A PANEL CLUSTERING APPROACH TO ANALYZING BUBBLE BEHAVIOR

By

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

COWLES FOUNDATION PAPER NO. 1862

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

http://cowles.yale.edu/

INTERNATIONAL ECONOMIC REVIEW
Vol. 64, No. 4, November 2023 DOI: 10.1111/iere.12647

A PANEL CLUSTERING APPROACH TO ANALYZING BUBBLE BEHAVIOR*

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

Yanbo Liu: Shandong University, China; Peter C. B. Phillips: Yale University, USA, and University of Auckland, New Zealand, and Singapore Management University, Singapore; Jun Yu: Singapore Management University, Singapore, and University of Macau, China

This study provides new mechanisms for identifying and estimating explosive bubbles in mixed-root panel autoregressions with a latent group structure. A postclustering approach is employed that combines k-means clustering with right-tailed panel-data testing. Uniform consistency of the k-means algorithm is established. Pivotal null limit distributions of the tests are introduced. A new method is proposed to consistently estimate the number of groups. Monte Carlo simulations show that the proposed methods perform well in finite samples; and empirical applications of the proposed methods identify bubbles in the U.S. and Chinese housing markets and the U.S. stock market.

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

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

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:

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

*Manuscript received March 2022; revised April 2023.

Thanks go to Jesus Fernandez-Villaverde (Editorial Board), three referees, Yong Bao, Timothy Christensen, Liyu Dou, Wayne Gao, Jia Li, Tassos Magdalinos, Morten Nielsen, Frank Schorfheide, Liangjun Su, Cheng Xu, Yichong Zhang, and the participants of the 2021 SH3, the 2022 SETA, and the 2022 AMES conferences, and the 2022 Symposium on Econometrics at XMU for helpful suggestions and discussions. Phillips acknowledges support from a Lee Kong Chian Fellowship at SMU, the Kelly Fund at the University of Auckland, and the NSF under Grant No. SES 18-50860. Yu acknowledges that this research/project is supported by the Ministry of Education, Singapore, under its Academic Research Fund (AcRF) Tier 2 (Award Number MOE-T2EP402A20-0002). Please address correspondence to: Peter C. B. Phillips, Cowles Foundation for Research in Economics, Yale University, New Haven, CT 0625, USA. E-mail: *peter.phillips@yale.edu*

TABLE 1

POWERS OF THE RIGHT-TAILED DF t- AND J-TESTS WHEN A BUBBLE IS SHORT LIVED OR GROWS SLOWLY

	Bubble Is Short Lived						Bubble Grows Slowly					
ρ		1.033			1.040			1.0009			1.0069	
T	10	20	30	10	20	30	50	100	200	50	100	200
t-test	0.1009	0.1203	0.1676	0.1043	0.1381	0.2202	0.0590	0.0590	0.0647	0.0801	0.1127	0.2913
J-test	0.0957	0.1202	0.1716	0.1042	0.1393	0.2261	0.0597	0.0590	0.0647	0.0808	0.1142	0.2991

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 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, 2015b).

Conventional econometric methods for bubble detection, including the Dickey–Fuller (DF) and augmented DF (ADF) tests (Diba and Grossman, 1988), the sup ADF (SADF) test (Phillips and Yu, 2011; Phillips et al., 2011), and the generalized sup ADF (GSADF) test (Phillips et al., 2015a, 2015b), 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. In order to illustrate the low-power problem of conventional single series tests when a bubble is short lived or grows slowly, we employed two experiments with simulated data from the following simple AR(1) design,

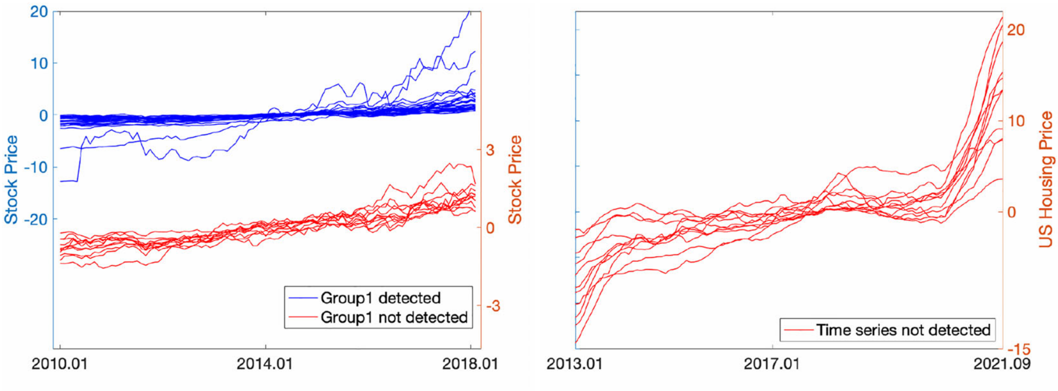
$$y_t = \rho y_{t-1} + u_t,\ y_0 = 0,\ u_t \sim \text{i.i.d.}\ \mathcal{N}(0,1),\ t = 1, 2, \ldots, T, \tag{2}$$

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

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

is the least-squares (LS) estimator of ρ, with $\bar{y} = (1/T)\sum_{s=1}^{T} y_s$, $\bar{y}_{-1} = (1/T)\sum_{s=1}^{T} y_{s-1}$, 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). The left panel of 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. The right panel of Table 1 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.

As an empirical illustration of low discriminatory power from single time-series regressions, the t-test was applied to each of the 146 stocks of the S&P 500 market in the United States



(a) US stock market (b) US housing market

NOTES: Left panel: Adjusted U.S. equity prices (specifically, the demeaned difference in levels between the monthly stock price and monthly dividends) for the stocks detected by the panel t-test but not by the single time-series t-test (in red); Right panel: Adjusted U.S. house price index (specifically, the demeaned ratio of the house price index to the rental CPI) for the cities detected by the panel t-test but not by the single time series t-test

FIGURE 1

PANEL AND TIME SERIES TESTING OF U.S. STOCK PRICES AND HOUSING PRICES FOR EXPLOSIVE BEHAVIOR

used in the empirical section of the article—see Section 6 for details. The results led to rejections of the unit root null in favor of an explosive alternative for 29 stocks, with tests for all other 117 stocks failing to reject. Figure 1(a) compares these test results with those of the new panel tests that make use of the clustering algorithm to identify clusters of stocks 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. Our clustered panel test detects the presence of an explosive root in 11 more stocks than the time-series test, as highlighted in red in Figure 1(a). An empirical example of the U.S. housing market is given in Figure 1(b). In this case, the clustered panel t-test helps to diagnose mildly explosive price behavior in U.S. city housing markets (again shown in red), where individual time-series tests reveal no evidence of such behavior, confirming the discriminatory power gains that arise from cross-section aggregation.

As indicated, 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 this article 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.

Under certain conditions it is often reasonable to expect homogeneity over cross-sectional units in economic and financial panel data (Hahn and Moon, 2010; Kong et al., 2019). In real estate markets, the dynamics in house prices may be related among cities with similar characteristics, including demographics, locational features, and level of urbanization. But the use of a general aggregated homogeneous model in estimating housing prices has been criticized due to its omission of individual elements, such as spatial correlation and heteroskedasticity (Goodman and Thibodeau, 1998). In fact, it is well acknowledged that housing properties are heterogeneous regarding such characteristics as location, structural characteristics, and environmental quality. To address this feature of the data, a standard modeling strategy in urban studies is market segmentation, where a housing market is divided into different submarkets, each of which may be considered homogeneous (Bourassa et al., 2007; Goodman and Thibodeau, 1998). Empirical researchers can therefore make use of a data-driven clustering algorithm to uncover the underlying submarket structure (Abraham et al., 1994).

In financial markets, individual stocks in the same industry may share similar fundamentals, and their dynamic behavior may be closely related and can be aggregated into groups (Bollerslev et al., 2022). Mutual funds may be grouped by the type of security, such as aggressive growth funds and income funds (Brown and Goetzmann, 1997). In the literature, data-driven clustering methods have been applied to stocks, bonds, hedge funds, and mutual funds in Ahn et al. (2009), Ludvigson and Ng (2007), Ludvigson and Ng (2009), Patton and Weller (2022), and Brown and Goetzmann (1997).

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, 2007b; 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., 2022) and the classification Lasso (C-Lasso) approach of Su et al. (2016). This article 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 postclustering 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 that 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 article makes six 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 (Bai and Ng, 2010; Chang and Song, 2009; Im et al., 2003). 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 this article to allow for latent groups with different factors that reflect differing behavior among the groups.

Second, the article contributes to the literature on latent membership and clustering algorithms. There are presently several clustering algorithms (Ando and Bai, 2016; Bonhomme and Manresa, 2015; Bonhomme et al., 2022; Leng et al., 2021; Okui and Wang, 2021; Su et al., 2016, 2019), and most of the available methods apply only to stationary data. An exception is the C-Lasso approach (Huang et al., 2020), 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 (Arellano and Bonhomme, 2012; Hsiao et al., 2002; Pesaran and Smith, 1995; Pesaran, 2006). 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.

A fourth contribution is 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., 2022; Su et al., 2016) 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). Both IC and residual-based inference approaches are found to underestimate the true number of groups in the mixed-root dynamic panel model. In order 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.

Fifth, the article adds to various existing empirical analyses by using a new clustering algorithm. For example, this article contributes to the urban economics literature by confirming the existence of latent membership in the U.S. and Chinese housing markets. In fact, our empirical results may be viewed as validating market segmentation theory (Bourassa et al., 1999; Goodman and Thibodeau, 1998). In a further application, the article contributes to the empirical finance literature on grouping and clustering the U.S. stock market. Although the k-means clustering algorithm has been successfully employed in classifying hedge funds, mutual funds, and individual stocks (Ahn et al., 2009; Brown and Goetzmann, 1997), to the best of our knowledge, this article is the first to document group-specific explosive behavior in the stock market using k-means clustering methods.

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) and nonlinear panels (Hahn and Newey, 2004). Bias correction methods include the use of explicit bias approximations (Hahn and Kuersteiner, 2002; Phillips and Moon, 1999), jackknife methods (Hahn and Newey, 2004), and indirect inference methods (Gouriéroux et al., 2010). The limit theory in this article 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 article 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 postclustering 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 U.S. real estate markets and the U.S. stock market. Section 7 concludes. Proofs of the main results are provided in the Appendix. Further technical details, some useful lemmas, and additional simulation findings are available in the Supporting Information that accompanies the article.

Throughout the article, the symbols I_d, $\ell_{d\times 1}$, $\mathbf{0}_{d\times d}$, $\to_p$, $\Rightarrow$, 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 $(n, T) \to \infty$; $A_{nT} \succ B_{nT}$ signifies that $B_{nT}/A_{nT} = o_p(1)$ as $(n, T) \to \infty$; $A_{nT} \sim B_{nT}$ signifies $\lim_{n,T\to\infty} A_{nT}/B_{nT} = 1$, $A_{nT} \sim_a B_{nT}$ denotes $\Pr(|A_{nT}/B_{nT}| \neq 1) \to 0$ as $(n, T) \to \infty$; the notation $A \asymp B$ denotes that both $A \preceq_p B$ and $B \preceq_p A$; 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

In order to capture explosive and mildly explosive behavior in panels, we use the following data generating process (DGP) based on the time-series model of Phillips and Magdalinos (2007b):

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

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 (29) 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 $\overline{\lambda}_i (:= \sum_{h=1}^{\infty} \mathbb{E}(u_{it}u_{i,t-h}))$, 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 accord-

[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 of the main article and details are provided in the supporting information to Liu et al. (2022).

ing 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., $i \in \{1, 2, \dots, n\}$) into specific groups for which $j \in \{1, \dots, G\}$ 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 in the following diagram organized in descending order of the values of the scale coefficients:

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

where $c_j \neq c_k$ for any $j \neq k$ with indices $j, k \in \{1, 2, \dots, G\}$. 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 $\{c_j > 0 | \ j = 1, \dots, g\}$; a single unit root group; and $(G - g - 1)$ mildly stationary groups, each with a different scale coefficient $\{c_j < 0 | \ j = g + 2, \dots, G\}$.

We begin by fixing notation. Denote the full set of n individuals by $\mathcal{I}_n := \{1, 2, \dots, n\}$ and membership indicators by the parameter vector $\delta (:= (g_1, g_2, \dots, g_n)')$. The true membership indicators are given by $\delta^0 (:= (g_1^0, g_2^0, \dots, g_n^0)')$. The estimated membership indicator, defined later, is $\widehat{\delta} (:= (\widehat{g}_1, \widehat{g}_2, \dots, \widehat{g}_n)')$. 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 $\mathcal{G} := \{1, 2, \dots, G\}$ 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 $\mathcal{G}^0 := \{1, 2, \dots, G^0\}$ 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 $\mathbb{R}^G$, $c(:= (c_1, \dots, c_j, \dots c_G)') \in \mathcal{C}_G$ be the distancing parameter vector, and $\rho := (\rho_1, \dots, \rho_j, \dots, \rho_G)' = (1 + c_1/T^\gamma, \dots, 1 + c_j/T^\gamma, \dots, 1 + c_G/T^\gamma)'$ 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 $\overline{\sigma}_{iu}^2 = \overline{\sigma}_{\ell u}^2 (=: \sigma_j^2)$, $\overline{\omega}_i^2 = \overline{\omega}_\ell^2 (=: \omega_j^2)$ and $\overline{\lambda}_i = \overline{\lambda}_\ell (=: \lambda_j)$. 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 by $\widehat{\sigma}_j^2$, $\widehat{\lambda}_j$, and $\widehat{\omega}_j^2$; and the post-clustering estimates, which rely on the estimated group identities, are denoted by $\check{\sigma}_j^2$, $\check{\lambda}_j$, and $\check{\omega}_j^2$ if the estimated membership is $\widehat{\delta} = (\widehat{g}_1, \dots, \widehat{g}_n)'$. Let the group-specific variances σ_j^2, λ_j, and ω_j^2 have true values denoted by $(\sigma_j^0)^2$, λ_j^0, and $(\omega_j^0)^2$. 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 article also considers the individual distancing parameters and estimates

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

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

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

REMARK 1. In the dynamic panel model (4), the mixed-root phenomenon is fully described by the group-specific distancing parameters c_{g_j} when the rate parameter γ is assumed to take a fixed common value. This assumption is imposed because the pair (c, γ) cannot be jointly identified, as discussed in Phillips (2023). An alternative modeling strategy is to let $\rho_{g_j} = (1 \pm 1/T^{\gamma_{g_j}})$, a nonlinear formulation used in Phillips (2023).[2] This alternative formulation requires estimation of the nonlinear rate parameters γ_{g_i} with different treatments and asymptotics for the cases $\gamma_{g_i} < 1$ and $\gamma_{g_i} \geq 1$, in the latter of which either ρ_{g_j} is local to unity when $\gamma_{g_i} = 1$ or even closer to unity when $\gamma_{g_i} > 1$. Use of this representation with heterogeneous rate parameters γ_{g_i} would involve major changes in the subsequent analysis and algorithm, so it is not pursued here. Further, in the dynamic panel model (4), the mixed-root phenomenon can be fully described by the group-specific distancing parameters c_j when the rate parameter γ is fixed, thereby providing a simple linear parametric framework of distancing. Determining latent membership via the scale parameters c_{g_i} is then empirically well suited to identifying both stationary and bubble directions of departure as well as unit roots.

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, postclustering 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 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, namely,

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

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

[2] Phillips (2023) shows that the model $y_t = \rho_T y_{t-1} + u_t$ with $\rho_T = 1 + c/T^\alpha$, $\alpha \in (0, 1)$, can be represented as $\rho_T = 1 + 1/T^{\gamma_T}$ with $\gamma_T = \alpha - (\log|c|)/(\log T) \in (0, 1)$ when $-T^\alpha < c < -1/T^{1-\alpha}$.

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.

(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 1-1 to update the estimates until convergence (say at step S). Define $\widehat{c}^* = c^{(S+1)}$ and $\widehat{\delta} = (g_1^{(S+1)}, \dots, g_n^{(S+1)})'$.

Instead of estimating c and δ simultaneously as in (8), which is numerically challenging, we follow Bonhomme and Manresa (2015) and use an iterative strategy to estimate c and δ recursively, as given in Algorithm 1 below. For convenience in the following derivations, knowledge of $\gamma \in (0, 1)$ is treated as prior information.[3] 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.[4] 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 $T \to \infty$.

REMARK 2. Algorithm 1 requires a well-specified initial value of δ to initiate the k-means clustering algorithm. Extensive simulations suggest that the estimated membership $\widehat{\delta}$ is highly accurate for different initial values in a large sample setting.[5] However, when the initial value is far from the correct value, although this does not lead to incorrect clustering outcomes, more iterations are needed to achieve convergence in membership estimation. Typically, when there is no prior information to guide the choice of the initial value, we recommend choosing nearly equally spaced initial draws of the time-series estimates, as is done in our simulations and empirics.

