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
[b] Cornell University and Xiamen University
[c] Sun Yat-Sen University

May 16, 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 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 null (zero) spectrum implied by no structural change. The proposed test is powerful against both smooth structural changes and abrupt structural breaks with a possibly unknown number of breaks and unknown break dates in factor loadings. It can detect a class of local alternatives at the rate of $T^{-1/2}N^{-1/2}$, where T and N are the numbers of time periods and cross-sectional units. As a result, the test is asymptotically more efficient than the existing tests in the factor model literature. Our test is easy to implement and tuning parameter-free. Also, it is robust to serial dependence and cross-sectional dependence of unknown form. Monte Carlo studies demonstrate its reasonable size and excellent power in detecting structural changes in factor loadings. In an application to Stock and Watson's (2012) U.S. macroeconomic data, we find significant and robust evidence against time-invariant factor loadings.

JEL Classification: C12, C14, C33, C38.

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

---

*Fu is from School of Economics, Fudan University, and Shanghai Institute of International Finance and Economics (SIIFE), Email: *zhfu@fudan.edu.cn*; Hong is from Department of Economics and Department of Statistics and Data Science, Cornell University, and Wang Yanan Institute for Studies in Economics (WISE) and MOE Key Laboratory of Econometrics, Xiamen University, Email: *yh20@cornell.edu*; Wang is from Lingnan College, Sun Yat-sen University, Email: *wangxia25@mail.sysu.edu.cn*. Wang acknowledges financial supports from National Science Foundation of China (No. 71401160, 71873151). We thank Liang Chen, Xu Han, Liangjun Su, and seminar participants in City University of Hong Kong, Fourth Guangzhou Econometrics workshop, Second Chinese Econometrican's Forum, and Fifth Annual Meeting of Young Econometricans in Asia-Pacific. Please address correspondence to Yongmiao Hong, Department of Economics and Department of Statistics and Data Science, Cornell University, 424 Uris Hall, Ithaca, NY 14850, USA, and Wang Yanan Institute for Studies in Economics, Economic Building, Xiamen University, Xiamen 361005, China; e-mail: *yh20@cornell.edu*;

1 Introduction

Factor models are useful for analyzing large dimensional macroeconomic and financial datasets. 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 factor loadings, which capture the relationship between economic variables and unobserved common factors, are time invariant. However, it is likely that the underlying structure of the data generating process changes over time when the time span is long. Even though Stock and Watson (2002, 2009) point out that the estimated factors by PCA are still consistent when factor loadings undergo small instabilities, it is difficult to believe that 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 affect the relationship between economic variables and 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 may not be consistent and the inference and forecasting based on such an assumption would lead to misleading conclusions. Furthermore, if factor loadings suffer from structural changes, most of the existing methods such as Bai and Ng (2002), Onatski (2009), and Ahn and Horenstein (2013) will 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 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 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 big 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 breaks in 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 a non-monotonic power problem. Cheng *et al.* (2016) consider the case in which both factor loadings and the number of common factors may change simultaneously. Although the aforementioned works provide useful econometric tools on detecting possible structural breaks in factor loadings, they focus on testing for abrupt structural breaks, especially a single structural break. In fact, the sources of structural changes, such as preference changes, technological progress, and institutional transformation, usually take effect gradually over time. Even if some policy switches occur immediately, it may take some time for economic agents 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 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 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 changes. Recently, Su and Wang (2017) propose an L_2-distance-based test to examine the stability of factor loadings. They estimate time-varying factor loadings and latent common factors by a local version of PCA and construct a test for the null hypothesis of no structural change by comparing the fitted values of 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 *et al.* (2018) in a time series context. Unlike the existing tests for structural changes in time series regressions that are based on time domain analysis, Fu *et al.* (2018) propose a novel method that investigates structural changes in frequency domain. Our test is constructed using a similar idea but in a different framework. Fu *et al.* (2018) consider linear time series regressions with observed regressors. In contrast, for large dimensional factor models, both factor loadings and common factors are unobservable. Moreover, the asymptotic analysis is substantively different, because, for large dimensional factor models, a joint asymptotic approach is called for to deal with the large N and large T problem, where N and T are the numbers of cross-sectional units and time periods, respectively. Among other things, we need to impose conditions on the relative speed between N and T in order to control the impact of estimating latent factors and factor loadings and achieve robustness to serially correlated and/or cross-sectionally dependent error terms. The intuition behind our test is straightforward. If factor loadings change over time, the conventional PCA will fail to capture the time-varying feature of true factor loadings. As a result, the estimated residuals based on the conventional PCA will contain the time-varying component. By a DFT, we can project the estimated residuals onto frequency domain and infer the existence of structural changes. Compared with the existing tests in the factor model literature, the proposed test has the following appealing features.

First, our test has power against a wide range of alternatives for structural changes, including smooth structural changes as well as abrupt structural breaks. For abrupt structural breaks, we require that neither the number of breaks nor break dates be known. This is different from 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, our 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. The rate of local alternatives that our test can detect is $T^{-1/2}N^{-1/2}$, 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 proposed by Su and Wang (2017). This is an appealing advantage of the DFT, which avoids smoothed nonparametric estimation. 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 existing parametric tests can only capture a single structural break with a rate of $T^{-1/2}$. More importantly, as mentioned by Chen *et al.* (2014), the order $T^{-1/2}N^{-1/2}$ is the upper bound of structural changes in factor loadings that guarantees consistent estimation of true factor space and the number of common factors. That is, if the order of magnitude of structural changes in factor loadings dominates $T^{-1/2}N^{-1/2}$, the estimated factor loadings and number of factors would be inconsistent. As a result, we can detect structural changes in factor loadings that may lead to inconsistent estimation of factor loadings and the number of common factors. Simulation studies also demonstrate the significant power improvement of our test over the existing tests in the factor model literature.

Third, our test is tuning parameter-free and so is practically simple. It avoids the delicate business of choosing a bandwidth and the arbitrariness of specifying a trimming parameter. In comparison, the power of the smoothed nonparametric test by Su and Wang (2017) depends on the choice of bandwidth h. While they propose a bootstrap test to relieve this problem, the power of their test is still sensitive to the choice of bandwidth h in finite samples. On the other hand, 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 a pre-specified trimming parameter and hence will miss structural changes in the boundary regions of the sample.

Finally, our test is robust to both cross-sectional dependence and serial correlation of unknown form in error terms. Su and Wang (2017) allow for cross-sectional dependence, but require error terms to be martingale difference sequences. Hence, it assumes that all serial dependence in mean in observed data is due to the small dimensional common factors. This may be restrictive for factor analysis with macroeconomic time series, including multi-country or multi-sector factor models. We relax this assumption to allow for serial dependence of unknown form in error terms, and hence broaden applicability of the proposed test.

The paper is organized as follows. We introduce our test in Section 2 and establish its asymptotic theory in Section 3. We then examine 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, 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 $\|A\|(\equiv [\sum_{i=1}^m \sum_{j=1}^n |A_{ij}|^2]^{1/2})$, its spectral norm as $\|A\|_{\text{sp}}(\equiv \sqrt{\mu_1(A'A)})$, where "$\equiv$" means "is defined as", and $\mu_s(\cdot)$ denotes the sth largest eigenvalue of a real symmetric matrix by counting eigenvalues of multiplicity multiple times. We use "$\xrightarrow{p}$" to denote convergence in probability, "$\xrightarrow{d}$" convergence in distribution, "$\Rightarrow$" weak convergence, "plim" the probability limit, and $C \in (0, \infty)$ a generic positive constant that may vary from case to case. We use $(T, N) \to \infty$ to denote that T and N go to infinity jointly.

2 Hypotheses and Test Statistic

In this section, we introduce the hypotheses of interest and propose 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 a panel dataset or the ith random variable in a multivariate time series dataset. 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 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

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

Obviously, under $\mathbb{H}_0$, λ_{it} is time-invariant, and model (2.1) degenerates to a conventional factor model with time-invariant factor loadings. This model has been elaborately studied in the factor model literature (e.g., Stock and Watson, 2002; Bai and Ng, 2002; Bai, 2003). However, since a dataset 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), Han and Inoue (2015), and Cheng *et al.* (2016). 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. 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 piecewise smooth function, Su and Wang (2017) allow for both smooth structural changes and abrupt structural breaks in factor loadings. In comparison, the setting of our test is more general. We do not need to assume that λ_{it} is a smooth deterministic function of scaled time ratio t/T. Thus, the alternative (2.3) allows for 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 and/or an unknown number of breaks.

2.2 Test Statistic

Under $\mathbb{H}_0$, we can follow Bai and Ng (2002) and Bai (2003) to apply the PCA method to estimate the following model

$$X_{it} = \lambda_{i0}' F_t + \varepsilon_{it}^{\dagger}, \tag{2.4}$$

where $\varepsilon_{it}^{\dagger} = \varepsilon_{it}$ under $\mathbb{H}_0$ and they are different under $\mathbb{H}_A$.

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

$$X = F\Lambda_0' + \varepsilon^{\dagger}.$$

The PCA method solves the following minimization problem:

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

under certain identification restrictions. 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} under the above identification restrictions. Put $\hat{F} = (\hat{F}_1, ..., \hat{F}_T)'$ and $\hat{\Lambda}_0 = (\hat{\lambda}_{10}, ..., \hat{\lambda}_{N0})'$. It is well known that $\hat{F}$ is $\sqrt{T}$ times the eigenvectors corresponding to the R largest eigenvalues of the $T \times T$ matrix XX', and $\hat{\Lambda}_0' = (\hat{F}'\hat{F})^{-1}\hat{F}'X = T^{-1}\hat{F}'X$. Under $\mathbb{H}_0$, $\hat{F}$ and $\hat{\Lambda}_0$ are consistent for F and Λ_0 up to a rotation matrix H as shown in Bai (2003). While under the alternative $\mathbb{H}_A$, $\hat{\Lambda}_0$ cannot capture the time-varying property of $\Lambda_t \equiv (\lambda_{1t}, ..., \lambda_{Nt})'$, and it is no longer consistent for the true common factor loadings Λ_t up to the rotation matrix H. Fortunately, unlike such existing tests as Su and Wang (2017), our approach does not require to estimate the unrestricted model under $\mathbb{H}_A$. Hence, we do not need to impose additional smoothness conditions on the time-varying factor loadings, which allows our test to capture a broader range of structural changes in factor loadings.

Having obtained the restricted estimators $\hat{F}_t$ and $\hat{\lambda}_{i0}$, we define the following complex-valued empirical process:

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

where $\hat{\varepsilon}_{it} = X_{it} - \hat{\lambda}_{i0}'\hat{F}_t$ is the estimated residuals from PCA. To construct $\hat{A}(u)$, we first perform a DFT of $\hat{F}_t\hat{\varepsilon}_{it}$ for each i, and then take an average over cross-sectional units. Our test is based on $\hat{A}(u)$. The intuition behind is straightforward: if the factor loadings suffer from structural changes, then PCA fails to capture the time-varying behavior of λ_{it}, and such information will be hidden in the estimated residuals $\hat{\varepsilon}_{it}$. With DFT, we can reveal such information in frequency domain, because the possible time-varying behavior of the factor loadings can be entirely captured by the DFT of $\hat{\varepsilon}_{it}$. By examining the pattern of DFT at each frequency, we can detect structural changes of unknown type. Compared to the existing tests that are based on time domain analysis, the DFT-based approach does not require 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 Su and Wang's (2017) test does not require to specify the change to be abrupt or smooth, it requires nonparametric local smoothing in time domain. In contrast, our DFT-based test is free of the aforementioned issues.

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

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

where we define

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

and $\hat{G}_t(u) = \left(e^{\mathbf{i}u2\pi t/T} - T^{-1}\sum_{t=1}^{T}\hat{F}_t\hat{F}_t'e^{\mathbf{i}u2\pi t/T}\right)\hat{F}_t$. Under regularity conditions (see assumptions in Proposition 1), we can show that $\hat{G}_t(u)$ is asymptotically equivalent to

$$G_t(u) \equiv H_0'F_t\left(e^{\mathbf{i}u2\pi t/T} - \int_0^1 e^{\mathbf{i}u2\pi\tau}d\tau\right),$$

where H_0 is the probability limit of the rotation matrix H as defined in Bai (2003). The first component $\hat{A}_1(u)$ captures the structural changes in factor loadings since it is asymptotically equivalent to a pseudo-covariance between λ_{it} and the Fourier basis function of time. Given the orthogonality conditions between F_t and ε_{it}, the second component $\hat{A}_2(u)$ is a pure noise term, and it determines the asymptotic distribution. Intuitively, the DFT $\hat{A}(u)$ is equivalent to a linear projection of X_{it} onto frequency domain. The projection vector $\hat{G}_t(u)$ 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 factor loadings are constant over time. If the unknown factor loadings have structural changes, *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 in X_{it}, the linear projection of X_{it} cannot pass the filter $\hat{G}_t(u)$ and will converge to a non-constant spectrum. In contrast, when there is no structural change, *i.e.,* λ_{it} is a constant function of time, the linear projection of X_{it} converges to a zero spectrum at all frequencies.

To ensure that our DFT approach can detect a wide range of structural changes of unknown type, we shall examine the deviation of $\hat{A}(u)$ from a zero spectrum at each frequency u. We consider the following test statistic:

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

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 possibly with 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. To avoid numerical integration in (2.5), we follow Hong *et al.* (2017) to use the $N(0,1)$ density function. Then the test statistic becomes:

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

3 Asymptotic Theory

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

3.1 Assumptions

Let $\gamma_N(s,t) = N^{-1} E(\varepsilon_s' \varepsilon_t)$, $\zeta_{st} = N^{-1}[\varepsilon_s' \varepsilon_t - E(\varepsilon_s' \varepsilon_t)]$, $\gamma_{N,FF}(s,t) = N^{-1} E(F_s \varepsilon_s' \varepsilon_t F_t')$, and $\tau_{ij,st} = E(\varepsilon_{it} \varepsilon_{js})$. 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) $E(F_t F_t') = \Sigma_F$ for some $R \times R$ positive definite matrix Σ_F; (ii) $\max_t E\|F_t\|^{8+\delta} \le C$ for some $\delta > 0$.

Assumption A.2 [Factor Loadings]: (i) $\{\lambda_{i0}, i = 1, ..., N\}$ 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 terms]: (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 $\alpha_i(\cdot)$, where $\alpha(\cdot) \equiv \max_i \alpha_i(\cdot)$ satisfies $\sum_{s=1}^{\infty} \alpha(s)^{\delta/(2+\delta)} \le C$ for some $\delta > 0$; (iii) $\max_t \sum_{s=1}^{T} |\gamma_N(s,t)| \le C$, $\max_{s,t} E\left|N^{1/2} \zeta_{st}\right|^4 \le C$, $\max_t E|N^{-1/2} \sum_{i=1}^{N} [\varepsilon_{it}^2 - E(\varepsilon_{it}^2)]|^4 \le C$; (iv) $\max_t \sum_{s=1}^{T} |\gamma_{N,FF}(s,t)| \le C$, $\max_{t \ne r} E\|N^{-1/2} F_t \varepsilon_t' \varepsilon_r F_r'\|^4 \le C$, and $N^{-1} T^{-1} \sum_{i,j=1}^{N} \sum_{s,t=1}^{T} |\tau_{ij,st}| \le C$; (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 a nonnegative, symmetric, continuous, and integrable weighting function; (ii) $\int_{\mathbb{R}} |u|^4 W(u) du \le C$.

