Unified Inference for Panel Autoregressive Models with Unobserved Grouped Heterogeneity*

Wenxin Huang[a], Liangjun Su[b], and Yiru Wang[c]

[a] Antai College of Economics and Management, Shanghai Jiao Tong University

[b] School of Economics and Management, Tsinghua University

[c] Department of Economics, University of Pittsburgh

February 26, 2023

Abstract

This paper considers unified estimation and inference in panel autoregressive (AR) models. The AR coefficients are assumed to contain a latent group structure that allows the degree of persistence for each time series to be heterogeneous and unknown. We propose a novel penalized weighted least squares approach to simultaneously identifying the unknown group membership and consistently estimating the AR coefficients, regardless of whether the underlying AR process is stationary, unit-root, near-integrated, or even explosive. Theoretically, we establish the classification consistency, oracle properties, and unified asymptotic normal distributions for the proposed Lasso-type estimators. Empirically, we apply our data-driven method to uncover the existence of firm-level hidden bubbles in the U.S. stock market that have been unaccounted for in previous studies.

JEL Codes: C13; C33; C38; C55; G19

Keywords: Panel autoregression; Latent group structures; Weighted least squares; Uniform inference; Financial bubbles

---

*The authors thank seminar participants at the University of Pittsburgh, Zhejiang University, the 16th International Symposium on Econometric Theory and Applications (SETA 2022), the Midwest Econometrics Group Conference (MEG 2022) for their valuable comments. Huang acknowledges the support from NSFC under the grant No.72003127. Su gratefully acknowledges the support from NSFC under the grant No.72133002. Address correspondence to: Liangjun Su, School of Economics and Management, Tsinghua University, Beijing, 100084, China; Phone: +86 10 6278 9506; E-mail: sulj@sem.tsinghua.edu.cn.

1 Introduction

The statistical inference of autoregressive (AR) coefficients has received considerable attention because of its theoretical challenges and empirical relevance. From a theoretical perspective, the asymptotic theorems of AR coefficients suffer from the issue of discontinuity as the underlying time series may exhibit stationary, (near) unit-root/integrated, or explosive behavior, see Phillips (1987), Phillips and Magdalinos (2007), Mikusheva (2007), and Chan, Li and Peng (2012). For example, the convergence rate of the least squares (LS) estimator changes from the standard rate ($\sqrt{T}$) to super-consistency (T) when the AR coefficient varies from less than unity to unity. From an empirical perspective, the AR coefficients measure the degree of persistence of the underlying time series and have meaningful economic insights. For example, the random walk process is linked to the fundamental values of asset prices in the efficient market hypothesis, and explosive deviations from those fundamental values are defined as bubbles (Fama (1970), Diba and Grossman (1988), and Phillips, Shi and Yu (2015a)). Although various time-series methods have been established to bridge the discontinuities among stationary, unit-root, and nearly explosive cases, such as grid bootstrap (Hansen (1999)), block formation (Phillips, Moon and Xiao (2001)), uniform normalization (Mikusheva (2007)), and weighted score equations (So and Shin (1999), Chan, Li and Peng (2012)), few studies employ the advantages of panel data to resolve the discontinuity problem in the AR coefficients, which could be considered as parameter heterogeneity in the panel framework.

This paper proposes a unified approach to estimate panel autoregressive (PAR) models when the degree of persistence is unknown for the underlying time series. Our approach combines advantages in weighted score equations (Chan, Li and Peng (2012)), grid bootstrap (Hansen (1999), Mikusheva (2007)) in time-series models, and latent grouped patterns in panel data (Su, Shi and Phillips (2016)) to detect the unknown heterogeneity in the PAR coefficients and effectively establish uniform asymptotic limits and confidence intervals for the

PAR coefficients. In particular, we relax the common homogeneity assumption in classical panel data models and allow the PAR coefficients to contain an unobserved group structure. This structure can provide different time-series properties across groups in which the underlying AR process captures stationary, (near) unit-root, or explosive behavior. This new framework also provides a novel approach to detecting the unobserved common bubbles in large-dimensional financial systems, which is in contrast to the existing time-series bubble detection models that rely on the recursive right-tailed unit-root tests, where unit-root periods capturing normal market behavior and explosive deviations mimicking market exuberance; see Phillips, Shi and Yu (2015a) and Phillips and Shi (2018). Although there are undoubtedly important findings to the time-series bubble detection methods such as Phillips, Wu and Yu (2011), Etienne, Irwin and Garcia (2015), and Anundsen et al. (2016), we argue that the existing methods ignore potential heterogeneity in large-dimensional financial systems. Since bubbles are only identified on aggregate-level time-series data such as the S&P 500 index, NASDAQ index, and country-level house prices, researchers may rush to bubble conclusions for the whole market, failing to recognize explosive deviations occurring in sub sectors. For example, the dot-com bubble in the late 1990s was driven by massive growth in internet-related companies. Therefore, our panel approach helps detect the firm-specific hidden bubbles, enriching heterogeneous time-series patterns across different assets.

In this context, we consider a PAR model $y_{i,t} = \mu_i + \beta_i y_{i,t-1} + u_{i,t}$ with latent group structures in β_i to capture the unknown persistent behavior in each time series. Specifically, we assume the AR coefficient β_i is heterogeneous across groups and can take different values around unity, such as $\beta_i < 1$, $\beta_i = 1 - \frac{c_i}{n}$ and $\beta_i > 1$. Doing so allows our PAR model to contain stationary, (nearly) unit-root, and explosive series across groups. While within a group we maintain the homogeneity assumption on the PAR coefficients to achieve estimation efficiency by pooling cross-section information, the estimation and asymptotic inference are complicated by the unobserved group structure and issue of discontinuity in autoregressive process with or near unit roots. We modify the objective function in the classifier-Lasso (C-Lasso) method to overcome these challenges with weighted score equations. The modified C-Lasso estimation considers a penalized weighted least squares (PWLS) objective function that combines the weighted least squares (WLS) function proposed by Chan, Li and Peng (2012) as a method to resolve the issue of discontinuity with a C-Lasso penalty term of Su, Shi and Phillips (2016) to detect latent group structures in the PAR coefficients. Minimizing the PWLS objective function simultaneously estimates the group-specific AR coefficients while recovering the unobserved group memberships on the AR coefficients. Given the estimated group structures, we obtain group-specific post-Lasso estimators of the AR coefficients by WLS. The Lasso-type estimators enjoy efficiency gains by pooling the within-group individuals while also having enough flexibility to capture different degrees of persistence for the underlying time series.

Theoretically, we establish the convergence rate, classification consistency, and uniform asymptotic normality for Lasso-type estimators. First, we show that the point-wise convergence rate of the PAR coefficients depends on each time series' unknown degree of persistence. We unify these convergence rates with a weighted second moment of the underlying time series $\{y_{i,t}\}$, which is denoted as $G(y_i, \delta) = \frac{1}{T}\sum_{t=1}^{T} \frac{y_{i,t-1}^2}{(\delta + y_{i,t-1}^2)^{1/2}}$, $\delta > 0$. This result resolves the issue of discontinuity in convergence rates. Second, we consistently recover the unknown group structure in which all individuals are classified into the correct groups with probability approaching 1 (w.p.a.1). Lastly, we uniformly establish the asymptotic normal distributions for both the modified C-Lasso and post-Lasso estimators, regardless of the degree of persistence in the time series, with a group-normalized uniform convergence rate for the AR coefficient in group k (see Theorem 3.3 below). Besides, when estimating the PAR models with drifts, we apply Zhu, Cai and Peng's (2014) "long-run difference" (LRD) transformation and provide theoretical results for this scenario. In addition, we employ the grid bootstrap proposed by Hansen (1999) to construct the confidence interval of the AR coefficients. Our simulation studies demonstrate good finite sample performance of our unified estimators.

In the empirical application, we employ our method to detect hidden bubbles in the U.S. stock market, which will allow for heterogeneous price behaviors across firms. In particular, we study the firm-level dividend-adjusted prices panel using our PAR models with latent group patterns. The sample covers all available common stocks in the NYSE, AMEX, and NASDAQ markets from January 1926 to March 2022. Using our novel framework, we first confirm heterogeneous asset price behaviors and group structures across firms in the U.S. stock market. By comparing the AR coefficients with some threshold values near unity, we uncover hidden bubbles that exist in some subset and cannot be recognized by the aggregate data using some existing bubble detection techniques. Furthermore, we explain our firm-level bubble results by examining the firms' fundamental characteristics, where stocks with better operating performance and fewer debts are more likely to be explosive in their asset prices. In sum, our panel framework provides an informative and flexible approach to investigating firm-level bubbles and their potential determinants.

Compared with the literature, the main contribution of this paper is to develop a panel methodology that consistently estimates the unknown group-specific AR coefficients, identifies latent group structures, and further detects the unobserved explosive behaviors in the sub-panels. To the best of our knowledge, no existing works have employed group patterns to model unobserved heterogeneity in the AR coefficients and resolve the issue of discontinuity. The existing methods to estimate unobserved group patterns of heterogeneity, for example, Su, Shi and Phillips (2016), Wang and Su (2021), or Huang et al. (2021), focus on either stationary or nonstationary panels, respectively. By combining the advantages of panel data and weighted score equations, we provide a unified approach to consistently estimating the PAR models and simultaneously identifying the unknown group membership. The by-products of the group-specific results further provide theoretical support for empirical researchers to detect hidden bubbles in sub-panels.

The rest of the paper is organized as follows. Section 2 introduces a PAR model with latent group structures and develops the PWLS estimation procedure. Section 3 provides the main asymptotic properties of the estimators for the PAR models without and with drifts. Section 4 reports the simulation results. Section 5 empirically detects the firm-level hidden bubble in the U.S. stock markets and explores the firm-specific of the stock market bubbles. Section 6 concludes. The online appendix provides the proofs of the main theoretical results in Section 3, along with some additional empirical results.

