Online Supplement to 'A Panel Clustering Approach to Analyzing Bubble Behavior'

Yanbo Liu
School of Economics, Shandong University

Peter C. B. Phillips
Singapore Management University, Yale University, University of Auckland, University of Southampton

Jun Yu
Singapore Management University

March 20, 2022

Abstract

This online supplement has five sections. Section A collects together technical lemmas that are used for membership estimation in the first stage. Section B collects the lemmas needed for post-clustering panel estimation and the bubble detection methods, specifically the post-clustering panel t- and J-tests. Section C collects results and proofs for selecting the number of groups. Section D overviews experimental designs and reports simulation findings. Section E contains tables.

We provide technical proofs for clustering, estimation and tests of the proposed two-stage grouping procedure. Throughout the supplement notations are the same as in the main paper. The technical lemmas given in the following sections play central roles in the proofs of the main theorems in the paper.

A Proofs for Stage 1: Recursive k-means Clustering

For any individual $i \in \mathcal{I}_n$, let $\widehat{g_i} := \widehat{g_i}(\widehat{c^*})$ denote the membership estimator of g_i^0 generated by the recursive k-means clustering algorithm. Note that $\widehat{c^*}\left(:= \left(\widehat{c_1^*}, \widehat{c_2^*}, ..., \widehat{c_{G^0}^*}\right)\right)$ is the first-stage estimator of the distancing parameter vector c.

To establish uniform consistency of the recursive k-means clustering algorithm, we first show consistency of $\widehat{c^*}$ in terms of the Hausdorff distance, which measures how far two compact subsets in a metric space are separated from each other, defined by

$$d_H(a,b) = \max\left\{\max_{j\in\{1,2,...,G^0\}}\left(\min_{\widetilde{j}\in\{1,2,...,G^0\}}\left(a_{\widetilde{j}} - b_j\right)^2\right), \max_{\widetilde{j}\in\{1,2,...,G^0\}}\left(\min_{j\in\{1,2,...,G^0\}}\left(a_{\widetilde{j}} - b_j\right)^2\right)\right\},$$

where $a := (a_1, a_2, ..., a_{G^0})$ and $b := (b_1, b_2, ..., b_{G^0})$.

Lemma A.1 *If Assumptions 1 and 2 hold, then,*

$$\sup_{(c,\delta)\in\mathcal{C}_{G^0}\times\Delta_{G^0}} T^{4\gamma}(\log T)^8\left|\widehat{Q}_{nT}(c,\delta)-\widetilde{Q}_{nT}(c,\delta)\right| = o_p(1),$$

where

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

$$\widetilde{Q}_{nT}(c,\delta) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\Upsilon_{iT}}\sum_{t=1}^{T}\left(\widetilde{y}_{i,t-1}\left(\overline{\rho}_i^0-\overline{\rho}_i\right)\right)^2+\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\Upsilon_{iT}}\sum_{t=1}^{T}\widetilde{u}_{it}^2,$$

with $\Upsilon_{iT}=\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2$, *and* $\overline{\rho}_i$ *and* $\overline{\rho}_i^0$ *defined as in the main paper,*

Proof of Lemma A.1: The difference between the two objective functions $\widehat{Q}_{nT}(\cdot,\cdot)$ and $\widetilde{Q}_{nT}(\cdot,\cdot)$ can be measured as

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

Based on the assumption of latent membership, we have

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\Upsilon_{iT}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\overline{c}_i^0=\sum_{\widetilde{j}=1}^{G^0}\mathbf{1}\left\{g_i^0=\widetilde{j}\right\}\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\Upsilon_{iT}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}c_{\widetilde{j}}^0.$$

Since $\left|c_{\widetilde{j}}^0\right|\leq c_{up}$, for any $\widetilde{j}\in\left\{1,2,\ldots,G^0\right\}$,

$$\left|\mathbf{1}\left\{g_i^0=\widetilde{j}\right\}\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\Upsilon_{iT}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}c_{\widetilde{j}}^0\right|\leq c_{up}\left|\mathbf{1}\left\{g_i^0=\widetilde{j}\right\}\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\Upsilon_{iT}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right|.$$

Therefore, by Phillips and Magdalinos (2007) and Phillips and Durlauf (1986), we have

$$\frac{2}{n}\left|\mathbf{1}\left\{g_i^0=\widetilde{j}\right\}\sum_{i=1}^{n}\frac{1}{\Upsilon_{iT}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{i,t}c_{\widetilde{j}}^0\right|\leq\begin{cases}O_p\left(\frac{1}{\left(\rho_{\widetilde{j}}^0\right)^T T^{\gamma}\sqrt{n}}\right) & \text{if the smallest } \widetilde{j} \text{ has } c_{\widetilde{j}}^0>0\\ O_p\left(\frac{1}{T}\right) & \text{if the smallest } \widetilde{j} \text{ has } c_{\widetilde{j}}^0=0\\ O_p\left(\frac{1}{\sqrt{n}T^{\frac{1+\gamma}{2}}}\right) & \text{if the smallest } \widetilde{j} \text{ has } c_{\widetilde{j}}^0<0\end{cases},\tag{1}$$

where $\rho_{\widetilde{j}}^0 = 1 + c_{\widetilde{j}}^0 / T^\gamma$. By applying a similar argument to $\frac{2}{n}\sum_{i=1}^n \frac{1}{\Upsilon_{iT}}\sum_{t=1}^T \widetilde{y}_{i,t-1}\widetilde{u}_{i,t}\overline{c}_i$, we can show that

$$\frac{2}{n}\left|\mathbf{1}\{g_i = j\}\sum_{i=1}^n \frac{1}{\Upsilon_{iT}}\sum_{t=1}^T \widetilde{y}_{i,t-1}\widetilde{u}_{i,t}\overline{c}_i\right| \\ \leq \begin{cases} O_p\left(\frac{1}{\sqrt{n}\left(\rho_{\widetilde{j}}^0\right)^T T^\gamma}\right) & \text{if} \quad \text{the smallest } \widetilde{j} \text{ has } c_{\widetilde{j}}^0 > 0 \\ O_p\left(\frac{1}{T}\right) & \text{if} \quad \text{the smallest } \widetilde{j} \text{ has } c_{\widetilde{j}}^0 = 0 \\ O_p\left(\frac{1}{\sqrt{n}T^{\frac{1+\gamma}{2}}}\right) & \text{if} \quad \text{the smallest } \widetilde{j} \text{ has } c_{\widetilde{j}}^0 < 0 \end{cases}, \tag{2}$$

for any $j = 1, 2, ..., G^0$. Finally, based on equations (1) and (2),

$$\sup_{(c,\delta)\in\mathcal{C}_{G^0}\times\Delta_{G^0}} T^\gamma \left|\widehat{Q}_{nT}(c,\delta) - \widetilde{Q}_{nT}(c,\delta)\right| = \begin{cases} O_p\left(\frac{1}{\left(\rho_{\widetilde{j}}^0\right)^T T^\gamma \sqrt{n}}\right) & \text{if} \quad \text{the smallest } \widetilde{j} \text{ has } c_{\widetilde{j}}^0 > 0 \\ O_p\left(\frac{1}{T}\right) & \text{if} \quad \text{the smallest } \widetilde{j} \text{ has } c_{\widetilde{j}}^0 = 0 \\ O_p\left(\frac{1}{\sqrt{n}T^{\frac{1+\gamma}{2}}}\right) & \text{if} \quad \text{the smallest } \widetilde{j} \text{ has } c_{\widetilde{j}}^0 < 0 \end{cases} \\ = o_p\left(\frac{1}{T^{3\gamma}(\log T)^8}\right),$$

by the rate restriction in Assumption 2. Therefore,

$$\sup_{(c,\delta)\in\mathcal{C}_{G^0}\times\Delta_{G^0}} T^{4\gamma}(\log T)^8 \left|\widehat{Q}_{nT}(c,\delta) - \widetilde{Q}_{nT}(c,\delta)\right| = o_p(1).$$

This concludes the proof. ■

Lemma A.2 *Suppose Assumptions 1 and 2 hold. Then, when* $(n,T)\to\infty$,

$$d_H\left(c^0, \widehat{c}^*\right) = o_p\left(T^{-2\gamma}(\log T)^{-8}\right). \tag{3}$$

Moreover, there exists a permutation $\tau : \left\{1, 2, ...G^0\right\} \to \left\{1, 2, ..., G^0\right\}$, *such that*

$$T^\gamma (\log T)^4 \left|\widehat{c}^*_{\tau(\widehat{j})} - c_j^0\right| \to_p 0.$$

If we relabel $\widehat{c}^*$ *by setting* $\tau\left(\widehat{j}\right) = j$, *then*

$$\left\|\widehat{c}^* - c^0\right\| = o_p\left(T^{-\gamma}(\log T)^{-4}\right). \tag{4}$$

Proof of Lemma A.2: By Lemma A.1,

$$\widetilde{Q}_{nT}\left(\widehat{c}^*, \widehat{\delta}\right) = \widehat{Q}_{nT}\left(\widehat{c}^*, \widehat{\delta}\right) + o_p\left(T^{-4\gamma}(\log T)^{-8}\right)$$

$$
\begin{aligned}
&\leq \widehat{Q}_{nT}\left(c^0,\delta^0\right)+o_p\left(T^{-4\gamma}(\log T)^{-8}\right)\\
&=\widetilde{Q}_{nT}\left(c^0,\delta^0\right)+o_p\left(T^{-4\gamma}(\log T)^{-8}\right).
\end{aligned}
$$

Because $\widetilde{Q}_{nT}(c,\delta)$ is minimized at $c=c^0$ and $\delta=\delta^0$, we have

$$
\widetilde{Q}_{nT}(\widehat{c}^*,\widehat{\delta})-\widetilde{Q}_{nT}\left(c^0,\delta^0\right)=o_p\left(T^{-4\gamma}(\log T)^{-8}\right).
$$

Then, for any c,

$$
\begin{aligned}
&\left|\widetilde{Q}_{nT}(c,\delta)-\widetilde{Q}_{nT}(c^0,\delta^0)\right|\\
&=\left|\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\Upsilon_{iT}}\sum_{t=1}^{T}\left(\widetilde{y}_{i,t-1}^2\left(\overline{\rho}_i^0-\overline{\rho}_i\right)^2\right)\right|\\
&=\sum_{j=1}^{G^0}\sum_{\widetilde{j}=1}^{G^0}\left|\left(\rho_j^0-\rho_{\widetilde{j}}\right)^2\right|\frac{1}{n}\sum_{i=1}^{n}\left[\left(\frac{1}{\Upsilon_{iT}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2\right)\mathbf{1}\left\{g_i^0=j\right\}\mathbf{1}\left\{g_i=\widetilde{j}\right\}\right]\\
&\geq\sum_{j=1}^{G^0}\sum_{\widetilde{j}=1}^{G^0}\left[\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\Upsilon_{iT}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2\right)\mathbf{1}\left\{g_i^0=j\right\}\mathbf{1}\left\{g_i=\widetilde{j}\right\}\right]\min_{1\leq\widetilde{j}\leq G^0}\left|\left(\rho_j^0-\rho_{\widetilde{j}}\right)^2\right|\\
&\geq\sum_{j=1}^{G^0}\max_{1\leq\widetilde{j}\leq G^0}\left[\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\Upsilon_{iT}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2\right)\mathbf{1}\left\{g_i^0=j\right\}\mathbf{1}\left\{g_i=\widetilde{j}\right\}\right]\min_{1\leq\widetilde{j}\leq G^0}\left|\left(\rho_j^0-\rho_{\widetilde{j}}\right)^2\right|\\
&\geq\min_{1\leq j\leq G^0}\max_{1\leq\widetilde{j}\leq G^0}\left[\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\Upsilon_{iT}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2\right)\mathbf{1}\left\{g_i^0=j\right\}\mathbf{1}\left\{g_i=\widetilde{j}\right\}\right]\sum_{j=1}^{G^0}\min_{1\leq\widetilde{j}\leq G^0}\left|\left(\rho_j^0-\rho_{\widetilde{j}}\right)^2\right|\\
&\geq\min_{1\leq j\leq G^0}\max_{1\leq\widetilde{j}\leq G^0}\left[\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{\Upsilon_{iT}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2\right)\mathbf{1}\left\{g_i^0=j\right\}\mathbf{1}\left\{g_i=\widetilde{j}\right\}\right]\max_{1\leq j\leq G^0}\min_{1\leq\widetilde{j}\leq G^0}\left|\left(\rho_j^0-\rho_{\widetilde{j}}\right)^2\right|\\
&\geq\max_{1\leq j\leq G^0}\min_{1\leq\widetilde{j}\leq G^0}\left|\left(\rho_j^0-\rho_{\widetilde{j}}\right)^2\right|.
\end{aligned}
$$

As a result,

$$
\max_{1\leq j\leq G^0}\left(\min_{1\leq\widetilde{j}\leq G^0}\left(\rho_j^0-\widehat{\rho}_{\widetilde{j}}^*\right)^2\right)=o_p\left(T^{-4\gamma}(\log T)^{-8}\right),\tag{5}
$$

or

$$
\max_{1\leq j\leq G^0}\left(\min_{1\leq\widetilde{j}\leq G^0}\left(c_j^0-\widehat{c}_{\widetilde{j}}^*\right)^2\right)=o_p\left(T^{-2\gamma}(\log T)^{-8}\right).\tag{6}
$$

Let $\tau(j)=\arg\min_{\widetilde{j}\in\{1,2,\ldots,G^0\}}\left(c_j^0-\widehat{c}_{\widetilde{j}}^*\right)^2$ and $\widehat{\rho}_j^*=1+\frac{\widehat{c}_j^*}{T^\gamma}$. For any $\widetilde{j}\neq j$,

$$
T^{2\gamma}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\Upsilon_{iT}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2\left(\widehat{\rho}_{\tau(j)}^*-\widehat{\rho}_{\tau(\widetilde{j})}^*\right)^2\right)
$$

$$\geq T^{2\gamma}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\Upsilon_{iT}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^{2}\left(\rho_{j}^{0}-\rho_{\widetilde{j}}^{0}\right)^{2}\right)-T^{2\gamma}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\Upsilon_{iT}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^{2}\left(\widehat{\rho}_{\tau(j)}^{*}-\rho_{j}^{0}\right)^{2}\right) \tag{7}$$

$$-T^{2\gamma}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\Upsilon_{iT}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^{2}\left(\widehat{\rho}_{\tau(\widetilde{j})}^{*}-\rho_{\widetilde{j}}^{0}\right)^{2}\right). \tag{8}$$

The first term of (7) is bounded away from zero since

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\Upsilon_{iT}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^{2}\to_{p}1,$$

under the joint asymptotic scheme $(n,T)\to\infty$ and $\left|c_{j}^{0}-c_{\widetilde{j}}^{0}\right|\geq\dot{c}>0$ for any $\widetilde{j}\neq j$. Therefore, the first term of (7) is nonzero. Due to (5) and (6), the second term of (7) and the term (8) are both $o_{p}(1)$. Therefore, we have $\tau(j)\neq\tau(\widetilde{j})$ with probability approaching one. We note that asymptotically τ is not only an 'onto' mapping but also 'one-to-one' mapping. Hence, τ has an inverse denoted as τ^{-1} and

$$\min_{1\leq j\leq G^{0}}\left(\rho_{j}^{0}-\widehat{\rho}_{\widetilde{j}}^{*}\right)^{2}\leq\left(\rho_{\tau^{-1}(\widetilde{j})}^{0}-\widehat{\rho}_{\widetilde{j}}^{*}\right)^{2}=\min_{1\leq h\leq G^{0}}\left(\rho_{\tau^{-1}(\widetilde{j})}^{0}-\widehat{\rho}_{h}^{*}\right)^{2}=o_{p}\left(T^{-4\gamma}(\log T)^{-8}\right),$$

where the last equality is due to (5) and (6). Then

$$\max_{1\leq\widetilde{j}\leq G^{0}}\left(\min_{1\leq j\leq G^{0}}\left(\rho_{j}^{0}-\widehat{\rho}_{\widetilde{j}}^{*}\right)^{2}\right)=o_{p}\left(T^{-4\gamma}(\log T)^{-8}\right). \tag{9}$$

Combining the results (5) and (9), we show equation (3). Then the proof is completed. ■

In the rest of the Online Supplement, we always relabel $\widehat{c}^{*}$ by setting $\tau\left(\widehat{j}\right)=j$. For any $\eta>0$, we define $\mathcal{N}_{\eta}$, $\widehat{g}_{i}(\widehat{c}^{*})$, and $\widehat{\delta}$ by

$$\mathcal{N}_{\eta}:=\left\{c\in\mathcal{C}_{G^{0}}:\left|c_{j}^{0}-c_{j}\right|<\eta,\ \forall j=1,2,\ldots,G^{0}\right\}, \tag{10}$$

$$\widehat{g}_{i}(\widehat{c}^{*}):=\arg\min_{j\in\{1,2,\ldots,G^{0}\}}\sum_{t=1}^{T}\left(\widetilde{y}_{it}-\widetilde{y}_{i,t-1}\exp\left(\frac{\widehat{c}_{j}^{*}}{T^{\gamma}}\right)\right)^{2},$$

$$\widehat{\delta}:=\left(\widehat{g}_{1}(\widehat{c}^{*}),\widehat{g}_{2}(\widehat{c}^{*}),\ldots,\widehat{g}_{n}(\widehat{c}^{*})\right),$$

where we treat the scaling parameter γ as given *a priori*.

Lemma A.3 *Suppose Assumptions 1 and 2 hold. Then, for any fixed $M>0$,*

(i) if $\overline{c}_{i}^{0}>0$,

$$\max_{i\in\mathcal{I}_{n}}\Pr\left(\frac{(\log T)^{2}}{\left(\overline{\rho}_{i}^{0}\right)^{2T}T^{\gamma}}\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right|\geq M\right)=o\left(\frac{1}{n}\right);$$

(ii) if $\overline{c}_i^0 = 0$,

$$\max_{i\in\mathcal{I}_n} \Pr\left(\frac{(\log_2(T))^2}{T^{2-\gamma}} \left| \sum_{t=1}^{T} \widetilde{y}_{i,t-1}\widetilde{u}_{it} \right| \geq M \right) = o\left(\frac{1}{n}\right);$$

(iii) if $\overline{c}_i^0 < 0$,

$$\max_{i\in\mathcal{I}_n} \Pr\left(\frac{1}{T} \left| \sum_{t=1}^{T} \widetilde{y}_{i,t-1}\widetilde{u}_{it} \right| \geq M \right) = o\left(\frac{1}{n}\right).$$

Proof of Lemma A.3: We use the standard decomposition

$$\sum_{t=1}^{T} \widetilde{y}_{i,t-1}\widetilde{u}_{it} = \sum_{t=1}^{T} y_{i,t-1}u_{it} - T\overline{y}_{i,-1}\overline{u}_i. \tag{11}$$

(i) If $\overline{c}_i^0 > 0$, fix an arbitrary $M > 0$ and we have

$$\begin{aligned}
&\max_{i\in\mathcal{I}_n} \Pr\left(\frac{(\log T)^2}{\left(\overline{\rho}_i^0\right)^{2T} T^{\gamma}} \left| \sum_{t=1}^{T} \widetilde{y}_{i,t-1}\widetilde{u}_{it} \right| \geq M \right) \\
&\leq \max_{i\in\mathcal{I}_n} \Pr\left(\frac{(\log T)^2}{\left(\overline{\rho}_i^0\right)^{2T} T^{\gamma}} \left| \sum_{t=1}^{T} y_{i,t-1}u_{it} \right| \geq \frac{M}{2} \right) + \max_{i\in\mathcal{I}_n} \Pr\left(\frac{(\log T)^2}{\left(\overline{\rho}_i^0\right)^{2T} T^{\gamma-1}} \left| \overline{y}_{i,-1}\overline{u}_i \right| \geq \frac{M}{2} \right).
\end{aligned} \tag{12}$$

For the first term of (12), by the Markov inequality, we have

$$\begin{aligned}
&n \max_{i\in\mathcal{I}_n} \Pr\left(\frac{(\log T)^2}{\left(\overline{\rho}_i^0\right)^{2T} T^{\gamma}} \left| \sum_{t=1}^{T} y_{i,t-1}u_{it} \right| \geq \frac{M}{2} \right) \\
&\leq n \max_{i\in\mathcal{I}_n} \frac{2(\log T)^2 \mathbb{E}\left|\sum_{t=1}^{T} y_{i,t-1}u_{it}\right|}{\left(\overline{\rho}_i^0\right)^{2T} T^{\gamma} M} \\
&\leq n \max_{i\in\mathcal{I}_n} \frac{2(\log T)^2 \left[\mathbb{E}\left(\sum_{t=1}^{T} y_{i,t-1}^2\right)\right]^{\frac{1}{2}} \left[\mathbb{E}\left(\sum_{t=1}^{T} u_{it}^2\right)\right]^{\frac{1}{2}}}{\left(\overline{\rho}_i^0\right)^{2T} T^{\gamma} M} \\
&\leq n \max_{i\in\mathcal{I}_n} K \frac{(\log T)^2 \left(\overline{\rho}_i^0\right)^{T} T^{\frac{1+2\gamma}{2}}}{\left(\overline{\rho}_i^0\right)^{2T} T^{\gamma}} = O\left(\frac{(\log T)^2 n T^{\frac{1}{2}}}{(\rho_{low})^T} \right) = o(1),
\end{aligned} \tag{13}$$

where K denotes a positive constant (a notation that recurs in later derivations) and the asymptotic negligibility in (13) is due to the dominance of the exponential rate. For the second term in (11), by the Markov inequality,

$$n \max_{i\in\mathcal{I}_n} \Pr\left(\frac{(\log T)^2}{\left(\overline{\rho}_i^0\right)^{2T} T^{\gamma-1}} \left| \overline{y}_{i,-1}\overline{u}_i \right| \geq \frac{M}{2} \right) \leq n \max_{i\in\mathcal{I}_n} \frac{2(\log T)^2 \mathbb{E}\left|\overline{y}_{i,-1}\overline{u}_i\right|}{\left(\overline{\rho}_i^0\right)^{2T} T^{\gamma-1} M}$$

$$\leq n \max_{i \in \mathcal{I}_n} \frac{2(\log T)^2 \mathbb{E}\left|\left(\sum_{t=1}^{T} y_{i,t-1}\right)\left(\sum_{t=1}^{T} u_{it}\right)\right|}{\left(\overline{\rho}_i^0\right)^{2T} T^{1+\gamma} M}$$

$$\leq n \max_{i \in \mathcal{I}_n} \frac{2(\log T)^2 \left(\mathbb{E}\left(\sum_{t=1}^{T} y_{i,t-1}\right)^2\right)^{\frac{1}{2}} \left(\mathbb{E}\left(\sum_{t=1}^{T} u_{it}\right)^2\right)^{\frac{1}{2}}}{\left(\overline{\rho}_i^0\right)^{2T} T^{1+\gamma} M}$$

$$\leq n \max_{i \in \mathcal{I}_n} K \frac{(\log T)^2 \left(\overline{\rho}_i^0\right)^T T^{\frac{3\gamma+1}{2}}}{\left(\overline{\rho}_i^0\right)^{2T} T^{1+\gamma}} = O\left(\frac{(\log T)^2 n T^{\frac{-1+\gamma}{2}}}{(\rho_{low})^T}\right) = o(1), \tag{14}$$

with the dominance of the exponential rate. Based on (13) and (14), when $\overline{c}_i^0 > 0$,

$$\max_{i \in \mathcal{I}_n} \Pr\left(\frac{(\log T)^2}{\left(\overline{\rho}_i^0\right)^{2T} T^{\gamma}} \left|\sum_{t=1}^{T} \widetilde{y}_{i,t-1} \widetilde{u}_{it}\right| \geq M\right) = o\left(\frac{1}{n}\right).$$