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 $E(F_t F_t') = \Sigma_F$ is homogeneous over t. This assumption implies that there is no structural change in the second moment of F_t. It greatly facilitates the derivation of asymptotic results and can be regarded as an identification condition. As is well known, latent common factors and 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 moments of common factors, we can always write that $\lambda_i' F_t = \lambda_i' \mathcal{L}(t/T)^{-1}\mathcal{L}(t/T)F_t = \lambda_{it}^{*\prime} F_t^*$ for any nonsingular matrix $\mathcal{L}(t/T)$ with $\lambda_{it}^* = \left[\mathcal{L}(t/T)^{-1}\right]' \lambda_i$ and $F_t^* = \mathcal{L}(t/T)F_t$ being time-varying factor loadings and common factors with time-varying second moments. 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 error terms and their interactions with 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 error terms. A.3(ii) requires the process $\{(\varepsilon_{it}, F_t), t = 1, 2, ...\}$ to be strong mixing with an algebraic mixing rate. With a more complicated notation, one could allow different individual time series to have various mixing rates and relax the mixing summability condition to $\limsup_N \frac{1}{N}\sum_{i=1}^N \sum_{s=1}^\infty \alpha_i(s)^{\delta/(1+\delta)} \leq C$. 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 satisfied 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). A variety of error terms satisfy these conditions, such as that (i) $\{\varepsilon_{it}\}$ is i.i.d. over t and cross-sectionally independent over i; (ii) $\{\varepsilon_{it}\}$ is i.i.d. over t but has cross-sectional dependence with $\max_{i,t}\sum_{j=1}^N |E(\varepsilon_{it}\varepsilon_{jt})| \leq C$; (iii) $\{\varepsilon_{it}\}$ is cross-sectionally independent and satisfies the α-mixing condition (*e.g.*, ARMA, bilinear, or ARCH); (iv) $\{\varepsilon_{it}\}$ has both serial dependence and cross-sectional dependence with $\max_{i,t}\sum_{j=1}^N \sum_{s=1}^T |E(\varepsilon_{it}\varepsilon_{js})| \leq C$ and the α-mixing condition. Assumption A.4 imposes mild conditions on the weighting function. It ensures that the integral in (2.5) is well-defined.

3.2 Asymptotic Null Distribution

When the unknown factor loadings are constant over time, the asymptotic results established by Bai (2003) hold. We now state the asymptotic distribution of $\hat{A}(u)$ under $\mathbb{H}_0$.

Proposition 1 *Suppose Assumptions A.1-A.3 and $\mathbb{H}_0 : \lambda_{it} = \lambda_{i0}$ for all i hold. Then as $(T, N) \to \infty$,*

$$\sqrt{NT}\hat{A}(u) = \frac{1}{\sqrt{NT}}\sum_{i=1}^N \sum_{t=1}^T G_t(u)B_i\varepsilon_{it} + O_P(N^{1/2}T^{-1}) + o_P(1),$$

where $B_i = 1 - \lambda_{i0}'\Sigma_{\Lambda_0}^{-1}\bar{\lambda}_0$, and $\bar{\lambda}_0 = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N \lambda_{i0}$. If in addition $T \propto N^\nu$ with $\nu > 1/2$, then

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

Under $\mathbb{H}_0$, the asymptotic behavior of $\hat{A}(u)$ depends on the relative speed between N and T. When $\sqrt{N}/T \to 0$, the dominant term of $\hat{A}(u)$ is a weighted average of error terms $\{\varepsilon_{it}\}$, which will converge to a zero-spectrum. 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) that the estimation impact on factors depends on the relative speed between N and T. We allow that N and T increase at the same rate. 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 dominant term now consists of two components, which are of the same order of magnitude. The first component is the same as the dominant term in the case of $\nu > 1/2$, while the second component arises due to the serial correlation in error terms and it is asymptotically equivalent to a long-run pseudo-covariance between the Fourier series and error terms. If we follow Su and Wang (2017) to impose the martingale difference sequence assumption for error terms, this second component of the dominant term will vanish up to a higher order. That is, if we rule out serial dependence in error terms, we do not need to impose any restriction on the relative speed between N and T. However, since serial correlation is common in macroeconomic and financial data, we allow for serial correlation in error terms and impose a sufficient condition on the relative speed between N and T when deriving the asymptotic distribution of the proposed test statistic.

Our asymptotic results are obtained with large N and large T. Theoretically speaking, the above relative speed between N and T also allows for the classical factor model (see Lawley and Maxwell, 1971; Anderson, 1984) and the approximate factor model (see Chamberlain and Rothschild, 1983) with large T and fixed N. However, as Anderson (1984) and Bai (2003) point out, with a fixed N, one can consistently estimate factor loadings but not common factors. Since the estimated common factors are used in $\hat{A}(u)$ and the test statistic $\hat{D}$, we do not consider this case here. In addition, for the case with fixed T and large N, under the assumption of the martingale difference sequence for error terms, we could 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. However, since the test statistic $\hat{D}$ is constructed based on the DFT over time domain, we establish the asymptotic distribution theory as $T \to \infty$. Hence, we rule out the cases of fixed T or fixed N in this paper.

We now derive the asymptotic distribution of the test statistic $\hat{D}$ under $\mathbb{H}_0$.

Theorem 1 *Suppose Assumptions A.1-A.4 hold, and $T \propto N^{\nu}$ with $\nu > 1/2$. Then under $\mathbb{H}_0 : \lambda_{it} = \lambda_{i0}$ for all i, as $(T, N) \to \infty$,*

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

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

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

and $M_t(u) = e^{\mathbf{i}u2\pi t/T} - \int_0^1 e^{\mathbf{i}u2\pi\tau}d\tau$ *is a demeaned Fourier process. Also,* $H_0 = \text{plim}\, H = Q_0^{-1}$*,* $Q_0 = V_0^{1/2}\Upsilon_0'\Sigma_{\Lambda_0}^{-1/2}$*,* V_0 *is an* $R \times R$ *diagonal matrix containing the* R *eigenvalues of* $\Sigma_{\Lambda_0}^{1/2}\Sigma_F\Sigma_{\Lambda_0}^{1/2}$ *in decreasing order, and* Υ_0 *is the* $R \times R$ *corresponding eigenvector matrix of* $\Sigma_{\Lambda_0}^{1/2}\Sigma_F\Sigma_{\Lambda_0}^{1/2}$ *such that* $\Upsilon_0'\Upsilon_0 = \mathbb{I}_R$.

Theorem 1 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 relative speed between N and T is important for our asymptotic theory. As shown in Proposition 1, if $T \propto N^\nu$ with $\nu \leq 1/2$, the dominant term of $\hat{A}(u)$ will contain an additional component with same ($\nu = 1/2$) or dominant ($\nu < 1/2$) order of magnitude, which will jointly determine the asymptotic distribution. For simplicity, we impose the condition that $\nu > 1/2$ so that the second term in $\hat{A}(u)$ becomes asymptotically negligible. This second component rises due to the existence of serial dependence in $\{\varepsilon_{it}\}$. Hence, if error terms are serially uncorrelated for all i, the condition on the relative speed between N and T is not necessary. We note that Bai (2003) requires $\nu > 1/2$ as well to ensure the asymptotical normality of estimated common factors. 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 these related works which impose restrictions on the upper bound of rate ν, we only impose a restriction on the low bound of ν, which is mild.

3.3 Asymptotic Local Power

When there exist structural changes, the estimated factor loading $\hat{\lambda}_{i0}$ is no longer consistent for the true factor loading λ_{it} up to a rotation matrix H defined in Bai (2003). Let V_{NT} be the $R \times R$ diagonal matrix of the first R largest eigenvalues of $\frac{1}{NT}XX'$ in decreasing order, and $\hat{F} = (\hat{F}_1, ..., \hat{F}_T)'$ be the corresponding $T \times R$ matrix that consists of the R largest eigenvectors of XX'. Following Bai (2003), we could show

$$\hat{F}_t - H_t'F_t \quad = \quad V_{NT}^{-1}\left(\frac{1}{T}\sum_{s=1}^T \hat{F}_s\gamma_N(s,t) + \frac{1}{T}\sum_{s=1}^T \hat{F}_s\zeta_{st} + \frac{1}{T}\sum_{s=1}^T \hat{F}_s\tilde{\eta}_{st} + \frac{1}{T}\sum_{s=1}^T \hat{F}_s\tilde{\xi}_{st}\right),$$

where $H_t = N^{-1}T^{-1}\Lambda_t'(\sum_{s=1}^T \Lambda_s F_s \hat{F}_s')V_{NT}^{-1}$ is an $R \times R$ matrix with $\Lambda_t = (\lambda_{1t}, \cdots, \lambda_{Nt})'$ being an $N \times R$ time-varying factor loading matrix, $\tilde{\eta}_{st} = F_s'\Lambda_s'\varepsilon_t/N$, and $\tilde{\xi}_{st} = F_t'\Lambda_t'\varepsilon_s/N$. Compared to Bai (2003), the rotation matrix H_t depends on time t because it contains the unknown factor loadings Λ_t. When there is no structural change, it is straightforward to show that $H_t = H \equiv (\Lambda_0'\Lambda_0/N)(F'\hat{F}/T)V_{NT}^{-1}$.

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

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

where $a_{NT} \to 0$ as $(T, N) \to \infty$. The rate a_{NT} controls the speed at which the local alternative $\mathbb{H}_A\left(a_{NT}\right)$ converges to the null hypothesis $\mathbb{H}_0$, and g_{it} is a deterministic function of time t for each i. We note that the local alternative $\mathbb{H}_A\left(a_{NT}\right)$ does not impose any smoothness condition on g_{it}. This is more general than the setting of Su and Wang (2017), who require g_{it} to be a piecewise 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 shall assume

$$\frac{1}{T}\sum_{t=1}^{T} g_{it}=\mathbf{0},\forall i,$$

for the purpose of location normalization, where $\mathbf{0}$ is an $R\times 1$ zero vector. If $g_{it}=g_i(t/T)$, it is equivalent to impose the following condition

$$\int_0^1 g_i(u)du=\mathbf{0},\forall i.$$

This normalization greatly simplifies local asymptotic power analysis. Both λ_{i0} and g_{it} can depend on N and T. For notational simplicity, we continue to write them as λ_{i0} and g_{it}.

Under $\mathbb{H}_A(a_{NT})$, Lemma A.2 in the appendix implies that $H_t=H+O_p(a_{NT})$. Thus, the PCA estimator $\hat{F}$ will be consistent for $H'F_t$ under the local alternative $\mathbb{H}_A(a_{NT})$. To study the asymptotic behavior of $\hat{A}(u)$, we add the following assumption:

Assumption A.5 [Local Alternative]: (i) The $R\times R$ matrix satisfies $E\left\|\frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\sum_{i=1}^{N}F_t g_{it}'\varepsilon_{it}\right\|^2<C$; (ii) $\max_{i,t}\|g_{it}\|<C$.

Assumption 5 restricts the size of local alternatives, but does not impose any restriction on the type of structural changes. The proposition below shows that $\sqrt{NT}\hat{A}(u)$ weakly converges to a complex-valued Gaussian process with a nonzero mean under $\mathbb{H}_A(a_{NT})$.

Proposition 2 *Suppose Assumptions A.1-A.3 and A.5 hold. Then under $\mathbb{H}_A(a_{NT})$ with $a_{NT}=(NT)^{-1/2}$, as $(T,N)\to\infty$,*

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

where $\psi(u)=Q_0\lim_{N,T\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}g_{it}e^{\mathbf{i}u2\pi t/T}$, and Q_0 and $\mathcal{G}(u)$ are defined in Proposition 1 and Theorem 1, respectively. In particular, if $g_{it}=g_i(\frac{t}{T})$, then $\psi(u)=Q_0\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\widetilde{cov}\left[e^{\mathbf{i}u2\pi\tau},g_i(\tau)\right]$, where $\widetilde{cov}[e^{\mathbf{i}u2\pi\tau},g_i(\tau)]=\int_0^1 e^{\mathbf{i}u2\pi\tau}g_i(\tau)\,d\tau$ is a pseudo-covariance.

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

Theorem 2 *Suppose Assumptions A.1-A.5 hold, and $T\propto N^{\nu}$ with $\nu>1/2$. Then under $\mathbb{H}_A(a_{NT})$ with $a_{NT}=(NT)^{-1/2}$, as $(T,N)\to\infty$,*

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

where $\psi(u)$ and $\mathcal{G}(u)$ are defined in Proposition 2 and Theorem 1, respectively.

Theorem 2 provides the asymptotic distribution of $\hat{D}$ under the local alternative $\mathbb{H}_A(a_{NT})$. It shows that our test can detect a class of local alternatives with $\psi(u) \neq 0$, at 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 can only detect the local alternative $\mathbb{H}_A(a_{NT})$ with rate $a_{NT} = T^{-1/2}N^{-1/4}h^{-1/4}$. That is an advantage of DFT, which avoids nonparametric smoothing over t/T.

In addition, we allow for various kinds of structural changes in factor loadings, including smooth structural changes, a single structural break, 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 factors and then test for the structural changes in such a linear regression by constructing the sup-Wald and sup-LM type 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 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 a class of local alternatives that converge to the null hypothesis at rate $T^{-1/2}$, which is slower than our rate $a_{NT} = T^{-1/2}N^{-1/2}$.

In fact, the rate $a_{NT} = T^{-1/2}N^{-1/2}$ is the upper bound of structural changes in factor loadings that guarantees consistency of the estimated number of common factors by Bai and Ng (2002) and estimated factors loadings by PCA. If the magnitude of structural changes is a higher order term of $T^{-1/2}N^{-1/2}$, then the estimated factor loadings given by PCA and the number of common factors determined by Bai and Ng (2002) are consistent. This order of magnitude corresponds to the definition of a small break by Chen *et al.* (2014). For such small structural changes, our test has no power. In contrast, if the magnitude of structural changes is dominant over $T^{-1/2}N^{-1/2}$, it leads to inconsistency of estimated factor loadings and the number of common factors. Thus, our test has nontrivial power to detect any structural changes that lead to inconsistent estimation of the number of common factors and factor loadings by PCA.

Furthermore, our test is tuning-parameter free. We require neither the smoothing parameter nor the trimming parameters. That is appealing in practice because there have been no criteria to choose the optimal bandwidth for the nonparametric smoothing test of Su and Wang (2017) or the trimming parameter for the aforementioned parametric tests. The result of a smoothed nonparametric test can be severely affected by the choice of the bandwidth. Even if one uses the bootstrap to correct size, the power of nonparametric smoothing test is still sensitive to the choice of the bandwidth. Moreover, the proposed test can detect 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 structural changes in the boundary regions.

3.4 A Bootstrap Version of the Test

The asymptotic null distribution of $\hat{D}$ is not pivotal, as it depends on the unknown data generating process. We can use resampling methods to obtain critical values in finite samples. As Su and Wang (2017) point out, the wild bootstrap may miss cross-sectional dependence in error terms. Hence, we follow Su and Wang (2017) to use a modified parametric bootstrap procedure that tries to mimic the cross-sectional dependence in error terms. Let $\hat{\varepsilon}_t = (\hat{\varepsilon}_{1t}, ..., \hat{\varepsilon}_{Nt})'$, $\hat{\Sigma}^0 = T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_t\hat{\varepsilon}_t'$. Let $\hat{\sigma}_{ij}^0$ denote the (i,j)th element of $\hat{\Sigma}^0$. Define the shrinkage version of $\hat{\Sigma}^0$ as $\hat{\Sigma}$, whose (i,j)th element is given by

$$\hat{\sigma}_{ij} = \hat{\sigma}_{ij}^0(1-\epsilon)^{|j-i|},$$

where ϵ is a small positive number (*e.g.*, 0.01) to ensure the maximum absolute column/row sum norm of $\hat{\Sigma}$ to be stochastically bounded provided $\max_{i,j}|\hat{\sigma}_{ij}^0|$ is stochastically bounded. By construction, $\hat{\Sigma}$ is also symmetric and positive semi-definite. Following Su and Wang (2017), we use the following bootstrap procedure:

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

(ii) For $i = 1, 2, ..., N$ and $t = 1, 2, ..., T$, obtain the bootstrap error terms $\varepsilon_t^b = \hat{\Sigma}^{1/2}\eta_t$, where $\eta_t = (\eta_{1t}, ..., \eta_{Nt})'$ with η_{it} being i.i.d. $N(0,1)$ across both i and t. Generate $X_{it}^b = \hat{\lambda}_{i0}'\hat{F}_t + \varepsilon_{it}^b$.

(iii) Run PCA on $\{X_{it}^b\}_{i=1,t=1}^{N,T}$ and compute the test statistic $\hat{D}^b$.

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

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

Following the proof of Theorem 4.5 in Su and Wang (2017), we could show that the above bootstrap procedure provides an asymptotic valid approximation to the limiting distribution of $\hat{D}$ under $\mathbb{H}_0$. However, this bootstrap could only mimic cross-sectional dependence in error terms. It works well if error terms do not exhibit serial correlation or only exhibit fairly weak serial correlation, but it tends to be oversized in the presence of moderate or strong serial correlation in error terms.

