# Supplemental Material to
# A Panel Clustering Approach to Analyzing Bubble Behavior

By

Yanbo Liu, Peter C.B. Phillips and Jun Yu

February 2022

COWLES FOUNDATION DISCUSSION PAPER NO. 2323S

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

# 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*

February 15, 2022

**Abstract**

This online supplement has six 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 extends the two-stage algorithm and the corresponding post-clustering statistics to the mixed-root panel autoregressive model with purely stationary, unit, and purely explosive roots. Section E overviews experimental designs and reports simulation findings. Section F 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

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

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

**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

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

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

$$= \frac{2}{n}\sum_{i=1}^{n} \frac{1}{\Upsilon_{iT}} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}\widetilde{u}_{it}\overline{c}_i^0 - \frac{2}{n}\sum_{i=1}^{n} \frac{1}{\Upsilon_{iT}} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}\widetilde{u}_{it}\overline{c}_i.$$

Based on the assumption of latent membership, we have

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

Since $\left|c_{\tilde{j}}^0\right| \le c_{up}$, for any $\tilde{j} \in \left\{1,2,\dots,G^0\right\}$,

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

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

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

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

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

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

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

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*

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

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

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

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 $\widetilde{j} \neq j$,

$$
\begin{aligned}
& T^{2\gamma}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\Upsilon_{iT}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2\left(\widehat{\rho}_{\tau(j)}^* - \widehat{\rho}_{\tau(\widetilde{j})}^*\right)^2\right) \\
& \geq T^{2\gamma}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\Upsilon_{iT}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2\left(\rho_j^0 - \rho_{\widetilde{j}}^0\right)^2\right) - T^{2\gamma}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\Upsilon_{iT}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2\left(\widehat{\rho}_{\tau(j)}^* - \rho_j^0\right)^2\right) \qquad (7) \\
& - T^{2\gamma}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\Upsilon_{iT}}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2\left(\widehat{\rho}_{\tau(\widetilde{j})}^* - \rho_{\widetilde{j}}^0\right)^2\right). \qquad (8)
\end{aligned}
$$

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 $\tau$ is not only an 'onto' mapping but also 'one-to-one' mapping. Hence, $\tau$ has an inverse denoted as $\tau^{-1}$ 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

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

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 $\tau(\widetilde{j}) = j$. For any $\eta > 0$, we define $\mathcal{N}_\eta$, $\widehat{g}_i(\widehat{c}^*)$, and $\widehat{\delta}$ by

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

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

$$
\widehat{\delta} := \left(\widehat{g}_1(\widehat{c}^*), \widehat{g}_2(\widehat{c}^*), \ldots, \widehat{g}_n(\widehat{c}^*)\right),
$$

where we treat the scaling parameter $\gamma$ as given *a priori*.

**Lemma A.3** *Suppose Assumptions 1 and 2 hold. Then, for any fixed $M > 0$,* *(i) if* $\bar{c}_i^0 > 0$,

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

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

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

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

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

**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 $\bar{c}_i^0 > 0$, fix an arbitrary $M > 0$ and we have

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

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

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

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,

$$
\begin{aligned}
&n \max_{i\in\mathcal{I}_n} \Pr\left( \frac{(\log T)^2}{\left(\overline{\rho}_i^0\right)^{2T} T^{\gamma-1}} \left|\overline{y}_{i,-1}\overline{u}_i\right| \geq \frac{M}{2} \right) \leq n \max_{i\in\mathcal{I}_n} \frac{2(\log T)^2\, \mathbb{E}\left|\overline{y}_{i,-1}\overline{u}_i\right|}{\left(\overline{\rho}_i^0\right)^{2T} T^{\gamma-1} M} \\
&\leq n \max_{i\in\mathcal{I}_n} \frac{2(\log T)^2\, \mathbb{E}\left|\left(\sum_{t=1}^T y_{i,t-1}\right)\left(\sum_{t=1}^T u_{it}\right)\right|}{\left(\overline{\rho}_i^0\right)^{2T} T^{1+\gamma} M} \\
&\leq n \max_{i\in\mathcal{I}_n} \frac{2(\log T)^2 \left(\mathbb{E}\left(\sum_{t=1}^T y_{i,t-1}\right)^2\right)^{\frac{1}{2}} \left(\mathbb{E}\left(\sum_{t=1}^T u_{it}\right)^2\right)^{\frac{1}{2}}}{\left(\overline{\rho}_i^0\right)^{2T} T^{1+\gamma} M} \\
&\leq n \max_{i\in\mathcal{I}_n} K \frac{(\log T)^2 \left(\overline{\rho}_i^0\right)^T T^{\frac{3\gamma+1}{2}}}{\left(\overline{\rho}_i^0\right)^{2T} T^{1+\gamma}} = O\left( \frac{(\log T)^2\, n T^{\frac{-1+\gamma}{2}}}{(\rho_{low})^T} \right) = o(1), \qquad (14)
\end{aligned}
$$

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

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

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

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

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

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

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

$$n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{(\log_2 T)^2}{T^{2-\gamma}}\left|F_i(1)\sum_{t=1}^{T}z_{1it}\right|\geq\frac{M}{12}\right)$$

$$
\begin{aligned}
&\leq n \max_{i\in\mathcal{I}_n} \Pr\left( \frac{(\log_2 T)^2}{T^{2-\gamma}} \left| F_i(1) \sum_{t=1}^{T} z_{1it} \right| \geq \frac{M}{12}, V_{iT} \leq v_{nT} \right) + n \max_{i\in\mathcal{I}_n} \Pr\left(V_{iT} > v_{nT}\right) \\
&\leq \exp\left( \frac{-M^2 T^{4-2\gamma} / \left(144 F_i^2(1) (\log_2 T)^4\right) + 2v_{nT}\log(n) + 4T^{2-\gamma} M d_{nT} \log(n) / \left(12 F_i(1) (\log_2 T)^2\right)}{2v_{nT} + 4T^{2-\gamma} M d_{nT} / \left(12 F_i(1) (\log_2 T)^2\right)} \right) \\
&+ o\left(nT^4 v_{nT}^{-2}\right) \\
&= o(1). \qquad (15)
\end{aligned}
$$

Moreover,

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

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

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

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

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

For (A.3), we have

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

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

$$
\begin{aligned}
& n \max_{i \in \mathcal{I}_n} \Pr\left( \left| \frac{(\log_2 T)^2 F_i(1)}{T^{2-\gamma}} \left( \sum_{s=1}^{T-1} \epsilon_{is} \right) \widetilde{\epsilon}_{i,T} \right| \geq \frac{M}{16} \right) \\
& \leq n \max_{i \in \mathcal{I}_n} \frac{256 (\log_2 T)^4 \mathbb{E} \left| F_i(1) \left( \sum_{s=1}^{T-1} \epsilon_{is} \right) \widetilde{\epsilon}_{i,T} \right|^2}{T^{4-2\gamma} M^2} \\
& \leq \frac{256 (\log_2 T)^4}{T^{4-2\gamma} M^2} \times n \max_{i \in \mathcal{I}_n} F_i^2(1) \mathbb{E} \left\{ \left( \sum_{s=1}^{T-1} \epsilon_{is} \right)^2 \widetilde{\epsilon}_{i,T}^2 \right\} \\
& = O\left( \frac{n (\log_2 T)^4}{T^{3-2\gamma}} \right) = o(1),
\end{aligned} \tag{19}
$$

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

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

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

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

where the rate restriction $\frac{T^{1-\gamma}}{(\log_2 T)^2} \succ n$ 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

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

$$= O\left(\frac{n(\log_2 T)^4}{T^{3-2\gamma}}\right) = o(1), \tag{21}$$

where the rate restriction $\frac{T^{3-2\gamma}}{(\log_2 T)^4} > n$ 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

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

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,

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

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

$$
\begin{aligned}
& n \max_{i \in \mathcal{I}_n} \Pr\left(\frac{(\log_2 T)^2}{T^{\frac{9}{4}-\gamma}} \left|\sum_{t=1}^{T} y_{i,t-1}\right| \geq \frac{\sqrt{2M}}{2}\right) \\
& \leq \underbrace{n \max_{i \in \mathcal{I}_n} \Pr\left(\frac{(\log_2 T)^2}{T^{\frac{9}{4}-\gamma}} \left|F_i(1) \sum_{s=0}^{T-1}\left(\sum_{t=s+1}^{T} 1\right) \epsilon_{is}\right| \geq \frac{\sqrt{2M}}{4}\right)}_{\text{(A.6.1)}} \\
& + \underbrace{n \max_{i \in \mathcal{I}_n} \Pr\left(\frac{(\log_2 T)^2}{T^{\frac{9}{4}-\gamma}} \left|\sum_{s=0}^{T-1}\left(\sum_{t=s+1}^{T} 1\right)(\widetilde{\epsilon}_{i,s-1} - \widetilde{\epsilon}_{i,s})\right| \geq \frac{\sqrt{2M}}{4}\right)}_{\text{(A.6.2)}},
\end{aligned}
$$

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 $\frac{T^{\frac{3}{2}-2\gamma}}{(\log_2 T)^4} > \sqrt{n}\log n$ and $\frac{T^{1-\gamma}}{(\log_2 T)^2} > n^{\frac{1}{4}}\log n$ 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 $\frac{T^{\frac{3}{2}-2\gamma}}{(\log_2 T)^4} > n$ 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 $\overline{c}_i^0 < 0$, for any $M > 0$, the demeaned numerator decomposes as

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

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

(iii.1) For the first term of (33), since $y_{i0} = 0$,

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T} y_{i,t-1}u_{it} \\
&= \frac{1}{T}\sum_{t=1}^{T}\left(\sum_{s=1}^{t-1}\left(\overline{\rho}_i^0\right)^{t-1-s}\left(F_i(1)\epsilon_{is} + \widetilde{\epsilon}_{i,s-1} - \widetilde{\epsilon}_{i,s}\right)\right)\left(F_i(1)\epsilon_{it} + \widetilde{\epsilon}_{i,t-1} - \widetilde{\epsilon}_{i,t}\right) \\
&= \underbrace{\frac{F_i(1)}{T}\sum_{t=1}^{T} y_{i,t-1}\epsilon_{it}}_{\text{(B.1)}} + \underbrace{\frac{F_i(1)}{T}\sum_{t=1}^{T}\left(\sum_{s=1}^{t-1}\left(\overline{\rho}_i^0\right)^{t-1-s}\epsilon_{is}\right)\left(\widetilde{\epsilon}_{i,t-1} - \widetilde{\epsilon}_{i,t}\right)}_{\text{(B.2)}} \\
&\quad \underbrace{+ \frac{1}{T}\sum_{t=1}^{T}\left(\sum_{s=1}^{t-1}\left(\overline{\rho}_i^0\right)^{t-1-s}\left(\widetilde{\epsilon}_{i,s-1} - \widetilde{\epsilon}_{is}\right)\right)\left(\widetilde{\epsilon}_{i,t-1} - \widetilde{\epsilon}_{i,t}\right)}_{\text{(B.3)}}.
\end{aligned}$$

For the term (B.1), by the exponential inequality of Freedman (1975), we have

$$n \max_{i\in\mathcal{I}_n} \Pr\left(\left|\frac{F_i(1)}{T}\sum_{t=1}^{T} y_{i,t-1}\epsilon_{it}\right| \geq \frac{M}{6}\right) = o(1),$$

where the rate restrictions $T^{1-\gamma} > \sqrt{n}\log n$ and $T^{\frac{3-2\gamma}{4}} > n^{\frac{1}{4}}\log n$ are assured by Assumption 2. For terms (B.2) and (B.3), by the Markov inequality, we have

$$n \max_{i\in\mathcal{I}_n} \Pr\left(\left|\frac{F_i(1)}{T}\sum_{t=1}^{T}\left(\sum_{s=1}^{t-1}\left(\overline{\rho}_i^0\right)^{t-1-s}\epsilon_{is}\right)\left(\widetilde{\epsilon}_{i,t-1} - \widetilde{\epsilon}_{i,t}\right)\right| \geq \frac{M}{6}\right) = o(1),$$

and

$$n \max_{i\in\mathcal{I}_n} \Pr\left(\left|\frac{1}{T}\sum_{t=1}^{T}\left(\sum_{s=1}^{t-1}\left(\overline{\rho}_i^0\right)^{t-1-s}\left(\widetilde{\epsilon}_{i,s-1} - \widetilde{\epsilon}_{is}\right)\right)\left(\widetilde{\epsilon}_{i,t-1} - \widetilde{\epsilon}_{i,t}\right)\right| \geq \frac{M}{6}\right) = o(1),$$

where the rate restriction $T^{1-\gamma} > n$ is assured by Assumption 2. Therefore, it follows that

$$n \max_{i\in\mathcal{I}_n} \Pr\left(\left|\frac{1}{T}\sum_{t=1}^{T} y_{i,t-1}u_{it}\right| \geq \frac{M}{2}\right) = o(1). \tag{34}$$

(iii.2) For the second term of (33), we have

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

$$
\begin{aligned}
&\le n \max_{i\in\mathcal{I}_n} \Pr\left(\frac{1}{T^2}\left|\left(\sum_{t=1}^{T} y_{i,t-1}\right)\left(\sum_{t=1}^{T} u_{i,t}\right)\right| \ge \frac{M}{2}, \frac{1}{T^{\frac{5}{4}}}\left|\sum_{t=1}^{T} y_{i,t-1}\right| \le \frac{\sqrt{2M}}{2}\right) \\
&+ n \max_{i\in\mathcal{I}_n} \Pr\left(\frac{1}{T^{\frac{5}{4}}}\left|\sum_{t=1}^{T} y_{i,t-1}\right| \ge \frac{\sqrt{2M}}{2}\right) \\
&\le \underbrace{n \max_{i\in\mathcal{I}_n} \Pr\left(\frac{1}{T^{\frac{3}{4}}}\left|\sum_{t=1}^{T} u_{i,t}\right| \ge \frac{\sqrt{2M}}{2}\right)}_{\text{(B.4)}} + \underbrace{n \max_{i\in\mathcal{I}_n} \Pr\left(\frac{1}{T^{\frac{5}{4}}}\left|\sum_{t=1}^{T} y_{i,t-1}\right| \ge \frac{\sqrt{2M}}{2}\right)}_{\text{(B.5)}}.
\end{aligned}
$$

For the term (B.4), similar to the derivation of the order of (A.5), we have

$$
\begin{aligned}
& n \max_{i\in\mathcal{I}_n} \Pr\left(\frac{1}{T^{\frac{3}{4}}}\left|\sum_{t=1}^{T} u_{it}\right| \ge \frac{\sqrt{2M}}{2}\right) \\
&\le n \max_{i\in\mathcal{I}_n} \Pr\left(\frac{1}{T^{\frac{3}{4}}}\left|F_i(1)\sum_{t=1}^{T} \epsilon_{i,t}\right| \ge \frac{\sqrt{2M}}{4}\right) + n \max_{i\in\mathcal{I}_n} \Pr\left(\frac{1}{T^{\frac{3}{4}}}\left|\widetilde{\epsilon}_{i,0} - \widetilde{\epsilon}_{i,T}\right| \ge \frac{\sqrt{2M}}{4}\right) \\
&= o(1),
\end{aligned}
$$

where the rate restrictions $T^{\frac{3}{2}} \succ n$ and $T \succ n(\log n)^2$ are assured by Assumption 2. For (B.5), by the Markov inequality, we have

$$
n \max_{i\in\mathcal{I}_n} \Pr\left(\frac{1}{T^{\frac{5}{4}}}\left|\sum_{t=1}^{T} y_{i,t-1}\right| \ge \frac{\sqrt{2M}}{2}\right) \le n \max_{i\in\mathcal{I}_n} \frac{2\mathbb{E}\left(\sum_{t=1}^{T} y_{i,t-1}\right)^2}{MT^{\frac{5}{2}}} \le O\left(T^{\gamma-\frac{3}{2}} n\right) = o(1),
$$

where the rate restriction $T^{\frac{3-2\gamma}{2}} \succ n$ is assured by Assumption 2. Therefore,

$$
n \max_{i\in\mathcal{I}_n} \Pr\left(\frac{1}{T^2}\left|\left(\sum_{t=1}^{T} y_{i,t-1}\right)\left(\sum_{t=1}^{T} u_{it}\right)\right| \ge \frac{M}{2}\right) = o(1). \tag{35}
$$