NOTATION. Let $\mathbf{1}\{\cdot\}$ be an indicator function. Let M denote a generic positive constant whose value can vary in different places. Let $\lfloor x \rfloor$ denote the largest integer less than or equal to x. The operators $\xrightarrow{p}$ and $\Rightarrow$ denote convergence in probability and weak convergence, respectively. Unless indicated otherwise, we use $(T, N) \to \infty$ to signify that T and N pass to infinity jointly and use $(T, N) \overset{\text{seq.}}{\to} \infty$ to signify that T and N pass to infinity sequentially.

2 Model and Estimation

This section introduces a PAR model with a latent group structure and unknown degree of persistence. The PWLS estimation method is then proposed to estimate the AR coefficients and unobserved group structure.

2.1 Model setup

Consider the following autoregressive process:

$$y_{i,t} = \mu_i + \beta_i y_{i,t-1} + u_{i,t}, \tag{2.1}$$

where $t = 1, \ldots, T$ and $i = 1, \ldots, N$, $y_{i,0} = 0$, $u_{i,t}$ is an error term with a mean of zero and variance of $\sigma^2_{u,i}$, and μ_i is a drift term and could be regarded as fixed effects in the panel setup. For each individual i, $y_{i,t}$ is a stochastic process whose degree of persistence depends on the unknown autoregressive coefficient β_i, which is typically non-negative in economics and finance. When $\beta_i < 1$ (fixed), $\beta_i = 1 - \frac{c_i}{T}$ for some $c_i \in \mathbb{R}$ and $\beta_i > 1$ (fixed), $y_{i,t}$ is stationary, (near) unit-root/integrated, and explosive, respectively. Here, "$\beta_i < 1$ (fixed)" means that $\beta_i < 1$ takes a fixed value that does not depend on N or T; this also holds for

"$\beta_i > 1$ (fixed)." Below, we will suppress the modifier "fixed." We will discuss the two cases without and with drifts separately, viz., $\mu_i = 0$ and $\mu_i \neq 0$.

To model the unobserved heterogeneity, the autoregressive coefficients β_i are assumed to have a latent group structure. That is, β_i are the same within each group but differ across groups. Specifically, we allow the true values of β_i, which are denoted as β_i^0, to follow an unknown group pattern:

$$\beta_i^0 = \sum_{k=1}^{K_0} \alpha_k^0 \mathbf{1}\left\{i \in G_k^0\right\}, \tag{2.2}$$

where $\alpha_j^0 \neq \alpha_k^0$ for any $j \neq k$, $\cup_{k=1}^{K_0} G_k^0 = \{1, 2, \cdots, N\}$, and $G_j^0 \cap G_k^0 = \varnothing$ for any $j \neq k$. All values of α_k^0 could be further classified into three categories: stationary, (near) unit-root, and explosive groups. We allow multiple groups in each category. Let $N_k = \#G_k^0$ denote the cardinality of the set G_k^0. Both the number of groups K_0 and group membership are unknown. Note that the time-varying group structures are not considered in this framework, so the model implicitly requires that individual group membership does not vary over time.

Let $\boldsymbol{\alpha} \equiv (\alpha_1, \cdots, \alpha_{K_0})$ and $\boldsymbol{\beta} \equiv (\beta_1, \cdots, \beta_N)$. We denote the corresponding true values as $\boldsymbol{\alpha}^0 \equiv (\alpha_1^0, \cdots, \alpha_{K_0}^0)$ and $\boldsymbol{\beta}^0 \equiv (\beta_1^0, \cdots, \beta_N^0)$. The goal of this paper is to develop a uniform estimation method for the group-specific coefficients $\boldsymbol{\alpha}^0$, regardless of whether the underlying autoregressive process is stationary, (near) unit-root, or explosive.

2.2 Penalized Weighted Least Squares Estimation

In what follows, we introduce a four-step estimation procedure to obtain the two Lasso-type estimators for the autoregressive coefficients.

Step 1: Data processing. Prepare the transformed sequence $\tilde{y}_{i,t}$ for estimation. If $\mu_i = 0$, that is, with no drift, then $\tilde{y}_{i,t} = y_{i,t}$. If $\mu_i \neq 0$, that is, with drifts, we consider the LRD transformation proposed by Zhu, Cai and Peng (2014) to eliminate the drift. More specifically, let $m = \lfloor T/2 \rfloor$, and denote $\tilde{y}_{i,t} = y_{i,t} - y_{i,t+m}$ for $t = 1, \cdots, m$, and each $i =$

$1, \cdots, N$. Then, the model in (2.1) becomes:

$$\tilde{y}_{i,t} = \beta_i \tilde{y}_{i,t-1} + \tilde{u}_{i,t}. \tag{2.3}$$

Step 2: Initial estimation. We focus on the autoregressive process of $\tilde{y}_{i,t}$ and obtain the initial estimates of $\tilde{\beta}_i$ from the WLS estimation method proposed by Chan, Li and Peng (2012). For simplicity, we define $\sum_t$ and $\overline{\sum}_t$ such that $\sum_t = \sum_{t=1}^T$, $\overline{\sum}_t = \frac{1}{T}\sum_{t=1}^T$ when using the original data and $\sum_t = \sum_{t=1}^m$, $\overline{\sum}_t = \frac{1}{m}\sum_{t=1}^m$ when using the long-run differenced data. For each time-series regression, we obtain the WLS estimator $\tilde{\beta}_i$ by minimizing a weighted average of squared residuals with data-dependent weights:

$$\tilde{\beta}_i = \arg\min_{\beta_i} \overline{\sum}_t \frac{(\tilde{y}_{i,t} - \beta_i \tilde{y}_{i,t-1})^2}{(\delta + \tilde{y}_{i,t-1}^2)^{1/2}} = \left(\sum_t \frac{\tilde{y}_{i,t-1}^2}{(\delta + \tilde{y}_{i,t-1}^2)^{1/2}} \right)^{-1} \left(\sum_t \frac{\tilde{y}_{i,t-1}\tilde{y}_{i,t}}{(\delta + \tilde{y}_{i,t-1}^2)^{1/2}} \right), \tag{2.4}$$

where δ is a non-negative constant. The WLS estimator degenerates to the Cauchy estimate proposed in So and Shin (1999) when $\delta = 0$. In this context, we set $\delta > 0$ and suggest $\delta = 1$, as in Chan, Li and Peng (2012), to which we refer the readers for more details regarding the choice of δ.

Step 3: Penalized WLS estimation. We combine the C-Lasso methodology of Su, Shi and Phillips (2016) with the WLS estimation method to uniformly obtain the group-specific autoregressive coefficient estimator $\hat{\alpha}_k$. At this moment, we assume the number of groups K_0 is known, which can either be determined by prior economic theory or by a data-driven information criterion, as proposed in Section 3.3. Using the initial estimates of $\tilde{\beta}_i$ as starting values, we minimize the following PWLS objective function:

$$\tilde{Q}_{NT}^{K_0,\lambda}(\boldsymbol{\beta}, \boldsymbol{\alpha}) = \frac{1}{N}\sum_{i=1}^N \overline{\sum}_t \frac{(\tilde{y}_{i,t} - \beta_i \tilde{y}_{i,t-1})^2}{(\delta + \tilde{y}_{i,t-1}^2)^{1/2}} + \frac{\lambda}{N}\sum_{i=1}^N \prod_{k=1}^{K_0} \|\beta_i - \alpha_k\|, \tag{2.5}$$

where $\lambda = \lambda(N, T)$ is a tuning parameter. Unlike Su, Shi and Phillips's (2016) penalized least squares criterion function, the first component in (2.5) is a weighted average of the squared residuals with data-dependent weights, thus facilitating unified AR coefficient estimates for time series with an unknown degree of persistence. Minimizing the criterion function in (2.5) produces our modified C-Lasso estimates $\{\hat{\beta}_i, \hat{\alpha}_k\}$, based on which, we define the group identities as follows: $\hat{G}_k = \{i \in \{1, 2, \ldots, N\} : \hat{\beta}_i = \hat{\alpha}_k\}$ for $k = 1, \cdots, K_0$. In finite samples, if $\sum_{k=1}^{K_0} \mathbf{1}\{\hat{\beta}_i = \hat{\alpha}_k\} = 0$, we classify $i \in \hat{G}_l$ for some $l = 1, \ldots, K_0$ if $|\hat{\beta}_i - \hat{\alpha}_l| = \min_{1 \leq k \leq K_0} |\hat{\beta}_i - \hat{\alpha}_k|$.

Step 4: Post-Lasso estimation. Given the estimated group identities $\{\hat{G}_k, k = 1, ..., K_0\}$, we can pool the observations within each estimated group to obtain the post-Lasso estimator:

$$\hat{\alpha}_{\hat{G}_k}^{post} = \left(\sum_{i \in \hat{G}_k} \sum_t \frac{\tilde{y}_{i,t-1}^2}{(\delta + \tilde{y}_{i,t-1}^2)^{1/2}} \right)^{-1} \left(\sum_{i \in \hat{G}_k} \sum_t \frac{\tilde{y}_{i,t-1} \tilde{y}_{i,t}}{(\delta + \tilde{y}_{i,t-1}^2)^{1/2}} \right) \tag{2.6}$$

for $k = 1, ..., K_0$. It is well known that the tests based on the asymptotic critical values do not properly control for Type I errors in the local-to-unity framework (see, e.g., Basawa et al. (1991), Stock (1991), and Mikusheva (2007)). Hence, we adopt Hansen's (1999) grid bootstrap method in the panel autoregressive setting and uniformly construct the confidence intervals for the group-specific autoregressive coefficients, regardless of the underlying time-series properties.

3 Asymptotic Results

In this section, we study the asymptotic properties of two Lasso-type estimators.

3.1 Main assumptions

Let $\mathcal{F}_{i,t} = \sigma(u_{i,t}, u_{i,t-1}, ...)$, the minimal sigma-field generated by $u_{i,t}, u_{i,t-1}, ...$ To study the asymptotic properties of our estimators, we first impose the following three assumptions.