(ii) If $\overline{c}_i^0 = 0$, for any $M > 0$, the following decomposition applies

$$n \max_{i \in \mathcal{I}_n} \Pr\left(\frac{(\log_2 T)^2}{T^{2-\gamma}} \left|\sum_{t=1}^{T} \widetilde{y}_{i,t-1} \widetilde{u}_{it}\right| \geq M\right) \leq n \max_{i \in \mathcal{I}_n} \Pr\left(\frac{(\log_2 T)^2}{T^{2-\gamma}} \left|\sum_{t=1}^{T} y_{i,t-1} u_{it}\right| \geq \frac{M}{2}\right)$$
$$+ n \max_{i \in \mathcal{I}_n} \Pr\left(\frac{(\log_2 T)^2}{T^{3-\gamma}} \left|\left(\sum_{t=1}^{T} y_{i,t-1}\right)\left(\sum_{t=1}^{T} u_{it}\right)\right| \geq \frac{M}{2}\right).$$

(ii.1) Since $y_{i0} = 0$ with $\overline{c}_i^0 = 0$, we have

$$\frac{(\log_2 T)^2}{T^{2-\gamma}} \sum_{t=1}^{T} y_{i,t-1} u_{it}$$
$$= \frac{(\log_2 T)^2}{T^{2-\gamma}} \sum_{t=1}^{T} \left(F_i(1) \sum_{s=1}^{t-1} \epsilon_{is} + \widetilde{\epsilon}_{i,0} - \widetilde{\epsilon}_{i,t-1}\right) \left(F_i(1) \epsilon_{it} + \widetilde{\epsilon}_{i,t-1} - \widetilde{\epsilon}_{i,t}\right)$$
$$= \underbrace{\frac{(\log_2 T)^2 F_i^2(1)}{T^{2-\gamma}} \sum_{t=1}^{T} \left(\sum_{s=1}^{t-1} \epsilon_{is}\right) \epsilon_{it}}_{\text{(A.1)}} + \underbrace{\frac{(\log_2 T)^2 F_i(1)}{T^{2-\gamma}} \sum_{t=1}^{T} (\widetilde{\epsilon}_{i,0} - \widetilde{\epsilon}_{i,t-1}) \epsilon_{it}}_{\text{(A.2)}}$$
$$+ \underbrace{\frac{(\log_2 T)^2 F_i(1)}{T^{2-\gamma}} \sum_{t=1}^{T} \left(\sum_{s=1}^{t-1} \epsilon_{is}\right) (\widetilde{\epsilon}_{i,t-1} - \widetilde{\epsilon}_{i,t})}_{\text{(A.3)}} + \underbrace{\frac{(\log_2 T)^2}{T^{2-\gamma}} \sum_{t=1}^{T} (\widetilde{\epsilon}_{i,0} - \widetilde{\epsilon}_{i,t-1})(\widetilde{\epsilon}_{i,t-1} - \widetilde{\epsilon}_{i,t})}_{\text{(A.4)}}.$$

For terms (A.1) and (A.2), we have

$$\Pr\left(\left|\frac{(\log_2 T)^2 F_i^2(1)}{T^{2-\gamma}} \sum_{t=1}^{T} \left(\sum_{s=1}^{t-1} \epsilon_{is}\right) \epsilon_{it} + \frac{(\log_2 T)^2 F_i(1)}{T^{2-\gamma}} \sum_{t=1}^{T} (\widetilde{\epsilon}_{i,0} - \widetilde{\epsilon}_{i,t-1}) \epsilon_{it}\right| \geq \frac{M}{4}\right)$$

$$= \Pr\left(\left|\frac{(\log_2 T)^2 F_i(1)}{T^{2-\gamma}}\sum_{t=1}^{T}\left(\sum_{s=1}^{t-1} u_{is}\right)\epsilon_{it}\right| \geq \frac{M}{4}\right).$$

Let $z_{it} = \left(\sum_{s=1}^{t-1} u_{is}\right)\epsilon_{it}$ and $\mathcal{F}_{i,t-1} = \sigma\{\varepsilon_{i,t-1}, \varepsilon_{i,t-2}, \ldots\}$ be the sigma-field generated by $\{\varepsilon_{i,t-1}, \varepsilon_{i,t-2}, \ldots\}$. Although $\{z_{it}\}$ is a martingale sequence as $\mathbb{E}(z_{it}|\mathcal{F}_{i,t-1}) = 0$, the exponential inequality (Freedman, 1975) is not directly applicable to $\{z_{it}\}$. Applying the following decomposition to z_{it}, we get

$$\begin{aligned}
&\frac{(\log_2 T)^2 F_i(1)}{T^{2-\gamma}}\sum_{t=1}^{T} z_{it} \\
&= \frac{(\log_2 T)^2 F_i(1)}{T^{2-\gamma}}\sum_{t=1}^{T} z_{1it} + \frac{(\log_2 T)^2 F_i(1)}{T^{2-\gamma}}\sum_{t=1}^{T} z_{2it} - \frac{(\log_2 T)^2 F_i(1)}{T^{2-\gamma}}\sum_{t=1}^{T}\mathbb{E}(z_{2it}|\mathcal{F}_{i,t-1}),
\end{aligned}$$

where $z_{1it} = z_{it}\mathbf{1}_{it} - \mathbb{E}[z_{it}\mathbf{1}_{it}|\mathcal{F}_{i,t-1}]$, $z_{2it} = z_{it}\overline{\mathbf{1}}_{it}$, $\mathbf{1}_{it} = \mathbf{1}\{|z_{it}| \leq d_{nT}\}$, $d_{nT} = n^{\frac{1}{4}}T^{\frac{3}{4}}$, and $\overline{\mathbf{1}}_{it} = 1 - \mathbf{1}_{it}$. Hence,

$$\begin{aligned}
&n\max_{1\leq i\leq n}\Pr\left(\frac{(\log_2 T)^2}{T^{2-\gamma}}\left|F_i(1)\sum_{t=1}^{T} z_{it}\right| \geq \frac{M}{4}\right) \\
&\leq n\max_{1\leq i\leq n}\Pr\left(\frac{(\log_2 T)^2}{T^{2-\gamma}}\left|F_i(1)\sum_{t=1}^{T} z_{1it}\right| \geq \frac{M}{12}\right) + n\max_{1\leq i\leq n}\Pr\left(\frac{(\log_2 T)^2}{T^{2-\gamma}}\left|F_i(1)\sum_{t=1}^{T} z_{2it}\right| \geq \frac{M}{12}\right) \\
&+ n\max_{1\leq i\leq n}\Pr\left(\frac{(\log_2 T)^2}{T^{2-\gamma}}\left|F_i(1)\sum_{t=1}^{T}\mathbb{E}(z_{2it}|\mathcal{F}_{i,t-1})\right| \geq \frac{M}{12}\right).
\end{aligned}$$

Let $V_{iT} = \sum_{t=1}^{T}\mathbb{E}\left(z_{it}^2|\mathcal{F}_{i,t-1}\right)$, $v_{nT} = \sqrt{n}T^2$. By Hölder's inequality and Jensen's inequality, we have

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

As $\mathbb{E}\{z_{1it}|\mathcal{F}_{i,t-1}\} = \mathbb{E}\{z_{it}\mathbf{1}_{it} - \mathbb{E}[z_{it}\mathbf{1}_{it}|\mathcal{F}_{i,t-1}]|\mathcal{F}_{i,t-1}\} = \mathbb{E}[z_{it}\mathbf{1}_{it}|\mathcal{F}_{i,t-1}] - \mathbb{E}[z_{it}\mathbf{1}_{it}|\mathcal{F}_{i,t-1}] = 0$, the exponential inequality of Freedman (1975) can be applied and leads to

$$\begin{aligned}
&n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{(\log_2 T)^2}{T^{2-\gamma}}\left|F_i(1)\sum_{t=1}^{T} z_{1it}\right| \geq \frac{M}{12}\right) \\
&\leq n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{(\log_2 T)^2}{T^{2-\gamma}}\left|F_i(1)\sum_{t=1}^{T} z_{1it}\right| \geq \frac{M}{12}, V_{iT} \leq v_{nT}\right) + n\max_{i\in\mathcal{I}_n}\Pr(V_{iT} > v_{nT}) \\
&\leq \exp\left(\frac{-M^2T^{4-2\gamma}/\left(144F_i^2(1)(\log_2 T)^4\right) + 2v_{nT}\log(n) + 4T^{2-\gamma}Md_{nT}\log(n)/\left(12F_i(1)(\log_2 T)^2\right)}{2v_{nT} + 4T^{2-\gamma}Md_{nT}/\left(12F_i(1)(\log_2 T)^2\right)}\right)
\end{aligned}$$

$$
\begin{aligned}
&+o\left(nT^4 v_{nT}^{-2}\right)\\
&=o\left(1\right). && (15)
\end{aligned}
$$

Moreover,

$$
\begin{aligned}
n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{\left(\log_2 T\right)^2}{T^{2-\gamma}}\left|F_i\left(1\right)\sum_{t=1}^{T}z_{2it}\right|\geq\frac{M}{12}\right)&\leq n\max_{i\in\mathcal{I}_n}\Pr\left(\max_{1\leq t\leq T}\left|z_{it}\right|\geq d_{nT}\right)\\
&\leq nT\max_{i\in\mathcal{I}_n}\max_{1\leq t\leq T}\Pr\left(\left|z_{it}\right|\geq d_{nT}\right)\\
&\leq\frac{nT}{d_{nT}^4}\max_{i\in\mathcal{I}_n}\max_{1\leq t\leq T}\mathbb{E}\left(\left|z_{it}\right|^4\mathbf{1}\left\{\left|z_{it}\right|\geq d_{nT}\right\}\right)\\
&=o\left(\frac{nT^3}{d_{nT}^4}\right)=o_p\left(1\right), && (16)
\end{aligned}
$$

where the rate restrictions $\frac{T^{2-2\gamma}}{\left(\log_2 T\right)^4}>n^{\frac{1}{2}}\log n$ and $\frac{T^{\frac{5}{4}-\gamma}}{\left(\log_2 T\right)^2}>n^{\frac{1}{4}}\log n$ are guaranteed by Assumption 2. Similarly, we can show that

$$
n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{\left(\log_2 T\right)^2}{T^{2-\gamma}}\left|F_i\left(1\right)\sum_{t=1}^{T}\mathbb{E}\left(z_{2it}|\mathcal{F}_{i,t-1}\right)\right|\geq\frac{M}{12}\right)=o\left(1\right). \tag{17}
$$

Combining the above results in (15) (16) and (17), we have

$$
n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{\left(\log_2 T\right)^2}{T^{2-\gamma}}\left|F_i\left(1\right)\sum_{t=1}^{T}z_{it}\right|\geq\frac{M}{4}\right)=o\left(1\right). \tag{18}
$$

For (A.3), we have

$$
\begin{aligned}
&\frac{\left(\log_2 T\right)^2 F_i\left(1\right)}{T^{2-\gamma}}\sum_{t=1}^{T}\left(\sum_{s=1}^{t-1}\epsilon_{is}\right)\left(\widetilde{\epsilon}_{i,t-1}-\widetilde{\epsilon}_{i,t}\right)\\
&=\frac{\left(\log_2 T\right)^2 F_i\left(1\right)}{T^{2-\gamma}}\sum_{s=1}^{T-1}\epsilon_{is}\left(\sum_{t=s+1}^{T}\left(\widetilde{\epsilon}_{i,t-1}-\widetilde{\epsilon}_{i,t}\right)\right)\\
&=\frac{\left(\log_2 T\right)^2 F_i\left(1\right)}{T^{2-\gamma}}\sum_{s=1}^{T-1}\epsilon_{is}\left(\widetilde{\epsilon}_{i,s}-\widetilde{\epsilon}_{i,T}\right)\\
&=\underbrace{\frac{\left(\log_2 T\right)^2 F_i\left(1\right)}{T^{2-\gamma}}\left(\sum_{s=1}^{T-1}\epsilon_{is}\right)\widetilde{\epsilon}_{i,s}}_{\text{(A.3.1)}}-\underbrace{\frac{\left(\log_2 T\right)^2 F_i\left(1\right)}{T^{2-\gamma}}\left(\sum_{s=1}^{T-1}\epsilon_{is}\right)\widetilde{\epsilon}_{i,T}}_{\text{(A.3.2)}}.
\end{aligned}
$$

For the term (A.3.2), by the Markov inequality, we have

$$
n\max_{i\in\mathcal{I}_n}\Pr\left(\left|\frac{\left(\log_2 T\right)^2 F_i\left(1\right)}{T^{2-\gamma}}\left(\sum_{s=1}^{T-1}\epsilon_{is}\right)\widetilde{\epsilon}_{i,T}\right|\geq\frac{M}{16}\right)
$$

$$
\begin{aligned}
&\le n \max_{i\in\mathcal{I}_n} \frac{256\left(\log_2 T\right)^4 \mathbb{E}\left|F_i(1)\left(\sum_{s=1}^{T-1}\epsilon_{is}\right)\widetilde{\epsilon}_{i,T}\right|^2}{T^{4-2\gamma}M^2} \\
&\le \frac{256\left(\log_2 T\right)^4}{T^{4-2\gamma}M^2} \times n \max_{i\in\mathcal{I}_n} F_i^2(1)\,\mathbb{E}\left\{\left(\sum_{s=1}^{T-1}\epsilon_{is}\right)^2 \widetilde{\epsilon}_{i,T}^2\right\} \\
&= O\left(\frac{n\left(\log_2 T\right)^4}{T^{3-2\gamma}}\right) = o(1),
\end{aligned} \tag{19}
$$

where the rate restriction $\frac{T^{3-2\gamma}}{\left(\log_2 T\right)^4} > n$ is ensured by Assumption 2. For the term (A.3.1), it follows that

$$
\begin{aligned}
&\frac{\left(\log_2 T\right)^2 F_i(1)}{T^{2-\gamma}}\left(\sum_{s=1}^{T-1}\epsilon_{is}\right)\widetilde{\epsilon}_{i,s} \\
&= \underbrace{\frac{\left(\log_2 T\right)^2 F_i(1)}{T^{2-\gamma}}\sum_{s=1}^{T-1}\left(\sigma^0\right)^2\widetilde{F}_i(1)}_{\text{(A.3.1.1)}} + \underbrace{\frac{\left(\log_2 T\right)^2 F_i(1)}{T^{2-\gamma}}\sum_{s=1}^{T-1}\left(\epsilon_{is}^2-\left(\sigma^0\right)^2\right)\widetilde{F}_i(1)}_{\text{(A.3.1.2)}} \\
&+ \underbrace{\frac{\left(\log_2 T\right)^2 F_i(1)}{T^{2-\gamma}}\sum_{s=1}^{T-1}\epsilon_{is}\left(\widetilde{\epsilon}_{i,s}-\widetilde{F}_i(1)\epsilon_{i,s}\right)}_{\text{(A.3.1.3)}}.
\end{aligned}
$$

As $n\max_{i\in\mathcal{I}_n}\left|\frac{\left(\log_2 T\right)^2 F_i(1)}{T^{2-\gamma}}\sum_{s=1}^{T-1}\left(\sigma^0\right)^2\widetilde{F}_i(1)\right| \le n\max_{i\in\mathcal{I}_n}\frac{\left(\log_2 T\right)^2}{T^{1-\gamma}}\left|F_i(1)\right|\left(\sigma^0\right)^2\left|\widetilde{F}_i(1)\right| = O\left(\frac{n\left(\log_2 T\right)^2}{T^{1-\gamma}}\right)$, we then have

$$
n\max_{i\in\mathcal{I}_n}\Pr\left(\left|\frac{\left(\log_2 T\right)^2 F_i(1)}{T^{2-\gamma}}\sum_{s=1}^{T-1}\left(\sigma^0\right)^2\widetilde{F}_i(1)\right| \ge \frac{M}{48}\right) = 0, \tag{20}
$$

where the rate restriction $\frac{T^{1-\gamma}}{\left(\log_2 T\right)^2} > n$ is ensured by Assumption 2. Note that $\mathbb{E}\left(\epsilon_{it}^2-\left(\sigma^0\right)^2|\mathcal{F}_{i,t-1}\right) = 0$ for all t. By the Markov inequality, we have

$$
\begin{aligned}
&n\max_{i\in\mathcal{I}_n}\Pr\left(\left|\frac{\left(\log_2 T\right)^2 F_i(1)}{T^{2-\gamma}}\sum_{s=1}^{T-1}\left(\epsilon_{is}^2-\left(\sigma^0\right)^2\right)\widetilde{F}_i(1)\right| \ge \frac{M}{48}\right) \\
&\le K\frac{n\left(\log_2 T\right)^4}{M^2 T^{4-2\gamma}}\max_{i\in\mathcal{I}_n}\mathbb{E}\left(F_i(1)\sum_{s=1}^{T-1}\left(\epsilon_{is}^2-\left(\sigma^0\right)^2\right)\widetilde{F}_i(1)\right)^2 \\
&= O\left(\frac{n\left(\log_2 T\right)^4}{T^{3-2\gamma}}\right) = o(1),
\end{aligned} \tag{21}
$$

where the rate restriction $\frac{T^{3-2\gamma}}{\left(\log_2 T\right)^4} > n$ is assured by Assumption 2. Similarly, by the

Hölder and Jensen inequalities, we have

$$n \max_{i\in\mathcal{I}_n} \Pr\left(\left|\frac{(\log_2 T)^2 F_i(1)}{T^{2-\gamma}}\sum_{s=1}^{T-1}\epsilon_{is}\left(\widetilde{\epsilon}_{i,s}-\widetilde{F}_i(1)\epsilon_{i,s}\right)\right| \geq \frac{M}{48}\right) = o(1), \tag{22}$$

with the rate restriction $\frac{T^{3-2\gamma}}{(\log_2 T)^4} > n$ assured by Assumption 2. Therefore, combining the results in (19) (20) (21) and (22), we have

$$n \max_{i\in\mathcal{I}_n} \Pr\left(\left|\frac{(\log_2 T)^2 F_i(1)}{T^{2-\gamma}}\sum_{t=1}^{T}\left(\sum_{s=1}^{t-1}\epsilon_{is}\right)(\widetilde{\epsilon}_{i,t-1}-\widetilde{\epsilon}_{i,t})\right| > \frac{M}{8}\right) = o(1). \tag{23}$$

For (A.4), we have the decomposition

$$\begin{aligned}
&\frac{(\log_2 T)^2}{T^{2-\gamma}}\sum_{t=1}^{T}(\widetilde{\epsilon}_{i,0}-\widetilde{\epsilon}_{i,t-1})(\widetilde{\epsilon}_{i,t-1}-\widetilde{\epsilon}_{i,t}) \\
&= \underbrace{\frac{(\log_2 T)^2}{T^{2-\gamma}}\widetilde{\epsilon}_{i,0}(\widetilde{\epsilon}_{i,0}-\widetilde{\epsilon}_{i,T})}_{\text{(A.4.1)}} - \underbrace{\frac{(\log_2 T)^2}{T^{2-\gamma}}\sum_{t=1}^{T}\widetilde{\epsilon}_{i,t-1}(\widetilde{\epsilon}_{i,t-1}-\widetilde{\epsilon}_{i,t})}_{\text{(A.4.2)}}.
\end{aligned}$$

For (A.4.1), applying the Markov inequality, we have

$$\begin{aligned}
&n \max_{i\in\mathcal{I}_n} \Pr\left(\left|\frac{(\log_2 T)^2}{T^{2-\gamma}}\widetilde{\epsilon}_{i,0}(\widetilde{\epsilon}_{i,0}-\widetilde{\epsilon}_{i,T})\right| \geq \frac{M}{16}\right) \\
&\leq \frac{256\cdot n(\log_2 T)^4}{T^{4-2\gamma}M^2}\max_{i\in\mathcal{I}_n}\mathbb{E}\left(\widetilde{\epsilon}_{i,0}^2(\widetilde{\epsilon}_{i,0}-\widetilde{\epsilon}_{i,T})^2\right) = o(1),
\end{aligned} \tag{24}$$

and

$$n \max_{i\in\mathcal{I}_n} \Pr\left(\left|\frac{(\log_2 T)^2}{T^{2-\gamma}}\sum_{t=1}^{T}\widetilde{\epsilon}_{i,t-1}(\widetilde{\epsilon}_{i,t-1}-\widetilde{\epsilon}_{i,t})\right| \geq \frac{M}{16}\right) = o(1), \tag{25}$$

where the rate restrictions $\frac{T^{4-2\gamma}}{(\log_2 T)^4} > n$ and $\frac{T^{2-2\gamma}}{(\log_2 T)^4} > n$ are assured by Assumption 2. Combining (24) and (25), we have

$$n \max_{i\in\mathcal{I}_n} \Pr\left(\left|\frac{(\log_2 T)^2}{T^{2-\gamma}}\sum_{t=1}^{T}(\widetilde{\epsilon}_{i,0}-\widetilde{\epsilon}_{i,t-1})(\widetilde{\epsilon}_{i,t-1}-\widetilde{\epsilon}_{i,t})\right| > \frac{M}{8}\right) = o(1). \tag{26}$$

