	0.1151	0.0673	0.1039	0.0667	0.0989
RMSE	1.6059	3.9177	1.5269	2.6315	1.5127	2.3263	1.5053	2.2091
c1=(c2=-3,c3=-9)	1		2		3		4	
N=60, T=250								
Detector	Panel	TS	Panel	TS	Panel	TS	Panel	TS
Mean	0.5096	0.563	0.4949	0.5158	0.491	0.5021	0.499	0.4947
Std	0.0677	0.1345	0.0678	0.1022	0.0668	0.0952	0.0662	0.091
RMSE	1.527	3.3184	1.5193	2.3102	1.505	2.1274	1.5	2.0356
c1=(c2=-3,c3=-9)	1		2		3		4	
N=60, T=350								
Detector	Panel	TS	Panel	TS	Panel	TS	Panel	TS
Mean	0.5059	0.5502	0.4948	0.5113	0.4914	0.4987	0.4899	0.4927
Std	0.0658	0.118	0.0644	0.0949	0.0636	0.0882	0.0632	0.0853
RMSE	1.477	2.8656	1.4436	2.1347	1.4326	1.9702	1.4292	1.9122

Table 22: Results on estimations of bubble origination date based on panel recursive detector and time-series recursive detector. Innovations are normal distributions. Dimension of cross-sectional units is 60.

c1=(c2=-3,c3=-9)	1		2		3		4	
N=120, T=150								
Detector	Panel	TS	Panel	TS	Panel	TS	Panel	TS
Mean	0.5103	0.5859	0.4955	0.5261	0.4913	0.5073	0.4895	0.4987
Std	0.0673	0.1529	0.0661	0.1113	0.065	0.0993	0.0645	0.0943
RMSE	1.5203	3.9177	1.4801	2.5535	1.4644	2.2233	1.4597	2.1077
c1=(c2=-3,c3=-9)	1		2		3		4	
N=120, T=250								
Detector	Panel	TS	Panel	TS	Panel	TS	Panel	TS
Mean	0.5059	0.563	0.4946	0.5158	0.4912	0.5021	0.4895	0.4947
Std	0.0685	0.1345	0.0655	0.1022	0.0646	0.0952	0.0642	0.091
RMSE	1.5353	3.3184	1.4673	2.3102	1.4566	2.1274	1.4528	2.0356
c1=(c2=-3,c3=-9)	1		2		3		4	
N=120, T=350								
Detector	Panel	TS	Panel	TS	Panel	TS	Panel	TS
Mean	0.5067	0.5502	0.4963	0.5113	0.4928	0.4987	0.4914	0.4927
Std	0.0569	0.118	0.0572	0.0949	0.058	0.0882	0.0577	0.0853
RMSE	1.2796	2.8656	1.2804	2.1347	1.3055	1.9702	1.3027	1.9122

Table 23: Results on estimations of bubble origination date based on panel recursive detector and time-series recursive detector. Innovations are normal distributions. Dimension of cross-sectional units is 120.

c1=(c2=-3,c3=-9)	1		2		3		4	
N=60, T=150								
Detector	Panel	TS	Panel	TS	Panel	TS	Panel	TS
Mean	0.5071	0.5744	0.4943	0.5271	0.4877	0.5041	0.4866	0.4954
Std	0.0744	0.1463	0.0733	0.1083	0.0753	0.102	0.0713	0.0971
RMSE	1.6687	3.6672	1.6434	2.494	1.7053	2.28	1.6215	2.1714
c1=(c2=-3,c3=-9)	1		2		3		4	
N=60, T=250								
Detector	Panel	TS	Panel	TS	Panel	TS	Panel	TS
Mean	0.5115	0.5601	0.4965	0.5127	0.4937	0.4984	0.493	0.4996
Std	0.0628	0.1287	0.0643	0.1047	0.0604	0.0981	0.0573	0.0943
RMSE	1.4269	3.1742	1.4385	2.3564	1.3574	2.1907	1.2889	2.1186
c1=(c2=-3,c3=-9)	1		2		3		4	
N=60, T=350								
Detector	Panel	TS	Panel	TS	Panel	TS	Panel	TS
Mean	0.507	0.5532	0.4985	0.5114	0.4914	0.5035	0.4909	0.4889
Std	0.0621	0.1067	0.0539	0.0936	0.0627	0.0811	0.0604	0.0903
RMSE	1.3957	2.6641	1.2049	2.1071	1.4137	1.8127	1.3653	2.0316

Table 24: Results on estimations of bubble origination date based on panel recursive detector and time-series recursive detector. Innovations are student's t distributions with degree of freedom 3. Dimension of cross-sectional units is 60.

.

c1=(c2=-3,c3=-9)	1		2		3		4	
N=120, T=150								
Detector	Panel	TS	Panel	TS	Panel	TS	Panel	TS
Mean	0.5059	0.5722	0.4978	0.5339	0.4914	0.5131	0.492	0.5005
Std	0.0734	0.148	0.0612	0.111	0.0658	0.1001	0.0603	0.0941
RMSE	1.6451	3.6791	1.3671	2.5926	1.4831	2.2546	1.3586	2.1022
c1=(c2=-3,c3=-9)	1		2		3		4	
N=120, T=250								
Detector	Panel	TS	Panel	TS	Panel	TS	Panel	TS
Mean	0.5041	0.5621	0.4927	0.5192	0.4912	0.5048	0.4969	0.4988
Std	0.0699	0.132	0.0688	0.1032	0.0644	0.0898	0.0687	0.0962
RMSE	1.5647	3.2586	1.545	2.344	1.4511	2.0084	1.5623	2.1627
c1=(c2=-3,c3=-9)	1		2		3		4	
N=120, T=350								
Detector	Panel	TS	Panel	TS	Panel	TS	Panel	TS
Mean	0.5044	0.5402	0.496	0.5092	0.4884	0.4967	0.4901	0.4935
Std	0.0609	0.1209	0.0576	0.0981	0.0673	0.0896	0.0602	0.0818
RMSE	1.3636	2.8475	1.2904	2.2006	1.5264	2.002	1.3619	1.8323

Table 25: Results on estimations of bubble origination date based on panel recursive detector and time-series recursive detector. Innovations are student's t distributions with degree of freedom 3. Dimension of cross-sectional units is 120.

B Empirical Illustrations

The classified membership is as follows,

Group 1: Anshan, Baoding, Changzhou, Chengdu, Chuzhou, Dalian, Dongguan, Foshan, Fuzhou, Hangzhou, Harbin, Hefei, Heyuan, Huaian, Huizhou, Huzhou, Jiaxing, Jieyang, Jinan, Jinhua, Jiujiang, Kaifeng, Kunming, Langfang, Leshan, Luoyang, Mianyang, Nanjing, Nanning, Nantong, Nanyang, Ningbo, Qingdao, Qingyuan, Qinhuangdao, Quanzhou, Rizhao, Shanghai, Shaoxing, Shijiazhuang, Songyuan, Suzhou, Tangshan, Tianjin, Tieling, Wenzhou, Wuhan, Wuhu, Wuxi, Xiamen, Xi'an, Xingtai, Xuancheng, Yingkou, Zaozhuang, Zhaoqing, Zhenjiang, Zhongshan;

Group 2: Anqing, Baotou, Beijing, Bengbu, Chaoyang, Changchun, Changde, Changji, Changsha, Chongqing, Dandong, Deyang, Fuzhou1, Guangzhou, Haikou, Hohhot, Jiangmen, Jiangyan, Jingdezhen, Luzhou, Luohe, Nanchang, Nanchong, Ningde, Pingxiang, Shangrao, Shantou, Shanwei, Shaoguan, Shenyang, Shenzhen, Suqian, Taizhou, Urumq, Wuludao, Xining, Xinxiang, Xinyu, Xuchang, Xuzhou, Yancheng, Yangzhou, Yichun, Zhangjiakou, Zhangzhou, Zhengzhou, Zhumadian;

Group 3: Huangshan, Puyang;

Group 4: Dezhou, Jilin, Lianyungang, Nanyang, Xilingol, Yangjiang.

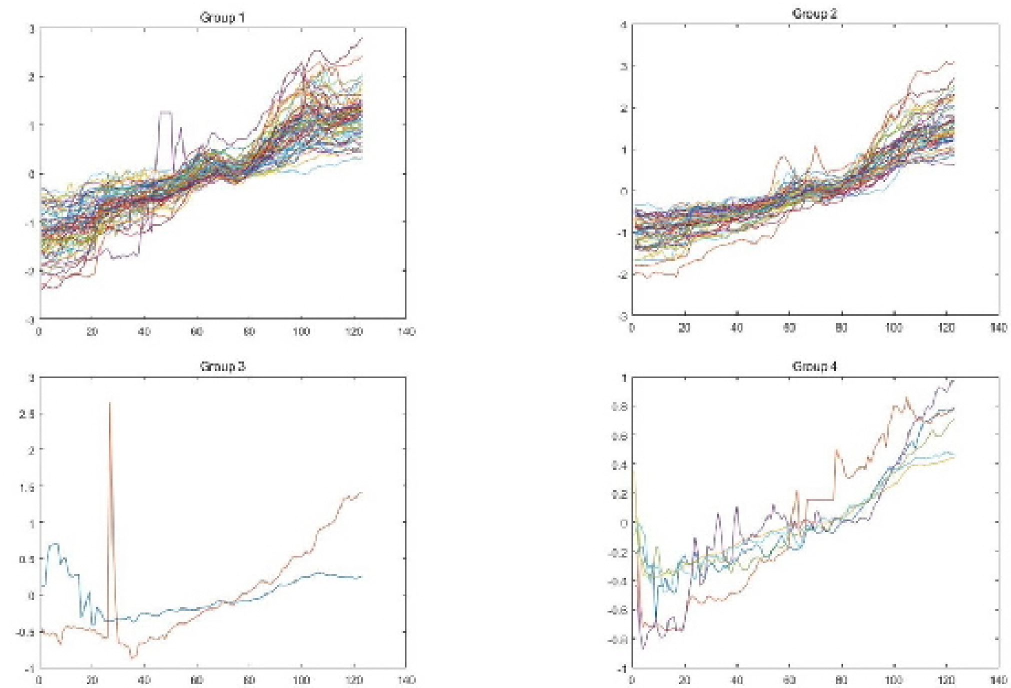

Figure 1: Classified Membership Using Modified K-means Algorithm

	K0=1	K0=2	K0=3	K0=4	K0=5	K0=6
BIC	8.6493	8.2803	7.9571	**7.782**	7.7884	7.7937

Table 26: BIC for Group Number

Membership	G1	G2	G3	G4
Slope Estimate	0.9962	1.0068	0.881	1.0017
Explosive Test	-271.097	486.4522	-28.9456	3.2209

Table 27: Estimates and Explosive Inference for Each Group

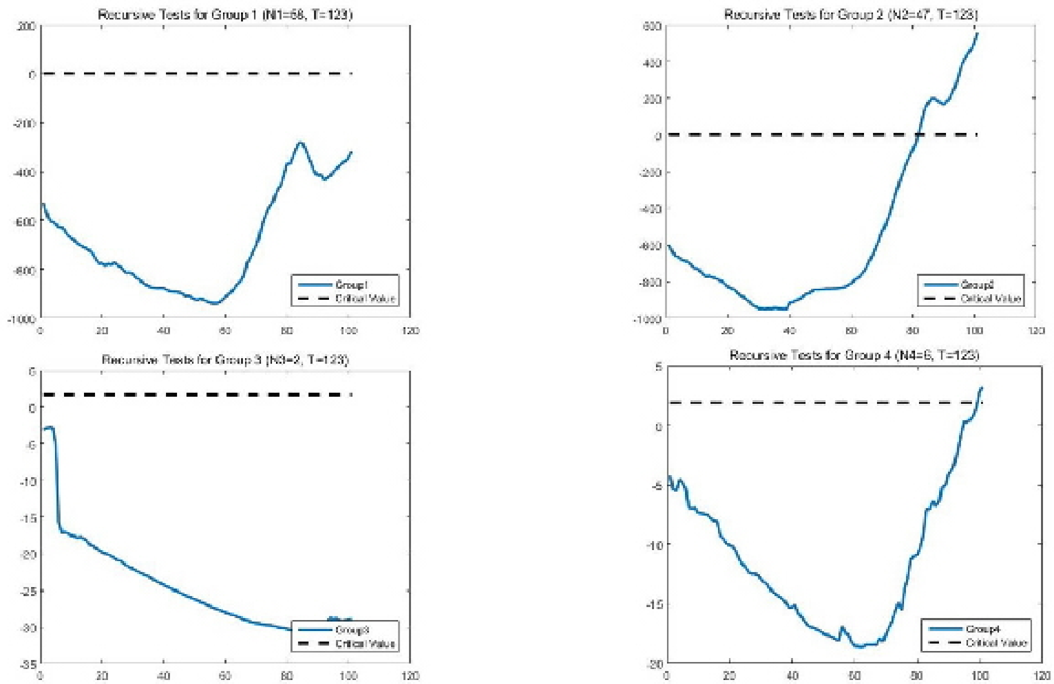

Figure 2: Panel Recursive Bubble Detector for Each Group

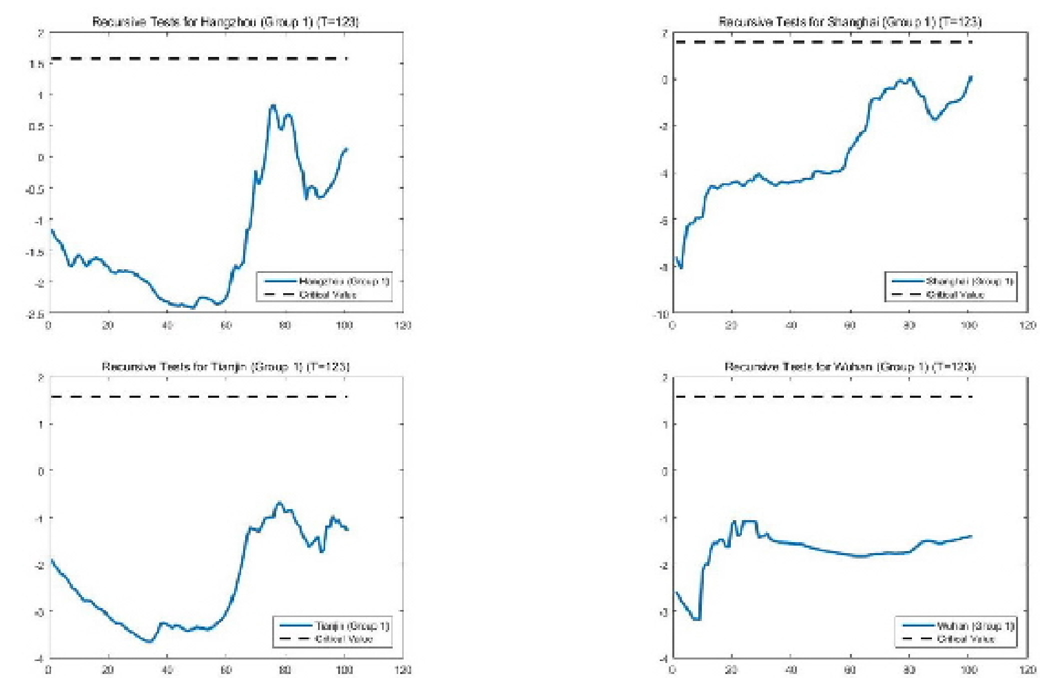

Figure 3: Robustness Check for Group 1

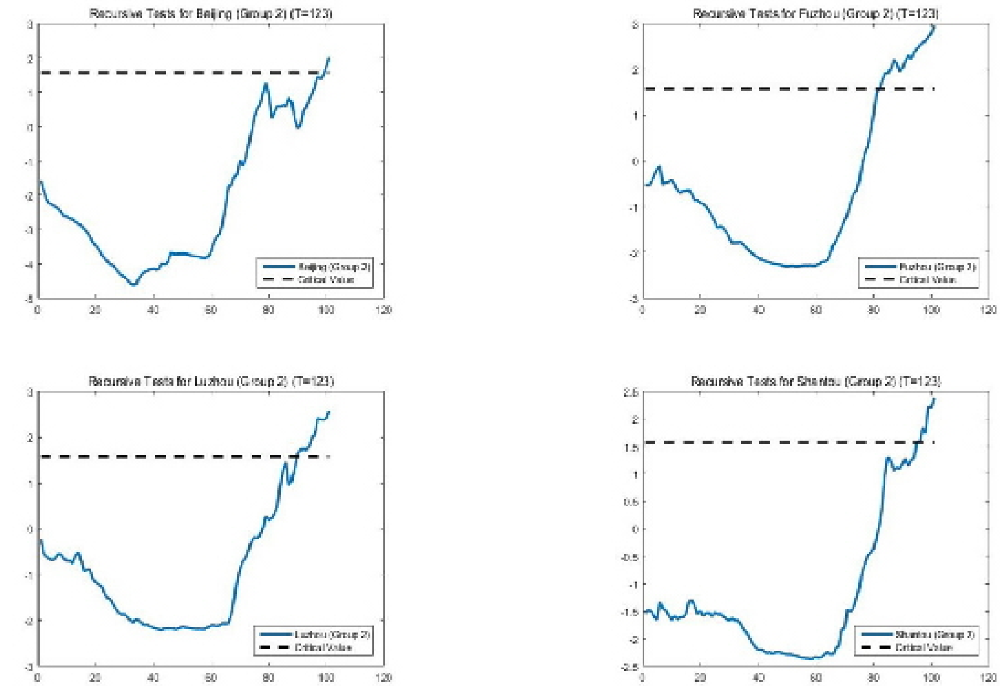

Figure 4: Robustness Check for Group 2

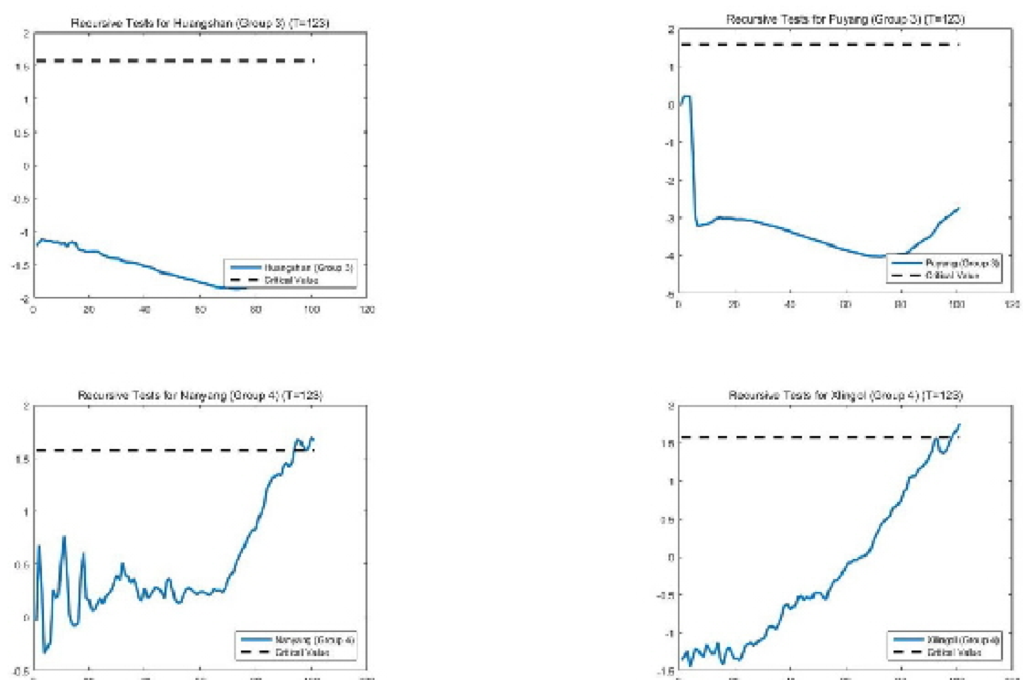

Figure 5: Robustness Check for Group 3 and 4

C Statistical Properties

We collect technical proofs for classifications and estimations on the recursive k-means classifications and the modified k-means classifications.

C.1 Recursive k-means classifications

The consistency of parameter estimations and uniform consistency of recursive k-means membership clustering are justified in the subsubsection of "consistency". The asymptotic equivalence between oracle estimator and post-classification estimator is verified in the subsubsection of "limiting distribution". Limiting distributions of the post-classification estimators are shown in the subsubsection of "limiting distribution".

C.1.1 Consistency

Lemma C.1 *If Assumption 1 holds,*

$$\sup_{(c,\delta)\in\Phi^n\times\Delta_{K^0}} T^{2\gamma}\left|\widehat{Q}_{nT}(c,\delta)-\widetilde{Q}_{nT}(c,\delta)\right| = o_p(1),$$

where

$$\widehat{Q}_{nT}(c,\delta) := \frac{1}{nT^{2\gamma}}\sum_{i=1}^{n}\frac{1}{\overline{\rho}_i^{2T}}\sum_{t=1}^{T}\left(\widetilde{y}_{it}-\widetilde{y}_{i,t-1}\rho_{g_i n}\right)^2,$$

and

$$\widetilde{Q}_{nT}(c,\delta) := \frac{1}{nT^{2\gamma}}\sum_{i=1}^{n}\frac{1}{\overline{\rho}_i^{2T}}\sum_{t=1}^{T}\left(\widetilde{y}_{i,t-1}\rho_{g_i n}\right)^2+\frac{1}{nT^{2\gamma}}\sum_{i=1}^{n}\frac{1}{\overline{\rho}_i^{2T}}\sum_{t=1}^{T}\widetilde{u}_{it}^2,$$

and $\{\overline{\rho}_i\}_{i=1}^n$ *are the collection of individual least square estimates and* $\rho_{g_i n} := \exp\left(\frac{c_{g_i n}}{T^\gamma}\right)$.

The Proof of Lemma C.1: Define $\rho^0_{g_i^0 n} := \exp\left(c^0_{g_i^0 n}/T^\gamma\right)$ and $\rho_{g_i n} := \exp\left(\frac{c_{g_i n}}{T^\gamma}\right)$. Observe the following argument as,

$$\begin{aligned}
T^{2\gamma}\left[\widehat{Q}_{nT}(c,\delta)-\widetilde{Q}_{nT}(c,\delta)\right] &= \frac{2}{n}\sum_{i=1}^{n}\frac{1}{\overline{\rho}_i^{2T}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\left(\rho^0_{g_i^0 n}-\rho_{g_i n}\right)\\
&= \frac{2}{nT^\gamma}\sum_{i=1}^{n}\frac{1}{\overline{\rho}_i^{2T}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\left(c^0_{g_i^0 n}-c_{g_i n}\right)\\
&= \frac{2}{nT^\gamma}\sum_{i=1}^{n}\frac{1}{\overline{\rho}_i^{2T}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}c^0_{g_i^0 n}\\
&\quad -\frac{2}{nT^\gamma}\sum_{i=1}^{n}\frac{1}{\overline{\rho}_i^{2T}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}c_{g_i n}.
\end{aligned}$$

We have

$$\frac{2}{nT^\gamma}\sum_{i=1}^{n}\frac{1}{\overline{\rho}_i^{2T}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}c^0_{g_i^0 n} = \sum_{\widetilde{g}=1}^{K^0}\mathbf{1}\left\{g_i^0=\widetilde{g}\right\}\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\overline{\rho}_i^{2T}T^\gamma}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}c^0_{\widetilde{g}n}.$$

For any $\widetilde{g} \in \{1, 2, ..., K^0\}$, we then have

$$\begin{aligned}
\mathbf{1}\left\{g_i^0 = \widetilde{g}\right\} \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\overline{\rho}_i^{2T} T^{\gamma}} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}\widetilde{u}_{it} c_{\widetilde{g}n} &\leq \left|c_{\widetilde{g}n}^0\right| \mathbf{1}\left\{g_i^0 = \widetilde{g}\right\} \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\overline{\rho}_i^{2T} T^{\gamma}} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}\widetilde{u}_{it} \\
&= O_p\left(\frac{1}{\rho_{\widetilde{g}n}^T \sqrt{n}}\right),
\end{aligned}$$

since $\left|c_{\widetilde{g}n}\right| \leq \overline{c}$ due to the compact support of distance parameters. Define $\rho_{\widetilde{g}n} := \exp\left(\frac{c_{\widetilde{g}n}}{T^{\gamma}}\right)$. Therefore

$$\frac{2}{nT^{\gamma}}\sum_{i=1}^{n} \frac{1}{\overline{\rho}_i^{2T}} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}\widetilde{u}_{it} c_{g_i^0 n}^0 = O_p\left(\frac{1}{\underline{\rho}^T \sqrt{n}}\right), \tag{22}$$

where we define $\underline{\rho} := \exp\left(\frac{\underline{c}}{T^{\gamma}}\right)$.