REMARK 3. The objective function in Equation (8) is stabilized by a self-normalized rate $O(n\Upsilon_{iT})$ that differs across units. Such a self-normalized rate presents a theoretical challenge that is not considered in the current panel clustering literature. For instance, both Bonhomme et al. (2022) and Su et al. (2016) applied the $O(nT)$ rate. However, the $O(nT)$ rate fails to stabilize the sample moments of the mixed-root individuals. In particular, the orders of different

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

[4] See Phillips (2023) for more details of parameter identification, estimation, and inference in mildly integrated and mildly explosive models.

[5] The robustness of membership estimation consistency even holds for choices such as the 0.1, 0.2, and 0.3 quantiles of the time-series estimates.

individuals in our mixed-root panel model are given below:

$$\sum_{t=1}^{T} \widetilde{y}_{i,t-1}^{2} \asymp \begin{cases} \left(\overline{\rho}_i^0\right)^{2T} T^{2\gamma} & \text{when } \overline{c}_i^0 > 0 \\ T^2 & \text{when } \overline{c}_i^0 = 0 \\ T^{1+\gamma} & \text{when } \overline{c}_i^0 < 0 \end{cases}. \tag{11}$$

The above rate conditions are derived in Phillips and Magdalinos (2007b) and Phillips (1987). It is clear from (11) that a common unified rate fails to normalize all mixed-root individuals. Instead, we use the self-normalized rate to address the challenge of stabilization.

3.2. *Stage 2: Postclustering Estimation and Testing.* Denote the true collection of members of the jth group as

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

Suppose the estimated membership indicator vector is $\widehat{\delta} = (\widehat{g}_1, \widehat{g}_2, \ldots, \widehat{g}_n)'$. Denote the estimated members of the jth group by

$$\hat{\mathcal{G}}(j) = \{i \in \mathcal{I}_n \mid \widehat{g}_i = j\} \forall j \in \mathcal{G}^0.$$

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

$$\widehat{\rho}_j = \frac{\sum_{i\in\mathcal{G}^0(j)} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}\widetilde{y}_{it}}{\sum_{i\in\mathcal{G}^0(j)} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}^2}. \tag{12}$$

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

$$\check{\rho}_j = \frac{\sum_{i\in\hat{\mathcal{G}}(j)} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}\widetilde{y}_{it}}{\sum_{i\in\hat{\mathcal{G}}(j)} \sum_{t=1}^{T} \widetilde{y}_{i,t-1}^2}. \tag{13}$$

Next define the following quantities:

$$\check{\sigma}_j^2 = \frac{1}{\check{n}_j} \sum_{i\in\hat{\mathcal{G}}(j)} \widehat{\widehat{\sigma}}_{iu}^2,\ \check{\omega}_j^2 = \frac{1}{\check{n}_j} \sum_{i\in\hat{\mathcal{G}}(j)} \widehat{\widehat{\omega}}_i^2,\ \check{\lambda}_j = \frac{1}{\check{n}_j} \sum_{i\in\hat{\mathcal{G}}(j)} \widehat{\widehat{\lambda}}_i,\ \check{E}_{j,nT} = \sum_{i\in\hat{\mathcal{G}}(j)} E_{i,nT}, \tag{14}$$

where

$$\widehat{\widehat{\omega}}_i^2 = \tfrac{1}{T}\sum_{t=1}^{T} \widehat{u}_{it}^2 + \tfrac{2}{T}\sum_{l=1}^{L}\sum_{t=l+1}^{T} w(l,L)\widehat{u}_{it}\widehat{u}_{i,t-l}, \tag{15}$$

$$\widehat{\widehat{\sigma}}_{iu}^2 = \tfrac{1}{T}\sum_{t=1}^{T} \widehat{u}_{it}^2,\quad \widehat{\widehat{\lambda}}_i = \tfrac{1}{T}\sum_{l=1}^{L}\sum_{t=l+1}^{T} w(l,L)\widehat{u}_{i,t-l}\widehat{u}_{it}, \tag{16}$$

$$\begin{gathered} E_{i,nT} = \sum_{t=1}^{T} \varphi_{i,t}^2 + 2\sum_{l=1}^{L}\sum_{t=l+1}^{T} w(l,L)\varphi_{i,t}\varphi_{i,t-l}, \\ \varphi_{i,t} = \widetilde{y}_{i,t-1}\widehat{\widehat{u}}_{i,t} - \overline{\varphi}_j, \overline{\varphi}_j = \tfrac{1}{\check{n}_j T}\sum_{i\in\hat{\mathcal{G}}(j)}\sum_{t=1}^{T} \widetilde{y}_{i,t-1}\widehat{\widehat{u}}_{i,t}, \\ \widehat{u}_{it} = y_{it} - \check{\rho}_j y_{i,t-1}, \widehat{\widehat{u}}_{i,t} = \widetilde{y}_{it} - \check{\rho}_j\widetilde{y}_{i,t-1},\ \text{with } i \in \hat{\mathcal{G}}(j), \\ w(l,L) = 1 - \tfrac{l}{L+1}. \end{gathered} \tag{17}$$

The quantities defined in (14) are variance, long-run variance, long-run one-sided covariance, and bias estimates, constructed from the regression residuals $\widehat{u}_{it}$ and $\widehat{\widehat{u}}_{i,t}$. 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. In order 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:

$$(18) \quad \widetilde{J}_j = \sqrt{\frac{\check{n}_j}{3}} T\left(\check{\rho}_j - 1 - \frac{\check{n}_j T \check{\lambda}_j}{\check{D}_{j,nT}} + \frac{\check{n}_j T \check{\omega}_j^2}{2\check{D}_{j,nT}}\right), \quad \widetilde{t}_j = \frac{\left(\check{\rho}_j - 1 - \frac{\check{n}_j T \check{\lambda}_j}{\check{D}_{j,nT}} + \frac{\check{n}_j T \check{\omega}_j^2}{2\check{D}_{j,nT}}\right)\check{D}_{j,nT}}{\check{\omega}_j \sqrt{\check{E}_{j,nT}}},$$

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 (17). Theorem 3 provides the limit theory for these statistics under both null and alternative hypotheses, showing test consistency.[6]

Note that there are two bias correction terms involved in these statistics, separately introduced by serial correlation and demeaned variable corrections in each test. The term $-(\check{n}_j T \check{\lambda}_j)/\check{D}_{j,nT}$ removes bias from serial correlation induced by stationary linear process errors in the autoregression; and the term $(\check{n}_j T \check{\omega}_j^2)/(2\check{D}_{j,nT})$ eliminates the bias caused by demeaning to remove fixed effects. To the best of our knowledge, the explicit forms of these bias terms are novel and are derived here for the first time in the panel context, 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 nonnegligible 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 PCA. 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 (18) 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 (Bai and Ng, 2004, 2010; Moon and Perron, 2004). Such extensions involve nontrivial 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

[6] Lemma B.1 in the Supporting Information and Lemma A10 provide supporting details for the proof of Theorem 3.

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

$$\text{IC}(G) = \ln\left(\frac{1}{nT}\sum_{j=1}^{G}\sum_{i\in\hat{\mathcal{G}}(j,G)}\sum_{t=1}^{T}(\widetilde{y}_{it} - \widetilde{y}_{i,t-1}\check{\rho}_j(G))^2\right) + \kappa_{nT}G, \tag{19}$$

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

$$\kappa_{nT} + \frac{1}{nT\kappa_{nT}} \to 0. \tag{20}$$

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

$$\widetilde{G} = \arg\min_{G=1,2,\ldots,G_{\max}} \text{IC}(G), \tag{21}$$

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

It is well known that the IC function defined in (19) 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 8 and proved in Theorem 4 below, when mildly stationary groups, a unit root group, and mildly explosive groups are all present in the panel, it turns out that $\widetilde{G} \le G^0$ with probability approaching 1, so that $\widetilde{G}$ may underestimate G^0 even in the limit.[7]

REMARK 4. It is well known in the clustering literature that the tuning parameter κ_{nT}, which penalizes the overspecification of group numbers, needs to be carefully chosen. A poor choice of κ_{nT} can lead to inconsistent estimation of group numbers, which further compromises the performance of the postclustering inference. In theory as long as $\kappa_{nT} + 1/(nT\kappa_{nT}) \to 0$, the IC estimator $\widetilde{G}$ does not overestimate the correct number of groups. But for practical work, we recommend setting $\kappa_{nT} \in [(nT)^{-0.7}, (nT)^{-0.6}]$, as guided by our simulation findings.

In order 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 $G \in \{\widetilde{G}, \widetilde{G}+1, \ldots, G_{\max}\}$. Assuming there are G groups, the recursive k-means algorithm is applied to each estimated group j for $j \in \{1, 2, \ldots, G\}$ 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 $h = 1, 2, \ldots, \overline{G}$. The notation becomes somewhat complex because of the groupings, subgroupings, and selection process. For precision, we let $\widehat{\delta}_j(G) = (\widehat{g}_{j,1}(G), \widehat{g}_{j,2}(G), \ldots, \widehat{g}_{j,\check{n}_j}(G))'$ be the estimated membership of subgroups in the estimated jth group, $\check{\boldsymbol{\rho}}_j(G) = (\check{\rho}_{j,1}(G), \check{\rho}_{j,2}(G), \ldots, \check{\rho}_{j,\overline{G}}(G))'$ 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 postclustering estimate with undemeaned variables as $\check{\rho}_j(G)$ and let

$$\check{\rho}_j(G) = \frac{\sum_{i\in\hat{\mathcal{G}}(j,G)}\sum_{t=1}^{T} y_{i,t-1}y_{it}}{\sum_{i\in\hat{\mathcal{G}}(j,G)}\sum_{t=1}^{T} y_{i,t-1}^2},\ \check{\rho}_{j,h}(G) = \frac{\sum_{i\in\hat{\mathcal{G}}(j,h,G)}\sum_{t=1}^{T} y_{i,t-1}y_{it}}{\sum_{i\in\hat{\mathcal{G}}(j,h,G)}\sum_{t=1}^{T} y_{i,t-1}^2},\ \text{and } \lim_{n,T}\frac{\check{n}_{j,h}}{\check{n}_j} \to \pi_{j,h}. \tag{22}$$

[7] The setting for $G_{\max}$ needs to be larger than the correct group number G^0. In our simulation study, we find that when the correct group number is only 2 or 3, although $G_{\max}$ is set as high as 7, $\widetilde{G}$ still underestimates G^0. This underestimation clearly comes from the IC procedure and not from the choice of $G_{\max}$. The simulations also show that $\widehat{G}$ is unaffected by the upper bound $G_{\max}$ provided it exceeds G^0.

ALGORITHM 2 RECURSIVE PROCEDURE TO COMPUTE $\widehat{G}$

(i) Optimize the IC function of equation (19) and estimate the lower bound of the group number $\widetilde{G}$ via equation (21).
(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 (26) for the estimated jth group. If the test statistic exceeds cv_{nT}, then set $G = G + 1$ and go back to Step 2. If the test statistic is smaller than cv_{nT}, then set $j = j + 1$ and re-iterate Step 2.
(v) If the null hypothesis of group-specific slope homogeneity cannot be rejected in each group $j = 1, 2, \ldots, G$ and $G < G_{\max}$, set $\widehat{G} = G$. If $G = G_{\max}$, set $\widehat{G} = G_{\max}$.

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

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

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

$$\sqrt{\check{n}_j T^{2\gamma}\Big(\rho_j^0\Big)^{2T}}\big(\check{\rho}_j(G) \cdot \ell_{\overline{G}\times 1} - \check{\boldsymbol{\rho}}_j(G)\big) \Rightarrow \mathcal{N}\Big(\mathbf{0}_{\overline{G}\times 1},\ 4\Big(c_j^0\Big)^2\Big(\widetilde{\pi}_j^{-1} - I_{\overline{G}}\Big)\Big),\ \text{if } c_j^0 > 0; \tag{25}$$

where $\widetilde{\pi}_j = \text{diag}\{\pi_{j,1}, \pi_{j,2}, \ldots, \pi_{j,\overline{G}}\}$. 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 as

$$W_j(G) := \big(\check{\rho}_j(G)\ell_{\overline{G}\times 1} - \check{\boldsymbol{\rho}}_j(G)\big)'\Big[\Big(-I_{\overline{G}} + \widetilde{\pi}_j^{-1}\Big)\check{\omega}_j^2 \check{D}_{j,nT}^{-1}\Big]^{-1}\big(\check{\rho}_j(G)\ell_{\overline{G}\times 1} - \check{\boldsymbol{\rho}}_j(G)\big)$$

$$\Rightarrow \chi^2\Big(\overline{G}\Big),\ \text{under the null hypothesis of slope homogeneity in the } j\text{th group}, \tag{26}$$

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 ($j = 1, 2, \ldots, G$) with a slowly diverging critical value of the form $cv_{nT} := (1 + b\log(nT))\chi^2_{0.95}(\overline{G})$, where $\chi^2_{0.95}(\overline{G})$ is the 95% critical value of $\chi^2(\overline{G})$ and b is some positive constant.[8] We define the new estimator of the group number as $\widehat{G}$ defined by

$$\widehat{G} = \inf_{\widetilde{G}\le G\le G_{\max}} \big\{G \mid W_j(G) \le cv_{nT},\ \text{for any } j = 1, 2, \ldots, G\big\}. \tag{27}$$

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

REMARK 5. The pivotal null distribution of the Hausman specification test relies heavily on the limit Gaussian distributions of the subgroup estimators defined in (22). Under cross-sectional independence and an increasing number of individuals for each subgroup, the limiting normal distributions of the subgroup estimators follow naturally. The condition of cross-

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

sectional independence is ensured by Assumption 1. Divergence of the dimension of the individuals in each subgroup is assured by the condition

$$\lim_{n,T} \frac{\check{n}_{j,h}}{\check{n}_j} \to \pi_{j,h},$$

for $h = 1, 2, \ldots, \overline{G}$ and $j = 1, 2, \ldots, G$. Under these conditions, the group number estimator based on IC and Hausman statistic determination is consistent for the correct group number. Extensive simulations reported in Section 5 and the Supporting Information show that this approach provides good finite-sample performance in group number estimation.

REMARK 6. The existing literature typically imposes a high-level condition to assist in determining the true number of groups—see, for example, Assumption A.4 of Su et al. (2016). This high-level condition can identify the signal of misspecification of the distancing parameter c_j through the mean squared errors of the residuals, thereby helping to determine the correct number of latent groups in most cases; but it fails to do so when there is a mildly stationary root—see Remark 8 and Lemma A11 below for details. If there are no mildly stationary individuals in the model, we can adopt this high-level condition and correctly estimate the true group number. Nonetheless, for practical purposes in bubble detection, a panel model with three individual types (mildly stationary, unit root, and mildly explosive) is more general and can be empirically more relevant than a model with only unit root and mildly explosive types, which makes a set of primitive conditions (especially Assumption 1(iv) that will be introduced in the next section) for validity more useful. The adoption of a high-level condition such as Assumption A.4 of Su et al. (2016) therefore places a constraint on the model and restricts its potential for the empirical analysis of bubble behavior in this wider context with multiple group types.

REMARK 7. Consistent estimation is feasible using only Hausman statistic determination by

$$\widehat{G}^* = \inf_{1 \le G \le G_{\max}} \left\{ G |\, W_j(G) \le cv_{nT},\ \text{for any } j = 1, 2, \ldots, G \right\}. \tag{28}$$

But while consistent, we expect that $\widehat{G}^*$ has a slower convergence rate to G^0 than $\widehat{G}$. The reason is that Hausman statistic determination employs a slowly diverging critical value cv_{nT} to remove Type I errors asymptotically. A side effect of this approach is to increase Type II errors, sacrifice discriminatory power, and reduce the convergence rate of $\widehat{G}^*$. For this reason, we retain the hybrid model specification approach of Algorithm 2.

REMARK 8. Consistency of the IC procedure (Bai and Ng, 2002; Bonhomme and Manresa, 2015; Su et al., 2016) and consistency of residual-based model specification tests (Lu and Su, 2017) both 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

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

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

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

where $\widehat{\overline{\rho}}_i^{TS} = 1 + \widehat{\overline{c}}_i^{TS}/T^{\gamma}$. Thus, when mildly stationary individual time series are missclustered 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, for example, 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.

REMARK 9. Regarding the construction of the statistic in (26), the model selection procedure and asymptotic theory are based on pooled least-squares (pooled LS hereafter) estimation in (22) instead of the panel within estimator in (13). The pooled LS estimator works because of its consistency in estimating the distance parameter c. Further, as the individual fixed effect is set to $O_p(1/T)$, the pooled LS estimator does not suffer from incidental parameter bias and converges to the correct value c^0 under joint asymptotics with a normal limit distribution. It is possible to develop a Hausman statistic approach based on the panel within-estimator in (13) or another estimator with an asymptotic normal distribution.

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 postclustering 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.

In order 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 $\mu_i = O_p(T^{-1})$.