Assumption 3.1 ***(MDS Errors)***

(i) For each i, let $\{u_{i,t}, \mathcal{F}_{i,t}\}$ be a martingale difference sequence with zero mean $\mathbb{E}(u_{i,t}|\mathcal{F}_{i,t-1}) = 0$ and variance $\mathbb{E}(u_{i,t}^2|\mathcal{F}_{i,t-1}) = \sigma_{u,i}^2$; $\sup_{N\geq 1} \max_{1\leq i\leq N} \mathbb{E}[\,|u_{i,t}|^{2q+\epsilon}] < M$ with $q \geq 4$ and ϵ is an arbitrarily small positive constant.

(ii) $\{u_{i,t}\}_{t\geq 1}$ are independent across i.

Assumption 3.2 ***(Parameters Separability and Tuning Parameter)***

(i) For each $k = 1, ..., K_0$, $N_k/N \to \tau_k \in (0,1)$ as $N \to \infty$.

(ii) $\min_{1\leq k\neq j\leq K_0} \left|\alpha_k^0 - \alpha_j^0\right| \geq \underline{c}_\alpha$ for some fixed $\underline{c}_\alpha > 0$.

(iii) As $(N,T) \to \infty$, $\lambda(\log T)^\epsilon \to 0$, $\lambda T^{1/2} N^{-1/q}/(\log N)^{(1+\epsilon)/2} \to \infty$, and $T^{-1/2}N^{1/q}(\log N)^{(1+\epsilon)/2} \to 0$, where $q \geq 4$.

Assumption 3.1 imposes conditions on the error term and requires that $u_{i,t}$ be a martingale difference sequence (MDS) across time and independently distributed across individuals. For cases in which there is no drift, we follow Mikusheva (2007) to relax the independence assumption along the time dimension, which is commonly assumed in autoregressive models (see Hansen (1999), Phillips and Magdalinos (2007), Phillips (2014), Zhu, Cai and Peng (2014)). Along the cross-section dimension, we retain the cross-sectional independence assumption and leave the case of cross-sectional dependence for future research.

Assumption 3.2(i) is the same as Assumption A1(vii) in Su, Shi and Phillips (2016); this implies that the number of individuals in each group grows at the same rate as $N \to \infty$; thus, no group is asymptotically negligible. Assumption 3.2(ii) is the same as Assumption A1(vi) in Su, Shi and Phillips (2016); it specifies that parameters of different groups are well separated from each other. Similar conditions are assumed in the panel literature with latent group patterns; see, e.g., Bonhomme and Manresa (2015), Ando and Bai (2016), and Huang et al. (2021). Assumption 3.2(iii) specifies the conditions on the permissible range of tuning parameter λ and the relative rates at which N and T pass to infinity. It ensures the pointwise consistency of the AR coefficients and classification consistency of the unknown group structures. In addition, the choice of tuning parameter is uniform to the underlying time series with different degree of persistence as we employ the data-dependent weights in our PWLS objective function. One can verify that the permissible range of values for λ that satisfy Assumption 3.2(iii) is $\lambda \propto T^{-a}$ for $a \in (0, \frac{1}{2})$. This range is the same as that in stationary panels but different from that in nonstationary panels (see, e.g., Su, Shi and Phillips (2016), Huang et al. (2021)).

3.2 Main theoretical results

3.2.1 The case with no drift

We first discuss the asymptotic results for the pure autoregressive model when $\mu_i = 0$ in (2.1).

Let $G(y_i, \delta) = \frac{1}{T}\sum_{t=1}^{T} \frac{y_{i,t-1}^2}{(\delta + y_{i,t-1}^2)^{1/2}}$. We can define three sets of groups, $\mathcal{G}_{stat}, \mathcal{G}_{ur}$, and $\mathcal{G}_{exp}$, which corresponds to the stationary, the (near) unit-root, and the explosive groups, respectively. Define

$$\Gamma(T, \alpha_k^0) = T^{-1/2}\mathbf{1}\left\{\alpha_k^0 < 1\right\} + T^{-1}\mathbf{1}\left\{\alpha_k^0 = 1 - \frac{c_k^0}{T}\right\} + T^{1/2}(\alpha_k^0)^{-(T-1)}\mathbf{1}\left\{\alpha_k^0 > 1\right\}.$$

The following theorem establishes the consistency of $\hat{\beta}_i$ and $\hat{\alpha}_k$.

Theorem 3.1 *Suppose that Assumptions 3.1 and 3.2 hold. Then*

(i) $G(y_i, \delta)\left(\hat{\beta}_i - \beta_i^0\right) = O_p(T^{-1/2} + \lambda)$, *for* $i = 1, ..., N$.

(ii) (a) $\frac{1}{N}\sum_{i=1}^{N} G(y_i, \delta)^2\left(\hat{\beta}_i - \beta_i^0\right)^2 = O_p(T^{-1})$;

(b) $\frac{1}{N}\sum_{i\in G_k^0}\left(\hat{\beta}_i - \beta_i^0\right)^2 = O_p(T^{-1})$ *for* $G_k^0 \in \mathcal{G}_{stat}$;

(c) $\frac{1}{N}\sum_{i\in G_k^0}\left(\hat{\beta}_i - \beta_i^0\right)^2 = O_p(T^{-2})$ *for* $G_k^0 \in \mathcal{G}_{ur}$;

(d) $\frac{1}{N}\sum_{i\in G_k^0}\left(\hat{\beta}_i - \beta_i^0\right)^2 = O_p(T(\alpha_k^0)^{-2(T-1)})$ *for* $G_k^0 \in \mathcal{G}_{exp}$.

(iii) $\hat{\alpha}_k - \alpha_k^0 = O_p(\sum_{l=1}^{K_0}\Gamma(T, \alpha_l^0))$ *for* $k = 1, \ldots, K_0$.

Theorem 3.1(i) and (ii) establishes the preliminary pointwise and mean-square convergence rates for $\{\hat{\beta}_i\}$, respectively. Compared with the standard pointwise and mean-square convergence rates in the stationary and nonstationary settings, Theorem 3.1 introduces a data-dependent "normalization function," $G(y_i, \delta)$, which exhibits different asymptotic behaviors for autoregressive series with different degrees of persistence. With $G(y_i, \delta)$, we are able to provide unified inference for $\{\beta_i^0\}$. Compared with the normalization function in Mikusheva (2007), our data-dependent "weight" is applied within the summation operator $\sum_t$. The pointwise convergence rate in Theorem 3.1(i) depends on $T^{-1/2} + \lambda$, which is consistent with existing C-Lasso limit theory in stationary panels (e.g., Su, Shi and Phillips (2016)). Theorem 3.1(ii) reports the mean-square convergence rates for three different groups, viz., stationary groups, (near) unit-root groups, and explosive groups. This theorem uncovers the conventional rates of consistency within each specific group, which are the same as those in the literature. For example, Theorem 3.1(ii)(b) indicates that the estimators have the standard T^{-1}-rate of mean-square convergence when the autoregressive coefficients lie in the stationary regime. Theorem 3.1(iii) indicates that the group-specific parameters $\{\alpha_k^0\}$ can be estimated consistently by $\{\hat{\alpha}_k\}$. As we remarked in the proof of the above theorem, the rates cannot be improved without showing the uniform classification consistency.

We now study classification consistency. Define $\hat{E}_{kNT,i} = \{i \notin \hat{G}_k | i \in G_k^0\}$ and $\hat{F}_{kNT,i} = \{i \notin G_k^0 | i \in \hat{G}_k\}$, where $i = 1, ..., N$ and $k = 1, ..., K_0$. Let $\hat{E}_{kNT} = \cup_{i \in \hat{G}_k} \hat{E}_{kNT,i}$ and $\hat{F}_{kNT} = \cup_{i \in \hat{G}_k} \hat{F}_{kNT,i}$. The events $\hat{E}_{kNT}$ and $\hat{F}_{kNT}$ mimic type I and type II errors in statistical tests. Following Su, Shi and Phillips (2016), we say that a classification method is individually consistent if $P(\hat{E}_{kNT,i}) \to 0$ as $(N, T) \to \infty$ for each $i \in G_k^0$ and $k = 1, ..., K$, and $P(\hat{F}_{kNT,i}) \to 0$ as $(N, T) \to \infty$ for each $i \in G_k^0$ and $k = 1, ..., K_0$. It is uniformly consistent if $P(\cup_{k=1}^{K_0} \hat{E}_{kNT}) \to 0$ and $P(\cup_{k=1}^{K_0} \hat{F}_{kNT}) \to 0$ as $(N, T) \to \infty$.

The following theorem establishes the uniform classification consistency.

Theorem 3.2 *Suppose that Assumptions 3.1 and 3.2 hold. Then* *(i)* $P(\cup_{k=1}^{K_0}\hat{E}_{kNT}) \leq \sum_{k=1}^{K_0} P(\hat{E}_{kNT}) \to 0$ *as* $(N,T) \to \infty$,

(ii) $P(\cup_{k=1}^{K_0}\hat{F}_{kNT}) \leq \sum_{k=1}^{K_0} P(\hat{F}_{kNT}) \to 0$ *as* $(N,T) \to \infty$.

Theorem 3.2 implies uniform classification consistency – all individuals within a certain group, say G_k^0, can be simultaneously and correctly classified into the same group (denoted as $\hat{G}_k$) w.p.a.1. Conversely, all individuals that are classified into the same group, say $\hat{G}_k$, simultaneously belong to the same group (G_k^0) w.p.a.1. Let $\hat{N}_k = \#\hat{G}_k$. One can easily show that $P(\hat{G}_k = G_k^0) \to 1$ so that $P(\hat{N}_k = N_k) \to 1$.

To ensure the asymptotic normality for the modified C-Lasso and post-Lasso estimators, we first specify the following covariance conditions.

Assumption 3.3 ***(Covariance)*** *Let* $Q_i = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \frac{y_{i,t-1}^2}{\delta+y_{i,t-1}^2}$.