Similar argument can be applied to the term $\frac{2}{nT^{\gamma}}\sum_{i=1}^{n} \frac{1}{\overline{\rho}_i^{2T}} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}\widetilde{u}_{it} c_{g_i n}$ and we can justify

$$\frac{2}{nT^{\gamma}}\sum_{i=1}^{n} \frac{1}{\overline{\rho}_i^{2T}} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}\widetilde{u}_{it} c_{g_i n} = O_p\left(\frac{1}{\underline{\rho}^T \sqrt{n}}\right). \tag{23}$$

By combining above results (22) and (23), we have

$$\sup_{(c,\delta)\in\Phi^n\times\Delta_{K^0}} T^{2\gamma}\left|\widehat{Q}_{nT}(c,\delta) - \widetilde{Q}_{nT}(c,\delta)\right| = O_p\left(\frac{1}{\underline{\rho}^T\sqrt{n}}\right) = o_p(1).$$

■

Proof of Theorem 4.1: Define $\rho_{g_i^0 n}^0 := \exp\left(c_{g_i^0 n}^0 / T^{\gamma}\right)$ and $\rho_{g_i n} := \exp\left(\frac{c_{g_i n}}{T^{\gamma}}\right)$. Observe the following argument as,

$$\widetilde{Q}_{nT}(\widehat{c},\widehat{\delta}) = \widehat{Q}_{nT}(\widehat{c},\widehat{\delta}) + o_p(T^{-2\gamma}) \leq \widehat{Q}_{nT}(c^0,\delta^0) + o_p(T^{-2\gamma}) = \widetilde{Q}_{nT}(c^0,\delta^0) + o_p(T^{-2\gamma}),$$

where the equalities come from Lemma C.1. Because $\widetilde{Q}_{nT}(c,\delta)$ is minimized at $c = c^0$ and $\delta = \delta^0$, we have

$$\widetilde{Q}_{nT}(\widehat{c},\widehat{\delta}) - \widetilde{Q}_{nT}(c^0,\delta^0) = o_p(T^{-2\gamma}).$$

On the other hand, for any c, we have

$$\begin{aligned}
o_p(T^{-2\gamma}) &= \widetilde{Q}_{nT}(c,\delta) - \widetilde{Q}_{nT}(c^0,\delta^0) \\
&= \frac{1}{nT^{2\gamma}}\sum_{i=1}^{n} \frac{1}{\overline{\rho}_i^{2T}} \sum_{t=1}^{T} \left(\widetilde{y}_{i,t-1}(\rho_{g_i^0 n}^0 - \rho_{g_i n})\right)^2 \\
&= (\rho_{gn}^0 - \rho_{\widetilde{g}n})^2 \left[\frac{1}{n}\sum_{i=1}^{n} \frac{1}{T^{2\gamma}\overline{\rho}_i^{2T}} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}^2\right] \\
&\geq (\rho_{gn}^0 - \rho_{\widetilde{g}n})^2 \frac{\sigma^2}{4\overline{c}^2} + o_p\left(\frac{1}{T^{2\gamma}}\right)
\end{aligned}$$

$$
\begin{aligned}
&\geq \frac{\sigma^2}{4\overline{c}^2}\frac{1}{T^{2\gamma}}\min_{\widetilde{g}\in\{1,2,\ldots,K^0\}}\left(c_{gn}^0-c_{\widetilde{g}n}\right)^2+o_p\left(\frac{1}{T^{2\gamma}}\right)\\
&\geq \frac{\sigma^2}{4\overline{c}^2}\frac{1}{T^{2\gamma}}\max_{g\in\{1,2,\ldots,K^0\}}\left(\min_{\widetilde{g}\in\{1,2,\ldots,K^0\}}\left(c_{gn}^0-c_{\widetilde{g}n}\right)^2\right)+o_p\left(\frac{1}{T^{2\gamma}}\right).
\end{aligned}
$$

Note that $\frac{\sigma^2}{4\overline{c}^2}$ is bounded away from zeros by Assumption 1. As a result, we have

$$
\max_{g\in\{1,2,\ldots,K^0\}}\left(\min_{\widetilde{g}\in\{1,2,\ldots,K^0\}}\left(\rho_{gn}^0-\rho_{\widetilde{g}n}\right)^2\right)=o_p\left(T^{-2\gamma}\right), \tag{24}
$$

or

$$
\max_{g\in\{1,2,\ldots,K^0\}}\left(\min_{\widetilde{g}\in\{1,2,\ldots,K^0\}}\left(c_{gn}^0-c_{\widetilde{g}n}\right)^2\right)=o_p\left(1\right). \tag{25}
$$

Let

$$
\tau(g)=\arg\min_{\widetilde{g}\in\{1,2,\ldots,K^0\}}\left(c_{gn}^0-c_{\widetilde{g}n}\right)^2.
$$

Then we have for $\widetilde{g}\neq g$,

$$
\begin{aligned}
&T^{2\gamma}\left(\frac{1}{nT^{2\gamma}}\sum_{i=1}^n\frac{1}{\rho_i^{2T}}\sum_{t=1}^T\widetilde{y}_{i,t-1}^2(\widehat{\rho}_{\tau(g)n}-\widehat{\rho}_{\tau(\widetilde{g})n})^2\right)^{\frac{1}{2}}\\
\geq\ &T^{2\gamma}\left(\frac{1}{nT^{2\gamma}}\sum_{i=1}^n\frac{1}{\rho_i^{2T}}\sum_{t=1}^T\widetilde{y}_{i,t-1}^2(\rho_{gn}^0-\rho_{\widetilde{g}n}^0)^2\right)^{\frac{1}{2}}\\
&-T^{2\gamma}\left(\frac{1}{nT^{2\gamma}}\sum_{i=1}^n\frac{1}{\rho_i^{2T}}\sum_{t=1}^T\widetilde{y}_{i,t-1}^2(\widehat{\rho}_{\tau(g)n}-\rho_{gn}^0)^2\right)^{\frac{1}{2}}\\
&-T^{2\gamma}\left(\frac{1}{nT^{2\gamma}}\sum_{i=1}^n\frac{1}{\rho_i^{2T}}\sum_{t=1}^T\widetilde{y}_{i,t-1}^2(\widehat{\rho}_{\tau(\widetilde{g})n}-\rho_{\widetilde{g}n}^0)^2\right)^{\frac{1}{2}}.
\end{aligned} \tag{26}
$$

The first term on the right hand side of (26) is bounded away from zero since

$$
\frac{1}{nT^{2\gamma}}\sum_{i=1}^n\frac{1}{\widetilde{\rho}_i^{2T}}\sum_{t=1}^T\widetilde{y}_{i,t-1}^2\geq\frac{\sigma^2}{4\overline{c}^2}+o_p\left(\frac{1}{T^{2\gamma}}\right),
$$

and $(c_{gn}^0-c_{\widetilde{g}n}^0)^2\geq c_{g,\widetilde{g}}^2>0$ as $\widetilde{g}\neq g$. Due to (24) or (25), we have the 2nd and 3rd term of (26) are $o_p(1)$. Therefore, we have $\tau(g)\neq\tau(\widetilde{g})$ with probability approaching one. We note that asymptotically τ is not only an onto mapping but also one-to-one mapping. Hence τ has the inverse mapping denoted as τ^{-1}. Then we have

$$
\min_{g\in\{1,2,\ldots,K^0\}}\left(\rho_{gn}^0-\widehat{\rho}_{\widetilde{g}n}\right)^2\geq\left(\rho_{\tau^{-1}(\widetilde{g})n}^0-\widehat{\rho}_{\widetilde{g}n}\right)^2=\min_{h\in\{1,2,\ldots,K^0\}}\left(\rho_{\tau^{-1}(h)n}^0-\widehat{\rho}_{hn}\right)^2=o_p\left(T^{-2\gamma}\right),
$$

where the last equality comes from (24) or (25). So we have

$$
\max_{\widetilde{g}\in\{1,2,\ldots,K^0\}}\left(\min_{g\in\{1,2,\ldots,K^0\}}\left(\rho_{gn}^0-\widehat{\rho}_{\widetilde{g}n}\right)^2\right)=o_p(T^{-2\gamma}). \tag{27}
$$

In all, by combining the results for two terms (24) and (27), we complete the proof. ■

Lemma C.2 *If Assumption 1 holds, for any $M > 0$,*

$$\max_{1 \le i \le n} \Pr\left(\frac{1}{\rho_i^{2T} T^{\gamma}} \left| \sum_{t=1}^{T} \widetilde{y}_{i,t-1} \widetilde{u}_{it} \right| \ge M \right) = o\left(\frac{1}{n}\right)$$

The Proof of Lemma C.2: Due to the uniform dominance of innovations over the fixed effect in (1), we consider the following decomposition as,

$$\frac{1}{\rho_i^{2T} T^{\gamma}} \sum_{t=1}^{T} \widetilde{y}_{i,t-1} \widetilde{u}_{it} = \frac{1}{\rho_i^{2T} T^{\gamma}} \sum_{t=1}^{T} y_{i,t-1} u_{it} - \frac{1}{\rho_i^{2T} T^{\gamma-1}} \overline{y}_{i,-1} \overline{u}_{i}. \tag{28}$$

It is equivalent to consider the decomposition of (28) in the model (1) with $\mu_i = 0$ for all $i = 1, 2, ..., n$. Fix $M > 0$. For the first term of (28), we apply the Markov inequality and derive the following

$$\begin{aligned} n \max_{1 \le i \le n} \Pr\left(\frac{1}{\rho_i^{2T} T^{\gamma}} \left| \sum_{t=1}^{T} y_{i,t-1} u_{it} \right| \ge \frac{M}{2} \right) &\le n \max_{1 \le i \le n} \frac{4\mathbb{E}\left(\sum_{t=1}^{T} y_{i,t-1} u_{it}\right)^2}{\rho_i^{4T} T^{2\gamma} M^2} \\ &= n \max_{1 \le i \le n} \frac{4\sigma^4}{\rho_i^{2T} M^2 c_i^2} \le \frac{4\sigma^4}{\underline{\rho}^{2T} M^2 \underline{c}^2} \\ &= O\left(\frac{n}{\underline{\rho}^{2T}}\right) = o(1), \end{aligned} \tag{29}$$

with the dominance of exponential rates. Define $\underline{\rho} := \exp\left(\frac{\underline{c}}{T^{\gamma}}\right)$. For the second term of (28), Markov inequality is applied as,

$$\begin{aligned} n \max_{1 \le i \le n} \Pr\left(\frac{1}{\rho_i^{2T} T^{\gamma-1}} \left| \overline{y}_{i,-1} \overline{u}_{i} \right| \ge \frac{M}{2} \right) &\le n \max_{1 \le i \le n} \frac{\mathbb{E}\left(\overline{y}_{i,-1} \overline{u}_{i}\right)}{\rho_i^{2T} T^{\gamma-1}} \\ &\le n \max_{1 \le i \le n} \left[\frac{4\sigma^2}{\rho_i^{T} T^{1-\gamma} c_i^2 M^2} + o\left(\frac{1}{\rho_i^{T} T^{1-\gamma}}\right) \right] \\ &= O\left(\frac{n}{T^{1-\gamma} \underline{\rho}^{T}}\right) = o(1), \end{aligned} \tag{30}$$

with the dominance of exponential rates. The second last equality of (30) is due to the following fact,

$$\begin{aligned} \mathbb{E}\left(\overline{y}_{i,-1} \overline{u}_{i}\right) &= \mathbb{E}\left(\frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{t-1} \left(\sum_{l=1}^{t-1} \rho_i^{t-1-l} u_{il} \right) u_{is} \right) = \mathbb{E}\left(\frac{1}{T^2} \sum_{t=1}^{T} \sum_{l=1}^{t-1} \rho_i^{t-1-l} u_{il}^2 \right) \\ &= \frac{\sigma^2 \rho_i^{T}}{\rho_i T^{2-2\gamma} c_i^2} - \frac{\sigma^2}{\rho_i T^{2-2\gamma} c_i^2} - \frac{\sigma^2}{T^{2-2\gamma} c_i}. \end{aligned}$$

By combining above results (29) and (30), we derive

$$\max_{1 \le i \le n} \Pr\left(\frac{1}{\rho_i^{2T} T^{\gamma}} \left| \sum_{t=1}^{T} \widetilde{y}_{i,t-1} \widetilde{u}_{it} \right| \ge M \right) = o\left(\frac{1}{n}\right).$$

■

Lemma C.3 *If Assumption 1 holds, for arbitrary $\widetilde{M} \geq \frac{5\sigma^2}{2\underline{c}^2}$,*

$$\max_{1\leq i\leq n} \Pr\left(\frac{1}{\rho_i^{2T}T^{\gamma}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2 \geq \widetilde{M}\right) = o\left(\frac{1}{n}\right)$$

The Proof of Lemma C.3: Due to the uniform dominance of innovations over the fixed effect in (1), we consider the following decomposition as,

$$\frac{1}{\rho_i^{2T}T^{2\gamma}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2 = \frac{1}{2c_i}\left\{\frac{\rho_i^{-2T}}{T^{\gamma}}y_{i,T}^2 - \frac{2\rho_i^{-2T+1}}{T^{\gamma}}\sum_{t=1}^{T}y_{i,t-1}u_{it} - \frac{\rho_i^{-2T}}{T^{\gamma}}\sum_{t=1}^{T}u_{it}^2\right\} - \frac{1}{\rho_i^{2T}T^{2\gamma-1}}\overline{y}_{i,-1}^2.$$

It is equivalent to consider the decomposition in (1) with $\mu_i = 0$ for all $i = 1, 2, ..., n$. Therefore we can derive the following uniform upper bound as

$$\begin{aligned}
n\max_{1\leq i\leq n} \Pr\left(\frac{1}{\rho_i^{2T}T^{2\gamma}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2 \geq \widetilde{M}\right) \leq\ & n\max_{1\leq i\leq n} \Pr\left(\left|\frac{1}{2c_i}\frac{\rho_i^{-2T}}{T^{\gamma}}\left(y_{i,T}^2 - \mathbb{E}y_{i,T}^2\right)\right| \geq \frac{\widetilde{M}}{5}\right) \\
& + n\max_{1\leq i\leq n} \Pr\left(\left|\frac{1}{2c_i}\frac{2\rho_i^{-2T+1}}{T^{\gamma}}\sum_{t=1}^{T}y_{i,t-1}u_{it}\right| \geq \frac{\widetilde{M}}{5}\right) \\
& + n\max_{1\leq i\leq n} \Pr\left(\left|\frac{1}{2c_i}\frac{\rho_i^{-2T}}{T^{\gamma}}\sum_{t=1}^{T}u_{it}^2\right| \geq \frac{\widetilde{M}}{5}\right) \\
& + n\max_{1\leq i\leq n} \Pr\left(\left|\frac{1}{2c_i}\frac{1}{\rho_i^{2T}T^{2\gamma-1}}\overline{y}_{i,-1}^2\right| \geq \frac{\widetilde{M}}{5}\right) \\
& + n\max_{1\leq i\leq n} \Pr\left(\frac{1}{2c_i}\frac{\rho_i^{-2T}}{T^{\gamma}}\mathbb{E}y_{i,T}^2 \geq \frac{\widetilde{M}}{5}\right). \qquad (31)
\end{aligned}$$

For the first term of (31), we can bound this term using the exponential inequality illustrated in Freedman (1975). By the martingale difference error process, we have

$$n\max_{1\leq i\leq n} \Pr\left(\left|\frac{1}{2c_i}\frac{\rho_i^{-2T}}{T^{\gamma}}\left(y_{i,T}^2 - \mathbb{E}y_{i,T}^2\right)\right| \geq \frac{\widetilde{M}}{5}\right) \leq 2n\max_{1\leq i\leq n} \Pr\left(\left|\frac{1}{2c_i}\frac{\rho_i^{-2T}}{T^{\gamma}}\sum_{t=1}^{T}\rho_i^{2(T-t)}\left(u_{it}^2 - \mathbb{E}u_{it}^2\right)\right| \geq \frac{\widetilde{M}}{10}\right). \qquad (32)$$

Note the fact that for each $i = 1, 2, ...n$, $\{u_{it}^2 - \mathbb{E}u_{it}^2\}_{t=1}^{T}$ is a martingale difference sequence adaptive to the filtration $\mathcal{F}_{i,t-1} := \sigma\{u_{i,t-1}, u_{i,t-2}, u_{i,t-3}, ...\}$ as $\mathbb{E}\{u_{it}^2 - \mathbb{E}u_{it}^2|\mathcal{F}_{i,t-1}\} = 0$ $\forall t = 1, 2, ...T$. Define $z_{it} := \rho_i^{2(T-t)}\left(u_{it}^2 - \mathbb{E}u_{it}^2\right) =: z_{1it} + z_{2it} - \mathbb{E}\{z_{2it}|\mathcal{F}_{i,t-1}\}$ with $z_{1it} := z_{it}\mathbf{1}_{it} - \mathbb{E}\{z_{it}\mathbf{1}_{it}|\mathcal{F}_{i,t-1}\}$, $\mathbf{1}_{it} := \mathbf{1}\{|z_{it}| \leq \theta_{i,nT}\}$ with $\theta_{i,nT} := \rho_i^{2T}T^{\frac{1}{4}}n^{\frac{1}{4}}$, $\overline{\mathbf{1}}_{it} := 1 - \mathbf{1}_{it}$, and $z_{2it} := z_{it}\overline{\mathbf{1}}_{it}$. Let $V_{iT} := \sum_{t=1}^{T}\mathbb{E}\left[z_{1it}^2\right]$ and $v_{i,nT} := \rho_i^{4T}T^{\frac{1+\gamma}{2}}\sqrt{n}$ as one truncation of V_{iT}.

$$\begin{aligned}
\mathbb{E}\left[V_{iT}^2\right] &= \mathbb{E}\left(\sum_{t=1}^{T}\mathbb{E}\left[z_{1it}^2|\mathcal{F}_{i,t-1}\right]\right)^2 \leq T\sum_{t=1}^{T}\mathbb{E}\left[\left\{\mathbb{E}\left[z_{1i,t-1}^2|\mathcal{F}_{i,t-1}\right]\right\}^2\right] \leq T\sum_{t=1}^{T}\mathbb{E}\left|z_{1it}\right|^4 \\
&\leq 16T\sum_{t=1}^{T}\mathbb{E}\left|z_{it}\right|^4 \leq DT^{1+\gamma}\rho_i^{8T},
\end{aligned}$$

where some positive constant $D > 0$. For the term related to z_{1it}, with any constant $d > 0$,

$$
\begin{aligned}
& n \max_{1\le i\le n} \Pr\left(\left|\frac{\rho_i^{-2T}}{T^\gamma}\sum_{t=1}^T z_{1it}\right| \ge d\right) \\
= \; & n \max_{1\le i\le n} \Pr\left(\left|\sum_{t=1}^T z_{1it}\right| \ge T^\gamma \rho_i^{2T} d,\ V_{iT} \le v_{i,nT}\right) + n \max_{1\le i\le n} \Pr\left(V_{iT} > v_{i,nT}\right) \qquad (33) \\
\le \; & \max_{1\le i\le n} \exp\left(\frac{-\rho_i^{4T}T^{2\gamma}d^2 + 2\left(\log n\right) v_{i,nT} + 2\left(\log n\right)\rho_i^{2T}T^\gamma d\theta_{i,nT}}{2v_{i,nT} + 2\rho_i^{2T}T^\gamma d\theta_{i,nT}}\right) + \max_{1\le i\le n} o\left(nv_{i,nT}^{-2}T^{1+\gamma}\rho_i^{8T}\right).
\end{aligned}
$$

The arguments for terms related to z_{2it} and $\mathbb{E}\{z_{2it}|\mathcal{F}_{i,t-1}\}$ are identical. Without losing generality, we just discuss z_{2it}.The uniform upper bound of z_{2it} is as,

$$
\begin{aligned}
n \max_{1\le i\le n} \Pr\left(\left|\frac{\rho_i^{-2T}}{T^\gamma}\sum_{t=1}^T z_{2it}\right| \ge d\right) \; & \le \; n \max_{1\le i\le n} \Pr\left(\max_{1\le t\le T}|z_{it}| \ge \theta_{i,nT}\right) \\
& \le \; \frac{nT}{\theta_{i,nT}^4} \max_{1\le i\le n}\max_{1\le t\le T} \mathbb{E}\left[|z_{it}|^4 \mathbf{1}\{|z_{it}| \ge \theta_{i,nT}\}\right] \\
& = \; \max_{1\le i\le n} o\left(\theta_{i,nT}^{-4} T n \rho_i^{8T}\right), \qquad (34)
\end{aligned}
$$

where the first inequality is due to the more relaxed restrictions of truncations and the second inequality is because of the Markov inequality. To make sure the three terms in both (33) and (34) are asymptotically negligible, we need the following restrictions: For each $i = 1, 2, ..., n$,

$$
\rho_i^{4T}T^{2\gamma}/\log n >> v_{i,nT},\ \rho_i^{2T}T^\gamma/\log n >> \theta_{i,nT},\ v_{i,nT} >> T^{\frac{1+\gamma}{2}}\rho_i^{8T}\sqrt{n},\ \theta_{i,nT} >> \rho_i^{2T}T^{\frac{1}{4}}n^{\frac{1}{4}}.
$$

Therefore the following rate restrictions are on request as, for each $i = 1, 2, ..., n$,

$$
T^{\frac{3\gamma-1}{2}} >> \sqrt{n}/\log n,\ \bar{\rho}^T T^{\gamma-\frac{1}{4}} >> n^{\frac{1}{4}}\log(n). \qquad (35)
$$

Rate restrictions (35) are ensured by Assumption 1 and the dominance of exponential rates. Therefore the choice of $d > 0$ is arbitrary, and we need $\widetilde{M}$ to be any positive value to make sure

$$
n \max_{1\le i\le n} \Pr\left(\left|\frac{1}{2c_i}\frac{\rho_i^{-2T}}{T^\gamma}\left(y_{i,T}^2 - \mathbb{E}y_{i,T}^2\right)\right| \ge \frac{\widetilde{M}}{5}\right).
$$

For the second term of (31), the Markov inequality is applied as,

$$
n \max_{1\le i\le n} \Pr\left(\left|\frac{\rho_i^{-2T+1}}{T^\gamma}\sum_{t=1}^T y_{i,t-1}u_{it}\right| \ge \frac{\widetilde{M}c_i}{5}\right) \le n \max_{1\le i\le n} \frac{25\mathbb{E}\left(\sum_{t=1}^T y_{i,t-1}u_{it}\right)^2}{\rho_i^{4T}T^{2\gamma}\widetilde{M}^2c_i^2} \le O\left(\frac{n}{\underline{\rho}^{2T}}\right) = o(1),
$$

due to the dominance of exponential rates. To make sure that the second term of (31) is converging to zero, $\widetilde{M}$ can be any positive constant.

For the third term of (31), similarly, by applying the Markov inequality, we have

$$n \max_{1\le i\le n} \Pr\left(\left|\frac{\rho_i^{-2T+1}}{T^\gamma}\sum_{t=1}^T u_{it}^2\right| \ge \frac{2\widetilde{M}c_i}{5}\right) \le n \max_{1\le i\le n} \frac{5\mathbb{E}\left(\sum_{t=1}^T u_{it}^2\right)}{2\rho_i^{2T}T^\gamma \widetilde{M}c_i} \le O\left(\frac{n}{\underline{\rho}^{2T}T^{\gamma-1}}\right) = o(1),$$

and $\widetilde{M}$ can be any positive constant. The asymptotically negligibility is implied by the dominant exponential rates.

