Manuscript submitted to *The Econometrics Journal*, pp. 1–20.

Detecting common breaks in the means of high dimensional cross-dependent panels

Lajos Horváth[†], Zhenya Liu[‡], Gregory Rice[*], Yuqian Zhao[§]

[†] *Department of Mathematics, University of Utah, Salt Lake City, USA.*
E-mail: horvath@math.utah.edu

[‡] *School of Finance, Renmin University of China, Beijing, China.*
China Financial Policy Research Center, Renmin University of China, Beijing, China.
CERGAM, Aix-Marseille University, 13090 Aix-en-Provence Cedex 02, France.
E-mail: zhenya.liu@ruc.edu.cn

[*] *Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, Canada.*
E-mail: grice@uwaterloo.ca

[‡] *Essex Business School, University of Essex, Colchester, UK*
E-mail: y.zhao@essex.ac.uk

Summary The problem of detecting change points in the mean of high dimensional panel data with potentially strong cross–sectional dependence is considered. Under the assumption that the cross–sectional dependence is captured by an unknown number of common factors, a new CUSUM type statistic is proposed. We derive its asymptotic properties under three scenarios depending on to what extent the common factors are asymptotically dominant. With panel data consisting of N cross sectional time series of length T, the asymptotic results hold under the mild assumption that $\min\{N, T\} \to \infty$, with an otherwise arbitrary relationship between N and T, allowing the results to apply to most panel data examples. Bootstrap procedures are proposed to approximate the sampling distribution of the test statistics. A Monte Carlo simulation study showed that our test outperforms several other existing tests in finite samples in a number of cases, particularly when N is much larger than T. The practical application of the proposed results are demonstrated with real data applications to detecting and estimating change points in the high dimensional FRED-MD macroeconomic data set.

Keywords: *High dimensional panel, Change points in mean, Asymptotic limit, Monte Carlo simulation, FRED-MD macroeconomic data.*

1. INTRODUCTION

Suppose we observe $y_{i,t}$, $1 \le i \le N$, $1 \le t \le T$, which are real valued observations of N financial or economic variables over T time units. Given the fast proliferation and accessibility of economic and financial data available today, usually at least one of N and T is large. This has spurred numerous efforts to develop inferential procedures for such data, and the primary features that distinguish these procedures concern the level of cross–sectional dependence assumed, and the relationship between N and T. The case when $N >> T$ or T is relatively small is commonly referred to as "panel data" in the econometrics literature. Other regimes describing the relationship between N and T generally fall within the scope of high-dimensional multivariate time series analysis.

The methods for the analysis of such data are impacted by how cross–sectional dependence is modeled. A relatively common practice is to assume that the cross–sectional time series are independent/uncorrleated, mainly to reduce some theoretical challenges, although this evidently limits the applicability of the corresponding methods. The most

common way to model dependence between the panels is to use common factors. High–dimensional factor models have experienced a rapid development over the last few decades, providing a theoretically tractable and relatively realistic model for cross–sectional dependence. We refer to Bai and Ng (2008) and Chudik and Pesaran (2015) for excellent surveys on the subject.

Often high–dimensional time series exhibit clear shocks or structural breaks that can be thought of as change points in the data generating mechanism. These are commonly attributable to policy changes, market crashes, or other events that affect all cross–sections simultaneously. Some of the first papers on change point tests in a panel data setting trace back to Joseph and Wolfson (1992, 1993). Bai (2010) followed their work in developing a least–squares type change point test for high dimensional panel data, and he assumed largely that the cross–sectional time series are independent. The absence of cross–sectional dependence is often unrealistic in empirical studies. Kim (2014) conducted several simulation studies that demonstrate that high dimensional change point statistics can provide misleading results if cross–sectional dependence is not adequately accounted for. Horváth and Husková (2012) proposed CUSUM–type tests to detect breaks in the means of high dimensional factor models, although it was shown in their theoretical and empirical analysis that their proposed CUSUM procedures break down if the dependence between the panels is too strong. Other than testing changes in the means, Kim (2011, 2014) and Baltagi et al. (2017) extended the least squares statistics of Bai (2010) to detect breaks in the parameters of linear panels with both serial and cross–sectional dependence.

The theoretical analysis of these procedures is usually carried out under specific conditions on the divergence rates of N and T, and a wide range of such conditions have been considered in the literature. For example, Westerlund (2018) and Antoch et al. (2019) proposed tests and estimators for common breaks in cross–correlated panel data with large N but small T. In finance and macroeconomic studies, the dimension of data can be massively large, e.g. the data examples in Stock and Watson (2012) and McCracken and Ng (2016). Horváth and Husková (2012), under the assumption of weakly correlated panels, allowed both T and N to be large but the condition $\sqrt{N}/T \to 0$ is assumed, which is a necessary condition to establish the asymptotic validity of their test. Related conditions and asymptotic results for the estimator of time of change can be found in Horváth et al. (2017). Furthermore, to test the stability of factor loadings across panels with large N and T, Chen et al. (2014) and Han and Inoue (2015) allow the condition that $\sqrt{T}/N \to 0$, and Baltagi et al. (2017) assume the condition $\log T/N \to 0$.

The purpose of this paper is to develop change point detection and estimation procedures under general dependence conditions between the cross-sections and the relative size of N and T. We propose a new CUSUM–type change point test statistic to detect common change points in the cross–sectional means. Since the behavior of the proposed statistics applied to high-dimensional panel data depend both on the relative divergence rates of N and T as well as the unknown strength of the cross–sectional dependence, we introduce new centering and normalizing methods for this CUSUM process that circumvent the need to know or estimate the underlying dependence structure. By assuming that the cross–sectional dependence can be modeled by a factor model, we derive the asymptotic distribution of the statistic under three scenarios depending on the strength of cross–sectional dependence represented by the common factors requiring only that $\min\{N, T\} \to \infty$. A bootstrap procedure is introduced and used to get critical values for the proposed tests. The finite sample performance of our test, as well as a compari-

son to a number of existing tests, is evaluated through a Monte Carlo simulation study, which shows that the proposed test achieves better size control when compared to other available methods under a wide range of dependence scenarios and relative sizes of N and T. In many cases the proposed test dramatically improves the power as well. Finally, to provide a practical demonstration of our test, we perform a change point analyses of the FRED-MD macroeconomic data set and detect multiple changes over the last two decades in the US.

The rest of paper is structured as the following. In Section 2 we introduce the new test statistic and state our main theoretical results. Extensions of these results to serially correlated errors are discussed in Section 3. In Section 5, we present the results of our Monte Carlo simulation study. Section 6 provides empirical applications and Section 7 concludes. The proofs of all technical results are given in Online supplement.

2. TEST STATISTICS AND MAIN RESULTS

We are interested in detecting changes in the means of high dimensional multivariate (panel) data with observations $y_{i,t}$, which follows the basic model

$$y_{i,t} = \mu_{i,\ell} + \boldsymbol{\lambda}_i^\top \boldsymbol{f}_t + e_{i,t}, \quad k_{\ell-1} < t \leq k_\ell, 1 \leq \ell \leq R+1, \tag{2.1}$$

where $1 \leq i \leq N$, where $1 \leq k_1 < k_2 < \ldots < k_R \leq T$ denote change points, and $\mu_{i,1}, \mu_{i,2}, \ldots, \mu_{i,R}$, are the means of the observations in the ith cross–section. The error terms $\{e_{i,t}, 1 \leq t \leq T\}$ can be thought of as linear or non–linear time series in t for each $1 \leq i \leq N$, in order to accommodate wider financial applications. The cross–sectional dependence is modeled by the common factors $\boldsymbol{f}_t, 1 \leq t \leq T$, $\boldsymbol{f}_t \in \mathbb{R}^p$, and the corresponding loadings $\boldsymbol{\lambda}_i \in \mathbb{R}^p$. Under the null hypothesis, there is no change in the means:

$$\mathcal{H}_0 : \ \mu_{i,1} = \cdots = \mu_{i,R}, \ \ 1 \leq i \leq N.$$

Under the alternative,

$$\mathcal{H}_A : \ \ \mu_{i,j} \neq \mu_{i,j+1}, \ j \in \{1, ..., R\}, \ \text{for some } i \in \{1, ..., N\}.$$

We discuss in Theorem 2.4 conditions on the sizes of the changes $\delta_{i,\ell} = \mu_{i,\ell+1} - \mu_{i,\ell}, 1 \leq \ell \leq R$ that can be detected. Let $\lfloor x \rfloor$ denote the integer part of x. To detect changes in the mean in any particular unit of the cross-section, it is natural to consider that unit's CUSUM process

$$S_i(u) = \sum_{t=1}^{\lfloor uT \rfloor} (y_{i,t} - \bar{y}_{i,T}), \quad \text{where} \quad \bar{y}_{i,T} = \frac{1}{T} \sum_{t=1}^{T} y_{i,t}.$$

Intuitively, $S_i(u)$ will be large in absolute value under $\mathcal{H}_A$ for a cross–section that undergoes a mean change when u is close to k_ℓ/T for some $1 \leq \ell \leq R$. In order to test $\mathcal{H}_0$ against $\mathcal{H}_A$, a natural idea is to aggregate the CUSUM processes across the N cross–sectional units. Horváth and Husková (2012) consider the L^2 aggregated process

$$\mathfrak{H}_{N,T}(u) = \frac{1}{TN^{1/2}} \sum_{i=1}^{N} \left(\frac{1}{\hat{\sigma}_{i,T}^2} S_i^2(u) - \frac{\lfloor Tu \rfloor (T - \lfloor Tu \rfloor)}{T^2} \right), \tag{2.2}$$

where $\hat{\sigma}_{i,T}^2$ is a consistent estimator of the (long–run) variance of each cross–section. By assuming the common factor loadings are zero or are sufficiently quickly decreasing

functions of N and $N^2/T \to 0$, Horváth and Husková (2012) prove that $\mathfrak{H}_{N,T}(u)$ converges weakly to a Gaussian process, which facilitates asymptotically consistent tests of $\mathcal{H}_0$. In the present paper, we wish to establish results of this type under much less restrictive conditions on the relative divergence rate of N and T and the strength of the cross–sectional dependence represented by the common factor loadings.

