Testing for Structural Changes in Large Dimensional Factor Models via Discrete Fourier Transform*

Zhonghao Fu[a], Yongmiao Hong[b], Xia Wang[c]

[a] Fudan University, China
[b] Cornell University, U.S. and Xiamen University, China
[c] Sun Yat-Sen University, China

March 24, 2019

Abstract

We propose a new test for structural changes in large dimensional factor models via a discrete Fourier transform (DFT) approach. If structural changes exist, the conventional principal component analysis (PCA) will fail to estimate common factors and factor loadings consistently. The estimated residuals will contain information about structural changes. Therefore, we can compare the DFT of the residuals with the zero spectrum implied by no structural change. By construction, the proposed test is powerful against both smooth structural changes and abrupt structural breaks with possibly unknown number of breaks and unknown break dates in factor loadings. It can detect a class of local alternatives at the rate $T^{-1/2}N^{-1/2}$, and so is asymptotically more efficient than the existing tests in the literature. Moreover, it is easy to implement and tuning parameter-free. And our test is robust to serial correlation and cross-sectional dependence of unknown form. Monte Carlo studies demonstrate its reasonable size and excellent power in detecting structural changes of unknown types in factor loadings. In an application to Stock and Watson's (2012) U.S. macroeconomic data, we find significant evidence against time-invariant factor loadings.

JEL Classification: C12, C14, C33, C38.

Key Words: Factor model, Structural changes, Discrete Fourier transform, Local power

*Wang acknowledges financial supports from the National Science Foundations of China (No. 71401160, 71873151). We thank Liang Chen, Liangjun Su, and seminar participants in the Fourth Guangzhou Econometrics workshop, the Second Chinese Econometrican's Forum, and the Fifth Annual Meeting of Young Econometricans in Asia-Pacific. Please address correspondence to: Xia Wang, Lingnan (University) College, Sun Yat-Sen University, Guangzhaou 510275, China, Phone +86 8411 1191. Email: *wangxia25@mail.sysu.edu.cn*.

1 Introduction

Factor models are useful for analyzing large dimensional macroeconomic and financial datasets. The Principal Component Analysis (PCA) has been extensively used to deal with latent factor models. Most existing works (*e.g.*, Stock and Watson, 2002; Bai and Ng, 2002; Bai, 2003) assume the factor loadings, which capture the relationship between economic variables and the unobserved common factors, are time invariant. However, it is likely that the underlying structure of the dataset changes over time when the time span is long. Even though Stock and Watson (2002, 2009) point out that the estimated factors by the PCA are still consistent when the factor loadings undergo small instabilities, it is difficult to believe that the factor loadings are time-invariant or only have small changes during a long sampling period for macroeconomic and financial data. The changing economic environment such as policy shifts, economic transition, preference changes and technological progress, may influence the relationship between economic variables and the unobserved common factors, which is expected to induce the time varying behavior of factor loadings. If the assumption of time-invariant factor loadings fails, the estimated common factors can be inconsistent and the inference and forecasting based on such an assumption may lead to misleading conclusions. Furthermore, if the factor loadings suffer from structural changes, most of the existing methods such as Bai and Ng (2002), Onatski (2009), Ahn and Horenstein (2013) tend to deliver a wrong number of common factors.

Testing for structural changes in time series models is pioneered by Chow (1960). In the past decade, along with the broad applications of factor models, a growing literature starts to focus on modeling and testing structural changes in factor models. Stock and Watson (2009) investigate the forecasting reliability when there exist abrupt structural breaks in factor loadings. Breitung and Eickmeier (2011) propose LR, LM and Wald tests to detect the existence of a single structural break in factor loadings. Chen *et al.* (2014) propose a two-stage procedure to detect a large break in factor loadings in which they first obtain the estimated common factors via PCA and then test parameter stability in a regression of one estimated factor on the remaining factors. Corradi and Swanson (2014) propose a test for structural stability in both factor loadings and factor-augmented forecasting regression coefficients. Han and Inoue (2015) propose a joint test for structural break of factor loadings by comparing the pre- and post-break subsample second moments of estimated factors. Yamamoto and Tanaka (2015) propose a modified version of Breitung and Eickmeier's (2011) test that avoids the non-monotonic power problem. Cheng *et al.* (2016) consider the case in which both factor loadings and the number of factors may change simultaneously. Although the aforementioned works provide useful econometric tools on detecting the possible structural breaks in factor loadings, they only focus on testing for abrupt structural breaks, especially a single structural break. The source of structural changes as preference changes, technological progress and institutional transformation usually take effect gradually over time. Even though some policy switches occur immediately, it may take some time for an economic agent to react. Due to price stickiness, for instance, a company may not be able to adjust the price of its product when facing a corporate tax increase. Thus, it is more realistic to assume smooth changes rather than abrupt breaks in many economic scenarios. In fact, several papers study time varying factor models, e.g., Stock and Watson (2002); Banerjee *et al.* (2008), Bates *et al.* (2013) and Eickmeier *et al.* (2015). All these papers model the time varying factor loadings as a random walk process or a vector autoregressive process and discuss the estimation problem. However, they do not consider the testing problem of structural change. Recently, Su and Wang (2017) propose an L_2-distance-based test statistic to check the stability of factor loadings. They estimate the time-varying factor loadings and the latent common factors by a local version of PCA, and construct statistic to check the null hypothesis of no structural change by comparing the fitted values of the common components with those estimated by the conventional PCA.

In this paper, we propose a new test for structural changes in large dimensional factor models via a discrete Fourier transform (DFT) approach that is first proposed in Fu, Hong and Wang (FHW, hereafter) (2018) under the framework of time series models. Unlike the related existing tests that are based on time domain analysis, FHW propose a novel method that investigate the structural changes in frequency domain. Our test is constructed using the similar idea but works for a different scenario. FHW's tests are proposed under the framework of linear time series regressions with observed regressors. While for the large dimensional factor models, both the factor loadings and common factors are unobservable. The intuition behind our test is quite straightforward. If factor loadings change over time, then the conventional PCA will fail to capture the time-varying feature of the true factor loadings. As a result, the estimated residuals based on the conventional PCA will contain the time-varying component. By the discrete Fourier transform, we can project the residuals onto the frequency domain and infer the existence of structural changes. Compared with the existing tests in the literature, the proposed test has the following appealing features.

First, our test is consistent against a wide range of alternatives of structural changes. the test is powerful against various kinds of smooth structural changes as well as abrupt structural breaks in factor loadings. For abrupt structural breaks, we require that neither the number of breaks nor the break dates to be known. This is contrast to the existing parametric tests for stability of factor loadings, most of which focus on abrupt structural breaks, especially the case with a single break point.

Second, the test can detect a class of local alternatives that converges to the null hypothesis at a faster rate than the existing tests for structural changes in factor models. Let N and T denote the numbers of cross-sectional units and time series periods. Then the rate of local alternatives that our test can detect is $N^{-1/2}T^{-1/2}$, which is faster than the rate of local alternatives for such parametric tests as Breitung and Eickmeier (2011), Chen *et al.* (2014), Han and Inoue (2015) and the nonparametric test by Su and Wang (2017). This is an advantage of using the discrete Fourier transform. In comparison, Su and Wang (2017) can only detect a class of local alternatives at a rate of $T^{-1/2}N^{-1/4}h^{-1/4}$, where h is a bandwidth, while the parametric tests can only capture single structural break with a rate of $T^{-1/2}$. More importantly, as proved by Chen *et al.* (2014), the order $N^{-1/2}T^{-1/2}$ is the upper bound of structural changes in factor loadings that guarantees the consistency of estimated common factors and the number of factor loadings. That is, if the order of magnitude of structural changes in factor loadings is larger than $N^{-1/2}T^{-1/2}$, the estimated number of common factors and the estimated common factors would be inconsistent. As a result, we could detect any structural changes in factor loadings that may lead to inconsistent estimation of the number of factors and the common factors. Simulation studies also demonstrate the significant power improvement of our test over the existing tests in the literature.

Third, our test is tuning parameter-free. It avoids the delicate business of choosing a bandwidth and the arbitrariness of specifying trimming parameters. More importantly, the power of the smoothed non-parametric test by Su and Wang (2017) depends on the choice of the bandwidth h. While they propose a bootstrap version test statistic to relieve this problem, the power of their test is still sensitive to the choice of bandwidth in finite samples. Furthermore, different choices of a bandwidth may lead to conflicting conclusions. The supremum-type tests of Breitung and Eickmeier (2011), Chen *et al.* (2014), Han and Inoue (2015) and Cheng *et al.* (2016) all rely on the pre-specified trimming parameter and hence would miss possible structural changes in the boundary regions.

Finally, our test allows for both cross-sectional dependence and temporal dependence of unknown forms in the error term. Su and Wang (2017) allows for cross-sectional dependence, but require the error term to be a martingale difference sequence. Hence, it assumes that all time series dependence in the observed data is due to the small dimensional common factors. This is rather restrictive for factor analysis with macroeconomic time series, or multi-country or multi-sector factor models. We relax this assumption to allow for time series dependence, and hence broaden the applicability of the proposed test.

The rest of this paper is organized as follows. We introduce our test in Section 2 and establish its asymptotic theory in Section 3. We then demonstrate its finite sample performance in Section 4 and provide an empirical application to U.S. Macroeconomic data in Section 5. We conclude in Section 6. Throughout this paper, we denote $\mathbf{i} = \sqrt{-1}$ to be an imaginary number. For an $m \times n$ real matrix A, we denote its transpose as A', its Euclidean norm as $\|A\|(\equiv [\sum_{i=1}^{m} \sum_{j=1}^{n} |A_{ij}|^2]^{1/2})$, where "$\equiv$"means "is defined as". The operator $\xrightarrow{p}$ denotes convergence in probability, $\xrightarrow{d}$ convergence in distribution, and plim the probability limit. We use $(N, T) \to \infty$ to denote that N and T pass to infinity jointly. Let $C \in (0, \infty)$ denote a generic positive constant that may vary from case to case.

2 Hypotheses and Test Statistic

In this section, we introduce the hypotheses of interest and show how to detect structural changes in factor models via a DFT approach.

2.1 Hypotheses

Let $\{X_{it}, i = 1, 2, \ldots, N;\ t = 1, 2, \ldots, T\}$ be an N-dimensional time series with T observations. The index i represents the ith cross-sectional unit in panel data set or the ith random variable in a multivariate time series data set. We assume that X_{it} is generated via the following factor model

$$X_{it} = \lambda_{it}' F_t + \varepsilon_{it}, \tag{2.1}$$

where F_t is an $R \times 1$ vector of unobserved common factors, λ_{it} is an $R \times 1$ vector of factor loadings that can admit abrupt and/or smooth structural changes over time, and ε_{it} is the idiosyncratic error term.

The null hypothesis of no structural change in the above factor model is:

$$\mathbb{H}_0 : \lambda_{it} = \lambda_{i0} \text{ for } i = 1, 2, \ldots, N \text{ and } t = 1, 2, \ldots, T. \tag{2.2}$$

The alternative hypothesis is

$$\mathbb{H}_1 : \lambda_{it} \neq \lambda_{i0} \text{ for some non-negligible values of } (i, t). \tag{2.3}$$

Obviously, under $\mathbb{H}_0$, λ_{it} is time-invariant and model (2.1) degenerates to the conventional factor model with time-invariant factor loadings. This model has been elaboratively studied in the literature (e.g., Stock and Watson, 2002; Bai and Ng, 2002; Bai, 2003). However, since a data may span a long time period, factor loadings may change over time during the sampling period. In this regard, testing for structural changes in factor models has drawn more and more attention. See, *e.g.*, Breitung and Eickmeier (2011), Chen *et al.* (2014), Cheng *et al.* (2016), and Han and Inoue (2015). Most existing works focus on testing for a single structural break in factor loadings by using some supremum-type test statistics. However, it is rather restrictive to assume only a single abrupt structural break in factor loadings, since usually no prior information about possible structural changes is available in practice. Recently, Su and Wang (2017) model $\lambda_{it} = \lambda_i(t/T)$, where $\lambda_i(\cdot)$ is a deterministic function of scaled time ratio t/T. By assuming λ_{it} to be a piece-wise smooth function, Su and Wang (2017) allow for both smooth structural changes and abrupt structural breaks in factor loadings. In this paper, we do not assume that λ_{it} is a smooth deterministic function of scaled time ratio t/T. Thus, the alternative (2.3) allows various kinds of structural changes in factor loadings, including smooth structural changes, a single structural break as well as multiple structural breaks, with possibly unknown break dates or unknown number of breaks. The setting of our test is rather general.

2.2 Test Statistic

Under the null hypothesis of no structural change in factor loadings, we can follow Bai and Ng (2002) and Bai (2003) and apply the PCA method to estimate the following model

$$X_{it} = \lambda'_{i0} F_t + \varepsilon^\dagger_{it}, \tag{2.4}$$

where $\varepsilon^\dagger_{it} = \varepsilon_{it}$ under $\mathbb{H}_0$ and they are distinct under $\mathbb{H}_1$.

Let $X_t \equiv (X_{1t}, \ldots, X_{Nt})'$, $\varepsilon_t \equiv (\varepsilon_{1t}, \ldots, \varepsilon_{Nt})'$, $\varepsilon^\dagger_t \equiv (\varepsilon^\dagger_{1t}, ..., \varepsilon^\dagger_{Nt})'$, $F \equiv (F_1, \ldots, F_T)'$, and $\Lambda_0 \equiv (\lambda_{10}, ..., \lambda_{N0})'$. Put $X = (X_1, ..., X_T)'$, $\varepsilon \equiv (\varepsilon_1, \ldots, \varepsilon_T)'$, $\varepsilon^\dagger \equiv (\varepsilon^\dagger_1, ..., \varepsilon^\dagger_T)'$. Then we can rewrite (2.4) in vector form

$$X = F\Lambda'_0 + \varepsilon^\dagger.$$

The PCA method solves the following minimization problem:

$$\min_{F,\Lambda} \operatorname{tr}\left(X - F\Lambda_0'\right)\left(X - F\Lambda_0'\right)' = \sum_{i=1}^{N}\sum_{t=1}^{T}\left(X_{it} - \lambda_{i0}'F_t\right)^2$$

under certain identification restrictions. In this paper, we follow Bai (2003) and consider the following identification restrictions:

$$T^{-1}F'F = \mathbb{I}_R \text{ and } \Lambda_0'\Lambda_0 \text{ is a diagonal matrix.}$$

Let $\hat{F}_t$ and $\hat{\lambda}_{i0}$ be the principal component estimators of F_t and λ_{i0}, respectively under the above identification restrictions. Let $\hat{F} = (\hat{F}_1, ..., \hat{F}_T)'$ and $\hat{\Lambda}_0 = (\hat{\lambda}_{10}, ..., \hat{\lambda}_{N0})'$. It is well known that $\hat{F}$ is $\sqrt{T}$ times eigenvectors corresponding to the R largest eigenvalues of the $T \times T$ matrix XX', and $\hat{\Lambda}_0' = (\hat{F}'\hat{F})^{-1}\hat{F}'X = T^{-1}\hat{F}'X$.

After obtaining the restricted estimators $\hat{F}_t$ and $\hat{\lambda}_{i0}$ of F_t and λ_{i0}, we now consider the following complex-valued empirical process:

$$\begin{aligned}\hat{A}(u) &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{F}_t\hat{\varepsilon}_{it}e^{\mathbf{i}u2\pi t/T} \\ &= \frac{1}{T}\sum_{t=1}^{T}\hat{F}_t\left(\frac{1}{N}\sum_{i=1}^{N}\hat{\varepsilon}_{it}\right)e^{\mathbf{i}u2\pi t/T},\end{aligned}$$

where $\hat{\varepsilon}_{it} = X_{it} - \hat{\lambda}_{i0}'\hat{F}_t$ is the residuals from PCA. To construct $\hat{A}(u)$, we first perform a discrete Fourier transform of $\hat{F}_t\hat{\varepsilon}_{it}$ for each i, and then take the average over cross-sectional units. The individual DFT can detect structural change at each i. Taking a cross-sectional average can combine individual DFTs together. Such a treatment is quite common in the literature. For instance, when testing for cross section dependence in panel data model, Pesaran (2004) first estimates pairwise serial correlations and then construct a test via their cross-sectional average; Levin *et al.* (2002) consider testing for unit root in panel data and their test statistics is constructed via double summation in both cross-sectional and time series indices; Chen and Huang (2018) test for smooth structural change in panel data models by comparing the difference between the consistent estimate under structural change and conventional least squares estimate; In factor models, Breitung and Eickmeier construct (2011) LR, LM, and Wald statistics at each cross-sectional unit i and then combine the individual statistics to obtain a pooled test for a single structural break, *etc.*

$\hat{A}(u)$ is equivalent to an average of discrete Fourier transforms at each cross-sectional unit i. If structural change exits, a test based on $\hat{A}(u)$ is consistent as long as the DFTs does not cancel with each other. The intuition behind our test is quite straightforward: if the factor loadings have structural changes, then the PCA fails to capture the time-varying behavior of λ_{it}, and such information will be hidden in the residuals $\hat{\varepsilon}_{it}$. By DFT, we can reveal such information in the frequency domain, because the possible time-varying behavior of the factor loadings can be completely captured by the DFT of $\hat{\varepsilon}_{it}$. By examining the pattern of the DFT at each frequency, we can detect structural change of unknown types. Compared to the existing tests that are based on time domain analysis, the DFT-based approach does not need prior information about the types of structural change. For instance, to apply the tests by Breitung and Eickmeier (2011), Chen *et al.* (2014), and Han and Inoue (2015), one needs to specify an abrupt type of structural change. On the other hand, while the consistent test by Su and Wang (2017) does not require to specify the change to be abrupt or smooth, it requires nonparametric local smoothing over the time domain. In contrast, our DFT-based test is free of the aforementioned issues.