Finally, based on (34) and (35),

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

The proof is then complete. ■

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

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

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

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

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

*(iii) if* $\bar{c}_i^0 < 0$ *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 $\bar{c}_i^0 > 0$, we decompose $\frac{1}{\left(\overline{\rho}_i^0\right)^{2T}T^{2\gamma}\left(\log T\right)^2}\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2\right|$ as

$$\begin{aligned}
& n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{\left(\overline{\rho}_i^0\right)^{2T}T^{2\gamma}\left(\log T\right)^2}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2 \geq \widetilde{M}_1\right) \\
& \leq n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{\left(\overline{\rho}_i^0\right)^{2T}T^{2\gamma}\left(\log T\right)^2}\sum_{t=1}^{T}y_{i,t-1}^2 \geq \frac{\widetilde{M}_1}{2}\right) \\
& + n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{\left(\overline{\rho}_i^0\right)^{2T}T^{2\gamma+1}\left(\log T\right)^2}\left(\sum_{t=1}^{T}y_{i,t-1}\right)^2 \geq \frac{\widetilde{M}_1}{2}\right) \\
& \leq \underbrace{n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{2\bar{c}_i^0\left(\overline{\rho}_i^0\right)^{2T}T^{\gamma}\left(\log T\right)^2}\left|y_{i,0}^2\right| \geq \frac{\widetilde{M}_1}{10}\right)}_{\text{(C.1)}} \\
& + \underbrace{n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{2\bar{c}_i^0\left(\overline{\rho}_i^0\right)^{2T}T^{\gamma}\left(\log T\right)^2}\left(y_{i,T}^2 - \mathbb{E}y_{i,T}^2\right) \geq \frac{\widetilde{M}_1}{10}\right)}_{\text{(C.2)}} \\
& + \underbrace{n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{\bar{c}_i^0\left(\overline{\rho}_i^0\right)^{2T-1}T^{\gamma}\left(\log T\right)^2}\left|\sum_{t=1}^{T}y_{i,t-1}u_{it}\right| \geq \frac{\widetilde{M}_1}{10}\right)}_{\text{(C.3)}} \\
& + \underbrace{n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{2\bar{c}_i^0\left(\overline{\rho}_i^0\right)^{2T}T^{\gamma}\left(\log T\right)^2}\left|\sum_{t=1}^{T}u_{it}^2\right| \geq \frac{\widetilde{M}_1}{10}\right)}_{\text{(C.4)}}
\end{aligned}$$

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

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

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

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

$$n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{2\bar{c}_i^0\left(\bar{\rho}_i^0\right)^{2T}T^{\gamma}(\log T)^2}\left|\sum_{t=1}^{T}u_{it}^2\right|\geq\frac{\widetilde{M}_1}{10}\right)\leq n\max_{i\in\mathcal{I}_n}\frac{10\cdot\mathbb{E}\left(\sum_{t=1}^{T}\left(u_{it}^2\right)\right)}{2\left(\bar{c}_i^0\right)\left(\bar{\rho}_i^0\right)^{2T}T^{\gamma}(\log T)^2\widetilde{M}_1}$$
$$=O\left(\max_{i\in\mathcal{I}_n}\frac{nT}{\left(\bar{\rho}_i^0\right)^{2T}T^{\gamma}(\log T)^2}\right)=O\left(\frac{nT^{1-\gamma}}{(\rho_{low})^{2T}(\log T)^2}\right)=o(1),\tag{36}$$

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

$$n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{2\bar{c}_i^0\left(\bar{\rho}_i^0\right)^{2T}T^{\gamma}(\log T)^2}\left|y_{i,0}^2\right|\geq\frac{\widetilde{M}_1}{10}\right)\leq n\max_{i\in\mathcal{I}_n}\frac{10\mathbb{E}y_{i,0}^2}{2\left(\bar{c}_i^0\right)\left(\bar{\rho}_i^0\right)^{2T}T^{\gamma}(\log T)^2\left(\widetilde{M}_1\right)}$$
$$\leq O\left(\max_{i\in\mathcal{I}_n}\frac{n}{\left(\bar{\rho}_i^0\right)^{2T}T^{\gamma}(\log T)^2}\right)=O\left(\frac{n}{(\rho_{low})^{2T}T^{\gamma}(\log T)^2}\right)=o(1),\tag{37}$$

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 $\chi^2$ variates (Laurent and Massart, 2000), we have

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

with constant $K$, since

$$\Pr\left(\max_{1\leq t\leq T}\left(y_{i,T}^2-\mathbb{E}y_{i,T}^2\right)\geq\sqrt{2M_T\mathbb{E}y_{i,T}^2}+2M_T\right)$$
$$\leq T\max_{1\leq t\leq T}\Pr\left(\left(y_{i,T}^2-\mathbb{E}y_{i,T}^2\right)\geq\sqrt{2M_T\mathbb{E}y_{i,T}^2}+2M_T\right)$$

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

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

$$n \max_{i\in\mathcal{I}_n} \Pr\left( \frac{1}{2\overline{c}_i^0 \left(\overline{\rho}_i^0\right)^{2T} T^\gamma (\log T)^2} \left(y_{i,T}^2 - \mathbb{E}y_{i,T}^2\right) \ge \frac{\widetilde{M}_1}{10} \right) = 0. \tag{38}$$

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

$$\begin{aligned}
& n \max_{i\in\mathcal{I}_n} \Pr\left( \frac{1}{2\overline{c}_i^0 \left(\overline{\rho}_i^0\right)^{2T} T^{2\gamma+1} (\log T)^2} \left( \sum_{t=1}^T y_{i,t-1} \right)^2 \ge \frac{\widetilde{M}_1}{6} \right) \\
\le\ & n \max_{i\in\mathcal{I}_n} \Pr\left( \frac{1}{\left(\overline{\rho}_i^0\right)^{T} T^{\gamma+\frac{1}{2}} (\log T)} \left| \sum_{t=1}^T y_{i,t-1} \right| \ge \frac{\sqrt{3\overline{c}_i^0 \widetilde{M}_1}}{3} \right) \\
\le\ & n \max_{i\in\mathcal{I}_n} \Pr\left( \frac{1}{\left(\overline{\rho}_i^0\right)^{T} T^{\gamma+\frac{1}{2}} (\log T)} \left| \frac{T^\gamma}{\overline{c}_i^0} \sum_{t=1}^{T-1} \left( \left(\overline{\rho}_i^0\right)^{T-t} - 1 \right) u_{it} + \sum_{t=1}^T \left(\overline{\rho}_i^0\right)^{t-1} y_{i0} \right| \ge \frac{\sqrt{3\overline{c}_i^0 \widetilde{M}_1}}{3} \right) \\
\le\ & n \max_{i\in\mathcal{I}_n} \Pr\left( \frac{1}{\overline{c}_i^0 T^{\frac{1}{2}} (\log T)} \left| \sum_{t=1}^{T-1} \left(\overline{\rho}_i^0\right)^{-t} u_{it} \right| \ge \frac{\sqrt{3\overline{c}_i^0 \widetilde{M}_1}}{6} \right) \\
& + n \max_{i\in\mathcal{I}_n} \Pr\left( \frac{1}{\left(\overline{\rho}_i^0\right)^{T} T^{\gamma+\frac{1}{2}} (\log T)} \left| \sum_{t=1}^T \left(\overline{\rho}_i^0\right)^{t-1} y_{i0} \right| \ge \frac{\sqrt{3\overline{c}_i^0 \widetilde{M}_1}}{12} \right) \\
& + n \max_{i\in\mathcal{I}_n} \Pr\left( \frac{1}{\overline{c}_i^0 \left(\overline{\rho}_i^0\right)^{T} T^{\frac{1}{2}} (\log T)} \left| \sum_{t=1}^{T-1} u_{it} \right| \ge \frac{\sqrt{3\overline{c}_i^0 \widetilde{M}_1}}{12} \right) \\
=\ & n \max_{i\in\mathcal{I}_n} \Pr\left( \frac{1}{\left(\overline{c}_i^0\right) \sqrt{T} (\log T)} \left| \sum_{t=1}^{T-1} \left(\overline{\rho}_i^0\right)^{-t} u_{it} \right| \ge \frac{\sqrt{3\overline{c}_i^0 \widetilde{M}_1}}{6} \right) + o(1) = o(1),
\end{aligned} \tag{39}$$

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

$$n \max_{i\in\mathcal{I}_n} \Pr\left( \frac{1}{\overline{c}_i^0 \sqrt{T} (\log T)} \left| \sum_{t=1}^{T-1} \left(\overline{\rho}_i^0\right)^{-t} u_{it} \right| \ge \frac{\sqrt{3\overline{c}_i^0 \widetilde{M}_1}}{6} \right)$$

$$\leq n \max_{i\in\mathcal{I}_n} \frac{K\mathbb{E}\left(\sum_{t=1}^{T-1}\left(\overline{\rho}_i^0\right)^{-t} u_{it}\right)^2}{\widetilde{M}_1 T (\log T)^2} = O\left(\frac{n}{T^{1-\gamma}(\log T)^2}\right),$$

where $K$ denotes some constant value,

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

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\overline{c}_i^0\left(\overline{\rho}_i^0\right)^{2T} T^\gamma(\log T)^2}\left|T\mu_i^2 + 2\mu_i\overline{\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\overline{c}_i^0\left(\overline{\rho}_i^0\right)^{2T} T^\gamma(\log T)^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\overline{c}_i^0\left(\overline{\rho}_i^0\right)^{2T-1} T^\gamma(\log T)^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\overline{c}_i^0\left(\overline{\rho}_i^0\right)^{2T} T^\gamma(\log T)^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\overline{c}_i^0\left(\overline{\rho}_i^0\right)^{2T} T^\gamma(\log T)^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(\overline{c}_i^0\right)^2\left(\overline{\rho}_i^0\right)^{T-1}(\log T)^2}\left|2\mu_i\sum_{t=1}^{T-1}\left(\overline{\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(\overline{c}_i^0\right)^2\left(\overline{\rho}_i^0\right)^{2T-1}(\log T)^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(\overline{c}_i^0\right)^2\left(\overline{\rho}_i^0\right)^{2T-1}(\log T)^2}\left|2\mu_i\sum_{t=1}^{T}\left(\overline{\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\overline{c}_i^0\left(\overline{\rho}_i^0\right)^{2T} T^\gamma(\log T)^2}\left|2\mu_i\sum_{t=1}^T u_{it}\right| \geq \frac{\widetilde{M}_1}{60}\right) = o(1), \qquad (40)
\end{aligned}$$

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

$$n \max_{i\in\mathcal{I}_n} \Pr\left(\frac{1}{2\overline{c}_i^0 \left(\overline{\rho}_i^0\right)^{2T} T^{\gamma} (\log T)^2} \mathbb{E} y_{i,T}^2 > \frac{\widetilde{M}_1}{3}\right) = 0. \tag{41}$$

Combining the results of (36)–(41) 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,

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

With the above uniform upper bound, it follows that

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

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:

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

For the term (D.1), we have

$$n \max_{i\in\mathcal{I}_n} \Pr\left(\frac{1}{-2\overline{c}_i^0 T} \left|\sum_{t=1}^{T} \mathbb{E} u_{it}^2\right| \geq \frac{\widetilde{M}_3}{3}\right) = 0, \tag{42}$$

where $\widetilde{M}_3 \geq \frac{\max_{i\in\mathcal{I}_n} 2\left(\overline{\sigma}_{iu}^0\right)^2}{c_{low}}$ and

$$\frac{1}{-2\overline{c}_i^0 T}\sum_{t=1}^{T}\mathbb{E}u_{it}^2 = -\frac{\left(\overline{\sigma}_{iu}^0\right)^2}{2\overline{c}_i^0} + o_p\left(\frac{1}{T^{\frac{\gamma}{2}-\frac{1}{q}}}\right),$$

by Phillips and Magdalinos (2007b). For the term (D.2), the decomposition for the difference between the sample moment and the population moment can be applied as

$$\begin{aligned}
&n\max_{i\in\mathcal{I}_n}\Pr\left(\left|\frac{1}{T^{1+\gamma}}\sum_{t=1}^{T}y_{i,t-1}^2 - \frac{1}{-2\overline{c}_i^0 T}\sum_{t=1}^{T}\mathbb{E}u_{it}^2\right| \geq \frac{\widetilde{M}_3}{3}\right)\\
&\leq \underbrace{n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{-2\overline{c}_i^0 T}\left|y_{i,0}^2\right| \geq \frac{\widetilde{M}_3}{15}\right)}_{\text{(D.2.1)}} + \underbrace{n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{-2\overline{c}_i^0 T}\left|y_{i,T}^2\right| \geq \frac{\widetilde{M}_3}{15}\right)}_{\text{(D.2.2)}}\\
&+\underbrace{n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{-\overline{c}_i^0 T}\left|\sum_{t=1}^{T}y_{i,t-1}u_{it}\right| \geq \frac{\widetilde{M}_3}{15}\right)}_{\text{(D.2.3)}} + \underbrace{n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{-2\overline{c}_i^0 T}\left|\sum_{t=1}^{T}\left(u_{it}^2 - \mathbb{E}u_{it}^2\right)\right| \geq \frac{\widetilde{M}_3}{15}\right)}_{\text{(D.2.4)}}\\
&+\underbrace{n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{-2\overline{c}_i^0 T}\left|T\mu_i^2 + 2\mu_i\overline{\rho}_i^0\sum_{t=1}^{T}y_{i,t-1} + 2\mu_i\sum_{t=1}^{T}u_{it}\right| \geq \frac{\widetilde{M}_3}{15}\right)}_{\text{(D.2.5)}}.
\end{aligned}$$

For (D.2.1) and (D.2.2), we have

$$n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{-2\overline{c}_i^0 T}\left|y_{i,0}^2\right| \geq \frac{\widetilde{M}_3}{15}\right) \leq n\max_{i\in\mathcal{I}_n}\frac{15\mathbb{E}\left(y_{i,0}^2\right)}{-2\left(\overline{c}_i^0\right)T\left(\widetilde{M}_3\right)} = o\left(\frac{n}{T}\right) = o(1),$$

and

$$n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{-2\overline{c}_i^0 T}\left|y_{i,T}^2\right| \geq \frac{\widetilde{M}_3}{15}\right) \leq n\max_{i\in\mathcal{I}_n}\frac{15\mathbb{E}\left(y_{i,T}^2\right)}{-2\left(\overline{c}_i^0\right)T\left(\widetilde{M}_3\right)} = O\left(\frac{nT^{\gamma}}{T}\right) = o(1),$$

where the rate restriction $T^{1-\gamma} > n$ is assured by Assumption 2. For (D.2.3), the corresponding asymptotic negligibility is obtained in Lemma A.3. For (D.2.4), we have

$$\begin{aligned}
&n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{-2\overline{c}_i^0 T}\left|\sum_{t=1}^{T}\left(u_{it}^2 - \mathbb{E}u_{it}^2\right)\right| \geq \frac{\widetilde{M}_3}{15}\right)\\
&\leq n\max_{i\in\mathcal{I}_n}\frac{225\mathbb{E}\left(\sum_{t=1}^{T}\left(u_{it}^2 - \mathbb{E}u_{it}^2\right)\right)^2}{4\left(\overline{c}_i^0\right)^2 T^2\left(\widetilde{M}_3\right)^2} = O\left(\frac{n}{T}\right) = o(1),
\end{aligned}$$

where $\mathbb{E}\left(\sum_{t=1}^{T}\left(u_{it}^2-\mathbb{E}u_{it}^2\right)\right)^2 \sim \mathbb{E}\left(\sum_{t=1}^{T}\left(F_i^2(1)\epsilon_{it}^2-\mathbb{E}F_i^2(1)\epsilon_{it}^2\right)^2\right)=O(T)$. For the term (D.2.5), we have

$$\begin{aligned}
& n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{-2\overline{c}_i^0 T}\left|T\mu_i^2+2\mu_i\overline{\rho}_i^0\sum_{t=1}^{T}y_{i,t-1}+2\mu_i\sum_{t=1}^{T}u_{it}\right|\geq\frac{\widetilde{M}_3}{15}\right)\\
&\leq n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{-2\overline{c}_i^0 T}\left|T\mu_i^2\right|\geq\frac{\widetilde{M}_3}{45}\right)+n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{-2\overline{c}_i^0 T}\left|2\mu_i\sum_{t=1}^{T}y_{i,t-1}\right|\geq\frac{\widetilde{M}_3}{45}\right)\\
&+n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{-2\overline{c}_i^0 T}\left|2\mu_i\sum_{t=1}^{T}u_{it}\right|\geq\frac{\widetilde{M}_3}{45}\right)\\
&\leq O\left(\frac{n}{T^5}\right)+O\left(\frac{n}{T^{3-2\gamma}}\right)+O\left(\frac{n}{T^3}\right)=o(1),
\end{aligned}$$

by the Markov inequality and Assumption 2. Based on above results, we can show the tail behavior of the term (D.2) as

$$n\max_{i\in\mathcal{I}_n}\Pr\left(\left|\frac{1}{T^{1+\gamma}}\sum_{t=1}^{T}y_{i,t-1}^2-\frac{1}{-2\overline{c}_i^0 T}\sum_{t=1}^{T}u_{it}^2\right|\geq\frac{\widetilde{M}_3}{3}\right)=o(1). \tag{43}$$