For the fourth term of (31), note that $\overline{y}_{i,t-1}$ is not a martingale difference sequence. Therefore Freedman exponential inequality is not directly applicable for $\frac{1}{T}\sum_{t=1}^T y_{i,t-1}$. The following transformation is readily justified as,

$$\begin{aligned}\sum_{t=1}^T y_{i,t-1} &= \sum_{t=1}^T\sum_{s=1}^{t-1}\rho_i^{t-1-s}u_{is} = \sum_{s=1}^T\left(\sum_{t=s+1}^T \rho_i^{t-1-s}\right)u_{is} \\ &= \frac{T^\gamma}{c_i}\sum_{s=1}^T \rho_i^{T-2s-2}u_{is} - \frac{T^\gamma}{c_i}\sum_{s=1}^T \rho_i^{-1-s}u_{is}. \end{aligned} \tag{36}$$

The two terms of (36) are summations of martingale difference sequence. The exponential inequality for martingale is applicable to the decomposition of (36). By (36), we have

$$\begin{aligned} & n\max_{1\le i\le n}\Pr\left(\left|\frac{\rho_i^{-2T+1}}{T^{2\gamma-1}}\overline{y}_{i,-1}^2\right| \ge \frac{2c_i\widetilde{M}}{5}\right) \\ \le\; & n\max_{1\le i\le n}\Pr\left(\left|\frac{\rho_i^{-T}}{T^\gamma\sqrt{T^{-1}}}\overline{y}_{i,-1}\right| \ge \sqrt{\frac{2c_i\widetilde{M}}{5}}\right) \\ \le\; & n\max_{1\le i\le n}\Pr\left(\left|\frac{1}{\rho_i^T\sqrt{T}}\sum_{s=1}^T\rho_i^{T-2s-2}u_{is}\right| \ge \frac{c_i}{2}\sqrt{\frac{2c_i\widetilde{M}}{5}}\right) \\ & +n\max_{1\le i\le n}\Pr\left(\left|\frac{1}{\rho_i^T\sqrt{T}}\sum_{s=1}^T\rho_i^{-s-2}u_{is}\right| \ge \frac{c_i}{2}\sqrt{\frac{2c_i\widetilde{M}}{5}}\right). \end{aligned} \tag{37}$$

The second term of (37) is trivially bounded by Markov inequality: Define $M^* := \frac{c_i}{2}\sqrt{\frac{2c_i\widetilde{M}}{5}}$, and we have

$$n\max_{1\le i\le n}\Pr\left(\left|\frac{1}{\rho_i^T\sqrt{T}}\sum_{s=1}^T\rho_i^{-3}u_{is}\right| \ge M^*\right) \le \max_{1\le i\le n}\frac{n\mathbb{E}\left(\sum_{s=1}^T\rho_i^{-s-2}u_{is}\right)^2}{\rho_i^{2T}T(M^*)^2} \le O\left(\frac{n}{\underline{\rho}^{2T}T^{1-\gamma}}\right) = o(1).$$

For $n\max_{1\le i\le n}\Pr\left(\left|\frac{1}{\sqrt{T}}\sum_{s=1}^T\rho_i^{-2s-2}u_{is}\right| \ge \frac{c_i}{2}\sqrt{\frac{2c_i\widetilde{M}}{5}}\right)$ we apply martingale exponential inequality in Freedman (1975). Note the fact that for each $i = 1, 2, \ldots n$, the process $\{\rho_i^{T-2s-2}u_{it}\}_{t=1}^T$ is a martingale difference sequence adaptive to the filtration $\mathcal{F}_{i,t-1} := \sigma\{u_{i,t-1}, u_{i,t-2}, u_{i,t-3}, \ldots\}$ as $\mathbb{E}\{\rho_i^{T-2s-2}u_{it}|\mathcal{F}_{i,t-1}\} = 0$ for any $t = 1, 2, \ldots T$. Define $x_{it} :=$

$\rho_i^{T-2s-2}u_{it} =: x_{1it} + x_{2it} - \mathbb{E}\{x_{2it}|\mathcal{F}_{i,t-1}\}$ with $x_{1it} := x_{it}\mathbf{1}_{it} - \mathbb{E}\{x_{it}\mathbf{1}_{it}|\mathcal{F}_{i,t-1}\}$, $\mathbf{1}_{it} := \mathbf{1}\{|x_{it}| \leq \varphi_{nT}\}$ with $\varphi_{nT} := n^{\frac{1}{4}}T^{\frac{1}{4}}$, $\overline{\mathbf{1}}_{it} := 1 - \mathbf{1}_{it}$, and $x_{2it} := x_{it}\overline{\mathbf{1}}_{it}$. Let $W_{iT} := \sum_{t=1}^{T}\mathbb{E}\left[x_{1it}^2\right]$ and $w_{nT} := T^{\frac{1+\gamma}{2}}\sqrt{n}$ is one truncation of W_{iT}.

$$\begin{aligned}\mathbb{E}\left[W_{iT}^2\right] &= \mathbb{E}\left(\sum_{t=1}^{T}\mathbb{E}\big[x_{1it}^2|\mathcal{F}_{i,t-1}\big]\right)^2 \leq T\sum_{t=1}^{T}\mathbb{E}\left[\left\{\mathbb{E}\left[x_{1i,t-1}^2|\mathcal{F}_{i,t-1}\right]\right\}^2\right] \leq T\sum_{t=1}^{T}\mathbb{E}\left|x_{1it}\right|^4 \\ &\leq 16T\sum_{t=1}^{T}\mathbb{E}\left|x_{it}\right|^4 \leq BT^{1+\gamma},\end{aligned}$$

with some positive $B > 0$. For the term related to x_{1it}, there exist some constant $\eta > 0$,

$$\begin{aligned}& n\max_{1\leq i\leq n}\Pr\left(\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}x_{1it}\right| \geq \eta\right) \\ =\ & n\max_{1\leq i\leq n}\Pr\left(\left|\sum_{t=1}^{T}z_{1it}\right| \geq \sqrt{T}\eta,\ W_{iT} \leq w_{nT}\right) + n\max_{1\leq i\leq n}\Pr\left(W_{iT} > w_{nT}\right) \qquad (38) \\ \leq\ & \max_{1\leq i\leq n}\exp\left(\frac{-T\eta^2 + 2\left(\log n\right)w_{nT} + 2\left(\log n\right)\sqrt{T}\eta\varphi_{nT}}{2w_{nT} + 2\sqrt{T}\eta\varphi_{nT}}\right) + \max_{1\leq i\leq n}o\left(nw_{nT}^{-2}T^{1+\gamma}\right).\end{aligned}$$

The arguments for terms related to x_{2it} and $\mathbb{E}\{x_{2it}|\mathcal{F}_{i,t-1}\}$ are identical. Without losing generality, we only discuss the case of x_{2it} as

$$\begin{aligned}n\max_{1\leq i\leq n}\Pr\left(\left|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}x_{2it}\right| \geq \eta_{nT}\right) &\leq n\max_{1\leq i\leq n}\Pr\left(\max_{1\leq t\leq T}|x_{it}| \geq \varphi_{nT}\right) \\ &\leq \frac{nT}{\varphi_{nT}^4}\max_{1\leq i\leq n}\max_{1\leq t\leq T}\mathbb{E}\left[|x_{it}|^4\,\mathbf{1}\{|x_{it}| \geq \varphi_{nT}\}\right] \\ &= \max_{1\leq i\leq n}o\left(\varphi_{nT}^{-4}Tn\right), \qquad (39)\end{aligned}$$

where the first inequality is due to the more relaxed restrictions of truncations and the second inequality is because of the Markov inequality. In order to make sure terms in both (38) and (39) are asymptotically negligible, the following several restrictions are on request: For each $i = 1, 2, ..., n$,

$$T/\log n >> w_{nT},\ \sqrt{T}/\log n >> \varphi_{nT},\ w_{nT} >> T^{\frac{1+\gamma}{2}}\sqrt{n},\ \varphi_{nT} >> T^{\frac{1}{4}}n^{\frac{1}{4}}.$$

These rate restrictions are ensured by Assumption 1. In order to make sure the fourth term of (31) is asymptotically negligible, just make sure that $\frac{c}{2}\sqrt{\frac{2c\widetilde{M}}{5}} > \eta$. Since η is an arbitrarily value, $\widetilde{M}$ can be any positive value.

For the fifth term of (31), if $\widetilde{M} > \frac{5\sigma^2}{2\underline{c}^2}$ or $\widetilde{M} < \frac{5\sigma^2}{8\overline{c}^2}$,

$$n\max_{1\leq i\leq n}\Pr\left(\left|\frac{1}{2c_i}\frac{\rho_i^{-2T}}{T^\gamma}\mathbb{E}y_{i,T}^2\right| \geq \frac{\widetilde{M}}{5}\right) = 0.$$

All in all, in order to make sure

$$\max_{1\le i\le n} \Pr\left(\frac{1}{\rho_i^{2T}T^{\gamma}}\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2\right| \ge \widetilde{M}\right) = o\left(\frac{1}{n}\right),$$

we just need to make sure $\widetilde{M}$ is large enough as $\widetilde{M} > \frac{5\sigma^2}{2\underline{c}^2}$. ■

Lemma C.4 *If Assumption 1 holds, for any $\overline{M}$ satisfying $\overline{M} \le \frac{5\sigma^2}{8\overline{c}^2}$,*

$$\max_{1\le i\le n} \Pr\left(\frac{1}{\rho_i^{2T}T^{\gamma}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2 \le \overline{M}\right) = o\left(\frac{1}{n}\right)$$

The Proof of Lemma C.4: The proof follows the fashion in Lemma C.3 and accommodates the following decomposition

$$\begin{aligned}
n\max_{1\le i\le n} \Pr\left(\frac{1}{\rho_i^{2T}T^{2\gamma}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2 \le \overline{M}\right) \le\ & n\max_{1\le i\le n} \Pr\left(\frac{1}{\rho_i^{2T}T^{2\gamma}}\sum_{t=1}^{T}\left(\widetilde{y}_{i,t-1}^2 - \mathbb{E}\widetilde{y}_{i,t-1}^2\right) \le -\overline{M}\right) \\
& + n\max_{1\le i\le n} \Pr\left(\frac{1}{\rho_i^{2T}T^{2\gamma}}\sum_{t=1}^{T}\mathbb{E}\widetilde{y}_{i,t-1}^2 \le 2\overline{M}\right) \\
\le\ & n\max_{1\le i\le n} \Pr\left(\left|\frac{1}{2c_i}\frac{\rho_i^{-2T}}{T^{\gamma}}\left(y_{i,T}^2 - \mathbb{E}y_{i,T}^2\right)\right| \ge \frac{\overline{M}}{4}\right) \\
& + n\max_{1\le i\le n} \Pr\left(\left|\frac{1}{2c_i}\frac{2\rho_i^{-2T+1}}{T^{\gamma}}\sum_{t=1}^{T}y_{i,t-1}u_{it}\right| \ge \frac{\overline{M}}{4}\right) \\
& + n\max_{1\le i\le n} \Pr\left(\left|\frac{1}{2c_i}\frac{\rho_i^{-2T}}{T^{\gamma}}\sum_{t=1}^{T}u_{it}^2\right| \ge \frac{\overline{M}}{4}\right) \\
& + n\max_{1\le i\le n} \Pr\left(\left|\frac{1}{2c_i}\frac{1}{\rho_i^{2T}T^{2\gamma-1}}\overline{y}_{i,-1}^2\right| \ge \frac{\overline{M}}{4}\right) \\
& + n\max_{1\le i\le n} \Pr\left(\frac{1}{2c_i}\frac{\rho_i^{-2T}}{T^{\gamma}}\mathbb{E}y_{i,T}^2 \le 2\overline{M}\right). \qquad (40)
\end{aligned}$$

For any positive constant $\overline{M} > 0$, the first four terms in (40) are $o(1)$. The asymptotic negligibility is due to Lemma C.3. If we define $2\overline{M} < \frac{5\sigma^2}{8\overline{c}^2}$, then

$$n\max_{1\le i\le n} \Pr\left(\frac{1}{2c_i}\frac{\rho_i^{-2T}}{T^{\gamma}}\mathbb{E}y_{i,T}^2 \le 2\overline{M}\right) = 0.$$

In all, we complete the proof. ■

Lemma C.5 *Suppose Assumption 1 holds. For some $\eta > 0$, under joint convergence $(n,T)\to\infty$,*

$$\sup_{c\in\mathcal{N}_\eta}\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\left\{\widehat{g}_i(c) \ne g_i^0\right\} \xrightarrow{p} 0.$$

The Proof of Lemma C.5: For the definition of $\widehat{g}_i(\cdot)$, we have for all $g \in \{1,2,...,K^0\}$,

$$\mathbf{1}\{\widehat{g}_i(c)=g\} \le \mathbf{1}\left\{\sum_{t=1}^T(\widetilde{y}_{it}-\widetilde{y}_{i,t-1}\rho_{gn})^2 \le \sum_{t=1}^T(\widetilde{y}_{it}-\widetilde{y}_{i,t-1}\rho_{g_i^0 n})^2\right\}.$$

Define $\rho_{g_i^0 n} := \exp\left(c_{g_i^0 n}/T^\gamma\right)$ and $\rho_{g_i n} := \exp\left(\frac{c_{g_i n}}{T^\gamma}\right)$. For simplicity, we write $\rho_i^0 \left(:= \exp\left(\frac{c_i^0}{T^\gamma}\right)\right)$ as ρ_i. We derive the following transformation

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^n \mathbf{1}\{\widehat{g}_i(c)=g\} &= \sum_{g=1}^{K^0}\frac{1}{n}\sum_{i=1}^n \mathbf{1}\{g_i^0\neq g\}\mathbf{1}\{\widehat{g}_i(c)=g\}\\
&\le \sum_{g=1}^{K^0}\frac{1}{n}\sum_{i=1}^n \mathbf{1}\{g_i^0\neq g\}\mathbf{1}\left\{\sum_{t=1}^T(\widetilde{y}_{it}-\widetilde{y}_{i,t-1}\rho_{gn})^2 \le \sum_{t=1}^T(\widetilde{y}_{it}-\widetilde{y}_{i,t-1}\rho_{g_i^0 n})^2\right\}\\
&=: \sum_{g=1}^{K^0}\frac{1}{n}\sum_{i=1}^n Z_{ig}(c),
\end{aligned}$$

where $Z_{ig}(c) := \mathbf{1}\{g_i^0\neq g\}\mathbf{1}\left\{\sum_{t=1}^T(\widetilde{y}_{it}-\widetilde{y}_{i,t-1}\rho_{gn})^2 \le \sum_{t=1}^T(\widetilde{y}_{it}-\widetilde{y}_{i,t-1}\rho_{g_i^0 n})^2\right\}$. We intend to bound $Z_{ig}(c)$ for all $c\in\mathcal{N}_\eta$ by the quantity irrelevant to c. Therefore for all i, it has

$$\begin{aligned}
Z_{ig}(c) &\le \max_{\widetilde{g}\neq g}\mathbf{1}\left\{\sum_{t=1}^T\left(\widetilde{y}_{it}-\widetilde{y}_{i,t-1}\rho_{gn}\right)^2 \le \sum_{t=1}^T\left(\widetilde{y}_{it}-\widetilde{y}_{i,t-1}\rho_{\widetilde{g}n}\right)^2\right\}\\
&= \max_{\widetilde{g}\neq g}\mathbf{1}\left\{\sum_{t=1}^T\widetilde{y}_{i,t-1}\left(\rho_{\widetilde{g}n}-\rho_{gn}\right)\left(2\widetilde{y}_{i,t-1}\rho^0_{\widetilde{g}n}+2\widetilde{u}_{it}-\widetilde{y}_{i,t-1}\left(\rho_{\widetilde{g}n}+\rho_{gn}\right)\right)\le 0\right\}.
\end{aligned}$$

Let us define

$$\begin{aligned}
H_T &:= \left|\begin{array}{c}\sum_{t=1}^T\widetilde{y}_{i,t-1}\left(\rho_{\widetilde{g}n}-\rho_{gn}\right)\left(2\widetilde{y}_{i,t-1}\rho^0_{\widetilde{g}n}+2\widetilde{u}_{it}-\widetilde{y}_{i,t-1}\left(\rho_{\widetilde{g}n}+\rho_{gn}\right)\right)\\ -\sum_{t=1}^T\widetilde{y}_{i,t-1}\left(\rho^0_{\widetilde{g}n}-\rho^0_{gn}\right)\left(2\widetilde{y}_{i,t-1}\rho^0_{\widetilde{g}n}+2\widetilde{u}_{it}-\widetilde{y}_{i,t-1}\left(\rho^0_{\widetilde{g}n}+\rho^0_{gn}\right)\right)\end{array}\right|\\
&\le \left|2\sum_{t=1}^T(\rho_{\widetilde{g}n}-\rho_{gn})\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right| + \left|2\sum_{t=1}^T(\rho^0_{\widetilde{g}n}-\rho^0_{gn})\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right|\\
&\quad+\left|\begin{array}{c}\sum_{t=1}^T\widetilde{y}_{i,t-1}\left(\rho_{\widetilde{g}n}-\rho_{gn}\right)\left(2\widetilde{y}_{i,t-1}\rho^0_{\widetilde{g}n}-\widetilde{y}_{i,t-1}\left(\rho_{\widetilde{g}n}+\rho_{gn}\right)\right)\\ -\sum_{t=1}^T\widetilde{y}_{i,t-1}\left(\rho^0_{\widetilde{g}n}-\rho^0_{gn}\right)\left(2\widetilde{y}_{i,t-1}\rho^0_{\widetilde{g}n}-\widetilde{y}_{i,t-1}\left(\rho^0_{\widetilde{g}n}+\rho^0_{gn}\right)\right)\end{array}\right|\\
&=: H_{1T}+H_{2T}+H_{3T},
\end{aligned}$$

where $H_{1T} := \left|2\sum_{t=1}^T(\rho_{\widetilde{g}n}-\rho_{gn})\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right|$, $H_{2T} := \left|2\sum_{t=1}^T(\rho^0_{\widetilde{g}n}-\rho^0_{gn})\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right|$ and

$$H_{3T} := \left|\begin{array}{c}\sum_{t=1}^T\widetilde{y}_{i,t-1}\left(\rho_{\widetilde{g}n}-\rho_{gn}\right)\left(2\widetilde{y}_{i,t-1}\rho^0_{\widetilde{g}n}-\widetilde{y}_{i,t-1}\left(\rho_{\widetilde{g}n}+\rho_{gn}\right)\right)\\ -\sum_{t=1}^T\widetilde{y}_{i,t-1}\left(\rho^0_{\widetilde{g}n}-\rho^0_{gn}\right)\left(2\widetilde{y}_{i,t-1}\rho^0_{\widetilde{g}n}-\widetilde{y}_{i,t-1}\left(\rho^0_{\widetilde{g}n}+\rho^0_{gn}\right)\right)\end{array}\right|.$$

By the compactness of the parameter support, we have

$$H_{1T}=\left|2\sum_{t=1}^{T}(\rho_{\widetilde{g}n}-\rho_{gn})\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right|\leq 2\left|\rho_{\widetilde{g}n}-\rho_{gn}\right|\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right|\leq\frac{B_1}{T^{\gamma}}\sqrt{\eta}\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right|,$$

where B_1 is a constant independent of η and T. We have the above argument by the definition of η. Similarly we can justify that $H_{2T}\leq\frac{B_2}{T^{\gamma}}\sqrt{\eta}\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right|$ where B_2 is a constant independent of η and T. For H_{3T}, we have with B_3 as a constant independent of η and T,

$$\begin{aligned}H_{3T} &= \left|\begin{array}{c}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\left(\rho_{\widetilde{g}n}-\rho_{gn}\right)\left(2\widetilde{y}_{i,t-1}\rho_{\widetilde{g}n}^{0}-\widetilde{y}_{i,t-1}\left(\rho_{\widetilde{g}n}+\rho_{gn}\right)\right)\\ -\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\left(\rho_{\widetilde{g}n}^{0}-\rho_{gn}^{0}\right)\left(2\widetilde{y}_{i,t-1}\rho_{\widetilde{g}n}^{0}-\widetilde{y}_{i,t-1}\left(\rho_{\widetilde{g}n}^{0}+\rho_{gn}^{0}\right)\right)\end{array}\right|\\ &= \left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^{2}\right|\left|\rho_{\widetilde{g}n}^{0}\left(\rho_{\widetilde{g}n}-\rho_{\widetilde{g}n}^{0}-\rho_{gn}+\rho_{gn}^{0}\right)+\frac{1}{2}\left(\left(\rho_{\widetilde{g}n}^{0}\right)^{2}-\rho_{\widetilde{g}n}^{2}+\rho_{gn}^{2}-\left(\rho_{gn}^{0}\right)^{2}\right)\right|\\ &\leq \frac{B_3}{T^{\gamma}}\sqrt{\eta}\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^{2}\right|.\end{aligned}$$

By combining available results, we obtain the following,

$$\begin{aligned}Z_{ig}(c) &\leq \max_{\widetilde{g}\neq g}\mathbf{1}\left\{\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\left(\rho_{\widetilde{g}n}-\rho_{gn}\right)\left(2\widetilde{y}_{i,t-1}\rho_{\widetilde{g}n}^{0}+2\widetilde{u}_{it}-\widetilde{y}_{i,t-1}\left(\rho_{\widetilde{g}n}+\rho_{gn}\right)\right)\leq 0\right\}\\ &\leq \max_{\widetilde{g}\neq g}\mathbf{1}\left\{\begin{array}{c}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\left(\rho_{\widetilde{g}n}^{0}-\rho_{gn}^{0}\right)\left(2\widetilde{y}_{i,t-1}\rho_{\widetilde{g}n}^{0}+2\widetilde{u}_{it}-\widetilde{y}_{i,t-1}\left(\rho_{\widetilde{g}n}^{0}+\rho_{gn}^{0}\right)\right)\\ \leq\frac{B_1+B_2}{T^{\gamma}}\sqrt{\eta}\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right|+\frac{B_3}{T^{\gamma}}\sqrt{\eta}\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^{2}\right|\end{array}\right\}.\end{aligned}$$

Based on the following fact that

$$\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\left(\rho_{\widetilde{g}n}^{0}-\rho_{gn}^{0}\right)\left(2\widetilde{y}_{i,t-1}\rho_{\widetilde{g}n}^{0}-\widetilde{y}_{i,t-1}\left(\rho_{\widetilde{g}n}^{0}+\rho_{gn}^{0}\right)\right)=\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^{2}\left(\rho_{\widetilde{g}n}^{0}-\rho_{gn}^{0}\right)^{2},$$

we define the following argument as

$$\widetilde{Z}_{ig}:=\max_{\widetilde{g}\neq g}\mathbf{1}\left\{\begin{array}{c}\left(\rho_{\widetilde{g}n}^{0}-\rho_{gn}^{0}\right)^{2}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^{2}+2\left(\rho_{\widetilde{g}n}^{0}-\rho_{gn}^{0}\right)\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\\ \leq\frac{B_1+B_2}{T^{\gamma}}\sqrt{\eta}\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right|+\frac{B_3}{T^{\gamma}}\sqrt{\eta}\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^{2}\right|\end{array}\right\}.$$

Consequently, we can bound $Z_{ig}(c)$ by $\sup_{c\in\mathcal{N}_{\eta}}Z_{ig}(c)\leq\widetilde{Z}_{ig}$. Note that

$$\sup_{c\in\mathcal{N}_{\eta}}\frac{1}{n}\sum_{i=1}^{n}\left\{\widehat{g}_i(c)\neq g_i^0\right\}\leq\frac{1}{n}\sum_{i=1}^{n}\sum_{g=1}^{K^0}\widetilde{Z}_{ig}.$$

For $i=1,2,...,n$, $g_i^0=\widetilde{g}\neq g$, we have equivalent representations $\rho_i=\rho_{\widetilde{g}n}$. For all $g\in\{1,2,...K^0\}$, we have

$$
\begin{aligned}
&\Pr\left(\widetilde{Z}_{ig}=1\right)\\
\le\ & \sum_{\widetilde{g}\neq g}\Pr\left(\begin{array}{c}\left(\rho^0_{\widetilde{g}n}-\rho^0_{gn}\right)^2\sum_{t=1}^T\widetilde{y}^2_{i,t-1}+2\left(\rho^0_{\widetilde{g}n}-\rho^0_{gn}\right)^2\sum_{t=1}^T\widetilde{y}_{i,t-1}\widetilde{u}_{it}\\ \le\frac{B_1+B_2}{T^\gamma}\sqrt{\eta}\left|\sum_{t=1}^T\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right|+\frac{B_3}{T^\gamma}\sqrt{\eta}\left|\sum_{t=1}^T\widetilde{y}^2_{i,t-1}\right|\end{array}\right)\\
\le\ & \sum_{\widetilde{g}\neq g}\Pr\left(\begin{array}{c}2\left(\rho^0_{\widetilde{g}n}-\rho^0_{gn}\right)\sum_{t=1}^T\widetilde{y}_{i,t-1}\widetilde{u}_{it}\le-\left(\rho^0_{\widetilde{g}n}-\rho^0_{gn}\right)^2\sum_{t=1}^T\widetilde{y}^2_{i,t-1}\\ +\frac{B_1+B_2}{T^\gamma}\sqrt{\eta}\left|\sum_{t=1}^T\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right|+\frac{B_3}{T^\gamma}\sqrt{\eta}\left|\sum_{t=1}^T\widetilde{y}^2_{i,t-1}\right|\end{array}\right)\\
\le\ & \sum_{\widetilde{g}\neq g}\Pr\left(\begin{array}{c}2\frac{\left(c^0_{\widetilde{g}n}-c^0_{gn}\right)}{T^\gamma}\sum_{t=1}^T\widetilde{y}_{i,t-1}\widetilde{u}_{it}\le-\left(c^0_{\widetilde{g}n}-c^0_{gn}\right)^2\left(\rho^0_{\widetilde{g}n}\right)^{2T}\overline{M}\\ +(B_1+B_2)\left(\rho^0_{\widetilde{g}n}\right)^{2T}\sqrt{\eta}M+B_3T^\gamma\left(\rho^0_{\widetilde{g}n}\right)^{2T}\sqrt{\eta}\widetilde{M}\end{array}\right)+\sum_{\widetilde{g}\neq g}\Pr\left(\frac{1}{\rho_i^{2T}T^\gamma}\sum_{t=1}^T\widetilde{y}^2_{i,t-1}\le\overline{M}\right)\\
&+\sum_{\widetilde{g}\neq g}\Pr\left(\frac{1}{\rho_i^{2T}T^\gamma}\left|\sum_{t=1}^T\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right|\ge M\right)+\sum_{\widetilde{g}\neq g}\Pr\left(\frac{1}{\rho_i^{2T}T^\gamma}\left|\sum_{t=1}^T\widetilde{y}^2_{i,t-1}\right|\ge\widetilde{M}\right)\\
\le\ & \sum_{\widetilde{g}\neq g}\Pr\left(\begin{array}{c}2\frac{c_{g,\widetilde{g}}}{T^\gamma}\sum_{t=1}^T\widetilde{y}_{i,t-1}\widetilde{u}_{it}\le-c^2_{g,\widetilde{g}}\left(\rho^0_{\widetilde{g}n}\right)^{2T}\overline{M}\\ +(B_1+B_2)\left(\rho^0_{\widetilde{g}n}\right)^{2T}\sqrt{\eta}M+B_3T^\gamma\left(\rho^0_{\widetilde{g}n}\right)^{2T}\sqrt{\eta}\widetilde{M}\end{array}\right)+\sum_{\widetilde{g}\neq g}\Pr\left(\frac{1}{\rho_i^{2T}T^\gamma}\sum_{t=1}^T\widetilde{y}^2_{i,t-1}\le\overline{M}\right)\\
&+\sum_{\widetilde{g}\neq g}\Pr\left(\frac{1}{\rho_i^{2T}T^\gamma}\left|\sum_{t=1}^T\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right|\ge M\right)+\sum_{\widetilde{g}\neq g}\Pr\left(\frac{1}{\rho_i^{2T}T^\gamma}\left|\sum_{t=1}^T\widetilde{y}^2_{i,t-1}\right|\ge\widetilde{M}\right). \qquad (41)
\end{aligned}
$$