To gain further insight into $\hat{A}(u)$, we decompose it as the following:

$$\begin{aligned}
\hat{A}(u) &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{F}_t\hat{\varepsilon}_{it}e^{\mathbf{i}u2\pi t/T} \\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\hat{F}_t\left(X_{it}-\hat{F}_t'\hat{\lambda}_{i0}\right)e^{\mathbf{i}u2\pi t/T}\right] \\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\hat{F}_t e^{\mathbf{i}u2\pi t/T}\left(X_{it}-\hat{F}_t'\left[\frac{1}{T}\sum_{t=1}^{T}\hat{F}_t X_{it}\right]\right)\right] \\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left\{\left[\hat{F}_t e^{\mathbf{i}u2\pi t/T}-\left(\frac{1}{T}\sum_{t=1}^{T}\hat{F}_t\hat{F}_t' e^{\mathbf{i}u2\pi t/T}\right)\hat{F}_t\right]X_{it}\right\} \\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}G_t(u)F_t'\lambda_{it}+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}G_t(u)\varepsilon_{it} \\
&\equiv \hat{A}_1(u)+\hat{A}_2(u),
\end{aligned}$$

where we define

$$\begin{aligned}
\hat{A}_1(u) &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}G_t(u)F_t'\lambda_{it}, \\
\hat{A}_2(u) &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}G_t(u)\varepsilon_{it}
\end{aligned}$$

and $G_t(u)=\hat{F}_t e^{\mathbf{i}u2\pi t/T}-\left(\frac{1}{T}\sum_{t=1}^{T}\hat{F}_t\hat{F}_t' e^{\mathbf{i}u2\pi t/T}\right)\hat{F}_t$. Under certain regularity conditions, *e.g.*, $\hat{F}_t$ is weakly stationary, we can show that $G_t(u)$ is asymptotically equivalent to

$$\hat{F}_t\left(e^{\mathbf{i}u2\pi t/T}-\int_0^1 e^{\mathbf{i}u2\pi\tau}d\tau\right),$$

which is a product of the estimated factor $\hat{F}_t$ and a demeaned Fourier series. The component $\hat{A}_1(u)$ captures the structural changes in factor loadings since it is asymptotically equivalent to pseudo-covariance of λ_{it} and Fourier basis function of time. The component $\hat{A}_2(u)$ is a pure noise term and it determines the asymptotic distribution given the orthogonality conditions between F_t and ε_{it}. Intuitively, the DFT $\hat{A}(u)$ is equivalent to a linear projection of X_{it} onto the frequency domain. The projection vector $G_t(u)$ can be viewed as a filter in the space spanned by $\hat{F}_t$ and time t/T. It is asymptotically orthogonal to X_{it} when the factor loadings are constant over time. By Euler's formula, $e^{\mathbf{i}u2\pi t/T} = \cos(u2\pi t/T) + \mathbf{i}\sin(u2\pi t/T)$. If the unknown factor loading has structural change, *i.e.,* λ_{it} is an unknown function of time, then it can be represented as an infinite sum of Fourier series. Since λ_{it} is contained by X_{it}, the linear projection X_{it} cannot pass the filter $G_t(u)$ and will converge to a non-constant spectrum. On the other hand, when there is no structural change, *i.e.,* λ_{it} is a constant function over time, the linear projection of X_{it} converges to a zero spectrum.

The test is based on Fourier series approximation of unknown λ_{it}. To ensure our DFT approach can detect structural change of unknown form, we need to examine the deviation of $\hat{A}(u)$ from a zero spectrum at each frequency u. Thus, we consider the following test statistic

$$\hat{D} = NT \int_{\mathbb{R}} \|\hat{A}(u)\|^2 W(u) du, \tag{2.5}$$

where $W : \mathbb{R} \to \mathbb{R}^+$ is a nonnegative symmetric weighting function of u. The use of $W(u)$ allows us to examine $\hat{A}(u)$ at all frequencies via different weights. If we choose a discrete probability mass function, then (2.5) degenerates to a weighted sum over various points of u. However, a discontinuous weighting function may adversely affect the power of the test. In practice, one may like to avoid numerical integration in (2.5) by choosing some suitable weighting functions. For example, if we follow Hong *et al.* (2017) to use the standard normal weighting function, then the test statistic could be written as:

$$\hat{D}_W = \frac{1}{NT} \sum_{i,j=1}^{N} \sum_{t,s=1}^{T} \hat{F}_t \hat{F}_s' \hat{\varepsilon}_{it} \hat{\varepsilon}_{js} \exp\{-2\pi^2 [(t-s)/T]^2\}$$

3 Asymptotic Properties of the Test Statistic

In this section, we derive the asymptotic null distribution of our test and investigate its asymptotic local power property. We also propose a block bootstrap procedure to improve the finite sample performance of the test.

3.1 Assumptions

Let $\gamma_N(s,t) = N^{-1}E(\varepsilon_s'\varepsilon_t)$, $\xi_{st} = N^{-1}[\varepsilon_s'\varepsilon_t - E(\varepsilon_s'\varepsilon_t)]$, $\gamma_{N,FF}(s,t) = N^{-1}E(F_s\varepsilon_s'\varepsilon_t F_t')$ and $\tau_{ij,st} = E(\varepsilon_{it}\varepsilon_{js}F_t'F_s)$. We use $\max_i$, $\max_t$, $\max_{i,t}$ and $\max_{s,t}$ to denote $\max_{1\le i\le N}$, $\max_{1\le t\le T}$, $\max_{1\le i\le N}\max_{1\le t\le T}$ and $\max_{1\le s,t\le T}$, respectively. Throughout, we make the following assumptions.

Assumption A.1 [Factors] (i) $E(F_tF_t') = \Sigma_F$ for some $R\times R$ positive definite matrix Σ_F; (ii) $\max_t E\|F_t\|^{8+\delta} < \infty$ for some $\delta > 0$; (iii) $E\|\hat{F}_t\|^{8+\delta} < \infty$ for some $\delta > 0$; (iv) $E(\hat{F}_tF_t')$ is finite and nonsingular.

Assumption A.2 [Factor Loadings] (i) λ_{i0} are nonrandom such that $\max_i \|\lambda_{i0}\| \le C$; (ii) $N^{-1}\Lambda_0'\Lambda_0 = N^{-1}\sum_{i=1}^N \lambda_{i0}\lambda_{i0}' \to \Sigma_{\Lambda_0}$ for some $R \times R$ positive definite matrix Σ_{Λ_0}; (iii) The eigenvalues of the $R \times R$ matrix $\Sigma_F\Sigma_{\Lambda_0}$ are distinct.

Assumption A.3 [Error term] (i) $E(\varepsilon_{it}) = 0$, $\max_{i,t} E|\varepsilon_{it}|^{8+\delta} \le C$ and $\max_{i,t} E\|F_t\varepsilon_{it}\|^{8+4\delta} \le C$ for some $\delta > 0$; (ii) For each $i = 1, 2, \ldots, N$, the process $\{(\varepsilon_{it}, F_t), t = 1, 2, \ldots\}$ is strong mixing with mixing coefficients $\alpha_i(\cdot)$. $\alpha(\cdot) \equiv \max_i \alpha_i(\cdot)$ satisfies $\sum_{s=1}^{\infty} \alpha(s)^{\delta/(2+\delta)} \leq C$ for some $\delta > 0$; (iii) $\max_t \sum_{s=1}^{T} |\gamma_N(s,t)| \leq C$, $\max_{s,t} E\left|N^{1/2}\xi_{st}\right|^4 \leq C$, $\max_t E|N^{-1/2}\sum_{i=1}^{N}[\varepsilon_{it}^2 - E\left(\varepsilon_{it}^2\right)]|^4 \leq C$; (iv) $\max_t \sum_{s=1}^{T} |\gamma_{N,FF}(s,t)| \leq C$, $\max_{t\neq r} E\|N^{-1/2}F_t\varepsilon_t'\varepsilon_r F_r'\|^4 \leq C$, and $N^{-1}T^{-1}\sum_{i,j=1}^{N}\sum_{s,t=1}^{T}|\tau_{ij,st}| \leq C$; (v) $\|\varepsilon\|_{\text{sp}} = O_P(N^{1/2} + T^{1/2})$.

Assumption A.4 [Weighting function] (i) $W(\cdot) : \mathbb{R} \to \mathbb{R}^+$ is nonnegative, symmetric, continuous and integrable weighting function; (ii) $\int_{\mathbb{R}} |u|^4 W(u)du < \infty$.

Assumption A.1 imposes conditions on the latent common factors. We follow Stock and Watson (2002), Breitung and Eickmeier (2011), Chen *et al.* (2014), Han and Inoue (2015) and Su and Wang (2017) and assume that $E\left(F_t F_t'\right) = \Sigma_F$ is homogeneous over t. This assumption assumes that there is no structural change on the second moment of F_t. It greatly facilitates the derivation of the asymptotic results and can be regarded as an identification condition. As is well known, the latent common factors and the factor loadings are not separately identifiable. A factor model with structural changes in common factors and time-invariant factor loadings is equivalent to a model with stationary common factors and time-varying factor loadings. In fact, even if there is no structural change in factor loadings and the second moment of common factors, we can always write that $\lambda_i' F_t = \lambda_i' Q(t/T)^{-1} Q(t/T) F_t = \lambda_{it}^{*'} F_t^*$ for any nonsingular matrix $Q(t/T)$ with $\lambda_{it}^* = Q(t/T)^{-1}\lambda_i$ and $F_t^* = Q(t/T)F_t$ being time varying factor loadings and common factors with time varying second moment. Assumption A.1(i) rules out this problem. Assumption A.2 ensures that each factor has a nontrivial contribution to the variance of X_t. Following Bai (2003) and Breitung and Eickmeier (2011), we assume that factor loadings are nonrandom for simplicity.

Assumption A.3 imposes moment conditions on the errors and their interactions with the factors and factor loadings. Assumptions A.3(i) and (iii) correspond to Assumptions C.1 and C.5 in Bai (2003). Compared to Su and Wang (2017), we allow for both serial correlation and cross-sectional dependence in the error terms. A.3(ii) requires the process $\{(\varepsilon_{it}, F_t), t = 1, 2, ...\}$ to be strong mixing with some algebraic mixing rate. With a more complicated notation, one could allow different individual time series to have various mixing rates and relax the summability mixing condition to $\limsup_N \frac{1}{N}\sum_{i=1}^{N}\sum_{s=1}^{\infty}\alpha_i(s)^{\delta/(1+\delta)} \leq C < \infty$. If the processes are strong mixing with a geometric rate (*e.g.*, $\alpha(s) = \rho^s$ for some $\rho \in [0,1)$), then the conditions on $\alpha(\cdot)$ can be met by specifying $T_0 = \lfloor C_0 \ln T \rfloor$ for some sufficiently large positive constant C_0. Assumptions A.3(iii) and (iv) control the cross-sectional dependence among $\{\varepsilon_{it}, i = 1, 2, \ldots, N\}$ and $\{F_t\varepsilon_{it}, i = 1, 2, \ldots, N\}$, respectively. Assumption A.3(v) is widely assumed in the factor model literature; see, *e.g.*, Moon and Weidner (2015), Su and Wang (2017), and Ma and Su (2017). Assumption A.4 imposes some mild conditions on the weighting function. It ensure the existence of the integral in (2.5).

3.2 Asymptotic Null Distribution

We now state the asymptotic distribution of $\hat{A}(u)$ under $\mathbb{H}_0$.

Proposition 3.1 *Under Assumptions A.1-A.3, and* $\mathbb{H}_0 : \lambda_{it} = \lambda_{i0}$ *holds,*

$$\hat{A}(u) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} G_t(u) B_i \varepsilon_{it} + O(T^{-3/2}) + o_p(1)$$

where $B_i = 1 - \lambda_{i0}' \left(\Lambda_0'\Lambda_0/N\right)^{-1} \bar{\lambda}_0$, *and* $\bar{\lambda}_0 = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \lambda_{i0}$. *Let* $T \propto N^{\nu}$ *with* $\nu > 1/2$, *then*

$$\hat{A}(u) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} G_t(u) B_i \varepsilon_{it} + o_p(1).$$

Under $\mathbb{H}_0$, the asymptotic behavior of the empirical process $\hat{A}(u)$ depends on the relative speeds between N and T. When $\sqrt{N}/T \to 0$, i.e., $\nu > 1/2$, the leading term of $\hat{A}(u)$ is a weighted average of error term ε_{it}, and it will converge to a zero-spectrum in frequency domain. The intuition is that when T grows faster than $\sqrt{N}$, the Fourier transform dominates the asymptotic behavior of $\hat{A}(u)$. It is consistent with the result in Bai (2003) since the estimation impact on the factors depends on the relative speeds between N and T. And it includes the general case that N and T are the same order of magnitude. On the other hand, when $\sqrt{N}$ grows faster than T, $\hat{A}(u)$ will become a degenerate statistic. Although it still converges to a zero-spectrum at rate $T^{-3/2}$ under $\mathbb{H}_0$, the leading term now consists of two components, which are the same order of magnitude. The first component is the same as the leading term in the case of $\nu > 1/2$, while the second component arises due to serial correlation of the error term and it is asymptotically equivalent to a pseudo-covariance between Fourier series and the long-run variance. If we follow Su and Wang (2017) to impose the martingale difference sequence assumption for the error term, this second component of the leading term disappears. That is, if we rule out the serial dependence in the error term, then we do not need to impose any restriction on the relative order of N and T. However, since serial correlation is common in macroeconomic and financial data, we allow for serial correlation in the error term and impose a condition on the relative speeds between N and T when we derive the asymptotic distribution of our test statistic.

In addition, the results of this paper are built under the framework of large N and large T. Theoretically, the above relative speed between N and T also allows for the classical factor model (see Lawley and Maxwell, 1971; Anderson, 1984) or the approximate factor model (see Chamberlain and Rothschild, 1983) with large T and fixed N. However, as mentioned by Anderson (1984) and Bai (2003), with a fixed N, one can consistently estimate factor loadings but not the common factors. Since the estimated common factors is contained in the process $\hat{A}(u)$ and our test statistic given bellow, we do not consider this case. In addition, under the assumption of the martingale difference error term, our test is suitable for the case of fixed T and large N. If we follow Bai (2003) to further impose the asymptotic homoskedasticity condition that $\frac{1}{N}\sum_{i=1}^{N} \varepsilon_{it}^2 \to \sigma^2$ for all t as $N \to \infty$, then the estimated common factors are consistent and our test is applicable. However, if T is fixed and small, it is meaningless to consider the structural change problem. Hence, we rule out the cases of fixed T and fixed N in this paper.

Let V_{NT} be the $R \times R$ diagonal matrix of the first R largest eigenvalues of $\frac{1}{NT}XX'$, and define $H_0 \equiv \text{plim}_{N,T\to\infty} H$, where $H = (\frac{1}{N}\sum_{i=1}^{N} \lambda_{i0}\lambda_{i0}')(\frac{1}{T}\sum_{t=1}^{T} \hat{F}_t F_t')V_{NT}$. Under $\mathbb{H}_1$, the following proposition shows that $\hat{A}(u)$ will converge to a non-constant spectrum.

