A PANEL CLUSTERING APPROACH TO ANALYZING BUBBLE BEHAVIOR

By

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

February 2022

A Panel Clustering Approach to Analyzing Bubble Behavior[*]

Yanbo Liu
School of Economics, Shandong University

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

Jun Yu
Singapore Management University

February 15, 2022

Abstract

This study provides new mechanisms for identifying and estimating explosive bubbles in mixed-root panel autoregressions with a latent group structure. A post-clustering approach is employed that combines a recursive k-means clustering algorithm with panel-data test statistics for testing the presence of explosive roots in time series trajectories. Uniform consistency of the k-means clustering algorithm is established, showing that the post-clustering estimate is asymptotically equivalent to the oracle counterpart that uses the true group identities. Based on the estimated group membership, right-tailed self-normalized t-tests and coefficient-based J-tests, each with pivotal limit distributions, are introduced to detect the explosive roots. The usual Information Criterion (IC) for selecting the correct number of groups is found to be inconsistent and a new method that combines IC with a Hausman-type specification test is proposed that consistently estimates the true number of groups. Extensive Monte Carlo simulations provide strong evidence that in finite samples, the recursive k-means clustering algorithm can correctly recover latent group membership in data of this type and the proposed post-clustering panel-data tests lead to substantial power gains compared with the time series approach. The proposed methods are used to identify bubble behavior in US and Chinese housing markets, and the US stock market, leading to new findings concerning speculative behavior in these markets.

JEL classification: C22, C33, C51, G01

1   Introduction

A key characteristic of financial bubbles such as the dot-com bubble of the 1990s and early 2000s is the presence of mildly explosive deviations of asset prices from their fundamental values during the expansive phase of the bubble. Divergence from market fundamentals can arise whenever there is widespread belief that ongoing robust price increases will continue. Sufficient market participants sharing this belief can drive up prices and produce expectations that ongoing price gains will continue, as argued by Shiller (2015) and others. This self-fulfilling mechanism can lead to price growth that becomes exponential (or explosive), resulting in a market that is progressively misaligned from its fundamentals.

This expansive phase of a financial bubble, although not the switching mechanism to bubble collapse, is partly captured by the standard present value model

$$P_t = \sum_{i=0}^{\infty} \left(\frac{1}{1+r_f} \right)^i \mathbb{E}_t \left(D_{t+i} \right) + B_t, \tag{1}$$

where P_t is the price of an asset at time t, D_t is the payoff of the asset, r_f is the risk-free interest rate, and B_t represents a potential bubble component which satisfies the following submartingale property

$$\mathbb{E}_t \left(B_{t+1} \right) = \left(1 + r_f \right) B_t > B_t.$$

When there is no bubble (i.e., $B_t = 0$), the asset price is completely determined by the aggregate of the discounted expected future payoffs, $\sum_{i=0}^{\infty} (1 + r_f)^{-i} \mathbb{E}_t(D_{t+i})$, which constitutes what is known as the fundamental value. Further, if D_{t+i} is a martingale or more generally an $I(1)$ (unit root) process, then asset prices P_t cannot be explosive. However, if there is a bubble (i.e., $B_t \neq 0$), then B_t and in consequence P_t are explosive. This implication of the model (1) explains why the econometric analysis of bubble behavior has focused on implementing right-tailed unit root tests to detect explosive behavior in asset prices adjusted by the fundamentals, as in Phillips et al. (2015a,b).

Conventional econometric methods for bubble detection, including the Dickey-Fuller (DF) and augmented DF (ADF) tests (Diba and Grossman, 1987, 1988), the SADF test (Phillips and Yu, 2009, 2011; Phillips et al., 2011) and the GSADF test (Phillips et al., 2015a,b), all proceed with a single time series to assess evidence. Single series methods do not always have good discriminatory power for bubble detection especially with short-lived or slow-growing bubbles; and such tests neglect the presence of any prevailing wider phenomena of market exuberance. To illustrate the low-power problem of conventional single series tests when a bubble is short-lived or grows slowly, we employed

2

two experiments with simulated data from the following simple AR(1) design,

$$y_t = \rho y_{t-1} + u_t, \ y_0 = 0, \ u_t \sim \text{i.i.d. } \mathcal{N}(0,1), \ t = 1, 2, ..., T, \tag{2}$$

using the DF t-statistic $[(\widehat{\rho} - 1)/se(\widehat{\rho})]$ and the DF J-statistic $[T(\widehat{\rho} - 1)]$, where

$$\widehat{\rho} = \sum_{t=1}^{T} (y_t - \bar{y})(y_{t-1} - \bar{y}_{-1}) / \sum_{t=1}^{T} (y_{t-1} - \bar{y}_{-1})^2, \tag{3}$$

is the least-squares (LS) estimator of ρ, with $\bar{y} = \frac{1}{T} \sum_{s=1}^{T} y_s$, $\bar{y}_{-1} = \frac{1}{T} \sum_{s=1}^{T} y_{s-1}$, and $se(\widehat{\rho})$ is the usual standard error of $\widehat{\rho}$ in the AR(1) model with intercept. The null hypothesis $\mathcal{H}_0 : \rho = 1$ is tested against the explosive alternative $\mathcal{H}_1 : \rho > 1$ based on right-tailed nominal 95% critical values. The first experiment uses the empirical estimates of ρ found in Phillips et al. (2011) as the true values (i.e., $\rho = 1.033, 1.040$) and sets $T = 10, 20, 30$. This experiment has small sample sizes so the bubble is short-lived but realistic, from the empirical findings in Phillips et al. (2015a,b). Table 1 reports the powers (i.e., empirical rejection frequency under $\mathcal{H}_1 : \rho > 1$ from 10,000 replications) of the right-tailed t and J tests rejecting the null hypothesis. Evidently, the power of these tests is low, ranging from 0.1009 to 0.2202 for the t-test and from 0.0957 to 0.2261 for the J-test. The second experiment uses true values of ρ that are much closer to unity ($\rho = 1.0009, 1.0069$), leading to very slow exponential growth during the expansive phase of the bubble and reflecting some of the growth rates actually found in the empirical results reported later in Section 6. Table 2 reports the powers of the right-tailed t and J tests in this case. The power of these tests is again found to be very low, ranging from 0.0590 to 0.2913 for the t-test and from 0.0597 to 0.2991 for the J-test.

[Insert Tables 1 and 2 Here]

As an empirical illustration of low discriminatory power from single time series regressions, the t-test was applied to each of the 107 city-level house price indices in China used in the empirical section of the paper – see Section 6 for details. The results led to rejections of the unit root null in favor of an explosive alternative for 58 cities, with tests for the remaining 49 cities failing to reject. Figure 2 compares these test results with those of the new panel tests that make use of the clustering algorithm to identify clusters of cities with the same autoregressive coefficient reflecting group time series behavior within the panel. The effects of clustering individual time series into common groups reveal the additional discriminatory power obtained by grouping. Clustered panel tests of this type can also detect the presence of explosive roots that fail detection in time series tests. Empirical examples are given in Figures 3 and 4 for the US housing and equity markets. More specifically, the clustered panel t-test introduced later helps to diagnose mildly explosive price behavior in U.S city housing markets where individual time series tests reveal no evidence of such behavior, confirming the discriminatory power gains that arise from cross section aggregation.

3

As indicated above, behavior such as speculative exuberance in financial, commodity, and real estate markets often manifests as a wider market phenomenon. Market prices of multiple assets in the same class are often available and can therefore be used in testing for exuberance. The primary contribution of the present paper is to propose the use of such panel data to enhance power in bubble detection algorithms. When there is homogeneity in behavior over certain cross section units in a particular group there are typically power advantages to pooling within that group for estimation and testing. If the group structure in a panel is known or can be reliably estimated and exuberance is expected, then pooling will sharpen statistical inference on the common explosive root compared to the use of single time series. It is often reasonable to expect some homogeneity over cross section units in economic and financial panel data (Hahn and Moon, 2010; Narayan et al., 2013; David et al., 2016; Kong et al., 2019; Wang et al., 2019; Bordt et al., 2020). In financial markets, for instance, stocks in the same industry may share similar fundamentals and their dynamic behavior may be closely related. Or in real estate markets, the dynamics in house prices may be related among cities with a similar characteristics, including demographics, locational features, and level of urbanization.

In practice, group structure is often unknown and has to be estimated from the panel itself. To do so in the context of asset bubble investigation, a mixed-root specification for the panel model is specified in which the individual time series are characterized as autoregressive with a mixture of roots, some mildly explosive, some near stationary, and some unit roots (Phillips and Magdalinos, 2007a,b; Phillips and Lee, 2013). In bubble detection it is typically far too restrictive to impose a homogeneous explosive root across all cross section units and instead more realistic to allow for some explosive bubble behavior in a proportion of these units so that parameter homogeneity is group specific. Then any underlying group structure must be recovered empirically. To achieve this end, econometric methods have recently been developed, including the k-means clustering algorithm (Bonhomme and Manresa, 2015; Bonhomme et al., 2017) and the classification Lasso (C-Lasso) approach of Su et al. (2016). The present paper uses the recursive k-means algorithm to uncover latent group membership. With grouping accomplished, bubble detection procedures can be implemented in the second step.

After estimating and determining the number of the latent groups, two right-tailed tests are proposed to detect explosive behavior. The tests are panel versions of the self-normalized and coefficient-based tests, which we subsequently refer to as the t and J statistics, and these are asymptotically pivotal with standard Gaussian distributions under the null hypothesis of a common unit root in the group. Under the alternative of a group-specific mildly explosive root, these post-clustering panel statistics diverge and the tests are consistent. For comparison with the single time series tests, the panel tests are applied to the same house price and stock market data discussed above.

The panel tests dominate the time series tests in two aspects. First, unlike the time series tests which have nonstandard limit theory, the panel tests have standard asymptotic

Gaussian distributions under the null and are convenient to implement. Second, under the condition of within group coefficient homogeneity, the tests have the advantage faster divergence rates than the time series tests by virtue of cross section information aggregation. Extensive Monte Carlo simulations demonstrate that the empirical powers of the panel tests are considerably higher than their time series counterparts.

This paper makes five contributions. First, it extends the literature on bubble detection by using statistical clustering and group averaging to raise test power of existing time series tests in a manner similar to the mechanism of the standard left-sided panel unit root tests (Im et al., 2003; Moon and Perron, 2004; Chang and Song, 2009; Bai and Ng, 2010). An alternative approach to panel modeling of financial bubbles is to allow for potentially explosive common factors and employ principal component analysis (PCA) to detect explosive behavior in some of the factors, as in Chen et al. (2022) who work with a model that has a single explosive factor and no latent groupings. An important advantage of the factor-based panel approach is the allowance for cross section dependence. That work can be extended by using the clustering methods of the present paper to allow for latent groups with different factors that reflect differing behavior among the groups.

Second, the paper contributes to the literature on latent membership and clustering algorithms. There are presently several clustering algorithms (Bonhomme and Manresa, 2015; Ando and Bai, 2016; Bonhomme et al., 2017; Su et al., 2016, 2019; Wang et al., 2018), and most of the available methods apply only to stationary data. An exception is the C-Lasso approach (Huang et al., 2020, 2021), which is not directly applicable to the mixed-root panel autoregressive model. To the best of our knowledge, this study is the first attempt to extend clustering algorithms to the context of a mixed-root panel model that incorporates the group-specific nonstationary phenomena. Our approach follows Bonhomme and Manresa (2015) by using a two stage procedure. In our case the procedure combines an estimation procedure for group identities to determine latent membership in the first stage and a bubble testing procedure in the second stage, both in the context of a mixed-root panel model.

Third, the present work contributes to random coefficient panel modeling where coefficient heterogeneity occurs across individuals in the panel (Pesaran and Smith, 1995; Hsiao et al., 2002; Pesaran, 2006; Arellano and Bonhomme, 2012; Hsiao, 2014). One strand of that literature deals with nonstationary panels where there is random autoregressive coefficient heterogeneity in a dynamic panel. In that framework unit root testing is conducted by pooling the cross section under the null hypothesis that the autoregressive coefficients have unit mean (Westerlund and Larsson, 2012). In our model the autoregressive coefficients may deviate from unity in a way that produces coefficient heterogeneity across groups and homogeneity within groups in the panel, thereby leading to a mixed-root panel framework. The nonstationary elements allow for unit roots and explosive roots in different clusters, so that subgroups of the panel can manifest very different time series behavior. The framework is therefore suited to large panels in which

some clusters may manifest wandering behavior of the unit root type and other groups manifest various degrees of explosiveness or near stationarity.

Fourth, the paper contributes to the literature on the estimation of the number of groups. At present, group number is usually selected using an Information Criterion (IC) and model specification tests. The IC, which balances model fitness and penalty, can consistently estimate the number of groups in both stationary panel models (Bonhomme and Manresa, 2015; Bonhomme et al., 2017; Su et al., 2016; Miao et al., 2020; Okui and Wang, 2021; Wang and Su, 2021) and panel cointegration models (Huang et al., 2020, 2021). Model specification tests can be categorized into two cases – multiple groups (Lu and Su, 2017) and a single group (Pesaran et al., 1996; Phillips and Sul, 2003; Pesaran and Yamagata, 2008; Su and Chen, 2013). Both IC and residual-based inference approaches are found to underestimate the true number of groups in the mixed-root dynamic panel model. To address this limitation and consistently select the true number of groups in this setting, a novel method is proposed that combines the IC approach with a Hausman-type specification test.

A final contribution relates to the literature of bias-corrected procedures in panel data models. The presence of incidental parameters is well known to produce bias in many panel settings, especially dynamic panels (Hahn and Kuersteiner, 2002; Hahn and Newey, 2004; Gouriéroux et al., 2010) and nonlinear panels (Hahn and Newey, 2004; Arellano et al., 2007). Bias correction methods include the use of explicit bias approximations (Phillips and Moon, 1999; Hahn and Kuersteiner, 2002), jackknife methods (Hahn and Newey, 2004; Dhaene and Jochmans, 2015), and indirect inference methods (Gouriéroux et al., 2010). The limit theory in the present paper involves a new asymptotic bias term that originates from the demeaning process and the presence of serially correlated errors. An explicit expression of this bias is obtained under the null hypothesis of a group-specific unit root, which enables the construction of asymptotically pivotal tests of explosive behavior in subgroups of the panel.

The rest of this paper is organized as follows. Section 2 discusses the model setup. Section 3 introduces a two stage method consisting of the recursive k-means clustering algorithm in the first stage and bubble testing statistics in the second stage. Section 4 derives the asymptotic properties of the two stage procedure and establishes pivotal limit theory for the post-clustering test statistics under the null hypothesis of a group-specific unit root. Section 5 reports simulation findings that explore the finite sample performance of the two stage procedure and tests. Section 6 provides empirical applications of the methodology to the Chinese and US real estate markets and the US stock market. Section 7 concludes.

Throughout the paper, the symbols I_d, $\ell_{d \times 1}$, $\mathbf{0}_{d \times d}$, $\rightarrow_p$, $\Rightarrow$ and $\Pr(A)$ denote the $d \times d$ identity matrix, a d-vector of ones, a $d \times d$ matrix of zeros, convergence in probability, weak convergence in Euclidean and function spaces, and the probability of event A. For two sequences A_{nT} and B_{nT}, the notation $A_{nT} \preceq B_{nT}$ signifies that A_{nT}/B_{nT} is either $O_p(1)$ or $o_p(1)$ as $(n, T) \rightarrow \infty$; $A_{nT} \succ B_{nT}$ signifies that $B_{nT}/A_{nT} = o_p(1)$ as $(n, T) \rightarrow \infty$; $A_{nT} \sim B_{nT}$

signifies $\lim_{T\to\infty} A_{nT}/B_{nT} = 1$, $A_{nT} \sim_a B_{nT}$ denotes $\Pr(|A_{nT}/B_{nT}| \neq 1) \to 0$ as $(n,T) \to \infty$; the notation $\log_2(\cdot)$ represents $\log(\log(\cdot))$, and a zero affix on a parameter, as in $\{a^0\}$, refers to the true value of the corresponding parameter $\{a\}$. The notations *Avar* and *Acov* represent asymptotic variance and asymptotic covariance.

2 Model Setup

To capture explosive and midly explosive behavior in panels we use the following data generating process (DGP) based on the time series model of Phillips and Magdalinos (2007b)

$$\begin{cases} y_{it} &= \mu_i + \rho_{g_i} y_{i,t-1} + u_{it}, \ i = 1,...n, \ t = 1,...,T, \\ \rho_{g_i} &= 1 + \frac{c_{g_i}}{T^\gamma}. \end{cases} \tag{4}$$

The rate exponent $\gamma \in (0,1)$[1] and the scale coefficients c_{g_i} both influence the extent of departure of the autoregressive coefficients ρ_{g_i} from unity, and g_i denotes the group membership of individual i, for which the group structure is defined later. The innovations u_{it} follow a stationary linear process (i.e., $I(0)$ for each i and are defined later in (27) of Assumption 1. Long run variances are given by $\overline{\omega}_i^2 = \sum_{h=-\infty}^{\infty} \mathbb{E}(u_{it}u_{i,t-h})$, one-sided long run covariances by $\overline{\lambda}_i \left(:= \sum_{h=1}^{\infty} \mathbb{E}(u_{it}u_{i,t-h})\right)$, and variances by $\overline{\sigma}_{iu}^2 = \mathbb{E}(u_{it}^2)$, so that $\overline{\omega}_i^2 = 2\overline{\lambda}_i + \overline{\sigma}_{iu}^2$ for each individual unit i.

Model (4) mixes three types of potential time series behavior depending on the sign and value of the autoregressive coefficient, covering mildly explosive roots (with $c_{g_i} > 0$ and $\rho_{g_i} > 1$), mildly integrated roots (with $c_{g_i} < 0$ and $\rho_{g_i} < 1$), and unit roots (with $c_{g_i} = 0$ and $\rho_{g_i} = 1$). Since the signs of the $\{c_{g_i}\}_{i=1}^n$ determine the presence or absence of bubble behavior it is convenient to assume a common unknown value of the rate coefficient γ in (4) and then heterogeneity in the autoregressive coefficients ρ_{g_i} arises through the localizing scale parameters $\{c_{g_i}\}_{i=1}^n$. Latent group membership of the ρ_{g_i} is therefore determined by the value of these localizing scale coefficients. The signs of the c_{g_i} and their magnitudes determine the nature and strength of the mildly explosive and mildly integrated character of the individual time series.