Finally, based on the results in (18) (23) and (26), we have

$$n \max_{i\in\mathcal{I}_n} \Pr\left(\left|\frac{(\log_2 T)^2}{T^{2-\gamma}}\sum_{t=1}^{T}y_{i,t-1}u_{it}\right| \geq \frac{M}{2}\right) = o(1). \tag{27}$$

(ii.2) For the demeaning term, the following decomposition applies

$$
\begin{aligned}
& n \max_{i \in \mathcal{I}_n} \Pr\left(\frac{(\log_2 T)^2}{T^{3-\gamma}} \left| \left(\sum_{t=1}^{T} y_{i,t-1} \right) \left(\sum_{t=1}^{T} u_{it} \right) \right| \geq \frac{M}{2} \right) \\
& \leq n \max_{i \in \mathcal{I}_n} \Pr\left(\frac{(\log_2 T)^2}{T^{3-\gamma}} \left| \left(\sum_{t=1}^{T} y_{i,t-1} \right) \left(\sum_{t=1}^{T} u_{it} \right) \right| \geq \frac{M}{2}, \frac{1}{T^{\frac{9}{4}-\gamma}} \left| \sum_{t=1}^{T} y_{i,t-1} \right| \leq \sqrt{\frac{M}{2}} \right) \\
& + n \max_{i \in \mathcal{I}_n} \Pr\left(\frac{1}{T^{\frac{9}{4}-\gamma}} \left| \sum_{t=1}^{T} y_{i,t-1} \right| \geq \sqrt{\frac{M}{2}} \right) \\
& \leq \underbrace{n \max_{i \in \mathcal{I}_n} \Pr\left(\frac{(\log_2 T)^2}{T^{\frac{3}{4}}} \left| \sum_{t=1}^{T} u_{i,t} \right| \geq \frac{\sqrt{2M}}{2} \right)}_{\text{(A.5)}} + \underbrace{n \max_{i \in \mathcal{I}_n} \Pr\left(\frac{1}{T^{\frac{9}{4}-\gamma}} \left| \sum_{t=1}^{T} y_{i,t-1} \right| \geq \frac{\sqrt{2M}}{2} \right)}_{\text{(A.6)}}.
\end{aligned}
$$

For (A.5), by the Beveridge-Nelson decomposition, we have

$$
\sum_{t=1}^{T} u_{it} = F_i(1) \sum_{t=1}^{T} \epsilon_{it} + \widetilde{\epsilon}_{i,0} - \widetilde{\epsilon}_{i,T},
$$

and hence,

$$
\begin{aligned}
& n \max_{i \in \mathcal{I}_n} \Pr\left(\frac{(\log_2 T)^2}{T^{\frac{3}{4}}} \left| \sum_{t=1}^{T} u_{i,t} \right| \geq \frac{\sqrt{2M}}{2} \right) \\
& \leq \underbrace{n \max_{i \in \mathcal{I}_n} \Pr\left(\frac{(\log_2 T)^2}{T^{\frac{3}{4}}} \left| F_i(1) \sum_{t=1}^{T} \epsilon_{i,t} \right| \geq \frac{\sqrt{2M}}{4} \right)}_{\text{(A.5.1)}} + \underbrace{n \max_{i \in \mathcal{I}_n} \Pr\left(\frac{(\log_2 T)^2}{T^{\frac{3}{4}}} \left| \widetilde{\epsilon}_{i,0} - \widetilde{\epsilon}_{i,T} \right| \geq \frac{\sqrt{2M}}{4} \right)}_{\text{(A.5.2)}}.
\end{aligned}
$$

For the term (A.5.1), by the exponential inequality of Freedman (1975), we have

$$
n \max_{i \in \mathcal{I}_n} \Pr\left(\frac{(\log_2 T)^2}{T^{\frac{3}{4}}} \left| F_i(1) \sum_{t=1}^{T} \epsilon_{i,t} \right| \geq \frac{\sqrt{2M}}{4} \right) = o(1), \tag{28}
$$

where the rate restrictions $\frac{\sqrt{T}}{(\log_2 T)^4} \succ \sqrt{n} \log n$ and $\frac{\sqrt{T}}{(\log_2 T)^2} \succ n^{\frac{1}{4}} \log n$ are ensured by Assumption 2. For the term (A.5.2), by the Markov inequality, we have

$$
n \max_{i \in \mathcal{I}_n} \Pr\left(\frac{(\log_2 T)^2}{T^{\frac{3}{4}}} \left| \widetilde{\epsilon}_{i,0} - \widetilde{\epsilon}_{i,T} \right| \geq \frac{\sqrt{2M}}{4} \right) = o(1), \tag{29}
$$

with the rate restriction $\frac{T^{\frac{3}{2}}}{(\log_2 T)^4} \succ n$ assured by Assumption 2. For the term (A.6), by the Beveridge-Nelson decomposition (Phillips and Solo, 1992), we prove a more restrictive case

$$
n \max_{i \in \mathcal{I}_n} \Pr\left(\frac{(\log_2 T)^2}{T^{\frac{9}{4}-\gamma}} \left| \sum_{t=1}^{T} y_{i,t-1} \right| \geq \frac{\sqrt{2M}}{2} \right)
$$

$$\leq \underbrace{n \max_{i \in \mathcal{I}_n} \Pr\left(\frac{(\log_2 T)^2}{T^{\frac{9}{4}-\gamma}} \left| F_i(1) \sum_{s=0}^{T-1} \left(\sum_{t=s+1}^{T} 1 \right) \epsilon_{is} \right| \geq \frac{\sqrt{2M}}{4} \right)}_{\text{(A.6.1)}}$$

$$+ \underbrace{n \max_{i \in \mathcal{I}_n} \Pr\left(\frac{(\log_2 T)^2}{T^{\frac{9}{4}-\gamma}} \left| \sum_{s=0}^{T-1} \left(\sum_{t=s+1}^{T} 1 \right) (\widetilde{\epsilon}_{i,s-1} - \widetilde{\epsilon}_{i,s}) \right| \geq \frac{\sqrt{2M}}{4} \right)}_{\text{(A.6.2)}},$$

where $y_{i0} = 0$. For the term (A.6.1), by the exponential inequality (Freedman, 1975), we have

$$n \max_{i \in \mathcal{I}_n} \Pr\left(\frac{(\log_2 T)^2}{T^{\frac{9}{4}-\gamma}} \left| F_i(1) \sum_{s=0}^{T-1} \left(\sum_{t=s+1}^{T} 1 \right) \epsilon_{is} \right| \geq \frac{\sqrt{2M}}{4} \right) = o(1), \tag{30}$$

where the rate restrictions $\frac{T^{\frac{3}{2}-2\gamma}}{(\log_2 T)^4} \succ \sqrt{n} \log n$ and $\frac{T^{1-\gamma}}{(\log_2 T)^2} \succ n^{\frac{1}{4}} \log n$ follow by Assumption 2. For (A.6.2), by the Markov inequality, we have

$$n \max_{i \in \mathcal{I}_n} \Pr\left(\frac{(\log_2 T)^2}{T^{\frac{9}{4}-\gamma}} \left| \sum_{s=0}^{T-1} \left(\sum_{t=s+1}^{T} 1 \right) (\widetilde{\epsilon}_{i,s-1} - \widetilde{\epsilon}_{i,s}) \right| \geq \frac{\sqrt{2M}}{4} \right) = o(1), \tag{31}$$

where the rate restriction $\frac{T^{\frac{3}{2}-2\gamma}}{(\log_2 T)^4} \succ n$ follows by Assumption 2. Therefore, based on (28)–(31), we have

$$n \max_{i \in \mathcal{I}_n} \Pr\left(\frac{(\log_2 T)^2}{T^{3-\gamma}} \left| \left(\sum_{t=1}^{T} y_{i,t-1} \right) \left(\sum_{t=1}^{T} u_{i,t} \right) \right| \geq \frac{M}{2} \right) = o(1). \tag{32}$$

At last, by (27) and (32), we have

$$\max_{i \in \mathcal{I}_n} \Pr\left(\frac{(\log_2(T))^2}{T^{2-\gamma}} \left| \sum_{t=1}^{T} \widetilde{y}_{i,t-1} \widetilde{u}_{it} \right| \geq M \right) = o\left(\frac{1}{n} \right).$$

(iii) When $\bar{c}_i^0 < 0$, for any $M > 0$, the demeaned numerator similarly decomposes as

$$\begin{aligned} & n \max_{i \in \mathcal{I}_n} \Pr\left(\frac{1}{T} \left| \sum_{t=1}^{T} \widetilde{y}_{i,t-1} \widetilde{u}_{it} \right| \geq M \right) \\ & \leq n \max_{i \in \mathcal{I}_n} \Pr\left(\frac{1}{T} \left| \sum_{t=1}^{T} y_{i,t-1} u_{it} \right| \geq \frac{M}{2} \right) + n \max_{i \in \mathcal{I}_n} \Pr\left(\frac{1}{T^2} \left| \left(\sum_{t=1}^{T} y_{i,t-1} \right) \left(\sum_{t=1}^{T} u_{it} \right) \right| \geq \frac{M}{2} \right), \end{aligned} \tag{33}$$

whose asymptotic negligibility follows by the Markov inequality and the exponential inequality (Freedman, 1975). The details are given in the proof of Lemma A.3 of the Online Supplement to Liu et al. (2022). ■

Lemma A.4 *Suppose that Assumptions 1 and 2 hold, then,*

(i) if $\overline{c}_i^0 > 0$ *and* $\widetilde{M}_1 \geq \frac{1}{c_{low}^2} \max_{i\in\mathcal{I}_n} \left(\overline{\omega}_i^0\right)^2$,

$$\max_{i\in\mathcal{I}_n} \Pr\left(\frac{1}{\left(\overline{\rho}_i^0\right)^{2T} T^{2\gamma} (\log T)^2} \left|\sum_{t=1}^{T} \widetilde{y}_{i,t-1}^2\right| \geq \widetilde{M}_1\right) = o\left(\frac{1}{n}\right);$$

(ii) if $\overline{c}_i^0 = 0$ *and* $\widetilde{M}_2 \geq \max_{i\in\mathcal{I}_n} \left(\overline{\omega}_i^0\right)^2$,

$$\max_{i\in\mathcal{I}_n} \Pr\left(\frac{1}{T^2 (\log_2 T)^2} \left|\sum_{t=1}^{T} \widetilde{y}_{i,t-1}^2\right| \geq \widetilde{M}_2\right) = o\left(\frac{1}{n}\right);$$

(iii) if $\overline{c}_i^0 < 0$ *and* $\widetilde{M}_3 \geq \frac{2\max_{i\in\mathcal{I}_n}\left(\overline{\sigma}_{iu}^0\right)^2}{c_{low}}$,

$$\max_{i\in\mathcal{I}_n} \Pr\left(\frac{1}{T^{1+\gamma}} \left|\sum_{t=1}^{T} \widetilde{y}_{i,t-1}^2\right| \geq \widetilde{M}_3\right) = o\left(\frac{1}{n}\right).$$

Proof of Lemma A.4: (i) If $\overline{c}_i^0 > 0$, we decompose $\frac{1}{\left(\overline{\rho}_i^0\right)^{2T} T^{2\gamma}(\log T)^2} \left|\sum_{t=1}^{T} \widetilde{y}_{i,t-1}^2\right|$ as

$$
\begin{aligned}
& n \max_{i\in\mathcal{I}_n} \Pr\left(\frac{1}{\left(\overline{\rho}_i^0\right)^{2T} T^{2\gamma} (\log T)^2} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}^2 \geq \widetilde{M}_1\right) \\
& \leq n \max_{i\in\mathcal{I}_n} \Pr\left(\frac{1}{\left(\overline{\rho}_i^0\right)^{2T} T^{2\gamma} (\log T)^2} \sum_{t=1}^{T} y_{i,t-1}^2 \geq \frac{\widetilde{M}_1}{2}\right) \\
& + n \max_{i\in\mathcal{I}_n} \Pr\left(\frac{1}{\left(\overline{\rho}_i^0\right)^{2T} T^{2\gamma+1} (\log T)^2} \left(\sum_{t=1}^{T} y_{i,t-1}\right)^2 \geq \frac{\widetilde{M}_1}{2}\right) \\
& \leq \underbrace{n \max_{i\in\mathcal{I}_n} \Pr\left(\frac{1}{2\overline{c}_i^0 \left(\overline{\rho}_i^0\right)^{2T} T^{\gamma} (\log T)^2} \left|y_{i,0}^2\right| \geq \frac{\widetilde{M}_1}{10}\right)}_{\text{(C.1)}} \\
& + \underbrace{n \max_{i\in\mathcal{I}_n} \Pr\left(\frac{1}{2\overline{c}_i^0 \left(\overline{\rho}_i^0\right)^{2T} T^{\gamma} (\log T)^2} \left(y_{i,T}^2 - \mathbb{E} y_{i,T}^2\right) \geq \frac{\widetilde{M}_1}{10}\right)}_{\text{(C.2)}} \\
& + \underbrace{n \max_{i\in\mathcal{I}_n} \Pr\left(\frac{1}{\overline{c}_i^0 \left(\overline{\rho}_i^0\right)^{2T-1} T^{\gamma} (\log T)^2} \left|\sum_{t=1}^{T} y_{i,t-1} u_{it}\right| \geq \frac{\widetilde{M}_1}{10}\right)}_{\text{(C.3)}}
\end{aligned}
$$

$$
\underbrace{+n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{2\bar{c}_i^0\left(\bar{\rho}_i^0\right)^{2T}T^{\gamma}\left(\log T\right)^2}\left|\sum_{t=1}^{T}u_{it}^2\right|\geq\frac{\widetilde{M}_1}{10}\right)}_{\text{(C.4)}}
$$

$$
\underbrace{+n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{2\bar{c}_i^0\left(\bar{\rho}_i^0\right)^{2T}T^{\gamma}\left(\log T\right)^2}\left|T\mu_i^2+2\bar{\rho}_i^0\mu_i\sum_{t=1}^{T}y_{i,t-1}+2\mu_i\sum_{t=1}^{T}u_{it}\right|\geq\frac{\widetilde{M}_1}{10}\right)}_{\text{(C.5)}}
$$

$$
\underbrace{+n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{\left(\bar{\rho}_i^0\right)^{2T}T^{2\gamma+1}\left(\log T\right)^2}\left(\sum_{t=1}^{T}y_{i,t-1}\right)^2\geq\frac{\widetilde{M}_1}{6}\right)}_{\text{(C.6)}}
$$

$$
\underbrace{+n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{2\bar{c}_i^0\left(\bar{\rho}_i^0\right)^{2T}T^{\gamma}\left(\log T\right)^2}\left|\mathbb{E}y_{i,T}^2\right|\geq\frac{\widetilde{M}_1}{3}\right)}_{\text{(C.7)}}.
$$

For (C.4), by the Markov inequality, we have

$$
\begin{aligned}
&n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{2\bar{c}_i^0\left(\bar{\rho}_i^0\right)^{2T}T^{\gamma}\left(\log T\right)^2}\left|\sum_{t=1}^{T}u_{it}^2\right|\geq\frac{\widetilde{M}_1}{10}\right)\leq n\max_{i\in\mathcal{I}_n}\frac{10\cdot\mathbb{E}\left(\sum_{t=1}^{T}\left(u_{it}^2\right)\right)}{2\left(\bar{c}_i^0\right)\left(\bar{\rho}_i^0\right)^{2T}T^{\gamma}\left(\log T\right)^2\widetilde{M}_1}\\
&=O\left(\max_{i\in\mathcal{I}_n}\frac{nT}{\left(\bar{\rho}_i^0\right)^{2T}T^{\gamma}\left(\log T\right)^2}\right)=O\left(\frac{nT^{1-\gamma}}{\left(\rho_{low}\right)^{2T}\left(\log T\right)^2}\right)=o\left(1\right),
\end{aligned}\tag{34}
$$

where the asymptotic negligibility is assured by the dominance of the exponential rates. For the term (C.1), we have

$$
\begin{aligned}
&n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{2\bar{c}_i^0\left(\bar{\rho}_i^0\right)^{2T}T^{\gamma}\left(\log T\right)^2}\left|y_{i,0}^2\right|\geq\frac{\widetilde{M}_1}{10}\right)\leq n\max_{i\in\mathcal{I}_n}\frac{10\mathbb{E}y_{i,0}^2}{2\left(\bar{c}_i^0\right)\left(\bar{\rho}_i^0\right)^{2T}T^{\gamma}\left(\log T\right)^2\left(\widetilde{M}_1\right)}\\
&\leq O\left(\max_{i\in\mathcal{I}_n}\frac{n}{\left(\bar{\rho}_i^0\right)^{2T}T^{\gamma}\left(\log T\right)^2}\right)=O\left(\frac{n}{\left(\rho_{low}\right)^{2T}T^{\gamma}\left(\log T\right)^2}\right)=o\left(1\right),
\end{aligned}\tag{35}
$$

where the exponential rates leads to asymptotic negligibility. Lemma A.3 proves the validity of (C.3) under Assumptions 1 and 2. For the term (C.2), by the exponential inequality for χ^2 variates (Laurent and Massart, 2000), we have

$$
\limsup_{T\to\infty}\frac{\left(y_{i,T}^2-\mathbb{E}y_{i,T}^2\right)}{\left(\bar{\rho}_i^0\right)^{2T}T^{\gamma}\left(\log T\right)^2}<K,
$$

with constant $K > 0$, since

$$\begin{aligned}
&\Pr\left(\max_{1\le t\le T}\left(y_{i,T}^2 - \mathbb{E}y_{i,T}^2\right) \ge \sqrt{2M_T\mathbb{E}y_{i,T}^2} + 2M_T\right)\\
&\le T\max_{1\le t\le T}\Pr\left(\left(y_{i,T}^2 - \mathbb{E}y_{i,T}^2\right) \ge \sqrt{2M_T\mathbb{E}y_{i,T}^2} + 2M_T\right)\\
&= T\max_{1\le t\le T}\Pr\left(\frac{\left(y_{i,T}^2 - \mathbb{E}y_{i,T}^2\right)}{\left(\overline{\rho}_i^0\right)^{2T}T^\gamma} \ge \frac{2M_T}{\left(\overline{\rho}_i^0\right)^{2T}T^\gamma} + \frac{\sqrt{2M_T\mathbb{E}y_{i,T}^2}}{\left(\overline{\rho}_i^0\right)^{2T}T^\gamma}\right) \le T\exp\left(-\frac{M_T}{\left(\overline{\rho}_i^0\right)^{2T}T^\gamma}\right) = o(1),
\end{aligned}$$

with $M_T = K\cdot\left(\overline{\rho}_i^0\right)^{2T}T^\gamma(\log T)^2$. Therefore,

$$n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{2\overline{c}_i^0\left(\overline{\rho}_i^0\right)^{2T}T^\gamma(\log T)^2}\left(y_{i,T}^2 - \mathbb{E}y_{i,T}^2\right) \ge \frac{\widetilde{M}_1}{10}\right) = 0. \tag{36}$$