Proposition 3.2 *Suppose Assumptions A.1-A.3 hold, and under* $\mathbb{H}_1 : \lambda_{it} \neq \lambda_{i0}$ *for at least some* i. *Then* *as* $N, T \to \infty$

$$\sup_{u \in \mathbb{R}} \|\hat{A}(u) - \tilde{A}(u)\| \xrightarrow{p} 0,$$

where $\tilde{A}(u) = H_0' E(F_t F_t') \lim_{N,T\to\infty} \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T} e^{iu2\pi t/T}\lambda_{it} - \frac{1}{T}\sum_{t=1}^{T} e^{iu2\pi t/T}\frac{1}{T}\sum_{t=1}^{T}\lambda_{it}\right)$. *In particular, when* $\lambda_{it} = \lambda_i(\frac{t}{T})$, *it follows that* $\tilde{A}(u) = H_0' E(F_t F_t') \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \widetilde{cov}(e^{iu2\pi\tau}, \lambda_i(\tau))$, *where* $\widetilde{cov}(e^{iu2\pi\tau}, \lambda_i(\tau))$ *is a pseudo-covariance such that*

$$\widetilde{cov}[e^{iu2\pi\tau}, \lambda_i(\tau)] = \int_0^1 e^{iu2\pi\tau}\lambda_i(\tau)\, d\tau - \int_0^1 e^{iu2\pi\tau} d\tau \int_0^1 \lambda_i(\tau)\, d\tau.$$

We observe that $\hat{A}(u)$ is asymptotically equivalent to a pseudo-covariance between the Fourier series $e^{\mathbf{i}u2\pi t/T}$ and λ_{it}. When structural changes exist such that $\hat{A}(u) \neq 0$ for all u, $\hat{A}(u)$ can capture the time-varying behavior of the factor loading λ_{it} and will converge to a non-zero spectrum. Therefore, by checking the behavior of $\hat{A}(u)$ at each frequency u, we can capture possible structural changes in factor loadings.

Theorem 3.3 *Suppose Assumptions A.1-A.4 hold, and* $T \propto N^{\nu}$ *with* $\nu > 1/2$. *Then under* $\mathbb{H}_0 : \lambda_{it} = \lambda_{i0}$,

$$\hat{D} \xrightarrow{d} \int_{\mathbb{R}} \|\mathcal{G}(u)\|^2 W(u) du,$$

where $\mathcal{G}(u)$ *is a complex-valued Gaussian process with a covariance-kernel* $\mathcal{K}(u_1, u_2)$

$$\mathcal{K}(u_1, u_2) = \lim_{N,T\to\infty} \frac{1}{NT} \sum_{s,t=1}^{T} \sum_{i,j=1}^{N} B_i B_j H_0' E\left[F_t F_s' \varepsilon_{it}\varepsilon_{js}\right] H_0 M_t(u_1) M_s(u_2)^*,$$

and $M_t(u) = e^{iu2\pi t/T} - \int_0^1 e^{iu2\pi\tau} d\tau$ *is a demeaned Fourier process.*

Theorem 3.3 provides the asymptotic null distribution of the test statistic $\hat{D}$, which is robust to both serial correlation and cross-sectional dependence of unknown form. The condition on the relate speeds of magnitude of N and T simplifies the derivation of our asymptotic result. As shown in Proposition 3.1, if $T \propto N^{\nu}$ with $\nu \leq 1/2$, the leading term of $\hat{A}(u)$ will contain two components with the same order of magnitude, which will determine the asymptotic distribution jointly. For simplicity, we impose the condition that $\nu > 1/2$. Recall that the second component of the leading term rises due to the existence of temporal dependence in ε_{it}. Hence, if the error term is serially uncorrelated, the condition on the relative speeds between N and T is not necessary. We note that Bai (2003) also require $\nu > 1/2$. Breitung and Eickmeier (2011), Chen *et al.* (2014) and Han and Inoue (2015) all require $\nu < 2$. Su and Wang (2017) imposes an even stronger condition: $Th/N \to 0$. If $h = O(T^{-1/5}h^{-1/10})$, then it implies $\nu < 11/8$. Unlike the related works which impose restrictions on the upper bound ν, we impose a restriction on the low bound of ν, which is mild.

3.3 Asymptotic Local Power

To gain insight into the asymptotic power property of $\hat{D}$, we now consider a class of local alternatives:

$$\mathbb{H}_1(a_{NT}) : \lambda_{it} = \lambda_{i0} + a_{NT} g_{it} \text{ for each } i \text{ and } t,$$

where $a_{NT} \to 0$ as $(N, T) \to \infty$. The rate a_{NT} controls the speed at which the local alternative converges to the null hypothesis, and g_{it} is a deterministic function of time t for each i. We note that the local alternative $\mathbb{H}_1(a_{NT})$ does not impose any smoothness condition on the alternative. This setting is more general than Su and Wang's (2017) setting, in which they require g_{it} to be a piece-wise smooth function of scaled time ratio $\frac{t}{T}$ for each i.

Noting that $\lambda_{i0} + a_{NT} g_{it} = (\lambda_{i0} + c_{i,NT}) + a_{NT}[g_{it} - c_{i,NT}/a_{NT}]$ for any $c_{i,NT} \in \mathbb{R}^R$, we will assume below that

$$\frac{1}{T}\sum_{t=1}^{T} g_{it} = 0$$

for the purpose of location normalization. It turns out such a normalization greatly simplifies local asymptotic power analysis. Both λ_{i0} and g_{it} can depend on the sample sizes N and T. For notational simplicity, we continue to write them as λ_{i0} and g_{it}.

Theorem 3.4 *Suppose Assumptions A.1-A.4 hold, $T \propto N^{\nu}$ with $\nu > 1/2$, and $\frac{1}{T}\sum_{t=1}^{T} \|g_{it}\|^2 < \infty$ under $\mathbb{H}_1(a_{NT})$ with $a_{NT} = (NT)^{-1/2}$. Then as $N, T \to \infty$,*

$$\hat{D} \quad \xrightarrow{d} \quad \int_{\mathbb{R}} \|\xi(u) + \mathcal{G}(u)\|^2 W(u) du,$$

where

$$\xi(u) \quad = \quad H_0' E(F_t F_t') \lim_{N,T\to\infty} \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T} e^{iu2\pi t/T} g_{it} - \frac{1}{T}\sum_{t=1}^{T} e^{iu2\pi t/T} \frac{1}{T}\sum_{t=1}^{T} g_{it}\right).$$

Theorem 3.4 provides the asymptotic distribution of $\hat{D}$ under the local alternative $\mathbb{H}_1(a_{NT})$. It shows that our test can detect a class of local alternatives with $\xi(u) \neq 0$ for all u, at the rate $a_{NT} = T^{-1/2}N^{-1/2}$. In terms of Pitman's criterion, it is asymptotically more efficient than the smoothed nonparametric test of Su and Wang (2017), which could only detect the local alternative $\mathbb{H}_1(a_{NT})$ with a rate of $T^{-1/2}N^{-1/4}h^{-1/4}$. This is an advantage of DFT which avoids nonparametric smoothing over t/T. Strictly speaking, our test is not a consistent test. Because when the average of pseudo-covariances converge to zero, our test has no power. However, such case is really rare since it requires the DFTs for each cross-sectional unit i have to cancel with each other.

In addition, the specification of g_{it} allows for various kinds of structural changes in factor loadings, including smooth structural changes, a single structural break, the multiple structural breaks, or mixtures of abrupt and smooth changes. The case of a single structural break overlaps with the alternative hypothesis considered by Breitung and Eickmeier (2011), Chen *et al.* (2014) and Han and Inoue (2015). The previous parametric tests all reduce the infinite dimensional problem to a finite dimensional one in various ways. For example, Breitung and Eickmeier (2011) propose three test statistics for each i; Chen *et al.* (2014) run the regression of one estimated factor on the remaining ones and then test for the structural changes in such a linear regression by constructing the sup-Wald and sup-LM statistics of Andrews (1993); Han and Inoue (2015) construct their sup-Wald and sup-LM statistics by comparing the pre-and post- break subsample second moments of the estimated factors. All these test statistics have the same asymptotic distribution and convergence rate as the conventional sup-Wald statistic of Andrews (1993). They could only detect the local alternative that converge to the null at the rate $T^{-1/2}$, which is slower than our rate $a_{NT} = T^{-1/2}N^{-1/2}$. The proposed test has power against a wide range of structural changes, including abrupt and smooth structural changes, when the average of the noncentrality process over all individuals is not a zero function.

In fact, the order $a_{NT} = T^{-1/2}N^{-1/2}$ is the upper bound of structural changes in factor loadings that guarantees consistency of the estimated number of common factors and the estimated common factor by PCA. If the order of magnitude of structural changes is smaller than $T^{-1/2}N^{-1/2}$, then the estimated common factors and the number of factor loadings are consistent. This order of magnitude corresponds to the definition of small break by Chen *et al.* (2014). For such small structural changes, our test has no power. In contrast, if the order of magnitude of structural changes is larger than $T^{-1/2}N^{-1/2}$, the estimated common factors and the number of factor loadings will not be consistent. Thus, our test has nontrivial power to detect any structural changes that lead to inconsistent estimation of the number of common factors and the common factor given by PCA.

Finally, our test is tuning-parameter free. We require neither the smoothing parameter nor the trimming parameters. That is appealing in practice, because there has been no criteria to choose the optimal bandwidth for the nonparametric smoothing test of Su and Wang (2017) and the trimming parameter for the aforementioned parametric tests. In fact, the result of a smoothed nonparametric test can be largely affected by the choice of the smoothing parameter. Even if one uses the bootstrap, the power of nonparametric smoothing tests is still sensitive to the choice of a bandwidth. Moreover, when the sample size (N, T) is sufficiently large, the proposed test can detect any structural changes that occur close to the starting and ending points of the sample period, because we do not need to trim the data. In contrast, Breitung and Eickmeier (2011), Chen *et al.* (2014), Han and Inoue (2015), Yamamoto and Tanaka (2015) and Cheng *et al.* (2016) all rely on a prespecified tuning parameter τ to trim out the first and last τT observations in the sample and hence would miss the possible structural changes in the boundary regions.

3.4 A Bootstrap Version of the Test

The asymptotic distribution of $\hat{D}$ is not pivotal, and it depends on the unknown data generating process. We need to use some resampling methods to obtain the critical values in finite samples. To account for the possible serial correlation and cross-sectional dependence of unknown forms in the error term, we follow Gonçlaves (2011) and propose the following moving blocks bootstrap (MBB) procedure. Let

$l_T = l(T) \in \mathbb{N}(1 \leq l_T < T)$ be a block length such that $l_T \to \infty$ and $l_T/T \to 0$ as $T \to \infty$.

- Step (i). Estimate the model via the conventional PCA and obtain the estimated common factors and factor loadings $\{\hat{F}_t\}_{t=1}^T$ and $\{\hat{\lambda}_{i0}\}_{i=1}^N$. Then we obtain the residuals $\hat{\varepsilon}_{it} = X_{it} - \hat{\lambda}_{i0}'\hat{F}_t, i = 1, 2, \cdots, N; t = 1, 2, \cdots, T$. Compute the test statistic $\hat{D}$.

- Step (ii). Let $\bar{\varepsilon}$ be the $N \times T$ demeaned residual matrix with each (i,t)th element being $\bar{\varepsilon}_{it} = \hat{\varepsilon}_{it} - \frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T \hat{\varepsilon}_{it}$. Divide the column vectors of $\bar{\varepsilon}$ into $T - l_T + 1$ blocks and generate a block dataset $\{\Xi_t\}_{t=1}^{T-l_T+1}$, where $\bar{\varepsilon}_t = [\bar{\varepsilon}_{1t}, ..., \bar{\varepsilon}_{Nt}]'$ is a $N \times 1$ column vector, and $\Xi_t = [\bar{\varepsilon}_t, \bar{\varepsilon}_{t+1} ..., \bar{\varepsilon}_{t+l_T-1}]$ is a $N \times l_T$ matrix. Resample $\{\Xi_t\}_{t=1}^{T-l_T+1}$ with replacement to form a bootstrap data set $\{\Xi_t^b\}_{t=1}^L$ satisfying $L = \lfloor T/l_T \rfloor + 1$; Let $\{\varepsilon_t^b\}_{t=1}^T$ be the first T column vectors of $\{\Xi_t^b\}_{t=1}^L$.

- Step (iii). Generate a bootstrap sample $\{X_{it}^b\}_{i=1,t=1}^{N,T}$ such that $X_{it}^b = \hat{\lambda}_{i0}'\hat{F}_t + \varepsilon_{it}^b$, where ε_{it}^b is the ith element of ε_t^b. Run PCA on $\{X_{it}^b\}_{i=1,t=1}^{N,T}$ and compute the test statistic $\hat{D}^b$.

- Step (iv). Repeat Step (ii)-(iii) B times to obtain B bootstrap test statistics $\{\hat{D}^b\}_{b=1}^B$.

- Step (v). Compute the p-value for $\hat{D}$ with $\hat{p} = B^{-1}\sum_{b=1}^B \mathbf{1}(\hat{D}^b > \hat{D})$.

We reject $\mathbb{H}_0$ when $\hat{p}$ is smaller than a pre-specified significance level. Choosing an appropriate block length is crucial and many approaches have been proposed (*e.g.*, Lahiri, 1999) in the literature. In this paper, we adopt Politis and White's (2004) automatic block-length selection procedure. The the following simulation studies demonstrate the excellent finite sample performance of the proposed MBB approach. We note that the MBB procedure proposed by Gonçlaves (2011) requires $T \to \infty$ faster than N, i.e., $\nu > 1$. However, this does not affect the theoretical applicability of our test to the case with serially correlated errors.

4 Monte Carlo Simulations

We now study the finite sample performance of the proposed test through Monte Carlo simulations. We compare our test with the tests of Breitung and Eickmeier (2011), Chen *et al.* (2014) and Han and Inoue (2015) for a single structural break with an unknown break date in factor loadings and Su and Wang's (2017) nonparametric smoothing test.

4.1 Data Generating Process

We generate data under the framework of large factor models with $R = 2$ common factors:

$$X_{it} = \lambda_{it}'F_t + \varepsilon_{it},$$

where $i = 1, ..., N$, $t = 1, ..., T$, $F_t \equiv (F_{1,t}, F_{2,t})'$, with $F_{1,t} = 0.6F_{1,t-1} + u_{1t}$, $u_{1t} \sim i.i.d.N(0, 1 - 0.6^2)$; $F_{2,t} = 0.3F_{2,t-1} + u_{2t}$, $u_{2t} \sim i.i.d.N(0, 1 - 0.3^2)$.

To examine size and power, we consider the following setups for the factor loading $\lambda_{it} \equiv (\lambda_{it,1}, \lambda_{it,2})'$:

DGP.S1: $\lambda_{it} = \lambda_{i0} \sim i.i.d.\ N(0, \mathbb{I}_2)$;

DGP.P1: $\lambda_{it,k} = \begin{cases} \lambda_{i0,k}, & \text{for } t = 1, 2, \ldots, T/2 \\ \lambda_{i0,k} + 0.2, & \text{for } t = T/2 + 1, \ldots, T \end{cases}$, $\lambda_{i0,k} \sim i.i.d.N(1, 1)$ for $k = 1, 2$;

DGP.P2: $\lambda_{it,1} = \begin{cases} \lambda_{i0,1}, & \text{for } 0.1T < t \leq 0.2T \text{ or } 0.7T < t \leq 0.8T \\ \lambda_{i0,1} + 0.2, & \text{for } 0.4T < t \leq 0.5T \\ \lambda_{i0,1} - 0.2, & \text{otherwise} \end{cases}$, $\lambda_{it,2} = \lambda_{i0,2} \sim i.i.d.N(0, 1)$;

DGP.P3: $\lambda_{it,1} = \mu_i + 0.5G(10t/T; 0.1, (1, 3, 7, 9)')$, $\mu_i \sim i.i.d.\ N(0, 1)$, $\lambda_{it,2} = \lambda_{i0,2} \sim i.i.d.\ N(0, 1)$;
where $G(z; \kappa, \gamma) = \{1 + \exp[-\kappa \prod_{l=1}^{p} (z - \gamma_l)]\}^{-1}$ denotes the Logistic function with a scale parameter κ and a location parameter $\gamma = (\gamma_1, ..., \gamma_p)'$.

For each DGP, we consider five cases for the error term ε_{it}: (i) i.i.d. case, $\varepsilon_{it} \sim i.i.d.N(0, 1)$, (ii) heteroskedastic case, $\varepsilon_{it} = \sigma_i v_{it}$, $\sigma_i \sim i.i.d.U(0.5, 1.5), v_{it} \sim i.i.d.N(0, 1)$; (iii) cross sectional dependence case, $\varepsilon_{\cdot t} \sim i.i.d.N(0, \Sigma_e)$, (iv) time series dependence case, $\varepsilon_{it} = 0.5\varepsilon_{it-1} + v_{it}, v_{it} \sim i.i.d.N(0, 1)$; (v) cross sectional and time series dependence case, $\varepsilon_{\cdot t} = 0.5\varepsilon_{\cdot t-1} + v_{\cdot t}, v_{\cdot t} \sim i.i.d.N(0, \Sigma_e)$, , where $\Sigma_e = (c_{ij})_{i,j=1,\ldots,N}$ with $c_{ij} = 0.5^{|i-j|}$ for cases (iii), (iv) and (v).

DGP.S1 satisfies the null hypothesis of time-invariant factor loadings and is used to study the size of all tests. We examine the performance of the tests under i.i.d., heteroskedasticity, cross-sectional dependence, temporal dependence, and both the cross-sectional and temporal dependence, respectively. DGP.P1-P3 describe various time-varying factor loadings. Among them, DGP.P1-P2 have a single abrupt structural break and multiple abrupt structural breaks, respectively, while DGP.P3 is a smooth structural change. We check the power of all the tests by using DGP.P1-P3 with various types of error terms.