The framework we adopt lies between a homogeneous panel (where $c_{g_i} = c$ for all i) and a fully heterogeneous panel (where $c_{g_i} \neq c_{g_\ell}$ for any $i \neq \ell$). Instead, we assume a group structure involving a fixed number $G < n$ of unknown separate groups that are classified according to the scale parameters c_{g_i}. The group membership variables are given by the $\{g_i\}_{i=1}^n$ which map individual units (i.e. $i \in \{1,2,...,n\}$) into specific groups for which $j \in \{1,...,G\}$ with $G < n$. This group structure allows for several possible mildly explosive and mildly integrated groups together with a unit root group. The mixed-root groups are determined by the signs and values of the scale coefficients and these are represented

---

[1]The case where $\gamma = 0$ is a specialization of the current model and remains suited to the two-stage algorithm developed here. The asymptotic theory of the two-stage algorithm when $\gamma = 0$ follows similar lines to that the main article and details are provided in the Online Supplement.

in the following diagram:

$$
\left\{
\begin{array}{lll}
\text{Explosive groups:} & \left\{
\begin{array}{ll}
\text{Group 1:} & c_1 > 0 \\
\text{Group 2:} & c_2 > 0 \\
\quad \vdots & \quad \vdots \\
\text{Group } g: & c_g > 0
\end{array}
\right. & \\
\text{Unit root group:} & \quad \text{Group } (g+1): \quad c_{g+1} = 0\,, & \qquad (5) \\
\text{Stationary groups:} & \left\{
\begin{array}{ll}
\text{Group } (g+2): & c_{g+2} < 0 \\
\text{Group } (g+3): & c_{g+3} < 0 \\
\quad \vdots & \quad \vdots \\
\text{Group } G: & c_G > 0
\end{array}
\right. &
\end{array}
\right.
$$

where $c_j \neq c_k$ for any $j \neq k$ with indices $j,k \in \{1,2,...,G\}$. The localizing scale coefficients are therefore homogeneous within each group but heterogeneous across groups. There are G groups in total: g mildly explosive groups each with a different scale coefficient $\{c_j > 0 | j = 1,...,g\}$; a single unit root group; and $(G - g - 1)$ mildly stationary groups, each with a different scale coefficient $\{c_j < 0 | j = g + 2,...,G\}$.

We begin by fixing notation. Denote the full set of n individuals by $\mathcal{I}_n := \{1,2,...,n\}$ and membership indicators by the parameter vector $\delta\left(:= (g_1,g_2,...,g_n)'\right)$. The true membership indicators are given by $\delta^0\left(:= \left(g_1^0, g_2^0,...,g_n^0\right)'\right)$. The estimated membership indicator, defined later, is $\widehat{\delta}\left(:= (\widehat{g}_1,\widehat{g}_2,...,\widehat{g}_n)'\right)$. Hence, for any individual subscript $i \in \mathcal{I}_n$, the membership indicators g_i, g_i^0, and $\widehat{g}_i$ all map from the set of individuals $\mathcal{I}_n$ to the set of group identities $\mathcal{G} := \{1,2,...,G\}$ with $\mathcal{G}(j)$ representing the jth group for any $j \in \mathcal{G}$. Let Δ_G be the set of all possible mappings from $\mathcal{I}_n$ to $\mathcal{G}$, so that $\delta, \delta^0, \widehat{\delta} \in \Delta_G$. As indicated, the notation $\mathcal{G}(j)$ is used to represent the jth group, with $\mathcal{G}^0(j)$ being the true jth group and $\hat{\mathcal{G}}(j)$ the estimated jth group to be defined later. We also define the collection $\mathcal{G}^0 := \left\{1,2,...,G^0\right\}$ where G^0 denotes the true number of groups.

Note that for any $j \in \mathcal{G}$, the distancing parameter in the group $\mathcal{G}(j)$ is c_j and the slope coefficient parameter is ρ_j. Let $\mathcal{C}_G$ be a compact subset of G-dimensional Euclidean space $\mathbb{R}^G$, $c\left(:= \left(c_1,...,c_j,...c_G\right)'\right) \in \mathcal{C}_G$ be the distancing parameter vector, and $\rho := \left(\rho_1,...,\rho_j,...,\rho_G\right)' = \left(1 + \frac{c_1}{T^\gamma},...,1 + \frac{c_j}{T^\gamma},...,1 + \frac{c_G}{T^\gamma}\right)'$ be the corresponding AR coefficient vector. Both c and ρ are G-dimensional vectors of group-specific parameters.

Within each group we impose an identical membership structure on the variances and covariances, so that for any $i,\ell \in \mathcal{I}_n$ with $g_i^0 = g_\ell^0 = j \in \mathcal{G}^0$, we have $\overline{\sigma}_{iu}^2 = \overline{\sigma}_{\ell u}^2 \left(=: \sigma_j^2\right)$, $\overline{\omega}_i^2 = \overline{\omega}_\ell^2 \left(=: \omega_j^2\right)$ and $\overline{\lambda}_i = \overline{\lambda}_\ell \left(=: \lambda_j\right)$. For the group-specific parameters σ_j^2, λ_j and ω_j^2, the oracle estimates, which rely on the true group identities, are denoted $\widehat{\sigma}_j^2$, $\widehat{\lambda}_j$ and $\widehat{\omega}_j^2$. Moreover, the post-clustering estimates, which rely on the estimated group identities, are denoted $\breve{\sigma}_j^2$, $\breve{\lambda}_j$ and $\breve{\omega}_j^2$ if the estimated membership is $\widehat{\delta} = (\widehat{g}_1,...,\widehat{g}_n)'$. Let the group-specific variances σ_j^2, λ_j, and ω_j^2 have true values $\left(\sigma_j^0\right)^2$, λ_j^0, and $\left(\omega_j^0\right)^2$. The true cardinality of the true jth group is given by n_j and the estimated cardinality of the estimated

jth group is given by $\check{n}_j := \sum_{i=1}^n \mathbf{1}_{\{\widehat{g}_i = j\}}$.

This paper also considers the individual distancing parameters and estimates

$$\bar{c}_i := c_{g_i}, \ \bar{c}_i^0 := c_{g_i^0}^0, \ \widehat{\bar{c}}_i := \check{c}_{\widehat{g}_i}, \ \forall i \in \mathcal{I}_n. \tag{6}$$

We define the following n-dimensional vectors of individual distancing parameters, their true values, and their estimated values as: $\bar{c} := (\bar{c}_1, ..., \bar{c}_n)'$, $\bar{c}^0 := \left(\bar{c}_1^0, ..., \bar{c}_n^0\right)'$, $\widehat{\bar{c}} := \left(\widehat{\bar{c}}_1, ..., \widehat{\bar{c}}_n\right)'$. Similar representations apply to $\bar{\rho}$, $\bar{\rho}^0$ and $\widehat{\bar{\rho}}$. For any $i \in \mathcal{I}_n$ and $j \in \mathcal{G}^0$, the estimates of individual variances are given by $\widehat{\bar{\sigma}}_{iu}^2$, $\widehat{\bar{\omega}}_i^2$ and $\widehat{\bar{\lambda}}_i$, which are time series variance estimates based on the post-clustering estimate $\check{\rho}_j$ with $\widehat{g}_i = j$. Correspondingly, the true values of the individual-specific variances, individual-specific one-sided and two-sided long run variances are denoted by

$$\left(\bar{\sigma}_{iu}^0\right)^2 := \left(\sigma_j^0\right)^2, \ \bar{\lambda}_i^0 := \lambda_j^0, \ \left(\bar{\omega}_i^0\right)^2 := \left(\omega_j^0\right)^2. \tag{7}$$

3 A Two Stage Approach

Econometric analysis of the model given in (4) and (5) employs a two stage approach. The first stage uses recursive k-means clustering to estimate the underlying group structure. In the second stage, post-clustering estimates of the parameters of interest are obtained and new tests for bubble detection are developed. It is convenient at first to assume that the true value of the number of groups, G^0, is known. A hybrid selection method that combines an information criterion (IC) and a Hausman-type specification test is designed later to enable consistent estimation of G^0, and thereby the full group structure.

3.1 Stage 1: a recursive k-means clustering algorithm

When groups are unobserved two types of parameters are considered in distinguishing membership – the group membership variable δ, which maps cross-sectional units into groups, and the G^0-dimensional distancing parameter vector c. Similar to Bonhomme and Manresa (2015), estimates $\widehat{c}^*$ and $\widehat{\delta}$ (and hence $\{\widehat{g}_i\}_{i=1}^n$) are obtained by extremum estimation, viz.,

$$\left(\widehat{c}^*, \widehat{\delta}\right) = \arg\min_{(c,\delta) \in \mathcal{C}_{G^0} \times \Delta_{G^0}} \frac{1}{n} \sum_{i=1}^n \frac{1}{\Upsilon_{iT}} \left[\sum_{t=1}^T \left(\widetilde{y}_{it} - \widetilde{y}_{i,t-1} \left(1 + \frac{c_{g_i}}{T^\gamma} \right) \right)^2 \right], \tag{8}$$

where $\Upsilon_{iT} := \sum_{t=1}^T \widetilde{y}_{i,t-1}^2$. The demeaned variables $\widetilde{y}_{it} \left(:= y_{it} - \bar{y}_i\right)$ and $\widetilde{y}_{i,t-1} \left(:= y_{it} - \bar{y}_{i,-1}\right)$, using the respective sample means $\bar{y}_i = \frac{1}{T} \sum_{t=1}^T y_{it}$ and $\bar{y}_{i,-1} = \frac{1}{T} \sum_{t=1}^T y_{i,t-1}$, are employed to eliminate fixed effects.

Instead of estimating c and δ simultaneously as in (8), which is numerically challenging, we use an iterative strategy (Bonhomme and Manresa, 2015), given in Algorithm 1 below, to estimate $\widehat{c}^*$ and $\widehat{\delta}$ recursively. For convenience in the following derivations,

knowledge of $\gamma \in (0,1)$ is treated as prior information.[2] This assumption is partly justified by the fact that in a time series sample with an autoregressive root $\rho = 1 + \frac{c}{T^\gamma}$ the localizing rate and localizing scale parameters (c,γ) are not jointly identifiable from ρ but each is clearly identified given the other.[3] Moreover, variation of c facilitates estimation and provides a full range of possibilities for the autoregressive coefficient ρ, while ensuring near unit root behavior when c is fixed as $T \to \infty$.

Algorithm 1 Recursive procedure to estimate c and δ

(i) Set $s = 0$. Obtain the individual time series estimates $\widehat{\overline{\rho}}_i^{TS}$ of the 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 γ. Using any relevant prior information or selective quantiles of the time series estimates, or by 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. For instance, in our later empirical analysis of the Chinese housing market, the initial values $\{c_j^{(0)}\}_{1 \le j \le 3}$ for the k-means algorithm are chosen as the 30%, 70%, and 90% quantiles of the individual time series estimates, $\{\widehat{\overline{\rho}}_i^{TS}\}_{1 \le i \le n}$.

(ii) Given $c^{(s)}$, for $i \in \mathcal{I}_n$ compute the extremum estimate of g_i

$$g_i^{(s+1)} = \arg\min_{j \in \mathcal{G}^0} \left[\sum_{t=1}^{T} \left(\widetilde{y}_{it} - \widetilde{y}_{i,t-1} \left(1 + \frac{c_j^{(s)}}{T^\gamma} \right) \right)^2 \right]. \tag{9}$$

(iii) Given $\{g_i^{(s+1)}\}_{i=1}^n$, compute the extremum estimate of c

$$c^{(s+1)} = \arg\min_{c \in \mathcal{C}_{G^0}} \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\Upsilon_{iT}} \left[\sum_{t=1}^{T} \left(\widetilde{y}_{it} - \widetilde{y}_{i,t-1} \left(1 + \frac{c_{g_i^{(s+1)}}}{T^\gamma} \right) \right)^2 \right]. \tag{10}$$

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

3.2 Stage 2: post-clustering estimation and testing

Denote the true collection of members of the jth group as

$$\mathcal{G}^0(j) = \left\{ i \in \mathcal{I}_n \mid g_i^0 = j \right\} \quad \forall j \in \mathcal{G}^0.$$

---

[2]Numerical simulations show that the estimated groupings are robust to various choices of the true scaling parameter γ^0.

[3]See Phillips (2021) for more details of parameter identification, estimation and inference in mildly integrated and mildly explosive models.

Suppose the estimated membership indicator vector is $\widehat{\delta} = (\widehat{g_1}, \widehat{g_2}, ..., \widehat{g_n})'$. Denote the estimated members of the jth group by

$$\hat{\mathcal{G}}(j) = \{i \in \mathcal{I}_n | \widehat{g_i} = j\} \; \forall j \in \mathcal{G}^0.$$

We consider two pooled LS estimators for ρ_j, namely, the oracle estimator $\widehat{\rho}_j$ and the post-clustering estimator $\check{\rho}_j$. The oracle estimator, that employs data from the true jth group $\mathcal{G}^0(j)$, is given by

$$\widehat{\rho}_j = \frac{\sum_{i \in \mathcal{G}^0(j)} \sum_{t=1}^{T} \widetilde{y}_{i,t-1} \widetilde{y}_{it}}{\sum_{i \in \mathcal{G}^0(j)} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}^2}. \tag{11}$$

The post-clustering estimator, that uses data from the estimated jth group $\hat{\mathcal{G}}(j)$, is given by

$$\check{\rho}_j = \frac{\sum_{i \in \hat{\mathcal{G}}(j)} \sum_{t=1}^{T} \widetilde{y}_{i,t-1} \widetilde{y}_{it}}{\sum_{i \in \hat{\mathcal{G}}(j)} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}^2}. \tag{12}$$

Next define the following quantities

$$\check{\sigma}_j^2 = \frac{1}{\check{n}_j} \sum_{i \in \hat{\mathcal{G}}(j)} \widehat{\sigma}_{iu}^2, \; \check{\omega}_j^2 = \frac{1}{\check{n}_j} \sum_{i \in \hat{\mathcal{G}}(j)} \widehat{\omega}_i^2, \; \check{\lambda}_j = \frac{1}{\check{n}_j} \sum_{i \in \hat{\mathcal{G}}(j)} \widehat{\lambda}_i, \; \check{E}_{j,nT} = \sum_{i \in \hat{\mathcal{G}}(j)} E_{i,nT}, \tag{13}$$

where

$$\widehat{\omega}_i^2 = \frac{1}{T} \sum_{t=1}^{T} \widehat{u}_{it}^2 + \frac{2}{T} \sum_{l=1}^{L} \sum_{t=l+1}^{T} w(l,L) \widehat{u}_{it} \widehat{u}_{i,t-l}, \tag{14}$$

$$\widehat{\sigma}_{iu}^2 = \frac{1}{T} \sum_{t=1}^{T} \widehat{u}_{it}^2, \quad \widehat{\lambda}_i = \frac{1}{T} \sum_{l=1}^{L} \sum_{t=l+1}^{T} w(l,L) \widehat{u}_{i,t-l} \widehat{u}_{it}, \tag{15}$$

$$E_{i,nT} = \sum_{t=1}^{T} \varphi_{i,t}^2 + 2 \sum_{l=1}^{L} \sum_{t=l+1}^{T} w(l,L) \varphi_{i,t} \varphi_{i,t-l}, \tag{16}$$

$$\varphi_{i,t} = \widetilde{y}_{i,t-1} \widetilde{\widehat{u}}_{i,t} - \overline{\varphi}_j, \overline{\varphi}_j = \frac{1}{\check{n}_j T} \sum_{i \in \hat{\mathcal{G}}(j)} \sum_{t=1}^{T} \widetilde{y}_{i,t-1} \widetilde{\widehat{u}}_{i,t},$$

$$\widehat{u}_{it} = y_{it} - \check{\rho}_j y_{i,t-1}, \widetilde{\widehat{u}}_{i,t} = \widetilde{y}_{it} - \check{\rho}_j \widetilde{y}_{i,t-1}, \text{ with } i \in \hat{\mathcal{G}}(j),$$

$$w(l,L) = 1 - \frac{l}{L+1}.$$

The quantities defined in (13) are variance, long run variance, long run one-sided covariance, and bias estimates, constructed from the regression residuals $\widehat{u}_{it}$ and $\widetilde{\widehat{u}}_{i,t}$. The long run quantities are constructed using the Bartlett window $w(l,L)$.

Based on the membership and variance estimates, we proceed to implement testing procedures to detect group-specific explosiveness. To test the null hypothesis $\mathcal{H}_0^{(j)} : c_j^0 = 0$

against $\mathcal{H}_1^{(j)} : c_j^0 > 0$ for any j, we provide two statistics, denoted by the following panel J and panel t statistics

$$\widetilde{J}_j = \sqrt{\frac{\check{n}_j}{3}} T \left(\check{\rho}_j - 1 - \frac{\check{n}_j T \check{\lambda}_j}{\check{D}_{j,nT}} + \frac{\check{n}_j T \check{\omega}_j^2}{2\check{D}_{j,nT}} \right), \ \widetilde{t}_j = \frac{\left(\check{\rho}_j - 1 - \frac{\check{n}_j T \check{\lambda}_j}{\check{D}_{j,nT}} + \frac{\check{n}_j T \check{\omega}_j^2}{2\check{D}_{j,nT}} \right) \check{D}_{j,nT}}{\check{\omega}_j \sqrt{\check{E}_{j,nT}}}, \tag{17}$$

in which $\check{D}_{j,nT} := \sum_{i \in \hat{\mathcal{G}}(j)} \sum_{t=1}^T \widetilde{y}_{i,t-1}^2$, and $\check{E}_{j,nT}$ is defined in (16).

Note that there are two bias correction terms separately introduced by serial correlations and demeaned variables in each test. The term $-(\check{n}_j T \check{\lambda}_j)/\check{D}_{j,nT}$ removes the bias of the stationary linear process while the term $(\check{n}_j T \check{\omega}_j^2)/(2\check{D}_{j,nT})$ eliminates the bias caused by demeaning variables. To the best of our knowledge, the explicit forms of these bias terms are novel and are derived here for the first time, although they have a clear precedent in Phillips and Magdalinos (2007b) in the time series context. The above findings help to enhance our understanding of bias correction procedures in dynamic panel models. In particular, beyond incidental parameter problems and the bias generated by the presence of nonlinear functions, serial correlation in the component innovations can lead to non-negligible additional bias, coupled with inferential issues that need treatment to ensure asymptotically pivotal tests. These adjustments are especially needed in near unit root cases.

A significant advantage of these panel tests is the potential power gains from cross section aggregation within a homogeneous cluster of individual time series that can enhance their discriminatory power for bubble detection. By comparison the recently developed panel approach of Chen et al. (2022) focuses on the possible presence of a common single explosive factor extracted by principal component analysis. This method has the advantage of allowing for individual weighting and cross section dependence but it does not enhance discriminatory power. However, the factor model approach might be modified by the use of clustering methods, similar to those used here, to gain power from group aggregation.

The test statistics in (17) are based on the entire sample. But further development of the methodology is possible to embed a real-time dating strategy for estimating the origination and collapse dates of financial bubbles analogous to the time series methods in Phillips et al. (2011, 2015a). As indicated above, cross section dependence (e.g., through interactive fixed effects) can be added to the mixed-root dynamic panel with latent membership, similar to second generation panel unit root testing (Moon and Perron, 2004; Bai and Ng, 2004, 2010; Westerlund, 2015). Such extensions involve non-trivial technical developments and are therefore left for future research.