To account for possible serial correlation and cross-sectional dependence of unknown form in error terms, 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 $T \to \infty$.

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

(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 an $N \times 1$ vector, and $\Xi_t = [\bar{\varepsilon}_t, \bar{\varepsilon}_{t+1}, ..., \bar{\varepsilon}_{t+l_T-1}]$ is an $N \times l_T$ matrix. Resample

$\{\Xi_t\}_{t=1}^{T-l_T+1}$ with replacement to form a bootstrap dataset $\{\Xi_t^b\}_{t=1}^{L}$ satisfying $L = \lfloor T/l_T \rfloor + 1$; Let $\{\varepsilon_t^b\}_{t=1}^{T}$ be the first T column vectors of $\{\Xi_t^b\}_{t=1}^{L}$.

(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 $\hat{D}^b$.

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

(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 simulation studies below demonstrate excellent finite sample performance of the proposed MBB procedure for our test. We note that the MBB procedure proposed by Gonçlaves (2011) requires $T \to \infty$ faster than N, i.e., $\nu > 1$. However, this additional restriction does not affect the applicability of our test to cases 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 Processes

We generate data under the framework of large dimensional 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 $F_{1,t} = 0.6F_{1,t-1} + u_{1t}$, $u_{1t} \sim i.i.d.N(0, 1 - 0.6^2)$; and $F_{2,t} = 0.3F_{2,t-1} + u_{2t}$, $u_{2t} \sim i.i.d.N(0, 1 - 0.3^2)$.

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

$\lambda_{it,2} = \lambda_{i0,2} \sim i.i.d.N(1, 1)$;

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 terms ε_{it}: (i) the i.i.d. case, where $\varepsilon_{it} \sim i.i.d.N(0,1)$; (ii) the heteroskedastic case, where $\varepsilon_{it} = \sigma_i v_{it}$, $\sigma_i \sim i.i.d.U(0.5, 1.5), v_{it} \sim i.i.d.N(0,1)$; (iii) the cross-sectionally dependent case, where $\varepsilon_{\cdot t} \sim i.i.d.N(0, \Sigma_\varepsilon)$; (iv) the serially correlated case, where $\varepsilon_{it} = 0.5\varepsilon_{it-1} + v_{it}, v_{it} \sim i.i.d.N(0,1)$; (v) the cross-sectionally dependent and serially correlated case, where $\varepsilon_{\cdot t} = 0.5\varepsilon_{\cdot t-1} + v_{\cdot t}, v_{\cdot t} \sim i.i.d.N(0, \Sigma_v)$, where $\Sigma_\varepsilon = \Sigma_v = (c_{ij})_{i,j=1,\ldots,N}$ with $c_{ij} = 0.5^{|i-j|}$ for cases (iii) 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, serial dependence, and both cross-sectional and serial dependence, respectively. DGPs.P1-P3 describe various time-varying factor loadings. Among them, DGPs.P1-P2 have a single abrupt structural break and multiple abrupt structural breaks, respectively, while DGP.P3 describes a smooth structural change. We check the power of all the tests by using DGPs.P1-P3 with various types of error terms.

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 their test statistics. 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 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 (Newey and West, 1987). We use the critical values presented in Andrews (1993) for the tests of Breitung and Eickmeier (2011), Chen *et al.* (2014), and Han and Innoe (2015). We apply bootstrap procedures suitable for Su and Wang's (2017) test and ours respectively. We set the number of bootstrap iterations $B = 200$.

For cases (i) to (iii) of serially uncorrelated error terms, we simulate 500 datasets with sample sizes $N = 100$, 200, and $T = 100$, 200, respectively. We apply bootstrap procedures proposed by Su and Wang (2017) for Su and Wang's (2017) test and our test. For the cases of (iv) and (v) with serially correlated error terms, we need to use the MBB procedure proposed by Gonçlaves (2011). As noted earlier, MBB requires that $T \to \infty$ faster than N. Hence, we consider the sample sizes $N = 40, 80$ and $T = 100, 200$ for cases (iv) and (v) with serially correlated error terms.

4.2 Results with Serially Uncorrelated Error Terms

Table 1 reports the size 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 for cases (i) to (iii) with serially uncorrelated error terms, where the number of common factors is fixed at $R = 2$. As shown in

Table 1: Size of tests under DGP.S1 with serially uncorrelated error terms 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 terms: $\varepsilon_{it} \sim i.i.d.N(0,1)$															
100	100	5.6	10.6	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	4.6	9.6	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	6.0	10.0	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	4.2	9.4	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 terms: $\varepsilon_{it} = \sigma_i v_{it}$, $\sigma_i \sim i.i.d.U(0.5, 1.5), v_{it} \sim i.i.d.N(0,1)$															
100	100	6.0	11.0	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	4.6	8.8	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	7.0	10.0	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	4.2	9.2	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-sectionally dependent error terms: $\varepsilon_{\cdot t} \sim i.i.d.N(0, \Sigma_\varepsilon)$															
100	100	8.2	11.0	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	8.6	14.2	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

Notes: (i) D_B denotes the bootstrap $\hat{D}$ test; (ii) SW denotes Su and Wang's (2017) bootstrap 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.

Table 1, both our test and Su and Wang's (2017) test have reasonable size using bootstrap critical values. Han and Inoue's (2015) sup-LM and sup-Wald tests tend to under-reject. Chen *et al.*'s (2014) sup-Wald test has reasonable size, but their sup-LM test also exhibits under-rejection. Breitung and Eickmeier's (2011) test suffers from slight under-rejection.

Table 2 reports the power of the tests under DGPs.P1-P3 at the 5% and 10% significance levels for cases (i) to (iii) with serially uncorrelated error terms when the number of common factors $R = 2$. Our test is most powerful in detecting all forms of time-varying factor loadings under DGPs.P1-P3, and its power increases as either T or N increases. Recall that DGPs.P1-P2 are factor models with abrupt structural breaks, and DGP.P3 is a factor model with smooth structural changes. The simulation results demonstrate the excellent performance of our test in detecting both abrupt structural breaks and smooth structural changes. Moreover, Su and Wang's (2017) test is also powerful in detecting structural changes under these DGPs, but the rejection rates are all lower than our test. These results 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 lower power against DGPs.P1-P3.

Since the exact number R of common factors is typically unknown in practice, one should determine the number of common factors before estimation and testing. In the literature on testing for structural breaks in factor loadings, the number of common factors is either determined using Bai and Ng's (2002)

Table 2: Power of tests under DGPs.P1-P3 with serially uncorrelated error terms 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 terms: $\varepsilon_{it} \sim i.i.d.N(0,1)$																
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
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 terms: $\varepsilon_{it} = \sigma_i v_{it}$, $\sigma_i \sim i.i.d.U(0.5, 1.5), v_{it} \sim i.i.d.N(0,1)$																
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
DGP.P3	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-sectionally dependent error terms: $\varepsilon_{\cdot t} \sim i.i.d.N(0, \Sigma_\varepsilon)$																
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
	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
DGP.P3	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

Notes: See the notes in Table 1.

information criteria (*e.g.*, Han and Inoue, 2015) or specified as some ad hoc fixed number, 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) nonparametric method to determine the number of factors which is robust to structural changes in factor loadings. In general, all the aforementioned methods can only select the correct number of factors consistently under $\mathbb{H}_0$ except Su and Wang (2017), which has been proven valid even under the alternative. Indeed, if we apply their 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 for possible misspecification of the number of factors under the alternative, we follow Han and Inoue (2015) and select the number of factors based on Bai and Ng's (2002) 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 2. To save space, we relegate the results based on IC_{p1} to the appendix. Moreover, we also examine the performance of our test as well as other various tests by setting the number of common factors to 3. The power of our bootstrap test is a bit lower than in the case of correctly specified number of factors as reported in Tables 2. 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. The results are reported in the appendix.

4.3 Results with Serially Correlated Error Terms

We now check the performance of our test using the MBB procedure proposed by Gonçlaves (2011) under serially correlated error terms. Table 3 reports the size 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 for cases (iv) and (v) with serially correlated error terms, where the number of common factors is fixed at the true value $R = 2$. As shown in Table 1, based on bootstrap critical values, our test tends to over-reject a bit but is still acceptable. Han and Inoue's (2015) sup-LM and sup-Wald tests also tend to under-reject. Chen *et al.*'s (2014) sup-Wald test has reasonable size, but their sup-LM test exhibits under-rejection. However, when error terms have serial correlation, Su and Wang's (2017) test displays serious over-rejection and the rejection rates even achieve 100%, which is not surprising since Su and Wang (2017) require error terms to be martingale difference sequences. On the other hand, Breitung and Eickmeier's (2011) test also suffers from severe over-rejection. In fact, Assumption 2 in Breitung and Eickmeier (2011) requires error terms to be serially independent, which does not hold for cases (iv) and (v) with serially correlated error terms. To compare our test with Su and Wang (2017) under a fair ground, we also apply MBB for Su and Wang's (2017) test. The results are given in the last two column in Table 3. Now, Su and Wang's (2017) test has reasonable size. Hence, if there exists serial correlation, one should use MBB for Su and Wang's (2017) test.

Table 4 reports the power performance of the tests under DGPs.P1-P3 at the 5% and 10% significance levels for cases (iv) and (v) with serially correlated error terms when the number of common factors is

Table 3: Size of tests under DGP.S1 with serially correlated error terms 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}		SW, MBB	
		5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
serially correlated error terms: $\varepsilon_{it} = 0.5\varepsilon_{it-1} + v_{it}, v_{it} \sim i.i.d.N(0,1)$																	
40	100	6.2	11.6	93.6	97.0	1.2	4.6	0.2	1.8	3.0	7.0	7.0	13.2	13.4	22.8	5.4	12.4
40	200	5.4	13.6	98.2	99.6	2.0	7.4	3.4	8.4	3.4	7.8	4.4	9.8	18.1	28.0	5.8	10.6
80	100	8.8	16.2	99.4	99.8	0.2	3.4	0.4	1.2	2.4	7.4	6.2	13.2	13.6	22.6	1.0	6.8
80	200	5.2	9.8	100	100	2.0	5.4	1.4	5.2	4.0	9.6	4.2	9.8	17.9	28.1	6.0	11.0
cross-sectionally dependent and serially correlated error terms: $\varepsilon_{\cdot t} = 0.5\varepsilon_{\cdot t-1} + v_{\cdot t}, v_{\cdot t} \sim i.i.d.N(0, \Sigma_v)$																	
40	100	6.2	12.8	87.0	95.2	1.4	4.6	0.6	2.6	3.8	10.0	5.8	11.0	13.6	22.6	7.0	15.6
40	200	6.4	12.2	93.8	97.2	2.4	6.6	2.8	7.0	3.0	8.2	4.8	10.4	18.0	28.2	6.6	13.6
80	100	8.8	14.2	98.6	99.4	1.4	2.6	0.6	1.4	2.4	7.8	6.0	11.6	13.7	23.0	3.4	8.0
80	200	6.2	11.0	99.0	99.8	2.0	4.0	1.8	6.0	4.0	10.8	5.6	10.8	17.9	28.1	4.2	10.6

Notes: (i) D_B denotes the $\hat{D}$ bootstrap test; (ii) SW and SW, MBB denotes Su and Wang's (2017) tests using modified parametric bootstrap and moving blocks bootstrap respectively; (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.

fixed at the true value $R = 2$. Similar to the results in Table 2, our test is powerful in detecting all forms of structural changes in factor loadings under DGPs.P1-P3, while the parametric tests of Breitung and Eickmeier (2011), Chen *et al.* (2014), and Han and Inoue (2015) have quite low power against these DGPs. Since Su and Wang's (2017) test using the modified parametric bootstrap could achieve unity rejection under DGP.S1, it is not surprising to see its high rejection rates for DGPs.P1-P3 when error terms have serial dependence. However, if we use MBB for Su and Wang's test, its rejection rates are all lower than those of our test, which is consistent with our theory on the relative efficiency between our test and Su and Wang's (2017) test.

5 Application to U.S. Macroeconomic Data

We now check whether the U.S. macroeconomic dynamics suffers from structural changes. The dataset, 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 series have been standardized to have zero mean and unit variance. For the details of data description and processing, we 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 5, where we see that different methods favor different numbers of common

Table 4: Power of tests under DGPs.P1-P3 with serially correlated error terms 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}		SW, MBB	
			5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
serially correlated error terms: $\varepsilon_{it} = 0.5\varepsilon_{it-1} + v_{it}, v_{it} \sim i.i.d.N(0, 1)$																		
DGP.P1	40	100	54.2	67.2	93.2	96.6	2.2	6.4	1.2	3.0	2.2	6.6	6.6	14.6	16.1	26.0	14.2	24.0
	40	200	81.4	88.8	99.8	99.8	3.2	8.8	5.6	13.4	3.6	7.6	6.8	12.4	24.7	35.6	27.0	42.6
	80	100	82.2	90.6	99.4	99.8	0.8	4.4	0.8	2.8	2.0	7.8	5.6	13.6	16.6	26.3	9.8	24.0
	80	200	98.6	99.6	100	100	5.2	10.6	4.8	12.8	3.6	7.8	4.6	12.2	24.8	35.8	40.8	57.8
DGP.P2	40	100	11.6	19.6	87.0	92.6	1.8	5.0	0.6	2.8	2.6	8.0	6.2	11.2	13.3	22.7	5.4	12.8
	40	200	17.6	30.4	98.6	99.4	2.2	6.6	3.4	7.4	3.2	7.8	4.0	8.4	18.4	28.8	6.0	11.6
	80	100	14.6	25.4	98.0	99.2	0.4	4.2	0.6	1.6	2.0	7.0	5.8	11.2	13.6	22.6	2.6	6.0
	80	200	31.0	45.8	99.8	100	2.2	6.8	1.0	6.6	3.8	9.4	3.0	9.8	18.5	28.6	6.2	13.8
DGP.P3	40	100	33.4	49.4	92.8	96.2	1.4	4.4	1.0	3.4	2.8	7.8	7.8	12.6	14.9	24.3	7.2	16.4
	40	200	67.0	80.6	99.6	100	2.6	7.0	3.4	9.2	2.4	7.4	6.0	11.6	20.4	31.1	16.8	27.6
	80	100	56.2	70.8	99.2	99.6	0.2	2.4	0.4	1.8	2.0	6.4	6.2	14.8	14.9	24.2	5.4	12.4
	80	200	91.8	97.6	100	100	1.6	4.6	2.0	9.2	3.4	7.2	5.4	10.8	20.8	31.7	22.0	35.4
cross-sectionally dependent and serially correlated error terms: $\varepsilon_{\cdot t} = 0.5\varepsilon_{\cdot t-1} + v_{\cdot t}, v_{\cdot t} \sim i.i.d.N(0, \Sigma_v)$																		
DGP.P1	40	100	37.6	50.0	90.8	95.6	2.2	7.6	1.0	3.8	1.8	8.2	7.0	14.4	15.9	25.6	13.8	21.8
	40	200	66.0	77.6	97.4	99.2	4.2	9.2	5.0	12.6	3.2	8.4	6.4	12.6	24.9	35.8	22.0	35.6
	80	100	64.6	75.6	98.8	99.6	1.0	5.0	0.4	3.8	3.4	7.4	7.8	11.8	16.7	26.6	9.4	19.6
	80	200	92.4	97.2	100	100	4.4	13.0	6.4	13.6	3.6	9.2	5.2	12.4	24.6	35.5	29.6	46.6
DGP.P2	40	100	8.8	16.0	84.6	91.6	2.4	6.2	0.8	2.2	2.2	9.0	7.2	12.6	13.2	22.3	8.0	15.4
	40	200	14.8	25.0	93.6	96.2	2.6	6.8	2.2	6.8	2.2	8.4	3.6	10.0	18.8	28.7	7.8	16.4
	80	100	13.0	23.2	98.0	98.6	0.6	2.8	0.6	1.6	3.4	7.4	5.4	10.0	14.1	23.2	2.8	8.2
	80	200	24.0	33.0	99.6	99.8	2.6	6.6	1.0	7.2	4.0	10.6	4.8	11.4	18.2	28.5	7.0	12.8
DGP.P3	40	100	23.4	36.0	85.6	91.4	1.6	4.8	0.8	3.0	2.4	6.8	7.8	13.4	14.6	23.9	9.4	17.0
	40	200	48.0	61.8	97.0	99.2	2.0	5.2	3.8	9.0	3.0	8.6	5.6	11.8	20.5	31.4	16.4	27.4
	80	100	40.2	55.2	98.4	99.4	0.6	2.2	0.4	1.4	2.2	6.2	7.0	13.6	14.9	24.6	6.0	13.4
	80	200	75.8	85.8	99.8	100	1.4	3.6	2.6	7.8	2.8	7.8	5.8	11.4	20.3	31.1	19.2	33.2

Notes: See the notes in Table 3.