(i) For each $k = 1,\cdots,K_0$, $\Sigma_k = \lim_{N\to\infty} \frac{1}{N_k}\sum_{i\in G_k^0} \sigma_{u,i}^2 Q_i$ *exists and* $\Sigma_k > 0$.

(ii) For each $k = 1,\cdots,K_0$, $\mathcal{Q}_k = \lim_{N\to\infty} \frac{1}{N_k}\sum_{i\in G_k^0} Q_i$ *exists and* $\mathcal{Q}_k \neq 0$.

We now examine the oracle properties of the two Lasso-type estimators. If the group membership is known, the oracle WLS estimator of α_k^0 is given by

$$\hat{\alpha}_{G_k^0} = \arg\min_{\alpha_k} \frac{1}{N_k T} \sum_{i\in G_k^0} \sum_{t=1}^{T} \frac{(y_{i,t} - \alpha_k y_{i,t-1})^2}{(\delta + y_{i,t-1}^2)^{1/2}}. \tag{3.1}$$

The following theorem reports the oracle properties and establishes the asymptotic normality for the modified C-Lasso estimator $\hat{\alpha}_k$ and post-Lasso estimator $\hat{\alpha}_{\hat{G}_k}^{post}$.

Theorem 3.3 *Suppose that Assumptions 3.1, 3.2, and 3.3 hold. Define*

$$\Upsilon_k = \left(\sum_{i\in\hat{G}_k}\sum_{t=1}^{T} \frac{y_{i,t-1}^2}{\delta + y_{i,t-1}^2}\right)^{-1/2} \left(\sum_{i\in\hat{G}_k}\sum_{t=1}^{T} \frac{y_{i,t-1}^2}{(\delta + y_{i,t-1}^2)^{1/2}}\right).$$

Then, as $(T,N) \overset{seq.}{\to} \infty$, *it holds that*

(i) $\Upsilon_k\,(\hat{\alpha}_k - \alpha_k^0) \Rightarrow N(0, \mathcal{Q}_k^{-1}\Sigma_k)$,

(ii) $\Upsilon_k \left(\hat{\alpha}^{post}_{\hat{G}_k} - \alpha^0_k \right) \Rightarrow N(0, \mathcal{Q}_k^{-1}\Sigma_k)$,

where $\mathcal{Q}_k$ *and* Σ_k *are defined in Assumption 3.3.*

Theorem 3.3 implies that, under Assumptions 3.1–3.3, the modified C-Lasso estimator $\hat{\alpha}_k$ and the post-Lasso estimator $\hat{\alpha}^{post}_{\hat{G}_k}$ are asymptotically equivalent and achieve the same limit distribution as the oracle estimator $\hat{\alpha}_{G^0_k}$, thus enjoying the asymptotic oracle property. Here, we introduce a data-dependent convergence rate, Υ_k. Much like with $G(y_i, \delta)$, Υ_k exhibits different asymptotic behaviors for autoregressive series with different degrees of persistence. Given the group-specific data-dependent rate Υ_k, we provide the unified asymptotic distribution for the estimators of the group-specific parameters.

3.2.2 The case with drifts

Now, we discuss the asymptotic results of the model with drifts, that is, $\mu_i \neq 0$ in (2.1), where we use the LRD for data transformation. We first lay out the modified assumption required in this scenario.

Assumption 3.4 ***(Independent Errors)*** $\{u_{i,t}\}$ *are independent across* i *and* t *such that* $\mathbb{E}(u_{i,t}) = 0$*,* $\mathbb{E}(u^2_{i,t}) = \sigma^2_{u,i}$*, and* $\sup_{N,T\geq 1} \max_{1\leq i\leq N} \max_{1\leq t\leq T} \mathbb{E}(u^4_{i,t}) < M$*.*

Assumption 3.4 strengthens Assumption 3.1 and imposes the error term to be independent but not identically distributed. Under the original MDS setting, the LRD transformation induces serial correlations, and the desired central limit theorem fails. Thus, a stronger assumption is required in the case with drifts. Similar independence conditions are required in the time-series autoregressive model with drifts (see, e.g., Hansen (1999), Phillips (2014)). In addition, we notice that Zhu, Cai and Peng (2014) relax the independence assumption to allow for serially correlated errors in the autoregressive predictors; however, the errors in the main predictive regression are still assumed to be independent.

Assumption 3.5 ***(Covariance-LRD)*** *Let* $\tilde{Q}_i = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{m} \frac{\tilde{y}^2_{i,t-1}}{\delta + \tilde{y}^2_{i,t-1}}$.

(i) For each $k=1,\cdots,K_0$, $\tilde{\Sigma}_k = \lim_{N\to\infty} \frac{1}{N_k}\sum_{i\in G_k^0} 2\sigma_{u,i}^2 \tilde{Q}_i$ exists and $\tilde{\Sigma}_k > 0$.

(ii) For each $k=1,\cdots,K_0$, $\tilde{\mathcal{Q}}_k = \lim_{N\to\infty} \frac{1}{N_k}\sum_{i\in G_k^0} \tilde{Q}_i$ exists and $\tilde{\mathcal{Q}}_k \neq 0$.

Assumption 3.5 is an analogy of Assumption 3.3. It specifies the conditions to ensure the asymptotic normality for the modified C-Lasso and post-Lasso estimators when we use the LRD in the data transformation.

Define

$$\tilde{\Gamma}(m,\alpha_k^0) = m^{-1/2}\mathbf{1}\left\{\alpha_k^0 < 1\right\} + m^{-3/2}\mathbf{1}\left\{\alpha_k^0 = 1 - \frac{c_k^0}{T}\right\} + m^{1/2}(\alpha_k^0)^{-2m}\mathbf{1}\left\{\alpha_k^0 > 1\right\}.$$

The following theorem establishes the consistency of $\hat{\beta}_i$ and $\hat{\alpha}_k$.

Theorem 3.4 *Suppose that Assumptions 3.2 and 3.4 hold. Let $G(\tilde{y}_i,\delta) = \frac{1}{m}\sum_{t=1}^m \frac{\tilde{y}_{i,t-1}^2}{(\delta+\tilde{y}_{i,t-1}^2)^{1/2}}$.*

(i) $G(\tilde{y}_i,\delta)\left(\hat{\beta}_i - \beta_i^0\right) = O_p(m^{-1/2}+\lambda)$, for $i=1,...,N$.

(ii) (a) $\frac{1}{N}\sum_{i=1}^N G(\tilde{y}_i,\delta)^2\left(\hat{\beta}_i - \beta_i^0\right)^2 = O_p(m^{-1})$;

(b) $\frac{1}{N}\sum_{i\in G_k^0}\left(\hat{\beta}_i - \beta_i^0\right)^2 = O_p(m^{-1})$ for $G_k^0 \in \mathcal{G}_{stat}$;

(c) $\frac{1}{N}\sum_{i\in G_k^0}\left(\hat{\beta}_i - \beta_i^0\right)^2 = O_p(m^{-3})$ for $G_k^0 \in \mathcal{G}_{ur}$;

(d) $\frac{1}{N}\sum_{i\in G_k^0}\left(\hat{\beta}_i - \beta_i^0\right)^2 = O_p(m(\alpha_k^0)^{-4m})$ for $G_k^0 \in \mathcal{G}_{exp}$.

(iii) $\hat{\alpha}_k - \alpha_k^0 = O_p(\sum_{l=1}^{K_0}\tilde{\Gamma}(m,\alpha_l^0))$ for $k=1,\ldots,K_0$.

The following theorem establishes the uniform classification consistency.

Theorem 3.5 *Suppose that Assumptions 3.2, 3.4, and 3.5 hold. Then*

(i) $P(\cup_{k=1}^{K_0}\hat{E}_{kNT}) \leq \sum_{k=1}^{K_0} P(\hat{E}_{kNT}) \to 0$ as $(N,T)\to\infty$,

(ii) $P(\cup_{k=1}^{K_0}\hat{F}_{kNT}) \leq \sum_{k=1}^{K_0} P(\hat{F}_{kNT}) \to 0$ as $(N,T)\to\infty$.

The following theorem reports the oracle property of the Lasso estimator $\hat{\alpha}_k$ and the post-Lasso estimator $\hat{\alpha}_{\hat{G}_k}^{post}$.

Theorem 3.6 *Suppose that Assumptions 3.4, 3.2, and 3.5 hold. Let*

$$\tilde{\Upsilon}_k = \left(\sum_{i \in \hat{G}_k} \sum_{t=1}^{m} \frac{\tilde{y}_{i,t-1}^2}{\delta + \tilde{y}_{i,t-1}^2} \right)^{-1/2} \left(\sum_{i \in \hat{G}_k} \sum_{t=1}^{m} \frac{\tilde{y}_{i,t-1}^2}{(\delta + \tilde{y}_{i,t-1}^2)^{1/2}} \right).$$

Then, as $(T, N) \overset{seq.}{\to} \infty$*, it holds that*

(i) $\tilde{\Upsilon}_k \left(\hat{\alpha}_k - \alpha_k^0 \right) \Rightarrow N(0, \tilde{\mathcal{Q}}_k^{-1} \tilde{\Sigma}_k)$,

(ii) $\tilde{\Upsilon}_k \left(\hat{\alpha}_{\hat{G}_k}^{post} - \alpha_k^0 \right) \Rightarrow N(0, \tilde{\mathcal{Q}}_k^{-1} \tilde{\Sigma}_k)$,

where $\tilde{\mathcal{Q}}_k$ *and* $\tilde{\Sigma}_k$ *are defined in Assumption 3.5.*

Theorems 3.4–3.6 parallel Theorems 3.1–3.3. They provide the asymptotics for the estimators when the LRD is used in the data transformation. In this scenario, the data-dependent scaling, $G(\tilde{y}_i, \delta)$ and $\tilde{\Upsilon}_k$, are functions of the transformed data $\tilde{y}_i$ instead. In particular, the rates of consistency of $\hat{\beta}_i$ for different groups are different from those in Theorem 3.1(ii)(b)–(d), resulting from the different asymptotic behaviors of unit-root processes with and without drifts. Our simulation studies also confirm the faster convergence rates in this scenario.