3.3 Estimation of the group number

So far we have assumed that the true number of groups G^0 is known. In practice G^0 is unknown. When the group number is set to G, the estimated quantities $\widehat{\delta}$, $\check{\rho}_j$ and $\hat{\mathcal{G}}(j)$ are all dependent on G. For clarity they are therefore denoted by $\widehat{\delta}(G)$, $\check{\rho}_j(G)$ and $\hat{\mathcal{G}}(j, G)$. We

propose to estimate the true number of groups using a new methodology that combines an IC and a Hausman-type model specification test. In particular, we use IC to select the lower bound of the group number in the first step, where the IC function is defined as

$$\text{IC}(G) = \ln\left(\frac{1}{nT} \sum_{j=1}^{G} \sum_{i \in \hat{\mathcal{G}}(j,G)} \sum_{t=1}^{T} \left(\widetilde{y}_{it} - \widetilde{y}_{i,t-1} \check{\rho}_j(G)\right)^2\right) + \kappa_{nT} G, \tag{18}$$

with penalty $\kappa_{nT} G$ depending on the number of groups G and a tuning parameter κ_{nT} that satisfies the rate restriction

$$\kappa_{nT} + \frac{1}{nT\kappa_{nT}} \to 0. \tag{19}$$

The lower bound of the group number is set to the minimizer of the IC, that is,

$$\widetilde{G} = \arg \min_{G=1,2,\dots,G_{\max}} \text{IC}(G), \tag{20}$$

where $G_{\max}$ is a generic upper bound of G.

It is well known that the IC function defined in (18) can consistently select the true number of groups or the true number of factors in many contexts (Bai and Ng, 2002; Bonhomme and Manresa, 2015). But as discussed in Remark 3.1 and proved in Theorem 4.4 below, when mildly stationary groups, a unit root group, and mildly explosive groups are present in the panel data, $\widetilde{G} \leq G^0$ with probability approaching 1, so that $\widetilde{G}$ may underestimate G^0 even in the limit.

To produce a consistent estimator of G^0, after $\widetilde{G}$ is obtained by IC, we propose a new Hausman-type specification test to assess slope homogeneity when G is assumed to be the number of groups for all $G \in \left\{\widetilde{G}, \widetilde{G}+1, \dots, G_{\max}\right\}$. Assuming there are G groups, the recursive k-means algorithm is applied to each estimated group j for $j \in \{1, 2, \dots, G\}$ by assuming that each group j has at most $\overline{G} := (G_{\max} - G + 1)$ subgroups. In the subgroup analysis of the estimated jth group, the variables of interest are denoted by $\widehat{\delta}_j(G)$, $\check{\rho}_{j,h}(G)$, $\hat{\mathcal{G}}(j,h,G)$, $\check{n}_{j,h}$ and $\pi_{j,h}$ with $h = 1, 2, \dots, \overline{G}$. The notation becomes somewhat complex because of the groupings, subgroupings, and selection process. For precision we let $\widehat{\delta}_j(G) = \left(\widehat{g}_{j,1}(G), \widehat{g}_{j,2}(G), \dots, \widehat{g}_{j,\check{n}_j}(G)\right)'$ be the estimated membership of subgroups in the estimated jth group, $\check{\rho}_j(G) = \left(\check{\rho}_{j,1}(G), \check{\rho}_{j,2}(G), \dots, \check{\rho}_{j,\overline{G}}(G)\right)'$ be the estimated slopes in subgroups of the estimated jth group, $\hat{\mathcal{G}}(j,h,G)$ be the estimated individuals in the estimated hth subgroup of the estimated jth group, and $\check{n}_{j,h}$ be the estimated dimension of individuals in the estimated hth subgroup of the estimated jth group. In a mild abuse of notation, we continue to denote the post-clustering estimate with undemeaned variables as $\check{\rho}_j(G)$ and let

$$\check{\rho}_j(G) = \frac{\sum_{i \in \hat{\mathcal{G}}(j,G)} \sum_{t=1}^{T} y_{i,t-1} y_{it}}{\sum_{i \in \hat{\mathcal{G}}(j,G)} \sum_{t=1}^{T} y_{i,t-1}^2}, \ \check{\rho}_{j,h}(G) = \frac{\sum_{i \in \hat{\mathcal{G}}(j,h,G)} \sum_{t=1}^{T} y_{i,t-1} y_{it}}{\sum_{i \in \hat{\mathcal{G}}(j,h,G)} \sum_{t=1}^{T} y_{i,t-1}^2}, \ \text{and} \ \lim_{n,T} \frac{\check{n}_{j,h}}{\check{n}_j} \to \pi_{j,h}. \tag{21}$$

The idea behind this Hausman-type test is to detect unspecified parameter heterogeneity across subgroups in a recursive manner. Without losing generality, we discuss the case in which $\overline{G}$ subgroups are specified in the estimated jth group since subgroup divisions with dimension smaller than $\overline{G}$ can be addressed in the same fashion as the following approach. Under the null hypothesis of slope homogeneity in the jth group, the joint asymptotic theory (i.e. $(n, T) \to \infty$) that will be discussed later in Section 4 shows

$$\sqrt{\check{n}_j T^{1+\gamma}} \left(\check{\rho}_j(G) \cdot \ell_{\overline{G} \times 1} - \check{\boldsymbol{\rho}}_j(G) \right) \Rightarrow \mathcal{N} \left(\mathbf{0}_{\overline{G} \times 1}, \ -2c_j^0 \left(\widetilde{\pi}_j^{-1} - I_{\overline{G}} \right) \right), \text{ if } c_j^0 < 0; \tag{22}$$

$$\sqrt{\check{n}_j T^2} \left(\check{\rho}_j(G) \cdot \ell_{\overline{G} \times 1} - \check{\boldsymbol{\rho}}_j(G) \right) \Rightarrow \mathcal{N} \left(\mathbf{0}_{\overline{G} \times 1}, \ 2 \left(\widetilde{\pi}_j^{-1} - I_{\overline{G}} \right) \right), \text{ if } c_j^0 = 0; \tag{23}$$

$$\sqrt{\check{n}_j T^{2\gamma} \left(\rho_j^0 \right)^{2T}} \left(\check{\rho}_j(G) \cdot \ell_{\overline{G} \times 1} - \check{\boldsymbol{\rho}}_j(G) \right) \Rightarrow \mathcal{N} \left(\mathbf{0}_{\overline{G} \times 1}, \ 4 \left(c_j^0 \right)^2 \left(\widetilde{\pi}_j^{-1} - I_{\overline{G}} \right) \right), \text{ if } c_j^0 > 0; \tag{24}$$

where $\widetilde{\pi}_j = \text{diag} \left\{ \pi_{j,1}, \pi_{j,2}, ..., \pi_{j,\overline{G}} \right\}$. We assume that $0 < \pi_{j,h} < 1$ for all $\{j, h\}$ under the null hypothesis of slope homogeneity in the jth group. Therefore, the Hausman-type statistic can be written

$$W_j(G) := \left(\check{\rho}_j(G) \ell_{\overline{G} \times 1} - \check{\boldsymbol{\rho}}_j(G) \right)' \left[\left(-I_{\overline{G}} + \widetilde{\pi}_j^{-1} \right) \check{\omega}_j^2 \check{D}_{j,nT}^{-1} \right]^{-1} \left(\check{\rho}_j(G) \ell_{\overline{G} \times 1} - \check{\boldsymbol{\rho}}_j(G) \right)$$
$$\Rightarrow \chi^2 \left(\overline{G} \right), \text{ under the null hypothesis of slope homogeneity in the } j\text{th group,} \tag{25}$$

where $\check{D}_{j,nT} := \sum_{i \in \hat{\mathcal{G}}(j,G)} \sum_{t=1}^{T} y_{i,t-1}^2$. Diminishing Type I error is achieved by implementing tests of slope homogeneity in each estimated group ($j = 1, 2, ..., G$) with a slowly diverging critical value of the form $cv_{nT} := (1 + b \log(nT)) \chi_{0.95}^2 \left(\overline{G} \right)$, where $\chi_{0.95}^2 \left(\overline{G} \right)$ is the 95% critical value of $\chi^2 \left(\overline{G} \right)$ and b is some positive constant.[4] We define the new estimator of the group number as $\widehat{G}$ defined by

$$\widehat{G} = \inf_{\overline{G} \leq G \leq G_{\max}} \left\{ G \mid W_j(G) \leq cv_{nT}, \text{ for any } j = 1, 2, ..., G \right\}. \tag{26}$$

For clarity, the procedure to estimate G^0 is summarized in Algorithm 2.

Remark 3.1 *Consistency of the IC procedure (Bai and Ng, 2002; Su et al., 2016; Bonhomme and Manresa, 2015) and residual-based model specification tests (Lu and Su, 2017) rely on the successful extraction of valid signals concerning potential model misspecifications from regression residuals. But the usual validity of these methods does not always hold for nonstationary models whose roots are close to unity. When the mixed-root panel model contains latent memberships, these procedures tend to underestimate the true number of groups. In particular, when individual time series follow a mildly stationary process, the usual theory and limit results change because the error variance can still be consistently estimated using an inconsistent estimate of the distance parameter $\overline{c}_i$. For instance, if $\overline{c}_i^0 < 0$ and an inconsistent time series estimator $\widehat{\overline{c}}_i^{TS}$ with $\widehat{\overline{c}}_i^{TS} - \overline{c}_i^0 = O_p(1)$ is employed, we have*

$$\frac{1}{T} \sum_{t=1}^{T} \left(y_{it} - \widehat{\overline{\rho}}_i^{TS} y_{i,t-1} \right)^2$$

---

[4] The setting $b = 5$ was found to work well in both the simulations and the empirical work.

Algorithm 2 Recursive procedure to compute $\widehat{G}$

   (i) Optimize the IC function of equation (18) and estimate the lower bound of group number $\widetilde{G}$ via equation (20).

   (ii) Let $G = \widetilde{G}$.

   (iii) Implement the recursive k-means algorithm in Algorithm 1. Set $j = 1$.

   (iv) Obtain the Hausman-type statistic of equation (25) for the estimated jth group. If the test statistic exceeds cv_{nT}, then set $G = G+1$ and go back to Step (iii). If the test statistic is smaller than cv_{nT}, then set $j = j+1$ and re-iterate Step (iv).

   (v) If the null hypothesis of group-specific slope homogeneity cannot be rejected in each group $j = 1, 2, ..., G$ and $G < G_{\max}$, set $\widehat{G} = G$. If $G = G_{\max}$, set $\widehat{G} = G_{\max}$.

$$= \frac{1}{T}\sum_{t=1}^{T} u_{it}^2 + \frac{2}{T^{1+\gamma}}\sum_{t=1}^{T} u_{it}y_{i,t-1}\left(\bar{c}_i^0 - \widehat{\bar{c}}_i^{TS}\right) + \frac{1}{T^{1+2\gamma}}\sum_{t=1}^{T} y_{i,t-1}^2\left(\bar{c}_i^0 - \widehat{\bar{c}}_i^{TS}\right)^2$$

$$= \frac{1}{T}\sum_{t=1}^{T} u_{it}^2 + O_p\left(\frac{1}{T^{\frac{1+\gamma}{2}}}\right) + O_p\left(\frac{1}{T^{\gamma}}\right) \to_p \left(\sigma_{iu}^0\right)^2,$$

where $\widehat{\bar{\rho}}_i^{TS} = 1 + \widehat{\bar{c}}_i^{TS}/T^{\gamma}$. Thus, when mildly stationary individual time series are miss-clustered into other groups, the sample variance of regression residuals can still consistently estimate the error variance. This property violates a key requirement of model selection, e.g., Assumption A.4 of Su et al. (2016), explaining the need to develop an alternative procedure based on a Hausman-type test that can correctly select the true number of groups.

4   Asymptotic Theory

This section develops the asymptotic properties of the two stage procedure for the mixed-root panel autoregressive model given in (4) and (5). We first establish the uniform consistency of the recursive k-means clustering method so that the estimated membership is asymptotically identical to the true membership. The post-clustering estimators of the AR coefficients are then shown to be asymptotically equivalent to the oracle estimators that employ the true group identities and the right-tailed panel J and t tests are shown to have pivotal limit distributions under the null hypothesis of a group-specific unit root. Consistency is demonstrated for the group number estimator based on the combined use of IC and Hausman-type tests, so that $\widehat{G}$ correctly selects the true number of groups G^0 in large samples.

    To establish the asymptotic theory of the two stage procedure, we first impose the following two assumptions to facilitate the development.

Assumption 1   *(i) For any $i \in \mathcal{I}_n$, the individual fixed effects $\mu_i = O_p\left(T^{-1}\right)$.*

(ii) The equation errors u_{it} follow stationary linear processes

$$u_{it} = \sum_{h=0}^{\infty} F_{ih}\epsilon_{i,t-h} = F_i(L)\epsilon_{it}, \tag{27}$$

in which the operator $F_i(z)\left(:= \sum_{h=0}^{\infty} F_{ih}z^h\right)$ contains a series of deterministic coefficients $\{F_{ih}\}_{h=0}^{\infty}$ with $F_{i0} = 1$, for any $i \in \mathcal{I}_n$. The innovations $\{\epsilon_{it}\}$ in (27) are $iid\left(0,\left(\sigma^0\right)^2\right)$ over t with $\left(\sigma^0\right)^2 > 0$ for each i and independent across i with uniformly bounded finite qth $(q \geq 4)$ moments, $\sup_i \mathbb{E}|\epsilon_{it}|^q < \infty$. The summability restrictions

$$\sum_{h=0}^{\infty} h|F_{ih}| < \infty, \tag{28}$$

hold uniformly over $i \in \mathcal{I}_n$. If $g_i^0 = g_\ell^0$ for any $i,\ell \in \mathcal{I}_n$, then $F_{ih} = F_{\ell h}$ for all h. For individuals $i \in \mathcal{I}_n$ with $\bar{c}_i^0 < 0$, then $F_{ih} = 0$ for all $h \geq 1$ and $u_{it} = \epsilon_{it}$.

(iii) Assume $y_{i0} = 0$ for any $i \in \mathcal{I}_n$ and $u_{is} = 0$ for any $i \in \mathcal{I}_n$ and $s \leq 0$.

(iv) There exist $c_{low}, c_{up} > 0$ for which $c_j^0 \in \mathcal{C} := \left[-c_{up}, -c_{low}\right] \cup \{0\} \cup \left[c_{low}, c_{up}\right], \forall j \in \mathcal{G}^0$.

(v) There exists a constant $\dot{c} \in (0,\infty)$ such that, for any $j, j' \in \mathcal{G}^0$,

$$\inf_{j \neq j'} \left|c_j^0 - c_{j'}^0\right| \geq \dot{c}.$$

Assumption 1(i) provides restrictions on the individual fixed effects that ensure drift effects from the equation intercept are asymptotically negligible in all cases. Assumption 1(ii) provides for linear process equation errors u_{it} with group-specific homogeneity that facilitates the limit theory (Phillips and Solo, 1992). Assumption 1 (ii) assumes cross-sectional independence and possible heterogeneity over $i \in \mathcal{I}_n$ with uniform moment conditions that facilitate the development of limit theory for cross section averages. Further enhancements to this framework that allow for cross section dependence are possible and will be considered in future work.

Under Assumption 1(ii), the error process $\{u_{it}\}$ admits the Beveridge-Nelson decomposition, namely,

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

where $\widetilde{\epsilon}_{it} = \sum_{h=0}^{\infty} \widetilde{F}_{ih}\epsilon_{i,t-h}$ and $\widetilde{F}_{ih} = \sum_{k=h+1}^{\infty} F_{ik}$. The summability condition $\sum_{h=0}^{\infty} |\widetilde{F}_{ih}| < \infty$ is satisfied by (28), which in turn assures functional laws hold for the partial sum processes $S_{it} = \sum_{s=1}^{t} u_{is}$ (Phillips and Solo, 1992)

$$B_{i,T^\gamma}(\cdot) = \frac{S_{i,\lfloor T^\gamma \cdot \rfloor}(\cdot)}{T^{\gamma/2}} = \frac{1}{T^{\gamma/2}} \sum_{s=1}^{\lfloor T^\gamma \cdot \rfloor} u_{is} \Rightarrow B_i(\cdot),$$

16

for all i where the $B_i(\cdot)$ are Brownian motions with variance $\left(\overline{\omega}_i^0\right)^2$.

Assumption 1(iii) details simple initial conditions and these may be considerably weakened without changing the limit theory, as shown in time series settings (Phillips and Magdalinos, 2009), at the expense of further notational complications. Assumption 1(iv) imposes an identification condition that ensures a bounded support for the distancing parameter vector c so that, employing the earlier support notation $\mathcal{C}_{G^0}$, we have

$$\mathcal{C}_{G^0} := \times_1^{G^0} \mathcal{C}, \ \text{ with } \ \mathcal{C} := \left[-c_{up}, -c_{low}\right] \cup \{0\} \cup \left[c_{low}, c_{up}\right].$$

Assumption 1(v) gives another identification condition in which the group-specific parameters are well separated and ensures that the recursive k-means algorithm can specify each group under the joint asymptotic scheme (i.e., with both $n, T \to \infty$).

Assumption 2   *(i) As $n \to \infty$,*

$$\frac{n_j}{n} \to \pi_j \in (0, 1],$$

where π_j is a constant value for any $j \in \mathcal{G}^0$. Moreover,

$$\inf_{j \in \mathcal{G}^0} \pi_j \geq \underline{M} > 0,$$

for some constant $\underline{M} > 0$.

(ii) The following rate restrictions hold[5]: $\frac{T^{\gamma \wedge (1-\gamma)}}{(\log_2 T)^2} \succ n$, $\frac{T^{3-4\gamma}}{(\log_2 T)^8} \succ n^2$, $\frac{T}{(\log_2 T)^8} \succ n(\log_2 n)^2$, $\gamma < \frac{1}{3}$ and $n \succ T^{5\gamma - 1}(\log T)^{16}$.

(iii) The truncation for the long run variance estimates, L, satisfies the condition that $L = o\left(T^{\frac{1}{3}}\right)$ and $L = o\left(T^{2\gamma}\right)$.

Assumption 2(i) implies that the cardinality of each group increases proportionally to the full dimension of the cross section. If we denote

$$\sigma_0^2 := \sum_{j=1}^{G^0} \pi_j \left(\sigma_j^0\right)^2,$$

then, under Assumption 2(i), we have

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \left(\overline{\sigma}_{iu}^0\right)^2 = \lim_{n \to \infty} \sum_{j=1}^{G^0} \frac{n_j}{n} \left(\frac{1}{n_j} \sum_{i \in \mathcal{G}^0(j)} \left(\sigma_j^0\right)^2\right) = \sum_{j=1}^{G^0} \pi_j \left(\sigma_j^0\right)^2 = \sigma_0^2 > 0. \qquad (29)$$

The reason $\sigma_0^2 > 0$ is that by assumption $\left(\sigma^0\right)^2 > 0$ which in turn ensures that $\left(\sigma_j^0\right)^2 > 0$ for all j.

Assumption 2(ii) imposes rate restrictions that ensure the clustering errors are negligible and the recursive k-means clustering algorithm is uniformly consistent. Assumption 2(iii) provides necessary conditions to ensure consistent estimates for the one-sided and two-sided long run variances.