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

$$u_{it} = \sum_{h=0}^{\infty} F_{ih}\epsilon_{i,t-h} = F_i(L)\epsilon_{it}, \tag{29}$$

in which the operator $F_i(z)(:= \sum_{h=0}^{\infty} F_{ih}z^h)$ 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 (29) are iid$(0, (\sigma^0)^2)$ over t with $(\sigma^0)^2 > 0$ 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

$$\sum_{h=0}^{\infty} h|F_{ih}| < \infty, \tag{30}$$

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 $\overline{c}_i^0 < 0$, 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* $c_j^0 \in \mathcal{C} := [-c_{up}, -c_{low}] \cup \{0\} \cup [c_{low}, c_{up}]$, $\forall j \in \mathcal{G}^0$.
(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 (30), which in turn assures functional laws hold for the partial sum processes $S_{it} = \sum_{s=1}^{t} u_{is}$ (Phillips and Solo, 1992)

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

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

REMARK 10. Bonhomme and Manresa (2015) assumed in their work that the errors were strong mixing. In our approach it is convenient to employ stationary linear process (Phillips and Solo, 1992) errors, which includes nearly all types of stationary autoregressive moving average (ARMA) processes and is a natural choice in the present linear panel dynamic setting, accommodating stationary endogenous control variables also. See Remarks 15 and 16 for further discussion.

Another modeling strategy for u_{it} that subsumes mixing and linear processes is to employ mixingale conditions and our framework accommodates this extension. Mixingales include martingale differences, ARMA processes, linear processes, and various mixing and near-epoch dependent series as special cases (Davidson, 1994) with known asymptotics and invariance principle properties (McLeish, 1977; Andrews, 1988). Recent uses of mixingale assumptions in panel models appear in Hahn et al. (2022) and Li et al. (2022). For use in the present case, we can employ a scalar L_q-mixingale array $\{u_{T,it}\} := \{u_{it}/\sqrt{T}\}$ with respect to a suitable filtration $\mathcal{F}_{it}$ satisfying the condition

$$\left\| \mathbb{E}[u_{T,it} | \mathcal{F}_{i,t-k}] \right\|_q \leq a_t \psi_k, \ \left\| u_{T,it} - \mathbb{E}[u_{T,it} | \mathcal{F}_{i,t+k}] \right\|_q \leq a_t \psi_{k+1}, \tag{31}$$

for any $k \geq 0$ and where $\|X\|_q := (\mathbb{E}\|X\|^q)^{\frac{1}{q}}$. The constants a_t and ψ_k control the magnitude and the dependence of $\{u_{it}\}$, respectively. With this approach we replace Assumption 1(ii) by the following condition.

ASSUMPTION 2LI AND LIAO (2020). *Let the error terms* $\{u_{T,it}\}$ *satisfy (31) for some* $q \geq 3$ *and for some positive sequence* a_t, $\sup_{1 \leq t \leq T} |a_t| \leq \overline{a}/T^{1/2}$ *with* $\overline{a}$ *being a constant and* $\sum_{k \geq 0} \psi_k < +\infty$. *For mildly stationary individuals,* $u_{it} = \epsilon_{it}$.

Assumption 2 allows us to approximate the partial sum of the mixingale $\{u_{T,it}\}$ using a martingale as

$$u_{T,it} = u^*_{T,it} + \widetilde{u}_{T,it} - \widetilde{u}_{T,i,t+1},$$

where $u^*_{T,it} = \sum_{s=-\infty}^{\infty}\{\mathbb{E}[u_{T,i,t+s}|\mathcal{F}_{it}] - \mathbb{E}[u_{T,i,t+s}|\mathcal{F}_{i,t-1}]\}$ forms a martingale difference sequence and the 'residual' $\widetilde{u}_{T,it}$ satisfies $\sup_{1\leq t\leq T}\|\widetilde{u}_{T,it}\|_2 = O_p(1/\sqrt{T})$. This representation allows the martingale concentration inequality (Freedman, 1975) to hold. The uniform upper bound then applies in the mixingale case and the difference between the martingale and mixingale components can be bounded by the Markov inequality. With this setup, no material change is needed for our framework to apply under mixingale errors.

REMARK 11. In the presence of heteroskedasticity no modification is needed for clustering as the convergence rates in parameter estimation stay the same. In that case, the clustering errors (CEs) remain asymptotically negligible, so that consistency of estimated group membership is retained.

Further, in the presence of heteroskedasticity over t as discussed later, our tests still have pivotal null limit distributions. In particular, postclustering estimates have limiting normal distributions under joint asymptotics; and for the unit root group the limit distributions are retained. For example, in the pure stationary AR(1) case, the standardized sample covariance $(1/\sqrt{T})\sum_{t=1}^{T}(x_{t-1}\epsilon_t)$ has asymptotic variance $\mathbb{E}[(x_{t-1}x'_{t-1})\cdot(\epsilon_t^2)]$, which cannot be factored to $\mathbb{E}(x_{t-1}x'_{t-1})\cdot\mathbb{E}(\epsilon_t^2)$, thereby playing a role in the limit theory. But under the group-specific unit root null hypothesis, partial sum processes converge to a limit Gaussian process with two-sided long-run variance. It follows that our statistics, which employ long-run variance estimators, are general enough to allow for time-series heteroskedasticity; and test statistics that allow for bias correction are robust to intertemporally correlated errors within the linear process error framework.

Our asymptotic theory does rely heavily on cross-sectional aggregation and independence across individuals. So heteroskedasticity across i brings substantial challenges. In the presence of such heterogeneity, the test statistics need to be scaled by a heteroskedasticity-robust variance estimate (White, 1980) instead of the current variance estimate. With the additional complexity, this extension would lead to a different test statistic, which involves further asymptotic analysis that is best left for future research.

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

$$\mathcal{C}_{G^0} := \mathrm{X}_1^{G^0}\mathcal{C}, \;\text{ with }\; \mathcal{C} := [-c_{up}, -c_{low}] \cup \{0\} \cup [c_{low}, c_{up}],$$

in which $c_{up} := \max_{j\in\mathcal{G}^0}|c_j^0 \cdot \mathbf{1}\{c_j^0 \neq 0\}|$ and $c_{low} := \min_{j\in\mathcal{G}^0}|c_j^0 \cdot \mathbf{1}\{c_j^0 \neq 0\}|$, with $\mathbf{1}\{\cdot\}$ being the indicator function. Assumption 1 gives another identification condition in which the group-specific parameters are well separated and ensures that the recursive k-means algorithm can correctly determine each group under the joint asymptotic scheme with both $n, T \to \infty$. The next assumption prescribes some useful rate conditions on these asymptotics.

ASSUMPTION 3.

(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:[9] $T^{\gamma \wedge (1-\gamma)} (\log_2 T)^{-2} \succ n$, $T^{3-4\gamma}(\log_2 T)^{-8} \succ n^2$, $T(\log_2 T)^{-8} \succ n(\log_2 n)^2$, $T^{2\gamma - 4/q - 4\varepsilon^*} \succ n$ *for any* $\varepsilon^* > 0$, *and* $n \succ T^{3\gamma + 2\xi - 1}(\log T)^4$ *in which the constant* ξ *satisfies* $\xi > 8c_{up}^2(1-\gamma)/(\min_{i \in \mathcal{I}_n}(\overline{\omega}_i^0)^2)$.

(iii) The lag truncation parameter in long-run variance estimations, L, satisfies the condition that $L = o(T^{\frac{1}{3}})$ *and* $L = o(T^{2\gamma})$.

Assumption 3(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 3(i), we have

$$\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} (\overline{\sigma}_{iu}^0)^2 = \lim_{n \to \infty} \sum_{j=1}^{G^0} \frac{n_j}{n} \left(\frac{1}{n_j} \sum_{i \in \mathcal{G}^0(j)} \left(\sigma_j^0\right)^2 \right) = \sum_{j=1}^{G^0} \pi_j \left(\sigma_j^0\right)^2 = \sigma_0^2 > 0, \tag{32}$$

in which the individual-specific variance $(\overline{\sigma}_{iu}^0)^2$ is defined in Equation (7). Positivity $\sigma_0^2 > 0$ is assured because all $\pi_j > 0$ by Assumption 3(i), the innovations $\{\epsilon_{it}\}$ have positive variance, and all group-specific true variances $\sigma_j^0 > 0$ by Assumption 1(ii). Assumption 3(iii) provides necessary conditions to ensure consistent estimates for the one-sided and two-sided long-run variances.

Assumption 3(ii) imposes rate restrictions on the joint divergence $(n, T) \to \infty$ that ensure the CEs are negligible and the recursive k-means clustering algorithm is uniformly consistent. The large-T conditions facilitate the application of concentration inequalities and ensure that the moment conditions for each individual are uniformly bounded. The large-n assumption guarantees that the mildly stationary individuals can be consistently allocated to their correct groups. These rate restrictions that ensure clustering consistency are all primitive conditions comparable to Assumption A.2 of Su et al. (2016) that bounds both n and T. Further, under the appropriate model framework, our rate restrictions imply the high-level conditions of clustering consistency in Bonhomme and Manresa (2015, Assumption A.2) and Okui and Wang (2021, Assumption A.3).

Assumption 3(ii) imposes rate restrictions ensuring that the first-stage parameter estimate $\widehat{c}^*$ converges at a faster rate than $O(T^{\gamma+\xi}(\log T)^2)$, so that the estimated group structure is consistent with the true latent membership for all types of nonstationary roots. Specifically, the large-T rate restrictions ensure that the tail probabilities of sample covariances are bounded. The large-n requirement $n \succ T^{3\gamma+2\xi-1}(\log T)^4$ ensures clustering consistency for all three types of nonstationary roots and assures that the upper bound of $\widehat{c}^*$ has convergence rate $O_p(1/(\sqrt{n}T^{\frac{1-\gamma}{2}}))$ faster than $O(T^{-\gamma-\xi}/(\log T)^2)$, the lower bound for clustering consistency. Assumption 3(ii) does not impose very restrictive conditions on the rate parameter γ. In practical work, the panel clustering approach should therefore have wide applicability in cases where mixed-root phenomena are present. The simulations reported later show that the recursive k-means clustering algorithm and postclustering test statistics are actually robust to γ choice. In particular, if the chosen value of γ is too large to satisfy the rate restrictions,

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

the clustering results are nearly identical to those obtained when the rate restrictions apply. This evidence suggests that the rate restrictions used to validate consistency of the clustering methodology may be stricter than necessary and supports the use of a reasonably high value of the localizing rate γ in empirical work, as recommended in the time-series studies of Kostakis et al. (2015) and employed in the empirical work and simulations of Liu et al. (2022).[10]

It is worth mentioning that, for mildly stationary groups, i.i.d. error components are assumed, which is convenient here to ensure clustering consistency for these groups and matches standard assumptions for stationary autoregressions. In mildly stationary time-series regression, serially correlated errors do affect the limit theory. In particular, Phillips and Magdalinos (2007b, Theorem 4.4) show that for mildly stationary individuals i, the time-series LS autoregressive estimator $\widehat{\overline{\rho}}_i^{TS}$ has second-order bias terms that appear in the limit theory

$$T^{\frac{1+\gamma}{2}}\left(\widehat{\overline{\rho}}_i^{TS}-\overline{\rho}_i^0-\frac{1}{T^\gamma}\frac{-2\overline{c}_i^0}{\left(\overline{\omega}_i^0\right)^2}\left(\overline{\lambda}_i^0+\frac{\overline{c}_i^0}{T^\gamma}M_{i,T}\right)\right)\Rightarrow\mathcal{N}\left(0,-2\overline{c}_i^0\right),$$

in which $M_{i,T}$ is defined in Theorem 2. When $\gamma>1/3$, the bias is simplified and the component $M_{i,T}$ is removed, leading to the modified asymptotics

$$T^{\frac{1+\gamma}{2}}\left(\widehat{\overline{\rho}}_i^{TS}-\overline{\rho}_i^0-\frac{1}{T^\gamma}\frac{-2\overline{c}_i^0\overline{\lambda}_i^0}{\left(\overline{\omega}_i^0\right)^2}\right)\Rightarrow\mathcal{N}\left(0,-2\overline{c}_i^0\right).$$

This simplification, which is achieved by $\gamma>1/3$, is convenient in conducting a correction for bias in estimation and inference. With $\gamma>1/3$ it would therefore be feasible to employ bias correction procedures within the k-means clustering algorithm, thereby allowing for serially correlated errors in mildly stationary groups. Such an extension is left for future work.

Thus, to attain clustering consistency for all group types without asymptotic bias correction, the limit theory here assumes i.i.d. errors for mildly stationary groups accompanied by conditions that include $n\succ T^{3\gamma+2\xi-1}(\log T)^4$. But if the practical focus is solely on mildly explosive and unit root groups then the i.i.d. condition can be eliminated. That approach is investigated in the Supporting Information and in such cases the size and power performance of postclustering bubble detection statistics are unaffected. Readers are referred to the Supporting Information for details.

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 1. *Suppose Assumptions 1 and 3 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 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 parameter supports are compact, the latent group membership can be consistently estimated via the recursive k-means clustering algorithm.

[10] In the simulation studies and the empirics we use the value 0.9 for the rate parameter, which has good performance in practice and is close to the value 0.95 suggested in Kostakis et al. (2015).

REMARK 12. Corrections for incidental parameters are known to cause endogeneity in dynamic panel models and disturb pivotal test procedures (Hahn and Kuersteiner, 2002). In the present framework, the bias terms generated by incidental parameter corrections still exist but they are not present in the limiting Gaussian distributions in all groups except for the unit root group. But in the unit root group the first-stage estimator $\widehat{c}_j^*$ still converges at a faster rate than $O_p(T^{\gamma+\xi}(\log T)^2)$ in the presence of incidental parameter bias, so the incidental parameter problems are no longer a concern. The disappearance of incidental parameter bias is well recognized in near-unity panel data models (Phillips, 2018). For instance, Moon and Phillips (2000) showed that incidental parameter bias terms are eliminated asymptotically under the rate condition $T/n \to \infty$. In this article, when the rate restriction $T^{1-\gamma}/n \to \infty$ holds,[11] incidental parameter bias is eliminated under joint asymptotics for the non–unit-root groups and the convergence rate of $\widehat{c}^*$ is faster than $O_p(T^{\gamma+\xi}(\log T)^2)$, guaranteeing the clustering consistency of the k-means algorithm.

As already noted, asymptotic bias from error serial correlations instead of incidental parameter bias can invalidate clustering consistency in the k-means algorithm. Under serial correlation the errors u_{it} induce asymptotic bias of order $O_p(nT)$. For the mildly explosive and unit root individuals, their group-specific signal matrices $\sum_{i\in G^0(j)}\sum_{t=1}^T \widetilde{y}_{i,t-1}^2$ are, respectively, of orders $O_p(n(\rho_j^0)^{2T}T^{2\gamma})$ and $O_p(nT^2)$. The first-stage estimates $\widehat{c}_j^*$ in these groups therefore provide faster convergence rates than $O_p(T^{\gamma+\xi}(\log T)^2)$, which is the lower bound to ensure consistent membership estimation for mildly explosive and unit root groups. But for mildly stationary groups, the first-stage estimates $\widehat{c}_j^*$ are inconsistent under serially correlated errors and i.i.d. errors $u_{it} = \epsilon_{it}$ are therefore assumed in the present work, as discussed, ensuring first-stage parameter estimation consistency and uniform consistency of the k-means clustering algorithm across all groups.

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 CEs are asymptotically negligible, it is easy to show that the postclustering 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 (12) and (13), then

$$\sqrt{n_j}\left(\rho_j^0\right)^T T^\gamma\left(\check{\rho}_j - \rho_j^0\right) = \sqrt{n_j}\left(\rho_j^0\right)^T T^\gamma\left(\widehat{\rho}_j - \rho_j^0\right) + o_p(1), \text{ if } c_j^0 > 0, \tag{33}$$

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

and

$$\begin{aligned}&\sqrt{n_j}T^{\frac{1+\gamma}{2}}\left(\check{\rho}_j - \rho_j^0 - \frac{1}{T^\gamma}\frac{-2c_j^0}{\left(\omega_j^0\right)^2}\left(\lambda_j^0 + \frac{c_j^0}{T^\gamma}\overline{m}_{j,T}\right)\right)\\ &= \sqrt{n_j}T^{\frac{1+\gamma}{2}}\left(\widehat{\rho}_j - \rho_j^0 - \frac{1}{T^\gamma}\frac{-2c_j^0}{\left(\omega_j^0\right)^2}\left(\lambda_j^0 + \frac{c_j^0}{T^\gamma}\overline{m}_{j,T}\right)\right) + o_p(1), \text{ if } c_j^0 < 0,\end{aligned} \tag{35}$$

where $\overline{m}_{j,T}$ in (35) denotes a nonnegligible bias element whose explicit form is given by (39) in Theorem 2. Based on (33), (34), and (35), 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 postclustering estimator $\check{\rho}_j$ of the jth estimated group.

[11] The rate restriction in Lemma A10 ensures that the demeaning terms arising from incidental parameter adjustments in mildly stationary and mildly explosive groups are asymptotically negligible.