Table 5: Number of common factors determined by various criteria

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 6: Tests for structural changes in the U.S. macroeconomic dynamics

	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) Each column under D_B, SM_B, LM, and $Wald$ reports the the values of the corresponding test statistics; (ii) Columns under "5%" report the corresponding bootstrap critical values (Our test and Su and Wang's (2017) test) or asymptotic critical values (Han and Inoue's (2015) test and Chen *et al.*'s (2014) test); (iii) c denotes the bandwidth parameter in Su and Wang's (2017) nonparametric test. Bold entries indicate significance at the 5% significance level.

factors. Below, we report the 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 Inoue's (2015) sup-LM and sup-Wald tests, as well as Chen *et al.*'s (2014) sup-LM and sup-Wald tests to investigate 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 as those used in our simulation studies. For our test and Su and Wang's (2017) test, we focus on bootstrap results based on $B = 1000$ bootstrap iterations.

Table 6 reports the results of various tests 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 bandwidths. With 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 hypothesis 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 relatively low power of the tests of Chen *et al.* (2014) and Han and Inoue (2015). Furthermore, it shows the robust performance of our test in detecting structural changes of unknown type.

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 using a discrete Fourier transform approach. Our test can capture a wide range of smooth and abrupt structural changes in factor loadings with unknown break dates and an unknown number of breaks. More importantly, the proposed test is asymptotically more powerful than all the existing related tests in the factor model literature. Our test is tuning parameter-free, so it has power for structural changes in the boundary regions of the sample period. Another appealing feature of our test is its robustness to serial correlation and cross-sectional dependence of unknown form, which greatly extends the scope of applicability of our test to various factor models. 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 of abrupt structural breaks and smooth structural changes in finite samples. We apply our test to check whether the U.S. macroeconomic dynamics suffers from structural changes, and document significant and robust evidence against the time invariance hypothesis for factor loadings. Other existing tests give mixed results.

There are several interesting topics for further research. For instance, when our test rejects the null hypothesis, one can further check the types of structural changes and distinguish smooth structural changes from abrupt structural breaks. This is an interesting and challenging issue and we leave it to further research.

References

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

Anderson, T.W. (1984) *An introduction to multivariate statistical analysis.* New York: Wiley.

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

Andrews, W.K. (1994) Empirical process methods in econometrics. *Handbook of econometrics* 4, 2247-2294.

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.

Chamberlain, G. & M. Rothschild (1983) Arbitrage, factor structure and mean-variance analysis in large asset markets. *Econometrica* 51(5), 1281-1304.

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

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

Chow, G.C. (1960) Tests of Equality between Sets of Coefficients in Two Linear Regressions. *Econometrica* 28, 591-605.

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

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.

Lawley, D.N. & A.E. Maxwell (1971) *Factor analysis in a statistical method.* London: Butterworth.

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.

Stewart, G. & J. Sun (1990) *Matrix Perturbation Theory.* Boston: Academic Press.

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.

Online Supplement for
"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
[b] Cornell University and Xiamen University
[c] Sun Yat-Sen University

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

A Mathematical Appendix

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

A.1 Technical Lemmas

Lemma A.1 *Suppose Assumption A.1-A.3 and A.5 hold, under $\mathbb{H}_A(a_{NT})$ with $a_{NT} = N^{-1/2}T^{-1/2}$, as $(T,N)\to\infty$,*

(i). $V_{NT} = T^{-1}\hat{F}'(XX'/NT)\hat{F} \xrightarrow{p} V_0$,

(ii). $(\hat{F}'F/T)(\Lambda_0'\Lambda_0/N)(F'\hat{F}/T) \xrightarrow{p} V_0$,

where V_{NT} is an $R\times R$ diagonal matrix consisting the R largest eigenvalues of (XX'/NT), and V_0 is an $R\times R$ matrix consisting of the R eigenvalues of $\Sigma_{\Lambda_0}\Sigma_F$, both arranged in descending orders.

Proof. By the definition of V_{NT} and $\hat{F}$, we have

$$\frac{1}{NT}XX'\hat{F} = \hat{F}V_{NT}.$$

By the identification condition that $\hat{F}'\hat{F}/T = \mathbb{I}_R$, it follows that $V_{NT} = T^{-1}\hat{F}'(XX'/NT)\hat{F}$. Since $X = F\Lambda_0' + \varepsilon^{\dagger}$, where $\varepsilon^{\dagger} = \varepsilon + a_{NT}g^{\dagger}$, $g^{\dagger} = (g_1^{\dagger\prime}, ..., g_T^{\dagger\prime})'$, and $g_t^{\dagger} = (g_{1t}'F_t, ..., g_{Nt}'F_t)'$, we have

$$\begin{aligned} V_{NT} &= \frac{\hat{F}'}{T}\frac{1}{NT}\left(F\Lambda_0'\Lambda_0F' + F\Lambda_0'\varepsilon^{\dagger\prime} + \varepsilon^{\dagger}\Lambda_0F' + \varepsilon^{\dagger}\varepsilon^{\dagger\prime}\right)\hat{F} \\ &= \left(\frac{\hat{F}'F}{T}\right)\left(\frac{\Lambda_0'\Lambda_0}{N}\right)\left(\frac{F'\hat{F}}{T}\right) + d_{NT}, \end{aligned}$$

where $d_{NT} = (T^{-1}\hat{F}'F)(N^{-1}T^{-1}\Lambda_0'\varepsilon^{\dagger\prime}\hat{F}) + (N^{-1}T^{-1}\hat{F}'\varepsilon^{\dagger}\Lambda_0)(T^{-1}F'\hat{F}) + T^{-2}N^{-1}\hat{F}'\varepsilon^{\dagger}\varepsilon^{\dagger\prime}\hat{F}$. (ii) holds by Lemma A.3 in Bai (2003). (i) holds if $d_{NT} = o_P(1)$. Note that under Assumptions A.1-A.3 and A.5, we have

$$\begin{aligned}
\left\|(T^{-1}\hat{F}'F)(N^{-1}T^{-1}\Lambda_0'\varepsilon^{\dagger\prime}\hat{F})\right\| &\leq \left(T^{-1}\left\|\hat{F}'F\right\|\right)N^{-1}T^{-1/2}\left(T^{-1/2}\left\|\hat{F}\right\|\right)(\|\varepsilon\Lambda_0\|+a_{NT}\|\bar{g}\Lambda_0\|) \\
&= O(N^{-1}T^{-1/2})O_P(1)O_P(N^{1/2}T^{1/2}+a_{NT}NT^{1/2}) \\
&= O_P(N^{-1/2}+a_{NT}), \\
\left\|T^{-2}N^{-1}\hat{F}'\varepsilon^{\dagger}\varepsilon^{\dagger\prime}\hat{F}\right\| &\leq 2N^{-1}T^{-1}\left(T^{-1}\left\|\hat{F}\right\|^2\right)\left(R\|\varepsilon\|_{sp}^2+a_{NT}^2\left\|g^{\dagger}\right\|^2\right) \\
&= O_P(T^{-1}+N^{-1}+a_{NT}^2).
\end{aligned}$$

Thus,

$$d_{NT}=O_P(N^{-1/2}+a_{NT}+T^{-1}+N^{-1}+a_{NT}^2)=o_P(1).$$

That completes the proof of Lemma A.1. ■

Lemma A.2 *Suppose Assumption A.1-A.3 and A.5 hold, under $\mathbb{H}_A(a_{NT})$ with $a_{NT}=N^{-1/2}T^{-1/2}$,*
(i) $T^{-1}\hat{F}'F=Q_0+O_P(C_{NT}^{-1})$,
(ii) $H=Q_0^{-1}+O_P(C_{NT}^{-1})$,
where the matrix $Q_0=V_0^{1/2}\Upsilon_0'\Sigma_{\Lambda_0}^{-1/2}$, and Υ_0 being the $R\times R$ corresponding eigenvector matrix of $\Sigma_{\Lambda_0}^{1/2}\Sigma_F\Sigma_{\Lambda_0}^{1/2}$ such that $\Upsilon_0'\Upsilon_0=\mathbb{I}_R$.

Proof. From PCA, we have

$$(NT)^{-1}XX'\hat{F}=\hat{F}V_{NT}$$

Premultiplying both sides of the above equation by $(N^{-1}\Lambda_0'\Lambda_0)^{1/2}(T^{-1}F')$ and plugging $X=F\Lambda_0'+\varepsilon^{\dagger}$, we have

$$\left(\frac{\Lambda_0'\Lambda_0}{N}\right)^{1/2}\left(\frac{F'F}{T}\right)\left(\frac{\Lambda_0'\Lambda_0}{N}\right)\left(\frac{F'\hat{F}}{T}\right)+\bar{d}_{NT}=\left(\frac{\Lambda_0'\Lambda_0}{N}\right)^{1/2}\left(\frac{F'\hat{F}}{T}\right)V_{NT}$$

where

$$\bar{d}_{NT}=\left(\frac{\Lambda_0'\Lambda_0}{N}\right)^{1/2}\left(\frac{F'F}{T}\right)\left(\frac{\Lambda_0'\varepsilon^{\dagger\prime}\hat{F}}{NT}\right)+\left(\frac{\Lambda_0'\Lambda_0}{N}\right)^{1/2}\left(\frac{F'\varepsilon^{\dagger}\Lambda_0}{NT}\right)\left(\frac{F'\hat{F}}{T}\right)+\left(\frac{\Lambda_0'\Lambda_0}{N}\right)^{1/2}\left(\frac{F'\varepsilon^{\dagger}\varepsilon^{\dagger\prime}\hat{F}}{NT^2}\right).$$

By Assumption A.2(ii), we have $N^{-1}\Lambda_0'\Lambda_0=\Sigma_{\Lambda_0}+O_P(N^{-1/2})$. In addition, by the error analysis in Riemann sum approximation of a definite integral and Assumption A.1(i), we have $T^{-1}F'F=\Sigma_F+O_P(T^{-1/2})$. According to the proof of Lemma A.1, we know

$$\begin{aligned}
\left\|\frac{\Lambda_0'\varepsilon^{\dagger\prime}\hat{F}}{NT}\right\| &= O_P(N^{-1/2}+a_{NT}), \\
\left\|\frac{F'\varepsilon^{\dagger}\varepsilon^{\dagger\prime}\hat{F}}{NT^2}\right\| &= O_P(T^{-1}+N^{-1}+a_{NT}^2).
\end{aligned}$$

Hence, we have $\bar{d}_{NT}=O_P(N^{-1/2})$. Let

$$B_{NT}=\left(\frac{\Lambda_0'\Lambda_0}{N}\right)^{1/2}\left(\frac{F'F}{T}\right)\left(\frac{\Lambda_0'\Lambda_0}{N}\right)^{1/2},$$

and

$$R_{NT} = \left(\frac{\Lambda_0'\Lambda_0}{N}\right)^{1/2}\left(\frac{F'\hat{F}}{T}\right),$$

it follows that

$$\begin{aligned}
\left[\left(\frac{\Lambda_0'\Lambda_0}{N}\right)^{1/2}\left(\frac{F'F}{T}\right)\left(\frac{\Lambda_0'\Lambda_0}{N}\right)^{1/2}\right]\left[\left(\frac{\Lambda_0'\Lambda_0}{N}\right)^{1/2}\left(\frac{F'\hat{F}}{T}\right)\right] + \bar{d}_{NT} &= \left[\left(\frac{\Lambda_0'\Lambda_0}{N}\right)^{1/2}\left(\frac{F'\hat{F}}{T}\right)\right]V_{NT} \\
\left[B_{NT} + \bar{d}_{NT}R_{NT}^{-1}\right]R_{NT} &= R_{NT}V_{NT}.
\end{aligned}$$

Thus each column of R_{NT}, though not of length 1, is an eigenvector of the matrix $\left[B_{NT} + \bar{d}_{NT}R_{NT}^{-1}\right]$. Let $\tilde{V}_{NT}$ be a diagonal matrix consisting of the diagonal elements of $R_{NT}'R_{NT}$. Denote $\Upsilon_{NT} = R_{NT}\tilde{V}_{NT}^{-1/2}$ so that each column of Υ_{NT} has a unit length, and it follows

$$\left[B_{NT} + \bar{d}_{NT}R_{NT}^{-1}\right]\Upsilon_{NT} = \Upsilon_{NT}V_{NT}.$$

Thus, V_{NT} contains the eigenvalues of $B_{NT} + \bar{d}_{NT}R_{NT}^{-1}$ with the corresponding normalized eigenvectors contained in Υ_{NT}. By analogous arguments in the proof of Proposition 1 of Bai (2003), we have

$$\left\|B_{NT} + \bar{d}_{NT}R_{NT}^{-1} - B_0\right\| = O_P(C_{NT}^{-1}),$$

where $B_0 = \lim_{N,T\to\infty} B_{NT} = \Sigma_{\Lambda_0}^{1/2}\Sigma_F\Sigma_{\Lambda_0}^{1/2}$. By the perturbation theory for eigenvalues of Hermitian matrices (e.g., Steward and Sun (1990, Ch. V)), we have

$$\left|\mu_j(B_{NT} + \bar{d}_{NT}R_{NT}^{-1}) - \mu_j(B_0)\right| \le \left\|B_{NT} + \bar{d}_{NT}R_{NT}^{-1} - B_0\right\| = O_P(C_{NT}^{-1}),$$

where $\mu_j(A)$ denotes the jth largest eigenvalue of a symmetric matrix A and $j = 1, 2, \cdots, R$. That is,

$$\|V_{NT} - V_0\| = O_P(C_{NT}^{-1}).$$

In addition, by the eigenvector perturbation theory that requires distinctness of eigenvalues, we have

$$\|\Upsilon_{NT} - \Upsilon_0\| = O_P(C_{NT}^{-1}),$$

where Υ_0 is the unique eigenvector matrix of B_0. By the definition of R_{NT},

$$\begin{aligned}
R_{NT} &= \Upsilon_{NT}\tilde{V}_{NT}^{1/2} \\
\left(\frac{F'\hat{F}}{T}\right) &= \left(\frac{\Lambda_0'\Lambda_0}{N}\right)^{-1/2}\Upsilon_{NT}\tilde{V}_{NT}^{1/2} \\
&= \Sigma_{\Lambda_0}^{-1/2}\Upsilon_0 V_0^{1/2} + O_P(C_{NT}^{-1}) = Q_0' + O_P(C_{NT}^{-1}).
\end{aligned}$$

It follows that

$$H = \left(\frac{\Lambda_0'\Lambda_0}{N}\right)\left(\frac{F'\hat{F}}{T}\right)V_{NT}^{-1} = \Sigma_{\Lambda_0}^{1/2}\Upsilon_0 V_0^{-1/2} + O_P(C_{NT}^{-1}) \equiv H_0 + O_P(C_{NT}^{-1}),$$

where $H_0 = Q_0^{-1}$. That completes the proof of Lemma A.2. ■

Lemma A.3 *Suppose Assumption A.1-A.3 and A.5 hold. Under $\mathbb{H}_A(a_{NT})$ with $a_{NT} = N^{-1/2}T^{-1/2}$,*

$$\frac{1}{T}\sum_{t=1}^{T}\|\hat{F}_t - H'F_t\|^2 = O_P(C_{NT}^{-2}).$$

Proof. According to PCA, we have

$$\begin{aligned}
\frac{1}{NT}(XX')\hat{F} &= \hat{F}V_{NT} \\
\hat{F} &= \frac{1}{NT}\left(F\Lambda_0'\Lambda_0F' + F\Lambda_0'\varepsilon^{\dagger\prime} + \varepsilon^{\dagger}\Lambda_0F' + \varepsilon^{\dagger}\varepsilon^{\dagger\prime}\right)\hat{F}V_{NT}^{-1} \\
\hat{F} - FH &= \frac{1}{NT}\left(F\Lambda_0'\varepsilon^{\dagger\prime} + \varepsilon^{\dagger}\Lambda_0F' + \varepsilon^{\dagger}\varepsilon^{\dagger\prime}\right)\hat{F}V_{NT}^{-1},
\end{aligned}$$

where $\varepsilon_{it}^{\dagger} = a_{NT}g_{it}'F_t + \varepsilon_{it} \equiv a_{NT}g_{it}^{\dagger} + \varepsilon_{it}$ and

$$H = \left(\frac{\Lambda_0'\Lambda_0}{N}\right)\left(\frac{F'\hat{F}}{T}\right)V_{NT}^{-1}.$$