---

[5]Recall that $A_{nT} \succ B_{nT}$ signifies that $B_{nT}/A_{nT} = o_p(1)$ as $(n, T) \to \infty$.

4.1 Clustering and estimation

First, uniform consistency of the recursive k-means clustering method is established, showing that the estimated membership is equivalent to the true membership under the joint asymptotic scheme, $(n, T) \to \infty$.

Theorem 4.1 *Suppose Assumptions 1 and 2 hold. When $(n, T) \to \infty$,*

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

Theorem 4.1 shows that we can correctly recover the latent group structures of the mixed-root panel autoregression model under the joint asymptotic framework. This result relies heavily on the identification conditions given in Assumptions 1(iv) and 1(v), based on which the computational algorithm of Bonhomme and Manresa (2015) applies to the present model. As long as the group-specific distancing parameters are well separated across groups and the support of parameters is compact, we can always estimate the latent membership of distancing parameters through the recursive k-means clustering algorithm.

With the clustered membership obtained from Stage 1, we employ the panel within estimate $\check{\rho}_j$ based on the estimated membership $\{\widehat{g}_i\}_{i=1}^n$. Since the clustering errors are asymptotically negligible, it is easy to show that the post-clustering estimator $\check{\rho}_j$ is asymptotically equivalent to the oracle estimator $\widehat{\rho}_j$ that uses the true group structure. Thus, for any $1 \le j \le G^0$, if $\widehat{\rho}_j$ and $\check{\rho}_j$ are defined as in (11) and (12), then

$$\sqrt{n_j}\left(\rho_j^0\right)^T T^\gamma \left(\check{\rho}_j - \rho_j^0\right) = \sqrt{n_j}\left(\rho_j^0\right)^T T^\gamma \left(\widehat{\rho}_j - \rho_j^0\right) + o_p(1), \text{ if } c_j^0 > 0, \tag{30}$$

$$\sqrt{n_j}T\left(\check{\rho}_j - 1 + \frac{3\left(\sigma_j^0\right)^2}{\left(\omega_j^0\right)^2}\frac{1}{T}\right) = \sqrt{n_j}T\left(\widehat{\rho}_j - 1 + \frac{3\left(\sigma_j^0\right)^2}{\left(\omega_j^0\right)^2}\frac{1}{T}\right) + o_p(1), \text{ if } c_j^0 = 0, \tag{31}$$

and

$$\sqrt{n_j}T^{\frac{1+\gamma}{2}}\left(\check{\rho}_j - \rho_j^0 - \frac{1}{T^\gamma}\frac{-2c_j^0}{\left(\omega_j^0\right)^2}\left(\lambda_j^0 + \frac{c_j^0}{T^\gamma}\overline{m}_{j,T}\right)\right)$$

$$= \sqrt{n_j}T^{\frac{1+\gamma}{2}}\left(\widehat{\rho}_j - \rho_j^0 - \frac{1}{T^\gamma}\frac{-2c_j^0}{\left(\omega_j^0\right)^2}\left(\lambda_j^0 + \frac{c_j^0}{T^\gamma}\overline{m}_{j,T}\right)\right) + o_p(1), \text{ if } c_j^0 < 0, \tag{32}$$

where $\overline{m}_{j,T}$ in (32) denotes a non-negligible bias element whose explicit form is given by (36) in Theorem 4.2. Based on (30), (31) and (32), the asymptotic theory may be obtained from that of the oracle estimator based on the true group membership. Under the joint convergence framework $(n, T) \to \infty$, the following theorem provides the Gaussian limit theory of the post-clustering estimator $\check{\rho}_j$ of the jth estimated group.

Theorem 4.2 *Suppose Assumptions 1 and 2 hold. When $(n, T) \to \infty$,*

$$\sqrt{n_j} \left(\rho_j^0 \right)^T T^\gamma \left(\check{\rho}_j - \rho_j^0 \right) \Rightarrow \mathcal{N} \left(0, 4 \left(c_j^0 \right)^2 \right), \text{ if } c_j^0 > 0; \tag{33}$$

$$\sqrt{n_j} T \left(\check{\rho}_j - 1 + \frac{3 \left(\sigma_j^0 \right)^2}{\left(\omega_j^0 \right)^2} \frac{1}{T} \right) \Rightarrow \mathcal{N}(0, 3), \text{ if } c_j^0 = 0; \tag{34}$$

$$\sqrt{n_j} T^{\frac{1+\gamma}{2}} \left(\check{\rho}_j - \rho_j^0 - \frac{1}{T^\gamma} \frac{-2c_j^0}{\left(\omega_j^0 \right)^2} \left(\lambda_j^0 + \frac{c_j^0}{T^\gamma} \overline{m}_{j,T} \right) \right) \Rightarrow \mathcal{N} \left(0, -2c_j^0 \right), \text{ if } c_j^0 < 0, \tag{35}$$

where

$$\overline{m}_{j,T} = \frac{1}{n_j} \sum_{i \in \mathcal{G}^0(j)} m_{i,T} \text{ and } m_{i,T} = \sum_{h=1}^\infty \left(\rho_j^0 \right)^{h-1} \mathbb{E} \left(\widetilde{\epsilon}_{it} u_{i,t-h} \right). \tag{36}$$

Remark 4.1 *Theorem 4.2 shows that $\check{\rho}_j$ can consistently estimate the true slope parameters in all three types of nonstationary roots. The asymptotic distributions of the post-clustering estimators are always Gaussian, regardless of the value of c_j^0 and show distinctively different behaviors from the time series case, as shown in Phillips and Magdalinos (2007b), in which the limiting distributions are Cauchy, Dickey-Fuller, and Gaussian when $c_j^0 > 0$, $= 0$, < 0, respectively. The above difference suggests that it is easier to test a hypothesis about the autoregressive coefficient in the panel context than in a time series context, as only the pivotal critical values are needed in practice.*

Remark 4.2 *The convergence rates of $\check{\rho}_j$ are $\sqrt{n_j} \left(\rho_j^0 \right)^T T^\gamma$, $\sqrt{n_j} T$ and $\sqrt{n_j} T^{\frac{1+\gamma}{2}}$ for the respective cases $c_j^0 > 0$, $c_j = 0$, and $c_j < 0$. These rates are $\sqrt{n_j}$ times the usual convergence rates for time series (Phillips and Magdalinos, 2007b). These enhanced rates in the panel model exploit the effects of cross section averaging and confirm that panel tests, which combine cross-section and time series information, are expected to have improved statistical power over tests that rely only on time series data.*

4.2 Testing for explosive roots

Theorem 4.3 *Suppose Assumptions 1 and 2 hold and $(n, T) \to \infty$. Under the null hypothesis $\mathcal{H}_0^{(j)} : c_j^0 = 0$, we have $\widetilde{t}_j \Rightarrow \mathcal{N}(0, 1)$, and $\widetilde{J}_j \Rightarrow \mathcal{N}(0, 1)$. Under the alternative hypothesis $\mathcal{H}_1^{(j)} : c_j^0 > 0$, we have $\widetilde{t}_j = O_p \left(\left(\rho_j^0 \right)^T \sqrt{n} \right)$, and $\widetilde{J}_j = O_p \left(\sqrt{n} T^{1-\gamma} \right)$.*

Remark 4.3 *In the explosive root groups, where $c_j^0 > 0$, the sample moment*

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

19

ensures asymptotically negligible bias terms, which diminish at a faster rate than the Gaussian distribution. Therefore,

$$\sqrt{n_j}\left(\rho_j^0\right)^T T^\gamma \left(\check{\rho}_j - \rho_j^0 + \frac{n_j T \left(\sigma_j^0\right)^2}{2 \sum_{i \in \mathcal{G}^0(j)} \sum_{t=1}^T \widetilde{y}_{i,t-1}^2}\right) \Rightarrow \mathcal{N}\left(0, 4\left(c_j^0\right)^2\right), \ \text{if } c_j^0 > 0; \qquad (37)$$

$$\sqrt{n_j} T \left(\check{\rho}_j - \rho_j^0 + \frac{n_j T \left(\sigma_j^0\right)^2}{2 \sum_{i \in \mathcal{G}^0(j)} \sum_{t=1}^T \widetilde{y}_{i,t-1}^2}\right) \Rightarrow \mathcal{N}\left(0, 3\right), \ \text{if } c_j^0 = 0. \qquad (38)$$

By Theorem 4.2 and Lemma B.1, when $(n, T) \to \infty$, we can simply replace the sample moment that relies on the true membership by the corresponding sample moment based on the estimated group structures. It follows that

$$\sqrt{n_j}\left(\rho_j^0\right)^T T^\gamma \left(\check{\rho}_j - \rho_j^0 + \frac{\check{n}_j T \check{\sigma}_j^2}{2 \sum_{i \in \hat{\mathcal{G}}(j)} \sum_{t=1}^T \widetilde{y}_{i,t-1}^2}\right) \Rightarrow \mathcal{N}\left(0, 4\left(c_j^0\right)^2\right), \ \text{if } c_j^0 > 0, \qquad (39)$$

$$\sqrt{n_j} T \left(\check{\rho}_j - \rho_j^0 + \frac{\check{n}_j T \check{\sigma}_j^2}{2 \sum_{i \in \hat{\mathcal{G}}(j)} \sum_{t=1}^T \widetilde{y}_{i,t-1}^2}\right) \Rightarrow \mathcal{N}\left(0, 3\right), \ \text{if } c_j^0 = 0. \qquad (40)$$

The consistency of the short-run variance estimate ensures standard normality of the post-clustering statistics $\widetilde{t}_j$ and $\widetilde{J}_j$ under $\mathcal{H}_0^{(j)} : c_j^0 = 0$.

Remark 4.4 *In comparison with statistics based on time series data, the t-statistic under the alternative hypothesis of an explosive root diverges at the rate $O\left(\left(\rho_j^0\right)^T\right)$, slower than the $O\left(\left(\rho_j^0\right)^T \sqrt{n}\right)$ rate of Theorem 4.3. The power deficiency of pure time series t-tests arises from this lower convergence rate of time series estimates under the alternative. Importantly, under the alternative $\mathcal{H}_1^{(j)} : c_j^0 > 0$ the t-statistic $\widetilde{t}_j = O_p\left(\left(\rho_j^0\right)^T \sqrt{n}\right)$ has a divergence rate that is faster by an exponential factor than the coefficient based test for which $\widetilde{J}_j = O_p\left(\sqrt{n} T^{1-\gamma}\right)$. This difference, which does not occur in stationary alternatives (for either time series or panel data tests), arises because $\widetilde{t}_j$ is constructed with a standard error in the denominator that shrinks at an exponential rate corresponding to the signal strength of the regressor in the mildly explosive case.*

4.3 Estimating the number of groups

The following Theorem shows that the IC estimator $\widetilde{G}$ is inconsistent when there are mildly stationary groups in the data.

Theorem 4.4 *Suppose Assumptions 1 and 2 hold and $(n, T) \to \infty$. When either (i) $\gamma \in (0, 1)$ and $\bar{c}_i^0 \geq 0$ or (ii) $\gamma = 0$, we have $\widetilde{G} \to_p G^0$. When $\gamma \in (0, 1)$, we have $\widetilde{G} \leq G^0$, with probability approaching 1.*

The first part of Theorem [4.4] indicates that when there is no mildly stationary group, $\widetilde{G}$ consistently estimates G^0. The second part of the theorem shows that $\widetilde{G}$ may underestimate G^0 with positive probability asymptotically when mildly stationary processes are present in the panel. The next result shows that the combined IC and Hausman test estimator $\widehat{G}$ delivers a consistent estimate of G^0.

Theorem 4.5 *Let Assumptions [1] and [2] hold and $\gamma \in (0,1)$. Then $\widehat{G} \to_p G^0$, as $(n,T) \to \infty$.*

Remark 4.5 *The idea of using a Hausman-type statistic in this context follows Phillips and Sul (2003). The procedure consistently tests for slope heterogeneity and possible mis-clustering of individuals in the panel, thereby providing a useful complement to IC group number selection, especially in cases like the present where IC is not consistent for all possible classifications. Pesaran et al. (1996) gave another Hausman-type statistic to test for a difference between panel within estimation $\widehat{\rho}^{FE}$ and mean group estimation $\widehat{\rho}^{MG}$, where $\widehat{\rho}^{MG} = \frac{1}{n} \sum_{i=1}^{n} \widehat{\widetilde{\rho}}_i^{TS}$ and $\widehat{\widetilde{\rho}}_i^{TS}$ is the time series estimate for individual i in the panel, as defined in equation [3]. But this procedure is not easily used in the present context since $\widehat{\rho}^{FE}$ and $\widehat{\rho}^{MG}$ are both asymptotically efficient estimates with $\text{Avar}\left(\widehat{\rho}^{FE}\right) = \text{Avar}\left(\widehat{\rho}^{MG}\right)$, so the Rao-Blackwell theorem is not applicable (Pesaran and Yamagata, 2008).*

Remark 4.6 *In the asymptotic analysis earlier in this section, the limit theory was obtained under the assumption that the true number of groups, G^0, was known. Theorem [4.5] shows that the estimator, $\widehat{G}$ consistently estimates G^0. Following common practice in the panel clustering literature (Bonhomme and Manresa, 2015; Su et al., 2016; Huang et al., 2021), we treat the consistent estimate $\widehat{G}$ as the true value G^0 in the practical work of implementation[6].*

5   Simulation Studies

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

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) iid errors ($\theta = 0$, $\epsilon_{it} \overset{i.i.d.}{\sim} \mathcal{N}(0, 0.01)$);[7] 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.,

---

[6]It is worth noting that a slow rate of convergence in determining the correct model specification can lead to invalid inference (Leeb and Pötscher, 2005, 2008) due to post-selection inference difficulties particularly in small samples. However, existing results in the literature (Liu et al., 2020; Dzemski and Okui, 2021) show that overidentification of the group number and resulting minor classification errors do not damage the asymptotic properties of the post-classification estimators. Extensions of such findings to nonstationary panel models like those of the present paper are a topic for further research.

[7]The *iid* error case is studied in the Online Supplement.

two groups) with $\pi_1 : \pi_2 = \frac{1}{2} : \frac{1}{2}$. The following parameter settings for c and γ were considered:

$$\left(c_1^0, c_2^0, c_3^0, \gamma\right) = \begin{cases} (-15, -8, -1, 0.6) & \text{for} \quad \text{DGP 0} \\ (1, \ 0, -6, 0.6) & \text{for} \quad \text{DGP 1} \\ (1, 0.2, -6, 0.6) & \text{for} \quad \text{DGP 2} \end{cases} \tag{41}$$

and

$$\left(c_1^0, c_2^0, \gamma\right) = (-1, 1, 0.6) \quad \text{for} \quad \text{DGP 3.} \tag{42}$$

These settings reflect those found in the empirical work in Section 6. The models considered allow for three groups (G^0=3) and two groups ($G^0 = 2$). With 3 groups and all $c_j^0 < 0$ DGP 0 helps to reveal the downward bias problem of IC when all groups are mildly stationary. DGPs 1–3 each has a mixed collection of groups and these collections are designed to show the accuracy of the hybrid model specification procedure, the consistency of the recursive k-means clustering algorithm, and the power improvements that result from cross section within group averaging in the panel inference procedures.

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

[Insert Figure 1 and Table 3 Here]

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

$$\text{PP } t\text{-test} = \frac{\left(\widehat{\overline{\rho}}_i^{TS} - 1 - \frac{T \cdot \widehat{\overline{\lambda}}_i^{TS}}{D_{i,T}}\right)\sqrt{D_{i,T}}}{\widehat{\overline{\omega}}_i^{TS}}, \quad \text{PP } J\text{-test} = T\left(\widehat{\overline{\rho}}_i^{TS} - 1 - \frac{T \cdot \widehat{\overline{\lambda}}_i^{TS}}{D_{i,T}}\right), \tag{43}$$

in which $\widehat{\overline{\rho}}_i^{TS}$ is the time series estimate of $\overline{\rho}_i^0$ defined in (3), 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 $\mathcal{H}_0 : \bar{c}_i^0 = 0$, it follows from standard theory that

$$\text{PP } t\text{-test} \Rightarrow \frac{\int_0^1 \underline{W}_i(r) dW_i(r)}{\left[\int_0^1 \underline{W}_i(r)^2 dr\right]^{\frac{1}{2}}}, \quad \text{PP } J\text{-test} \Rightarrow \frac{\int_0^1 \underline{W}_i(r) dW_i(r)}{\int_0^1 \underline{W}_i(r)^2 dr}, \tag{44}$$

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

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 time series 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)). The bandwidth for the long run variance estimates in (14) and (15) was set at $L = \lfloor T^{0.3} \rfloor$, the bandwidth for the variance estimate in (16) was set to $L = \lfloor T^{0.1} \rfloor$, and for the time series statistics in (43) the bandwidth for the long run variance and covariance components was set to $\lfloor T^{0.3} \rfloor$.[8] These bandwidth choices are consistent with the rate restrictions in the theory development and they are used in the empirical analysis. In all cases in the numerical simulations the number of replications was $1,000$.

The performance of the group number estimate $\widehat{G}$ is considered first. The penalty κ_{nT} of IC is $(nT)^{-0.35}$ and the upper bound $G_{\max}$ is 5. The critical value of the Hausman test is set as $cv_{nT} = (1 + 5\log(nT)) \chi^2(\overline{G})$ and $\overline{G} = (G_{\max} - G + 1)$. Tables 4–7 report the empirical frequency of $\widehat{G}$ in (26). 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 consistency in group number estimation. By comparison the downward bias of IC is evident in nearly every case, corroborating the asymptotic theory.

[Insert Tables 4–7 Here]

Next, we checked the finite sample performance of the recursive k-means clustering algorithm and post-clustering estimation, assuming the true group number G^0 is known. Tables 8–10 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

---

[8]Bandwidths are selected based on the simulation findings in the mixed-root panel model. When the bandwidth of (14) and (15) is smaller than $\lfloor T^{0.3} \rfloor$, the panel t test statistic overrejects and leads to size distortion. When the bandwidths used in (16) exceed $\lfloor T^{0.1} \rfloor$, the panel t test statistic is too small and test power in rejecting the null of a group-specific unit root is reduced. Automated bandwidth choices could be obtained by cross-validation, as in Phillips et al. (2017), and this approach is left for future investigation.

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

[Insert Tables 8–10 Here]

According to Tables 8–10, the CE decreases to zero as T increases. The RMSE and bias of the oracle estimates are smaller than those of the post-clustering estimates. For the post-clustering estimates of the nonstationary groups, the magnitude of the RMSE and bias also generally decreases when $T \to \infty$. For all DGPs, the difference between the oracle and the post-clustering estimates is negligible when $T \geq 150$. The diminishing differences suggest asymptotic equivalence between these two sets of estimates. This property is due to the uniform consistency of the recursive k-means clustering algorithm, as shown in the theory development.