For term (C.6), as $\overline{c}_i^0$ is a postive constant, it is equivalent to show

$$\begin{aligned}
&n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{2\overline{c}_i^0\left(\overline{\rho}_i^0\right)^{2T}T^{2\gamma+1}(\log T)^2}\left(\sum_{t=1}^T y_{i,t-1}\right)^2 \ge \frac{\widetilde{M}_1}{6}\right)\\
&\le n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{\left(\overline{\rho}_i^0\right)^{T}T^{\gamma+\frac{1}{2}}(\log T)}\left|\sum_{t=1}^T y_{i,t-1}\right| \ge \frac{\sqrt{3\overline{c}_i^0\widetilde{M}_1}}{3}\right)\\
&\le n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{\left(\overline{\rho}_i^0\right)^{T}T^{\gamma+\frac{1}{2}}(\log T)}\left|\frac{T^\gamma}{\overline{c}_i^0}\sum_{t=1}^{T-1}\left(\left(\overline{\rho}_i^0\right)^{T-t}-1\right)u_{it} + \sum_{t=1}^T\left(\overline{\rho}_i^0\right)^{t-1}y_{i0}\right| \ge \frac{\sqrt{3\overline{c}_i^0\widetilde{M}_1}}{3}\right)\\
&\le n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{\overline{c}_i^0 T^{\frac{1}{2}}(\log T)}\left|\sum_{t=1}^{T-1}\left(\overline{\rho}_i^0\right)^{-t}u_{it}\right| \ge \frac{\sqrt{3\overline{c}_i^0\widetilde{M}_1}}{6}\right)\\
&+ n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{\left(\overline{\rho}_i^0\right)^{T}T^{\gamma+\frac{1}{2}}(\log T)}\left|\sum_{t=1}^T\left(\overline{\rho}_i^0\right)^{t-1}y_{i0}\right| \ge \frac{\sqrt{3\overline{c}_i^0\widetilde{M}_1}}{12}\right)\\
&+ n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{\overline{c}_i^0\left(\overline{\rho}_i^0\right)^{T}T^{\frac{1}{2}}(\log T)}\left|\sum_{t=1}^{T-1}u_{it}\right| \ge \frac{\sqrt{3\overline{c}_i^0\widetilde{M}_1}}{12}\right)\\
&= n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{\left(\overline{c}_i^0\right)\sqrt{T}(\log T)}\left|\sum_{t=1}^{T-1}\left(\overline{\rho}_i^0\right)^{-t}u_{it}\right| \ge \frac{\sqrt{3\overline{c}_i^0\widetilde{M}_1}}{6}\right) + o(1) = o(1),
\end{aligned}\tag{37}$$

by the dominance of the exponential rates, the Markov inequality, the rate restriction $T^{1-\gamma} > n$ and the following results:

$$n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{\left(\overline{\rho}_i^0\right)^{T}T^{\gamma+\frac{1}{2}}(\log T)}\left|\sum_{t=1}^T\left(\overline{\rho}_i^0\right)^{t-1}y_{i0}\right| \ge \frac{\sqrt{3\overline{c}_i^0\widetilde{M}_1}}{12}\right) \le O\left(\frac{n}{T^{3+\gamma}(\log T)^2}\right) = o(1),$$

and

$$\begin{aligned}
&n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{\bar{c}_i^0\sqrt{T}\left(\log T\right)}\left|\sum_{t=1}^{T-1}\left(\bar{\rho}_i^0\right)^{-t}u_{it}\right|\geq\frac{\sqrt{3\bar{c}_i^0\widetilde{M}_1}}{6}\right)\\
&\leq n\max_{i\in\mathcal{I}_n}\frac{K\mathbb{E}\left(\sum_{t=1}^{T-1}\left(\bar{\rho}_i^0\right)^{-t}u_{it}\right)^2}{\widetilde{M}_1T\left(\log T\right)^2}=O\left(\frac{n}{T^{1-\gamma}\left(\log T\right)^2}\right),
\end{aligned}$$

where K denotes some constant value,

$$\mathbb{E}\left(\sum_{t=1}^{T-1}\left(\bar{\rho}_i^0\right)^{-t}u_{it}\right)^2\sim\sum_{t=1}^{T-1}\left(\bar{\rho}_i^0\right)^{-2t}F_i^2\left(1\right)\mathbb{E}\epsilon_{it}^2=O\left(T^{\gamma}\right),$$

due to the Beveridge-Nelson decomposition (Phillips and Solo, 1992). For the term (C.5), we have

$$\begin{aligned}
&n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{2\bar{c}_i^0\left(\bar{\rho}_i^0\right)^{2T}T^{\gamma}\left(\log T\right)^2}\left|T\mu_i^2+2\mu_i\bar{\rho}_i^0\sum_{t=1}^{T}y_{i,t-1}+2\mu_i\sum_{t=1}^{T}u_{it}\right|\geq\frac{\widetilde{M}_1}{10}\right)\\
&\leq n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{2\bar{c}_i^0\left(\bar{\rho}_i^0\right)^{2T}T^{\gamma}\left(\log T\right)^2}\left|T\mu_i^2\right|\geq\frac{\widetilde{M}_1}{30}\right)\\
&+n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{2\bar{c}_i^0\left(\bar{\rho}_i^0\right)^{2T-1}T^{\gamma}\left(\log T\right)^2}\left|2\mu_i\sum_{t=1}^{T}y_{i,t-1}\right|\geq\frac{\widetilde{M}_1}{20}\right)\\
&+n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{2\bar{c}_i^0\left(\bar{\rho}_i^0\right)^{2T}T^{\gamma}\left(\log T\right)^2}\left|2\mu_i\sum_{t=1}^{T}u_{it}\right|\geq\frac{\widetilde{M}_1}{60}\right)\\
&\leq n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{2\bar{c}_i^0\left(\bar{\rho}_i^0\right)^{2T}T^{\gamma}\left(\log T\right)^2}\left|T\mu_i^2\right|\geq\frac{\widetilde{M}_1}{30}\right)\\
&+n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{2\left(\bar{c}_i^0\right)^2\left(\bar{\rho}_i^0\right)^{T-1}\left(\log T\right)^2}\left|2\mu_i\sum_{t=1}^{T-1}\left(\bar{\rho}_i^0\right)^{-t}u_{it}\right|\geq\frac{\widetilde{M}_1}{60}\right)\\
&+n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{2\left(\bar{c}_i^0\right)^2\left(\bar{\rho}_i^0\right)^{2T-1}\left(\log T\right)^2}\left|2\mu_i\sum_{t=1}^{T-1}u_{it}\right|\geq\frac{\widetilde{M}_1}{60}\right)\\
&+n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{2\left(\bar{c}_i^0\right)^2\left(\bar{\rho}_i^0\right)^{2T-1}\left(\log T\right)^2}\left|2\mu_i\sum_{t=1}^{T}\left(\bar{\rho}_i^0\right)^{t-1}y_{i0}\right|\geq\frac{\widetilde{M}_1}{60}\right)\\
&+n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{2\bar{c}_i^0\left(\bar{\rho}_i^0\right)^{2T}T^{\gamma}\left(\log T\right)^2}\left|2\mu_i\sum_{t=1}^{T}u_{it}\right|\geq\frac{\widetilde{M}_1}{60}\right)=o\left(1\right),
\end{aligned}\tag{38}$$

by the Markov inequality and the dominance of the exponential rates over the polynomial rate. For the term (C.7), since

$$\frac{1}{2\overline{c}_i^0\left(\overline{\rho}_i^0\right)^{2T}T^{\gamma}}\mathbb{E}y_{i,T}^2 \sim \frac{1}{2\overline{c}_i^0\left(\overline{\rho}_i^0\right)^{2T}T^{\gamma}}\left(\frac{\left(\overline{\rho}_i^0\right)^{2T}-1}{\left(\overline{\rho}_i^0\right)^{2}-1}\right)\left(\overline{\omega}_i^0\right)^2 \to \frac{\left(\overline{\omega}_i^0\right)^2}{4\left(\overline{c}_i^0\right)^2},$$

if $\widetilde{M}_1 \geq \frac{1}{c_{low}^2}\max_{i\in\mathcal{I}_n}\left(\overline{\omega}_i^0\right)^2$, we have

$$n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{2\overline{c}_i^0\left(\overline{\rho}_i^0\right)^{2T}T^{\gamma}(\log T)^2}\mathbb{E}y_{i,T}^2 > \frac{\widetilde{M}_1}{3}\right) = 0. \tag{39}$$

Combining the results of (34)–(39) and Lemma A.3, we can show that

$$\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{\left(\overline{\rho}_i^0\right)^{2T}T^{2\gamma}(\log T)^2}\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2\right| \geq \widetilde{M}_1\right) = o\left(\frac{1}{n}\right).$$

(ii) If $\overline{c}_i^0 = 0$, by Lemma A.3 of Huang et al. (2021),

$$\limsup_{T\to\infty}\left(\frac{1}{T^2(\log_2 T)^2}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2\right) \leq \left(\frac{1}{2}+\varepsilon\right)\left(\overline{\omega}_i^0\right)^2,$$

for any $\varepsilon > 0$. Therefore,

$$\limsup_{T\to\infty}\left(\frac{1}{T^2(\log_2 T)^2}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2\right) \leq \left(\frac{1}{2}+\varepsilon\right)\left(\overline{\omega}_i^0\right)^2, \ \ \forall i \in \mathcal{I}_n.$$

With the above uniform upper bound, it follows that

$$n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{T^2(\log_2 T)^2}\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2\right| \geq \widetilde{M}_2\right) = 0,$$

where $\widetilde{M}_2 \geq \max_{i\in\mathcal{I}_n}\left(\overline{\omega}_i^0\right)^2$.
(iii) If $\overline{c}_i^0 < 0$, the following decomposition can be applied to the sample moment:

$$\begin{aligned}
&n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{T^{1+\gamma}}\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2\right| \geq \widetilde{M}_3\right)\\
&\leq \underbrace{n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{-2\overline{c}_i^0 T}\left|\sum_{t=1}^{T}\mathbb{E}u_{it}^2\right| \geq \frac{\widetilde{M}_3}{3}\right)}_{\text{(D.1)}} + \underbrace{n\max_{i\in\mathcal{I}_n}\Pr\left(\left|\frac{1}{T^{1+\gamma}}\sum_{t=1}^{T}y_{i,t-1}^2 - \frac{1}{-2\overline{c}_i^0 T}\sum_{t=1}^{T}\mathbb{E}u_{it}^2\right| \geq \frac{\widetilde{M}_3}{3}\right)}_{\text{(D.2)}}
\end{aligned}$$

$$\underbrace{+n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{T^{2+\gamma}}\left(\sum_{t=1}^{T}y_{i,t-1}\right)^2\geq\frac{\widetilde{M}_3}{3}\right)}_{(\text{D.3})}.$$

The asymptotic negligiblity of (D.1), (D.2) and (D.3) is similarly proved using the Markov inequality and exponential inequality (Freedman, 1975). The details are given in the proof of Lemma A.4 of the Online Supplement to Liu et al. (2022). Combining the above results, we have

$$n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{T^{1+\gamma}}\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2\right|\geq\widetilde{M}_3\right)=o(1),$$

with $\widetilde{M}_3\geq\frac{2\max_{i\in\mathcal{I}_n}\left(\overline{\sigma}_{iu}^0\right)^2}{c_{low}}$ and $\overline{c}_i^0<0$. The proof is then complete. ■

Lemma A.5 *Suppose Assumptions 1 and 2 hold.*

(i) If $\overline{c}_i^0>0$ *and* $0<\overline{M}_1\leq\frac{1}{16c_{up}^2}\min_{i\in\mathcal{I}_n}\left(\overline{\omega}_i^0\right)^2$,

$$\max_{i\in\mathcal{I}_n}\Pr\left(\frac{(\log T)^2}{\left(\overline{\rho}_i^0\right)^{2T}T^{2\gamma}}\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2\right|\leq\overline{M}_1\right)=o\left(\frac{1}{n}\right).$$

(ii) If $\overline{c}_i^0=0$ *and* $0<\overline{M}_2\leq\min_{i\in\mathcal{I}_n}\frac{\left(\overline{\omega}_i^0\right)^2}{24}$,

$$\max_{i\in\mathcal{I}_n}\Pr\left(\frac{(\log_2 T)^2}{T^2}\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2\right|\leq\overline{M}_2\right)=o\left(\frac{1}{n}\right).$$

(iii) If $\overline{c}_i^0<0$ *and* $0<\overline{M}_3\leq\frac{\min_{i\in\mathcal{I}_n}\left(\overline{\sigma}_{iu}^0\right)^2}{8c_{up}}$,

$$\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{T^{1+\gamma}}\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2\right|\leq\overline{M}_3\right)=o\left(\frac{1}{n}\right).$$

Proof of Lemma A.5: The proof of Lemma A.5(iii) follows that of Lemma A.4(iii). For (ii), if $\overline{c}_i^0=0$, by Lemma A.3 of Huang et al. (2021),

$$\lim_{T\to\infty}\inf\left(\frac{(\log_2 T)^2}{T^2}\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2\right|\right)\geq\frac{\left(\overline{\omega}_i^0\right)^2}{12},\quad\forall i=1,2,...,n.$$

Therefore,

$$n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{(\log_2 T)^2}{T^2}\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2\right|\leq\overline{M}_2\right)=0,\quad\forall i=1,2,...,n.$$

The proof of Lemma A.5 (i) follows that of Lemma A.4 (i), the iterated logarithm law for quadratic forms of martingales (Fernholz and Teicher, 1980; Donsker and Varadhan, 1977), and the fact that

$$\begin{aligned}\inf_{1\le t\le T} y_{iT}^2 &= -\max_{1\le t\le T}\left(\mathbb{E}y_{i,T}^2 - y_{i,T}^2\right) + \inf_{1\le t\le T}\mathbb{E}y_{i,T}^2\\ &\le K\cdot\left[\left(\overline{\rho}_i^0\right)^{2T} T^{2\gamma}(\log T)^2 + \left(\overline{\rho}_i^0\right)^{2T} T^{2\gamma}\right],\end{aligned}$$

by Lemma 1 of Laurent and Massart (2000) and Donsker and Varadhan (1977, Formula (4.6)), which justifies the lim inf for the quadratic form of martingales. The proof is then complete. ■

Lemma A.6 *Suppose Assumptions 1 and 2 hold. Let $\eta = O\left(\frac{1}{T^\gamma(\log T)^4}\right)$. Then, when $(n,T)\to\infty$,*

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

where $\mathcal{N}_\eta$ is defined in (10).

Proof of Lemma A.6: By the definition of $\widehat{g}_i(\cdot)$, we have

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

for any $j=1,2,\dots,G^0$. Note that

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

where $Z_{ij}(c) = \mathbf{1}\left\{g_i^0\neq j\right\}\mathbf{1}\left\{\sum_{t=1}^T(\widetilde{y}_{it}-\widetilde{y}_{i,t-1}\rho_j)^2\le\sum_{t=1}^T(\widetilde{y}_{it}-\widetilde{y}_{i,t-1}\rho_{g_i^0})^2\right\}$, $\rho_j = 1+c_j/T^\gamma$ and $\rho_{g_i^0} = 1+c_{g_i^0}/T^\gamma$. We intend to bound $Z_{ij}(c)$ for all $c\in\mathcal{N}_\eta$ by the arguments that are free of parameter c. Therefore, for any $i\in\mathcal{I}_n$,

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

Define

$$
\begin{aligned}
H_T &= \left| \begin{array}{c} \sum_{t=1}^{T} \widetilde{y}_{i,t-1} \left(\rho_{\widetilde{j}} - \rho_j\right)\left(2\widetilde{y}_{i,t-1}\rho_{\widetilde{j}}^0 + 2\widetilde{u}_{it} - \widetilde{y}_{i,t-1}\left(\rho_{\widetilde{j}} + \rho_j\right)\right) \\ -\sum_{t=1}^{T} \widetilde{y}_{i,t-1} \left(\rho_{\widetilde{j}}^0 - \rho_j^0\right)\left(2\widetilde{y}_{i,t-1}\rho_{\widetilde{j}}^0 + 2\widetilde{u}_{it} - \widetilde{y}_{i,t-1}\left(\rho_{\widetilde{j}}^0 + \rho_j^0\right)\right) \end{array} \right| \\
&\leq \left| 2\sum_{t=1}^{T} (\rho_{\widetilde{j}} - \rho_j)\widetilde{y}_{i,t-1}\widetilde{u}_{it} - 2\sum_{t=1}^{T} (\rho_{\widetilde{j}}^0 - \rho_j^0)\widetilde{y}_{i,t-1}\widetilde{u}_{it} \right| \\
&+ \left| \begin{array}{c} \sum_{t=1}^{T} \widetilde{y}_{i,t-1} \left(\rho_{\widetilde{j}} - \rho_j\right)\left(2\widetilde{y}_{i,t-1}\rho_{\widetilde{j}}^0 - \widetilde{y}_{i,t-1}\left(\rho_{\widetilde{j}} + \rho_j\right)\right) \\ -\sum_{t=1}^{T} \widetilde{y}_{i,t-1} \left(\rho_{\widetilde{j}}^0 - \rho_j^0\right)\left(2\widetilde{y}_{i,t-1}\rho_{\widetilde{j}}^0 - \widetilde{y}_{i,t-1}\left(\rho_{\widetilde{j}}^0 + \rho_j^0\right)\right) \end{array} \right| \\
&=: H_{1T} + H_{2T},
\end{aligned}
$$

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

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

By compactness of the parameter space and the definition of η,

$$
H_{1T} = \left| 2\sum_{t=1}^{T} (\rho_{\widetilde{j}} - \rho_j)\widetilde{y}_{i,t-1}\widetilde{u}_{it} - 2\sum_{t=1}^{T} (\rho_{\widetilde{j}}^0 - \rho_j^0)\widetilde{y}_{i,t-1}\widetilde{u}_{it} \right| \leq \frac{B_1}{T^\gamma}\eta \left| \sum_{t=1}^{T} \widetilde{y}_{i,t-1}\widetilde{u}_{it} \right|,
$$

where B_1 is a constant independent of η and T. For H_{2T}, we have, with B_2 as a constant independent of η and T,

$$
\begin{aligned}
H_{2T} &= \left| \begin{array}{c} \sum_{t=1}^{T} \widetilde{y}_{i,t-1} \left(\rho_{\widetilde{j}} - \rho_j\right)\left(2\widetilde{y}_{i,t-1}\rho_{\widetilde{j}}^0 - \widetilde{y}_{i,t-1}\left(\rho_{\widetilde{j}} + \rho_j\right)\right) \\ -\sum_{t=1}^{T} \widetilde{y}_{i,t-1} \left(\rho_{\widetilde{j}}^0 - \rho_j^0\right)\left(2\widetilde{y}_{i,t-1}\rho_{\widetilde{j}}^0 - \widetilde{y}_{i,t-1}\left(\rho_{\widetilde{j}}^0 + \rho_j^0\right)\right) \end{array} \right| \\
&= 2\left| \sum_{t=1}^{T} \widetilde{y}_{i,t-1}^2 \right| \left| \rho_{\widetilde{j}}^0\left(\rho_{\widetilde{j}} - \rho_{\widetilde{j}}^0 - \rho_j + \rho_j^0\right) + \frac{1}{2}\left((\rho_{\widetilde{j}}^0)^2 - \rho_{\widetilde{j}}^2 + \rho_j^2 - \left(\rho_j^0\right)^2\right) \right| \\
&\leq \frac{B_2}{T^\gamma}\eta \left| \sum_{t=1}^{T} \widetilde{y}_{i,t-1}^2 \right|.
\end{aligned}
$$

Combining the above results, we obtain

$$
\begin{aligned}
Z_{ij}(c) &\leq \max_{\widetilde{j} \neq j} \mathbf{1}\left\{ \sum_{t=1}^{T} \widetilde{y}_{i,t-1}\left(\rho_{\widetilde{j}} - \rho_j\right)\left(2\widetilde{y}_{i,t-1}\rho_{\widetilde{j}}^0 + 2\widetilde{u}_{it} - \widetilde{y}_{i,t-1}\left(\rho_{\widetilde{j}} + \rho_j\right)\right) \leq 0 \right\} \\
&\leq \max_{\widetilde{j} \neq j} \mathbf{1}\left\{ \sum_{t=1}^{T} \widetilde{y}_{i,t-1}\left(\rho_{\widetilde{j}}^0 - \rho_j^0\right)\left(2\widetilde{y}_{i,t-1}\rho_{\widetilde{j}}^0 + 2\widetilde{u}_{it} - \widetilde{y}_{i,t-1}\left(\rho_{\widetilde{j}}^0 + \rho_j^0\right)\right) \right. \\
&\qquad \left. \leq \frac{B_1}{T^\gamma}\eta \left| \sum_{t=1}^{T} \widetilde{y}_{i,t-1}\widetilde{u}_{it} \right| + \frac{B_2}{T^\gamma}\eta \left| \sum_{t=1}^{T} \widetilde{y}_{i,t-1}^2 \right| \right\}.
\end{aligned}
$$

Since

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

we define

$$\widetilde{Z}_{ij}=\max_{\widetilde{j}\neq j}\mathbf{1}\left\{\left(\rho_{\widetilde{j}}^{0}-\rho_{j}^{0}\right)^{2}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^{2}+2\left(\rho_{\widetilde{j}}^{0}-\rho_{j}^{0}\right)\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\leq\frac{B_1}{T^{\gamma}}\eta\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right|+\frac{B_2}{T^{\gamma}}\eta\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^{2}\right|\right\}.$$

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

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

In the following, we intend to bound the clustering error in three cases: Case I ($c_{\widetilde{j}}^0>0$), Case II ($c_{\widetilde{j}}^0=0$), Case III ($c_{\widetilde{j}}^0<0$).

Case I ($c_{\widetilde{j}}^0>0$): For any $j,\widetilde{j}=1,2,\ldots,G^0$ and $g_i^0=\widetilde{j}$, we have

$$\begin{aligned}
&\Pr\left(\widetilde{Z}_{ij}=1\right)\\
&\leq\sum_{\substack{\widetilde{j}\neq j,\\ c_{\widetilde{j}}^0>0}}\Pr\left(\left(\rho_{\widetilde{j}}^{0}-\rho_{j}^{0}\right)^{2}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^{2}+2\left(\rho_{\widetilde{j}}^{0}-\rho_{j}^{0}\right)\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\leq\frac{B_1}{T^{\gamma}}\eta\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right|+\frac{B_2}{T^{\gamma}}\eta\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^{2}\right|\right)\\
&\leq\sum_{\substack{\widetilde{j}\neq j,\\ c_{\widetilde{j}}^0>0}}\Pr\left(2\left(\rho_{\widetilde{j}}^{0}-\rho_{j}^{0}\right)\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\leq-\left(\rho_{\widetilde{j}}^{0}-\rho_{j}^{0}\right)^{2}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^{2}+\frac{B_1}{T^{\gamma}}\eta\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right|+\frac{B_2}{T^{\gamma}}\eta\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^{2}\right|\right)\\
&\leq\sum_{\substack{\widetilde{j}\neq j,\\ c_{\widetilde{j}}^0>0}}\Pr\left(2\frac{\left(c_{\widetilde{j}}^{0}-c_{j}^{0}\right)}{T^{\gamma}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\leq-\frac{\left(c_{\widetilde{j}}^{0}-c_{j}^{0}\right)^{2}\left(\rho_{\widetilde{j}}^{0}\right)^{2T}}{(\log T)^{2}}\overline{M}_1+\frac{(B_1)\left(\rho_{\widetilde{j}}^{0}\right)^{2T}}{(\log T)^{2}}\eta M_1+\frac{B_2T^{\gamma}\left(\rho_{\widetilde{j}}^{0}\right)^{2T}\eta}{(\log T)^{-2}}\widetilde{M}_1\right)\\
&+\sum_{\substack{\widetilde{j}\neq j,\\ c_{\widetilde{j}}^0>0}}\Pr\left(\frac{(\log T)^{2}}{\left(\overline{\rho}_{i}^{0}\right)^{2T}T^{2\gamma}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^{2}\leq\overline{M}_1\right)+\sum_{\substack{\widetilde{j}\neq j,\\ c_{\widetilde{j}}^0>0}}\Pr\left(\frac{(\log T)^{2}}{\left(\overline{\rho}_{i}^{0}\right)^{2T}T^{\gamma}}\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right|\geq M_1\right)\\
&+\sum_{\substack{\widetilde{j}\neq j,\\ c_{\widetilde{j}}^0>0}}\Pr\left(\frac{1}{\left(\overline{\rho}_{i}^{0}\right)^{2T}T^{2\gamma}(\log T)^{2}}\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^{2}\right|\geq\widetilde{M}_1\right)\\
&\leq\sum_{\substack{\widetilde{j}\neq j,\\ c_{\widetilde{j}}^0>0}}\Pr\left(2\frac{\dot{c}}{T^{\gamma}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\leq-\frac{(\dot{c})^{2}\left(\rho_{\widetilde{j}}^{0}\right)^{2T}}{(\log T)^{2}}\overline{M}_1+\frac{(B_1)}{(\log T)^{2}}\left(\rho_{\widetilde{j}}^{0}\right)^{2T}\eta M_1+B_2T^{\gamma}\left(\rho_{\widetilde{j}}^{0}\right)^{2T}(\log T)^{2}\eta\widetilde{M}_1\right)
\end{aligned}$$

$$+\sum_{\substack{\widetilde{j}\neq j,\\ c^0_{\widetilde{j}}>0}}\Pr\left(\frac{(\log T)^2}{\left(\rho^0_{\widetilde{j}}\right)^{2T}T^{2\gamma}}\sum_{t=1}^T\widetilde{y}^2_{i,t-1}\le\overline{M}_1\right)+\sum_{\substack{\widetilde{j}\neq j,\\ c^0_{\widetilde{j}}>0}}\Pr\left(\frac{(\log T)^2}{\left(\rho^0_{\widetilde{j}}\right)^{2T}T^{\gamma}}\left|\sum_{t=1}^T\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right|\ge M_1\right)$$

$$+\sum_{\substack{\widetilde{j}\neq j,\\ c^0_{\widetilde{j}}>0}}\Pr\left(\frac{1}{\left(\rho^0_{\widetilde{j}}\right)^{2T}T^{2\gamma}(\log T)^2}\left|\sum_{t=1}^T\widetilde{y}^2_{i,t-1}\right|\ge\widetilde{M}_1\right). \tag{40}$$