3.3 Determination of the Number of Groups

In practice, the exact number K_0 of groups is typically unknown. When K_0 is unknown, we assume K_0 is bounded from above by a finite integer $K_{\max}$ and determine the number of groups using some information criterion (IC). In this subsection, we discuss the general framework with and without drifts. Recall that $\tilde{y}_{i,t} = y_{i,t}$ if $\mu_i = 0$ and $\tilde{y}_{i,t} = y_{i,t} - y_{i,t+m}$ if $\mu_i \neq 0$. Similarly, we use $\tilde{\Upsilon}_k$ to denote the scaling for both cases.

Replacing K_0 in the PWLS criterion function in (2.5) by K, we obtain the modified C-Lasso estimators $\{\hat{\beta}_i(K, \lambda), \hat{\alpha}_k(K, \lambda)\}$ of (β_i, α_k), which depends on (K, λ). We classify individual i into group $\hat{G}_k(K, \lambda)$ if and only if $\hat{\beta}_i(K, \lambda) = \hat{\alpha}_k(K, \lambda)$. Let $\hat{G}(K, \lambda) \equiv$

$\{\hat{G}_1(K,\lambda),\cdots,\hat{G}_K(K,\lambda)\}$. The post-Lasso WLS estimate of α_k^0 can be denoted as follows:

$$\hat{\alpha}^{post}_{\hat{G}_k(K,\lambda)} = \left(\sum_{i\in\hat{G}_k(K,\lambda)}\sum_t \frac{\tilde{y}^2_{i,t-1}}{(\delta+\tilde{y}^2_{i,t-1})^{1/2}}\right)^{-1}\left(\sum_{i\in\hat{G}_k(K,\lambda)}\sum_t \frac{\tilde{y}_{i,t-1}\tilde{y}_{i,t}}{(\delta+\tilde{y}^2_{i,t-1})^{1/2}}\right).$$

We propose selecting K by minimizing the following IC:

$$IC(K,\lambda) = \log\hat{\sigma}^2_{\hat{G}_k(K,\lambda)} + \rho_{NT}pK, \tag{3.2}$$

where $\hat{\sigma}^2_{\hat{G}_k(K,\lambda)} \equiv \frac{1}{N}\sum_{k=1}^K\sum_{i\in\hat{G}_k(K,\lambda)}\overline{\sum}_t(\tilde{y}_{i,t}-\hat{\alpha}^{post}_{\hat{G}_k(K,\lambda)}\tilde{y}_{i,t-1})^2$ and ρ_{NT} is a tuning parameter. Then, the number of groups is determined as $\hat{K}(\lambda)\equiv\arg\min_{1\le K\le K_{\max}}IC(K,\lambda)$.

Let $G^{(K)}\equiv(G_{K,1},\cdots,G_{K,K})$ be any K-partition of $\{1,2,\cdots,N\}$ and $\mathcal{G}_K$ a collection of all such partitions. Let $\hat{\sigma}^2_{G^{(K)}}\equiv\frac{1}{N}\sum_{k=1}^K\sum_{i\in G_{K,k}}\overline{\sum}_t(\tilde{y}_{i,t}-\hat{\alpha}_{G_{K,k}}\tilde{y}_{i,t-1})^2$, where $\hat{\alpha}_{G_{K,k}}\equiv\arg\min_{\alpha_k}\frac{1}{N_k}\sum_{i\in G_{K,k}}\overline{\sum}_t\frac{(\tilde{y}_{i,t}-\alpha_k\tilde{y}_{i,t-1})^2}{(\delta+\tilde{y}^2_{i,t-1})^{1/2}}$. We add the following two assumptions.

Assumption 3.6 *As* $(N,T)\to\infty$, $\min_{1\le K<K_0}\inf_{G^{(K)}\in\mathcal{G}_K}\hat{\sigma}^2_{G^{(K)}}\xrightarrow{p}\underline{\sigma}^2>\sigma_0^2$, *where* $\sigma_0^2\equiv\lim_{(N,T)\to\infty}\frac{1}{N}\sum_{k=1}^{K_0}\sum_{i\in G_k^0}\overline{\sum}_t(\tilde{y}_{i,t}-\alpha_k^0\tilde{y}_{i,t-1})^2$.

Assumption 3.7 *As* $(N,T)\to\infty$, $\rho_{NT}\to 0$ *and* $\rho_{NT}\left(\min_{k=1}^{K_0}\tilde{\Upsilon}_k^2\right)\to\infty$.

The following theorem justifies the use of (3.2) as a selector criterion for K.

Theorem 3.7 *Suppose Assumptions 3.1 or 3.4, 3.2, 3.3 or 3.5, 3.6, and 3.7 hold. Then,* $P\left(\hat{K}(\lambda)=K_0\right)\to 1$ *as* $(T,N)\overset{seq.}{\to}\infty$.

Assumptions 3.6–3.7 are adapted from Assumptions A4–A5 in Su, Shi and Phillips (2016). Assumption 3.6 requires that asymptotic mean squared errors of all under-fitted models are larger than that of the true models. Assumption 3.7 imposes conditions on ρ_{NT}, hence ensuring the consistency of model selection. Different from Su, Shi and Phillips (2016), the convergence rate of ρ_{NT} in this unified framework depends on $\tilde{\Upsilon}_k$. Theorem 3.7 indicates that, as long as λ satisfies Assumption 3.2(iii) and ρ_{NT} satisfies Assumption 3.7, we have

$\inf_{1\leq K\leq K_{\max},K\neq K_0} IC(K,\lambda) > IC(K_0,\lambda)$ as $(N,T)\to\infty$. Theoretically, the minimizer of $IC(K,\lambda)$ with respect to K selects the true K_0 with the probability approaching 1 for a variety of choices of λ. In practice, we can further specify the constant multiplier in λ over a finite grid of values to minimize $IC(K^{\lambda},\lambda)$ to search for the number of groups. The simulation section provides more details.

4 Monte Carlo Simulations

In this section, we conduct simulations to evaluate the finite sample performance of the group classification, estimation accuracy, and bubble detection.

4.1 Finite sample performance of the classification and estimation

We consider a panel autoregressive model with unobserved grouped heterogeneous AR coefficients. The observations in each of these DGPs are generated based on the following design. For $i = 1, ..., N$ and $t = 1, ..., T$,

$$y_{i,t} = \mu_i + \beta_i y_{i,t-1} + u_{i,t},$$

where $u_{i,t} \overset{i.i.d}{\sim} (0,1)$, and $y_{i,0} \overset{i.i.d}{\sim} N(0,1)$.

DGP 1: Panel AR with no drift. We consider a panel autoregression with no drift such that $\mu_i = 0$ and the AR coefficients β_i are fixed parameters drawn from three groups with the group-specific number of observations $N_1 : N_2 : N_3 = 0.3 : 0.4 : 0.3$ and the group-specific coefficients $\alpha_k = \{0.5, 1, 1.1\}$.

DGP 2: Panel AR with drifts. We consider a panel autoregression with drifts such that $\mu_i \overset{i.i.d.}{\sim} N(0.2, 1)$ and the AR coefficients β_i are the same as those in DGP 1.

Our simulation studies include two exercises. The first exercise assesses how well the proposed information criterion in (3.2) selects the correct number of groups, as both the classification consistency and asymptotic properties hinge on the correct number of groups. Although, all sequences ρ_{NT} work asymptotically as long as Assumption 3.6 is satisfied. In practice, we use $\rho_{NT} = 1/3 \left(\min_{k=1}^{K_0} \tilde{\Upsilon}_k^2 \right)^{-1/2}$ throughout the simulations. The second exercise assesses the classification consistency and modified C-Lasso and post-Lasso estimation accuracies given the true number of groups. Regarding the modified C-Lasso tuning parameter, we specify $\lambda = c * T^{-1/3}$ throughout the simulations.[1] In each exercise, we consider four combinations of sample sizes with $N = 100, 200$ and $T = 100, 200$ in the "with no drift" cases and $N = 100, 200$ and $T = 200, 400$ in the "with drifts" cases as we apply the LRD transformation and lose half of the time dimensional sample. All simulation results are obtained from 500 replications.

Table 1 reports the probability that a particular number of groups from 1 to 6 is selected according to the IC in (3.2) when the true number of groups is $K_0 = 3$. The results demonstrate the usefulness of the IC for various DGPs and sample sizes. For DGP 1, the correct determination rate is above 90% when $(N, T) = (100, 100)$, and the IC achieves perfect selection of the true group number with a moderately large time dimension when $T = 200$. For DGP 2, when we apply the LRD transformation and use half of the sample for estimation, the correct determination rate is above 90% when $(N, T) = (100, 200)$, and the IC achieves perfect selection with $(N, T) = (200, 400)$.

Table 2 shows the performance of the classification and modified C-Lasso and post-Lasso estimators discussed in Section 2. Column 4 of Table 2 reports the average correct classification rate of the N individuals over all replications. For each replication, the correct classification rate is calculated as $\frac{1}{N} \sum_{k=1}^{K_0} \sum_{i \in \hat{G}_k} \mathbf{1}\{\beta_i^0 = \alpha_k^0\}$. The results imply that the percentage of correct group classification is approaching 100% as (N, T) increase for various DGPs, demonstrating the uniform classification consistency in Theorem 3.2. The remaining columns summarize the performance of the modified C-Lasso and post-Lasso estimators for

[1] The choice of the constant c has only minor effects on the classification consistency and estimation accuracy. Because of space limits, we only report the results for $c = 0.1$. Robust results are available upon request.

each coefficient, including the average root-mean-squared error (RMSE), bias, and coverage rate of the two-sided nominal 95% confidence interval over all replications. The confidence interval is the panel version of the grid bootstrap procedure proposed by Hansen (1999). For comparison, we also report the corresponding statistics of the oracle estimator, which uses the true group membership for estimation. As expected, the oracle estimators typically outperform the modified C-Lasso and post-Lasso estimators, that is, with smaller RMSE and bias and larger coverage rates. In addition, the post-Lasso estimators typically perform better than the C-Lasso estimators. For all DGPs, as (N, T) increase, the modified C-Lasso and post-Lasso estimators' RMSEs and biases decrease and the coverage rates approach 95%. In addition, the estimators of the unit-root and explosive coefficients generally outperform the stationary coefficients, which is consistent with the different group-specific convergence rates discussed in Theorems 3.1–3.3. For the explosive case, both RMSE and bias are negligible and all values are times 1000. In particular, the unit-root and explosive coefficient estimators' performances are generally comparable to that of the oracle estimators.