4.2 Test Statistics and Simulation Results

For each DGP, we simulate 500 data sets with sample sizes $N = 100$, 200, and $T = 100$, 200, respectively. In addition to our test, we also consider Breitung and Eickmeier's (2011) sup-LM N-variable specific test, Chen *et al.*'s (2014) sup-LM and sup-Wald tests, Han and Inoue's (2015) sup-LM and sup-Wald tests and Su and Wang's (2017) nonparametric test. Following Su and Wang (2017), we use the Epanechnikov kernel and Silverman's rule-of-thumb bandwidth $h = (2.35/\sqrt{12})T^{-1/5}N^{-1/10}$ for Su and Wang's (2017) test. We set the trimming parameter $\tau = 0.15$ for the parametric tests, which is a common choice in the literature. We also examine the performance of these tests with $\tau = 0.1$ and 0.25 and find the results are quite similar. The tests of Chen *et al.* (2014) and Han and Inoue (2015) involve long-run variance estimation. We follow the HAC literature by setting the truncation parameter $m = \lfloor T^{1/3} \rfloor$ and choosing the Bartlett kernel to estimate the long-run variance. The critical values presented in Andrews (1993) are used for the tests of Breitung and Eickmeier (2011), Chen *et al.* (2014) and Han and Innoe (2015). We apply bootstrap procedures for Su and Wang's (2017) test and our test. We set the number of bootstrap $B = 200$.

Table 1 reports the sizes of our test as well as the tests of Breitung and Eickmeier (2011), Chen *et al.* (2014), Han and Inoue (2015) and Su and Wang (2017) at the 5% and 10% significance levels when the number of common factors are fixed and the true value $R = 2$. As shown in Table 1, our test has reasonable

Table 1: Size of tests under DGP.S1 when the number of factors is fixed to the true value

N	T	D_B		SW		HI_{LM}		HI_W		CDG_{LM}		CDG_W		BE_{LM}	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
i.i.d. error term: $\varepsilon_{it} \sim i.i.d.N(0,1)$															
100	100	4.6	9.2	5.6	12.6	1.0	4.4	0.2	1.4	2.0	6.2	5.2	12.8	2.7	6.3
100	200	4.0	10.8	6.6	12.8	2.2	7.0	2.2	6.6	3.4	8.0	4.8	11.6	3.5	7.5
200	100	5.6	11.0	5.0	11.2	1.2	5.0	0.6	2.6	2.2	6.2	7.6	11.8	2.8	6.4
200	200	5.8	10.8	5.6	11.4	3.4	7.6	3.0	8.6	2.4	7.2	6.6	9.8	3.3	7.3
heteroskedastic error term: $\varepsilon_{it} = \sigma_i v_{it}$, $\sigma_i \sim i.i.d.U(0.5, 1.5), v_{it} \sim i.i.d.N(0,1)$															
100	100	5.2	10.6	5.4	12.0	0.8	4.4	0.2	1.6	1.4	7.0	5.6	12.2	2.8	6.3
100	200	5.4	11.4	6.8	14.6	2.2	7.2	2.4	6.8	3.4	8.6	4.6	11.2	3.5	7.4
200	100	4.8	11.0	5.6	10.0	1.4	5.2	0.6	2.6	2.0	5.8	7.2	11.8	2.8	6.4
200	200	5.0	10.2	5.4	11.8	3.2	7.6	3.0	8.4	2.0	7.0	5.8	9.2	3.3	7.3
cross sectional dependence error term: $\varepsilon_{\cdot t} \sim i.i.d.N(0, \Sigma_e)$															
100	100	5.8	10.2	5.4	10.2	1.0	4.0	0.2	1.0	1.6	7.0	6.6	12.8	2.7	6.3
100	200	5.0	10.8	4.2	8.8	2.0	6.4	2.0	4.6	2.0	7.2	5.0	10.6	3.4	7.4
200	100	8.0	12.2	5.8	12.4	1.6	6.0	0.8	3.0	1.6	5.8	7.0	12.0	2.8	6.4
200	200	5.4	9.2	5.4	11.2	3.2	7.6	3.4	8.0	2.6	7.2	5.8	9.4	3.5	7.6
time series dependence error term: $\varepsilon_{it} = 0.5\varepsilon_{it-1} + v_{it}, v_{it} \sim i.i.d.N(0,1)$															
100	100	7.6	13.8	100	100	1.0	3.2	0.2	1.8	1.6	7.8	6.0	10.8	13.6	22.7
100	200	4.0	10.4	100	100	2.6	6.8	2.4	8.2	4.2	9.4	6.4	12.6	17.6	27.8
200	100	5.2	13.0	100	100	1.6	3.8	0.8	2.6	2.2	7.2	6.8	11.8	13.3	22.6
200	200	6.0	11.4	100	100	3.4	7.2	2.4	9.4	4.8	10.8	6.4	12.0	17.7	27.8
cross sectional and time series dependence error term: $\varepsilon_{\cdot t} = 0.5\varepsilon_{\cdot t-1} + v_{\cdot t}, v_{\cdot t} \sim i.i.d.N(0, \Sigma_e)$															
100	100	6.8	13.0	100	100	0.4	3.2	0.4	2.0	2.4	8.0	4.4	11.2	13.6	22.5
100	200	7.2	12.6	99.8	100	2.4	6.0	3.4	8.0	3.2	8.2	4.8	9.4	18.1	28.4
200	100	6.6	12.4	100	100	1.2	3.8	0.4	2.0	2.0	5.4	6.0	12.0	13.9	23.1
200	200	7.0	13.4	100	100	3.2	6.8	2.6	9.6	4.8	11.6	6.8	12.4	17.8	27.8

Notes: (i) D_B denotes the results of our D test using bootstrap critical values; (ii) SW denotes the results of Su and Wang's (2017) bootstrap-based test; (iii) HI_{LM} and HI_W denote Han and Inoue's (2015) sup-LM and Wald tests; (iv) CDG_{LM} and CDG_W denote Chen *et al.*'s (2014) sup-LM and Wald tests; (v) BE_{LM} denotes Breitung and Eickmeier's (2011) N variable-specific sup-LM test. The main entries report the average percentage of rejections.

sizes using bootstrap critical values. For the case of both cross-sectional and temporal dependence, our test tends to over-reject a bit but is still acceptable. Han and Inoue's (2014) sup-LM and sup-Wald tests tend to under-reject. Chen *et al.*'s (2014) sup-Wald test has reasonable sizes, but their sup-LM test also exhibits under-rejection. Su and Wang's (2017) test tends to over-reject slightly, but is still acceptable for the first three cases. However, when the error term has serial correlation, it displays serious over-rejection and the rejection rates even achieve 100%. It is not surprising to see the poor size performance of Su and Wang's (2017) test since it requires the error term to be a martingale difference sequence. On the other hand, Breitung and Eickmeier's (2011) test suffers from slight under-rejection for the first three cases, and severe over-rejection for the last two cases. In fact, Assumption 2 in Breitung and Eickmeier (2011) requires that the error term be serially independent, which does not hold in the last two cases.

Table 2: Power of tests under DGP.P1-P2 when the number of factors is fixed to the true value

	N	T	D_B		SW		HI_{LM}		HI_W		CDG_{LM}		CDG_W		BE_{LM}	
			5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
i.i.d. error term: $\varepsilon_{it} \sim i.i.d.N(0,1)$																
DGP.P1	100	100	99.0	99.4	71.0	80.2	1.6	5.6	0.0	2.0	2.2	7.6	6.0	12.6	5.4	10.8
	100	200	100	100	97.6	99.0	4.6	9.8	5.6	12.2	3.6	7.6	6.0	11.4	10.5	17.7
	200	100	100	100	88.6	93.2	1.4	5.4	1.2	3.2	2.4	9.0	7.6	11.2	5.4	10.7
	200	200	100	100	100	100	5.4	12.6	5.6	12.8	4.0	8.8	3.6	10.6	10.7	17.9
DGP.P2	100	100	26.2	37.4	10.2	17.8	1.2	4.4	0.2	1.6	1.8	7.6	4.0	10.2	2.9	6.6
	100	200	62.6	76.6	21.4	31.0	2.2	7.4	2.2	5.8	3.8	8.6	4.8	9.4	3.9	8.3
	200	100	42.0	59.4	11.4	20.0	1.2	5.4	0.4	3.0	3.2	8.4	6.6	10.4	2.9	6.7
	200	200	87.2	93.4	24.6	37.4	3.4	8.0	2.4	8.2	4.2	9.2	2.8	9.8	3.8	8.2
heteroskedastic error term: $\varepsilon_{it} = \sigma_i v_{it}$, $\sigma_i \sim i.i.d.U(0.5, 1.5), v_{it} \sim i.i.d.N(0,1)$																
DGP.P1	100	100	99.0	99.8	73.6	81.4	1.6	5.6	0.0	2.0	2.2	7.6	5.6	12.2	6.7	12.3
	100	200	100	100	98.2	98.6	4.6	9.8	5.6	12.2	3.6	7.8	6.4	11.6	13.2	20.5
	200	100	100	100	91.4	95.8	1.4	5.2	1.2	3.0	2.6	9.0	7.4	11.6	6.7	12.2
	200	200	100	100	100	100	5.8	13.0	6.0	12.8	4.0	9.2	3.8	10.4	13.5	20.7
DGP.P2	100	100	26.6	40.6	10.2	18.8	1.2	4.2	0.2	1.6	1.8	7.8	3.8	10.2	2.9	6.6
	100	200	62.8	75.4	21.2	33.8	2.6	6.8	2.4	6.0	4.2	8.6	4.8	10.2	4.1	8.6
	200	100	45.6	62.8	13.2	20.2	1.4	5.4	0.4	3.2	3.2	9.2	6.4	10.4	3.0	6.8
	200	200	86.4	92.6	28.2	39.4	3.2	7.6	2.4	8.2	4.0	9.6	3.0	9.8	4.0	8.5
cross sectional dependence error term: $\varepsilon_{\cdot t} \sim i.i.d.N(0, \Sigma_e)$																
DGP.P1	100	100	98.6	99.4	66.2	76.4	1.2	4.6	0.0	1.2	2.4	7.4	5.6	13.0	5.4	10.6
	100	200	100	100	97.6	98.6	5.2	9.8	4.4	12.4	3.0	8.4	5.6	10.8	10.9	18.0
	200	100	100	100	88.0	94.6	1.4	5.4	1.0	4.0	2.2	8.4	6.4	11.4	5.5	10.8
	200	200	100	100	100	100	5.8	13.2	5.8	12.0	4.4	8.8	4.6	10.6	10.9	18.1
DGP.P2	100	100	26.4	38.0	10.4	17.2	1.2	5.0	0.0	0.8	2.6	7.2	4.4	10.2	2.9	6.6
	100	200	62.8	77.6	17.4	27.4	2.6	7.4	1.6	6.6	4.0	9.2	4.6	8.8	3.8	8.1
	200	100	40.6	56.4	12.4	19.8	1.6	5.0	0.6	3.2	2.2	7.8	5.8	10.2	3.0	6.7
	200	200	88.6	94.6	27.6	42.8	3.4	7.8	3.2	7.8	4.2	9.2	2.6	9.8	4.0	8.5
time series dependence error term: $\varepsilon_{it} = 0.5\varepsilon_{it-1} + v_{it}, v_{it} \sim i.i.d.N(0,1)$																
DGP.P1	100	100	89.6	93.2	100	100	1.2	5.0	0.4	3.4	1.6	6.0	4.6	10.8	16.4	26.1
	100	200	100	100	100	100	5.4	10.4	5.4	14.0	4.6	9.6	5.6	11.6	24.5	35.4
	200	100	100	100	100	100	1.8	4.8	0.8	3.2	2.0	6.4	5.0	12.2	16.4	26.1
	200	200	100	100	100	100	5.8	12.6	5.0	15.0	5.8	9.4	7.4	12.0	24.6	35.6
DGP.P2	100	100	16.4	26.4	99.0	99.4	0.8	4.4	0.2	1.8	1.4	6.2	3.2	9.8	13.7	22.8
	100	200	41.2	58.8	100	100	2.8	7.4	2.8	6.6	3.8	10.4	4.0	11.2	18.2	28.3
	200	100	24.4	38.0	99.8	99.8	1.6	4.6	0.4	2.0	1.4	6.2	3.4	8.8	13.5	22.6
	200	200	65.2	79.8	100	100	3.2	7.4	2.6	8.4	4.6	10.8	4.4	10.2	18.2	28.4
cross sectional and time series dependence error term: $\varepsilon_{\cdot t} = 0.5\varepsilon_{\cdot t-1} + v_{\cdot t}, v_{\cdot t} \sim i.i.d.N(0, \Sigma_e)$																
DGP.P1	100	100	90.2	94.6	100	100	1.0	4.8	0.4	3.0	2.6	7.6	4.4	10.2	16.4	26.2
	100	200	100	100	100	100	5.0	10.8	5.6	13.8	4.2	8.8	6.0	11.6	25.0	35.8
	200	100	100	100	100	100	1.6	4.4	0.6	3.2	2.0	5.6	4.6	12.4	16.8	26.6
	200	200	100	100	100	100	4.8	12.4	5.2	14.4	5.2	9.8	7.0	12.2	24.6	35.4
DGP.P2	100	100	14.8	23.6	99.8	100	0.8	3.2	0.4	1.6	1.8	6.8	3.8	8.8	13.5	22.7
	100	200	39.4	53.2	100	100	2.6	7.2	3.2	8.2	4.4	10.4	4.6	11.0	18.5	29.0
	200	100	25.6	40.4	99.8	100	1.2	3.8	0.4	2.2	1.4	6.2	4.0	9.0	14.0	23.0
	200	200	65.8	78.8	100	100	2.8	6.6	2.2	8.4	4.6	10.8	5.0	10.4	18.2	28.5

Note: See the notes in Table 1.

Tables 2 and 3 report the power performance of the tests under DGP.P1-P3 at the 5% and 10% significance levels when the number of common factors is fixed and the true value $R = 2$. Our test is most powerful in detecting all forms of time-varying factor loadings given by DGP.P1-P3 and its power increases as either T or N increases. Recall that DGP.P1-P2 are factor models with abrupt structural breaks, while DGP.P3 is the factor model with smooth structural changes. The simulation results demonstrate the excellent performance of our test in detecting both a finite number of abrupt structural breaks and smooth structural changes. Moreover, Su and Wang's (2017) test is also powerful in detecting all these DGPs, but the rejection rates are all lower than our new test except for the last two cases of serially dependent error term. Since Su and Wang's (2017) test could even achieve unity rejection under DGP.S1, it is not surprise to see its high rejection rate for DGP.P1-P3 when the error term has serial dependence. The results for other cases are consistent with our analysis on the relative efficiency between our test and Su and Wang's (2017) test. In contrast, Han and Inoue's (2015) sup-LM and sup-Wald tests, Chen *et al.*'s (2014) sup-LM and sup-Wald tests and Breitung and Eickmeier's (2011) N-variable-specific sup-LM test all have relatively quite low power against DGP.P1-P3, which exhibit either abrupt structural breaks or smooth structural changes in factor loadings.

As the exact number R of common factors is typically unknown in practice, one should determine the number of common factors before estimating and testing. In the literature on testing for structural breaks in factor loadings, the number of common factors is either determined by Bai and Ng's (2002, BN hereafter) information criteria (*e.g.*, Han and Inoue, 2015) or specified by some fixed numbers, which may be equal to, less than, or greater than the correct number of factors (*e.g.*, Chen *et al.*, 2014). Of course, one can also consider applying the testing procedures of Onatski (2009, 2010) or Ahn and Horenstein (2013) to determine the number of factors, which have been shown to work well in the presence of moderate or strong cross-sectional dependence. Alternatively, one can apply Su and Wang's (2017, SW hereafter) nonparametric method to determine the number of factors that is robust to the presence of structural changes in factor loadings. In general, all the aforementioned methods can select the correct number of factors consistently under the null hypothesis of no structural change, but only SW's method has been proven valid even under the alternative. Indeed, if we apply SW's method to determine the number of factors, the size and power performance of all tests will be similar to those in Tables 1 and 2. To allow the possible misspecification of the number of factors under the alternative, we follow Han and Inoue (2015) and select the number of factors based on BN's information criteria IC_{p1} and IC_{p2}. The simulation results based on IC_{p1} and IC_{p2} are also similar to those reported in Tables 1 and 3. In fact, for all DGPs, our simulation studies show that Bai and Ng's (2002) IC_{p1} only tends to overparameterize slightly, and the problem alleviates as the sample size increases. To save space, we relegate the results based on IC_{p1} to the online supplement. Moreover, we also examine the performance of the proposed test as well as other various tests by setting the number of common factors as 3. The power of our bootstrap-based test is a bit lower than in the case of correctly specified factors as reported in Tables 2 and 3. However, our test still has reasonable power that increases as either T or N increases. More importantly, it is still the most powerful test among all the tests under consideration. For space, we do not report the results for this case here.