Finally, for the term (D.3), we have

$$\begin{aligned}
& n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{T^{2+\gamma}}\left(\sum_{t=1}^{T}y_{i,t-1}\right)^2\geq\frac{\widetilde{M}_3}{3}\right)\\
&\leq n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{T^{1+\frac{\gamma}{2}}}\left|\sum_{t=1}^{T}y_{i,t-1}\right|\geq\frac{\sqrt{3\widetilde{M}_3}}{3}\right)\\
&\leq n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{-\overline{c}_i^0 T^{1-\frac{\gamma}{2}}}\left|\sum_{t=1}^{T}u_{it}\right|\geq\frac{\sqrt{3\widetilde{M}_3}}{6}\right)\\
&+n\max_{i\in\mathcal{I}_n}\Pr\left(\frac{1}{-\overline{c}_i^0 T^{1-\frac{\gamma}{2}}}\left|\sum_{t=1}^{T}\left(\overline{\rho}_i^0\right)^{T-t}u_{it}\right|\geq\frac{\sqrt{3\widetilde{M}_3}}{6}\right)\\
&\leq O\left(\frac{nT}{T^{2-\gamma}}\right)+O\left(\frac{nT^{\gamma}}{T^{2-\gamma}}\right)=o(1),
\end{aligned} \tag{44}$$

where the rate restrictions $T^{1-\gamma}>n$ and $T^{2-2\gamma}>n$ are again assured by Assumption 2.

Combining the results in (42) (43) and (44), 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),

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

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

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

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

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

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

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

where $Z_{ij}(c) = \mathbf{1}\left\{g_i^0 \neq j\right\}\mathbf{1}\left\{\sum_{t=1}^{T}(\widetilde{y}_{it} - \widetilde{y}_{i,t-1}\rho_j)^2 \leq \sum_{t=1}^{T}(\widetilde{y}_{it} - \widetilde{y}_{i,t-1}\rho_{g_i^0})^2\right\}$, $\rho_j = 1 + c_j/T^\gamma$ 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 $i \in \mathcal{I}_n$,

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

Define

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

$$=: H_{1T} + H_{2T},$$

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

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

By compactness of the parameter space and the definition of $\eta$,

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

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

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

Combining the above results, we obtain

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

Since

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

we define

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

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

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

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

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

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

$$+\sum_{\substack{\widetilde{j}\neq j,\\ c^0_{\widetilde{j}}>0}} \Pr\left(\frac{1}{\left(\rho^0_{\widetilde{j}}\right)^{2T} T^{2\gamma}(\log T)^2}\left|\sum_{t=1}^{T}\widetilde{y}^2_{i,t-1}\right| \geq \widetilde{M}_1\right). \tag{45}$$

By Lemmas A.3, A.4 and A.5, the second, third, and fourth terms in (45) are all $o\left(\frac{1}{n}\right)$. We can bound $\eta$ 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

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

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

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

$$\leq\sum_{\substack{\widetilde{j}\neq j,\\ c^0_{\widetilde{j}}>0}} \Pr\left(\frac{(\log T)^2}{T^{\gamma}\left(\rho^0_{\widetilde{j}}\right)^{2T}}\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right| \geq \frac{\dot{c}\overline{M}_1}{4}\right) = o\left(\frac{1}{n}\right). \tag{46}$$

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 $j,\widetilde{j}=1,2,\ldots,G^0$ and $g^0_i=\widetilde{j}$, we have

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

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

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

$$+B_2T^{\gamma}\eta\left|\sum_{t=1}^{T}\widetilde{y}^2_{i,t-1}\right|\Bigg)$$

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

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

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

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

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

$$+ \sum_{\substack{\widetilde{j}\neq j,\\ c^0_{\widetilde{j}}=0}} \Pr\left(\frac{1}{T^2(\log_2 T)^2}\left|\sum_{t=1}^{T}\widetilde{y}^2_{i,t-1}\right| \geq \widetilde{M}_2\right). \tag{47}$$

Similarly, we can bound $\eta$ 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

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

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

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

$$\leq \sum_{\substack{\widetilde{j}\neq j,\\ c^0_{\widetilde{j}}=0}} \Pr\left(\frac{(\log_2 T)^2}{T^{2-\gamma}}\left|\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right| \geq \frac{\dot{c}\overline{M}_2}{4}\right) = o\left(\frac{1}{n}\right). \tag{48}$$

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

$$
\begin{aligned}
&\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\Bigg(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| \\
&\qquad\qquad + B_2 T^\gamma \eta \left|\sum_{t=1}^{T} \widetilde{y}^2_{i,t-1}\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 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).
\end{aligned}
$$

We can bound $\eta$ 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

$$\begin{aligned}
&\sum_{\substack{\widetilde{j}\neq j,\\ c^0_{\widetilde{j}}<0}}\Pr\left(2\dot{c}\frac{1}{T}\sum_{t=1}^T\widetilde{y}_{i,t-1}\widetilde{u}_{it}\leq -(\dot{c})^2\overline{M}_3+(B_1)\eta M_3+B_2T^\gamma\eta\widetilde{M}_3\right)\\
&\leq\sum_{\substack{\widetilde{j}\neq j,\\ c^0_{\widetilde{j}}<0}}\Pr\left(2\frac{\dot{c}}{T}\sum_{t=1}^T\widetilde{y}_{i,t-1}\widetilde{u}_{it}\leq\frac{-(\dot{c})^2\overline{M}_3}{2}\right)\\
&\leq\sum_{\substack{\widetilde{j}\neq j,\\ c^0_{\widetilde{j}}<0}}\Pr\left(\frac{1}{T}\left|\sum_{t=1}^T\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right|\geq\frac{\dot{c}\overline{M}_3}{4}\right)=o\left(\frac{1}{n}\right). \qquad (49)
\end{aligned}$$

The last equality is due to Lemma A.3.

Combining the results in (46), (48) and (49), we obtain $\Pr\left(\widetilde{Z}_{ij}=1\right)=o\left(\frac{1}{n}\right)$, which in turn implies that

$$\begin{aligned}
\mathbb{E}\left[\sup_{c\in\mathcal{N}_\eta}\frac{1}{n}\sum_{i=1}^n\mathbf{1}\left\{\widehat{g}_i(c)\neq g_i^0\right\}\right]&\leq\frac{1}{n}\sum_{j=1}^{G^0}\sum_{i=1}^n\mathbb{E}\widetilde{Z}_{ij}=\frac{1}{n}\sum_{j=1}^{G^0}\sum_{i=1}^n\Pr\left(\widetilde{Z}_{ij}=1\right)\\
&=G^0\left(G^0-1\right)o\left(\frac{1}{n}\right)=o\left(\frac{1}{n}\right), \qquad (50)
\end{aligned}$$

where $\eta=O\left(\frac{1}{(\log T)^4T^\gamma}\right)$. Weak convergence holds due to the Markov inequality and (50). 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

$$\widehat{E}_{j,i}:=\left\{\widehat{g}_i\neq j|g_i^0=j\right\}\text{ and }\widehat{F}_{j,i}:=\left\{g_i^0\neq j|\widehat{g}_i=j\right\}, \qquad (51)$$

for any $j=1,2,\ldots,G^0$ 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)* $\Pr\left(\bigcup_{j=1}^{G^0}\widehat{E}_{j,nT}\right)\leq\sum_{j=1}^{G^0}\Pr\left(\widehat{E}_{j,nT}\right)\to 0$;

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

**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

$$\sum_{j=1}^{G^0}\sum_{i\in\mathcal{G}^0(j)}\Pr\left(\widehat{E}_{j,i}\right)\leq n\max_{i\in\mathcal{I}_n}\mathbb{E}\mathbf{1}\{\widehat{g}_i(\widehat{c}^*)\neq g_i^0\}\leq n\max_{i\in\mathcal{I}_n}\Pr\{\left|\widehat{g}_i(\widehat{c}^*)-g_i^0\right|>0\}$$
$$\leq n\max_{i\in\mathcal{I}_n}\sup_{c\in\mathcal{N}_\eta}\Pr\{\left|\widehat{g}_i(c)-g_i^0\right|>0\}+n\max_{1\leq j\leq G^0}\Pr\{\left|\widehat{c}_j^*-c_j^0\right|>\eta\}$$
$$=o(1)+n\max_{1\leq j\leq G^0}\Pr\{\left|\widehat{c}_j^*-c_j^0\right|>\eta\}=o(1), \tag{52}$$

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 $j=1,2,...,G^0$ with $c_j^0\geq 0$.*

**Proof of Lemma B.1:** Without losing generality we only prove consistency of $\check{\omega}_j^2$ under the joint convergence framework $(n,T)\to\infty$. Similar arguments give the results for $\check{\sigma}_j^2$ and $\check{\lambda}_j$. By virtue of the uniform consistency of the clustering algorithm we show that

$$\frac{1}{\check{n}_j}\sum_{i\in\hat{\mathcal{G}}(j)}\widehat{\overline{\omega}}_i^2=\frac{1}{n_j}\sum_{i\in\mathcal{G}^0(j)}\widehat{\overline{\omega}}_i^2+o_p(1).$$

The oracle estimates $\widehat{\omega}_j^2$, $\widehat{\lambda}_j$, and $\widehat{\sigma}_j^2$ employ the true group identities and are immune to clustering errors, for any $j=1,2,...,G^0$. For any $i\in\mathcal{I}_n$, we have

$$\left|\left(\overline{\omega}_i^0\right)^2-\widehat{\overline{\omega}}_i^2\right|\leq\left|\widehat{\overline{\omega}}_i^2-\left[\frac{1}{T}\sum_{t=1}^{T}u_{it}^2+\frac{1}{T}\sum_{l=1}^{L}\sum_{t=l+1}^{T}w(l,L)\left(u_{it}u_{i,t-l}+u_{i,t-l}u_{it}\right)\right]\right|$$
$$+\left|\frac{1}{T}\sum_{t=1}^{T}\left(u_{it}^2-\mathbb{E}u_{it}^2\right)+\frac{2}{T}\sum_{l=1}^{L}\sum_{t=l+1}^{T}w(l,L)\left(u_{it}u_{i,t-l}-\mathbb{E}u_{it}u_{i,t-l}\right)\right|$$

$$
\begin{aligned}
&+\left|\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}u_{it}^2+\frac{2}{T}\sum_{l=1}^{L}\sum_{t=l+1}^{T}w(l,L)\,\mathbb{E}u_{it}u_{i,t-l}-\left(\overline{\omega}_i^0\right)^2\right| \\
&\le \underbrace{\left|\widehat{\overline{\omega}}_i^2-\left[\frac{1}{T}\sum_{t=1}^{T}u_{it}^2+\frac{1}{T}\sum_{l=1}^{L}\sum_{t=l+1}^{T}w(l,L)\left(u_{it}u_{i,t-l}+u_{i,t-l}u_{it}\right)\right]\right|}_{\text{(F.1)}} \\
&+\underbrace{\left|\frac{1}{T}\sum_{t=1}^{T}\left(u_{it}^2-\mathbb{E}u_{it}^2\right)+\frac{2}{T}\sum_{l=1}^{L}\sum_{t=l+1}^{T}w(l,L)\left(u_{it}u_{i,t-l}-\mathbb{E}u_{it}u_{i,t-l}\right)\right|}_{\text{(F.2)}} \\
&+\underbrace{\frac{2}{T}\sum_{l=1}^{L}|w(l,L)-1|\sum_{t=l+1}^{T}\left|\mathbb{E}u_{it}u_{i,t-l}\right|}_{\text{(F.3)}}+\underbrace{\frac{2}{T}\sum_{l=L+1}^{T-1}\sum_{t=l+1}^{T}\left|\mathbb{E}u_{it}u_{i,t-l}\right|}_{\text{(F.4)}}.
\end{aligned}
$$

For the term (F.4),

$$
\frac{2}{T}\sum_{l=L+1}^{T-1}\sum_{t=l+1}^{T}\left|\mathbb{E}u_{it}u_{i,t-l}\right|\to 0,\ \text{ as } T\to\infty,\ \ L\to\infty\ \text{ and }\ L=o(T).
$$

For the term (F.3), the condition that

$$
\frac{1}{T}\sum_{l=1}^{L}\sum_{t=l+1}^{T}\left|\mathbb{E}u_{it}u_{i,t-l}\right|=O_p(1),
$$

is needed to derive asymptotic negligibility. By Phillips and Solo (1992),

$$
\frac{1}{T}\sum_{l=1}^{L}\sum_{t=l+1}^{T}\left|\mathbb{E}u_{it}u_{i,t-l}\right|\le\frac{1}{T}\sum_{l=1}^{L}\sum_{t=1}^{T}\left|\mathbb{E}u_{it}u_{i,t-l}\right|\le\sum_{l=1}^{\infty}\left|\mathbb{E}u_{it}u_{i,t-l}\right|<\infty.
$$

Moreover, $\lim_{T\to\infty}w(l,L)=1$ for any $l$, so that (F.3) diminishes. For the term (F.2), Theorem 3.7 of Phillips and Solo (1992) ensures that $\frac{1}{T}\sum_{t=1}^{T}\left(u_{it}^2-\mathbb{E}u_{it}^2\right)=o_p(1)$ for any $i\in\mathcal{I}_n$. Note that $\sum_{l=1}^{L}|w(l,L)|\le L$. By the Markov inequality, for any $\varepsilon>0$,

$$
\begin{aligned}
&\Pr\left(\left|\frac{1}{T}\sum_{l=1}^{L}\sum_{t=l+1}^{T}w(l,L)\left(u_{it}u_{i,t-l}-\mathbb{E}u_{it}u_{i,t-l}\right)\right|>\varepsilon\right) \\
&\le\Pr\left(\sum_{l=1}^{L}|w(l,L)|\left|\frac{1}{T}\sum_{t=l+1}^{T}\left(u_{it}u_{i,t-l}-\mathbb{E}u_{it}u_{i,t-l}\right)\right|>\varepsilon\right) \\
&\le\sum_{l=1}^{L}\Pr\left(\left|\frac{1}{T}\sum_{t=l+1}^{T}\left(u_{it}u_{i,t-l}-\mathbb{E}u_{it}u_{i,t-l}\right)\right|>\frac{\varepsilon}{L}\right) \\
&\le\sum_{l=1}^{L}\left(\frac{L}{\varepsilon}\right)^2\mathbb{E}\left[\sum_{t=l+1}^{T}\left(u_{it}u_{i,t-l}-\mathbb{E}u_{it}u_{i,t-l}\right)\right]^2\frac{1}{T^2}
\end{aligned}
$$

$$\leq \sum_{l=1}^{L}\left(\frac{L}{\varepsilon}\right)^2 (T-L)\frac{1}{T^2}Var\left(F_i^2(1)\epsilon_{it}\epsilon_{i,t-l} - F_i^2(1)\mathbb{E}\epsilon_{it}\epsilon_{i,t-l}\right).$$