Based on Lemma C.2, C.3 and C.4 we can argue the 2nd, 3rd and 4th terms of (41) are $o\left(\frac{1}{n}\right)$. We restrict η as $\eta\le\frac{c^2_{g,\widetilde{g}}\overline{M}}{2B_3T^\gamma\widetilde{M}}$. For example, we can set $\eta=\frac{c^2_{g,\widetilde{g}}\overline{M}}{4B_3T^\gamma\widetilde{M}}$. Therefore we have

$$
\begin{aligned}
\sum_{\widetilde{g}\neq g}\Pr\left(\begin{array}{c}2\frac{c_{g,\widetilde{g}}}{T^\gamma}\sum_{t=1}^T\widetilde{y}_{i,t-1}\widetilde{u}_{it}\le-c^2_{g,\widetilde{g}}\rho^{2T}_{\widetilde{g}n}\overline{M}\\ +(B_1+B_2)\rho^{2T}_{\widetilde{g}n}\sqrt{\eta}M+B_3T^\gamma\rho^{2T}_{\widetilde{g}n}\sqrt{\eta}\widetilde{M}\end{array}\right) &= \sum_{\widetilde{g}\neq g}\Pr\left(\begin{array}{c}2\frac{c_{g,\widetilde{g}}}{T^\gamma\left(\rho^0_{\widetilde{g}n}\right)^{2T}}\sum_{t=1}^T\widetilde{y}_{i,t-1}\widetilde{u}_{it}\le-c^2_{g,\widetilde{g}}\overline{M}\\ +(B_1+B_2)\sqrt{\eta}M+B_3T^\gamma\sqrt{\eta}\widetilde{M}\end{array}\right)\\
&\le \sum_{\widetilde{g}\neq g}\Pr\left(2\frac{c_{g,\widetilde{g}}}{T^\gamma\rho_i^{2T}}\sum_{t=1}^T\widetilde{y}_{i,t-1}\widetilde{u}_{it}\le\frac{-c^2_{g,\widetilde{g}}\overline{M}}{2}\right)\\
&\le \sum_{\widetilde{g}\neq g}\Pr\left(\frac{1}{T^\gamma\rho_i^{2T}}\left|\sum_{t=1}^T\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right|\ge\frac{c_{g,\widetilde{g}}\overline{M}}{4}\right)\\
&= o\left(\frac{1}{n}\right).
\end{aligned}
$$

The last equality is due to Lemma C.2. Combining the above results, we obtain $\Pr\left(\widetilde{Z}_{ig}=1\right)=o\left(\frac{1}{n}\right)$. This implies that

$$
\sup_{c\in\mathcal{N}_\eta}\frac{1}{n}\sum_{i=1}^n\mathbf{1}\left\{\widehat{g}_i(c)\neq g_i^0\right\}\le\frac{1}{n}\sum_{g=1}^{K^0}\sum_{i=1}^n\mathbb{E}\widetilde{Z}_{ig}=\frac{1}{n}\sum_{g=1}^{K^0}\sum_{i=1}^n\Pr\left(\widetilde{Z}_{ig}=1\right)=K^0\left(K^0-1\right)o\left(\frac{1}{n}\right)=o\left(\frac{1}{n}\right).
$$

■

The Proof of Theorem 4.2: (i) Note the definition of $\widehat{g}_i$, and we have

$$\Pr\left(\max_{1\le i\le n}\left|\widehat{g}_i(\widehat{c})-g_i^0\right|>0\right)\le \Pr\left(\widehat{c}\notin\mathcal{N}_\eta\right)+\mathbb{E}\left[\sup_{c\in\mathcal{N}_\eta}\Pr\left(\max_{1\le i\le n}\left|\widehat{g}_i(c)-g_i^0\right|>0\right)\right].$$

Based on the proof of Lemma C.5, we know that $\eta=O\left(\frac{1}{T^\gamma}\right)$ asymptotically. The convergence rates of our estimates are fast enough to satisfy this condition as $\widehat{c}_g-c_g^0=O_p\left(\frac{1}{\sqrt{n}\underline{\rho}^T}\right)=o_p\left(\frac{1}{T^\gamma}\right)$ for $g=1,2,...,K^0$. Therefore we derive the following argument as,

$$\Pr\left(\widehat{c}\notin\mathcal{N}_\eta\right)=o(1).$$

Besides,

$$\begin{aligned}\sup_{c\in\mathcal{N}_\eta}\Pr\left(\max_{1\le i\le n}\left|\widehat{g}_i(c)-g_i^0\right|>0\right) &\le n\sup_{c\in\mathcal{N}_\eta}\max_{1\le i\le n}\Pr\left(\left|\widehat{g}_i(c)-g_i^0\right|>0\right)\\ &= n\max_{1\le i\le n}\sup_{c\in\mathcal{N}_\eta}\Pr\left(\left|\widehat{g}_i(c)-g_i^0\right|>0\right)\\ &= n\cdot o\left(\frac{1}{n}\right)=o\left(1\right).\end{aligned}$$

We completed the proof.■

The Proof of Theorem 4.2: For the uniform consistency, observe that $\Pr\left(\bigcup_{g=1}^{K^0}\widehat{E}_{g,nT}\right)\le\sum_{g=1}^{K^0}\Pr\left(\widehat{E}_{g,nT}\right)\le\sum_{g=1}^{K^0}\sum_{i\in G_g^0}\Pr\left(\widehat{E}_{g,i}\right)$. We have

$$\begin{aligned}\sum_{g=1}^{K^0}\sum_{i\in G_g^0}\Pr\left(\widehat{E}_{g,i}\right) &\le n\max_{1\le i\le n}\mathbb{E}\mathbf{1}\{\widehat{g}_i(\widehat{c})\ne g_i^0\}=n\max_{1\le i\le n}\Pr\{\left|\widehat{g}_i(\widehat{c})-g_i^0\right|>0\}\\ &\le n\max_{1\le i\le n}\sup_{c\in\mathcal{N}_\eta}\Pr\{\left|\widehat{g}_i(c)-g_i^0\right|>0\}+n\max_{1\le i\le n}\Pr\{\left|\widehat{c}_i-c_i^0\right|>\eta\}\\ &= o(1)+n\max_{1\le i\le n}\Pr\{\left|\widehat{c}_i-c_i^0\right|>\eta\}.\end{aligned}\tag{42}$$

It remains to show that the second term of (42) is asymptotically negligible,

$$\begin{aligned}&n\max_{1\le i\le n}\Pr\{\left|\widehat{c}_i-c_i^0\right|>\eta\}\\ =\;& n\max_{1\le i\le n}\Pr\left\{\left|\left[\sum_{t=1}^T\widetilde{y}_{i,t-1}^2\right]^{-1}\left[\sum_{t=1}^T\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right]\right|\ge\eta\right\}\\ \le\;& n\max_{1\le i\le n}\Pr\left\{\frac{1}{\rho_i^{2T}T^{2\gamma}}\sum_{t=1}^T\widetilde{y}_{i,t-1}^2\le\frac{5\sigma^2}{8\overline{c}^2}\right\}\\ &+n\max_{1\le i\le n}\Pr\left\{\left|\frac{1}{\rho_i^{2T}T^\gamma}\sum_{t=1}^T\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right|\ge\eta T^\gamma\frac{5\sigma^2}{8\overline{c}^2}\right\}\\ =\;& o(1).\end{aligned}$$

where the last equality is based on Lemma C.2, C.3 and C.4. Note the fact that $\eta = O\left(\frac{1}{T^{\gamma}}\right)$. Then we successfully justify the argument of (i);

For (ii), we basically follow the derivations in Su et al. (2016). By Baye's theorem, we have the following decomposition as,

$$
\begin{aligned}
\Pr\left(\widehat{F}_{g,i}\right) &= 1-\Pr\left(i\in G_g^0 | i\in\widehat{G}_g\right)\\
&= \frac{\sum_{l=1,l\neq g}^{K^0}\Pr\left(i\in\widehat{G}_g|i\in G_g^0\right)\Pr\left(i\in G_l^0\right)}{\Pr\left(i\in\widehat{G}_g|i\in G_g^0\right)\Pr\left(i\in G_g^0\right)+\sum_{l=1,l\neq g}^{K^0}\Pr\left(i\in\widehat{G}_g|i\in G_l^0\right)\Pr\left(i\in G_l^0\right)}.
\end{aligned}
$$

We have

$$
\sum_{l=1,l\neq g}^{K^0}\Pr\left(i\in\widehat{G}_g|i\in G_l^0\right)\Pr\left(i\in G_l^0\right)\le\left(K^0-1\right)\sum_{l=1}^{K^0}\Pr\left(i\notin\widehat{G}_l|i\in G_l^0\right)=o(1).
$$

here the last equality is for the uniform consistency of (i). Note the fact that $\Pr\left(i\in\widehat{G}_g|i\in G_g^0\right) = 1-\Pr\left(i\notin\widehat{G}_g|i\in G_g^0\right)=1-o(1)$ uniformly in i and g by (i), and we have

$$
\Pr\left(i\in\widehat{G}_g|i\in G_g^0\right)\Pr\left(i\in G_g^0\right)+\sum_{l=1,l\neq g}^{K^0}\Pr\left(i\in\widehat{G}_g|i\in G_l^0\right)\Pr\left(i\in G_l^0\right)\ge\Pr\left(i\in G_g^0\right)/2,
$$

asymptotically. Therefore we have the following results

$$
\begin{aligned}
\Pr\left(\bigcup_{g=1}^{K^0}\widehat{F}_{g,nT}\right) &\le \sum_{g=1}^{K^0}\Pr\left(\widehat{F}_{g,nT}\right)\le\sum_{g=1}^{K^0}\sum_{i\in\widehat{G}_g}\Pr\left(\widehat{F}_{g,i}\right)\\
&\le \frac{\sum_{l=1,l\neq g}^{K^0}\Pr\left(i\in\widehat{G}_g|i\in G_g^0\right)\Pr\left(i\in G_l^0\right)}{\min_{1\le i\le n}\min_{1\le g\le K^0}\Pr\left(i\in G_g^0\right)/2}\\
&= \frac{o(1)}{\min_{1\le g\le K^0}\tau_g}=o(1),
\end{aligned}
$$

based on the fact that $\frac{n_g}{n}\to\pi_g$ for $g=1,2,...,K^0$. ■

C.1.2 Limiting distribution

The Proof of Theorem 4.4: Based on the definition of the post-classification estimator, we have for each $g\in\left\{1,2,...,K^0\right\}$ and $c_{gn}>0$

$$
\sqrt{n_g}T^{\gamma}(\rho_{gn}^0)^T\left(\widehat{\rho}_{\widehat{g}n}-\rho_{gn}^0\right)=\frac{\frac{1}{\sqrt{n_g}T^{\gamma}(\rho_{gn}^0)^T}\sum_{i\in\widehat{G}_{\widehat{g}}}\sum_{t=1}^T\widetilde{y}_{i,t-1}\widetilde{u}_{it}}{\frac{1}{n_gT^{2\gamma}(\rho_{gn}^0)^{2T}}\sum_{i\in\widehat{G}_{\widehat{g}}}\sum_{t=1}^T\widetilde{y}_{i,t-1}^2}.
$$

The numerator and denominator can be decomposed as

$$
\frac{1}{\sqrt{n_g}T^{\gamma}(\rho_{gn}^0)^T}\sum_{i\in\widehat{G}_{\widehat{g}}}\sum_{t=1}^T\widetilde{y}_{i,t-1}\widetilde{u}_{it}
$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{n_g}T^{\gamma}(\rho_{gn}^0)^T}\sum_{i\in G_g^0}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it} + \frac{1}{\sqrt{n_g}T^{\gamma}(\rho_{gn}^0)^T}\sum_{\substack{\widetilde{g}=1,\\ \widetilde{g}\neq g}}^{K^0}\sum_{\substack{i\in\widehat{G}_{\widehat{g}}\backslash G_g^0\\ i\in G_{\widetilde{g}}^0,\alpha_{\widetilde{g}}^0<\alpha_g^0}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\\
&\quad+\frac{1}{\sqrt{n_g}T^{\gamma}(\rho_{gn}^0)^T}\sum_{\substack{\widetilde{g}=1,\\ \widetilde{g}\neq g}}^{K^0}\sum_{\substack{i\in\widehat{G}_{\widehat{g}}\backslash G_g^0\\ i\in G_{\widetilde{g}}^0,\alpha_{\widetilde{g}}^0\geq\alpha_g^0}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it} - \frac{1}{\sqrt{n_g}T^{\gamma}(\rho_{gn}^0)^T}\sum_{i\in G_g^0\backslash\widehat{G}_{\widehat{g}}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it},
\end{aligned}
$$

and

$$
\begin{aligned}
&\frac{1}{n_gT^{2\gamma}(\rho_{gn}^0)^{2T}}\sum_{i\in\widehat{G}_g}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2\\
&= \frac{1}{n_gT^{2\gamma}(\rho_{gn}^0)^{2T}}\sum_{i\in G_g^0}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2 + \frac{1}{n_gT^{2\gamma}(\rho_{gn}^0)^{2T}}\sum_{\substack{\widetilde{g}=1,\\ \widetilde{g}\neq g}}^{K^0}\sum_{\substack{i\in\widehat{G}_g\backslash G_g^0\\ i\in G_{\widetilde{g}}^0,\alpha_{\widetilde{g}}^0<\alpha_g^0}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2\\
&\quad+\frac{1}{n_gT^{2\gamma}(\rho_{gn}^0)^{2T}}\sum_{\substack{\widetilde{g}=1,\\ \widetilde{g}\neq g}}^{K^0}\sum_{\substack{i\in\widehat{G}_g\backslash G_g^0\\ i\in G_{\widetilde{g}}^0,\alpha_{\widetilde{g}}^0\geq\alpha_g^0}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2 - \frac{1}{n_gT^{2\gamma}(\rho_{gn}^0)^{2T}}\sum_{i\in G_g^0\backslash\widehat{G}_g}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2.
\end{aligned}
$$

Therefore it remains to justify these arguments:

(i) For any $g = 1, 2, ..., K^0$,

$$
\begin{aligned}
\frac{1}{\sqrt{n_g}T(\rho_{gn}^0)^T}\sum_{\substack{i\in\widehat{G}_g\backslash G_g^0\\ i\in G_{\widetilde{g}}^0,\alpha_{\widetilde{g}}^0<\alpha_g^0}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it} &= o_p(1),\\
\frac{1}{n_gT^{2\gamma}(\rho_{gn}^0)^{2T}}\sum_{\substack{i\in\widehat{G}_g\backslash G_g^0\\ i\in G_{\widetilde{g}}^0,\alpha_{\widetilde{g}}^0<\alpha_g^0}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2 &= o_p(1);
\end{aligned}
$$

(ii) For any $g = 1, 2, ..., K^0$,

$$
\begin{aligned}
\frac{1}{\sqrt{n_g}T^{\gamma}(\rho_{gn}^0)^T}\sum_{\substack{i\in\widehat{G}_g\backslash G_g^0\\ i\in G_{\widetilde{g}}^0,\alpha_{\widetilde{g}}^0\geq\alpha_g^0}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it} &= o_p(1),\\
\frac{1}{n_gT^{2\gamma}(\rho_{gn}^0)^{2T}}\sum_{\substack{i\in\widehat{G}_g\backslash G_g^0\\ i\in G_{\widetilde{g}}^0,\alpha_{\widetilde{g}}^0\geq\alpha_g^0}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2 &= o_p(1);
\end{aligned}
$$

(iii) For any $g = 1, 2, ..., K^0$,

$$
\frac{1}{\sqrt{n_g}T^{\gamma}(\rho_{gn}^0)^T}\sum_{i\in G_g^0\backslash\widehat{G}_{\widehat{g}}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it} = o_p(1),
$$

$$\frac{1}{n_g T^{2\gamma}(\rho_{gn}^0)^{2T}} \sum_{i\in G_g^0\backslash\widehat{G}_{\widetilde{g}}} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}^2 \quad = \quad o_p(1).$$

For (i) (ii) and (iii), second terms can be proved identically as the first ones. Without losing generality we just focus on the first terms. For (iii), we have for any $\varepsilon > 0$, and for any $g = 1, 2, ..., K^0$

$$\Pr\left(\left|\frac{1}{\sqrt{n_g} T^{\gamma}(\rho_{gn}^0)^T} \sum_{i\in G_g^0\backslash\widehat{G}_g} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}\widetilde{u}_{it}\right| > \varepsilon\right) \le \Pr\left(\widehat{E}_{g,nT}\right) \to 0,$$

under joint asymptotics. For (i), for any $\widetilde{g} \neq g = 1, 2, ..., K^0$ and for any $\varepsilon > 0$, we have

$$\Pr\left(\left|\frac{1}{\sqrt{n_g} T^{\gamma}(\rho_{gn}^0)^T} \sum_{\substack{i\in\widehat{G}_g\backslash G_g^0 \\ i\in G_{\widetilde{g}}^0, \alpha_{\widetilde{g}}^0<\alpha_g^0}} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}\widetilde{u}_{it}\right| > \varepsilon\right) \le \Pr\left(\sum_{\substack{\widetilde{g}=1, \\ \widetilde{g}\neq g}}^{K^0} \widehat{F}_{\widetilde{g},nT}\right) \to 0,$$

under joint asymptotic framework. For (ii), for any $\widetilde{g} \neq g = 1, 2, ..., K^0$ and for any $\varepsilon > 0$, we have

$$\Pr\left(\left|\frac{1}{\sqrt{n_g} T^{\gamma}(\rho_{gn}^0)^T} \sum_{\substack{\widetilde{g}=1, \\ \widetilde{g}\neq g}}^{K^0} \sum_{\substack{i\in\widehat{G}_g\backslash G_g^0 \\ i\in G_{\widetilde{g}}^0, \alpha_{\widetilde{g}}^0\ge\alpha_g^0}} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}\widetilde{u}_{it}\right| > \varepsilon\right) \le \Pr\left(\sum_{\substack{\widetilde{g}=1, \\ \widetilde{g}\neq g}}^{K^0} \widehat{F}_{\widetilde{g},nT}\right) \to 0.$$

Summarizing above results, we have

$$\frac{1}{\sqrt{n_g} T^{\gamma}(\rho_{gn}^0)^T} \sum_{i\in\widehat{G}_g} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}\widetilde{u}_{it} = \frac{1}{\sqrt{n_g} T^{\gamma}(\rho_{gn}^0)^T} \sum_{i\in G_g^0} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}\widetilde{u}_{it} + o_p(1).$$

and

$$\frac{1}{n_g T^{2\gamma}(\rho_{gn}^0)^{2T}} \sum_{i\in\widehat{G}_g} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}^2 = \frac{1}{n_g T^{2\gamma}(\rho_{gn}^0)^{2T}} \sum_{i\in G_g^0} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}^2 + o_p(1).$$

The asymptotics for the post-classification estimator is asymptotically equivalent to the infeasible within estimator. The limiting behaviors of oracle estimators are established in the online supplement.■

C.2 Modified k-means classifications

Lemma C.6 *If $\gamma > \frac{3}{13}$ and Assumption 1 hold,*

$$\sup_{1\le i\le n, c_i>0} |\widehat{c}_i - c_i| = O_p\left((\overline{\rho})^{-T} T^{\frac{3-3\gamma}{4}}\right),$$

where $\overline{\rho} := \exp\left(\frac{\overline{c}}{T^{\gamma}}\right)$.

The Proof of Lemma C.6: If $c_i > 0$, there is one explosive root. The time series estimator is as,

$$\widehat{c}_i - c_i = T^{\gamma} \left(\sum_{t=1}^{T} \widetilde{y}_{i,t-1}^2 \right)^{-1} \left(\sum_{t=1}^{T} \widetilde{y}_{i,t-1} \widetilde{u}_{it} \right),$$

and

$$\begin{aligned} \sup_{1 \le i \le n} \left| \widehat{c}_i - c_i \right| &= T^{\gamma} \left(\inf_{1 \le i \le n} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}^2 \right)^{-1} \left(\sup_{1 \le i \le n} \left| \sum_{t=1}^{T} \widetilde{y}_{i,t-1} \widetilde{u}_{it} \right| \right) \\ &\le T^{\gamma} \left(\inf_{1 \le i \le n} \sum_{t=1}^{T} \mathbb{E} y_{i,t-1}^2 - \sup_{1 \le i \le n} \sum_{t=1}^{T} \left(\mathbb{E} y_{i,t-1}^2 - \widetilde{y}_{i,t-1}^2 \right) \right)^{-1} \left(\sup_{1 \le i \le n} \left| \sum_{t=1}^{T} \widetilde{y}_{i,t-1} \widetilde{u}_{it} \right| \right). \end{aligned}$$

(i) To justify the uniform upper bound for $\left| \sum_{t=1}^{T} \widetilde{y}_{i,t-1} \widetilde{u}_{it} \right|$, we apply the exponential inequality of Freedman (1975) as

$$\begin{aligned} & \Pr \left(\sup_{1 \le i \le n} \left| \sum_{t=1}^{T} \widetilde{y}_{i,t-1} \widetilde{u}_{it} \right| \ge M_{nT} \right) \\ \le & \Pr \left(\sup_{1 \le i \le n} \left| \sum_{t=1}^{T} y_{i,t-1} u_{it} \right| \ge \frac{M_{nT}}{2} \right) + \Pr \left(\sup_{1 \le i \le n} \left| T \overline{y}_{i,-1} \overline{u}_i \right| \ge \frac{M_{nT}}{2} \right) \\ \le & \underbrace{n \sup_{1 \le i \le n} \Pr \left(\left| \sum_{t=1}^{T} y_{i,t-1} u_{it} \right| \ge \frac{M_{nT}}{2} \right)}_{\text{(A.1)}} + \underbrace{n \sup_{1 \le i \le n} \Pr \left(\left| T \overline{y}_{i,-1} \overline{u}_i \right| \ge \frac{M_{nT}}{2} \right)}_{\text{(A.2)}}, \end{aligned}$$

where the innovations uniformly dominate the fixed effect. For (A.1) we apply the exponential inequality of Freedman (1975). The process $\{y_{i,t-1} u_{it}\}_{t=1}^{T}$ is a martingale difference sequence as $\mathbb{E}\left(y_{i,t-1} u_{it} | \mathcal{F}_{i,t-1}\right) = 0$ with $\mathcal{F}_{i,t-1} := \sigma \left\{ u_{i,t-1}, u_{i,t-2}, ... \right\}$. We set a truncation rate $d_{i,nT} >> \rho_i^T n^{\frac{1}{4}} T^{\frac{1}{4}}$. We define $z_{it} := y_{i,t-1} u_{it}$ and make the following decomposition $\sum_{t=1}^{T} z_{it} = \sum_{t=1}^{T} z_{1it} + \sum_{t=1}^{T} z_{2it} - \sum_{t=1}^{T} \mathbb{E}\left[z_{2it} | \mathcal{F}_{i,t-1} \right]$. Define $z_{1it} := z_{it} \mathbf{1}_{it} - \mathbb{E}\left[z_{it} \mathbf{1}_{it} | \mathcal{F}_{i,t-1} \right]$ and $z_{2it} := z_{it} \overline{\mathbf{1}}_{it}$. Define $\mathbf{1}_{it} := \mathbf{1} \left\{ |z_{it}| \le d_{i,nt} \right\}$ and $\overline{\mathbf{1}}_{it} = 1 - \mathbf{1}_{it}$. It suffices to show that

$$\begin{aligned} n \sup_{1 \le i \le n} \Pr \left(\left| \sum_{t=1}^{T} z_{1it} \right| \ge \frac{M_{nT}}{6} \right) &= o(1), \\ n \sup_{1 \le i \le n} \Pr \left(\left| \sum_{t=1}^{T} z_{2it} \right| \ge \frac{M_{nT}}{6} \right) &= o(1), \\ n \sup_{1 \le i \le n} \Pr \left(\left| \sum_{t=1}^{T} \mathbb{E}\left[z_{2it} | \mathcal{F}_{i,t-1} \right] \right| \ge \frac{M_{nT}}{6} \right) &= o(1). \end{aligned} \tag{43}$$

The second and third arguments of (43) share identical derivations, and without losing generality we only focus on the second term. We define $V_{iT} := \sum_{t=1}^{T} \mathbb{E}\left[z_{1it}^2 | \mathcal{F}_{i,t-1} \right]$, and

$v_{i,nT} >> \rho_i^{2T} n^{\frac{1}{2}} T^{\frac{1}{2}+\frac{\gamma}{2}}$ as a truncation rate for V_{iT}.

$$\begin{aligned}
\mathbb{E}\left[V_{iT}^2\right] &= \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{E}\left[z_{1it}^2 | \mathcal{F}_{i,t-1}\right]\right]^2 \leq T\sum_{t=1}^{T} \mathbb{E}\left[z_{1it}^4\right] \leq 16T\sum_{t=1}^{T} \mathbb{E}\left[z_{it}^4\right] \\
&\leq CT\sum_{t=1}^{T} \rho_i^{4t} = O_p\left(T^{1+\gamma}\rho_i^{4T}\right).
\end{aligned}$$

By Proposition 2.1 in Freedman (1975), we have

$$\begin{aligned}
& n \sup_{1\leq i\leq n} \Pr\left(\left|\sum_{t=1}^{T} z_{1it}\right| \geq \frac{M_{nT}}{6}\right) \\
\leq\; & n \sup_{1\leq i\leq n} \Pr\left(\left|\sum_{t=1}^{T} z_{1it}\right| \geq \frac{M_{nT}}{6}, V_{iT} \leq v_{i,nT}\right) + n \sup_{1\leq i\leq n} \Pr\left(V_{iT} > v_{i,nT}\right) \\
\leq\; & \sup_{1\leq i\leq n} \exp\left(\frac{-M_{nT}^2/36 + 2v_{i,nT}\log(n) + 4M_{nT}d_{i,nT}\log(n)/6}{2v_{i,nT} + 4M_{nT}d_{i,nT}/6}\right) + \sup_{1\leq i\leq n} o\left(nT^{1+\gamma}\rho_i^{4T} v_{i,nT}^{-2}\right) \\
=\; & o(1). \qquad (44)
\end{aligned}$$

To show asymptotic negligibility of (44), we need $M_{nT} >> \sup_{1\leq i\leq n} \rho_i^T n^{\frac{1}{4}} T^{\frac{1+\gamma}{4}} \sqrt{\log(n)}$ and $M_{nT} >> \sup_{1\leq i\leq n} \rho_i^T n^{\frac{1}{4}} T^{\frac{1}{4}} \log(n)$. The assumption for $v_{i,nT}$ ensures $o\left(nT^{1+\gamma}\rho_i^{4T} v_{i,nT}^{-2}\right) = o(1)$. For the second term of (43), we have

$$\begin{aligned}
& n \sup_{1\leq i\leq n} \Pr\left(\left|\sum_{t=1}^{T} z_{2it}\right| \geq \frac{M_{nT}}{6}\right) \\
\leq\; & n \sup_{1\leq i\leq n} \Pr\left(\max_{1\leq t\leq T} |z_{it}| \geq d_{i,nT}\right) \leq \frac{nT}{d_{i,nT}^4} \sup_{1\leq i\leq n} \max_{1\leq t\leq T} \mathbb{E}\left[|z_{it}|^4 1\left\{|z_{it}| > d_{i,nT}\right\}\right] \\
=\; & \sup_{1\leq i\leq n} \max_{1\leq t\leq T} o\left(\frac{nT}{d_{i,nT}^4}\rho_i^{4t}\right) = o(1),
\end{aligned}$$

which is guaranteed by our assumption for $d_{i,nT}$. It is easily justified that $\sup_{1\leq i\leq n} \rho_i^T n^{\frac{1}{4}} T^{\frac{1+\gamma}{4}} \sqrt{\log(n)} >> \sup_{1\leq i\leq n} \rho_i^T n^{\frac{1}{4}} T^{\frac{1}{4}} \log(n)$. Therefore, for (A.1) we need $M_{nT} >> \bar{\rho}^T n^{\frac{1}{4}} T^{\frac{1+\gamma}{4}+\epsilon} \sqrt{\log(n)}$ for any $\epsilon > 0$.