A clear challenge that arises here is that neither the normalizing term $TN^{1/2}$ nor the centering process $\lfloor Tu \rfloor (T - \lfloor Tu \rfloor)/T^2$ are adequate to define an aggregated CUSUM process with a tractable asymptotic distribution in general under (2.1) when only $\min\{N, T\} \to \infty$ is assumed. In fact, the appropriate centering sequences in general might depend on the properties of the unknown, and in many cases inestimable, common factor loadings and the variance matrix of the common factors.

To overcome these issues, we use random centering and study test statistics based on functionals of

$$V_{N,T}(u) = \sum_{i=1}^{N} \left(S_i^2(u) - \frac{\lfloor uT \rfloor (T - \lfloor uT \rfloor)}{\lfloor \tau T \rfloor (T - \lfloor \tau T \rfloor)} S_i^2(\tau) \right) \mathbb{1}\{uT \geq 1\}, \tag{2.3}$$

where τ is a constant satisfying

Assumption 2.1. $0 < \tau < 1$.

This process is similar to $\mathfrak{H}_{N,T}(u)$ in (2.2), but in this case the centering process is replaced by functionals of the cross–sectional CUSUM processes themselves. The aim of doing this is to adapt the centering to the unknown cross–sectional dependence structure. This idea is similar in spirit to techniques in "self-normalized" or "ratio-type" statistics; see, Halungaa and Osborn (2012) and Shao (2015). The parameter τ acts as a tuning parameter that is used to properly center the process. We study the choice of this parameter in the simulation studies with $T \geq 100$ below, and in summary recommend choosing $\tau = 0.1$ or $\tau = 0.9$, in order to improve the power of the test. For $T < 100$, we suggest to choose $\tau = 0.3$ or $\tau = 0.7$. We comment on the asymptotic behavior of $V_{N,T}(u)$ as a function of u and τ in Remark 2.1.

2.1. Main theoretical asymptotic results

In order to derive the asymptotic properties of the process $V_{N,T}$, we make use of the following basic assumptions. First, we assume that the cross-sectional dependence between the panels is fully explained by the common factors, so that the idiosyncratic errors satisfy:

Assumption 2.2. *$\{e_{i,t}, -\infty < t < \infty\}$ are independent sequences, $1 \leq i \leq N$.*

Assumption 2.3. *For each panel i, $\{e_{i,t}, -\infty < t < \infty\}$ is a stationary sequence with $Ee_{i,t} = 0, Ee_{i,t}e_{i,s} = 0$ for $t \neq s, 1 \leq i \leq N$, and there are ν_0 and $\kappa > 4$ such that $E|\sum_{t=1}^{\ell} e_{i,t}|^{\kappa} \leq \nu_0 \ell^{\kappa/2}$.*

The Prokhorov-Lévy distance between measures and the corresponding processes is denoted by ρ_{PL}. Let

$$s_i(u) = \sum_{t=1}^{\lfloor Tu \rfloor} e_{i,t}, \tag{2.4}$$

and W be the Wiener process (standard Brownian motion). For the definition of the Prokhorov–Lévy measure we refer to Billingsley (1968, p. 238). The next assumption supposes that the partial sums of the errors in each panel satisfy a functional central limit theorem:

ASSUMPTION 2.4. $\lim_{\min(N,T)\to\infty} \max_{1\le i\le N} \rho_{\text{PL}}(T^{-1/2}s_i(\cdot), \sigma_i W(\cdot)) = 0$, *where* $\sigma_i^2 = Ee_{i,t}^2$ *and* $\liminf_{i\to\infty} \sigma_i^2 > 0$.

The fact that $\rho_{PL}(T^{-1/2}s_i(\cdot), \sigma_i W(\cdot)) \to 0$ holds for many stationary time series processes, including ARCH, and symmetric and non–symmetric GARCH processes, see Example 8-9 in Wu (2011). Assumption 2.4 supposes that this convergence is uniform over all cross-sections. For additional discussion on related assumptions, we refer to Carrasco and Chen (2002), Nzé and Doukhan (2004), Doukhan and Wintenberger (2007) and Hörmann (2008). The results in the above papers show that the convergence of the partial sum process of variables following many parametric models is uniform if the parameters are in naturally defined compact set.

Next we define

$$\bar{\sigma}^4 = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \sigma_i^4, \tag{2.5}$$

which we assume exists, and

$$g(u,v) = 2u^2(1-v)^2, \quad 0 \le u \le v \le 1. \tag{2.6}$$

Note that Assumption 2.3 implies that $\bar{\sigma}^4 < \infty$, while Assumption 2.4 yields that $\bar{\sigma}^4 > 0$. We also require the mild assumption that the factor loadings are essentially bounded as $N \to \infty$. Given that $\|\cdot\|$ denotes the Euclidean norm,

ASSUMPTION 2.5. $\boldsymbol{\lambda}_i = \boldsymbol{\lambda}_{i,N}, 1 \le i \le N$ *and* $\limsup_{N\to\infty} \max_{1\le i\le N} \|\boldsymbol{\lambda}_i\| < \infty$.

Let

$$\mathbf{Q} = \lim_{N\to\infty} \left(\sum_{i=1}^{N} \|\boldsymbol{\lambda}_i\|^2 \right)^{-1} \sum_{i=1}^{N} \boldsymbol{\lambda}_i \boldsymbol{\lambda}_i^\top \tag{2.7}$$

and

$$c_* = \lim_{N\to\infty} \frac{1}{N^{1/2}} \sum_{i=1}^{N} \|\boldsymbol{\lambda}_i\|^2, \text{ for } c_* \in [0,\infty]. \tag{2.8}$$

We note that $\mathbf{Q}$ is a finite, symmetric and non–negative definite matrix. The common factors can be arbitrary, perhaps serially correlated, stationary vector valued processes, but we assume that they too satisfy the functional central limit theorem. Let $\mathbf{I}_p$ denote the $p \times p$ identity matrix.

ASSUMPTION 2.6. *$E\mathbf{f}_t = \mathbf{0}$, $E\mathbf{f}_t\mathbf{f}_t^\top = \mathbf{I}_p$ and*

$$T^{-1/2}\sum_{t=1}^{\lfloor Tu\rfloor}\mathbf{f}_t \xrightarrow{\mathcal{D}[0,1]} \mathbf{W}_{\mathbf{\Sigma}}(u),$$

where $\mathbf{W}_{\mathbf{\Sigma}}(u) \in \mathbb{R}^p$ is a Gaussian process with $E\mathbf{W}_{\mathbf{\Sigma}}(u) = \mathbf{0}$ and $E\mathbf{W}_{\mathbf{\Sigma}}(u)\mathbf{W}_{\mathbf{\Sigma}}^\top(v) = \mathbf{\Sigma}\min(u,v)$ and $\mathbf{\Sigma}$ is a non–negative definite matrix.

We note that Assumption 2.6 can be written as the Prokhorov-Lévy distance between the measures generated by the partial sum process and the Gaussian process. The matrix $\mathbf{\Sigma}$ is the long run covariance matrix of the partial sum of the $\mathbf{f}_t$'s.

The following three theorems show that the asymptotic behavior of the process $V_{N,T}$ depends crucially on the constant c_* in (2.8). Three different behaviors are observed depending on if $c_* = 0$, $c_* \in (0,\infty)$, or $c_* = \infty$.

THEOREM 2.1. *If $\mathcal{H}_0$, and Assumptions 2.1–2.6 hold, and $c_* = 0$, then we have that as $\min\{N,T\} \to \infty$,*

$$\frac{1}{TN^{1/2}}V_{N,T}(u) \xrightarrow{\mathcal{D}[0,1]} \Gamma(u),$$

where $\Gamma(u)$ is a Gaussian process with $E\Gamma(u) = 0$ and

$$E\Gamma(u)\Gamma(v) = \bar{\sigma}^4\left(g(u,v) - \frac{u(1-u)}{\tau(1-\tau)}g(\tau,v) - \frac{v(1-v)}{\tau(1-\tau)}g(\tau,u) + \frac{u(1-u)}{\tau(1-\tau)}\frac{v(1-v)}{\tau(1-\tau)}g(\tau,\tau)\right).$$

To state the next theorem we define

$$\mathbf{B}_{\mathbf{\Sigma}}(u) = \mathbf{W}_{\mathbf{\Sigma}}(u) - u\mathbf{W}_{\mathbf{\Sigma}}(1), \quad 0 \le u \le 1, \tag{2.9}$$

to be the generalized p-dimensional Brownian bridge.