Furthermore,

$$Var\left(F_i^2(1)\epsilon_{it}\epsilon_{i,t-l} - F_i^2(1)\mathbb{E}\epsilon_{it}\epsilon_{i,t-l}\right) = O(1).$$

By Assumption 2(iii), we have

$$\begin{aligned}
&\sum_{l=1}^{L}\left(\frac{L}{\varepsilon}\right)^2 (T-L)\frac{1}{T^2}Var\left(F_i^2(1)\epsilon_{it}\epsilon_{i,t-l} - F_i^2(1)\mathbb{E}\epsilon_{it}\epsilon_{i,t-l}\right) \\
&\leq \sum_{l=1}^{L}\left(\frac{L}{\varepsilon}\right)^2 \frac{1}{T}Var\left(F_i^2(1)\epsilon_{it}\epsilon_{i,t-l} - F_i^2(1)\mathbb{E}\epsilon_{it}\epsilon_{i,t-l}\right) \\
&= O\left(\frac{L^3}{T}\right) = o(1).
\end{aligned}$$

For the term (F.1), we have the following decomposition:

$$\begin{aligned}
&\left|\widehat{\omega}_i^2 - \left[\frac{1}{T}\sum_{t=1}^{T}u_{it}^2 + \frac{1}{T}\sum_{l=1}^{L}\sum_{t=l+1}^{T}w(l,L)(u_{it}u_{i,t-l} + u_{i,t-l}u_{it})\right]\right| \\
&= \left|\frac{1}{T}\sum_{t=1}^{T}\left(\widehat{u}_{it}^2 - u_{it}^2\right) + \frac{2}{T}\sum_{l=1}^{L}\sum_{t=l+1}^{T}w(l,L)(u_{it}u_{i,t-l} - \widehat{u}_{it}\widehat{u}_{i,t-l})\right| \\
&\leq \left|\frac{1}{T}\sum_{t=1}^{T}\left(\widehat{u}_{it}^2 - u_{it}^2\right)\right| + \frac{2}{T}\sum_{l=1}^{L}|w(l,L)|\left|\sum_{t=l+1}^{T}(u_{it}u_{i,t-l} - \widehat{u}_{it}\widehat{u}_{i,t-l})\right|.
\end{aligned}$$

From the definition of $\widehat{u}_{it}$

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}\widehat{u}_{it}^2 = &\frac{1}{T}\sum_{t=1}^{T}u_{it}^2 + \frac{1}{T}\left(\overline{\rho}_i^0 - \widehat{\overline{\rho}}_i\right)^2\sum_{t=1}^{T}y_{i,t-1}^2 + (\mu_i)^2 + \frac{2}{T}(\mu_i)\sum_{t=1}^{T}u_{it} \\
&+ \frac{2}{T}(\mu_i)\sum_{t=1}^{T}y_{i,t-1}\left(\overline{\rho}_i^0 - \widehat{\overline{\rho}}_i\right) + \frac{2}{T}\left(\overline{\rho}_i^0 - \widehat{\overline{\rho}}_i\right)\sum_{t=1}^{T}y_{i,t-1}u_{it}, \qquad (53)
\end{aligned}$$

in which

$$\frac{1}{T}\left(\overline{\rho}_i^0 - \widehat{\overline{\rho}}_i\right)^2\sum_{t=1}^{T}y_{i,t-1}^2 = \begin{cases} O_p\left(\frac{1}{nT}\right) & \text{for} \quad \overline{c}_i^0 > 0 \\ O_p\left(\frac{1}{T}\right) & \text{for} \quad \overline{c}_i^0 = 0 \end{cases},$$

$$\frac{2}{T}(\mu_i)\sum_{t=1}^{T}u_{it} \leq \begin{cases} O_p\left(\frac{1}{T^{\frac{3}{2}}}\right) & \text{for} \quad \overline{c}_i^0 > 0 \\ O_p\left(\frac{1}{T^{\frac{3}{2}}}\right) & \text{for} \quad \overline{c}_i^0 = 0 \end{cases},$$

$$(\mu_i)^2 \leq \begin{cases} O_p\left(\frac{1}{T^2}\right) & \text{for} \quad \overline{c}_i^0 > 0 \\ O_p\left(\frac{1}{T^2}\right) & \text{for} \quad \overline{c}_i^0 = 0 \end{cases},$$

$$\frac{2}{T}(\mu_i)\sum_{t=1}^{T} y_{i,t-1}\left(\overline{\rho}_i^0 - \widehat{\overline{\rho}}_i\right) \leq \begin{cases} O_p\left(\frac{1}{\sqrt{n}T^{2-\frac{\gamma}{2}}}\right) & \text{for} \quad \overline{c}_i^0 > 0 \\ O_p\left(\frac{1}{T^{\frac{3}{2}}}\right) & \text{for} \quad \overline{c}_i^0 = 0 \end{cases},$$

$$\frac{2}{T}\left(\overline{\rho}_i^0 - \widehat{\overline{\rho}}_i\right)\sum_{t=1}^{T} y_{i,t-1}u_{it} = \begin{cases} O_p\left(\frac{1}{\sqrt{n}T}\right) & \text{for} \quad \overline{c}_i^0 > 0 \\ O_p\left(\frac{1}{T}\right) & \text{for} \quad \overline{c}_i^0 = 0 \end{cases}.$$

Therefore, $\frac{1}{T}\sum_{t=1}^{T}\widehat{u}_{it}^2 = \frac{1}{T}\sum_{t=1}^{T}u_{it}^2 + o_p\left(\frac{1}{L}\right)$. Similarly, we have $\frac{1}{T}\sum_{t=1}^{T}\widehat{u}_{it}\widehat{u}_{i,t-l} = \frac{1}{T}\sum_{t=1}^{T}u_{it}u_{i,t-l} + o_p\left(\frac{1}{L}\right)$. Hence, by consistency of the variance and covariance estimates, asymptotic negligibility of the term (F.1) is established, so that $\left(\overline{\omega}_i^0\right)^2 - \widehat{\overline{\omega}}_i^2 = o_p(1)$ for any $i \in \mathcal{I}_n$. As $T \to \infty$, $\widehat{\overline{\omega}}_i^2 - \left(\overline{\omega}_i^0\right)^2 = o_p(1)$ for any $i \in \mathcal{I}_n$. This completes the proof of Lemma B.1. ■

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

$$\frac{1}{n_j T^2}\sum_{i\in\mathcal{G}^0(j)}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2 \to_p \frac{\left(\omega_j^0\right)^2}{6}, \text{ if } c_j^0 = 0; \tag{54}$$

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

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

**Proof of Lemma B.2:** For this proof, we assume $g_i^0 = j$ with $j = 1, 2, \ldots, G^0$. 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 $c_j^0 = 0$,

$$\frac{1}{n_j T^2}\sum_{i\in\mathcal{G}^0(j)}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2 \overset{T\to\infty}{\Rightarrow} \frac{1}{n_j}\sum_{i\in\mathcal{G}^0(j)}\int_0^1 \widetilde{B}_i^2(r)\,dr \overset{n\to\infty}{\to_p} \mathbb{E}\int_0^1 \widetilde{B}_i^2(r)\,dr, \tag{55}$$

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

$$\begin{aligned}\mathbb{E}\int_0^1 \widetilde{B}_i^2(r)\,dr &= \int_0^1 \mathbb{E}B_i^2(r)\,dr - \int_0^1\int_0^1 \mathbb{E}(B_i(r)B_i(s))\,dr\,ds \\ &= \left(\omega_j^0\right)^2\int_0^1 r\,dr - \left(\omega_j^0\right)^2\int_0^1\int_0^1 (r\wedge s)\,dr\,ds = \frac{\left(\omega_j^0\right)^2}{6}.\end{aligned} \tag{56}$$

Combining (55) and (56), we have

$$\frac{1}{n_j T^2} \sum_{i \in \mathcal{G}^0(j)} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}^2 \to_p \frac{\left(\omega_j^0\right)^2}{6}, \tag{57}$$

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 (55), equation (57) 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

$$\frac{1}{n_j \left(\rho_j^0\right)^{2T} T^{2\gamma}} \sum_{i \in \mathcal{G}^0(j)} \sum_{t=1}^{T} y_{i,t-1}^2 \overset{T\to\infty}{\Rightarrow} \frac{1}{2c_j^0 n_j} \sum_{i \in \mathcal{G}^0(j)} \left(X_i^\star\right)^2 \overset{n\to\infty}{\to_p} \frac{1}{2c_j^0} \left(\frac{\left(\omega_j^0\right)^2}{2c_j^0}\right), \tag{58}$$

where $\left(Y_i^\star, X_i^\star\right)' \sim \mathcal{N}\left(\mathbf{0}_{2\times 1}, diag\left(\frac{\left(\omega_j^0\right)^2}{2c_j^0}, \frac{\left(\omega_j^0\right)^2}{2c_j^0}\right)\right)$. 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 (58), 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 $c_j^0 < 0$,

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

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

$$\frac{1}{n_j T^{1+\gamma}} \sum_{i\in\mathcal{G}^0(j)} \sum_{t=1}^{T} y_{i,t-1}^2 \overset{T\to\infty}{\Rightarrow} \frac{1}{n_j} \sum_{i\in\mathcal{G}^0(j)} \left(\frac{\left(\omega_j^0\right)^2}{-2c_j^0}\right) \overset{n\to\infty}{\to_p} \frac{\left(\omega_j^0\right)^2}{-2c_j^0}. \tag{59}$$

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 (59), the following joint convergence result holds:

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

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 $(n,T)\to\infty$,*

$$\frac{1}{\sqrt{n_j}T} \sum_{i\in\mathcal{G}^0(j)} \sum_{t=1}^{T} \left(\widetilde{y}_{i,t-1}\widetilde{u}_{i,t} - \widehat{\lambda}_j + \frac{\widehat{\omega}_j^2}{2}\right) \Rightarrow \mathcal{N}\left(0, \frac{\left(\omega_j^0\right)^4}{12}\right), \text{ if } c_j^0 = 0; \tag{60}$$

$$\frac{1}{\sqrt{n_j}T^{\gamma}(\rho_j^0)^T} \sum_{i\in\mathcal{G}^0(j)} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}\widetilde{u}_{it} \Rightarrow \mathcal{N}\left(0, \frac{\left(\omega_j^0\right)^4}{4\left(c_j^0\right)^2}\right), \text{ if } c_j^0 > 0; \tag{61}$$

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

*where*

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

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

$$\frac{1}{\sqrt{n_j}T} \sum_{i\in\mathcal{G}^0(j)} \sum_{t=1}^{T} \left(\widetilde{y}_{i,t-1}\widetilde{u}_{i,t} - \widehat{\lambda}_j + \frac{\widehat{\omega}_j^2}{2}\right)$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{n_j}} \sum_{i \in \mathcal{G}^0(j)} \left[ \frac{1}{T} \sum_{t=1}^{T} \left( y_{i,t-1} u_{i,t} - \widehat{\lambda}_j \right) - \left( \overline{y}_{i,-1} \overline{u}_i - \frac{\widehat{\omega}_j^2}{2} \right) \right] \\
&\overset{T\to\infty}{\Rightarrow} \frac{1}{\sqrt{n_j}} \sum_{i \in \mathcal{G}^0(j)} \left[ \int_0^1 B_i(r)\, dB_i(r) - \left( \int_0^1 B_i(r)\, dr \right) B_i(1) + \frac{\left(\omega_j^0\right)^2}{2} \right] \\
&=: \frac{1}{\sqrt{n_j}} \sum_{i \in \mathcal{G}^0(j)} \left( N_{1,i} - N_{2,i} \right), \qquad (62)
\end{aligned}
$$

in which

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

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

$$
\begin{aligned}
\frac{1}{\sqrt{n_j}} \sum_{i \in \mathcal{G}^0(j)} \left( N_{1,i} - N_{2,i} \right) &\overset{n\to\infty}{\Rightarrow} \mathcal{N}\left(0,\ \mathbb{E}N_{1,i}^2 + \mathbb{E}N_{2,i}^2 - 2\mathbb{E}N_{1,i}N_{2,i}\right) \\
&=_d \mathcal{N}\left(0,\ \frac{\left(\omega_j^0\right)^4}{12}\right), \qquad (63)
\end{aligned}
$$

where (63) holds in view of the following:

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

For term (a), we have

$$
\mathbb{E}\left(N_{1,i}^2\right) = \mathbb{E}\left[ \left( \int_0^1 B_i(r)\, dB_i(r) \right)^2 \right] = \left(\omega_j^0\right)^2 \mathbb{E}\left[ \int_0^1 B_i^2(r)\, dr \right] = \left(\omega_j^0\right)^4 \int_0^1 r\, dr = \frac{\left(\omega_j^0\right)^4}{2}. \qquad (64)
$$

For term (b), we have

$$
\begin{aligned}
&\mathbb{E}\left(N_{2,i}^2\right) \\
&= \mathbb{E}\left[ \left( \int_0^1 B_i(r)\, dr B_i(1) - \frac{\left(\omega_j^0\right)^2}{2} \right)^2 \right] \\
&= \mathbb{E}\left( \int_0^1 \int_0^1 B_i(r) B_i(s) B_i^2(1)\, dr\, ds \right) - \frac{\left(\omega_j^0\right)^4}{4}
\end{aligned}
$$

$$
\begin{aligned}
&= \left(\omega_j^0\right)^4 \int_0^1 \int_0^1 \left[3(r\wedge s)^2 + (r\wedge s)(1-r\wedge s) + 2(r\wedge s)|r-s|\right] drds - \frac{\left(\omega_j^0\right)^4}{4} \\
&= \frac{7}{12}\left(\omega_j^0\right)^4,
\end{aligned} \tag{65}
$$

in which

$$
\begin{aligned}
\int_0^1 \int_0^1 (r\wedge s)\, drds &= \int_0^1 \int_0^s r drds + \int_0^1 \int_s^1 s drds = \frac{1}{3}, \\
2\int_0^1 \int_0^1 (r\wedge s)^2 drds &= 2\int_0^1 \int_0^s r^2 drds + 2\int_0^1 \int_s^1 s^2 drds = \frac{1}{3},
\end{aligned}
$$

and

$$
\begin{aligned}
2\int_0^1 \int_0^1 (r\wedge s)|r-s|\, drds &= 2\int_0^1 \int_0^s r(s-r)\, drds + 2\int_0^1 \int_s^1 s(r-s)\, drds \\
&= 2\left(\int_0^1 r dr\right)^2 - 2\int_0^1 \int_0^1 (r\wedge s)^2 drds \\
&= \frac{1}{2} - \frac{1}{3} = \frac{1}{6}.
\end{aligned}
$$

For term (c), we have

$$
\begin{aligned}
&\mathbb{E}(N_{1,i}N_{2,i}) \\
&= \mathbb{E}\left[\left(\int_0^1 B_i(r)\, dB_i(r)\right)\left(\int_0^1 B_i(r)\, drB_i(1) - \frac{\left(\omega_j^0\right)^2}{2}\right)\right] \\
&= \mathbb{E}\left[\left(\frac{1}{2}B_i^2(1) - \frac{1}{2}\left(\omega_j^0\right)^2\right)\left(\int_0^1 B_i(r)\, drB_i(1) - \frac{\left(\omega_j^0\right)^2}{2}\right)\right] \\
&= \mathbb{E}\left[\frac{1}{2}\int_0^1 B_i(r) B_i^3(1)\, dr\right] - \frac{\left(\omega_j^0\right)^2}{2}\mathbb{E}\left(\int_0^1 B_i(r)\, drB_i(1)\right) - \frac{\left(\omega_j^0\right)^2}{4}\mathbb{E}\left(B_i^2(1)\right) + \frac{\left(\omega_j^0\right)^4}{4} \\
&= \mathbb{E}\left[\frac{1}{2}\int_0^1 B_i(r) B_i^3(1)\, dr\right] - \frac{\left(\omega_j^0\right)^4}{4} = \frac{\left(\omega_j^0\right)^4}{2}\int_0^1 \left(3r^2 + 3r(1-r)\right) dr - \frac{\left(\omega_j^0\right)^4}{4} \\
&= \frac{3\left(\omega_j^0\right)^4}{2}\int_0^1 r dr - \frac{\left(\omega_j^0\right)^4}{4} = \frac{\left(\omega_j^0\right)^4}{2}.
\end{aligned} \tag{66}
$$

Based on (64) (65) and (66), the validity of (60) is confirmed under sequential asymptotics $(n,T)_{seq} \to \infty$.