Based on the estimated membership $\widehat{\delta}$, the performance of the post-clustering panel t and J tests for detecting explosive roots is analyzed and compared with the time series counterparts. The nominal levels are all set at 5%, accompanied by the right-tail 95% critical values of the standard normal distribution and standard unit root limit distributions. We obtain the empirical rejection rates of the PP t and J tests when $n = 1$ and the empirical rejection rates of the post-clustering panel t and J tests when $n > 1$, which are presented in Table 11. If the distancing parameter c_j^0 is zero (as in the null hypothesis) the empirical rejection rate gives test size; and when c_j^0 is nonzero, the empirical rejection rate gives test power.

[Insert Table 11 Here]

Evidently the size distortion of both panel tests is small when $n \geq 60$ and $T \geq 150$, although size distortion of the panel tests is slightly larger than that of the time series counterparts. This is unsurprising as the asymptotics require the use of cross section central limit theory, which inevitably introduces approximation errors in finite samples, particularly small samples that arise in group subsamples. This loss is counterbalanced by a substantial improvement in the power of the panel tests over the time series tests. For instance, when $c_j^0 = 0.2$ (the corresponding ρ_j^0 is 1.0126, 1.0099 and 1.0083 when $T = 100, 150, 200$, which are empirically plausible values based on our empirical work), the power performances of the post-clustering panel tests are much larger than those of the time series tests. If $T = 100$, the post-clustering panel t-test raises the power of the time series t-test from 0.175 to 0.599 when $n_j = 10$, to 0.807 when $n_j = 20$, and to 0.917 when $n_j = 30$. The post-clustering panel t test with $T = 100$ has substantially greater power than the time series t test with $T = 200$ (0.917 versus 0.382). Moreover, it is interesting to note that the panel t test has greater power than the panel J-test that is based on the estimated membership, corroborating the different divergence rates in asymptotic theory of Theorem 4.3 under the mildly explosive alternative.

24

6 Empirical Applications

6.1 Housing prices in China

It is well known that housing prices in China have experienced unprecedented growth over the last 20 years. Using data on thirty five major Chinese cities, Chen and Wen (2017) found that housing prices have substantially outgrown income in these cities, leading them to interpret China's housing boom as a rational bubble. Such an interpretation is important to subject to a formal assessment of the empirical evidence using rigorous methods to detect potential explosive behavior. Our first empirical study applies the methods of this paper to a panel of monthly housing indices from 107 ($n = 107$) cities in China obtained from Fang et al. (2016). The sample period is from January 2003 to December 2012 and contains 120 monthly observations ($T = 120$). Ideally housing rental prices in these cities would be useful to measure fundamental values in these real estate markets. But it is difficult to find reliable rental indices at the city level in China. Instead, as a proxy, we use the monthly national-level Consumer Price Index (CPI) for rentals to approximate fundamental values.[9] For this application we let $\{y_{it}\}$ be the ratio between the nominal housing price index for city i and the CPI for rentals in month t.

The model in (4) and (5) was fitted to $\{y_{it}\}$ using the proposed methods. Cross section heterogeneity exists because different cities have different characteristics and may, for example, experience different levels of urbanization. Nonetheless, a group structure in the evolution of house prices may exist because of similarities in the driving mechanism underlying the house price dynamics in some cities and commonalities that exist in supply and demand factors, leading to the co-existence of possible groupings of cities into mildly explosive groups, a unit root efficient market group, and mildly stationary groups.

[Insert Table 12 Here]

With tuning parameter[10] $\kappa_{nT} = (nT)^{-0.7}$, Table 12 reports the values of the computed ICs for $G = 1,...,5$. According to the IC selection $\widetilde{G} = 3$. The Hausman-type test algorithm is applied using the critical value $cv_{nT} = (1 + 5\log(nT))\chi^2\left(\overline{G}\right)$ and this procedure leads to the same estimate $\widehat{G} = 3$. The recursive k-means clustering algorithm is then implemented based on (8), giving the post-clustering estimate (12). This two stage procedure provides the clustered group structure[11,12]. There are 27 cities in Group 1 including one

---

[9]The CPI for rentals is available on the official website of National Bureau of Statistics, China, http://www.stats.gov.cn/

[10]Here and in the following empirical applications the IC penalty tuning parameter κ_{nT} was chosen to satisify the rate condition (19). The results reported were found to be robust to tuning parameter choices in the range $\kappa_{nT} \in [(nT)^{-0.7}, (nT)^{-0.6}]$. In particular, for $\kappa_{nT} = (nT)^{-0.62}$, membership estimation, post-clustering parameter estimation and inferences concerning explosive behavior are unchanged in all the applications of the paper.

[11]The initial values used to start the k-means algorithm were chosen as the 30%, 70%, and 90% quantiles of the individual time series estimates. The rate parameter $\gamma = 0.6$ was used. The estimated groups were found to be robust to various choices of the initial values and the rate parameter.

[12]The names of the cities in each estimated group are reported in the Online Supplement.

tier-1 city, Beijing. Another 42 cities are included in Group 2 including two tier-1 cities, Guangzhou and Shenzhen. Comparatively, there are 38 cities in Group 3 including the other tier-1 city, Shanghai, whose time series estimate of the slope coefficient is 0.9959. Figure 5 gives the time series plots for all three groups.

[Insert Table 13 Here]

For each identified group, we report the panel-within estimate $\check{\rho}_j$, the number of cities in each estimated group, the post-clustering t- and J-statistics for the null hypothesis of the group-specific unit root, as in Table 13. For Group 1, the estimated $\check{\rho}_1$ is 1.011. Both the panel t-test and the panel J-test which are based on the estimated groups suggest that c_1^0 is significantly larger than zero at the 1% significance level. For Group 2, the estimated $\check{\rho}_2$ is 1.003, which is very close to unity. Nonetheless, the post-clustering t-test suggests that the true scale coefficient is bigger than zero at the 1% level while the panel J-test rejects the null hypothesis of the group-specific unit root at the 5% significance level. For Group 3, the post-clustering estimate $\check{\rho}_3$ is just below unity and the unit root null hypothesis is not rejected in this group. The post-clustering t and J tests therefore indicate explosive behavior in Groups 1 and 2 but unit root behavior in Group 3.

[Insert Figure 5 Here]

6.2 Housing prices in the US

In recent years, strong surges in house prices have occurred in many US cities. Possible reasons for these surges include near-zero interest rates and rising inflation expectations. To examine whether rising fundamental values justify these developments the two-stage algorithm of the present paper was applied to a panel of monthly housing indices for 11 ($n = 11$) US cities obtained from the official website of the Federal Reserve Bank of St. Louis[13]. Monthly observations of $T = 105$ time series for each series were used, covering the period from January 2013 to September 2021. To measure fundamental values monthly city-specific Consumer Price Index (CPI) data for rentals was employed.[14] In the application, $\{y_{it}\}$ was set as the ratio of the nominal housing price index to the CPI for rentals for city i in month t.

[Insert Table 14 Here]

Using the penalty parameter setting $\kappa_{nT} = (nT)^{-0.65}$, Table 14 reports calculated values of the ICs for $G = 1, ..., 5$, leading to the choice $\widetilde{G} = 2$. The combined IC-Hausman test procedure with critical value $cv_{nT} = (1 + 5\log(nT))\chi^2(\overline{G})$ gave the same estimated value $\widehat{G} = 2$. Using this estimated group number the clustering algorithm based on (8) was implemented, giving post-clustering estimates from (12) and the corresponding clustered

---

[13]https://fred.stlouisfed.org/
[14]For cities whose city-specific CPI for rentals data were unavailable, fundamental values were approximated using the national CPI for rentals.

26

group structure.[15,16] The two stage procedure produced 7 cities in Group 1 and and 4 cities in Group 2. Time series plots of these two groups are provided in Figure 6.

[Insert Figure 6 Here]

For each identified group of cities in the US housing market, Table 15 reports the panel within-group estimates $\breve{\rho}_j$, the number of cities in each estimated group, and the post-clustering t- and J-statistics for the null hypothesis of the group-specific unit root. According to both the post-clustering t and J statistics, the explosive root $\breve{\rho}_1 = 1.0432$ of Group 1 is statistically significant at the 1% level, indicating the presence of a housing bubble in this group. For Group 2, the post-clustering estimate $\breve{\rho}_2 = 1.0113$ also exceeds unity and the unit root group null is rejected at the 5% level. These post-clustering panel tests suggest that explosive price bubbles are present in the data for both Groups 1 and 2, although Group 2 has 4 cities with a considerably weaker common explosive root in housing prices. In consequence, there are clear differences in detection between the panel clustered series and the individual time series. Figure 3 shows the time series of US house prices[17] for the cities where explosive behavior was detected by the panel t-test but not by individual time series t-tests. In fact, none of the 11 cities were found to have explosive behavior in house prices in the individual tests at the 5% level.

[Insert Table 15 Here]

6.3 Equity prices in the US

In a further application the methodology is applied to a panel of equity prices in the US stock market. Whereas analysis of a general stock price index may indicate the presence of an explosive root as in the historical study of Phillips et al. (2015a), such a finding does not mean that all of the component stocks manifest explosive features. Narayan et al. (2013) found evidence of group-specific heterogeneity in 589 stocks from nine different sectors, so it is natural to incorporate group-specific heterogeneity in the analysis of stock market bubbles. Our application employs the proposed two stage approach in which the panel variables $\{y_{it}\}$ are set as the difference in levels between the monthly price and monthly dividends of stock i in period t. The presence of a significant explosive common root in any group is then indicative of a stock price bubble in that group.

[Insert Table 16 Here]

Monthly data for the S&P 500 component stocks were sourced from the Wharton Research Data Service (WRDS), covering around 500 stocks in different sampling periods. For this study we selected a panel of 146 stocks giving 98 monthly observations

---

[15]The initial values were obtained from the 30% and 80% quantiles of the individual time series estimates and the localizing rate parameter was set to $\gamma = 0.6$. The empirical results were found to be robust to various initial values and localizing rate parameters.

[16]The names of the US cities in each estimated group are reported in the Online Supplement.

[17]Specifically, the plotted data in Figure 3 is the ratio of the city price index to the CPI of the city rentals

$(n = 146, T = 98)$ taken over the common period between January 2010 and February 2018.

Using the penalty parameter $\kappa_{nT} = (nT)^{-0.6}$ Table 16 reports IC values for $G = 1, ..., 5$, leading to the estimate $\widetilde{G} = 2$. The combined IC-Hausman test procedure with critical value $cv_{nT} = (1 + 5\log(nT))\chi^2(\overline{G})$ produced the same estimate $\widehat{G} = 2$. The clustering algorithm was implemented using (8) and post-clustering estimates from (12). The two stage procedure provided the group structure[18,19]. The results gave 40 stocks in Group 1 and 106 stocks in Group 2, in which high-tech stocks such as IT and biotech stocks and energy stocks usually manifest mildly explosive roots. Time series plots of these two groups are provided in Figure 7.

[Insert Table 17 Here]

For each identified group, Table 17 reports the panel within-group estimates $\check{\rho}_j$, the number of stocks in each estimated group, and the post-clustering t- and J-statistics for the null hypothesis of the group-specific unit root. According to both the post-clustering t and J statistics, the explosive root $\check{\rho}_1 = 1.012$ of Group 1 is statistically significant at the 5% level, indicating the presence of a price bubble in this group. For Group 2, the post-clustering estimate $\check{\rho}_2 = 0.967$ is smaller than unity and we cannot reject the null hypothesis of a group-specific unit root. These post-clustering panel tests therefore suggest that an explosive price bubble is manifest in Group 1 with 40 stocks, whereas Group 2 has 106 stocks with near unit root behavior indicative of an efficient market. The time series in the two groups are displayed in Figure 7. In individual time series tests at the 5% level, only 29 of the stocks in Group 1 were found to be explosive and no explosive behavior was supported in any of the 109 stocks in Group 2. Clustering the time series therefore assists in the detecting 11 further stocks manifesting explosive behavior.

[Insert Figure 7 Here]

7   Conclusions

The existence of explosive phenomena is conveniently captured in time series autoregression by an autoregressive root that exceeds unity. Phillips and Magdalinos (2007a) introduced the concept of mildly explosive roots, which have proved particularly useful in empirical research because they are amenable to estimation and inference with pivotal asymptotic theory, confidence interval construction (Phillips, 2021) and recursive testing algorithms. This mechanism of detection has assisted in determining the presence of asset price bubbles in financial assets like stocks and real assets like housing. The present paper extends this mechanism to allow for latent group structures within a

---

[18]The initial values were obtained from the 30% and 80% quantiles of the individual time series estimates and the localizing rate parameter was set to $\gamma = 0.6$. The empirical results were found to be robust to various initial values and localizing rate parameters.

[19]The individual stock symbols in each estimated group are reported in the Online Supplement.

dynamic panel model so that the individual time series may have mixed roots that fall into three general categories, some that are mildly explosive, some that are mildly stationary, and some with a unit root. This extension is appealing in wide panels where behavior may vary within each of these general classifications. The framework then allows for subgroups with different autoregressive coefficients within a particular class such as those with mildly explosive roots, which assists in modeling several forms of explosive behavior. The paper develops a clustering algorithm that accommodates this framework and enables detection of the clusters and estimation of their respective coefficients, taking advantage of cross section averaging within each cluster. In particular, a two stage approach is proposed to detect explosive behavior, incorporating a recursive k-means clustering algorithm in the first stage and the panel approach to bubble analysis in the second stage. Both asymptotic theory and numerical simulations show that the post-clustering testing procedures attain better power performance in bubble detection than a time series approach; and the clustering algorithm is uniformly consistent in recovering the latent group membership.

Several extensions of the present research are possible. First, the framework of this paper only accommodates time-invariant parameters and does not allow for structural breaks. Hence, the origination and termination of an explosive episode in data are not included within the present framework. However, the model and methods can be extended to a wider panel setup that includes the real-time bubble dating strategy developed in Phillips et al. (2011, 2015a) and more recent research on dating methods. Second, cross section independence was imposed in the present framework to facilitate the development of the asymptotic theory of clustering. A natural extension of this framework is to employ models with panel interactive fixed effects (Bai, 2009) or panel group fixed effects (Bonhomme and Manresa, 2015; Bonhomme et al., 2017) to accommodate some of the features that are often present in panel data, particularly those where common factors play a role in determining episodes of exuberance in the data (c.f., Chen et al. (2022)). Such extensions of the present model involve considerable complexities, especially when different groups involve different break dates and real-time analysis is need for practical implementation. Some of these complications are currently under investigation and will be reported in future work.

Appendix: Proofs

Throughout the following proofs we use the same notation as in the paper. The technical lemmas listed in the next section are proved in the Online Supplement and play central roles in the proofs of the main theorems in the paper.

A   Proofs for Stage 1: Recursive k-means Clustering

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 for any individual $i \in \mathcal{I}_n$. 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 demonstrate uniform consistency of the recursive k-means clustering algorithm, we first establish consistency of the parameter estimate $\widehat{c}^*$ in terms of the Hausdorff distance that measures how far two compact subsets in a metric space are separated from each other. This distance is defined as

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

in which $a := (a_1, a_2, ..., a_{G^0})$ and $b := (b_1, b_2, ..., b_{G^0})$. The proof of uniform consistency makes use of the following lemmas which are stated first.

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

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

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

with $\Upsilon_{iT} = \sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2$ and $\overline{\rho}_i$ and $\overline{\rho}_i^0$ defined in (6).

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

$$d_H(c^0,\widehat{c}^*) = o_p\left(T^{-2\gamma}(\log T)^{-8}\right). \tag{45}$$

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

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

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

$$\left\|\widehat{c}^* - c^0\right\| = o_p\left(T^{-\gamma}(\log T)^{-4}\right). \tag{46}$$

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

$$\mathcal{N}_{\eta} := \left\{c\in\mathcal{C}_{G^0}: \left|c_j^0 - c_j\right| < \eta, \ \forall j = 1,2,...,G^0\right\}, \tag{47}$$

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

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

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

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

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

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

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

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

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

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

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

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

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

Lemma A.5 *Suppose Assumptions 1 and 2 hold, then:*

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

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

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

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

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

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

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 (47).

For any $j \in \mathcal{G}^0$ and $i \in \mathcal{I}_n$, let

$$\widehat{E}_{j,i} := \left\{\widehat{g}_i \neq j | g_i^0 = j\right\} \text{ and } \widehat{F}_{j,i} := \left\{g_i^0 \neq j | \widehat{g}_i = j\right\}. \tag{48}$$

Moreover, let $\widehat{E}_{j,nT} := \bigcup_{i \in \mathcal{G}^0(j)} \widehat{E}_{j,i}$ and $\widehat{F}_{j,nT} := \bigcup_{i \in \widehat{\mathcal{G}}(j)} \widehat{F}_{j,i}$. To show uniform consistency of the recursive k-means clustering algorithm, we use the following lemma.

Lemma A.7 *(Uniform Consistency of Clustering) Suppose Assumptions 1 and 2 hold. Then, 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 Theorem 4.1: We use Lemmas A.7(i) and A.7(ii). To establish uniform consistency of the recursive k-means clustering algorithm, we first bound the clustering error by

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

Then, it follows that

$$\begin{aligned}
\sum_{j=1}^{G^0} \sum_{i \in \mathcal{G}^0(j)} \Pr\left(\widehat{E}_{j,i}\right) &\leq n \max_{i \in \mathcal{I}_n} \mathbb{E}\mathbf{1}\{\widehat{g}_i(\widehat{c}^*) \neq g_i^0\} \leq n \max_{i \in \mathcal{I}_n} \Pr\{\left|\widehat{g}_i(\widehat{c}^*) - g_i^0\right| > 0\} \\
&\leq n \max_{i \in \mathcal{I}_n} \sup_{c \in \mathcal{N}_{\eta}} \Pr\{\left|\widehat{g}_i(c) - g_i^0\right| > 0\} + n \max_{1 \leq j \leq G^0} \Pr\{\left|\widehat{c}_j^* - c_j^0\right| > \eta\} \\
&= o(1) + n \max_{1 \leq j \leq G^0} \Pr\{\left|\widehat{c}_j^* - c_j^0\right| > \eta\} \\
&= o(1), \tag{49}
\end{aligned}$$

where the last step is due to the Markov inequality, equation (46) in Lemma A.1 and the rate restriction in Assumption 2. This proves Lemma A.7(i). For Lemma A.7(ii), we can follow the proof of Theorem 2.2(ii) in Su et al. (2016). The results of Theorem 4.1 are then extensions of Lemma A.7 and the proof is complete. ∎

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

We need a lemma that establishes consistency of the variance estimates $\breve{\omega}_j^2$, $\breve{\sigma}_j^2$ and $\breve{\lambda}_j$ and lemmas that show the limits of various sample moments. These lemmas are stated first.