THEOREM 2.2. *If $\mathcal{H}_0$, and Assumptions 2.1–2.6 hold, and $c_* = \infty$, then we have that as $\min\{N,T\} \to \infty$,*

$$\frac{1}{T}\left(\sum_{i=1}^N \|\boldsymbol{\lambda}_i\|^2\right)^{-1} V_{N,T}(u) \xrightarrow{\mathcal{D}[0,1]} \operatorname{trace}\left(\mathbf{Q}\left(\mathbf{B}_{\mathbf{\Sigma}}(u)\mathbf{B}_{\mathbf{\Sigma}}^\top(u) - \frac{u(1-u)}{\tau(1-\tau)}\mathbf{B}_{\mathbf{\Sigma}}(\tau)\mathbf{B}_{\mathbf{\Sigma}}^\top(\tau)\right)\right). \tag{2.10}$$

where $\operatorname{trace}(\cdot)$ *denotes the trace of square matrices.*

The difference between Theorems 2.1 and 2.2 arises because of the different "strengths" of the common factors. In Theorem 2.1, the effect of the common factor is asymptotically negligible, and hence the limiting process is Gaussian, while in Theorem 2.2 the common factor dominates and the limit is a quadratic form of an $\mathbb{R}^p$ valued Gaussian process. Using the vec operator (cf. Abadir and Magnus (2005, p. 283)), we get that

$$\begin{aligned}&\operatorname{trace}\left(\mathbf{Q}\left(\mathbf{B}_{\mathbf{\Sigma}}(u)\mathbf{B}_{\mathbf{\Sigma}}^\top(u) - \frac{u(1-u)}{\tau(1-\tau)}\mathbf{B}_{\mathbf{\Sigma}}(\tau)\mathbf{B}_{\mathbf{\Sigma}}^\top(\tau)\right)\right)\\ &\quad = (\operatorname{vec}(\mathbf{Q}))^\top\left(\operatorname{vec}(\mathbf{B}_{\mathbf{\Sigma}}(u)) - \frac{u(1-u)}{\tau(1-\tau)}\operatorname{vec}(\mathbf{B}_{\mathbf{\Sigma}}(\tau))\right).\end{aligned}$$

We also note if $p = 1$, then $\mathbf{Q} = 1$, so in this case

$$\left\{\text{trace}\left(\mathbf{Q}\left(\mathbf{B}_{\mathbf{\Sigma}}(u)\mathbf{B}_{\mathbf{\Sigma}}^{\top}(u) - \frac{u(1-u)}{\tau(1-\tau)}\mathbf{B}_{\mathbf{\Sigma}}(\tau)\mathbf{B}_{\mathbf{\Sigma}}^{\top}(\tau)\right)\right), 0 \le u \le 1\right\}$$
$$\stackrel{\mathcal{D}}{=} \left\{\sigma_{1,f}^2\left(B^2(u) - \frac{u(1-u)}{\tau(1-\tau)}B^2(\tau)\right), 0 \le u \le 1\right\},$$

where $\sigma_{1,f}^2$ is the long run variance of the one dimensional common factors, and B is a Brownian bridge.

Thirdly, we consider the case when the errors and the common factors are of the same order. Since in this case both the error and common factor processes affect the limit, we have to specify their joint behavior:

ASSUMPTION 2.7. *$\{e_{i,t}, 1 \le i \le N, -\infty < t < \infty\}$ and $\{\mathbf{f}_t, -\infty < t < \infty\}$ are independent.*

THEOREM 2.3. *If $\mathcal{H}_0$, and Assumptions 2.1–2.7 hold, and $0 < c_* < \infty$, then we have that as $\min\{N, T\} \to \infty$,*

$$\frac{1}{TN^{1/2}}V_{N,T}(u) \stackrel{\mathcal{D}[0,1]}{\longrightarrow} \Gamma(u) + c_*\text{trace}\left(\mathbf{Q}\left(\mathbf{B}_{\mathbf{\Sigma}}(u)\mathbf{B}_{\mathbf{\Sigma}}^{\top}(u) - \frac{u(1-u)}{\tau(1-\tau)}\mathbf{B}_{\mathbf{\Sigma}}(\tau)\mathbf{B}_{\mathbf{\Sigma}}^{\top}(\tau)\right)\right),$$

where $\{\Gamma(u), 0 \le u \le 1\}$ and $\{\mathbf{B}_{\mathbf{\Sigma}}(u), 0 \le u \le 1\}$ are independent Gaussian processes defined in Theorem 2.1 and (2.9), respectively.

Based on these results, the behavior of $V_{N,T}$ and its functionals depend on the magnitude of the unknown common factors measured by c_*.

REMARK 2.1. The simulation study in Section 5 shows that a fixed τ often provides good power under the alternative. One may also consider $V_{N,T}(u)$ as a two parameter process $V_{N,T}(u, \tau)$ of $(t, \tau) \in [0, 1] \times [\delta, 1 - \delta]$, $0 < \delta < 1/2$. With some modifications of our proofs one can show that under $\mathcal{H}_0$ the suitably normalized $V_{N,T}(u, \tau)$ converges weakly in $\mathcal{D}([0, 1] \times [\delta, 1 - \delta])$, $0 < \delta < 1/2$, to a two parameter Gaussian process under the assumptions of Theorems 2.1–2.3.

Lastly, we provide a discussion of the behavior of supremum functionals of $V_{N,T}(u)$ under the change point alternative $\mathcal{H}_A$. Consistency of tests based on other functionals, like the square integral, can be discussed similarly. We consider the case when the times of the changes are proportional to the number of the observations in the panels:

ASSUMPTION 2.8. *$k_\ell = \lfloor T\theta_\ell \rfloor, 1 \le \ell \le R$ with some $0 < \theta_1 < \theta_2 \ldots < \theta_R < 1$.*

The drift function of the ith panel is

$$\bar{v}_i(u) = v_i^2(u) - \frac{u(1-u)}{\tau(1-\tau)}v_i^2(\tau),$$

where

$$v_i(u) = \sum_{m=1}^{\ell-1}\mu_{i,m}(\theta_m - \theta_{m-1}) + (u - \theta_{\ell-1})\mu_{i,\ell} - u\sum_{m=1}^{R+1}\mu_{i,m}(\theta_m - \theta_{m-1}), \quad 1 \le \ell \le R+1.$$

Also, we let

$$Q_N = \sup_{0\le u\le 1}\left|\sum_{i=1}^{N}\bar{v}_i(u)\right|.$$

THEOREM 2.4. *We assume that Assumptions 2.1–2.8 hold.*
(i) If $0 \le c_* < \infty$ *and*

$$\frac{TQ_N}{N^{1/2}} \to \infty,$$

then

$$\frac{1}{TN^{1/2}}\sup_{0\le u\le 1}|V_{N,T}(u)| \xrightarrow{P} \infty. \tag{2.11}$$

(ii) If $c_* = \infty$ *and*

$$\frac{TQ_N}{\sum_{i=1}^{N}\|\boldsymbol{\lambda}_i\|^2} \to \infty,$$

then

$$\frac{1}{T}\left(\sum_{i=1}^{N}\|\boldsymbol{\lambda}_i\|^2\right)^{-1}\sup_{0\le u\le 1}|V_{N,T}(u)| \xrightarrow{P} \infty. \tag{2.12}$$

Simple conditions can be given when $R = 1$. The next corollary illustrates the role of the change magnitude and the factor loadings in the consistency. We recall (2.1) that $\delta_i = \delta_{i,1}$ denotes the size of the change in the mean in the ith cross section.

COROLLARY 2.1. *We assume that Assumptions 2.1–2.8, and* $R = 1$ *are satisfied.*
(i) If $0 \le c_* < \infty$ *and*

$$\frac{T}{N^{1/2}}\sum_{i=1}^{N}\delta_i^2 \to \infty,$$

then (2.11) *holds.*
(ii) If $c_* = \infty$ *and*

$$\frac{T\sum_{i=1}^{N}\delta_i^2}{\sum_{i=1}^{N}\|\boldsymbol{\lambda}_i\|^2} \to \infty,$$

then (2.12) *holds.*

We note that if the change magnitudes δ_i are bounded away from zero, i.e., $\liminf_{N\to\infty}\min_{1\le i\le N}|\delta_i| > 0$, along with the condition $\min\{N,T\} \to \infty$, then Corollary 2.1 implies that the normalized supremum functional of $V_{N,T}$ diverges to infinity in probability under $\mathcal{H}_A$.