For (A.2) term, the uniform upper bound follows decompositions as

$$\sup_{1\leq i\leq n} \left|T\overline{y}_{i,-1}\overline{u}_i\right| \leq T \sup_{1\leq i\leq n} \left|\overline{y}_{i,-1}\right| \sup_{1\leq i\leq n} |\overline{u}_i|.$$

For $\sup_{1\leq i\leq n} |\overline{u}_i|$ term, the exponential inequality of Freedman (1975) is easily applied, and shows $\sup_{1\leq i\leq n} |\overline{u}_i| = O_p\left(T^{-\frac{1}{2}} n^{\frac{1}{4}} \sqrt{\log n}\right)$. For $\sup_{1\leq i\leq n} \left|\overline{y}_{i,-1}\right|$, we firstly apply the following transformation as

$$\sup_{1\leq i\leq n} \left|\overline{y}_{i,-1}\right| = \sup_{1\leq i\leq n} \left|\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{s=0}^{t-1} \rho_i^{t-1-s} u_{is}\right)\right| = \sup_{1\leq i\leq n} \left|\frac{1}{T}\sum_{s=0}^{T-1}\left(\sum_{t=s+1}^{T} \rho_i^{t-1-s}\right) u_{is}\right|,$$

where the innovations dominate the fixed effects. Hence the exponential inequality for $\frac{1}{T}\sum_{s=0}^{T-1}\left(\sum_{t=s+1}^{T}\rho_i^{t-1-s}\right)u_{is}$ is applicable and demonstrates that

$$\sup_{1\leq i\leq n}\left|\overline{y}_{i,-1}\right| = O_p\left(\overline{\rho}^T n^{\frac{1}{4}} T^{\frac{7\gamma-1}{4}}\sqrt{\log n}\right).$$

Combining above results, we have

$$\sup_{1\leq i\leq n}\left|T\overline{y}_{i,-1}\overline{u}_i\right| = O_p\left(\overline{\rho}^T n^{\frac{1}{2}} T^{\frac{7\gamma+1}{4}}\log n\right) >> \overline{\rho}^T n^{\frac{1}{4}} T^{\frac{1+\gamma}{4}}\sqrt{\log(n).}$$

Based on derivations for (A.1) and (A.2), for any $\epsilon > 0$ we have

$$\sup_{1\leq i\leq n}\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right| = O_p\left(\overline{\rho}^T n^{\frac{1}{2}} T^{\frac{7\gamma+1}{4}+\varepsilon}\log n\right). \tag{45}$$

(ii) To justify the uniform lower bound for the denominator $\inf_{1\leq i\leq n}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2$, we employ the following decomposition as

$$\inf_{1\leq i\leq n}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2 = \underbrace{\inf_{1\leq i\leq n}\sum_{t=1}^{T}\mathbb{E}y_{i,t-1}^2}_{\text{(B.1)}} - \underbrace{\sup_{1\leq i\leq n}\sum_{t=1}^{T}\left(\mathbb{E}y_{i,t-1}^2 - \widetilde{y}_{i,t-1}^2\right)}_{\text{(B.2)}}.$$

For (B.1) term, expectation operator removes randomness. We have

$$\inf_{1\leq i\leq n}\sum_{t=1}^{T}\mathbb{E}y_{i,t-1}^2 = O\left(\underline{\rho}^{2T}T^{2\gamma}\right).$$

For (B.2) term, the martingale exponential inequality is not directly applicable. Therefore, we employ the following decomposition,

$$\begin{aligned}
&\Pr\left(\sup_{1\leq i\leq n}\sum_{t=1}^{T}\left(\mathbb{E}y_{i,t-1}^2 - \widetilde{y}_{i,t-1}^2\right) \geq \widetilde{M}_{nT}\right)\\
\leq\ & n\sup_{1\leq i\leq n}\Pr\left(\sum_{t=1}^{T}\left(\mathbb{E}y_{i,t-1}^2 - \widetilde{y}_{i,t-1}^2\right) \geq \widetilde{M}_{nT}\right)\\
\leq\ & \underbrace{n\sup_{1\leq i\leq n}\Pr\left(\left|\frac{T^\gamma}{2c_i}\left(y_{i,T}^2 - \mathbb{E}y_{i,T}^2\right)\right| \geq \widetilde{M}_{nT}\right)}_{\text{(B.2.1)}} + \underbrace{n\sup_{1\leq i\leq n}\Pr\left(\left|\frac{\rho_i^T T^\gamma}{2c_i}\sum_{t=1}^{T}y_{i,t-1}u_{it}\right| \geq \widetilde{M}_{nT}\right)}_{\text{(B.2.2)}}\\
&+\underbrace{n\sup_{1\leq i\leq n}\Pr\left(\left|\frac{T^\gamma}{2c_i}\sum_{t=1}^{T}u_{it}^2\right| \geq \widetilde{M}_{nT}\right)}_{\text{(B.2.3)}} + \underbrace{n\sup_{1\leq i\leq n}\Pr\left(\left|\frac{T^{\gamma+1}}{2c_i}\overline{y}_{i,-1}^2\right| \geq \widetilde{M}_{nT}\right)}_{\text{(B.2.4)}},
\end{aligned}$$

where the innovations dominate the fixed effect uniformly. For (B.2.1) term, we have

$$n\sup_{1\leq i\leq n}\Pr\left(\left|\frac{T^\gamma}{2c_i}\left(y_{i,T}^2 - \mathbb{E}y_{i,T}^2\right)\right| \geq \widetilde{M}_{nT}\right) \leq n\sup_{1\leq i\leq n}\Pr\left(\left|\frac{T^\gamma}{2c_i}\sum_{t=1}^{T}\rho_i^{2(T-t)}\left(u_{it}^2 - \mathbb{E}u_{it}^2\right)\right| \geq \widetilde{M}_{nT}\right).$$

Note that $\left\{\rho_i^{2(T-t)}\left(u_{it}^2-\mathbb{E}u_{it}^2\right)\right\}_{t=1}^T$ is a martingale difference sequence as $\mathbb{E}\left(\rho_i^{2(T-t)}\left(u_{it}^2-\mathbb{E}u_{it}^2\right)|\mathcal{F}_{i,t-1}\right)=0$ with $\mathcal{F}_{i,t-1}:=\sigma\left\{u_{i,t-1},u_{i,t-2},...\right\}$. We set a truncation rate $\phi_{i,nT}>>\rho_i^{2T}n^{\frac{1}{4}}T^{\frac{1}{4}+\gamma}$. We define $x_{it}:=\rho_i^{2(T-t)}\left(u_{it}^2-\mathbb{E}u_{it}^2\right)$ and make the following decomposition $\sum\limits_{t=1}^T x_{it}=\sum\limits_{t=1}^T x_{it}\mathbf{1}_{it}+\sum\limits_{t=1}^T x_{2it}-\sum\limits_{t=1}^T\mathbb{E}\left[x_{2it}|\mathcal{F}_{i,t-1}\right]$. Define $x_{1it}:=x_{it}\mathbf{1}_{it}-\mathbb{E}\left[x_{it}\mathbf{1}_{it}|\mathcal{F}_{i,t-1}\right]$ and $x_{2it}:=x_{it}\overline{\mathbf{1}}_{it}$. Define $\mathbf{1}_{it}:=\mathbf{1}\left\{|x_{it}|\leq\phi_{i,nt}\right\}$ and $\overline{\mathbf{1}}_{it}=1-\mathbf{1}_{it}$. It suffices to show that

$$
\begin{aligned}
n\sup_{1\leq i\leq n}\Pr\left(\left|\sum_{t=1}^T x_{1it}\right|\geq\widetilde{M}_{nT}\right) &= o(1),\\
n\sup_{1\leq i\leq n}\Pr\left(\left|\sum_{t=1}^T x_{2it}\right|\geq\widetilde{M}_{nT}\right) &= o(1),\\
n\sup_{1\leq i\leq n}\Pr\left(\left|\sum_{t=1}^T\mathbb{E}\left[x_{2it}|\mathcal{F}_{i,t-1}\right]\right|\geq\widetilde{M}_{nT}\right) &= o(1). \qquad (46)
\end{aligned}
$$

The second and third arguments of (46) share identical derivations, and without losing generality we only focus on the second term. We define $\widetilde{V}_{iT}:=\sum_{t=1}^T\mathbb{E}\left[x_{1it}^2|\mathcal{F}_{i,t-1}\right]$, and $\widetilde{v}_{i,nT}>>\rho_i^{4T}n^{\frac{1}{2}}T^{\frac{1}{2}+\frac{\gamma}{2}}$ is a truncation rate for $\widetilde{V}_{iT}$. With some constant $\widetilde{C}>0$, we have

$$
\begin{aligned}
\mathbb{E}\left[\widetilde{V}_{iT}^2\right] &= \mathbb{E}\left[\sum_{t=1}^T\mathbb{E}\left[x_{1it}^2|\mathcal{F}_{i,t-1}\right]\right]^2\leq T\sum_{t=1}^T\mathbb{E}\left[x_{1it}^4\right]\leq 16T\sum_{t=1}^T\mathbb{E}\left[x_{it}^4\right]\\
&\leq \widetilde{C}T\sum_{t=1}^T\rho_i^{8t}=O_p\left(T^{1+\gamma}\rho_i^{8T}\right).
\end{aligned}
$$

By Proposition 2.1 in Freedman (1975), we have

$$
\begin{aligned}
& n\sup_{1\leq i\leq n}\Pr\left(\frac{T^\gamma}{2c_i}\left|\sum_{t=1}^T x_{1it}\right|\geq\widetilde{M}_{nT}\right)\\
\leq\ & n\sup_{1\leq i\leq n}\Pr\left(\frac{T^\gamma}{2c_i}\left|\sum_{t=1}^T x_{1it}\right|\geq\widetilde{M}_{nT},\widetilde{V}_{iT}\leq\widetilde{v}_{i,nT}\right)+n\sup_{1\leq i\leq n}\Pr\left(\widetilde{V}_{iT}>\widetilde{v}_{i,nT}\right)\\
\leq\ & \sup_{1\leq i\leq n}\exp\left(\frac{-4c_i^2\widetilde{M}_{nT}^2/T^{2\gamma}+2\widetilde{v}_{i,nT}\log(n)+4c_i\widetilde{M}_{nT}\phi_{i,nt}\log(n)/T^\gamma}{2\widetilde{v}_{i,nT}+4c_i\widetilde{M}_{nT}\phi_{i,nt}/T^\gamma}\right)+\sup_{1\leq i\leq n}o\left(nT^{1+\gamma}\rho_i^{8T}\widetilde{v}_{i,nT}^{-2}\right)\\
=\ & o(1). \qquad (47)
\end{aligned}
$$

By our assumptions for $\widetilde{v}_{i,nT}$, we have $\sup_{1\leq i\leq n}o\left(nT^{1+\gamma}\rho_i^{8T}\widetilde{v}_{i,nT}^{-2}\right)=o(1)$. The asymptotic negligibility of the exponential term in (47) follows

$$
\begin{aligned}
\widetilde{M}_{nT} &>> \sup_{1\leq i\leq n}T^\gamma\sqrt{\log(n)}T^{\frac{1+\gamma}{4}}\rho_i^{2T}n^{\frac{1}{4}},\\
\widetilde{M}_{nT} &>> \sup_{1\leq i\leq n}T^\gamma\log(n)T^{\frac{1}{4}}\rho_i^{2T}n^{\frac{1}{4}}.
\end{aligned}
$$

Since $\sup_{1\le i\le n} T^{\gamma}\sqrt{\log(n)}T^{\frac{1+\gamma}{4}}\rho_i^{2T}n^{\frac{1}{4}} >> \sup_{1\le i\le n} T^{\gamma}\log(n)T^{\frac{1}{4}}\rho_i^{2T}n^{\frac{1}{4}}$, we only require $\widetilde{M}_{nT} >> \sup_{1\le i\le n} T^{\gamma}\sqrt{\log(n)}T^{\frac{1+\gamma}{4}}\rho_i^{2T}n^{\frac{1}{4}}$. For the second term of (46), we show

$$\begin{aligned} n\sup_{1\le i\le n}\Pr\left(\frac{T^{\gamma}}{2c_i}\left|\sum_{t=1}^{T}x_{2it}\right|\ge \widetilde{M}_{nT}\right) &\le n\sup_{1\le i\le n}\Pr\left(\max_{1\le t\le T}|x_{it}|\ge\phi_{i,nt}\right)\\ &\le \frac{nT}{\phi_{i,nT}^4}\sup_{1\le i\le n}\max_{1\le t\le T}\mathbb{E}\left[|x_{it}|^4 1\left\{|x_{it}|>\phi_{i,nT}\right\}\right]\\ &= \sup_{1\le i\le n} o\left(\frac{nT}{\phi_{i,nT}^4}\rho_i^{8T}\right). \end{aligned}$$

The assumption for $\phi_{i,nt}$ ensures that $\sup_{1\le i\le n} o\left(\frac{nT}{\phi_{i,nT}^4}\rho_i^{8T}\right)=o(1)$.

Therefore, for (B.2.1) term, we have

$$\sup_{1\le i\le n}\left|\frac{T^{\gamma}}{2c_i}\left(y_{i,T}^2-\mathbb{E}y_{i,T}^2\right)\right| = O_p\left(\sup_{1\le i\le n}T^{\gamma+\epsilon}\sqrt{\log(n)}T^{\frac{1+\gamma}{4}}\rho_i^{2T}n^{\frac{1}{4}}\right),$$

for any $\epsilon>0$.

For (B.2.2) term, we have, for any $\epsilon>0$,

$$\sup_{1\le i\le n}\left|\frac{\rho_i^T T^{\gamma}}{2c_i}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right| = O_p\left(T^{\gamma}\overline{\rho}^{2T}n^{\frac{1}{4}}T^{\frac{1+\gamma}{4}+\epsilon}\sqrt{\log(n)}\right),$$

based on our derivations for the numerator.

For (B.2.3) term, we have

$$\sup_{1\le i\le n}\frac{T^{\gamma}}{2c_i}\sum_{t=1}^{T}u_{it}^2 = \sup_{1\le i\le n}\frac{T^{\gamma}}{2c_i}\sum_{t=1}^{T}\mathbb{E}u_{it}^2 - \sup_{1\le i\le n}\frac{T^{\gamma}}{2c_i}\sum_{t=1}^{T}\left(\mathbb{E}u_{it}^2-u_{it}^2\right).$$

We easily show $\sup_{1\le i\le n}\frac{T^{\gamma}}{2c_i}\sum_{t=1}^{T}\mathbb{E}u_{it}^2 = O\left(T^{1+\gamma}\right)$. Similarly by the law of iterated logarithm, we show $\sup_{1\le i\le n}\frac{T^{\gamma}}{2c_i}\sum_{t=1}^{T}\left(\mathbb{E}u_{it}^2-u_{it}^2\right)=o_p\left(T^{1+\gamma}\right)$. Therefore we show

$$\sup_{1\le i\le n}\frac{T^{\gamma}}{2c_i}\sum_{t=1}^{T}u_{it}^2 = O_p\left(T^{1+\gamma}\right),$$

dominated by exponential rates.

For (B.2.4) term, we have the following decomposition as

$$\begin{aligned} n\sup_{1\le i\le n}\Pr\left(\left|\frac{T^{\gamma+1}}{2c_i}\overline{y}_{i,-1}^2\right|\ge\overline{M}_{nT}\right) &\le n\sup_{1\le i\le n}\Pr\left(\left|\frac{T^{\frac{\gamma+1}{2}}}{\sqrt{2c_i}}\overline{y}_{i,-1}\right|\ge\sqrt{\overline{M}_{nT}}\right)\\ &= n\sup_{1\le i\le n}\Pr\left(\left|\frac{T^{\frac{\gamma-1}{2}}}{\sqrt{2c_i}}\sum_{t=1}^{T}\left(\sum_{s=0}^{t-1}\rho_i^{t-1-s}u_{is}\right)\right|\ge\sqrt{\overline{M}_{nT}}\right)\\ &= n\sup_{1\le i\le n}\Pr\left(\left|\frac{T^{\frac{\gamma-1}{2}}}{\sqrt{2c_i}}\sum_{s=0}^{T-1}\left(\sum_{t=s+1}^{T}\rho_i^{t-1-s}\right)u_{is}\right|\ge\sqrt{\overline{M}_{nT}}\right), \end{aligned}$$

where innovations dominate the fixed effect. Note that $\left\{\left(\sum_{t=s+1}^{T}\rho_i^{t-1-s}\right)u_{is}\right\}_{s=1}^{T}$ is a martingale difference sequence as $\mathbb{E}\left(\left(\sum_{t=s+1}^{T}\rho_i^{t-1-s}\right)u_{is}|\mathcal{F}_{i,s-1}\right) = 0$ with $\mathcal{F}_{i,s-1} := \sigma\left\{u_{i,s-1},u_{i,s-2},...\right\}$. We set a truncation rate $o_{i,nT} >> \rho_i^T n^{\frac{1}{4}}T^{\frac{1}{4}+\gamma}$. We define $\widetilde{x}_{it} := \left(\sum_{t=s+1}^{T}\rho_i^{t-1-s}\right)u_{is}$ and make the following decomposition $\sum\limits_{t=1}^{T}\widetilde{x}_{it} = \sum\limits_{t=1}^{T}\widetilde{x}_{1it} + \sum\limits_{t=1}^{T}\widetilde{x}_{2it} - \sum\limits_{t=1}^{T}\mathbb{E}\left[\widetilde{x}_{2it}|\mathcal{F}_{i,t-1}\right]$. Define $\widetilde{x}_{1it} := \widetilde{x}_{it}\mathbf{1}_{it} - \mathbb{E}\left[\widetilde{x}_{it}\mathbf{1}_{it}|\mathcal{F}_{i,t-1}\right]$ and $\widetilde{x}_{2it} := \widetilde{x}_{it}\overline{\mathbf{1}}_{it}$. Define $\mathbf{1}_{it} := \mathbf{1}\left\{|\widetilde{x}_{it}| \leq o_{i,nt}\right\}$ and $\overline{\mathbf{1}}_{it} = 1-\mathbf{1}_{it}$. It suffices to show that

$$\begin{aligned}
n\sup_{1\leq i\leq n}\Pr\left(\frac{T^{\frac{\gamma+1}{2}}}{\sqrt{2c_i}}\left|\sum_{t=1}^{T}\widetilde{x}_{1it}\right| \geq \sqrt{\overline{M}_{nT}}\right) &= o(1), \\
n\sup_{1\leq i\leq n}\Pr\left(\left|\sum_{t=1}^{T}\widetilde{x}_{2it}\right| \geq \sqrt{\overline{M}_{nT}}\right) &= o(1), \\
n\sup_{1\leq i\leq n}\Pr\left(\left|\sum_{t=1}^{T}\mathbb{E}\left[\widetilde{x}_{2it}|\mathcal{F}_{i,t-1}\right]\right| \geq \sqrt{\overline{M}_{nT}}\right) &= o(1). \qquad (48)
\end{aligned}$$

The second and third arguments of (48) share identical derivations, and without losing generality we only focus on the second term. We define $\overline{V}_{iT} := \sum_{t=1}^{T}\mathbb{E}\left[\widetilde{x}_{1it}^2|\mathcal{F}_{i,t-1}\right]$, and $\overline{v}_{i,nT} >> \rho_i^{2T}n^{\frac{1}{2}}T^{\frac{1}{2}+\frac{5\gamma}{2}}$ is a truncation rate for $\overline{V}_{iT}$. With some constant $\overline{C}>0$, we have

$$\begin{aligned}
\mathbb{E}\left[\overline{V}_{iT}^2\right] &= \mathbb{E}\left[\sum_{t=1}^{T}\mathbb{E}\left[\widetilde{x}_{1it}^2|\mathcal{F}_{i,t-1}\right]\right]^2 \leq T\sum_{t=1}^{T}\mathbb{E}\left[\widetilde{x}_{1it}^4\right] \leq 16T\sum_{t=1}^{T}\mathbb{E}\left[\widetilde{x}_{it}^4\right] \\
&\leq \overline{C}T^{1+4\gamma}\sum_{t=1}^{T}\rho_i^{4t} = O_p\left(T^{1+5\gamma}\rho_i^{4T}\right).
\end{aligned}$$

By Proposition 2.1 in Freedman (1975), we have

$$\begin{aligned}
& n\sup_{1\leq i\leq n}\Pr\left(\frac{T^{\frac{\gamma-1}{2}}}{\sqrt{2c_i}}\left|\sum_{t=1}^{T}\widetilde{x}_{1it}\right| \geq \sqrt{\overline{M}_{nT}}\right) \\
\leq\ & n\sup_{1\leq i\leq n}\Pr\left(\frac{T^{\frac{\gamma-1}{2}}}{\sqrt{2c_i}}\left|\sum_{t=1}^{T}\widetilde{x}_{1it}\right| \geq \sqrt{\overline{M}_{nT}}, \overline{V}_{iT}\leq \overline{v}_{i,nT}\right) + n\sup_{1\leq i\leq n}\Pr\left(\overline{V}_{iT} > \overline{v}_{i,nT}\right) \\
\leq\ & \sup_{1\leq i\leq n}\exp\left(\frac{-2c_i\overline{M}_{nT}T^{1-\gamma} + 2\overline{v}_{i,nT}\log(n) + 2\sqrt{2c_i}\sqrt{\overline{M}_{nT}}o_{i,nt}\log(n)T^{1-\gamma}}{2\overline{v}_{i,nT} + 2\sqrt{2c_i}\sqrt{\overline{M}_{nT}}o_{i,nt}T^{1-\gamma}}\right) \\
& + \sup_{1\leq i\leq n} o\left(nT^{1+5\gamma}\rho_i^{4T}\overline{v}_{i,nT}^{-2}\right) \\
=\ & o(1). \qquad (49)
\end{aligned}$$

By our assumptions for $\widetilde{v}_{i,nT}$, we have $\sup_{1\leq i\leq n} o\left(nT^{1+5\gamma}\rho_i^{4T}\overline{v}_{i,nT}^{-2}\right) = o(1)$. The asymptotic negligibility of the exponential term (49) is ensured by the following facts,

$$\overline{M}_{nT} >> \sup_{1\leq i\leq n} T^{\gamma-1}\left(\log(n)\right)^2 T^{\frac{1}{2}+2\gamma}\rho_i^{2T}n^{\frac{1}{2}},$$

$$\overline{M}_{nT} >> \sup_{1\le i\le n} T^{\gamma-1}\log(n)T^{\frac{5\gamma+1}{2}}\rho_i^{2T}n^{\frac{1}{2}}.$$

Since $\sup_{1\le i\le n} T^{\gamma-1}\log(n)T^{\frac{5\gamma+1}{2}}\rho_i^{2T}n^{\frac{1}{2}} >> \sup_{1\le i\le n} T^{\gamma-1}\left(\log(n)\right)^2 T^{\frac{1}{2}+2\gamma}\rho_i^{2T}n^{\frac{1}{2}}$, we only require $\overline{M}_{nT} >> \sup_{1\le i\le n} T^{\gamma-1}\log(n)T^{\frac{5\gamma+1}{2}}\rho_i^{2T}n^{\frac{1}{2}}$. For the second term of (48), we show

$$\begin{aligned} n\sup_{1\le i\le n}\Pr\left(\frac{T^{\frac{1-\gamma}{2}}}{2c_i}\left|\sum_{t=1}^{T}\widetilde{x}_{2it}\right| \ge \sqrt{\overline{M}_{nT}}\right) &\le n\sup_{1\le i\le n}\Pr\left(\max_{1\le t\le T}|\widetilde{x}_{it}| \ge o_{i,nt}\right) \\ &\le \frac{nT}{o_{i,nT}^4}\sup_{1\le i\le n}\max_{1\le t\le T}\mathbb{E}\left[|\widetilde{x}_{it}|^4 1\left\{|\widetilde{x}_{it}| > o_{i,nT}\right\}\right] \\ &= \sup_{1\le i\le n} o\left(\frac{nT}{o_{i,nT}^4}\rho_i^{4T}T^{4\gamma}\right). \end{aligned}$$

The assumption for $o_{i,nt}$ ensures that $\sup_{1\le i\le n} o\left(\frac{nT}{o_{i,nT}^4}\rho_i^{4T}T^{4\gamma}\right) = o(1)$.