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

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

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

$$
\frac{1}{n_j T^2} \sum_{i \in \mathcal{G}^0(j)} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}^2 \to_p \frac{\left(\omega_j^0\right)^2}{6}, \quad \text{if } c_j^0 = 0; \tag{50}
$$

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

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

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

$$
\frac{1}{\sqrt{n_j} T} \sum_{i \in \mathcal{G}^0(j)} \sum_{t=1}^{T} \left(\widetilde{y}_{i,t-1} \widetilde{u}_{i,t} - \widehat{\lambda}_j + \frac{\widehat{\omega}_j^2}{2} \right) \Rightarrow \mathcal{N}\left(0, \frac{\left(\omega_j^0\right)^4}{12} \right), \quad \text{if } c_j^0 = 0; \tag{51}
$$

$$
\frac{1}{\sqrt{n_j} T^{\gamma} (\rho_j^0)^{T}} \sum_{i \in \mathcal{G}^0(j)} \sum_{t=1}^{T} \widetilde{y}_{i,t-1} \widetilde{u}_{it} \Rightarrow \mathcal{N}\left(0, \frac{\left(\omega_j^0\right)^4}{4\left(c_j^0\right)^2} \right), \quad \text{if } c_j^0 > 0; \tag{52}
$$

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

where

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

Proof of Theorem 4.2: When $c_j^0 > 0$, the following decompositions apply to the numerator and denominator of the post-clustering estimate, $\breve{\rho}_j$:

$$\frac{1}{\sqrt{n_j}T^{\gamma}(\rho_j^0)^T} \sum_{i\in\hat{\mathcal{G}}(j)} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}\widetilde{u}_{it}$$

$$= \frac{1}{\sqrt{n_j}T^{\gamma}(\rho_j^0)^T} \sum_{i\in\mathcal{G}^0(j)} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}\widetilde{u}_{it} + \frac{1}{\sqrt{n_j}T^{\gamma}(\rho_j^0)^T} \sum_{\widetilde{j}\neq j} \sum_{i\in\hat{\mathcal{G}}(j)\cap\mathcal{G}^0(\widetilde{j})} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}\widetilde{u}_{it}$$

$$- \frac{1}{\sqrt{n_j}T^{\gamma}(\rho_j^0)^T} \sum_{\widetilde{j}\neq j} \sum_{i\in\hat{\mathcal{G}}(\widetilde{j})\cap\mathcal{G}^0(j)} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}\widetilde{u}_{it},$$

and

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

$$= \frac{1}{n_j T^{2\gamma}(\rho_j^0)^{2T}} \sum_{i\in\mathcal{G}^0(j)} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}^2 + \frac{1}{n_j T^{2\gamma}(\rho_j^0)^{2T}} \sum_{\widetilde{j}\neq j} \sum_{i\in\hat{\mathcal{G}}(j)\cap\mathcal{G}^0(\widetilde{j})} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}^2$$

$$- \frac{1}{n_j T^{2\gamma}(\rho_j^0)^{2T}} \sum_{\widetilde{j}\neq j} \sum_{i\in\hat{\mathcal{G}}(\widetilde{j})\cap\mathcal{G}^0(j)} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}^2.$$

To demonstrate the asymptotic equivalence between the post-clustering estimates and the oracle estimates, we need to show the following:

(i) for any $j = 1, 2, ..., G^0$,

$$\frac{1}{\sqrt{n_j}T^{\gamma}(\rho_j^0)^T} \sum_{\widetilde{j}\neq j} \sum_{i\in\hat{\mathcal{G}}(j)\cap\mathcal{G}^0(\widetilde{j})} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}\widetilde{u}_{it} = o_p(1),$$

$$\frac{1}{n_j T^{2\gamma}(\rho_j^0)^{2T}} \sum_{\widetilde{j}\neq j} \sum_{i\in\hat{\mathcal{G}}(j)\cap\mathcal{G}^0(\widetilde{j})} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}^2 = o_p(1);$$

(ii) for any $j = 1, 2, ..., G^0$,

$$\frac{1}{\sqrt{n_j}T^{\gamma}(\rho_j^0)^T} \sum_{\widetilde{j}\neq j} \sum_{i\in\hat{\mathcal{G}}(\widetilde{j})\cap\mathcal{G}^0(j)} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}\widetilde{u}_{it} = o_p(1),$$

$$\frac{1}{n_j T^{2\gamma}(\rho_j^0)^{2T}} \sum_{\widetilde{j}\neq j} \sum_{i\in\hat{\mathcal{G}}(\widetilde{j})\cap\mathcal{G}^0(j)} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}^2 = o_p(1).$$

34

Since the treatment of the denominator is identical to that of the numerator, we need only focus on the numerator. In terms of (ii), for any $\widetilde{j} \neq j$ and any $\varepsilon > 0$,

$$\Pr\left(\left\|\frac{1}{\sqrt{n_j}T^\gamma(\rho_j^0)^T}\sum_{\widetilde{j}\neq j}\sum_{i\in\hat{\mathcal{G}}(\widetilde{j})\cap\mathcal{G}^0(j)}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right\| > \varepsilon\right) \leq \Pr\left(\sum_{j=1}^{G^0}\widehat{E}_{j,nT}\right) \to 0, \tag{53}$$

when $(n,T) \to \infty$. Similarly, in terms of (i), for any $\widetilde{j} \neq j$ and any $\varepsilon > 0$,

$$\Pr\left(\left\|\frac{1}{\sqrt{n_j}T^\gamma(\rho_j^0)^T}\sum_{\widetilde{j}\neq j}\sum_{i\in\hat{\mathcal{G}}(j)\cap\mathcal{G}^0(\widetilde{j})}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right\| > \varepsilon\right) \leq \Pr\left(\sum_{j=1}^{G^0}\widehat{F}_{j,nT}\right) \to 0, \tag{54}$$

when $(n,T) \to \infty$. Combining (53) and (54), we have

$$\frac{1}{\sqrt{n_j}T^\gamma(\rho_j^0)^T}\sum_{i\in\hat{\mathcal{G}}(j)}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it} = \frac{1}{\sqrt{n_j}T^\gamma(\rho_j^0)^T}\sum_{i\in\mathcal{G}^0(j)}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it} + o_p(1), \tag{55}$$

and

$$\frac{1}{n_j T^{2\gamma}(\rho_j^0)^{2T}}\sum_{i\in\hat{\mathcal{G}}(j)}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2 = \frac{1}{n_j T^{2\gamma}(\rho_j^0)^{2T}}\sum_{i\in\mathcal{G}^0(j)}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2 + o_p(1). \tag{56}$$

Based on (55) and (56), the asymptotic theory for the post-clustering estimator is equivalent to that of the infeasible estimator, which is

$$\sqrt{n_j}T^\gamma(\rho_j^0)^T\left(\check{\rho}_j - \rho_j^0\right) = \sqrt{n_j}T^\gamma(\rho_j^0)^T\left(\widehat{\rho}_j - \rho_j^0\right) + o_p(1). \tag{57}$$

Similarly, asymptotic equivalence can be obtained in the other cases. For example, when $c_j^0 = 0$,

$$\sqrt{n_j}T\left(\check{\rho}_j - \rho_j^0 - \frac{1}{T}\frac{6\lambda_j^0}{\left(\omega_j^0\right)^2} + \frac{1}{T}\frac{3\left(\omega_j^0\right)^2}{\left(\omega_j^0\right)^2}\right) = \sqrt{n_j}T\left(\widehat{\rho}_j - \rho_j^0 - \frac{1}{T}\frac{6\lambda_j^0}{\left(\omega_j^0\right)^2} + \frac{1}{T}\frac{3\left(\omega_j^0\right)^2}{\left(\omega_j^0\right)^2}\right) + o_p(1); \tag{58}$$

when $c_j^0 < 0$,

$$\sqrt{n_j}T^{\frac{1+\gamma}{2}}\left(\check{\rho}_j - \rho_j^0 - \frac{1}{T^\gamma}\frac{-2c_j^0}{\left(\omega_j^0\right)^2}\left(\lambda_j^0 + \frac{c_j^0}{T^\gamma}\overline{m}_{j,T}\right)\right)$$

$$= \sqrt{n_j}T^{\frac{1+\gamma}{2}}\left(\widehat{\rho}_j - \rho_j^0 - \frac{1}{T^\gamma}\frac{-2c_j^0}{\left(\omega_j^0\right)^2}\left(\lambda_j^0 + \frac{c_j^0}{T^\gamma}\overline{m}_{j,T}\right)\right) + o_p(1), \tag{59}$$

where $\overline{m}_{j,T} = \frac{1}{n_j}\sum_{i\in\mathcal{G}^0(j)} m_{i,T}$ and $m_{i,T} = \sum_{h=1}^{\infty}\left(\rho_j^0\right)^{h-1}\mathbb{E}\left(\widetilde{\epsilon}_{it}u_{i,t-h}\right)$. Therefore, we only need to derive the limit distribution of the infeasible estimator in all cases. From Lemmas B.2 and B.3, the limit distributions of the oracle estimators that employ the true group identities are derived and this completes the proof. ∎

Proof of Theorem 4.3: By Lemma B.1 and Theorem 4.2, when $(n, T) \to \infty$,

$$\sqrt{n_j}\left(\rho_j^0\right)^T T^\gamma \left(\check{\rho}_j - \rho_j^0 + \frac{\check{n}_j T \check{\sigma}_j^2}{2\sum_{i\in\hat{\mathcal{G}}(j)}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2}\right) \Rightarrow \mathcal{N}\left(0, 4\left(c_j^0\right)^2\right), \text{ if } c_j^0 > 0;$$

$$\sqrt{n_j}T\left(\check{\rho}_j - \rho_j^0 + \frac{\check{n}_j T \check{\sigma}_j^2}{2\sum_{i\in\hat{\mathcal{G}}(j)}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2}\right) \Rightarrow \mathcal{N}\left(0, 3\right), \text{ if } c_j^0 = 0,$$

So both the post-clustering t- and J-statistics follow a pivotal distribution upon standardizations, under the null hypothesis of group-specific unit root behavior.

Similarly, under the alternative hypothesis of the group-specific explosive root, for the post-clustering t-test, we have

$$\frac{\left(\check{\rho}_j - 1 + \frac{\check{n}_j T \check{\sigma}_j^2}{2\sum_{i\in\hat{\mathcal{G}}(j)}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2}\right)\check{D}_{j,nT}}{\check{\omega}_j\sqrt{\check{E}_{j,nT}}}$$

$$= \frac{\left(\check{\rho}_j - \rho_j^0 + \frac{\check{n}_j T \check{\sigma}_j^2}{2\sum_{i\in\hat{\mathcal{G}}(j)}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2}\right) - \left(1 - \rho_j^0\right)}{\check{\omega}_j} \times \frac{\left(\check{D}_{j,nT}\right)}{\sqrt{\check{E}_{j,nT}}}$$

$$= \frac{\left(\check{\rho}_j - \rho_j^0 + \frac{\check{n}_j T \check{\sigma}_j^2}{2\sum_{i\in\hat{\mathcal{G}}(j)}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2}\right)\left(\check{D}_{j,nT}\right)}{\check{\omega}_j\sqrt{\check{E}_{j,nT}}} - \frac{\left(1 - \rho_j^0\right)\check{D}_{j,nT}}{\check{\omega}_j\sqrt{\check{E}_{j,nT}}}$$

$$= O_p\left(1\right) + O_p\left(\sqrt{n}\left(\rho_j^0\right)^T\right),$$

where the last equality is due to Theorem 4.2. For the post-clustering J-test, we have

$$\sqrt{\frac{\check{n}_j}{3}}T\left(\check{\rho}_j - 1 + \frac{\check{n}_j T \check{\sigma}_j^2}{2\sum_{i\in\hat{\mathcal{G}}(j)}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2}\right)$$

$$= \sqrt{\frac{\check{n}_j}{3}}T\left(\check{\rho}_j - \rho_j^0 + \frac{\check{n}_j T \check{\sigma}_j^2}{2\sum_{i\in\hat{\mathcal{G}}(j)}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2}\right) - \sqrt{\frac{\check{n}_j}{3}}T\left(1 - \rho_j^0\right)$$

$$= O_p\left(\left(\rho_j^0\right)^{-T} T^{1-\gamma}\right) + O_p\left(\sqrt{n}T^{1-\gamma}\right),$$

where the last equality is due to Theorem 4.2. This concludes the proof. ∎

C Proofs for the Estimation of Group Numbers

Proofs of Equations (22)–(24): We first discuss the results of equation (22). When $G = G^0$ and $c_j^0 < 0$, we have

$$\sqrt{\check{n}_j T^{1+\gamma}}\left(\check{\rho}_j(G) - \rho_j^0\right) \Rightarrow \mathcal{N}\left(0, -2c_j^0\right),$$

$$\sqrt{\check{n}_{j,h} T^{1+\gamma}}\left(\check{\rho}_{j,h}(G) - \rho_j^0\right) \Rightarrow \mathcal{N}\left(0, -2c_j^0\right), \text{ for any } h = 1, 2, ..., \overline{G},$$

whose derivations are similar to Theorem 4.2. Since $\check{n}_{j,h}/\check{n}_j \to \pi_{j,h}$, then we have

$$\sqrt{\check{n}_j T^{1+\gamma}}\left(\check{\rho}_j(G) - \rho_j^0\right) \Rightarrow \mathcal{N}\left(0, -2c_j^0\right),$$

$$\sqrt{\check{n}_j T^{1+\gamma}}\left(\check{\rho}_{j,h}(G) - \rho_j^0\right) \Rightarrow \mathcal{N}\left(0, -2c_j^0 \pi_{j,h}^{-1}\right), \text{ for any } h = 1, 2, ..., \overline{G}.$$

Therefore, it follows that

$$\sqrt{\check{n}_j T^{1+\gamma}}\left(\check{\rho}_j(G) - \check{\rho}_{j,h}(G)\right) \sim_a \mathcal{N}\left(0, \text{ Avar}\left(\sqrt{\check{n}_j T^{1+\gamma}}\left(\check{\rho}_j(G) - \check{\rho}_{j,h}(G)\right)\right)\right),$$

in which

$$\text{Avar}\left(\sqrt{\check{n}_j T^{1+\gamma}}\left(\check{\rho}_j(G) - \check{\rho}_{j,h}(G)\right)\right)$$

$$= \text{Avar}\left(\sqrt{\check{n}_j T^{1+\gamma}}\left(\check{\rho}_j(G) - \rho_j^0\right)\right) + \text{Avar}\left(\sqrt{\check{n}_j T^{1+\gamma}}\left(\check{\rho}_{j,h}(G) - \rho_j^0\right)\right)$$

$$- 2\,\text{Acov}\left(\sqrt{\check{n}_j T^{1+\gamma}}\left(\check{\rho}_{j,h}(G) - \rho_j^0\right), \sqrt{\check{n}_j T^{1+\gamma}}\left(\check{\rho}_j(G) - \rho_j^0\right)\right).$$

For any $h = 1, 2, ..., \overline{G}$, we consider the asymptotic covariance as

$$\text{Acov}\left(\frac{1}{\sqrt{\check{n}_j T^{1+\gamma}}} \sum_{i \in \hat{\mathcal{G}}(j,G)} \sum_{t=1}^{T} y_{i,t-1} u_{i,t}, \frac{1}{\sqrt{\check{n}_j T^{1+\gamma}}} \sum_{i \in \hat{\mathcal{G}}(j,h,G)} \sum_{t=1}^{T} y_{i,t-1} u_{i,t}\right)$$

$$= \text{Acov}\left(\frac{1}{\sqrt{\check{n}_j T^{1+\gamma}}} \sum_{i \in \hat{\mathcal{G}}(j,h,G)} \sum_{t=1}^{T} y_{i,t-1} u_{i,t}, \frac{1}{\sqrt{\check{n}_j T^{1+\gamma}}} \sum_{i \in \hat{\mathcal{G}}(j,h,G)} \sum_{t=1}^{T} y_{i,t-1} u_{i,t}\right)$$

$$\sim_a \text{Acov}\left(\frac{1}{\sqrt{\check{n}_j T^{1+\gamma}}} \sum_{i \in \hat{\mathcal{G}}(j,h,G)} \sum_{t=1}^{T} y_{i,t-1} \epsilon_{i,t}, \frac{1}{\sqrt{\check{n}_j T^{1+\gamma}}} \sum_{i \in \hat{\mathcal{G}}(j,h,G)} \sum_{t=1}^{T} y_{i,t-1} \epsilon_{i,t}\right)$$

$$= \text{Avar}\left(\frac{1}{\sqrt{\check{n}_j T^{1+\gamma}}} \sum_{i \in \hat{\mathcal{G}}(j,h,G)} \sum_{t=1}^{T} y_{i,t-1} \epsilon_{i,t}\right)$$

$$= \text{Avar}\left(\frac{1}{\sqrt{\check{n}_j T^{1+\gamma}}} \sum_{i \in \hat{\mathcal{G}}(j,h,G)} \sum_{t=1}^{T} y_{i,t-1} \epsilon_{i,t}\right)$$

$$= \pi_{j,h} \cdot \text{Avar} \left(\frac{1}{\sqrt{\check{n}_{j,h}T^{1+\gamma}}} \sum_{i \in \hat{\mathcal{G}}(j,h,G)} \sum_{t=1}^{T} y_{i,t-1}\epsilon_{i,t} \right)$$

as $(n,T) \to \infty$, from which we deduce that

$$\text{Avar}\left(\sqrt{\check{n}_j T^{1+\gamma}} \left(\check{\rho}_j(G) - \check{\rho}_{j,h}(G) \right) \right)$$

$$= -\text{Avar}\left(\sqrt{\check{n}_j T^{1+\gamma}} \left(\check{\rho}_j(G) - \rho_j^0 \right) \right) + \text{Avar}\left(\sqrt{\check{n}_j T^{1+\gamma}} \left(\check{\rho}_{j,h}(G) - \rho_j^0 \right) \right)$$

$$= -2c_j^0 \left(\pi_{j,h}^{-1} - 1 \right).$$

Further, based on the assumption of cross-sectional independence, we prove equation (22). The proof of equation (24) follows a procedure similar to equation (22). For equation (23), the impact of the asymptotic bias needs to be considered as

$$\sqrt{\check{n}_j T^2} \left(\check{\rho}_j(G) - \rho_j^0 - \frac{2\left(\lambda_j^0\right)}{\left(\omega_j^0\right)^2} \frac{1}{T} \right) \Rightarrow \mathcal{N}(0,2),$$

$$\sqrt{\check{n}_{j,h} T^2} \left(\check{\rho}_{j,h}(G) - \rho_j^0 - \frac{2\left(\lambda_j^0\right)}{\left(\omega_j^0\right)^2} \frac{1}{T} \right) \Rightarrow \mathcal{N}(0,2), \text{ for any } h = 1,2,...,\overline{G}.$$

Then the stochastic distance between $\check{\rho}_j(G)$ and $\check{\rho}_{j,h}(G)$ can be similarly formulated as

$$\sqrt{\check{n}_j T^2} \left(\check{\rho}_j(G) - \check{\rho}_{j,h}(G) \right) \Rightarrow \mathcal{N}\left(0, \; 2\left(\pi_{j,h}^{-1} - 1 \right) \right).$$