By Lemmas A.3, A.4 and A.5, the second, third, and fourth terms in (40) are all $o\left(\frac{1}{n}\right)$. We can bound η by $\frac{(\dot{c})^2\overline{M}_1}{2\left(B_2T^\gamma(\log T)^4\widetilde{M}_1+B_1M_1\right)}$. For instance, we can set $\eta=\frac{(\dot{c})^2\overline{M}_1}{4\left(B_2T^\gamma(\log T)^4\widetilde{M}_1+B_1M_1\right)}$. Thus, it follows that

$$\sum_{\substack{\widetilde{j}\neq j,\\ c^0_{\widetilde{j}}>0}}\Pr\left(2\frac{\dot{c}(\log T)^2}{T^\gamma}\sum_{t=1}^T\widetilde{y}_{i,t-1}\widetilde{u}_{it}\le-(\dot{c})^2\left(\rho^0_{\widetilde{j}}\right)^{2T}\overline{M}_1+(B_1)\left(\rho^0_{\widetilde{j}}\right)^{2T}\eta M_1+B_2T^\gamma\left(\rho^0_{\widetilde{j}}\right)^{2T}(\log T)^4\eta\widetilde{M}_1\right)$$

$$=\sum_{\substack{\widetilde{j}\neq j,\\ c^0_{\widetilde{j}}>0}}\Pr\left(2\frac{\dot{c}(\log T)^2}{T^\gamma\left(\rho^0_{\widetilde{j}}\right)^{2T}}\sum_{t=1}^T\widetilde{y}_{i,t-1}\widetilde{u}_{it}\le-(\dot{c})^2\overline{M}_1+B_1\eta M_1+B_2T^\gamma(\log T)^4\eta\widetilde{M}_1\right)$$

$$\le\sum_{\substack{\widetilde{j}\neq j,\\ c^0_{\widetilde{j}}>0}}\Pr\left(\frac{2\dot{c}(\log T)^2}{T^\gamma\left(\rho^0_{\widetilde{j}}\right)^{2T}}\sum_{t=1}^T\widetilde{y}_{i,t-1}\widetilde{u}_{it}\le\frac{-(\dot{c})^2\overline{M}_1}{2}\right)$$

$$\le\sum_{\substack{\widetilde{j}\neq j,\\ c^0_{\widetilde{j}}>0}}\Pr\left(\frac{(\log T)^2}{T^\gamma\left(\rho^0_{\widetilde{j}}\right)^{2T}}\left|\sum_{t=1}^T\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right|\ge\frac{\dot{c}\overline{M}_1}{4}\right)=o\left(\frac{1}{n}\right). \tag{41}$$

The last equality is due to Lemma A.3 and the fact that G^0 is a finite value.

Case II ($c^0_{\widetilde{j}}=0$): For any $j,\widetilde{j}=1,2,...,G^0$ and $g^0_i=\widetilde{j}$, we have

$$\Pr\left(\widetilde{Z}_{ij}=1\right)$$

$$\le\sum_{\substack{\widetilde{j}\neq j,\\ c^0_{\widetilde{j}}=0}}\Pr\left(\left(\rho^0_{\widetilde{j}}-\rho^0_j\right)^2\sum_{t=1}^T\widetilde{y}^2_{i,t-1}+2\left(\rho^0_{\widetilde{j}}-\rho^0_j\right)\sum_{t=1}^T\widetilde{y}_{i,t-1}\widetilde{u}_{it}\le\frac{B_1}{T^\gamma}\eta\left|\sum_{t=1}^T\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right|+\frac{B_2}{T^\gamma}\eta\left|\sum_{t=1}^T\widetilde{y}^2_{i,t-1}\right|\right)$$

$$\le\sum_{\substack{\widetilde{j}\neq j,\\ c^0_{\widetilde{j}}=0}}\Pr\left(2\left(c^0_{\widetilde{j}}-c^0_j\right)T^\gamma\sum_{t=1}^T\widetilde{y}_{i,t-1}\widetilde{u}_{it}\le-\left(c^0_{\widetilde{j}}-c^0_j\right)^2\sum_{t=1}^T\widetilde{y}^2_{i,t-1}+(B_1)T^\gamma\eta\left|\sum_{t=1}^T\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right|\right.$$

$$+B_2 T^{\gamma} \eta \left| \sum_{t=1}^{T} \widetilde{y}_{i,t-1}^2 \right| \Bigg)$$

$$\leq \sum_{\substack{\widetilde{j} \neq j, \\ c^0_{\widetilde{j}}=0}} \Pr\left(2\left(c^0_{\widetilde{j}} - c^0_j\right) T^{\gamma} \sum_{t=1}^{T} \widetilde{y}_{i,t-1} \widetilde{u}_{it} \leq -\left(c^0_{\widetilde{j}} - c^0_j\right)^2 \frac{T^2}{(\log_2 T)^2} \overline{M}_2 + (B_1) \frac{T^2}{(\log_2 T)^2} \eta M_2 \right.$$

$$\left. + B_2 T^{2+\gamma} (\log_2 T)^2 \eta \widetilde{M}_2 \right) + \sum_{\substack{\widetilde{j} \neq j, \\ c^0_{\widetilde{j}}=0}} \Pr\left(\frac{(\log_2 T)^2}{T^2} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}^2 \leq \overline{M}_2 \right)$$

$$+ \sum_{\substack{\widetilde{j} \neq j, \\ c^0_{\widetilde{j}}=0}} \Pr\left(\frac{(\log_2 T)^2}{T^{2-\gamma}} \left| \sum_{t=1}^{T} \widetilde{y}_{i,t-1} \widetilde{u}_{it} \right| \geq M_2 \right) + \sum_{\substack{\widetilde{j} \neq j, \\ c^0_{\widetilde{j}}=0}} \Pr\left(\frac{1}{T^2 (\log_2 T)^2} \left| \sum_{t=1}^{T} \widetilde{y}_{i,t-1}^2 \right| \geq \widetilde{M}_2 \right)$$

$$\leq \sum_{\substack{\widetilde{j} \neq j, \\ c^0_{\widetilde{j}}=0}} \Pr\left(2\dot{c} \frac{(\log_2 T)^2}{T^{2-\gamma}} \sum_{t=1}^{T} \widetilde{y}_{i,t-1} \widetilde{u}_{it} \leq -(\dot{c})^2 \overline{M}_2 + (B_1)\eta M_2 + B_2 T^{\gamma} (\log_2 T)^4 \eta \widetilde{M}_2 \right)$$

$$+ \sum_{\substack{\widetilde{j} \neq j, \\ c^0_{\widetilde{j}}=0}} \Pr\left(\frac{(\log_2 T)^2}{T^2} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}^2 \leq \overline{M}_2 \right) + \sum_{\substack{\widetilde{j} \neq j, \\ c^0_{\widetilde{j}}=0}} \Pr\left(\frac{(\log_2 T)^2}{T^{2-\gamma}} \left| \sum_{t=1}^{T} \widetilde{y}_{i,t-1} \widetilde{u}_{it} \right| \geq M_2 \right)$$

$$+ \sum_{\substack{\widetilde{j} \neq j, \\ c^0_{\widetilde{j}}=0}} \Pr\left(\frac{1}{T^2 (\log_2 T)^2} \left| \sum_{t=1}^{T} \widetilde{y}_{i,t-1}^2 \right| \geq \widetilde{M}_2 \right). \tag{42}$$

Similarly, we can bound η by setting $\eta = \frac{(\dot{c})^2 \overline{M}_2}{4\left(B_2 T^{\gamma} (\log_2 T)^4 \widetilde{M}_2 + B_1 M_2\right)}$. Therefore, we have

$$\sum_{\substack{\widetilde{j} \neq j, \\ c^0_{\widetilde{j}}=0}} \Pr\left(\frac{2\left(c^0_{\widetilde{j}} - c^0_j\right)}{T^{-\gamma}} \sum_{t=1}^{T} \widetilde{y}_{i,t-1} \widetilde{u}_{it} \leq \frac{-\left(c^0_{\widetilde{j}} - c^0_j\right)^2 T^2}{(\log_2 T)^2} \overline{M}_2 + \frac{(B_1) T^2}{(\log_2 T)^2} \eta M_2 + B_2 T^{2+\gamma} (\log_2 T)^2 \eta \widetilde{M}_2 \right)$$

$$\leq \sum_{\substack{\widetilde{j} \neq j, \\ c^0_{\widetilde{j}}=0}} \Pr\left(2\dot{c} \frac{(\log_2 T)^2}{T^{2-\gamma}} \sum_{t=1}^{T} \widetilde{y}_{i,t-1} \widetilde{u}_{it} \leq -(\dot{c})^2 \overline{M}_2 + (B_1) \eta M_2 + B_2 T^{\gamma} (\log_2 T)^4 \eta \widetilde{M}_2 \right)$$

$$\leq \sum_{\substack{\widetilde{j} \neq j, \\ c^0_{\widetilde{j}}=0}} \Pr\left(\frac{2\dot{c} (\log_2 T)^2}{T^{2-\gamma}} \sum_{t=1}^{T} \widetilde{y}_{i,t-1} \widetilde{u}_{it} \leq \frac{-(\dot{c})^2 \overline{M}_2}{2} \right)$$

$$\leq \sum_{\substack{\widetilde{j}\neq j,\\ c^0_{\widetilde{j}}=0}} \Pr\left(\frac{(\log_2 T)^2}{T^{2-\gamma}}\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right| \geq \frac{\dot{c}\overline{M}_2}{4}\right) = o\left(\frac{1}{n}\right). \tag{43}$$

The last equality is due to Lemma A.3.

Case III ($c^0_{\widetilde{j}} < 0$): For any $j, \widetilde{j} = 1, 2, ..., G^0$ and $g^0_i = \widetilde{j}$, we have

$$\Pr\left(\widetilde{Z}_{ij} = 1\right)$$

$$\leq \sum_{\substack{\widetilde{j}\neq j,\\ c^0_{\widetilde{j}}<0}} \Pr\left(\left(\rho^0_{\widetilde{j}} - \rho^0_j\right)^2 \sum_{t=1}^{T}\widetilde{y}^2_{i,t-1} + 2\left(\rho^0_{\widetilde{j}} - \rho^0_j\right)\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it} \leq \frac{B_1}{T^\gamma}\eta\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right| + \frac{B_2}{T^\gamma}\eta\left|\sum_{t=1}^{T}\widetilde{y}^2_{i,t-1}\right|\right)$$

$$\leq \sum_{\substack{\widetilde{j}\neq j,\\ c^0_{\widetilde{j}}<0}} \Pr\left(2\left(c^0_{\widetilde{j}} - c^0_j\right)T^\gamma \sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it} \leq -\left(c^0_{\widetilde{j}} - c^0_j\right)^2 \sum_{t=1}^{T}\widetilde{y}^2_{i,t-1} + (B_1)T^\gamma\eta\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right| \right.$$
$$\left. + B_2 T^\gamma \eta\left|\sum_{t=1}^{T}\widetilde{y}^2_{i,t-1}\right|\right)$$

$$\leq \sum_{\substack{\widetilde{j}\neq j,\\ c^0_{\widetilde{j}}<0}} \Pr\left(2\left(c^0_{\widetilde{j}} - c^0_j\right)T^\gamma \sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it} \leq -\left(c^0_{\widetilde{j}} - c^0_j\right)^2 T^{1+\gamma}\overline{M}_3 + (B_1)T^{1+\gamma}\eta M_3 + B_2 T^{1+2\gamma}\eta\widetilde{M}_3\right)$$

$$+ \sum_{\substack{\widetilde{j}\neq j,\\ c^0_{\widetilde{j}}<0}} \Pr\left(\frac{1}{T^{1+\gamma}}\sum_{t=1}^{T}\widetilde{y}^2_{i,t-1} \leq \overline{M}_3\right) + \sum_{\substack{\widetilde{j}\neq j,\\ c^0_{\widetilde{j}}<0}} \Pr\left(\frac{1}{T}\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right| \geq M_3\right)$$

$$+ \sum_{\substack{\widetilde{j}\neq j,\\ c^0_{\widetilde{j}}<0}} \Pr\left(\frac{1}{T^{1+\gamma}}\left|\sum_{t=1}^{T}\widetilde{y}^2_{i,t-1}\right| \geq \widetilde{M}_3\right)$$

$$\leq \sum_{\substack{\widetilde{j}\neq j,\\ c^0_{\widetilde{j}}<0}} \Pr\left(2\dot{c}\frac{1}{T}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it} \leq -(\dot{c})^2\overline{M}_3 + (B_1)\eta M_3 + B_2 T^\gamma \eta\widetilde{M}_3\right)$$

$$+ \sum_{\substack{\widetilde{j}\neq j,\\ c^0_{\widetilde{j}}<0}} \Pr\left(\frac{1}{T^{1+\gamma}}\sum_{t=1}^{T}\widetilde{y}^2_{i,t-1} \leq \overline{M}_3\right) + \sum_{\substack{\widetilde{j}\neq j,\\ c^0_{\widetilde{j}}<0}} \Pr\left(\frac{1}{T}\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right| \geq M_3\right)$$

$$+ \sum_{\substack{\widetilde{j}\neq j,\\ c^0_{\widetilde{j}}<0}} \Pr\left(\frac{1}{T^{1+\gamma}}\left|\sum_{t=1}^{T}\widetilde{y}^2_{i,t-1}\right| \geq \widetilde{M}_3\right).$$

We can bound η by $\frac{(\dot{c})^2\overline{M}_3}{2\left(B_2T^\gamma\widetilde{M}_3+B_1M_3\right)}$. Set $\eta=\frac{(\dot{c})^2\overline{M}_3}{4\left(B_2T^\gamma\widetilde{M}_3+B_1M_3\right)}$ and then

$$\begin{aligned}
&\sum_{\substack{\tilde{j}\neq j,\\ c^0_{\tilde{j}}<0}}\Pr\left(2\dot{c}\frac{1}{T}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\leq-(\dot{c})^2\overline{M}_3+(B_1)\eta M_3+B_2T^\gamma\eta\widetilde{M}_3\right)\\
&\leq\sum_{\substack{\tilde{j}\neq j,\\ c^0_{\tilde{j}}<0}}\Pr\left(2\frac{\dot{c}}{T}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\leq\frac{-(\dot{c})^2\overline{M}_3}{2}\right)\\
&\leq\sum_{\substack{\tilde{j}\neq j,\\ c^0_{\tilde{j}}<0}}\Pr\left(\frac{1}{T}\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right|\geq\frac{\dot{c}\overline{M}_3}{4}\right)=o\left(\frac{1}{n}\right).
\end{aligned}\tag{44}$$

The last equality is due to Lemma A.3.

Combining the results in (41), (43) and (44), we obtain $\Pr\left(\widetilde{Z}_{ij}=1\right)=o\left(\frac{1}{n}\right)$, which in turn implies that

$$\begin{aligned}
\mathbb{E}\left[\sup_{c\in\mathcal{N}_\eta}\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\left\{\widehat{g}_i(c)\neq g_i^0\right\}\right]&\leq\frac{1}{n}\sum_{j=1}^{G^0}\sum_{i=1}^{n}\mathbb{E}\widetilde{Z}_{ij}=\frac{1}{n}\sum_{j=1}^{G^0}\sum_{i=1}^{n}\Pr\left(\widetilde{Z}_{ij}=1\right)\\
&=G^0\left(G^0-1\right)o\left(\frac{1}{n}\right)=o\left(\frac{1}{n}\right),
\end{aligned}\tag{45}$$

where $\eta=O\left(\frac{1}{(\log T)^4T^\gamma}\right)$. Weak convergence holds due to the Markov inequality and (45). The proof of Lemma A.6 is then complete. ■

To establish uniform consistency of the recursive k-means clustering algorithm, we first define the following sequences of events

$$\widehat{E}_{j,i}:=\left\{\widehat{g}_i\neq j|g_i^0=j\right\}\text{ and }\widehat{F}_{j,i}:=\left\{g_i^0\neq j|\widehat{g}_i=j\right\},\tag{46}$$

for any $j=1,2,...,G^0$ and $i\in\mathcal{I}_n$. Let $\widehat{E}_{j,nT}:=\bigcup_{i\in\mathcal{G}^0(j)}\widehat{E}_{j,i}$ and $\widehat{F}_{j,nT}:=\bigcup_{i\in\hat{\mathcal{G}}(j)}\widehat{F}_{j,i}$. Uniform consistency of the clustering algorithm is shown in the following lemma.

Lemma A.7 *(Uniform Consistency of Clustering) Let Assumption 1 and 2 hold. When $(n,T)\to\infty$,*

(i) $\Pr\left(\bigcup_{j=1}^{G^0}\widehat{E}_{j,nT}\right)\leq\sum_{j=1}^{G^0}\Pr\left(\widehat{E}_{j,nT}\right)\to0$;

(ii) $\Pr\left(\bigcup_{j=1}^{G^0}\widehat{F}_{j,nT}\right)\leq\sum_{j=1}^{G^0}\Pr\left(\widehat{F}_{j,nT}\right)\to0$.

Proof of Lemma A.7: To establish the uniform consistency of the recursive k-means clustering algorithm, we bound the clustering error as

$$\Pr\left(\bigcup_{j=1}^{G^0}\widehat{E}_{j,nT}\right)\leq\sum_{j=1}^{G^0}\Pr\left(\widehat{E}_{j,nT}\right)\leq\sum_{j=1}^{G^0}\sum_{i\in\mathcal{G}^0(j)}\Pr\left(\widehat{E}_{j,i}\right).$$

It then follows that

$$\sum_{j=1}^{G^0}\sum_{i\in\mathcal{G}^0(j)}\Pr\left(\widehat{E}_{j,i}\right)\le n\max_{i\in\mathcal{I}_n}\mathbb{E}\mathbf{1}\{\widehat{g}_i(\widehat{c}^*)\neq g_i^0\}\le n\max_{i\in\mathcal{I}_n}\Pr\{\left|\widehat{g}_i(\widehat{c}^*)-g_i^0\right|>0\}$$
$$\le n\max_{i\in\mathcal{I}_n}\sup_{c\in\mathcal{N}_\eta}\Pr\{\left|\widehat{g}_i(c)-g_i^0\right|>0\}+n\max_{1\le j\le G^0}\Pr\{\left|\widehat{c}_j^*-c_j^0\right|>\eta\}$$
$$=o(1)+n\max_{1\le j\le G^0}\Pr\{\left|\widehat{c}_j^*-c_j^0\right|>\eta\}=o(1),\tag{47}$$

where the last step is due to the Markov inequality, equation (4) in Lemma A.1 and Assumption 2. The above derivations prove Lemma A.7(i). For Lemma A.7(ii), we can follow the proof of Theorem 2.2 (ii) in Su et al. (2016). This completes the proof. ■

B Proofs for Stage 2: Post-clustering Estimation and Testing

We need the following lemma that shows the consistency of the variance and covariance estimates $\check{\omega}_j^2$, $\check{\sigma}_j^2$ and $\check{\lambda}_j$, which are essential for inference to ensure that the test statistics are properly centred and scaled.

Lemma B.1 *Suppose Assumptions 1 and 2 hold. When $(n,T)\to\infty$,*

$$\check{\omega}_j^2\to_p\left(\omega_j^0\right)^2,\quad\widehat{\omega}_j^2\to_p\left(\omega_j^0\right)^2,$$
$$\check{\sigma}_j^2\to_p\left(\sigma_j^0\right)^2,\quad\widehat{\sigma}_j^2\to_p\left(\sigma_j^0\right)^2,$$
$$\check{\lambda}_j\to_p\lambda_j^0,\qquad\widehat{\lambda}_j\to_p\lambda_j^0,$$

for any $j=1,2,\ldots,G^0$ with $c_j^0\ge 0$.

Proof of Lemma B.1: Details are given in the proof of Lemma B.1 of the Online Supplement to Liu et al. (2022).

Lemma B.2 *Suppose Assumptions 1 and 2 hold. Then, for any $j\in\mathcal{G}^0$, when $(n,T)\to\infty$,*

$$\frac{1}{n_jT^2}\sum_{i\in\mathcal{G}^0(j)}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2\to_p\frac{\left(\omega_j^0\right)^2}{6},\ \textit{if}\ c_j^0=0;\tag{48}$$

$$\frac{1}{n_jT^{2\gamma}(\rho_j^0)^{2T}}\sum_{i\in\mathcal{G}^0(j)}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2\to_p\frac{1}{2c_j^0}\left(\frac{\left(\omega_j^0\right)^2}{2c_j^0}\right),\ \textit{if}\ c_j^0>0;$$

$$\frac{1}{n_jT^{1+\gamma}}\sum_{i\in\mathcal{G}^0(j)}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2\to_p\frac{\left(\omega_j^0\right)^2}{-2c_j^0},\ \textit{if}\ c_j^0<0.$$

Proof of Lemma B.2: For this proof, we assume $g_i^0 = j$ with $j = 1, 2, ..., G^0$. For the denominator of the panel within estimator $\widehat{\rho}_j$, three cases remain to be discussed: (i) the unit root group ($c_j^0 = 0$); (ii) the explosive groups ($c_j^0 > 0$); and (iii) the stationary groups ($c_j^0 < 0$).