It follows

$$\begin{aligned}
\hat{F}_t - H'F_t &= V_{NT}^{-1}\frac{1}{NT}\left(\sum_{s=1}^{T}\hat{F}_s\varepsilon_s^{\dagger\prime}\Lambda_0F_t + \sum_{s=1}^{T}\hat{F}_sF_s'\Lambda_0'\varepsilon_t^{\dagger} + \sum_{s=1}^{T}\hat{F}_s\varepsilon_s^{\dagger\prime}\varepsilon_t^{\dagger}\right) \\
&= V_{NT}^{-1}\left(\frac{1}{T}\sum_{s=1}^{T}\hat{F}_s\gamma_N^{\dagger}(s,t) + \frac{1}{T}\sum_{s=1}^{T}\hat{F}_s\zeta_{st}^{\dagger} + \frac{1}{T}\sum_{s=1}^{T}\hat{F}_s\eta_{st}^{\dagger} + \frac{1}{T}\sum_{s=1}^{T}\hat{F}_s\xi_{st}^{\dagger}\right) \\
&\equiv V_{NT}^{-1}(a_t + b_t + c_t + d_t),
\end{aligned}$$

where $\zeta_{st}^{\dagger} = \frac{\varepsilon_s^{\dagger\prime}\varepsilon_t^{\dagger}}{N} - \gamma_N^{\dagger}(s,t)$, $\gamma_N^{\dagger}(s,t) = E(\frac{\varepsilon_s^{\dagger\prime}\varepsilon_t^{\dagger}}{N})$, $\eta_{st}^{\dagger} = F_s'\Lambda_0'\varepsilon_t^{\dagger}/N$, and $\xi_{st}^{\dagger} = F_t'\Lambda_0'\varepsilon_s^{\dagger}/N$. Since $(x+y+z+w)^2 \leq 4(x^2+y^2+z^2+w^2)$, it follows

$$\frac{1}{T}\sum_{t=1}^{T}\|\hat{F}_t - H'F_t\|^2 \leq \frac{4}{T}\|V_{NT}^{-1}\|^2\sum_{t=1}^{T}[\|a_t\|^2 + \|b_t\|^2 + \|c_t\|^2 + \|d_t\|^2]$$

Given Lemma A.1, $V_{NT} = O_P(1)$, it suffices to find the orders of $\frac{1}{T}\sum_{t=1}^{T}\|a_t\|^2$, $\frac{1}{T}\sum_{t=1}^{T}\|b_t\|^2$, $\frac{1}{T}\sum_{t=1}^{T}\|c_t\|^2$, and $\frac{1}{T}\sum_{t=1}^{T}\|d_t\|^2$. Since

$$\begin{aligned}
\varepsilon_s^{\dagger\prime}\varepsilon_t^{\dagger} &= (\varepsilon_s + a_{NT}g_s^{\dagger})'(\varepsilon_t + a_{NT}g_t^{\dagger}) \\
&= \varepsilon_s'\varepsilon_t + a_{NT}\varepsilon_s g_t^{\dagger} + a_{NT}g_s^{\dagger\prime}\varepsilon_t + a_{NT}^2 g_s^{\dagger\prime}g_t^{\dagger},
\end{aligned}$$

by Cauchy-Schwarz (CS hereafter) inequality,

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}[\gamma_N^{\dagger}(s,t)]^2 \leq \frac{4}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}[E(\varepsilon_s'\varepsilon_t/N)]^2 + \frac{8a_{NT}^2}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\left[E(\varepsilon_s'g_t^{\dagger}/N)\right]^2 + \frac{4a_{NT}^4}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\left[E(g_s^{\dagger\prime}g_t^{\dagger}/N)\right]^2.$$

By Assumption A.3, the first term is bounded above by $4\max_{s,t}\gamma_N(s,t)\max_t\sum_{t=1}^{T}|\gamma_N(s,t)| = O(1)$. By Davydov inequality and Assumptions A.1(ii), A.3(i), (iii), and A.5(i)

$$\begin{aligned}
\frac{a_{NT}^2}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\left[E(\varepsilon_s'g_t^{\dagger}/N)\right]^2 &= \frac{a_{NT}^2}{N^2T}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}E(\varepsilon_{is}F_t'g_{it})E(\varepsilon_{js}F_t'g_{jt}) \\
&\leq C\max_{i,s}E(\varepsilon_{is}F_t'g_{it})\|\varepsilon_{is}\|_{2+\delta}\|F_t\|_{2+\delta}\frac{8a_{NT}^2}{N^2T}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_j(|t-s|)^{\frac{\delta}{2+\delta}}
\end{aligned}$$

$$
\begin{aligned}
&\leq C \max_{i,s} E(\varepsilon_{is} F_t' g_{it}) \|\varepsilon_{is}\|_{2+\delta} \|F_t\|_{2+\delta} \frac{8a_{NT}^2}{N} \sum_{j=1}^{N} \sum_{s=1}^{\infty} \alpha_j(s)^{\frac{\delta}{2+\delta}} \\
&= O(a_{NT}^2) = o(1).
\end{aligned}
$$

By Assumption A.1(ii) and A.5(i),

$$
\frac{4a_{NT}^4}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \left[E(g_s^{\dagger\prime} g_t^{\dagger}/N) \right]^2 = O(a_{NT}^4 T) = o(1).
$$

Thus, $\frac{1}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} [\gamma_N^{\dagger}(s,t)]^2 \leq \frac{4}{T} \sum_{t=1}^{T} \sum_{s=1}^{T} \left[E(\varepsilon_s' \varepsilon_t / N) \right]^2 + o(1) = O(1)$. By the sub-multiplicative property of the Frobenius norm, CS inequality, and the fact that $T^{-1} \hat{F}' \hat{F} = \mathbb{I}_R$, it follows

$$
\begin{aligned}
\frac{1}{T} \sum_{t=1}^{T} \|a_t\|^2 &= \frac{1}{T} \sum_{t=1}^{T} \left\| \frac{1}{T} \sum_{s=1}^{T} \hat{F}_s \gamma_N(s,t) \right\|^2 \\
&\leq \frac{1}{T} \sum_{t=1}^{T} \left\{ \frac{1}{T} \sum_{s=1}^{T} \left\| \hat{F}_s \gamma_N(s,t) \right\| \right\}^2 \\
&\leq \frac{1}{T} \sum_{t=1}^{T} \left\| \hat{F}_t \right\|^2 \frac{1}{T^2} \sum_{s=1}^{T} \sum_{t=1}^{T} [\gamma_N^{\dagger}(s,t)]^2 \\
&= O_P(1) O(T^{-1}) = O_P(T^{-1}).
\end{aligned}
$$

The second term

$$
\begin{aligned}
\frac{1}{T} \sum_{t=1}^{T} \|b_t\|^2 &= \frac{1}{T} \sum_{t=1}^{T} \left\| \frac{1}{T} \sum_{s=1}^{T} \hat{F}_s \zeta_{st}^{\dagger} \right\|^2 \\
&= \frac{1}{T^3} \sum_{t=1}^{T} \sum_{s=1}^{T} \sum_{r=1}^{T} \hat{F}_s' \hat{F}_r \zeta_{st}^{\dagger} \zeta_{rt}^{\dagger} \\
&\leq \frac{1}{T} \left[\frac{1}{T^2} \sum_{s=1}^{T} \sum_{r=1}^{T} \left(\hat{F}_s' \hat{F}_r \right)^2 \right]^{1/2} \left[\frac{1}{T^2} \sum_{s=1}^{T} \sum_{r=1}^{T} \left(\sum_{t=1}^{T} \zeta_{st}^{\dagger} \zeta_{rt}^{\dagger} \right)^2 \right]^{1/2} \\
&\leq \frac{1}{T} \left[\frac{1}{T} \sum_{s=1}^{T} \left\| \hat{F}_s \right\|^2 \right] \left[\frac{1}{T^2} \sum_{s=1}^{T} \sum_{r=1}^{T} \left(\sum_{t=1}^{T} \zeta_{st}^{\dagger} \zeta_{rt}^{\dagger} \right)^2 \right]^{1/2}.
\end{aligned}
$$

Given $\zeta_{st}^{\dagger} = \frac{\varepsilon_s^{\dagger\prime} \varepsilon_t^{\dagger}}{N} - \gamma_N^{\dagger}(s,t)$, we can show that $\frac{1}{T^2} \sum_{s=1}^{T} \sum_{r=1}^{T} E \left(\sum_{t=1}^{T} \zeta_{st}^{\dagger} \zeta_{rt}^{\dagger} \right)^2 = O(T^2 N^{-2})$ under Assumptions A.3 and A.5. Since $\frac{1}{T} \sum_{s=1}^{T} \left\| \hat{F}_s \right\|^2 = O_P(1)$, it follows that $\frac{1}{T} \sum_{t=1}^{T} \|b_t\|^2 = O_P(N^{-1})$.

For the third term,

$$
\begin{aligned}
\frac{1}{T} \sum_{t=1}^{T} \|c_t\|^2 &= \frac{1}{T} \sum_{t=1}^{T} \left\| \frac{1}{T} \sum_{s=1}^{T} \hat{F}_s \eta_{st}^{\dagger} \right\|^2 \\
&= \frac{1}{T} \sum_{t=1}^{T} \left\| \frac{1}{T} \sum_{s=1}^{T} \hat{F}_s F_s' \Lambda_0' \varepsilon_t^{\dagger} / N \right\|^2
\end{aligned}
$$

$$
\begin{aligned}
&\leq \frac{1}{T}\sum_{t=1}^{T}\left\|\Lambda_0'\varepsilon_t^{\dagger}/N\right\|^2\left[\frac{1}{T}\sum_{s=1}^{T}\left\|\hat{F}_s\right\|^2\right]\left[\frac{1}{T}\sum_{s=1}^{T}\|F_s\|^2\right] \\
&= \left[\frac{1}{T}\sum_{t=1}^{T}\|\Lambda_0'\varepsilon_t/N\|^2+\frac{a_{NT}^2}{T}\sum_{t=1}^{T}\left\|\Lambda_0'g_t^{\dagger}/N\right\|^2\right]\left[\frac{1}{T}\sum_{s=1}^{T}\left\|\hat{F}_s\right\|^2\right]\left[\frac{1}{T}\sum_{s=1}^{T}\|F_s\|^2\right].
\end{aligned}
$$

Given $\|\Lambda_0'\varepsilon_t/\sqrt{N}\|^2 = O_P(1)$, we have $\frac{1}{T}\sum_{t=1}^{T}\|\Lambda_0'\varepsilon_t/N\|^2 = O_P(N^{-1})$. And

$$
\begin{aligned}
\frac{a_{NT}^2}{T}\sum_{t=1}^{T}\left\|\Lambda_0'g_t^{\dagger}/N\right\|^2 &= \frac{a_{NT}^2}{T}\sum_{t=1}^{T}\left\|\frac{1}{N}\sum_{i=1}^{N}\lambda_{i0}F_t'g_{it}\right\|^2 \\
&\leq \frac{a_{NT}^2}{T}\sum_{t=1}^{T}\left\|\frac{1}{N}\sum_{i=1}^{N}F_t'g_{it}\right\|^2\left\|\frac{1}{N}\sum_{i=1}^{N}\lambda_{i0}\right\|^2 \\
&= O(a_{NT}^2)O_P(1)O(1) = O_P(a_{NT}^2),
\end{aligned}
$$

where the last equality is implied by Assumption A.2(i) and A.5(i). Thus, it follows that $\frac{1}{T}\sum_{t=1}^{T}\|c_t\|^2 = O_P(N^{-1})$. For the fourth term

$$
\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}\|d_t\|^2 &= \frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{T}\sum_{s=1}^{T}\hat{F}_s\xi_{st}^{\dagger}\right\|^2 \\
&= \frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{T}\sum_{s=1}^{T}\hat{F}_sF_t'\Lambda_0'\varepsilon_s^{\dagger}/N\right\|^2 \\
&\leq \frac{1}{T^3}\sum_{t=1}^{T}\left[\sum_{s=1}^{T}\left\|\hat{F}_sF_t'\right\|^2\sum_{s=1}^{T}\|\Lambda_0'\varepsilon_s^{\dagger}/N\|^2\right] \\
&= \left[\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T}\left\|\hat{F}_sF_t'\right\|^2\right]\left[\frac{1}{T}\sum_{s=1}^{T}\|\Lambda_0'\varepsilon_s^{\dagger}/N\|^2\right] \\
&\leq \left[\frac{1}{T}\sum_{s=1}^{T}\left\|\hat{F}_s\right\|^2\right]\left[\frac{1}{T}\sum_{t=1}^{T}\|F_t\|^2\right]\left[\frac{1}{T}\sum_{t=1}^{T}\|\Lambda_0'\varepsilon_t/N\|^2+\frac{a_{NT}^2}{T}\sum_{t=1}^{T}\left\|\Lambda_0'g_t^{\dagger}/N\right\|^2\right] \\
&= O_P(1)O_P(1)\left[O_P(N^{-1})+O_P(a_{NT}^2)\right],
\end{aligned}
$$

by analogous proof in $\frac{1}{T}\sum_{t=1}^{T}\|c_t\|^2$. Thus, $\frac{1}{T}\sum_{t=1}^{T}\|d_t\|^2 = O_P(N^{-1})$. Combining these results, we have $\frac{1}{T}\sum_{t=1}^{T}\|\hat{F}_t - H'F_t\|^2 = O_P(C_{NT}^{-2})$. ■

Lemma A.4 *Suppose Assumption A.1-A.3 and A.5 hold, under $\mathbb{H}_A(a_{NT})$,*

$$
\hat{F}_t - H'F_t = O_P(C_{NT}^{-2}+a_{NT}) = O_P(C_{NT}^{-2}).
$$

Proof. Following Bai (2003), we have

$$
\begin{aligned}
&\hat{F}_t - H'F_t \\
&= V_{NT}^{-1}\frac{1}{NT}\left(\sum_{s=1}^{T}\hat{F}_s\varepsilon_s^{\dagger\prime}\Lambda_0F_t+\sum_{s=1}^{T}\hat{F}_sF_s'\Lambda_0'\varepsilon_t^{\dagger}+\sum_{s=1}^{T}\hat{F}_s\varepsilon_s^{\dagger\prime}\varepsilon_t^{\dagger}\right) \\
&= V_{NT}^{-1}\Bigg[\frac{1}{T}\sum_{s=1}^{T}\hat{F}_s\gamma_N(s,t)+\frac{1}{T}\sum_{s=1}^{T}\hat{F}_s\zeta_{st}+\frac{1}{T}\sum_{s=1}^{T}\hat{F}_s\eta_{st}+\frac{1}{T}\sum_{s=1}^{T}\hat{F}_s\xi_{st}+\frac{a_{NT}}{T}\sum_{s=1}^{T}\hat{F}_s\left(\frac{g_s^{\dagger\prime}\Lambda_0F_t}{N}\right)
\end{aligned}
$$

$$+\frac{a_{NT}}{T}\sum_{s=1}^{T}\hat{F}_s\left(\frac{g_t^{\dagger\prime}\Lambda_0F_s}{N}\right)+\frac{a_{NT}^2}{T}\sum_{s=1}^{T}\hat{F}_s\left(\frac{g_s^{\dagger\prime}g_t^{\dagger}}{N}\right)+\frac{a_{NT}}{T}\sum_{s=1}^{T}\hat{F}_s\left(\frac{\varepsilon_s'g_t^{\dagger}}{N}\right)+\frac{a_{NT}}{T}\sum_{s=1}^{T}\hat{F}_s\left(\frac{\varepsilon_t'g_s^{\dagger}}{N}\right)\Bigg],$$