The proof is now complete. ∎

Proof of Theorem 4.5: Assume any G with $\widetilde{G} \leq G \leq G_{\max}$ where $\widetilde{G}$ estimated by IC. Also assume any $j = 1,2,...,G$. Under the null hypothesis of slope homogeneity in the jth group, the test statistic (26) converges to $\chi^2\left(\overline{G}\right)$ based on equations (22)–(24). Since the critical values $cv_{nT} := (1 + b\log(nT)) \cdot \chi^2_{0.95}\left(\overline{G}\right) \to \infty$, the probability of the Type I error shrinks to zero asymptotically. Under the alternative hypothesis of a nonzero fraction of slope heterogeneity in the jth group, the test statistic (26) asymptotically diverges to infinity since

$$\min\left\{ \sqrt{\check{n}_j T^{1+\gamma}}, \sqrt{\check{n}_j T^2}, \sqrt{\check{n}_j \left(\rho_j^0\right)^{2T} T^{2\gamma}} \right\} > T^{\gamma}.$$

Aso since

$$\min\left\{ \check{n}_j T^{1-\gamma}, \check{n}_j T^{2-2\gamma}, \check{n}_j \left(\rho_j^0\right)^{2T} \right\} > cv_{nT},$$

power converges to unity asymptotically. Then the estimator $\widehat{G}$ of (26) is consistent and the proof is complete. ∎

D Simulation & Empirical Results

D.1 Simulation results

Table 1: Powers of the right-tailed DF t- and J-tests when a bubble is short-lived

ρ	1.033			1.040		
T	10	20	30	10	20	30
Power of DF t-test	0.1009	0.1203	0.1676	0.1043	0.1381	0.2202
Power of DF J-test	0.0957	0.1202	0.1716	0.1042	0.1393	0.2261

Table 2: Powers of the right-tailed DF t- and J-tests when the bubble grows slowly

ρ	1.0009			1.0069		
T	50	100	200	50	100	200
Power of DF t-test	0.0590	0.0590	0.0647	0.0801	0.1127	0.2913
Power of DF J-test	0.0597	0.0590	0.0647	0.0808	0.1142	0.2991

Table 3: SNRs for mildly explosive, unit root and mildly stationary groups

| n	T	DGP 1			DGP 2			DGP 2	
		$c_1(\times 10^5)$	$c_2(\times 10^2)$	$c_3(\times 1)$	$c_1(\times 10^5)$	$c_2(\times 10^2)$	$c_3(\times 1)$	$c_1(\times 10^5)$	$c_2(\times 1)$
30	100	6.798	2.792	8.375	6.814	7.338	8.493	6.726	47.096
30	150	82.066	4.935	13.139	82.202	16.354	12.926	81.978	71.135
30	200	562.425	6.942	15.426	562.771	28.344	15.426	559.628	85.447
60	100	6.595	2.859	8.881	6.606	7.458	8.623	6.693	47.220
60	150	81.352	5.009	13.496	81.506	16.580	13.088	81.527	71.234
60	200	562.555	6.974	15.686	563.078	28.350	15.584	561.495	85.547
90	100	6.668	2.852	9.026	6.676	7.434	8.642	6.724	47.204
90	150	81.778	5.077	13.408	81.935	16.816	13.155	82.117	71.359
90	200	565.714	7.013	15.754	566.288	28.469	15.641	561.739	85.688
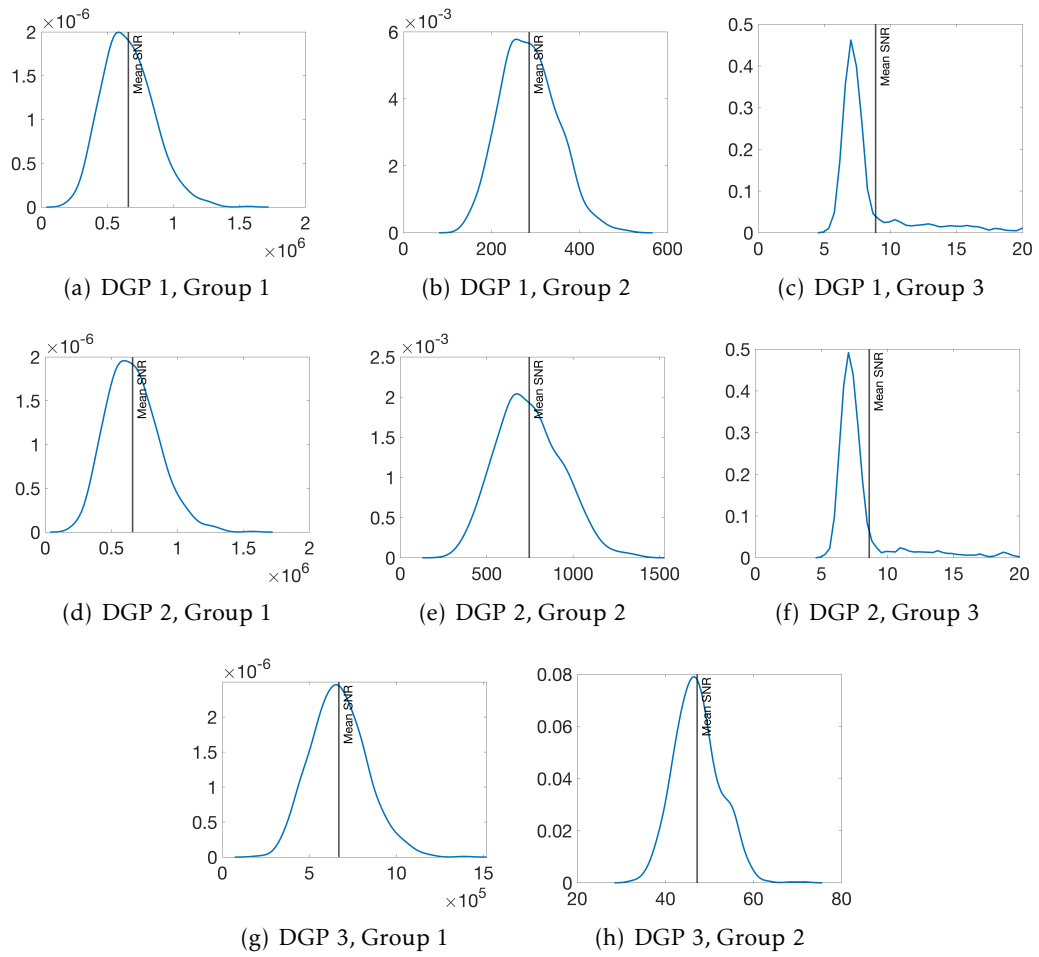
Figure 1: Empirical frequency distribution and sample average (shown by the vertical line) of the signal-to-noise ratio in each group of DGPs 1–3.

Table 4: 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 5: 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 6: 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 7: 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 8: 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 9: 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 10: 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 11: 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

D.2 Empirical results
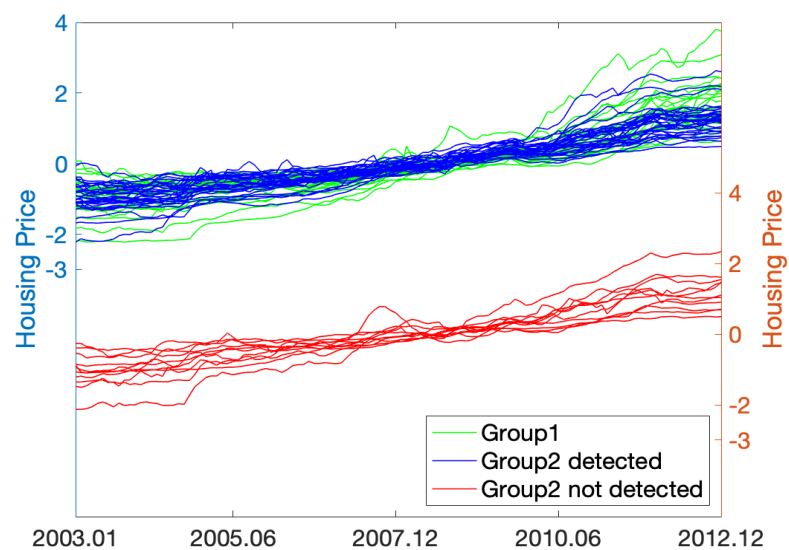
Figure 2: Chinese house price index (demeaned ratio of price index to rental CPI - see the text) for all cities detected by the panel t-test, separated into those cities also detected by the time series test and cities not detected by the individual time series t-test.
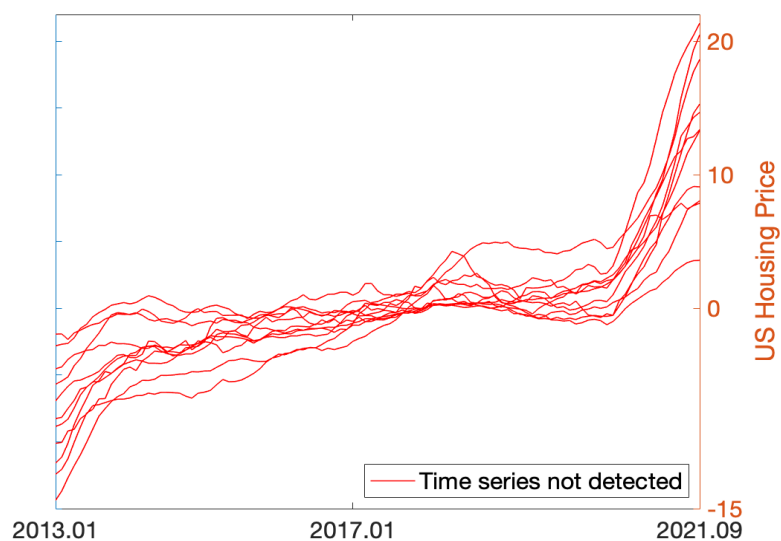

Figure 3: US house price index for the cities detected by the panel t-test but not by the single time series t-test.
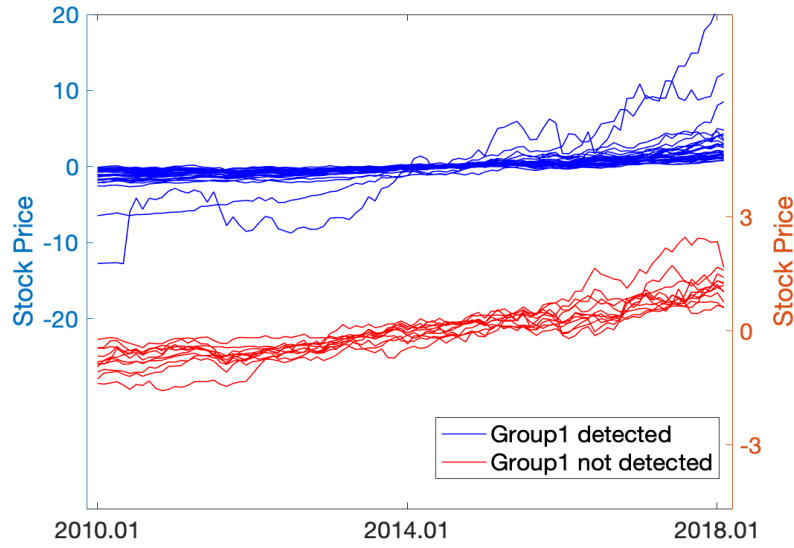
Figure 4: The US stock price time series for the cities detected by the panel t-test (blue lines) but not by the single time series t-test (red lines).

Table 12: Selection by IC of the number of groups G in Chinese city housing markets

G	1	2	3	4	5
IC	-14.8792	-14.8777	**-14.8817**	-14.8810	-14.8793

Note: The smallest IF value is in boldface.

Table 13: Post-clustering estimates and panel t- and J-tests in China's housing market

Groups	Group 1	Group 2	Group 3
$\breve{n}_j$	27	42	38
$\breve{\rho}_j$	1.011	1.003	0.995
t-test	7.140***	4.409***	-0.480
J-test	4.743***	2.313**	-0.356

Note: *, ** and *** imply rejection of the null hypothesis at the 10%, 5% and 1% levels.

Table 14: Selection of G by IC in the US housing market

G	1	2	3	4	5
IC	-11.1665	**-11.1854**	-11.1823	-11.1752	-11.1658

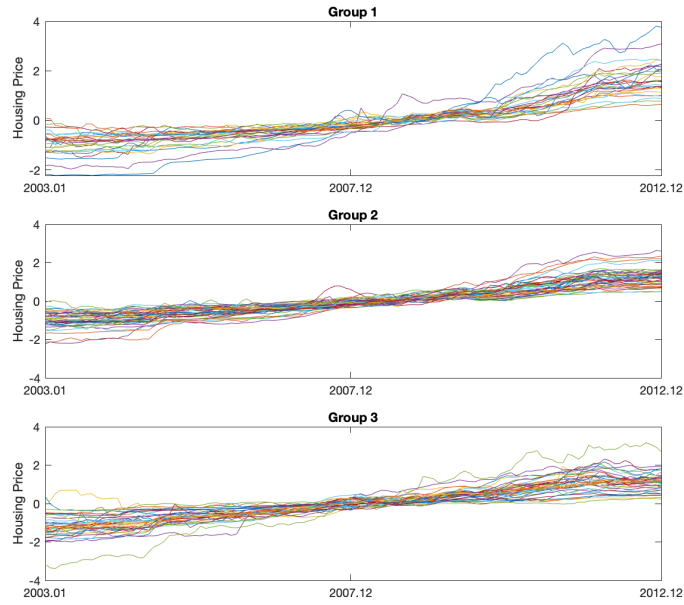Note: The smallest IF value is in boldface.

Figure 5: Times series plots of the demeaned ratio of Chinese city house price indices to the rental CPI (vertical axis) for the estimated groups.
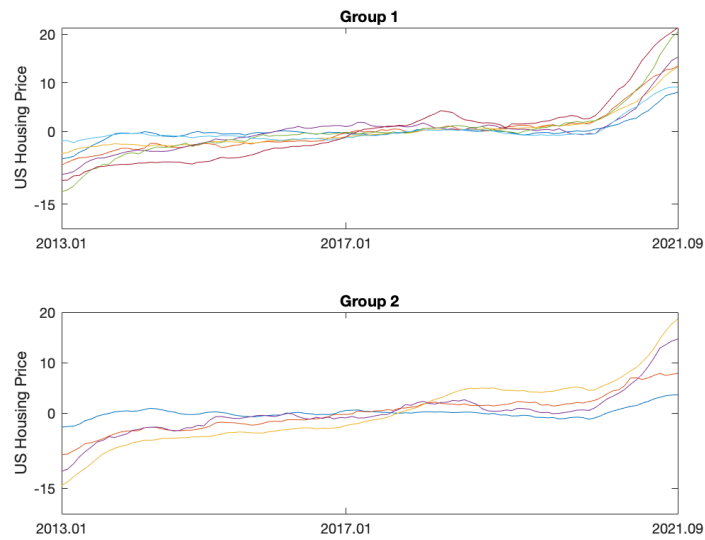

Figure 6: Time series plots of the demeaned ratio of the housing price to the rental CPI (vertical axis) in the two estimated groups of the US housing market

Table 15: Post-clustering estimate and the panel t- and J-tests in the US housing market

Groups	Group 1	Group 2
$\widecheck{n}_j$	7	4
$\widecheck{\rho}_j$	1.0432	1.0113
t-test	31.9880***	1.8617*
J-test	35.5062***	2.3412**

Note: *, ** and *** imply rejection of the null hypothesis at the 10%, 5% and 1% levels.

Table 16: Selection of the group number G by IC in the US equity market

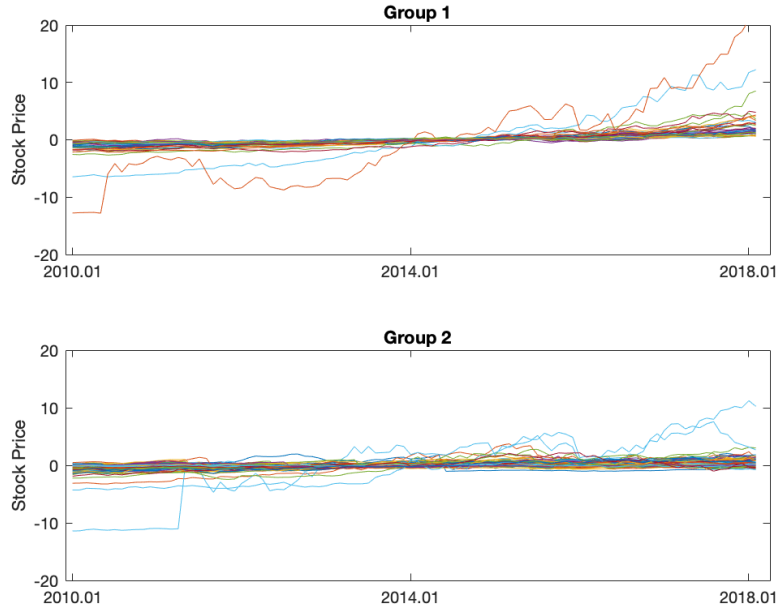| G	1	2	3	4	5
IC	-2.9545	**-2.9639**	-2.9631	-2.9615	-2.9588

Note: The smallest IF value is in boldface.

Figure 7: Time series plots of the demeaned difference between S&P 500 stock prices and dividends in the two estimated groups.

Table 17: Post-clustering estimates and panel t- and J-tests in the US equity market

Groups	Group 1	Group 2
$\widecheck{n}_j$	40	106
$\widecheck{\rho}_j$	1.012	0.967
t-test	2.236**	0.141
J-test	8.908***	1.330

Note: *, ** and *** imply rejection of the null hypothesis at the 10%, 5% and 1% levels.