2.2. Normalization of the test statistics

Theorems 2.1–2.3 provide a full description of the asymptotic properties of the aggregated CUSUM process $V_{N,T}(u)$. However, the normalization and the form of the limiting process depend on the rate of divergence of N and T, as well as the unknown factor loadings measured by c_*. In this subsection, we describe the appropriate order of the normalizing sequence that are consistent regardless of the value of c_*. This makes it possible to devise a bootstrapping scheme for obtaining critical values of functionals of $V_{N,T}$. Let

$$z_{i,t} = (y_{i,t} - \bar{y}_{i,T})^2 - \frac{1}{T}\sum_{s=1}^{T}(y_{i,s} - \bar{y}_{i,T})^2,$$

and define the corresponding empirical cross covariances as

$$\hat{\gamma}_{i,j}(h) = \frac{1}{T-h}\sum_{s=1}^{T-h} z_{i,s}z_{j,s+h}.$$

Our estimator for the normalizing sequence is

$$w_{N,T} = \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{h=-H}^{H} K\left(\frac{h}{T}\right)\hat{\gamma}_{i,j}(h),$$

where the kernel function $K(\cdot)$ and the bandwidth parameter H satisfy standard conditions on the kernel/lag window assumed on spectral density estimators in the literature:

ASSUMPTION 2.9. *(i) $K(0) = 1$, (ii) there is a constant $b > 0$ such that $K(u) = 0$ if $|u| \geq b$, (iii) K is Lipschitz continuous on $[-b, b]$.*

ASSUMPTION 2.10. *$H = H(T)$, and $1/H + H/T^{1/2} \to 0$ as $T \to \infty$.*

Then, we define

$$\mathbf{R}_t = \mathrm{vec}\left(\mathbf{f}_t\mathbf{f}_t^\top - \mathbf{I}_p\right).$$

Due to Assumption 2.6 we have that $E\mathbf{R}_t = \mathbf{0}$. We assume that the auto-correlation matrix of the $\mathbf{R}_t$'s decays quickly as a function of the lag parameter:

ASSUMPTION 2.11. *$\{\mathbf{f}_t, -\infty < t < \infty\}$ is a stationary sequence, $E\mathbf{f}_0 = \mathbf{0}, E\|\mathbf{f}_0\|^4 < \infty$ and there are $\nu_0 > 0$ and $\alpha_0 > 2$ such that*

$$\left\|E\mathbf{R}_0\mathbf{R}_h^\top\right\| \leq \nu_0(|h|+1)^{-\alpha_0}.$$

We further define

$$\mathfrak{R} = \sum_{h=-\infty}^{\infty} E\mathbf{R}_0\mathbf{R}_h^\top.$$

Similarly we need assumptions on the behavior of the second moments of the $e_{i,t}$'s:

ASSUMPTION 2.12. *$Ee_{i,0} = 0, E|e_{i,0}|^4 < \infty$ and there are ν_0 and $\alpha_0 > 6$ such that*

$$|\mathrm{cov}(e_{i,0}^2, e_{i,h}^2)| \leq \nu_0(|h|+1)^{-\alpha_0}, \;\; 1 \leq i \leq N.$$

The following results also involve

$$\mathfrak{s}_i^2 = \sum_{h=-\infty}^{\infty} \mathrm{cov}(e_{i,0}^2, e_{i,h}^2).$$

THEOREM 2.5. *We assume that $\mathcal{H}_0$ and Assumptions 2.2–2.12 hold.*
(i) If $c_* = 0$ then we have that

$$\frac{w_{N,T}}{NT} \xrightarrow{P} a_0, \quad \textit{where } a_0 = \lim_{N\to\infty} \sum_{i=1}^{N} \mathfrak{s}_i^2.$$

(ii) If $0 < c_* < \infty$ then we have that

$$\frac{w_{N,T}}{NT} \xrightarrow{P} a_0 + c_* \mathbf{q}^\top \mathfrak{R}\mathbf{q}, \quad \textit{where } \mathbf{q} = \lim_{N\to\infty} \left(\sum_{j=1}^{N} \|\boldsymbol{\lambda}_j\|^2 \right)^{-1} \sum_{i=1}^{N} \mathrm{vec}(\boldsymbol{\lambda}_i \boldsymbol{\lambda}_i^\top).$$

(iii)If $c_* = \infty$ then we have that

$$\frac{w_{N,T}}{T\left(\sum_{i=1}^{N} \|\boldsymbol{\lambda}_i\|^2 \right)^2} \xrightarrow{P} \mathbf{q}^\top \mathfrak{R}\mathbf{q}.$$

The next result is an immediate consequence of Theorems 2.1–2.3 and 2.5.

COROLLARY 2.2. *If $\mathcal{H}_0$ and Assumptions 2.1–2.5 hold, then we have that*

$$\frac{1}{(T w_{N,T})^{1/2}} V_{N,T}(u) \xrightarrow{\mathcal{D}[0,1]} \Delta(u),$$

where

$$\Delta(u) = \begin{cases} a_0^{-1/2} \Gamma(u), \ \textit{if } c_* = 0 \\ (a_0 + c_* \mathbf{q}^\top \mathfrak{R}\mathbf{q})^{-1/2} \Big[\Gamma(u) \\ \quad + c_* \mathrm{trace}\left(\mathbf{Q} \left(\mathbf{B}_{\boldsymbol{\Sigma}}(u) \mathbf{B}_{\boldsymbol{\Sigma}}^\top(u) - \frac{u(1-u)}{\tau(1-\tau)} \mathbf{B}_{\boldsymbol{\Sigma}}(\tau) \mathbf{B}_{\boldsymbol{\Sigma}}^\top(\tau) \right) \right) \Big], \ \textit{if } 0 < c_* < \infty \\ (\mathbf{q}^\top \mathfrak{R}\mathbf{q})^{-1/2} \mathrm{trace} \left(\mathbf{Q} \left(\mathbf{B}_{\boldsymbol{\Sigma}}(u) \mathbf{B}_{\boldsymbol{\Sigma}}^\top(u) - \frac{u(1-u)}{\tau(1-\tau)} \mathbf{B}_{\boldsymbol{\Sigma}}(\tau) \mathbf{B}_{\boldsymbol{\Sigma}}^\top(\tau) \right) \right), \ \textit{if } c_* = \infty, \end{cases}$$

and the processes Γ and $\mathbf{B}_{\boldsymbol{\Sigma}}(u)$ are defined in Theorem 2.1 and (2.9), respectively.

The normalizing sequence in Corollary 2.2 works well under $\mathcal{H}_0$ but without modification tends to overestimate under $\mathcal{H}_A$, thereby reducing the power of the tests. This is due to fact that centering each series by $\bar{y}_{i,T}$ tends to overestimate the variances of the idiosyncratic errors under $\mathcal{H}_A$; we refer to Vogelsang (1999) for a discussion on this issue. In order to mitigate this problem, when computing $w_{N,T}$ we center each cross section by taking into account a potential change point in the mean. To do this, in each cross section we estimate a potential change point using a standard CUSUM estimator (Csörgő and Horváth, 1997). We then calculate the normalization term $\bar{w}_{N,T}$ after recentering

each cross section taking into account this potential change point. A similar procedure is employed in this setting in Cho (2016).

Under the conditions of Corollary 2.2 we have

$$\frac{1}{(T\bar{w}_{N,T})^{1/2}} V_{N,T}(u) \quad \stackrel{\mathcal{D}[0,1]}{\longrightarrow} \quad \Delta(u) \tag{2.13}$$

under $\mathcal{H}_0$ and

$$\frac{1}{(T\bar{w}_{N,T})^{1/2}} \sup_{0\le u\le 1} |V_{N,T}(u)| \quad \stackrel{P}{\to} \quad \infty \tag{2.14}$$

under $\mathcal{H}_A$, if the assumptions of Corollary 2.1 are satisfied. We then consider the detector statistic

$$v_{N,T} = \frac{1}{(T\bar{w}_{N,T})^{1/2}} \sup_{0\le u\le 1} |V_{N,T}(u)|. \tag{2.15}$$

Based on this process we can estimate the change point with

$$\hat{k}_V^* = \lfloor T\hat{\theta}_{N,T} \rfloor, \tag{2.16}$$

where $\hat{\theta}_{N,T} = \min\left\{u : \ |V_{N,T}(u)| = \sup_{0\le v\le 1} |V_{N,T}(v)|\right\}$. If multiple changes are thought to be present in the data, we apply standard binary segmentation to estimate all change points (see, Csörgő and Horváth, 1997). The distribution of the estimator of the time of change in high-dimensional panel data is discussed in Horváth et al. (2017).

Remark 2.2. Our tests are tailored to find changes in the means of panels during the observation period. If the loadings change within the sample but the means remain the same, it may be shown that the null hypothesis will not be rejected with probability tending to one as long as

$$\limsup_{T\to\infty} \max_{1\le i\le N} \max_{1\le t\le T} \|\boldsymbol{\lambda}_{i,t}\| < \infty,$$

where $\boldsymbol{\lambda}_{i,t}$ is the loading in the ith cross-section at time t. We refer to Breitung and Eickmeier (2017) for tests on detecting changes in the number of common factors and loadings. Similarly, if the variances of $e_{i,t}$ change, $\mathcal{H}_0$ will not be rejected with probability tending to one if there are no changes in the means. In this sense this procedure is robust to potential heterogeneity of the error variance or factor loadings.

3. EXTENSION TO CORRELATED ERRORS

In Section 2.1, we investigated panels where the idiosyncratic errors are serially uncorrelated (cf. Assumption 2.3), which covers errors generated from many commonly used volatility processes. In this section, we replace Assumption 2.3 with the following more general assumption that allows for serial correlation.