THEOREM 2. *Suppose Assumptions 1 and 3 hold. When* $(n, T) \to \infty$,

$$\sqrt{n_j}\left(\rho_j^0\right)^T T^{\gamma}\left(\check{\rho}_j - \rho_j^0\right) \Rightarrow \mathcal{N}\left(0, 4\left(c_j^0\right)^2\right),\ \mathit{if}\, c_j^0 > 0; \tag{36}$$

$$\sqrt{n_j}T\left(\check{\rho}_j - 1 + \frac{3\left(\sigma_j^0\right)^2}{\left(\omega_j^0\right)^2}\frac{1}{T}\right) \Rightarrow \mathcal{N}(0, 3),\ \mathit{if}\, c_j^0 = 0; \tag{37}$$

$$\sqrt{n_j}T^{\frac{1+\gamma}{2}}\left(\check{\rho}_j - \rho_j^0 - \frac{1}{T^{\gamma}}\frac{-2c_j^0}{\left(\omega_j^0\right)^2}\left(\lambda_j^0 + \frac{c_j^0}{T^{\gamma}}\overline{m}_{j,T}\right)\right) \Rightarrow \mathcal{N}\left(0, -2c_j^0\right),\ \mathit{if}\, c_j^0 < 0, \tag{38}$$

where

$$\overline{m}_{j,T} = \tfrac{1}{n_j}\textstyle\sum_{i\in\mathcal{G}^0(j)} m_{i,T} \ \mathit{and}\ m_{i,T} = \sum_{h=1}^{\infty}\left(\rho_j^0\right)^{h-1}\mathbb{E}(\widetilde{\epsilon}_{it}u_{i,t-h}). \tag{39}$$

REMARK 13. Theorem 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 postclustering 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, DF, 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 14. The convergence rates of $\check{\rho}_j$ are $\sqrt{n_j}(\rho_j^0)^T T^{\gamma}$, $\sqrt{n_j}T$, and $\sqrt{n_j}T^{\frac{1+\gamma}{2}}$ for the respective cases $c_j^0 > 0$, $c_j^0 = 0$, and $c_j^0 < 0$. 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.

REMARK 15. Additional control variables have been used in the dynamic panel literature. For example, Moon and Phillips (2004) and Norkutė and Westerlund (2021) modeled the near-unit-root panel model by including exogenous variables as in the following system:

$$z_{i,t} = \beta' x_{i,t} + y_{i,t},\ \ y_{i,t} = \alpha_i + \rho_{g_i}y_{i,t-1} + u_{i,t}\ \text{ with } \rho_{g_i} = 1 + \frac{c_{g_i}}{T^{\gamma}}, \tag{40}$$

where the z_{it} are asset prices and x_{it} contains d-dimensional market fundamentals. Following the quasi-maximum likelihood estimator (QMLE) approach of Norkutė and Westerlund (2021) or the PSY-IVX approach of Shi and Phillips (2021), we can concentrate out x_{it} and obtain the residuals $\{\widehat{y}_{it}\}_{i=1,t=1}^{n,T}$. The proposed two-stage procedure can then be applied to $\widehat{y}_{it}$ without further changes. In the empirical work of this article, market fundamentals are concentrated out and our method is applied to (4). A further extension that incorporates multidimensional clusterings as in Cheng et al. (2019) and Leng et al. (2023) to allow both β and ρ to differ across groups may be potentially useful but is not considered in the present work.

REMARK 16. Predetermined variables have also been used in the dynamic panel literature. For example, when testing for the presence of unit roots in a panel setting, Chang (2002) included predetermined variables. When $\alpha_i = 0$ in (40), the mixed-root panel model can be decomposed as

$$(41) \qquad (1-\rho_{g_i}L)y_{it} = \left(\sum_{h=0}^{\infty} F_{ih}L^h\right)\epsilon_{it} = F_i(L)\epsilon_{it} = F_{i,1}(L)\cdot F_{i,2}(L)\epsilon_{it},$$

where some factorization $F_i(L) = F_{i,1}(L)F_{i,2}(L)$ applies. Then the model can be transformed as

$$(42) \qquad (1-\rho_{g_i}L)F_{i,1}(L)^{-1}y_{it} = F_{i,2}(L)\epsilon_{it},$$

and lagged variables generated as endogenous controls. Clearly, by allowing u_{it} to be a general stationary linear process, (4) characterizes the impact of such endogenous controls in a different manner.

REMARK 17. Theorem 2 gives closed-form expressions of the second-order asymptotic biases that can arise in our panel model framework and connects to earlier work on endogeneity issues in the panel unit root literature. For instance, the traditional model setup (Chang and Song, 2009) includes endogenous regressors to eliminate omitted variable bias and is essentially equivalent to estimating the following augmented model:

$$(43) \qquad (1-\rho_{g_i}L)F_i(L)^{-1}y_{it} = \epsilon_{it},$$

by ordinary least squares (OLS). Or, explicit forms of asymptotic bias as in the general framework of Phillips and Magdalinos (2007b) can be employed with bias correction procedures. In our model framework, the focus is on distinguishing explosive, unit root, and mildly integrated groups, so the regression approach is much simpler and better suited to practical situations where it is useful to distinguish such groups in empirical research.

4.2. *Testing for Explosive Roots.* THEOREM 3. *Suppose Assumptions 1 and 3 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 $\widetilde{t}_j \asymp ((\rho_j^0)^T\sqrt{n})$, and $\widetilde{J}_j \asymp (\sqrt{n}T^{1-\gamma})$.*

REMARK 18. In the explosive root groups, where $c_j^0 > 0$, the sample moment

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

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

$$(44) \qquad \sqrt{n_j}\left(\rho_j^0\right)^T T^\gamma\left(\check{\rho}_j - \rho_j^0 + \frac{n_jT\left(\sigma_j^0\right)^2}{2\sum_{i\in\mathcal{G}^0(j)}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2}\right) \Rightarrow \mathcal{N}\left(0, 4\left(c_j^0\right)^2\right), \;\; \text{if } c_j^0 > 0;$$

$$(45) \qquad \sqrt{n_j}T\left(\check{\rho}_j - \rho_j^0 + \frac{n_jT\left(\sigma_j^0\right)^2}{2\sum_{i\in\mathcal{G}^0(j)}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2}\right) \Rightarrow \mathcal{N}(0, 3), \;\; \text{if } c_j^0 = 0.$$

By Theorem 2 and Lemma A8, 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

$$\sqrt{n_j}\left(\rho_j^0\right)^T T^\gamma\left(\check{\rho}_j-\rho_j^0+\frac{\check{n}_j T\check{\sigma}_j^2}{2\sum_{i\in\hat{\mathcal{G}}(j)}\sum_{t=1}^T\widetilde{y}_{i,t-1}^2}\right)\Rightarrow\mathcal{N}\left(0,4\left(c_j^0\right)^2\right),\ \text{ if } c_j^0>0, \tag{46}$$

$$\sqrt{n_j}T\left(\check{\rho}_j-\rho_j^0+\frac{\check{n}_j T\check{\sigma}_j^2}{2\sum_{i\in\hat{\mathcal{G}}(j)}\sum_{t=1}^T\widetilde{y}_{i,t-1}^2}\right)\Rightarrow\mathcal{N}(0,3),\ \text{ if } c_j^0=0. \tag{47}$$

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

REMARK 19. In comparison with statistics based on time-series data, the t-statistic under the alternative hypothesis of an explosive root diverges at the rate $O((\rho_j^0)^T)$, slower than the $O((\rho_j^0)^T\sqrt{n})$ rate of Theorem 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 $\widetilde{t}_j$ has a divergence rate that is faster by an exponential factor than the coefficient-based test for which $\widetilde{J}_j=O_p(\sqrt{n}T^{1-\gamma})$. 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 result shows that the IC estimator $\widetilde{G}$ is inconsistent when there are mildly stationary groups in the data.

THEOREM 4. *Suppose Assumptions 1 and 3 hold and $(n,T)\to\infty$. When either (i) $\gamma\in(0,1)$ and $\overline{c}_i^0\geq 0$ 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 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 5. *Let Assumptions 1 and 3 hold and $\gamma\in(0,1)$. Then, $\widehat{G}\to_p G^0$, as $(n,T)\to\infty$.*

REMARK 20. 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 misclustering 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 $\widehat{\rho}^{MG}=(1/n)\sum_{i=1}^n\widehat{\rho}_i^{TS}$ and $\widehat{\rho}_i^{TS}$ 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 $\text{Avar}(\widehat{\rho}^{FE})=\text{Avar}(\widehat{\rho}^{MG})$ when mildly stationary groups exist, so the Rao-Blackwell theorem is not applicable (Pesaran and Yamagata, 2008).

REMARK 21. 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 5 shows that the estimator $\widehat{G}$ consistently estimates G^0. Following common practice in the panel clustering literature (Bonhomme and Manresa, 2015; Huang et al., 2021; Su et al., 2016), we treat the consistent estimate $\widehat{G}$ as the true value G^0 in the practical work of implementation.[12]

5. SIMULATION STUDIES

Several numerical experiments were designed to check the finite sample performance of the procedures developed above. These include the following: the group number estimate in (27); the membership estimate generated by the recursive k-means clustering algorithm in (8); the postclustering estimates in (13); and the size and power performances of the proposed tests in (18).

The following DGP was used for the simulations: individual fixed effects $\mu_i \overset{i.i.d.}{\sim} T^{-1}\mathcal{N}(0,1)$; errors $u_{it} = \theta u_{i,t-1} + \epsilon_{it}$ with either (i) serially correlated errors ($\theta = 0.5$, $\epsilon_{it} \overset{i.i.d.}{\sim} \mathcal{N}(0, 0.1)$) or (ii) $i.i.d.$ errors ($\theta = 0$, $\epsilon_{it} \overset{i.i.d.}{\sim} \mathcal{N}(0, 0.1)$)[13]; sample sizes $n = 32$, 48, 64, 72, 96, 120, 144, 150, 192, and $T = 150$, 200, 250, 400, 600, 800; group number $G^0 = 3$ (i.e., three groups) with $\pi_1 : \pi_2 : \pi_3 = (1/3) : (1/3) : (1/3)$ or $G^0 = 2$ (i.e., two groups) with $\pi_1 : \pi_2 = (1/2) : (1/2)$. The parameter settings for c and γ were as follows:

$$(48) \qquad \left(c_1^0, c_2^0, c_3^0, \gamma\right) = \begin{cases} (-1, \ -5, \ -10, 0.4) & \text{for DGP 0} \\ (0.5, 0.04, -0.06, 0.1) & \text{for DGP 1} \end{cases},$$

and

$$(49) \qquad \left(c_1^0, c_2^0, \gamma\right) = \begin{cases} (0.5, 0.04, 0.1) & \text{for DGP 2} \\ (0.5, \quad 0, 0.1) & \text{for DGP 3} \\ (0.5, 0.01, 0.1) & \text{for DGP 4} \end{cases},$$

in which only a small true value $\gamma = 0.1$ was used, while checking the robustness of the clustering accuracy with a misspecified $\gamma^* > \gamma$. Simulation results with a larger true value (e.g., $\gamma > 0.5$) are given in Liu et al. (2022) to which readers are referred.

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. In DGPs 1–4, the AR(1) parameter values are close to the empirical estimates. All these DGPs have a mixed collection of groups and 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. In order to explore the advantages of the postclustering panel tests, we draw comparisons with standard semiparametric time-series test statistics, namely, PP t and J tests (Phillips, 1987; Phillips and Perron, 1988), which follow nonstandard limit distributions[14] under the null $\mathcal{H}_0 : \overline{c}_i^0 = 0$.

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 (15) and (16) was set at $L = \lfloor T^{0.3} \rfloor$, the bandwidth for the variance estimate in (17) was set to $L = \lfloor T^{0.1} \rfloor$, and for the PP

[12] 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 postselection inference difficulties particularly in small samples. However, existing results in the literature (Liu et al., 2020) show that overidentification of the group number and resulting minor classification errors do not damage the consistency of the postclassification estimators. Extensions of such findings to nonstationary panel models like those of this article are a topic for further research.

[13] The Supporting Information gives results for $i.i.d.$ errors and signal-to-noise ratios for each group in DGPs 0–2.

[14] Explicit forms and null limit distributions of the t and J tests are given in equations (43) and (44) of Liu et al. (2022).

TABLE 2
EMPIRICAL FREQUENCY OF MODEL SELECTION UNDER DGP 0 ($\theta = 0.5$)

		IC Estimator $\widetilde{G}$							Proposed Estimator $\widehat{G}$						
n	T	1	2	3	4	5	6	7	1	2	3	4	5	6	7
120	200	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000
120	400	0.000	0.999	0.001	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000
120	600	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000
120	800	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000
150	200	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.999	0.000	0.000	0.000	0.001
150	400	0.000	0.947	0.053	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000
150	600	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000
150	800	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000

time-series statistics the bandwidth for the long-run variance and covariance components was set to $\lfloor T^{0.3} \rfloor$.[15] 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 tuning parameter κ_{nT} in the IC penalty is $(nT)^{-0.6}$ or $(nT)^{-0.7}$ and the upper bound $G_{\max}$ is 7.[16] Although IC has a downward bias, it does help to raise the convergence rate of the procedure and provides some improvement in finite-sample performance. The penalty can be set in a flexible way (e.g., in a domain such as $\kappa_{nT} \in [(nT)^{-0.7}, (nT)^{-0.6}]$).[17] The critical value of the Hausman test is set as $cv_{nT} = (1 + 5\log(nT))\chi^2_{0.95}(\overline{G})$ and $\overline{G} = (G_{\max} - G + 1)$. Table 2 reports the empirical frequency of the IC and proposed estimators of G (i.e., $\widetilde{G}$ in (21) and $\widehat{G}$ in (27)) for DGP 0. As T increases, the performance of $\widehat{G}$ steadily improves, so that when T is larger than 200 $\widehat{G}$ 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. The Supporting Information reports the performance of $\widetilde{G}$ and $\widehat{G}$ for other DGPs.

Next, finite sample performance of the recursive k-means clustering algorithm and postclustering estimation was checked, assuming the true group number G^0 was known. Table 3 reports the CE, root mean squared error (RMSE), and bias of the postclustering estimates of DGP 1 and results for other DGPs are given in the Supporting Information. The CE is defined as

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

The RMSE is the square root of the sample moment of the squared differences between the postclustering estimates and the true values. The bias is based on the averaged differences between the postclustering 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.

[15] Bandwidths are selected based on the simulation findings in the mixed-root panel model. When the bandwidths of (15) and (16) are smaller than $\lfloor T^{0.3} \rfloor$, the panel t-test statistic overrejects and leads to size distortion. When the bandwidths used in (17) 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.

[16] The generic upper bound does not impact the performance of either IC or Hausman-type statistics once it is set to a large enough value.

[17] The main article uses $(nT)^{-0.7}$ and results with the setting $(nT)^{-0.6}$ appear in the Supporting Information.

TABLE 3

CLUSTERING AND ESTIMATION BY TWO STAGE PROCEDURE UNDER DGP 1 AND CORRECT γ^* VALUE ($\gamma^* = 0.1$, WITH $\theta = 0.5$)

			Group 1				Group 2				Group 3			
			Clustering		Oracle		Clustering		Oracle		Clustering		Oracle	
n	T	CE	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias
48	150	0.029	0.000	0.000	0.000	0.000	0.002	−0.002	0.002	−0.002	0.034	0.033	0.033	0.033
48	200	0.011	0.000	0.000	0.000	0.000	0.001	−0.001	0.001	−0.001	0.035	0.035	0.035	0.035
48	250	0.004	0.000	0.000	0.000	0.000	0.000	−0.000	0.000	−0.000	0.036	0.036	0.036	0.036
72	150	0.029	0.000	0.000	0.000	0.000	0.002	−0.002	0.002	−0.002	0.034	0.033	0.033	0.033
72	200	0.011	0.000	0.000	0.000	0.000	0.001	−0.001	0.001	−0.001	0.035	0.035	0.035	0.035
72	250	0.004	0.000	0.000	0.000	0.000	0.000	−0.000	0.000	−0.000	0.036	0.036	0.036	0.036
96	150	0.029	0.000	0.000	0.000	0.000	0.002	−0.002	0.002	−0.002	0.034	0.033	0.033	0.033
96	200	0.011	0.000	0.000	0.000	0.000	0.001	−0.001	0.001	−0.001	0.035	0.035	0.035	0.035
96	250	0.004	0.000	0.000	0.000	0.000	0.000	−0.000	0.000	−0.000	0.036	0.036	0.036	0.036

According to Table 3, the CE decreases to zero as T increases. The RMSE and bias of the oracle estimates are smaller than those of the postclustering estimates. For the postclustering estimates of the nonstationary groups, the magnitude of the RMSE and bias also generally decreases when $T \to \infty$. For DGP 1, the difference between the oracle and the postclustering estimates is negligible when $T \geq 200$. 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. Further, clustering results are explored when the employed value of the rate parameter γ^* exceeds the true generating value $\gamma = 0.1$. Table 8 of the Supporting Information shows that the estimated membership based on the use of γ^* larger than $\gamma = 0.1$ is identical to membership estimation based on the value $\gamma^* = 0.1$, again corroborating the use of a larger rate parameter. Further extensive simulations were conducted with various values of γ^*,[18] all showing that use of larger values of γ^* are robust in recovering correct group membership. This robustness of clustering performance to γ^* raises the practical question of a suitable choice of this parameter in practice. Based on our simulation results, we recommend the use of $\gamma^* = 0.9$ in practice and this value was employed in our empirical work.