Notation Glossary

Symbol	Description
G	Number of groups;
G^0	True number of groups;
$\widehat{G}$	Estimated number of groups generated by the combined method based on the use of IC and the Hausman test
j	Subscript for group identities, namely $1 \leq j \leq G^0$ or $1 \leq j \leq G$;
i	Subscript for individuals, namely $1 \leq i \leq n$;
$\mathcal{I}_n$	Set of individual subscripts $\{1, 2, ..., n\}$;
$\mathcal{G}$	Set of group identities $\{1, 2, ..., G\}$;
$\mathcal{G}^0$	Set of group identities $\{1, 2, ..., G^0\}$;
g_i	Membership indicator mapping from individuals $\mathcal{I}_n$ into group identities $\mathcal{G}^0$;
g_i^0	True membership indicator for (g_i);
$\widehat{g_i}$	Estimated membership indicator for (g_i) generated by recursive k-means clustering;
δ	Collection of membership indicators $(:= (g_1, g_2, ..., g_n)')$;
δ^0	True value of δ as $\delta^0 \left(:= \left(g_1^0, g_2^0, ..., g_n^0\right)'\right)$;
$\widehat{\delta}$	Estimation of δ as $\widehat{\delta}\left(:= (\widehat{g_1}, \widehat{g_2}, ..., \widehat{g_n})'\right)$;
Δ_{G^0}	Set of all possible δ, so that $\delta \in \Delta_{G^0}$;
$\mathcal{G}(j)$	Individuals of the jth group; for instance, $\mathcal{G}(j) = \{i \in \mathcal{I}_n | g_i = j\}$;
$\mathcal{G}^0(j)$	Individuals of the true jth group; for instance, $\mathcal{G}^0(j) = \left\{i \in \mathcal{I}_n | g_i^0 = j\right\}$;
$\hat{\mathcal{G}}(j)$	Individuals of the estimated jth group; for instance, $\hat{\mathcal{G}}(j) = \{i \in \mathcal{I}_n | \widehat{g_i} = j\}$;
c_j	Group-specific distancing parameter; for instance, $c_j \left(:= c_{g_i}\right)$ if $g_i = j$;
c_j^0	True value of group-specific distancing parameter c_j;
$\widehat{c}_j^*$	First-stage estimate of group-specific parameter c_j by recursive k-means clustering;
$\widehat{c}_j$	Oracle estimate of group-specific parameter c_j based on the true membership δ^0;
$\check{c}_j$	Post-clustering estimate of group-specific parameter c_j based on the estimation $\widehat{\delta}$;
$\widehat{\overline{c}}_i^{TS}$	Time series estimate of individual parameter $\bar{c}_i$;
$\mathcal{C}_G$ and $\mathcal{C}_{G^0}$	Set of all possible G-dimensional or G^0-dimensional distance parameter c;
ρ_j	Group-specific slope coefficient $\rho_j = 1 + c_j/T^\gamma$;

Symbol	Description
ρ_j^0	True value of group-specific slope coefficient $\rho_j^0 = 1 + c_j^0/T^\gamma$;
$\widetilde{\rho}_j^*$	First-stage estimate of group-specific slope coefficient $\widetilde{\rho}_j^* = 1 + \widetilde{c}_j^*/T^\gamma$;
$\widehat{\rho}_j$	Oralce estimate of group-specific slope coefficient $\widehat{\rho}_j = 1 + \widehat{c}_j/T^\gamma$;
$\check{\rho}_j$	Post-clustering estimate of group-specific slope coefficient $\check{\rho}_j = 1 + \check{c}_j/T^\gamma$;
$\widehat{\overline{\rho}}_i^{TS}$	Time series estimate of individual parameter $\overline{\rho}_i$;
c, ρ	G^0-dimensional parameters $c = (c_1, c_2, ..., c_{G^0})'$ and $\rho = (\rho_1, \rho_2, ..., \rho_{G^0})'$;
c^0, ρ^0	G^0-dimensional true values $c^0 = \left(c_1^0, c_2^0, ..., c_{G^0}^0\right)'$ and $\rho^0 = \left(\rho_1^0, \rho_2^0, ..., \rho_{G^0}^0\right)'$;
$\widetilde{c}^*, \widetilde{\rho}^*$	First-stage estimate of c, ρ as $\widetilde{c}^* = \left(\widetilde{c}_1^*, \widetilde{c}_2^*, ..., \widetilde{c}_{G^0}^*\right)'$ and $\widetilde{\rho}^* = \left(\widetilde{\rho}_1^*, \widetilde{\rho}_2^*, ..., \widetilde{\rho}_{G^0}^*\right)'$;
$\widehat{c}, \widehat{\rho}$	Oracle estimate of c, ρ as $\widehat{c} = (\widehat{c}_1, \widehat{c}_2, ..., \widehat{c}_{G^0})'$ and $\widehat{\rho} = (\widehat{\rho}_1, \widehat{\rho}_2, ..., \widehat{\rho}_{G^0})'$;
$\check{c}, \check{\rho}$	Post-clustering estimate of c, ρ as $\check{c} = (\check{c}_1, \check{c}_2, ..., \check{c}_{G^0})'$ and $\check{\rho} = (\check{\rho}_1, \check{\rho}_2, ..., \check{\rho}_{G^0})'$;
c_{low}, c_{up}	Bounds for c_j; for instance, $c_j \in \left[-c_{up}, -c_{low}\right] \cup \{0\} \cup \left[c_{low}, c_{up}\right]$;
ρ_{low}, ρ_{up}	Bounds for ρ_j; for instance $\rho_{up} = 1 + \frac{c_{up}}{T^\gamma}$ and $\rho_{low} = 1 + \frac{c_{low}}{T^\gamma}$;
$\dot{c}$	Separation for c_j; for instance $\inf_{j \neq \widetilde{j}}\left|c_j - c_{\widetilde{j}}\right| \geq \dot{c}$;
$\sigma_j^2, \lambda_j, \omega_j^2$	Group-specific parameters for variances, one-sided and two-sided long run variances in the jth group;
n_j	Number of individuals in the jth true group;
$\check{n}_j$	Number of individuals in the jth estimated group;
$\left(\sigma_j^0\right)^2, \lambda_j^0, \left(\omega_j^0\right)^2$	True values for variances, one-sided and two-sided long run variances in the true jth group;
$\left(\check{\sigma}_j\right)^2, \check{\lambda}_j, \left(\check{\omega}_j\right)^2$	Post-clustering estimates for variances, one-sided and two-sided long run variances in the estimated jth group;
$\overline{c}_i, \overline{\rho}_i$	Individual distancing parameter and slope coefficient for the ith individual; namely, $\overline{\rho}_i = 1 + \frac{\overline{c}_i}{T^\gamma}$;
$\overline{c}_i^0, \overline{\rho}_i^0$	True values for $\overline{c}_i$ and $\overline{\rho}_i$; namely, $\overline{\rho}_i^0 = 1 + \frac{\overline{c}_i^0}{T^\gamma}$;
$\widehat{\overline{c}}_i, \widehat{\overline{\rho}}_i$	Post-clustering estimations for $\overline{c}_i$ and $\overline{\rho}_i$; namely, $\widehat{\overline{\rho}}_i = 1 + \frac{\widehat{\overline{c}}_i}{T^\gamma}$;
$\overline{c}, \overline{\rho}$	$\overline{c} = (\overline{c}_1, \overline{c}_2, ..., \overline{c}_n)'$ and $\overline{\rho} = \left(\overline{\rho}_1, \overline{\rho}_2, ..., \overline{\rho}_n\right)'$;
$\overline{c}^0, \overline{\rho}^0$	$\overline{c}^0 = \left(\overline{c}_1^0, \overline{c}_2^0, ..., \overline{c}_n^0\right)'$ and $\overline{\rho}^0 = \left(\overline{\rho}_1^0, \overline{\rho}_2^0, ..., \overline{\rho}_n^0\right)'$;
$\widehat{\overline{c}}, \widehat{\overline{\rho}}$	$\widehat{\overline{c}} = \left(\widehat{\overline{c}}_1, \widehat{\overline{c}}_2, ..., \widehat{\overline{c}}_n\right)'$ and $\widehat{\overline{\rho}} = \left(\widehat{\overline{\rho}}_1, \widehat{\overline{\rho}}_2, ..., \widehat{\overline{\rho}}_n\right)'$;

Symbol	Description
$\sigma^2, \left(\sigma^0\right)^2$	Parameter representation and true value of $Var(\epsilon_{it})$;
$\overline{\sigma}_{iu}^2, \overline{\lambda}_i, \overline{\omega}_i^2$	Individual parameters for variances, one-sided and two-sided long run variances of the ith individual; for instance, $\overline{\sigma}_{iu}^2 := \sigma_{g_i}^2$, $\overline{\lambda}_i := \lambda_{g_i}$ and $\overline{\omega}_i^2 := \omega_{g_i}^2$;
$\left(\overline{\sigma}_{iu}^0\right)^2, \overline{\lambda}_i^0, \left(\overline{\omega}_i^0\right)^2$	True values for variances, one-sided and two-sided long run variances of the ith individual; for instance, $\left(\overline{\sigma}_{iu}^0\right)^2 := \left(\sigma_{g_i^0}^0\right)^2$, $\overline{\lambda}_i^0 := \lambda_{g_i^0}^0$ and $\left(\overline{\omega}_i^0\right)^2 := \left(\omega_{g_i^0}^0\right)^2$;
$\left(\widehat{\overline{\sigma}}_{iu}\right)^2, \widehat{\overline{\lambda}}_i, \left(\widehat{\overline{\omega}}_i\right)^2$	Individual time series estimates for variances, one-sided and two-sided long run variances of the ith individual, based on the post-clustering estimate $\check{\rho}_j$ with $\widehat{g}_i = j$;
$I_d, \mathbf{0}_{d \times d}$	A $d \times d$ identity matrix; A $d \times d$ matrix of zeros;
$\to_p, \Rightarrow$	Convergence in probability; Weak convergence in the Euclidean space or functional space;
$A_{nT} \preceq B_{nT}$	A_{nT}/B_{nT} is either $O_p(1)$ or $o_p(1)$ as $(n,T) \to \infty$;
$A_{nT} \succ B_{nT}$	B_{nT}/A_{nT} is $o_p(1)$ as $(n,T) \to \infty$;
$A_{nT} \sim_a B_{nT}$	$\Pr(|B_{nT}/A_{nT}| \neq 1) \to 0$ as $(n,T) \to \infty$;
$A_{nT} \sim B_{nT}$	$|B_{nT}/A_{nT}| \to 1$ as $(n,T) \to \infty$;
$G_{\max}$	Generic upper bound for group number G;
$\overline{G} := G_{\max} - G + 1$	The maximum number of subgroups in each group $j = 1,2,...,G$;
$\widetilde{G}$	The lower bound of group number selected by IC as in (20);
$\check{\rho}_j(G), \check{\rho}_{j,h}(G), h = 1,...\overline{G}$	Post-clustering pooled LS estimators for slopes in the estimated j group and in the estimated hth subgroup of the estimated jth group when assuming G groups, as in (21);
$\widehat{g}_{j,h}, h = 1,...,\overline{G}$	The estimated membership of the estimated hth subgroup of the estimated jth group;
$\widehat{\delta}_j(G)$	The estimated membership of subgroups in the estimated jth group, and $\widehat{\delta}_j(G) = \left(\widehat{g}_{j,1}(G), \widehat{g}_{j,2}(G),...,\widehat{g}_{j,\overline{G}}(G)\right)'$;
$\check{\boldsymbol{\rho}}_j(G)$	The estimated slopes of subgroups in the estimated jth group, and $\check{\boldsymbol{\rho}}_j(G) = \left(\check{\rho}_{j,1}(G), \check{\rho}_{j,2}(G),...,\check{\rho}_{j,\overline{G}}(G)\right)'$;
$\hat{\mathcal{G}}(j,h,G), h = 1,2,...,\overline{G}$	Individuals in the estimated hth subgroup of the estimated jth group;
$\check{n}_{j,h}, h = 1,2,...,\overline{G}$	The dimension of the estimated hth subgroup in the estimated jth group, assuming G groups;
$\pi_{j,h}$	$\lim_n \frac{\check{n}_{j,h}}{\check{n}_j} \to \pi_{j,h}$;
$\widetilde{\pi}_j$	$\widetilde{\pi}_j = \text{diag}\left\{\pi_{j,1}, \pi_{j,2},...,\pi_{j,\overline{G}}\right\}$.

References

ANDO, T. AND J. BAI (2016): "Panel data models with grouped factor structure under unknown group membership," *Journal of Applied Econometrics*, 31, 163–191.

ARELLANO, M. AND S. BONHOMME (2012): "Identifying distributional characteristics in random coefficients panel data models," *The Review of Economic Studies*, 79, 987–1020.

ARELLANO, M., J. HAHN, ET AL. (2007): "Understanding bias in nonlinear panel models: Some recent developments," *Econometric Society Monographs*, 43, 381.

BAI, J. (2009): "Panel data models with interactive fixed effects," *Econometrica*, 77, 1229–1279.

BAI, J. AND S. NG (2002): "Determining the number of factors in approximate factor models," *Econometrica*, 70, 191–221.

——— (2004): "A PANIC attack on unit roots and cointegration," *Econometrica*, 72, 1127–1177.

——— (2010): "Panel unit root tests with cross-section dependence: a further investigation," *Econometric Theory*, 26, 1088–1114.

BONHOMME, S., T. LAMADON, AND E. MANRESA (2017): "Discretizing unobserved heterogeneity," *University of Chicago, Becker Friedman Institute for Economics Working Paper*.

BONHOMME, S. AND E. MANRESA (2015): "Grouped patterns of heterogeneity in panel data," *Econometrica*, 83, 1147–1184.

BORDT, S., H. FARBMACHER, AND H. KÖGEL (2020): "Estimating grouped patterns of heterogeneity in repeated public goods experiments," Tech. rep., Working paper.

CHANG, Y. AND W. SONG (2009): "Testing for unit roots in small panels with short-run and long-run cross-sectional dependencies," *The Review of Economic Studies*, 76, 903–935.

CHEN, K. AND Y. WEN (2017): "The great housing boom of China," *American Economic Journal: Macroeconomics*, 9, 73–114.

CHEN, Y. Z., P. C. B. PHILLIPS, AND S. SHI (2022): "Common bubble detection in large dimensional financial systems," *Journal of Financial Econometrics*, *doi.org/10.1093/jjfinec/nbab027*.

DAVID, H., A. MANNING, AND C. L. SMITH (2016): "The contribution of the minimum wage to US wage inequality over three decades: a reassessment," *American Economic Journal: Applied Economics*, 8, 58–99.

DHAENE, G. AND K. JOCHMANS (2015): "Split-panel jackknife estimation of fixed-effect models," *The Review of Economic Studies*, 82, 991–1030.

DIBA, B. T. AND H. I. GROSSMAN (1987): "On the inception of rational bubbles," *The Quarterly Journal of Economics*, 102, 697–700.

——— (1988): "Explosive rational bubbles in stock prices?" *The American Economic Review*, 78, 520–530.

Dzemski, A. and R. Okui (2021): "Convergence rate of estimators of clustered panel models with misclassification," *Economics Letters*, 203, 109844.

Fang, H., Q. Gu, W. Xiong, and L.-A. Zhou (2016): "Demystifying the Chinese housing boom," *NBER Macroeconomics Annual*, 30, 105–166.

Gouriéroux, C., P. C. B. Phillips, and J. Yu (2010): "Indirect inference for dynamic panel models," *Journal of Econometrics*, 157, 68–77.

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

Hahn, J. and H. R. Moon (2010): "Panel data models with finite number of multiple equilibria," *Econometric Theory*, 26, 863–881.

Hahn, J. and W. Newey (2004): "Jackknife and analytical bias reduction for nonlinear panel models," *Econometrica*, 72, 1295–1319.

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

Hsiao, C. (2014): *Analysis of panel data*, 54, Cambridge university press.

Hsiao, C., M. H. Pesaran, and A. K. Tahmiscioglu (2002): "Maximum likelihood estimation of fixed effects dynamic panel data models covering short time periods," *Journal of Econometrics*, 109, 107–150.

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.

Huang, W., S. Jin, and L. Su (2020): "Identifying latent grouped patterns in cointegrated panels," *Econometric Theory*, 36, 410–456.

Im, K. S., M. H. Pesaran, and Y. Shin (2003): "Testing for unit roots in heterogeneous panels," *Journal of Econometrics*, 115, 53–74.

Kong, J., P. C. Phillips, and D. Sul (2019): "Weak σ-convergence: Theory and applications," *Journal of Econometrics*, 209, 185–207.

Leeb, H. and B. M. Pötscher (2005): "Model selection and inference: Facts and fiction," *Econometric Theory*, 21, 21–59.

——— (2008): "Sparse estimators and the oracle property, or the return of Hodges' estimator," *Journal of Econometrics*, 142, 201–211.

Liu, R., Z. Shang, Y. Zhang, and Q. Zhou (2020): "Identification and estimation in panel models with overspecified number of groups," *Journal of Econometrics*, 215, 574–590.

Liu, Y. (2019): "Panel approaches to econometric analysis of bubble behaviour," *Job Market Paper, Singapore Management University*.

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

Miao, K., L. Su, and W. Wang (2020): "Panel threshold regressions with latent group structures," *Journal of Econometrics*, 214, 451–481.

Moon, H. R. and B. Perron (2004): "Testing for a unit root in panels with dynamic factors," *Journal of Econometrics*, 122, 81–126.

Narayan, P. K., S. Mishra, S. Sharma, and R. Liu (2013): "Determinants of stock price bubbles," *Economic Modelling*, 35, 661–667.

Okui, R. and W. Wang (2021): "Heterogeneous structural breaks in panel data models," *Journal of Econometrics*, 220, 447–473.

Pesaran, H., R. Smith, and K. S. Im (1996): "Dynamic linear models for heterogenous panels," in *The econometrics of panel data*, Springer, 145–195.

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

Pesaran, M. H. and R. Smith (1995): "Estimating long-run relationships from dynamic heterogeneous panels," *Journal of Econometrics*, 68, 79–113.

Pesaran, M. H. and T. Yamagata (2008): "Testing slope homogeneity in large panels," *Journal of Econometrics*, 142, 50–93.

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 J. H. Lee (2013): "Predictive regression under various degrees of persistence and robust long-horizon regression," *Journal of Econometrics*, 177, 250–264.

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.

——— (2009): "Unit root and cointegrating limit theory when initialization is in the infinite past," *Econometric Theory*, 25, 1682–1715.

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., S. SHI, AND J. YU (2015a): "Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500," *International Economic Review*, 56, 1043–1078.

——— (2015b): "Testing for multiple bubbles: Limit theory of real-time detectors," *International Economic Review*, 56, 1079–1134.

PHILLIPS, P. C. B. AND V. SOLO (1992): "Asymptotics for linear processes," *The Annals of Statistics*, 971–1001.

PHILLIPS, P. C. B. AND D. SUL (2003): "Dynamic panel estimation and homogeneity testing under cross section dependence," *The Econometrics Journal*, 6, 217–259.

PHILLIPS, P. C. B., Y. WU, AND J. YU (2011): "Explosive behavior in the 1990s Nasdaq: When did exuberance escalate asset values?" *International Economic Review*, 52, 201–226.

PHILLIPS, P. C. B. AND J. YU (2009): "Limit theory for dating the origination and collapse of mildly explosive periods in time series data," *Working paper, School of Economics, Singapore Management University*, 2009–09.

——— (2011): "Dating the timeline of financial bubbles during the subprime crisis," *Quantitative Economics*, 2, 455–491.

SHILLER, R. J. (2015): *Irrational exuberance*, Princeton University Press.

SU, L. AND Q. CHEN (2013): "Testing homogeneity in panel data models with interactive fixed effects," *Econometric Theory*, 29, 1079–1135.

SU, L., Z. SHI, AND P. C. B. PHILLIPS (2016): "Identifying latent structures in panel data," *Econometrica*, 84, 2215–2264.

SU, L., W. WANG, AND Y. ZHANG (2019): "Strong consistency of spectral clustering for stochastic block models," *IEEE Transactions on Information Theory*, 66, 324–338.

WANG, W., P. C. B. PHILLIPS, AND L. SU (2018): "Homogeneity pursuit in panel data models: Theory and application," *Journal of Applied Econometrics*, 33, 797–815.

——— (2019): "The heterogeneous effects of the minimum wage on employment across states," *Economics Letters*, 174, 179–185.

WANG, W. AND L. SU (2021): "Identifying latent group structures in nonlinear panels," *Journal of Econometrics*, 220, 272–295.

WESTERLUND, J. (2015): "The power of PANIC," *Journal of Econometrics*, 185, 495–509.

WESTERLUND, J. AND R. LARSSON (2012): "Testing for a unit root in a random coefficient panel data model," *Journal of Econometrics*, 167, 254–273.