Table 1: Frequency of selecting $K = 1, ..., 6$ groups when $K_0 = 3$

The table reports the frequency of selecting the correct number of groups for DGPs 1–2, where the true number of groups $K_0 = 3$ and estimated number of groups is obtained by minimizing the information criterion function (3.2).

DGP 1: Panel AR without fixed effects							
N	T	1	2	**3**	4	5	6
100	100	0	0.018	0.942	0.040	0	0
200	100	0	0.034	0.922	0.044	0	0
100	200	0	0	1	0	0	0
200	200	0	0	1	0	0	0
DGP 2: Panel AR with fixed effects (long-run difference transformation)							
N	T	1	2	**3**	4	5	6
100	200	0	0	0.906	0.094	0	0
200	200	0	0	0.914	0.086	0	0
100	400	0	0	0.996	0.004	0	0
200	400	0	0	1	0	0	0

Table 2: Group classification and point estimation for two Lasso-types and oracle estimates

The table reports the percentage of correct group classification (% Correct classification) and the point estimation results in terms of RMSE, bias, and percentage of coverage probability (% Coverage probability) for the 95% confidence intervals for DGPs 1–2 without or with drifts in panel autoregressive models. For the explosive case, all values of RMSE and bias are times 1000. We consider both the C-Lasso estimates based on the objective function in eq.(2.5) and the post-Lasso estimates based on eq.(2.6). The oracle estimates are obtained by eq.(3.1).

(N,T)		% Correct Specification	Stationary group: $\alpha_s = 0.5$ RMSE	Bias	Coverage	Unit-root group: $\alpha_u = 1$ RMSE	Bias	Coverage	Explosive group: $\alpha_e = 1.1$ RMSE $\times 1000$	Bias $\times 1000$	Coverage
DGP 1: Panel AR with no drift											
(100, 100)	C-Lasso	97.17	0.0386	-0.0318		0.0404	-0.0380		2.0146	-1.7487	
	post-Lasso	97.17	0.0233	-0.0122	85.20	0.0092	-0.0042	80.00	0.0293	-0.0101	61.80
	Oracle		0.0160	0.0000	94.40	0.0027	-0.0003	97.80	0.0071	-0.0007	98.60
(200, 100)	C-Lasso	97.40	0.0357	-0.0319		0.0410	-0.0398		4.1694	-3.8027	
	post-Lasso	97.40	0.0189	-0.0120	76.40	0.0060	-0.0036	71.60	0.0239	-0.0077	53.60
	Oracle		0.0119	-0.0004	93.60	0.0021	-0.0002	95.80	0.0049	0.0001	99.20
(100, 200)	C-Lasso	99.87	0.0217	-0.0156		0.0152	-0.0145		2.5582	-2.3197	
	post-Lasso	99.87	0.0125	-0.0012	91.20	0.0015	-0.0003	93.20	0.0000	0.0000	99.20
	Oracle		0.0120	-0.0004	93.20	0.0015	-0.0001	95.80	0.0000	0.0000	100.00
(200, 200)	C-Lasso	99.85	0.0195	-0.0164		0.0184	-0.0180		3.5440	-3.3681	
	post-Lasso	99.85	0.0087	-0.0014	94.60	0.0011	-0.0003	92.40	0.0000	0.0000	100.00
	Oracle		0.0083	-0.0004	95.00	0.0010	-0.0001	95.40	0.0000	0.0000	100.00
DGP 2: Panel AR model with drifts (long-run difference transformation)											
(100, 200)	C-Lasso	98.31	0.0392	-0.0329		0.0080	-0.0069		1.9094	-1.6138	
	post-Lasso	98.31	0.0228	-0.0120	87.60	0.0005	-0.0003	97.00	0.0000	0.0000	78.00
	Oracle		0.0163	-0.0003	94.00	0.0003	-0.0001	99.20	0.0000	0.0000	100.00
(200, 200)	C-Lasso	98.35	0.0367	-0.0330		0.0113	-0.0107		4.3198	-3.9274	
	post-Lasso	98.35	0.0185	-0.0116	82.20	0.0004	-0.0003	89.80	0.0000	0.0000	75.60
	Oracle		0.0123	-0.0004	93.60	0.0002	-0.0001	99.80	0.0000	0.0000	100.00
(100, 400)	C-Lasso	99.78	0.0223	-0.0169		0.0056	-0.0051		2.7584	-2.4098	
	post-Lasso	99.78	0.0124	-0.0016	95.00	0.0001	0.0000	94.80	0.0000	0.0000	94.20
	Oracle		0.0118	-0.0002	96.20	0.0001	0.0000	99.20	0.0000	0.0000	94.20
(200, 400)	C-Lasso	99.80	0.0198	-0.0170		0.0115	-0.0106		5.9453	-5.5825	
	post-Lasso	99.80	0.0087	-0.0013	92.40	0.0001	0.0000	95.40	0.0000	0.0000	91.20
	Oracle		0.0084	0.0000	93.40	0.0001	0.0000	99.40	0.0000	0.0000	91.40

4.2 Bubble Detection

In this subsection, we consider a DGP that is motivated by our empirical study, in which we apply our framework to detect hidden bubbles in the U.S. stock market (Section 5). We detect the unknown firm-specific bubbles when their dividend-adjusted price series exhibit an explosive pattern. Therefore, when the group-specific AR coefficient estimator is greater than 1, those firms are classified into the "bubble group." The following simulations show the classification and estimation performance of the "bubble group" when T is short with respect to N.

DGP 3: Mimicking the empirical set-up. We consider panel data of $N = 1000$ individuals covering $T = 100$ and $T = 200$ periods with fixed effects such that $\mu_i \overset{i.i.d.}{\sim} N(0.05, 0.25)$, $u_{i,t} \overset{i.i.d.}{\sim} N(0, 0.01)$, and $y_{i,0} = 0$. The AR coefficients β_i are fixed parameters drawn from five groups with $N_1 : N_2 : N_3 : N_4 : N_5 = 0.04 : 0.12 : 0.25 : 0.22 : 0.21$ and $\alpha_k^0 = \{0.79, 0.91, 0.96, 0.99, 1.01\}$. The sample sizes, $(N, T) = (1000, 100)$ and $(N, T) = (1000, 200)$, approximate the average dimensions of each rolling window using half-year and yearly historical daily stock prices panels, respectively. The group-specific AR coefficient (α_k^0), as well as group-specific number of individuals (N_k), are calculated as the average estimates across the rolling windows. Compared with DGPs 1 and 2, the ratios of T/N are 0.1 and 0.2, respectively, which are relatively small here in DGP 3, and the parameters of some groups are not separated far apart from each other. Especially, the group-specific AR coefficients α_k^0 for $k = 3, 4$ are very close, and both groups can be regarded as the "local-to-unity" group for $T = 100, 200$.[2] It is hard to identify these groups in this set-up; however, our empirical study focuses on the "bubble group," that is, the firms with explosive autoregressive coefficients. Thus, we show next that, in this specific setting, our method works well in identifying and estimating the bubble group with group-specific explosive autoregressive coefficients.

[2] In our framework, local-to-unity individuals are in the same group as the unit-root ones. Theoretically, we cannot distinguish between group 3 ($\alpha_k^0 = 0.96$) and group 4 ($\alpha_k^0 = 0.99$), which correspond to $1 - \frac{c}{T}$ with $c = 4$ and 1 for $T = 100$, respectively.

Table 3 summarizes the classification estimation results from our modified C-Lasso method discussed in Section 2 and time-series right-tailed sup ADF tests in Phillips, Shi and Yu (2015b). Rows 2 and 11 report the overall correct classification rates of the N individuals classified into a specific group, which is the same measure we report in column 4 of Table 2. The results imply that the percentage for correct group classification among all groups is not very good when our methods fail to distinguish between the two "local-to-unity" groups ($\alpha_k^0 = 0.96, 0.99$). When we combine the two "local-to-unity" groups and focus on how the N individuals are classified into the combined four groups, rows 3 and 12 show that the overall correct classification rates improve from 63% to 85%. Moreover, rows 6–7 and 15–16 summarize the correct and incorrect classification rates of individuals in the "bubble group" over all replications. For each replication, the correct and incorrect bubble detection rates are calculated as $\frac{1}{N_e}\sum_{i\in\hat{G}_e}\mathbf{1}\{\beta_i^0 = 1.01\}$ and $\frac{1}{N-N_e}\sum_{i\in\hat{G}_e}\mathbf{1}\{\beta_i^0 \neq 1.01\}$, respectively. Even though the overall classification results are not ideal, the percentage of correct bubble detection is approaching 100% when $T = 100, 200$. Moreover, the percentage of incorrect bubble detection is quite small. To demonstrate the advantages of pooling cross-section information, we also report the correct and incorrect classification rates obtained from the time-series sup ADF tests. For each time series in our simulated panel, we implement the right-tailed sup ADF tests and use the 95% critical value to determine bubble behaviors. Specifically, if the calculated sup ADF test statistic is large than the 95% critical value, we mark this time series as an explosive one and vice versa. Combining these results, we notice that the classification performance of our panel approach dominates the time-series test in both size and power. Lastly, the RMSE and bias of the explosive AR coefficient, as reported in rows 4–5 and 13–14 in the table, further indicate good finite sample performance of the "bubble group." This simulation results validate our data-driven methodology to detect hidden bubbles in the U.S. stock market.