An interesting feature of Table 3 is that the RMSE and bias are almost zero for the explosive group with $c_1^0 = 0.5$. This outcome is due to the fact that for a mildly explosive process the estimation errors decay at an exponential rate that is influenced by the true distance parameter. In DGP 1, since $c_1^0 = 0.5$ is large, the estimation of the distancing parameter is highly accurate and the resulting RMSE and bias are nearly zero.

Based on the estimated membership $\widehat{\delta}$, the performance of the postclustering panel t and J tests for detecting explosive roots was 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 postclustering panel t and J tests when $n > 1$, which are presented in Table 4. 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.

Evidently the size distortion of both panel tests is small when $n \geq 48$ 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 the small samples that arise in group subsamples. This loss is counterbalanced by a substantial improvement

[18] Some of these simulation results are reported in the Supporting Information and show robustness of the k-means clustering algorithm to an overspecified rate parameter $\gamma^*(> \gamma)$.

TABLE 4
POWER OF TESTS FOR DETECTING EXPLOSIVENESS ($\theta = 0.5$)

n	T	DGP 3				DGP 4			
		$c_1 = 0.5$		$c_2 = 0$		$c_1 = 0.5$		$c_2 = 0.01$	
		t-Test	J-Test	t-Test	J-Test	t-Test	J-Test	t-Test	J-Test
1	150	1.000	1.000	0.064	0.066	1.000	1.000	0.156	0.155
1	200	1.000	1.000	0.069	0.069	1.000	1.000	0.203	0.204
1	250	1.000	1.000	0.054	0.055	1.000	1.000	0.286	0.291
32	150	1.000	1.000	0.063	0.070	1.000	1.000	0.326	0.166
32	200	1.000	1.000	0.049	0.051	1.000	1.000	0.564	0.278
32	250	1.000	1.000	0.056	0.052	1.000	1.000	0.888	0.611
48	150	1.000	1.000	0.053	0.058	1.000	1.000	0.418	0.238
48	200	1.000	1.000	0.049	0.054	1.000	1.000	0.703	0.425
48	250	1.000	1.000	0.056	0.063	1.000	1.000	0.958	0.828
64	150	1.000	1.000	0.063	0.061	1.000	1.000	0.477	0.295
64	200	1.000	1.000	0.041	0.049	1.000	1.000	0.797	0.562
64	250	1.000	1.000	0.058	0.057	1.000	1.000	0.985	0.914

in the power of the panel tests over the time-series tests. For instance, when $c_j^0 = 0.01$ (the corresponding ρ_j^0 is 1.0061, 1.0059, and 1.0058 when $T = 150, 200, 250$, which are empirically plausible based on the later empirical outcomes), the power performances of the postclustering panel tests are much larger than those of the time-series tests. If $T = 250$, the postclustering panel t-test raises the power of the time series t-test from 0.286 to 0.888 when $n_j = 16$, to 0.958 when $n_j = 24$, and to 0.985 when $n_j = 32$. The postclustering panel t test with $T = 200$ has substantially greater power than the time-series t test with $T = 250$ (0.797 versus 0.291). Moreover, 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 3 under the mildly explosive alternative.

6. EMPIRICAL APPLICATIONS

Many empirical studies have investigated the existence of financial asset bubbles in real estate (Giglio et al., 2016; Phillips, 2023; Phillips and Yu, 2011) and equity markets (Diba and Grossman, 1988; Phillips et al., 2015a). Most of these have used time-series data. This section applies our methodology to housing markets in China and the United States and the U.S. equity market.

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 35 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 article to a panel of monthly housing indices from 83 ($n = 83$) cities in China obtained from Fang et al. (2016). The sample period is from January 2003 to June 2012 and contains 114 monthly observations ($T = 114$). 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.[19] For this appli-

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

TABLE 5
SELECTION BY IC OF THE NUMBER OF GROUPS G IN THREE EMPIRICAL STUDIES

G	1	2	3	4	5	6	7
IC for Chinese housing	−14.8514	**−14.8595**	−14.8586	−14.8583	−14.8564	−14.8560	−14.8545
IC for U.S. housing	−11.1695	**−11.1914**	−11.1914	−11.1873	−11.1810	−11.1732	−11.1667
IC for U.S. equty	−2.9545	**−2.9639**	−2.9631	−2.9615	−2.9588	−2.9562	−2.9531

The smallest IF value is in boldface.

TABLE 6
POSTCLUSTERING ESTIMATES AND PANEL t- AND J-TESTS IN THREE EMPIRICAL STUDIES

	Chinese Housing		U.S. Housing		U.S. Equity	
	Group 1	Group 2	Group 1	Group 2	Group 1	Group 2
$\check{n}_j$	40	43	7	4	40	106
$\check{\rho}_j$	1.011	0.996	1.0432	1.0113	1.012	0.967
t-test	7.964***	0.383	31.9880***	1.8617*	2.236**	0.141
J-test	5.608***	0.313	35.5062***	2.3412**	8.908***	1.330

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

cation, 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 price dynamics in some cities and commonalities that exist in supply and demand factors, leading to the coexistence of possible groupings of cities into mildly explosive groups, a unit root efficient market group, and mildly stationary groups.

With tuning parameter $\kappa_{nT} = (nT)^{-0.7}$,[20] the first row of Table 5 reports the values of the computed ICs for $G = 1, \ldots, 7$. According to the IC selection $\widetilde{G} = 2$. The Hausman-type test algorithm is applied using the critical value $cv_{nT} = (1 + 5\log(nT))\chi^2_{0.95}(\overline{G})$ and this procedure leads to the same estimate $\widehat{G} = 2$. The recursive k-means clustering algorithm is then implemented based on (8), giving the postclustering estimate (13). This two-stage procedure provides the clustered group structure.[21,22] There are 40 cities in Group 1 including three tier 1 cities, Beijing, Guangzhou, and Shenzhen. Comparatively, there are 43 cities in Group 2. Figure 2 gives the time-series plots for both groups.

For each identified group, we report the panel-within estimate $\check{\rho}_j$, the number of cities in each estimated group, the postclustering t- and J-statistics for the null hypothesis of the group-specific unit root, as in the first panel of Table 6. For Group 1, the estimated $\check{\rho}_1$ is 1.011. Both the panel t-test and the panel J-test suggest that c_1^0 is significantly larger than zero at the 1% significance level. For Group 2, the postclustering estimate $\check{\rho}_2$ is just below unity and the unit root null hypothesis is not rejected in this group. The postclustering t and J tests therefore indicate explosive behavior in Group 1 but unit root behavior in Group 2.

The recursive k-means clustering procedure evidently finds a two-cluster structure in the China real estate market: one fast-growing group and one relatively slow-growing cluster, in

[20] Here and in the following empirical applications the IC penalty tuning parameter κ_{nT} was chosen to satisfy the rate condition (20). The IC 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 $\kappa_{nT} = (nT)^{-0.7}$ or $\kappa_{nT} = (nT)^{-0.6}$, membership estimation, postclustering parameter estimation and inferences concerning explosive behavior are unchanged in all the applications of the article.

[21] The initial values used to start the k-means algorithm were chosen as the 0.3 and 0.8 quantiles of the individual time-series estimates. The rate parameter $\gamma^* = 0.9$ was used. The estimated groups were found to be robust to various choices of the initial values and the rate parameter.

[22] The names of the cities in each estimated group are reported in the Supporting Information.

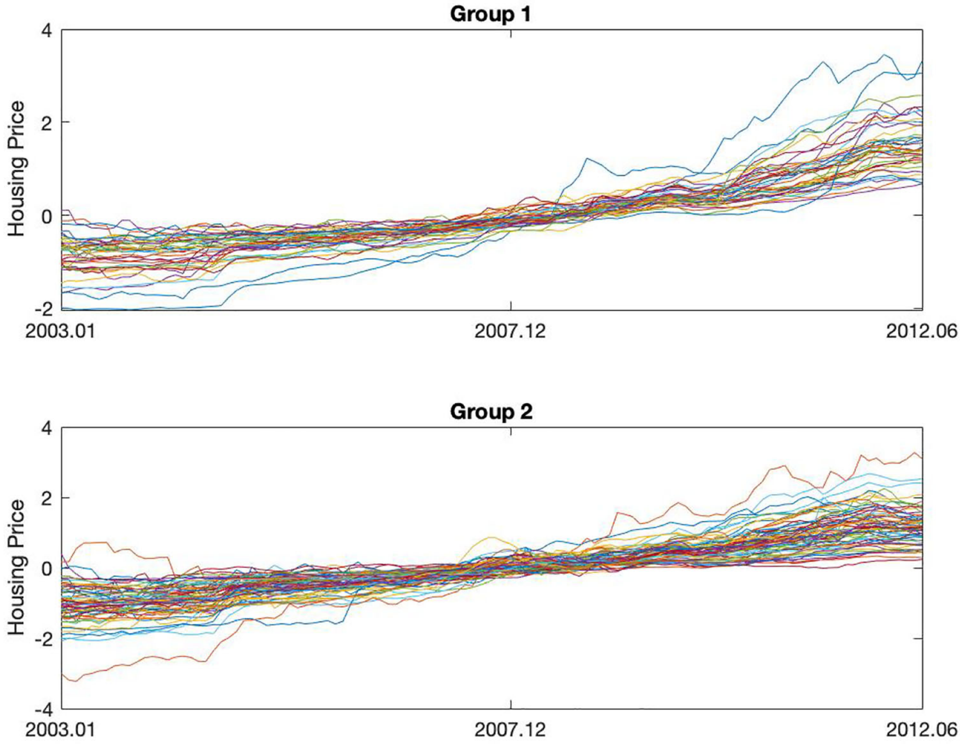


FIGURE 2

TIMES-SERIES PLOTS OF THE DEMEANED RATIO OF CHINESE CITY HOUSE PRICE INDICES TO THE RENTAL CPI (VERTICAL AXIS) FOR THE ESTIMATED GROUPS

which the fast-growing one shows strong signals of explosive roots, supported by the postclustering statistics. This two-cluster estimated membership thereby demonstrates the feature of real estate market segmentation that has been long investigated in the literature. For instance, it is well acknowledged that housing submarkets can be due to such features as spatial attributes, policy-related issues, and population expansion (Goodman, 1978; Rosen, 1974).

Fang et al. (2016) showed that house price appreciation of the first-tier cities in China (such as Beijing, Guangzhou, and Shenzhen) is significant relative to increases in household disposable income, whereas house price growth in second-tier and third-tier cities is slower. For instance, price appreciation had an average annual real growth rate larger than 7.9% throughout the entire country between 2003 and 2013, which included the much higher average growth rate of 13.1% in the first-tier cities. The housing price growth was accompanied by growth in household disposable income, which includes an average annual real growth rate of about 9.0% at the national level and a lower average growth rate of 6.6% in the first-tier cities. Another reason for this divergence may relate to demographics and the urbanization process. In particular, the population of the four first-tier cities grew from 48 million in 2004 to almost 70 million in 2012, whereas the population of the second-tier and third-tier cities experienced a much smaller increase: The total population of the second-tier cities, for instance, grew from 220 million in 2004 to around 260 million in 2012.

These clustering results support a general finding of group divergence: The estimated group with the largest explosive root, including tier-one and several tier-two cities, reveals a clear grouping phenomenon, with the remaining cities classified into a single group where house price appreciation is relatively lower than that of household disposable income. Notably, the clustering approach corroborates the distinct high house price appreciation phenomenon generally acknowledged among tier 1 and certain tier 2 cities by direct application of bubble detection statistical tests.

6.2. *Housing Prices in the United States.* In recent years, strong surges in house prices have occurred in many U.S. cities. Possible reasons for these surges include near-zero interest rates and rising inflation expectations. In order to examine whether rising fundamental values justify these developments, the two-stage algorithm of this article was applied to a panel of monthly housing indices for 11 ($n = 11$) U.S. cities obtained from the official website of the Federal Reserve Bank of St. Louis.[23] Monthly observations of $T = 105$ time series for each series were used, covering the period from January 2013 to September 2021. In order to measure fundamental values monthly city-specific CPI data for rentals was employed.[24] 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.

Using the tuning parameter setting $\kappa_{nT} = (nT)^{-0.7}$, the second row of Table 5 reports calculated values of the ICs for $G = 1, \ldots, 7$, leading to the choice $\widetilde{G} = 2$. The combined IC-Hausman test procedure with critical value $cv_{nT} = (1 + 5\log(nT))\chi^2_{0.95}(\overline{G})$ gave the same estimated value $\widehat{G} = 2$. Using this estimated group number the clustering algorithm based on (8) was implemented, giving postclustering estimates from (13) and the corresponding clustered group structure.[25,26] The two-stage procedure produced seven cities in Group 1 and four cities in Group 2. Time-series plots of these two groups are provided in Figure 3.

For each identified group of cities in the U.S. housing market, the second panel of Table 6 reports the panel within-group estimates $\check{\rho}_j$, the number of cities in each estimated group, and the postclustering t- and J-statistics for the null hypothesis of the group-specific unit root. According to both the postclustering t- and J-statistics, the explosive root $\check{\rho}_1 = 1.0432$ of Group 1 is statistically significant at the 1% level, indicating the presence of a housing bubble in this group. For Group 2, the postclustering estimate $\check{\rho}_2 = 1.0113$ also exceeds unity and the unit root group null is rejected at the 5% level. These postclustering panel tests suggest that explosive price bubbles are present in the data for both Groups 1 and 2, although Group 2 has four 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 1(b) shows the time series of U.S. house prices[27] 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.

Our recursive k-means clustering procedure finds a two-cluster structure also in the U.S. real estate market. A fast-growing group and a relatively slow-growing cluster are detected. Both groups show strong signals of explosive roots, which are supported by the postclustering statistical tests. This group estimation outcome accords with earlier evidence in Abraham et al. (1994), where a k-means clustering algorithm was applied to identify structural relationships in the U.S. housing market and a bootstrap test procedure was used to test for statistical significance of a three-group structure. Abraham et al. (1994) employed a sample of 30 U.S. cities between 1977 and 1992, whereas our data are for only 11 U.S. cities, so that a partial grouping structure of that in Abraham et al. (1994) is to be expected.

The regional structure of the United States has a tremendous impact on direct and indirect investment in the housing market (Abraham et al., 1994; Bourassa et al., 1999; Goetzmann

[23] https://fred.stlouisfed.org/

[24] For cities whose city-specific CPI for rentals data were unavailable, fundamental values were approximated using the national CPI for rentals.

[25] The initial values were obtained from the 0.3 and 0.8 quantiles of the individual time-series estimates and the localizing rate parameter was set to $\gamma^* = 0.9$. The empirical results were found to be robust to various initial values and localizing rate parameters.

[26] The names of the U.S. cities in each estimated group are reported in the Supporting Information.

[27] Specifically, the plotted data in Figure 1(b) are the ratio of the city price index to the CPI of the city rentals

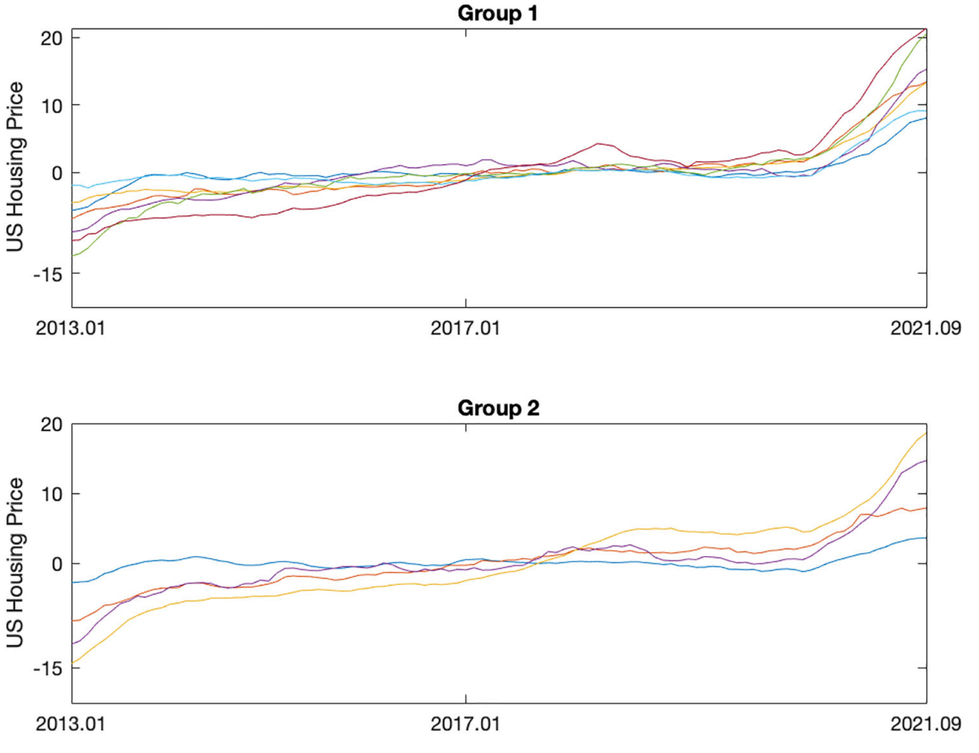


FIGURE 3

TIME-SERIES PLOTS OF THE DEMEANED RATIO OF THE HOUSING PRICE TO THE RENTAL CPI (VERTICAL AXIS) IN THE TWO ESTIMATED GROUPS OF THE U.S. HOUSING MARKET

and Wachter, 1995) with systematic differences across cities that reduce portfolio risk in real estate. In general, the greater the topographical distance there is between cities, the higher is the level of diversification, with fewer common factors driving house prices being shared across cities. Topographical distance also tends to produce market division into submarkets in which cities may share certain common factors (such as environmental quality, living facilities, crime incidence, or extremities of weather) that can serve as proxies for one or the other (Bourassa et al., 1999; Goodman and Thibodeau, 2003). These levels of aggregation may be reasonably approximated in the econometric modeling of latent membership used in this article. Estimated groups identified here might then be treated as submarkets in which similar features are shared.