First, when $c_j^0 = 0$,

$$\frac{1}{n_j T^2} \sum_{i \in \mathcal{G}^0(j)} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}^2 \overset{T\to\infty}{\Rightarrow} \frac{1}{n_j} \sum_{i \in \mathcal{G}^0(j)} \int_0^1 \widetilde{B}_i^2(r)\, dr \overset{n\to\infty}{\to_p} \mathbb{E} \int_0^1 \widetilde{B}_i^2(r)\, dr, \tag{49}$$

where $\widetilde{B}_i(\cdot) := B_i(\cdot) - \int_0^1 B_i(s)\, ds$ and $B_i(\cdot)$ is the limit Brownian motion associated with partial sums of the $u_{i,t}$. Standard calculations lead to

$$\begin{aligned} \mathbb{E} \int_0^1 \widetilde{B}_i^2(r)\, dr &= \int_0^1 \mathbb{E} B_i^2(r)\, dr - \int_0^1 \int_0^1 \mathbb{E}(B_i(r) B_i(s))\, dr ds \\ &= \left(\omega_j^0\right)^2 \int_0^1 r dr - \left(\omega_j^0\right)^2 \int_0^1 \int_0^1 (r \wedge s)\, dr ds = \frac{\left(\omega_j^0\right)^2}{6}. \end{aligned} \tag{50}$$

Combining (49) and (50), we have

$$\frac{1}{n_j T^2} \sum_{i \in \mathcal{G}^0(j)} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}^2 \to_p \frac{\left(\omega_j^0\right)^2}{6}, \tag{51}$$

under $(n, T)_{seq} \to \infty$. Due to the fact that $\lim_{n_j \to \infty} \frac{1}{n_j} \sum_{i \in \mathcal{G}^0(j)} \mathbb{E} \int_0^1 \widetilde{B}_i^2(r)\, dr = \frac{\left(\omega_j^0\right)^2}{6} < \infty$ and in view of the sequential limit (49), equation (51) also holds under the joint asymptotic scheme $(n, T) \to \infty$ (Phillips and Moon, 1999, Theorem 1).

Second, when $c_j^0 > 0$,

$$\begin{aligned} \frac{1}{n_j \left(\rho_j^0\right)^{2T} T^{2\gamma}} \sum_{i \in \mathcal{G}^0(j)} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}^2 &= \frac{1}{n_j \left(\rho_j^0\right)^{2T} T^{2\gamma}} \sum_{i \in \mathcal{G}^0(j)} \sum_{t=1}^{T} y_{i,t-1}^2 - \frac{1}{n_j \left(\rho_j^0\right)^{2T} T^{2\gamma}} \sum_{i \in \mathcal{G}^0(j)} \left(T \cdot \overline{y}_{i,-1}^2\right) \\ &= \frac{1}{n_j \left(\rho_j^0\right)^{2T} T^{2\gamma}} \sum_{i \in \mathcal{G}^0(j)} \sum_{t=1}^{T} y_{i,t-1}^2 + O_p\left(\frac{1}{T^{1-\gamma}}\right), \end{aligned}$$

since $\overline{y}_{i,-1} = \frac{1}{T} \sum_{s=1}^{T} y_{i,s-1} = O_p\left(\left(\rho_j^0\right)^T T^{\frac{3\gamma}{2}-1}\right)$. Therefore, it follows that

$$\frac{1}{n_j \left(\rho_j^0\right)^{2T} T^{2\gamma}} \sum_{i \in \mathcal{G}^0(j)} \sum_{t=1}^{T} y_{i,t-1}^2 \overset{T\to\infty}{\Rightarrow} \frac{1}{2c_j^0 n_j} \sum_{i \in \mathcal{G}^0(j)} \left(X_i^\star\right)^2 \overset{n\to\infty}{\to_p} \frac{1}{2c_j^0} \left(\frac{\left(\omega_j^0\right)^2}{2c_j^0}\right), \tag{52}$$

where $\left(Y_i^\star, X_i^\star\right)' \sim \mathcal{N}\left(\mathbf{0}_{2\times1}, diag\left(\frac{\left(\omega_j^0\right)^2}{2c_j^0}, \frac{\left(\omega_j^0\right)^2}{2c_j^0}\right)\right)$. Due to the fact that

$$\lim_{n_j\to\infty} \frac{1}{2c_j^0 n_j} \sum_{i \in \mathcal{G}^0(j)} \mathbb{E}\left(X_i^\star\right)^2 = \frac{1}{2c_j^0} \left(\frac{\left(\omega_j^0\right)^2}{2c_j^0}\right) < \infty,$$

and in view of the sequential limit in (52), the following joint convergence result holds:

$$\frac{1}{n_j\left(\rho_j^0\right)^{2T} T^{2\gamma}} \sum_{i\in\mathcal{G}^0(j)} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}^2 \to_p \frac{1}{2c_j^0}\left(\frac{\left(\omega_j^0\right)^2}{2c_j^0}\right),$$

under $(n,T)\to\infty$.

Finally, when $c_j^0 < 0$,

$$\begin{aligned}\frac{1}{n_jT^{1+\gamma}} \sum_{i\in\mathcal{G}^0(j)} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}^2 &= \frac{1}{n_jT^{1+\gamma}} \sum_{i\in\mathcal{G}^0(j)} \sum_{t=1}^{T} y_{i,t-1}^2 - \frac{1}{n_jT^{1+\gamma}} \sum_{i\in\mathcal{G}^0(j)} \left(T\cdot \overline{y}_{i,-1}^2\right)\\ &= \frac{1}{n_jT^{1+\gamma}} \sum_{i\in\mathcal{G}^0(j)} \sum_{t=1}^{T} y_{i,t-1}^2 + O_p\left(\frac{1}{T^{1-\gamma}}\right),\end{aligned}$$

since $\overline{y}_{i,-1} = \frac{1}{T}\sum_{s=1}^{T} y_{i,s-1} = O_p\left(T^{\gamma-\frac{1}{2}}\right)$. Therefore, under the sequential asymptotic scheme, the leading term of the denominator

$$\frac{1}{n_jT^{1+\gamma}} \sum_{i\in\mathcal{G}^0(j)} \sum_{t=1}^{T} y_{i,t-1}^2 \overset{T\to\infty}{\Rightarrow} \frac{1}{n_j}\sum_{i\in\mathcal{G}^0(j)}\left(\frac{\left(\omega_j^0\right)^2}{-2c_j^0}\right) \overset{n\to\infty}{\to_p} \frac{\left(\omega_j^0\right)^2}{-2c_j^0}. \tag{53}$$

Due to the fact that

$$\lim_{n_j\to\infty}\frac{1}{n_j}\sum_{i\in\mathcal{G}^0(j)}\mathbb{E}\left[\frac{\left(\omega_j^0\right)^2}{-2c_j^0}\right] = \frac{\left(\omega_j^0\right)^2}{-2c_j^0} < \infty,$$

and in view of the sequential limit (53), the following joint convergence result holds:

$$\frac{1}{n_jT^{1+\gamma}} \sum_{i\in\mathcal{G}^0(j)} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}^2 \to_p \frac{\left(\omega_j^0\right)^2}{-2c_j^0},$$

under $(n,T)\to\infty$. This completes the proof. ■

Lemma B.3 *Suppose Assumptions 1 and 2 hold. Then, for any $j\in\mathcal{G}^0$, when $(n,T)\to\infty$,*

$$\frac{1}{\sqrt{n_j}T} \sum_{i\in\mathcal{G}^0(j)} \sum_{t=1}^{T}\left(\widetilde{y}_{i,t-1}\widetilde{u}_{i,t} - \widehat{\lambda}_j + \frac{\widehat{\omega}_j^2}{2}\right) \Rightarrow \mathcal{N}\left(0, \frac{\left(\omega_j^0\right)^4}{12}\right), \text{ if } c_j^0 = 0; \tag{54}$$

$$\frac{1}{\sqrt{n_j}T^{\gamma}(\rho_j^0)^T} \sum_{i\in\mathcal{G}^0(j)} \sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it} \Rightarrow \mathcal{N}\left(0, \frac{\left(\omega_j^0\right)^4}{4\left(c_j^0\right)^2}\right), \text{ if } c_j^0 > 0; \tag{55}$$

$$\frac{1}{\sqrt{n_j}T^{\frac{1+\gamma}{2}}}\sum_{i\in\mathcal{G}^0(j)}\sum_{t=1}^{T}\left(\widetilde{y}_{i,t-1}\widetilde{u}_{it}-\lambda_j^0-\overline{m}_{j,T}\frac{c_j^0}{T^\gamma}\right)\Rightarrow\mathcal{N}\left(0,\frac{\left(\omega_j^0\right)^4}{-2c_j^0}\right),\ \textit{if}\ c_j^0<0,$$

where

$$\overline{m}_{j,T}=\frac{1}{n_j}\sum_{i\in\mathcal{G}^0(j)}m_{i,T},\ \textit{and}\ m_{i,T}=\sum_{h=1}^{\infty}\left(\rho_j^0\right)^{h-1}\mathbb{E}\left(\widetilde{\epsilon}_{it}u_{i,t-h}\right).$$

Proof of Lemma B.3: For this proof, we assume that $g_i^0=j$ with $j=1,2,...,G^0$. First we discuss the unit root case ($c_j^0=0$). When $T\to\infty$ and n is fixed, we have

$$\begin{aligned}
&\frac{1}{\sqrt{n_j}T}\sum_{i\in\mathcal{G}^0(j)}\sum_{t=1}^{T}\left(\widetilde{y}_{i,t-1}\widetilde{u}_{i,t}-\widehat{\lambda}_j+\frac{\widehat{\omega}_j^2}{2}\right)\\
&=\frac{1}{\sqrt{n_j}}\sum_{i\in\mathcal{G}^0(j)}\left[\frac{1}{T}\sum_{t=1}^{T}\left(y_{i,t-1}u_{i,t}-\widehat{\lambda}_j\right)-\left(\overline{y}_{i,-1}\overline{u}_i-\frac{\widehat{\omega}_j^2}{2}\right)\right]\\
&\overset{T\to\infty}{\Rightarrow}\frac{1}{\sqrt{n_j}}\sum_{i\in\mathcal{G}^0(j)}\left[\int_0^1 B_i(r)\,dB_i(r)-\left(\int_0^1 B_i(r)\,dr\right)B_i(1)+\frac{\left(\omega_j^0\right)^2}{2}\right]\\
&=:\frac{1}{\sqrt{n_j}}\sum_{i\in\mathcal{G}^0(j)}\left(N_{1,i}-N_{2,i}\right),
\end{aligned}\tag{56}$$

in which

$$N_{1i}:=\int_0^1 B_i(r)\,dB_i(r),\ \ N_{2i}:=\left(\int_0^1 B_i(r)\,dr\right)B_i(1)-\frac{\left(\omega_j^0\right)^2}{2}.$$

Note that $\mathbb{E}N_{1,i}=\mathbb{E}N_{2,i}=0$. When $T\to\infty$ followed by $n\to\infty$, we have

$$\begin{aligned}
\frac{1}{\sqrt{n_j}}\sum_{i\in\mathcal{G}^0(j)}\left(N_{1,i}-N_{2,i}\right)&\overset{n\to\infty}{\Rightarrow}\mathcal{N}\left(0,\ \mathbb{E}N_{1,i}^2+\mathbb{E}N_{2,i}^2-2\mathbb{E}N_{1,i}N_{2,i}\right)\\
&=_d\mathcal{N}\left(0,\ \frac{\left(\omega_j^0\right)^4}{12}\right),
\end{aligned}\tag{57}$$

where (57) holds in view of the following facts:

(a) $\mathbb{E}N_{1,i}^2=\frac{\left(\omega_j^0\right)^4}{2}$;

(b) $\mathbb{E}N_{2,i}^2=\frac{7\left(\omega_j^0\right)^4}{12}$;

(c) $\mathbb{E}\left(N_{1,i}N_{2,i}\right)=\frac{\left(\omega_j^0\right)^4}{2}$.

Based on (a) (b) and (c), whose calculations are given in the Online Supplement to Liu et al. (2022), the validity of (54) is confirmed under sequential asymptotics $(n,T)_{seq} \to \infty$. Checking the sufficient conditions then assures joint convergence limit theory using the results in Phillips and Moon (1999), whose full derivations are given in the same reference. Joint weak convergence (54) then follows irrespective of the divergence rates of $n, T \to \infty$.

Second, for the group of explosive roots ($c_j^0 > 0$),

$$\begin{aligned}
&\frac{1}{\sqrt{n_j}T^{\gamma}(\rho_j^0)^T}\sum_{i\in\mathcal{G}^0(j)}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it} \\
&\sim_a \frac{1}{\sqrt{n_j}T^{\gamma}(\rho_j^0)^T}\sum_{i\in\mathcal{G}^0(j)}\sum_{t=1}^{T}y_{i,t-1}u_{it} && (58)\\
&\overset{T\to\infty}{\Rightarrow} \frac{1}{\sqrt{n_j}}\sum_{i\in\mathcal{G}^0(j)}\left(Y_i^{\star}X_i^{\star}\right) && (59)\\
&\overset{n\to\infty}{\Rightarrow} \mathcal{N}\left(0, \frac{\left(\omega_j^0\right)^2}{2c_j^0}\left(\frac{\left(\omega_j^0\right)^2}{2c_j^0}\right)\right),
\end{aligned}$$

where (58) holds from the fact that $\overline{y}_{i,t-1} = O_p\left(T^{\frac{3\gamma}{2}-1}\left(\rho_j^0\right)^T\right)$ and $T^{1-\gamma} > n$ by virtue of Assumption 2, and (59) holds since $\left(Y_i^{\star}, X_i^{\star}\right)' \sim \mathcal{N}\left(\mathbf{0}_{2\times1}, diag\left\{\frac{\left(\omega_j^0\right)^2}{2c_j^0}, \frac{\left(\omega_j^0\right)^2}{2c_j^0}\right\}\right)$. Joint convergence is verified by checking the sufficient conditions of Phillips and Moon (1999) and the details follow those of the unit root case. So when $(n,T) \to \infty$, we have the joint convergence

$$\frac{1}{\sqrt{n_j}T^{\gamma}(\rho_j^0)^T}\sum_{i\in\mathcal{G}^0(j)}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it} \Rightarrow \mathcal{N}\left(0, \frac{\left(\omega_j^0\right)^2}{2c_j^0}\left(\frac{\left(\omega_j^0\right)^2}{2c_j^0}\right)\right).$$

Last, for the group of stationary roots ($c_j^0 < 0$),

$$\begin{aligned}
&\frac{1}{\sqrt{n_j}T^{\frac{1+\gamma}{2}}}\sum_{i\in\mathcal{G}^0(j)}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it} \\
&= \frac{1}{\sqrt{n_j}T^{\frac{1+\gamma}{2}}}\sum_{i\in\mathcal{G}^0(j)}\sum_{t=1}^{T}y_{i,t-1}u_{it} - \frac{1}{\sqrt{n_j}T^{\frac{\gamma-1}{2}}}\sum_{i\in\mathcal{G}^0(j)}\overline{y}_{i,-1}\overline{u}_i \\
&= \frac{1}{\sqrt{n_j}T^{\frac{1+\gamma}{2}}}\sum_{i\in\mathcal{G}^0(j)}\sum_{t=1}^{T}y_{i,t-1}\left(F_i(1)\epsilon_{it} + \widetilde{\epsilon}_{i,t-1} - \widetilde{\epsilon}_{it}\right) + o_p(1) \\
&= \frac{1}{\sqrt{n_j}T^{\frac{1+\gamma}{2}}}\sum_{i\in\mathcal{G}^0(j)}F_i(1)\sum_{t=1}^{T}y_{i,t-1}\epsilon_{it} - \frac{1}{\sqrt{n_j}T^{\frac{1+\gamma}{2}}}\sum_{i\in\mathcal{G}^0(j)}y_{iT}\widetilde{\epsilon}_{iT}
\end{aligned}$$

$$+\frac{1}{\sqrt{n_j}T^{\frac{1+\gamma}{2}}}\sum_{i\in\mathcal{G}^0(j)}\sum_{t=1}^{T}\left\{\frac{c_j^0}{T^\gamma}y_{i,t-1}+u_{it}+\mu_i\right\}\widetilde{\epsilon}_{it}+o_p(1)$$

$$=\frac{1}{\sqrt{n_j}T^{\frac{1+\gamma}{2}}}\sum_{i\in\mathcal{G}^0(j)}F_i(1)\sum_{t=1}^{T}y_{i,t-1}\epsilon_{it}-\frac{1}{\sqrt{n_j}T^{\frac{1+\gamma}{2}}}\sum_{i\in\mathcal{G}^0(j)}y_{iT}\widetilde{\epsilon}_{iT}$$

$$+\frac{1}{\sqrt{n_j}T^{\frac{1+3\gamma}{2}}}\sum_{i\in\mathcal{G}^0(j)}c_j^0\sum_{t=1}^{T}y_{i,t-1}\widetilde{\epsilon}_{it}+\frac{1}{\sqrt{n_j}T^{\frac{1+\gamma}{2}}}\sum_{i\in\mathcal{G}^0(j)}\sum_{t=1}^{T}u_{it}\widetilde{\epsilon}_{it}+o_p(1),$$

where $\frac{1}{\sqrt{n_j}T^{\frac{1+\gamma}{2}}}\left|\sum_{i\in\mathcal{G}^0(j)}y_{iT}\widetilde{\epsilon}_{iT}\right|\leq\frac{\sqrt{n_j}}{T^{\frac{1+\gamma}{2}}}\sup_{i\in\mathcal{G}^0(j)}\left|y_{iT}\right|\sup_{i\in\mathcal{G}^0(j)}|\widetilde{\epsilon}_{iT}|=O_p\left(\sqrt{\frac{n}{T}}\right)$. Since $T^{1-\gamma}>n$ from Assumption 2, and $\sum_{t=1}^{T}y_{i,t-1}=O_p\left(T^{\gamma+\frac{1}{2}}\right)$ and $\sum_{t=1}^{T}u_{it}=O_p\left(\sqrt{T}\right)$, we have

$$\frac{1}{\sqrt{n_j}T^{\frac{\gamma-1}{2}}}\sum_{i\in\mathcal{G}^0(j)}\overline{y}_{i,-1}\overline{u}_i=o_p(1).$$

By Phillips and Magdalinos (2007),

$$\frac{1}{\sqrt{n_j}T^{\frac{1+\gamma}{2}}}\left|\sum_{i\in\mathcal{G}^0(j)}\sum_{t=1}^{T}\left(u_{it}\widetilde{\epsilon}_{it}-\overline{\lambda}_i^0\right)\right|=o_p(1),$$

$$\frac{1}{\sqrt{n_j}T^{\frac{1+3\gamma}{2}}}\left|\sum_{i\in\mathcal{G}^0(j)}c_j^0\sum_{t=1}^{T}\left(y_{i,t-1}\widetilde{\epsilon}_{it}-m_{iT}\right)\right|=o_p(1).$$

Thus, under the sequential limit $(n,T)_{seq}\to\infty$, we have

$$\frac{1}{\sqrt{n_j}T^{\frac{1+\gamma}{2}}}\sum_{i\in\mathcal{G}^0(j)}\sum_{t=1}^{T}\left(y_{i,t-1}u_{it}-\lambda_j^0-\overline{m}_{j,T}\frac{c_j^0}{T^\gamma}\right)$$

$$=\frac{1}{\sqrt{n_j}T^{\frac{1+\gamma}{2}}}\sum_{i\in\mathcal{G}^0(j)}F_i(1)\sum_{t=1}^{T}y_{i,t-1}\epsilon_{it}+\frac{1}{\sqrt{n_j}T^{\frac{1+3\gamma}{2}}}\sum_{i\in\mathcal{G}^0(j)}c_j^0\sum_{t=1}^{T}\left(y_{i,t-1}\widetilde{\epsilon}_{it}-\overline{m}_{j,T}\right)$$

$$+\frac{1}{\sqrt{n_j}T^{\frac{1+\gamma}{2}}}\sum_{i\in\mathcal{G}^0(j)}\sum_{t=1}^{T}\left(u_{it}\widetilde{\epsilon}_{it}-\lambda_j^0\right)+o_p(1)$$

$$=\frac{1}{\sqrt{n_j}T^{\frac{1+\gamma}{2}}}\sum_{i\in\mathcal{G}^0(j)}F_i(1)\sum_{t=1}^{T}y_{i,t-1}\epsilon_{it}+o_p(1)$$

$$\overset{T\to\infty}{\Rightarrow}\frac{1}{\sqrt{n_j}}\sum_{i\in\mathcal{G}^0(j)}\mathcal{N}\left(0,\frac{\left(\omega_j^0\right)^4}{-2c_j^0}\right)\overset{n\to\infty}{\Rightarrow}\mathcal{N}\left(0,\frac{\left(\omega_j^0\right)^4}{-2c_j^0}\right).$$

Joint convergence is verified by following the same procedure as the unit root case. Joint convergence of the numerator of the stationary groups is then established and the proof is complete. ■

C Proofs for the Estimation of the Group Numbers

Lemma C.1 *Suppose Assumptions 1 and 2 hold. Let* $(n,T)\to\infty$*. When (i)* $\gamma>0$ *and* $\bar{c}_i^0\geq 0$ *or (ii)* $\gamma=0$*, we have*

$$\min_{1\leq G<G^0}\inf_{\widehat{\delta}(G)\in\Delta_G}\check{\sigma}^2_{\widehat{\delta}(G)}>\sigma_0^2,\ \textit{with probability approaching 1,}\tag{60}$$

and

$$\check{\sigma}^2_{\widehat{\delta}(G^0)}\rightarrow_p\sigma_0^2,\tag{61}$$

where σ_0^2 *is defined in equation (29) of the main paper,*

$$\check{\sigma}^2_{\widehat{\delta}(G)}:=\frac{1}{nT}\sum_{j=1}^{G}\sum_{i\in\hat{\mathcal{G}}(j,G)}\sum_{t=1}^{T}\left(\widetilde{y}_{it}-\widetilde{y}_{i,t-1}\check{\rho}_j(G)\right)^2,$$

and $\widehat{\delta}(G)\left(:=\left(\widehat{g}_1^{(G)},\widehat{g}_2^{(G)},\ldots,\widehat{g}_n^{(G)}\right)\right)$ *is the vectorized membership estimate, assuming there are G groups*[1].