Combining results of (B.1), (B.2.1), (B.2.2), (B.2.3), and (B.2.4), for any $\epsilon > 0$ and $\gamma > \frac{3}{13}$, we have

$$\inf_{1\le i\le n}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2 = O_p\left(T^{\gamma-1+\epsilon}\log(n)T^{\frac{5\gamma+1}{2}}\overline{\rho}^{2T}n^{\frac{1}{2}}\right). \tag{50}$$

All in all, based on (50) and (45), we have

$$\sup_{1\le i\le n}|\widehat{c}_i - c_i| = O_p\left((\overline{\rho})^{-T}T^{\frac{3-3\gamma}{4}}\right).$$

This concludes our proof.■

Lemma C.7 *If $\gamma < \frac{2}{3}$ and Assumption 1 hold,*

$$\sup_{1\le i\le n, c_i<0}|\widehat{c}_i - c_i| = O_p\left(\frac{n^{\frac{1}{4}+\epsilon}\sqrt{\log(n)}}{T^{\frac{\gamma}{2}+\frac{3}{4}}}\right) =: O_p(\delta_{nT}),$$

with arbitrary $\epsilon > 0$.

The Proof of Lemma C.7: Similarly, we demonstrate the uniform upper bounds and uniform lower bound of several terms as

$$\underbrace{\inf_{1\le i\le n}\sum_{t=1}^{T}\mathbb{E}\widetilde{y}_{i,t-1}^2,}_{\text{(C.1)}}\ \underbrace{\sup_{1\le i\le n}\sum_{t=1}^{T}\left|\mathbb{E}\widetilde{y}_{i,t-1}^2 - \widetilde{y}_{i,t-1}^2\right|,}_{\text{(C.2)}}\ \underbrace{\sup_{1\le i\le n}\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right|}_{\text{(C.3)}}.$$

For the term of (C.1), we have

$$\inf_{1\le i\le n}\sum_{t=1}^{T}\mathbb{E}\widetilde{y}_{i,t-1}^2 = O\left(T^{1+\gamma}\right),$$

since the expectation operator removes randomness. For (C.2), by applying the exponential inequality of Freedman (1975) or as in Su et al. (2016), we can readily show

$$\sup_{1\le i\le n}\sum_{t=1}^{T}\left|\mathbb{E}y_{i,t-1}^{2}-\widetilde{y}_{i,t-1}^{2}\right|=o_{p}\left(T^{1+\gamma}\right).$$

Therefore we show

$$\begin{aligned}\inf_{1\le i\le n}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^{2} &= \inf_{1\le i\le n}\sum_{t=1}^{T}\mathbb{E}\widetilde{y}_{i,t-1}^{2}-\sup_{1\le i\le n}\sum_{t=1}^{T}\left(\mathbb{E}\widetilde{y}_{i,t-1}^{2}-\widetilde{y}_{i,t-1}^{2}\right)\\ &= O_{p}\left(T^{1+\gamma}\right).\end{aligned}$$

For the term of (C.3), we accommodates the following decomposition as,

$$\begin{aligned}&\Pr\left(\sup_{1\le i\le n}\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right|\ge\check{M}_{nT}\right)\\ \le\ &\Pr\left(\sup_{1\le i\le n}\left|\sum_{t=1}^{T}y_{i,t-1}u_{it}\right|\ge\check{M}_{nT}\right)+\Pr\left(\sup_{1\le i\le n}\left|T\overline{y}_{i,-1}\overline{u}_{i}\right|\ge\check{M}_{nT}\right)\\ \le\ &\underbrace{n\sup_{1\le i\le n}\Pr\left(\left|\sum_{t=1}^{T}y_{i,t-1}u_{it}\right|\ge\check{M}_{nT}\right)}_{\text{(C.3.1)}}+\underbrace{n\sup_{1\le i\le n}\Pr\left(\left|T\overline{y}_{i,-1}\overline{u}_{i}\right|\ge\check{M}_{nT}\right)}_{\text{(C.3.2)}},\end{aligned}$$

where the innovations dominate the fixed effect uniformly. For (C.3.1) we apply the exponential inequality of Freedman (1975). The process $\{y_{i,t-1}u_{it}\}_{t=1}^{T}$ is a martingale difference sequence as $\mathbb{E}\left(y_{i,t-1}u_{it}|\mathcal{F}_{i,t-1}\right)=0$ with $\mathcal{F}_{i,t-1}:=\sigma\left\{u_{i,t-1},u_{i,t-2},...\right\}$. We set a truncation rate $\check{d}_{i,nT}>>n^{\frac{1}{4}}T^{\frac{1+\gamma}{4}}$. We define $\check{z}_{it}:=y_{i,t-1}u_{it}$ and make the following decomposition $\sum_{t=1}^{T}\check{z}_{it}=\sum_{t=1}^{T}\check{z}_{1it}+\sum_{t=1}^{T}\check{z}_{2it}-\sum_{t=1}^{T}\mathbb{E}\left[\check{z}_{2it}|\mathcal{F}_{i,t-1}\right]$. Define $\check{z}_{1it}:=\check{z}_{it}\mathbf{1}_{it}-\mathbb{E}\left[\check{z}_{it}\mathbf{1}_{it}|\mathcal{F}_{i,t-1}\right]$ and $\check{z}_{2it}:=\check{z}_{it}\overline{\mathbf{1}}_{it}$. Define $\mathbf{1}_{it}:=\mathbf{1}\left\{|\check{z}_{it}|\le\check{d}_{i,nT}\right\}$ and $\overline{\mathbf{1}}_{it}=1-\mathbf{1}_{it}$. It suffices to show that

$$\begin{aligned}n\sup_{1\le i\le n}\Pr\left(\left|\sum_{t=1}^{T}\check{z}_{1it}\right|\ge\check{M}_{nT}\right)&=o(1),\\ n\sup_{1\le i\le n}\Pr\left(\left|\sum_{t=1}^{T}\check{z}_{2it}\right|\ge\check{M}_{nT}\right)&=o(1),\\ n\sup_{1\le i\le n}\Pr\left(\left|\sum_{t=1}^{T}\mathbb{E}\left[\check{z}_{2it}|\mathcal{F}_{i,t-1}\right]\right|\ge\check{M}_{nT}\right)&=o(1).\end{aligned}\tag{51}$$

The second and third arguments of (51) share identical derivations, and without losing generality we only focus on the second term. We define $\check{V}_{iT}:=\sum_{t=1}^{T}\mathbb{E}\left[\check{z}_{1it}^{2}|\mathcal{F}_{i,t-1}\right]$, and $\check{v}_{i,nT}>>n^{\frac{1}{2}}T^{\frac{1}{2}+\gamma}$ is a truncation rate for $\check{V}_{iT}$.

$$\mathbb{E}\left[\check{V}_{iT}^{2}\right]=\mathbb{E}\left[\sum_{t=1}^{T}\mathbb{E}\left[\check{z}_{1it}^{2}|\mathcal{F}_{i,t-1}\right]\right]^{2}\le T\sum_{t=1}^{T}\mathbb{E}\left[\check{z}_{1it}^{4}\right]\le 16T\sum_{t=1}^{T}\mathbb{E}\left[\check{z}_{it}^{4}\right]$$

$$\leq \quad CT^{2\gamma+1} = O_p\left(T^{2\gamma+1}\right).$$

By Proposition 2.1 in Freedman (1975), we have

$$\begin{aligned}
& n \sup_{1\leq i\leq n} \Pr\left(\left|\sum_{t=1}^{T} \check{z}_{1it}\right| \geq \check{M}_{nT}\right) \\
\leq \quad & n \sup_{1\leq i\leq n} \Pr\left(\left|\sum_{t=1}^{T} \check{z}_{1it}\right| \geq \check{M}_{nT}, \check{V}_{iT} \leq \check{v}_{i,nT}\right) + n \sup_{1\leq i\leq n} \Pr\left(\check{V}_{iT} > \check{v}_{i,nT}\right) \\
\leq \quad & \sup_{1\leq i\leq n} \exp\left(\frac{-\check{M}_{nT}^2 + 2\check{v}_{i,nT}\log(n) + 4\check{M}_{nT}\check{d}_{i,nT}\log(n)}{2\check{v}_{i,nT} + 4\check{M}_{nT}\check{d}_{i,nT}}\right) + \sup_{1\leq i\leq n} o\left(nT^{1+2\gamma}\check{v}_{i,nT}^{-2}\right) \\
= \quad & o(1). \qquad (52)
\end{aligned}$$

To show asymptotic negligibility of (52), we need $M_{nT} >> n^{\frac{1}{4}}T^{\frac{1+2\gamma}{4}}\sqrt{\log(n)}$ and $M_{nT} >> n^{\frac{1}{4}}T^{\frac{1+\gamma}{4}}\log(n)$. Our assumption for $v_{i,nT}$ ensures $\sup_{1\leq i\leq n} o\left(nT^{1+2\gamma}\check{v}_{i,nT}^{-2}\right) = o(1)$. For the second term of (51), we have

$$\begin{aligned}
& n \sup_{1\leq i\leq n} \Pr\left(\left|\sum_{t=1}^{T} \check{z}_{2it}\right| \geq \check{M}_{nT}\right) \\
\leq \quad & n \sup_{1\leq i\leq n} \Pr\left(\max_{1\leq t\leq T} |\check{z}_{it}| \geq \check{d}_{i,nT}\right) \leq \frac{nT}{\check{d}_{i,nT}^4} \sup_{1\leq i\leq n} \max_{1\leq t\leq T} \mathbb{E}\left[|z_{it}|^4 1\left\{|z_{it}| > \check{d}_{i,nT}\right\}\right] \\
= \quad & \sup_{1\leq i\leq n} \max_{1\leq t\leq T} o\left(\frac{nT^{1+\gamma}}{\check{d}_{i,nT}^4}\right) = o(1),
\end{aligned}$$

which is guaranteed by our assumption for $d_{i,nT}$. Since $n^{\frac{1}{4}}T^{\frac{1+2\gamma}{4}}\sqrt{\log(n)} >> n^{\frac{1}{4}}T^{\frac{1+\gamma}{4}}\log(n)$, we need $\check{M}_{nT} >> n^{\frac{1}{4}}T^{\frac{1+2\gamma}{4}}\sqrt{\log(n)}$. Therefore we have, for any $\epsilon > 0$,

$$\sup_{1\leq i\leq n}\left|\sum_{t=1}^{T} y_{i,t-1}u_{it}\right| = O_p\left(n^{\frac{1}{4}}T^{\frac{1+2\gamma}{4}+\epsilon}\sqrt{\log(n)}\right).$$

For (C.3.2) term, we have

$$\sup_{1\leq i\leq n}\left|T\overline{y}_{i,-1}\overline{u}_i\right| = T\sup_{1\leq i\leq n}\left|\overline{y}_{i,-1}\right|\sup_{1\leq i\leq n}|\overline{u}_i|.$$

By applying exponential inequality, we have $\sup_{1\leq i\leq n}|T\overline{u}_i| = O_p\left(\sqrt{\log n}\, n^{\frac{1}{4}}T^{\frac{1}{2}}\right)$. Note that $\left\{\left(\sum_{t=s+1}^{T}\rho_i^{t-1-s}\right)u_{is}\right\}_{s=1}^{T}$ is a martingale difference sequence as

$$\mathbb{E}\left(\left(\sum_{t=s+1}^{T}\rho_i^{t-1-s}\right)u_{is}|\mathcal{F}_{i,s-1}\right) = 0,$$

with $\mathcal{F}_{i,s-1} := \sigma\{u_{i,s-1}, u_{i,s-2}, \ldots\}$. We set a truncation rate $\check{o}_{i,nT} >> n^{\frac{1}{4}}T^{\frac{1}{4}+\gamma}$. We define $\check{x}_{it} := \left(\sum_{t=s+1}^{T}\rho_i^{t-1-s}\right)u_{is}$ and make the following decomposition $\sum\limits_{t=1}^{T}\check{x}_{it} = \sum\limits_{t=1}^{T}\check{x}_{1it} +$

$\sum_{t=1}^{T} \check{x}_{2it} - \sum_{t=1}^{T} \mathbb{E}\left[\check{x}_{2it}|\mathcal{F}_{i,t-1}\right]$. Define $\check{x}_{1it} := \check{x}_{it}\mathbf{1}_{it} - \mathbb{E}\left[\check{x}_{it}\mathbf{1}_{it}|\mathcal{F}_{i,t-1}\right]$ and $\check{x}_{2it} := \check{x}_{it}\overline{\mathbf{1}}_{it}$. Define $\mathbf{1}_{it} := \mathbf{1}\left\{|\check{x}_{it}| \leq \check{o}_{i,nt}\right\}$ and $\overline{\mathbf{1}}_{it} = 1 - \mathbf{1}_{it}$. It suffices to show that

$$
\begin{aligned}
n \sup_{1\leq i\leq n} \Pr\left(\left|\sum_{t=1}^{T}\check{x}_{1it}\right| \geq T\check{f}_{nT}\right) &= o\left(1\right), \\
n \sup_{1\leq i\leq n} \Pr\left(\left|\sum_{t=1}^{T}\check{x}_{2it}\right| \geq T\check{f}_{nT}\right) &= o\left(1\right), \\
n \sup_{1\leq i\leq n} \Pr\left(\left|\sum_{t=1}^{T}\mathbb{E}\left[\check{x}_{2it}|\mathcal{F}_{i,t-1}\right]\right| \geq T\check{f}_{nT}\right) &= o\left(1\right). \qquad (53)
\end{aligned}
$$

The second and third arguments of (53) share identical derivations, and without losing generality we only focus on the second term. We define $\check{W}_{iT} := \sum_{t=1}^{T} \mathbb{E}\left[\check{x}_{1it}^{2}|\mathcal{F}_{i,t-1}\right]$, and $\check{w}_{i,nT} >> n^{\frac{1}{2}}T^{\frac{1}{2}+\frac{5\gamma}{2}}$ is a truncation rate for $\check{W}_{iT}$. With some constant $\check{C} > 0$, we have

$$
\begin{aligned}
\mathbb{E}\left[\check{W}_{iT}^{2}\right] &= \mathbb{E}\left[\sum_{t=1}^{T}\mathbb{E}\left[\check{x}_{1it}^{2}|\mathcal{F}_{i,t-1}\right]\right]^{2} \leq T\sum_{t=1}^{T}\mathbb{E}\left[\check{x}_{1it}^{4}\right] \leq 16T\sum_{t=1}^{T}\mathbb{E}\left[\check{x}_{it}^{4}\right] \\
&\leq \check{C}T^{1+4\gamma}\sum_{t=1}^{T}\rho_{i}^{4t} = O_{p}\left(T^{1+5\gamma}\right).
\end{aligned}
$$

By Proposition 2.1 in Freedman (1975), we have

$$
\begin{aligned}
& n \sup_{1\leq i\leq n} \Pr\left(\left|\sum_{t=1}^{T}\check{x}_{1it}\right| \geq T\check{f}_{nT}\right) \\
\leq\ & n \sup_{1\leq i\leq n} \Pr\left(\left|\sum_{t=1}^{T}\check{x}_{1it}\right| \geq T\check{f}_{nT}, \check{W}_{iT} \leq \check{w}_{i,nT}\right) + n \sup_{1\leq i\leq n} \Pr\left(\check{W}_{iT} > \check{w}_{i,nT}\right) \\
\leq\ & \sup_{1\leq i\leq n} \exp\left(\frac{-T^{2}\check{f}_{nT}^{2} + 2\check{w}_{i,nT}\log\left(n\right) + 4T\check{f}_{nT}\check{o}_{i,nt}\log\left(n\right)}{2\check{w}_{i,nT} + 4T\check{f}_{nT}\check{o}_{i,nt}}\right) + \sup_{1\leq i\leq n} o\left(nT^{1+5\gamma}\check{w}_{i,nT}^{-2}\right) \\
=\ & o(1). \qquad (54)
\end{aligned}
$$

To show asymptotic negligibility of (54), we need $\check{f}_{nT} >> n^{\frac{1}{4}}T^{\frac{5\gamma-3}{4}}\sqrt{\log\left(n\right)}$ and $\check{f}_{nT} >> n^{\frac{1}{4}}T^{\gamma-\frac{3}{4}}\log\left(n\right)$. Our assumption for $\check{w}_{i,nT}$ ensures $\sup_{1\leq i\leq n} o\left(nT^{1+5\gamma}\check{w}_{i,nT}^{-2}\right) = o(1)$. For the second term of (53), we have

$$
\begin{aligned}
& n \sup_{1\leq i\leq n} \Pr\left(\left|\sum_{t=1}^{T}\check{z}_{2it}\right| \geq T\check{f}_{nT}\right) \\
\leq\ & n \sup_{1\leq i\leq n} \Pr\left(\max_{1\leq t\leq T}\left|\check{x}_{it}\right| \geq \check{o}_{i,nt}\right) \leq \frac{nT}{\check{o}_{i,nT}^{4}} \sup_{1\leq i\leq n}\max_{1\leq t\leq T}\mathbb{E}\left[\left|\check{x}_{it}\right|^{4} 1\left\{\left|\check{x}_{it}\right| > \check{o}_{i,nt}\right\}\right] \\
=\ & \sup_{1\leq i\leq n}\max_{1\leq t\leq T} o\left(\frac{nT^{1+4\gamma}}{\check{o}_{i,nT}^{4}}\right) = o\left(1\right),
\end{aligned}
$$

which is guaranteed by our assumption for $\breve{o}_{i,nt}$. Since $n^{\frac{1}{4}}T^{\frac{5\gamma-3}{4}}\sqrt{\log(n)} >> n^{\frac{1}{4}}T^{\gamma-\frac{3}{4}}\log(n)$, we need $\breve{f}_{nT} >> n^{\frac{1}{4}}T^{\frac{5\gamma-3}{4}}\sqrt{\log(n)}$. This shows

$$\sup_{1\le i\le n}\left|T\overline{y}_{i,-1}\overline{u}_i\right| \le T\sup_{1\le i\le n}\left|\overline{y}_{i,-1}\right|\sup_{1\le i\le n}\left|\overline{u}_i\right| = O_p\left((\log n)\, n^{\frac{1}{2}}T^{\frac{5\gamma-1}{4}}\right).$$

With rate restriction $\gamma < \frac{2}{3}$, we have

$$\sup_{1\le i\le n}\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right| = O_p\left(n^{\frac{1}{4}+\epsilon}T^{\frac{1+2\gamma}{4}}\sqrt{\log(n)}\right),$$

and

$$\sup_{1\le i\le n, c_i<0}\left|\widehat{c}_i - c_i\right| = O_p\left(\frac{n^{\frac{1}{4}+\epsilon}\sqrt{\log(n)}}{T^{\frac{\gamma}{2}+\frac{3}{4}}}\right),$$

with arbitrary $\epsilon > 0$.■

Lemma C.8 *If $\frac{3}{13} < \gamma < \frac{2}{3}$ and suppose Assumption 1 hold,*

$$\sup_{1\le i\le n}\left|\widehat{c}_i - c_i\right| = O_p\left(\frac{n^{\frac{1}{4}+\epsilon}\sqrt{\log(n)}}{T^{\frac{\gamma}{2}+\frac{3}{4}}}\right) =: O_p(\delta_{nT}),$$

for arbitrary $\epsilon > 0$.

The Proof of Lemma C.8: The uniform convergence rate of individual least square estimator is shown as

$$\begin{aligned}
\sup_{1\le i\le n}\left|\widehat{c}_i - c_i\right| &= \max\left\{\sup_{1\le i\le n, c_i>0}\left|\widehat{c}_i - c_i\right|, \sup_{1\le i\le n, c_i<0}\left|\widehat{c}_i - c_i\right|\right\} \\
&= \max\left\{O_p\left(\frac{1}{\rho^T}\right), O_p\left(\frac{n^{\frac{1}{4}}\sqrt{\log(n)}}{T^{\frac{3}{4}+\frac{\gamma}{2}}}\right)\right\} \\
&= O_p\left(\frac{n^{\frac{1}{4}+\epsilon}\sqrt{\log(n)}}{T^{\frac{3}{4}+\frac{\gamma}{2}}}\right),
\end{aligned}$$

by Lemma C.6, Lemma C.7 and dominance of exponential rates. We conclude our proof. ■

The Proof of Lemma 4.1: (This proof follows derivations in Su et al. (2019)). Let $Q_n(\alpha) = \sum_{g=1}^{K^0}\min_{1\le l\le K}\left(c_{gn}^0 - \alpha_l\right)^2\pi_g$.

(i) Firstly we derive the convergence rate of $\widehat{Q}_n(\alpha) - Q_n(\alpha)$ uniformly over $\alpha \in \mathcal{M} := \left\{(\alpha_1, ..., \alpha_{K^0}) : \sup_{1\le g\le K^0}|\alpha_g| \le 2\overline{c}\right\}$ for some upper bound $\overline{c}$. Let $R_n = \sup_{1\le i\le n}\left|\widehat{c}_i - c_{g_i^0 n}^0\right|$. Then by assumptions, we have

$$R_n = \sup_{1\le i\le n}\left|\widehat{c}_i - c_{g_i^0 n}^0\right| \le O_p(\delta_{nT}) \le \overline{c} \text{ a.s.},$$

by our previous derivations. In addition, we have,

$$
\begin{aligned}
\left(\widehat{c}_i-\alpha_l\right)^2 &\geq \left(c_{g_i^0 n}^0-\alpha_l\right)^2-2\left|\left(c_{g_i^0 n}^0-\widehat{c}_i\right)\left(c_{g_i^0 n}^0-\alpha_l\right)\right|-\left(c_{g_i^0 n}^0-\widehat{c}_i\right)^2 \\
&\geq \left(c_{g_i^0 n}^0-\alpha_l\right)^2-2\left|\left(c_{g_i^0 n}^0-\widehat{c}_i\right)\left(c_{g_i^0 n}^0-\alpha_l\right)\right|-\left(\sup_{1\leq i\leq n}\left(c_{g_i^0 n}^0-\widehat{c}_i\right)\right)^2 \\
&\geq \left(c_{g_i^0 n}^0-\alpha_l\right)^2-2\left(\sup_{1\leq i\leq n}\left|c_{g_i^0 n}^0-\widehat{c}_i\right|\right)\left(\left|c_{g_i^0 n}^0-\alpha_l\right|\right)-\left(\sup_{1\leq i\leq n}\left(c_{g_i^0 n}^0-\widehat{c}_i\right)\right)^2 \\
&\geq \left(c_{g_i^0 n}^0-\alpha_l\right)^2-2\left(\sup_{1\leq i\leq n}\left|c_{g_i^0 n}^0-\widehat{c}_i\right|\right)\left(\left|c_{g_i^0 n}^0\right|+\left|\alpha_l\right|\right)-\left(\sup_{1\leq i\leq n}\left(c_{g_i^0 n}^0-\widehat{c}_i\right)\right)^2 .
\end{aligned}
$$

Taking $\min_{1\leq l\leq K^0}$ on both sides and take average over i, we have

$$
\frac{1}{n}\sum_{i=1}^{n}\min_{1\leq l\leq K^0}\left(\widehat{c}_i-\alpha_l\right)^2 \geq \frac{1}{n}\sum_{i=1}^{n}\min_{1\leq l\leq K^0}\left(c_{g_i^0 n}^0-\alpha_l\right)^2-2\delta_{nT}\left(2\overline{c}+\overline{c}\right)-\overline{c}\delta_{nT},
$$

where $\left(\sup_{1\leq i\leq n}\left(c_{g_i^0 n}^0-\widehat{c}_i\right)\right)^2\leq\overline{c}\delta_{nT}$ is of lower order term. Therefore we have $\widehat{Q}_n\left(\alpha\right)\geq Q_n\left(\alpha\right)-7\overline{c}\delta_{nT}$. Similarly we have

$$
\left(\widehat{c}_i-\alpha_l\right)^2 \leq \left(c_{g_i^0 n}^0-\alpha_l\right)^2+2\left(\sup_{1\leq i\leq n}\left|c_{g_i^0 n}^0-\widehat{c}_i\right|\right)\left(\left|c_{g_i^0 n}^0\right|+\left|\alpha_l\right|\right)+\left(\sup_{1\leq i\leq n}\left(c_{g_i^0 n}^0-\widehat{c}_i\right)\right)^2 .
$$

Hence we derive $\widehat{Q}_n\left(\alpha\right)\leq Q\left(\alpha\right)+7\overline{c}\delta_{nT}$ and $\sup_{\alpha\in\mathcal{M}}\left|\widehat{Q}_n\left(\alpha\right)-Q_n\left(\alpha\right)\right|\leq 7\overline{c}\delta_{nT}$.