Now, we prove that
(a). $\frac{1}{T}\sum_{s=1}^{T}\hat{F}_s\gamma_N(s,t)=O_P(C_{NT}^{-1}T^{-1/2})$;
(b). $\frac{1}{T}\sum_{s=1}^{T}\hat{F}_s\zeta_{st}=O_P(C_{NT}^{-1}N^{-1/2})$;
(c). $\frac{1}{T}\sum_{s=1}^{T}\hat{F}_s\eta_{st}=O_P(N^{-1/2})$;
(d). $\frac{1}{T}\sum_{s=1}^{T}\hat{F}_s\xi_{st}=O_P(C_{NT}^{-1}N^{-1/2})$;
(e). $\frac{a_{NT}}{T}\sum_{s=1}^{T}\hat{F}_s(g_s^{\dagger\prime}\Lambda_0F_t/N)=O_P(a_{NT})$;
(f). $\frac{a_{NT}}{T}\sum_{s=1}^{T}\hat{F}_s(g_t^{\dagger\prime}\Lambda_0F_s/N)=O_P(a_{NT})$;
(g). $\frac{a_{NT}^2}{T}\sum_{s=1}^{T}\hat{F}_s(g_s^{\dagger\prime}g_t^{\dagger}/N)=O_P(a_{NT}^2)$;
(h). $\frac{a_{NT}}{T}\sum_{s=1}^{T}\hat{F}_s(\varepsilon_s'g_t^{\dagger}/N)=O_P(a_{NT}C_{NT}^{-1}N^{-1/2})$;
(i). $\frac{a_{NT}}{T}\sum_{s=1}^{T}\hat{F}_s(\varepsilon_t'g_s^{\dagger}/N)=O_P(a_{NT}N^{-1/2})$,
where $\zeta_{st}=\frac{\varepsilon_s'\varepsilon_t}{N}-\gamma_N(s,t)$, $\gamma_N(s,t)=E(\frac{\varepsilon_s'\varepsilon_t}{N})$, $\eta_{st}=F_s'\Lambda_0'\varepsilon_t/N$, and $\xi_{st}=F_t'\Lambda_0'\varepsilon_s/N$.

Given Lemma A.1(i), under Assumptions A.1-A.3, we can show (a)-(d) following analogous proof of Lemma A.2 in Bai (2003). Consider part (e),

$$\frac{a_{NT}}{T}\sum_{s=1}^{T}\hat{F}_s(g_s^{\dagger\prime}\Lambda_0F_t/N) = \frac{a_{NT}}{T}\sum_{s=1}^{T}(\hat{F}_s-H'F_s)(g_s^{\dagger\prime}\Lambda_0F_t/N)+\frac{a_{NT}}{T}\sum_{s=1}^{T}H'F_s(g_s^{\dagger\prime}\Lambda_0F_t/N).$$

The first term of (e)

$$\begin{aligned}\left\|\frac{a_{NT}}{T}\sum_{s=1}^{T}(\hat{F}_s-H'F_s)(g_s^{\dagger\prime}\Lambda_0F_t/N)\right\| &\le a_{NT}\left(\frac{1}{T}\sum_{s=1}^{T}\left\|\hat{F}_s-H'F_s\right\|^2\right)^{1/2}\left(\frac{1}{T}\sum_{s=1}^{T}\left\|\frac{g_s^{\dagger\prime}\Lambda_0}{N}\right\|^2\right)^{1/2}\|F_t\|\\ &= O(a_{NT})O_P(C_{NT}^{-1})O_P(1)O_P(1)\\ &= O_P(a_{NT}C_{NT}^{-1}).\end{aligned}$$

For the second term of (e),

$$\frac{a_{NT}}{T}\sum_{s=1}^{T}H'F_s(g_s^{\dagger\prime}\Lambda_0F_t/N)\le a_{NT}\|H\|\left(\frac{1}{T}\sum_{s=1}^{T}\|F_s\|^2\right)^{1/2}\left(\frac{1}{T}\sum_{s=1}^{T}\left\|\frac{g_s^{\dagger\prime}\Lambda_0}{N}\right\|^2\right)^{1/2}\|F_t\|=O_P(a_{NT}).$$

Thus, (e) is $O_P(a_{NT})$. For (f), we have

$$\frac{a_{NT}}{T}\sum_{s=1}^{T}\hat{F}_s(g_t^{\dagger\prime}\Lambda_0F_s/N) = \frac{a_{NT}}{T}\sum_{s=1}^{T}(\hat{F}_s-H'F_s)(g_t^{\dagger\prime}\Lambda_0F_s/N)+\frac{a_{NT}}{T}\sum_{s=1}^{T}H'F_s(g_t^{\dagger\prime}\Lambda_0F_s/N).$$

For the second term of (f), it holds that

$$a_{NT}\left(\frac{1}{T}\sum_{s=1}^{T}F_sF_s'\right)(\Lambda_0'g_t^{\dagger}/N) = a_{NT}\left(\frac{1}{T}\sum_{s=1}^{T}F_sF_s'\right)\left(\frac{1}{N}\sum_{i=1}^{N}\lambda_{i0}g_{it}'F_t\right)=O_P(a_{NT})O_P(1)O_P(1).$$

The first term has

$$\left\|\frac{a_{NT}}{T}\sum_{s=1}^{T}(\hat{F}_s-H'F_s)(F_s'\Lambda_0'g_t^{\dagger}/N)\right\| \le a_{NT}\left(\frac{1}{T}\sum_{s=1}^{T}\left\|\hat{F}_s-H'F_s\right\|^2\right)^{1/2}\left(\frac{1}{T}\sum_{s=1}^{T}\|F_s\|^2\right)^{1/2}\left\|\frac{g_s^{\dagger\prime}\Lambda_0}{N}\right\|$$

$$= \quad O(a_{NT})O_P(C_{NT}^{-1})O_P(1).$$

Therefore, we have (f) is $O_P(a_{NT})$. Now we consider part (g)

$$\frac{a_{NT}^2}{T}\sum_{s=1}^{T}\hat{F}_s(g_s^{\dagger\prime}g_t^{\dagger}/N) \quad = \quad \frac{a_{NT}^2}{T}\sum_{s=1}^{T}(\hat{F}_s - H'F_s)(g_s^{\dagger\prime}g_t^{\dagger}/N) + \frac{a_{NT}^2}{T}\sum_{s=1}^{T}H'F_s(g_s^{\dagger\prime}g_t^{\dagger}/N).$$

The first term has

$$\begin{aligned} &\left\|\frac{a_{NT}^2}{T}\sum_{s=1}^{T}(\hat{F}_s - H'F_s)(g_s^{\dagger\prime}g_t^{\dagger}/N)\right\| \\ \leq\ & a_{NT}^2\left(\frac{1}{T}\sum_{s=1}^{T}\left\|\hat{F}_s - H'F_s\right\|^2\right)^{1/2}\left(\frac{1}{T}\sum_{s=1}^{T}\left\|g_s^{\dagger\prime}g_t^{\dagger}/N\right\|^2\right)^{1/2}. \end{aligned}$$

Since $\|g_s^{\dagger\prime}g_t^{\dagger}/N\|^2 = \|\frac{1}{N}\sum_{i=1}^{N}g_{is}'F_sg_{it}'F_t\|^2 \leq \left(\frac{1}{N}\sum_{i=1}^{N}\|g_{is}'F_s\|^2\right)\left(\frac{1}{N}\sum_{i=1}^{N}\|g_{it}'F_t\|^2\right) = O_P(1)$ by CS inequality and Assumption A.1, A.2, and A.5. Thus, it follows the first term of (g) is $O_p(a_{NT}^2C_{NT}^{-1})$. By analogous arguments, we can show the second term of (g) is $O_p(a_{NT}^2)$, and thus (g) is $O_p(a_{NT}^2)$.

For part (h), we have

$$\frac{a_{NT}}{T}\sum_{s=1}^{T}\hat{F}_s(\varepsilon_s'g_t^{\dagger}/N) \quad = \quad \frac{a_{NT}}{T}\sum_{s=1}^{T}(\hat{F}_s - H'F_s)(\varepsilon_s'g_t^{\dagger}/N) + \frac{a_{NT}}{T}\sum_{s=1}^{T}H'F_s(\varepsilon_s'g_t^{\dagger}/N).$$

Consider the first term

$$\left\|\frac{a_{NT}}{T}\sum_{s=1}^{T}(\hat{F}_s - H'F_s)(\varepsilon_s'g_t^{\dagger}/N)\right\| \leq a_{NT}\left(\frac{1}{T}\sum_{s=1}^{T}\left\|\hat{F}_s - H'F_s\right\|^2\right)^{1/2}\left(\frac{1}{T}\sum_{s=1}^{T}\left\|\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{is}g_{it}'F_t\right\|^2\right)^{1/2}.$$

We can show that for each (s,t), $\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{is}g_{it}'F_t = O_p(1/\sqrt{N})$, since $E(\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{is}g_{it}'F_t) = 0$ and $E(\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{is}g_{it}'F_t)^2 = O(N^{-1})$ by Assumption A.3 (iv). Thus, the first term of (h) is $O_P(a_{NT}C_{NT}^{-1}N^{-1/2})$. For the second term of (h), we have

$$\begin{aligned} \frac{a_{NT}}{T}\sum_{s=1}^{T}H'F_s(\varepsilon_s'g_t^{\dagger}/N) &= a_{NT}H'\left(\frac{1}{NT}\sum_{s=1}^{T}\sum_{i=1}^{N}F_s\varepsilon_{is}g_{it}'\right)F_t \\ &= O_P(a_{NT}N^{-1/2}T^{-1/2}). \end{aligned}$$

Therefore, we have (h) is $O_P(a_{NT}C_{NT}^{-1}N^{-1/2})$. Lastly, we consider part (i)

$$\frac{a_{NT}}{T}\sum_{s=1}^{T}\hat{F}_s(\varepsilon_t'g_s^{\dagger}/N) \quad = \quad \frac{a_{NT}}{T}\sum_{s=1}^{T}(\hat{F}_s - H'F_s)(\varepsilon_t'g_s^{\dagger}/N) + \frac{a_{NT}}{T}\sum_{s=1}^{T}H'F_s(\varepsilon_t'g_s^{\dagger}/N).$$

Since we have shown that $\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{is}g_{it}'F_t = O_p(1/\sqrt{N})$ for each (s,t)

$$\begin{aligned} \left\|\frac{a_{NT}}{T}\sum_{s=1}^{T}(\hat{F}_s - H'F_s)(\varepsilon_t'g_s^{\dagger}/N)\right\| &\leq a_{NT}\left(\frac{1}{T}\sum_{s=1}^{T}\left\|\hat{F}_s - H'F_s\right\|^2\right)^{1/2}\left(\frac{1}{T}\sum_{s=1}^{T}\left\|\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{it}g_{is}'F_s\right\|^2\right)^{1/2} \\ &= O_P(a_{NT}C_{NT}^{-1}N^{-1/2}). \end{aligned}$$

For the second term of (i), we have

$$\begin{aligned}\frac{a_{NT}}{T}\sum_{s=1}^{T}H'F_s(\varepsilon_t' g_s^{\dagger}/N) &= \frac{a_{NT}}{T}\sum_{s=1}^{T}H'F_s\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{it}g_{is}'F_s \\ &= a_{NT}H'\left(\frac{1}{T}\sum_{s=1}^{T}F_sF_s'\right)\left(\frac{1}{N}\sum_{i=1}^{N}g_{is}\varepsilon_{is}\right) \\ &= O_P(a_{NT}N^{-1/2}).\end{aligned}$$

Thus, we have (i) is $O_P(a_{NT}N^{-1/2})$. ■

Lemma A.5 *Suppose Assumptions A.1 to A.3 and A.5 hold,*

$$\frac{1}{T}\sum_{t=1}^{T}\left(\hat{F}_t\hat{F}_t' - \mathbb{I}_R\right)e^{\mathbf{i}u2\pi t/T}$$

and

$$\frac{1}{T}\sum_{t=1}^{T}\left(\hat{F}_tF_t' - Q_0\right)e^{\mathbf{i}u2\pi t/T}$$

are stochastically equicontinuous, where $Q_0 = V_0^{1/2}\Upsilon_0'\Sigma_{\Lambda_0}^{-1/2}$ *is as defined in Lemma A.2.*

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

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

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

$$\begin{aligned}&\lim_{T,N\to\infty}P\left[\sup_{u_1,u_2\in\mathbb{R}:|u_1-u_2|<\delta}\left\|\frac{1}{T}\sum_{t=1}^{T}\left(\hat{F}_t\hat{F}_t' - \mathbb{I}_R\right)\left(e^{\mathbf{i}u_12\pi t/T} - e^{\mathbf{i}u_22\pi t/T}\right)\right\| > \kappa\right] \\ =\;& \lim_{T,N\to\infty}P\left[\sup_{u_1,u_2\in\mathbb{R}:|u_1-u_2|<\delta}\left\|\frac{1}{T}\sum_{t=1}^{T}\left(\hat{F}_t\hat{F}_t' - \mathbb{I}_R\right)\left(\mathbf{i}2\pi t/Te^{\mathbf{i}\bar{u}2\pi t/T}\right)(u_1 - u_2)\right\| > \kappa\right] \\ \leq\;& \lim_{T,N\to\infty}P\left[\sup_{\bar{u}\in\mathbb{R}}\left\|\frac{1}{T}\sum_{t=1}^{T}\left(\hat{F}_t\hat{F}_t' - \mathbb{I}_R\right)\left(\mathbf{i}2\pi t/Te^{\mathbf{i}\bar{u}2\pi t/T}\right)\right\| > \kappa/\delta\right] \\ \leq\;& \lim_{T,N\to\infty}P\left[\sup_{\bar{u}\in\mathbb{R}}\frac{1}{T}\sum_{t=1}^{T}\left\|\left(\hat{F}_t\hat{F}_t' - \mathbb{I}_R\right)\left(\mathbf{i}2\pi t/Te^{\mathbf{i}\bar{u}2\pi t/T}\right)\right\| > \kappa/\delta\right] \\ \leq\;& \lim_{T,N\to\infty}P\left[\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left\|\left(\hat{F}_t\hat{F}_t' - \mathbb{I}_R\right)\right\|^2}\sup_{\bar{u}\in\mathbb{R}}\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left\|\left(\mathbf{i}2\pi t/Te^{\mathbf{i}\bar{u}2\pi t/T}\right)\right\|^2} > \kappa/\delta\right]\end{aligned}$$

$$
\begin{aligned}
&= \lim_{T,N\to\infty} P\left[\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left\|\left(\hat{F}_t\hat{F}_t'-\mathbb{I}_R\right)\right\|^2}\sqrt{\frac{1}{T}\sum_{t=1}^{T}4\pi t^2/T^2}>\kappa/\delta\right] \\
&< \epsilon,
\end{aligned}
$$

where the third to last inequality is by triangle inequality and the second to last is by CS 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 $O_P(1)$, and $\frac{1}{T}\sum_{t=1}^{T}4\pi t^2/T^2$ is $O(1)$. By analogous argument, we can show $\frac{1}{T}\sum_{t=1}^{T}\left(\hat{F}_tF_t'-Q_0\right)e^{\mathbf{i}u2\pi t/T}$ is also stochastically equicontinuous, where $Q_0 = \text{plim}\,\frac{1}{T}\sum_{t=1}^{T}\hat{F}_tF_t'$. Note that these results hold under both H_0 and $H_A(a_{NT})$. ■

Lemma A.6 *Suppose Assumptions A.1 to A.3 and A.5 hold, then as* $(T,N)\to\infty$

$$
\begin{aligned}
&\sup_{u\in\mathbb{R}}\left\|\tfrac{1}{T}\textstyle\sum_{t=1}^{T}\hat{F}_t\hat{F}_t'e^{\mathbf{i}u2\pi t/T}-\mathbb{I}_R\int e^{\mathbf{i}u2\pi\tau}d\tau\right\| \xrightarrow{p} 0, \\
&\sup_{u\in\mathbb{R}}\left\|\tfrac{1}{T}\textstyle\sum_{t=1}^{T}\hat{F}_tF_t'e^{\mathbf{i}u2\pi t/T}-Q_0\int e^{\mathbf{i}u2\pi\tau}d\tau\right\| \xrightarrow{p} 0.
\end{aligned}
$$

Proof.