Sufficient conditions then assure joint convergence limit theory using the results in Phillips and Moon (1999). In particular, by Lemma 5(b) of Phillips and Moon (1999), once sequential asymptotic theory $X_{n_j,T} \overset{T\to\infty}{\Rightarrow} X_{n_j} \overset{n\to\infty}{\Rightarrow} X$ is given, joint asymptotics hold if and only if

$$
\lim_{n_j,T\to\infty} \sup \left|\mathbb{E}f\left(X_{n_j,T}\right) - \mathbb{E}f\left(X_{n_j}\right)\right| = 0, \tag{67}
$$

for all bounded, continuous real functions $f$ on $\mathbb{R}^2$. To match the notations in Phillips and Moon (1999), the intermediate workhorse elements here are defined as

$$X_{n_j,T} = \frac{1}{\sqrt{n_j}} \sum_{i \in \mathcal{G}^0(j)} Y_{iT},\ X_{n_j} = \frac{1}{\sqrt{n_j}} \sum_{i \in \mathcal{G}^0(j)} Y_i,$$

$$Y_{iT} := \frac{1}{T} \sum_{t=1}^{T} \left( \widetilde{y}_{i,t-1} \widetilde{u}_{i,t} - \widehat{\lambda}_j + \frac{\widehat{\omega}_j^2}{2} \right),$$

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

By Lemma 5(b) of Phillips and Moon (1999), the sufficient condition (67) holds if

(i) $\limsup_{n,T\to\infty} \frac{1}{n_j} \sum_{i \in \mathcal{G}^0(j)} \mathbb{E}\|Y_{iT}\| < \infty$;

(ii) $\limsup_{n,T\to\infty} \frac{1}{n_j} \sum_{i \in \mathcal{G}^0(j)} \mathbb{E}\|Y_{iT} - \mathbb{E}Y_i\| < \infty$;

(iii) $\limsup_{n,T\to\infty} \frac{1}{n_j} \sum_{i \in \mathcal{G}^0(j)} \mathbb{E}\|Y_{iT}\| \mathbf{1}_{\{\|Y_{iT}\| > n_j \varepsilon\}} = 0$ for any $\varepsilon > 0$;

(iv) $\limsup_{n\to\infty} \frac{1}{n_j} \sum_{i \in \mathcal{G}^0(j)} \mathbb{E}\|Y_i\| \mathbf{1}_{\{\|Y_i\| > n_j \varepsilon\}} = 0$ for any $\varepsilon > 0$.

We proceed to verify conditions (i)-(iv) for $Y_{iT}$ and $Y_i$. First, the integrability of $Y_{iT}$ is readily verified since

$$\mathbb{E}|Y_{iT}| \le \mathbb{E}\left| \frac{1}{T} \sum_{t=1}^{T} y_{i,t-1} u_{i,t} \right| + \mathbb{E}\left| \left( \frac{1}{T^{\frac{3}{2}}} \sum_{t=1}^{T} y_{i,t-1} \right) \left( \frac{1}{\sqrt{T}} \sum_{s=1}^{T} u_{i,s} \right) \right| + \mathbb{E}\left| \frac{\widehat{\sigma}_j^2}{2} \right| < \infty, \qquad (68)$$

when $T \to \infty$ for any $i \in \mathcal{G}^0(j)$ (Phillips, 1987). Then (i) holds due to

$$\limsup_{n,T\to\infty} \frac{1}{n_j} \sum_{i \in \mathcal{G}^0(j)} \mathbb{E}\|Y_{iT}\| = \limsup_{T\to\infty} \mathbb{E}\|Y_{iT}\| < \infty,$$

for any $i \in \mathcal{G}^0(j)$. To show (ii) holds, simply note that $\mathbb{E}Y_{iT} \sim \mathbb{E}Y_i = 0$ for any $i \in \mathcal{G}^0(j)$. For (iii), it follows that

$$\limsup_{n,T\to\infty} \frac{1}{n_j} \sum_{i \in \mathcal{G}^0(j)} \mathbb{E}\|Y_{iT}\| \mathbf{1}_{\{\|Y_{iT}\| > n_j \varepsilon\}} = \limsup_{n,T\to\infty} \mathbb{E}\|Y_{iT}\| \mathbf{1}_{\{\|Y_{iT}\| > n_j \varepsilon\}} = 0,$$

due to (68), for any $i \in \mathcal{G}^0(j)$. Finally, (iv) holds by

$$\limsup_{n\to\infty} \frac{1}{n_j} \sum_{i \in \mathcal{G}^0(j)} \mathbb{E}\|Y_i\| \mathbf{1}_{\{\|Y_i\| > n_j \varepsilon\}} = \limsup_{n\to\infty} \mathbb{E}\|Y_i\| \mathbf{1}_{\{\|Y_i\| > n_j \varepsilon\}} = 0,$$

for any $i \in \mathcal{G}^0(j)$ since $\mathbb{E}\|Y_i\| < \infty$ as shown in the sequential limit. Hence, (67) holds and we have the joint weak convergence (60) irrespective of the divergence rates of $n$ and $T$ to infinity.

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

$$
\begin{aligned}
&\frac{1}{\sqrt{n_j}T^{\gamma}(\rho_j^0)^T}\sum_{i\in\mathcal{G}^0(j)}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\\
&\sim_a \frac{1}{\sqrt{n_j}T^{\gamma}(\rho_j^0)^T}\sum_{i\in\mathcal{G}^0(j)}\sum_{t=1}^{T}y_{i,t-1}u_{it} && (69)\\
&\overset{T\to\infty}{\Rightarrow} \frac{1}{\sqrt{n_j}}\sum_{i\in\mathcal{G}^0(j)}\left(Y_i^{\star}X_i^{\star}\right) && (70)\\
&\overset{n\to\infty}{\Rightarrow} \mathcal{N}\left(0,\frac{\left(\omega_j^0\right)^2}{2c_j^0}\left(\frac{\left(\omega_j^0\right)^2}{2c_j^0}\right)\right),
\end{aligned}
$$

where (69) 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 (70) holds since $\left(Y_i^{\star}, X_i^{\star}\right)' \sim \mathcal{N}\left(\mathbf{0}_{2\times 1}, diag\left\{\frac{\left(\omega_j^0\right)^2}{2c_j^0}, \frac{\left(\omega_j^0\right)^2}{2c_j^0}\right\}\right)$. Joint convergence is verified by checking equation (67) 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 ($c_j^0 < 0$),

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

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

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 (2007b),

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

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

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

$$\begin{aligned}
&\frac{1}{\sqrt{n_j}T^{\frac{1+\gamma}{2}}}\sum_{i\in\mathcal{G}^0(j)}\sum_{t=1}^{T}\left(y_{i,t-1}u_{it}-\lambda_j^0-\overline{m}_{j,T}\frac{c_j^0}{T^\gamma}\right)\\
&=\frac{1}{\sqrt{n_j}T^{\frac{1+\gamma}{2}}}\sum_{i\in\mathcal{G}^0(j)}F_i(1)\sum_{t=1}^{T}y_{i,t-1}\epsilon_{it}+\frac{c_j^0}{\sqrt{n_j}T^{\frac{1+3\gamma}{2}}}\sum_{i\in\mathcal{G}^0(j)}\sum_{t=1}^{T}\left(y_{i,t-1}\widetilde{\epsilon}_{it}-\overline{m}_{j,T}\right)\\
&+\frac{1}{\sqrt{n_j}T^{\frac{1+\gamma}{2}}}\sum_{i\in\mathcal{G}^0(j)}\sum_{t=1}^{T}\left(u_{it}\widetilde{\epsilon}_{it}-\lambda_j^0\right)+o_p(1)\\
&=\frac{1}{\sqrt{n_j}T^{\frac{1+\gamma}{2}}}\sum_{i\in\mathcal{G}^0(j)}F_i(1)\sum_{t=1}^{T}y_{i,t-1}\epsilon_{it}+o_p(1)\\
&\overset{T\to\infty}{\Rightarrow}\frac{1}{\sqrt{n_j}}\sum_{i\in\mathcal{G}^0(j)}\mathcal{N}\left(0,\frac{\left(\omega_j^0\right)^4}{-2c_j^0}\right)\overset{n\to\infty}{\Rightarrow}\mathcal{N}\left(0,\frac{\left(\omega_j^0\right)^4}{-2c_j^0}\right).
\end{aligned}$$

The joint convergence can then be verified by checking (67) and the remaining details follow those of 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 $\overline{c}_i^0\geq 0$ or (ii) $\gamma=0$, we have*

$$\min_{1\leq G<G^0}\inf_{\widehat{\delta}(G)\in\Delta_G}\check{\sigma}^2_{\widehat{\delta}(G)}>\sigma_0^2,\ \textit{with probability approaching 1,}\tag{71}$$

*and*

$$\check{\sigma}^2_{\widehat{\delta}(G^0)} \to_p \sigma_0^2, \tag{72}$$

*where* $\sigma_0^2$ *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 (71) it is sufficient to show

$$\inf_{\widehat{\delta}(G)\in\Delta_G}\check{\sigma}^2_{\widehat{\delta}(G)} > \sigma_0^2, \text{ with probability approaching } 1, \tag{73}$$

for all $G<G^0$. The above relationship (73) 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

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

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

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

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

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

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

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

in which $Q$ is a positive constant. Then equation (71) holds. Moreover, equation (72) holds due to the consistency of $k$-means clustering and post-clustering estimates:

$$\begin{aligned}
\check{\sigma}^2_{\widehat{\delta}(G^0)} &:= \frac{1}{nT}\sum_{j=1}^{G^0}\sum_{i\in\hat{\mathcal{G}}(j,G^0)}\sum_{t=1}^{T}\left(\widetilde{y}_{it}-\widetilde{y}_{i,t-1}\check{\rho}_j\left(G^0\right)\right)^2 = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\widetilde{\widetilde{u}}_{it}^2 \\
&\sim_a \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\widetilde{u}_{it}^2 = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}u_{it}^2 + O_p\left(\frac{1}{T}\right) \\
&\to_p \sigma_0^2, \qquad (74)
\end{aligned}$$

where the last line (74) holds due to equation (29) of the main paper. The proof is now complete. ■

**Proof of Theorem 4.4:** From Theorems 4.1 and 4.2, it follows that

$$\begin{aligned}
\mathrm{IC}\left(G^0\right) &= \ln\left(\check{\sigma}^2_{\widehat{\delta}(G^0)}\right) + G^0\kappa_{nT} \\
&= \ln\left[\frac{1}{nT}\sum_{j=1}^{G^0}\sum_{i\in\hat{\mathcal{G}}(j,G^0)}\sum_{t=1}^{T}\left(\widetilde{y}_{it}-\widetilde{y}_{i,t-1}\check{\rho}_j\left(G^0\right)\right)^2\right] + o(1) \to \ln\left(\sigma_0^2\right).
\end{aligned}$$

Moreover, for an underfitted model with $G < G^0$, note that

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

$$= \min_{1 \le G < G^0} \inf_{\widehat{\delta}(G) \in \Delta_G} \check{\sigma}^2_{\widehat{\delta}(G)}.$$

Under the imposed assumptions (i) $\gamma > 0$ and $\bar{c}_i^0 \ge 0$ or (ii) $\gamma = 0$, as $(n,T) \to \infty$,

$$\min_{1 \le G < G^0} \mathrm{IC}(G) \ge \min_{1 \le G < G^0} \inf_{\widehat{\delta}(G)} \ln \check{\sigma}^2_{\widehat{\delta}(G)} + G \cdot \kappa_{nT} > \ln\left(\sigma_0^2\right).$$

It follows that, when $(n,T) \to \infty$,

$$\Pr\left(\min_{1 \le G < G^0} \mathrm{IC}(G) > \mathrm{IC}\left(G^0\right)\right) \to 1.$$

Lastly, for an overfitted model with $G^0 < G \le G_{\max}$, as $(n,T) \to \infty$,

$$\begin{aligned}
&\Pr\left(\min_{G^0 < G \le G_{\max}} \mathrm{IC}(G) > \mathrm{IC}\left(G^0\right)\right) \\
&= \Pr\left(\min_{G^0 < G \le G_{\max}} \left(nT \ln\left(\check{\sigma}^2_{\widehat{\delta}(G)} / \check{\sigma}^2_{\widehat{\delta}(G^0)}\right) + nT\left(G - G^0\right) \cdot \kappa_{nT}\right) > 0\right) \\
&= \Pr\left(\min_{G^0 < G \le G_{\max}} \left(nT\left(\frac{\check{\sigma}^2_{\widehat{\delta}(G)} - \check{\sigma}^2_{\widehat{\delta}(G^0)}}{\check{\sigma}^2_{\widehat{\delta}(G^0)}}\right) + nT\left(G - G^0\right) \cdot \kappa_{nT}\right) > 0\right) \to 1.
\end{aligned}$$

The proof is now complete. ■

# D Extensions to Fixed Mixed-Root Panel Autoregression

Because the estimated explosive slopes in financial and real estate markets are often very close to unity (e.g., $\check{\rho}_j = 1.011$ or $1.003$ in the Chinese housing market), it is natural to characterize the group-specific slopes in the fashion of a mildly integrated process (Phillips and Magdalinos, 2007a) whose slopes lie within a small region with a radius that shrinks to unity as the sample size $T$ increases (e.g., $\rho_j = 1 + c_j/T^{\gamma}$). As discussed in the main paper, the two-stage algorithm can achieve clustering consistency in such a model. Correspondingly, the associated post-clustering test statistics follow pivotal distribution under the group-specific unit root null hypothesis and diverge under the alternative hypothesis of a mildly explosive root.

When the scaling parameter $\gamma = 0$, the mixed-root panel autoregression has a change in membership character that involves fixed departures from unity: pure stationary groups (with $c_j < 0$ and $\gamma = 0$) and pure explosive groups (with $c_j > 0$ and $\gamma = 0$) in addition to a possible unit root group (with $c_j = 0$ and $\gamma = 0$) . Such a model with fixed slopes helps eliminate potential joint identification issues concerning the parameters $\{c_j, \gamma\}$ and is well-suited to cases where $\gamma$ is estimated to have a value close to zero - see Phillips (2021) for further details. For completeness in our theory development we demonstrate here the continuing applicability of our two-stage selection algorithm.

The new fixed coefficient data generating process for $y_{it}$ is given by

$$y_{it} = \mu_i + \rho_j y_{i,t-1} + u_{it}, \text{ in which } \rho_j = 1 + c_j, \tag{75}$$

in which the main paper is followed and group-specific slope and distance parameters are defined as $\rho_j = \rho_{g_i}$ and $c_j = c_{g_i}$ with $g_i = j$. Other notation listed in the notational glossary of the main paper applies in the same fashion. The mixed-root groups are then determined by the signs and values of the scale coefficients and these are represented in the following diagram:

$$\begin{cases} \text{Explosive groups:} & \begin{cases} \text{Group 1:} & c_1 > 0 \\ \text{Group 2:} & c_2 > 0 \\ \vdots & \vdots \\ \text{Group } g\text{:} & c_g > 0 \end{cases} \\ \text{Unit root group:} & \text{Group } (g+1)\text{:} \quad c_{g+1} = 0 \, , \\ \text{Stationary groups:} & \begin{cases} \text{Group } (g+2)\text{:} & c_{g+2} < 0 \\ \text{Group } (g+3)\text{:} & c_{g+3} < 0 \\ \vdots & \vdots \\ \text{Group } G\text{:} & c_G < 0 \end{cases} \end{cases} \tag{76}$$

where $c_j \neq c_k$ for any $j \neq k$ with indices $j,k \in \{1,2,...,G\}$. The remainder of this section validates the two-stage algorithm and the corresponding test statistics for use in the new model (75) with the group-specific mixed-root structure given by (76).