(ii) Secondly, we show that $\widehat{\alpha}\in\mathcal{M}$. Denote $\widehat{\alpha}:=\{\widehat{\alpha}_1,\widehat{\alpha}_1,...,\widehat{\alpha}_{K^0}\}$. By our assumption, we have

$$
\sup_{1\leq i\leq n}\left|\widehat{c}_i\right|\leq\sup_{1\leq i\leq n}\left|\widehat{c}_i-c_{g_i^0 n}^0\right|+\sup_{1\leq i\leq n}\left|c_{g_i^0 n}^0\right|\leq 2\overline{c}.
$$

Denote $I_n\left(g\right):=\{i:g=\arg\min_{1\leq g\leq K}\left|\widehat{c}_i-\widehat{\alpha}_l\right|\}$ for some $g\leq K^0$. Here we use contradictions to demonstrate our results:

(ii.1) If $\left|\alpha_g\right|>2\overline{c}$ and $I_n\left(g\right)=\varnothing$, then we choose $\widehat{\alpha}^{'}:=\left\{\widehat{\alpha}_1,...,\widehat{\alpha}_{g-1},\widehat{\alpha}_g^{'},\widehat{\alpha}_{g+1}...,\widehat{\alpha}_{K^0}\right\}$, with $\widehat{\alpha}_g^{'}=\widehat{c}_i$ for each $i\in\{1,2,...,n\}$. Therefore we can get $\left|\widehat{\alpha}_g^{'}\right|\leq 2\overline{c}<\left|\widehat{\alpha}_g\right|$ and $\widehat{Q}_n\left(\widehat{\alpha}^{'}\right)<\widehat{Q}_n\left(\widehat{\alpha}\right)$. This demonstrates a contradiction.

(ii.2) If $\left|\alpha_g\right|>2\overline{c}$ and $I_n\left(g\right)\neq\varnothing$, then we choose $\widehat{\alpha}^{'}:=\left\{\widehat{\alpha}_1,...,\widehat{\alpha}_{g-1},\widehat{\alpha}_g^{'},\widehat{\alpha}_{g+1}...,\widehat{\alpha}_K\right\}$, with $\widehat{\alpha}_g^{'}=\frac{1}{\left|I_n(g)\right|}\sum_{i\in I_n(g)}\widehat{c}_i$ for each $i\in\{1,2,...,n\}$. Here $\left|I_n\left(g\right)\right|$ denotes the cardinality of $I_n\left(g\right)$. This shows $\left|\widehat{\alpha}_g^{'}\right|\leq 2\overline{c}<\left|\widehat{\alpha}_g\right|$ and $\widehat{Q}_n\left(\widehat{\alpha}^{'}\right)<\widehat{Q}_n\left(\widehat{\alpha}\right)$. This is a contradiction too.

Based on (ii.1) and (ii.2), $\left|\widehat{\alpha}_g\right|\leq 2\overline{c}$ for each $g\in\left\{1,2,...,K^0\right\}$.

(iii) We show for any $\eta>0$ $\inf_{\alpha:d_H\left(\alpha,c_n^0\right)>\eta}Q_n\left(\alpha\right)\geq\underline{M}\min\left\{\eta^2,c_{g,\widetilde{g}}^2\right\}$ where $c_n^0:=\left\{c_{1n}^0,c_{2n}^0,...,c_{K^0 n}^0\right\}$ and $\underline{M}\leq\pi_g\leq 1$ for $g=1,2,...,K^0$. If there exists some $l_o\in\left\{1,2,...,K^0\right\}$ and two indexes g_1 and g_2 such that

$$
l_o=\arg\min_{1\leq l\leq K}\left|c_{g_1 n}^0-\alpha_{l_o}\right|=\arg\min_{1\leq l\leq K}\left|c_{g_2 n}^0-\alpha_{l_o}\right|,
$$

then we have

$$\begin{aligned}
Q_n(\alpha) &\geq \pi_{g_1}\left(c^0_{g_1 n} - \alpha_{l_o}\right)^2 + \pi_{g_2}\left(c^0_{g_2 n} - \alpha_{l_o}\right)^2 \\
&\geq \underline{M}\left(\left|c^0_{g_1 n} - \alpha_{l_o}\right| + \left|c^0_{g_2 n} - \alpha_{l_o}\right|\right)^2 \geq \underline{M}\left(c^0_{g_1 n} - c^0_{g_2 n}\right)^2 \\
&\geq \underline{M} c^2_{g,\widetilde{g}}.
\end{aligned}$$

Besides, if there does not exist such an l_o, then there is one-to-one mapping h: $\left\{1, 2, ..., K^0\right\} \to \left\{1, 2, ..., K^0\right\}$ such that

$$h(g) = \arg\min_{1 \leq l \leq K} \left|c^0_{gn} - \alpha_l\right|.$$

Thus $Q_n(\alpha) = \sum_{g=1}^{K^0} \pi_g \left(c^0_{gn} - \alpha_{h(g)}\right)^2 \geq \left(\inf_{1 \leq g \leq K^0} \pi_g\right) d^2_H\left(\alpha, c^0_n\right) \geq \underline{M}\eta^2$. Then we show (iii).

(iv) Lastly, we show that $d_H\left(\widehat{\alpha}, c^0_n\right) \leq \frac{\sqrt{15\overline{c}\delta_{nT}}}{\sqrt{\underline{M}}}$ for some constant $D > 0$ and arbitrary $\epsilon > 0$. We have

$$\begin{aligned}
&\Pr\left(d_H\left(\widehat{\alpha}, c^0_n\right) \geq \frac{\sqrt{15\overline{c}\delta_{nT}}}{\sqrt{\underline{M}}}\right) \\
= &\Pr\left(d_H\left(\widehat{\alpha}, c^0_n\right) \geq \frac{\sqrt{15\overline{c}\delta_{nT}}}{\sqrt{\underline{M}}}, Q_n(\widehat{\alpha}) \geq Q_n\left(c^0_n\right) + \min\left\{\underline{M}c^2_{g,\widetilde{g}}, 15\overline{c}\delta_{nT}\right\}\right) \\
\leq &\Pr\left(Q_n(\widehat{\alpha}) \geq Q_n\left(c^0_n\right) + \min\left\{\underline{M}c^2_{g,\widetilde{g}}, 15\overline{c}\delta_{nT}\right\}\right) \\
\leq &\Pr\left(\widehat{Q}_n(\widehat{\alpha}) + R_n \geq \widehat{Q}_n\left(c^0_n\right) - R_n + \min\left\{\underline{M}c^2_{g,\widetilde{g}}, 15\overline{c}\delta_{nT}\right\}\right) \\
= &\Pr\left(2R_n \geq \widehat{Q}_n\left(c^0_n\right) - \widehat{Q}_n(\widehat{\alpha}) + \min\left\{\underline{M}c^2_{g,\widetilde{g}}, 15\overline{c}\delta_{nT}\right\}\right) \\
\leq &\Pr\left(2R_n \geq \min\left\{\underline{M}c^2_{g,\widetilde{g}}, 15\overline{c}\delta_{nT}\right\}\right) = o(1),
\end{aligned}$$

since $\widehat{Q}_n\left(c^0_n\right) - \widehat{Q}_n(\widehat{\alpha}) \geq 0$. Note the fact that for large enough n and T, we have

$$2R_n \leq 2 \cdot 7\overline{c}\delta_{nT} < 15\overline{c}\delta_{nT} < \underline{M}c^2_{g,\widetilde{g}}.$$

We complete the whole proof.■

The Proof of Theorem 4.5: (This proof follows derivations in Su et al. (2019)). By Assumption 1, under joint convergence framework $(n, T) \to \infty$, there is one-to-one mapping $F_n : \left\{1, 2, ..., K^0\right\} \to \left\{1, 2, ..., K^0\right\}$, such that

$$\sup_g \left|\widehat{\alpha}_g - c^0_{F_n(g)n}\right| \leq O_p\left(\frac{\sqrt{15\overline{c}\delta_{nT}}}{\sqrt{\underline{M}}}\right) = O_p\left(\delta^{\frac{1}{2}}_{nT}\right).$$

Without losing generality, we can assume that $F_n(g) = g$ such that

$$\overline{R}_n = \sup_g \left|\widehat{\alpha}_g - c^0_{F_n(g)n}\right| = \sup_g \left|\widehat{\alpha}_g - c^0_{gn}\right| \leq O_p\left(\frac{\sqrt{15\overline{c}\delta_{nT}}}{\sqrt{\underline{M}}}\right) = O_p\left(\delta^{\frac{1}{2}}_{nT}\right).$$

If $\widehat{g}_i \neq g^0_i$, then $\left|\widehat{c}_i - \widehat{\alpha}_{\widehat{g}_i}\right| \leq \left|\widehat{c}_i - \widehat{\alpha}_{g^0_i}\right|$. This, in conjunction with triangle inequality, implies that

$$\left|\widehat{\alpha}_{\widehat{g}_i} - \widehat{\alpha}_{g^0_i}\right| - \left|\widehat{c}_i - \widehat{\alpha}_{g^0_i}\right| \leq \left|\widehat{\alpha}_{\widehat{g}_i} - \widehat{c}_i\right| \leq \left|\widehat{c}_i - \widehat{\alpha}_{g^0_i}\right|.$$

It follows that $\left|\widehat{c}_i - \widehat{\alpha}_{g_i^0}\right| \geq \frac{1}{2}\left|\widehat{\alpha}_{\widehat{g}_i} - \widehat{\alpha}_{g_i^0}\right|$. Therefore we have,

$$\begin{aligned}
\delta_{nT} + \overline{R}_n &\geq \left|\widehat{c}_i - c^0_{g_i^0 n}\right| + \left|c^0_{g_i^0 n} - \widehat{\alpha}_{g_i^0}\right| \geq \left|\widehat{c}_i - \widehat{\alpha}_{g_i^0}\right| \geq \frac{1}{2}\left|\widehat{\alpha}_{\widehat{g}_i} - \widehat{\alpha}_{g_i^0}\right| \\
&\geq \frac{1}{2}\left|\left(\widehat{c}_{\widehat{g}_i n} - c^0_{g_i^0 n}\right) + \left(\widehat{\alpha}_{\widehat{g}_i} - c^0_{\widehat{g}_i n}\right) + \left(c^0_{g_i^0 n} - \widehat{\alpha}_{g_i^0}\right)\right| \\
&\geq \frac{1}{2}\left|\left(\widehat{c}_{\widehat{g}_i n} - c^0_{g_i^0 n}\right)\right| - \frac{1}{2}\left|\left(\widehat{\alpha}_{\widehat{g}_i} - c^0_{\widehat{g}_i n}\right)\right| - \frac{1}{2}\left|\left(c^0_{g_i^0 n} - \widehat{\alpha}_{g_i^0}\right)\right| \\
&\geq \frac{1}{2}\left|\left(\widehat{c}_{\widehat{g}_i n} - c^0_{g_i^0 n}\right)\right| - \sup_g \left|\left(\widehat{\alpha}_g - c^0_{gn}\right)\right| \\
&= \frac{1}{2}\left|\left(\widehat{c}_{\widehat{g}_i n} - c^0_{g_i^0 n}\right)\right| - \overline{R}_n \geq \frac{c_{g,\widetilde{g}}}{2} - \overline{R}_n.
\end{aligned}$$

This implies that $\mathbf{1}\left\{\widehat{g}_i \neq g_i^0\right\} \leq \mathbf{1}\left\{2\overline{R}_n + \delta_{nT} \geq \frac{c_{g,\widetilde{g}}}{2}\right\}$. Noting that the right hand side of the above term is independent of i, we have

$$\begin{aligned}
\Pr\left\{\sup_{1\leq i\leq n} \mathbf{1}\left\{\widehat{g}_i \neq g_i^0\right\} > 0\right\} &\leq \Pr\left\{2\overline{R}_n + \delta_{nT} \geq \frac{c_{g,\widetilde{g}}}{2}\right\} \\
&\leq \Pr\left\{\delta_{nT} + 2\delta_{nT}^{\frac{1}{2}} \geq \frac{c_{g,\widetilde{g}}}{2}\right\} \xrightarrow{p} 0,
\end{aligned}$$

as $(n,T) \to \infty$. This concludes our proof.■

The Proof of Theorem 4.6: For uniform consistency, it is sufficient to show

$$\sum_{g=1}^{K^0} \Pr\left(\widehat{E}_{g,nT}\right) = o(1).$$

The $\sum_{g=1}^{K^0} \Pr\left(\widehat{F}_{g,nT}\right) = o(1)$ follows the exact derivations as in Theorem 4.3. By Theorem 4.5 we know $\left|\widehat{c}_i - \widehat{\alpha}_{\widehat{g}_i}\right| \leq \left|\widehat{c}_i - \widehat{\alpha}_{g_i^0}\right|$. This implies that

$$\mathbf{1}\left\{\widehat{g}_i \neq g_i^0\right\} \leq \mathbf{1}\left\{\delta_{nT} + 2\overline{R}_n \geq \frac{c_{g,\widetilde{g}}}{2}\right\},$$

where $R_{nT} = \frac{\sqrt{15\bar{c}\delta_{nT}}}{\sqrt{\underline{M}}}$. Since we intend to show $n\sup_{1\leq i\leq n} \Pr\left(\left|\widehat{g}_i - g_i^0\right| > 0\right) = o(1)$, we need to show

$$n\sup_{1\leq i\leq n} \Pr\left(\left|\widehat{g}_i - g_i^0\right| > 0\right) \leq n\sup_{1\leq i\leq n} \Pr\left(\delta_{nT} + 2\overline{R}_n \geq \frac{c_{g,\widetilde{g}}}{2}\right) = o(1).$$

Indeed we have

$$\begin{aligned}
& n\sup_{1\leq i\leq n} \Pr\left(\left|\widehat{g}_i - g_i^0\right| > 0\right) \\
\leq\ & n\sup_{1\leq i\leq n} \Pr\left(\delta_{nT} \geq \frac{c_{g,\widetilde{g}}}{4}\right) + n\sup_{1\leq i\leq n} \Pr\left(R_n \geq \frac{c_{g,\widetilde{g}}}{8}\right) \\
=\ & o(1),
\end{aligned}$$

under the joint convergence $(n,T) \to \infty$. The reason why we can have the asymptotic negligibility is for the properties of deterministic trends. Hence, the type I error of modified k-means classification is asymptotically diminishing as

$$\begin{aligned}\sum_{g=1}^{K^0} \Pr\left(\widehat{E}_{g,nT}\right) &= \sum_{g=1}^{K^0} \sum_{i \in G_g^0} \Pr\left(\widehat{E}_{gnT,i}\right) \leq \sum_{g=1}^{K^0} \sum_{i \in G_g^0} \Pr\left(i \notin \widehat{G}_g | i \in G_g^0\right) \\ &\leq n \sup_{1 \leq i \leq n} \Pr\left(\left|\widehat{g}_i - g_i^0\right| > 0\right) = o\left(1\right).\end{aligned}$$

The proof for $\sum_{g=1}^{K^0} \Pr\left(\widehat{F}_{g,nT}\right) = o\left(1\right)$ is identical to Theorem 4.3. Then we complete the proof.■

The Proof of Theorem 4.7: For mixed-roots panel with latent groups, the difference between the post-classification estimator and the oracle estimator is asymptotically diminishing. For the explosive group, the proof here is identical to Theorem 4.4. For the stationary group, derivations follow similar procedures. Similar results are also shown in Phillips (2014).■

D Inference procedures & model selection

This section provides the proofs for inference procedures and model selection on the true group numbers.

D.1 Inference procedures: explosiveness detections

The Proof of Theorem 5.1: For any $g = 1, 2, ..., K^0$ with $\alpha_g^0 > 0$, we have

$$\begin{aligned}\widehat{\widetilde{u}}_{it} &= \widetilde{y}_{it} - \widehat{\rho}_{\widehat{g}n} \widetilde{y}_{i,t-1}, \\ \widetilde{u}_{it} &= \widetilde{y}_{it} - \rho_{gn}^0 \widetilde{y}_{i,t-1}.\end{aligned}$$

Therefore we have

$$\begin{aligned}\frac{1}{2n_{\widehat{g}}T} \sum_{i \in \widehat{G}_g} \sum_{t=1}^{T} \widehat{\widetilde{u}}_{it}^2 &= \frac{1}{2n_{\widehat{g}}T} \sum_{i \in \widehat{G}_{\widehat{g}}} \sum_{t=1}^{T} \widetilde{u}_{it}^2 + (\widehat{\rho}_{\widehat{g}n} - \rho_{gn}^0)^2 \frac{1}{2n_{\widehat{g}}T} \sum_{i \in \widehat{G}_g} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}^2 \\ &\quad - (\widehat{\rho}_{\widehat{g}n} - \rho_{gn}^0) \frac{1}{2n_{\widehat{g}}T} \sum_{i \in \widehat{G}_g} \sum_{t=1}^{T} \widetilde{y}_{i,t-1} \widetilde{u}_{it} \\ &= \frac{1}{2n_{\widehat{g}}T} \sum_{i \in G_g^0} \sum_{t=1}^{T} \widetilde{u}_{it}^2 + \frac{1}{2n_{\widehat{g}}T} \sum_{i \in \widehat{G}_g \backslash G_g^0} \sum_{t=1}^{T} \widetilde{u}_{it}^2 - \frac{1}{2n_{\widehat{g}}T} \sum_{i \in G_g^0 \backslash \widehat{G}_g} \sum_{t=1}^{T} \widetilde{u}_{it}^2 \\ &\quad + (\widehat{\rho}_{\widehat{g}n} - \rho_{gn}^0)^2 \frac{1}{2n_{\widehat{g}}T} \sum_{i \in G_g^0} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}^2 + (\widehat{\rho}_{\widehat{g}n} - \rho_{gn}^0)^2 \frac{1}{2n_{\widehat{g}}T} \sum_{i \in \widehat{G}_g \backslash G_g^0} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}^2\end{aligned}$$

$$-(\widehat{\rho}_{\widehat{g}n}-\rho_{gn}^{0})^{2}\frac{1}{2n_{\widehat{g}}T}\sum_{i\in G_{g}^{0}\backslash\widehat{G}_{g}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^{2}-(\widehat{\rho}_{\widehat{g}n}-\rho_{gn}^{0})\frac{1}{2n_{\widehat{g}}T}\sum_{i\in G_{g}^{0}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}$$

$$-(\widehat{\rho}_{\widehat{g}n}-\rho_{g}^{0})\frac{1}{2n_{\widehat{g}}T}\sum_{i\in\widehat{G}_{g}\backslash G_{g}^{0}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}+(\widehat{\rho}_{\widehat{g}n}-\rho_{gn}^{0})\frac{1}{2n_{\widehat{g}}T}\sum_{i\in G_{g}^{0}\backslash\widehat{G}_{g}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it} \quad (55)$$

To demonstrate the dominance of the first term in (55), we first show that $n_{\widehat{g}}-n_g=o_p(1)$. Note the fact that $\mathbf{1}\left\{i\in\widehat{G}_g\right\}-\mathbf{1}\left\{i\in G_g^0\right\}=\mathbf{1}\left\{i\in\widehat{G}_g\backslash G_g^0\right\}-\mathbf{1}\left\{i\in G_g^0\backslash\widehat{G}_g\right\}$. By Markov inequality, for any $\epsilon>0$, we have

$$\begin{aligned}\Pr\left(\left|n_{\widehat{g}}-n_g\right|>2\epsilon\right) &\leq \Pr\left(\sum_{i=1}^{n}\mathbf{1}\left\{i\in\widehat{G}_g\backslash G_g^0\right\}\geq\epsilon\right)+\Pr\left(\sum_{i=1}^{n}\mathbf{1}\left\{i\in G_g^0\backslash\widehat{G}_g\right\}\geq\epsilon\right)\\ &\leq \frac{1}{\varepsilon}\sum_{i=1}^{n}\Pr\left(\widehat{F}_{g,i}\right)+\frac{1}{\varepsilon}\sum_{i=1}^{n}\Pr\left(\widehat{E}_{g,i}\right)\\ &= \frac{1}{\varepsilon}\sum_{g=1}^{K^0}\sum_{i\in G_g^0}\Pr\left(\widehat{F}_{g,i}\right)+\frac{1}{\varepsilon}\sum_{g=1}^{K}\sum_{i\in G_g^0}\Pr\left(\widehat{E}_{g,i}\right)\\ &= o(1)\end{aligned}$$

The asymptotic negligibility of $\frac{1}{2n_{\widehat{g}}T}\sum_{i\in G_g^0\backslash\widehat{G}_g}\sum_{t=1}^{T}\widetilde{u}_{it}^2$, $(\widehat{\rho}_{\widehat{g}n}-\rho_{gn}^0)^2\frac{1}{2n_{\widehat{g}}T}\sum_{i\in G_g^0\backslash\widehat{G}_g}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2$ and $(\widehat{\rho}_{\widehat{g}n}-\rho_{gn}^0)\frac{1}{2n_gT}\sum_{i\in G_g^0\backslash\widehat{G}_g}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}$ follows the same techniques as in (iii) of the Theorem 4.4. The asymptotic negligibility of $\frac{1}{2n_{\widehat{g}}T}\sum_{i\in\widehat{G}_g\backslash G_g^0}\sum_{t=1}^{T}\widetilde{u}_{it}^2$, $(\widehat{\rho}_{\widehat{g}n}-\rho_{gn}^0)^2\frac{1}{2n_{\widehat{g}}T}\sum_{i\in\widehat{G}_g\backslash G_g^0}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2$ and $(\widehat{\rho}_{\widehat{g}n}-\rho_{gn}^0)\frac{1}{2n_{\widehat{g}}T}\sum_{i\in\widehat{G}_g\backslash G_g^0}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}$ follows the same techniques as in (i) (ii) of the Theorem 4.4. The asymptotic negligibility of $(\widehat{\rho}_{\widehat{g}n}-\rho_{gn}^0)^2\frac{1}{2n_{\widehat{g}}T}\sum_{i\in G_g^0}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2$ and $(\widehat{\rho}_{\widehat{g}n}-\rho_{gn}^0)\frac{1}{2n_{\widehat{g}}T}\sum_{i\in G_g^0}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}$ follows identically as the proof of the oracle estimator. Therefore, under the joint convergence $(n,T)\to\infty$, we have

$$\frac{1}{2n_{\widehat{g}}T}\sum_{i\in\widehat{G}_g}\sum_{t=1}^{T}\widehat{\widetilde{u}}_{it}^2=\frac{1}{2n_gT}\sum_{i\in G_g^0}\sum_{t=1}^{T}\widetilde{u}_{it}^2+o_p(1).$$

For the stationary group, the consistency of the estimator follows Phillips (2014).■

The Proof of Theorem 5.2: Based on Theorem 4.4 and 5.1, the results are trivial.■

The Proof of Theorem 5.3: Under the joint convergence $(n,T)\to\infty$ and the null hypothesis, the $\widetilde{W}_{\widehat{g}}$ statistics follows chi-square distribution. Under the alternative hypothesis of explosive roots, we have

$$\frac{\left(\widehat{\rho}_{gn}-1\right)\left(\sqrt{\sum_{i=1}^{n}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2}\right)}{\widehat{\sigma}}=\frac{\left(\widehat{\rho}_{gn}-\rho_{gn}^0\right)-\left(1-\rho_{gn}^0\right)}{\widehat{\sigma}}\sqrt{\sum_{i=1}^{n}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2}$$

$$=\frac{\left(\widehat{\rho}_{gn}-\rho_{gn}^0\right)\left(\sqrt{\sum_{i=1}^{n}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2}\right)}{\widehat{\sigma}}-\frac{\left(1-\rho_{gn}^0\right)\left(\sqrt{\sum_{i=1}^{n}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2}\right)}{\widehat{\sigma}}=O_p\left(1\right)+O_p\left(\sqrt{n}\left(\rho_{gn}^0\right)^T\right).$$

Therefore we conclude our proof. ■

D.1.1 Inference procedures: detection of bubble origination

The Proof of Theorem 5.4: As $(n,T)\to\infty$, we have $\beta_{Tn}\to 0$ and $cv_{\beta_{Tn}}\to\infty$. Since $\widetilde{W}_g \xrightarrow{d} \chi^2(1)$ under the null hypothesis of no bubble episode for $g=1,2,...K^0$, we have

$$\lim_{(n,T)\to\infty} \Pr\left(\widetilde{W}_g(\cdot) > cv_{\beta_{Tn}}\right) = \Pr\left(\chi^2(1)=\infty\right)=0,$$

where $cv_{\beta_{Tn}}\to\infty$. Hence, with joint convergence $(n,T)\to\infty$, no origination point for an explosive model in the data will be detected under the null hypothesis.■

The Proof of Theorem 5.5: We fix $g=1,2,...,K^0$**.** Due to the uniform consistency of the modified k-means classifiers, $\widehat{G}_g$ is equivalent to G_g^0, and $\widehat{g}_i$ is equivalent to g_i^0 asymptotically. For all individual in G_g^0, the time series data is sampled from

$$y_{it} = \mu_i(1-\rho_{it}) + \rho_{it} y_{i,t-1} + u_{it},\ i=1,2,...,n; i=1,2,...,n, \tag{56}$$

where $\rho_{it} = 1+\frac{c_{it}}{T^\gamma}$, and $c_{it} = c_{1i}^0 \mathbf{1}\left\{t<\tau_{g_i}^e\right\} + c_{2i}^0\mathbf{1}\left\{t\geq\tau_{g_i}^e\right\}$. For the g-th group, the true distance parameter of explosive episode is α_{2g}^0. The true value of the slope in the g-th group is $\rho_{2,gn}^0$, and we write as $\rho_{2,gn}$ for simplicity. The true value of the explosive slope in the i-th individual is ρ_{2i}^0, and we write as ρ_{2i} for simplicity. For any $i\in G_g^0$, we have $g_i=g$. For $r_g<r_g^e$, we have the convergence in distribution,

$$\widetilde{W}_g(r_g) \xrightarrow{d} \chi^2(1),$$

under the model of (56). Therefore, under the alternative hypothesis, we have for $r_g<r_g^e$, $\Pr\left(\widehat{r}_g^e<r_g^e\right)\to 0$ as $(n,T)\to\infty$.