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

We now show $R_1 = o_P(1)$.

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

Next, we show that $R_{11} = o_P(1)$. 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,

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

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

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). $\frac{1}{T}\sum_{t=1}^{T}\left(\hat{F}_t\hat{F}_t' - \mathbb{I}_R\right)\theta(\tau) = o_P(1)$ for any $\theta \in \mathbb{D}$.

For (i):

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

by analogous arguments in the proof of Lemma A.5. The point-wise convergence (ii) is easy to verify, since under both $\mathbb{H}_0$ and $\mathbb{H}_A(a_{NT})$, by Lemma A.4,

$$
\begin{aligned}
\hat{F}_t\hat{F}_t' &= \left(\hat{F}_t - H'F_t\right)\left(\hat{F}_t - H'F_t\right)' + \left(\hat{F}_t - H'F_t\right)F_t'H + H'F_t\left(\hat{F}_t - H'F_t\right)' + H'F_tF_t'H \\
&= O_P(N^{-1}) + O_P(N^{-1/2}) + O_P(N^{-1/2}) + H'F_tF_t'H.
\end{aligned}
$$

By Lemma A.2(ii), we know $H = Q_0^{-1}$. Thus $E(H'F_tF_t'H) = E(Q_0^{-1\prime}F_tF_t'Q_0^{-1}) + O_P(C_{NT}^{-2}) = Q_0^{-1\prime}\Sigma_F Q_0^{-1} + O_P(C_{NT}^{-2}) = V_0^{-1/2}\Upsilon_0'\Sigma_{\Lambda_0}^{1/2}\Sigma_F\Sigma_{\Lambda_0}^{1/2}\Upsilon_0 V_0^{-1/2} = V_0^{-1/2}V_0V_0^{-1/2} = \mathbb{I}_R$. Together with the boundedness condition of $\mathbb{D}$, we have that $\frac{1}{T}\sum_{t=1}^{T}\left(\hat{F}_t\hat{F}_t' - \mathbb{I}_R\right)\theta(\tau) = o_P(1)$ for any $\theta \in \mathbb{D}$. By analogous proof, we can show that $R_{12} = o_P(1)$. Therefore, $R_1 = o_P(1)$.

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

$$
\sup_{u\in\mathbb{R}}\left\|\frac{1}{T}\sum_{t=1}^{T}\hat{F}_t F_t' e^{\mathbf{i}u2\pi t/T} - Q_0\int e^{\mathbf{i}u2\pi\tau}d\tau\right\| \xrightarrow{p} 0,
$$

where Q_0 is defined in Lemma A.2.

■

Lemma A.7 *Suppose Assumptions A.1 to A.3 and A.5 hold, then as* $(T, N) \to \infty$,

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

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

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

where $H_0 = Q_0^{-1}$, and $M_t(u)$ 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 the definition of $\hat{A}_2(u)$ and Lemma A.6, we have

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

where

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

It follows

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

It is straightforward to show that

$$\begin{aligned}
&\left\| \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(\hat{F}_t - H'F_t)\varepsilon_{it}M_t(u) \right\| \\
\le& \left(\frac{1}{T}\sum_{t=1}^{T}\left\|\hat{F}_t - H'F_t\right\|^2\right)^{1/2}\left(\frac{1}{T}\sum_{t=1}^{T}\left\|\frac{1}{N}\sum_{i=1}^{N}\varepsilon_{it}\right\|^2\|M_t(u)\|^2\right)^{1/2} \\
=& O_P(C_{NT}^{-1})O_P(N^{-1/2}).
\end{aligned}$$

Therefore, we have $\hat{A}_{21}(u) = O_P(C_{NT}^{-2}) = o_P(1)$. It follows that the dominant term in $\hat{A}_2(u)$ is $\hat{A}_{22}(u)$. By Lemma A.2(ii), $H = H_0 + O_P(C_{NT}^{-1})$,

$$\hat{A}_{22}(u) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}H_0'F_t\varepsilon_{it}M_t(u) + o_P(1)$$

Under Assumption A.1 to A.3, we have for each fixed $u \in \mathbb{R}$

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

where

$$\begin{aligned}
\mathcal{M}(u,u) &= \text{avar}[\sqrt{NT}\hat{A}_{22}(u)] \\
&= \lim_{N,T\to\infty} \frac{1}{NT} \text{ var}\left[\sum_{i=1}^{N}\sum_{t=1}^{T} H_0' F_t \varepsilon_{it} M_t(u)\right] \\
&= H_0' \left[\lim_{N,T\to\infty} \frac{1}{NT} \sum_{s,t=1}^{T}\sum_{i,j=1}^{N} E\left[F_t F_s' \varepsilon_{it}\varepsilon_{js}\right] M_t(u) M_s(u)^*\right] H_0,
\end{aligned}$$

where the last equation is obtained by Lemma A.2. Furthermore, given the stochastic equi-continuity results established in Lemma A.5, 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

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

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

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

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 Proofs of the Theorems and Propositions in Section 3

Proof of Proposition 1. Under $\mathbb{H}_0 : \lambda_{it} = \lambda_{i0}$, we have

$$\begin{aligned}
\hat{A}_1(u) &= \frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{T}\sum_{t=1}^{T}\hat{G}_t(u) F_t' \lambda_{i0}\right] \\
&= \left[\frac{1}{T}\sum_{t=1}^{T}\hat{G}_t(u) F_t'\right]\left[\frac{1}{N}\sum_{i=1}^{N}\lambda_{i0}\right] \\
&= \left[\frac{1}{T}\sum_{t=1}^{T}\hat{G}_t(u)(F_t - H'^{-1}\hat{F}_t)'\right]\bar{\lambda}_{N,0} + \left[\frac{1}{T}\sum_{t=1}^{T}\hat{G}_t(u)\hat{F}_t' H^{-1}\right]\bar{\lambda}_{N,0} \\
&\equiv \hat{A}_{11}(u) + \hat{A}_{12}(u),
\end{aligned}$$

where $\bar{\lambda}_{N,0} = \left[\frac{1}{N}\sum_{i=1}^{N}\lambda_{i0}\right]$.

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

$$
\begin{aligned}
\hat{A}_{12}(u) &= \left[\frac{1}{T}\sum_{t=1}^{T}\hat{G}_t(u)\hat{F}_t' H^{-1}\right]\bar{\lambda}_{N,0} \\
&= \left[\frac{1}{T}\sum_{t=1}^{T}\hat{F}_t e^{\mathbf{i}u2\pi t/T}\hat{F}_t' H^{-1} - \left(\frac{1}{T}\sum_{t=1}^{T}\hat{F}_t\hat{F}_t' e^{\mathbf{i}u2\pi t/T}\right)\left(\frac{1}{T}\sum_{t=1}^{T}\hat{F}_t\hat{F}_t'\right) H^{-1}\right]\bar{\lambda}_{N,0} \\
&= 0,
\end{aligned}
$$

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 Lemma A.2 of 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

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

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 dominant terms. Let $T \propto N^{\nu}$, we have the following three cases

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

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

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

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

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

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

When $\nu > 1/2$,

$$\begin{aligned}
\hat{A}_{11}(u) &= \left[\frac{1}{T}\sum_{t=1}^{T}\hat{G}_t(u)\left(-H'^{-1}V_{NT}^{-1}\frac{1}{T}\sum_{s=1}^{T}\hat{F}_s\eta_{st}\right)'\right]\bar{\lambda}_{N,0} + o_P(1) \\
&= \left\{\frac{1}{T}\sum_{t=1}^{T}\hat{G}_t(u)\left[-H'^{-1}V_{NT}^{-1}\left(\frac{1}{T}\sum_{s=1}^{T}\hat{F}_sF_s'\right)\left(\frac{1}{N}\sum_{i=1}^{N}\lambda_{i0}\varepsilon_{it}\right)\right]'\right\}\bar{\lambda}_{N,0} + o_P(1) \\
&= \left\{-\frac{1}{T}\sum_{t=1}^{T}\hat{G}_t(u)\left(\frac{1}{N}\sum_{i=1}^{N}\lambda_{i0}\varepsilon_{it}\right)'\left(\frac{1}{T}\sum_{s=1}^{T}\hat{F}_sF_s'\right)'V_{NT}^{-1}H^{-1}\right\}\bar{\lambda}_{N,0} + o_P(1) \\
&= -\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{G}_t(u)\varepsilon_{it}\lambda_{i0}'\left(\frac{1}{T}\sum_{s=1}^{T}\hat{F}_sF_s'\right)'V_{NT}^{-1}H^{-1}\bar{\lambda}_{N,0} + o_P(1) \\
&= -\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{G}_t(u)\varepsilon_{it}\lambda_{i0}'\left(\Lambda_0'\Lambda_0/N\right)^{-1}\bar{\lambda}_{N,0} + o_P(1).
\end{aligned}$$

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

$$\hat{A}(u) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{G}_t(u)\hat{B}_i\varepsilon_{it} + o_P(1),$$

where

$$\hat{B}_i = 1 - \lambda_{i0}'\left(\Lambda_0'\Lambda_0/N\right)^{-1}\bar{\lambda}_{N,0}.$$

Given

$$\begin{aligned}
\hat{F}_t\hat{F}_t' &= \left(\hat{F}_t - H'F_t\right)\left(\hat{F}_t - H'F_t\right)' + \left(\hat{F}_t - H'F_t\right)F_t'H + H'F_t\left(\hat{F}_t - H'F_t\right)' + H'F_tF_t'H \\
&= O_P(N^{-1}) + O_P(N^{-1/2}) + O_P(N^{-1/2}) + H'F_tF_t'H,
\end{aligned}$$

it follows

$$\frac{1}{T}\sum_{t=1}^{T}\hat{F}_t\hat{F}_t'e^{\mathbf{i}u2\pi t/T} = H'\left(\frac{1}{T}\sum_{t=1}^{T}F_tF_t'e^{\mathbf{i}u2\pi t/T}\right)H + O_P(N^{-1/2}).$$

Therefore, for each fixed $u \in \mathbb{R}$,

$$\begin{aligned}
\hat{G}_t(u) &= \hat{F}_te^{\mathbf{i}u2\pi t/T} - \left(\frac{1}{T}\sum_{t=1}^{T}\hat{F}_t\hat{F}_t'e^{\mathbf{i}u2\pi t/T}\right)\hat{F}_t \\
&= \left(\hat{F}_t - H'F_t\right)e^{\mathbf{i}u2\pi t/T} - H'\left(\frac{1}{T}\sum_{t=1}^{T}F_tF_t'e^{\mathbf{i}u2\pi t/T}\right)H\left(\hat{F}_t - H'F_t\right) + O_P(N^{-1/2})
\end{aligned}$$

$$
\begin{aligned}
&= H'F_t e^{\mathbf{i}u2\pi t/T} - H'\left[\frac{1}{T}\sum_{t=1}^{T}(F_tF_t' - \Sigma_F)e^{\mathbf{i}u2\pi t/T}\right]HH'F_t - H'\Sigma_F HH'F_t\left[\frac{1}{T}\sum_{t=1}^{T}e^{\mathbf{i}u2\pi t/T}\right] + O_P(N^{-1/2}) \\
&= H'F_t e^{\mathbf{i}u2\pi t/T} + O_P(T^{-1/2}) - H'\Sigma_F HH'F_t\left[\frac{1}{T}\sum_{t=1}^{T}e^{\mathbf{i}u2\pi t/T}\right] + O_P(N^{-1/2}) \\
&= H_0'F_t e^{\mathbf{i}u2\pi t/T} - H_0'\Sigma_F H_0 H_0'F_t\left(\int_0^1 e^{\mathbf{i}u2\pi\tau}d\tau\right) + O_P(C_{NT}^{-1}) \\
&= H_0'F_t\left(e^{\mathbf{i}u2\pi t/T} - \int_0^1 e^{\mathbf{i}u2\pi\tau}d\tau\right) + O_P(C_{NT}^{-1}) \\
&\equiv G_t(u) + O_P(C_{NT}^{-1}).
\end{aligned}
$$

where the third to last equality is due to that Lemma A.2 establishes that $H = H_0 + O_P(C_{NT}^{-1})$. Furthermore, it is straightforward to show that $\hat{B}_i = B_i + O_P(N^{-1/2})$,where $B_i = 1 - \lambda_{i0}'\Sigma_{\Lambda_0}\bar{\lambda}_0$ and $\bar{\lambda}_0 = \lim_{N\to\infty}\lambda_{N,0}$. As a result, it follows that

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

When $\nu < 1/2$,

$$
\begin{aligned}
\hat{A}_{11}(u) &= \left[\frac{1}{T}\sum_{t=1}^{T}\hat{G}_t(u)\left(-H'^{-1}V_{NT}^{-1}\frac{1}{T}\sum_{s=1}^{T}\hat{F}_s\gamma_N(s,t)\right)'\right]\bar{\lambda}_{N,0} + o_P(1) \\
&= \left[\frac{1}{T}\sum_{t=1}^{T}\hat{G}_t(u)\left(-H'^{-1}V_{NT}^{-1}\frac{1}{T}\sum_{s=1}^{T}\hat{F}_s\left(\frac{1}{N}\sum_{i=1}^{N}E(\varepsilon_{is}\varepsilon_{it})\right)\right)'\right]\bar{\lambda}_{N,0} + o_P(1) \\
&= -\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{s,t=1}^{T}\hat{G}_t(u)\hat{F}_s'E(\varepsilon_{is}\varepsilon_{it})V_{NT}^{-1}H^{-1}\bar{\lambda}_{N,0} + o_P(1) \\
&= -\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{G}_t(u)\hat{F}_t'E(\varepsilon_{it}^2)V_{NT}^{-1}H^{-1}\bar{\lambda}_{N,0} \\
&\quad -\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{s\neq t}^{T}\hat{G}_t(u)\hat{F}_s'E(\varepsilon_{is}\varepsilon_{it})V_{NT}^{-1}H^{-1}\bar{\lambda}_{N,0} + o_P(1) \\
&= \hat{A}_{111}(u) + \hat{A}_{112}(u) + o_P(1), \text{ say.}
\end{aligned}
$$

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

$$
\begin{aligned}
\hat{A}_{111}(u) &= -\frac{1}{T}\sum_{t=1}^{T}\hat{G}_t(u)\hat{F}_tV_{NT}^{-1}H^{-1}\bar{\lambda}_{N,0}\left(\frac{1}{N}\sum_{i=1}^{N}\sigma_i^2\right) \\
&= -\left[\frac{1}{T}\sum_{t=1}^{T}\hat{F}_te^{\mathbf{i}u2\pi t/T}\hat{F}_t' - \left(\frac{1}{T}\sum_{t=1}^{T}\hat{F}_t\hat{F}_t'e^{\mathbf{i}u2\pi t/T}\right)\left(\frac{1}{T}\sum_{t=1}^{T}\hat{F}_t\hat{F}_t'\right)\right]V_{NT}^{-1}H^{-1}\bar{\lambda}_{N,0}\left(\frac{1}{N}\sum_{i=1}^{N}\sigma_i^2\right) \\
&= 0.
\end{aligned}
$$

Since $\nu < 1/2$, $\hat{F}_t - H'F_t = O_P(T^{-1})$. Then we can show

$$\frac{1}{T}\sum_{t=1}^{T}\hat{F}_t\hat{F}_t' e^{\mathbf{i}u2\pi t/T} = H'\left(\frac{1}{T}\sum_{t=1}^{T}F_tF_t'e^{\mathbf{i}u2\pi t/T}\right)H + O_P(T^{-1}),$$

and therefore it follows $\hat{G}_t(u) = G_t(u) + O_P(T^{-1/2})$. For $\hat{A}_{112}(u)$,

$$\begin{aligned}
\hat{A}_{112}(u) &= -\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{s\neq t}^{T}\hat{G}_t(u)\hat{F}_s' E(\varepsilon_{is}\varepsilon_{it})V_{NT}^{-1}H^{-1}\bar{\lambda}_{N,0} \\
&= -\frac{1}{T^2}\sum_{s\neq t}^{T}G_t(u)F_s'H_0\gamma_N(s,t)V_0^{-1}H_0^{-1}\bar{\lambda}_0 + o_P(1) \\
&= -\frac{1}{T^2}\sum_{s\neq t}^{T}H_0'F_tF_s'H_0M_t(u)\gamma_N(s,t)V_0^{-1}H_0^{-1}\bar{\lambda}_0 + o_P(1) \\
&= -\frac{2}{T^2}\sum_{j=1}^{T-1}\sum_{t=1}^{T-j}H_0'F_tF_{t+j}'H_0M_t(u)\gamma_N(t,t+j)V_0^{-1}H_0^{-1}\bar{\lambda}_0 + o_P(1) \\
&= -\frac{2}{T^2}\sum_{j=1}^{T-1}\sum_{t=1}^{T-j}H_0'E(F_tF_{t+j}')H_0M_t(u)\gamma_N(t,t+j)V_0^{-1}H_0^{-1}\bar{\lambda}_N + o_P(1) \\
&\quad -\frac{2}{T^2}\sum_{j=1}^{T-1}\sum_{t=1}^{T-j}H_0'\left[F_tF_{t+j}' - E(F_tF_{t+j}')\right]H_0M_t(u)\gamma_N(t,t+j)V_0^{-1}H_0^{-1}\bar{\lambda}_0 + o_P(1) \\
&= -\frac{2}{T}\sum_{j=1}^{T-1}H_0'F_tF_{t+j}'H_0\gamma_N(t,t+j)\left[\frac{1}{T}\sum_{t=1}^{T-j}M_t(u)\right]V_0^{-1}H_0^{-1}\bar{\lambda}_0 + o_P(1) \\
&= O_P(T^{-1}) * O_P(T^{-1/2}),
\end{aligned}$$

where we use the fact that

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

and

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

.