Proof of Lemma C.1: To show equation (60) it is sufficient to show

$$\inf_{\widehat{\delta}(G)\in\Delta_G}\check{\sigma}^2_{\widehat{\delta}(G)}>\sigma_0^2,\ \text{with probability approaching 1,}\tag{62}$$

for all $G<G^0$. The above relationship (62) holds since the true number of groups, G^0, is finite. Without losing generality, we discuss the case in which $G=G^0-d$ and $d=1$. Treatment for the case $d\geq 2$ is similar to the case $d=1$ and is omitted.

When the minimum value of $\check{\sigma}^2_{\widehat{\delta}(G)}$ is attained, the individuals of G groups are correctly clustered to their true membership, and the individuals of the remaining group, namely the individuals for the j^*-th true group, are wrongly allocated to the group whose group-specific distance parameter $c^0_{\widetilde{j}}$ is close to the group-specific parameter of these individuals, $c^0_{j^*}$. Without losing generality we assume $\widetilde{j}\leq G$ and still call the union of the $\widetilde{j}$-th and j^*-th true groups as the $\widetilde{j}$-th estimated group of the G-group partition.

Therefore, $\left\{\hat{\mathcal{G}}(j,G)\right\}_{1\leq j\leq G}$ and $\left\{\hat{\mathcal{G}}\left(j,G^0\right)\right\}_{1\leq j\leq G^0}$ share $(G-1)$ common groups. Apart from these $(G-1)$ common groups, the remaining group in $\left\{\hat{\mathcal{G}}(j,G)\right\}_{1\leq j\leq G}$ is the union of the two remaining groups in $\left\{\hat{\mathcal{G}}\left(j,G^0\right)\right\}_{1\leq j\leq G^0}$. By the consistency of the post-clustering estimates, we have

$$\begin{aligned}&\check{\sigma}^2_{\widehat{\delta}(G)}-\check{\sigma}^2_{\widehat{\delta}(G^0)}\\&=\frac{1}{nT}\sum_{j=1}^{G}\sum_{i\in\hat{\mathcal{G}}(j,G)}\sum_{t=1}^{T}\left(\widetilde{y}_{it}-\widetilde{y}_{i,t-1}\check{\rho}_j(G)\right)^2-\frac{1}{nT}\sum_{j=1}^{G^0}\sum_{i\in\hat{\mathcal{G}}(j,G^0)}\sum_{t=1}^{T}\left(\widetilde{y}_{it}-\widetilde{y}_{i,t-1}\check{\rho}_j\left(G^0\right)\right)^2\end{aligned}$$

[1] $\check{\rho}_j(G)$ can follow the definition given either in (12) or in (21) of the main paper.

$$
\begin{aligned}
&\sim_a -\frac{2}{nT}\sum_{j=\widetilde{j},j^*}\sum_{i\in\hat{\mathcal{G}}(j,G^0)}\sum_{t=1}^{T}\left(\widetilde{y}_{it}-\widetilde{y}_{i,t-1}\check{\rho}_j\left(G^0\right)\right)\widetilde{y}_{i,t-1}\left(\check{\rho}_{\widetilde{j}}(G)-\check{\rho}_j\left(G^0\right)\right)\\
&+\frac{1}{nT}\sum_{j=\widetilde{j},j^*}\sum_{i\in\hat{\mathcal{G}}(j,G^0)}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2\left(\check{\rho}_{\widetilde{j}}(G)-\check{\rho}_j\left(G^0\right)\right)^2\\
&\sim_a \frac{1}{nT}\sum_{j=\widetilde{j},j^*}\sum_{i\in\hat{\mathcal{G}}(j,G^0)}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2\left(\check{\rho}_{\widetilde{j}}(G)-\check{\rho}_j\left(G^0\right)\right)^2.
\end{aligned}
$$

where the asymptotic equivalence holds by the consistency of the post-clustering estimates $\check{\rho}_j\left(G^0\right)$ and $\check{\rho}_j\left(G\right)$. When (i) $\overline{c}_i^0 \geq 0$ and $\gamma\in(0,1)$ for all $i\in\mathcal{I}_n$

$$
\frac{1}{nT}\sum_{j=\widetilde{j},j^*}\sum_{i\in\hat{\mathcal{G}}(j,G^0)}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2\left(\check{\rho}_{\widetilde{j}}(G)-\check{\rho}_j\left(G^0\right)\right)^2 \sim_a \begin{cases} O_p\left(T^{1-2\gamma}\right) & \text{if } \ c_{j^*}^0=0 \\ O_p\left(\left(\rho_{j^*}^0\right)^{2T}T^{-1}\right) & \text{if } \ c_{j^*}^0>0 \end{cases};
$$

or when (ii) $\gamma=0$,

$$
\frac{1}{nT}\sum_{j=\widetilde{j},j^*}\sum_{i\in\hat{\mathcal{G}}(j,G^0)}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2\left(\check{\rho}_{\widetilde{j}}(G)-\check{\rho}_j\left(G^0\right)\right)^2 \sim_a \begin{cases} Q>0 & \text{if } \ c_{j^*}^0<0 \\ O_p\left(T\right) & \text{if } \ c_{j^*}^0=0 \\ O_p\left(\left(\rho_{j^*}^0\right)^{2T}T^{-1}\right) & \text{if } \ c_{j^*}^0>0 \end{cases},
$$

in which Q is a positive constant. Then equation (60) holds. Moreover, equation (61) holds due to the consistency of k-means clustering and post-clustering estimates:

$$
\begin{aligned}
\check{\sigma}_{\widehat{\delta}(G^0)}^2 &:= \frac{1}{nT}\sum_{j=1}^{G^0}\sum_{i\in\hat{\mathcal{G}}(j,G^0)}\sum_{t=1}^{T}\left(\widetilde{y}_{it}-\widetilde{y}_{i,t-1}\check{\rho}_j\left(G^0\right)\right)^2=\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\widetilde{\widetilde{u}}_{it}^2\\
&\sim_a \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\widetilde{u}_{it}^2=\frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}u_{it}^2+O_p\left(\frac{1}{T}\right)\\
&\to_p \sigma_0^2, \qquad (63)
\end{aligned}
$$

where the last line (63) holds due to equation (29) of the main paper. The proof is now complete. ■

Proof of Theorem 4.4: Details are given in the Proof of Theorem 4.4 of the Online Supplement to Liu et al. (2022).

D Simulation Studies

We designed several numerical experiments to check the finite sample performance of the procedures developed above. These include: the group number estimate in (26) of the main paper; the membership estimate generated by the recursive k-means clustering algorithm in (8) of the main paper; the post-clustering estimates in (12) of the main paper; and the size and power performances of the proposed tests in (17) of the main paper.

The following model setup was used to generate the simulated data: the individual fixed effects $\mu_i \overset{i.i.d.}{\sim} T^{-1}\mathcal{N}(0,0.1)$; the error process $u_{it} = \theta u_{i,t-1} + \epsilon_{it}$ with (i) serially correlated errors ($\theta = 0.5$, $\epsilon_{it} \overset{i.i.d.}{\sim} \mathcal{N}(0,0.01)$) or (ii) martingale differences ($\theta = 0$, $\epsilon_{it} \overset{i.i.d.}{\sim} \mathcal{N}(0,0.01)$); sample sizes $n = 30, 60, 90, 120, 150$, and $T = 100, 150, 200, 250, 350, 450, 550$; and group number $G^0 = 3$ (i.e., three groups) with $\pi_1 : \pi_2 : \pi_3 = \frac{1}{3} : \frac{1}{3} : \frac{1}{3}$ or $G^0 = 2$ (i.e., two groups) with $\pi_1 : \pi_2 = \frac{1}{2} : \frac{1}{2}$. The following parameter settings for c and γ were considered:

$$\left(c_1^0, c_2^0, c_3^0, \gamma\right) = \begin{cases} (-15,-8,-1,0.6) & \text{for} \quad \text{DGP 0} \\ (1,\ 0,-6,0.6) & \text{for} \quad \text{DGP 1}\,, \\ (1,0.2,-6,0.6) & \text{for} \quad \text{DGP 2} \end{cases} \tag{64}$$

and

$$\left(c_1^0, c_2^0, \gamma\right) = (-1,1,0.6) \quad \text{for} \quad \text{DGP 3}. \tag{65}$$

In all DGPs we set the distance parameters c, rate scaling parameter γ, group division δ, and error process u_{it} to approximate the fitted parameters in the empirical results of the housing and equity markets in Section 6 of the main paper. The explosive root signals are often weak in the empirical data, so group-specific parameters in the explosive groups are set with smaller values (i.e., $c_j^0 = 0.2$ or 1); and near-unit root behavior is present in all markets, so the unit root group is considered in most DGPs. Since model misspecifications can lead to invalid clustering and inference, we investigate cases where there are either three groups (G^0=3) or two groups ($G^0 = 2$). Specifically, DGP 0 is designed to reveal the downward bias of IC in the panel clustering model of all mildly stationary groups. DGPs 1–3 are designed to assess the accuracy of the hybrid model specification procedure, the consistency of the recursive k-means clustering algorithm, and the power improvement of the panel inference procedures.

Figure 1 plots both the empirical density and the sample average of the signal-to-noise ratio (SNR) for each group in DGPs 1–3. The SNR is measured by the ratio of the sample variance of $(1 + \frac{\check{c}_j}{T^\gamma})y_{i,t-1}$ to the sample variance of $\widehat{u}_{it}$ with sample sizes $n = 60$, $T = 100$ and $\widehat{g}_i = j$ obtained from the simulated sample paths. For all DGPs the SNR of the mildly explosive group is far greater than the SNR of the unit root group, which in turn is much larger than the SNR of the mildly stationary group. These results, which are also evident in Table 1, are to be expected in view of the different divergence rates ($O_p\left(T^{2\gamma-1}\left(\rho_j^0\right)^{2T}\right)$, $O_p(T)$ and $O_p(T^\gamma)$, respectively) of the SNRs in these groups. This heterogeneity in rates across groups facilitates recovery of the latent group structure by the recursive k-means clustering algorithm and the overall performance in selection is strengthened in the panel regressions because of cross section averaging within groups.

[Insert Figure 1 Here]

To explore the advantages of the post-clustering panel tests, we make comparisons with the behavior of the usual semiparametric time series test statistics (Phillips, 1987; Phillips and Perron, 1988):

$$\text{PP } t\text{-test} = \frac{\left(\widehat{\overline{\rho}}_i^{TS} - 1 - \frac{T\cdot\widehat{\overline{\lambda}}_i^{TS}}{D_{i,T}}\right)\sqrt{D_{i,T}}}{\widehat{\overline{\omega}}_i^{TS}},\ \ \text{PP } J\text{-test} = T\left(\widehat{\overline{\rho}}_i^{TS} - 1 - \frac{T\cdot\widehat{\overline{\lambda}}_i^{TS}}{D_{i,T}}\right), \tag{66}$$

in which $\widehat{\overline{\rho}}_i^{TS}$ is the time series estimate of $\overline{\rho}_i^0$ defined in (3) of the main paper, the long run covariance estimate $\widehat{\overline{\lambda}}_i^{TS}$ and long run variance estimate $\left(\widehat{\overline{\omega}}_i^{TS}\right)^2$ are based on $\widehat{\overline{\rho}}_i^{TS}$, and the sample moment $D_{i,T} := \sum_{t=1}^{T} \widetilde{y}_{i,t-1}^2$. Under the null hypothesis $\mathcal{H}_0 : \overline{c}_i^0 = 0$, it follows from standard theory that

$$\text{PP } t\text{-test} \Rightarrow \frac{\int_0^1 \underset{\sim}{W}_i(r)dW_i(r)}{\left[\int_0^1 \underset{\sim}{W}(r)^2 dr\right]^{\frac{1}{2}}},\ \ \text{PP } J\text{-test} \Rightarrow \frac{\int_0^1 \underset{\sim}{W}_i(r)dW_i(r)}{\int_0^1 \underset{\sim}{W}(r)^2 dr}, \tag{67}$$

where the $W_i(\cdot)$ are standard Brownian motions and $\underset{\sim}{W}_i(r) = W_i(r) - \int_0^1 W_i(s)ds$.

According to the pivotal distributions of the panel t- and J-tests under the null hypothesis, the right-tailed 95% critical value is 1.64. For the PP t- and J-tests, the right-tailed 95% critical values are set at −0.07 and −0.13, respectively (e.g., Tables B.5-B.6 in Hamilton (1994)). Bandwidths are selected based on simulated performance in the mixed-root panel model[2]. The bandwidth for the long run variance estimates in (14) and (15) of the main paper is set at $L = \lfloor T^{0.3} \rfloor$ and the bandwidth for the variance estimate in (16) of the main paper was set at $L = \lfloor T^{0.1} \rfloor$. In addition, the bandwidth for the long run variance and covariance components of the time series statistics in (66) is set as $\lfloor T^{0.3} \rfloor$. These bandwidth choices are all consistent with the rate restrictions in the theory development. The number of replications was 1,000 in all experiments.

i Correlated errors ($\theta = 0.5$)

The performance of the group number estimate $\widehat{G}$ is first considered. The penalty κ_{nT} of IC is $(nT)^{-0.35}$ and the upper bound $G_{\max}$ is 5. The critical value of the Hausman test is set as $cv_{nT} = (1 + 5\log(nT))\chi^2\left(\overline{G}\right)$ and $\overline{G} = (G_{\max} - G + 1)$. Tables 2–5 report the empirical frequency of $\widehat{G}$ in (26) of the main paper. As T increases, the performance of the estimator $\widehat{G}$ steadily improves, so that when T is larger than 350 $\widehat{G}$ successfully identifies the true G^0 with only small errors involving overestimation, revealing evidence of its consistency in estimating the true number of groups. By comparison the downward bias of IC is evident in nearly every case, corroborating the asymptotic theory.

[Insert Tables 2–5 Here]

[2]In future research, cross-validation (CV) methods could also be employed, as in Phillips et al. (2017).

Next, we check the performance of the recursive k-means clustering algorithm and the post-clustering estimate while assuming the true group number G^0 is known. Tables 6–8 report the clustering error (CE), root mean squared error (RMSE), and bias of the post-clustering estimates. The CE is defined as

$$\frac{1}{n}\sum_{j=1}^{G^0}\sum_{i\in\hat{\mathcal{G}}(j)}\mathbf{1}\left\{\widehat{g_i}\neq g_i^0\right\}.$$

The RMSE is the square root of the sample moment of the squared differences between the post-clustering estimates and the true values. The bias is the averaged differences between the post-clustering estimates and the true values. For comparison we also report the CE, RMSE, and bias of the oracle estimates where it is assumed that the true group membership δ^0 is known.

[Insert Tables 6–8 Here]

According to Tables 6–8, the CE decreases to zero as T increases. The RMSE and bias of the oracle estimates are smaller than those of the post-clustering estimates. For the post-clustering estimates of the nonstationary groups, the magnitude of the RMSE and bias also generally decreases when $T \to \infty$. For all DGPs, the difference between the oracle and the post-clustering estimates is negligible when $T \geq 150$. The diminishing differences suggest asymptotic equivalence between these two sets of estimates. This property is due to the uniform consistency of the recursive k-means clustering algorithm, as shown in the theory development.

Based on the estimated membership $\widehat{\delta}$, the performance of the post-clustering panel t and J tests for detecting explosive roots is analyzed and compared with the time series counterparts. The nominal levels are all set at 5%, accompanied by the right-tail 95% critical values of the standard normal distribution and standard unit root limit distributions. We obtain the empirical rejection rates of the PP t and J tests when $n = 1$ and the empirical rejection rates of the post-clustering panel t and J tests when $n > 1$, as presented in Table 9. If the distancing parameter c_j^0 is zero, as in the null hypothesis, the empirical rejection rate is the empirical size. If the relevant c_j^0 is nonzero, the empirical rejection rate is defined as empirical power.

[Insert Table 9 Here]

Evidently the size distortion of both panel tests is small when $n \geq 60$ and $T \geq 150$, although size distortion of the panel tests is slightly larger than that of the time series counterparts. This is unsurprising as the asymptotics require the use of cross section central limit theory, which inevitably introduces approximation errors in finite samples, particularly small samples that arise in group subsamples. This loss is counterbalanced by a substantial improvement in the power of the panel tests over the time series tests. For instance, when $c_j^0 = 0.2$ (the corresponding ρ_j^0 is 1.0126, 1.0099 and 1.0083 when $T =$

$100, 150, 200$, which are empirically plausible values based on our empirical work), the power performances of the post-clustering panel tests are much larger than those of the time series tests. If $T = 100$, the post-clustering panel t-test raises the power of the time series t-test from 0.175 to 0.599 when $n_j = 10$, to 0.807 when $n_j = 20$, and to 0.917 when $n_j = 30$. The post-clustering panel t test with $T = 100$ has substantially greater power than the time series t test with $T = 200$ (0.917 versus 0.382). Moreover, it is interesting to note that the panel t test has greater power than the panel J-test that is based on the estimated membership, corroborating the different divergence rates in asymptotic theory of Theorem 4.3 of the main paper under the mildly explosive alternative.

ii Uncorrelated errors ($\theta = 0$)

First, the performance of the group number estimate $\widehat{G}$ in (26) of the main paper is studied and its empirical frequency distribution is reported in Tables 10–12, for the case of no error autocorrelation ($\theta = 0$). We set that the IC penalty as $\kappa_{nT} = (nT)^{-0.35}$ and $G_{\max} = 5$. The critical value of the Hausman test is set to $cv_{nT} = (1 + 5\log(nT))\,\chi^2\left(\overline{G}\right)$ and $\overline{G} = (G_{\max} - G + 1)$. As before, the IC procedure has a clear tendency to underestimate the true number of groups G^0, although the magnitude of the error rate declines as T rises. The combined IC-Hausman procedure shows good performance in selecting the true number of groups, with rapidly diminishing error rates as T increases.

[Insert Tables 10–12 Here]

Next, the performance of the recursive k-means clustering algorithm and the associated post-clustering estimate is checked when $\theta = 0$ and the true number of groups is known. Tables 14–16 report the CE, RMSE, and bias of the post-clustering estimates. From Tables 14–16 it is clear that the difference between the oracle and post-clustering estimates decreases as the sample size increases, corroborating the asymptotic equivalence between these two sets of estimators under the joint convergence framework. It is also evident that the CE with $\theta = 0$ is smaller than the value with $\theta = 0.5$, as expected.

[Insert Tables 14-16 Here]

Finally, Table 17 reports the empirical rejection rates of the PP t and J tests when $n = 1$ (the time series case) and the empirical rejection rates of the corresponding post-clustering panel tests when $n > 1$. The panel t test has conservative size in finite samples, whereas the panel J test shows mild oversizing that diminishes as both n and T get larger. Interestingly, both the time series t and J tests are conservative when $\theta = 0$. As in the case of no serial correlation, there is a substantial improvement in the power of the two post-clustering panel tests over that of the time series tests. For instance, when $c_2^0 = 0.2$, an empirically plausible value, the power of the panel test is much larger than that of the two time series tests. For instance, if $T = 100$ the panel t test raises the power of the time series t test from 0.0112 to 0.441 when $n_j = 10$, to 0.712 when $n_j = 20$, and to 0.878 when

$n_j = 30$; and the panel t test with $T = 100$ has substantially greater power than the time series t test with $T = 200$ (0.878 versus 0.292). These findings corroborate the power enhancements introduced by cross section information and statistical averaging.