Assumption 3.1. *for each i, $\{e_{i,t}, -\infty < t < \infty\}$ is a stationary sequence. Also, $Ee_{i,t} = 0$, and there are constants $\nu_0, \bar{\nu}_0, \kappa > 4, \beta > 3$ such that $E|\sum_{t=1}^{\ell} e_{i,t}|^{\kappa} \le \nu_0 \ell^{\kappa/2}$ and $|Ee_{i,0}e_{i,h}| \le \bar{\nu}_0(|h|+1)^{-\beta}$.*

Under Assumption 3.1, the long run variances of the sums of the $e_{i,t}$'s are well defined and we denote them as $\omega_i^2 = \sum_{h=-\infty}^{\infty} Ee_{i,0}e_{i,h}$. Since ω_i^2 represents the limiting variance of the partial sums of the errors, we need to modify Assumption 2.4 as

ASSUMPTION 3.2. $\lim_{\min(N,T)\to\infty} \max_{1\le i\le N} \rho_{\mathrm{PL}}(T^{-1/2}s_i(\cdot), \omega_i W(\cdot)) = 0$ *and* $\liminf_{i\to\infty} \omega_i^2 > 0$, *where* W *denotes the standard Wiener measure.*

Similarly to $\bar{\sigma}^4$ of (2.5), let

$$\bar{\omega}^2 = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \omega_i^4. \tag{3.17}$$

THEOREM 3.1. *If* $\mathcal{H}_0$*, Assumptions 2.1, 2.3–2.6, 3.1–3.2 hold,* $c_* = 0$ *and*

$$\frac{N}{T^2} \to 0, \tag{3.18}$$

then we have that $\frac{1}{TN^{1/2}} V_{N,T}(u) \overset{\mathcal{D}[0,1]}{\longrightarrow} \Gamma^*(u)$*, where* $\Gamma^*(u)$ *is a Gaussian process with* $E\Gamma^*(u) = 0$ *and*

$$E\Gamma^*(u)\Gamma^*(v) = \bar{\omega}^2 \left(g(u,v) - \frac{u(1-u)}{\tau(1-\tau)} g(\tau,v) - \frac{v(1-v)}{\tau(1-\tau)} g(\tau,u) + \frac{u(1-u)}{\tau(1-\tau)} \frac{v(1-v)}{\tau(1-\tau)} g(\tau,\tau) \right),$$

where $g(u,v)$ *and* $\bar{\omega}^2$ *are defined in* (2.6) *and* (3.17)*, respectively.*

THEOREM 3.2. *If* $\mathcal{H}_0$*, Assumptions 2.1, 2.3–2.6, 3.1–3.2 hold,* $c_* = \infty$ *and*

$$\frac{N^{1/2}}{T \sum_{i=1}^{N} \|\boldsymbol{\lambda}_i\|^2} \to 0, \tag{3.19}$$

then Theorem 2.2 remains true.

THEOREM 3.3. *If* $\mathcal{H}_0$*, Assumptions 2.1, 2.3–2.7, 3.1–3.2, and condition* (3.18) *hold,* $0 < c_* < \infty$*, then we have that*

$$\frac{1}{TN^{1/2}} V_{N,T}(u) \overset{\mathcal{D}[0,1]}{\longrightarrow} \Gamma^*(u) + c_* \mathrm{trace} \left(\mathbf{Q} \left(\mathbf{B}_{\boldsymbol{\Sigma}}(u) \mathbf{B}_{\boldsymbol{\Sigma}}^{\top}(u) - \frac{u(1-u)}{\tau(1-\tau)} \mathbf{B}_{\boldsymbol{\Sigma}}(\tau) \mathbf{B}_{\boldsymbol{\Sigma}}^{\top}(\tau) \right) \right),$$

where $\{\Gamma^*(u), 0 \le u \le 1\}$ *and* $\{\mathbf{B}_{\boldsymbol{\Sigma}}(u), 0 \le u \le 1\}$ *are independent Gaussian processes defined in Theorem 3.1 and* (2.9)*, respectively.*

Furthermore, we may also consider applying weight functions to the proposed test statistic to improve the power of the test against changes that occur early or late in the sample. We provide detail in this direction in Online supplement.

4. BOOTSTRAP FOR THE CRITICAL VALUES

In this section, we provide some details regarding the implementation of the proposed tests. Following the results in Section 2.2, we reject $\mathcal{H}_0$ for large values of the detector $v_{N,T}$ defined in (2.15). For a specified significance level α, we define the asymptotic critical value $\mathfrak{c}(\alpha)$ such that

$$P\left\{ \sup_{0\le u\le 1} |\Delta(u)| > \mathfrak{c}(\alpha) \right\} = \alpha, \tag{4.20}$$

where the process $\Delta(u)$ is defined in Corollary 2.2. Direct computation of the critical value $\mathfrak{c}(\alpha)$ is difficult since the value of c_*, as well as the long run variances and factor loadings are unknown. As such, we propose bootstrap procedures to determine the empirical distribution of $v_{N,T}$ and obtain approximate critical values from the sample.

We consider two new semi–parametric bootstrap approaches. Some similar procedures have been developed in the literature to date. For example, Cho (2016) proposed a semi–parametric method based on fitting a generalized dynamic factor model (GDFM), and produced bootstrap samples by simulating from this model. Jirák (2015) considered a wide range of bootstrap procedures with perhaps the most effective among them to simply simulate independent and identically distributed (IID) standard Gaussian random variables in each cross section for producing a bootstrap sample of the test statistic. We generally found their methods are not suitable in our set up due to the more general serially dependent processes considered.

Our approaches are tailored to accommodate serial dependence in (2.1). The first method is a factor model based bootstrap, similar in spirit to the GDFM bootstrap of Cho (2016), but adapted to the assumptions of Theorems 2.1–2.3. It is detailed in Algorithm 1, which aims to produce bootstrap samples that share the same relevant features affecting the limiting distribution of $v_{N,T}$.

Following this, and inspired by Jirák (2015), we also perform a bootstrap to approximate the limit distribution by simulating panels with idiosyncratic errors following an IID standard Gaussian distribution after controlling for cross–sectional dependence using a common factor model. The second method is described as Algorithm 2.

Algorithm 1: Bootstrap I

Input: $y_{i,t}$, $1 \leq i \leq N$, $1 \leq t \leq T$.

Output: estimator $\mathfrak{c}_{N,T}(\alpha)$, $0 < \alpha < 1$, the critical value of $v_{N,T}$.

Step 1. estimate p, the number of common factors by using the information criteria proposed by Bai and Ng (2002), obtain the estimators $\hat{\boldsymbol{\lambda}}_i$ and $\hat{\boldsymbol{f}}_t$ of the loadings and the common factors;

Step 2. estimate $\bar{\sigma}$ of 2.5, by taking the square root of the averaged squares of long–run variances of cross–sectional residuals $y_{i,t} - \hat{\boldsymbol{\lambda}}_i^\top \hat{\boldsymbol{f}}_t$;

Step 3. simulate a $T \times p$ matrix $\tilde{\boldsymbol{f}}_t$ from the multivariate normal distribution having the same long–run covariance structure as estimated by $\hat{\boldsymbol{f}}_t$;

Step 4. simulate sequences $\tilde{e}_{i,t}$, $1 \leq i \leq N$, and $1 \leq t \leq T$, that follow an IID normal distribution with zero mean and standard deviation $\bar{\sigma}$;

Step 5. construct a bootstrap sample with observations $B_{i,t}^{(1)} = \hat{\boldsymbol{\lambda}}_i^\top \tilde{\boldsymbol{f}}_t + \tilde{e}_{i,t}$. Repeat Steps 4–5 a large number of times, we used 10,000 below. Based on these bootstrap samples compute the empirical distribution of the bootstrap versions of $v_{N,T}$. Set $\mathfrak{c}_{N,T}(\alpha)$ to be the critical value at significance level α.

Establishing the asymptotic validity of these bootstrap approaches is beyond the scope of the paper, although we emphasize that this is an important issue worthy of further study. The simulation results in Section 5 confirm that both bootstrap methods perform reasonably well under a wide variety of settings. We also tried to use traditional block and permutation bootstraps, which are also considered in Jirák (2015). However, the

Algorithm 2: Bootstrap II

Input: $y_{i,t}$, $1 \le i \le N$, $1 \le t \le T$.
Output: estimator $\mathfrak{c}_{N,T}(\alpha)$, $0 < \alpha < 1$, the critical value of $v_{N,T}$.
Step 1. estimate p, the number of common factors by using the information criteria proposed by Bai and Ng (2002), obtain the estimators of the loadings $\hat{\boldsymbol{\lambda}}_i$ and the common factors $\hat{\boldsymbol{f}}_t$;
Step 2. obtain $\tilde{\boldsymbol{f}}_t$ by block–bootstrapping $\hat{\boldsymbol{f}}_t$ with block length $\lfloor \log T \rfloor$;
Step 3. simulate IID sequences $\tilde{e}_{i,t}$ from the standard normal distribution;
Step 4. Construct a bootstrap sample with observations $B_{i,t}^{(2)} = \hat{\boldsymbol{\lambda}}_i^\top \tilde{\boldsymbol{f}}_t + \tilde{e}_{i,t}$; repeat Steps 2–4 a large number of times. Based on these bootstrap samples compute the empirical distribution of the bootstrap versions of $v_{N,T}$. Set $\mathfrak{c}_{N,T}(\alpha)$ to be the critical value at significance level α.

results indicated the resulting test are quite under–sized when N is large. Similar poor performance was also observed when using sub–sample bootstrap approaches. These results are omitted to save space.