Abraham et al. (1994) applied the k-means algorithm and found three groups: West Coast, Middle, and East Coast. The West Coast group contains the Bay Area and LA, the Middle group includes cities of the Great Lakes Region, and the East Coast group has the regions on the east coast. Their detected group structure is similar to ours, but our estimated group structure aggregates the Middle and West Coast groups into one estimated group. Based on our estimated group structure, postclustering test statistic outcomes indicate explosive bubble signals for both groups and support a broad finding of a housing price bubble. These results have implications for regional economic assessments of housing price trends and for specialist financial services requiring more detailed analysis and accuracy in real estate asset pricing.

6.3. *Equity Prices in the United States.* In a further application, the methodology is applied to a panel of equity prices in the U.S. 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 fea-

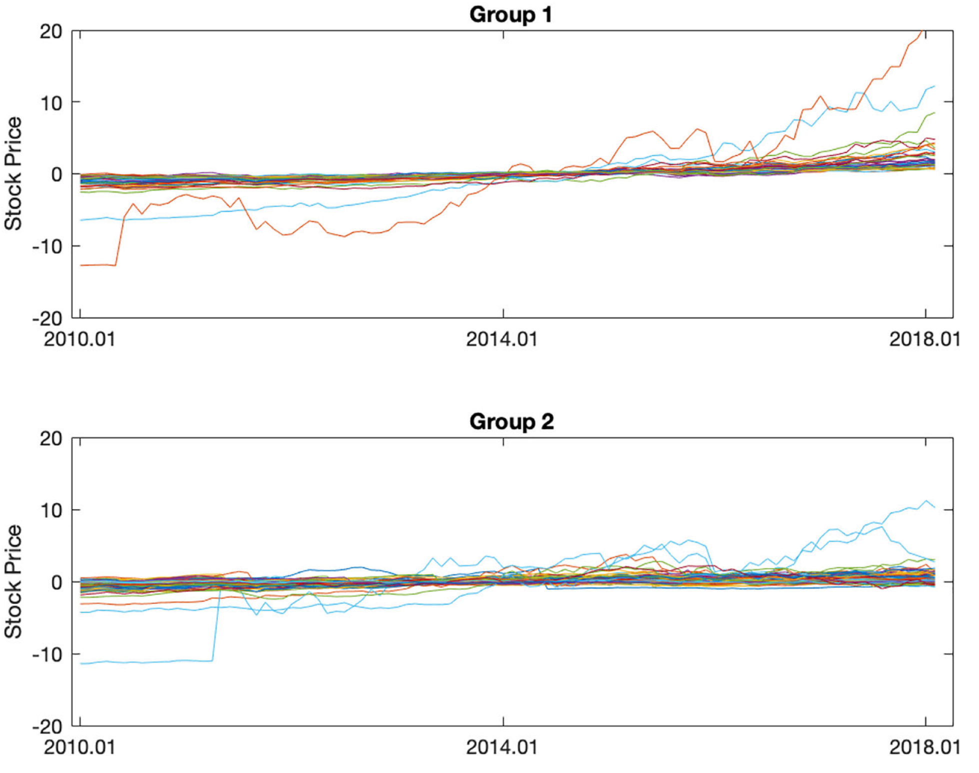


FIGURE 4

TIME-SERIES PLOTS OF THE DEMEANED DIFFERENCE BETWEEN S&P 500 STOCK PRICES AND DIVIDENDS IN THE TWO ESTIMATED GROUPS

tures. 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.

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

Using the penalty parameter $\kappa_{nT} = (nT)^{-0.6}$ the third row of Table 5 reports IC values for $G = 1, \ldots, 7$, leading to the estimate $\widetilde{G} = 2$. The combined IC–Hausman test procedure with critical value $cv_{nT} = (1 + 5\log(nT))\chi^2_{0.95}(\overline{G})$ produced the same estimate $\widehat{G} = 2$. The clustering algorithm was implemented using (8) and postclustering estimates from (13). The two-stage procedure provided the group structure.[28,29] 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 4.

[28] The initial values were obtained from the 0.3 and 0.8 quantiles of the individual time series estimates and the localizing rate parameter was set to $\gamma^* = 0.9$. The empirical results were found to be robust to various initial values and localizing rate parameters.

[29] The individual stock symbols in each estimated group are reported in the Supporting Information.

For each identified group, the third panel of Table 6 reports the panel within-group estimates $\check{\rho}_j$, the number of stocks in each estimated group, and the postclustering t- and J-statistics for the null hypothesis of the group-specific unit root. According to both the postclustering 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 postclustering estimate $\check{\rho}_2 = 0.967$ is smaller than unity and we cannot reject the null hypothesis of a group-specific unit root. These postclustering 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 4. 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 106 stocks in Group 2. Clustering the time series therefore assists in the detecting 11 further stocks manifesting explosive behavior.

Our recursive k-means clustering procedure therefore reveals a two-cluster structure in the US S&P500 market. Again, one fast-growing group and one relatively slow-growing cluster are detected, but only the fast-growing group shows a strong signal of an explosive root that is supported by postclustering statistical tests. These results relate to earlier studies by Harvey et al. (2016), Kozak et al. (2020), and Feng et al. (2020) where grouping and clustering methods were used to classify asset returns. A key question in the application of k-means clustering in financial markets that needs to be addressed is whether group separations are statistically significant, thereby justifying the cluster outcomes. Oh and Patton (2023) and Patton and Weller (2022) tested for statistical significance in group separations in the S&P 100 stock market and U.S. mutual funds. Rejection of the hypothesis of homogeneous parameters gives evidence of group-specific heterogeneity, a conclusion that has implications for financial investment. More specifically, in our application of postclustering testing, we find statistically significant signals of an explosive price bubble in the group composed of high-tech firms and energy giants and fail to detect explosive roots in the other group.

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, 2023), 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. This article extends this mechanism to allow for latent group structures within a 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 article 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 postclustering 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 article 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. Some related research is in Okui and Wang (2021) and Lumsdaine et al. (2023). 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., 2022) 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 (cf., 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 needed for practical implementation. Some of these complications are currently under investigation and will be reported in future work.

APPENDIX A: PROOFS

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

A.1. *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 $\widehat{c}^* (:= (\widehat{c}_1^*, \widehat{c}_2^*, \dots, \widehat{c}_{G^0}^*))$ is the first-stage estimator of the distancing parameter vector c.

In order 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

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

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

LEMMA A1. *If Assumptions 1 and 3 hold, then*

$$\sup_{(c,\delta)\in\mathcal{C}_{G^0}\times\Delta_{G^0}} T^{4\gamma+2\xi}(\log T)^4\left|\widehat{Q}_{nT}(c,\delta) - \widetilde{Q}_{nT}(c,\delta)\right| = o_p(1),$$

where

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

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

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 A2. *Suppose Assumptions 1 and 3 hold. Then, when* $(n,T)\to\infty$,

$$d_H(c^0, \widehat{c}^*) = o_p\left(T^{-2\gamma-2\xi}(\log T)^{-4}\right). \quad \text{(A.1)}$$

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

$$T^{\gamma+\xi}(\log T)^2 \left| \widehat{c}^*_{\tau(\widehat{j})} - c^0_j \right| \to_p 0.$$

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

$$\left\| \widehat{c}^* - c^0 \right\| = o_p\Big(T^{-\gamma-\xi}(\log T)^{-2} \Big). \tag{A.2}$$

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

$$\mathcal{N}_\eta := \left\{ c \in \mathcal{C}_{G^0} : \left| c^0_j - c_j \right| < \eta,\ \forall j = 1, 2, \ldots, G^0 \right\}, \tag{A.3}$$

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

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

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

LEMMA A3. *Suppose Assumptions 1 and 3 hold. Then, for any fixed* $M > 0$,

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

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

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

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

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

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

LEMMA A4. *Suppose that Assumptions 1 and 3 hold, then:*

(i) if $\overline{c}^0_i > 0$ *and* $\widetilde{M}_1 \geq 3(c^{-2}_{low}) \max_{i \in \mathcal{I}_n} (\overline{\omega}^0_i)^2$, *then*

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

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

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

(iii) if $\overline{c}_i^0 < 0$ *and* $\widetilde{M}_3 \geq 2(c_{low}^{-1}) \max_{i\in\mathcal{I}_n}(\overline{\sigma}_{iu}^0)^2$,

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

LEMMA A5. *Suppose Assumptions 1 and 3 hold, then:*

(i) if $\overline{c}_i^0 > 0$, $\overline{M}_1 > 0$ *and* $\xi > 8c_{up}^2(1-\gamma)/(\min_{i\in\mathcal{I}_n}(\overline{\omega}_i^0)^2)$,

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

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

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

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

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

LEMMA A6. *Suppose Assumptions 1 and 3 hold. Let* $\eta = O(T^{-\gamma-\xi}(\log T)^{-2})$*. Then, when* $(n, T) \to \infty$,

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

where $\mathcal{N}_\eta$ *is defined in (A.3).*

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

$$\widehat{E}_{j,i} := \{\widehat{g}_i \neq j | g_i^0 = j\} \text{ and } \widehat{F}_{j,i} := \{g_i^0 \neq j | \widehat{g}_i = j\}. \tag{A.4}$$

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

LEMMA A7. *(Uniform Consistency of Clustering) Suppose Assumptions 1 and 3 hold. Then, when* $(n, T) \to \infty$,

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

Proof of Theorem 1: We use Lemmas A7(i) and A7(ii). In order to establish uniform consistency of the recursive k-means clustering algorithm, we first bound the CE 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

$$
\begin{aligned}
\sum_{j=1}^{G^0}\sum_{i\in\mathcal{G}^0(j)}\Pr\left(\widehat{E}_{j,i}\right) &\le n\max_{i\in\mathcal{I}_n}\mathbb{E}\mathbf{1}\{\widehat{g}_i(\widehat{c}^*)\neq g_i^0\}\le n\max_{i\in\mathcal{I}_n}\Pr\{\left|\widehat{g}_i(\widehat{c}^*)-g_i^0\right|>0\}\\
&\le n\max_{i\in\mathcal{I}_n}\sup_{c\in\mathcal{N}_\eta}\Pr\{\left|\widehat{g}_i(c)-g_i^0\right|>0\}+n\max_{1\le j\le G^0}\Pr\{\left|\widehat{c}_j^*-c_j^0\right|>\eta\}\\
&=o(1)+n\max_{1\le j\le G^0}\Pr\{\left|\widehat{c}_j^*-c_j^0\right|>\eta\}\\
&=o(1), \qquad\qquad \text{(A.5)}
\end{aligned}
$$

where the last step is due to the Markov inequality, Equation (A.2) in Lemma A1, and the rate restriction in Assumption 3. This proves Lemma A7(i). For Lemma A7(ii), we can follow the proof of Theorem 2.2(ii) in Su et al. (2016). The results of Theorem 1 are then extensions of Lemma A7 and the proof is complete. ■

A.2. *Proofs for Stage 2: Postclustering Estimation and Testing.* We need a lemma that establishes consistency of the variance estimates $\check{\omega}_j^2$, $\check{\sigma}_j^2$, and $\check{\lambda}_j$ and lemmas that show the limits of various sample moments. These lemmas are stated first.

LEMMA A8. *Suppose Assumptions 1 and 3 hold. Then, for any $j\in\mathcal{G}^0$, if $c_j^0\ge 0$, when $(n,T)\to\infty$,*

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

LEMMA A9. *Suppose Assumptions 1 and 3 hold. Then, for any $j\in\mathcal{G}^0$, when $(n,T)\to\infty$,*

$$
\tfrac{1}{n_jT^2}\textstyle\sum_{i\in\mathcal{G}^0(j)}\sum_{t=1}^T\widetilde{y}_{i,t-1}^2\to_p\frac{\left(\omega_j^0\right)^2}{6},\ if\ c_j^0=0; \qquad \text{(A.6)}
$$

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

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

LEMMA A10. *Suppose Assumptions 1 and 3 hold. Then, for any $j\in\mathcal{G}^0$, when $(n,T)\to\infty$,*

$$
\frac{1}{\sqrt{n_j}T}\sum_{i\in\mathcal{G}^0(j)}\sum_{t=1}^T\left(\widetilde{y}_{i,t-1}\widetilde{u}_{i,t}-\widehat{\lambda}_j+\frac{\widehat{\omega}_j^2}{2}\right)\Rightarrow\mathcal{N}\left(0,\ \frac{\left(\omega_j^0\right)^4}{12}\right),\ if\ c_j^0=0; \qquad \text{(A.7)}
$$

$$
\tfrac{1}{\sqrt{n_j}T^{\gamma}(\rho_j^0)^T}\textstyle\sum_{i\in\mathcal{G}^0(j)}\sum_{t=1}^T\widetilde{y}_{i,t-1}\widetilde{u}_{it}\Rightarrow\mathcal{N}\left(0,\frac{\left(\omega_j^0\right)^4}{4\left(c_j^0\right)^2}\right),\ if\ c_j^0>0; \qquad \text{(A.8)}
$$

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

where

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

Proof of Theorem 2: When $c_j^0>0$, the following decompositions apply to the numerator and denominator of the postclustering estimate, $\check{\rho}_j$:

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

and

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

In order to demonstrate the asymptotic equivalence between the postclustering estimates and the oracle estimates, we need to show the following:

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

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

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

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

Since the treatment of the denominator is identical to that of the numerator, we need to only focus on the numerator. In terms of A.2, for any $\widetilde{j} \neq j$ and any $\varepsilon > 0$,

$$\text{(A.9)} \qquad \Pr\left(\left|\frac{1}{\sqrt{n_j}T^{\gamma}(\rho_j^0)^T}\sum_{\widetilde{j}\neq j}\sum_{i\in\hat{\mathcal{G}}(\widetilde{j})\cap\mathcal{G}^0(j)}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right| > \varepsilon\right) \leq \Pr\left(\sum_{j=1}^{G^0}\widehat{E}_{j,nT}\right) \to 0,$$

when $(n, T) \to \infty$. Similarly, in terms of A.2, for any $\widetilde{j} \neq j$ and any $\varepsilon > 0$,

$$\text{(A.10)} \qquad \Pr\left(\left|\frac{1}{\sqrt{n_j}T^{\gamma}(\rho_j^0)^T}\sum_{\widetilde{j}\neq j}\sum_{i\in\hat{\mathcal{G}}(j)\cap\mathcal{G}^0(\widetilde{j})}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it}\right| > \varepsilon\right) \leq \Pr\left(\sum_{j=1}^{G^0}\widehat{F}_{j,nT}\right) \to 0,$$

when $(n, T) \to \infty$. Combining (A.9) and (A.10), we have

$$\text{(A.11)} \qquad \frac{1}{\sqrt{n_j}T^{\gamma}(\rho_j^0)^T}\sum_{i\in\hat{\mathcal{G}}(j)}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it} = \frac{1}{\sqrt{n_j}T^{\gamma}(\rho_j^0)^T}\sum_{i\in\mathcal{G}^0(j)}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}\widetilde{u}_{it} + o_p(1)$$

and

$$\text{(A.12)} \qquad \frac{1}{n_jT^{2\gamma}(\rho_j^0)^{2T}}\sum_{i\in\hat{\mathcal{G}}(j)}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2 = \frac{1}{n_jT^{2\gamma}(\rho_j^0)^{2T}}\sum_{i\in\mathcal{G}^0(j)}\sum_{t=1}^{T}\widetilde{y}_{i,t-1}^2 + o_p(1).$$

Based on (A.11) and (A.12), the asymptotic theory for the postclustering estimator is equivalent to that of the infeasible estimator, which is

$$\text{(A.13)} \qquad \sqrt{n_j}T^{\gamma}(\rho_j^0)^T\left(\check{\rho}_j - \rho_j^0\right) = \sqrt{n_j}T^{\gamma}(\rho_j^0)^T\left(\widehat{\rho}_j - \rho_j^0\right) + o_p(1).$$

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

$$\text{(A.14)} \quad \sqrt{n_j}T\left(\check{\rho}_j - \rho_j^0 - \frac{1}{T}\frac{6\lambda_j^0}{\left(\omega_j^0\right)^2} + \frac{1}{T}\frac{3\left(\omega_j^0\right)^2}{\left(\omega_j^0\right)^2}\right) = \sqrt{n_j}T\left(\widehat{\rho}_j - \rho_j^0 - \frac{1}{T}\frac{6\lambda_j^0}{\left(\omega_j^0\right)^2} + \frac{1}{T}\frac{3\left(\omega_j^0\right)^2}{\left(\omega_j^0\right)^2}\right) + o_p(1);$$

when $c_j^0 < 0$,

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

where $\overline{m}_{j,T} = (1/n_j)\sum_{i\in\mathcal{G}^0(j)} m_{i,T}$ and $m_{i,T} = \sum_{h=1}^{\infty}(\rho_j^0)^{h-1}\mathbb{E}(\widetilde{\epsilon}_{it}u_{i,t-h})$. Therefore, we only need to derive the limit distribution of the infeasible estimator in all cases. From Lemmas A9 and A10, the limit distributions of the oracle estimators that employ the true group identities are derived and this completes the proof. ■

Proof of Theorem 3: By Lemma A8 and Theorem 2, when $(n, T) \to \infty$,

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

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

So both the postclustering 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

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

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

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

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

where the last equality is due to Theorem 2 and the $O_p(\cdot)$ rate is sharp. For the postclustering 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 2 and the $O_p(\cdot)$ rate is again sharp. This concludes the proof. ■

A.3. *Proofs for the Estimation of Group Numbers.* LEMMA A11. *Suppose Assumptions 1 and 3 hold. Let $(n,T)\to\infty$. When (i) $\gamma>0$ and $\bar{c}_i^0 \ge 0$ or (ii) $\gamma = 0$, we have*

$$\min_{1\le G<G^0}\inf_{\widehat{\delta}(G)\in\Delta_G}\check{\sigma}^2_{\widehat{\delta}(G)} > \sigma_0^2, \text{ with probability approaching 1,} \tag{A.15}$$

and

$$\check{\sigma}^2_{\widehat{\delta}(G^0)} \to_p \sigma_0^2, \tag{A.16}$$

where σ_0^2 is defined in Equation (32),

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

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

Proof of Lemma A11: In order to show Equation (A.15), it is sufficient to show

$$\inf_{\widehat{\delta}(G)\in\Delta_G} \check{\sigma}^2_{\widehat{\delta}(G)} > \sigma_0^2, \text{ with probability approaching 1,} \tag{A.17}$$

for all $G < G^0$. The above relationship (A.17) holds since the true number of groups, G^0, is finite. Without losing generality, we discuss the case in which $G = G^0 - d$ and $d = 1$. Treatment for the case $d \geq 2$ is similar to the case $d = 1$ and is omitted.