Table 3: Power of tests under DGP.P3 when the number of factors is fixed to the true value

	N	T	D_B		SW		HI_{LM}		HI_W		CDG_{LM}		CDG_W		BE_{LM}	
			5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
			i.i.d. error term: $\varepsilon_{it} \sim i.i.d.N(0,1)$													
DGP.P3	100	100	82.8	91.8	36.8	48.2	0.6	3.6	0.4	2.4	2.2	8.0	5.8	11.8	3.7	8.3
	100	200	99.6	100	72.0	80.8	1.6	5.6	3.8	9.8	4.2	8.6	4.6	11.8	5.6	11.1
	200	100	91.2	97.2	42.8	54.2	1.6	4.6	1.0	3.4	1.8	4.6	10.0	15.8	3.7	8.2
	200	200	99.8	100	90.4	93.6	2.0	6.0	4.0	11.6	2.8	7.4	7.6	13.4	5.7	11.4
			heteroskedastic error term: $\varepsilon_{it} = \sigma_i v_{it}$, $\sigma_i \sim i.i.d.U(0.5, 1.5), v_{it} \sim i.i.d.N(0,1)$													
	100	100	80.0	91.4	31.2	43.2	0.6	3.2	0.4	2.6	2.4	8.4	5.6	12.4	4.2	8.9
	100	200	98.8	99.8	63.6	73.4	1.4	5.6	3.8	9.8	4.0	8.4	4.8	11.6	6.8	12.5
	200	100	87.6	97.2	38.0	49.8	1.6	4.8	1.0	3.4	1.6	4.6	10.0	15.2	4.1	8.7
	200	200	99.6	100	83.2	90.6	2.0	6.2	4.0	11.4	3.0	7.4	7.6	13.6	6.9	12.9
			cross sectional dependence error term: $\varepsilon_{\cdot t} \sim i.i.d.N(0, \Sigma_e)$													
	100	100	83.0	91.6	31.2	43.4	0.8	3.8	0.6	2.4	1.8	8.8	8.0	13.4	3.6	7.9
	100	200	99.2	100	74.0	82.4	1.4	5.4	2.6	9.0	3.6	8.6	5.4	11.4	5.6	11.2
	200	100	92.2	97.8	45.6	57.8	1.4	5.4	1.0	3.6	1.8	5.6	10.6	15.6	3.8	8.1
	200	200	100	100	92.2	95.4	1.8	5.6	3.6	11.0	2.8	7.4	8.2	13.6	5.9	11.6
			time series dependence error term: $\varepsilon_{it} = 0.5\varepsilon_{it-1} + v_{it}, v_{it} \sim i.i.d.N(0,1)$													
	100	100	57.2	74.8	99.2	100	0.4	3.4	0.2	3.4	1.8	8.0	6.4	12.8	14.6	24.1
	100	200	93.2	97.0	100	100	2.2	5.6	4.0	10.2	2.8	9.0	5.0	11.6	20.4	31.1
	200	100	80.2	90.2	99.8	100	1.4	3.6	0.8	2.8	1.6	6.0	7.2	14.8	14.6	24.1
	200	200	99.0	99.8	100	100	2.2	5.2	2.8	11.8	2.6	7.4	9.0	15.2	20.4	31.3
			cross sectional and time series dependence error term: $\varepsilon_{\cdot t} = 0.5\varepsilon_{\cdot t-1} + v_{\cdot t}, v_{\cdot t} \sim i.i.d.N(0, \Sigma_e)$													
	100	100	58.4	76.4	99.4	100	0.4	2.2	0.4	3.4	1.8	7.4	5.8	14.4	14.7	21.2
	100	200	92.8	97.4	100	100	1.8	5.6	3.8	8.4	2.8	8.6	5.0	11.0	21.0	31.7
	200	100	79.6	90.8	99.6	100	0.8	3.8	0.2	2.6	1.8	5.6	8.4	13.6	15.0	24.5
	200	200	99.0	99.6	100	100	1.8	4.8	3.2	11.4	2.8	7.8	7.4	15.4	20.6	31.3

Note: See the notes in Table 1.

Table 4: Tests of structural changes in the U.S. economy

Number of selected factors	1	2	3	4
Criterion functions	Ona, ER, GR	IC_{h1}, IC_{h2}	PC_{p1}, PC_{p2}	IC_{p1} ,IC_{p2}

Notes: (i) PC_{p1}, PC_{p2}, IC_{p1} and IC_{p2} denote Bai and Ng's (2002) information criteria; (ii) Ona denotes the results of Onatski's (2009) test; (iii) ER and GR denote Ahn and Horenstein's (2013) criteria; (iv) IC_{h1} and IC_{h2} denote the information criteria proposed by Su and Wang (2017).

Table 5: Tests for Structural Changes for the U.S. Economy

	D_B		SW, $c = 0.5$		SW, $c = 1$		SW, $c = 2$		Han and Inoue (2015)			Chen *et al.* (2014)		
	D_B	5%	SM_B	5%	SM_B	5%	SM_B	5%	LM	Wald	5%	LM	Wald	5%
$r = 1$	**11.27**	5.48	**-3.20**	-7.45	**6.15**	1.59	**10.30**	1.74	**11.43**	6.49	8.85	-	-	-
$r = 2$	**10.63**	5.52	-3.60	-0.12	**18.56**	8.50	**32.06**	5.88	**21.26**	10.05	14.15	4.67	1.89	8.85
$r = 3$	**14.70**	8.02	0.96	6.53	**32.57**	16.60	**53.74**	10.70	**25.43**	12.69	20.26	3.26	10.45	11.79
$r = 4$	**12.68**	8.03	1.70	11.77	**32.11**	23.56	**53.63**	14.46	**28.79**	**27.37**	27.03	**24.39**	**24.13**	14.15

Notes (i) D_B denotes the results of our D test using bootstrap critical value based on $B = 1000$ iterations; (ii) 5% and 10% denote the corresponding significance level.

5 An Empirical Application to U.S. Macroeconomic Data Set

We now apply our test to check whether the U.S. macroeconomic dynamics suffers from structural changes. The data set, firstly constructed by Stock and Watson (2012), and then extended by Cheng *et al.* (2016), consists of $N = 102$ series of monthly macroeconomic and financial indicators, spanning from 1985:M1 to 2013:M1 ($T = 337$). All the data have been standardized to have zero mean and unit variance. For the details of the data description and processing, one can refer to Stock and Watson (2012) and Cheng *et al.* (2016).

We first determine the appropriate number of common factors. The maximum number of common factors is set to be 8 in this empirical study. We use Bai and Ng's (2002) information criteria $PC_{p1}, PC_{p2}, IC_{p1}, IC_{p2}$, Onatski's (2009) testing procedure, Ahn and Horenstein's (2013) criterion functions ER and GR and Su and Wang's (2017) local information criteria IC_{h1}, IC_{h2} to determine the number of common factors. The results are reported in Table 4, where we see that different methods choose deliver different numbers of common factors. Below, we report the test results for the cases of one to four common factors respectively.

We apply our test $\hat{D}_B$, Su and Wang's (2017) nonparametric test SW, Han and Innoue's (2014) sup-LM and sup-Wald tests, as well as Chen *et al.*'s (2014) sup-LM and sup-Wald tests to investigate the possible structural changes in factor loadings. For Su and Wang's (2017) test, we choose the bandwidth $h = ch^*$ with $h^* = (2.35/\sqrt{12})T^{-1/5}N^{-1/10}$ given in their paper. By choosing $c = 0.5, 1, 2$, we consider the effect of different bandwidths on the results of Su and Wang's (2017) test. The other settings, including the kernel functions and tuning parameters, are all the same to those used in our simulation studies. For our test and Su and Wang's (2017) test, we focus on the bootstrap results based on $B = 1000$ bootstrap replications.

Table 5 reports the results of various tests and the corresponding critical values at the 5% significant level. Our test clearly rejects the null hypothesis of no structural changes for all the cases of one to four common factors. Su and Wang's (2017) results are sensitive to the choice of bandwidth. By using different bandwidths, different results arise, and so the evidence is mixed. Moreover, Chen *et al.*'s (2014) sup-LM and sup-Wald tests can only reject the null for the case of $R = 4$, while Han and Inoue's (2015) results are mixed. Their sup-LM test rejects the null hypothesis for all cases while the sup-Wald test can only reject the null hypothesis for the case of four common factors. This result is consistent with our simulation studies that indicate the relative low power of the tests given by Chen *et al.* (2014) and Han and Inoue (2014).

6 Conclusion

Conventional factor models assume factor loadings, which capture the relationship between observed random variables and the latent common factors, to be time-invariant. In fact, since macroeconomic data usually have a long time span, it is difficult to assume that factor loadings are constant over time. In this paper, we propose a new test for structural changes in large dimensional factor models via a discrete Fourier transform approach. Compared to FHW, our test is constructed in large dimensional factor models where the regressors are unobservable. By construction, our test can capture a wide range of smooth and abrupt structural changes in factor loadings with unknown break dates and unknown number of breaks. More importantly, the proposed test is asymptotically more powerful than all the existing related tests in the literature. Our test is tuning parameter-free, and it is robust to serial correlation and cross-sectional dependence of unknown forms, which greatly extends the scope of applicability of our test. Simulation studies show that in comparison with the tests of Breitung and Eickmeier (2011), Chen *et al.* (2014), Han and Inoue (2015) and Su and Wang (2017), the proposed test has both reasonable size and excellent power against various alternatives in finite samples. We apply our test to check whether the U.S. macroeconomic dynamics suffers from structural changes, and document significant evidence to against the time invariance property of factor loadings.

There are several interesting topics for further research. For instance, when our test rejects the null hypothesis, one can further check the type of structural changes, *i.e.*, distinguishing smooth structural changes from abrupt structural breaks. That is an interesting and challenging issue, but it is out of scope of the present paper. We will leave it to subsequent studies.

References

Ahn, S. & A. Horenstein (2013) Eigenvalue ratio test for the number of factors. *Econometrica* 81, 1203-1227.

Andrews, W.K. (1993) Tests for parameter instability and structural change with unknown change point. *Econometrica* 61, 821-856.

Bai, J. (2003) Inferential theory for factor models of large dimensions. *Econometrica* 71, 135-172.

Bai, J. & S. Ng (2002) Determining the number of factors in approximate factor models. *Econometrica* 90, 191-221.

Banerjee, A., M. Marcellino, & I. Masten (2008) Forecasting macroeconomic variables using diffusion indexes in short samples with structural change. *Frontiers of Economics and Globalization* 3, 149-194.

Bates, B.J., M. Plagborg-Møller, J.H. Stock, & M.W. Watson (2013) Consistent factor estimation in dynamic factor models with structural instability. *Journal of Econometrics* 177, 289-304.

Breitung, J. & S. Eickmeier (2011) Testing for structural breaks in dynamic factor models. *Journal of Econometrics* 163, 71-84.

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

Corradi, V. & N.R. Swanson (2014) Testing for structural stability of factor augmented forecasting models. *Journal of Econometrics* 182, 100-118.

Cheng, X., Z. Liao, & F. Schorfheide (2016) Shrinkage estimation of high-dimensional factor models with structural instabilities. *Review of Economic Studies* 83, 1511-1543.

Eickmeier, S., W. Lemke, & M. Marcellino (2015) Classical time varying factor-augmented vector auto-regressive models-estimation, forecasting and structural analysis. *Journal of Royal Statistical Society: Series A (Statistics in Society)* 178, 493-533.

Fu, Z., Y. Hong, & X. Wang (2018) Consistent testing for structural changes in time series models via discrete fourier transform. *Working paper.*

Gonçalves, S. (2011) The moving blocks bootstrap for panel linear regression models with individual fixed effects. *Econometric Theory* 27, 1048-1082.

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

Hong, Y., X. Wang & S. Wang (2017) Tests strict stationarity with applications to macroeconomic time series. *International Economic Review* 58, 1227-1277.

Lahiri, S. (1999) Theoretical comparisons of block bootstrap methods. *Annals of Statistics*, 27, 386-404.

Ma, S. & L. Su (2017) Estimation of large dimensional factor models with an unknown number of breaks. *Working Paper,* Singapore Management University.

Moon, H. & M. Weidner (2015) Linear regression for panel with unknown number of factors as interactive fixed effects. *Econometrica* 83, 1543-1579.

Onatski, A. (2009) Testing hypotheses about the number of factors in large factor models. *Econometrica* 77, 1447-1479.

Politis, D. & H. White (2004) Automatic block-length selection for the dependent bootstrap. *Econometric Reviews* 23(1), 53-70.

Stock, J.H. & M.W. Watson (2002) Forecasting using principal components from a large number of predictors. *Journal of American Statistical Association* 97, 1167-1179.

Stock, J.H. & M.W. Watson (2009) Forecasting in dynamic factor models subject to structural instability, In J. Castle & N. Shephard (eds.), *Methodology and Practice of Econometrics, A Festschrift in Honour of Professor David F. Hendry*, pp. 1-57, Oxford University Press, Oxford.

Stock, J.H. & M.W. Watson (2012) Disentangling the channels of the 2007-09 recession, *Brookings Papers on Economic Actibity*, 81-156.

Su, L. & X. Wang (2017) On time varying factor models: estimation and testing. *Journal of Econometrics* 198, 84-101.

Yamamoto, Y. & S. Tanaka (2015) Testing for factor loading structural change under common breaks. *Journal of Econometrics* 189, 187-206.

Zhang, T. & W. Wu (2011) Testing parametric assumptions of trends of a nonstationary time series. *Biometrica* 98, 599-614.

Online Supplement for
"Testing for Structural Changes in Factor Models via Discrete Fourier Transform"

Zhonghao Fu[a], Yongmiao Hong[b], Xia Wang[c]
[a] Fudan University, China
[b] Cornell University, U.S. and Xiamen University, China
[c] Sun Yat-Sen University, China

This Online Supplement contains two appendices. Appendix A is a mathematical appendix that contains some technical lemmas and the proofs of the theorems and lemmas in the paper. Appendix B contains some additional simulation results.

A Mathematical Appendix

Notations: Denote $\gamma_N(s,t) = E(\varepsilon_s'\varepsilon_t/N) = E(\frac{1}{N}\sum_{i=1}^N \varepsilon_{is}\varepsilon_{it})$, $\zeta_{st} = \frac{1}{N}\varepsilon_s'\varepsilon_t - \gamma_N(s,t)$, $\eta_{st} = F_s'\Lambda_0'\varepsilon_t/N$, $\xi_{st} = F_t'\Lambda_0'\varepsilon_s/N$. Let V_{NT} denote the $R\times R$ diagonal matrices of the first R largest eigenvalues of $(NT)^{-1}XX'$ in decreasing order and $H = (\Lambda_0'\Lambda_0/N)(F'\hat{F}/T)V_{NT}^{-1}$. Let $C_{NT} = \min\{\sqrt{N}, \sqrt{T}\}$.

A.1 Technical Lemmas

Lemma A.1 *Suppose Assumptions A.1 to A.4 hold,*

$$\frac{1}{T}\sum_{t=1}^{T}\left(\hat{F}_t\hat{F}_t' - \mathbb{I}_R\right)e^{iu2\pi t/T},$$

and

$$\frac{1}{T}\sum_{t=1}^{T}\left(\hat{F}_tF_t' - E(\hat{F}_tF_t')\right)e^{iu2\pi t/T}$$

are stochastically equicontinuous.