5. A MONTE CARLO SIMULATION STUDY

We now present the results of a Monte Carlo simulation study that aimed to evaluate the finite sample properties of the proposed statistic and bootstrapping procedures, and compare the proposed testing procedure to others available in the literature. In particular, we compared our results to the test statistics and bootstrap procedures in Horváth and Hušková (2012), Jirák (2015), and Cho (2016).

In order to generate panel data following the model in (2.1), we considered two data generating processes (DGPs) for the error sequence. Let $v_{i,t}$, $1 \le i \le N$, $1 \le t \le T$, denote IID standard normal random variables, and we gained

1 (**ARMA**) $e_{i,t}$ follows a stationary ARMA(1,1) process $e_{i,t} = 0.2e_{i,t-1} - 0.3e_{i,t-2} - 0.1v_{i,t} + 0.2v_{i,t-1}$.

2 (**GARCH**) $e_{i,t}$ follows a stationary GARCH(1,1) process $e_{i,t} = \sigma_{i,t}v_{i,t}$, and $\sigma_{i,t}^2 = 0.05 + 0.01e_{i,t-1}^2 + 0.9\sigma_{i,t-1}^2$.

We assume that there is at most one change, i.e. $R = 1$. The size of the change in the means is $\delta = \delta_i, 1 \le i \le N$. In each case the error series of length T are generated after discarding a burn–in sample of length 200. These errors are combined with common factors to produce panels according to three different schemes with varying levels of cross–sectional dependence:

1 (**IND**) Cross-sectionally independent panels

$$y_{i,t} = 0.1 + \delta \mathbb{1}\{t > \lfloor T/2 \rfloor\} + e_{i,t},\ 1 \le i \le N, 1 \le t \le T;$$

2 (**CF–W–λ**) Common factor model with weak cross sectional dependence

$$y_{i,t} = 0.1 + \delta \mathbb{1}\{t > \lfloor T/2 \rfloor\} + \lambda_i \mathbb{1}\{i \in \mathcal{J}\} f_t + e_{i,t},\ 1 \le i \le N, 1 \le t \le T,$$

where $\mathcal{J}$ is a random subset of $\{1, 2, \ldots, N\}$ of size $\lfloor\sqrt{N}\rfloor$. The common factor f_t is taken to be an IID standard normal sequence.

3 (**CF**–**S**–λ) Common factor model with strong cross sectional dependence

$$y_{i,t} = 0.1 + \delta \mathbb{1}\{t > \lfloor T/2 \rfloor\} + \lambda_i f_t + e_{i,t},\ 1 \leq i \leq N, 1 \leq t \leq T.$$

The common factor f_t is again taken to be an IID standard normal sequence, and the factor loadings $\lambda_i, 1 \leq i \leq N$, are all nonzero.

The size of change $\delta = 0$ under $\mathcal{H}_0$. Under $\mathcal{H}_A$, we consider small and large changes, for choosing $\delta = 0.15$ and 0.3, respectively. The cross–sectional dependence is determined by the coefficient $\lambda_i, 1 \leq i \leq N$, selected from $\{0.25, 0.5, 1.0\}$. For each setting of these parameters and dimensions, we generate the panels independently 400 times, and carry out tests of $\mathcal{H}_0$ of size $\alpha = 0.05$. The critical values of the test statistic are evaluated from 200 bootstraps. We considered in limited situations a larger number (10,000) of bootstraps, and found that the basic patterns observed in the simulations were the same, and so we used 200 bootstraps in order to reduce the significant computational time involved in carrying out the bootstrap in Algorithm 1.

In total, we compared six tests. The first two are based on the statistic v_{NT} in (2.15) with the tunning parameter $\tau = 0.1$, and the critical values are computed from bootstraps in Algorithm 1 and Algorithm 2. The bandwidth H used to define the normalizing sequence and estimate the long run covariance of the common factors in the bootstrap Algorithm 1 was $H = \lfloor \log T \rfloor$, and the kernel used was the flat–top kernel. We label these tests as $\mathcal{V}^{(1)}_{N,T}$ and $\mathcal{V}^{(2)}_{N,T}$, respectively. In unreported simulations we tried numerous values of the parameter τ, including $\tau = 0.05, 0.1, 0.25, 0.75, 0.9$, and 0.95. We generally found that the empirical sizes of the tests with varying values of τ were relatively similar, but the power was in general reduced if τ was close to the change point, which encourages taking τ to be relatively close to either 0 or 1 to avoid this.

The third and fourth tests are modified CUSUM statistics as proposed in Horváth and Hušková (2012). As discussed in their paper, the original test statistic they proposed is not suitable for change point detection when the cross–sections are highly correlated and/or $N >> T$. We therefore consider a modified version of their process

$$\mathfrak{H}^{mod}_{N,T}(u) = \frac{1}{(T\bar{w}_{N,T})^{1/2}} \sum_{i=1}^{N} \left(S_i^2(u) - \frac{\lfloor Tu \rfloor (T - \lfloor Tu \rfloor)}{T^2} \right),$$

and the statistic as well as change point estimator are

$$\mathfrak{H}^{mod}_{N,T} = \sup_{0 \leq u \leq 1} |\mathfrak{H}^{mod}_{N,T}(u)|,\ \ \hat{k}_{\mathfrak{H}} = \lfloor T\hat{\theta}_{\mathfrak{H}} \rfloor,$$

where $\hat{\theta}_{\mathfrak{H}}$ is the location of the maximum of $|\mathfrak{H}^{mod}_{N,T}(u)|$. Tests based on $\mathfrak{H}^{mod}_{N,T}$ are then carried out at a specified size using Algorithms 1 and 2. The resulting tests are denoted by $\mathfrak{H}^{(1)}_{N,T}$ and $\mathfrak{H}^{(2)}_{N,T}$. Additionally, we also considered Jirák's (2015) test using the recommended parametric bootstrap method that was shown to work well for large N, and we denote this test as $\mathfrak{J}_{N,T}$. A comparison was also conducted to Cho's (2016) double CUSUM test, carried out using the GDFM bootstrap method, denoted as $\mathfrak{C}_{N,T}$.

To evaluate the performances of competing tests, we compared their empirical size, power, and the accuracy of their corresponding change point estimators for various values of N and T. In particular, we considered six combinations between T and N that are

Table 1. Rejection rates under $\mathcal{H}_0$ at 5% significance level when the errors are generated from the **ARMA** process

	$\mathcal{V}^{(1)}_{N,T}$	$\mathcal{V}^{(2)}_{N,T}$	$\mathfrak{H}^{(1)}_{N,T}$	$\mathfrak{H}^{(2)}_{N,T}$	$\mathfrak{J}_{N,T}$	$\mathfrak{C}_{N,T}$	$\mathcal{V}^{(1)}_{N,T}$	$\mathcal{V}^{(2)}_{N,T}$	$\mathfrak{H}^{(1)}_{N,T}$	$\mathfrak{H}^{(2)}_{N,T}$	$\mathfrak{J}_{N,T}$	$\mathfrak{C}_{N,T}$
	T=100, N=100						T=100, N=250					
IN	0.02	0.02	0.03	0.02	0.01	0.00	0.03	0.03	0.01	0.01	0.02	0.00
CF-W-0.25	0.03	0.03	0.02	0.02	0.02	0.00	0.03	0.02	0.02	0.01	0.02	0.00
CF-W-0.5	0.03	0.03	0.02	0.02	0.02	0.00	0.02	0.02	0.01	0.01	0.02	0.00
CF-W-1	0.04	0.04	0.01	0.02	0.02	0.00	0.04	0.02	0.01	0.00	0.02	0.00
CF-S-0.25	0.03	0.05	0.02	0.02	0.01	0.00	0.03	0.02	0.02	0.02	0.01	0.00
CF-S-0.5	0.04	0.03	0.03	0.03	0.01	0.04	0.06	0.06	0.03	0.03	0.01	0.06
CF-S-1	0.06	0.05	0.03	0.03	0.02	0.10	0.06	0.07	0.02	0.02	0.02	0.14
	T=250, N=100						T=250, N=250					
IN	0.02	0.02	0.01	0.00	0.01	0.00	0.01	0.01	0.01	0.00	0.01	0.00
CF–**W**–0.25	0.03	0.03	0.01	0.01	0.01	0.00	0.01	0.01	0.00	0.00	0.00	0.00
CF–**W**–0.5	0.02	0.02	0.01	0.01	0.01	0.00	0.01	0.01	0.01	0.00	0.00	0.00
CF–**W**–1	0.02	0.04	0.01	0.02	0.01	0.00	0.03	0.02	0.01	0.03	0.00	0.00
CF–**S**–0.25	0.04	0.01	0.02	0.02	0.01	0.00	0.05	0.00	0.03	0.03	0.00	0.00
CF–**S**–0.5	0.05	0.04	0.04	0.04	0.01	0.06	0.06	0.04	0.04	0.04	0.01	0.03
CF–**S**–1	0.03	0.07	0.04	0.04	0.01	0.09	0.07	0.04	0.04	0.04	0.01	0.06
	T=500, N=100						T=100, N=500					
IN	0.01	0.01	0.00	0.00	0.01	0.00	0.03	0.03	0.00	0.00	0.01	0.00
CF–**W**–0.25	0.03	0.03	0.01	0.01	0.00	0.00	0.05	0.04	0.03	0.02	0.01	0.00
CF–**W**–0.5	0.03	0.03	0.01	0.01	0.00	0.00	0.06	0.06	0.02	0.02	0.01	0.00
CF–**W**–1	0.04	0.01	0.01	0.07	0.01	0.00	0.06	0.06	0.01	0.01	0.01	0.00
CF–**S**–0.25	0.04	0.01	0.03	0.03	0.01	0.00	0.03	0.03	0.02	0.01	0.01	0.00
CF–**S**–0.5	0.03	0.05	0.04	0.04	0.01	0.05	0.04	0.06	0.01	0.01	0.01	0.05
CF–**S**–1	0.04	0.06	0.03	0.03	0.02	0.06	0.04	0.06	0.02	0.02	0.02	0.11