Table 3: Classification and estimation performance of bubble group

The table compares the percentage of correct (incorrect) bubble detection (% correct (incorrect) bubble detection) between the proposed panel C-Lasso method for heterogeneous AR coefficients and the time-series sup ADF test. For the panel C-Lasso method, the table reports the percentage of correct group classification (% Correct specification and % Correct specification (combine) when two local-to-unit groups are combined)), RMSE, and bias of the explosive AR coefficients.

N=1000, T=100		
Panel C-Lasso method	% Correct specification	63.58
	% Correct specification (combine)	84.07
	RMSE of $\alpha_k^0 = 1.01$	0.0002
	Bias $\alpha_k^0 = 1.01$	-0.0001
	% Correct bubble detection	98.35
	% Incorrect bubble detection	3.15
Time-series sup ADF test	% Correct bubble detection	76.80
	% Incorrect bubble detection	10.10
N=1000, T=200		
Panel C-Lasso method	% Correct specification	63.79
	% Correct specification (combine)	85.47
	RMSE $\alpha_k^0 = 1.01$	0.0001
	Bias $\alpha_k^0 = 1.01$	0.0000
	% Correct bubble detection	99.64
	% Incorrect bubble detection	2.67
Time-series sup ADF test	% Correct bubble detection	96.98
	% Incorrect bubble detection	6.00

5 Empirical Study

Bubble detection is a longstanding question in financial econometrics; here bubbles are modeled by mildly explosive behaviors in prices series. A "benchmark" model in the literature is Phillips, Shi and Yu (2015a) who used several right-tailed sup ADF tests and empirically identified bubble periods in the S&P 500 stock market index. However, the bubble detection algorithms of Phillips, Shi and Yu (2015a) are designed for aggregate time series rather than panels where individual stock price series may exhibit heterogeneous time-series properties in terms of stationarity, nonstationarity, or even explosiveness and where useful information may be lost when applying the time-series tests on a single stock market index series to identify the bubbles.

In this empirical study, we apply our data-driven panel methodology to reinvestigate whether there are bubbles in the stock market by allowing for heterogeneous price behaviors across firms. Our results are more informative in identifying firm-specific unknown bubble behaviors. In particular, we impose latent group structures on the autoregressive coefficients of each stock price series. Our empirical analyses include several stages and answer the following questions: (i) Do bubbles exist in the stock market? If yes, can the latent group structure reflect the hidden bubbles that may exist in some subset and cannot be recognized by the aggregate data using the existing techniques? (ii) What firm-level characteristics can explain the bubble behaviors? In addition, we investigate how the the detected stock market bubbles respond to the aggregate shocks. The detailed procedure and results for this stage are reported in the online appendix.

We consider an extended panel dataset of historical daily adjusted prices of all available common stocks listed on NYSE, AMEX, and NASDAQ. Our data are from the Center for Research in Security Prices (CRSP) in Wharton Research Data Services (WRDS). The sample is an unbalanced panel sampling daily from January 1926 to March 2022.[3] To calculate the adjusted prices, we use the daily close prices and cumulative factor to adjusted prices before then taking the logarithm of all individual stock prices.

5.1 Detecting bubbles in the stock market

We apply our novel methodology in the first stage to detect firm-level hidden bubbles using a rolling window scheme. We explicitly allow the presence of individual effects (μ_i) for each time series. At each month in time, we use data from the most recent three months, which have been long-run differenced using data from the next three months, for the PWLS estimation and repeat this procedure. More precisely, in the first estimation, we use daily

[3]Following the literature, we drop all firms with daily prices of less than 5 dollars. The number of firms differs across time, ranging from 89 to 5587.

data from January 1926 to March 1926, apply the LRD transformation using the data from April 1926 to June 1926, estimate the panel autoregressions of 89 individual time series with a latent group structure, and obtain the group-specific AR coefficients, as well as the unknown group memberships. At each month t, we obtain $\hat{\alpha}_k^{post}$, that is, $\hat{\beta}_i$, for each series i. We claim that a bubble exists in series i at month t, denoted as $B_{i,t} = 1$, if there are explosive coefficients, that is, $\max_k \hat{\alpha}_k^{post}$ is above some thresholds, here following Phillips, Wu and Yu's (2011) guidance on the parameter values of explosive trajectories in practice. All firms with explosive coefficients are classified into a "bubble group." As a robustness check, we obtain results of the "bubble group," which is defined by $\hat{\alpha}^e = \max_k \hat{\alpha}_k^{post} > \{1.005, 1.002, 1\}$, respectively.[4] In addition, we claim that a bubble exists in the market at month t if $B_{i,t} = 1$ for some i.

We further compute the *bubble fraction*, that is, the ratio of the number of firms (N_t^B) in the "bubble group" and total number of the firms (N_t) for estimation at that point in time,

$$\textit{bubble fraction } \pi_t = \frac{\#\text{ firms classified into the bubble group } N_t^B}{\#\text{ firms with classification estimates } N_t} * 100\% \tag{5.1}$$

to measure the fraction of the firms in the stock market classified into the "bubble group." This fraction reflects the spread of bubbles in the stock market at each point in time. The higher *bubble fraction* ratio indicates that more stocks' adjusted price series exhibit explosive behavior over the past three months.

Figure 1 reports the fraction of the firms with bubbles (π_t) and number of the firms with bubbles (N_t^B) at each month in time (in red bars with the y-axis on the left) with different threshold values $\hat{\alpha}^e = \max_k \hat{\alpha}_k^{post} > \{1.005, 1.002, 1\}$, together with the results of Phillips, Shi and Yu's (2015a) bubble detection algorithms applied on the log of the S&P 500 price index for comparison. Phillips, Shi and Yu's (2015a) results are the backward sup ADF (BSADF, hereafter) test sequence of statistics (in blue dashed line with the y-axis on

[4]Phillips, Wu and Yu (2011) point out that, for economic and financial data, typical values of the explosive slope coefficient are in the region $[1.005, 1.05]$.

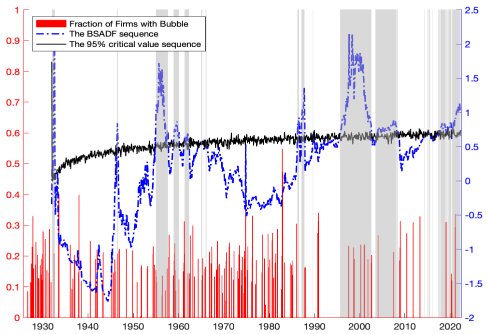
the right) along with its 95% critical values sequences (in black line with the y-axis on the right), and we denote the bubble periods identified by Phillips, Shi and Yu's (2015a) bubble detection algorithms in shaded areas. The red bars imply that bubbles exist in the stock market over time. As we can see from Figure 1, the numbers of firms with bubbles increase over time, while the factions of firms with bubbles fluctuate at around 20%, with the highest fraction reaching 55% around February 1983 for all three threshold values. Compared with the bubble periods identified by Phillips, Shi and Yu's (2015a) method when applied to the aggregate stock market index time series, our method not only detects bubbles in or around the shaded areas, but it also uncovers more bubbles in the panel of firm-level stock prices than the shaded areas. This indicates that hidden bubbles exist in subsets of the firms, which cannot be recognized by the aggregate data when using Phillips, Shi and Yu's (2015a) method. In this sense, our results add to Phillips, Shi and Yu's (2015a) results by uncovering the heterogeneous patterns in bubbles across firms. In addition, the fraction of firms with bubbles reflects the overall severity of bubbles in the stock market, which can be seen as more informative than the bubble indicator provided by the BSADF test.

5.2 What can explain the bubble behavior?

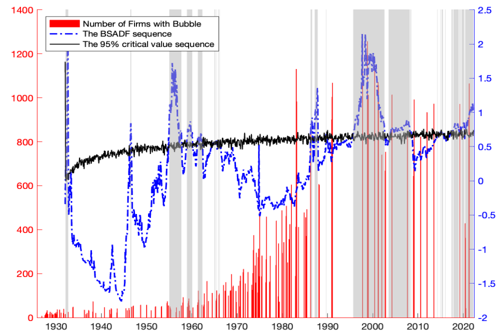
In the second stage, we investigate the potential firm-level explanatory variables of the bubbles. To explore this issue, we first construct a yearly explained variable, *bubble ratio*,

$$\textit{bubble ratio} = \frac{\#\text{ months a stock is classified into the bubble group}}{\#\text{ months with classification estimates}} * 100\% \tag{5.2}$$

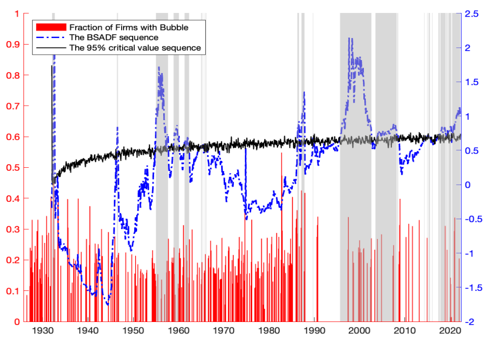
to measure how often a stock is classified into the "bubble group" in a year. The *bubble ratio* measure is the percentage of months in a year that a stock is classified into the "bubble group." A higher *bubble ratio* implies that the adjusted prices of the stocks are more explosive in that year. From theoretical models of Scheinkman and Xiong (2003), we note that asset price bubbles may arise from investors' disagreements on the firm's fundamental information, and



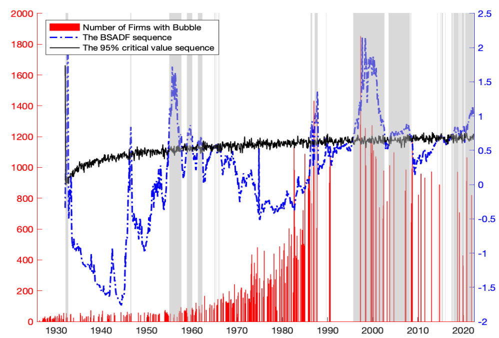
(a) Fraction of firms with bubble, π_t, $\hat{\alpha}^e > 1.005$