An attractive feature of the two-stage algorithm is its robustness to the model (75) and (76). Without any material changes, the latent membership of the slope coefficients can be consistently recovered by the recursive $k$-means algorithm and the estimated membership is asymptotically equivalent to the true membership. Further, the post-clustering $t$ and $J$ statistics can be constructed identically as in (17) of the main article. We also establish the pivotal limit theory of the post-clustering $t$ and $J$ statistics under the null hypothesis of group-specific unity and consistency under the alternative of the model (75) and (76) with fixed slope coefficients in explosive groups.

## i Applicability of the two-stage algorithm

The proposed two-stage algorithm continues to work for (75) and (76) conditional on the zero scaling parameter $\gamma$. When $\gamma = 0$, we have

$$\left(\widehat{\rho}_j^* - 1\right) T^{\gamma} = \left(\widehat{\rho}_j^* - 1\right) = \widehat{c}_j^*,$$

under which the validity of the recursive $k$-means algorithm is established. The following modification to Stage 1 of the recursive $k$-means algorithm is used.

The post-clustering estimates and panel test statistics of Stage 2 do not rely on a scaling parameter $\gamma$. So we continue to employ the expressions (12) and (17) of the main paper. In the following discussion, we prove the uniform consistency of the recursive $k$-means clusterings and establish the asymptotic behavior of the post-clustering estimates and test statistics under the joint $(n,T) \to \infty$ asymptotic framework.

**Algorithm 1** Recursive procedure to estimate $c$ and $\delta$

(i) Set $s = 0$. Obtain the time series estimates $\widehat{\overline{\rho}}_i^{TS}$ of the individual slope coefficients $\overline{\rho}_i$ for all $i \in \mathcal{I}_n$ and the corresponding estimates of the localizing coefficients $\widehat{\overline{c}}_i^{TS}$, given $\gamma = 0$. Using any relevant prior information or by pure random assignment, choose $G^0$ estimates of the distancing parameters $c_j^{(0)}$ to form a $G^0$-dimensional vector $c^{(0)}$ as the initial value.

(ii) Given $c^{(s)}$, for $i \in \mathcal{I}_n$ implement the following optimization problem:

$$g_i^{(s+1)} = \arg\min_{j \in \mathcal{G}^0} \left[ \sum_{t=1}^{T} \left( \widetilde{y}_{it} - \widetilde{y}_{i,t-1} \left( 1 + c_j^{(s)} \right) \right)^2 \right]. \tag{77}$$

(iii) Given $\left\{ g_i^{(s+1)} \right\}_{i=1}^{n}$, implement the following optimization problem:

$$c^{(s+1)} = \arg\min_{c \in \mathcal{C}_{G^0}} \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\Upsilon_{iT}} \left[ \sum_{t=1}^{T} \left( \widetilde{y}_{it} - \widetilde{y}_{i,t-1} \left( 1 + c_{g_i^{(s+1)}} \right) \right)^2 \right]. \tag{78}$$

(iv) Let $s = s + 1$ and repeat steps (ii)-(iii) until convergence (say at step $S$). Define $\widehat{c}^* = c^{(S+1)}$ and $\widehat{\delta} = \left( g_1^{(S+1)}, \cdots, g_n^{(S+1)} \right)'$.

## ii Uniform consistency of membership estimation

Under joint asymptotics the estimated group membership is asymptotically equivalent to the true group membership. To justify the uniform consistency of clustering the rate restrictions in Assumption 2(iii) of the main article are modified to $\frac{T}{(\log_2 T)^8} > n(\log_2 n)^2$, and Assumption 1(ii) is modified to 'independent Gaussian errors are imposed for the purely explosive groups'. Then, with $\bar{c}_i^0 > 0$, $u_{it} = \epsilon_{it}$ and $\epsilon_{it} \overset{i.i.d.}{\sim} \mathcal{N}\left(0, \left(\sigma^0\right)^2\right)$. Therefore, for individual $i$ with $\bar{c}_i^0 > 0$, $\left(\bar{\omega}_i^0\right)^2 = \left(\sigma^0\right)^2$. The conditions for the errors of the stationary and unit-root groups are unchanged in Assumption 1(ii). With these modifications to Assumptions 1 and 2 assumed in what follows, the central result on clustering consistency can be established.

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

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

First, the validity of Theorem D.1 holds due to estimation consistency in the first-stage parameter estimate $\widehat{c}^*$.

**Lemma D.1** *Suppose Assumptions 1 and 2 hold. Let $\varepsilon$ take any positive value. Then, when $(n, T) \to \infty$,*

$$d_H\left(c^0, \widehat{c}^*\right) = o_p\left((\log T)^{-\varepsilon}\right). \tag{79}$$

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

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

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

$$\left\|\widehat{c}^* - c^0\right\| = o_p\left((\log T)^{-\varepsilon}\right). \tag{80}$$

Identical to the results for $\gamma \in (0,1)$, Lemma D.1 delivers estimation consistency of $\widehat{c}^*$, which in turn ensures the uniform consistency of membership estimation. The proof of Lemma D.1 is identical to that of Lemma A.2 and, as there, needs to confirm that

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

in which $\widehat{Q}(\cdot,\cdot)$ and $\widetilde{Q}(\cdot,\cdot)$ are as defined as in Lemma A.1.

Second, Theorem D.1 is valid since the tail behavior of the sample variance and covariance of all the individual elements $i = 1, 2, ..., n$ are well controlled asymptotically as shown in the following results.

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

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

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

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

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

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

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

**Lemma D.3** *Suppose that Assumptions 1 and 2 hold, then,*

(i) if $\overline{c}_i^0 > 0$ and $\widetilde{M}_1 \geq \frac{5}{\left(\rho_{low}^2-1\right)^2}\max_{i\in\mathcal{I}_n}\left(\overline{\omega}_i^0\right)^2$,

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

(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 $\widetilde{M}_3 \geq \frac{5\max_{i\in\mathcal{I}_n}\left(\overline{\sigma}_{iu}^0\right)^2}{1-\rho_{low}^2}$,

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

and

**Lemma D.4** *Suppose Assumptions 1 and 2 hold.*

(i) If $\overline{c}_i^0 > 0$ and $0 < \overline{M}_1 \leq \frac{1}{5\left(\rho_{up}^2-1\right)^2}\min_{i\in\mathcal{I}_n}\left(\overline{\omega}_i^0\right)^2$,

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

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

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

*(iii) If* $\overline{c}_i^0 < 0$ *and* $0 < \overline{M}_3 \leq \frac{\min_{i \in \mathcal{I}_n} (\overline{\sigma}_{iu}^0)^2}{5(1-\rho_{up}^2)}$,

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

To establish tail behavior of the unit root individuals we use the iterated logarithm law, as in Huang et al. (2021). To prove tail behavior of the purely stationary individuals, we use the exponential inequality for martingale differences (Freedman, 1975). To prove tail behavior of the purely explosive individuals, we employ the exponential inequality of a $\chi^2$ sequence (Laurent and Massart, 2000) and the dominance of the exponential rates. The details of these proofs are identical to those given in Lemmas A.3, A.4 and A.5 and are omitted here. Based on Lemmas D.2, D.3 and D.4, we can bound the clustering error once the first-step estimate $\widehat{c}^*$ and the true value $c^0$ are close enough, as shown below.

**Lemma D.5** *Suppose Assumptions 1 and 2 hold. Let* $\eta = o\left((\log T)^{-4}\right)$*. 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).

Notice in Lemma D.5 that the radius $\eta$ is much wider than that in Lemma A.6. The reason is that when $\gamma > 0$, the difference $\rho_j - \rho_k = O(T^{-\gamma})$ is a moving quantity that diminishes in large samples. In such cases, the convergence rate in estimating $c_j$ has to be much faster to distinguish differences between $c_j$ and $c_k$ more accurately from the corresponding slope coefficient estimates.

Finally, combining the above results confirms that Theorem D.1 holds for the fixed coefficient model (75) and (76). In particular

$$\begin{aligned}
&n \max_{i \in \mathcal{I}_n} \mathbb{E}\mathbf{1}\{\widehat{g}_i(\widehat{c}^*) \neq g_i^0\} \leq n \max_{i \in \mathcal{I}_n} \Pr\{\left|\widehat{g}_i(\widehat{c}^*) - g_i^0\right| > 0\} \\
\leq& n \max_{i \in \mathcal{I}_n} \sup_{c \in \mathcal{N}_\eta} \Pr\{\left|\widehat{g}_i(c) - g_i^0\right| > 0\} + n \max_{1 \leq j \leq G^0} \Pr\{\left|\widehat{c}_j^* - c_j^0\right| > \eta\} \\
=& o(1) + n \max_{1 \leq j \leq G^0} \Pr\{\left|\widehat{c}_j^* - c_j^0\right| > \eta\} \\
=& o(1),
\end{aligned} \tag{81}$$

where the third line is due to Lemma D.5 and the last line is due to Lemma D.1.

## iii Post-clustering estimates and panel test statistics

Uniform consistency of membership estimation enables the oracle property of the post-clustering estimates to be recovered. Specifically, for any $1 \leq j \leq G^0$, if $c_j^0 \geq 0$ and $\widehat{\rho_j}$ and $\check{\rho}_j$ are defined as in (11) and (12) of the main article, then we have:

$$\sqrt{n_j}\left(\rho_j^0\right)^T\left(\check{\rho}_j - \rho_j^0\right) = \sqrt{n_j}\left(\rho_j^0\right)^T\left(\widehat{\rho_j} - \rho_j^0\right) + o_p(1), \text{ if } c_j^0 > 0; \tag{82}$$

$$\sqrt{n_j}T\left(\check{\rho}_j - 1 + \frac{3\left(\sigma_j^0\right)^2}{\left(\omega_j^0\right)^2}\frac{1}{T}\right) = \sqrt{n_j}T\left(\widehat{\rho_j} - 1 + \frac{3\left(\sigma_j^0\right)^2}{\left(\omega_j^0\right)^2}\frac{1}{T}\right) + o_p(1), \text{ if } c_j^0 = 0. \tag{83}$$

To justify the null limit behavior of the post-clustering panel $t$ and $J$ test statistics, we use the limiting distributions of the $\check{\rho}_j$ when $c_j^0 = 0$, which are identical to the results in the main article where $\gamma \in (0,1)$. Then the pivotal null distributions of the corresponding test statistics are recovered and test consistency follows in a similar fashion, as stated in the following result.

**Theorem D.2** *Suppose that Assumptions 1 and 2 hold and $(n,T) \to \infty$. Under the null hypothesis $\mathcal{H}_0^{(j)} : c_j^0 = 0$, we have*

$$\widetilde{t}_j \Rightarrow \mathcal{N}(0,1),$$
$$\widetilde{J}_j \Rightarrow \mathcal{N}(0,1).$$

*Under the alternative hypothesis $\mathcal{H}_1^{(j)} : c_j^0 > 0$, we have*

$$\widetilde{t}_j = O_p\left(\left(\rho_j^0\right)^T\sqrt{n}\right),$$
$$\widetilde{J}_j = O_p\left(\sqrt{n}T\right).$$

From Theorem D.2 it is readily checked that the post-clustering panel tests still detect explosive roots at a higher rate due to the improved convergence rate obtained from clustered cross section aggregation of the panel data.

## iv Selection of the number of groups

Theorem 4.4 in the main article shows that IC correctly selects asymptotically the real number of groups $G^0$ under the joint convergence framework $(n,T) \to \infty$ when there is a unit root group and mildly explosive root groups. This theorem continues to hold when the explosive root groups involve purely explosive roots with $c_j > 0$ and $\gamma = 0$ rather than mildly explosive groups with $c_j > 0$ and $\gamma \in (0,1)$. Furthermore, a modification of the arguments in the proof of Theorem 4.4 shows that IC correctly selects asymptotically the real number of groups $G^0$ under the joint convergence framework $(n,T) \to \infty$ when there are purely stationary groups (with $c_j < 0$ and $\gamma = 0$) in place of the mildly stationary case with $\gamma > 0$. This result for fixed departures from unity in the stationary direction differs from the result in the main paper where IC tends to underestimate $G^0$. Hence, when $\gamma = 0$, IC delivers correct model selection asymptotically for each of the three possible groups involving fixed departures from a unit root in the explosive and stationary directions in addition to a pure unit root group.

# E 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 $\gamma$ were considered:

$$\left(c_1^0, c_2^0, c_3^0, \gamma\right) = \begin{cases} (-15,-8,-1,0.6) & \text{for} \quad \text{DGP } 0 \\ (1,\ 0,-6,0.6) & \text{for} \quad \text{DGP } 1\ , \\ (1,0.2,-6,0.6) & \text{for} \quad \text{DGP } 2 \end{cases} \tag{84}$$

and

$$\left(c_1^0, c_2^0, \gamma\right) = (-1,1,0.6) \quad \text{for} \quad \text{DGP } 3. \tag{85}$$

In all DGPs we set the distance parameters $c$, rate scaling parameter $\gamma$, group division $\delta$, 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.

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):

$$\text{PP } t\text{-test} = \frac{\left(\widehat{\overline{\rho}}_i^{TS} - 1 - \frac{T \cdot \widehat{\overline{\lambda}}_i^{TS}}{D_{i,T}}\right)\sqrt{D_{i,T}}}{\widehat{\overline{\omega}}_i^{TS}}, \quad \text{PP } J\text{-test} = T\left(\widehat{\overline{\rho}}_i^{TS} - 1 - \frac{T \cdot \widehat{\overline{\lambda}}_i^{TS}}{D_{i,T}}\right), \tag{86}$$

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

$$\text{PP } t\text{-test} \Rightarrow \frac{\int_0^1 \underset{\sim}{W}_i(r)dW_i(r)}{\left[\int_0^1 \underset{\sim}{W}(r)^2 dr\right]^{\frac{1}{2}}}, \quad \text{PP } J\text{-test} \Rightarrow \frac{\int_0^1 \underset{\sim}{W}_i(r)dW_i(r)}{\int_0^1 \underset{\sim}{W}(r)^2 dr}, \tag{87}$$

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 (86) 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 $\kappa_{nT}$ 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 1–4 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 1–4 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 5–7 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 $\delta^0$ is known.