When the minimum value of $\check{\sigma}^2_{\widehat{\delta}(G)}$ is attained, the individuals of G groups are correctly clustered to their true membership, and the individuals of the remaining group, namely, the individuals for the j^*-th true group, are wrongly allocated to the group whose group-specific distance parameter $c^0_{\widetilde{j}}$ is close to the group-specific parameter of these individuals, $c^0_{j^*}$.[30] Without losing generality, we assume $\widetilde{j} \leq G$ and $\widetilde{j} \leq j^*$ and still call the union of the $\widetilde{j}$-th and j^*-th true groups as the $\widetilde{j}$-th estimated group of the G-group partition.

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

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

$$\sim_a \frac{1}{nT}\sum_{j=\widetilde{j},j^*}\sum_{i\in\hat{\mathcal{G}}(j,G^0)}\sum_{t=1}^{T}\widetilde{y}^2_{i,t-1}\left(\check{\rho}_{\widetilde{j}}(G) - \check{\rho}_j(G^0)\right)^2, \tag{A.18}$$

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

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

[30] When only one group is misallocated, the leading term of the distance between the MSEs of the residuals is affected by the distance between group-specific slope coefficients, as shown by Equation (A.18). So once the difference between these slope coefficients is smaller, the distance between the MSEs of residuals decreases. Then it is evident that when the misallocated individuals are all put into the group with the nearest distancing parameter, the MSE of residuals attains its minimum.

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

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

in which Q is a positive constant. The case in which $j^* \leq \widetilde{j}$ follows a similar procedure to that mentioned above and is omitted. Then Equation (A.15) holds. Moreover, Equation (A.16) holds due to the consistency of k-means clustering and postclustering estimates:

$$\begin{aligned}\check{\sigma}^2_{\widehat{\delta}(G^0)} &:= \frac{1}{nT}\sum_{j=1}^{G^0}\sum_{i\in\hat{\mathcal{G}}(j,G^0)}\sum_{t=1}^{T}\left(\widetilde{y}_{it}-\widetilde{y}_{i,t-1}\check{\rho}_j\left(G^0\right)\right)^2 = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\widehat{\widetilde{u}}_{it}^2 \\ &\sim_a \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\widetilde{u}_{it}^2 = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}u_{it}^2 + O_p\left(\frac{1}{T}\right) \\ &\to_p \sigma_0^2, \end{aligned} \tag{A.19}$$

where the last line (A.19) holds due to Equation (32). The proof is now complete. ■

Proof of Theorem 4: From Theorems 1 and 2, it follows that

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

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

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

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

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

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

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

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

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

The proof is now complete. ■

Proofs of Equations (23)–(25): We first discuss the results of Equation (23). When $G = G^0$ and $c_j^0 < 0$, we have

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

whose derivations are similar to Theorem 2. Since $\check{n}_{j,h}/\check{n}_j \to \pi_{j,h}$, then we have

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

Therefore, it follows that

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

in which

$$
\begin{aligned}
&\mathrm{Avar}\left(\sqrt{\check{n}_j T^{1+\gamma}}(\check{\rho}_j(G) - \check{\rho}_{j,h}(G))\right) \\
&= \mathrm{Avar}\left(\sqrt{\check{n}_j T^{1+\gamma}}\left(\check{\rho}_j(G) - \rho_j^0\right)\right) + \mathrm{Avar}\left(\sqrt{\check{n}_j T^{1+\gamma}}\left(\check{\rho}_{j,h}(G) - \rho_j^0\right)\right) \\
&-2\,\mathrm{Acov}\left(\sqrt{\check{n}_j T^{1+\gamma}}\left(\check{\rho}_{j,h}(G) - \rho_j^0\right),\ \sqrt{\check{n}_j T^{1+\gamma}}\left(\check{\rho}_j(G) - \rho_j^0\right)\right).
\end{aligned}
$$

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

$$
\begin{aligned}
&\mathrm{Acov}\left(\frac{1}{\sqrt{\check{n}_j T^{1+\gamma}}} \sum_{i\in\hat{\mathcal{G}}(j,G)} \sum_{t=1}^{T} y_{i,t-1}u_{i,t},\ \frac{1}{\sqrt{\check{n}_j T^{1+\gamma}}} \sum_{i\in\hat{\mathcal{G}}(j,h,G)} \sum_{t=1}^{T} y_{i,t-1}u_{i,t}\right) \\
&= \mathrm{Acov}\left(\frac{1}{\sqrt{\check{n}_j T^{1+\gamma}}} \sum_{i\in\hat{\mathcal{G}}(j,h,G)} \sum_{t=1}^{T} y_{i,t-1}u_{i,t},\ \frac{1}{\sqrt{\check{n}_j T^{1+\gamma}}} \sum_{i\in\hat{\mathcal{G}}(j,h,G)} \sum_{t=1}^{T} y_{i,t-1}u_{i,t}\right) \\
&= \pi_{j,h}\cdot \mathrm{Avar}\left(\frac{1}{\sqrt{\check{n}_{j,h} T^{1+\gamma}}} \sum_{i\in\hat{\mathcal{G}}(j,h,G)} \sum_{t=1}^{T} y_{i,t-1}\epsilon_{i,t}\right),
\end{aligned}
$$

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

$$\begin{aligned}
&\text{Avar}\left(\sqrt{\check{n}_j T^{1+\gamma}}(\check{\rho}_j(G) - \check{\rho}_{j,h}(G))\right) \\
&= -\text{Avar}\left(\sqrt{\check{n}_j T^{1+\gamma}}\left(\check{\rho}_j(G) - \rho_j^0\right)\right) + \text{Avar}\left(\sqrt{\check{n}_j T^{1+\gamma}}\left(\check{\rho}_{j,h}(G) - \rho_j^0\right)\right) \\
&= -2c_j^0\left(\pi_{j,h}^{-1} - 1\right).
\end{aligned}$$

Further, based on the assumption of cross-sectional independence, we prove Equation (23). The proofs of (24) and (25) follow a procedure similar to (23). ■

Proof of Theorem 5: Assume any G with $\widetilde{G} \leq G \leq G_{\max}$ where $\widetilde{G}$ is estimated by IC. Also assume any $j = 1, 2, \ldots, G$. Under the null hypothesis of slope homogeneity in the jth group, the test statistic (27) converges to $\chi^2(\overline{G})$ based on Equations (23)–(25). Since the critical values $cv_{nT} := (1 + b\log(nT)) \cdot \chi^2_{0.95}(\overline{G}) \to \infty$, 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 (27) asymptotically diverges since

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

Also, since $\min\{\check{n}_j T^{1-\gamma}, \check{n}_j T^{2-2\gamma}, \check{n}_j(\rho_j^0)^{2T}\} \succ cv_{nT}$, power converges to unity and the estimator $\widehat{G}$ of (27) is consistent. The proof is complete. ■

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Table 1: Clustering errors with an underestimated group number
Table 2: Empirical frequency of model selection under DGP 0 ($\theta = 0.5 and k_{nT} = (nT) - 0.7$)
Table 3: Empirical frequency of model selection under DGP 1 ($\theta = 0.5 and k_{nT} = ((nT) - 0.7$)
Table 4: Empirical frequency of model selection under DGP 2 ($\theta = 0.5$ and $k_{nT} = ((nT)$-0.7)
Table 5: Empirical frequency of model selection under DGP 1 ($\theta = 0.5$ and $k_{nT} = ((nT)$-0.6)
Table 6: Clustering and estimation by the two stage procedure under DGP 1 and correct γ^*($\theta = 0.5$ and $\gamma^* = 0.1$)
Table 7: Clustering and estimation by the two stage procedure under DGP 2 and correct γ^*($\theta = 0.5$ and $\gamma^* = 0.1$)
Table 8: Clustering and estimation by the two stage procedure under DGP 1 and incorrect γ^*($\theta = 0.5$ and $\gamma^* = 0.4$)
Table 9: Clustering and estimation by the two stage procedure under DGP 2 and incorrect γ^*($\theta = 0.5$ and $\gamma^* = 0.4$)
Figure 1: Empirical frequency distribution and sample average (shown by the vertical line) of the signal-to-noise ratio in each group of DGPs 0-2.
Table 10: Empirical frequency of the clustering errors in the mixed-root model with stationary linear processes (c = (0.5,0) and $\theta = 0.5$) when $\gamma^* = 0.9$
Table 11: Power of tests for detecting explosiveness ($\theta = 0.5$)
Table 12: Empirical frequency of model selection under DGP 1 ($\theta = 0$ and $k_{nT} = ((nT)$-0.7)
Table 13: Empirical frequency of model selection under DGP 2 ($\theta = 0$ and $k_{nT} = ((nT)$-0.7)
Table 14: Empirical frequency of model selection under DGP 1 ($\theta = 0$ and $k_{nT} = ((nT)$-0.6)
Table 15: Clustering and estimation by the two stage procedure under DGP 1 and correct γ^*($\theta = 0$ and $\gamma^* = 0.1$)

REFERENCES

Abraham, J. M., W. N. Goetzmann, and S. M. Wachter, "Homogeneous Groupings of Metropolitan Housing Markets," *Journal of Housing Economics* 3 (1994), 186–206.

Ahn, D.-H., J. Conrad, and R. F. Dittmar, "Basis Assets," *The Review of Financial Studies* 22 (2009), 5133–74.

Ando, T., and J. Bai, "Panel Data Models with Grouped Factor Structure Under Unknown Group Membership," *Journal of Applied Econometrics* 31 (2016), 163–91.

Andrews, D. W., "Laws of Large Numbers for Dependent Non-Identically Distributed Random Variables," *Econometric Theory* 4 (1988), 458–67.

Arellano, M., and S. Bonhomme, "Identifying Distributional Characteristics in Random Coefficients Panel Data Models," *The Review of Economic Studies* 79 (2012), 987–1020.

Bai, J., "Panel Data Models with Interactive Fixed Effects," *Econometrica* 77 (2009), 1229–79.

———, and S. Ng, "Determining the Number of Factors in Approximate Factor Models," *Econometrica* 70 (2002), 191–221.

———, and ———, "A PANIC Attack on Unit Roots and Cointegration," *Econometrica* 72 (2004), 1127–77.

———, and ———, "Panel Unit Root Tests with Cross-Section Dependence: A Further Investigation," *Econometric Theory* 26 (2010), 1088–114.

Bollerslev, T., A. J. Patton, and H. Zhang, "Equity Clusters Through the Lens of Realized Semicorrelations," *Economics Letters* 211 (2022), 110245.

Bonhomme, S., T. Lamadon, and E. Manresa, "Discretizing Unobserved Heterogeneity," *Econometrica* 90 (2022), 625–43.

———, and E. Manresa, "Grouped Patterns of Heterogeneity in Panel Data," *Econometrica* 83 (2015), 1147–84.

Bourassa, S. C., E. Cantoni, and M. Hoesli, "Spatial Dependence, Housing Submarkets, and House Price Prediction," *The Journal of Real Estate Finance and Economics* 35 (2007), 143–60.

———, F. Hamelink, ———, and B. D. MacGregor, "Defining Housing Submarkets," *Journal of Housing Economics* 8 (1999), 160–83.

Brown, S. J., and W. N. Goetzmann, "Mutual Fund Styles," *Journal of Financial Economics* 43 (1997), 373–99.

Chang, Y., "Nonlinear IV Unit Root Tests in Panels with Cross-Sectional Dependency," *Journal of Econometrics* 110 (2002), 261–92.

———, and W. Song, "Testing for Unit Roots in Small Panels with Short-Run and Long-Run Cross-Sectional Dependencies," *The Review of Economic Studies* 76 (2009), 903–35.

Chen, K., and Y. Wen, "The Great Housing Boom of China," *American Economic Journal: Macroeconomics* 9 (2017), 73–114.

Chen, Y. Z., P. C. B. Phillips, and S. Shi, "Common Bubble Detection in Large Dimensional Financial Systems," *Journal of Financial Econometrics*, https://doi.org/10.1093/jjfinec/nbab027 (2022).

Cheng, X., F. Schorfheide, and P. Shao, "Clustering for Multi-Dimensional Heterogeneity," Working Paper, University of Pennsylvania, 2019.

Davidson, J., *Stochastic Limit Theory: An Introduction for Econometricians* (Oxford: OUP, 1994).

Diba, B. T., and H. I. Grossman, "Explosive Rational Bubbles in Stock Prices?," *The American Economic Review* 78 (1988), 520–30.

Fang, H., Q. Gu, W. Xiong, and L.-A. Zhou, "Demystifying the Chinese Housing Boom," *NBER Macroeconomics Annual* 30 (2016), 105–66.

Feng, G., S. Giglio, and D. Xiu, "Taming the Factor Zoo: A Test of New Factors," *The Journal of Finance* 75 (2020), 1327–70.

Freedman, D. A., "On Tail Probabilities for Martingales," *Annals of Probability* 3 (1975), 100–118.

Giglio, S., M. Maggiori, and J. Stroebel, "No-Bubble Condition: Model-Free Tests in Housing Markets," *Econometrica* 84 (2016), 1047–91.

Goetzmann, W. N., and S. M. Wachter, "Clustering Methods for Real Estate Portfolios," *Real Estate Economics* 23 (1995), 271–310.

Goodman, A. C., "Hedonic Prices, Price Indices and Housing Markets," *Journal of Urban Economics* 5 (1978), 471–84.

———, and T. G. Thibodeau, "Housing Market Segmentation," *Journal of Housing Economics* 7 (1998), 121–43.

———, and ———, "Housing Market Segmentation and Hedonic Prediction Accuracy," *Journal of Housing Economics* 12 (2003), 181–201.

Gouriéroux, C., P. C. B. Phillips, and J. Yu, "Indirect Inference for Dynamic Panel Models," *Journal of Econometrics* 157 (2010), 68–77.

Hahn, J., and G. Kuersteiner, "Asymptotically Unbiased Inference for a Dynamic Panel Model with Fixed Effects When Both n and T Are Large," *Econometrica* 70 (2002), 1639–57.

———, ———, and M. Mazzocco, "Joint Time-Series and Cross-Section Limit Theory under Mixingale Assumptions," *Econometric Theory* 38 (2022), 942–58.

———, and H. R. Moon, "Panel Data Models with Finite Number of Multiple Equilibria," *Econometric Theory* 26 (2010), 863–81.

———, and W. Newey, "Jackknife and Analytical Bias Reduction for Nonlinear Panel Models," *Econometrica* 72 (2004), 1295–319.

Hamilton, J. D., *Time Series Analysis* (Princeton, NJ: Princeton University Press, 1994).

Harvey, C. R., Y. Liu, and H. Zhu, "$\cdots$ and the Cross-Section of Expected Returns," *The Review of Financial Studies* 29 (2016), 5–68.

Hsiao, C., M. H. Pesaran, and A. K. Tahmiscioglu, "Maximum Likelihood Estimation of Fixed Effects Dynamic Panel Data Models Covering Short Time Periods," *Journal of Econometrics* 109 (2002), 107–50.