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

$$\hat{A}_1(u) = -\frac{2}{T}\sum_{j=1}^{T-1}H_0'F_tF_{t+j}'H_0\gamma_N(t,t+j)\left[\frac{1}{T}\sum_{t=1}^{T-j}M_t(u)\right]V_0^{-1}H_0^{-1}\bar{\lambda}_0 + o_P(1),$$

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

$$\hat{A}(u) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}G_t(u)B_i\varepsilon_{it} - \frac{2}{T}\sum_{j=1}^{T-1}H_0'F_tF_{t+j}'H_0\gamma_N(t,t+j)\left[\frac{1}{T}\sum_{t=1}^{T-j}M_t(u)\right]V_0^{-1}H_0^{-1}\bar{\lambda}_0 + o_P(1).$$

■

Proof of Proposition 2. Under $\mathbb{H}_A(a_{NT}) : \lambda_{it} = \lambda_{i0} + a_{NT} g_{it}$ we can show that

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

By Proposition 1, the first term is $O_P(T^{-3/2} + N^{-1/2}T^{-1/2})$. The second term is

$$\begin{aligned}&\frac{a_{NT}}{N}\sum_{i=1}^{N}\left[\frac{1}{T}\sum_{t=1}^{T}\hat{G}_t(u)F_t' g_{it}\right] \\ = &\frac{a_{NT}}{N}\sum_{i=1}^{N}\left[\frac{1}{T}\sum_{t=1}^{T}\hat{F}_t F_t' e^{\mathbf{i}u2\pi t/T} g_{it} - \left(\frac{1}{T}\sum_{t=1}^{T}\hat{F}_t\hat{F}_t' e^{\mathbf{i}u2\pi t/T}\right)\frac{1}{T}\sum_{t=1}^{T}\hat{F}_t F_t' g_{it}\right].\end{aligned}$$

For the first term, we have

$$\begin{aligned}&\frac{1}{T}\sum_{t=1}^{T}\hat{F}_t F_t' e^{\mathbf{i}u2\pi t/T} g_{it} \\ &= \frac{1}{T}\sum_{t=1}^{T}\left(\hat{F}_t F_t' - Q_0\right) e^{\mathbf{i}u2\pi t/T} g_{it} + \frac{1}{T}\sum_{t=1}^{T} Q_0 e^{\mathbf{i}u2\pi t/T} g_{it}\end{aligned}$$

By Lemma A.2(i), A.5, A.6, and Assumption A.5,we know $\frac{1}{T}\sum_{t=1}^{T}\left(\hat{F}_t F_t' - Q_0\right) e^{\mathbf{i}u2\pi t/T} g_{it} = O_P(C_{NT}^{-1})$. By law of large numbers, the second term

$$\frac{1}{T}\sum_{t=1}^{T} Q_0 e^{\mathbf{i}u2\pi t/T} g_{it} \Rightarrow Q_0 \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T} e^{\mathbf{i}u2\pi t/T} g_{it}.$$

Similarly, for the second term, we know $\frac{1}{T}\sum_{t=1}^{T}\hat{F}_t\hat{F}_t' e^{\mathbf{i}u2\pi t/T} = O_P(1)$, and

$$\frac{1}{T}\sum_{t=1}^{T}\hat{F}_t F_t' g_{it} = \frac{1}{T}\sum_{t=1}^{T}\left(\hat{F}_t - H'F_t\right)F_t' g_{it} + \frac{1}{T}\sum_{t=1}^{T} H'F_t F_t' g_{it}.$$

By Lemma A.4 and Assumption 5, we know the term $\frac{1}{T}\sum_{t=1}^{T}\left(\hat{F}_t - H'F_t\right)F_t' g_{it} = O_P(C_{NT}^{-2}) = o_P(1)$, and the term $\frac{1}{T}\sum_{t=1}^{T} H'F_t F_t' g_{it} = H'\Sigma_F \frac{1}{T}\sum_{t=1}^{T} g_{it} + o_P(1) = o_P(1)$.

Therefore, given $a_{NT} = N^{-1/2}T^{-1/2}$, and $\nu > 1/2$,

$$\begin{aligned}\sqrt{NT}\hat{A}_1(u) &\Rightarrow Q_0 \lim_{N,T\to\infty}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} e^{\mathbf{i}u2\pi t/T} g_{it} + o_P(1) \\ &= Q_0 \lim_{N,T\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T} e^{\mathbf{i}u2\pi t/T} g_{it} - \frac{1}{T}\sum_{t=1}^{T} e^{\mathbf{i}u2\pi t/T}\frac{1}{T}\sum_{t=1}^{T} g_{it}\right) + o_P(1).\end{aligned}$$

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

$$\sqrt{NT}\hat{A}_1(u) \Rightarrow Q_0 \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \widetilde{cov}(e^{\mathbf{i}u2\pi\tau}, g_i(\tau)),$$

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

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

By Lemma A.2, $\sqrt{NT}\hat{A}_2(u) \Rightarrow \mathcal{G}(u)$. Thus, we proved the result. ■

Proof of Theorem 1. Under $\mathbb{H}_0 : \lambda_{it} = \lambda_{i0}$, Proposition 1 shows

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

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

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

By Lemma A.7,

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

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

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

By Lemma A.5, 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 2. Under $\mathbb{H}_A(a_{NT}) : \lambda_{it} = \lambda_{i0} + a_{NT} g_{it}$,

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

By Proposition 1, we have

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} \hat{G}_t(u) F_t' \lambda_{i0}$$

$$= \quad -\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{G}_t(u)\varepsilon_{it}\lambda_{i0}'(\Lambda_0'\Lambda_0/N)^{-1}\bar{\lambda}_0 + O_P(T^{-3/2}) + o_P(1).$$

Given $a_{NT} = \frac{1}{\sqrt{NT}}$, by Proposition 2,

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

where

$$\psi(u) \quad = \quad Q_0 \lim_{N,T\to\infty}\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}e^{\mathbf{i}u2\pi t/T}g_{it} - \frac{1}{T}\sum_{t=1}^{T}e^{\mathbf{i}u2\pi t/T}\frac{1}{T}\sum_{t=1}^{T}g_{it}\right).$$

Given Theorem 1, it follows

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

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

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

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

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

■

References

Andrews, D. W. (1994) Empirical process methods in econometrics. *Handbook of econometrics* 4, 2247-2294.

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

Stewart, G.W. & J.G., Sun (1990) *Matrix Perturbation Theory.* Boston: Academic Press.

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 4. For cases (i)-(iii), the results of our test are based on Su and Wang's (2017) modified parametric bootstrap procedure, while for cases (iv) and (v), the results of our test are based on the MBB given in section 3.4.

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}		SW, MBB	
		5%	10%	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	5.6	10.6	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	4.6	9.6	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	6.0	10.0	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	4.2	9.4	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 terms: $\varepsilon_{it} = \sigma_i v_{it}$, $\sigma_i \sim i.i.d.U(0.5, 1.5), v_{it} \sim i.i.d.N(0,1)$																	
100	100	6.0	11.0	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	4.6	8.8	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	7.0	10.0	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	4.2	9.2	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-sectionally dependent error terms: $\varepsilon_{\cdot t} \sim i.i.d.N(0, \Sigma_\varepsilon)$																	
100	100	8.2	11.0	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	8.6	14.2	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	–	–
serially correlated error terms: $\varepsilon_{it} = 0.5\varepsilon_{it-1} + v_{it}, v_{it} \sim i.i.d.N(0,1)$																	
40	100	6.2	11.6	93.6	97.0	1.2	4.6	0.2	1.8	3.0	7.0	7.0	13.2	13.4	22.8	5.4	12.4
40	200	5.4	13.6	98.2	99.6	2.0	7.4	3.4	8.4	3.4	7.8	4.4	9.8	18.1	28.0	5.8	10.6
80	100	8.8	16.2	99.4	99.8	0.2	3.4	0.4	1.2	2.4	7.4	6.2	13.2	13.6	22.6	1.0	6.8
80	200	5.2	9.8	100	100	2.0	5.4	1.4	5.2	4.0	9.6	4.2	9.8	17.9	28.1	6.0	11.0
cross-sectionally dependent and serially correlated error terms: $\varepsilon_{\cdot t} = 0.5\varepsilon_{\cdot t-1} + v_{\cdot t}, v_{\cdot t} \sim i.i.d.N(0, \Sigma_v)$																	
40	100	6.2	12.8	87.0	95.2	1.4	4.6	0.6	2.6	3.8	10.0	5.8	11.0	13.6	22.6	7.0	15.6
40	200	6.4	12.2	93.8	97.2	2.4	6.6	2.8	7.0	3.0	8.2	4.8	10.4	18.0	28.2	6.6	13.6
80	100	8.8	14.2	98.6	99.4	1.4	2.6	0.6	1.4	2.4	7.8	6.0	11.6	13.7	23.0	3.4	8.0
80	200	6.2	11.0	99.0	99.8	2.0	4.0	1.8	6.0	4.0	10.8	5.6	10.8	17.9	28.1	4.2	10.6

Notes: See the notes in Table 3.

Table A.2 Power of tests under DGP.P1-P3 with serially uncorrelated error terms 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 terms: $\varepsilon_{it} \sim i.i.d.N(0,1)$													
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
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 terms: $\varepsilon_{it} = \sigma_i v_{it}$, $\sigma_i \sim i.i.d.U(0.5, 1.5), v_{it} \sim i.i.d.N(0,1)$													
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
DGP.P3	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-sectionally dependent error terms: $\varepsilon_{\cdot t} \sim i.i.d.N(0, \Sigma_\varepsilon)$													
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
DGP.P3	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

Notes: See the notes in Table 3.

Table A.3 Power of tests under DGP.P1-P3 with serially correlated error terms 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}		SW,MBB	
			5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
serially correlated error terms: $\varepsilon_{it} = 0.5\varepsilon_{it-1} + v_{it}, v_{it} \sim i.i.d.N(0,1)$																		
DGP.P1	40	100	47.6	60.6	94.4	96.8	1.0	4.8	1.0	2.4	2.0	6.0	8.0	16.2	17.4	27.6	14.0	26.4
	40	200	80.2	88.4	99.8	99.8	3.2	8.6	6.0	13.8	4.2	8.0	7.0	12.6	24.9	35.8	26.2	40.6
	80	100	81.6	90.0	99.6	100	0.8	4.4	0.8	2.8	2.2	8.6	6.0	14.4	16.8	26.7	9.2	23.6
	80	200	98.6	99.6	100	100	5.2	10.8	5.0	13.0	3.8	8.0	4.8	12.4	24.8	35.8	40.8	57.8
DGP.P2	40	100	12.0	20.4	89.0	92.8	1.4	4.0	0.6	2.2	1.8	6.8	7.6	11.8	14.4	24.0	6.6	15.6
	40	200	17.6	30.0	98.6	99.4	2.2	6.6	3.2	7.0	3.2	8.0	4.0	8.2	18.5	28.9	6.0	11.4
	80	100	14.4	25.4	98.2	99.4	0.4	4.0	0.6	1.6	2.0	6.6	5.6	11.2	13.8	22.9	2.6	6.0
	80	200	31.0	45.8	99.8	100	2.2	7.0	1.2	6.8	3.8	9.6	3.0	10.0	18.5	28.6	6.2	13.8
DGP.P3	40	100	30.0	44.6	93.2	96.4	1.4	4.0	1.0	2.4	2.6	7.4	8.6	13.8	15.6	25.2	8.0	18.2
	40	200	65.4	78.8	99.6	100	2.6	7.0	3.4	9.6	2.6	7.8	6.0	12.0	20.5	31.1	16.4	26.6
	80	100	52.4	66.2	99.2	99.6	0.2	2.4	0.4	1.6	2.0	6.4	6.4	15.0	15.1	24.5	5.6	12.6
	80	200	91.4	97.2	100	100	1.6	4.8	2.2	9.4	3.4	7.0	5.6	11.0	20.8	31.7	21.8	35.2
cross-sectionally dependent and serially correlated error terms: $\varepsilon_{\cdot t} = 0.5\varepsilon_{\cdot t-1} + v_{\cdot t}, v_{\cdot t} \sim i.i.d.N(0,\Sigma_v)$																		
DGP.P1	40	100	17.6	28.2	100	100	0.0	0.0	0.0	0.0	0.4	1.4	28.2	39.4	24.2	37.3	51.8	69.4
	40	200	25.4	37.8	100	100	0.0	0.2	0.6	2.6	1.8	6.8	12.6	22.4	34.0	47.7	26.8	46.6
	80	100	41.6	57.4	99.4	99.6	0.0	0.4	0.0	0.6	0.4	3.8	21.2	30.6	24.8	37.5	39.6	56.6
	80	200	67.2	76.8	100	100	1.0	5.8	3.8	9.6	5.4	12.0	13.6	24.0	32.0	45.0	19.2	37.4
DGP.P2	40	100	8.2	15.6	100	100	0.2	0.2	0.0	0.2	0.0	0.8	24.2	33.6	22.4	35.1	51.6	71.2
	40	200	9.8	18.2	100	100	0.0	0.4	0.2	0.4	1.6	4.2	8.4	15.2	29.7	43.2	26.2	47.4
	80	100	12.0	21.8	100	100	0.0	0.6	0.4	0.4	0.6	3.0	16.0	25.0	22.6	35.1	34.8	55.4
	80	200	16.2	25.6	99.8	99.8	0.6	3.0	0.6	3.4	2.6	6.8	7.8	15.2	26.7	39.1	12.0	24.6
DGP.P3	40	100	10.4	18.4	100	100	0.0	0.0	0.0	0.2	0.2	1.4	27.8	35.4	22.0	35.2	46.0	67.0
	40	200	11.6	20.6	100	100	0.0	0.2	0.0	0.4	1.2	4.6	10.8	18.4	30.6	44.1	23.2	41.6
	80	100	18.6	30.0	99.6	100	0.0	0.2	0.4	0.4	0.8	3.2	18.4	28.4	23.1	36.2	36.6	53.8
	80	200	27.6	38.2	100	100	0.0	0.8	0.6	1.6	1.6	5.4	7.8	14.8	27.8	40.7	10.4	24.8

Notes: See the notes in Table 3.