[Insert Tables 5–7 Here]

According to Tables 5–7, 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 both DGPs, the differences 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 8. 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 8 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 $\rho_j^0$ 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 9–12, 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 9–12 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 13–15 report the CE, RMSE, and bias of the post-clustering estimates. From Tables 13–15 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 13-15 Here]

Finally, Table 16 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 16 Here]

# F Tables

Table 1: Empirical frequency of model selection under DGP 0 ($\theta = 0.5$)

| IC | $n$ | $T$ | $G = 1$ | $G = 2$ | $G = 3$ | $G = 4$ | $G = 5$ |
|---|---|---|---|---|---|---|---|
| | 120 | 150 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 |
| | 120 | 250 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 |
| | 120 | 350 | 0.015 | 0.985 | 0.000 | 0.000 | 0.000 |
| | 120 | 450 | 0.476 | 0.524 | 0.000 | 0.000 | 0.000 |
| | 150 | 150 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 |
| | 150 | 250 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 |
| | 150 | 350 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 |
| | 150 | 450 | 0.014 | 0.986 | 0.000 | 0.000 | 0.000 |
| Hausman Test | $n$ | $T$ | $G = 1$ | $G = 2$ | $G = 3$ | $G = 4$ | $G = 5$ |
| | 120 | 150 | 0.000 | 0.008 | 0.980 | 0.012 | 0.000 |
| | 120 | 250 | 0.000 | 0.005 | 0.993 | 0.001 | 0.001 |
| | 120 | 350 | 0.000 | 0.004 | 0.996 | 0.000 | 0.000 |
| | 120 | 450 | 0.000 | 0.004 | 0.996 | 0.000 | 0.000 |
| | 150 | 150 | 0.000 | 0.000 | 0.992 | 0.006 | 0.002 |
| | 150 | 250 | 0.000 | 0.000 | 0.998 | 0.001 | 0.001 |
| | 150 | 350 | 0.000 | 0.000 | 1.000 | 0.000 | 0.000 |
| | 150 | 450 | 0.000 | 0.000 | 1.000 | 0.000 | 0.000 |

Table 2: Empirical frequency of model selection under DGP 1 ($\theta = 0.5$)

| IC | $n$ | $T$ | $G = 1$ | $G = 2$ | $G = 3$ | $G = 4$ | $G = 5$ |
|---|---|---|---|---|---|---|---|
| | 120 | 150 | 0.000 | 0.976 | 0.024 | 0.000 | 0.000 |
| | 120 | 250 | 0.000 | 0.891 | 0.109 | 0.000 | 0.000 |
| | 120 | 350 | 0.000 | 0.774 | 0.226 | 0.000 | 0.000 |
| | 120 | 450 | 0.000 | 0.610 | 0.390 | 0.000 | 0.000 |
| | 150 | 150 | 0.000 | 0.994 | 0.006 | 0.000 | 0.000 |
| | 150 | 250 | 0.000 | 0.931 | 0.069 | 0.000 | 0.000 |
| | 150 | 350 | 0.000 | 0.829 | 0.171 | 0.000 | 0.000 |
| | 150 | 450 | 0.000 | 0.671 | 0.329 | 0.000 | 0.000 |
| Hausman Test | $n$ | $T$ | $G = 1$ | $G = 2$ | $G = 3$ | $G = 4$ | $G = 5$ |
| | 120 | 150 | 0.000 | 0.000 | 0.726 | 0.067 | 0.207 |
| | 120 | 250 | 0.000 | 0.000 | 0.892 | 0.041 | 0.067 |
| | 120 | 350 | 0.000 | 0.000 | 0.963 | 0.011 | 0.026 |
| | 120 | 450 | 0.000 | 0.000 | 0.982 | 0.008 | 0.010 |
| | 150 | 150 | 0.000 | 0.000 | 0.669 | 0.085 | 0.246 |
| | 150 | 250 | 0.000 | 0.000 | 0.884 | 0.037 | 0.079 |
| | 150 | 350 | 0.000 | 0.000 | 0.966 | 0.008 | 0.026 |
| | 150 | 450 | 0.000 | 0.000 | 0.977 | 0.010 | 0.013 |

Table 3: Empirical frequency of model selection under DGP 2 ($\theta = 0.5$)

| IC | $n$ | $T$ | $G = 1$ | $G = 2$ | $G = 3$ | $G = 4$ | $G = 5$ |
|---|---|---|---|---|---|---|---|
| | 120 | 150 | 0.000 | 0.625 | 0.375 | 0.000 | 0.000 |
| | 120 | 250 | 0.000 | 0.259 | 0.741 | 0.000 | 0.000 |
| | 120 | 350 | 0.000 | 0.091 | 0.909 | 0.000 | 0.000 |
| | 120 | 450 | 0.000 | 0.017 | 0.983 | 0.000 | 0.000 |
| | 150 | 150 | 0.000 | 0.700 | 0.300 | 0.000 | 0.000 |
| | 150 | 250 | 0.000 | 0.271 | 0.729 | 0.000 | 0.000 |
| | 150 | 350 | 0.000 | 0.094 | 0.906 | 0.000 | 0.000 |
| | 150 | 450 | 0.000 | 0.017 | 0.983 | 0.000 | 0.000 |
| Hausman Test | $n$ | $T$ | $G = 1$ | $G = 2$ | $G = 3$ | $G = 4$ | $G = 5$ |
| | 120 | 150 | 0.000 | 0.000 | 0.743 | 0.058 | 0.199 |
| | 120 | 250 | 0.000 | 0.000 | 0.911 | 0.023 | 0.066 |
| | 120 | 350 | 0.000 | 0.000 | 0.957 | 0.012 | 0.031 |
| | 120 | 450 | 0.000 | 0.000 | 0.984 | 0.006 | 0.010 |
| | 150 | 150 | 0.000 | 0.000 | 0.678 | 0.078 | 0.244 |
| | 150 | 250 | 0.000 | 0.000 | 0.914 | 0.020 | 0.066 |
| | 150 | 350 | 0.000 | 0.000 | 0.948 | 0.017 | 0.035 |
| | 150 | 450 | 0.000 | 0.000 | 0.984 | 0.007 | 0.009 |

Table 4: Empirical frequency of model selection under DGP 3 ($\theta = 0.5$)

| IC | $n$ | $T$ | $G=1$ | $G=2$ | $G=3$ | $G=4$ | $G=5$ |
|---|---|---|---|---|---|---|---|
| | 120 | 150 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 |
| | 120 | 250 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 |
| | 120 | 350 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 |
| | 120 | 450 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 |
| | 150 | 150 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 |
| | 150 | 250 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 |
| | 150 | 350 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 |
| | 150 | 450 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 |
| Hausman Test | $n$ | $T$ | $G=1$ | $G=2$ | $G=3$ | $G=4$ | $G=5$ |
| | 120 | 150 | 0.000 | 0.738 | 0.003 | 0.233 | 0.026 |
| | 120 | 250 | 0.000 | 0.928 | 0.000 | 0.065 | 0.007 |
| | 120 | 350 | 0.000 | 0.967 | 0.000 | 0.032 | 0.001 |
| | 120 | 450 | 0.000 | 0.986 | 0.000 | 0.014 | 0.000 |
| | 150 | 150 | 0.000 | 0.703 | 0.003 | 0.248 | 0.046 |
| | 150 | 250 | 0.000 | 0.904 | 0.000 | 0.093 | 0.003 |
| | 150 | 350 | 0.000 | 0.957 | 0.000 | 0.042 | 0.001 |
| | 150 | 450 | 0.000 | 0.988 | 0.000 | 0.012 | 0.000 |

Table 5: Clustering and estimation by the two stage procedure under DGP 1 ($\theta = 0.5$)

| | $n$ | $T$ | CE | Post-clustering | | Oracle | |
|---|---|---|---|---|---|---|---|
| | | | | RMSE | Bias | RMSE | Bias |
| Group 1 | 60 | 100 | 0.015 | 0.003 | -0.002 | 0.003 | -0.002 |
| | 60 | 150 | 0.007 | 0.001 | -0.001 | 0.001 | -0.001 |
| | 60 | 200 | 0.004 | 0.000 | -0.000 | 0.000 | -0.000 |
| | 90 | 100 | 0.014 | 0.003 | -0.002 | 0.003 | -0.002 |
| | 90 | 150 | 0.007 | 0.001 | -0.001 | 0.001 | -0.001 |
| | 90 | 200 | 0.003 | 0.000 | -0.000 | 0.000 | -0.000 |
| Group 2 | 60 | 100 | 0.015 | 0.180 | -0.164 | 0.175 | -0.160 |
| | 60 | 150 | 0.007 | 0.154 | -0.142 | 0.151 | -0.139 |
| | 60 | 200 | 0.004 | 0.138 | -0.128 | 0.138 | -0.127 |
| | 90 | 100 | 0.014 | 0.173 | -0.163 | 0.168 | -0.158 |
| | 90 | 150 | 0.007 | 0.148 | -0.140 | 0.146 | -0.137 |
| | 90 | 200 | 0.003 | 0.134 | -0.127 | 0.133 | -0.126 |
| Group 3 | 60 | 100 | 0.015 | 3.600 | 3.596 | 3.568 | 3.565 |
| | 60 | 150 | 0.007 | 3.697 | 3.695 | 3.680 | 3.678 |
| | 60 | 200 | 0.004 | 3.747 | 3.745 | 3.739 | 3.737 |
| | 90 | 100 | 0.014 | 3.603 | 3.600 | 3.569 | 3.567 |
| | 90 | 150 | 0.007 | 3.695 | 3.693 | 3.679 | 3.678 |
| | 90 | 200 | 0.003 | 3.748 | 3.746 | 3.740 | 3.739 |

Table 6: Clustering and estimation by the two stage procedure under DGP 2 ($\theta = 0.5$)

| | $n$ | $T$ | CE | Post-clustering | | Oracle | |
|---|---|---|---|---|---|---|---|
| | | | | RMSE | Bias | RMSE | Bias |
| Group 1 | 60 | 100 | 0.013 | 0.003 | -0.002 | 0.003 | -0.002 |
| | 60 | 150 | 0.005 | 0.001 | -0.001 | 0.001 | -0.001 |
| | 60 | 200 | 0.003 | 0.000 | -0.000 | 0.000 | -0.000 |
| | 90 | 100 | 0.012 | 0.003 | -0.002 | 0.003 | -0.002 |
| | 90 | 150 | 0.005 | 0.001 | -0.001 | 0.001 | -0.001 |
| | 90 | 200 | 0.003 | 0.000 | -0.000 | 0.000 | -0.000 |
| Group 2 | 60 | 100 | 0.013 | 0.177 | -0.168 | 0.174 | -0.166 |
| | 60 | 150 | 0.005 | 0.139 | -0.133 | 0.138 | -0.132 |
| | 60 | 200 | 0.003 | 0.114 | -0.109 | 0.114 | -0.109 |
| | 90 | 100 | 0.012 | 0.172 | -0.166 | 0.170 | -0.164 |
| | 90 | 150 | 0.005 | 0.134 | -0.130 | 0.133 | -0.129 |
| | 90 | 200 | 0.003 | 0.110 | -0.107 | 0.110 | -0.107 |
| Group 3 | 60 | 100 | 0.013 | 3.599 | 3.595 | 3.568 | 3.565 |
| | 60 | 150 | 0.005 | 3.697 | 3.694 | 3.680 | 3.678 |
| | 60 | 200 | 0.003 | 3.748 | 3.745 | 3.739 | 3.737 |
| | 90 | 100 | 0.012 | 3.601 | 3.597 | 3.569 | 3.567 |
| | 90 | 150 | 0.005 | 3.699 | 3.697 | 3.679 | 3.678 |
| | 90 | 200 | 0.003 | 3.749 | 3.748 | 3.740 | 3.739 |

Table 7: Clustering and estimation by the two stage procedure under DGP 3 ($\theta = 0.5$)

| | $n$ | $T$ | CE | Post-clustering | | Oracle | |
|---|---|---|---|---|---|---|---|
| | | | | RMSE | Bias | RMSE | Bias |
| Group 1 | 60 | 100 | 0.005 | 0.003 | -0.002 | 0.003 | -0.002 |
| | 60 | 150 | 0.002 | 0.001 | -0.001 | 0.001 | -0.001 |
| | 60 | 200 | 0.001 | 0.000 | -0.000 | 0.000 | -0.000 |
| | 90 | 100 | 0.006 | 0.002 | -0.002 | 0.002 | -0.002 |
| | 90 | 150 | 0.002 | 0.001 | -0.001 | 0.001 | -0.001 |
| | 90 | 200 | 0.001 | 0.000 | -0.000 | 0.000 | -0.000 |
| Group 2 | 60 | 100 | 0.005 | 0.562 | 0.559 | 0.563 | 0.560 |
| | 60 | 150 | 0.002 | 0.586 | 0.584 | 0.586 | 0.584 |
| | 60 | 200 | 0.001 | 0.600 | 0.598 | 0.600 | 0.598 |
| | 90 | 100 | 0.006 | 0.575 | 0.556 | 0.563 | 0.561 |
| | 90 | 150 | 0.002 | 0.587 | 0.585 | 0.587 | 0.585 |
| | 90 | 200 | 0.001 | 0.600 | 0.599 | 0.600 | 0.599 |

Table 8: Tests for detecting explosiveness ($\theta = 0.5$)

| DGP 1 | $n$ | $T$ | $c_1 = 1$ | | $c_2 = 0$ | |
|---|---|---|---|---|---|---|
| | | | $t$-test | $J$-test | $t$-test | $J$-test |
| | 1 | 100 | 0.994 | 0.994 | 0.060 | 0.059 |
| | 1 | 150 | 0.996 | 0.996 | 0.064 | 0.066 |
| | 1 | 200 | 0.999 | 0.999 | 0.069 | 0.069 |
| | 30 | 100 | 1.000 | 1.000 | 0.045 | 0.047 |
| | 30 | 150 | 1.000 | 1.000 | 0.062 | 0.064 |
| | 30 | 200 | 1.000 | 1.000 | 0.042 | 0.041 |
| | 60 | 100 | 1.000 | 1.000 | 0.070 | 0.057 |
| | 60 | 150 | 1.000 | 1.000 | 0.054 | 0.057 |
| | 60 | 200 | 1.000 | 1.000 | 0.058 | 0.059 |
| | 90 | 100 | 1.000 | 1.000 | 0.082 | 0.079 |
| | 90 | 150 | 1.000 | 1.000 | 0.052 | 0.053 |
| | 90 | 200 | 1.000 | 1.000 | 0.040 | 0.042 |
| DGP 2 | $n$ | $T$ | $c_1 = 1$ | | $c_2 = 0.2$ | |
| | | | $t$-test | $J$-test | $t$-test | $J$-test |
| | 1 | 100 | 0.994 | 0.994 | 0.175 | 0.175 |
| | 1 | 150 | 0.996 | 0.996 | 0.293 | 0.297 |
| | 1 | 200 | 0.999 | 0.999 | 0.382 | 0.382 |
| | 30 | 100 | 1.000 | 1.000 | 0.599 | 0.262 |
| | 30 | 150 | 1.000 | 1.000 | 0.797 | 0.441 |
| | 30 | 200 | 1.000 | 1.000 | 0.920 | 0.657 |
| | 60 | 100 | 1.000 | 1.000 | 0.807 | 0.527 |
| | 60 | 150 | 1.000 | 1.000 | 0.961 | 0.785 |
| | 60 | 200 | 1.000 | 1.000 | 0.994 | 0.951 |
| | 90 | 100 | 1.000 | 1.000 | 0.917 | 0.705 |
| | 90 | 150 | 1.000 | 1.000 | 0.992 | 0.938 |
| | 90 | 200 | 1.000 | 1.000 | 0.999 | 0.996 |
| DGP 3 | $n$ | $T$ | $c_1 = 1$ | | $c_2 = -1$ | |
| | | | $t$-test | $J$-test | $t$-test | $J$-test |
| | 30 | 100 | 1.000 | 1.000 | 0.010 | 0.226 |
| | 30 | 150 | 1.000 | 1.000 | 0.006 | 0.195 |
| | 30 | 200 | 1.000 | 1.000 | 0.003 | 0.179 |
| | 60 | 100 | 1.000 | 1.000 | 0.027 | 0.513 |
| | 60 | 150 | 1.000 | 1.000 | 0.014 | 0.454 |
| | 60 | 200 | 1.000 | 1.000 | 0.005 | 0.473 |
| | 90 | 100 | 1.000 | 1.000 | 0.089 | 0.766 |
| | 90 | 150 | 1.000 | 1.000 | 0.040 | 0.719 |
| | 90 | 200 | 1.000 | 1.000 | 0.029 | 0.734 |