Proof. We first show that $\frac{1}{T}\sum_{t=1}^{T}\left(\hat{F}_t\hat{F}_t' - \mathbb{I}_R\right)e^{\mathbf{i}u2\pi t/T}$ is stochastically equicontinuous, i.e., we need to show that, for any $\epsilon > 0$ and $\kappa > 0$, there exists $\delta > 0$ such that

$$\lim_{T\to\infty} P\left[\sup_{u_1,u_2\in\mathbb{R}:|u_1-u_2|<\delta}\left\|\frac{1}{T}\sum_{t=1}^{T}\left(\hat{F}_t\hat{F}_t' - \mathbb{I}_R\right)\left(e^{\mathbf{i}u_12\pi t/T} - e^{\mathbf{i}u_22\pi t/T}\right)\right\| > \kappa\right] < \epsilon.$$

Let $\bar{u} = au_1 + (1-a)u_2$ for some $a\in(0,1)$ and

$$e^{\mathbf{i}u_12\pi t/T} = e^{\mathbf{i}u_22\pi t/T} + \mathbf{i}2\pi t/Te^{\mathbf{i}\bar{u}2\pi t/T}(u_1-u_2),$$

then

$$\begin{aligned}
& \lim_{T\to\infty} P\left[\sup_{u_1,u_2\in\mathbb{R}:|u_1-u_2|<\delta}\left\|\frac{1}{T}\sum_{t=1}^{T}\left(\hat{F}_t\hat{F}_t'-\mathbb{I}_R\right)\left(e^{\mathbf{i}u_1 2\pi t/T}-e^{\mathbf{i}u_2 2\pi t/T}\right)\right\|>\kappa\right] \\
= & \lim_{T\to\infty} P\left[\sup_{u_1,u_2\in\mathbb{R}:|u_1-u_2|<\delta}\left\|\frac{1}{T}\sum_{t=1}^{T}\left(\hat{F}_t\hat{F}_t'-\mathbb{I}_R\right)\left(\mathbf{i}2\pi t/Te^{\mathbf{i}\bar{u}2\pi t/T}\right)(u_1-u_2)\right\|>\kappa\right] \\
\leq & \lim_{T\to\infty} P\left[\sup_{\bar{u}\in\mathbb{R}}\left\|\frac{1}{T}\sum_{t=1}^{T}\left(\hat{F}_t\hat{F}_t'-\mathbb{I}_R\right)\left(\mathbf{i}2\pi t/Te^{\mathbf{i}\bar{u}2\pi t/T}\right)\right\|>\kappa/\delta\right] \\
\leq & \lim_{T\to\infty} P\left[\sup_{\bar{u}\in\mathbb{R}}\frac{1}{T}\sum_{t=1}^{T}\left\|\left(\hat{F}_t\hat{F}_t'-\mathbb{I}_R\right)\left(\mathbf{i}2\pi t/Te^{\mathbf{i}\bar{u}2\pi t/T}\right)\right\|>\kappa/\delta\right] \\
\leq & \lim_{T\to\infty} P\left[\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left\|\left(\hat{F}_t\hat{F}_t'-\mathbb{I}_R\right)\right\|^2}\sup_{\bar{u}\in\mathbb{R}}\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left\|\left(\mathbf{i}2\pi t/Te^{\mathbf{i}\bar{u}2\pi t/T}\right)\right\|^2}>\kappa/\delta\right] \\
= & \lim_{T\to\infty} P\left[\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left\|\left(\hat{F}_t\hat{F}_t'-\mathbb{I}_R\right)\right\|^2}\sqrt{\frac{1}{T}\sum_{t=1}^{T}4\pi t^2/T^2}>\kappa/\delta\right] \\
< & \epsilon,
\end{aligned}$$

where the third to last inequality is by triangle inequality and the second to last is by Cauchy-Swartz inequality. As is shown in Andrews (1994), the last inequality holds since we always find a $\delta > 0$ small enough given $\frac{1}{T}\sum_{t=1}^{T}\left\|\left(\hat{F}_t\hat{F}_t'-\mathbb{I}_R\right)\right\|^2$ is $Op(1)$, and $\frac{1}{T}\sum_{t=1}^{T}4\pi t^2/T^2$ is $O(1)$. By analogous argument, we can show $\frac{1}{T}\sum_{t=1}^{T}\left(\hat{F}_tF_t'-E(\hat{F}_tF_t')\right)e^{\mathbf{i}u2\pi t/T}$ is also stochastically equicontinuous. ■

Lemma A.2 *Suppose Assumptions A.1 to A.4 hold, then as $T\to\infty$*

$$\sup_{u\in\mathbb{R}}\left\|\tfrac{1}{T}\sum_{t=1}^{T}\hat{F}_t\hat{F}_t'e^{\mathbf{i}u2\pi t/T}-\mathbb{I}_R\int e^{\mathbf{i}u2\pi\tau}d\tau\right\|\xrightarrow{p}0,$$
$$\sup_{u\in\mathbb{R}}\left\|\tfrac{1}{T}\sum_{t=1}^{T}\hat{F}_tF_t'e^{\mathbf{i}u2\pi t/T}-E(\hat{F}_tF_t')\int e^{\mathbf{i}u2\pi\tau}d\tau\right\|\xrightarrow{p}0.$$

Proof.

$$\begin{aligned}
& \sup_{u\in\mathbb{R}}\left\|\frac{1}{T}\sum_{t=1}^{T}\hat{F}_t\hat{F}_t'e^{\mathbf{i}u2\pi t/T}-\mathbb{I}_R\int e^{\mathbf{i}u2\pi\tau}d\tau\right\| \\
= & \sup_{u\in\mathbb{R}}\left\|\frac{1}{T}\sum_{t=1}^{T}\hat{F}_t\hat{F}_t'e^{\mathbf{i}u2\pi t/T}-\mathbb{I}_R\frac{1}{T}\sum_{t=1}^{T}e^{\mathbf{i}u2\pi t/T}+\mathbb{I}_R\frac{1}{T}\sum_{t=1}^{T}e^{\mathbf{i}u2\pi t/T}-\mathbb{I}_R\int e^{\mathbf{i}u2\pi\tau}d\tau\right\| \\
\leq & \sup_{u\in\mathbb{R}}\left\|\frac{1}{T}\sum_{t=1}^{T}\hat{F}_t\hat{F}_t'e^{\mathbf{i}u2\pi t/T}-\mathbb{I}_R\frac{1}{T}\sum_{t=1}^{T}e^{\mathbf{i}u2\pi t/T}\right\|+\sup_{u\in\mathbb{R}}\left\|\mathbb{I}_R\frac{1}{T}\sum_{t=1}^{T}e^{\mathbf{i}u2\pi t/T}-\mathbb{I}_R\int e^{\mathbf{i}u2\pi\tau}d\tau\right\| \\
= & \sup_{u\in\mathbb{R}}\left\|\frac{1}{T}\sum_{t=1}^{T}\left(\hat{F}_t\hat{F}_t'-\mathbb{I}_R\right)e^{\mathbf{i}u2\pi t/T}\right\|+\mathbb{I}_R\sup_{u\in\mathbb{R}}\left\|\frac{1}{T}\sum_{t=1}^{T}\left(e^{\mathbf{i}u2\pi t/T}-\int e^{\mathbf{i}u2\pi\tau}d\tau\right)\right\| \\
= & R_1+R_2,
\end{aligned}$$

We now show $D_1 = op(1)$.

$$
\begin{aligned}
& sup_{u\in\mathbb{R}} \left\| \frac{1}{T}\sum_{t=1}^{T}\left(\hat{F}_t\hat{F}_t' - \mathbb{I}_R\right) e^{\mathbf{i}u2\pi t/T} \right\| \\
\leq\ & \sup_{u\in\mathbb{R}} \left\| \frac{1}{T}\sum_{t=1}^{T}\left(\hat{F}_t\hat{F}_t' - \mathbb{I}_R\right)\cos(u2\pi t/T) \right\| + \sup_{u\in\mathbb{R}} \left\| \frac{1}{T}\sum_{t=1}^{T}\left(\hat{F}_t\hat{F}_t' - \mathbb{I}_R\right)\sin(u2\pi t/T) \right\| \\
\equiv\ & R_{11} + R_{12},\ \text{say}.
\end{aligned}
$$

Next, we show that $R_{11} = op(1)$. Let $\mathbb{D}$ be the space of all functions $\theta : [0,1] \to [-1,1]$, where $\theta(\tau) = \cos(u2\pi\tau)$ for $\tau = t/T$, $t = 1,2,...,T$, and $\theta(\tau) = 0$ otherwise. Therefore,

$$
\begin{aligned}
\lim_{T\to\infty} P(R_{11} > 2\kappa) \ &\leq\ \lim_{T\to\infty} P\left(\sup_{\theta\in\mathbb{D}} \left\| \frac{1}{T}\sum_{t=1}^{T}\left(\hat{F}_t\hat{F}_t' - \mathbb{I}_R\right)\theta(\tau) \right\| > 2\kappa\right) \\
&\leq\ \lim_{T\to\infty} P\Bigg(\max_{j\leq J}\ \sup_{\tilde{\theta}\in B(\theta_j,\delta)} \Bigg\{ \left\| \frac{1}{T}\sum_{t=1}^{T}\left(\hat{F}_t\hat{F}_t' - \mathbb{I}_R\right)\left[\tilde{\theta}(\tau) - \theta_j(\tau)\right] \right\| \\
&\qquad + \left\| \frac{1}{T}\sum_{t=1}^{T}\left(\hat{F}_t\hat{F}_t' - \mathbb{I}_R\right)\theta_j(\tau) \right\| \Bigg\} > 2\kappa\Bigg) \\
&\leq\ \lim_{T\to\infty} P\left(\sup_{\theta\in\mathbb{D}}\ \sup_{\tilde{\theta}\in B(\theta_j,\delta)} \left\| \frac{1}{T}\sum_{t=1}^{T}\left(\hat{F}_t\hat{F}_t' - \mathbb{I}_R\right)\left[\tilde{\theta}(\tau) - \theta_j(\tau)\right] \right\| > \kappa\right) \\
&\qquad + \lim_{T\to\infty} P\left(\max_{j\leq J} \left\| \frac{1}{T}\sum_{t=1}^{T}\left(\hat{F}_t\hat{F}_t' - \mathbb{I}_R\right)\theta_j(\tau) \right\| \Bigg\} > \kappa\right) \\
&<\ \kappa,
\end{aligned}
$$

where we let $\{B(\theta_j,\delta) : j = 1,2,...,J\}$ be a finite cover of $\mathbb{D}$ such that $\theta \in B(\theta_j,\delta)$ if and only if $d(\theta,\theta_j) \equiv \sqrt{\int_0^1 |\theta(\tau) - \theta_j(\tau)|^2 d\tau} \leq \delta$. To let the last inequality hold, we need to show: (i). $\frac{1}{T}\sum_{t=1}^{T}\left(\hat{F}_t\hat{F}_t' - \mathbb{I}_R\right)\theta(\tau)$ is stochastically equicontinuous; and (ii). $\frac{1}{T}\sum_{t=1}^{T}\left(\hat{F}_t\hat{F}_t' - \mathbb{I}_R\right)\theta(\tau) = op(1)$ for any $\theta \in \mathbb{D}$.

For (i):

$$
\begin{aligned}
& \lim_{T\to\infty} P\left[\sup_{\theta_1,\theta_2\in\mathbb{D}:d(\theta_1,\theta_2)<\delta} \left\| \frac{1}{T}\sum_{t=1}^{T}\left(\hat{F}_t\hat{F}_t' - \mathbb{I}_R\right)(\theta_1(\tau) - \theta_2(\tau)) \right\| > \kappa\right] \\
\leq\ & \lim_{T\to\infty} P\left[\left\| \frac{1}{T}\sum_{t=1}^{T}\left(\hat{F}_t\hat{F}_t' - \mathbb{I}_R\right) \right\| > \kappa/\delta\right] < \epsilon,
\end{aligned}
$$

by analogous arguments in the proof of Lemma A.1. The point-wise convergence (ii) is easy to verify given $E(\hat{F}_t\hat{F}') = \mathbb{I}_R$, and the boundedness condition of $\mathbb{D}$. By analogous proof, we can show that $R_{12} = op(1)$. Therefore, $R_1 = op(1)$.

By Riemann approximation of an integral, we can show that R_2 is $o(1)$. Thus, we have shown

$$
\sup_{u\in\mathbb{R}} \left\| \frac{1}{T}\sum_{t=1}^{T}\hat{F}_t\hat{F}_t' e^{\mathbf{i}u2\pi t/T} - \mathbb{I}_R \int e^{\mathbf{i}u2\pi\tau} d\tau \right\| \xrightarrow{p} 0.
$$

By analogous proof, we can also show

$$\sup_{u\in\mathbb{R}}\left\|\frac{1}{T}\sum_{t=1}^{T}\hat{F}_tF_t'e^{\mathbf{i}u2\pi t/T}-E(\hat{F}_tF_t')\int e^{\mathbf{i}u2\pi\tau}d\tau\right\|\xrightarrow{p}0.$$

■

Lemma A.3 *Suppose Assumptions A.1 to A.4, and $\mathbb{H}_0$ hold, then as $N,T\to\infty$,*

$$\sqrt{NT}\hat{A}_2(u)\Rightarrow\mathcal{N}(u),$$

where $\mathcal{N}(u)$ is a complex-valued Gaussian process with covariance-kernel $\mathcal{M}(u_1,u_2)$

$$\mathcal{M}(u_1,u_2)=\lim_{N,T\to\infty}\frac{1}{NT}\sum_{s,t=1}^{T}\sum_{i,j=1}^{N}H_0'E\left[F_tF_s'\varepsilon_{it}\varepsilon_{js}\right]H_0M_t(u_1)M_s(u_2)^*,$$

where $M_t(u)$ is a demeaned Fourier process such that

$$M_t(u)=e^{\mathbf{i}u2\pi t/T}-\int e^{\mathbf{i}u2\pi\tau}d\tau.$$

Proof. By Lemma A.2, we have

$$\begin{aligned}
\hat{A}_2(u) &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}G_t(u)\varepsilon_{it}\\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{F}_te^{\mathbf{i}u_12\pi t/T}\varepsilon_{it}-\frac{1}{NT}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}\hat{F}_t\hat{F}_t'e^{\mathbf{i}u_22\pi t/T}\right)\sum_{t=1}^{T}\hat{F}_t\varepsilon_{it}\\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{F}_te^{\mathbf{i}u_12\pi t/T}\varepsilon_{it}-\frac{1}{NT}\sum_{i=1}^{N}\mathbb{I}_R\int e^{\mathbf{i}u2\pi\tau}d\tau\sum_{t=1}^{T}\hat{F}_t\varepsilon_{it}+op(1)\\
&= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{F}_t\varepsilon_{it}M_t(u)+op(1),
\end{aligned}$$

where

$$M_t(u)=e^{\mathbf{i}u2\pi t/T}-\int e^{\mathbf{i}u2\pi\tau}d\tau.$$

It follows

$$\begin{aligned}
\hat{A}_2(u) &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\hat{F}_t-H'F_t)\varepsilon_{it}M_t(u)+\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}H'F_t\varepsilon_{it}M_t(u)+op(1)\\
&\equiv \hat{A}_{21}(u)+\hat{A}_{22}(u)+op(1),\ \text{say.}
\end{aligned}$$

According to Bai (2003), $\hat{F}_t-H'F_t=op(1)$, thus $\hat{A}_{21}(u)=op(1)$ since $M_t(u)$ is bounded for all u. Therefore, the leading term in $\hat{A}_2(u)$ is $\hat{A}_{22}(u)$. Under Assumption A.1 to A.4, it follows that for each fixed $u\in\mathbb{R}$

$$\sqrt{NT}\hat{A}_2(u)\xrightarrow{d}N(0,\mathcal{M}(u,u)),$$

where

$$\mathcal{M}(u,u) = \text{avar}[\sqrt{NT}\hat{A}_{22}(u)]$$

$$= \lim_{N,T\to\infty} \frac{1}{NT} \operatorname{var}\left[\sum_{i=1}^{N}\sum_{t=1}^{T} H' F_t \varepsilon_{it} M_t(u)\right]$$

$$= \lim_{N,T\to\infty} \frac{1}{NT} \sum_{s,t=1}^{T}\sum_{i,j=1}^{N} H_0' E\left[F_t F_s' \varepsilon_{it}\varepsilon_{js}\right] H_0 M_t(u) M_s(u)^*.$$

Furthermore, given the stochastic equi-continuity result established in Lemma A.1, we can show

$$\sqrt{NT}\hat{A}_2(u) \Rightarrow \mathcal{N}(u),$$

where $\mathcal{N}(u)$ is a complex-valued Gaussian process with covariance-kernel $\mathcal{M}(u_1,u_2)$

$$\mathcal{M}(u_1,u_2) = \lim_{N,T\to\infty} \frac{1}{NT} \sum_{s,t=1}^{T}\sum_{i,j=1}^{N} H_0' E\left[F_t F_s' \varepsilon_{it}\varepsilon_{js}\right] H_0 M_t(u_1) M_s(u_2)^*.$$

Especially, when ε_{it} is a MDS, then

$$\mathcal{M}(u_1,u_2) = \lim_{N\to\infty} \frac{1}{N} \sum_{i,j=1}^{N} H_0' E(F_t F_t' \varepsilon_{it}\varepsilon_{jt}) H_0 \left[\int e^{\mathbf{i}2\pi\tau(u_1-u_2)} d\tau - \iint e^{\mathbf{i}2\pi(\tau u_1 - \lambda u_2)} d\tau d\lambda\right].$$

Furthermore, if ε_{it} is also cross-sectional uncorrelated,

$$\mathcal{M}(u_1,u_2) = H_0' E(F_t F_t' \varepsilon_{it}^2) H_0 \left[\int e^{\mathbf{i}2\pi\tau(u_1-u_2)} d\tau - \iint e^{\mathbf{i}2\pi(\tau u_1 - \lambda u_2)} d\tau d\lambda\right].$$

■

A.2 Proof of the Theorems

Proof of Proposition 3.1

Proof. Under $\mathbb{H}_0 : \lambda_{it} = \lambda_{i0}$, we have

$$\begin{aligned}
\hat{A}_1(u) &= \frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{T}\sum_{t=1}^{T} G_t(u) F_t' \lambda_{i0}\right] \\
&= \left[\frac{1}{T}\sum_{t=1}^{T} G_t(u) F_t'\right]\left[\frac{1}{N}\sum_{i=1}^{N}\lambda_{i0}\right] \\
&= \left[\frac{1}{T}\sum_{t=1}^{T} G_t(u)(F_t - H'^{-1}\hat{F}_t)'\right]\bar{\lambda}_0 + \left[\frac{1}{T}\sum_{t=1}^{T} G_t(u)\hat{F}_t' H^{-1}\right]\bar{\lambda}_0 \\
&\equiv \hat{A}_{11}(u) + \hat{A}_{12}(u),
\end{aligned}$$

where $\bar{\lambda}_0 = \left[\frac{1}{N}\sum_{i=1}^{N}\lambda_{i0}\right]$.