Next suppose the data is sampled over $t=1,2,...,\tau_{g_i}=[Tr_{g_i}]$, for each $i\in G_g^0$. In this case, the data $\left\{y_{it}: t=\tau_{g_i}^e,...,\tau_{g_i}\right\}$ satisfies

$$y_{it} = \mu_i(1-\rho_{2i}) + \rho_{2i}y_{i,t-1} + u_{it} = \sum_{j=0}^{t-\tau_{g_i}^e} \rho_{2i}^j u_{i,t-j} + \sum_{j=0}^{t-\tau_{g_i}^e} \rho_{2i}^j \mu_i(1-\rho_{2i}) + \rho_{2i}^{t-\tau_{g_i}^e+1} y_{i,\tau_{g_i}^e-1}.$$

As $t-\tau_{g_i}^e\to\infty$, the following asymptotic theory holds,

$$\frac{1}{T^{\frac{\gamma}{2}}}\sum_{j=0}^{t-\tau_{g_i}^e} \rho_{2i}^{j-\left(t-\tau_{g_i}^e\right)} u_{i,t-j} \xrightarrow{d} N\left(0,\frac{\sigma^2}{2c_{2i}^0}\right),$$

where $\frac{1}{\sqrt{T}}y_{i,\tau_{g_i}^e-1} \Rightarrow B_i\left(\tau_{g_i}^e\right)$ by the functional central limit theorem. Then as $t-\tau_{g_i}^e\to\infty$, we have

$$\begin{aligned}\frac{\rho_{2i}^{-\left(t-\tau_{g_i}^e\right)}}{\sqrt{T}} y_{i,t} &= \frac{1}{T^{\frac{1-\gamma}{2}}}\frac{1}{\sqrt{T^\gamma}}\sum_{j=0}^{t-\tau_{g_i}^e} \rho_{2i}^{-\left(t-\tau_{g_i}^e-j\right)} u_{i,t-j} + \frac{\rho_{2i}}{\sqrt{T}} y_{i,\tau_{g_i}^e-1} \\ &\quad + \frac{1}{T^{\frac{1-\gamma}{2}}}\frac{1}{\sqrt{T^\gamma}}\sum_{j=0}^{t-\tau_{g_i}^e} \rho_{2i}^{-\left(t-\tau_{g_i}^e-j\right)} \mu_i(1-\rho_{2i})\end{aligned}$$

$$\Rightarrow \quad B_i\left(r_{g_i}^e\right).$$

So that for each $i \in G_g^0$, we have $y_{it} \sim \sqrt{T}\rho_{2i}^{\left(t-\tau_{g_i}^e\right)} B_i\left(r_{g_i}^e\right)$, for all $t - \tau_{g_i}^e \to \infty$.

Now consider the centered quantities $\widetilde{y}_{it} = y_{it} - \frac{1}{\tau}\sum_{j=1}^{\tau} y_{i,j}$. Note the fact that $g_i = g$. For $\tau_g = [Tr_g]$ (or $\tau_{g_i} = [Tr_{g_i}]$) and $r_g^e < r_g$ (or $r_{g_i}^e < r_{g_i}$), we have

$$\frac{1}{\tau_{g_i}\sqrt{T}}\sum_{j=1}^{\tau_{g_i}} y_{ij} = \frac{1}{\tau_{g_i}\sqrt{T}}\sum_{j=\tau_{g_i}^e}^{\tau_{g_i}} y_{ij} + \frac{\tau_{g_i}^e}{\tau_{g_i}}\frac{1}{\tau_{g_i}^e}\sum_{j=1}^{\tau_{g_i}^e} y_{ij} \sim \frac{1}{\tau_{g_i}\sqrt{T}}\sum_{j=\tau_{g_i}^e}^{\tau_{g_i}} y_{ij} + \frac{r_{g_i}^e}{r_{g_i}}\int_0^1 B_i(s)\,ds,$$

and

$$\begin{aligned}
\frac{1}{\tau_{g_i}}\sum_{j=\tau_{g_i}^e}^{\tau_{g_i}} y_{ij} &= \frac{1}{\tau_{g_i}}\sum_{j=\tau_{g_i}^e}^{\tau_{g_i}} \rho_{2i}^{\left(j-\tau_{g_i}^e\right)}\left(\rho_{2i}^{-\left(j-\tau_{g_i}^e\right)} y_{ij}\right) = \frac{y_{i,\tau_{g_i}^e}}{\tau_{g_i}}\sum_{k=0}^{t-\tau_{g_i}^e} \rho_{2i}^k\left(1+o(1)\right) \\
&= \frac{y_{i,\tau_{g_i}^e}}{\tau_{g_i}}\frac{\rho_{2i}^{t-\tau_{g_i}^e+1}-1}{\tau_{g_i}\left(\rho_{2i}-1\right)}\left(1+o\left(1\right)\right) \\
&= \frac{y_{i,\tau_{g_i}^e}}{\tau_{g_i}}\frac{T^\gamma\rho_{2i}^{t-\tau_{g_i}^e+1}}{\tau_{g_i}c_{2i}^0}\left(1+o\left(1\right)\right).
\end{aligned}$$

It follows that

$$\widetilde{y}_{it} = y_{it} - \frac{1}{\tau_{g_i}}\sum_{j=1}^{\tau_{g_i}} y_{i,j} = \left[\rho_{2i}^{\left(t-\tau_{g_i}^e\right)} - \frac{T^\gamma\rho_{2i}^{\left(t-\tau_{g_i}^e\right)}}{\tau_{g_i}c_{2i}^0}\right] y_{i,\tau_{g_i}^e}\left\{1+o\left(1\right)\right\}.$$

Therefore we have the following asymptotics for the sample moment in the post-classification estimator as,

$$\sum_{i\in G_g^0}\sum_{j=1}^{\tau_{g_i}^e} \widetilde{y}_{it}^2 = \frac{nT^{2\gamma}\tau_g^e\rho_{2,gn}^{2\left(t-\tau_g^e\right)}}{\tau_g^2\left(\alpha_{2g}^0\right)^2}\mathbb{E}y_{i,\tau_g^e}^2\left\{1+o\left(1\right)\right\}.$$

Using these results in conjuction with the standard unit root limit theory, we have

$$\sum_{i\in G_g^0}\sum_{j=1}^{\tau_{g_i}} \widetilde{y}_{i,j-1}^2 = \sum_{i\in G_g^0}\sum_{j=1}^{\tau_{g_i}^e-1} \widetilde{y}_{i,j-1}^2 + \sum_{i\in G_g^0}\sum_{j=\tau_{g_i}^e}^{\tau_{g_i}} \widetilde{y}_{i,j-1}^2 \sim \sum_{i\in G_g^0}\sum_{j=\tau_{g_i}^e}^{\tau_{g_i}} \widetilde{y}_{i,j-1}^2, \tag{57}$$

and

$$\sum_{i\in G_g^0}\sum_{j=1}^{\tau_{g_i}} \widetilde{y}_{i,t-1}\left(y_{i,j}-\rho_{2i}y_{i,j-1}\right) = \sum_{i\in G_g^0}\sum_{j=1}^{\tau_{g_i}^e-1} \widetilde{y}_{i,t-1}\left(u_{i,j} - \frac{c_{2i}^0}{T^{2\gamma}}y_{i,j-1}\right) + \sum_{i\in G_g^0}\sum_{j=1}^{\tau_{g_i}^e-1} \widetilde{y}_{i,t-1}u_{i,j}\left(1+o\left(1\right)\right). \tag{58}$$

Explicit probability limits of (57) and (58) are as,

$$\sum_{i\in G_g^0}\sum_{j=\tau_{g_i}^e}^{\tau_{g_i}} \widetilde{y}_{i,j-1}^2 \quad = \quad \sum_{i\in G_g^0}\sum_{j=\tau_{g_i}^e}^{\tau_{g_i}} \rho_{2i}^{2\left(j-\tau_{g_i}^e\right)} y_{i,\tau_{g_i}^e}^2\left(1+o\left(1\right)\right) = \frac{n\rho_{2,gn}^{2\left(t-\tau_g^e+1\right)}}{\rho_{2,gn}^2-1}\mathbb{E}y_{i,\tau_g^e}^2\left\{1+o\left(1\right)\right\}$$

$$= \frac{nT^{\gamma}\rho_{2,gn}^{2\left(t-\tau_g^e\right)}}{2\alpha_{2g}^0}\mathbb{E}y_{i,\tau_g^e}^2\left\{1+o\left(1\right)\right\},$$

which dominates $\sum_{i\in G_g^0}\sum_{j=1}^{\tau_{g_i}^e}\widetilde{y}_{i,j-1}^2$. The above derivations are due to the uniform consistency of the modified k-means classifiers.

Besides, we have for each i,

$$\begin{aligned}
\sum_{j=\tau_{g_i}^e}^{\tau_{g_i}}\widetilde{y}_{i,j-1}\widetilde{u}_{i,j} &= \sum_{j=\tau_{g_i}^e}^{\tau_{g_i}}\rho_{2i}^{j-1-\tau_{g_i}^e}y_{i,\tau_{g_i}^e}u_{i,j}\left(1+o_p\left(1\right)\right) \\
&= T^{\frac{\gamma}{2}}\rho_{2i}^{\tau_{g_i}-\tau_{g_i}^e}y_{i,\tau_{g_i}^e}\left[\frac{1}{T^{\frac{\gamma}{2}}}\sum_{j=\tau_{g_i}^e}^{\tau_{g_i}}\rho_{2i}^{-\left(\tau_{g_i}-j+1\right)}u_{i,j}\right]\left(1+o_p\left(1\right)\right) \\
&\sim T^{\frac{\gamma+1}{2}}\rho_{2,gn}^{\tau_g-\tau_g^e}\cdot N\left(0,\frac{\sigma^2}{2\alpha_{2g}^0}\right)\cdot N\left(0,\sigma^2\right).
\end{aligned}$$

The terms $\frac{1}{\sqrt{T}}y_{i,\tau_{g_i}^e}$ and $\left[\frac{1}{T^{\frac{\gamma}{2}}}\sum_{j=\tau_{g_i}^e}^{\tau_{g_i}}\rho_{2i}^{-(\tau-j+1)}u_{i,j}\right]$ are uncorrelated due to the martingale difference property of innovations. Therefore the two limiting Guassian processes are independent. Under the joint convergence $(n,T)\to\infty$, we have

$$\sum_{i\in G_g^0}\sum_{j=\tau_{g_i}^e}^{\tau_{g_i}}\widetilde{y}_{i,j-1}\widetilde{u}_{i,j}\sim\sqrt{n}T^{\frac{\gamma+1}{2}}\rho_{2,gn}^{\tau_g-\tau_g^e}\cdot N\left(0,\frac{\sigma^4}{2\alpha_{2g}^0}\right).$$

It follows that for $\tau_g=[Tr_g]$ and $r_g>r_g^e$ we have

$$\frac{\sqrt{n}T^{\frac{1+\gamma}{2}}\rho_{2,gn}^{\tau_g-\tau_g^e}}{2\alpha_{2g}^0}\left(\widehat{\rho}_{2,gn}-\rho_{2,gn}\right)\xrightarrow{d}N\left(0,2\alpha_{2g}^0\right),$$

where under $(n,T)\to\infty$, we have

$$\frac{\rho_{2,gn}^{-\left(\tau_g-\tau_g^e\right)}}{\sqrt{n}T^{\frac{1+\gamma}{2}}}\sum_{i\in G_g^0}\sum_{j=\tau_{g_i}^e}^{\tau_{g_i}}\widetilde{y}_{i,j-1}\widetilde{u}_{i,j}\xrightarrow{d}N\left(0,\frac{\sigma^4}{2\alpha_{2g}^0}\right),$$

and

$$\frac{2\alpha_{2g}^0\rho_{2,gn}^{-2\left(\tau_g-\tau_g^e\right)}}{nT^{1+\gamma}}\sum_{i\in G_g^0}\sum_{j=\tau_{g_i}^e}^{\tau_{g_i}}\widetilde{y}_{i,j-1}^2\xrightarrow{p}\sigma^2.$$

The regression residuals variance estimate is as, for any $r_g\in[r_0,1]$, we have

$$\begin{aligned}
\widetilde{\sigma}_g^2\left(r_g\right) &= \frac{1}{2n_g\tau_g}\sum_{i\in G_g^0}\sum_{j=1}^{\tau_g}\left(\widetilde{y}_{i,j}-\widehat{\rho}_{2i}\left(r_g\right)\widetilde{y}_{i,j-1}\right)^2 \\
&= \frac{1}{2n_g\tau_g}\sum_{i\in G_g^0}\sum_{j=1}^{\tau_g}\left(\widetilde{u}_{i,j}-\left(\widehat{\rho}_{2i}\left(r_g\right)-\rho_{2i}\right)\widetilde{y}_{i,j-1}\mathbf{1}\left\{j\geq\tau_{g_i}^e\right\}-\left(\widehat{\rho}_{2i}\left(r_g\right)-1\right)\widetilde{y}_{i,j-1}\mathbf{1}\left\{j<\tau_{g_i}^e\right\}\right)^2
\end{aligned}$$

$$\sim \frac{\tau_g^e \rho_{2,gn}^{2\left(\tau_g-\tau_g^e\right)}}{2\left(\alpha_{2g}^0\right)^2\tau_g^3}\mathbb{E}y_{i,\tau_g^e}^2\left(1+o_p\left(1\right)\right),$$

due to the fact that $\frac{1}{2n_g\tau_g}\sum_{i\in G_g^0}\sum_{j=1}^{\tau_g}\widetilde{y}_{i,j-1}^2=\frac{\tau_g^e \rho_{2,gn}^{2\left(\tau_g-\tau_g^e\right)}}{2\left(\alpha_{2g}^0\right)^2\tau^3}\mathbb{E}y_{i,\tau_g^e}^2\left(1+o_p\left(1\right)\right).$

Collecting all results, the $\widetilde{W}_g^2\left(r_g\right)$ statistics for any $r_g\in\left[r_g^e,1\right]$, is as

$$\begin{aligned}
\widetilde{W}_g\left(r_g\right) &= \frac{\left(\widehat{\rho}_{2,gn}-1\right)^2\left(\sum_{i\in\widehat{G}_g}\sum_{j=1}^{\tau_g}\widetilde{y}_{i,j-1}^2\right)}{\widetilde{\sigma}_g^2\left(r_g\right)}\\
&= \frac{\left(\widehat{\rho}_{2,gn}-1\right)^2\tau_g^2\left(\frac{1}{\tau_g^2}\sum_{i\in G_g^0}\sum_{j=1}^{\tau_g}\widetilde{y}_{i,j-1}^2\right)}{\widetilde{\sigma}_g^2\left(r_g\right)}\\
&= \frac{T^{2-2\gamma}\left(r_g\cdot\alpha_{2,g}^0\right)^2\left(1+o_p\left(1\right)\right)\left(\frac{1}{\tau_g^2}\sum_{i\in G_g^0}\sum_{j=1}^{\tau_g}\widetilde{y}_{i,j-1}^2\right)}{\widetilde{\sigma}_g^2\left(r_g\right)}\\
&= \frac{T^{2-2\gamma}\left(r_g\cdot\alpha_{2,g}^0\right)^2\left(1+o_p\left(1\right)\right)\left(\frac{nT^{\gamma}\rho_{2,gn}^{2\left(\tau_g-\tau_g^e\right)}}{2\tau_g^2\alpha_{2g}^0}\mathbb{E}y_{i,\tau_g^e}^2\right)}{\frac{r_g^e\rho_{2,gn}^{2\left(\tau_g-\tau_g^e\right)}}{2r_g\left(\alpha_{2g}^0\right)^2\tau_g^2}\mathbb{E}y_{i,\tau_g^e}^2\left(1+o_p\left(1\right)\right)}\\
&= \frac{\left(nT^{\gamma}\alpha_{2,g}^0 r_g\right)T^{2-2\gamma}\left(r_g\cdot\alpha_{2,g}^0\right)^2\left(1+o_p\left(1\right)\right)}{r_g^e}=O\left(nT^{2-\gamma}\right)=O\left(P_{Tn}\right).
\end{aligned}$$

This argument shows that the statistics is consistent under the alternative hypothesis. ■

D.1.2 Model selection

The Proof of Theorem 5.6: Based on Theorem 4.4 and 5.1, we have

$$\begin{aligned}
\mathrm{BIC}(K^0) &= \ln\left(\widetilde{\sigma}_{\widehat{G}_g(K^0)}^2\right)+\widetilde{\sigma}_{\widehat{G}_g(K^0)}^2\frac{K^0+n}{nT}\ln(nT)\\
&= \ln\left[\frac{1}{nT}\sum_{g=1}^{K^0}\sum_{i\in\widehat{G}_g(K^0)}\sum_{t=1}^{T}\left(\widetilde{y}_{it}-\widehat{\rho}_{\widehat{G}_g(K)n}\widetilde{y}_{i,t-1}\right)^2\right]+o(1)\\
&\to \sigma^2.
\end{aligned}$$

Moreover, under the under-fitted model with $K^*<K^0$, note that

$$\begin{aligned}
\widetilde{\sigma}_{\widehat{G}(K^*)}^2 &= \ln\left[\frac{1}{nT}\sum_{g=1}^{K^*}\sum_{i\in\widehat{G}_g(K^*)}\sum_{t=1}^{T}\left(\widetilde{y}_{it}-\widehat{\rho}_{\widehat{G}_g(K^*)}n\widetilde{y}_{i,t-1}\right)^2\right]\\
&\geq \min_{1\leq K^*<K^0}\inf_{G(K^*)}\frac{1}{nT}\sum_{g=1}^{K^*}\sum_{i\in G_g(K^*)}\sum_{t=1}^{T}\left(\widetilde{y}_{it}-\widehat{\rho}_{G_g(K^*)n}\widetilde{y}_{i,t-1}\right)^2
\end{aligned}$$

$$= \min_{1\leq K^*<K^0} \inf_{G(K^*)} \widetilde{\sigma}^2_{G(K^*)}.$$

Under Assumption 1, we have

$$\begin{aligned}
\min_{1\leq K^*<K} \text{BIC}\left(K^*\right) &\geq \min_{1\leq K^*<K} \inf_{G(K^*)\in\mathcal{G}(K^*)} \ln\left(\widetilde{\sigma}^2_{G(K^*)}\right) + \widetilde{\sigma}^2_{G(K^*)} \frac{K^0+n}{nT} \ln\left(nT\right) \\
\xrightarrow{p} \ln\left(\underline{\sigma}^2\right) &> \ln\left(\sigma^2\right),
\end{aligned}$$

and it follows that

$$\Pr\left(\min_{1\leq K^*<K} \text{BIC}\left(K^*\right) > BIC\left(K\right)\right) \to 1.$$

Lastly, under the overfitted moel, and in this case $K^0 < K^* \leq K_{\max}$,

$$\begin{aligned}
& \Pr\left(\min_{K^0<K^*\leq K_{\max}} \text{BIC}\left(K^*\right) > \text{BIC}\left(K\right)\right) \\
= & \Pr\left(\min_{K^0<K^*\leq K_{\max}} nT\ln\left(\widetilde{\sigma}^2_{G(K^*)}/\widetilde{\sigma}^2_{G(K)}\right) + \left(K^*-K^0\right)\ln\left(nT\right) > 0\right) \\
= & \Pr\left(\min_{K^0<K^*\leq K_{\max}} nT\ln\left(\frac{\widetilde{\sigma}^2_{G(K^*)} - \widetilde{\sigma}^2_{G(K)}}{\widetilde{\sigma}^2_{G(K)}}\right) + \left(K^*-K^0\right)\ln\left(nT\right) > 0\right) \\
& \to 1,
\end{aligned}$$

as $(n, T) \to \infty$.■

References

Bai, J., & Ng, S. (2002). Determining the number of factors in approximate factor models. Econometrica, 70(1), 191-221.

Bai, J. (2003). Inferential theory for factor models of large dimensions. Econometrica, 71(1), 135-171.

Bai, J. (2009). Panel data models with interactive fixed effects. Econometrica, 77(4), 1229-1279.

Blanchard, O. J., & Watson, M. W. (1982). Bubbles, rational expectations and financial markets. NBER working paper.

Bonhomme, S., & Manresa, E. (2015). Grouped patterns of heterogeneity in panel data. Econometrica, 83(3), 1147-1184.

Chen, K., & Wen, Y. (2017). The great housing boom of China. American Economic Journal: Macroeconomics, 9(2), 73-114.

Fang, H., Gu, Q., Xiong, W., & Zhou, L. A. (2016). Demystifying the Chinese housing boom. NBER macroeconomics annual, 30(1), 105-166.

Freedman, D. A. (1975). On tail probabilities for martingales. the Annals of Probability, 3(1), 100-118.

Grout, P. A. (2003). Public and private sector discount rates in public–private partnerships. The Economic Journal, 113(486), C62-C68.

Hahn, J., & Kuersteiner, G. (2002). Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and T are large. Econometrica, 70(4), 1639-1657.

Harrison, G. W., Lau, M. I., & Williams, M. B. (2002). Estimating individual discount rates in Denmark: A field experiment. American economic review, 92(5), 1606-1617.

Harvey, D. I., Leybourne, S. J., & Sollis, R. (2013). Recursive right-tailed unit root tests for an explosive asset price bubble. Journal of Financial Econometrics, 13(1), 166-187.

Harvey, D. I., Leybourne, S. J., Sollis, R., & Taylor, A. R. (2016). Tests for explosive financial bubbles in the presence of non-stationary volatility. Journal of Empirical Finance, 38, 548-574.

Homm, U., & Breitung, J. (2012). Testing for speculative bubbles in stock markets: a comparison of alternative methods. Journal of Financial Econometrics, 10(1), 198-231.

Huang, W., S. Jin, and L. Su. (2019). Panel Cointegration with Latent Group Structures and an Application to the PPP Theory. Econometric Theory, forthcoming.

Ke, Z. T., Fan, J., & Wu, Y. (2015). Homogeneity pursuit. Journal of the American Statistical Association, 110(509), 175-194.

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

Narayan, P. K., Mishra, S., Sharma, S., & Liu, R. (2013). Determinants of stock price bubbles. Economic Modelling, 35, 661-667.

Pavlidis, E., Yusupova, A., Paya, I., Peel, D., Martínez-García, E., Mack, A., & Grossman, V. (2016). Episodes of exuberance in housing markets: in search of the smoking gun. The Journal of Real Estate Finance and Economics, 53(4), 419-449.

Phllips. P. C. (2013). Estimation of the Localizing Rate for Mildly Integrated and Mildly Explosive Processes. Technical Report: Working Paper, Yale University.

Phillips, P. C., & Magdalinos, T. (2007). Limit theory for moderate deviations from a unit root. Journal of Econometrics, 136(1), 115-130.

Phillips, P. C., & Moon, H. R. (1999). Linear regression limit theory for nonstationary panel data. Econometrica, 67(5), 1057-1111.

Phillips, P. C. (2014). Dynamic panel GMM with roots near unity. Technical Report: Working Paper, Yale University.

Phillips, P. C., Shi, S., & Yu, J. (2015a). Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500. International economic review, 56(4), 1043-1078.

Phillips, P. C., Shi, S., & Yu, J. (2015b). Testing for multiple bubbles: Limit theory of real-time detectors. International Economic Review, 56(4), 1079-1134.

Phillips, P. C., Wu, Y., & Yu, J. (2011). Explosive behavior in the 1990s Nasdaq: When did exuberance escalate asset values?. International economic review, 52(1), 201-226.

Santos, M. S., & Woodford, M. (1997). Rational asset pricing bubbles. Econometrica: Journal of the Econometric Society, 19–57.

Sarafidis, V., & Weber, N. (2015). A partially heterogeneous framework for analyzing panel data. Oxford Bulletin of Economics and Statistics, 77(2), 274-296.

Shi, S. (2017). Speculative bubbles or market fundamentals? An investigation of US regional housing markets. Economic Modelling, 66, 101-111.

Su, L., Shi, Z., & Phillips, P. C. (2016). Identifying latent structures in panel data. Econometrica, 84(6), 2215-2264.

Su, L., Wang, W., & Zhang, Y. (2019). Strong consistency of spectral clustering for stochastic block models. IEEE Transactions on Information Theory, forthcoming.