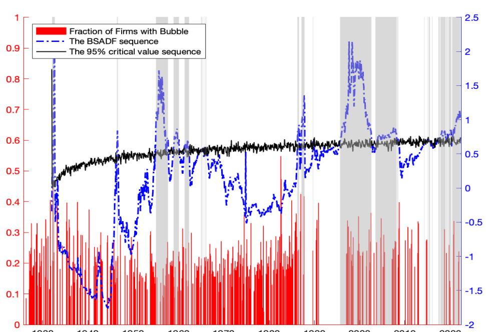
(b) Number of firms with bubble, N_t^B, $\hat{\alpha}^e > 1.005$



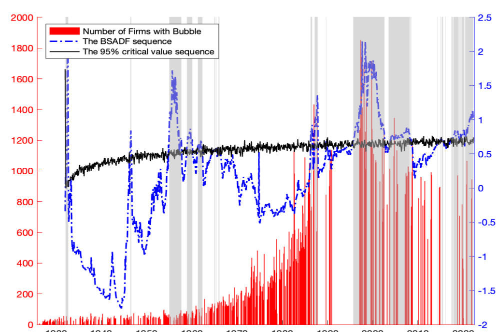
(c) Fraction of firms with bubble, π_t, $\hat{\alpha}^e > 1.002$



(d) Number of firms with bubble, N_t^B, $\hat{\alpha}^e > 1.002$



(e) Fraction of firms with bubble, π_t, $\hat{\alpha}^e > 1$



(f) Number of firms with bubble, N_t^B, $\hat{\alpha}^e > 1$

Note: This figure plots the bubble fraction π_t and number of the firms with bubbles N_t^B defined with $\hat{\alpha}^e = \{1.005, 1.002, 1\}$ across time (in red bars with the y-axis on the left), together with Phillips, Shi and Yu's (2015a) BSADF sequence of statistics (in blue dashed line with the y-axis on the right) and its 95% critical values sequences (in black line with the y-axis on the right) based on the S&P 500 price index.

Figure 1: Bubble detection results with various $\hat{\alpha}^e$

the size of a bubble increases with the degree of the investors' overconfidence and fundamental volatility of the asset. Therefore, we regress *bubble ratio* on a wide array of firm-specific fundamentals, such as firm size, valuation, leverage, profitability, and financial soundness. Considering potential endogeneity issues among asset price bubbles and firm characteristics, we employ the one-year lagged variables for all determinants used in the regression.

Table 4 reports the regression results of *bubble ratio* on a battery of firm-specific characteristics, where the "bubble group" are defined by different threshold values, that is, $\hat{\alpha}^e = \max_k \hat{\alpha}_k^{post} > \{1.005, 1.002, 1\}$. We find that larger firms, i.e., firms with higher fractions of tangible assets and good operating performance, are more often classified into the "bubble group." The sales-to-equity ratio, which measures the efficiency of utilizing shareholders' equity to enhance sale growth, is also positively related to our *bubble ratio*. Moreover, the higher debt ratio is associated with a lower *bubble ratio*, which is consistent with our expectation that firms that borrow more capital from the market to fund their operations are more likely in the case of a business downturn and less likely to be classified into "bubble group." In sum, the results in Table 4 show that our data-driven bubble results are in accordance with the underlying firm-specific fundamentals and consistent with the predictions from the theoretical models. Therefore, our panel autoregressive model with unobserved group structures helps identify the hidden bubbles at the firm level and uncover potential fundamental determinants of those bubbles.

6 Conclusion

We propose a unified approach to estimation and inference for panel autoregressive models with latent group structures in which the degree of persistence in the AR coefficient is unknown. Our methods simultaneously provide the group-specific Lasso-type estimators for the AR coefficients and identify the unknown group structures. We establish the convergence rate, classification consistency, and uniform asymptotic normality for two Lasso-type estima-

Table 4: Multivariate tests of the determinants for bubble group classification

This table presents the panel regression results of the *bubble ratio*, that is, the fraction of months in a year that stocks are grouped into "bubble group," on the explanatory variables: firm size, book-to-market (BM) ratio, sales-to-equity ratio, % tangible assets, leverage, operating profit, debt ratio, and price-to-earnings (P/E). The definitions of these variables are illustrated in the online appendix. The estimated classifications are different from column (1) to column (3) by using different threshold values in defining explosive behavior. *, **, *** denote statistical significance at 10%, 5%, and 1%, respectively, with associated t -statistics in parentheses. The sample period is from 1970 to 2021 because of the availability of firm-level characteristics.

	Bubble ratio		
	(1)	(2)	(3)
	$\hat{\alpha}^e > 1.005$	$\hat{\alpha}^e > 1.002$	$\hat{\alpha}^e > 1$
Size	0.057**	0.210***	0.217***
	(0.023)	(0.030)	(0.037)
BM	-0.023***	-0.012	-0.025**
	(0.009)	(0.010)	(0.012)
Sales/Equity	0.029**	0.034***	0.028*
	(0.014)	(0.013)	(0.014)
% Tangible assets	0.604***	0.657***	0.849***
	(0.179)	(0.239)	(0.286)
Operating profit	0.023***	0.038***	0.045***
	(0.001)	(0.003)	(0.005)
Debt ratio	-0.004***	-0.005***	-0.005
	(0.000)	(0.001)	(0.003)
P/E	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)
Constant	0.611***	0.534***	1.486***
	(0.155)	(0.202)	(0.245)
Firm FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes
Observations	114,818	114,818	114,818
Adj. R^2	0.245	0.3	0.322

tors. The simulation results suggest that our method has good finite-sample performance. An empirical application of detecting hidden bubbles in the U.S. stock market reveals the advantages of this unified estimation procedure for persistent and even explosive data. There are several interesting topics for further research. First, it may be interesting to consider factor structures in the error term, given the extensive literature on modeling cross-sectional dependence and serial correlation using factor structures. Second, it may be interesting to relax the assumption on the distance of coefficients of different groups so that mildly explosive trajectories can be distinguished from explosive ones. Third, our method can be extended to panel vector autoregressive models. We leave these topics for future research.

REFERENCES

Ando, T. and Bai, J. (2016), "Panel data models with grouped factor structure under unknown group membership," *Journal of Applied Econometrics* 31(1), 163-191.

Anundsen, A.K., Gerdrup, K., Hansen, F. and Kragh-Sørensen, K. (2016), "Bubbles and crises: The role of house prices and credit," *Journal of Applied Econometrics* 31(7), 1291–1311.

Basawa, I.V., Mallik, A.K., McCormick, W.P., Reeves, J.H. and Taylor, R.L. (1991), "Bootstrapping unstable first-order autoregressive processes," *The Annals of Statistics* 19(2), 1098–1101.

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

Chan, N.H., Li, D. and Peng, L. (2012), "Toward a unified interval estimation of autoregressions," *Econometric Theory*, 28(3) 705–717.

Diba, B.T. and Grossman, H.I. (1988), "Explosive rational bubbles in stock prices?" *American Economic Review* 78(3), 520–530.

Etienne, X.L., Irwin, S.H. and Garcia, P. (2015), "Price explosiveness, speculation, and grain futures prices," *American Journal of Agricultural Economics* 97(1), 65–87.

Fama, E.F. (1970), "Efficient capital markets: A review of theory and empirical work," *The Journal of Finance* 25(2), 383–417.

Hansen, B.E. (1999), "The grid bootstrap and the autoregressive model," *Review of Economic and Statistics* 81(4), 594–607.

Huang, W., Jin, S., Phillips, P.C. and Su, L. (2021), "Nonstationary panel models with latent group structures and cross-section dependence," *Journal of Econometrics* 221(1), 198-222.

Mikusheva, A. (2007), "Uniform inference in autoregressive models," *Econometrica* 75(5), 1411–1452.

Phillips, P.C. (1987), "Towards a unified asymptotic theory for autoregression," *Biometrika* 74(3), 535–547.

Phillips, P.C. (2014), "On confidence intervals for autoregressive roots and predictive regression," *Econometrica* 82(3), 1177–1195.

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

Phillips, P.C., Moon, H.R. and Xiao, Z. (2001), "How to estimate autoregressive roots near unity," *Econometric Theory* 17(1), 29–69.

Phillips, P.C. and Shi, S.-P. (2018), "Financial bubble implosion and reverse regression," *Econometric Theory* 34(4), 705–753.

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

Phillips, P.C., Shi, S. and Yu, J. (2015b), "Testing for multiple bubbles: Limit theory of realtime detectors," *International Economic Review* 56(4), 1079–1134.

Phillips, P.C., Wu, Y. and Yu, J. (2011), "Explosive behavior in the 1990s nasdaq: When did exuberance escalate asset values?" *International Economic Review* 52(1), 201–226.

Scheinkman, J.A. and Xiong, W. (2003), "Overconfidence and speculative bubbles," *Journal of Political Economy* 111(6), 1183–1220.

So, B.S. and Shin, D.W. (1999), "Cauchy estimators for autoregressive processes with applications to unit root tests and confidence intervals," *Econometric Theory* 15(2), 165–176.

Stock, J.H. (1991), "Confidence intervals for the largest autoregressive root in us macroeconomic time series," *Journal of Monetary Economics* 28(3), 435–459.

Su, L., Shi, Z. and Phillips, P.C. (2016), "Identifying latent structures in panel data," *Econometrica* 84(6), 2215–2264.

Wang W. and Su L. (2021), "Identifying latent group structures in nonlinear panels," *Journal of Econometrics* 220(2), 272–295.

Zhu, F., Cai, Z. and Peng, L. (2014), "Predictive regressions for macroeconomic data," *The Annals of Applied Statistics* 8(1), 577–594.