For $\hat{A}_{12}(u)$, it follows that

$$\begin{aligned}
\hat{A}_{12}(u) &= \left[\frac{1}{T}\sum_{t=1}^{T} G_t(u)\hat{F}_t' H^{-1}\right]\bar{\lambda}_0 \\
&= \left[\frac{1}{T}\sum_{t=1}^{T} \hat{F}_t e^{\mathbf{i}u2\pi t/T}\hat{F}_t' H^{-1} - \left(\frac{1}{T}\sum_{t=1}^{T}\hat{F}_t\hat{F}_t' e^{\mathbf{i}u2\pi t/T}\right)\left(\frac{1}{T}\sum_{t=1}^{T}\hat{F}_t\hat{F}_t'\right) H^{-1}\right]\bar{\lambda}_0
\end{aligned}$$

$$= \quad 0,$$

where the last equality comes from the fact that $\frac{1}{T}\sum_{t=1}^{T}\hat{F}_t\hat{F}_t' = \mathbb{I}_R$.

By Bai (2003),

$$\hat{F}_t - H'F_t = V_{NT}^{-1}\left(\frac{1}{T}\sum_{s=1}^{T}\hat{F}_s\gamma_N(s,t) + \frac{1}{T}\sum_{s=1}^{T}\hat{F}_s\zeta_{st} + \frac{1}{T}\sum_{s=1}^{T}\hat{F}_s\eta_{st} + \frac{1}{T}\sum_{s=1}^{T}\hat{F}_s\xi_{st}\right).$$

Then, it follows

$$F_t - H'^{-1}\hat{F}_t = -H'^{-1}V_{NT}^{-1}\left(\frac{1}{T}\sum_{s=1}^{T}\hat{F}_s\gamma_N(s,t) + \frac{1}{T}\sum_{s=1}^{T}\hat{F}_s\zeta_{st} + \frac{1}{T}\sum_{s=1}^{T}\hat{F}_s\eta_{st} + \frac{1}{T}\sum_{s=1}^{T}\hat{F}_s\xi_{st}\right),$$

and

$$\begin{aligned}
\frac{1}{T}\sum_{s=1}^{T}\hat{F}_s\gamma_N(s,t) &= Op\left(\frac{1}{\sqrt{T}C_{NT}}\right) \\
\frac{1}{T}\sum_{s=1}^{T}\hat{F}_s\zeta_{st} &= Op\left(\frac{1}{\sqrt{N}C_{NT}}\right) \\
\frac{1}{T}\sum_{s=1}^{T}\hat{F}_s\eta_{st} &= Op\left(\frac{1}{\sqrt{N}}\right) \\
\frac{1}{T}\sum_{s=1}^{T}\hat{F}_s\xi_{st} &= Op\left(\frac{1}{\sqrt{N}C_{NT}}\right),
\end{aligned}$$

where

$$C_{NT} = \min\{\sqrt{N}, \sqrt{T}\}.$$

We claim only $\frac{1}{T}\sum_{s=1}^{T}\hat{F}_s\gamma_N(s,t)$ and $\frac{1}{T}\sum_{s=1}^{T}\hat{F}_s\eta_{st}$ can be the leading terms. Let $T \propto N^{\nu}$, we have the following three cases

- **CASE 1:** If $\nu > 1/2$, then the leading term is

$$\frac{1}{T}\sum_{s=1}^{T}\hat{F}_s\eta_{st} = Op\left(\frac{1}{\sqrt{N}}\right).$$

- **CASE 2:** If $\nu = 1/2$, then the leading terms will be

$$\frac{1}{T}\sum_{s=1}^{T}\hat{F}_s\eta_{st} + \frac{1}{T}\sum_{s=1}^{T}\hat{F}_s\gamma_N(s,t) = Op\left(\frac{1}{\sqrt{N}}\right) = Op\left(\frac{1}{T}\right).$$

- **CASE 3:** If $\nu < 1/2$, then the leading term will be

$$\frac{1}{T}\sum_{s=1}^{T}\hat{F}_s\gamma_N(s,t) = Op\left(\frac{1}{T}\right).$$

When $\nu > 1/2$,

$$\begin{aligned}
\hat{A}_{11}(u) &= \left[\frac{1}{T}\sum_{t=1}^{T} G_t(u)\left(-H'^{-1}V_{NT}^{-1}\frac{1}{T}\sum_{s=1}^{T}\hat{F}_s\eta_{st}\right)'\right]\bar{\lambda}_0 + op(1) \\
&= \left\{\frac{1}{T}\sum_{t=1}^{T} G_t(u)\left[-H'^{-1}V_{NT}^{-1}\left(\frac{1}{T}\sum_{s=1}^{T}\hat{F}_sF_s'\right)\left(\frac{1}{N}\sum_{i=1}^{N}\lambda_{i0}\varepsilon_{it}\right)\right]'\right\}\bar{\lambda}_0 + op(1) \\
&= \left\{-\frac{1}{T}\sum_{t=1}^{T} G_t(u)\left(\frac{1}{N}\sum_{i=1}^{N}\lambda_{i0}\varepsilon_{it}\right)'\left(\frac{1}{T}\sum_{s=1}^{T}\hat{F}_sF_s'\right)' V_{NT}^{-1}H^{-1}\right\}\bar{\lambda}_0 + op(1) \\
&= -\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} G_t(u)\varepsilon_{it}\lambda_{i0}'\left(\frac{1}{T}\sum_{s=1}^{T}\hat{F}_sF_s'\right)' V_{NT}^{-1}H^{-1}\bar{\lambda}_0 + op(1) \\
&= -\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} G_t(u)\varepsilon_{it}\lambda_{i0}'\left(\Lambda_0'\Lambda_0/N\right)^{-1}\bar{\lambda}_0 + op(1).
\end{aligned}$$

Therefore, we combine $\hat{A}_1(u)$ and $\hat{A}_2(u)$ and get

$$\hat{A}(u) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} G_t(u)B_i\varepsilon_{it} + op(1),$$

where

$$B_i = 1 - \lambda_{i0}'\left(\Lambda_0'\Lambda_0/N\right)^{-1}\bar{\lambda}_0.$$

When $\nu < 1/2$,

$$\begin{aligned}
\hat{A}_{11}(u) &= \left[\frac{1}{T}\sum_{t=1}^{T} G_t(u)\left(-H'^{-1}V_{NT}^{-1}\frac{1}{T}\sum_{s=1}^{T}\hat{F}_s\gamma_N(s,t)\right)'\right]\bar{\lambda}_0 + op(1) \\
&= \left[\frac{1}{T}\sum_{t=1}^{T} G_t(u)\left(-H'^{-1}V_{NT}^{-1}\frac{1}{T}\sum_{s=1}^{T}\hat{F}_s\left(\frac{1}{N}\sum_{i=1}^{N}E(\varepsilon_{is}\varepsilon_{it})\right)\right)'\right]\bar{\lambda}_0 + op(1) \\
&= -\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{s,t=1}^{T} G_t(u)\hat{F}_s'E(\varepsilon_{is}\varepsilon_{it})V_{NT}^{-1}H^{-1}\bar{\lambda}_0 + op(1) \\
&= -\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T} G_t(u)\hat{F}_t'E(\varepsilon_{it}^2)V_{NT}^{-1}H^{-1}\bar{\lambda}_0 \\
&\quad -\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{s\neq t}^{T} G_t(u)\hat{F}_s'E(\varepsilon_{is}\varepsilon_{it})V_{NT}^{-1}H^{-1}\bar{\lambda}_0 + op(1) \\
&= \hat{A}_{111}(u) + \hat{A}_{112}(u) + op(1), \text{ say.}
\end{aligned}$$

When ε_{it} is weakly stationary, then $E(\varepsilon_{it}^2) = \sigma_i^2$, and

$$\hat{A}_{111}(u) = -\frac{1}{T}\sum_{t=1}^{T} G_t(u)\hat{F}_tV_{NT}^{-1}H^{-1}\bar{\lambda}_0\left(\frac{1}{N}\sum_{i=1}^{N}\sigma_i^2\right)$$

$$= -\left[\frac{1}{T}\sum_{t=1}^{T}\hat{F}_t e^{\mathbf{i}u2\pi t/T}\hat{F}_t' - \left(\frac{1}{T}\sum_{t=1}^{T}\hat{F}_t\hat{F}_t' e^{\mathbf{i}u2\pi t/T}\right)\left(\frac{1}{T}\sum_{t=1}^{T}\hat{F}_t\hat{F}_t'\right)\right] V_{NT}^{-1}H^{-1}\bar{\lambda}_0\left(\frac{1}{N}\sum_{i=1}^{N}\sigma_i^2\right)$$
$$= 0.$$

For $\hat{A}_{112}(u)$,

$$\begin{aligned}
\hat{A}_{112}(u) &= -\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{s\neq t}^{T} G_t(u)\hat{F}_s' E(\varepsilon_{is}\varepsilon_{it})V_{NT}^{-1}H^{-1}\bar{\lambda}_0 \\
&= -\frac{1}{T^2}\sum_{s\neq t}^{T} G_t(u)\hat{F}_s'\gamma_N(s,t)V_{NT}^{-1}H^{-1}\bar{\lambda}_0 \\
&= -\frac{1}{T^2}\sum_{s\neq t}^{T}\hat{F}_t\hat{F}_s' M_t(u)\gamma_N(s,t)V_{NT}^{-1}H^{-1}\bar{\lambda}_0 + op(1) \\
&= -\frac{2}{T^2}\sum_{j=1}^{T-1}\sum_{t=1}^{T-j}\hat{F}_t\hat{F}_{t+j}M_t(u)\gamma_N(t,t+j)V_{NT}^{-1}H^{-1}\bar{\lambda}_0 + op(1) \\
&= -\frac{2}{T^2}\sum_{j=1}^{T-1}\sum_{t=1}^{T-j}E(\hat{F}_t\hat{F}_{t+j})M_t(u)\gamma_N(t,t+j)V_{NT}^{-1}H^{-1}\bar{\lambda}_0 \\
&\quad -\frac{2}{T^2}\sum_{j=1}^{T-1}\sum_{t=1}^{T-j}\left[\hat{F}_t\hat{F}_{t+j} - E(\hat{F}_t\hat{F}_{t+j})\right] M_t(u)\gamma_N(t,t+j)V_{NT}^{-1}H^{-1}\bar{\lambda}_0 + op(1) \\
&= -\frac{2}{T}\sum_{j=1}^{T-1}E(\hat{F}_t\hat{F}_{t+j})\gamma_N(t,t+j)\left[\frac{1}{T}\sum_{t=1}^{T-j}M_t(u)\right]V_{NT}^{-1}H^{-1}\bar{\lambda}_0 + op(1) \\
&= O(T^{-1}) * O(T^{-1/2}),
\end{aligned}$$

where we use the fact that

$$\frac{1}{T}\sum_{t=1}^{T}M_t(u) = O(T^{-1/2}),$$

and

$$\sum_{j=-\infty}^{\infty}E(F_tF_{t+j})\gamma_N(t,t+j) < C < \infty$$

.

Therefore, $\hat{A}_{11}(u) = Op(T^{-3/2})$. And

$$\hat{A}(u) = -\frac{2}{T}\sum_{j=1}^{T-1}E(\hat{F}_t\hat{F}_{t+j})\gamma_N(t,t+j)\left[\frac{1}{T}\sum_{t=1}^{T-j}M_t(u)\right]V_{NT}^{-1}H^{-1}\bar{\lambda}_0 + op(1),$$

which is a degenerate statistic. When $\nu = 1/2$, it is straightforward to show that

$$\hat{A}(u) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}G_t(u)B_i\varepsilon_{it} - \frac{2}{T}\sum_{j=1}^{T-1}E(\hat{F}_t\hat{F}_{t+j})\gamma_N(t,t+j)\left[\frac{1}{T}\sum_{t=1}^{T-j}M_t(u)\right]V_{NT}^{-1}H^{-1}\bar{\lambda}_0 + op(1).$$

where

$$B_i = 1 - \lambda_{i0}'\left(\Lambda_0'\Lambda_0/N\right)^{-1}\bar{\lambda}_0.$$

■

Proof of Proposition 3.2
Proof. Under $\mathbb{H}_1$, we can show that

$$\begin{aligned}\hat{A}_1(u) &= \frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{T}\sum_{t=1}^{T}G_t(u)F_t'\lambda_{it}\right]\\ &= \frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{T}\sum_{t=1}^{T}\hat{F}_tF_t'e^{\mathbf{i}u2\pi t/T}\lambda_{it}-\left(\frac{1}{T}\sum_{t=1}^{T}\hat{F}_t\hat{F}_t'e^{\mathbf{i}u2\pi t/T}\right)\frac{1}{T}\sum_{t=1}^{T}\hat{F}_tF_t'\lambda_{it}\right]\end{aligned}$$

By Lemma A.1, we have

$$\frac{1}{T}\sum_{t=1}^{T}\hat{F}_tF_t'e^{\mathbf{i}u2\pi t/T}\lambda_{it}\Rightarrow E(\hat{F}_tF_t')\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}e^{\mathbf{i}u2\pi t/T}\lambda_{it},$$

and

$$\left(\frac{1}{T}\sum_{t=1}^{T}\hat{F}_t\hat{F}_t'e^{\mathbf{i}u2\pi t/T}\right)\frac{1}{T}\sum_{t=1}^{T}\hat{F}_tF_t'\lambda_{it}\Rightarrow E(\hat{F}_tF_t')\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}e^{\mathbf{i}u2\pi t/T}\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\lambda_{it}.$$

Thus

$$\hat{A}_1(u)\Rightarrow E(\hat{F}_tF_t')\lim_{N,T\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}e^{\mathbf{i}u2\pi t/T}\lambda_{it}-\frac{1}{T}\sum_{t=1}^{T}e^{\mathbf{i}u2\pi t/T}\frac{1}{T}\sum_{t=1}^{T}\lambda_{it}\right).$$

Furthermore, if $\lambda_{it}=\lambda_i(\frac{t}{T})$, we have

$$\hat{A}_1(u)\Rightarrow E(\hat{F}_tF_t')\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\widetilde{cov}(e^{\mathbf{i}u2\pi\tau},\lambda_i(\tau)),$$

where $\widetilde{cov}(e^{\mathbf{i}u2\pi\tau},\lambda_i(\tau))$ is a pseudo-covariance such that

$$\widetilde{cov}[e^{\mathbf{i}u2\pi\tau},\lambda_i(\tau)]=\int_0^1 e^{\mathbf{i}u2\pi\tau}\lambda_i(\tau)\,d\tau-\int_0^1 e^{\mathbf{i}u2\pi\tau}d\tau\int_0^1\lambda_i(\tau)\,d\tau.$$

By Lemma A.2, $\hat{A}_2(u)=Op\left(\frac{1}{\sqrt{NT}}\right)$. Thus, we proved the result. ■

Proof of Theorem 3.3
Proof. Under $\mathbb{H}_0:\lambda_{it}=\lambda_{i0}$, Proposition 3.1 shows

$$\hat{A}(u)=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}G_t(u)B_i\varepsilon_{it}+O(T^{-3/2})+op(1).$$

Given $\sqrt{N}/T\to 0$,

$$\sqrt{NT}\hat{A}(u)=\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}G_t(u)B_i\varepsilon_{it}+op(1)$$

By Lemma A.3,

$$\sqrt{NT}\hat{A}(u) \Rightarrow \mathcal{G}(u),$$

where $\mathcal{G}(u)$ is a complex-valued Gaussian process with covariance-kernel $\mathcal{K}(u_1, u_2)$ such that

$$\mathcal{K}(u_1, u_2) = \lim_{N,T\to\infty} \frac{1}{NT} \sum_{s,t=1}^{T} \sum_{i,j=1}^{N} B_i B_j H_0' E\left[F_t F_s' \varepsilon_{it} \varepsilon_{js}\right] H_0 M_t(u_1) M_s(u_2)^*.$$

By Lemma A.1, it is straightforward to show that $\hat{A}(u)$ is stochastically equicontinuous over $u \in \mathbb{R}$, under Assumption A.4 and continuous mapping theorem, we have

$$\hat{D} \xrightarrow{d} \int_{\mathbb{R}} \|\mathcal{G}(u)\|^2 W(u) du.$$

■

Proof of Theorem 3.4

Proof. Under $H_1(a_{NT}) : \lambda_{it} = \lambda_{i0} + a_{NT} g_{it}$,

$$\begin{aligned}\hat{A}_1(u) &= \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} G_t(u) F_t' \lambda_{it} \\ &= \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} G_t(u) F_t' \lambda_{i0} + \frac{a_{NT}}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} G_t(u) F_t' g_{it}.\end{aligned}$$