Table 9: Empirical frequency of model selection under DGP 0 ($\theta = 0$)

| IC | $n$ | $T$ | $G=1$ | $G=2$ | $G=3$ | $G=4$ | $G=5$ |
|---|---|---|---|---|---|---|---|
| | 120 | 150 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 |
| | 120 | 250 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 |
| | 120 | 350 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 |
| | 120 | 450 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 |
| | 150 | 150 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 |
| | 150 | 250 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 |
| | 150 | 350 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 |
| | 150 | 450 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 |
| Hausman Test | $n$ | $T$ | $G=1$ | $G=2$ | $G=3$ | $G=4$ | $G=5$ |
| | 120 | 150 | 0.000 | 0.642 | 0.286 | 0.052 | 0.020 |
| | 120 | 250 | 0.000 | 0.027 | 0.791 | 0.037 | 0.145 |
| | 120 | 350 | 0.000 | 0.000 | 0.932 | 0.016 | 0.052 |
| | 120 | 450 | 0.000 | 0.000 | 0.987 | 0.004 | 0.009 |
| | 150 | 150 | 0.000 | 0.454 | 0.388 | 0.090 | 0.068 |
| | 150 | 250 | 0.000 | 0.006 | 0.637 | 0.042 | 0.315 |
| | 150 | 350 | 0.000 | 0.000 | 0.842 | 0.015 | 0.143 |
| | 150 | 450 | 0.000 | 0.000 | 0.972 | 0.012 | 0.016 |

Table 10: Empirical frequency of model selection under DGP 1 ($\theta = 0$)

| IC | $n$ | $T$ | $G=1$ | $G=2$ | $G=3$ | $G=4$ | $G=5$ |
|---|---|---|---|---|---|---|---|
| | 120 | 150 | 0.000 | 0.005 | 0.995 | 0.000 | 0.000 |
| | 120 | 250 | 0.000 | 0.014 | 0.986 | 0.000 | 0.000 |
| | 120 | 350 | 0.000 | 0.023 | 0.977 | 0.000 | 0.000 |
| | 120 | 450 | 0.000 | 0.033 | 0.967 | 0.000 | 0.000 |
| | 150 | 150 | 0.000 | 0.000 | 1.000 | 0.000 | 0.000 |
| | 150 | 250 | 0.000 | 0.000 | 1.000 | 0.000 | 0.000 |
| | 150 | 350 | 0.000 | 0.000 | 1.000 | 0.000 | 0.000 |
| | 150 | 450 | 0.000 | 0.000 | 1.000 | 0.000 | 0.000 |
| Hausman Test | $n$ | $T$ | $G=1$ | $G=2$ | $G=3$ | $G=4$ | $G=5$ |
| | 120 | 150 | 0.000 | 0.000 | 0.797 | 0.068 | 0.135 |
| | 120 | 250 | 0.000 | 0.001 | 0.934 | 0.023 | 0.042 |
| | 120 | 350 | 0.000 | 0.002 | 0.975 | 0.009 | 0.014 |
| | 120 | 450 | 0.000 | 0.002 | 0.992 | 0.002 | 0.004 |
| | 150 | 150 | 0.000 | 0.000 | 0.738 | 0.072 | 0.190 |
| | 150 | 250 | 0.000 | 0.000 | 0.898 | 0.029 | 0.073 |
| | 150 | 350 | 0.000 | 0.000 | 0.972 | 0.008 | 0.020 |
| | 150 | 450 | 0.000 | 0.000 | 0.992 | 0.000 | 0.008 |

Table 11: Empirical frequency of model selection under DGP 2 ($\theta = 0$)

| IC | $n$ | $T$ | $G = 1$ | $G = 2$ | $G = 3$ | $G = 4$ | $G = 5$ |
|---|---|---|---|---|---|---|---|
| | 120 | 150 | 0.000 | 0.000 | 1.000 | 0.000 | 0.000 |
| | 120 | 250 | 0.000 | 0.000 | 1.000 | 0.000 | 0.000 |
| | 120 | 350 | 0.000 | 0.000 | 1.000 | 0.000 | 0.000 |
| | 120 | 450 | 0.000 | 0.000 | 1.000 | 0.000 | 0.000 |
| | 150 | 150 | 0.000 | 0.000 | 1.000 | 0.000 | 0.000 |
| | 150 | 250 | 0.000 | 0.000 | 1.000 | 0.000 | 0.000 |
| | 150 | 350 | 0.000 | 0.000 | 1.000 | 0.000 | 0.000 |
| | 150 | 450 | 0.000 | 0.000 | 1.000 | 0.000 | 0.000 |
| Hausman Test | $n$ | $T$ | $G = 1$ | $G = 2$ | $G = 3$ | $G = 4$ | $G = 5$ |
| | 120 | 150 | 0.000 | 0.000 | 0.792 | 0.082 | 0.126 |
| | 120 | 250 | 0.000 | 0.000 | 0.861 | 0.088 | 0.051 |
| | 120 | 350 | 0.000 | 0.000 | 0.834 | 0.119 | 0.047 |
| | 120 | 450 | 0.000 | 0.000 | 0.808 | 0.126 | 0.066 |
| | 150 | 150 | 0.000 | 0.000 | 0.732 | 0.098 | 0.170 |
| | 150 | 250 | 0.000 | 0.000 | 0.804 | 0.117 | 0.079 |
| | 150 | 350 | 0.000 | 0.000 | 0.727 | 0.186 | 0.087 |
| | 150 | 450 | 0.000 | 0.000 | 0.705 | 0.185 | 0.110 |

Table 12: Empirical frequency of model selection under DGP 3 ($\theta = 0$)

| IC | $n$ | $T$ | $G = 1$ | $G = 2$ | $G = 3$ | $G = 4$ | $G = 5$ |
|---|---|---|---|---|---|---|---|
| | 120 | 150 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 |
| | 120 | 250 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 |
| | 120 | 350 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 |
| | 120 | 450 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 |
| | 150 | 150 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 |
| | 150 | 250 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 |
| | 150 | 350 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 |
| | 150 | 450 | 0.000 | 1.000 | 0.000 | 0.000 | 0.000 |
| Hausman Test | $n$ | $T$ | $G = 1$ | $G = 2$ | $G = 3$ | $G = 4$ | $G = 5$ |
| | 120 | 150 | 0.000 | 0.817 | 0.000 | 0.173 | 0.010 |
| | 120 | 250 | 0.000 | 0.949 | 0.000 | 0.050 | 0.001 |
| | 120 | 350 | 0.000 | 0.986 | 0.000 | 0.013 | 0.001 |
| | 120 | 450 | 0.000 | 0.991 | 0.000 | 0.009 | 0.000 |
| | 150 | 150 | 0.000 | 0.791 | 0.000 | 0.194 | 0.015 |
| | 150 | 250 | 0.000 | 0.913 | 0.000 | 0.085 | 0.002 |
| | 150 | 350 | 0.000 | 0.971 | 0.000 | 0.029 | 0.000 |
| | 150 | 450 | 0.000 | 0.989 | 0.000 | 0.011 | 0.000 |

Table 13: Clustering and estimation by the two stage procedure under DGP 1 ($\theta = 0$)

| | $n$ | $T$ | CE | Post-clustering RMSE | Post-clustering Bias | Oracle RMSE | Oracle Bias |
|---|---|---|---|---|---|---|---|
| Group 1 | 60 | 100 | 0.011 | 0.003 | -0.002 | 0.003 | -0.002 |
| | 60 | 150 | 0.004 | 0.001 | -0.001 | 0.001 | -0.001 |
| | 60 | 200 | 0.002 | 0.000 | -0.000 | 0.000 | -0.000 |
| | 90 | 100 | 0.011 | 0.003 | -0.002 | 0.003 | -0.002 |
| | 90 | 150 | 0.004 | 0.001 | -0.001 | 0.001 | -0.001 |
| | 90 | 200 | 0.002 | 0.000 | -0.000 | 0.000 | -0.000 |
| Group 2 | 60 | 100 | 0.011 | 0.477 | -0.462 | 0.476 | -0.461 |
| | 60 | 150 | 0.004 | 0.421 | -0.409 | 0.420 | -0.409 |
| | 60 | 200 | 0.002 | 0.379 | -0.369 | 0.379 | -0.369 |
| | 90 | 100 | 0.011 | 0.466 | -0.457 | 0.467 | -0.458 |
| | 90 | 150 | 0.004 | 0.408 | -0.400 | 0.408 | -0.400 |
| | 90 | 200 | 0.002 | 0.371 | -0.364 | 0.372 | -0.364 |
| Group 3 | 60 | 100 | 0.011 | 0.384 | -0.220 | 0.406 | -0.289 |
| | 60 | 150 | 0.004 | 0.352 | -0.215 | 0.358 | -0.243 |
| | 60 | 200 | 0.002 | 0.342 | -0.211 | 0.341 | -0.224 |
| | 90 | 100 | 0.011 | 0.352 | -0.212 | 0.379 | -0.288 |
| | 90 | 150 | 0.004 | 0.322 | -0.218 | 0.332 | -0.248 |
| | 90 | 200 | 0.002 | 0.305 | -0.207 | 0.308 | -0.221 |

Table 14: Clustering and estimation by the two stage procedure under DGP 2 ($\theta = 0$)

| | $n$ | $T$ | CE | Post-clustering RMSE | Post-clustering Bias | Oracle RMSE | Oracle Bias |
|---|---|---|---|---|---|---|---|
| Group 1 | 60 | 100 | 0.009 | 0.003 | -0.002 | 0.003 | -0.002 |
| | 60 | 150 | 0.003 | 0.001 | -0.001 | 0.001 | -0.001 |
| | 60 | 200 | 0.002 | 0.000 | -0.000 | 0.000 | -0.000 |
| | 90 | 100 | 0.009 | 0.003 | -0.002 | 0.003 | -0.002 |
| | 90 | 150 | 0.003 | 0.001 | -0.001 | 0.001 | -0.001 |
| | 90 | 200 | 0.002 | 0.000 | -0.000 | 0.000 | -0.000 |
| Group 2 | 60 | 100 | 0.009 | 0.292 | -0.281 | 0.292 | -0.282 |
| | 60 | 150 | 0.003 | 0.221 | -0.214 | 0.221 | -0.214 |
| | 60 | 200 | 0.002 | 0.173 | -0.168 | 0.174 | -0.168 |
| | 90 | 100 | 0.009 | 0.284 | -0.278 | 0.285 | -0.278 |
| | 90 | 150 | 0.003 | 0.211 | -0.207 | 0.212 | -0.208 |
| | 90 | 200 | 0.002 | 0.168 | -0.164 | 0.168 | -0.164 |
| Group 3 | 60 | 100 | 0.009 | 0.392 | -0.234 | 0.406 | -0.289 |
| | 60 | 150 | 0.003 | 0.355 | -0.219 | 0.358 | -0.243 |
| | 60 | 200 | 0.002 | 0.342 | -0.214 | 0.341 | -0.224 |
| | 90 | 100 | 0.009 | 0.356 | -0.226 | 0.379 | -0.288 |
| | 90 | 150 | 0.003 | 0.321 | -0.223 | 0.332 | -0.248 |
| | 90 | 200 | 0.002 | 0.304 | -0.209 | 0.308 | -0.221 |

Table 15: Clustering and estimation by the two stage procedure under DGP 3 ($\theta = 0$)

| | $n$ | $T$ | CE | Post-clustering | | Oracle | |
|---|---|---|---|---|---|---|---|
| | | | | RMSE | Bias | RMSE | Bias |
| Group 1 | 60 | 100 | 0.004 | 0.002 | -0.002 | 0.002 | -0.002 |
| | 60 | 150 | 0.001 | 0.001 | -0.001 | 0.001 | -0.001 |
| | 60 | 200 | 0.001 | 0.000 | -0.000 | 0.000 | -0.000 |
| | 90 | 100 | 0.005 | 0.002 | -0.002 | 0.002 | -0.002 |
| | 90 | 150 | 0.001 | 0.001 | -0.001 | 0.001 | -0.001 |
| | 90 | 200 | 0.001 | 0.000 | -0.000 | 0.000 | -0.000 |
| Group 2 | 60 | 100 | 0.004 | 0.416 | -0.392 | 0.414 | -0.392 |
| | 60 | 150 | 0.001 | 0.346 | -0.324 | 0.345 | -0.323 |
| | 60 | 200 | 0.001 | 0.306 | -0.284 | 0.305 | -0.284 |
| | 90 | 100 | 0.005 | 0.404 | -0.388 | 0.403 | -0.388 |
| | 90 | 150 | 0.001 | 0.336 | -0.320 | 0.335 | -0.320 |
| | 90 | 200 | 0.001 | 0.297 | -0.283 | 0.297 | -0.282 |

Table 16: Tests for detecting explosiveness ($\theta = 0$)

| DGP 1 | $n$ | $T$ | $c_1 = 1$ | | $c_2 = 0$ | |
|---|---|---|---|---|---|---|
| | | | $t$-test | $J$-test | $t$-test | $J$-test |
| | 1 | 100 | 0.992 | 0.992 | 0.025 | 0.025 |
| | 1 | 150 | 0.996 | 0.996 | 0.022 | 0.020 |
| | 1 | 200 | 0.999 | 0.999 | 0.027 | 0.027 |
| | 30 | 100 | 1.000 | 1.000 | 0.012 | 0.062 |
| | 30 | 150 | 1.000 | 1.000 | 0.020 | 0.068 |
| | 30 | 200 | 1.000 | 1.000 | 0.010 | 0.055 |
| | 60 | 100 | 1.000 | 1.000 | 0.019 | 0.093 |
| | 60 | 150 | 1.000 | 1.000 | 0.017 | 0.083 |
| | 60 | 200 | 1.000 | 1.000 | 0.012 | 0.067 |
| | 90 | 100 | 1.000 | 1.000 | 0.040 | 0.129 |
| | 90 | 150 | 1.000 | 1.000 | 0.015 | 0.088 |
| | 90 | 200 | 1.000 | 1.000 | 0.005 | 0.058 |
| DGP 2 | $n$ | $T$ | $c_1 = 1$ | | $c_2 = 0.2$ | |
| | | | $t$-test | $J$-test | $t$-test | $J$-test |
| | 1 | 100 | 0.992 | 0.992 | 0.112 | 0.111 |
| | 1 | 150 | 0.996 | 0.996 | 0.211 | 0.211 |
| | 1 | 200 | 0.999 | 0.999 | 0.292 | 0.296 |
| | 30 | 100 | 1.000 | 1.000 | 0.441 | 0.324 |
| | 30 | 150 | 1.000 | 1.000 | 0.665 | 0.479 |
| | 30 | 200 | 1.000 | 1.000 | 0.847 | 0.689 |
| | 60 | 100 | 1.000 | 1.000 | 0.712 | 0.622 |
| | 60 | 150 | 1.000 | 1.000 | 0.926 | 0.826 |
| | 60 | 200 | 1.000 | 1.000 | 0.986 | 0.964 |
| | 90 | 100 | 1.000 | 1.000 | 0.878 | 0.815 |
| | 90 | 150 | 1.000 | 1.000 | 0.987 | 0.960 |
| | 90 | 200 | 1.000 | 1.000 | 0.999 | 0.998 |
| DGP 3 | $n$ | $T$ | $c_1 = 1$ | | $c_2 = -1$ | |
| | | | $t$-test | $J$-test | $t$-test | $J$-test |
| | 30 | 100 | 1.000 | 1.000 | 0.000 | 0.098 |
| | 30 | 150 | 1.000 | 1.000 | 0.000 | 0.087 |
| | 30 | 200 | 1.000 | 1.000 | 0.000 | 0.084 |
| | 60 | 100 | 1.000 | 1.000 | 0.000 | 0.238 |
| | 60 | 150 | 1.000 | 1.000 | 0.000 | 0.179 |
| | 60 | 200 | 1.000 | 1.000 | 0.000 | 0.190 |
| | 90 | 100 | 1.000 | 1.000 | 0.000 | 0.389 |
| | 90 | 150 | 1.000 | 1.000 | 0.000 | 0.311 |
| | 90 | 200 | 1.000 | 1.000 | 0.000 | 0.336 |

Table 17: 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 18: 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 19: 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. |

## 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.

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

——— (2021): "Estimation and Inference with Near Unit Roots," *Cowles Foundation Discussion Paper No. 2304, Yale University*.

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 (2007a): "Limit theory for moderate deviations from a unit root," *Journal of Econometrics*, 136, 115–130.

——— (2007b): "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.