Table 2. Accuracy of change point estimation at 5% significance level for DGP **CF-S**-0.5 when the errors are generated from the **ARMA** process

	$\mathcal{V}^{(1)}_{N,T}$	$\mathcal{V}^{(2)}_{N,T}$	$\mathfrak{H}^{(1)}_{N,T}$	$\mathfrak{H}^{(2)}_{N,T}$	$\mathfrak{J}_{N,T}$	$\mathfrak{C}_{N,T}$	$\mathcal{V}^{(1)}_{N,T}$	$\mathcal{V}^{(2)}_{N,T}$	$\mathfrak{H}^{(1)}_{N,T}$	$\mathfrak{H}^{(2)}_{N,T}$	$\mathfrak{J}_{N,T}$	$\mathfrak{C}_{N,T}$
	$\delta = 0.15$						$\delta = 0.30$					
T=100,N=100	0.59	0.57	0.60	0.55	0.47	0.38	0.75	0.73	0.76	0.75	0.42	0.56
T=100,N=250	0.56	0.54	0.52	0.58	0.29	0.34	0.68	0.71	0.73	0.75	0.38	0.61
T=250,N=100	0.38	0.42	0.45	0.41	0.19	0.37	0.73	0.67	0.74	0.70	0.36	0.67
T=250,N=250	0.40	0.51	0.42	0.49	0.18	0.30	0.74	0.74	0.75	0.76	0.35	0.66
T=500,N=100	0.33	0.42	0.37	0.42	0.19	0.33	0.67	0.73	0.67	0.73	0.32	0.68
T=100,N=500	0.53	0.51	0.64	0.54	0.24	0.37	0.70	0.70	0.77	0.73	0.39	0.65

selected from 100, 250, and 500. The accuracy of the change point estimator is defined as the percentage of estimators $\hat{k}_{N,T}$, conditionally on rejecting $\mathcal{H}_0$ at level $\alpha = 0.05$, that lies within the interval of $[T/2 - \lfloor \log T \rfloor, T/2 + \lfloor \log T \rfloor]$. The results for ARMA errors were generally similar, when compared to the GARCH case, and so we omit similar results for the GARCH case. The percentage of simulations that reject $\mathcal{H}_0$ at the 5% significance level are reported in Tables 1, and the comparison of power of these procedures are displayed in terms of power curves in Figures 1 and 2.

We summarize these results as follows. In general, all tests exhibited reasonably good empirical size under $\mathcal{H}_0$, with a slight tendency to be under-sized under weak serial/cross sectional dependence. The tests $\mathcal{V}^{(1)}_{N,T}$ and $\mathcal{V}^{(2)}_{N,T}$ overall performed the best in terms of size, although the improvement over the other tests was fairly modest. The two proposed bootstraps generally performed similarly, although the bootstrap Algorithm 2 started to show some poor performance under strong cross–sectional dependence. The test $\mathfrak{J}_{N,T}$

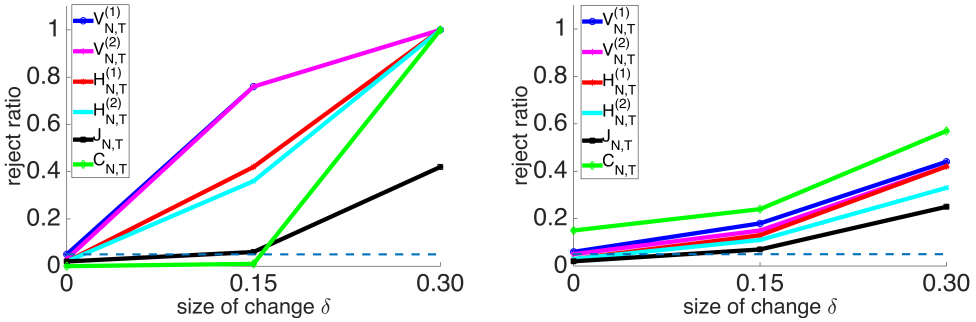

Figure 1. Plots of the empirical power as a function of the size of the change δ for the tests $\mathcal{V}_{N,T}^{(1)}$, $\mathcal{V}_{N,T}^{(2)}$, $\mathfrak{H}_{N,T}^{(1)}$, $\mathfrak{H}_{N,T}^{(2)}$, $\mathfrak{J}_{N,T}$ and $\mathfrak{C}_{N,T}$. The left panel is based on data generated according to **IN** with $N = 250$ and $T = 100$, and the right panel is based on data generated according to **CF-S**-1 with $N = 250$ and $T = 100$.

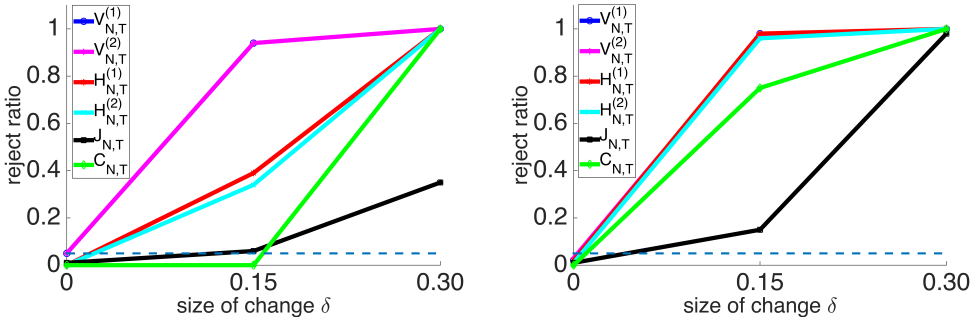

Figure 2. Plots of the empirical power as a function of the size of the change δ for the tests $\mathcal{V}_{N,T}^{(1)}$ (blue), $\mathcal{V}_{N,T}^{(2)}$, $\mathfrak{H}_{N,T}^{(1)}$, $\mathfrak{H}_{N,T}^{(2)}$, $\mathfrak{J}_{N,T}$ and $\mathfrak{C}_{N,T}$. The left panel is based on data generated according to **CF-W**-1 with $N = 500$ and $T = 100$, and the right panel is based on data generated according to **CF-W**-1 with $N = 100$ and $T = 250$.

was under-sized which is consistent with the results presented in Jirák (2015). Similarly, the test $\mathfrak{C}_{N,T}$ was under-sized in independent panels, and was somewhat over-sized in highly dependent panels.

In terms of power, Figures 1 and 2 show power curves as a function of the change magnitude $\delta = 0.15$ and 0.30 when the significance level is set at 5% in the case of ARMA errors for several settings. In general, all tests gain more power as expected when N and T increase, and they attain nearly unit power in dimensions of $(T = 250, N = 100)$, $(T = 250, N = 250)$, and $(T = 500, N = 100)$ if the change is large. In many cases, the tests $\mathcal{V}_{N,T}^{(1)}$ and $\mathcal{V}_{N,T}^{(2)}$ outperformed all other tests across dimensions and dependence levels. Furthermore, the $\mathcal{V}_{N,T}^{(1)}$ and $\mathcal{V}_{N,T}^{(2)}$ tests almost uniformly outperformed the $\mathfrak{H}_{N,T}^{mod}$ tests, which showed the improvement generated by applying the proposed centering scheme defining $V_{N,T}(u)$ in (2.3). One exception to the superiority of the $\mathcal{V}_{N,T}^{(1)}$ and $\mathcal{V}_{N,T}^{(2)}$ tests is the case of large N and highly cross–sectionally dependent panels, in which the test $\mathfrak{C}_{N,T}$

tended to perform better. The test $\mathfrak{J}_{N,T}$ showed poor performance in our simulations, although this is likely due to the relatively small changes in the mean under strong dependence, while this test should perform comparatively better in the case when the changes are larger but only in a few cross-sections. Although our tests are not designed for a panel with either small N or T, in an unreported simulation, the tests $\mathcal{V}^{(1)}_{N,T}$ and $\mathcal{V}^{(2)}_{N,T}$ still maintain good power in the case that $N >> T$, e.g., $N = 2000$ and $T = 30$. The power of the other tests considered were low in this scenario.