By Proposition 3.1, we have

$$\begin{aligned}&\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} G_t(u) F_t' \lambda_{i0} \\ =\ &-\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} G_t(u) \varepsilon_{it} \lambda_{i0}' (\Lambda_0' \Lambda_0 / N)^{-1} \bar{\lambda}_0 + Op(T^{-3/2}) + op(1).\end{aligned}$$

Given $a_{NT} = \frac{1}{\sqrt{NT}}$, by Proposition 3.2,

$$\sqrt{NT} \frac{a_{NT}}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} G_t(u) F_t' g_{it} \Rightarrow \xi(u),$$

where

$$\xi(u) = E(\hat{F}_t F_t') \lim_{N,T\to\infty} \frac{1}{N} \sum_{i=1}^{N} \left(\frac{1}{T} \sum_{t=1}^{T} e^{\mathbf{i}u2\pi t/T} g_{it} - \frac{1}{T} \sum_{t=1}^{T} e^{\mathbf{i}u2\pi t/T} \frac{1}{T} \sum_{t=1}^{T} g_{it} \right).$$

Given Theorem 3.3, it follows

$$\sqrt{NT}\hat{A}(u) \Rightarrow \xi(u) + \mathcal{G}(u)$$

when $\sqrt{N}/T \to 0$. By continuous mapping theorem, we have

$$\hat{D} \xrightarrow{d} \int \|\xi(u) + \mathcal{G}(u)\|^2 W(u) du.$$

Table A.1 Size of tests under DGP.S1 when the number of factors is determined from the data

N	T	D_B		SW		HI_{LM}		HI_W		CDG_{LM}		CDG_W		BE_{LM}	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
i.i.d. error term: $\varepsilon_{it} \sim i.i.d.N(0,1)$															
100	100	4.6	9.2	5.6	12.6	1.0	4.4	0.2	1.4	2.0	6.2	5.2	12.8	2.7	6.3
100	200	4.0	10.8	6.6	12.8	2.2	7.0	2.2	6.6	3.4	8.0	4.8	11.6	3.5	7.5
200	100	5.6	11.0	5.0	11.2	1.2	5.0	0.6	2.6	2.2	6.2	7.6	11.8	2.8	6.4
200	200	5.8	10.8	5.6	11.4	3.4	7.6	3.0	8.6	2.4	7.2	6.6	9.8	3.3	7.3
heteroskedastic error term: $\varepsilon_{it} = \sigma_i v_{it}$, $\sigma_i \sim i.i.d.U(0.5, 1.5), v_{it} \sim i.i.d.N(0,1)$															
100	100	5.2	10.6	5.4	12.0	0.8	4.4	0.2	1.6	1.4	7.0	5.6	12.2	2.8	6.3
100	200	5.4	11.4	6.8	14.6	2.2	7.2	2.4	6.8	3.4	8.6	4.6	11.2	3.5	7.4
200	100	4.8	11.0	5.6	10.0	1.4	5.2	0.6	2.6	2.0	5.8	7.2	11.8	2.8	6.4
200	200	5.0	10.2	5.4	11.8	3.2	7.6	3.0	8.4	2.0	7.0	5.8	9.2	3.3	7.3
cross sectional dependence error term: $\varepsilon_{\cdot t} \sim i.i.d.N(0, \Sigma_e)$															
100	100	5.8	10.2	5.4	10.2	1.0	4.0	0.2	1.0	1.6	7.0	6.6	12.8	2.7	6.3
100	200	5.0	10.8	4.2	8.8	2.0	6.4	2.0	4.6	2.0	7.2	5.0	10.6	3.4	7.4
200	100	8.0	12.2	5.8	12.4	1.6	6.0	0.8	3.0	1.6	5.8	7.0	12.0	2.8	6.4
200	200	5.4	9.2	5.4	11.2	3.2	7.6	3.4	8.0	2.6	7.2	5.8	9.4	3.5	7.6
time series dependence error term: $\varepsilon_{it} = 0.5\varepsilon_{it-1} + v_{it}, v_{it} \sim i.i.d.N(0,1)$															
100	100	7.6	13.8	100	100	1.0	4.0	0.2	1.0	1.6	7.0	6.6	12.8	2.7	6.4
100	200	4.0	10.4	100	100	2.0	6.4	2.0	4.6	2.0	7.2	5.0	10.6	3.4	7.4
200	100	5.2	13.0	100	100	1.6	6.0	0.8	3.0	1.6	5.8	7.0	12.0	2.8	6.4
200	200	6.0	11.4	100	100	3.2	7.6	3.4	8.0	2.6	7.2	5.8	9.4	3.5	7.6
cross sectional and time series dependence error term: $\varepsilon_{\cdot t} = 0.5\varepsilon_{\cdot t-1} + v_{\cdot t}, v_{\cdot t} \sim i.i.d.N(0, \Sigma_e)$															
100	100	6.8	13.0	100	100	0.4	3.2	0.4	2.0	2.4	8.0	4.4	11.2	13.6	22.5
100	200	7.2	12.6	100	100	2.4	6.0	3.4	8.0	3.2	8.2	4.8	9.4	18.1	28.4
200	100	6.6	12.4	100	100	1.2	3.8	.4	2.0	2.0	5.4	6.0	12.0	13.9	23.1
200	200	7.0	13.4	100	100	3.2	6.8	2.6	9.6	4.8	11.6	6.8	12.4	17.8	27.8

Note: See the note in Table 1.

Furthermore, when $\sqrt{N}/T \to \infty$,

$$\sqrt{NT}\hat{A}(u) \Rightarrow \infty.$$

■

B Some Additional Simulation Results

In this appendix, we report some additional simulation results. Tables A.1 to A.3 report the size and power performance of various tests at the 5% and 10% significant levels when the number of factors is determined by Bai and Ng's (2002) IC_{p1}. The results are similar to those reported in Tables 1 to 3.

Table A.2 Power of tests under DGP.P1-P2 when the number of factors is determined from the data

	N	T	D_B		SW		HI_{LM}		HI_W		CDG_{LM}		CDG_W		BE_{LM}	
			5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
i.i.d. error term: $\varepsilon_{it} \sim i.i.d.N(0,1)$																
DGP.P1	100	100	99.0	99.4	71.0	80.2	1.6	5.6	0.0	2.0	2.2	7.6	6.0	12.6	5.4	10.8
	100	200	100	100	97.6	99.0	4.6	9.8	5.6	12.2	3.6	7.6	6.0	11.4	10.5	17.7
	200	100	100	100	88.6	93.2	1.4	5.4	1.2	3.2	2.4	9.0	7.6	11.2	5.4	10.7
	200	200	100	100	100	100	5.4	12.6	5.6	12.8	4.0	8.8	3.6	10.6	10.7	17.9
DGP.P2	100	100	26.2	37.4	10.2	17.8	1.2	4.4	0.2	1.6	1.8	7.6	4.0	10.2	2.9	6.6
	100	200	62.6	76.6	21.4	31.0	2.2	7.4	2.2	5.8	3.8	8.6	4.8	9.4	4.0	8.3
	200	100	42.0	59.4	11.4	20.0	1.2	5.4	0.4	3.0	3.2	8.4	6.6	10.4	2.9	6.7
	200	200	87.2	93.4	24.6	37.4	3.4	8.0	2.4	8.2	4.2	9.2	2.8	9.8	3.8	8.2
heteroskedastic error term: $\varepsilon_{it} = \sigma_i v_{it}$, $\sigma_i \sim i.i.d.U(0.5, 1.5), v_{it} \sim i.i.d.N(0,1)$																
DGP.P1	100	100	99.0	99.8	73.6	81.4	1.6	5.6	0.0	0.2	2.2	7.6	5.6	12.2	6.7	12.3
	100	200	100	100	98.2	98.6	4.6	9.8	5.6	12.2	3.6	7.8	6.4	11.6	13.2	20.5
	200	100	100	100	91.4	95.8	1.4	5.2	1.2	3.0	2.6	9.0	7.4	11.6	6.7	12.2
	200	200	100	100	100	100	5.8	13.0	6.0	12.8	4.0	9.2	3.8	10.4	13.5	20.7
DGP.P2	100	100	26.6	40.6	10.2	18.8	1.2	4.2	0.2	1.6	1.8	7.8	3.8	10.2	2.9	6.6
	100	200	62.8	75.4	21.2	33.8	2.6	6.8	2.4	6.0	4.2	8.6	4.8	10.2	4.1	8.6
	200	100	45.6	62.8	13.2	20.2	1.4	5.4	0.4	3.2	3.2	9.2	6.4	10.4	3.0	6.8
	200	200	86.4	92.6	28.2	39.4	3.2	7.6	2.4	8.2	4.0	9.6	3.0	9.8	4.0	8.5
cross sectional dependence error term: $\varepsilon_{\cdot t} \sim i.i.d.N(0, \Sigma_e)$																
DGP.P1	100	100	98.6	99.4	66.2	76.4	1.2	4.6	0.0	1.2	2.4	7.4	5.6	13.0	5.4	10.6
	100	200	100	100	97.6	98.6	5.2	9.8	4.4	12.4	3.0	8.4	5.6	10.8	10.9	18.0
	200	100	100	100	88.0	94.6	1.4	5.4	1.0	4.0	2.2	8.4	6.4	11.4	5.5	10.8
	200	200	100	100	100	100	5.8	13.2	5.8	12.0	4.4	8.8	4.6	10.6	10.9	18.1
DGP.P2	100	100	26.4	38.0	10.4	17.2	1.2	5.0	0.0	0.8	2.6	7.2	4.4	10.2	2.9	6.6
	100	200	62.8	77.6	17.4	27.4	2.6	7.4	1.6	6.6	4.0	9.2	4.6	8.8	3.8	8.1
	200	100	40.6	56.4	12.4	19.8	1.6	5.0	0.6	3.2	2.2	7.8	5.8	10.2	3.0	6.7
	200	200	88.6	94.6	27.6	42.8	3.4	7.8	3.2	7.8	4.2	9.2	2.6	9.8	3.9	8.5
time series dependence error term: $\varepsilon_{it} = 0.5\varepsilon_{it-1} + v_{it}, v_{it} \sim i.i.d.N(0,1)$																
DGP.P1	100	100	89.4	92.8	99.8	100	1.2	5.0	0.4	3.4	1.6	6.2	4.6	11.0	16.5	26.3
	100	200	100	100	100	100	5.4	10.4	5.4	14.0	4.6	9.6	5.6	11.6	24.5	35.4
	200	100	98.6	99.2	100	100	1.8	4.8	0.8	3.2	2.0	6.4	4.8	12.0	16.4	26.2
	200	200	100	100	100	100	5.8	12.6	5.0	15.0	5.8	9.4	7.4	12.0	24.6	35.6
DGP.P2	100	100	16.4	26.4	99.0	99.4	0.8	4.4	0.2	1.8	1.4	6.2	3.2	9.8	13.7	22.8
	100	200	41.2	58.8	100	100	2.8	7.4	2.8	6.6	3.8	10.4	4.0	11.2	18.2	28.3
	200	100	24.4	38.0	100	100	1.6	4.6	0.4	2.0	1.4	6.2	3.4	8.8	13.5	22.6
	200	200	65.2	79.8	100	100	3.2	7.4	2.6	8.4	4.6	10.8	4.4	10.2	18.2	28.4
cross sectional and time series dependence error term: $\varepsilon_{\cdot t} = 0.5\varepsilon_{\cdot t-1} + v_{\cdot t}, v_{\cdot t} \sim i.i.d.N(0, \Sigma_e)$																
DGP.P1	100	100	89.4	94.4	100	100	1.0	4.8	0.4	3.0	2.6	7.8	4.8	10.6	16.5	26.3
	100	200	99.0	99.2	100	100	5.0	10.8	5.6	13.8	4.2	8.8	6.0	11.6	25.0	35.8
	200	100	99.0	100	100	100	1.6	4.4	0.6	3.2	2.0	5.6	4.6	12.4	16.8	26.6
	200	200	100	100	100	100	4.8	12.4	5.2	14.4	5.2	9.8	7.0	12.2	24.6	35.4
DGP.P2	100	100	14.8	23.4	99.8	100	0.8	3.2	0.4	1.6	1.8	6.8	4.0	9.0	13.6	22.8
	100	200	39.4	53.2	100	100	2.6	7.2	3.2	8.2	4.4	10.4	4.6	11.0	18.5	29.0
	200	100	25.6	40.4	100	100	1.2	3.8	0.4	2.2	1.4	6.2	4.0	9.0	14.0	23.0
	200	200	65.8	78.8	100	100	2.8	6.6	2.2	8.4	4.6	10.8	5.0	10.4	18.2	28.5

Note: See the note in Table 1.

Table A.3 Power of tests under DGP.P3 when the number of factors is determined from the data

	N	T	D_B		SW		HI_{LM}		HI_W		CDG_{LM}		CDG_W		BE_{LM}	
			5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
			i.i.d. error term: $\varepsilon_{it} \sim i.i.d.N(0,1)$													
DGP.P3	100	100	82.2	91.2	37.0	48.4	0.6	3.6	0.4	2.4	2.2	8.0	5.8	12.0	3.7	8.2
	100	200	98.2	98.6	72.0	80.8	1.6	5.6	3.8	9.8	4.2	8.8	4.8	12.0	5.5	11.0
	200	100	89.6	95.6	42.8	54.4	1.6	4.6	1.0	3.4	1.8	4.6	9.8	15.6	3.7	8.2
	200	200	100	100	89.8	93.4	1.8	5.6	3.6	11.2	2.6	7.0	7.4	12.6	5.5	11.0
			heteroskedastic error term: $\varepsilon_{it} = \sigma_i v_{it}$, $\sigma_i \sim i.i.d.U(0.5, 1.5), v_{it} \sim i.i.d.N(0,1)$													
	100	100	78.4	89.6	31.6	43.4	0.6	3.0	0.4	2.6	2.2	8.4	7.8	13.2	3.9	8.2
	100	200	96.0	97.4	63.8	73.4	1.2	5.0	3.8	10.4	3.2	7.6	5.2	12.6	5.7	11.0
	200	100	86.6	96.0	37.8	49.6	1.4	4.6	1.0	3.6	1.6	4.8	9.8	14.6	3.8	8.1
	200	200	100	100	83.0	90.6	2.0	5.6	4.2	11.2	2.6	7.6	8.2	13.8	5.7	11.1
			cross sectional dependence error term: $\varepsilon_{\cdot t} \sim i.i.d.N(0, \Sigma_e)$													
	100	100	82.2	90.6	31.4	43.6	0.8	3.2	0.8	2.8	2.2	8.4	8.6	14.0	3.4	7.5
	100	200	98.4	99.2	74.0	82.4	1.4	5.4	3.2	9.6	3.4	7.8	5.0	12.0	5.0	10.2
	200	100	90.6	96.2	45.8	57.8	1.4	4.6	1.2	4.0	1.6	5.2	10.4	15.2	3.5	7.7
	200	200	98.8	99.6	91.8	95.0	1.6	5.0	3.8	11.4	2.2	7.6	7.8	15.0	5.3	10.4
			time series dependence error term: $\varepsilon_{it} = 0.5\varepsilon_{it-1} + v_{it}, v_{it} \sim i.i.d.N(0,1)$													
	100	100	55.0	71.8	99.2	100	0.4	3.2	0.4	3.4	1.8	6.8	7.8	13.8	14.5	24.0
	100	200	91.0	94.8	100	100	2.4	4.6	4.0	10.8	2.8	8.2	5.4	13.0	19.7	30.1
	200	100	76.0	86.2	99.8	100	1.2	3.0	0.8	2.8	1.6	6.4	7.6	14.8	14.5	24.0
	200	200	96.2	97.0	100	100	1.8	3.8	2.6	12.2	3.0	8.0	9.0	15.4	19.9	30.4
			cross sectional and time series dependence error term: $\varepsilon_{\cdot t} = 0.5\varepsilon_{\cdot t-1} + v_{\cdot t}, v_{\cdot t} \sim i.i.d.N(0, \Sigma_e)$													
	100	100	55.4	71.2	100	100	0.2	2.2	0.4	3.4	1.8	6.6	7.2	14.4	14.7	24.2
	100	200	91.2	95.8	100	100	1.4	5.4	3.2	9.0	2.4	7.8	4.8	11.4	20.3	30.8
	200	100	76.0	87.0	100	100	0.8	3.6	0.6	2.6	1.8	5.4	7.2	13.6	14.8	24.2
	200	200	96.0	96.6	100	100	1.8	4.2	3.4	11.2	2.8	6.8	8.2	15.4	20.0	30.5

Note: See the note in Table 1.