Huang, W., S. Jin, P. C. B. Phillips, and L. Su, "Nonstationary Panel Models with Latent Group Structures and Cross-Section Dependence," *Journal of Econometrics* 221 (2021), 198–222.

———, ———, and L. Su, "Identifying Latent Grouped Patterns in Cointegrated Panels," *Econometric Theory* 36 (2020), 410–56.

Im, K. S., M. H. Pesaran, and Y. Shin, "Testing for Unit Roots in Heterogeneous Panels," *Journal of Econometrics* 115 (2003), 53–74.

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

Kostakis, A., T. Magdalinos, and M. P. Stamatogiannis, "Robust Econometric Inference for Stock Return Predictability," *The Review of Financial Studies* 28 (2015), 1506–53.

Kozak, S., S. Nagel, and S. Santosh, "Shrinking the Cross-Section," *Journal of Financial Economics* 135 (2020), 271–92.

Leeb, H., and B. M. Pötscher, "Model Selection and Inference: Facts and Fiction," *Econometric Theory* 21 (2005), 21–59.

———, and ———, "Sparse Estimators and the Oracle Property, or the Return of Hodges' Estimator," *Journal of Econometrics* 142 (2008), 201–11.

Leng, X., H. Chen, and W. Wang, "Multi-Dimensional Latent Group Structures with Heterogeneous Distributions," *Journal of Econometrics* 233 (2023), 1–21.

Li, J., and Z. Liao, "Uniform Nonparametric Inference for Time Series," *Journal of Econometrics* 219 (2020), 38–51.

———, ———, and W. Zhou, "Uniform Nonparametric Inference for Spatially Dependent Panel Data," Working Paper, UCLA & SMU, 2022.

Liu, R., Z. Shang, Y. Zhang, and Q. Zhou, "Identification and Estimation in Panel Models with Overspecified Number of Groups," *Journal of Econometrics* 215 (2020), 574–90.

Liu, Y., P. C. B. Phillips, and J. Yu, "Panel Clustering Approach to Analyzing Bubble Behavior," Cowles Foundation Discussion Paper No. 2323, 2022.

Lu, X., and L. Su, "Determining the Number of Groups in Latent Panel Structures with an Application to Income and Democracy," *Quantitative Economics* 8 (2017), 729–60.

Ludvigson, S. C., and S. Ng, "The Empirical Risk–Return Relation: A Factor Analysis Approach," *Journal of Financial Economics* 83 (2007), 171–222.

———, and ———, "Macro Factors in Bond Risk Premia," *The Review of Financial Studies* 22 (2009), 5027–67.

Lumsdaine, R. L., R. Okui, and W. Wang, "Estimation of Panel Group Structure Models with Structural Breaks in Group Memberships and Coefficients," *Journal of Econometrics* 233 (2023), 45–65.

McLeish, D., "On the Invariance Principle for Nonstationary Mixingales," *The Annals of Probability* 5 (1977), 616–21.

Moon, H. R., and B. Perron, "Testing for a Unit Root in Panels with Dynamic Factors," *Journal of Econometrics* 122 (2004), 81–126.

———, and P. C. B. Phillips, "Estimation of Autoregressive Roots Near Unity Using Panel Data," *Econometric Theory* 16 (2000), 927–97.

———, and ———, "GMM Estimation of Autoregressive Roots Near Unity with Panel Data," *Econometrica* 72 (2004), 467–522.

Narayan, P. K., S. Mishra, S. Sharma, and R. Liu, "Determinants of Stock Price Bubbles," *Economic Modelling* 35 (2013), 661–67.

Norkutė, M., and J. Westerlund, "The Factor Analytical Approach in Near Unit Root Interactive Effects Panels," *Journal of Econometrics* 221 (2021), 569–90.

Oh, D. H., and A. J. Patton, "Dynamic Factor Copula Models with Estimated Cluster Assignments," *Journal of Econometrics* forthcoming, 2023. https://doi.org/10.1016/j.jeconom.2022.07.012

Okui, R., and W. Wang, "Heterogeneous Structural Breaks in Panel Data Models," *Journal of Econometrics* 220 (2021), 447–73.

Patton, A. J., and B. M. Weller, "Testing for Unobserved Heterogeneity via k-Means Clustering," *Journal of Business & Economic Statistics*, forthcoming, (2022), 1–35. https://doi.org/10.1080/07350015.2022.2061983

Pesaran, H., R. Smith, and K. S. Im, "Dynamic Linear Models for Heterogenous Panels," in *The Econometrics of Panel Data* (Springer Nature, Dordrecht, Netherlands: Springer, 1996), 145–95.

Pesaran, M. H., "Estimation and Inference in Large Heterogeneous Panels with A Multifactor Error Structure," *Econometrica* 74 (2006), 967–1012.

———, and ———, "Estimating Long-Run Relationships from Dynamic Heterogeneous Panels," *Journal of Econometrics* 68 (1995), 79–113.

———, and T. Yamagata, "Testing Slope Homogeneity in Large Panels," *Journal of Econometrics* 142 (2008), 50–93.

Phillips, P. C. B., "Time Series Regression with a Unit Root," *Econometrica* 55 (1987), 277–301.

———, "Dynamic Panel Anderson-Hsiao Estimation with Roots Near Unity," *Econometric Theory* 34 (2018), 253–76.

———, "Estimation and Inference with Near Unit Roots," *Econometric Theory* 39 (2023), 221–63.

———, and J. H. Lee, "Predictive Regression Under Various Degrees of Persistence and Robust Long-Horizon Regression," *Journal of Econometrics* 177 (2013), 250–64.

———, D. Li, and J. Gao, "Estimating Smooth Structural Change in Cointegration Models," *Journal of Econometrics* 196 (2017), 180–95.

———, and T. Magdalinos, "Limit Theory for Moderate Deviations from a Unit Root," *Journal of Econometrics* 136 (2007a), 115–30.

———, and ———, "Limit Theory for Moderate Deviations from a Unit Root Under Weak Dependence," in L. Mátyás and P. Sevestre, eds., G. D. A. Phillips and E. Tzavalis, eds., *The Refinement of Econometric Estimation and Test Procedures: Finite Sample and Asymptotic Analysis* (Cambridge UK: Cambridge Univesity Press, 2007b), 123–62.

———, and ———, "Unit Root and Cointegrating Limit Theory When Initialization Is in the Infinite Past," *Econometric Theory* 25 (2009), 1682–715.

———, and H. R. Moon, "Linear Regression Limit Theory for Nonstationary Panel Data," *Econometrica* 67 (1999), 1057–111.

———, and P. Perron, "Testing for a Unit Root in Time Series Regression," *Biometrika* 75 (1988), 335–46.

———, S. Shi, and J. Yu, "Testing for Multiple Bubbles: Historical Episodes of Exuberance and Collapse in the S&P 500," *International Economic Review* 56 (2015a), 1043–78.

———, ———, and ———, "Testing for Multiple Bubbles: Limit Theory of Real-Time Detectors," *International Economic Review* 56 (2015b), 1079–134.

———, and V. Solo, "Asymptotics for Linear Processes," *The Annals of Statistics* 20 (1992), 971–1001.

———, and D. Sul, "Dynamic Panel Estimation and Homogeneity Testing under Cross Section Dependence," *The Econometrics Journal* 6 (2003), 217–59.

———, Y. Wu, and ———, "Explosive Behavior in the 1990s Nasdaq: When Did Exuberance Escalate Asset Values?" *International Economic Review* 52 (2011), 201–26.

———, and J. Yu, "Dating the Timeline of Financial Bubbles During the Subprime Crisis," *Quantitative Economics* 2 (2011), 455–91.

Rosen, S., "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," *Journal of Political Economy* 82 (1974), 34–55.

Shi, S., and P. C. B. Phillips, "Diagnosing Housing Fever with an Econometric Thermometer," *Journal of Economic Surveys* 37 (2021), 159–86.

Shiller, R. J., *Irrational Exuberance* (Princeton, NJ: Princeton University Press, 2015).

Su, L., Z. Shi, and P. C. B. Phillips, "Identifying Latent Structures in Panel Data," *Econometrica* 84 (2016), 2215–64.

———, W. Wang, and Y. Zhang, "Strong Consistency of Spectral Clustering for Stochastic Block Models," *IEEE Transactions on Information Theory* 66 (2019), 324–38.

Westerlund, J., and R. Larsson, "Testing for a Unit Root in a Random Coefficient Panel Data Model," *Journal of Econometrics* 167 (2012), 254–73.

White, H., "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica* 48 (1980), 817–38.

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, $1 \le j \le G^0$ or $1 \le j \le G$
i	Subscript for individuals, namely, $1 \le i \le n$
$\mathcal{I}_n$	Set of individual subscripts $\{1, 2, \ldots, n\}$
$\mathcal{G}$	Set of group identities $\{1, 2, \ldots, G\}$
$\mathcal{G}^0$	Set of group identities $\{1, 2, \ldots, G^0\}$
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 $(:= (g_1, g_2, \ldots, g_n)')$
δ^0	True value of δ as $\delta^0(:= (g_1^0, g_2^0, \ldots, g_n^0)')$
$\widehat{\delta}$	Estimation of δ as $\widehat{\delta}(:= (\widehat{g}_1, \widehat{g}_2, \ldots, \widehat{g}_n)')$
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, $\mathcal{G}^0(j) = \{i \in \mathcal{I}_n \| g_i^0 = j\}$
$\hat{\mathcal{G}}(j)$	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, $c_j(:= c_{g_i})$ 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 $\overline{c}_i$
$\mathcal{C}_G$ and $\mathcal{C}_{G^0}$	Set of all possible G-dimensional or G^0-dimensional distance parameter c
ρ_j	Group-specific slope coefficient $\rho_j = 1 + c_j/T^\gamma$
ρ_j^0	True value of group-specific slope coefficient $\rho_j^0 = 1 + c_j^0/T^\gamma$
$\widehat{\rho}_j^*$	First-stage estimate of group-specific slope coefficient $\widehat{\rho}_j^* = 1 + \widehat{c}_j^*/T^\gamma$
$\widehat{\rho}_j$	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 $\check{\rho}_j = 1 + \check{c}_j/T^\gamma$

Symbol	Description
$\widehat{\overline{\rho}}_i^{TS}$	Time series estimate of individual parameter $\overline{\rho}_i$
c, ρ	G^0-dimensional parameters $c = (c_1, c_2, \ldots, c_{G^0})'$ and $\rho = (\rho_1, \rho_2, \ldots, \rho_{G^0})'$
c^0, ρ^0	G^0-dimensional true values $c^0 = (c_1^0, c_2^0, \ldots, c_{G^0}^0)'$ and $\rho^0 = (\rho_1^0, \rho_2^0, \ldots, \rho_{G^0}^0)'$
$\widehat{c}^*, \widehat{\rho}^*$	First-stage estimate of c, ρ as $\widehat{c}^* = (\widehat{c}_1^*, \widehat{c}_2^*, \ldots, \widehat{c}_{G^0}^*)'$ and $\widehat{\rho}^* = (\widehat{\rho}_1^*, \widehat{\rho}_2^*, \ldots, \widehat{\rho}_{G^0}^*)'$
$\widehat{c}, \widehat{\rho}$	Oracle estimate of c, ρ as $\widehat{c} = (\widehat{c}_1, \widehat{c}_2, \ldots, \widehat{c}_{G^0})'$ and $\widehat{\rho} = (\widehat{\rho}_1, \widehat{\rho}_2, \ldots, \widehat{\rho}_{G^0})'$
$\check{c}, \check{\rho}$	Post-clustering estimate of c, ρ as $\check{c} = (\check{c}_1, \check{c}_2, \ldots, \check{c}_{G^0})'$ and $\check{\rho} = (\check{\rho}_1, \check{\rho}_2, \ldots, \check{\rho}_{G^0})'$
c_{low}, c_{up}	Bounds for c_j; for instance, $c_j \in [-c_{up}, -c_{low}] \cup \{0\} \cup [c_{low}, c_{up}]$ in which we define $c_{up} := \max_{j\in\mathcal{G}^0} \lvert c_j^0 \cdot \mathbf{1}\{c_j^0 \neq 0\}\rvert$ and $c_{low} := \min_{j\in\mathcal{G}^0} \lvert c_j^0 \cdot \mathbf{1}\{c_j^0 \neq 0\}\rvert$
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 $\inf_{j\neq\tilde{j}} \lvert c_j - c_{\tilde{j}}\rvert \geq \dot{c}$
$\sigma_j^2, \lambda_j, \omega_j^2$	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
$(\sigma_j^0)^2, \lambda_j^0, (\omega_j^0)^2$	True values for variances, one-sided and two-sided long-run variances in the true jth group
$(\check{\sigma}_j)^2, \check{\lambda}_j, (\check{\omega}_j)^2$	Postclustering 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, $\overline{\rho}_i^0 = 1 + \frac{\overline{c}_i^0}{T^\gamma}$
$\widehat{\overline{c}}_i, \widehat{\overline{\rho}}_i$	Postclustering estimations for $\overline{c}_i$ and $\overline{\rho}_i$; namely, $\widehat{\overline{\rho}}_i = 1 + \frac{\widehat{\overline{c}}_i}{T^\gamma}$
$\overline{c}, \overline{\rho}$	$\overline{c} = (\overline{c}_1, \overline{c}_2, \ldots, \overline{c}_n)'$ and $\overline{\rho} = (\overline{\rho}_1, \overline{\rho}_2, \ldots, \overline{\rho}_n)'$
$\overline{c}^0, \overline{\rho}^0$	$\overline{c}^0 = (\overline{c}_1^0, \overline{c}_2^0, \ldots, \overline{c}_n^0)'$ and $\overline{\rho}^0 = (\overline{\rho}_1^0, \overline{\rho}_2^0, \ldots, \overline{\rho}_n^0)'$
$\widehat{\overline{c}}, \widehat{\overline{\rho}}$	$\widehat{\overline{c}} = (\widehat{\overline{c}}_1, \widehat{\overline{c}}_2, \ldots, \widehat{\overline{c}}_n)'$ and $\widehat{\overline{\rho}} = (\widehat{\overline{\rho}}_1, \widehat{\overline{\rho}}_2, \ldots, \widehat{\overline{\rho}}_n)'$
$\sigma^2, (\sigma^0)^2$	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 $\overline{\omega}_i^2 := \omega_{g_i}^2$
$(\overline{\sigma}_{iu}^0)^2, \overline{\lambda}_i^0, (\overline{\omega}_i^0)^2$	True values for variances, one-sided and two-sided long-run variances of the ith individual; for instance, $(\overline{\sigma}_{iu}^0)^2 := (\sigma_j^0)^2, \overline{\lambda}_i^0 := \lambda_j^0$ and $(\overline{\omega}_i^0)^2 := (\omega_j^0)^2$ when $g_i^0 = j$
$(\widehat{\overline{\sigma}}_{iu})^2, \widehat{\overline{\lambda}}_i, (\widehat{\overline{\omega}}_i)^2$	Individual time-series estimates for variances, one-sided and two-sided long-run variances of the ith individual, based on the postclustering 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 $(n, T) \to \infty$
$A_{nT} \sim_a B_{nT}$	$\Pr(\lvert B_{nT}/A_{nT}\rvert \neq 1) \to 0$ as $(n, T) \to \infty$
$A_{nT} \sim B_{nT}$	$\lvert B_{nT}/A_{nT}\rvert \to 1$ 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 $j = 1, 2, \ldots, G$
$\widetilde{G}$	The lower bound of group number selected by IC as in (21)
$\check{\rho}_j(G), \check{\rho}_{j,h}(G), h = 1, \ldots \overline{G}$	Postclustering 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 (22)
$\widehat{g}_{j,k}, k = 1, \ldots, \check{n}_j$	The estimated membership indicators of individuals in the estimated jth group
$\widehat{\delta}_j(G)$	The estimated membership of subgroups in the estimated jth group, and $\widehat{\delta}_j(G) = (\widehat{g}_{j,1}(G), \widehat{g}_{j,2}(G), \ldots, \widehat{g}_{j,\check{n}_j}(G))'$
$\check{\boldsymbol{\rho}}_j(G)$	The estimated slopes of subgroups in the estimated jth group, and $\check{\boldsymbol{\rho}}_j(G) = (\check{\rho}_{j,1}(G), \check{\rho}_{j,2}(G), \ldots, \check{\rho}_{j,\overline{G}}(G))'$;
$\hat{\mathcal{G}}(j, h, G), h = 1, 2, \ldots, \overline{G}$	Individuals in the estimated hth subgroup of the estimated jth group
$\check{n}_{j,h}, h = 1, 2, \ldots, \overline{G}$	The dimension of the estimated hth subgroup in the estimated jth group, assuming G groups
$\pi_{j,h}$	$\lim_n \frac{\check{n}_{j,h}}{\check{n}_j} \to \pi_{j,h}$;
$\widetilde{\pi}_j$	$\widetilde{\pi}_j = \mathrm{diag}\{\pi_{j,1}, \pi_{j,2}, \ldots, \pi_{j,\overline{G}}\}$.
γ	True value of the rate parameter
γ^*	Employed value of the rate parameter use in clustering.