[Insert Table 17 Here]

E Tables & Figures

Table 1: SNRs for mildly explosive, unit root and mildly stationary groups

n	T	DGP 1			DGP 2			DGP 2	
		$c_1(\times 10^5)$	$c_2(\times 10^2)$	$c_3(\times 1)$	$c_1(\times 10^5)$	$c_2(\times 10^2)$	$c_3(\times 1)$	$c_1(\times 10^5)$	$c_2(\times 1)$
30	100	6.798	2.792	8.375	6.814	7.338	8.493	6.726	47.096
30	150	82.066	4.935	13.139	82.202	16.354	12.926	81.978	71.135
30	200	562.425	6.942	15.426	562.771	28.344	15.426	559.628	85.447
60	100	6.595	2.859	8.881	6.606	7.458	8.623	6.693	47.220
60	150	81.352	5.009	13.496	81.506	16.580	13.088	81.527	71.234
60	200	562.555	6.974	15.686	563.078	28.350	15.584	561.495	85.547
90	100	6.668	2.852	9.026	6.676	7.434	8.642	6.724	47.204
90	150	81.778	5.077	13.408	81.935	16.816	13.155	82.117	71.359
90	200	565.714	7.013	15.754	566.288	28.469	15.641	561.739	85.688

Table 2: Empirical frequency of model selection under DGP 0 ($\theta = 0.5$)

		IC Estimator $\widetilde{G}$					Proposed Estimator $\widehat{G}$				
n	T	$G=1$	$G=2$	$G=3$	$G=4$	$G=5$	$G=1$	$G=2$	$G=3$	$G=4$	$G=5$
120	150	.000	1.000	.000	.000	.000	.000	.008	.980	.012	.000
120	250	.000	1.000	.000	.000	.000	.000	.005	.993	.001	.001
120	350	.015	.985	.000	.000	.000	.000	.004	.996	.000	.000
120	450	.476	.524	.000	.000	.000	.000	.004	.996	.000	.000
150	150	.000	1.000	.000	.000	.000	.000	.000	.992	.006	.002
150	250	.000	1.000	.000	.000	.000	.000	.000	.998	.001	.001
150	350	.000	1.000	.000	.000	.000	.000	.000	1.000	.000	.000
150	450	.014	.986	.000	.000	.000	.000	.000	1.000	.000	.000

Table 3: Empirical frequency of model selection under DGP 1 ($\theta = 0.5$)

		IC Estimator $\widetilde{G}$					Proposed Estimator $\widehat{G}$				
n	T	$G=1$	$G=2$	$G=3$	$G=4$	$G=5$	$G=1$	$G=2$	$G=3$	$G=4$	$G=5$
120	150	.000	.976	.024	.000	.000	.000	.000	.726	.067	.207
120	250	.000	.891	.109	.000	.000	.000	.000	.892	.041	.067
120	350	.000	.774	.226	.000	.000	.000	.000	.963	.011	.026
120	450	.000	.610	.390	.000	.000	.000	.000	.982	.008	.010
150	150	.000	.994	.006	.000	.000	.000	.000	.669	.085	.246
150	250	.000	.931	.069	.000	.000	.000	.000	.884	.037	.079
150	350	.000	.829	.171	.000	.000	.000	.000	.966	.008	.026
150	450	.000	.671	.329	.000	.000	.000	.000	.977	.010	.013

Table 4: Empirical frequency of model selection under DGP 2 ($\theta = 0.5$)

		IC Estimator $\widetilde{G}$					Proposed Estimator $\widehat{G}$				
n	T	$G=1$	$G=2$	$G=3$	$G=4$	$G=5$	$G=1$	$G=2$	$G=3$	$G=4$	$G=5$
120	150	.000	.625	.375	.000	.000	.000	.000	.743	.058	.199
120	250	.000	.259	.741	.000	.000	.000	.000	.911	.023	.066
120	350	.000	.091	.909	.000	.000	.000	.000	.957	.012	.031
120	450	.000	.017	.983	.000	.000	.000	.000	.984	.006	.010
150	150	.000	.700	.300	.000	.000	.000	.000	.678	.078	.244
150	250	.000	.271	.729	.000	.000	.000	.000	.914	.020	.066
150	350	.000	.094	.906	.000	.000	.000	.000	.948	.017	.035
150	450	.000	.017	.983	.000	.000	.000	.000	.984	.007	.009

Table 5: Empirical frequency of model selection under DGP 3 ($\theta = 0.5$)

		IC Estimator $\widetilde{G}$					Proposed Estimator $\widehat{G}$				
n	T	$G=1$	$G=2$	$G=3$	$G=4$	$G=5$	$G=1$	$G=2$	$G=3$	$G=4$	$G=5$
120	150	.000	1.000	.000	.000	.000	.000	.738	.003	.233	.026
120	250	.000	1.000	.000	.000	.000	.000	.928	.000	.065	.007
120	350	.000	1.000	.000	.000	.000	.000	.967	.000	.032	.001
120	450	.000	1.000	.000	.000	.000	.000	.986	.000	.014	.000
150	150	.000	1.000	.000	.000	.000	.000	.703	.003	.248	.046
150	250	.000	1.000	.000	.000	.000	.000	.904	.000	.093	.003
150	350	.000	1.000	.000	.000	.000	.000	.957	.000	.042	.001
150	450	.000	1.000	.000	.000	.000	.000	.988	.000	.012	.000

Table 6: Clustering and estimation by the two stage procedure under DGP 1 ($\theta = 0.5$)

			Group 1				Group 2				Group 3			
			Clustering		Oracle		Clustering		Oracle		Clustering		Oracle	
T	T	CE	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias
60	100	.015	.003	-.002	.003	-.002	.180	-.164	.175	-.160	3.600	3.596	3.568	3.565
60	150	.007	.001	-.001	.001	-.001	.154	-.142	.151	-.139	3.697	3.695	3.680	3.678
60	200	.004	.000	-.000	.000	-.000	.138	-.128	.138	-.127	3.747	3.745	3.739	3.737
90	100	.014	.003	-.002	.003	-.002	.173	-.163	.168	-.158	3.603	3.600	3.569	3.567
90	150	.007	.001	-.001	.001	-.001	.148	-.140	.146	-.137	3.695	3.693	3.679	3.678
90	200	.003	.000	-.000	.000	-.000	.134	-.127	.133	-.126	3.748	3.746	3.740	3.739

Table 7: Clustering and estimation by the two stage procedure under DGP 2 ($\theta = 0.5$)

			Group 1				Group 2				Group 3			
			Clustering		Oracle		Clustering		Oracle		Clustering		Oracle	
T	T	CE	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias
60	100	.013	.003	-.002	.003	-.002	.177	-.168	.174	-.166	3.599	3.595	3.568	3.565
60	150	.005	.001	-.001	.001	-.001	.139	-.133	.138	-.132	3.697	3.694	3.680	3.678
60	200	.003	.000	-.000	.000	-.000	.114	-.109	.114	-.109	3.748	3.745	3.739	3.737
90	100	.012	.003	-.002	.003	-.002	.172	-.166	.170	-.164	3.601	3.597	3.569	3.567
90	150	.005	.001	-.001	.001	-.001	.134	-.130	.133	-.129	3.699	3.697	3.679	3.678
90	200	.003	.000	-.000	.000	-.000	.110	-.107	.110	-.107	3.749	3.748	3.740	3.739

Table 8: Clustering and estimation by the two stage procedure under DGP 3 ($\theta = 0.5$)

			Group 1				Group 2			
			Clustering		Oracle		Clustering		Oracle	
T	T	CE	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias
60	100	.005	.003	-.002	.003	-.002	.562	.559	.563	.560
60	150	.002	.001	-.001	.001	-.001	.586	.584	.586	.584
60	200	.001	.000	-.000	.000	-.000	.600	.598	.600	.598
90	100	.006	.002	-.002	.002	-.002	.575	.556	.563	.561
90	150	.002	.001	-.001	.001	-.001	.587	.585	.587	.585
90	200	.001	.000	-.000	.000	-.000	.600	.599	.600	.599

Table 9: Power of tests for detecting explosiveness ($\theta = 0.5$)

n	T	DGP 1				DGP 2				DGP 3			
		$c_1 = 1$		$c_2 = 0$		$c_1 = 1$		$c_2 = 0.1$		$c_1 = 1$		$c_2 = -1$	
		t-test	J-test	t-test	J-test	t-test	J-test	t-test	J-test	t-test	J-test	t-test	J-test
1	100	.994	.994	.060	.059	.994	.994	.175	.175	-	-	-	-
1	150	.996	.996	.064	.066	.996	.996	.293	.297	-	-	-	-
1	200	.999	.999	.069	.069	.999	.999	.382	.382	-	-	-	-
30	100	1.000	1.000	.045	.047	1.000	1.000	.599	.262	1.000	1.000	.010	.226
30	150	1.000	1.000	.062	.064	1.000	1.000	.797	.441	1.000	1.000	.006	.195
30	200	1.000	1.000	.042	.041	1.000	1.000	.920	.657	1.000	1.000	.003	.179
60	100	1.000	1.000	.070	.057	1.000	1.000	.807	.527	1.000	1.000	.027	.513
60	150	1.000	1.000	.054	.057	1.000	1.000	.961	.785	1.000	1.000	.014	.454
60	200	1.000	1.000	.058	.059	1.000	1.000	.994	.951	1.000	1.000	.005	.473
90	100	1.000	1.000	.082	.079	1.000	1.000	.917	.705	1.000	1.000	.089	.766
90	150	1.000	1.000	.052	.053	1.000	1.000	.992	.938	1.000	1.000	.040	.719
90	200	1.000	1.000	.040	.042	1.000	1.000	.999	.996	1.000	1.000	.029	.734

Table 10: Empirical frequency of model selection under DGP 0 ($\theta = 0$)

		IC Estimator $\widetilde{G}$					Proposed Estimator $\widehat{G}$				
n	T	$G=1$	$G=2$	$G=3$	$G=4$	$G=5$	$G=1$	$G=2$	$G=3$	$G=4$	$G=5$
120	150	.000	1.000	.000	.000	.000	.000	.642	.286	.052	.020
120	250	.000	1.000	.000	.000	.000	.000	.027	.791	.037	.145
120	350	.000	1.000	.000	.000	.000	.000	.000	.932	.016	.052
120	450	.000	1.000	.000	.000	.000	.000	.000	.987	.004	.009
150	150	.000	1.000	.000	.000	.000	.000	.454	.388	.090	.068
150	250	.000	1.000	.000	.000	.000	.000	.006	.637	.042	.315
150	350	.000	1.000	.000	.000	.000	.000	.000	.842	.015	.143
150	450	.000	1.000	.000	.000	.000	.000	.000	.972	.012	.016

Table 11: Empirical frequency of model selection under DGP 1 ($\theta = 0$)

		IC Estimator $\widetilde{G}$					Proposed Estimator $\widehat{G}$				
n	T	$G=1$	$G=2$	$G=3$	$G=4$	$G=5$	$G=1$	$G=2$	$G=3$	$G=4$	$G=5$
120	150	.000	.005	.995	.000	.000	.000	.000	.797	.068	.135
120	250	.000	.014	.986	.000	.000	.000	.001	.934	.023	.042
120	350	.000	.023	.977	.000	.000	.000	.002	.975	.009	.014
120	450	.000	.033	.967	.000	.000	.000	.002	.992	.002	.004
150	150	.000	.000	1.000	.000	.000	.000	.000	.738	.072	.190
150	250	.000	.000	1.000	.000	.000	.000	.000	.898	.029	.073
150	350	.000	.000	1.000	.000	.000	.000	.000	.972	.008	.020
150	450	.000	.000	1.000	.000	.000	.000	.000	.992	.000	.008

Table 12: Empirical frequency of model selection under DGP 2 ($\theta = 0$)

		IC Estimator $\widetilde{G}$					Proposed Estimator $\widehat{G}$				
n	T	$G=1$	$G=2$	$G=3$	$G=4$	$G=5$	$G=1$	$G=2$	$G=3$	$G=4$	$G=5$
120	150	.000	.000	1.000	.000	.000	.000	.000	.792	.082	.126
120	250	.000	.000	1.000	.000	.000	.000	.000	.861	.088	.051
120	350	.000	.000	1.000	.000	.000	.000	.000	.834	.119	.047
120	450	.000	.000	1.000	.000	.000	.000	.000	.808	.126	.066
150	150	.000	.000	1.000	.000	.000	.000	.000	.732	.098	.170
150	250	.000	.000	1.000	.000	.000	.000	.000	.804	.117	.079
150	350	.000	.000	1.000	.000	.000	.000	.000	.727	.186	.087
150	450	.000	.000	1.000	.000	.000	.000	.000	.705	.185	.110

Table 13: Empirical frequency of model selection under DGP 3 ($\theta = 0$)

		IC Estimator $\widetilde{G}$					Proposed Estimator $\widehat{G}$				
n	T	$G=1$	$G=2$	$G=3$	$G=4$	$G=5$	$G=1$	$G=2$	$G=3$	$G=4$	$G=5$
120	150	.000	1.000	.000	.000	.000	.000	.817	.000	.173	.010
120	250	.000	1.000	.000	.000	.000	.000	.949	.000	.050	.001
120	350	.000	1.000	.000	.000	.000	.000	.986	.000	.013	.001
120	450	.000	1.000	.000	.000	.000	.000	.991	.000	.009	.000
150	150	.000	1.000	.000	.000	.000	.000	.791	.000	.194	.015
150	250	.000	1.000	.000	.000	.000	.000	.913	.000	.085	.002
150	350	.000	1.000	.000	.000	.000	.000	.971	.000	.029	.000
150	450	.000	1.000	.000	.000	.000	.000	.989	.000	.011	.000

Table 14: Clustering and estimation by the two stage procedure under DGP 1 ($\theta = 0$)

			Group 1				Group 2				Group 3			
			Clustering		Oracle		Clustering		Oracle		Clustering		Oracle	
T	T	CE	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias
60	100	.011	.003	-.002	.003	-.002	.477	-.462	.476	-.461	.384	-.220	.406	-.289
60	150	.004	.001	-.001	.001	-.001	.421	-.409	.420	-.409	.352	-.215	.358	-.243
60	200	.002	.000	-.000	.000	-.000	.379	-.369	.379	-.369	.342	-.211	.341	-.224
90	100	.011	.003	-.002	.003	-.002	.466	-.457	.467	-.458	.352	-.212	.379	-.288
90	150	.004	.001	-.001	.001	-.001	.408	-.400	.408	-.400	.322	-.218	.332	-.248
90	200	.002	.000	-.000	.000	-.000	.371	-.364	.372	-.364	.305	-.207	.308	-.221

Table 15: Clustering and estimation by the two stage procedure under DGP 2 ($\theta = 0$)

			Group 1				Group 2				Group 3			
			Clustering		Oracle		Clustering		Oracle		Clustering		Oracle	
T	T	CE	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias
60	100	.009	.003	-.002	.003	-.002	.292	-.281	.292	-.282	.392	-.234	.406	-.289
60	150	.003	.001	-.001	.001	-.001	.221	-.214	.221	-.214	.355	-.219	.358	-.243
60	200	.002	.000	-.000	.000	-.000	.173	-.168	.174	-.168	.342	-.214	.341	-.224
90	100	.009	.003	-.002	.003	-.002	.284	-.278	.285	-.278	.356	-.226	.379	-.288
90	150	.003	.001	-.001	.001	-.001	.211	-.207	.212	-.208	.321	-.223	.332	-.248
90	200	.002	.000	-.000	.000	-.000	.168	-.164	.168	-.164	.304	-.209	.308	-.221

Table 16: Clustering and estimation by the two stage procedure under DGP 3 ($\theta = 0$)

			Group 1				Group 2			
			Clustering		Oracle		Clustering		Oracle	
T	T	CE	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias
60	100	.004	.002	-.002	.002	-.002	.416	-.392	.414	-.392
60	150	.001	.001	-.001	.001	-.001	.346	-.324	.345	-.323
60	200	.001	.000	-.000	.000	-.000	.306	-.284	.305	-.284
90	100	.005	.002	-.002	.002	-.002	.404	-.388	.403	-.388
90	150	.001	.001	-.001	.001	-.001	.336	-.320	.335	-.320
90	200	.001	.000	-.000	.000	-.000	.297	-.283	.297	-.282

Table 17: Power of tests for detecting explosiveness ($\theta = 0$)

		DGP 1				DGP 2				DGP 3			
		$c_1 = 1$		$c_2 = 0$		$c_1 = 1$		$c_2 = 0.1$		$c_1 = 1$		$c_2 = -1$	
n	T	t-test	J-test	t-test	J-test	t-test	J-test	t-test	J-test	t-test	J-test	t-test	J-test
1	100	.992	.992	.025	.025	.992	.992	.112	.111	-	-	-	-
1	150	.996	.996	.022	.020	.996	.996	.211	.211	-	-	-	-
1	200	.999	.999	.027	.027	.999	.999	.292	.296	-	-	-	-
30	100	1.000	1.000	.012	.062	1.000	1.000	.441	.324	1.000	1.000	.000	.098
30	150	1.000	1.000	.020	.068	1.000	1.000	.665	.479	1.000	1.000	.000	.087
30	200	1.000	1.000	.010	.055	1.000	1.000	.847	.689	1.000	1.000	.000	.084
60	100	1.000	1.000	.019	.093	1.000	1.000	.712	.622	1.000	1.000	.000	.238
60	150	1.000	1.000	.017	.083	1.000	1.000	.926	.826	1.000	1.000	.000	.179
60	200	1.000	1.000	.012	.067	1.000	1.000	.986	.964	1.000	1.000	.000	.190
90	100	1.000	1.000	.040	.129	1.000	1.000	.878	.815	1.000	1.000	.000	.389
90	150	1.000	1.000	.015	.088	1.000	1.000	.987	.960	1.000	1.000	.000	.311
90	200	1.000	1.000	.005	.058	1.000	1.000	.999	.998	1.000	1.000	.000	.336

Table 18: Estimated group structure of the housing price indices in China

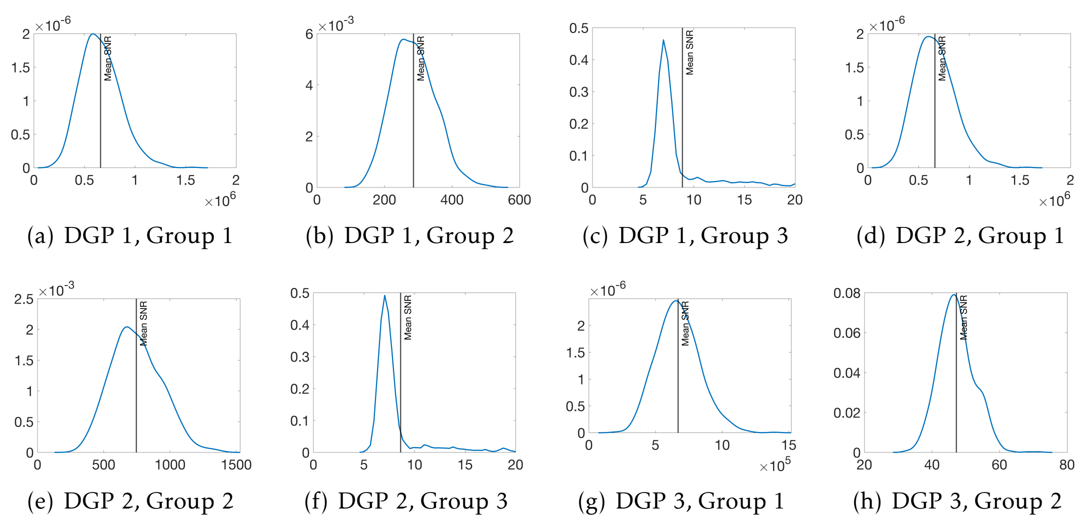
Group 1	beijing, xinlingol, chaoyang, wuludao, xuzhou, yangzhou, jiangyan, taizhou, bengbu, zhangzhou, ningde, nanchang, jingdezhen, pingxiang, xinyu, yichun, shangrao, fuzhou_jx, dezhou, zhumadian, changde, guangzhou, shantou, luzhou, xining, Urumuqi, changji;
Group 2	xingtai, baoding, zhangjiakou, hohhot, baotou, shenyang, dandong, changchun, songyuan, harbin, nanjing, yancheng, suqian, hangzhou, shaoxing, hefei, anqing, xuancheng, jiujiang, zhengzhou, kaifeng, luoyang, xinxiang, xuchang, luohe, nanyang, changsha, shenzhen, jiangmen, zhaoqing, huizhou, shanwei, yangjiang, jieyang, nanning, haikou, chongqing, deyang, leshan, nanchong, kunming, xi'an;
Group 3	tianjin, shijiazhuang, tangshan, qinhuangdao, langfang, dalian, anshan, yingkou, tieling, shanghai, wuxi, changzhou, suzhou, nantong, lianyungang, huai'an, zhenjiang, ningbo, wenzhou, jiaxing, huzhou, jinhua, wuhu, huangshan, chuzhou, fuzhou, xiamen, jinan, qingdao, zaozhuang, rizhao, foshan, heyuan, qingyuan, dongguan, zhongshan, chengdu, mianyang.

Table 19: Estimated group structure of the prices for the U.S. housing market

Group 1	Atlanta, Boston, Charlotte, Dallas, Miami, New York, Seattle ;
Group 2	Chicago, Detroit, Las Vegas, San Francisco .

Table 20: Estimated group structure of the prices for S&P 500 stocks

Group 1	ACN, ADBE, ADP, AMAT, AMZN, ATO, AVY, BDX, BK, BLK, BSX, CDNS, CI, CMA, COST, CSCO, CTAS, DE, DGX, DPZ, EA, EQIX, ETFC, FDX, FIS, FITB, GD, HAS, HBAN, HD, HIG, HON, IEX, INTC, INTU, JBHT, JKHY, JNJ, JPM, KEY;
Group 2	APPLE, AEE, AEP, AES, AFL, AIG, AIZ, ALB, ALK, AMD, AMGN, APA, APD, AXP, BAX, BIIB, BKR, BLL, BXP, C, CAG, CAT, CB, CCL, CE, CERN, CHD, CHRW, CL, CME, CMI, CMS, CNP, COF, COP, CPB, CPRT, CSX, CTSH, CTXS, CVS, CVX, DHR, DIS, DISH, DLR, DLTR, DOV, DRI, DTE, DXC, EBAY, ED, EFX, EL, EMR, EOG, ES, ESS, ETN, ETR, EVRG, EW, EXC, FFIV, FISV, FLIR, FLS, FRT, GILD, GIS, GL, GLW, GPN, GRMN, GS, HAL, HES, HFC, HOG, HOLX, HRL, HSIC, HST, HSY, HWM, IDXX, IFF, ILMN, INCY, IP, IPG, J, JCI, JNPR, JWN, K, KIM, KLAC, KMB, KMX, KO, KR, KSS, LB, LEG.



(a) DGP 1, Group 1 (b) DGP 1, Group 2 (c) DGP 1, Group 3 (d) DGP 2, Group 1

(e) DGP 2, Group 2 (f) DGP 2, Group 3 (g) DGP 3, Group 1 (h) DGP 3, Group 2

Figure 1: Empirical frequency distribution and sample average (shown by the vertical line) of the signal-to-noise ratio in each group of DGPs 1–3.

References

Donsker, M. and S. Varadhan (1977): "On laws of the iterated logarithm for local times," *Communications on Pure and Applied Mathematics*, 30, 707–753.

Fernholz, L. T. and H. Teicher (1980): "Stability of random variables and iterated logarithm laws for martingales and quadratic forms," *The Annals of Probability*, 765–774.

Freedman, D. A. (1975): "On tail probabilities for martingales," *Annals of Probability*, 100–118.

Hamilton, J. D. (1994): *Time Series Analysis*, Princeton university press.

Huang, W., S. Jin, P. C. B. Phillips, and L. Su (2021): "Nonstationary panel models with latent group structures and cross-section dependence," *Journal of Econometrics*, 221, 198–222.

Laurent, B. and P. Massart (2000): "Adaptive estimation of a quadratic functional by model selection," *Annals of Statistics*, 1302–1338.

Liu, Y., P. C. B. Phillips, and J. Yu (2022): "Panel clustering approach to analyzing bubble behavior," *Cowles Foundation Discussion Paper No. 2323*.

Phillips, P. C. B. (1987): "Time series regression with a unit root," *Econometrica*, 277–301.

Phillips, P. C. B. and S. N. Durlauf (1986): "Multiple time series regression with integrated processes," *The Review of Economic Studies*, 53, 473–495.

Phillips, P. C. B., D. Li, and J. Gao (2017): "Estimating smooth structural change in cointegration models," *Journal of Econometrics*, 196, 180–195.

Phillips, P. C. B. and T. Magdalinos (2007): "Limit theory for moderate deviations from a unit root under weak dependence," *in G. D. A. Phillips and E. Tzavalis, eds., The Refinement of Econometric Estimation and Test Procedures: Finite Sample and Asymptotic Analysis*, 123–162.

Phillips, P. C. B. and H. R. Moon (1999): "Linear regression limit theory for nonstationary panel data," *Econometrica*, 67, 1057–1111.

Phillips, P. C. B. and P. Perron (1988): "Testing for a unit root in time series regression," *Biometrika*, 75, 335–346.

Phillips, P. C. B. and V. Solo (1992): "Asymptotics for linear processes," *The Annals of Statistics*, 971–1001.

Su, L., Z. Shi, and P. C. B. Phillips (2016): "Identifying latent structures in panel data," *Econometrica*, 84, 2215–2264.