The results on change point accuracy are provided in Tables 2 in the case of **CF**–**S**–0.5 with ARMA errors only, in order to save space. The chosen DGP with stronger dependence structure exhibits an overall change point estimation accuracy for all methods that is much lower than the other DGPs we consider, but the pattern that emerges here was also observed for the other DGPs. The statistics $\mathcal{V}^{(1)}_{N,T}$, $\mathcal{V}^{(2)}_{N,T}$, $\mathfrak{H}^{(1)}_{N,T}$, $\mathfrak{H}^{(2)}_{NT}$, and $\mathfrak{C}_{N,T}$ perform similarly well in all dimensions of panels. The $\mathfrak{H}^{(1)}_{N,T}$ and $\mathfrak{H}^{(2)}_{N,T}$ typically perform the best. This is to some extent expected given the way the process $V_{N,T}(u)$ is centered.

6. APPLICATIONS TO US MACROECONOMIC DATA

In this section, we demonstrate the applicability of the proposed test to detect and analyze structural breaks in large panels of macroeconomic data related to the United States (US). We focus on the FRED–MD data set, which comprises monthly resolution data on 128 macroeconomic variables available from the United States federal reserve economic database (FRED). The analysis of high–dimensional macroeconomic panel data has drawn a great deal of attention in the last decade. One of the most influential papers on this area of research is Stock and Watson (2012), who used up to 200 time series of macroeconomic variables to investigate the dynamics of the great recession during 2007–09.

In total, the 128 time series are taken from the period from June 1999 to June-2019, with information related to nine areas, including output and income, labor markets, consumption and orders, orders and inventories, money and credit, interest rate and exchange rates, prices, and the stock market. McCracken and Ng (2016) provided detailed descriptions of this dataset. They also suggested transformations of each series towards stationarity so the data are suitable for a factor analysis, which we follow. They found that the transformed data have a factor structure similar to the model considered in Stock and Watson (2012). Analysis of this panel data using the criteria proposed by Bai and Ng (2002) suggests that the cross–sectional dependence is well explained by eight factors.

We take as the goal of the analysis to evaluate for structural breaks in the means of this panel over the observation period. Change points in the means of the cross-sections may represent changing phases of the US economy. To detect change points in the panel, we applied the tests discussed in Section 5. We used a binary segmentation method, which entails the segmentation of the data into pre- and post–break segments, to estimate and detect additional changes in the means.

We applied all six change point tests used in Section 5 to the FRED–MD data. Each method identified three change points at relatively similar time points. All the tests $\mathcal{V}^{(1)}_{N,T}$, $\mathcal{V}^{(2)}_{N,T}$, $\mathfrak{H}^{(1)}_{N,T}$ and $\mathfrak{H}^{(2)}_{NT}$ found that the biggest change in the macroeconomic structure occurred in March 2008, which corresponds to when the US government bailout began. However, different dates were found in the second and third step of the binary segmenta-

tion. The tests $\mathcal{V}_{N,T}^{(1)}$ and $V_{N,T}^{(2)}$ detected changes in June 2006 and January 2016, whereas $\mathfrak{H}_{N,T}^{(1)}$ and $\mathfrak{H}_{N,T}^{(2)}$ found that August 2003 and August 2012 are the most likely times of changes in the means. Instead of finding the first change in 2008, the tests $\mathfrak{J}_{N,T}$ and $\mathfrak{C}_{N,T}$ detected changes on December 2011 and August 2007, respectively. The change in 2007 seems plausible due to the "sub–prime mortgage crisis", but it is rather difficult to interpret the detected change in 2011. This finding is likely due to the properties of $\mathfrak{J}_{N,T}$ that is tailored for sparse data with weak cross sectional dependence.

7. CONCLUSION

In this paper, we consider the changes in the means of high dimensional panel data with potentially strong cross–sectional dependence. A new CUSUM-type statistic is proposed to distinguish the null of no change point from the alternative of multiple changes. To establish the asymptotic results, we only require the mild condition on the relative divergence rate of the cross-sectional dimension N and time series observations T that $min(N, T) \to \infty$. Assuming that the cross–sectional dependence is captured by latent common factors, we show that the statistic converges to its limits differently under three scenarios: the common factors are asymptotically negligible, i.e., no or weak cross–sectional dependence; the common factors are asymptotically dominating, i.e., strong dependence; and a status between these two. The main results are also extended and shown to be valid even if the error sequences are serially correlated. We suggest using semi-parametric bootstrap methods to obtain critical values. The finite sample performance of the proposed test is assessed in a Monte Carlo simulation study. Our tests enjoy good rejection rates in terms of size and power in arbitrary relationship between N and T, and they nearly always outperform the existing tests. Two data applications are also provided. We detect breaks in means of the FRED–MD microeconomic data set.

REFERENCES

Abadir, K.M. and Magnus, J.R. (2005). *Matrix Algebra*, Cambridge University Press, Cambridge, U.K.

Antoch, J., Hanousek, J., Horváth, L., Husková, M. and Wang, S. (2019). Structural breaks in panel data: large number of panels and short length time series. *Econometric Reviews 38*, 828–855.

Bai, J. (2010). Common breaks in means and variances for panel. *Journal of Econometrics 157*, 78–92.

Bai, J., Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica 70*, 191–221.

Bai, J. and Ng, S. (2008). Large dimensional factor analysis. *Foundations and Trends in Econometrics 3*, 89–163.

Billingsley, P. (1968). *Convergence of Probability Measures*, Wiley, New York.

Baltagi, B.H., Kao, C. and Wang, F. (2017). Identification and estimation of a large factor model with structural instability. *Journal of Econometrics 197*, 87–100.

Breitung, J. and Eickmeier, S. (2017). Testing for structural breaks in dynamic factor models. *Journal of Econometrics 163*, 71–84.

Carrasco, M. and Chen, X. (2002). Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory 18*, 17–39.

Chan, J., Horváth, L. and Hušková, M., (2013). Darling–Erdös limit results for change-point detection in panel data. *Journal of Statistical Planning and Inference 143*, 955–970.

Cho, H. (2016). Change-point detection in panel data via double CUSUM statistic. *Electronic Journal of Statistics 10*, 2000–2038.

Chudik, A. and Pesaran, M.H. (2015). Large panel data models with cross–sectional dependence: a survey. In: *The Oxford Handbook on Panel Data*, The Oxford University Press, New York, pp. 3–45.

Chen, L., Dolado, J.J. and Gonzalo, J. (2014). Detecting big structural breaks in large factor models. *Journal of Econometrics 180*, 30–48.

Csörgő, M. and Horváth, L. (1997). *Limit Theorems in Change–Point Analysis.* Wiley, New York, 1997.

Doukhan, P. and Wintenberger, O. (2007). An invariance principle for weakly dependent stationary general models. *Probability and Mathematical Statistics 27*, 45–73.

Halungaa, A. and Osborn, D. (2012). Ratio–based estimators for a change point in persistence. *Journal of Econometrics 171*, 24–31.

Han, X. and Inoue, A. (2015). Tests for parameter instability in dynamic factor models. *Econometric Theory 31*, 1117–1152.

Hörmann, S. (2008). Augmented GARCH sequences: dependence structure and asymptotics. *Bernoulli 14*, 543–561.

Horváth, L. and Hušková, M. (2012). Change–point detection in panel data. *Journal of Time Series Analysis 33*, 631–648

Horváth, L., Husková, M., Rice, G. and Wang, J. (2017). Asymptotic properties of the CUSUM estimator for the time of change in linear panel data models. *Econometric Theory 33*, 366–412.

Jirák, M. (2015). Uniform change point tests in high dimension. *The Annals of Statistics 43*, 2451–2483.

Joseph, L. and Wolfson, D.B. (1992). Estimation in multi–path change-point problems. *Communications in Statistics–Theory and Methods 21*, 897–913.

Joseph, L. and Wolfson, D.B. (1993). Maximum likelihood estimation in the multi–path changepoint problem. *Annals of the Institute of Statistical Mathematics 45*, 511–530.

Kim, D. (2011). Estimating a common deterministic time trend break in large panels with cross sectional dependence. *Journal of Econometrics 164*, 310–330.

Kim, D. (2014). Common breaks in time trends for large panel data with a factor structure. *Econometrics Journal 17*, 301–337.

McCracken, M. W., Ng, S. (2016). FRED-MD: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics 34*, 574–589.

Nzé, P.A. and Doukhan, P. (2004). Weak dependence: Models and applications to econometrics. *Econometric Theory 20*, 995–1045.

Shao, X. (2015). Self-normalization for time series: A review of recent developments. *Journal of the American Statistical Association 110*, 1797–1817.

Stock, J.H. and Watson, M.W. (2012). Disentangling the channels of the 2007–2009 recession (No. w18094). National Bureau of Economic Research.

Vogelsang, T.J. (1999). Sources of nonmonotonic power when testing for a shift in mean of a dynamic time series *Journal of Econometrics 88*, 283–299.

Westerlund, J. (2019). Common Breaks in Means for Cross–Correlated Fixed–T Panel Data. *Journal of Time Series Analysis 40*, 248–255.

Wu, W. B. (2011). Asymptotic theory for stationary processes. *Statistics and its Interface 4*, 207–226.

McCracken, M. W. (2021). FRED-MD and FRED-QD: Monthly and Quarterly Databases for Macroeconomic Research. *https://research.stlouisfed.org/econ/mccracken/fred-databases/ (last accessed: 6 June 2021).*